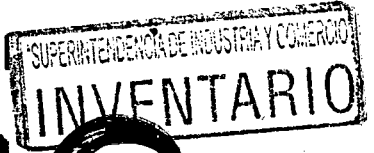




Industria y Comercio
SUPERINTENDENCIA

PGT

Delegatura de Propiedad Industrial
División de Nuevas Creaciones



10-151612

SOLICITUD

PATENTE DE INVENCION



21. EXPEDIENTE No. _____

54. TÍTULO **COMPOSICIONES Y MÉTODOS QUE COMPRENDEN
VARIANTES DE CELULASA CON AFINIDAD REDUCIDA PARA
MATERIALES NO CELULOSICOS.**

51. CLASIFICACIÓN INTERNACIONAL _____

71. SOLICITANTE **DANISCO US INC.**

DOMICILIO **925 Page Mill Road, Palo Alto, California 94304, E. U. A.**

74. APODERADO **JERÓNIMO CASTILLO BONILLA**

22. BOGOTÁ, D.C., _____

En
06-01-2011

PETITORIO			
 Industria y Comercio SUPERINTENDENCIA		SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO  No. 10-151612- -00000-0000 Fecha: 2010-12-01 17:17:51 Dep. 2020 DIR.NUEVASCR Tra. 11 - PATENTEIYII Eve: 378 FASENACIONALI Act. 411 PRESENTACION Folios: 190	
FORMULARIO ÚNICO DE SOLICITUD DE PATENTES PCT ENTRADA EN FASE NACIONAL			
1		SOLICITUD DE:	
☒ Patente de Invención		☐ Patente de Modelo de Utilidad	
☒ Capítulo I		☐ Capítulo II	
2 SOLICITANTE		IDENTIFICACIÓN	
Nombre: DANISCO US INC. Dirección: 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de America Nacionalidad o Domicilio: Estadounidense Lugar de Constitución: Delaware Teléfono: Fax: E-mail: 		☐ C.C. ☐ NIT ☐ C.E. ☐ Otro Cuál Número 	
3 REPRESENTANTE O APODERADO		IDENTIFICACIÓN	
Nombre: JERÓNIMO CASTILLO BONILLA Dirección: Calle 93A No. 14-17 Oficina 311 Bogotá, D.C. Teléfono: 236 34 93 Fax: 218 23 51 E-mail: info@pcastillopineda.com		☒ C.C. ☐ NIT ☐ C.E. ☐ Otro Cuál Número 79.462.057 TP 104167	
En caso de haber presentado poder general para asuntos que se adelanten ante la Delegatura de Propiedad Industrial, sírvase indicar el numero de radicación (del poder). _____			
4 INVENTOR (ES) (72)		Nombre: LUIS G. CASCAO - PEREIRA Dirección: 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de America Nacionalidad o Domicilio: Brasileño	
5 PCT (86) Solicitud Internacional No. <u>PCT/US2009/046159</u> Fecha <u>Junio 3, 2009</u> Publicación internacional No. <u>WO 2009/149202 A2</u> Fecha <u>Diciembre 10, 2009</u>			
6 Título (54) <u>COMPOSICIONES Y METODOS QUE COMPRENDEN VARIANTES DE CELULASA</u> <u>CON AFINIDAD REDUCIDA PARA MATERIALES NO CELULOSICOS</u>			
7 Clasificación internacional (51)			
8 Prioridad ☒ SI ☐ NO		(33)País de Origen	(32) Fecha
		US	Junio 6, 2008
9 Comprobante de pago No. _____ Fecha _____			

(Continuación Inventores)

Nombre : Thijs KAPER
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América.
Nacionalidad : holandesa

Nombre : Bradley R. KELEMEN
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América.
Nacionalidad : estadounidense

Nombre : Amy LIU
Dirección : 925 Page Mill Road, Palo Alto, California 94304, Estados Unidos de América.
Nacionalidad : Canadiense

State of California

SECRETARY OF STATE

Requested for use in Colombia.

Not for use within the United States of America.

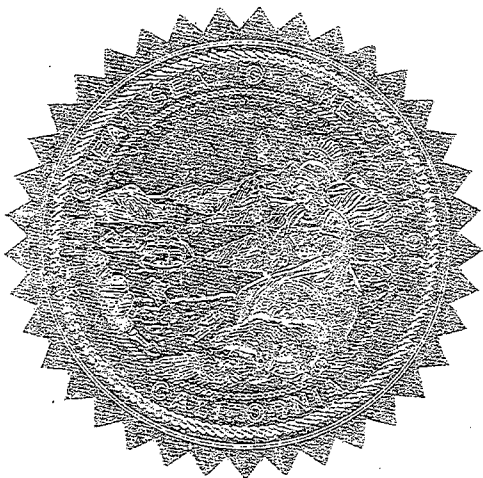
The purpose of the Apostille is to certify the authenticity of the signature of the official signing the document, the capacity in which the official signing the document has acted, and, where appropriate, the identity of the seal or stamp

APOSTILLE

(Convention de La Haye du 5 octobre 1961)

1. *Country: United States of America*
This public document
2. *has been signed by Carol A. See*
3. *acting in the capacity of Notary Public, State of California*
4. *bears the seal/stamp of Carol A. See, Notary Public.*
State of California
5. *At Sacramento California*
6. *the 1st day of September 2009*
7. *by Deputy Secretary of State, State of California*
8. *No. 723031*
9. *Seal/Stamp:*

CERTIFIED



10. *Signature*

John Bowen
Secretary of State

BY *[Signature]*



LAW OFFICES
PEDRO CASTILLO PINEDA &
ASSOCIATES
P.O. BOX 48362
Bogotá, Colombia / South America

IMPORTANT NOTE

The lodging of this GENERAL POWER with the Patent Office obviates the need of presenting special powers with each individual trade mark application.

FOR PATENTS AND TRADE MARKS IN
COLOMBIA

Poder General

General Power of Attorney

1 Applicant's (1)
name

(1) Danisco US Inc.

2 Applicant's domiciliado en (2)
address

domiciled in (2)
925 Page Mill Road, Palo Alto
California 94304
U.S.A.

3 Date
4 Applicant's
signature

nombramos al Dr. JERONIMO CASTILLO BONILLA y/o Dra. XIMENA BONILLA PEREIRA y/o Dr. ALVARO CASTILLO GRAU Calle 93A N° 14-17 Oficina 311 en Bogotá, Colombia, como sus representantes con poder especial, amplio y suficiente, para recabar de las oficinas y autoridades Colombianas, administrativas, judiciales o de policía, en primera o segunda instancia, la obtención de Patentes de Invención, Registro de Marcas, de Modelos y de Dibujos Industriales, de Nombres Comerciales y Enseñas, Lemas Comerciales y Denominaciones de Origen, lo mismo que trasposos, extensiones, renovaciones, cambios de nombre, registro de licencias de explotación y registro de licencias de uso referentes a tales actuaciones, cancelación de Registros de marca, a cuyo efecto les faculta para dar ante dichas autoridades todos los pasos necesarios al objeto indicado, elevar solicitudes, formular descripciones, enmiendas, protestas, declaraciones, apelaciones y reclamos, solicitar la aplicación de medidas cautelares; solicitar la nulidad de Patentes de Invención, modelos de utilidad, diseños industriales, así como la de registros de marcas ante la autoridad nacional competente; aceptar en nuestro nombre y representación la Cesión de derechos de invención y de patente de invención que se nos haga por cualquier persona, abonar todos los impuestos, cuotas y pagos determinados por la ley, recibir todos los documentos y valores dando el descargo respectivo, llenar cualesquiera otros requisitos, desistir y tomar en fin todas las medidas que creyeren conducentes al resguardo de nuestros intereses y en caso de presentarse oposiciones y/o observaciones para que intervengan como demandantes o como demandados, con todas las demás facultades necesarias, así como iniciar y tramitar procesos jurisdiccionales y/o administrativos relativos a competencia desleal. Este poder comprende además las facultades de recibir, desistir, transigir, comprometer, sustituir, revocar sustituciones, conciliar, renunciar, recibir notificaciones y nombrar apoderados judiciales o extrajudiciales.
Dado y firmado en (3)

hereby appoint, JERONIMO CASTILLO BONILLA and/or XIMENA BONILLA PEREIRA and/or ALVARO CASTILLO GRAU, Calle 93A N° 14-17 Office 311 en Bogotá, Colombia, as our representatives with special, ample and sufficient power to obtain from the offices and Colombian authorities, administrative, judicial and police, in first or in second instance, Privileged Patents of invention, Registration of Trade Marks, of Models and Industrial Drawings, of Commercial Names and ensigns, Commercial slogans and Origin Denominations as well as assignments, extensions, renewals, changes of name, registration of license of exploitation and of license of use referred to said actions, Cancellation Actions against trademark Registrations, who are empowered to take before said authorities all of the necessary steps for the purposes indicated, to file applications, to formulate descriptions, corrections, protests, declarations, appeals, and claims, apply for precautionary measures; apply for the nullity of Patent Inventions, Utility Models, Industrial Designs, as well as for Trademark Registrations before the competent national authority; to accept on our behalf and representation the assignment of rights of inventions and titles of invention that may be done to us by any person, to pay all taxes, quotas and payments prescribed by law, to receive all the documents and proceeds giving the respective acquittance or receipt, to fulfill any other requisites to desist and take in fact any other measures that may be deemed conducive to the safeguard of our interests, and in case that oppositions and/or observations are presented to intervene as complainant and defendant, with all necessary legal faculties, as well as to file jurisdictional processes and/or administrative related to disloyal competition.

This power of attorney includes the faculties to receive, desist, compromise, transact substitute it in all or in part, to revoke substitutions, to conciliate, to renounce, to receive notifications and to appoint extra-judicial or judicial attorneys.

Given and signed in (3)
Palo Alto, California, U.S.A.
27th day of August, 2009



4) Signature (s)
X *Soonhee Jang*
/ Soonhee Jang

DECLARACION NOTARIAL
(Individual)

A los días del mes de
 compareció y/o comparecieron

personalmente ante mi, Notario Público

quien (es) doy fe de conocer y me consta que es
(son) la(s) persona(s) descrita(s) y firmante(s) del
documento que precede, declarando ante mi que
otorga(n) el mismo para los fines y propósitos que
en él se expresan.

Notario Público,

(Sociedades)

El suscrito Notario Certifica: que la firma que
antecede y dice

es auténtica; que el signatario desempeña
actualmente el cargo de

de,

que tiene facultad para otorgar este poder; y que
dicha compañía esta domiciliada en

y actualmente existe y esta organizada de acuerdo
a las leyes de

Todos los hechos anteriores le constan por haber
examinado los documentos respectivos y de todo
ello da fe.

Dado y firmado en

El Notario Público,

(Con legalización Consular)

NOTARIAL CERTIFICATE
(Individual)

On this day of of
personally appeared before me

a Notary Public

to me known, and known to me to be the
person(s) described in and who executed the
foregoing document, and he (they) duly
acknowledged to me that he (they) executed
same for the uses and purposes therein expresed.

Notary Public.

(Corporations)

The undersigned Notary Public certifies: that the
preceding signature reading
Soonhee Jang

is authentic; that the signer at present fills the post
of Vice President, IP Strategy & Chief IP Counsel

of Danisco US Inc.

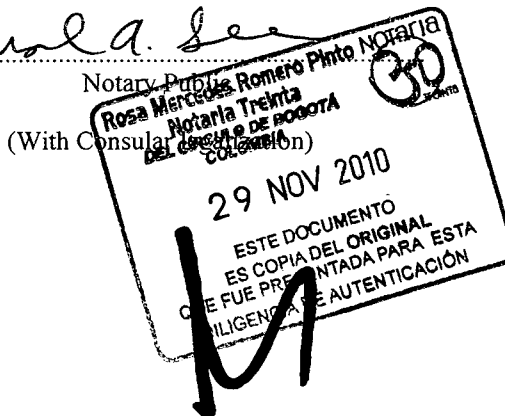
and that she is empowered to grant this power; and
that said company is domiciled in
Palo Alto, California, U.S.A.

and actually exists and is organized in accordance
with the laws of Delaware, USA

All the foregoing facts are known to the Notary
due to her having examined the respective
documents to which she, the Notary attests..

Given and signed in Palo Alto, California, USA
Carol A. See, Notary Public

Carol A. See



IN WITNESS WHEREOF I undersign as follows:

Date: August 27, 2009

Soonhee Jang
Soonhee Jang,
Vice President, IP Strategy &
Chief IP Counsel

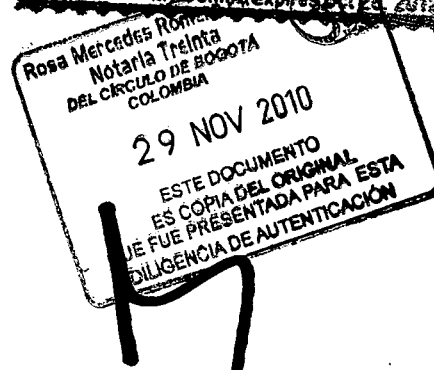
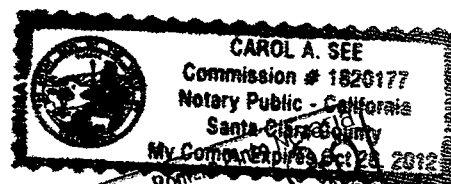
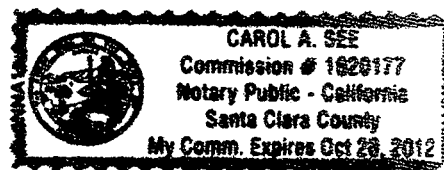
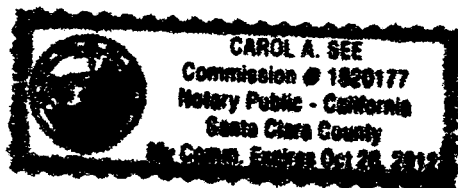
STATE OF CALIFORNIA
COUNTY OF SANTA CLARA

On 8/27/09, before me, Carol A. See, Notary Public, personally appeared, Soonhee Jang, who proved to me on the basis of satisfactory evidence to be the person(s) whose name(s) is/are subscribed to the within instrument and acknowledged to me that he/she/they executed the same in his/her/their authorized capacity(ies), and that by his/her/their signature(s) on the instrument the person(s), or the entity upon behalf of which the person(s) acted, executed the instrument. I certify under PENALTY OF PERJURY under the laws of the State of California that the foregoing paragraph is true and correct.

WITNESS my hand and official seal.

Signature Carol A. See

(Seal)



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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 December 2009 (10.12.2009)



PCT



(10) International Publication Number
WO 2009/149202 A2

(51) International Patent Classification:
C12N 9/42 (2006.01)

(21) International Application Number:

PCT/US2009/046159

(22) International Filing Date:

3 June 2009 (03.06.2009)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/059,506

6 June 2008 (06.06.2008)

US

(71) Applicant (for all designated States except US): **DANISCO US INC., GENENCOR DIVISION** [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CASCAO-PEREIRA, Luis, G.** [BR/US]; Danisco US Inc., Genencor Division, 925 Page Mill Road, Palo Alto, CA 94304 (US). **KAPER, Thijs** [NL/US]; Danisco US Inc., Genencor Division, 925 Page Mill Road, Palo Alto, CA 94304 (US). **KELEMEN, Bradley, R.** [US/US]; Danisco US Inc., Genencor Division, 925 Page Mill Road, Palo Alto, CA 94304 (US). **LIU, Amy** [CA/US]; Danisco US Inc., Genencor Division, 925 Page Mill Road, Palo Alto, CA 94304 (US).

(74) Agent: **KOLMAN, Michael, F.**; Danisco US Inc., Genencor Division, 925 Page Mill Road, Palo Alto, CA 94304 (US).

(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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

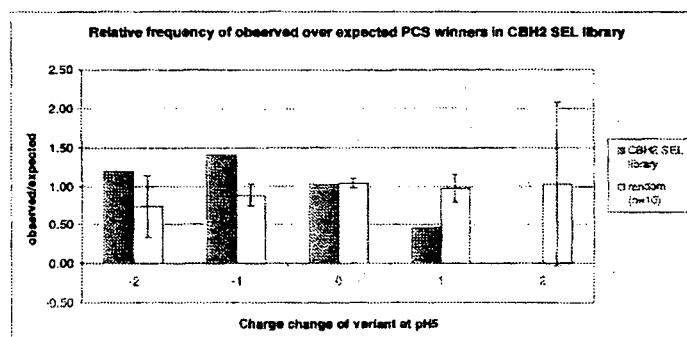
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT,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

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: COMPOSITIONS AND METHODS COMPRISING CELLULASE VARIANTS WITH REDUCED AFFINITY TO NON-CELLULOSIC MATERIALS

Figure 4



(57) Abstract: The present disclosure relates to cellulase variants. In particular the present disclosure relates to cellulase variants having reduced binding to non-cellulosic materials. Also described are nucleic acids encoding the cellulase, compositions comprising said cellulase, methods of identifying cellulase variants and methods of using the compositions.

WO 2009/149202 A2

COMPOSITIONS AND METHODS COMPRISING CELLULASE VARIANTS WITH REDUCED AFFINITY TO NON-CELLULOSIC MATERIALS

I. STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY 5 SPONSORED RESEARCH AND DEVELOPMENT

[01] This invention was made with Government support under conditional award no: DE-FC36-08GO18078 awarded by the Department of Energy. The Government has certain rights in this invention.

10 II. CROSS-REFERENCE TO RELATED APPLICATION

[02] This application claims the benefit of U.S. Provisional Application No. 61/059,506, filed June 6, 2008, which is incorporated herein by reference.

III. FIELD

15 [03] The present disclosure relates to enzymes and in particular cellulase variants. Also described are nucleic acids encoding the cellulase variants, compositions comprising the cellulase variants, methods of identifying additional useful cellulase variants and methods of using the compositions.

20 IV. BACKGROUND

[04] Cellulose and hemicellulose are the most abundant plant materials produced by photosynthesis. They can be degraded and used as an energy source by numerous microorganisms (*e.g.*, bacteria, yeast and fungi) that produce extracellular enzymes capable of hydrolysis of the polymeric substrates to monomeric sugars (Aro et al., J Biol Chem, 276: 24309-24314, 2001). As the limits of non-renewable resources approach, the potential of cellulose to become a major renewable energy resource is enormous (Krishna et al., Bioresource Tech, 77: 193-196, 2001). The effective utilization of cellulose through biological processes is one approach to overcoming the shortage of foods, feeds, and fuels (Ohmiya et al., Biotechnol Gen Engineer Rev, 14: 365-414, 1997).

25 [05] Cellulases are enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellooligosaccharides, and the like. Cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-

30

glucosidases ([beta]-D-glucoside glucohydrolase; EC 3.2.1.21) ("BG"). (Knowles et al., TIBTECH 5: 255-261, 1987; and Schulein, Methods Enzymol, 160: 234-243, 1988).

Endoglucanases act mainly on the amorphous parts of the cellulose fibre, whereas cellobiohydrolases are also able to degrade crystalline cellulose (Nevalainen and Penttila, Mycota, 303-319, 1995). Thus, the presence of a cellobiohydrolase in a cellulase system is required for efficient solubilization of crystalline cellulose (Suurnakki et al., Cellulose 7: 189-209, 2000). Beta-glucosidase acts to liberate D-glucose units from cellobiose, cello-oligosaccharides, and other glucosides (Freer, J Biol Chem, 268: 9337-9342, 1993).

[06] Cellulases are known to be produced by a large number of bacteria, yeast and fungi. Certain fungi produce a complete cellulase system capable of degrading crystalline forms of cellulose, such that the cellulases are readily produced in large quantities via fermentation. Filamentous fungi play a special role since many yeast, such as *Saccharomyces cerevisiae*, lack the ability to hydrolyze cellulose (See, e.g., Wood et al., Methods in Enzymology, 160: 87-116, 1988).

[07] The fungal cellulase classifications of CBH, EG and BG can be further expanded to include multiple components within each classification. For example, multiple CBHs, EGs and BGs have been isolated from a variety of fungal sources including *Trichoderma reesei* (also referred to as *Hypocrea jecorina*), which contains known genes for two CBHs, i.e., CBH I ("CBH1") and CBH II ("CBH2"), at least 8 EGs, i.e., EG I, EG II, EG III, EGIV, EGV, EGVII and EGVIII, and at least 5 BGs, i.e., BG1, BG2, BG3, BG4 and BG5. EGIV, EGVII and EGVIII also have xyloglucanase activity.

[08] In order to efficiently convert crystalline cellulose to glucose the complete cellulase system comprising components from each of the CBH, EG and BG classifications is required, with isolated components less effective in hydrolyzing crystalline cellulose (Filho et al., Can J Microbiol, 42:1-5, 1996). A synergistic relationship has been observed between cellulase components from different classifications. In particular, the EG-type cellulases and CBH-type cellulases synergistically interact to more efficiently degrade cellulose.

[09] Cellulases are known in the art to be useful in the treatment of textiles for the purposes of enhancing the cleaning ability of detergent compositions, for use as a softening agent, for improving the feel and appearance of cotton fabrics, and the like (Kumar et al., Textile Chemist and Colorist, 29:37-42, 1997). Cellulase-containing detergent compositions with improved cleaning performance (U.S. Pat. No. 4,435,307; GB App. Nos. 2,095,275 and 2,094,826) and for use in the treatment of fabric to improve the feel and appearance of the textile (U.S. Pat. Nos. 5,648,263, 5,691,178, and 5,776,757; and GB App. No. 1,358,599),

have been described. Hence, cellulases produced in fungi and bacteria have received significant attention. In particular, fermentation of *Trichoderma* spp. (e.g., *Trichoderma longibrachiatum* or *Trichoderma reesei*) has been shown to produce a complete cellulase system capable of degrading crystalline forms of cellulose.

- 5 [10] Although cellulase compositions have been previously described, there remains a need for new and improved cellulase compositions. Improved cellulase compositions find used in household detergents, textile treatments, biomass conversion and paper manufacturing. Cellulases that exhibit improved performance are of particular interest.

10 V. SUMMARY

- [11] The present teachings relates to cellulase variants modified to reduce binding to non-cellulosic materials. In general, the cellulase variants have increased cellulolytic activity in the presence of non-cellulosic materials in comparison to wild type cellulases. In some embodiments the cellulase variants have a decreased net charge (i.e. is more negative) in
15 comparison to wild type cellulases. In some embodiments, the cellulase variants are less positively charged than wild type cellulases. In some embodiments, a cellulase is modified by removing one or more positive charges. In some embodiments, a cellulase is modified by adding one or more negative charges. In some embodiments, a cellulase is modified by removing one or more positive charges and adding one or more negative charges.
- 20 [12] In some embodiments, the present teachings relate to cellobiohydrolase I (CBH1) or cellobiohydrolase II (CBH2) variants. In some embodiments the cellulase variant is a mature form having cellulase activity and a substitution at one or more positions selected from the group consisting of 63, 77, 129, 147, 153, 157, 161, 194, 197, 203, 237, 239, 247, 254, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, and 382, wherein the positions are numbered by
25 correspondence to a reference (e.g., wild type *Hypocrea jecorina* CBH2) cellulase having the amino acid sequence of SEQ ID NO:3, and wherein the substitution at one or more positions causes the cellulase variant to have a more negative net charge in comparison to the reference cellulase. In some embodiments, CBH2 is modified by removing one or more positive charges, which in some embodiments entails a replacement of a lysine or an arginine with a
30 neutral amino acid (e.g., K or R replaced by N or Q or other neutral residue). In some embodiments, CBH2 is modified by adding one or more negative charges, which in some embodiments entails a replacement of a neutral amino acid with a negatively charged amino acid (e.g., N or Q or other neutral residue replaced by D or E). In some embodiments, CBH2 is modified by removing one or more positive charges and adding one or more

negative charges, which in some embodiments entails a replacement of a lysine or an arginine with a negatively charged amino acid (e.g., K or R replaced by D or E). In general, the CBH2 variant has increased cellulolytic activity in the presence of lignin in comparison to the wild type *Hypocrea jecorina* CBH2 having the amino acid sequence of SEQ ID NO:3. The present teachings further provide CBH2 variants comprising one or more substitutions selected from the group consisting of K129E, K157E, K194E, K288E, K327E, K356E, R63Q, R77Q, R153Q, R203Q, R294Q, R378Q, N161D, N197D, N237D, N247D, N254D, N285D, N289D, N339D, N344D, N382D, Q147E, Q204E, Q239E, Q281E, D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, and E244Q, in the mature form of CBH2, wherein said substitutions are numbered according to the mature form of *Hypocrea jecorina* CBH2 of SEQ ID NO:3. In some embodiments, the variant comprises a further substitution at one or more further positions selected from the group consisting of 146, 151, 189, 208, 211, 244, 277 and 405, wherein the further positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3. In some embodiments, the further substitution at one or more further positions comprises a replacement of aspartic acid or glutamic acid with a neutral amino acid (e.g., D or E replaced by N or Q or other neutral residue). In some embodiments, the further substitution at one or more further positions comprises one or more of the group consisting of D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, and E244Q, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3. In some preferred embodiments, the substitution at one or more positions is selected from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 positions. In some preferred embodiments, the cellulase variant is derived from a parent cellulase selected from the group consisting of *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, and *Cellulomonas fimi* CenA EG. In some preferred embodiments, the cellulase variant is derived from a parent cellulase whose amino acid sequence is at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a member of the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, and SEQ ID NO:13. In some embodiments, the more negative net charge is a -1 or -2 in comparison to the reference CBH2.

[13] The present disclosure further provides cellulase variants, wherein the variant is a mature form having cellulase activity and comprising a chemical modification of a lysine residue to remove positive charge of the lysine residue. In some preferred embodiments, the chemical modification comprises a treatment with a compound selected from the group consisting of succinic anhydride, acetoxysuccinic anhydride, maleic anhydride, tartaric anhydride, phthalic anhydride, trimellitic anhydride, cis-aconitic anhydride, t-nitrophthalic anhydride, acetic anhydride, butyric anhydride, isobutyric anhydride, hexanoic anhydride, valeric anhydride, isovaleric anhydride, and pivalic anhydride. In some preferred embodiments, the cellulase variant is derived from a parent cellulase selected from the group consisting of a *Hypocrea jecorina* cellobiohydrolase I, *Hypocrea jecorina* cellobiohydrolase II, *Hypocrea jecorina* endoglucanase I, *Hypocrea jecorina* endoglucanase II, and *Hypocrea jecorina* beta-glucosidase. In some preferred embodiments, the cellulase variant is derived from a parent cellulase selected from the group consisting of *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, and *Cellulomonas fimi* CenA EG. Also provided are cellulase variants derived from a parent cellulase whose amino acid sequence is at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a member of the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, and SEQ ID NO:13. In some embodiments, the cellulase variant comprises a substitution at one or more positions selected from the group consisting of 63, 77, 129, 147, 153, 157, 161, 194, 197, 203, 237, 239, 247, 254, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, and 382, wherein the positions are numbered by correspondence with the amino acid sequence of a reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

[14] The present teachings further relates to CBH2 variant comprising from one to twenty six substitutions selected from the group consisting of K129E, K157E, K194E, K288E, K327E, K356E, R63Q, R77Q, R153Q, R203Q, R294Q, R378Q, N161D, N197D, N237D, N247D, N254D, N285D, N289D, N339D, N344D, N382D, Q147E, Q204E, Q239E, and Q281E. In some embodiments, the CBH2 variant comprises a combination of substitutions selected from the group consisting of: i) K157E / K129E; ii) K157E / K129E / K288E / K194E; iii) K157E / K129E / K288E / K194E / K356E / K327E; iv) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q; v) K157E / K129E / K288E /

K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q / N382D / N344D / N327D / N339D; vi) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q / N382D / N344D / N327D / N339D / N289D / N161D / Q204E / Q147E; vii) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q N382D / N344D / N327D / N339D / N289D / N161D / Q204E / Q147E / N285D / N197D / N254D / N247D; and viii) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q N382D / N344D / N327D / N339D / N289D / N161D / Q204E / Q147E / N285D / N197D / N254D / N247D / Q239E / Q281E / R63Q / R77Q.

[15] In some embodiments, the CBH2 variant comprises from one to eight substitutions selected from the group consisting of D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, and E244Q. In some embodiments, the CBH2 variant comprises a combination of substitutions selected from the group consisting: i) D189N / E208Q / D211N / D405; and ii) D189N / E208Q / D211N / D405 / E244Q / D277N / D151 / E146Q.

[16] Also described are isolated nucleic acids encoding a CBH2 variant having cellobiohydrolase activity as described in the preceding paragraphs. In a first aspect, the disclosure encompasses an isolated nucleic acid encoding a polypeptide having cellobiohydrolase activity, which polypeptide is a variant of a glycosyl hydrolase of family 6, and wherein said nucleic acid encodes a substitution at a residue which decreases the net charge in comparison to the wild type *Hypocrea jecorina* CBH2.

[17] In another aspect, the disclosure is directed to an isolated nucleic acid encoding a CBH2 variant, wherein said variant comprises a substitution at a position selected from the group consisting of K129E, K157E, K194E, K288E, K327E, K356E, R63Q, R77Q, R153Q, R203Q, R294Q, R378Q, N161D, N197D, N237D, N247D, N254D, N285D, N289D, N339D, N344D, N382D, Q147E, Q204E, Q239E, Q281E, D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, and E244Q, in the mature form of CBH2, wherein said substitutions are numbered according to the mature form of *Hypocrea jecorina* CBH2 of SEQ ID NO:3.

[18] In some embodiments, the disclosure is directed to an expression cassette comprising a nucleic acid encoding a CBH2 variant, a constructs comprising the nucleic acid of encoding the CBH2 variant operably linked to a regulatory sequence, a vector comprising a nucleic acid encoding a CBH2 variant, and host cell transformed with the vector comprising a nucleic acid encoding a CBH2 variant. The present teachings further provide methods producing a CBH2 variant by culturing the host cells expressing a CBH2 variant in a culture medium under suitable conditions to produce the CBH2 variant.

[19] Also provided are compositions comprising the cellulase variant of the preceding paragraphs. In some preferred embodiments, the composition further comprises at least one additional enzyme selected from the group consisting of a subtilisin, a neutral metalloprotease, a lipase, a cutinase, an amylase, a carbohydrase, a pectinase, a mannanase, an arabinase, a galactanase, a xylanase, an oxidase, and a peroxidase

[20] Provided herein, are methods of converting biomass to sugars comprising contacting said biomass with a cellulase variant. Also provided are methods of producing a fuel by contacting a biomass composition with an enzymatic composition comprising the cellulase variant to yield a sugar solution and culturing with a fermentative microorganism under conditions sufficient to produce a fuel.

[21] Also provided are compositions comprising cellulase variants including detergent compositions, feed additives for example, and methods of cleaning or fabric care by contacting a surface and/or an article comprising a fabric with the detergent composition. Also, provided are methods of fabric care treatment, including depilling and surface finishing, by contacting a surface and/or an article comprising a fabric with a cellulase variant.

[22] Other objects, features and advantages of the present disclosure will become apparent from the following detailed description. It should be understood, however, that the detailed description and specific examples, while indicating preferred embodiments of the disclosure, are given by way of illustration only, since various changes and modifications within the scope and spirit of the disclosure will become apparent to one skilled in the art from this detailed description.

VI. BRIEF DESCRIPTION OF THE DRAWINGS

[23] FIG. 1 illustrates saccharification of APB by modified (squares) and unmodified (circles) *Trichoderma* sp. cellulase preparations in the presence of increasing amounts of lignin inhibitor. FIG. 1A and FIG. 1B shows results after 24 and 48 hour incubations, respectively.

[24] FIG. 2A illustrates saccharification comparing modified cellulases and FIG. 2B shows the difference of saccharification using modified and unmodified cellulases.

[25] FIG. 3 provides an alignment of the amino acid sequences of the mature form of various cellulases: *Hypocrea jecorina* (also known as *T. reesei*) CBH2 (SEQ ID NO:3), *Hypocrea koningii* CBH2 (SEQ ID NO:4), *Humicola insolens* CBH2 (SEQ ID NO:5), *Acremonium cellulolyticus* CBH2 (SEQ ID NO:6), *Agaricus bisporus* CBH2 (SEQ ID NO:7),

Fusarium osysporum EG (SEQ ID NO:8), *Phanerochaete chrysosporium* CBH2 (SEQ ID NO:9), *Talaromyces emersonii* CBH2 (SEQ ID NO:10), *Thermobifida fusca* 6B/E3 CBH2 (SEQ ID NO:11), *Thermobifida fusca* 6A/E2 EG (SEQ ID NO:12), and *Cellulomonas fimi* CenA EG (SEQ ID NO:13).

- 5 [26] FIG. 4 provides a graph of the relative frequency of observed over expected pretreated corn stover (PCS) assay winners of the CBH2 variant SELs as a product of charge change. Decreasing CBH2 charge results in a significantly higher frequency of PCS winners.
- [27] FIG. 5 provides a plasmid map of pTTTpyr-cbh2.

10 VII. DETAILED DESCRIPTION OF VARIOUS EMBODIMENTS

- [28] The present teachings relates to cellulase variants modified to reduce binding to non-cellulosic materials. In general, the cellulase variant has increased cellulolytic activity in the presence of non-cellulosic materials in comparison to the wild type cellulase. In some embodiments the variant cellulase has a decreased net charge (*i.e.* is more negative) in
- 15 comparison to the wild type cellulase. In some embodiments, the cellulase variants are less positively charged than wild type cellulase. In some embodiments, a cellulase is modified by removing one or more positive charges. In some embodiments, a cellulase is modified by adding one or more negative charges. In some embodiments, a cellulase is modified by removing one or more positive charges and adding one or more negative charges.
- 20 [29] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the compositions and methods described herein. Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. In this application, the use of the
- 25 singular includes the plural unless specifically stated otherwise. The use of "or" means "and/or" unless stated otherwise. Likewise, the terms "comprise," "comprising," "comprises," "include," "including" and "includes" are not intended to be limiting. All patents and publications, including all amino acid and nucleotide sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference. The
- 30 headings provided herein are not limitations of the various aspects or embodiments of the disclosure which can be had by reference to the specification as a whole. Accordingly, the terms herein are more fully defined by reference to the specification as a whole.
- [30] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this

disclosure belongs. Singleton, et al., **DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY**, 2D ED., John Wiley and Sons, New York (1994), and Hale & Marham, **THE HARPER COLLINS DICTIONARY OF BIOLOGY**, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this disclosure. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, the preferred methods and materials are described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxyl orientation, respectively.

Practitioners are particularly directed to Sambrook et al., **MOLECULAR CLONING: A LABORATORY MANUAL** (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989, and Ausubel FM et al., **Current Protocols in Molecular Biology**, John Wiley & Sons, New York, N.Y., 1993, for definitions and terms of the art. It is to be understood that this disclosure is not limited to the particular methodology, protocols, and reagents described, as these may vary.

I. Definitions

[31] The terms below are more fully defined by reference to the specification as a whole.

[32] The term "polypeptide" as used herein refers to a compound made up of a single chain of amino acid residues linked by peptide bonds. The term "protein" as used herein may be synonymous with the term "polypeptide".

[33] "Variant" means a protein which is derived from a precursor protein (*e.g.*, the native protein) by addition of one or more amino acids to either or both the C- and N-terminal end, substitution of one or more amino acids at one or a number of different sites in the amino acid sequence, or deletion of one or more amino acids at either or both ends of the protein or at one or more sites in the amino acid sequence, or one or more amino acids is modified by changing the charge (*i.e.* by removing a positive charge, adding a negative charge, or by both removing a positive charge and adding a negative charge). The preparation of a cellulase variant may be performed by any means known in the art, including chemical modification of amino acids, by modifying a DNA sequence which encodes for the native protein, transformation of the modified DNA sequence into a suitable host, and expression of the modified DNA sequence to form the variant enzyme. The variant cellulase of the disclosure includes peptides comprising altered amino acid sequences in comparison with a precursor enzyme amino acid sequence wherein the variant cellulase retains the characteristic

cellulolytic nature of the precursor enzyme but which may have altered properties in some specific aspect. For example, a variant cellulase may have an increased pH optimum or increased temperature or oxidative stability or decreased affinity or binding to non-cellulosic materials but will retain its characteristic cellulolytic activity. It is contemplated that the

5 variants according to the present disclosure may be derived from a DNA fragment encoding a cellulase variant wherein the functional activity of the expressed cellulase variant is retained. For example, a DNA fragment encoding a cellulase may further include a DNA sequence or portion thereof encoding a hinge or linker attached to the cellulase DNA sequence at either the 5' or 3' end wherein the functional activity of the encoded cellulase domain is retained.

10 The terms variant and derivative may be used interchangeably herein.

[34] "Equivalent residues" may also be defined by determining homology at the level of tertiary structure for a precursor cellulase whose tertiary structure has been determined by x-ray crystallography. Equivalent residues are defined as those for which the atomic coordinates of two or more of the main chain atoms of a particular amino acid residue of a cellulase and *Hypocrea jecorina* CBH2 (N on N, CA on CA, C on C and O on O) are within
15 0.13 nm and preferably 0.1 nm after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the cellulase in question to the *H. jecorina* CBH2. The best model is the crystallographic model giving the lowest R factor for experimental diffraction data at
20 the highest resolution available see for examples US 2006/0205042.

[35] Equivalent residues which are functionally analogous to a specific residue of *H. jecorina* CBH2 are defined as those amino acids of a cellulase which may adopt a conformation such that they either alter, modify or contribute to protein structure, substrate binding or catalysis in a manner defined and attributed to a specific residue of the *H. jecorina*
25 CBH2. Further, they are those residues of the cellulase (for which a tertiary structure has been obtained by x-ray crystallography) which occupy an analogous position to the extent that, although the main chain atoms of the given residue may not satisfy the criteria of equivalence on the basis of occupying a homologous position, the atomic coordinates of at least two of the side chain atoms of the residue lie within 0.13 nm of the corresponding side
30 chain atoms of *H. jecorina* CBH2. The crystal structure of *H. jecorina* CBH2 is shown in Zou et al. (1999) (Ref. 5, supra).

[36] The term "nucleic acid molecule" includes RNA, DNA and cDNA molecules. It will be understood that, as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding a given protein such as CBH2 and/or variants thereof may be produced.

The present disclosure contemplates every possible variant nucleotide sequence, encoding variant cellulase such as CBH2, all of which are possible given the degeneracy of the genetic code.

[37] A "heterologous" nucleic acid construct or sequence has a portion of the sequence which is not native to the cell in which it is expressed. Heterologous, with respect to a control sequence refers to a control sequence (i.e. promoter or enhancer) that does not function in nature to regulate the same gene the expression of which it is currently regulating. Generally, heterologous nucleic acid sequences are not endogenous to the cell or part of the genome in which they are present, and have been added to the cell, by infection, transfection, transformation, microinjection, electroporation, or the like. A "heterologous" nucleic acid construct may contain a control sequence/DNA coding sequence combination that is the same as, or different from a control sequence/DNA coding sequence combination found in the native cell.

[38] As used herein, the term "vector" refers to a nucleic acid construct designed for transfer between different host cells. An "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

[39] Accordingly, an "expression cassette" or "expression vector" is a nucleic acid construct generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a target cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid sequence to be transcribed and a promoter.

[40] As used herein, the term "plasmid" refers to a circular double-stranded (ds) DNA construct used as a cloning vector, and which forms an extrachromosomal self-replicating genetic element in many bacteria and some eukaryotes.

[41] As used herein, the term "selectable marker-encoding nucleotide sequence" refers to a nucleotide sequence which is capable of expression in cells and where expression of the selectable marker confers to cells containing the expressed gene the ability to grow in the presence of a corresponding selective agent, or under corresponding selective growth conditions.

[42] As used herein, the term "promoter" refers to a nucleic acid sequence that functions to direct transcription of a downstream gene. The promoter will generally be appropriate to the host cell in which the target gene is being expressed. The promoter together with other transcriptional and translational regulatory nucleic acid sequences (also termed "control sequences") are necessary to express a given gene. In general, the transcriptional and translational regulatory sequences include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences.

[43] "Chimeric gene" or "heterologous nucleic acid construct", as defined herein refers to a non-native gene (i.e., one that has been introduced into a host) that may be composed of parts of different genes, including regulatory elements. A chimeric gene construct for transformation of a host cell is typically composed of a transcriptional regulatory region (promoter) operably linked to a heterologous protein coding sequence, or, in a selectable marker chimeric gene, to a selectable marker gene encoding a protein conferring, for example, antibiotic resistance to transformed cells. A typical chimeric gene of the present disclosure, for transformation into a host cell, includes a transcriptional regulatory region that is constitutive or inducible, a protein coding sequence, and a terminator sequence. A chimeric gene construct may also include a second DNA sequence encoding a signal peptide if secretion of the target protein is desired.

[44] A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA encoding a secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading frame. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors, linkers or primers for PCR are used in accordance with conventional practice.

[45] As used herein, the term "gene" means the segment of DNA involved in producing a polypeptide chain, that may or may not include regions preceding and following the coding region, e.g. 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).

[46] In general, nucleic acid molecules which encode the variant cellulase such as CBH2 will hybridize, under moderate to high stringency conditions to the wild type sequence such as provided herein as SEQ ID NO:1. However, in some cases a CBH2-encoding nucleotide sequence is employed that possesses a substantially different codon usage, while the protein encoded by the CBH2-encoding nucleotide sequence has the same or substantially the same amino acid sequence as the native protein. For example, the coding sequence may be modified to facilitate faster expression of CBH2 in a particular prokaryotic or eukaryotic expression system, in accordance with the frequency with which a particular codon is utilized by the host (Te'o et al., FEMS Microbiology Letters, 190: 13-19, 2000, for example, describes the optimization of genes for expression in filamentous fungi).

[47] A nucleic acid sequence is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Hybridization conditions are based on the melting temperature (T_m) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about $T_m - 5^\circ\text{C}$ (5°C below the T_m of the probe); "high stringency" at about $5-10^\circ\text{C}$ below the T_m ; "moderate" or "intermediate stringency" at about $10-20^\circ\text{C}$ below the T_m of the probe; and "low stringency" at about $20-25^\circ\text{C}$ below the T_m . Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify sequences having about 80% or more sequence identity with the probe.

[48] Moderate and high stringency hybridization conditions are well known in the art (see, for example, Sambrook, et al, 1989, Chapters 9 and 11, and in Ausubel, F. M., et al., 1993, expressly incorporated by reference herein). An example of high stringency conditions includes hybridization at about 42°C in 50% formamide, 5xSSC, 5timesDenhardt's solution, 0.5% SDS and 100 ug/ml denatured carrier DNA followed by washing two times in 2timesSSC and 0.5% SDS at room temperature and two additional times in 0.1x SSC and 0.5% SDS at 42°C .

[49] The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form

of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

[50] As used herein, the terms "transformed", "stably transformed" or "transgenic" with reference to a cell means the cell has a non-native (heterologous) nucleic acid sequence integrated into its genome or as an episomal plasmid that is maintained through multiple generations.

[51] As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation.

10 [52] The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell where the nucleic acid sequence may be incorporated into the genome of the cell (for example, chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (for example, transfected mRNA).

[53] It follows that the term "CBH2 expression" refers to transcription and translation of the *cbh2* gene or variants thereof, the products of which include precursor RNA, mRNA, polypeptide, post-translationally processed polypeptides, and derivatives thereof, including CBH2 from related species such as *Trichoderma koningii*, *Hypocrea jecorina* (also known as
20 *Trichoderma longibrachiatum*, *Trichoderma reesei* or *Trichoderma viride*) and *Hypocrea schweinitzii*. By way of example, assays for CBH2 expression include Western blot for CBH2 protein, Northern blot analysis and reverse transcriptase polymerase chain reaction (RT-PCR) assays for *cbh2* mRNA, and Phosphoric Acid Swollen Cellulose (PASC) and p-hydroxybenzoic acid hydrazide (PAHBAH) assays as described in the following: (a) PASC: (Karlsson, J. et al. (2001), Eur. J. Biochem, 268, 6498-6507, Wood, T. (1988) in Methods in Enzymology, Vol.160. Biomass Part a Cellulose and Hemicellulose (Wood, W. & Kellog, S. Eds.), pp.19-25, Academic Press, San Diego, Calif., USA) and (b) PAHBAH: (Lever, M. (1972) Analytical Biochemistry, 47, 273, Blakeney, A. B. & Mutton, L. L. (1980) Journal of Science of Food and Agriculture, 31, 889, Henry, R. J. (1984) Journal of the Institute of
25 Brewing, 90, 37).

[54] The term "alternative splicing" refers to the process whereby multiple polypeptide isoforms are generated from a single gene, and involves the splicing together of nonconsecutive exons during the processing of some, but not all, transcripts of the gene. Thus a particular exon may be connected to any one of several alternative exons to form messenger

RNAs. The alternatively-spliced mRNAs produce polypeptides ("splice variants") in which some parts are common while other parts are different.

[55] The term "signal sequence" refers to a sequence of amino acids at the N-terminal portion of a protein that facilitates the secretion of the mature form of the protein outside the cell. The mature form of the extracellular protein lacks the signal sequence that is cleaved off during the secretion process.

[56] By the term "host cell" is meant a cell that contains a vector and supports the replication, and/or transcription or transcription and translation (expression) of the expression construct. Host cells for use in the present disclosure can be prokaryotic cells, such as *E. coli*, or eukaryotic cells such as yeast, plant, insect, amphibian, or mammalian cells. In general, host cells are filamentous fungi.

[57] The term "filamentous fungi" means any and all filamentous fungi recognized by those of skill in the art. A preferred fungus is selected from the group consisting of *Aspergillus*, *Trichoderma*, *Fusarium*, *Chrysosporium*, *Penicillium*, *Humicola*, *Neurospora*, or alternative sexual forms thereof such as *Emericella*, *Hypocrea*. It has now been demonstrated that the asexual industrial fungus *Trichoderma reesei* is a clonal derivative of the ascomycete *Hypocrea jecorina* (See, Kuhls et al., PNAS, 93:7755-7760, 1996).

[58] The term "cellooligosaccharide" refers to oligosaccharide groups containing from 2-8 glucose units and having beta-1,4 linkages, e.g., cellobiose.

[59] The terms "cellulase" "cellulolytic enzymes" or "cellulase enzymes" refer to a category of enzymes capable of hydrolyzing cellulose polymers to shorter cello-oligosaccharide oligomers, cellobiose and/or glucose. Numerous examples of cellulases, such as exoglucanases, exocellobiohydrolases, endoglucanases, and glucosidases have been obtained from cellulolytic organisms, particularly including fungi, plants and bacteria. The enzymes made by these microbes are mixtures of proteins with three types of actions useful in the conversion of cellulose to glucose: endoglucanases (EG), cellobiohydrolases (CBH), and beta-glucosidase (BGL or Bglu). These three different types of cellulase enzymes act synergistically to convert cellulose and its derivatives to glucose.

[60] Many microbes make enzymes that hydrolyze cellulose, including the wood rotting fungus *Trichoderma*, the compost bacteria *Thermomonospora*, *Bacillus*, and *Cellulomonas*; *Streptomyces*; and the fungi *Humicola*, *Aspergillus* and *Fusarium*.

[61] CBH2 from *Hypocrea jecorina* is a member of the Glycosyl Hydrolase Family 6 (hence Cel6) and, specifically, was the first member of that family identified in *Hypocrea jecorina* (hence Cel6A). The Glycosyl Hydrolase Family 6 contains both Endoglucanases and

Cellobiohydrolases/exoglucanases, and that CBH2 is the latter. Thus, the phrases CBH2, CBH2-type protein and Cel6 cellobiohydrolases may be used interchangeably herein.

[62] The term "cellulose binding domain" as used herein refers to portion of the amino acid sequence of a cellulase or a region of the enzyme that is involved in the cellulose binding activity of a cellulase or derivative thereof. Cellulose binding domains generally function by non-covalently binding the cellulase to cellulose, a cellulose derivative or other polysaccharide equivalent thereof. Cellulose binding domains permit or facilitate hydrolysis of cellulose fibers by the structurally distinct catalytic core region, and typically function independent of the catalytic core. Thus, a cellulose binding domain will not possess the significant hydrolytic activity attributable to a catalytic core. In other words, a cellulose binding domain is a structural element of the cellulase enzyme protein tertiary structure that is distinct from the structural element which possesses catalytic activity. Cellulose binding domain and cellulose binding module may be used interchangeably herein.

[63] As used herein, the term "surfactant" refers to any compound generally recognized in the art as having surface active qualities. Thus, for example, surfactants comprise anionic, cationic and nonionic surfactants such as those commonly found in detergents. Anionic surfactants include linear or branched alkylbenzenesulfonates; alkyl or alkenyl ether sulfates having linear or branched alkyl groups or alkenyl groups; alkyl or alkenyl sulfates; olefinsulfonates; and alkanesulfonates. Ampholytic surfactants include quaternary ammonium salt sulfonates, and betaine-type ampholytic surfactants. Such ampholytic surfactants have both the positive and negative charged groups in the same molecule. Nonionic surfactants may comprise polyoxyalkylene ethers, as well as higher fatty acid alkanolamides or alkylene oxide adduct thereof, fatty acid glycerine monoesters, and the like.

[64] As used herein, the term "cellulose containing fabric" refers to any sewn or unsewn fabrics, yarns or fibers made of cotton or non-cotton containing cellulose or cotton or non-cotton containing cellulose blends including natural cellulosics and manmade cellulosics (such as jute, flax, ramie, rayon, and lyocell).

[65] As used herein, the term "cotton-containing fabric" refers to sewn or unsewn fabrics, yarns or fibers made of pure cotton or cotton blends including cotton woven fabrics, cotton knits, cotton denims, cotton yarns, raw cotton and the like.

[66] As used herein, the term "stonewashing composition" refers to a formulation for use in stonewashing cellulose containing fabrics. Stonewashing compositions are used to modify cellulose containing fabrics prior to sale, i.e., during the manufacturing process. In contrast,

detergent compositions are intended for the cleaning of soiled garments and are not used during the manufacturing process.

[67] As used herein, the term "detergent composition" refers to a mixture which is intended for use in a wash medium for the laundering of soiled cellulose containing fabrics.

5 In the context of the present disclosure, such compositions may include, in addition to cellulases and surfactants, additional hydrolytic enzymes, builders, bleaching agents, bleach activators, bluing agents and fluorescent dyes, caking inhibitors, masking agents, cellulase activators, antioxidants, and solubilizers.

[68] As used herein, the term "decrease or elimination in expression of the cbh2 gene" means that either that the cbh2 gene has been deleted from the genome and therefore cannot be expressed by the recombinant host microorganism; or that the cbh2 gene or transcript has been modified such that a functional CBH2 enzyme is not produced by the host microorganism or at levels that are significantly less than the unmodified cbh2 gene or transcript.

15 [69] The term "variant cbh2 gene" means that the nucleic acid sequence of the cbh2 gene from *H. jecorina* has been altered by removing, adding, and/or manipulating the coding sequence.

[70] As used herein, the term "purifying" generally refers to subjecting transgenic nucleic acid or protein containing cells to biochemical purification and/or column chromatography.

20 [71] As used herein, the terms "active" and "biologically active" refer to a biological activity associated with a particular protein and are used interchangeably herein. For example, the enzymatic activity associated with a protease is proteolysis and, thus, an active protease has proteolytic activity. It follows that the biological activity of a given protein refers to any biological activity typically attributed to that protein by those of skill in the art.

25 [72] As used herein, the term "enriched" means that the cellulase such as CBH2 is found in a concentration that is greater relative to the CBH2 concentration found in a wild-type, or naturally occurring, fungal cellulase composition. The terms enriched, elevated and enhanced may be used interchangeably herein.

[73] A wild type fungal cellulase composition is one produced by a naturally occurring fungal source and which comprises one or more BGL, CBH and EG components wherein each of these components is found at the ratio produced by the fungal source. Thus, an enriched CBH composition would have CBH at an altered ratio wherein the ratio of CBH to other cellulase components (i.e., EGs, beta-glucosidases and other endoglucanases) is

elevated. This ratio may be increased by either increasing CBH or decreasing (or eliminating) at least one other component by any means known in the art.

[74] The term "isolated" or "purified" as used herein refers to a nucleic acid or amino acid that is removed from at least one component with which it is naturally associated.

5 [75] Thus, to illustrate, a naturally occurring cellulase system may be purified into substantially pure components by recognized separation techniques well published in the literature, including ion exchange chromatography at a suitable pH, affinity chromatography, size exclusion and the like. For example, in ion exchange chromatography (usually anion exchange chromatography), it is possible to separate the cellulase components by eluting with
10 a pH gradient, or a salt gradient, or both a pH and a salt gradient. The purified CBH may then be added to the enzymatic solution resulting in an enriched CBH solution. It is also possible to elevate the amount of CBH produced by a microbe using molecular genetics methods to overexpress the gene encoding CBH, possibly in conjunction with deletion of one or more genes encoding other cellulases.

15 [76] Fungal cellulases may contain more than one CBH component. The different components generally have different isoelectric points which allow for their separation via ion exchange chromatography and the like. Either a single CBH component or a combination of CBH components may be employed in an enzymatic solution.

[77] When employed in enzymatic solutions, the homolog or variant CBH2 component is
20 generally added in an amount sufficient to allow the highest rate of release of soluble sugars from the biomass. The amount of homolog or variant CBH2 component added depends, upon the type of biomass to be saccharified, which can be readily determined by the skilled artisan when employed, the weight percent of the homolog or variant CBH2 component present in the cellulase composition is from preferably between 1 and 100 with illustrative examples
25 being about 1, preferably about 5, preferably about 10, preferably about 15, or preferably about 20 weight percent to preferably about 25, preferably about 30, preferably about 35, preferably about 40, preferably about 45 or preferably about 50 weight percent. Furthermore, preferred ranges may be about 0.5 to about 15 weight percent, about 0.5 to about 20 weight percent, from about 1 to about 10 weight percent, from about 1 to about 15 weight percent,
30 from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from

about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45 weight percent, from about 10 to about 50 weight percent, from about 15 to about 60 weight percent, from about 15 to about 65 weight percent, from about 15 to about 70 weight percent, from about 15 to about 75 weight percent, from about 15 to about 80 weight percent, from about 15 to about 85 weight percent, from about 15 to about 95 weight percent. However, when employed, the weight percent of the homolog or variant CBH2 component relative to any EG type components present in the cellulase composition is from preferably about 1, preferably about 5, preferably about 10, preferably about 15, or preferably about 20 weight percent to preferably about 25, preferably about 30, preferably about 35, preferably about 40, preferably about 45 or preferably about 50 weight percent. Furthermore, preferred ranges may be about 0.5 to about 15 weight percent, about 0.5 to about 20 weight percent, from about 1 to about 10 weight percent, from about 1 to about 15 weight percent, from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45 weight percent, from about 10 to about 50 weight percent, from about 15 to about 20 weight percent, from about 15 to about 25 weight percent, from about 15 to about 30 weight percent, from about 15 to about 35 weight percent, from about 15 to about 40 weight percent, from about 15 to about 45 weight percent, from about 15 to about 50 weight percent.

II. Cellulases

[78] Cellulases are known in the art as enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellobiosaccharides, and the like. As set forth above, cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases (EC 3.2.1.21) ("BG").

[79] Certain fungi produce complete cellulase systems which include exo-cellobiohydrolases or CBH-type cellulases, endoglucanases or EG-type cellulases and beta-glucosidases or BG-type cellulases. However, sometimes these systems lack CBH-type

cellulases and bacterial cellulases also typically include little or no CBH-type cellulases. In addition, it has been shown that the EG components and CBH components synergistically interact to more efficiently degrade cellulose. The different components, i.e., the various endoglucanases and exocellobiohydrolases in a multi-component or complete cellulase system, generally have different properties, such as isoelectric point, molecular weight, degree of glycosylation, substrate specificity and enzymatic action patterns.

[80] It is believed that endoglucanase-type cellulases hydrolyze internal beta-1,4-glucosidic bonds in regions of low crystallinity of the cellulose and exo-cellobiohydrolase-type cellulases hydrolyze cellobiose from the reducing or non-reducing end of cellulose. It follows that the action of endoglucanase components can greatly facilitate the action of exo-cellobiohydrolases by creating new chain ends which are recognized by exo-cellobiohydrolase components. Further, beta-glucosidase-type cellulases have been shown to catalyze the hydrolysis of alkyl and/or aryl beta.-D-glucosides such as methyl beta.-D-glucoside and p-nitrophenyl glucoside as well as glycosides containing only carbohydrate residues, such as cellobiose. This yields glucose as the sole product for the microorganism and reduces or eliminates cellobiose which inhibits cellobiohydrolases and endoglucanases.

[81] Cellulases also find a number of uses in detergent compositions including to enhance cleaning ability, as a softening agent and to improve the feel of cotton fabrics (Hemmpel, ITB Dyeing/Printing/Finishing 3:5-14, 1991; Tyndall, Textile Chemist and Colorist 24:23-26, 1992; and Kumar et al., Textile Chemist and Colorist, 29:37-42, 1997). While the mechanism is not part of the disclosure, softening and color restoration properties of cellulase have been attributed to the alkaline endoglucanase components in cellulase compositions, as exemplified by U.S. Pat. Nos. 5,648,263, 5,691,178, and 5,776,757, which disclose that detergent compositions containing a cellulase composition enriched in a specified alkaline endoglucanase component impart color restoration and improved softening to treated garments as compared to cellulase compositions not enriched in such a component. In addition, the use of such alkaline endoglucanase components in detergent compositions has been shown to complement the pH requirements of the detergent composition (e.g., by exhibiting maximal activity at an alkaline pH of 7.5 to 10, as described in U.S. Pat. Nos. 5,648,263, 5,691,178, and 5,776,757).

[82] Cellulase compositions have also been shown to degrade cotton-containing fabrics, resulting in reduced strength loss in the fabric (U.S. Pat. No. 4,822,516), contributing to reluctance to use cellulase compositions in commercial detergent applications. Cellulase compositions comprising endoglucanase components have been suggested to exhibit reduced

strength loss for cotton-containing fabrics as compared to compositions comprising a complete cellulase system.

[83] Cellulases have also been shown to be useful in degradation of cellulase biomass to ethanol (wherein the cellulase degrades cellulose to glucose and yeast or other microbes further ferment the glucose into ethanol), in the treatment of mechanical pulp (Pere et al., In Proc. Tappi Pulping Conf., Nashville, Tenn., 27-31, pp. 693-696, 1996), for use as a feed additive (WO 91/04673) and in grain wet milling.

[84] Most CBHs and EGs have a multidomain structure consisting of a core domain separated from a cellulose binding domain (CBD) by a linker peptide (Suurnakki et al., 2000). The core domain contains the active site whereas the CBD interacts with cellulose by binding the enzyme to it (van Tilbeurgh et al., FEBS Lett. 204:223-227, 1986; Tomme et al., Eur. J. Biochem. 170:575-581, 1988). The CBDs are particularly important in the hydrolysis of crystalline cellulose. It has been shown that the ability of cellobiohydrolases to degrade crystalline cellulose clearly decreases when the CBD is absent (Linder and Teeri, J.

Biotechnol. 57:15-28, 1997). However, the exact role and action mechanism of CBDs is still a matter of speculation. It has been suggested that the CBD enhances the enzymatic activity merely by increasing the effective enzyme concentration at the surface of cellulose (Stahlberg et al., Bio/Technol. 9:286-290, 1991), and/or by loosening single cellulose chains from the cellulose surface (Tormo et al., EMBO J. vol. 15, no. 21, pp. 5739-5751, 1996). Most studies concerning the effects of cellulase domains on different substrates have been carried out with core proteins of cellobiohydrolases, as their core proteins can easily be produced by limited proteolysis with papain (Tomme et al., 1988). Numerous cellulases have been described in the scientific literature, examples of which include: from *Trichoderma reesei*: Shoemaker, S. et al., Bio/Technology, 1:691-696, 1983, which discloses CBH1; Teeri, T. et al., Gene, 51:43-52, 1987, which discloses CBH2. Cellulases from species other than *Trichoderma* have also been described e.g., Ooi et al., Nucleic Acids Research, vol. 18, no. 19, 1990, which discloses the cDNA sequence coding for endoglucanase F1-CMC produced by *Aspergillus aculeatus*; Kawaguchi T et al., Gene 173(2):287-8, 1996, which discloses the cloning and sequencing of the cDNA encoding beta-glucosidase 1 from *Aspergillus aculeatus*; Sakamoto et al., Curr. Genet. 27:435-439, 1995, which discloses the cDNA sequence encoding the endoglucanase CMCase-1 from *Aspergillus kawachii* IFO 4308; Saarilahti et al., Gene 90:9-14, 1990, which discloses an endoglucanase from *Erwinia carotovora*; Spilliaert R, et al., Eur J Biochem. 224(3):923-30, 1994, which discloses the cloning and sequencing of bglA, coding for a thermostable beta-glucanase from *Rhodothermus marinus*; and Halldorsdottir S et al., Appl

Microbiol Biotechnol. 49(3):277-84, 1998, which discloses the cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12. However, there remains a need for identification and characterization of novel cellulases, with improved properties, such as improved performance
5 under conditions of thermal stress or in the presence of surfactants, increased specific activity, altered substrate cleavage pattern, and/or high level expression in vitro.

[85] The development of new and improved cellulase compositions that comprise varying amounts CBH-type, EG-type and BG-type cellulases is of interest for use: (1) in compositions for degrading wood pulp or other biomass into sugars (e.g., for biochemicals
10 production such as bio-fuels); (2) in detergent compositions that exhibit enhanced cleaning ability (3) function as a softening agent and/or improve the feel of cotton fabrics (e.g., "stone washing" or "biopolishing"); and/or (3) in feed compositions, for example.

[86] Also provided herein are whole cellulase preparation comprising cellulase variants. As used herein, the phrase "whole cellulase preparation" refers to both naturally occurring
15 and non-naturally occurring cellulase containing compositions. A "naturally occurring" composition is one produced by a naturally occurring source and which comprises one or more cellobiohydrolase-type, one or more endoglucanase-type, and one or more beta-glucosidase components wherein each of these components is found at the ratio produced by the source. A naturally occurring composition is one that is produced by an organism
20 unmodified with respect to the cellulolytic enzymes such that the ratio of the component enzymes is unaltered from that produced by the native organism. A "non-naturally occurring" composition encompasses those compositions produced by: (1) combining component cellulolytic enzymes either in a naturally occurring ratio or non-naturally occurring, i.e., altered, ratio; or (2) modifying an organism to overexpress or underexpress
25 one or more cellulolytic enzyme; or (3) modifying an organism such that at least one cellulolytic enzyme is deleted. Accordingly, in some embodiments, the whole cellulase preparation can have one or more of the various EGs and/or CBHs, and/or beta-glucosidase deleted. For example, EG1 may be deleted alone or in combination with other EGs and/or CBHs.

[87] In general, the whole cellulase preparation includes enzymes including, but are not
30 limited to: (i) endoglucanases (EG) or 1,4- β -D-glucan-4-glucanohydrolases (EC 3.2.1.4), (ii) exoglucanases, including 1,4- β -D-glucan glucanohydrolases (also known as cellodextrinases) (EC 3.2.1.74) and 1,4- β -D-glucan cellobiohydrolases (exo-cellobiohydrolases, CBH) (EC 3.2.1.91), and (iii) β -glucosidase (BG) or β -glucoside glucohydrolases (EC 3.2.1.21).

[88] In the present disclosure, the whole cellulase preparation can be from any microorganism that is useful for the hydrolysis of a cellulosic material. In some embodiments, the whole cellulase preparation is a filamentous fungi whole cellulase. "Filamentous fungi" include all filamentous forms of the subdivision *Eumycota* and *Oomycota*. In some embodiments, the whole cellulase preparation is a *Acremonium*, *Aspergillus*, *Emericella*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Scytalidium*, *Thielavia*, *Tolypocladium*, or *Trichoderma* species, whole cellulase. In some embodiments, the whole cellulase preparation is an *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, or *Aspergillus oryzae* whole cellulase. In another aspect, whole cellulase preparation is a *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, or *Fusarium venenatum* whole cellulase. In another aspect, the whole cellulase preparation is a *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Scytalidium thermophilum*, or *Thielavia terrestris* whole cellulase. In another aspect, the whole cellulase preparation a *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* e.g., RL-P37 (Sheir-Neiss et al., Appl. Microbiol. Biotechnology, 20 (1984) pp. 46-53; Montenecourt B.S., Can., 1-20, 1987), QM9414 (ATCC No. 26921), NRRL 15709, ATCC 13631, 56764, 56466, 56767, or *Trichoderma viride* e.g., ATCC 32098 and 32086, whole cellulase. In some embodiments, the whole cellulase preparation is a *Trichoderma reesei* RutC30 whole cellulase, which is available from the American Type Culture Collection as *Trichoderma reesei* ATCC 56765.

[89] Examples of commercial cellulase preparations suitable for use in the present disclosure include, for example, CELLUCLASTTM (available from Novozymes A/S) and LAMINEXTM, IndiAgeTM and PrimafastTM LAMINEX BG enzyme, ACCELLERASETM 100 and ACCELLERASETM 1500 (available Genencor Division, Danisco US. Inc.)

[90] In the present disclosure, the whole cellulase preparation can be from any microorganism cultivation method known in the art resulting in the expression of enzymes capable of hydrolyzing a cellulosic material. Fermentation can include shake flask cultivation, small- or large-scale fermentation, such as continuous, batch, fed-batch, or solid

state fermentations in laboratory or industrial fermenters performed in a suitable medium and under conditions allowing the cellulase to be expressed or isolated.

[91] Generally, the microorganism is cultivated in a cell culture medium suitable for production of enzymes capable of hydrolyzing a cellulosic material. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable culture media, temperature ranges and other conditions suitable for growth and cellulase production are known in the art. As a non-limiting example, the normal temperature range for the production of cellulases by *Trichoderma reesei* is 24°C to 28°C.

[92] Generally, the whole cellulase preparation is used as is produced by fermentation with no or minimal recovery and/or purification. For example, once cellulases are secreted by a cell into the cell culture medium, the cell culture medium containing the cellulases can be used. In some embodiments the whole cellulase preparation comprises the unfractionated contents of fermentation material, including cell culture medium, extracellular enzymes and cells. Alternatively, the whole cellulase preparation can be processed by any convenient method, e.g., by precipitation, centrifugation, affinity, filtration or any other method known in the art. In some embodiments, the whole cellulase preparation can be concentrated, for example, and then used without further purification. In some embodiments the whole cellulase preparation comprises chemical agents that decrease cell viability or kills the cells. In some embodiments, the cells are lysed or permeabilized using methods known in the art.

III. Molecular Biology

[93] In one embodiment this disclosure provides for the expression of variant *cbh2* genes under control of a promoter functional in a filamentous fungus. Therefore, this disclosure relies on routine techniques in the field of recombinant genetics (See, e.g., Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd ed., 1989; Kriegler, *Gene Transfer and Expression: A Laboratory Manual*, 1990; and Ausubel et al., eds., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, Greene Publishing and Wiley-Interscience, New York, 1994).

Methods of Mutating *cbh2* Nucleic Acid Sequences

[94] Any method known in the art that can introduce mutations is contemplated by the present disclosure.

[95] The present disclosure relates to the expression, purification and/or isolation and use of variant CBH2. These enzymes are preferably prepared by recombinant methods utilizing

the *cbh2* gene from *H. jecorina*. The fermentation broth may be used with or without purification.

[96] After the isolation and cloning of the *cbh2* gene from *H. jecorina*, other methods known in the art, such as site directed mutagenesis, are used to make the substitutions, additions or deletions that correspond to substituted amino acids in the expressed CBH2 variant. Again, site directed mutagenesis and other methods of incorporating amino acid changes in expressed proteins at the DNA level are known in the art (Sambrook et al., supra; and Ausubel et al., supra).

[97] DNA encoding an amino acid sequence variant of the *H. jecorina* CBH2 is prepared by a variety of methods known in the art. These methods include, but are not limited to, preparation by site-directed (or oligonucleotide-mediated) mutagenesis, PCR mutagenesis, and cassette mutagenesis of an earlier prepared DNA encoding the *H. jecorina* CBH2.

[98] Site-directed mutagenesis is a preferred method for preparing substitution variants. This technique is well known in the art (see, e.g., Carter et al. Nucleic Acids Res. 13:4431-4443 (1985) and Kunkel et al., Proc. Natl. Acad. Sci. USA 82:488 (1987)). Briefly, in carrying out site-directed mutagenesis of DNA, the starting DNA is altered by first hybridizing an oligonucleotide encoding the desired mutation to a single strand of such starting DNA. After hybridization, a DNA polymerase is used to synthesize an entire second strand, using the hybridized oligonucleotide as a primer, and using the single strand of the starting DNA as a template. Thus, the oligonucleotide encoding the desired mutation is incorporated in the resulting double-stranded DNA.

[99] PCR mutagenesis is also suitable for making amino acid sequence variants of the starting polypeptide, i.e., *H. jecorina* CBH2. See Higuchi, in PCR Protocols, pp. 177-183 (Academic Press, 1990); and Vallette et al., Nuc. Acids Res. 17:723-733 (1989). See, also, for example Cadwell et al., PCR Methods and Applications, Vol 2, 28-33 (1992). Briefly, when small amounts of template DNA are used as starting material in a PCR, primers that differ slightly in sequence from the corresponding region in a template DNA can be used to generate relatively large quantities of a specific DNA fragment that differs from the template sequence only at the positions where the primers differ from the template.

[100] Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al., Gene 34:315-323 (1985). The starting material is the plasmid (or other vector) comprising the starting polypeptide DNA to be mutated. The codon(s) in the starting DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites

exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the starting polypeptide DNA. The plasmid DNA is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but
5 containing the desired mutation(s) is synthesized using standard procedures, wherein the two strands of the oligonucleotide are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 5' and 3' ends that are compatible with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the
10 mutated DNA sequence.

[101] Alternatively, or additionally, the desired amino acid sequence encoding a variant CBH2 can be determined, and a nucleic acid sequence encoding such amino acid sequence variant can be generated synthetically.

[102] The variant CBH2(s) so prepared may be subjected to further modifications,
15 oftentimes depending on the intended use of the cellulase. Such modifications may involve further alteration of the amino acid sequence, fusion to heterologous polypeptide(s) and/or covalent modifications.

IV. *cbh2* Nucleic Acids And CBH2 Polypeptides

A. Variant *cbh2*-Type Nucleic acids

[103] The nucleic acid sequence for the wild type *cbh2* is shown in SEQ ID NO:1. The disclosure encompasses a nucleic acid molecule encoding the variant cellulases described herein. The nucleic acid may be a DNA molecule.

[104] After DNA sequences that encode the CBH2 variants have been cloned into DNA
25 constructs, the DNA is used to transform microorganisms. The microorganism to be transformed for the purpose of expressing a variant CBH2 according to the present disclosure may advantageously comprise a strain derived from *Trichoderma sp.* Thus, a preferred mode for preparing variant CBH2 cellulases according to the present disclosure comprises
30 transforming a *Trichoderma sp.* host cell with a DNA construct comprising at least a fragment of DNA encoding a portion or all of the variant CBH2. The DNA construct will generally be functionally attached to a promoter. The transformed host cell is then grown under conditions so as to express the desired protein. Subsequently, the desired protein product may be purified to substantial homogeneity.

[105] However, it may in fact be that the best expression vehicle for a given DNA encoding a variant CBH2 may differ from *H. jecorina*. Thus, it may be that it will be most advantageous to express a protein in a transformation host that bears phylogenetic similarity to the source organism for the variant CBH2. In an alternative embodiment, *Aspergillus niger* can be used as an expression vehicle. For a description of transformation techniques with *A. niger*, see WO 98/31821, the disclosure of which is incorporated by reference in its entirety.

[106] Accordingly, the present description of an *Aspergillus* spp. expression system is provided for illustrative purposes only and as one option for expressing the variant CBH2 of the disclosure. One of skill in the art, however, may be inclined to express the DNA encoding variant CBH2 in a different host cell if appropriate and it should be understood that the source of the variant CBH2 should be considered in determining the optimal expression host. Additionally, the skilled worker in the field will be capable of selecting the best expression system for a particular gene through routine techniques utilizing the tools available in the art.

B. Variant CBH2 Polypeptides

[107] The variant CBH2's of this disclosure have amino acid sequences that are derived from the amino acid sequence of a precursor CBH2. The amino acid sequence of the CBH2 variant differs from the precursor CBH2 amino acid sequence by the substitution, deletion or insertion of one or more amino acids of the precursor amino acid sequence. In a preferred embodiment, the precursor CBH2 is *Hypocrea jecorina* CBH2. The mature amino acid sequence of *H. jecorina* CBH2 is shown in SEQ ID NO:3. Thus, this disclosure is directed to CBH2 variants which contain amino acid residues at positions which are equivalent to the particular identified residue in *H. jecorina* CBH2. A residue (amino acid) of an CBH2 homolog is equivalent to a residue of *Hypocrea jecorina* CBH2 if it is either homologous (i.e., corresponding in position in either primary or tertiary structure) or is functionally analogous to a specific residue or portion of that residue in *Hypocrea jecorina* CBH2 (i.e., having the same or similar functional capacity to combine, react, or interact chemically or structurally). As used herein, numbering is intended to correspond to that of the mature CBH2 amino acid sequence (SEQ ID NO:3).

[108] Alignment of amino acid sequences to determine homology is preferably determined by using a "sequence comparison algorithm." Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson &

Lipman, Proc. Nat'l Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection, Visual inspection may utilize graphics packages such as, for example, MOE by
5 Chemical Computing Group, Montreal Canada.

[109] An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul, et al., J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). This algorithm involves first
10 identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the
15 cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved value; the cumulative score goes to zero or below; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix
20 (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M⁵, N⁻⁴, and a comparison of both strands.

[110] The BLAST algorithm then performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum
25 probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, an amino acid sequence is considered similar to a protease if the smallest sum probability in a comparison of the test amino acid sequence to a protease amino acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

30 [111] For purposes of the present disclosure, the degree of identity may be suitably determined by means of computer programs known in the art, such as GAP provided in the GCG program package (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wis., USA 53711) (Needleman, S. B. and Wunsch, C. D., (1970), Journal of Molecular Biology, 48, 443-45),

using GAP with the following settings for polynucleotide sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3.

[112] A structural alignment between a *T. reesei* CBH2 and other cellulases may be used to identify equivalent/corresponding positions in other cellulases having a moderate to high degree of homology, e.g., about 50%, 55%, 60 %, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or even 99%, with *T. reesei* CBH2 (SEQ ID NO: 3). One method of obtaining said structural alignment is to use the Pile Up programme from the GCG package using default values of gap penalties, i.e., a gap creation penalty of 3.0 and gap extension penalty of 0.1. Other structural alignment methods include the hydrophobic cluster analysis (Gaboriaud *et al.*, FEBS Letters, 224:149-155, 1987) and reverse threading (Huber and Torda, Protein Science, 7:142-149, 1998).

[113] An exemplary alignment of the mature form of various reference cellulases is provided as Figure 3. The reference cellulases include: *Hypocrea jecorina* (also known as *T. reesei*) CBH2 (SEQ ID NO:3), *Hypocrea koningii* CBH2 (SEQ ID NO:4), *Humicola insolens* CBH2 (SEQ ID NO:5), *Acremonium cellulolyticus* CBH2 (SEQ ID NO:6), *Agaricus bisporus* CBH2 (SEQ ID NO:7), *Fusarium osysporum* EG (SEQ ID NO:8), *Phanerochaete chrysosporium* CBH2 (SEQ ID NO:9), *Talaromyces emersonii* CBH2 (SEQ ID NO:10), *Thermobifida fusca* 6B/E3 CBH2 (SEQ ID NO:11), *Thermobifida fusca* 6A/E2 EG (SEQ ID NO:12), and *Cellulomonas fimi* CenA EG (SEQ ID NO:13). Sequences were aligned using the ClustalW and the MUSCLE multiple sequence alignment algorithms. A matrix showing the percent identity of cellulases of the sequence alignment of Figure 3 is provided in Table 1.

Table 1. Cellulase Percent Identity Matrix*

Percent_ID	3	4	5	6	7	8	9	10	11	12	13
3	100	95.5	62.3	64.7	59.6	63.1	55.4	63.4	31.9	13.5	27
4	95.5	100	61.6	64	59.1	63.6	54.7	63	32.9	13.5	26.8
5	62.3	61.6	100	59.1	57.6	61.3	54	58.8	31.9	15.9	26.6
6	64.7	64	59.1	100	58.6	56.4	54	72.6	32.8	13.5	29.2
7	59.6	59.1	57.6	58.6	100	55.8	69.1	58.1	34.9	17.5	27.6
8	63.1	63.6	61.3	56.4	55.8	100	48.7	54.8	31.1	13.9	25.2
9	55.4	54.7	54	54	69.1	48.7	100	52.6	32.4	15.4	25.6
10	63.4	63	58.8	72.6	58.1	54.8	52.6	100	33.9	13.2	27.3
11	31.9	32.9	31.9	32.8	34.9	31.1	32.4	33.9	100	15.9	36.3
12	13.5	13.5	15.9	13.5	17.5	13.9	15.4	13.2	15.9	100	12.8
13	27	26.8	26.6	29.2	27.6	25.2	25.6	27.3	36.3	12.8	100

*Numbers in the top row and left column correspond to the SEQ ID NOS of the aligned sequences of Figure 3.

5 [114] Sequence searches are typically carried out using the BLASTN program when evaluating a given nucleic acid sequence relative to nucleic acid sequences in the GenBank DNA Sequences and other public databases. The BLASTX program is preferred for searching nucleic acid sequences that have been translated in all reading frames against amino acid sequences in the GenBank Protein Sequences and other public databases. Both BLASTN and BLASTX are run using default parameters of an open gap penalty of 11.0, and
10 an extended gap penalty of 1.0, and utilize the BLOSUM-62 matrix. (See, e.g., Altschul, et al., 1997.)

V. Expression Of Recombinant CBH2 Variants

[115] The methods of the disclosure rely on the use cells to express variant CBH2, with no
15 particular method of CBH2 expression required. The variant CBH2 is preferably secreted from the cells. The disclosure provides host cells which have been transduced, transformed or transfected with an expression vector comprising a variant CBH2-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the parental host cell prior to transduction, transformation or transfection and will be
20 apparent to those skilled in the art.

[116] In one approach, a filamentous fungal cell or yeast cell is transfected with an expression vector having a promoter or biologically active promoter fragment or one or more (e.g., a series) of enhancers which functions in the host cell line, operably linked to a DNA segment encoding variant CBH2, such that variant CBH2 is expressed in the cell line.
25

A. Nucleic Acid Constructs/Expression Vectors.

[117] Natural or synthetic polynucleotide fragments encoding variant CBH2 ("CBH2-encoding nucleic acid sequences") may be incorporated into heterologous nucleic acid constructs or vectors, capable of introduction into, and replication in, a filamentous fungal or
30 yeast cell. The vectors and methods disclosed herein are suitable for use in host cells for the expression of variant CBH2. Any vector may be used as long as it is replicable and viable in the cells into which it is introduced. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Cloning and expression vectors are also described in Sambrook et al., 1989, Ausubel F M et al., 1989, and Strathern

et al., The Molecular Biology of the Yeast *Saccharomyces*, 1981, each of which is expressly incorporated by reference herein. Appropriate expression vectors for fungi are described in van den Hondel, C. A. M. J. J. et al. (1991) In: Bennett, J. W. and Lasure, L. L. (eds.) *More Gene Manipulations in Fungi*. Academic Press, pp. 396-428. The appropriate DNA sequence
5 may be inserted into a plasmid or vector (collectively referred to herein as "vectors") by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by standard procedures. Such procedures and related sub-cloning procedures are deemed to be within the scope of knowledge of those skilled in the art.

[118] Recombinant filamentous fungi comprising the coding sequence for variant CBH2
10 may be produced by introducing a heterologous nucleic acid construct comprising the variant CBH2 coding sequence into the cells of a selected strain of the filamentous fungi.

[119] Once the desired form of a variant *cbh2* nucleic acid sequence is obtained, it may be modified in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions,
15 transversions, deletions, and insertions may be performed on the naturally occurring sequence.

[120] A selected variant *cbh2* coding sequence may be inserted into a suitable vector according to well-known recombinant techniques and used to transform filamentous fungi capable of CBH2 expression. Due to the inherent degeneracy of the genetic code, other
20 nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used to clone and express variant CBH2. Therefore it is appreciated that such substitutions in the coding region fall within the sequence variants covered by the present disclosure. Any and all of these sequence variants can be utilized in the same way as described herein for a parent CBH2-encoding nucleic acid sequence.

[121] The present disclosure also includes recombinant nucleic acid constructs comprising
25 one or more of the variant CBH2-encoding nucleic acid sequences as described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the disclosure has been inserted, in a forward or reverse orientation.

[122] Heterologous nucleic acid constructs may include the coding sequence for variant
30 *cbh2*. (i) in isolation; (ii) in combination with additional coding sequences; such as fusion protein or signal peptide coding sequences, where the *cbh2* coding sequence is the dominant coding sequence; (iii) in combination with non-coding sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions,

effective for expression of the coding sequence in a suitable host; and/or (iv) in a vector or host environment in which the cbh2 coding sequence is a heterologous gene.

[123] In one aspect of the present disclosure, a heterologous nucleic acid construct is employed to transfer a variant CBH2-encoding nucleic acid sequence into a cell in vitro, with established filamentous fungal and yeast lines preferred. For long-term, production of variant CBH2, stable expression is preferred. It follows that any method effective to generate stable transformants may be used in practicing the disclosure.

[124] Appropriate vectors are typically equipped with a selectable marker-encoding nucleic acid sequence, insertion sites, and suitable control elements, such as promoter and termination sequences. The vector may comprise regulatory sequences, including, for example, non-coding sequences, such as introns and control elements, i.e., promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in host cells (and/or in a vector or host cell environment in which a modified soluble protein antigen coding sequence is not normally expressed), operably linked to the coding sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, many of which are commercially available and/or are described in Sambrook, et al., (supra).

[125] Exemplary promoters include both constitutive promoters and inducible promoters, examples of which include a CMV promoter, an SV40 early promoter, an RSV promoter, an EF-1.alpha. promoter, a promoter containing the tet responsive element (TRE) in the tet-on or tet-off system as described (ClonTech and BASF), the beta actin promoter and the metallothionine promoter that can upregulated by addition of certain metal salts. A promoter sequence is a DNA sequence which is recognized by the particular filamentous fungus for expression purposes. It is operably linked to DNA sequence encoding a variant CBH2 polypeptide. Such linkage comprises positioning of the promoter with respect to the initiation codon of the DNA sequence encoding the variant CBH2 polypeptide in the disclosed expression vectors. The promoter sequence contains transcription and translation control sequence which mediate the expression of the variant CBH2 polypeptide. Examples include the promoters from the *Aspergillus niger*, *A. awamori* or *A. oryzae* glucoamylase, alpha-amylase, or alpha-glucosidase encoding genes; the *A. nidulans* gpdA or trpC Genes; the *Neurospora crassa* cbh1 or trp1 genes; the *A. niger* or *Rhizomucor miehei* aspartic proteinase encoding genes; the *H. jecorina* (*T. reesei*) cbh1, cbh2, egl1, egl2, or other cellulase encoding genes.

[126] The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art. Typical selectable marker genes include *argB* from *A. nidulans* or *T. reesei*, *amdS* from *A. nidulans*, *pyr4* from *Neurospora crassa* or *T. reesei*, *pyrG* from *Aspergillus niger* or *A. nidulans*. Additional
5 exemplary selectable markers include, but are not limited to *trpc*, *trp1*, *oliC31*, *niaD* or *leu2*, which are included in heterologous nucleic acid constructs used to transform a mutant strain such as *trp-*, *pyr-*, *leu-* and the like.

[127] Such selectable markers confer to transformants the ability to utilize a metabolite that is usually not metabolized by the filamentous fungi. For example, the *amdS* gene from *H. jecorina* which encodes the enzyme acetamidase that allows transformant cells to grow on acetamide as a nitrogen source. The selectable marker (e.g. *pyrG*) may restore the ability of
10 an auxotrophic mutant strain to grow on a selective minimal medium or the selectable marker (e.g. *olic31*) may confer to transformants the ability to grow in the presence of an inhibitory drug or antibiotic.

[128] The selectable marker coding sequence is cloned into any suitable plasmid using methods generally employed in the art. Exemplary plasmids include pUC18, pBR322, pRAX and pUC100. The pRAX plasmid contains AMAL sequences from *A. nidulans*, which make it possible to replicate in *A. niger*.
15

[129] The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Sambrook et al., 1989; Freshney, Animal Cell Culture, 1987; Ausubel, et al., 1993; and Coligan et al., Current Protocols in Immunology, 1991.
20

25 B. Host Cells and Culture Conditions For CBH2 Production

(i) Filamentous Fungi

[130] Thus, the present disclosure provides filamentous fungi comprising cells which have been modified, selected and cultured in a manner effective to result in variant CBH2 production or expression relative to the corresponding non-transformed parental fungi.

[131] Examples of species of parental filamentous fungi that may be treated and/or modified for variant CBH2 expression include, but are not limited to *Trichoderma*, e.g., *Trichoderma reesei*, *Trichoderma longibrachiatum*, *Trichoderma viride*, *Trichoderma koningii*; *Penicillium* sp., *Humicola* sp., including *Humicola insolens*; *Aspergillus* sp., *Chrysosporium* sp., *Fusarium* sp., *Hypocrea* sp., and *Emericella* sp.
30

[132] CBH2 expressing cells are cultured under conditions typically employed to culture the parental fungal line. Generally, cells are cultured in a standard medium containing physiological salts and nutrients, such as described in Pourquie, J. et al., Biochemistry and Genetics of Cellulose Degradation, eds. Aubert, J. P. et al., Academic Press, pp. 71-86, 1988 and Ilmen, M. et al., Appl. Environ. Microbiol. 63:1298-1306, 1997. Culture conditions are also standard, e.g., cultures are incubated at 28.degree. C. in shaker cultures or fermenters until desired levels of CBH2 expression are achieved.

[133] Preferred culture conditions for a given filamentous fungus may be found in the scientific literature and/or from the source of the fungi such as the American Type Culture Collection (ATCC; www.atcc.org/). After fungal growth has been established, the cells are exposed to conditions effective to cause or permit the expression of variant CBH2.

[134] In cases where a CBH2 coding sequence is under the control of an inducible promoter, the inducing agent, e.g., a sugar, metal salt or antibiotics, is added to the medium at a concentration effective to induce CBH2 expression.

[135] In one embodiment, the strain comprises *Aspergillus niger*, which is a useful strain for obtaining overexpressed protein. For example *A. niger* var *awamori* dgr246 is known to secrete elevated amounts of secreted cellulases (Goedegebuur et al., Curr. Genet (2002) 41: 89-98). Other strains of *Aspergillus niger* var *awamori* such as GCDAP3, GCDAP4 and GAP3-4 are known (Ward et al., Appl. Microbiol. Biotechnol. 39:738-743, 1993).

[136] In another embodiment, the strain comprises *Trichoderma reesei*, which is a useful strain for obtaining overexpressed protein. For example, RL-P37, described by Sheir-Neiss, et al., Appl. Microbiol. Biotechnol. 20:46-53 (1984) is known to secrete elevated amounts of cellulase enzymes. Functional equivalents of RL-P37 include *Trichoderma reesei* strain RUT-C30 (ATCC No. 56765) and strain QM9414 (ATCC No. 26921). It is contemplated that these strains would also be useful in overexpressing variant CBH2.

[137] Where it is desired to obtain the variant CBH2 in the absence of potentially detrimental native cellulolytic activity, it is useful to obtain a *Trichoderma* host cell strain which has had one or more cellulase genes deleted prior to introduction of a DNA construct or plasmid containing the DNA fragment encoding the variant CBH2. Such strains may be prepared by the method disclosed in U.S. Pat. No. 5,246,853 and WO 92/06209, which disclosures are hereby incorporated by reference. By expressing a variant CBH2 cellulase in a host microorganism that is missing one or more cellulase genes, the identification and subsequent purification procedures are simplified. Any gene from *Trichoderma* sp. which has been cloned can be deleted, for example, the *cbh1*, *cbh2*, *egl1*, and *egl2* genes as well as those

encoding EG III and/or EGV protein (see e.g., U.S. Pat. No. 5,475,101 and WO 94/28117, respectively).

[138] Gene deletion may be accomplished by inserting a form of the desired gene to be deleted or disrupted into a plasmid by methods known in the art. The deletion plasmid is then cut at an appropriate restriction enzyme site(s), internal to the desired gene coding region, and the gene coding sequence or part thereof replaced with a selectable marker. Flanking DNA sequences from the locus of the gene to be deleted or disrupted, preferably between about 0.5 to 2.0 kb, remain on either side of the selectable marker gene. An appropriate deletion plasmid will generally have unique restriction enzyme sites present therein to enable the fragment containing the deleted gene, including flanking DNA sequences, and the selectable marker gene to be removed as a single linear piece.

[139] A selectable marker must be chosen so as to enable detection of the transformed microorganism. Any selectable marker gene that is expressed in the selected microorganism will be suitable. For example, with *Aspergillus* sp., the selectable marker is chosen so that the presence of the selectable marker in the transformants will not significantly affect the properties thereof. Such a selectable marker may be a gene that encodes an assayable product. For example, a functional copy of a *Aspergillus* sp. gene may be used which if lacking in the host strain results in the host strain displaying an auxotrophic phenotype. Similarly, selectable markers exist for *Trichoderma* sp.

[140] In one embodiment, a *pyrG*- derivative strain of *Aspergillus* sp. is transformed with a functional *pyrG* gene, which thus provides a selectable marker for transformation. A *pyrG*- derivative strain may be obtained by selection of *Aspergillus* sp. strains that are resistant to fluoroorotic acid (FOA). The *pyrG* gene encodes orotidine-5'-monophosphate decarboxylase, an enzyme required for the biosynthesis of uridine. Strains with an intact *pyrG* gene grow in a medium lacking uridine but are sensitive to fluoroorotic acid. It is possible to select *pyrG*- derivative strains that lack a functional orotidine monophosphate decarboxylase enzyme and require uridine for growth by selecting for FOA resistance. Using the FOA selection technique it is also possible to obtain uridine-requiring strains which lack a functional orotate pyrophosphoribosyl transferase. It is possible to transform these cells with a functional copy of the gene encoding this enzyme (Berges & Barreau, *Curr. Genet.* 19:359-365 (1991), and van Hartingsveldt et al., (1986) Development of a homologous transformation system for *Aspergillus niger* based on the *pyrG* gene. *Mol. Gen. Genet.* 206:71-75). Selection of derivative strains is easily performed using the FOA resistance technique referred to above, and thus, the *pyrG* gene is preferably employed as a selectable marker.

[141] In a second embodiment, a pyr4⁻ derivative strain of *Hyprocrea* sp. (*Trichoderma* sp.) is transformed with a functional pyr4 gene, which thus provides a selectable marker for transformation. A pyr4^{sup}- derivative strain may be obtained by selection of *Hyprocrea* sp. (*Trichoderma* sp.) strains that are resistant to fluoroorotic acid (FOA). The pyr4 gene encodes orotidine-5'-monophosphate decarboxylase, an enzyme required for the biosynthesis of uridine. Strains with an intact pyr4 gene grow in a medium lacking uridine but are sensitive to fluoroorotic acid. It is possible to select pyr4^{sup}- derivative strains that lack a functional orotidine monophosphate decarboxylase enzyme and require uridine for growth by selecting for FOA resistance. Using the FOA selection technique it is also possible to obtain uridine-requiring strains which lack a functional orotate pyrophosphoribosyl transferase. It is possible to transform these cells with a functional copy of the gene encoding this enzyme (Berges & Barreau, 1991). Selection of derivative strains is easily performed using the FOA resistance technique referred to above, and thus, the pyr4 gene is preferably employed as a selectable marker.

[142] To transform pyrG⁻ *Aspergillus* sp. or pyr4⁻ *Hyprocrea* sp. (*Trichoderma* sp.) so as to be lacking in the ability to express one or more cellulase genes, a single DNA fragment comprising a disrupted or deleted cellulase gene is then isolated from the deletion plasmid and used to transform an appropriate pyr⁻ *Aspergillus* or pyr⁻ *Trichoderma* host. Transformants are then identified and selected based on their ability to express the pyrG or pyr4, respectively, gene product and thus compliment the uridine auxotrophy of the host strain. Southern blot analysis is then carried out on the resultant transformants to identify and confirm a double crossover integration event that replaces part or all of the coding region of the genomic copy of the gene to be deleted with the appropriate pyr selectable markers.

[143] Although the specific plasmid vectors described above relate to preparation of pyr⁻ transformants, the present disclosure is not limited to these vectors. Various genes can be deleted and replaced in the *Aspergillus* sp. or *Hyprocrea* sp. (*Trichoderma* sp.) strain using the above techniques. In addition, any available selectable markers can be used, as discussed above. In fact, any host, e.g., *Aspergillus* sp. or *Hyprocrea* sp., gene that has been cloned, and thus identified, can be deleted from the genome using the above-described strategy.

[144] As stated above, the host strains used may be derivatives of *Hyprocrea* sp. (*Trichoderma* sp.) that lack or have a nonfunctional gene or genes corresponding to the selectable marker chosen. For example, if the selectable marker of pyrG is chosen for *Aspergillus* sp., then a specific pyrG⁻ derivative strain is used as a recipient in the transformation procedure. Also, for example, if the selectable marker of pyr4 is chosen for a

Hypocrea sp., then a specific pyr4- derivative strain is used as a recipient in the transformation procedure. Similarly, selectable markers comprising *Hypocrea sp.* (*Trichoderma sp.*) genes equivalent to the *Aspergillus nidulans* genes amdS, argB, trpC, niaD may be used. The corresponding recipient strain must therefore be a derivative strain such as argB-, trpC-, niaD-, respectively.

[145] DNA encoding the CBH2 variant is then prepared for insertion into an appropriate microorganism. According to the present disclosure, DNA encoding a CBH2 variant comprises the DNA necessary to encode for a protein that has functional cellulolytic activity. The DNA fragment encoding the CBH2 variant may be functionally attached to a fungal promoter sequence, for example, the promoter of the glaA gene in *Aspergillus* or the promoter of the cbh1 or egl1 genes in *Trichoderma*.

[146] It is also contemplated that more than one copy of DNA encoding a CBH2 variant may be recombined into the strain to facilitate overexpression. The DNA encoding the CBH2 variant may be prepared by the construction of an expression vector carrying the DNA encoding the variant. The expression vector carrying the inserted DNA fragment encoding the CBH2 variant may be any vector which is capable of replicating autonomously in a given host organism or of integrating into the DNA of the host, typically a plasmid. In preferred embodiments two types of expression vectors for obtaining expression of genes are contemplated. The first contains DNA sequences in which the promoter, gene-coding region, and terminator sequence all originate from the gene to be expressed. Gene truncation may be obtained where desired by deleting undesired DNA sequences (e.g., coding for unwanted domains) to leave the domain to be expressed under control of its own transcriptional and translational regulatory sequences. A selectable marker may also be contained on the vector allowing the selection for integration into the host of multiple copies of the novel gene sequences.

[147] The second type of expression vector is preassembled and contains sequences required for high-level transcription and a selectable marker. It is contemplated that the coding region for a gene or part thereof can be inserted into this general-purpose expression vector such that it is under the transcriptional control of the expression cassettes promoter and terminator sequences.

[148] For example, in *Aspergillus*, pRAX is such a general-purpose expression vector. Genes or part thereof can be inserted downstream of the strong glaA promoter.

[149] For example, in *Hypocrea*, pTEX is such a general-purpose expression vector. Genes or part thereof can be inserted downstream of the strong cbh1 promoter.

[150] In the vector, the DNA sequence encoding the CBH2 variant of the present disclosure should be operably linked to transcriptional and translational sequences, i.e., a suitable promoter sequence and signal sequence in reading frame to the structural gene. The promoter may be any DNA sequence that shows transcriptional activity in the host cell and may be
5 derived from genes encoding proteins either homologous or heterologous to the host cell. An optional signal peptide provides for extracellular production of the CBH2 variant. The DNA encoding the signal sequence is preferably that which is naturally associated with the gene to be expressed, however the signal sequence from any suitable source, for example an exo-cellobiohydrolase or endoglucanase from *Trichoderma*, is contemplated in the present
10 disclosure.

[151] The procedures used to ligate the DNA sequences coding for the variant CBH2 of the present disclosure with the promoter, and insertion into suitable vectors are well known in the art.

[152] The DNA vector or construct described above may be introduced in the host cell in
15 accordance with known techniques such as transformation, transfection, microinjection, microporation, biolistic bombardment and the like.

[153] In the preferred transformation technique, it must be taken into account that the permeability of the cell wall to DNA in *Hyprocrea* sp. (*Trichoderma* sp.) is very low. Accordingly, uptake of the desired DNA sequence, gene or gene fragment is at best minimal.
20 There are a number of methods to increase the permeability of the *Hyprocrea* sp. (*Trichoderma* sp.) cell wall in the derivative strain (i.e., lacking a functional gene corresponding to the used selectable marker) prior to the transformation process.

[154] The preferred method in the present disclosure to prepare *Aspergillus* sp. or *Hyprocrea* sp. (*Trichoderma* sp.) for transformation involves the preparation of protoplasts
25 from fungal mycelium. See Campbell et al. Improved transformation efficiency of *A. niger* using homologous *niaD* gene for nitrate reductase. *Curr. Genet.* 16:53-56; 1989. The mycelium can be obtained from germinated vegetative spores. The mycelium is treated with an enzyme that digests the cell wall resulting in protoplasts. The protoplasts are then protected by the presence of an osmotic stabilizer in the suspending medium. These
30 stabilizers include sorbitol, mannitol, potassium chloride, magnesium sulfate and the like. Usually the concentration of these stabilizers varies between 0.8 M and 1.2 M. It is preferable to use about a 1.2 M solution of sorbitol in the suspension medium.

[155] Uptake of the DNA into the host strain, (*Aspergillus* sp. or *Hyprocrea* sp. (*Trichoderma* sp.)), is dependent upon the calcium ion concentration. Generally between about

10 mM CaCl.sub.2 and 50 mM CaCl.sub.2 is used in an uptake solution. Besides the need for the calcium ion in the uptake solution, other items generally included are a buffering system such as TE buffer (10 mM Tris, pH 7.4; 1 mM EDTA) or 10 mM MOPS, pH 6.0 buffer (morpholinepropanesulfonic acid) and polyethylene glycol (PEG). It is believed that the polyethylene glycol acts to fuse the cell membranes thus permitting the contents of the medium to be delivered into the cytoplasm of the host cell, by way of example either *Aspergillus* sp. or *Hyprocrea* sp. strain, and the plasmid DNA is transferred to the nucleus. This fusion frequently leaves multiple copies of the plasmid DNA integrated into the host chromosome.

10 [156] Usually a suspension containing the *Aspergillus* sp. protoplasts or cells that have been subjected to a permeability treatment at a density of 10.sup.5 to 10.sup.6/mL, preferably 2.times.10.sup.5/mL are used in transformation. Similarly, a suspension containing the *Hyprocrea* sp. (*Trichoderma* sp.) protoplasts or cells that have been subjected to a permeability treatment at a density of 10.sup.8 to 10.sup.9/mL, preferably 2.times.10.sup.8/mL are used in transformation. A volume of 100 .mu.L of these protoplasts or cells in an appropriate solution (e.g., 1.2 M sorbitol; 50 mM CaCl.sub.2) are mixed with the desired DNA. Generally a high concentration of PEG is added to the uptake solution. From 0.1 to 1 volume of 25% PEG 4000 can be added to the protoplast suspension. However, it is preferable to add about 0.25 volumes to the protoplast suspension. Additives such as dimethyl sulfoxide, heparin, spermidine, potassium chloride and the like may also be added to the uptake solution and aid in transformation.

[157] Generally, the mixture is then incubated at approximately 0.degree. C. for a period of between 10 to 30 minutes. Additional PEG is then added to the mixture to further enhance the uptake of the desired gene or DNA sequence. The 25% PEG 4000 is generally added in volumes of 5 to 15 times the volume of the transformation mixture; however, greater and lesser volumes may be suitable. The 25% PEG 4000 is preferably about 10 times the volume of the transformation mixture. After the PEG is added, the transformation mixture is then incubated either at room temperature or on ice before the addition of a sorbitol and CaCl.sub.2 solution. The protoplast suspension is then further added to molten aliquots of a growth medium. This growth medium permits the growth of transformants only. Any growth medium can be used in the present disclosure that is suitable to grow the desired transformants. However, if Pyr.sup.+ transformants are being selected it is preferable to use a growth medium that contains no uridine. The subsequent colonies are transferred and purified on a growth medium depleted of uridine.

[158] At this stage, stable transformants may be distinguished from unstable transformants by their faster growth rate and, in *Trichoderma*, for example, the formation of circular colonies with a smooth, rather than ragged outline on solid culture medium lacking uridine. Additionally, in some cases a further test of stability may be made by growing the transformants on solid non-selective medium (i.e. containing uridine), harvesting spores from this culture medium and determining the percentage of these spores which will subsequently germinate and grow on selective medium lacking uridine.

[159] In a particular embodiment of the above method, the CBH2 variant(s) are recovered in active form from the host cell after growth in liquid media as a result of the appropriate post translational processing of the CBH2 variant.

(ii) Yeast

[160] The present disclosure also contemplates the use of yeast as a host cell for CBH2 production. Several other genes encoding hydrolytic enzymes have been expressed in various strains of the yeast *S. cerevisiae*. These include sequences encoding for two endoglucanases (Penttila et al., *Yeast* vol. 3, pp 175-185, 1987), two cellobiohydrolases (Penttila et al., *Gene*, 63: 103-112, 1988) and one beta-glucosidase from *Trichoderma reesei* (Cummings and Fowler, *Curr. Genet.* 29:227-233, 1996), a xylanase from *Aureobasidium pullulans* (Li and Ljungdahl, *Appl. Environ. Microbiol.* 62, no.1, pp. 209-213, 1996), an alpha-amylase from wheat (Rothstein et al., *Gene* 55:353-356, 1987), etc. In addition, a cellulase gene cassette encoding the *Butyrivibrio fibrisolvens* endo-[beta]-1,4-glucanase (END1), *Phanerochaete chrysosporium* cellobiohydrolase (CBH1), the *Ruminococcus flavefaciens* cellodextrinase (CEL1) and the *Endomyces fibrilizer* cellobiase (Bgl1) was successfully expressed in a laboratory strain of *S. cerevisiae* (Van Rensburg et al., *Yeast*, vol. 14, pp. 67-76, 1998).

C. Introduction of a CBH2-Encoding Nucleic Acid Sequence into Host Cells.

[161] The disclosure further provides cells and cell compositions which have been genetically modified to comprise an exogenously provided variant CBH2-encoding nucleic acid sequence. A parental cell or cell line may be genetically modified (i.e., transduced, transformed or transfected) with a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc., as further described above.

[162] The methods of transformation of the present disclosure may result in the stable integration of all or part of the transformation vector into the genome of the filamentous

fungus. However, transformation resulting in the maintenance of a self-replicating extra-chromosomal transformation vector is also contemplated.

[163] Many standard transfection methods can be used to produce *Trichoderma reesei* cell lines that express large quantities of the heterologous protein. Some of the published methods for the introduction of DNA constructs into cellulase-producing strains of *Trichoderma* include Lorito, Hayes, DiPietro and Harman, 1993, *Curr. Genet.* 24: 349-356; Goldman, VanMontagu and Herrera-Estrella, 1990, *Curr. Genet.* 17:169-174; Penttila, Nevalainen, Ratto, Salminen and Knowles, 1987, *Gene* 6: 155-164, for *Aspergillus* Yelton, Hamer and Timberlake, 1984, *Proc. Natl. Acad. Sci. USA* 81: 1470-1474, for *Fusarium* Bajar, Podila and Kolattukudy, 1991, *Proc. Natl. Acad. Sci. USA* 88: 8202-8212, for *Streptomyces* Hopwood et al., 1985, The John Innes Foundation, Norwich, UK and for *Bacillus* Brigidi, DeRossi, Bertarini, Riccardi and Matteuzzi, 1990, *FEMS Microbiol. Lett.* 55: 135-138).

[164] Other methods for introducing a heterologous nucleic acid construct (expression vector) into filamentous fungi (e.g., *H. jecorina*) include, but are not limited to the use of a particle or gene gun, permeabilization of filamentous fungi cells walls prior to the transformation process (e.g., by use of high concentrations of alkali, e.g., 0.05 M to 0.4 M CaCl_2 or lithium acetate), protoplast fusion or *Agrobacterium* mediated transformation. An exemplary method for transformation of filamentous fungi by treatment of protoplasts or spheroplasts with polyethylene glycol and CaCl_2 is described in Campbell, E. I. et al., *Curr. Genet.* 16:53-56, 1989 and Penttila, M. et al., *Gene*, 63:11-22, 1988.

[165] Any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, biolistics, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). Also of use is the *Agrobacterium*-mediated transfection method described in U.S. Pat. No. 6,255,115. It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the heterologous gene.

[166] In addition, heterologous nucleic acid constructs comprising a variant CBH2-encoding nucleic acid sequence can be transcribed in vitro, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection.

[167] The disclosure further includes novel and useful transformants of filamentous fungi such as *H. jecorina* and *A. niger* for use in producing fungal cellulase compositions. The

disclosure includes transformants of filamentous fungi especially fungi comprising the variant CBH2 coding sequence, or deletion of the endogenous cbh coding sequence.

[168] Following introduction of a heterologous nucleic acid construct comprising the coding sequence for a variant cbh2, the genetically modified cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying expression of a variant CBH2-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the host cell selected for expression, and will be apparent to those skilled in the art.

[169] The progeny of cells into which such heterologous nucleic acid constructs have been introduced are generally considered to comprise the variant CBH2-encoding nucleic acid sequence found in the heterologous nucleic acid construct.

[170] The disclosure further includes novel and useful transformants of filamentous fungi such as *H. jecorina* for use in producing fungal cellulase compositions. *Aspergillus niger* may also be used in producing the variant CBH2. The disclosure includes transformants of filamentous fungi especially fungi comprising the variant cbh 2 coding sequence, or deletion of the endogenous cbh2 coding sequence.

[171] Stable transformants of filamentous fungi can generally be distinguished from unstable transformants by their faster growth rate and, in *Trichoderma*, for example, the formation of circular colonies with a smooth rather than ragged outline on solid culture medium. Additionally, in some cases, a further test of stability can be made by growing the transformants on solid non-selective medium, harvesting the spores from this culture medium and determining the percentage of these spores which will subsequently germinate and grow on selective medium.

VI. Isolation And Purification Of Recombinant CBH2 Protein

[172] In general, a variant CBH2 protein produced in cell culture is secreted into the medium and may be purified or isolated, e.g., by removing unwanted components from the cell culture medium. However, in some cases, a variant CBH2 protein may be produced in a cellular form necessitating recovery from a cell lysate. In such cases the variant CBH2 protein is purified from the cells in which it was produced using techniques routinely employed by those of skill in the art. Examples include, but are not limited to, affinity chromatography (Tilbeurgh et al., FEBS Lett. 16:215, 1984), ion-exchange chromatographic methods (Goyal et al., Bioresource Technol. 36:37-50, 1991; Fliess et al., Eur. J. Appl. Microbiol. Biotechnol. 17:314-318, 1983; Bhikhabhai et al., J. Appl. Biochem. 6:336-345,

1984; Ellouz et al., J. Chromatography 396:307-317, 1987), including ion-exchange using materials with high resolution power (Medve et al., J. Chromatography A 808:153-165, 1998), hydrophobic interaction chromatography (Tomaz and Queiroz, J. Chromatography A 865:123-128, 1999), and two-phase partitioning (Brumbauer, et al., Bioseparation 7:287-295, 1999).

[173] Typically, the variant CBH2 protein is fractionated to segregate proteins having selected properties, such as binding affinity to particular binding agents, e.g., antibodies or receptors; or which have a selected molecular weight range, or range of isoelectric points.

[174] Once expression of a given variant CBH2 protein is achieved, the CBH2 protein thereby produced is purified from the cells or cell culture. Exemplary procedures suitable for such purification include the following: antibody-affinity column chromatography, ion exchange chromatography; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; and gel filtration using, e.g., Sephadex G-75. Various methods of protein purification may be employed and such methods are known in the art and described e.g. in Deutscher, Methods in Enzymology, vol. 182, no. 57, pp. 779, 1990; Scopes, Methods Enzymol. 90: 479-91, 1982. The purification step(s) selected will depend, e.g., on the nature of the production process used and the particular protein produced.

VII. Utility of cbh2 and CBH2

[175] It can be appreciated that the variant cbh nucleic acids, the variant CBH2 protein and compositions comprising variant CBH2 protein activity find utility in a wide variety applications, some of which are described below.

[176] New and improved cellulase compositions that comprise varying amounts BG-type, EG-type and variant CBH-type cellulases find utility in detergent compositions that exhibit enhanced cleaning ability, function as a softening agent and/or improve the feel of cotton fabrics (e.g., "stone washing" or "biopolishing"), in compositions for degrading wood pulp into sugars (e.g., for bio-ethanol production), and/or in feed compositions. The isolation and characterization of cellulase of each type provides the ability to control the aspects of such compositions.

[177] In one approach, the cellulase of the disclosure finds utility in detergent compositions or in the treatment of fabrics to improve the feel and appearance.

[178] Since the rate of hydrolysis of cellulosic products may be increased by using a transformant having at least one additional copy of the cbh gene inserted into the genome,

products that contain cellulose or heteroglycans can be degraded at a faster rate and to a greater extent. Products made from cellulose such as paper, cotton, cellulosic diapers and the like can be degraded more efficiently in a landfill. Thus, the fermentation product obtainable from the transformants or the transformants alone may be used in compositions to help

5 degrade by liquefaction a variety of cellulose products that add to the overcrowded landfills.

[179] Separate saccharification and fermentation is a process whereby cellulose present in biomass, e.g., corn stover, is converted to glucose and subsequently yeast strains convert glucose into ethanol. Simultaneous saccharification and fermentation is a process whereby cellulose present in biomass, e.g., corn stover, is converted to glucose and, at the same time and in the same reactor, yeast strains convert glucose into ethanol. Thus, in another approach, the variant CBH type cellulase of the disclosure finds utility in the degradation of biomass to ethanol. Ethanol production from readily available sources of cellulose provides a stable, renewable fuel source.

10 [180] Cellulose-based feedstocks are comprised of agricultural wastes, grasses and woods and other low-value biomass such as municipal waste (e.g., recycled paper, yard clippings, etc.). Ethanol may be produced from the fermentation of any of these cellulosic feedstocks. However, the cellulose must first be converted to sugars before there can be conversion to ethanol.

15 [181] A large variety of feedstocks may be used with the inventive variant CBH and the one selected for use may depend on the region where the conversion is being done. For example, in the Midwestern United States agricultural wastes such as wheat straw, corn stover and bagasse may predominate while in California rice straw may predominate. However, it should be understood that any available cellulosic biomass may be used in any region.

20 [182] The methods of the present disclosure can be used in the production of monosaccharides, disaccharides, and polysaccharides as chemical or fermentation feedstocks for microorganism for the production of organic products, chemicals and fuels, plastics, and other products or intermediates. In particular, the value of processing residues (dried distillers grain, spent grains from brewing, sugarcane bagasse, etc.) can be increased by partial or complete solubilization of cellulose or hemicellulose. In addition to ethanol, some chemicals that can be produced from cellulose and hemicellulose include, acetone, acetate, 25 glycine, lysine, organic acids (e.g., lactic acid), 1,3-propanediol, butanediol, glycerol, ethylene glycol, furfural, polyhydroxyalkanoates, cis, cis-muconic acid, animal feed and xylose.

30

[183] A cellulase composition containing an enhanced amount of cellobiohydrolase finds utility in ethanol production. Ethanol from this process can be further used as an octane enhancer or directly as a fuel in lieu of gasoline which is advantageous because ethanol as a fuel source is more environmentally friendly than petroleum derived products. It is known that the use of ethanol will improve air quality and possibly reduce local ozone levels and smog. Moreover, utilization of ethanol in lieu of gasoline can be of strategic importance in buffering the impact of sudden shifts in non-renewable energy and petrochemical supplies.

[184] Ethanol can be produced via saccharification and fermentation processes from cellulosic biomass such as trees, herbaceous plants, municipal solid waste and agricultural and forestry residues. However, the ratio of individual cellulase enzymes within a naturally occurring cellulase mixture produced by a microbe may not be the most efficient for rapid conversion of cellulose in biomass to glucose. It is known that endoglucanases act to produce new cellulose chain ends which themselves are substrates for the action of cellobiohydrolases and thereby improve the efficiency of hydrolysis of the entire cellulase system. Therefore, the use of increased or optimized cellobiohydrolase activity may greatly enhance the production of ethanol.

[185] Thus, the inventive cellobiohydrolase finds use in the hydrolysis of cellulose to its sugar components. In one embodiment, a variant cellobiohydrolase is added to the biomass prior to the addition of a fermentative organism. In a second embodiment, a variant cellobiohydrolase is added to the biomass at the same time as a fermentative organism. Optionally, there may be other cellulase components present in either embodiment.

[186] In another embodiment the cellulosic feedstock may be pretreated. Pretreatment may be by elevated temperature and the addition of either of dilute acid, concentrated acid or dilute alkali solution. The pretreatment solution is added for a time sufficient to at least partially hydrolyze the hemicellulose components and then neutralized.

[187] The major product of CBH2 action on cellulose is cellobiose which is available for conversion to glucose by BG activity (for instance in a fungal cellulase product). Either by the pretreatment of the cellulosic biomass or by the enzymatic action on the biomass, other sugars, in addition to glucose and cellobiose, can be made available from the biomass. The hemicellulose content of the biomass can be converted (by hemicellulases) to sugars such as xylose, galactose, mannose and arabinose. Thus, in a biomass conversion process, enzymatic saccharification can produce sugars that are made available for biological or chemical conversions to other intermediates or end-products. Therefore, the sugars generated from biomass find use in a variety of processes in addition to the generation of ethanol. Examples

of such conversions are fermentation of glucose to ethanol (as reviewed by M. E. Himmel et al. pp 2-45, in "Fuels and Chemicals from Biomass", ACS Symposium Series 666, ed B. C. Saha and J. Woodward, 1997) and other biological conversions of glucose to 2,5-diketo-D-gluconate (U.S. Pat. No. 6,599,722), lactic acid (R. Datta and S-P. Tsai pp 224-236; *ibid*), succinate (R. R. Gokarn, M. A. Eiteman and J. Sridhar pp 237-263, *ibid*), 1,3-propanediol (A-P. Zheng, H. Biebl and W-D. Deckwer pp 264-279, *ibid*), 2,3-butanediol (C. S. Gong, N. Cao and G. T. Tsao pp 280-293, *ibid*), and the chemical and biological conversions of xylose to xylitol (B. C. Saha and R. J. Bothast pp 307-319, *ibid*). See also, for example, WO 98/21339.

[188] The detergent compositions of this disclosure may employ besides the cellulase composition (irrespective of the cellobiohydrolase content, i.e., cellobiohydrolase-free, substantially cellobiohydrolase-free, or cellobiohydrolase enhanced), a surfactant, including anionic, non-ionic and ampholytic surfactants, a hydrolase, building agents, bleaching agents, bluing agents and fluorescent dyes, caking inhibitors, solubilizers, cationic surfactants and the like. All of these components are known in the detergent art. The cellulase composition as described above can be added to the detergent composition either in a liquid diluent, in granules, in emulsions, in gels, in pastes, and the like. Such forms are well known to the skilled artisan. When a solid detergent composition is employed, the cellulase composition is preferably formulated as granules. Preferably, the granules can be formulated so as to contain a cellulase protecting agent. For a more thorough discussion, see U.S. Pat. No. 6,162,782 entitled "Detergent compositions containing cellulase compositions deficient in CBH2 type components," which is incorporated herein by reference.

[189] Preferably the cellulase compositions are employed from about 0.00005 weight percent to about 5 weight percent relative to the total detergent composition. More preferably, the cellulase compositions are employed from about 0.0002 weight percent to about 2 weight percent relative to the total detergent composition.

[190] In addition the variant CBH2 nucleic acid sequence finds utility in the identification and characterization of related nucleic acid sequences. A number of techniques useful for determining (predicting or confirming) the function of related genes or gene products include, but are not limited to, (A) DNA/RNA analysis, such as (1) overexpression, ectopic expression, and expression in other species; (2) gene knock-out (reverse genetics, targeted knock-out, viral induced gene silencing (VIGS, see Baulcombe, 100 Years of Virology, Calisher and Horzinek eds., Springer-Verlag, New York, N.Y. 15:189-201, 1999); (3) analysis of the methylation status of the gene, especially flanking regulatory regions; and (4) in situ hybridization; (B) gene product analysis such as (1) recombinant protein expression;

(2) antisera production, (3) immunolocalization; (4) biochemical assays for catalytic or other activity; (5) phosphorylation status; and (6) interaction with other proteins via yeast two-hybrid analysis; (C) pathway analysis, such as placing a gene or gene product within a particular biochemical or signaling pathway based on its overexpression phenotype or by sequence homology with related genes; and (D) other analyses which may also be performed to determine or confirm the participation of the isolated gene and its product in a particular metabolic or signaling pathway, and help determine gene function.

EXPERIMENTAL

[191] The present disclosure is described in further detail in the following Examples which are not in any way intended to limit the scope of the disclosure as claimed. The attached Figures are meant to be considered as integral parts of the specification and description of the disclosure. The following Examples are offered to illustrate, but not to limit the claimed disclosure

[192] In the experimental disclosure which follows, the following abbreviations apply: M (molar); mM (millimolar); μ M (micromolar); nM (nanomolar); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); gm (grams); mg (milligrams); μ g (micrograms); pg (picograms); L (liters); ml and mL (milliliters); μ l and μ L (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); U (units); V (volts); MW (molecular weight); sec (seconds); min(s) (minute/minutes); h(s) and hr(s) (hour/hours); $^{\circ}$ C (degrees Centigrade); QS (quantity sufficient); ND (not done); NA (not applicable); rpm (revolutions per minute); H₂O (water); dH₂O (deionized water); HCl (hydrochloric acid); aa (amino acid); bp (base pair); kb (kilobase pair); kD (kilodaltons); cDNA (copy or complementary DNA); DNA (deoxyribonucleic acid); ssDNA (single stranded DNA); dsDNA (double stranded DNA); dNTP (deoxyribonucleotide triphosphate); RNA (ribonucleic acid); MgCl₂ (magnesium chloride); NaCl (sodium chloride); w/v (weight to volume); v/v (volume to volume); g (gravity); OD (optical density); CNPG (chloro-nitro-phenyl-beta-D-glucoside); CNP (2-chloro-4-nitrophenol); APB (acid-pretreated bagasse); PASC (phosphoric acid swollen cellulose) PCS (acid-pretreated corn stover); Pi or PI (performance index); PAGE (polyacrylamide gel electrophoresis); PCR (polymerase chain reaction); RT-PCR (reverse transcription PCR); and HPLC (high pressure liquid chromatography).

EXAMPLE 1

Chemical Modification of a *Trichoderma* sp. Cellulase Preparation and Assays For Testing CBH2 Variants

[193] This Example describes the treatment of a commercial *Trichoderma* sp. cellulase preparation (LAMINEX BG enzyme complex (Genencor Division, Danisco US, Inc.) with succinic anhydride to acetylate the lysine residues. Acetylation of the lysine residues of the LAMINEX BG enzyme complex alters the net charge of the proteins (e.g., increased negative charge). Other similar chemical modifications can be used to also convert the positive charge of the lysine to a negative charged group (for example with acetoxysuccinic anhydride, maleic anhydride, tartaric anhydride and phthalic anhydride treatment) or even to two negative charges (for example with trimellitic anhydride, cis-aconitic anhydride and 4-nitrophthalic anhydride treatment). Other chemical modifications can be used to remove the positive charges of lysine residues resulting in a noncharged residue (for example acetic anhydride, butyric anhydride, isobutyric anhydride, hexanoic anhydride, valeric anhydride, isovaleric anhydride and pivalic anhydride treatment).

[194] Lysine residues on a cellulase preparation were modified using succinic anhydride, using a variation of published methods (Lundblad, Chemical Reagents for Protein Modification, Editor: R. Lundblad, 3rd Edition CRC press, 1984). For this reaction, a 236 mg sample of LAMINEX BG enzyme complex was prepared in 1 mL of 500 mM HEPES buffer pH 8. A succinic anhydride (Aldrich) solution was prepared by dissolving the powder in DMSO to a 500mg/mL final concentration before addition of the enzyme complex. An aliquot of succinic anhydride was added such that a ratio of > 1:100 lysine to succinic acid was achieved in the reaction tube. Another reaction tube was set up with DMSO and enzyme only, using similar volumes, to serve as the unmodified protein control. The tubes were vortexed and left at room temp overnight. The following day, a 1:10 volume of 1M glycine pH 3 was added to each tube to quench the succinic anhydride reaction.

[195] Chemical modification was confirmed by comparing modified and unmodified proteins on native gels. Aliquots from each reaction (chemically-modified and unmodified) were analyzed on gradient 8-25 %, native gels run at pH 8.8 at 100 volts (Phast System gels, GE Healthcare). Proteins were visualized after Coomassie blue staining of the gel, to confirm that the modification was successful. Staining revealed shifts in protein band migration, confirming the changes in charge of the various protein components of the cellulase preparation. Modified samples of *Trichoderma* sp. cellulase preparation were more negativity charged than unmodified samples.

[196] To isolate the modified and unmodified (control) proteins, 80µl aliquots of each sample were desalted using spin desalt columns (Pierce). The absorbance at 280nm of desalted samples (including the control without modification) was measured using a NanoDrop™ spectrophotometer (Thermo), in duplicates after a 1:10 sample dilution to
 5 determine the total protein concentration of the samples.

Zeta Potential Determinations

[197] This Example describes determining the zeta potential of an enzyme and a substrate. The presence of a charge on the surface of a particle influences the distribution of ions in the
 10 surrounding interfacial region. The result is an increased concentration of counter ions of opposite charge to that of the particle near the particle surface. As one moves away from the particle surface, the heterogeneous distribution of ions will eventually become homogeneous. The distance at which a homogenous distribution is obtained is called the Debye length ($1/\kappa$) or screening distance, and is dependent upon the ionic strength as shown in the expression
 15 below, where ϵ_0 is the permittivity of free space ($8.854 \times 10^{-12} \text{ F m}^{-1}$), ϵ_r is the permittivity of the liquid, k is the Boltzmann constant ($1.38 \times 10^{-23} \text{ J K}^{-1}$), T is the temperature in Kelvin, e is the electronic charge ($1.6022 \times 10^{-19} \text{ C}$), I is the molar ionic strength, and NA is Avogadro's constant ($6.022 \times 10^{23} \text{ mol}^{-1}$).

$$\text{Debye Length} = \frac{1}{\kappa} = \sqrt{\frac{\epsilon_0 \epsilon_r k T}{NA e^2 I}} \quad (0.1)$$

20 The molar ionic strength can be calculated from the following equation, where C_i is the ionic species concentration and Z_i is the valency.

$$I = \frac{1}{2} \sum C_i Z_i^2 \quad (0.2)$$

For water at 298 K, the Debye length expression reduces to the following form.

$$\kappa^{-1} = 0.304 (I^{-0.5}) \quad (0.3)$$

25 The liquid layer surrounding the particle exists as two parts; an inner region (Stern layer) where the ions are strongly bound and an outer (diffuse) region where they are less firmly associated. Within the diffuse layer there is a boundary inside which the ions and particles form a stable entity. When a particle moves, ions within this boundary move with it. Those ions beyond the boundary do not travel with the particle. The electric potential at this
 30 boundary, also called the surface of hydrodynamic shear, is defined as the zeta potential.

[198] In electrophoretic light scattering, the zeta potential ζ is calculated from the measured electrophoretic mobility u using the Henry equation shown below, where ϵ is the dielectric

constant, η is the solution viscosity, κ is the inverse Debye length, a is the particle radius, and $f(ka)$ is the Henry function.

$$u = \frac{2ez}{3\eta} f(ka) \quad (0.4)$$

[199] The units of κ are reciprocal length, with $1/\kappa$ being the "thickness" of the electrical double layer (Debye length). The parameter a refers to the radius of the particle, and therefore, ka is the ratio of the particle radius to the electrical double layer thickness. The Henry function, $f(ka)$ depends on particle shape, but is known for a sphere. In the expression above it ranges from $f(0) = 1$ (Hückel limit) to $f(\infty) = 1.5$ (Smoluchowski limit). For small particles such as proteins in a low dielectric (or low ionic strength) medium, the Hückel limit of $f(ka) = 1$ is the more appropriate model.

[200] Zeta potentials of proteins were measured with the Zetasizer NS (Malvern Instruments, UK) according to the principle outlined above. Zeta potentials of BMI-stained fabrics were measured with the SurPass (Anton-Paar, Austria) using the streaming potential implementation of the above principle. From the definition of surface charge, usually expressed in Coulombs:

$$q_s = 4\pi e_o e_r a (1 + ka) z \quad (0.5)$$

This can also be expressed as a net charge z multiplied by the elementary charge e $1.6 \times 10^{-19} \text{C}$:

$$q_s = ze \quad (0.6)$$

Therefore the expected change in zeta potential due to a net charge increment is given by:

$$\frac{Dz}{Dz} = \frac{e}{4\pi e_o e_r a (1 + ka)} \quad (0.7)$$

It is also possible to measure zeta potentials using a native gel technique (Sparks et al., Journal of Lipid Research, 33:123-130, 1992) as described in Example 1. Electrophoretic mobility measured with native gels is usually less than in solution due to retardation caused by the gel matrix. Zeta potentials calculated this way are usually lower compared to solution-based methods. We therefore refer to them as apparent zeta potentials when obtained via the native gel technique.

[201] The effective charge in a given formulation is enumerated as its zeta potential. The use of zeta potential as a common charge scale allows comparison of enzyme variants having different folds (e.g., serine proteases, metalloproteases, etc.), as well as interactions with

different substrates (e.g., BMI microswatch) under the conditions of interest (e.g., AATCC HDL detergent). Although zeta potential is preferred for comparing different protein folds, electrophoretic mobilities or measured charges also provide an absolute scale and are adequate for comparisons. BMI performance as a function of enzyme zeta potential is well described by a standard normal distribution, indicated by the solid line, with a mean μ equal to -9.68 mV, standard deviation σ of 11.39 mV and peak value of 0.4056 [A600-background]. This distribution is indicated in standard reduced coordinates by the BMI activity divided by the peak value on the right Y-axis as a function of the Z score on the top X-axis. The Z score is defined as usual as $(X-\mu)/\sigma$ where X in this case is zeta potential.

Table 1-1. BMI Microswatch Activity of Proteases

Performance Level	Z Score	Zeta Potential ζ Window* For BMI Microswatch Activity
90%	± 0.46	$-14.92 < \zeta < -4.44$
80%	± 0.65	$-17.08 < \zeta < -2.28$
70%	± 0.84	$-19.25 < \zeta < -0.11$
60%	± 1.00	$-21.07 < \zeta < +1.71$
50%	± 1.18	$-23.12 < \zeta < +3.76$

*Mean μ = -9.68 mV, standard deviation σ = 11.39 mV, Zeta potential $\zeta = Z \cdot \sigma + \mu$

Reference buffer: 5mM HEPES pH 8.0, 2.5mM NaCl

[202] The normal distribution is unique to each substrate stain under given reaction conditions (pH, conductivity, type of salt, detergent chelators, etc.). Different benefits or favorable outcomes follow a normal distribution with a physical property that holds across enzymes from various folds, as is for instance, the case of expression levels and zeta potentials for ASP and NprE charge ladder variants. In a normal distribution the peak value occurs at the mean. Comparison of enzyme and substrate charges on a common zeta potential scale reveals that optimum BMI performance occurs when the mean enzyme zeta potential in this case -9.68 mV, essentially matches the substrate stain zeta potential, in this case -8.97 mV, measured under the same conditions.

[203] Performance levels of standard normal distributions are conveniently described in terms of their z scores as indicated in Table 1-1 (See, Abramowitz and Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover, New

York, 9th Ed., 1964). Conversion to zeta potentials is straightforward given knowledge of the mean and standard deviation defining the distribution for a given application. In this example measured cleaning performance for a protein fold is confined to zeta potential values between -40 mV and + 20 mV. Variants with a cleaning performance above 80% of their fold optimum (*i.e.*, $z = \pm 0.65$), are confined to zeta potential values between -17.08 mV and -2.28 mV. Variants with a cleaning performance above 90% of their fold optimum (*i.e.*, $z = \pm 0.46$), are confined to zeta potential values between -14.92 mV and -4.44 mV.

[204] Different substrate stains (*e.g.*, grass, body soils, tomato) have different zeta potentials under the same formulation and the same substrate stain has different zeta potentials under different formulations (*e.g.*, North American HDL, European powder dishwashing detergent). Regardless, while the substrate stain charge varies, the standard deviation of the normal distribution is expected to remain constant. Knowledge of enzyme and substrate zeta potentials in a given detergent formulation allows rapid identification of the expected performance level for that variant, as well as the direction and magnitude of charge change needed in order to achieve optimal performance levels. Measurement of the substrate zeta potential in the desired reaction medium allows optimization of the enzyme reaction on the particulate substrate in that medium. Any enzyme reaction in any medium can be optimized using a similar process.

20 **Optimizing Reactions On Substrates Exhibiting Variable Charge**

[205] Cellulose conversion was evaluated by techniques known in the art (See, *e.g.*, Baker et al., Appl Biochem Biotechnol, 70-72:395-403, 1998). A standard cellulosic conversion assay was used in the experiments. In this assay enzyme and buffered substrate were placed in containers and incubated at a temperature over time. The reaction was quenched with enough 100 mM glycine, pH 11, to bring the pH of the reaction mixture to at least pH 10. Once the reaction was quenched, an aliquot of the reaction mixture was filtered through a 0.2 micron membrane to remove solids. The filtered solution was then assayed for soluble sugars by HPLC according to the methods described in Baker et al., above.

30 **Determination of Zeta-Potential of Pretreated Corn Stover (PCS)**

[206] Corn stover was pretreated with 2% w/w H₂SO₄ as described (Schell *et al.*, J Appl Biochem Biotechnol, 105:69 – 86 [2003]) and followed by multiple washes with deionized water to obtain a pH of 4.5. Sodium acetate was added to make a final concentration of

50mM and the solution was titrated to pH 5.0. The cellulose concentration in the reaction mixture was approximately 7%.

[207] PCS aliquots before and after saccharification by a commercial cellulase mixture, Spezyme CP and Indigae 44L, were dosed into 1.5 mL Eppendorf centrifugation tubes and occupied about one third of the volume. Samples were centrifuged at 6,000 rpm for 5 min, the supernatant exchanged for Milli-Q™ water and the process repeated 5 times. A 100 mg/mL stock solution in Milli-Q™ water was prepared from the rinsed corn stover. This stock was diluted to 1 mg/mL into a 50 mM sodium acetate buffer pH 5.0 for zeta potential measurements. 1 mL aliquot of each substrate sample was transferred to a clean Malvern Instruments (UK) disposable Zetasizer NST™ cuvette.

[208] Table 1-2 indicates that throughout the course of the saccharification reaction the PCS substrate charge, expressed as zeta potential, nearly became twice as negative. Without being bound by theory, there are many explanations for a net negative charge increase including but not limited to enrichment in lignin, the non-reactive portion of this substrate, as well as non-productive binding or fouling of whole cellulases and other proteins. There is an optimal enzyme zeta potential for performance (e.g., extent of reaction and reaction rate), which matches the substrate zeta potential under reaction medium conditions. Different biomass pretreatments may dramatically influence initial substrate charge. If the enzyme or the substrate become zeta potential mismatched throughout the course of the reaction, the enzyme-substrate interaction will no longer be optimal. This effect will be dramatic for changes of nearly 10 mV, which are the case for biomass conversion.

[209] Strategies to remedy this situation include but are not limited to supplying an enzyme blend spanning various charges; a fed-batch process approach where enzymes possessing different charges at the new optimum are supplied at different reaction times and/or extents of conversion; control of substrate surface charge through addition of formulation agents, particularly surfactants (ionic and non-ionic) or other proteins; control of substrate surface charge through pH adjustments; ionic strength adjustments throughout the reaction in order to shift the enzyme charge optima; membrane filtration, particularly reverse osmosis and nanofiltration, to control ionic strength throughout the reaction; addition of chelators to control ionic strength through elimination of salts; and control of biomass substrate charge through pretreatment processes.

Table 1-2. Zeta Potential Of Acid Pretreated Corn Stover	
PCS Condition	Zeta Potential
Initial (before saccharification)	-12.0 ± 7.00 mV
After saccharification by Spezyme CP	-22.2 ± 8.67 mV
After saccharification by Indigae 44L	-22.7 ± 6.84 mV

[210] The following assays were used in the examples described below. Any deviations from the protocols provided below are indicated in the examples. In these experiments, a spectrophotometer was used to measure the absorbance of the products formed after the completion of the reactions.

Hexokinase Assay For Measurement Of Residual Glucose

[211] Residual glucose from *H. jecorina* culture supernatants expressing CBH2 variants was measured using a hexokinase assay. A volume of 5 µl of supernatant was added to 195 µl glucose hexokinase assay (Instrumentation Laboratory, Breda, Netherlands) in a 96-well microtiterplate (Costar Flat Bottom PS). The plates were incubated at room temperature for 15 min. Following incubation, the absorbance was measured at 340 nm OD. Supernatants of cultures expressing residual glucose were excluded from pooling for further studies.

HPLC Assay for Protein Content Determination

[212] The concentration of CBH2 variant proteins from pooled culture supernatants was determined using an Agilent 1100 (Hewlet Packard) HPLC equipped with a Proswift RP 2H column (Dionex). Ten microliters of sample, mixed with 50 µl of 10% acetonitrile in filtered demineralized water was injected following equilibration of the HPLC column with 10% acetonitrile containing 0.01% trifluoroacetic acid. Compounds were eluted using a gradient of 10% to 30% acetonitrile from 0.3 min to 1 min, followed by a gradient of 30% to 65% from 1 min to 4 mins. Protein concentrations of CBH2 variants were determined from a calibration curve generated using purified wild-type CBH2 (6.25, 12.5, 25, 50 µg/ml). To calculate performance index (P_i or PI), the ratio of the (average) total protein produced by a variant and (average) total protein produced by the wild-type at the same dose were averaged.

Specific Activity Determination by Phosphoric Acid Swollen Cellulose (PASC) Hydrolysis Assay

[213] *Cellulose hydrolysis*: Phosphoric acid swollen cellulose (PASC) was prepared from Avicel according to a published method (Walseth, Tappi 35:228, 1971; and Wood, Biochem J, 121:353-362, 1971). This material was diluted with buffer and water to achieve a 1% w/v mixture such that the final concentration of sodium acetate was 50mM, pH 5.0. One hundred
5 microliters of a 1% suspension of PASC in 50 mM sodium acetate buffer (pH5.0) was dispensed in a 96-well microtiterplate (Costar Flat Bottom PS). Ten microliters of a 5 mg/ml culture supernatant from a CBH2 deleted strain was added to the PASC, and 5, 10, 15, or 20 μ l of pooled culture supernatants from *H. jecorina* cells expressing either wild-type CBH2 or CBH2 variants were added to it. Deletion of the CBH2 gene from *Hypocrea jecorina* (also
10 referred to as *Trichoderma reesei*) is described in U. S. Pat. Nos. 5,861,271 and 5,650,322. Compensating volumes of acetate buffer were added to make up for differences in total volume. The microtiterplate was sealed and incubated in a thermostatted incubator at 50°C under continuous shaking at 900 rpm. After two hours, the hydrolysis reaction was stopped by the addition of 100 μ l glycine buffer, pH 10 to each well. The hydrolysis reaction
15 products were analyzed with the PAHBAH assay.

[214] *PAHBAH assay*: Aliquots of 150 μ l of PAHBAH reducing sugar reagent (5% w/v p-hydroxybenzoic acid hydrazide (PAHBAH, Sigma # H9882, dissolved in 0.5 N HCl), (Lever, Anal Biochem, 47:273-279, 1972) were added to all wells of an empty microtiter plate. Ten
20 microliters of the hydrolysis reaction supernatants were added to the PABA reaction plate. All plates were sealed and incubated at 69°C under continuous shaking of 900 rpm. After one hour the plates were placed on ice for five minutes and centrifuged at 720 x g at room temperature for five minutes. Samples of 80 μ L of the developed PAHBAH reaction
mixtures were transferred to a fresh (read) plate and absorbance was measured at 410 nm in a spectrophotometer. A cellobiose standard was included as control. A dose response curve
25 was generated for wild-type CBH2 protein. To calculate performance index (P_i or PI), the ratio of the (average) total sugar produced by a variant and (average) total sugar produced by the wild-type at the same dose were averaged.

Specific Activity Determination By Hydrolysis Of Dilute Acid Pretreated Corn Stover (PCS)

30

[215] *Pretreated Corn Stover (PCS)*: Corn stover was pretreated with 2% w/w H_2SO_4 as described (Schell et al., J Appl Biochem Biotechnol, 105:69-86, 2003) and followed by multiple washes with deionized water to obtain a paste having a pH of 4.5. Sodium acetate buffer (pH 5.0) was then added to a final concentration of 50mM sodium acetate and, if

necessary, this mixture was then titrated to pH 5.0 using 1N NaOH. The cellulose concentration in the reaction mixture was approximately 7%. Sixty-five microliters of this cellulose suspension was added per well to a 96-well microtiterplate (Nunc Flat Bottom PS). Ten microliters of a 5 mg/ml culture supernatant from a CBH2 deleted strain was added to the PCS, and 5, 10, 15, or 20 μ l of pooled culture supernatants from *H. jecorina* cells expressing either wild-type CBH2 or CBH2 variants were added to it. Compensating volumes of acetate buffer were added to make up for differences in total volume. After sealing of the plate, the plates were placed in a thermostatted incubator at 50°C under continuous shaking of 1300 rpm for 5 minutes. The plates were then incubated at 50°C while shaking at 220 rpm under 80% humidity to prevent drying. After 7 days the plates were put on ice for 5 min and the hydrolysis reaction was stopped by the addition of 100 μ l glycine buffer, pH 10 to each well. The hydrolysis reaction products were analyzed with the PAHBAH assay.

[216] **PAHBAH assay:** Aliquots of 150 μ l of PAHBAH reducing sugar reagent (5% w/v p-hydroxybenzoic acid hydrazide (PAHBAH, Sigma # H9882, dissolved in 0.5 N HCl), (Lever, Anal Biochem, 47:273-279, 1972) were added to all wells of an empty microtiter plate. Ten microliters of the hydrolysis reaction supernatants were added to the PABA reaction plate. All plates were sealed and incubated at 69°C under continuous shaking of 900 rpm. After one hour the plates were placed on ice for five minutes and centrifuged at 720 x g at room temperature for five minutes. Samples of 80 μ L of the developed PAHBAH reaction mixtures were transferred to a fresh (read) plate and absorbance was measured at 410 nm in a spectrophotometer. A cellobiose standard was included as control. A dose response curve was generated for wild-type CBH2 protein. To calculate performance index (P_i or PI), the ratio of the (average) total sugar produced by a variant and (average) total sugar produced by the wild-type at the same dose were averaged.

Stability Of CBH2 Variants In Presence Of Ethanol

[217] The stability of wild-type CBH2 and CBH2 variants was tested in the presence of 4.5% ethanol (EtOH) at 49°C. Pooled culture supernatants (80 μ L) of *H. jecorina* cells expressing CBH2 variants were added to a 96-well plate (Greiner V-bottom PS) containing 10 μ l of 40.5% EtOH per well. The plates were sealed and incubated in a thermostatted incubator at 49°C for 16 hours with shaking at 900 rpm. Following incubation, the plates were placed on ice for 5 minutes. Residual CBH2 activity was determined using the phosphoric acid swollen cellulose (PASC) hydrolysis assay as described above.

[218] To calculate residual activity, the value of the product formed by the addition of 5, 10, 15 and 20 μ l of EtOH-incubated CBH2 to the residual activity PASC assay was divided by the value of the product formed by the addition of 5, 10, 15 and 20 μ l of EtOH-free CBH2 to the PASC assay. The individual values of these four ratios were then averaged to give the average residual activity. To determine PI value for the variant, the value of average residual activity for the variants was then divided by the average of the residual activity values of the wild-type CBH2 controls.

Thermostability Of CBH2 Variants

[219] The thermostability of wild-type CBH2 and CBH2 variants was tested at 53°C. Pooled culture supernatant (80 μ L) of *H. jecorina* cells expressing CBH2 variants were added to a 96-well plate (Greiner V-bottom PS). The plates were sealed and incubated in a thermostatted incubator at 53°C for 16 hours with shaking at 900 rpm. Following incubation, the plates were placed on ice for 5 minutes. Residual CBH2 activity was determined using the phosphoric acid swollen cellulose (PASC) hydrolysis assay as described above.

[220] To calculate residual activity, the value of the product formed by the addition of 5, 10, 15 and 20 μ l of heat-treated CBH2 to the residual activity PASC assay was divided by the value of the product formed by the addition of 5, 10, 15 and 20 μ l of unheated CBH2 to the PASC assay. The individual values of these four ratios were then averaged to give the average residual activity. To determine PI value for the variant, the value of average residual activity for the variants was then divided by the average of the residual activity values of the wild-type CBH2 controls.

EXAMPLE 2

Evaluation of Lignin Binding

[221] Lignin, a complex biopolymer of phenylpropanoid, is the chief non-carbohydrate constituent of wood that binds to cellulose fibers to harden and strengthen cell walls of plants. Because it is cross-linked to other cell wall components, lignin minimizes the accessibility of cellulose and hemicellulose to cellulose degrading enzymes. Hence, lignin is generally associated with reduced digestibility of all plant biomass. In particular the binding of cellulases to lignin reduces the degradation of cellulose by cellulases. Lignin is hydrophobic and apparently negatively charged. Thus the addition of negative charges to cellulases is contemplated to reduce their binding to lignin.

[222] As described herein a reaction was set up to measure the effect of chemical modification on the ability of a *Trichoderma* sp. cellulase preparation to bind a component of plant polymers, namely lignin. Lignin was recovered by extensive digestion of acid pretreated sugar cane bagasse by cellulases (100 mg Laminex BG/ g of cellulose) followed by hydrolysis of the cellulases by nonspecific serine protease exactly as described in Berlin et al. (Applied Biochemistry and Biotechnology, 121:163-170, 2005), except that sonication, drying, grinding and screening were not done and an acid wash (0.1 N HCl) to remove the protease followed by repeated washes with acetate buffer (50 mM sodium acetate pH 5) to return the sample to a pH 5 were added to the procedure. Briefly 50 μ L of 1.16% lignin (recovered from complete saccharification of bagasse) prepared in 50mM sodium acetate buffer at pH 5 was combined with 4 μ L of a desalted modified or an unmodified *Trichoderma* sp. cellulase preparation. Microfuge tubes containing the reaction mixture were incubated at room temperature for 1 hour, and then centrifuged at high speed to separate soluble from insoluble materials. Ten μ L of the supernatant from each tube was collected. The reaction tubes were re-mixed and incubated for an additional 2 hours after which second 10 μ L aliquots of the supernatant from each tube were collected. The supernatant samples were analyzed by SDS-PAGE. Reduction of the band intensity in modified *Trichoderma* sp. cellulase preparations was indicative of a reduction in lignin binding.

EXAMPLE 3

Evaluation of Bagasse Binding

[223] Bagasse is the biomass that remains after sugarcane has been crushed to extract its juice. A solution containing 2% cellulose of bagasse (acid treated, 28% solid, 57% glycan) was prepared in 50 mM sodium acetate at pH 5. Samples of unmodified or chemically-modified *Trichoderma* sp. cellulase preparations were diluted ten fold in the same sodium acetate buffer. Aliquots of the diluted enzymes were mixed with either bagasse solution or buffer alone and incubated for 1hr at room temperature. The supernatant was collected and assayed for activity of a component of cellulase, namely beta-glucosidase.

[224] Beta-glucosidase activity was measured using the chloro-nitro-phenyl-beta-D-glucoside (CNPG) assay. The CNPG assay is a kinetic assay in which β -glucosidase converts CNPG to the colored product 2-chloro-4-nitrophenol (CNP). OD is measured at 405 nm over a period of 10 minutes at 37°C. Rates are obtained as Vmax using the SpectraMax software and then converted to specific activity (μ M CNP/sec/mg Protein). Briefly, 200 μ L of

50 mM sodium acetate buffer pH 5.0 was added to each well of a 96-well microtiter plate. The plate was covered and placed in an Eppendorf Thermomixer at 37°C for 15 minutes to allow it to equilibrate to temperature. Five µl of the enzyme samples, serially diluted in 50 mM sodium acetate buffer, pH 5.0 were added to each well after equilibration. A 10 mM CNPG stock solution was diluted 1:5 using 50 mM sodium acetate buffer, pH 5.0, then 20 µl of the diluted CNPG solution (2 mM) was added to each well containing enzyme samples. The microtiter plate was transferred to a spectrophotometer (SpectraMAX, type 340; Molecular Devices) set at 37°C and OD was read at 405 nm for 0-15 min, reading at ≤ 9 sec intervals.

- 10 [225] The amount of beta-glucosidase activity of the cellulase enzyme samples that remained unbound to the bagasse substrate was considerably greater in the case of the chemically-modified *Trichoderma* sp. cellulase preparation. In particular as determined by the CNPG assay, less than 50% of the unmodified beta-glucosidase remained unbound (e.g., 50% bound) to the bagasse substrate, while nearly 80% of the modified bglu remained
- 15 unbound (e.g., 20% bound) to the bagasse substrate. Taken together the modified cellulase binding data indicate that reducing the positive charges on cellulase leads to reduced binding to a more negatively-charged plant polymer substrate. In this case the plant polymer substrate was lignin remaining in acid treated biomass. Acid treated biomass from corn stover, a plant biopolymer of similar chemical composition, was demonstrated to adopt an
- 20 increasingly more negative charged during the course of saccharification, as determined by measurement of zeta potential (See, Table 1-2).

EXAMPLE 4

Saccharification of Acid-Pretreated Bagasse

- 25 [226] Saccharification of cellulose present in acid-pretreated bagasse (APB) containing varying amounts of additional lignin was evaluated using chemically-modified and unmodified *Trichoderma* sp. cellulase preparations and assayed by HPLC to monitor release of sugars, DP1 to DP7. The results are shown on in FIG. 1 as percent conversion of the polymer substrate. In a microtiter plate, 200 µL of APB (3.5% glucan) was prepared in 50
- 30 mM sodium acetate buffer at pH 5, and adjusted to varying amounts of lignin. Twenty microliters of cellulase enzyme solution (unmodified or modified LAMINEX BG) was added to the wells. The plates were covered with aluminum plate sealers and placed in incubators at 50°C, with shaking, for 24 hrs or 48 hrs. The reaction was terminated by adding 100 µl 100

mM glycine pH 10 to each well. Following thorough mixing, the contents of the microtiter plate wells were filtered through a Millipore 96-well filter plate (0.45 μ m, PES). The filtrate was diluted into a plate containing 100 μ l 10 mM Glycine pH 10 and the amount of soluble sugars (DP1 through DP7) produced measured by HPLC. The Agilent 1100 series HPLC
5 was equipped with a de-ashing/guard column (Biorad Catalog No. 125-0118) and an Aminex lead based carbohydrate column (Aminex Catalog No. HPX-87P). The mobile phase used was water with a 0.6 ml/min flow rate. Soluble sugar standards (DP1-DP7) obtained from Sigma were all diluted in Milli-Q water to 100 mg/mL and used for converting peak area for the individual sugars to actual sugar concentrations. The percent of conversion was
10 calculated by dividing the sugars measured from HPLC by 100% conversion of cellulose to glucose.

[227] Cellulase binding to lignin will decrease its efficiency of degrading cellulose. This is demonstrated as a reduction in cellulose conversion in the presence of increasing amounts of lignin present in the saccharification reactions. This trend persists in the modified cellulase
15 preparations. However, there is a 10% increase in cellulose conversion in the modified cellulase samples as compared to unmodified cellulase samples. This result indicates that increasing negative charge of the cellulase reduces the nonproductive binding of cellulase to lignin.

EXAMPLE 5

Chemical Modified CBH2 Increased Saccharification of APB

[228] Purified Trichoderma CBH1, CBH2 variant, EG1, EG2 and beta-glucosidase were chemically modified as described in example 1. CBH2 variant used in this experiment has multiple substitutions (P98L/M134V/T154A/I2112V/S316P/S413Y with numbers
25 corresponding to the wild type mature CBH2 cellulase) as described in US Pub. No. 2006/0205042. The amino acid sequence of the mature CBH2 variant is as follows:

QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAA
STTSRVSPSTSRSSSATPPPGSTTTRVPPVSGGTATYSGNPFVGVTLWANAYYASEVS
SLAIPSLTGAMATAAAVAKVPSFVWLDTLDKTPLMEQTLADIRAANKNGGNYAG
30 QFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDVRTLLVIEPDS
LANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWPNQ
DPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPPYTQGNNAVYNEKLYIHA
IGPLLANHGWSN AFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLD

SFVWVKPGGECGTSDSSAPRFDYHCALPDALQPAPQAGAWFQAYFVQLLTNANPS
FL (SEQ ID NO:14).

[229] Chemical modifications of CBH1, CBH2 variant, EG1, EG2 and beta-glucosidase were verified by their shifted mobility on the native gel compared to the unmodified proteins.

5 Modified CBH1, CBH2 variant, EG1, EG2 and beta-glucosidase have more negative charges.

All the protein concentrations were measured using a NanoDrop™ spectrophotometer (Thermo). A saccharification reaction was set up in microtiter plate, in each well of a microtiter plate, 150 uL of APB (7% glucan prepared as described above for PCS) was prepared in 50 mM sodium acetate buffer at pH 5, 20μl of enzyme mix of 21μg total protein

10 was added so that the final protein to cellulose ratio in each well is 20mg/g. Six enzyme mixes were made by adding purified modified or unmodified Bglu, CBH2 variant, EG1, or EG2 to a *T. reesei* background in which the genes encoding cellobiohydrolase I (CBHI, Cel7a), cellobiohydrolase II (CBHII, Cel6a), endoglucanase I (EGI, Cel7b), and endoglucanase II (EGII, Cel5a) have been inactivated (See, US 2007/0128690). In each mix,

15 72.5% *T. reesei* background, 2.5% Bglu, 15% CBH2 variant, 5% EG1, 5% EG2 were added, the first four mixes have one protein that is not modified, the fifth mix has all the protein not modified, and the sixth mix has all the protein modified. The plate was incubated at 50C for 72 hours. The reaction was terminated by adding 100 μl 100 mM glycine pH 10 to each well.

Following thorough mixing, the contents of the microtiter plate wells were filtered through a
20 Millipore 96-well filter plate (0.45 μm, PES). The filtrate was diluted into a plate containing 100 μl 10 mM Glycine pH 10 and the amount of soluble sugars (DP1 through DP7) produced measured by HPLC. The Agilent 1100 series HPLC was equipped with a de-ashing/guard column (Biorad Catalog No. 125-0118) and an Aminex lead based carbohydrate column (Aminex Catalog No. HPX-87P). The mobile phase used was water with a 0.6 ml/min flow

25 rate. Soluble sugar standards (DP1-DP7) obtained from Sigma were all diluted in Milli-Q water to 100 mg/mL and used for converting peak area for the individual sugars to actual sugar concentrations. The percent of conversion was calculated by dividing the sugars measured from HPLC by 100% conversion of cellulose to glucose.

[230] FIG. 2 A shows that for the sixth enzyme mix (modified EG2, EG1, CBH2 variant
30 and beta-glucosidase) with all protein modified has the highest cellulose conversion, and fifth enzyme mix with all protein unmodified has the lowest conversion. Comparing the first four enzyme mixes, the second enzyme mix with the unmodified CBH2 gave next lowest conversion. FIG. 2B shows the advantages of modified proteins over unmodified proteins in cellulose conversion.

EXAMPLE 6**Preparation of *T. reesei* CBH2 Charge Ladder Variants**

[231] As determined during development of the present disclosure, succinylation of surface lysine residues of CBH2 improved performance on APB, and on pretreated corn stover. The charge of modified CBH2 variant was about -17 compared to unmodified CBH2 variant. With this in mind, a charge ladder of CBH2 was designed for determination of the optimal surface charge in cellulase performance applications.

[232] SEQ ID NO:1 sets forth the reference *Hypocrea jecorina* CBH2 coding DNA sequence:

10 atgattgtcggcattctcaccacgctggctacgtggccacactcgcagctagtgtgcctctagaggagcggcaagcttgctcaagcgt
ctggggccaatgtggtggccagaattggcgggtccgacttgctgtgcttcgggaagcacatgcgtctactccaacgactattactccc
agtgtcttcccggcgtgcaagctcaagctcgtccacgcgcgcgcgtcgacgacttctcgagtatccccacaacatcccggtcga
gtccgcgacgcctccacctggttctactactaccagagtacctccagtcggatcgggaaccgctacgtattcaggcaaccctttgtg
gggtcactccttggccaatgcatattacgcctctgaagttagcagcctcgtattcctagcttgactggagccatggccactgctgcag
15 cagctgtcgaaagggtccctctttatgtggctagatactcttgacaagaccctctcatggagcaaacccttgccgacatccgcaccg
ccaacaagaatggcggtaactatgccggacagtttgggtgatgacttgccggatcgcgattgcgctgcccttgccctgaatggcgaa
tactctattgccgatggtggcgtcgccaaatataagaactatcgacaccattcgtcaaattgctgtgaatattccgatatccggaccct
cctgggtattgagcctgactctcttgccaacctgggtaccaacctcgtactccaaagtgtgccaatgctcagtcagcctaccttgagtgc
atcaactacgccgtcacacagctgaaccttccaaatgttgcgatgtattggacgctggccatgcaggatggctggctggccggcaaa
20 ccaagaccggccgctcagctatttgcaaattgttacaagaatgcatcgtctccgagagctcttcgcggttggaaccaatgtcgcca
actacaacgggtggaacattaccagcccccatcgtacacgcaaggcaacgctgtctacaacgagaagctgtacatccacgctattg
gacctcttctgccaatcacggctgtgccaacgcctcttcatcactgatcaaggctgatcgggaaagcagcctaccggacagcaaca
gtggggagactggtgcaatgtgatcggcaccggatttggtattcgccatccgcaaacactggggactcgttgctggattcgtttgtctg
ggtcaagccaggcggcgagtgtgacggcaccagcgacagcagtgcgccacgatttgactcccactgtgcgtcccagatgccttgc
25 aaccggcgcctcaagctggtgcttgggtccaagcctactttgtgcagcttctcacaacgcaaacccatcgttctgtaa*

[233] SEQ ID NO:2 sets forth the *Hypocrea jecorina* CBH2 full length protein sequence:

MIVGILTTLATLATLAASVPLEERQACSSVWGQCGQNWSGPTCCASGSTCVYSNDYYSQCL
PGAASSSSSTRAASTTSRVSPITSRSSSATPPGSTTTTRVPPVSGGTATYSGNPFVGVTPWA
30 NAYYASEVSSLAIPSLTGAMATAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTKANGGNY
AGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLA
NLVTNLGTPKCANAQSAYLECINYAVTQNLNPNVAMYLDAGHAGWLGWPANQDPAAQLFANV
YKNASSPRALRGLATNVANYNGWNITSPPSYTQGNVYNEKLYIHAIGPLLANHGWSSNAFFI
TDQGRSGKQPTGQQQWGDWCNVI GTGFGIRPSANTGDSL LDSFVWVKPGGECDGTSDSSAPR
35 FDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL*

[234] SEQ ID NO:3 sets forth the *Hypocrea jecorina* CBH2 mature protein sequence:

QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPPTS
RSSHATPPPGSTTTRVPPVGSSTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAA
5 AAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYS
IADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLNLVTNLGTPKCANAQSAYLECINY
AVTQLNLPNVAMYLDAGHAGWLGWPNQDPAAQLFANVYKNASSPRALRGLATNVANYNGWN
ITSPPSYTQGNVYNEKLYIHAIGPLLANHGWSSNAFFITDQGRSGKQPTGQQQWGDWCNVIG
TGFGIRPSANTGDSLSDSFVWVKPGGECDDTSDSSAPRFDSDHCLPDALQPAPQAGAWFQAY
10 FVQLLTNANPSFL*

[235] Residues selected to be mutagenized included non-conserved, exposed lysine, arginine, asparagine, and glutamine residues, which were selected for substitution to introduce negative charges. Succinylated lysines in modified CBH2 were identified by mass spectrometry and selected for mutagenesis to glutamate, resulting in a -2 charge difference per substitution. Other residues were selected for substitution by analysis of CBH2 three-dimensional structure combined with amino acids alignment of homologous CBH2 sequences (See, e.g., Figure 3 of US Pub. No. US 2006/0205042, herein incorporated by reference). Surface residues that were highly variable in the CBH2 amino acid sequence alignment were candidates for mutagenesis. However, accumulation of substitutions in close proximity was avoided. Arginine was replaced with glutamine (charge -1), and glutamine and asparagine were substituted with the respective carboxyl variants (charge -1). In addition, aspartate and glutamate residues were selected for substitution to the respective amine residues for completion of the charge ladder (charge +1). Specific CBH2 substitutions are shown in Table 6-1, with all positions shown with the exception of R63 and R77, located in the CBH2 catalytic domain. A net positive charge can be created by either removal of a negatively charged residue or by introduction of a positively charged residue. Likewise a net negative charge can be created by either removal of a positively charged residue or by introduction of a negatively charged residue.

Table 6-1. CBH2 Substitutions					
Introduction of negative charge and removal of	Removal of positive charge	Introduction of negative charge	Introduction of negative charge	Removal of negative charge	Removal of negative charge

Table 6-1. CBH2 Substitutions					
positive charge					
Lysine	Arginine	Asparagine	Glutamine	Aspartate	Glutamate
K157E	R153Q	N382D	Q204E	D189N	E208Q
K129E	R294Q	N344D	Q147E	D211N	E244Q
K288E	R203Q	N237D	Q239E	D405N	E146Q
K194E	R378Q	N339D	Q281E	D277N	
K356E	R63Q	N289D		D151N	
K327E	R77Q	N161D			
		N285D			
		N197D			
		N254D			
		N247D			

[236] For preparation of a CBH2 charge ladder, ten CBH2 charge variants (C-1 to C-10) were designed spanning a charge range of +8 to -32, compared to the wild-type CBH2 in steps of 4 as shown in Table 6-2.

5

Table 6-2. CBH2 Charge Ladder										
C-1	C-2	CBH2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
8	4	0	-4	-8	-12	-16	-20	-24	-28	-32
D189N	D189N	WT	K157E	K157E	K157E	K157E	K157E	K157E	K157E	K157E
E208Q	E208Q		K129E	K129E	K129E	K129E	K129E	K129E	K129E	K129E
D211N	D211N			K288E	K288E	K288E	K288E	K288E	K288E	K288E
D405N	D405N			K194E	K194E	K194E	K194E	K194E	K194E	K194E
E244Q					K356E	K356E	K356E	K356E	K356E	K356E
D277N					K327E	K327E	K327E	K327E	K327E	K327E
D151N						R153Q	R153Q	R153Q	R153Q	R153Q
E146Q						R294Q	R294Q	R294Q	R294Q	R294Q
						R203Q	R203Q	R203Q	R203Q	R203Q
						R378Q	R378Q	R378Q	R378Q	R378Q
							N382D	N382D	N382D	N382D
							N344D	N344D	N344D	N344D
							N237D	N237D	N237D	N237D

Table 6-2. CBH2 Charge Ladder										
							N339D	N339D	N339D	N339D
								N289D	N289D	N289D
								N161D	N161D	N161D
								Q204E	Q204E	Q204E
								Q147E	Q147E	Q147E
									N285D	N285D
									N197D	N197D
									N254D	N254D
									N247D	N247D
										Q239E
										Q281E
										R63Q
										R77Q

[237] The amino acid sequences of the variants were back translated to DNA and codon optimized for expression in *Trichoderma reesei* using GeneDesigner software (DNA2.0). The codon-optimized cbh2 variant genes were synthesized and the DNA of the CBH2 surface charge variants (SCVs) was amplified from the DNA2.0 constructs by PCR with using primers: GGHTK22 forward 5'-CACCATGATCGTGGGAATTCTTACTACTC-3' (SEQ ID NO:15); and GGTHK23 reverse 5'-CTACAAAAACGAAGGGTTCGCATT-3' (SEQ ID NO:16). In one experiment, site directed mutagenesis was used to introduce K129E and K157E mutations (cbh2 charge variant C-3) in the genomic DNA of wild type CBH2. CBH2 charge variant C-3 was cloned into pTrex3GM and expressed as described below.

[238] The PCR products were purified and cloned into pENTR/TOPO for transformation of *E. coli* TOP10 cells. Plasmid DNA was isolated from single colonies and the correct sequence was verified. CBH2 SCVs were cloned into pTrex3GM, pTTTpyr(pcbh1), and pTTTpyr(pstp1) as shown in Table 6-3.

Table 6-3. Expression Clones of CBH2 Surface Charge Variants			
CBH2 Variant	Destination vectors		
	pTrex3gM	pTTTpyr (P _{cbh1})	pTTTpyr(P _{stp1})
C-1 (pTK354a)	1	11	21
C-2 (pTK355a)	2	12	22

C-3 (pTK356a)	3	13	23
C-4 (pTK357a)	4	14	24
C-5 (pTK358a)	5	15	25
C-8 (pTK361a)	6	16	26
C-6 (pTK359a)	(7)	(17)	(27)
C-7 (pTK360a)	8	18	28
C-9 (pTK362a)	9	19	29
C-10 (pTK363b)	10	20	30

[239] CBH2 surface charge variants (SCV) in *T. reesei*. Biolistic transformation of *T. reesei* with the pTrex3gM expression vector containing the cbh2 charge variant C3 (with K129E and K157E mutations) open reading frame was performed using the following

5 protocol. *T. reesei* in which the genes encoding cellobiohydrolase I (CBHI, Cel7a), cellobiohydrolase II (CBHII, Cel6a), endoglucanase I (EGI, Cel7b), and endoglucanase II (EGII, Cel5a) have been inactivated was used.). Transformation of the *Trichoderma reesei* strain by the biolistic transformation method was accomplished using a Biolistic® PDS-1000/he Particle Delivery System from Bio-Rad (Hercules, CA) following the manufacturer's

10 instructions (See WO 05/001036 and US 2006/0003408). Transformants were transferred to new acetamide selection plates. Stable transformants were inoculated into filter microtiter plates (Millipore), containing 200ul/well of glycine minimal media (6.0 g/L glycine; 4.7 g/L (NH₄)₂SO₄; 5.0 g/L KH₂PO₄; 1.0 g/L MgSO₄·7H₂O; 33.0 g/L PIPPS; p.H. 5.5) with post sterile addition of ~2% glucose/sophorose mixture as the carbon source, 10 ml/L of 100g/L of

15 CaCl₂, 2.5 ml/L of *T. reesei* trace elements (400X): 175g/L Citric acid anhydrous; 200g/L FeSO₄·7H₂O; 16g/L ZnSO₄·7H₂O; 3.2 g/L CuSO₄·5H₂O; 1.4 g/L MnSO₄·H₂O; 0.8 g/L H₃BO₃. Transformants were grown in liquid culture for 5 days in O₂ rich chamber housed in a 28° C incubator. The supernatant samples from the filter microtiter plate were obtained by using a vacuum manifold. Samples were run on 4-12% NuPAGE gels (Invitrogen) according

20 to the manufactures instructions. The gel was stained with Simply Blue stain (Invitrogen). Expression of additional CHB2 surface charge variants may be accomplished using this method.

EXAMPLE 7

Generation of CBH2 Charge Variants in *T. reesei*

[240] The pTTTPyr-cbh2 plasmid containing the *Hypocrea jecorina* CBH2 protein encoding sequence (SEQ ID NO:1) was sent to BASEClear (Leiden, The Netherlands), GeneArt AG (Regensburg, Germany), and Sloning BioTechnology GmbH (Puchheim, Germany) for the generation of Site Evaluation Libraries (SELs). The plasmid map of pTTTPyr-cbh2 is shown in Figure 5. A request was made to the vendors for the generation of positional libraries at each of the sites in *Hypocrea jecorina* CBH2 mature protein (SEQ ID NO:3). The amino acid sequence of CBH2 full length protein is shown in SEQ ID NO:2.

[241] SEQ ID NO:1 sets forth the reference *Hypocrea jecorina* CBH2 coding DNA sequence:

10 atgattgtcggcattctcaccacgttggtacgtggccacactcgcagctagtgtgcctctagaggagcggcaagcttgctcaagcgt
ctggggccaatgtggtggccagaattggtcgggtccgacttgctgtgcttccggaagcacatgcgtctactccaacgactattactccc
agtgtcttcccggcgctgcaagctcaagctcgtccacgcgcgcgcgtcgacgacttctcgagtatccccacacatcccggtcga
gctccgcgacgcctccacctggttctactactaccagagtacctccagtcggatcgggaaccgctacgtattcaggcaaaccttttgttg
gggtcactccttgggccaatgcatattacgcctctgaagttagcagcctcgctattcctagcttgactggagccatggccactgctgcag
15 cagctgtcgaaggttcctcttttatgtggctagatactcttgacaagaccctctcatggagcaaaccctggccgacatccgcaccg
ccaacaagaatggcggtaactatgccggacagtttggtgtatgacttgccggatcgcgattgcgctgcccttgccctgaatggcgaa
tactctattgccgatggtggcgtcgccaaatataagaactatcgacaccattcgtaaatgtcgtggaatattccgatatccggaccct
cctggttattgagcctgactctcttgccaacctgggtgaccaacctcggtactccaaagtgtgccaatgctcagtcagcctaccttgagtgc
atcaactacgccgtcacacagctgaaccttccaaatgttgcatgtatttgacgctggccatgcaggatggcttggtggccggcaaa
20 ccaagaccggcgcgtcagctatttgcaaatgtttacaagaatgcatcgtctccgagagctcttcgcggattggcaaccaatgtcgcca
actacaacgggtggaacattaccagcccccatcgtaacgcaaggcaacgctgtctacaacgagaagctgtacatccacgctattg
gacctcttcttgccaatcacggctggtccaacgccttctcatcactgatcaaggctgatcgggaaagcagcctaccggacagcaaca
gtggggagactggtgcaatgtgatcggcaccggatttggtattcgcccatccgcaaacactggggactcgttgctggattcgtttgtctg
ggtcaagccaggcggcgagtgtgacggcaccagcgacagcagtgccacgatttgactcccactgtgcgctcccagatgccttgc
25 aaccggcgcctcaagctggtgcttggttccaagcctactttgtgcagcttctcacaacgcaaacccatcgttcctgtaa*

[242] SEQ ID NO:2 sets forth the *Hypocrea jecorina* CBH2 full length protein sequence:

MIVGILTTLATLATLAASVPLEERQACSSVWQCGGQNSGPTCCASGSTCVYSNDYYSQCL
PGAASSSSSTRAASTTSRVSPITSRSSSATPPPGSTTTRVPPVGSATYSGNPFVGVTPWA
30 NAYYASEVSSLAIPSLTGAMATAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNY
AGQFVVYDLPRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLA
NLVTNLGTPKCANASAYLECINYAVTQLNLPNVAMYLDAGHAGWLGPANQDPAAQLFANV
YKNASSPRALRGLATNVANYNGWNITSPPSYTQGNVYNEKLYIHAIGPLLANHGWSNAFFI
TDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLDSFVWVKPGGECDDGTSDSSAPR
35 FDSHCALPDALQAPQAGAWFQAYFVQLLTNANPSFL*

[243] SEQ ID NO:3 sets forth the *Hypocrea jecorina* CBH2 mature protein sequence:

QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPPTS
 RSSSATPPPGSTTTTRVPPVGSSTATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAA
 5 AAVAKVPSFMWLDLTDLKTPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYS
 IADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLNLVTNLGTPKCANAQSAYLECIN
 AVTQLNLPNVAMYLDAGHAGWLGWPNQDPAAQLFANVYKNASSPRALRGLATNVANYNGWN
 ITSPPSYTQGNVYNEKLYIHAIGPLLANHGWSNFAFFITDQGRSGKQPTGQQQWGDWCNVIG
 TGFGIRPSANTGDSLDSFVWVKPGGECGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAY
 10 FVQLLTNANPSFL*

[244] Purified pTTTpyr-cbh2 plasmids (p_{cbh1}, Amp^R, Acetamide^R) containing open reading frames encoding CBH2 variant sequences were obtained from the vendors specified above.

Protoplasts of *H. jecorina* strain (Δ eg1, Δ eg2, Δ cbh1 Δ cbh2) were transformed with the
 15 pTTTpyr constructs and grown on selective agar containing acetamide at 28°C for 7 days. Briefly, biolistic transformation of *H. jecorina* was performed using the following protocol and a strain in which the genes encoding cellobiohydrolase I (CBHI, Cel7a), cellobiohydrolase II (CBHII, Cel6a), endoglucanase I (EGI, Cel7b), and endoglucanase II (EGII, Cel5a) have been inactivated. Transformation of the *H. jecorina* by the biolistic
 20 transformation method was accomplished using a Biolistic® PDS-1000/he Particle Delivery System from Bio-Rad (Hercules, CA) following the manufacturer's instructions (See WO 05/001036 and US 2006/0003408). Spores were harvested, replated on acetamide agar, and incubated at 28°C for 7 days. Spores were harvested in 15% glycerol and stored at -20°C for further use. For CBH2 variant protein production, a volume of 10 μ l spore suspension was
 25 added to 200 μ l glycine minimal medium supplemented with 2% glucose/sophorose mixture in a PVDF filter plate: 6.0 g/L glycine, 4.7 g/L (NH₄)₂SO₄; 5.0 g/L KH₂PO₄; 1.0 g/L MgSO₄·7H₂O; 33.0 g/L PIPPS; pH 5.5; with post sterile addition of ~2% glucose/sophorose mixture as the carbon source, 10 ml/L of 100g/L of CaCl₂, 2.5 ml/L of *T. reesei* trace elements (400X): 175g/L Citric acid anhydrous; 200g/L FeSO₄·7H₂O; 16g/L ZnSO₄·7H₂O;
 30 3.2 g/L CuSO₄·5H₂O; 1.4 g/L MnSO₄·H₂O; 0.8 g/L H₃BO₃. Each CBH2 variant was grown in quadruplicate. After sealing the plate with an oxygen permeable membrane, the plates were incubated at 28°C for 6 days, while shaking at 220 rpm. Supernatant was harvested by transferring the culture medium to a microtiter plate under low pressure and tested for residual glucose using the hexokinase assay as described in Example 1.

35

EXAMPLE 8

Expression, Activity and Performance of CBH2 Variants

[245] *H. jecorina* CBH2 charge variants were tested for various properties of interest. In particular, the cellulase variants were tested for protein expression using the HPLC assay (HPLC), specific activity using the PASC hydrolysis assay (Act. PASC) and the PCS hydrolysis assay (Act. PCS), stability in the presence of ethanol (EtOH ratio) and thermostability (heat ratio) as described in Example 1. Performance data for CBH2 charge variants are shown in Table 8-1. Performance index (PI) is the ratio of performance of the variant cellulase to the parent or reference cellulase. Various terms set forth below are used to describe the mutation: up mutations have a $PI > 1$; neutral mutations have a $PI \geq 0.5$, non-deleterious mutations have a $PI > 0.05$; deleterious mutations have a $PI = 0.05$; combinable mutations are those mutations for which the variant has Performance index values ≥ 0.5 for at least one property. Combinable mutations are mutations that can be combined to deliver proteins with appropriate performance indices for one or more desired properties. Positions at which mutations occur are classed as follows: Non-restrictive positions have $\geq 20\%$ neutral mutations for at least one property; and Restrictive positions have $< 20\%$ neutral mutations for activity and stability. Fully Restrictive positions have no neutral mutations for activity or stability.

[246] These data may be used to engineer any CBH2. Even if the CBH2 to be engineered has an amino acid different from that of *Hypocrea jecorina* CBH2 at a particular position, these data may be used to find a substitution that will alter the desired properties by identifying the best choice substitution, including substitution to the *Hypocrea jecorina* CBH2 wild type amino acid.

[247] Table 8-11 shows performance index values (Pi or PI) for variants of *Hypocrea jecorina* CBH2. Performance indices less than or equal to 0.05 were fixed to 0.05 and indicated in bold italics in the table.

Table 8-1: Performance Indexes of CBH2 Charge Variants					
Variant	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
R63A	1.21	1.03	0.94	0.97	1.00
R63C	1.47	1.02	1.05	0.96	1.06
R63D	0.12	0.11	0.19	0.05	1.02
R63E	0.88	0.94	0.98	0.96	1.05
R63F	0.91	0.95	0.85	0.94	1.09
R63G	0.71	0.96	0.95	1.08	1.20

Table 8-1: Performance Indexes of CBH2 Charge Variants

Variant	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
R63I	0.81	0.93	1.11	1.03	1.09
R63L	0.94	0.96	0.91	1.06	1.29
R63M	1.22	0.99	1.10	1.05	1.09
R63N	1.23	1.02	1.11	1.03	1.07
R63P	0.96	0.99	1.17	1.04	1.09
R63Q	0.87	0.99	1.01	0.99	1.13
R63S	0.59	0.92	1.09	1.00	0.94
R63T	0.43	0.87	0.94	0.91	0.84
R63V	1.00	0.99	0.89	1.03	1.06
R63W	0.98	0.97	0.91	1.08	1.10
R63Y	1.29	1.03	0.99	1.06	1.07
R77F	0.25	0.81	0.58	0.92	0.81
R77G	0.41	1.02	0.75	1.00	0.93
R77L	0.14	0.53	0.36	0.94	0.79
R77N	0.83	0.99	0.87	1.00	1.02
K129A	0.23	1.03	0.81	0.74	0.57
K129L	0.80	0.94	0.80	0.82	0.71
K129N	0.15	0.93	0.65	0.73	0.61
K129Q	0.41	1.30	0.89	1.01	0.99
K129S	0.37	1.05	0.92	0.84	0.84
K129T	0.16	0.87	0.37	0.81	0.69
K129V	0.17	0.93	0.91	0.63	0.45
K129Y	0.35	0.95	0.82	0.68	0.46
Q147E	0.86	1.00	0.88	1.07	1.05
R153A	0.23	0.50	0.72	0.63	0.64
R153C	0.26	0.51	0.86	0.94	0.88
R153D	0.35	0.35	0.54	0.75	0.72
R153E	0.27	0.69	0.75	0.79	0.80
R153F	0.21	0.38	0.69	0.68	0.58
R153G	0.21	0.31	0.59	0.62	0.92
R153I	0.22	0.28	0.37	0.63	1.09
R153L	0.19	0.30	0.40	0.94	0.44
R153M	0.31	0.70	0.86	0.92	1.05
R153N	0.22	0.20	0.42	1.12	1.23
R153P	0.23	0.14	0.10	0.39	2.17
R153Q	0.24	0.52	0.99	0.77	1.05
R153S	0.41	0.78	0.99	0.93	0.93
R153T	0.31	0.63	0.74	0.84	0.94
R153V	0.23	0.28	0.59	0.63	0.81
R153W	0.42	0.77	1.07	0.90	0.87
R153Y	0.32	0.60	0.73	0.84	0.90
K157A	0.68	1.04	0.93	1.01	0.91
K157D	0.69	0.96	1.01	1.12	0.90
K157E	0.10	0.92	0.75	1.03	0.73
K157F	0.40	1.01	0.94	1.07	0.95

Table 8-1: Performance Indexes of CBH2 Charge Variants

Variant	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
K157G	0.50	0.98	0.76	1.04	0.93
K157I	0.71	1.00	0.81	1.05	0.94
K157L	0.18	0.88	0.60	1.04	0.75
K157M	0.33	0.99	0.77	0.99	0.88
K157P	0.11	0.57	0.63	3.04	1.86
K157Q	0.50	1.00	0.70	1.00	0.88
K157T	0.14	0.79	0.47	1.02	0.83
K157V	0.46	0.99	0.86	0.99	0.91
K157W	1.06	0.95	0.84	1.04	0.93
K157Y	0.77	0.98	0.92	1.00	0.84
N161E	0.76	1.03	1.07	1.12	1.12
D189A	0.29	0.77	0.73	0.79	0.69
D189C	0.38	0.89	0.82	0.90	0.76
D189E	0.44	0.98	0.84	0.94	0.89
D189F	0.39	0.82	0.64	0.94	0.99
D189G	0.27	0.87	0.86	0.88	0.71
D189H	0.39	0.91	0.52	0.91	0.83
D189I	0.16	0.52	0.39	0.63	0.61
D189K	0.37	0.96	0.78	0.89	0.73
D189L	0.33	1.04	0.82	0.89	0.85
D189N	0.55	0.95	1.05	1.00	0.96
D189P	0.12	0.30	0.22	0.72	0.76
D189Q	0.73	0.98	0.72	0.93	0.86
D189R	0.38	0.96	0.82	0.94	0.88
D189S	0.14	0.60	0.32	0.86	1.22
D189T	0.21	0.84	0.84	0.85	0.76
D189V	0.36	1.03	0.99	1.03	0.91
D189W	0.13	0.18	0.34	0.45	1.06
D189Y	0.32	0.52	0.46	1.30	1.34
K194A	0.30	0.67	0.86	0.95	1.00
K194C	0.22	1.08	1.72	0.88	0.88
K194D	0.61	0.98	1.07	1.00	1.03
K194E	0.44	1.00	1.14	1.02	1.14
K194F	0.67	1.03	0.94	1.04	1.13
K194G	0.43	0.92	0.65	0.92	0.96
K194I	0.24	0.38	0.49	0.50	0.59
K194L	0.61	0.98	0.90	1.01	1.22
K194M	0.61	1.02	0.86	1.03	1.17
K194N	0.54	0.99	1.02	1.04	1.17
K194P	0.32	0.39	0.55	0.46	0.38
K194Q	0.64	1.00	1.00	1.03	1.17
K194S	0.66	1.00	0.83	1.12	1.03
K194T	0.08	0.11	1.09	1.17	9.80
K194V	0.23	0.30	0.71	0.55	0.62
K194W	0.43	0.93	0.98	0.98	1.01

Table 8-1: Performance Indexes of CBH2 Charge Variants

Variant	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
K194Y	0.08	0.11	0.36	0.60	2.65
N197D	1.50	1.12	1.08	1.06	1.07
R203A	0.71	0.93	0.75	0.69	0.88
R203F	0.85	0.92	0.99	0.67	0.87
R203G	0.60	1.07	1.13	0.66	1.39
R203I	0.39	0.70	0.64	1.46	3.02
R203L	0.46	1.14	0.54	1.08	1.24
R203M	0.58	0.88	0.91	0.99	1.25
R203N	0.36	1.08	0.82	1.06	1.20
R203P	0.29	0.49	0.28	2.85	4.58
R203Q	0.80	0.99	0.78	0.84	0.86
R203S	0.54	1.19	0.53	1.99	1.96
R203T	0.66	1.04	0.82	0.99	0.91
R203V	0.85	0.94	0.89	0.73	0.78
R203W	0.39	1.10	0.63	2.28	2.42
R203Y	0.57	1.09	1.03	1.03	1.18
Q204D	1.23	1.04	0.73	1.00	1.00
Q204E	1.08	1.12	1.05	0.93	1.04
Q239D	0.59	0.98	0.91	0.92	0.92
Q239E	0.70	1.10	1.16	1.00	1.04
N247D	1.18	0.76	0.67	0.58	0.96
N254E	0.20	0.84	0.56	0.76	0.91
Q281D	1.10	0.60	0.43	1.33	0.95
Q281E	2.50	0.89	0.85	1.10	0.63
N285D	0.35	0.70	0.59	1.99	2.24
K288A	0.41	0.61	0.68	1.52	2.84
K288C	0.41	0.95	0.63	0.72	1.49
K288D	0.23	0.23	0.05	5.77	5.10
K288E	0.38	0.89	0.81	0.65	1.60
K288F	0.36	0.75	0.29	1.53	2.74
K288G	0.23	0.40	0.12	2.22	4.09
K288H	0.32	0.90	0.48	1.09	1.54
K288I	0.40	0.75	0.37	1.51	2.48
K288L	0.80	0.96	0.81	0.60	0.82
K288N	0.71	0.98	0.84	1.08	1.68
K288P	0.26	0.34	0.05	5.00	5.39
K288Q	0.23	0.80	0.18	1.23	2.51
K288S	0.19	0.71	0.21	2.02	2.72
K288T	0.25	0.71	0.35	1.27	2.96
K288V	0.31	0.84	0.38	1.67	1.87
N289D	0.84	0.96	0.87	0.93	0.92
N289E	0.13	0.26	0.63	1.03	1.48
R294A	0.97	1.06	0.89	0.92	0.77
R294C	0.70	1.15	0.74	1.01	1.16
R294D	0.75	0.94	0.85	0.94	1.09

Table 8-1: Performance Indexes of CBH2 Charge Variants

Variant	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
R294E	0.73	0.82	0.66	1.02	1.20
R294F	0.37	0.52	0.54	0.90	1.56
R294G	1.25	0.93	0.84	0.71	0.87
R294I	0.77	0.84	0.72	0.98	1.01
R294L	0.54	1.06	1.03	1.05	1.22
R294M	0.34	0.76	0.78	0.81	1.78
R294N	0.50	0.80	0.64	1.30	1.06
R294P	0.48	0.86	0.74	0.74	0.99
R294Q	0.50	0.79	0.75	1.23	1.30
R294S	0.64	1.00	0.57	1.32	1.47
R294T	1.12	1.03	0.97	1.02	1.10
R294V	0.83	0.96	0.83	0.82	1.03
R294W	1.08	1.01	0.57	0.56	0.58
R294Y	0.26	0.43	0.31	2.39	3.19
K327A	0.86	1.03	0.98	0.76	0.76
K327C	0.42	0.93	1.25	1.28	1.02
K327D	0.31	0.73	0.76	1.13	1.81
K327E	0.73	0.84	0.85	1.05	0.87
K327F	0.24	0.61	0.44	1.91	2.14
K327G	0.24	0.83	0.16	1.65	1.30
K327I	0.31	0.74	0.23	1.44	1.37
K327L	0.26	0.94	0.60	0.89	1.30
K327M	0.41	0.90	0.90	0.87	1.44
K327N	0.41	0.90	1.04	0.74	1.33
K327P	0.21	0.58	0.05	1.65	2.77
K327Q	1.32	1.06	1.42	1.15	0.77
K327S	0.20	0.63	0.58	2.39	2.92
K327T	0.21	0.29	0.09	3.69	7.57
K327V	0.28	0.86	0.58	0.66	1.26
K327W	0.28	1.05	0.74	0.93	1.57
K327Y	0.93	1.04	1.08	0.49	0.60
N339D	0.92	1.04	1.05	1.04	1.10
N339E	0.94	1.03	0.83	1.08	1.13
N344D	0.29	0.85	1.10	1.43	1.31
K356A	1.22	1.03	0.88	0.23	0.25
K356C	0.62	0.92	0.98	0.29	0.45
K356D	0.42	0.87	0.57	0.22	0.42
K356E	0.43	0.98	0.71	0.28	0.43
K356F	0.69	0.99	0.45	0.15	0.39
K356G	0.97	0.98	0.20	0.20	0.40
K356I	0.76	0.96	0.60	0.26	0.46
K356L	0.67	1.00	0.68	0.45	0.39
K356M	0.68	0.99	0.88	0.54	0.53
K356N	0.98	1.04	0.72	0.19	0.37
K356P	0.23	0.52	0.33	1.09	0.99

Table 8-1: Performance Indexes of CBH2 Charge Variants

Variant	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
K356Q	0.77	0.99	0.77	0.41	0.43
K356S	0.58	0.88	0.48	0.22	0.41
K356T	0.63	1.00	0.49	0.28	0.43
K356V	0.91	0.90	0.75	0.35	0.31
K356W	0.39	0.90	0.42	0.24	0.47
K356Y	0.52	0.96	0.47	0.20	0.51
R378A	0.21	0.76	0.79	0.25	0.54
R378C	0.25	0.34	0.42	0.74	1.01
R378D	0.19	0.23	0.33	0.49	0.58
R378E	0.17	0.18	0.30	0.92	0.94
R378F	0.13	0.39	1.34	0.12	0.38
R378G	0.08	0.22	1.48	0.19	0.82
R378I	0.19	0.42	0.33	0.38	0.69
R378L	0.21	0.75	0.79	0.33	0.53
R378M	0.14	0.52	0.77	0.37	0.47
R378N	0.12	0.19	0.48	0.38	0.81
R378P	0.17	0.65	0.54	0.48	0.62
R378Q	0.20	0.82	0.94	0.56	0.59
R378S	0.18	0.65	0.47	0.45	0.64
R378T	0.18	0.52	1.03	0.25	0.43
R378V	0.36	1.01	0.98	1.06	1.19
R378W	0.09	0.10	0.64	0.34	0.30
R378Y	0.24	0.60	0.62	0.23	0.53
N382D	0.71	0.81	N.D.	1.17	1.15
N382E	2.21	0.16	N.D.	6.19	5.56
D405A	1.75	0.82	0.33	0.32	0.60
D405C	0.45	0.63	0.57	0.72	0.97
D405E	0.85	0.81	0.63	0.50	0.65
D405F	0.79	0.95	0.24	0.56	0.82
D405G	0.77	0.87	0.16	0.56	0.86
D405H	0.27	0.38	0.25	3.35	3.59
D405I	0.62	1.04	0.49	0.83	1.02
D405K	0.63	0.65	0.24	0.43	0.72
D405L	0.41	0.95	0.76	0.38	0.43
D405M	0.52	1.31	0.58	0.59	0.69
D405N	0.74	1.04	0.78	1.34	1.32
D405P	0.77	0.80	0.19	0.99	1.02
D405Q	0.71	1.12	0.39	1.10	1.21
D405R	0.93	1.20	0.36	0.56	0.61
D405S	0.95	1.24	0.40	0.75	0.96
D405T	0.74	1.17	0.32	0.64	0.83
D405V	0.39	1.05	0.60	1.32	1.24
D405W	1.12	1.17	0.40	0.42	0.48
D405Y	1.56	1.07	0.39	0.31	0.07

EXAMPLE 9

Effect of Charge Change on the Activity of CBH2 Variants

[248] In this example, the effect of charge change on the activity of CBH2 in a pretreated corn stover (PCS) assay of cellulase activity was assessed. Briefly, the number of PCS winners in the CBH2 SELs was determined as a property of net charge change. In Table 9-1, the ratio of observed to expected (o/e) winners was determined in the PCS assay. Values in bold italics are significantly different from the average of 10 random distributions plus or minus the number of standard deviations (sd) listed in the respective columns.

Table 9-1: Charge Effect on Activity of CBH2 Variants in a PCS Assay				
PCS	o/e 1 sd	o/e 2 sd	o/e 3 sd	Results
Charge change				
-2.00	<i>1.20</i>	<i>1.20</i>	<i>1.20</i>	>90% confident more than expected
-1.00	<i>1.40</i>	<i>1.40</i>	<i>1.40</i>	>99% confident more than expected
0.00	<i>1.03</i>	<i>1.03</i>	<i>1.03</i>	as expected
1.00	<i>0.46</i>	<i>0.46</i>	<i>0.46</i>	>95% confident less than expected
2.00	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	as expected

[249] As shown in Table 9-1 and Figure 4, decreasing charge (e.g., -1, -2) results in a significantly higher frequency of CBH2 winners in the PCS assay, while increasing charge (e.g., +1) results in a significantly lower frequency of CBH2 winners in the PCS assay. In conclusion, CBH2 activity on PCS correlates with decrease in charge.

CLAIMS

We claim:

- 5 1. An isolated cellulase variant, wherein said variant is a mature form having cellulase activity and comprising a substitution at one or more positions selected from the group consisting of 63, 77, 129, 147, 153, 157, 161, 194, 197, 203, 237, 239, 247, 254, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, and 382, wherein the positions are numbered by
10 correspondence with the amino acid sequence of a reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3, and wherein the substitution at one or more positions causes the cellulase variant to have a more negative net charge in comparison to the reference CBH2.
2. The isolated cellulase variant of Claim 1, wherein the substitution at one or more
15 positions comprises removal of one or more positive charges.
3. The isolated cellulase variant of Claim 2, wherein the removal of one or more
 positive charges comprises a replacement of a lysine or an arginine with a neutral amino acid.
4. The isolated cellulase variant of Claim 1, wherein the substitution at one or more
20 positions comprises addition of one or more negative charges.
5. The isolated cellulase variant of Claim 4, wherein the addition of one or more
25 negative charges comprises a replacement of a neutral amino acid with a negatively charged amino acid.
6. The isolated cellulase variant of Claim 1, wherein the substitution at one or more
 positions comprises removal of one or more positive charges and addition of one or more
 negative charges.
- 30 7. The isolated cellulase variant of Claim 6, wherein the removal of one or more
 positive charges and addition of one or more negative charges comprises a placement of a
 lysine or an arginine with a negatively charged amino acid.

8. The isolated cellulase variant of Claim 1, wherein the substitution at one or more positions comprises one or more of the group consisting of K129E, K157E, K194E, K288E, K327E, K356E, R63Q, R77Q, R153Q, R203Q, R294Q, R378Q, N161D, N197D, N237D, N247D, N254D, N285D, N289D, N339D, N344D, N382D, Q147E, Q204E, Q239E, and
5 Q281E, wherein the positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

9. The isolated cellulase variant of Claim 1, wherein the variant comprises a further substitution at one or more further positions selected from the group consisting of 146, 151,
10 189, 208, 211, 244, 277 and 405, wherein the further positions are numbered by correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

10. The isolated cellulase variant of Claim 9, wherein the further substitution at one
15 or more further positions comprises a replacement of aspartic acid or glutamic acid with a neutral amino acid.

11. The isolated cellulase variant of Claim 9, wherein the further substitution at one or more further positions comprises one or more of the group consisting of D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, and E244Q, wherein the positions are numbered by
20 correspondence with the amino acid sequence of the reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

12. The isolated cellulase variant of Claim 1, wherein the substitution at one or more
25 positions is selected from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 positions.

13. The isolated cellulase variant of Claim 1, wherein the cellulase variant is derived from a parent cellulase selected from the group consisting of *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2,
30 *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, and *Cellulomonas fimi* CenA EG.

14. The isolated cellulase variant of Claim 1, wherein the cellulase variant is derived from a parent cellulase whose amino acid sequence is at least 75% identical to a member of the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, and
5 SEQ ID NO:13.

15. The isolated cellulase variant of Claim 1, wherein the more negative net charge is a -1 or -2 in comparison to the reference CBH2.

10 16. A method of converting biomass to sugars comprising contacting said biomass with the cellulase variant of Claim 1.

17. A method of producing a fuel comprising:
contacting a biomass composition with an enzymatic composition comprising the
15 cellulase variant of Claim 1 to yield a sugar solution; and
culturing the sugar solution with a fermentative microorganism under conditions sufficient to produce a fuel.

20 18. A cellulase variant, wherein said variant is a mature form having cellulase activity and comprising a chemical modification of a lysine residue to remove positive charge of the lysine residue.

19. The cellulase variant of Claim 18, wherein the chemical modification comprises a treatment with a compound selected from the group consisting of succinic anhydride,
25 acetoxysuccinic anhydride, maleic anhydride, tartaric anhydride, phthalic anhydride, trimetallitic anhydride, cis-aconitic anhydride, t-nitrophthalic anhydride, acetic anhydride, butyric anhydride, isobutyric anhydride, hexanoic anhydride, valeric anhydride, isovaleric anhydride, and pivalic anhydride.

30 20. The cellulase variant of Claim 18, wherein the cellulase variant is derived from a parent cellulase selected from the group consisting of a *Hypocrea jecorina* cellobiohydrolase I, *Hypocrea jecorina* cellobiohydrolase II, *Hypocrea jecorina* endoglucanase I, *Hypocrea jecorina* endoglucanase II, and *Hypocrea jecorina* beta-glucosidase.

21. The cellulase variant of Claim 18, wherein the cellulase variant is derived from a parent cellulase selected from the group consisting of *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, and *Cellulomonas fimi* CenA EG.

22. The isolated cellulase variant of Claim 18, wherein the cellulase variant is derived from a parent cellulase whose amino acid sequence is at least 75% identical to a member of the group consisting of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, and SEQ ID NO:13.

23. The isolated cellulase variant of Claim 18, wherein the cellulase variant comprises a substitution at one or more positions selected from the group consisting of 63, 77, 129, 147, 153, 157, 161, 194, 197, 203, 237, 239, 247, 254, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, and 382, wherein the positions are numbered by correspondence with the amino acid sequence of a reference cellobiohydrolase II (CBH2) set forth as SEQ ID NO:3.

FIG. 1A

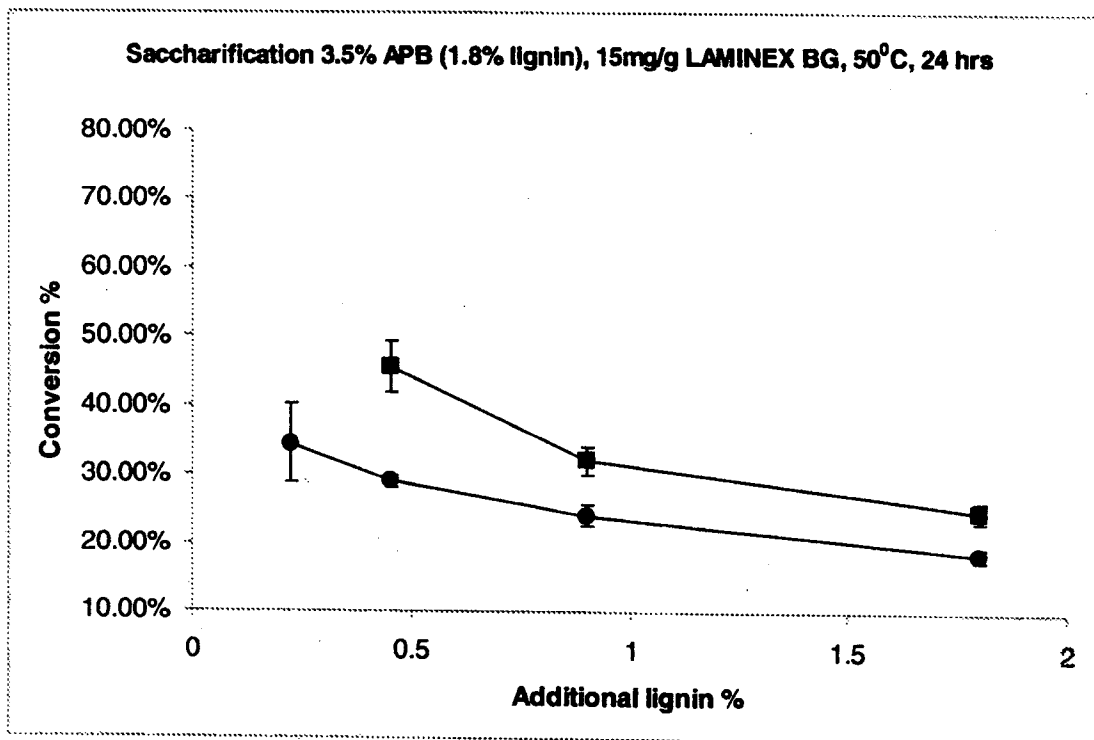


FIG. 1B

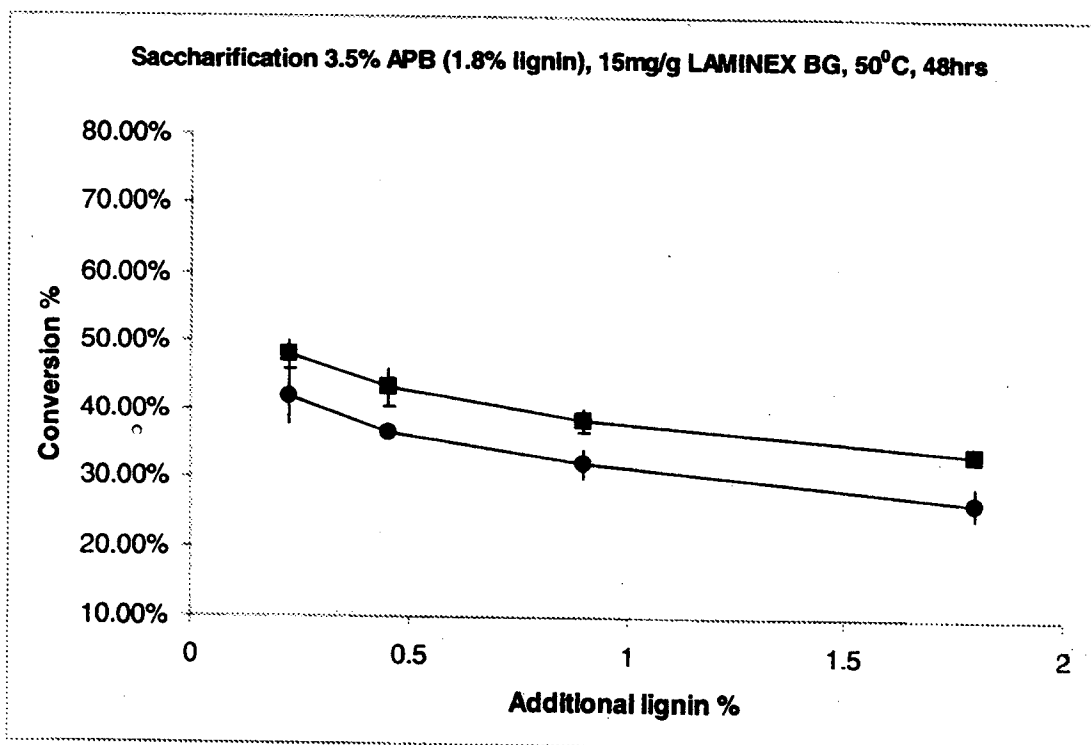


FIG. 2A

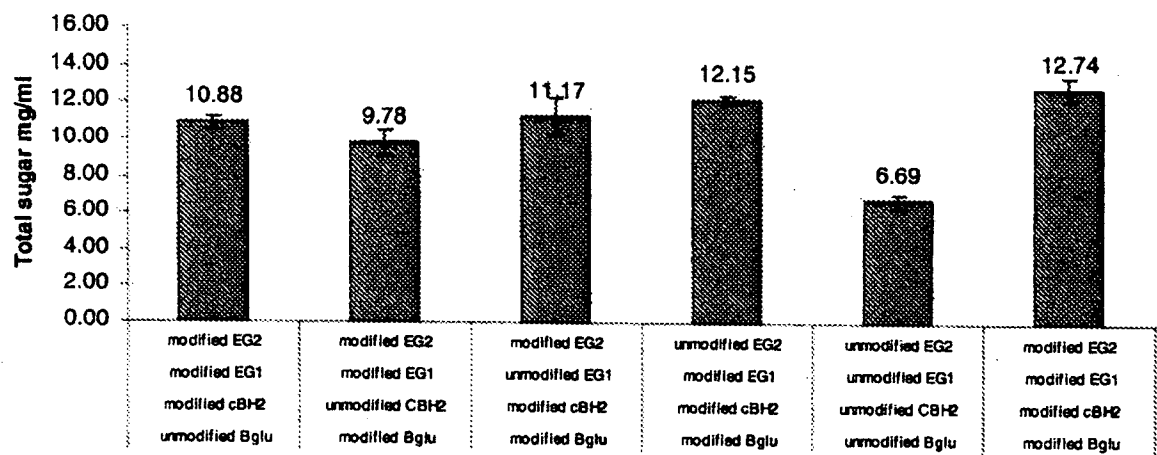


FIG. 2B

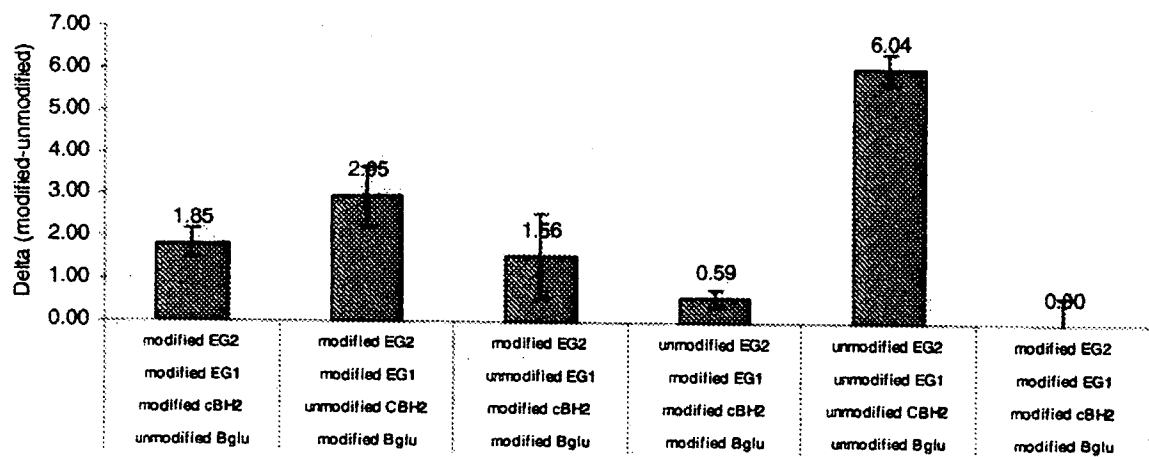


FIG. 3

```
T. reesei CBH #3 -----QACSSVWGQCGGQN----W
H. koningii CBH #4 -----QACSSVWGQCGGQN----W
H. insolens CBH #5 -----APVVEERQNCAPTWGQCGGIG----F
A. cellulolyticus CBH #6 -----MLRYLSIVAATAILTGVEAQQSVWGQCGGQG----W
A. bisporus CBH #7 -----QSPV-----WGQCGGNG----W
F. oxysporum EG #8 -----APVEERQSCSNGVWAQCGGQN----W
P. chrysosporium CBH #9 -----MKSTAFFAALVTLLPAYVAGQASE--WGQCGGIG----W
T. emersonii CBH #10 -----MRNLLALAPAALLVGAAEAQQSLWGQCGGSS----W
T. fusc CBH #11 -----AGCSVDYTVN--SWGTFGTANVTITNLGSAINGWTLWDFPGNQQVTNLW
T. fusc EG #12 -----NDSPFFVNPNNMSSAEWVRNNPND-----
C. fimi EG #13 APGCRVDYAVTNQWPGGFGANVTITNLGDPVSSWKLDWTYTAGQRIQQLW
*

T. reesei CBH #3 SGPTCCASGSTCVYSNDYYSQCLP--GA---ASSSSSTRAAST----TSR
H. koningii CBH #4 SGPTCCASGSTCVYSNDYYSQCLP--GA-----ASSSSSTRASSTTA
H. insolens CBH #5 NGPTCCQSGSTCVKQNDWYSQCLP--GSQVTTTSTTSTSSST----TSR
A. cellulolyticus CBH #6 SGATSCAAGSTCSTLNPYYAQCI--GT-----ATSTTLV----KTT
A. bisporus CBH #7 TGPTTCASGSTCVKQNDFYSCLP--NN-----Q
F. oxysporum EG #8 SGTPCCTSGNKCCKLNDFYSCQP--GS-----AEPS--STA
P. chrysosporium CBH #9 TGPTTCVSGTTCTVLNPYYSCLP--GS-AVTTTTSVITSHSSS----VSS
T. emersonii CBH #10 TGATSCAAGATCSTINPYAQCV--AT-----ATPTTLT----TTT
T. fusc CBH #11 NGTYTQSGQHVSNSAPYNASI--PANGTVEFGFNGSYSGSNDIPSSFKLN
T. fusc EG #12 PRTPVIRDRIASVPQGTWFAHNP--GQ-----ITGQVDA----LMS
C. fimi EG #13 NGTASTNGGQVSVTSLPWNGSI--PTGGTASFNGSWAGSNPTPASFSLN
*

T. reesei CBH #3 VSPTTSR---SSSATPPPGSTTTTRVPP-VGSGTATYSGNPFVGVTPWANA
H. koningii CBH #4 RASSTTS---RSSATPPPGSSTTRVPP-VGSGTATYSGNPFVGVTPWANA
H. insolens CBH #5 ATSTTRT---GGVTSITTAPTRTVTI--PGGATTATSYNGNPFEGVQLWANN
A. cellulolyticus CBH #6 SSTS VGT-----TSPPTTTTTKAST--TATTTAAASGNPFSGYQLYANP
A. bisporus CBH #7 APPSTTT---QPGTTPPATTTSCGTGP--TSGA-----GNPYTGKTVWLSP
F. oxysporum EG #8 AGPSSTT---ATKTTATGGSSTTAGGS--VTSAPPAASDNPYAGVDLWANN
P. chrysosporium CBH #9 VSSHSGS---STSTSSPTGPTGTNPPP--PPSA-----NNPWTGFQIFLSP
T. emersonii CBH #10 KPTSTGG---AAPTTPPPTTTGTTTSP--VVTRPASASGNPFEGYQLYANP
T. fusc CBH #11 GVTCDGSDDPDEPSPSPSPSPSPTDPDEPGGPTNPPNPGKVD---NP
T. fusc EG #12 AAQAAGK---IPILVVYNAPGRDCGNH--SSGGAPSHS-----AYRSWIDE
C. fimi EG #13 GTTCTGT---VPTTSPTPTPTPTPTPT--TPTPTPTPT--PTVTPQPTSGF

T. reesei CBH #3 YYASEVSSLAIPSLTG-AMATAAAVAKVPSFMWLDL-----DKTP-LM
H. koningii CBH #4 YYASEVSSLAIPSLTG-AMATAAAVAKVPSMMLDTF-----DKTP-LM
H. insolens CBH #5 YYRSEVHTLAIPQITDPALRAAASAVAEVPSFQWLDRN-----VTVDTL
A. cellulolyticus CBH #6 YYSSEVHTLAIPSLTG-SLAAAATKAAEIPSFVWLDTA-----AKVP-TM
A. bisporus CBH #7 FYADEVAQ-AAADISNPSLATKAASVAKIPTFVWFDTV-----AKVP-DL
F. oxysporum EG #8 YYRSEVMNLAVPKLSG-AKATAAAKVADVPSFQWMDTY-----DHIS-LM
P. chrysosporium CBH #9 YYANEVAA-AAKQITDPTLSSKAASVANIPTFTWLDSV-----AKIP-DL
T. emersonii CBH #10 YYASEVISLAIPSLSS-ELVPKASEVAKVPSFVWLDAQ-----AKVP-SM
T. fusc CBH #11 FEGAKLYVNPVWSAKA-AAEPGGSANAVANESTAVWLDRIGAEIGNDSPTTG
T. fusc EG #12 F-AAGLKNRPAYIIVEPDLISLMSSCMQHVVQVEVLETM-----AYAGKAL
C. fimi EG #13 YVDPTTQGYRAWQAASGTDKALLEKIALTPQAYWVGWN-----ADAS---
```

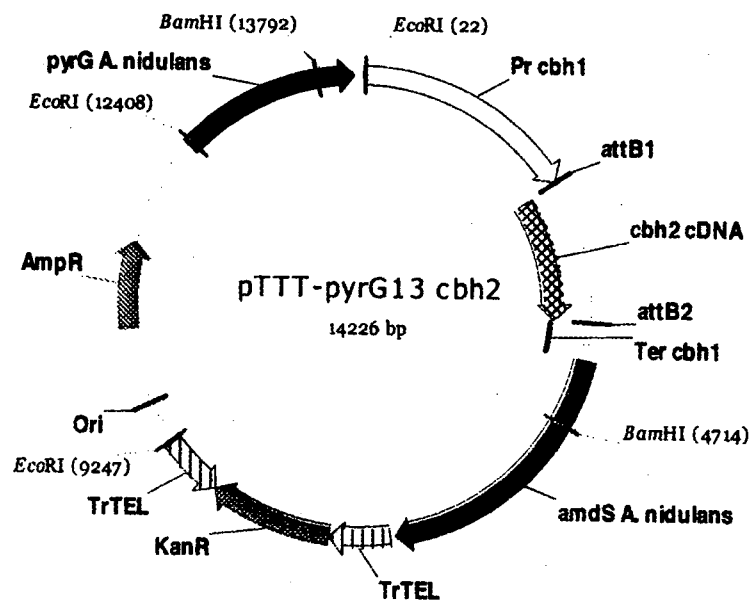
Figure 3 continued

T. reesei CBH #3	EQTLAD-IRTANKNGGN--YAGQFVVYDLPDRDCAALASNGEYSIADGGV
H. koningii CBH #4	EQTLAD-IRTANKNGGN--YAGQFVVYDLPDRDCAALASNGEYSIADGGV
H. insolens CBH #5	VETLSE-IRAANQAGANPPYAAQIVVYDLPDRDCAAAASNGEWAIANNGA
A. cellulolyticus CBH #6	GTylan-IEAANKAGASPPiAGIFVVYDLPDRDCAAAASNGEYTVANNGV
A. bisporus CBH #7	GGYLAD-ARSKNQ-----LVQIVVYDLPDRDCAALASNGEFSLANDGL
F. oxysporum EG #8	EDTLAD-IRKANKAGGK--YAGQFVVYDLPNRDCAAAASNGEYSLDKDGA
P. chrysosporium CBH #9	GTyLAS--ASALGKSTGKQLVQIVYDLPDRDCAAKASNGEFSIANNGQ
T. emersonii CBH #10	GDYLD- IQSONAAGADPPIAGIFVVYDLPDRDCAAAASNGEFSIANNGV
T. fusc CBH #11	SMGLRDHLEEAVRQSGGDPLTIQVVIYNLPGRDCAALASNGELGP--DEL
T. fusc EG #12	KAGSSQ-ARIYFDAGHSAWHSPAQMASWL----QQADISNSAHGIA-TNT
C. fimi EG #13	-HAQAE-VADYTGRAVAAGKTPMLVVYAI PGRDCGSHSGGG-----VSE
	: : : ...
T. reesei CBH #3	AKYKN-YIDTIRQIVVEYSD---IRTLLVIEPDSLNLVTNLGTPK----
H. koningii CBH #4	DKYKN-YIDTIRQIVVEYSD---IRTLLVIEPDSLNLVTNLGTPK----
H. insolens CBH #5	NNYKG-YINRIREILISFSD---VRTILVIEPDSLNLVTNLNVAK----
A. cellulolyticus CBH #6	ANYKA-YIDSIVAQLKAYPD---VHTILIEPDSLNLVTNLSTAK----
A. bisporus CBH #7	NKYKN-YVDQIAAQIKQFPD---VSVVAVIEPDSLNLVTNLNVQK----
F. oxysporum EG #8	NKYKA-YIAKIKGILQNYSD---TKVILVIEPDSLNLVTNLNVQK----
P. chrysosporium CBH #9	ANYEN-YIDQIVAQIQFPD---VRVAVIEPDSLNLVTNLNVQK----
T. emersonii CBH #10	ALYKQ-YIDSIREQLTTYSD---VHTILVIEPDSLNLVTNLNVQK----
T. fusc CBH #11	DRYKSEYIDPIADIMWDFADYENLRIVAIIEIDSLPNLVTNLVGGNGGTEL
T. fusc EG #12	SNYRW-TADEVA-----YAK----AVLSAIGNPSLRAVIDTSRNGNG----
C. fimi EG #13	SEYAR-WVDTVAQGIK-----GNPIVILEPDALAQLGD-----
	* : : : . . .
T. reesei CBH #3	CA--NAQSAYLECINYAVTQL-NLPNVAMYLDAGHAGWLGWPANQDPAAQ
H. koningii CBH #4	CA--NAQSAYLECINYAVTQL-NLPNVAMYLDAGHAGWLGWPANQDPAAQ
H. insolens CBH #5	CS--GAASTYRELTIIYALKQL-DLPHVAMYMDAGHAGWLGWPANIQPAAE
A. cellulolyticus CBH #6	CA--EAQSAYYECVNYALINL-NLANVAMYIDAGHAGWLGWPANLSPAAQ
A. bisporus CBH #7	CA--NAQSAYKEGVIYAVQKL-NAVGVITMYIDAGHAGWLGWPANLSPAAQ
F. oxysporum EG #8	CA--KAESAYKELTVYAIKEL-NLPNVSMYLDAGHGGWLGWPANIGPAAK
P. chrysosporium CBH #9	CA--NAKTTYLACVNYALTNL-AKVGVMYMDAGHAGWLGWPANLSPAAQ
T. emersonii CBH #10	CA--NAQDAYLECINYAITQL-DLPNVAMYLDAGHAGWLGWQANLAPAAQ
T. fusc CBH #11	CAYMKQNGGYVNGVGYALRKLGEIPNVYNYIDAAHHGWIGWDSNFGPSVD
T. fusc EG #12	----PAGNEWCDPSGRAIGTPSTTNTGDPMDAFL--WIKLPGEADGCIA
C. fimi EG #13	CS---GQGDVRVGLKYAAKSL-TLKGARVYIDAGHAKWLSVDTPVNRLNQ
	* : ** *
T. reesei CBH #3	LFANVYKNAS-SPRALRGLATNVANYNGWNIT-----SPPSY
H. koningii CBH #4	LFANVYKNAS-SPRALRGLATNVANYNGWNIT-----SPPSY
H. insolens CBH #5	LFakiYEDAG-KPRAVRGLATNVANYNAWSIS-----SPPPY
A. cellulolyticus CBH #6	LFATVYKNAS-APASLRGLATNVANYNAWSIS-----SPPSY
A. bisporus CBH #7	LFAQIYRDAG-SPRNLRLGIATNVANFNALRAS-----SPDPI
F. oxysporum EG #8	LYAQIYKDAG-KPSRVRLVTVNSNYNGWKLS-----TKPDY
P. chrysosporium CBH #9	LFTQVWQWAG-KSPFIKGLATNVANYNALQAA-----SPDPI
T. emersonii CBH #10	LFASVYKNAS-SPASVRGLATNVANYNAWSIS-----RCPSY
T. fusc CBH #11	IFYEAAANASGSTVDYVHGFI SNTANY SATVEPYLDVNGTVNGQLIRQSKW
T. fusc EG #12	GAGQFVPQAAYEMAIAAGGTNPENPNP--TP-----TPTPT
C. fimi EG #13	V-----GFEYAVGFALNTSNYQT-----
	* * *

Figure 3 continued

T. reesei CBH #3	TQGNVYNEKLYIHAIGPLLANHGWSN-AFFITDQGRSGK----	QPTGQQ
H. koningii CBH #4	TQGNVYNEQLYIHAIGPLLANHGWSN-AFFITDQGRSGK----	QPTGQQ
H. insolens CBH #5	TSPNPYDEKHYIEAFRPLLEARGFP--AQFIVDQGRSGK----	QPTGQQ
A. cellulolyticus CBH #6	TSGDSNYDEKLYINALSPLLTSTNGWPN-AHFIMDTSRNGV----	QPTKQQ
A. bisporus CBH #7	TQGNSNYDEIHYIEALAPMLSNAGFP--AHFIVDQGRSGV----	QNIRD
F. oxysporum EG #8	TESNPYDEQRYINAFAPLLAQEGWSN-VKFIVDQGRSGK----	QPTGQK
P. chrysosporium CBH #9	TQGNPNYDEIHYINALAPLLQQAGWD--ATFIVDQGRSGV----	QNIRQ
T. emersonii CBH #10	TQGDANCDEEDYVNALGPLFQEQGFP--AYFIIDTSRNGV----	RPTKQS
T. fusc CBH #11	VDWNQYVDELSEFVQDLRQALIAKGRSDIGMLIDTSRNGWGGPNRPTGPS	
T. fusc EG #12	PTPPGSSGACTATYTTIANEWNDGFQATVTVTANQNITGW-----	TV
C. fimi EG #13	-----TADSKAYGQQISQRLGG-----KKFVIDTSRNGN-----	G
	: . . . *	
T. reesei CBH #3	QWGD-----WCNVIGTGFGIRPSANTGDSLLDSFVWV	
H. koningii CBH #4	QWGD-----WCNVIGTGFGIRPSANTGDSLLDSFVWI	
H. insolens CBH #5	EWGH-----WCNAIGTGFGMRPTANTGHQYVDAFVWV	
A. cellulolyticus CBH #6	AWGD-----WCNVIGTGFGVQPTTNTGDPLEDAFVWV	
A. bisporus CBH #7	QWGD-----WCNVKGAGFGQRPTTNTGSSSLIDAFVWV	
F. oxysporum EG #8	AQGD-----WCNAKGTGFLRPSTNTGDALADAFVWV	
P. chrysosporium CBH #9	QWGD-----WCNIKAGFGTRPTTNTGTSQFIDSIVWV	
T. emersonii CBH #10	QWGD-----WCNVIGTGFGVRPTTDTGNPLEDAFVWV	
T. fusc CBH #11	SSTDNLNTYVDESRIIDRIHPGNWCNQAGAGLGERPTVNPA-PGVDAFVWV	
T. fusc EG #12	TWT-----FTDGTITNAWNADVSTSGSSVTARNVG	
C. fimi EG #13	SNGE-----WCNPRGRALGERPVAVNDGSGLDALLWV	
	: : . . . :	
T. reesei CBH #3	KPGGECDGTSDDSS----APRFDHCA-----LPDALQPAPQAG	
H. koningii CBH #4	KPGGECDGTSDDSS----APRFDHCA-----LPDALQPAPQAG	
H. insolens CBH #5	KPGGECDGTSDDTT----AARYDYHCG-----LEDALKPAPEAG	
A. cellulolyticus CBH #6	KPGGESDGTSSNS----ATRYDFHCG-----YSDALQPAPQAG	
A. bisporus CBH #7	KPGGECDGTSDDNS----SPRFDHSCS-----LSDAHQPAPQAG	
F. oxysporum EG #8	KPGGESDGTSDTS----AARYDYHCG-----LDDALKPAPEAG	
P. chrysosporium CBH #9	KPGGECDGTSNS--SPRYDSTCS-----LPDAAQPAPQAG	
T. emersonii CBH #10	KPGGESDGTSTNT--SPRYDYHCG-----LSDALQPAPQAG	
T. fusc CBH #11	KPPGESDGASEEIPNDEGKGFDRCMDPTYQGNARNGNPNPSGALPNAPISG	
T. fusc EG #12	HNGTLSQGASTEF-----GFVG	
C. fimi EG #13	KLPGESDGACNGG-----PAAG	
	: . . . *	
T. reesei CBH #3	AWFQAYFVQLLTNANPSFL	
H. koningii CBH #4	AWFQAYFVQLLTNANPSFL	
H. insolens CBH #5	QWFQAYFEQLLRNANPPF-	
A. cellulolyticus CBH #6	TWFQAYFVQLLTNANPALV	
A. bisporus CBH #7	TWFQAYFETLVANANPAL-	
F. oxysporum EG #8	TWFQAYFEQLLDNANPSFL	
P. chrysosporium CBH #9	TWFQAYFQTLVSAANPPL-	
T. emersonii CBH #10	TWFQAYFEQLLTNANPLF-	
T. fusc CBH #11	HWFSAQFRELLANAYPPL-	
T. fusc EG #12	SKGNSNSVPTLTCAAS---	
C. fimi EG #13	QWWQEIALEMARNARW---	
	*	

Figure 5



COMPOSICIONES Y MÉTODOS QUE COMPRENDEN VARIANTES DE CELULASA CON AFINIDAD REDUCIDA PARA MATERIALES NO CELULÓSICOS

I. DECLARACIÓN RELATIVA A LOS DERECHOS EN INVENCIONES REALIZADAS MEDIANTE INVESTIGACIÓN Y DESARROLLO FINANCIADOS CON RECURSOS FEDERALES

[01] Esta invención se realizó con apoyo del Gobierno bajo subvención condicional No: DE-FC36-08GO18078, otorgada por el Departamento de Energía. El Gobierno tiene ciertos derechos en esta invención.

II. REFERENCIA CRUZADA A UNA SOLICITUD RELACIONADA

[02] Esta solicitud reclama el beneficio de la Solicitud Provisional de los Estados Unidos No. 61/059,506, presentada el 6 de junio de 2008, la cual se incorpora en esta especificación por referencia.

III. CAMPO

[03] La presente divulgación se refiere a enzimas y, en particular, a variantes de celulasa. Se describen asimismo ácidos nucleicos que codifican las variantes de celulasa, composiciones que contienen las variantes de celulasa, métodos útiles adicionales de identificar variantes de celulasa y métodos de usar las composiciones.

IV. ARTE PREVIO

[04] La celulosa y la hemicelulosa son los materiales vegetales más abundantes producidos por fotosíntesis. Pueden degradarse y usarse como una fuente de energía por numerosos microorganismos (v.g., bacterias, levadura y hongos), que producen enzimas extracelulares capaces de hidrolizar los sustratos poliméricos en azúcares monoméricos (Aro y col., J Biol Chem, 276: 24309-24314, 2001). A medida que se aproximan los límites de los recursos no renovables, el potencial de la celulosa de convertirse en un importante recurso energético renovable es enorme (Krishna y col., Bioresource Tech, 77: 193-196, 2001). La utilización efectiva de celulosa por procesos biológicos es un enfoque para superar la escasez de alimentos, forrajes, y combustibles (Ohmiya y col., Biotechnol Gen Engineer Rev, 14: 365-414, 1997).

[05] Las celulasas son enzimas que hidrolizan celulosa (enlaces beta-1,4-glucano o beta D-glucosídicos), lo cual resulta en la formación de glucosa, celobiosa, celooligosacáridos, y similares. Tradicionalmente las celulasas se han dividido en tres clases principales: endoglucanasas (EC 3.2.1.4) ("EG"), exoglucanasas o celobiohidrolasas (EC 3.2.1.91) ("CBH") y beta-glucosidasas ([beta]-D-glucósido glucohidrolasa; EC 3.2.1.21) ("BG"). (Knowles y col., TIBTECH 5: 255-261, 1987; y Schulein, Methods Enzymol, 160: 234-243, 1988). Las endoglucanasas actúan principalmente sobre las partes amorfas de la fibra de celulosa, mientras que las celobiohidrolasas son también capaces de degradar la celulosa cristalina (Nevalainen y Penttila, Mycota, 303-319, 1995). Por consiguiente, la presencia de una celobiohidrolasa en un sistema de celulasas se requiere para una eficiente solubilización de la celulosa cristalina (Suurnakki y col., Cellulose 7: 189-209, 2000). La beta-glucosidasa actúa para liberar unidades de D-glucosa de la celobiosa, celo-oligosacáridos, y otros glucósidos (Freer, J Biol Chem, 268: 9337-9342, 1993).

[06] Se sabe que las celulasas son producidas por un gran número de bacterias, levaduras y hongos. Ciertos hongos producen un sistema completo de celulasas capaz de degradar formas cristalinas de celulosa, de modo que las celulasas se produzcan fácilmente en grandes cantidades mediante fermentación. Los hongos filamentosos juegan un papel especial puesto que muchas levaduras, tales como *Saccharomyces cerevisiae*, carecen de la capacidad de hidrolizar celulosa (véase, v.g., Wood y col., Methods in Enzymology, 160: 87-116, 1988).

[07] Las clasificaciones de CBH, EG y BG de las celulasas fúngicas pueden ampliarse además para incluir múltiples componentes dentro de cada clasificación. Por ejemplo, múltiples CBHs, EGs y BGs se han aislado de una variedad de fuentes fúngicas incluyendo *Trichoderma reesei* (también denominado *Hypocrea jecorina*), el cual contiene genes conocidos para dos CBHs, es decir, CBH I ("CBH1") y CBH II ("CBH2"), por lo menos 8 EGs, es decir, EG I, EG II, EG III, EGIV, EGV, EGVI, EGVII y EGVIII, y por lo menos 5 BGs, es decir, BG1, BG2, BG3, BG4 y BG5. EGIV, EGVI y EGVIII tienen igualmente actividad de xiloglucanasa.

[08] Con el fin de convertir eficientemente celulosa cristalina en glucosa, se requiere el sistema completo de celulasas constituido por componentes de cada una de las clasificaciones CBH, EG y BG, con componentes aislados menos efectivos en hidrolizar celulosa cristalina (Filho y col., Can J Microbiol, 42:1-5,

1996). Se ha observado una relación sinérgica entre componentes de celulasa de diferentes clasificaciones. En particular, las celulastas tipo EG y las celulastas tipo CBH interactúan sinérgicamente para degradar celulosa de manera más eficiente.

[09] Las celulastas son conocidas en el arte por ser útiles en el tratamiento de textiles para los propósitos de potenciar la capacidad limpiadora de composiciones detergentes, para uso como un agente suavizante, para mejorar el tacto y apariencia de las telas de algodón, y similares (Kumar y col., Textile Chemist and Colorist, 29:37-42, 1997). Se han descrito composiciones detergentes que contienen celulasa con rendimiento limpiador mejorado (Patente de los Estados Unidos No. 4,435,307; Solicitudes de Gran Bretaña Nos. 2,095,275 y 2,094,826) y para uso en el tratamiento de telas para mejorar el tacto y apariencia de textiles (Patentes de los Estados Unidos Nos. 5,648,263, 5,691,178, y 5,776,757; y Solicitud de la Gran Bretaña No. 1,358,599). Por consiguiente, celulastas producidas en hongos y bacterias han recibido una atención significativa. En particular, la fermentación de *Trichoderma spp.* (v.g., *Trichoderma longibrachiatum* o *Trichoderma reesei*) ha demostrado producir un sistema completo de celulastas capaz de degradar formas cristalinas de celulosa.

[10] Aunque previamente se han descrito composiciones de celulasa, subsiste una necesidad de composiciones de celulasa nuevas y mejoradas. Composiciones de celulasa mejoradas se usan en detergentes de uso en el hogar, tratamientos para textiles, conversión de biomasa y fabricación de papel. Son de particular interés las celulastas que exhiben un rendimiento mejorado.

V. SUMARIO

[11] La presente enseñanza se refiere a variantes de celulasa modificadas para reducir la unión a materiales no celulósicos. En general, las variantes de celulasa tienen una actividad celulolítica incrementada en la presencia de materiales no celulósicos, en comparación con celulastas naturales ("**wild type cellulases**"). En algunas incorporaciones las variantes de celulasa tienen una carga neta disminuída (es decir, es más negativa) en comparación con celulastas naturales. En algunas incorporaciones, las variantes de celulasa están cargadas menos positivamente que las celulastas naturales. En algunas incorporaciones, una celulasa se modifica eliminando una o más cargas positivas. En algunas incorporaciones, una celulasa se modifica agregando una o más cargas negativas. En algunas incorporaciones, una celulasa se modifica eliminando una o más cargas positivas y agregando una o más cargas negativas.

[12] En algunas incorporaciones, las presentes enseñanzas se refieren a variantes de celobiohidrolasa I (CBH1) o celobiohidrolasa II (CBH2). En algunas incorporaciones la variante de celulasa es una forma madura que tiene actividad de celulasa y una sustitución en una o más posiciones seleccionadas del grupo conformado por 63, 77, 129, 147, 153, 157, 161, 194, 197, 203, 237, 239, 247, 254, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, y 382, en donde las posiciones se numeran por correspondencia a una celulasa de referencia (v.g., *Hypocrea jecorina* CBH2 natural) que tiene la secuencia de aminoácidos de SEQ ID NO:3, y en donde la sustitución en una o más posiciones hace que la variante de celulasa tenga una carga neta más negativa en comparación con la celulasa de referencia. En algunas incorporaciones, CBH2 se modifica eliminando una o más cargas positivas, lo cual en algunas incorporaciones implica un reemplazo de una lisina o una arginina con un aminoácido neutro (v.g., K o R reemplazado por N o Q u otro residuo neutro). En algunas incorporaciones, CBH2 se modifica agregando una o más cargas negativas, lo cual en algunas incorporaciones implica un reemplazo de un aminoácido neutro con un aminoácido cargado negativamente (v.g., N o Q u otro residuo neutro reemplazado por D o E). En algunas incorporaciones, CBH2 se modifica eliminando una o más cargas positivas y agregando una o más cargas negativas, lo cual en algunas incorporaciones implica un reemplazo de una lisina o una arginina con un aminoácido cargado negativamente (v.g., K o R reemplazado con D o E). En general, la variante CBH2 tiene una actividad celulolítica incrementada en la presencia de lignina en comparación con la *Hypocrea jecorina* CBH2 natural que tiene la secuencia de aminoácidos de SEQ ID NO:3. Las presentes enseñanzas proporcionan además variantes CBH2 que incluyen una o más sustituciones seleccionadas del grupo conformado por K129E, K157E, K194E, K288E, K327E, K356E, R63Q, R77Q, R153Q, R203Q, R294Q, R378Q, N161D, N197D, N237D, N247D, N254D, N285D, N289D, N339D, N344D, N382D, Q147E, Q204E, Q239E, Q281E, D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, y E244Q, en la forma madura de CBH2, en donde las citadas sustituciones se numeran de acuerdo con la forma madura de *Hypocrea jecorina* CBH2 de SEQ ID NO:3. En algunas incorporaciones, la variante incluye una sustitución adicional en una o más posiciones adicionales seleccionadas del grupo conformado por 146, 151, 189, 208, 211, 244, 277 y 405, en donde las posiciones adicionales se numeran por correspondencia con la secuencia de aminoácidos de la celobiohidrolasa II (CBH2) de referencia divulgada como SEQ ID NO:3. En algunas incorporaciones, la sustitución adicional en una o más posiciones adicionales incluye un reemplazo de ácido aspártico o ácido glutámico con un aminoácido neutro (v.g., D o E reemplazado por N o Q u

otro residuo neutro). En algunas incorporaciones, la sustitución adicional en una o más posiciones adicionales incluye una o más del grupo conformado por D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, y E244Q, en donde las posiciones se numeran por correspondencia con la secuencia de aminoácidos de la celobiohidrolasa II (CBH2) de referencia divulgada como SEQ ID NO:3. En algunas incorporaciones preferidas, la sustitución en una o más posiciones se selecciona del grupo conformado por las posiciones 1, 2, 3, 4, 5, 6, 7, 8, 9 y 10. En algunas incorporaciones preferidas, la variante de celulasa se deriva de una celulasa progenitora seleccionada del grupo conformado por *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, y *Cellulomonas fimi* CenA EG. En algunas incorporaciones preferidas, la variante de celulasa se deriva de una celulasa progenitora cuya secuencia de aminoácidos es por lo menos 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% o 99% idéntica a un miembro del grupo conformado por SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO: 10, SEQ ID NO:11, SEQ ID NO:12, y SEQ ID NO:13. En algunas incorporaciones, la carga neta más negativa es -1 o -2 en comparación con la CBH2 de referencia.

[13] La presente divulgación provee también variantes de celulasa, en donde la variante es una forma madura que tiene actividad de celulasa e incluye una modificación química de un residuo lisina para eliminar la carga positiva del residuo lisina. En algunas incorporaciones preferidas, la modificación química comprende un tratamiento con un compuesto seleccionado del grupo conformado por anhídrido succínico, anhídrido acetoxisuccínico, anhídrido maleico, anhídrido tartárico, anhídrido ftálico, anhídrido trimelítico, anhídrido cis-aconítico, anhídrido t-nitroftálico, anhídrido acético, anhídrido butírico, anhídrido isobutírico, anhídrido hexanoico, anhídrido valérico, anhídrido isovalérico y anhídrido piválico. En algunas incorporaciones preferidas, la variante de celulasa se deriva de una celulasa progenitora seleccionada del grupo conformado por una *Hypocrea jecorina* celobiohidrolasa I, *Hypocrea jecorina* celobiohidrolasa II, *Hypocrea jecorina* endoglucanasa I, *Hypocrea jecorina* endoglucanasa II, e *Hypocrea jecorina* beta-glucosidasa. En algunas incorporaciones preferidas, la variante de celulasa se deriva de una celulasa progenitora seleccionada del grupo conformado por *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2,

Thermobifida fusca 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, y *Cellulomonas fimi* CenA EG. De igual manera, se proveen variantes de celulasa derivadas de una celulasa progenitora cuya secuencia de aminoácidos es por lo menos 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% o 99% idéntica a un miembro del grupo conformado por SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, y SEQ ID NO:13. En algunas incorporaciones, la variante de celulasa comprende una sustitución en una o más posiciones seleccionadas del grupo conformado por 63, 77, 129, 147, 153, 157, 161, 194, 197, 203, 237, 239, 247, 254, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, y 382, en donde las posiciones se numeran por correspondencia con la secuencia de aminoácidos de una celobiohidrolasa II (CBH2) de referencia divulgada como SEQ ID NO:3.

[14] Las presentes enseñanzas se refieren además a una variante CBH2 que incluye de una a veintiséis sustituciones seleccionadas del grupo conformado por K129E, K157E, K194E, K288E, K327E, K356E, R63Q, R77Q, R153Q, R203Q, R294Q, R378Q, N161D, N197D, N237D, N247D, N254D, N285D, N289D, N339D, N344D, N382D, Q147E, Q204E, Q239E, y Q281E. En algunas incorporaciones, la variante CBH2 comprende una combinación de sustituciones seleccionadas del grupo conformado por: i) K157E / K129E; ii) K157E / K129E / K288E / K194E; iii) K157E / K129E / K288E / K194E / K356E / K327E; iv) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q; v) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q / N382D / N344D / N327D / N339D; vi) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q / N382D / N344D / N327D / N339D / N289D / N161D / Q204E / Q147E; vii) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q / N382D / N344D / N327D / N339D / N289D / N161D / Q204E / Q147E / N285D / N197D / N254D / N247D; y viii) K157E / K129E / K288E / K194E / K356E / K327E / R153Q / R294Q / R203Q / R378Q / N382D / N344D / N327D / N339D / N289D / N161D / Q204E / Q147E / N285D / N197D / N254D / N247D / Q239E / Q281E / R63Q / R77Q.

[15] En algunas incorporaciones, la variante CBH2 incluye de una a ocho sustituciones seleccionadas del grupo conformado por D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, y E244Q. En algunas incorporaciones, la variante CBH2 incluye una combinación de sustituciones seleccionadas del grupo constituido por: i) D189N / E208Q / D211N / D405; y ii) D189N / E208Q / D211N / D405 / E244Q / D277N / D151 / E146Q.

[16] Se describen asimismo ácidos nucleicos aislados que codifican una variante CBH2 que posee actividad de celobiohidrolasa, como se describe en los párrafos precedentes. En un primer aspecto, la divulgación abarca un ácido nucleico aislado que codifica un polipéptido que posee actividad de celobiohidrolasa, el cual polipéptido es una variante de una glicosil hidrolasa de la familia 6, y en donde el citado ácido nucleico codifica una sustitución en un residuo que disminuye la carga neta en comparación con la *Hypocrea jecorina* CBH2 natural.

[17] En otro aspecto, la divulgación se dirige a un ácido nucleico aislado que codifica una variante CBH2, en donde la citada variante comprende una sustitución en una posición seleccionada del grupo conformado por K129E, K157E, K194E, K288E, K327E, K356E, R63Q, R77Q, R153Q, R203Q, R294Q, R378Q, N161D, N197D, N237D, N247D, N254D, N285D, N289D, N339D, N344D, N382D, Q147E, Q204E, Q239E, Q281E, D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, y E244Q, en la forma madura de CBH2, en donde las citadas sustituciones se numeran de acuerdo con la forma madura de *Hypocrea jecorina* CBH2 de SEQ ID NO:3.

[18] En algunas incorporaciones, la divulgación se dirige a un casete de expresión que comprende un ácido nucleico que codifica una variante CBH2, un constructo que comprende el ácido nucleico que codifica la variante CBH2 enlazado operacionalmente a una secuencia reguladora, un vector que comprende un ácido nucleico que codifica una variante CBH2, y una célula hospedera transformada con el vector que comprende un ácido nucleico que codifica una variante CBH2. Las presentes enseñanzas ofrecen además métodos para producir una variante CBH2 cultivando las células hospederas que expresan una variante CBH2 en un medio de cultivo bajo condiciones apropiadas para producir la variante CBH2.

[19] Se proveen igualmente composiciones que comprenden la variante de celulasa de los párrafos precedentes. En algunas incorporaciones preferidas, la composición comprende además por lo menos una enzima adicional seleccionada del grupo conformado por una subtilisina, una metaloproteasa neutra, una lipasa, una cutinasa, una amilasa, una carbohidrasa, una pectinasa, una mannanasa, una arabinasa, una galactanasa, una xilanasa, una oxidasa, y una peroxidasa.

[20] En esta especificación se ofrecen métodos de convertir biomasa en azúcares, métodos que comprenden contactar la citada biomasa con una variante de celulasa. Se proveen además métodos de producir un combustible contactando una composición de biomasa con una composición enzimática que comprende la

variante de celulasa para producir una solución de azúcar y cultivar con un microorganismo fermentador bajo condiciones suficientes para producir un combustible.

[21] Se proporcionan igualmente composiciones que comprenden variantes de celulasa incluyendo composiciones detergentes, aditivos de forrajes por ejemplo, y métodos de limpieza o cuidado de telas contactando una superficie y/o un artículo, que comprende una tela, con la composición detergente. Asimismo, se facilitan métodos de tratamiento para el cuidado de telas, incluyendo la eliminación de bolitas formadas por agrupación de fibras en los tejidos ("*depilling*") y acabado superficial, contactando una superficie y/o un artículo, que comprende una tela, con una variante de celulasa.

[22] Otros objetos, características y ventajas de la presente divulgación se harán evidentes a partir de la siguiente descripción detallada. Debe entenderse, sin embargo, que la descripción detallada y los ejemplos específicos, aunque indican incorporaciones preferidas de la divulgación, se dan a manera de ilustración únicamente, puesto que varios cambios y modificaciones dentro del ámbito y espíritu de la divulgación se harán evidentes a un experto en el arte a partir de esta descripción detallada.

VI. BREVE DESCRIPCIÓN DE LSO DIBUJOS

[23] FIG. 1 ilustra la sacarificación de APB (*acid pretreated bagasse* – bagazo pretratado con ácido) por preparaciones de celulasa de *Trichoderma sp.* modificada (cuadrados) y no modificada (círculos) en la presencia de cantidades crecientes de inhibidor de lignina. FIG. 1A y FIG. 1B muestran los resultados después de incubaciones de 24 y 48 horas, respectivamente.

[24] FIG. 2A ilustra la sacarificación comparando celulasas modificadas y FIG. 2B muestra la diferencia de la sacarificación usando celulasas modificadas y no modificadas.

[25] FIG. 3 presenta un alineamiento de las secuencias de aminoácidos de la forma madura de varias celulasas: *Hypocrea jecorina* (también conocida como *T. reesei*) CBH2 (SEQ ID NO:3), *Hypocrea koningii* CBH2 (SEQ ID NO:4), *Humicola insolens* CBH2 (SEQ ID NO:5), *Acremonium cellulolyticus* CBH2 (SEQ ID NO:6), *Agaricus bisporus* CBH2 (SEQ ID NO:7), *Fusarium osysporum* EG (SEQ ID NO:8), *Phanerochaete chrysosporium* CBH2 (SEQ ID NO:9), *Talaromyces emersonii*

CBH2 (SEQ ID NO: 10), *Thermobifida fusca* 6B/E3 CBH2 (SEQ ID NO: 11), *Thermobifida fusca* 6A/E2 EG (SEQ ID NO: 12), y *Cellulomonas fimi* CenA EG (SEQ ID NO: 13).

[26] FIG. 4 ilustra un gráfico de la frecuencia relativa de los ganadores observados sobre los esperados en un análisis con rastreo de maíz pretratado (PCS - *pretreated corn stover*) de la variante CBH2 SELs como un producto del cambio de carga. Disminuir la carga de CBH2 resulta en una frecuencia significativamente mayor de ganadores PCS.

[27] FIG. 5 presenta un mapa de un plásmido de pTTTpyr-cbh2.

VII. DESCRIPCIÓN DETALLADA DE VARIAS INCORPORACIONES

[28] Las presentes enseñanzas se refieren a variantes de celulasa modificada para reducir la unión a materiales no celulósicos. En general, la variante de celulasa tiene una mayor actividad celulolítica en la presencia de materiales no celulósicos, en comparación con la celulasa natural. En algunas incorporaciones la celulasa variante tiene una carga neta disminuída (*es decir*, es más negativa) en comparación con la celulasa natural. En algunas incorporaciones, las variantes de celulasa están cargadas menos positivamente que la celulasa natural. En algunas incorporaciones, una celulasa se modifica eliminando una o más cargas positivas. En algunas incorporaciones, una celulasa se modifica agregando una o más cargas negativas. En algunas incorporaciones, una celulasa se modifica eliminando una o más cargas positivas y agregando una o más cargas negativas.

[29] Hay que entender que tanto la anterior descripción general como la siguiente descripción detallada son ejemplares y explicativas únicamente y no son restrictivas de las composiciones y métodos descritos en esta especificación. A menos que se defina lo contrario en esta especificación, todos los términos técnicos y científicos empleados aquí tienen el mismo significado que se entiende comúnmente por una persona de habilidad ordinaria en el arte al cual pertenece esta divulgación. En esta solicitud, el uso del singular incluye el plural a menos que específicamente se manifieste lo contrario. El uso de "o" significa "y/o", a menos que se indique otra cosa. Los términos "comprenden," "que comprende," "comprende," "incluyen," "incluyendo," e "incluye" no pretenden ser limitantes. Todas las patentes y publicaciones, incluyendo todas las secuencias de aminoácidos y las secuencias nucleotídicas divulgadas en tales patentes y publicaciones, mencionadas en esta especificación, se incorporan expresamente

por referencia. Los encabezamientos provistos en esta especificación no constituyen limitaciones de los diferentes aspectos o incorporaciones de la divulgación que puedan ser tenidas por referencia a la especificación como un todo. Por consiguiente, los términos contenidos aquí se definen más completamente por referencia a la especificación como un todo.

[30] A menos que se defina lo contrario en esta especificación, todos los términos técnicos y científicos empleados aquí tienen el mismo significado que se entiende comúnmente por una persona de habilidad ordinaria en el arte al cual pertenece esta divulgación. Singleton, y col., *DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY*, 2D ED., John Wiley and Sons, New York (1994), y Hale & Marham, *THE HARPER COLLINS DICTIONARY OF BIOLOGY*, Harper Perennial, NY (1991) ofrecen a una persona experta un diccionario general de muchos de los términos usados en la divulgación. Aunque algunos métodos y materiales similares o equivalentes a los descritos en esta especificación pueden usarse en la práctica o experimentación de la presente divulgación, se describen los métodos y materiales preferidos. Los intervalos numéricos son inclusivos de los números que definen el intervalo. A menos que se indique otra cosa, los ácidos nucleicos se escriben de izquierda a derecha en orientación 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en orientación amino a carboxilo, respectivamente. A los profesionales se los remite particularmente a Sambrook y col., *MOLECULAR CLONING: A LABORATORY MANUAL* (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989, y Ausubel FM y col., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y., 1993, en lo que respecta a las definiciones y términos del arte. Hay que entender que esta divulgación no se limita a la metodología, protocolos, y reactivos particulares descritos, ya que éstos pueden variar.

I. Definiciones

[31] Los siguientes términos se definen más completamente por referencia a la especificación como un todo.

[32] El término "polipéptido", tal como se emplea en esta especificación, se refiere a un compuesto constituido por una cadena única de residuos aminoácidos conectados por enlaces peptídicos. El término "proteína", tal como se emplea en esta especificación, puede ser sinónimo con el término "polipéptido".

[33] "Variante" significa una proteína que se deriva de una proteína precursora

(v.g., la proteína nativa) por adición de uno o más aminoácidos a uno cualquiera o a ambos extremos C- y N-terminal, sustitución de uno o más aminoácidos en uno o un número de diferentes sitios en la secuencia de aminoácidos, o eliminación de uno o más aminoácidos en uno cualquiera o en ambos extremos de la proteína o en uno o más sitios en la secuencia de aminoácidos, o uno o más aminoácidos se modifican cambiando la carga (es decir, suprimiendo una carga positiva, agregando una carga negativa, o ambos, suprimiendo una carga positiva y agregando una carga negativa). La preparación de una variante de celulasa puede llevarse a cabo por cualquier medio conocido en el arte, incluyendo modificación química de aminoácidos, modificando una secuencia de DNA que codifica la proteína nativa, transformación de la secuencia de DNA modificado en un hospedero apropiado, y expresión de la secuencia de DNA modificado para formar una enzima variante. La celulasa variante de la divulgación incluye péptidos que comprenden secuencias de aminoácidos alteradas en comparación con una secuencia de aminoácidos de la enzima precursora en donde la celulasa variante retiene la naturaleza celulolítica característica de la enzima precursora, pero que puede tener propiedades alteradas en algún aspecto específico. Por ejemplo, una celulasa variante puede tener un pH óptimo o una mayor temperatura o estabilidad oxidativa o afinidad disminuída o unión a materiales no celulósicos, pero conservará su actividad celulolítica característica. Se contempla que las variantes de acuerdo con la presente divulgación pueden derivarse de un fragmento de DNA que codifica una variante de celulasa en donde se conserva la actividad funcional de la variante de celulasa expresada. Por ejemplo, un fragmento de DNA que codifica una celulasa puede incluir además una secuencia de DNA o porción de la misma que codifica una región flexible o conector adherido a la secuencia de DNA de la celulasa en uno cualquiera de los extremos 5' o 3' en donde se conserva la actividad funcional del dominio de la celulasa codificada. Los términos variante y derivado pueden usarse indistintamente en esta especificación.

[34] "Residuos equivalentes" pueden definirse igualmente determinando homología en el nivel de la estructura terciaria para una celulasa precursora cuya estructura terciaria se ha determinado por cristalografía de rayos X. Residuos equivalentes se definen como aquellos para los cuales las coordenadas atómicas de dos o más de los átomos de la cadena principal de un residuo aminoácido particular de una celulasa e *Hypocrea jecorina* CBH2 (N en N, CA en CA, C en C y O en O) están en 0.13 nm y, preferiblemente, 0.1 nm después del alineamiento. El alineamiento se obtiene después de que se ha orientado y posicionado el mejor modelo para producir la superposición máxima de las coordenadas atómicas de

átomos de proteínas distintos de hidrógeno de la celulasa en cuestión con respecto a la *H. jecorina* CBH2. El mejor modelo es el modelo cristalográfico que produce el factor R mínimo para datos de difracción experimentales con la máxima resolución disponible, véase por ejemplo US 2006/0205042.

[35] Residuos equivalentes, que son funcionalmente análogos a un residuo específico de *H. jecorina* CBH2, se definen como aquellos aminoácidos de una celulasa que pueden adoptar una conformación tal que puedan alterar, modificar o contribuir a la estructura de la proteína, unión al sustrato o catálisis en una manera definida y atribuida a un residuo específico de la *H. jecorina* CBH2. Además, son aquellos residuos de la celulasa (para la cual se ha obtenido una estructura terciaria por cristalografía de rayos X) los cuales ocupan una posición análoga en la medida que, aunque los átomos de la cadena principal de un residuo dado pueden no satisfacer los criterios de equivalencia sobre la base de ocupar una posición homóloga, las coordenadas atómicas de por lo menos dos de los átomos de la cadena lateral del residuo están a 0.13 nm de los átomos de la cadena lateral correspondiente de *H. jecorina* CBH2. La estructura cristalina de *H. jecorina* CBH2 se muestra en Zou y col. (1999) (Ref. 5, supra).

[36] El término "molécula de ácido nucleico" incluye moléculas de RNA, DNA y cDNA. Se entenderá que, como resultado de la degeneración del código genético, puede producirse una multitud de secuencias nucleotídicas que codifican una proteína dada tal como CBH2 y/o variantes de la misma.

La presente divulgación contempla toda secuencia nucleotídica de una variante posible, que codifica la celulasa variante tal como CBH2, todas las cuales son posibles, dada la degeneración del código genético.

[37] Un constructo de ácido nucleico "heterólogo" o secuencia tiene una porción de la secuencia que no es nativa a la célula en la cual se expresa. Heterólogo, con respecto a una secuencia de control se refiere a una secuencia de control (es decir, promotora o potenciadora) que no funciona en la naturaleza para regular el mismo gen cuya expresión está regulando actualmente. Por lo general, secuencias de ácido nucleico heterólogo son no endógenas a la célula o parte del genoma en el cual están presentes, y se han añadido a la célula, por infección, transfección, transformación, microinyección, electroporación, o similares. Un constructo de ácido nucleico "heterólogo" puede contener una combinación de secuencia de control/secuencia codificante de DNA que es la misma que, o diferente de, una combinación de secuencia de control/secuencia codificante de

DNA hallada en la célula nativa.

[38] Tal como se emplea en esta especificación, el término "vector" se refiere a un constructo de ácido nucleico diseñado para transferencia entre diferentes células hospederas. Un "vector de expresión" se refiere a un vector que tiene la capacidad de incorporar y expresar fragmentos de DNA heterólogo en una célula extraña. En el comercio se dispone de muchos vectores de expresión procarióticos y eucarióticos. La selección de los vectores de expresión apropiados está dentro del conocimiento de los expertos en el arte.

[39] Por consiguiente, un "casete de expresión" o "vector de expresión" es un constructo de ácido nucleico generado recombinante o sintéticamente, con una serie de elementos especificados de ácido nucleico que permiten la transcripción de un ácido nucleico particular en una célula diana. El casete de expresión recombinante puede incorporarse en un DNA plasmídico, cromosómico, mitocondrial, plastídico, virus, o fragmento de ácido nucleico. Normalmente, la porción del casete de expresión recombinante de un vector de expresión incluye, entre otras secuencias, una secuencia de ácido nucleico que se va a transcribir y un promotor.

[40] Tal como se emplea en esta especificación, el término "plásmido" se refiere a un constructo de DNA de hebra doble (ds - *double-stranded*) usado como un vector de clonación, y el cual forma un elemento genético auto-replicante en muchas bacterias y algunas eucariotas.

[41] Tal como se emplea en esta especificación, el término "secuencia nucleotídica que codifica un marcador seleccionable" se refiere a una secuencia nucleotídica que es capaz de expresión en células y donde la expresión del marcador seleccionable confiere a las células, que contienen el gen expresado, la capacidad de crecer en la presencia de un agente selectivo correspondiente, o bajo condiciones correspondientes de crecimiento selectivo.

[42] Tal como se emplea en esta especificación, el término "promotor" se refiere a una secuencia de ácido nucleico que funciona para dirigir la transcripción de un gen hacia 3' ("*a downstream gene*"). Generalmente el promotor será apropiado a la célula hospedera en la cual se está expresando el gen diana. El promotor, junto con otras secuencias de ácido nucleico reguladoras transcripcionales y traduccionales (denominadas asimismo "secuencias de control"), son necesarios para expresar un gen dado. En general, las secuencias reguladoras

transcripcionales y traduccionales incluyen, pero no se limitan a, secuencias promotoras, sitios de unión ribosómica, secuencias transcripcionales de iniciación y terminación, secuencias traduccionales de iniciación y terminación, y secuencias potenciadoras o activadoras.

[43] "Gen quimérico" o "constructo de ácido nucleico heterólogo", como se define en esta especificación, se refiere a un gen no nativo (es decir, uno que se ha introducido en un hospedero), que puede estar compuesto de partes de diferentes genes, incluyendo elementos reguladores. Un constructo de gen quimérico para transformación de una célula hospedera se compone normalmente de una región reguladora transcripcional (promotor) enlazada operacionalmente a una secuencia codificante de una proteína heteróloga, o, en un gen quimérico marcador seleccionable, a un gen marcador seleccionable que codifica una proteína que confiere, por ejemplo, resistencia antibiótica a células transformadas. Un gen quimérico típico de la presente divulgación, para transformación en una célula hospedera, incluye una región reguladora transcripcional que es constitutiva o inducible, una secuencia codificante de una proteína, y una secuencia terminadora. Un constructo de un gen quimérico puede incluir además una segunda secuencia de DNA que codifica un péptido señal si se desea la secreción de la proteína diana.

[44] Un ácido nucleico está "operacionalmente enlazado" cuando se pone en relación funcional con otra secuencia de ácido nucleico. Por ejemplo, DNA que codifica un líder secretor está operacionalmente enlazado a DNA para un polipéptido si se expresa como una pre-proteína que participa en la secreción del polipéptido; un promotor o potenciador está operacionalmente enlazado a una secuencia codificante si afecta la transcripción de la secuencia; o un sitio de unión a ribosoma está operacionalmente enlazado a una secuencia codificante si está posicionado de tal modo que facilite la traducción. Generalmente, "operacionalmente enlazado" significa que las secuencias de DNA que se están enlazando son contiguas, y, en el caso de un líder secretor, contiguas y en el marco de lectura. No obstante, los potenciadores no tienen que estar contiguos. El enlazamiento se lleva a cabo por ligación en sitios de restricción convenientes. Si tales sitios no existen, se usan de acuerdo con la práctica convencional los adaptadores oligonucleotídicos sintéticos, conectores o cebadores para PCR.

[45] Tal como se emplea en esta especificación, el término "gen" significa el segmento de DNA involucrado en producir una cadena polipeptídica, que puede o no puede incluir regiones que precedan y sigan a la región codificante, v.g.

secuencias 5' no traducidas (5' UTR) o secuencias "líder" y secuencias 3' UTR o secuencias "remolque", así como secuencias intervinientes (intrones) entre segmentos codificantes individuales (exones).

[46] En general, moléculas de ácido nucleico, que codifican la celulasa variante tal como CBH2, se hibridarán, bajo condiciones de astringencia moderada a alta, a la secuencia natural tal como la provista en esta especificación como SEQ ID NO:1. No obstante, en algunos casos, se emplea una secuencia nucleotídica codificante de CBH2 que posea un empleo de codón sustancialmente diferente, mientras que la proteína codificada por la secuencia nucleotídica codificante de CBH2 tiene la misma o sustancialmente la misma secuencia de aminoácidos que la proteína nativa. Por ejemplo, la secuencia codificante puede modificarse para facilitar una expresión más rápida de CBH2 en un sistema particular de expresión procariota o eucariota, de acuerdo con la frecuencia con la cual un codón particular se utiliza por el hospedero (Te'o y col., FEMS Microbiology Letters, 190: 13-19, 2000, por ejemplo, describe la optimización de genes para expresión en hongos filamentosos).

[47] Se considera que una secuencia de ácido nucleico es "hibrizable selectivamente" a una secuencia de ácido nucleico de referencia si las dos secuencias se hibridan específicamente una a otra bajo condiciones de astringencia moderada a alta y condiciones de lavado. Las condiciones de hibridación se basan en la temperatura de fusión (T_m) del complejo de unión a ácido nucleico o sonda. Por ejemplo, la "astringencia máxima" ocurre normalmente a alrededor de $T_m - 5^\circ\text{C}$ (5°C por debajo de la T_m de la sonda); "alta astringencia" a alrededor de $5-10^\circ\text{C}$ por debajo de la T_m ; "astringencia moderada" o "intermedia" a alrededor de $10-20^\circ\text{C}$ por debajo de la T_m de la sonda; y "baja astringencia" a alrededor de $20-25^\circ\text{C}$ por debajo de la T_m . Funcionalmente, pueden usarse condiciones de máxima astringencia para identificar secuencias que tienen una identidad estricta o una identidad casi estricta con la sonda de hibridización; mientras que las condiciones de alta astringencia se usan para identificar secuencias que tienen alrededor de 80% o más de identidad de secuencia con la sonda.

[48] Condiciones de hibridación de astringencia moderada a alta son bien conocidas en el arte (véase, por ejemplo, Sambrook, y col., 1989, Capítulos 9 y 11, y en Ausubel, F. M., y col., 1993, expresamente incorporados por referencia en esta especificación). Un ejemplo de condiciones de alta astringencia incluye hibridación a alrededor de 42°C en 50% formamida, 5xSSC, 5x solución de

Denhardt, 0.5% SDS y 100 ug/ml de DNA transportador desnaturalizado seguido por lavado dos veces en 2xSSC y 0.5% SDS a temperatura ambiente y dos veces más en 0.1x SSC y 0.5% SDS a 42°C.

[49] El término "recombinante" cuando se emplea con referencia, v.g., a una célula, o ácido nucleico, proteína, o vector, indica que la célula, ácido nucleico, proteína o vector, se han modificado por la introducción de un ácido nucleico heterólogo o proteína o la alteración de un ácido nucleico nativo o proteína, o que la célula se deriva de una célula así modificada. Por consiguiente, por ejemplo, células recombinantes expresan genes que no se encuentran en la forma nativa (no recombinante) de la célula o expresan genes nativos que, de otro modo, se expresan anormalmente, se subexpresan o no se expresan en absoluto.

[50] Tal como se emplea en esta especificación, los términos "transformado", "establemente transformado " o "transgénico" con referencia a una célula significan que la célula tiene una secuencia de ácido nucleico no nativo (heterólogo) integrada en su genoma o como un plásmido episómico que se mantiene a través de múltiples generaciones.

[51] Tal como se emplea en esta especificación, el término "expresión" se refiere al proceso por el cual se produce un polipéptido basándose en la secuencia de ácido nucleico de un gen. El proceso incluye tanto la transcripción como la traducción.

[52] El término "introducido" en el contexto de insertar una secuencia de ácido nucleico en una célula, significa "transfección", o "transformación" o "transducción" e incluye referencia a la incorporación de una secuencia de ácido nucleico en una célula eucariota o procariota, donde la secuencia de ácido nucleico puede incorporarse en el genoma de la célula (por ejemplo, DNA cromosómico, plasmídico, plastídico, o mitocondrial), convertirse en un replicón autónomo, o expresarse transitoriamente (por ejemplo, mRNA transfectado).

[53] Se sigue que el término "expresión de CBH2" se refiere a la transcripción y traducción del gen *cbh2* o variantes del mismo, cuyos productos incluyen RNA precursor, mRNA, polipéptido, polipéptidos procesados pos-traduccionalmente, y derivados de los mismos, incluyendo CBH2 de especies relacionadas tales como *Trichoderma koningii*, *Hypocrea jecorina* (también conocida como *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*) e *Hypocrea schweinitzii*. A modo de ejemplo, los análisis para la expresión de CBH2 incluyen

Western blot para la proteína CBH2, análisis Northern blot y análisis de reacción en cadena de la polimerasa con transcriptasa inversa (RT-PCR) para *cbh2* mRNA, y análisis de Celulosa Esponjada con Ácido Fosfórico (PASC - *Phosphoric Acid Swollen Cellulose*) y análisis de hidrazida de ácido p-hidrobenczoico (PAHBAH - *p-hydroxybenzoic acid hydrazide*) como se describe en los siguientes: (a) PASC: (Karlsson, J. y col., (2001), Eur. J. Biochem, 268, 6498-6507, Wood, T. (1988) en *Methods in Enzymology*, Vol.160. Biomass Part a Cellulose and Hemicellulose (Wood, W. & Kellog, S. Eds.), páginas 19-25, Academic Press, San Diego, Calif., USA) y (b) PAHBAH: (Lever, M. (1972) *Analytical Biochemistry*, 47, 273, Blakeney, A. B. & Mutton, L. L. (1980) *Journal of Science of Food and Agriculture*, 31, 889, Henry, R. J. (1984) *Journal of the Institute of Brewing*, 90, 37).

[54] El término "empalme alternativo" se refiere al proceso mediante el cual múltiples isoformas polipeptídicas se generan de un único gen, e implica el corte y empalme conjunto de exones no consecutivos durante el procesamiento de algunos, mas no todos, los transcritos del gen. Por consiguiente un exón particular puede conectarse a uno cualquiera de varios exones alternativos para formar RNAs mensajeros. Los mRNAs alternativamente empalmados producen polipéptidos ("variantes de empalme") en las cuales algunas partes son comunes mientras que otras partes son diferentes.

[55] El término "secuencia señal" se refiere a una secuencia de aminoácidos en la porción N-terminal de una proteína que facilita la secreción de la forma madura de la proteína fuera de la célula. La forma madura de la proteína extracelular carece de la secuencia señal que se escinde durante el proceso de secreción.

[56] Por el término "célula hospedera" se da a entender una célula que contiene un vector y apoya la replicación, y/o transcripción o la transcripción y traducción (expresión) del constructo de expresión. Células hospederas para uso en la presente divulgación pueden ser células procariotas, tales como células de *E. coli*, o células eucariotas tales como células de levadura, células vegetales, células de insectos, células anfibias, o células mamíferas. En general, las células hospederas son hongos filamentosos.

[57] El término "hongos filamentosos" significa cualquiera y todos los hongos filamentosos reconocidos por aquellas personas de habilidad en el arte. Un hongo preferido se selecciona del grupo conformado por *Aspergillus*, *Trichoderma*, *Fusarium*, *Chrysosporium*, *Penicillium*, *Humicola*, *Neurospora*, o formas sexuales alternativas de los mismos tales como *Emericella*, *Hypocrea*. Actualmente se ha

demostrado que el hongo industrial asexual *Trichoderma reesei* es un derivado clonal del ascomicete *Hypocrea jecorina* (véase, Kuhls y col., PNAS, 93:7755-7760, 1996).

[58] El término "celooligosacárido" se refiere a grupos oligosacáridos que contienen desde 2-8 unidades de glucosa y tienen enlaces beta-1,4, v.g., celobiosa.

[59] Los términos "celulasa", "enzimas celulolíticas" o "enzimas de celulasa" se refieren a una categoría de enzimas capaces de hidrolizar polímeros de celulosa en oligómeros celooligosacáridos más cortos, celobiosa y/o glucosa. Numerosos ejemplos de celulasas, tales como exoglucanasas, exocelobiohidrolasas, endoglucanasas, y glucosidasas se han obtenido de organismos celulolíticos, incluyendo particularmente hongos, plantas y bacterias. Las enzimas elaboradas por estos microbios son mezclas de proteínas con tres tipos de acciones útiles en la conversión de celulosa en glucosa: endoglucanasas (EG), celobiohidrolasas (CBH), y beta-glucosidasa (BGL o Bglu). Estos tres diferentes tipos de enzimas de celulasa actúan sinérgicamente para convertir celulosa y sus derivados en glucosa.

[60] Muchos microbios elaboran enzimas que hidrolizan celulosa, incluyendo el hongo de pudrición de la madera *Trichoderma*, las bacterias del compost *Thermomonospora*, *Bacillus*, y *Cellulomonas*; *Streptomyces*; y los hongos *Humicola*, *Aspergillus* y *Fusarium*.

[61] *Hypocrea jecorina* CBH2 es un miembro de la Familia 6 Glicosil Hidrolasa (de ahí Cel6) y, específicamente, fue el primer miembro de esa familia identificado en *Hypocrea jecorina* (de ahí Cel6A). La Familia 6 Glicosil Hidrolasa contiene tanto endoglucanasas y celobiohidrolasas/exoglucanasas, y esa CBH2 es la última. Por consiguiente, las expresiones CBH2, proteína tipo CBH2 y Cel6 celobiohidrolasas pueden usarse indistintamente en esta especificación.

[62] El término "dominio de unión a celulosa", tal como se emplea en esta especificación, se refiere a una porción de la secuencia de aminoácidos de una celulasa o a una región de la enzima que está involucrada en la actividad de unión a celulosa de una celulasa o derivado de la misma. Dominios de unión a celulosa generalmente funcionan uniendo no covalentemente la celulasa a la celulosa, un derivado de celulosa u otro polisacárido equivalente de la misma. Dominios de unión a celulosa permiten o facilitan la hidrólisis de las fibras de celulosa por la

región central catalítica estructuralmente distinta, y normalmente funcionan independientes del centro catalítico. Por consiguiente, un dominio de unión a celulosa no poseerá la actividad hidrolítica significativa atribuible a una parte central catalítica. En otras palabras, un dominio de unión a celulosa es un elemento estructural de la estructura terciaria proteica de la enzima de celulasa, que es distinto del elemento estructural que posee actividad catalítica. Dominio de unión a celulosa y módulo de unión a celulosa pueden emplearse indistintamente en esta especificación.

[63] Tal como se emplea en esta especificación, el término "tensioactivo" se refiere a cualquier compuesto generalmente reconocido en el arte como poseedor de cualidades tensioactivas. Así, por ejemplo, los tensioactivos comprenden tensioactivos aniónicos, catiónicos y no iónicos tales como los hallados comúnmente en detergentes. Los tensioactivos aniónicos incluyen alquilbencenosulfonatos lineales o ramificados; alquil o alquenil éter sulfatos con grupos alquilo lineales o ramificados o grupos alquenilo; alquil o alquenil sulfatos; olefinsulfonatos; y alcanosulfonatos. Tensioactivos anfólicos incluyen sulfonatos de sales de amonio cuaternario, y tensioactivos anfólicos tipo betaína. Tales tensioactivos anfólicos tienen tanto los grupos cargados positivos como los negativos en la misma molécula. Los tensioactivos no iónicos pueden incluir éteres de polioxialquileo, así como alcanolamidas de ácido graso más altas o su aducto de óxido de alquileo, monoésteres de glicerina de ácido graso, y similares.

[64] Tal como se emplea en esta especificación, el término "tela con contenido de celulosa" se refiere a cualesquiera telas cosidas o no cosidas, hilos o fibras hechos de celulosa con contenido de algodón o sin contenido de algodón o mezclas de celulosa con contenido de algodón o sin contenido de algodón incluyendo celulósicos naturales y celulósicos sintéticos (tales como yute, lino, ramio, rayón, y lyocell).

[65] Tal como se emplea en esta especificación, el término "tela con contenido de algodón" se refiere a telas cosidas o no cosidas, hilos o fibras hechos de solo algodón o mezclas de algodón incluyendo telas tejidas de algodón, tejidos de punto de algodón, telas de algodón para jeans, hilos de algodón, algodón en bruto y similares.

[66] Tal como se emplea en esta especificación, el término "composición de lavado a la piedra" se refiere a una formulación para uso en lavar a la piedra telas con contenido de celulosa. Las composiciones de lavado a la piedra se utilizan

para modificar telas con contenido de celulosa antes de la venta, es decir, durante el proceso de fabricación. En contraste, las composiciones detergentes están previstas para la limpieza de prendas sucias y no se emplean durante el proceso de fabricación.

[67] Tal como se emplea en esta especificación, el término "composición detergente" se refiere a una mezcla prevista para uso en un medio de lavado para lavar y blanquear telas sucias que contienen celulosa. En el contexto de la presente divulgación, tales composiciones pueden incluir, además de celulasas y tensioactivos, enzimas hidrolíticas adicionales, mejoradores de detergente, agentes blanqueadores, activadores de blanqueado, agentes azulantes y colorantes fluorescentes, inhibidores de formación de grumos, agentes enmascarantes, activadores de celulasa, antioxidantes, y solubilizantes.

[68] Tal como se emplea en esta especificación, el término "disminución o eliminación en la expresión del gen *cbh2*" significa que o bien que el gen *cbh2* se ha eliminado del genoma y, por consiguiente, no puede expresarse por el microorganismo hospedero recombinante; o bien que el gen *cbh2* o transcrito se ha modificado de tal modo que una enzima CBH2 funcional no se produce por el microorganismo hospedero o en niveles que son significativamente menores que el gen *cbh2* no modificado o transcrito.

[69] El término "gen *cbh2* variante" significa que la secuencia de ácido nucleico del gen *cbh2* de *H. jecorina* ha sido alterado retirando, agregando, y/o manipulando la secuencia codificante.

[70] Tal como se emplea en esta especificación, el término "purificar" se refiere generalmente a someter células, que contienen ácido nucleico transgénico o proteína, a la purificación bioquímica y/o cromatografía en columna.

[71] Tal como se emplea en esta especificación, los términos "activo" y "biológicamente activo" se refieren a una actividad biológica asociada con una proteína particular y se emplean indistintamente en esta especificación. Por ejemplo, la actividad enzimática asociada con una proteasa es proteólisis y, en consecuencia, una proteasa activa tiene actividad proteolítica. Se sigue que la actividad biológica de una proteína dada se refiere a cualquier actividad biológica atribuida normalmente a esa proteína por los expertos en el arte.

[72] Tal como se emplea en esta especificación, el término "enriquecido" significa

que la celulasa tal como CBH2 se encuentra en una concentración que es mayor en relación con la concentración de CBH2 hallada en una composición de celulasa fúngica, natural o de origen natural. Los términos enriquecido, elevado y mejorado pueden emplearse indistintamente en esta especificación.

[73] Una composición de celulasa fúngica natural es una producida por una fuente fúngica de origen natural y que consta de uno o más componentes BGL, CBH y EG en donde cada uno de estos componentes se halla en la relación producida por la fuente fúngica. En consecuencia, una composición CBH enriquecida tendría CBH en una relación alterada, en donde la relación de CBH a otros componentes de celulasa (es decir, EGs, beta-glucosidasas y otras endoglucanasas) es elevada. Esta relación puede aumentarse incrementando la CBH o disminuyendo (o eliminando) por lo menos otro componente por cualquier medio conocido en el arte.

[74] El término "aislado" o "purificado", tal como se emplea en esta especificación, se refiere a un ácido nucleico o aminoácido que se retira de por lo menos un componente con el cual está asociado naturalmente.

[75] Así, para ilustrar, un sistema de celulasas de origen natural puede purificarse en componentes sustancialmente puros por técnicas reconocidas de separación publicadas adecuadamente en la literatura, incluyendo cromatografía de intercambio iónico con un pH apropiado, cromatografía de afinidad, por exclusión de tamaño y similares. Por ejemplo, en la cromatografía de intercambio iónico (usualmente cromatografía de intercambio aniónico), es posible separar los componentes de celulasa eluyendo con un gradiente de pH, o un gradiente salino, o ambos, un gradiente de pH y un gradiente salino. La CBH purificada puede agregarse entonces a la solución enzimática resultando en una solución de CBH enriquecida. Es igualmente posible elevar la cantidad de CBH producida por un microbio usando métodos de genética molecular para sobreexpresar el gen que codifica CBH, posiblemente en conjunción con la eliminación de uno o más genes que codifican otras celulasas.

[76] Celulasas fúngicas pueden contener más de un componente CBH. Los diferentes componentes tienen generalmente diferentes puntos isoeléctricos los cuales permiten su separación mediante cromatografía de intercambio iónico, y similares. Un componente CBH único o una combinación de componentes CBH pueden emplearse en una solución enzimática.

[77] Cuando se emplea en soluciones enzimáticas, el componente CBH2 variante u homólogo se agrega generalmente en una cantidad suficiente para permitir la tasa más alta de liberación de azúcares solubles de la biomasa. La cantidad de componente CBH2 variante u homólogo agregado depende del tipo de biomasa que se va a sacarificar, la cual puede determinarse fácilmente por el artesano experto. Cuando se emplea, el porcentaje en peso del componente CBH2 variante u homólogo presente en la composición de celulasa es desde preferiblemente entre 1 y 100, siendo ejemplos ilustrativos alrededor de 1, preferiblemente alrededor de 5, preferiblemente alrededor de 10, preferiblemente alrededor de 15, o preferiblemente alrededor de 20 por ciento en peso a preferiblemente alrededor de 25, preferiblemente alrededor de 30, preferiblemente alrededor de 35, preferiblemente alrededor de 40, preferiblemente alrededor de 45 o preferiblemente alrededor de 50 por ciento en peso. Más aún, intervalos preferidos pueden ser alrededor de 0.5 a alrededor de 15 por ciento en peso, alrededor de 0.5 a alrededor de 20 por ciento en peso, desde alrededor de 1 a alrededor de 10 por ciento en peso, desde alrededor de 1 a alrededor de 15 por ciento en peso, desde alrededor de 1 a alrededor de 20 por ciento en peso, desde alrededor de 1 a alrededor de 25 por ciento en peso, desde alrededor de 5 a alrededor de 20 por ciento en peso, desde alrededor de 5 a alrededor de 25 por ciento en peso, desde alrededor de 5 a alrededor de 30 por ciento en peso, desde alrededor de 5 a alrededor de 35 por ciento en peso, desde alrededor de 5 a alrededor de 40 por ciento en peso, desde alrededor de 5 a alrededor de 45 por ciento en peso, desde alrededor de 5 a alrededor de 50 por ciento en peso, desde alrededor de 10 a alrededor de 20 por ciento en peso, desde alrededor de 10 a alrededor de 25 por ciento en peso, desde alrededor de 10 a alrededor de 30 por ciento en peso, desde alrededor de 10 a alrededor de 35 por ciento en peso, desde alrededor de 10 a alrededor de 40 por ciento en peso, desde alrededor de 10 a alrededor de 45 por ciento en peso, desde alrededor de 10 a alrededor de 50 por ciento en peso, desde alrededor de 15 a alrededor de 60 por ciento en peso, desde alrededor de 15 a alrededor de 70 por ciento en peso, desde alrededor de 15 a alrededor de 75 por ciento en peso, desde alrededor de 15 a alrededor de 80 por ciento en peso, desde alrededor de 15 a alrededor de 85 por ciento en peso, desde alrededor de 15 a alrededor de 95 por ciento en peso. No obstante, cuando se emplea, el porcentaje en peso del componente de CBH2 variante u homólogo, relativo a cualesquiera componentes tipo EG presentes en la composición de celulasa es desde preferiblemente alrededor de 1, preferiblemente alrededor de 5, preferiblemente alrededor de 10, preferiblemente alrededor de 15, o preferiblemente alrededor de 20 por ciento en peso a preferiblemente alrededor de 25, preferiblemente alrededor de 30,

preferiblemente alrededor de 35, preferiblemente alrededor de 40, preferiblemente alrededor de 45 o preferiblemente alrededor de 50 por ciento en peso. Más aún, intervalos preferidos pueden ser alrededor de 0.5 a alrededor de 15 por ciento en peso, alrededor de 0.5 a alrededor de 20 por ciento en peso, desde alrededor de 1 a alrededor de 10 por ciento en peso, desde alrededor de 1 a alrededor de 15 por ciento en peso, desde alrededor de 1 a alrededor de 20 por ciento en peso, desde alrededor de 1 a alrededor de 25 por ciento en peso, desde alrededor de 5 a alrededor de 20 por ciento en peso, desde alrededor de 5 a alrededor de 25 por ciento en peso, desde alrededor de 5 a alrededor de 30 por ciento en peso, desde alrededor de 5 a alrededor de 35 por ciento en peso, desde alrededor de 5 a alrededor de 40 por ciento en peso, desde alrededor de 5 a alrededor de 45 por ciento en peso, desde alrededor de 5 a alrededor de 50 por ciento en peso, desde alrededor de 10 a alrededor de 20 por ciento en peso, desde alrededor de 10 a alrededor de 25 por ciento en peso, desde alrededor de 10 a alrededor de 30 por ciento en peso, desde alrededor de 10 a alrededor de 35 por ciento en peso, desde alrededor de 10 a alrededor de 40 por ciento en peso, desde alrededor de 10 a alrededor de 45 por ciento en peso, desde alrededor de 10 a alrededor de 50 por ciento en peso, desde alrededor de 15 a alrededor de 20 por ciento en peso, desde alrededor de 15 a alrededor de 25 por ciento en peso, desde alrededor de 15 a alrededor de 30 por ciento en peso, desde alrededor de 15 a alrededor de 35 por ciento en peso, desde alrededor de 15 a alrededor de 40 por ciento en peso, desde alrededor de 15 a alrededor de 45 por ciento en peso, desde alrededor de 15 a alrededor de 50 por ciento en peso.

II. Celulasas

[78] Las celulasas son conocidas en el arte como enzimas que hidrolizan celulosa (enlaces beta-1,4-glucano o beta D-glucosídicos) lo cual resulta en la formación de glucosa, celobiosa, celooligosacáridos, y similares. Como ya se expuso, las celulasas se han dividido tradicionalmente en tres clases principales: endoglucanasas (EC 3.2.1.4) ("EG"), exoglucanasas o celobiohidrolasas (EC 3.2.1.91) ("CBH") y beta-glucosidasas (EC 3.2.1.21) ("BG").

[79] Ciertos hongos producen sistemas completos de celulasas, los cuales incluyen exo-celobiohidrolasas o celulasas tipo CBH, endoglucanasas o celulasas tipo EG y beta-glucosidasas o celulasas tipo BG. Sin embargo, a veces estos sistemas carecen de celulasas tipo CBH, y las celulasas bacterianas incluyen además normalmente pocas o ningunas celulasas tipo CBH. Adicionalmente, se ha demostrado que los componentes EG y los componentes CBH interactúan

sinérgicamente para degradar la celulosa de manera más eficiente. Los diferentes componentes, es decir, las varias endoglucanasas y exocelobiohidrolasas en un sistema de celulasas multi-componente o completo, generalmente tienen diferentes propiedades, tales como punto isoeléctrico, peso molecular, grado de glicosilación, especificidad del sustrato y patrones de acción enzimática.

[80] Se cree que las celulasas tipo endoglucanasa hidrolizan enlaces beta-1,4-glucosídicos internos en regiones de baja cristalinidad de la celulosa y que las celulasas tipo exo-celobiohidrolasa hidrolizan celobiosa del extremo reductor o no reductor de la celulosa. Se sigue que la acción de los componentes endoglucanasa puede facilitar en gran medida la acción de las exocelobiohidrolasas creando nuevos extremos de cadena, los cuales son reconocidos por componentes exo-celobiohidrolasa. Además, se ha demostrado que las celulasas tipo beta-glucosidasa catalizan la hidrólisis de alquil y/o aril beta-D-glucósidos tales como metil beta-D-glucósido y p-nitrofenil glucósido, así como glicósidos que contienen únicamente residuos carbohidratos, tales como celobiosa. Esto produce glucosa como el solo producto para el microorganismo y reduce o elimina la celobiosa, la cual inhibe las celobiohidrolasas y las endoglucanasas.

[81] Las celulasas encuentran asimismo un sinnúmero de usos en composiciones detergentes incluyendo potenciar la capacidad limpiadora, como agente suavizante y para mejorar el tacto de las telas de algodón (Hemmpel, ITB Dyeing/Printing/Finishing 3:5-14, 1991; Tyndall, Textile Chemist and Colorist 24:23-26, 1992; y Kumar y col., Textile Chemist and Colorist, 29:37-42, 1997). Aunque el mecanismo no es parte de la divulgación, las propiedades de suavizado y restauración del color de la celulosa se ha atribuido a los componentes alcalinos de la endoglucanasa en composiciones de celulosa, como se ejemplifica por las Patentes de los Estados Unidos Nos. 5,648,263, 5,691,178, y 5,776,757, las cuales divulgan que composiciones detergentes que contienen una composición de celulosa enriquecida con un componente alcalino de endoglucanasa especificado imparten restauración del color y una suavidad mejorada a las prendas tratadas, en comparación con composiciones de celulosa no enriquecidas con tal componente. Además, se ha demostrado que el uso de tales componentes alcalinos de endoglucanasa en composiciones detergentes complementan los requerimientos de pH de la composición detergente (v.g., exhibiendo actividad máxima con un pH alcalino de 7.5 a 10, como se describe en las Patentes de los Estados Unidos Nos. 5,648,263, 5,691,178, y 5,776,757).

[82] Se ha demostrado igualmente que las composiciones de celulasa degradan telas que contienen algodón, con el resultado de una pérdida reducida de resistencia en la tela (Patente de los Estados Unidos No. 4,822,516), lo cual contribuye a la renuencia a usar composiciones de celulasa en aplicaciones detergentes comerciales. Se ha sugerido que las composiciones de celulasa que comprenden componentes de endoglucanasa exhiben una pérdida reducida de resistencia para telas con contenido de algodón, comparadas con composiciones que comprenden un sistema completo de celulasas.

[83] De igual modo, se ha demostrado que las celulasas son útiles en la degradación de biomasa de celulasa a etanol (en donde la celulasa degrada celulosa a glucosa, y levadura u otros microbios fermentan posteriormente la glucosa en etanol), en el tratamiento de pulpa mecánica (Pere y col., In Proc. Tappi Pulping Conf., Nashville, Tenn., 27-31, páginas 693-696, 1996), para uso como aditivo de forrajes (WO 91/04673) y en la molienda húmeda de granos.

[84] La mayoría de las CBHs y EGs tienen una estructura multi-dominio conformada por un dominio central, separado de un dominio de unión a celulosa (CBD – *cellulose binding domain*) por un péptido conector (Suurnakki y col., 2000). El dominio central contiene el sitio activo mientras que el CBD interactúa con celulosa uniendo la enzima a él (van Tilbeurgh y col., FEBS Lett. 204:223-227, 1986; Tomme y col., Eur. J. Biochem. 170:575-581, 1988). Los CBDs son particularmente importantes en la hidrólisis de celulosa cristalina. Se ha demostrado que la capacidad de las celobiohidrolasas de degradar celulosa cristalina disminuye claramente cuando el CBD está ausente (Linder y Teeri, J. Biotechnol. 57:15-28, 1997). No obstante, el papel exacto y el mecanismo de acción de los CBDs es aún materia de especulación. Se ha sugerido que el CBD potencia la actividad enzimática aumentando simplemente la concentración enzimática efectiva en la superficie de la celulosa (Stahlberg y col., Bio/Technol. 9:286-290, 1991), y/o aflojando cadenas simples de celulosa de la superficie de la celulosa (Torino y col., EMBO J. vol. 15, no. 21, páginas 5739-5751, 1996). La mayoría de los estudios concernientes a los efectos de los dominios de celulasa sobre diferentes sustratos se han llevado a cabo con proteínas centrales ("**core proteins**") de celobiohidrolasas, ya que sus proteínas centrales pueden producirse fácilmente por proteólisis limitada con papaína (Tomme y col., 1988). Numerosas celulasas se han descrito en la literatura científica, ejemplos de las cuales incluyen: de *Trichoderma reesei*: Shoemaker, S. y col., Bio/Technology, 1:691-696, 1983, que divulga CBH1; Teeri, T. y col., Gene, 51:43-52, 1987, que divulga CBH2. Asimismo se han descrito celulasas de especies distintas a

Trichoderma v.g., Ooi y col., Nucleic Acids Research, vol. 18, No. 19, 1990, que divulga la secuencia de cDNA que codifica la endoglucanasa F1-CMC producida por *Aspergillus aculeatus*; Kawaguchi T y col., Gene 173(2):287-8, 1996, que divulga la clonación y secuenciación del cDNA que codifica beta-glucosidasa 1 de *Aspergillus aculeatus*; Sakamoto y col., Curr. Genet. 27:435-439, 1995, que divulga la secuencia de cDNA que codifica la endoglucanasa CMCase-1 de *Aspergillus kawachii* IFO 4308; Saarilahti y col., Gene 90:9-14, 1990, que divulga una endoglucanasa de *Erwinia carotovora*; Spilliaert R, y col., Eur J Biochem. 224(3):923-30, 1994, que divulga la clonación y secuenciación de bg1A, que codifica una beta-glucanasa termoestable de *Rhodothermus marinus*; y Halldorsdottir S y col., Appl Microbiol Biotechnol. 49(3):277-84, 1998, que divulga la clonación, secuenciación y sobreexpresión de un gen *Rhodothermus marinus* que codifica una celulasa termoestable de la familia 12 glicosil hidrolasa. Sin embargo, subsiste una necesidad de identificación y caracterización de celulasas novedosas, con propiedades mejoradas, tales como un mayor rendimiento bajo condiciones de estrés térmico o en la presencia de tensioactivos, una mayor actividad específica, un patrón alterado de escisión del sustrato, y/o alto nivel de expresión *in vitro*.

[85] El desarrollo de composiciones de celulasa novedosas y mejoradas que contienen cantidades variables de celulasas tipo CBH, tipo EG y tipo BG es de interés para uso: (1) en composiciones para degradar pulpa de madera u otra biomasa en azúcares (v.g., para la producción de bioquímicos tales como biocombustibles); (2) en composiciones detergentes que exhiben una capacidad limpiadora mejorada (3) funcionan como agente suavizante y/o mejoran el tacto de telas de algodón (v.g., "lavado a la piedra" o "biodesfibrilación") ("**stone washing**" or "**biopolishing**"); y/o (3) en composiciones para forrajes, por ejemplo.

[86] También se proveen en esta especificación preparaciones de celulasa entera que comprenden variantes de celulasa. Tal como se emplea en esta especificación, la expresión "preparación de celulasa entera" se refiere a composiciones que contienen celulasa de origen natural y de origen no natural. Una composición "de origen natural" es una producida por una fuente de origen natural y que comprende uno o más componentes tipo celobiohidrolasa, uno o más componentes tipo endoglucanasa, y uno o más componentes tipo beta-glucosidasa, en donde cada uno de estos componentes se encuentra en la relación producida por la fuente. Una composición de origen natural es una que es producida por un organismo no modificado con respecto a las enzimas celulolíticas, de modo que la relación de las enzimas componentes permanece

inalterada de las producidas por el organismo natural. Una composición "de origen no natural" abarca aquellas composiciones producidas: (1) combinando enzimas celulolíticas componentes en una relación de origen natural o en una relación de origen no natural, es decir, alterada; o (2) modificando un organismo para sobreexpresar o subexpresar una o más enzimas celulolíticas; o (3) modificando un organismo de modo que por lo menos se elimine una enzima celulolítica. Por consiguiente, en algunas incorporaciones, la preparación de celulosa entera puede tener eliminadas una o más de las diferentes EGs y/o CBHs, y/o beta-glucosidasa. Por ejemplo, EG1 puede eliminarse sola o en combinación con otras EGs y/o CBHs.

[87] En general, la preparación de celulosa entera contiene enzimas incluyendo, pero no limitadas a: (i) endoglucanasas (EG) o 1,4- β -D-glucan-4-glucanohidrolasas (EC 3.2.1.4), (ii) exoglucanasas, incluyendo 1,4- β -D-glucano glucanohidrolasas (también conocidas como celodextrinasas) (EC 3.2.1.74) y 1,4- β -D-glucano celobiohidrolasas (exo-celobiohidrolasas, CBH) (EC 3.2.1.91), y (iii) β -glucosidasa (BG) o β -glucósido glucohidrolasas (EC 3.2.1.21).

[88] En la presente divulgación, la preparación de celulosa entera puede ser de cualquier microorganismo que sea útil para la hidrólisis de un material celulósico. En algunas incorporaciones, la preparación de celulosa entera es una celulosa entera de hongos filamentosos. "Hongos filamentosos" incluyen todas las formas filamentosas de la subdivisión *Eumycota* y *Oomycota*. En algunas incorporaciones, la preparación de celulosa entera es una celulosa entera de la especie *Acremonium*, *Aspergillus*, *Emericella*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Scytalidium*, *Thielavia*, *Tolypocladium*, o *Trichoderma*. En algunas incorporaciones, la preparación de celulosa entera es una celulosa entera de *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, o *Aspergillus oryzae*. En otro aspecto, la preparación de celulosa entera es una celulosa entera de *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, o *Fusarium venenatum*. En otro aspecto, la preparación de celulosa entera es una celulosa entera de *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Scytalidium thermophilum*, o *Thielavia terrestris*. En otro aspecto, la preparación de celulosa entera es una celulosa de

Trichoderma harzianum, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* v.g., RL-P37 (Sheir-Neiss y col., Appl. Microbiol. Biotechnology, 20 (1984) páginas 46-53; Montenecourt B.S., Can., 1-20, 1987), QM9414 (ATCC No. 26921), NRRL 15709, ATCC 13631, 56764, 56466, 56767, o *Trichoderma viride* v.g., ATCC 32098 y 32086. En algunas incorporaciones, la preparación de celulasa entera es una celulasa entera *Trichoderma reesei* RutC30, la cual se encuentra disponible de la American Type Culture Collection como *Trichoderma reesei* ATCC 56765.

[89] Ejemplos de preparaciones comerciales de celulasa apropiadas para uso en la presente divulgación incluyen, por ejemplo, CELLUCLAST™ (disponible de Novozymes A/S) y la enzima LAMINEX™, IndiAge™ y Primafast™ LAMINEX BG, ACCELLERASE™ 100 y ACCELLERASE™ 1500 (disponibles de Genencor Division, Danisco US. Inc.)

[90] En la presente divulgación, la preparación de celulasa entera puede provenir de cualquier método de cultivo de microorganismos conocido en el arte, resultante en la expresión de enzimas capaces de hidrolizar un material celulósico. La fermentación puede incluir cultivo en frascos de agitación, fermentación a pequeña o gran escala, tal como fermentaciones continuas, en lote, en lote alimentado o fermentaciones en estado sólido en fermentadores de laboratorio o fermentadores industriales, realizadas en un medio apropiado y bajo condiciones que permitan expresar o aislar la celulasa.

[91] Generalmente, el microorganismo se cultiva en un medio de cultivo celular apropiado para la producción de enzimas capaces de hidrolizar un material celulósico. El cultivo tiene lugar en un nutriente apropiado constituido por fuentes de carbono y nitrógeno y sales inorgánicas, usando procedimientos conocidos en el arte. Medios de cultivo apropiados, intervalos de temperatura y otras condiciones apropiadas para crecimiento y producción de celulasa son conocidos en el arte. Como ejemplo no limitante, el intervalo de temperatura normal para la producción de celulasas por *Trichoderma reesei* es 24°C a 28°C.

[92] Generalmente, la preparación de celulasa entera se usa como se produce por fermentación sin ninguna o mínima recuperación y/o purificación. Por ejemplo, una vez que las celulasas son secretadas por una célula en el medio de cultivo celular, puede usarse el medio de cultivo celular que contiene las celulasas. En algunas incorporaciones la preparación de celulasa entera comprende el contenido no fraccionado de material de fermentación, incluyendo medio de cultivo celular,

enzimas extracelulares y células. Alternativamente, la preparación de celulasa entera puede procesarse por cualquier método conveniente, v.g., por precipitación, centrifugación, afinidad, filtración o cualquier otro método conocido en el arte. En algunas incorporaciones, la preparación de celulasa entera puede concentrarse, por ejemplo, y luego utilizarse sin purificación posterior. En algunas incorporaciones la preparación de celulasa entera contiene agentes químicos que disminuyen la viabilidad celular o eliminan las células. En algunas incorporaciones, las células se lisan o permeabilizan usando métodos conocidos en el arte.

III. Biología Molecular

[93] En una incorporación esta divulgación posibilita la expresión de genes *cbh2* variantes bajo el control de un promotor funcional en un hongo filamentoso. Por consiguiente, esta divulgación se basa en técnicas rutinarias en el campo de la genética recombinante (véase, v.g., Sambrook y col., *Molecular Cloning, A Laboratory Manual*, 2nd ed., 1989; Kriegler, *Gene Transfer and Expression: A Laboratory Manual*, 1990; y Ausubel y col., eds., *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, Greene Publishing and Wiley-Interscience, New York, 1994).

Métodos o de Mutar Secuencias de Ácido Nucleico *cbh2*

[94] Cualquier método conocido en el arte que pueda introducir mutaciones está contemplado por la presente divulgación.

[95] La presente divulgación se refiere a la expresión, purificación y/o aislamiento y uso de la variante CBH2. Estas enzimas se preparan preferiblemente por métodos recombinantes utilizando el gen *cbh2* de *H. jecorina*. El caldo de fermentación puede usarse con o sin purificación.

[96] Luego del aislamiento y clonación del gen *cbh2* de *H. jecorina*, se emplean otros métodos conocidos en el arte, tales como mutagénesis dirigida al sitio, para hacer las sustituciones, adiciones o eliminaciones que correspondan a los aminoácidos sustituidos en la variante CBH2 expresada. Nuevamente, mutagénesis dirigida al sitio y otros métodos de incorporar cambios de aminoácidos en proteínas expresadas en el nivel DNA son conocidos en el arte (Sambrook y col., supra; y Ausubel y col., supra).

[97] DNA que codifica una variante de secuencia de aminoácidos de la *H. jecorina*

CBH2 se prepara por una variedad de métodos conocidos en el arte. Estos métodos incluyen, pero no se limitan a, la preparación por mutagénesis dirigida al sitio (o mediada por oligonucleótidos), mutagénesis por PCR, y mutagénesis de casete de un DNA preparado anteriormente que codifica la *H. jecorina* CBH2.

[98] La mutagénesis dirigida al sitio es un método preferido para preparar variantes de sustitución. Esta técnica es bien conocida en el arte (véase, v.g., Carter y col., *Nucleic Acids Res.* 13:4431-4443 (1985) y Kunkel y col., *Proc. Natl. Acad. Sci. USA* 82:488 (1987)). En síntesis, cuando se lleva a cabo mutagénesis de DNA dirigida al sitio, el DNA inicial se altera hibridando primeramente un oligonucleótido, que codifica la mutación deseada, a una sola hebra de tal DNA inicial. Luego de la hibridación, se emplea una DNA polimerasa para sintetizar una segunda hebra completa, usando el oligonucleótido hibridado como cebador, y usando la hebra única del DNA inicial como molde. Por consiguiente, el oligonucleótido que codifica la mutación deseada se incorpora en el DNA de doble hebra resultante.

[99] Las mutagénesis por PCR es igualmente apropiada para elaborar variantes de secuencias de aminoácidos del polipéptido inicial, es decir, *H. jecorina* CBH2. Véase Higuchi, en *PCR Protocols*, páginas 177-183 (Academic Press, 1990); y Vallette y col., *Nuc. Acids Res.* 17:723-733 (1989). Véase, asimismo, por ejemplo Cadwell y col., *PCR Methods and Applications*, Vol 2, 28-33 (1992). En síntesis, cuando pequeñas cantidades de DNA molde se utilizan como material de partida en una PCR, pueden emplearse cebadores que difieren ligeramente en secuencia de la región correspondiente en un DNA molde para generar cantidades relativamente grandes de un fragmento de DNA específico, que difiere de la secuencia molde únicamente en las posiciones en las que los cebadores difieren del molde.

[100] Otro método para preparar variantes, la mutagénesis de casete, se basa en la técnica descrita por Wells y col., *Gene* 34:315-323 (1985). El material de partida es el plásmido (u otro vector) que contiene el DNA polipeptídico inicial que se va a mutar. Se identifica(n) el codón (los codones) en el DNA de partida que se va a mutar. Tiene que haber un sitio único de endonucleasa de restricción en cada lado del sitio(s) de mutación identificado(s). Si no existen tales sitios de restricción, éstos pueden generarse usando el método ya descrito de mutagénesis mediada por oligonucleótido para introducirlos en localizaciones apropiadas en el DNA polipeptídico inicial. El DNA plasmídico se corta en estos sitios para linearizarlo. Un oligonucleótido de doble hebra, que codifica la secuencia del DNA entre los

sitios de restricción pero contiene la(s) mutación (mutaciones) deseada(s), se sintetiza usando procedimientos estándar, en donde las dos hebras del oligonucleótido se sintetizan separadamente y luego se hibridan entre sí empleando técnicas estándar. Este oligonucleótido de doble hebra se denomina el casete. Este casete se diseña para tener extremos 5' y 3', que son compatibles con los extremos del plásmido linearizado, de modo que pueda ligarse directamente al plásmido. Este plásmido contiene ahora la secuencia de DNA mutado.

[101] Alternativa, o adicionalmente, puede determinarse la secuencia de aminoácidos deseada que codifica una variante CBH2, y una secuencia de ácido nucleico que codifica tal secuencia de aminoácidos variante puede generarse sintéticamente.

[102] La (las) variante(s) CBH2(s) así preparada(s) puede(n) someterse a modificaciones adicionales, dependiendo a menudo del uso previsto de la celulasas. Tales modificaciones pueden involucrar alteración adicional de la secuencia de aminoácidos, fusión a un polipéptido(s) heterólogo(s) y/o modificaciones covalentes.

IV. Ácidos Nucleicos *cbh2* y Polipéptidos CBH2

A. Ácidos Nucleicos de Tipo *cbh2* Variante

[103] La secuencia de ácido nucleico para el *cbh2* natural se muestra en SEQ ID NO:1. La divulgación abarca una molécula de ácido nucleico que codifica las celulasas variantes descritas en esta especificación. El ácido nucleico puede ser una molécula de DNA.

[104] Una vez que las secuencias de DNA, que codifican las variantes CBH2, se han clonado en constructos de DNA, el DNA se usa para transformar microorganismos. El microorganismo que se va a transformar con el propósito de expresar una variante CBH2 de acuerdo con la presente divulgación puede incluir ventajosamente una cepa derivada de *Trichoderma sp.* Por consiguiente, un modo preferido para preparar celulasas variantes CBH2 de acuerdo con la presente divulgación incluye transformar una célula hospedera *Trichoderma sp.* con un constructo de DNA, que contiene por lo menos un fragmento de DNA que codifica una porción o la totalidad de la variante CBH2. Por lo general, el constructo de DNA estará fijado funcionalmente a un promotor. La célula hospedera

transformada se cultiva entonces bajo condiciones tales que expresen la proteína deseada. Posteriormente, el producto proteína deseada puede purificarse a una homogeneidad sustancial.

[105] No obstante, en realidad, puede ser que el mejor vehículo de expresión para un DNA dado, que codifica una variante CBH2, difiera de *H. jecorina*. Por lo tanto, es posible que sea muchísimo más ventajoso expresar una proteína en un hospedero de transformación que tenga similitud filogenética con el organismo fuente para la variante CBH2. En una incorporación alternativa, puede usarse *Aspergillus niger* como un vehículo de expresión. Para una descripción de las técnicas de transformación con *A. niger*, véase WO 98/31821, cuya divulgación se incorpora por referencia en su totalidad.

[106] Por consiguiente, se provee la presente descripción de un sistema de expresión de *Aspergillus* spp. con fines de ilustración únicamente y como una opción para expresar la variante CBH2 de la divulgación. Una persona experta en el arte, sin embargo, puede estar inclinada a expresar el DNA que codifica la variante CBH2 en una célula hospedera diferente si es apropiado, y debe entenderse que la fuente de la variante CBH2 debe considerarse al determinar el hospedero de expresión óptimo. Adicionalmente, el trabajador experto en el campo estará en capacidad de seleccionar el mejor sistema de expresión para un gen particular mediante técnicas rutinarias, utilizando las herramientas disponibles en el arte.

B. Polipéptidos de CBH2 Variante

[107] Las CBH2 variantes de esta divulgación tienen secuencias de aminoácidos que se derivan de la secuencia de aminoácidos de una CBH2 precursora. La secuencia de aminoácidos de la variante CBH2 difiere de la secuencia de aminoácidos de la CBH2 precursora por la sustitución, eliminación o inserción de uno o más aminoácidos de la secuencia de aminoácidos precursora. En una incorporación preferida, la CBH2 precursora es *Hypocrea jecorina* CBH2. La secuencia de aminoácidos madura de *H. jecorina* CBH2 se muestra en SEQ ID NO:3. Por lo tanto, esta divulgación se dirige a CBH2 variantes que contienen residuos aminoácidos en posiciones que son equivalentes al residuo particular identificado en *H. jecorina* CBH2. Un residuo (aminoácido) de un homólogo CBH2 es equivalente a un residuo de *Hypocrea jecorina* CBH2 si es homólogo (es decir, correspondiente en posición en cualquiera estructura primaria o terciaria) o es funcionalmente análogo a un residuo específico o porción de ese residuo en

Hypocrea jecorina CBH2 (es decir, que tiene la misma capacidad funcional o similar de combinar, reaccionar, o interactuar química o estructuralmente). Tal como se emplea en esta especificación, se pretende que la numeración corresponda a la de la secuencia de aminoácidos de CBH2 madura (SEQ ID NO:3).

[108] El alineamiento de secuencias de aminoácidos para determinar la homología se determina preferiblemente usando un "algoritmo de comparación de secuencias." El alineamiento óptimo de secuencias para comparación puede conducirse, v.g., por el algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2:482 (1981), por el algoritmo de alineamiento de homología de Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), por el método de búsqueda de similitud de Pearson & Lipman, Proc. Nat'l Acad. Sci. USA 85:2444 (1988), por implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA, y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), o mediante inspección visual. La inspección visual puede utilizar paquetes de gráficos tales como, por ejemplo, MOE por Chemical Computing Group, Montreal, Canadá.

[109] Un ejemplo de un algoritmo que es apropiado para determinar la similitud de secuencias es el algoritmo BLAST, el cual se describe en Altschul, y col., J. Mol. Biol. 215:403-410 (1990). El software para efectuar análisis BLAST se encuentra públicamente disponible a través del National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). Este algoritmo involucra primeramente identificar pares de secuencias de alta puntuación (HSPs - *high scoring sequence pairs*) identificando palabras cortas de longitud W en la secuencia problema que coincidan o satisfagan alguna puntuación T umbral de valor positivo cuando se alineen con una palabra de la misma longitud en una secuencia de base de datos. Estos aciertos de palabras de vecindad inicial actúan como puntos de partida para encontrar HSPs más largos que los contengan. Los aciertos de palabras se extienden en ambas direcciones a lo largo de cada una de las dos secuencias que se están comparando, en la medida en que la puntuación de alineamiento acumulativo pueda aumentarse. La extensión de los aciertos de palabras se suspende cuando: el puntaje de alineamiento acumulativo cae fuera de la cantidad X de un valor máximo alcanzado; el puntaje acumulativo tiende a cero o por debajo; o se alcanza el extremo de cualquiera de las dos secuencias. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad del alineamiento. El programa BLAST utiliza como valores predeterminados una longitud de palabra (W) de 11, la matriz de puntuación BLOSUM62 (véase

Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)) las alineaciones (B) de 50, expectativa (E) de 10, M'5, N'-4, y una comparación de ambas hebras.

[110] El algoritmo BLAST realiza entonces un análisis estadístico de la similitud entre dos secuencias (véase, v.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). Una medida de similitud provista por el algoritmo BLAST es la probabilidad de suma mínima (P(N)), que ofrece una indicación de la probabilidad mediante la cual ocurrirá casualmente una coincidencia entre dos secuencias de nucleótidos o aminoácidos. Por ejemplo, una secuencia de aminoácidos se considera similar a una proteasa si la probabilidad de suma mínima en una comparación de la secuencia de aminoácidos de prueba a una secuencia de aminoácidos de la proteasa es menor de alrededor de 0.1, más preferiblemente menor de alrededor de 0.01 y, muy preferiblemente, menor de alrededor de 0.001.

[111] Para efectos de la presente divulgación, el grado de identidad puede determinarse apropiadamente por medio de programas de computador conocidos en el arte, tales como GAP provisto en paquete de programas GCG (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wis., USA 53711) (Needleman, S. B. y Wunsch, C. D., (1970), Journal of Molecular Biology, 48, 443-45), usando GAP con las siguientes configuraciones para comparación de secuencias polinucleotídicas: penalización de creación GAP de 5.0 y penalización de extensión GAP de 0.3.

[112] Un alineamiento estructural entre una *T. reesei* CBH2 y otras celulasas puede usarse para identificar posiciones equivalentes/correspondientes en otras celulasas que tienen un grado de homología moderado a alto, v.g., alrededor de 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o incluso 99%, con *T. reesei* CBH2 (SEQ ID NO: 3). Un método de obtener el citado alineamiento estructural es usar el programa Pile Up del paquete GCG utilizando valores predeterminados de penalizaciones de espacios, es decir, una penalización por creación de espacios de 3.0 y una penalización por extensión de espacios de 0.1. Otros métodos de alineamiento estructural incluyen el análisis de clúster hidrofóbico (Gaboriaud y col., FEBS Letters, 224:149-155, 1987) y concatenado inverso ("**reverse threading**") (Huber y Torda, Protein Science, 7:142-149, 1998).

[113] Un alineamiento ejemplar de la forma madura de varias celulasas de referencia se provee como Figura 3. Las celulasas de referencia incluyen:

Hypocrea jecorina (también conocida como *T. reesei*) CBH2 (SEQ ID NO:3), *Hypocrea koningii* CBH2 (SEQ ID NO:4), *Humicola insolens* CBH2 (SEQ ID NO:5), *Acremonium cellulolyticus* CBH2 (SEQ ID NO:6), *Agaricus bisporus* CBH2 (SEQ ID NO:7), *Fusarium osysporum* EG (SEQ ID NO:8), *Phanerochaete chrysosporium* CBH2 (SEQ ID NO:9), *Talaromyces emersonii* CBH2 (SEQ ID NO: 10), *Thermobifida fusca* 6B/E3 CBH2 (SEQ ID NO: 11), *Thermobifida fusca* 6A/E2 EG (SEQ ID NO: 12), y *Cellulomonas fimi* CenA EG (SEQ ID NO: 13). Las secuencias se alinearon empleando los algoritmos de alineamiento de secuencias múltiples ClustalW y MUSCLE. Una matriz que muestra la identidad porcentual de celulasas del alineamiento de secuencias de Figura 3 se provee en la Tabla 1.

Tabla 1. Matriz de Identidad Porcentual de Celulasa*

ID porcentual	3	4	5	6	7	8	9	10	11	12	13
3	100	95.5	62.3	64.7	59.6	63.1	55.4	63.4	31.9	13.5	27
4	95.5	100	61.6	64	59.1	63.6	54.7	63	32.9	13.5	26.8
5	62.3	61.6	100	59.1	57.6	61.3	54	58.8	31.9	15.9	26.6
6	64.7	64	59.1	100	58.6	56.4	54	72.6	32.8	13.5	29.2
7	59.6	59.1	57.6	58.6	100	55.8	69.1	58.1	34.9	17.5	27.6
8	63.1	63.6	61.3	56.4	55.8	100	48.7	54.8	31.1	13.9	25.2
9	55.4	54.7	54	54	69.1	48.7	100	52.6	32.4	15.4	25.6
10	63.4	63	58.8	72.6	58.1	54.8	52.6	100	33.9	13.2	27.3
11	31.9	32.9	31.9	32.8	34.9	31.1	32.4	33.9	100	15.9	36.3
12	13.5	13.5	15.9	13.5	17.5	13.9	15.4	13.2	15.9	100	12.8
13	27	26.8	26.6	29.2	27.6	25.2	25.6	27.3	36.3	12.8	100

*Los números en la fila superior y la columna izquierda corresponden a las SEQ ID NOS de las secuencias alineadas de Figura 3.

[114] Normalmente las búsquedas de secuencias se llevan a cabo empleando el programa BLASTN cuando se evalúa una secuencia de ácido nucleico dada, relativa a secuencias de ácido nucleico en el GenBank DNA Sequences y otras bases de datos públicas. Se prefiere el programa BLASTX para buscar secuencias de ácido nucleico que se han traducido en todos los marcos de lectura contra secuencias de aminoácidos en el GenBank Protein Sequences y otras bases de datos públicas. Tanto BLASTN como BLASTX se ejecutan usando parámetros predeterminados de una penalización de espacio abierta de 11.0, y una penalización de espacio extendida de 1.0, y utilizan la matriz BLOSUM-62. (Véase, v.g., Altschul, y col., 1997.)

V. Expresión de Variantes CBH2 Recombinantes

[115] Los métodos de la divulgación se basan en el uso de células para expresar

la variante CBH2, no requiriéndose un método particular de expresión de CBH2. La variante CBH2 se secreta preferiblemente de las células. La divulgación provee células hospederas que se han transducido, transformado o transfectado con un factor de expresión que comprende por una secuencia de ácido nucleico que codifica la variante CBH2. Las condiciones de cultivo, tales como temperatura, pH y similares, son las previamente utilizadas para la célula hospedera progenitora antes de la transducción, transformación o transfección y serán evidentes para los expertos en el arte.

[116] En un método, una célula fúngica filamentosa o una célula de levadura se transfecta con un vector de expresión que tiene un promotor o un fragmento de promotor biológicamente activo de uno o más (v.g., una serie) de potenciadores que funciona en la línea de células hospederas, enlazados operacionalmente a un segmento de DNA que codifica la variante CBH2, de modo que la variante CBH2 se exprese en la línea de células.

A. Constructos de Ácido Nucleico/Vectores de Expresión

[117] Fragmentos polinucleotídicos naturales o sintéticos que codifican una variante CBH2 ("secuencias de ácido nucleico que codifican CBH2") pueden incorporarse en constructos de ácido nucleico heterólogo o vectores, capaces de introducción y replicación en una célula fúngica filamentosa o una célula de levadura. Los vectores y métodos divulgados en esta especificación son apropiados para uso en células hospederas para la expresión de una variante CBH2. Puede emplearse cualquier vector siempre y cuando sea replicable y viable en células en las cuales se introduce. Grandes números de vectores y promotores apropiados son conocidos por los expertos en el arte, y se encuentran disponibles en el comercio. La clonación y los vectores de expresión se describen igualmente en Sambrook y col., 1989, Ausubel F M y col., 1989, y Strathern y col., The Molecular Biology of the Yeast *Saccharomyces*, 1981, cada uno de los cuales se incorpora expresamente por referencia en esta especificación. Vectores de expresión apropiados para hongos se describen en van den Hondel, C. A. M. J. J. y col., (1991) In: Bennett, J. W. y Lasure, L. L. (eds.) *More Gene Manipulations in Fungi*. Academic Press, páginas 396-428. La secuencia de DNA apropiada puede insertarse en un plásmido o vector (denominados colectivamente en esta especificación "vectores") por una variedad de procedimientos. En general, la secuencia de DNA se inserta en un sitio(s) apropiado(s) de endonucleasa de restricción por procedimientos estándar. Los citados procedimientos y procedimientos relacionados de subclonación se consideran dentro del campo de

acción del conocimiento de los expertos en el arte.

[118] Hongos filamentosos recombinantes que incluyen la secuencia codificante para la variante CBH2 pueden producirse introduciendo un constructo de ácido nucleico heterólogo que comprende la secuencia que codifica la variante CBH2 en células de una cepa seleccionada de los hongos filamentosos.

[119] Una vez obtenida la forma deseada de una secuencia de ácido nucleico del cbh2 variante, ésta puede modificarse en una variedad de modos. Donde la secuencia involucra regiones flanqueantes no codificantes, las regiones flanqueantes pueden someterse a resección, mutagénesis, etc. Por lo tanto, pueden realizarse transiciones, transversiones, eliminaciones e inserciones en la secuencia de origen natural.

[120] Una secuencia seleccionada que codifique un cbh2 variante puede insertarse en un vector apropiado de acuerdo con técnicas recombinantes bien conocidas y usarse para transformar hongos filamentosos capaces de expresión de CBH2. Debido a la degeneración inherente del código genético, otras secuencias de ácido nucleico, las cuales codifican sustancialmente la misma secuencia de aminoácidos o una secuencia funcionalmente equivalente, pueden usarse para clonar y expresar la variante CBH2. Por consiguiente, se observa que tales sustituciones en la región codificante caen dentro de las variantes de secuencias cubiertas por la presente divulgación. Cualquiera y todas estas variantes de secuencias pueden utilizarse en la misma manera como se describe en esta especificación para una secuencia de ácido nucleico que codifica una CBH2 progenitora.

[121] La presente divulgación incluye asimismo constructos de ácido nucleico recombinante constituidos por una o más de las secuencias de ácido nucleico que codifican la CBH2 variante, como ya se describió. Los constructos contienen un vector, tal como un plásmido o vector vírico, en el cual se ha insertado una secuencia de la divulgación, en una orientación directa o inversa.

[122] Los constructos de ácido nucleico heterólogo pueden incluir la secuencia codificante para el cbh2 variante (i) en aislamiento; (ii) en combinación con secuencias codificantes adicionales; tales como secuencias codificantes de proteína de fusión o péptido señal, donde la secuencia que codifica el cbh2 es la secuencia codificante dominante; (iii) en combinación con secuencias no codificantes, tales como intrones y elementos de control, tales como elementos

promotores y terminadores o regiones no traducidas 5' y/o 3', efectivos para expresión de la secuencia codificante en un hospedero apropiado; y/o (iv) en un entorno vector u hospedero en el cual la secuencia que codifica el *cbh2* es un gen heterólogo.

[123] En un aspecto de la presente divulgación, se emplea un constructo de ácido nucleico heterólogo para transferir una secuencia de ácido nucleico, que codifica la CBH2 variante, a una célula *in vitro*, prefiriéndose las líneas fúngicas filamentosas y de levaduras. Para la producción a largo plazo de la CBH2 variante, se prefiere la expresión estable. Se sigue que, para poner en práctica la divulgación, puede usarse cualquier método efectivo para generar transformantes estables.

[124] Vectores apropiados se equipan normalmente con una secuencia de ácido nucleico que codifica un marcador seleccionable, sitios de inserción, y elementos de control apropiados, tales como secuencias promotoras y terminadoras. El vector puede comprender secuencias reguladoras, incluyendo, por ejemplo, secuencias no codificantes, tales como intrones y elementos de control, es decir, elementos promotores y terminadores o regiones no traducidas 5' y/o 3', efectivos para la expresión de la secuencia codificante en células hospederas (y/o en un entorno vector o célula hospedera en el cual una secuencia codificante de antígeno de proteína soluble modificada no se expresa normalmente), enlazados operacionalmente con la secuencia codificante. Grandes números de vectores y promotores apropiados son conocidos por los expertos en el arte, muchos de los cuales se encuentran disponibles en el comercio y/o se describe en Sambrook, y col., (supra).

[125] Promotores ejemplares incluyen promotores constitutivos y promotores inducibles, ejemplos de los cuales incluyen un promotor CMV, un promotor temprano del virus SV40, un promotor RSV, un promotor EF-1 α , un promotor que contiene el elemento sensible a la tetraciclina (TRE - *tet responsive element*) en el sistema de activación o desactivación de tetraciclina ("*tet-on or tet-off system*"), como se describe (ClonTech y BASF), el promotor de beta actina y el promotor de metalotionina que pueden regularse por aumento mediante la adición de ciertas sales de metales. Una secuencia promotora es una secuencia de DNA que es reconocida por el hongo filamentosos particular para propósitos de expresión. Está operacionalmente conectada a la secuencia de DNA que codifica un polipéptido CBH2 variante. Tal conexión incluye posicionamiento del promotor con respecto al codón de iniciación de la secuencia de DNA que codifica el polipéptido de CBH2

variante en los vectores de expresión divulgados. La secuencia promotora contiene secuencias de control de transcripción y traducción que median la expresión del polipéptido de CBH2 variante. Ejemplos incluyen los promotores de los genes que codifican la glucoamilasa, α -amilasa, o α -glucosidasa de *Aspergillus niger*, *A. awamori* o *A. oryzae*; los genes de *A. nidulans* *gpdA* o *trpC*; los genes de *Neurospora crassa* *cbh1* o *trpI*; los genes que codifican la proteinasa aspártica de *A. niger* o *Rhizomucor miehei*; los genes *cbh1*, *cbh2*, *egl1*, *egl2* de *H. jecorina* (*T. reesei*), u otros genes que codifican celulasa.

[126] La selección del marcador seleccionable apropiado dependerá de la célula hospedera, y marcadores apropiados para diferentes hospederos son bien conocidos en el arte. Genes marcadores seleccionables típicos incluyen *argB* de *A. nidulans* o *T. reesei*, *amdS* de *A. nidulans*, *pyr4* de *Neurospora crassa* o *T. reesei*, *pyrG* de *Aspergillus niger* o *A. nidulans*. Marcadores seleccionables ejemplares adicionales incluyen, pero no se limitan a *trpc*, *trp1*, *oliC31*, *niaD* o *leu2*, los cuales se incluyen en constructos de ácido nucleico heterólogo, usados para transformar una cepa mutante tal como *trp*-, *pyr*-, *leu*- y similares.

[127] Los citados marcadores seleccionables confieren a los transformantes la capacidad de utilizar un metabolito que usualmente no es metabolizado por los hongos filamentosos. Por ejemplo, el gen *amdS* de *H. jecorina*, que codifica la enzima acetamidasa, que permite a células transformantes crecer en acetamida como fuente de nitrógeno. El marcador seleccionable (v.g. *pyrG*) puede restaurar la capacidad de una cepa mutante auxotrófica de crecer en un medio mínimo selectivo o el marcador seleccionable (v.g. *olic31*) puede conferir a los transformantes la capacidad de crecer en la presencia de un fármaco inhibitorio o antibiótico.

[128] La secuencia que codifica el marcador seleccionable se clona en cualquier plásmido apropiado usando métodos generalmente empleados en el arte. Plásmidos ejemplares incluyen *pUC18*, *pBR322*, *pRAX* y *pUC100*. El plásmido *pRAX* contienen las secuencias AMAL de *A. nidulans*, que hacen posible replicarlo en *A. niger*.

[129] La práctica de la presente divulgación empleará, a menos que se indique lo contrario, técnicas convencionales de biología molecular, microbiología, DNA recombinante, e inmunología, las cuales se encuentran al alcance de la habilidad en el arte. Tales técnicas se explican completamente en la literatura. Véase, por ejemplo, Sambrook y col., 1989; Freshney, *Animal Cell Culture*, 1987; Ausubel, y

col., 1993; y Coligan y col., Current Protocols in Immunology, 1991.

B. Células Hospederas Y Condiciones de Cultivo para la Producción de CBH2

(i) Hongos Filamentosos

[130] Por lo tanto, la presente divulgación provee hongos filamentosos que comprenden células que se han modificado, seleccionado y cultivado en una manera efectiva para resultar en la producción de CBH2 variante o expresión relativa a los hongos progenitores no transformados correspondientes.

[131] Ejemplos de especies de hongos filamentosos progenitores que pueden tratarse y/o modificarse para expresión de CBH2 variante incluyen, pero no se limitan a *Trichoderma*, v.g., *Trichoderma reesei*, *Trichoderma longibrachiatum*, *Trichoderma viride*, *Trichoderma koningii*; *Penicillium* sp., *Humicola* sp., incluyendo *Humicola insolens*; *Aspergillus* sp., *Chrysosporium* sp., *Fusarium* sp., *Hypocrea* sp., y *Emericella* sp.

[132] Células que expresan CBH2 se cultivan bajo condiciones empleadas normalmente para cultivar la línea fúngica progenitora. Generalmente, las células se cultivan en medio estándar que contienen sales fisiológicas y nutrientes, tal como se describe en Pourquie, J. y col., Biochemistry and Genetics of Cellulose Degradation, eds. Aubert, J. P. y col., Academic Press, páginas 71-86, 1988 e Ilmen, M. y col., Appl. Environ. Microbiol. 63:1298-1306, 1997. Las condiciones de cultivo son también estándar, v.g., los cultivos se incuban a 28° C en cultivos en frascos de agitación o fermentadores, hasta que se consiguen niveles deseados de expresión de CBH2.

[133] Condiciones de cultivo preferidas para un hongo filamentoso dado pueden encontrarse en la literatura científica y/o de la fuente de los hongos tales como la American Type Culture Collection (ATCC; www.atcc.org/). Una vez se ha establecido el crecimiento fúngico, las células se exponen a condiciones efectivas para provocar o permitir la expresión de CBH2 variante.

[134] En casos en los que una secuencia codificante de CBH2 está bajo control de un promotor inducible, el agente inductor, v.g., un azúcar, una sal de metal o antibiótico, se agrega al medio en una concentración efectiva para inducir la expresión de CBH2.

[135] En una incorporación, la cepa comprende *Aspergillus niger*, la cual es una cepa útil para obtener una proteína sobreexpresada. Por ejemplo dgr246 de *A. niger var awamori* es conocida por secretar cantidades elevadas de celulasas secretadas (Goedegebuur y col., Curr. Genet (2002) 41: 89-98). Son conocidas otras cepas de *Aspergillus niger var awamori* tales como GCDAP3, GCDAP4 y GAP3-4 (Ward y col., Appl. Microbiol. Biotechnol. 39:738-743, 1993).

[136] En otra incorporación, la cepa consta de *Trichoderma reesei*, la cual es una cepa útil para obtener proteína sobreexpresada. Por ejemplo, RL-P37, descrita por Sheir-Neiss, y col., Appl. Microbiol. Biotechnol. 20:46-53 (1984), es conocida por secretar cantidades elevadas de enzimas de celulasa. Equivalentes funcionales de RL-P37 incluyen la cepa RUT-C30 de *Trichoderma reesei* (ATCC No. 56765) y la cepa QM9414 (ATCC No. 26921). Se contempla que estas cepas serán igualmente útiles en sobreexpresar la CBH2 variante.

[137] Donde se desea obtener la CBH2 variante en la ausencia de actividad celulolítica natural potencialmente perjudicial, es útil obtener una cepa de células hospederas de *Trichoderma* que haya tenido eliminados uno o más genes de celulasa antes de la introducción del constructo o plásmido de DNA que contiene el fragmento de DNA que codifica la CBH2 variante. Tales cepas pueden prepararse por el método divulgado en la Patente de los Estados Unidos No. 5,246,853 y WO 92/06209, divulgaciones que se incorporan por este medio por referencia en su totalidad. Expresando una celulasa CBH2 variante en un microorganismo hospedero que carece de uno o más genes de celulasa, se simplifican los procedimientos de identificación y purificación subsiguiente. Puede eliminarse cualquier gen de *Trichoderma sp.* que se haya clonado, por ejemplo, los genes *cbh1*, *cbh2*, *egl1*, y *egl2*, así como aquellos que codifican la proteína EG III y/o EGV (véase v.g., la Patente de los Estados Unidos No. 5,475,101 y WO 94/28117, respectivamente).

[138] La eliminación de los genes puede llevarse a cabo insertando una forma del gen deseado, que se va a eliminar o alterar, en un plásmido por métodos conocidos en el arte. El plásmido de eliminación se corta entonces en un sitio(s) apropiado(s) de enzimas de restricción, interno(s) a la región codificante del gen deseado, y la secuencia que codifica el gen o parte del mismo se reemplaza con un marcador seleccionable. Secuencias de DNA flanqueantes del locus del gen que se va a eliminar o alterar, preferiblemente entre alrededor de 0.5 a 2.0 kb, permanecen a cualquiera de los lados del gen marcador seleccionable. Un

plásmido de eliminación apropiado tendrá generalmente sitios únicos de enzimas de restricción presentes en él para permitir que el fragmento que contiene el gen eliminado, incluyendo secuencias de DNA flanqueantes, y el gen marcador seleccionable se extraigan como un elemento lineal único.

[139] Un marcador seleccionable tiene que elegirse de modo que permita la detección del microorganismo transformado. Será apropiado cualquier gen marcador seleccionable que se exprese en el microorganismo seleccionado. Por ejemplo, con *Aspergillus sp.*, el marcador seleccionable se elige de modo que la presencia del marcador seleccionable en los transformantes no afecte significativamente las propiedades de los mismos. Tal marcador seleccionable puede ser un gen que codifica un producto analizable. Por ejemplo, puede usarse una copia funcional de un gen de *Aspergillus sp.* que, si falta en la cepa hospedera, resulta en la cepa hospedera que despliega un fenotipo auxotrófico. Análogamente, existen marcadores seleccionables para *Trichoderma sp.*

[140] En una incorporación, una cepa derivada pyrG de *Aspergillus sp.* se transforma con un gen pyrG funcional, el cual provee así un marcador seleccionable para transformación. Una cepa derivada pyrG puede obtenerse por selección de cepas de *Aspergillus sp.*, que son resistentes a ácido fluoroorótico (FOA - *fluoroorotic acid*). El gen pyrG codifica orotidina-5'-monofosfato decarboxilasa, una enzima requerida para la biosíntesis de uridina. Cepas con un gen pyrG intacto crecen en un medio que carece de uridina, pero son sensibles a ácido fluoroorótico. Es posible seleccionar cepas derivadas pyrG que carecen de una enzima orotidina monofosfato decarboxilasa funcional y requieren uridina para crecimiento seleccionando con respecto a resistencia FOA. Usando la técnica de selección FOA es también posible obtener cepas que requieren uridina, las cuales carecen de una orotato pirofosforribosil transferasa funcional. Es posible igualmente transformar estas células con una copia funcional del gen que codifica esta enzima (Berges & Barreau, Curr. Genet. 19:359-365 (1991), y van Hartingsveldt y col., (1986) Development of a homologous transformation system for *Aspergillus niger* based on the pyrG gene. Mol. Gen. Genet. 206:71-75). La selección de cepas derivadas se lleva a cabo fácilmente usando la técnica de resistencia FOA ya mencionada y, por consiguiente, el gen pyrG se emplea preferiblemente como un marcador seleccionable.

[141] En una segunda incorporación, una cepa derivada pyr4 de *Hyprocrea sp.* (*Hyprocrea sp.* (*Trichoderma sp.*)) se transforma con un gen pyr4 funcional, el cual provee así un marcador seleccionable para transformación. Una cepa derivada

pyr4 puede obtenerse por selección de cepas de *Hyprocrea sp.* (*Trichoderma sp.*), que son resistentes a ácido fluoroorótico (FOA). El gen pyr4 codifica orotidina-5'-monofosfato decarboxilasa, una enzima requerida para la biosíntesis de uridina. Cepas con un gen pyr4 intacto crecen en un medio que carece de uridina pero son sensibles a ácido fluoroorótico. Es posible seleccionar cepas derivadas pyr4, que carecen de una enzima orotidina monofosfato decarboxilasa funcional y requieren uridina para crecimiento, seleccionando con respecto a resistencia FOA. Usando la técnica de selección FOA es igualmente posible obtener cepas que requieren uridina, las cuales carecen de una orotato pirofosforribosil transferasa funcional. Es posible transformar estas células con una copia funcional del gen que codifica esta enzima (Berges & Barreau, 1991). La selección de cepas derivadas se realiza fácilmente usando la técnica de resistencia FOA ya mencionada y, por consiguiente, el gen pyr4 se emplea preferiblemente como un marcador seleccionable.

[142] Para transformar pyrG - *Aspergillus sp.* o pyr4 - *Hyprocrea sp.* (*Trichoderma sp.*) de modo que carezcan de la capacidad de expresar uno o más genes de celulasa, un fragmento único de DNA que comprende un gen de celulasa alterado o eliminado se aísla entonces del plásmido de eliminación y se usa para transformar el hospedero apropiado pyr-*Aspergillus* o pyr-*Trichoderma*. Los transformantes se identifican entonces y se seleccionan basándose en su capacidad de expresar el producto génico pyrG o pyr4, respectivamente y, por consiguiente, satisfacer la auxotrofia de uridina de la cepa hospedera. Análisis Southern blot se lleva entonces a cabo en los transformantes resultantes para identificar y confirmar un caso de integración de cruce doble, que reemplaza parte o la totalidad de la región codificante de la copia genómica del gen que se va a eliminar con los marcadores seleccionables pyr apropiados.

[143] Aunque los vectores plasmídicos específicos ya descritos se refieren a la preparación de transformantes pyr, la presente divulgación no se limita a estos vectores. Varios genes pueden eliminarse y reemplazarse en la cepa *Aspergillus sp.* o *Hyprocrea sp.* (*Trichoderma sp.*) utilizando las técnicas anteriores. Adicionalmente, pueden usarse cualesquiera marcadores seleccionables, como ya se discutió. De hecho, cualquier hospedero, v.g., el gen *Aspergillus sp.* o *Hyprocrea sp.*, que se ha clonado e identificado así, puede eliminarse del genoma empleando la estrategia ya descrita.

[144] Como se especifica anteriormente, las cepas hospederas utilizadas pueden ser los derivados de *Hyprocrea sp.* (*Trichoderma sp.*), que carecen o tienen un

gen o genes no funcionales, correspondientes al marcador seleccionable elegido. Por ejemplo, si el marcador seleccionable de *pyrG* se escoge para *Aspergillus sp.*, entonces se usa una cepa específica derivada *pyrG* como receptor en el procedimiento de transformación. Asimismo, por ejemplo, si el marcador seleccionable de *pyr4* se escoge para una *Hyprocrea sp.*, entonces la cepa específica derivada *pyr4* se usa en un receptor en el procedimiento de transformación. Análogamente, pueden emplearse marcadores seleccionables que comprenden genes de *Hyprocrea sp.* (*Trichoderma sp.*), equivalentes a los genes *amdS*, *argB*, *trpC*, *niaD* de *Aspergillus nidulans*. Por lo tanto, la correspondiente cepa receptora tiene que ser una cepa derivada tal como *argB*-, *trpC*-, *niaD*-, respectivamente.

[145] DNA que codifica la CBH2 variante se prepara entonces para inserción en un microorganismo apropiado. De acuerdo con la presente divulgación, DNA que codifica una CBH2 variante incluye el DNA necesario para codificar una proteína que tiene actividad celulolítica funcional. El fragmento de DNA que codifica la CBH2 variante puede fijarse funcionalmente a una secuencia promotora fúngica, por ejemplo, el promotor del gen *glaA* en *Aspergillus* o el promotor de los genes *cbh1* o *egl1* en *Trichoderma*.

[146] Se contempla asimismo que más de una copia del DNA que codifica una CBH2 variante pueda recombinarse en la cepa para facilitar la sobreexpresión. El DNA que codifica la CBH2 variante puede prepararse por la construcción de un vector de expresión que transporta el DNA que codifica la variante. El vector de expresión que transporta el fragmento de DNA insertado que codifica la CBH2 variante puede ser cualquier vector que es capaz de replicarse autónomamente en un organismo hospedero dado o de integrarse en el DNA del hospedero, normalmente un plásmido. En incorporaciones preferidas se contemplan dos tipos de vectores de expresión para obtener la expresión de genes. El primero contiene secuencias de DNA en las cuales el promotor, la región codificante del gen, y la secuencia terminadora se originan todos del gen que se va a expresar. El truncado de genes puede obtenerse donde se desee, eliminando secuencias indeseadas de DNA (v.g., que codifican dominios no deseados) para dejar el dominio que se va a expresar bajo control de sus propias secuencias reguladoras transcripcionales y traduccionales. Un marcador seleccionable puede estar contenido igualmente en el vector, permitiendo la selección para integración en el hospedero de múltiples copias de las novedosas secuencias génicas.

[147] El segundo tipo de vector de expresión se pre-ensambla y contiene

secuencias requeridas para transcripción de alto nivel y un marcador seleccionable. Se contempla que la región codificante de un gen o parte del mismo pueda insertarse en este vector de expresión de propósito general, de modo que esté bajo el control transcripcional de las secuencias promotoras y terminadoras de casetes de expresión.

[148] Por ejemplo, en *Aspergillus*, pRAX es tal vector de expresión de propósito general. Genes o parte de los mismos pueden insertarse hacia 3' del promotor glaa fuerte.

[149] Por ejemplo, en *Hypocrea*, pTEX es tal vector de expresión de propósito general. Genes o parte de los mismos pueden insertarse hacia 3' del promotor cbh1 fuerte.

[150] En el vector, la secuencia de DNA que codifica la CBH2 variante de la presente divulgación debe conectarse operacionalmente a secuencias transcripcionales y traduccionales, es decir, una secuencia promotora apropiada y una secuencia señal en el marco de lectura al gen estructural. El promotor puede ser cualquier secuencia de DNA que muestre actividad transcripcional en la célula hospedera y pueda derivarse de genes que codifican proteínas homólogas o heterólogas a la célula hospedera. Un péptido señal opcional posibilita la producción extracelular de la CBH2 variante. El DNA que codifica la secuencia señal es preferiblemente aquel que se asocia naturalmente con el gen que se va a expresar; no obstante, se contempla en la presente divulgación la secuencia señal de cualquier fuente conveniente, por ejemplo una exo-celobiohidrolasa o endoglucanasa de *Trichoderma*.

[151] Los procedimientos usados para ligar las secuencias de DNA que codifican la CBH2 variante de la presente divulgación con el promotor, y la inserción en vectores apropiados son bien conocidos en el arte.

[152] El vector o constructo de DNA ya descritos pueden introducirse en la célula hospedera de acuerdo con técnicas conocidas, tales como transformación, transfección, microinyección, microporación, bombardeo biolístico y similares.

[153] En la técnica de transformación preferida, hay que tener en cuenta que la permeabilidad de la pared celular a DNA en *Hypocrea sp.* (*Trichoderma sp.*) es muy baja. Por consiguiente, la absorción de la secuencia deseada de DNA, gen o fragmento génico es a lo mejor mínima. Existen muchos métodos de aumentar la

permeabilidad de la pared celular de *Hyprocrea sp.* (*Trichoderma sp.*) en la cepa derivada (es decir, que carece de un gen funcional correspondiente al marcador seleccionable usado) antes del proceso de transformación.

[154] El método preferido en la presente divulgación para preparar *Aspergillus sp.* o *Hyprocrea sp.* (*Trichoderma sp.*) para transformación involucra la preparación de protoplastos del micelio fúngico. Véase Campbell y col., Improved transformation efficiency of *A. niger* using homologous *niaD* gene for nitrate reductase. Curr. Genet. 16:53-56; 1989. El micelio puede obtenerse de esporas vegetativas germinadas. El micelio se trata con una enzima que digiere la pared celular resultando en protoplastos. A continuación los protoplastos se protegen por la presencia de un estabilizante osmótico en el medio de suspensión. Estos estabilizantes incluyen sorbitol, manitol, cloruro de potasio, sulfato de magnesio y similares. Usualmente la concentración de estos estabilizantes varía entre 0.8 M y 1.2 M. Es preferible emplear alrededor de una solución 1.2 M de sorbitol en el medio de suspensión.

[155] La absorción del DNA en la cepa hospedera, (*Aspergillus sp.* o *Hyprocrea sp.* (*Trichoderma sp.*), depende de la concentración de los iones de calcio. Generalmente se usan entre alrededor de 10 mM CaCl_2 y 50 mM CaCl_2 en una solución de absorción. Además de la necesidad del ión de calcio en la solución de absorción, otros ítems generalmente incluidos son un sistema amortiguador tal como amortiguador TE (10 Mm Tris, pH 7.4; 1 mM EDTA) o amortiguador 10 mM MOPS, pH 6.0 (ácido morfolinopropanosulfónico) y polietilén glicol (PEG). Se cree que el polietilén glicol actúa para fusionar las membranas celulares permitiendo así que el contenido del medio se transfiera al citoplasma de la célula hospedera, a modo de ejemplo la cepa *Aspergillus sp.* o *Hyprocrea sp.*, y DNA plasmídico se transfiera al núcleo. Frecuentemente, esta fusión deja múltiples copias del DNA del plásmido integradas en el cromosoma hospedero.

[156] Usualmente una suspensión que contiene los protoplastos de *Aspergillus sp.* o células que se han sometido a un tratamiento de permeabilidad con una densidad de 10^5 a $10^6/\text{mL}$, preferiblemente $2 \times 10^5/\text{mL}$ se emplea en la transformación. Análogamente, una suspensión que contiene los protoplastos de *Hyprocrea sp.* (*Trichoderma sp.*) o células que se han sometido a un tratamiento de permeabilidad con una densidad de 10^8 a $10^9/\text{mL}$, preferiblemente $2 \times 10^8/\text{mL}$ se usa en la transformación. Un volumen de 100 μL de estos protoplastos o células en una solución apropiada (v.g., 1.2 M sorbitol; 50 mM CaCl_2) se mezclan con el DNA deseado. Generalmente se agrega una alta concentración de PEG a la

solución de absorción. De 0.1 a 1 volumen de 25% PEG 4000 puede agregarse a la suspensión de protoplastos. Sin embargo, es preferible agregar alrededor de 0.25 volúmenes a la suspensión de protoplastos. Aditivos tales como sulfóxido de dimetilo, heparina, espermidina, cloruro de potasio y similares pueden agregarse asimismo a la solución de absorción y ayudar en la transformación.

[157] Generalmente, la mezcla se incuba luego a aproximadamente 0°C durante un período de entre 10 a 30 minutos. A continuación se agrega PEG adicional a la mezcla para potenciar aún más la absorción del gen deseado o secuencia de DNA. El 25% PEG 4000 se agrega generalmente en volúmenes de 5 a 15 veces el volumen de la mezcla de transformación; no obstante, pueden ser apropiados volúmenes mayores y menores. El 25% PEG 4000 es preferiblemente alrededor de 10 veces el volumen de la mezcla de transformación. Una vez agregado el PEG, la mezcla de transformación se incuba entonces a temperatura ambiente o sobre hielo antes de la adición de una solución de sorbitol y CaCl_2 . La suspensión de protoplastos se agrega entonces a alícuotas líquidas de un medio de crecimiento. Este medio permite el crecimiento de transformantes únicamente. Puede usarse cualquier medio de crecimiento en la presente divulgación que sea apropiado para desarrollar los transformantes deseados. No obstante, si se están seleccionando transformantes Pyr^+ , es preferible usar un medio de crecimiento que no contiene uridina. Las colonias subsiguientes se transfieren y purifican en un medio de crecimiento agotado de uridina.

[158] En esta fase, transformantes estables pueden distinguirse de transformantes inestables por su tasa de crecimiento más rápida y, en *Trichoderma*, por ejemplo, la formación de colonias circulares con un perfil suavizado, más que irregular, en medio de cultivo sólido que carece de uridina. Adicionalmente, en algunos casos puede hacerse una prueba adicional de estabilidad desarrollando los transformantes en medio sólido no selectivo (es decir, que contiene uridina), recolectando las esporas de este medio de cultivo y determinando el porcentaje de estas esporas que germinarán posteriormente y crecerán en medio selectivo que carece de uridina.

[159] En una incorporación particular del método ya mencionado, la(s) CBH2(s) variante(s) se recuperan en forma activa de la célula hospedera luego del crecimiento en medio líquido como resultado del procesamiento postraducciona apropiado de la CBH2 variante.

(ii) Levadura

[160] La presente divulgación contempla igualmente el uso de levadura como célula hospedera para la producción de CBH2. Varios otros genes que codifican enzimas hidrolíticas se han expresado en varias cepas de la levadura *S. cerevisiae*. Éstos incluyen secuencias que codifican dos endoglucanasas (Penttilä y col., *Yeast* vol. 3, páginas 175-185, 1987), dos celobiohidrolasas (Penttilä y col., *Gene*, 63: 103-112, 1988) y una beta-glucosidasa de *Trichoderma reesei* (Cummings y Fowler, *Curr. Genet.* 29:227-233, 1996), una xilanasa de *Aureobasidium pullulans* (Li y Ljungdahl, *Appl. Environ. Microbiol.* 62, no. 1, páginas 209-213, 1996), una alpha-amilasa del trigo (Rothstein y col., *Gene* 55:353-356, 1987), etc. Además, un casete del gen de celulasa que codifica la *Butyrivibrio fibrisolvens* endo-[beta]-1,4-glucanasa (END1), la *Phanerochaete chrysosporium* celobiohidrolasa (CBH1), la *Ruminococcus flavefaciens* celodextrinasa (CEL1) y la *Endomyces fibuliger* celobiasa (BgII) se expresó exitosamente en una cepa de laboratorio de *S. cerevisiae* (Van Rensburg y col., *Yeast*, vol. 14, páginas 67-76, 1998).

C. Introducción de una Secuencia de Ácido Nucleico que Codifica una CBH2 en Células Hospederas

[161] La divulgación provee además células y composiciones de células que se han modificado genéticamente para incluir una secuencia de ácido nucleico que codifica una CBH2 variante provista exógenamente. Una célula progenitora o línea celular puede modificarse genéticamente (es decir, transducirse, transformarse o transfectarse) con un vector de clonación o un vector de expresión. El vector puede estar, por ejemplo, en la forma de un plásmido, una partícula vírica, un bacteriófago, etc, como se describe con más detalle anteriormente.

[162] Los métodos de transformación de la presente divulgación pueden resultar en la integración estable de la totalidad o parte del vector de transformación en el genoma del hongo filamentoso. Sin embargo, se contempla también la transformación resultante en el mantenimiento de un vector de transformación extra-cromosómico auto-replicante.

[163] Muchos métodos de transfección estándar pueden usarse para producir líneas celulares de *Trichoderma reesei* que expresan grandes cantidades de la proteína heteróloga. Algunos de los métodos publicados para la introducción de constructos de DNA en cepas productoras de celulasa de *Trichoderma* incluyen Lorito, Hayes, DiPietro y Harman, 1993, *Curr. Genet.* 24: 349-356; Goldman,

VanMontagu y Herrera-Estrella, 1990, Curr. Genet. 17:169-174; Penttilä, Nevalainen, Ratto, Salminen y Knowles, 1987, Gene 6: 155-164, para *Aspergillus* Yelton, Hamer y Timberlake, 1984, Proc. Natl. Acad. Sci. USA 81: 1470-1474, para *Fusarium* Bajar, Podila y Kolattukudy, 1991, Proc. Natl. Acad. Sci. USA 88: 8202-8212, para *Streptomyces* Hopwood y col., 1985, The John Innes Foundation, Norwich, UK y para *Bacillus Brigidi*, DeRossi, Bertarini, Riccardi y Matteuzzi, 1990, FEMS Microbiol. Lett. 55: 135-138).

[164] Otros métodos para introducir un constructo de ácido nucleico heterólogo (vector de expresión) en hongos filamentosos (v.g., *H. jecorina*) incluyen, pero no se limitan a, el uso de una pistola de partículas o genes, permeabilización de las paredes celulares de hongos filamentosos antes del proceso de transformación (v.g., mediante el empleo de elevadas concentraciones de álcali, v.g., 0.05 M a 0.4 M CaCl_2 o acetato de litio), fusión de protoplastos o transformación mediada por *Agrobacterium*. Un método ejemplar para transformación de hongos filamentosos por tratamiento de protoplastos o esferoplastos con polietilén glicol y CaCl_2 se describe en Campbell, E. I. y col., Curr. Genet. 16:53-56, 1989 y Penttilä, M. y col., Gene, 63:11-22, 1988.

[165] Puede emplearse cualquiera de los procedimientos bien conocidos para introducir secuencias nucleotídicas extrañas en células hospederas. Éstos incluyen el uso de transfección con fosfato de calcio, polibreno, fusión de protoplastos, electroporación, biolística, liposomas, microinyección, vectores de plasma, vectores víricos y cualquiera de los demás métodos bien conocidos para introducir DNA genómico clonado, cDNA, DNA sintético u otro material genético extraño en una célula hospedera (véase, v.g., Sambrook y col., supra). También útil es el método de transfección mediado por *Agrobacterium*, descrito en la Patente de los Estados Unidos No. 6,255,115. Sólo es necesario que el procedimiento de ingeniería genética particular utilizado pueda introducir exitosamente por lo menos un gen en la célula hospedera, capaz de expresar el gen heterólogo.

[166] Además, constructos de ácido nucleico heterólogo que contienen una secuencia de ácido nucleico que codifica una CBH2 variante pueden transcribirse *in vitro*, y el RNA resultante introducirse en la célula hospedera por métodos bien conocidos, v.g., mediante inyección.

[167] La divulgación incluye asimismo transformantes novedosos y útiles de hongos filamentosos tales como *H. jecorina* y *A. niger* para uso en producir

composiciones de celulasas fúngicas. La divulgación incluye transformantes de hongos filamentosos, especialmente hongos que comprenden la secuencia que codifica la CBH2 variante, o eliminación de la secuencia que codifica el cbh endógeno.

[168] Luego de la introducción de un constructo de ácido nucleico heterólogo que comprende la secuencia codificante para un cbh2 variante, las células genéticamente modificadas pueden cultivarse en medio nutriente convencional modificado según sea apropiado para activar promotores, seleccionar transformantes o amplificar la expresión de una secuencia de ácido nucleico que codifica una CBH2 variante. Las condiciones de cultivo, tales como temperatura, pH y similares, son aquellas previamente usadas para la célula hospedera seleccionada para expresión, y serán evidentes a los expertos en el arte.

[169] Generalmente se considera que la progenie de células, en las cuales se han introducido tales constructos de ácido nucleico heterólogo, comprende la secuencia de ácido nucleico que codifica la CBH2 variante hallada en el constructo de ácido nucleico heterólogo.

[170] La divulgación incluye además transformantes útiles y novedosos de hongos filamentosos tales como *H. jecorina* para uso en producir composiciones de celulasas fúngicas. De igual forma *Aspergillus niger* puede usarse en producir la CBH2 variante. La divulgación incluye transformantes de hongos filamentosos, especialmente hongos que comprenden la secuencia que codifica el cbh2 variante, o eliminación de la secuencia que codifica el cbh2 endógeno.

[171] Transformantes estables de hongos filamentosos pueden distinguirse generalmente de transformantes inestables por su tasa de crecimiento más rápida y, en *Trichoderma*, por ejemplo, la formación de colonias circulares con un perfil suave, más bien que irregular, en medio de cultivo sólido. Adicionalmente, en algunos casos, puede realizarse una prueba de estabilidad desarrollando los transformantes en medio no selectivo sólido, recolectando las esporas de este medio de cultivo y determinando el porcentaje de estas esporas que germinarán posteriormente y crecerán sobre el medio selectivo.

VI. Aislamiento y Purificación de Proteína CBH2 Recombinante

[172] En general, una proteína CBH2 variante producida en cultivo celular se secreta en el medio y puede purificarse o aislarse, v.g., eliminando componentes

indeseados del medio de cultivo celular. Sin embargo, en algunos casos, una proteína CBH2 variante puede producirse en una forma celular que requiera recuperación de un lisado de células. En tales casos la proteína CBH2 variante se purifica de las células en las cuales se produjo usando técnicas empleadas rutinariamente por los expertos en el arte. Los ejemplos incluyen, pero no se limitan a, cromatografía de afinidad (Tilbeurgh y col., FEBS Lett. 16:215, 1984), métodos cromatográficos de intercambio iónico (Goyal y col., Bioresource Technol. 36:37-50, 1991; Fliess y col., Eur. J. Appl. Microbiol. Biotechnol. 17:314-318, 1983; Bhikhabhai y col., J. Appl. Biochem. 6:336-345, 1984; Ellouz y col., J. Chromatography 396:307-317, 1987), incluyendo usar materiales de intercambio iónico con alto poder de resolución (Medve y col., J. Chromatography A 808:153-165, 1998), cromatografía de interacción hidrofóbica (Tomaz y Queiroz, J. Chromatography A 865:123-128, 1999), y partición de dos fases (Brumbauer, y col., Bioseparation 7:287-295, 1999).

[173] Normalmente, la proteína CBH2 variante se fracciona para segregar proteínas que tengan propiedades seleccionadas, tales como afinidad de unión a agentes de unión particulares, v.g., anticuerpos o receptores; o que tengan un intervalo de peso molecular seleccionado, o intervalo de puntos isoeléctricos.

[174] Una vez se ha logrado la expresión de una proteína CBH2 variante dada, la proteína CBH2 así producida se purifica de las células o del cultivo celular. Procedimientos ejemplares apropiados para la citada purificación incluyen los siguientes: cromatografía en columna de afinidad de anticuerpo, cromatografía de intercambio iónico; precipitación en etanol; HPLC de fase inversa; cromatografía en sílice o en una resina de intercambio catiónico tal como DEAE; cromatografía de exclusión; SDS-PAGE; precipitación en sulfato de amonio; y filtración en gel usando, v.g., Sephadex G-75. Pueden emplearse varios métodos de purificación de proteínas, y tales métodos son conocidos en el arte y se describen v.g. en Deutscher, *Methods in Enzymology*, vol. 182, No. 57, páginas. 779, 1990; Scopes, *Methods Enzymol.* 90: 479-91, 1982. Los pasos de purificación seleccionados dependerán, v.g., de la naturaleza del proceso de producción utilizado y la proteína particular producida.

VII. Utilidad de cbh2 y CBH2

[175] Puede observarse que los ácidos nucleicos de cbh variante, la proteína CBH2 variante y composiciones que incluyen actividad de proteína CBH2 variante tienen utilidad en una amplia variedad de aplicaciones, algunas de las cuales se

describen a continuación.

[176] Composiciones de celulasa novedosas y mejoradas que incluyen cantidades variables de celulasa tipo BG, tipo EG y tipo CBH variante tienen utilidad en composiciones detergentes que exhiben una capacidad limpiadora mejorada, funcionan como agente suavizante y/o mejoran el tacto de telas de algodón (v.g., "lavado a la piedra" o "biodesfibrilación"), en composiciones para degradar pulpa de madera en azúcares (v.g., para producción de bio-etanol), y/o en composiciones de forrajes. El aislamiento y caracterización de celulasa de cada tipo ofrece la posibilidad de controlar los aspectos de tales composiciones.

[177] En un enfoque, la celulasa de la divulgación tiene utilidad en composiciones detergentes o en el tratamiento de telas para mejorar el tacto y la apariencia.

[178] Puesto que la tasa de hidrólisis de productos celulósicos puede aumentarse usando un transformante, que tiene por lo menos una copia adicional del gen *cbh* insertado en el genoma, productos que contienen celulosa o heteroglucanos pueden degradarse a una tasa más rápida y en una mayor medida. Productos hechos de celulosa tales como papel, algodón, pañales celulósicos y similares pueden degradarse más eficientemente en un relleno sanitario. Por consiguiente, el producto de fermentación obtenible de los transformantes o los transformantes solos pueden usarse en composiciones para ayudar a degradar por licuefacción una variedad de productos de celulosa que se agregan a los rellenos sanitarios atestados.

[179] La sacarificación y fermentación separada es un proceso mediante el cual celulosa presente en biomasa, v.g., rastrojo de maíz, se convierte en glucosa y, posteriormente, cepas de levadura convierten glucosa en etanol. La sacarificación y fermentación simultánea es un proceso por el cual celulosa presente en biomasa, v.g., rastrojo de maíz, se convierte en glucosa y, al mismo tiempo y en el mismo reactor, cepas de levadura convierten glucosa en etanol. Por consiguiente, en otro enfoque, la celulasa tipo CBH2 variante de la divulgación tiene utilidad en la degradación de biomasa en etanol. La producción de etanol de fuentes fácilmente disponibles de celulosa ofrece una fuente estable de combustibles renovables.

[180] Materias primas a base de celulosa están constituidas por desechos agrícolas, pastos y maderas y demás biomasa de bajo valor, tales como desechos municipales (v.g., papel reciclado, residuos vegetales, etc.). Etanol puede

producirse de la fermentación de cualquiera de esas materias primas celulósicas. No obstante, la celulosa tiene que convertirse primero en azúcares antes de que pueda haber conversión en etanol.

[181] Una gran variedad de materias primas puede usarse con la CBH variante inventiva, y la seleccionada para uso puede depender de la región en donde se está haciendo la conversión. Por ejemplo, en el Medio Oeste de los Estados Unidos pueden predominar desechos agrícolas tales como paja de trigo, rastrojo de maíz y bagazo, mientras que en California puede predominar la paja de arroz. No obstante, debe entenderse que cualquier biomasa celulósica disponible puede usarse en cualquier región.

[182] Los métodos de la presente divulgación pueden emplearse en la producción de monosacáridos, disacáridos, y polisacáridos como materias primas químicas o de fermentación con microorganismos para la producción de productos orgánicos, químicos y combustibles, plásticos, y otros productos o intermedios. En particular, el valor de residuos de procesamiento (grano seco de destilerías, granos agotados de cervecerías, bagazo de caña de azúcar, etc.) puede aumentarse por solubilización parcial o completa de celulosa o hemicelulosa. Aparte de etanol, algunos químicos que pueden producirse de celulosa y hemicelulosa incluyen acetona, acetato, glicina, lisina, ácidos orgánicos (v.g., ácido láctico), 1,3-propanodiol, butanodiol, glicerol, etilen glicol, furfural, polihidroxialcanoatos, cis, ácido cis-mucónico, forrajes para animales y xilosa.

[183] Una composición de celulasa, que contiene una cantidad mejorada de celobiohidrolasa, tiene utilidad en la producción de etanol. El etanol de este proceso puede usarse además como un potenciador de octanaje o directamente como combustible en lugar de gasolina, lo cual es ventajoso porque el etanol como una fuente de combustible es más amigable con el medio ambiente que los productos derivados del petróleo. Se sabe que el uso de etanol mejorará la calidad del aire y reducirá posiblemente los niveles locales de ozono y smog. Más aún, la utilización de etanol en lugar de gasolina puede ser de importancia estratégica para amortiguar el impacto de los cambios súbitos en la energía no renovable y los suministros petroquímicos.

[184] Etanol puede producirse mediante procesos de sacarificación y fermentación de biomasa celulósica como árboles, plantas herbáceas, desechos sólidos municipales y residuos agrícolas y forestales. Sin embargo, la relación de enzimas de celulasa individuales con una mezcla de celulasas de origen natural,

producidas por un microbio, puede no ser la más eficiente para la rápida conversión en glucosa de la celulosa contenida en la biomasa. Se sabe que las endoglucanasas actúan para producir nuevos extremos de cadena de celulosa, los cuales son ellos mismos sustratos para la acción de celobiohidrolasas y mejoran así la eficiencia de hidrólisis del sistema completo de celulasas. Por consiguiente, el uso de una actividad de celobiohidrolasa incrementada u optimizada puede mejorar en gran medida la producción de etanol.

[185] Por lo tanto, la celobiohidrolasa inventiva encuentra uso en la hidrólisis de celulosa en sus componentes de azúcar. En una incorporación, una celobiohidrolasa variante se agrega a la biomasa antes de la adición de un organismo fermentador. En una segunda incorporación, una celobiohidrolasa variante se agrega a la biomasa al mismo tiempo que un organismo fermentador. Opcionalmente, pueden estar presentes otros componentes de celulosa en cualquiera de las dos incorporaciones.

[186] En otra incorporación, la materia prima celulósica puede pretratarse. El pretratamiento puede ser por temperatura elevada y la adición de cualquiera de ácido diluido, ácido concentrado o solución de álcali diluido. La solución de pretratamiento se agrega durante un tiempo suficiente para hidrolizar por lo menos parcialmente los componentes de hemicelulosa y luego se neutralizan.

[187] El principal producto de la acción de CBH2 sobre la celulosa es celobiosa la cual está disponible para conversión en glucosa mediante actividad de BG (por ejemplo en un producto de celulosa fúngica). Ya sea por el pretratamiento de la biomasa celulósica o por la acción enzimática sobre la biomasa, pueden obtenerse de la biomasa otros azúcares, aparte de la glucosa y la celobiosa. El contenido de hemicelulosa de la biomasa puede convertirse (por hemicelulasas) en azúcares tales como xilosa, galactosa, mannososa y arabinosa. Por consiguiente, en un proceso de conversión de biomasa, la sacarificación enzimática puede producir azúcares que se hacen disponibles para conversiones biológicas o químicas en otros intermedios o productos finales. En consecuencia, los azúcares generados de biomasa tienen uso en una variedad de procesos, aparte de la generación de etanol. Ejemplos de tales conversiones son fermentación de glucosa en etanol (como se reseña por M. E. Himmel y col., páginas 2-45, en "Fuels and Chemicals from Biomass", ACS Symposium Series 666, ed B. C. Saha y J. Woodward, 1997) y otras conversiones biológicas de glucosa en 2,5-diaceto-D-gluconato (Patente de los Estados Unidos No. 6,599,722), ácido láctico (R. Datta y S-P. Tsai páginas 224-236, *ibid*), succinato (R. R. Gokarn, M. A. Eiteman y J. Sridhar páginas 237-

263, *ibid*), 1,3-propanodiol (A-P. Zheng, H. Biebl y W-D. Deckwer páginas 264-279, *ibid*), 2,3-butanodiol (C. S. Gong, N. Cao y G. T. Tsao páginas 280-293, *ibid*), y las conversiones biológicas y químicas de xilosa en xilitol (B. C. Saha y R. J. Bothast páginas 307-319, *ibid*). Véase asimismo, por ejemplo, WO 98/21339.

[188] Las composiciones detergentes de esta divulgación pueden emplear, además de la composición de celulasa (independientemente del contenido de celobiohidrolasa, es decir, libre de celobiohidrolasa, sustancialmente libre de celobiohidrolasa, o mejorada con celobiohidrolasa), un tensioactivo, incluyendo tensioactivos aniónicos, no iónicos y anfotéricos, una hidrolasa, agentes mejoradores de detergente, agentes blanqueadores, agentes azulantes y colorantes fluorescentes, inhibidores de formación de grumos, solubilizantes, tensioactivos catiónicos y similares. Todos estos componentes son conocidos en el arte de los detergentes. Como ya se describió, la composición de celulasa puede agregarse a la composición detergente en un diluyente líquido, en gránulos, en emulsiones, en geles, en pastas, y similares. Las citadas formas son bien conocidas por el artesano experto. Cuando se emplea una composición detergente sólida, la composición de celulasa se formula preferiblemente en forma de gránulos. Preferiblemente, los gránulos pueden formularse con el fin de contener un agente protector de celulasa. Para una discusión más exhaustiva, véase la Patente de los Estados Unidos No. 6,162,782 titulada "Detergent compositions containing cellulase compositions deficient in CBH2 type components," la cual se incorpora aquí por referencia.

[189] Preferiblemente las composiciones de celulasa se emplean desde alrededor de 0.00005 por ciento en peso a alrededor de 5 por ciento en peso relativo a la composición detergente total. Más preferiblemente, las composiciones de celulasa se emplean desde alrededor de 0.0002 por ciento en peso a alrededor de 2 por ciento en peso relativo a la composición detergente total.

[190] Adicionalmente la secuencia de ácido nucleico de CBH2 variante tiene utilidad en la identificación y caracterización de secuencias de ácido nucleico relacionadas. Un número de técnicas útiles para determinar (predecir o confirmar) la función de genes relacionados o productos génicos incluyen, pero no se limitan a, (A) análisis DNA/RNA, tales como (1) sobreexpresión, expresión ectópica, y expresión en otras especies; (2) bloqueo de genes (genética inversa, bloqueo direccionado, silenciamiento génico inducido por virus (VIGS - *viral induced gene silencing*, véase Baulcombe, 100 Years of Virology, Calisher y Horzinek eds., Springer-Verlag, New York, N.Y. 15:189-201, 1999); (3) análisis del estado de

metilación del gen, especialmente regiones reguladoras flanqueantes; y (4) hibridación *in situ*; (B) análisis de productos génicos tales como (1) expresión de proteínas recombinantes; (2) producción de antisueros, (3) inmunolocalización; (4) análisis bioquímicos para actividad catalítica u otra actividad; (5) estado de fosforilación; y (6) interacción con otras proteínas mediante análisis de dos híbridos de levadura; (C) análisis de vías, tal como poner un gen o un producto génico en una vía particular bioquímica o vía de señalización, basándose en su fenotipo de sobreexpresión, o por homología de secuencias con los genes relacionados; y (D) otros análisis que pueden realizarse igualmente para determinar o confirmar la participación del gen aislado y su producto en una vía metabólica particular o de señalización, y ayudar a determinar la función del gen.

DIVULGACIÓN EXPERIMENTAL

[191] La presente divulgación se describe con mayor detalle en los siguientes Ejemplos, los cuales en ninguna forma pretenden limitar el ámbito de la divulgación, tal como se reivindica. La finalidad de las Figuras es ser parte integrante de la especificación y describir la divulgación. Los siguientes Ejemplos se dan con el fin de ilustrar, no para limitar, la divulgación reivindicada.

[192] En la siguiente divulgación experimental se aplican las siguientes abreviaturas: M (molar); mM (milimolar); μ M (micromolar); nM (nanomolar); mol (moles); mmol (milimoles); μ mol (micromoles); nmol (nanomoles); gm (gramos); mg (miligramos); μ g (microgramos); pg (picogramos); L (litros); ml y mL (mililitros); μ l y μ L (microlitros); cm (centímetros); mm (milímetros); μ m (micrómetros); nm (nanómetros); U (unidades); V (voltios); MW (peso molecular); sec (segundos); min(s) (minuto/minutos); h(s) y hr(s) (hora/horas); °C (grados Centígrados); QS (cantidad suficiente); ND (no hecho); NA (no aplicable); rpm (revoluciones por minuto); H₂O (agua); dH₂O (agua desionizada); HCl (ácido clorhídrico); aa (aminoácido); bp (par de bases); kb (par de kilobases); kD (kilodaltons); cDNA (DNA copia o complementario); DNA (ácido desoxirribonucleico); ssDNA (DNA de una sola hebra); dsDNA (DNA de doble hebra); dNTP (desoxirribonucleótido trifosfato); RNA (ácido ribonucleico); MgCl₂ (cloruro de magnesio); NaCl (cloruro de sodio); w/v (peso a volumen); v/v (volumen a volumen); g (gravedad); OD (densidad óptica); CNPG (cloro-nitro-fenil-beta-D-glucósido); CNP (2-cloro-4-nitrofenol); APB (bagazo pretratado con ácido - *acid-pretreated bagasse*); PASC (celulosa esponjada con ácido fosfórico - *phosphoric acid swollen cellulose*); PCS (rastreo de maíz pre-tratado con ácido - *acid-pretreated corn stover*); Pi o PI (índice de rendimiento - *performance index*); PAGE (electroforesis en gel de

poliacrilamida - *polyacrylamide gel electrophoresis*); PCR (reacción en cadena de la polimerasa - *polymerase chain reaction*); RT-PCR (PCR de transcripción inversa); y HPLC (cromatografía líquida de alta presión - *high pressure liquid chromatography*).

EJEMPLO 1

Modificación Química de una Preparación de Celulasa *Trichoderma sp.* y Análisis para Examinar CBH2 Variantes

[193] Este Ejemplo describe el tratamiento de una preparación comercial de celulasa *Trichoderma sp.* (complejo enzimático BG LAMINEX (Genencor Division, Danisco US, Inc.) con anhídrido succínico para acetilar los residuos lisina. La acetilación de residuos lisina del complejo enzimático BG LAMINEX altera la carga neta de las proteínas (v.g., mayor carga negativa). Otras modificaciones químicas similares pueden usarse igualmente para convertir la carga positiva de la lisina en un grupo de carga negativa (por ejemplo con tratamiento con anhídrido acetoxisuccínico, anhídrido maleico, anhídrido tartárico y anhídrido ftálico) o incluso a dos cargas negativas (por ejemplo con tratamiento con anhídrido trimelítico, anhídrido cis-aconítico y anhídrido 4-nitroftálico). Otras modificaciones químicas pueden emplearse para eliminar las cargas positivas de los residuos lisina, resultantes en un residuo no cargado (por ejemplo, con tratamiento con anhídrido acético, anhídrido butírico, anhídrido isobutírico, anhídrido hexanoico, anhídrido valérico, anhídrido isovalérico y anhídrido pivalico).

[194] Los residuos lisina en una preparación de celulasa se modificaron usando anhídrido succínico, empleando una variedad de métodos publicados (Lundblad, Chemical Reagents for Protein Modification, Editor: R. Lundblad, 3rd Edition CRC press, 1984). Para esta reacción, se preparó una muestra de 236 mg del complejo enzimático BG LAMINEX en 1 mL de amortiguador 500 mM HEPES pH 8. Una solución de anhídrido succínico (Aldrich) se preparó disolviendo el polvo en DMSO a una concentración final de 500mg/mL antes de la adición del complejo enzimático. Se agregó una alícuota de anhídrido succínico, de modo que se lograra una relación de > 1:100 lisina a ácido succínico en el tubo de reacción. Otro tubo de reacción se preparó con DMSO y enzima únicamente, usando volúmenes similares, para servir como control de proteína no modificada. Los tubos se sometieron a rotación en espiral ("**were vortexed**") y se dejaron a temperatura ambiente de un día para otro. Al día siguiente, un volumen 1:10 de 1M de glicina, con pH 3, se agregó a cada tubo para extinguir la reacción de

anhídrido succínico.

[195] Se confirmó la modificación química comparando proteínas modificadas y no modificadas en geles nativos. Se analizaron alícuotas de cada reacción (modificada y no modificada químicamente) en un gradiente 8-25 %, geles nativos corridos con pH 8.8 a 100 voltios (Phast System gels, GE Healthcare). Las proteínas se visualizaron luego de la tinción del gel con azul de Coomassie, para confirmar que la modificación había sido exitosa. La tinción reveló cambios en la migración de bandas de proteína, confirmando los cambios en carga de los diferentes componentes de proteína de la preparación de celulasa. Las muestras modificadas de la preparación de celulasa de *Trichoderma sp.* fueron más negativamente cargadas que las muestras no modificadas.

[196] Para aislar las proteínas modificadas y no modificadas (control), se desalinizaron alícuotas de 80µl de cada muestra empleando columnas de desalinización giratorias ("**spin desalt columns**") (Pierce). Se midió la absorbencia a 280nm de las muestras desalinizadas (incluyendo el control sin modificación) usando un espectrofotómetro NanoDrop™ (Thermo), en duplicados después de una dilución 1:10 de la muestra para determinar la concentración total de proteína de las muestras.

Determinaciones del Potencial Zeta

[197] Este Ejemplo describe cómo determinar el potencial zeta de una enzima y un sustrato. La presencia de una carga en la superficie de una partícula influye en la distribución de iones en la región interfacial circundante. El resultado es una concentración creciente de iones contrarios de carga opuesta a la de la partícula cerca de la superficie de la partícula. A medida que se aleja de la superficie de la partícula, la distribución heterogénea de iones se hará eventualmente homogénea. La distancia a la cual se obtiene una distribución homogénea se denomina la longitud de Debye ($1/\kappa$) o distancia de apantallamiento ("**screening distance**"), y depende de la resistencia iónica, como se muestra en la siguiente expresión, donde ϵ_0 es la permitividad de espacio libre ($8.854 \times 10^{-12} \text{ F m}^{-1}$), ϵ_r es la permitividad del líquido, k es la constante de Boltzmann ($1.38 \times 10^{-23} \text{ J K}^{-1}$), T es la temperatura en Kelvin, e es la carga electrónica ($1.6022 \times 10^{-19} \text{ C}$), I es la fuerza iónica molar, y NA es la constante de Avogadro ($6.022 \times 10^{23} \text{ mol}^{-1}$).

$$\text{Longitud de Debye } \frac{1}{\kappa} = \sqrt{\frac{\epsilon_0 \epsilon_r k T}{NA e^2 I}} \quad (0.1)$$

La fuerza iónica molar puede calcularse a partir de la siguiente ecuación, en donde C es la concentración de especie iónica y Z es la valencia.

$$I = \frac{1}{2} \sum C_i Z_i^2 \quad (0.2)$$

Para agua a 298 K, la expresión longitud de Debye se reduce a la siguiente forma.

$$\kappa^{-1} = 0.304(I^{-0.5}) \quad (0.3)$$

La capa líquida que circunda la partícula existe como dos partes; una región interna (capa de Stern) en la que los iones están unidos fuertemente y una región externa (difusa) en la que están asociados menos firmemente. En la capa difusa existe una frontera dentro de la cual los iones y partículas forman una entidad estable. Cuando una partícula se mueve, los iones en esta frontera se mueven con ella. Aquellos iones más allá de la frontera no se desplazan con la partícula. El potencial eléctrico en esta frontera, también llamado la superficie de corte hidrodinámico, se define como el potencial zeta.

[198] En la dispersión electroforética de la luz, el potencial zeta ζ se calcula a partir de la movilidad electroforética medida u usando la ecuación de Henry mostrada a continuación, en la que ϵ es la constante dieléctrica, h es la viscosidad de la solución, κ es la longitud inversa de Debye, a es el radio de la partícula, y $f(ka)$ es la función de Henry.

$$u = \frac{2e\zeta}{3h} f(ka) \quad (0.4)$$

[199] Las unidades de κ son longitud recíproca, siendo $1/\kappa$ el "espesor" de la doble capa eléctrica (longitud de Debye). El parámetro a se refiere al radio de la partícula y, por consiguiente, κa es la relación del radio de la partícula al espesor de la doble capa eléctrica. La función de Henry, $f(ka)$, depende de la forma de la partícula, pero se conoce para una esfera. En la expresión anterior, varía desde $f(0) = 1$ (límite de Hückel) a $f(\infty) = 1.5$ (límite de Smoluchowski). Para partículas pequeñas tales como proteínas en un medio dieléctrico bajo (o fuerza iónica baja), el límite de Hückel de $f(ka) = 1$ es el modelo más apropiado.

[200] Se midieron potenciales zeta de proteínas con el Zetasizer NS (Malvern Instruments, UK) de acuerdo con el principio descrito más arriba. Potenciales zeta de telas manchadas con sangre/leche/tinta (BMI - *blood/milk/ink*) se midieron con el SurPass (Anton-Paar, Austria) usando la implementación de potencial de flujo del principio anterior. De la definición de carga superficial, usualmente expresada en Culombios:

$$q_s = 4pe_o e_r a (1 + ka) z \quad (0.5)$$

Esto puede expresarse igualmente como la carga neta z multiplicada por la carga elemental e $1.6 \cdot 10^{-19} \text{C}$:

$$q_s = ze \quad (0.6)$$

Por consiguiente, el cambio esperado en el potencial zeta debido a un incremento de carga neta viene dado por:

$$\frac{Dz}{Dz} = \frac{e}{4pe_o e_r a (1 + ka)} \quad (0.7)$$

Es también posible medir potenciales zeta usando una técnica de geles nativos (Sparks y col., Journal of Lipid Research, 33:123-130, 1992), como se describe en el Ejemplo 1. La movilidad electroforética medida con geles nativos es usualmente menor que en solución debido al retardo causado por la matriz del gel. Potenciales zeta calculados de esta manera son usualmente menores, comparados con métodos a base de solución. Por consiguiente, nos referimos a ellos como potenciales zeta aparentes, cuando se obtienen mediante la técnica de geles nativos.

[201] La carga efectiva en una formulación dada se expresa como su potencial zeta. El uso del potencial zeta como una escala común de carga permite la comparación de variantes de enzima con diferentes pliegues (v.g., serina proteasas, metaloproteasas, etc.), así como interacciones con diferentes sustratos (v.g., BMI microswatch - micromuestra de tela manchada con sangre/leche/ tinta) bajo las condiciones de interés (v.g., detergente HDL AATCC). Aunque se prefiere el potencial zeta para comparar diferentes pliegues de proteína, las movilidades electroforéticas o cargas medidas proporcionan asimismo una escala absoluta y son adecuadas para comparaciones. El rendimiento BMI, como una función del potencial zeta de la enzima, está perfectamente descrito como una distribución normal estándar, indicada por la línea sólida, con una media μ igual a -9.68 mV , desviación estándar σ de 11.39 mV y un valor pico de $0.4056 [\text{A600-background}]$. Esta distribución es indicada en coordenadas reducidas estándar por la actividad BMI dividida por el valor pico en el eje Y derecho, como una función de la puntuación Z en el eje X superior. La puntuación Z se define usualmente como $(X - \mu)/\sigma$ en donde X en este caso es el potencial zeta.

Tabla 1-1. Actividad BMI Microswatch de Proteasas		
Nivel de Rendimiento	Puntuación Z	Ventana de Potencial Zeta ζ * para Actividad BMI Microswatch
90%	± 0.46	$-14.92 < \zeta < -4.44$
80%	± 0.65	$-17.08 < \zeta < -2.28$
70%	± 0.84	$-19.25 < \zeta < -0.11$
60%	± 1.00	$-21.07 < \zeta < +1.71$
50%	± 1.18	$-23.12 < \zeta < +3.76$

*Media μ = -9.68 mV, desviación estándar σ = 11.39 mV, potencial zeta ζ = $Z \cdot \sigma + \mu$

Amortiguador de referencia: 5mM HEPES pH 8.0, 2.5mM NaCl

[202] La distribución normal es única para cada mancha de sustrato bajo condiciones de reacción dadas (pH, conductividad, tipo de sal, quelantes de detergente, etc.). Diferentes beneficios o resultados favorables siguen una distribución normal con una propiedad física que se mantiene a través de enzimas de varios pliegues, como es por ejemplo el caso de niveles de expresión y potenciales zeta para variantes de escalera de cargas ASP y NprE. En una distribución normal, el valor pico ocurre en la media. La comparación de cargas de enzima y sustrato en una escala común de potencial zeta revela que el rendimiento BMI óptimo ocurre cuando el potencial zeta medio de la enzima, en este caso -9.68 mV, coincide esencialmente con el potencial zeta de la mancha sustrato, en este caso -8.97 mV, medido bajo las mismas condiciones.

[203] Niveles de rendimiento de distribuciones normales estándar se describen convenientemente en términos de sus puntuaciones z como se indica en la Tabla 1-1 (Véase, Abramowitz y Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover, New York, 9th Ed., 1964). La conversión en potenciales zeta es un conocimiento claro dado de la media y la desviación estándar, que define la distribución para una aplicación dada. En este ejemplo el rendimiento de limpieza medido para un pliegue de proteína se limita a valores de potencial zeta entre -40 mV y + 20 mV. Variantes con un rendimiento de limpieza por encima de 80% de su pliegue óptimo (*es decir*, $z = + 0.65$), se limitan a valores de potencial zeta entre -17.08 mV y -2.28 mV. Variantes con un rendimiento de limpieza por encima de 90% de su pliegue óptimo (*es decir*, $z = + 0.46$), se limitan a valores de potencial zeta entre -14.92 mV y -4.44 mV.

[204] Diferentes manchas de sustrato (v.g., pasto, fluidos corporales, tomate) tienen diferentes potenciales zeta bajo la misma formulación, y la misma mancha de sustrato tiene diferentes potenciales zeta bajo diferentes formulaciones (v.g., North American HDL, European powder dishwashing detergent). No obstante, aunque la carga de la mancha del sustrato varía, se espera que la desviación estándar de la distribución normal permanezca constante. El conocimiento de los potenciales zeta de enzima y sustrato en una formulación detergente dada permite la rápida identificación del nivel de rendimiento esperado para esa variante, así como la dirección y magnitud del cambio de carga necesario, con el fin de alcanzar niveles óptimos de rendimiento. La medición del potencial zeta del sustrato en el medio de reacción deseado permite la optimización de la reacción enzimática sobre el sustrato particulado en ese medio. Cualquier reacción enzimática en cualquier medio puede optimizarse usando un proceso similar.

Optimización de Reacciones en Sustratos que Exhiben Carga Variable

[205] La conversión de celulosa se evaluó por técnicas conocidas en el arte (véase, v.g., Baker y col., Appl Biochem Biotechnol, 70-72:395-403, 1998). En los experimentos se empleó un análisis de conversión celulósica estándar. En este análisis, enzima y sustrato amortiguado se pusieron en recipientes y se incubaron a una temperatura durante un tiempo. La reacción se extinguió con 100 mM de glicina suficiente, pH 11, para llevar el pH de la mezcla de reacción a por lo menos pH 10. Una vez se extinguió la reacción, una alícuota de la mezcla de reacción se filtró a través de una membrana de 0.2 micrones para retirar los sólidos. La solución filtrada se analizó entonces para detectar la presencia de azúcares solubles mediante HPLC, de acuerdo con los métodos descritos en Baker y col., más arriba.

Determinación del Potencial Zeta de Rastrojo de Maíz Pretratado (PCS)

[206] Rastrojo de maíz se pretrató con 2% w/w H₂SO₄, como se describe (Schell y col., J Appl Biochem Biotechnol, 105:69 - 86 [2003]) y se siguió por múltiples lavados con agua desionizada para obtener un pH de 4.5. Para obtener una concentración final de 50mM se agregó acetato sódico, y la solución se valoró volumétricamente a pH 5.0. La concentración de celulosa en la mezcla de reacción fue aproximadamente 7%.

[207] Alícuotas de PCS antes y después de la sacarificación por una mezcla de celulasas comercial, Spezyme CP e Indigage 44L, se dosificaron en tubos de

centrifugación Eppendorf de 1.5 mL y ocuparon alrededor de una tercera parte del volumen. Las muestras se centrifugaron a 6.000 rpm durante 5 min, el sobrenadante se cambió por agua Milli-Q™, y el proceso se repitió 5 veces. Se preparó una solución patrón concentrada de 100 mg/mL en agua Milli-Q™ del rastrojo de maíz lavado ("*the rinsed corn stover*"). Esta solución patrón se diluyó a 1 mg/mL en un amortiguador 50 mM de acetato sódico, pH 5.0, para mediciones de potencial zeta. Una alícuota de 1 mL de cada muestra de sustrato se transfirió a una cubeta limpia, desechable, Zetasizer NS™ de Malvern Instruments (RU).

[208] Tabla 1-2 indica que durante el curso de la reacción de sacarificación la carga del sustrato PCS, expresada como potencial zeta, se hizo casi dos veces negativa. Sin entrar en limitaciones teóricas, existen muchas explicaciones para un aumento de la carga negativa neta incluyendo, pero lo limitado a, enriquecimiento en lignina, la porción no reactiva de este sustrato, así como unión no productora o acumulación indeseada de depósitos ("*fouling*") de celulasas enteras y otras proteínas. Existe un potencial zeta óptimo de la enzima para el rendimiento (v.g., grado de reacción y tasa de reacción), que coincide con el potencial zeta del sustrato bajo condiciones del medio de reacción. Diferentes pretratamientos de biomasa pueden influir dramáticamente la carga inicial del sustrato. Si la enzima o el sustrato resultan incompatibles en cuanto al potencial zeta durante el curso de la reacción, la interacción enzima-sustrato ya no será óptima. Este efecto será dramático para cambios de aproximadamente 10 mV, los cuales son el caso para conversión de biomasa.

[209] Estrategias para remediar esta situación incluyen, pero no se limitan a, suministrar una mezcla de enzimas que abarque varias cargas; un método de proceso con lote alimentado, en el que enzimas que poseen diferentes cargas con el nuevo óptimo se suministran en diferentes tiempos de reacción y/o grados de conversión; control de la carga superficial del sustrato mediante la adición de agentes de formulación, particularmente tensioactivos (iónicos y no iónicos) u otras proteínas; control de la carga superficial del sustrato mediante ajustes de pH; ajustes de fuerza iónica durante la reacción con el fin de cambiar los óptimos de carga de la enzima; filtración de membrana, particularmente ósmosis inversa y nanofiltración, para controlar la fuerza iónica de principio a fin de la reacción; adición de quelantes para controlar la fuerza iónica mediante la eliminación de sales; y control de la carga del sustrato biomasa mediante procesos de pretratamiento.

Tabla 1-2. Potencial Zeta de Rastrojo de Maíz Pretratado con Ácido	
Condición PCS	Potencial Zeta
Inicial (antes de la sacarificación)	-12.0 ± 7.00 mV
Después de la sacarificación por Spezyme CP	-22.2 ± 8.67 mV
Después de la sacarificación por Indiage 44L	-22.7 ± 6.84 mV

[210] En los ejemplos descritos a continuación se usaron los siguientes análisis. Todas las desviaciones de los protocolos provistos más adelante se indican en los ejemplos. En estos experimentos, se utilizó un espectrofotómetro para medir la absorbencia de los productos formados luego de concluidas de las reacciones.

Análisis de Hexocinasa para la Medición de Glucosa Residual

[211] La glucosa residual de sobrenadantes de cultivo de *H. jecorina* que expresan CBH2 variantes se midió empleando un análisis de hexocinasa. Un volumen de 5 µl de sobrenadante se agregó a un análisis de hexocinasa de 195 µl glucosa (Instrumentation Laboratory, Breda, Países Bajos) en una placa de microtitulación de 96 pozos (Costar Flat Bottom PS). Las placas se incubaron a temperatura ambiente durante 15 min. Luego de la incubación, se midió la absorbencia a 340 nm OD. Los sobrenadantes de cultivos que expresan glucosa residual, se excluyeron de la combinación ("**pooling**") para estudios posteriores.

Análisis HPLC para Determinación del Contenido de Proteína

[212] La concentración de proteínas CBH2 variantes de sobrenadantes de cultivo combinados se determinó usando una HPLC Agilent 1100 (Hewlet Packard) equipada con una columna Proswift RP 2H (Dionex). 10 microlitros de muestra, mezclados con 50 µl de 10% acetonitrilo en agua desmineralizada filtrada, se inyectaron luego de equilibrar la columna HPLC con 10% acetonitrilo, que contenía 0.01% ácido trifluoroacético. Los compuestos se eluyeron usando un gradiente de 10% a 30% acetonitrilo de 0.3 min a 1 min, seguido por un gradiente de 30% a 65% de 1 min a 4 mins. Se determinaron las concentraciones de proteína de CBH2 variantes de una curva de calibración generada, usando CBH2 natural purificada (6.25, 12.5, 25, 50 µg/ml). Para calcular el índice de rendimiento (P_i o PI), se promediaron la relación de la proteína total (promedio), producida por una variante, y la proteína total (promedio), producida por la natural con la misma dosis.

Determinación de la Actividad Específica por Análisis de Hidrólisis con Celulosa Esponjada con Ácido Fosfórico (PASC)

[213] *Hidrólisis de Celulosa*: Se preparó celulosa esponjada con ácido fosfórico (PASC) de Avicel, de acuerdo con un método publicado (Walseth, Tappi 35:228, 1971; y Wood, Biochem J, 121:353-362, 1971). Este material se diluyó con amortiguador y agua para obtener una mezcla 1% w/v, de modo que la concentración final de acetato sódico fue 50mM, pH 5.0. 100 microlitros de una suspensión al 1% de PASC en 50 mM amortiguador de acetato sódico (pH 5.0) se dispensaron en una placa de microtitulación de 96 pozos (Costar Flat Bottom PS). 10 microlitros de un sobrenadante de cultivo 5 mg/ml de una cepa eliminada de CBH2 se agregó a la PASC, y 5, 10, 15, o 20 μ l de sobrenadantes de cultivo combinados de células de *H. jecorina*, que expresan CBH2 naturales o CBH2 variantes, se agregaron a ella. La eliminación del gen CBH2 de *Hypocrea jecorina* (también conocida como *Trichoderma reesei*) se describe en las Patentes de los Estados Unidos Nos. 5,861,271 y 5,650,322. Se agregaron volúmenes compensadores de amortiguador acetato para compensar diferencias en el volumen total. La placa de microtitulación se selló y se incubó en un incubador con termoestato a 50°C bajo agitación continua a 900 rpm. Después de dos horas, se detuvo la reacción de hidrólisis por la adición de 100 μ l de amortiguador glicina, pH 10, a cada pozo. Los productos de la reacción de hidrólisis se examinaron con el análisis PAHBAH.

[214] *Análisis PAHBAH*: Alícuotas de 150 μ l de reactivo de azúcar reductor PAHBAH (5% w/v hidrazida de ácido p-hidroxibenzoico (PAHBAH, Sigma # H9882, disuelto en 0.5 N HCl), (Lever, Anal Biochem, 47:273-279, 1972) se agregaron a todos los pozos de una placa de microtitulación vacía. 10 microlitros de los sobrenadantes de la reacción de hidrólisis se agregaron a la placa de reacción PABA. Todas las placas se sellaron e incubaron a 69°C bajo agitación continua de 900 rpm. Después de una hora, las placas se pusieron sobre hielo durante cinco minutos y se centrifugaron a 720 x g a temperatura ambiente durante cinco minutos. Muestras de 80 μ L de las mezclas de reacción PAHBAH desarrolladas se transfirieron a una placa fresca (leída), y se midió la absorbencia a 410 nm en un espectrofotómetro. Un estándar de celobiosa se incluyó como control. Se generó una curva de respuesta a la dosis para la proteína CBH2 natural. Para calcular el índice de rendimiento (P_i o PI), se promediaron la relación del azúcar total (promedio) producido por una variante y el azúcar total (promedio) producido por la natural con la misma dosis.

Determinación de la Actividad Específica por Hidrólisis de Rastrojo de Maíz Pretratado (PCS) con Ácido Diluido

[215] *Rastrojo de Maíz Pretratado (PCS - Pretreated Corn Stover)*: se pretrató rastrojo de maíz con 2% w/w H_2SO_4 como se describe (Schell y col., J Appl Biochem Biotechnol, 105:69-86, 2003) y se siguió por múltiples lavados con agua desionizada para obtener una pasta que tenía un pH de 4.5. A continuación se agregó amortiguador acetato sódico (pH 5.0) a una concentración final de 50mM acetato sódico y, si era necesario, esta mezcla se valoró entonces volumétricamente a un pH 5.0 usando 1N NaOH. La concentración de celulosa en la mezcla de reacción fue aproximadamente 7%. Se agregaron 65 microlitros de esta suspensión de celulosa por pozo a una placa de microtitulación de 96 pozos (Nunc Flat Bottom PS). Se agregaron al PCS 10 microlitros de un sobrenadante de cultivo 5 mg/ml de una cepa eliminada de CBH2, y 5, 10, 15, o 20 μ l de sobrenadantes de cultivo agrupados de células de *H. jecorina* que expresan CBH2 naturales o CBH2 variantes se agregaron a ella. Para compensar diferencias en el volumen total, se agregaron volúmenes compensadores de amortiguador acetato. Después de sellarlas, las placas se pusieron en un incubador con termostato a 50°C bajo agitación continua de 1.300 rpm durante 5 minutos. Las placas se incubaron luego a 50°C agitando al mismo tiempo a 220 rpm bajo 80% de humedad, para evitar la desecación. Después de 7 días las placas se pusieron sobre hielo durante 5 min y se detuvo la reacción de hidrólisis por la adición de 100 μ l de amortiguador glicina, pH 10, a cada pozo. Los productos de reacción de hidrólisis se examinaron con el análisis PAHBAH.

[216] *Análisis PAHBAH*: alícuotas de 150 μ l de reactivo de azúcar reductor PAHBAH (5% w/v hidrazida de ácido p-hidroxibenzoico (PAHBAH, Sigma # H9882, disuelta en 0.5 N HCl), (Lever, Anal Biochem, 47:273-279, 1972), se agregaron a todos los pozos de una placa de microtitulación vacía. Se agregaron 10 microlitros de los sobrenadantes de la reacción de hidrólisis a la placa de reacción PABA. Todas las placas se sellaron e incubaron a 69°C bajo agitación continua de 900 rpm. Después de una hora las placas se pusieron sobre hielo durante cinco minutos y se centrifugaron a 720 x g a temperatura ambiente durante cinco minutos. Muestras de 80 μ L de las mezclas de reacción PAHBAH desarrolladas se transfirieron a una placa fresca (leída), y se midió la absorbencia a 410 nm en un espectrofotómetro. Un estándar de celobiosa se incluyó como control. Se generó una curva de respuesta a la dosis para la proteína CBH2 natural. Para calcular el índice de rendimiento (P_i o PI), se promediaron la relación del azúcar total (promedio) producido por una variante y el azúcar total (promedio)

producido por la natural con la misma dosis.

Estabilidad de CBH2 Variantes en Presencia de Etanol

[217] Se examinó la estabilidad de CBH2 naturales y CBH2 variantes en la presencia de 4.5% etanol (EtOH) a 49°C. Sobrenadantes de cultivo combinados (80 µL) de células *H. jecorina*, que expresan CBH2 variantes, se agregaron a una placa de 96 pozos (Greiner V-bottom PS), que contenía 10 µl de 40.5% EtOH por pozo. Las placas se sellaron e incubaron en un incubador con termostato a 49°C durante 16 horas con agitación a 900 rpm. Luego de la incubación, las placas se pusieron sobre hielo durante 5 minutos. Se determinó la actividad de CBH2 residual usando el análisis de hidrólisis en celulosa esponjada con ácido fosfórico (PASC), como ya se describió.

[218] Para calcular la actividad residual, el valor del producto formado por la adición de 5, 10, 15 y 20 µl de CBH2 incubada en EtOH al análisis PASC de actividad residual se dividió por el valor del producto formado por la adición de 5, 10, 15 y 20 µl CBH2 libre de EtOH al análisis PASC. Los valores individuales de estas cuatro relaciones se promediaron entonces para producir la actividad residual promedio. Para determinar el valor PI para la variante, el valor de la actividad residual promedio para las variantes se dividió entonces por el promedio de los valores de actividad residual de los controles CBH2 natural.

Termoestabilidad de CBH2 Variantes

[219] Se analizó la termoestabilidad de CBH2 naturales y CBH2 variantes a 53°C. Sobrenadante de cultivo combinado (80 µL) de células *H. jecorina*, que expresan CBH2 variantes, se agregó a una placa de 96 pozos (Greiner V-bottom PS). Las placas se sellaron e incubaron en un incubador con termostato a 53°C durante 16 horas con agitación a 900 rpm. Luego de la incubación, las placas se pusieron sobre hielo durante 5 minutos. Se determinó la actividad de CBH2 residual usando el análisis de hidrólisis con celulosa esponjada con ácido fosfórico (PASC), como se describe más arriba.

[220] Para calcular la actividad residual, el valor del producto formado por la adición de 5, 10, 15 y 20 µl de CBH2 tratada térmicamente al análisis PASC de actividad residual se dividió por el valor del producto formado por la adición de 5, 10, 15 y 20 µl CBH2 no calentada al análisis PASC. Los valores individuales de estas cuatro relaciones se promediaron entonces para producir la actividad

residual promedio. Para determinar el valor PI para la variante, el valor de la actividad residual promedio para las variantes se dividió entonces por el promedio de los valores de actividad residual de los controles CBH2 naturales.

EJEMPLO 2

Evaluación de la Unión de Lignina

[221] Lignina, un biopolímero complejo de fenilpropanoide, es el principal constituyente no carbohidrato de la madera, el cual se une a fibras de celulosa para endurecer y fortalecer las paredes celulares de las plantas. Debido a que se enlaza cruzadamente con otros componentes de la pared celular, la lignina minimiza la accesibilidad de la celulosa y la hemicelulosa a enzimas que degradan la celulosa. Por consiguiente, la lignina se asocia generalmente con una digestibilidad reducida de toda la biomasa vegetal. En particular, la unión de celulasas a lignina reduce la degradación de celulosa por las celulasas. La lignina es hidrófoba y, aparentemente, de carga negativa. Por lo tanto, se contempla la adición de cargas negativas a las celulasas para reducir su unión a lignina.

[222] Como se describe en esta especificación, se preparó una reacción para cuantificar el efecto de la modificación química sobre la capacidad de una preparación de celulasa *Trichoderma sp.* de unir un componente de polímeros vegetales, a saber, lignina. Se recuperó lignina mediante una digestión extensiva de bagazo de caña de azúcar, pretratado con ácido, por celulasas (100 mg Laminex BG/ g de celulosa), seguida por hidrólisis de las celulasas por serina proteasa no específica, exactamente como se describe en Berlin y col., (Applied Biochemistry and Biotechnology, 121:163-170, 2005), excepto que no se realizaron sonicación, secamiento, molienda y cribado, y se agregó al procedimiento un lavado con ácido (0.1 N HCl) para extraer la proteasa, seguido por lavados repetidos con amortiguador acetato (50 mM acetato sódico, pH 5) para devolver la muestra a un pH 5. En síntesis, 50 µL de 1.16% lignina (recuperada de la sacarificación completa del bagazo), preparados en 50mM amortiguador acetato sódico a pH 5 se combinaron con 4 µl de una preparación de celulasa *Trichoderma sp.* modificada o no modificada, desalinizada. Tubos de microcentrífuga, que contenían la muestra de reacción, se incubaron a temperatura ambiente durante 1 hora, y luego se centrifugaron a alta velocidad para separar los materiales solubles de los insolubles. Luego se recogieron 10 µl del sobrenadante de cada tubo. Los tubos de reacción se re-mezclaron e incubaron durante 2 horas más, luego de las cuales se recogieron segundas alícuotas de 10 µl del sobrenadante de cada tubo. Las muestras de sobrenadante

se analizaron por electroforesis en gel de poliacrilamida SDS (SDS-PAGE). La reducción de la intensidad de la banda en preparaciones de celulasa *Trichoderma sp.* fue indicativa de una reducción en la unión de lignina.

EJEMPLO 3

Evaluación de la Unión de Bagazo

[223] El bagazo es la biomasa que queda luego de que la caña de azúcar se ha triturado para extraer su jugo. Una solución que contenía 2% de celulosa de bagazo (tratado con ácido, 28% de sólidos, 57% glucano) se preparó en 50 mM acetato sódico a pH 5. Muestras de preparaciones de celulasa de *Trichoderma sp.* modificada químicamente o no modificada se diluyeron 10 veces en el mismo amortiguador acetato sódico. Alícuotas de las enzimas diluídas se mezclaron con solución de bagazo o amortiguador solo y se incubaron durante 1 hr a temperatura ambiente. Se recogió el sobrenadante y se analizó para detectar actividad de un componente de celulasa, es decir, beta-glucosidasa.

[224] Se midió la actividad de beta-glucosidasa usando el análisis cloro-nitro-fenil-beta-D-glucósido (CNPG – *chloro-nitro-phenyl-beta-D-glucoside*). El análisis CNPG es un análisis cinético en el cual β -glucosidasa convierte CNPG en el producto coloreado 2-cloro-4-nitrofenol (CNP – *2-chloro-4-nitrophenol*). La OD se mide a 405 nm durante un período de 10 minutos a 37°C. Las tasas se obtienen como Vmax usando el software SpectraMax y luego se convierten a actividad específica (μ M CNP/sec/mg Proteína). En síntesis, 200 μ l de 50 mM amortiguador acetato sódico, pH 5.0, se agregan a cada pozo de una placa de microtitulación de 96 pozos. La placa se cubrió y se puso en un Thermomixer Eppendorf a 37°C durante 15 minutos para permitir equilibrar a temperatura. Luego de equilibrar, se agregaron 5 μ l de las muestras de enzima, diluída serialmente en 50 mM amortiguador acetato sódico, pH 5.0. Una solución patrón concentrada 10 mM CNPG se diluyó 1:5 usando 50 mM amortiguador acetato sódico, pH 5.0, luego 20 μ l de la solución CNPG diluída (2 mM) se agregaron a cada pozo que contenía muestras de enzima. La placa de microtitulación se transfirió a un espectrofotómetro (SpectraMAX, tipo 340; Molecular Devices) ajustado a 37°C, y se leyó la OD a 405 nm durante 0-15 min, leyendo en intervalos ≤ 9 sec.

[225] La cantidad de actividad de beta-glucosidasa de las muestras de enzima de celulasa, que quedó sin unir al sustrato bagazo, fue considerablemente mayor en el caso de la preparación de celulasa *Trichoderma sp.* químicamente modificada. En particular, como se determinó mediante el análisis CNPG, menos de 50% de la

beta-glucosidasa no modificada quedó sin unir (v.g., 50% unida) al sustrato bagazo, mientras que aproximadamente 80% de la bglu quedó sin unir (v.g., 20% unida) al sustrato bagazo. Tomados conjuntamente, los datos de unión de celulasa modificada indican que disminuir las cargas positivas en la celulasa lleva a una unión reducida a un sustrato polimérico de carga más negativa. En este caso, el sustrato polimérico vegetal fue lignina, que quedó en la biomasa tratada con ácido. Se demostró que biomasa tratada con ácido de rastrojo de maíz, un biopolímero vegetal de composición química similar, adopta una carga crecientemente más negativa durante el curso de la sacarificación, como se determinó por la medición del potencial zeta (véase, Tabla 1-2).

EJEMPLO 4

Sacarificación de Bagazo Pretratado con Ácido

[226] Se evaluó la sacarificación de la celulosa presente en bagazo pretratado con ácido (APB - *acid-pretreated bagasse*) que contenía diferentes cantidades de lignina adicional usando preparaciones de celulasa *Trichoderma sp.* químicamente modificada y no modificada, y se analizó mediante HPLC para monitorear la liberación de azúcares, DP1 a DP7. Los resultados se muestran en FIG. 1 como conversión porcentual del sustrato polimérico. En una placa de microtitulación, 200 μ L de APB (3.5% glucano) se prepararon en 50 mM amortiguador acetato sódico a pH 5, y se ajustaron a cantidades variables de lignina. Se agregaron a los pozos 20 microlitros de solución de enzimas de celulasa (LAMINEX BG modificada y no modificada). Las placas se cubrieron con selladores de aluminio y se pusieron en incubadores a 50°C, con agitación, durante 24 hrs o 48 hrs. La reacción se terminó agregando 100 μ l 100 mM glicina, pH 10, a cada pozo. Luego de mezclar concienzudamente, los contenidos de los pozos de la placa de microtitulación se filtraron a través de una placa filtrante Millipore de 96 pozos (0.45 μ m, PES). El filtrado se diluyó en una placa que contenía 100 μ l 10 mM glicina, pH 10, y se midió por HPLC la cantidad de azúcares solubles (DP1 a DP7) producidos. La HPLC Agilent serie 1100 se equipó con una columna de desionización/protección (Biorad Catálogo No. 125-0118) y una columna para carbohidratos con resina iónica a base de plomo Aminex (Aminex Catálogo No. HPX-87P). La fase móvil usada fue agua, con un caudal de 0.6 ml/min. Estándares de azúcares solubles (DP1-DP7) obtenidos de Sigma se diluyeron todos en agua Milli-Q a 100 mg/mL y se usaron para convertir el área pico para los azúcares individuales en concentraciones efectivas de azúcar. El porcentaje de conversión se calculó dividiendo los azúcares medidos de la HPLC por conversión 100% de celulosa en glucosa.

[227] La unión de celulasa a lignina disminuirá su eficiencia de degradar celulosa. Esto se demuestra como una reducción en la conversión de celulosa en la presencia de cantidades crecientes de lignina presente en las reacciones de sacarificación. Esta tendencia persiste en las preparaciones de celulasa no modificada. Sin embargo, hay un incremento del 10% en la conversión de celulosa en las muestras de celulasa modificadas, comparadas con las muestras de celulasa no modificada. Este resultado indica que aumentar la carga negativa de la celulasa reduce la unión no productiva de celulasa a lignina.

EJEMPLO 5

CBH2 Modificada Químicamente Aumentó la Sacarificación de APB

[228] *Trichoderma* CBH1 purificada, CBH2 variante, EG1, EG2 y beta-glucosidasa se modificaron químicamente, como se describe en el Ejemplo 1. La CBH2 variante usada en este experimento tiene múltiples sustituciones P98L/M134V/T154A/I2112V/S316P/S413Y, con números que corresponden a la celulasa CBH2 madura natural), como se describe en la Publicación de los Estados Unidos No. 2006/0205042. La secuencia de aminoácidos de la CBH2 variante madura es como sigue:

QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAA
STTSRVSPPTSRSSSATPPPGSTTTTRVPPVSGTATYSGNPFVGVTLWANAYYASEVS
SLAIPSLTGAMATAAAVAKVPSFVWLDTLDKTPLEQTLADIRAANKNGGNYAG
QFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDEVRTLTVIEPDS
LANLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGWLPANQ
DPAAQLFANVYKNASSPRALRGLATNVANYNGWNITSPPPYTQGNVYNEKLYIHA
IGPLLANHGWNAFFITDQGRSGKQPTGQQQWGDWCNVIGTGFGRPSANTGDSLDD
SFVWVKPGGECGDTSDSSAPRFDYHICALPDALQPAPQAGAWFQAYFVQLLTNANPS
FL (SEQ ID NO:14).

[229] Modificaciones químicas de CBH1, CBH2 variante, EG1, EG2 y beta-glucosidasa se verificaron por su movilidad cambiada en el gel nativo, comparadas con las proteínas no modificadas. CBH1, CBH2 variante, EG1, EG2 y beta-glucosidasa modificadas tienen cargas más negativas. Todas las concentraciones de proteína se midieron usando un espectrofotómetro NanoDrop™ (Thermo). Se estableció una reacción de sacarificación en una placa de microtitulación, en cada pozo de una placa de microtitulación, 150 µL de APB (7% glucano preparada como se describe más arriba para PCS) se prepararon en 50 mM amortiguador acetato sódico a pH 5, 20 µL de la mezcla de enzimas de 21 µg de proteína total se

agregaron de modo que la relación de proteína final a celulosa en cada pozo es 20mg/g. Se elaboraron 6 mezclas de enzimas agregando Bglu modificada o no modificada purificada, CBH2 variante, EG1, o EG2 a un fondo de *T. reesei* en el cual los genes que codifican celobiohidrolasa I (CBHI, Cel7a), celobiohidrolasa II (CBHII, Cel6a), endoglucanasa I (EGI, Cel7b), y endoglucanasa II (EGII, Cel5a) se habían inactivado (véase, US 2007/0128690). En cada mezcla, se agregaron 72.5% de fondo *T. reesei*, 2.5% Bglu, 15% CBH2 variante, 5% EGI, 5% EG2, las cuatro primeras mezclas tienen una proteína que no es modificada, la quinta mezcla tiene toda la proteína no modificada, y la sexta mezcla tiene toda la proteína modificada. La placa se incubó a 50°C durante 72 horas. Se terminó la reacción agregando 100 µl 100 mM glicina, pH 10, a cada pozo. Luego de mezclar concienzudamente, los contenidos de los pozos de la placa de microtitulación se filtraron a través de una placa filtrante Millipore de 96 pozos (0.45 µm, PES). El filtrado se diluyó en una placa que contenía 100 µl 10 mM glicina, pH 10, y se midió por HPLC la cantidad de azúcares solubles (DP1 a DP7) producidos. La HPLC Agilent serie 1100 se equipó con una columna de desionización/protección (Biorad Catálogo No. 125-0118) y una columna para carbohidratos con resina iónica a base de plomo Aminex (Aminex Catálogo No. HPX-87P). La fase móvil usada fue agua con un caudal de 0.6 ml/min. Estándares de azúcares solubles (DP1-DP7) obtenidos de Sigma se diluyeron todos en agua Milli-Q a 100 mg/mL y se usaron para convertir el área pico para los azúcares individuales en concentraciones efectivas de azúcar. El porcentaje de conversión se calculó dividiendo los azúcares medidos de la HPLC por conversión 100% de celulosa en glucosa.

[230] FIG. 2 A muestra que para la sexta mezcla de enzimas (EG2, EG1, CBH2 variante y beta-glucosidasa modificadas) con toda la proteína modificada tiene la máxima conversión de celulosa, y la quinta mezcla de enzimas con toda la proteína no modificada tiene la conversión mínima. Comparando las primeras cuatro mezclas de enzimas, la segunda mezcla de enzimas con la CBH2 no modificada arrojó la concentración mínima. FIG. 2B muestra las ventajas de proteínas modificadas sobre proteínas no modificadas en la conversión de celulosa.

EJEMPLO 6

Preparación de Variantes de Escalera de Cargas CBH2 *T. reesei*

[231] Como se determinó durante el desarrollo de la presente divulgación, la succinilación de residuos de lisina superficial de CBH2 mejoró el rendimiento

sobre APB, y sobre rastrojo de maíz pretratado. La carga de CBH2 modificada fue alrededor de -17 comparada con la CBH2 variante no modificada.

Con esto en mente, se diseñó una escalera de cargas de CBH2 para la determinación de la carga superficial óptima en aplicaciones de rendimiento de celulasa.

[232] SEQ ID NO: 1 divulga la secuencia de DNA que codifica la *Hypocrea jecorina* CBH2 de referencia:

atgattgtcggaattctaccacgctggctacgctggccacactcgagctagtgtgcctctagaggagcggcaagcttgcctaacggt
ctggggccaatgtggtggccagaattggtcggtccgacttctgtgtctccggaagcacatgcgtctactccaacgactattactccc
agtgtcttccggcgctgcaagctcaagctcgctccacgcgccgctcgacgacttctcgagtatccccacaacatcccggctga
gctccggagacgctccacctgggttctactactaccagagtacctccagtcggatcgggaaccgctacgtattcaggcaaccctttgtg
gggtcactccttgggccaatgcatattacgctctgaagtttagcagcctcgtattcctagcttgactggagccatggccactgctgcag
cagctgtcgcaaaggctccctctttatgtggctagatactctgacaagacccctctcatggagcaaaccttggccgacatccgcaccg
ccaacaagaatggcggttaactatgccggacagtttgggtgatgacttgcggatcgcgattgcgctgcccttgctcgaatggcgaa
tactctattgccgatggtggctcgccaaatataagaactatatacgacacattcgtcaaattgctgtggaatattccgatatccggacctt
cctggttattgagcctgactcttggccaacctgggtgaccaacctcggtaactccaaagtgtgccaatgctcagtcagcctaccttgagtgc
atcaactacgcccgcacacagctgaaccttccaaatgttgcgatgtattggacgctggccatgcaggatggcttggctggccggcaaa
ccaagacccggcgctcagctatttgcaaattgttacaagaatgcatcgctccgagagctcttcgaggattggcaaccaatgtcgcca
actacaacgggtggaacattaccagccccccatcgtacacgcaaggcaacgctgtctacaacgagaagctgtacatccacgctattg
gacctcttcttggcaatcacggctgggtccaacgccttcttcatcactgatcaaggctgatcgggaagcagcctaccggacagcaaca
gtggggagactgggtgcaatgtgatcggcaccggatttggattcggccatccgcaaacactggggactcgttgcgtgattcgttgcgtg
ggtaagccaggcgggcgagtgtgacggcaccagcgacagcagtgcgccacgatttgactcccactgtgcgtcccagatgccttgc
aaccggcgccctcaagctgggtgcttgggtccaagcctacttggcagcttctcacaacgcaaacccatcgttctgttaa*

[233] SEQ ID NO:2 divulga la secuencia de la proteína de longitud total *Hypocrea jecorina* CBH2:

MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCL
PGAASSSSSTRAASTTSRVSPPTSRSSSATPPPGSTTTRVPPVGSATYSGNPFVGVTPWA
NAYYASEVSSLAIPSLTGAMATAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNY
AGQFVVYDLPDRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSIDRTLLVIEPDSLA
NLVTNLGTPKCANAQSAYLECINYAVTQNLNPNVAMYLDAGHAGWLGWPANQDPAAQLFANV
YKNASSPRALRGLATNVANYNGWNITSPPSYTGQNAVYNEKLYIHAIGPLLANHGSNAFFI
TDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGECDDGTSDSSAPR
FDSHCALPDALQPAPQACAWFQAYFVQLLTNANPSFL*

[234] SEQ ID NO:3 divulga la secuencia de la proteína madura *Hypocrea jecorina* CBH2:

QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPPTS
 RSSSATPPPGSTTTTRVPPVGSATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAA
 AAVAKVPSFMWLDTLTKPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYS
 IADGGVAKYKNYIDTIRQIVVEYSDIRTLLEVIEDSLANLVTNLGTPKCANAQSAYLECINY
 AVTQLNLPNVAMYLDAGHAGWLGWPAHQDPAAQLFANVYKNASSPRALRGLATNVANYNGWN
 ITSPPSYTQCNVYNEKLYIHAIGPLLANEGWSNAFFITDQGRSGKQPTGQQQWGDWCNVIG
 TGFGRPSANTGDSLLDSFVWVKPGGECGTSDSSAPREFDSHCALPDALQAPQAGAWFQAY
 FVQLLTNANPSFL*

[235] Residuos seleccionados para inducir mutagénesis incluían residuos lisina, arginina, asparagina, y glutamina expuestos, no conservados, los cuales se seleccionaron para sustitución para introducir cargas negativas. Lisinas succiniladas en CBH2 modificada se identificaron por espectrometría de masas y se seleccionaron para mutagénesis a glutamato, resultando en una diferencia de carga de -2 por cada sustitución. Otros residuos se seleccionaron para sustitución mediante análisis de la estructura tridimensional de CBH2 combinada con alineamiento de aminoácidos de secuencias CBH2 homólogas (véase, v.g., Figura 3 de la Publicación de los Estados Unidos No. US 2006/0205042, que se incorpora en esta especificación por referencia). Residuos superficiales, que fueron altamente variables en el alineamiento de secuencias de aminoácidos CBH2, fueron candidatos para mutagénesis. No obstante, se evitó la acumulación de sustituciones en estrecha proximidad. La arginina se reemplazó con glutamina (carga -1), y glutamina y asparagina se sustituyeron con las respectivas variantes carboxilo (carga -1). Adicionalmente, residuos aspartato y glutamato se seleccionaron para sustitución a los respectivos residuos amina para completar la escalera de cargas (carga +1). Sustituciones específicas CBH2 se relacionan en Tabla 6-1, con todas las posiciones mostradas, con la excepción de R63 y R77, localizadas en el dominio catalítico CBH2. Una carga positiva neta puede crearse por eliminación de un residuo de carga negativa, o por la introducción de un residuo de carga positiva. Análogamente puede crearse una carga negativa neta por eliminación de un residuo de carga positiva o por la introducción de un residuo de carga negativa.

Tabla 6-1. Sustituciones CBH2					
Introducción de carga negativa y eliminación de carga positiva	Eliminación de carga positiva	Introducción de carga negativa	Introducción de carga negativa	Eliminación de carga negativa	Eliminación de carga negativa
Lisina	Arginina	Asparagina	Glutamina	Aspartato	Glutamato
K157E	R153Q	N382D	Q204E	D189N	E208Q
K129E	R294Q	N344D	Q147E	D211N	E244Q
K288E	R203Q	N237D	Q239E	D405N	E146Q
K194E	R378Q	N339D	Q281E	D277N	
K356E	R63Q	N289D		D151N	
K327E	R77Q	N161D			
		N285D			
		N197D			
		N254D			
		N247D			

[236] Para la preparación de una escalera de cargas CBH2, se diseñaron 10 variantes de carga CBH2 (C-1 a C-10), que abarcan un intervalo de +8 a -32, comparadas con CBH2 natural en pasos de 4, como se muestra en Tabla 6-2.

Tabla 6-2. Escalera de Cargas CBH2										
C-1	C-2	CBH2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10
8	4	0	-4	-8	-12	-16	-20	-24	-28	-32
D189N	D189N	WT	K157E	K157E	K157E	K157E	K157E	K157E	K157E	K157E
E208Q	E208Q		K129E	K129E	K129E	K129E	K129E	K129E	K129E	K129E
D211N	D211N			K288E	K288E	K288E	K288E	K288E	K288E	K288E
D405N	D405N			K194E	K194E	K194E	K194E	K194E	K194E	K194E
E244Q					K356E	K356E	K356E	K356E	K356E	K356E
D277N					K327E	K327E	K327E	K327E	K327E	K327E
D151N						R153Q	R153Q	R153Q	R153Q	R153Q
E146Q						R294Q	R294Q	R294Q	R294Q	R294Q
						R203Q	R203Q	R203Q	R203Q	R203Q
						R378Q	R378Q	R378Q	R378Q	R378Q
							N382D	N382D	N382D	N382D
							N344D	N344D	N344D	N344D
							N237D	N237D	N237D	N237D
							N339D	N339D	N339D	N339D
								N289D	N289D	N289D
								N161D	N161D	N161D
								Q204E	Q204E	Q204E

Tabla 6-2. Escalera de Cargas CBH2									
								Q147E	Q147E
									N285D
									N197D
									N254D
									N247D
									Q239E
									Q281E
									R63Q
									R77Q

[237] Las secuencias de aminoácidos de las variantes se retradujeron a DNA y codón optimizado para expresión en *Trichoderma reesei* usando software GeneDesigner (DNA2.0). Se sintetizaron los genes variantes cbh2 optimizados en el codón, y se amplificó el DNA de las variantes de carga superficial CBH2 (SCVs - *surface charge variants*) de los constructos DNA2.0 por PCR usando cebadores: GGHTK22 directo 5'-CACCATGATCGTGGGAATTCTTACTACTC-3' (SEQ ID NO: 15); y GGTHK23 inverso 5'-CTACAAAACGAAGGGTTTCGCATT-3' (SEQ ID NO: 16). En un experimento, se utilizó mutagénesis dirigida al sitio para introducir mutaciones K129E y K157E (cbh2 variante de carga C-3) en el DNA genómico de CBH2 natural. CBH2 variante de carga C-3 se clonó en pTrex3GM y se expresó como se describe a continuación.

[238] Los productos PCR se purificaron y clonaron en pENTR/TOPO para transformación de células TOP10 de *E. coli*. DNA plasmídico se aisló de colonias individuales, y se verificó la secuencia correcta. CBH2 SCVs se clonaron en pTrex3GM, pTTTpyr(pcbh1), y pTTTpyr(pstpl) como se muestra en la Tabla 6-3.

Tabla 6-3. Clones de Expresión de Variantes de Carga Superficial CBH2			
CBH2 Variante	Vectores de Destino		
	pTrex3gM	pTTTpyr (P _{cbh1})	pTTTpyr(P _{stpl})
C-1 (pTK354a)	1	11	21
C-2 (pTK355a)	2	12	22
C-3 (pTK356a)	3	13	23
C-4 (pTK357a)	4	14	24
C-5 (pTK358a)	5	15	25
C-8 (pTK361a)	6	16	26
C-6 (pTK359a)	(7)	(17)	(27)
C-7 (pTK360a)	8	18	28
C-9 (pTK362a)	9	19	29

C-10 (pTK363b)	10	20	30
----------------	----	----	----

[239] Variantes de carga superficial CBH2 (SCV - *surface charge variants*) en *T. reesei*. Se realizó transformación biolística de *T. reesei* con el vector de expresión pTrex3gM, que contenía el marco de lectura abierto de la variante de carga cbh2 C3 (con mutaciones K129E y K157E), usando el siguiente protocolo. Se empleó *T. reesei* en el cual se habían inactivado los genes que codifican la celobiohidrolasa I (CBHI, Cel7a), la celobiohidrolasa II (CBHII, Cel6a), la endoglucanasa I (EGI, Cel7b) y la endoglucanasa II (EGII, Cel5a)). La transformación de la cepa *Trichoderma reesei* mediante el método de transformación biolística se llevó a cabo usando un Sistema de Transferencia de Partículas Biolistic® PDS-1000/he de Bio-Rad (Hercules, CA), siguiendo las instrucciones del fabricante (véase WO 05/001036 y US 2006/0003408). Los transformantes se transfirieron a placas nuevas de selección de acetamida. Se inocularon transformantes estables en placas de microtitulación filtrantes (Millipore), que contenían 200ul/pozo de medio mínimo de glicina (6.0 g/L glicina; 4.7 g/L (NH₄)₂SO₄; 5.0 g/L KH₂PO₄; 1.0 g/L MgSO₄·7H₂O; 33.0 g/L PIPPS; p.H. 5.5) con adición pos-estéril de ~2% de una mezcla de glucosa/soforosa como fuente de carbono, 10 ml/L de 100g/L de CaCl₂, 2.5 ml/L de elementos trazadores de *T. reesei* (400X): 175g/L ácido cítrico anhidro; 200g/L FeSO₄·7H₂O; 16g/L ZnSO₄·7H₂O; 3.2 g/L CuSO₄·5H₂O; 1.4 g/L MnSO₄·H₂O; 0.8 g/L H₃BO₃. Los transformantes se desarrollaron en cultivo líquido durante 5 días en una cámara rica en O₂, alojada en un incubador a 28° C. Las muestras de sobrenadante de la placa de microtitulación filtrante se obtuvieron usando un múltiple de vacío. Las muestras se corrieron en geles 4-12% NuPAGE (Invitrogen), de acuerdo con las instrucciones del fabricante. Se tiñó el gel con tinte Simply Blue (Invitrogen). La expresión de variantes de carga CHB2 adicionales puede realizarse empleando este método.

EJEMPLO 7

Generación de Variantes de Carga CBH2 en *T. reesei*

[240] El plásmido pTTTpyr-cbh2 que contenía la secuencia codificante de la proteína *Hypocrea jecorina* CBH2 (SEQ ID NO:1) se envió a BASEClear (Leiden, Países Bajos), Gene Art AG (Regensburg, Alemania), y Sloning BioTechnology GmbH (Puchheim, Alemania) para la generación de Bibliotecas de Evaluación de Sitios (SELS - *Site Evaluation Libraries*). El mapa del plásmido de pTTTpyr-cbh2 se muestra en Figura 5. Se hizo una solicitud a los proveedores para la generación de bibliotecas posicionales en cada uno de los sitios en la proteína madura *Hypocrea jecorina* CBH2 (SEQ IDNO:3). La secuencia de aminoácidos de

la proteína de longitud total CBH2 se muestra en SEQ ID NO:2.

[241] SEQ ID NO: 1 divulga la secuencia de DNA que codifica *Hypocrea jecorina* CBH2:

atgattgtcggcattctcaccacgctggctacgctggccacactcgcagctagtgtgcctctagaggagcggcaagcttgctcaagcgt
ctggggccaatgtggtggccagaattggtcgggtccgacttgctgtgcttcgggaagcacatgcttactccaacgactattactccc
agtgtcttccggcgctgcaagctcaagctcgtccacgcgcgcgcgtcgacgacttctcgagtatccccacacatcccggtcga
gtcccgcgacgcctccacctgggttctactactaccagagtacctccagtcggatcgggaaccgctacgtattcaggcaaccctttgttg
gggtcactccttgggccaatgcataattacgcctctgaagttagcagcctcgtattcctagcttgactggagccatggccactgctgcag
cagctgtcgcgaaggttccctctttatgtggctagatactctgacaagacccctctcatggagcaaaccttggccgacatccgcaccg
ccaacaagaatggcggtaactatgccggacagttgtggtgatgacttgccggatcgcgattgcgctgcccttgccctcgaatggcgaa
tactctattgccgatgggtggcgtcgccaaatataagaactatcgacaccattcgtcaaattgtcgtggaatattccgatatccggacct
cctggttattgagcctgactcttggcaacctggtgaccaacctcggctacccaaagtgtgccaatgctcagtcagcctaccttgagtgc
atcaactacgccgtcacacagctgaaccttccaaatgttgcatgtatttgacgctggccatgcaggatggcttggctggccggcaaa
ccaagaccggcgctcagctatttgcaaatgtttacaagaatgcatcgtctccgagagctcttcgggattggcaaccaatgtcgcca
actacaacgggtggaacattaccagccccccatcgtaacgcaaggcaacgctgctacaacgagaagctgtacatccacgctattg
gacctcttcttggcaatcacggctggtccaacgccttcttcatcactgatcaaggctgatcgggaaagcagcctaccggacagcaaca
gtggggagactggtgcaatgtgatcggcaccggatttggtattcgccatccgcaaacactggggactcgttgctggattcgttgtctg
ggtaagccagggcgagtgtagcggcaccagcgacagcagtcgccacgatttgactcccactgtgcgctcccagatgccttgc
aacggcgccctcaagctggtgcttgggtccaagcctactttgtgcagcttctcacaacgcaaacccatcgttctgtaa*

[242] SEQ ID NO:2 divulga la secuencia de proteína de longitud total *Hypocrea jecorina* CBH2

MIVGILTTLATLATLAASVPLEERQACSSVWGQCGGQNSGPTCCASGSTCVYSNDYYSQCL
PGAASSSSSTRAASTTSRVSPITTSRSSSATPPPGSTTTTRVPPVSGTATYSGNPFVGVTPWA
NAYYASEVSSLAIPSLTGAMATAAAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGNY
AGQFVVYDLPRDCAALASNGEYSIADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLA
NLVTNLGTPKCANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGW PANQDPAAQLFANV
YKNASSPRALRGLATNVANYNGWNITSPPSYTGQNAVYNEKLYIHAIGPLL ANHGWSNAFFI
TDQGRSGKQPTGQQQWGDWCNVIGTGFGIRPSANTGDSLLDSFVWVKPGGEC DGTSDSSAPR
FDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL*

[243] SEQ ID NO:3 divulga la secuencia de proteína madura *Hypocrea jecorina* CBH2:

QACSSVWGQCGGQNWSGPTCCASGSTCVYSNDYYSQCLPGAASSSSSTRAASTTSRVSPPTS
 RSSSATPPPGSTTTTRVPPVGSATYSGNPFVGVTPWANAYYASEVSSLAIPSLTGAMATAA
 AAVAKVPSFMWLDTLDKTPLMEQTLADIRTANKNGGNYAGQFVVYDLPDRDCAALASNGEYS
 IADGGVAKYKNYIDTIRQIVVEYSDIRTLLVIEPDSLNLVTNLGTPKCANAQSAYLECINY
 AVTQLNLPNVAMYLDAGHAGWLGWPANQDPAAQLFANVYKNASSPRALRGLATNVANYNGWN
 IITSPPSYTQGNVYNEKLYIHAIGPLLANHGWSNFAFITDQGRSGKQPTGQQQWGDWCNVIG
 TGFGIRPSANTGDSLDSFVWVKPGGECDDGTSDSSAPRFDSHCALPDALQPAPQAGAWFQAY
 FVQLLTNANPSFL*

[244] Los plásmidos pTTTpyr-cbh2 purificados (p_{cbh1} , Amp^R, Acetamide^R) que contenían los marcos de lectura abiertos, que codifican las secuencias de CBH2 variante, se obtuvieron de los proveedores especificados anteriormente. Protoplastos de la cepa *H. jecorina* ($\Delta eg1$, $\Delta eg2$, $\Delta cbh1$ $\Delta cbh2$) se transformaron con los constructos pTTTpyr y se desarrollaron en agar selectivo, que contenía acetamida a 28°C durante 7 días. En síntesis, se llevó a cabo transformación biolística de *H. jecorina* empleando el siguiente protocolo y una cepa en la cual se habían inactivado los genes que codifican la celobiohidrolasa I (CBHI, Cel7a), la celobiohidrolasa II (CBHII, Cel6a), la endoglucanasa I (EGI, Cel7b) y la endoglucanasa II (EGII, Cel5a). La transformación de la *H. jecorina* por el método de transformación biolística se realizó usando un Sistema de Transferencia de Partículas Biolistic® PDS-1000/he de Bio-Rad (Hercules, CA), siguiendo las instrucciones del fabricante (véase WO 05/001036 y US 2006/0003408). Se cosecharon las esporas, se esparcieron en agar de acetamida, y se incubaron a 28°C durante 7 días. Se cosecharon las esporas en 15% glicerol y se almacenaron a -20°C para uso posterior. Para la producción de la proteína CBH2 variante, se agregó un volumen de 10 µl de suspensión de esporas a 200 µl de medio mínimo glicina suplementado con una mezcla al 2% de glucosa/soforosa en una placa de filtro PVDF: 6.0 g/L glicina, 4.7 g/L (NH₄)₂SO₄; 5.0 g/L KH₂PO₄; 1.0 g/L MgSO₄·7H₂O; 33.0 g/L PIPPS; pH 5.5; con adición pos-estéril de una mezcla de ~2% glucosa/soforosa como fuente de carbono, 10 ml/L de 100g/L de CaCl₂, 2.5 ml/L de elementos trazadores de *T. reesei* (400X): 175g/L ácido cítrico anhidro; 200g/L FeSO₄·7H₂O; 16g/L ZnSO₄·7H₂O; 3.2 g/L CuSO₄·5H₂O; 1.4 g/L MnSO₄·H₂O; 0.8 g/L H₃BO₃. Cada CBH2 variante se desarrolló en cuadruplicado. Luego de sellar la placa con una membrana permeable al oxígeno, las placas se incubaron a 28°C durante 6 días, agitando simultáneamente a 220 rpm. Se recogió el sobrenadante transfiriendo el medio de cultivo a una placa de microtitulación bajo presión reducida y se examinó para detectar la presencia de glucosa residual usando el análisis de hexocinasa, como se describe en el Ejemplo 1.

EJEMPLO 8

Expresión, Actividad y Rendimiento de CBH2 Variantes

[245] Se examinaron variantes de carga *H. jecorina* CBH2 con respecto a varias propiedades de interés. En particular, se analizaron las variantes de celulasa con respecto a la expresión de proteína utilizando el análisis HPLC (HPLC), actividad específica usando el análisis de hidrólisis PASC (Act. PASC) y el análisis de hidrólisis PCS (Act. PCS), estabilidad en la presencia de etanol (relación de EtOH) y termoestabilidad (relación térmica), como se describe en el Ejemplo 1. Los datos de rendimiento para las variantes de carga CBH2 se muestran en la Tabla 8-1. El índice de desempeño (PI - *Performance Index*) es la relación de rendimiento de la celulasa variante a la celulasa progenitora o de referencia. Varios términos expuestos más adelante se emplean para describir la mutación: mutaciones "up" tienen un $PI > 1$; mutaciones neutras tienen un $PI \geq 0.5$, mutaciones no nocivas tienen un $PI > 0.05$; mutaciones nocivas tienen un $PI = 0.05$; mutaciones combinables son aquellas mutaciones para las cuales la variante tiene valores de índice de Rendimiento ≥ 0.5 para por lo menos una propiedad. Las mutaciones combinables son mutaciones que pueden combinarse para proporcionar proteínas con índices de desempeño apropiados para una o más propiedades deseadas. Las posiciones en las cuales ocurren mutaciones se clasifican en la siguiente forma: Posiciones no restrictivas tienen $\geq 20\%$ de mutaciones neutras para por lo menos una propiedad; y las posiciones restrictivas tienen $< 20\%$ de mutaciones neutras para actividad y estabilidad. Las posiciones completamente restrictivas no tienen mutaciones neutras para actividad o estabilidad.

[246] Estos datos pueden usarse para genomanipular cualquier CBH2. Incluso si la CBH2 que se va a genomanipular tiene un aminoácido diferente del de *Hypocrea jecorina* CBH2 en una posición particular, estos datos pueden emplearse para encontrar una sustitución que alterará las propiedades deseadas, identificando la sustitución de mejor elección, incluyendo una sustitución al amino de *Hypocrea jecorina* CBH2 natural.

[247] Tabla 8-11 muestra valores de Índice de Rendimiento (Pi o PI) para variantes de *Hypocrea jecorina* CBH2. Los Índices de Rendimiento menores de o iguales a 0.05 se fijaron en 0.05 y se indican en cursiva negrita en la Tabla.

Tabla 8-1: Índices de Rendimiento de Variantes de Carga CBH2					
Variante	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
R63A	1.21	1.03	0.94	0.97	1.00

Tabla 8-1: Índices de Rendimiento de Variantes de Carga CBH2

Variante	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
R63C	1.47	1.02	1.05	0.96	1.06
R63D	0.12	0.11	0.19	0.05	1.02
R63E	0.88	0.94	0.98	0.96	1.05
R63F	0.91	0.95	0.85	0.94	1.09
R63G	0.71	0.96	0.95	1.08	1.20
R63I	0.81	0.93	1.11	1.03	1.09
R63L	0.94	0.96	0.91	1.06	1.29
R63M	1.22	0.99	1.10	1.05	1.09
R63N	1.23	1.02	1.11	1.03	1.07
R63P	0.96	0.99	1.17	1.04	1.09
R63Q	0.87	0.99	1.01	0.99	1.13
R63S	0.59	0.92	1.09	1.00	0.94
R63T	0.43	0.87	0.94	0.91	0.84
R63V	1.00	0.99	0.89	1.03	1.06
R63W	0.98	0.97	0.91	1.08	1.10
R63Y	1.29	1.03	0.99	1.06	1.07
R77F	0.25	0.81	0.58	0.92	0.81
R77G	0.41	1.02	0.75	1.00	0.93
R77L	0.14	0.53	0.36	0.94	0.79
R77N	0.83	0.99	0.87	1.00	1.02
K129A	0.23	1.03	0.81	0.74	0.57
K129L	0.80	0.94	0.80	0.82	0.71
K129N	0.15	0.93	0.65	0.73	0.61
K129Q	0.41	1.30	0.89	1.01	0.99
K129S	0.37	1.05	0.92	0.84	0.84
K129T	0.16	0.87	0.37	0.81	0.69
K129V	0.17	0.93	0.91	0.63	0.45
K129Y	0.35	0.95	0.82	0.68	0.46
Q147E	0.86	1.00	0.88	1.07	1.05
R153A	0.23	0.50	0.72	0.63	0.64
R153C	0.26	0.51	0.86	0.94	0.88
R153D	0.35	0.35	0.54	0.75	0.72
R153E	0.27	0.69	0.75	0.79	0.80
R153F	0.21	0.38	0.69	0.68	0.58
R153G	0.21	0.31	0.59	0.62	0.92
R153I	0.22	0.28	0.37	0.63	1.09
R153L	0.19	0.30	0.40	0.94	0.44
R153M	0.31	0.70	0.86	0.92	1.05
R153N	0.22	0.20	0.42	1.12	1.23
R153P	0.23	0.14	0.10	0.39	2.17
R153Q	0.24	0.52	0.99	0.77	1.05
R153S	0.41	0.78	0.99	0.93	0.93
R153T	0.31	0.63	0.74	0.84	0.94
R153V	0.23	0.28	0.59	0.63	0.81
R153W	0.42	0.77	1.07	0.90	0.87
R153Y	0.32	0.60	0.73	0.84	0.90
K157A	0.68	1.04	0.93	1.01	0.91
K157D	0.69	0.96	1.01	1.12	0.90
K157E	0.10	0.92	0.75	1.03	0.73
K157F	0.40	1.01	0.94	1.07	0.95
K157G	0.50	0.98	0.76	1.04	0.93
K157I	0.71	1.00	0.81	1.05	0.94
K157L	0.18	0.88	0.60	1.04	0.75
K157M	0.33	0.99	0.77	0.99	0.88
K157P	0.11	0.57	0.63	3.04	1.86
K157Q	0.50	1.00	0.70	1.00	0.88

Tabla 8-1: Índices de Rendimiento de Variantes de Carga CBH2

Variante	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
K157T	0.14	0.79	0.47	1.02	0.83
K157V	0.46	0.99	0.86	0.99	0.91
K157W	1.06	0.95	0.84	1.04	0.93
K157Y	0.77	0.98	0.92	1.00	0.84
N161E	0.76	1.03	1.07	1.12	1.12
D189A	0.29	0.77	0.73	0.79	0.69
D189C	0.38	0.89	0.82	0.90	0.76
D189E	0.44	0.98	0.84	0.94	0.89
D189F	0.39	0.82	0.64	0.94	0.99
D189G	0.27	0.87	0.86	0.88	0.71
D189H	0.39	0.91	0.52	0.91	0.83
D189I	0.16	0.52	0.39	0.63	0.61
D189K	0.37	0.96	0.78	0.89	0.73
D189L	0.33	1.04	0.82	0.89	0.85
D189N	0.55	0.95	1.05	1.00	0.96
D189P	0.12	0.30	0.22	0.72	0.76
D189Q	0.73	0.98	0.72	0.93	0.86
D189R	0.38	0.96	0.82	0.94	0.88
D189S	0.14	0.60	0.32	0.86	1.22
D189T	0.21	0.84	0.84	0.85	0.76
D189V	0.36	1.03	0.99	1.03	0.91
D189W	0.13	0.18	0.34	0.45	1.06
D189Y	0.32	0.52	0.46	1.30	1.34
K194A	0.30	0.67	0.86	0.95	1.00
K194C	0.22	1.08	1.72	0.88	0.88
K194D	0.61	0.98	1.07	1.00	1.03
K194E	0.44	1.00	1.14	1.02	1.14
K194F	0.67	1.03	0.94	1.04	1.13
K194G	0.43	0.92	0.65	0.92	0.96
K194I	0.24	0.38	0.49	0.50	0.59
K194L	0.61	0.98	0.90	1.01	1.22
K194M	0.61	1.02	0.86	1.03	1.17
K194N	0.54	0.99	1.02	1.04	1.17
K194P	0.32	0.39	0.55	0.46	0.38
K194Q	0.64	1.00	1.00	1.03	1.17
K194S	0.66	1.00	0.83	1.12	1.03
K194T	0.08	0.11	1.09	1.17	9.80
K194V	0.23	0.30	0.71	0.55	0.62
K194W	0.43	0.93	0.98	0.98	1.01
K194Y	0.08	0.11	0.36	0.60	2.65
N197D	1.50	1.12	1.08	1.06	1.07
R203A	0.71	0.93	0.75	0.69	0.88
R203F	0.85	0.92	0.99	0.67	0.87
R203G	0.60	1.07	1.13	0.66	1.39
R203I	0.39	0.70	0.64	1.46	3.02
R203L	0.46	1.14	0.54	1.08	1.24
R203M	0.58	0.88	0.91	0.99	1.25
R203N	0.36	1.08	0.82	1.06	1.20
R203P	0.29	0.49	0.28	2.85	4.58
R203Q	0.80	0.99	0.78	0.84	0.86
R203S	0.54	1.19	0.53	1.99	1.96
R203T	0.66	1.04	0.82	0.99	0.91
R203V	0.85	0.94	0.89	0.73	0.78
R203W	0.39	1.10	0.63	2.28	2.42
R203Y	0.57	1.09	1.03	1.03	1.18
Q204D	1.23	1.04	0.73	1.00	1.00

Tabla 8-1: Índices de Rendimiento de Variantes de Carga CBH2					
Variante	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
Q204E	1.08	1.12	1.05	0.93	1.04
Q239D	0.59	0.98	0.91	0.92	0.92
Q239E	0.70	1.10	1.16	1.00	1.04
N247D	1.18	0.76	0.67	0.58	0.96
N254E	0.20	0.84	0.56	0.76	0.91
Q281D	1.10	0.60	0.43	1.33	0.95
Q281E	2.50	0.89	0.85	1.10	0.63
N285D	0.35	0.70	0.59	1.99	2.24
K288A	0.41	0.61	0.68	1.52	2.84
K288C	0.41	0.95	0.63	0.72	1.49
K288D	0.23	0.23	0.05	5.77	5.10
K288E	0.38	0.89	0.81	0.65	1.60
K288F	0.36	0.75	0.29	1.53	2.74
K288G	0.23	0.40	0.12	2.22	4.09
K288H	0.32	0.90	0.48	1.09	1.54
K2 88I	0.40	0.75	0.37	1.51	2.48
K288L	0.80	0.96	0.81	0.60	0.82
K288N	0.71	0.98	0.84	1.08	1.68
K288P	0.26	0.34	0.05	5.00	5.39
K288Q	0.23	0.80	0.18	1.23	2.51
K288S	0.19	0.71	0.21	2.02	2.72
K288T	0.25	0.71	0.35	1.27	2.96
K288V	0.31	0.84	0.38	1.67	1.87
N289D	0.84	0.96	0.87	0.93	0.92
N289E	0.13	0.26	0.63	1.03	1.48
R294A	0.97	1.06	0.89	0.92	0.77
R294C	0.70	1.15	0.74	1.01	1.16
R294D	0.75	0.94	0.85	0.94	1.09
R294E	0.73	0.82	0.66	1.02	1.20
R294F	0.37	0.52	0.54	0.90	1.56
R294G	1.25	0.93	0.84	0.71	0.87
R294I	0.77	0.84	0.72	0.98	1.01
R294L	0.54	1.06	1.03	1.05	1.22
R294M	0.34	0.76	0.78	0.81	1.78
R294N	0.50	0.80	0.64	1.30	1.06
R294P	0.48	0.86	0.74	0.74	0.99
R294Q	0.50	0.79	0.75	1.23	1.30
R294S	0.64	1.00	0.57	1.32	1.47
R294T	1.12	1.03	0.97	1.02	1.10
R294V	0.83	0.96	0.83	0.82	1.03
R294W	1.08	1.01	0.57	0.56	0.58
R294Y	0.26	0.43	0.31	2.39	3.19
K327A	0.86	1.03	0.98	0.76	0.76
K327C	0.42	0.93	1.25	1.28	1.02
K327D	0.31	0.73	0.76	1.13	1.81
K327E	0.73	0.84	0.85	1.05	0.87
K327F	0.24	0.61	0.44	1.91	2.14
K327G	0.24	0.83	0.16	1.65	1.30
K327I	0.31	0.74	0.23	1.44	1.37
K327L	0.26	0.94	0.60	0.89	1.30
K327M	0.41	0.90	0.90	0.87	1.44
K327N	0.41	0.90	1.04	0.74	1.33
K327P	0.21	0.58	0.05	1.65	2.77
K327Q	1.32	1.06	1.42	1.15	0.77
K327S	0.20	0.63	0.58	2.39	2.92
K327T	0.21	0.29	0.09	3.69	7.57

Tabla 8-1: Índices de Rendimiento de Variantes de Carga CBH2					
Variante	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
K327V	0.28	0.86	0.58	0.66	1.26
K327W	0.28	1.05	0.74	0.93	1.57
K327Y	0.93	1.04	1.08	0.49	0.60
N339D	0.92	1.04	1.05	1.04	1.10
N339E	0.94	1.03	0.83	1.08	1.13
N344D	0.29	0.85	1.10	1.43	1.31
K356A	1.22	1.03	0.88	0.23	0.25
K356C	0.62	0.92	0.98	0.29	0.45
K356D	0.42	0.87	0.57	0.22	0.42
K356E	0.43	0.98	0.71	0.28	0.43
K356F	0.69	0.99	0.45	0.15	0.39
K356G	0.97	0.98	0.20	0.20	0.40
K356I	0.76	0.96	0.60	0.26	0.46
K356L	0.67	1.00	0.68	0.45	0.39
K356M	0.68	0.99	0.88	0.54	0.53
K356N	0.98	1.04	0.72	0.19	0.37
K356P	0.23	0.52	0.33	1.09	0.99
K356Q	0.77	0.99	0.77	0.41	0.43
K356S	0.58	0.88	0.48	0.22	0.41
K356T	0.63	1.00	0.49	0.28	0.43
K356V	0.91	0.90	0.75	0.35	0.31
K356W	0.39	0.90	0.42	0.24	0.47
K356Y	0.52	0.96	0.47	0.20	0.51
R378A	0.21	0.76	0.79	0.25	0.54
R378C	0.25	0.34	0.42	0.74	1.01
R378D	0.19	0.23	0.33	0.49	0.58
R378E	0.17	0.18	0.30	0.92	0.94
R378F	0.13	0.39	1.34	0.12	0.38
R378G	0.08	0.22	1.48	0.19	0.82
R378I	0.19	0.42	0.33	0.38	0.69
R378L	0.21	0.75	0.79	0.33	0.53
R378M	0.14	0.52	0.77	0.37	0.47
R378N	0.12	0.19	0.48	0.38	0.81
R378P	0.17	0.65	0.54	0.48	0.62
R378Q	0.20	0.82	0.94	0.56	0.59
R378S	0.18	0.65	0.47	0.45	0.64
R378T	0.18	0.52	1.03	0.25	0.43
R378V	0.36	1.01	0.98	1.06	1.19
R378W	0.09	0.10	0.64	0.34	0.30
R378Y	0.24	0.60	0.62	0.23	0.53
N382D	0.71	0.81	N.D.	1.17	1.15
N382E	2.21	0.16	N.D.	6.19	5.56
D405A	1.75	0.82	0.33	0.32	0.60
D405C	0.45	0.63	0.57	0.72	0.97
D405E	0.85	0.81	0.63	0.50	0.65
D405F	0.79	0.95	0.24	0.56	0.82
D405G	0.77	0.87	0.16	0.56	0.86
D405H	0.27	0.38	0.25	3.35	3.59
D405I	0.62	1.04	0.49	0.83	1.02
D405K	0.63	0.65	0.24	0.43	0.72
D405L	0.41	0.95	0.76	0.38	0.43
D405M	0.52	1.31	0.58	0.59	0.69
D405N	0.74	1.04	0.78	1.34	1.32
D405P	0.77	0.80	0.19	0.99	1.02
D405Q	0.71	1.12	0.39	1.10	1.21
D405R	0.93	1.20	0.36	0.56	0.61

Tabla 8-1: Índices de Rendimiento de Variantes de Carga CBH2					
Variante	HPLC	PASC sp.ac.	PCS sp.ac.	Res EtOH	Res Heat
D405S	0.95	1.24	0.40	0.75	0.96
D405T	0.74	1.17	0.32	0.64	0.83
D405V	0.39	1.05	0.60	1.32	1.24
D405W	1.12	1.17	0.40	0.42	0.48
D405Y	1.56	1.07	0.39	0.31	0.07

EJEMPLO 9

Efecto del Cambio de Carga sobre la Actividad de Variantes CBH2

[248] En este ejemplo, se evaluó el efecto del cambio de carga sobre la actividad de CBH2 en un análisis, con rastrojo de maíz pretratado (PCS), de la actividad de la celulasa. En síntesis, se determinó el número de ganadores PCS en las CBH2 SELs como una propiedad del cambio de carga. En Tabla 9-1, se determinó la relación de los ganadores observados a esperados (*o/e - observed to expected*) en el análisis PCS. Los valores en cursiva negrita son significativamente diferentes del promedio de 10 distribuciones aleatorias más o menos el número de desviaciones estándar (*sd - standard deviations*) listadas en las columnas respectivas.

Tabla 9-1: Efecto de la Carga sobre la Actividad de Variantes CBH2 en un Análisis PCS				
PCS	o/e 1 sd	o/e 2 sd	o/e 3 sd	Resultados
Cambio de carga				
-2.00	1.20	<i>1.20</i>	<i>1.20</i>	>90% de confianza más de lo esperado
-1.00	1.40	1.40	1.40	>99% de confianza más de lo esperado
0.00	<i>1.03</i>	<i>1.03</i>	<i>1.03</i>	como se esperaba
1.00	0.46	0.46	<i>0.46</i>	>95% de confianza menos de lo esperado
2.00	<i>0.00</i>	<i>0.00</i>	<i>0.00</i>	como se esperaba

[249] Como se muestra en Tabla 9-1 y Figura 4, disminuir la carga (v.g., -1, -2) resulta en una frecuencia significativamente más alta de CBH2 ganadoras en el análisis PCS, mientras que aumentar la carga (v.g., +1) resulta en una frecuencia significativamente menor de CBH2 ganadoras en el análisis PCS. En conclusión, la actividad CBH2 sobre PCS se correlaciona con la reducción en carga.

REIVINDICACIONES

Reivindicamos:

1. Una variante de celulasa aislada, en donde la citada variante es una forma madura que tiene actividad de celulasa e incluye una sustitución en una o más posiciones seleccionadas del grupo conformado por 63, 77, 129, 147, 153, 157, 161, 194, 197, 203, 237, 239, 247, 254, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, y 382, en donde las posiciones se numeran por correspondencia con la secuencia de aminoácidos de una celobiohidrolasa II (CBH2) de referencia divulgada como SEQ ID NO:3, y en donde la sustitución en una o más posiciones hace que la variante de celulasa tenga una carga neta más negativa en comparación con la CBH2 de referencia.

2. La variante de celulasa aislada de la reivindicación 1, en donde la sustitución en una o más posiciones comprende la eliminación de una o más cargas positivas.

3. La variante de celulasa aislada de la reivindicación 2, en donde la eliminación de una o más cargas positivas comprende un reemplazo de una lisina o una arginina con un aminoácido neutro.

4. La variante de celulasa aislada de la reivindicación 1, en donde la sustitución en una o más posiciones comprende la adición una o más cargas negativas.

5. La variante de celulasa aislada de la reivindicación 4, en donde la adición de una o más cargas negativas comprende un reemplazo de un aminoácido neutro con un aminoácido cargado negativamente.

6. La variante de celulasa aislada de la reivindicación 1, en donde la sustitución en una o más posiciones comprende la eliminación de una o más cargas positivas y la adición de una o más cargas negativas.

7. La variante de celulasa aislada de la reivindicación 6, en donde la eliminación de una o más cargas positivas y la adición de una o más cargas negativas comprende un reemplazo de una lisina o una arginina con un aminoácido cargado negativamente.

8. La variante de celulasa aislada de la reivindicación 1, en donde la sustitución en una o más posiciones comprende una o más del grupo conformado

por K129E, K157E, K194E, K288E, K327E, K356E, R63Q, R77Q, R153Q, R203Q, R294Q, R378Q, N161D, N197D, N237D, N247D, N254D, N285D, N289D, N339D, N344D, N382D, Q147E, Q204E, Q239E, y Q281E, en donde las posiciones se numeran por correspondencia con la secuencia de aminoácidos de la celobiohidrolasa II (CBH2) de referencia divulgada como SEQ ID NO:3.

9. La variante de celulasa aislada de la reivindicación 1, en donde la variante comprende una sustitución adicional en una o más posiciones adicionales seleccionadas del grupo conformado por 146, 151, 189, 208, 211, 244, 277 y 405, en donde las posiciones adicionales se numeran por correspondencia con la secuencia de aminoácidos de la celobiohidrolasa II (CBH2) de referencia divulgada como SEQ ID NO:3.

10. La variante de celulasa aislada de la reivindicación 9, en donde la sustitución adicional en una o más posiciones adicionales comprende un reemplazo de ácido aspártico o ácido glutámico con un aminoácido neutro.

11. La variante de celulasa aislada de la reivindicación 9, en donde la sustitución adicional en una o más posiciones adicionales comprende una o más del grupo conformado por D151N, D189N, D211N, D277N, D405N, E146Q, E208Q, y E244Q, en donde las posiciones se numeran por correspondencia con la secuencia de aminoácidos de la celobiohidrolasa II (CBH2) de referencia divulgada como SEQ ID NO:3.

12. La variante de celulasa aislada de la reivindicación 1, en donde la sustitución en una o más posiciones se selecciona del grupo conformado por las posiciones 1, 2, 3, 4, 5, 6, 7, 8, 9 y 10.

13. La variante de celulasa aislada de la reivindicación 1, en donde la variante de celulasa se deriva de una celulasa progenitora seleccionada del grupo conformado por *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, y *Cellulomonas fimi* CenA EG.

14. La variante de celulasa aislada de la reivindicación 1, en donde la variante de celulasa se deriva de una celulasa progenitora cuya secuencia de aminoácidos es por lo menos 75% idéntica a un miembro del grupo conformado por SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, y SEQ ID

NO: 13.

15. La variante de celulasa aislada de la reivindicación 1, en donde la carga neta más negativa es -1 o -2, en comparación con la CBH2 de referencia.

16. Un método de convertir biomasa en azúcares, que comprende contactar la citada biomasa con la variante de celulasa de la reivindicación 1.

17. Un método de producir un combustible que comprende:

contactar una composición de biomasa con una composición enzimática que comprende la variante de celulasa de la reivindicación 1 para producir una solución de azúcar; y

cultivar la solución de azúcar con un microorganismo fermentador bajo condiciones suficientes para producir un combustible.

18. Una variante de celulasa, en donde la citada variante es una forma madura que tiene actividad de celulasa y comprende una modificación química de un residuo lisina para remover carga positiva del residuo lisina.

19. La variante de celulasa de la reivindicación 18, en donde la modificación química comprende un tratamiento con un compuesto seleccionado del grupo conformado por anhídrido succínico, anhídrido acetoxisuccínico, anhídrido maleico, anhídrido tartárico, anhídrido ftálico, anhídrido trimelítico, anhídrido cis-aconítico, anhídrido t-nitroftálico, anhídrido acético, anhídrido butírico, anhídrido isobutírico, anhídrido hexanoico, anhídrido valérico, anhídrido isovalérico, y anhídrido piválico.

20. La variante de celulasa de la reivindicación 18, en donde la variante de celulasa se deriva de una celulasa progenitora seleccionada del grupo conformado por una *Hypocrea jecorina* celobiohidrolasa I, *Hypocrea jecorina* celobiohidrolasa II, *Hypocrea jecorina* endoglucanasa I, *Hypocrea jecorina* endoglucanasa II, e *Hypocrea jecorina* beta-glucosidasa.

21. La variante de celulasa de la reivindicación 18, en donde la variante de celulasa se deriva de una celulasa progenitora seleccionada del grupo conformado por *Hypocrea jecorina* CBH2, *Hypocrea koningii* CBH2, *Humicola insolens* CBH2, *Acremonium cellulolyticus* CBH2, *Agaricus bisporus* CBH2, *Fusarium osysporum* EG, *Phanerochaete chrysosporium* CBH2, *Talaromyces emersonii* CBH2, *Thermobifida fusca* 6B/E3 CBH2, *Thermobifida fusca* 6A/E2 EG, y *Cellulomonas*

fimi CenA EG.

22. La variante de celulasa aislada de la reivindicación 18, en donde la variante de celulasa se deriva de una celulasa progenitora cuya secuencia de aminoácidos es por lo menos 75% idéntica a un miembro del grupo conformado por SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, y SEQ ID NO: 13.

23. La variante de celulasa aislada de la reivindicación 18, en donde la variante de celulasa incluye una sustitución en una o más posiciones seleccionadas del grupo conformado por 63, 77, 129, 147, 153, 157, 161, 194, 197, 203, 237, 239, 247, 254, 281, 285, 288, 289, 294, 327, 339, 344, 356, 378, y 382, en donde las posiciones se numeran por correspondencia con la secuencia de aminoácidos de una celobiohidrolasa II (CBH2) de referencia, divulgada como SEQ ID NO:3.

FIG. 1A

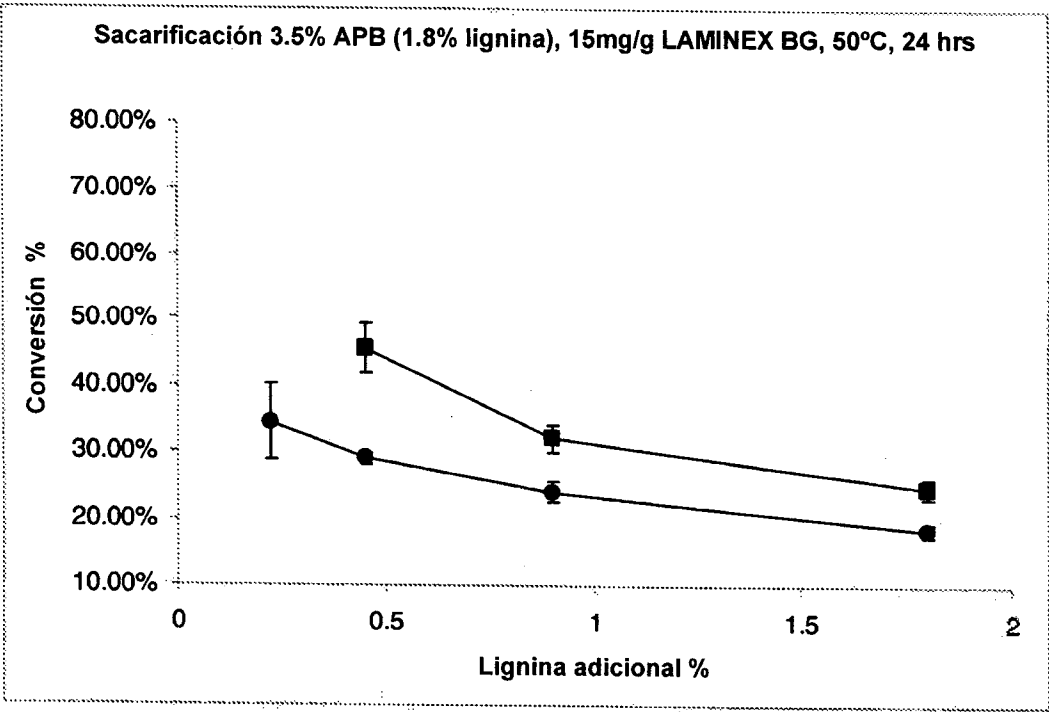


FIG. 1B

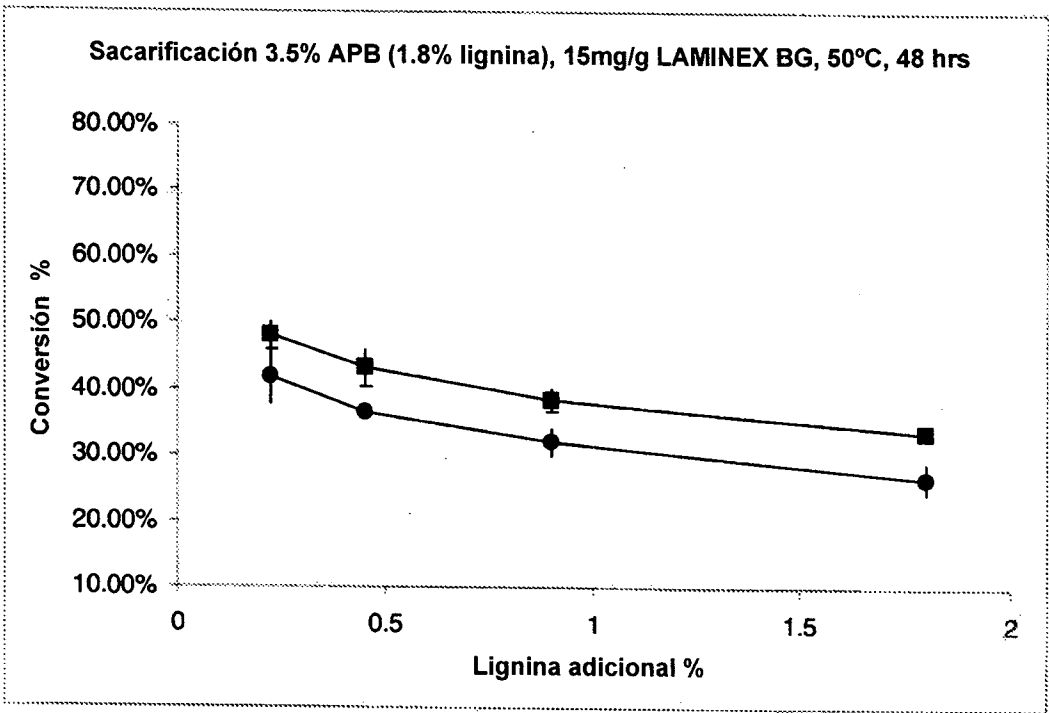


FIG. 2A

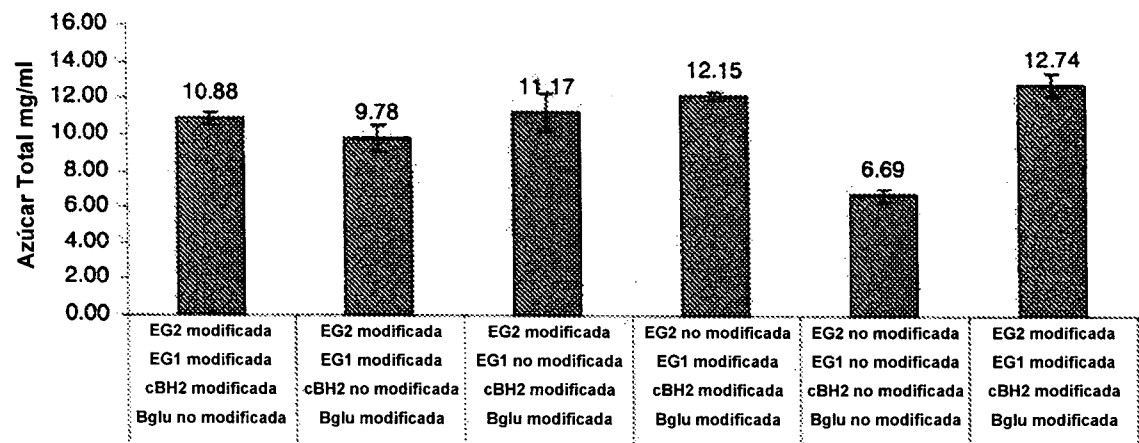
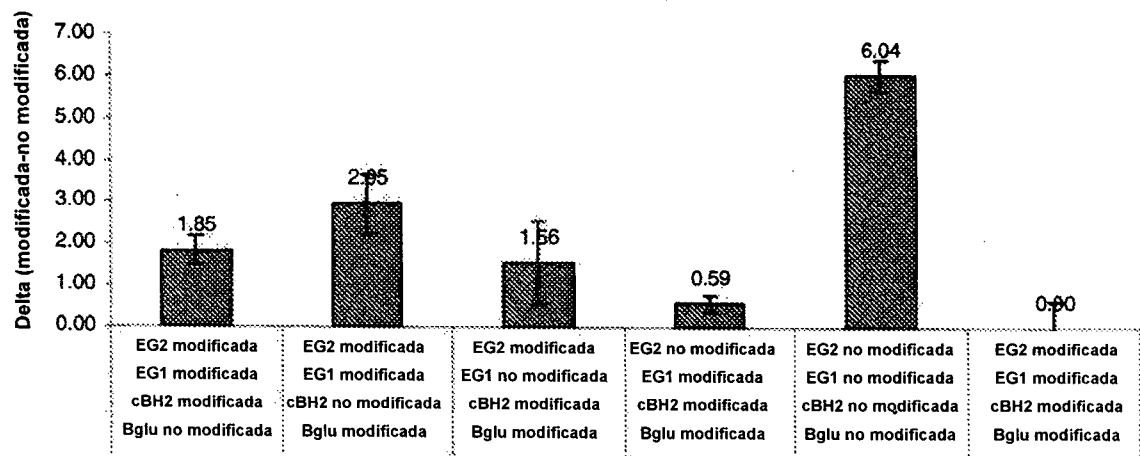


FIG. 2B



3 / 7

FIG. 3

T. reesei CBK #3
H. koningii CBH #4
H. insolens CBH #5
A. cellulolyticus CBH #6
A. bisporus CBH #7
F. oxysporum EG #8
P. chrysosporium CBH #9
T. emersonii CBH #10
T. fusc CBH #11
T. fusc EG #12
C. fimi EG #13

```

-----QACSSVWGQCGGQN-----W
-----QACSSVWGQCGGQN-----W
-----APVVEERQNCAPTWGQCGGIG-----F
-----MLRYLSIVAATAILTGVEAQSSVWGQCGGQG-----W
-----QSPV-----WGQCGGNG-----W
-----APVEERQSCSNGVWAQCGGQN-----W
-----MKSTAFFAALVTLLPAYVAGQASE-----WGQCGGIG-----W
-----MRNLLALAPAAALLVGAAEAQQSLWGQCGGSS-----W
-AGCSVDYTVN-SWGTGFTANVTITNLGSAINGWTLEWDFPGNQVVTNLW
-NDSPFYVNPNMSSAEVVRNPNPD
APGCRVDYAVTNQWPGGFGANVTITNLGDPVSSWKLDWYTAGQRIQQLW

```

T. reesei CBH #3
H. koningii CBH #4
H. insolens CBH #5
A. cellulolyticus CBH #6
A. bisporus CBH #7
F. oxysporum EG #8
P. chrysosporium CBH #9
T. emersonii CBH #10
T. fusc CBH #11
T. fusc EG #12
C. fimi EG #13

SGPTCCASGSTCVYSNDYYSQCLP--GA---ASSSSSTRAAST---TSR
SGPTCCASGSTCVYSNDYYSQCLP--GA-----ASSSSSTRASSTA
NGPTCCQSGSTCVKQNDWYSQCLP--GSQVTTTSTTSTSSSSST---TSR
SGATSCAAGSTCSTLNPYYAQCP--GT-----ATSTTLV---KTT
TGPTTCAAGSTCVKQNDFYSCQLP--NN-----Q
SGTPCCTSGNKCVKLNDFYSCQP--GS-----AEPS---STA
TGPTTCVSGTTCTVLNPYYSCQLP--GS-AVTTTTSVITSHSSS---VSS
TGATSCAAGSTCSTINPYAQCP--AT-----ATPTTLT---TTT
NGTYTQSGQHVSVSNAPNYASI--PANGTVEFGFNGSYSGSNDIPSSFKLN
PRTPVTRDRIASVPQGTWFAHNP--GQ-----ITGQVDA---LMS
NGTASTNGGQVSVTSLPWNGSI--PTGQTASFPGNSGWSNPTPASFSLN

T. reesei CBH #3
H. koningii CBH #4
H. insolens CBH #5
A. cellulolyticus CBH #6
A. bisporus CBH #7
F. oxysporum EG #8
P. chrysosporium CBH #9
T. emersonii CBH #10
T. fusc CBH #11
T. fusc EG #12
C. fimi EG #13

VSPTTSR---SSSATPPPGSTTTRVPP-VGSGTATYSGNPFVGVTPWANA
RASSTTS---RSSATPPPGSSTTRVPP-VGSGTATYSGNPFVGVTPWANA
ATSTTRT---GGVTSITAPTRTVTIPGGATTASYNNGNPFEGVQLWANN
SSTS SVGT-----TSPTPTTTTAKST-TATTTAAASGNPFSGYQLYANP
APPSTTT---QPGTTPPATTSGGTGP-TSGA-----GNPYTGKTVWLSP
AGPSSTT---ATKTTATGGSSTTAGGS-VTSAPPAASDNPNYAGVDLWANN
VSSHSGS---STSTSSPTGPTGTNPPP-PPSA-----NNPWTGFQIFLSP
KPTSTGG---AAPTTPPTPTTGTTTSP-VVTRPASASGNPFEGYQLYANP
GVTCDSGDDPDEPPSPSPSPSPDDEPGGPTNPPTNPGEKVD---NP
AAQAACK---IPILVVYNAPGRDCGNH-SSGAPSHS-----AYRSWIDE
GTTCTGT---VPTTSPTPTPTPTPTPT-TPPTPTPTT---PTVTPOPTSGF

T. reesei CBH #3
H. koningii CBH #4
H. insolens CBH #5
A. cellulolyticus CBH #6
b. bisporus CBH #7
F. oxysporum EG #8
P. chrysosporium CBH #9
T. emersonii CBH #10
T. fusc CBH #11
T. fusc EG #12
C. fimi EG #13

YYASEVSSLAIPSLTG-AMATAAAAVAKVPSFMWLDTL-----DKTP-LM
YYASEVSSLAIPSLTG-AMATAAAAVAKVPSMMLDTF-----DKTP-LM
YYRSEVHTLAIPQITDPALRAAASAVAEVPSFQWLDRN-----VTVDTL
YYSEVHTLAIPSLTG-SLAAAATKAAEIPSFVWLDTA-----AKVP-TM
FYADEVAQ-AAADISNPSLATKAASVAKIPTFVWFDTV-----AKVP-DL
YYSEVNNLAVPKLGS-AKATAAAKVADVPSFQWMDTY-----DHIS-LM
YYANEVAA-AAKQITDPTLSSKAAASVANIPTFTWLDSV-----AKIP-DL
YYASEVISLAIPSLSS-ELVPKASEVAKVPSFVWLDAQ-----AKVP-SM
FEGAKLVNPNVWSAKA-AAEPGSSAVANESATVWLDRIGAIEGNDSPPTG
F-AAGLKNNRPAYIIIVEPDLISLSSSCMQHVQOEVLRTM-----AYACKAL
YVDPPTTQGYRAWOAASGTDKALLEKIALTPQAYVVGNGW-----ADAS---

Figura 3 (continuación)

T. reesei CBH #3	EQTLAD-IRTANKNGGN--YAGQFVVYDLPDRDCAALASNGEYSIADGGV
H. koningii CBH #4	EQTLAD-IRTANKNGGN--YAGQFVVYDLPDRDCAALASNGEYSIADGGV
H. insolens CBH #5	VETLSE-IRAANQAGANPPYAAQIVVYDLPDRDCAAAAASNGEWAIANNGA
A. cellulolyticus CBH #6	GTYLAN-IEAANKAGASPPFIAGIFVVYDLPDRDCAAAAASNGEYTVANNGV
A. bisporus CBH #7	GGYLAD-ARSKNQ-----LVQIVVYDLPDRDCAALASNGEFSLANDGL
F. oxysporum EG #8	EDTLAD-IRKANKAGGK--YAGQFVVYDLPDRDCAAAAASNGEYSLDKDGA
P. chrysosporium CBH #9	GTYLAS--ASALGKSTGKQLVQIVYDLPDRDCAAKASNGEFSIANNGQ
T. emersonii CBH #10	GDYKLD-IQSQNAAGADPPIAGIFVVYDLPDRDCAAAAASNGEFSIANNGV
T. fusc CBH #11	SMGLRDHLEEAVRQSGGDPLTIQVVIYNLPGRDCAALASNGELGP--DEL
T. fusc EG #12	KAGSSQ-ARIYFDAGHSAWHSPAQMASWL----QQADISNSAHGIA-TNT
C. fimi. EG #13	-HAQAE-VADYTGRAVAAGKTPMLVVYAI PGRDCGSHSGGG-----VSE
	: : : . . .
T. reesei CBH #3	AKYKN-YIDTIRQIVVEYSD---IRTLVIEPDSLANTLVNLTGTPK----
H. koningii CBH #4	DKYKN-YIDTIRQIVVEYSD---IRTLVIEPDSLANTLVNLTGTPK----
H. insolens CBH #5	NNYKG-YINRIREILISFSD---VRTILVIEPDSLANTLVNLTGTPK----
A. cellulolyticus CBH #6	ANYKA-YIDSIVAQLKAYPD---VHTILVIEPDSLANTLVNLTGTPK----
A. bisporus CBH #7	NKYKN-YVDQIAAQIKQFPD---VSVVAVIEPDSLANTLVNLTGTPK----
F. oxysporum EG #8	NKYKA-YIAKIKGILQNYSD---TKVILVIEPDSLANTLVNLTGTPK----
P. chrysosporium CBH #9	ANYEN-YIDQIVAQIQQFPD---VRVAVIEPDSLANTLVNLTGTPK----
T. emersonii CBH #10	ALYKQ-YIDSIREQLTTSYSD---VHTILVIEPDSLANTLVNLTGTPK----
T. fusc CBH #11	DRYKSEYIDFIADIMWDFADYENLRIVAI IEIDSLPNLVNLTGTPK----
T. fusc EG #12	SNYRW-TADEVA-----YAK---AVLSAIGNPSLRVIDTSTRNGNG
C. fimi EG #13	SEYAR-WVDTVAQGIK-----GNPILVIEPDALALQGLD-----
	* : : . . .
T. reesei CBH #3	CA--NAQSAYLECIYAVTQL-NLPNVAMYLDAGHAGWLGWPANQDPAAQ
H. koningii CBH #4	CA--NAQSAYLECIYAVTQL-NLPNVAMYLDAGHAGWLGWPANQDPAAQ
H. insolens CBH #5	CS--GAASYRELTIYALKQL-DLPHVAMYMDAGHAGWLGWPANIQPAAE
A. cellulolyticus CBH #6	CA--EAQSAYYECVNYALINL-NLANVAMYIDAGHAGWLGWSANLSPAAQ
A. bisporus CBH #7	CA--NAQSAYKEGVYAVQKL-NAVGVMTYIDAGHAGWLGWPANLSPAAQ
F. oxysporum EG #8	CA--KAESAYKELTVYAIKEL-NLPNVSMYLDAGHGGWLGWPANIGPAAK
P. chrysosporium CBH #9	CA--NAKTYYLACVNYALTNL-AKVGVMYMDAGHAGWLGWPANLSPAAQ
T. emersonii CBH #10	CA--NAQDAYLECIYAVTQL-DLPNVAMYLDAGHAGWLGWQANLAPAAQ
T. fusc CBH #11	CAYMKQNGGYVNGVGYALRLGEIPNVYNYIDAHHGWIGWDSNFGPSVD
T. fusc EG #12	----PAGNEWCDPSGRAIGTPSTTNTGDPMDAFL--WIKLPGEADGCI A
C. fimi EG #13	CS--GQGDRVGFGLKYAAKSL-TLKGARVYIDAGHAKWLSVDTTPVNRLNQ
	* : ** *
T. reesei CBH #3	LFANVYKNAS-SPRALRGLATNVANYNGWNIT-----SPPSY
H. koningii CBH #4	LFANVYKNAS-SPRALRGLATNVANYNGWNIT-----SPPSY
H. insolens CBH #5	LFAKIYEDAG-KPRAVRGLATNVANYNAWSIS-----SPPPY
A. cellulolyticus CBH #6	LFATVYKNAS-APASLRGLATNVANYNAWSIS-----SPPSY
A. bisporus CBH #7	LFAQIYRDAG-SPRNLRLGIATNVANFNALRAS-----SPDPI
F. oxysporum EG #8	LYAQIYKDAG-KPSRVRLVLTNVSNYNGWKLK-----TKPDY
P. chrysosporium CBH #9	LFTQVWQWAG-KSPFIKGLATNVANYNALQAA-----SPDPI
T. emersonii CBH #10	LFASVYKNAS-SPASVRGLATNVANYNAWSIS-----RCPSY
T. fusc CBH #11	IFYEAAANASGSTVDYVHGFISNTANYSATVEPYLDVNGTVNGQLIRQSKW
T. fusc EG #12	GAGQFVPQAAAYEMAIAAGGTNPNNPNP--TP-----TPTPT
C. fimi EG #13	V-----GFEYAVGFALNTSNYQT-----
	* . *

5 / 7

Figura 3 (continuación)

T. reesei CBH #3
H. koningii CBH #4
H. insolens CBH #5
A. cellulolyticus CBH #6
A. bisporus CBH #7
F. oxysporum EG #8
P. chrysosporium CBH #9
T. emersonii CBH #10
T. fusc CBH #11
T. fusc EG #12
C. fimi EG #13

TQGNVAVNEKLYIHAIGPLLANHGWSN-AFFITDQGRSGK---OPTGQQ
TQGNVAVNEQLYIHAIGPLLANHGWSN-AFFITDQGRSGK---OPTGQQ
TSPNPNYDEKHYIEAFRPLLEARGFP--AQFIVDQGRSGK---OPTGQK
TSGDSNYDEKLYINALSPLLTNSNGWPN-AHFIMDTSRNGV---OPTKQQ
TQGNSNYDEIHYIEALAPMLSNAGFP--AHFIVDQGRSGV---QNIRD
TESNPNYDEQRYINAFAPLLAQEGWSN-VKFIVDQGRSGK---OPTGQK
TQGNPNYDEIHYINALAPLLQQAGWD--ATFIVDQGRSGV---QNIRQ
TQGDANCDEEDYVVALGPLFQEQGFP--AYFIIDTSRNGV---RPTKQS
VDWNQYVDELSFVQDLRQALIAKGFRSDIGMLIDTSRNGWGGPNRPTGPS
PTPPPGSSGACTATYTIANEWNDGFQATVTVTANQNTITGW-----TV
-----TADSKAYGQQISQRLGG-----KKFVIDTSRNGN-----G

T. reesei CBH #3
H. koningii CBH #4
H. insolens CBH #5
A. cellulolyticus CBH #6
A. bisporus CBH #7
F. oxysporum EG #8
P. chrysosporium CBH #9
T. emersonii CBH #10
T. fusc CBH #11
T. fusc EG #12
C. fimi EG #13

QWGD-----WCNVIGTGFGIRPSANTGDSLLDSFVWV
QWGD-----WCNVIGTGFGIRPSANTGDSLLDSFVWI
EWGH-----WCNAIGTCFGMRPTANTGHQYVDAFVWV
AWGD-----WCNVIGTGFGVQPTTNTGDFLEDAFVWV
QWGD-----WCNVKGAGFGQRPTTNTGSSSIDAIVWV
AQGD-----WCNAKGTGFGLRPSTNTGDALADAFVWV
QWGD-----WCNIKAGFGTRPTTNTGSGQFIDSIVWV
QWGD-----WCNVIGTGFGVRPTTDTGNPLEDAFVWV
SSTDNLNTYVDESRIIDRIHGPWNWCNQAGAGLGERPTVNPA-PGVDAIVWV
TWT-----FTDQGTITNAWNADVSTSGSSVTARNVG
SNGE-----WCNPRGRALGERPVAVNDGSGLDALLWV

T. reesei CBH #3
H. koningii CBH #4
H. insolens CBH #5
A. cellulolyticus CBH #6
A. bisporus CBH #7
F. oxysporum EG #8
P. chrysosporium CBH #9
T. emersonii CBH #10
T. fusc CBH #11
T. fusc EG #12
C. fimi EG #13

KPGGECDDTSDSS-----APRFDHCA-----LPDALQPAPOAG
KPGGECDDTSDSS-----APRFDHCA-----LPDALQPAPOAG
KPGGECDDTSDTT-----AARYDYHCG-----LEDALKPAPEAG
KPGGESDGTSSNS-----ATRYDFHCG-----YSDALQPAPEAG
KPGGECDDTSDNS-----SPRFDHCS-----LSDAHQPAPEAG
KPGGESDGTSDTS-----AARYDYHCG-----LDDALKPAPEAG
KPGGECDDTSSNS-----SPRYDSTCS-----LPDAAQPAPEAG
KPGGESDGTSTNT-----SPRYDYHCG-----LSDALQPAPEAG
KPPGESDGAEEIIPNDEGKGFDRCMDPTYQGNARNGNPNPSGALPNAPISG
HNGTSLSQGASTE-----GFVG
KLPGESDGCACNG-----PAAG
: . : * . *

T. reesei CBH #3
H. koningii CBH #4
H. insolens CBH #5
A. cellulolyticus CBH #6
A. bisporus CBH #7
F. oxysporum EG #8
P. chrysosporium CBH #9
T. emersonii CBH #10
T. fusc CBH #11
T. fusc EG #12
C. fimi EG #13

AWFQAYFVQLLTNANPSFL
AWFQAYFVQLLTNANPSFL
QWFQAYFEQLLRNANPPF-
TWFOAYFVQLLTNANPALV
TWFOAYFETLVANANPAL-
TWFOAYFEQLLDNANPSFL
TWFOAYFOTLVSAANPPL-
TWFOAYFEQLLTNANPLF-
HWFSAQFRELLANAYPPL-
SKGNSNSVPTLTCAAS---
QWWQEIALEMARNARW---

Figura 4

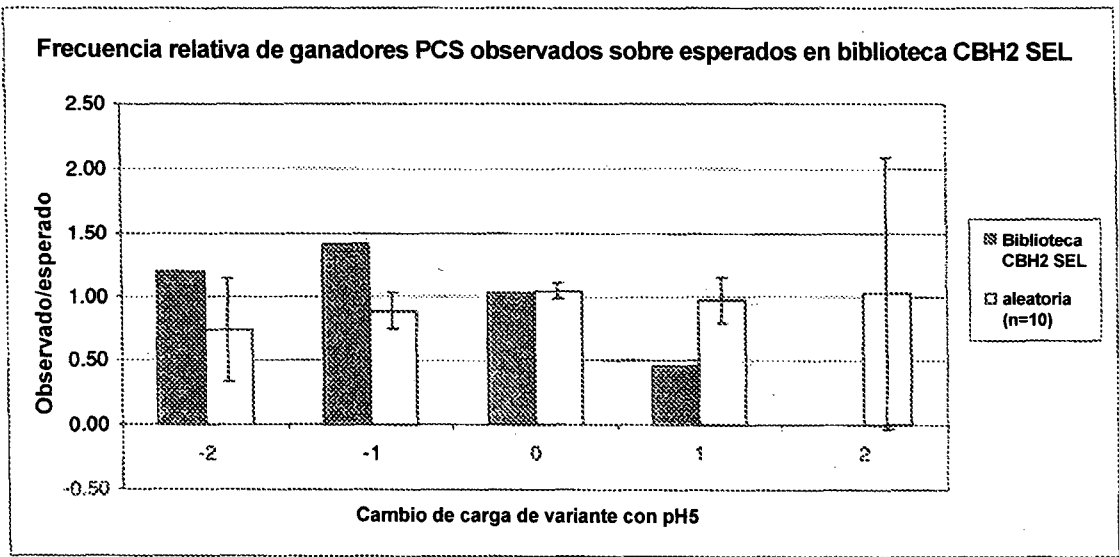
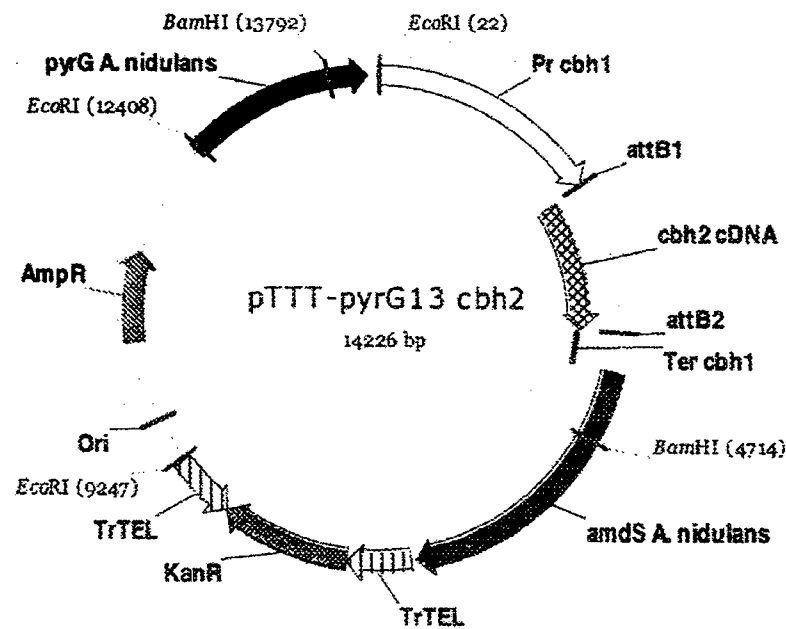


Figura 5



SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800176089-2

RECIBO OFICIAL DE CAJA : 10 - 102,472
FECHA : DICIEMBRE 1 DE 2010

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 10-151612-00000-0000

Fecha: 2010-12-01 17:17:51 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 411 PRESENTACION Folios: 190

***** CONSIGNACION *****

DEPOSITANTE	TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PAGO	Vr. PAGO
CASTILLO PINEDA	CONSIGNACION	BANCO DE BOGOTA	062754387	383679450	01/12/2010	10,128,000.00

***** CONCEPTO *****

CANT. RENTISTICO	CONCEPTO	TOTAL CONCEPTO
1 50005-01-01 SOLICITUDES	1 TRAMITES DE SOL. DE PATENTE DE IN	554,000.00
	TOTAL :	\$ 554,000.00

SON: QUINIENTOS CINCUENTA Y CUATRO MIL PESOS

RESPONSABLE :

RECIBO DE CAJA APLICADO AL EXPEDIENTE No. _____

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800176089-2

RECIBO OFICIAL DE CAJA : 10 - 102,474
FECHA : DICIEMBRE 1 DE 2010



***** CONSIGNACION *****

DEPOSITANTE	TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PAGO	Vr. PAGO
CASTILLO PINEDA	CONSIGNACION	BANCO DE BOGOTA	062754387	383679450	01/12/2010	10,128,000.00

***** CONCEPTO *****

CANT. RENTISTICO	CONCEPTO	TOTAL CONCEPTO
13 50005-01-01 SOLICITUDES	*** REIVINDICACION UNITARIA ADICIONAL	338,000.00
	TOTAL :	\$ 338,000.00

SON: TRESCIENTOS TREINTA Y OCHO MIL PESOS

RESPONSABLE : _____

RECIBO DE CAJA APLICADO AL EXPEDIENTE No. _____



No. 10-151612-00000-0000

Fecha: 2010-12-01 17:17:51
Tra. 11 PATENTEIYII
Act. 411 PRESENTACION

Dep. 2020 DIR.NUEVASCR
Eve: 378 FASENACIONALI
Folios: 190

PATENTE DE INVENCION ☐

MODELO DE U

- ☐ Indicación que se solicita una patente.
- ☐ Datos de identificación del solicitante o de la persona que presenta la solicitud
- ☐ Descripción de la invención
- ☐ Dibujos de ser estos pertinentes
- ☐ Comprobante de pago de las tasas establecidas (De ser el caso formato de descuento)
- Completa ☐ Incompleta ☐

PATENTE DE INVENCION PCT ☒

MODELO DE UTILIDAD PCT ☐

Art.33 Decisión 486/00, Circular Única

- ☒ Indicación que se solicita una PCT
- ☒ Copia de la solicitud en español, tal como fue presentada inicialmente (capítulo descriptivo, reivindicatorio, resumen)
- ☒ Dibujos de ser estos pertinentes
- ☒ Comprobante de pago de las tasas establecidas (de ser el caso formato de descuento)
- Completa ☒ Incompleta ☐

DISEÑO INDUSTRIAL ☐

(Art. 119 Decisión 486/00)

- ☐ Indicación que se solicita Diseño industrial
- ☐ Datos de identificación del solicitante o de la persona que presenta la solicitud
- ☐ Representación gráfica y fotográfica del Diseño industrial o muestra del material que incorpora el diseño
- ☐ Comprobante de pago de las tasas establecidas
- Completa ☐ Incompleta ☐

ESQUEMA DE TRAZADO ☐

(Art. 92 Decisión 486/00)

- ☐ Indicación que se solicita un esquema de trazado
- ☐ Datos de identificación del solicitante o de la persona que presenta la solicitud
- ☐ Representación gráfica de un esquema de trazado
- ☐ Comprobante de pago de las tasas establecidas
- Completa ☐ incompleta ☐



Bogotá D.C.,
2020

Doctor(a)
CASTILLO BONILLA JERONIMO
Apoderado(a) y/o Representante de
DANISCO US INC

REFERENCIA	Radicación No.	10 151612
	Solicitud internacional	pct/us2009/046159
	Fecha inicio fase nacional	06/01/2011
	Trámite	11
	Evento	378
	Actuación	430
	Oficio No.	0 22 03

Como resultado del examen de forma, realizado con fundamento en el artículo 38 de la Decisión 486 de la Comisión de la Comunidad Andina, artículo 27 del Tratado de Cooperación en Materia de Patentes (PCT), regla 51 bis del Reglamento y Circular Única de la Superintendencia de Industria y Comercio, se le comunica que:

1. La solicitud no contiene la copia del documento en que consta la cesión del derecho a la patente del inventor al solicitante o a su causante, conforme a lo establecido por el artículo 26 literal k) de la Decisión 486 de la Comisión de la Comunidad Andina. Cuando se trate de documentos otorgados en el extranjero, estos deberán legalizarse de conformidad con lo dispuesto por los artículos 259 y 260 del Código de Procedimiento Civil. Cuando se trate de las declaraciones a que hace referencia la Regla 4.17 ii) del Tratado de Cooperación en Materia de Patentes PCT., el alcance de las tales declaraciones se determinará según indiquen claramente una cesión de derechos del inventor al solicitante o su causante, de conformidad con lo dispuesto en el numeral 5.2.15 del Capítulo Quinto, Título X de la Circular Única de Superintendencia de Industria y Comercio.

En cumplimiento de lo dispuesto por el artículo 39 de la Decisión 486, se le requiere para que dentro del término de dos meses siguientes a la fecha de notificación del presente oficio, complete los requisitos anteriormente enunciados. En caso de no dar cumplimiento al presente requerimiento, la solicitud se considerará abandonada y perderá su prelación.

Notifíquese el presente oficio conforme a lo dispuesto en el numeral 5.2, literal e), del capítulo quinto del título primero de la Circular Única No.10 de 2001.

JOSE LUIS SALAZAR LOPEZ
Director de Nuevas Creaciones (E)

El anterior requerimiento fue notificado por fijación en lista de fecha 11 FEB. 2011
adpa11