



(86) Date de dépôt PCT/PCT Filing Date: 2010/05/28
(87) Date publication PCT/PCT Publication Date: 2010/12/02
(85) Entrée phase nationale/National Entry: 2011/11/29
(86) N° demande PCT/PCT Application No.: CA 2010/000807
(87) N° publication PCT/PCT Publication No.: 2010/135836
(30) Priorité/Priority: 2009/05/29 (US61/182,275)

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

(71) Demandeur/Applicant:
IOGEN ENERGY CORPORATION, CA

(72) Inventeurs/Inventors:
LAVIGNE, JAMES, CA;
LIU, CHENGSONG, CA;
SCOTT, BRIAN R., CA;
TOMASHEK, JOHN J., CA

(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : NOUVELLES ENZYMES BETA-GLUCOSIDASES

(54) Title: NOVEL BETA-GLUCOSIDASE ENZYMES

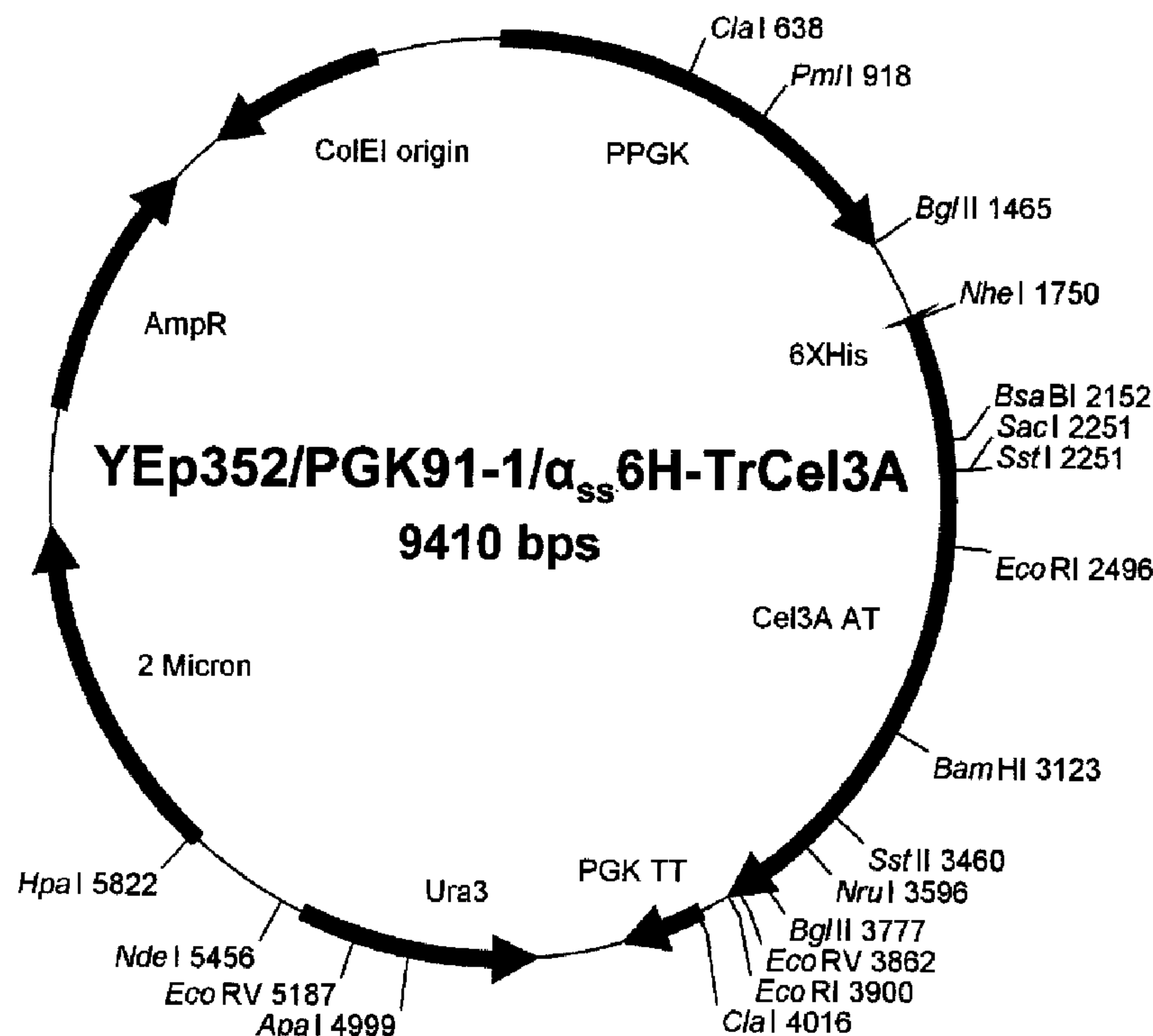


Figure 1

(57) Abrégé/Abstract:

Provided are modified beta-glucosidase enzymes, derived from the *Trichoderma reesei* CeBA beta-glucosidase, that exhibit improvements in one or more kinetic parameters (K_G , K_{G2} , k_{cat}) comprising amino acid substitutions at one or more of positions 43,



(57) **Abrégé(suite)/Abstract(continued):**

101, 260 and 543. Also provided are genetic constructs comprising nucleotide sequences encoding for modified beta-glucosidase enzymes, methods for the production of modified beta-glucosidase enzymes from host strains and the use of the modified beta-glucosidase enzymes in the hydrolysis of cellulose.

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

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

PCT



(10) International Publication Number

WO 2010/135836 A8

(51) International Patent Classification:

C12N 15/56 (2006.01) *C12N 9/42* (2006.01)
C12N 1/15 (2006.01) *C12P 19/14* (2006.01)
C12N 1/19 (2006.01)

(21) International Application Number:

PCT/CA2010/000807

(22) International Filing Date:

28 May 2010 (28.05.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/182,275 29 May 2009 (29.05.2009) US

(71) Applicant (for all designated States except US): IOGEN

ENERGY CORPORATION [CA/CA]; 310 Hunt Club
 Road East, Ottawa, Ontario K1V 1C1 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SCOTT, Brian R.

[CA/CA]; 6157 Malakoff Road, RR #3, Richmond, Ontario K0A 2Z0 (CA). LIU, Chengsong [CA/CA]; 1510 Rosebella Avenue, Ottawa, Ontario K1T 1E8 (CA). LAVIGNE, James [CA/CA]; 3554 Cambrian Road, Ottawa, Ontario K2J 0T6 (CA). TOMASHEK, John J. [CA/CA]; 157 Daly Avenue #1, Ottawa, Ontario K1N 6E8 (CA).

(74) Agents: SECHLEY, Konrad A. et al.; Gowling Lafleur

Henderson LLP, 550 Burrard Street, Suite 2300, Bentall 5, Vancouver, British Columbia, V6C 2B5 (CA).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,

[Continued on next page]

(54) Title: NOVEL BETA-GLUCOSIDASE ENZYMES

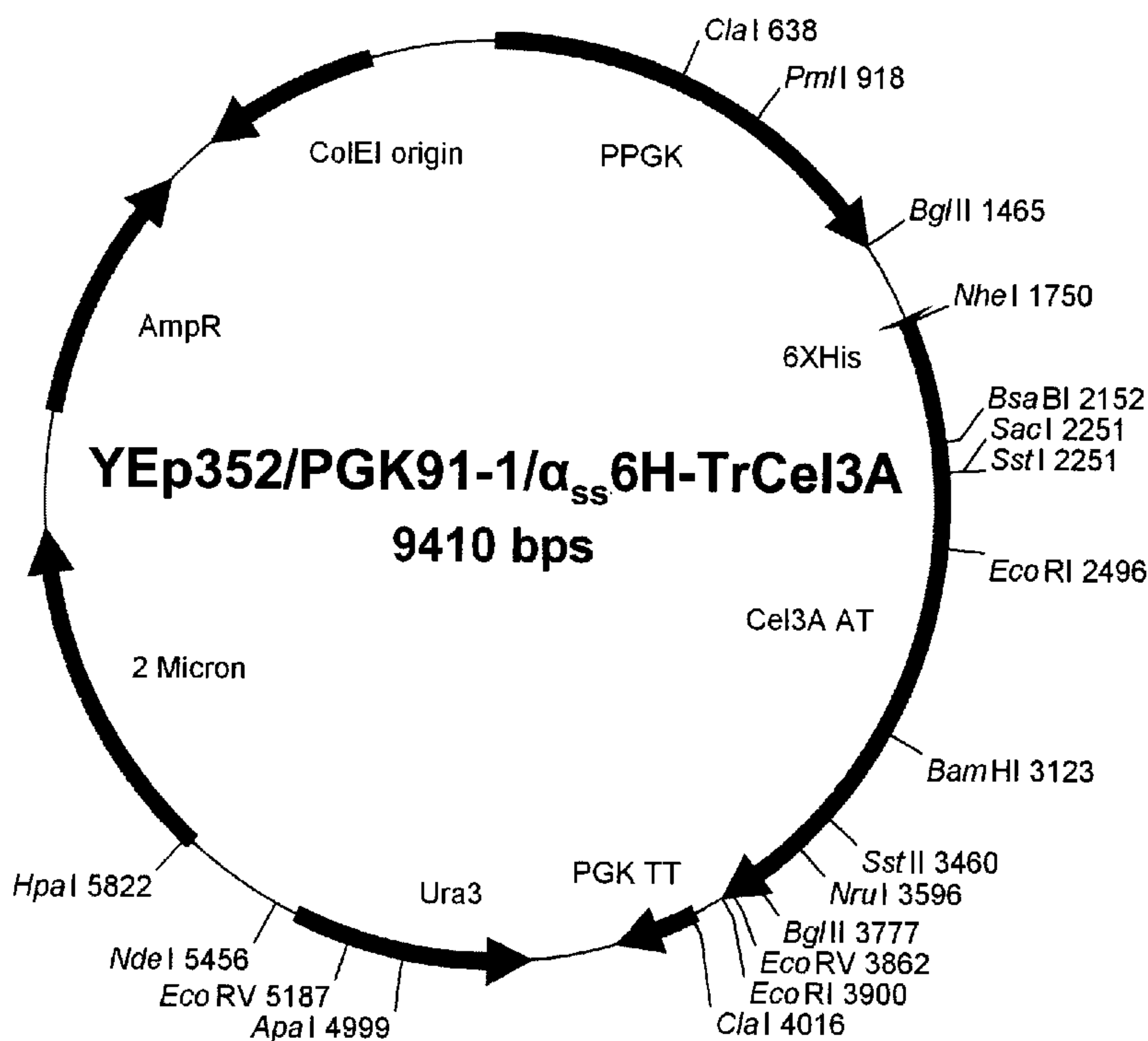


Figure 1

(57) Abstract: Provided are modified beta-glucosidase enzymes, derived from the *Trichoderma reesei* CeBA beta-glucosidase, that exhibit improvements in one or more kinetic parameters (K_G , K_{G2} , k_{cat}) comprising amino acid substitutions at one or more of positions 43, 101, 260 and 543. Also provided are genetic constructs comprising nucleotide sequences encoding for modified beta-glucosidase enzymes, methods for the production of modified beta-glucosidase enzymes from host strains and the use of the modified beta-glucosidase enzymes in the hydrolysis of cellulose.

WO 2010/135836 A8

NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

- *with international search report (Art. 21(3))*
- *with amended claims (Art. 19(1))*
- *with sequence listing part of description (Rule 5.2(a))*

(48) Date of publication of this corrected version:

21 July 2011

(15) Information about Correction:

see Notice of 21 July 2011

V83281WO**NOVEL BETA-GLUCOSIDASE ENZYMES****FIELD OF THE INVENTION**

[0001] The present invention relates to modified beta-glucosidases. More specifically, the invention relates to modified beta-glucosidases with improved kinetic parameters for the conversion of cellobiose to glucose. The present invention also relates to genetic constructs comprising nucleotide sequences encoding for modified beta-glucosidases, methods for the production of the modified beta-glucosidases from host strains and the use of the modified beta-glucosidase in industrial applications, including the hydrolysis of cellulose.

BACKGROUND OF THE INVENTION

[0002] Lignocellulosic feedstocks are a promising alternative to corn starch for the production of fuel ethanol. These raw materials are widely available, inexpensive and several studies have concluded that cellulosic ethanol generates close to zero greenhouse gas emissions.

[0003] However, these feedstocks are not easily broken down into their composite sugar molecules. Recalcitrance of lignocellulose can be partially overcome by physical and/or chemical pretreatment. An example of a chemical pretreatment is steam explosion in the presence of dilute sulfuric acid (U.S. Patent No. 4,461,648). This process removes most of the hemicellulose, but there is little conversion of the cellulose to glucose. The pretreated material may then be hydrolyzed by cellulase enzymes.

[0004] The term cellulase broadly refers to enzymes that catalyze the hydrolysis of the beta-1,4-glucosidic bonds joining individual glucose units in the cellulose polymer. The catalytic mechanism involves the synergistic actions of endoglucanases (E.C. 3.2.1.4), cellobiohydrolases (E.C. 3.2.1.91) and beta-glucosidases (E.C. 3.2.1.21) (Henrissat et al, 1994; Knowles et al., 1987; Lynd et al., 2002; Teeri, 1997; Wood and Garcia-Campayo, 1990; Zhang and Lynd, 2004). Endoglucanases hydrolyze accessible glucosidic bonds in the middle of the cellulose chain, while cellobiohydrolases processively release cellobiose from these chain ends.

Beta-glucosidases hydrolyze cellobiose to glucose thus minimizing product inhibition of the cellobiohydrolases and endoglucanases.

[0005] Beta-glucosidases are produced by many organisms occurring in all five living kingdoms. Generally these enzymes hydrolyze aryl-beta-glucosides, among which is included cellobiose (gluco-beta-(1,4)-glucoside). Some also catalyze transglycosylation reactions to varying extents.

[0006] Filamentous fungi, including *Trichoderma* ssp., *Aspergillus* ssp., *Hypocrea* ssp., *Humicola* ssp., *Neurospora* ssp., *Orpinomyces* ssp., *Gibberella* ssp., *Emmericella* ssp., *Chaetomium* ssp., *Fusarium* ssp., *Penicillium* ssp., *Magnaporthe* ssp., *Chrysosporium* ssp., *Myceliophthora* ssp., *Theilavia* ssp., *Sporotrichum* ssp. and *Phanerochaete* ssp. are effective producers of cellulase enzymes. Many of these organisms secrete beta-glucosidase enzymes. *Trichoderma* spp. (*Trichoderma longibrachiatum* or *Trichoderma reesei*) secrete small amounts of beta-glucosidase I or Cel3A (Chirico et al., 1987) and likely also secrete two other beta-glucosidases, Cel3B and Cel3E (Foreman et al., 2003).

[0007] The enzymatic hydrolysis of pretreated lignocellulosic feedstocks is an inefficient step in the production of cellulosic ethanol and its cost constitutes one of the major barriers to commercial viability. Improving enzymatic activity has been widely regarded as an opportunity for significant cost savings.

[0008] Cellobiohydrolases are strongly inhibited by cellobiose and to a lesser degree by glucose. Conversion of cellobiose to glucose is a rate-limiting step in cellulose hydrolysis because filamentous fungi, such as *Trichoderma reesei*, produce very low levels of beta-glucosidase and beta-glucosidases are highly sensitive to glucose inhibition (Chirico et al., 1987; Berghem et al., 1974). One technique for reducing cellulase inhibition is to increase the amount of beta-glucosidase in the system (U.S. Patent No. 6,015,703), as cellobiose is more inhibitory to cellulases than glucose (Holtzapple et al., 1990; Teleman et al., 1995). However, over-expressing a beta-glucosidase in an organism such as *Trichoderma* may reduce the production of other cellulase enzymes and, in turn, may limit the rate of cellulose conversion to cellobiose. In addition, this approach does not specifically address the effect of glucose inhibition on beta-glucosidase activity. A complementary approach would be to use a beta-glucosidase with a

higher specific activity which is also less sensitive to glucose inhibition. This enzyme would mitigate cellobiose product inhibition, but do so with lower amounts of beta-glucosidase (relative to the amount of cellulase(s)) and maintain its catalytic efficiency in the presence of high glucose concentrations.

[0009] Beta-glucosidases from most fungi have binding constants for cellobiose (K_{G2}) that range from 0.2-2.0 mM (Chirico et al., 1987; Berghem et al., 1974; Enari et al., 1981; Christakopoulos et al., 1994). These enzymes are highly sensitive to glucose inhibition; K_G values for glucose ranging from 0.6-8.0 mM have been reported for these enzymes. Several microbial beta-glucosidases with higher tolerance to glucose inhibition ($K_G > 8.0$ mM) have been reported (Riou et al., 1998; U.S. Patent 6,087,131; Saha et al., 1996; U.S. Patent 5,747,320; Gueguen et al., 1995; Li et al., 1991; Perez-Pons et al., 1994; Chen et al., 1994; US Patent 6,184,018 B1). However, these enzymes generally have a lower affinity for cellobiose (i.e., higher K_{G2} values). As a result, the concentration of cellobiose at steady state would be higher using these beta-glucosidases, increasing the degree of cellobiose inhibition on cellulase activity. Therefore, these particular glucose tolerant beta-glucosidase enzymes have limited utility for the production of cellulosic ethanol.

[0010] In spite of much research effort, there remains a need for improved beta-glucosidase enzymes in order to generate enzyme mixtures with higher sustained hydrolysis activity on pretreated lignocellulosic feedstock. The absence of such improved beta-glucosidase enzymes represents a large hurdle in the commercialization of cellulose conversion to glucose and other soluble fermentable sugars for the production of ethanol and other products.

SUMMARY OF THE INVENTION

[0011] The present invention relates to modified beta-glucosidases. More specifically, the invention relates to modified beta-glucosidases with improved kinetic parameters for the conversion of cellobiose to glucose. Beta-glucosidases of the present invention find utility in industrial processes requiring efficient conversion of cellobiose to glucose in the presence of glucose concentrations that would otherwise inhibit a parental beta-glucosidase.

[0012] An embodiment of the invention relates to a modified beta-glucosidase of *Trichoderma reesei* produced by substitution of the amino acid at one or more of positions 43, 101, 260 and 543 in the beta-glucosidase I or TrCel3A sequence (SEQ ID NO: 1) and comprising an amino acid sequence that is from about 80% to 99.9% to that TrCel3A amino acid sequence defined by SEQ ID NO: 1.

[0013] The modified TrCel3A beta-glucosidase may be derived from a parental TrCel3A beta-glucosidase that is otherwise identical to the modified TrCel3A beta-glucosidase and includes the substitution of the naturally occurring amino acid at one or more of positions 43, 101, 260 and 543. For example, the modified TrCel3A beta-glucosidase may contain one or more amino acid substitutions at positions other than at positions 43, 101, 260 and 543, provided that the amino acid sequence of the modified TrCel3A is from about 80% to about 99.9% identical to the TrCel3A amino acid sequence (SEQ ID NO: 1). For example, this invention includes the modified TrCel3A as defined above and further comprising an amino acid substitution at one or more of positions 66, 72, 96, 235, 248 and 369.

[0014] The present invention also relates to a modified TrCel3A beta-glucosidase comprising an amino acid sequence that is from about 80% to about 99.9% identical to that of the wild-type TrCel3A of SEQ ID NO: 1 and which exhibits (a) at least a 20% increase in the K_G , (b) at least a 20% decrease in K_{G2} , or (c) at least a 10% increase in k_{cat} for cellobiose relative to the K_G , K_{G2} and/or k_{cat} of a parental TrCel3A beta-glucosidase from which is derived.

[0015] The present invention also relates to a modified TrCel3A consisting of the amino acid sequence defined by:

SEQ ID NO: 2 (TrCel3A-V43I);

SEQ ID NO: 3 (TrCel3A-V43C);

SEQ ID NO: 4 (TrCel3A-V101A);

SEQ ID NO: 5 (TrCel3A-V101G);

SEQ ID NO: 6 (TrCel3A-F260I);

SEQ ID NO: 7 (TrCel3A-F260V);

SEQ ID NO: 8 (TrCel3A-F260Q);

SEQ ID NO: 9 (TrCel3A-F260D);

SEQ ID NO: 10 (TrCel3A-I543N);
SEQ ID NO: 11 (TrCel3A-I543W);
SEQ ID NO: 12 (TrCel3A-I543A);
SEQ ID NO: 13 (TrCel3A-I543S);
SEQ ID NO: 14 (TrCel3A-I543G);
SEQ ID NO: 15 (TrCel3A-I543L);
SEQ ID NO: 16 (TrCel3A-S72N-V101M-F260I);
SEQ ID NO: 17 (TrCel3A-V43I-S72N-V101M);
SEQ ID NO: 18 (TrCel3A-S72N-V101M-I543N);
SEQ ID NO: 19 (TrCel3A-S72N-V101M-I543D);
SEQ ID NO: 20 (TrCel3A-S72N-V101M-I543L);
SEQ ID NO: 21 (TrCel3A-V43I-S72N-V101M-F260I);
SEQ ID NO: 22 (TrCel3A-V43I-S72N-V101M-F260I-I543N);
SEQ ID NO: 23 (TrCel3A-V43I-S72E-V101M-I543N);
SEQ ID NO: 69 (TrCel3A-S72N-V101M-F260I-I543N); or
SEQ ID NO: 70 (TrCel3A-S72N-V101M-F260I-I543L).

[0016] The genetic constructs of the present invention comprise a nucleic acid sequence encoding a modified TrCel3A with an amino acid sequence that is from about 80% to about 99.9% amino acid sequence identity to SEQ ID NO: 1 and that comprises an amino acid substitution at one or more of positions 43, 101, 260 and 543, which nucleic acid sequence is operably linked to nucleic acid sequences regulating its expression and secretion from a host microbe. For example, the nucleic acid sequences regulating the expression and secretion of the modified TrCel3A beta-glucosidase may be derived from the host microbe used for expression of the modified TrCel3A beta-glucosidase. The host microbe may be a yeast, such as *Saccharomyces cerevisiae*, or a filamentous fungus, such as *Trichoderma reesei*.

[0017] The invention also relates to a genetic construct as defined above, wherein the modified TrCel3A beta-glucosidase comprises an amino acid sequence that is from about 90% to about 99.9% identical to SEQ ID NO: 1. The modified TrCel3A beta-glucosidase may further comprise substitutions at one or more of positions 66, 72, 96, 235, 248, 369 and 386 or any other additional mutations.

[0018] The invention also relates to a genetically modified microbe comprising a genetic construct encoding the modified TrCel3A beta-glucosidase.. For example, the genetically modified microbe may be capable of expression and secretion of the modified TrCel3A beta-glucosidase further comprising substitution at one or more of positions 66, 72, 96, 235, 248, 369 and 386 or any other additional mutations. The genetically modified microbe may be a yeast or filamentous fungus. For example, the genetically modified microbe may be a species of *Saccharomyces*, *Pichia*, *Hansenula*, *Trichoderma*, *Hypocrea*, *Aspergillus*, *Fusarium*, *Humicola*, *Chrysosporium*, *Myceliophthora*, *Thielavia*, *Sporotrichum* or *Neurospora*.

[0019] The present invention also relates to the use of the modified TrCel3A beta-glucosidase in a hydrolysis reaction containing a cellulosic substrate and a cellulase mixture comprising the modified TrCel3A beta-glucosidase.

[0020] The invention also relates to a process of producing the modified TrCel3A beta-glucosidase as defined above, including providing a yeast or fungal host with a genetic construct comprising a nucleic acid sequence encoding the modified TrCel3A beta-glucosidase, selection of recombinant yeast or fungal strains expressing the modified TrCel3A beta-glucosidase, culturing the selected recombinant strains in submerged liquid fermentations under conditions that induce the expression of the modified TrCel3A beta-glucosidase and recovering the modified TrCel3A beta-glucosidase.

[0021] Such modified TrCel3A beta-glucosidases find use in a variety of applications in industrial processes requiring enzymes that can retain high activity in the presence of normally inhibitory concentrations of the glucose. For example, modified TrCel3A beta-glucosidases, as described herein, may be used for the purposes of saccharification of lignocellulosic feedstocks for the production of fermentable sugars or in the production of compounds such as those used in the medical and food industries.

[0022] In another embodiment, the invention relates to a modified Family 3 beta-glycosidase comprising one or more of the amino acid substitutions selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543A, I543S, I543G and I543L and having an amino acid sequence that is at least 80% identical to the amino acid sequence of a parental Family 3 beta-glycosidase from which it is derived. The positions of the

amino acid substitution(s) are determined from sequence alignment of the Family 3 beta-glycosidase with a *Trichoderma reesei* Cel3A amino acid sequence as defined in SEQ ID NO: 1. The modified Family 3 beta-glycosidase of the present invention exhibits (a) an increase in the K_p , (b) a decrease in K_s , or (c) an increase in k_{cat} relative to the K_p , K_s or k_{cat} of the parental Family 3 beta-glycosidase from which it is derived.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] **FIGURE 1** depicts plasmid vector YEp352/PGK91-1/ α_{ss} 6H-TrCel3A directing the expression and secretion of native and modified TrCel3A from recombinant *Saccharomyces cerevisiae*.

[0024] **FIGURE 2** shows a *Saccharomyces cerevisiae* growth plate containing synthetic complete medium plus 0.1% esculin hydrate and 0.03% FeCl₃. Colonies which turned black after incubation for 3-4 days at 30°C were identified as expressing active TrCel3A beta-glucosidase.

[0025] **FIGURE 3** is a scatter plot of beta-glucosidase activity in assay 1 (30 mM cellobiose) versus beta-glucosidase activity in assay 2 (5 mM cellobiose + 1.25 mM glucose). The data relate to the screening of one 96-well culture plate containing wild-type TrCel3A (wt), filtrates from empty vector transformants (EV) and modified TrCel3As (variants). The wild-type TrCel3A data were fit by linear regression in which the y-intercept was fixed to zero.

[0026] **FIGURE 4** shows an SDS-PAGE gel of the parental (wt) and modified TrCel3A beta-glucosidases expressed from yeast and purified as described in Example 7, as well as a cellulase enzyme mixture from *Trichoderma reesei* (cellulase) and the wild-type TrCel3A purified from *T. reesei* cellulase (Cel3A). After SDS-PAGE separation, the proteins were visualized by Coomassie Blue stain.

[0027] **FIGURE 5** is a diagram illustrating the reaction scheme of the *p*NPG competitive substrate/inhibitor kinetic assay.

[0028] **FIGURE 6** shows the competitive substrate/inhibitor assay results for parental TrCel3A (Wt).

[0029] **FIGURE 7** shows the competitive substrate/inhibitor assay results for the modified TrCel3A-V43I beta-glucosidase.

[0030] **FIGURE 8** shows the competitive substrate/inhibitor assay results for the modified TrCel3A-V101A beta-glucosidase.

[0031] **FIGURE 9** shows the competitive substrate/inhibitor assay results for the modified TrCel3A-F260I beta-glucosidase.

[0032] **FIGURE 10** shows the competitive substrate/inhibitor assay results for the modified TrCel3A-I543N beta-glucosidase.

[0033] **FIGURE 11** shows an alignment of the amino acid sequences of 45 fungal Family 3 beta-glucosidases, including the parental TrCel3A of SEQ ID NO: 1, a consensus Family 3 beta-glucosidase sequence, and the % sequence identity of each amino acid sequence to that of TrCel3A. The positions of V43, V101, F260, and I543 are indicated by asterisks (*); the positions of the catalytic amino acids D236 and E447 are indicated by arrows (↓). A graphical representation of the frequency of occurrence of the amino acid at each position of the consensus Family 3 beta-glucosidase sequence of Figure 11 among the 45 fungal Family 3 beta-glucosidases is provided below the aligned sequences.

[0034] **FIGURE 12** is a Michaelis-Menton plot comparing the rates of cellobiose hydrolysis by wild-type parental TrCel3A and modified TrCel3A-F260I at different substrate concentrations.

[0035] **FIGURE 13** depicts plasmid vector pc/xCel3A-S72N-V101M-F260I-pyr4-TV directing the expression and secretion of modified TrCel3A beta-glucosidase from genetically modified *Trichoderma reesei*.

[0036] **FIGURE 14** shows the relative abundance of modified TrCel3A beta-glucosidases in cellulase mixtures produced by genetically modified *Trichoderma reesei*.

[0037] **FIGURE 15** is a plot showing the conversion of cellulose in a pretreated lignocellulosic substrate by cellulase mixtures comprising parental (WT TrCel3A) or modified (Modified TrCel3A) beta-glucosidases produced by genetically modified *Trichoderma reesei*.

Leu Ser Tyr Thr Lys Phe
580

<210> 5

<211> 582

<212> PRT

<213> Trichoderma reesei

<400> 5

Val Val Pro Pro Ala Gly Thr Pro Trp Gly Thr Ala Tyr Asp Lys Ala
1 5 10 15

Lys Ala Ala Leu Ala Lys Leu Asn Leu Gln Asp Lys Val Gly Ile Val
20 25 30

Ser Gly Val Gly Trp Asn Gly Gly Pro Cys Val Gly Asn Thr Ser Pro
35 40 45

Ala Ser Lys Ile Ser Tyr Pro Ser Leu Cys Leu Gln Asp Gly Pro Leu
50 55 60

Gly Val Arg Tyr Ser Thr Gly Ser Thr Ala Phe Thr Pro Gly Val Gln
65 70 75 80

Ala Ala Ser Thr Trp Asp Val Asn Leu Ile Arg Glu Arg Gly Gln Phe
85 90 95

Ile Gly Glu Glu Gly Lys Ala Ser Gly Ile His Val Ile Leu Gly Pro
100 105 110

Val Ala Gly Pro Leu Gly Lys Thr Pro Gln Gly Gly Arg Asn Trp Glu
115 120 125

Gly Phe Gly Val Asp Pro Tyr Leu Thr Gly Ile Ala Met Gly Gln Thr
130 135 140

Ile Asn Gly Ile Gln Ser Val Gly Val Gln Ala Thr Ala Lys His Tyr
145 150 155 160

Ile Leu Asn Glu Gln Glu Leu Asn Arg Glu Thr Ile Ser Ser Asn Pro
165 170 175

Asp Asp Arg Thr Leu His Glu Leu Tyr Thr Trp Pro Phe Ala Asp Ala
180 185 190

Val Gln Ala Asn Val Ala Ser Val Met Cys Ser Tyr Asn Lys Val Asn
195 200 205

Thr Thr Trp Ala Cys Glu Asp Gln Tyr Thr Leu Gln Thr Val Leu Lys
210 215 220

Asp Gln Leu Gly Phe Pro Gly Tyr Val Met Thr Asp Trp Asn Ala Gln
225 230 235 240

His Thr Thr Val Gln Ser Ala Asn Ser Gly Leu Asp Met Ser Met Pro
245 250 255

Gly Thr Asp Phe Asn Gly Asn Asn Arg Leu Trp Gly Pro Ala Leu Thr
260 265 270

[0041] Beta-glucosidases are a subset of beta-glycosidases belonging to glycoside hydrolase (GH) Families 1 and 3, using the classification system developed by Henrissat and coworkers (Henrissat, B. (1991); Henrissat, B. and Bairoch, A. (1996)). There are currently over 115 GH families that have been identified using this classification system, which are listed in the database of Carbohydrate Active Enzymes (CAZy) (see <http://afmb.cnrs-mrs.fr/CAZY/index.html> for reference). Family 1 comprises beta-glycosidases from archaeobacteria, plants and animals. Beta-glycosidases from some bacteria, mold and yeast belong to Family 3. For the purpose of this invention, a “beta-glycosidase” is therefore defined as any protein that is categorized as a Family 3 glycoside hydrolase according to the CAZy system.

[0042] The three dimensional structure of beta-D-glucan exohydrolase, a Family 3 glycosyl hydrolase, was described by Varghese et al. (1999). The structure was of a two domain globular protein comprising a N-terminal (α/β)₈ TIM-barrel domain and a C-terminal a six-stranded beta-sandwich, which contains a beta-sheet of five parallel beta-strands and one antiparallel beta-strand, with three alpha-helices on either side of the sheet. This structure is likely shared by other Family 3 enzymes.

[0043] As shown in Figure 11, the primary amino acid sequence of Family 3 beta-glucosidases show a high degree of similarity. Multiple alignment across 45 Family 3 beta-glucosidase amino acid sequences shows that the most naturally occurring Family 3 beta-glucosidases of fungal origin show from about 40% to about 100% amino acid sequence identity to the amino acid sequence of TrCel3A (Figure 11). In particular, there are several regions of very high amino acid sequence conservation within the Family 3 beta-glucosidases including, for example, from amino acids 225-256 and 439-459, containing the catalytic amino acids D236 and E447, respectively.

[0044] By “TrCel3A beta-glucosidase” or “TrCel3A” it is meant the Family 3 glycosyl hydrolase produced by *Trichoderma reesei* defined by the amino acid sequence of SEQ ID NO: 1. TrCel3A beta-glucosidase is also known as *Trichoderma reesei* beta-glucosidase or BGL1. By “native” or “wild type” TrCel3A (also annotated as TrCel3A^{wt}), it is meant the TrCel3A of SEQ ID NO: 1 without any amino acid substitutions.

[0045] By “modified TrCel3A beta-glucosidase” or “modified TrCel3A”, it is meant a TrCel3A beta-glucosidase which comprises one or more of the amino acid substitutions, introduced by genetic engineering techniques, selected from the group consisting of V43X, V101X, F260X, and I543X. For example, the modified modified TrCel3A beta-glucosidase may comprising one or more amino acid substitutions selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543A, I543S, I543G and I543L.

[0046] Genetic engineering techniques for altering amino acid sequences include, but are not limited to, site-directed mutagenesis, cassette mutagenesis, random mutagenesis, synthetic oligonucleotide construction, cloning and other genetic engineering techniques as would be known by those of skill in the art (Eijsink VG, et al. 2005). Modified TrCel3A beta-glucosidases of the present invention include those comprising amino acid substitutions at any one of V43X, V101X, F260X and I543X, at any two of V43X, V101X, F260X and I543X, any three of V43X, V101X, F260X and I543X, or all four of V43X, V101X, F260X and I543X.

[0047] It will be understood that the modified TrCel3A beta-glucosidase may be derived from wild-type TrCel3A beta-glucosidase of SEQ ID NO: 1 or from a TrCel3A beta-glucosidase that contains other amino acid substitutions. For example, the modified TrCel3A beta-glucosidase may contain amino acid substitution at one or more of positions 66, 73, 96, 235, 248, and 369. Alternatively, after production of the modified TrCel3A beta-glucosidase comprising mutations at one or more of positions 43, 101, 260 and 543, it may be subsequently further modified to contain additional amino acid substitutions, including but not limited to those set forth above.

[0048] As used herein in respect of modified TrCel3A beta-glucosidase amino acid sequences, “derived from” refers to the isolation of a target nucleic acid sequence element encoding the desired modified TrCel3A beta-glucosidase using genetic material or nucleic acid or amino acid sequence information specific to the parental TrCel3A beta-glucosidase. As is known by one of skill in the art, such material or sequence information can be used to generate a nucleic acid sequence encoding the desired modified TrCel3A beta-glucosidase using one or more molecular biology techniques including, but not limited to, cloning, sub-cloning, amplification by PCR, *in vitro* synthesis, and the like.

[0049] In one embodiment of the invention, the amino acid sequence of the modified TrCel3A beta-glucosidase is from about 80% to about 99.9% identical to SEQ ID NO: 1. For example, the amino acid sequence of the modified TrCel3A beta-glucosidase may be from about 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 99.9% identical to SEQ ID NO: 1. In other words, the number of amino acid substitutions in the modified TrCel3A beta-glucosidase may not exceed 20% of the total number amino acids in the parental TrCel3A beta-glucosidase sequence.

[0050] In another embodiment of the invention, the amino acid sequence of modified TrCel3A beta-glucosidase may be from about 90% to about 99.9% identical to SEQ ID NO: 1. For example, the amino acid sequence of the modified TrCel3A beta-glucosidase may be from about 95% to about 100% identical to SEQ ID NO: 1.

[0051] In another embodiment, the amino acid sequence of the modified TrCel3A beta-glucosidase may be from about 80% to about 99.9% identical to SEQ ID NO: 1 and the modified TrCel3A beta-glucosidase may exhibit (a) at least about a 20% increase in K_G , (b) at least about a 20% decrease in K_{G2} , or (c) at least about a 10% increase in k_{cat} for cellobiose relative to the K_G , K_{G2} and/or k_{cat} of a parental TrCel3A beta-glucosidase from which it is derived. For example, the modified TrCel3A beta-glucosidase may exhibit (a) from about a 20% to about a 800% increase in K_G , or any increase therebetween, (b) from about a 20% to about an 80% decrease in K_{G2} , any decrease therebetween, or (c) from about a 10% to about a 50% increase in k_{cat} for cellobiose relative to the K_G , K_{G2} and/or k_{cat} of a parental TrCel3A beta-glucosidase from which is derived.

[0052] By “parental TrCel3A beta-glucosidase” or “parental TrCel3A”, it is meant a TrCel3A beta-glucosidase that does not contain a substitution of its original amino acid(s) at positions 43, 101, 260 or 543. For example, the parental TrCel3A beta-glucosidase may comprise amino acid substitutions at one or more of positions 66, 72, 96, 235, 248, and 369.

[0053] In order to assist one of skill in the art regarding those amino acid positions of the TrCel3A beta-glucosidase at which amino acid substitutions (other than V43X, V101X, F260X, and I543X) may be made and produce an active beta-glucosidase, an alignment of 45 Family 3 beta-glucosidases derived from fungal sources along with a consensus beta-glucosidase sequence

consisting of the amino acids that naturally occur with the highest frequency at each position is provided in Figure 11 along with a graph showing the frequency of occurrence of each amino acid of the consensus sequence at each position. Using the information provided in Figure 11, one of skill in the art would recognize regions of low sequence conservation to other Family 3 beta-glucosidases and choose such regions for introduction of amino acid substitutions that are not likely to compromise significantly the function of the enzyme. Non-limiting examples of such regions include, for example, the regions between positions 1-20, 303-323 and 403-414 and select amino acid positions within these regions.

[0054] As described in more detail herein, several modified TrCel3A beta-glucosidases have been prepared that exhibit (a) at least a 20% increase in K_G , (b) at least a 20% decrease in K_{G2} , or (c) at least a 10% increase in k_{cat} for cellobiose relative to the K_G , K_{G2} and/or k_{cat} of a parental TrCel3A beta-glucosidase from which is derived. A list of several modified TrCel3A beta-glucosidases, which is not to be considered limiting in any manner, is presented in Table 1.

Table 1: TrCel3A beta-glucosidases with improved catalytic efficiency

Modified TrCel3A beta-glucosidase	SEQ ID NO:
TrCel3A-V43I	2
TrCel3A-V43C	3
TrCel3A-V101A	4
TrCel3A-V101G	5
TrCel3A-F260I	6
TrCel3A-F260V	7
TrCel3A-F260Q	8
TrCel3A-F260D	9
TrCel3A-I543N	10
TrCel3A-I543W	11
TrCel3A-I543A	12
TrCel3A-I543S	13
TrCel3A-I543G	14
TrCel3A-I543L	15

TrCel3A-S72N-V101M-F260I	16
TrCel3A--V43I-S72N-V101M	17
TrCel3A-S72N-V101M-I543N	18
TrCel3A-S72N-V101M-I543D	19
TrCel3A- S72N-V101M-I543L	20
TrCel3A-V43I-S72N-V101M-F260I	21
TrCel3A-V43I-S72N-V101M-F260I- I543N	22
TrCel3A-V43I-S72N-V101M- I543N	23
TrCel3A-S72N-V101M-F260I-I543N	69
TrCel3A-S72N-V101M-F260I-I543L	70

Modified TrCel3A beta-glucosidases Improved Kinetic Parameters

[0055] The modified TrCel3A beta-glucosidases of the present invention exhibit improvements in at least one of the following kinetic parameters: K_G , K_{G2} and k_{cat} . K_G is defined as the concentration of glucose which reduces the enzymatic activity of the beta-glucosidase by 50%. K_{G2} is defined as the concentration of cellobiose at which the beta-glucosidase exhibits half its maximal rate. The k_{cat} is the catalytic rate constant for the hydrolysis of cellobiose. Example 8 details an assay for measuring the K_G and K_{G2} of parental and modified TrCel3A beta-glucosidases. Example 9 details an assay for measuring the k_{cat} of parental and modified TrCel3A beta-glucosidases.

[0056] K_G of the parental and modified TrCel3A beta-glucosidases can be determined by measuring the rate of hydrolysis of a chromogenic substrate, such as *p*-nitrophenyl-beta-D-glucopyranoside (pNPG), in the presence of various concentrations of glucose as described in Example 8. The K_G is the concentration of glucose that reduces the rate of *p*-nitrophenol (pNP) release from pNPG by 50% compared to the rate of pNPG hydrolysis in the absence of glucose. The K_{G2} constants for parental and modified TrCel3A beta-glucosidases can be determined by measuring the rate of hydrolysis of cellobiose in reactions containing increasing concentrations of cellobiose or, alternatively, by measuring the rate of hydrolysis of a chromogenic substrate, such as pNPG, in the presence of various concentrations of a cellobiose as described in Example 8. The K_{G2} is the concentration of cellobiose that reduces the rate of pNP release from pNPG by

50% compared to the rate of pNPG hydrolysis in the absence of cellobiose. The k_{cat} values for parental and modified TrCel3A beta-glucosidases can be determined by measuring the rate of cellobiose hydrolysis at varying concentrations of a cellobiose substrate, for example, as described in Example 9.

[0057] The effect of amino acid substitutions at positions 43, 101, 260 and 543 were determined by a comparative study of the modified and parental TrCel3A beta-glucosidases. The relative values of K_G , K_{G2} and k_{cat} for the parental and modified TrCel3A beta-glucosidases are shown in Table 2, below. Reaction curves for the hydrolysis of pNPG substrate alone and in the presence of glucose or cellobiose by parental and modified TrCel3A beta-glucosidases are shown in Figures 6 through 10. Reaction curves for the hydrolysis of cellobiose substrate by parental and modified TrCel3A beta-glucosidases are shown in Figure 12.

Table 2: Modified TrCel3A Beta-glucosidases with Improved Kinetic Parameters

Amino acid substitution	Relative K_G	Relative K_{G2}	Relative K_G/ K_{G2}	Relative k_{cat}
None (TrCel3A)	1.00	1.00	1	1
V43I	0.72	0.53	1.38	0.75
V43C	7.24	2.03	3.58	ND
V101A	0.93	0.69	1.34	0.81
F260I	1.59	1.04	1.52	1.24
F260D	1.21	1.11	1.09	ND
F260Q	1.19	1.14	1.05	ND
F260V	1.24	1.08	1.14	ND
I543N	1.50	1.18	1.27	0.73
I543W	0.97	0.81	1.20	0.51
I543S	1.09	0.92	1.19	1.09
I543A	1.17	0.98	1.19	1.00
I543G	0.93	0.86	1.08	0.90
I543L	1.05	0.96	1.10	1.29
S72N-V101M-F260I	0.76	0.58	1.32	0.95

V43I-S72N-V101M	0.59	0.39	1.50	0.85
S72N-V101M-I543N	0.74	0.54	1.37	0.87
S72N-V101M-I543D	0.62	0.45	1.38	ND
S72N-V101M-I543L	0.72	0.58	1.24	0.96
V43I-S72N-V101M-F260I	0.59	0.31	1.92	0.78
S72N-V101M-F260I-I543N	1.10	0.71	1.55	1.02
V43I-S72N-V101M-F260I-I543N	0.71	0.32	2.20	0.69
V43I-S72N-V101M- I543N	0.52	0.27	1.91	0.73
S72N-V101M-F260I-I543L	0.86	0.56	1.54	1.00

Genetic Constructs Encoding Modified TrCel3A Beta-glucosidases

[0058] The present invention also relates to genetic constructs comprising a nucleic acid sequence encoding the modified TrCel3A beta-glucosidase operably linked to regulatory nucleic acid sequences directing the expression and secretion of the modified TrCel3A beta-glucosidase from a host microbe. By “regulatory nucleic acid sequences” it is meant nucleic acid sequences directing the transcription and translation of the modified TrCel3A-encoding nucleic acid sequence and a nucleic acid sequence encoding a secretion signal peptide capable of directing the secretion of the modified TrCel3A beta-glucosidase from a host microbe. The regulatory nucleic acid sequences may be derived from genes that are highly expressed and secreted in the host microbe under industrial fermentation conditions. For example, the regulatory nucleic acid sequences may be derived from any one or more of the *Trichoderma reesei* cellulase or hemicellulase genes.

[0059] The genetic construct may further comprise a selectable marker gene to enable isolation of a genetically modified microbe transformed with the construct as is commonly known to those of skill in the art. The selectable marker gene may confer resistance to an antibiotic or the ability to grow on medium lacking a specific nutrient to the host organism that otherwise could not grow under these conditions. The present invention is not limited by the choice of selectable marker gene, and one of skill in the art may readily determine an appropriate gene. For example, the selectable marker gene may confer resistance to hygromycin,

phleomycin, kanamycin, geneticin, or G418, or may complement a deficiency of the host microbe in one of the *trp*, *arg*, *leu*, *pyr4*, *pyr*, *ura3*, *ura5*, *his*, or *ade* genes or may confer the ability to grow on acetamide as a sole nitrogen source.

[0060] The genetic construct may further comprise other nucleic acid sequences as is commonly known to those of skill in the art, for example, transcriptional terminators, nucleic acid sequences encoding peptide tags, synthetic sequences to link the various other nucleic acid sequences together, origins of replication, and the like. The practice of the present invention is not limited by the presence of any one or more of these other nucleic acid sequences.

Genetically Modified Microbes Expressing Modified TrCel3A Beta-glucosidases

[0061] The modified TrCel3A beta-glucosidase may be expressed and secreted from a genetically modified microbe produced by transformation of a host microbe with a genetic construct encoding the modified TrCel3A beta-glucosidase. The host microbe may be a yeast or a filamentous fungus, particularly those classified as Ascomycota. Genera of yeasts useful as host microbes for the expression of modified TrCel3A beta-glucosidases of the present invention include *Saccharomyces*, *Pichia*, *Hansenula*, *Kluyveromyces*, *Yarrowia*, and *Arxula*. Genera of fungi useful as microbes for the expression of modified TrCel3A beta-glucosidases of the present invention include *Trichoderma*, *Hypocrea*, *Aspergillus*, *Fusarium*, *Humicola*, *Neurospora*, *Chrysosporium*, *Myceliophthora*, *Thielavia*, *Sporotrichum* and *Penicillium*. For example, the host microbe may be an industrial strain of *Trichoderma reesei*. Typically, the host microbe is one which does not express a parental TrCel3A beta-glucosidase.

[0062] The genetic construct may be introduced into the host microbe by any number of methods known by one skilled in the art of microbial transformation, including but not limited to, treatment of cells with CaCl_2 , electroporation, biolistic bombardment, PEG-mediated fusion of protoplasts (e.g. White et al., WO 2005/093072, which is incorporated herein by reference). After selecting the recombinant fungal strains expressing the modified TrCel3A, the selected recombinant strains may be cultured in submerged liquid fermentations under conditions that induce the expression of the modified TrCel3A.

Production of Modified TrCel3A Beta-glucosidases

[0063] The modified TrCel3A beta-glucosidase of the present invention may be produced in a fermentation process in which a genetically modified microbe comprising a genetic construct encoding the modified TrCel3A beta-glucosidase is grown in submerged liquid culture fermentation.

[0064] Submerged liquid fermentations of microorganisms, including *Trichoderma* and related filamentous fungi, are typically conducted as a batch, fed-batch or continuous process. In a batch process, all the necessary materials, with the exception of oxygen for aerobic processes, are placed in a reactor at the start of the operation and the fermentation is allowed to proceed until completion, at which point the product is harvested. A batch process for producing the modified TrCel3A beta-glucosidase of the present invention may be carried out in a shake-flask or a bioreactor.

[0065] In a fed-batch process, the culture is fed continuously or sequentially with one or more media components without the removal of the culture fluid. In a continuous process, fresh medium is supplied and culture fluid is removed continuously at volumetrically equal rates to maintain the culture at a steady growth rate.

[0066] One of skill in the art is aware that fermentation medium comprises a carbon source, a nitrogen source as well as other nutrients, vitamins and minerals which can be added to the fermentation media to improve growth and enzyme production of the host cell. These other media components may be added prior to, simultaneously with or after inoculation of the culture with the host cell.

[0067] For the process for producing the modified TrCel3A beta-glucosidase of the present invention, the carbon source may comprise a carbohydrate that will induce the expression of the modified TrCel3A beta-glucosidase from a genetic construct in the genetically modified microbe. For example, if the genetically modified microbe is a strain of a cellulolytic fungus such as *Trichoderma*, the carbon source may comprise one or more of cellulose, cellobiose, sophorose, xylan, xylose, xylobiose and related oligo- or poly-saccharides known to induce expression of cellulases and beta-glucosidase in such cellulolytic fungi.

[0068] In the case of batch fermentation, the carbon source may be added to the fermentation medium prior to or simultaneously with inoculation. In the cases of fed-batch or continuous operations, the carbon source may also be supplied continuously or intermittently during the fermentation process. For example, when the genetically modified microbe is a strain of *Trichoderma*, the carbon feed rate is between 0.2 and 2.5 g carbon/L of culture/h, or any amount therebetween.

[0069] The process for producing the modified TrCel3A beta-glucosidase of the present invention may be carried at a temperature from about 20°C to about 40°C, or any temperature therebetween, for example from about 25°C to about 37°C, or any temperature therebetween, or from 20, 22, 25, 26, 27, 28, 29, 30, 32, 35, 37, 40°C or any temperature therebetween.

[0070] The process for producing the modified TrCel3A beta-glucosidase of the present invention may be carried out at a pH from about 3.0 to 6.5, or any pH therebetween, for example from about pH 3.5 to pH 5.5, or any pH therebetween, for example from about pH 3.0, 3.2, 3.4, 3.5, 3.7, 3.8, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5.0, 5.2, 5.4, 5.5, 5.7, 5.8, 6.0, 6.2, 6.5 or any pH therebetween.

[0071] Following fermentation, the fermentation broth containing the modified TrCel3A beta-glucosidase may be used directly, or the modified TrCel3A beta-glucosidase may be separated from the fungal cells, for example by filtration or centrifugation. Low molecular solutes such as unconsumed components of the fermentation medium may be removed by ultrafiltration. The modified TrCel3A beta-glucosidase may be concentrated, for example, by evaporation, precipitation, sedimentation or filtration. Chemicals such as glycerol, sucrose, sorbitol and the like may be added to stabilize the modified TrCel3A beta-glucosidase. Other chemicals, such as sodium benzoate or potassium sorbate, may be added to the modified TrCel3A beta-glucosidase to prevent growth of microbial contamination.

The Use of Modified TrCel3A Beta-glucosidases

[0072] The modified TrCel3A beta-glucosidase of the present invention may be used in the hydrolysis of cellulose or in the production of compounds such as those used in the medical and food industries

[0073] For use in the enzymatic hydrolysis of cellulose, such as in the production of fermentable sugars from a pretreated lignocellulosic feedstock, the modified TrCel3A beta-glucosidase of the invention may be combined with one or more cellulases to produce a cellulase mixture. In one embodiment of the invention, the modified TrCel3A beta-glucosidase is one of many proteins expressed from a host cell, including, but not limited to, cellulase enzymes. The one or more cellulases in the cellulase enzyme mixture and the modified TrCel3A beta-glucosidase may be secreted from a single genetically modified microbe or by different microbes in combined or separate fermentations. Similarly, the one or more cellulases in the cellulase enzymes mixture with which the modified TrCel3A beta-glucosidase may be combined may be expressed individually or in sub-groups from different strains of different organisms and the enzymes combined to make the cellulase enzyme mixture. It is also contemplated that the enzyme mixture may be expressed individually or in sub-groups from different strains of a single organism, such as from different strains of *Saccharomyces*, *Pichia*, *Hansenula*, *Trichoderma*, *Hyprocrea*, *Aspergillus*, *Fusarium*, *Humicola*, *Chrysosporium*, *Myceliophthora*, *Thielavia*, *Sporotrichum* or *Neurospora*, and the enzymes combined to make the cellulase enzyme mixture. Preferably, all of the enzymes are expressed from a single host organism, such as a strain of cellulolytic fungus belonging to a species of *Trichoderma*, *Hyprocrea*, *Aspergillus*, *Fusarium*, *Humicola*, *Chrysosporium*, *Myceliophthora*, *Thielavia*, *Sporotrichum* or *Neurospora*.

[0074] It is further contemplated that the cellulase mixture may comprise two or more of such modified beta-glucosidases as described here in, each with a unique set of improved kinetic parameters. Such a cellulase mixture would be expected to maintain a constant beta-glucosidase activity across a broad range of conditions. For example, a cellulase mixture may comprise one modified TrCel3A beta-glucosidase with low substrate affinity and low product inhibition (i.e., higher values of K_{G2} and K_G than the parental beta-glucosidase) and one modified TrCel3A beta-glucosidase with high substrate affinity and high product inhibition (i.e., lower values of K_{G2} and K_G than the parental beta-glucosidase). Such a cellulase mixture would exhibit a near level apparent beta-glucosidase activity across a wide range of cellobiose and glucose concentrations. Many possible combinations of two or more beta-glucosidase enzymes might be envisioned to maintain a constant activity across a variety of conditions that could occur across many different processes and applications.

[0075] A pretreated lignocellulosic feedstock is a material of plant origin that, prior to pretreatment, contains at least 20% cellulose (dry weight), more preferably greater than about 30% cellulose, even more preferably greater than 40% cellulose, for example 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 55, 60, 65, 70, 75, 80, 85, 90% or any percent therebetween, and at least 10% lignin (dry wt), more typically at least 12% (dry wt) and that has been subjected to physical and/or chemical processes to make the fiber more accessible and/or receptive to the actions of cellulolytic enzymes. After pretreatment, the lignocellulosic feedstock may contain higher levels of cellulose. For example, if acid pretreatment is employed, the hemicellulose component is hydrolyzed, which increases the relative level of cellulose. In this case, the pretreated feedstock may contain greater than about 20% cellulose and greater than about 12% lignin.

[0076] Lignocellulosic feedstocks that may be used in the invention include, but are not limited to, agricultural residues such as corn stover, wheat straw, barley straw, rice straw, oat straw, canola straw, sugarcane straw and soybean stover; fiber process residues such as corn fiber, sugar beet pulp, pulp mill fines and rejects or sugar cane bagasse; forestry residues such as aspen wood, other hardwoods, softwood, and sawdust; or grasses such as switch grass, miscanthus, cord grass, and reed canary grass. The lignocellulosic feedstock may be first subjected to size reduction by methods including, but not limited to, milling, grinding, agitation, shredding, compression/expansion, or other types of mechanical action. Size reduction by mechanical action can be performed by any type of equipment adapted for the purpose, for example, but not limited to, a hammer mill.

[0077] The enzymatic hydrolysis of cellulose using a cellulase enzyme mixture, as defined above, comprising the modified TrCel3A beta-glucosidase may be batch hydrolysis, continuous hydrolysis, or a combination thereof. The hydrolysis may be agitated, unmixed, or a combination thereof.

[0078] The enzymatic hydrolysis may be carried out at a temperature of about 30°C to about 80°C, or any temperature therebetween, for example a temperature of 30, 35, 40, 45, 50, 55, 60, 65, 70, 75°C, 80°C or any temperature therebetween, and a pH of about 3.0 to about 8.0, or any pH therebetween, for example at a pH of 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0

or pH therebetween. The initial concentration of cellulose in the hydrolysis reactor, prior to the start of hydrolysis, is preferably about 2% (w/w) to about 15% (w/w), or any amount therebetween, for example 2, 4, 6, 8, 10, 12, 14, 15% or any amount therebetween.

[0079] The dosage of the cellulase enzyme mixture comprising the modified TrCel3A beta-glucosidase may be about 0.1 to about 100 mg protein per gram cellulose, or any amount therebetween, for example 0.1, 0.5, 1, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100 mg protein per gram cellulose or any amount therebetween. The hydrolysis may be carried out for a time period of about 1 hours to about 200 hours, or any time therebetween; for example, the hydrolysis may be carried out for a period of 15 hours to 100 hours, or any time therebetween, or it may be carried out for 1, 2, 4, 8, 12, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 120, 140, 160, 180, 200 or any time therebetween. It should be appreciated that the reaction conditions are not meant to limit the invention in any manner and may be adjusted as desired by those of skill in the art.

[0080] In practice, the enzymatic hydrolysis is typically carried out in a hydrolysis reactor. The enzyme mixture is added to the pretreated lignocellulosic feedstock (also referred to as the “substrate”) prior to, during, or after the addition of the substrate to the hydrolysis reactor.

Modified Family 3 Beta-glycosidases

[0081] Beta-glucosidases are just one or several classes of hydrolytic enzymes belong to glycoside hydrolase Family 3. For example, Family 3 includes other enzymes that catalyse the hydrolysis of beta-glycosidic bonds such as xylan 1,4-beta-xylosidase (EC 3.2.1.37), beta-N-acetylhexosaminidase (EC 3.2.1.52), glucan 1,3- beta -glucosidase (EC 3.2.1.58), and glucan 1,4- beta -glucosidase (EC 3.2.1.74). For the purposes of the present invention, a “Family 3 beta-glycosidase” is any xylan 1,4-beta-xylosidase (EC 3.2.1.37), beta-N-acetylhexosaminidase (EC 3.2.1.52), glucan 1,3- beta -glucosidase (EC 3.2.1.58), and glucan 1,4- beta -glucosidase (EC 3.2.1.74) that is classified as a Family 3 glycoside hydrolase under the CAZy system (see URL afmb.cnrs-mrs.fr/CAZY/index.html for reference).

[0082] By “modified Family 3 beta-glycosidase”, it is meant a Family 3 beta-glycosidase which comprises one or more of the amino acid substitutions, introduced by genetic engineering

techniques, selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543A, I543S, I543G and I543L (TrCel3A numbering) and which amino acid sequence is at least 80% identical to the amino acid sequence of the parental Family 3 beta-glycosidase from which it is derived. For example, the amino acid sequence of the modified Family 3 beta-glycosidase may be from about 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 99.9% identical to the amino acid sequence of the parental Family 3 beta-glycosidase from which it is derived.

[0083] Sequence identity can be readily determined by alignment of the amino acids of the two sequences, either using manual alignment, or any sequence alignment algorithm as known to one of skill in the art, for example but not limited to, BLAST algorithm (BLAST and BLAST 2.0; Altschul et al., 1997 and 1990), the algorithm disclosed by Smith & Waterman (1981), by the homology alignment algorithm of Needleman & Wunsch (1970), by the search for similarity method of Pearson & Lipman (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection. An alignment of 45 Family 3 beta-glycosidase sequences is provided in Figure 11.

[0084] It will be understood that the modified Family 3 beta-glycosidase may be derived from wild-type Family 3 beta-glycosidase or from a Family 3 beta-glycosidase that contains other amino acid substitutions. Alternatively, after production of the modified Family 3 beta-glycosidase comprising mutations selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543A, I543S, I543G and I543L, it may be subsequently further modified to contain additional amino acid substitutions, including but not limited to those set forth above.

[0085] By "TrCel3A numbering" it is meant the numbering corresponding to the position of amino acids based on the amino acid sequence of TrCel3A (SEQ ID NO:1) based on alignment of the amino acid sequence of the Family 3 beta-glycosidase with the TrCel3A amino acid sequence. An example of the alignment of 44 other Family 3 beta-glycosidase amino acid sequences with the TrCel3A beta-glucosidase amino acid sequence is provided in Figure 11.

[0086] As used herein in respect of modified Family 3 beta-glycosidase amino acid sequences, “derived from” refers to the isolation of a target nucleic acid sequence element encoding the desired modified Family 3 beta-glycosidase using genetic material or nucleic acid or amino acid sequence information specific to the corresponding parental Family 3 beta-glycosidase. As is known by one of skill in the art, such material or sequence information can be used to generate a nucleic acid sequence encoding the desired modified Family 3 beta-glycosidase using one or more molecular biology techniques including, but not limited to, cloning, sub-cloning, amplification by PCR, *in vitro* synthesis, and the like.

[0087] In one embodiment of the invention of the invention, the amino acid sequence of the modified Family 3 beta-glycosidase is from about 80% to about 99.9% identical to the amino acid sequence of the parental Family 3 beta-glycosidase from which it is derived. For example, the amino acid sequence of the Family 3 beta-glycosidase may be from about 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 99.9% identical to the amino acid sequence of the parental Family 3 beta-glycosidase from which it is derived. In other words, the number of amino acid substitutions in the modified Family 3 beta-glycosidase does not exceed 20% of the total number amino acids in the parental Family 3 beta-glycosidase sequence.

[0088] The modified Family 3 beta-glycosidase of the present invention exhibit improvements in at least one of the following kinetic parameters: K_p , K_s and k_{cat} . K_p is defined as the concentration of product which reduces the enzymatic activity of the Family 3 beta-glycosidase by 50%. K_s is defined as the concentration of substrate at which the Family 3 beta-glycosidase exhibits half its maximal rate. The k_{cat} is the catalytic rate constant for the hydrolysis of substrate.

[0089] In another embodiment of the invention, the amino acid sequence of the modified Family 3 beta-glycosidase is from about 90% to about 99.9% identical to the amino acid sequence of the parental Family 3 beta-glycosidase from which it is derived. For example, the amino acid sequence of the modified Family 3 beta-glycosidase may be from about 95% to about 100% identical to SEQ ID NO: 1.

[0090] In another embodiment, the amino acid sequence of the modified Family 3 beta-glycosidase may be from about 80% to about 99.9% identical to the amino acid sequence of the

parental Family 3 beta-glycosidase from which it is derived, and the modified Family 3 beta-glycosidase may exhibit (a) an increase in the K_p , (b) a decrease in K_s , or (c) an increase in k_{cat} relative to the K_p , K_s and/or k_{cat} of a parental Family 3 beta-glycosidase from which is derived.

[0091] By “parental Family 3 beta-glycosidase”, it is meant a Family 3 beta-glycosidase that does not contain:

isoleucine or cysteine at position 43,

alanine or glycine at position 101;

isoleucine, valine, glutamine, aspartic acid at position 260; or

asparagine, alanine, serine, glycine or leucine at position 543.

[0092] The modified Family 3 beta-glycosidase may be derived from a parental Family 3 beta-glycosidase that comprises one or more naturally-occurring amino acid(s) at the substituted positions corresponding to that of the modified Family 3 beta-glycosidase, but that is otherwise identical to the modified Family 3 beta-glycosidase, for example a native Family 3 beta-glycosidase from *A. nidulans* –AN1804.2, *B. fuckeliana*, *T. aurantiacus levisporus*. The parental Family 3 beta-glycosidase may contain one or more amino acid substitutions at other positions, given that these substitutions are also present in the corresponding modified Family 3 beta-glycosidase. Family 3 beta-glycosidases suitable as parental beta-glycosidases from which modified Family 3 beta-glycosidases may be derived are provided in Table 3.

Table 3: Family 3 beta-glycosidases

Organism	Protein	GenPept Accession Number	SEQ ID NO: 1	Identity with TrCel3A (%)
<i>Aspergillus aculeatus</i> F-50	b-glucosidase 1 (Bgl1)	BAA10968.1	72	44.1
<i>Aspergillus fumigatus</i> Af293	b-glucosidase (AFUA_1G17410;Afulg17410)	EAL91070.1	73	35.2
<i>Aspergillus fumigatus</i> Af293	b-glucosidase (AFUA_1G05770;Afulg05770)	EAL88289.1	74	43.1
<i>Aspergillus kawachii</i> ifo4308	b-glucosidase (BglA)	BAA19913.1	75	42.6
<i>Aspergillus nidulans</i> FGSC A4	b-glucosidase (AN1804.2)	EAA64969.1	76	13.2
<i>Aspergillus nidulans</i> FGSC A4	b-glucosidase (AN0712.2)	EAA65189.1	77	19.4

<i>Aspergillus niger B1</i>	b-glucosidase / tannase (Bgl1;BG3;BGs;SP188)	CAB75696.1	78	42.8
<i>Aspergillus niger CBS 513.88</i>	An18g03570 (Bgl1)	CAK48740.1	79	42.6
<i>Aspergillus oryzae</i>	b-glucosidase	CAD67686.1	80	42.2
<i>Aspergillus oryzae RIB 40</i>	b-glucosidase (AO090009000356)	BAE54829.1	81	42.3
<i>Aspergillus oryzae RIB 40</i>	b-glucosidase 5 (Bgl5;AO090001000544)	BAE57053.1	82	43.4
<i>Aspergillus oryzae RIB 40</i>	b-glucosidase 1 (Bgl1;AO090003001511)	BAE58551.1	83	31.6
<i>Botryotinia fuckeliana</i>	b-glucosidase (Sap1)	CAB61489.1	84	35.9
<i>Coccidioides posadasii C735</i>	b-glucosidase/ exo-b-1,3- glucosidase (Bgl2)	AAF21242.1	85	41.4
<i>Coccidioides posadasii C735</i>	b-glucosidase (Bgl1)	AAB67972.1	86	40.4
<i>Dictyostelium discoideum Ax3</i>	b-glucosidase	AAA74233.1	87	16.6
<i>Hordeum vulgare</i>	exo-1,3-glucanase II (EII;ExoII)	AAC49170.1	88	19.5
<i>Hypocrea jecorina QM6a</i>	Cel3c - Cel3C	AAP57756.1	89	19.4
<i>Hypocrea jecorina QM6a</i>	Cel3e - Cel3E	AAP57760.1	90	36.5
<i>Hypocrea jecorina QM6a</i>	Cel3d (fragment) - Cel3D	AAP57759.1	91	12.8
<i>Hypocrea jecorina QM6a</i>	b-glucosidase - Cel3B	AAP57755.1	92	41.5
<i>Hypocrea jecorina RutC-30</i>	b-xylosidase (Bx11)	CAA93248.1	93	19.2
<i>Kluyveromyces fragilis</i>	b-glucosidase	CAA29353.1	94	18.0
<i>Kuraishia capsulata 35M5N</i>	b-glucosidase (BglN)	AAA91297.1	95	33.0
<i>Nicotiana tabacum</i>	b-glucosidase	BAA33065.1	96	20.0
<i>Periconia sp. BCC 2871</i>	b-glucosidase	ABX84365.1	97	41.9
<i>Phaeosphaeria avenaria WAC1293</i>	b-glucosidase (Bgl1)	CAB82861.1	98	42.7
<i>Phanerochaete chrysosporium K-3</i>	glucan b-1,3-glucosidase (Bgl)	BAB85988.1	99	52.6
<i>Phanerochaete chrysosporium OGC101</i>	glucan 1,3-b-glucosidase (CbgL) - Bgl1A	AAC26489.1	100	52.6
<i>Pichia anomala var. acetaetherius</i>	b-glucosidase	CAA26662.1	101	36.4
<i>Piromyces sp. E2</i>	b-glucosidase (Cel3A) - Cel3A	AAO41704.1	102	38.3
<i>Saccharomycopsis fibuligera</i>	b-glucosidase 2 (Bgl2)	AAA34315.1	103	38.2
<i>Saccharomycopsis fibuligera</i>	b-glucosidase 1 (Bgl1)	AAA34314.1	104	38.3
<i>Schizosaccharomyces pombe 972h-</i>	b-glucosidase (SPBC1683.04)	CAB91166.1	105	17.4
<i>Septoria lycopersici</i>	b-glucosidase (tomatinase;B2Tom)	AAB08445.1	106	37.2
<i>Talaromyces emersonii</i>	b-glucosidase - Cel3A	AAL69548.3	107	43.2
<i>Thermoascus aurantiacus</i>	b-glucosidase (Bg2;BGII)	AAY33982.1	108	45
<i>Thermoascus aurantiacus IFO 9748</i>	b-glucosidase (Bgl1;Bgl)	AAZ95587.1	109	44.3
<i>Thermoascus aurantiacus var. levisporus</i>	b-1,4-glucosidase (Bgl1)	ABX79552.1	110	44.3
<i>Thermoascus aurantiacus var. levisporus</i>	b-1,4-glucosidase (Bgl2)	ABX56926.1	111	45
<i>Trichoderma viride AS 3.3711</i>	b-D-glucoside glucohydrolase (Bgl1)	AAQ76093.1	112	98.3
<i>Tropaeolum majus</i>	b-glucosidase	CAA07070.1	113	20.3
<i>Uromyces viciae-fabae</i>	b-glucosidase (Bgl1)	CAE01320.1	114	39.8
<i>Volvariella volvacea V14</i>	b-glucosidase (Bgl;BGL-II)	AAG59831.1	115	18.8

EXAMPLES

[0093] The present invention will be further illustrated in the following examples. However, it is to be understood that these examples are for illustrative purposes only and should not be used to limit the scope of the present invention in any manner.

[0094] Example 1 describes the strains and vectors used in the following examples. Example 2 describes the cloning of the TrCel3A gene and transformation in yeast. Example 3 summarizes the preparation of the error prone-PCR and site-saturation mutagenesis libraries of TrCel3A. Example 4 describes the selection and expression of wild-type and modified TrCel3A beta-glucosidases from microculture. Example 5 describes the high-throughput screening assays to identify modified TrCel3A beta-glucosidases with improved kinetic parameters. Example 6 describes the construction of modified TrCel3A beta-glucosidases comprising multiple amino acid substitutions. Example 7 describes the expression and purification of wild-type and modified TrCel3A beta-glucosidases from large scale flask cultures. Example 8 describes the measurement of the K_G and K_{G2} of modified and native TrCel3A beta-glucosidases and the calculation of kinetic parameters. Example 9 describes the measurement and calculation of k_{cat} for wild-type and modified TrCel3A beta-glucosidases. Example 10 describes the construction of genetically modified *Trichoderma reesei* strains expressing cellulase mixtures comprising modified TrCel3A beta-glucosidases and the production of such cellulase mixtures in submerged liquid culture fermentation. Example 11 describes the hydrolysis of pretreated lignocellulosic substrates with cellulase mixtures comprising parental and modified TrCel3A beta-glucosidases.

Example 1: Strains and Vectors

[0095] *Saccharomyces cerevisiae* strain BJ3505 (pep4::HIS3 prb- Δ 1.6R HIS3 lys2-208 trp1- Δ 101 ura3-52 gal2 can1) was obtained from Sigma and was a part of the Amino-Terminal Yeast FLAG Expression Kit. The YEp352/PGK91-1 vector was obtained from the National Institute of Health. The pGEM T-easy vector was obtained from Promega. The vector pC/XBG1-TV is described in U.S. Patent No. 6,105,703.

Example 2: Cloning of the TrCel3A gene into YEp352/PGK91-1 and transformation in yeast

[0096] The TrCel3A gene (SEQ ID NO: 44) contains two introns. One intron is located in the secretion signal at position 323 bp to 391 bp, while the other is located within the gene at position 2152 bp to 2215 bp. The TrCel3A gene contains a unique *NheI* site located at position 1203 bp. In order to facilitate expression from yeast and cloning using *NheI* and *KpnI* restriction enzymes, the unique *NheI* located within TrCel3A at position 1203 bp and the second intron were removed by a three step PCR. The TrCel3A gene was amplified in three segments from a plasmid containing a genomic subclone of the coding region, including introns, of the mature TrCel3A beta-glucosidase, pC/XBG1-TV using iPROOF DNA polymerase (BioRad). The first fragment (A) was amplified using primers which introduced an *NheI* site at the 5' end of the gene downstream of the secretion signal (AT048) and which removed the internal *NheI* site (AT051). The second fragment (B) was amplified using primers which removed the internal *NheI* site (AT050) and the intron at position 2152 to 2215 bp (AT053). The third fragment (C) was amplified using primers which removed the intron at position 2152 to 2215 bp (AT052) and introduced a *KpnI* site at the 3' end of the gene, downstream of the stop codon (AT049). Gene products B and C were joined together (to make gene product D) using PCR with primers AT050 and AT049. Gene product D was joined with gene product A using PCR with primers AT048 and AT049 to obtain TrCel3A without introns and with unique *NheI* and *KpnI* sites at the 5' and 3' ends, respectively. The final gene product was cloned into the pGEM T-easy vector (Promega) as per the manufacturer's instructions to make plasmid pGEM-TrCel3A. Primer sequences are shown below:

AT048: 5' CGC CAG GCT AGC GTT GTA CCT CCT GC (SEQ ID NO: 24)
 AT049: 5' CTG AGG GTA CCG CTA CGC TAC CGA C (SEQ ID NO: 25)
 AT050: 5' CCC GCT AGT ATT GCC GTC GTT GGA TC (SEQ ID NO: 26)
 AT051: 5' CCA ACG ACG GCA ATA CTA GCG GGC TTC (SEQ ID NO: 27)
 AT052: 5' GTT CGG CTA TGG ACT GTC TTA CAC CAA GTT CAA CTA C (SEQ ID NO: 28)
 AT053: 5' GTT GAA CTT GGT GTA AGA CAG TCC ATA GCC GAA CTC (SEQ ID NO: 29)

[0097] Preparation of YEp352/PGK91-1/ α_{ss} TrCel3A was conducted as follows. A DNA adapter containing *NheI*, *KpnI*, and *EcoRI* restriction sites was prepared by annealing primers AT046 and AT047 together. The adapter was inserted into a YEp based-plasmid containing the *pgk1* promoter, alpha mating factor secretion signal, and *pgk1* terminator sequences to make

plasmid YEp352/PGK91-1/ α_{ss} NKE. Specifically, the adapter was inserted as an *NheI/EcoRI* fragment into the *NheI* and *EcoRI* sites located downstream of the alpha mating factor secretion signal and upstream of the *pgk1* terminator. Primer sequences are shown below:

AT046: 5' CTA GCT GAT CAC TGA GGT ACC G (SEQ ID NO: 30)

AT047: 5' AAT TCG GTA CCT CAG TGA TCA G (SEQ ID NO: 31)

[0098] Plasmid pGEM-TrCel3A was digested with *NheI* and *EcoRI* to release the 2235 bp TrCel3A gene. The fragment was purified and ligated into the *NheI* and *EcoRI* sites of YEp352/PGK91-1/ α_{ss} NKE to obtain YEp352/PGK91-1/ α_{ss} TrCel3A. The resulting vector YEp352/PGK91-1/ α_{ss} -TrCel3A was transformed in yeast strain BJ3505 using the procedure described by Gietz, R. D. and Woods, R. A. (2002).

[0099] Preparation of YEp352/PGK91-1/ α_{ss} 6H-TrCel3A was conducted as follows. A DNA adapter containing *SpeI*, *NheI*, *KpnI*, and *EcoRI* restriction sites was prepared by annealing primers AT044 and AT045 together. The adapter contains sequences coding for six histidine residues downstream of the *SpeI* site and upstream of the *NheI* site. The adapter was inserted into a YEp based-plasmid containing the *pgk1* promoter, alpha mating factor secretion signal, and *pgk1* terminator sequences to make plasmid YEp352/PGK91-1/ α_{ss} 6HNKE. Specifically, the adapter was inserted as an *NheI/EcoRI* fragment into the *NheI* and *EcoRI* sites located downstream of the alpha mating factor secretion signal and upstream of the *pgk1* terminator. Primer sequences are shown below:

AT044: 5' CTA GTC ATC ACC ATC ACC ATC ACG CTA GCT GAT CAC TGA GGT ACC G
(SEQ ID NO: 32)

AT045: 5' AAT TCG GTA CCT CAG TGA TCA GCT AGC GTG ATG GTG ATG GTG ATG A
(SEQ ID NO: 33)

[00100] Plasmid pGEM-TrCel3A was digested with *NheI* and *EcoRI* to release the 2235 bp TrCel3A gene. The fragment was purified and ligated into the *NheI* and *EcoRI* sites of YEp352/PGK91-1/ α_{ss} 6HNKE to obtain YEp352/PGK91-1/ α_{ss} 6H-TrCel3A. The resulting vector YEp352/PGK91-1/ α_{ss} 6H-TrCel3A (Figure 1) was transformed in yeast strain BJ3505 using the procedure described by Gietz, R. D. and Woods, R. A. (2002).

Example 3: Random Mutagenesis of TrCel3A

a. Error Prone-PCR

[00101] A random mutagenesis library was generated by error-prone PCR using a Mutazyme® II DNA polymerase method. A series of four independent PCRs was performed using 5, 10, 15, 20 ng of YEp352/PGK91-1/ α_{ss} 6H-TrCel3A vector and the Mutazyme® II DNA polymerase with primers Y α N21 and 3'PGK-term. Annealing temperature was set to 50°C. The amplification was done for 20 cycles. The four PCR products were pooled and diluted to 16 ng/ μ L. The YEp352/PGK91-1/ α_{ss} 6H-TrCel3A vector was digested with NheI and KpnI and the empty vector fragment was isolated. This linear fragment and the final amplicon were transformed simultaneously and cloned by *in vivo* recombination into yeast strain BJ3505 (Butler, T. and Alcalde, M., 2003).

Y α N21: 5'AGC ACA AAT AAC GGG TTA TTG (SEQ ID NO: 34)

3'PGK-term: 5'GCA ACA CCT GGC AAT TCC TTA CC (SEQ ID NO: 35)

b. Site-Saturation Mutagenesis

[00102] Four TrCel3A libraries were created using site-saturation mutagenesis (SSM) with degenerate primers (NNS) targeting amino acid positions V43, V101, F260, and I543. SSM was performed using a two-step PCR method involving megaprimer synthesis followed by PCR-mediated overlap extension. PCR reactions were carried out using the High Fidelity iProof *Taq* Polymerase (BioRad). YEp352/PGK91-1/ α_{ss} 6H-TrCel3A was used as the template for the V43X, F260X, and I543X libraries, while YEp352/PGK91-1/ α_{ss} 6H-TrCel3A (S72N, F96L, V101M, N369K, A386T) served as the template for the V101X library.

[00103] For each SSM library, MegaPrimer A was amplified using the external primer Y α N21 with an internal reverse primer, while MegaPrimer B was derived by combining the external primer PGKterm with an internal forward primer. The internal forward primers contained a degenerate codon sequence to introduce random amino acid substitutions within their target sites. The megaprimers were purified using the Wizard® SV Gel and PCR Clean-Up System.

YαN21: 5'AGC ACA AAT AAC GGG TTA TTG (SEQ ID NO: 34)
 PGKterm: 5'GCA ACA CCT GGC CCT TAC C (SEQ ID NO: 35)
 V43X-F: 5'TGG AAC GGC GGT CCT TGC NNS GGA AAC ACA TCT CCG GCC TC
 (SEQ ID NO: 36)
 V43X-R: 5'GCA AGG ACC GCC GTT CCA (SEQ ID NO: 37)
 M101X-F: 5'C GGT GAG GAG NNS AAG GCC TCG G (SEQ ID NO: 38)
 M101X-R: 5'ATG AAC TGT CCA CGT TCG CGG (SEQ ID NO: 39)
 F260X-F: 5'TG TCA ATG CCT GGC ACA GAC NNS AAC GGT AAC AAT CGG (SEQ ID NO:
 40)
 F260X-R: 5'GT CTG TGC CAG GCA TTG ACA (SEQ ID NO: 41)
 I543X-F: 5'CCC AAT GAC TAT AAC ACT CGC NNS GTT TCC GGC GGC AGT GAC (SEQ
 ID NO: 42)
 I543X-R: 5'GCG AGT GTT ATA GTC ATT GGG (SEQ ID NO: 43)

[00104] In the second round of PCR, both megaprimers for a given SSM library were allowed to anneal and extend for 10 cycles to generate the final template. The external primers YαN21 and PGKterm were then added for another 25 cycles to amplify the final product, which was subsequently purified using the Wizard® SV Gel and PCR Clean-Up System. Both the purified PCR product and the linearized vector YEp352/PGK91-1α_{ss}-6H-TrCel3A (*NheI* + *NruI*) were transformed and cloned via *in vivo* recombination within the *BJ3505* yeast strain using the procedure described by Gietz, R. D. and Woods (2002).

Example 4: Selection and growth of yeast strains expressing parental and modified TrCel3A beta-glucosidases.

[00105] This example describes the selection and expression of parental (TrCel3A^{Wt}) and modified TrCel3A beta-glucosidase from *Saccharomyces cerevisiae* for use in high-throughput screening assays.

[00106] *S. cerevisiae* transformants were grown on plates containing synthetic complete medium (SC: 2% agar w/v, 0.17% yeast nitrogen base w/v, 0.192% -Ura drop-out supplement w/v, 2% glucose w/v, 2% casamino acids w/v, 0.5% ammonium sulfate w/v, pH 5.5) for 3-4 days at 30°C. Each growth plate was replicated by transferring a portion of each colony, using sterilized velvet, to a screen-out plate containing SC medium plus 0.1% esculin hydrate and 0.03% FeCl₃. Colonies which turned black after incubation for 3-4 days at 30°C were identified

as expressing active beta-glucosidase (Figure 2). Colonies were correlated back to their original growth plate and selected for liquid media expression cultures by toothpick inoculation of 1 mL SC media in sterile 96-deepwell plates containing one glass bead (1.5 - 2.0 mm diameter). Expression cultures were grown for 3 days at 30°C and 245 rpm with humidity control (New Brunswick Scientific Innova[®]44 incubator shaker series). Glycerol stocks were prepared by transferring 50 µL of liquid culture to the corresponding wells of a microplate containing 50 µL of 40% glycerol and stored at -80°C. Expression culture plates were centrifuged at 1600 X g for 5 minutes to pellet cells and supernatant was aspirated for screening assays (Example 5).

Example 5: Screening of TrCel3A Gene Libraries for Modified TrCel3A beta-glucosidases with Higher Catalytic Efficiency

[00107] This example describes the screening of modified TrCel3A beta-glucosidases for increased higher catalytic efficiency by comparison to parental TrCel3A that had been cloned into *Saccharomyces cerevisiae*.

[00108] Modified TrCel3A beta-glucosidases expressed from yeast as described in Example 4 were tested in two 80 µL citrate buffered (pH 5) cellobiose hydrolysis assays using a 96-well PCR plate format. A 40 µL aliquot of supernatant containing a parental or modified TrCel3A beta-glucosidase was incubated with 30 mM cellobiose (Assay 1) and 5.0 mM cellobiose plus 1.25 mM glucose (Assay 2) for 5, 10, 20 and 40 min at 50°C in an MJ Research Tetrad[™]2 Peltier Thermal Cycler. Contained in each 96-well PCR plate were six parental TrCel3A controls used for comparison. Enzyme activity was measured through the detection of glucose using a glucose oxidase- peroxidase coupled assay (Trinder P., 1969). Exogenous glucose (1.25 mM) included in Assay 2 was subtracted from the total amount of glucose measured following the incubation. An Assay 2/Assay 1 enzyme activity ratio was calculated for the parental (TrCel3A^{Wt} or Wt) and all modified TrCel3A beta-glucosidases by dividing the enzyme activity in Assay 2 by the enzyme activity in Assay 1. The Assay 2/Assay 1 activity ratio for each modified TrCel3A beta-glucosidase was then compared to that of the average of six parental TrCel3A beta-glucosidase controls on a particular microplate and positives were selected at the 95% confidence level using a t-test. . All positive modified TrCel3A beta-glucosidases were produced again in microculture and re-screened to reduce the number of false positives (Figure 3). Table 4 lists the positive

modified TrCel3A beta-glucosidases obtained from screening the error-prone and site-saturation libraries (Example 3) and the Assay 2/Assay 1 enzyme activity ratios compared to the parental, wild-type TrCel3A beta-glucosidase.

Table 4. Modified TrCel3A beta-glucosidases selected random libraries.

Modified TrCel3A	Assay 2 / Assay 1 Enzyme Activity Ratio (normalized to TrCel3A^{wt})
Wt	1.00
V43I	1.20
V43C	1.36
V101A	1.17
V101G	1.17
F260I	1.14
I543N	1.12
I543W	1.26
I543A	1.23
I543G	1.22
I543L	1.21
I543S	1.25

Example 6: Construction and expression of modified TrCel3A beta-glucosidases with multiple amino acid substitutions

[00109] Using YEp352/PGK91-1/ α_{ss} 6H-TrCel3A(S72N-F96L-V101M (U.S. Publication No. 2010/0093040A1) as a template, additional mutations were introduced using a two-step PCR method involving megaprimer synthesis followed by megaprimer PCR using High Fidelity iProof *Taq* Polymerase (Table 5). The internal primers were modified to introduce the desired amino acid substitutions into the TrCel3A construct. The external plasmid primers (Y α N21 and PGKterm) were used to amplify the final product. The megaprimers and final products were purified using the Wizard® SV Gel and PCR Clean-Up System.

Table 5: Generation of modified TrCel3A enzymes by PCR.

PCR	Step	Template	Primer 1	Primer 2	Amplicon
1	1	YEp352/PGK91-1- α_{ss} -6H- <i>TrCel3A(S72N-F96L-V101M)</i>	Y α N21	DK068	PCR 1 Step 1
	1	YEp352/PGK91-1- α_{ss} -6H- <i>TrCel3A(S72N-F96L-V101M)</i>	DK067	PGK-Term	PCR 1 Step 1
	2	Both PCR 1 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(S72N-V101M)</i>
2	1	<i>TrCel3A(S72N-V101M)</i>	Y α N21	DK222	PCR 2 Step 1
	1	<i>TrCel3A(S72N-V101M)</i>	DK221	PGK-Term	PCR 2 Step 1
	2	Both PCR 2 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(V43I-S72N-V101M)</i>
3	1	<i>TrCel3A(S72N-V101M)</i>	Y α N21	DK106	PCR 3 Step 1
	1	<i>TrCel3A(S72N-V101M)</i>	DK105	PGK-Term	PCR 3 Step 1
	2	Both PCR 3 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(S72N-V101M-F260I)</i>
4	1	<i>TrCel3A(S72N-V101M)</i>	Y α N21	DK224	PCR 4 Step 1
	1	<i>TrCel3A(S72N-V101M)</i>	DK223	PGK-Term	PCR 4 Step 1
	2	Both PCR 4 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(S72N-V101M-I543D)</i>
5	1	<i>TrCel3A(S72N-V101M)</i>	Y α N21	DK232	PCR 5 Step 1
	1	<i>TrCel3A(S72N-V101M)</i>	DK231	PGK-Term	PCR 5 Step 1
	2	Both PCR 5 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(S72N-V101M-I543L)</i>
6	1	<i>TrCel3A(S72N-V101M)</i>	Y α N21	DK230	PCR 6 Step 1
	1	<i>TrCel3A(S72N-V101M)</i>	DK229	PGK-Term	PCR 6 Step 1
	2	Both PCR 6 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(S72N-V101M-I543N)</i>
7	1	<i>TrCel3A(V43I-S72N-V101M)</i>	Y α N21	DK230	PCR 7 Step 1
	1	<i>TrCel3A(V43I-S72N-V101M)</i>	DK229	PGK-Term	PCR 7 Step 1
	2	Both PCR 7 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(V43I-S72N-V101M-I543N)</i>
8	1	<i>TrCel3A(S72N-V101M-F260I)</i>	Y α N21	DK222	PCR 8 Step 1
	1	<i>TrCel3A(S72N-V101M-F260I)</i>	DK221	PGK-Term	PCR 8 Step 1
	2	Both PCR 8 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(V43I-S72N-V101M-F260I)</i>
9	1	<i>TrCel3A(S72N-V101M-F260I)</i>	Y α N21	DK232	PCR 9 Step 1
	1	<i>TrCel3A(S72N-V101M-F260I)</i>	DK231	PGK-Term	PCR 9 Step 1
	2	Both PCR 9 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(S72N-V101M-F260I-I543L)</i>
10	1	<i>TrCel3A(S72N-V101M-F260I)</i>	Y α N21	DK230	PCR 10 Step 1
	1	<i>TrCel3A(S72N-V101M-F260I)</i>	DK229	PGK-Term	PCR 10 Step 1
	2	Both PCR 10 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(S72N-V101M-F260I-I543N)</i>
11	1	<i>TrCel3A(S72N-V101M-F260I-I543N)</i>	Y α N21	DK222	PCR 11 Step 1
	1	<i>TrCel3A(S72N-V101M-F260I-I543N)</i>	DK221	PGK-Term	PCR 11 Step 1
	2	Both PCR 11 Step 1 megaprimers	Y α N21	PGK-Term	<i>TrCel3A(V43I-S72N-V101M-F260I-I543N)</i>

[00110] To facilitate cloning, the final product was digested with *NheI* + *KpnI* and ligated into vector YEp352/PGK91-1/ α_{ss} 6H-TrCel3A linearized with *NheI* + *KpnI*. The ligation mix was transformed into *DH5 α* chemically-competent *E. coli* cells, plasmid extracted, and sequenced. Plasmids encoding the modified beta-glucosidases were transformed into yeast strain *BJ3505*.

Y α N21 5'-AGCACAAATAACGGGTATTG-3' (SEQ ID NO: 34)
 3'PGKterm 5'-GCAACACCTGGCCCTTACC-3' (SEQ ID NO: 35)
 5'DK067 5'-CGCGAACGTGGACAGTTCATCGGTGAGGAGATG-3' (SEQ ID NO: 45)
 3'DK068 5'-CATCTCCTCACCGATGAACTGTCCACGTTTCGCG-3' (SEQ ID NO: 46)
 5'DK105 5'-CAATGCCTGGCACAGACATCAACGGTAACAATC-3' (SEQ ID NO: 47)
 3'DK106 5'-GATTGTTACCGTTGATGTCTGTGCCAGGCATTG-3' (SEQ ID NO: 48)
 5'DK221 5'-GGCGGTCCTTGCATTGGAAACACAT-3' (SEQ ID NO: 49)
 3'DK222 5'-ATGTGTTTCCAATGCAAGGACCGCC-3' (SEQ ID NO: 50)
 5'DK223 5'-GACTATAACACTCGCGACGTTTCCGGCGGCAG-3' (SEQ ID NO: 51)
 3'DK224 5'-CTGCCGCCGGAAACGTCGCGAGTGTTATAGTC-3' (SEQ ID NO: 52)
 5'DK229 5'-GACTATAACACTCGCAACGTTTCCGGCGGCAG-3' (SEQ ID NO: 53)
 3'DK230 5'-CTGCCGCCGGAAACGTTGCGAGTGTTATAGTC-3' (SEQ ID NO: 54)
 5'DK231 5'-GACTATAACACTCGCCTGGTTTCCGGCGGCAG-3' (SEQ ID NO: 55)
 3'DK232 5'-CTGCCGCCGGAAACAGGCGAGTGTTATAGTC-3' (SEQ ID NO: 56)

Example 7: Purification of Modified TrCel3A Beta-glucosidases

[00111] Modified TrCel3A beta-glucosidases that passed the selection criteria in Example 5, along with the modified TrCel3A-F260X beta-glucosidases produced by site-saturation mutagenesis at position 260 (Example 4) or by combining two or more amino acid substitutions (Example 6), were purified for further analysis. For each modified TrCel3A beta-glucosidase, 50 mL of sterile YPD medium (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose) was inoculated with 10 mL of overnight cultures of transformed *Saccharomyces cerevisiae* grown from cells freshly picked from an agar plate. The cultures were then incubated for 96 hours at 30°C with shaking at 200 rpm.

[00112] After incubation, the broth from each culture was centrifuged for 10 minutes at 9000 rpm and the pellet (containing yeast cells) discarded. The pH of the supernatant was adjusted to 5.0. The TrCel3A in each spent culture medium was then purified by immobilized metal affinity chromatography (IMAC) using His-Trap NTA/ Ni^{2+} columns from GE Healthcare (catalogue #17-5247-01). Purified proteins were concentrated and buffer exchanged using Vivaspin 20 centrifugal concentrators (Sartorius Stedim Biotech, catalogue No. VS2012). Protein concentrations were measured using the method of Bradford (1976) and stored at -20°C . Samples of each purified Modified TrCel3A were separated by SDS-PAGE and visualized by Coomassie Blue stain (Figure 4).

Example 8: Measuring the K_G and K_{G2} constants of Parental and Modified TrCel3A Beta-glucosidases

[00113] The K_G and K_{G2} constants of each modified TrCel3A beta-glucosidase were determined using a *p*-nitrophenyl-beta-D-glucopyranoside (pNPG) competitive substrate/inhibitor real-time kinetic assay. Each modified TrCel3A (3 μg /reaction) was incubated with 0.4 mM *p*NPG in a stirred cuvette; the total reaction volume was 3 mL. Assays were buffered using 50 mM citrate, pH 5.0. Incubations were done at 50°C for up to 40 min in a Varian Cary UV/Vis spectrophotometer. Absorbance measurements collected at 340 nm during the time course were converted to *p*-nitrophenol (pNP) concentration using Equation 1.

$$pNP = \frac{Abs_{340nm} - \epsilon_{pNPG} \cdot pNPG}{\epsilon_{pNP} - \epsilon_{pNPG}} \quad \text{Equation 1}$$

where,

*p*NP is the concentration of *p*-nitrophenol (mM),

Abs_{340nm} is the absorbance at 340 nm,

ϵ_{pNPG} is the the extinction coefficient at 340 nm of pNPG at pH 5.0 ($3.33 \text{ L mol}^{-1} \text{ cm}^{-1} \pm 0.04$)

pNPG is the initial concentration of pNPG (mM),

ϵ_{pNP} is the extinction coefficient at 340 nm of pNP at pH 5.0 ($5.96 \text{ L mol}^{-1} \text{ cm}^{-1} \pm 0.02$)

[00114] Three different incubations were done for each modified TrCel3A: 1) with *p*NPG alone, 2) with *p*NPG and 3 mM cellobiose, and 3) with *p*NPG and 5 mM glucose. The *p*NP concentration as a function of time in each of the three reactions was modeled according to

Equation 2 using a fourth order Runge-Kutta workbook in MS Excel and using the method of least squares.

$$\frac{dpNP}{dt} = \frac{k_{pNPG} \cdot E \cdot pNPG}{pNPG + K_{pNPG} \left(1 + \frac{G2}{K_{G2}} + \frac{G}{K_G} \right)} \quad \text{Equation 2}$$

where,

dpNP/dt is the rate of conversion of pNPG to pNP (mM/min),

k_{pNPG} is the catalytic rate constant for the conversion of pNPG to pNP ($\mu\text{mol}/\text{min}/\text{mg}$ protein),

E is the concentration of TrCel3A (mg/mL),

pNPG is the concentration of p-nitrophenyl-beta-D-glucopyranoside (mM),

K_{pNPG} is the Michaelis constant (or K_m) for pNPG (mM),

G2 is the concentration of cellobiose (mM),

K_{G2} is the Michaelis constant (or K_m) for cellobiose (mM),

G is the concentration of glucose (mM),

K_G is the glucose inhibition constant (mM).

[00115] The reaction scheme for this assay is shown in Figure 5. In this model, Cel3A hydrolyzes pNPG according to Michaelis-Menten kinetics. Cel3A activity is assumed to be inhibited competitively by glucose as described by the inhibition constant, K_G . Therefore, when glucose was added to a cuvette containing pNPG, the rate of pNPG catalysis decreased. The decrease in the rate of pNPG hydrolysis is accounted for by the K_G parameter. Modified TrCel3A beta-glucosidases with a higher K_G value are less affected by glucose, compared to the parental TrCel3A^{Wt}, and will have relatively higher rates of pNPG hydrolysis under these conditions. Similarly, when cellobiose was included in the reaction with pNPG, the rate of pNPG hydrolysis decreased. Modified TrCel3A beta-glucosidases with a lower K_{G2} value are more affected by the addition of cellobiose, compared to TrCel3A^{Wt}, and will have relatively lower rates of pNPG hydrolysis under these conditions.

[00116] The rates of pNPG hydrolysis were assayed for each modified TrCel3A in each of the three conditions, pNPG alone, pNPG + G2 and pNPG + G, by using a global fit of these three data sets to the parameters k_{pNPG} , K_{pNPG} , k_{cat} , K_{G2} and K_G in manner known by one of skill in the art. A K_G/K_{G2} ratio was also calculated using the values of K_G and K_{G2} from each global fit of the three data sets for each modified TrCel3A.

Each modified TrCel3A was assayed in this manner between two and five times. The average K_G , K_{G2} and K_G/K_{G2} values determined in this manner and their standard deviations are shown in Table 6. Student's t-test was used to identify modified TrCel3A beta-glucosidases with statistically significant improvements in K_G , K_{G2} and K_G/K_{G2} ($P < 0.05$) compared to TrCel3A^{Wt}. Graphs showing representative pNPG hydrolysis data and model fits for TrCel3A-WT (Figure 6), TrCel3A-V43I (Figure 7), TrCel3A-V101A (Figure 8), TrCel3A-F260I (Figure 9) and TrCel3A-I543N (Figure 10) are also shown.

[00117] The K_G values of TrCel3A-V43C (4.20 mM), TrCel3A-F260I (0.92 mM), TrCel3A-F260D (0.70 mM), TrCel3A-F260Q (0.69 mM), TrCel3A-F260V (0.72 mM) and TrCel3A-I543N (0.94 mM) were higher than that of TrCel3A^{WT} (0.58 mM). This indicates that the activity of each of these modified TrCel3A beta-glucosidases is significantly less inhibited by glucose and that they maintain relatively higher activity in the presence of glucose than does the parental TrCel3A beta-glucosidase. The values of K_{G2} of several modified TrCel3As with single amino acid substitutions, such as TrCel3A-V43I (0.62 mM) and TrCel3A-V101A (0.82 mM), were significantly (< 0.001 and 0.001 , respectively) lower than the K_{G2} of wild-type TrCel3A (1.18 mM). Similarly, most of the modified TrCel3As with different combinations of more than one amino acid substitution had lower K_{G2} than wild-type. These modified TrCel3A beta-glucosidases exhibit maximum activity at lower concentrations of cellobiose, indicating that they have a higher affinity for cellobiose. In a cellulose hydrolysis system, such as the conversion of cellulose to fermentable sugars utilizing cellulase such as that from *Trichoderma reesei* or other cellulolytic fungi, the use of a TrCel3A with a lower K_{G2} would contribute to lowering steady-state concentrations of cellobiose and lower product inhibition of cellulase enzymes.

[00118] The value of K_G , K_{G2} , K_G/K_{G2} and k_{G2} for each modified TrCel3A was divided by the value of the corresponding parameter for the parental TrCel3A in order to calculate the relative values shown in Table 2. These results show that the TrCel3A-V43I (1.41), TrCel3A-V101A (1.34), TrCel3A-F260I (1.52) and TrCel3A-I543N (1.35) beta-glucosidases have substantially improved K_G/K_{G2} ; improvements ranged from 35-52%, relative to TrCel3A^{Wt}.

Example 9: Measuring the Catalytic Rate Constant of Parental and Modified TrCel3A beta-glucosidases.

[00119] Initial rate assays were used to measure the catalytic rate constant (k_{cat}) of the parental and each modified TrCel3A beta-glucosidase on cellobiose. Purified wild-type parental and modified TrCel3A beta-glucosidases were incubated with 12 concentrations of cellobiose, ranging from 0.3 to 40 mM. The protein concentration in each of the reactions was 1 μ g/mL. Samples were incubated at 50°C for 15 min in deep well plates and then placed in a boiling water bath for 10 min to stop the reaction. The concentration of glucose produced at each concentration of substrate was measured as described in Example 7.

[00120] The rate of cellobiose consumption for the parental and each modified TrCel3A beta-glucosidase was plotted as a function of cellobiose concentration. As the cellobiose concentration increases from 0.4 mM to 10 mM, the reaction rate of wild-type TrCel3A increases until it reaches an apparent maximum reaction rate (Figure 12). Further increasing the substrate concentration results in a gradual decrease in the reaction rate, a phenomenon that is reportedly due to substrate inhibition (Cascalheira et al., 1999). As a result, data for the reaction rate as a function of cellobiose concentration were modeled using a modified form of the Michaelis-Menten equation which incorporates a K_{si} term for uncompetitive substrate inhibition (Equation 3).

$$\frac{dG2}{dt} = - \frac{\frac{k_{cat} \cdot E \cdot G2}{K_{G2}}}{1 + \frac{G2}{K_{G2}} + \frac{G2^2}{K_{G2} \cdot K_{si}}} \quad \text{Equation 3}$$

where,

$dG2/dt$ is the rate of conversion of cellobiose (G2) to two glucose molecules (2G) (mM G2 consumed/min),
 k_{cat} is the catalytic rate constant for the conversion of cellobiose to glucose (μ mol of G2 consumed/min/mg protein),
 E is the concentration of TrCel3A (mg/mL),
 $G2$ is the concentration of cellobiose (mM),
 K_{G2} is the Michaelis constant (or K_m) for cellobiose (mM),
 K_{si} is the cellobiose substrate inhibition constant (mM)

[00121] The k_{cat} is the TrCel3A rate constant on cellobiose and K_{si} is the parameter that describes the substrate inhibition. The data for the parental and each modified TrCel3A beta-glucosidase were fit to this model by non-linear regression using the method of least squares as

known to those of skill in the art. The parental and each modified TrCel3A beta-glucosidase were assayed in triplicate on three different occasions. The mean values of k_{cat} , K_{si} and their standard deviations are shown in Table 6.

[00122] The k_{cat} of several modified TrCel3As, including TrCel3A-F260I (11.06 $\mu\text{mol}/\text{min}/\text{mg}$), TrCel3A-I543S (9.76 $\mu\text{mol}/\text{min}/\text{mg}$) and TrCel3A-I543L (11.50 $\mu\text{mol}/\text{min}/\text{mg}$) was significantly higher than that of wild-type TrCel3A (8.92 $\mu\text{mol}/\text{min}/\text{mg}$) (Table 6 and Figure 12). Therefore, these modified beta-glucosidases catalyze the conversion of cellobiose to two glucose molecules at a faster rate than does wild-type TrCel3A. In a cellulose hydrolysis system, such as the conversion of cellulose to fermentable sugars utilizing cellulase such as that from *Trichoderma reesei*, the use of a TrCel3A with a higher k_{cat} would contribute to lowering steady-state concentrations of cellobiose and lower product inhibition of cellulase enzymes.

Table 6: Kinetic Parameters of Parental and Modified TrCel3A beta-glucosidases. The K_G , K_{G2} , K_G/K_{G2} , k_{cat} and K_{Si} parameters for parental (TrCel3A-Wt) and modified beta-glucosidases were determined as described in Examples 8 and 9. Significant improvements (P value < 0.05) in any of these parameters are shown in bold text.

TrCel3A-	K_G (mM)	P -value (K_G)	K_{G2} (mM)	P -value (K_{G2})	K_G/K_{G2}	P -value (K_G/K_{G2})	k_{cat} ($\mu\text{mol}/\text{min}/\text{mg}$)	P -value (k_{cat})	K_{Si} (mM)
V43I	0.42 $\pm$ 0.04	<0.001	0.65$\pm$0.07	<0.001	0.66$\pm$0.06	<0.001	6.73 $\pm$ 0.13	0.018	64.1
V43C	4.20$\pm$0.57	<0.001	2.39 $\pm$ 0.33	<0.001	1.79$\pm$0.38	<0.001	ND ^a	-	ND ^a
V101A	0.54 $\pm$ 0.18	0.593	0.82$\pm$0.16	0.001	0.65$\pm$0.09	0.006	7.22 $\pm$ 0.18	0.049	52.5
F260I	0.92$\pm$0.04	<0.001	1.23 $\pm$ 0.03	0.22	0.75$\pm$0.02	<0.001	11.06$\pm$0.29	0.017	60.8
F260D	0.70	-	1.31	-	0.53	-	ND ^a	-	ND ^a
F260Q	0.69	-	1.34	-	0.51	-	ND ^a	-	ND ^a
F260V	0.72	-	1.28	-	0.56	-	ND ^a	-	ND ^a
I543N	0.87$\pm$0.12	0.001	1.39 $\pm$ 0.21	0.064	0.64$\pm$0.11	0.019	6.53 $\pm$ 0.17	0.006	85.5
I543W	0.56 $\pm$ 0.08	0.635	0.95$\pm$0.08	0.003	0.59$\pm$0.06	0.041	4.53 $\pm$ 0.24	<0.01	42.3
I543S	0.63	-	1.08	-	0.59	-	9.76$\pm$0.15	<0.01	102.7
I543A	0.68	-	1.16	-	0.59	-	8.96 $\pm$ 0.19	0.944	94.5
I543G	0.54	-	1.02	-	0.53	-	8.02 $\pm$ 0.14	<0.01	98.7
I543L	0.61	-	1.13	-	0.55	-	11.50$\pm$0.12	<0.01	129.4
S72N- V101M- F260I	0.44 $\pm$ 0.03	<0.001	0.68$\pm$0.04	<0.001	0.65$\pm$0.05	0.019	8.45 $\pm$ 0.35	0.605	38.3
V43I-S72N- V101M	0.34 $\pm$ 0.08	0.001	0.46$\pm$0.04	<0.001	0.73$\pm$0.04	<0.001	7.55 $\pm$ 0.19	0.034	38.0
S72N- V101M- I543N	0.43 $\pm$ 0.05	0.001	0.64$\pm$0.04	<0.001	0.66$\pm$0.03	<0.001	7.73 $\pm$ 0.13	<0.01	40.8
S72N- V101M- I543D	0.36 $\pm$ 0.02	<0.001	0.53$\pm$0.04	<0.001	0.68$\pm$0.03	0.005	ND	-	ND
S72N-	0.42 $\pm$ 0.04	<0.001	0.69$\pm$0.06	<0.001	0.61$\pm$0.06	0.004	8.57 $\pm$ 0.30	0.638	35.0

V101M-I543L									
V43I-S72N-V101M-F260I	0.34±0.09	0.009	0.36±0.09	<0.001	0.93±0.08	<0.001	6.98±0.25	0.029	34.6
S72N-V101M-F260I-I543N	0.64±0.04	0.108	0.84±0.04	<0.001	0.77±0.08	0.004	9.12±0.14	0.580	31.8
V43I-S72N-V101M-F260I-I543N	0.41±0.02	<0.001	0.38±0.04	<0.001	1.10±0.11	0.001	6.14±0.04	<0.001	49.3
V43I-S72N-V101M-I543N	0.30±0.01	<0.001	0.32±0.03	<0.001	0.95±0.09	0.001	6.51±0.07	<0.001	51.7
S72N-V101M-F260I-I543L	0.50±0.03	0.010	0.66±0.06	<0.001	0.77±0.10	0.008	8.88±0.24	0.816	53.6
Wt (parental)	0.58 ±0.09	-	1.18 ± 0.12	-	0.49±0.06	-	8.92±0.29	-	66.1

^a ND=not determined

Example 10. Expression of modified Cel3A in *Trichoderma reesei*

10.1: *Trichoderma reesei* strains

[00123] Strain P59G is a genetically modified strain that produces and secretes high levels of the beta-glucosidase encoded by *T. reesei bglI* as described in U.S. Patent No. 6,015,703. The parent strain of P59G and modified Cel3A over-expressing transformant 4115A, is strain BTR213aux. The strain BTR213 is a derivative of RutC30 (ATCC #56765; Montenecourt and Eveleigh, 1979) produced by random mutagenesis and first selected for ability to produce larger clearing zones on minimal media agar containing 1% acid swollen cellulose and 4 g L⁻¹ 2-deoxyglucose and then selected for the ability to grow on lactose media containing 0.2 µg/mL carbendazim. A uridine auxotroph of BTR213, BTR213aux, was obtained through selection of mutants spontaneously resistant to 0.15% w/v 5-fluoroorotic-acid (FOA).

10.2: Generation of *T. reesei* transformation vector

[00124] The *T. reesei* expression vector was generated using pUC19 vector (Fermentas) as a backbone. To introduce spacers and cloning sites required for cloning of selection and expression cassettes two DNA fragments were amplified using pCAMBIA1301 plasmid (see URL: cambia.org/daisy/cambia/materials/vectors.525.html#dsy585) as a template and two pairs

of primers AC166/AC167 and AC168/AC169 (Table 7). The first fragment was cloned into the *EcoRI/SacI* sites of pUC19 introducing two new *PacI/AflIII* restriction sites. The second fragment was cloned into the *SacI/BamHI* sites introducing *NotI/MluI* restriction sites and generating pUC19-GDR vector.

[00125] For the construction of the TrCel3A expression cassette, a fragment containing the TrCel7A promoter and xylanase 2 secretion signal (Pcel7a-Xyn2ss fragment) was amplified using primers AC230 / AC231 (Table 7) and pC/XBG -TV vector (US 6,015,703) as template. The gene encoding TrCel3A-S72N-V101M-F260I (described in Example 6) was amplified using primers AC232 and AC233 (Table 7). The Pcel7a-Xyn2ss fragment was ligated to the TrCel3A-S72N-V101M-F260I encoding gene in two subsequent PCR reactions using primers AC230 and AC233 (Table 7) to produce the resulting c/xCel3A- S72N-V101M-F260I fragment. A fragment comprising the *cel6a* terminator (Tcel6a fragment) was amplified from the pC/XBG -TV (US 6,015,703) template using primers Tcel6a-F and Tcel6a-R (Table 7), which introduced *BamHI* / *KpnI* restriction sites, respectively. The c/xCel3A- S72N-V101M-F260I and Tcel6a fragments were cloned into pGEM-Teasy vectors generating vectors pGEM-c/xCel3A- S72N-V101M-F260I and pGEM-Tcel6a which were then digested with *MluI/KpnI* and *BamHI/KpnI* restriction enzymes to release the c/xCel3A- S72N-V101M-F260I and Tcel6a fragments, respectively. Both fragments were gel isolated and cloned by three fragment ligation into *MluI/ BamHI* sites of the pUC-GDR vector, generating pc/xCel3A- S72N-V101M-F260I. To eliminate the *SbfI* restriction site, pc/xCel3A- S72N-V101M-F260I was digested with *XbaI/ SphI* and the ends modified by removal of the 5' overhang and filling in of the 3' overhang. The linear plasmid was then ligated back together generating pc/xCel3A- S72N-V101M-F260I -Sbf. Next, a 2.2 kb fragment containing the *N. crassa* orotidine-5'-phosphate decarboxylase (*pyr4*) gene was amplified from pNcBgl (U.S. Patent No.6,939,704) containing *N. crassa pyr4* gene under control of its native promoter and terminator using primers AC323 and AC343 (Table 7). The *pyr4* cassette was cloned into pGEM-T-easy, digested with *PacI/NotI* restriction enzymes, gel purified and cloned into *PacI/NotI* sites of pc/xCel3A- S72N-V101M-F260I -Sbf generating final transformation vector pc/xCel3A-S72N-V101M-F260I-pyr4-TV (Figure 13).

Table 7. Primers used for construction of *T. reesei* transformation vector pc/xCel3A-AT012;F260I-pyr4-TV.

Primer name	Primer sequence
AC166	ACTGAATTCTTAATTAAGAACCGACGACTCGTCCGTC (SEQ ID NO: 57)
AC167	GTGGAGCTCCTTAAGGTGACATCGGCTTCAAATGGC (SEQ ID NO: 58)
AC168	GCAGAGCTCGCGGCCGCGAACCGACGACTCGTCCGTC (SEQ ID NO: 59)
AC169	CTGGGATCCGATATCACGCGTGTGACATCGGCTTCAAATGGC (SEQ ID NO: 60)
AC230	TTTACGCGTGATTATGGCGTACTAGAGAGCGG (SEQ ID NO: 61)
AC231	CTGCAGGAGGTACAACCTGGCGCTTCTCCACAGCCACGG (SEQ ID NO: 62)
AC232	GTGGAGAAGCGCCAGGTTGTACCTCCTGCAGGGACTCCATG (SEQ ID NO: 63)
AC233	TTTGGTACCCTACGCTACCGACAGAGTGCTCG (SEQ ID NO: 64)
AC323	TTTGCGGCCGCCATCATTCGTGCTTTTCGG (SEQ ID NO: 65)
AC343	TTCGATCGACTATACCACCACCCACCG (SEQ ID NO: 66)
Tcel6A-F	CTGGGTACCGGCTTTTCGTGACCGGGCTTC (SEQ ID NO: 67)
Tcel6A-R	CTGGGATCCGATGGACTAGTACAGCCATG (SEQ ID NO: 68)

10.3: *Trichoderma reesei* transformation and characterization of transformants

[00126] *Trichoderma* strain 4115A was generated by transformation of pc/xCel3A-S72N-V101M-F260I-pyr4-TV into strain BTR213aux by biolistic gold particle bombardment using the PDS-1000/He system (BioRad; E.I. DuPont de Nemours and Company). Gold particles (median diameter of 0.6 μ m, BioRad Cat. No. 1652262) were used as microcarriers. The following parameters were used in the optimization of the transformation: a rupture pressure of 1100 psi, a helium pressure of 29 mm Hg, a gap distance of 6 mm and a target distance of 6 cm. The spore suspension was prepared by washing *T. reesei* spores from PDA plates incubated for 4-5 days at 30°C with sterile water. About 5×10^7 spores were plated on 60 mm diameter plates containing minimal media agar (MM). After particle delivery, all transformation plates were incubated at 30°C for 5-10 days. All transformants were transferred to minimal media agar and incubated at 30°C.

Minimal medium (MM) agar:*

Component	Amount for 1L of medium
KH ₂ PO ₄	10 g
(NH ₄) ₂ SO ₄	6 g
Na ₃ Citrate•2H ₂ O	3 g
FeSO ₄ •7H ₂ O	5 mg
MnSO ₄ •H ₂ O	1.6 mg
ZnSO ₄ •7H ₂ O	1.4 mg
CaCl ₂ •2H ₂ O	2 mg
Agar	20 g
20% Glucose f.s.	50 ml
1 M MgSO ₄ •7H ₂ O f.s.	4 mL
pH to 5.5	

*MMSS agar contains the same components as MM agar plus 1.2 M sorbitol, 4 mM MgSO₄, 1 g/L YNB (Yeast Nitrogen Base w/o Amino Acids from DIFCO Cat. No.291940) and 0.12 g/L amino acids (-Ura DO Supplement from CLONTECH Cat. No.8601-1).

[00127] All *T. reesei* transformants were pre-screened for production of active modified TrCel3A beta-glucosidase using Esculin (β -D-glucose-6,7-dihydroxycoumarin) plate assay. An esculin stock solution was made by mixing 2 g of Esculin and 0.6 g FeCl₃ in 200 mL of deionized water. The mixture was heated until dissolved, cooled to approximately 40°C and filter sterilized. Transformants were plated on minimal media agar plates containing 1% cellobiose and grown for 3 to 4 days at 30°C. The esculin stock solution was diluted four-fold with 250 mM citrate buffer, pH 4.8 and 15 mL of the diluted esculin solution were overlaid onto the plates. Plates were incubated at 30°C for one hour. Positive transformants were selected according to formation of black precipitate formed around *T. reesei* colonies.

[00128] Individual *Trichoderma* colonies selected for the expression of active modified TrCel3A beta-glucosidases were transferred to potato dextrose agar (PDA) (Difco) plates and allowed to sporulate. At that time, about 10⁴-10⁶ spores of each individual transformant, parental and P59G strains were used to inoculate 1 mL of microculture medium in 24-well micro-plates.

Microculture medium

Component	g/L
(NH ₄) ₂ SO ₄	12.7
KH ₂ PO ₄	8.0
MgSO ₄ •7H ₂ O	4.0
CaCl ₂ •2H ₂ O	1.02
CSL	5.0
CaCO ₃	20.0
Carbon source**	30-35
Trace elements*	2mL/L
	pH 5.5

*Trace elements solution contains 5 g/L FeSO₄•7H₂O; 1.6 g/L MnSO₄•H₂O; 1.4 g/L ZnSO₄•7H₂O.

** glucose, Solka floc, lactose, cellobiose, sophorose, corn syrup, or Avicel. The carbon source can be sterilized separately as an aqueous solution at pH 2 to 7 and added to the remaining media initially or through the course of the fermentation.

[00129] The cultures were incubated at a temperature of 30°C with shaking at 250 rpm for 6 days. The biomass was separated from growth media containing the secreted protein by centrifugation at 12000 rpm. The protein concentration was determined using the Bio-Rad Protein Assay (Cat. No. 500-0001).

[00130] The concentration of Cel3A in supernatants from *Trichoderma reesei* microcultures was determined by ELISA (Figure 14). Filtrate and purified component standard were diluted 0.01-10 µg/mL (based on total protein) in phosphate-buffered saline, pH 7.2 (PBS) and incubated overnight at 4°C in microtitre plates (Costar EIA #9018). These plates were washed with PBS containing 0.1% Tween-20 (PBS/Tween) and then incubated in PBS containing 1% bovine serum albumin (PBS/BSA) for 1 hr at room temperature. Blocked microtitre wells were washed with PBS/Tween. Rabbit polyclonal antisera specific for TrCel3A was diluted (1:8,000) in PBS/BSA, added to separate microtitre plates and incubated for 2 h at room temperature. Plates were washed and incubated with a goat anti-rabbit antibody coupled to horseradish peroxidase (Sigma #A6154), diluted 1/2000 in PBS/BSA, for 1 h at room temperature. After washing, tetramethylbenzidine was added to each plate and incubated for 30 min at room temperature. The absorbance at 360 nm was measured in each well and converted into protein concentration using a TrCel3A standard curve.

10.4: *Trichoderma reesei* pilot fermentation

[00131] For 14 L pilot fermentations, *T. reesei* strains were grown on Potato Dextrose Agar at 28-30°C until a confluent lawn of spores was obtained. Spores were collected and used to inoculate 750 ml of Berkeley media (10 g/L glucose, 1.4 g/L (NH₄)₂SO₄, 2.0 g/L KH₂PO₄, 0.31 g/L MgSO₄•7H₂O, 0.53 g/L CaCl₂, 5.1 g/L dry corn steep, 5 mg/L FeSO₄•7H₂O, 0.8 mg/L MnSO₄•H₂O, 0.7 mg/L ZnSO₄•7H₂O) in a 2 L baffled flask. After 3 days of growth at 28°C and 150 rpm, this culture was used to inoculate 10 L of fermentation medium with the following initial composition: 13 g/l glucose, 2.2 g/l (NH₄)₂SO₄, 1.39 g/l KH₂PO₄, 0.7 g/l MgSO₄•7H₂O, 0.185 g/l CaCl₂, 6 g/l dry corn steep, 1.75 mg/l FeSO₄•7H₂O, 0.56 mg/l MnSO₄•H₂O, 0.49 g/l ZnSO₄•7H₂O. A fed-batch aerobic fermentation using an inducing carbohydrate source was run for 6 days at pH 5 and 28°C in a 14 L New Brunswick Microferm fermentor. After 6 days, the culture was filtered over Harborlite 1500S and the culture filtrate saved.

[00132] The concentration of parental or modified TrCel3A beta-glucosidases in fermentation filtrate from *Trichoderma reesei* was determined by ELISA as described above (Example 10.3).

Example 11: Measuring the cellulose hydrolysis activity of a whole cellulase secreted by a strain of *Trichoderma* that expresses high levels of a modified beta-glucosidase.

[00133] A whole enzyme produced by a strain of *Trichoderma* that expresses high concentrations of a modified beta-glucosidase, TrCel3A-S72N-V101M-F260I, was compared to that from a strain that expresses similar concentrations of the parental, wild-type TrCel3A in an extended hydrolysis time course assay on a lignocellulosic substrate. The concentrations of the wild-type and the modified beta-glucosidase in their respective whole cellulase mixtures were 31.1±1.7% and 28.2±1.3% of total protein, respectively. Whole *Trichoderma* cellulase mixtures containing these TrCel3A beta-glucosidases were incubated with pretreated wheat straw at a concentration of 25 g/L cellulose at a dose of 10 mg of total cellulase mixture per gram of cellulose. Triplicate assays were performed for each cellulase mixture under the same conditions. The hydrolysis assays were buffered in 50 mM citrate, pH 5.0 containing 0.1% sodium benzoate. The assay was conducted at 50°C for 96 hr with continuous orbital shaking.

Aliquots of 0.7 mL were taken at various time points from each flask and the glucose concentration in the soluble portion was assayed and converted into a measure of fractional cellulose conversion. The conversion data were then fit with a rectangular hyperbola with an additional linear term using minimization of the sum of squared residuals of fit. The equation was of the following form: $\text{conversion} = (\text{max} \cdot \text{time}) / (\text{halfmax} + \text{time}) + c \cdot \text{time}$. Both sets of data were fit globally with unique max and halfmax values and a shared value of the variable c.

[00134] The results are shown in Figure 15. This figure demonstrates that the fractional cellulose conversion measured for the cellulase mixture containing the modified TrCel3A beta-glucosidase was higher at 96 hr (0.99 ± 0.03) than for the cellulase mixture containing the wild-type TrCel3A beta-glucosidase (0.89 ± 0.02). This increase in fractional cellulose conversion was statistically significant ($P < 0.05$, Student's T-Test).

References

- Berghem, L.E. and Pettersson, L.G. (1974) The mechanism of enzymatic cellulose degradation. Isolation and some properties of a beta-glucosidase from *Trichoderma viride*. *European Journal of Biochemistry*, 46(2):295-305.
- Bhatia, Y., Mishra, S. and Bisaria, V.S. (2002) Microbial beta-Glucosidases: Cloning, Properties and Applications. *Critical Reviews in Biotechnology*, 22(4):375-407)
- Bradford, M.M. (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72:248-254.
- Butler, T. and Alcalde, M. (2003) In *Methods in Molecular Biology*, vol. 231: (F. H. Arnold and G. Georgiou, editors), Humana Press Inc. Totowa (New Jersey), pages 17-22.
- Cascalheira, J.F., Queiroz, J.A. (1999) Kinetic study of the cellobiase activity of *Trichoderma reesei* cellulase complex at high substrate concentrations. *Biotechnology Letters*, 21(8):651-655
- Chen, H., Li, X. and Ljungdahl, L.G. (1994) Isolation and properties of an extracellular beta-glucosidase from the polycentric rumen fungus *Orpinomyces* sp. strain PC-2. *Applied and Environmental Microbiology*, 60(1):64-70.
- Chirico, W.J. and Brown, R.D. Jr. (1987) Purification and characterization of a beta-glucosidase from *Trichoderma reesei*. *European Journal of Biochemistry*, 165(2):333-341.
- Christakopoulos, P., Goodenough, P.W., Kekos, D., Macris, B.J., Claeysens, M. and Bhat, M.K. (1994) Purification and characterization of an extracellular beta-glucosidase with transglycosylation and exo-glucosidase activities from *Fusarium oxysporum*. *European Journal of Biochemistry*, 224(2):279-385.
- Eijsink VG, Gaseidnes C., Borchert TV, van den Burg B. 2005. Directed Evolution of Enzyme Stability. *Biomol. Eng.* 22:21-30

- Enari, T.-M., Niku-Paavola, M.-L., Harju, L., Lappalainen, A. and Nummi, M. (1981) Purification of *Trichoderma reesei* and *Aspergillus niger* beta-glucosidase. *Journal of Applied Biochemistry*, 3:157-163.
- Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.J., Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C., Olivares, H.A., Teunissen, P.J., Yao, H. and Ward, M. (2003) Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *Journal of Biological Chemistry*, 278(34):31988-31997.
- Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by the Liac/ss carrier DNA/PEG method. In *Methods in Enzymology*, 350:87-96.
- Gueguen, Y., Chemardin, P., Arnaud, A. and Galzy, P. (1995) Purification and characterization of an intracellular beta-glucosidase from *Botrytis cinerea*. *Enzyme and Microbial Technology*, 78:900-906.
- Henrissat, B. (1991) A classification of glycosyl-dhydrolases based on amino acid sequence similarities. *Biochemical Journal*, 293:781-788;
- Henrissat, B. (1994) Cellulases and their interaction with cellulose. *Cellulose*, 1:169–196.
- Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal*, 316:695-696
- Holtzapple, M., Cognata, M., Shu, Y., and Hendrickson, C. (1990) Inhibition of *Trichoderma reesei* cellulase by sugars and solvents. *Biotechnology and Bioengineering*, 36:275.
- Knowles, J., Lehtovaara, P. and Teeri, T. Cellulase families and their genes. (1987) *Trends in Biotechnology*, 5:255-261.
- Li, X.L. and Calza, R.E. (1991) Purification and characterization of an extracellular beta-glucosidase from the Rumen Fungus *Neocallimastix frontalis* EB188. *Enzyme and Microbial Technology*, 13:622-628.

- Lynd, L.R., Weimer, P.J., van Zyl, W.H. and Pretorius I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, 66:506-577.
- Perez-Pons, J.A., Cayetano, A., Rebordosa, X., Lloberas, J., Guasch, A. and Querol, E. (1994) A beta-glucosidase gene (bgl3) from *Streptomyces* sp. strain QM-B814. Molecular cloning, nucleotide sequence, purification and characterization of the encoded enzyme, a new member of family 1 glycosyl hydrolases. *European Journal of Biochemistry*, 223(2):557-565.
- Riou, C., Salmon, J.M., Vallier, M.J., Günata, Z. and Barre, P. (1998) Purification, characterization and substrate specificity of a novel highly glucose-tolerant beta-glucosidase from *Aspergillus oryzae*. *Applied and Environmental Microbiology*, 64(10):3607-3614.
- Saha, B.C. and Bothast, R.J. (1996) Production, purification and characterization of a highly glucose-tolerant novel beta-glucosidase from *Candida peltata*. *Applied and Environmental Microbiology*, 62(9):3165-3170.
- Teeri, T.T. (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in Biotechnology*, 15(5):160-167.
- Teleman, A., Koivula, A., Reinikainen, T., Valkeajärvi, A., Teeri, T.T., Drakenberg, T., and Teleman, O. (1995) Progress-curve analysis shows that glucose inhibits the cellotriose hydrolysis catalysed by cellobiohydrolase II from *Trichoderma reesei*. *European Journal of Biochemistry*, 231:250.
- Trinder, P. (1969) Determination of glucose in blood using glucose oxidase with an alternative oxygen acceptor. *Annals of Clinical Biochemistry*, 6:24-27.
- Varghese, J.N., Hrmova, M., Fincher, G.B. (1999) Three-dimensional structure of a barley beta-D-glucan exohydrolase, a family 3 glycosyl hydrolase. *Structure Fold.Des.* 7: 179-190
- Wood, T.M. and Garcia-Campayo, V. (1990) Enzymology of cellulose degradation. *Biodegradation*, 1:147-161.
- Zhang, Y.H. and Lynd, L.R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose. *Biotechnology and Bioengineering*, 88(7):797-824.

EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A modified *Trichoderma reesei* TrCel3A beta-glucosidase comprising one or more amino acid substitutions selected from the group consisting of V43X, V101X, F260X and I543X, the modified TrCel3A beta-glucosidase comprising an amino acid sequence which is from about 80% to about 99.9% identical to SEQ ID NO: 1 or 116.
2. The modified *Trichoderma reesei* TrCel3A beta-glucosidase of claim 1, comprising an amino acid sequence which from about 90% to about 99.9% identical to SEQ ID NO: 1 or 116.
3. The modified *Trichoderma reesei* TrCel3A beta-glucosidase of claim 1 or 2, wherein the one or more of the amino acid substitutions is selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543W, I543A, I543S, I543G, and I543L.
4. The modified *Trichoderma reesei* TrCel3A beta-glucosidase of any one of claims 1 to 3, wherein the modified *Trichoderma reesei* TrCel3A beta-glucosidase exhibits:
 - a. at least a 20% increase in K_G ,
 - b. at least a 20% decrease in K_{G2} , or
 - c. at least a 10% increase in k_{cat}

relative to the K_G , K_{G2} , or k_{cat} of a parental *Trichoderma reesei* TrCel3A from which the modified *Trichoderma reesei* TrCel3A is derived.

5. The modified *Trichoderma reesei* TrCel3A beta-glucosidase of claim 4, wherein the modified *Trichoderma reesei* TrCel3A beta-glucosidase exhibits:
 - a. from about a 30% increase in K_G ,
 - b. from about a 30% decrease in K_{G2} , or
 - c. from about a 20% increase in k_{cat}

relative to the K_G , K_{G2} , or k_{cat} of a parental *Trichoderma reesei* TrCel3A from which the modified *Trichoderma reesei* TrCel3A is derived.

6. The modified *Trichoderma reesei* TrCel3A beta-glucosidase of any one of claims 1 to 5, further comprising one or more amino acid substitutions selected from the group consisting of V66X, S72X, F96X, T235X, N248X, N369X and A386X.

7. An isolated genetic construct comprising a nucleic acid sequence encoding a modified *Trichoderma reesei* TrCel3A beta-glucosidase comprising one or more of the amino acid substitutions selected from the group consisting of V43X, V101X, F260X and I543X, wherein the amino acid sequence of the modified TrCel3A beta-glucosidase encoded by the construct is from about 80% to about 99.9% identical to SEQ ID NO: 1.

8. An isolated genetic construct comprising a nucleic acid encoding a modified *Trichoderma reesei* TrCel3A beta-glucosidase comprising one or more of the amino acid substitutions selected from the group consisting of V43X, V101X, F260X and I543X, wherein the amino acid sequence of the modified TrCel3A beta-glucosidase encoded by the construct is from about 80% to about 99.9% identical to SEQ ID NO: 1, the modified *Trichoderma reesei* TrCel3A beta-glucosidase exhibiting:

- a. at least a 20% increase in K_G ,
- b. at least a 20% decrease in K_{G2} , or
- c. at least a 10% increase in k_{cat}

relative to a parental *Trichoderma reesei* TrCel3A from which the modified *Trichoderma reesei* TrCel3A is derived.

9. The isolated genetic construct of claim 8, wherein the modified TrCel3A beta-glucosidase encoded by the construct exhibits:

- a. at least a 30% increase in K_G ,
- b. at least a 30% decrease in K_{G2} , or

c. at least a 20% increase in k_{cat}

relative to a parental *Trichoderma reesei* TrCel3A from which the modified *Trichoderma reesei* TrCel3A is derived.

10. An isolated genetically modified microbe comprising the genetic construct of any one of claims 7 to 9.

11. The isolated genetically modified microbe of claim 10, wherein said microbe is a species of yeast or filamentous fungus.

12. A process for producing a modified *Trichoderma reesei* TrCel3A beta-glucosidase, comprising the steps of (i) providing a genetically modified microbe comprising the genetic construct of any one of claims 7 to 9; (ii) culturing the genetically modified microbe in submerged liquid fermentations under conditions that induce the expression of the modified *Trichoderma reesei* TrCel3A beta-glucosidase; and (iii) recovering the modified *Trichoderma reesei* TrCel3A beta-glucosidase.

13. A process for enzymatic hydrolysis of a cellulose substrate comprising contacting the substrate with one or more cellulase enzymes and the modified *Trichoderma reesei* TrCel3A beta-glucosidase of any one of claims 1 to 6.

14. The process of claim 13, wherein the cellulose substrate is a pretreated lignocellulosic feedstock and wherein the process produces fermentable sugars.

15. The process of claim 14, wherein the pretreated lignocellulose feedstock is selected from the group consisting of corn stover, wheat straw, barley straw, rice straw, oat straw, canola straw, sugarcane straw, soybean stover, corn fiber, sugar beet pulp, pulp mill fines and rejects, sugar cane bagasse, hardwood, softwood, sawdust, switch grass, miscanthus, cord grass, and reed canary grass.

16. A modified *Trichoderma reesei* TrCel3A beta-glucosidase comprising the amino acid sequence of one of:

SEQ ID NO: 2 (TrCel3A-V43I);
SEQ ID NO: 3 (TrCel3A-V43C);
SEQ ID NO: 4 (TrCel3A-V101A);
SEQ ID NO: 5 (TrCel3A-V101G);
SEQ ID NO: 6 (TrCel3A-F260I);
SEQ ID NO: 7 (TrCel3A-F260V);
SEQ ID NO: 8 (TrCel3A-F260Q);
SEQ ID NO: 9 (TrCel3A-F260D);
SEQ ID NO: 10 (TrCel3A-I543N);
SEQ ID NO: 11 (TrCel3A-I543W);
SEQ ID NO: 12 (TrCel3A-I543A);
SEQ ID NO: 13 (TrCel3A-I543S);
SEQ ID NO: 14 (TrCel3A-I543G);
SEQ ID NO: 15 (TrCel3A-I543L);
SEQ ID NO: 16 (TrCel3A-S72N-V101M-F260I);
SEQ ID NO: 17 (TrCel3A-V43I-S72N-V101M);
SEQ ID NO: 18 (TrCel3A-S72N-V101M-I543N);
SEQ ID NO: 19 (TrCel3A-S72N-V101M-I543D);
SEQ ID NO: 20 (TrCel3A-S72N-V101M-I543L);
SEQ ID NO: 21 (TrCel3A-V43I-S72N-V101M-F260I);
SEQ ID NO: 22 (TrCel3A-V43I-S72N-V101M-F260I-I543N);
SEQ ID NO: 23 (TrCel3A-V43I-S72E-V101M-I543N);
SEQ ID NO: 69 (TrCel3A-S72N-V101M-F260I-I543N); or
SEQ ID NO: 70 (TrCel3A-S72N-V101M-F260I-I543L).

17. A modified Family 3 beta-glycosidase comprising one or more of the amino acid substitutions at a position selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543W, I543A, I543S, I543G, and I543L, the position determined from alignment of a parental Family 3 beta-glycosidase with the TrCel3A amino acid sequence as defined in SEQ ID NO: 1, wherein the amino acid sequence of the modified Family 3 beta-glycosidase comprises a sequence which is from about 80% to about 99.9% identical to an

amino acid sequence of a parental Family 3 beta-glycosidase from which the modified Family 3 beta-glycosidase is derived.

18. The modified Family 3 beta-glycosidase of claim 17, wherein the amino acid sequence of the modified Family 3 beta-glycosidase is from about 80% to about 99.9% identical to an amino acid sequence of a parental Family 3 beta-glycosidase from which the modified Family 3 beta-glycosidase is derived, the modified Family 3 beta-glycosidase exhibiting

- a. an increase in K_p ,
- b. a decrease in K_s , or
- c. an increase in k_{cat} .

relative to the K_p , K_s , or k_{cat} of a parental Family 3 beta-glycosidase from which the modified Family 3 beta-glycosidase is derived.

19. A modified *Trichoderma reesei* TrCel3A beta-glucosidase comprising the amino acid sequence of one of:

SEQ ID NO: 117 (TrCel3A-V43I full);
SEQ ID NO: 118 (TrCel3A-V43C full);
SEQ ID NO: 119 (TrCel3A-V101A full);
SEQ ID NO: 120 (TrCel3A-V101G full);
SEQ ID NO: 121 (TrCel3A-F260I full);
SEQ ID NO: 122 (TrCel3A-F260V full);
SEQ ID NO: 123 (TrCel3A-F260Q full);
SEQ ID NO: 124 (TrCel3A-F260D full);
SEQ ID NO: 125 (TrCel3A-I543N full);
SEQ ID NO: 126 (TrCel3A-I543W full);
SEQ ID NO: 127 (TrCel3A-I543A full);
SEQ ID NO: 128 (TrCel3A-I543S full);
SEQ ID NO: 129 (TrCel3A-I543G full);

SEQ ID NO: 130 (TrCel3A-I543L full);
SEQ ID NO: 131 (TrCel3A-S72N-V101M-F260I full);
SEQ ID NO: 132 (TrCel3A-V43I-S72N-V101M full);
SEQ ID NO: 133 (TrCel3A-S72N-V101M-I543N full);
SEQ ID NO: 134 (TrCel3A-S72N-V101M-I543D full);
SEQ ID NO: 135 (TrCel3A-S72N-V101M-I543L full);
SEQ ID NO: 136 (TrCel3A-V43I-S72N-V101M-F260I full);
SEQ ID NO: 137 (TrCel3A-V43I-S72N-V101M-F260I-I543N full);
SEQ ID NO: 138 (TrCel3A-V43I-S72E-V101M-I543N full);
SEQ ID NO: 139 (TrCel3A-S72N-V101M-F260I-I543N full); or
SEQ ID NO: 140 (TrCel3A-S72N-V101M-F260I-I543L full).

[0012] An embodiment of the invention relates to a modified beta-glucosidase of *Trichoderma reesei* produced by substitution of the amino acid at one or more of positions 43, 101, 260 and 543 in the beta-glucosidase I or TrCel3A sequence (SEQ ID NO: 116) and comprising an amino acid sequence that is from about 80% to 99.9% to that TrCel3A amino acid sequence defined by SEQ ID NO: 1 or 116.

[0013] The modified TrCel3A beta-glucosidase may be derived from a parental TrCel3A beta-glucosidase that is otherwise identical to the modified TrCel3A beta-glucosidase and includes the substitution of the naturally occurring amino acid at one or more of positions 43, 101, 260 and 543. For example, the modified TrCel3A beta-glucosidase may contain one or more amino acid substitutions at positions other than at positions 43, 101, 260 and 543, provided that the amino acid sequence of the modified TrCel3A is from about 80% to about 99.9% identical to the TrCel3A amino acid sequence (SEQ ID NO: 1 or 116). For example, this invention includes the modified TrCel3A as defined above and further comprising an amino acid substitution at one or more of positions 66, 72, 96, 235, 248 and 369.

[0014] The present invention also relates to a modified TrCel3A beta-glucosidase comprising an amino acid sequence that is from about 80% to about 99.9% identical to that of the wild-type TrCel3A of SEQ ID NO: 1 or 116 and which exhibits (a) at least a 20% increase in the K_G , (b) at least a 20% decrease in K_{G2} , or (c) at least a 10% increase in k_{cat} for cellobiose relative to the K_G , K_{G2} and/or k_{cat} of a parental TrCel3A beta-glucosidase from which is derived.

[0015] The present invention also relates to a modified TrCel3A consisting of the amino acid sequence defined by:

SEQ ID NO: 2 or SEQ ID NO: 117;
SEQ ID NO: 3 or SEQ ID NO: 118;
SEQ ID NO: 4 or SEQ ID NO: 119;
SEQ ID NO: 5 or SEQ ID NO: 120;
SEQ ID NO: 6 or SEQ ID NO: 121;
SEQ ID NO: 7 or SEQ ID NO: 122;
SEQ ID NO: 8 or SEQ ID NO: 123;
SEQ ID NO: 9 or SEQ ID NO: 124;

SEQ ID NO: 10 or SEQ ID NO: 125;
SEQ ID NO: 11 or SEQ ID NO: 126;
SEQ ID NO: 12 or SEQ ID NO: 127;
SEQ ID NO: 13 or SEQ ID NO: 128;
SEQ ID NO: 14 or SEQ ID NO: 129;
SEQ ID NO: 15 or SEQ ID NO: 130;
SEQ ID NO: 16 or SEQ ID NO: 131;
SEQ ID NO: 17 or SEQ ID NO: 132;
SEQ ID NO: 18 or SEQ ID NO: 133;
SEQ ID NO: 19 or SEQ ID NO: 134;
SEQ ID NO: 20 or SEQ ID NO: 135;
SEQ ID NO: 21 or SEQ ID NO: 136;
SEQ ID NO: 22 or SEQ ID NO: 137;
SEQ ID NO: 23 or SEQ ID NO: 138;
SEQ ID NO: 69 or SEQ ID NO: 139; or
SEQ ID NO: 70 or SEQ ID NO: 140.

[0016] The genetic constructs of the present invention comprise a nucleic acid sequence encoding a modified TrCel3A with an amino acid sequence that is from about 80% to about 99.9% amino acid sequence identity to SEQ ID NO: 1 or 116 and that comprises an amino acid substitution at one or more of positions 43, 101, 260 and 543, which nucleic acid sequence is operably linked to nucleic acid sequences regulating its expression and secretion from a host microbe. For example, the nucleic acid sequences regulating the expression and secretion of the modified TrCel3A beta-glucosidase may be derived from the host microbe used for expression of the modified TrCel3A beta-glucosidase. The host microbe may be a yeast, such as *Saccharomyces cerevisiae*, or a filamentous fungus, such as *Trichoderma reesei*.

[0017] The invention also relates to a genetic construct as defined above, wherein the modified TrCel3A beta-glucosidase comprises an amino acid sequence that is from about 90% to about 99.9% identical to SEQ ID NO: 1 or 116. The modified TrCel3A beta-glucosidase may further comprise substitutions at one or more of positions 66, 72, 96, 235, 248, 369 and 386 or any other additional mutations.

[0041] Beta-glucosidases are a subset of beta-glycosidases belonging to glycoside hydrolase (GH) Families 1 and 3, using the classification system developed by Henrissat and coworkers (Henrissat, B. (1991); Henrissat, B. and Bairoch, A. (1996)). There are currently over 115 GH families that have been identified using this classification system, which are listed in the database of Carbohydrate Active Enzymes (CAZy) (see <http://afmb.cnrs-mrs.fr/CAZY/index.html> for reference). Family 1 comprises beta-glycosidases from archaeobacteria, plants and animals. Beta-glycosidases from some bacteria, mold and yeast belong to Family 3. For the purpose of this invention, a "beta-glycosidase" is therefore defined as any protein that is categorized as a Family 3 glycoside hydrolase according to the CAZy system.

[0042] The three dimensional structure of beta-D-glucan exohydrolase, a Family 3 glycosyl hydrolase, was described by Varghese et al. (1999). The structure was of a two domain globular protein comprising a N-terminal $(\alpha/\beta)_8$ TIM-barrel domain and a C-terminal a six-stranded beta-sandwich, which contains a beta-sheet of five parallel beta-strands and one antiparallel beta-strand, with three alpha-helices on either side of the sheet. This structure is likely shared by other Family 3 enzymes.

[0043] As shown in Figure 11, the primary amino acid sequence of Family 3 beta-glucosidases show a high degree of similarity. Multiple alignment across 45 Family 3 beta-glucosidase amino acid sequences shows that the most naturally occurring Family 3 beta-glucosidases of fungal origin show from about 40% to about 100% amino acid sequence identity to the amino acid sequence of TrCel3A (Figure 11). In particular, there are several regions of very high amino acid sequence conservation within the Family 3 beta-glucosidases including, for example, from amino acids 225-256 and 439-459, containing the catalytic amino acids D236 and E447, respectively.

[0044] By "TrCel3A beta-glucosidase" or "TrCel3A" it is meant the Family 3 glycosyl hydrolase produced by *Trichoderma reesei* defined by the amino acid sequence of SEQ ID NO: 116. TrCel3A beta-glucosidase is also known as *Trichoderma reesei* beta-glucosidase or BGL1. By "native" or "wild type" TrCel3A (also annotated as TrCel3A^{wt}), it is meant the TrCel3A of SEQ ID NO: 116 without any amino acid substitutions.

[0045] By “modified TrCel3A beta-glucosidase” or “modified TrCel3A”, it is meant a TrCel3A beta-glucosidase which comprises one or more of the amino acid substitutions, introduced by genetic engineering techniques, selected from the group consisting of V43X, V101X, F260X, and I543X. For example, the modified modified TrCel3A beta-glucosidase may comprising one or more amino acid substitutions selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543A, I543S, I543G and I543L.

[0046] Genetic engineering techniques for altering amino acid sequences include, but are not limited to, site-directed mutagenesis, cassette mutagenesis, random mutagenesis, synthetic oligonucleotide construction, cloning and other genetic engineering techniques as would be known by those of skill in the art (Eijsink VG, et al. 2005). Modified TrCel3A beta-glucosidases of the present invention include those comprising amino acid substitutions at any one of V43X, V101X, F260X and I543X, at any two of V43X, V101X, F260X and I543X, any three of V43X, V101X, F260X and I543X, or all four of V43X, V101X, F260X and I543X.

[0047] It will be understood that the modified TrCel3A beta-glucosidase may be derived from wild-type TrCel3A beta-glucosidase of SEQ ID NO: 116 or from a TrCel3A beta-glucosidase that contains other amino acid substitutions. For example, the modified TrCel3A beta-glucosidase may contain amino acid substitution at one or more of positions 66, 73, 96, 235, 248, and 369. Alternatively, after production of the modified TrCel3A beta-glucosidase comprising mutations at one or more of positions 43, 101, 260 and 543, it may be subsequently further modified to contain additional amino acid substitutions, including but not limited to those set forth above.

[0048] As used herein in respect of modified TrCel3A beta-glucosidase amino acid sequences, “derived from” refers to the isolation of a target nucleic acid sequence element encoding the desired modified TrCel3A beta-glucosidase using genetic material or nucleic acid or amino acid sequence information specific to the parental TrCel3A beta-glucosidase. As is known by one of skill in the art, such material or sequence information can be used to generate a nucleic acid sequence encoding the desired modified TrCel3A beta-glucosidase using one or more molecular biology techniques including, but not limited to, cloning, sub-cloning, amplification by PCR, *in vitro* synthesis, and the like.

[0049] In one embodiment of the invention, the amino acid sequence of the modified TrCel3A beta-glucosidase is from about 80% to about 99.9% identical to SEQ ID NO: 1 or 116. For example, the amino acid sequence of the modified TrCel3A beta-glucosidase may be from about 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 99.9% identical to SEQ ID NO: 1 or 116. In other words, the number of amino acid substitutions in the modified TrCel3A beta-glucosidase may not exceed 20% of the total number amino acids in the parental TrCel3A beta-glucosidase sequence.

[0050] In another embodiment of the invention, the amino acid sequence of modified TrCel3A beta-glucosidase may be from about 90% to about 99.9% identical to SEQ ID NO: 1. For example, the amino acid sequence of the modified TrCel3A beta-glucosidase may be from about 95% to about 100% identical to SEQ ID NO: 1 or 116.

[0051] In another embodiment, the amino acid sequence of the modified TrCel3A beta-glucosidase may be from about 80% to about 99.9% identical to SEQ ID NO: 1 or 116 and the modified TrCel3A beta-glucosidase may exhibit (a) at least about a 20% increase in K_G , (b) at least about a 20% decrease in K_{G2} , or (c) at least about a 10% increase in k_{cat} for cellobiose relative to the K_G , K_{G2} and/or k_{cat} of a parental TrCel3A beta-glucosidase from which it is derived. For example, the modified TrCel3A beta-glucosidase may exhibit (a) from about a 20% to about a 800% increase in K_G , or any increase therebetween, (b) from about a 20% to about an 80% decrease in K_{G2} , any decrease therebetween, or (c) from about a 10% to about a 50% increase in k_{cat} for cellobiose relative to the K_G , K_{G2} and/or k_{cat} of a parental TrCel3A beta-glucosidase from which is derived.

[0052] By "parental TrCel3A beta-glucosidase" or "parental TrCel3A", it is meant a TrCel3A beta-glucosidase that does not contain a substitution of its original amino acid(s) at positions 43, 101, 260 or 543. For example, the parental TrCel3A beta-glucosidase may comprise amino acid substitutions at one or more of positions 66, 72, 96, 235, 248, and 369.

[0053] In order to assist one of skill in the art regarding those amino acid positions of the TrCel3A beta-glucosidase at which amino acid substitutions (other than V43X, V101X, F260X, and I543X) may be made and produce an active beta-glucosidase, an alignment of 45 Family 3 beta-glucosidases derived from fungal sources along with a consensus beta-glucosidase sequence consisting of the amino acids that naturally occur with the highest frequency at each position is provided in Figure 11 along with a graph showing the frequency

of occurrence of each amino acid of the consensus sequence at each position. Using the information provided in Figure 11, one of skill in the art would recognize regions of low sequence conservation to other Family 3 beta-glucosidases and choose such regions for introduction of amino acid substitutions that are not likely to compromise significantly the function of the enzyme. Non-limiting examples of such regions include, for example, the regions between positions 1-20, 303-323 and 403-414 and select amino acid positions within these regions.

[0054] As described in more detail herein, several modified TrCel3A beta-glucosidases have been prepared that exhibit (a) at least a 20% increase in K_G , (b) at least a 20% decrease in K_{G2} , or (c) at least a 10% increase in k_{cat} for cellobiose relative to the K_G , K_{G2} and/or k_{cat} of a parental TrCel3A beta-glucosidase from which is derived. A list of several modified TrCel3A beta-glucosidases, which is not to be considered limiting in any manner, is presented in Table 1.

Table 1: TrCel3A beta-glucosidases with improved catalytic efficiency

Modified TrCel3A beta-glucosidase	SEQ ID NO:	Modified TrCel3A beta-glucosidase	SEQ ID NO:
TrCel3A-V43I	2	TrCel3A-V43I full	117
TrCel3A-V43C	3	TrCel3A-V43C full	118
TrCel3A-V101A	4	TrCel3A-V101A full	119
TrCel3A-V101G	5	TrCel3A-V101G full	120
TrCel3A-F260I	6	TrCel3A-F260I full	121
TrCel3A-F260V	7	TrCel3A-F260V full	122
TrCel3A-F260Q	8	TrCel3A-F260Q full	123
TrCel3A-F260D	9	TrCel3A-F260D full	124
TrCel3A-I543N	10	TrCel3A-I543N full	125
TrCel3A-I543W	11	TrCel3A-I543W full	126
TrCel3A-I543A	12	TrCel3A-I543A full	127
TrCel3A-I543S	13	TrCel3A-I543S full	128
TrCel3A-I543G	14	TrCel3A-I543G full	129
TrCel3A-I543L	15	TrCel3A-I543L full	130
TrCel3A-S72N-V101M-F260I	16	TrCel3A-S72N-V101M-F260I full	131
TrCel3A-V43I-S72N-V101M	17	TrCel3A-V43I-S72N-V101M full	132
TrCel3A-S72N-V101M-I543N	18	TrCel3A-S72N-V101M-I543N full	133
TrCel3A-S72N-V101M-I543D	19	TrCel3A-S72N-V101M-I543D full	134

TrCel3A-S72N-V101M-I543L	20	TrCel3A-S72N-V101M-I543L full	135
TrCel3A-V43I-S72N-V101M-F260I	21	TrCel3A-V43I-S72N-V101M-F260I full	136
TrCel3A-V43I-S72N-V101M-F260I-I543N	22	TrCel3A-V43I-S72N-V101M-F260I-I543N full	137
TrCel3A-V43I-S72N-V101M-I543N	23	TrCel3A-V43I-S72N-V101M-I543N full	138
TrCel3A-S72N-V101M-F260I-I543N	69	TrCel3A-S72N-V101M-F260I-I543N full	139
TrCel3A-S72N-V101M-F260I-I543L	70	TrCel3A-S72N-V101M-F260I-I543L full	140

Modified TrCel3A beta-glucosidases Improved Kinetic Parameters

[0055] The modified TrCel3A beta-glucosidases of the present invention exhibit improvements in at least one of the following kinetic parameters: K_G , K_{G2} and k_{cat} . K_G is defined as the concentration of glucose which reduces the enzymatic activity of the beta-glucosidase by 50%. K_{G2} is defined as the concentration of cellobiose at which the beta-glucosidase exhibits half its maximal rate. The k_{cat} is the catalytic rate constant for the hydrolysis of cellobiose. Example 8 details an assay for measuring the K_G and K_{G2} of parental and modified TrCel3A beta-glucosidases. Example 9 details an assay for measuring the k_{cat} of parental and modified TrCel3A beta-glucosidases.

[0056] K_G of the parental and modified TrCel3A beta-glucosidases can be determined by measuring the rate of hydrolysis of a chromogenic substrate, such as *p*-nitrophenyl-beta-D-glucopyranoside (pNPG), in the presence of various concentrations of glucose as described in Example 8. The K_G is the concentration of glucose that reduces the rate of *p*-nitrophenol (pNP) release from pNPG by 50% compared to the rate of pNPG hydrolysis in the absence of glucose. The K_{G2} constants for parental and modified TrCel3A beta-glucosidases can be determined by measuring the rate of hydrolysis of cellobiose in reactions containing increasing concentrations of cellobiose or, alternatively, by measuring the rate of hydrolysis of a chromogenic substrate, such as pNPG, in the presence of various concentrations of a cellobiose as described in Example 8. The K_{G2} is the concentration of cellobiose that reduces the rate of pNP release from pNPG by

techniques, selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543A, I543S, I543G and I543L (TrCel3A numbering) and which amino acid sequence is at least 80% identical to the amino acid sequence of the parental Family 3 beta-glycosidase from which it is derived. For example, the amino acid sequence of the modified Family 3 beta-glycosidase may be from about 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 99.9% identical to the amino acid sequence of the parental Family 3 beta-glycosidase from which it is derived.

[0083] Sequence identity can be readily determined by alignment of the amino acids of the two sequences, either using manual alignment, or any sequence alignment algorithm as known to one of skill in the art, for example but not limited to, BLAST algorithm (BLAST and BLAST 2.0; Altschul et al., 1997 and 1990), the algorithm disclosed by Smith & Waterman (1981), by the homology alignment algorithm of Needleman & Wunsch (1970), by the search for similarity method of Pearson & Lipman (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection. An alignment of 45 Family 3 beta-glycosidase sequences is provided in Figure 11.

[0084] It will be understood that the modified Family 3 beta-glycosidase may be derived from wild-type Family 3 beta-glycosidase or from a Family 3 beta-glycosidase that contains other amino acid substitutions. Alternatively, after production of the modified Family 3 beta-glycosidase comprising mutations selected from the group consisting of V43I, V43C, V101A, V101G, F260I, F260V, F260Q, F260D, I543N, I543A, I543S, I543G and I543L, it may be subsequently further modified to contain additional amino acid substitutions, including but not limited to those set forth above.

[0085] By "TrCel3A numbering" it is meant the numbering corresponding to the position of amino acids based on the amino acid sequence of TrCel3A (SEQ ID NO:116) based on alignment of the amino acid sequence of the Family 3 beta-glycosidase with the TrCel3A amino acid sequence. An example of the alignment of 44 other Family 3 beta-glycosidase amino acid sequences with the TrCel3A beta-glucosidase amino acid sequence is provided in Figure 11.

parental Family 3 beta-glycosidase from which it is derived, and the modified Family 3 beta-glycosidase may exhibit (a) an increase in the K_P , (b) a decrease in K_S , or (c) an increase in k_{cat} relative to the K_P , K_S and/or k_{cat} of a parental Family 3 beta-glycosidase from which is derived.

[0091] By "parental Family 3 beta-glycosidase", it is meant a Family 3 beta-glycosidase that does not contain:

isoleucine or cysteine at position 43,

alanine or glycine at position 101;

isoleucine, valine, glutamine, aspartic acid at position 260; or

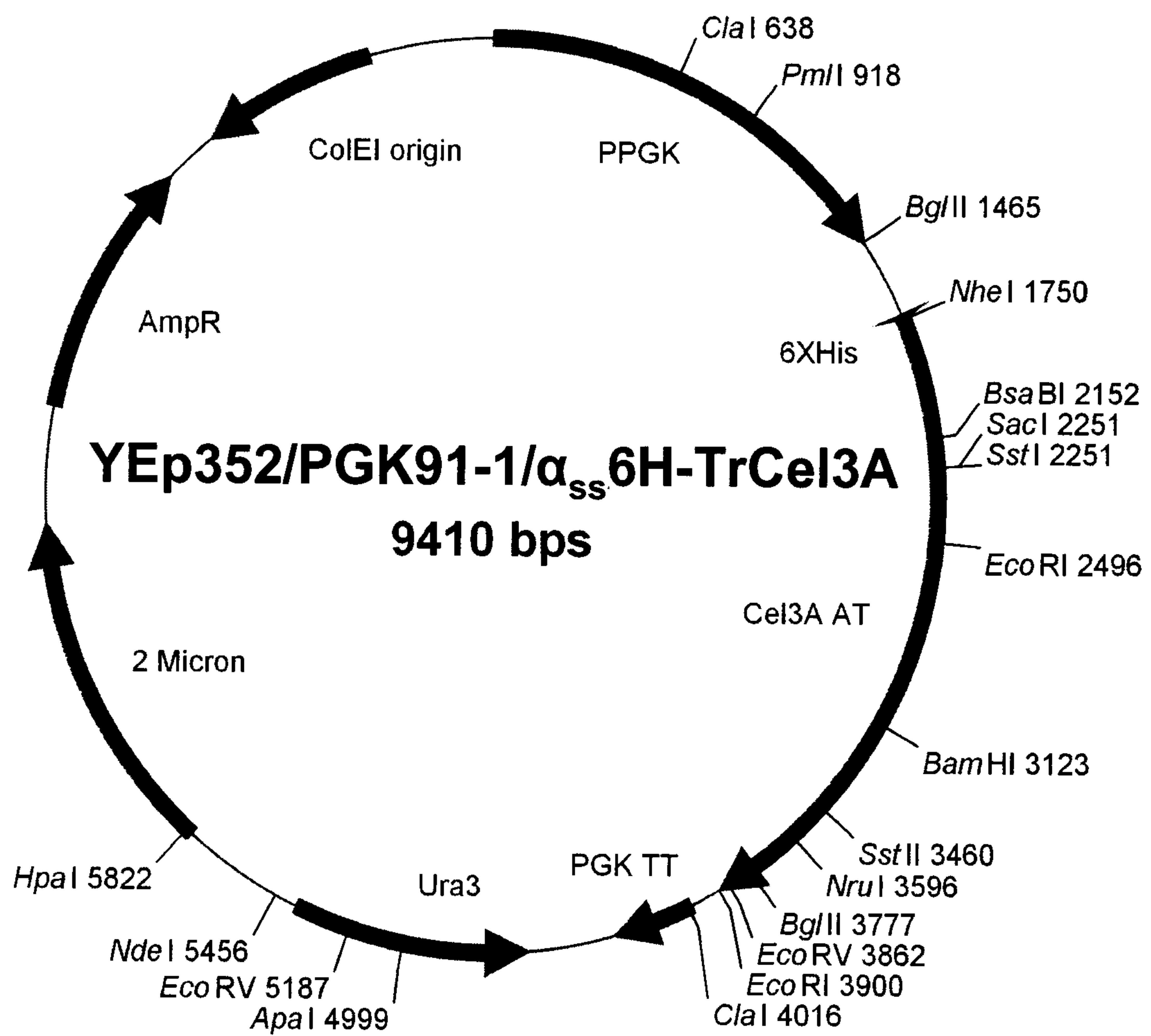
asparagine, alanine, serine, glycine or leucine at position 543.

[0092] The modified Family 3 beta-glycosidase may be derived from a parental Family 3 beta-glycosidase that comprises one or more naturally-occurring amino acid(s) at the substituted positions corresponding to that of the modified Family 3 beta-glycosidase, but that is otherwise identical to the modified Family 3 beta-glycosidase, for example a native Family 3 beta-glycosidase from *A. nidulans* -AN1804.2, *B. fuckeliana*, *T. aurantiacus levisporus*. The parental Family 3 beta-glycosidase may contain one or more amino acid substitutions at other positions, given that these substitutions are also present in the corresponding modified Family 3 beta-glycosidase. Family 3 beta-glycosidases suitable as parental beta-glycosidases from which modified Family 3 beta-glycosidases may be derived are provided in Table 3.

Table 3: Family 3 beta-glycosidases

Organism	Protein	GenPept Accession Number	SEQ ID NO:	Identity with aa 1-582 of TrCel3A (%)
<i>Aspergillus aculeatus</i> F-50	b-glucosidase I (BglI)	BAA10968.1	141	44.1
<i>Aspergillus fumigatus</i> Af293	b-glucosidase (AFUA_IG17410;Afulg17410)	EAL91070.1	142	35.2
<i>Aspergillus fumigatus</i> Af293	b-glucosidase (AFUA_IG05770;Afulg05770)	EAL88289.1	143	43.1
<i>Aspergillus kawachii</i> ifo4308	b-glucosidase (BglA)	BAA19913.1	144	42.6
<i>Aspergillus nidulans</i> FGSC A4	b-glucosidase (AN1804.2)	EAA64969.1	76	13.2
<i>Aspergillus nidulans</i> FGSC A4	b-glucosidase (AN0712.2)	EAA65189.1	145	19.4

<i>Aspergillus niger</i> B1	b-glucosidase / tannase (Bgl1;BG3;BGs;SP188)	CAB75696.1	146	42.8
<i>Aspergillus niger</i> CBS 513.88	An18g03570 (Bgl1)	CAK48740.1	147	42.6
<i>Aspergillus oryzae</i>	b-glucosidase	CAD67686.1	148	42.2
<i>Aspergillus oryzae</i> RIB 40	b-glucosidase (AO090009000356)	BAE54829.1	149	42.3
<i>Aspergillus oryzae</i> RIB 40	b-glucosidase 5 (Bgl5;AO090001000544)	BAE57053.1	150	43.4
<i>Aspergillus oryzae</i> RIB 40	b-glucosidase 1 (Bgl1;AO090003001511)	BAE58551.1	151	31.6
<i>Botryotinia fuckeliana</i>	b-glucosidase (Sap1)	CAB61489.1	152	35.9
<i>Coccidioides posadasii</i> C735	b-glucosidase/ exo-b-1,3- glucosidase (Bgl2)	AAF21242.1	153	41.4
<i>Coccidioides posadasii</i> C735	b-glucosidase (Bgl1)	AAB67972.1	154	40.4
<i>Dictyostelium discoideum</i> Ax3	b-glucosidase	AAA74233.1	155	16.6
<i>Hordeum vulgare</i>	exo-1,3-glucanase II (EII;ExoII)	AAC49170.1	156	19.5
<i>Hypocrea jecorina</i> QM6a	Cel3c - Cel3C	AAP57756.1	157	19.4
<i>Hypocrea jecorina</i> QM6a	Cel3e - Cel3E	AAP57760.1	158	36.5
<i>Hypocrea jecorina</i> QM6a	Cel3d (fragment) - Cel3D	AAP57759.1	159	12.8
<i>Hypocrea jecorina</i> QM6a	b-glucosidase - Cel3B	AAP57755.1	160	41.5
<i>Hypocrea jecorina</i> RutC-30	b-xylosidase (Bxl1)	CAA93248.1	161	19.2
<i>Kluyveromyces fragilis</i>	b-glucosidase	CAA29353.1	162	18.0
<i>Kuraishia capsulata</i> 35M5N	b-glucosidase (BglN)	AAA91297.1	163	33.0
<i>Nicotiana tabacum</i>	b-glucosidase	BAA33065.1	96	20.0
<i>Periconia</i> sp. BCC 2871	b-glucosidase	ABX84365.1	164	41.9
<i>Phaeosphaeria avenaria</i> WAC1293	b-glucosidase (Bgl1)	CAB82861.1	165	42.7
<i>Phanerochaete chrysosporium</i> K-3	glucan b-1,3-glucosidase (Bgl)	BAB85988.1	166	52.6
<i>Phanerochaete chrysosporium</i> OGC101	glucan 1,3-b-glucosidase (CbgL) - Bgl1A	AAC26489.1	167	52.6
<i>Pichia anomala</i> var. <i>acetabularis</i>	b-glucosidase	CAA26662.1	168	36.4
<i>Piromyces</i> sp. F2	b-glucosidase (Cel3A) - Cel3A	AAO41704.1	169	38.3
<i>Saccharomycopsis fibuligera</i>	b-glucosidase 2 (Bgl2)	AAA34315.1	170	38.2
<i>Saccharomycopsis fibuligera</i>	b-glucosidase 1 (Bgl1)	AAA34314.1	171	38.3
<i>Schizosaccharomyces pombe</i> 972h-	b-glucosidase (SPBC1683.04)	CAB91166.1	172	17.4
<i>Septoria lycopersici</i>	b-glucosidase (tomatinase;B2Tom)	AAB08445.1	173	37.2
<i>Talaromyces emersonii</i>	b-glucosidase - Cel3A	AAL69548.3	174	43.2
<i>Thermoascus aurantiacus</i>	b-glucosidase (Bg2;BGII)	AAV33982.1	175	45
<i>Thermoascus aurantiacus</i> IFO 9748	b-glucosidase (Bgl1;Bgl1)	AAZ95587.1	176	44.3
<i>Thermoascus aurantiacus</i> var. <i>levisporus</i>	b-1,4-glucosidase (Bgl1)	ABX79552.1	177	44.3
<i>Thermoascus aurantiacus</i> var. <i>levisporus</i>	b-1,4-glucosidase (Bgl2)	ABX56926.1	178	45
<i>Trichoderma viride</i> AS 3.3711	b-D-glucoside glucohydrolase (Bgl1)	AAQ76093.1	179	98.3
<i>Tropaeolum majus</i>	b-glucosidase	CAA07070.1	180	20.3
<i>Uromyces viciae-fabae</i>	b-glucosidase (Bgl1)	CAE01320.1	181	39.8
<i>Volvariella volvacea</i> V14	b-glucosidase (Bgl;BGL-II)	AAG59831.1	182	18.8

**Figure 1**

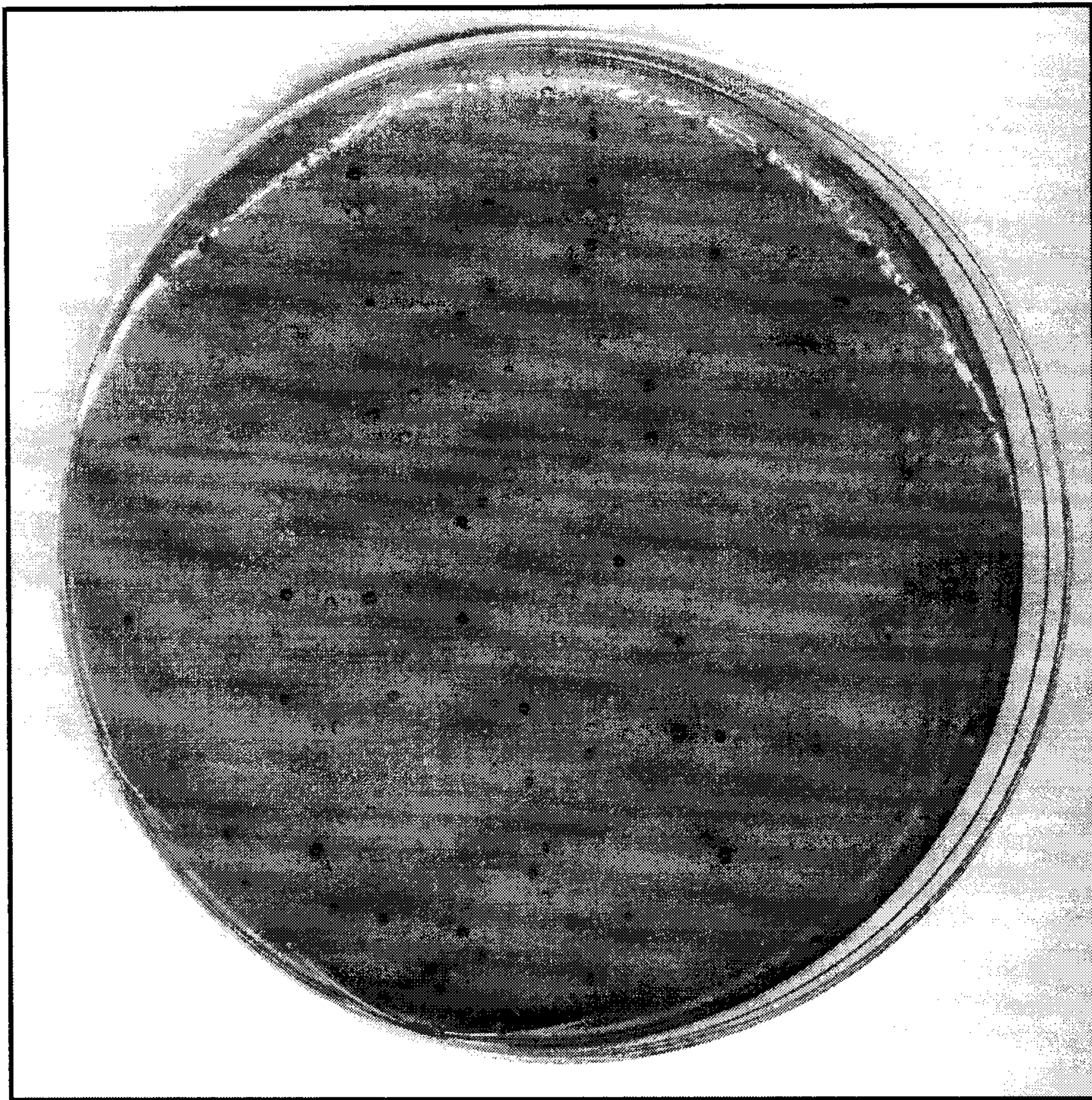
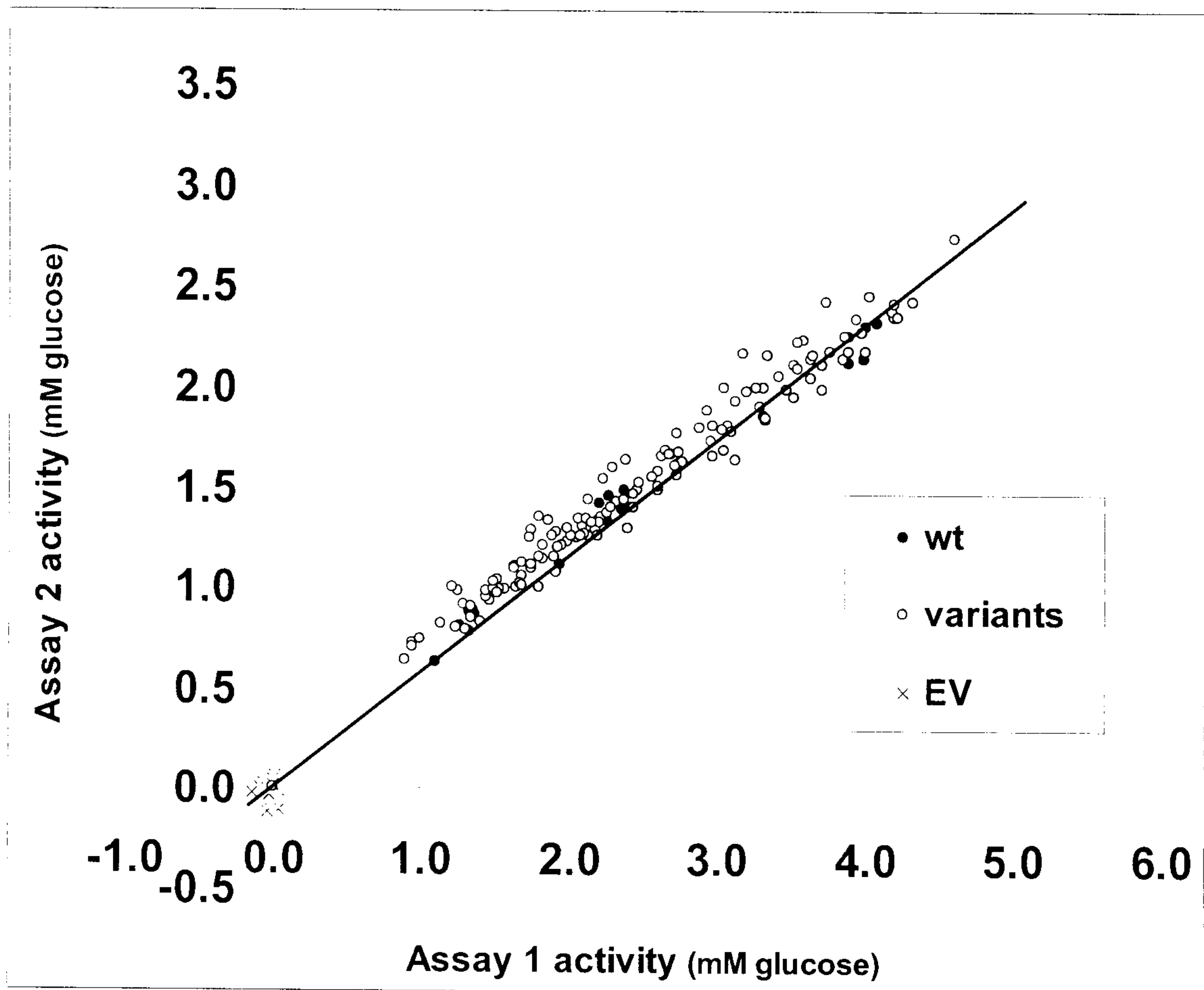
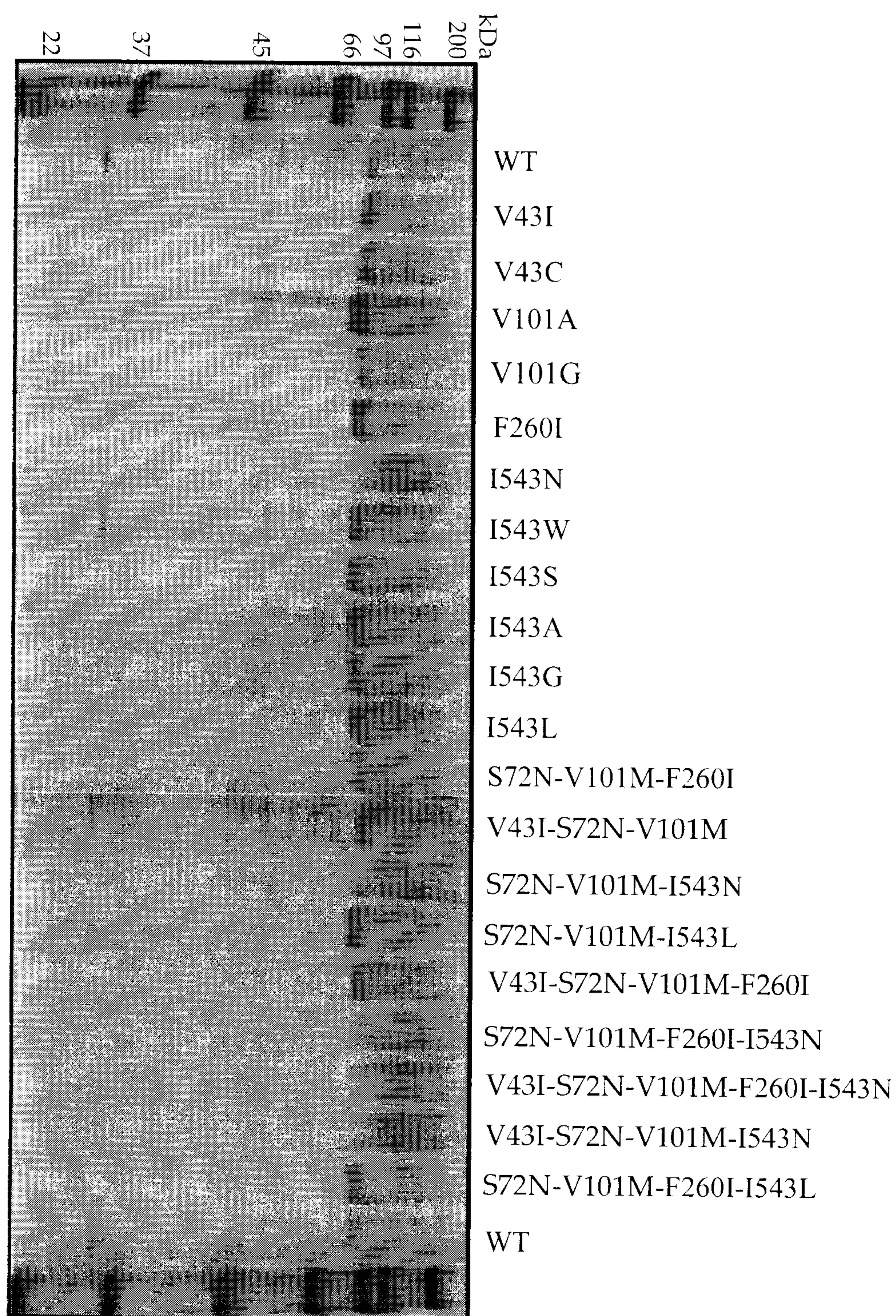


Figure 2

**Figure 3**

**Figure 4**

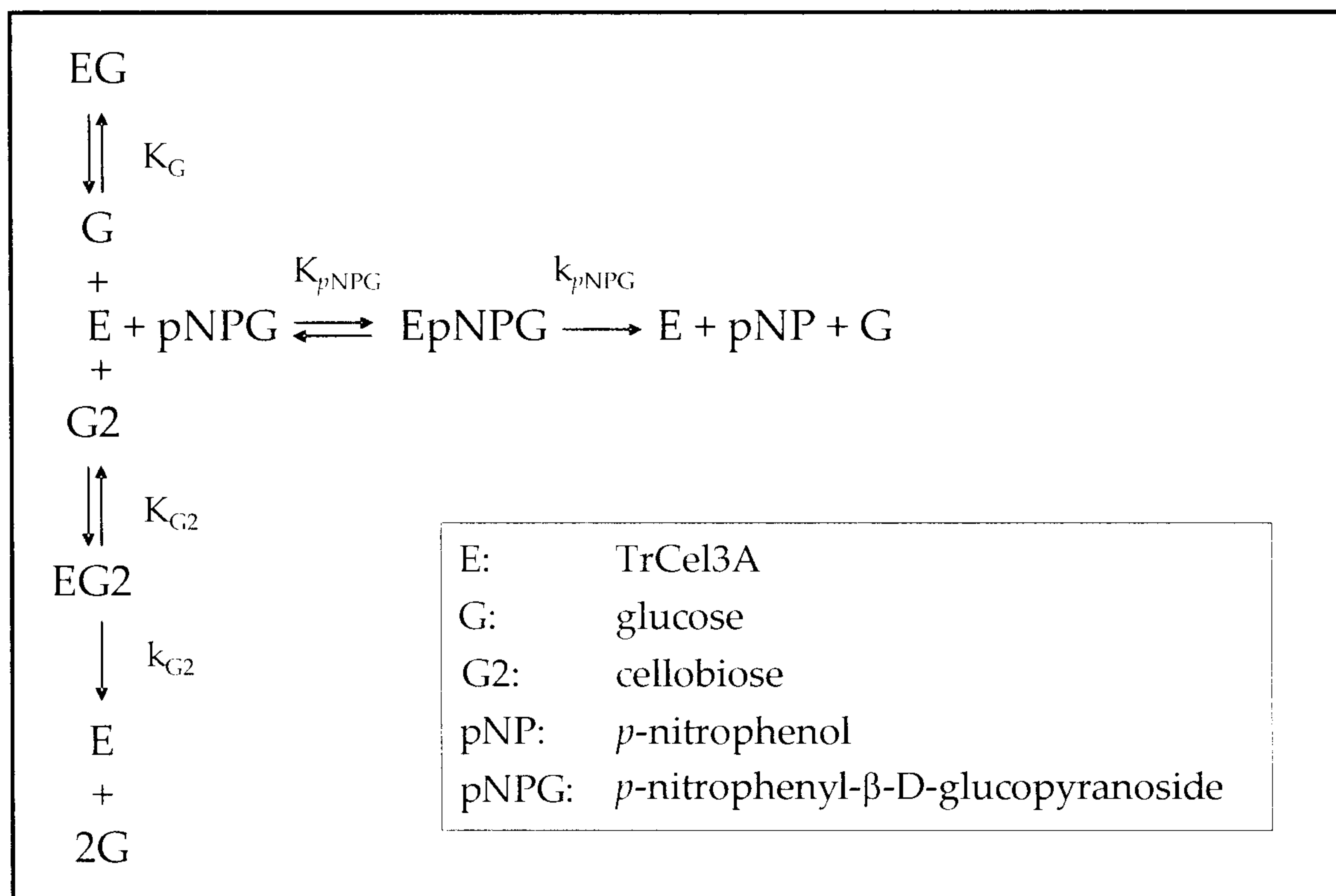
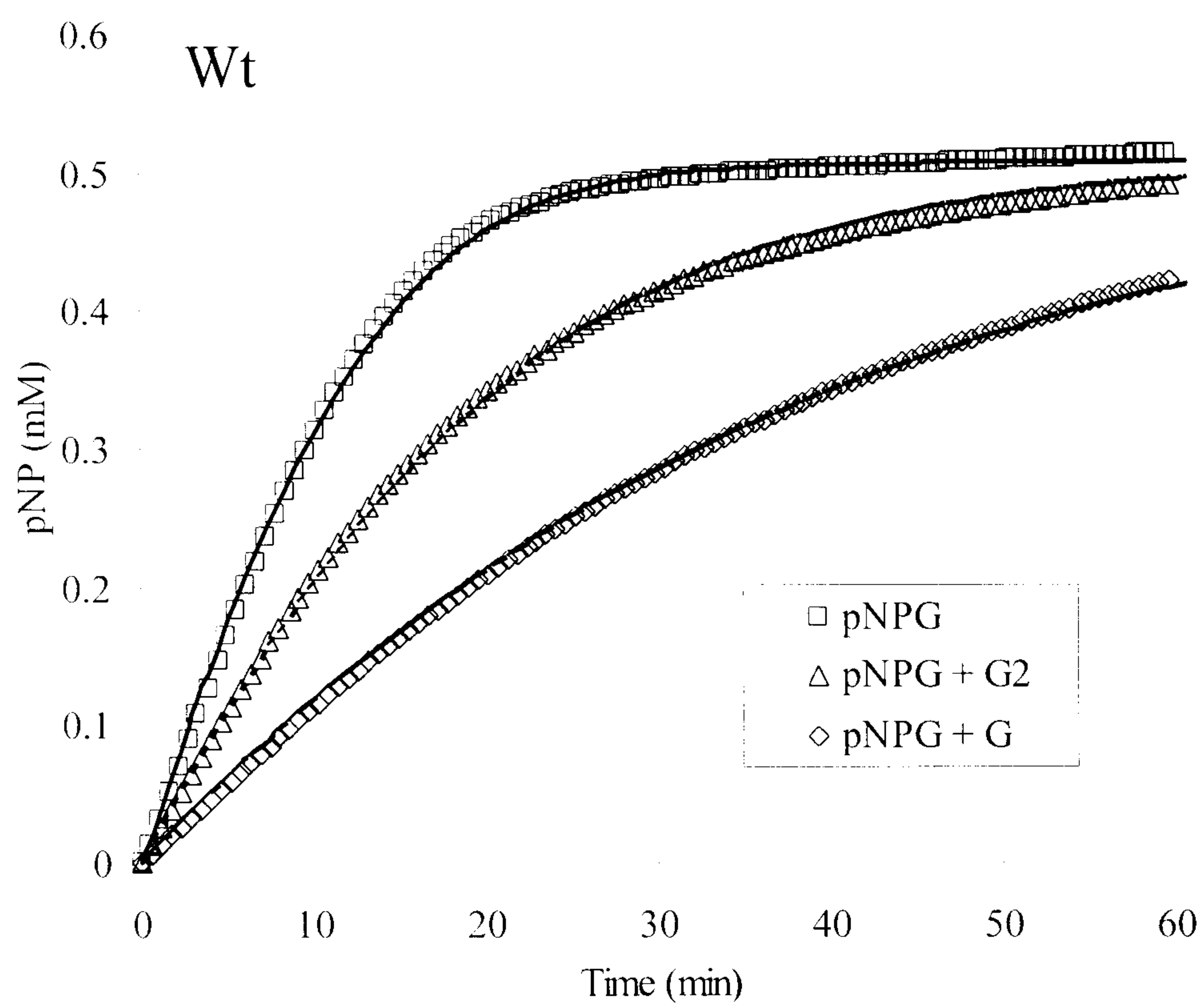
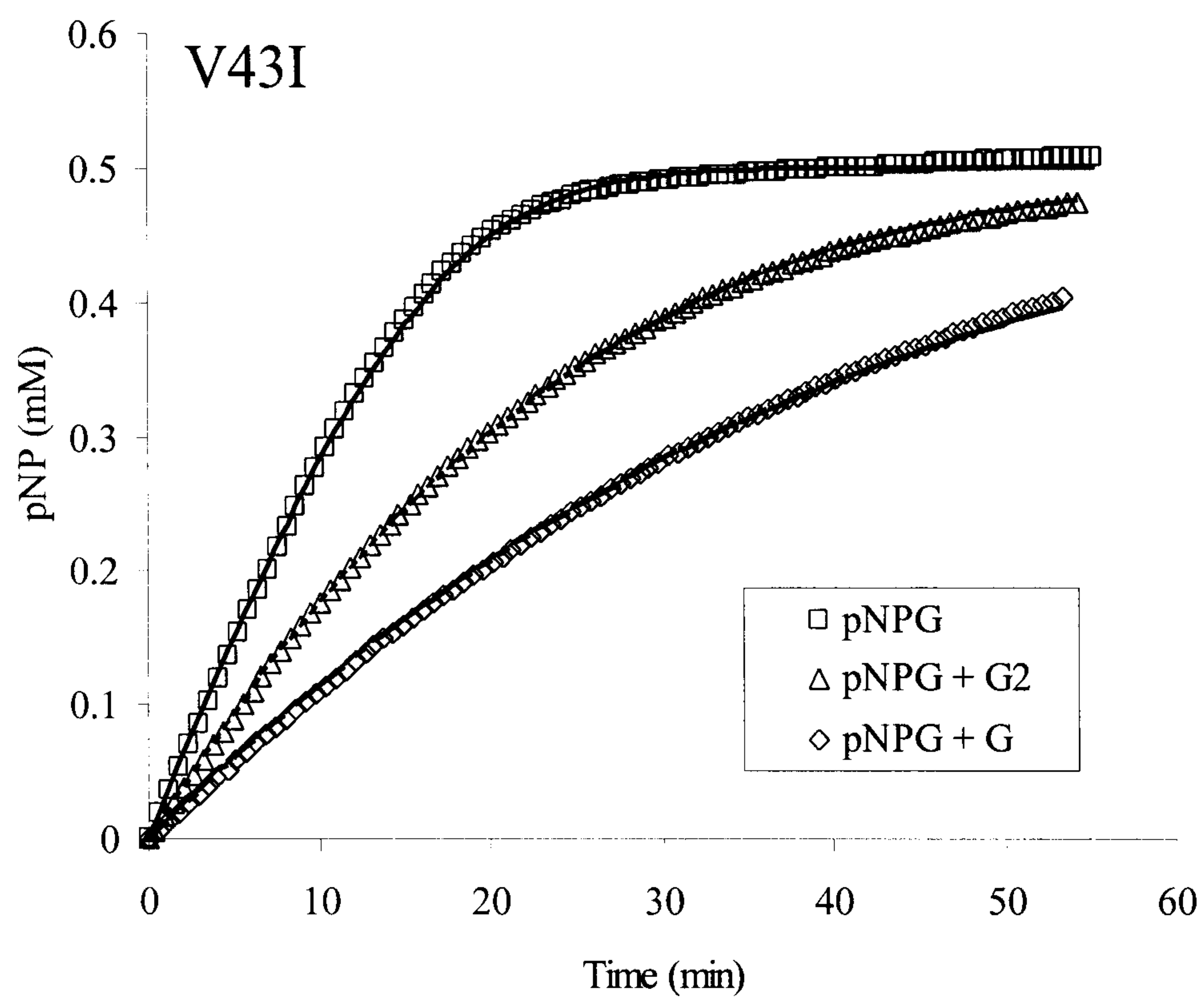
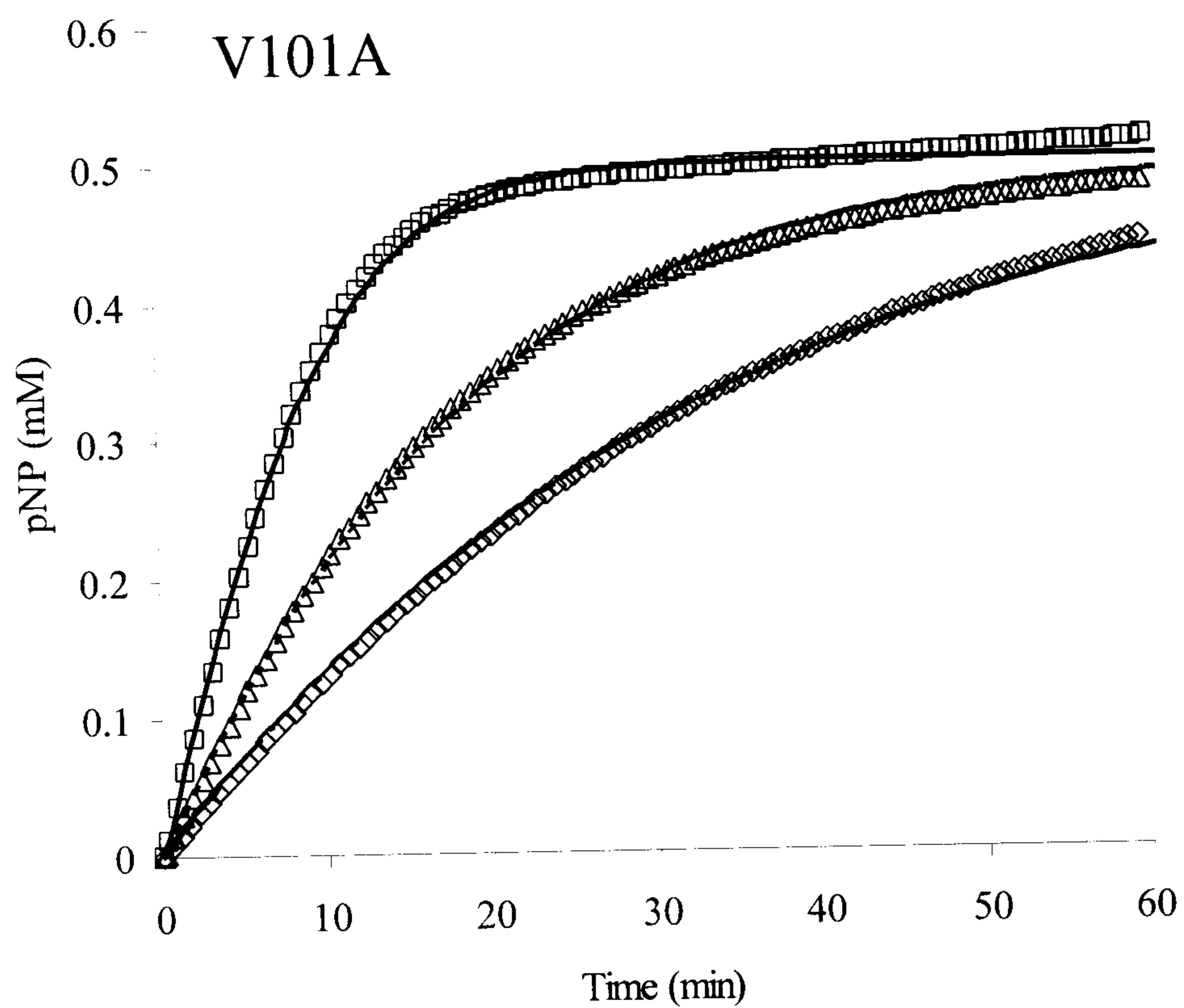
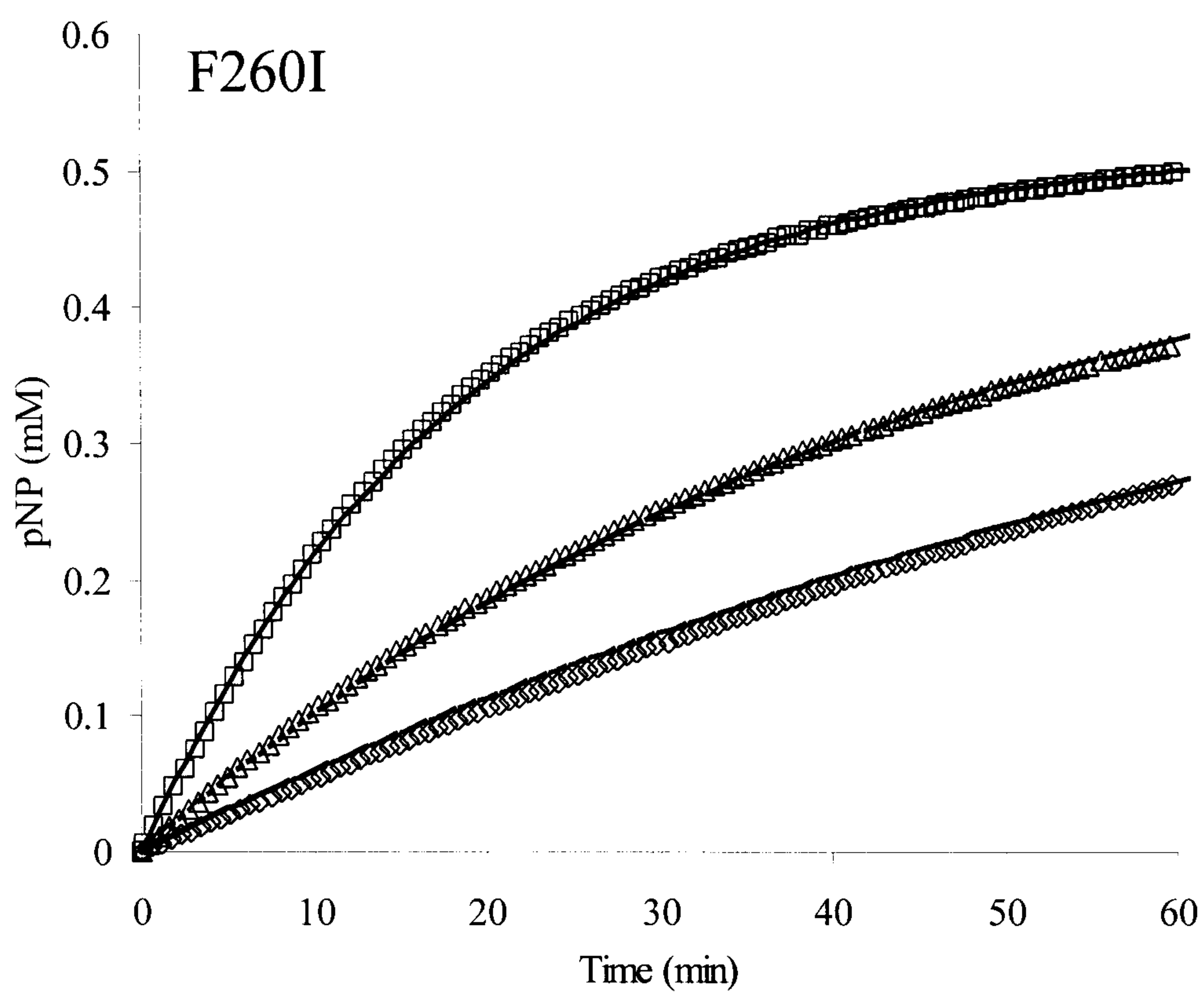


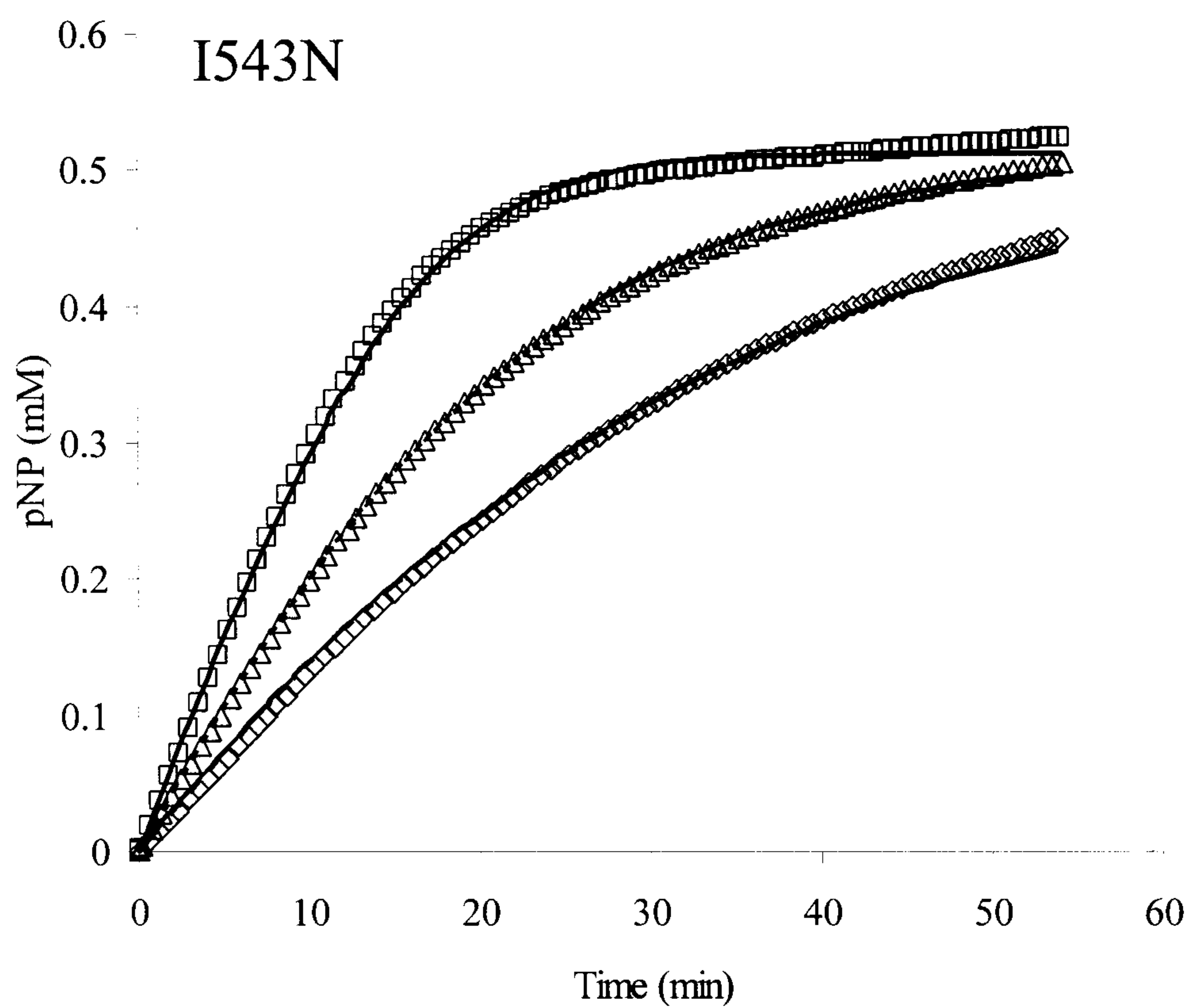
Figure 5

**Figure 6**

**Figure 7**

**Figure 8**

**Figure 9**

**Figure 10**

SEQ ID NO: 1	1	-----VPPAGTPWGTAYDKAKAALAKLNLQDKV---	29
CONSENSUS	1	-----W--G-QGEWAEAYARAKDLVSQMTLAEKV---	26
SEQ ID NO: 72	1	-----PWANG-QGEWAEAYQRAVAIVSQMTLDEKV---	29
SEQ ID NO: 73	1	-----GSWAAAYAKAKKFVAQLTPEEKV---	23
SEQ ID NO: 74	1	-----PWADG-QGEWADAHRRAVEIVSQMTLAEKV---	29
SEQ ID NO: 75	1	-----PWANG-QGDWAQAYQRAVDIVSQMTLAEKV---	29
SEQ ID NO: 76	1	-----ALPIYKNASYCVDERVERDLLSRMTLEEKAGQL	32
SEQ ID NO: 77	1	-----AQIESVLSKLTLEEKI---	16
SEQ ID NO: 78	1	-----PWANG-QGDWAQAYQRAVDIVSQMTLDEKV---	29
SEQ ID NO: 79	1	-----PWANG-QGDWAEAYQRAVDIVSQMTLAEKV---	29
SEQ ID NO: 80	1	-----PWADG-QGEWAEVYKRAVDIVSQMTLTEKV---	29
SEQ ID NO: 81	1	-----PWADG-QGEWAEVYKRAVDIVSQMTLTEKV---	29
SEQ ID NO: 82	1	-----PNGGWISEWASAYEKAHRVVSNMTLAEKV---	29
SEQ ID NO: 83	1	-----PETKGLGDWEEAFTKARSLVAQMTDKEKN---	29
SEQ ID NO: 84	1	-----RASTSFQNWTHAHLALTFVNQLNITEKS---	29
SEQ ID NO: 85	1	-----PWASG-QGEWSEAYNKAREFVSQTLTEKV---	29
SEQ ID NO: 86	1	-----EAYSKAHTVVSKMTLAGKV---	19
SEQ ID NO: 87	1	SQIYPKKQLNQENINFMSARDTFVDNLMSKMSITEKIGQM	40
SEQ ID NO: 88	1	-----DYLKYKDPKQPLGVRIKDLLGRMTLAEKIGQM	32
SEQ ID NO: 89	1	-----VEAILKKLTAEKV---	14
SEQ ID NO: 90	1	-----GPWAHAYRRAEKLVRQMTLEEKA---	23
SEQ ID NO: 91	1	-----	1
SEQ ID NO: 92	1	-----WMDPSAPGWEQAYAQAKEFVSGLTLEKV---	29
SEQ ID NO: 93	1	-----KNNLVCDSSAGYVERAQUALISLFTLEELILN-	31
SEQ ID NO: 94	1	-----VEQLLSELNQDEKI---	14
SEQ ID NO: 95	1	-----PEISGLGDWQFAYQRAREIVALMTNEEKT---	29
SEQ ID NO: 96	1	-----EYVKYKDPKQPVGARIKDLMKRMTLEEKIGQM	32
SEQ ID NO: 97	1	-----PWMDPSAAGWAEAYTKAQAFVRQLTLEKV---	30
SEQ ID NO: 98	1	-----PWIEG-LGDWEAAYQKAQAFVSQTLTLEKV---	29
SEQ ID NO: 99	1	-----TVANVSPEWAAAYVKAQAQAAVAKLSVTDMV---	29
SEQ ID NO: 100	1	-----TVANVSPEWAAAYVKAQAQAAVAKLSVTDMV---	29
SEQ ID NO: 101	1	-----GKWQAIFYRARELVDQMSIAEKV---	23
SEQ ID NO: 102	1	-----VSQVFATSWSEADEKAKSFMSDLSESEKI---	29
SEQ ID NO: 103	1	-----QGGRLQGVWQDAYTKAKALVSQMTIVEKV---	29
SEQ ID NO: 104	1	-----QGGRLQDVWQEAYARAKAIVGQMTIVEKV---	29
SEQ ID NO: 105	1	-----MMEHDVEDLINQLDISEKA---	19
SEQ ID NO: 106	1	-----WKAAFEKAADAVSRLNLTQKV---	21
SEQ ID NO: 107	1	-----PWANG-QGDWAEAYQKAVQFVSQTLAEKV---	29
SEQ ID NO: 108	1	-----WVSDWAPAYAKAYEVVSNMTLAEKV---	25
SEQ ID NO: 109	1	-----PWMDG-NGEWAEAYRRAVD FVSQTLAEKV---	29
SEQ ID NO: 110	1	-----PWMNG-NGEWAEAYRRAVD FVSQTLAEKV---	29
SEQ ID NO: 111	1	-----PPGGWVSDWAPAYAKAYEVVSNMTLAEKV---	29
SEQ ID NO: 112	1	-----VVLPAAGTPWGTAYDKAKAALAKLNLQDKV---	29
SEQ ID NO: 113	1	-----EYMRYKDPKKPLNVRIKDLMSRMTLAEKIGQM	32
SEQ ID NO: 114	1	-----PVQSGLSPWSESIVRARAFVAQLTIEEKV---	29
SEQ ID NO: 115	1	-----MPPSDFAKANIDEIVEQLTLDEAI---	24

Figure 11/1

SEQ ID NO: 1	-----GIVSG	34
CONSENSUS	-----NLTTG	31
SEQ ID NO: 72	-----NLTTG	34
SEQ ID NO: 73	-----NLTAG	28
SEQ ID NO: 74	-----NLTTG	34
SEQ ID NO: 75	-----NLTTG	34
SEQ ID NO: 76	FHKQLSEGPLDDDDSSGNSTETMIGKKHMTNFNLASDITNA	72
SEQ ID NO: 77	-----SLLAG	21
SEQ ID NO: 78	-----NLTTG	34
SEQ ID NO: 79	-----NLTTG	34
SEQ ID NO: 80	-----NLTTG	34
SEQ ID NO: 81	-----NLTTG	34
SEQ ID NO: 82	-----NLTS	34
SEQ ID NO: 83	-----NITYG	34
SEQ ID NO: 84	-----NIVTG	34
SEQ ID NO: 85	-----NLTTG	34
SEQ ID NO: 86	-----NLTTG	24
SEQ ID NO: 87	TQLDITTLTSPNTITINETTLAYYAKTYIGSYLNSPVSG	80
SEQ ID NO: 88	TQIERENATAEAMSKY-----FIGSVLSG	56
SEQ ID NO: 89	-----DLLAG	19
SEQ ID NO: 90	-----NITRG	28
SEQ ID NO: 91	-----	1
SEQ ID NO: 92	-----NLTTG	34
SEQ ID NO: 93	-----TQNSG	36
SEQ ID NO: 94	-----SLLSA	19
SEQ ID NO: 95	-----NLTFG	34
SEQ ID NO: 96	TQIERKVATADVMMQN-----FIGSVLSG	56
SEQ ID NO: 97	-----NLTTG	35
SEQ ID NO: 98	-----NLTTG	34
SEQ ID NO: 99	-----NLATG	34
SEQ ID NO: 100	-----NLATG	34
SEQ ID NO: 101	-----NLTTG	28
SEQ ID NO: 102	-----DIVTG	34
SEQ ID NO: 103	-----NLTTG	34
SEQ ID NO: 104	-----NLTTG	34
SEQ ID NO: 105	-----MLLSG	24
SEQ ID NO: 106	-----ALTTG	26
SEQ ID NO: 107	-----NLTTG	34
SEQ ID NO: 108	-----NMTTG	30
SEQ ID NO: 109	-----NLTTG	34
SEQ ID NO: 110	-----NLTTG	34
SEQ ID NO: 111	-----NMTTG	34
SEQ ID NO: 112	-----GIVSG	34
SEQ ID NO: 113	TQIERKEATPDVISKY-----FIGSVLSG	56
SEQ ID NO: 114	-----NLTTG	34
SEQ ID NO: 115	-----SLTAG	29

Figure 11/2

SEQ ID NO: 1	VGWNGGP-----CVGNTSPASKIS-YP---SLCLQDGPL	64
CONSENSUS	TGWELGP-----CVGNTGSVPRLGIP----GLCLQDGPL	61
SEQ ID NO: 72	TGWELEK-----CVGQTGGVPRLNIG----GMCLQDSPL	64
SEQ ID NO: 73	-TDANN-----GCSGNIAAIPRLNFP----GLCVSDAGN	57
SEQ ID NO: 74	TGWEMDR-----CVGQTGSVPRLGINW---GLCGQDSPL	65
SEQ ID NO: 75	TGWELEL-----CVGQTGGVPRLGVP----GMCLQDSPL	64
SEQ ID NO: 76	TQTAEFIN-----LIQKRALQTRLGIP-----ITISTDPR	102
SEQ ID NO: 77	KNFWETQ-----DYPEKGV-----PVKTS DGPN	45
SEQ ID NO: 78	TGWELEL-----CVGQTGGVPRLGVP----GMCLQDSPL	64
SEQ ID NO: 79	TGWELEL-----CVGQTGGVPRLGIP----GMCAQDSPL	64
SEQ ID NO: 80	TGWQLER-----CVGQTGSVPRLNIP----SLCLQDSPL	64
SEQ ID NO: 81	TGWQLER-----CVGQTGSVPRLNIP----SLCLQDSPL	64
SEQ ID NO: 82	TGIYMGP-----CAGQTGSVPRFGIPN----LCLHDSPL	64
SEQ ID NO: 83	YSSTAN-----GCGGTSGGVPRLGFP----GICLQDAGN	64
SEQ ID NO: 84	TYHTSTLDGKPMNCIGNIGPIRLNFT----GICLNDGPA	70
SEQ ID NO: 85	VGWMQEA-----CVGNVGSIPRLGFR----SLCMQDGPL	64
SEQ ID NO: 86	TGFLMA-----LVGQTGSALRFGIPR----LCLQDGPL	53
SEQ ID NO: 87	GLAGDIHHINSSVWLDMINTIQTIVIEGSPNKIPMIYGLD	120
SEQ ID NO: 88	GGSVPS PQASAAWQSMVNEMQK GALSTRL-GIPMIYGID	95
SEQ ID NO: 89	IDFWHTK-----ALPKHGV-----SLRFTDGP	43
SEQ ID NO: 90	-FTGDN-----VCAGNTGSVPRLGWP----GMCVHDAGN	57
SEQ ID NO: 91	-----	1
SEQ ID NO: 92	VGWMGEK-----CVGNVGTVPRLGMR----SLCMQDGPL	64
SEQ ID NO: 93	PGVPRLG-----LPNYQVWN-EALHGLDRANF	62
SEQ ID NO: 94	VDFWHTK-----KIERLGIP----AVRVSDGP	43
SEQ ID NO: 95	-SSGDT-----GCSGMISDVDPDVPF----GLCLQDAGN	63
SEQ ID NO: 96	GGSVPA PKASAQVWTNMVDEIQKGS LSTRL-GIPMIYGID	95
SEQ ID NO: 97	VGWEGEA-----CVGNTGSIPRLGFP----GFCTQDSPL	65
SEQ ID NO: 98	TGWQSDH-----CVGNTGGVPRLNFT----GICNQDAPL	64
SEQ ID NO: 99	VQWEKGP-----CVGNTPAISSIPGFT---GLCLQDSPV	65
SEQ ID NO: 100	VQWQKGP-----CVGNTPAISSIPGFT---GLCLQDSPV	65
SEQ ID NO: 101	VGSASGP-----CSGNTGSVPRLNIS----SICVQDGPL	58
SEQ ID NO: 102	YMNMQGT-----CVGNIKPLDRKNFK----GLCLQDGPA	64
SEQ ID NO: 103	TGWQLGP-----CVGNTGSVPRFGIP----NLCLQDGPL	64
SEQ ID NO: 104	TGWQLDP-----CVGNTGSVPRFGIP----NLCLQDGPL	64
SEQ ID NO: 105	TDLWHTA-----AIPRLNIP----SIRLS DGPN	48
SEQ ID NO: 106	TTAG-----LSCNGNIAPIPEINFS----GLCLADGPV	55
SEQ ID NO: 107	TGWEQDR-----CVGQVGSIPRLGFP----GLCMQDSPL	64
SEQ ID NO: 108	TGIFMGP-----CVGQTGSALRFGIPN----LCLMDSPL	60
SEQ ID NO: 109	VGWMQEK-----CVGETGSIPRLGFR----GLCLQDSPL	64
SEQ ID NO: 110	VGWMQEK-----CVGETGSIPRLGFR----GLCLQDSPL	64
SEQ ID NO: 111	TGIFMGP-----CVGQTGSALRFGIPN----LCLMDSPL	64
SEQ ID NO: 112	VGWNGGP-----CVGNTSPASKFS-YP---SLCLQDGPL	64
SEQ ID NO: 113	GGSVPA PKASPEAWVDLVNGMQKAALSTRL-GIPMIYGID	95
SEQ ID NO: 114	AG-TQGR-----CVGETGTVPRLGFNQ---PICLQDGPV	64
SEQ ID NO: 115	VGFWHTH-----AIERLGVP----AVKVS DGPN	53

Figure 11/3

SEQ ID NO: 1	GVRYST-----GSTAFTPGVQAASTWDVNLIRERG--	94
CONSENSUS	GVRGTD-----YNSAFPAGVNVAATWDKNLAYLRG--	91
SEQ ID NO: 72	GIRDS-----YNSAFPAGVNVAATWDKNLAYLRG--	94
SEQ ID NO: 73	GLRGTD-----YVSSWPSGLHVGASWNKALARQRA--	87
SEQ ID NO: 74	GIRFSD-----LNSAFPAGTNVAATWDKTLAYLRG--	95
SEQ ID NO: 75	GVRDS-----YNSAFPAGMNVAATWDKNLAYLRG--	94
SEQ ID NO: 76	HSFTENVGTGFQAGVFSQWPESLGLAALRDPQLVREFA--	140
SEQ ID NO: 77	GARGATFKG---GVTAACFPASSLLAATWDLDAKHIG--	80
SEQ ID NO: 78	GVRDS-----YNSAFPAGMNVAATWDKNLAYLRG--	94
SEQ ID NO: 79	GVRDS-----YNSAFPAGVNVAATWDKNLAYLRG--	94
SEQ ID NO: 80	GIRFSD-----YNSAFPAGVNVAATWDKTLAYLRG--	94
SEQ ID NO: 81	GIRFSD-----YNSAFPAGVNVAATWDKTLAYLRG--	94
SEQ ID NO: 82	GVRNSD-----HNTAFPAGITVGATFDKDLMYERG--	94
SEQ ID NO: 83	GVRGTD-----MVNSYASGVHVGASWNRDLTYSRA--	94
SEQ ID NO: 84	LLLVRD-----LISVFPSGVTLAATWDRELVIQSY--	100
SEQ ID NO: 85	GIRFAD-----HVSAFPAGINVGATWSKSLAYLRG--	94
SEQ ID NO: 86	GLRNTD-----HNTAFPAGISVGATFDKKLMYERG--	83
SEQ ID NO: 87	SVHGAN--Y---VHKATLEPHNTGLAATENIEHATTAA--	153
SEQ ID NO: 88	AVHGHNN-----VYKATIFPHNVGLGATRDPMLVKRIG--	128
SEQ ID NO: 89	GVRGTFKFFN---GVPAACFPCTSLGSTFNQTLLEEAG--	78
SEQ ID NO: 90	GVRATD-----LVNSYPSGIHVGASWDRNLTYERG--	87
SEQ ID NO: 91	-----	1
SEQ ID NO: 92	GLRENT-----YNSAFSVGLTAAASWSRHLWVDRG--	94
SEQ ID NO: 93	ATKGGQ-----FEWATSFPMPILTAAALNRTLHIQIADI	96
SEQ ID NO: 94	GIRGTFKFFD---GVPSGCFPNGTGLASTFDRDLLETAG--	78
SEQ ID NO: 95	GVRGTD-----MVNAYASGLHVGASWNRQLAYDRA--	93
SEQ ID NO: 96	AVHGHNN-----VYGATIFPHNVGLGVTRDPDLVKRIG--	128
SEQ ID NO: 97	GVRFAD-----YVSAFTAGGTIAASWDRSEFYRRG--	95
SEQ ID NO: 98	GVRFAD-----YVSAFPSSGTIAAAWDRGEWYLRG--	94
SEQ ID NO: 99	GVRYAD-----GTSVFPPEINVAATWNRTLMRQRG--	95
SEQ ID NO: 100	GVRYAD-----GTSVFPPEINVAATWNRTLMRQRG--	95
SEQ ID NO: 101	SVRAAD-----LTDVFPCGMAASSSFNKQLIYDRA--	88
SEQ ID NO: 102	GVRFNCG-----TSTTWQAGINNAATFNKDLKYKIG--	95
SEQ ID NO: 103	GVRLTD-----FSTGYPSGMATGATFNKDLFLQRG--	94
SEQ ID NO: 104	GVRFAD-----FVTGYPSGLATGATFNKDLFLQRG--	94
SEQ ID NO: 105	GIRGTSFFN---SSPSACFPCTALGATFDKKLLFEVG--	83
SEQ ID NO: 106	SVRIAD-----LATVFPAGLTAAATWDRQLIYERA--	85
SEQ ID NO: 107	GVRDTD-----YNSAFPAGVNVAATWDRNLAYRRG--	94
SEQ ID NO: 108	GIRNTD-----HNTAFPAGITVGATFDKELMYARG--	90
SEQ ID NO: 109	GVRFAD-----YVSAFPAGVNVAATWDKNLAYLRG--	94
SEQ ID NO: 110	GVRFAD-----YISAFPAGVNVAATWDKNLAYLRG--	94
SEQ ID NO: 111	GIRNTD-----HNTAFPAGITVGATFDKELMYARG--	94
SEQ ID NO: 112	GVRYST-----GSTAFTPGVQAASTWDVNLIRERG--	94
SEQ ID NO: 113	AVHGHNN-----VYNATIFPHNVGLGVTRDPALIKRIG--	128
SEQ ID NO: 114	GIRYTD-----FNSVFPAAINVAATFDKQLMFKRA--	94
SEQ ID NO: 115	GIRGNHFFM---GTPAKCLPSSTALGATWDPEVVEEVGL-	89

Figure 11/4

SEQ ID NO: 1	-QFIGEEVKASGIH--VILGPVAGPLGKTPQGGRNWEGFG	131
CONSENSUS	-QAMGEEFRGKGVD--VQLGPVAGPLGRSPDGGRNWEGFS	128
SEQ ID NO: 72	-QAMGQEFSDKGID--VQLGPAAGPLGRSPDGGRNWEGFS	131
SEQ ID NO: 73	-VQMATEFRKKGVN--VLLGPVVGPLGRVAEAGRNWEGFS	124
SEQ ID NO: 74	-KAMGEEFNDKGVD--ILLGPAAGPLGKYPDGGRIWEGFS	132
SEQ ID NO: 75	-KAMGQEFSDKGAD--IQLGPAAGPLGRSPDGGRNWEGFS	131
SEQ ID NO: 76	-EVAREEYLAVGIR--AALHPQVD-LSTEPRWARISGTWG	176
SEQ ID NO: 77	-EALADETRSKGAR--VLLAPTVC-IHRHPLGGRNFESFS	116
SEQ ID NO: 78	-KAMGQEFSDKGAD--IQLGPAAGPLGRSPDGGRNWEGFS	131
SEQ ID NO: 79	-QAMGQEFSDKGAD--IQLGPAAGPLGRSPDGGRNWEGFS	131
SEQ ID NO: 80	-QAMGEEFSDKGID--VQLGPAAGPLGAHPDGGRNWEGFS	131
SEQ ID NO: 81	-QAMGEEFSDKGID--VQLGPAAGPLGAHPDGGRNWEGFS	131
SEQ ID NO: 82	-VGLGEEARGKGIN--VLLGPSVGPIGRKPRGGRNWEGFG	131
SEQ ID NO: 83	-QYMGAEFKRKGVN--VALGPVAGPIGRIARGGRNWEGFS	131
SEQ ID NO: 84	-KALGEEFRGKGTH--VALGPVAGPLGRHPLGGRNWEGFS	137
SEQ ID NO: 85	-KAMGEEHRDKGVD--VQLGPAVGPLGRSPDGGRNWEGFS	131
SEQ ID NO: 86	-CAMGEEFRGKGAN--VHLGPSVGPLGRKPRGGRNWEGFG	120
SEQ ID NO: 87	-QITSKDTVAVGIP--WVFAPVLG-IGVQPLWSRIYETFG	189
SEQ ID NO: 88	-EATALEVRATGIP--YAFAPCIA-VCRDPRWGRCYESYS	164
SEQ ID NO: 89	-KMMGKEAIAKSAH--VILGPTIN-MQRSPLGGRGFESIG	114
SEQ ID NO: 90	-LHMGGEFKAKGVN--VPLGPNAGPLGRTPLGGRNWEGFS	124
SEQ ID NO: 91	-----	1
SEQ ID NO: 92	-TALGSEAKGKGVD--VLLGPVAGPLGRNPNGGRNVEGFG	131
SEQ ID NO: 93	ISTQARAFSNSGRYGLDVYAPNVN-GFRSPLWGRGQETPG	135
SEQ ID NO: 94	-KLMAKESIAKNAA--VILGPTTN-MQRGPLGGRGFESFS	114
SEQ ID NO: 95	-VYMGAEFRHKGVN--VLLGPVVGPIGRVATGGRNWEGFT	130
SEQ ID NO: 96	-AATALEVRATGIP--YAFAPCIA-VCRNPRWGRCYESYS	164
SEQ ID NO: 97	-YQMGVEHRGKGVD--VQLGPVVGPIGRHHPKGGRNWEGFS	132
SEQ ID NO: 98	-YQMGSEHRGKGVD--VQLGPVVGPLGRNPKGGRNWEGFS	131
SEQ ID NO: 99	-AAMGAEFKKGKVH--VALGPMMN-LMRVPAAGRNWEGGG	131
SEQ ID NO: 100	-AAMGAEFKKGKVH--VALGPMMN-LMRVPAAGRNWEGGG	131
SEQ ID NO: 101	-VAIGSEFKKGKAD--AILGPVYGPMGVKAAGGRGWEGHG	125
SEQ ID NO: 102	-KDQGAEFYAKGIN--IALAPSMN-ILRAPASGRVWENFG	131
SEQ ID NO: 103	-QALGHEFNSKGVH--IALGPAVGPLGVKARGGRNFEEAFG	131
SEQ ID NO: 104	-QALGHEFNSKGVH--IALGPAVGPLGVKARGGRNFEEAFG	131
SEQ ID NO: 105	-EYLAEEAKAKGVS--VVLGPTVN-IHRGPLNGRGFESFS	119
SEQ ID NO: 106	-RALGSEFRGKGSQ--VHLGPASGALGRHPLGGRNWESFS	122
SEQ ID NO: 107	-VAMGEEHRGKGVD--VQLGPVAGPLGRSPDAGRNWEGFA	131
SEQ ID NO: 108	-VALGEEARGKGVN--VLMGPMVGPIGRKPRGGRNWEGFG	127
SEQ ID NO: 109	-KAMGEEHRGKGVD--VQLGPVAGPLGRHPDGGRNWEGFS	131
SEQ ID NO: 110	-KAMGEEHRGKGVD--VQLGPVAGPLGRHPDGGRNWEGFS	131
SEQ ID NO: 111	-VALGEEARGKGVN--VLMGPMVGPIGRKPRGGRNWEGFG	131
SEQ ID NO: 112	-QFIGEEVKASGIH--VILGPVAGPLGKTPQGGRNWEGFG	131
SEQ ID NO: 113	-EATALECRATGIP--YAFAPCIA-VCRDPRWGRCYESYS	164
SEQ ID NO: 114	-QAMAEFRGKGAN--VVLAPMTN-LMRTPOAGRAWEGYG	130
SEQ ID NO: 115	-KLLAPEAKLRAAS--LVLAPTSN-IQRNPLGGRSFESFS	125

Figure 11/5

SEQ ID NO: 1	VD-----PYLTGIAMGQTINGIQSVG-----	152
CONSENSUS	PD-----PYLSGVLAAETIKGIQSAG-----	149
SEQ ID NO: 72	PD-----PALTGVLFAETIKGIQDAG-----	152
SEQ ID NO: 73	ND-----PYLSGALVYETVDGAQSVG-----	145
SEQ ID NO: 74	PD-----PALTGVLFAETIKGIQDAG-----	153
SEQ ID NO: 75	PD-----PALSGVLFAETIKGIQDAG-----	152
SEQ ID NO: 76	EN-----STLTSELIVEYIKGFQGEGLGP----	201
SEQ ID NO: 77	ED-----PFLAGKLAAQYIKGLQGNG-----	137
SEQ ID NO: 78	PD-----PALSGVLFAETIKGIQDAG-----	152
SEQ ID NO: 79	PD-----PALSGVLFAETIKGIQDAG-----	152
SEQ ID NO: 80	PD-----PALTGVLFAETIKGIQDAG-----	152
SEQ ID NO: 81	PD-----PALTGVLFAETIKGIQDAG-----	152
SEQ ID NO: 82	AD-----PSLQAFGGSLTIKGMQSTG-----	152
SEQ ID NO: 83	ND-----PYLSGALTGDTVRLQES-----	151
SEQ ID NO: 84	PD-----PYLTGHLMTASITGMQSVG-----	158
SEQ ID NO: 85	PD-----PVLSGYLVAETIKGIQDAG-----	152
SEQ ID NO: 86	SD-----PSLQAIAAVETIKGVQSKG-----	141
SEQ ID NO: 87	ED-----PYVASMMGAAAVRGFQGGNNSFDGPIN	218
SEQ ID NO: 88	EDPKVVQSMTTLISGLQGDVPAGSE-GRPYVGGSKK----	199
SEQ ID NO: 89	ED-----PFLAGLGAAALIRGIQSTG-----	135
SEQ ID NO: 90	ID-----PYLSGQLNAETITGMQDAG-----	145
SEQ ID NO: 91	-----MILGCESTG-----	9
SEQ ID NO: 92	SD-----PYLAGLALADTVTGIQNAG-----	152
SEQ ID NO: 93	ED-----AFFLSSAYTYEYITGIQGGVDP-----E	160
SEQ ID NO: 94	ED-----PYLAGMATSSSVKGMQGE-----	135
SEQ ID NO: 95	ND-----PYLAGALVYETTKGIQEN-----	150
SEQ ID NO: 96	EDHRIVRSMTETIIPGLQGDLPKSKNGVPYVGGKTK----	200
SEQ ID NO: 97	PD-----PVLSGIAVAETVKGIQDAG-----	153
SEQ ID NO: 98	PD-----PYLSGIIASAESVRGIQDAG-----	152
SEQ ID NO: 99	GD-----PFLSGELAFETITGIQSSG-----	152
SEQ ID NO: 100	GD-----PFLSGEVAFETITGIQSSG-----	152
SEQ ID NO: 101	PD-----PYLEGVIAYLQTIGIQSQG-----	146
SEQ ID NO: 102	ED-----PYLSGVCQAQITKGYQDSG-----	152
SEQ ID NO: 103	SD-----PYLQGIAAAATIKGLQENN-----	152
SEQ ID NO: 104	SD-----PYLQGTAAAATIKGLQENN-----	152
SEQ ID NO: 105	ED-----STLSGLAASYVILGLQSKN-----	140
SEQ ID NO: 106	PD-----PYLSGVAMDFSIRGIQEMG-----	143
SEQ ID NO: 107	PD-----PVLTGNNMASTIQGIQDAG-----	152
SEQ ID NO: 108	AD-----PTLQAIGGAQTIKGMQSTG-----	148
SEQ ID NO: 109	PD-----PVLTGVLMAETIKGIQDAG-----	152
SEQ ID NO: 110	PD-----PVLTGVLMAETIKGIQDAG-----	152
SEQ ID NO: 111	AD-----PTLQAIGGAQTIKGMQSTG-----	152
SEQ ID NO: 112	VD-----PYLTGIAMGQTINGIQSVG-----	152
SEQ ID NO: 113	EDHTIVQAMTEIIPGLQGDVPPDVKKGVFPVGGKTK----	200
SEQ ID NO: 114	SD-----PYLSGVATVQSVLGIQSTR-----	151
SEQ ID NO: 115	ED-----PYLSGIIISASYVNGVQKGG-----	146

Figure 11/6

SEQ ID NO: 1	---VQATAKHYILNEQEELNR-----E	170
CONSENSUS	---VIAYAKHFIGNEQEHRQAS-----E	170
SEQ ID NO: 72	---VVATAKHYILNEQEHRQVAE-----AAGYGFNISD	183
SEQ ID NO: 73	---VATCTKHILNEQETNRNP-G-----MEDGVEVA	173
SEQ ID NO: 74	---VIATAKHYILNEQEHRQVGE-----AQGYGYNITE	184
SEQ ID NO: 75	---VVATAKHYIAYEQEHRQAPE-----AQGYGFNISE	183
SEQ ID NO: 76	-KSVKTVTKHFPGGGPMENGEDSH-----FY	226
SEQ ID NO: 77	---VAATIKHYAANEQETCRFTVN-----	158
SEQ ID NO: 78	---VVATAKHYIAYEQEHRQAPE-----AQGFGFNISE	183
SEQ ID NO: 79	---VVATAKHYIAYEQEHRQAPE-----AQGYGFNITE	183
SEQ ID NO: 80	---VIATAKHYIMNEQEHRQQPE-----AAGYGFNVSD	183
SEQ ID NO: 81	---VIATAKHYIMNEQEHRQQPE-----AAGYGFNVSD	183
SEQ ID NO: 82	---AIASLKHLIGNEQEHRMSS-----VITQ	176
SEQ ID NO: 83	---VIACVKHLIGNEQETHRSTPS-----MLANSRNQ	180
SEQ ID NO: 84	---VQTSSKHFIGNEQETQRSNSP-----LPDGTNID	187
SEQ ID NO: 85	---VIACVKHFIVNEQERFQAPE-----AQGYGFNISE	183
SEQ ID NO: 86	---VIATIKHLVGNEQEMYRMTN-----IVQR	165
SEQ ID NO: 87	APSAVCTAKHYFGYSNPTSGKDRT-----	242
SEQ ID NO: 88	---VAACAKHYVGDGGTFMGINEN-----	220
SEQ ID NO: 89	---VQATIKHFCLCNDQEDRRMMVQ-----	156
SEQ ID NO: 90	---VIANIKHFIANEQETLRRP-----YFGVE	169
SEQ ID NO: 91	---VISAVKHFFVANDQEHERRAVD-----	30
SEQ ID NO: 92	---TIACAKHFLNEQEHRQVGE-----ANGYGYPITE	183
SEQ ID NO: 93	HLKVAATVKHFAGYDLENWNNQSR-----L	185
SEQ ID NO: 94	---IAATVKHFVVCNDLEDQRFSSN-----	156
SEQ ID NO: 95	---VIACKHFIGNEQETNRNP-----SGTYNQ	175
SEQ ID NO: 96	---VAACAKHFVGDGGTLHGVDDES-----	221
SEQ ID NO: 97	---VIACKHFILNEQEHRQPGN-----VGDFGF--VD	182
SEQ ID NO: 98	---VIACKHYIMNEQEHRQPGN-----FEDQGF--VD	181
SEQ ID NO: 99	---AQACAKHFINNEQEHR-----D	170
SEQ ID NO: 100	---AQACAKHFINNEQEHR-----D	170
SEQ ID NO: 101	---VVSTAKHLIGNEQEHRFAKKDKHAGKIDPGMENTSS	183
SEQ ID NO: 102	---VIVAACHYVANDIEHNR-----E	170
SEQ ID NO: 103	---VMACVKHFIGNEQDIYRQPSNSKVDPEYDP---ATKE	186
SEQ ID NO: 104	---VMACVKHFIGNEQEKYRQPDD--INPATNQ---TTKE	184
SEQ ID NO: 105	---VQACIKHFVCNDMEDERNVS-----	161
SEQ ID NO: 106	---VQANRKHFIGNEQETQRSNTF-----TDDGTEIQ	172
SEQ ID NO: 107	---VIACAKHFILYEQEHRQG-----AQDGYDISD	180
SEQ ID NO: 108	---VIATLKHFIGNEQEMYRMTS-----VIQR	172
SEQ ID NO: 109	---VIACAKHFIGNEMEHFRQASE-----AVGYGFDITE	183
SEQ ID NO: 110	---VIACAKHFIGNEMEHFRQASE-----AVGYGFDITE	183
SEQ ID NO: 111	---VIATLKHFIGNEQEMYRMTS-----VIQR	176
SEQ ID NO: 112	---VQATAKHYILNEQEELNR-----E	170
SEQ ID NO: 113	---VAACAKHFVGDGGTTKGIDEN-----	221
SEQ ID NO: 114	---ASACVKHYIGNEQEHRGGS-----GAT-	174
SEQ ID NO: 115	---IGATIKHFVGNDKEDDRQGYD-----	167

Figure 11/6/1

16/20/1

SEQ ID NO: 1	TISSNPDDRTLHELYTWPFADAVQAN-VASVMCSYNKVNT	209
CONSENSUS	S-SSNIDDRTLHELYLWPFADAVRAG-VGSVMCSYNQVNN	208
SEQ ID NO: 72	TISSNVDDKTIHEMYLWPFADAVRAG-VGAIMCSYNQINN	222
SEQ ID NO: 73	AVSSNIDDKTMHELYLWPFQDAVLG-SASIMCSYNRVNN	212
SEQ ID NO: 74	TISSNVDDKTMHELYLWPFADAVRAG-VGAVMCSYNQINN	223
SEQ ID NO: 75	SGSANLDDKTMHELYLWPFADAIRAG-AGAVMCSYNQINN	222
SEQ ID NO: 76	YGKNQTPGNNIDEHLIPFKAALAAG-ATEIMPYYSRPIG	265
SEQ ID NO: 77	---EHITERALREIYLKPFIEIAIKESNPLAVMTAYNIVNG	195
SEQ ID NO: 78	SGSANLDDKTMHELYLWPFADAIRAG-AGAVMCSYNQINN	222
SEQ ID NO: 79	SGSANLDDKTMHELYLWPFADAIRAG-AGAVMCSYNQINN	222
SEQ ID NO: 80	SLSSNVDDKTMHELYLWPFADAVRAG-VGAVMCSYNQINN	222
SEQ ID NO: 81	SLSSNVDDKTMHELYLWPFADAVRAG-VGAVMCSYNQINN	222
SEQ ID NO: 82	GYSSNIDDRTLHELYLWPFSAESVRAG-AGSVMIAYNNDVNR	215
SEQ ID NO: 83	SSSSNLDDKTMHELYLWPFQDAVKAG-AGSVMCSYNRINN	219
SEQ ID NO: 84	AISSNIDDRTLHELYLWPFADAVRAG-TTSILCSYNRINE	226
SEQ ID NO: 85	SSSSNVDDVTMHELYLWPFADAVRAG-VGSVMCSYNQINN	222
SEQ ID NO: 86	AYSANIDDRTMHELYLWPFSAESVRAG-VGAVMMAYNDVNG	204
SEQ ID NO: 87	--AAWIPERMLRRYFLPSFAEAITGAGAGTIMINSGEVNG	280
SEQ ID NO: 88	--DTIIDAHLMTIHPAYYNSIIRG-VSTVMTSYSSWNG	257
SEQ ID NO: 89	---SIVTERALREIYALPFQIAVRDSQPGAFMTAYNGING	193
SEQ ID NO: 90	AVSANIDDRTLHELYLWPFMDSVHAG-VGSVMCSYNRINN	208
SEQ ID NO: 91	---CLITQRALREVYLRFQIVARDARPGALMTSYNKVNG	67
SEQ ID NO: 92	ALSSNVDDKTIHEVYGWPFQDAVKAG-VGSFMCSYNQVNN	222
SEQ ID NO: 93	GFDAIITQQDLSEYYTPQFLAAARYAKSRSLMCAYNSVNG	225
SEQ ID NO: 94	---SIVSERALREIYLEPFRLAVKHANPVCIMTAYNKVNG	193
SEQ ID NO: 95	SVSANIDDKTMHELYLWPFQDSVRAG-LGSIMGSYNRVNN	214
SEQ ID NO: 96	--NTVISSNSLFSIHMPAYYDSLRRG-VATVMVSYSSWNG	258
SEQ ID NO: 97	AVSANLADKTLHELYLWPFADAVRAG-TGSIMCSYNKANN	221
SEQ ID NO: 98	ALSSNLDDKTLHELYLWPFADAVRAG-TGSIMCSYNKVNN	220
SEQ ID NO: 99	SSSSNVDDRTEHELYGHPFLRSVQAN-VASVMCSYNQING	209
SEQ ID NO: 100	SSSSNVDDRTEHELYGHPFLRSVQAN-VASVMCSYNQING	209
SEQ ID NO: 101	SLSSEIDDRAMHEIYLWPFSAEAVRGG-VSSIMCSYNKLNG	222
SEQ ID NO: 102	ASSSNMDDQTLMEIHVEPFYRTIKDGDAGSVMASYNVNN	210
SEQ ID NO: 103	SISANIPDRAMHELYLWPFADSIRAG-VGSVMCSYNRVNN	225
SEQ ID NO: 104	AISANIPDRAMHALYLWPFADSVRAG-VGSVMCSYNRVNN	223
SEQ ID NO: 105	---IDVSQRALREVYLMFQLACKYSNFKSLMTSYNKVNG	198
SEQ ID NO: 106	AISSNIDDRTMHELYLWPFANAVRSG-VASVMCSYNRLNQ	211
SEQ ID NO: 107	SISANADDKTMHELYLWPFADAVRAG-VGSVMCSYNQVNN	219
SEQ ID NO: 108	GYSSNIDDRTLHELYLWPFSAEVRAG-VGSVMAYNDVNG	211
SEQ ID NO: 109	SVSSNIDDKTLHELYLWPFADAVRAG-VGSFMCSYNQVNN	222
SEQ ID NO: 110	SVSSNIDDKTLHELYLWPFADAVRAG-VGSFMCSYNQVNN	222
SEQ ID NO: 111	GYSSNIDDRTLHELYLWPFSAEVRAG-VGSVMAYNDVNG	215
SEQ ID NO: 112	TISSNPDDRTLHELYTWPFADAVQAN-VASVMCSYNKVNT	209
SEQ ID NO: 113	--NTVIDSRGLFSIHMPAYHDSIKKG-VATVMVSYSSWNG	258
SEQ ID NO: 114	ASSSNIDDRTLRELYEWPFSAEAIHAG-VDYIMCSYNRVNQ	213
SEQ ID NO: 115	---SIISERALREIYLLPFMLTQKYAAPWAIMTAYNRVNG	204

Figure 11/6/2

16/20/2

SEQ ID NO: 1	T-----WACEDQYTLQTVLKDQLGF--PGYVMTDWNAQHT	242
CONSENSUS	S-----YACQNSKLLNGLLKGEELGF--QGFVMSDWGAQHS	241
SEQ ID NO: 72	S-----YGCQNSYTLNKLLKAELGF--QGFVMSDWGAHHS	255
SEQ ID NO: 73	S-----YGCQNSKTLNGLLKTELGF--QGYVMTDWGAQHA	245
SEQ ID NO: 74	S-----YGCQNSQTLNKLLKAELGF--QGFVMSDWSAHHS	256
SEQ ID NO: 75	S-----YGCQNSYTLNKLLKAELGF--QGFVMSDWAAHHA	255
SEQ ID NO: 76	TNWEAVGFSFNKEIVTDLLRGELGF--DGIVLTDWGLITD	303
SEQ ID NO: 77	T-----HADSNNFLLRDVLRGEWGW--KGLVMSDWGGTNS	228
SEQ ID NO: 78	S-----YGCQNSYTLNKLLKAELGF--QGFVMSDWAAHHA	255
SEQ ID NO: 79	S-----YGCQNSYTLNKLLKAELGF--QGFVMSDWAAHHA	255
SEQ ID NO: 80	S-----YGCENSETLNKLLKAELGF--QGFVMSDWTAHHS	255
SEQ ID NO: 81	S-----YGCENSETLNKLLKAELGF--QGFVMSDWTAHHS	255
SEQ ID NO: 82	S-----ACSQNSKLINGILKDELGF--QGFVVTDWLAHIG	248
SEQ ID NO: 83	S-----YGCQNSKAMNGLLKGEELGF--QGFVVS DWGAQHT	252
SEQ ID NO: 84	T-----YACENPHLLNNILKGEELGF--QGYVVS DWFATHS	259
SEQ ID NO: 85	S-----YGCNSYTLNKLLKGEELGF--QGFIMSDWQAHS	255
SEQ ID NO: 86	S-----ASCQNSKLINGILKDELGF--QGFVMTDWYAQIG	237
SEQ ID NO: 87	V-----PMHTSYKYLTEVLRGELQF--EGVAVTDWQDIEK	313
SEQ ID NO: 88	K-----KMHANHFLVTDFLKNKLKF--RGFVISDWQGIDR	290
SEQ ID NO: 89	V-----SCSENPKYLDGMLRKEWGW--DGLIMSDWYGYS	226
SEQ ID NO: 90	T-----YGCMNDKLMNGILKAELGF--QGFVMLDWNAQHD	241
SEQ ID NO: 91	K-----HVADSAEFLQGILRTEWNW--DPLIVSDWYGTYT	100
SEQ ID NO: 92	S-----YACQNSKLINGLLKEEYGF--QGFVMSDWQAQHT	255
SEQ ID NO: 93	V-----PSCANSFFLQTLRESWGFPEWGYVSSDCDAVYN	260
SEQ ID NO: 94	D-----HCSQSKLLIDILRDEWKW--DGMLMSDWFGTYT	226
SEQ ID NO: 95	S-----YACKNSKVLNGLLKSELGF--QGFVVS DWGGQHT	247
SEQ ID NO: 96	R-----KMHANRDLVTGFLKDKLKF--RGFVISDWQGIDR	291
SEQ ID NO: 97	S-----QVCQNSYLQNYILKGEELGF--QGFTMSDWDAQHS	254
SEQ ID NO: 98	S-----QACQNSYLQNYILKGEELGF--QGFIMSDWDAQHS	253
SEQ ID NO: 99	T-----FSCENEKTLGSLKGEYGF--QGYVMSDWWATHS	242
SEQ ID NO: 100	T-----FSCENEKTLGSLKGEYGF--QGYVMSDWWATHS	242
SEQ ID NO: 101	S-----HACQNSYLLNYLLKEELGF--QGFVMTDWGALYS	255
SEQ ID NO: 102	I-----YVVQNKKVLTEILKEGIGF--QGFVMSDWWAIHD	243
SEQ ID NO: 103	T-----YSCENSYMINHLLKEELGF--QGFVVS DWAAQMS	258
SEQ ID NO: 104	T-----YACENSYMMNHLLKEELGF--QGFVVS DWGAQLS	256
SEQ ID NO: 105	E-----HVSQSRILLDNILRKEWEW--KGTIISDWFGTYS	231
SEQ ID NO: 106	T-----YACENSKLMNGILKGEELGF--QGYVVS DWYATHS	244
SEQ ID NO: 107	S-----YACNSYTMNKLLKSELGF--QGFVMTDWGGHHS	252
SEQ ID NO: 108	S-----ACSQNSKLINGILKDELGF--QGFVMTDWLTQIG	244
SEQ ID NO: 109	S-----YSCNSYLLNKLLKSELDF--QGFVMSDWGAHHS	255
SEQ ID NO: 110	S-----YSCNSYLLNKLLKSELDF--QGFVMSDWGAHHS	255
SEQ ID NO: 111	S-----ACSQNSKLINGILKDELGF--QGFVMTDWLTQIG	248
SEQ ID NO: 112	T-----WACEDQYTLQTVLKDQLGF--PGYVMTDWNAQHA	242
SEQ ID NO: 113	L-----RMHANRDLVTGYLKNKLKF--RGFVISDWEGIDR	291
SEQ ID NO: 114	T-----YACENSKLINGIAKGEHKF--QGVMTDWAAAES	246
SEQ ID NO: 115	V-----HVAEDPFLKQVLRNEWKY--KGLIMSDWFGMYS	237

Figure 11/6/3

16/20/3

SEQ ID NO: 1	TVQSANSGLDMSMP-----G-TDFNG----NNRLW	267
CONSENSUS	GVGSALAGLDMSMP-----GD--F-----GTSYW	263
SEQ ID NO: 72	GVGSALAGLDMSMP-----GDITFDS----ATSF	281
SEQ ID NO: 73	GIAGANAGLDMVMP-----STETWG	265
SEQ ID NO: 74	GVGAALAGLDMSMP-----GDISFDD----GLSFW	282
SEQ ID NO: 75	GVSGALAGLDMSMP-----GDVDYDS----GTSYW	281
SEQ ID NO: 76	TYIGNQYMPARAWGVEYLSELQRAARILDAGCDQFGGEER	343
SEQ ID NO: 77	TADALNAGLDLEMPG-----PTRWR	248
SEQ ID NO: 78	GVSGALAGLDMSMP-----GDVDYDS----GTSYW	281
SEQ ID NO: 79	GVSGALAGLDMSMP-----GDVDYDS----GTSYW	281
SEQ ID NO: 80	GVGAALAGLDMSMP-----GDVTFDS----GTSFW	281
SEQ ID NO: 81	GVGAALAGLDMSMP-----GDVTFDS----GTSFW	281
SEQ ID NO: 82	GVSSALAGLDMSMP-----GDGAIP---LLGTSYW	275
SEQ ID NO: 83	GIASAAAGLDMAMP-----SSSYWE	272
SEQ ID NO: 84	GYPANAGLDMDMPG-----YIS--QSAINTGETYF	288
SEQ ID NO: 85	GVGDDLALAGLDMSMP-----GDTLFLT----GKSYW	281
SEQ ID NO: 86	GVSSALAGLDMSMP-----GDGSPV---LSGTSFW	264
SEQ ID NO: 87	LVYFHHTAGSAEEAIL-----QALDAGIICLCHDLLS	345
SEQ ID NO: 88	ITSPPGVNYSYSVEAG-----VGAGIDMIMVPFAYTEF	323
SEQ ID NO: 89	TTEAVVAGLDLEMPG-----PPRFR	246
SEQ ID NO: 90	-LQSANAGLDMVMP-----LGGSWG	260
SEQ ID NO: 91	TIDAIKAGLDLEMPG-----VSRYR	120
SEQ ID NO: 92	GVASAVAGLDMTMP-----GDTAFNT----GASYF	281
SEQ ID NO: 93	VFNPHDYASNQSSAAAS-----SLRAGTDIDCGQTY	291
SEQ ID NO: 94	TAAAIKNGLDIEFPG-----PTRWR	246
SEQ ID NO: 95	GIASANAGLDMAMP-----SSTYWE	267
SEQ ID NO: 96	ITDPPHANYSYSVQAG-----IMAGIDMIMVPENYREF	324
SEQ ID NO: 97	GVASTLAGLDNMMP-----GDTDFDS----GFSEFW	280
SEQ ID NO: 98	GVASTFAGLDMTMP-----GDTDFNS----GKTFW	279
SEQ ID NO: 99	GAPAVNAGLDMTMP-----GDETTNS----GTTYF	268
SEQ ID NO: 100	GAPAVNAGLDMTMP-----GDETLSS----GTTYF	268
SEQ ID NO: 101	GIDAANAGLDMDMP-----CEAQ-----YF	275
SEQ ID NO: 102	LEGSFNAGMDNMMP-----GGKAWGP-DYVNNSFW	272
SEQ ID NO: 103	GAYSAISGLDMSMP-----GELLGGW--NTGKSYW	286
SEQ ID NO: 104	GVYSAISGLDMSMP-----GEVYGGW--NTGTSFW	284
SEQ ID NO: 105	LKKAIDAGLDLEMPG-----KPRFR	251
SEQ ID NO: 106	GVESVNAGLDMTMPG-----PLDSPSTALRPPPSYL	275
SEQ ID NO: 107	GVGSALAGLDMSMP-----GDIAFDS----GTSFW	278
SEQ ID NO: 108	GVSSALAGLDMAMP-----GDGPIP---LFGDSYW	271
SEQ ID NO: 109	GVGAALAGLDMSMP-----GDTAFGT----GKSFW	281
SEQ ID NO: 110	GVGAALAGLDMSMP-----GDTAFGT----GKSFW	281
SEQ ID NO: 111	GVSSALAGLDMAMP-----GDGPIP---LFGDSYW	275
SEQ ID NO: 112	TVQSANSGLDVSMMP-----G-TDFNG----NNRLW	267
SEQ ID NO: 113	ITDPPGRNYSYSVEAG-----VGAGIDMIMVPEDFTKF	324
SEQ ID NO: 114	GVRTALAGTDMNMPGFMA-----YGQPSEPNPSTANGSYW	281
SEQ ID NO: 115	VDHGIKAGLDLEMPG-----INKWR	257

Figure 11/6/4

16/20/4

SEQ ID NO: 1	GPALTNAVNSNQVPTSR--VDDMVTRILAAWYLTGQDQ--	303
CONSENSUS	GTNLTNVAVLNGTVPQWR--LDDMATRILAAYYKVGQD--D	299
SEQ ID NO: 72	GTNLTI AVLNGTVPQWR--VDDMAVRIMAAYYKVGRDR-L	318
SEQ ID NO: 73	AN-LTTAISNGTMDASR--LDDMATRIIASWYQMNQD-SD	301
SEQ ID NO: 74	GTNLTVSVLNGTVPQWR--VDDMAVRIMTAYYKVGRDR-L	319
SEQ ID NO: 75	GTNLTVSVLNGTVPQWR--VDDMAVRIMAAYYKVGRDR-L	318
SEQ ID NO: 76	PELIVQLVREGTISED-R--IDVSVARLLKEKFLGLGFDNP	381
SEQ ID NO: 77	KVDEVLA VVKSGAVLEE-TIDERARNVLELLAKLN-----	282
SEQ ID NO: 78	GTNLTI SVLNGTVPQWR--VDDMAVRIMAAYYKVGRDR-L	318
SEQ ID NO: 79	GTNLTI SVLNGTVPQWR--VDDMAVRIMAAYYKVGRDR-L	318
SEQ ID NO: 80	GANLTVGV LNGTIPQWR--VDDMAVRIMAAYYKVGRDT-K	318
SEQ ID NO: 81	GANLTVGV LNGTIPQWR--VDDMAVRIMAAYYKVGRDT-K	318
SEQ ID NO: 82	SWELSR SVLNGSVPVER--LNDMVTRIVATWYKMGQDK-D	312
SEQ ID NO: 83	NGTLALAVKNESLPSTR--LDDMATRIVATWYKY----AE	306
SEQ ID NO: 84	GPHLISAIQAGNMTEDR--LDDMVTRIMTSYFLLNQS-SN	325
SEQ ID NO: 85	GPNLTI AVTNGTIPQWR--LDDMAVRIMAAYYKVRRDQ-T	318
SEQ ID NO: 86	ASELSRSILNGTVALDR--LNDMVTRIVATWFKFGQDK-D	301
SEQ ID NO: 87	QLFSLEILAAGTVPESR--LDLSVRRILNLKYALGLFSNP	383
SEQ ID NO: 88	IDDLTYQVKNNIIPMSR--INDAVYRILRVKFTMGLFESP	361
SEQ ID NO: 89	GETLKFNVSNGKPFH--VIDQRAREVLQFVKKCAA----	280
SEQ ID NO: 90	KN-LTDAVANGTVSESR--ITDMATRIIAAWYLVGQDGNN	297
SEQ ID NO: 91	-GKYIESALQARLLKQS-TIDERARRVLRFAQKASH--LK	156
SEQ ID NO: 92	GSNLTLAVLNGTVPQWR--IDDMVMRIMAPFFKVGKTVDS	319
SEQ ID NO: 93	PWHLNESFVAGEVSRGE--IERSVTRLVYANLVRLGYFDKK	329
SEQ ID NO: 94	TRALVSHSLNSREQITTEDVDDRVRQVLKMIKFVVDNLEK	286
SEQ ID NO: 95	EG-LIEAVKNGTVDQSR--LDDMATRIIAAWYKY----AR	300
SEQ ID NO: 96	IDTLTSQVKANIIPMSR--IDDAVKRILRVKFVMGLFENP	362
SEQ ID NO: 97	GPNMTLSIINGTVPEWR--LDDAATRIMAAYYLVGRDR-H	317
SEQ ID NO: 98	GTNFTTSILNGTVPQWR--LDDAVTRIMAAFYVGRDK-A	316
SEQ ID NO: 99	GQNLVNAVNSGQVSQAR--IKDMATRILAAWYLLGQDQNF	306
SEQ ID NO: 100	GQNLVNAVNSGQVSQAR--VKDMATRILAAWYLLGQDQNF	306
SEQ ID NO: 101	GGNLTTAVLNGTLPQDR--LDDMATRILSALIYSGVHNPD	313
SEQ ID NO: 102	GSNISNAIRSGQVSSSR--LDDAVRRIIRTLYRFDQMS-G	309
SEQ ID NO: 103	GQNLTKAVYNETVPIER--LDDMATRILAAALYATNSFPTK	324
SEQ ID NO: 104	GQNLTKAIYNETVPIER--LDDMATRILAAALYATNSFPTK	322
SEQ ID NO: 105	NVNTIQHLVGSKELSES-ILDERAKNV LKLVKHS-----WQ	286
SEQ ID NO: 106	GGNLTEAVLNGTIPPEAR--VDDMARRILMPYFFLGQD-TD	312
SEQ ID NO: 107	GTNLTVAVLNGSIPQWR--VDDMAVRIMSAYYKVGRDR-Y	315
SEQ ID NO: 108	GSELSRAVLNGTVPVDR--LNDMVTRIVATWYKFGQDK-D	308
SEQ ID NO: 109	GTNLTI AVLNGTVPQWR--VDDMAVRIMAAFYKVGRDR-Y	318
SEQ ID NO: 110	GTNLTI AVLNGTVPQWR--VDDMAVRIMAAFYKVGRDR-Y	318
SEQ ID NO: 111	GSELSRAVLNGTVPVDR--LNDMVTRIVATWYKFGQDK-D	312
SEQ ID NO: 112	GPALTNAVNSNQVPTSR--VDDMVTRILAAWYLTGQDQ--	303
SEQ ID NO: 113	LNELTSQVKNNIIPMSR--IDDAVKRILRVKFVMGLFESP	362
SEQ ID NO: 114	GLRMIEAVKNGTVPMER--LDDMVTRVISTYYKQGQDKSD	319
SEQ ID NO: 115	TLDLVNRTIQARKLTPR-DIKDRARVVLELVKKCAQG--A	294

Figure 11/6/5

16/20/5

SEQ ID NO: 1	---AGYPSFNISR-----NVQ	316
CONSENSUS	YPPPNFSSWTRDEYGYEY-----VNQ-VDVQ	324
SEQ ID NO: 72	YQPPNFSSWTRDEYGFKYFYYPQEGP----YEKVNHFVNVQ	354
SEQ ID NO: 73	FPSPGAGMPSDMYAPHQR-----VIGRDAS	326
SEQ ID NO: 74	RIPPNFSSWTRDEYGWEHSAVSEGA----WTKVNDFVNVQ	355
SEQ ID NO: 75	WTPPNFSSWTRDEYGYKYYYVSEGP----YEKVNHYVNVQ	354
SEQ ID NO: 76	FVNASAANNIVGNEHFVN-----	399
SEQ ID NO: 77	CFENPTIPEEKAINRPEHQKLIRSVGSQGLVLLKNEGDVL	322
SEQ ID NO: 78	WTPPNFSSWTRDEYGYKYYYVSEGP----YEKVNQYVNVQ	354
SEQ ID NO: 79	WTPPNFSSWTRDEYGFKYFYYPQEGP----YEKVNQFVNVQ	354
SEQ ID NO: 80	YTPPNFSSWTRDEYGFAHNHVSEGA----YERVNEFVDVQ	354
SEQ ID NO: 81	YTPPNFSSWTRDEYGFAHNHVSEGA----YERVNEFVDVQ	354
SEQ ID NO: 82	YPLPNFSSNTEDETGPLYPGALFSP----SGIVNQYVNVQ	348
SEQ ID NO: 83	IENPGHGLPYSLAPHNL-----TDARDPK	331
SEQ ID NO: 84	YPTPDPSQTYVMANMYGY-----DYGAQIPARDVR	355
SEQ ID NO: 85	QVPINFNSWTRDEFGYLHAGGQEG-----YGRVNQMVNVR	353
SEQ ID NO: 86	FPLPNFSSYTQNAKGLLYPGALFSP----LGVVNQFVNVQ	337
SEQ ID NO: 87	YPNPNAAIVDITIGQVQD-----	400
SEQ ID NO: 88	YADPSLVGELGKQEHF-----	377
SEQ ID NO: 89	SGVTENGPETTVNNTPETAALLRKVGNEGIVLLKNENNVL	320
SEQ ID NO: 90	FPVPGIGLK-QLTKPHEQ-----VDARDPA	321
SEQ ID NO: 91	VSEVEQGRD-FPED----RVLNRQICGSSIVLLKNENSIL	191
SEQ ID NO: 92	LIDTNFDSWTNGEYGYVQAAVNEN----WEKVNYGVDVR	354
SEQ ID NO: 93	NQYRSLGWKDVVKTDWN-----	347
SEQ ID NO: 94	TGIVENGPESTSNTKETSDLLREIAADSIVLLKNKNNYL	326
SEQ ID NO: 95	LDDPGFGMPVSLAEDHEL-----VDARDPA	325
SEQ ID NO: 96	MSDPSLANQLGSQEHF-----	378
SEQ ID NO: 97	AVPVNFNSWSKDTYGYQHAYAKVG-----YGLINQHVVDVR	352
SEQ ID NO: 98	RIPVNFDSWSRDTYGFHDHYGKAG-----YSQINSHVDVR	351
SEQ ID NO: 99	PA-VNFNSWNSGQG-----QHVNVS	325
SEQ ID NO: 100	PA-VNFNSWNSGQG-----QHVNVS	325
SEQ ID NO: 101	G--PNYNAQTFLTEGHEYFKQQEGD----IVVLNKHVDVR	347
SEQ ID NO: 102	YPNVNLKAPSMHADTN-----	325
SEQ ID NO: 103	DRLPNFSSFTTKEYGNEFFVDKTSP----VVKVNHFVDPS	360
SEQ ID NO: 104	DHLPNFSSWTTKEYGNKYADNTTE----IVKVNYNVDPS	358
SEQ ID NO: 105	NTEAENHCE-LNNDSSCLREALKKFASQSIVLLKNKKKLL	325
SEQ ID NO: 106	FPTVDPSTGFVFARTYNYPDEYLTGGLDPYNPPPARVDVR	352
SEQ ID NO: 107	SVPINFDSWTLDTYGPEHYAVGQG-----QTKINEHVVDVR	350
SEQ ID NO: 108	FPLPNFSSNTDAATGLLYPGAVFSP----IGVVNQFVDVQ	344
SEQ ID NO: 109	QVPVNFDSWTKDEYGYEHALVGQN-----YVKVNDKVDVR	353
SEQ ID NO: 110	QVPVNFDSWTKDEHGYEHALVGQD-----YVKVNDKVDVR	353
SEQ ID NO: 111	FPLPNFSSNTDAATGLLYPGAVFSP----IGVVNQFVDVQ	348
SEQ ID NO: 112	---AGYPSFNISR-----NVQ	316
SEQ ID NO: 113	LADYSLANQLGSQEHF-----	378
SEQ ID NO: 114	YPKLNFMMSGQGTP-----AE----QAVSNHHVNVQ	346
SEQ ID NO: 115	PEILDGDGEERTVELESDKLLMRRIASESIVLLKNDN-VL	333

Figure 11/6/6

16/20/6

SEQ ID NO: 1	GNH-----KTNVRAIARDGIVLLKN-DANIL	341
CONSENSUS	GNH-----ADLIRRIAAASTVLLKN-DNALP	349
SEQ ID NO: 72	RNH-----SEVIRKLGADSTVLLKN-NNALP	379
SEQ ID NO: 73	S-----KQTLRLGAIEGHVLVKNHNSALP	350
SEQ ID NO: 74	RSH-----SQIIREIGAASTVLLKN-TGALP	380
SEQ ID NO: 75	RNH-----SELIRRIGADSTVLLKN-DGALP	379
SEQ ID NO: 76	-----LGRDAQRRSYTLLTNNQITLP	420
SEQ ID NO: 77	PLRKEILTNKKVALLG-FAREALIHGGGSASVNAHYRVTP	361
SEQ ID NO: 78	RNH-----SELIRRIGADSTVLLKN-DGALP	379
SEQ ID NO: 79	RNH-----SELIRRIGADSTVLLKN-DGALP	379
SEQ ID NO: 80	RDH-----ADLIRRIGAQSTVLLKN-KGALP	379
SEQ ID NO: 81	RDH-----ADLIRRIGAQSTVLLKN-KGALP	379
SEQ ID NO: 82	GNH-----NVTARAIARDAITLLKNNENVLP	374
SEQ ID NO: 83	S-----KSTILQGAVEGHVLVKNNTNNALP	355
SEQ ID NO: 84	ANH-----SSLIRQIGSAGTVLLKNTNNILP	381
SEQ ID NO: 85	GRH-----AVIARKVASASTVLLKN-RGVLP	378
SEQ ID NO: 86	ADH-----HKLARVIARESITLLKNEDNLLP	363
SEQ ID NO: 87	-----REAAAATAEESITLLLFKNNILP	423
SEQ ID NO: 88	-----DLAREAVRKSLVLLKNGKSAST	399
SEQ ID NO: 89	PLSK----KKKTLIVGPNAKQATYHGGGSAAALRAYYAVTP	356
SEQ ID NO: 90	S-----KPVLLLEGAIAGHVLVKNENNALP	345
SEQ ID NO: 91	PLPKS---VKKVALVGSHVRLPAISGGGSASLVPYYAISL	228
SEQ ID NO: 92	ANH-----ANHIREVGAKGTVIFKN-NGILP	379
SEQ ID NO: 93	-----ISYEAAVEGIVLLKN-DGTLP	367
SEQ ID NO: 94	TSKE----RRQYHVIGPNAKAKTSSGGGSASMNSYYVVSP	362
SEQ ID NO: 95	A-----ASTIFQGAVEGHVLVKNEN-ALP	348
SEQ ID NO: 96	-----ELAREAVRKSLVLLKNGKTPSQ	400
SEQ ID NO: 97	ADH-----FKSIRTAATAKSTVLLKN-NGVLP	377
SEQ ID NO: 98	ADH-----FRSIRRTAAMSTVLLKN-EGALP	376
SEQ ID NO: 99	GNH-----ASLIRTIGAASQILLKN-VNGAL	350
SEQ ID NO: 100	GNH-----ASLIRTIGAASQILLKN-VNSAL	350
SEQ ID NO: 101	SDIN-----RAVALRSAVEGVLLKNEHETLP	374
SEQ ID NO: 102	-----RQAAIESSVLLKNADDILP	344
SEQ ID NO: 103	NDFT-----EDTALKVAEESIVLLKNEKNTLP	387
SEQ ID NO: 104	NDFT-----EDTALKVAEESIVLLKNNENTLP	385
SEQ ID NO: 105	PLSR----KGTFAVIGPNAKVCNYSGGGSANLKPYYTVSM	361
SEQ ID NO: 106	GNH-----SDIVRKVAAAGTVLLKNVNNVLP	378
SEQ ID NO: 107	GNH-----AEIIHEIGAASAVLLKN-KGGLP	375
SEQ ID NO: 108	GDH-----KVVARAIARDAITLLKNEDNALP	370
SEQ ID NO: 109	ADH-----ADIIRQIGSASVLLKN-DGGLP	378
SEQ ID NO: 110	ADH-----ADIIRQIGSASVLLKN-DGGLP	378
SEQ ID NO: 111	GDH-----KVVARAIARDAITLLKNEDNALP	374
SEQ ID NO: 112	GNH-----KTNVRAIARDGIVLLKN-DANIL	341
SEQ ID NO: 113	-----DLAREAVRKSLVLLKNGESADK	400
SEQ ID NO: 114	KDH-----YLIIRQIATASTILLKNVNHTLP	372
SEQ ID NO: 115	PLKPEGGAIKKIAVVGGNAKAQVLSGGGSAAALKASYFISP	373

Figure 11/6/7

16/20/7

SEQ ID NO: 1	PLKKPASIAVVGSAAIIGNHARNS-----	365
CONSENSUS	LKGKEKKIAVFGED--AGSNPYGA-----	371
SEQ ID NO: 72	LTGKERKVAILGED--AGSNSYGA-----	401
SEQ ID NO: 73	LKSPQ-LLSVFGYDAKGPNAALKQN-----	373
SEQ ID NO: 74	LTGKEVKVGVLGED--AGSNPWGA-----	402
SEQ ID NO: 75	LTGKERLVALIGED--AGSNPYGA-----	401
SEQ ID NO: 76	LAKPGEGRFYIEGFDSAFAFMSARN-----	444
SEQ ID NO: 77	EEGLRAALGDT-VEFEYAKGAHTFRQLPLMSDNVVNLEGQ	400
SEQ ID NO: 78	LTGKERLVALIGED--AGSNPYGA-----	401
SEQ ID NO: 79	LTGKERLVALIGED--AGSNPYGA-----	401
SEQ ID NO: 80	LSRKEKLVALLGED--AGSNSWGA-----	401
SEQ ID NO: 81	LSRKEKLVALLGED--AGSNSWGA-----	401
SEQ ID NO: 82	LKRND---TLKIFGTDAGTNSDGI-----	395
SEQ ID NO: 83	LKKPQ-FLSLFGYDAVAAARNTMDDLWNMWSMGYDNSLT	394
SEQ ID NO: 84	LSKPL-NIGVFGNSAPDPTDG-----	401
SEQ ID NO: 85	LKGKEKLTAVIGED--AGPNLWGP-----	400
SEQ ID NO: 86	LDPNR---AIKYSEQMPGTNPRGI-----	384
SEQ ID NO: 87	LNTNTIKNVLLTGPSADSIRNLNG-----	447
SEQ ID NO: 88	PLLPLPKKAGKILVAGSHADDLGN-----	423
SEQ ID NO: 89	FDGLSKQLET---PPSYTVGAYTTVPP-ILGEQCLTPDGA	392
SEQ ID NO: 90	FNKKLTMISVFGYDATIPRTKNTDILFQLGYTSSPE----	381
SEQ ID NO: 91	YDAVSEVLAGA--TITHEVGAYAHQMLPVIDAMISN----	262
SEQ ID NO: 92	LK-KPKFLTIVIGED--AGGNPAGP-----	400
SEQ ID NO: 93	LSKKVRSIALIGPWANATTQMQGN-----	391
SEQ ID NO: 94	YEGIVNKL GK---EVDYTVGAYSHKSIGGLAESSLIDAAK	399
SEQ ID NO: 95	LKKPK-YISLFGYDGVSTDVNTVG-----	371
SEQ ID NO: 96	PLLPLPKKAPKILVAGTHADNLGY-----	424
SEQ ID NO: 97	LKGTEKYTAVFGND--AGEAQYGP-----	399
SEQ ID NO: 98	LTGSEKWTAVFGDD--AGEGQLGP-----	398
SEQ ID NO: 99	PLKKPKTIGIIGNG--AGSNPSGP-----	372
SEQ ID NO: 100	PLKKPKTIGIIGNG--AGSNPNGP-----	372
SEQ ID NO: 101	LG-REKVKRISILGQAAGDDSKG-----	396
SEQ ID NO: 102	LTKKYRKIAIIGKDADKAQS-----	364
SEQ ID NO: 103	IS-PNKVRKLLLSGIAAGPDPKG-----	409
SEQ ID NO: 104	IS-PEKAKRLLLSGIAAGPDPIG-----	407
SEQ ID NO: 105	YDGIAAKIDG---VPEYALGCHNYLNLPNIANLLVNPRTG	398
SEQ ID NO: 106	LKEPK-SVGIFGNGAADVTEG-----	398
SEQ ID NO: 107	LTGTERFVGVFGKD--AGSNPWGV-----	397
SEQ ID NO: 108	LKRND---SLKIFGTDAGTNPdGI-----	391
SEQ ID NO: 109	LTGYEKFTGVFGED--AGSNRWGA-----	400
SEQ ID NO: 110	LTGYEKFTGVFGED--AGSNRWGA-----	400
SEQ ID NO: 111	LKRND---SLKIFGTDAGTNPdGI-----	395
SEQ ID NO: 112	PLKKPASIAVVGSAAIIGNHARNS-----	365
SEQ ID NO: 113	PFVPLPKNAKKILVAGSHADNLGR-----	424
SEQ ID NO: 114	LKSPDKMRSVVVVGSDAGDNPQGP-----	396
SEQ ID NO: 115	YDGIKAALPHGVEVTFSEGARAYKTLPTLEWDLETETGE	413

Figure 11/6/8

16/20/8

SEQ ID NO: 1	-----	365
CONSENSUS	-----	371
SEQ ID NO: 72	-----	401
SEQ ID NO: 73	-----	373
SEQ ID NO: 74	-----	402
SEQ ID NO: 75	-----	401
SEQ ID NO: 76	-----	444
SEQ ID NO: 77	P-----GWTLDFFADEEP---NGEPGSSISSEQPSYIPL	431
SEQ ID NO: 78	-----	401
SEQ ID NO: 79	-----	401
SEQ ID NO: 80	-----	401
SEQ ID NO: 81	-----	401
SEQ ID NO: 82	-----	395
SEQ ID NO: 83	Y-----	395
SEQ ID NO: 84	-----	401
SEQ ID NO: 85	-----	400
SEQ ID NO: 86	-----	384
SEQ ID NO: 87	-----	447
SEQ ID NO: 88	-----	423
SEQ ID NO: 89	P-----GMRWRVFNEPP---GTPNRQHIDELFFTKTDM	422
SEQ ID NO: 90	-----	381
SEQ ID NO: 91	-----AVIHFYNDPI---DVKDRKLLGSENVSTSF	290
SEQ ID NO: 92	-----	400
SEQ ID NO: 93	-----	391
SEQ ID NO: 94	PADAENAGLIAKFYSNPVEE--RSEDEEPFHVTKVNRSNV	437
SEQ ID NO: 95	-----	371
SEQ ID NO: 96	-----	424
SEQ ID NO: 97	-----	399
SEQ ID NO: 98	-----	398
SEQ ID NO: 99	-----	372
SEQ ID NO: 100	-----	372
SEQ ID NO: 101	-----	396
SEQ ID NO: 102	-----	364
SEQ ID NO: 103	-----	409
SEQ ID NO: 104	-----	407
SEQ ID NO: 105	K-----HGYVAKFYLEPA---TSENRTLIDDYDLEDGVV	429
SEQ ID NO: 106	-----	398
SEQ ID NO: 107	-----	397
SEQ ID NO: 108	-----	391
SEQ ID NO: 109	-----	400
SEQ ID NO: 110	-----	400
SEQ ID NO: 111	-----	395
SEQ ID NO: 112	-----	365
SEQ ID NO: 113	-----	424
SEQ ID NO: 114	-----	396
SEQ ID NO: 115	R-----GWIGTWHTHESDDSM TALDQPFIA PRLVDETRI	447

Figure 11/6/9

16/20/9

SEQ ID NO: 1	-----	365
CONSENSUS	-----	371
SEQ ID NO: 72	-----	401
SEQ ID NO: 73	-----FNWLS-----	378
SEQ ID NO: 74	-----	402
SEQ ID NO: 75	-----	401
SEQ ID NO: 76	-----	444
SEQ ID NO: 77	FVKESWG-----SVRASAHFTPTQ-SGKHYFGMSG	460
SEQ ID NO: 78	-----	401
SEQ ID NO: 79	-----	401
SEQ ID NO: 80	-----	401
SEQ ID NO: 81	-----	401
SEQ ID NO: 82	-----	395
SEQ ID NO: 83	-----P-----NGSAVDAMMLKYIFLSSANP	416
SEQ ID NO: 84	-----LTWPQSD--	408
SEQ ID NO: 85	-----	400
SEQ ID NO: 86	-----	384
SEQ ID NO: 87	-----	447
SEQ ID NO: 88	-----	423
SEQ ID NO: 89	HLVDYYHPKAADT---WYADMEGTYTADE-DCTYELGLVV	458
SEQ ID NO: 90	-----MAQAVLGN---	389
SEQ ID NO: 91	QLMDYNNIPTLNKA-MFWGTLVGEFIPTA-TGIWEFGLSV	328
SEQ ID NO: 92	-----	400
SEQ ID NO: 93	-----	391
SEQ ID NO: 94	HLFDFKHEKVDPKNPYFFVTLTGQYVPQE-DGDYIFSLQV	476
SEQ ID NO: 95	-----GG-----FSFFS-----	378
SEQ ID NO: 96	-----	424
SEQ ID NO: 97	-----	399
SEQ ID NO: 98	-----	398
SEQ ID NO: 99	-----	372
SEQ ID NO: 100	-----	372
SEQ ID NO: 101	-----	396
SEQ ID NO: 102	-----	364
SEQ ID NO: 103	-----	409
SEQ ID NO: 104	-----	407
SEQ ID NO: 105	RFYDYCN-DKMKDG-YFYIDIEGYLIPDE-DAVYEFGISV	466
SEQ ID NO: 106	-----LTFTGDDS-	406
SEQ ID NO: 107	-----	397
SEQ ID NO: 108	-----	391
SEQ ID NO: 109	-----	400
SEQ ID NO: 110	-----	400
SEQ ID NO: 111	-----	395
SEQ ID NO: 112	-----	365
SEQ ID NO: 113	-----	424
SEQ ID NO: 114	-----	396
SEQ ID NO: 115	FISTSYPKGITKR---WTMRLKGYLKPREKDTNFEFGLIA	484

Figure 11/6/10

16/20/10

SEQ ID NO: 1	PSCNDKGCDDGALGMWGSGAVNYP-YFVAPYDAINT-RA	403
CONSENSUS	NGCSDRGCDNGTLAMWGSGTANFP-YLVTPEEAIQN--V	408
SEQ ID NO: 72	NGCSDRGCDNGTLAMAWGSGTAEFP-YLVTPEQAIQA-EV	439
SEQ ID NO: 73	--YSPAIQENHTLWVGGSANNAA-YIDAPIDAIQR--Q	413
SEQ ID NO: 74	NGCPDRGCDNGTLAMAWGSGTANFP-YLVTPEQAIQR-EV	440
SEQ ID NO: 75	NGCSDRGCDNGTLAMWGSGTANFP-YLVTPEQAIQN-EV	439
SEQ ID NO: 76	-----YTVVNTTEEADFALLRYNAPYEPRN---	469
SEQ ID NO: 77	LGRSKLLIDGEVIYEQKLNCPDSMGFLLGGVEEPEIQYSF	500
SEQ ID NO: 78	NGCSDRGCDNGTLAMWGSGTANFP-YLVTPEQAIQN-EV	439
SEQ ID NO: 79	NGCSDRGCDNGTLAMWGSGTANFP-YLVTPEQAIQN-EV	439
SEQ ID NO: 80	NGCDDRGCDDGTLAMAWGSGTANFP-YLVTPEQAIQN-EV	439
SEQ ID NO: 81	NGCDDRGCDDGTLAMAWGSGTANFP-YLVTPEQAIQN-EV	439
SEQ ID NO: 82	NSCTDKGCNKGVLTMGWGSGTSRLP-YLITPQEAIAN--I	432
SEQ ID NO: 83	SAFGPGVALNATTITGGGSGASTAS-YIDAPFNAFQR--Q	453
SEQ ID NO: 84	---LQTGFDIGTLDIGGGSGSARHT-TLISPLTALRT--R	442
SEQ ID NO: 85	NGCPDRGCANGTLAMWGSGTADFP-YLVTPEQAIEN-EV	438
SEQ ID NO: 86	NACPDKGCNKGVLTMGWGSGTSNLP-YLVTPEDAIRN--I	421
SEQ ID NO: 87	-----GWSVHWQGAIEDSEFPFGTSILTGLREIT	476
SEQ ID NO: 88	-----QCGGWTITWQGQTGNDKTAGTTILSAIKS---	452
SEQ ID NO: 89	CGTAKAYVDDQLVVDNATKQVPGDAFFGSATREETGRINL	498
SEQ ID NO: 90	EAHFDQAAKGGTIMTGGRAGANAPS-YIDDPLAAIQR--R	426
SEQ ID NO: 91	FGTADLYIDNELVIENTTHQTRGTAFFGKGTTEKVATRRM	368
SEQ ID NO: 92	NGCGDRGCDDGTLAMEWGSGTTNFP-YLVTPEAALQS-QA	438
SEQ ID NO: 93	-----YYGPAPYLISPLEAAKKAGY	411
SEQ ID NO: 94	YGSGLFYLNDELIIDQKHNQERGSFCFGAGTKERTKKLTL	516
SEQ ID NO: 95	--FDVKAIENKTILISGGGSGTNTPS-YVDAPFNAFVA--K	413
SEQ ID NO: 96	-----QCGGWTIEWQGVAGNDLTIGTTILTAIKK---	453
SEQ ID NO: 97	NGCADHGCDNGTLAMWGSGTADYP-YLVTPELAIKR-TV	437
SEQ ID NO: 98	NGFPDHGGNNGTLAMWGSGTSDYP-YLVTPLESIKA-TV	436
SEQ ID NO: 99	NAFSDRAGDVGV LALGWGSGTANFP-YLVAPVDAITA-RA	410
SEQ ID NO: 100	NAFSDRAGDVGV LALGWGSGTANFP-YLVAPVDAITA-RA	410
SEQ ID NO: 101	TSCSLRGCGSGAIGTG YGSG-AGTF SYFVTPADGIGA--R	433
SEQ ID NO: 102	--CTDTACSGGNI IQWGSGTTDFT-GISDPITAIKN-RA	400
SEQ ID NO: 103	YECSDQSCVDGALFEGWGSGSVGYPKYQVTPFEEISA--N	447
SEQ ID NO: 104	YQCEDQSC TNGALFQWGSGSVGSPKYQVTPFEEISY--L	445
SEQ ID NO: 105	FGTALLFIDDVLLIDNKTKQTPNHTFEFGTIEERNSIYL	506
SEQ ID NO: 106	---GPWGADIGALSVGGGSGAGRHT-HLVSPLAIRK--R	440
SEQ ID NO: 107	NGCSDRGCDNGTLAMWGSGTANFP-YLVTPEQAIQR-EV	435
SEQ ID NO: 108	NSCADKGCCKGVLTMGWGSGTSKLP-YLNTPEQAIAN--A	428
SEQ ID NO: 109	DGCSDRGCDNGTLAMWGSGTADFP-YLVTPEQAIQN-EI	438
SEQ ID NO: 110	DGCSDRGCDNGTLAMWGSGTADFP-YLVTPEQAIQN-EI	438
SEQ ID NO: 111	NSCADKGCCKGVLTMGWGSGTSKLP-YLNTPEQAIAN--A	432
SEQ ID NO: 112	PSCNDKGCDDGALGMWGSGAVNYP-YFVAPYDAINT-RA	403
SEQ ID NO: 113	-----QCGGWTIEWQGVNGNDLTGTTLILNAIKK---	453
SEQ ID NO: 114	NSCVDRGCNRRGILAIWGSGTANFA-HLTAPATSIQNYLL	435
SEQ ID NO: 115	AGRAKLWVDGQLVIDNWTRQRRGEAFFGSGSQEETGVYLL	524

Figure 11/6/11

16/20/11

SEQ ID NO: 1	SSQGTQVTLSEN	TDN-----TSSGAS-	423	
CONSENSUS	LSN-KGNVSSV	TDN-----AL-QIE	426	
SEQ ID NO: 72	LKH-KGSVYAI	TDN-----WALSQVE	459	
SEQ ID NO: 73	AYEDGTSVLYD	IS-----SED	429	
SEQ ID NO: 74	ISN-GGNVFAV	TDN-----GALSQMA	460	
SEQ ID NO: 75	LKN-KNGVFTAT	DN-----WAIDQIE	459	
SEQ ID NO: 76	---GTFEANFH	AGS-----	480	
SEQ ID NO: 77	EAGKTYAVEVVS	VKP----TSKGGLALLDGF	IGFRLGFM	536
SEQ ID NO: 78	LKH-KNGVFTAT	DN-----WAIDQIE	459	
SEQ ID NO: 79	LKN-KNGVFTAT	DN-----WAIDQIE	459	
SEQ ID NO: 80	LQG-RGNVFAVT	DS-----WALDKIA	459	
SEQ ID NO: 81	LQG-RGNVFAVT	DS-----WALDKIA	459	
SEQ ID NO: 82	SSNAEFHITDT	FPL-----GVT----	449	
SEQ ID NO: 83	AYDDDTFLAWD	FA-----SQN	469	
SEQ ID NO: 84	AATY-ARVQYL	TNNA-----LLSSSSSSSF	466	
SEQ ID NO: 85	ITKGVGEAMSV	FDN-----YATSQIE	459	
SEQ ID NO: 86	SKNTEFHITDK	FPN-----NVQ----	438	
SEQ ID NO: 87	NDTADFNIQYT	IG-----HEI	492	
SEQ ID NO: 88	TVDPSTEVVFS	ENP-----DSA	469	
SEQ ID NO: 89	VKGNTYKFKIE	FGSA-PTYTLKGD-TIVPGHGS	LRVGGCK	536
SEQ ID NO: 90	ARKDDTWVNWD	LD-----SFN	442	
SEQ ID NO: 91	VAGSTYKLRLE	FGSA-NTTKMETTG	VVNFGGGAVHLGAC	407
SEQ ID NO: 92	LQD-GTRYESIL	SN-----YAISQIQ	458	
SEQ ID NO: 93	HVNFELGTEIA	GN-----TT	427	
SEQ ID NO: 94	KKGQVYNVRVE	YSG-PTSGLVG----	EFGAGGFQAGVIK	551
SEQ ID NO: 95	AREDNTFLSWD	FT-----SAE	429	
SEQ ID NO: 96	TVDPSTQVVYQ	QNP-----DAN	470	
SEQ ID NO: 97	GDH-GGVIASV	TDN-----YAFSQIM	457	
SEQ ID NO: 98	AQN-GGIVTSV	TDN-----WAYTQIQ	456	
SEQ ID NO: 99	SQDGTTVSSSL	SDT-----DLTGAAN	431	
SEQ ID NO: 100	SQDGTTVSSSL	SDT-----DLTGAAN	431	
SEQ ID NO: 101	AQQEKISYEF	IGDS-----WNQAAAM	454	
SEQ ID NO: 102	SKEGISIVSSI	SDS-----ANEG	418	
SEQ ID NO: 103	ARKNKMQFDY	IRES-----FDLTQVS	468	
SEQ ID NO: 104	ARKNKMQFDY	IRES-----YDLAQVT	466	
SEQ ID NO: 105	KKGRKYNVRVE	YSGA-ATYTLSTN-LSPSTGG	RYSIGCVK	544
SEQ ID NO: 106	TESVGGRVQYL	LSNS-----RIVN---DDF	462	
SEQ ID NO: 107	LSR-NGTFTGI	TDN-----GALAEMA	455	
SEQ ID NO: 108	SSNAEFFVTDS	FP-----NVN----	445	
SEQ ID NO: 109	LSKGKGLVSAV	TDN-----GALDQME	459	
SEQ ID NO: 110	LSKGKGLVSAV	TDN-----GALDQME	459	
SEQ ID NO: 111	SSNAEFFVTDS	FP-----NVN----	449	
SEQ ID NO: 112	SSQGTQVTLSEN	TDN-----TSSGAS-	423	
SEQ ID NO: 113	TVDPPTQVIYN	ENP-----DSN	470	
SEQ ID NO: 114	QSNPTITYRSI	FDD-----YAYDEIA	456	
SEQ ID NO: 115	KAGKKHEIYVE	YCNVRAPADGDEDEA	IMDSNPGVRLGGAE	564

Figure 11/6/12

16/20/12

SEQ ID NO: 1	AARGKD--VAIVFITADSGEGYITVEGNAGDR-----	453
CONSENSUS	AVASQA-DVAIVFVNADSGEGYITVDGNEGDR-----	457
SEQ ID NO: 72	TLAKQA-SVSLVFNADSGEGYISVDGNEGDR-----	490
SEQ ID NO: 73	PEVDPTTDACLVFINSYATEG-----WDR-----	453
SEQ ID NO: 74	DVASQS-SVSLVFNADSGEGFISVDGNEGDR-----	491
SEQ ID NO: 75	ALAKTA-SVSLVFNADSGEGYINVDGNLGDR-----	490
SEQ ID NO: 76	-----	480
SEQ ID NO: 77	EEEHNRDLLSEAVDVAKRSDIAIVFTGHTPDWETEGQDQI	576
SEQ ID NO: 78	ALAKTA-SVSLVFNADSGEGYINVDGNLGDR-----	490
SEQ ID NO: 79	ALAKTA-SVSLVFNADSGEGYINVDGNLGDR-----	490
SEQ ID NO: 80	AAARQA-SVSLVFNADSGEGYLSVDGNEGDR-----	490
SEQ ID NO: 81	AAARQA-SVSLVFNADSGEGYLSVDGNEGDR-----	490
SEQ ID NO: 82	--AGPD-DIAIVFINSDSGENYITVDGNPGDRTL-----	480
SEQ ID NO: 83	PLVNPASDACIVFINEQSSEG-----WDR-----	493
SEQ ID NO: 84	AGIYPIPEICLVFLKTFSSSEG-----VDR-----	490
SEQ ID NO: 85	SVVSQA-TVSLVFNAGAGEGFISVDGNEGDR-----	490
SEQ ID NO: 86	--PGPD-DVAIVFVNADSGENYIIVESNPGDRTV-----	469
SEQ ID NO: 87	GVPTNQTSIDEAVELAQSSDVVVVIGELPEAETPG---D	529
SEQ ID NO: 88	AVDSGKYDYAIVVVGEPPYAETFGDN-----	495
SEQ ID NO: 89	VIDDQ-AEIEKSVALAKEHDQVIICAGLNADWETEGADRA	575
SEQ ID NO: 90	PEVNAASDACLVFINAIAATEG-----WDR-----	466
SEQ ID NO: 91	KVDPQ-EMIAHAVKAAADADYTIICTGLSGEWESEGFDRP	446
SEQ ID NO: 92	ALVSQPDIAIAIVFANSDSGEGYINVDGNEGDR-----	490
SEQ ID NO: 93	GFAKAIAAAKKSDAIIYLGIDNTIEQEGADR-----	459
SEQ ID NO: 94	AIDDD-EEIRNAAELAAKHDKAVLIIGLNGEWETEGYDRE	590
SEQ ID NO: 95	PVANPASDACIDFINAAASEG-----YDR-----	453
SEQ ID NO: 96	FVKSNKFSYAIVVVGEVPYAEMFGDS-----	496
SEQ ID NO: 97	ALAKQA-THAIVFVNADSGEGYITVDGNEGDR-----	488
SEQ ID NO: 98	TLAKQA-SVAIVFVNADSGEGYITVDGNAGDR-----	487
SEQ ID NO: 99	TATGKD--VAMVFITADSGEGYLTVEGNAGDR-----	461
SEQ ID NO: 100	TATGKD--VAMVFITADSGEGYLTVEGNAGDR-----	461
SEQ ID NO: 101	DSALYA-DAAIEVANSVAGEEIGDVDGNYGDL-----	485
SEQ ID NO: 102	ANVAKDADVAVVFRATSGEYIIVVDNNKGDR-----	450
SEQ ID NO: 103	TVASDA-HMSIVVVSASVSGEGYLIIDGNRGDK-----	499
SEQ ID NO: 104	KVASDA-HLSIVVVSASVSGEGYITVDGNQGDR-----	497
SEQ ID NO: 105	VIDPE-TEIDYAVRVAKSVDCVILCVGLTAEWETEGEDRK	583
SEQ ID NO: 106	TSIYPTPEVCLVFLKTWAREG-----TDR-----	486
SEQ ID NO: 107	AAASQA-DTCLVFANADSGEGYITVDGNEGDR-----	486
SEQ ID NO: 108	--ANPE-DIAIVFINADSGENYITVEGNYGDRSA-----	476
SEQ ID NO: 109	QVASQA-SVSIVFVNADSGEGYINVDGNEGDR-----	490
SEQ ID NO: 110	QVASQA-SVSIVFVNADSGEGYINVDGNEGDR-----	490
SEQ ID NO: 111	--ANPE-DIAIVFINADSGENYITVEGNYGDRSA-----	480
SEQ ID NO: 112	AARGKD--VAIVFITADSGEGYITVEGNAGDR-----	453
SEQ ID NO: 113	YVKTNSEFDYAIVVVGEPPYAEMQGDS-----	496
SEQ ID NO: 114	KAASTA-DVSIVHVSSDSGEGYLTVEGNQGDR-----	487
SEQ ID NO: 115	VANAD-DLLSEAVKLASEADAVIAVVGLNADWETEGNDRR	603

Figure 11/6/13

16/20/13

SEQ ID NO: 1	-NNLDPWHNGNALVQAVAGANS-NVIVVVHSGVGAIIIEQI	491
CONSENSUS	-NNLTLWHNGDNLIKAVASNCN-NTVVVIHSGVGPVLVEEW	495
SEQ ID NO: 72	-NNLTLWKNGDNLIKAAANNCN-NTIVVIHSGVGPVLVDEW	528
SEQ ID NO: 73	-PGLADNSSDTLVKN-VARKCA-NTIVTIHNAGIRVVGW	490
SEQ ID NO: 74	-KNLTLWKNGEAVIDTVVSHCN-NTIVVIHSGVGPVLIDRW	529
SEQ ID NO: 75	-KNLTLWRNGDNVIKAAASNCN-NTIVIIHSGVGPVLVNEW	528
SEQ ID NO: 76	---LAFNATEKARQAKIYSSLP--TIVDIILDRPAVIPEV	515
SEQ ID NO: 77	SFHLPSNGSQDRLVAAVGAANP-NTVVVNCTGVAVAMP--	613
SEQ ID NO: 78	-RNLTLWRNGDNVIKAAASNCN-NTIVVIHSGVGPVLVNEW	528
SEQ ID NO: 79	-RNLTLWRNGDNVIKAAASNCN-NTIVIIHSGVGPVLVNEW	528
SEQ ID NO: 80	-NNITLWKNGDNVVKTAANNCN-NTVVIIHSGVGPVLIDEW	528
SEQ ID NO: 81	-NNITLWKNGDNVVKTAANNCN-NTVVIIHSGVGPVLIDEW	528
SEQ ID NO: 82	-AGLHAWHNGDNLVKAAAEKFS-NVVVVVHTVGPILMEEW	518
SEQ ID NO: 83	-PYLADPYSDTLVQN-VASQCS-NTMVVIHNAGVRLVDRW	530
SEQ ID NO: 84	-TSYPLDWNSTLVVNNVASFCQGNTIVITHSSGINTLP-F	528
SEQ ID NO: 85	-KNLTLWKNGDELIKTVASMEN-NTVVVMHTAGPVLVNKW	528
SEQ ID NO: 86	-AQMKLWHNGDELIESAAKKFSNVVVVVVHTVGPIMEKW	508
SEQ ID NO: 87	IYDLSMDPNEVLLLQQLVDTGK-PVVLILVEARPRILPP-	567
SEQ ID NO: 88	-LNLTIAPAGPSVIQNVCKSVR--CVVVLISGRPLVVEP-	531
SEQ ID NO: 89	SMKLP--GVLDQLIADVAAANP-NTVVVMQTGTPEEMP--	610
SEQ ID NO: 90	-DGLHDDFSDGLVLN-VAANCS-NTIVVVHAAGTRLVDQW	503
SEQ ID NO: 91	HMDLP--PGVDTMISQVLDAAP-NAVNVNQSGTPVTMS--	481
SEQ ID NO: 92	-KNLTLWKNGDDLKTVAAVNP-KTIVVIHSTGPPVILKDY	528
SEQ ID NO: 93	-TDIAWPGNQLDLIKQLSEVGK-PLVVLQMGGGQVDSSSL	497
SEQ ID NO: 94	NMDLP--KRTNELVRAVLKANP-NTVIVNQSGTPVEFP--	625
SEQ ID NO: 95	-PNLADKYSDKLVEA-VASQCS-NTIVVIHNAGIRLVDNW	490
SEQ ID NO: 96	-SNLTIAEPGPSTISNICGSVK--CVVVVSGRPVPLEP-	532
SEQ ID NO: 97	-NNLTLWQNGEELVRNVSGYCN-NTIVVIHSGVGPVLVDSF	526
SEQ ID NO: 98	-NNLTLWQDGDTLIKNVSSLCN-NTIVVIHSGVGPVLVNSF	525
SEQ ID NO: 99	-NDLQAWHGGDALVQQVASHNK-NTIVVINSVGPINMEAW	499
SEQ ID NO: 100	-NDLQAWHGGDALVQQVASHNK-NTIVVINSVGPINMEAW	499
SEQ ID NO: 101	-NNLTLWHNAVPLIKNISSINN-NTIVIVTSGQQIDLEPF	523
SEQ ID NO: 102	-NNLDLWHGGNDLVKSVAAVNK-NTVVVIHAPATVNLP-F	487
SEQ ID NO: 103	-NNVTLWHNSDNLIKAVAENCA-NTVVVITSTGQVDVESF	537
SEQ ID NO: 104	-KNLTLWNNGDKLIETVAENCA-NTVVVVTSTGQINFEGF	535
SEQ ID NO: 105	TMTLP--SLSDKLVYSILQSNP-NTVVVTQSGTPIEMP--	618
SEQ ID NO: 106	-LSYENDWNSTAVVNNVARRCP-NTIVVTHSGGINTMP-W	523
SEQ ID NO: 107	-KNLTLWQGADQVIHNVSANCN-NTVVVLHTVGPVLIDDW	524
SEQ ID NO: 108	-AGLYAWHNGDDLKAAAKFS-KVVVVVHTVGPILLENW	514
SEQ ID NO: 109	-KNLTLWKGGEEVIKTVAANCN-NTIVVMHTVGPVLIDEW	528
SEQ ID NO: 110	-KNLTLWKGGEEVIKTVAANCN-NTIVVMHTVGPVLIDEW	528
SEQ ID NO: 111	-AGLYAWHNGDDLKAAAKFS-KVVVVVHTVGPILLENW	518
SEQ ID NO: 112	-NNLDPWHNGNALVQAVAGANS-NVIVVVHSGVGAIIIEQI	491
SEQ ID NO: 113	-FNLTIPEPGPTTISVCGAVK--CVVVVISGRPVVLQP-	532
SEQ ID NO: 114	-SNTSLWNKGDELILKAAEACN-NVVVVVIHSGVPVMEAW	525
SEQ ID NO: 115	TLALP--GRTDELVEKVAKVNS-KTVVVTQAGSAITLP--	638

Figure 11/6/14

16/20/14

SEQ ID NO: 1	LALPQVKAVVWAGLPSQESGNALVDVLWGDVSPSGKLVYT	531
CONSENSUS	YDHPNVTAIVWAGLPGQESGNALADVLYGRVNPSGKLPFT	535
SEQ ID NO: 72	YDHPNVTAILWAGLPGQESGNSLADVLYGRVNPGAKSPFT	568
SEQ ID NO: 73	IDHENVTAIVIFAHLPQGDSGRALVELLYGRANPSGKLPYT	530
SEQ ID NO: 74	YDNPNTAIIWAGLPGQESGNSLVDVLYGRVNPSAKTPFT	569
SEQ ID NO: 75	YDNPNTAILWAGLPGQESGNSLADVLYGRVNPGAKSPFT	568
SEQ ID NO: 76	VEQAQAVLASYGSD-----SEAFLDVVFGVSKPEGKLPFD	550
SEQ ID NO: 77	-WLDKVKAVVQAWFPGQEAGNAIADVLTGAVNPSGRLPVS	652
SEQ ID NO: 78	YDNPNTAILWAGLPGQESGNSLADVLYGRVNPGAKSPFT	568
SEQ ID NO: 79	YDNPNTAILWAGLPGQESGNSLADVLYGRVNPGAKSPFT	568
SEQ ID NO: 80	YDHPNVTGILWAGLPGQESGNSIADVLYGRVNPGAKSPFT	568
SEQ ID NO: 81	YDHPNVTGILWAGLPGQESGNSIADVLYGRVNPGAKSPFT	568
SEQ ID NO: 82	IDLDSVKAVLVAHLPGQEAGWSLTDILFGDYSPSGHLPYT	558
SEQ ID NO: 83	IENDNITAVIYAHLPQGDSGRALVEVMYKQSPSGRLPYT	570
SEQ ID NO: 84	AQNPVNSAILLAHYPGQESGNSIVDVLFVGVNPSGKLPYT	568
SEQ ID NO: 85	YDHPNVTAILWAGLPGQESGNALGDVIYGRVNPGAKSPFT	568
SEQ ID NO: 86	IDLLRSR---VSCLPDFQDKKLEILLISCSETSVRVAAS	545
SEQ ID NO: 87	DLVYSCAAVLMAYLPGSEGKPIANILMGNVNPSGRLPLT	607
SEQ ID NO: 88	-YISAMDAFVAAWLPGSE-GQGVADVLFGDYGFSGKLART	569
SEQ ID NO: 89	-WLDATPAVIQAWYGGNETGNSIADVVFVDYNPSGKLSLS	649
SEQ ID NO: 90	IEHPNVTAAVIAHLPGQDSGRALVKLLYGEANFSGKLPYT	543
SEQ ID NO: 91	-WAHKAKAIVQAWYGGNETGHGISDVLFGNVNPSGKLSLS	520
SEQ ID NO: 92	ANHPNISAILWAGAPGQESGNSLVDILYKQSPG-RTPFT	567
SEQ ID NO: 93	KSNKKVNSLVWGGYPGQSGGVALFDILSGKRAPAGRLVTT	537
SEQ ID NO: 94	-WLEEANALVQAWYGGNELGNALADVLYGDVVPNGKLSLS	664
SEQ ID NO: 95	IEHENVTGIVLAHLPGQDTGTSLIEVLYGNQSPSGRLPYT	530
SEQ ID NO: 96	-YVSKMDALVAAWLPGTE-GQGVADALFGDYGFTGKLART	570
SEQ ID NO: 97	NNSPNVSAIILWAGLPGQESGNALITDVLYGRVNPGGKLPFT	566
SEQ ID NO: 98	YDSENVTAILWAGLPGQESGNALIDILYGRHNPGGKLPFT	565
SEQ ID NO: 99	VNHPNVTAIVWSGLPGQEAGNAVTDVLFVAVNPGGKLPFT	539
SEQ ID NO: 100	VNHPNVTAIVWSGLPGQEAGNAVTDVLFVAVNPGGKLPFT	539
SEQ ID NO: 101	IDNENVTAIVYSSYLGQDFGTVLAKVLFVDENPSGKLPFT	563
SEQ ID NO: 102	LNN--VKAIHAGMPGAESGNALIASILFGDSNPSGHLPT	525
SEQ ID NO: 103	ADHPNVTAIVWAGPLGDRSGTAIANILFGNANPSGHLPT	577
SEQ ID NO: 104	ADHPNVTAIVWAGPLGDRSGTAIANILFGKANPSGHLPT	575
SEQ ID NO: 105	-WISEAHTLLHIWYNGNELGNALANIIIFGEQNPCGKLPT	657
SEQ ID NO: 106	ADNANVTAILAAHYPGQENGNSIMDILYGDVNPSGRLPYT	563
SEQ ID NO: 107	YDHPNVTAILWAGLPGQESGNSLVDVLYGRVNPG-KTPFT	563
SEQ ID NO: 108	IDLPSVKSVVFAHLPGQEAGDSLVDVLFVDYSPSGHLPYT	554
SEQ ID NO: 109	YDNPNTAIIVWAGLPGQESGNSLVDVLYGRVSPGGKTPFT	568
SEQ ID NO: 110	YDNPNTAIIVWAGLPGQESGNSLVDVLYGRVSPGGKTPFT	568
SEQ ID NO: 111	IDLPSVKSVVFAHLPGQEAGDSLVDVLFVDYSPSGHLPYT	558
SEQ ID NO: 112	LALPQVKAVVWTGLPSQESGNALVDVLWGDVSPSGKLVYT	531
SEQ ID NO: 113	-YVSMDALVAAWLPGTE-GQGVTDVLFVDYGFTGKLART	570
SEQ ID NO: 114	INHPNVTAIVLLAGLPGQEAGSAEVDVLWGSTNPSGRLPYT	565
SEQ ID NO: 115	-WLDSVAAVVHAWYLGATGDAIADVLFVGKQNPSTGKLSLT	677

Figure 11/6/15

16/20/15

SEQ ID NO: 1	IAKSPND-YNTRIVS-----GGG-----DSFSEGLFI	557
CONSENSUS	WGKTRED-YGPSLVT-P-----NG-G-PQDDFTEGVFI	564
SEQ ID NO: 72	WGKTREA-YGDYLVRELN-----NGNGAPQDDFSEGVFI	601
SEQ ID NO: 73	VAKKVEDYGSLLHPSLPE-----TPYGL-FPQSDFDSEGVFI	565
SEQ ID NO: 74	WGKTRES-YGAPLLTEPN-----NGNGAPQDDFNEGVFI	602
SEQ ID NO: 75	WGKTREA-YQDYLVTEPN-----NGNGAPQEDFVEGVFI	601
SEQ ID NO: 76	LPRSMDA-----VEAQA	562
SEQ ID NO: 77	FPRAIED--APAHGNFPGDYTDGKDNRRHLEVITYKEGVFV	690
SEQ ID NO: 78	WGKTREA-YQDYLVTEPN-----NGNGAPQEDFVEGVFI	601
SEQ ID NO: 79	WGKTREA-YQDYLYTEPN-----NGNGAPQEDFVEGVFI	601
SEQ ID NO: 80	WGKTRES-YGSPLVKDAN-----NGNGAPQSDFTQGVFI	601
SEQ ID NO: 81	WGKTRES-YGSPLVKDAN-----NGNGAPQSDFTQGVFI	601
SEQ ID NO: 82	IPHSES--DYPESVGLIA-----QPFQIQDDYTEGLYI	590
SEQ ID NO: 83	VAKNESDYGSLLNPVIQS-----GTDDIYYPQDNFTEGVFI	606
SEQ ID NO: 84	IPVHESDMHIPVVNLSTSEVLAAHQSNQNAWQSNFTEKLAI	608
SEQ ID NO: 85	WAATSED-YGVVSILKEPN-----AATKAPQIDFEEGIFI	601
SEQ ID NO: 86	IYDTESTRIGLSDSVSLIN-----QRFQIQDFTFTEGLFI	579
SEQ ID NO: 87	YPGTTGD-----IGV	617
SEQ ID NO: 88	WFKSADQ-----LPMNV	581
SEQ ID NO: 89	FPKRLQD--NPAFLNFR--TEAGR-----TLYGEDVYV	678
SEQ ID NO: 90	IAKNESDY-SVYTPCQRR----SPEDT-DPQCDFTEGVYL	577
SEQ ID NO: 91	WPVDVKH--NPAYLNYA--SVGGR-----VLYGEDVYV	549
SEQ ID NO: 92	WGPSLES-YGVSVMTTPN-----NGNGAPQDNFNEGAFI	600
SEQ ID NO: 93	QYPAEYVHQFPQNDMNLR-----PDGKSN	561
SEQ ID NO: 94	WPFKLQD--NPAFLNFK--TEFGR-----VVYGEDIFV	693
SEQ ID NO: 95	VAKKASDYGGLLWPTEPE----GDLDLYFPQSNFTEGVFI	566
SEQ ID NO: 96	WFKRVDQ-----LPMNF	582
SEQ ID NO: 97	IGKSAEE-YGPDIIYEPT-----AGHGSPQANFEEGVFI	599
SEQ ID NO: 98	IGSDAAE-YGPDIIYEPT-----NNSSSPQDNFEEGVFI	598
SEQ ID NO: 99	IGKSISD-YSAQIIT-----TGSGIVPIPYNEGLFI	569
SEQ ID NO: 100	IGKSISD-YSAQIIT-----TGSGIVPIPYNEGLFI	569
SEQ ID NO: 101	IAKDVND-YIPVIEK-----VDVPDPVDKFTEESIYV	593
SEQ ID NO: 102	WAAREDYCCDVSYPALP-----HGGNSKTAYDYKEGLFV	560
SEQ ID NO: 103	VAKSNDD-YIPIVTYNPP-----NGEPEDNTLAEHDLV	610
SEQ ID NO: 104	IAKTDDD-YIPIETYSPP-----SGEPEDNHLVENDLLV	608
SEQ ID NO: 105	FPKKLKD--NPAYLSFR--SSRGH-----CVYGEDVYV	686
SEQ ID NO: 106	IPKLATDYDFPVVNITN-----EAQDPYVWQADFTEGLLI	598
SEQ ID NO: 107	WGRARDD-YGAPLIVKPN-----NGKGAPQQDFTEGIFI	596
SEQ ID NO: 108	IPRSED--QYPSSVSLIN-----QPFQIQDFTFTEGLYI	586
SEQ ID NO: 109	WGKTRES-YGAPLLTKPN-----NGKGAPQDDFTEGVFI	601
SEQ ID NO: 110	WGKTRES-YGAPLLTKPN-----NGKGAPQDDFTEGVFI	601
SEQ ID NO: 111	IPRSED--QYPSSVSLIN-----QPFQIQDFTFIEGLYI	590
SEQ ID NO: 112	IAKSPND-YNTRIVS-----GGG-----DSFSEGLFI	557
SEQ ID NO: 113	WFKTVQD-----LPMNV	582
SEQ ID NO: 114	IAKKPS--DYPAL-LYE-----SNMTVPQINYSERLNI	596
SEQ ID NO: 115	FPKRLED--VPSHGFG--SENGK-----VRYAEDLFV	706

Figure 11/6/16

16/20/16

SEQ ID NO:			% Ident. with TrCel3A
1	DYKHFDNAN-----ITPRYEFYGLSYTKF---	582	
CONSENSUS	DYRHFDKAN-----ITPRYEFYGLSYTTF---	589	52.1
72	DYRGFDKRN-----ETPIYEFYGHGLSYTTF---	626	44.1
73	DYRAFDNAN-----ITAQFEFGFGLSYTSF---	590	35.2
74	DYRHFDKRN-----ETPIYEFYGHGLSYTTF---	627	43.1
75	DYRGFDKRN-----ETPIYEFYGLSYTTF---	626	42.6
76	EDLPFDNEN-----PVPFYGHGLEYNEN---	585	13.2
77	GYRHYDLSEANR-----AKVLFPPFYGLSYTTF---	718	19.4
78	DYRGFDKRN-----ETPIYEFYGLSYTTF---	626	42.8
79	DYRGFDKRN-----ETPIYEFYGLSYTTF---	626	42.6
80	DYRHFDKEN-----ETPIYEFYGLSYTTF---	626	42.2
81	DYRHFDKEN-----ETPIYEFYGLSYTTF---	626	42.3
82	DYRHFLKAN-----ITPRYPFGHGLSYTTF---	615	43.4
83	DYKAFVAN-----ITPRYEFYGLTYSTF---	631	31.6
84	DYRGFEMKN-----VTPLYEFYGHGLSYTTF---	633	35.9
85	DYRAFDKSN-----TKPIYEFYGLSYTTF---	626	41.4
86	DYRHFDKEN-----ITPRYHFGYGLSYTTF---	604	40.4
87	PYYHKYSENG-----VTTPLFQFGDGLSYTTF---	644	16.6
88	GDKHY-----DPLFPFGFGLTTEA-----	600	19.5
89	GYRYEFADK-----DVNFPPFGHGLSYTTF---	703	19.4
90	DYRAFDANN-----MTPRFEFGYGLSYTSF---	602	36.5
91	GYKFYDKTER-----EVLFPPFGHGLSYATF---	574	12.8
92	DYRYFDKVAPGKPRSSDKAPTYEFGFGLSWSTF---	633	41.5
93	PGQTYIWYTG-----KPVYEFYSGFLFYTTF---	586	19.2
94	GYRYEKLQR-----KVAFPFYGLSYTTF---	718	18.0
95	DYKYFIQKN-----ITPRYEFYGLTYTTF---	591	33.0
96	DDAHV-----DPLFPFGFGITTKPVKGY---	605	20.0
97	DY-----	601	41.9
98	DYRAFDKQN-----VTPIYEFYGLSYTKF---	623	42.7
99	DYRHFDQAG-----IAPRFEFGFGLSYTTF---	594	52.6
100	DYRHFDQAG-----IAPRFEFGFGLSYTTF---	594	52.6
101	DYRYFDKYN-----KPVRYEFYGLSYSNF---	618	36.4
102	GYRWFDKKN-----KTPIFPFGHGLSYTTF---	585	38.3
103	DYRYFEEKN-----IEPRYAFYGLSYNEY---	635	38.2
104	DYRYFEEKN-----IEPRYAFYGLSYNEY---	633	38.3
105	GYKYEAEVER-----EVLFPPFYGLSYTTF---	711	17.4
106	DYRHFDARN-----ITPLYEFYGLSYTTF---	623	37.2
107	DYRRFDKYN-----ITPIYEFYGLSYTTF---	621	43.2
108	DYRHFLHAN-----LTPRYPFYGHGLSYTTF---	611	45.0
109	DYRRFDKYN-----ETPIYEFYGLSYTTF---	626	44.3
110	DYRRFDKYN-----ETPIYEFYGLSYTTF---	626	44.3
111	DYRHFLHAN-----LTPRYPFYGHGLSYTTF---	615	45.0
112	DYKHFDNAN-----ITPRYEFYGL-----	577	98.3
113	GDKHY-----DPLFPFGFGLTTKPS---	602	20.3
114	DYRHFDNEN-----IEPRFEFGFGLSYTTF---	621	39.8
115	GYKHYHHRNI-----EPLFPFGFGLSYTTF---	731	18.8

Figure 11/6/17

16/20/17

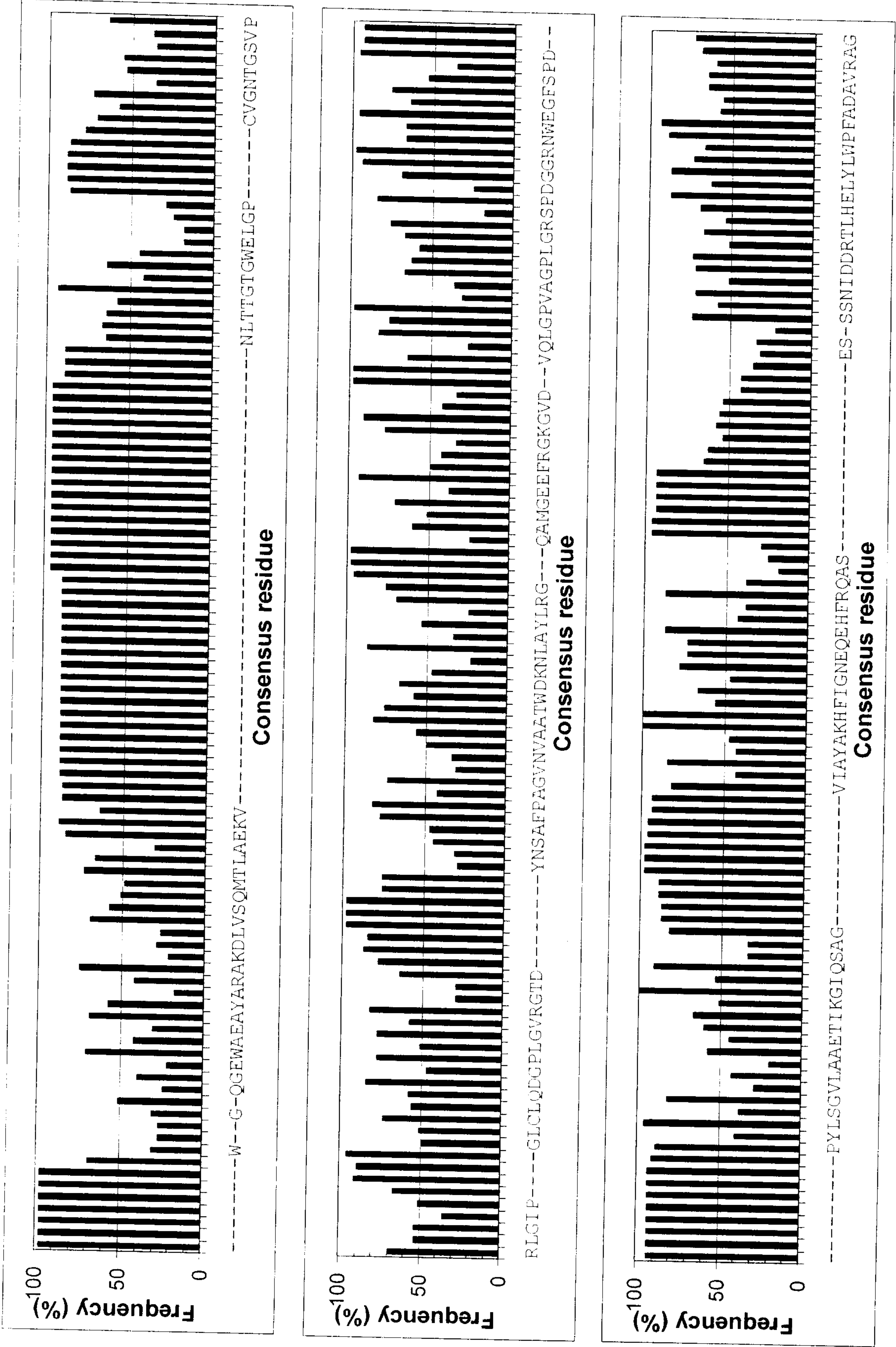


Figure 11/6/18

16/20/18

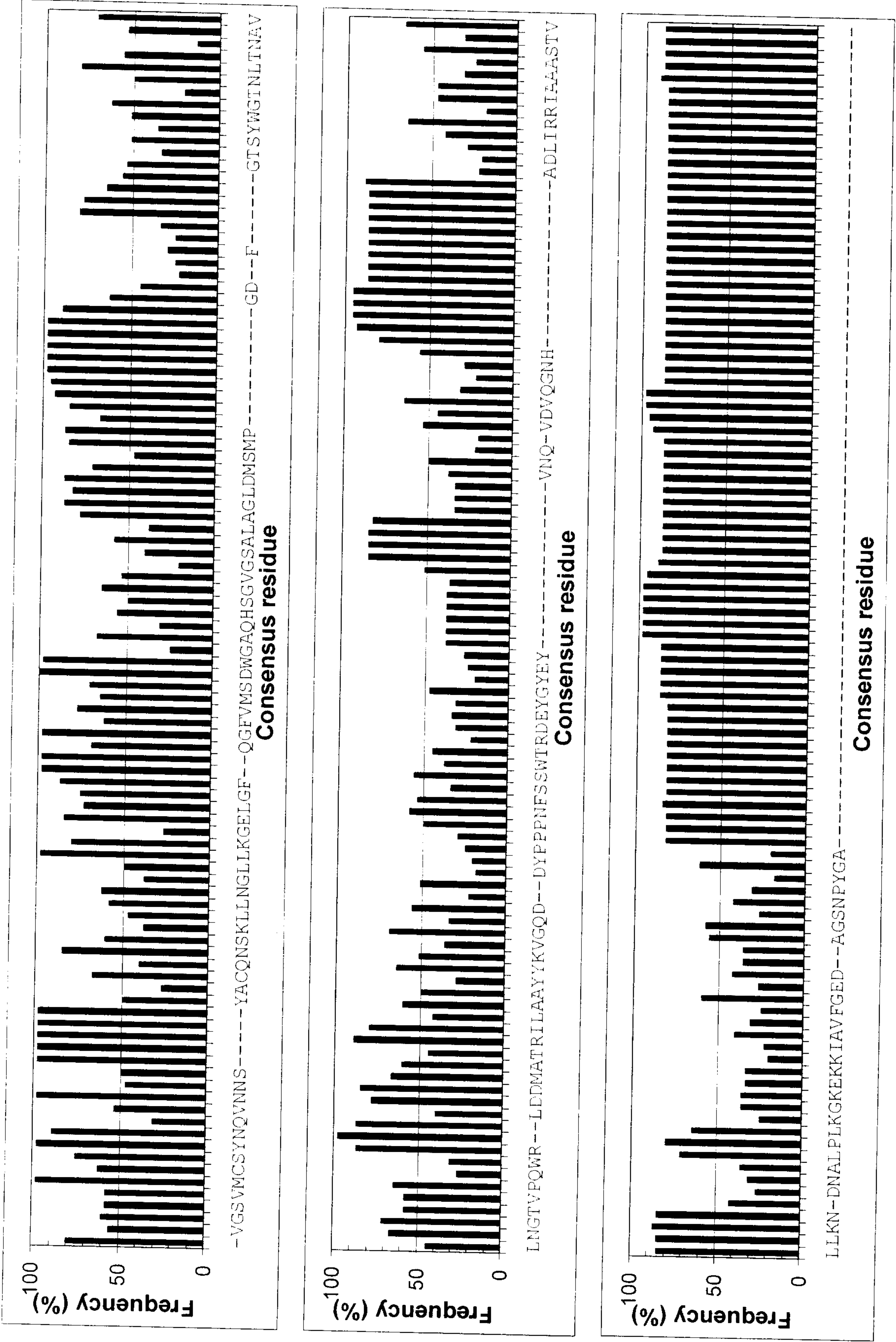


Figure 11/6/19

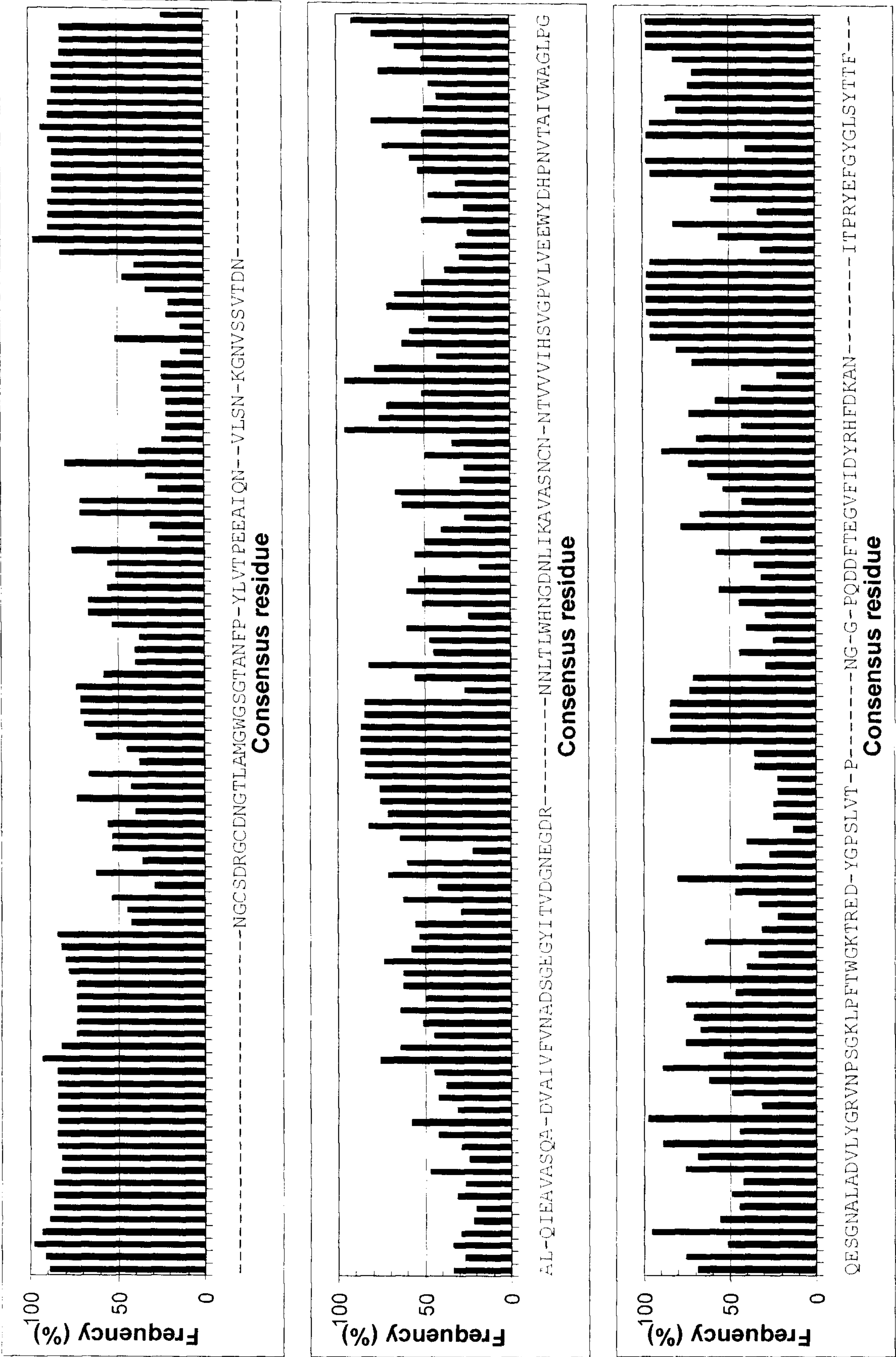
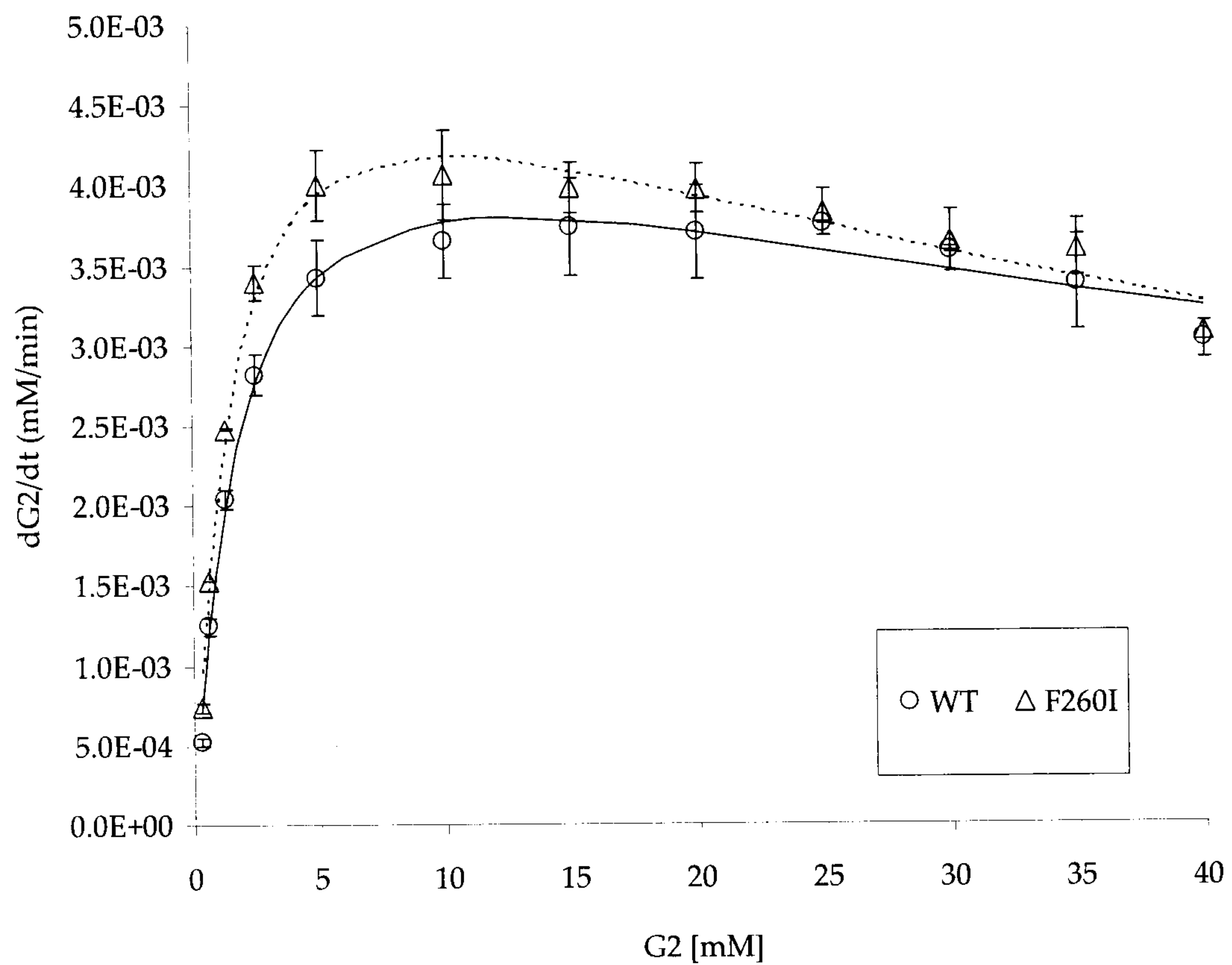


Figure 11/6/20

16/20/20

**Figure 12**

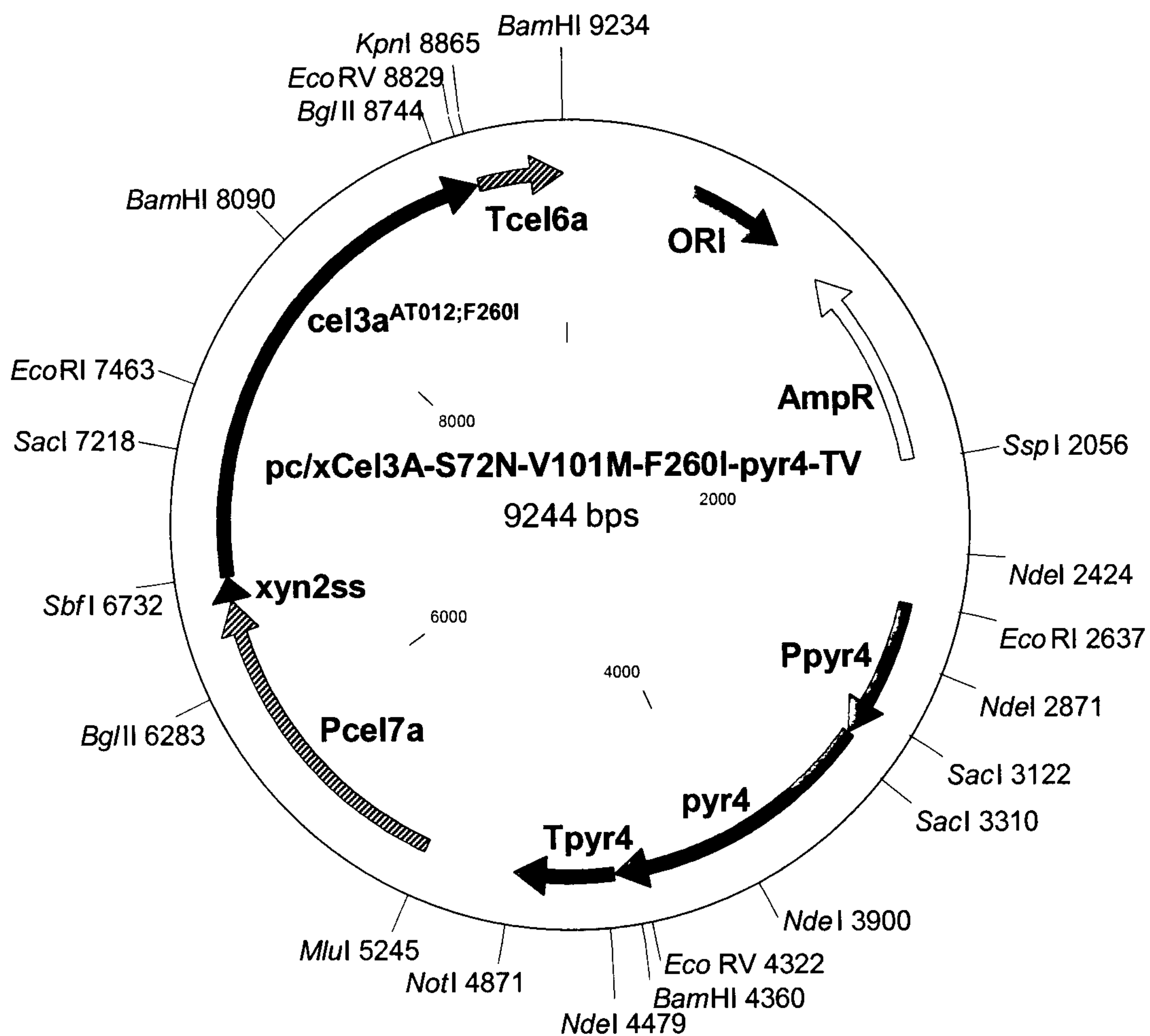
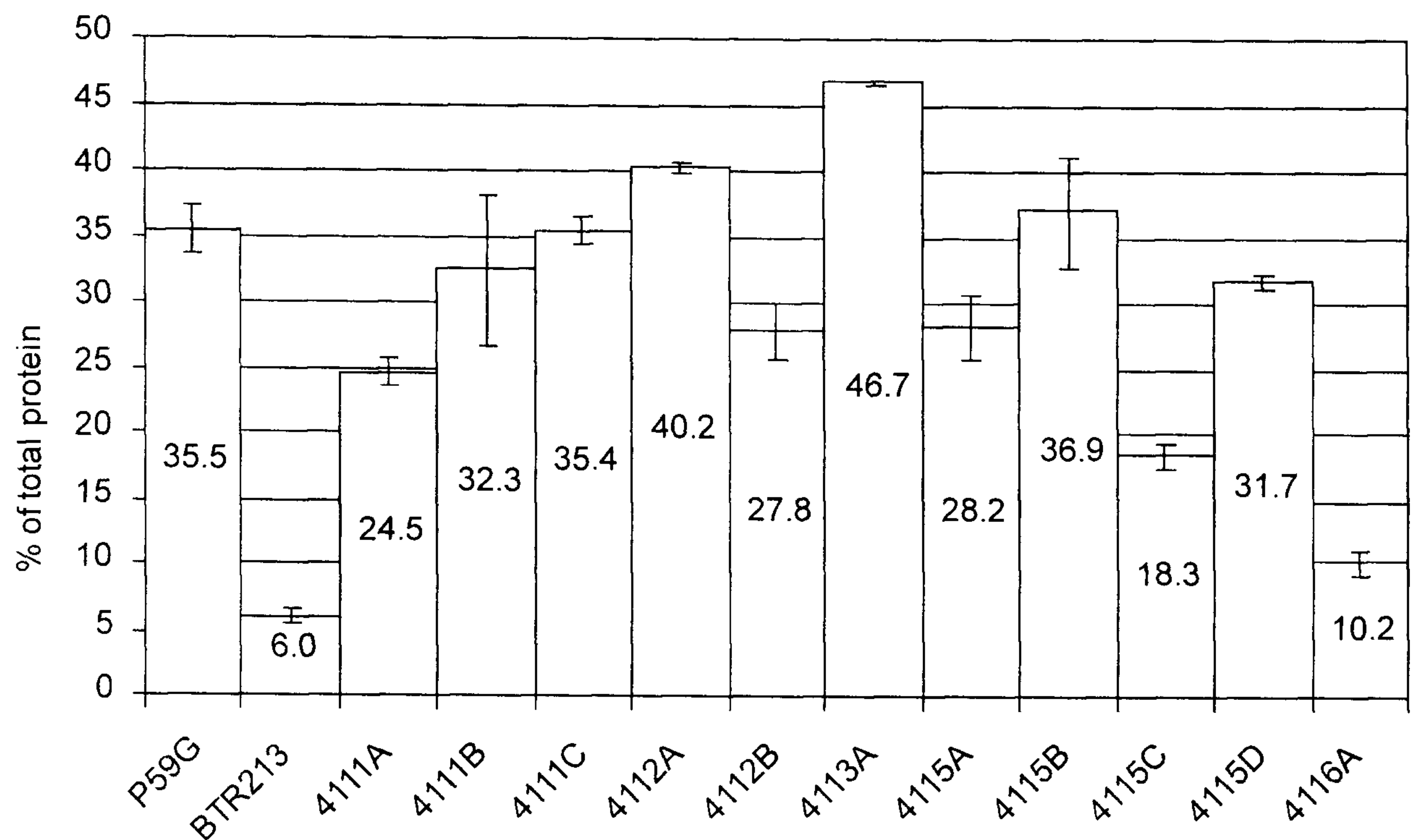
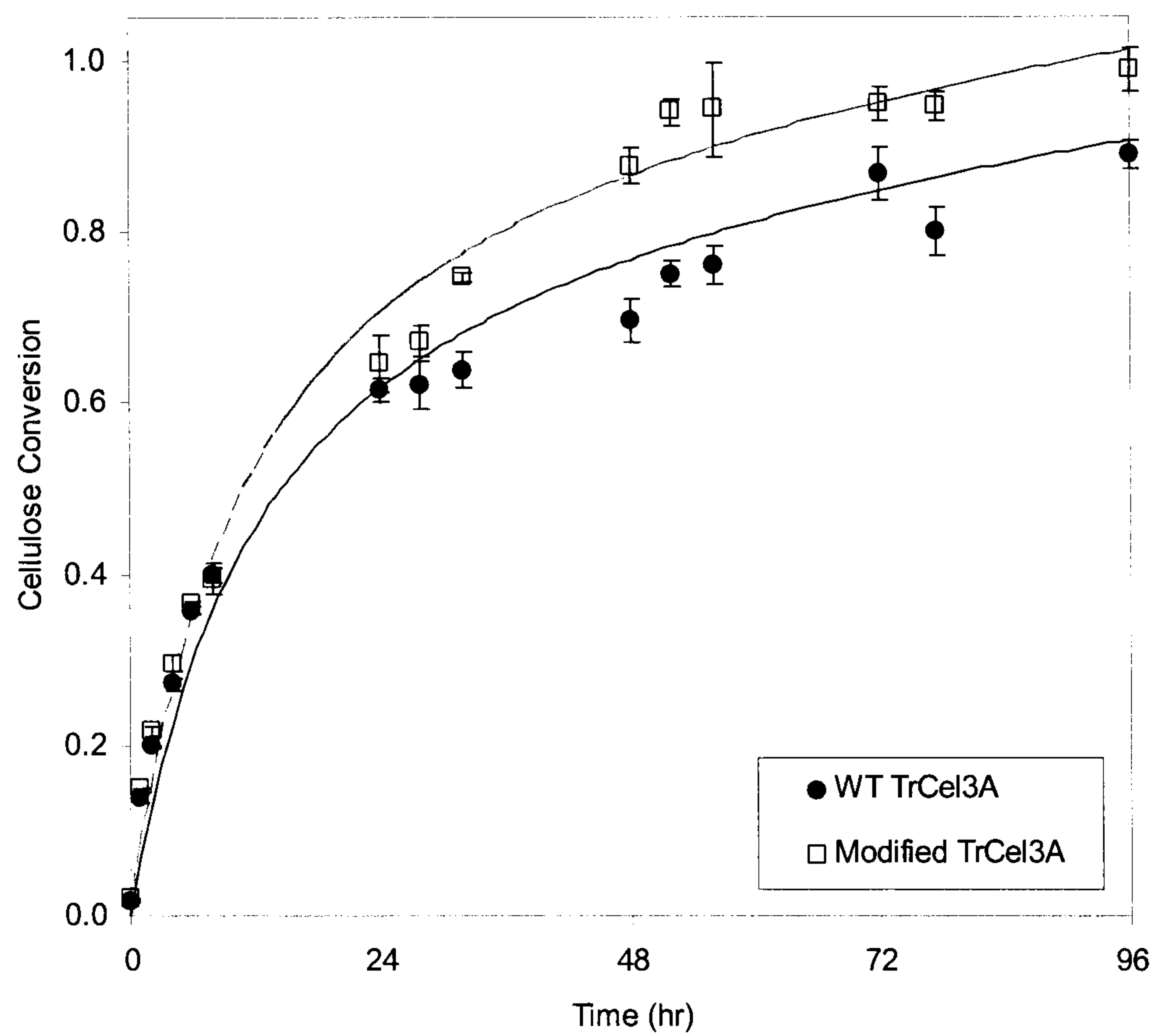


Figure 13

**Figure 14**

**Figure 15**

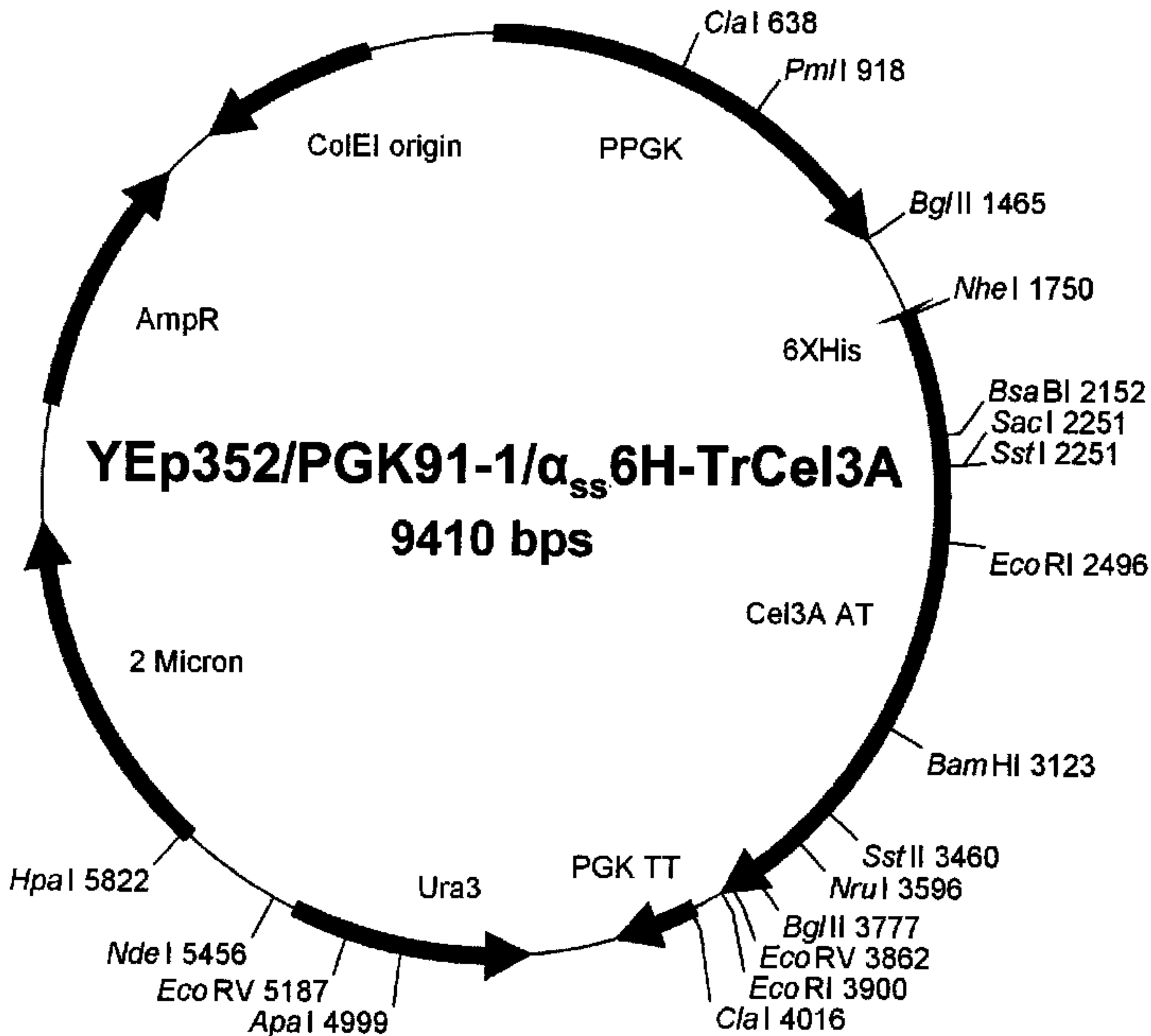


Figure 1