



(86) **Date de dépôt PCT/PCT Filing Date:** 2009/11/23
(87) **Date publication PCT/PCT Publication Date:** 2010/05/27
(45) **Date de délivrance/Issue Date:** 2017/05/30
(85) **Entrée phase nationale/National Entry:** 2011/05/20
(86) **N° demande PCT/PCT Application No.:** US 2009/065571
(87) **N° publication PCT/PCT Publication No.:** 2010/060056
(30) **Priorité/Priority:** 2008/11/21 (US61/116,981)

(51) **Cl.Int./Int.Cl. C12N 1/19** (2006.01),
C12N 15/00 (2006.01), **C12N 15/56** (2006.01),
C12N 15/81 (2006.01), **C12P 19/14** (2006.01),
C12P 7/06 (2006.01), C12N 9/42 (2006.01)

(72) **Inventeurs/Inventors:**
MCBRIDE, JOHN, US;
BREVNOVA, ELENA, US;
GHANDI, CHHAYAL, US;
MELLON, MARK, US;

...

(73) **Propriétaires/Owners:**
UNIVERSITEIT STELLENBOSCH, ZA;

(54) **Titre : LEVURE EXPRIMANT DES CELLULASES POUR SACCHARIFICATION ET FERMENTATION SIMULTANÉES UTILISANT LA CELLULOSE**
(54) **Title: YEAST EXPRESSING CELLULASES FOR SIMULTANEOUS SACCHARIFICATION AND FERMENTATION USING CELLULOSE**

(57) **Abrégé/Abstract:**

Host cells, comprising *Kluveryomyces* expressing heterologous cellulases produce ethanol from cellulose. In addition, multiple host cells expressing different heterologous cellulases can be co-cultured together and used to produce ethanol from cellulose. The recombinant yeast strains and co-cultures of the yeast strains can be used to produce ethanol on their own, or can also be used in combination with externally added cellulases to increase the efficiency of saccharification and fermentation processes.



(72) **Inventeurs(suite)/Inventors(continued):** FROEHLICH, ALAN, US; DELEAULT, KRISTEN, US; RAJGARHIA, VINEET, US; FLATT, JIM, US; DEN HAAN, RIAAN, ZA; LAGRANGE, DANIE, ZA; ROSE, SHAUNITA, ZA; PENTTILA, MERJA, FI; ILMEN, MARJA, FI; SIIKA-AHO, MATTI, FI; UUSITALO, JAANA, FI; HAU, HEIDI HANSON, US; RICE, CHARLES, US; VILLARI, JEFF, US; STONEHOUSE, EMILY A., US; GILBERT, ALAN, US; KEATING, JEFFREY D., US; XU, HAOWEN, US; WILLES, DEIDRE, US; SHIKHARE, INDRANEEL, US; THORNGREN, NAOMI, US; WARNER, ANNE K., US; MURPHY, DAN, US; VAN ZYL, EMILE, ZA

(73) **Propriétaires(suite)/Owners(continued):** LALLEMAND HUNGARY LIQUIDITY MANAGEMENT LLC, HU

(74) **Agent:** GOWLING WLG (CANADA) LLP

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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
27 May 2010 (27.05.2010)(10) International Publication Number
WO 2010/060056 A3

(51) International Patent Classification:

C12N 15/09 (2006.01) *C12P 21/00* (2006.01)
C12N 15/81 (2006.01)

(21) International Application Number:

PCT/US2009/065571

(22) International Filing Date:

23 November 2009 (23.11.2009)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/116,981 21 November 2008 (21.11.2008) US

(71) Applicant (for all designated States except US): **MASCOMA CORPORATION** [US/US]; 67 Etna Road, Suite 300, Lebanon, New Hampshire 03766 (US).

(72) Inventors; and

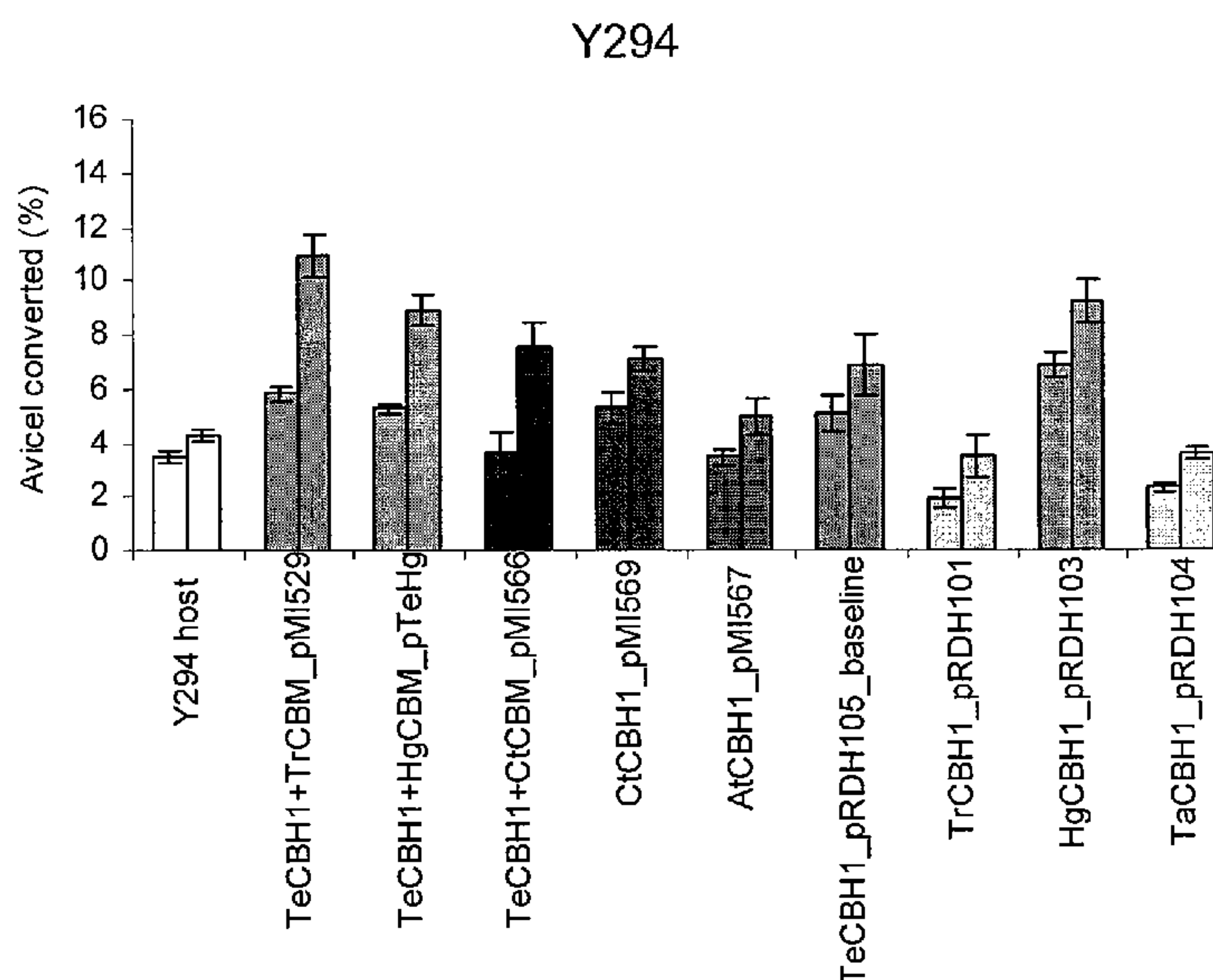
(75) Inventors/Applicants (for US only): **MCBRIDE, John** [US/US]; 174 Dartmouth College Highway, Lebanon, New Hampshire 03768 (US). **BREVENOVA, Elena** [RU/US]; 35 Moss Road, Lebanon, New Hampshire 03766 (US). **GHANDI, Chhayal** [IN/US]; 19 Spencer Street, Apt. #307, Lebanon, New Hampshire 03766 (US). **MELLON, Mark** [US/US]; P.O. Box 1791, Grantham, New Hampshire 03753 (US). **FROEHLICH, Alan** [US/US]; 5 Prospect Street, Lebanon, New Hampshire 03766 (US).

DELEAULT, Kristen [US/US]; 409 Fernwood Farms Road, Canaan, New Hampshire 03741 (US). **RAJGHARIA, Vineet** [IN/US]; 21 Spencer Street, Apt.# 201, Lebanon, New Hampshire 03766 (US). **FLATT, Jim** [US/US]; 1664 Oakmoor Heights, Colorado Springs, Colorado 80906 (US). **VAN ZYL, Emile** [ZA/ZA]; 12 Kaneel Crescent, Die Boord, 7600 Stellenbosch (ZA). **DEN HAAN, Riaan** [ZA/ZA]; 3 Gulden Street, 7550 Durbanville (ZA). **LAGRANGE, Danie** [ZA/ZA]; Monteverdi 6, Pascali Street, 7550 Durbanville (ZA). **ROSE, Shaunita** [ZA/ZA]; 6 Wade Street, 7140 Strand (ZA). **PENTTILA, Merja** [FI/FI]; Vattuniemenkatu 2 A 72, FI-00210 Helsinki (FI). **ILMEN, Marja** [FI/FI]; Sankariatie 7 A 9, FI-00320 Helsinki (FI). **SIIKA-AHO, Matti** [FI/FI]; Hiihtomaentie 16 E 49, FI-00810 Helsinki (FI). **UUSITALO, Jaana** [FI/FI]; Kaitaankulma 6 B 4, FI-02360 Espoo (FI). **HAU, Heidi Hanson** [US/US]; 21 Spencer Street, Apt. 218, Lebanon, New Hampshire 03766 (US). **RICE, Charles** [US/US]; 76 Schiller Street, Manchester, New Hampshire 03102 (US). **VILLARI, Jeff** [US/US]; 222 Pero Hill Road, Thetford Center, Vermont 05075 (US). **STONEHOUSE, Emily A.** [US/US]; 35 Parkhurst, Apt. 2, Lebanon, New Hampshire 03766 (US). **GILBERT, Alan** [US/US]; 21 Spencer Street, Apt. 405, Lebanon, New Hampshire 03766 (US). **KEATING, Jeffrey D.** [CA/US]; 417-19 Spencer Street, Lebanon, New Hampshire 03766 (US). **XU, Haowen** [CN/US]; 42 Wolf Road, Apt. 622, Lebanon, New Hampshire 03766

[Continued on next page]

(54) Title: YEAST EXPRESSING CELLULASES FOR SIMULTANEOUS SACCHARIFICATION AND FERMENTATION USING CELLULOSE

Figure 38A



(57) Abstract: Host cells, comprising Kluyveromyces expressing heterologous cellulases produce ethanol from cellulose. In addition, multiple host cells expressing different heterologous cellulases can be co-cultured together and used to produce ethanol from cellulose. The recombinant yeast strains and co-cultures of the yeast strains can be used to produce ethanol on their own, or can also be used in combination with externally added cellulases to increase the efficiency of saccharification and fermentation processes.

WO 2010/060056 A3

(US). **WILLES, Deidre** [US/US]; 153 Slayton Hill Road, Lebanon, New Hampshire 03766 (US). **SHIKHARE, Indraneel** [IN/US]; 7 Kimball Street, Apt. 16, Lebanon, New Hampshire 03766 (US). **THORNGREN, Naomi** [US/US]; 106 Ladeau Road, Norwich, Vermont 05055 (US). **WARNER, Anne K.** [US/US]; 42 Wolf Road, Unit 832, Lebanon, New Hampshire 03766 (US). **MURPHY, Dan** [US/US]; 30 Wolf Road, Unit 213, Lebanon, New Hampshire 03766 (US).

(74) **Agents:** **JACKMAN, Peter A.** et al.; Sterne, Kessler, Goldstein & Fox P.L.L.C., 1100 New York Avenue, N.W., Washington, District of Columbia 20005 (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, PE, 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, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description (Rules 13bis.4(d)(i) and 48.2(a)(viii))*
- *with sequence listing part of description (Rule 5.2(a))*

(88) **Date of publication of the international search report:**

15 July 2010

YEAST EXPRESSING CELLULASES FOR SIMULTANEOUS SACCHARIFICATION AND FERMENTATION USING CELLULOSE

Background of the Invention

[0001] Lignocellulosic biomass is widely recognized as a promising source of raw material for production of renewable fuels and chemicals. The primary obstacle impeding the more widespread production of energy from biomass feedstocks is the general absence of low-cost technology for overcoming the recalcitrance of these materials to conversion into useful fuels. Lignocellulosic biomass contains carbohydrate fractions (e.g., cellulose and hemicellulose) that can be converted into ethanol. In order to convert these fractions, the cellulose and hemicellulose must ultimately be converted or hydrolyzed into monosaccharides; it is the hydrolysis that has historically proven to be problematic.

[0002] Biologically mediated processes are promising for energy conversion, in particular for the conversion of lignocellulosic biomass into fuels. Biomass processing schemes involving enzymatic or microbial hydrolysis commonly involve four biologically mediated transformations: (1) the production of saccharolytic enzymes (cellulases and hemicellulases); (2) the hydrolysis of carbohydrate components present in pretreated biomass to sugars; (3) the fermentation of hexose sugars (e.g., glucose, mannose, and galactose); and (4) the fermentation of pentose sugars (e.g., xylose and arabinose). These four transformations occur in a single step in a process configuration called consolidated bioprocessing (CBP), which is distinguished from other less highly integrated configurations in that it does not involve a dedicated process step for cellulase and/or hemicellulase production.

[0003] CBP offers the potential for lower cost and higher efficiency than processes featuring dedicated cellulase production. The benefits result in part from avoided capital costs, substrate and other raw materials, and utilities associated with cellulase production. In addition, several factors support the realization of higher rates of hydrolysis, and hence reduced reactor volume and capital investment using CBP, including enzyme-microbe synergy and the use of thermophilic organisms and/or complexed cellulase systems. Moreover, cellulose-adherent cellulolytic microorganisms are likely to compete

successfully for products of cellulose hydrolysis with non-adhered microbes, e.g., contaminants, which could increase the stability of industrial processes based on microbial cellulose utilization. Progress in developing CBP-enabling microorganisms is being made through two strategies: engineering naturally occurring cellulolytic microorganisms to improve product-related properties, such as yield and titer, and engineering non-cellulolytic organisms that exhibit high product yields and titers to express a heterologous cellulase and hemicellulase system enabling cellulose and hemicellulose utilization.

[0004] Three major types of enzymatic activities are required for native cellulose degradation: The first type are endoglucanases (1,4- β -D-glucan 4-glucanohydrolases; EC 3.2.1.4). Endoglucanases cut at random in the cellulose polysaccharide chain of amorphous cellulose, generating oligosaccharides of varying lengths and consequently new chain ends. The second type are exoglucanases, including cellodextrinases (1,4- β -D-glucan glucanohydrolases; EC 3.2.1.74) and cellobiohydrolases (1,4- β -D-glucan cellobiohydrolases; EC 3.2.1.91). Exoglucanases act in a processive manner on the reducing or non-reducing ends of cellulose polysaccharide chains, liberating either glucose (glucanohydrolases) or cellobiose (cellobiohydrolase) as major products. Exoglucanases can also act on microcrystalline cellulose, presumably peeling cellulose chains from the microcrystalline structure. The third type are β -glucosidases (β -glucoside glucohydrolases; EC 3.2.1.21). β -Glucosidases hydrolyze soluble cellodextrins and cellobiose to glucose units.

[0005] Bakers' yeast (*Saccharomyces cerevisiae*) remains the preferred micro-organism for the production of ethanol (Hahn-Hägerdal, B., *et al.*, *Adv. Biochem. Eng. Biotechnol.* 73, 53–84 (2001)). Favorable attributes of this microbe include (i) high productivity at close to theoretical yields (0.51 g ethanol produced / g glucose used), (ii) high osmo- and ethanol tolerance, (iii) natural robustness in industrial processes, and (iv) being generally regarded as safe (GRAS) due to its long association with wine and bread making, and beer brewing. Furthermore, *S. cerevisiae* exhibits tolerance to inhibitors commonly found in hydrolyzates resulting from biomass pretreatment.

[0006] One major shortcoming of *S. cerevisiae* is its inability to utilize complex polysaccharides such as cellulose, or its break-down products, such as cellobiose and cellodextrins. In attempt to address this problem, several heterologous cellulases from

bacterial and fungal sources have been transferred to *S. cerevisiae*, enabling the degradation of cellulosic derivatives (Van Rensburg, P., *et al.*, *Yeast* 14, 67-76 (1998)), or growth on cellobiose (Van Rooyen, R., *et al.*, *J. Biotech.* 120, 284-295 (2005)); McBride, J.E., *et al.*, *Enzyme Microb. Techol.* 37, 93-101 (2005)). However, current levels of expression and specific activity of cellulases heterologously expressed in yeast are still not sufficient to enable efficient growth and ethanol production by yeast on cellulosic substrates without externally added enzymes. There remains a significant need for improvement in the amount of cellulase activity in order to attain the goal of achieving a consolidated bioprocessing (CBP) system capable of efficiently and cost-effectively converting cellulosic substrates to ethanol.

[0007] Another major shortcoming of the use of *S. cerevisiae* is that externally produced cellulases function optimally at a higher temperature than the temperature at which *S. cerevisiae* function optimally. Thus, either the processing must be carried out in a two step process at two different temperatures or one temperature can be selected where both processes function to some extent, but at least one of the processes does not occur at optimal efficiency.

[0008] In order to address these limitations, the present invention provides for heterologous expression of wild-type and codon-optimized combinations of heterologous cellulases in yeast that allows efficient production of ethanol from cellulose sources. The invention also provides for expression of such heterologous cellulases in thermotolerant yeast and methods of using such transformed yeast for ethanol production.

Brief Description of the Invention

[0009] The present invention is directed to cellulytic host cells. The host cells of the invention expressing heterologous cellulases and are able to produce ethanol from cellulose.

[0010] In particular, in some embodiments, the invention provides a transformed thermotolerant yeast host cell comprising at least one heterologous polynucleotide comprising a nucleic acid encoding a cellulase, wherein the yeast host cell is capable of producing ethanol when grown using cellulose as a carbon source.

[0011] In another embodiment, the invention provides a transformed thermotolerant yeast host cell comprising: (a) at least one heterologous polynucleotide comprising a nucleic

acid which encodes an endoglucanase; (b) at least one heterologous polynucleotide comprising a nucleic acid which encodes a β -glucosidase; (c) at least one heterologous polynucleotide comprising a nucleic acid which encodes a first cellobiohydrolase; and (d) at least one heterologous polynucleotide comprising a nucleic acid which encodes a second cellobiohydrolase.

[0012] In another embodiment, the invention provides a transformed yeast host cell comprising: (a) at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellulase which is an endoglucanase; (b) at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellulase which is a β -glucosidase; (c) at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellulase which is a first cellobiohydrolase; and (d) at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellulase which is a second cellobiohydrolase, wherein at least two of the cellulases are secreted by the cell.

[0013] In yet another embodiment, the invention provides a transformed yeast host cell comprising at least six heterologous polynucleotides, wherein each heterologous polynucleotide comprises a nucleic acid which encodes a cellulase.

[0014] In yet another embodiment, the invention provides a transformed yeast host cell comprising at least four heterologous polynucleotides, wherein each heterologous polynucleotide comprises a nucleic acid which encodes an endoglucanase.

[0015] In still another embodiment, the invention provides a co-culture comprising at least two yeast host cells wherein (a) at least one of the host cells comprises a first heterologous polynucleotide comprising a nucleic acid which encodes a cellulase which is an endoglucanase; (b) at least one of the host cells comprises a second heterologous polynucleotide comprising a nucleic acid which encodes a cellulase which is a β -glucosidase; (c) at least one of the host cells comprises a third heterologous polynucleotide comprising a nucleic acid which encodes a cellulase which is a first cellobiohydrolase; (d) at least one of the host cells comprises a fourth heterologous polynucleotide comprising a nucleic acid which encodes a cellulase which is a second cellobiohydrolase; wherein the first polynucleotide, the second polynucleotide, the third polynucleotide and the fourth polynucleotide are not in the same host cell; and wherein the co-culture is capable of producing ethanol from cellulose.

- [0016] In some particular embodiments of the invention, the cellulose carbon source is insoluble cellulose, crystalline cellulose, cellulose derived from lignocellulose, hardwood, phosphoric acid swollen cellulose or microcrystalline cellulose.
- [0017] In some embodiments, the host cells of the invention comprise a heterologous polynucleotide comprising a nucleic acid encoding a first cellobiohydrolase, a polynucleotide comprising a nucleic acid encoding an endoglucanase, a polynucleotide comprising a nucleic acid encoding a β -glucosidase and/or a polynucleotide comprising a nucleic acid encoding a second cellobiohydrolase.
- [0018] In some embodiments, the cellulase, endoglucanase, β -glucosidase or cellobiohydrolase is a *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, or *Arabidopsis thaliana* cellulase, endoglucanase, β -glucosidase or cellobiohydrolase.
- [0019] In some particular embodiments, the cellobiohydrolase is an *H. grisea* CBH1, a *T. aurantiacus* CBH1, a *T. emersonii* CBH1, a *T. reesei* CBH1, a *T. emersonii* CBH2, a *C. lucknowense* CBH2 or a *T. reesei* CBH2. In some embodiments, the heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase, encodes a fusion protein comprising a cellobiohydrolase and a cellulose binding module (CBM). In some particular embodiments, the CBM is the CBM of *T. reesei* CBH2, the CBM of *T. reesei* CBH1 or the CBM of *C. lucknowense* CBH2b. In some particular embodiments, the CBM is fused to the cellobiohydrolase via a linker sequence. In some particular embodiments, the host cell expresses a first and a second cellobiohydrolase, wherein the first cellobiohydrolase is a *T. emersonii* CBH1 and CBD fusion, and the second cellobiohydrolase is a *C. lucknowense* CBH2b.
- [0020] In other particular embodiments, the β -glucosidase is a *S. fibuligera* β -glucosidase. In another particular embodiment, the endoglucanase is a *C. formosanus*

endoglucanase. In another particular embodiment, the endoglucanase is a *T. reesei* endoglucanase, e.g. *T. reesei* EG2.

[0021] In some embodiments of the invention, at least one or at least two of the cellulases is tethered. In other embodiments of the invention, at least one of the cellulases is secreted. In another embodiment, at least one of the cellulases is tethered and at least one of the cellulases is secreted. In another embodiment, all of the cellulases are secreted.

[0022] In some embodiments of the invention, the nucleic acid encoding a cellulase is codon optimized.

[0023] In some embodiments, the host cell can be a thermotolerant host cell. In some embodiments, the host cell is a *Issatchenkia orientalis*, *Pichia mississippiensis*, *Pichia mexicana*, *Pichia farinosa*, *Clavispora opuntiae*, *Clavispora lusitaniae*, *Candida mexicana*, *Hansenula polymorpha* or *Kluveryomyces* host cell. For example, in some embodiments, the host cell is a *K. lactis* or *K. marxianus* host cell. In some embodiments, the thermotolerant host cell is an *S. cerevisiae* host cell, wherein the *S. cerevisiae* is selected to be thermotolerant.

[0024] In some embodiments, the host cell can be an oleaginous yeast cell. In some particular embodiments, the oleaginous yeast cell is a *Blakeslea*, *Candida*, *Cryptococcus*, *Cunninghamella*, *Lipomyces*, *Mortierella*, *Mucor*, *Phycomces*, *Pythium*, *Rhodospiridium*, *Rhodotorula*, *Trichosporon* or *Yarrowia cell*.

[0025] In some embodiments, the host cell is a *Saccharomyces cerevisiae* cell.

[0026] In some particular embodiments, the host cell can produce ethanol from cellulose at temperatures above about 30° C, 35 °C, 37° C, 42° C, 45° C or 50° C.

[0027] In another particular embodiment, the host cell can produce ethanol at a rate of at least about 10 mg per hour per liter, at least about 30 mg per hour per liter, at least about 40 mg per hour per liter, at least about 50 mg per hour per liter, at least about 60 mg per hour per liter, at least about 70 mg per hour per liter, at least about 80 mg per hour per liter, at least about 90 mg per hour per liter, at least about 100 mg per hour per liter, at least about 200 mg per hour per liter, at least about 300 mg per hour per liter, at least about 400 mg per hour per liter, at least about 500 mg per hour per liter, at least about 600 mg per hour per liter, at least about 700 mg per hour per liter, at least about 800 mg per hour per liter, at least about 900 mg per hour per liter, or at least about 1 g per hour per liter.

[0028] The present invention also provides methods of using the host cells and co-cultures of the invention. For example, the present invention is also directed to a method for hydrolyzing a cellulosic substrate, comprising contacting said cellulosic substrate with a host cell or co-culture of the invention. The invention is also directed to a method of fermenting cellulose comprising culturing a host cell or co-culture of the invention in medium that contains insoluble cellulose under suitable conditions for a period sufficient to allow saccharification and fermentation of the cellulose. In some particular embodiments, the methods further comprise contacting the cellulosic substrate with externally produced cellulase enzymes.

[0029] In some particular methods of the invention, the cellulosic substrate is a lignocellulosic biomass selected from the group consisting of grass, switch grass, cord grass, rye grass, reed canary grass, miscanthus, sugar-processing residues, sugarcane bagasse, agricultural wastes, rice straw, rice hulls, barley straw, corn cobs, cereal straw, wheat straw, canola straw, oat straw, oat hulls, corn fiber, stover, soybean stover, corn stover, forestry wastes, recycled wood pulp fiber, paper sludge, sawdust, hardwood, softwood, Agave, and combinations thereof.

[0030] In some particular methods of the invention, the host cell or co-culture produces ethanol. The ethanol can be produced at a rate of at least about 10 mg per hour per liter, at least about 30 mg per hour per liter, at least about 40 mg per hour per liter, at least about 50 mg per hour per liter, at least about 60 mg per hour per liter, at least about 70 mg per hour per liter, at least about 80 mg per hour per liter, at least about 90 mg per hour per liter, at least about 100 mg per hour per liter, at least about 200 mg per hour per liter, at least about 300 mg per hour per liter, at least about 400 mg per hour per liter, at least about 500 mg per hour per liter, at least about 600 mg per hour per liter, at least about 700 mg per hour per liter, at least about 800 mg per hour per liter, at least about 900 mg per hour per liter, or at least about 1 g per hour per liter.

[0031] In other particular methods of the invention, the host cell or co-cultures contact a cellulosic substance at a temperature of at least about 37 °C, least about 42 °C, from about 42 °C to about 45 °C, or from about 42 °C to about 50 °C.

- 7a -

Various embodiments of the present invention relate to a thermotolerant yeast host cell comprising: at least one heterologous polynucleotide comprising a nucleic acid which encodes an endoglucanase; at least one heterologous polynucleotide comprising a nucleic acid which encodes a β -glucosidase; at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase I; and at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase II; wherein the thermotolerant yeast host cell grows at a temperature of at least 35°C or above, wherein the endoglucanase, β -glucosidase, cellobiohydrolase I, and cellobiohydrolase II are secreted; and wherein the host cell is capable of producing ethanol from AvicelTM. Various embodiments relate to a method of fermenting cellulose comprising culturing the host cell in medium that contains cellulose to allow saccharification and fermentation of the cellulose. Various embodiments relate to a culture comprising the host cell and cellulose.

Various embodiments of the present invention relate to a thermotolerant yeast host cell comprising: (a) at least one heterologous polynucleotide comprising a nucleic acid which encodes an endoglucanase; (b) at least one heterologous polynucleotide comprising a nucleic acid which encodes a β -glucosidase; (c) at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase I; and (d) at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase II; wherein the thermotolerant yeast host cell grows at a temperature of at least 35°C or above, wherein the cellobiohydrolase I, the cellobiohydrolase II, or both is fused to a cellulose binding module (CBM); and wherein the host cell is capable of producing ethanol from AvicelTM. Various embodiments relate to a method of fermenting cellulose comprising culturing the host cell in medium that contains cellulose to allow saccharification and fermentation of the cellulose. Various embodiments relate to a culture comprising the host cell and cellulose.

Brief Description of the Drawings

- [0032] Figure 1 shows an image of a CMC plate assay to detect endoglucanase I activity in *K. lactis* (colonies numbered 1-8) and *K. marxianus* strains (colonies numbered 9-16) transformed with heterologous cellulases. Strains 8 and 16 are untransformed negative controls. The plate on the left shows colony growth, and the plate on the right shows CMCase activity, indicated by the presence of a clearance zone. Clearance zones appear as white spots in the image.
- [0033] Figure 2 depicts the results of an MU-lac assay to detect CBH1 activity in *K. marxianus* strains transformed with heterologous cellulases.
- [0034] Figure 3 depicts the percent of Avicel™ converted by several strains of *K. marxianus* expressing heterologous cellulases.
- [0035] Figure 4 depicts the ethanol production/consumption from Avicel™ by several strains of *K. marxianus* expressing heterologous cellulases.
- [0036] Figure 5 depicts the growth of *S. cerevisiae* expressing heterologous cellulases on bacterial microcrystalline cellulose (BMCC).
- [0037] Figure 6 depicts the ethanol production from Avicel™ by an *S. cerevisiae* strain expressing heterologous cellulases.
- [0038] Figure 7 depicts the ethanol production from pretreated hardwood (5% based on a dry weight percentage) by an *S. cerevisiae* strain expressing heterologous cellulases.
- [0039] Figure 8 depicts the ethanol production from pretreated hardwood (5% based on a dry weight percentage) by an *S. cerevisiae* expressing heterologous cellulases in the presence of various concentrations of exogenously added cellulases.
- [0040] Figure 9 depicts the ethanol production from Avicel™ by MO288 (circles) and a control strain (triangles) in both YP media and YNB media.
- [0041] Figure 10 depicts the ethanol yield from Avicel™ (15% based on a dry weight percentage) by a small scale simultaneous saccharification and fermentation (SSF) process using *S. cerevisiae* supplemented with external cellulases. The yield from a yeast strain expressing heterologous cellulases (MO288) is compared to the yield from a control strain (MO249) at a variety of external cellulase concentrations over 150 hours. (100% cellulase loading indicates 25 mg/g total solids; initial solids concentration was 15%.)

- [0042] Figure 11 depicts the theoretical ethanol yield from a simultaneous saccharification and fermentation (SSF) process using *S. cerevisiae* supplemented with external cellulases. The yield from a yeast strain expressing heterologous cellulases (MO288) is compared to the yield from a control strain (MO249).
- [0043] Figure 12 illustrates the predicted cellulase enzyme savings based on ethanol yield at 168 hours of simultaneous saccharification and fermentation (SSF) process.
- [0044] Figure 13 shows the activity of an artificial cellulase in the Avicel conversion assay as described in Example 9. The MO429 strain was transformed the CBH1 consensus sequence "CBH1cons," and the MO419 strain was transformed with empty pMU451 vector as a negative control. Descriptions of other strains are found in Table 8 of Example 9.
- [0045] Figure 14 demonstrates the activity of yeast expressing various combinations of CBH1 and CBH2 enzymes on Avicel as described in Example 10.
- [0046] Figure 15 demonstrates the activity of yeast expressing various cellulase enzymes on Avicel as described in Example 10.
- [0047] Figure 16 depicts the ethanol production from Avicel by a co-culture of five *S. cerevisiae* strains expressing heterologous cellulases.
- [0048] Figure 17 depicts the ethanol production from Avicel by a co-culture of four *S. cerevisiae* strains expressing heterologous cellulases as well as the ethanol production from strain MO288, which is expressing four cellulases.
- [0049] Figure 18 depicts the ethanol production from Avicel by a co-culture of four *S. cerevisiae* strains expressing heterologous cellulases in combination with externally added cellulase.
- [0050] Figure 19 depicts the calculated enzyme savings using a co-culture of four *S. cerevisiae* strains expressing heterologous cellulases or MO288 as compared to untransformed *S. cerevisiae*.
- [0051] Figure 20 depicts the xylose utilization and ethanol production of M0509 freezer stock, YPX-isolate and YPD-isolate.
- [0052] Figure 21 depicts the growth of M1105 (labeled "colony C2") and MO1046 in the presence of the same medium and 8 g/L acetate at 40 °C.
- [0053] Figure 22 depicts the ethanol production by M1105 (triangles) and M1088 (squares) on 18% TS MS419. The experiment with M1105 had 10% lower enzyme dose

and half the inoculated cell density, but produced a higher ethanol titer. The experiment with MO1105 was performed at 40° C, and the experiment with M1088 was performed at 35°C.

- [0054] Figure 23 depicts the ethanol production of M1105 where the fermentation was only inoculated with 0.15 g/L DCW and resulted in some sugar accumulation and 29 g/L ethanol.
- [0055] Figure 24 depicts the ethanol production of M1254 in standard IFM (circles) and low ammonium IFM (squares) conditions.
- [0056] Figure 25 depicts the specific growth rate of single colonies compared to M1254 and M1339 on complex xylose medium supplemented with a synthetic inhibitor mixture (which included 8 g/L acetate) at 40° C. The single colonies were screened at the same conditions as the evolution occurred. Colony C1 was renamed M1360.
- [0057] Figure 26 depicts the fermentation performance of M1360 at 40° C on industrially relevant fermentation medium supplemented with glucose. The fermentation was inoculated with 60 mg/L dry cell weight of M1360.
- [0058] Figure 27 depicts the ethanol production in SSF runs on PHW (18% solids, unwashed MS149) at 35° C and 40° C by several strains. All reactions were loaded with 4mg/g “zoomerose” (Novozyme 22c).
- [0059] Figure 28 depicts cultures spotted on SC^{-URA} plates containing 0.2% of either CMC or lichenin or barley-β-glucan. The top two rows of each plate were Y294 based cultures, and the bottom two rows contained MO749 based strains. Numbers indicate the plasmid contained by each strain. pMU471 contains the *C.f.*EG and served as positive control. Plates were incubated for 24 hours at 30° C (pictured on the left), after which colonies were washed off and the plates were stained with 0.1% congo red and destained with 1% NaCl (pictured on the right).
- [0060] Figure 29 depicts SDS-PAGE analysis of the supernatants of Cel5 cellulase producing strains. A strain containing a plasmid with no foreign gene was used as reference strain (REF). The strain containing the plasmid pMU471 expressing *C.f.*EG, the most successful EG previously found was also included.
- [0061] Figure 30 depicts the activity of strains expressing EGs on (A) PASC (2 hours) and (B) avicel (24 hours). A strain containing a plasmid with no foreign gene was used as

reference strain (REF) and the strain expressing *C.f.*EG (pMU471) was included as positive control.

[0062] Figure 31 depicts the distribution of avicel conversion ability of yeast supernatants from transformation with *Tr*EG2 and additional *Te*CBH1w/*Tr*CBD. M1088 conversion is presented as a dark vertical line, and the dotted lines flanking this line represent the standard deviation of the measurement.

[0063] Figure 32 depicts the conversion of Avicel in the HTP avicel assay (48 hour time point) by supernatants of cellulase expressing yeast strains. M0509 is the negative control expressing no cellulases. Strain 1088 is the parental strain expressing only CBH1, CBH2, and BGL, whereas 1179, 1180, and 1181 are transformants of 1088 also expressing *Tr*EG2.

[0064] Figure 33 depicts ethanol production in paper sludge CBP/SSF with cellulolytic strain M1403 and non-cellulolytic background strain M1254 with various amounts of commercial enzyme supplementation. Experimental conditions: 30% solids fed batch, 10g/l cell inoculation, pH 5.5 and temperature 40° C, Zoom = Novozymes 22C cellulase preparation, BGL = AB Enzymes EL2008044L BGL preparation, Xyl = AB Enzymes EL2007020L xylanase preparation.

[0065] Figure 34 depicts fermentation of two types of paper sludge by CBP yeast (M1179) and a control strain M0509, not expressing cellulases. Experimental conditions: 18% solids, cells loaded at 10 or 1 g/L, pH 5.5, Temp: 35C, 1 mg/g BGL and 1mg/g Xyl loaded. BGL = AB Enzymes EL2008044L BGL preparation, Xyl = AB Enzymes EL2007020L xylanase preparation.

[0066] Figure 35 depicts the performance of cellulolytic yeast strain M0963 and non-cellulolytic control strain (M0509) on 22% unwashed solids of pretreated hardwood (PHW) (MS149) at various external cellulase concentrations. Experimental conditions: 22% solids fed batch, pH 5.4, temperature 35°C, all enzyme protein (EP) was “zoomerose” (Novozymes 22C).

[0067] Figure 36 depicts the performance of cellulolytic yeast strain M1284 on 30% solids of washed pretreated hardwood at various initial cell loadings. Experimental conditions: 30% solids fed batch, pH 5.0, temperature 35°C, 4mg EP = 0.25mg BGL + 0.25mg Xylanase + 0.25mg Pectinase + 3.25 mg Zoomerose, 20mg EP = 1mg BGL + 1mg Xylanase + 1mg Pectinase + 16.7mg Zoomerose. Zoomerose = Novozymes 22C cellulase

preparation, BGL = AB Enzymes EL2008044L BGL preparation, Xyl = AB Enzymes EL2007020L xylanase preparation, Pectinase = Genencor Multifect pectinase FE.

[0068] Figure 37 depicts the ethanol production in washed corn stover CBP/SSF with cellulolytic strain M1284 and non-cellulolytic background strain M0509 with various amounts of commercial enzyme supplementation. Experimental conditions: 18% solids fed batch, 10g/l cell inoculation, pH 5.0 and temperature 35°C, 1 mg/g BGL and 1 mg/g xylanase loaded in each case. BGL = AB Enzymes EL2008044L BGL preparation, Xyl = AB Enzymes EL2007020L xylanase preparation.

[0069] Figure 38 depicts the activity on Avicel (A, B) or MULac (C, D) of yeast culture supernatants expressing different CBH1 genes, and estimated CBH1 concentration (mg/L, E, F) based on MULac. The host strain was either Y294 or M0749. The CBH1 genes are: *Te*, *Talaromyces emersonii*; *Ct*, *Chaetomium thermophilum*; *At*, *Acremonium thermophilum*; *Tr*, *Trichoderma reesei*; *Hg*, *Humicola grisea*; *Ta*, *Thermoascus aurantiacus*. The plasmid names are indicated. Yeast were cultivated in YPD in triplicate for 3 days. The data are means $\pm$ standard deviation.

[0070] Figure 39 shows the genes modified in yeast strain M0509.

[0071] Figure 40 shows the yeast strains used to construct M0509 and the relevant genetic modifications.

[0072] Figure 41 shows the genealogy of yeast strain M1105.

[0073] Figure 42 shows the genealogy of yeast strain M1254.

Detailed Description of the Invention

[0074] The disclosed methods and materials are useful generally in the field of engineered yeast.

Definitions

[0075] A "vector," *e.g.*, a "plasmid" or "YAC" (yeast artificial chromosome) refers to an extrachromosomal element often carrying one or more genes that are not part of the central metabolism of the cell, and is usually in the form of a circular double-stranded DNA molecule. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear, circular, or supercoiled, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is

capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. Preferably, the plasmids or vectors of the present invention are stable and self-replicating.

[0076] An "expression vector" is a vector that is capable of directing the expression of genes to which it is operably associated.

[0077] The term "heterologous" as used herein refers to an element of a vector, plasmid or host cell that is derived from a source other than the endogenous source. Thus, for example, a heterologous sequence could be a sequence that is derived from a different gene or plasmid from the same host, from a different strain of host cell, or from an organism of a different taxonomic group (*e.g.*, different kingdom, phylum, class, order, family genus, or species, or any subgroup within one of these classifications). The term "heterologous" is also used synonymously herein with the term "exogenous."

[0078] The term "domain" as used herein refers to a part of a molecule or structure that shares common physical or chemical features, for example hydrophobic, polar, globular, helical domains or properties, *e.g.*, a DNA binding domain or an ATP binding domain. Domains can be identified by their homology to conserved structural or functional motifs. Examples of cellobiohydrolase (CBH) domains include the catalytic domain (CD) and the cellulose binding domain (CBD).

[0079] A "nucleic acid," "polynucleotide," or "nucleic acid molecule" is a polymeric compound comprised of covalently linked subunits called nucleotides. Nucleic acid includes polyribonucleic acid (RNA) and polydeoxyribonucleic acid (DNA), both of which may be single-stranded or double-stranded. DNA includes cDNA, genomic DNA, synthetic DNA, and semi-synthetic DNA.

[0080] An "isolated nucleic acid molecule" or "isolated nucleic acid fragment" refers to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, inter

alia, in linear or circular DNA molecules (e.g., restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA).

[0081] A "gene" refers to an assembly of nucleotides that encode a polypeptide, and includes cDNA and genomic DNA nucleic acids. "Gene" also refers to a nucleic acid fragment that expresses a specific protein, including intervening sequences (introns) between individual coding segments (exons), as well as regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences.

[0082] A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified, *e.g.*, in Sambrook, J., Fritsch, E. F. and Maniatis, T. MOLECULAR CLONING: A LABORATORY MANUAL, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein (hereinafter "Maniatis", entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. For more stringent conditions, washes are performed at higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS are increased to 60°C. Another set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C. An additional set of highly stringent conditions are

defined by hybridization at 0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS.

[0083] Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T_m have been derived (*see, e.g.*, Maniatis at 9.50-9.51). For hybridizations with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (*see, e.g.*, Maniatis, at 11.7-11.8). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferably a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

[0084] The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences.

[0085] As known in the art, "similarity" between two polypeptides is determined by comparing the amino acid sequence and conserved amino acid substitutes thereto of the polypeptide to the sequence of a second polypeptide.

[0086] "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, NY (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, NY (1993); Computer Analysis of Sequence Data,

Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, NJ (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, NY (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, Wis.). Multiple alignments of the sequences disclosed herein were performed using the Clustal method of alignment (Higgins and Sharp (1989) CABIOS. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

[0087] Suitable nucleic acid sequences or fragments thereof (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% to 75% identical to the amino acid sequences reported herein, at least about 80%, 85%, or 90% identical to the amino acid sequences reported herein, or at least about 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments are at least about 70%, 75%, or 80% identical to the nucleic acid sequences reported herein, at least about 80%, 85%, or 90% identical to the nucleic acid sequences reported herein, or at least about 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleic acid sequences reported herein. Suitable nucleic acid fragments not only have the above identities/similarities but typically encode a polypeptide having at least 50 amino acids, at least 100 amino acids, at least 150 amino acids, at least 200 amino acids, or at least 250 amino acids.

[0088] A DNA or RNA "coding region" is a DNA or RNA molecule which is transcribed and/or translated into a polypeptide in a cell *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. "Suitable regulatory regions" refer to nucleic acid regions located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding region, and which influence the transcription, RNA processing or stability, or translation of the associated coding region. Regulatory regions may include promoters, translation leader sequences, RNA processing site, effector

binding site and stem-loop structure. The boundaries of the coding region are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding region can include, but is not limited to, prokaryotic regions, cDNA from mRNA, genomic DNA molecules, synthetic DNA molecules, or RNA molecules. If the coding region is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding region.

[0089] An "isoform" is a protein that has the same function as another protein but which is encoded by a different gene and may have small differences in its sequence.

[0090] A "paralogue" is a protein encoded by a gene related by duplication within a genome.

[0091] An "orthologue" is gene from a different species that has evolved from a common ancestral gene by speciation. Normally, orthologues retain the same function in the course of evolution as the ancestral gene.

[0092] "Open reading frame" is abbreviated ORF and means a length of nucleic acid, either DNA, cDNA or RNA, that comprises a translation start signal or initiation codon, such as an ATG or AUG, and a termination codon and can be potentially translated into a polypeptide sequence.

[0093] "Promoter" refers to a DNA fragment capable of controlling the expression of a coding sequence or functional RNA. In general, a coding region is located 3' to a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity. A promoter is generally bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above

background. Within the promoter will be found a transcription initiation site (conveniently defined for example, by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

[0094] A coding region is "under the control" of transcriptional and translational control elements in a cell when RNA polymerase transcribes the coding region into mRNA, which is then trans-RNA spliced (if the coding region contains introns) and translated into the protein encoded by the coding region.

[0095] "Transcriptional and translational control regions" are DNA regulatory regions, such as promoters, enhancers, terminators, and the like, that provide for the expression of a coding region in a host cell. In eukaryotic cells, polyadenylation signals are control regions.

[0096] The term "operably associated" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably associated with a coding region when it is capable of affecting the expression of that coding region (*i.e.*, that the coding region is under the transcriptional control of the promoter). Coding regions can be operably associated to regulatory regions in sense or antisense orientation.

[0097] The term "expression," as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

Host Cells Expressing Heterologous Cellulases

[0098] In order to address the limitations of the previous systems, the present invention provides host cells expressing heterologous cellulases that can be effectively and efficiently utilized to produce ethanol from cellulose. In some embodiments, the host cells can be a yeast. According to the present invention the yeast host cell can be, for example, from the genera *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*. Yeast species as host cells may include, for example, *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus*, or *K. fragilis*. In some embodiments, the yeast is selected from the group consisting of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Candida albicans*, *Pichia pastoris*, *Pichia stipitis*, *Yarrowia lipolytica*, *Hansenula polymorpha*, *Phaffia rhodozyma*, *Candida utilis*,

Arxula adenivorans, *Debaryomyces hansenii*, *Debaryomyces polymorphus*, *Schizosaccharomyces pombe* and *Schwanniomyces occidentalis*. In one particular embodiment, the yeast is *Saccharomyces cerevisiae*. In another embodiment, the yeast is a thermotolerant *Saccharomyces cerevisiae*. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

[0099] In some embodiments of the present invention, the host cell is an oleaginous cell. According to the present invention, the oleaginous host cell can be an oleaginous yeast cell. For example, the oleaginous yeast host cell can be from the genera *Blakeslea*, *Candida*, *Cryptococcus*, *Cunninghamella*, *Lipomyces*, *Mortierella*, *Mucor*, *Phycomyces*, *Pythium*, *Rhodosporidium*, *Rhodotorula*, *Trichosporon* or *Yarrowia*. According to the present invention, the oleaginous host cell can be an oleaginous microalgae host cell. For example, the oleaginous microalgae host cell can be from the genera *Thraustochytrium* or *Schizochytrium*. Biodiesel could then be produced from the triglyceride produced by the oleaginous organisms using conventional lipid transesterification processes. In some particular embodiments, the oleaginous host cells can be induced to secrete synthesized lipids. Embodiments using oleaginous host cells are advantageous because they can produce biodiesel from lignocellulosic feedstocks which, relative to oilseed substrates, are cheaper, can be grown more densely, show lower life cycle carbon dioxide emissions, and can be cultivated on marginal lands.

[0100] In some embodiments of the present invention, the host cell is a thermotolerant host cell. Thermotolerant host cells can be particularly useful in simultaneous saccharification and fermentation processes by allowing externally produced cellulases and ethanol-producing host cells to perform optimally in similar temperature ranges.

[0101] Thermotolerant host cells of the invention can include, for example, *Issatchenkia orientalis*, *Pichia mississippiensis*, *Pichia mexicana*, *Pichia farinosa*, *Clavispora opuntiae*, *Clavispora lusitaniae*, *Candida mexicana*, *Hansenula polymorpha* and *Kluyveromyces* host cells. In some embodiments, the thermotolerant cell is an *S. cerevisiae* strain, or other yeast strain, that has been adapted to grow in high temperatures, for example, by selection for growth at high temperatures in a cytostat.

[0102] In some particular embodiments of the present invention, the host cell is a *Kluyveromyces* host cell. For example, the *Kluyveromyces* host cell can be a *K. lactis*, *K. marxianus*, *K. blattae*, *K. phaffii*, *K. yarrowii*, *K. aestuarii*, *K. dobzhanskii*, *K.*

wickerhamii *K. thermotolerans*, or *K. waltii* host cell. In one embodiment, the host cell is a *K. lactis*, or *K. marxianus* host cell. In another embodiment, the host cell is a *K. marxianus* host cell.

[0103] In some embodiments of the present invention the thermotolerant host cell can grow at temperatures above about 30° C, about 31° C, about 32° C, about 33° C, about 34° C, about 35° C, about 36° C, about 37° C, about 38° C, about 39° C, about 40° C, about 41° C or about 42° C. In some embodiments of the present invention the thermotolerant host cell can produce ethanol from cellulose at temperatures above about 30° C, about 31° C, about 32° C, about 33° C, about 34° C, about 35° C, about 36° C, about 37° C, about 38° C, about 39° C, about 40° C, about 41° C, about 42° C, or about 43 °C, or about 44 °C, or about 45 °C, or about 50° C.

[0104] In some embodiments of the present invention, the thermotolerant host cell can grow at temperatures from about 30° C to 60° C, about 30° C to 55° C, about 30° C to 50° C, about 40° C to 60° C, about 40° C to 55° C or about 40° C to 50° C. In some embodiments of the present invention, the thermotolerant host cell can produce ethanol from cellulose at temperatures from about 30° C to 60° C, about 30° C to 55° C, about 30° C to 50° C, about 40° C to 60° C, about 40° C to 55° C or about 40° C to 50° C.

[0105] In some methods described herein, the host cell has the ability to metabolize xylose. Detailed information regarding the development of the xylose-utilizing technology can be found in the following publications: Kuyper M *et al. FEMS Yeast Res.* 4: 655-64 (2004), Kuyper M *et al. FEMS Yeast Res.* 5:399-409 (2005), and Kuyper M *et al. FEMS Yeast Res.* 5:925-34 (2005), which are herein incorporated by reference in their entirety. For example, xylose-utilization can be accomplished in *S. cerevisiae* by heterologously expressing the xylose isomerase gene, *XylA*, *e.g.* from the anaerobic fungus *Piromyces sp. E2*, overexpressing five *S. cerevisiae* enzymes involved in the conversion of xylulose to glycolytic intermediates (xylulokinase, ribulose 5-phosphate isomerase, ribulose 5-phosphate epimerase, transketolase and transaldolase) and deleting the GRE3 gene encoding aldose reductase to minimise xylitol production.

[0106] According to the methods described herein, the host cells can contain antibiotic markers or can contain no antibiotic markers.

[0107] Host cells are genetically engineered (transduced or transformed or transfected) with the polynucleotides encoding cellulases of this invention which are described in

more detail below. The polynucleotides encoding cellulases can be introduced to the host cell on a vector of the invention, which may be, for example, a cloning vector or an expression vector comprising a sequence encoding a heterologous cellulase. The host cells can comprise polynucleotides of the invention as integrated copies or plasmid copies.

[0108] In certain aspects, the present invention relates to host cells containing the polynucleotide constructs described below. The host cells of the present invention can express one or more heterologous cellulase polypeptides. In some embodiments, the host cell comprises a combination of polynucleotides that encode heterologous cellulases or fragments, variants or derivatives thereof. The host cell can, for example, comprise multiple copies of the same nucleic acid sequence, for example, to increase expression levels, or the host cell can comprise a combination of unique polynucleotides. In other embodiments, the host cell comprises a single polynucleotide that encodes a heterologous cellulase or a fragment, variant or derivative thereof. In particular, such host cells expressing a single heterologous cellulase can be used in co-culture with other host cells of the invention comprising a polynucleotide that encodes at least one other heterologous cellulase or fragment, variant or derivative thereof.

[0109] Introduction of a polynucleotide encoding a heterologous cellulase into a host cell can be done by methods known in the art. Introduction of polynucleotides encoding heterologous cellulases into, for example yeast host cells, can be effected by lithium acetate transformation, spheroplast transformation, or transformation by electroporation, as described in Current Protocols in Molecular Biology, 13.7.1-13.7.10. Introduction of the construct in other host cells can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation. (Davis, L., *et al.*, Basic Methods in Molecular Biology, (1986)).

[0110] The transformed host cells or cell cultures, as described above, can be examined for endoglucanase, cellobiohydrolase and/or β glucosidase protein content. For the use of secreted heterologous cellulases, protein content can be determined by analyzing the host (*e.g.*, yeast) cell supernatants. In certain embodiments, high molecular weight material can be recovered from the yeast cell supernatant either by acetone precipitation or by buffering the samples with disposable de-salting cartridges. Proteins, including tethered heterologous cellulases, can also be recovered and purified from recombinant yeast cell

cultures by methods including spheroplast preparation and lysis, cell disruption using glass beads, and cell disruption using liquid nitrogen for example. Additional protein purification methods include ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, gel filtration, and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

[0111] Protein analysis methods include methods such as the traditional Lowry method or the protein assay method according to BioRad's manufacturer's protocol. Using such methods, the protein content of saccharolytic enzymes can be estimated. Additionally, to accurately measure protein concentration a heterologous cellulase can be expressed with a tag, for example a His-tag or HA-tag and purified by standard methods using, for example, antibodies against the tag, a standard nickel resin purification technique or similar approach.

[0112] The transformed host cells or cell cultures, as described above, can be further analyzed for hydrolysis of cellulose (*e.g.*, by a sugar detection assay), for a particular type of cellulase activity (*e.g.*, by measuring the individual endoglucanase, cellobiohydrolase or β glucosidase activity) or for total cellulase activity. Endoglucanase activity can be determined, for example, by measuring an increase of reducing ends in an endoglucanase specific CMC substrate. Cellobiohydrolase activity can be measured, for example, by using insoluble cellulosic substrates such as the amorphous substrate phosphoric acid swollen cellulose (PASC) or microcrystalline cellulose (Avicel) and determining the extent of the substrate's hydrolysis. β -glucosidase activity can be measured by a variety of assays, *e.g.*, using cellobiose.

[0113] A total cellulase activity, which includes the activity of endoglucanase, cellobiohydrolase and β -glucosidase, can hydrolyze crystalline cellulose synergistically. Total cellulase activity can thus be measured using insoluble substrates including pure cellulosic substrates such as Whatman No. 1 filter paper, cotton linter, microcrystalline cellulose, bacterial cellulose, algal cellulose, and cellulose-containing substrates such as dyed cellulose, alpha-cellulose or pretreated lignocellulose. Specific activity of cellulases can also be detected by methods known to one of ordinary skill in the art, such as by the

Avicel assay (described supra) that would be normalized by protein (cellulase) concentration measured for the sample.

[0114] One aspect of the invention is thus related to the efficient production of cellulases to aid in the digestion of cellulose and generation of ethanol. A cellulase can be any enzyme involved in cellulase digestion, metabolism and/or hydrolysis, including an endoglucanase, exoglucanase, or β -glucosidase.

[0115] In additional embodiments, the transformed host cells or cell cultures are assayed for ethanol production. Ethanol production can be measured by techniques known to one of ordinary skill in the art *e.g.* by a standard HPLC refractive index method.

Heterologous Cellulases

[0116] According to the present invention the expression of heterologous cellulases in a host cell can be used advantageously to produce ethanol from cellulosic sources. Cellulases from a variety of sources can be heterologously expressed to successfully increase efficiency of ethanol production. For example, the cellulases can be from fungi, bacteria, plant, protozoan or termite sources. In some embodiments, the cellulase is a *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowensis*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, or *Arabidopsis thaliana* cellulase.

[0117] In some embodiments of the invention, multiple cellulases from a single organism are co-expressed in the same host cell. In some embodiments of the invention, multiple cellulases from different organisms are co-expressed in the same host cell. In particular, cellulases from two, three, four, five, six, seven, eight, nine or more organisms can be co-expressed in the same host cell. Similarly, the invention can encompass co-cultures of yeast strains, wherein the yeast strains express different cellulases. Co-cultures can include yeast strains expressing heterologous cellulases from the same organisms or from

different organisms. Co-cultures can include yeast strains expressing cellulases from two, three, four, five, six, seven, eight, nine or more organisms.

[0118] Cellulases of the present invention include both endoglucanases or exoglucanases. The cellulases can be, for example, endoglucanases, β -glucosidases or cellobiohydrolases.

[0119] In certain embodiments of the invention, the endoglucanase(s) can be an endoglucanase I or an endoglucanase II isoform, paralogue or orthologue. In some embodiments, the endoglucanase expressed by the host cells of the present invention can be recombinant endo-1,4- β -glucanase. In particular embodiments, the endoglucanase is a *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *R. speratus*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces*, *Irpex lacteus*, *C. lucknowense*, *C. globosum*, *Aspergillus terreus*, *Aspergillus fumigatus*, *Neurospora crassa* or *Acremonium thermophilum* endoglucanase. In one particular embodiment, the endoglucanase comprises an amino acid sequence selected from SEQ ID NOs: 30-39 or 52-56, as shown in Table 1 below. In certain other embodiments, the endoglucanase comprises an amino acid sequence that is at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to an amino acid sequence selected from SEQ ID NOs: 30-39 or 52-56.

[0120] As a practical matter, whether any polypeptide is at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identical to a polypeptide of the present invention can be determined conventionally using known computer programs. Methods for determining percent identity, as discussed in more detail below in relation to polynucleotide identity, are also relevant for evaluating polypeptide sequence identity.

[0121] In one particular embodiment, the endoglucanase is an endoglucanase I ("egl") from *Trichoderma reesei*. In certain embodiments, the endoglucanase comprises an amino acid sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:39.

[0122] In another particular embodiment, the endoglucanase is an endoglucanase from *C. formosanus*. In certain embodiments, the endoglucanase comprises an amino acid sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:31.

[0123] In another particular embodiment, the the endoglucanase is an endoglucanase from *H. jecorina*. In certain embodiments, the endoglucanase comprises an amino acid

sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:54.

[0124] In certain embodiments, the β -glucosidase is a β -glucosidase I or a β -glucosidase II isoform, paralogue or orthologue. In certain embodiments of the present invention the β -glucosidase is derived from *Saccharomycopsis fibuligera*. In particular embodiments, the β -glucosidase comprises an amino acid sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:40.

[0125] In certain embodiments of the invention, the cellobiohydrolase(s) can be a cellobiohydrolase I and/or a cellobiohydrolase II isoform, paralogue or orthologue. In one particular embodiment, the cellobiohydrolase comprises an amino acid sequence selected from SEQ ID NOs: 21-29 or 46, as shown in Table 1 below. In particular embodiments of the present invention the cellobiohydrolase is a cellobiohydrolase I or II from *Trichoderma reesei*. In another embodiment, the cellobiohydrolase comprises a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:27 or SEQ ID NO:28.

[0126] In other particular embodiments of the present invention the cellobiohydrolase is a cellobiohydrolase I or II from *T. emersonii*. In another embodiment, the cellobiohydrolase comprises a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:23 or SEQ ID NO:24.

[0127] In another embodiment, the cellobiohydrolase of the invention is a *C. lucknowense* cellobiohydrolase. In a particular embodiment, the cellobiohydrolase is *C. lucknowense* cellobiohydrolase Cbh2b. In one embodiment, the cellobiohydrolase comprises a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:25.

[0128] In some particular embodiments of the invention, the cellulase comprises a sequence selected from the sequences in Table 1 below. The cellulases of the invention also include cellulases that comprise a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99 or 100% identical to the sequences of Table 1.

[0129] Some embodiments of the invention encompass a polypeptide comprising at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, or 500 or more consecutive amino

acids of any of SEQ ID NOs:21-40, 46 or 52-56, or domains, fragments, variants, or derivatives thereof.

Table 1: Cellulases used in Examples 1-11 as described below.

Donor organism/ Gene	Codon-Optimized DNA sequence used	Accession number and amino acid sequence
<i>Cellulohydrolases</i>		
<i>Humicola grisea cbh1</i>	GAATTCATGAGAACCGCTAAGTTCGCTACCTTGGCTGCCTTGGTTGCCTCT GCTGCTGCTCAACAAGCCTGTTCCTTGACTACTGAACGTCAACCCATCTTTG TCTTGGAACAAGTGACTGCTGGTGCTCAATGTCAAAGTGTCCAAAGCCTCC ATCACTTTGGACTCTAATTGGAGATGGACCCACCAGTCTCTGCTAGTACT AACTGTTACACCGGTAATAAGTGGGACACTTCTATTGTACTGACGCTAA GTCTTGCTCAAAATTGTTGTGTGATGGTGCTGATTACACCTCCACTTA TGGTATTACCACCAACGGTGACTCTTTGTCCCTTGAAGTTCGTTACTAAAGG TCAACATTCCACCAACGTCGGTTCTAGAACCTACTTAATGGACGGTGAAG ACAAGTACCAAACCTTCGAATTGTTGGGTAAATGAATTTACCTTCGATGTCG ATGTGCTAACATCGGTTGTGCTCGTTTGAACGGTGCTTTATACTTCGTTTCTAT GGACGCCGACGGTGTGTTGCTCGTTACCCAGGTAAATAGGCTGGTGCCA AGTATGGTACCGGTTACTGTGATGCTCAATGCCCAAGAGACATTAAAGTTC ATCAACGGTGAAGCTAACATTGAAGGTGGACTGGTTCTACCAACGACCC AAACGCTGGCGCGGTAGATAACGGTACCTGTTTCCGAAATGGACATTT GGGAAGCCAAACAATGGCTACTGCTTTTACTCCACACCCATGTACCATC ATTGGTCAATCCAGATGTGAAGGTGACTCCTGTGGCGGTACCTACTCCAA CGAAAGATACGCTGGTGTGTTGTGATCCAGACGGTTGTGACTTCAACTCCTA CAGACAAGGTACAAGACTTTCTATGTTAAGGTATGACTGTGCTGATACCA CCAAGAAGATCACCGTCGTACCCCAATTTCTTGAAGGACGCTAACGGTGAT TTAGGTGAATTAAGAATTCTACGTCCAAGATGGTAAGATCATCCCAA CTCTGAATCTACCATTCAGGTGTGTTGAAGGTAAATTCATCACTCAAGACTG GTGTGACAGACAAAAGGTGCTTCGGTGATATTGACGACTTCAACAGAA AGGGTGGTATGAAGCAAATGGGTAAAGGCTTTGGCCGGTCCAAATGGTCTTG GTTATGTCTATTGGGACGATCACGCTTCCAACATGTTGTGTTGGACTCC ACCTTCCAGTTGATGCTGCTGGTAAGCCAGGTGCCGAAAGAGGTGCTTG TCCAACACTACTCCGGTGTCCAGCTGAAGTTGAAGCCGAAGCTCCAAATT CTAACGTTGTCTTCTTAACATCAGATTCGGTCCAATCGGTTCCACAGTCTG	CAA35159 MRTAKFATLAALVASAAQA CSLTTERHPSLSWNKCTAGGQC QTVQASITLDSNWRWTHQVSGS TNCYTGKNKWDTSICTDAKSCAQ NCCVDGADYTSTYGITNGDSLS LKFVTKGQHSTNVGSRTYLMDG EDKYQTFELLGNEFTFDVDVSN GCGLNGALYFVSMDADGGLSR YPGNKAGAKYGTGYCDAQCP DIKFINGEANIEGWTGSTNDPNA GAGRYGTCCSEMDIWEANNMA TAFTPHPCTIIGQSRCEGDSGGT YSNERYAGVCDPDGCDFNSYRQ GNKTFYGKGMTVDTTKKITVVT QFLKDANGDLGEIKRFYVQDGK IIPNSESTIPGVEGNSITQDWCDR QKVAFGDIDDFNRKGGMKQMG KALAGPMVLVMSIWDDHASNM LWLDSTFPVDAAGKPGAERGAC PTTSGVPAEVEAEAPNSNVFSN IRFGPIGSTVAGLPGAGNGGNG GNPPPPTTTTSSAPATTTTASAGP KAGRWQCCGGIGFTGPTQCCEEP YICTKLNDWYSQCL (SEQ ID NO:21)

	CTGGTTTGCCAGGTGCTGGTAATGGTGGTAATAACGGTGGTAACCCACCA CCACCACCACTACCACTTCTTCTGCCCCAGCTACTACCACCCGCTTCT GCTGGTCCAAAGGCTGTAAGATGGCAACAATGTGGTGGAATTGGTTTCAC CGGTCCAAACCAATGTGAAGAACCATACATCTGTACCAAGTTGAACGACT GGTACTCTCAATGTTTATAACTCGAG (SEQ ID NO:1)	
<i>Thermoascus aurantiacus cbh1</i>	GAATTCAATGTAACCAAGAGCTCTATTGTTCTCCTTCTTGGCCGCCGCT AGAGCTCATGAAGCCGGTACTGTCAACCGCCGAAACCAACCCATCCTTGAC TTGGCAACAATGTTCCCTCTGGTGGTTCTTGTAATACTCAAAACGGGAAGGT TGTTATTGACGCTAACTGGAGATGGGTTACACTACCTCCGGTTACACCAA CTGTTACACTGGTAACACTTGGGATACTTCCATCTGTCCAGACGACGTTAC CTGTGCTCAAAACTGTGCTTTGGACGGTGCTGACTACTCCGGTACTTACGG TGTCACTACCTCTGGCAACGCGTTGAGATTGAACCTTCGTCACCCAAATCTTC TGTAAGAAACATCGGTTCTAGATTGTACTTGTGCAAGACGATACTACTTA CCAAATCTTCAAGTTGTTGGGTCAAGAGTTCACCTTCGACGTTGATGTTTC CAACTTGCCTTGTGTTTGAACGGTGCTTTGTACTTCGTTGCTATGGACGC CGACGGTAACCTTATCCAAAGTACCCAGGTAAACAAGCCGGTGCCAAAGTACG GTACCGGTTACTGTGATTCTCAATGTCCAAGAGACCTAAAATTCAATTAAACG GTCAAGCTAACGTGCAAGGTTGGCAACCATCTGTAAACGATCCAAACGCC GGTGTCGGTAATCACGGTTCCTCCTGTGCTGAAATGGACGTTTGGGAAGC TAACTCTATCTCCACCGCCGTCCTCCACATCCATGTGATAACCCAGGTCA AACCATGTGTCAAGGTGATGATTGTGGTGATACCTACTCTTCCACTAGATA CGCTGGTACCTGTGACACCGACGGTTGTGATTTCAACCCATACCAACCCAG GTAAACCACTCTTTCTACGGTCCAGGTAAAGATTGTCGATACTTCTTAAGT TCACTGTTGTCACTCAATTCAATTACCGACGATGGTACCCCATCTGGTACCC TAACTGAAATTAAAGAGATTCTACGTCCAAACGGTAAAGTCAATTCACAA TCCGAAAGCACCAATTTCCGGTGTTACCGGTAACTCCATCACCACCTGAATAC TGTAACCGCTCAAAAGGCCGCCCTTTGACAACACCGGTTTCTTCAACCCATGGT GGTTTGCAAAAGATTCTCAAGCCTTGGCTCAAGGTATGGTTTGGTTCATG TCCTTGTGGGATGACCACGCTGCTAACATGTTGTGGTTGGATTCTACTTAC CCAACTGACGCTGATCCAGACACCCAGGTGTTGCTAGAGGTACTTGTCC AACCACCTTCTGGTGTCCAGCTGACGTGGAATCTCAAAACCCTAACCTCTTA CGTTATCTACTTAACATCAAGGTGGGTCCAATTAACTCCACCTTCACTGC	AAL83303 MYQRALLFSFFLAARAHEAGT VTAEHNHPSLTWQQCSSGGSC QNGKVVIDANWRWVHTTSGYT NCYTGNTWDTSICPDDVTCAQN CALDGADYSGTYGVTTSGNALR LNFVTQSSGKNIGSRLYLLQDDT TYQIFKLLGQEFTHFDVDSNLPC GLNGALYFVAMDADGNLSKYP GNKAGAKYGTGYCDSQCPRDL KFINGQANVEGWQPSANDPNAG VGNHGSSCAEMDVWEANSISTA VTPHPCDTPGQTMCGDDCCGGT YSSTRYAGTCDTDGCDFNPYQP GNHSFYGPGKIVDTSSKFTVVTQ FITDDGTPSGTLTEIKRFYVQNG KVIPQSESTISGVTGNSITTEYCT AQKAAFDNTGFFTHGGLQKISQ ALAQGMVLVMSLWDDHAANM LWLDSTYPTDADPDTPGVARGT CPTTSGVPADVESQNPNSYVIYS NIKVGPINSTFTAN (SEQ ID NO:22)

	TAACTAACTCGAG (SEQ ID NO:2)	GAATTCATGCTAAGAAGAGCTTTACTATTGAGCTCTTCTGCTATCTTGGCC GTTAAGGCTCAACAAGCCGTACCGCTACTGCTGAAACCAACCCTCCATT GACCTGGCAAGATGTACCGCTCCAGGTTCTTGTACCAACCACAAACGGTG CTGTCGTCTTGACGCTAACTGGAGATGGGTCCACGACGTCAACGGTTAC ACTAACTGTTACACCGGTAAACACCTGGGACCCAACTTACTGTCCAGACGA CGAAACTTGCGCTCAAAACTGTGCCCTTGGAACGGTGCTGACTACGAAGGTA CTTACGGTGTTACCTCCTCTGCTGTTCTTCCCTTGAAGTTGAACCTTCGTCACTGG TTCTAACGTCGGTTCAGATTGTATTGTTGCAAGATGACTCCACTTACCA AATCTTCAAGTTGTTGAACAGAGAAATTTTCTTTTCGACGTGATGTGTCCAA CTTGCCCTTGTTGAACGGTGCTCTATACTTCGTTGCTATGGACGCTGA TGGTGGTGTTTCCAAGTACCCAAACAACAAGGCTGGTGCCAAATACGGTA CTGGTTACTGTGACTCTCAATGTCCACGTGACTTGAAGTTTATTGATGGTG AAGCTAATGTGCAAGGTTGGCAACCATCTTCTAACAAACGCTAACACTGGC ATCGGTGACCACGGTTCTTGCTGTGCCGAATGGACGTTTGGGAAGCCAA CTCCATTTCCAACGCCGTCACCTCCACACCCATGTGACACTCCAGGTCAAAC TATGTGTTCCGGCGATGACTGTGGTGTAATCTTACTCTAACGATAGATACGC TGGTACCTGTGATCCAGACGGTTGCGACTTCAATCCATACAGAAATGGGTA ACACTTCCTTTACGGTCCAGGCAAGATCATCGACACTACTAAGCCATTCA CTGTTGTCACCCAAATTTCTTGACCGACGATGGTACTGATACCGGTACTTTGT CCGAAATCAAGAGATTCTACATCCAAAACTCTAACGTCAATCCACAACCA AATTCGGACATCTCTGGTGTCACCTGGTAACCTCCATTACCAACCGAATTTGT ACCGCCCAAAAGCAAGCTTTCGGTGACACCCGACGACTTCTCTCAACACGG TGGTTTGGCTAAGATGGGTGCTGCTATGCAACAAGGTATGGTTTGGTTCAT GTCTTTGTGGGACGACTACGCTGCTCAAATGTTGTGGTTGGACTCCGATTA CCCAACCGATGCCGACCCAAACCAACCCCTGGTATCGCTAGAGGTACCTGTC CAACTGACTCTGGTGTTCATCTGACGTGCAATCCCAATCTCCAAACTCCT ACGTCACTTACTCCAACATTAAATTCGGTGCCAATCAACTCCACTTTCACCTG CTTCTTAACTCGAG (SEQ ID NO:3)	MLRRALLSSSAILAVKAQQAG TATAENHPPLTWQECTAPGSCCT QNGAVVLDANWRWVHDVNGY TNCYTGNTWDPTYCPDDETCAQ NCALDGADYEGTYGVTSSGSSL KLNFTVGSNVGSRLYLLQDDST YQIFKLLNREFSFDVDVSNLPCG LNGALYFVAMDADGGVSKYPN NKAGAKYGTGYCDSQCPRDLKF IDGEANVEGWQPSSNNANTGIG DHGSCCAEMDVWEANSISNAV PHPCDTPGQTMCSGDDCCGGTYS NDRYAGTCDPDGCDFNPYRMG NTSFYGPCKIIDTTKP FTVVVTQFLTDDGTDGTGLSEIKR FYIQNSNVIPQPNSDISGVTGNSI TTEFCTAQKQAFGDTDDFSQHG GLAKMGAAMQQGMVLVMSLW DDYAAQMLWLDSDYPTDADPT TPGIARGTCPTDSGVPSDVESQSP NSYVTYSNIKFGPINSTFTAS (SEQ ID NO:23)
Talaromyces emersonii cbh1			
Talaromyces emersonii cbh2	GAATTCATGCGTAACCTTGTGGCCTTGGCTCCAGCCGCTTTGTTGGTTGGT GCTGCCGAAGCTCAACAAATCCTTGTGGGGTCAATGCGGTGTTCTCCTCTG GACTGGTGCAACTTCTGTGCCGCTGGTGCCACCTGTTCACCATTAACCC	AAL78165 MRNLLALAPALLVGAAEAQQS LWGQCGGSSWTGATSCAAGAT	

	ATACTACGCTCAATGTGTTCCAGCCACTGCCACTCCAACTACCTTGACTAC CACCATAAGCCAACCTCCACCGGTGGTGCTGCTCCAACCACTCCACCAC CAACTACTACCGGTACTACCACCTCTCCAGTCGTCAACAGACCTGCCCTCCG CCTCCGGTAATCCATTTCGAAGGTTATCAATTGTACGCTAACCTTACTACG CTTCTGAAGTCATTTCCCTTGGCTATCCCATCTTTGAGCTCCGAGTTGGTCCC AAAGGCTCCGAAGTTGCTAAGTCCCTTCAATTGTCTGGTTAGATCAAGC TGCCAAAGGTTCCATCTATGGGTGATTACTTGAAAGGATATTCAATCTCAAAA CGCTGCTGGTGCTGATCCACCAATCGCCGGTATTTTCGTTGTTACGATTI GCCAGATAGAGACTGTGCCGCCGCTGCTTCTAACGGTGAAATTTTCTATCGC CAACAAACGGTGTCGCTTTATACAACAATATATCGATTCCATTAGAGAAC AATTAAACCACTTACTCCGACGTCCATACCATCTTGTTATCGAACCAAGACT CTTTGGCTAACGTTGTCATACTTGAACGTTCCAAAATGTGCTAACGCTC AAGATGCTTACTTGGAATGTATACTAACTACGCTATTACCCAAATTGGACTTGC CAAAACGTTGCTATGTACTTGACGCTGGTCAACGCCGTTGTTGGTTGGCTTGGC AAGCCAACTTGGCCCCAGCTGCTCAATTATTGCTTCTGTTTACAAGAACG CCTCTTCCCCAGCCTCTGTAGAGGTTTGCTACCAACGTGGCTAACTACA ACGCCCTGGTCCATTTCTAGATGTCCATCCTACACTCAAGGTGACGCTAACT GTGATGAAGAAGATTACGTTAACGCTTTGGGTCCATTGTTCCAAGAACAA GGTTTCCCAGCTTACTTCATCATCGACACTTCCCCTAACGGTGTGACAGCCA ACTAAGCAATCTCAATGGGGTGACTGGTGAACGTTATTGGTACCGGTTTC GGTGTTAGACCAACCAACCGACACTGGTAACCCATTGGAAGACGCTTTCGT TTGGGTCAAGCCAGGTGGTGAATCCGACGGTACCTCCAACACTACTAGCC CACGTTACGATTACCACTGTGGTTTGTCTGACGCTTTGCAACCAAGCTCCAG AAGCTGGTACCTGGTTCCAAGCCTACTTCGAACAATTGTTGACTAACGCC AACCCATTGTTCTAACTCGAG (SEQ ID NO:4)	CSTINPYAQCVPATATPTTLTT TTKPTSTGGAAPTTPPPTTTGTTT SPVVTRPASASGNPFEGYQLYAN PYYASEVISLAIPSLSELVPKAS EVAKVPSFVWLDQAQAKVPSMG DYLKDIQSQNAAGADPPIAGIFV VYDLPRDRDCAAAAASNGEFSIAN NGVALYKQYIDSIREQLTTYSDV HTILVIEPDSLANNVVTNLNVPKC ANAQDAYLECINYAITQLDLPNV AMYLDAGHAGWLGWQANLAP AAQLFASVYKNASSPASVRGLA TNVANYNAWSISRCPSTYTQGDA NCDEEDYVNALGPLFQEQQGPPA YFIIDTSRNGVRPTKQSQWGDW CNVIGTGFGVRPTTDTGNPLEDA FVWVKPGGESDGTSTNTTSPRYD YHCGLSDALQPAPEAGTWFQAY FEQLLTNANPLF (SEQ ID NO:24)
<i>Chryso- sporium lucknowense CBH2b</i>	ATGGCCAAGAAGTTGTTTCATTACCGCTGCCTTAGCTGCCGCAGTGCTTGCT GCACCAAGTGATCGAAGAGAGACAAATTTGCGGAGCCGTCTGGACACACAGT GCGGAGGCAACGGCTGGCAAGGCCCAACATGTTGTGCTTCTGGCTCAACG TGCGTGGCACAGAACGAGTGGTATTCCCAGTGCCTTCCAAACTCCCAGGT GACTTCTTCAACAACCCCCAGCTCAACGTCTACTTCACAGAGATCCACAA GTACCTCTTCTAGCACAAACCAGAGTGGCTCATCTCATCTAGCAGTACG ACCCCTCACCCGTATCAAGTCTGTACAGGATATCCCTGGCGGAGCAAC	MAKKLFITAAALAAAVLAAPVIEE RQNCGAVWTQCGGNWQGPTC CASGSTCVAQNEWYSQCLPNSQ VTSSTTPSSSTSTQRSTSTSSSTR SGSSSSSTTPPPVSSPVTsipgga TSTASYSGNPFSGVRLFANDYYR

	CTCAACAGCCAGTTATTCCGGCAATCCTTTCTCTGGAGTGAGATTATTTGC AAACGACTATTATAGATCAGAGGTTCAACAACCTTGCAATTCTCTATGAC GGGAACCCTAGCCGCA AAGGCTTCCGCCGTAGCAGAAGTCCCTAGTTTCC AATGGCTTGACAGAAACGTTACAATAGATACACTTATGGTACAGACTTTA TCTCAGGTTAGAGCTTTGAATAAGGCCGGTGCCAACCCACCTTATGCTGCC CAATTAGTAGTCTATGACTTGCCAGATAGAGACTGTGCTGCCGACGCTTCT AATGGTGAAATTTTCCATCGCA AATGGCGGAGCTGCAAACTATAGATCATA CATTGATGCAATAAGAAACACATCATTTGAGTATTCTGATATTAGAATAA TCCTTGTTGATTGAACCCAGACTCCATGGCTAATATGGTTACCAACATGAATG TAGCCAAAGTGTCTAACGCAGCTTCCACATACCATGAGCTAACCGTATAT GCATTA AACA AACTGAATCTACCTAACGTTGCTATGTACTTAGATGCCGGT CATGCCGATGGTTGGGCTGGCCTGCA AATATCCAACCCGACGCTGAATT GTTTCGCTGGAACTACAACGACGCCGGA AAGCCCGCTGCCGTTAGAGGCT TAGCCACA AATGTTGCA AATTACAACGCTTGGTCAATTGCTAGTGCCCTT CTTATACCTCACCAATCCTAACTACGATGAGAAACATTACATAGAAGCA TTTTCCCAATTGTTAACTCCGCTGGATTCCCTGCCAGATTTCATCGTGGAT ACCGGTAGAAACGGCA AACAACCAACTGGACAACAACAATGGGAGATT GGTGTAACGTCAAGGGAACCGGCTTCGGCGTCAGGCCTACGGCAACACACC GGACACGAGCTAGTCGACGCTTTTGATGGGTTAAGCCAGGTGGCGAAAG TGACGGAACAAGTGACACGAGTGCTGCAAGATACGATTACCACTGTGGTC TGTC CGACGCTTACAGCCCCCGAGGCTGGACAATGGTTCCAGGCT TATTTTGAAACAATTGTTAACGAAACGCAAAATCCACCATTTCTAA (SEQ ID NO:5)	SEVHNLAIPSMTGTLAAKASAV AEVPSFQWLDNRNVTIDTLMVQT LSQVRALNKAGANPPYAAQLVV YDLPRDCAAAASNGEFSIANG GAANYRSYIDAIRKHIEYS DIRII LVIEPDSMANMVTN MNVAKCS NAASTYHELTVYALKQLNLPNV AMYLDAGHAGWLGW PANIQPA AELFAGIYNDAGKPAAVRGLAT NVANYNAWSIASAPSYTSPNPN YDEKHYIEAFSPLLNSAGFPARFI VDTGRNGKQPTGQQQWGDWC NVKGTGFGVRPTANTGHEL VDA FVWVKP GGESDGTSDTSAARYD YHCGLSDALQPAEAGQWFAQY FEQLLTNANPPF (SEQ ID NO:25)
<i>Talaromyces emersonii cbh1 with CBD</i>	ATGCTAAGAAGAGCTTTACTATTGAGCTCTTCTGCTATCTTGGCCGTTAAG GCTCAACAAGCCGGTACCGCTACTGCTGA A AACCACCTCCATTGACCTG GCAAGAATGTACCGCTCCAGGTTCTTGTA CCA CCA A A A CCGTGCTGTCG TCTTGGACGCTAACTGGAGATGGGTCCACGACGTCAACGGTTACACTAAC TGTTACACCCGGTAACACCTGGGACCCA ACTTACTGTCCAGACGACGAAAC TTGCGCTCA A AACTGTGCCTTGACCGGTGCTGACTACGAAGGTACTTACG GTGTTACCTCCTCTGGTTCTTCCTTGAAGTTGA ACTTCGTCACGTGTTCTAA CGTCGGTTCAGATTGTATTTGTTGCAAGATGACTCCACTTACCAATCTT CAAGTTGTTGAACAGAGAATTTTCTTTCGACGTCGATGTGTCCAACTTGCC	MLRRALLSSSAILAVKAQQAG TATAENHPPLTWQECTAPGSCTT QNGAVVLDANWRWVHDVNGY TNCYTGNTWDPTYCPDDETC AQ NCALDGADYEGTYGVTSSGSSL KLNFTGSNVGSRLYLLQDDST YQIFKLLNREFSFDVDVSNLPCG LNGALYFVAMDADGGVSKYPN NKAGAKYGTGYCDSQCPRDLKF

	TTGTGGTTTGAAACGGTGCTCTATACTTCGTTGCTATGGACGCTGATGGTGG TGTTTCCAAAGTACCCAAACAACAAGGCTGGTGCCAAATAACGGTACTGGTT ACTGTGACTCTCAATGTCCACGTGACTTGAAGTTTATTGATGGTGAAGCTA ATGTCGAAGGTTGGCAACCATCTTCTAACAAACGCTAACACTGGCATCGGT GACCACGGTTCTTGCTGTGCCGAAATGGACGTTTGGGAAGCCAACTCCCAT TTCCAACGCCGTCACTCCACACCCATGTGACACTCCAGGTCAAACTATGTG TTCCGGCGATGACTGTGTGGTACTTACTCTAACGATAGATACGCTGGTAC CTGTGATCCAGACGGTTGCGACTTCAATCCATACAGAAATGGGTAAACACTT CCTTTTACGGTCCAGGCAAGATCATCGACACTACTAAGCCATTCACACTGTTG TCACCCAAATTCTTGACCGACGATGCTACTGATACCGGTACTTTGTCCGAAA TCAAGAGATTCTACATCCAAAACCTCTAACGTCAATCCCAACCAAAATTCC GACATCTCTGGTGTCACCTGGTAACCTCCATTACCAACCGAATTTTGTACCGCC CAAAAGCAAAGCTTTCGGTGACACCCGACGACTTCTCTCAACACGGTGGTTT GGCTAAGATGGGTGCTGCTATGCAACAAGGTATGGTTTGGTCAATGTCTTT GTGGGACGACTACGCTGCTCAAATGTTGTGGTTGGACTCCGATTACCCAA CCGATGCCGACCCAAACCAACCCCTGGTATCGCTAGAGGTACCTGTCCAACT GACTCTGGTGTCCATCTGACGTCCGAATCCCAATCTCCAACCTCCTACGTC ACTTACTCCAACATTAAATTTCGGTCCAATCAACTCCACTTTCACCTGCTTCT AACCCTCCAGGTGTTAACAGAGGTACTACCCTACTCTCGTAGGCCAGCTAC TACAACGTGTTCTTCCCCAGGCCCAACCCCAATCCCACTACGGTCAATGTGG TGGTATCGGTTACTCTGGTCCAAACCGTCTGTGCTTCTGGTACTACCTGTCA AGTTTTAAACCCATACTACTCTCAATGTTTGTAG (SEQ ID NO:6)	IDGEANVEGWQPSSNNANTGIG DHGSCCAEMDVWEANSISNAV PHPCDTPGQTMCSGDDCGGTYS NDRYAGTCDPDGCDFNPYRMG NTSFYGPCKIIDTTKPFVVTQFL TDDGTDGTLSEIKRFYIQNSNVI PQPNSDISGVTGNSITTEFCTAQK QAFGDTDDFSQHGGGLAKMGAA MQQGMVLVMSLWDDYAAQML WLDSDYPTDADPTTPGIARGTCP TDSGVPSDVESQSPNSYVTYSNI KFGPINSTFTASNPPGGNRGTTT RRPATTTGSSPGPTQSHYGQCGG IGYSGPTVCASGTTCCQVLNPYYS QCL (SEQ ID NO:26)
Trichoderma reesei CBH1	ACCESSION NO.: CAA49596 MVSFTSLLAGVAAISGVLAAPA AEVEPVAVEKREAEAEQAQSACT LQSETHPPLTWQKCSSGGTCTQ QTGSVVIDANWRWTHATNSSTN CYDGNTWSSTLCPDNETCAKNC CLDGAAYASTYGVTTSGNSLSIG FVTQSAQKNVGARLYLMASDIT YQEFTLLGNEFSFDVDVSQLPCG	

	TGTTGACGTTTCCCAATTGCCATGTGGCTTGAACGGTGCTTTGTACTTTGTCTCTATGGATGCTGACGGTGGTGTTTCTAAGTACCCAACTAACACTGCCCGGTGCTAAGTACGGTACTGTTACTGTGATTCTCAATGTCCACGTGACTTGAA GTTCATTAACGGTCAAGCCAACGTCAAGTTGGGAACCATCCTCCAACA ACGCTAACACCGGTATCGGTGGTCACGGTTCCTGTTGTTCCGAATGGAC ATCTGGGAAGCTAACAGTATTTCTGAAGCTTTGACACCAACCATGCAC CACTGTCGGTCAAGAAATTTGTGAAGTGATGGATGTGGTGGAACCTACT CTGATAACAGATACGGTGGTACTTGTGACCCAGACGGTTGTGACTGGAAC CCATACAGATTGGGTAAACACTTCTTTCTATGGTCCAGGTTCTTCTTTCACCT TGGATACCAACAAGTTGACTGTTGTTACCCAATTTCGAAACTTCTCTGGTG CTATCAACAGATACTACGTTCAAAACGGTGTCACCTTCCAAACAACCAAAC GCTGAATTGGGTTCTTACTCTGGTAATGAATTGAACGACGACTACTGTACC GCTGAAGAAGCTGAATTTGGTGTTCCCTCTTTCTCCGACAAAGGGTGTTTG ACCCAATTCAAGAAGGCTACCTCCGGTGGTATGGTTTGGTTATGTCCCTTG TGGGATGATTACTACGCAAAACATGTTATGGTTAGACAGTACTTACCCAAC TAACGAAACCTCCTCTACTCCAGGTGCTGTCAGAGGTTCCCTGTCTTCTACCTC TTCTGGTGTCCAGCTCAAGTTGAATCTCAATCTCCAACGCTAAGGTCAC TTTCTCCAACATCAAGTTCGGTCCAATCGGTTCCACTGGTAATCCATCTGG TGGAACCCCTCCAGGTGGTAACAGAGGTACTACCACTACTCGTAGGCCAG CTACTACAACCTGTTCTTCCCCAGGCCCAACCCAATCCCACACTACGGTCAAT GTGGTGGTATCGGTTACTCTGGTCCAACCGTCTGTGCTTCTGTGTAACCT GTCAAGTTTTAAACCCATACTACTCTCAATGTTTGTA (SEQ ID NO:7)	LNGALYFVSMADADGGVSKYPTN TAGAKYGTGYCDSQCPRDLKFI NGQANVEGWEPSSNNANTGIGG HGSCCSEMDIWEANSISEALTPH PCTTVGQEICEGDCGCGGTYS DN RYGGTCDPDGC DWNPYRLGNTS FYGPGSSFLLD TTKKLT VVTQFE TSGAINRYVYVQNGVTFFQQPNAE LGSYSGNELNDDYCTAEAEAFG GSSFSDKGGLTQFKKATSGGMV LVMSLWDDYYANMLWLDSTYP TNETSSTPGAVRGSCSTSSGVPA QVESQSPNAKVTF SNIKFGPIGST GNPSGGNPPGGNRGTTTTRRPAT TTGSSPGPTQSHYGQCGGIGYSG PTVCASGTTCCQVLNPHYYSQCL (SEQ ID NO:27) [SECRETION SIGNAL : 1 – 33 CATALYTIC DOMAIN : 41 – 465 CELLULOSE-BINDING DOMAIN : 503 – 535]
Trichoderma reesei CBH2	ACCESSION NO.: AAA72922AAA34210 MIVGILTTLATLATLAASVPLEE RQACSSVWGQC GGQNWSGPTC CASGSTCVYSNDYYSQCCLPGAA SSSSSTRAASTTSRVSP TTSRSSS ATPPPGSTTTRVPPVSGGTATYS GNPFVGVTPWANAYYASEVSSL AIPSLTGAMATAAAA VAKVPSF	

	GCTGTCGCCAAAGTTCCATCCTTCATGTGTTGGACACCTTGGACAAAACCT CCATTATGGAACAACCTTGGCAGACATAAGGACTGCTAACAGAACG GCGGTAACTACGCTGGTCAATTGTGTGTACGACTTGCCAGACAGAGAC TGTGCTGCTTTGGCTTCCAACGGTGAATACTCCATCGCTGACGGTGTGTC GCCAAGTACAAGAACTACATTGATACCATAGACAAATCGTTGTGCGAATA CTCTGACATCAGAACCTTGTTAGTCATCGAACCAAGATTCTTTAGCCAATTT AGTCACCAACTTGGGTACTCCAAGTGTGCTAACGCTCAATCTGCCTACTT AGAATGTATCAATTATGCAGTTACCCAATTGAACCTTGCCAAACGTTGCTAT GTACTTGGACGCTGGTCACGCCGGTTGGTTGGTTGGCCAGCTAACCAAG ACCCAGCCGCTCAATTATTCGCCAACGTTTACAAGAAATGCCTCTTCTCCTA GAGCCTTGCGTGGTTGGCTACTAACGTCGCTAACTACAACGGTTGGAAC ATCACTTCTCCACCATCTTACACCCAAGTAACGCTGTTTACAACGAAAA GTTGTACATTACGCTATCGGTCCATTATTGGCTAACCAATGGTTGGTCTAA CGCCTTCTTCATCACCGACCAAGGTAGATCCGGTAAACAACCAACTGGTC AACAACAATGGGGTGATTGGTGTAACGTCAATCGGTACTGGTTTCGGTATC AGACCATCCGCTAACACTGGTGATTCCCTTGTGGATTCCCTTCGCTCGGTT AAGCCAGGTGGTGAATGTGATGGCACCTCTGATTCCCTCTGCTCCAAGATTC GATTCCCACTGCGCCTTGCCAGACGCTTTTGCAACCAAGCCCAAGCTGG TGCAATGGTTCCAAGCTTACTTTGTCCAATTGTTGACCAACGCTAACCCATC TTTCTTGTAA (SEQ ID NO:8)	MWLDTLDKTPLMEQTLADIRTA NKNNGNYAGQFVVYDLPDRDC AALASNGEYSIADGGVAKYKNY IDTIRQIVVEYSDIRTLLVIEPDSL ANLVTNLGTPKCANAQSAYLEC INYAVTQLNLPNVAMYLDAGHA GWLGWPANQDPAQLFANVYK NASSPRALRGLATNVANYNGW NITSPPSYTQGNADVYNEKLYIHAI GRLLANHGWSNAFFITDQGRSG KQPTGQQQWGDWCNVIGTGFGI RPSANTGDSLLDSFVWVKPGGE CDGTSDSAPRFDSHCALPDALQ PAAQAGAWFQAYFVQLLTNAN PSFL (SEQ ID NO:28)
<i>Chaetomium thermo- philum CBH1</i>	TTAATTAAACAATGATGTACAAGAAATTTGCAGCCCTAGCTGCTTTAGTTG CAGGAGCTTCCGCTCAACAGGCATGTTTCATTGACTGCCGAAATCATCCA TCCTTAACGTGGAAGAGATGCACGTCAGGAGGTTTCATGCTCCACTGTAAA CGGAGCTGTCACAATAGATGCAAAATTGGAGATGGACCCACACTGTGTCCG GTAGTACAACTGCTACACCGGTAATCAATGGGATACGTCTTTGTGTACA GATGGAAAGTCATGCGCTCAGACCTGTTGCGTGATGGAGCAGACTACTC TTCTACTTACGGAAATCACGACATCAGGTGACAGTCTTAATTTGAAATTCTGT AACCAGCACCAAGTACGGAAACAATGTAGGCTCCAGAGTGTAATTAAATGG AGAACGATACCAAATATCAAATGTTTCGAGTTATTAGGCAATGAGTTTACC TTTGACGTAGACGTTAGCAAATTTGGGTTGCGGATTAAACGGCGCCCTTTAC TTCGTGTCTATGGATGCTGACGGAGGTATGTCAAAGTATTCTGTTAACAA AGCCGAGCAAGTACGGTACAGGTTATTGTGACGCTCAGTGCCCTAGAG	AM711862 MMYKKFAALAAALVAGASAQQA CSLTAENHPSLTWKRCTSGGSCS TVNGAVTIDANWRWTH TVSGSTNCYTGNGQWDTSLCTDG KSCAQTCVVDGADYSSTYGHTS GDSLNLKFVTKHQYG TNVGSRVYLMENDTKYQMFELL GNEFTFDVDVSNLGCGLNGALY FVSMDADGGMSKYSNG KAGAKYGTGYCDAQCPDLKFI NGEANVGNWTPSTNDANAGFG

	ATTGAAGTTTATCAACGGAGAAGCCAACGTTGGTA ACTGGACGCCAAGT ACTAACGACGCAAAACGCTGGATTTCGGCAGATACGGTAGTTGTTGCTCAGA AATGGACGTGTGGAGGCCAATAACATGGCAACCGCTTTTACTCCTCACCC CATGTACAACCTGTTGGACAATCTAGATGTGAAGCCGACACGTGCGGTGGC ACCTACAGTAGCGATAGGTATGCAGGAGTATGTGATCCTGACGGTTGCCGA TTTCAATGCTTATAGACAAGGAGACAAAACGTTTATGTTAAAGGTATGA CCGTCGATACTAACAAAGAAGATGACTGTGTTACCCAGTTCACACAAGAAC TCAGCTGGAGTATTGTCTGAAATTAAAGATTCTACGTCCAGGATGGAAA GATTATTGCTAATGCCGAGAGTAAGATACCAGGTAAACCTGGAAATAGTA TCACACAGGAATACTGTGACGCTCAGAAGGTAGCTTTTAGCAACACCGAT GACTTCAATAGAAAAGGTGGAATGGCTCAAATGAGTAAGGCTTTAGCCGG TCCAATGGTGTGTGATGTCTGTGTTGGGATGATCACTATGCAAACATGCT TTGGCTTGACAGCACCTATCCTATCGACCAAGCCGAGCCCCAGGTGCTG AAAGGGGTGCATGTCCAACCCAGAGTGGTGTGCCCGCCGAGATTGAAGCT CAAGTGCCTAATAGTAACGTTATCTTTTCCAAATATAAGATTCCGACCAATC GGATCCACTGTTCCAGGTTTGATGGATCTAATCCTGGCAACCCAAACAAC CACGGTAGTCCCTCCAGCTTCAACTCCACAAGTAGACCAACAAGTTCAA CGTCCAGTCCAGTGTCTACTCCTACCGGACAACCAAGGAGGCTGTACCACT CAGAAATGGGGTCAATGCGGTGGAATTGGCTATACAGGTTGTACGAATTG CGTTGCAGGAACCACTTGTACACAGTTAAACCCTTGGTACTCACAAATGCCT ATAAGGCGCGCC (SEQ ID NO:9)	RYGSCCSEMDVWEANNM ATAFTPHPCTTTVGQSRCEADTCG GTYSDDRYAGVCDPDGCDFNAY RQGDKTFYGKGMTVD TNKKMTVVVTQFHKNSAGVLSEI KRFYVQDGKILANAESKIPGNPG NSITQEYCD AQK VAF SNTDDFNRKGGMAQMSKALAG PMVLVMSVWDDHYANMLWLD STYPIDQAGAPGAERGACP TTSGVPAEIEAQVPNSNVIFS NIR FGPIGSTVPGLDGSPNPNPTTV VPPASTSTRPTS STSSPVSTPTGQPGGCTTQKWGQ CGGIGYTGCTNCVAGTTCTQLN PWYSQCL (SEQ ID NO:29)
<i>Accremonium thermo- philum CBH1</i>	ATGTATACCAAATTTGCTGCATTTGGCCGCTTTAGTTGCAACAGTAAGAGGT CAAGCCGCTTGTCTCTAAACCGCAGAAACTCACCCCATCTCTACAATGGCA GAAATGCACAGCCCTGGATCTTGTACAACTGTCTCCGGCCAAAGTCAACCA TTGACGCTAATTGGAGATGGCTTCACCAAACTAACTCTTCAACGAAATTGTT ATACCGGTACGAAATGGGATACTTCCATAATGTTCAATCCGATACAGACTGC GCAACGAAATGTTGTTTAGATGGAGCAGACTATACGGGAACCTTATGTTGT TACAGCCTCAGGTAAATCCCTAAACCTTAAGTTTCGTAACCTCAAGGACCAT ATAGTAAGAAATATCGGCTCTAGAAATGTACTTGATGGAAGTGAGAGCAAA TATCAGGGTTTACGTTATTGGGACAAGAGTTTACATTTGATGTTGATGTG AGTAACCTTAGGTTGCGGCTTAAACGGCGCCTTGTACTTCGTTTCTATGGAT CTTGATGAGGTGTATCAAAATACACGACCAACAGGCTGGAGCCAAATA	MYTKFAALAAALVATVRGQAAC SLTAETHPSLQWQKCTAPGSCCT VSGQVTIDANWRWLHQTNSSTN CYTGNEWDT SICSSD TDCATKC CLDGADYTGTYGVTASGNSLNL KFVTQGPYSKNIGSRMYLMESES KYQGFTLLGQEFTFDVDVSNLG CGLNGALYFVSMDLDGGVSKYT TNKAGAKYGTGYCDSQCPRDLK FINGQANIDGWQPSSNDANAGL GNHGSCCSEMDIWEANKVSAAY

	TGGTACGGGATATTGTGACAGCCAATGCCCTAGAGACTTAAAGTTCATTA ACGGTCAGGCAATAATTGACGGCTGGCAACCAAGCAGTAACGACGCTAAT GCCGGAAGTAACCATGGCTCATGTTGTTCCGAAATGGATATCTGGGA AGCCAATAAGGTGCCGTGCCTACACCCCATCCATGCACGACAATCG GTCAGACAAATGTGTACCGGTGATGACTGTGGAGGCACATACTCAAGTGAT AGGTACGCCGGTATATGTGATCCTGACGGTTGCCGATTTCAACTCTTATAGA ATGGGAGATACATCCTTTTACGGCCCCGGTAAACAAGTTGATACGGGTAG TAGTTCACTGTTGTTACTCAGTTCTTAACAGGTTCAGACGGCAATCTTAG TGAAATCAAAAGATTCTACGTTCAGAATGGAAAGTCAATTCCTAATTCCG AGAGTAAGATTGCTGCTGTCTGTAAACAGTATCACGACCGACTTCTGT ACCGCCCAAAGACTGCCCTTTGGAGATACGAATGTTTTCGAGGAAAGGGG CGGCTTGCTCAAATGGGCAAGGCTTTGGCCGAACCAATGGTATTAGTCC TATCCGTTTGGGATGATCATGCAGTGATATATGCTTTGGCTTGATAGCACCT ACCTACTGACAGCACCAGCCAGGAGTGCCAGAGTGACTGCTCCTATC ACAAGTGGCGTTCCAGCAGATGTAGAGAGCCAAAGCTCCAACCTCCAATGT GATCTATTCTAACATCAGATTGGCCCCATTAAATAGTACCTATACAGGAAC GCCCTCTGGTGAACCCCTCCAGCGGAGGCCACCACTACACGACCA CAACGACTTCAAAGCCTTCTGGCCCTACGACAACCTACCAATCCTTCCGGA CCACAGCAAACCTCACTGGGGTCAGTGTGGAGGCCAAGGATGACGGGTC CTACCGTGTCAATCACCTTACACATGCAAAATACAGTAATGACTGGTACT CTCAGTGTTTATAA (SEQ ID NO:45)	TPHPCTTIGQTMCTGDDDCGGTYS SDRYAGICDPDGCDFNSYRMGD TSFYGPCKTVDTGSKFVVVTQFL TGSDGNLSEIKRFYVQNGKVIPN SESKIAGVSGNSITTDFTAQKT AFGDTNVFEERGGLAQMGKAL AEPMLVLVSVWDDHAVNMLWL DSTYPTDSTKPGAARGDCPITSG VPADVESQAPNSNVVYSNIRFGPI NSTYTGTPSGGNPPGGTTTFTT TTTSKPSGPTTTTNPSPGPOQTHW GQCGGQGWGTGPTVCQSPYTCK YSNDWYSQCL (SEQ ID NO:46)
<i>Endoglucanases</i> <i>Coptotermes</i> <i>lacteus EG</i>	ATGAGATTTCCTTCCATATTACCCGCTGTTTGTTCGCAGCCTCAAGTGCTT TAGCAGAAATGTACTAAGGGTGGATGTACTAACAGAATGGATACATAGTT CATGATAAGCACGTGCGTGACATCCAGAAATAGAGACACTTTGGACCCCTCC AGACTTAGATTATGAATAAGGACGTGGGAGTAACCGTGTCCGGTGGAACCC TTAGTCAAAGATTAGTCTCAACTTGGAACGGTAAGAAAGTCGTGGGAAGT AGATTGTATATTGTGGACGAAGCCGACGAGAAATATCAATTATTCACATTT GTCGGTAAGGAGTTCACCTATACCGTTGATATGTCCCAGATCCAATGTGGA ATCAATGCCGCATTATACACAGTGGAATGCCTGCCGCTGGAAAGACCCC TGGAGGTGTTAAGTATGGATATGGATATTGTGATGCCAACTGCGTGGATG GAGATTGTTGTATGGAGTTCGATATCCAAGAAGCTTCTAACAAAGGCAATC	MRFPSIFTAVLFAASSALAECK GGCTNKNGYIVHDKHVGDIQNR DTLDPPDLDYEKDVGVTVSGGT LSQRLVSTWNGKKVVGSRLYIV DEADEKYQLFTFVGKEFTYTVD MSQIQCCGINAALYTVEMPAAGK TPGGVKYGYGYCDANCVGDGC CMEFDIQEASNKAIVYTTTHSCQS QTSGCCDTSGCGYNPYRDSGDKA FWGTTINVNQPVTVTIVTQFIGSGSS

	GTTTACACCACCCATTCTGTCAAAGTCAAACCTTCAGGTTGCGATACCTCA GGATGCGGTTACAACCCCTTACAGAGACAGTGGTGACAAAGGCATTTCTGGGG AACAACTATAAACGTAAACCAGCCTGTGACAAATTGTAAACACAGTTTATCG GTTCTGGTAGTTCCTTAACTGAAGTCAAAAGATTGTGCGTGCAAGGTGGA AAGACCTTCCCTCCAGCCAAATCATTAACCGACAGTTATTGTAAATGCCAAC GACTATAGAAAGTTTGAGAACTATGGGTGCATCCATGGCTAGAGGACACAGT TGTTGTGTTTCTTTGTGGGATTCTAATGGTATGAGTTGGATGGATGGAGG TAACGCCGTCCTTGTACCTCATATAATATTGAATCTTTGGAAATCCAGTCA GCCAAACTTAAAGGTCACATGGTCAACGTGAATAACGGAGAGATCGATT CTCCTTATTAA (SEQ ID NO:10)	LTEVKRLCVQGGKTFPPAKSLT DSYCNANDYRSLRTMGASMAR GHVVVFSLWDSNMGMSWMDGG NAGPCTSYNIESLESSQPNLKVT WSNVKYGEIDSPY (SEQ ID NO:30)
<i>Coptotermes formosanus EG</i>	ATGAGATTCCCTTCCATTTCACACTGCTGTTTGTTCGCAGCCTCAAGTGCTT TAGCAGCCTATGACTACAAGACAGTATTGAAGAACTCCTTGTGTTCTACG AAGCTCAAAGAGTGGAATAATGCCTGCAGACCAGAAAGTGACCTGGAG AAAAGATTCCGCATTAAACGACAAGGACAGAAAGGAGAGACTTAAC GGAGGTATTACGACGCCGGAGACTTTGTGAAGTTCGGTTTCCCAATGGCA TACACAGTTACCGTGTGGCCTGGGTTTAGTCGATTATGAATCTGCTTAC AGTACTGCGGTGCCCTTGGATGATGTTAGAAAGGCCTTGAAATGGGTAC AGATTATTCTTGAAAGCACATACCGCTGCCAATGAGTTTACGGACAGGT GGGTCAGGGAGATGTGGATCATGCTTACTGGGACGTCCTGAGGACATGA CTATGTCTAGACCAGCTTACAAGATCGATACATCAAAACCTGGTAGTGACT TAGCTGCAGAAACAGCAGCCGCTTAGCAGCAACCGCAATAGCTTACAAG TCAGCCGATTCTACCTACAGTAACAACTTAATTACTCATGCAAAAGCAGTTG TTCGATTTTGCAAAACAATTATAGAGGAAAGTACTCTGATAGTATTACCGAT GCCAAGAAATTCTATGCATCCGGTGATTATAAGGACGAAATTAGTATGGGCT GCAGCCTGGTTGTATAGAGCTACAAATGATAACACTTACTTAACCAAAGC CGAATCATTTGTATAATGAATTGGTTTAGGATCTTGGAACGGTGCAATTC TTGGGATAACAAGATATCCGGAGTTCAGGTCTTATTAGCCAAATTGACATC CAACAAGCATACAAGATAAAGTTCAGGGTTATGTTGATTACTTAGTCTC CTCTCAAAAGAAAACCTCCAAGGGATTGGTCTATATTGACCAATGGGGAA CCTTAAGACACGCAGCTAATAGTGCCCTTGATCGCTTACAGGCCGCTGATT TGGGTATAACGCTGCTAGTTATAGACAAATACGCAAGAAGCAAAATTGAT TATGCCCTTAGGTGACGGAGGTCGTTCTTACGTGTCGGATTTCGGAACTAAC	BAB40697 MRFPSIFTAVLFAASSALAAVDY KTVLKNSLLFYEAQRSGLPAD QKVTVWRKDSALNDKGQKGEDL TGGYVDAGDFVKFGFPMAYTV TVLAWGLVDYESAYSTAGALD DGRKALKWGTDYFLKAHTAAN EFYGQVGGQDVEDHAYWGRPED MTMSRPAVKIDTSKPGSDLA TAAALAAATAIAYKSADSTYSNN LITHAKQLFDFANNYRGKYSDSI TDAKNFYASGDYKDELVWAAA WLYRATNDNTYLTKAESLYNEF GLGSWNGAFNWDNKGISVQVL LAKLTSKQAYKDKVQGYVDYL VSSQKKTPKGLVYIDQWGTLRH AANSALIALQAADLGINAASYR QYAKKQIDYALGDGGRSYVVG FGTNPPVVRPHRSHSSCPDAPAC DWNNTYNSAGPNAHVLTGALVG GPDSNDSYTDSDYISNEVATD YNAGFQSAVAGLLKAGV (SEQ

	CCTCCAGTAAGACCTCATCATAGATCCAGTTCCTGTCTGACGCACCAGCC GCTTGCGACTGGAATACTTACAACCTCTGCCGGACCAAAATGCCACGTCTTG ACCGGAGCCTTAGTAGGTGGACCAGATTCCAACGATAGTTACACAGATTC ACGTTCTGATTATATCAGTAACGAAAGTCGCTACTGATTACAATGCCGGTTT CCAATCTGCAGTTGCTGTTGTTGTTGAAAGCCGGAGTATAA (SEQ ID NO:11)	ID NO:31)
<i>Nasutitermes takasa- goensis EG</i>	ATGAGATTTCATCTATTTTCACTGCCGTCTTATTTGCAGCCTCCAGTGCAAT TAGCAGCCTATGATTATAAACAAAGTTTGTGAGAGATTCCCTTATTGTTCTACG AAGCTCAGAGAACGGTAGATTACCAGCAGACCAGAGGTCACTTGGAG AAAAGATTACGCCTTGAAATGATCAGGGAGATCAAGGTCAAGACTTAACCG GAGGTTATTTTGACGCCGTGATTTTGTGAATTTTGGTTTCCCAATGGCAT ATACTGCTACCGTCTTGCCCTGGGTTTAAATCGATTTTGAGGCAGGATACA GTTCCGCTGGTGCCTTGGATGACGGTAGAAAGCAGTAAAGTGGCAACT GATTACTTTATAAAGGCCACACTTCACAGAAATGAGTTTACGGACAAGTC GGTCAGGGTGACGCTGATCACGCTTCTGGGACGTCCTGAAGATATGAC CATGGCTAGACCAGCCTACAAGATTGACACACAGCAGACAGGTAGTGACT TAGCGGGTGAAACCGCAGCGGCAATTGGCAGCTGCCAGTATCGTGTTTAGA AATGTTGATGGTACATACTCTAACAACTTACTTACTCATGCCAGACAATTA TTTGACTTTTGCAAATAACTACAGAGGAAATACTCAGATTCCATAACCGA CGCTAGAAACTTTTACGCCAGTGCAGATTACCGTGACGAATTGGTTTGGGC TGCCGCAATGGTTGTACAGAGCTACAAATGACAACACTTACTTGAATACCG CAGAAATCCTTGTAATGATGAATTTGGATTGCAGAACTGGGTGGAGGGTTA AACTGGGATTCAAAAGGTGTCTGGTGTCCAGGTCTTGTAGCAAAATTGACC AACAAACAGGCTTACAAGATACTGTGCAGTCTTACGTGAATTACCTGATT AATAACCAGCAAAAGACCCCAAAAGGATTGTTATACATTGATATGTGGGG TACATTGAGACACGCCGCAATGCTGCATTCATCATGTTGGAAGCTGCCG AGTTGGGTTTATCCGCATCATCTTACAGACAGTTTGTCTCAAACTCAGATCG ACTACGCTTTGGGTGACGGTGGAAGAAGTTTCGTCTGTGGTTTGGTTCAA ACCTCCTACAGACCACATCATCGTTCTTCCAGTTGCCCGCCTGCCCCAG CAACTTGACTGGAATACATTCAACTCACTGACCCAAATTACCACGTGT TATCTGGAGCTTTGGTAGGAGGACCAGATCAAAACGATAATTATGTGGAT GATAGATCCGACTACGTCCATAACGAAAGTGGCAACCGACTACAACGCCGG ATTTCAGAGTGCTTTGGCAGCCTTAGTTGCTTTGGGTTATTAA (SEQ ID	MRFPSIFTAVLFAASSALAAAYDY KQVLRDSLFLFYEAQSRSLPAD QKVTWRKDSALNDQDQGQDL TGGYFDAGDFVKFGFPMAYTAT VLAWGLIDFEAGYSSAGALDDG RKA VKWATDYFIKAHTSQNEFY GQVGQGDADHAFWGRPDMT MARPAYKIDTSRPGSDLAGETA AALAAASIVFRNVDTYSNNLL THARQLFDFANNYRGKYSDSIT DARNFYASADYRDELVWAAAW LYRATNDNTYLTAEPLYDEF LQNWGGGLNWDSKVSQVQVLL AKLTNKQAYKDTVQSYVNYLIN NQKTPKGLLYIDMWGTLRHA ANAAFIMLEAAELGLSASSYRQF AQTQIDYALGDGGRSFVCGFGS NPPTRPHHRSSCPPAPATCDWN TFNSPDPNHYVLSGALVGGPDQ NDNYVDDRSDYVHNEVATDYN AGFQSALAAALVALGY (SEQ ID NO:32)

	NO:12)	ATGAGATTCCCTAGTATTTTCACTGCCGCTCTTATTTGCAGCCAGTTCTGCTT TAGCCGCATATGATTATACCACAGTTTGTGAAGAAGTTCCTTATTGTTCTACG AAGCTCAAAGATCCGGTAAGTTGCCAGCCGACCAGAAAGGTCACCTTGGAGA AAAGATTACGATTAGACGATAAAGGAATAAATGGAGAGGACTTAACAG GAGGTTATTATGACGCTGGTGATTTTGTGAAGTTTGGTTTTCCTTTAGCATA CACCGCTACTGTTTAGCCTGGGTTTGGTGGACTATGAAGCGGGTTACTC ATCCGCTGGAGCCACAGATGACGGTAGAAAGGCAGTGAAATGGGCAACC GACTATTTGTGAAGGCACATACTGCCGCTACCGAGTTATACGGACAGGTC GGGACGGTGACGCCGATCACGCATATTGGGGACGTCCTGAAGATATGAC TATGGCTAGACCAGCATACAAGATCGACGCTAGCAGACCAGGATCTGACT TAGCGGGTGAAACCGCTGCCGCTTTAGCCGCTGCATCCATAGTTTTCAAAG GTGTAGATTCTTCATATTCTGACAACCTTGTTAGCTCACGCTAACACGTTAT TTGATTTGCTGACAAATTATAGAGGAAATAACAGTGATTCCTATAACACAA GCTTCAAACCTTTACGCCCTCCGGAGATTACAAAGACGAGTTAGTCTGGGCT GCCACTTGGTTGTACAGAGCAACCAACGATAATACATAATTTGACCAAAGC AGAATCCTTGTAACAACGAGTTCGGATTAGGAAACTGGAACGGAGCCTTTA ATTGGGACAACAAGGTGTCCGGTGTTCAGGTGTTGTAGCCAAATTGACCT CCAAGCAGGCTTATAAAGACACCGTTCAAGGATACGTCGATTATTTGATTA ACAATCAGCAAAGACCCCAAAGGGTTTGTATACATAGACCAATGGGGG ACCTTGAGACACGCAGCTAATGCTGCCTTAATAATCTTACAGGCTGCTGAT TTGGGTATTTCTGCCGACAGTTATAGACAATTTCGCAAAGAAAGCAAATAGA TTACGCTTTAGGTGACGGAGGTAGATCATATGTAGTTGGTTTGGAGACAA TCCTCCAACACATCCTCATCACCGTTCTTCCTCATGCCCTGACGCCCCAGC AGTATGCGGATTGGAATACTTTCAATTCACTGATCCAAACTTTCATGTCTT AACC GGAGCTTTAGTGGGAGGTCCCTGATCAGAACGATAACTACGTTGATG ATCGTTCGACTACGTGTCCAACGAGGTTGCAACCGACTATAATGCAGGAT TCCAAAGTGCTGTGGCCGCTTTAGTTACTTTAGGAGTTTAA (SEQ ID NO:13) ATGAGATTCCCAAGTATATTTACTGCTGTTTGTTCGCAGCCAGTTCTGCTT TAGCAGCCTATGATTACAATGACGTATTAAACCAAAGTTTGTGTTCTACG AAGCTCAAAGATCCGGTAAGTTACCTTCTGATCAGAAAGTCACCTGGAGA AAAGATTACGATTAAACGATAAGGGACAAATGGTGAGGACTTAACTGG	MRFPSIFTAVLFAASSALAAAYDY TTVLKSSLLFYEAQQRSGKLPADQ KVTWRKDSALDDKGNNGEDLT GGYYDAGDFVKFGFPLAYTATV LAWGLVDYEAGYSSAGATDDG RKAVKWATDYLLKAHTAATEL YGQVGDGDADHAYWGRPEDM TMARPAYKIDASRPGSDLAGET AAALAAASIVFKGVDSSYSDNL LAHAKQLDFADNRYRGKYSDSI TQASNFYASGDYKDELVWAAAT WLYRATNDNTYLTKAESLYNEF GLGNWNGAFNWDNKVSGVQV LLAKLTSKQAYKDTVQGYVDY LINNQKTPKGLLYIDQWGTLR HAANAALIHQAADLGISADSYR QFAKKQIDYALGDGGRSYVVGF GDNPPTHPHRSSSCPDAVVC DWNTFNSPDPNFHVLTGALVGG PDQNDNYVDDRSDYVSNEVAT DYNAGFQSAVAALVTLGV (SEQ ID NO:33)
Coptotermes acinaciformis EG		MRFPSIFTAVLFAASSALAAAYDY NDVLTKSLLFYEAQQRSGKLP SD QKVTWRKDSALNDKQNGEDL TGGYYDAGDYVKFGFPMAYTA	
Mastotermes darwinensis EG			

	TGGATATTATGACGCCGGTGATTACGTGAAGTTTGGTTTCCAAATGGCATA TACTGCTACCGTTTGGCTTGGGGTTTAGTGGACCATCCTGCCGGATACAG TTCTGCGGTGCTTGGATGATGGTAGAAAAGCTGTGAAGTGGGTACCG ATTACTTAATCAAAGCCACGTATCAAAGAACGAATTATACGGACAGGTC GGTGACGGTGACGAGATCACGCTTATTGGGGACGTCCAGAGGATATGAC AATGGCAAGACCAGCATACAAAATAGACACTTCAAGACCAGGTTCCGACT TAGCGGGTGAACCCGACGGCATTTGGCTGCTGCATCTATTGTGTTTAAGT CAACAGATTCTAATTACGCCAACACCTTATTGACCCACGCAAAACAATTAT TCGACTTTGCCAATAACTATAGAGGTAAGTATAGTGATTCCATAACACAG GCATCTAATTTCTACAGTAGTTCCGACTATAAAGATGAATTGTTTGGGCA GCTGTATGGTTGTACAGAGCCACTAACGATCAGACCTATTTGACAACTGCA GAGAA GTTATACTCAGACTTGGGATTACAGTCCTGGAACGGAGGTTTCAC ATGGGACACCAA AATTAGTGGAGTAGAAGTGTTATTGGCTAAGATTACTG GTA AACAGGCATATAAGGACAAAGTAAAGGATATTGTGATTATATCTCA GGATCTCAGCAGAAACACCTAAAGGATTAGTTTACATAGATAAGTGGGG TTCCTTAAGAAATGGCCGCAAAACGCCGATATATTGCGCTGTAGCCGCAGA CGTCGGAATCAGTTC AACAGCTTACAGACAGTTCGCCAAACACAGATT ATTACATATTGGGTGATGCCGGACGTTCTTTTGTGTTGTTACGGAAACA ACCCACCTACACACCCACATCACAGATCCAGTTCATGTCCTGACGCCCCAG CAACATGCGATTGGAATAACTACAACAGTGCTAACCCCTAATCCACATGTTT TATACGGTGCA TTAGTTGGTGACCAAGATTCCAACGATAATTATCAAGACT TAAGATCAGATTATGTCGCCAACGAAGTGCGCAACAGACTACAAATGCAGCC TTCCAGTCATTGTTAGCATTAATCGTGGA CTTAGGTTTGTA A (SEQ ID NO:14)	TVLAWGLVDHPAGYSSAGVLD DGRKAVKWVTDYLIKAHVSKN ELYGQVGDGDADHAYWGRP ED MTMARPAYKIDTSRPGSDLA GE TAAALAAASIVFKSTDSNY ANT LLTHAKQLDFANNYRGKYS DS ITQASNFYSSSDYKDELVWA AV WLYRATNDQTYLTTAEKLY SDL GLQSWNGGFTWDTKISGVE VLL AKITGKQAYKDKVKGYCDY ISG SQQKTPKGLVYIDKWGSLR MA ANAA YICAVAADVGISSTA YRQ FAKTQINYILGDAGRSFVV GYG NNPPTHPHRSSSCPDA PAT CD WNNYNSANPNPHVLYGAL VGG PDSNDNYQDLRSDYVANE VAT DYNAAFQSLALIVDLGL (SEQ ID NO:34)
<i>Nasutitermes walkeri EG</i>	ATGAGATTTCCATCTATTTTCACTGCCGCTCTTATTG CAGCCTCAAGTGCTT TAGCAGCCTATGATTACAAACAAGTATTGAGAGATTCTCTATTGTTCTACG AAGCTCAGAGAAGCGGTAGATTACCAGCAGACCAGAGGTCACTGGAG AAAAGATTCCGCCTTGAAATGATCAGGGAGAGCAAGGTCAAGACTTAACCG GAGGTTATTTTGACGCCGGTGATTTTGTGAAGTTTGGATTCCCAATGGCTT ATACAGCAACCGTTTGGCCTGGGGTTTAATCGACTTTGAAGCCGGTTACT CTTCTGCTGGTGCCTTGGACGATGGTAGAAAAGCAGTAAAGTGGGCTACT GATTACTTTATAAAAGCCCATACTTCTCAAAACGAGTTTACGGACAAGTC	MRFPSIFTAVLFAASSALAA YDY KQVLRDSL LFYE AQRSGRLPAD QKVTWRKDSALNDQGEQQDL TGGYFDAGDFVKFGFPMAYTAT VLAWGLIDFEAGYSSAGALDDG RKAVKWATDYFIKAHTSQNEFY GOVGQGDVDHAYWGRP EDMT MARPAYKIDTSRPGSDLA GETA

	GGTCAGGGTGACGTAGATCACGCATATTGGGGACGTCCTGAAGATATGAC AATGGCTAGACCAGCCTACAGATTGATACCAAGCAGACCAGGTAGTGACT TAGCAGGAGAACTGCTGCAGCTTTGGCTGCCGCATCCATCGTTTCAAGA ATGTAGATGTTACATATTCCAACAACCTTACTTACTCATGCTAGACAGTTGT TTGATTTCCGCCAACAAATTACAGAGGAAATACTCTGATAGTATTACCGATG CAAGAAACTTTACGCTAGTGCCGACTATAGAGATGAGTTAGTCTGGGCA GCTGCCCTGGTTACAGAGCAACCAACGACAATTCTTACTTGAAACACTGCT GAATCATTATACAACGAGTTTGGAATTGCAAAATTGGGTGGAGGTTAAA CTGGGATTCTAAAGTGAGTGGTGTTCAGTTTGTAGCCAAAGTTGACCAA CAACAAGAGTATAAGGACACTATTCAATCATACGTGAATTACTTAATCA ATAACCAACAGAAACTCCAAAGGATTGTATACATTGACATGTGGGGG ACCTTGAGACACGCAGCTAACGCAGCCTTTATAATGTTAGAAGCTGCCGA CTTAGGTTTATCCGCTTCATCTTATAGACAGTTCGCCCAACACAATAAGA CTACGCATTTGGGGACGGTGACGTTCTTTGTCTGTGTTTCGGTTCCTAA TCCTCCAACTAGACCTCATCATAGATCCAGTTCATGCCCGCTGCTCCAGC TACCTGTGATTGGAATACATTCAATTCTCCTGACCCAAACTACAATGTTT ATCCGGTGCCCTGGTGGTCTCTGACCAAGATGATACTACGTGGACG ATAGAAGTGATTATGTCCATAATGAGGTAGCAACTGACTACAAATGCCGGT TTCCCAATCAGCCCTTAGCCGCTTAGTCGCCCTTAGGTTACTAA (SEQ ID NO:15)	AALAAASIVFKNVDGTYSNLL THARQLDFANNYRGKYSDSIT DARNFYASADYRDELVWAAAW LYRATNDNSYLNTAESLYNEFG LQNWGGGLNWDSKVSQVQVLL AKLTNKQEKDITIQSYVNYLIN NQOKTPKGLLYIDMWGTLRHA ANAAFIMLEAADLGLSASSYRQ FAQTQIDYALGDGGRSFVCGFG SNPPTRPHRSSSCPPAPATCDW NTFNSPDPNYNVLSGALVGGPD QNDNYVDDRSDYVHNEVATDY NAGFQSALAAALVALGY (SEQ ID NO:35)
<i>Reticuli- termes speratus EG</i>	ATGAGATTCCCAAGTATATTACTGCCGCTCTATTTCAGCCTCCAGTGCA TTAGCCGCTTATGACTACAAAACAGTATTGTCCAAATTCCTTGTTGTTCTAC GAAGCTCAAAGATCCGGTAAGTTACCTTCTGACCAGAAAGTGACCTGGAG AAAGGATTCAGCATTAACGACAAAGGACAAAGGGTGAGGACTTAACC GGTGATATTACGACGCCGGAGACTTTGTGAAATTTGGTTTCCCAATGGCT TACACAGTTACCGTATTGGCATGGGTGTTATTGATTACGAATCCGCCTAC TCTGCCGCAGGAGCTTTAGATTCAGGTAGAAAGGCCCTTGAAATATGGGAC CGACTATTTCTTAAAGGCACATACAGCAGCTAACGAGTTTACGGACACAGG TGGGTCAAAGGTGACGTTGACCACGCATACTGGGGACGTCCTGAAGATATG ACCATGAGCAGACCAGCATACAAATAGACACTTCTAAGCCTGGTTCCCGA CTTAGCTGCAGAGACTGCAGCTGCATTAGCAGCCACAGCTATTGTCATACA AATCTGCCGATGCAACATATTCCAAACAAATTTGATAACACATGCAAACAA	AB019095 MRFPSIFTAVLFAASSALAAVDY KTVLSNLLFYEAQRSGKLPSDQ KVTWRKDSALNDKGQKGEDLT GGYYDAGDFVKFGFPMAYTVT VLAWGVIDYESAYSAAALDSG RKALKYGTDYFLKAHTAANEFY GQVGQGDVDHAYWGRPDMT MSRPA YKIDTSKPGSDLAETA AALAAATAIAYKSADATYSNNLIT HAKQLDFANNYRGKYSDSITD

	TTATTCGACTTTGCCAACAAATTACAGAGGAAATAATTC CGATAGTATTACC GATGCCAAGAACTTTATGCTTCTGCTGATTACAAAGACGAATTGGTATGG GCCGCTGCATGGTTGTACAGAGCAACCAATGACAAACACATATTTGACTAA GGCAGAAATCCTTATACAAATGAATTTGGTTTGGGAAACTTCAATGGTGCCCTT CAATTGGGATAACAAAGTCTCCGGAGTCCAGGTGTTATTGGCCAAAGTTAA CCTCAAAACAAGTGATAAGGATAAGGTACAGTCTTACGTGGACTATTTG ATCTCCTCACAAAAAAGACACCAAAAGGTTTAGTGACATCGATCAATG GGGTACTTTAAGACACGCAGCTAATTCTGCTTTGATCGCTTTGCAGGCAGC TGACTTAGGAATTAAACGCTGCTACTTACAGAGCCTACGCAAGAAGCAAA TCGACTATGCTTTGGGTGATGGTGGAAGATCCTATGTTATTGGATTTGGGA CCAACCTCCAGTAAGACCAACATCACAGAAAGTTTCATCTTGCCCAGATGCA CCAGCTGTCTGCGATTGGAAACACCTATAACTCCGCTGGTCCAACGCCAC GTGTTAACCGGTGCATTGGTTGGAGGACCTGATAGTAATGATAGTTATACC GATGCTCGTTCTGACTACATATCCAACGAAGTGCAACTGATTACAAATGCG GGTTTCCAATCCGCTGTCGCTGGATTATTGAAGCGGGTGCTCTAA (SEQ ID NO:16)	AKNFYASGDYKDELVWAAAWL YRATNDNTYLTKAESLYNEFGL GNFNGAFNWDNKVSGVQVLLA KLTSKQVYKDKVQSYVDYLISS QKKT PKGLVYIDQWGTLRHAA NSALIALQAADLGINAATYRAY AKKQIDYALGDGGRSYVIGFGT NPPVRPHHRSSSCPDAVCDW NTYNSAGPNAHVLTGALVGGPD SNDSYTDARSDYISNEVATDYN AGFQSAVAGLLKAGV (SEQ ID NO:36)
<i>Neosartorya fischeri</i> EG	ATGAGATTTCCATCTATTTTCACTGCAGTTTGTTCGCAGCCAGTTC CGCTT TGGCCCAACAGATCGGGTCCATCGCCGAAAAATCATCCTGAGTTGACAACC TATAGATGCTCCTCTCAAGCTGGATGCGTAGCACAGAGTACTTCCGTCGTG TTAGATAATTAAACGCTCATTTGGATTTCATCAAAACGGTGCCCAACAAGTTGC ACTACCTCAAGTGGATTGGACCCCTTCATTGTGCCCTGATAAAGTCACCTGT TCTCAGAACTGCGTAGTCGAAGGAATAACCGACTACTCATCTTTTGGTGTG CAAACTCCGGAGATGCAATGACATTAAGACAGTATCAAGTTCAAAAATGG ACAGATCAAAACATTGCGTCCCTAGAGTGTACTTGTAGCTGAGGATGGAA TCAATTACTCCAATAATTGCAGTTGTTGAACCAAGAGTTTACTTTCGATGTGG ACGCTTCCAATAATTGCCCTTGTGGTATGAATGGAGCTTTATATTGTCAGAAA TGGAATGCTTCTGTGGACGTTCTGCTTGAACCCAGCGGTGCCACATATG GAACAGGTTACTGTGATGCCCAGTGCTTCAACCCAGGTCCATGGATAAAT GGAGAGCAAAATACTGCTGGAGCCGGTGCAATGTTGCCAAGAGATGGACTT ATGGGAAGCCAACTCCCGTTCTACCATTTTCAGTCCCTCACCCATGTACAAC TGCGGGTTTGATGCTGTACTGGAGCTGAGTGCTACTCAATCTGTGACGG TTATGTTGCACTTACAACCCCTTATGAATTAGGAGCCAAAGATTACTATGG	XM_001258277 MRFPSIFTAVLFAASSALAQQIG SIAENHPELT TYRCSQAGCVAQ STSVVLDINAHWHIQNGAQTSC TTSSGLDPSLCPDKVTC SQNCVV EGITDYSSFVQNSGDAMTLRQ YQVQNGQIKTLRPRVYLLAEDG INYSKLQLLNQEFTFDVASKLP CGMNGALYLSEMDASGGRSAL NPAGATYGTGYCDAQCFNPGP WINGEANTAGAGACCQEMDLW EANSRSTIFSPHPCTTAGLYACT GAECYSICDGYGCTYNPYELGA KDYGYGLTIDTAKPITVVTQF MTADNTATGTLAEIRRLYVQDG KVIGNTAVAMTEAFCSSTRFEE

	TTACGGTTTGACTATTGACACCGCAAGCCAAATAACAGTGGTTACTCAGTT TATGACCGCTGATAATACAGCAACCGGTACATTAGCAGAGATCAGAAGAT TATATGTTCAAGATGGTAAAGTAATCGGAATACAGCCGTGGCCATGACC GAGGCATTTTGTAGTTCTAGTAGAACATTTGAAGAGTTAGGTGGTTTGCAA AGAAATGGGAGAAAGCTTTAGGTAGAGGAATGGTGCCAGTTTCTCAATATG GGACGATCCTGGTTTGTGGATGCATTGGTTAGATTCTGACGGTGCAGGACC TTGTGGTAATACTGAAGGTGATCCTGCCCTTCATTCAGGCTAACTACCCAAA TACCGCCGTAAACATTCTCCAAGGTGAGATGGGAGATATCGGTTCTACCTA TAGTTCTTAA (SEQ ID NO:17)	LGGLQRMGEALGRGMVPVFSI WDDPGLWMHWLSDGAGPCG NTEGDPAFIQANYPNTAVTFSKV RWGDIGSTYSS (SEQ ID NO:37)
<i>Reticuli- termes flavipes EG</i>	ATGAGATTTCCATCTATTTTCACTGCTGTTTGTTCGCAGCCTCAAGTGCTT TAGCACAATGGATGCAGATCGGTGGTAAGCAGAAATATCCTGCCCTTTAAG CCAGGTGCTAAGTACGGAAGAGGTTATTGTGACGGACAGTGCCCTCACGA CATGAAGGTGTCTAGTGAAGAGCAAAACGTTGACGGATGGAAGCCACAA GACAACGACGAAAATAGTGGAATGGAAAATTGGGTACATGTTGCTGGGA GATGGATATATGGGAAGGAAACTTAGTGTCCCAAGCCTACACCGTTCACG CTGGTTCCAAGTCCGGACAATATGAGTGTACTGGAACACAAATGCGGTGAC ACCGACAGTGGTGAAAGATTCAAGGGAACATGCGATAAAGATGTTGTGA TTTCGCAAGTTACAGATGGGGAGCTACAGACTATTACGGTCCTGGAAAGA CCGTGGACACCAACACAGCCAATGACAGTCGTGACCCAGTTTCAATTGGTGAC CCTTTGACTGAGATAAAGAGAGTTTATGTACAAGGAGGAAAAGTCATAAA CAATTCCAACAACATCTAACTTAGGTTCAAGTGTACGATTCTTTGACTGAGGC CTTCTGCGATGACACCAACACAGGTTACAGGTGATACAAATGACTTTAAGG CTAAGGAGGTATGTCTGGATTCTCCAAGAACTTAGACACCCACAAAGTTT TGGTGATGTCTTTATGGGATGACCATACAGCTAATATGTTATGGTTAGATT CTACTTATCCTACCGATAGTACAAAGCCAGGTGCCGCAAGAGGTACTTGT GCCGTCACCTCCGGGACCCCTAAAGATGTGGAATCCAAGCAAGCCAACTC TCAGGTAGTTTACAGTGACATTAAAGTTTGGTCCTATTAAATTCACATACAA AGCAAAATTAA (SEQ ID NO:18)	DQ014512 MRFPSIFTAVLFAASSALAQWM QIGGKQKYPAFKPGAKYGRGYC DGQCPHDMKVSSGRANVDGWK PQNDENSGNGKLGTCCEWEMD IWEGNLVSQAYTVHAGSKSGQY ECTGTQCGDTSGERFKGTC DK DGCDFASYRWGATDYYPGPKT VDTKQPMTVVTQFIGDPLTEIKR VYVQGGKVINNSKTSNLGSVYD SLTEAFCDDTKQVTGDTNDFKA KGGMSGFSKNLDTQPQLVMSL WDDHTANMLWLDSTYPTDSTK PGAARGTCAVTSGDPKDVESKQ ANSQVVYSDIKFGPINSTYKAN (SEQ ID NO:38)

<i>Trichoderma reesei</i> EGI	ATGGTCTCCTTCACCTCCCTGCTGGCCGGCGTTGCCGCTATCTCTGGTGTCC TAGCAGCCCCCTGCCGCAGAAAGTTGAACCTGTGCGCAGTTGAGAAACGTGAG GCCGAAGCAGAAAGCTCAACAACCAAGGAACATCAACACCAGAAAGTCCATC CAAAGTTAACAAACCTATAAATGTACTAAGAGTGGAGGGTGTGTAGCGCAG GACACAAGTGTGCTTAGACTGGAATTATCGTTGGATGCATGATGCCAAT TATAATTCTGTACTGTTAACGGCGGTGTTAACACTACGTTATGCCCCCGAT GAAGCGACTTGTGGTAAGAAATTGTTTATTGAAGGGGTTGACTACGCCGCT AGTGGTGTACGACGAGTGGGTCACTCTGTCTCCAAGGCTGAATCAATACATGCCT TCTTCTAGTGGTGGTATTCTCTCTGTGTCTCCAAGGCTGATTTATTGGATT CCGATGGGGAATATGTTATGTTAAATTAAATGGGCAAGAACTGAGTTT GATGTGATCTATCTGCATTACCTTGTGGAGAAATGGTAGTCTTTATTTA TCACAAATGGACGAAACGGCGGAGCCAATCAGTACAATACAGCTGGTGC TAATTATGGTTCAGGCTATTGTGATGCTCAATGTCCAGTGCAAGACTTGGAG GAATGGCACCTTAAACACATCACATCAAGGATTTTGTGTGAACGAAATGG ACATAATTAGAAAGTAATTCAAGAGCTAATGCACCTAACTCCGCACTCTTGTA CTGCGACCGCATGTGATTCTGCCGGTTGTGGTTTCAACCTTATGTTCTG GTTATAAGAGTTACTACGGTCCGGAGACACCGTGATACGTCAAAGACC TTCACATAAATCACTCAGTTTAAACACAGATAACGGATCTCCGAGTGGTAAT TTGGTGAGTATTACTAGGAATAATCAGCAGAACGGTGTGATATTCCGTCC GCGCAGCCAGGCGGTGACACTATATCTAGCTGTCTTCCGCCAGTGCCTAT GGCGGACTTGCTACAATGGGTAAAGGCATTGTCTCCTCAGGTATGGTCTAGTA TTTTCTATTGGAAATGATAATTCACAATACATGAATTGGCTGGATTCTGGT AATGCAGGCCCTTGCTCCTCTACAGAAAGGTAAACCAAGCAATATACTAGC TAATAACCCAAATACTCATGTTGTCTTAGTAATATTAGATGGGCGGATAT AGGTAGCACTACGAACAGTACCGCACCTCCTCCTCCACCTGCTAGCTCCAC GACATTTTCCACTACTAGAAGGTCCAGCACTACCAGCTCATCACCATCTTG TACTCAAACCCATTGGGGACAGTGTGGGTATAGGTTACAGCGGTTGCA AAACTTGCACATCTGGTACTACATGCCAATACAGTAATGACTATTACTCAC AATGTTAA (SEQ ID NO:19)	AB003694 MVSFTSLLAGVAAISGVLAAPA AEVEPVAVEKREAEAEAQPGT STPEVHPKLTITYKCTKSGCVA QDTSVVLDWNYRWMHDANYN SCTVNGGVNTTLCPDEATCGKN CFIEGVDYAASGVTTSGSSLTMN QYMPSSGGYSSVSPRLYLDS GEYVMLKLNGQELSFVDLSAL PCGENGSLYLSQMDENGGANQ YNTAGANYGSGYCD AQCPVQT WRNGTLNTSHQGFCCNEMDILE GNSRANALTPHSCTATACDSAG CGFNPYGSYKSYYPGDTVDT SKTFTIITQFNTDNGSPSGNLVSI TRKYQQNGVDIPSAQPGDITSS CPSASAYGGLATMGKALSSGM VLVFSIWNDNSQYMNWLDSGN AGPCSSTEGNPSNILANNPNTHV VFSNIRWGDIGSTTNSTAPPPPA SSTTFSTTRRSSTTSSSPSCTQTH WGQCGGIGYSGCKTCTSGTTCQ YSNDYYSQC (SEQ ID NO:39)
-------------------------------	--	--

<i>Aspergillus kawachii EgA</i>	TTAATTAAAATGAGAAATTTCTAACTTGATTGTTGCTGCTTCTGCTGCTACTA TGGTTTCTGCTTTGCCATCTAGACAAATGAAAAGAGGGATTCTGGTTTAA AATGGGTTGGTACTTCTGAATCTGGTGCTGAATTTGGTTCTGCTTTACCAAG GTACTTTGGGTACTGATTATACTTGGCCAGAAACTTCTAAAAATTCAAGTTT TGAGAAACAAGGGTATGAACATTTTGAATAACCATCTCTTGATGGAAAGA TTAACTCCAGATGGTTTGACTGGTTCTTTTGCTTCTACTTACTTGTCTGATT TGAAGTCAACTGTTGAATTTGTTACTAATTCTGGTGCTTATGCTGTTTTAGA TCCACATAATTACGGTAGATTTCGATGGTTCTATTATTGAATCTACTTCTGAT TTTAAGACTTGGTGGAATAATGTTGCTACTGAATTTGCTGATAACGATAAG GTTAATTTTCGATACAAACAACGAATATCATGATATGGAACAATCTTTGGTT TTGAATTTGAACCAAGCTGCTATTAAATGGTATTAGAGCTGCTGGTGCTACT ACTCAATACATTTTCGTTGAAGGTAAATGCTTATACTGGTGCTTGGGATTGG ACTACTTACAAATGATGATTTGTCTGGTTAACTGATTTCTGAAGATAAGATA ATATACGAAATGCATCAATACTTGGAATTTCTGATTTCTTGTTACATCTGAA ACTTGTGTTTCTTCTACTATTGGTAAAGAAAGAAATTGAAAAGGCTACTGAA TGGTTGAAAACTAACAAGCAAGGTATTATTGGTGAAATTTTGCAGGTGG TGTTAATTCTGTTTGTGAAGAGGCTGTTGAAGGAATGTTGGCTTATATGTC TGAAAATTCTGATGTTTGGGTTGGTGCTTCTTGGTGCTGCTGCTCCATG GTGGGTACTTACATGTATTCTTTGGAAACCAACTGATGTTACTGCTTATTC TACTTATTGGCCAAATTTTGAAAAAATACTTCCCATCTGGTGATGCTTCATCA TCTTCATCTGCTTCAGCTTCAGTTGCAGCCGCTACTTCTGCTGTTTCTACTA CTACTACAGCTGCATTTGAACAAACTACTACTCCAGCTACTCAAGTTGAAA TTGCTTCTTCTTCATCTTCATCATCAGCTGTTGCTGCTTCACAAACTACTTT GTCTAAGGTTAAGTCTAAATCTAAATCTCCATGTAAATTGTCACTCTGCTAC TTCATCTGCTGTTTCATCAGCTGCTGCAGTTACTACACCTGCAGTTGCAGCT ACAACTCCAGCTGCTGCTCCAACCTTCTTCTGTTGCTTTTGGCTACTACTT CTGTTTACGTTCCAACACTACTACTGCTGCTGCACCATCTCAAGTTTCATCTC AGCTGCAGCTTCATCTTCAGGTGTTGTTGGTGTTCATCAATTAACCAATGTGTGG ATCTGCTACTAATTCTGCTGGTGAAGTTAATCAATATTACCAATGTGTGG TATTAAATTGGACTGGTCCAACTGTTTGTGCTTCTCCATATACTTGTAAAGGTT CAAAACGATTACTACTATCAATGTGTGCTGAATTATAAGGCGCGCC (SEQ ID NO:47)	MRISNLIVAAASAAATMVSALPSRQ MKKRDSGFKWVGTSESGAEFGS ALPGTLGTDYTWPETSKIQVLR NKGMINIFRIPFLMERLTPDGLTG SFASTYLSDLKSTVEFVTNSGAY AVLDPHNYGRFDGSIESTSDFK TWWKNVATEFADNDKVIFDTN NEYHDMEQSLVLNLNQAAINGI RAAGATTQYIFVEGNAYTGAW DWTTYNDDDL SGLTDSEDIKIYE MHQYLDSDSSGTSETCVSSTIGK ERIEKATEWLKTNNKQGIIEFA GGVNSVCEEAVEGMLAYMSEN SDVWVGASWWSAGPWWTGYM YSLEPTDGTAYSTYLPILEKYFPS GDASSSSASASVAAATSAVSTT TTAAFEQTTTPATQVELASSSSSS SAVAASQTTL SKVKSKSKSPCKL SSATSSAVSSAAAATTTPAVAAAT PAAAPTSSSVAFATTSVYVPTTI AAAPSQVSSSAAASSSGVVGV DPQGPSATNSAGEVNQYQCGG INWTGPTVCASPYTCKVQNDYY YQCVAE (SEQ ID NO:52)
---	--	--

<i>Heterodera schachtii</i> <i>Engl</i>	TTAATTAAATGCATTGGGCTGATGTTGCTTGTTCTAGACCACCATGGCCA AGAGATTCTGTAAAGCTTTGAAGTGTAATTGGAACGCTAATGTTATTAGA GGTGCTATGGGTGTTGATGAAGGTGGTTATTGTCTGATGCTAATACTGCT TACAAATTGATGGTTGCTGTTATTGAAGCTGCTATTCTAATGGTATCTACG TTATTGTTGATTGGCATGCTCATAAATGCTCATCCAGATGAAGCTGTTAAAT TCTTTACTAGAAATTGCTCAAGCTTATGGTTCTTACTTGCAATATTTGTACGA AGATTTCAATGAACCATTTGGATGTTTCTTGGA CTGATGTTTGGTTCCATA CCATAAAAGTTATTGCTGCCATTAGAGCTATTGATAAGAAAGACGTTA TTATCTTGGGTACTCCAAATGGTCACAAAGATGTTGATGTTGCTTCTCAAA ATCCAATTAAAGGATTACCAAAACTTGATGTACACTTTGCATTTTTACGCTT CATCTCATTTTACATCTGATTTGGGTGCTAAATTGAATACTGCTGTTAACA ATGGTTTGCCAGTTTGTTACTGAATATGGTACTTGTGAAGCTTCTGGTA ATGGTAATTTGAATACTGATTCTATGTCACTTGGTGGA CTTGTGATTGATTC TTTGAAAATTCTTACGCTAATTGGGCTATTCTGATAAATCTGAAGCTTGT TCTGCTTTGTCTCCAGGTACTACTGCTGTTAATGTTGGTGTCTTCTCTAGAT GGACTTCTTCTGGTAATATGGTTGCTTCTTACTACAAAAGTCCACTG GTATTTCTTGTCTGGTAGTTCTTCAGGTCTTCAAGTGGTTCATCTAGTGG TTCTTCCGGTACATCTTCTGGTCTAGTGGTTCATCTAGTGGTAGTTCTTCC GGTAGTTCTAGTGGTAGTTCTGGTTCAGGTCTGTTCTCCTCTGCTGTTCTG GTTCTGCATCTATTCTGTGTTCCATCTAATACTTGGAAATGGTGGTGGTAG AGTTAATTTTGAAATTAAAGAACACTGGTCTGTCTGTCCATTGTGTGGTGTGTT TTTTCTGTCTTTGGCCATCTGGTACTACTTTGGGTGGTCTTGGAATATGG AATCTGCTGGTTCTGGTCAATATTCTTTACCATCTTGGGTAGAAATTGAAG CTGGTAAATCTTCTAAAGATGCTGGTTTGACTTTTAATGGTAAAGATAAGC CAACTGCTAAAATTGTTACCACCAAGAAAGTGCTTATAAGGCGCGCC (SEQ ID NO:48)	MHWADVACSRPPWPRDSVKAL KCNWNAVIRGAMGVDEGGYL SDANTAYNLMVAVIEAAISNGIY VVDWHAHNAHPDEAVKFFTRI AQAYGSYLHILYEDFNEPLDVS WTDVLVPYHKKVIAAIRAIDKK NVIILGTPKWSQDQDVASQNPIK DYQNLMYTLHFYASSHFTSDLG AKLKTAVNNGLPVFVTEYGTCE ASGNLNLNTDSMSSWWTL LDS LKISYANWAISDKSEACSA LSPG TTAVNVGVSSRWTS SGNMVAS YYKKKSTGISCSSSSSGSSSS GSSGTSSSGSSSGSSSGSSSGS SGSSSGSSSGSGSASISVVP SNTW NGGGRVNF EIKNTGSVPLCGVV FSVSLPSGTTLGGSWNMESAGS GQYSLPSWVRIEAGKSSKDAGL TFNGKDKPTAKIVTTKKC (SEQ ID NO:53)
--	---	---

<i>Hypocrea jecorina</i> (anamorph: <i>Trichoderma reesei</i>) Eg2	TTAATTAAATGAACAAGTCTGTGCTCCATTGTTGTTGGCTGCTTCTATT TGTAATGGTGTGCTGTGCTCAACAACAACTGTTTGGGGTCAATGTGGTGGTA TTGGTTGGTCTGTCCAACTAATTGTGCTCCAGGTTCTGCTTGTCTACTTT GAATCCATAATTATGCTCAATGTATTCCAGGTGCTACTACTATTACTACTTCT ACTAGACCACCATCTGGTCCAAACAATACTACTAGAGCTACTTCTACATCT TCTTCTACTCCACCAACTTCATCTGTTGTTAGATTGCTGGTGTAAACATTG CTGGTTTGTGATTTTGGTTGTACTACTGATGGTACTTGTGTTACTTCTAAAGT TTACCCACCATTGAAAAATTTCACTGTTCTAACAAATTATCCAGATGGTAT TGGTCAAAATGCAACATTTTGTAAACGAAGATGGTATGACTATTTTTAGATT GCCAGTTGGTTGGCAATATTGGTTAACAACAATTTGGGTGGTAATTGGA TTCTACTTCTATTCTAAGTACGATCAATTGGTTCAAGTTGTTGTCTTTG GGTGCTTACTGTATTGTGATATTCTATAATTATGCTAGATGGAATGGTGGT ATTATTGGTCAAGGTGGTCCAAACAATGCTCAATTTACTTCTTTTGTGGTCA CAATTGGCTTCAAAATATGCTTCTCAATCTAGAGTTTGGTTTGGTATTATG AATGAACCACATGATGTTAACATTAATACTTGGGCTGCTACTGTTCAAGAA GTTGTTACTGCTATTAGAAATGCTGGTGCTACTTCTCAATTCAATTTCTTTC CAGGTAATGATTGGCAATCTGCTGGTGCTTTTATTCTGATGGTTCTGCTGC TGCTTTGTCTCAAGTTACTAATCCAGATGGTTCTACTACTAATTGATCTTC GATGTTCATAAGTACTTGGATTCTGATAATTCTGGTACTCATGCTGAATGT ACTACAAACAATAATTGATGGTGCTTTTCTCCATTGGCTACTTGGTTGAGA CAAAACAATAGACAAGCTATTTTGACTGAAACTGGTGGTGGTAATGTTCA ATCTTGATATCCAAGATATGTGCCAACAAATTCAATACTTGAACCAAAATTC TGATGTTTATTGGGTACGTTGGTTGGGTGCTGTTCTTTTGATTTCTACT TACGTTTAACTGAAACTCCAACCTTCTTCTGGTAATTCTTGACTGATACTT CTTTGGTTTCTTCATGTTTGGCTAGAAAGTTATAAGGCGCGCC (SEQ ID NO:49)	MNKSVAPLLAAASILYGGAVAQ QTVWQCGGIGWSGPTNCAPGS ACSTLNPYYAQCPGATTTTSTR PPSGPTTTTRATSTSSSTPPTSSG VRFAGVNIAGFDGCTTDGTCV TSKVYPPPLKNFTGSNNYPDGIGQ MQHFVNEDGMTIFRLPVGWQY LVNNNLGGNLDSTSISKYDQLV QGCLSLGAYCIVDIHNYARWNG GIIGQGGPTNAQFTSLWSQLASK YASQSRVWFGIMNEPHDVNINT WAATVQEVVTAIRNAGATSQFI SLPGNDWQAGAFISDGSAAAL SQVTNPDGSTTNLIFDVHKYLD DNSGTHAECTTNIDGAFSPLAT WLRQNNRQAILTETGGGNVQSC IQDMCQQIQYLNQNSDVYLYV GWGAGSFDSTYVLTETPTSSGN SWTDTSLVSSCLARK (SEQ ID NO:54)
---	--	---

<i>Orpinomyces</i> <i>sp.PC-2 CelB</i>	TTAATTAAAATGAAGTTCTTGAACCTCTTTTGCTCTTTGTTGGGTTTGGTTATTGCTGGTTGTGAAGCTATGAGAAACATTTCTTCTAAAGAAATTGGTTAAAGAATTGACTATTGGTTGGTCTTTGGGTAATACTTTGGATGCTTCTTGTTGAAACTTTGAACTACTCTAAGATCAAACTGCTTCTGAAACTTGTTGGGGTAATGTTAAAACTACTCAAGAAATTGTACTACAAATTGTCTGATTTGGGTTTCAATACTTTCAGAAATACCAACTACTTGGTCTGGTCATTTTGGTGATGCTCCAGATTACAATAATTTCTGATGTTTGGATGAAGAGTTACGAAAGTTGTTGATTATGCTTTGAATACTGGTTACGCTATTTTGAACATTCATCATGAACCTTGGAATTACGCTTTTCAAAAGAAATTTGGAAATCTGCTAAAGAAATTTGGTTGCTATTGGAAACAATTTGCTGCTGAATTTGGTGATTACGATGAACATTTGATTTTGAAGGTATGAATGAACCAAGAAAGTTGGTGATCCAGCTGAATGGACTGTGGTGATCAAGAGGTTGGAATTTTGTTAATGAATGAACGCTTTGTTCTGTAAAACTATTAGAGCTACTGGTGTAACAATGCTAATAGACATTTGATGATTCCAACTTATGCTGCTTCTGTTAATGATGGTTCTATTAAACAATTTAAGTACCCAAATGGTGATGATAAAGTTATTGTTTCTTTGCAATTCCTACTCTCCATACAATTTTGCTTTGAACAATGGTCCAGGTGCTATTCTAATTTCTACGATGGTAACGAAATTGATTGGGTTATGAACACTATTAACTCTTCATTCATTTCTAAGGGTATTCAGTTATTATTGGTGAATTTGTGCTATGAACAGAGATAATGAAGATGATAGAGAAAGATGGCAAGAAATACTACATTAAAGGCTACTGCTTTGGGTATTCATGTTATTTGGGATAATGGTTATTTGAAGGTGAAGGTGAAGATTGGTATTATGATAGAAAGTCTTTGAACGTTATTTTCCCAAAGTTGATTAATGGTTTGATGAAGGTTTGGGTGATGAAGAAACCAAACTACTATTAGAAGAACTACTACTACAGTTCAAGTTCAACCACTATTAAACAACGAATGTTTCTACTAGATTGGGTTATTCTTGTGTAAATGGTTTCGATGTTTGTACACTGATAATGATGGTCAATGGGTGTTGAAAATGGTAATTGGTGTGTATTAAATCTTCTTGTGTTAACAATCAAGACAAATGTTGGTCTGAAAGATTAGGTTATCCCATGTTGTCAATACACTACTAATGCTGAATATACAGACAACGACGGTAGATGGGTGTAGAAAACGGTAACTGGTGCGGAATATACTTGTAAGCGCGCC (SEQ ID NO:50)	MKFLNSLLGLVIAGCEAMRNI SSKELVKELTIGWSLGNLTLDASC VETLNYSKDQTASETCWGNVKT TQELYYKLSDLGFNTFRIPTTWS GHFGDAPDYKISDVVWMKRVHE VVDYALNTGGYAILNIHHETWN YAFQKNLESARKILVAIWKQLA AEFGDYDEHLIFEGMNEPRKVG DPAEWTGGDQEGWNVFNEMN ALFVKTIRATGGNNANRHLMIP TYAASVNDGSINNFKYPNGDDK VIVSLHSYSPYNFALNNGPGAIS NFYDGNEDWVMNTINSSFISKG IPVIIGEFVAMNRDNEDDRERW QEYVIKKATALGIPCVIWDNGYF EGEGERFGHIDRKSLNVIFPKLIN GLMKGLGDEKPKTTIRRTTTT VQVQPTINNECFSTRLGYSCCNG FDVLYTDNDGQWGVENGWNC GIKSSCGNNQRQCWSERLGYP CQYTTNAEYTDNDGRWGVENG NWCGIY (SEQ ID NO:55)
---	---	---

<i>Irpex lacteus</i> <i>En1</i>	TTAATTAAATGAAGTCTTTGTTGTTGCTGCTGCTACTTTGGCTTTAT CTACTCCAGCTTTTCTGTTTCTGTTTGGGGTCAATGTGGTGATTTGGTTT TACTGGTTCTACTACTTGATGCTGCTGCTACTTCTTGTTTCATTTGAACGAT TACTACTTTCAATGTCAACCAGGTGCTGCTACTTCTACTGTTCAACCAACT ACTACTGCTTCTTCTACTTCTTCTGCTGCAGCTCCATCTTCTTCAGGTAATG CTGTTTGTCTGCTACTAGAAACAAGTTTAAAGTTCTTCGGTGTTAATGAAT CTGGTGCTGAATTTGGTAACAATGTTATTCCAGGTACTTTGGGTACTGATT ATACTTGGCCATCTCCATCTTCTATTGATTTTTCGTTGGTAAGGGTTTAA TACTTTCAGAGTTCCATTTTGTGATGGAAGATTGTCTCCACCTGCTACTGGT TTGACTGGTCCATTTGATTCTACTTATTGCAAGGTTTGAAAACTATTGTTT CTTACATTACTGGTAAAGGTGGTTATGCTTTGGTTGATCCACATAACTTTA TGATTTACAACGGTGCTACTATTCTGATACTAATGCTTTTCAAACCTTGGTG GCAAAATTTGGCTGCTCAATTTAAGACTGATTCTCATGTTGTTTTCGATGTT ATGAATGAACCACATGATATTCCAGCTCAAACTGTTTAACTTGAACCAA GCTGCTATTAAAGAAATTAGAGCTTCTGGTGCTACTTCTCAATCTATTTTGG TTGAAGGTACTTCTTATACTGGTGCTTGGACTTGGACTACTACTTCTGGTA ATTCTCAAGTTTGTGGTGCTATTCTATGATCCAAACAACAATGTTGCTATTG AAATGCATCAATACTTGGATTCTGATGGTTCTGCTGCTCTCCAACTTGTG TTTCTCCAACATAATTGGTGCTGAAAGATTGCAAGCTGCTACTCAATGGTTGC AACAAAAACAATTTGAAAGGTTTCTTGGGTGAAATTGGTGCTGGTTCTAATG CTGATTGTATTCTGCTGTTCAAGGTGCTTTGTGTGAATGCAACAATCTG ATGTTTGGTTGGGTGCTTTGTGGTGGGCTGCTGGTCCATGGTGGGTGATT ATTTTCAATCTATTGAACCACCATCTGGTGTTGCTGTTTCTTCTATTTTGCC ACAAGCTTTGGAACCATTTTGTGTTATAAGGCGCGCC (SEQ ID NO:51)	MKSLLSAAATLALSTPAFSVSV WGQCGGIGFTGSTTCDAGTSCV HLNDYYFQCQPGAAATSTVQPTT TASSTSSAAAPSSSGNAVCSGTR NKFKFFGVNESGAIEFGNNVIPGT LGTDYTWPSPSIDFFVGKGFNT FRVPFLMERLSPPATGLTGPFDS TYLQGLKTIVSYITGKGGYALV DPHNFMIYNGATISDTNAFQTW WQNLAQFKTDSHVVFDMNE PHDIPAQTVFNLNQAAINRIRAS GATSQSILVEGTSYTGAWTWTT TSGNSQVFGAIHDPNNNVAIEM HQYLDSDSGTSPTCVSPTIGAE RLQAAATQWLQQNNLKGFLGEIG AGSNADCISAVQGALCEMQQSD VWLGALWWAAGPWWGDYFQS IEPPSGVAVSSILPQALEPFL (SEQ ID NO:56)
<i>β-Glucosidases</i>		

<i>S. f. BGLI</i>	ATGGTCTCCTTACCTCCCTCCTCGCCGGCGTCTGCCGCCATCTCGGGCGTC TTGGCCGCTCCCGCCGAGGTCTGAATCCGTGGCTGTGGAGAAAGCGCTC GGACTCGGAGTCCCAATTCAAAATATACCCAGTCTCCATCCCAGAGAG ATGAGAGCTCCCAATGGGTGAGCCCGCATTTATCCAACTCCACAAGGT GGTAGGCTCCAAGACGCTCTGGCAAGAAGCATATGCTAGAGCAAAAGCCAT CGTTGGCCAGATGACTATTGTTGAAAAGGTCAATTTGACCACTGGTACCGG TTGGCAATTAGATCCATGTGTGTTGTAATACCGGTTCTGTTCCAAGATTCTGG CATCCCAACCTTTGCCCTACAAGATGGGCCATTGGGTGTTCTGATTCGCTGA CTTTGTTACTGGCTATCCATCCGGTCTTGCTACTGGTGCAACGTTCAATAA GGATTTGTTCTTCAAAGAGGTCAAGCTCTCGGTCAAGTTCAACAGCAA AGGTGTACATAATTGCGTTGGGCCCTGCTGTTGGCCCACTTGGTGTCAAAAGC CAGAGGTGCAGAAATTTGGAAGCCTTTGGTTCGACCCATATCTCCAAG GTA CTGCTGCTGTGCAACCATCAAGGTCTCCAGAGAAATAATGTTATG GCTTGTGTCAAGCACTTTATTGTTACGAACAACTACTAAAGAACTATTAGTG AGATGACATAAACCCCTGCCACCAACCAACTACTAAAGAACTATTAGTG CCAACATTCAGACAGGCCATGCATGAGTTGTACTTGTGGCCATTTGCCG ATTGCGTTCGAGCAGGTGTGTTCTGTTATGTGCTCTTATAACAGAGTCA ACAACACTTACGCTTGCGAAACTCTTACATGATGAACCACTTGCTTAAAG AAGAGTTGGGTTTCAAGGCTTTGTTGTTTCGGACTGGGTGCACAAATTAA GTGGGTTTATAGCGCTATCTCGGCTTAGATATGTCTATGCCTGGTGAAG TGTATGGGGGATGGAACACCGGCACGTCTTTCTGGGTCAAAACTTGACG AAAGCTATTTACAATGAGACTGTTCCGATTGAAAGATTAGATGATATGGC AACCAGGATCTTGGCTGCTTTGTATGCTACCAATAGTTTCCCAACAGAAGA TCACCTTCCAAATTTTCTTCATGGACAACGAAAGAATATGGCAATAAATA TTATGCTGACAACACTACCGAGATTGTCAAAGTCAACTACCATGTGGACCC ATCAAATGACTTTACGGAGGACACAGCTTTGAAGGTTGCTGAGGAATCTA TTGTGCTTTTAAAAAATGAAAAACAACACTTTGCCAAATTTCTCCCGAAAAGG CTAAAAGATTACTATTGTCGGGTATTGCTGCAGGCCCTGATCCGATAGGTT ATCAGTGTGAAGATCAATCTTGCACAAATGGCGCTTTGTTTCAAGGTTGGG GTTCTGGCAGTGTGTTCTCCAAAATATCAAGTCACTCCATTTGAGGAAA TTTCTTATCTTGCAAGAAAAACAAGATGCAATTTGATTATATTCGGGAGT CTTACGACTTAGCTCAAGTTACTAAAGTAGCTTCCGATGCTCATTTGTCTA	FJ028723 MVSFTSLLAGVAAISGVLAAPA AEVESVAVEKRSDSRVPIONYT QSPSQRDESSQWVSPHYYPQPQ GGRLQDVWQEA YARAKAIVGQ MTIVEKVNLTGTGTGWQLDPCVG NTGSVPRFGIPNLCLODGPLGVR FADFVTGYPSGLATGATFNKDL FLQRQALGHEFNSKGVHIALG PAVGPLGVKARGGRNFEAFGSD PYLQGTAAAATIKGLQENNVMA CVKHFIGNEQEKYRQDDINPAT NQTTKEAISANIPDRAMHEL YL WPFADSVRAGVGSVMCSYNRV NNTYACENS YMMNHLLKEELG FQGFVSDWGAQLSGVYSAISG LDMSPGEVYGGWNTGTTSFWG QNLTKA IYNETVPIERLDDMATR ILAALYATNSFPTEDHLPNFSSW TTKEYGNKYADNTTEIVKVNY HVDPSNDFTEDTALKVAEESIVL LKNENNTLPISEKAKRLLLSGIA AGDPPIGYQCEDQSC TNGALFQ GWGSGSVGSPKYQVTPFEEISYL ARKNKMQFDYIRESYDLAQVTK VASDAHLSIVVVSAA SGEYITV DGNQGDRRNLTLWNNGDKLIET VAENCANTVVVVVSTGQINFEG FADHPNVT AIVWAGPLGDRSGT AIANILFGKANPSGHL PFTIAKTD DDYIPIETYSPSSGEPEDNHLVEN DLLVDYRYFEEKNIEPRYAFGY
-------------------	--	--

	TAGTTGTTGCTCTGCTGCAAGCGGTGAGGGTTATATAACCGTTGACGGTA ACCAAGGTGACAGAGAAGAAATCTCACTTTGTGGAAACAACGGTGATAAATTG ATTGAAACAGTTGCTGAAAACACTGTGCCAATACTGTTGTTGTTACTTCT ACTGGTCAAAATTAATTTGAAGGCTTTGCTGATCACCCCAAATGTTACCGCA ATTGTCTGGCCGCCCATTAGGTGACAGATCCGGGACTGCTATCGCCCAAT ATTCTTTTGGTAAGCGAACCCATCAGGTCATCTTCCATTCACATAATTGCTA AGACTGACGATGATTACATTCCAAATTGAAACCTACAGTCCATCGAGTGGT GAACCTGAAGACAACCACTTGGTTGAATAATGACTTGCTTGTTGACTATAG ATATTTTGAAGAGAGAAGAAATATTGAGCCCAAGATACGCATTTGGTTATGGCTT GTCTTACAATGAGTATGAAGTTAGCAATGCAAAAGGTCTCGGCAGCCAAA AAGTTGATGAGGAGTTGCCCTGAACCACTACCTACTTATCGGAGTTTAGCT ATCAAAATGCAAAAGACAGCAAAAATCCAAAGTGATGCTTTTGCTCCAACA GATTTAAACAGAGTTAATGAGTACCTTTATCCATATTTAGATAGCAATGTT ACCTTAAAGACGGAAACTATGAGTATCCGATGGCTACAGCACTGAGCA AAGAACACACCTATCCAACCTGGGGCGGCTTGGGAGGCAACGATGCTT TGTGGGAGGTCGCTTATAAGTTGAAGTGGACGTTCAAAACTTGGGTAACT TCCACTGATAAGTTTGTTCACACAGTTGTATTGAAACACCCCTGAGGATGGC AAGTTTGAAACCCCTATTCAATTGAGAGGGTTTGAAAAGGTTGAGTTGTCC CCGGGTGAGAAGAGACAGTTGAGTTTGAAGCTTTTGAGAAGAGATCTTAG TGTGTGGGATACCACGACAGTCTTGGATCGTTGAATCTGGTACTTATGA GGCCTTAATTGGTGTGCTGTTAATGATATCAAGACATCTGTCCCTGTTTACT ATT (SEQ ID NO:20)	GLSYNEYEVSNAKVSAAKKVDE ELPEPATYLSEFSYQNAKDSKNP SDAFAPTDLNRVNEYLYPYLDS NVTCLKDGNIEYYPDGYSTEQRTT PIOPGGGLGGNDALWEVAYKVE VDVQNLGNSTDKFPQPLYLKH EDGKFETPIQLRGFEKVELSPGE KKTVEFELLRRDLSVWDTRQS WIVESGTYEALIGVAVNDIKTSV LFTI (SEQ ID NO:40)
--	---	--

- [0130] In certain aspects of the invention, the polypeptides and polynucleotides of the present invention are provided in an isolated form, *e.g.*, purified to homogeneity.
- [0131] The present invention also encompasses polypeptides which comprise, or alternatively consist of, an amino acid sequence which is at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% similar to the polypeptide of any of SEQ ID NOs: 21-40, 46, or 52-56 and to portions of such polypeptide with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.
- [0132] As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and conserved amino acid substitutes thereto of the polypeptide to the sequence of a second polypeptide.
- [0133] The present invention further relates to a domain, fragment, variant, derivative, or analog of the polypeptide of any of SEQ ID NOs: 21-40, 46, or 52-56.
- [0134] Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis. Therefore, the fragments may be employed as intermediates for producing the full-length polypeptides.
- [0135] Fragments of cellobiohydrolase, endoglucanase or beta-glucosidase polypeptides encompass domains, proteolytic fragments, deletion fragments and in particular, fragments of *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, or *Arabidopsis thaliana* cellobiohydrolase, endoglucanase or beta-glucosidase polypeptides which retain any specific biological activity of the cellobiohydrolase, endoglucanase or beta-glucosidase proteins. Polypeptide fragments further include any portion of the polypeptide which retains a catalytic activity of cellobiohydrolase, endoglucanase or beta-glucosidase proteins.

[0136] The variant, derivative or analog of the polypeptide of any of SEQ ID NOs: 21-40, 46, or 52-56, may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide for purification of the polypeptide or (v) one in which a fragment of the polypeptide is soluble, i.e., not membrane bound, yet still binds ligands to the membrane bound receptor. Such variants, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

[0137] The polypeptides of the present invention further include variants of the polypeptides. A "variant" of the polypeptide can be a conservative variant, or an allelic variant. As used herein, a conservative variant refers to alterations in the amino acid sequence that do not adversely affect the biological functions of the protein. A substitution, insertion or deletion is said to adversely affect the protein when the altered sequence prevents or disrupts a biological function associated with the protein. For example, the overall charge, structure or hydrophobic-hydrophilic properties of the protein can be altered without adversely affecting a biological activity. Accordingly, the amino acid sequence can be altered, for example to render the peptide more hydrophobic or hydrophilic, without adversely affecting the biological activities of the protein.

[0138] By an "allelic variant" is intended alternate forms of a gene occupying a given locus on a chromosome of an organism. Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985). Non-naturally occurring variants may be produced using art-known mutagenesis techniques. Allelic variants, though possessing a slightly different amino acid sequence than those recited above, will still have the same or similar biological functions associated with the *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus*

kawachii, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, or *Arabidopsis thaliana* cellobiohydrolase, endoglucanase or beta-glucosidase protein.

[0139] The allelic variants, the conservative substitution variants, and members of the endoglucanase, cellobiohydrolase or β -glucosidase protein families, can have an amino acid sequence having at least 75%, at least 80%, at least 90%, at least 95% amino acid sequence identity with a *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *R. flavipes*, or *Neosartorya fischeri* cellobiohydrolase, endoglucanase or beta-glucosidase amino acid sequence set forth in any one of SEQ ID NOs: 21-40, 46, or 52-56. Identity or homology with respect to such sequences is defined herein as the percentage of amino acid residues in the candidate sequence that are identical with the known peptides, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent homology, and not considering any conservative substitutions as part of the sequence identity. N-terminal, C-terminal or internal extensions, deletions, or insertions into the peptide sequence shall not be construed as affecting homology.

[0140] Thus, the proteins and peptides of the present invention include molecules comprising the amino acid sequence of SEQ ID NOs: 21-40, 46 and 52-56 or fragments thereof having a consecutive sequence of at least about 3, 4, 5, 6, 10, 15, 20, 25, 30, 35 or more amino acid residues of the *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *R. flavipes*, or *Neosartorya fischeri* cellobiohydrolase, endoglucanase or beta-glucosidase polypeptide sequences; amino acid sequence variants of such sequences wherein at least one amino acid residue has been inserted N- or C- terminal to, or within, the disclosed sequence; amino acid sequence variants of the disclosed sequences, or their fragments as defined above, that have been substituted by another residue. Contemplated variants further

include those containing predetermined mutations by, *e.g.*, homologous recombination, site-directed or PCR mutagenesis, and the corresponding proteins of other animal species, including but not limited to bacterial, fungal, insect, rabbit, rat, porcine, bovine, ovine, equine and non-human primate species, the alleles or other naturally occurring variants of the family of proteins; and derivatives wherein the protein has been covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid (for example, a detectable moiety such as an enzyme or radioisotope).

[0141] Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the cellulase polypeptides. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function.

[0142] Thus, the invention further includes *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes* or *Arabidopsis thaliana* cellobiohydrolase, endoglucanase or beta-glucosidase polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity.

[0143] The skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (*e.g.*, replacing one aliphatic amino acid with a second aliphatic amino acid), as further described below.

[0144] For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

- [0145] The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.
- [0146] The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, *Science* 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.
- [0147] As the authors state, these two strategies have revealed that proteins are often surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.
- [0148] The terms "derivative" and "analog" refer to a polypeptide differing from the *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces* sp., *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*,

Aspergillus fumigatus, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, or *Arabidopsis thaliana* cellobiohydrolase, endoglucanase or beta-glucosidase polypeptide, but retaining essential properties thereof. Generally, derivatives and analogs are overall closely similar, and, in many regions, identical to the *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, or *Arabidopsis thaliana* cellobiohydrolase, endoglucanase or beta-glucosidase polypeptides. The terms "derivative" and "analog" when referring to *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes* or *Arabidopsis thaliana* cellobiohydrolase, endoglucanase or beta-glucosidase polypeptides include any polypeptides which retain at least some of the activity of the corresponding native polypeptide, e.g., the exoglucanase activity, or the activity of the catalytic domain.

[0149] Derivatives of *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R.*

flavipes, or *Arabidopsis thaliana* cellobiohydrolase, endoglucanase or beta-glucosidase polypeptides, are polypeptides which have been altered so as to exhibit additional features not found on the native polypeptide. Derivatives can be covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid (for example, a detectable moiety such as an enzyme or radioisotope). Examples of derivatives include fusion proteins.

[0150] An analog is another form of a *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. lucknowense*, *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces sp.*, *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, or *Arabidopsis thaliana* cellobiohydrolase, endoglucanase or beta-glucosidase polypeptide of the present invention. An "analog" also retains substantially the same biological function or activity as the polypeptide of interest, e.g., functions as a cellobiohydrolase. An analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

[0151] The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide. In some particular embodiments, the polypeptide is a recombinant polypeptide.

[0152] Also provided in the present invention are allelic variants, orthologs, and/or species homologs. Procedures known in the art can be used to obtain full-length genes, allelic variants, splice variants, full-length coding portions, orthologs, and/or species homologs of genes corresponding to any of SEQ ID NOs: 1-40, using information from the sequences disclosed herein or the clones deposited with the ATCC. For example, allelic variants and/or species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for allelic variants and/or the desired homologue.

Consensus Sequence Cellulases

[0153] In some embodiments of the present invention, the host cells express at least one heterologous cellulase that is not derived from any one particular organism, but instead has an artificial amino acid sequence that is a consensus cellulase sequence. The consensus cellulase sequence can be an endoglucanase consensus sequence, a β -glucosidase consensus sequence, or a cellobiohydrolase consensus sequence.

[0154] In one particular embodiment, the heterologous cellulase is a CBH1 consensus sequence. Therefore, in one embodiment, the invention is directed to a polypeptide sequence which comprises a sequence that is at least 80%, 85%, 90%, 95%, 98% or 99% identical to the consensus CBH1 sequence of SEQ ID NO: 43. In some embodiments, the invention is directed to a polypeptide which comprises the sequence of SEQ ID NO: 43.

[0155] The invention is also directed to host cells that comprise a polypeptide sequence which comprises a sequence that is at least 80%, 85%, 90%, 95%, 98% 99% or 100% identical to the consensus CBH1 sequence of SEQ ID NO: 43. The invention further directed to host cells that comprise a polynucleotide that encodes a polypeptide sequence which comprises a sequence that is at least 80%, 85%, 90%, 95%, 98% 99% or 100% identical to the consensus CBH1 sequence of SEQ ID NO: 43. In some embodiments the host cell comprises at least one polynucleotide that encodes a polypeptide sequence which comprises a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99% or 100% identical to the consensus CBH1 sequence of SEQ ID NO: 43 and at least a second polynucleotide that encodes a heterologous cellulase. The second polynucleotide can encode a endoglucanase, a β -glucosidase, a cellobiohydrolase, an endoglucanase consensus sequence, a β -glucosidase consensus sequence, or a cellobiohydrolase consensus sequence. In some embodiments the host cell comprising the polynucleotide that encodes a polypeptide sequence which comprises a sequence that is at least 80%, 85%, 90%, 95%, 98%, 99%, or 100% identical to the consensus CBH1 sequence of SEQ ID NO: 43 is capable of producing ethanol when grown using cellulose as a carbon source.

Combinations of Cellulases

[0156] In some embodiments of the present invention the host cells express a combination of heterologous cellulases. For example, the host cell can contain at least two heterologous cellulases, at least three heterologous cellulases, at least four heterologous cellulases, at least five heterologous cellulases, at least six heterologous cellulases, at least seven heterologous cellulases, at least eight heterologous cellulases, at

least nine heterologous cellulases, at least ten heterologous cellulases, at least eleven heterologous cellulases, at least twelve heterologous cellulases, at least thirteen heterologous cellulases, at least fourteen heterologous cellulases or at least fifteen heterologous cellulases. The heterologous cellulases in the host cell can be from the same or from different species.

[0157] In some embodiments of the present invention, the host cells express a combination of heterologous cellulases which includes at least one endoglucanase, at least one β -glucosidase and at least one cellobiohydrolase. In another embodiment of the invention, the host cells express a combination of heterologous cellulases which includes at least one endoglucanase, at least one β -glucosidase and at least two cellobiohydrolases. The at least two cellobiohydrolases can be both be cellobiohydrolase I, can both be cellobiohydrolase II, or can be one cellobiohydrolase I and one cellobiohydrolase II.

[0158] In one particular embodiment of the invention, the host cells express a combination of cellulases that includes a *C. formosanus* endoglucanase I and an *S. fibuligera* β -glucosidase I. In another embodiment of the invention, the host cells express a combination of cellulases that includes a *T. emersonii* cellobiohydrolase I, and a *T. reesei* cellobiohydrolase II.

[0159] In yet another embodiment the host cells express a combination of cellulases that includes a *C. formosanus* endoglucanase I, an *S. fibuligera* β -glucosidase I, a *T. emersonii* cellobiohydrolase I, and a *C. lucknowense* cellobiohydrolase IIb. In still another embodiment, the host cells express a combination of cellulases that includes a *C. formosanus* endoglucanase I, an *S. fibuligera* β -glucosidase I, a *T. emersonii* cellobiohydrolase I, and a *T. reesei* cellobiohydrolase II. In still another embodiment, the host cells express a combination of cellulases that includes an *H. jecorina* endogluconase 2, an *S. fibuligera* β -glucosidase I, a *T. emersonii* cellobiohydrolase I, and a *T. reesei* cellobiohydrolase II. In still another embodiment, the host cells express a combination of cellulases that includes an *H. jecorina* endogluconase 2, an *S. fibuligera* β -glucosidase I, a *T. emersonii* cellobiohydrolase I, and a *C. lucknowense* cellobiohydrolase II.

Tethered and Secreted Cellulases

[0160] According to the present invention, the cellulases may be either tethered or secreted. As used herein, a protein is "tethered" to an organism's cell surface if at least one terminus of the protein is bound, covalently and/or electrostatically for example, to

the cell membrane or cell wall. It will be appreciated that a tethered protein may include one or more enzymatic regions that may be joined to one or more other types of regions at the nucleic acid and/or protein levels (e.g., a promoter, a terminator, an anchoring domain, a linker, a signaling region, etc.). While the one or more enzymatic regions may not be directly bound to the cell membrane or cell wall (e.g., such as when binding occurs via an anchoring domain), the protein is nonetheless considered a "tethered enzyme" according to the present specification.

[0161] Tethering may, for example, be accomplished by incorporation of an anchoring domain into a recombinant protein that is heterologously expressed by a cell, or by prenylation, fatty acyl linkage, glycosyl phosphatidyl inositol anchors or other suitable molecular anchors which may anchor the tethered protein to the cell membrane or cell wall of the host cell. A tethered protein maybe tethered at its amino terminal end or optionally at its carboxy terminal end.

[0162] As used herein, "secreted" means released into the extracellular milieu, for example into the media. Although tethered proteins may have secretion signals as part of their immature amino acid sequence, they are maintained as attached to the cell surface, and do not fall within the scope of secreted proteins as used herein.

[0163] As used herein, "flexible linker sequence" refers to an amino acid sequence which links two amino acid sequences, for example, a cell wall anchoring amino acid sequence with an amino acid sequence that contains the desired enzymatic activity. The flexible linker sequence allows for necessary freedom for the amino acid sequence that contains the desired enzymatic activity to have reduced steric hindrance with respect to proximity to the cell and may also facilitate proper folding of the amino acid sequence that contains the desired enzymatic activity.

[0164] In some embodiments of the present invention, the tethered cellulase enzymes are tethered by a flexible linker sequence linked to an anchoring domain. In some embodiments, the anchoring domain is of CWP2 (for carboxy terminal anchoring) or FLO1 (for amino terminal anchoring) from *S. cerevisiae*.

[0165] In some embodiments, heterologous secretion signals may be added to the expression vectors of the present invention to facilitate the extra-cellular expression of cellulase proteins. In some embodiments, the heterologous secretion signal is the secretion signal from *T. reesei* Xyn2.

Fusion Proteins Comprising Cellulases

- [0166] The present invention also encompasses fusion proteins. For example, the fusion proteins can be a fusion of a heterologous cellulase and a second peptide. The heterologous cellulase and the second peptide can be fused directly or indirectly, for example, through a linker sequence. The fusion protein can comprise for example, a second peptide that is N-terminal to the heterologous cellulase and/or a second peptide that is C-terminal to the heterologous cellulase. Thus, in certain embodiments, the polypeptide of the present invention comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises a heterologous cellulase.
- [0167] According to the present invention, the fusion protein can comprise a first and second polypeptide wherein the first polypeptide comprises a heterologous cellulase and the second polypeptide comprises a signal sequence. According to another embodiment, the fusion protein can comprise a first and second polypeptide, wherein the first polypeptide comprises a heterologous cellulase and the second polypeptide comprises a polypeptide used to facilitate purification or identification or a reporter peptide. The polypeptide used to facilitate purification or identification or the reporter peptide can be, for example, a HIS-tag, a GST-tag, an HA-tag, a FLAG-tag, a MYC-tag, or a fluorescent protein.
- [0168] According to yet another embodiment, the fusion protein can comprise a first and second polypeptide, wherein the first polypeptide comprises a heterologous cellulase and the second polypeptide comprises an anchoring peptide. In some embodiments, the anchoring domain is of CWP2 (for carboxy terminal anchoring) or FLO1 (for amino terminal anchoring) from *S. cerevisiae*.
- [0169] According to yet another embodiment, the fusion protein can comprise a first and second polypeptide, wherein the first polypeptide comprises a heterologous cellulase and the second polypeptide comprises a cellulose binding module (CBM). In some embodiments, the CBM is from, for example, *T. reesei* Cbh1 or Cbh2, from *H. grisea* Cbh1, or from *C. lucknowense* Cbh2b. In some particular embodiments, the CBM is fused to a cellobiohydrolase. In one particular embodiment, the fusion protein comprises a first and second polypeptide, wherein the first polypeptide comprises a heterologous cellobiohydrolase and the second polypeptide comprises a CBM. In yet another particular embodiment, the cellobiohydrolase is *T. emersonii* cellobiohydrolase I and the

CBM is a *T. reesei* cellobiohydrolase CBM. In yet another particular embodiment, the cellobiohydrolase is *T. emersonii* cellobiohydrolase I and the CBM is a *H. grisea* cellobiohydrolase CBM. In some embodiments, the CBM of *H. grisea* comprises amino acids 492-525 of SEQ ID NO: 21.

[0170] In certain embodiments, the polypeptide of the present invention encompasses a fusion protein comprising a first polypeptide and a second polypeptide, wherein the first polypeptide is a cellobiohydrolase, and the second polypeptide is a domain or fragment of a cellobiohydrolase. In certain embodiments, the polypeptide of the present invention encompasses a fusion protein comprising a first polypeptide, where the first polypeptide is a *T. emersonii* Cbh1, *H. grisea* Cbh1, *T. aurantiacusi* Cbh1, *T. emersonii* Cbh2, *T. reesei* Cbh1, *T. reesei* Cbh2, *C. lucknowense* Cbh2b, or domain, fragment, variant, or derivative thereof, and a second polypeptide, where the second polypeptide is a *T. emersonii* Cbh1, *H. grisea* Cbh1, or *T. aurantiacusi* Cbh1, *T. emersonii* Cbh2, *T. reesei* Cbh1 or *T. reesei* Cbh2, *C. lucknowense* Cbh2b, or domain, fragment, variant, or derivative thereof. In particular embodiments the first polypeptide is *T. emersonii* Cbh1 and the second polynucleotide is a CBM from *T. reesei* Cbh1 or Cbh2 or from *C. lucknowense* Cbh2b. In additional embodiments, the first polypeptide is either N-terminal or C-terminal to the second polypeptide. In certain other embodiments, the first polypeptide and/or the second polypeptide are encoded by codon-optimized polynucleotides, for example, polynucleotides codon-optimized for *S. cerevisiae* or *Kluyveromyces*. In particular embodiments, the first polynucleotide is a codon-optimized *T. emersonii* cbh1 and the second polynucleotide encodes for a codon-optimized CBM from *T. reesei* Cbh1 or Cbh2. In another particular embodiment, the first polynucleotide is a codon-optimized *T. emersonii* cbh1 and the second polynucleotide encodes for a codon-optimized CBM from *C. lucknowense* or Cbh2b.

[0171] In certain other embodiments, the first polypeptide and the second polypeptide are fused via a linker sequence. The linker sequence can, in some embodiments, be encoded by a codon-optimized polynucleotide. (Codon-optimized polynucleotides are described in more detail below.) An amino acid sequence corresponding to a codon-optimized linker 1 according to the invention is a flexible linker - strep tag - TEV site - FLAG - flexible linker fusion and corresponds to GGGGSGGGGS AWHPQFGG ENLYFQG DYKDDDK GGGGSGGGGS (SEQ ID NO:57)

[0172] The DNA sequence is as follows:

GGAGGAGGTGGTTCAGGAGGTGGTGGGTCTGCTTGGCAT
CCACAATTTGGAGGAGGCGGTGGTGAAAATCTGTATTTC
CAGGGAGGCGGAGGTGATTACAAGGATGACGACAAAGG
AGGTGGTGGATCAGGAGGTGGTGGCTCC (SEQ ID NO:41)

[0173] An amino acid sequence corresponding to optimized linker 2 is a flexible linker - strep tag -linker- TEV site - flexible linker and corresponds to GGGGSGGGGS WSH PQFEK GG ENLYFQG GGGGSGGGGS (SEQ ID NO:58). The DNA sequence is as follows:

[0174] ggtggcgggtgatctggaggaggcggttcttggtctcaccacaatttgaaaagggtggagaaaacttgactttcaa
ggcgggtggtggaggttctggcggaggtggctccggctca (SEQ ID NO:42)

Co-Cultures

[0175] The present invention is also directed to co-cultures comprising at least two yeast host cells wherein the at least two yeast host cells each comprise an isolated polynucleotide encoding a heterologous cellulase. As used herein, "co-culture" refers to growing two different strains or species of host cells together in the same vessel. In some embodiments of the invention, at least one host cell of the co-culture comprises a heterologous polynucleotide comprising a nucleic acid which encodes an endoglucanase, at least one host cell of the co-culture comprises a heterologous polynucleotide comprising a nucleic acid which encodes a β -glucosidase and at least one host cell comprises a heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase. In a further embodiment, the co-culture further comprises a host cell comprising a heterologous polynucleotide comprising a nucleic acid which encodes a second cellobiohydrolase.

[0176] The co-culture can comprise two or more strains of yeast host cells and the heterologous cellulases can be expressed in any combination in the two or more strains of host cells. For example, according to the present invention, the co-culture can comprise two strains: one strain of host cells that expresses an endoglucanase and a second strain of host cells that expresses a β -glucosidase, a cellobiohydrolase and a second cellobiohydrolase. According to the present invention, the co-culture can also comprise four strains: one strain of host cells which expresses an endoglucanase, one strain of host cells that expresses a β -glucosidase, one strain of host cells which expresses a first cellobiohydrolase, and one strain of host cells which expressess a second

cellobiohydrolase. Similarly, the co-culture can comprise one strain of host cells that expresses two cellulases, for example an endoglucanase and a beta-glucosidase and a second strain of host cells that expresses one or more cellulases, for example one or more cellobiohydrolases. The co-culture can, in addition to the at least two host cells comprising heterologous cellulases, also include other host cells which do not comprise heterologous cellulases.

[0177] The various host cell strains in the co-culture can be present in equal numbers, or one strain or species of host cell can significantly outnumber another second strain or species of host cells. For example, in a co-culture comprising two strains or species of host cells the ratio of one host cell to another can be about 1:1, 1:2, 1:3, 1:4, 1:5, 1:10, 1:100, 1:500 or 1:1000. Similarly, in a co-culture comprising three or more strains or species of host cells, the strains or species of host cells may be present in equal or unequal numbers.

[0178] The co-cultures of the present invention can include tethered cellulases, secreted cellulases or both tethered and secreted cellulases. For example, in some embodiments of the invention, the co-culture comprises at least one yeast host cell comprising a polynucleotide encoding a secreted heterologous cellulase. In another embodiment, the co-culture comprises at least one yeast host cell comprising a polynucleotide encoding a tethered heterologous cellulase. In one embodiment, all of the heterologous cellulases in the co-culture are secreted, and in another embodiment, all of the heterologous cellulases in the co-culture are tethered. In addition, other cellulases, such as externally added cellulases may be present in the co-culture.

Polynucleotides Encoding Heterologous Cellulases

[0179] The present invention also includes isolated polynucleotides encoding cellulases of the present invention. Thus, the polynucleotides of the invention can encode endoglucanases or exoglucanases. The polynucleotides can encode endoglucanases, β -glucosidases or cellobiohydrolases.

[0180] In some particular embodiments of the invention, the polynucleotide encodes an endoglucanase which is an endo-1,4- β -glucanase. In particular embodiments, the polynucleotide encodes an endoglucanase I from *Trichoderma reesei*. In certain other embodiments, the endoglucanase is encoded by a polynucleotide comprising a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or

100% identical to SEQ ID NO:19. In particular embodiments, the polynucleotide encodes an endoglucanase I from *C. formosanus*. In certain other embodiments, the endoglucanase is encoded by a polynucleotide comprising a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:11. In particular embodiments, the polynucleotide encodes an endoglucanase I from *Trichoderma reesei*. In certain other embodiments, the endoglucanase is encoded by a polynucleotide comprising a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:19. In particular embodiments, the polynucleotide encodes an endoglucanase 2 from *H. jecorina*. In certain other embodiments, the endoglucanase is encoded by a polynucleotide comprising a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:54.

[0181] In certain embodiments, the polynucleotide encodes a β -glucosidase I or a β -glucosidase II isoform, paralogue or orthologue. In certain embodiments of the present invention the polynucleotide encodes a β -glucosidase derived from *Saccharomycopsis fibuligera*. In particular embodiments, the β -glucosidase is encoded by a polynucleotide comprising a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:20.

[0182] In certain embodiments of the invention, the polynucleotide encodes a cellobiohydrolase I and/or an cellobiohydrolase II isoform, paralogue or orthologue. In particular embodiments of the present invention, the polynucleotide encodes the cellobiohydrolase I or II from *Trichoderma reesei*. In particular embodiments of the present invention, the polynucleotide encodes the cellobiohydrolase I or II from *Trichoderma emersonii*. In another embodiment, the cellobiohydrolase is encoded by a polynucleotide comprising a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:7 or SEQ ID NO:8. In particular embodiments of the present invention, the polynucleotide encodes a cellobiohydrolase from *C. lucknowense*. In another embodiment, the cellobiohydrolase is encoded by a polynucleotide comprising a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or 100% identical to SEQ ID NO:5.

[0183] In further embodiments the polynucleotide is a polypeptide comprising a sequence at least about 70, about 80, about 90, about 95, about 96, about 97, about 98, about 99, or

100% identical to a nucleotide sequence listed in Table 1. In certain aspects the polynucleotide can encode an endoglucanase, cellobiohydrolase or β -glucosidase derived from, for example, a fungal, bacterial, protozoan or termite source.

[0184] In certain aspects, the present invention relates to a polynucleotide comprising a nucleic acid encoding a functional or structural domain of *T. emersonii*, *H. grisea*, *T. aurantiacus*, *C. lucknowense* or *T. reesei* Cbh1 or Cbh2. For example, the domains of *T. reesei* Cbh 1 include, without limitation: (1) a signal sequence, from amino acid 1 to 33 of SEQ ID NO: 27; (2) a catalytic domain (CD) from about amino acid 41 to about amino acid 465 of SEQ ID NO: 27; and (3) a cellulose binding module (CBM) from about amino acid 503 to about amino acid 535 of SEQ ID NO: 27. The domains of *T. reesei* Cbh 2 include, without limitation: (1) a signal sequence, from amino acid 1 to 33 of SEQ ID NO: 27; (2) a catalytic domain (CD) from about amino acid 145 to about amino acid 458 of SEQ ID NO: 27; and (3) a cellulose binding module (CBM) from about amino acid 52 to about amino acid 83 of SEQ ID NO: 27.

[0185] The present invention also encompasses an isolated polynucleotide comprising a nucleic acid that is at least about 70%, 75%, or 80% identical, at least about 90% to about 95% identical, or at least about 96%, 97%, 98%, 99% or 100% identical to a nucleic acid encoding a *T. emersonii*, *H. grisea*, *T. aurantiacus*, *C. lucknowense* or *T. reesei* Cbh1 or Cbh2 domain, as described above.

[0186] The present invention also encompasses variants of the cellulase genes, as described above. Variants may contain alterations in the coding regions, non-coding regions, or both. Examples are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. In certain embodiments, nucleotide variants are produced by silent substitutions due to the degeneracy of the genetic code. In further embodiments, *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense* *R. speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces* sp., *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium*

thermophilum, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, and *Arabidopsis thaliana* cellulase polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host. Codon-optimized polynucleotides of the present invention are discussed further below.

[0187] The present invention also encompasses an isolated polynucleotide encoding a fusion protein. In certain embodiments, the nucleic acid encoding a fusion protein comprises a first polynucleotide encoding for a *T. emersonii* cbh1, *H. grisea* cbh1, *T. aurantiacusi* cbh1 or *T. emersonii* cbh1 and a second polynucleotide encoding for the CBM domain of *T. reesei* cbh1 or *T. reesei* cbh2 or *C. lucknowense* cbh2b. In particular embodiments of the nucleic acid encoding a fusion protein, the first polynucleotide encodes *T. emersonii* cbh1 and the second polynucleotide encodes for a CBM from *T. reesei* Cbh1 or Cbh2.

[0188] In further embodiments, the first and second polynucleotides are in the same orientation, or the second polynucleotide is in the reverse orientation of the first polynucleotide. In additional embodiments, the first polynucleotide encodes a polypeptide that is either N-terminal or C-terminal to the polypeptide encoded by the second polynucleotide. In certain other embodiments, the first polynucleotide and/or the second polynucleotide are encoded by codon-optimized polynucleotides, for example, polynucleotides codon-optimized for *S. cerevisiae*, *Kluyveromyces* or for both *S. cerevisiae* and *Kluyveromyces*. In particular embodiments of the nucleic acid encoding a fusion protein, the first polynucleotide is a codon-optimized *T. emersonii* cbh1 and the second polynucleotide encodes for a codon-optimized CBM from *T. reesei* Cbh1 or Cbh2.

[0189] Also provided in the present invention are allelic variants, orthologs, and/or species homologs. Procedures known in the art can be used to obtain full-length genes, allelic variants, splice variants, full-length coding portions, orthologs, and/or species homologs of genes corresponding to any of SEQ ID NOs: 1-20, using information from the sequences disclosed herein or the clones deposited with the ATCC. For example, allelic variants and/or species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for allelic variants and/or the desired homologue.

[0190] By a nucleic acid having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that

the nucleotide sequence of the nucleic acid is identical to the reference sequence except that the nucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the particular polypeptide. In other words, to obtain a nucleic acid having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown of any of SEQ ID NOs:1-20, or any fragment or domain specified as described herein.

[0191] As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence or polypeptide of the present invention can be determined conventionally using known computer programs. A method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245.) In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

[0192] If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent

identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

[0193] For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

[0194] Some embodiments of the invention encompass a nucleic acid molecule comprising at least 10, 20, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, or 800 consecutive nucleotides or more of any of SEQ ID NOs:1-20, or domains, fragments, variants, or derivatives thereof.

[0195] The polynucleotide of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double stranded or single-stranded, and if single stranded can be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature polypeptide can be identical to the coding sequence encoding SEQ ID NO:21-40, 46, or 52-56, or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature polypeptide as the DNA of any one of SEQ ID NOs:21-40, 46, or 52-56.

- [0196] In certain embodiments, the present invention provides an isolated polynucleotide comprising a nucleic acid fragment which encodes at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 95, or at least 100 or more contiguous amino acids of SEQ ID NOs: 21-40, 46, or 52-56.
- [0197] The polynucleotide encoding for the mature polypeptide of SEQ ID NOs: 21-40, 46, or 52-56 or may include: only the coding sequence for the mature polypeptide; the coding sequence of any domain of the mature polypeptide; and the coding sequence for the mature polypeptide (or domain-encoding sequence) together with non coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature polypeptide.
- [0198] Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only sequences encoding for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequences.
- [0199] In further aspects of the invention, nucleic acid molecules having sequences at least about 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequences disclosed herein, encode a polypeptide having cellobiohydrolase ("Cbh"), endoglucanase ("Eg") or beta-gluconase ("Bgl") functional activity. By "a polypeptide having Cbh, Eg or Bgl functional activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to a functional activity of the Cbh, Eg or Bgl polypeptides of the present invention, as measured, for example, in a particular biological assay. For example, a Cbh, Eg or Bgl functional activity can routinely be measured by determining the ability of a Cbh, Eg or Bgl polypeptide to hydrolyze cellulose, or by measuring the level of Cbh, Eg or Bgl activity.
- [0200] Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large portion of the nucleic acid molecules having a sequence at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the nucleic acid sequence of any of SEQ ID NOs: 1-20, or fragments thereof, will encode polypeptides having Cbh, Eg or Bgl functional activity. In fact, since degenerate variants of any of these nucleotide sequences all encode the same polypeptide, in many instances, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are

not degenerate variants, a reasonable number will also encode a polypeptide having Cbh, Eg or Bgl functional activity.

[0201] The polynucleotides of the present invention also comprise nucleic acids encoding a *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces* sp., *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes* or *Arabidopsis thaliana* cellulase, or domain, fragment, variant, or derivative thereof, fused to a polynucleotide encoding a marker sequence which allows for detection of the polynucleotide of the present invention. In one embodiment of the invention, expression of the marker is independent from expression of the cellulase. The marker sequence may be a yeast selectable marker selected from the group consisting of URA3, HIS3, LEU2, TRP1, LYS2 or ADE2. Casey, G.P. *et al.*, "A convenient dominant selection marker for gene transfer in industrial strains of *Saccharomyces* yeast: SMR1 encoded resistance to the herbicide sulfometuron methyl," *J. Inst. Brew.* 94:93-97 (1988).

Codon Optimized Polynucleotides

[0202] According to one embodiment of the invention, the polynucleotides encoding heterologous cellulases can be codon-optimized. As used herein the term "codon-optimized coding region" means a nucleic acid coding region that has been adapted for expression in the cells of a given organism by replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that organism.

[0203] In general, highly expressed genes in an organism are biased towards codons that are recognized by the most abundant tRNA species in that organism. One measure of this bias is the "codon adaptation index" or "CAI," which measures the extent to which the codons used to encode each amino acid in a particular gene are those which occur most frequently in a reference set of highly expressed genes from an organism.

[0204] The CAI of codon optimized sequences of the present invention corresponds to between about 0.8 and 1.0, between about 0.8 and 0.9, or about 1.0. A codon optimized sequence may be further modified for expression in a particular organism, depending on that organism's biological constraints. For example, large runs of "As" or "Ts" (e.g., runs greater than 4, 4, 5, 6, 7, 8, 9, or 10 consecutive bases) can be removed from the sequences if these are known to effect transcription negatively. Furthermore, specific restriction enzyme sites may be removed for molecular cloning purposes. Examples of such restriction enzyme sites include PacI, AscI, BamHI, BglII, EcoRI and XhoI. Additionally, the DNA sequence can be checked for direct repeats, inverted repeats and mirror repeats with lengths of ten bases or longer, which can be modified manually by replacing codons with "second best" codons, *i.e.*, codons that occur at the second highest frequency within the particular organism for which the sequence is being optimized.

[0205] Deviations in the nucleotide sequence that comprise the codons encoding the amino acids of any polypeptide chain allow for variations in the sequence coding for the gene. Since each codon consists of three nucleotides, and the nucleotides comprising DNA are restricted to four specific bases, there are 64 possible combinations of nucleotides, 61 of which encode amino acids (the remaining three codons encode signals ending translation). The "genetic code" which shows which codons encode which amino acids is reproduced herein as Table 2. As a result, many amino acids are designated by more than one codon. For example, the amino acids alanine and proline are coded for by four triplets, serine and arginine by six, whereas tryptophan and methionine are coded by just one triplet. This degeneracy allows for DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA.

TABLE 2: The Standard Genetic Code

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC " TAA Ter TAG Ter	TGT Cys (C) TGC TGA Ter TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " ATG Met (M)	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys (K) AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

[0206] Many organisms display a bias for use of particular codons to code for insertion of a particular amino acid in a growing peptide chain. Codon preference or codon bias, differences in codon usage between organisms, is afforded by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, inter alia, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

[0207] Given the large number of gene sequences available for a wide variety of animal, plant and microbial species, it is possible to calculate the relative frequencies of codon usage. Codon usage tables are readily available

and these tables can be adapted in a number of ways. See Nakamura, Y., *et al.* "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" *Nucl. Acids Res.* 28:292 (2000). Codon usage tables for yeast, calculated from GenBank Release 128.0 [15 February 2002], are reproduced below as Table 3. This table uses mRNA nomenclature, and so instead of thymine (T) which is found in DNA, the tables use uracil (U) which is found in RNA. The Table has been adapted so that frequencies are calculated for each amino acid, rather than for all 64 codons.

TABLE 3: Codon Usage Table for *Saccharomyces cerevisiae* Genes

Amino Acid	Codon	Number	Frequency per hundred
Phe	UUU	170666	26.1
Phe	UUC	120510	18.4
Total			
Leu	UUA	170884	26.2
Leu	UUG	177573	27.2
Leu	CUU	80076	12.3
Leu	CUC	35545	5.4
Leu	CUA	87619	13.4
Leu	CUG	68494	10.5
Total			
Ile	AUU	196893	30.1
Ile	AUC	112176	17.2
Ile	AUA	116254	17.8
Total			
Met	AUG	136805	20.9
Total			
Val	GUU	144243	22.1
Val	GUC	76947	11.8
Val	GUA	76927	11.8
Val	GUG	70337	10.8
Total			
Ser	UCU	153557	23.5
Ser	UCC	92923	14.2
Ser	UCA	122028	18.7
Ser	UCG	55951	8.6
Ser	AGU	92466	14.2

Amino Acid	Codon	Number	Frequency per hundred
Ser	AGC	63726	9.8
Total			
Pro	CCU	88263	13.5
Pro	CCC	44309	6.8
Pro	CCA	119641	18.3
Pro	CCG	34597	5.3
Total			
Thr	ACU	132522	20.3
Thr	ACC	83207	12.7
Thr	ACA	116084	17.8
Thr	ACG	52045	8.0
Total			
Ala	GCU	138358	21.2
Ala	GCC	82357	12.6
Ala	GCA	105910	16.2
Ala	GCG	40358	6.2
Total			
Tyr	UAU	122728	18.8
Tyr	UAC	96596	14.8
Total			
His	CAU	89007	13.6
His	CAC	50785	7.8
Total			
Gln	CAA	178251	27.3
Gln	CAG	79121	12.1
Total			
Asn	AAU	233124	35.7
Asn	AAC	162199	24.8
Total			
Lys	AAA	273618	41.9
Lys	AAG	201361	30.8
Total			
Asp	GAU	245641	37.6
Asp	GAC	132048	20.2
Total			

Amino Acid	Codon	Number	Frequency per hundred
Glu	GAA	297944	45.6
Glu	GAG	125717	19.2
Total			
Cys	UGU	52903	8.1
Cys	UGC	31095	4.8
Total			
Trp	UGG	67789	10.4
Total			
Arg	CGU	41791	6.4
Arg	CGC	16993	2.6
Arg	CGA	19562	3.0
Arg	CGG	11351	1.7
Arg	AGA	139081	21.3
Arg	AGG	60289	9.2
Total			
Gly	GGU	156109	23.9
Gly	GGC	63903	9.8
Gly	GGA	71216	10.9
Gly	GGG	39359	6.0
Total			
Stop	UAA	6913	1.1
Stop	UAG	3312	0.5
Stop	UGA	4447	0.7

[0208] By utilizing this or similar tables, one of ordinary skill in the art can apply the frequencies to any given polypeptide sequence, and produce a nucleic acid fragment of a codon-optimized coding region which encodes the polypeptide, but which uses codons optimal for a given species. Codon-optimized coding regions can be designed by various different methods.

[0209] In one method, a codon usage table is used to find the single most frequent codon used for any given amino acid, and that codon is used each time that particular amino acid appears in the polypeptide sequence. For example, referring to Table 3 above, for

leucine, the most frequent codon is UUG, which is used 27.2% of the time. Thus all the leucine residues in a given amino acid sequence would be assigned the codon UUG.

[0210] In another method, the actual frequencies of the codons are distributed randomly throughout the coding sequence. Thus, using this method for optimization, if a hypothetical polypeptide sequence had 100 leucine residues, referring to Table 3 for frequency of usage in the *S. cerevisiae*, about 5, or 5% of the leucine codons would be CUC, about 11, or 11% of the leucine codons would be CUG, about 12, or 12% of the leucine codons would be CUU, about 13, or 13% of the leucine codons would be CUA, about 26, or 26% of the leucine codons would be UUA, and about 27, or 27% of the leucine codons would be UUG.

[0211] These frequencies would be distributed randomly throughout the leucine codons in the coding region encoding the hypothetical polypeptide. As will be understood by those of ordinary skill in the art, the distribution of codons in the sequence can vary significantly using this method; however, the sequence always encodes the same polypeptide.

[0212] When using the methods above, the term "about" is used precisely to account for fractional percentages of codon frequencies for a given amino acid. As used herein, "about" is defined as one amino acid more or one amino acid less than the value given. The whole number value of amino acids is rounded up if the fractional frequency of usage is 0.50 or greater, and is rounded down if the fractional frequency of use is 0.49 or less. Using again the example of the frequency of usage of leucine in human genes for a hypothetical polypeptide having 62 leucine residues, the fractional frequency of codon usage would be calculated by multiplying 62 by the frequencies for the various codons. Thus, 7.28 percent of 62 equals 4.51 UUA codons, or "about 5," i.e., 4, 5, or 6 UUA codons, 12.66 percent of 62 equals 7.85 UUG codons or "about 8," i.e., 7, 8, or 9 UUG codons, 12.87 percent of 62 equals 7.98 CUU codons, or "about 8," i.e., 7, 8, or 9 CUU codons, 19.56 percent of 62 equals 12.13 CUC codons or "about 12," i.e., 11, 12, or 13 CUC codons, 7.00 percent of 62 equals 4.34 CUA codons or "about 4," i.e., 3, 4, or 5 CUA codons, and 40.62 percent of 62 equals 25.19 CUG codons, or "about 25," i.e., 24, 25, or 26 CUG codons.

[0213] Randomly assigning codons at an optimized frequency to encode a given polypeptide sequence, can be done manually by calculating codon frequencies for each

amino acid, and then assigning the codons to the polypeptide sequence randomly. Additionally, various algorithms and computer software programs are readily available to those of ordinary skill in the art. For example, the "EditSeq" function in the Lasergene Package, available from DNASTar, Inc., Madison, WI, the backtranslation function in the VectorNTI Suite, available from InforMax, Inc., Bethesda, MD, and the "backtranslate" function in the GCG--Wisconsin Package, available from Accelrys, Inc., San Diego, CA. In addition, various resources are publicly available to codon-optimize coding region sequences. Constructing a rudimentary algorithm to assign codons based on a given frequency can also easily be accomplished with basic mathematical functions by one of ordinary skill in the art.

[0214] A number of options are available for synthesizing codon optimized coding regions designed by any of the methods described above, using standard and routine molecular biological manipulations well known to those of ordinary skill in the art. In one approach, a series of complementary oligonucleotide pairs of 80-90 nucleotides each in length and spanning the length of the desired sequence is synthesized by standard methods. These oligonucleotide pairs are synthesized such that upon annealing, they form double stranded fragments of 80-90 base pairs, containing cohesive ends, e.g., each oligonucleotide in the pair is synthesized to extend 3, 4, 5, 6, 7, 8, 9, 10, or more bases beyond the region that is complementary to the other oligonucleotide in the pair. The single-stranded ends of each pair of oligonucleotides is designed to anneal with the single-stranded end of another pair of oligonucleotides. The oligonucleotide pairs are allowed to anneal, and approximately five to six of these double-stranded fragments are then allowed to anneal together via the cohesive single stranded ends, and then they ligated together and cloned into a standard bacterial cloning vector, for example, a TOPO[®] vector available from Invitrogen Corporation, Carlsbad, CA. The construct is then sequenced by standard methods. Several of these constructs consisting of 5 to 6 fragments of 80 to 90 base pair fragments ligated together, i.e., fragments of about 500 base pairs, are prepared, such that the entire desired sequence is represented in a series of plasmid constructs. The inserts of these plasmids are then cut with appropriate restriction

enzymes and ligated together to form the final construct. The final construct is then cloned into a standard bacterial cloning vector, and sequenced. Additional methods would be immediately apparent to the skilled artisan. In addition, gene synthesis is readily available commercially.

[0215] In certain embodiments, an entire polypeptide sequence, or fragment, variant, or derivative thereof is codon optimized by any of the methods described herein. Various desired fragments, variants or derivatives are designed, and each is then codon-optimized individually. In addition, partially codon-optimized coding regions of the present invention can be designed and constructed. For example, the invention includes a nucleic acid fragment of a codon-optimized coding region encoding a polypeptide in which at least about 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the codon positions have been codon-optimized for a given species. That is, they contain a codon that is preferentially used in the genes of a desired species, *e.g.*, a yeast species such as *Saccharomyces cerevisiae* or *Kluveromyces*, in place of a codon that is normally used in the native nucleic acid sequence.

[0216] In additional embodiments, a full-length polypeptide sequence is codon-optimized for a given species resulting in a codon-optimized coding region encoding the entire polypeptide, and then nucleic acid fragments of the codon-optimized coding region, which encode fragments, variants, and derivatives of the polypeptide are made from the original codon-optimized coding region. As would be well understood by those of ordinary skill in the art, if codons have been randomly assigned to the full-length coding region based on their frequency of use in a given species, nucleic acid fragments encoding fragments, variants, and derivatives would not necessarily be fully codon optimized for the given species. However, such sequences are still much closer to the codon usage of the desired species than the native codon usage. The advantage of this approach is that synthesizing codon-optimized nucleic acid fragments encoding each fragment, variant, and derivative of a given polypeptide, although routine, would be time consuming and would result in significant expense.

[0217] The codon-optimized coding regions can be, for example, versions encoding a cellobiohydrolase, endoglucanase or beta-glucosidase from *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M.*

darwinensis, *N. walkeri*, *S. fibuligera*, *C. luckowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces* sp., *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, or *Arabidopsis thaliana*, or domains, fragments, variants, or derivatives thereof.

[0218] Codon optimization is carried out for a particular species by methods described herein, for example, in certain embodiments codon-optimized coding regions encoding polypeptides of *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces* sp., *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, or *Arabidopsis thaliana* cellulases, or domains, fragments, variants, or derivatives thereof are optimized according to yeast codon usage, e.g., *Saccharomyces cerevisiae*, *Kluyveromyces lactis* and/or *Kluyveromyces marxianus*. Also provided are polynucleotides, vectors, and other expression constructs comprising codon-optimized coding regions encoding polypeptides of *H. grisea*, *T. aurantiacus*, *T. emersonii*, *T. reesei*, *C. lacteus*, *C. formosanus*, *N. takasagoensis*, *C. acinaciformis*, *M. darwinensis*, *N. walkeri*, *S. fibuligera*, *C. luckowense*, *R. speratus*, *Thermobifida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Aspergillus kawachii*, *Heterodera schachtii*, *H. jecorina*, *Orpinomyces* sp., *Irpex lacteus*, *Acremonium thermophilum*, *Neosartorya fischeri*, *Chaetomium globosum*, *Chaetomium thermophilum*, *Aspergillus fumigatus*, *Aspergillus terreus*, *Neurospora Crassa*, *R. flavipes*, or *Arabidopsis thaliana* cellulases or domains, fragments, variants, or

derivatives thereof, and various methods of using such polynucleotides, vectors and other expression constructs.

[0219] In certain embodiments described herein, a codon-optimized coding region encoding any of SEQ ID NOs:21-40, 46, or 52-56 or domain, fragment, variant, or derivative thereof, is optimized according to codon usage in yeast (*Saccharomyces cerevisiae*, *Kluyveromyces lactis* or *Kluyveromyces marxianus*). In some embodiments, the sequences are codon-optimized specifically for expression in *Saccharomyces cerevisiae*. In some embodiments, the sequences are codon-optimized for expression in *Kluyveromyces*. In some embodiments, a sequence is simultaneously codon-optimized for optimal expression in both *Saccharomyces cerevisiae* and in *Kluyveromyces*. Alternatively, a codon-optimized coding region encoding any of SEQ ID NOs: 21-40, 46, or 52-56 may be optimized according to codon usage in any plant, animal, or microbial species.

Vectors and Methods of Using Vectors in Host Cells

[0220] The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

[0221] Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, 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. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

[0222] The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, *e.g.*, derivatives of SV40; bacterial plasmids; and yeast plasmids. However, any other vector may be used as long as it is replicable and viable in the host.

[0223] The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

[0224] The DNA sequence in the expression vector is operatively associated with an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. Representative examples of such promoters are as follows:

Gene	Organism	Systematic name	Reason for use/benefits
PGK1	<i>S. cerevisiae</i>	YCR012W	Strong constitutive promoter
ENO1	<i>S. cerevisiae</i>	YGR254W	Strong constitutive promoter
TDH3	<i>S. cerevisiae</i>	YGR192C	Strong constitutive promoter
TDH2	<i>S. cerevisiae</i>	YJR009C	Strong constitutive promoter
TDH1	<i>S. cerevisiae</i>	YJL052W	Strong constitutive promoter
ENO2	<i>S. cerevisiae</i>	YHR174W	Strong constitutive promoter
GPM1	<i>S. cerevisiae</i>	YKL152C	Strong constitutive promoter
TPI1	<i>S. cerevisiae</i>	YDR050C	Strong constitutive promoter

[0225] Additionally, promoter sequences from stress and starvation response genes are useful in the present invention. In some embodiments, promoter regions from the *S. cerevisiae* genes *GAC1*, *GET3*, *GLC7*, *GSH1*, *GSH2*, *HSF1*, *HSP12*, *LCB5*, *LRE1*, *LSP1*, *NBP2*, *PIL1*, *PIM1*, *SGT2*, *SLG1*, *WHI2*, *WSC2*, *WSC3*, *WSC4*, *YAP1*, *YDC1*, *HSP104*, *HSP26*, *ENA1*, *MSN2*, *MSN4*, *SIP2*, *SIP4*, *SIP5*, *DPL1*, *IRS4*, *KOG1*, *PEP4*, *HAP4*, *PRB1*, *TAX4*, *ZPR1*, *ATG1*, *ATG2*, *ATG10*, *ATG11*, *ATG12*, *ATG13*, *ATG14*, *ATG15*, *ATG16*, *ATG17*, *ATG18*, and *ATG19* may be used. Any suitable promoter to drive gene expression in the host cells of the invention may be used. Additionally the *E. coli*, lac or trp, and other promoters known to control expression of genes in prokaryotic or lower eukaryotic cells can be used.

[0226] In addition, the expression vectors may contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as *URA3*, *HIS3*, *LEU2*, *TRP1*, *LYS2* or *ADE2*, dihydrofolate reductase, neomycin (G418) resistance or zeocin resistance for eukaryotic cell culture, or tetracycline or ampicillin resistance in *E. coli*.

[0227] The expression vector may also contain a ribosome binding site for translation initiation and/or a transcription terminator. The vector may also include appropriate sequences for amplifying expression, or may include additional regulatory regions.

[0228] The vector containing the appropriate DNA sequence as herein, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

[0229] Thus, in certain aspects, the present invention relates to host cells containing the above-described constructs. The host cell can be a host cell as described elsewhere in the application. The host cell can be, for example, a lower eukaryotic cell, such as a yeast cell, *e.g.*, *Saccharomyces cerevisiae* or *Kluyveromyces*, or the host cell can be a prokaryotic cell, such as a bacterial cell.

[0230] As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli*, *Streptomyces*, *Salmonella typhimurium*; thermophilic or mesophilic bacteria; fungal cells, such as yeast; and plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

[0231] Appropriate fungal hosts include yeast. In certain aspects of the invention the yeast is selected from the group consisting of *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Schizosaccharomyces pombe*, *Candida albicans*, *Pichia pastoris*, *Pichia stipitis*, *Yarrowia lipolytica*, *Hansenula polymorpha*, *Phaffia rhodozyma*, *Candida utilis*, *Arxula adeninivorans*, *Debaryomyces hansenii*, *Debaryomyces polymorphus*, *Schwanniomyces occidentalis*, *Issatchenkia orientalis*, *Kluyveromyces marxianus*, *Blakeslea*, *Candida*, *Cryptococcus*, *Cunninghamella*, *Lipomyces*, *Mortierella*, *Mucor*, *Phycomces*, *Pythium*, *Rhodosporidium*, *Rhodotorula*, *Trichosporon* and *Yarrowia*.

Methods of Using Host Cells to Produce Ethanol

[0232] The present invention is also directed to use of host cells and co-cultures to produce ethanol from cellulosic substrates. Such methods can be accomplished, for example, by contacting a cellulosic substrate with a host cell or a co-culture of the present invention.

[0233] Numerous cellulosic substrates can be used in accordance with the present invention. Substrates for cellulose activity assays can be divided into two categories, soluble and insoluble, based on their solubility in water. Soluble substrates include cellodextrins or derivatives, carboxymethyl cellulose (CMC), or hydroxyethyl cellulose (HEC). Insoluble substrates include crystalline cellulose, microcrystalline cellulose

(Avicel), amorphous cellulose, such as phosphoric acid swollen cellulose (PASC), dyed or fluorescent cellulose, and pretreated lignocellulosic biomass. These substrates are generally highly ordered cellulosic material and thus only sparingly soluble.

[0234] It will be appreciated that suitable lignocellulosic material may be any feedstock that contains soluble and/or insoluble cellulose, where the insoluble cellulose may be in a crystalline or non-crystalline form. In various embodiments, the lignocellulosic biomass comprises, for example, wood, corn, corn stover, sawdust, bark, leaves, agricultural and forestry residues, grasses such as switchgrass, ruminant digestion products, municipal wastes, paper mill effluent, newspaper, cardboard or combinations thereof.

[0235] In some embodiments, the invention is directed to a method for hydrolyzing a cellulosic substrate, for example a cellulosic substrate as described above, by contacting the cellulosic substrate with a host cell of the invention. In some embodiments, the invention is directed to a method for hydrolyzing a cellulosic substrate, for example a cellulosic substrate as described above, by contacting the cellulosic substrate with a co-culture comprising yeast cells expressing heterologous cellulases.

[0236] In some embodiments, the invention is directed to a method for fermenting cellulose. Such methods can be accomplished, for example, by culturing a host cell or co-culture in a medium that contains insoluble cellulose to allow saccharification and fermentation of the cellulose.

[0237] The production of ethanol can, according to the present invention, be performed at temperatures of at least about 30° C, about 31° C, about 32° C, about 33° C, about 34° C, about 35° C, about 36° C, about 37° C, about 38° C, about 39° C, about 40° C, about 41° C, about 42° C, about 43 °C, about 44 °C, about 45 °C, about 46 °C, about 47 °C, about 48 °C, about 49 °C, or about 50° C. In some embodiments of the present invention the thermotolerant host cell can produce ethanol from cellulose at temperatures above about 30° C, about 31° C, about 32° C, about 33° C, about 34° C, about 35° C, about 36° C, about 37° C, about 38° C, about 39° C, about 40° C, about 41° C, about 42° C, or about 43 °C, or about 44 °C, or about 45 °C, or about 50° C. In some embodiments of the present invention, the thermotolerant host cell can produce ethanol from cellulose at temperatures from about 30° C to 60° C, about 30° C to 55° C, about 30° C to 50° C, about 40° C to 60° C, about 40° C to 55° C or about 40° C to 50° C.

[0238] In some embodiments, methods of producing ethanol can comprise contacting a cellulosic substrate with a host cell or co-culture of the invention and additionally contacting the cellulosic substrate with externally produced cellulase enzymes. Exemplary externally produced cellulase enzymes are commercially available and are known to those of skill in the art.

[0239] Therefore, the invention is also directed to methods of reducing the amount of externally produced cellulase enzymes required to produce a given amount of ethanol from cellulose comprising contacting the cellulose with externally produced cellulases and with a host cell or co-culture of the invention. In some embodiments, the same amount of ethanol production can be achieved using at least about 5%, 10%, 15%, 20%, 25%, 30%, or 50% less externally produced cellulases. In some embodiments, no external cellulase is added, or less than about 5% of the cellulase is externally added cellulase, or less than about 10% of the cellulase is externally added cellulase, or less than about 15% of the cellulase is externally added cellulase.

[0240] In some embodiments, the methods comprise producing ethanol at a particular rate. For example, in some embodiments, ethanol is produced at a rate of at least about 0.1 mg per hour per liter, at least about 0.25 mg per hour per liter, at least about 0.5 mg per hour per liter, at least about 0.75 mg per hour per liter, at least about 1.0 mg per hour per liter, at least about 2.0 mg per hour per liter, at least about 5.0 mg per hour per liter, at least about 10 mg per hour per liter, at least about 15 mg per hour per liter, at least about 20.0 mg per hour per liter, at least about 25 mg per hour per liter, at least about 30 mg per hour per liter, at least about 50 mg per hour per liter, at least about 100 mg per hour per liter, at least about 200 mg per hour per liter, at least about 300 mg per hour per liter, at least about 400 mg per hour per liter, or at least about 500 mg per hour per liter.

[0241] In some embodiments, the host cells of the present invention can produce ethanol at a rate of at least about 0.1 mg per hour per liter, at least about 0.25 mg per hour per liter, at least about 0.5 mg per hour per liter, at least about 0.75 mg per hour per liter, at least about 1.0 mg per hour per liter, at least about 2.0 mg per hour per liter, at least about 5.0 mg per hour per liter, at least about 10 mg per hour per liter, at least about 15 mg per hour per liter, at least about 20.0 mg per hour per liter, at least about 25 mg per hour per liter, at least about 30 mg per hour per liter, at least about 50 mg per hour per liter, at least about 100 mg per hour per liter, at least about 200 mg per hour per liter, at least about 300

mg per hour per liter, at least about 400 mg per hour per liter, or at least about 500 mg per hour per liter more than a control strain (lacking heterologous cellulases) and grown under the same conditions. In some embodiments, the ethanol can be produced in the absence of any externally added cellulases.

[0242] Ethanol production can be measured using any method known in the art. For example, the quantity of ethanol in fermentation samples can be assessed using HPLC analysis. Many ethanol assay kits are commercially available that use, for example, alcohol oxidase enzyme based assays. Methods of determining ethanol production are within the scope of those skilled in the art from the teachings herein.

[0243] The following embodiments of the invention will now be described in more detail by way of these non-limiting examples.

Examples

[0244] The present invention presents a number of important steps forward for creating a yeast capable of consolidated bioprocessing. It describes improved cellulolytic yeast created by expressing combinations of heterologous cellulases. The present invention demonstrates for the first time, the ability of transformed *Kluyveromyces* to produce ethanol from cellulose, the ability of yeast strains expressing only secreted heterologous cellulases to produce ethanol from cellulose, and the ability of co-cultures of multiple yeast strains expressing different cellulases to produce ethanol from cellulose. In addition such yeast strains and co-cultures of yeast strains can increase the efficiency of simultaneous saccharification and fermentation (SSF) processes.

General Protocols

General Strain Cultivation and Media

[0245] *Escherichia coli* strain DH5 α (Invitrogen), or NEB 5 α (New England Biolabs) was used for plasmid transformation and propagation. Cells were grown in LB medium (5 g/L yeast extract, 5 g/L NaCl, 10 g/L tryptone) supplemented with ampicillin (100 mg/L), kanamycin (50mg/L), or zeocin (20 mg/L). When zeocin selection was desired LB was adjusted to pH 7.0. Also, 15 g/L agar was added when solid media was desired.

[0246] Yeast strains were routinely grown in YPD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose), YPC (10 g/L yeast extract, 20 g/L peptone, 20 g/L cellobiose), or YNB + glucose (6.7 g/L Yeast Nitrogen Base without amino acids, and supplemented with appropriate amino acids for strain, 20 g/L glucose) media with either G418 (250 mg/L unless specified) or zeocin (20 mg/L unless specified) for selection. 15 g/L agar was added for solid media.

Molecular Methods

[0247] Standard protocols were followed for DNA manipulations (Sambrook et al. 1989). PCR was performed using Phusion polymerase (New England Biolabs) for cloning, and Taq polymerase (New England Biolabs) for screening transformants, and in some cases Advantage Polymerase (Clontech) for PCR of genes for correcting auxotrophies. Manufacturers guidelines were followed as supplied. Restriction enzymes were purchased from New England Biolabs and digests were set up according to the supplied guidelines. Ligations were performed using the Quick ligation kit (New England Biolabs) as specified by the manufacturer. Gel purification was performed using either Qiagen or Zymo research kits, PCR product and digest purifications were performed using Zymo research kits, and Qiagen midi and miniprep kits were used for purification of plasmid DNA. Sequencing was performed by the Molecular Biology Core Facility at Dartmouth College. Yeast mediated ligation (YML) was used to create some constructs (Ma *et al. Gene* 58:201-216 (1987)). This was done by creating DNA fragments to be cloned with 20-40bp of homology with the other pieces to be combined and/or the backbone vector. A backbone vector (pRS426), able to replicate in yeast, and with the Ura3 gene for selection, was then transformed into yeast by standard methods with the target sequences for cloning. Transformed yeast recombine these fragments to form a whole construct and the resulting plasmid allows selection on media without uracil.

Vectors

[0248] Plasmid constructs vectors in the experiments detailed below are summarized in Table 4, and the primers used in vector construction are shown in Table 5.

Table 4. Plasmids used.

Plasmid	Genotype
pBKD1-BGLI	<i>bla KanMX PGK1_{P-S.f.} bgl1- PGK1_T</i>
pBKD2-sEGI	<i>bla KanMX ENO1_{P-sT.r.} egl- ENO1_T</i>
pBKD1-BGLI-sEGI	<i>bla KanMX ENO1_{P-sT.r.} egl- ENO1_T & PGK1_{P-S.f.} bgl1- PGK1_T</i>
YE _p ENO-BBH	<i>bla URA3 ENO1_PT</i>
pJC1 La grange et al. (1996)	<i>bla URA3 PGK_PT</i>
pRDH101	<i>bla URA3 ENO1_{P-sT.r.} cbh1- ENO1_T</i>
pRDH103	<i>bla URA3 ENO1_{P-sH.g.} cbh1- ENO1_T</i>
pRDH104	<i>bla URA3 ENO1_{P-sT.a.} cbh1- ENO1_T</i>
pRDH105	<i>bla URA3 ENO1_{P-sT.e.} cbh1- ENO1_T</i>
pRDH106	<i>bla URA3 ENO1_{P-sT.e.} cbh2- ENO1_T</i>
pRDH107	<i>bla URA3 PGK1_{P-sT.r.} cbh2- PGK1_T</i>
pRDH108	<i>bla URA3 PGK1_{P-sT.r.} cbh2- PGK1_T & ENO1_{P-sT.e.} cbh1- ENO1_T</i>
pRDH118	<i>bla URA3 PGK1_{P-sT.r.} cbh2- PGK1_T & ENO1_{P-sH.g.} cbh1- ENO1_T</i>
pRDH120	<i>bla URA3 PGK1_{P-sT.r.} cbh2- PGK1_T & ENO1_{P-sT.a.} cbh1- ENO1_T</i>
pDF1 La Grange et al. (1996)	<i>bla fur1::LEU2</i>
pCEL5 Den Haan et al. 2007	2 micron vector for expression of SfBGLI and TrEGI (native sequence)
pMU185	pUG66 (loxp-zeo-loxp)
pKLAC1 New England Biolabs	K. lactis expression vector for integration at the lac4 locus, acetamide selection
pRS426	2 micron vector for yeast mediated ligation (YML)
pMU289	pRS426 with portion of pKLAC1 for insertion of TrEG1 (from pBKD_11621, as detailed in example 1) into lac4 locus created by YML
pMU291	pRS426 with portion of pKLAC1 for insertion of TrCBH2 (from pBZD_20641, as detailed in example 1) into lac4 locus created by YML
pMU398	<i>ENO1_{P-sT.e.} cbh1- ENO1_T</i> from pRDH105 into pMU289 (cloning by YML)
pMU451	pRDH105 with PacI/AscI linker (formed using primers) inserted into EcoRI/XhoI
pMU458	synthetic construct for N.f. EG inserted into

	pMU451 (PacI/AscI digest of both pieces)
pMU463	TrEG1 from pBKD1-BGLI-sEG1 into pMU451 (PacI/AscI digest of both pieces)
pMU465	synthetic construct for C.l.(a) EG inserted into pMU451 (PacI/AscI digest of both pieces)
pMU469	synthetic construct for R.f. EG inserted into pMU451 (PacI/AscI digest of both pieces)
pMU471	synthetic construct for C.f. EG inserted into pMU451 (PacI/AscI digest of both pieces)
pMU472	synthetic construct for N.t. EG inserted into pMU451 (PacI/AscI digest of both pieces)
pMU473	synthetic construct for C.a. EG inserted into pMU451 (PacI/AscI digest of both pieces)
pMU475	synthetic construct for T.r. CBH2 derived from pBKD_20641 with tether removed (from example 1) inserted into pMU451 (PacI/AscI digest of both pieces)
pMU499	synthetic construct for M.d. EG inserted into pMU451 (PacI/AscI digest of both pieces)
pMU500	synthetic construct for R.s. EG inserted into pMU451 (PacI/AscI digest of both pieces)
pMU503	synthetic construct for N.w. EG inserted into pMU451 (PacI/AscI digest of both pieces)
pMU624/pMI529	2 micron vector for expression of T.e. CBH1 w/CBD (PCR fragments for chimeric enzyme with <i>PmlI</i> - <i>XhoI</i> digested pRDH105)
pMU326	synthetic construct for R.s. EG from Codon Devices
pMU784/pMI574	2 micron vector for expression of C.l.(b) CBH2 (synthetic construct for C.l.(b) CBH2 inserted into PacI/AscI digested pMU624)
pMU562	pBKD_2 with loxp-zeo-loxp inserted (NotI digest of both pieces)
pMU576	ENO1p- T.r.cbh1-ENO1 _T (from pMU291) in pMU562 (PacI/AscI digest of both pieces)
pMU577	ENO1p- T.e.cbh1-ENO1 _T in (from pMU398) pMU562 (PacI/AscI digest of both pieces)
pMU661	ENO1p- T.r. EG1-ENO1 _T (from pMU463) in pMU562 (PacI/AscI digest of both pieces)
pMU662	ENO1p- C.l.(a) EG1-ENO1 _T (from pMU465) in pMU562 (PacI/AscI digest of both pieces)
pMU663	ENO1p- C.f. EG1-ENO1 _T (from pMU471) in pMU562 (PacI/AscI digest of both pieces)
pMU664	ENO1p- N.t. EG1-ENO1 _T (from pMU472) in pMU562 (PacI/AscI digest of both pieces)
pMU665	ENO1p- C.a. EG1-ENO1 _T (from pMU473) in pMU562 (PacI/AscI digest of both pieces)

pMU666	ENO1p- T.r.CBH2-ENO1 _T (from pMU475) in pMU562 (PacI/AscI digest of both pieces)
pMU667	ENO1p- M.d.-EG1-ENO1 _T (from pMU499) in pMU562 (PacI/AscI digest of both pieces)
pMU668	ENO1p- N.w.-EG1-ENO1 _T (from pMU503) in pMU562 (PacI/AscI digest of both pieces)
pMU755	ENO1p- T.e.CBH1 w/CBD-ENO1 _T (from pMU624) in pMU562 (PacI/AscI digest of both pieces)
pMU750	ENO1p- R.s.-EG2-ENO1 _T (from pMU326) in pMU562 (PacI/AscI digest of both pieces)
pMU809	ENO1p- C.l.(b) CBH2b-ENO1 _T (from pMU784) in pMU562 (PacI/AscI digest of both pieces)
pMU721	pMU562 with hph gene (hygromycin resistance marker) replacing zeocin marker (NotI digest for both fragments)
pMU760	ENO1p- T.e.CBH1 w/CBD-ENO1 _T from pMU624 in pMU721 (MheI/AscI digest for both fragments)
pMU761	ENO1p- T.r.CBH2-ENO1 _T from pMU291 in pMU721 (PacI/AscI digest for both fragments)
pMI553	2 micron vector for expression of T.r. CBH2 and T.e. CBH1+CBM
pMI568	2 micron vector for expression of T.r. EG1, please see text for description of how this construct was built.
pMI574	2 micron vector for expression of C.l.(b) CBH2
pMI577	2 micron vector for expression of T.r. CBH2 and H.g. CBH1
pMI578	2 micron vector for expression of T.r. CBH2 and T.e. CBH1
pMI579	2 micron vector for expression of T.r. CBH2 and C.l.(b) CBH1
pMI580	2 micron vector for expression of C.l.(b) CBH2 and T.e. CBH1+CBM
pMI581	2 micron vector for expression of C.l.(b) CBH2 and T.e. CBH1
pMI582	2 micron vector for expression of C.l.(b) CBH2 and H.g. CBH1
pMI583	2 micron vector for expression of C.l.(b) CBH2 and C.t. CBH1

[0249] Abbreviations: *ENO1*_{P/T} = Enolase 1 gene promoter/terminator; *PGK1*_{P/T} = phosphoglycerate kinase 1 gene promoter & terminator; *T.r.* = *Trichoderma reesei*; *H.g.* =

Humicola grisea; *T.a.* = *Thermoascus aurantiacus*; *T.e.* = *Talaromyces emersonii*, *S.f.* = *Saccharomycopsis fibuligera*; *C.l. (a)* = *Coptotermes lacteus*; *C.f.* = *Coptotermes formosanus*; *N.t.* = *Nasutitermes takasagoensis*; *C.a.* = *Coptotermes acinaciformis*; *M.d.* = *Mastotermes darwinensis*; *N.w.* = *Nasutitermes walkeri*; *R.s.* = *Reticulitermes speratus*; *C.l. (b)* = *Chrysosporium lucknowense*; *N.f.* = *Neosartorya fischeri*; *R.f.* = *Reticulitermes flavipes*; *C.t.* = *Chaetomium thermophilum*

Table 5: Primers Used

sCBH1/2-L	GACTGAATTCATAATGGTCTCCTTCACCTCC
sCBH1-R	GACTCTCGAGTTACAAACATTGAGAGTAGTATGG
sCBH2-R	CAGTCTCGAGTTACAAGAAAGATGGGTTAGC
395 Te cbh1 Synt1 PacI-ATG	GCGTTGGTACCGTTTAAACGGGGCCCTTAATTAAACAAT GCTAAGAAGAGCTTTACTATTGAG
398 Te cbh1 synt core SmaI	CCTCCCCCGGGTTAGAAGCAGTGAAAGTGGAGTTGATTG G
399Trcbh1 synt CBM5 MlyIHincII	GCGACGAGTCAACCC TCCAGGTGGTAACAG AGGTACTACCAC
400 Trcbh1 synt CBM AscIXhoI	GCGACTCGAGGGCGCGCCTACAAACATTGAGAGTAGTA TGGGTTTA
379 ScPGK1prom -786 SacI+ApaI	GCGTTGAGCTCGGGCCCTAATTTTATTTTAGATTCCTGA CTTCAAC
380 ScPGK1prom EcoRI-PacI	GCGTTGAATTCTTAATTAAGTAAAAAGTAGATAATTACT TCCTTG
381 CBH2 WT EcoRI-PacI-ATG	GCGTTGAATTCTTAATTAACAATGATTGTCGGCATTCT CACCACGC
386 CBH2 WT TAA-AscI-EcoRI	gcgatgaattcggcgcgccTTACAGGAACGATGGGTTTGCGTTTG

[0250] The yeast expression vector YEpENO-BBH was created to facilitate heterologous expression under control of the *S. cerevisiae* enolase 1 (*ENO1*) gene promoter and terminator. The vector was also useful because the expression cassette from this vector could be simply excised using a BamHI, BglII digest. YEpENO1 (Den Haan *et al.*, *Metabolic Engineering*. 9: 87-942007) contains the YEp352 backbone with the ENO1 gene promoter and terminator sequences cloned into the BamHI and HindIII sites. This plasmid was digested with BamHI and the overhang filled in with Klenow polymerase and dNTPs to remove the BamHI site. The plasmid was re-ligated to generate YEpENO-B. Using the same method, the BglII and then the HindIII sites were subsequently destroyed to create YEpENO-BBHtemplate. YEpENO-BBHtemplate was used as

template for a PCR reaction with primers ENOBB-left (5'-GATCGGATCCCAATTAATGTGAGTTACCTCA-3') and ENOBB-right (5'-GTACAAGCTTAGATCTCCTATGCGGTGTGAAATA-3') in which the ENO1 cassette was amplified together with a 150 bp flanking region upstream and 220bp downstream. This product was digested with BamHI and HindIII and the over hangs filled in by treatment with Klenow polymerase and dNTPs and cloned between the two PvuII sites on yENO1 effectively replacing the original ENO1 cassette and generating YEpENO-BBH.

[0251] Codon optimized versions of *Humicola grisea cbh1* (*Hgcbh1*), *Thermoascus aurantiacus cbh1* (*Tacbh1*) and *Talaromyces emersonii cbh1* and *cbh2* (*Tecbh1* and *Tecbh2*) were designed and synthetic genes were ordered from GenScript Corporation (Piscataway, NJ, USA). These four synthetic *cbh* encoding genes received from GenScript Corporation were cloned onto the plasmid pUC57. The resulting vectors were digested with EcoRI and XhoI to excise the *cbh* genes which were subsequently cloned into an EcoRI and XhoI digested YEpENO-BBH. This created the plasmids pRDH103 (with *Hgcbh1*), pRDH104 (with *Tacbh1*), pRDH105 (with *Tecbh1*) and pRDH106 (with *Tecbh2*) with the *cbh* encoding genes under transcriptional control of the *ENO1* promoter and terminator. Additionally, pRDH101 was created to express the *T. reesei* CBH1 from pBZD_10631_20641. Takara ExTaq enzyme was used as directed and to amplify the sTrcbh1 from pBZD_10631_20641 using primers sCBH1/2 L and sCBH1R. The fragment was then isolated and digested with EcoRI and XhoI. YEpENO-BBH was also digested with EcoRI and XhoI and the relevant bands were isolated and ligated. A 1494bp fragment encoding the *T. reesei cbh2* gene was amplified from the plasmid pBZD_10631_20641, with primers sCBH1/2-L and sCBH2 R (5'-CAGTCTCGAGTTACAAGAAAGATGGGTTAGC-3'), digested with EcoRI and XhoI and cloned into the EcoRI and XhoI sites of pJC1 (Crouse *et al.*, *Curr. Gen.* 28: 467 - 473 (1995)) placing it under transcriptional control of *S. cerevisiae* phosphoglycerate kinase 1 (PGK1) gene promoter and terminator. This plasmid was designated pRDH107. Subsequently the expression cassettes from pRDH103, pRDH104 and pRDH105 were excised with BamHI and BglII digestion and cloned into the BamHI site of pRDH107 to yield pRDH118, pRDH120, pRDH108 and pRDH109, respectively. pRDH109 contains the same expression cassettes as pRDH108 but in pRDH108 the gene expression cassettes

are in the reverse orientation relative to each other. These plasmids and their basic genotypes are summarized in Table 4.

[0252] Two additional 2-micron vectors for expression of *Chrysosporium lucknowense* CBH2b and the *T. emersonii* CBH1 with a c-terminal fusion of the CBM of *T. reesei* CBH1 were also created. The fusion between *T. emersonii cbh1* and the CBM of *T. reesei cbh1* was generated by ligation of three fragments. Table 5 lists the oligonucleotides used for these constructs. A PCR product was amplified with the oligonucleotides 395 *Te cbh1 Synt1* PacI-ATG and 398 *Te cbh1 synt core* SmaI using pRDH105 as the template, digested with PmlI and SmaI and the 800 bp fragment was isolated. A second PCR product was amplified with oligonucleotides 399 *Trcbh1 synt* CBM5 MlyIHincII and 400 *Trcbh1 synt* CBM AscIXhoI with pRDH101 as the template, digested with MlyI and XhoI and the 180 bp fragment was isolated. The two PCR fragments were ligated with the 6.9 kb PmlI-XhoI fragment of pRDH105 resulting in pMU624.

[0253] The genomic 3900 bp DNA sequence of *Chrysosporium lucknowense cbh2b* gene (described in Published United States Patent Application No: 2007/0238155) was analyzed for putative introns using the NetAspGene 1.0 Server. Removal of the predicted introns from the genomic sequence resulted in an open reading frame of 482 amino acids which was synthesized at Codon Devices and codon optimized for expression in *S. cerevisiae* and cloned into pUC57 vector. Plasmid pAJ401 (Saloheimo et al. Mol. Microbiol. 13:219-228, 1994), which contains the PGK1 promoter and terminator, was modified for expression of *T. reesei cbh2* between PacI and AscI restrictions sites. The *PGK1* promoter was amplified with primers 379 *ScPGK1prom* -786 SacI+ApaI and 380 *ScPGK1prom* EcoRI-PacI and pAJ410 as the template and digested with PacI and EcoRI. The *T. reesei cbh2* ORF was amplified from pTTc01 (Teeri et al., Gene 51:43-52, 1987) with oligonucleotides 381 CBH2 WT EcoRI-PacI-ATG and 386 CBH2 WT TAA-AscI-EcoRI, digested with PacI and EcoRI, and ligated with the SacI-EcoRI digested pAJ401 resulting in pMI508. The PacI-AscI fragment in pMI508 was replaced by a synthetic 1.4 kb *T. reesei egl1* gene resulting in pMI522. The 1.9 kb fragment of pMI522 was digested with PmlI and XhoI and ligated to the 6.4 kb PmlI-XhoI fragment of pRDH107 resulting in pMI568. pMI568 was digested with PacI and AscI and the 7 kb fragment was ligated

to the 1.5 kb fragment of pMI558 producing pMU784 for the expression of *C. lucknowense cbh2b*.

[0254] A set of 2-micron vectors was also constructed for the expression of endoglucanases in *S. cerevisiae*, as well as related plasmids to act as controls. pMU451 was created as a control vector and for cloning the cellulases under control of the *ENO1* promoter and terminator. This was done by adding a PacI/AscI linker into the EcoRI/XhoI site of pMU451. Synthetic genes ordered from Codon Devices and received in pUC57 were cloned into this vector as PacI/AscI fragments. Vectors created this way and listed in Table 4 are: pMU458, pMU463, pMU465, pMU469, pMU471, pMU472, pMU473, pMU475, pMU499, pMU500, and pMU503.

[0255] Vectors for integrating secreted versions of cellulases at the delta integration sites in *S. cerevisiae*, or for integration into the genome of *K. marxianus* were created from the pBKD_1 and pBKD_2 constructs. The *S. fibuligera* BGL1 (SfBGLI) was cloned by PCR from ySFI (van Rooyen *et al.*, *J. Biotechnol.* 120: 284-95 (2005)). The endoglucanase (TrEGI) used was the sequence give in Table 1. The cellulase encoding genes were cloned via PCR (using PacI and AscI sites) into pBKD_1 and pBKD_2 - to create pBKD1-BGL1 and pBKD2-sEG1. The ENO1P-sEG1-ENO1T cassette from pBKD2-sEG1 was subsequently sub cloned as a SpeI, NotI fragment to pBKD1-BGL1 to create pBKD1-BGL1-sEG1.

[0256] pMU562, used for integrating cellulases into *K. marxianus*, was generated by cutting with pMU185 (pUG66) with NotI and isolating a 1190 bp lox P ZeoR containing insert. This insert was ligated into a NotI digested 4.5Kb delta-integration vector to produce pMU562. pMU576 was generated by cutting *T. reesei* CBH2 containing plasmid pMU291 with AscI/PacI, isolating a 1491 bp CBH2 gene and ligating it into delta-integration vector pMU562 cut with AscI/PacI. pMU577 was generated by cutting *T. emersonii* CBH1 from pMU398 with AscI/ PacI, isolating a 1380 bp CBH1 gene and ligating into delta-integration vector pMU562 cut with AscI/PacI. Similarly, a set of recombinant cellulase constructs (pMU661 to pMU668 and pMU750, pMU755, pMU809—see Table 4), including a variety of endoglucanases and cellobiohydrolases, was incorporated into pMU562 for co-transformation. Synthetic sequences for these cellulase genes were originally obtained from Codon Devices and subsequently cloned into 2μ expression vectors for use in *S. cerevisiae*. They were then transferred from these

vectors to the integrating vectors as detailed (including digests used) in Table X. Together these constructs formed a library that could be transformed separately or together and then screened by activity assay. Constructs were digested with enzymes that cut inside of, or very closely outside of, the delta sequences for integration. Similar constructs for integrating cellulases using the hygromycin marker (pMU721, pMU760, and pMU761) were also built.

Yeast Transformation

[0257] For routine transformation of whole plasmids in *S. cerevisiae*, standard chemical transformation was used (*Sambrook et al. Molecular cloning: A laboratory manual*. New York: Cold Spring Harbor Laboratory Press (1989)). For some transformations, a modified protocol described by Hill *et al. (Nucleic Acids Res. 19: 5791 (1991))* was used.

[0258] A protocol for electrotransformation of yeast was developed based on Cho *et al.* (1999) and on Ausubel *et al.* (1994). Linear fragments of DNA were created by digesting pBD1-BGL1-sEG1 with AccI. AccI has a unique site in the δ sequence. The fragments were purified by precipitation with 3M NaAc and ice cold ethanol, subsequent washing with 70% ethanol, and resuspension in USB dH₂O (DNAse and RNAse free, sterile water) after drying in a 70 °C vacuum oven.

[0259] *S. cerevisiae* cells for transformation were prepared by growing to saturation in 5mL YPD cultures. 4 mL of the culture was sampled, washed 2X with cold distilled water, and resuspended in 640 μ L cold distilled water. 80 μ L of 100 mM Tris-HCl, 10 mM EDTA, pH 7.5 (10X TE buffer—filter sterilized) and 80 μ L of 1 M lithium acetate, pH 7.5 (10X LiAc—filter sterilized) were added, and the cell suspension was incubated at 30 °C for 45 min. with gentle shaking. 20 μ L of 1M DTT was added and incubation continued for 15 min. The cells were then centrifuged, washed once with cold distilled water, and once with electroporation buffer (1M sorbitol, 20mM HEPES), and finally resuspended in 267 μ L electroporation buffer. The same protocol was used for transforming *K. lactis* and *K. marxianus* strains, except that 50 mLs of YPD was inoculated with 0.5 mL from an overnight culture, grown for 4 hours at 37 °C, and then centrifuged and prepared as above. Additionally, incubations and recovery steps were carried out at 37 °C.

[0260] For electroporation, 10 μ g of linearized DNA (measured by estimation on a gel) was combined with 50 μ L of the cell suspension in a sterile 1.5 mL microcentrifuge tube.

The mixture was then transferred to a 0.2 cm electroporation cuvette, and a pulse of 1.4 kV (200 μ s, 25 μ F) was applied to the sample using the Biorad Gene Pulser™ device. 1mL of YPD with 1M sorbitol adjusted to pH 7.0 (YPDS) was placed in the cuvette and the cells were allowed to recover for ~3 hrs. 100-200 μ L cell suspension were spread out on YPDS agar plates with appropriate antibiotic, which were incubated at 30 °C for 3-4 days until colonies appeared.

Yeast Strains

[0261] The yeast strains listed in Table 6 were created using the vectors and transformation protocols as described.

Table 6: Yeast Strains.

Name	Background strain	Genes expressed and/or knocked out	Constructs
M0013	<i>Saccharomyces cerevisiae</i> Y294 (ATCC 201160)	Genotype: α , <i>leu2-3,112 ura3-52 his3 trp1-289</i>	None
M0243	M0013	SfBGLI, TrEGI	pBKD1-BGLI-sEGI
M0244	M0013	SfBGLI, TrEGI (native sequence)	pCEL5
M0247	M0013	TeCBH1; delta FUR1	pRDH105
M0248	M0013	TrCBH2, TeCBH1; delta FUR1	pRDH108; pDF1
M0249	M0013	None (control); delta FUR1	pJC1; pDF1
M0265	M0013	HgCBH1; delta FUR1	pRDH103; pDF1
M0266	M0013	TaCBH1; delta FUR1	pRDH104; pDF1
M0282	M0248	SfBGLI, TrEGI, TrCBH2, TeCBH1; delta FUR1	pBKD1-BGLI-sEGI; pRDH108; pDF1
M0284	M0243	SfBGLI, TrEGI, TrCBH2, HgCBH1; delta FUR1	pBKD1-BGLI-sEGI; pRDH118; pDF1
M0286	M0243	SfBGLI, TrEGI, TrCBH2, TaCBH1; delta FUR1	pBKD1-BGLI-sEGI; pRDH120; pDF1
M0288	M0243	SfBGLI, TrEGI, TrCBH2, TeCBH1; delta FUR1	pBKD1-BGLI-sEGI; pRDH108; pDF1
M0289	M0013	TrCBH2, HgCBH1; delta FUR1	pRDH118; pDF1
M0291	M0013	TrCBH2, TaCBH1; delta FUR1	pRDH120; pDF1
M0358	M0282	SfBGLI, TrEGI, TrCBH2, TeCBH1; delta FUR1; Trp1; His3	pBKD1-BGLI-sEGI; pRDH108; pDF1
M0359	M0288	SfBGLI, TrEGI, TrCBH2, TeCBH1; delta FUR1; Trp1; His3	pBKD1-BGLI-sEGI; pRDH108; pDF1
M0361	M0249	None (control); delta FUR1; Trp1; His3	pJC1; pDF1
M0157	<i>Kluyveromyces marxianus</i> (ATCC	None	None

	#10606)		
M0158	<i>Kluyveromyces lactis</i> (ATCC #34440)	None	None
M0411	M0158 (colony #1)	SfBGLI, TrEGI	pBKD1-BGLI-sEGI;
M0412	M0158 (colony #2)	SfBGLI, TrEGI	pBKD1-BGLI-sEGI;
M0413	M0157 (Colony #1)	SfBGLI, TrEGI	pBKD1-BGLI-sEGI;
M0414	M0157 (Colony #2)	SfBGLI, TrEGI	pBKD1-BGLI-sEGI;
M0491	M0414	SfBGLI, TrEGI, TeCBH1, TrCBH2	pBKD1-BGLI-sEGI; pMU576 and pMU577
M0599	M0414	SfBGLI, TrEGI, TeCBH1, TrCBH2	pBKD1-BGLI-sEGI; pMU760 and pMU761
M0600	M0414	SfBGLI, TrEGI, TeCBH1, TrCBH2	pBKD1-BGLI-sEGI; pMU760 and pMU761
M0601 to M0604; M0611 to M0617	M0414 (11 colonies displaying highest avicelase activity)	SfBGLI, TrEGI, Cl(a)EG, CfEG, NtEG, CaEG, MdEG, NwEG, RsEG, TeCBH1, TeCBH1+CBD, TrCBH2, Cl(b)CBH2	pBKD1-BGLI-sEGI; pMU663, pMU755, pMU809, pMU576, pMU661, pMU662, pMU664, pMU665, pMU667, pMU668, pMU750, pMU577
M0618 to M0625	M0157(8 colonies displaying highest avicelase activity)	Cl(a)EG, CfEG, NtEG, CaEG, MdEG, NwEG, RsEG, TeCBH1, TeCBH1+CBD, TrCBH2, Cl(b)CBH2	pMU663, pMU755, pMU809, pMU576, pMU661, pMU662, pMU664, pMU665, pMU667, pMU668, pMU750, pMU577
yENO1	M0013	ENO1 P/T	YE _p ENO-BBH; pDF1
M0419	M0013	ENO1 P/T	pMU451
M0420	M0013	TeCBH1	pMU272
M0423	M0013	TrEG1	pMU463
M0424	M0013	SfBGL1	pMU464
M0426	M0013	RfEG	pMU469
M0446	M0013	Cl(a)EG	pMU465
M0449	M0013	CfEG	pMU471
M0450	M0013	NtEG	pMU472
M0460	M0013	MdEG	pMU499
M0461	M0013	RsEG	pMU500
M0464	M0013	NwEG	pMU503
M0476	M0013	NfEG	pMU458
Y294/pMI529 fur1Δ	M0013	TeCBH1+CBM	pMU624
Y294/pMI553 fur1Δ	M0013	TrCBH2, TeCBH1+CBM	pMI553

Y294/pMI574 fur1Δ	M0013	Cl(b)CBH2	pMI574
Y294/pMI577 fur1Δ	M0013	TrCBH2, HgCBH1	pMI577
Y294/pMI578 fur1Δ	M0013	TrCBH2, TeCBH1	pMI578
Y294/pMI579 fur1Δ	M0013	TrCBH2, Cl(b)CBH1	pMI579
Y294/pMI580 fur1Δ	M0013	Cl(b)CBH2, TeCBH1+CBM	pMI580
Y294/pMI581 fur1Δ	M0013	Cl(b)CBH2, TeCBH1	pMI581
Y294/pMI582 fur1Δ	M0013	Cl(b)CBH2, HgCBH1	pMI582
Y294/pMI583 fur1Δ	M0013	Cl(b)CBH2, Cl(b)CBH1	pMI583

[0262] The plasmid pBKD1-BGL1-sEG1 (pMU276) was digested with AccI and transformed to *S. cerevisiae* Y294 by electrotransformation to create a strain with delta integrated copies of the *SfBGLI* and *TrEGI*, designated M0243. Episomal plasmids were then transformed to *S. cerevisiae* Y294 and/or M0243.

[0263] To create autoselective *S. cerevisiae* strains, *i.e.* strains that can be grown in medium without requiring selective pressure to maintain the episomal plasmid, strains were transformed with NsiI & NcoI digested pDF1 and selected on SC-ura-leu plates. This lead to the disruption of the FUR1 gene of *S. cerevisiae*. PCR was used to confirm FUR1 disruption with primers FUR1-left (5'-ATTTCTTCTTGAACCATGAAC-3') and FUR1-right (5'-CTTAATCAAGACTTCTGTAGCC-3'), where a 2568bp indicated a disruption.

[0264] M0282 was created by transforming M0248 with AccI digested pBKD1-BGLI-sEGI, as described above, except that the transformation mixture was spread on plated containing 10 g/L BMCC with 10g/L yeast extract and 20 g/L peptone.

[0265] The presence of integrated genes was verified by colony PCR for *Kluyveromyces* strains. Selected yeast strains were made prototrophic by transforming with PCR products for genes to complement their auxotrophies.

Cellulosic Substrates for Enzyme Assays

[0266] Bacterial microcrystalline cellulose (BMCC) was a gift from CP Kelco company. BMCC as received was stirred O/N at 4C in water. After the substrate was rehydrated, it

was washed 6 times with water and resuspended in water. The dry weight of the substrate was measured by drying samples at 105°C until constant weight was obtained.

[0267] Avicel PH105 (FMC Biopolymers) was used as provided by the manufacturer.

[0268] Pretreated mixed hardwoods were generated by autohydrolysis of the substrate at 160 PSI for 10 minutes. Pretreated material was washed 5 times to remove inhibitors and soluble sugars and resuspended in distilled water. Samples were dried overnight at 105°C to determine the dry weight. Analysis of sugar content by quantitative saccharification showed a 50% glucan content.

[0269] Phosphoric acid swollen cellulose (PASC) was prepared as in Zhang and Lynd (2006), with only slight modifications. Avicel PH105 (10 g) was wetted with 100 mL of distilled water in a 4L flask. 800 mL of 86.2% phosphoric acid was added slowly to the flask with a first addition of 300mL followed by mixing and subsequent additions of 50mL aliquots. The transparent solution was kept at 4°C for 1 hour to allow complete solubilization of the cellulose, until no lumps remained in the reaction mixture. Next, 2L of ice-cooled distilled water was added in 500mL aliquots with mixing between additions. 300 mL aliquots of the mixture were centrifuged at 5,000 rpm for 20 minutes at 2°C and the supernatant removed. Addition of 300 mL cold distilled water and subsequent centrifugation was repeated 4X. 4.2 mL of 2M sodium carbonate and 300 mL of water were added to the cellulose, followed by 2 or 3 washes with distilled water, until the final pH was ~6. Samples were dried to constant weight in a 105°C oven to measure the dry weight.

Enzyme Assays

[0270] β -glucosidase activity was measured in a manner similar to McBride, J.E., *et al.*, (*Enzyme Microb. Techol.* 37: 93-101 (2005)), except that the volume of the assay was decreased and the reaction performed in a microtiter plate. Briefly, yeast strains were grown to saturation in YPD or YPC media with or without appropriate antibiotics, the optical density at 600nm (OD(600)) was measured, and an 0.5 mL sample of the cultures was taken. This sample was centrifuged, the supernatant was separated and saved, and the cell pellet was washed 2X 50 mM citrate buffer, pH 5.0. Reactions for supernatants were made up of 50 μ L sample, 50 μ L citrate buffer, and 50 μ L 20 mM p-nitrophenyl- β -D-glucopyranoside (PNPG) substrate. Reactions with washed cells consisted of 25 μ L of cells, 75 μ L citrate buffer, and 50 μ L PNPG substrate. If the activity was too high for the

range of the standard curve, a lower cell concentration was used and the assay was re-run. The standard curve consisted of a 2-fold dilution series of nitrophenol (PNP) standards, starting at 500 nM, and ending at 7.8 nM, and a buffer blank was included. After appropriate dilutions of supernatant or cells were prepared, the microtiter plate was incubated at 37 °C for 10 minutes along with the reaction substrate. The reaction was carried out by adding the substrate, incubating for 30 min., and stopping with 150 µL of 2M Na₂CO₃. The plate was then centrifuged at 2500 rpm for 5 minutes, and 150 µL of supernatant was transferred to another plate. The absorbance at 405 nm was read for each well.

[0271] Endoglucanase activity was qualitatively detected by observing clearing zones on synthetic complete media (as above, but including 20 g/L glucose) plates with 0.1% carboxymethyl cellulose (CMC) stained with congo red (Beguin, *Anal. Biochem.* 131: 333-6 (1983)). Cells were grown for 2-3 days on the plates and were washed off the plate with 1M Tris-HCL buffer pH 7.5. The plates were then stained for 10 minutes with a 0.1% Congo red solution, and extra dye was subsequently washed off with 1M NaCl.

[0272] CBH1 activity was detected using the substrate 4-Methylumbelliferyl-β-D-lactoside (MULac). Assays were carried out by mixing 50 µL of yeast supernatant with 50 µL of a 4mM MULac substrate solution made in 50 mM citrate buffer pH 5.5. The reaction was allowed to proceed for 30 minutes and then stopped with 1M Na₂CO₃. The fluorescence in each well was read in a microtiter plate reader (ex. 355 nm and em. 460 nm).

Quantification of Enzyme Activity

[0273] Enzyme activity on PASC and Avicel were measured using the protocol described in Den Haan *et al.*, *Enzyme and Microbial Technology* 40: 1291-1299 (2007). Briefly, yeast supernatants were incubated with cellulose at 4 °C to bind the cellulase. The cellulose was then filtered from the yeast supernatant, resuspended in citrate buffer and sodium azide, and incubated at 37 °C. Accumulation of sugar was measured in the reaction by sampling and performing a phenol-sulfuric acid assay. (See Example 10 and Table 9.)

[0274] Avicel activity levels were also generated using a 96-well plate method. (See Example 2.) Strains to be tested were grown in YPD in deep-well 96 well plates at 35°C with shaking at 900 RPM. After growing, plates were centrifuged at 4000 rpm for 10

min. 300 μ L substrate (2% avicel, 50mM sodium acetate buffer, 0.02% sodium azide, β -glucosidase—1 μ L per mL) was added to a new 96-well deep well plate, without allowing the avicel to settle. 300 μ L of yeast supernatant was added to this substrate, and 100 μ L was taken for an initial sample. The assay plate is incubated at 35°C, with shaking at 800 rpm, and samples were taken at 24 and 48 hours. Samples were placed in 96-well PCR plates, and spun at 2000 rpm for 2 minutes. 50 μ L of supernatant was then added to 100 μ L of DNS reagent previously placed in a separate 96 well PCR plate, mixed, and heated to 99°C for 5 minutes in a PCR machine, followed by cooling to 4 °C. 50 μ L was transferred to a microtiter plate and the absorbance was measured at 565 nm. The conversion of avicel was calculated as follows:

$$Y = \frac{(\text{OD}(T=24\text{or}48) - \text{OD}(T=0)) \times 100\%}{S \times A} = \frac{\Delta\text{OD} \times 100}{0.1 \times 10}$$

Y - % of Avicel converted at 24 or 48 hrs

S – DNS/glucose calibration slope that is 0.1 for DNS at 565 nm

A – Avicel concentration at T=0 that is 10 g/L for 1% Avicel .

Example 1: Production of *Kluyveromyces* expressing heterologous β -glucosidase and endoglucanase

[0275] In order to test the ability of *Kluyveromyces* to express functional heterologous cellulases, two *Kluyveromyces* strains, *Kluyveromyces marxianus* (ATCC strain #10606; MO157) and *Kluyveromyces lactis* (ATCC strain # 34440), were transformed with vectors encoding heterologous cellulases.

[0276] Vectors containing yeast delta integration sequences, the KanMX marker and sequences encoding *S.f.* BGLI and *T.r.* EGI (pBKD-BFLI-sEG1) were transformed into *Kluyveromyces* according to the yeast transformation protocol as described above, and selected on G418. Transformants were verified by PCR and then tested by CMC assay. The results are shown in Figure 1. The presence of the heterologous cellulase activity is indicated by a clearing zone on the CMC plate. As shown in Figure 1, neither an untransformed *K. lactis* strain (colony 8) or an untransformed *K. marxianus* strain (colony 16) showed endoglucanase activity. However, 6 of 7 transformed *K. lactis* colonies showed CMCase activity, and all 7 transformed *K. marxianus* colonies showed CMCase

activity. MO413 and MO414 were identified as two *K. marxianus* colonies showing CMCase activity.

Example 2: Production of *Kluyveromyces* expressing CBH1 and CBH2

[0277] The ability of *Kluyveromyces* to express functional heterologous cellobiohydrolases was also examined. In these experiments, *K. marxianus* (MO157) was transformed with constructs containing *T. reesei* CBH2, *T. emersonii* CBH1 or both. Similarly, MO414 (*K. marxianus* transformed with *S.f.* BGLI and *T.r.* EGI) was transformed with constructs containing *T. reesei* CBH2, *T. emersonii* CBH1 or both.

[0278] Transformations were performed as described in above. CBH1 activity was then detected using the substrate 4-Methylumbelliferyl- β -D-lactoside (MU-Lac) as described above. The assay was performed on eight colonies of each transformant and the three colonies showing the highest activity were averaged. The results are shown in Figure 2 and demonstrate that strains transformed with *T. emersonii* CBH1 had high MU-lac activity.

[0279] The activity of *Kluyveromyces* strains expressing heterologous cellobiohydrolases on Avicel was also assessed. In one experiment, MO413 was transformed with vectors containing *T. reesei* CBH2 and *T. emersonii* CBH1 coding sequences along with a zeocin marker. Novel strain MO491 was created by this transformation and showed MU-lactoside activity. In a second experiment, MO413 was transformed with vectors containing *T. reesei* CBH2 and *T. emersonii* CBH1 coding sequences along with a hygromycin marker, and strains MO599 and MO600 were isolated from this transformation. Activity on Avicel was assessed at 48 hours as described above, and the results, shown in Figure 3, demonstrate that *Kluyveromyces* expressing heterologous cellulases have Avicelase activity at 35 °C. Avicelase activity at 45° C was also demonstrated (data not shown).

Example 3: Production of *Kluyveromyces* expressing a library of cellulases

[0280] *Kluyveromyces* strains were also created by transforming yeast with a library of cellulases (creation of library was described above). For example, MO413 was transformed with a library of cellulases containing a zeocin marker to produce novel

strains MO601-MO604 and MO611-MO617. In addition, MO157 (*K. marxianus*) was transformed with the same library and novel strains MO618-MO625 were identified. Activity on Avicel was assessed at 48 hours as described above, and the results, shown in Figure 3, demonstrate that *Kluveryomces* transformed with a library of heterologous cellulases also have Avicelase activity at 35° C. Transformants of MO157 with the library showed the highest activity. Avicelase activity at 45° C was also demonstrated (data not shown).

Example 4: Ethanol production by transformed *Kluyveromyces*

[0281] In order to determine if *Kluyveromyces* expressing heterologous cellulases could produce ethanol from Avicel, precultures were grown in for 24 hours in YPD (YPD as above, with 20 g/L glucose; 25 mL in a 250 mL shake flask) with shaking at 300 rpm at 35 °C. After 24 and 48 hours, 40 g/L of additional glucose was added. At 72 hours, the pH of the cultures was adjusted to ~5.0 with citrate buffer (initial pH of buffer was 5.5, final concentration was 50 mM), and the culture was added to a sealed plastic shake flask containing 5.5 grams of Avicel (final concentration 10% (w/v). Avicel PH105 (FMC Biopolymers) was used as provided by the manufacturer. The culture was incubated at 35 °C with shaking at 150 rpm.

[0282] Quantification of ethanol in fermentation samples was carried out by HPLC analysis, and initial ethanol concentrations in bottles (from precultures) was subtracted from all subsequent data points (initial ethanol concentrations ranged between 0 and about 6 g/L). The initial glucose concentration for all strains except MO603 was 0.000 g/L. For this strain it was 0.069 g/L, which would result in a maximum in 0.035 g/L of ethanol from the initial sugar.

[0283] The results, as shown in Figure 4, demonstrate that Engineered *K. marxianus* strains were also able to produce ethanol directly from Avicel. Strain MO157, the untransformed control, showed a steady decrease in ethanol concentration over the course of the experiment. This is due to ethanol consumption by the strain because of the presence of a small amount of oxygen in the flasks.

[0284] Of the two strains transformed with *T. reesei* CBH2 and *T. emersonii* CBH1 with the hygromycin marker (MO599 and M600), one (MO599) showed ethanol production. In addition, of the five strains transformed with *T. reesei* CBH2 and *T. emersonii* CBH1

with the zeocin marker, four (MO601, MO602, MO604 and MO491) showed ethanol production. This demonstrates that engineered thermotolerant *K. marxianus* are capable of producing ethanol directly from the recalcitrant crystalline cellulose, Avicel.

Example 5: Production of *S. cerevisiae* expressing heterologous cellulases

[0285] *S. cerevisiae* expressing heterologous cellulases were also produced and tested for their ability to grow on media containing bacterial microcrystalline cellulose (BMCC). In these experiments, microaerobic conditions were maintained by growing strains on BMCC in sealed hungate tubes with an air atmosphere.

[0286] Strains expressing *T. emersonii* CBH1 and *T. reesei* CBH2 (MO248) were transformed with a construct allowing *T. reesei* EGI and *S. fibuligera* BGLI expression (pKD-BGLI-sEGI). That transformation was plated on a BMCC solid agar plate and five colonies appeared on the plate after seven days (data not shown). Yeast from the largest of the five colonies was isolated as strain MO282. (MO282 is described in more detail above.) The three control strains were tested for growth on the same plates. One strain expressed with *T. emersonii* CBH1 and *T. reesei* CBH2, and two strains expressed *T. reesei* EGI and *S. fibuligera* BGLI. No colonies appeared on plates with control yeast strains (data not shown).

[0287] The ability of MO282 to grow on BMCC was also tested using liquid media. Figure 5 shows that MO282, which expresses all 4 secreted cellulases grew to a much greater extent on BMCC than a plasmid only control (MO249), a strain expressing only *T. emersonii* CBH1 and *T. reesei* CBH2 (MO249), and a strain expressing 4 tethered cellulases (MO144).

[0288] These results indicate that yeast expressing secreted *T. emersonii* CBH1, *T. reesei* CBH2, *T. reesei* EGI and *S. fibuligera* BGLI heterologously are able to grow on bacterial microcrystalline cellulose.

Example 6: *S. cerevisiae* expressing heterologous cellulases can produce ethanol from Avicel and Pretreated Hardwood

[0289] In order to determine if transformed *S. cerevisiae* can produce ethanol directly from cellulose without exogenously added cellulase enzymes, transformed strains were

grown on Avicel as the sole carbon source. Avicel PH105 (FMC Biopolymers) was used as provided by the manufacturer.

[0290] Avicel media was made using the non-glucose components of synthetic complete medium for yeast including, yeast nitrogen base without amino acids—6.7 g/L, and supplemented with a complete amino acid mix (complete supplemental mixture). In some cases yeast extract (10 g/L) and peptone (20 g/L) (YP) were used as supplements in growth experiments. Cultivation conditions were anaerobic and were maintained by flushing sealed glass bottles with N₂ after carbon source addition and before autoclaving. Non-carbon media components were added as 10X solutions by filter sterilizing after autoclaving. Inoculation into Avicel cultures was done at 20% by volume. Quantification of ethanol in fermentation samples was carried out by HPLC analysis, and initial ethanol concentrations in bottles (from precultures) was subtracted from all subsequent data points.

[0291] As shown in Figure 6, Strain M0288 (expressing *S. fibuligera* BGLI, *T. reesei* EGI, *T. reesei* CBH2, and *T. emersonii* CBH1) was able to produce ethanol directly from avicel PH105 as compared to the control strain (M0249) when YNB media components were used.

[0292] The ability of MO288 to produce ethanol from cellulose was also demonstrated using pretreated hardwoods. Pretreated mixed hardwoods were generated by autohydrolysis of the substrate at 160 PSI for 10 minutes. Pretreated material was washed 5 times to remove inhibitors and soluble sugars and resuspended in distilled water. Samples were dried overnight at 105 °C to determine the dry weight. Analysis of sugar content by quantitative saccharification showed a 50% glucan content. Media and culture conditions were as described above for Avicel experiments except that cultures were inoculated at 10% by volume.

[0293] The data presented in Figure 7 demonstrates that MO288 was also able to make ethanol from pretreated hardwoods without added enzyme. The strain made ~0.5 g/L more than the control when YP was used as media, and ~0.2 g/L when YNB was used.

[0294] These data demonstrate that yeast expressing secreted *T. emersonii* CBH1, *T. reesei* CBH2, *T. reesei* EGI and *S. fibuligera* BGLI heterologously are able to produce ethanol from cellulose without the addition of any exogenous cellulases.

Example 7: Transformed yeast strains and externally added cellulases act synergistically to produce ethanol from pretreated mixed hardwoods.

- [0295] Production of ethanol from biomass is currently achieved using an SSF type of process where cellulase enzymes are added exogenously to a reaction containing pretreated cellulosic biomass, yeast growth media, and yeast. In order to determine if yeast expressing recombinant cellulases could improve this process, recombinant yeast expressing secreted cellulases were cultured in the presence of a range of exogenously added cellulase concentrations. Growth and media conditions were as described in previous examples.
- [0296] In these experiments, a recombinant yeast strain expressing four secreted cellulases (MO288) was compared directly to the control strain (MO249) under the same conditions. External cellulases were added at concentrations of 25 mg cellulase per gram cellulose (100%), 22.5 mg cellulase per gram cellulose (90%), 18.75 mg cellulase per gram cellulose (75%) or 6.25 mg cellulase per gram cellulose (25%). Experiments were also performed without adding any external cellulases (0%). Pretreated mixed hardwoods (prepared as described in examples above) at an initial solids concentration of 5% were used as a cellulose source. The data is presented in Figure 8. From this data, it is clear that the strain producing cellulases makes additional ethanol relative to the control strain for each of the cellulase loading concentrations tested.
- [0297] In order to examine this effect in more detail, ethanol production at different external cellulase concentrations was evaluated in two different types of media using pretreated mixed hardwood. The results are shown in Figure 9. In YP media, MO288 makes 6-9% more ethanol at the higher cellulase loadings, only 1% more at a 25% loading, and 100% more when no cellulase is loaded. In YNB media MO288 makes 20-40% more ethanol at low cellulase loadings, and ~10% more ethanol at higher cellulase loadings. These results can be used to determine the amount of cellulase that can be removed from the process with the same overall ethanol yield being achieved. For YP media cellulase loading can be reduced ~15% compared to the control, and for YNB media, cellulase loading can be reduced ~5%. At non-zero cellulase loadings ethanol productivity was increased between 5 and 20% for strains expressing cellulases in YP media as compared to the control. It was increased between 10 and 20% for strains cultured in YNB media compared to the control.

[0298] These data demonstrate that previous SSF processes can be improved in terms of ethanol yield from biomass and ethanol productivity if strains expressing secreted cellulases are used in combination with exogenously added cellulases. Similarly, cellulase loadings required to achieve a particular percentage of theoretical ethanol yield can be reduced when strains expressing recombinant cellulases are added.

Example 8: Transformed yeast strains also increase efficiency of externally added cellulases in the production of ethanol from Avicel.

[0299] To test whether this same trend would hold at high substrate concentrations these experiments were repeated using 15% Avicel PH105 as substrate instead of 5% pretreated mixed hardwood. The results are shown in Figures 10 and 11. The strain making cellulases (MO288) routinely produced more ethanol from Avicel than the control yeast strain (MO249) under identical conditions, even at increased ethanol concentrations (Figure 10). For example, when 25 mg cellulase per gram cellulose was loaded in the SSF reaction, the test strain (M0288) produced 54 g/L, while the control (M0249) produced 50 g/L.

[0300] To examine cellulase displacement the percentage of theoretical ethanol yield achieved at different cellulase loadings was determined. The results presented in Figure 12 were repeated in triplicate for M0288 and M0249, allowing standard deviations for the increased ethanol yields to be calculated. The data that can be used for calculating cellulase displacement is presented in Figure 12. Figure 12 presents cellulase enzyme savings based on theoretical ethanol yield at 168 hours in an SSF experiment. SSF was performed in 30 ml of nitrogen purged YP + 15% Avicel in pressure bottles. External cellulase mix at a ratio of 5 Spyzme:1 Novozyme-188 was used. The experiment was continued for 168 hours and sampling was done each day for ethanol estimation by HPLC. The arrows in the figure depict the necessary cellulase loading needed to achieve the same ethanol production from cellulose as the control. This loading is consistently lower than for the control (*i.e.* the ethanol yield is consistently higher). For data at 168 hours, the average cellulase displacement (amount less that needs to be loaded) is 13.3% $\pm$ 4.9%.

Example 9: Use of Artificial Cbh1 to Produce Ethanol

[0301] In order to design a CBH1 protein with efficient cellulase activity, 17 CBH1 protein sequences from NCBI database (Table 7) were aligned.

Table 7. Fungal CBH1 genes used for alignment.

Organism	Genbank#
Neosartorya fischeri	<u>XM_001258277</u>
Gibberella zeae	<u>AY196784</u>
Penicillium janthinellum	<u>X59054</u>
Nectria haematococca	<u>AY502070</u>
Fusarium poae	<u>AY706934</u>
Chaetomium thermophilum	<u>AY861347</u>
Aspergillus terreus	<u>XM_001214180</u>
Penicillium chrysogenum	<u>AY790330</u>
Neurospora crassa	<u>X77778</u>
Trichoderma viride	<u>AY368686</u>
Humicola grisea	<u>X17258</u>
Thermoascus aurantiacus	<u>AF421954</u>
Talaromyces emersonii	<u>AAL89553</u>
Trichoderma reesei	<u>P62694</u>
Phanerochaete chrysosporium	<u>Z29653</u>
Aspergillus niger	<u>XM_001391971</u>
Aspergillus niger	<u>XM_001389539</u>

[0302] The artificial protein sequence was designed as a consensus (the most common) sequence for these proteins. The predicted signal sequence was exchanged by *S.cerevisiae* alpha mating factor pre signal sequence, and the sequence of the consensus CBH1 protein is shown below. Capital letters indicate the *S.cerevisiae* alpha mating factor pre signal sequence.

MRFPSIFTAVLFAASSALAqqaglttaethpsltwqkctsggscttvngsvvidanwrwvhatsgst
ncytgntwdttlcpddvtcaqncaldgadyssstygvttsgnslrlnfvtqgsqknvgsrlylmeddttymfklg
qeftfdvdvsnlpcglngalyfvamdadggmskypgnkagakygtgycdsqcprdlkfingqanvegweps
sndanagignhgccaemdiweansistaftphpcdtigqtmcegdscggyssdryggtdcpdgcdfnpyr
mgntkfygpgktvdttkkvttvtqfitgssgtlseikrfyvqngkvipnsestisgvsgnsittdfctaaktafgdtd
dfakkgglegmkgalaqgmvlvmslwdhhaanmlwldstyptdatsstpgaargscdtssgvpadvansp
nsyvtfsnikfgpigstftg (SEQ ID NO:43).

[0303] An *S.cerevisiae* and *K. lactis* codon optimized sequence for expressing the CBH1 consensus sequence (SEQ ID NO:44) was developed and is shown below.

atgagatttccttcaatcttcaactgctgtttgttcgcagcctcaagtgttttagcacaacaggccggaacattgacag
cagaaactcatccttccttaacctggcaaaagtgcacttctggagggtcatgcactacagtgaatggatctgtcgtg
atcgatgcaaactggagatgggttcacgcaacttcaggttctaccaactgttataccggaacacttgggacacca
cattgtgccagatgacgtcacgtgcgctcagaactgtgctttggatggagctgattacagttcaacctatggtgta
actacatccggaactctttgagattaaacttcgttactcaaggaagtcaaaagaacgttggttctagattgtacttaa
tggaggacgatacaacctatcaaattgtcaaattgtaggtcaggagttcacctttgacgtagatgtcagtaacttgc
catgtgggttaaaccggagctttatactttgtggcaatggatgctgacgggtggaatgtccaagtatccaggaaacaa
agccggtgcaaagtacgggtacaggatattgtgattcacagtgccttagagatttgaagttcattaacgggtcaagca
aatgtggagggttgggaacctctagtaacgatccaatgcgggtattggaatcatgggtcctgttgcgctgaga
tggatatctgggaggccaactcaatatctactgcctttaccctcaccatgcgatacaattggtcaaactatgtgcg
agggtgattcatgtgttgaacctactcctctgatagatacggaggtacatgcgatccagatggttgcgactttaat
ccatacagaatgggaaacaaaaccttttacggctcctggaaagacagttgatactaccaagaaagtaacagtcgtg
accagtttatcaccggtagtcttggaaaccttatccgaaatcaaaagattctacgttcagaacggtaaagtaattcca
aacagtgaatctacaatttcaggagtgagtggtaattctattactaccgactttgtacagctcagaaaacagcatttg
gtgacaccgatgactttgctaagaagggtggattagaaggtatgggttaaagctttggcccagggaatggtgttagt
tatgtctttatgggatgatcacgccgcaaatatgttatggttggattcaacatatccaactgatgccacaagtagtaca
cctggagctgccagaggttcttgtgatacatctccggtgttcagccgatgtagaagcaaattctcctaactcctat
gttaccttctccaatataaagtttgggtccaatcggttcaacattcactgggttaa (SEQ ID NO:44)

[0304] The codon optimized sequence was inserted into the episomal yeast expression vector (pMU451) under control of ENO1 promoter and terminator into PacI/AscI sites. The resulting expression constructs (pMU505) was transformed into MO375 host strain that derived from Y294 (MO013) in which His3 and Trp1 auxotrophies were rescued by transformation with *S.cerevisiae* His3 and Trp1 PCR products. The resulting strain expressing the CBH1 consensus sequence was named MO429.

[0305] In order to determine if MO429 had cellulase activity, an Avicel conversion assay was performed as described above and measured at 24 hours. As shown in Figure 14, *S. cerevisiae* expressing the consensus Cbh1 sequence (MO429) showed cellulase activity as compared to a negative control transformed with an empty vector (MO419). The cellulase activity of MO429 was also compared to that of yeast strains expressing other heterologous cellulases. The strains tested are summarized in Table 8 below.

Table 8: Cellulytic Strains Used in Avicel Conversion Assay

Strain #	Description	Cellulase	Family	Organism	Activity	Signal
M0419	MO375+pM U451	none			none	
M0420	MO375+pM U272	CBH1	Fungi	Talaromyces emersonii	exo	native
M0429	MO375+pM U505	fungal CBH1 consensus	Fungi	N/A	exo	S.c. α MFpre
M0445	MO375+pM U459	CBH1	Fungi	Neosartorya fischeri	exo	S.c. α MFpre
M0456	MO375+pM U495	CBH1	Fungi	Chaetomium thermophilum	exo	S.c. α MFpre
M0457	MO375+pM U496	CBH1	Fungi	Aspergillus terreus	exo	S.c. α MFpre
M0458	MO375+pM U497	CBH1	Fungi	Penicillium chrysogenum	exo	S.c. α MFpre

[0306] All of the strains in Table 8 were derived from the same parental MO375 strain and were transformed with an episomal yeast vector. MO420, MO429, MO445, MO456, MO457 and MO458 were created using episomal yeast vectors containing the heterologous cellulase genes as listed in the table which were codon optimized for expression in *S. cerevisiae* and *K. lactis*. The cellulases in MO429, MO445, MO456, MO457 and MO458 were expressed under control of *S. cerevisiae* ENO1 promoter and terminator. *T. emersonii* CBH1 was expressed with its own native signal sequence. As shown in Figure 14, the secreted activity on Avicel of the consensus CBH1 was comparable with activity of other fungal CBH1s expressed in the same vector and in the same host strain.

Example 10: Comparison of Cellulase Activity in *S. cerevisiae*

[0307] *S. cerevisiae* were transformed with polynucleotides encoding a number of different heterologous cellobiohydrolases and their activity on PASC and Avicel was assessed as described above. The results are shown in the table below:

Table 9: Cellobiohydrolase activity in *S. cerevisiae*.

Plasmid	Expression Cassette(s)	Act. (PASC) (mU/gDCW)	Act. (Avicel) (mU/gDCW)
yENO1	<i>ENO1</i> p/t	2.68±1.1	2.99±0.7
M0265	<i>ENO1</i> p/t- <i>sH.g.cbh1</i>	32.82±6.5	34.85±2.0
M0266	<i>ENO1</i> p/t- <i>sT.a.cbh1</i>	38.56±5.9	38.15±4.1
M0247	<i>ENO1</i> p/t- <i>sT.e.cbh1</i>	75.60±13.1	21.42±6.1
M0248	<i>PGK1</i> p/t- <i>sT.r.cbh2</i> & <i>ENO1</i> p/t- <i>sT.e.cbh1</i>	174.35±6.5	40.5±4.9
M0289	<i>PGK1</i> p/t- <i>sT.r.cbh2</i> & <i>ENO1</i> p/t- <i>sH.g.cbh1</i>	Not measured	106.2±6.8
M0291	<i>PGK1</i> p/t- <i>sT.r.cbh2</i> & <i>ENO1</i> p/t- <i>sT.a.cbh1</i>	Not measured	32.7±5.7

- [0308] In addition, activity on Avicel was assayed using a 96-plate assay, and the results are shown in Figure 14. In the Figure, for each strain, the first bar indicates the sugar released at 24 hours, and the second bar indicates the sugar released by 48 hours. CBH1s expressed individually, or in combination with *T. reesei* CBH2 showed some avicel activity—reaching 10% conversion of avicel in 48 hours. Combinations of CBH1 with CBH2 from *C. lucknowense* reached much higher avicel conversions of about 22% conversion in 48 hours in combination with *T. emersonii* CBH1 with CBD attached.
- [0309] The avicel activity data for endoglucanases tested in *S. cerevisiae* is shown in Figure 15. The data demonstrate that among the EGs tested the *C. formosanus* EG demonstrated the highest avicel activity when expressed in *S. cerevisiae*.

Example 11: Co-cultures of yeast strains expressing different heterologous cellulases produce ethanol from Avicel

- [0310] A co-culture of a number of cellulase producing yeast strains also showed the ability to make ethanol from Avicel PH105 in YNB media (Figure 16). In this experiment 5 strains independently producing *T. emersonii* CBH1 (M0247), *T. aurantiacus* CBH1 (M0266), *H. grisea* CBH1 (M0265), a combination of *T. emersonii* CBH1 and *T. reesei* CBH2 (M0248), and a combination of *T. reesei* EGI and *S. fibuligera* BGLI (M0244) were mixed in equal proportion by volume and then inoculated at 20% by volume. Each of the heterologously expressed cellulases in each of these strains was secreted. Media and culture conditions were as described above for Avicel experiments. The data in

Figure 16 demonstrate that heterologous cellulases do not need to be expressed in an individual yeast strain in order to produce ethanol from cellulose. Instead, yeast strains expressing different secreted heterologous cellulases can be cultured together in order to produce ethanol from cellulose without the addition of any exogenous cellulases.

[0311] A co-culture using a different combination of cellulases was also evaluated. In this set of co-culture experiments, four yeast strains were cultured together: M0566 (M0424 with FUR deletion): Secreted SfBGLI; M0592 (M0449 with FUR deletion): Secreted CfEGI; M0563 (same as Y294/pMI574 fur1 Δ): Secreted Cl CBH2b; and M0567 (same as Y294/pMI529 fur1 Δ): Secreted TeCBH1+CBD. These strains were grown in liquid YPD for 3 days, until the culture was saturate for pre-culture. At this point they were used to inoculate experiments where avicel (10%) was used as the substrate, and the 4 strains were mixed at equal volume prior to inoculation.

[0312] Figure 17 demonstrates that the co-cultured strains are capable of producing ethanol directly from avicel in the absence of any added cellulase enzyme. The co-culture produces about 4-fold more ethanol after 168 hours as compared to the control strain, and about 3-fold more than M0288.

[0313] This co-culture was also used in SSF experiments where Zoomerace cellulase enzyme cocktail was used at 5 different loadings (10 mg protein/g avicel, 7.5 mg/g, 5 mg/g, and 2.5 mg/g, and 0 mg/g), and strains were inoculated at 10% by volume.

[0314] Figure 18 presents the raw data for ethanol production at a variety of cellulase loadings by the co-culture, M0288, and M0249. Figure 18A shows that at all cellulase loadings tested, the co-cultured strains produced significantly more ethanol than a control not producing cellulase. Figure 18B shows that at all cellulase loadings tested, the co-culture produced more ethanol than the previously tested strain M0288. Figure 19 shows the percentage of the theoretical yield of ethanol that could be achieved with each of these cultures after 168 hours of SSF using a variety of cellulase loadings. The data demonstrate that the co-cultured strains would achieve about a 2-fold reduction in cellulase relative to the control strain, and approximately a 35% reduction compared to M0288.

[0315] These data demonstrate that the combination of cellulases in this co-culture is highly efficient in the production of ethanol.

Example 12: Construction of a Robust Xylose-Utilizing Strain

[0316] M0509 (ATCC deposit designation _____, deposited on November 23, 2009) is a strain of *Saccharomyces cerevisiae* that combines the ability to metabolize xylose with the robustness required to ferment sugars in the presence of pretreated hardwood inhibitors. M0509 was created in a three-step process. First, industrial strains of *S. cerevisiae* were benchmarked to identify strains possessing a level of robustness/hardiness sufficient for simultaneous saccharification and fermentation (SSF) of pretreated mixed hardwood substrates. Strain M0086, a diploid strain of strain of *S. cerevisiae*, satisfied this first requirement. Second, M0086 was genetically engineered with the ability to utilize xylose, resulting in strain M0407. Third, M0407 was adapted for several weeks in a chemostat containing xylose media with pretreatment inhibitors, generating strain M0509.

[0317] Strain M0407 was genetically engineered from M0086 to utilize xylose. This engineering required seven genetic modifications. The primary modification was the functional expression of the heterologous xylose isomerase gene, *XylA*, isolated from the anaerobic fungus *Piromyces sp. E2*. The *S. cerevisiae* structural genes coding for all five enzymes involved in the conversion of xylulose to glycolytic intermediates were also overexpressed: xylulokinase, ribulose 5-phosphate isomerase, ribulose 5-phosphate epimerase, transketolase and transaldolase. In addition, the *GRE3* gene encoding aldose reductase was deleted to minimise xylitol production. The seven modified genes are listed in Figure 39. The genetic modifications at the *GRE3*, *RKI1*, *RPE1*, *TAL1*, and *TKL1* loci were designed to leave behind minimal vector DNA and no antibiotic markers. Each locus' DNA was sequenced to confirm the expected results. Each of the seven genetic modifications were sequentially introduced into strain M0086. Figure 40 shows the progression of modifications from top to bottom together with the designations for the strain at each step in the process, starting with M0086 and finishing with M0407.

[0318] The deletion of *GRE3* and the increased expression of *RKI1*, *RPE1*, *TAL1*, and *TKL1* involve modifications of the endogenous *S. cerevisiae* loci. In the case of *GRE3*, both alleles were deleted. For the other four loci, only a single allele was modified. All of the modifications of endogenous loci required the use of selectable antibiotic markers including *kan^r* from the *Escherichia coli* transposon Tn903 (confers resistance to G418), *nat1* from *Streptomyces noursei* (confers resistance to clonNAT/nourseothricin), and

dsdA from *Escherichia coli* (confers resistance to D-serine.) After selection for a desired genomic modification, the antibiotic marker was excised from the genome using the loxP/*cre* recombinase system. The *cre* recombinase was carried on plasmid pMU210 which contains a zeocin resistance marker. Loss of pMU210 as well as all antibiotic markers was tested on the appropriate selective media. Subsequent PCR genotyping and DNA sequencing confirmed removal of the antibiotic markers from the modified genomic loci.

[0319] The overexpression of *RKII*, *RPE1*, *TAL1*, and *TKL1* was achieved by placing the *S. cerevisiae* triose phosphate isomerase promoter, *TPI*, immediately 5' of each of the four ORFs. For *TAL1* and *RKII*, small portions of their endogenous promoters were deleted. To avoid disruption of adjacent ORFs and possible transcriptional regulatory elements, the introduction of the *TPI* promoter at the *RPE1* and *TKL1* loci was done such that the *RPE1* and *TKL1* loci were duplicated with the duplicate copies of both loci being regulated by the *TPI* promoter.

[0320] In order to boost M0407's xylose-utilization and increase its pretreatment inhibitor tolerance, the strain was maintained in a chemostat for four weeks under the following sequential conditions described in Table 10.

Table 10: Conditions to Improve M0407.

Duration (days)	Residence Time (h)	Media
5	24	YPX, 20 g/L xylose
5	18	YPX, 20 g/L xylose
7	24	YPX + 25% of a 30% MS129 washate (21.5 g/L xylose)
14	24	YPX + 75% of a 30% MS129 washate (~22% solids equivalent)

[0321] An aliquot of the adapted chemostat culture was plated on YPXi50% and nine M0407 "adapted" colonies were screened in YPDXi media (100 g/L glucose, 50 g/L xylose, 25% MS149 pressate). M0407 and M0228 (a xylose-utilizing strain created at Mascoma containing XlyA and XKS1 on plasmids) were included as controls. At 24 hours, the glucose had been entirely consumed by all strains. M0407 and M0228 had utilized 30 and 25 g/L of xylose respectively. All nine M0407 "adapted" colonies had utilized more than 44 g/L of xylose. The highest amount of xylose consumed was 48 g/L. This strain was designated M0509.

[0322] 18S rDNA sequencing was used to confirm strain M0509 as *Saccharomyces cerevisiae* (Kurtzman CP and Robnett, CJ ; FEMS Yeast Research 3 (2003) 417-432). A

1774 bp fragment spanning the 18S rDNA was amplified from M0509 genomic DNA and sent for sequencing. The 1753 bp of M0509 18S rDNA sequence exhibited a 100% match to the NCBI sequence for *S. cerevisiae* 18S (nucleotide accession #Z75578).

[0323] Since strain M0509 was obtained by cultivating M0407 in a chemostat for four weeks, the length of cultivation separating the two strains provides a means to assess the stability of the engineered genetic modifications. Comparison of the DNA sequence of M0407 and M0509 at the *GRE3*, *RKII*, *RPE1*, *TAL1*, and *TKL1* loci showed no changes. This suggests that the genetic modifications at these loci are genetically stable, at least under the growth conditions used.

[0324] Real Time PCR analysis was used to estimate the copy number of integrations of the *XylA/XKS1* vector. M0407 has approximately 10 copies of the vector, whereas M0509 has approximately 20 copies. This suggests that the copy number of the *XylA/XKS1* vector can be increased by extended cultivation on xylose media.

[0325] To further assess the stability of the *XylA/XKS1* integrations, M0509 was cultivated for ~50 generations in liquid media with either glucose or xylose as the sole carbon source. After 50 generations, an individual colony was isolated from each culture and the number of *XylA/XKS1* integrations quantified and compared to the original M0509 freezer stock. The colony isolated from the xylose-culture had ~20 copies of *XylA/XKS1*, the same as the freezer stock. The glucose-cultured colony exhibited a slightly decreased copy number, ~ 16.

[0326] The slight decrease in *XylA/XKS1* copy number of the glucose-colony raises the question of the strain's performance. To partially address this question, xylose consumption was compared between the xylose-isolate, glucose-isolate, and freezer stock. The freezer and xylose-propagated isolates utilized all of the xylose in 24 hours and produced identical amounts of ethanol, but the glucose-propagated strain consumed only half as much xylose. Figure 20.

Example 13: Selection of a Thermotolerant, Robust, Xylose-Utilizing Strain

[0327] M1105 is capable of fermentation at temperatures above 40° C in the presence of 8 g/L acetate. M1105 was constructed in a M0509 background and is therefore an industrially robust strain capable of converting both glucose and xylose into ethanol.

- [0328] M1105 was isolated following four rounds of selection / adaptation in a cyostat as outlined in Figure 41 and described as follows. The temperature was increased from 38° C to 41° C during the course of the experiment. M1017 (ATCC deposit designation _____, deposited on November 23, 2009) was isolated from this first cyostat run and was later confirmed by PCR of the GRE3 locus to be a descendant of M0509. M1017 was used to inoculate a second cyostat run using YMX media (yeast nitrogen base, 2g/L xylose) at 41° C. M1046 was isolated from this second cyostat run. At 42° C on YPX50, M1046 grew slowly yet with a doubling time 36% shorter than M1017. M1080 was isolated from a cyostat inoculated with M1046 and YMX media at 40° C. M1080 grew with a specific growth rate of 0.22 h⁻¹ on YMX at 40° C. M1105 was isolated from M1080 based on selection in the cyostat using YPD2X10+acetate media (2 g/L glucose, 10 g/L xylose, 8 g/L acetate, pH 5.4) at 39° C.
- [0329] M1105 grows 10-20% faster than M0509 in rich media at 35° C. In addition, M1105 has increased acetate tolerance as the strain can grow more quickly than its ancestral strains in the presence of acetate. Figure 21. While the parental strains required glucose for tolerance to acetate at high temperatures, M1105 does not require glucose or complex medium components to grow in the presence of 7 g/L acetate at pH 5.4.
- [0330] To test fermentation performance, M1105 was inoculated at approximately 0.7 g/L DCW in 18% MS419 using 3.8 mg Zoomerose/g feedstock at 40° C. M1105 produced 3.55% (w/v) ethanol by 168 hours. The time course is presented in Figure 22 along with a similar run performed with M1088 (described below) for comparison. A similar run using only 0.15 g/L DCW for inoculum resulted in 2.9% (w/v) ethanol and some sugar accumulation during the experiment. Figure 23.

Example 14: Adaptation of a Thermotolerant, Robust, Xylose-Utilizing Strain

- [0331] M1254 is capable of fermentation at temperatures above 40°C in the presence of 12 g/L acetate, exhibiting an increased robustness relative to the thermotolerant strain M1105.
- [0332] M1254 was isolated following three rounds of selection / adaptation in a cyostat as outlined in Table 11 and Figure 42 and described as follows. The first cyostat run was inoculated with M1105. YMX media (yeast nitrogen base w/o amino acids, 20 g/L xylose) plus 8g/L acetate was used at pH 5.5 and 40° C. M1155 was isolated from this

first cytostat run and used to inoculate a second cytostat containing YPD media (yeast extract, peptone, 20 g/L glucose) plus 12 g/L acetate at pH 5.4 and 41° C. M1202 was isolated from this second cytostat run. M1254 was isolated from a third cytostat run inoculated with both M1155 and M1202 in yeast nitrogen base w/o amino acids + 5% solids equivalent MS419 hydrolysate media at pH 4.8, 39° C.

Table 11: Evolutionary Conditions to Generate M1254 from M1105.

Parental Strain(s)	Evolutionary Condition	New Strain
M1105	Xylose minimal + 8 g/L acetate, pH 5.5, 40°C	M1155
M1155	Complex glucose + 12 g/L acetate, pH 5.4, 41°C	M1202
M1155+M1202	5% solids equivalent MS419 hydrolysate + yeast nitrogen base w/o amino acids, pH 4.8, 39°C	M1254

[0333] M1254 grows 7.3±0.9% faster than M1202 and 17±2.0% faster than M1155 in 5% solids equivalent MS419 hydrolysate, which is the condition under which strain M1254 was selected. However, standard fermentation medium limits fermentation performance. Accordingly, use of this strain should be with lower ammonium concentrations, such as 1.1 g/L diammonium phosphate (DAP) or lower than 3 g/L DAP. Figure 24 demonstrates the higher fermentation rate using the lower DAP concentration. The fermentations were performed using 18% MS149, 4 mg external cellulase/g TS, 40° C, 0.5 g/L inoculation DCW M1254 and pH 5.4. The pH was controlled using 5 M potassium hydroxide, and 1 g/L magnesium carbonate was fed with each solids feeding. All enzyme was front loaded, while the solids were fed at five time points (0, 3, 6, 24, and 48 hours) in equal size feedings of 3.6% TS.

[0334] MO1360 was created from M1254 using the evolutionary conditions described in Table 12 below.

Table 12: Conditions to Generate M1360 from M1254.

Parental Strain(s)	Evolutionary Condition	New Strain
M1254	Complex with low xylose + 8 g/L acetate, pH 5.4, 40°C	M1339
M1339	Complex xylose + synthetic inhibitor mixture (including 8 g/L acetate) , pH 5.4, 40°C	M1360

[0335] M1360, while still substantially inhibited by the synthetic inhibitor mixture, grows at 40° C with a doubling time of approximately 5 hours. Figure 25. In industrially

relevant medium, M1360 is able to generate over 60 g/L ethanol from glucose along with 5 g/L dry cell weight in 48 h at 40°C beginning with only 60 mg/L dry cell weight.
Figure 26.

[0336] Enzyme activity is known to increase as temperature increases, and thus it is desirable to have thermotolerant *S. cerevisiae strains*. Figure 27 shows three equivalent SSFs with 18% PHW solids loaded. The reactions carried out at 40° C show approximately 17% more ethanol produced than the control reaction carried out at 35° C, when both reactions were carried out at the same external enzyme loading (4mg/g). This increased performance represents a substantial cost savings for the process.

Example 15: Expression of Cellulases in a Robust Xylose-Utilizing Strain

[0337] M1088 is capable of secreting three distinct cellulolytic enzymes: β-glucosidase from *S. fibuligera* (SfBGL), cellobiohydrolase 2b from *C. lucknowense* (ClCBH2b), and cellobiohydrolase I from *T. emersonii* fused to the *T. reesei* cellobiohydrolase I cellulose binding domain (TeCBH1+ CBDTrCBH1). The M1088 genome also contains genes that encode for polypeptides capable of providing resistance to the following antibiotics: kanamycin, nourseothricin, and hygromycin B. Plasmid pMU624, which is also present in M1088, contains a gene encoding for a polypeptide capable of providing resistance to ampicillin. The steps used to generate M1088 and M0963 from M0509 are summarized in Table 13 below.

Table 13: Strains Used to Generate Strains M1088 and M0963

Strain	Genotype	Parent	Description
M0509	<i>gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta</i>		
M0539	<i>URA-3/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta</i>	M0509	A single copy of the genomic URA-3 gene was deleted and replaced with a kanMX cassette. The KanMX gene cassette provides resistance to kanamycin (an aminoglycoside antibiotic).
M0544	<i>ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1</i>	M0539	The second copy of the genomic URA-3 gene was deleted and replaced with a kanMX cassette.

	<i>TKL+/loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta</i>		
M0749	<i>ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta (pMU782)fur1::nat</i>	M0544	A single copy of the genomic FUR-1 gene was deleted and replaced with a <i>Streptomyces noursei</i> nat1 cassette. The nat1 gene cassette provides resistance to the antibiotic nourseothricin/clonNAT (an aminoglycoside antibiotic).
M0867	<i>FUR-1/fur-1::nat ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta (pMU782)fur1::nat; [pMU624]</i>	M0749	The plasmid pMU624 was transformed into the strain. pMU624 can replicate in <i>S. cerevisiae</i> (2 micron ori and URA-3) and <i>E. coli</i> (pBMR ori and ampicillin resistance gene: beta-lactam antibiotic). pMU624 also carries the <i>T. emersonii</i> CBH1+CBDBTrCBH1 gene regulated by the ENO1 promoter and terminator.
M0759	<i>fur-1::hyg/fur-1::nat ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta (pMU782) fur1::nat; [pMU624]; (pMU1037) fur1::hyg</i>	M0867	The second copy of the genomic FUR-1 gene was deleted and replaced with a hygMX cassette. The hygMX gene cassette encodes for a hygromycin B phosphotransferase that confers resistance to hygromycin B (an aminoglycoside antibiotic).
M1088	<i>fur-1::hyg/fur-1::nat ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta [pMU624] (pMU1260) delta::PGKprom-SfBGL-PGKterm, ENO1prom-TeCBH+TrCBD-ENO1term (pMU1169) delta::PGKprom-SfBGL-PGKterm, ENO1prom-C1CBH2-ENO1term</i>	M0759	Two distinct integration cassettes were transformed into the strain and multiple copies were integrated into the genome at delta site. One cassette contained the cellulolytic genes <i>S. fibuligeria</i> BGL and <i>C. lucknowense</i> CBH2b. The other cassette contained the cellulolytic genes <i>S. fibuligeria</i> BGL and a <i>T. emersonii</i> chimeric CBH1.
M0963	<i>fur-1::hyg/fur-1::nat ura-3::kanMX/ura-3::kanMX gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PADH1-XKS::delta</i>	M0759	Linear DNA from the 4 plasmids shown was transformed into M0759. 24 of the resulting colonies were then passaged for a week in YPD media containing zeocin at a low level (50 ug/mL), and assayed. The resulting strain M0963 was the best of those

(pMU782) fur1::nat; [pMU624]; (pMU1037) fur1::hyg; (pMU755) delta::ZeoMX, ENO1prom-TeCBH1w/TrCBD- ENO1term; (pMU809) delta::ZeoMX, ENO1prom- ClCBH2b-ENO1term; (pMU663) delta::ZeoMX, ENO1prom-CfEG-ENO1term; (pMU864) delta::ZeoMX, ENO1prom-SfBGL-ENO1term		found in the avicel assay.
--	--	----------------------------

Example 16: Selection of an Endoglucanase for Expression in a Robust Xylose-Utilizing Strain

[0338] Endoglucanases augment the activity of cellobiohydrolases, and therefore, the ability of family 5 endoglucanases to complement the previously identified CBH1 and CBH2 was investigated. Five family 5 endoglucanases were selected and cloned under control of the ENO1 promoter/terminator using the pRDH122 expression plasmid as shown in Table 14.

Table 14: Family 5 endoglucanases expressed in *S. cerevisiae*.

Organism & Gene:	CBM domain:	Expression plasmid:	Theoretical enzyme size Da*
<i>Aspergillus kawachii</i> egA	C-terminal CBM1	pRDH145	55034.58
<i>Heterodera schachtii</i> eng1	C-terminal CBM2	pRDH146	43739.46
<i>Hypocrea jecorina</i> (anamorph: <i>Trichoderma reesei</i>) eg2	N-terminal CBM1	pRDH147	44226.91
<i>Orpinomyces</i> sp.PC-2 celB	2x C-terminal CBM10	pRDH148	53103.40
<i>Irpex lacteus</i> en1	N-terminal CBM1	pRDH149	42357.15

[0339] All plasmids expressing the 5 new EG2-type cellulases were transformed to Y294 (a lab strain) and M0749 (robust xylose utilizing strain; described above) and transformants were confirmed via PCR. Figure 28 shows several of the M0749 strains that were spotted on SC^{-URA} plates containing 0.2% of either CMC or lichenin or barley-β-glucan. As can be seen in figure 28, the M0749 reference strain yielded small zones on the CMC containing plates. Both pMU471 (*Coptotermes formosanus* EG) and pRDH147 based strains yielded very good clearing zones on all the tested substrates.

[0340] Along with the reference strain and a strain expressing the *Coptotermes formosanus* EG (pMU471), the five *eg2* expressing strains were tested for avicel and PASC hydrolysis while the *cbh2* expressing strains were tested for activity on avicel. The strains were grown in double strength SC^{URA} medium (3.4 g/L YNB; 3 g/L amino acid dropout pool without uracil; 10 g/L ammonium sulfate; 20 g/L glucose) that was buffered to pH 6 (20 g/L succinic acid; 12 g/L NaOH, set pH to 6 with NaOH). 10 mL Cultures in 125 mL Erlenmeyer flasks were grown at 30° C for three days. Three flasks were inoculated for each strain. After incubation, samples were taken for gel analysis and activity measurement. After centrifugation of the samples, 12 µl of each was taken, added to 5 µl of protein loading buffer and boiled for 5 minutes. The samples were subsequently loaded on a 10% SDS-PAGE and separated, followed by silver staining. The results are shown in Figure 29. Not all strains produced visible bands in the expected size range. The *C.f.*EG appeared as a band of about 55 kDa as previously seen but the band produced by M0749 seems to be slightly larger than the one produced by Y294. No bands were visible for the *H. schachtii eng1*, *Orpinomyces celB*, or *I. lacteus en1* products. The *H. jecorina* EG2 produced by Y294 and M0749 was visible as ~57 kDa bands. The increased weight compared to the predicted 44 kDa size may represent hyperglycosylation. The *A. kawachii* EGA produced by Y294 was visible as a ~42 kDa band. However, the *A. kawachii* EGA produced by M0749 was clearly visible as a ~120 kDa band. The extra weight may signify hyperglycosylation.

[0341] All strains were tested for activity using the high-throughput avicel conversion method as prescribed. Strains expressing endoglucanases were also tested for activity on PASC. The DNS used for the assay procedure contained phenol which, according to literature, renders greater sensitivity. Activity data can be seen in Figure 30.

[0342] The M0749 strain expressing *H.j.eg2* (pRDH147) produced the highest levels of secreted activity as measured on PASC or avicel of the EG2s tested. The activity of this enzyme was higher on PASC and avicel than *C.f.EG* (pMU471). The synthetic *A.k.*EGA (pRDH145) also gave appreciable activity on both substrates. This product seems to have been produced at higher levels in M0749 than in Y294 and yielded greater activity than *C.f.EG* on avicel and PASC when produced in this strain.

Example 17: Expression of an Endogluconase in Robust Xylose-Utilizing Yeast

- [0343] Several strains were created to test the impact of co-expressing *TrEG2* with CBHs in a robust xylose utilizing strain background. M1088 was transformed with a construct to integrate *TrEG2* at the rDNA locus using the *Sh-ble* gene as a marker (pMU1409). A similar transformation was done, but integrating *TeCBH1w/TrCBD* to increase the copy number of that gene. 43 transformants from both transformations along with duplicate M1088 cultures were grown in 20 ug/mL zeocin containing YPD and the avicel assay was performed. Figure 31 shows the results of those assays. The data show that a very large proportion of strains transformed with the *TrEG2* construct had significantly increased avicel conversion ability, while transformants with additional *TeCBH1w/TrCBD* copies had only marginal improvements in avicel hydrolysis.
- [0344] Of the strains assayed, the top 9 candidates were chosen and restreaked for single colonies. These single colonies were then grown in YPD with 2 transfers to equal a total of 18 generations. The final transfer (passaged data in Figure 32) was compared to the first YPD culture (original data in figure 32). The data confirms that there is an ~50% increase in ability of the yeast supernatant to convert avicel when *TrEG2* is overexpressed.
- [0345] In addition, strain M1403, which contains heterologous genes encoding *S. fibuligera* (*SfBGL*), cellobiohydrolase 2b from *C. lucknowense* (*ClCBH2b*), cellobiohydrolase I from *T. emersonii* fused to the *T. reesei* cellobiohydrolase I cellulose binding domain (*TeCBH1+ CBDTrCBH1*), and *Heterodera schachtii eng1* was produced in the M1254 background. Strain M1284, which contains heterologous genes encoding those same four cellulases was produced in the M0509 background. Strains M1284 and M1403 are described in more detail in Table 15.

Table 15: Endogluconase Expressing Yeast Strains.

Strain	Genotype	Parent	Description
M1403	<i>(pMU1339) delta::MET3prom-SfBGL-PGKterm, ENO1prom-TeCBH1+TrCBD-ENO1term</i> <i>(pMU1260) delta::PGKprom-SfBGL-PGKterm, ENO1prom-TeCBH+TrCBD-ENO1term</i> <i>(pMU1169) delta::PGKprom-SfBGL-PGKterm, ENO1prom-ClCBH2-</i>	M1254	Linear DNA cassettes created by restriction digests of plasmids were integrated in multiple copies into the genome at the Ty1 delta sites and rDNA sites.

	<i>ENO1term</i> (pMU1409) <i>rDNA::ZeoMX, ENO1prom-HjEG2-ENO1term</i>		
M0991	<i>gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PTPI-XKS LEU2/leu2D::hph</i>	M0509	A single copy of the genomic LEU-2 gene was deleted and replaced with a hygMX cassette.
M0992	<i>gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PTPI-XKS leu2D::hph/leu2D::nat</i>	M0991	The second copy of the genomic LEU-2 gene was deleted and replaced with a <i>Streptomyces noursei nat1</i> cassette.
M1162	<i>gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PTPI-XKS leu2D::hph/leu2D::nat (pMU1379) delta::leu2-19, ENO1prom-TeCBH+TrCBD-ENO1term</i>	M0992	A linear DNA cassette created by restriction digests of plasmid pMU1379 was integrated in multiple copies into the genome at the Ty1 delta sites.
M1284	<i>gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RKI1+/loxP-PTPI-RKI1 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xylA PTPI-XKS leu2D::hph/leu2D::nat (pMU1379) delta::leu2-19, ENO1prom-TeCBH+TrCBD-ENO1term (pMU1169) delta::PGKprom-SfBGL-PGKterm, ENO1prom-CICBH2-ENO1term (pMU1409) rDNA::ZeoMX, ENO1prom-HjEG2-ENO1term</i>	M1162	Linear DNA cassettes created by restriction digests of plasmids pMU1169 and pMU1409 were integrated in multiple copies into the genome at the Ty1 delta sites and rDNA sites.

Example 18: Conversion of lignocellulosic substrates via CBP yeast strains

[0346] Expression of cellulases in yeast, particularly CBH1 (*T. emersonii* CBH1 w/ *T. reesei* CBD attached), CBH2 (*C. lucknowense* CBH2b), EG2 (*T. reesei* EG2), and BGL (*S. fibuligera* BGL) dramatically reduces the need for externally added enzymes during enzymatic conversion of lignocellulose to ethanol. To test the effect of overexpressing these enzymes, several strains were constructed and tested on a number of substrates.

- [0347] Figure 33 presents data from a CBP fermentation of paper sludge by an engineered thermotolerant *S. cerevisiae* host strain (parent strain M1254, cellulolytic derivative M1403). The data for M1254 alone demonstrates that the addition of cellulase (*i.e.* zoomerase) is required for ethanol production from paper sludge. The data for M1430 where no external cellulase is added (filled orange squares), demonstrates that this strain can convert a substantial fraction (~80%) of the “convertible” substrate by virtue of its expressed cellulases. Fermentations with additional external cellulase added to the M1403 strain demonstrate the ultimate potential of enzymatic conversion for the paper sludge substrate. Visual inspections demonstrated that the non-CBP strain was not able to liquefy the substrate, whereas the CBP strain was.
- [0348] Furthermore, the CBP strain M1179, which expresses CBH1, CBH2, EG2, and BGL can convert paper sludge to a large extent without added cellulase enzyme. Figure 34. The control strain in this reaction, M0509, made only a small amount of ethanol during this reaction. The data also show that M1179 can convert this material when loaded at lower cell density (1 g/L) as opposed to the higher cell density (10 g/L) used in other reactions. This implies that the strain is able to grow and produce cellulase throughout the fermentation experiments.
- [0349] Pretreated hardwood (PHW) can also be converted by CBP strains. Figure 35, shows the effect of using a cellulase expressing strain (M0963), compared to a control strain not expressing cellulases (M0509) during fermentation of PHW. The comparison demonstrates that the CBP strain can achieve the same yield of ethanol from PHW when only 2 mg/g of external enzyme are loaded compared to when 4 mg/g of M0509 are loaded in the process. This 2-fold reduction in external enzyme needed represents a large potential cost reduction in the process.
- [0350] CBP strains are capable of producing high ethanol titers from PHW as well. Figure 36 shows that a 30% washed solids fermentation can generate titers of ethanol up to about 70 g/L with minimal external enzyme loaded 4mg/g and a relatively low cell inoculum (2 g/L). The ability of the low cell density cultivation to eventually catch up to and pass the high cell density culture indicates that the strain grows and continues to make enzyme throughout the fermentation.
- [0351] In addition to PHW, corn stover has been implicated as good substrate for conversion to ethanol via an enzymatic saccharification. Figure 37 demonstrates that

pretreated corn stover can be converted well by CBP yeast strains. The CBP strain in this experiment was able to convert about 82% of what was converted with a high enzyme loading (15 FPU, or about 20 mg/g) could achieve. The non-CBP strain made about 60% of the ethanol that the CBP strain was able to achieve.

Example 19: Comparison of CBH1 Cellulases

[0352] In order to provide additional data on the expression levels of different CBH1 enzymes, selected strains were grown in YPD-medium and activities on MULac and Avicel™ were assayed. Both Y294 and M0749 transformants were studied, and the results are shown in Figure 38.

* * *

[0353] These examples illustrate possible embodiments of the present invention. The scope of the claims should not be limited by the preferred embodiments set forth in the examples, but should be given the broadest interpretation consistent with the description as a whole.

WHAT IS CLAIMED IS:

1. A thermotolerant yeast host cell comprising:
 1. at least one heterologous polynucleotide comprising a nucleic acid which encodes an endoglucanase;
 2. at least one heterologous polynucleotide comprising a nucleic acid which encodes a β -glucosidase;
 3. at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase I; and
 4. at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase II;wherein the thermotolerant yeast host cell grows at a temperature of at least 35°C or above,
wherein the endoglucanase, β -glucosidase, cellobiohydrolase I, and cellobiohydrolase II are secreted; and
wherein the host cell is capable of producing ethanol from AvicelTM.
2. The thermotolerant yeast host cell of claim 1, wherein the β -glucosidase is a *Saccharomycopsis fibuligera* β -glucosidase comprising the amino acids of SEQ ID NO:40.
3. The thermotolerant yeast host cell of claim 1 or 2, wherein the cellobiohydrolase I is a *Talaromyces emersonii* cellobiohydrolase I comprising the amino acids of SEQ ID NO:23.
4. The thermotolerant yeast host cell of claim 3, wherein the *Talaromyces emersonii* cellobiohydrolase I is fused to a cellulose binding module (CBM).
5. The thermotolerant yeast host cell of claim 4, wherein the CBM is the CBM of *Trichoderma reesei* CBH1 comprising amino acids 503 to 535 of SEQ ID NO:27 or *Humicola grisea* CBH1 comprising amino acids 492 to 525 of SEQ ID NO:21.

6. The thermotolerant yeast host cell of any one of claims 1 to 5, wherein the cellobiohydrolase II is a *Chrysosporium lucknowense* cellobiohydrolase II comprising the amino acids of SEQ ID NO:25.
7. The thermotolerant yeast host cell of any one of claims 1 to 6, wherein the endoglucanase is an *Humicola grisea*, *Thermoascus aurantiacus*, *Talaromyces emersonii*, *Trichoderma reesei*, *Coptotermes lacteus*, *Coptotermes formosanus*, *Nasutitermes takasagoensis*, *Coptotermes acinaciformis*, *Mastotermes darwinensis*, *Nasutitermes walkeri*, *Saccharomycopsis fibuligera*, *Chrysosporium lucknowense*, *Reticulitermes speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*, *Clostridium josui*, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Chaetomium thermophilum*, *Aspergillus terreus*, *Neurospora crassa*, *Reticulitermes flavipes*, or *Arabidopsis thaliana* endoglucanase.
8. The thermotolerant yeast host cell of any one of claims 1 to 7, further comprising a heterologous polynucleotide comprising a nucleic acid which encodes a second endoglucanase.
9. The thermotolerant yeast host cell of claim 1, wherein the β -glucosidase is a *Saccharomycopsis fibuligera* β -glucosidase comprising the amino acids of SEQ ID NO:40, the cellobiohydrolase I is a *Talaromyces emersonii* cellobiohydrolase I comprising the amino acids of SEQ ID NO:23, and the cellobiohydrolase II is a *Chrysosporium lucknowense* cellobiohydrolase II comprising the amino acids of SEQ ID NO:25.
10. The thermotolerant yeast host cell of claim 9, further comprising a second heterologous polynucleotide comprising a nucleic acid which encodes a second endoglucanase, and wherein the endoglucanase and the second endoglucanase are selected from the group consisting of *Humicola grisea*, *Thermoascus aurantiacus*, *Talaromyces emersonii*, *Trichoderma reesei*, *Coptotermes lacteus*, *Coptotermes formosanus*, *Nasutitermes takasagoensis*, *Coptotermes acinaciformis*, *Mastotermes darwinensis*, *Nasutitermes walkeri*, *Saccharomycopsis fibuligera*, *Chrysosporium lucknowense*, *Reticulitermes speratus*, *Thermobfida fusca*, *Clostridium thermocellum*, *Clostridium cellulolyticum*,

Clostridium josui, *Bacillus pumilis*, *Cellulomonas fimi*, *Saccharophagus degradans*, *Piromyces equii*, *Neocallimastix patricarum*, *Neosartorya fischeri*, *Chaetomium thermophilum*, *Aspergillus terreus*, *Neurospora crassa*, *Reticulitermes flavipes*, and *Arabidopsis thaliana* endoglucanases.

11. The thermotolerant yeast host cell of any one of claims 1 to 10, wherein the host cell is a *Kluyveromyces* host cell.
12. The thermotolerant yeast host cell of any one of claims 1 to 10, wherein the host cell is a *Saccharomyces cerevisiae* host cell.
13. The thermotolerant yeast host cell of any one of claims 1 to 12, wherein the host cell is capable of fermenting pentose sugars.
14. The thermotolerant yeast host cell of any one of claims 1 to 13, wherein the host cell is capable of fermenting xylose.
15. The thermotolerant yeast host cell of claim 14, wherein the host cell expresses xylose isomerase, xylulokinase, ribulose 5-phosphate isomerase, ribulose 5-phosphate epimerase, transketolase, transaldolase or a combination thereof.
16. The thermotolerant yeast host cell of claim 15, wherein the host cell expresses xylose isomerase, xylulokinase, ribulose 5-phosphate isomerase, ribulose 5-phosphate epimerase, transketolase and transaldolase.
17. The thermotolerant yeast host cell of claim 13, wherein the host cell does not express a GRE3 gene encoding aldose reductase.
18. The thermotolerant yeast host cell of claim 1, wherein the endoglucanase is an *Aspergillus kawachii*, *Heterodera schachtii*, *Hypocrea jecorina*, *Orpinomyces* sp., *Irpex lacteus*, *Acremonium thermophilum*, *Chaetomium globosum*, or *Aspergillus fumigatus* endoglucanase.
19. The thermotolerant yeast host cell of claim 9, further comprising a second heterologous polynucleotide comprising a nucleic acid which encodes a second endoglucanase, and

wherein the endoglucanase and the second endoglucanase are selected from the group consisting of *Aspergillus kawachii*, *Heterodera schachtii*, *Hypocrea jecorina*, *Orpinomyces* sp., *Irpex lacteus*, *Acremonium thermophilum*, *Chaetomium globosum*, and *Aspergillus fumigatus* endoglucanase.

20. The thermotolerant yeast host cell of claim 10, wherein the endoglucanase is a *Neosartorya fischeri* endoglucanase comprising the amino acids of SEQ ID NO:37.
21. The thermotolerant yeast host cell of any one of claims 1 to 20, wherein the thermotolerant yeast host cell grows at temperatures of at least 38°C or above.
22. The thermotolerant yeast host cell of any one of claims 1 to 20, wherein the thermotolerant yeast host cell grows at temperatures of at least 40°C or above.
23. The thermotolerant yeast host cell of any one of claims 1 to 20, wherein the thermotolerant yeast host cell grows at a temperature from 40°C to 50°C.
24. The thermotolerant yeast host cell of claim 9, wherein the *Talaromyces emersonii* cellobiohydrolase I is fused to a cellulose binding module (CBM).
25. A thermotolerant yeast host cell comprising:
 - (a) at least one heterologous polynucleotide comprising a nucleic acid which encodes an endoglucanase;
 - (b) at least one heterologous polynucleotide comprising a nucleic acid which encodes a β -glucosidase;
 - (c) at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase I; and
 - (d) at least one heterologous polynucleotide comprising a nucleic acid which encodes a cellobiohydrolase II;wherein the thermotolerant yeast host cell grows at a temperature of at least 35°C or above,
wherein the cellobiohydrolase I, the cellobiohydrolase II, or both is fused to a cellulose binding module (CBM); and
wherein the host cell is capable of producing ethanol from AvicelTM.

26. A method of fermenting cellulose comprising culturing the host cell of any one of claims 1 to 25 in medium that contains cellulose to allow saccharification and fermentation of the cellulose.
27. A culture comprising the host cell of any one of claims 1 to 25 and cellulose.

1/42

Figure 1

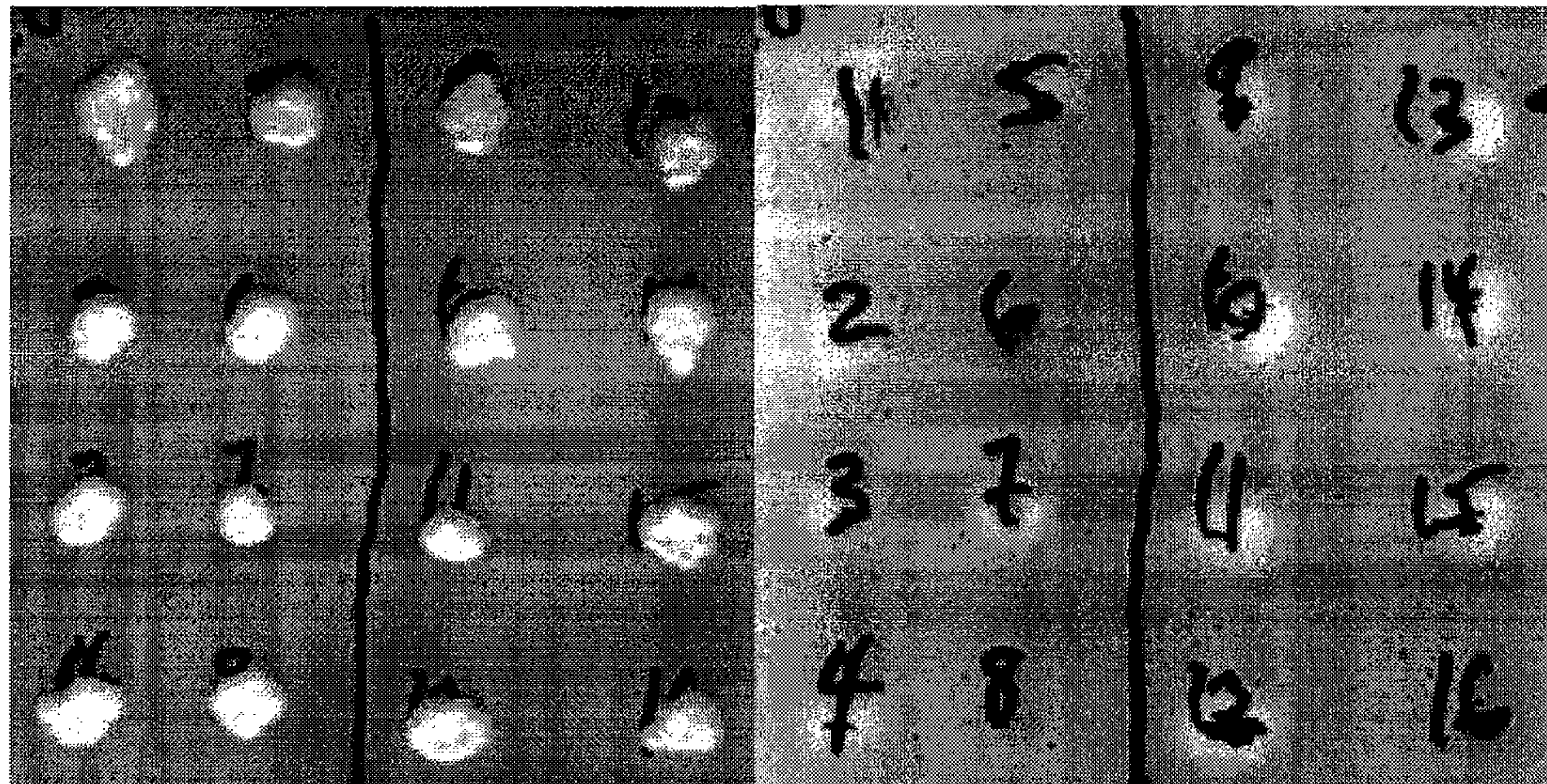
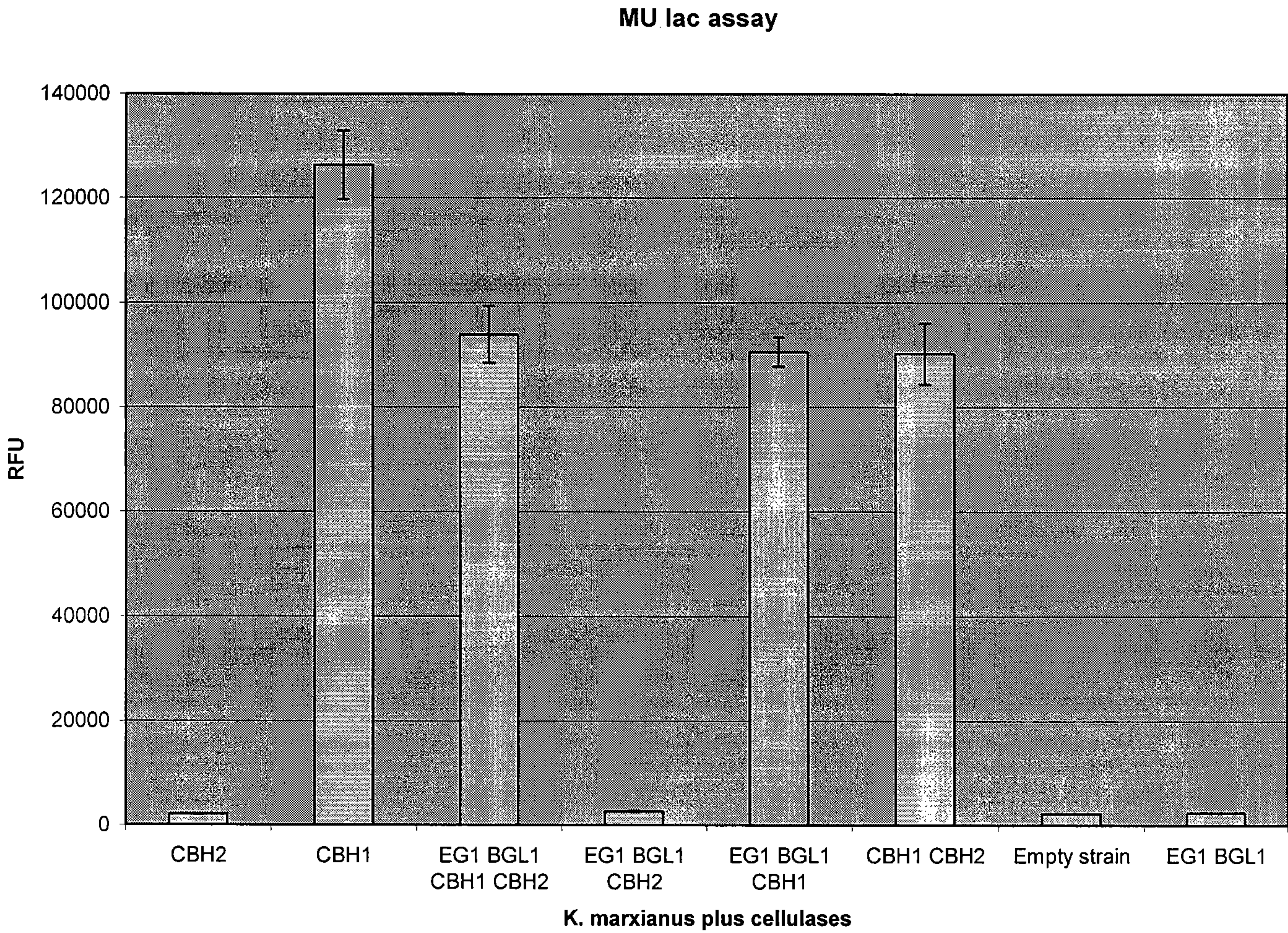


Figure 2



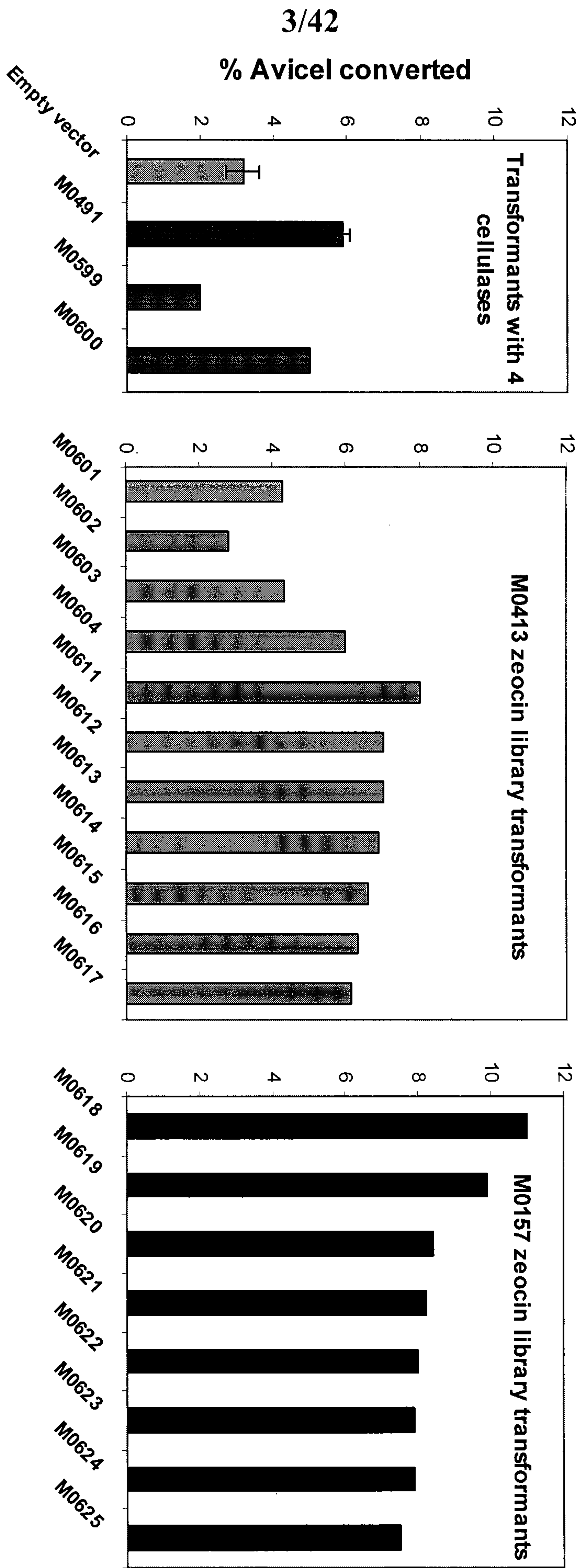


Figure 3

Figure 4

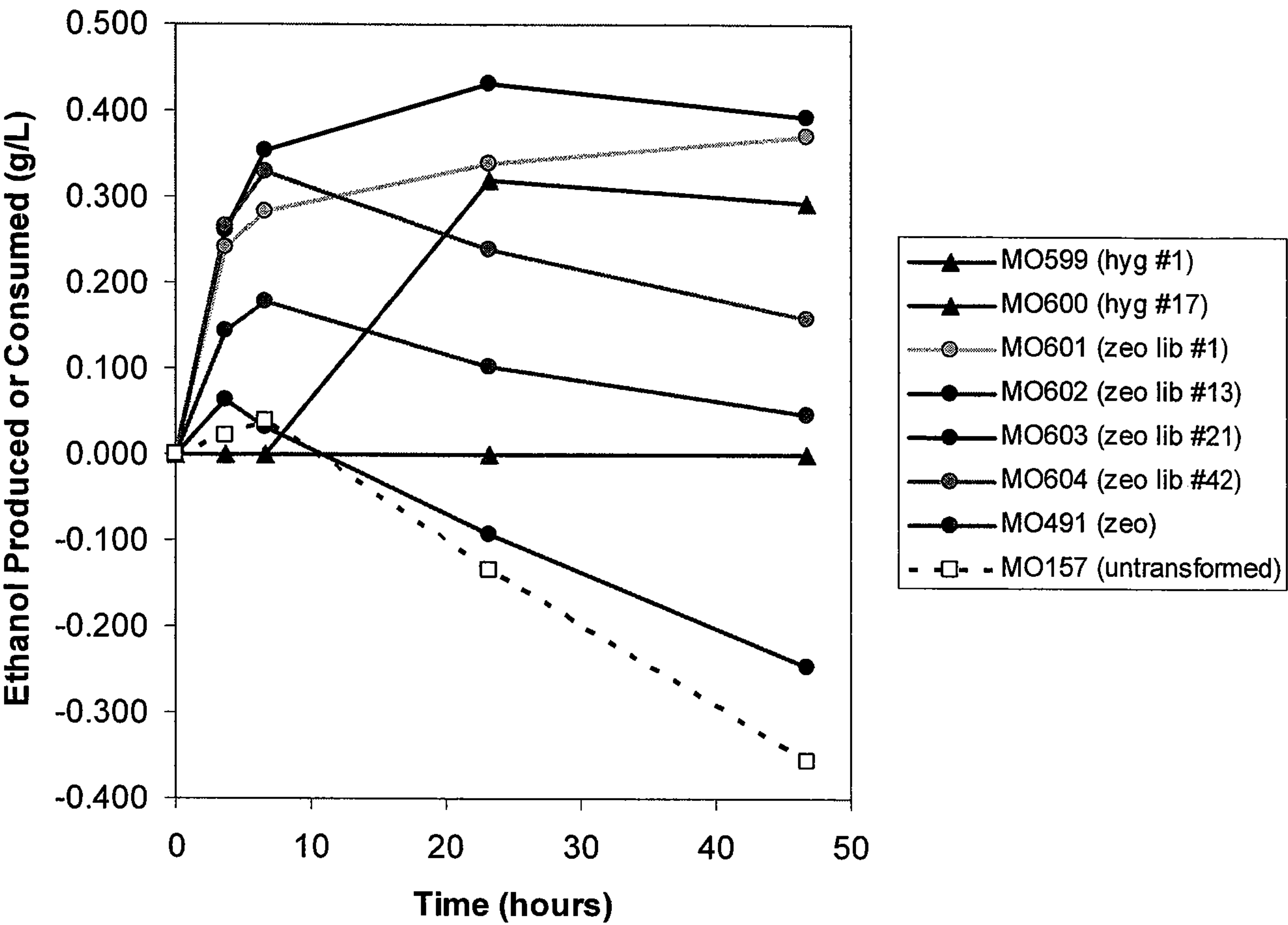
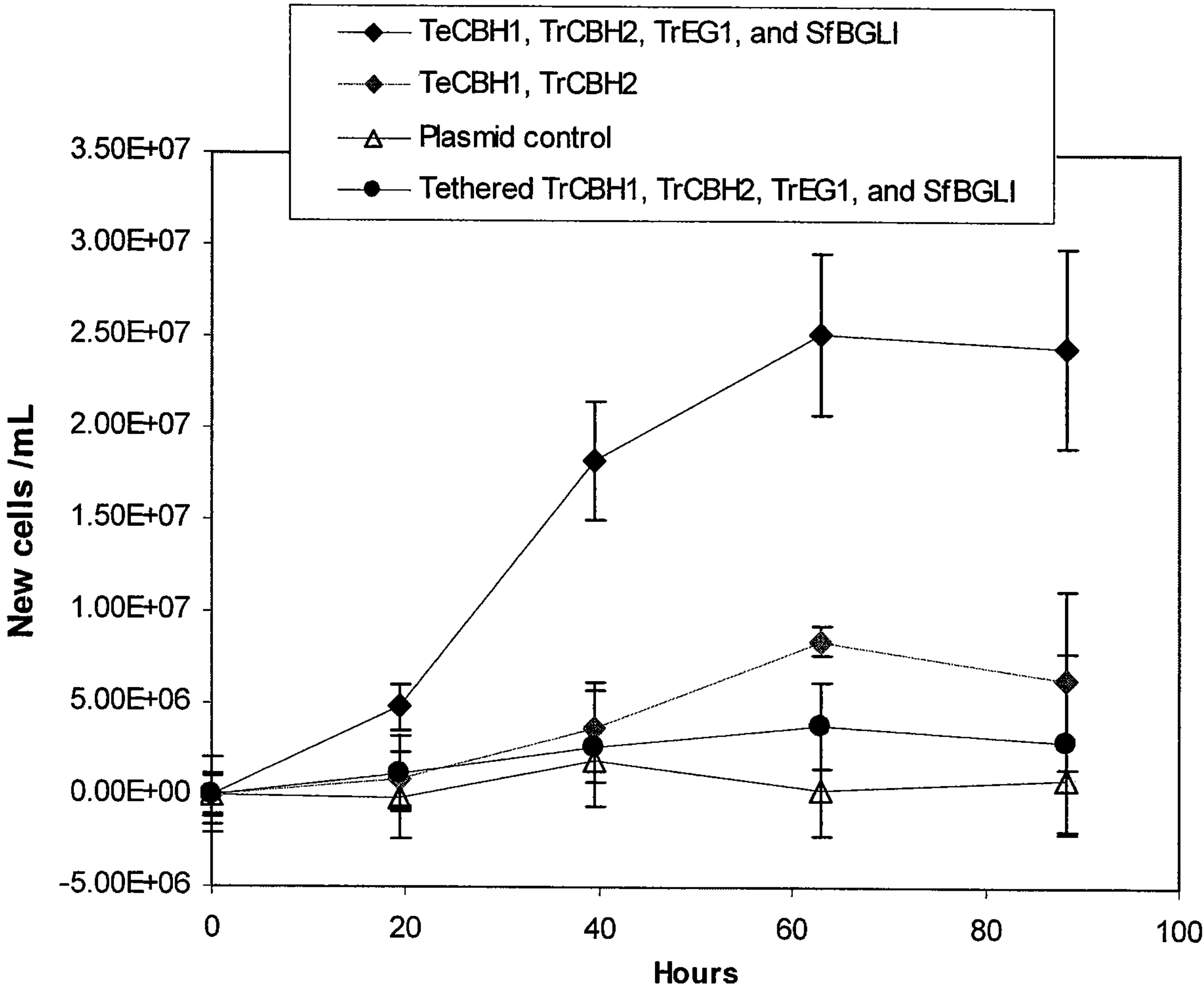


Figure 5



6/42

Figure 6

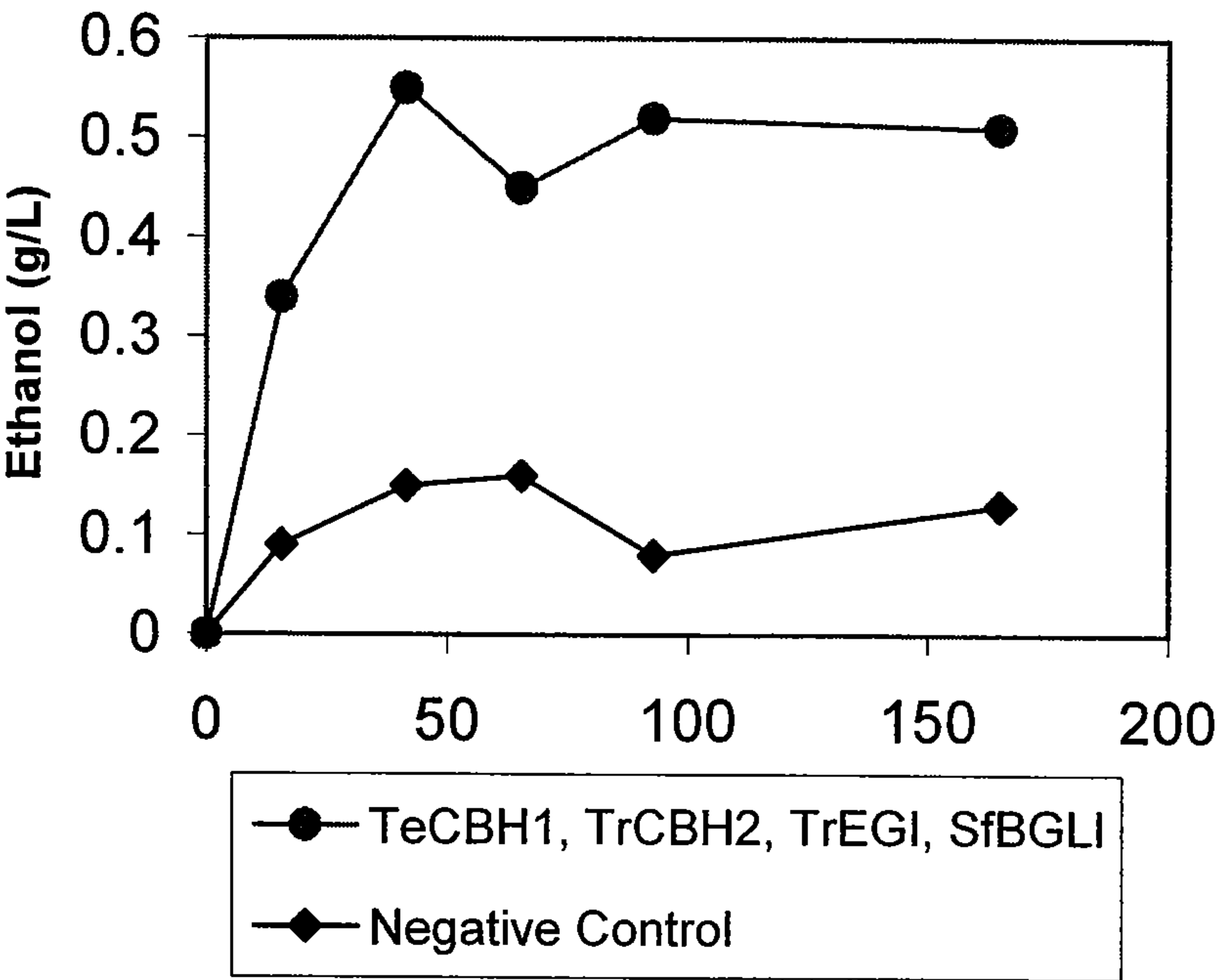


Figure 7

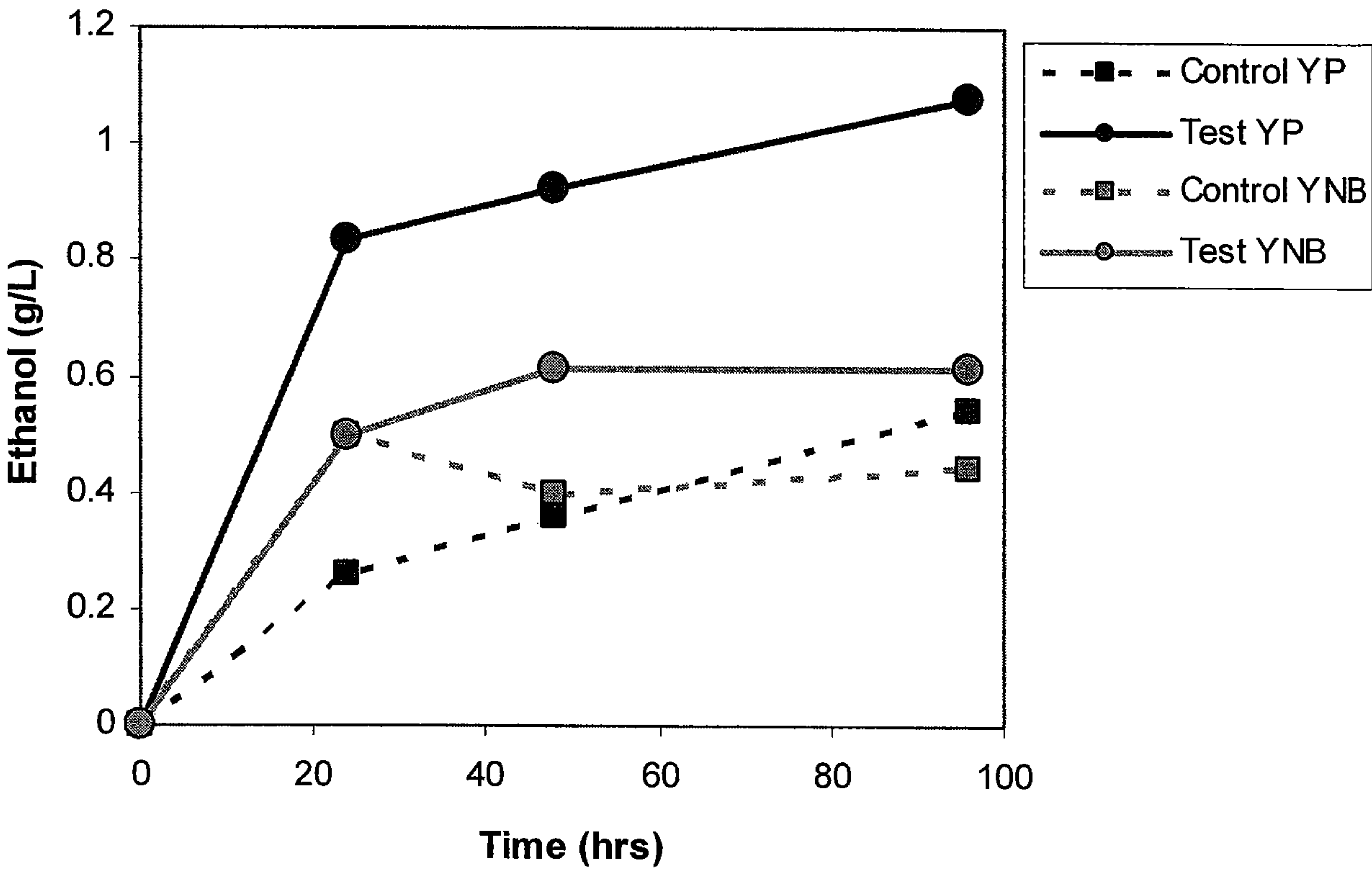


Figure 8

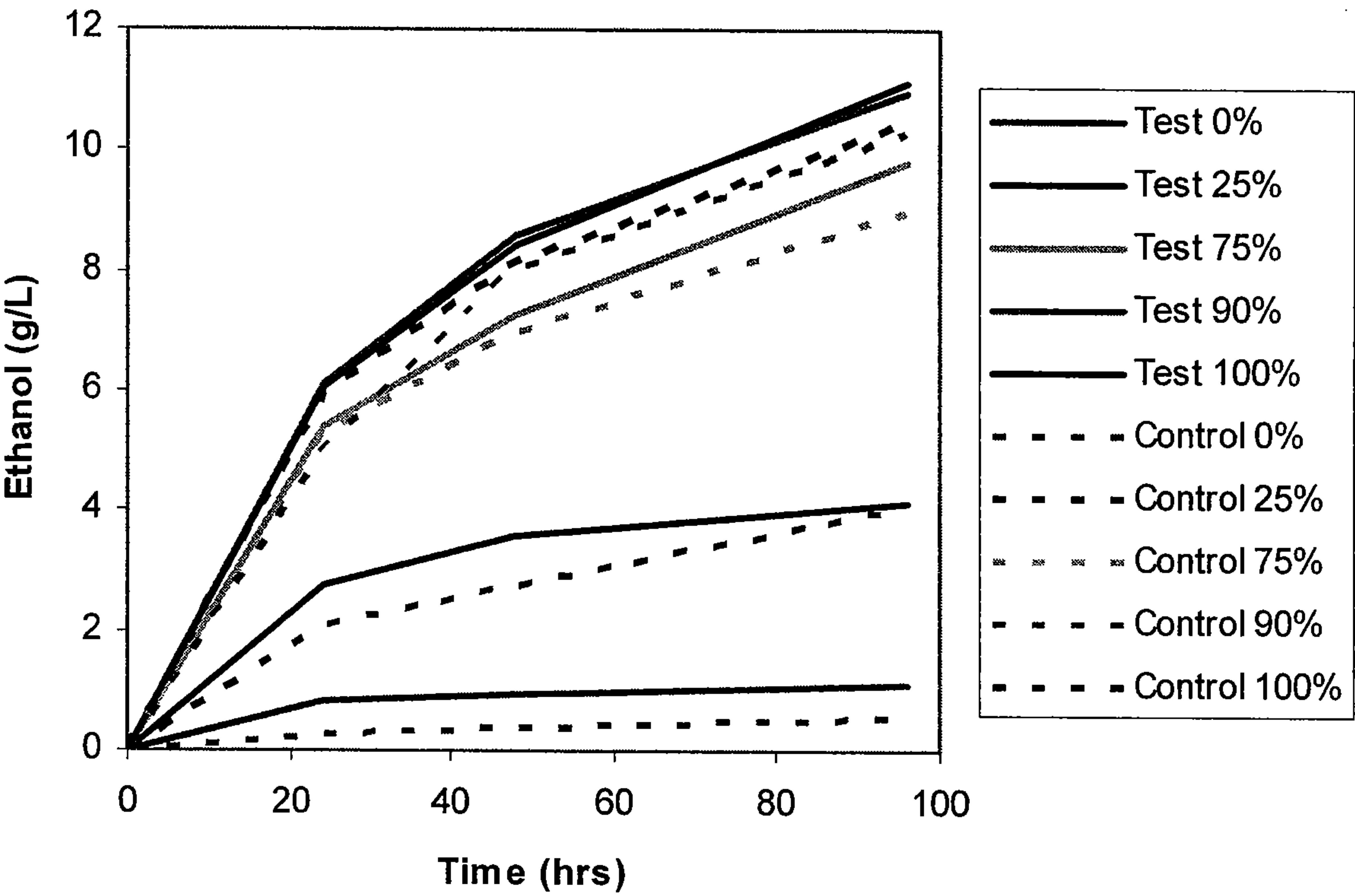


Figure 9

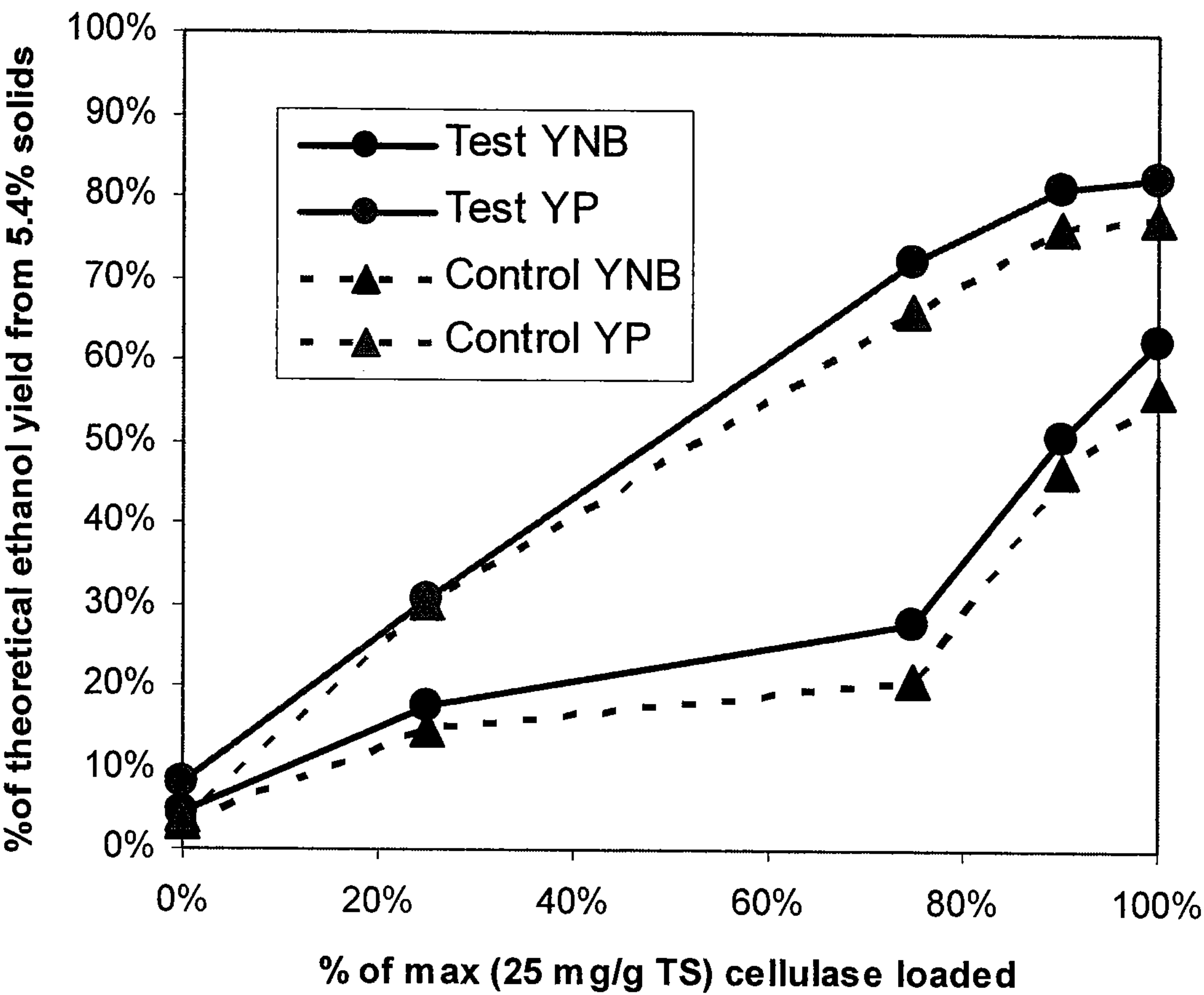


Figure 10

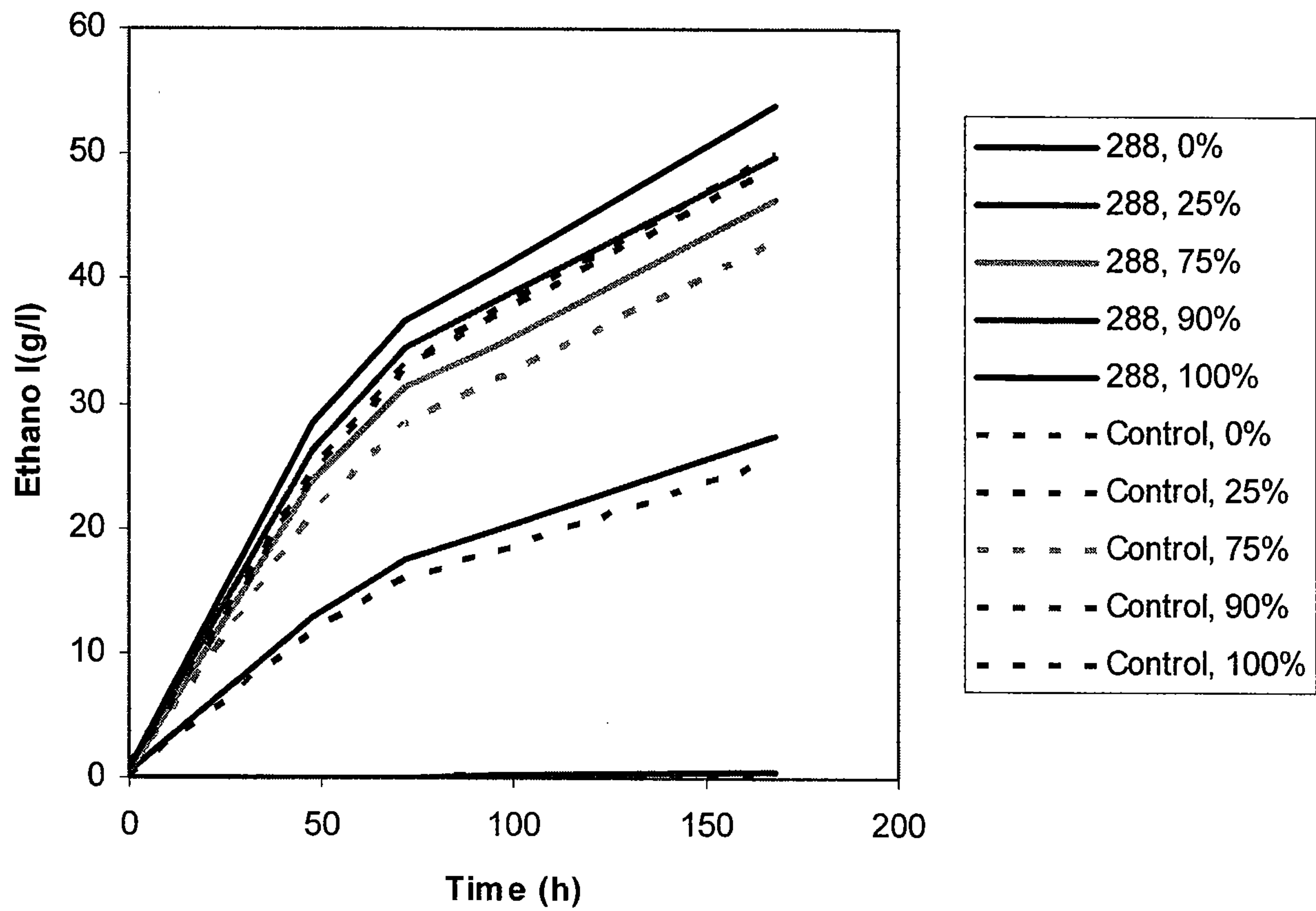


Figure 11

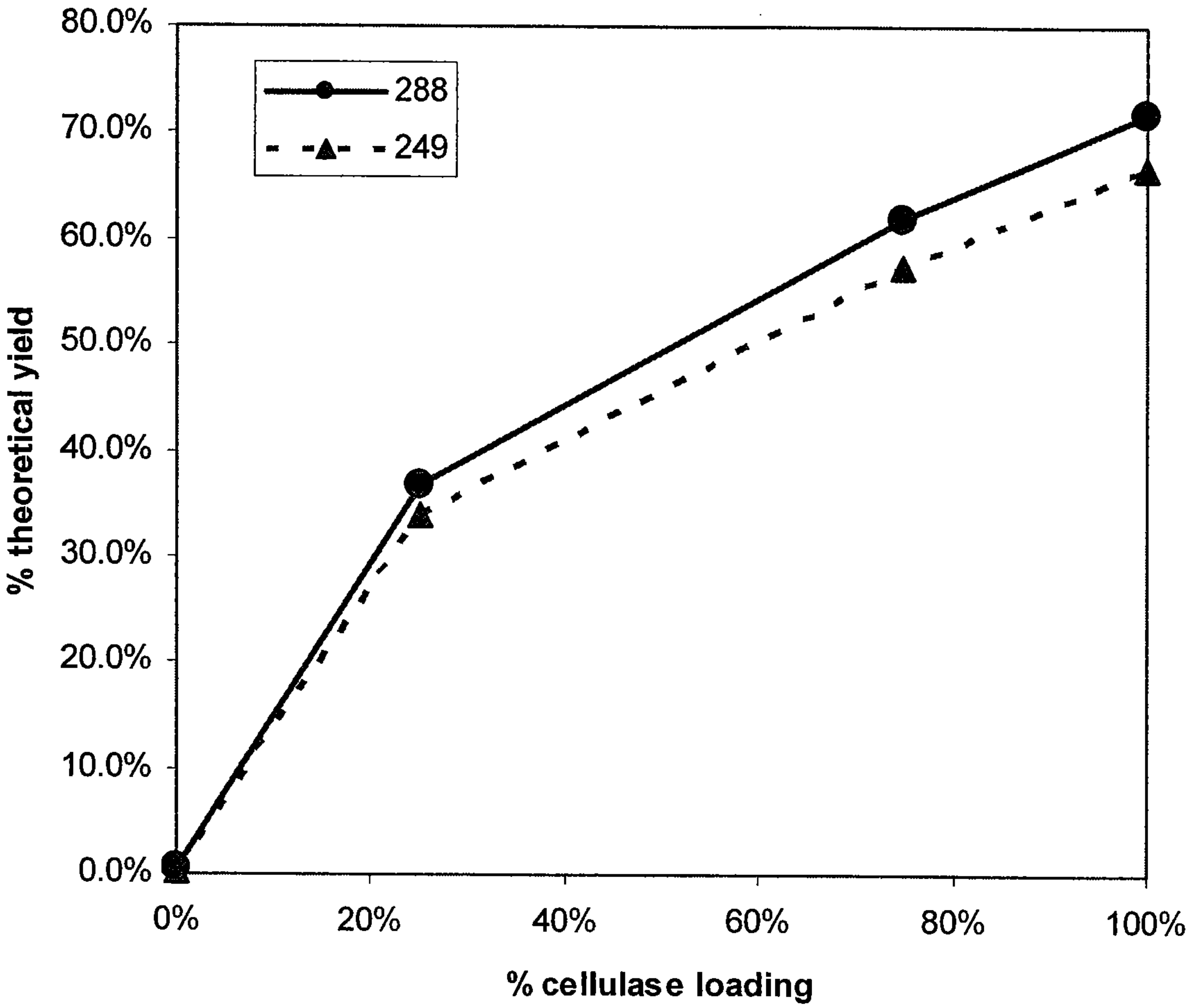
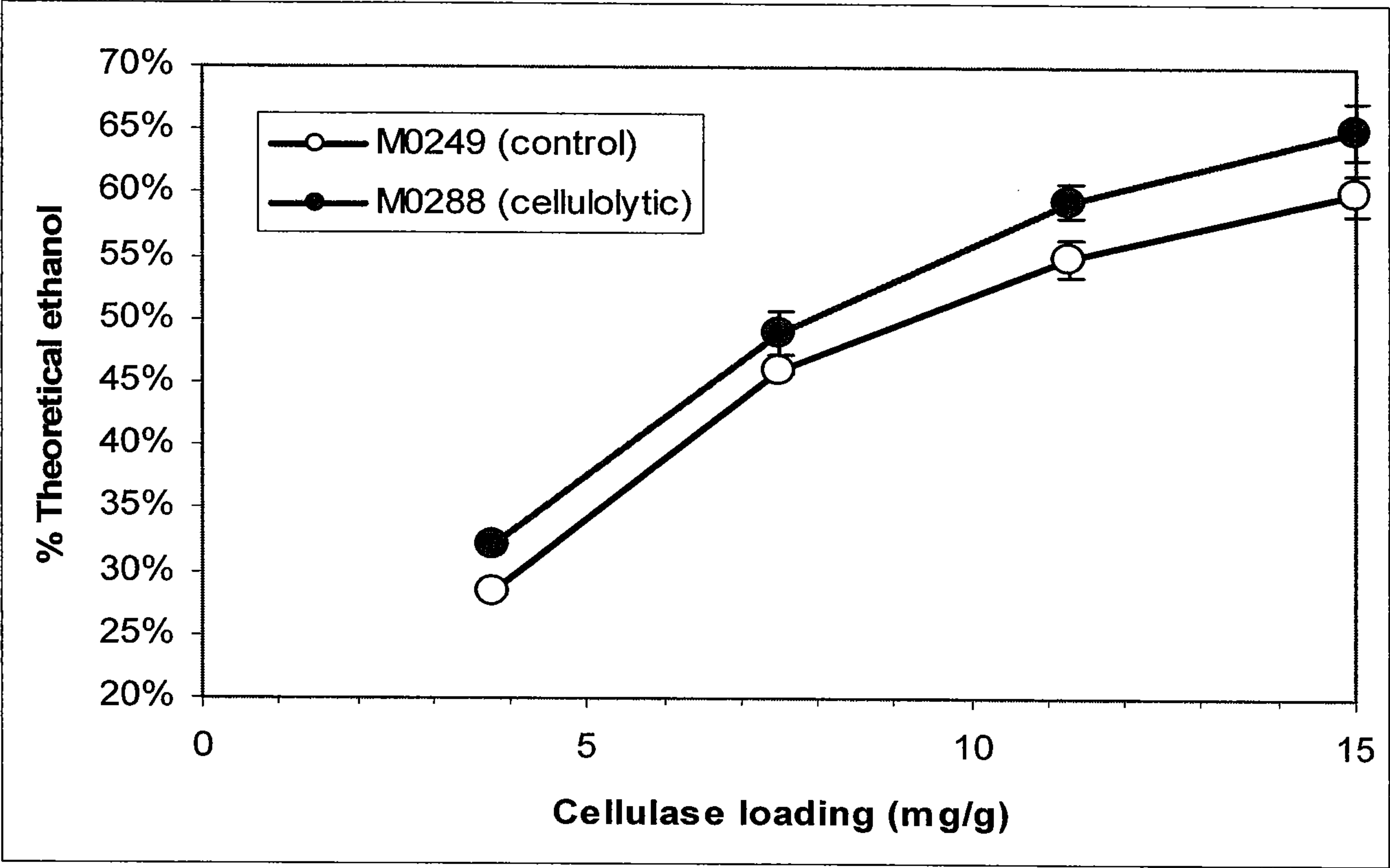


Figure 12



13/42

Figure 13

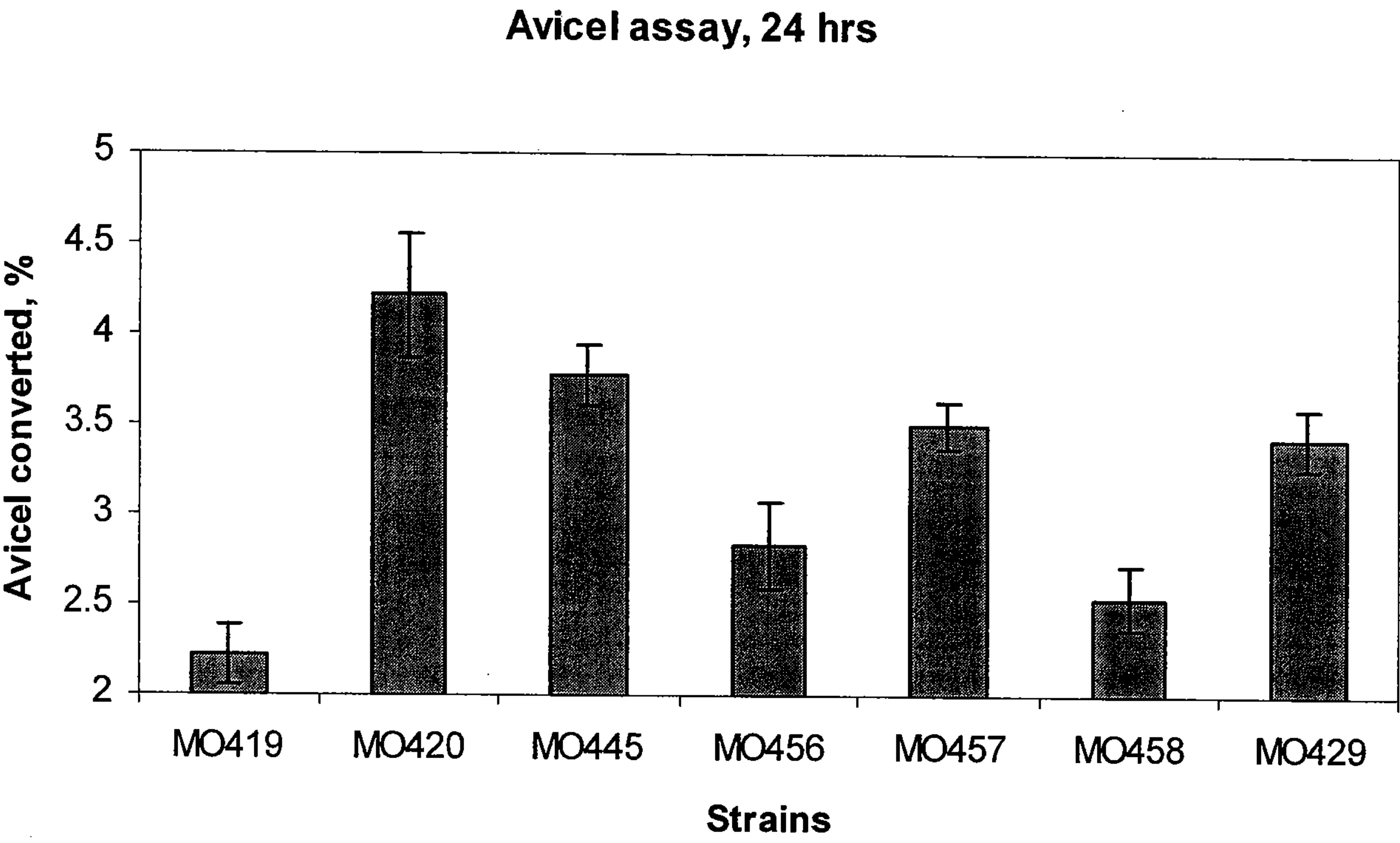


Figure 14

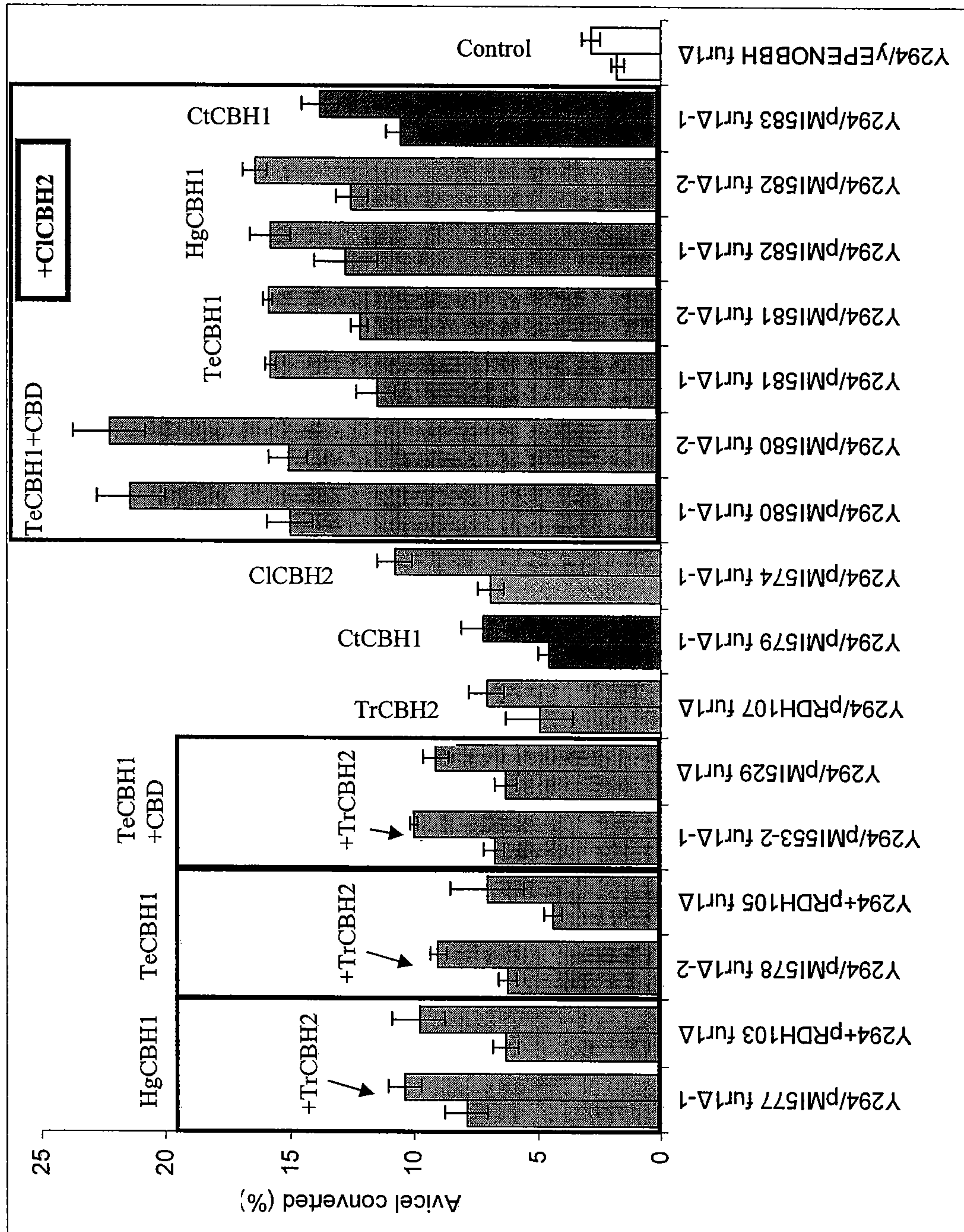
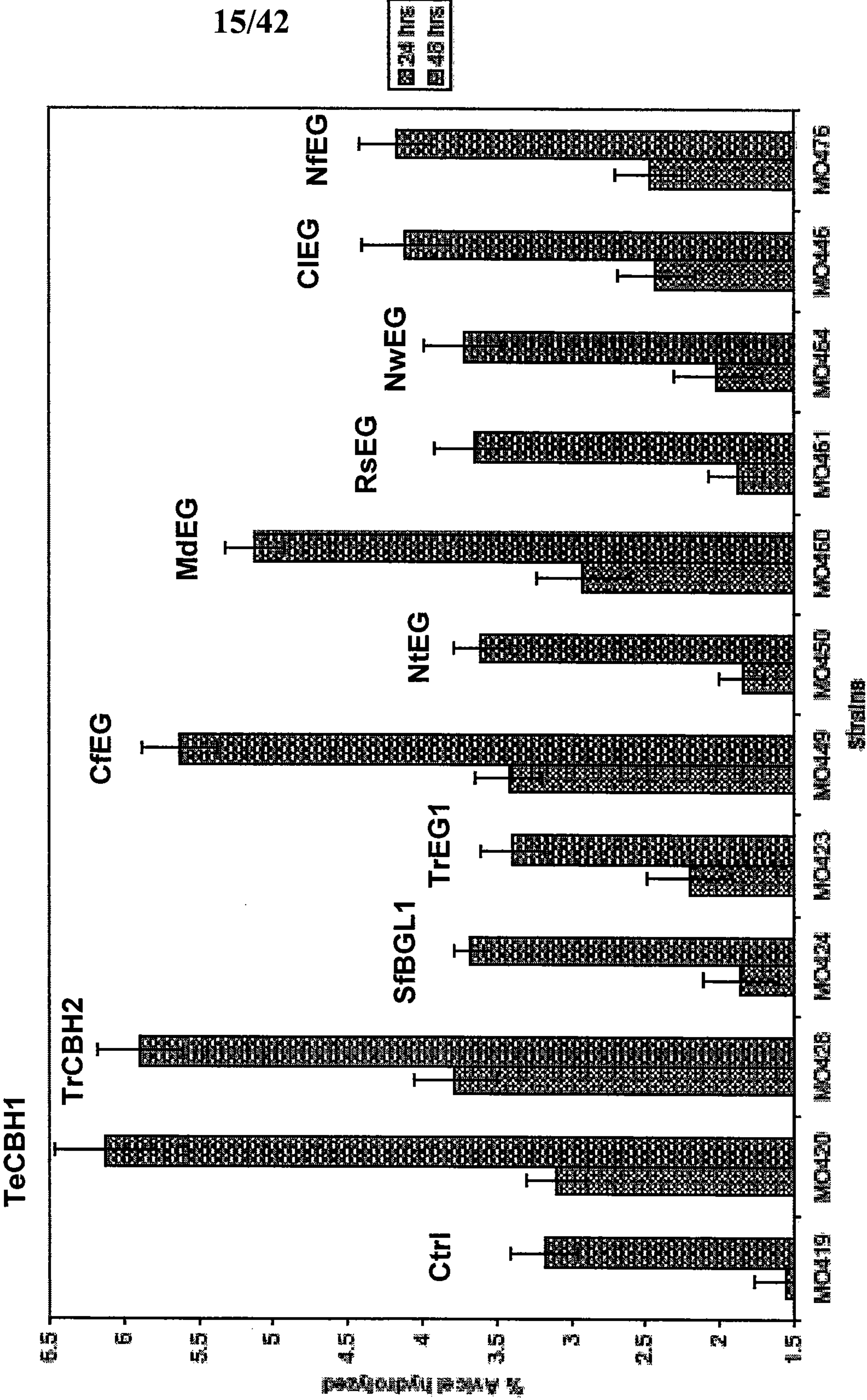


Figure 15



16/42

Figure 16

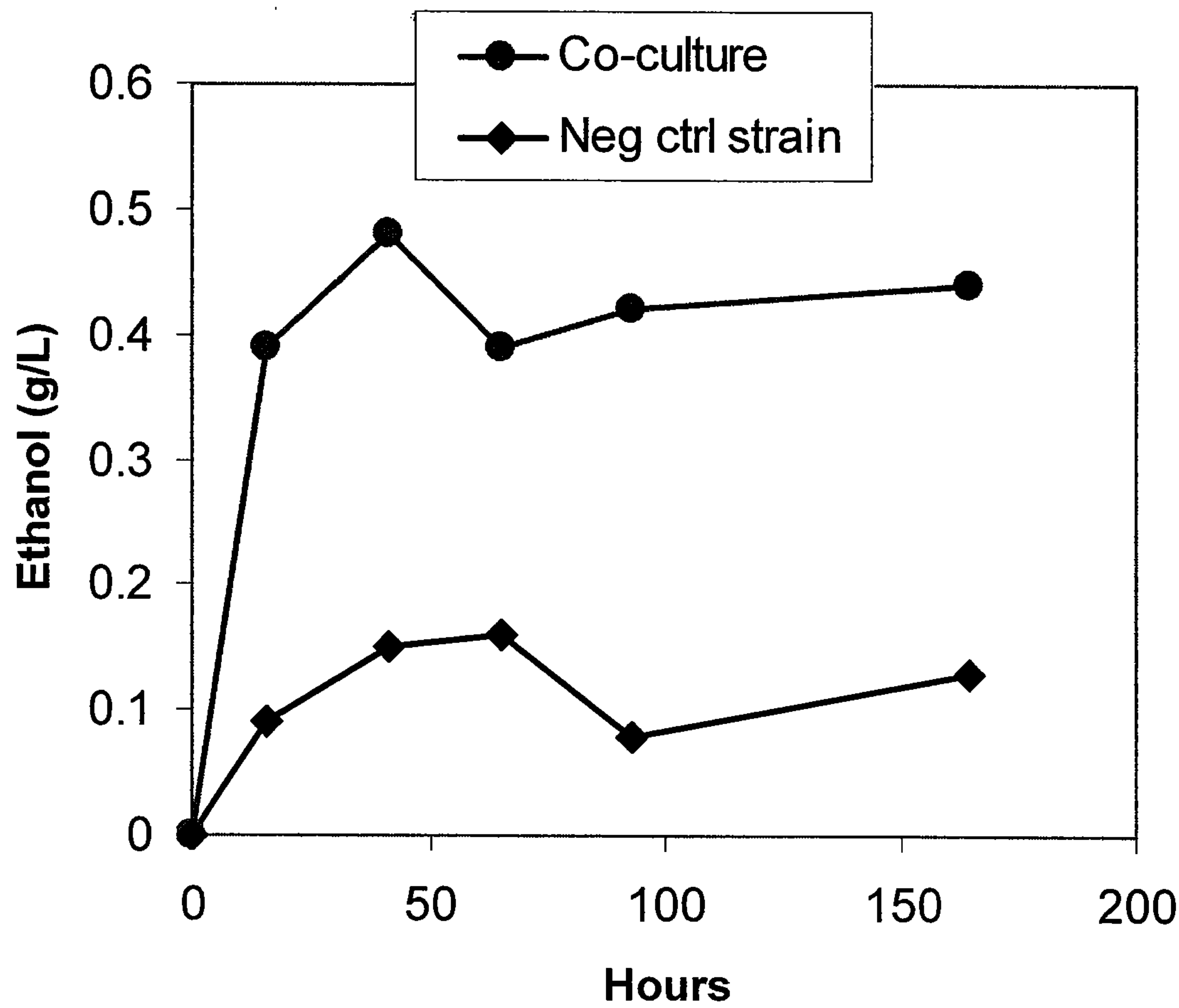


Figure 17

Zero cellulase loading ethanol production in SSF

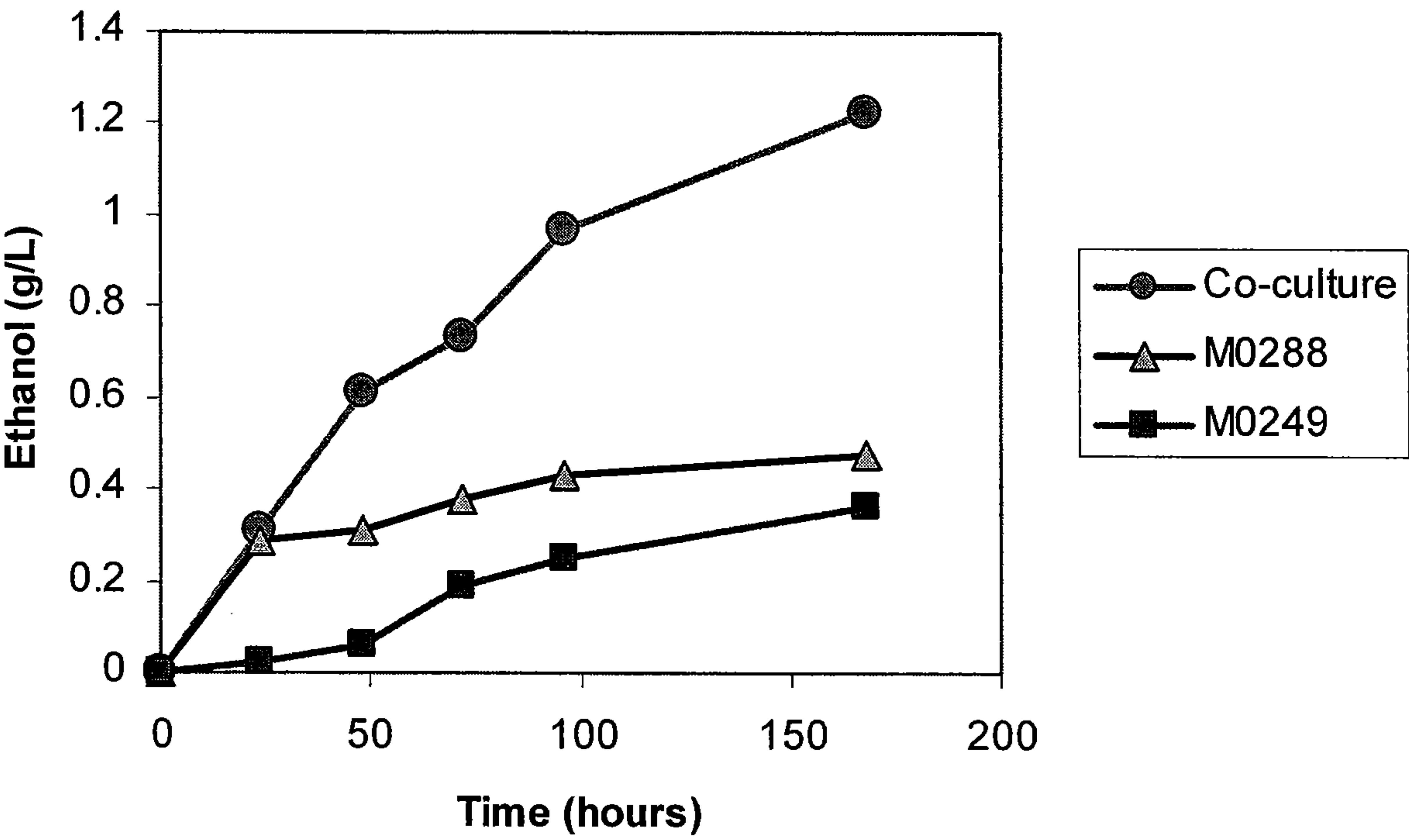
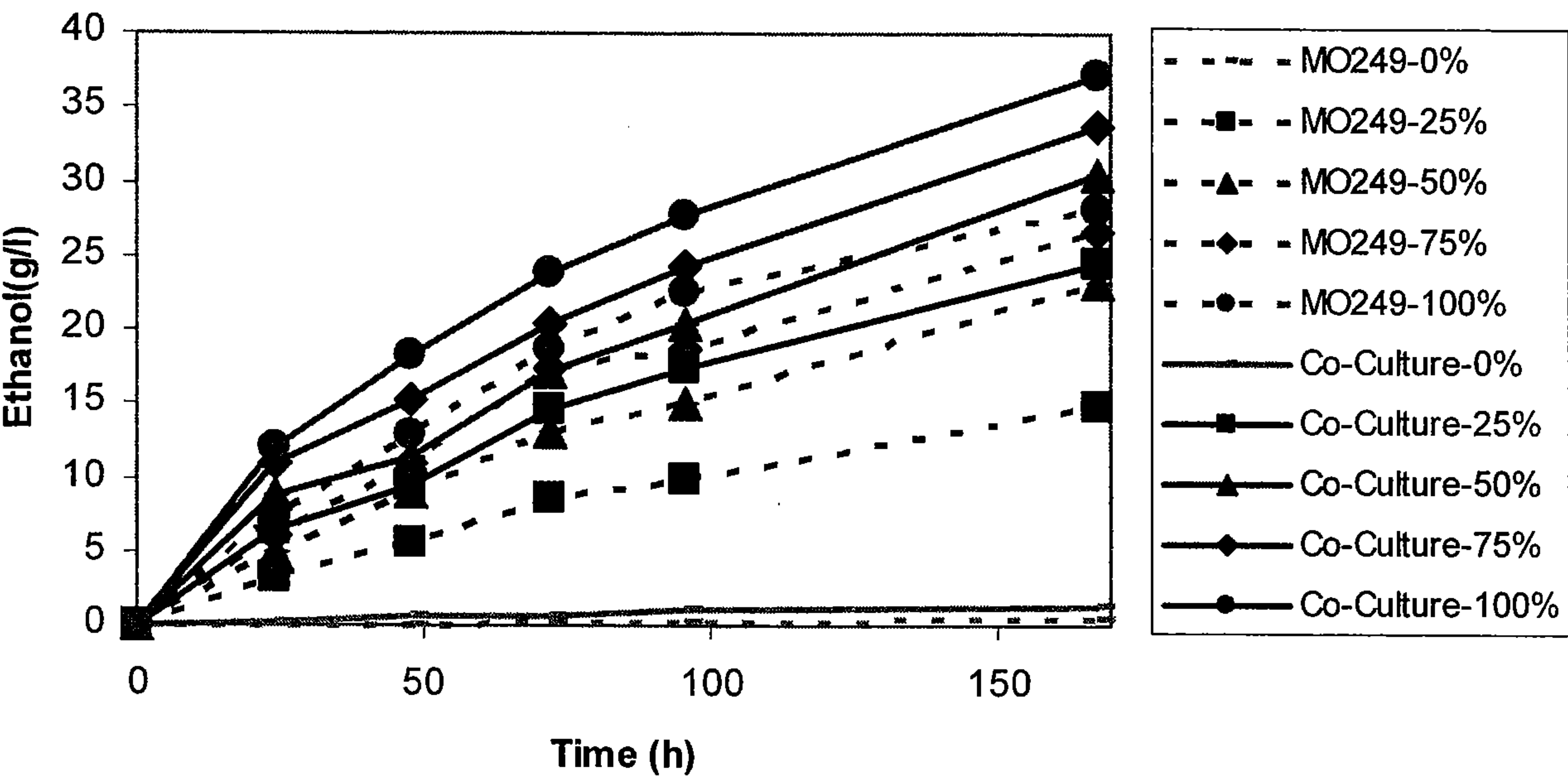
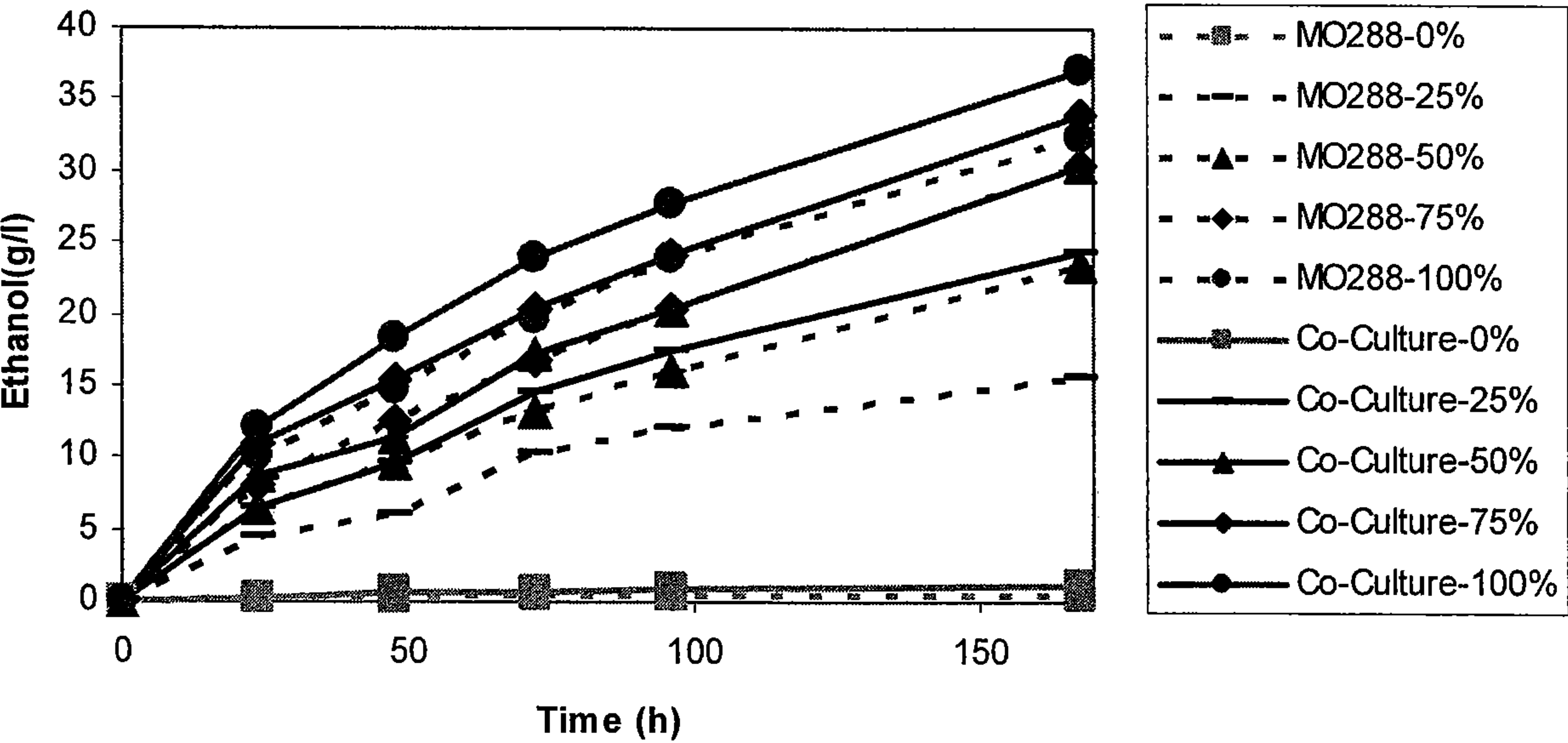


Figure 18

SSF 10mg/g of cellulase=100%, 10% Avicel substrate



SSF 10mg/g of cellulase=100%, 10% Avicel substrate



19/42

Figure 19

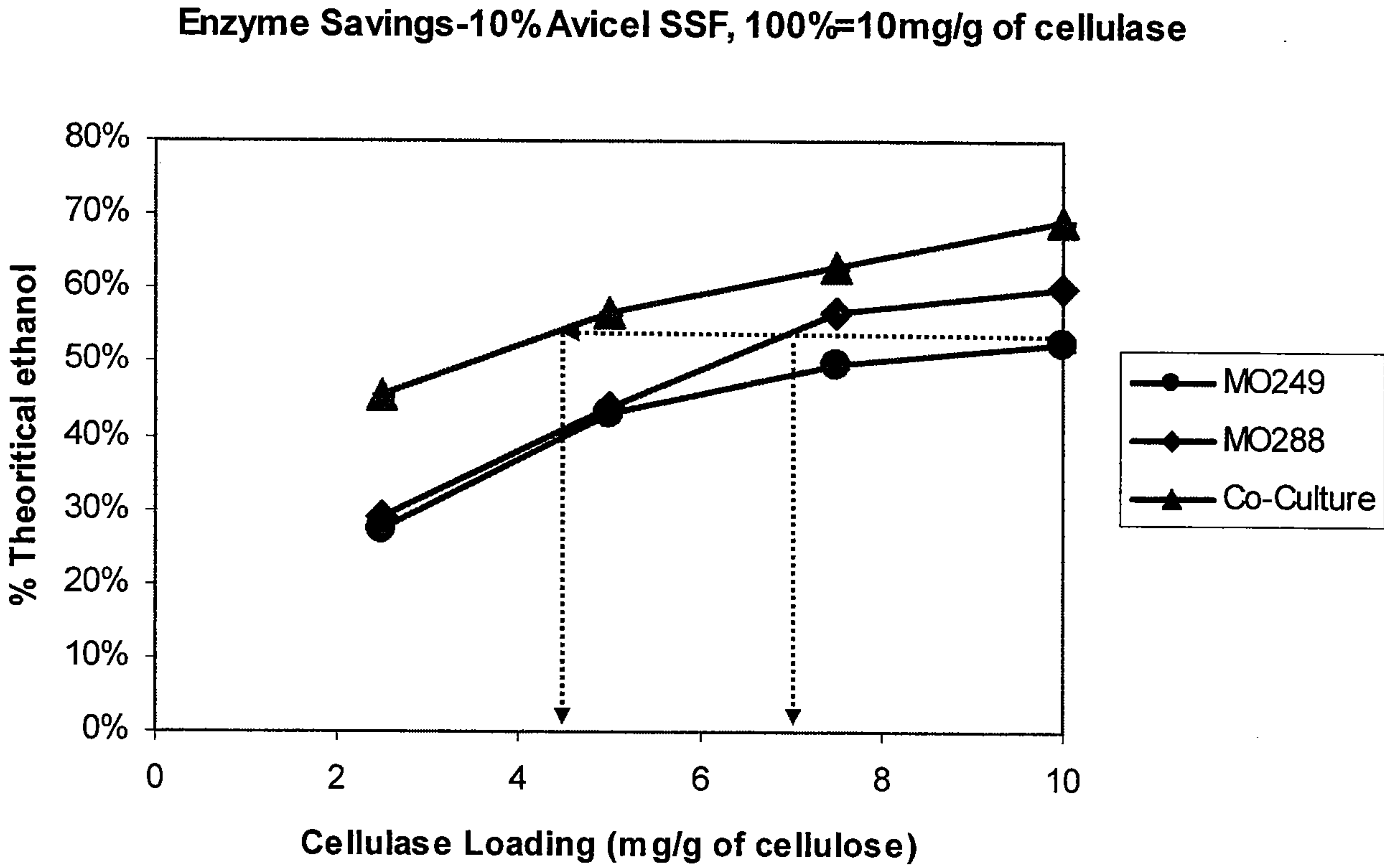


Figure 20

20/42

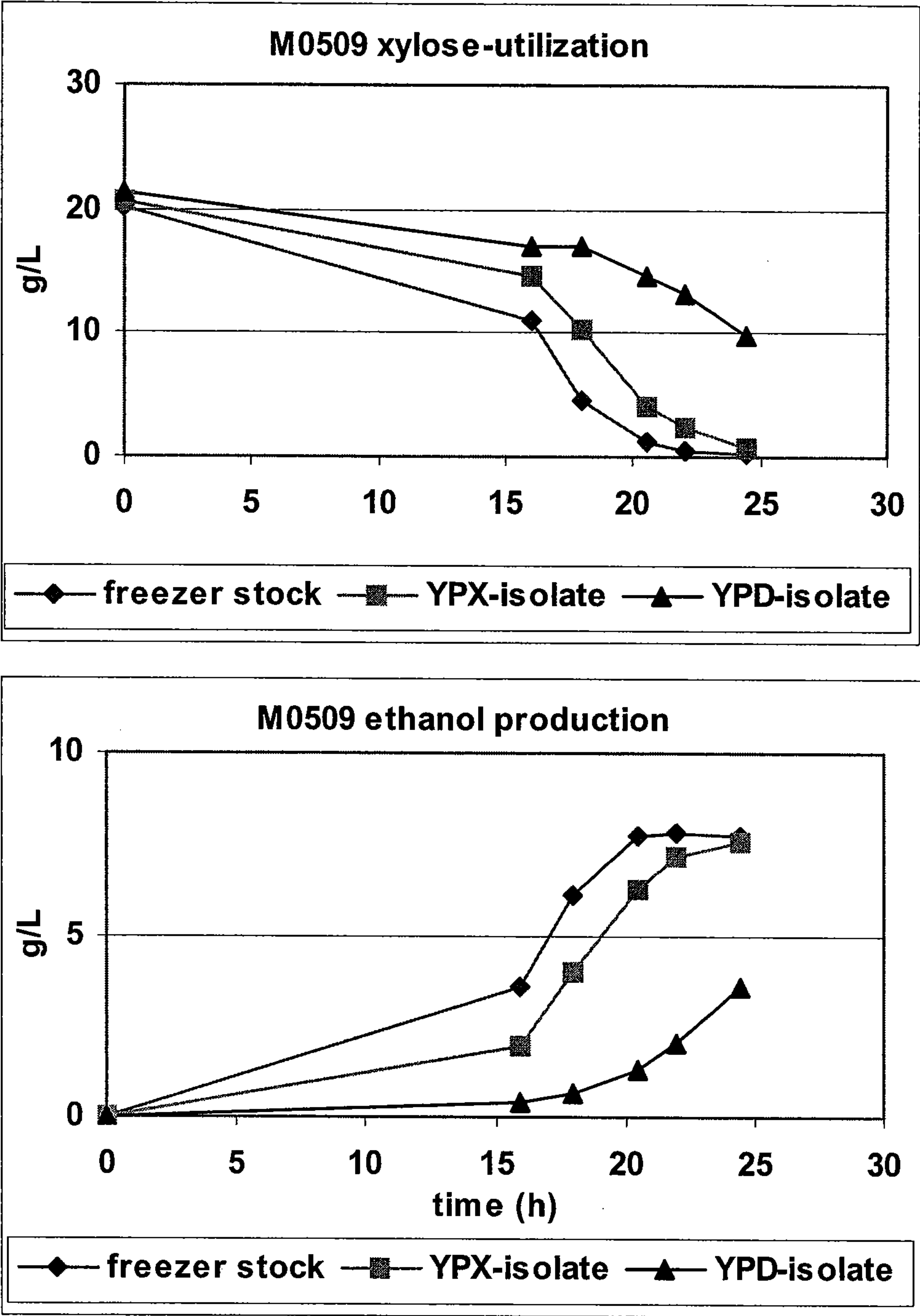
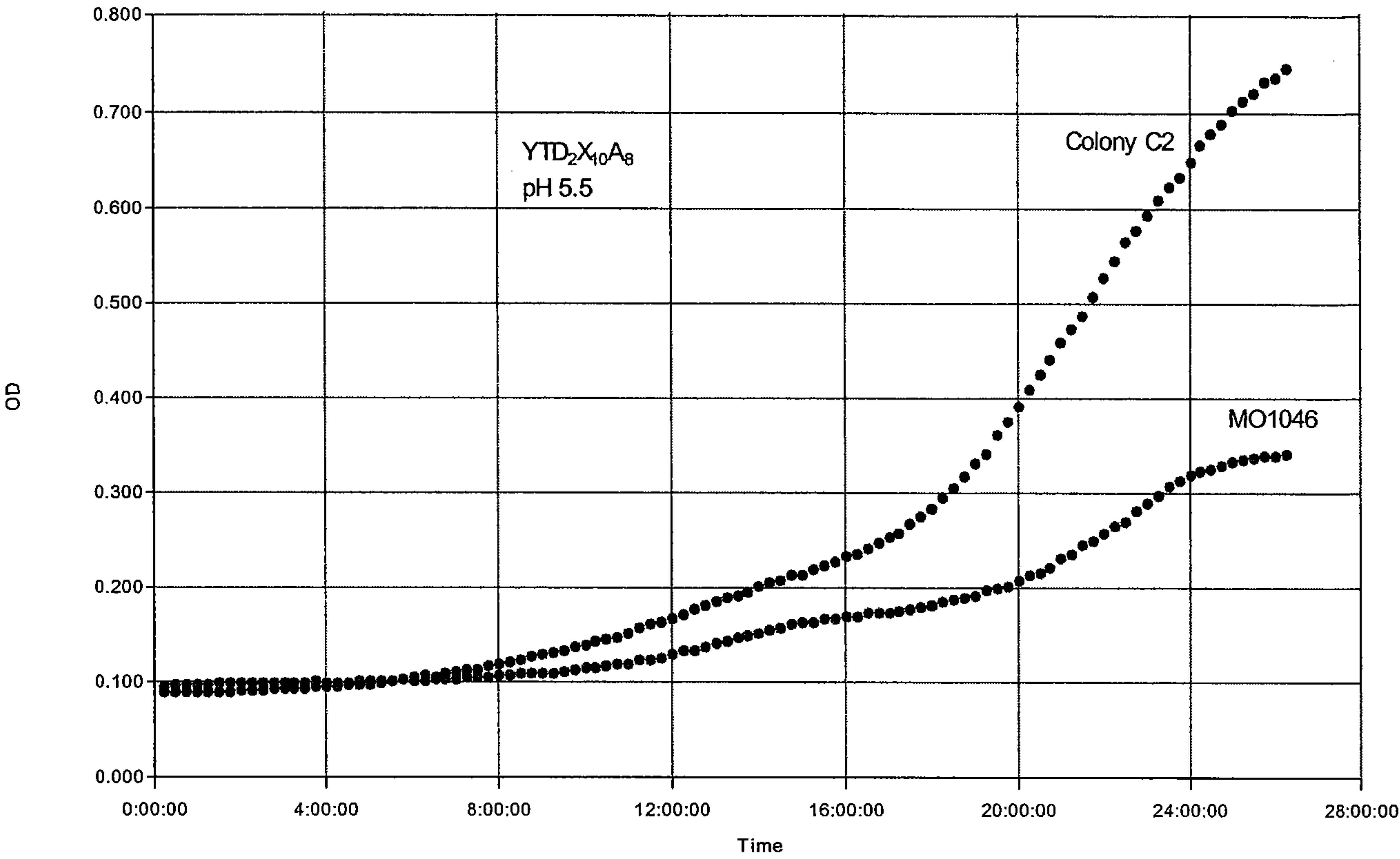


Figure 21



22/42

Figure 22

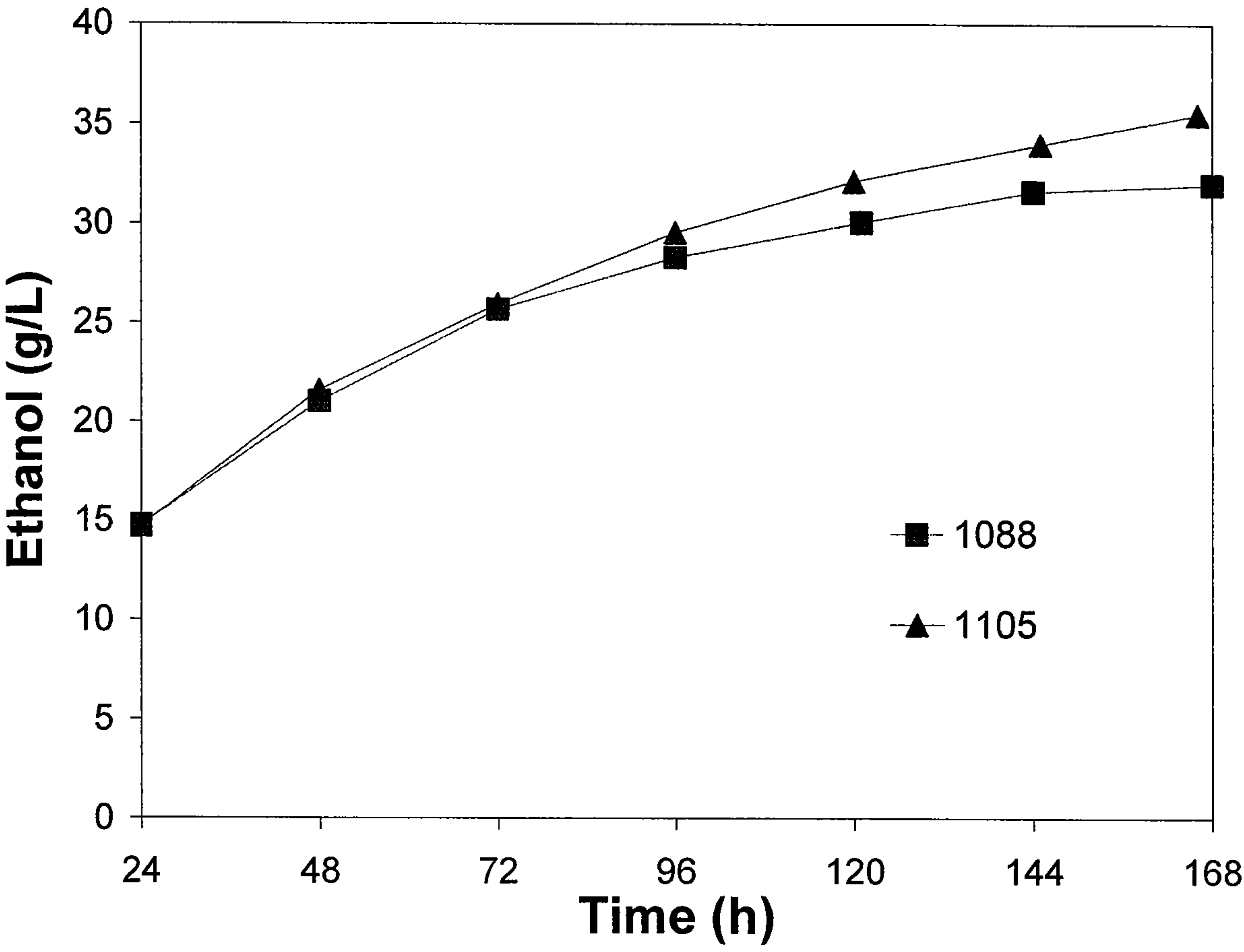
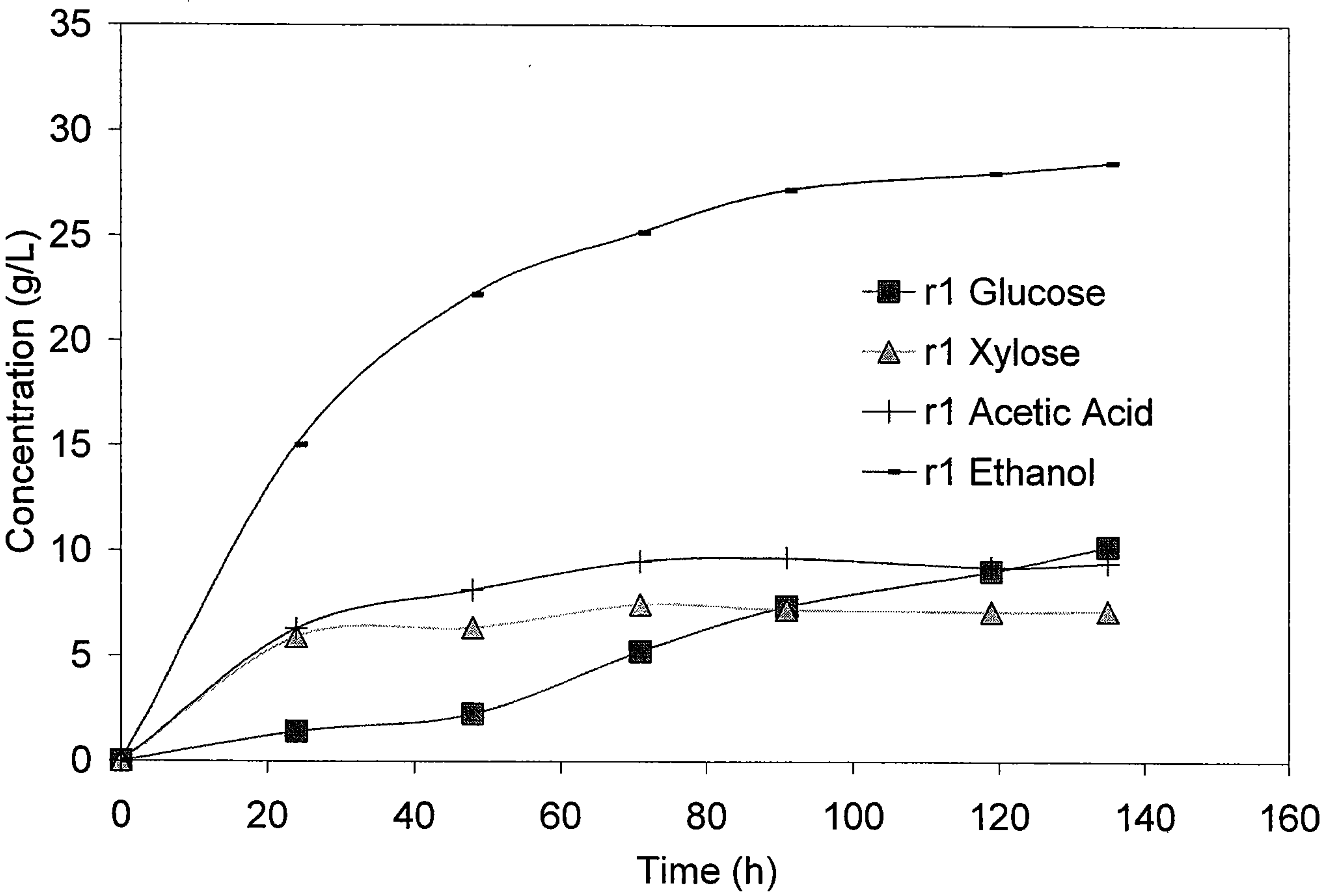


Figure 23



24/42

Figure 24

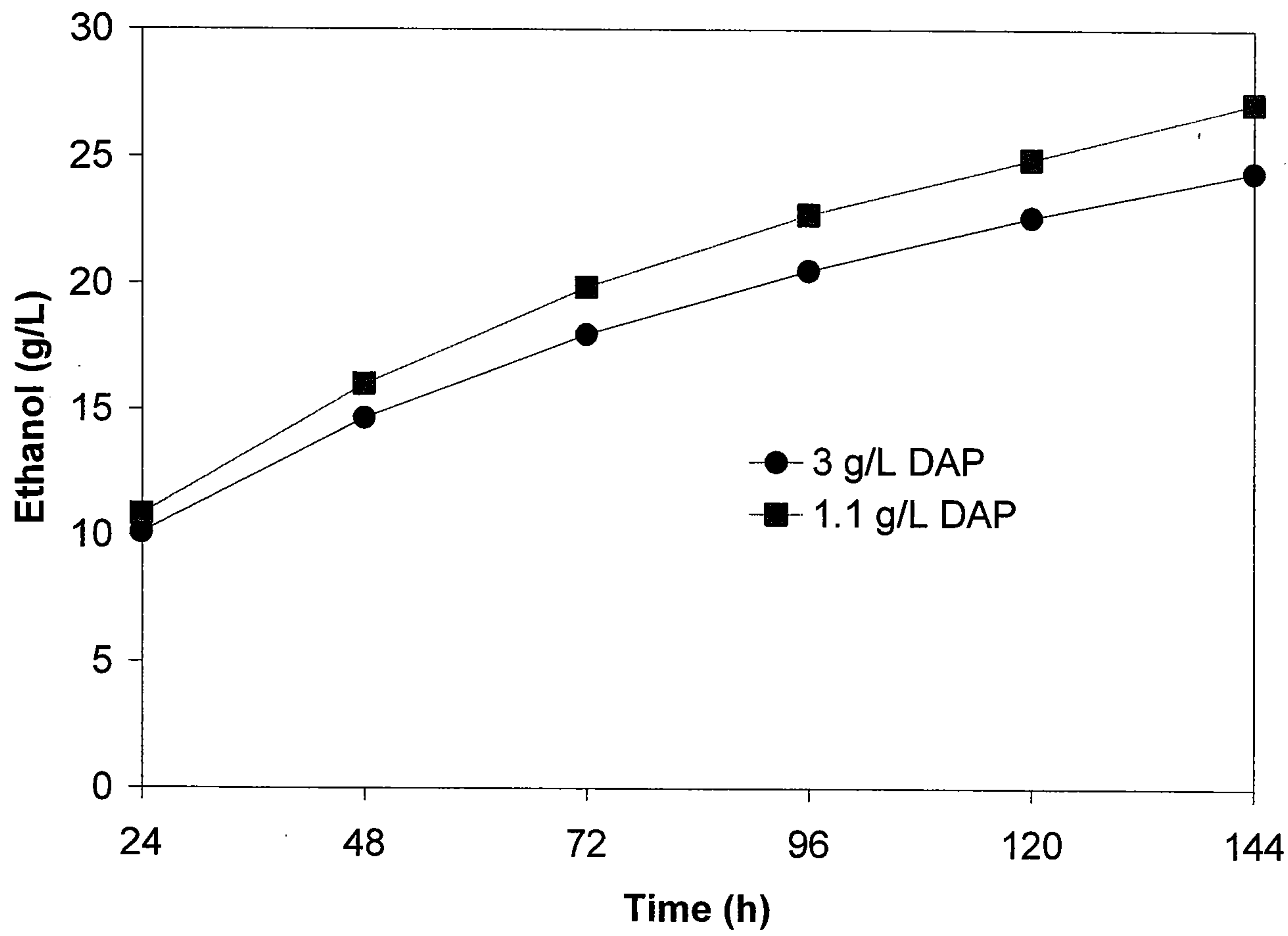
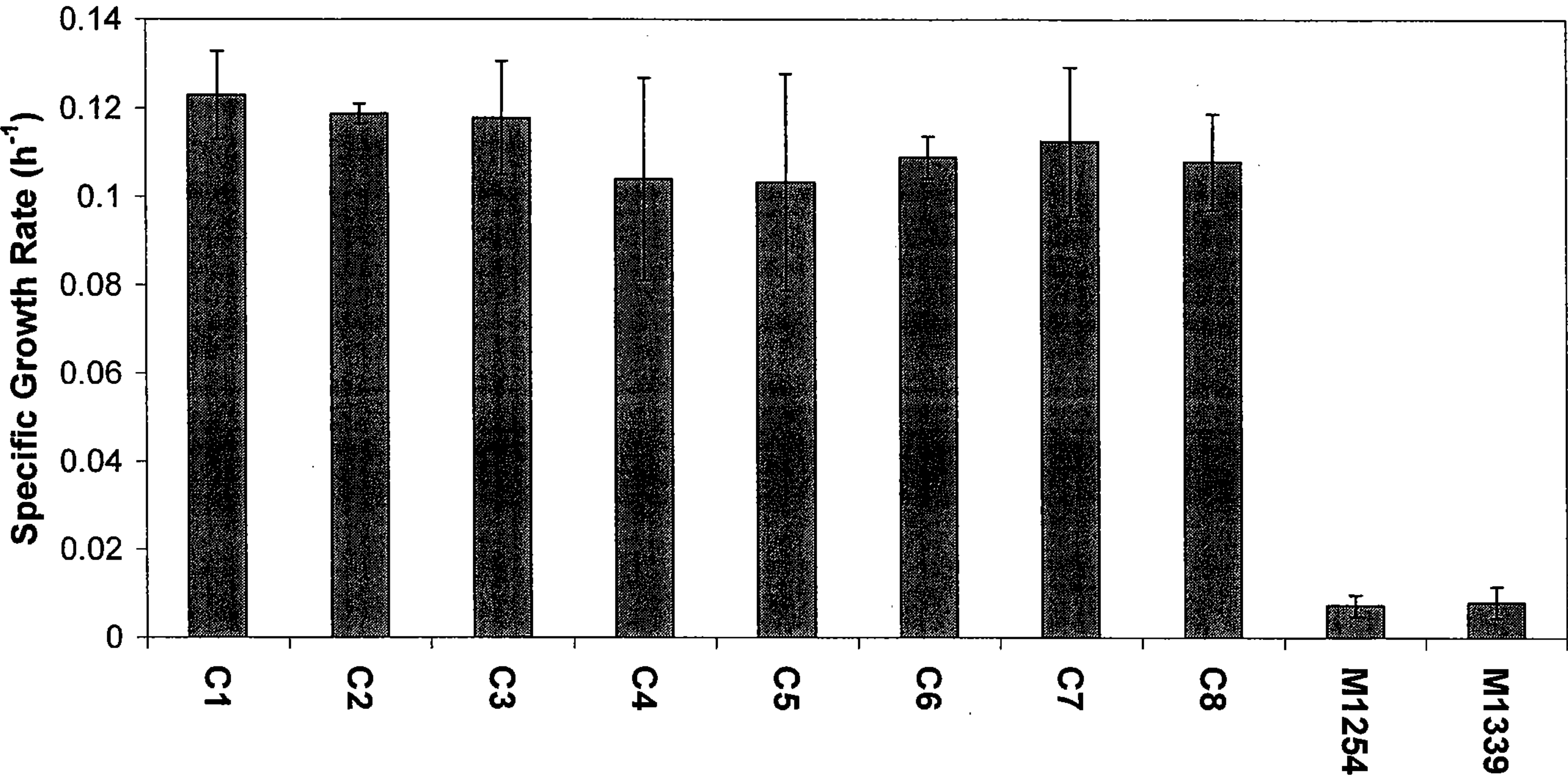
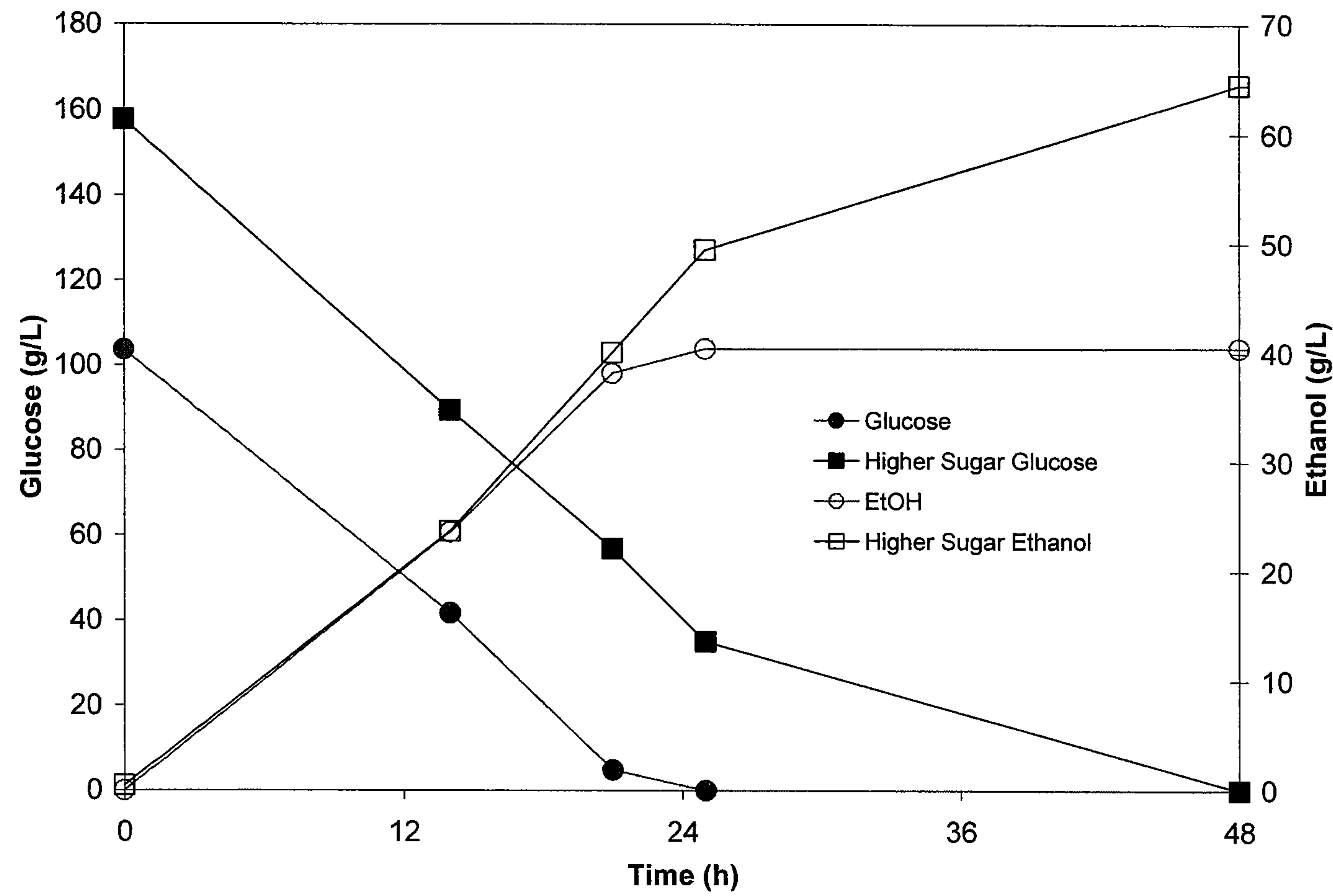


Figure 25



26/42

Figure 26



27/42

Figure 27

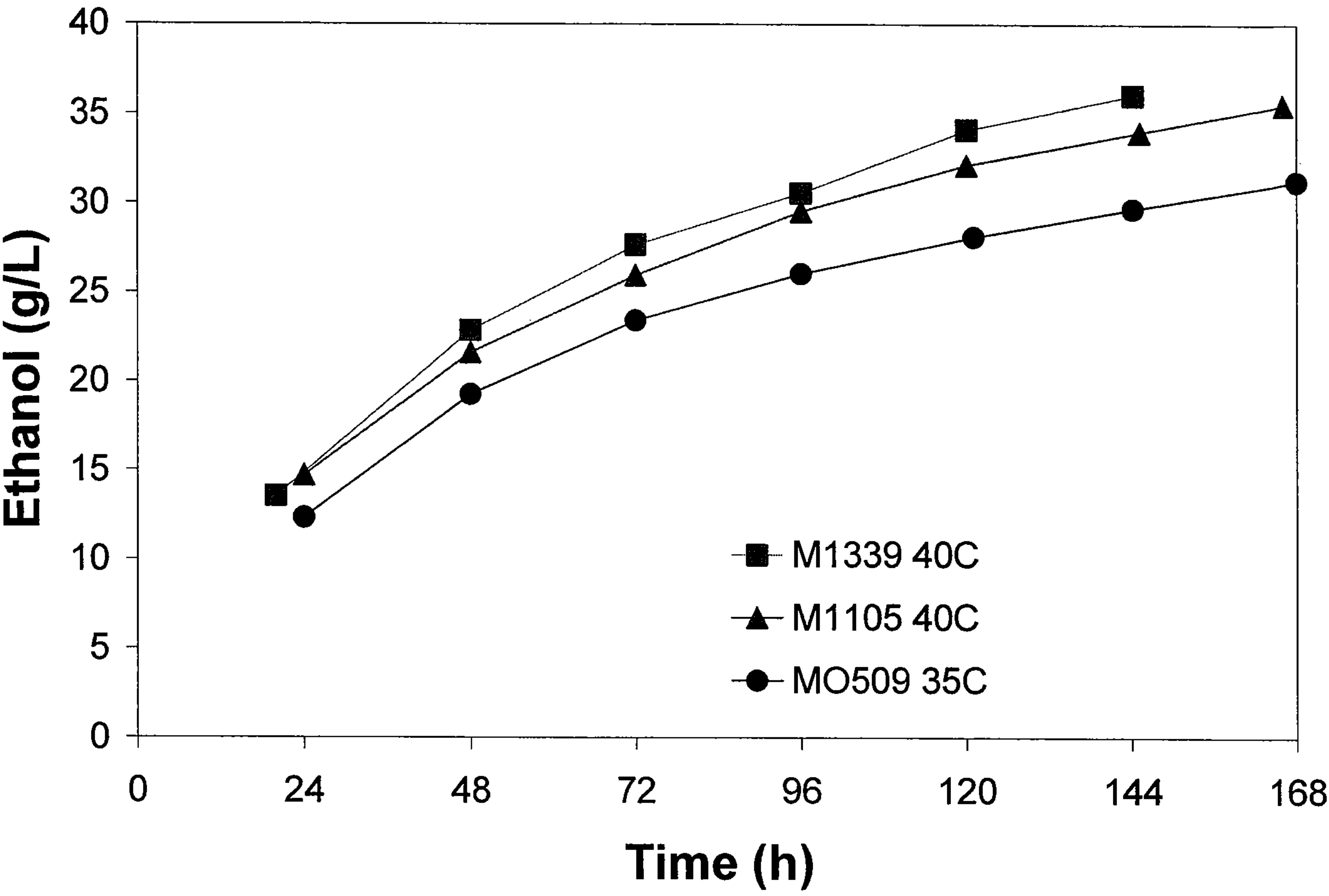
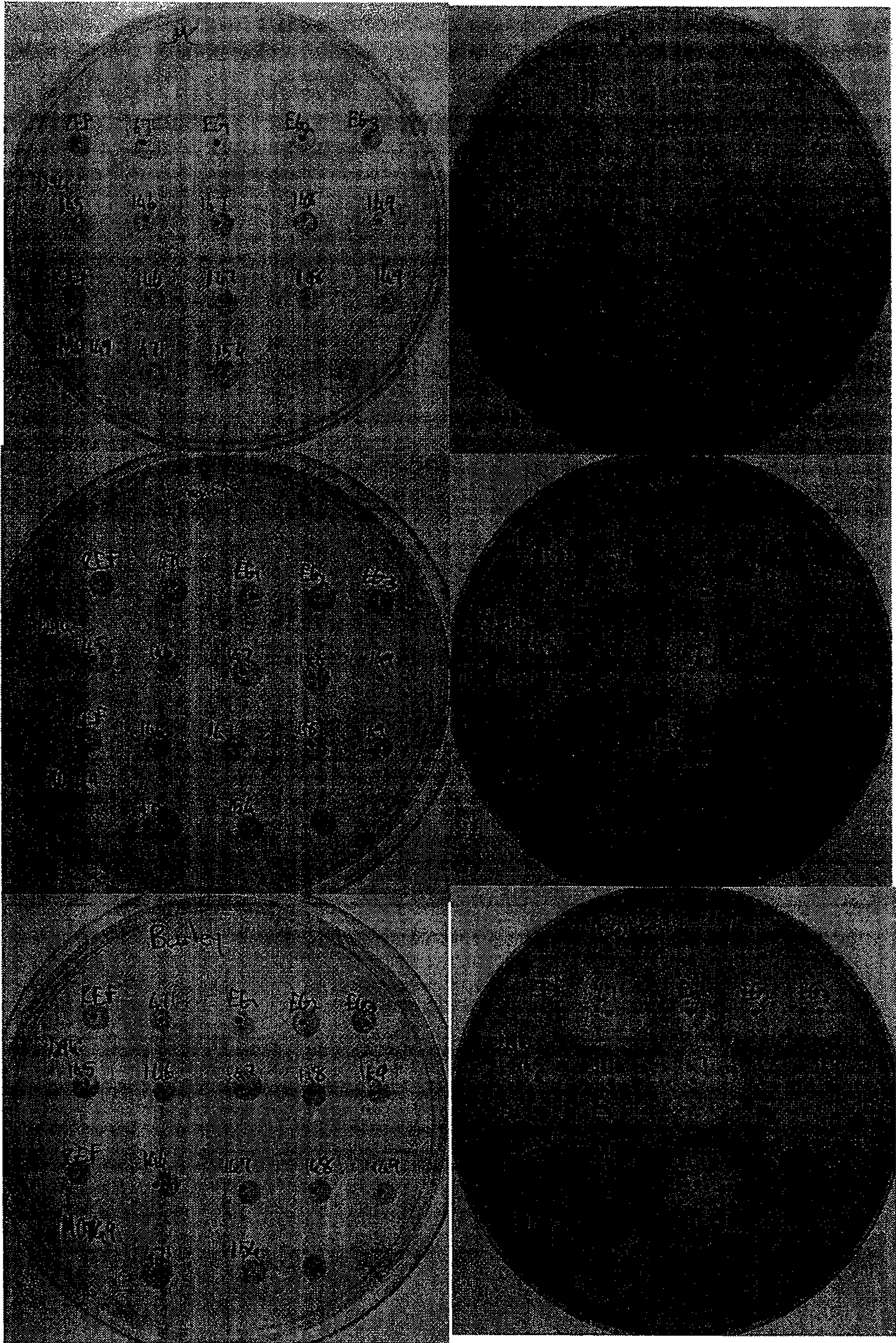
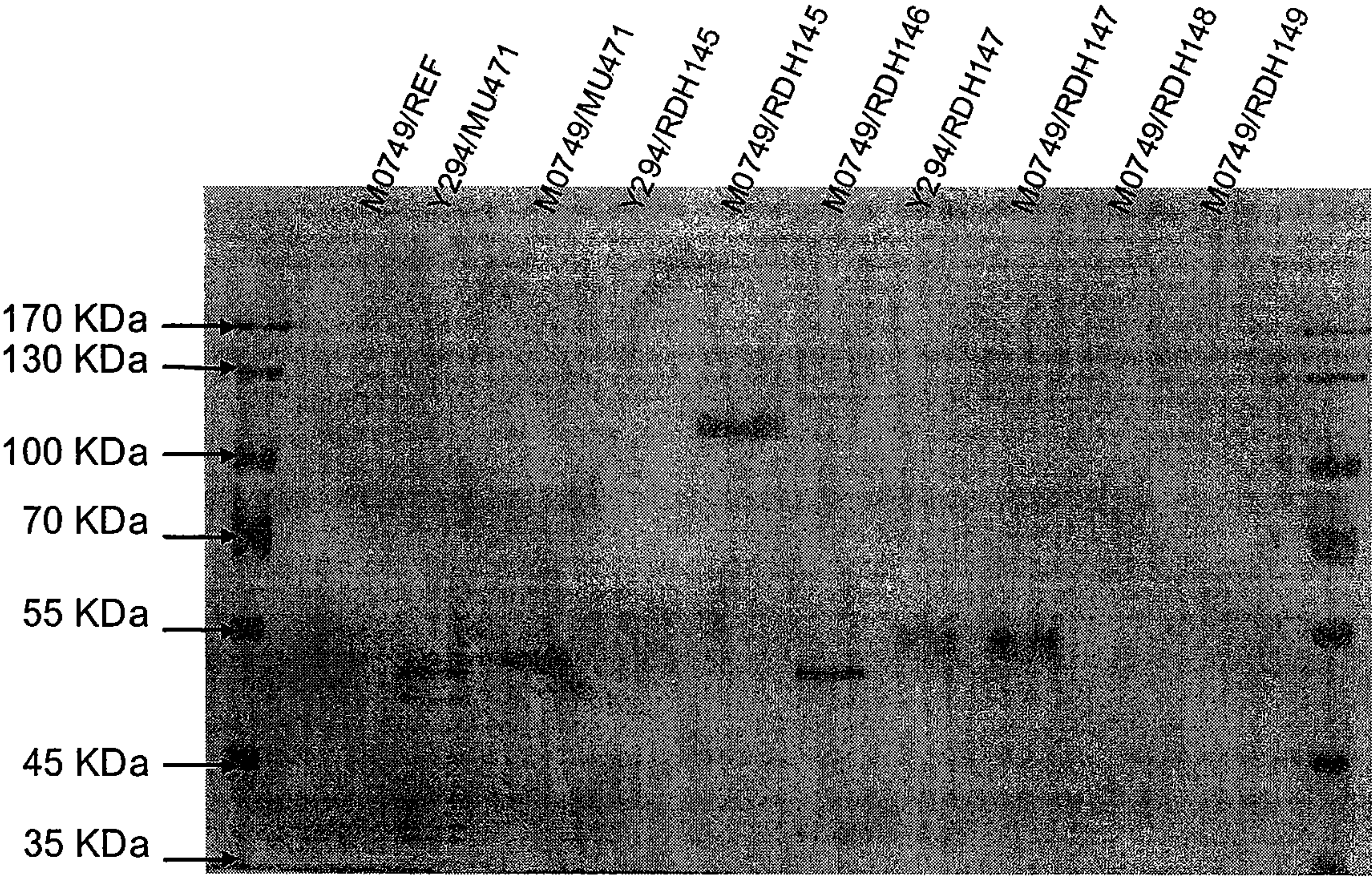


Figure 28



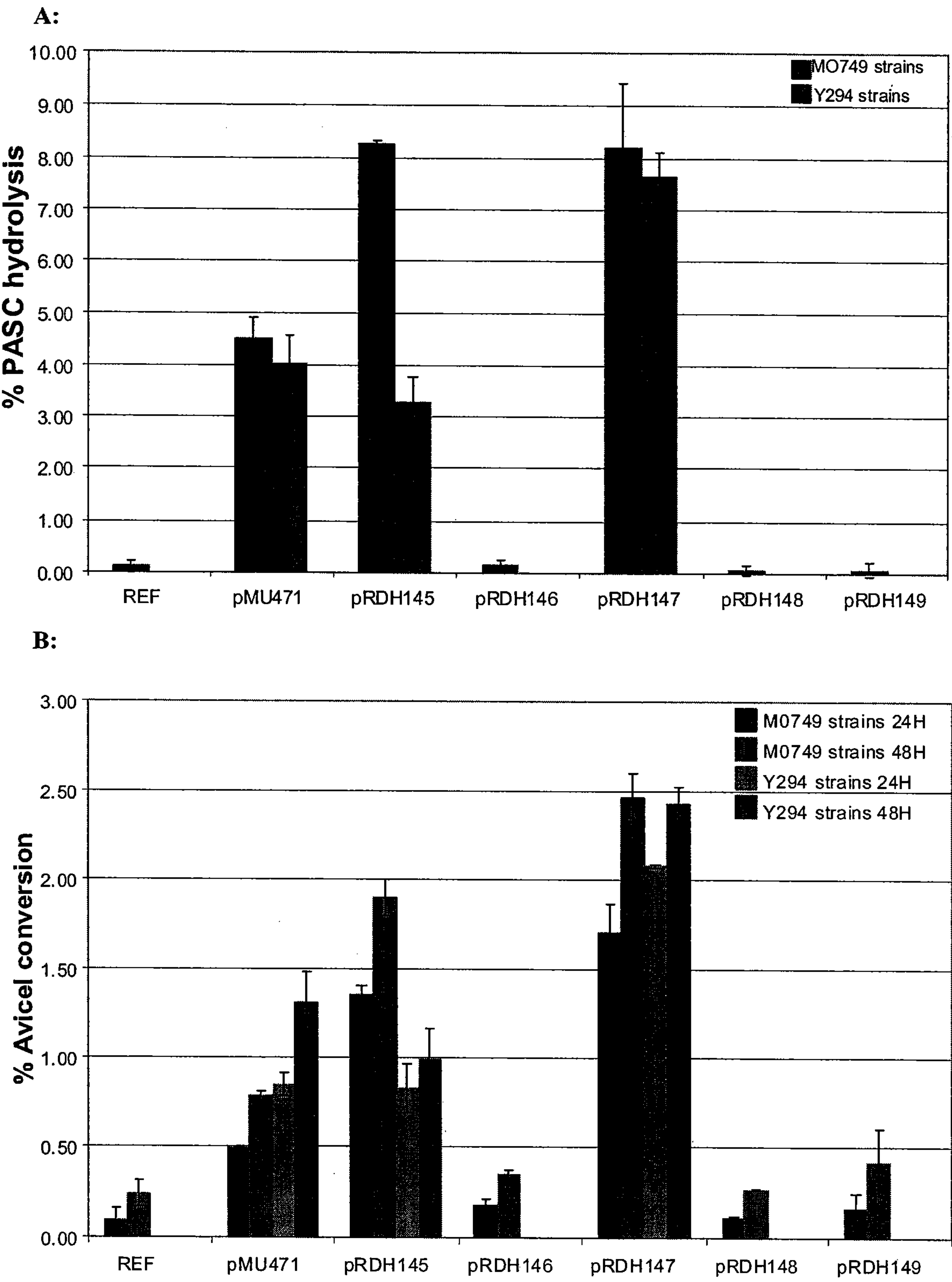
29/42

Figure 29



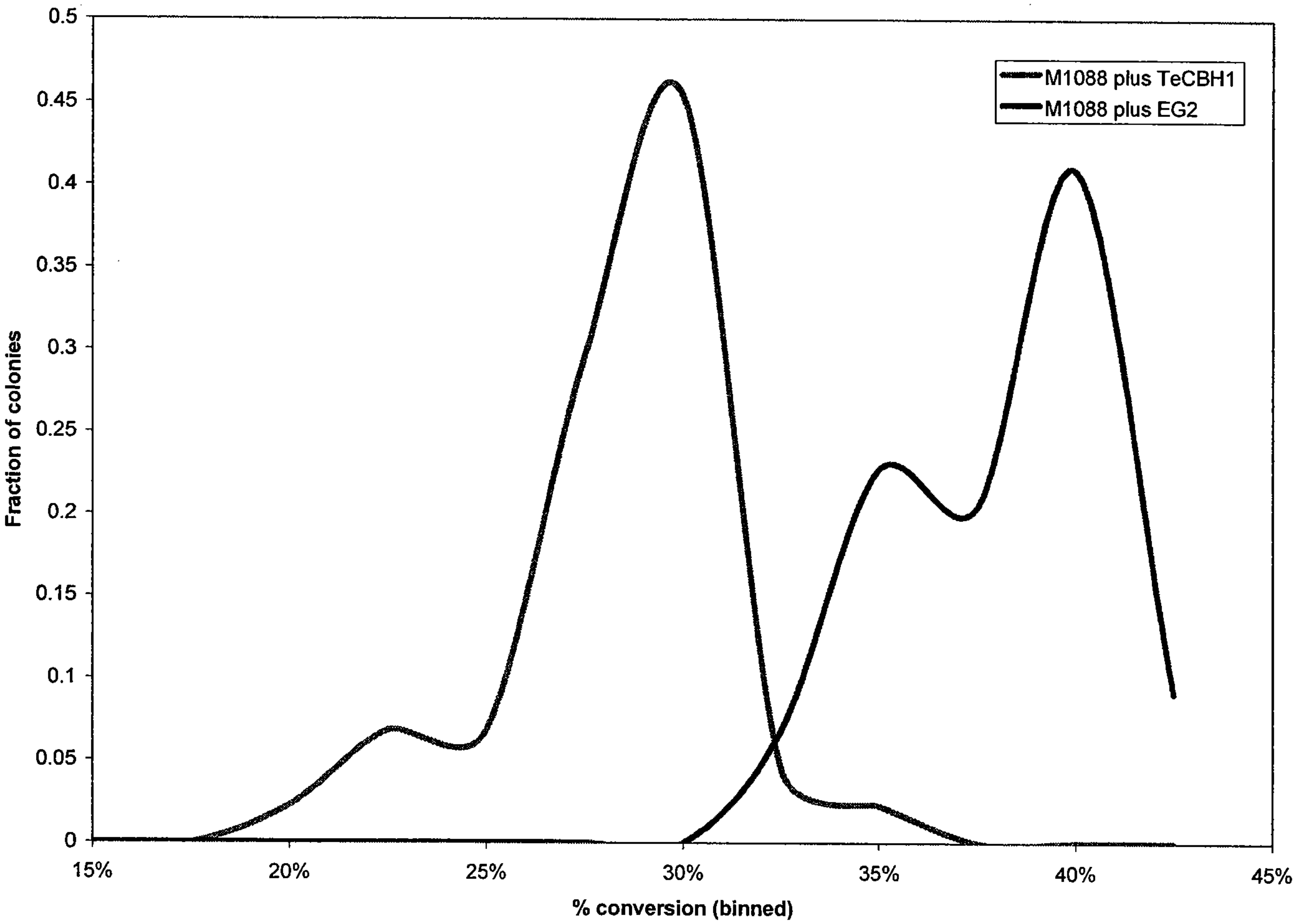
30/42

Figure 30



31/42

Figure 31



32/42

Figure 32

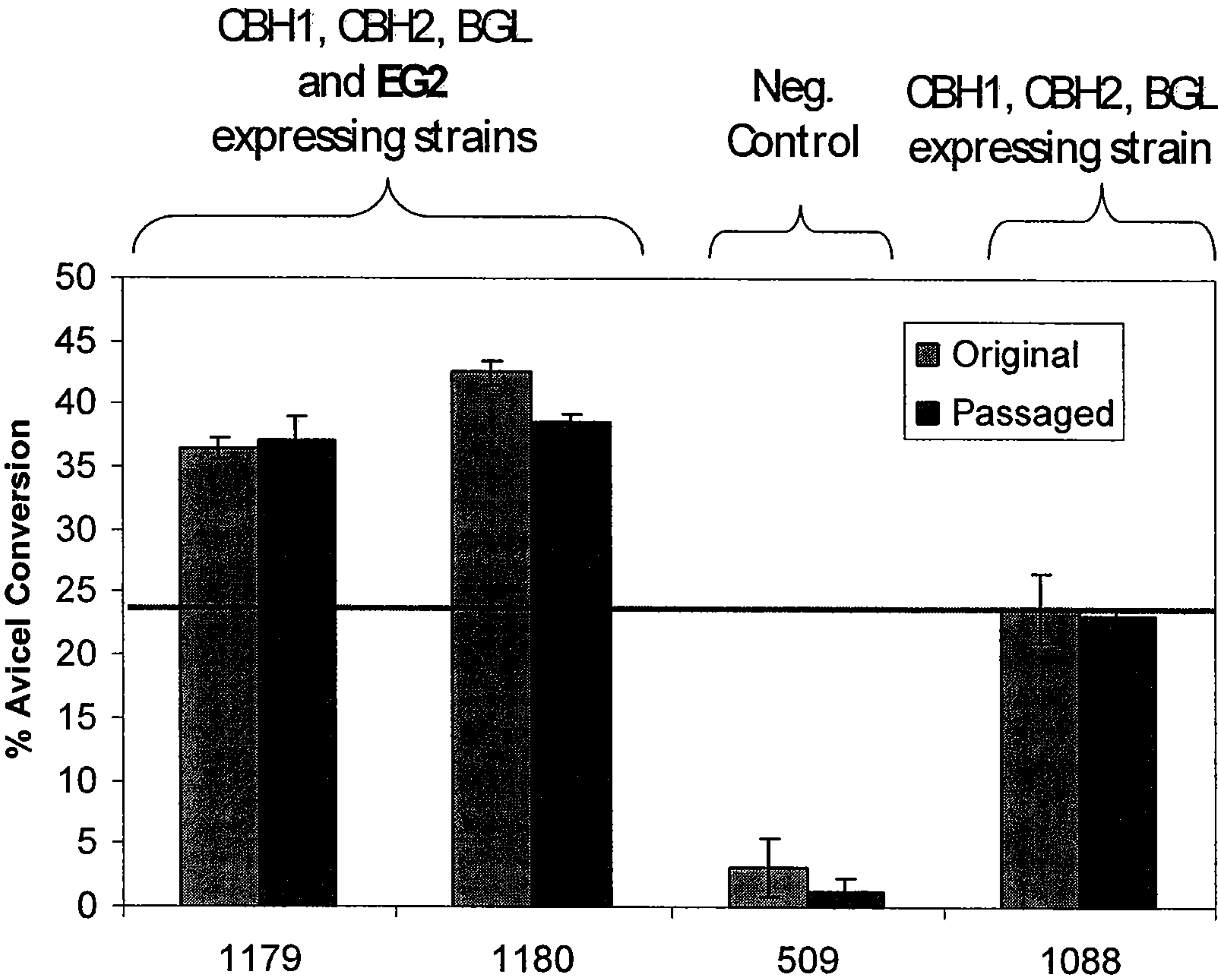


Figure 33

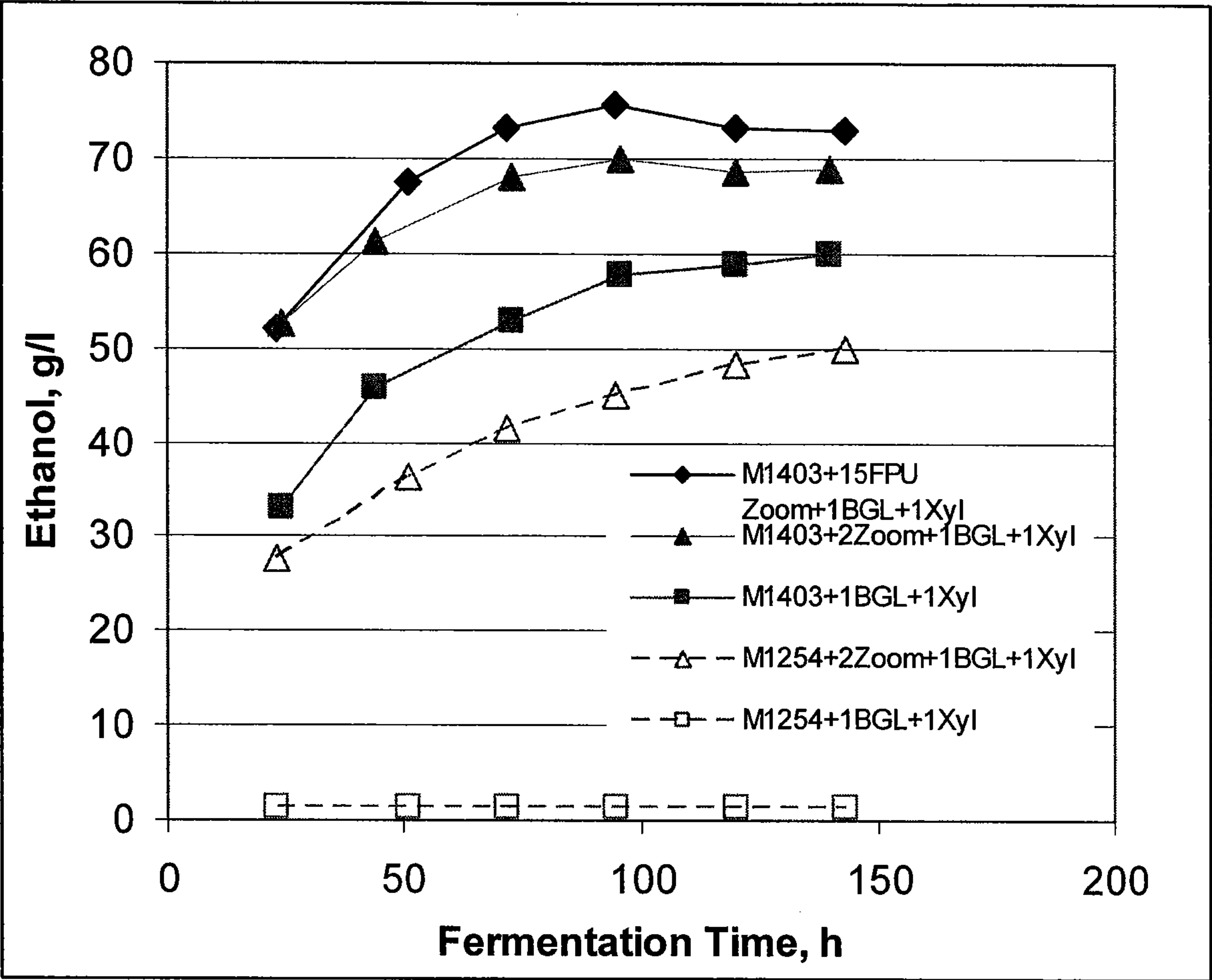


Figure 34

34/42

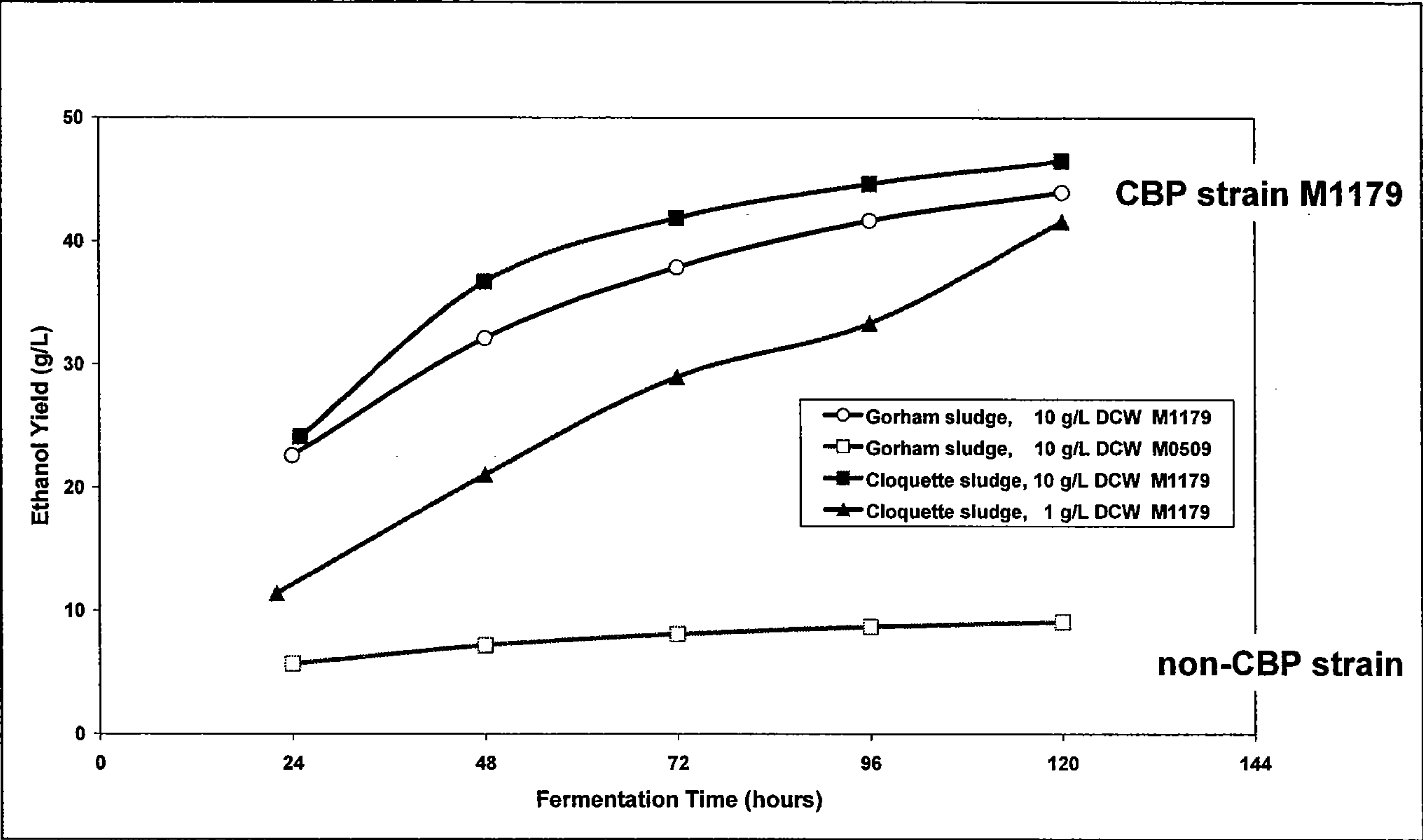


Figure 35

35/42

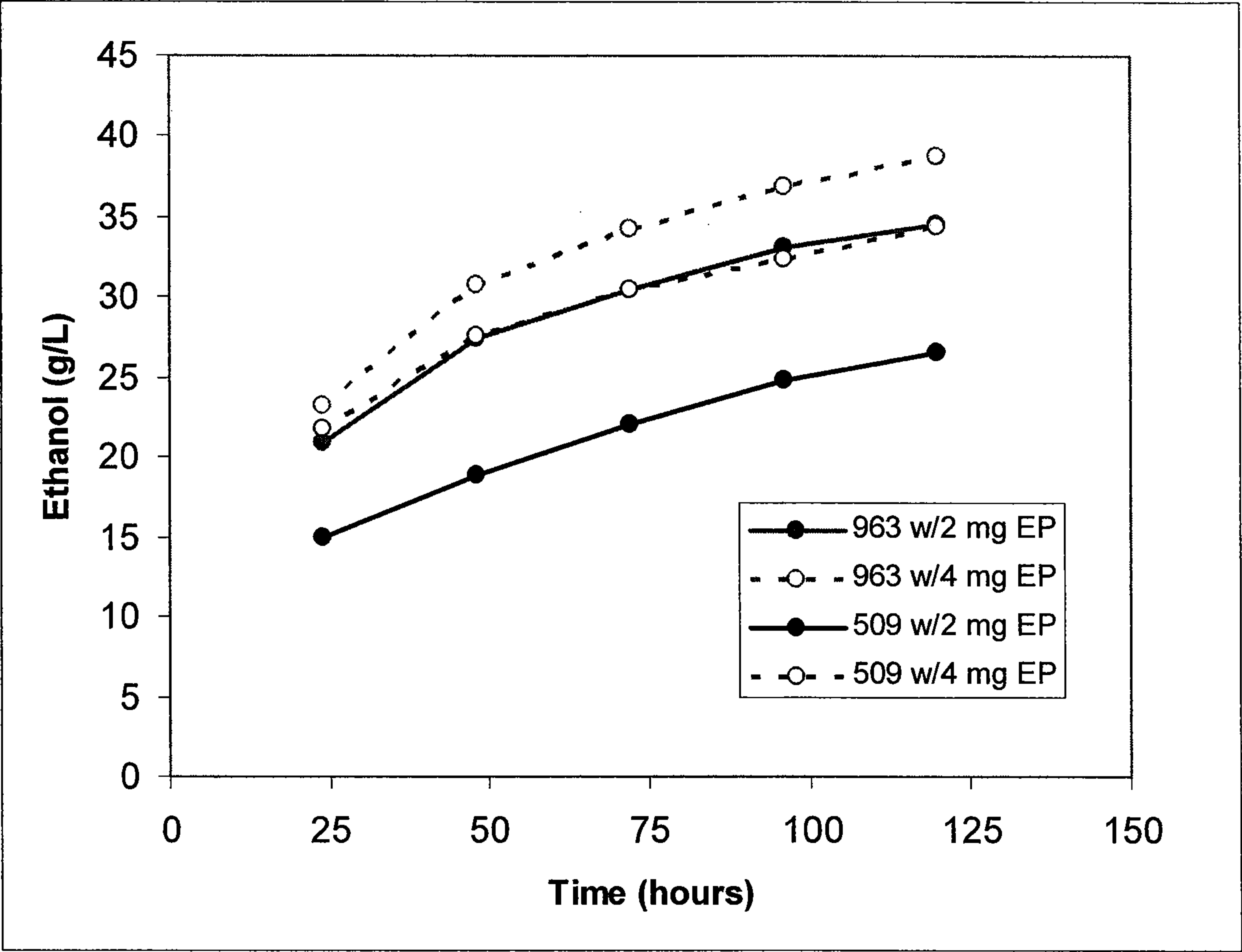


Figure 36

36/42

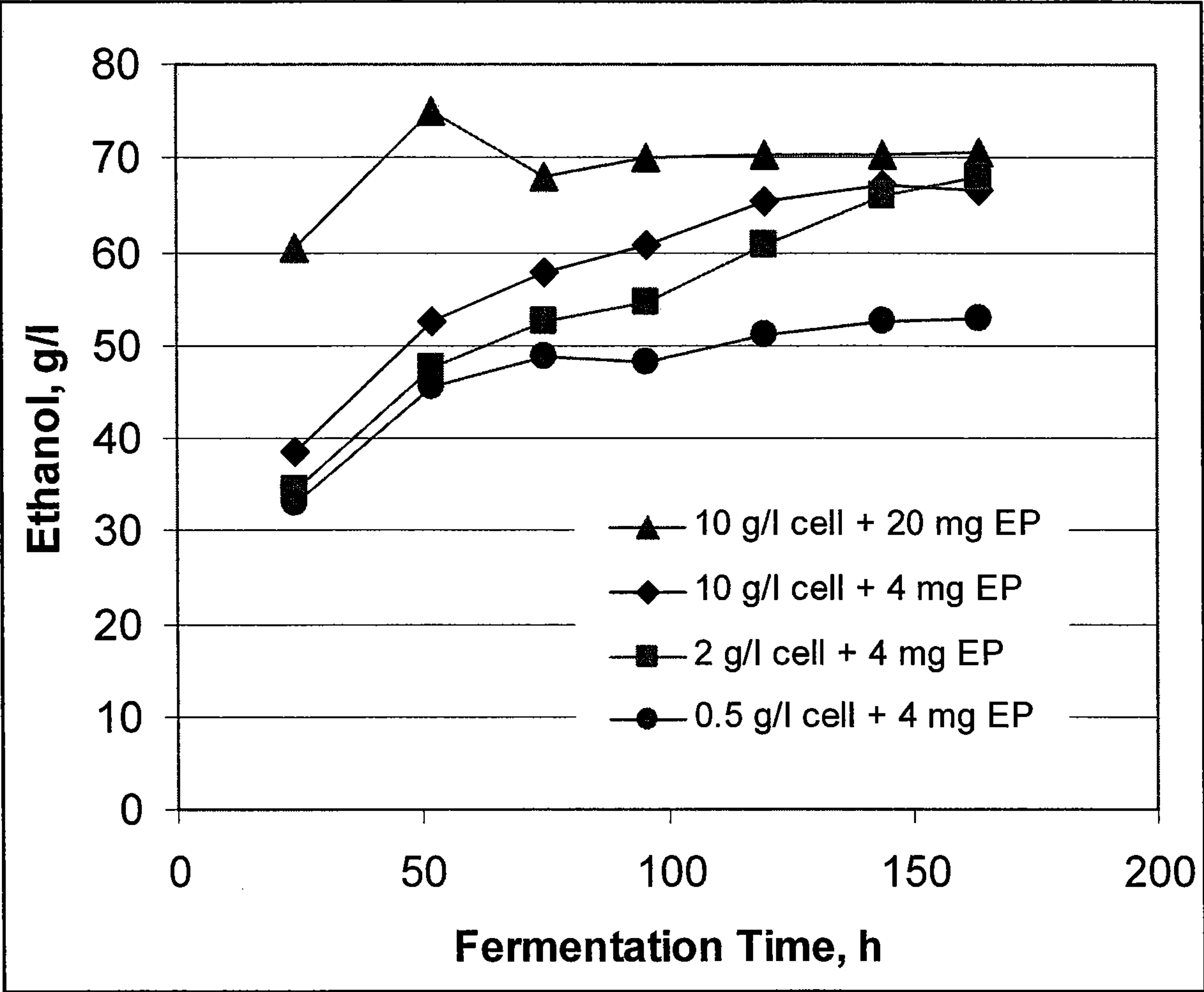


Figure 37

37/42

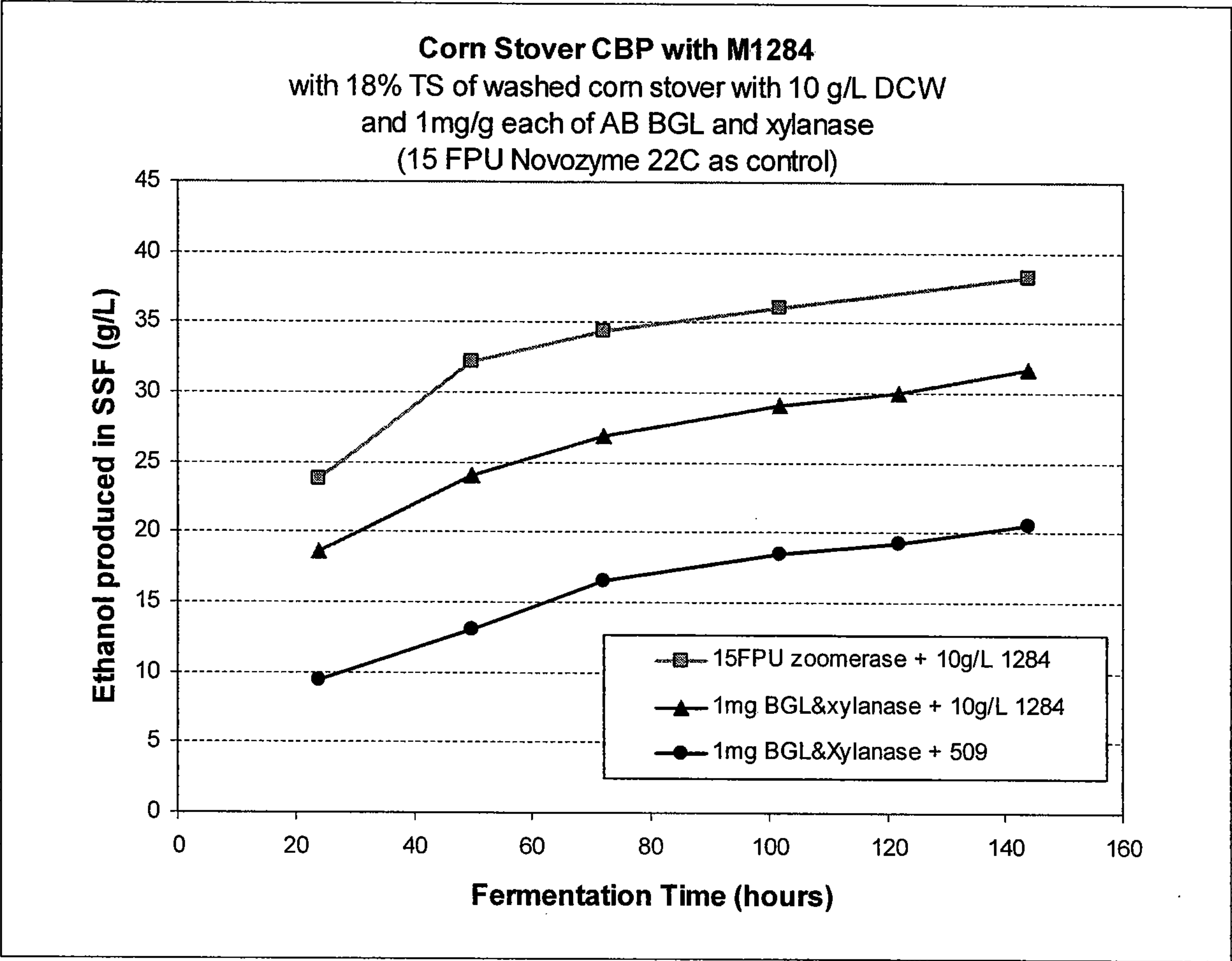


Figure 38A

38/42

Y294

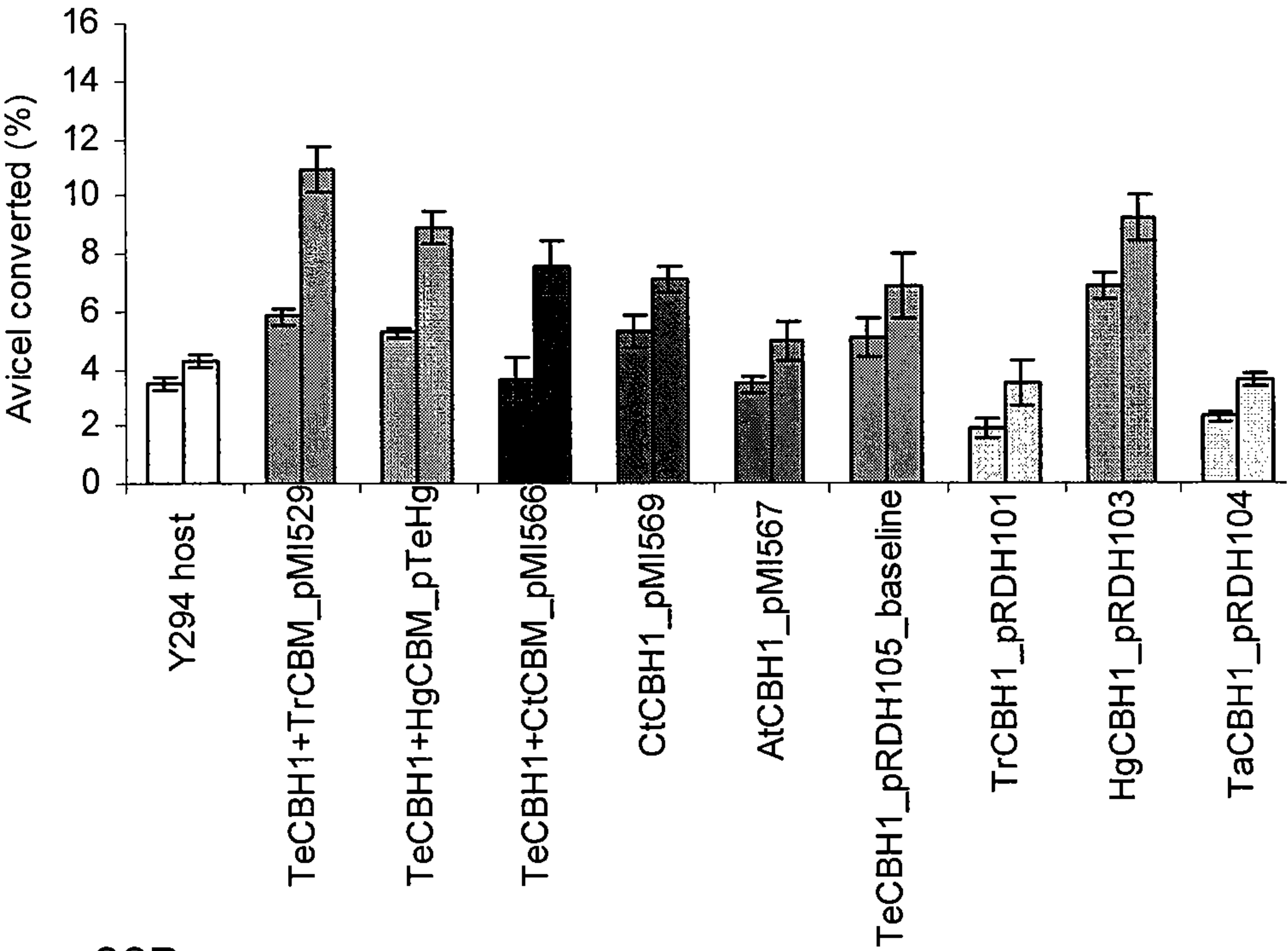


Figure 38B

M0749

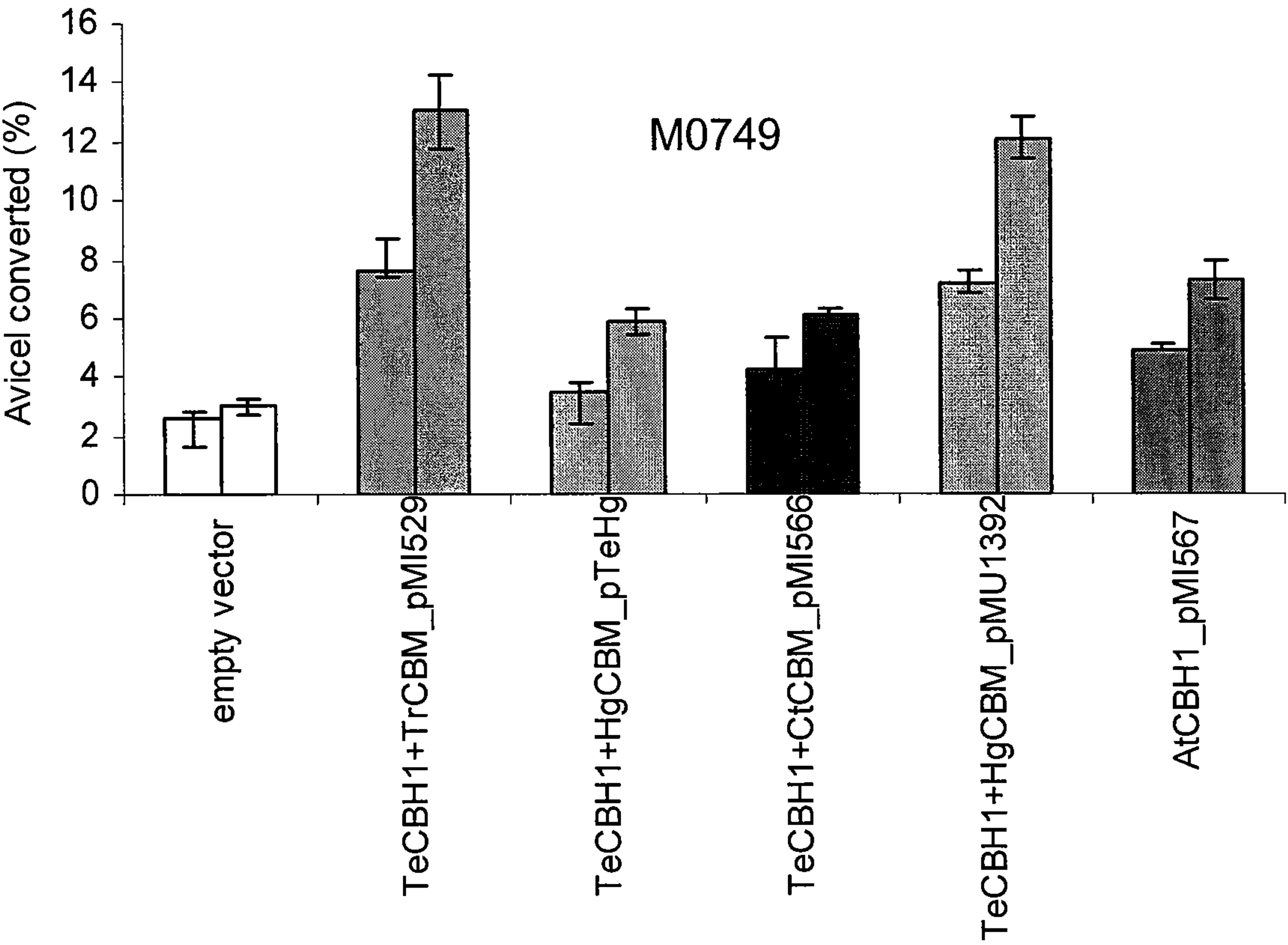


Figure 38C 39/42

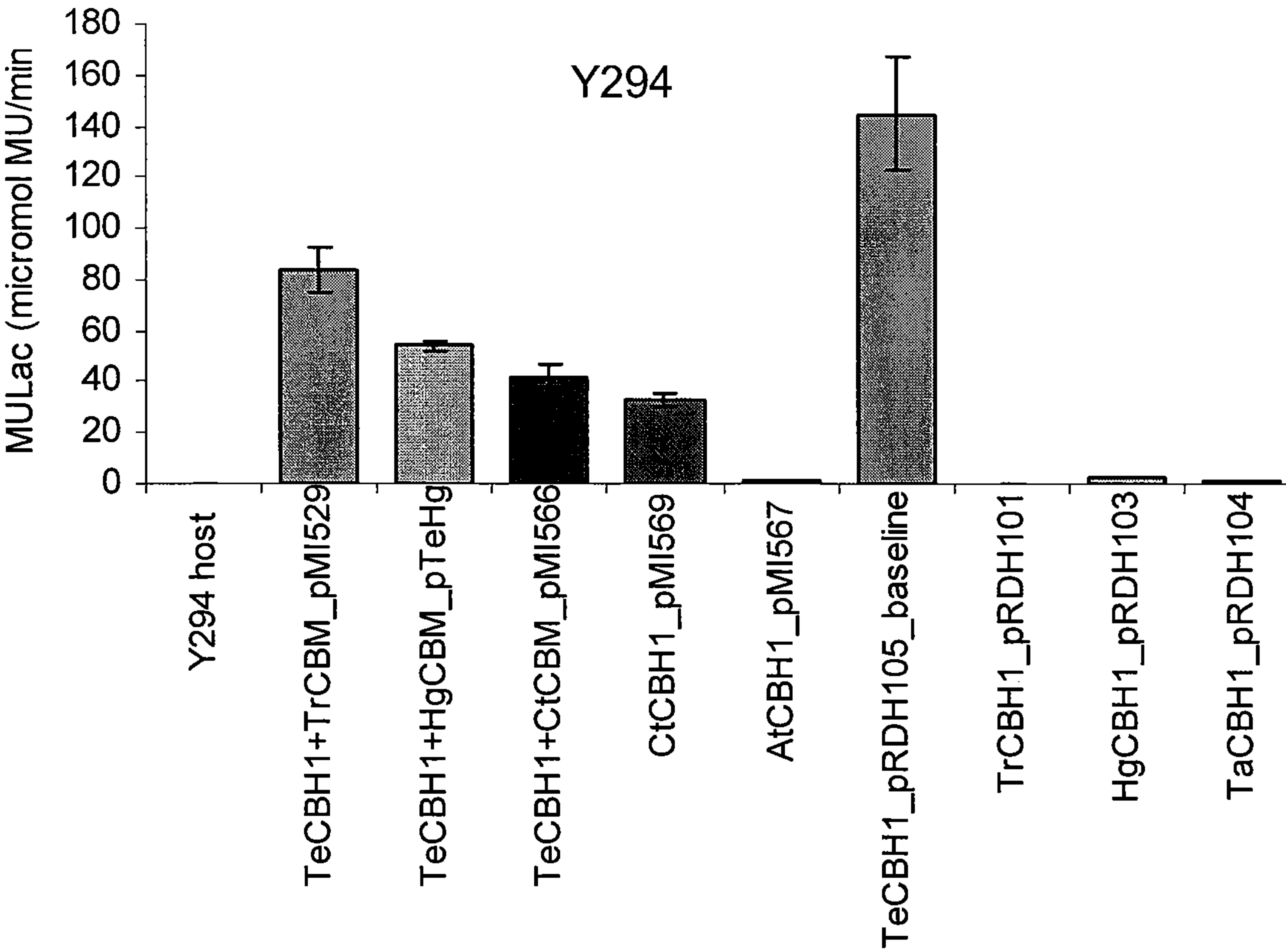


Figure 38D

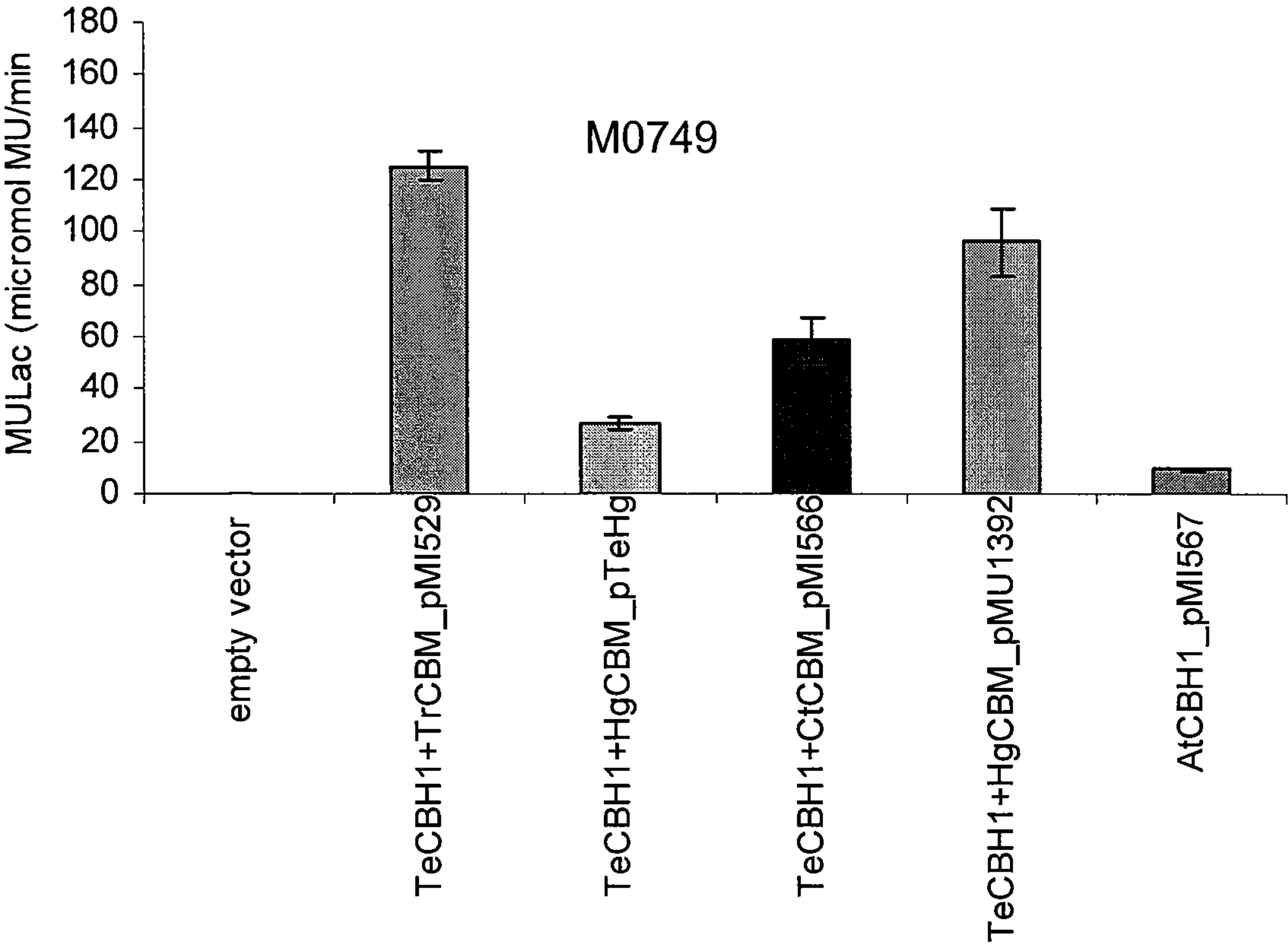


Figure 38E 40/42

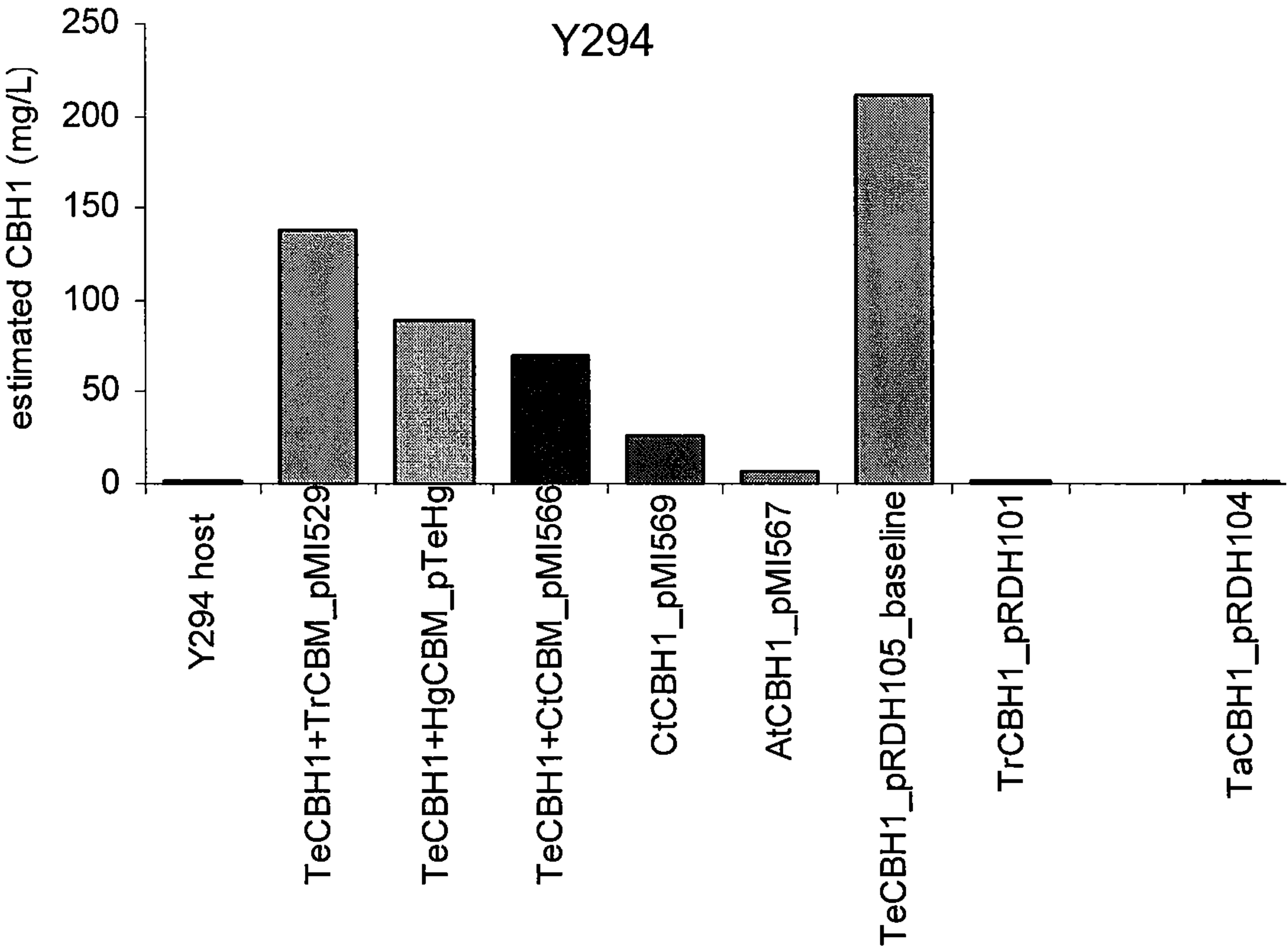


Figure 38F

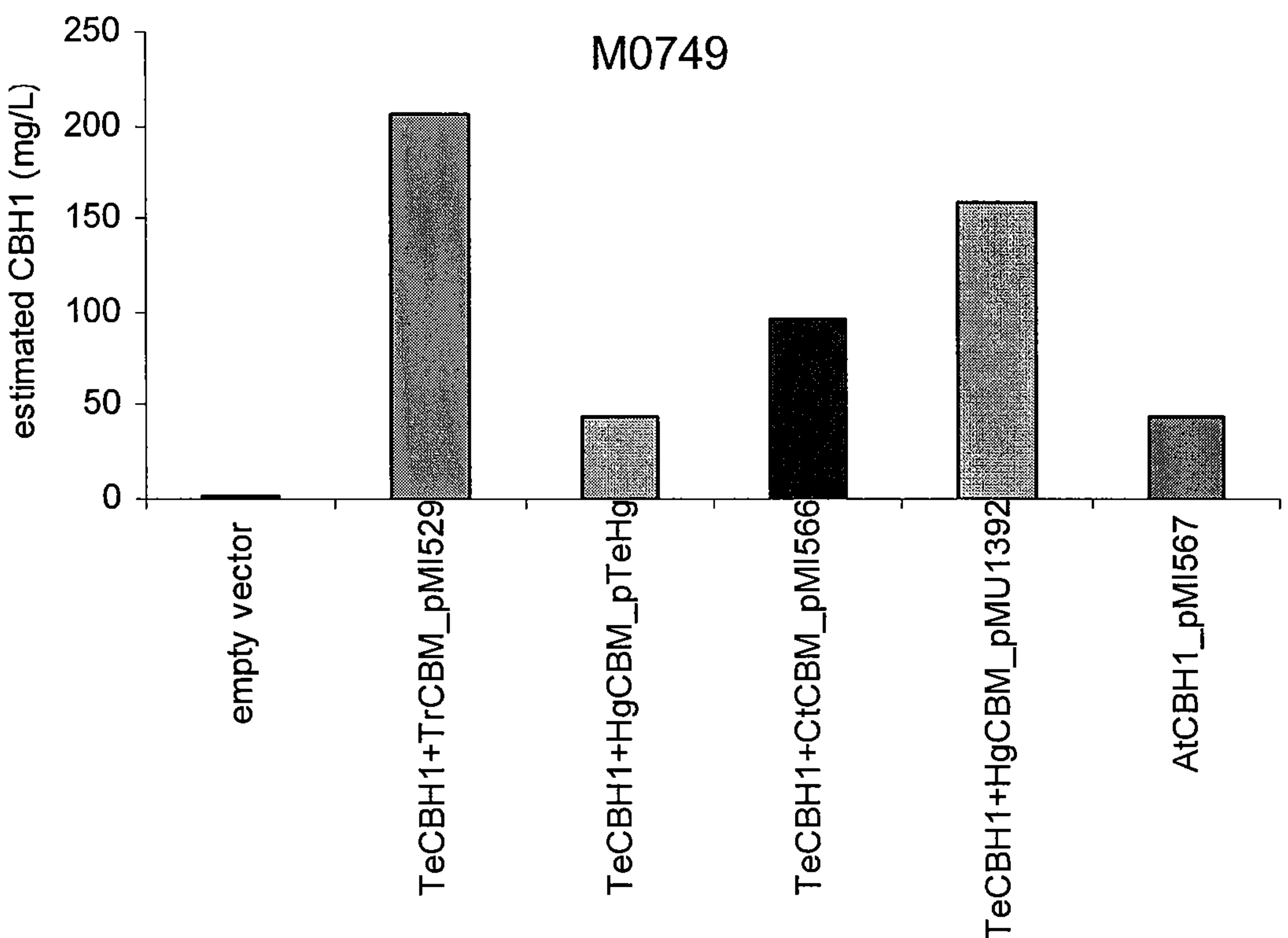


Figure 39
•Genes Modified in M0509.

Gene	Enzyme	EC number	Organism	Modification	Promoter	Terminator
XylA	xylose isomerase	5.3.1.5	Piromyces sp. E2	multi-copy integration	TPI	CYC1
XKS1	xylulokinase	2.7.1.17	S. cerevisiae	multi-copy integration	ADH1	CYC1
TAL1	transaldolase	2.2.1.2	S. cerevisiae	increased expression	TPI	-----
TKL1	transketolase	2.2.1.1	S. cerevisiae	increased expression	TPI	-----
RPE1	ribulose-5-phosphate epimerase	5.1.3.1	S. cerevisiae	increased expression	TPI	-----
RKI1	ribulose-5-phosphate isomerase	5.3.1.6	S. cerevisiae	increased expression	TPI	-----
GRE3	aldose reductase	1.1.1.21	S. cerevisiae	deleted	-----	-----

ADH1 = S. cerevisiae alcohol dehydrogenase1, TPI = S. cerevisiae triose phosphate isomerase, CYC1 = S. cerevisiae cytochrome c, isoform 1

Figure 40
•Strains used to construct M0509 and relevant genetic modifications

Strain	Genetic Loci		RKI1	RPE1	TAL1	TKL1	XylA & XKS1
M0086	GRE3	+/+	+/+	+/+	+/+	+/+	
M0150	GRE3+/gre3::loxP-kan						
M0207	gre3::loxP-dsdA/gre3::loxP-kan						
M0222	gre3::loxP/gre3::loxP						
M0227	gre3::loxP/gre3::loxP	RKI1+/loxP-kan-P _{TPI} -RKI1					
M0237	gre3::loxP/gre3::loxP	RKI1+/loxP-kan-P _{TPI} -RKI1	RPE1+/loxP-nat-P _{TPI} -RPE1				
M0252	gre3::loxP/gre3::loxP	RKI1+/loxP-P _{TPI} -RKI1	RPE1+/loxP-P _{TPI} -RPE1				
M0267	gre3::loxP/gre3::loxP	RKI1+/loxP-P _{TPI} -RKI1	RPE1+/loxP-P _{TPI} -RPE1	TAL1+/loxP-kan-P _{TPI} -TAL1			
M0277	gre3::loxP/gre3::loxP	RKI1+/loxP-P _{TPI} -RKI1	RPE1+/loxP-P _{TPI} -RPE1	TAL1+/loxP-kan-P _{TPI} -TAL1	TKL+/loxP-nat-P _{TPI} -TKL		
M0280	gre3::loxP/gre3::loxP	RKI1+/loxP-P _{TPI} -RKI1	RPE1+/loxP-P _{TPI} -RPE1	TAL1+/loxP-P _{TPI} -TAL1	TKL+/loxP-P _{TPI} -TKL		
M0407	gre3::loxP/gre3::loxP	RKI1+/loxP-P _{TPI} -RKI1	RPE1+/loxP-P _{TPI} -RPE1	TAL1+/loxP-P _{TPI} -TAL1	TKL+/loxP-P _{TPI} -TKL	delta::P _{ADH1} -XKS P _{TPI} -XylA	~10 copies
M0509	gre3::loxP/gre3::loxP	RKI1+/loxP-P _{TPI} -RKI1	RPE1+/loxP-P _{TPI} -RPE1	TAL1+/loxP-P _{TPI} -TAL1	TKL+/loxP-P _{TPI} -TKL	delta::P _{ADH1} -XKS P _{TPI} -XylA	~20 copies

Figure 41

Strain M1105 Genealogy

Strain	Selection	Parent	Genotype
M0509			<i>gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RK11+/loxP-PTPI-RK11 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xyIA PADH1-XKS</i>
M1017	YPX, 41°C	M0509	genotype not confirmed following cyostat evolution
M1046	YMX, 41°C	M1017	genotype not confirmed following cyostat evolution
M1080	YMX, 40°C	M1046	genotype not confirmed following cyostat evolution
M1105	YPD _X +acetate, 39°C	M1080	genotype not confirmed following cyostat evolution

Figure 42

Strain 1254 Genealogy

Strain	Selection	Parent	Genotype
M0509			<i>gre3::loxP/gre3::loxP TAL1+/loxP-PTPI-TAL1 RK11+/loxP-PTPI-RK11 RPE1+/loxP-PTPI-RPE1 TKL+/loxP-PTPI-TKL delta::PTPI-xyIA PADH1-XKS</i>
M1017	YPX, 41°C	M0509	genotype not confirmed following cyostat evolution
M1046	YMX, 41°C	M1017	genotype not confirmed following cyostat evolution
M1080	YMX, 40°C	M1046	genotype not confirmed following cyostat evolution
M1105	YPD _X +acetate, 39°C	M1080	genotype not confirmed following cyostat evolution
M1155	Xylose minimal + 8 g/L acetate, pH 5.5, 40°	M1105	genotype not confirmed following cyostat evolution
M1202	Complex glucose + 12 g/L acetate, pH 5.4, 41°C	M1155	genotype not confirmed following cyostat evolution
M1254	5% solids equivalent MS419 hydrolysate + yeast nitrogen base w/o amino acids, pH 4.8, 39°C	M1202+ M1155	genotype not confirmed following cyostat evolution