



(86) Date de dépôt PCT/PCT Filing Date: 2011/10/06
(87) Date publication PCT/PCT Publication Date: 2012/04/12
(85) Entrée phase nationale/National Entry: 2013/03/19
(86) N° demande PCT/PCT Application No.: US 2011/055181
(87) N° publication PCT/PCT Publication No.: 2012/048171
(30) Priorité/Priority: 2010/10/06 (US61/390,392)

(51) Cl.Int./Int.Cl. *C07K 14/435* (2006.01),
C07K 14/47 (2006.01), *C12N 9/00* (2006.01)
(71) Demandeur/Applicant:
BP CORPORATION NORTH AMERICA INC., US
(72) Inventeurs/Inventors:
STEGE, JUSTIN T., US;
VARVAK, ALEXANDER, IL;
POLAND, JOHN, US;
LYON, CHRIS S., US;
HEALEY, SHAUN, US;
LUGINBUHL, PETER, US
(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : POLYPEPTIDES VARIABLES DE LA CBH I DONT L'INHIBITION DES PRODUITS EST REDUITE
(54) Title: VARIANT CBH I POLYPEPTIDES WITH REDUCED PRODUCT INHIBITION

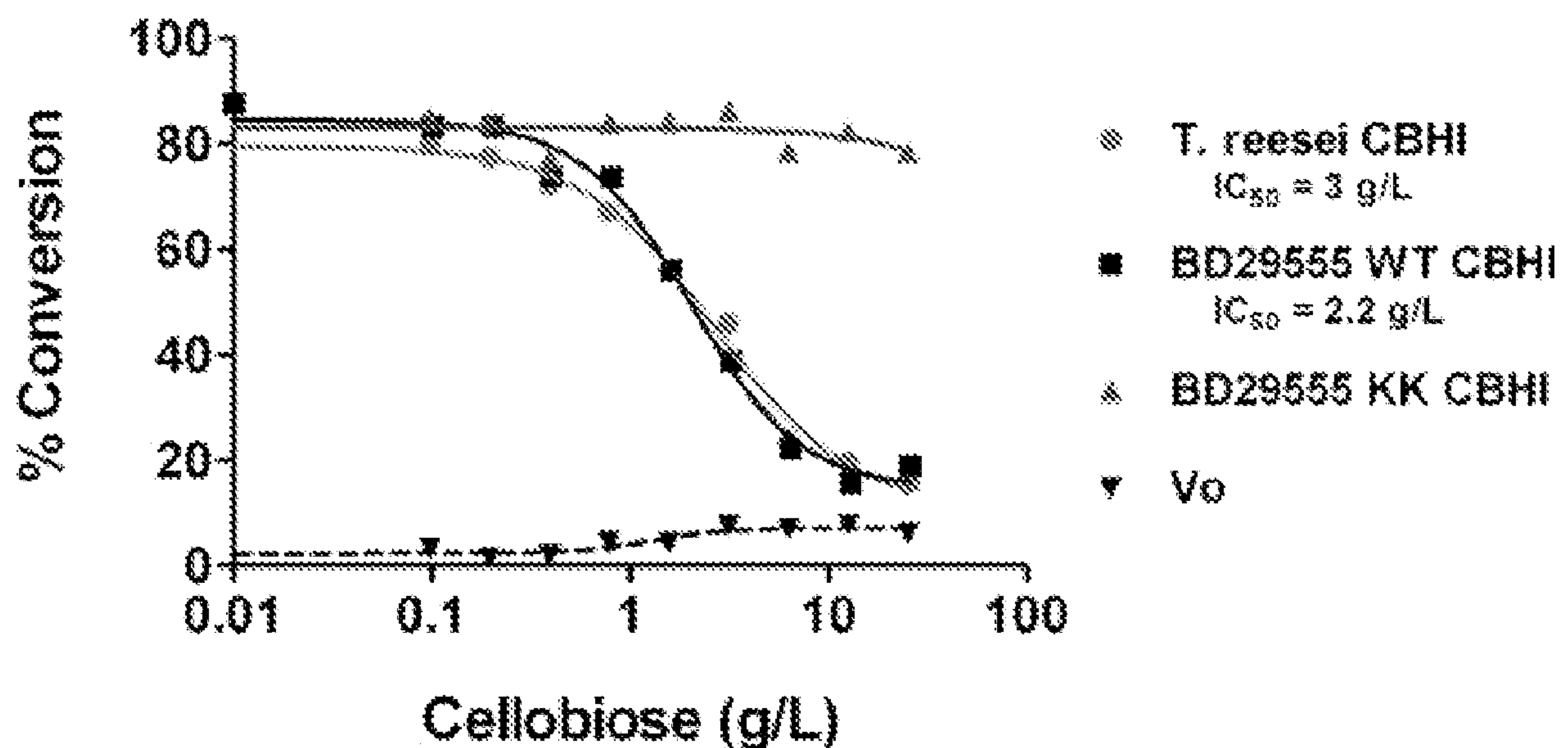


FIGURE 3

(57) Abrégé/Abstract:

The present disclosure relates to variant CBH I polypeptides that have reduced product inhibition, and compositions, e.g., cellulase compositions, comprising variant CBH I polypeptides. The variant CBH I polypeptides and related compositions can be used in variety of agricultural and industrial applications. The present disclosure further relates to nucleic acids encoding variant CBH I polypeptides and host cells that recombinantly express the variant CBH I polypeptides.

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

CORRECTED VERSION

(19) World Intellectual Property
Organization
International Bureau



WIPO | PCT



(10) International Publication Number
WO 2012/048171 A9

(43) International Publication Date
12 April 2012 (12.04.2012)

(51) International Patent Classification:

C12N 9/42 (2006.01) C12N 9/00 (2006.01)
C07K 14/47 (2006.01)

(21) International Application Number:

PCT/US2011/055181

(22) International Filing Date:

6 October 2011 (06.10.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/390,392 6 October 2010 (06.10.2010) US

(71) Applicant (for all designated States except US): **BP CORPORATION NORTH AMERICA INC.** [US/US]; 501 Westlake Park Boulevard, Houston, Texas 77079 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **STEGE, Justin T.** [US/US]; 11355 Denning Place, San Diego, California 92126 (US). **VARVAK, Alexander** [US/IL]; Harav Kook 47/28, 47/28 Netanya (IL). **POLAND, John** [US/US]; Unit E, 2929 Cowley Way, San Diego, California 92117 (US). **LYON, Chris S.** [US/US]; 2808 Copley Avenue, San Diego, California 92116 (US). **HEALEY, Shaun** [US/US]; #7a, 3380 Harding Street, Carlsbad, California 92008 (US). **LUGINBUHL, Peter** [CH/US]; 12802 Calle de la Siena, San Diego, California 92130 (US).

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

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

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

19 July 2012

[Continued on next page]

(54) Title: VARIANT CBH I POLYPEPTIDES WITH REDUCED PRODUCT INHIBITION

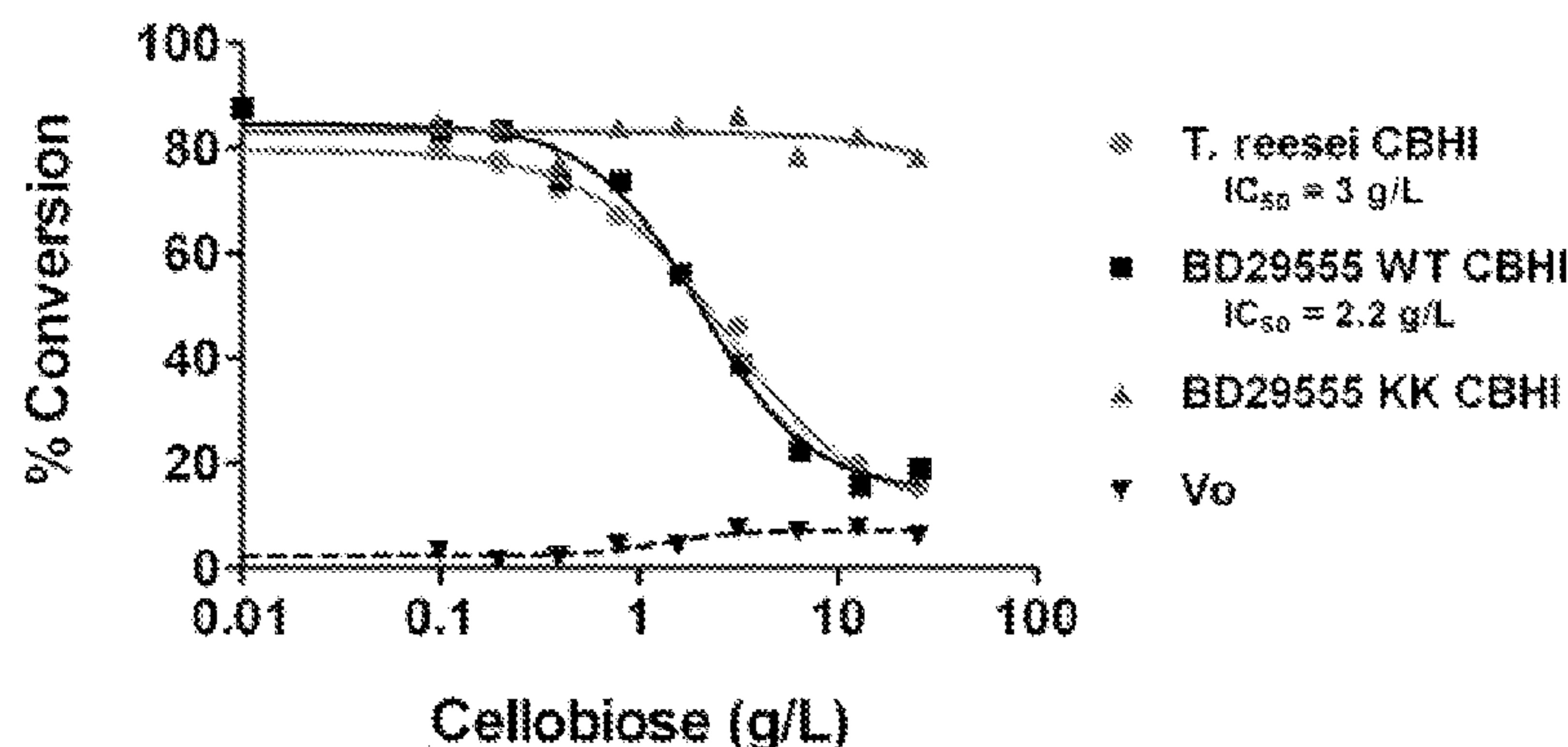


FIGURE 3

(57) Abstract: The present disclosure relates to variant CBH I polypeptides that have reduced product inhibition, and compositions, e.g., cellulase compositions, comprising variant CBH I polypeptides. The variant CBH I polypeptides and related compositions can be used in variety of agricultural and industrial applications. The present disclosure further relates to nucleic acids encoding variant CBH I polypeptides and host cells that recombinantly express the variant CBH I polypeptides.

WO 2012/048171 A9

(48) Date of publication of this corrected version:

30 August 2012

(15) Information about Correction:

see Notice of 30 August 2012

VARIANT CBH I POLYPEPTIDES WITH REDUCED PRODUCT INHIBITION

1. CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application no. 61/390,392, filed October 6, 2010, the contents of which are incorporated herein in their entireties by reference thereto.

2. BACKGROUND

[0002] Cellulose is an unbranched polymer of glucose linked by $\beta(1\rightarrow4)$ -glycosidic bonds. Cellulose chains can interact with each other via hydrogen bonding to form a crystalline solid of high mechanical strength and chemical stability. The cellulose chains are depolymerized into glucose and short oligosaccharides before organisms, such as the fermenting microbes used in ethanol production, can use them as metabolic fuel. Cellulase enzymes catalyze the hydrolysis of the cellulose (hydrolysis of β -1,4-D-glucan linkages) in the biomass into products such as glucose, cellobiose, and other cellooligosaccharides. Cellulase is a generic term denoting a multienzyme mixture comprising exo-acting cellobiohydrolases (CBHs), endoglucanases (EGs) and β -glucosidases (BGs) that can be produced by a number of plants and microorganisms. Enzymes in the cellulase of *Trichoderma reesei* include CBH I (more generally, Cel7A), CBH2 (Cel6A), EG1 (Cel7B), EG2 (Cel5), EG3 (Cel12), EG4 (Cel61A), EG5 (Cel45A), EG6 (Cel74A), Cip1, Cip2, β -glucosidases (including, *e.g.*, Cel3A), acetyl xylan esterase, β -mannanase, and swollenin.

[0003] Cellulase enzymes work synergistically to hydrolyze cellulose to glucose. CBH I and CBH II act on opposing ends of cellulose chains (Barr *et al.*, 1996, Biochemistry 35:586-92), while the endoglucanases act at internal locations in the cellulose. The primary product of these enzymes is cellobiose, which is further hydrolyzed to glucose by one or more β -glucosidases.

[0004] The cellobiohydrolases are subject to inhibition by their direct product, cellobiose, which results in a slowing down of saccharification reactions as product accumulates. There is a need for new and improved cellobiohydrolases with improved productivity that maintain

of cellulose into fermentable sugars and for related fields of cellulosic material processing such as pulp and paper, textiles and animal feeds.

3. SUMMARY

[0005] The present disclosure relates to variant CBH I polypeptides. Most naturally occurring CBH I polypeptides have arginines at positions corresponding to R268 and R411 of *T. reesei* CBH I (SEQ ID NO:2). The variant CBH I polypeptides of the present disclosure include a substitution at either or both positions resulting in a reduction or decrease in product (*e.g.*, cellobiose) inhibition. Such variants are sometimes referred to herein as “product tolerant.”

[0006] The variant CBH I polypeptides of the disclosure minimally contain at least a CBH I catalytic domain, comprising (a) a substitution at the amino acid position corresponding to R268 of *T. reesei* CBH I (“R268 substitution”); (b) a substitution at the amino acid position corresponding to R411 of *T. reesei* CBH I (“R411 substitution”); or (c) both an R268 substitution and an R411 substitution. The amino acid positions of exemplary CBH I polypeptides into which R268 and/or R411 substitutions can be introduced are shown in Table 1, and the amino acid positions corresponding to R268 and/or R411 in these exemplary CBH I polypeptides are shown in Table 2.

[0007] R268 and/or R411 substituents can include lysines and/or alanines. Accordingly, the present disclosure provides a variant CBH I polypeptide comprising a CBH I catalytic domain with one of the following amino acid substitutions or pairs of R268 and/or R411 substitutions: (a) R268K and R411K; (b) R268K and R411A; (c) R268A and R411K; (d) R268A and R411A; (e) R268A; (f) R268K; (g) R411A; and (h) R411K. In some embodiments, however, the amino acid sequence of the variant CBH I polypeptide does not comprise or consist of SEQ ID NO:299, SEQ ID NO:300, SEQ ID NO:301, or SEQ ID NO:302.

[0008] The variant CBH I polypeptides of the disclosure typically include a CD comprising an amino acid sequence having at least 50% sequence identity to a CD of a reference CBH I exemplified in Table 1. The CD portions of the CBH I polypeptides exemplified in Table 1 are delineated in Table 3. The variant CBH I polypeptides can have a cellulose binding

exemplified in Table 1. The CD portions of the CBH I polypeptides exemplified in Table 1 are delineated in Table 3. The variant CBH I polypeptides can have a cellulose binding domain (“CBD”) sequence in addition to the catalytic domain (“CD”) sequence. The CBD can be N- or C-terminal to the CD, and the CBD and CD are optionally connected via a linker sequence.

[0009] The variant CBH I polypeptides can be mature polypeptides or they may further comprise a signal sequence.

[0010] Additional embodiments of the variant CBH I polypeptides are provided in Section 5.1.

[0011] The variant CBH I polypeptides of the disclosure typically exhibit reduced product inhibition by cellobiose. In certain embodiments, the IC_{50} of cellobiose towards a variant CBH I polypeptide of the disclosure is at least 1.2-fold, at least 1.5-fold, or at least 2-fold the IC_{50} of cellobiose towards a reference CBH I lacking the R268 substitution and/or R411 substitution present in the variant. Additional embodiments of the product inhibition characteristics of the variant CBH I polypeptides are provided in Section 5.1.

[0012] The variant CBH I polypeptides of the disclosure typically retain some cellobiohydrolase activity. In certain embodiments, a variant CBH I polypeptide retains at least 50% the CBH I activity of a reference CBH I lacking the R268 substitution and/or R411 substitution present in the variant. Additional embodiments of cellobiohydrolase activity of the variant CBH I polypeptides are provided in Section 5.1.

[0013] The present disclosure further provides compositions (including cellulase compositions, *e.g.*, whole cellulase compositions, and fermentation broths) comprising variant CBH I polypeptides. Additional embodiments of compositions comprising variant CBH I polypeptides are provided in Section 5.3. The variant CBH I polypeptides and compositions comprising them can be used, *inter alia*, in processes for saccharifying biomass. Additional details of saccharification reactions, and additional applications of the variant CBH I polypeptides, are provided in Section 5.4.

[0014] The present disclosure further provides nucleic acids (*e.g.*, vectors) comprising nucleotide sequences encoding variant CBH I polypeptides as described herein, and recombinant cells engineered to express the variant CBH I polypeptides. The recombinant cell can be a prokaryotic (*e.g.*, bacterial) or eukaryotic (*e.g.*, yeast or filamentous fungal) cell. Further provided are methods of producing and optionally recovering the variant CBH I polypeptides. Additional embodiments of the recombinant expression system suitable for expression and production of the variant CBH I polypeptides are provided in Section 5.2.

4. BRIEF DESCRIPTION OF THE FIGURES AND TABLES

[0015] **FIGURE 1A-1B:** Cellobiose dose-response curves using a 4-MUL assay for a wild-type CBH I (BD29555; Figure 1A) and a R268K/R411K variant CBH I (BD29555 with the substitutions R273K/R422K; Figure 1B).

[0016] **FIGURE 2A-2B:** The effect of cellobiose accumulation on the activity of wild-type CBH I and a R268K/R411K variant CBH I, based on percent conversion of glucan after 72 hours in the bagasse assay. Figure 2A shows relative activity in the presence (+) and absence (-) of β -glucosidase (BG), where relative activity is normalized to wild type activity with BG (WT+ = 1). Figure 2B shows tolerance to cellobiose as a function of the ratio of activity in the absence vs. presence of β -glucosidase (activity ratio = Activity -BG/Activity +BG).

[0017] **FIGURE 3:** Cellobiose dose-response curves using PASC assay for a R268K/R411K variant CBH I polypeptide as compared to two wild type CBH I polypeptides.

[0018] **FIGURE 4:** The effect of cellobiose accumulation on the activity of a wild-type CBH I and a R268K/R411K variant CBH I based on percent conversion of glucan after 72 hours in the bagasse assay in the presence (+) and absence (-) of β -glucosidase (BG). Activity is normalized to wild type activity with BG (WT+ = 1).

[0019] **FIGURE 5:** Characterization of cellobiose product tolerance of variant CBH I polypeptides, based on percent conversion of glucan after 72 hours in the absence and presence of β -glucosidase (BG) in the bagasse assay; tolerance is evaluated as a function of the ratio of activity in the absence vs. presence of β -glucosidase.

database. These amino acid sequences are mostly wild type, with the exception of some sequences from the PDB database which contain mutations to facilitate protein crystallization.

[0021] TABLE 2: Amino acid positions in the exemplary reference CBH I polypeptides that correspond to R268 and R411 in *T. reesei* CBH I. Database descriptors are as for Table 1.

[0022] TABLE 3: Approximate amino acid positions of CBH I polypeptide domains. Abbreviations used: SS is signal sequence; CD is catalytic domain; and CBD is cellulose binding domain. Database descriptors are as for Table 1.

[0023] TABLE 4: Table 4 shows a segment within the catalytic domain of each exemplary reference CBH I polypeptide containing the active site loop (shown in bold, underlined text) and the catalytic residues (glutamates in most CBH I polypeptides) (shown in bold, double underlined text). Database descriptors are as for Table 1.

[0024] TABLE 5: MUL and bagasse assay results for variants of BD29555. ND means not determined. $\pm$ %Activity (+/- cellobiose) = [(Activity with cellobiose)/(Activity without cellobiose)] * 100. $\% \text{Activity} (-/+ \text{BG}) = [(\text{Activity without BG})/(\text{Activity with BG})] * 100$

[0025] TABLE 6: MUL and bagasse assay results for variants of *T. reesei* CBH I. ND means not determined. $\pm$ %Activity (+/- cellobiose) = [(Activity with cellobiose)/(Activity without cellobiose)] * 100. $\% \text{Activity} (-/+ \text{BG}) = [(\text{Activity without BG})/(\text{Activity with BG})] * 100$.

[0026] TABLE 7: Informal sequence listing. SEQ ID NO:1-149 correspond to the exemplary reference CBH I polypeptides. SEQ ID NO:299 corresponds to mature *T. reesei* CBH I (amino acids 26-529 of SEQ ID NO:2) with an R268A substitution. SEQ ID NO:300 corresponds to mature *T. reesei* CBH I (amino acids 26-529 of SEQ ID NO:2) with an R411A substitution. SEQ ID NO:301 corresponds to full length BD29555 with both an R268K substitution and an R411K substitution. SEQ ID NO:302 corresponds to mature BD29555 with both an R268K substitution and an R411K substitution.

corresponds to mature *T. reesei* CBH I (amino acids 26-529 of SEQ ID NO:2) with an R411A substitution. SEQ ID NO:152 corresponds to full length BD29555 with both an R268K substitution and an R411K substitution. SEQ ID NO:153 corresponds to mature BD29555 with both an R268K substitution and an R411K substitution.

5. DETAILED DESCRIPTION

[0027] The present disclosure relates to variant CBH I polypeptides. Most naturally occurring CBH I polypeptides have arginines at positions corresponding to R268 and R411 of *T. reesei* CBH I (SEQ ID NO:2). The variant CBH I polypeptides of the present disclosure include a substitution at either or both positions resulting in a reduction of product (*e.g.*, cellobiose) inhibition. The following subsections describe in greater detail the variant CBH I polypeptides and exemplary methods of their production, exemplary cellulase compositions comprising them, and some industrial applications of the polypeptides and cellulase compositions.

5.1. VARIANT CBH I POLYPEPTIDES

[0028] The present disclosure provides variant CBH I polypeptides comprising at least one amino acid substitution that results in reduced product inhibition. “Variant” means a polypeptide which differs in sequence from a reference polypeptide by substitution of one or more amino acids at one or a number of different sites in the amino acid sequence.

Exemplary reference CBH I polypeptides are shown in Table 1.

[0029] The variant CBH I polypeptides of the disclosure have an amino acid substitution at the amino acid position corresponding to R268 of *T. reesei* CBH I (SEQ ID NO:2) (an “R268 substitution”), (b) a substitution at the amino acid position corresponding to R411 of *T. reesei* CBH I (“R411 substitution”); or (c) both an R268 substitution and an R411 substitution, as compared to a reference CBH I polypeptide. It is noted that the R268 and R411 numbering is made by reference to the full length *T. reesei* CBH I, which includes a signal sequence that is generally absent from the mature enzyme. The corresponding numbering in the mature *T. reesei* CBH I (see, *e.g.*, SEQ ID NO:4) is R251 and R394, respectively.

[0030] Accordingly, the present disclosure provides variant CBH I polypeptides in which at least one of the amino acid positions corresponding to R268 and R411 of *T. reesei* CBH I, and optionally both the amino acid positions corresponding to R268 and R411 of *T. reesei* CBH I, is not an arginine.

[0031] The amino acid positions in the reference polypeptides of Table 1 that correspond to R268 and R411 in *T. reesei* CBH I are shown in Table 2. Amino acid positions in other CBH I polypeptides that correspond to R268 and R411 can be identified through alignment of their sequences with *T. reesei* CBH I using a sequence comparison algorithm. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, 1981, Adv. Appl. Math. 2:482-89; by the homology alignment algorithm of Needleman & Wunsch, 1970, J. Mol. Biol. 48:443-53; by the search for similarity method of Pearson & Lipman, 1988, Proc. Nat'l Acad. Sci. USA 85:2444-48, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection.

[0032] The R268 and/or R411 substitutions are preferably selected from (a) R268K and R411K; (b) R268K and R411A; (c) R268A and R411K; (d) R268A and R411A; (e) R268A; (f) R268K; (g) R411A; and (h) R411K.

[0033] CBH I polypeptides belong to the glycosyl hydrolase family 7 ("GH7"). The glycosyl hydrolases of this family include endoglucanases and cellobiohydrolases (exoglucanases). The cellobiohydrolases act processively from the reducing ends of cellulose chains to generate cellobiose. Cellulases of bacterial and fungal origin characteristically have a small cellulose-binding domain ("CBD") connected to either the N or the C terminus of the catalytic domain ("CD") via a linker peptide (see Suumakki *et al.*, 2000, Cellulose 7: 189-209). The CD contains the active site whereas the CBD interacts with cellulose by binding the enzyme to it (van Tilbeurgh *et al.*, 1986, FEBS Lett. 204(2): 223-227; Tomme *et al.*, 1988, Eur. J. Biochem. 170:575-581). The three-dimensional structure of the catalytic domain of *T. reesei* CBH I has been solved (Divne *et al.*, 1994, Science 265:524-528). The CD consists of two β -sheets that pack face-to-face to form a β -sandwich.

Most of the remaining amino acids in the CD are loops connecting the β -sheets. Some loops are elongated and bend around the active site, forming cellulose-binding tunnel of (~ 50 Å). In contrast, endoglucanases have an open substrate binding cleft/groove rather than a tunnel. Typically, the catalytic residues are glutamic acids corresponding to E229 and E234 of *T. reesei* CBH I.

[0034] The loops characteristic of the active sites (“the active site loops”) of reference CBH I polypeptides, which are absent from GH7 family endoglucanases, as well as catalytic glutamate residues of the reference CBH I polypeptides, are shown in Table 4. The variant CBH I polypeptides of the disclosure preferably retain the catalytic glutamate residues or may include a glutamine instead at the position corresponding to E234, as for SEQ ID NO:4. In some embodiments, the variant CBH I polypeptides contain no substitutions or only conservative substitutions in the active site loops relative to the reference CBH I polypeptides from which the variants are derived.

[0035] Many CBH I polypeptides do not have a CBD, and most studies concerning the activity of cellulase domains on different substrates have been carried out with only the catalytic domains of CBH I polypeptides. Because CDs with cellobiohydrolase activity can be generated by limited proteolysis of mature CBH I by papain (see, *e.g.*, Chen *et al.*, 1993, Biochem. Mol. Biol. Int. 30(5):901-10), they are often referred to as “core” domains. Accordingly, a variant CBH I can include only the CD “core” of CBH I. Exemplary reference CDs comprise amino acid sequences corresponding to positions 26 to 455 of SEQ ID NO:1, positions 18 to 444 of SEQ ID NO:2, positions 26 to 455 of SEQ ID NO:3, positions 1 to 427 of SEQ ID NO:4, positions 24 to 457 of SEQ ID NO:5, positions 18 to 448 of SEQ ID NO:6, positions 27 to 460 of SEQ ID NO:7, positions 27 to 460 of SEQ ID NO:8, positions 20 to 449 of SEQ ID NO:9, positions 1 to 424 of SEQ ID NO:10, positions 18 to 447 of SEQ ID NO:11, positions 18 to 434 of SEQ ID NO:12, positions 18 to 445 of SEQ ID NO:13, positions 19 to 454 of SEQ ID NO:14, positions 19 to 443 of SEQ ID NO:15, positions 2 to 426 of SEQ ID NO:16, positions 23 to 446 of SEQ ID NO:17, positions 19 to 449 of SEQ ID NO:18, positions 23 to 446 of SEQ ID NO:19, positions 19 to 449 of SEQ ID NO:20, positions 2 to 416 of SEQ ID NO:21, positions 19 to 454 of SEQ ID NO:22,

positions 19 to 447 of SEQ ID NO:23, positions 19 to 447 of SEQ ID NO:24, positions 20 to 443 of SEQ ID NO:25, positions 18 to 447 of SEQ ID NO:26, positions 19 to 442 of SEQ ID NO:27, positions 18 to 451 of SEQ ID NO:28, positions 23 to 446 of SEQ ID NO:29, positions 18 to 444 of SEQ ID NO:30, positions 18 to 451 of SEQ ID NO:31, positions 18 to 447 of SEQ ID NO:32, positions 19 to 449 of SEQ ID NO:33, positions 18 to 447 of SEQ ID NO:34, positions 26 to 459 of SEQ ID NO:35, positions 19 to 450 of SEQ ID NO:36, positions 19 to 453 of SEQ ID NO:37, positions 18 to 448 of SEQ ID NO:38, positions 19 to 443 of SEQ ID NO:39, positions 19 to 442 of SEQ ID NO:40, positions 18 to 444 of SEQ ID NO:41, positions 24 to 457 of SEQ ID NO:42, positions 18 to 449 of SEQ ID NO:43, positions 19 to 453 of SEQ ID NO:44, positions 26 to 456 of SEQ ID NO:45, positions 19 to 451 of SEQ ID NO:46, positions 18 to 443 of SEQ ID NO:47, positions 18 to 448 of SEQ ID NO:48, positions 19 to 451 of SEQ ID NO:49, positions 18 to 444 of SEQ ID NO:50, positions 2 to 419 of SEQ ID NO:51, positions 27 to 461 of SEQ ID NO:52, positions 21 to 445 of SEQ ID NO:53, positions 19 to 449 of SEQ ID NO:54, positions 19 to 448 of SEQ ID NO:55, positions 18 to 443 of SEQ ID NO:56, positions 20 to 443 of SEQ ID NO:57, positions 18 to 448 of SEQ ID NO:58, positions 18 to 447 of SEQ ID NO:59, positions 26 to 455 of SEQ ID NO:60, positions 19 to 449 of SEQ ID NO:61, positions 19 to 449 of SEQ ID NO:62, positions 26 to 460 of SEQ ID NO:63, positions 18 to 448 of SEQ ID NO:64, positions 19 to 451 of SEQ ID NO:65, positions 19 to 447 of SEQ ID NO:66, positions 1 to 424 of SEQ ID NO:67, positions 19 to 448 of SEQ ID NO:68, positions 19 to 443 of SEQ ID NO:69, positions 23 to 447 of SEQ ID NO:70, positions 17 to 448 of SEQ ID NO:71, positions 19 to 449 of SEQ ID NO:72, positions 18 to 444 of SEQ ID NO:73, positions 23 to 458 of SEQ ID NO:74, positions 20 to 452 of SEQ ID NO:75, positions 18 to 435 of SEQ ID NO:76, positions 18 to 446 of SEQ ID NO:77, positions 22 to 457 of SEQ ID NO:78, positions 18 to 448 of SEQ ID NO:79, positions 1 to 431 of SEQ ID NO:80, positions 19 to 453 of SEQ ID NO:81, positions 21 to 440 of SEQ ID NO:82, positions 19 to 442 of SEQ ID NO:83, positions 18 to 448 of SEQ ID NO:84, positions 17 to 446 of SEQ ID NO:85, positions 18 to 447 of SEQ ID NO:86, positions 18 to 443 of SEQ ID NO:87, positions 23 to 448 of SEQ ID NO:88, positions 18 to 451 of SEQ ID NO:89, positions 21 to 447 of SEQ ID

NO:90, positions 18 to 444 of SEQ ID NO:91, positions 19 to 442 of SEQ ID NO:92, positions 20 to 436 of SEQ ID NO:93, positions 18 to 450 of SEQ ID NO:94, positions 22 to 453 of SEQ ID NO:95, positions 16 to 472 of SEQ ID NO:96, positions 21 to 445 of SEQ ID NO:97, positions 19 to 447 of SEQ ID NO:98, positions 19 to 450 of SEQ ID NO:99, positions 19 to 451 of SEQ ID NO:100, positions 18 to 448 of SEQ ID NO:101, positions 19 to 442 of SEQ ID NO:102, positions 20 to 457 of SEQ ID NO:103, positions 19 to 454 of SEQ ID NO:104, positions 18 to 440 of SEQ ID NO:105, positions 18 to 439 of SEQ ID NO:106, positions 27 to 460 of SEQ ID NO:107, positions 23 to 446 of SEQ ID NO:108, positions 17 to 446 of SEQ ID NO:109, positions 21 to 447 of SEQ ID NO:110, positions 19 to 447 of SEQ ID NO:111, positions 18 to 449 of SEQ ID NO:112, positions 22 to 457 of SEQ ID NO:113, positions 18 to 445 of SEQ ID NO:114, positions 18 to 448 of SEQ ID NO:115, positions 18 to 448 of SEQ ID NO:116, positions 23 to 435 of SEQ ID NO:117, positions 21 to 442 of SEQ ID NO:118, positions 23 to 435 of SEQ ID NO:119, positions 20 to 445 of SEQ ID NO:120, positions 21 to 443 of SEQ ID NO:121, positions 20 to 445 of SEQ ID NO:122, positions 23 to 443 of SEQ ID NO:123, positions 20 to 445 of SEQ ID NO:124, positions 21 to 435 of SEQ ID NO:125, positions 20 to 437 of SEQ ID NO:126, positions 21 to 442 of SEQ ID NO:127, positions 23 to 434 of SEQ ID NO:128, positions 20 to 444 of SEQ ID NO:129, positions 21 to 435 of SEQ ID NO:130, positions 20 to 445 of SEQ ID NO:131, positions 21 to 446 of SEQ ID NO:132, positions 21 to 435 of SEQ ID NO:133, positions 22 to 448 of SEQ ID NO:134, positions 23 to 433 of SEQ ID NO:135, positions 23 to 434 of SEQ ID NO:136, positions 23 to 435 of SEQ ID NO:137, positions 23 to 435 of SEQ ID NO:138, positions 20 to 445 of SEQ ID NO:139, positions 20 to 437 of SEQ ID NO:140, positions 21 to 435 of SEQ ID NO:141, positions 20 to 437 of SEQ ID NO:142, positions 21 to 435 of SEQ ID NO:143, positions 26 to 435 of SEQ ID NO:144, positions 23 to 435 of SEQ ID NO:145, positions 24 to 443 of SEQ ID NO:146, positions 20 to 445 of SEQ ID NO:147, positions 21 to 441 of SEQ ID NO:148, and positions 20 to 437 of SEQ ID NO:149.

[0036] The CBDs are particularly involved in the hydrolysis of crystalline cellulose. It has been shown that the ability of cellobiohydrolases to degrade crystalline cellulose decreases

when the CBD is absent (Linder and Teeri, 1997, Journal of Biotechnol. 57:15-28). The variant CBH I polypeptides of the disclosure can further include a CBD. Exemplary CBDs comprise amino acid sequences corresponding to positions 494 to 529 of SEQ ID NO:1, positions 480 to 514 of SEQ ID NO:2, positions 494 to 529 of SEQ ID NO:3, positions 491 to 526 of SEQ ID NO:5, positions 477 to 512 of SEQ ID NO:6, positions 497 to 532 of SEQ ID NO:7, positions 504 to 539 of SEQ ID NO:8, positions 486 to 521 of SEQ ID NO:13, positions 556 to 596 of SEQ ID NO:15, positions 490 to 525 of SEQ ID NO:18, positions 495 to 530 of SEQ ID NO:20, positions 471 to 506 of SEQ ID NO:23, positions 481 to 516 of SEQ ID NO:27, positions 480 to 514 of SEQ ID NO:30, positions 495 to 529 of SEQ ID NO:35, positions 493 to 528 of SEQ ID NO:36, positions 477 to 512 of SEQ ID NO:38, positions 547 to 586 of SEQ ID NO:39, positions 475 to 510 of SEQ ID NO:40, positions 479 to 513 of SEQ ID NO:41, positions 506 to 541 of SEQ ID NO:42, positions 481 to 516 of SEQ ID NO:43, positions 503 to 537 of SEQ ID NO:45, positions 488 to 523 of SEQ ID NO:46, positions 476 to 511 of SEQ ID NO:48, positions 488 to 523 of SEQ ID NO:49, positions 479 to 513 of SEQ ID NO:50, positions 500 to 535 of SEQ ID NO:52, positions 493 to 528 of SEQ ID NO:55, positions 479 to 514 of SEQ ID NO:58, positions 494 to 529 of SEQ ID NO:60, positions 490 to 525 of SEQ ID NO:61, positions 497 to 532 of SEQ ID NO:62, positions 475 to 510 of SEQ ID NO:64, positions 477 to 512 of SEQ ID NO:65, positions 486 to 521 of SEQ ID NO:66, positions 470 to 505 of SEQ ID NO:67, positions 491 to 526 of SEQ ID NO:68, positions 476 to 511 of SEQ ID NO:69, positions 480 to 514 of SEQ ID NO:73, positions 506 to 540 of SEQ ID NO:74, positions 471 to 504 of SEQ ID NO:76, positions 501 to 536 of SEQ ID NO:78, positions 473 to 508 of SEQ ID NO:79, positions 481 to 516 of SEQ ID NO:83, positions 488 to 523 of SEQ ID NO:86, positions 475 to 510 of SEQ ID NO:92, positions 468 to 504 of SEQ ID NO:93, positions 501 to 536 of SEQ ID NO:96, positions 482 to 517 of SEQ ID NO:98, positions 481 to 516 of SEQ ID NO:99, positions 488 to 523 of SEQ ID NO:100, positions 472 to 507 of SEQ ID NO:101, positions 481 to 516 of SEQ ID NO:102, positions 471 to 505 of SEQ ID NO:105, positions 481 to 516 of SEQ ID NO:106, positions 495 to 530 of SEQ ID NO:107, positions 488 to 523 of SEQ ID NO:111, positions 478 to 513 of SEQ ID NO:112, positions 501 to 536 of

SEQ ID NO:113, positions 491 to 526 of SEQ ID NO:115, and positions 503 to 538 of SEQ ID NO:116.

[0037] The CD and CBD are often connected via a linker. Exemplary linker sequences correspond to positions 456 to 493 of SEQ ID NO:1, positions 445 to 479 of SEQ ID NO:2, positions 456 to 493 of SEQ ID NO:3, positions 458 to 490 of SEQ ID NO:5, positions 449 to 476 of SEQ ID NO:6, positions 461 to 496 of SEQ ID NO:7, positions 461 to 503 of SEQ ID NO:8, positions 446 to 485 of SEQ ID NO:13, positions 444 to 555 of SEQ ID NO:15, positions 450 to 489 of SEQ ID NO:18, positions 450 to 494 of SEQ ID NO:20, positions 448 to 470 of SEQ ID NO:23, positions 443 to 480 of SEQ ID NO:27, positions 445 to 479 of SEQ ID NO:30, positions 460 to 494 of SEQ ID NO:35, positions 451 to 492 of SEQ ID NO:36, positions 449 to 476 of SEQ ID NO:38, positions 444 to 546 of SEQ ID NO:39, positions 443 to 474 of SEQ ID NO:40, positions 445 to 478 of SEQ ID NO:41, positions 458 to 505 of SEQ ID NO:42, positions 450 to 480 of SEQ ID NO:43, positions 457 to 502 of SEQ ID NO:45, positions 452 to 487 of SEQ ID NO:46, positions 449 to 475 of SEQ ID NO:48, positions 452 to 487 of SEQ ID NO:49, positions 445 to 478 of SEQ ID NO:50, positions 462 to 499 of SEQ ID NO:52, positions 449 to 492 of SEQ ID NO:55, positions 449 to 478 of SEQ ID NO:58, positions 456 to 493 of SEQ ID NO:60, positions 450 to 489 of SEQ ID NO:61, positions 450 to 496 of SEQ ID NO:62, positions 449 to 474 of SEQ ID NO:64, positions 452 to 476 of SEQ ID NO:65, positions 448 to 485 of SEQ ID NO:66, positions 425 to 469 of SEQ ID NO:67, positions 449 to 490 of SEQ ID NO:68, positions 444 to 475 of SEQ ID NO:69, positions 445 to 479 of SEQ ID NO:73, positions 459 to 505 of SEQ ID NO:74, positions 436 to 470 of SEQ ID NO:76, positions 458 to 500 of SEQ ID NO:78, positions 449 to 472 of SEQ ID NO:79, positions 443 to 480 of SEQ ID NO:83, positions 448 to 487 of SEQ ID NO:86, positions 443 to 474 of SEQ ID NO:92, positions 437 to 467 of SEQ ID NO:93, positions 473 to 500 of SEQ ID NO:96, positions 448 to 481 of SEQ ID NO:98, positions 451 to 480 of SEQ ID NO:99, positions 452 to 487 of SEQ ID NO:100, positions 449 to 471 of SEQ ID NO:101, positions 443 to 480 of SEQ ID NO:102, positions 441 to 470 of SEQ ID NO:105, positions 440 to 480 of SEQ ID NO:106, positions 461 to 494 of SEQ ID NO:107, positions 448 to 487 of SEQ ID NO:111, positions 450 to

478 of SEQ ID NO:112, positions 458 to 500 of SEQ ID NO:113, positions 449 to 490 of SEQ ID NO:115, and positions 449 to 502 of SEQ ID NO:116.

[0038] Because CBH I polypeptides are modular, the CBDs, CDs and linkers of different CBH I polypeptides, such as the exemplary CBH I polypeptides of Table 1, can be used interchangeably. However, in a preferred embodiment, the CBDs, CDs and linkers of a variant CBH I of the disclosure originate from the same polypeptide.

[0039] The variant CBH I polypeptides of the disclosure preferably have at least a two-fold reduction of product inhibition, such that cellobiose has an IC_{50} towards the variant CBH I that is at least 2-fold the IC_{50} of the corresponding reference CBH I, *e.g.*, CBH I lacking the R268 substitution and/or R411 substitution. More preferably the IC_{50} of cellobiose towards the variant CBH I is at least 3-fold, at least 5-fold, at least 8-fold, at least 10-fold, at least 12-fold or at least 15-fold the IC_{50} of the corresponding reference CBH I. In specific embodiments the IC_{50} of cellobiose towards the variant CBH I is ranges from 2-fold to 15-fold, from 2-fold to 10-fold, from 3-fold to 10-fold, from 5-fold to 12-fold, from 4-fold to 12-fold, from 5-fold to 10-fold, from 5-fold to 12-fold, from 2-fold to 8-fold, or from 8-fold to 20-fold the IC_{50} of the corresponding reference CBH I. The IC_{50} can be determined in a phosphoric acid swollen cellulose ("PASC") assay (Du *et al.*, 2010, Applied Biochemistry and Biotechnology 161:313-317) or a methylumbelliferyl lactoside ("MUL") assay (van Tilbeurgh and Claeysens, 1985, FEBS Letts. 187(2):283-288), as exemplified in the Examples below.

[0040] The variant CBH I polypeptides of the disclosure preferably have a cellobiohydrolase activity that is at least 30% the cellobiohydrolase activity of the corresponding reference CBH I, *e.g.*, CBH I lacking the R268 substitution and/or R411 substitution. More preferably, the cellobiohydrolase activity of the variant CBH I is at least 40%, at least 50%, at least 60% or at least 70% the cellobiohydrolase activity of the corresponding reference CBH I. In specific embodiments the IC_{50} cellobiohydrolase activity of the variant CBH I is ranges from 30% to 80%, from 40% to 70%, 30% to 60%, from 50% to 80% or from 60% to 80% of the cellobiohydrolase activity of the corresponding reference CBH I. Assays for cellobiohydrolase activity are described, for example, in Becker *et al.*, 2011, Biochem J.

356:19-30 and Mitsuishi *et al.*, 1990, FEBS Letts. 275:135-138, each of which is expressly incorporated by reference herein. The ability of CBH I to hydrolyze isolated soluble and insoluble substrates can also be measured using assays described in Srisodsuk *et al.*, 1997, J. Biotech. 57:4957 and Nidetzky and Claeysens, 1994, Biotech. Bioeng. 44:961-966. Substrates useful for assaying cellobiohydrolase activity include crystalline cellulose, filter paper, phosphoric acid swollen cellulose, celooligosaccharides, methylumbelliferyl lactoside, methylumbelliferyl cellobioside, orthonitrophenyl lactoside, paranitrophenyl lactoside, orthonitrophenyl cellobioside, paranitrophenyl cellobioside. Cellobiohydrolase activity can be measured in an assay utilizing PASC as the substrate and a calcofluor white detection method (Du *et al.*, 2010, Applied Biochemistry and Biotechnology 161:313-317). PASC can be prepared as described by Walseth, 1952, TAPPI 35:228-235 and Wood, 1971, Biochem. J. 121:353-362.

[0041] Other than said R268 and/or R411 substitution, the variant CBH I polypeptides of the disclosure preferably:

- comprise an amino acid sequence having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to a CD of a reference CBH I exemplified in Table 1 (*i.e.*, a CD comprising an amino acid sequence corresponding to positions 26 to 455 of SEQ ID NO:1, positions 18 to 444 of SEQ ID NO:2, positions 26 to 455 of SEQ ID NO:3, positions 1 to 427 of SEQ ID NO:4, positions 24 to 457 of SEQ ID NO:5, positions 18 to 448 of SEQ ID NO:6, positions 27 to 460 of SEQ ID NO:7, positions 27 to 460 of SEQ ID NO:8, positions 20 to 449 of SEQ ID NO:9, positions 1 to 424 of SEQ ID NO:10, positions 18 to 447 of SEQ ID NO:11, positions 18 to 434 of SEQ ID NO:12, positions 18 to 445 of SEQ ID NO:13, positions 19 to 454 of SEQ ID NO:14, positions 19 to 443 of SEQ ID NO:15, positions 2 to 426 of SEQ ID NO:16, positions 23 to 446 of SEQ ID NO:17, positions 19 to 449 of SEQ ID NO:18, positions 23 to 446 of SEQ ID NO:19, positions 19 to

449 of SEQ ID NO:20, positions 2 to 416 of SEQ ID NO:21, positions 19 to 454 of SEQ ID NO:22, positions 19 to 447 of SEQ ID NO:23, positions 19 to 447 of SEQ ID NO:24, positions 20 to 443 of SEQ ID NO:25, positions 18 to 447 of SEQ ID NO:26, positions 19 to 442 of SEQ ID NO:27, positions 18 to 451 of SEQ ID NO:28, positions 23 to 446 of SEQ ID NO:29, positions 18 to 444 of SEQ ID NO:30, positions 18 to 451 of SEQ ID NO:31, positions 18 to 447 of SEQ ID NO:32, positions 19 to 449 of SEQ ID NO:33, positions 18 to 447 of SEQ ID NO:34, positions 26 to 459 of SEQ ID NO:35, positions 19 to 450 of SEQ ID NO:36, positions 19 to 453 of SEQ ID NO:37, positions 18 to 448 of SEQ ID NO:38, positions 19 to 443 of SEQ ID NO:39, positions 19 to 442 of SEQ ID NO:40, positions 18 to 444 of SEQ ID NO:41, positions 24 to 457 of SEQ ID NO:42, positions 18 to 449 of SEQ ID NO:43, positions 19 to 453 of SEQ ID NO:44, positions 26 to 456 of SEQ ID NO:45, positions 19 to 451 of SEQ ID NO:46, positions 18 to 443 of SEQ ID NO:47, positions 18 to 448 of SEQ ID NO:48, positions 19 to 451 of SEQ ID NO:49, positions 18 to 444 of SEQ ID NO:50, positions 2 to 419 of SEQ ID NO:51, positions 27 to 461 of SEQ ID NO:52, positions 21 to 445 of SEQ ID NO:53, positions 19 to 449 of SEQ ID NO:54, positions 19 to 448 of SEQ ID NO:55, positions 18 to 443 of SEQ ID NO:56, positions 20 to 443 of SEQ ID NO:57, positions 18 to 448 of SEQ ID NO:58, positions 18 to 447 of SEQ ID NO:59, positions 26 to 455 of SEQ ID NO:60, positions 19 to 449 of SEQ ID NO:61, positions 19 to 449 of SEQ ID NO:62, positions 26 to 460 of SEQ ID NO:63, positions 18 to 448 of SEQ ID NO:64, positions 19 to 451 of SEQ ID NO:65, positions 19 to 447 of SEQ ID NO:66, positions 1 to 424 of SEQ ID NO:67, positions 19 to 448 of SEQ ID NO:68, positions 19 to 443 of SEQ ID NO:69, positions 23 to 447 of SEQ ID NO:70, positions 17 to 448 of SEQ ID NO:71, positions 19 to 449 of SEQ ID NO:72, positions 18 to 444 of SEQ ID NO:73, positions 23 to 458 of SEQ ID NO:74, positions 20 to 452 of SEQ ID NO:75, positions 18 to 435 of SEQ ID NO:76, positions 18 to 446 of SEQ ID NO:77, positions 22 to 457 of SEQ ID NO:78, positions 18 to 448 of SEQ ID NO:79, positions 1 to 431 of SEQ ID NO:80, positions

19 to 453 of SEQ ID NO:81, positions 21 to 440 of SEQ ID NO:82, positions 19 to 442 of SEQ ID NO:83, positions 18 to 448 of SEQ ID NO:84, positions 17 to 446 of SEQ ID NO:85, positions 18 to 447 of SEQ ID NO:86, positions 18 to 443 of SEQ ID NO:87, positions 23 to 448 of SEQ ID NO:88, positions 18 to 451 of SEQ ID NO:89, positions 21 to 447 of SEQ ID NO:90, positions 18 to 444 of SEQ ID NO:91, positions 19 to 442 of SEQ ID NO:92, positions 20 to 436 of SEQ ID NO:93, positions 18 to 450 of SEQ ID NO:94, positions 22 to 453 of SEQ ID NO:95, positions 16 to 472 of SEQ ID NO:96, positions 21 to 445 of SEQ ID NO:97, positions 19 to 447 of SEQ ID NO:98, positions 19 to 450 of SEQ ID NO:99, positions 19 to 451 of SEQ ID NO:100, positions 18 to 448 of SEQ ID NO:101, positions 19 to 442 of SEQ ID NO:102, positions 20 to 457 of SEQ ID NO:103, positions 19 to 454 of SEQ ID NO:104, positions 18 to 440 of SEQ ID NO:105, positions 18 to 439 of SEQ ID NO:106, positions 27 to 460 of SEQ ID NO:107, positions 23 to 446 of SEQ ID NO:108, positions 17 to 446 of SEQ ID NO:109, positions 21 to 447 of SEQ ID NO:110, positions 19 to 447 of SEQ ID NO:111, positions 18 to 449 of SEQ ID NO:112, positions 22 to 457 of SEQ ID NO:113, positions 18 to 445 of SEQ ID NO:114, positions 18 to 448 of SEQ ID NO:115, positions 18 to 448 of SEQ ID NO:116, positions 23 to 435 of SEQ ID NO:117, positions 21 to 442 of SEQ ID NO:118, positions 23 to 435 of SEQ ID NO:119, positions 20 to 445 of SEQ ID NO:120, positions 21 to 443 of SEQ ID NO:121, positions 20 to 445 of SEQ ID NO:122, positions 23 to 443 of SEQ ID NO:123, positions 20 to 445 of SEQ ID NO:124, positions 21 to 435 of SEQ ID NO:125, positions 20 to 437 of SEQ ID NO:126, positions 21 to 442 of SEQ ID NO:127, positions 23 to 434 of SEQ ID NO:128, positions 20 to 444 of SEQ ID NO:129, positions 21 to 435 of SEQ ID NO:130, positions 20 to 445 of SEQ ID NO:131, positions 21 to 446 of SEQ ID NO:132, positions 21 to 435 of SEQ ID NO:133, positions 22 to 448 of SEQ ID NO:134, positions 23 to 433 of SEQ ID NO:135, positions 23 to 434 of SEQ ID NO:136, positions 23 to 435 of SEQ ID NO:137, positions 23 to 435 of SEQ ID NO:138, positions 20 to 445 of SEQ ID NO:139,

positions 20 to 437 of SEQ ID NO:140, positions 21 to 435 of SEQ ID NO:141, positions 20 to 437 of SEQ ID NO:142, positions 21 to 435 of SEQ ID NO:143, positions 26 to 435 of SEQ ID NO:144, positions 23 to 435 of SEQ ID NO:145, positions 24 to 443 of SEQ ID NO:146, positions 20 to 445 of SEQ ID NO:147, positions 21 to 441 of SEQ ID NO:148, and positions 20 to 437 of SEQ ID NO:149 (preferably the CD corresponding to positions 26-455 of SEQ ID NO:1 or 18-444 of SEQ ID NO:2); and/or

- comprise an amino acid sequence having at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity to a mature polypeptide of a reference CBH I exemplified in Table 1 (*i.e.*, a mature protein comprising an amino acid sequence corresponding to positions 26 to 529 of SEQ ID NO:1, positions 18 to 514 of SEQ ID NO:2, positions 26 to 529 of SEQ ID NO:3, positions 1 to 427 of SEQ ID NO:4, positions 24 to 526 of SEQ ID NO:5, positions 18 to 512 of SEQ ID NO:6, positions 27 to 532 of SEQ ID NO:7, positions 27 to 539 of SEQ ID NO:8, positions 20 to 449 of SEQ ID NO:9, positions 1 to 424 of SEQ ID NO:10, positions 18 to 447 of SEQ ID NO:11, positions 18 to 434 of SEQ ID NO:12, positions 18 to 521 of SEQ ID NO:13, positions 19 to 454 of SEQ ID NO:14, positions 19 to 596 of SEQ ID NO:15, positions 2 to 426 of SEQ ID NO:16, positions 23 to 446 of SEQ ID NO:17, positions 19 to 525 of SEQ ID NO:18, positions 23 to 446 of SEQ ID NO:19, positions 19 to 530 of SEQ ID NO:20, positions 2 to 416 of SEQ ID NO:21, positions 19 to 454 of SEQ ID NO:22, positions 19 to 506 of SEQ ID NO:23, positions 19 to 447 of SEQ ID NO:24, positions 20 to 443 of SEQ ID NO:25, positions 18 to 447 of SEQ ID NO:26, positions 19 to 516 of SEQ ID NO:27, positions 18 to 451 of SEQ ID NO:28, positions 23 to 446 of SEQ ID NO:29, positions 18 to 514 of SEQ ID NO:30, positions 18 to 451 of SEQ ID NO:31, positions 18 to 447 of SEQ ID NO:32, positions 19 to 449 of SEQ ID NO:33, positions 18 to 447 of SEQ ID NO:34,

positions 26 to 529 of SEQ ID NO:35, positions 19 to 528 of SEQ ID NO:36, positions 19 to 453 of SEQ ID NO:37, positions 18 to 512 of SEQ ID NO:38, positions 19 to 586 of SEQ ID NO:39, positions 19 to 510 of SEQ ID NO:40, positions 18 to 513 of SEQ ID NO:41, positions 24 to 541 of SEQ ID NO:42, positions 18 to 516 of SEQ ID NO:43, positions 19 to 453 of SEQ ID NO:44, positions 26 to 537 of SEQ ID NO:45, positions 19 to 523 of SEQ ID NO:46, positions 18 to 443 of SEQ ID NO:47, positions 18 to 511 of SEQ ID NO:48, positions 19 to 523 of SEQ ID NO:49, positions 18 to 513 of SEQ ID NO:50, positions 2 to 419 of SEQ ID NO:51, positions 27 to 535 of SEQ ID NO:52, positions 21 to 445 of SEQ ID NO:53, positions 19 to 449 of SEQ ID NO:54, positions 19 to 528 of SEQ ID NO:55, positions 18 to 443 of SEQ ID NO:56, positions 20 to 443 of SEQ ID NO:57, positions 18 to 514 of SEQ ID NO:58, positions 18 to 447 of SEQ ID NO:59, positions 26 to 529 of SEQ ID NO:60, positions 19 to 525 of SEQ ID NO:61, positions 19 to 532 of SEQ ID NO:62, positions 26 to 460 of SEQ ID NO:63, positions 18 to 510 of SEQ ID NO:64, positions 19 to 512 of SEQ ID NO:65, positions 19 to 521 of SEQ ID NO:66, positions 1 to 505 of SEQ ID NO:67, positions 19 to 526 of SEQ ID NO:68, positions 19 to 511 of SEQ ID NO:69, positions 23 to 447 of SEQ ID NO:70, positions 17 to 448 of SEQ ID NO:71, positions 19 to 449 of SEQ ID NO:72, positions 18 to 514 of SEQ ID NO:73, positions 23 to 540 of SEQ ID NO:74, positions 20 to 452 of SEQ ID NO:75, positions 18 to 504 of SEQ ID NO:76, positions 18 to 446 of SEQ ID NO:77, positions 22 to 536 of SEQ ID NO:78, positions 18 to 508 of SEQ ID NO:79, positions 1 to 431 of SEQ ID NO:80, positions 19 to 453 of SEQ ID NO:81, positions 21 to 440 of SEQ ID NO:82, positions 19 to 516 of SEQ ID NO:83, positions 18 to 448 of SEQ ID NO:84, positions 17 to 446 of SEQ ID NO:85, positions 18 to 523 of SEQ ID NO:86, positions 18 to 443 of SEQ ID NO:87, positions 23 to 448 of SEQ ID NO:88, positions 18 to 451 of SEQ ID NO:89, positions 21 to 447 of SEQ ID NO:90, positions 18 to 444 of SEQ ID NO:91, positions 19 to 510 of SEQ ID NO:92, positions 20 to 504 of SEQ ID NO:93, positions 18 to 450 of SEQ ID NO:94, positions 22 to 453 of SEQ ID NO:95,

positions 16 to 536 of SEQ ID NO:96, positions 21 to 445 of SEQ ID NO:97, positions 19 to 517 of SEQ ID NO:98, positions 19 to 516 of SEQ ID NO:99, positions 19 to 523 of SEQ ID NO:100, positions 18 to 507 of SEQ ID NO:101, positions 19 to 516 of SEQ ID NO:102, positions 20 to 457 of SEQ ID NO:103, positions 19 to 454 of SEQ ID NO:104, positions 18 to 505 of SEQ ID NO:105, positions 18 to 516 of SEQ ID NO:106, positions 27 to 530 of SEQ ID NO:107, positions 23 to 446 of SEQ ID NO:108, positions 17 to 446 of SEQ ID NO:109, positions 21 to 447 of SEQ ID NO:110, positions 19 to 523 of SEQ ID NO:111, positions 18 to 513 of SEQ ID NO:112, positions 22 to 536 of SEQ ID NO:113, positions 18 to 445 of SEQ ID NO:114, positions 18 to 526 of SEQ ID NO:115, positions 18 to 538 of SEQ ID NO:116, positions 23 to 435 of SEQ ID NO:117, positions 21 to 442 of SEQ ID NO:118, positions 23 to 435 of SEQ ID NO:119, positions 20 to 445 of SEQ ID NO:120, positions 21 to 443 of SEQ ID NO:121, positions 20 to 445 of SEQ ID NO:122, positions 23 to 443 of SEQ ID NO:123, positions 20 to 445 of SEQ ID NO:124, positions 21 to 435 of SEQ ID NO:125, positions 20 to 437 of SEQ ID NO:126, positions 21 to 442 of SEQ ID NO:127, positions 23 to 434 of SEQ ID NO:128, positions 20 to 444 of SEQ ID NO:129, positions 21 to 435 of SEQ ID NO:130, positions 20 to 445 of SEQ ID NO:131, positions 21 to 446 of SEQ ID NO:132, positions 21 to 435 of SEQ ID NO:133, positions 22 to 448 of SEQ ID NO:134, positions 23 to 433 of SEQ ID NO:135, positions 23 to 434 of SEQ ID NO:136, positions 23 to 435 of SEQ ID NO:137, positions 23 to 435 of SEQ ID NO:138, positions 20 to 445, of SEQ ID NO:139, positions 20 to 437 of SEQ ID NO:140, positions 21 to 435 of SEQ ID NO:141, positions 20 to 437 of SEQ ID NO:142, positions 21 to 435 of SEQ ID NO:143, positions 26 to 435 of SEQ ID NO:144, positions 23 to 435 of SEQ ID NO:145, positions 24 to 443 of SEQ ID NO:146, positions 20 to 445 of SEQ ID NO:147, positions 21 to 441 of SEQ ID NO:148, and positions 20 to 437 of SEQ ID NO:149, preferably the mature polypeptide corresponding to positions 26-529 of SEQ ID NO:1 or 18-514 of SEQ ID NO:2).

[0042] An example of an algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, 1990, J. Mol. Biol. 215:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. These initial neighborhood word hits act as starting points to find longer HSPs containing them. The word hits are expanded in both directions along each of the two sequences being compared for as far as the cumulative alignment score can be increased. Extension of the word hits is stopped when: the cumulative alignment score falls off by the quantity X from a maximum achieved value; the cumulative score goes to zero or below; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, 1992, Proc. Nat'l. Acad. Sci. USA 89:10915-10919) alignments (B) of 50, expectation (E) of 10, M'5, N'-4, and a comparison of both strands.

[0043] Most CBH I polypeptides are secreted and are therefore expressed with a signal sequence that is cleaved upon secretion of the polypeptide from the cell. Accordingly, in certain aspects, the variant CBH I polypeptides of the disclosure further include a signal sequence. Exemplary signal sequences comprise amino acid sequences corresponding to positions 1 to 25 of SEQ ID NO:1, positions 1 to 17 of SEQ ID NO:2, positions 1 to 25 of SEQ ID NO:3, positions 1 to 23 of SEQ ID NO:5, positions 1 to 17 of SEQ ID NO:6, positions 1 to 26 of SEQ ID NO:7, positions 1 to 27 of SEQ ID NO:8, positions 1 to 19 of SEQ ID NO:9, positions 1 to 17 of SEQ ID NO:11, positions 1 to 17 of SEQ ID NO:12, positions 1 to 17 of SEQ ID NO:13, positions 1 to 18 of SEQ ID NO:14, positions 1 to 18 of SEQ ID NO:15, positions 1 to 22 of SEQ ID NO:17, positions 1 to 18 of SEQ ID NO:18, positions 1 to 22 of SEQ ID NO:19, positions 1 to 18 of SEQ ID NO:20, positions 1 to 18 of SEQ ID NO:22, positions 1 to 18 of SEQ ID NO:23, positions 1 to 18 of SEQ ID NO:24, positions 1 to 19 of SEQ ID NO:25, positions 1 to 17 of SEQ ID,NO:26, positions 1 to 18 of

SEQ ID NO:27, positions 1 to 17 of SEQ ID NO:28, positions 1 to 22 of SEQ ID NO:29, positions 1 to 18 of SEQ ID NO:30, positions 1 to 17 of SEQ ID NO:31, positions 1 to 17 of SEQ ID NO:32, positions 1 to 18 of SEQ ID NO:33, positions 1 to 17 of SEQ ID NO:34, positions 1 to 25 of SEQ ID NO:35, positions 1 to 18 of SEQ ID NO:36, positions 1 to 18 of SEQ ID NO:37, positions 1 to 17 of SEQ ID NO:38, positions 1 to 18 of SEQ ID NO:39, positions 1 to 18 of SEQ ID NO:40, positions 1 to 17 of SEQ ID NO:41, positions 1 to 23 of SEQ ID NO:42, positions 1 to 17 of SEQ ID NO:43, positions 1 to 18 of SEQ ID NO:44, positions 1 to 25 of SEQ ID NO:45, positions 1 to 18 of SEQ ID NO:46, positions 1 to 17 of SEQ ID NO:47, positions 1 to 17 of SEQ ID NO:48, positions 1 to 18 of SEQ ID NO:49, positions 1 to 17 of SEQ ID NO:50, positions 1 to 26 of SEQ ID NO:52, positions 1 to 20 of SEQ ID NO:53, positions 1 to 18 of SEQ ID NO:54, positions 1 to 18 of SEQ ID NO:55, positions 1 to 17 of SEQ ID NO:56, positions 1 to 19 of SEQ ID NO:57, positions 1 to 17 of SEQ ID NO:58, positions 1 to 17 of SEQ ID NO:59, positions 1 to 25 of SEQ ID NO:60, positions 1 to 18 of SEQ ID NO:61, positions 1 to 18 of SEQ ID NO:62, positions 1 to 25 of SEQ ID NO:63, positions 1 to 17 of SEQ ID NO:64, positions 1 to 18 of SEQ ID NO:65, positions 1 to 18 of SEQ ID NO:66, positions 1 to 18 of SEQ ID NO:68, positions 1 to 18 of SEQ ID NO:69, positions 1 to 23 of SEQ ID NO:70, positions 1 to 17 of SEQ ID NO:71, positions 1 to 18 of SEQ ID NO:72, positions 1 to 17 of SEQ ID NO:73, positions 1 to 22 of SEQ ID NO:74, positions 1 to 19 of SEQ ID NO:75, positions 1 to 17 of SEQ ID NO:76, positions 1 to 17 of SEQ ID NO:77, positions 1 to 21 of SEQ ID NO:78, positions 1 to 18 of SEQ ID NO:79, positions 1 to 18 of SEQ ID NO:81, positions 1 to 20 of SEQ ID NO:82, positions 1 to 18 of SEQ ID NO:83, positions 1 to 17 of SEQ ID NO:84, positions 1 to 16 of SEQ ID NO:85, positions 1 to 17 of SEQ ID NO:86, positions 1 to 17 of SEQ ID NO:87, positions 1 to 22 of SEQ ID NO:88, positions 1 to 17 of SEQ ID NO:89, positions 1 to 20 of SEQ ID NO:90, positions 1 to 17 of SEQ ID NO:91, positions 1 to 18 of SEQ ID NO:92, positions 1 to 19 of SEQ ID NO:93, positions 1 to 17 of SEQ ID NO:94, positions 1 to 21 of SEQ ID NO:95, positions 1 to 15 of SEQ ID NO:96, positions 1 to 20 of SEQ ID NO:97, positions 1 to 18 of SEQ ID NO:98, positions 1 to 18 of SEQ ID NO:99, positions 1 to 18 of SEQ ID NO:100, positions 1 to 17 of SEQ ID NO:101, positions 1 to 18 of SEQ ID NO:102,

positions 1 to 19 of SEQ ID NO:103, positions 1 to 18 of SEQ ID NO:104, positions 1 to 17 of SEQ ID NO:105, positions 1 to 17 of SEQ ID NO:106, positions 1 to 26 of SEQ ID NO:107, positions 1 to 22 of SEQ ID NO:108, positions 1 to 16 of SEQ ID NO:109, positions 1 to 20 of SEQ ID NO:110, positions 1 to 18 of SEQ ID NO:111, positions 1 to 17 of SEQ ID NO:112, positions 1 to 21 of SEQ ID NO:113, positions 1 to 17 of SEQ ID NO:114, positions 1 to 17 of SEQ ID NO:115, positions 1 to 18 of SEQ ID NO:116, positions 1 to 22 of SEQ ID NO:117, positions 1 to 20 of SEQ ID NO:118, positions 1 to 22 of SEQ ID NO:119, positions 1 to 19 of SEQ ID NO:120, positions 1 to 20 of SEQ ID NO:121, positions 1 to 19 of SEQ ID NO:122, positions 1 to 22 of SEQ ID NO:123, positions 1 to 19 of SEQ ID NO:124, positions 1 to 20 of SEQ ID NO:125, positions 1 to 19 of SEQ ID NO:126, positions 1 to 21 of SEQ ID NO:127, positions 1 to 22 of SEQ ID NO:128, positions 1 to 19 of SEQ ID NO:129, positions 1 to 20 of SEQ ID NO:130, positions 1 to 19 of SEQ ID NO:131, positions 1 to 20 of SEQ ID NO:132, positions 1 to 20 of SEQ ID NO:133, positions 1 to 21 of SEQ ID NO:134, positions 1 to 22 of SEQ ID NO:135, positions 1 to 22 of SEQ ID NO:136, positions 1 to 22 of SEQ ID NO:137, positions 1 to 22 of SEQ ID NO:138, positions 1 to 19 of SEQ ID NO:139, positions 1 to 19 of SEQ ID NO:140, positions 1 to 20 of SEQ ID NO:141, positions 1 to 19 of SEQ ID NO:142, positions 1 to 20 of SEQ ID NO:143, positions 1 to 25 of SEQ ID NO:144, positions 1 to 22 of SEQ ID NO:145, positions 1 to 23 of SEQ ID NO:146, positions 1 to 19 of SEQ ID NO:147, positions 1 to 20 of SEQ ID NO:148, and positions 1 to 19 of SEQ ID NO:149.

5.2. RECOMBINANT EXPRESSION OF VARIANT CBH I POLYPEPTIDES

5.2.1. Cell Culture Systems

[0044] The disclosure also provides recombinant cells engineered to express variant CBH I polypeptides. Suitably, the variant CBH I polypeptide is encoded by a nucleic acid operably linked to a promoter.

[0045] Where recombinant expression in a filamentous fungal host is desired, the promoter can be a filamentous fungal promoter. The nucleic acids can be, for example, under the

control of heterologous promoters. The variant CBH I polypeptides can also be expressed under the control of constitutive or inducible promoters. Examples of promoters that can be used include, but are not limited to, a cellulase promoter, a xylanase promoter, the 1818 promoter (previously identified as a highly expressed protein by EST mapping *Trichoderma*). For example, the promoter can suitably be a cellobiohydrolase, endoglucanase, or β -glucosidase promoter. A particularly suitable promoter can be, for example, a *T. reesei* cellobiohydrolase, endoglucanase, or β -glucosidase promoter. Non-limiting examples of promoters include a cbhl, cbh2, egl1, egl2, egl3, egl4, egl5, pki1, gpdl, xyn1, or xyn2 promoter.

[0046] Suitable host cells include cells of any microorganism (e.g., cells of a bacterium, a protist, an alga, a fungus (e.g., a yeast or filamentous fungus), or other microbe), and are preferably cells of a bacterium, a yeast, or a filamentous fungus.

[0047] Suitable host cells of the bacterial genera include, but are not limited to, cells of *Escherichia*, *Bacillus*, *Lactobacillus*, *Pseudomonas*, and *Streptomyces*. Suitable cells of bacterial species include, but are not limited to, cells of *Escherichia coli*, *Bacillus subtilis*, *Bacillus licheniformis*, *Lactobacillus brevis*, *Pseudomonas aeruginosa*, and *Streptomyces lividans*.

[0048] Suitable host cells of the genera of yeast include, but are not limited to, cells of *Saccharomyces*, *Schizosaccharomyces*, *Candida*, *Hansenula*, *Pichia*, *Kluyveromyces*, and *Phaffia*. Suitable cells of yeast species include, but are not limited to, cells of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Candida albicans*, *Hansenula polymorpha*, *Pichia pastoris*, *P. canadensis*, *Kluyveromyces marxianus*, and *Phaffia rhodozyma*.

[0049] Suitable host cells of filamentous fungi include all filamentous forms of the subdivision Eumycotina. Suitable cells of filamentous fungal genera include, but are not limited to, cells of *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Chrysosporium*, *Coprinus*, *Coriolus*, *Corynascus*, *Chaetomium*, *Cryptococcus*, *Filobasidium*, *Fusarium*, *Gibberella*, *Humicola*, *Hypocrea*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Mucor*,

Neocallimastix, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Phlebia*, *Piromyces*, *Pleurotus*, *Scytalidium*, *Schizophyllum*, *Sporotrichum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trametes*, and *Trichoderma*. More preferably, the recombinant cell is a *Trichoderma* sp. (e.g., *Trichoderma reesei*), *Penicillium* sp., *Humicola* sp. (e.g., *Humicola insolens*); *Aspergillus* sp. (e.g., *Aspergillus niger*), *Chrysosporium* sp., *Fusarium* sp., or *Hypocrea* sp. Suitable cells can also include cells of various anamorph and teleomorph forms of these filamentous fungal genera.

[0050] Suitable cells of filamentous fungal species include, but are not limited to, cells of *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Chrysosporium lucknowense*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, *Ceriporiopsis subvermisporea*, *Coprinus cinereus*, *Coriolus hirsutus*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Neurospora intermedia*, *Penicillium purpurogenum*, *Penicillium canescens*, *Penicillium solitum*, *Penicillium funiculosum*, *Phanerochaete chrysosporium*, *Phlebia radiata*, *Pleurotus eryngii*, *Talaromyces flavus*, *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, and *Trichoderma viride*.

[0051] The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the nucleic acid sequence encoding the variant CBH I polypeptide. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art. As noted, many references are available for the culture

and production of many cells, including cells of bacterial and fungal origin. Cell culture media in general are set forth in Atlas and Parks (eds.), 1993, *The Handbook of Microbiological Media*, CRC Press, Boca Raton, FL, which is incorporated herein by reference. For recombinant expression in filamentous fungal cells, the cells are cultured in a standard medium containing physiological salts and nutrients, such as described in Pourquie *et al.*, 1988, *Biochemistry and Genetics of Cellulose Degradation*, eds. Aubert, *et al.*, Academic Press, pp. 71-86; and Ilmen *et al.*, 1997, *Appl. Environ. Microbiol.* 63:1298-1306. Culture conditions are also standard, *e.g.*, cultures are incubated at 28°C in shaker cultures or fermenters until desired levels of variant CBH I expression are achieved. Preferred culture conditions for a given filamentous fungus may be found in the scientific literature and/or from the source of the fungi such as the American Type Culture Collection (ATCC). After fungal growth has been established, the cells are exposed to conditions effective to cause or permit the expression of a variant CBH I.

[0052] In cases where a variant CBH I coding sequence is under the control of an inducible promoter, the inducing agent, *e.g.*, a sugar, metal salt or antibiotics, is added to the medium at a concentration effective to induce variant CBH I expression.

[0053] In one embodiment, the recombinant cell is an *Aspergillus niger*, which is a useful strain for obtaining overexpressed polypeptide. For example *A. niger var. awamori* dgr246 is known to product elevated amounts of secreted cellulases (Goedegebuur *et al.*, 2002, *Curr. Genet.* 41:89-98). Other strains of *Aspergillus niger var awamori* such as GCDAP3, GCDAP4 and GAP3-4 are known (Ward *et al.*, 1993, *Appl. Microbiol. Biotechnol.* 39:738-743).

[0054] In another embodiment, the recombinant cell is a *Trichoderma reesei*, which is a useful strain for obtaining overexpressed polypeptide. For example, RL-P37, described by Sheir-Neiss *et al.*, 1984, *Appl. Microbiol. Biotechnol.* 20:46-53, is known to secrete elevated amounts of cellulase enzymes. Functional equivalents of RL-P37 include *Trichoderma reesei* strain RUT-C30 (ATCC No. 56765) and strain QM9414 (ATCC No. 26921). It is contemplated that these strains would also be useful in overexpressing variant CBH I polypeptides.

[0055] Cells expressing the variant CBH I polypeptides of the disclosure can be grown under batch, fed-batch or continuous fermentations conditions. Classical batch fermentation is a closed system, wherein the compositions of the medium is set at the beginning of the fermentation and is not subject to artificial alternations during the fermentation. A variation of the batch system is a fed-batch fermentation in which the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is likely to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Batch and fed-batch fermentations are common and well known in the art. Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation systems strive to maintain steady state growth conditions. Methods for modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

5.2.2. Recombinant Expression in Plants

[0056] The disclosure provides transgenic plants and seeds that recombinantly express a variant CBH I polypeptide. The disclosure also provides plant products, *e.g.*, oils, seeds, leaves, extracts and the like, comprising a variant CBH I polypeptide.

[0057] The transgenic plant can be dicotyledonous (a dicot) or monocotyledonous (a monocot). The disclosure also provides methods of making and using these transgenic plants and seeds. The transgenic plant or plant cell expressing a variant CBH I can be constructed in accordance with any method known in the art. See, for example, U.S. Patent No. 6,309,872. *T. reesei* CBH I has been successfully expressed in transgenic tobacco (*Nicotiana tabaccum*) and potato (*Solanum tuberosum*). See Hooker *et al.*, 2000, in *Glycosyl Hydrolases for Biomass Conversion*, ACS Symposium Series, Vol. 769, Chapter 4, pp. 55–90.

[0058] In a particular aspect, the present disclosure provides for the expression of CBH I variants in transgenic plants or plant organs and methods for the production thereof. DNA expression constructs are provided for the transformation of plants with a nucleic acid encoding the variant CBH I polypeptide, preferably under the control of regulatory sequences which are capable of directing expression of the variant CBH I polypeptide. These regulatory sequences include sequences capable of directing transcription in plants, either constitutively, or in stage and/or tissue specific manners.

[0059] The expression of variant CBH I polypeptides in plants can be achieved by a variety of means. Specifically, for example, technologies are available for transforming a large number of plant species, including dicotyledonous species (*e.g.*, tobacco, potato, tomato, *Petunia*, *Brassica*) and monocot species. Additionally, for example, strategies for the expression of foreign genes in plants are available. Additionally still, regulatory sequences from plant genes have been identified that are serviceable for the construction of chimeric genes that can be functionally expressed in plants and in plant cells (*e.g.*, Klee, 1987, *Ann. Rev. of Plant Phys.* 38:467-486; Clark *et al.*, 1990, *Virology* 179(2):640-7; Smith *et al.*, 1990, *Mol. Gen. Genet.* 224(3):477-81.

[0060] The introduction of nucleic acids into plants can be achieved using several technologies including transformation with *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*. Non-limiting examples of plant tissues that can be transformed include protoplasts, microspores or pollen, and explants such as leaves, stems, roots, hypocotyls, and cotyls. Furthermore, DNA encoding a variant CBH I can be introduced directly into protoplasts and plant cells or tissues by microinjection, electroporation, particle bombardment, and direct DNA uptake.

[0061] Variant CBH I polypeptides can be produced in plants by a variety of expression systems. For instance, the use of a constitutive promoter such as the 35S promoter of Cauliflower Mosaic Virus (Guilley *et al.*, 1982, *Cell* 30:763-73) is serviceable for the accumulation of the expressed protein in virtually all organs of the transgenic plant. Alternatively, promoters that are tissue-specific and/or stage-specific can be used (Higgins, 1984, *Annu. Rev. Plant Physiol.* 35:191-221; Shotwell and Larkins, 1989, In: The

Biochemistry of Plants Vol. 15 (Academic Press, San Diego: Stumpf and Conn, eds.), p. 297), permit expression of variant CBH I polypeptides in a target tissue and/or during a desired stage of development.

5.3. COMPOSITIONS OF VARIANT CBH I POLYPEPTIDES

[0062] In general, a variant CBH I polypeptide produced in cell culture is secreted into the medium and may be purified or isolated, *e.g.*, by removing unwanted components from the cell culture medium. However, in some cases, a variant CBH I polypeptide may be produced in a cellular form necessitating recovery from a cell lysate. In such cases the variant CBH I polypeptide is purified from the cells in which it was produced using techniques routinely employed by those of skill in the art. Examples include, but are not limited to, affinity chromatography (Van Tilbeurgh *et al.*, 1984, FEBS Lett. 169(2):215-218), ion-exchange chromatographic methods (Goyal *et al.*, 1991, Bioresource Technology, 36:37-50; Fliess *et al.*, 1983, Eur. J. Appl. Microbiol. Biotechnol. 17:314-318; Bhikhabhai *et al.*, 1984, J. Appl. Biochem. 6:336-345; Ellouz *et al.*, 1987, Journal of Chromatography, 396:307-317), including ion-exchange using materials with high resolution power (Medve *et al.*, 1998, J. Chromatography A, 808:153-165), hydrophobic interaction chromatography (Tomaz and Queiroz, 1999, J. Chromatography A, 865:123-128), and two-phase partitioning (Brumbauer *et al.*, 1999, Bioseparation 7:287-295).

[0063] The variant CBH I polypeptides of the disclosure are suitably used in cellulase compositions. Cellulases are known in the art as enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellooligosaccharides, and the like. Cellulase enzymes have been traditionally divided into three major classes: endoglucanases ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases (EC 3.2.1.21) ("BG") (Knowles *et al.*, 1987, TIBTECH 5:255-261; Schulein, 1988, Methods in Enzymology 160(25):234-243).

[0064] Certain fungi produce complete cellulase systems which include exo-cellobiohydrolases or CBH-type cellulases, endoglucanases or EG-type cellulases and β -glucosidases or BG-type cellulases (Schulein, 1988, Methods in Enzymology 160(25):234-243). Such cellulase compositions are referred to herein as "whole" cellulases. However,

sometimes these systems lack CBH-type cellulases and bacterial cellulases also typically include little or no CBH-type cellulases. In addition, it has been shown that the EG components and CBH components synergistically interact to more efficiently degrade cellulose. See, *e.g.*, Wood, 1985, Biochemical Society Transactions 13(2):407-410.

[0065] The cellulase compositions of the disclosure typically include, in addition to a variant CBH I polypeptide, one or more cellobiohydrolases, endoglucanases and/or β -glucosidases. In their crudest form, cellulase compositions contain the microorganism culture that produced the enzyme components. "Cellulase compositions" also refers to a crude fermentation product of the microorganisms. A crude fermentation is preferably a fermentation broth that has been separated from the microorganism cells and/or cellular debris (*e.g.*, by centrifugation and/or filtration). In some cases, the enzymes in the broth can be optionally diluted, concentrated, partially purified or purified and/or dried. The variant CBH I polypeptide can be co-expressed with one or more of the other components of the cellulase composition or it can be expressed separately, optionally purified and combined with a composition comprising one or more of the other cellulase components.

[0066] When employed in cellulase compositions, the variant CBH I is generally present in an amount sufficient to allow release of soluble sugars from the biomass. The amount of variant CBH I enzymes added depends upon the type of biomass to be saccharified which can be readily determined by the skilled artisan. In certain embodiments, the weight percent of variant CBH I polypeptide is suitably at least 1, at least 5, at least 10, or at least 20 weight percent of the total polypeptides in a cellulase composition. Exemplary cellulase compositions include a variant CBH I of the disclosure in an amount ranging from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45 weight percent, from about 10 to about

50 weight percent, from about 15 to about 20 weight percent, from about 15 to about 25 weight percent, from about 15 to about 30 weight percent, from about 15 to about 35 weight percent, from about 15 to about 30 weight percent, from about 15 to about 45 weight percent, or from about 15 to about 50 weight percent of the total polypeptides in the composition.

5.4. UTILITY OF VARIANT CBH I POLYPEPTIDES

[0067] It can be appreciated that the variant CBH I polypeptides of the disclosure and compositions comprising the variant CBH I polypeptides find utility in a wide variety applications, for example detergent compositions that exhibit enhanced cleaning ability, function as a softening agent and/or improve the feel of cotton fabrics (*e.g.*, “stone washing” or “biopolishing”), or in cellulase compositions for degrading wood pulp into sugars (*e.g.*, for bio-ethanol production). Other applications include the treatment of mechanical pulp (Pere *et al.*, 1996, Tappi Pulping Conference, pp. 693-696 (Nashville, TN, Oct. 27-31, 1996)), for use as a feed additive (see, *e.g.*, WO 91/04673) and in grain wet milling.

5.4.1. Saccharification Reactions

[0068] Ethanol can be produced via saccharification and fermentation processes from cellulosic biomass such as trees, herbaceous plants, municipal solid waste and agricultural and forestry residues. However, the ratio of individual cellulase enzymes within a naturally occurring cellulase mixture produced by a microbe may not be the most efficient for rapid conversion of cellulose in biomass to glucose. It is known that endoglucanases act to produce new cellulose chain ends which themselves are substrates for the action of cellobiohydrolases and thereby improve the efficiency of hydrolysis of the entire cellulase system. The use of optimized cellobiohydrolase activity may greatly enhance the production of ethanol.

[0069] Cellulase compositions comprising one or more of the variant CBH I polypeptides of the disclosure can be used in saccharification reaction to produce simple sugars for fermentation. Accordingly, the present disclosure provides methods for saccharification comprising contacting biomass with a cellulase composition comprising a variant CBH I polypeptide of the disclosure and, optionally, subjecting the resulting sugars to fermentation by a microorganism.

[0070] The term “biomass,” as used herein, refers to any composition comprising cellulose (optionally also hemicellulose and/or lignin). As used herein, biomass includes, without limitation, seeds, grains, tubers, plant waste or byproducts of food processing or industrial processing (*e.g.*, stalks), corn (including, *e.g.*, cobs, stover, and the like), grasses (including, *e.g.*, Indian grass, such as *Sorghastrum nutans*; or, switchgrass, *e.g.*, *Panicum* species, such as *Panicum virgatum*), wood (including, *e.g.*, wood chips, processing waste), paper, pulp, and recycled paper (including, *e.g.*, newspaper, printer paper, and the like). Other biomass materials include, without limitation, potatoes, soybean (*e.g.*, rapeseed), barley, rye, oats, wheat, beets, and sugar cane bagasse.

[0071] The saccharified biomass (*e.g.*, lignocellulosic material processed by enzymes of the disclosure) can be made into a number of bio-based products, via processes such as, *e.g.*, microbial fermentation and/or chemical synthesis. As used herein, “microbial fermentation” refers to a process of growing and harvesting fermenting microorganisms under suitable conditions. The fermenting microorganism can be any microorganism suitable for use in a desired fermentation process for the production of bio-based products. Suitable fermenting microorganisms include, without limitation, filamentous fungi, yeast, and bacteria. The saccharified biomass can, for example, be made it into a fuel (*e.g.*, a biofuel such as a bioethanol, biobutanol, biomethanol, a biopropanol, a biodiesel, a jet fuel, or the like) via fermentation and/or chemical synthesis. The saccharified biomass can, for example, also be made into a commodity chemical (*e.g.*, ascorbic acid, isoprene, 1,3-propanediol), lipids, amino acids, polypeptides, and enzymes, via fermentation and/or chemical synthesis.

[0072] Thus, in certain aspects, the variant CBH I polypeptides of the disclosure find utility in the generation of ethanol from biomass in either separate or simultaneous saccharification and fermentation processes. Separate saccharification and fermentation is a process whereby cellulose present in biomass is saccharified into simple sugars (*e.g.*, glucose) and the simple sugars subsequently fermented by microorganisms (*e.g.*, yeast) into ethanol. Simultaneous saccharification and fermentation is a process whereby cellulose present in biomass is saccharified into simple sugars (*e.g.*, glucose) and, at the same time and in the same reactor, microorganisms (*e.g.*, yeast) ferment the simple sugars into ethanol.

[0073] Prior to saccharification, biomass is preferably subject to one or more pretreatment step(s) in order to render cellulose material more accessible or susceptible to enzymes and thus more amenable to hydrolysis by the variant CBH I polypeptides of the disclosure.

[0074] In an exemplary embodiment, the pretreatment entails subjecting biomass material to a catalyst comprising a dilute solution of a strong acid and a metal salt in a reactor. The biomass material can, *e.g.*, be a raw material or a dried material. This pretreatment can lower the activation energy, or the temperature, of cellulose hydrolysis, ultimately allowing higher yields of fermentable sugars. See, *e.g.*, U.S. Patent Nos. 6,660,506; 6,423,145.

[0075] Another exemplary pretreatment method entails hydrolyzing biomass by subjecting the biomass material to a first hydrolysis step in an aqueous medium at a temperature and a pressure chosen to effectuate primarily depolymerization of hemicellulose without achieving significant depolymerization of cellulose into glucose. This step yields a slurry in which the liquid aqueous phase contains dissolved monosaccharides resulting from depolymerization of hemicellulose, and a solid phase containing cellulose and lignin. The slurry is then subject to a second hydrolysis step under conditions that allow a major portion of the cellulose to be depolymerized, yielding a liquid aqueous phase containing dissolved/soluble depolymerization products of cellulose. See, *e.g.*, U.S. Patent No. 5,536,325.

[0076] A further exemplary method involves processing a biomass material by one or more stages of dilute acid hydrolysis using about 0.4% to about 2% of a strong acid; followed by treating the unreacted solid lignocellulosic component of the acid hydrolyzed material with alkaline delignification. See, *e.g.*, U.S. Patent No. 6,409,841. Another exemplary pretreatment method comprises prehydrolyzing biomass (*e.g.*, lignocellulosic materials) in a prehydrolysis reactor; adding an acidic liquid to the solid lignocellulosic material to make a mixture; heating the mixture to reaction temperature; maintaining reaction temperature for a period of time sufficient to fractionate the lignocellulosic material into a solubilized portion containing at least about 20% of the lignin from the lignocellulosic material, and a solid fraction containing cellulose; separating the solubilized portion from the solid fraction, and removing the solubilized portion while at or near reaction temperature; and recovering the solubilized portion. The cellulose in the solid fraction is rendered more amenable to

enzymatic digestion. See, *e.g.*, U.S. Patent No. 5,705,369. Further pretreatment methods can involve the use of hydrogen peroxide H_2O_2 . See Gould, 1984, Biotech, and Bioengr. 26:46-52.

[0077] Pretreatment can also comprise contacting a biomass material with stoichiometric amounts of sodium hydroxide and ammonium hydroxide at a very low concentration. See Teixeira *et al.*, 1999, Appl. Biochem. and Biotech. 77-79:19-34. Pretreatment can also comprise contacting a lignocellulose with a chemical (*e.g.*, a base, such as sodium carbonate or potassium hydroxide) at a pH of about 9 to about 14 at moderate temperature, pressure, and pH. See PCT Publication WO2004/081185.

[0078] Ammonia pretreatment can also be used. Such a pretreatment method comprises subjecting a biomass material to low ammonia concentration under conditions of high solids. See, *e.g.*, U.S. Patent Publication No. 20070031918 and PCT publication WO 06/110901.

5.4.2. Detergent Compositions Comprising Variant CBH I Proteins

[0079] The present disclosure also provides detergent compositions comprising a variant CBH I polypeptide of the disclosure. The detergent compositions may employ besides the variant CBH I polypeptide one or more of a surfactant, including anionic, non-ionic and ampholytic surfactants; a hydrolase; a bleaching agents; a bluing agent; a caking inhibitors; a solubilizer; and a cationic surfactant. All of these components are known in the detergent art.

[0080] The variant CBH I polypeptide is preferably provided as part of cellulase composition. The cellulase composition can be employed from about 0.00005 weight percent to about 5 weight percent or from about 0.0002 weight percent to about 2 weight percent of the total detergent composition. The cellulase composition can be in the form of a liquid diluent, granule, emulsion, gel, paste, and the like. Such forms are known to the skilled artisan. When a solid detergent composition is employed, the cellulase composition is preferably formulated as granules.

6. EXAMPLES

6.1. MATERIALS AND METHODS

6.1.1. Preparation Of CBH I Polypeptides For Biochemical Characterization

[0081] Protein expression was carried out in an *Aspergillus niger* host strain that had been transformed using PEG-mediated transformation with expression constructs for CBHI that included the hygromycin resistance gene as a selectable marker, in which the full length CBH I sequences (signal sequence, catalytic domain, linker and cellulose binding domain) were under the control of the glyceraldehyde-3-phosphate dehydrogenase (gpd) promoter. Transformants were selected on the regeneration medium based on resistance to hygromycin. The selected transformants were cultured in *Aspergillus* salts medium, pH 6.2 supplemented with the antibiotics penicillin, streptomycin, and hygromycin, and 80g/L glycerol, 20g/L soytone, 10mM uridine, 20g/L MES) in baffled shake flasks at 30°C, 170 rpm. After five days of incubation, the total secreted protein supernatant was recovered, and then subjected to hollow fiber filtration to concentrate and exchange the sample into acetate buffer (50 mM NaAc, pH 5). CBH I protein represented over 90% of the total protein in these samples. Protein purity was analyzed by SDS-PAGE. Protein concentration was determined by gel densitometry and/or HPLC analysis. All CBH I protein concentrations were normalized before assay and concentrated to 1-2.5 mg/ml..

6.1.2. CBH I Activity Assays

[0082] **4- Methylumbelliferyl Lactoside (4-MUL) Assay:** This assay measures the activity of CBH I on the fluorogenic substrate 4-MUL (also known as MUL). Assays were run in a costar 96-well black bottom plate, where reactions were initiated by the addition of 4-MUL to enzyme in buffer (2mM 4-MUL in 200mM MES pH 6). Enzymatic rates were monitored by fluorescent readouts over five minutes on a SPECTRAMAXTM plate reader (ex/em 365/450 nm). Data in the linear range was used to calculate initial rates (V_o).

[0083] **Phosphoric Acid Swollen Cellulose (PASC) Assay:** This assay measures the activity of CBH I using PASC as the substrate. During the assay, the concentration of PASC is monitored by a fluorescent signal derived from calcofluor binding to PASC (ex/em

365/440 nm). The assay is initiated by mixing enzyme (15 μ l) and reaction buffer (85 μ l of 0.2% PASC, 200 mM MES, pH 6), and then incubating at 35°C while shaking at 225 RPM. After 2 hours, one reaction volume of calcofluor stop solution (100 μ g/ml in 500 mM glycine pH 10) is added and fluorescence read-outs obtained (ex/em 365/440 nm).

[0084] Bagasse Assay: This assay measures the activity of CBH I on bagasse, a lignocellulosic substrate. Reactions were run in 10 ml vials with 5% dilute acid pretreated bagasse (250 mg solids per 5 ml reaction). Each reaction contained 4 mg CBH I enzyme/g solids, 200 mM MES pH 6, kanamycin, and chloramphenicol. Reactions were incubated at 35°C in hybridization incubators (Robbins Scientific), rotating at 20 RPM. Time points were taken by transferring a sample of homogenous slurry (150 μ l) into a 96-well deep well plate and quenching the reaction with stop buffer (450 μ l of 500 mM sodium carbonate, pH 10). Time point measurements were taken every 24 hours for 72 hours.

[0085] Cellobiose Tolerance Assays (or Cellobiose Inhibition Assays): Tolerance to cellobiose (or inhibition caused by cellobiose) was tested in two ways in the CBH I assays. A direct-dose tolerance method can be applied to all of the CBH I assays (*i.e.*, 4-MUL, PASC, and/or bagasse assays), and entails the exogenous addition of a known amount of cellobiose into assay mixtures. A different indirect method entails the addition of an excess amount of β -glucosidase (BG) to PASC and bagasse assays (typically, 1 mg β -glucosidase/g solids loaded). BG will enzymatically hydrolyze the cellobiose generated during these assays; therefore, CBH I activity in the presence of BG can be taken as a measure of activity in the absence of cellobiose. Furthermore, when activity in the presence and absence of BG are similar, this indicates tolerance to cellobiose. Notably, in cases where BG activity is undesired, but may be present in crude CBH I enzyme preparations, the BG inhibitor gluconolactone can be added into CBH I assays to prevent cellobiose breakdown.

6.2. LIBRARY SCREENING ASSAYS

[0086] The wild type CBH I polypeptide BD29555 was mutagenized to identify variants with improved product tolerance. A small (60-member) library of BD29555 variants was designed to identify variant CBH I polypeptides with reduced product inhibition. This product-release-site library was designed based on residues directly interacting with the

cellobiose product in an attempt to identify variants with weakened interactions with cellobiose from which the product would be released more readily than the wild type enzyme. The 60-member evolution library contained wild-type residues and mutations at positions R273, W405, and R422 of BD29555 (SEQ ID NO:1), and included the following substitutions: R273 (WT), R273Q, R273K, R273A, W405 (WT), W405Q, W405H, R422 (WT), R422Q, R422K, R422L, and R422E (4 variants at position 273 X 3 variants at position 405 X 5 variants at position 422 equals 60 variants in total). All members of the library were screened using the 4-MUL assay in the presence and absence of 250g/L cellobiose and using gluconolactone to inhibit any BG activity. The R273A, R273Q, and R273K/R422K variants showed enhanced product tolerance. The R273K/R422K variant showed greatest activity among the variants and cellobiose tolerance at 250 mg/L. Due to low expression, the R273K variant was not tested for product inhibition.

6.3. CHARACTERIZATION OF PRODUCT TOLERANT VARIANTS OF BD29555

[0087] The R273K/R422K substitutions were characterized in both a wild type BD29555 background and also in combination with the substitutions Y274Q, D281K, Y410H, P411G, which were identified in a screen of an expanded product release site evolution library.

[0088] The wild type, the R273K/R422K variant and the R273K/Y274Q/D281K/Y410H/P411G/R422K variants were tested for activity on 4-MUL in the presence and absence of 250mg/L cellobiose, and the R273K/R422K variant was also tested in the bagasse assay in the presence and absence of BG. The results are summarized in Table 5.

[0089] The results from these activity assays were converted into the percentage of activity remaining with and without cellobiose present, where values close to 100% indicated cellobiose tolerance. The percent of activity remaining in the MUL assay in the presence cellobiose versus in the absence of cellobiose shows that the R273K/R422K variant was the most tolerant, followed by the R273K/Y274Q/D281K/Y410H/P411G/R422K variant, and then wild-type, at 95%, 78%, and 25% activity, respectively.

[0090] Cellobiose dose response curves of the wild-type and R273K/R422K variant of BD29555 were obtained during the 4-MUL assay. Enzyme rates (V_o) were measured in the presence of different concentrations of cellobiose (200 mM MES pH 6, 25°C). Rates were measured in quadruplicate. The results are shown in Figure 1A-1B. Figure 1A shows that wild type BD2955 is inhibited by cellobiose, with a half maximal inhibitory concentration (IC_{50} value) of 60 mg/L. Figure 1B shows that the R273K/R422K variant is tolerant to cellobiose up to 250 mg/L.

[0091] The bagasse assay results shown in Table 5, which lists the percentage of activity remaining in the absence vs. presence of BG, also demonstrate that the percentage activity of the wild type BD29555 is lower than the percentage activity of the R273K/R422K variant, indicating that the R273K/R422K variant is less sensitive to the presence of cellobiose than the wild type. Figure 2A-2B shows bar graph data for the bagasse assay of BD29555 vs. the R273K/R422K variant. In Figure 2A, bars represent relative activity, which has been normalized to wild type activity in the absence of cellobiose (WT +BG = uninhibited activity = 1). In Figure 2B, bars indicate tolerance to cellobiose, as represented by the ratio of activity in the presence of cellobiose (-BG) to that of activity in the absence of cellobiose (+BG); ratios close to 1 indicate greater tolerance to cellobiose. These data again demonstrate that the R273K/R422K variant of BD29555 is more tolerant to cellobiose than the wild type BD29555.

[0092] The wild type and R273K/R422K variant were also characterized in the PASC assay. Results are shown in Figure 3. The activities of both wild type BD29555 (SEQ ID NO:1) and wild type *T. reesei* CBH I (SEQ ID NO:2) were inhibited by cellobiose concentrations starting around 1 g/L (with IC_{50} values of 2.2 and 3 g/L, respectively), whereas the R273K/R422K variant showed little inhibition in the presence of 10 g/L cellobiose.

6.4. CHARACTERIZATION OF PRODUCT TOLERANT VARIANTS OF *T. REESEI* CBH I

[0093] Cellobiose product tolerant substitutions were introduced into *T. reesei* CBH I (SEQ ID NO:2). A panel of variants with single and double alanine and lysine substitutions at R268 and R411 were expressed and analyzed. The variants were tested for activity on 4-

MUL in the presence and absence of 250mg/L cellobiose and also in the bagasse assay in the absence and presence of BG. The results from these assays were converted into the percentage activity remaining in the presence and absence of cellobiose and BG, respectively. Values are summarized in Table 6.

[0094] The 4-MUL assay results shown in Table 6 demonstrate that the activity of the wild type *T. reesei* CBH I was reduced to 23% in the presence of cellobiose, whereas the double mutants at R268 and R411 retained more than 90% of their activity under the same conditions.

[0095] The bagasse assay results shown in Table 6 demonstrate that the activity of the wild type *T. reesei* CBH I is more significantly impacted by the presence of BG than is the activity of the single or double substitution variants, indicating that the variants are less sensitive to the accumulation of cellobiose than the wild type. Figures 4 and 5 show bar graph data for the bagasse assay of wild type *T. reesei* CBH I vs. the variants. In Figure 4, bars represent relative activity, normalized to wild type activity in the absence of cellobiose (WT +BG = 1). In Figure 5, bars represent tolerance to cellobiose, as represented by the ratio of activity in the presence of accumulating cellobiose (-BG) to that of activity in the absence of cellobiose (+BG); ratios close to 1 indicate greater tolerance to cellobiose.

7. SPECIFIC EMBODIMENTS AND INCORPORATION BY REFERENCE

[0096] All publications, patents, patent applications and other documents cited in this application are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference for all purposes.

[0097] While various specific embodiments have been illustrated and described, it will be appreciated that various changes can be made without departing from the spirit and scope of the invention(s).

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:1	BD29555*	Unknown	MSALNSFNMY KSAIILGSL L ATAGAQQIGT YTAETHPSLS WSTCKSGGSC TTNSGAITLD ANWRWVHG VN TSTNCYTGNT WNTAICDTDA SCAQDCALDG ADYSGTYGIT TSGNSLRLNF VTGSNVGSRT YLMADNTHYQ IFDLLNQEFT FTVDVSHLPC GLNGALYFVT MDADGGVSKY PNNKAGAQYG VGYCDSQCPR DLKFIAGQAN VEGWTPSSNN ANTGLGNHGA CCAELDIWEA NSISEALTPH PCDTPGLSVC TTDACGGTYS SDRYAGTCDP DGCDFNPYRL GVTDFYGS GK TVDTTKPITV VTQFVTDDGT STGTLSEIRR YYVQNGVVIP QPSSKISGVS GNVINSDFCD AEISTFGETA SFSKHGGLAK MGAGMEAGMV LVMSLWDDYS VNMLWLDSTY PTNATGTPGA ARGSCPTTSG DPKTVESQSG SSYVTFSDIR VGPFNSTFSG GSSTGGSSTT TASGTTTTKA SSTSTSSTST GTGVAAHWGQ CGGQGWGTGPT TCASGTTCTV VNPYYSQCL	
SEQ ID NO:2	340514556	<i>Trichoderma reesei</i>	MYRKLAVIS A FLATARAQSA CTIQSETHPP LTWQKCSSGG TCTQQTGSVV IDANWRWTHA TNSSTNCYDG NTWSSSTLC PD NETCAKNCC L DGAAYASTYG VTTSGNSLSI GFVTQSAQKN VGARLYIMAS DTTYQEFTLL GNEFSFDVDV SQLPCGLNGA LYFVSMADAG GVSKYPTNTA GAKYGTGYCD SQCPRLKFI NGQANVEGWE PSSNNANTGI GGHGSCCSEM DIWEANSISE ALTPHPCTTV GQEICEGDGC GGTYS DNRYG GTCDDPDGCDW NPYRLGNTSF YGPGSSFTLD TTKKLTVV TQ FETSGAINRY YVQNGVTFQ Q PNAELGSYG NELNDDYCTA EEAEFGGSSF SDKGGLTQFK KATSGGMVLV MSLWDDYYAN MLWLDSTYPT NETSSTPGAV RGSCSTSSGV PAQVESQSPN AKVTF SNIKE GPIGSTGNPS GGNPPGGNPP GTTTTTRRPAT TTGSSSPGPTQ SHYGQCGGIG YSGPTVCASG TTCQVLNPHY SQCL	
SEQ ID NO:3	51243029	<i>Penicillium occitanis</i>	MSALNSFNMY KSAIILGSL L ATAGAQQIGT YTAETHPSLS WSTCKSGGSC TTNSGAITLD ANWRWVHG VN TSTNCYTGNT WNSAICDTDA SCAQDCALDG ADYSGTYGIT TSGNSLRLNF VTGSNVGSRT YLMADNTHYQ IFDLLNQEFT FTVDVSHLPC GLNGALYFVT MDADGGVSKY PNNKAGAQYG VGYCDSQCPR DLKFIAGQAN VEGWTPSANN ANTGIGNHGA CCAELDIWEA NSISEALTPH PCDTPGLSVC TTDACGGTYS SDRYAGTCDP DGCDFNPYRL GVTDFYGS GK TVDTTKPFTV VTQFVTNDGT STGSLSEIRR YYVQNGVVIP QPSSKISGIS GNVINSDYCA AEISTFGGTA SFNKHGGLTN MAAGMEAGMV LVMSLWDDYA VNMLWLDSTY PTNATGTPGA ARGTCATTSG DPKTVESQSG SSYVTFSDIR VGPFNSTFSG GSSTGGSTTT TASRTTTTSA SSTSTSSTST GTGVAGHWGQ CGGQGWGTGPT TCVSGTTCTV VNPYYSQCL	
SEQ ID NO:4	7cel (PDB) &	<i>Trichoderma reesei</i>	ESACTLQSET HPPLTWQKCS SGGTCTQQTG SVIDANWRW THATNSSTNC YDGNTWSSTL CPDNETCAKN CCLDGAAYAS TYGVTTSGNS LSIDFVTQSA QKNVGARLYL MASDTTYQEF TLLGNEFSFD VDVSQLPCGL NGALYFVSM D ADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KFRINGQANVE GWEPSSNNAN TGIGGHGSCC SEMDIWQANS ISEALTPHPC TTVGQEICEG DCGGGTYS DN RYGGTCDDPG CDWNPYRLGN TSFYGP GSSF TLDTTKKLTV VTQFETSGAI NRYVQNGVT FQQPNAELGS YSGNELNDDY CTAEAEAFGG SSFSDKGGLT QFKKATSGGM VLVMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS SGVPAQVESQ SPNAKVTFSN IKEGPIGSTG NPSG	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:5	67516425	<i>Aspergillus nidulans</i> FGSC A4	MASSFQLYKA LLFFSSLLSA VQAQKVGTOQ AEVHPGLTWQ TCTSSGSCTT VNGEVTIDAN WRWLHTVNGY TNCYTGNEWD TSICTSNEVC AEQCAVDGAN YASTYGITTS GSSLRLNFVT QSQQKNIGSR VYLMDDEDTY TMFYLLNKEF TFDVDVSELP CGLNGAVYFV SMDADGGKSR YATNEAGAKY GTGYCDSQCP RDLKFFINGVA NVEGWESSDT NPNGGVGNHG SCCAEMDIWE ANSISTAFTP HPCDTPGQTL CTGDSCGGTY SNDRYGGTCD PDGCDFNSYR QGNKTFYGP G LTVDTNSPVT VVTQFLTDDN TDTGTLSEIK RFYVQNGVVI PNSESTYPAN PGNSITTEFC ESQKELFGDV DVFSAHGGMA GMGAAL EQGM VLVL SLWDDN YSNMLWLDN YPTDADPTQP GIARGTCPTD SGVPSEVEAQ YPNAYVVYSN IKFGPIGSTF GNGGGSGPPT TVTSTATST TSSATSTATG QAQHWEQCGG NGWTGPTVCA SPWACTVVNS WYSQCL	
SEQ ID NO:6	46107376	<i>Gibberella zeae</i> PH-1	MYRAIATASA LIAAVRAQQV CSLTQESKPS LNWSKCTSSG CSNVKGSVTI DANWRWTHQV SGSTNCYTGN KWDTSVCTSG KVCAEKCCLD GADYASTYGI TSSGDQLSL S FVTKGPYSTN IGSRTYIMED ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSMDADG GKAKYPGNKA GAKYGTGYCD AQCPRDVKFI NGQANSDGWQ PSDSDVNGGI GNLGTCCEM DIWEANSIST AYTPHPCTKL TQHSCTGDSC GGTYSNDRYG GTCDDADG CDF NSYRQGNKTF YGPGSGFNVD TTKKVTVV TQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGVPGNSLTA DFCTKQKKVF NDPDDFTKKG AWSGMSDALE APMVLVMSLW HDHHSNMLWL DSTYPTDSTK LGSQRGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YKSDGTTPTN PTNPSEPSNT ANPNPGTV DQ WGQCGGSNYS GPTACKSGFT CKKINDFYSQ CQ	
SEQ ID NO:7	70992391	<i>Aspergillus fumigatus</i> Af293	MLASTFSYRM YKTALIL AAL LGSGQAQQVG TSQAEVHPSM TWQSCTAGGS CTNNGKVVI DANWRWVHKV GDYTNCYTGN TWDTTICPDD ATCASNCAL E GANYESTYGV TASGNSLR LN FVTTSQQKNI GSRLYMMKDD STYEMFKLLN QEFTFDVDVS NLP CGLNGAL YFVAMDADGG MSKYPTNKAG AKYGTGYCDS QCPRDLKFIN GQANVEGWQP SSNDANAGTG NHGSCCAEMD IWEANSISTA FTPHPCDTPG QVMCTGDACG GTYSSDRYGG TCDPDGCDEN SFRQGNKTFY GPGMTVDTKS KFTVV TQFIT DDGTSSGTLK EIKRFYVQNG KVIPNSESTW TGVSGNSITT EYCTAQKSLF QDQNVFEKHG GLEGMGAALA QGMVLVMSLW DDHSANMLWL DSNYP TTASS TTPGVARGTC DISSGVPADV EANHPPDAYVV YSNIKVGPIG STFNSGGSNP GGGTTTTTTT QPTTTTTTAG NPGGTGVAQH YGQCGGIGWT GPTTCASPYT CQKLNDYYSQ CL	
SEQ ID NO:8	121699984	<i>Aspergillus clavatus</i> NRRL 1	MLPSTISYRI YKNALFFAAL FGAVQAQKVG TSKAEVHPSM AWQCAADGT CTTKNGKVVI DANWRWVHDV KGYTNCYTGN TWNAELCPDN ESCAENCALE GADYAAATYGA TTSGNALSLK FVTQSQQKNI GSRLYMMKDD NTYETFKLLN QEFTFDVDVS NLP CGLNGAL YFVSMDADGG LSRYTGNEAG AKYGTGYCDS QCPRDLKFIN GLANVEGWTP SSSDANAGNG GHGSCCAEMD IWEANSISTA YTPHPCDTPG QAMCNGDSCG GTYSSDRYGG TCDPDGCDEN SYRQGNKSFY GPGMTVDTKK KMTVV TQFLT NDGTATGTL S EIKRFYVQDG KVIANSESTW PNLGNSLTN DFCKAQKTVF GMDTFSKHG GMEGMGAALA EGMVLVMSLW DDHNSNMLWL DSNSP TTGTS TTPGVARGSC DISSGDPKDL EANHPPDASVV YSNIKVGPIG STFNSGGSNP GGS TTTTKPA TSTTTTKATT TATTNTTGPT GTGVAQPW AQ CGGIGYSGPT QCAAPYTCTK QNDYYSQCL	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:9	1906845	<i>Claviceps purpurea</i>	MHPSLQTILL SALTFTAHAAQ QACSSKPEETH PPLSWSRCSR SGRSRVQGAV TVDANWLWTT VDGSQLNCYTG NRWDTASICSS EKTCSESCCI DGADYAGTYG VTTTGDALSL KRVQQGPYSK NVGSRLYLMK DESRYEMFTL LGNEFTFDVD VSKLGCGLNG ALYFVSMDED GGMKRFPMNK AGAKFGTGYC DSQCPRDVKE INGMANSKDW IPSKSDANAG IGSLGACCRI MDIWEANNIA SAFTPHPCNK SAYHSCTGDG CGGTYSKNRY SGDCDPPDGGD FNSYRLGNIT FYGPGPKFTI DTRKISVVT QFLKGRDGL REIKRFYVQN GKVIPNSVSR VRGVPGNSIT QGFCNAQKKM FGAHESFNAK GGMKGMSAAV SKPMVLVMSL WDDHNSNMLW LDSTYPTNSR QRGSKRGSCP ASSGRPTDVE SSAPDSTVVF SNIKFGPIGS TFSRGK	
SEQ ID NO:10	1gpi (PDB) &	<i>Phanerochaete chrysosporium</i>	ESACTLQSET HPPLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC YDGNWTWSSL CPDNETCAKN CCLDGAAYAS TYGVTTSGNS LSIDEVTTQSA QKNVGARLYL MASDTTYQEF TLLGNEFSFD VDVSQLPCGL NGALYFVSM DADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KFINQANVE GWEPSSNNAN TGIGGHGSCC SEMDIWQANS ISEALTPHPC TTVGQEICEG DCGGTYS DN RYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTTKKLTV VTQFETSGAI NRYVQNGVT FQQPNAELGS YSGNELNDDY CTAEAEAFGG SSFSDKGGLT QFKKATSGGM VLVMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS SGVPAQVESQ SPNAKVTFSN IKFGPIGSTG NPSG	
SEQ ID NO:11	119468034	<i>Neosartorya fischeri</i> NRRL 181	MHQRALLFSA LAVAANAQV GTQKPETHPP LTWQKCTAAG SCSQSGSVV IDANWRWLHS TKDTTNCYTG NTWNTELCPD NESCAQNCAV DGADYAGTYG VTTSGSELKL SEVTGANVGS RLYLMQDDET YQHFNLINNE FTFDQVSVNL PCGLNGALYF VAMDADGMS KYP SNKAGAK YGTGYCDSQC PRDLKFINGM ANVEGWKPSS NDKNAGVGGH GSCCPMDIW EANSISTAVT PHPCDDVSQT MCSGDACGGT YSATRYAGTC DPGCDFNPF RMGNESFYGP GKIVDTKSEM TVVTQFITAD GTDTGALSEI KRLYVQNGKV IANSVSNVAD VSGNSISSDF CTAQKKAFGD EDIFAKHGGL SGMGKALSEM VLIMSIWDDH HSSMMWLDST YPTDADPSKP GVARGTCEHG AGDPEKVESQ HPDASVTFSN IKFGPIGSTY KA	
SEQ ID NO:12	7804883	<i>Leptosphaeria maculans</i>	MYRSLIFATS LSLAKGQLV GNLYCKGSC T AKNGKVVIDA NWRWLHVKG YTNCTGTNEW NATACPDNKS CATNCAIDGA DYRRLRHYCE RQLLGTEVHH QGLYSTNIGS RYLMQDDST YQLEKFTGSQ EFTFDVDLSN LPCGLNGALY FVSM DADGGL KKYPTNKAGA KYGTGYCDAQ CPRDLKFING EGNVEGWQPS KNDQNAVGVG HGSCCAEMDI WEANSVSTAV TPHSCSTIEQ SRC DGDGCGG TYSADRYAGV CDPDGCDFNS YRMGVKDFYG KGKTVDTSKK FTVVTQFIGS GDAMEIKRFY VQNGKTIPQP DSTIPGV TGN SITTFECDAQ KKA FGDKYTF KDKGGMANMP STCNMGVLMV SLWDDHYSNM LWLDSTYPTD KNPDTDAGSG RGECAITSGV PADVESQHPD ASVIYSNIKE GPINTTFG	

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:13	85108032	<i>Neurospora crassa</i> N150 (OR74A)	MLAKFAALAA LVASANAQAV CSLTAETHPS LNWSKCTSSG CTNVAGSITV DANWRWTHIT SGSTNCYSGN EWDTSLCSTN TDCATKCCVD GAEYSSTYGI QTSGNSLSLQ FVTKGSYSTN IGSRTYLMNG ADAYQGFE LL GNEFTFDVDV SGTGCGLNGA LYFVSMIDL DG GKAKYTN NKA GAKYGTGYCD AQCPRDLK YI NGIANVEGWT PSTNDANAGI GDHGTCCSEM DIWEANKVST AFTPHPCTTI EQHMCEGDSC GGTYSDDDRYG GTCDDADG CDF NSYRMGNTTF YGEGKTV DTS SKFTVV TQFI KDSAGDLAEI KRFYVQNGKV IENSQSNVDG VSGNSITQSF CNAQKTAFGD IDDFNKKGGL KQMGKALAKP MVLVMSIWDD HAANMLW LDS TYPVEGGPGA YRGECP TTS G VPAEVEANAP NSKVIFSNIK FGPIGSTFSG GSSGTPPSNP SSSVKPVTST AKPSSTSTAS NPSGTGA AHW AQCGGIGFSG PTTCQSPYTC QKINDYYSQC V
SEQ ID NO:14	169859458	<i>Coprinopsis cinerea okayama</i>	MEKKVALTAL CFLAVAQAQQ VGREVAENHP RLPWQRCTRN GGCQTVSNGQ VVLDANWRWL HVT DGYTNCY TGN SWNSTVC SDPTTCAQRC ALEGANYQQT YGITTNGDAL TIKFLTRSQQ TNVGARVYLM ENENRYQM FN LLNKEFTFDV DVSKVPCGIN GALYFIQMDA DGM SKQPNN RAGAKYGTGY CDSQCPRDIK FIDGVANSAD WTPSETDPNA GRGRYGICCA EMDIWEANSI SNAYTPHPCR TQNDGGYQRC EGRDCNQPRY EGLCDPDGCD YNPFRMGNKD FYGPGKTVDT NRKMTVV TQF ITHDNTDTGT LVDIRRLYVQ DGRVIANPPT NFPGLMPAHD SITEQFCTDQ KNLFGDYSSF ARDGGLAHMG RSLAKGHVLA LSIWN DHGAH MLW LDSNYPT DADPNKPGIA RGTCPTTGGT PRETEQNHPD AQVIFSNIKE GDIGSTFSGY
SEQ ID NO:15	154292161	<i>Botryotinia fuckeliana</i> B05-10	MYSAAVLATF SFLLGAG AQQ VGTSTAETHP ALT VQKCAAG GTCTDES DSI VLDANWRWLH STSGSTNCYT GNTWDTTLC P DAATCTTNCA LGADYEGTY GITTSGDSLK ISFVTGSNVG SRTY LMDSET TYKEFALLGN EFTFTVDVSK LPCGLNGALY FVPMDADGGM SKYPTNKAGA KYGTGYCDAQ CPQDMKFVNG TANVEGWVPD SNSANS GTGN IGSCCSEFDV WEANSM SQAL TPHVCTVDSQ TACTGDDCAS NTGVCDDGDC DFNPYRMGNT TFYGSGMTID TSKPFSVV TQ FITDDGTETG TLTEIKRFYV QDDVVYEQPS SDISGVSGNS ITDDFCAAQK TAFGDTDYFT QNGGMAAMGK KMADGMVLVL SIWDDYNVNM LWLDS DYPTT KDASTPGVSR GSCATDSGVP ATVEAASGSA YVTFESSIKYG PIGSTFNAPA DSSSSVSASS SPAPIASSSS SASIAPVSSV VAAIVSSSAQ AISSAAPVVS SSAQAISAA PVVSSVVSSA APVATSSTKS KCSKVSSTLK TSVAAPATSA TSAAVVATSS AASSTGSVPL YGNCTGGKTC SEGT CVVQND YYSQC VASS
SEQ ID NO:16	169615761 #	<i>Phaeosphaeria nodorum</i> SN15	MTWQRCTGTG GSSCTNVNGE IVIDANRWRI HATGGYT NCF DGNEWNKTAC PSNAACTKNC AIEGSDYRGT YGITTSGNSL TLKFITKGQY STNVGSR TYL MKDTNNYEMF NLIGNEFTFD VDLSQLPCGL NGALYFVSMP EKGQGT PGAK YGTGKLSQCS VHISKTLTDA CARDLK FVGG EANADGWQAS TSDPNAGV GK KGACCAEMDV WEANSMSTAL TPHSCQPEGY AVCEESNCGG TYSIDRYAGT CDANGCDFNP YRVGNKDFY G KGKTVDTSKK MTVV TQFLGT GSDLTELKRF YVQDGKVISN PEPTIPGMTG NSITQKWCDT QKEVFKEEYV PFNQWGGMAS MGKGMAQGMV LVMSLWDDHY SNMLWLDSTY PTD RDPESP G AARGECAITS GAPAEVEANN PDASVMFSNI KFGPIGSTEQ QPA

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:17	4883502	<i>Humicola grisea</i>	MQIKSYIQYL AAALPLLSSV AAQQAQTITA ENHPRMTWKR CSGPGNCQTV QGEVVIDANW RWLHNNQNC YEGNKWTSQC SSATDCAQRC ALDGANYQST YGASTSGDSL TLKFVTKHEY GTNIGSRFYL MANQNKYQMF TLMNNEFAFD VDLKVECGI NSALYFVAME EDGGMASYPs NRAGAKYGTG YCDAQCARDL KFIGGKANIE GWRPSTNDPN AGVGPMGACC AEIDVWESNA YAYAFTPHAC GSKNRYHICE TNNCGGTYSD DRFAGYCDAN GCDYNPYRMG NKDFYGKGKT VDTNRKFTVV SRFERNRLSQ FFVQDGRKIE VPPPTWPGLP NSADITPELC DAQFRVFDDR NRFAETGGFD ALNEALTIPM VLVMSIWDDH HSNMLWLDSS YPPEKAGLPG GDRGPCPTTS GVPAEVEAQY PNAQVVWSNI RFGPIGSTVN V	
SEQ ID NO:18	950686	<i>Humicola grisea</i>	MRTAKFATLA ALVASAAAQQ ACSLTTERHP SLSWKKCTAG GQCQTVQASI TLDSNWRWTH QVSGSTNCYT GNKWDTSICT DAKSCAQNCC VDGADYTSTY GITTINGDLSL LKFVTKGQYS TNVGSRTYLM DGEDKYQTFE LLGNEFTFDV DVSNIGCGLN GALYFVSMDA DGLSRYPGN KAGAKYGTGY CDAQCPRDIK FINGEANIEG WTGSTNDPNA GAGRYGTCCS EMDIWEANNM ATAFTHPCT IIGQSRCEGD SCGGTYSNER YAGVCDPDGC DFNSYRQGNK TFYGKGMTVD TTKKITVVTTQ FLKDANGDLG EIKRFYVQDG KIIPNSESTI PGVEGNSITQ DWCDRQKVAF GDIDDFNRKG GMKQMGKALA GPMVLVMSIW DDHASNMLWL DSTFPVDAAG KPGAERGACP TTSGVPAEVE AEAPNSNVVF SNIRFGPIGS TVAGLPGAGN GGNNGGNPPP PTTTSSAPA TTTTASAGPK AGRWQQCGGI GFTGPTQCEE PYTCTKLNDW YSQCL	
SEQ ID NO:19	124491660	<i>Chaetomium thermophilum</i>	MQIKQYLQYL AAALPLVNMA AAQRAGTQQT ETHPRLSWKR CSSGGNCQTV NAEIVIDANW RWLHDSNYQN CYDGNRWTSA CSSATDCAQK CYLEGANYGS TYGVSTSGDA LTLKFVTKHE YGTNIGSRVY LMNGSDKYQM FTLMNNEFAF DVDLSKVECG LNSALYFVAM EEDGGMRSYS SNKAGAKYGT GYCDAQCARD LKFVGGKANI EGWRPSTNDA NAGVGPIYAC CAEIDVWESN AYAFAFTPHG CLNNNYHVCE TSNCGGTYSE DRFGGLCDAN GCDYNPYRMG NKDFYGKGKT VDTSRKFTVV TRFEENKLTQ FFIQDGRKID IPPPTWPGLP NSSAITPELC TNLSKVFFDR DRYEETGGFR TINEALRIPM VLVMSIWDGH YANMLWLDSV YPPEKAGQPG AERGPCAPTS GVPAEVEAQF PNAQVIWSNI RFGPIGSTYQ V	
SEQ ID NO:20	58045187	<i>Chaetomium thermophilum</i>	MMYKKFAALA ALVAGAAAQQ ACSLTTETHP RLTWKRCTSG GNCSTVNGAV TIDANWRWTH TVSGSTNCYT GNEWDTSICS DGKSCAQTC VDGADYSSTY GITTSGDSL N LKFVTKHHQG TNVGSRVYLM ENDTKYQMFE LLGNEFTFDV DVSNLGCGLN GALYFVSMDA DGM SKYS GN KAGAKYGTGY CDAQCPRDLK FINGEANIEN WTPSTNDANA GFGRYGS CCS EMDIWDANNM ATAFTHPCT IIGQSRCEGN SCGGTYSSE R YAGVCDPDGC DFNAYRQGD K TFYGKGMTVD TTKKMTVVTTQ FHKNSAGVLS EIKRFYVQDG KI IANAESKI PGNPGNSITQ EWCDAQKVAF GDIDDFNRKG GMAQMSKALE GPMVLVMSVW DDHYANMLWL DSTYPIDKAG TPGAERGACP TTSGVPAEIE AQVPNSNVIF SNIRFGPIGS TVPGLDGSTP SNPTATVAPP TSTTTSVRSS TTQISTPTSQ PGGCTTQKWG QCGGIGYTGC TNCVAGTTCT ELPWYSQCL	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:21	169601100 #	<i>Phaeosphaeria nodorum</i> SN15	MYRNFLYAAS LLSVARSQLV GTQTTETHPG MTWQSCTAKG SCTTCSDNKA CASNCAVDGA DYKGTYGITA SGNLQLKFI TKGSYSTNIG SRTYLMASDT AYQMFKFDGN KEFTFDVDLS GLPCGFNGAL YFVSMDEDDGG LKKYSGNKAG AKYGTGYCDA QCPRDLKFIN GEGNVEGWKP SDNDANAGVG GHGSCCAEMD IWEANSISTA VTPHACSTIE QTRCDGDGCG GTYSADRYAG VCDPDGCDEN AYRMGVKNFY KGGMTVDTSK KFTVVVTQFIG TGDAMEIKRF YVQGGKTIEQ PASTIPGVEG NSITTKFCDQ QKQVFGDRYT YKEKGGTANM AKALAQGMVL VMSLWDDHYS NMLWLDSTYP TDKNPDTDLG SGRGSCDVKS GAPADVESKS PDATVIYSNI KFGPLNSTY	
SEQ ID NO:22	169870197	<i>Coprinopsis cinerea okayama</i>	MLGKIAIASL SFLAIAKQQ VGREVAENHP RLPWQRCTRN GGCQTVSNGQ VVL DANWRWL HVT DGYTNCY TGN SWNSSVC SDGTTCAQRC ALEGANYQQT YGITTSGNSL TMKFLTRSQG TNVGGRVYLM ENENRYQM FN LLNKEFTFDV DVSKVPCGIN GALYFIQMDA DGMSSQPNN RAGAKYGTGY CDSQCPRDIK FIDGVANSVG WEPSETDSNA GRGRYGICCA EMDIWEANSI SNAYTPHPCR TQNDGGYQRC EGRDCNQPRY EGLCDPDGCD YNPFRMGNKD FYGPGKTIDT NRKMTVVVTQF ITHDNTDTGT LVDIRRLYVQ DGRVIANPPT NFPGLMPAHD SITEQFCTDQ KNLFGDYSSE ARDGGLAHMG RSLAKGHVLA LSIWNHDGAH MLWLD SNIYPT DADPNKPGIA RGTCPPTTGGT PRETEQNHPD AQVIF SNIKF GDIGSTFSGY	
SEQ ID NO:23	3913806	<i>Agaricus bisporus</i>	MFPRSILLAL SLTAVALGQQ VGTNMAENHP SLTWQRCTSS GCQNVNGKVT LDANWRWTHR INDETN CYTG NEWDTSICPD GVTCAENCAL DGADYAGTYG VTSSGTALT L KFEVTESQQKN IGSRLYLMAD DSNYEIFNLL NKEFTFDVDV SKLPCGLNGA LYFSEMAADG GMSSTNTAGA KYGTGYCDSQ CPRDIKFIDG EANSEGWE GS PNDVNAGTGN FGACCGEMDI WEANSISSAY TPHPCREPGL QRCEGNTCSV NDRYATECDP DGCD FNSFRM GDKSFYGP GM TVDTNQPITV VTQFTDNGS DGNLQEI RR IYVQNGQVIQ NSNVNIPGID SGN SISA EFC DQAKEAFGDE RSFQDRGGLS GMGSALDRGM VLVLSIWDDH AVNMLWLDSD YPLDASPSQP GISRGTC SRD SGKPEDVEAN AGGVQVVYSN IKFGDINSTF NNNGGGGGNP SP TTTTRPN SP AQTMWGQCGG QGWTGPTACQ SPSTCHVIND FYSQCF	
SEQ ID NO:24	169611094	<i>Phaeosphaeria nodorum</i> SN15	MYRN LALASL SLFGAARAQQ AGTVTTETHP SLSWKTC TGT GGTSCTTKAG KITLDANWRW THVTTGYTNC YDGN SWNTTA CPDGATCTKN CAVDGADYSG TYGITTSSNS LSIKFVTKGS NSANIGSRTY LMESDTKYQM FN LIGQEEFF DVDVSKLPCG L NGALYFVEM AADGGIGKN NKAGAKYGTG YCDSQC PHDI K FINGKANVE GWNPSDADPN AGSGKIGACC PEMDIWEANS ISTAYTPHPC KGTGLQECTD DVSCGDGSNR YSGLCDK DGC DFNSYRMGVK DFYGP GATLD TTKKMTVVVTQ FLGSGSTLSE IKRFYVQNGK VFKNSDSAIE GVTGNSITES FCAAQKTA FG DTNSFKTLGG L NEMGASLAR GHVLVMSLWD DHAVNMLWLD STYPTNSTKL GAQRGTCAID SGKPEDVEKN HPDATVVVFS D IKFGPIGSTF QQPS	
SEQ ID NO:25	3131	<i>Phanerochaete chrysosporium</i>	MVDIQIATFL LLGVVGVA AQ QVGTYIPENH PLLATQ SCTA SGGCTTSSSK IVLDANRRWI HSTLGTTSCL TANGWDPTLC PDGITCANYC ALDGVSY SST YGITTSGSAL RLQFVTGTNI GSRVFLMADD THYRTFQLLN QELAFD VDVS KLPCGLNGAL YFVAMDADGG KSKYPGNRAG AKYGTGYCDS QCPRDVQFIN GQANVQGWNA TSATTGTGSY GSCCTELDIW EANSNAAALT PHTCTNNAQT R CSGSNCTSN TGFC DADGCD FNSFRLGNTT FLGAGMSVD T TKFTFVVVTQF ITSDNTSTGN L TEIRRFYVQ NGNVIPNSV V NVTGIGAVNS ITDPFCSQQK KAFIETNYFA QHGG LAQLGQ ALRTGMVLAF SISDDDPANHM L WLD SNEFP PS ANPAVPGVAR GMCSITSGNP ADVGILNPSP YVSFLNIKFG SIGTFRPA	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:26	70991503	<i>Aspergillus fumigatus</i> Af293	MHQRALLFSA LAVAANAQOV GTQTPETHPP LTWQKCTAAG SCSQQSGSVV IDANWRWLHS TKDTTNCYTG NTWNTELCPD NESCAQNICAL DGADYAGTYG VTTSGSELKL SFVTGANVGS RLYLMQDDET YQHFNLLNHE FTFDVDVSNL PCGLNGALYF VAMDADGGMS KYPSNKAAGK YGTGYCDSQC PRDLKFINGM ANVEGWEPSS SDKNAGVGGH GSCCPEMDIW EANSISTAVT PHPCDDVSQT MCSGDACGGT YSESRYAGTC DPDGCCDFNPF RMGNESFYGP GKIVDTKSKM TVVTQFITAD GTDSGALSEI KRLYVQNGKV IANSVSNVAG VSGNSITSDF CTAQKKAFGD EDIFAKHGG LSGMGKALSEM VLIMSIWDDH HSSMMWLDST YPTDADPSKP GVARGTCEHG AGDPENVESQ HPDASVTFSN IKFGPIGSTY EG	
SEQ ID NO:27	294196	<i>Phanerochaete chrysosporium</i>	MERTATLLAF TMAAMVFGQQ VGTNTAENHR TLTSQKCTKS GGCNLTNTKI VLDANWRWLH STSGYTNCYT GNQWDATLCP DGKTCANCA LDGADYTGTY GITASGSSLK LQFVTGSNVG SRVYLMADDT HYQMFQLLNQ EFTFDVDMSN LPCGLNGALY LSAMDADGGM AKYPTNKAGA KYGTGYCDSQ CPRDIKFING EANVEGWNAT SANAGTGNYG TCCTEMDIWE ANNDAAAATP HPCTTNAQTR CSGSDCTRDT GLCDADGDEF NSFRMGDQTF LGKGLTVDTS KPFTVVTQFI TNDGTSAGTL TEIRRLYVQN GKVIQNSSVK IPGIDPVNSI TDNFCSQQKT AFGDTNYFAQ HGGLKQVGEA LRTGMVLALS IWDDYAAANML WLDSNYPTNK DPSTPGVARG TCATTSGVPA QIEAQSPNAY VVFSNIKFGD LNTTYTGTVS SSSVSSSSHSS TSTSSSHSS STPPTQPTGV TVPQWGGQCGG IGYTGSTTCA SPYTCHVLNP YYSQCY	
SEQ ID NO:28	18997123	<i>Thermoplasma acidophilum</i>	MYQRALLFSF FLAARAHEA GTVTAENHPS LTWQQCSSGG SCTTQNGKVV IDANWRWVHT TSGYTNCYTG NTWDTSICPD DVTCAQNICAL DGADYSGTYG VTTSGNALRL NEVTQSSGKN IGSRLYLQD DTTYQIFKLL GQEFFDQDV SNLPCGLNGA LYFVAMDADG NLSKYPGNKA GAKYGTGYCD SQCPRDLKFI NGQANVEGWQ PSANDPNAGV GNHGSSCAEM DVWEANSIST AVTPHPCDTP GQTMCCQDDC GGTYSSTRYA GTCDDPDGPDF NPYQPGNHSF YGPGKIVDTS SKFTVVTQFI TDDGTPSGTL TEIKRFYVQN GKVIPQSEST ISGVTGNSIT TEYCTAQKAA FGDNTGFFTH GGLQKISQAL AQGMVLVMSL WDDHAANMLW LDSTYPTDAD PDTIPGVARGT CPTTSGVPAD VESQNPNSYV IYSNIKVGPI NSTFTAN	
SEQ ID NO:29	4204214	<i>Humicola grisea var thermoides</i>	MQIKSYIQYL AAALPLLSSV AAQAGTITA ENHPRMTWKR CSGPGNCQTV QGEVVIDANW RWLHNNQNC YEGNKWTSQC SSATDCAQRC ALDGANYQST YGASTSGDSL TLKFVTKHEY GTNIGSRFYL MANQNKYQMF TLMNNEFAFD VDLSKVECGI NSALYFVAME EDGGMASYPs NRAGAKYGTG YCDAQCARDL KFIGGKANIE GWRPSTNDPN AGVGPMGACC AEIDVWESNA YAYAFTPHAC GSKNRYHICE TNNCGGTYS D DRFAGYCDAN GCDYNPYRMG NKDFYGGKGT VDTNRKFTVV SRFERNRLSQ FFVQDGRKIE VPPPTWPGLP NSADITPELC DAQFRVFD DR NRFAETGGFD ALNEALTIPM VLVMSIWDDH HSNMLWLDSS YPPEKAGLPG GDRGPCPTTS GVPAEVEAQY PDAQVVWSNI RFGPIGSTVN V	

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:30	34582632	<i>Trichoderma viride</i> (also known as <i>Hypochlrea rufa</i>)	MYRKLAVISA FLATARAQSA CTLQSETHPP LTWQKCSSGG TCTQQTGSVV IDANWRWTHA TNSSTNCYDG NTWSSTLCPD NETCAKNCCCL DGAAYASTYG VTTSGNSLSI GFVTQSAQKN VGARLYLMAS DTTYQEFTLL GNEFSFDVDV SQLPCGLNGA LYFVSMDDADG GVSKEYPTNTA GAKYGTGYCD SQCPRDLKFI NGQANVEGWE PSSNNANTGI GGHGSCCSEM DIWEANSISE ALTPHPCTTV GQEICEGDGC GGTYSNRYG GTCDDPDGCDW DPYRLGNTSF YGPGSSFTLD TTKKLTVVVTQ FETSGAINRY YVQNGVTFQQ PNAELGSYSG NGLNDDYCTA EEAFFGGSSF SDKGGLTQFK KATSGGMVLV MSLWDDYYAN MLWLDSTYPT NETSSTPGAV RGSCSTSSGV PAQVESQSPN AKVTFSNIKE GPIGSTGDPS GGNPPPGGNPP GTTTTTRRPAT TTGSSSPGPTQ SHYGQCGGIG YSGPTVCASG TTCQVLNPYY SQCL
SEQ ID NO:31	156712284	<i>Thermoascus aurantiacus</i>	MYQRALLFSF FLAARAQQA GTVTAENHPS LTWQQCSSGG SCTTQNGKVV IDANWRWVHT TSGYTNCTYG NTWDTSICPD DVTCAQNICAL DGADYSGTYG VTTSGNALRL NEVTQSSGKN IGSRLYLQD DTTYQIFKLL GQEFTFDVDV SNLPCGLNGA LYFVAMDADG GLSKYPGNKA GAKYGTGYCD SQCPRDLKFI NGQANVEGWQ PSANDPNAGV GNHGSCCAEM DVWEANSIST AVTPHPCDTP GQTMCGDDC GGTYSSTRYA GTCDDPDGCDF NPYRQGNHSF YGPGQIVDTS SKFTVVVTQFI TDDGTPSGTL TEIKRFYVQN GKVIPQSEST ISGVTGNSIT TEYCTAQKAA FGDNTGFFTH GGLQKISQAL AQGMVLVMSL WDDHAANMLW LDSTYPTDAD PDTPGVARGT CPTTSGVPAD VESQYPNSYV IYSNIKVGPI NSTFTAN
SEQ ID NO:32	39977899	<i>Magnaporthe grisea</i> (<i>oryzae</i>) 70-15	MIRKITTLLA LVGVVRGQAA CSLTAETHPS LTWQKCSSGG SCTNVAGSVT IDANWRWTHT TSGYTNCTYG NKWDTSICST NADCASKCCV DGANYQQTYG ASTSGNALSL QYVTQSSGKN VGSRLYLLES ENKYQMFNLL GNEFTFDVDA SKLGCGLNGA VYFVSMDDADG GQSKYSGNKA GAKYGTGYCD SQCPRDLKYI NGAANVEGWQ PSSGDANSV GNMGSCCAEM DIWEANSIST AYTPHPCSNN AQHSCCKGDDC GGTYSVRYA GDCDDPDGCDF NSYRQGNRTF YGPGSNFNVD SSKKVTVVVTQ FISSGGQLTD IKRFYVQNGK VIPNSQSTIT GVTGNSVTQD YCDKQKTAFG DQNVFNQRGG LRQMGDALAK GMVLVMSVWD DHHSQMLWLD STYPTTSTAP GAARGSCSTS SGKPSDVQSQ TPGATVVYSN IKFGPIGSTF KSS
SEQ ID NO:33	20986705	<i>Talaromyces emersonii</i>	MLRRALLLSS SAILAVKAQQ AGTATAENHP PLTWQECTAP GSCTTQNGAV VLDANWRWVH DVNGYTNCTY GNTWDPTYCP DDETCQNCA LDGADYEGTY GVTSSGSSLK INFVTGSNVG SRLYLLQDDS TYQIFKLLNR EFSFDVDVSN LPCGLNGALY FVAMDADGGV SKYPNNKAGA KYGTGYCDSQ CPRDLKFIDG EANVEGWQPS SNNANTGIGD HGSCCAEMDV WEANSISNAV TPHPCDTPGQ TMCSGDDCGG TYSNDRYAGT CDDPDGCDENP YRMGNTSFYG PGKIIDTTKP FTVVTQFLTD DGDTGTLSE IKRFYIQNSN VIPQPNSDIS GVTGNSITTE FCTAQKQAFG DTDDFSQHGG LAKMGAAMQQ GMVLVMSLWD DYAAQMLWLD SDYPTDADPT TPGIARGTGP TDSGVVPSDVE SQSPNSYVTV SNIKFGPINS TFTAS

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:34	22138843	<i>Aspergillus oryzae</i>	MHQRALLFSA FWTAVQAQQA GTLTAEHPS LTWQKCAAGG TCTEQKGSVV LDSNWRWLHS VDGSTNCYTG NTWDATLCPD NESCASNCAL DGADYEGTYG VTTSGDALTL QFVTGANIGS RLYLMADDDDE SYQTFNLLNN EFTFDVDASK LPCGLNGAVY FVSMDADGGV AKYSTNKAGA KYGTGYCDSQ CPRDLKFING QVRKGWEPST SDKNAGVGGH GSCCPQMDIW EANSISTAYT PHPCDDTAQT MCEGDTCCGT YSSERYAGTC DPDGCDENAY RMGNESFYGP SKLVDSSSPV TVVTQFITAD GTDSGALSEI KRIFYVQGGKV IANAASNVDG VTGNSITADF CTAQKKAFGD DDIFAQHGGI QGMGNALSSM VLTLSIWDDH HSSMMWLDSS YPEDADATAP GVARGTCEPH AGDPEKVESQ SGSATVTYSN IKYGPIGSTF DAPA	
SEQ ID NO:35	55775695	<i>Penicillium chrysogenum</i>	MASTLSFKIY KNALLLAAFL GAAQAQQVGT STAEVHPSLT WQKCTAGGSC TSQSGKVVID SNWRWVHNTG GYTNCYTGND WDRTLCPDDV TCATNCALDG ADYKGTYGVV ASGSSLRINF VTQASQKNIG SRLYLMADDS KYEMFQLLNQ EFTFDVDVSN LPCGLNGALY FVAMDEDGGM ARYPNTKAGA KYGTGYCDAQ CPRDLKFING QANVEGWEPs SSDVNGGTGN YGSCCAEMDI WEANSISTAF TPHPCDDPAQ TRCTGDSGG TYSSDRYGGT CDPDGCDENP YRMGNQSFYG PSKIVDTESP FTVVTQFITN DGTSTGTLSE IKRFYVQNGK VIPQSVSTIS AVTGNsITDS FCSAQKTAfK DTDVFAKHGG MAGMAGLAE GMVLVMSLWD DHAANMLWLD STYPTSASST TPGAARGSCD ISSGEPsDVE ANHSNAYVVY SNIKVGPLGS TFGSTDsGSg TTTTKVTTTT ATKTTTTTGP STTGAAHYAQ CGGQNWtGPT TCASPYTCQR QGDYYSQCL	
SEQ ID NO:36	171676762	<i>Podospora anserina</i>	MVSAKFAALA ALVASASAAQ VCSLTPESHP PLTWQRCSAG GSCTNVAGSV TLDsNWRWTH TLQGSTNCYS GNEWDTsICT TGTKCAQNCC VEGAEYAATY GITTsGNQLN LKFVTEGKYS TNVGSRTYLM ENATKYQGfN LLGNEFTFDV DVSNIgCGLN GAlYFVSMDL DGGLAKYSgN KAGAKYGTGY CDAQCPRDIK FINGEANIEG WNPSTNDVNA GAGRYGTCCS EMDIWEANNM ATAYTPHsCT ILDQSRCEGE sCGGTYSsDR YGGVCDPDGC DFNSYRMGNK EFYgKGKTVD TTKKMTVVtQ FLKNAAGELS EIKRFYVQNG VVIPNSVSSI PGVPNQNSIT QDWCDAQKIA FGDPDDNTAK GGLRQMGLAL DKPMVLVMSI WNDHAAHMLW LDSTYPVDAA GRPGAERGAC PTTSGVPSEV EAEAPNSNVA FSNIKFGPIG STFNSGSTNP NPISsSTATT PTSTRVSSTS TAAQTPTSAP GGTVPRWGQC GGQYTGPTQ CVAPYTCVVS NQWYSQCL	
SEQ ID NO:37	146350520	<i>Pleurotus sp Florida</i>	MEPYIALVSF SFLSVVLAQQ VGTILTAETHP QLTVQQCTRg GSCTTQQRsV VLDGNWRWLH STSGsNNCYT GNTWDTsLCP DAATCSRNCA LDGADYSGTy GITSSGNALT LKFVTHGPYS TNIGSRVYLL ADDSHYQMfN LKNKEFTFDV DVSQLPCGLN GAlYFSQMDA DGGTGRFPNN KAGAKYGTGY CDSQCPhDIK FINGEANVQG WQPSPNDSNA GKGQYGSCCA EMDIWEANSM ASAYTPHPCT VTPTRCQGN DCGDGDNRyG GVCdKDGCDf NSFRMGDKNF lGPgKTVNTN SKFTVVtQfL TSDNTTSgTL SEIRRLyVQN GRVIONSKVN lPGMASTLDS lTESFCSTQK TVFGDTNSFA SKGGLRAMGN AFDKGMVLVL SIWDDHEAKM lWLDSNYPLD KSASAPGVAR GTCATTSGEP KDVESQSPNA QVIFsNIKYG DIGSTYSN	

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:38	37732123	<i>Gibberella zeae</i>	MYRAIATASA LIAAVRAQOV CSLTQESKPS LNWSKCTSSG CSNVKGSVTI DANWRWTHQV SGSTNCYTGKWDTSVCTSG KVCAERCCLD GADYASTYGI TSSGDQLSL FVTKGPYSTN IGSRTYLMED ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSMADAG GKAKYPGNKA GAKYGTGYCD AQCPRDVKFI NGQANSDGWQ PSDSDVNNGI GNLGTCCPEM DIWEANSIST AYTPHPCTKL TQHSCTGDSG GGTYSNDRYG GTCDDADGCD NSYRQGNKTF YGPGSGFNVD TTKKVTVTQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGVPGNSLTA DFCTKQKKVF NDPDFTKKG AWSGMSDALE APMVLVMSLW HDHHSNMLWL DSTYPTDSTK LGSQRGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YKSDGTTPTN PTNPSEPSNT ANPNPGTVDQ WGQC GGSNYSGPTACKSGFT CKKINDFYSQ CQ
SEQ ID NO:39	156055188	<i>Sclerotinia sclerotiorum</i> 1980	MYSAAVLATF SFLLGAGAQV VGTLKTESHP PLTIQKCAAG GTCTDEADSV VLDANWRWLH STSGSTNCYT GNTWDTTLCP DAATCTANCA FDGADYEGTY GITSSGDSLK LSFVTGSNVG SRTYLMDS ETYKEFALLGN EFTFTVDVSK LPCGLNGALY FVPMDADGGM SKYPTNKAGA KYGTGYCDAQ CPQDMKFVSG GANNEG WVPD SNSANS GTGN IGSCCSEFDV WEANSMSQAL TPHTCTVDGQ TACTGDDCAG NTGVCDDAGC DENPYRMGNT TFYGS GKTID TTKPFVS VVTQ FITDDGTETG TLTEIKRFYV QDDVVEQPN SDISGVSGNS ITDDFCTAQK TAFGDTDYS QKGGMAAMGK KMADGMVLVL SIWDDYNVNM LWLDSDYPTT KDASTPGVSR GSCATTSGVP ATVEAASGSA YVTFSSIKYG PIGSTFKAPA DSSSPV VASS SPAAVA AVVS TSSAQAVPSH PAVSSSQAAV STPEAVSSAP EVPASSAAQ SVAPTSTKPK CSKVSQSSTL ATSVAAPATT ATSAAVAATS AASSSGSVPL YGNCTGGKTC SEGT CVVQNP WYSQC VASS
SEQ ID NO:40	453224	<i>Phanerochaete chrysosporium</i>	MFRAAALLAF TCLAMVSGQQ AGNTAENHP QLQSQCTTS GGCKPLSTKV VLDSNWRWVH STSGYTNCYT GNEWDTSLCP DGKTC AANCA LDGADYSGTY GITSTGTALT LKFVTGSNVG SRVYLMADDDT HYQLLKLLNQ EFTFDVDMSN LPCGLNGALY LSAMDADGGM SKYPGNKAGA KYGTGYCDSQ CPKDIKFING EANVGNWTET GSNTGTGSYG TCCSEMDIWE ANNDAAFTP HPCTTTGQTR CSGDDCARNT GLCDGDG CDF NSFRMGDKTF LGKGMTVDTS KPFTVVTQFL TNDNTSTGTL SEIRRIYIQN GKVIQNSVAN IPGVDPVNSI TDNFC AQQKT AFGDTNWFAQ KGLKQMG EA LGNGMVLALS IWDDH AANML WLDSDYPTDK DPSAPGVARG TCATTSGVPS DVESQVPNSQ VVFSNIKFGD IGSTFSGTSS PNPPGGSTTS SPVTTSP TTP FTGPTVPQWG QCGGIGYSGS TTCASPYTCH VLNPIYSQCY
SEQ ID NO:41	50402144	<i>Trichoderma reesei</i>	MYRKLAVIS A FLATARAQSA CTLQSETHPP LTWQKCSSGG TCTQQTGSVV IDANWRWTHA TNSSTNCYDG NTWSSTLCPD NETCAKNCC L DGAAYASTYG VTTSGNSLSI GFVTQSAQKN VGARLYLMAS DTTYQEFTLL GNEFSFDVDV SQLPCGLNGA LYFVSMADAG GVSKYPTNTA GAKYGTGYCD SQCPRDIKFI NGQANVEGWE PSSNNANTGI GGHGSCCSEM DIWEANSISE ALTPHPCTTV GQEICEGDGC GGTYSNDRYG GTCDDPDGCDW NPYRLGNTSF YGPGSSFTLD TTKKLT VVTQ FETSGAINRY YVQNGVTFQQ PNAELGSYSG NELNDDYCTA EEA EFGGSSF SDKGGLTQFK KATSGGMVLV MSLWDDYYAN MLWLDSTYPT NETSSTPGAV RGSCSTSSGV PAQVESQSPN AKVTFSNIKF GPIGSTGNPS GGNPPGGNRG TTTTTRRPATT TGSSPGPTQS HYGQC GGIGYSGPTVCASGT TCQVLNPYYS QCL

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:42	115397177	<i>Aspergillus terreus</i> NIH2624	MPSTYDIYKK LLLLASFLSA SQAQQVGTSK AEVHPSLTWQ TCTSGGSCTT VNGKVVDAN WRVHNVVDGY NNCYTGNTWD TTLCPDDETC ASNCALEGAD YSGTYGVTTT GNSLRINFVT QASQKNIGSR LYLMEDDSTY KMFKLLNQEF TFDVDVSNLP CGLNGAVYFV SMDADGGMAK YPANKAGAKY GTGYCDSQCP RDLKFINGMA NVEGWEPSAN DANAGTGNHG SCCAEMDIWE ANSISTAYTP HPCDTPGQVM CTGDSCGGTY SSDRYGGTCD PDGCDFNSYR QGNKTFYGP G MTVDTKSKIT VVTQFLTNDG TASGTLSEIK RFYVQNGKVI PNSESTWSGV SGNSTITTAYC NAQKTLFGDT DVFTKHGGME GMGAALAEGM VLVLWLDDH NSNMLWLDN YPTDKPSTTP GVARGSCDIS SGDPKDV EAN DANAYVVYSN IKVGPIGSTF SGSTGGGSSS STTATSKTTT TSATKTTTTT TKTTTTTSAS STSTGGAQHW AQCGGIGWTG PTTCVAPYTC QKQNDYYSQC L	
SEQ ID NO:43	154312003	<i>Botryotinia fuckeliana</i> B05-10	MISKVLAFTS LLAARAQQA GTLTTETHPP LSVSQCTASG CT TSAQSI VV DANRWLHST TGSTNCYTGN TWDKTLCPDG ATCAANCALD GADYSGVYGI TTSGNSIKLN FVTKGANTNV GSRTYLM AAG STTYQMLKL INQEFTFDVD VSNLPCGLNG ALYFAAMDAD GGLSRFP TNK AGAKYGTGYC DAQCPQDIKE INGVANSVGW TPSSNDVNAG AGQYGSCCSE MDIWEANKIS AAYTPHPCSV DTQTRCTGTD CGIGARYSSL CDADGCDFNS YRQGN TSFYG AGLTVNTNKV FTVVTQFITN DGTASGTLKE IRRFYVQNGV VIPNSQSTIA GVPGNSITDS FCAAQKTAFG DTNEFATKGG LATMSKALAK GMVLVMSIWD DHTANMLWLD APYPATKSPS APGVTRGSCS ATSGNPVDVE ANSPGSSVTF SNIKWGPINS TYTGSGAAPS VPGTTTVSSA PASTATSGAG GVAKYAQCGG SGYSGATACV SGSTCVALNP YYSQQ	
SEQ ID NO:44	49333365	<i>Volvariella volvacea</i>	MEPAATLFAF SLFAAVYGQQ VGTQLAETHP RL TWQKCTRS GGCQTQNSGA IVLDANWRWV HNVGGYTNCY TGN TWNTSLC PDGATCAKNC ALDGANYQST YGITTSGNAL TLKFVTQSEQ KNIGSRVYLL ESDTKYQLFN PLNQEFTFDV DVSQ LPCGLN GAVYFSAMDA DGGMSKFPNN AAGAKYGTGY CDSQCPRDIK FINGEANVQG WQPSPNDTNA GTGNYGACCN EMDVWEANSI STAYTPHPCT QQGLVRCSGT ACGGGSNRYG SICDDPDG CDF NSFRMGDKSF YGPGLTVNTQ QKFTVVTQFL TNNSSSGTL REIRRLYVQN GRVIQNSKVN IPGMPSTMDS VTTEFCNAQK TAFNDTFSFQ QKGGMANMSE ALRRGMVLVL SIWDDHAANM LWLDSNYPTD RPASQPGVAR GTCPTSSGKP SDVENSTANS QVIYSNIKFG DIGSTYSA	
SEQ ID NO:45	729650	<i>Penicillium janthinellum</i>	MKGSISYQIY KGALLLSALL NSVSAQQVGT LTAETHPALT WSKCTAGXCS QVSGSVVIDA NWPXVHSTSG STNCYTGNTW DATLCPDDVT CAANCAVDGA RRQHLRVTTT GNSLRINFVT TASQKNIGSR LYLLENDTTY QKENLLNQEF TFDVDVSNLP CGLNGALYFV DMDADGGMAK YPTNKAGAKY GTGYCDSQCP RDLKFINGQA NVDGWTPSKN DVNSGIGNHG SCCAEMDIWE ANSISNAVTP HPCDTPSQTM CTGQRCGGTY STD RYGGTCD PDGCDFNPYR MGV TNFYGP G ETIDTKSPFT VVTQFLTNDG TSTGTLSEIK RFYVQGGKVI GNPQSTIVG SGN SITDSWC NAQKSAFGDT NEFSKHGGMA GMGAGLADGM VLVMSLWDDH ASDMLWLDST YPTNATSTTP GAKRGTCDIS RRPNTVESTY PNAYVIYSNI KTGPLNSTFT GGTTSSSSTT TTTSKSTSTS SSSKTTTTVT TTTTSSGSSG TGARDWAQCG GNGWTGPTTC VSPYTCTKQN DWYSQCL	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:46	146424871	<i>Pleurotus sp Florida</i>	MERTAALTAF TLAAVVLGQQ VGTLTAENHP ALSIQQCTAS GCTTQQKSVV LDSNWRWTHS LPVHTNCYTG NAWDASLCPD PTTCATNCAI DGADYSGTYG ITTSGNALTL RFVTNGPYSK NIGSRVYLLD DADHYKMFDL KNQEFTFDVD MSGLPCGLNG ALYFSEMPAD GGKAAHTSNK AGAKYGTGYC DAQCPHDIKW INGEANILDW SASATDANAG NGRYGACCAE MDIWEANSEA TAYTPHVCRD EGLYRCSGTE CGDGDNRYGG VCDKDGCDFN SYRMGDKNFL GRGKTIDTTK KITVVVTQFIT DDNTSSGNLV EIRRVYVQDG VTYQNSFSTF PLSLSQYNSIS DDFCVAQKTL FGDNQYYNTH GGTEKMGDAM ANGMVLIMSL WSDHAAHMLW LDSDYPLDKS PSEPGVSRGA CATTTGDDPD VVANHPNASV TFSNIKYGPI GSTYGGSTPP VSSGNTSAPP VTSTTSSGPT TPTGPTGTVP KWGQCGGNGY SGPTTCVAGS TCTYSNDWYS QCL	
SEQ ID NO:47	67538012	<i>Aspergillus nidulans</i> FGSC A4	MYQRALLFSA LLSVSRQQQA GTAQEEVHPS LTWQRCEASG SCTEVAGSVV LDSNWRWTHS VDGYTNCYTG NEWDATLCPD NESCAQNCAV DGADYEATYG ITSNGDSLTL KEVTGSNVGS RYVLMEDDET YQMFDDLINNE FTFDVDVSNL PCGLNGALYF TSMDADGGLS KYEGNTAGAK YGTGYCDSQC PRDIKFINGL GNVEGWEPD SDANAGVGGM GTCCPEMDIW EANSISTAYT PHPCDSVEQT MCEGDSGCGT YSDDRYGGTC DFDGCCDFNSY RMGNTSFYGP GAIIDTSSKE TVVTQFIADG GSLSEIKRFY VQNGEVIPNS ESNISGVEGN SITSEFCTAQ KTAFGDEDIF AQHGGLSAMG DAASAMVLIL SIWDDHHSSM MWLDSSTPYTD ADPSQPGVAR GTCEQGAGDP DVVESEHADA SVTFESNIKEG PIGTF	
SEQ ID NO:48	62006162	<i>Fusarium poae</i>	MYRAIATASA LIAAVRAQQV CSLTTETKPA LTWSKCTSSG CSNVQGSVTI DANWRWTHQV SGSTNCHTGN KWDTSVCTSG KVCAEKCCVD GADYASTYGI TSSGNQLSLS FVTKGSYGTN IGSRTYIMED ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSMADAG GKAKYPGNKA GAKYGTGYCD AQCPRDVKFI NGQANSDGWE PSKSDVNGGI GNLGTCCPEM DIWEANSIST AYTPHPCTKL TQHACTGDSC GGTYSNDRYG GTCDADGCDF NAYRQGNKTF YGPGSGFNVD TTKKVTVVTTQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGNPGSSLTS DFCTTQKKVF GDIDDFAKKG AWNGMSDALE APMVLVMSLW HDHHSNMLWL DSTYPTDSTA LGSQRGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YNKEGTQPQP TNPTNPNPTN PTNPGTVDQW GQCGGTNYSG PTACKSPFTC KKINDEYSQC Q	
SEQ ID NO:49	146424873	<i>Pleurotus sp Florida</i>	MERTAALTAF TLAAVVLGQQ VGTLAAENHP ALSIQQCTAS GCTTQQKSVV LDSNWRWTHS TAGATNCYTG NAWDSSLCPN PTTCATNCAI DGADYSGTYG ITTSGNSLTL RFVTNGQYSE NIGSRVYLLD DADHYKLFNL KNQEFTFDVD MSGLPCGLNG ALYFSEMAAD GGKAAHTGNN AGAKYGTGYC DAQCPHDIKW INGEANILDW SGSATDPNAG NGRYGACCAE MDIWEANSEA TAYTPHVCRD EGLYRCSGTE CGDGDNRYGG VCDKDGCDFN SYRMGDKNFL GRGKTIDTTK KITVVVTQFIT DDNTPTGNLV EIRRVYVQDG VTYQNSFSTF PLSLSQYNSIS DDFCVAQKTL FGDNQYYNTH GGTEKMGDSL ANGMVLIMSL WSDHAAHMLW LDSDYPLDKS PSEPGVSRGA CATTTGDDPD VVANHPNASV TFSNIKYGPI GSTYGGSTPP VSSGNTSVPP VTSTTSSGPT TPTGPTGTVP KWGQCGGIGY SGPTSCVAGS TCTYSNEWYS QCL	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:50	295937	<i>Trichoderma viride</i>	MYQKLALISA FLATARAQSA CTLQAEHPP LTWQKCSSGG TCTQQTGSVV IDANWRWTHA TNSSTNCYDG NTWSSTLCPD NETCAKNCCCL DGAAYASTYG VTTSADSLSI GFVTQSAQKN VGARLYLMAS DTTYQEFTLL GNEFSFDVDV SQLPCGLNGA LYFVSMDDADG GVTKYPTNTA GAKYGTGYCD SQCPRDlkFI NGQANVEGWE PSSNNANTGI GGHGSCCSEM DIWEANSISE ALTPHPCTTV GQEICEGDSC GGTYSGDRYG GTCDDPDGCDW NPYRLGNTSF YGPGSSFTLD TTKKLTVVtQ FETSGAINRY YVQNGVTFQQ PNAELGDYSG NSLDDDDYCAA EEAefGGSSF SDKGGLTQFK KATSGGMVLV MSLWDDYYAN MLWLDSTYPT DETSSTPGAV RGSSTSSGV PAQLESNSPN AKVVYSNIKE GPIGSTGNPS GGNPPPGGNPP GTTTPRPATS TGSSSPGPTQT HYGQC GGIGY IGPTVCASGS TCQVLNPYYS QCL	
SEQ ID NO:51	6179889 #	<i>Alternaria alternata</i>	MTWQSCtAKG SCTNKNgKIV IDANWRWlHK KEGYDNCYTg NEWDATAcPD NKACAANCav DGADYSGTYG ITAGNSLKL KfITKGSYST NIGSRtYlMK DDTtYEMfKF TGNQEFtFDV DVSNLPCGFN GALYFVSMdA DGGLKKYSTN KAGAKYGTGY CDAQCPRDlK FINGEGNVEG WKPSSNDANA GVGGHGSCCA EMDIWEANSV STAVtPHSCS TIEQSRCDGD GCGGTYSADR YAGVCDPDGC DFNSYRMGVK DFYgKGKtVD TSKKfTVtTQ fIGtGDAMEI KRfYVQNGKt IAQPASAVPG VEGNSItTKF CDQqKAVFGD TyTFKDKGGM ANMAKALANG MVLVMSlWDD HYSNMLWlDS TyPTDKNPdT DLGTGRGECE TSSGVPADVE SQHADATVVY SNIKFGPLNS TFG	
SEQ ID NO:52	119483864	<i>Neosartorya fischeri</i> NRRL 181	MASAIStQVY RSALILSAFL PSItQAQqIG TyTtEtHPSM TWETcTSgGS CATNQGSVVm DANWRWVHQV GSTtNCYtGN TWdTSICdTD EtCATECAVD GADYEstYGV TtSGSQIRlN FVTQNSNGAN VGSRLYMMAD NTHYQMFkLL NQEFtFDVDV SNlPCGLNGA LYfVtMDEdG GVSKYPNNKA GAQYGVGYCD SQCPRDlKfI QGQANVEGWT PSSNNENTGL GNyGSCCAEL DIWESNSISQ ALtPHPCDtA TNTMCTGDAC GGTYSsDRYA GTCDDPDGCDf NPYRMGNtTF YGPgKtIDtN SPFTVVtQfI TDDGtDtGtL SEIRRYyVQN GVtYAQPDSd ISGITGNAIN ADYcTAENTV FDGPgTfAKH GGfSAmSEAM STGMVLVMSL WDDYYADMLW LDStYPtNAS SSStPGAVRGs cStDSGVPAT IESESPDSyV TySNIKVGPI GSStfSSGSgS GSSGSgSSGS ASStStStKT TAAtStStAV AQHYSQCgGQ DWtGPtTCVS PYtCQVQNAY YSQCL	
SEQ ID NO:53	85083281	<i>Neurospora crassa</i> OR74A	MKAYfEYLVA ALPlLGIATA QQVGKQtTET HPKLSWKKCT GKANCtVNt EVVIDSNWRW LHDSSGKNcY DGNKWTsACS SATDCASKCQ LDGANyGtTY GASTSGDALt lKfVtKHEyG TNIGSRfYlM NGASKYQMfT lMNNEFADFV DLStVECGlN AAlYfVAMEE DGMASySSN KAGAKYGTGY CDAQCARDlK fVGGKANIEG WtPStNDANA GVGPyGGCCA EIDVWESNAH SFAftPHACK TNKYHVcERD NCGGTySEDR FAGLCdANGC DYNPYRMGNt DFYgKGKtVD TSKKfTVVSR FEENKlTQff VQNGQKIEIP GPKWDGIPSD NANItPEfCS AQfQAFGDRD RfAEVGGfAQ lNSALRMPMV lVMSIWDDHY ANMLWlDSVY PPEKEGQPGA ARGDCPQSSG VPAEVESQYA NSKVvYSNIR fGPVGStVNv	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:54	3913803	<i>Cryphonectria parasitica</i>	MESKFALTGS LLAGAVNAQG VGTQQTETHP QMTWQSC TSP SSCTTNQGEV VIDSNWRWVH DKDGYVNCYT GNTWNTTLC PDDKTC AANCV LDGADYSSTY GITTSGNALS LQFVTQSSGK NIGSRTYLME SSTKYHLFDL I GNEFAFDVD LSKLPCGLNG ALYFVTMDAD GGM AKYSTNT AGAEYGTGYC DSQCPRDLKF INGQGNVEGW TPSTNDANAG VGGLGCCSE MDVWEANSMD MAYTPHPCET AAQHSCNADE CGGTYS SRY AGDCD PDGCD WNPFRMGNKD FYGSGDTVDT SQKFTVVTF HSGSSLTEI SQYIQGGTK IQQPNSTWPT LTGYNSITDD FCKAQKVEFN DTDVFSEKGG LAQMGA GMAD GMVLVMSLWD DHYANMLWLD STY PVDADAS SPGKQRGTCA TTSGVPADVE SSDASATVIY SNIKFGPIGA TY	
SEQ ID NO:55	60729633	<i>Corticium rolfsii</i>	MFPAAALLSF TLLAVAS AQQ IGTNTAEVHP SLTVSQCTTS GGCTSS TQSI VLDANWRWLH STSGYTNCYT GNQWNSDLC PDPDTCATNCA LDGASYESTY GISTDGN AVT LNFVTQGSQT NVGSRVYLLS DDTHYQTFSL LNKEFSFDVD ASNIGCGING AVYFVQMDAD GGLSKYSSNK AGAQYGTGYC DSQCPQDIKF INGEANLLDW NATSANS GTG SYGSCCPEMD IWEANKYAAA YTPHPCSVSG QTRCTGTSCG AGSERYDGYC DKDGCDFNSW RMGNETFLGP GMTIDTNKKE TIVTQFITDD NTANGTLSEI RRLYVQGGTV IQNSVANQPN IPKVNSITDS FCTAQKTEFG DQDYFGTIGG LSQMGKAMSD MVLVMSIWDD YDAEMLWLD S NYPTSGSAST PGISRGP CSA TSGLPATVES QQASASVTYS NIKWGDIGST YSGSGSSGSS SSSSSSAASA STSTHTSAAA TATSSAAAAAT GSPVPAYGQC GGQSYTGSTT CASPYVCKVS NAYYSQCLPA	
SEQ ID NO:56	39971383	<i>Magnaporthe grisea</i> 70-15	MKRALCASLS LLA AAVAQQV GTNEPEVHPK MTWKKCSSGG SCSTVNGEV V IDGNWRWIHN IGGYENCYSG NKWTSVCSTN ADCATKCAME GAKYQETYGV STSGDAL TLK FVQQNSSGKN VGS RMYLMNG ANKYQMFTLK NNEFAFDVDL SSVECGMNSA LYFVPMKEDG GMSTEPNKA GAKYGTGYCD AQCARDLKF I GGKGNIEGWQ PSSTDSSAGI GAQGACCAEI DIWESNKNAF AFTPHPCENN EYHVCTE PNC GGTYADDRYG GGCDANGCDY NPYRMGNPDF YGPGKTIDTN RKFTVISRFE NNRNYQILMQ DGV AHRIPGP KFDGLEGETG ELNEQFCTDQ FTVFDERNRF NEVGGW SKLN AAYEIPMVIV MSIWSDHFAN MLWLDSTYPP EKAGQPGSAR GPCPADGGDP NGVVNQYPNA KVIWSNVRF G PIGSTYQVD	
SEQ ID NO:57	39973029	<i>Magnaporthe grisea</i> 70-15	MQLTKAGVFL GALMGGA AQQ QVGTQTAENH PKMTWKKCTG KASCTTVNGE VVIDANWRWL HDASSKNCYD GNRWTDSCRT ASDCAAKCSL EGADYAKTYG ASTSGDALSL KEVTRHDYGT NIGSRFYLMN GASKYQMFSL LGNEFAFDVD LSTIECGLNS ALYFVAMEED GGMKSYSSNK AGAKYGTGYC DAQCARDLKF VGGKANIEGW KPSSNDANAG VGPYGACCAE IDVWESNAHA FAFTPHPCTD NKYHVCQDSN CGGTYSDDR F AGKCDANGCD INPYRLGNTD FYGKGKTVD T SKKFTVVTRF ERDALTQFFV QNNKRIDMPS PALEGLPATG AITAEYCTNV FNVFGDRNRF DEVGGWSQLQ QALSLPMVIV MSIWDDHYSN MLWLD SVYPP DKEGSPGAAR GDCPQDSGVP SEVESQIPGA TVVWSNIRFG PVGSTVNV	

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:58	1170141	<i>Fusarium oxysporum</i>	MYRIVATASA LIAAARAQQV CSLNTETKPA LTWSKCTSSG CSDVKGSVVI DANWRWTHQT SGSTNCYTGN KWDTSICTDG KTCAEKCCLD GADYSGTYGI TSSGNQLSLG FVTNGPYSKN IGSRTYLMEN ENTYQMFQLL GNEFTFDVDV SGIGCGLNGA PHFVSMDEDG GKAKYSGNKA GAKYGTGYCD AQCPRDVKFI NGVANSEGWK PSDSDVNAGV GNLGTCCPEM DIWEANSIST AFTPHPCTKL TQHSCTGDSG GGTYSDDRYG GTCDDADGCD F NAYRQGNKTF YGPGSNFNID TTKKMTVVTTQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGNPGSSLTS DFCSKQKSVF GDIDDFSCKG GWNMGSDALS APMVLVMSLW HDHHSNMLWL DSTYPTDSTK VGSQRGSCAT TSGKPSDLER DVPNSKVSFS NIKFGPIGST YKSDGTTPNP PASSSTTGSS TPTNPPAGSV DQWGQCGGQN YSGPTTCKSP FTCKKINDFY SQCQ
SEQ ID NO:59	121710012	<i>Aspergillus clavatus</i> NRRL 1	MYQRALLFSA LATAVSAQQV GTQKAEVHPA LTWQKCTAAG SCTDQKGSVV IDANWRWLHS TEDTTNCYTG NEWNAELCPD NEACAKNCAL DGADYSGTYG VTADGSSIKL NEVTSANVGS RLYLMEDDET YQMFNLLNNE FTEDVDVSNL PCGLNGALYF VSMDADGGLS KYPGNKAGAK YGTGYCDSQC PRDLKFINGE ANVEGWKPSD NDKNAGVGGY GSCCPEMDIW EANSISTAYT PHPCDGMEOQ RCDGNDCCGT YSSTRYAGTC DFDGCCDFNSF RMGNESFYGP GGLVDTKSPI TVVTQFVTAG GTDSGALKEI RRVYVQGGKV IGNSASNAG VEGDSITSDF CTAQKKAFGD EDIFSKHGGL EGMGKALNKM ALIVSIWDDH ASSMMWLDST YPVDADASTP GVARGTCEHG LGDPETVESQ HPDASVTFSN IKFGPIGSTY KSV
SEQ ID NO:60	17902580	<i>Penicillium funiculosum</i>	MSALNSFNMY KSALILGSLT ATAGAQQIGT YTAETHPSLS WSTCKSGGSC TTNSGAITLD ANWRWVHG VN TSTNCYTGNT WNTAICDTDA SCAQDCALDG ADYSGTYGIT TSGNSLRNLF VTGSNVGSRT YLMADNTHYQ IFDLLNQEFT FTVDVSNLPC GLNGALYFVT MDADGGVSKY PNNKAGAQYG VGYCDSQCPR DLKFLAQQAN VEGWTPSTNN SNTGIGNHGS CCAELDIWEA NSISEALT PH PCDTPGLTVC TADDCGGTYS SNRYAGTCDP DGCDFNPYRL GVTDFYGS GK TVDTTKPFTV VTQFVTDDGT SSGSLSEIRR YYVQNGVVIP QPSSKISGIS GNVINSDFCA AELSAFGETA SFTNHGGLKN MGSALEAGMV LVMSLWDDYS VNMLWLDSTY PANETGTPGA ARGSCPTTSG NPKTVESQSG SSYVVFSDIK VGFENSTFSG GTSTGGSTTT TASGTTSTKA STTSTSSSTST GTGVAAHWGQ CGGQGWGTGPT TCASGTTCTV VNPYYSQCL
SEQ ID NO:61	1346226	<i>Humicola grisea</i> var <i>thermoidea</i>	MRTAKFATLA ALVASAAAQQ ACSLTTERHP SLSWNKCTAG GQCQTVQASI TLDSNWRWTH QVSGSTNCYT GNKWDTSICT DAKSCAQNCC VDGADYTSTY GITTNGDSLS LKFVTKGQHS TNVGSRTYLM DGEDKYQTFE LLGNEFTFDV DVSNIGCGLN GALYFVSM DA DGGLSRYPGN KAGAKYGTGY CDAQCPRDIK FINGEANIEG WTGSTNDPNA GAGRYGTCCS EMDIWEANNM ATAFTHPCT IIGQSRCEGD SCGGTYSNER YAGVCDPDGC DENSYRQGNK TFYKGGMTVD TTKKITVVTTQ FLKDANGDLG EIKRFYVQDG KIIPNSESTI PGVEGNSITQ DWCDRQKVA F GDIDDENRKG G MKQMGKALA GPMVLVMSIW DDHASNMLWL DSTFPVDAAG KPGAERGACP TTSGVP AEVE AEAPNSNVVF SNIRFGPIGS TVAGLP GAGN GGNNGGNPPP FTTTSSAPA TTTTASAGPK AGRWQCGGI GFTGPTQCEE PYICTKLN DW YSQCL

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:62	156712282	<i>Chaetomium thermophilum</i>	MMYKKFAALA ALVAGASAAQ ACSLTAENHP SLTWKRCTSG GSCSTVNGAV TIDANWRWTH TVSGSTNCYT GNQWDTSLCT DGKSCAQTC VDGADYSSTY GITTSGDSL N LKFVTKHQYG TNVGSRVYLM ENDTKYQMFE LLGNEFTFDV DVSNLGCGLN GALYFVSM DA DGGMSKYS GN KAGAKYGTGY CDAQCP RDLK FINGEANVGN WTPSTNDANA GFGRYGSCCS EMDVWEANNM ATAFTHPCT TVGQSRCEAD TCGGTYSDDR YAGVCDPDGC DENAYRQGD K TFYKGKMTVD TNKKMTVV TQ FHKNSAGVLS EIKRFYVQDG KI IANAESKI PGNPGNSITQ EYCDAAQKVAF SNTDDFN RKG GMAQMSKALA GPMVLVMSVW DDHYANMLWL DSTYPIDQAG APGAERGACP TTSGVPAEIE AQVPNSNVIF SNIRFGPIGS TVPGLDGSNP GNPTTTVVPP ASTSTSRPTS STSSPVSTPT GQPGGCTTQK WQCGGIGYT GCTNCVAGTT CTQLNPWYSQ CL	
SEQ ID NO:63	169768818	<i>Aspergillus oryzae</i> RIB40	MASLSLKIC RNALILSSVL STAQGQQVGT YQTETHPSMT WQTCGNGGSC STNQGSVVLD ANWRWVHQTG SSSNCYTGNK WDTSYCSTND ACAQKCALDG ADYSNTYGIT TSGSEVRLNF VTSNSNGKNV GSRVYMMADD THYEVYKLLN QEFTFDVDVS KLPCGLNGAL YFVVMADAGG VSKYPNNKAG AKYGTGYCDS QCPRDLKFIQ GQANVEGWVS STNNANTGTG NHGSCCAELD IWESNISQA LTPHPCDTPT NTLCTGDACG GTYSSDRYSG TCDPDGCDFN PYRVGN TTFY GPGKTIDTNK PITVVVTQFIT DDGTSSG TLS EIKRFYVQDG VTYPQPSADV SGLSGNTINS EYCTAENTLF EGSGSFAKHG GLAGMGEAMS TGMVLVMSLW DDYANMLWL DSNYPTNEST SKPGVARGTC STSSGV PSEV EASNPSAYVA YSNIKVGPIG STFKS	
SEQ ID NO:64	46241270	<i>Gibberella pulicaris</i>	MYRAIATASA LIAAVRAQQV CSLTPETKPA LSWSKCTSSG CSNVQGSVTI DANWRWTHQL SGSTNCYTGN KWDTSICTSG KVCAEKCCID GA EYASTYGI TSSGNQLSLS FVTKGAYGTN IGSRTYIMED ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSM DADG GKAKYPGNKA GAKYGTGYCD AQCP RDKFI NGQANS D GWQ PSKSDVNAGI GNMGTCCPEM DIWEANSIST AYTPHPCTKL TQHSCTGDSC GGTYSNDRYG GTC DADG CDF NAYRQGNKTF YGPGSGFNVD TTKKVTVV TQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGVPGSSLTP EFCTAQKKVF GDTDDFAKKG AWSGMSDALE APMVLVMSLW HDHHSNMLWL DSTYPTDSTK LGAQRGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YKEGVPEPTN PTNPTNPTNP TNPGTV DQWA QCGGTNYSGP TACKSPFTCK KINDEFYSQQ	
SEQ ID NO:65	49333363	<i>Volvariella volvacea</i>	MFPKSSLLVL SFLATAYAQQ VGTQTAEVHP SLNWARCTSS GCTNVAGSVT LDANWRWLHT TSGYTNCYTG NSWNTTLC PD GATCAQN CAL DGANYQSTCG ITTSGNALT L KFTVQGEQKN IGSRVYLMA S ESRYEMFGLL NKEFTFDVDV SNLPCGLNGA LYFSSMDADG GMAKNPGNKA GAKYGTGYCD SQCP RDIKFI NGEANVAGWN GSPNDTNAGT GNWGACCNEM DIWEANSISA AYTPHPCTVQ GLSRCSGTAC GTNDRYGTVC DPGCDFNSY RMGDKTY YGP GGTGVDTRSK FTVVTQFLT N NNSSSGT LSE IRRLYVQNGR VVQNSKVNIP GMSNTLDSIT TGFCDSQKTA FGDTRSFQNK GGMSAMQAL GAGMVLVLSV WDDHAANMLW LDSNYPVDAD PSKPGIARGT CSTTSGKPTD VEQSAANSSV TFSNIKFGDI GTTYTGGSVT TTPGNPGTTT STAPGAVQTK WQCGGQGW T GPTRCESGST CTVVNQWYSQ CI	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:66	46395332	<i>Irpex lacteus</i>	MERKAALLAF SFLAIAHQQQ VGTNQAENHP SLPSQHCTAS GCTTSSTSVV LDANWRWVHT TTGYTNCYTG QTWDAASICPD GVTCAKACAL DGADYSGTYG ITTSGNALTL QFVKGNTVGS RVYLLQDASN YQLFKLINQE FTFDVDMSNL PCGLNGAVYL SQMDQDGGVS RFPTNTAGAK YGTGYCDSQC PRDIKFINGE ANVAGWTGSS SDPNSGTGNY GTCCSEMDIW EANSVAAAYT PHPCSVNQQT RCTGADCGQD ANRYKGVCDP DGCDFNSFRM GDQTFLGKGL TVDTSRKFTI VTQFISDDGT SSGNLAEIRR FYVQDGKVIP NSKVNIAGCD AVNSITDKFC TQOKTAFGDT NRFADQGGLK QMGAALKSGM VLALSLLWDDH AANMLWLDS YPTTADASKP GVARGTCPNT SGVPKDVESQ SGSATVTYSN IKWGDLNSTF SGTASNPTGP SSSPSGPPSS SSSTAGSQPT QPSSGSVAQW GQCGGIGYSG ATGCVSPYTC HVVNPYYSQC Y	
SEQ ID NO:67	50844407 #	<i>Chaetomium thermophilum</i> var <i>thermophilum</i>	TETHPRLTWK RCTSGGNCST VNGAVTIDAN WRWHTTVSGS TNCYTGNEWD TSICSDGKSC AQTCCVDGAD YSSTYGITTS GDSLNLFVTT KHQHGNTVGS RVYIMENDTK YQMFELLGNE FTFDVDVSNL GCGLNGALYF VSMDADGGMS KYSGNKAGAK YGTGYCDAQC PRDLKFINGE ANIENWTPST NDANAGFGRY GSCCSEMDIW EANNMATAFT PHPCTIIGQS RCEGNSCGGT YSSERYAGVC DPDGCDFNAY RQDKTFYGK GMTVDTTKKM TVVTQFHKNS AGVLSEIKRF YVQDGKIIAN AESKIPGNPG NSITQEWCD A QKVAFGDIDD FNRKGGMAQM SKALEGPMVL VMSVWDDHYA NMLWLDSTYP IDKAGTPGAE RGACPTTSGV PAEIEAQVPN SNVIFSNIRF GPIGSTVPGL DGSTPSNPTA TVAPPTSTTT SVRSSTTQIS TPTSQPGGCT TQKWGQCGGI GYTGCTNCVA GTTCTELNPW YSQCL	
SEQ ID NO:68	4586347	<i>Irpex lacteus</i>	MEHKAVLVAF SLVTIVHGQQ AGTQTAENHP QLSSQKCTAG GSCTSASTSV VLDSNWRWVH TTSGYTNCYT GNTWDASICS DPVSCAQNCA LDGADYAGTY GITTSGDALT LKFVTGSNVG SRVYLMEDET NYQMFKLMNQ EFTFDVDVSN LPCGLNGAVY FVQMDQDGGT SKFPNNKAGA KFGTGYCDSQ CPQDIKFING EANIVDWTAS AGDANSGTGS FGTCCEMDI WEANSISAAY TPHPCTVTEQ TRCSGSDCGQ GSDRFNGICD PDGCDFNSFR MGNTEFFYGK LTVDTSQKFT IVTQFISDDG TADGNLAEIR RFYVQNGKVI PNSVVQITGI DPVNSITEDF CTQOKTVFGD TNNFAAKGGL KQMG EAVKNG MVLALSLLWDD DYPTTADPSQ PGVARGTCPT TSGVPSQVEG QEGSSSVIYS NIKFGDLNST FTGTLTNPSS PAGPPVTSSP SEPSQSTQPS QPAQPTQPAG TAAQWAQCGG MGFTGPTVCA SPFTCHVLNP YYSQCY	
SEQ ID NO:69	3980202	<i>Phanerochaete chrysosporium</i>	MFRAAALLAF TCLAMVSGQQ AGNTAENHP QLQSQQCTTS GGCKPLSTKV VLDSNWRWVH STSGYTNCYT GNEWNTSLCP DGKTC AANCA LDGADYSGTY GITSTGTALT LKFVTGSNVG SRVYLMADDT HYQLLKLINQ EFTFDVDMSN LPCGLNGALY LSAMDADGGM SKYPGNKAGA KYGTGYCDSQ CPKDIKFING EANVGNWTET GSNTGTGSYG TCCSEMDIWE ANNDAAAFTP HPCTTTGQTR CSGDDCARNT GLCDHGDGCD FNSFRMGDKT FLGKGMTVDT SKPFTDVTQF LTNDNTSTGT LSEIRRIYIQ NGKVIQNSVA NIPGVDPVNS ITDNFCAQQK TAFGDTNWFA QKGGLKQMG E ALNGMVLAL SIWDDH AANM LWLSDDYPTD KDPSAPGVAR GTCATTSGVP SDVESQVPNS QVVF SNIKFG DIGSTFSGTS SPNPPGGSTT SSPVTTSPTP PPTGPTVPQW GQCGGIGYSG STTCASPYTC HVLNPYYSQC Y	

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:70	27125837	<i>Melanocarpus albomyces</i>	MMMKQYLQYL AAALPLVGLA AGQRAGNETP ENHPPLTWQR CTAPGNCQTV NAEVVIDANW RWLHDDNMQN CYDGNQWTNA CSTATDCAEK CMIEGAGDYL GTYGASTSGD ALTLKFVTKH EYGTNVGSRF YLMNGPDKYQ MFNLMGNELA FDVDLSTVEC GINSALYFVA MEEDGGMASY PSNQAGARYG TGYCDAQCAR DLKFFVGGKAN IEGWKSSTSD PNAGVGPYGS CCAEIDVWES NAYAFAFTPH ACTTNEYHVC ETTNCGGTYS EDRFAGKCD A NGCDYNPYRM GNPDFYGGK TLDTSRKFTV VSRFEENKLS QYFIQDGRKI EIPPTWEGM PNSSEITPEL CSTMFDVFN D RNRFEEVGGE EQLNNALRVP MVLVMSIWDD HYANMLWLD IYPEKEGQP GAARGDCPTD SGVPAEVEAQ FPD AQVVWSN IRFGPIGSTY DF
SEQ ID NO:71	171696102	<i>Podospora anserina</i>	MYRSATFLTF ASLVLGQQVG TYTAERHPSM PIQVCTAPGQ CTRESTEVVL DANWRWTHIT NGYTNCYTGN EWNATACPDG ATCAKNCAVD GADYSGTYGI TTPSSGALRL QFVKKNDNGQ NVGSRVYIMA SSDKYKLFNL LNKEFTFDVD VSKLPCGLNG AVYFSEMLED GGLKSFSGNK AGAKYGTGYC DSQCPQDIKF INGEANVEGW GGADGNSGTG KYGICCAEMD IWEANSDATA YTPHVCVNE QTRCEGVDCG AGSDRYNSIC DKDGCDFENSY RLGNREFYGP GKTVDTTTRPF TIVTQFVTDD GTDSGNLKSI HRYVYQDGNV IPNSVTEVAG VDQTNFISEG FCEQQKSAFG DNNYFGQLGG MRAMGESLKK MVLVLSIWDD HAVNMNWLD IFFNDADPEQ PGVARGRCDDP ADGVPATIEA AHPDAYVIYS NIKFGAINST FTAN
SEQ ID NO:72	3913802	<i>Cochliobolus carbonum</i>	MYRTLAFASL SLYGAARAQQ VGTSTAENHP KLTWQTCGTG GTTNC SNKSG SVVLDSNWRW AHNVGGYTNC YTGNSWSTQY CPDGD SCKTN CAIDGADYSG TYGITTSNNA LSLKFVTKGS FSSNIGSRTY LMETDTKYQM FNLINKEFTF DVDVSKLPCG LNgALYFVEM AADGGIGKN NKAGAKYGTG YCDSQC PHDI KFINGKANVE GWNPSDADPN GGAGKIGACC PEMDIWEANS ISTAYTPHPC RGVGLQECSD AASC GDG SNR YDGQCDK DGC DENSYRMGVK DFYGP GATLD TTKKMTVITQ FLGSGSSLSE IKRFYVQNGK VYKNSQSAVA GVTGNSITES FCTAQKKAFG DTSSFAALGG L NEMGASLAR GHVLIMSLWG DHAVNMMLWLD STYPTDADPS KPGAARGTCP TTSGKPEDVE KNSPDATVVF SNIKFGPIGS TFAQPA
SEQ ID NO:73	50403723	<i>Trichoderma viride</i>	MYQKLALISA FLATARAQSA CTIQAETHPP LTWQKCSSGG TCTQQTGSVV IDANWRWTHA TNSSTNCYDG NTWSSTLCPD NETCAKNCC L DGAAYASTYG VTT SADSLSI GFVTQSAQKN VGARLYIMAS DTTYQEFTLL GNEFSFDVDV SQLPCGLNGA LYFVSM DADG GVS KYPTNTA GAKYGTGYCD SQ CPRDLKFI NGQANVEGWE PSSNNANTGI GGHGSCCSEM DIWEANSISE ALTPHPCTTV GQEICDGDSC GGTYS GDRYG GTCDDPDGCDW NPYRLGNTSF YGPGSSFTLD TTKKLTVV TQ FETSGAINRY YVQNGVTFQQ PNAELGDYSG NSLDDDYCAA EEAEFGGSSF SDKGGLTQFK KATSGGMVLV MSLWDDYYAN MLWLDSTYPT NETSSTPGAV RGSCSTSSGV PAQLESNSPN AKVVYSNIKF GPIGSTGNSS GGNPPGGNPP GTTTTTRRPAT STGSSSPGPTQ THYGQCGGIG YSGPTVCASG STCQVLNPHY SQCL

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:74	3913798	<i>Aspergillus aculeatus</i>	MVDSFSIYKT ALLLSMLATS NAQQVGTYTA ETHPSLTWQT CSGSGSCTTT SGSVVIDANW RWVHEVGGYT NCYSGNTWDS SICSTDTTCA SECALEGATY ESTYGVTTSG SSLRLNFVTT ASQKNIGSRL YLLADDSTYE TKLLENREFT FDVDVSNLPC GLNGALYFVS MDADGGVSRF PTNKAGAKYG TGYCDSQCPR DLKFIDGQAN IEGWEPSTD VNAGTGNHGS CCPEDIWEA NSISSAFTAH PCDSVQQTMC TGDTCGGTYS DTTDRYSGTC DPDGCDENPY RFGNTNFYGP GKTVDNSKPF TVVTQFITHD GTDTGTLTEI RRLYVQNGVV IGNGPSTYTA ASGNSITESF CKAECTLFGD TNVFETHGGL SAMGDALGDG MVLVLSLWDD HAADMWLWDS DYPTTSCASS PGVARGTCPT TTGNATYVEA NYPNSYVTYS NIKFGTLNST YSGTSSGGSS SSSTTLLTTKA STSTTSKTT TTTSKTSTTS SSSTNVAQLY GQCGGQGW TG PTTCASGTCTKQNDYYSQCL
SEQ ID NO:75	66828465	<i>Dictyostelium discoideum</i>	MYRILKSFIL LSLVNMSLSQ KIGKLTPEVH PPMTFQKCE GGSCETIQGE VVVDANWRWV HSAQGQNCYT GNTWNPTICP DDETCACENCY LDGANYESVY GVTTSEDSVR INFVTQSQ GK NIGSRLFLMS NESNYQLFHV LGQEFTFDVD VSNLDCGLNG ALYLVSMDS D GGSARFPTNE AGAKYGTGYC DAQCPRDLKF ISGSANVDGW IPSTNNPNTG YGNLGSCCAE MDLWEANNMA TAVTPHPCDT SSQSVCCKSDS CGGAASNRY GGICDPDGCD YNPYRMGNTS FFGPNKMIDT NSVITVVTFQF ITDDGSSDGK LTSIKRLYVQ DGNVISQSVS TIDGVEGNEV NEEFCTNQKK VFGDEDSFTK HGGLAKMGEA LKDGMLVLVS LWDDYQANML WLDSSYPPTS SPTDPGVARG SCPTTSGVPS KVEQNPYNAY VVYSNIKVGP IDSTYKK
SEQ ID NO:76	156060391	<i>Sclerotinia sclerotiorum</i> 1980	MISRVLAISS LLAARAQQI GTNTAEVHPA LTSIVIDANW RWLHTTSGYT NCYTGNSWDA TLCPDAVTCA ANCALDGADY SGTYGITTSG NSLKLNFVTK GANTNVGSR T YLMAAGSKTQ YQLLKLIGQE FTFDVDVSNL PCGLNGALYF AEMDADGGVS RFPTNKAGAAQ YGTGYCDAQC PQDIKFINQ ANSVGWT P SS NDVNTGTGQY GSCCSEMDIW EANKISAA Y T PHPCSVDGQT RCTGTDCGIG ARYSSLCDAD GCDFNSYRMG DTGFYAGALT VDTSKVFTVV TQFITNDGTT SGTLS EIRRF YVQNGKVIPN SQSKVTGVSG NSITDSFCAA QKTAFGDTNE FATKGGLATM SKALAKGMVL VMSIWDDHSA NMLWLDAPY ASKSPSAAGV SRGSCSASSG VPADVEANSP GASVTYSNIK WGPINSTYSA GTGSNTGSGS GSTTTLVSSV PSSTPTSTTG VPKYGCQCGS GYTGPTNCIG STCVSMGQYY SQCQ
SEQ ID NO:77	116181754	<i>Chaetomium globosum</i> CBS 148-51	MYRQVATALS FASLVLGQQV GTLTAETHPS LPIEVCTAPG SCTKEDTTVV LDANWRWTHV TDGYTNCYTG NAWNETACPD GKTCAANCAI DGAEYEKTYG ITTPEEGALR LNFVTESNVG SRVYLMAGED KYRLFNLLNK EFTMDVDVSN LPCGLNGAVY FSEMEDDGGM SRFEGNKAGA KYGTGYCDSQ CPRDIKFING EANSEGWGE DGNSGTGKYG TCCAEMDIWE ANLDATA Y T P HPCKVTEQTR CEDDTECGAG DARYEGLCDR DGCDFNSFRL GNKEFYGPEK TVDTSKPFTL VTQFVTA DGT DTGALQSIRR FVVDGTVIP NSETVVEGVD PTNEITDDFC AQQKTAFGDN NHFKTIGGLP AMGKSLEKMV LVLSIWDDHA VYMNWLD S NY PTDADPTKPG VARGRCDPEA GVPETVEAAH PDAYVIYSNI KIGALNSTFA AA

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:78	145230535	<i>Aspergillus niger</i>	MSSFQVYRAA LLLSILATAN AQQVGTYTTE THPSLTWQTC TSDGSCTTND GEVIDANWR WVHSTSSATN CYTGNEWDTS ICTDDVTCAA NCALDGATYE ATYGVTTSGS ELRLNFVTQG SSKNIGSRLY LMSDDSNYEL FKL LGQEFTF DVDVSNLPCG LNgALYFVAM DADGGTSEYS GNKAGAKYGT GYCDSQCPRD LKFINGEANC DGWEPSSNNV NTGVGDHGSC CAEMDVWEAN SISNAFTAHP CDSVSQTMCD GDSCGGTYSA SGDRYSGTCD PDGCDYNPYR LGNTDFYGPg LTVDTNSPFT VVTQFITDDG TSSGTLTEIK RLYVQNGEVI ANGASTYSSV NGSSITSAFC ESEKTLFGDE NVFDKHGGLE GMGEAMAKGM VLVLslWDDY AADMLWLDSD YPVNSSASTP GVARGTCSTD SGVPATVEAE SPNAYVTYSN IKFGPIGSTY SSGSSSGSGS SSSSSSTTTK ATSTTLKTTs TTSSGSSSSTS AAQAYGQCGG QGWTGPTTCV SGYTCTYENA YYSQCL
SEQ ID NO:79	46241266	<i>Nectria haematococca</i> mpVI	MYRAIATASA LLATARAQQV CTlNTENKPA LTWAKCTSSG CSNVRGSVVV DANWRWAHST SSSTNCYTGN TWDKTlCPDG KTCADKCCLD GADYSGTYGV TSSGNQLNLK FVTVGPySTN VGSRLYIMED ENNYQMFDLL GNEFTFDVDV NNIGCGLNGA LYFVSMDDKG GKSrFSTNKA GAKYGTGYCD AQCPRDVKFI NGVANSDEWK PSDSDKNAGV GKYGTCCPEM DIWEANKIST AYTPHPCKSL TQOCEGDAC GGTYSATRYA GTCDPDGCDf NPYRQGNKTF YGPGSGFNVD TTKKVTVVTQ FIKGSDGKLS EIKRLYVQNG KVIgNPQSEI ANNPgSSVTD SFCKAQKVAF NDPDDFNKKG GWSGMSDALA KPMVLVMSLW HDHYANMLWL DSTYPKGSKT PGSARGSCPE DSGDPDTLEK EVPNsgVSFS NIKFGPIGST YTGtGGSNPd PEEPEEPEEP VGTVPYGQC GGINYSGPtA CVSPYKCNKI NDFYSQCQ
SEQ ID NO:80	1q9h (PDB) #	<i>Talaromyces emersonii</i>	EQAGTATAEN HPPLTWQECT APGSCTTQNG AVVLDANWRW VHDVNGYTNC YTGNTWDPTY CPDDEtCAQN CALDGADYEG TYGVTSSGSS LKlNFVTGSN VGSRLYLLQD DSTYQIFKLL NREFSFDVDV SNLPCGLNGA LYFVAMDADG GVSKYPNNKA GAKYGTGYCD SQCPRDlKFI DGEANVEGWQ PSSNNANTGI GDHGSCCAEM DVWEANSISN AVTPHPCDTP GQTMCSGDDC GGTYSNDRYA GTCDPDGCDf NPYRMGNTSF YGPGKIIDTT KPFTVVtQFL TDDGTDtGTL SEIKRFYIQN SNVIPQPNSD IsgVTGNSIT TEfCTAQKQA FGDTDDfSQH GGLAKMGAAm QQGMVLVMSL WDDYAAQMLW LDSDYPTDAD PTTPGIARGT CPTDSGVPSD VESQSPNSYV TYSNIKFGPI NSTFTAS
SEQ ID NO:81	157362170	<i>Polyporus arcularius</i>	MEPTLAlVSL SFLAIAYGQQ VGtLTAETHP KLSVSQCTAG GSCTTVQRSV VLDSNRWLH DVGGSTNCYT GNTWDDSLCP DPTTCAANCA LDGADYSGTY GITTSGNALS LKFVTQGPYS TNIGSRVYLL SEDDSTYEMF NLKNQEFTFD VDMSALPCGL NGALYFVEMD KDGGSGRFPT NKAGSKYGTG YCDTQCPhDI KfINGEANVL DWAGSSNDPN AGTGHYGTCC NEMDIWEANS MGAAVTPHVC TVQgQTRCEG TDCGDGDERY DGICDKDGCd FNSWRMGDQt FLGPgKTVDT SSKFTVVtQF ItADNTTSGD lSEIRRLYVQ NGKVIANSKT QIAGMDAYDS ITDDFCNAQK TTFGDtNTFE QMGGLATMGD AfETGMVLVM SIWDDHEAKM lWLDSDYPTD ADASAPGVSR GPCPTTSgDP TDVESQSPGA TVIFsNIKtG PIGSTFTS

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:82	7804885	<i>Leptosphaeria maculans</i>	MLSASKAAAI LAFCAHTASA WVVGDDQQTET HPKLNWQRCT GKGRSSCTNV NGEVVIDANW RWLAHRSGYT NCYTGSEWNQ SACPNNEACT KNCAIEGSDY AGTYGITTSG NQMNIKFITK RPYSTNIGAR TYLMKDEQNY EMFQLIGNEF TFDVDLSQRC GMNGALYFVS MPQKGQGAPG AKYGTGYCDA QCARDLKFFVR GSANAEGWTK SASDPNSGVG KKGACCAQMD VWEANSAATA LTPHSCQPAG YSVCEDTNCG GTYSEDRIYAG TCDANGCDFN PFRVGVKDFY GKGTVDVTTK KMTVVTFQVG SGNQLSEIKR FYVQDGKVIA NPEPTIPGME WCNTQKKVFQ EEAYPFNEFG GMASMSEGMS QGMVLVMSLW DDHYANMLWL DSNWPREADP AKPGVARRDC PTSGGKPSKV EAAANPNAQVM FSNIKFGPIG STFAHAA	
SEQ ID NO:83	121852	<i>Phanerochaete chrysosporium</i>	MERTATLLAF TMAAMVFGQQ VGTNTARSHP ALTSQKCTKS GGCSNLNTKI VLDANWRWLH STSGYTNCYT GNQWDATLCP DGKTCAANCA LDGADYTGTY GITASGSSLK LQFVTGSNVG SRVYLMADDT HYQMFQLLNQ EFTFDVDMSN IPCGLNGALY LSAMDADGGM AKYPTNKAGA KYGTGYCDSQ CPRDIKFIN G EANVEGWNAT SANAGTGNYG TCCTEMDIWE ANNDAAAATP HPCCTTNAQTR CSGSDCTRDT GLCDADG CDF NSFRMGDQTF LGKGLTV DTS KPFTVVTFQFI TNDGTSAGTL TEIRRLYVQN GKVIQNSSVK IPGIDPVNSI TDNFC SQQKT AFGDTNYFAQ HGGLKQVGEA LRTGMVLALS IWDDYAAANML WLDSNYPTNK DPSTPGVARG TCATTSGVPA QIEAQSPNAY VVFSNIKF GD LNTTYTGTVS SSSVSSSSHSS TSTSSSHSS STPPTQPTGV TVPQWGGQCGG IGYTGSTTCA SPYTCHVLNP YYSQCY	
SEQ ID NO:84	126013214	<i>Penicillium decumbens</i>	MYQRALLFSA IMAGVSAQQV GTQKPETHPP LAWKECTSSG CTSKDGSVVI DANWRWVHSV DGYKNCYTGN EWDSTLCPDD ATCATNCAVD GADYAGTYGA TTEGDSLSIN FVTGSNIGSR FYLMEDENKY QMFKLLNKEF TFDVDVSTLP CGLNGALYFV SMDADGGMSK YETNKAGAKY GTGYCDSQCP RDLKFINGKG NVEGWKPSAN DKNAGVGP HG SCCAEMDIWE ANSISTALT P HPCDTNGQTI CEGDSCGGTY STTRYAGTCD PDGCD FNPFR MGNESFYGPG KMVDTKSKMT VVTQFITSDG TDTGSLKEIK RVYVQNGKVI ANSASDVSGI TGNSITSDFC TAQKKTFGDE DVFNKHGGLS GMGDALGEGM VLVMSLWDDH NSNMLWLDGE KYPTDAAASK AGVSRGTCST DSGKPSTVES ESGSAKV VVS NIKVGSIGST FSA	
SEQ ID NO:85	156048578	<i>Sclerotinia sclerotiorum</i> 1980	MTSKIALASL FAAA YGQQIG TYTTETHPSL TWQSC TAKGS CTTQSGSIVL DGNWRWTHST TSSTNCYTGN TWDATLCPDD ATCAQNCALD GADYSGTYGI TTSGDSLRLN FVTQTANKNV GSRVYLLADN THYKTFNLLN QEFTFDVDVS NLP CGLNGAV YFANLPADGG ISSTNKAGAQ YGTGYCDSQC PRDGKFINGK ANVDGWVPSS NNPNTGVGNY GSCCAEMDIW EANSISTAVT PHSCDTVTQT VCTGDNCGGT YSTTRYAGTC DPDGCDFNPY RQGNESFYGP GKTVDTNSVF TIVTQFLTTD GTSSGTLNEI KRFYVQNGKV IPNSESTISG VTGNSITTPF CTAQKTAFGD PTSFSDHGGL ASMSAAFEAG MVLVLSLWDD YYANMLW LDS TYPTTKTGAG GPRGTCSTSS GVPASVEASS PNA YVVYSNI KVGAINSTFG	

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:86	156712278	<i>Acremonium thermophilum</i>	MYTKFAALAA LVATVRGQAA CSLTAETHPS LQWQKCTAPG SCTTVSGQVT IDANWRWLHQ TNSSTNCYTG NEWDTSICSS DTDCA TKCCL DGADYTGTYG VTASGNSLNL KFVTQGPYSK NIGSRMYLME SESKYQGFTL LGQEFTFDVD VSNLGCGLNG ALYFVSMDL D GGVSKYTTNK AGAKYGTGYC DSQCPRDLKF INGQANIDGW QPSSNDANAG LGNHGSCCSE MDIWEANKVS AAYTPHPCTT IGQTMCTGDD CGGTYSSDRY AGICDPDGCD FNSYRMGDT S FYGPGKTVDT GSKFTVVTQF LTGSDGNLSE IKRFYVQNGK VIPNSES KIA GVS GNSITTD FCTAQKTA FG DTNVFEERG G LAQMGKALAE PMVLVLSVWD DHAVNMLWLD STYPTDSTKP GAARGDCPIT SGVPADVESQ APNSNVIYSN IRFGPINSTY TGTPSGGNPP GGGTTTTTTT TTSKPSGPPT TTNPSGPQQT HWGQCGGQGW TGPTVCQSPY TCKYSNDWYS QCL
SEQ ID NO:87	21449327	<i>Aspergillus nidulans</i> (also known as <i>Emericella nidulans</i>)	MYQRALLFSA LLSVSR AQQA GTAQEEVHPS LTWQRCEASG SCTEVAGSVV LDSNWRWTHS VDG YTN CYTG NEWDATLC PD NESCAQNCAV DGADYEATYG IT SNGDSLTL KFVTGSNVGS RYLMEDDET YQMF D L L N N E FTFDVDVSNF PCGLNGALYF TSMDADGGLS KYEGNTAGAK YGTGYCDSQC PRDIKFINGL GNVEGWEP S D SDANAGVGGM GTCCPEMDIW EANSISTAYT PHPCDSVEQT MCEGDS CGGT YSDDRYGGTC DPGCCDFNSY RMGNTRFYGP GAIIDTSSKE TVVTQFIADG GSLSEIKRFY VQNGEVIPNS ESNISGVEGN SITSEFCTAQ KTAFGDEDIF AQHGGLS AMG DAASAMVLIL SIWDDHHSSM MWLDSSSYPTD ADPSQPGVAR GTCEQGAGDP DVVESEHADA SVTF SNIKEG PIGTF
SEQ ID NO:88	171683762	<i>Podospora anserine</i> (<i>S mat</i> +)	MMMKYLIQYL AAGSLMTGLV AGQGVGTQQT ETHPRITWKR CTGKANC'TTV QAEVVIDSNW RWIHTSGGTN CYDGNAWN TA ACSTATDCAS KCLMEGAGNY QQTYGASTSG D S L T L K F V T K HEYGTNVGSR FYLMNGASKY QMFTLMNNEF TFDVDLSTVE CGLNSALYFV AMEEDGGMRS YPTNKAGAKY GTGYCDAQCA RDLKFVGGKA NIEGWRESSN DENAGVGPYG GCCAEIDVWE SNAHAYAFTP HACENNNYHV CERDTCGGTY SEDRFAGGCD ANGCDYNPYR MGNPDFYGKG KTVDTTKKFT VVTRFQDDNL EQFFVQNGQK ILAPAPTFDG IPASPNLTPE FCSTQFDVFT DRNRFREVGD FPQLNAALRI PMVLVMSIWA DHYANMLWLD SVYPPEKEGE PGAARGPCAQ DSGVPSEVKA NYPNAKV VWS NIRFGPIGST VNV
SEQ ID NO:89	56718412	<i>Thermoascus aurantiacus</i> var <i>levisporus</i>	MYQRALLFSF FLAARAQQA GTVTAENHPS LTWQQCSSGG SCTTQNGKVV IDANWRWVHT TSGYTN CYTG NTWDT SICPD DVTCAQNCAL DGADYSGTYG VTTSGNALRL NEVTQSSGKN IGSRLYLQD DTTYQIFKLL GQEFTFDVDV SNLPCGLNGA LYFVAMDADG GLSKYPGNKA GAKYGTGYCD SQCP RDLKFI NGQANVEGWQ PSANDPNAGV GNHGSCCAEM DVWEANSIST AVTPHPCDTP GQTMCGDDC GGTYSSTRYA GTCDDPDG CDF NPYRQGNHSF YGPGKIVDTS SKFTVVTQFI TDDGTPSGTL TEIKRFYVQN GKVIPQSEST ISGVTGNSIT TEYCTAQKAA FGDNTGFFTH GGLQKISQAL AQGMVLVMSL WDDHAANMLW IDSTYPTDAD PDTPGVARGT CPTTSGVPAD VESQNPNSYV IYSNIKV GPI NSTFTAN

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:90	15824273	<i>Pseudotrichonympha grassii</i>	MEAIVLLGLT RSLGTGTNQA ENHPSLSWQN CRSGGSCTQT SGSVVLDSNW RWTHDSSLTN CYDGENEWSSS LCPDPKTCSD NCLIDGADYS GTYGITSSGN SLKLVFVTNG PYSTNIGSRV YLLKDESHYQ IFDLKNKEFT FTVDDSNLDC GLNGALYFVS MDEDGGTSRF SSNKAGAKYG TGYCDAQCPH DIKFINGEAN VENWKPQTND ENAGNGRYGA CCTEMDIWEA NKYATAYTPH ICTVNGEYRC DGSECDDTDS GNRYGGVCDK DGCDFNSYRM GNTSFWGPGL IIDTGKPVTV VTQFVTKDGT DNGQLSEIRR KYVQGGKVIE NTVVNIAGMS SGNSTITDDFC NEQSAFGDT NDFEKKGGLS GLGKAFDYGM VLVLSLWDDH QVNMLWLDSI YPTDQPASQP GVKRGPCCATS SGAPSDVESQ HPDSSVTFSD IRFGPIDSTY	
SEQ ID NO:91	115390801	<i>Aspergillus terreus</i> NIH2624	MHQRALLFSA LVGAVRAQQA GTLTEEVHPP LTWQKCTADG SCTEQSGSVV IDSNWRWLHS TNGSTNCYTG NTWDESLCPD NEACAANCAL DGADYESTYG ITTSGDALTL TFVTGENVGS RVYLMAEDDE SYQTFDLVGN EFTFDVDVSN LPCGLNGALY FTSMADGGV SKYPANKAGA KYGTGYCDSQ CPRDLKFING MANVEGWTPS DNDKNAGVGG HGSCCPELDI WEANSISSAF TPHPCDDLQ TMCSGDDCGG TYSETRYAGT CDPDGCDFENA YRMGNTSYG PDKIVDTNSV MTVVTFQFIGD GGSLSIEIKRL YVQNGKVIAN AQSNVDGVGT NSITSDFCTA QKTAFGDQDI FSKHGGLSGM GDAMSAMVLI LSIWDDHNSS MMWLDSTYPE DADASEPGVA RGTCEHGVGD PETVESQHPG ATVTFSKIKE GPIGSTYSSN STA	
SEQ ID NO:92	453223	<i>Phanerochaete chrysosporium</i>	MFRAAALLAF TCLAMVSGQQ AGTNTAENHP QLQSQQCTTS GGCKPLSTKV VLDSNWRWVH STSGYTNCYT GNEWDTSLCP DGKTCANCA LDGADYSGTY GITSTGTALT LKFVTGSNVG SRVYLMADDT HYQLLKLINQ EFTFDVDMSN LPCGLNGALY LSAMDADGGM SKYPGNKAGA KYGTGYCDSQ CPKDIKFING EANVGNWTET GSNTGTGSYG TCCSEMDIWE ANNDAAAFTP HPCTTTGQTR CSGDDCARNT GLCDGDGCD F NSFRMGDKTF LGKGMTVDTS KPFTVVVTQFL TNDNTSTGTL SEIRRIYIQN GKVIQNSVAN IPGVDPVNSI TDNFCAQQKT AFGDTNWFAQ KGLKQMG EA LGNGMVLALS IWDDHAAANML WLDSDYPTDK DPSAPGVARG TCATTSGVPS DVESQVPNSQ VVFSNIKFGD IGSTFSGTSS PNPPGGSTTS SPVTTSPTPP PTGPTVPQWG QCGGIGYSGS TTCASPYTCH VLNPCESILS LQRSSNADQY LQTTRSATKR RLDTALQPRK	
SEQ ID NO:93	3132	<i>Phanerochaete chrysosporium</i>	MRTALALILA IAAFSAVSAQ QAGTITAETH PTLTIQQCTQ SGGCAPLTTK VVLDVNWRWI HSTTGYTNCY SGNTWDAILC PDPVTCAANC ALDGADYTGT FGILPSGTSV TLRPVDGLGL RLFLLLADDSH YQMFQLLNKE FTFDVEMPNM RCGSSGAIHL TAMDADGGLA KYPGNQAGAK YGTGFCSAQC PKGVKFINGQ ANVEGWLGTT ATTGTFGFFGS CCTDIALWEA NDNSASFAPH PCTTNSQTRC SGSDCTADSG LCDADGCNEN SFRMGNTTFF GAGMSVDTTK LFTVVVTQFIT SDNTSMGALV EIHRLYIQNG QVIQNSVVNI PGINPAT SIT DDLC AQENAA FGGTSSFAQH GGLAQVGEAL RSGMVLALS I VNSAADTLWL DSNYPADADP SAPGVARGTC PQDSASIPEA PTPSVVFSNI KLGDIGTTFG AGSALFSGRS PPGVPVGSAP ASSATATAPP FGSQC GGLGY AGPTGVCPSP YTCQALNIYY SQCI	

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:94	16304152	<i>Thermoascus aurantiacus</i>	MYQRALLFSF FLAAARAHEA GTVTAENHPS LTWQQCSSGG SCTTQNGKVV IDANWRWVHT TSGYTNCYTG NTWDTSLICPD DVTCAQNICAL DGADYSGTYG VTTSGNALRL NEVTQSSGKN IGSRLYLLQD DTTYQIFKLL GQEFTFDVDV SNLPCGLNGA LYFVAMDADG NLSKYPGNKA GAKYGTGYCD SQCPRLKFI NGQANVEGWQ PSANDPNAGV GNHGSSCAEM DVWEANSIST AVTPHPCDTP GQTMCGDDC GGTYSSTRYA GTCDDTDGCD FNPYQPGNHSF YGPGKIVDTS SKFTVVVTQFI TDDGTPSGTL TEIKRFYVQN GKVIPQSEST ISGVTGNSIT TEYCTAQKAA FDNLTGFFTHG GLQKISQALA QGMVLVMSLW DDHAANMLWL DSTYPTDADP DTPGVARGTC PTTSGVPADV ESQNPNSYVI YSNIKVGPIN STFTAN
SEQ ID NO:95	156712280	<i>Acremonium thermophilum</i>	MHKRAATLSA LVVAAAGFEAR GQGVGTQOTE THPKLTFQKC SAAGSCTTON GEVVIDANWR WVHDKNGYTN CYTGNEWN TT ICADAAASCAS NCVVDGADYQ GTYGASTSGN ALTLKFVTKG SYATNIGSRM YLMASPTKYA MFTLLGHEFA FDVDLSKLPC GLNGAVYFVS MDEDGGTSKY PSNKAGAKYG TGYCDSQCPR DLKFIDGKAN SASWQPSSND QNAGVGGMGS CCAEMDIWEA NSVSAAYTPH PCQNYQQHSC SGDDCGGTYS ATRFAGDCDP DGCDWNAYRM GVHDFYGN GK TVDTGKKFSI VTQFKGSGST LTEIKQFYVQ DGRKIENPNA TWPGLEPFNS ITPDFCKAQK QVFGDPPDRFN DMGGFTNMAK ALANPMVLVL SLWDDHYSNM LWLDSTYPTD ADPSAPGKGR GTCDTSSGVP SDVESKNGDA TVIYSNIKFG PLDSTYTAS
SEQ ID NO:96	5231154	<i>Volvariella volvacea</i>	MRASLLAFSL NSAAQQQAGT LQTKNHPSLT SQKCRQGGCP QVNTTIVLDA NWRWTHSTSG STNCYTGN TW QATLCPDGKT CAANCALDGA DYTGTYGVTT SGNSLTLQFV TQSNVGARLG YLMADDTTYQ MFNLLNQEFW FDVDMSNLPC GLNGALYFSA MARTAAWMPM VVCASTPLIS TRRSTARLLR LPVPPRSRYG RGICDSQCPR DIKFINGEAN VQGWQPSND TNAGTGNYGA CCNKMDVWEA NSISTAYTPH PCTQRGLVRC SGTACGGGSN RYGSICDHDG LGFQNLFGMG RTRVRARVGR VKQFNRSRV VEPISWTKQT TLHLGNLPWK SADCNVQNGR VIQNSKVNIP GMPSTMDSVT TEFCNAQKTA FNDTFSFQOK GGMANMSEAL RRGMLVLVLSI WDDHAANMLW LDSITSAAAC RSTPSEVHAT PLRESQIRSS HSRQTRYVTF TNIKFGPFNS TGTTYTTGSV PTTSTSTGTT GSSTPPQPTG VTVPQGQCGG IGYTGPTTCA SPTTCHVLNP YYSQCY
SEQ ID NO:97	116200349	<i>Chaetomium globosum</i> CBS 148-51	MKQYLQYLAA ALPLMSIVSA QGVGTSTSET HPKITWKKCS SGGSCSTVNA EVVIDANWRW LHNADSKNCY DGNWTDACT SSDDCTSKCV LEGAEYGKTY GASTSGDSLS LKFLTCKHEY TNIGSRFYLM NGASKYQMFT LMNNEFAFDV DLSTVECGLN SALYFVAMEE DGMASYSTN KAGAKYGTGY CDAQCARDLK FVGKANYDG WTPSSNDANA GVGALGGCCA EIDVWESNAH AFAFTPHACE NNNYHVCEDT TCGGTYSEDR FAGDCDANGC DYNPYRVGNT DFYGKGMTVD TSKKFTVVSQ FQENKLTQFF VQNGKKIEIP GPKHEGLPTE SSDITPELCS AMPEVFGDRD RFAEVGGFDA LNKALAVPMV LVMSIWDDHY ANMLWIDSSY PPEKAGTPGG DRGPCAQDSG VPSEVESQYP DATVVWSNIR FGPIGSTVQV

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:98	4586343	<i>Irpex lacteus</i>	MFPKASLLIAL SFIAAVYGQQ VGTQMAEVHP KLPSQLCTKS GCTNQNTAVV LDANWRWLHT TSGYTNCTGT NSWDATLCPD ATTCANCAV DGADYSGTYG ITTSGNALTL KFKTGTNVGS RVYLMQTDTA YQMFQLLNQE FTFDVDMSNL PCGLNGALYL SQMDQDGGLS KFTPNKAGAK YGTGYCDSQC PHDIKFINGM ANVAGWAGSA SDPNAGSGTL GTCCSEMDIW EANNDAAAFT PHPCSVDGQT QCSGTQCGDD DERYSGLCDK DGCDFNSFRM GDKSFLGKGM TVDTSRKFTV VTQFVTTDGT TNGDLHEIRR LYVQDGKVIQ NSVVSIPGID AVDSITDNFC AQKSVFGDT NYFATLGGKL KMGAAALKSGM VLAMSVWDDH AASMQLDSN YPADGDATKP GVARGTCSAD SGLPTNVESQ SASASVTFSN IKWGDINTTF TGTGSTSPSS PAGPVSSSTS VASQPTQPAQ GTVAQWQGCG GTGFTGPTVC ASPFTCHVVNPYYSQCY
SEQ ID NO:99	15321718	<i>Lentimula edodes</i>	MERTAALLSF AYLAVVYGQQ AGTSTAETHP PLTWEQCTSG GSCTTQSSSV VLDSNWRWTH VVGGYTNCTY GNEWNTTVCP DGTTCANCA LDGADYEGTY GISTSGNALT LKFVTASAQT NVGSRVYIMA PGSETEYQMF NPLNQEFTFD VDVSAALPCGL NGALYFSEMD ADGGLSEYPT NKAGAKYGTG YCDSQCPRDI KFIGKANVE GWTPSSTSPN AGTGGTGICC NEMDIWEANS ISEALTPHPC TAQGGTACTG DSCSSPNSTA GICDQAGCDF NSFRMGDTSF YGPGLTVDTT SKITVVTQFI TSDNTTTGDL TAIRRIYVQN GQVIQNSMSN IAGVTPPTNEI TTDFCDQQKT AFGDTNTFSE KGGLTGMGAA FSRGMVLVLS IWDDDAEAEML WLDSTYPVGK TGPGAARGTC ATTSGQPDQV ETQSPNAQVV FSNIKFGAIG STFSSTGTGT GTGTGTGTGT GTTSSAPAA TQTKYGGQCGG QGWGTATVCA SGSTCTSSGP YYSQCL
SEQ ID NO:100	146424875	<i>Pleurotus sp Florida</i>	MERTAALTAF TFAAVVLGQQ VGTLTTHENHP ALSIQQCTAT GCTTQQKSVV LDSNWRWTHS TAGATNCTGT NAWDPALCPD PATCATNCAI DGADYSGTYG ITTSGNALTL REVTVNGQYSQ NIGSRVYLLD DADHYKLFDL KNQEFTFDVD MSGLPCGLNG ALYFSEMAAD GGKAAHAGNN AGAKYGTGYC DAQCPHDIKW INGEANVLDW SASATDDNAG NGRYGACCAE MDIWEANSEA TAYTPHVCRD EGLYRCSGTE CGDGNTRYGG VCDKDGCDEN SYRMGDKNFL GRGKTIDTTK KVTVVVTQFIT DNNTPTGNLV EIRRVYVQNG VVYQNSFSTF PSLSQYNSIS DEFCVAQKTL FGDNQYYNTH GGTTKMGDAF DNGMVLIMSL WSDHAAHMLW LDSDYPLDKS PSEPGVSRGA CPTSSGDDPD VVANHPNASV TFSNIKYGPI GSTFGGSTPP VSSGGSSVPP VTSTTSSGTT TPTGPTGTVP KWQCQGGIGY SGPTACVAGS TCTYSNDWYS QCL
SEQ ID NO:101	62006158	<i>Fusarium venenatum</i>	MYRAIATASA LIAAVRAQQV CSLTPETKPA LWSKCTSSG CSNVQGSVTI DANWRWTHQL SGSTNCTGN KWDTSICTSG KVCAEKCCID GAEYASTYGI TSSGNQLSLS FVTKGTYGTN IGSRTYLMED ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSMADAG GKAKYPGNKA GAKYGTGYCD AQCPRDVKFI NGQANSDGWQ PSKSDVNGGI GNLGTCCPEM DIWEANSIST AHTPHPCTKL TQHSCTGDSC GGTYSEDRYG GTCDAADGCD NAYRQGNKTF YPGSGGFNVD TTKKVTVVTQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGVPGSSLTP EFCTAQKKVF GDIDDFEKKG AWGMSDALE APMVLVMSLW HDHHSNMLWL DSTYPTDSTK LGAQRGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YKEGQPEPTN PTNPNTTPG GTVDQWQGCG GTNYSGPTAC KSPFTCKKIN DFYSQCQ

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:102	296027	<i>Phanerochaete chrysosporium</i>	MERTATLLAF TMAAMVFGQQ VGTNTAENHR TLTSQKCKTS GGCSNLNTKI VLDANWRWLH STSGYTNCYT GNQWDATLCP DGKTCANCA LDGADYTGTY GITASGSSLK LQFVTGSNVG SRVYLMADDT HYQMFQLLNQ EFTFDVDMSN LPCGLNGALY LSAMDADGGM AKYPTNKAGA KYGTGYCDSQ CPRDIKFING EANVEGWNAT SANAGTGNYG TCCTEMDIWE ANNDAAAATP HPCTTNAQTR CSGSDCTRDT GLCDADGCFE NSFRMGDQTF LGKGLTVDTs KPFTVVVTQFI TNDGTSAGTL TEIRRLYVQN GKVIQNSSVK IPGIDLVNSI TDNFCsQQKT AFGDTNYFAQ HGGLKQVGEA LRTGMVLALS IWDDYAAANML WLDSNYPTNK DPSTPGVARG TCATTSGVPA QIEAQSPNAY VVFSNIKFGD LNTTYTGTVS SSSVSSSSHSS TSTSSSHSSS STPPTQPTGV TVPQWGGCGG IGYTGSTTCA SPYTCHVLNP YYSQCY	
SEQ ID NO:103	154449709	<i>Fusicoccum sp</i> BCC4124	MYQTSLLASL SELLATSQAQ QVGTQTAETH PKLTTQKCTT AGGCTDQSTS IVLDANWRWL HTVDGYTNCY TGQEWDTsIC TDGKTCAEK ALDGADYEST YGISTSGNAL TMNFVTKSSQ TNIGGRVYLL AADsDDTYEL FKLKNQEFFE DVDVSNLPCG LNgALYFSEM DSDGGLSKYT TNKAGAKYGT GYCDTQCPHD IKFINGEANV QNWTASSTDK NAGTGHYGsC CNEMDIWEAN SQATAFTPHV CEAKVEGQYR CEGTECGDGD NRYGGVCDKD GCDFNSYRMG NETFYGSNGS TIDTTKKFTV VTQFITADNT ATGALTEIRR KYVQNDVVIE NSYADYETLS KFNSITDDEF AAQKTLsGDT NDFKTKGGIA RMGESFERGM VLVMSVWDDH AANALWLDSS YPTDADASKP GVKRGPCSTS SGVPSDVEAN DADSSVIYSN IRYGDIGSTF NKTA	
SEQ ID NO:104	169859460	<i>Coprinopsis cinerea okayama</i>	MFSKVALTAL CFLAVAQAQQ VGREVAENHP RLPWQRCTRN GGCQTVsNGQ VVLDANWRWL HVTdGYTNCY TGNAWNssVC SDGATCAQRC ALEGANYQQT YGITTSGDAL TIKFLTRSEQ TNIGARVYLM ENEDRYQMfN LLNKEFTFDV DVSKVPCGIN GAlYFIQMDA DGGLssQPNn RAGAKYGTGY CDSQCPRDIK FINGEANSVG WEPSETDPNA GKGQYGICCA EMDIWEANSI SNAYTPHPCQ TVNDGGYQRC QGRDCNQPRY EGLCDPDGCD YNPFRMGNKD FYGPGKTVDT NRKMTVVVTQF ITHDNTDTGT LVDIRRLYVQ DGRVIANPPT NFPGLMPAHD SITQEFCDDA KRAFEDNDsF GRNGGLAHMG RSLAKGHVLA LSIWNDHTAH MLWLDsNYPT DADPNKPGIA RGTCPTTGGS PRDTEQNHPD AQVIFsNIKF GDIGSTFSGN	
SEQ ID NO:105	50400675	<i>Trichoderma harzianum</i> (anamorph of <i>Hypocrea lixii</i>)	MYRKLAVISa FLAAARAQQV CTQQAETHPP LTWQKCTASG CTPQQGSVVL DANWRWTHDT KSTTNCYDGN TWSSTLCpDD ATCAKNCCLD GANYSGTYGV TTSGDALTLQ FVTASNVGSR LYLMANDsTY QEFTLSGNEF sFDVDVSQLP CGLNGALYFV SMDADGGQSK YPGNAAGAKY GTGYCDsQCP RDLKFINGQA NVEGWEPSsN NANTGVGGHG SCCSEMDIWE ANsISEALTp HPCETVgQTM CSGDSCGGTY SNDRYGGTCd PDGCDWNPYR LGNTsFYGPG SSFALDttKK LTVVtQFATD GSISRYYVQN GVKFQQPNAQ VGSYSGNTIN TDYCAAEQTA FGGTsFTDKG GLAQINKAFQ GGMVLVMSLW DDYAVNMLWL DsTYPTNATA STPGAKRGsC STSSGVPAQV EAQSPNSKVI YSNIRFGPIG STGGNTGSNP PGtSTtTRAPP sSTGSSPTAT QTHYGGCcgGT GWTGPTRCAs GYTCQVLNPF YSQCL	

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:106	729649	<i>Neurospora crassa</i> (OR74A)	MRASLLAFSL AAAVAGGQQA GTLTAKRHPS LTWQKCTRGG CPTLNTTMVL DANWRWTHAT SGSTKCYTGN KWQATLCPDG KSCAANCALD GADYTGTYGI TSGGWSLTLQ FVTDNVGARA YLMADDTQYQ MLELLNQELW FDVDMSNIPC GLNGALYLSA MDADGGM RKY PTNKAGAKYA TGYCDAQCPR DLKYINGIAN VEGWTPSTND ANGIGDHGSC CSEMDIWEAN KVSTAFTPHP CTTIEQHMCE GDSCGGTYS D DRYGVLCDAD GCDFNSYRMG NTTYFYGEGKT VDTSSKFTVV TQFIKDSAGD LAEIKAFYVQ NGKVIENSQS NVDGVSGNSI TQSFCKSQKT AFGDIDDFNK KGLKQMGKA LAQAMVLVMS IWDDHAAANML WLDSTYPVPK VPGAYRGSGP TTSGVPAEVD ANAPNSKVAF SNIKFGHLGI SPFSGGSSGT PPSNPSSSAS PTSSTAKPSS TSTASNPSGT GAAHWAQCGG IGFSGPPTCP EPYTCAKDHD IYSQCV
SEQ ID NO:107	119472134	<i>Neosartorya fischeri</i> NRRL 181	MLASTFSYRM YKTALILAAL LGSGQAQQVG TSQAEVHPSM TWQSCTAGGS CTTNNGKVVI DANWRWVHKV GDYTNICYTGN TWDKTLCRDD ATCASNCAL E GANYQSTYGA TTSGDSLRLN FVTTSQQKNI GSRLYMK added TTYEMFKLLN QEFTFDVDVS NLP CGLNGAL YFVAMDADGG MSKYPTNKAG AKYGTGYCDS QCPRDLKFIN GQANVEGWQP SSNDANAGTG NHGSCCAEMD IWEANSISTA FTHPCDTPG QVMCTGDACG GTYSSDRYGG TCDPDGCDFN SFRQGNKTFY GPGMTVDTKS KFTVVVTQFIT DDGTASGTLK EIKRFYVQNG KVIPNSESTW SGVGGNSITN DYCTAQKSLE KDQNVFAKHG GMEGMGAALA QGMVLVMSLW DDHAAANMLWL DSNYPPTASS STPGVARGTC DISSGVPADV EANHPPDASVV YSNIKVGPIG STFNSSGGSNP GGGTTTTAKP TTTTTTTAGSP GGTGVAQHYG QCGNGWQGP TTCASPYTCQ KLNDFYSQCL
SEQ ID NO:108	117935080	<i>Chaetomium thermophilum</i>	MQIKQYLQYL AAALPLVNMA AAQRAGTQQT ETHPRLSWKR CSSGGNCQTV NAEIVIDANW RWLHDSNYQN CYDGNRWTS A CSSATDCAQK CYLEGANYGS TYGVSTSGDA LTLKFVTKHE YGTNIGSRVY LMNGSDKYQM FTLMNNEFAF DVDLSKVECG LNSALYFVAM EEDGGMRSYS SNKAGAKYGT GYCDAQCARD LKFVGGKANI EGWRPSTNDA NAGVGPIYGAC CAEIDVWESN AYAF AFTPHG CLNNNYHVCE TSNCGGTYSE DRFGGLCDAN GCDYNPYRMG NKDFYGKGKT VDTSRKFTVV TRFEENKLTQ FFIQDGRKID IPPPTWPGLP NSSAITPELC TNLSKVFD DR DRYEETGGFR TINEALRIPM VLVMSIWDGH YASMLWLDV YPPEKAGQPG AERGPCAPTS GVPAEVEAQF PNAQVIWSNI RFGPIGSTYQ V
SEQ ID NO:109	154300584	<i>Botryotinia fuckeliana</i> B05-10	MTSRIALVSL FAAVYGQQVG TYQTETHPSL TWQSCTAKGS CTTNTGSIVL DGNWRWTHGV GTSTNCYTGN TWDATLCRDD ATCAQNCALE GADYSGTYGI TTSGNSLRLN FVTQSANKNI GSRVYLMADT THYKTFNLLN QEFTFDVDVS NLP CGLNGAV YFANLPADGG ISSTNTAGAE YGTGYCDSQC PRDMKFIKGQ ANVDGWVPSS NNANTGVGNH GSCCAEMDIW EANSISTAVT PHSCDVTQT VCTGDDCCGGT YSSSRYAGTC DPDGCDFNSY RMGDETFYGP GKTVDTNSVF TVVTQFLTTD GTASGTLNEI KRFYVQDGKV IPNSYSTISG VSGNSITTPF CDAQKTAFGD PTSFSDHGGL ASMSAAFEAG MVLVLSLWDD YYANMLWLDS TYPVGKTSAG GPRGTCDTSS GVPASVEASS PNAVYVYSNI KVGAINSTYG

TABLE 1			
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence
SEQ ID NO:110	15824271	<i>Pseudotriconympha grassii</i>	MEVFVLLWLT QSLGTGTNQA ENHPSLSWQN CRSGGSCTQT SGSVVLDSNW RWTHDSSLTN CYDGENEWSSS LCPDPKTCSD NCLIDGADYS GTYGITSSGN SLKLVFVTNG PYSTNIGSRV YLLKDESHYQ IFDLKNKEFT FTVDDSNLDC GLNGALYFVS MDEDGGTSRF SSNKAGAKYG TGYCDAQCPH DIKFINGEAN VENWKPQTND ENAGNGRYGA CCTEMDIWEA NKYATAYTPH ICTVNGEYRC DGSECGDTS GNRYGGVCDK DGCDFNSYRM GNTSFWGPGL IIDTGKPVTV VTQFVTKDGT DNGQLSEIRR KYVQGGKVIE NTVVNIAGMS SGNSTITDDFC NEQSAFGDT NDFEKKGGLS GLGKAFDYGM VLVLSLWDDH QVNMLWLD SI YPTDQPASQP GVKRGPCCATS SGAPSDVESQ HPDSSVTFSD IRFGPIDSTY
SEQ ID NO:111	4586345	<i>Irpex lacteus</i>	MFRKAALLAF SFLAIAHQQQ VGTNQAEHNP SLPSQKCTAS GCTTSSTSVV LDANWRWVHT TTGYTNCTGTG QTWDAASICPD GVTCAKACAL DGADYSGTYG ITTSGNALTL QFVKGTNVGS RVYLLQDASN YQMFQLINQE FTFDVIDMSNL PCGLNGAVYL SQMDQDGGVS RFPNTAGAK YGTGYCDSQC PRDIKFINGE ANVEGWTGSS TDSNSGTGNY GTCCSEMDIW EANSVAAAYT PHPCSVNQQT RCTGADCGQG DDRYDGVCDP DGCDFNSFRM GDQTFLGKGL TVDTSRKFTI VTQFISDDGT TSGNLAEIRR FYVQDGNVIP NSKVSIAGID AVNSITDDFC TQQTAFAGDT NRFAAQGGLK QMGAALKSGM VLALS L WDDH AANMLWLDSD YPTTADASN P GVARGTCPTT SGFPRDVESQ SGSATVTYSN IKWGD LNSTF TGTLTTPSGS SSPSPASTS GSSTSASSA SVPTQSGTVA QWAQCGGIGY SGATTVCVSPY TCHVVNAYYS QCY
SEQ ID NO:112	46241268	<i>Gibberella avenacea</i>	MYRAIATASA LIAAARAQQV CTILTETKPA LTWSKCTSSG CTDVKGSGVI DANWRWTHQT SSSTNCYTG N KWDTSVCTSG ETCAQKCCLD GADYAGTYGI TSSGNQLSLG FVTKGSFSTN IGSRTYIMEN ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSMADAG GKARYPANKA AQCPRDVKFI NGKANSDGWK PSDSDINAGI GNMGTCCPEM DIWEANSIST AFTPHPCTKL TQHACTGDSC GGTYSNDRYG GTCDADGCD F NSYRQGNKTF YGRGSDFNVD TTKKVTVV TQ FKKG SNGRLS EITRLYVQNG KVIANSESKI PGNSGSSLTA DFCSKQKSVF GDIDDFSKG GWSGMSDALE SPPMVLVMSL WHD HSNMLW LDSTYPTDST KLGAQRGSCA TTSGVPSDLE RDVPNSKVSE SNIKEGP IGS TYSSGTTNPP PSSTD TSTTP TNPPTGGTVG QYGQCGGQTY TGP KDCSPY TCKKINDFYS QCQ
SEQ ID NO:113	6164684	<i>Aspergillus niger</i>	MSSFQIYRAA LLLSILATAN AQQVGTYTTE THPSLTWQTC TSDGSC TTND GEVIDANWR WVHSTSSATN CYTGNEWDTS ICTDDVTCAA NCALDGATYE ATYGVTTSGS ELRLNFVTQG SSKNIGSRLY LMSDDSNYEL FKLLGQEFTF DVDVSNLPCG L NGALYFVAM DADGGTSEYS GNKAGAKYGT GYCDSQCPRD LKFINGEANC DGWEPSSNNV NTGVGDHGSC CAEMDVWEAN SISNAFTAHP CDSVSQTMCD GDSCGGTYS A SGDRYSGTCD PDGCDYNPYR IGNTDFYGP G LTVD TNSPFT VVTQFITDDG TSSGTITEIK RLYVQNGEVI ANGASTYSSV NGSSITSAFC ESEKTLFGDE NVFDKHG GLE GMGEAMAKGM VLVLSLWDDY AADMLWLDSD YPVNSSASTP GVARGTCSTD SGVPATVEAE SPNAYVTYSN IKFGPIGSTY SSGSSSGSGS SSSSSSTTTK ATSTTLKTT S TTSSGSSSTS AAQAYQC GG QGWTGPTTCV SGYTCTYENA YYSQCL

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:114	6164682	<i>Aspergillus niger</i>	MHQRAALLFSA LLTAVRAQQA GTLTAEVHPS LTWQKCTSEG SCTEQSGSVV IDSNWRWTHS VNDSTNCYTG NTWDATLCPD DETCAANCAL DGADYESTYG VTTDGDLSLTL KFVTGSGNVGS RLYLMDTSD E GYQTFNLLDA EFTFDVDVSN LPCGLNGALY FTAMDADGGV SKYPANKAGA KYGTGYCDSQ CPRDLKFIDG QANVDGWEP S SNNDNTGIGN HGSCCPEMDI WEANKISTAL TPHPCD SSEQ TMCEGNDCGG TYSDDRYGGT CDPDGCDFNP YRMGNDSEFYG PGKTIDTGSK MTVVVTQFITD GSGSLSEIKR YVQNGNVIA NADSNISGVT GNSITTD ECT AQKKAFGEDED IFAEHNGLAG ISDAMSSMVL ILSLWDDYYA SMEWLDS DYP ENATATDPGV ARGTC DSESG VPATVEGAHP DSSVTFESNIK FGPINSTFSA SA	
SEQ ID NO:115	33733371	<i>Chrysosporium lucknowense</i> US6573086-10	MYAKFATLAA LVAGAAAQNA CTLTAENHPS LTWSKCTSGG SCTSVQGSIT IDANWRWTHR TDSATNCYEG NKWDTSYCSD GPSCASKCCI DGADYSSTYG ITTSGNSLNL KFVTKGQYST NIGSRTYLME SDTKYQMFQL LGNEFTFDVD VSNLGCGLNG ALYFVSM DAD GGMSKYSGNK AGAKYGTGYC DSQCPRDLKF INGEANVENW QSSTNDANAG TGKYGSCCSE MDVWEANNMA AAFTPHPCXV IGQSRCEGDS CGGTYSTDRY AGICDPDGCD FNSYRQGNKT FYGKGMTVD T TKKITVV TQF LKNSAGELSE IKRFYVQNGK VIPNSESTIP GVEGNSITQD WCDRQKAAFG DVTDXQDKGG MVQMGKALAG PMVLVMSIWD DHAVNMLWLD STWPIDGAGK PGAERGACPT TSGVPAEVEA EAPNSNVIFS NIRFGPIGST VSGLPDGGSG NPNPVSSST PVPSSSTTSS GSSGPTGGTG VAKHYEQCGG IGFTGPTQCE SPYTCTKLND WYSQCL	
SEQ ID NO:116	29160311	<i>Thielavia australiensis</i>	MYAKFATLAA LVAGASAAQAV CSILTAETHPS LTWQKCTAPG SCTNVAGSIT IDANWRWTHQ TSSATNCYSG SKWDSSICTT GTDCASKCCI DGAEYSSTYG ITTSGNALNL KFVTKGQYST NIGSRTYLME SDTKYQMFKL LGNEFTFDVD VSNLGCGLNG ALYFVSM DAD GGMSKYSGNK AGAKYGTGYC DAQCPRDLKF INGEANVEGW ESSTNDANAG SGKYGSCC TE MDVWEANNMA TAFTPHPCTT IGQTRCEGDT CGGTYS SDRY AGVCDPDGCD FNSYRQGNKT FYGKGMTVD T TKKITVV TQF LKNSAGELSE IKRFY AQDGK VIPNSESTIA GIPGNSITKA YCDAQKT V FQ NTDDFTAKGG LVQMGKALAG DMVLVMSVWD DHAVNMLWLD STYPTDQGV AGAERGACPT TSGVPSDVEA NAPNSNVIFS NIRFGPIGST VQGLPSSGGT SSSSSAAPQS TSTKASTTTS AVRTTSTATT KTTSSAPAQG TNTAKHWQQC GGNGWTGPTV CESPYKCTKQ NDWYSQCL	
SEQ ID NO:117	146197087	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLTLVYFLLS LVVSLEIGTQ QSEDHPKLTW QNGSSSVSGS IVLDSNWRWV HDSGTTNCYD GNLWSKDLCP SSDTCSQKCY IEGADYSGTY GIQSSGSKLT LKFVTKGSYS TNIGSRVYLL KDENTYESFK LKNKEFTFTV DDSKLNCGLN GALYFVAMDA DGGKAKYSSF KPGAKYGMGY CDAQCPHDMK FISGKANVDD WKPQDN DENS GNGKLGTC CS EMDIWEGNMK SQAYTVHACT KSGQYECTGQ QCGD TDSGDR FKGTC DKGC DYASWRWGDQ SFYGE GKTVD TKQPVTVVTQ FIGDPLTEIR RLYVQGGKTI NNSKTSNLAD TYDSITDKFC DATKEASGDT NDFKAKGAMS GFSTNLNNGQ VLVMSLWDDH YPTDSSDSTA QRGPCPTSSG VPKDVESQH G DATVVFSDIK FGAINSTFKY N	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:118	146197237	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MLAAALTFEA CSVGVGTKTP ENHPKLNWQN CASKGSCSQV SGEVTMDSNW RWTHDGNNGKN CYDGNWTWISS LCPDDKTCSD KCVLDGAHEYQ ATYGIQNGT ALTLKFVTHG SYSTNIGSRL YLLKDKSTYY VFKLNNKEFT FSVDVSKLPC GLNGALYFVE MDADGGKAKY AGAKPGAHEY LGYCDAQCPD DLKFINGEAN SEGWKPQSGD KNAGNGKYG CCSEMDVWES NSQATALTPH VCKTTGQQRG SGKSECGGQD GQDRFAGLCD EDGCDNFNNWR MGDKTFFGPG LIVDTKSPFV VVTQFYGSPV TEIRRKYVQN GKVIENSKSN IPGIDATAAI SDHFCEQQKK AFGDTNDEKN KGGEAKLGQV FDRGMVLVLS LWDDHQVAML WLDSTYPTNK DKSQPGVDRG PCPTSSGKPD DVESASADAT VVYGNIKFGA LDSTY	
SEQ ID NO:119	146197067	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLTLVYFLLS LVVSLEIGTQ QSEDHPKLTW QNGSSSVSGS IVLDSNWRWV HDSGTTNCYD GNLWSKDLC P SSNTCSQKCY IEGADYSGTY GIQSSGSKLT LKFVTKGYSY TNIGSRVYLL KIDENTYESFK LKNKEFTFTV DDSKLNCGLN GALYFVAMDA DGGKAKYSSF KPGAKYMGY CDAQCPHDMK FISGKANVDD WKPQDNDENS GNGKLGTCSS EMDIWEGNMK SQAYTVHACT KSGQYECTGQ QCGLDSDGDR FKGTCDDKDG DYASWRWGDQ SFYGEGETVD TKQPVTVVTQ FIGDPLTEIR RLYVQGGKTI NNSKTSNLAD TYDSITDKFC DATKEASGDT NDFKAKGAMS GFSTNLNNGQ VLVMSLWDDH TANMLWLDST YPTDSTKTGA SRGPCAVSSG VPKDVESQYG DATVIYSDIK FGAINSTFKW N	
SEQ ID NO:120	146197407	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MILALLSLAK SLGIATNQAE THPKLTWTRY QSKGSGQTVN GEIVLDSNWR WTHHSGTNCY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGISTSGNSL KLGFTVTHGSY STNIGSRVYL LRDSKNYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPQSN DENSGRGYACC TEMDIWEANS MATAYTPHVC TVTGLRRC EG TECGDTDANQ RYNGICDDKG CDFNSYRLGD KTFFGVGKTV DSSKPVTVVT QFVTSNGQDS GTLSEIRRKY VQGGKVIENS KVNIAGITAG NSVTDTFCNE QKKAFGDNND FEKKGGLGAL SKQLDAGMVL VLSLWDDHSV NMLWLDSTYP TNAAAGALGT ERGACATSSG APSDVESQSP DATVTFSDIK FGPIDSTY	
SEQ ID NO:121	146197157	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	MLVIALILRG LSVGTGTQQS ETHPSLSWQQ TSKGSGSQSV SGSVVLDNSW RWHTTDTGTT NCYDGN EWSS DLCPDASTCS SNCVLEGADY SGTYGITGSG SSLKLGFVTK GSYSTNIGSR VYLLGDESHY KLFKLENNEF TFTVDDSNLE CGLNGALYFV AMDEDDGGASK YSGAKPGAKY GMGYCDAQCP HDMKFINGDA NVEGWKPSDN DENAGTGKWG ACCTEM DIWE ANKYATAYTP HICTKNGEYR CEGTDCGDTK DNNRYGGVCD KDGCDFN SWR MGNQSFWGPG LIIDTGKPV T VVTQFLADGG SLSEIRRKYV QGGKVIENTV TKISGMDEFD SITDEF CNQQ KKAFRDTNDF EKKGGLKGLG TAVDAGVVIV LSLWDDHDVN MLWLDSIYPT DSGSKAGADR GPCATSSGVP KDVESNYASA SVTFSDIKEG PIDSTY	
SEQ ID NO:122	146197403	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MLLALFAFGK SLGIATNQAE NHPKLTWTRY QSKGSGQTVN GEIVLDSNWR WTHHSGTNCY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGISSSGNSL KLGFTVTHGSY STNIGSRVYL LRDSKNYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPQSN DENSGRGYACC TEMDIWEANS MATAYTPHVC TVTGIRRC EG TECGDTDANQ RYNGICDDKG CDFNSYRLGD KSFFGVGKTV DSSKPVTVVT QFVTSNGQDS GTLSEIRRKY VQGGKVIENS KVNIAGMAAG NSITD TFCNE QKKAFGDNND FEKKGGLGAL SKQLDSGMVL VLSLWDDHSV NMLWLDSTYP TNAAAGALGT ERGACATSSG APSDVESQSP DATVTFSDIK FGPIDSTY	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:123	146197081	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLASVVYLVSLVVSLEIGTQ QSEEHPKLTW QNGSSSVSGS IVLDSNWRWL HDSGTTNCYD GNLWSDDLCP NADTCSSKCY IEGADYSGTY GITSSGSKVT LKFVTKGSYS TNIGSRIYLL KIDENTYETFK LKNKEFTFTV DDSKLDICGLN GALYFVAMDA DGGKAKYSSF KPGAKYMGY CDAQCPHDMK FISGKANVDD WKPQDNDENS GDGKLGTCSS EMDIWEGNAK SQAYTVHACS KSGQYECTGQ QCGDTSGDR FKGTCDDKGC DYASWRWGDQ SFYGEGETVD TKSPVTVVTTQ FIGDPLTEIR RVYVQGGKTI NNSKTSNLAD TYDSITDKFC DATKDATGDT NDFKAKGAMA GFSTNLNTAQ VLVSVHCGMI IQPICCGLIR RIQRIQQKV QAVDRVLCRR VFQRMKLKASM VMLQSRTRTL SLELSTRPLV GISPAGRLEF F	
SEQ ID NO:124	146197413	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MILALLVLGK SLGIATNQAE THPKLTWTRY QSKGSGSTVN GEIVLDSNWR WTHHSGTNCY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGISTSGNSL KLGFTVTHGSY STNIGSRVYL LKDTKSYLEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPPQSNDE SGNRGYGACC TEMDIWEANS MATAYTPHVC TVTGLRRCEG TECGDTDNDQ RYNGICDDKG CDFNSYRLGD KSFFGVGKTV DSSKPVTVVT QFVTSNGQDS GTLSEIRRKY VQGGKVIENS KVNVAGITAG NSVTDTFCNE QKKAFGDNND FEKKGGLGAL SKQLDAGMVL VLSLWDDHSV NMLWLDSTYP TNAAGALGT ERGACATSSG KPDSVESQSP DATVTFSDIK FGPIDSTY	
SEQ ID NO:125	146197309	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MLCIGLISFV YSLGVGTNTA ETHPKLTWKN GGQTVNGEVT VDSNWRWTHT KGSTKNCYDG NLWSKDLCPD AATCGKNCVL EGADYSGTYG VTSSGNALT L KFTVTHGSYST NVGSRLYLLK DEKTYQMFNL NGKEFTFTVD VSNLPCGLNG ALYHVNMD E GGTRYPDNE AGAKYGTGYC DAQCPTDLKF INGIPNSDGW KPQSNDKNSG NGKYGSCCSE MDIWEANS IC SAVTPHVCDN LQTRCQGT A CGENGGSRF GSSCDPDGCD FNSWRMGNKT FYGPGLIVDT KSKFTVVTTQF VGNPVTEIKR KYVQNGKVI E NSYSNIEGMD KFNSVSDKFC TAQKKAFGDT DSFTKHGGFK QLGSAKAGM VLVLSLWDDH TVNMLWLDSV YPTNSKKAGS DRGPCPTTSG VPADVESKSA DANVIYSDIR FGAIDSTYK	
SEQ ID NO:126	146197227	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MLGALVALAS CIGVGTNTPE KHPDLKWTNG GSSVSGSIVV DSNWRWTHIK GETKNCYDGN LWSDKYCPDA ATCGKNCVLE GADYSGTYGV TTSGDAATLK FVTHGQYSTN VGSRLYLLKD EKTYQMFNLV GKEFTFTVDV SNLPCGLNGA LYFVQMDSDG GMAKYPDNQA GAKYGTGYCD AQCPDCLKFI NGIPNSDGWK PQNDKNSGN GKYGSCCSEM DIWEANSMAT AYTPHVCDKL EQTRCSGSAC GQNGGGDRFS SSCDDPDGCD F NSWRMGNKTF WGPGLIVDTK KPVQVVTTQFV GSGGSVTEIK GHVLVLSLWD DHDVNMLWLD SKQYNSVSDE FCQAQKKAFG DNDSFTKHGG FRQLGATLSK GHVLVLSLWD DHDVNMLWLD SVYPTNSNKP GADRGPKCTS SGVPSDVESQ NADSTVKYSD IREGAIDSTY SK	
SEQ ID NO:127	146197253	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MLAAALFTFA CSVGVGTKTT ETHPKLNWQQ CACKGSCSQV SGEVTMDSNW RWTHDGNNGN CYDGNTWISS LCPDDKTCSD KCVLDGA EYQ ATYGIQNGT ALTPKFVTHG SYSTNIGSRL YLLKDKSTYY VFQLNNKEFT FSVDVSKLPC GLNGALYFVE MDADGGKSKY AGAKPGA EYG LGYCDAQCP S DLKFINGEAN SEGWK PQSGD KNAGNGKYGS CCSEMDVWES NSMATALTPH VCKTTGQTRC SGKSECGGD QGDRFAGNCD EDGCDFNWR MGDKTFFGPG LTVDTKSPFV VVTQFYGSPV TEIRRKYVQN GKVIENAKSN IPGIDATNAI SDTFCEQKK AFGDTNDFKN KGGFTKIGSV FSRGMVLVLS LWDDHQVAML DKSVPGVDRG PCPTSSGKPD DVESASGDAT VVYGNIKFGA LDSTY	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:128	146197099	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MEGFLLSLEA LQFALEIGTQ TSESHPSITW ELNGARQSGQ IVIDSNWRWL HDSGTTNCYD GNTWSSDLCP DPEKCSQNCY LEGADYSGTY GISASGSQLT LGFVTHGYSY TNIGSRVYLL KIDENTYPMFK LKNKEFTFTV DVSNLPCGLN GALYFVAMPS DGGKAKYPLA KPGAKYGMGY CDAQCPHDMK FINGEANVLD WKPQSN DENA GTGRYGTCC T EMDIWEANSQ ATAYTVHACS KNARCEGTEC GDD SASQRYN GICDKDGCDF NSWRWGNKTF FGPGLTV DSS KPVTVVTQFI GDPLTEIRRI WVQGGKVIQN SFTNVSGITS VDSITNTFC D ESKVATGDTN DEKAKGMSG FSKALDTEVV LVLSLWDDHT ANMLWLDSTY PTDSTAIGAS RGPCATSSGD PKDVESASAN ASVKFSDIKE GALDSTY	
SEQ ID NO:129	146197409	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MLASLLPLSN SLGTASNQAE THPKLTWTQY TKGAGQTVN GEIVLDSNWR WTHKDG TN CY DGNTWSSSLC PDPTTCSNNC NLDGADYPGT YGITTSGNQL KLG FVTHGSY STNIGSRVYL LRDSKNYQMF KLKNKEFTFT VDDSKLPCGL NGAVYFVAMD EDGGTAKHSI NKAGA QYGTG YCDAQCPHDM KFINGEANVL DWKPQSN DEN SGNR WGARC TEMDIWEANS RATA YTPHIC TKTGLYRCEG TECGDSDTNR YGGVCDK DGC DFNSYRMGDK SFFGQGKTVD SSKPVTVV TQ FITDNNQDSG KLTEIRRK YV QGGKVIDNSK VNIAGITAGN PITDTFCDEA KKAFGDNNDF EKKGGLSALG TQLEAGFVLV LSLWDDH SVN MLWLDSTYPT NASPGALGVE RGDCAITSGV PADVESQSAD ASVTFSDIKE GPIDSTY	
SEQ ID NO:130	146197315	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MLCIGLISFV YSLGVGTNTA ETHPKLTWKN GGQTVNGEVT VDSNWRWTH T KGSTKNCYDG NLWSKDLCPD AATCGKN CVL EGADYSGTYG VTSSGNALTL K FVTHGSYST NVGSRLYLLK DEKTYQMFNL NGKEFTFTVD VSNLPCGLSG ALYHVNMD ED GTKRYPDNE AGAKYGTGYC DAQCPTDLKF INGIPNSD GW KPQSN DKNSG NGKYGSCCSE MDIWEANSIC SAVTPHVCDN LQTRCQGA A CGENGGS RF GSSCDPDGCD FNSWGMGNKT FYGPGLIVDT KSKFTVV TQF VGNPVTEIKR KYVQNGKVIE NSYSNIEGMD KFNSVSDKFC TAQKKAFGDT DSFTKHGGFK QLGSA LAKGM VLVL SLWDDH TVNMLWLD SV YPTNSKKAGS DRGPCPTTSG VPADVESKSA DANVIYSDIR FGAIDSTYK	
SEQ ID NO:131	146197411	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MILALLVLGK SLGIATNQAE THPKLTWTRY QSKGSGSTVN GEIVLDSNWR WTHHSGTN CY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGISTSGNSL KLG FVTHGSY STNIGSRVYL LRDSKNYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGA QYGTG YCDAQCPHDM KFINGEANVL DWKPQSN DEN SGNGRYGACC TEMDIWEANS MATA YTPHVC TVTGLRRCEG TECGDTDNDQ RYNGICDKDG CDFNSYRLGD KSFFGVGKT V DSSKPVTVV T QFVTSNGQDS GILSETRRKY VQGGKVIENS KVN VAGITAG NSVTD TFCNE QKKAFGDNND FEKKGGLGAL SKQLDAGMVL VL SLWDDH SV NMLWLDSTY T TNAAGALGT ERGACATSSG KPSDVESQSP DATVTFSDIK FGPIDSTY	
SEQ ID NO:132	146197161	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	MIGIVLIQTV FGIGVGTQQS ESHPSLSWQQ CSKGGSCTSV SGSIVLDSNW RWTHIPDGT T NCYDGN EWSS DLCPDPTTCS NNCVLEGADY SGTYGISTSG SSAKLGFVTK GSYSTNIGSR VYLLGDESHY KIFDLKNKEF TFTVDDSNLE CGLNGALYFV AMDEDGGASR FTLAKPGAKY GTGYCDAQCP HDIKFINGEA NVQDWKPSDN DDNAGTGHYG ACCTEMDIWE ANKYATAYTP HICTENGEYR CEGKSCGDS S DDRYGGVCDK DGCDFNSWRL GNQSFWGPGL IIDTGKPVTV VTQFVTKDGT DSGALSEIRR KYVQGGKTIE NTVVKISGID EVDSITDEF C NQKQAFGDT NDFEKKGGLS GLGKAFDYGV VLVLSLWDDH DVNMLWLD SV YPTNPAGKAG ADRGPCATSS GDPKEVEDKY ASASVTFSDI KFGPIDSTY	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:133	146197323	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MLVFGIVSEV YSIGVGTNTA ETHPKLTWKN GGSTTNGEVT VDSNWRWTHT KGSTKNCYDG NLWSKDLCPD AATCGKNCVL EGADYSGTYG VTSSGDALT L KFTVTHGSYST NVGSRLYLLK DEKTYQMFNL NGKEFTFTVD VSQ L P C G L N G A L Y F V C M D Q D G G M S R Y P D N Q A G A K Y G T G Y C D A Q C P T D L K F I N G L P N S D G W K P Q S N D K N S G N G K Y G S C C S E M D I W E A N S L A T A V T P H V C D Q V G Q T R C E G R A C G E N G G G D R F G S I C D P D G C D F N S W R M G N K T F W G P G L I I D T K K P V T V V T Q F I G S P V T E I K R E Y V Q G G K V I E N S Y T N I E G M D K F N S I S D K F C T A Q K K A F G D N D S F T K H G G F S K L G Q S F T K G Q V L V L S L W D D H T V N M L W L D S V Y P T N S K K L G S D R G P C P T S S G V P A D V E S K N A D S S V K Y S D I R F G S I D S T Y K	
SEQ ID NO:134	146197077	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	M L S F V F L L G F G V S L E I G T Q Q S E N H P T L S W Q Q C T S S G S C T S Q S G S I V L D S N W R W H D S G T T N C Y D G N E W S S D L C P D P E T C S K N C Y L D G A D Y S G T Y G I T S N G S S L K L G F V T E G S Y S T N I G S R V Y L K K D T N T Y Q I F K L K N H E F T F T V D V S N L P C G L N G A L Y F V E M E A D G G K G K Y P L A K P G A Q Y G M G Y C D A Q C P H M K F I N G N A N V L D W K P Q E T D E N S G N G R Y G T C C T E M D I W E A N S Q A T A Y T P H I C T K D G Q Y Q C E G T E C G D S D A N Q R Y N G V C D K D G C D F N S Y R L G N K T F F G P G L I V D S K K P V T V V T Q F I T S N G Q D S G D L T E I R R I Y V Q G G K T I Q N S F T N I A G L T S V D S I T E A F C D E S K D L F G D T N D F K A K G G F T A M G K S L D T G V V L V L S L W D D H S V N M L W L D S T Y P T D A A A G A L G T Q R G P C A T S S G A P S D V E S Q S P D A S V T F S D I K F G P L D S T Y	
SEQ ID NO:135	146197089	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	M L T L V V Y L L S L V S L E I G T Q Q S E S H P A L T W Q R E G S S A S G S I V L D S N W R W V H D S G T T N C Y D G N E W S T D L C P S S D T C T Q K C Y I E G A D Y S G T Y G I T T S G S K I T L K F V T K G S Y S T N I G S R V Y L L K D E N T Y E T F K L K N K E F T F T V D D S K L D C G L N G A L Y F V A M D A D G G K Q K Y S S F K P G A K Y G M G Y C D A Q C P H D M K F I S G K A N V E D W K P Q D N D E N S G N G K L G T C C S E M D I W E G N A K S Q A Y T V H A C T K S G Q Y E C T G T D C G D S D S R Y Q G T C D K D G C D Y A S Y R W G D H S F Y G E G K T V D T K Q P I T V V T Q F I G D P L T E I R R L Y I Q G G K V I N N S K T Q N L A S V Y D S I T D A F C D A T K A A S G D T N D F K A K G A M A G F S K N L D T P Q V L V L S L W D D H T A N M L W L D S T Y P T D S R D A T A E R G P C A T S S G V P K D V E S N Q A D A S V V F S D I K F G A I N S T Y S Y N	
SEQ ID NO:136	146197091	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	M F G F L L S L F A L Q F A L E I G T Q T S E S H P S I T W E L N G A R Q S G Q I V I D S N W R W L H D S G T T N C Y D G N T W S S D L C P D P E K C S Q N C Y I E G A D Y S G T Y G I S A S G S Q L T L G F V T K G S Y S T N I G S R V Y L L K D E N T Y Q M F K L K N K E F T F T V D V S N L P C G L N G A L Y F V A M P S D G G K A K Y P L A K P G A K Y G M G Y C D A Q C P H D M K F I N G E A N V L D W K P Q S N D E N A G T G R Y G T C C T E M D I W E A N S Q A T A Y T V H A C S K N A R C E G T E C G D S A S Q R Y N G I C D K D G C D F N S W R W G N K T F F G P G L T V D S S K P V T V V T Q F I G D P L T E I R R I W V Q G G K V I Q N S F T N V S G I T S V D S I T N T F C D E S K V A T G D T N D F K A K G G M S G F S K A L D T E V V L V L S L W D D H T A N M L W L D S T Y P S N S T A I G A T R G P C A T S S G D P K N V E S A S A N A S V K F S D I K F G A F D S T Y	
SEQ ID NO:137	146197097	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	M L A L V Y F L L S L V V S L E I G T Q Q S E D H P K L T W Q N G S S S V S G S I V L D S N W R W V H D S G T T N C Y D G N I W S T D L C P S S D T C T S K C Y I E G A D Y S G T Y G I T S S G S K V T L K F V T K G S Y S T N I G S R I Y L L K D E N T Y E T F K L K N K E F T F T V D D S Q L N C G L N G A L Y F V A M D A D G G K A K Y S S F K P G A K Y G M G Y C D A Q C P H D M K F I S G K A N V D D W K P Q D N D E N S G N G K L G T C C S E M D I W E G N A K S Q A Y T V H A C T K S G Q Y E C T G Q Q C G D T D S G D R F K G T C D K D G C D Y A S W R W G D Q S F Y G E G K T V D T K Q P V T V V T Q F I G D P L T E I R R I Y V Q G G K T I N N S K T S N L A D T Y D S I T D K F C D A T K E A S G D T N D F K A K G A M S G F S T N L N T A Q V L V L S L W D D H T A N M L W L D S T Y P T D S T K T G A S R G P C A V T S G V P K D V E S Q Y G S A Q V V Y S D I K F G A I N S T Y	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:138	146197095	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLALVYFLLS FVVSLEIGTQ QSEDHPPKLTW QNGSSSVSGS IVLDSNWRWV HDSGTTNCYD GNLWSTDLCG SS DTCSSKCY IEGADYSGTY GISASGSKLT LKFVTKGSYS TNIGSRVYLL KIDENTYETFK LKGKEFTFTV DDSKLD CGLN GALYFVAMDA DGGKAKYSSF KPGAKYMGY CDAQCPHDMK FISGKANVDD WKPQDN DENS GNGKLG TCCS EMDIWE GNAK SQAYTVHACT KSGQYECTGQ QCGD TDSGDR FKGTC DKDGC DYASWRWGDQ SFYGE GKTID TKQPVT VVTQ FIGDPLTEIR RVYVQGGKVI NNSKTSNLAN VYDSITDKFC DDTKDATGDT NDFKAKGAMS GFSTNLNTAQ VLVMSLWDDH TANMLWLDST YPTDSTKTGA SRGPCAVLSG VPKNVESQHG DATVIYS DIK FGAINSTFSY N	
SEQ ID NO:139	146197401	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MFLALFVLGK SLGIATNQAE NHPKLTWTRY QSKGSGQTVN GEVVLDSNWR WTHHSGTNCY DGNTWSTSLC PDPQTCSSNC DLDGADYPGT YGISSSGNSL KLG FVTHGSY STNIGSRVYL LRDSKNYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAME EDGGA KNSI NKAGA QYGTG YCDAQCPHDM KFINGEANVL DWKPQSN DEN SGNRGY GACC IEMDIWEANS MATAYTPHVC TVTGIHRC EG TECGDT DANQ RYNGICDKDG CDFNSYRMGD KSFFGVGKTV DSSKPVTVVT QFVTSNGQDG GTLSEIKRKY VQGGKVIENS KVN IAGITAV NSITDTFCNE QKKAFGDNND FEKKGGLGAL SKQLDLGMVL VLSLWDDHSV NMLWLDSTYP TDAAGALGT ERGACATSSG KPSDVESQSP DASVTFSDIK FGPIDSTY	
SEQ ID NO:140	146197225	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MLLCLLSIAN SLGVGTNTAE NHPKLSWKNG GSSVSGSVTV DANWRWTHIK GETKNCYDGN LWSDKYCPDA ATCGKNCVIE GADYQGTYG V SSGDGLTIT FVTHGQYSTN VGSRLYLMKD EKYQMFNLN GKEFTFTVDV SNLPCGLNG LYFVQMSDVG GMAKYPDNQA GAKYGTGYCD AQCP TDLKFI NGIPNSDGWK PQKNDKNSGN GKYGSCCSEM DIWEANSQAT AYTPHVCDKL EQTRCSGSSC GHTGGERFS SSCDPDGCDF NSWRMGNKTF WGPGLIVDTK KPVQVVTQFV GSGNSCTEIK RKYVQGGKVI DNSMSNIAGM SKQYNSVSD D FCQAQKKAFG DNDSFTKHGG FRQLGATLGK GHVLVLSLWD DHVNNMLWLD SVYPTNSNKP GS DRGPCKTS SGIPADVESQ AASSSVKYSD IRFGAIDSTY K	
SEQ ID NO:141	146197317	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MLCIGLISFV YSLGVGTNTA ETHPKLTWKN GGQTVNGEVT VDSNWRWTHT KGSTKNCYDG NLWSKDLCPD AATCGKNCVL EGADYSGTYG VTSSGNALT L K FVTHGSYST NVGSRLYLMK DEKTYQMFNL NGKEFTFTVD VSNLPCGLNG ALYHVNMD ED GGTKRYPDNE AGAKYGTGYC DAQCPTDLKF INGIPNSD GW KPQSN DKNSG NGKYGSCCSE MDIWEANSIC SAVTPHVCDT LQTRCQGT A CGENGGS RF GSSCDPDGCD FNSWRMGNKT FYGPGLIVDT KSKFTVVTQF VGSPVTEIKR KYVQNGKVIE NSF SNIEGMD KFNSISDKFC TAQKKAFGDT DSFTKHGGFK QLGSA LAKGM VLVL SLWDDH TVNMLWLD SV YPTNSKKAGS DRGPCPTTSG VPADVESKSA NANVIYSDIR FGAIDSTYK	
SEQ ID NO:142	146197251	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MLLCLLGIAS SLDAGTNTAE NHPQLSWKNG GSSVSGSVTV DANWRWTHIK GETKNCYDGN LWSDKYCPDA ATCGQNCVIE GADYQGTYG V SASGNALTIT FVTHGQYSTN VGSRLYLLKD EKYQIFNLI GKEFTFTVDV SNLPCGLNG LYFVQMDADG GTAKYSDNKA GAKYGTGYCD AQCP TDLKFI NGIPNSDGWK PQKNDKNSGN GRYGSCCSEM DVWEANSLAT AYTPHVCDKL AYTPHVCDKL EQVRCDGRAC GQNGGDRFS SSCDPDGCDF NSWRLGNKTF WGPGLIVDTK QPVQVVTQWV GSGTSVTEIK RKYVQGGKVI DNSFTKLDSL TKQYNSVSDE FCVAQKKAFG DNDSFTKHGG FRQLGATLAK GHVLVLSLWD DHVNNMLWLD SVYPTNSNKP GADRGPCCKTS SGVPADVESQ AASSSVKYSD IRFGAIDSTY K	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:143	146197319	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MLGIGFVCIV YSLGVGTNTA ENHPKLTWKN SGSTTNGEVT VDSNWRWTHT KGTTKNCYDG NLWSKDLCPD AATCGKNCVL EGADYSGTYG VTSSGDALT L KFTVTHGSYST NVGSRLYLLK DEKTYQIFNL NGKEFTFTVD VSNLPCGLNG ALYFVNMDAD GGTRYPDNQ AGAKYGTGYC DAQCPTDLKF INGIPNSDGW KPQSNDKNSG NGKYGSCCSE MDIWEANSLA TAVTPHVCDQ VGQTRCEGRA CGENGGGDRF GSSCDPDGCD FNSWRLGNKT FWGPGLIVDT KKPVTVVTFE VGSPVTEIKR KYVQGGKVIE NSYTNIEGLD KFNSISDKFC TAQKKAFGDN DSFIKHGGER QLQGSFTKGQ VLVLSLWDDH TVNMLWLDSV YPTNSKKPGA DRGPCPTSSG VPADVESKNA GSSVKYSDIR FGSIDSTYK	
SEQ ID NO:144	146197071	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MATLVGILVS LFALEVALEI GTQTSESHPS LSWEELNGQRQ TGSIVIDSNW RWLHDSGTTN CYDGENEWSSD LCPDPEKCSQ NCYLEGADYS GTYGISSGN SLQLGFVTKG SYSTNIGSRV YLLKDENTYA TFKLKNKEFT FTADVSNLPC GLNGALYFVA MPADGGKSKY PLAKPGA KYG MGYCDAQCPH DMKFINGEAN ILDWKPSSND ENAGAGRYGT CCTEMDIWEA NSQATAYTVH ACSKNARCEG TECGDDDDGRY NGICDKDGCD FNSWRWGNKT FFGPNLIVDS SKPVTVVTFE IGDPLTEIRR IYVQGGKVIQ NSFTNISGVA SVDSITDAFC NENKVATGDT NDFKAKGMS GFSAKALDTEV VLVLSLWDDH TANMLWLDST YPTDSSSALGA SRGPCAITSG EPKDVESASA NASVKFSDIK FGAIDSTY	
SEQ ID NO:145	146197075	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	MLTLVYFLLS LVVLEIGTQ QSESHPQLSW QNGSSSVSGS IVLDSNWRWV HDSGTTNCYD GNLWSTDLCP SSDTCTSKCY IEGADYSGTY GITSSGSKLT LKFVTKGSYS TNIGSRVYLL KDENTYETFK LKNKEFTFTV DDSKLDCCGLN GALYFVAMDA DGGKAKYSSF KPGA KYGMGY CDAQCPHDMK FISGKANVDD WKPQDNDENS GNGKLGTCSS EMDIWEAGNAK SQAYTVHACT KSGQYECTGQ QCAGTDSGDR FKGTCDKDG C DYASWRWGDQ SFYGEKGKTV TDQPLTVVVTQ FVGDP LTEIR RVYVQGGKTI NNSKTSNLAD TYDSITDKFC DATKEASGDT NDFKAKGAMS GFSTNLNNTAQ VLVMSLWDDH TANMLWLDST YPTDSTKTGA SRGPCAVSSG VPKDVESQHG DATVIYSDIK FGAINSTFKW N	
SEQ ID NO:146	146197159	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	MLSLVSIFLV GLGFSLGVGT QQSESHPSLS WQNC SAKGSC QSVSGSIVLD SNWRWLHDSG TTNCYD GNEW STDLCPDAST CDKN CYIEGA DYSGTYGITS SGAQLKLGFV TKGSYSTNIG SRVYLLRDES HYQLFKLKNH EFTFTVDDSQ LPCGLNGALY FVEM AEDGGA KPGAQYGMGY CDAQCPHDMK FITGEANVKD WKPQETDENA GNH YGACCT EMDIWEANSQ ATAYTPHICS KTGIYRCEGT ECGDNDANQR YNGVCDKDG C DFNSYRLGNK TFWGPGGLTVD SNKAMIVVTQ FTTSNNQDSG ELSEIRRIYV QGGKTIQNSD TNVQGITTN KITQAFCD E T KVTFGDTNDF KAKGGFSGLS KSLESGAVLV LSLWDDHDSVN MLWLDSTYPT DSAGKPGADR GPCAITSGDP KDVESQSPNA SVTFSDIKEG PIDSTY	
SEQ ID NO:147	146197405	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	MILALLVLGK SLGIATNQAE THPKLTWTRY QSKGSGSTVN GEIVLDSNWR WTHSGTNCY DGNTWSTSIC PDPTTCSNNC DLDGADYPGT YGISTSGNSL KLGFTVTHGSY STNIGSRVYL LKDTKS YEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPQSN DEN SGNGRYGACC TEMDIWEANS MATAYTPHVC TVTGLRRC EG TECGDTDNDQ RYNGICDKDG CDFNSYRLGD KSFFGVGKTV DSSKPVTVVT QFVTSNGQDS GTLSEIRRKY VQGGKVIENS KVNVA GITAG NSVTD TFCNE QKKAFGDNND FEKKGFGAL SKQLVAGMVL NMLWLDSTYP TNAAGALGT ERGACATSSG KPSDVESQSP DATVTFSDIK FGPIDSTY	

TABLE 1				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence	
SEQ ID NO:148	146197327	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	MLCVGLFGLV YSIGVGTNTQ ETHPKLSWKQ CSSGGSCCTQ QGSVVIDSNW RWTHSTKDLT NCYDGNLWDS TLCPDGTTCS KNCVLEGADY SGTYGITSSG DSLTLKFVTH GSYSTNVGSR LYLLKDDNNY QIFNLAGEF TFTVDVSNLP CGLNGALYFV EMDQDGGKGK HKENEAGAKY GTGYCDAQCP TDLKFIDGIA NSDGWKPQDN DENSGNGKYG SCCSEMDIWE ANSLATAYTP HVCDTKGQKR CQGTACGENG GGDRFGSECD PDGCDFNWR QGNKSFWGP G LIIDTKKSVQ VVTQFIGSGS SVTEIRRKYV QNGKVIENSY STISGTEKYN SISDDYCNAQ KKAFGDTNSF ENHGGFKRFS QHIQDMVLVL SLWDDHTVNM LWLDSVYPTN SNKPGADRG P CETSSGVPAD VESKSASASV KYSDIRFGPI DSTYK	
SEQ ID NO:149	146197261	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	MLLCLWSIAY SLGVGTNTAE NHPKLSWKNG GSSVSGSVTV DANWRWTHIK GETKNCYDGN LWSDKYCPDA ATCGKNCVIE GADYQGTYGV SASGDGLTLT FVTHGQYSTN VGSRLYLMKD EKYQIFNLN GKEFTFTVDV SNLPCGLNGA LYFVQMDS DG GMAKYPDNQA GAKYGTGYCD AQCPDCLKFI NGIPNSDGWK PQKNDKNSGN GKYGSCCSEM DIWEANSQAT AYTPHVCDKL EQTRCSGSAC GHTGGERFS SSCDPDGCDF NSWRMGNKTF WGPGLIVDTK KPVQVVTQFV GSGNSCTEIK RKYVQGGKVI DNSMSNIAGM TKQYNSVSDD FCQAQKKAFG DNDSFTKHGG FRQLGATLGK GHVLVLSLWD DHDVNMLWLD SVYPTNSNKP GS DRGPKCTS SGIPADVESQ AASSVKYSD IRFGAIDSTY K	

TABLE 2				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO:1	BD29555*	Unknown	273	422
SEQ ID NO:2	340514556	<i>Trichoderma reesei</i>	268	411
SEQ ID NO:3	51243029	<i>Penicillium occitanis</i>	273	422
SEQ ID NO:4	7cel (PDB) &	<i>Trichoderma reesei</i>	251	394
SEQ ID NO:5	67516425	<i>Aspergillus nidulans</i> FGSC A4	274	424
SEQ ID NO:6	46107376	<i>Gibberella zeae</i> PH-1	268	415
SEQ ID NO:7	70992391	<i>Aspergillus fumigatus</i> Af293	277	427
SEQ ID NO:8	121699984	<i>Aspergillus clavatus</i> NRRL 1	277	427
SEQ ID NO:9	1906845	<i>Claviceps purpurea</i>	269	416
SEQ ID NO:10	1gpi (PDB) &	<i>Phanerochaete chrysosporium</i>	240	391
SEQ ID NO:11	119468034	<i>Neosartorya fischeri</i> NRRL 181	265	414
SEQ ID NO:12	7804883	<i>Leptosphaeria maculans</i>	256	401
SEQ ID NO:13	85108032	<i>Neurospora crassa</i> N150	268	412
SEQ ID NO:14	169859458	<i>Coprinopsis cinerea</i> okayama	270	421
SEQ ID NO:15	154292161	<i>Botryotinia fuckeliana</i> B05-10	-	410
SEQ ID NO:16	169615761 #	<i>Phaeosphaeria nodorum</i> SN15	246	393
SEQ ID NO:17	4883502	<i>Humicola grisea</i>	272	413
SEQ ID NO:18	950686	<i>Humicola grisea</i>	270	416
SEQ ID NO:19	124491660	<i>Chaetomium thermophilum</i>	272	413
SEQ ID NO:20	58045187	<i>Chaetomium thermophilum</i>	270	416
SEQ ID NO:21	169601100 #	<i>Phaeosphaeria nodorum</i> SN15	237	383
SEQ ID NO:22	169870197	<i>Coprinopsis cinerea</i> okayama	269	421
SEQ ID NO:23	3913806	<i>Agaricus bisporus</i>	263	414
SEQ ID NO:24	169611094	<i>Phaeosphaeria nodorum</i> SN15	270	414
SEQ ID NO:25	3131	<i>Phanerochaete chrysosporium</i>	-	410
SEQ ID NO:26	70991503	<i>Aspergillus fumigatus</i> Af293	265	414
SEQ ID NO:27	294196	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO:28	18997123	<i>Thermoascus aurantiacus</i>	268	418
SEQ ID NO:29	4204214	<i>Humicola grisea</i> var <i>thermoidea</i>	272	413
SEQ ID NO:30	34582632	<i>Trichoderma viride</i> (also known as <i>Hypochrea rufa</i>)	268	411
SEQ ID NO:31	156712284	<i>Thermoascus aurantiacus</i>	268	418

TABLE 2				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO:32	39977899	<i>Magnaporthe grisea (oryzae) 70-15</i>	268	414
SEQ ID NO:33	20986705	<i>Talaromyces emersonii</i>	266	416
SEQ ID NO:34	22138843	<i>Aspergillus oryzae</i>	265	414
SEQ ID NO:35	55775695	<i>Penicillium chrysogenum</i>	276	426
SEQ ID NO:36	171676762	<i>Podospora anserina</i>	270	417
SEQ ID NO:37	146350520	<i>Pleurotus sp Florida</i>	268	420
SEQ ID NO:38	37732123	<i>Gibberella zeae</i>	268	415
SEQ ID NO:39	156055188	<i>Sclerotinia sclerotiorum 1980</i>	-	410
SEQ ID NO:40	453224	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO:41	50402144	<i>Trichoderma reesei</i>	268	411
SEQ ID NO:42	115397177	<i>Aspergillus terreus NIH2624</i>	274	424
SEQ ID NO:43	154312003	<i>Botryotinia fuckeliana B05-10</i>	266	416
SEQ ID NO:44	49333365	<i>Volvariella volvacea</i>	268	420
SEQ ID NO:45	729650	<i>Penicillium janthinellum</i>	274	424
SEQ ID NO:46	146424871	<i>Pleurotus sp Florida</i>	267	418
SEQ ID NO:47	67538012	<i>Aspergillus nidulans FGSC A4</i>	265	410
SEQ ID NO:48	62006162	<i>Fusarium poae</i>	268	415
SEQ ID NO:49	146424873	<i>Pleurotus sp Florida</i>	267	418
SEQ ID NO:50	295937	<i>Trichoderma viride</i>	268	411
SEQ ID NO:51	6179889 #	<i>Alternaria alternata</i>	240	386
SEQ ID NO:52	119483864	<i>Neosartorya fischeri NRRL 181</i>	278	428
SEQ ID NO:53	85083281	<i>Neurospora crassa OR74A</i>	270	412
SEQ ID NO:54	3913803	<i>Cryphonectria parasitica</i>	269	416
SEQ ID NO:55	60729633	<i>Corticium rolfsii</i>	265	415
SEQ ID NO:56	39971383	<i>Magnaporthe grisea 70-15</i>	268	410
SEQ ID NO:57	39973029	<i>Magnaporthe grisea 70-15</i>	269	410
SEQ ID NO:58	1170141	<i>Fusarium oxysporum</i>	268	415
SEQ ID NO:59	121710012	<i>Aspergillus clavatus NRRL 1</i>	265	414
SEQ ID NO:60	17902580	<i>Penicillium funiculosum</i>	273	422
SEQ ID NO:61	1346226	<i>Humicola grisea var thermoidea</i>	270	416
SEQ ID NO:62	156712282	<i>Chaetomium thermophilum</i>	270	416

TABLE 2				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO:63	169768818	<i>Aspergillus oryzae</i> RIB40	277	427
SEQ ID NO:64	46241270	<i>Gibberella pulicaris</i>	268	415
SEQ ID NO:65	49333363	<i>Volvariella volvacea</i>	265	418
SEQ ID NO:66	46395332	<i>Irpex lacteus</i>	263	414
SEQ ID NO:67	50844407 #	<i>Chaetomium thermophilum</i> var <i>thermophilum</i>	245	391
SEQ ID NO:68	4586347	<i>Irpex lacteus</i>	264	415
SEQ ID NO:69	3980202	<i>Phanerochaete chrysosporium</i>	258	410
SEQ ID NO:70	27125837	<i>Melanocarpus albomyces</i>	273	414
SEQ ID NO:71	171696102	<i>Podospora anserina</i>	265	415
SEQ ID NO:72	3913802	<i>Cochliobolus carbonum</i>	270	416
SEQ ID NO:73	50403723	<i>Trichoderma viride</i>	268	411
SEQ ID NO:74	3913798	<i>Aspergillus aculeatus</i>	275	425
SEQ ID NO:75	66828465	<i>Dictyostelium discoideum</i>	269	419
SEQ ID NO:76	156060391	<i>Sclerotinia sclerotiorum</i> 1980	252	402
SEQ ID NO:77	116181754	<i>Chaetomium globosum</i> CBS 148-51	263	413
SEQ ID NO:78	145230535	<i>Aspergillus niger</i>	274	424
SEQ ID NO:79	46241266	<i>Nectria haematococca</i> mpVI	268	415
SEQ ID NO:80	1q9h (PDB) #	<i>Talaromyces emersonii</i>	248	398
SEQ ID NO:81	157362170	<i>Polyporus arcularius</i>	269	420
SEQ ID NO:82	7804885	<i>Leptosphaeria maculans</i>	267	407
SEQ ID NO:83	121852	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO:84	126013214	<i>Penicillium decumbens</i>	264	415
SEQ ID NO:85	156048578	<i>Sclerotinia sclerotiorum</i> 1980	265	413
SEQ ID NO:86	156712278	<i>Acremonium thermophilum</i>	269	414
SEQ ID NO:87	21449327	<i>Aspergillus nidulans</i>	265	410
SEQ ID NO:88	171683762	<i>Podospora anserina</i>	274	415
SEQ ID NO:89	56718412	<i>Thermoascus aurantiacus</i> var <i>levisporus</i>	268	418
SEQ ID NO:90	15824273	<i>Pseudotriconympha grassii</i>	263	414
SEQ ID NO:91	115390801	<i>Aspergillus terreus</i> NIH2624	266	411
SEQ ID NO:92	453223	<i>Phanerochaete chrysosporium</i>	258	409

TABLE 2				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO:93	3132	<i>Phanerochaete chrysosporium</i>	-	407
SEQ ID NO:94	16304152	<i>Thermoascus aurantiacus</i>	268	417
SEQ ID NO:95	156712280	<i>Acremonium thermophilum</i>	273	420
SEQ ID NO:96	5231154	<i>Volvariella volvacea</i>	281	438
SEQ ID NO:97	116200349	<i>Chaetomium globosum</i> CBS 148-51	270	412
SEQ ID NO:98	4586343	<i>Irpex lacteus</i>	263	414
SEQ ID NO:99	15321718	<i>Lentinula edodes</i>	-	417
SEQ ID NO:100	146424875	<i>Pleurotus</i> sp Florida	267	418
SEQ ID NO:101	62006158	<i>Fusarium venenatum</i>	268	415
SEQ ID NO:102	296027	<i>Phanerochaete chrysosporium</i>	258	409
SEQ ID NO:103	154449709	<i>Fusicoccum</i> sp BCC4124	272	424
SEQ ID NO:104	169859460	<i>Coprinopsis cinerea</i> okayama	269	421
SEQ ID NO:105	50400675	<i>Trichoderma harzianum</i>	264	407
SEQ ID NO:106	729649	<i>Neurospora crassa</i>	262	406
SEQ ID NO:107	119472134	<i>Neosartorya fischeri</i> NRRL 181	277	427
SEQ ID NO:108	117935080	<i>Chaetomium thermophilum</i>	272	413
SEQ ID NO:109	154300584	<i>Botryotinia fuckeliana</i> B05-10	265	413
SEQ ID NO:110	15824271	<i>Pseudotriconympha grassii</i>	263	414
SEQ ID NO:111	4586345	<i>Irpex lacteus</i>	263	414
SEQ ID NO:112	46241268	<i>Gibberella avenacea</i>	268	416
SEQ ID NO:113	6164684	<i>Aspergillus niger</i>	274	424
SEQ ID NO:114	6164682	<i>Aspergillus niger</i>	266	412
SEQ ID NO:115	33733371	<i>Chrysosporium lucknowense</i> US6573086-10	269	415
SEQ ID NO:116	29160311	<i>Thielavia australiensis</i>	269	415
SEQ ID NO:117	146197087	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO:118	146197237	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	264	409
SEQ ID NO:119	146197067	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO:120	146197407	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412

TABLE 2				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO:121	146197157	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	264	410
SEQ ID NO:122	146197403	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO:123	146197081	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	410
SEQ ID NO:124	146197413	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO:125	146197309	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO:126	146197227	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	258	404
SEQ ID NO:127	146197253	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	264	409
SEQ ID NO:128	146197099	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	258	401
SEQ ID NO:129	146197409	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	260	411
SEQ ID NO:130	146197315	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO:131	146197411	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO:132	146197161	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	263	413
SEQ ID NO:133	146197323	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO:134	146197077	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	264	415
SEQ ID NO:135	146197089	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	258	400
SEQ ID NO:136	146197091	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	258	401
SEQ ID NO:137	146197097	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO:138	146197095	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO:139	146197401	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO:140	146197225	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	258	404

TABLE 2				
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Position corresponding to position 268	Position corresponding to position 411
SEQ ID NO:141	146197317	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO:142	146197251	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	258	404
SEQ ID NO:143	146197319	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	259	402
SEQ ID NO:144	146197071	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	259	402
SEQ ID NO:145	146197075	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	260	402
SEQ ID NO:146	146197159	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	260	410
SEQ ID NO:147	146197405	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	261	412
SEQ ID NO:148	146197327	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	264	408
SEQ ID NO:149	146197261	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	258	404

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:1	BD29555*	Unknown	1-25	26-455	456-493	494-529
SEQ ID NO:2	340514556	<i>Trichoderma reesei</i>	1-17	18-444	445-479	480-514
SEQ ID NO:3	51243029	<i>Penicillium occitanis</i>	1-25	26-455	456-493	494-529
SEQ ID NO:4	7cel (PDB) &	<i>Trichoderma reesei</i>	N/A	1-427	N/A	N/A
SEQ ID NO:5	67516425	<i>Aspergillus nidulans</i> FGSC A4	1-23	24-457	458-490	491-526
SEQ ID NO:6	46107376	<i>Gibberella zeae</i> PH-1	1-17	18-448	449-476	477-512
SEQ ID NO:7	70992391	<i>Aspergillus fumigatus</i> Af293	1-26	27-460	461-496	497-532
SEQ ID NO:8	121699984	<i>Aspergillus clavatus</i> NRRL 1	1-27	27-460	461-503	504-539
SEQ ID NO:9	1906845	<i>Claviceps purpurea</i>	1-19	20-449	N/A	N/A
SEQ ID NO:10	1gpi (PDB) &	<i>Phanerochaete chrysosporium</i>	N/A	1-424	N/A	N/A
SEQ ID NO:11	119468034	<i>Neosartorya fischeri</i> NRRL 181	1-17	18-447	N/A	N/A
SEQ ID NO:12	7804883	<i>Leptosphaeria maculans</i>	1-17	18-434	N/A	N/A
SEQ ID NO:13	85108032	<i>Neurospora crassa</i> N150	1-17	18-445	446-485	486-521
SEQ ID NO:14	169859458	<i>Coprinopsis cinerea</i> okayama	1-18	19-454	N/A	N/A
SEQ ID NO:15	154292161	<i>Botryotinia fuckeliana</i> B05-10	1-18	19-443	444-555	556-596
SEQ ID NO:16	169615761 #	<i>Phaeosphaeria nodorum</i> SN15	1	2-426	N/A	N/A
SEQ ID NO:17	4883502	<i>Humicola grisea</i>	1-22	23-446	N/A	N/A
SEQ ID NO:18	950686	<i>Humicola grisea</i>	1-18	19-449	450-489	490-525
SEQ ID NO:19	124491660	<i>Chaetomium thermophilum</i>	1-22	23-446	N/A	N/A
SEQ ID NO:20	58045187	<i>Chaetomium thermophilum</i>	1-18	19-449	450-494	495-530
SEQ ID NO:21	169601100 #	<i>Phaeosphaeria nodorum</i> SN15	1	2-416	N/A	N/A

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:22	169870197	<i>Coprinopsis cinerea okayama</i>	1-18	19-454	N/A	N/A
SEQ ID NO:23	3913806	<i>Agaricus bisporus</i>	1-18	19-447	448-470	471-506
SEQ ID NO:24	169611094	<i>Phaeosphaeria nodorum</i> SN15	1-18	19-447	N/A	N/A
SEQ ID NO:25	3131	<i>Phanerochaete chrysosporium</i>	1-19	20-443	N/A	N/A
SEQ ID NO:26	70991503	<i>Aspergillus fumigatus</i> Af293	1-17	18-447	N/A	N/A
SEQ ID NO:27	294196	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-480	481-516
SEQ ID NO:28	18997123	<i>Thermoascus aurantiacus</i>	1-17	18-451	N/A	N/A
SEQ ID NO:29	4204214	<i>Humicola grisea</i> var <i>thermoidea</i>	1-22	23-446	N/A	N/A
SEQ ID NO:30	34582632	<i>Trichoderma viride</i> (also known as <i>Hypochrea rufa</i>)	1-18	18-444	445-479	480-514
SEQ ID NO:31	156712284	<i>Thermoascus aurantiacus</i>	1-17	18-451	N/A	N/A
SEQ ID NO:32	39977899	<i>Magnaporthe grisea</i> (oryzae) 70-15	1-17	18-447	N/A	N/A
SEQ ID NO:33	20986705	<i>Talaromyces emersonii</i>	1-18	19-449	N/A	N/A
SEQ ID NO:34	22138843	<i>Aspergillus oryzae</i>	1-17	18-447	N/A	N/A
SEQ ID NO:35	55775695	<i>Penicillium chrysogenum</i>	1-25	26-459	460-494	495-529
SEQ ID NO:36	171676762	<i>Podospira anserina</i>	1-18	19-450	451-492	493-528
SEQ ID NO:37	146350520	<i>Pleurotus</i> sp <i>Florida</i>	1-18	19-453	N/A	N/A
SEQ ID NO:38	37732123	<i>Gibberella zeae</i>	1-17	18-448	449-476	477-512
SEQ ID NO:39	156055188	<i>Sclerotinia sclerotiorum</i> 1980	1-18	19-443	444-546	547-586
SEQ ID NO:40	453224	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-474	475-510
SEQ ID NO:41	50402144	<i>Trichoderma reesei</i>	1-17	18-444	445-478	479-513
SEQ ID NO:42	115397177	<i>Aspergillus terreus</i> NIH2624	1-23	24-457	458-505	506-541

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:43	154312003	<i>Botryotinia fuckeliana</i> B05-10	1-17	18-449	450-480	481-516
SEQ ID NO:44	49333365	<i>Volvariella volvacea</i>	1-18	19-453	N/A	N/A
SEQ ID NO:45	729650	<i>Penicillium janthinellum</i>	1-25	26-456	457-502	503-537
SEQ ID NO:46	146424871	<i>Pleurotus sp Florida</i>	1-18	19-451	452-487	488-523
SEQ ID NO:47	67538012	<i>Aspergillus nidulans</i> FGSC A4	1-17	18-443	N/A	N/A
SEQ ID NO:48	62006162	<i>Fusarium poae</i>	1-17	18-448	449-475	476-511
SEQ ID NO:49	146424873	<i>Pleurotus sp Florida</i>	1-18	19-451	452-487	488-523
SEQ ID NO:50	295937	<i>Trichoderma viride</i>	1-17	18-444	445-478	479-513
SEQ ID NO:51	6179889 #	<i>Alternaria alternata</i>	1	2-419	N/A	N/A
SEQ ID NO:52	119483864	<i>Neosartorya fischeri</i> NRRL 181	1-26	27-461	462-499	500-535
SEQ ID NO:53	85083281	<i>Neurospora crassa</i> OR74A	1-20	21-445	N/A	N/A
SEQ ID NO:54	3913803	<i>Cryphonectria parasitica</i>	1-18	19-449	N/A	N/A
SEQ ID NO:55	60729633	<i>Corticium rolszii</i>	1-18	19-448	449-492	493-528
SEQ ID NO:56	39971383	<i>Magnaporthe grisea</i> 70-15	1-17	18-443	N/A	N/A
SEQ ID NO:57	39973029	<i>Magnaporthe grisea</i> 70-15	1-19	20-443	N/A	N/A
SEQ ID NO:58	1170141	<i>Fusarium oxysporum</i>	1-17	18-448	449-478	479-514
SEQ ID NO:59	121710012	<i>Aspergillus clavatus</i> NRRL 1	1-17	18-447	N/A	N/A
SEQ ID NO:60	17902580	<i>Penicillium funiculosum</i>	1-25	26-455	456-493	494-529
SEQ ID NO:61	1346226	<i>Humicola grisea var thermoidea</i>	1-18	19-449	450-489	490-525
SEQ ID NO:62	156712282	<i>Chaetomium thermophilum</i>	1-18	19-449	450-496	497-532
SEQ ID NO:63	169768818	<i>Aspergillus oryzae</i> RIB40	1-25	26-460	N/A	N/A
SEQ ID NO:64	46241270	<i>Gibberella pulicaris</i>	1-17	18-448	449-474	475-510
SEQ ID NO:65	49333363	<i>Volvariella volvacea</i>	1-18	19-451	452-476	477-512

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:66	46395332	<i>Irpex lacteus</i>	1-18	19-447	448-485	486-521
SEQ ID NO:67	50844407 #	<i>Chaetomium thermophilum</i> var <i>thermophilum</i>	N/A	1-424	425-469	470-505
SEQ ID NO:68	4586347	<i>Irpex lacteus</i>	1-18	19-448	449-490	491-526
SEQ ID NO:69	3980202	<i>Phanerochaete chrysosporium</i>	1-18	19-443	444-475	476-511
SEQ ID NO:70	27125837	<i>Melanocarpus albomyces</i>	1-23	23-447	N/A	N/A
SEQ ID NO:71	171696102	<i>Podospora anserina</i>	1-17	17-448	N/A	N/A
SEQ ID NO:72	3913802	<i>Cochliobolus carbonum</i>	1-18	19-449	N/A	N/A
SEQ ID NO:73	50403723	<i>Trichoderma viride</i>	1-17	18-444	445-479	480-514
SEQ ID NO:74	3913798	<i>Aspergillus aculeatus</i>	1-22	23-458	459-505	506-540
SEQ ID NO:75	66828465	<i>Dictyostelium discoideum</i>	1-19	20-452	N/A	N/A
SEQ ID NO:76	156060391	<i>Sclerotinia sclerotiorum</i> 1980	1-17	18-435	436-470	471-504
SEQ ID NO:77	116181754	<i>Chaetomium globosum</i> CBS 148-51	1-17	18-446	N/A	N/A
SEQ ID NO:78	145230535	<i>Aspergillus niger</i>	1-21	22-457	458-500	501-536
SEQ ID NO:79	46241266	<i>Nectria haematococca</i> mpVI	1-18	18-448	449-472	473-508
SEQ ID NO:80	1q9h (PDB) #	<i>Talaromyces emersonii</i>	N/A	1-431	N/A	N/A
SEQ ID NO:81	157362170	<i>Polyporus arcularius</i>	1-18	19-453	N/A	N/A
SEQ ID NO:82	7804885	<i>Leptosphaeria maculans</i>	1-20	21-440	N/A	N/A
SEQ ID NO:83	121852	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-480	481-516
SEQ ID NO:84	126013214	<i>Penicillium decumbens</i>	1-17	18-448	N/A	N/A
SEQ ID NO:85	156048578	<i>Sclerotinia sclerotiorum</i> 1980	1-16	17-446	N/A	N/A
SEQ ID NO:86	156712278	<i>Acremonium thermophilum</i>	1-17	18-447	448-487	488-523

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:87	21449327	<i>Aspergillus nidulans</i>	1-17	18-443	N/A	N/A
SEQ ID NO:88	171683762	<i>Podospora anserina</i>	1-22	23-448	N/A	N/A
SEQ ID NO:89	56718412	<i>Thermoascus aurantiacus</i> var <i>levisporus</i>	1-17	18-451	N/A	N/A
SEQ ID NO:90	15824273	<i>Pseudotriconympha grassii</i>	1-20	21-447	N/A	N/A
SEQ ID NO:91	115390801	<i>Aspergillus terreus</i> NIH2624	1-17	18-444	N/A	N/A
SEQ ID NO:92	453223	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-474	475-510
SEQ ID NO:93	3132	<i>Phanerochaete chrysosporium</i>	1-19	20-436	437-467	468-504
SEQ ID NO:94	16304152	<i>Thermoascus aurantiacus</i>	1-17	18-450	N/A	N/A
SEQ ID NO:95	156712280	<i>Acremonium thermophilum</i>	1-21	22-453	N/A	N/A
SEQ ID NO:96	5231154	<i>Volvariella volvacea</i>	1-15	16-472	473-500	501-536
SEQ ID NO:97	116200349	<i>Chaetomium globosum</i> CBS 148-51	1-20	21-445	N/A	N/A
SEQ ID NO:98	4586343	<i>Irpex lacteus</i>	1-18	19-447	448-481	482-517
SEQ ID NO:99	15321718	<i>Lentinula edodes</i>	1-18	19-450	451-480	481-516
SEQ ID NO:100	146424875	<i>Pleurotus</i> sp <i>Florida</i>	1-18	19-451	452-487	488-523
SEQ ID NO:101	62006158	<i>Fusarium venenatum</i>	1-17	18-448	449-471	472-507
SEQ ID NO:102	296027	<i>Phanerochaete chrysosporium</i>	1-18	19-442	443-480	481-516
SEQ ID NO:103	154449709	<i>Fusicoccum</i> sp BCC4124	1-19	20-457	N/A	N/A
SEQ ID NO:104	169859460	<i>Coprinopsis cinerea</i> okayama	1-18	19-454	N/A	N/A
SEQ ID NO:105	50400675	<i>Trichoderma harzianum</i>	1-17	18-440	441-470	471-505
SEQ ID NO:106	729649	<i>Neurospora crassa</i>	1-17	18-439	440-480	481-516
SEQ ID NO:107	119472134	<i>Neosartorya fischeri</i> NRRL 181	1-26	27-460	461-494	495-530
SEQ ID NO:108	117935080	<i>Chaetomium thermophilum</i>	1-22	23-446	N/A	N/A

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:109	154300584	<i>Botryotinia fuckeliana</i> B05-10	1-16	17-446	N/A	N/A
SEQ ID NO:110	15824271	<i>Pseudotriconympha grassii</i>	1-20	21-447	N/A	N/A
SEQ ID NO:111	4586345	<i>Irpex lacteus</i>	1-18	19-447	448-487	488-523
SEQ ID NO:112	46241268	<i>Gibberella avenacea</i>	1-17	18-449	450-478	478-513
SEQ ID NO:113	6164684	<i>Aspergillus niger</i>	1-21	22-457	458-500	501-536
SEQ ID NO:114	6164682	<i>Aspergillus niger</i>	1-17	18-445	N/A	N/A
SEQ ID NO:115	33733371	<i>Chrysosporium lucknowense</i> US6573086-10	1-17	18-448	449-490	491-526
SEQ ID NO:116	29160311	<i>Thielavia australiensis</i>	1-18	18-448	449-502	503-538
SEQ ID NO:117	146197087	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO:118	146197237	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-20	21-442	N/A	N/A
SEQ ID NO:119	146197067	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO:120	146197407	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO:121	146197157	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	1-20	21-443	N/A	N/A
SEQ ID NO:122	146197403	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO:123	146197081	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-443	N/A	N/A

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:124	146197413	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO:125	146197309	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO:126	146197227	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-19	20-437	N/A	N/A
SEQ ID NO:127	146197253	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-21	21-442	N/A	N/A
SEQ ID NO:128	146197099	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-434	N/A	N/A
SEQ ID NO:129	146197409	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-444	N/A	N/A
SEQ ID NO:130	146197315	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO:131	146197411	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO:132	146197161	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	1-20	21-446	N/A	N/A
SEQ ID NO:133	146197323	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO:134	146197077	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-21	22-448	N/A	N/A

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:135	146197089	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-433	N/A	N/A
SEQ ID NO:136	146197091	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-434	N/A	N/A
SEQ ID NO:137	146197097	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO:138	146197095	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A
SEQ ID NO:139	146197401	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO:140	146197225	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-19	20-437	N/A	N/A
SEQ ID NO:141	146197317	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO:142	146197251	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-19	20-437	N/A	N/A
SEQ ID NO:143	146197319	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-435	N/A	N/A
SEQ ID NO:144	146197071	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-25	26-435	N/A	N/A
SEQ ID NO:145	146197075	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	1-22	23-435	N/A	N/A

TABLE 3						
SEQ ID NO:	Database Accession Number	Species of Origin	Signal sequence (SS) start and end position	Catalytic Domain (CD) start and end position	Linker start and end position	Cellulose Binding Domain (CBD) start and end
SEQ ID NO:146	146197159	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	1-23	24-443	N/A	N/A
SEQ ID NO:147	146197405	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	1-19	20-445	N/A	N/A
SEQ ID NO:148	146197327	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	1-20	21-441	N/A	N/A
SEQ ID NO:149	146197261	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	1-19	20-437	N/A	N/A

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:150	BD29555*	Unknown	NVEGWTPSSNNANTGLGNHGACCAELDIWEANS	210-242	214-226	234, 239
SEQ ID NO:151	340514556	<i>Trichoderma reesei</i>	NVEGWTPSANNANTGIGNHGACCAELDIWEANS	205-237	209-221	229, 234
SEQ ID NO:152	51243029	<i>Penicillium occitanis</i>	NVEGWEPSNNANTGIGGHGSCCSEMDIWEANS	210-242	214-226	234, 239
SEQ ID NO:153	7cel (PDB) &	<i>Trichoderma reesei</i>	NVEGWEPSNNANTGIGGHGSCCSEMDIWEANS	188-220	192-204	212, 217
SEQ ID NO:154	67516425	<i>Aspergillus nidulans</i> FGSC A4	NVEGWESDTPNGGVGNHGSCCAEMDIWEANS	211-243	215-227	235, 240
SEQ ID NO:155	46107376	<i>Gibberella zeae</i> PH-1	NSDGWQPSDSDVNGGIGNLGTCPCPEMDIWEANS	205-237	209-221	229, 234
SEQ ID NO:156	70992391	<i>Aspergillus fumigatus</i> Af293	NVEGWQPSNDANAGTGNHGSCCAEMDIWEANS	214-246	218-230	238, 243
SEQ ID NO:157	121699984	<i>Aspergillus clavatus</i> NRRL 1	NVEGWTPSSSDANAGNGGHGSCCAEMDIWEANS	214-246	218-230	238, 243
SEQ ID NO:158	1906845	<i>Claviceps purpurea</i>	NSKDWIPSKSDANAGIGSLGACCCREMDIWEANN	206-238	210-222	230, 235
SEQ ID NO:159	Igpi (PDB) &	<i>Phanerochaete chrysosporium</i>	NVGNWTETG---SNTGTGSYGTCSCSEMDIWEANN	185-215	189-199	207, 212
SEQ ID NO:160	119468034	<i>Neosartorya fischeri</i> NRRL 181	NVEGWKPSSNDKNAGVGGHGSCCPEMDIWEANS	202-234	206-218	226, 231
SEQ ID NO:161	7804883	<i>Leptosphaeria maculans</i>	NVEGWQPSKNDQONAGVGGHGSCCAEMDIWEANS	193-225	197-209	217, 222
SEQ ID NO:162	85108032	<i>Neurospora crassa</i> N150 (OR74A)	NVEGWTPSTNDANAGIGDHGTCSCSEMDIWEANK	205-237	209-221	229, 234

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:163	169859458	<i>Coprinopsis cinerea okayama</i>	NSADWTPSETDPNAGRGRYGCCAEMDIWEANS	207-239	211-223	231, 236
SEQ ID NO:164	154292161	<i>Botryotinia fuckeliana B05-10</i>	NVEGWVPSNSANS ~~GTG~~ NI ~~GS~~ CCSE ~~FDV~~ WEANS	203-235	207-219	227, 232
SEQ ID NO:165	169615761 #	<i>Phaeosphaeria nodorum</i> SN15	NADGWQASTSDPNAGVGKKGACCAEMDVWEANS	183-215	187-199	207, 212
SEQ ID NO:166	4883502	<i>Humicola grisea</i>	NIEGWRPSTNDPNAGVGPMGACCAEIDVWE ~~S~~ NA	208-240	212-224	232, 237
SEQ ID NO:167	950686	<i>Humicola grisea</i>	NIEGWTGSTNDPNAGAGRYGTCCSEMDIWE ~~A~~ NN	207-239	211-223	231, 236
SEQ ID NO:168	124491660	<i>Chaetomium thermophilum</i>	NIEGWRPSTNDANAGVGPPYGACCAEIDVWE ~~S~~ NA	209-241	213-225	233, 238
SEQ ID NO:169	58045187	<i>Chaetomium thermophilum</i>	NIENWTPSTNDANAGFGRYGSCCSEMDIWE ~~A~~ NN	207-239	211-223	231, 236
SEQ ID NO:170	169601100 #	<i>Phaeosphaeria nodorum</i> SN15	NVEGWKPSDNDANAGVGGHGSCCAEMDIWEANS	174-206	178-190	198, 203
SEQ ID NO:171	169870197	<i>Coprinopsis cinerea okayama</i>	NSVGWEPSETDSNAGRGRYGCCAEMDIWEANS	207-239	211-223	231, 236
SEQ ID NO:172	3913806	<i>Agaricus bisporus</i>	NSEGWE ~~G~~ SPNDVNAGTGNFGACCGEMDIWEANS	203-235	207-219	227, 232
SEQ ID NO:173	169611094	<i>Phaeosphaeria nodorum</i> SN15	NVEGWNPSDADPNAGSGKIGACCP ~~EM~~ DIWEANS	208-240	212-224	232, 237
SEQ ID NO:174	3131	<i>Phanerochaete chrysosporium</i>	NVQGN ~~N~~ ATS--ATTGTGSYGSCCTE ~~L~~ DIWEANS	204-234	208-218	226, 231
SEQ ID NO:175	70991503	<i>Aspergillus fumigatus</i> A ~~7~~ 93	NVEGWE ~~P~~ SSSDKNAGVGGHGSCCP ~~EM~~ DIWEANS	202-234	206-218	226, 231
SEQ ID NO:176	294196	<i>Phanerochaete chrysosporium</i>	NVEGWN ~~N~~ ATS--ANAGTGNYGTCCTE ~~M~~ DIWEANN	203-233	207-217	225, 230

TABLE 4						
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:177	18997123	<i>Thermoascus aurantiacus</i>	NVEGWQPSANDPNAGVGNHGSSCAEIMDVWEANS	205-237	209-221	229, 234
SEQ ID NO:178	4204214	<i>Humicola grisea</i> var <i>thermoidea</i>	NIEGWRPSTNDPNAGVGPAGACCAEIDVWEANA	208-240	212-224	232, 237
SEQ ID NO:179	34582632	<i>Trichoderma viride</i> (also known as <i>Hypochrea rufa</i>)	NVEGWEPSNNANTGIGGHGSCCSEMDIWEANS	205-237	209-221	229, 234
SEQ ID NO:180	156712284	<i>Thermoascus aurantiacus</i>	NVEGWQPSANDPNAGVGNHGSSCAEIMDVWEANS	205-237	209-221	229, 234
SEQ ID NO:181	39977899	<i>Magnaporthe grisea</i> (oryzae) 70-15	NVEGWQPSGDNAGVGNMGSCCAEIMDIWEANS	205-237	209-221	229, 234
SEQ ID NO:182	20986705	<i>Talaromyces emersonii</i>	NVEGWQPSNNANTGIGDHGSCCAEIMDVWEANS	203-235	207-219	227, 232
SEQ ID NO:183	22138843	<i>Aspergillus oryzae</i>	R-KGWEPSDSKDNAGVGGHGSCCPQMDIWEANS	203-234	206-218	226, 231
SEQ ID NO:184	55775695	<i>Penicillium chrysogenum</i>	NVEGWEPSSSDVNGGTGNYGSCCAEIMDIWEANS	213-245	217-229	237, 242
SEQ ID NO:185	171676762	<i>Podospora anserina</i>	NIEGWNPTNDVNAGAGRYGTCCSEMDIWEANN	207-239	211-223	231, 236
SEQ ID NO:186	146350520	<i>Pleurotus sp</i> Florida	NVQGWQPSPNDSNAGKGQYGSCCAEIMDIWEANS	207-239	211-223	231, 236
SEQ ID NO:187	37732123	<i>Gibberella zeae</i>	NSDCWQPSDSVNGGIGNLGTCCPEMDIWEANS	205-237	209-221	229, 234
SEQ ID NO:188	156055188	<i>Sclerotinia sclerotiorum</i> 1980	NNEGWVPSNSANSNGTGNIGSCCSEFVDWEANS	203-235	207-219	227, 232
SEQ ID NO:189	453224	<i>Phanerochaete chrysosporium</i>	NVGNWTETG--SNTGTGSYGTCCSEMDIWEANN	203-233	207-217	225, 230

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:190	50402144	<i>Trichoderma reesei</i>	NVEGWEPSNNANTGIGGHGSCCSEMDIWEANS	205-237	209-221	229, 234
SEQ ID NO:191	115397177	<i>Aspergillus terreus</i> NIH2624	NVEGWEPSANDANAGTGNHGSCCAEMDIWEANS	211-243	215-227	235, 240
SEQ ID NO:192	154312003	<i>Botryotinia fuckeliana</i> B05-10	NSVGWTFSSNDVNAGAGQYGSCCSEMDIWEANK	206-238	210-222	230, 235
SEQ ID NO:193	49333365	<i>Volvariella volvacea</i>	NVQGWQPSNDTNAGTGNYGACCNEMDVWEANS	207-239	211-223	231, 236
SEQ ID NO:194	729650	<i>Penicillium janthinellum</i>	NVDGWTPSKNDVNSGIGNHGSCCAEMDIWEANS	211-243	215-227	235, 240
SEQ ID NO:195	146424871	<i>Pleurotus sp</i> Florida	NILDWSASATDANAGNGRYGACCAEMDIWEANS	206-238	210-222	230, 235
SEQ ID NO:196	67538012	<i>Aspergillus nidulans</i> FGSC A4	NVEGWEPSDSDANAGVGGMGTCCPEMDIWEANS	202-234	206-218	226, 231
SEQ ID NO:197	62006162	<i>Fusarium poae</i>	NSDGWEPSKSDVNGGIGNLGTCCPEMDIWEANS	205-237	209-221	229, 234
SEQ ID NO:198	146424873	<i>Pleurotus sp</i> Florida	NILDWSGSATDPNAGNGRYGACCAEMDIWEANS	206-238	210-222	230, 235
SEQ ID NO:199	295937	<i>Trichoderma viride</i>	NVEGWEPSNNANTGIGGHGSCCSEMDIWEANS	205-237	209-221	229, 234
SEQ ID NO:200	6179889 #	<i>Alternaria alternata</i>	NVEGWKPSSNDANAGVGGHGSCCAEMDIWEANS	177-209	181-193	201, 206
SEQ ID NO:201	119483864	<i>Neosartorya fischeri</i> NRRL 181	NVEGWTPSSNNENTGLGNYGSCCAELDIWEANS	215-247	219-231	239, 244
SEQ ID NO:202	85083281	<i>Neurospora crassa</i> OR74A	NIEGWTPSTNDANAGVGPYGGCCAEIDVWESNA	207-239	211-223	231, 236

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:203	3913803	<i>Cryphonectria parasitica</i>	NVEGWTPTNDANAGVGLGCCSEMDVWEANS	206-238	210-222	230, 235
SEQ ID NO:204	60729633	<i>Corticium rolfii</i>	NLLDWNATS--ANSGTGSYGSCCP ~~EMDIWE~~ ANK	206-236	210-220	228, 233
SEQ ID NO:205	39971383	<i>Magnaporthe grisea</i> 70-15	NIEGWQPSSTDSSAGIGAQGACCAEID ~~IWE~~ SNK	205-237	209-221	229, 234
SEQ ID NO:206	39973029	<i>Magnaporthe grisea</i> 70-15	NIEGWKPPSSNDANAGVGPYGACCAEIDV ~~WE~~ SNA	206-238	210-222	230, 235
SEQ ID NO:207	1170141	<i>Fusarium oxysporum</i>	NSEGWKPPSDSDVNAGVGNLGTCCP ~~EMDIWE~~ ANS	205-237	209-221	229, 234
SEQ ID NO:208	121710012	<i>Aspergillus clavatus</i> NRRL 1	NVEGWKPPSDNKNAGVGGYGSCCP ~~EMDIWE~~ ANS	202-234	206-218	226, 231
SEQ ID NO:209	17902580	<i>Penicillium funiculosum</i>	NVEGWTPTSTNNSNTGIGNHGSCCAE ~~LDIWE~~ ANS	210-242	214-226	234, 239
SEQ ID NO:210	1346226	<i>Humicola grisea</i> var <i>thermoidea</i>	NIEGWTGSTNDPNAGAGRYGTCCSEMDI ~~WE~~ ANN	207-239	211-223	231, 236
SEQ ID NO:211	156712282	<i>Chaetomium thermophilum</i>	NVGNWTPTSTNDANAGFGRYGSCCSEMDV ~~WE~~ ANN	207-239	211-223	231, 236
SEQ ID NO:212	169768818	<i>Aspergillus oryzae</i> RIB40	NVEGWVSSSTNNANTGTGNHGSCCAE ~~LDIWE~~ SNS	214-246	218-230	238, 243
SEQ ID NO:213	46241270	<i>Gibberella pulicaris</i>	NSDGWQPSKSDVNAGIGNMGTC ~~CP~~ EMDI ~~WE~~ ANS	205-237	209-221	229, 234
SEQ ID NO:214	49333363	<i>Volvariella volvacea</i>	NVAGWNGSPNDTNAGTGNW ~~GACC~~ NEMDI ~~WE~~ ANS	205-237	209-221	229, 234
SEQ ID NO:215	46395332	<i>Irpex lacteus</i>	NVAGWTGSSSPNSGTGN ~~YGTC~~ CCSEMDI ~~WE~~ ANS	202-234	206-218	226, 231

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:216	50844407 #	<i>Chaetomium thermophilum</i> var <i>thermophilum</i>	N I E N W T P S T N D A N A G F G R Y G S C C S E M D I W E A N N	182-214	186-198	206, 211
SEQ ID NO:217	4586347	<i>Irpex lacteus</i>	N I V D W T A S A G D A N S G T G S F G T C C Q E M D I W E A N S	203-235	207-219	227, 232
SEQ ID NO:218	3980202	<i>Phanerochaete chrysosporium</i>	N V G N W T E T G - - S N T G T G S Y G T C C S E M D I W E A N N	203-233	207-217	225, 230
SEQ ID NO:219	27125837	<i>Melanocarpus albomyces</i>	N I E G W K S S T S D P N A G V G P Y G S C C A E I D V W E S N A	210-242	214-226	234, 239
SEQ ID NO:220	171696102	<i>Podospora anserina</i>	N V E G W G G A D - - G N S G T G K Y G I C C A E M D I W E A N S	206-236	210-220	228, 233
SEQ ID NO:221	3913802	<i>Cochliobolus carbonum</i>	N V E G W N P S D A D P N G G A G K I G A C C P E M D I W E A N S	208-240	212-224	232, 237
SEQ ID NO:222	50403723	<i>Trichoderma viride</i>	N V E G W E P S S N I A N T G I G G H G S C C S E M D I W E A N S	205-237	209-221	229, 234
SEQ ID NO:223	3913798	<i>Aspergillus aculeatus</i>	N I E G W E P S S T D V N A G T G N H G S C C P E M D I W E A N S	210-242	214-226	234, 239
SEQ ID NO:224	66828465	<i>Dictyostelium discoideum</i>	N V D G W I P S T N P N T G Y G N L G S C C A E M D I W E A N N	206-238	210-222	230, 235
SEQ ID NO:225	156060391	<i>Sclerotinia sclerotiorum</i> 1980	N S V G W T P S S N D V N T G T G Q Y G S C C S E M D I W E A N K	192-224	196-208	216, 221
SEQ ID NO:226	116181754	<i>Chaetomium globosum</i> CBS 148-51	N S E G W G G E D - - G N S G T G K Y G T C C A E M D I W E A N L	203-233	207-217	225, 230
SEQ ID NO:227	145230535	<i>Aspergillus niger</i>	N C D G W E P S S N N V N T G V G D H G S C C A E M D V W E A N S	209-241	213-225	233, 238
SEQ ID NO:228	46241266	<i>Nectria haematococca</i> mpVI	N S D E W K P S D S D K N A G V G K Y G T C C P E M D I W E A N K	205-237	209-221	229, 234

TABLE 4						
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:229	1q9h (PDB) #	<i>Talaromyces emersonii</i>	NVEGWQPSNNANTGIGDHGSCCAE <u>MDVWE</u> ANS	185-217	189-201	209, 214
SEQ ID NO:230	157362170	<i>Polyporus arcularius</i>	NVLDWAGSSNDPNAGTGHYGTCCN <u>EMDIWE</u> ANS	208-240	212-224	232, 237
SEQ ID NO:231	7804885	<i>Leptosphaeria maculans</i>	NAEGWTKSADPNSSGVGKKGACCAQMDV <u>WE</u> ANS	204-236	208-220	228, 233
SEQ ID NO:232	121852	<i>Phanerochaete chrysosporium</i>	NVEGWNATS - <u>ANAGTGN</u> YGTCT <u>EMDIWE</u> ANN	203-233	207-217	225, 230
SEQ ID NO:233	126013214	<i>Penicillium decumbens</i>	NVEGWKPSANDKNAGVPHGSCCAE <u>MDIWE</u> ANS	201-233	205-217	225, 230
SEQ ID NO:234	156048578	<i>Sclerotinia sclerotiorum</i> 1980	NVDGWPSSNNPNTGVGNYGSCCAE <u>MDIWE</u> ANS	202-234	206-218	226, 231
SEQ ID NO:235	156712278	<i>Acremonium thermophilum</i>	NIDGWQPSNDANAGLGNHGSCC <u>SEMDIWE</u> ANK	206-238	210-222	230, 235
SEQ ID NO:236	21449327	<i>Aspergillus nidulans</i> (also known as <i>Emicella nidulans</i>)	NVEGWEPSSDANAGVGGMGTCPP <u>EMDIWE</u> ANS	202-234	206-218	226, 231
SEQ ID NO:237	171683762	<i>Podospora anserina</i> (S mat++)	NIEGWRESSNDENAGVGPYGGCCAE <u>IDVWE</u> SNA	211-243	215-227	235, 240
SEQ ID NO:238	56718412	<i>Thermoscypha aurantiacus</i> var <i>levisporus</i>	NVEGWQPSANDPNAGVGNHGSCCAE <u>MDVWE</u> ANS	205-237	209-221	229, 234
SEQ ID NO:239	15824273	<i>Pseudotrachynema ha grassii</i>	NVENWKPQINDENAGNGRYGACCT <u>EMDIWE</u> ANK	200-232	204-216	224, 229

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:240	115390801	<i>Aspergillus terreus</i> NIH2624	NVEGWT <u>PSDNDKNAGVGGHGCCCP</u> ELDIWE <u>ANS</u>	203-235	207-219	227, 232
SEQ ID NO:241	453223	<i>Phanerochaete chrysosporium</i>	NVGNW <u>TETG</u> --SNTGTGSYGTCCTCCSEMDIWE <u>ANN</u>	203-233	207-217	225, 230
SEQ ID NO:242	3132	<i>Phanerochaete chrysosporium</i>	NVEGWL <u>GT</u> T--ATTGTGFFGSCCTDIALWE <u>AND</u>	202-232	206-216	224, 229
SEQ ID NO:243	16304152	<i>Thermoascus aurantiacus</i>	NVEGWQPSANDPNAGVGNHGSSCAE <u>MDVWE</u> ANS	205-237	209-221	229, 234
SEQ ID NO:244	156712280	<i>Acremonium thermophilum</i>	NSASWQPSNDQAGVGGMGSCCAE <u>MDIWE</u> ANS	210-242	214-226	234, 239
SEQ ID NO:245	5231154	<i>Volvariella volvacea</i>	NVQGWQPSPNDTNAGTGNYGACCNKMDVWE <u>ANS</u>	220-252	224-236	244, 249
SEQ ID NO:246	116200349	<i>Chaetomium globosum</i> CBS 148-51	NYDGWTPSSNDANAGVGALGGCCAEIDVWE <u>SNA</u>	207-239	211-223	231, 236
SEQ ID NO:247	4586343	<i>Irpex lacteus</i>	NVAGWAGSADPNAGSGTLGTCCSEMDIWE <u>ANN</u>	202-234	206-218	226, 231
SEQ ID NO:248	15321718	<i>Lentinula edodes</i>	NVEGWT <u>PSSTSPNAGTGGTGICCN</u> EMDIWE <u>ANS</u>	208-240	212-224	232, 237
SEQ ID NO:249	146424875	<i>Pleurotus sp</i> Florida	NVLDWSASATDDNAGNGRYGACCAE <u>MDIWE</u> ANS	206-238	210-222	230, 235
SEQ ID NO:250	62006158	<i>Fusarium venenatum</i>	NSDGWQPSKSDVNGGIGNLGTCCTCCPEMDIWE <u>ANS</u>	205-237	209-221	229, 234
SEQ ID NO:251	296027	<i>Phanerochaete chrysosporium</i>	NVEGWN <u>ATS</u> --ANAGTGNYGTCCTEMDIWE <u>ANN</u>	203-233	207-217	225, 230
SEQ ID NO:252	154449709	<i>Fusicoccum sp</i> BCC4124	NVQNWTA <u>SSFDKNAGTGHYGSCCN</u> EMDIWE <u>ANS</u>	209-241	213-225	233, 238

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:253	169859460	<i>Coprinopsis cinerea okayama</i>	NSVGWEPSETPNAGKGQYGICCAE <u>MDIWE</u> ANS	207-239	211-223	231, 236
SEQ ID NO:254	50400675	<i>Trichoderma harzianum</i> (anamorph of <i>Hypocrea livii</i>)	NVEGWEPSSNRANTGVGGHSGCC <u>SEMDIWE</u> ANS	201-233	205-217	225, 230
SEQ ID NO:255	729649	<i>Neurospora crassa</i> (OR74A)	NVEGWTPSTNDAN-GIGDHGSCC <u>SEMDIWE</u> ANK	200-231	204-215	223, 228
SEQ ID NO:256	119472134	<i>Neosartorya fischeri</i> NRRL 181	NVEGWQPSSNDANAGTGNHGSCCA <u>E</u> MDIWEANS	214-246	218-230	238, 243
SEQ ID NO:257	117935080	<i>Chaetomium thermophilum</i>	NIEGWRPSTNDANAGVGPYGACCA <u>E</u> IDVW <u>ESNA</u>	209-241	213-225	233, 238
SEQ ID NO:258	154300584	<i>Botryotinia fuckeliana</i> B05-10	NVDGWVPSSNNANTGVGNHGSCCA <u>E</u> MDIWEANS	202-234	206-218	226, 231
SEQ ID NO:259	15824271	<i>Pseudotriconomyces ha grassii</i>	NVENWKPQTDENAGNGRYGACCT <u>E</u> MDIWEANK	200-232	204-216	224, 229
SEQ ID NO:260	4586345	<i>Irpex lacteus</i>	NVEGWTGSSTDSNSGTGNYGTCC <u>SEMDIWE</u> ANS	202-234	206-218	226, 231
SEQ ID NO:261	46241268	<i>Gibberella avenacea</i>	NSDGWKPSSDSDINAGIGNMGTCCTCC <u>PEMDIWE</u> ANS	205-237	209-221	229, 234
SEQ ID NO:262	6164684	<i>Aspergillus niger</i>	NCDGWEPSSNNVNTGVGDHGSCCA <u>E</u> MDVW <u>ESNA</u>	209-241	213-225	233, 238
SEQ ID NO:263	6164682	<i>Aspergillus niger</i>	NVDGWEPSSNNNDNTGIGNHGSCC <u>PEMDIWE</u> ANK	203-235	207-219	227, 232
SEQ ID NO:264	33733371	<i>Chrysosporium lucknowense</i> US6573086-10	NVENWQSSSTNDANAGTGKYGSCC <u>SEMDVWE</u> ANN	206-238	210-222	230, 235
SEQ ID NO:265	29160311	<i>Thielavia australiensis</i>	NVEGWESSTNDANAGSGKYGSCCT <u>E</u> MDVW <u>ESNA</u>	206-238	210-222	230, 235

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:266	146197087	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWKPQDNDENSGNGKLGTCCEMDIWE ²²¹ GNM	197-229	201-213	221, 226
SEQ ID NO:267	146197237	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	NSEGWK ²²⁴ PQSGDKNAGNGKYGCCCEMDVWE ²²⁹ SN	200-232	204-216	224, 229
SEQ ID NO:268	146197067	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWKPQDNDENSGNGKLGTCCEMDIWE ²²¹ GNM	197-229	201-213	221, 226
SEQ ID NO:269	146197407	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWKPQSDNENSGNGRYGACCTEMDIWE ²²² ANS	198-230	202-214	222, 227
SEQ ID NO:270	146197157	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	NVEGWKPSDNDENAGTGKKGACCTEMDIWE ²²⁵ ANK	201-233	205-217	225, 230
SEQ ID NO:271	146197403	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWKPQSDNENSGNGRYGACCTEMDIWE ²²² ANS	198-230	202-214	222, 227
SEQ ID NO:272	146197081	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWKPQDNDENSGDGKLGTCCEMDIWE ²²¹ GNA	197-229	201-213	221, 226

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:273	146197413	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWK <u>PQ</u> NDENSGNGRYGACCTEMDIWEANS	198-230	202-214	222, 227
SEQ ID NO:274	146197309	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDGWK <u>PQ</u> NDKNSGNGKYGSCCSEMDIWEANS	196-228	200-212	220, 225
SEQ ID NO:275	146197227	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	NSDGWK <u>PQ</u> NDKNSGNGKYGSCCSEMDIWEANS	195-227	199-211	219, 224
SEQ ID NO:276	146197253	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	NSEGWK <u>PQ</u> SGDKNAGNGKYGSCCSEMDVWEANS	200-232	204-216	224, 229
SEQ ID NO:277	146197099	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVLDWK <u>PQ</u> NDENAGTGRYGTCTCTEMDIWEANS	197-229	201-213	221, 226
SEQ ID NO:278	146197409	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWK <u>PQ</u> NDENSGNGRWGARCTEMDIWEANS	198-230	202-214	222, 227
SEQ ID NO:279	146197315	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDGWK <u>PQ</u> NDKNSGNGKYGSCCSEMDIWEANS	196-228	200-212	220, 225

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:280	146197411	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWK <u>PQSN</u> DENSGNGRYGACCTE <u>M</u> DIWE <u>A</u> NS	198-230	202-214	222, 227
SEQ ID NO:281	146197161	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	NVQDWK <u>P</u> SDNDNDNAGTGHYGACCTE <u>M</u> DIWE <u>A</u> NK	201-233	205-217	225, 230
SEQ ID NO:282	146197323	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDGWK <u>PQSN</u> DKNSGNGKYGSCCSE <u>M</u> DIWE <u>A</u> NS	196-228	200-212	220, 225
SEQ ID NO:283	146197077	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVLDWK <u>PQ</u> ETDENSGNGRYGTCCTE <u>M</u> DIWE <u>A</u> NS	201-233	205-217	225, 230
SEQ ID NO:284	146197089	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVEDWK <u>PQ</u> DNDENSGNGKLGTCSCSE <u>M</u> DIWE <u>G</u> NA	197-229	201-213	221, 226
SEQ ID NO:285	146197091	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVLDWK <u>PQSN</u> DENAGTGRYGTCTE <u>M</u> DIWE <u>A</u> NS	197-229	201-213	221, 226
SEQ ID NO:286	146197097	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWK <u>PQ</u> DNDENSGNGKLGTCSCSE <u>M</u> DIWE <u>G</u> NA	197-229	201-213	221, 226

TABLE 4

Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:287	146197095	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDD <u>WKPKQND</u> ENSGNGKLGTCCEMDIWE <u>GNA</u>	197-229	201-213	221, 226
SEQ ID NO:288	146197401	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLD <u>WKPKQ</u> NDENSGNGRYGACCIEMDIWE <u>ANS</u>	198-230	202-214	222, 227
SEQ ID NO:289	146197225	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	NSDG <u>WKPKQ</u> NDKNSGNGKYGSCCEMDIWE <u>ANS</u>	195-227	199-211	219, 224
SEQ ID NO:290	146197317	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDG <u>WKPKQ</u> NDKNSGNGKYGSCCEMDIWE <u>ANS</u>	196-228	200-212	220, 225
SEQ ID NO:291	146197251	uncultured symbiotic protist of <i>Neotermes koshunensis</i>	NSDG <u>WKPKQ</u> NDKNSGNGRYGSCCEMDVWE <u>ANS</u>	195-227	199-211	219, 224
SEQ ID NO:292	146197319	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDG <u>WKPKQ</u> NDKNSGNGKYGSCCEMDIWE <u>ANS</u>	196-228	200-212	220, 225
SEQ ID NO:293	146197071	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NILD <u>WKPKSS</u> NDENAGAGRYGTCCTE <u>MDIWE</u> ANS	200-232	204-216	224, 229

TABLE 4						
Sequence Identifier (SEQ ID NO:)	Database Accession Number	Species of Origin	Amino acid sequence of fragment of catalytic domain including loop and catalytic residue	Amino acid positions of fragment in sequence identifier	Amino acid positions of active site loop in sequence identifier	Position of catalytic residues in sequence identifier
SEQ ID NO:294	146197075	uncultured symbiotic protist of <i>Reticulitermes speratus</i>	NVDDWKPQDNDENSGNGKLGTCCEMDIWE <u>GN</u> A	197-229	201-213	221, 226
SEQ ID NO:295	146197159	uncultured symbiotic protist of <i>Hodotermopsis sjoestedti</i>	NVKDWKPQETDENAGNGHYGACCTE <u>MDIWE</u> ANS	197-229	201-213	221, 226
SEQ ID NO:296	146197405	uncultured symbiotic protist of <i>Cryptocercus punctulatus</i>	NVLDWKPQSNIDENSGNGRYGACCTE <u>MDIWE</u> ANS	198-230	202-214	222, 227
SEQ ID NO:297	146197327	uncultured symbiotic protist of <i>Mastotermes darwiniensis</i>	NSDGWKPQDNDENSGNGKYGSCCSE <u>MDIWE</u> ANS	201-233	205-217	225, 230
SEQ ID NO:298	146197261	uncultured symbiotic protist of <i>Neotermes kosshunensis</i>	NSDGWKPQKQIDKNSGNGKYGSCCSE <u>MDIWE</u> ANS	195-227	199-211	219, 224

TABLE 5		
	Tolerance to 250mg/L cellobiose	Tolerance to cellobiose accumulation
Substitution(s)	% Activity in 4-MUL Assay (+/- Cellobiose)[±]	% Activity in Bagasse Assay (+/- BG)[±]
None	25%	60%
R273K/R422K	95%	84%
R273K/Y274Q/D281K/Y410H/P411G/R422K	78%	ND

TABLE 6		
	Tolerance to 250mg/L cellobiose	Tolerance to cellobiose accumulation
Substitution(s)	% Activity in 4-MUL Assay (+/- Cellobiose)[±]	% Activity in Bagasse Assay (+/- BG)[±]
None	23%	74%
R268K/R411K	92%	94%
R268A/R411A	92%	95%
R268A/R411K	97%	94%
R268K/R411A	97%	102%
R268K	ND	92%
R268A	ND	86%
R411K	ND	89%
R411A	ND	94%

TABLE 7

TABLE 7																				
SEQ ID NO.	Amino acid sequence																			
SEQ ID NO:1	MSALNSFNMY	KSALILGSL	ATAGAAQQIGT	YTAETHPSLS	WSTCKSGGSC	TTNSGAITLD	ANRWVHGVN	TSTNCYTGNT	WNTAICDTDA											
	SCAQDCALDG	ADYSGTYGIT	TSGNSLRINF	VTGSNVGSRT	YLMADNTHYQ	IFDLLNQEFT	FTVDVSHLPC	GLNGALYFVT	MDADGGVSKY											
	PNNKAGAQYG	VGYCDSQCPR	DLKFIAGQAN	VEGWTPSSNN	ANTGLGNHGA	CCAELDIWEA	NSISEALTPH	PCDTPGLSVC	TTDACGGTYS											
	SDRYAGTCDP	DGCDENPYRL	GVTDFFYGSBK	TVDTTKPITV	VTQFVTDGDT	STGTLSEIRR	YYVQNGVVIP	QPSSKISGVS	GNVINSDFCD											
	AEISTFGETA	SFSKHGGLAK	MGAGMEAGMV	LVMSLWDDYS	VNMLWLDSTY	PTNATGTPGA	ARGSCTPTSG	DPKTVESQSG	SSYVTFSDIR											
	VGPFNSTFSG	GSSTGGSSTT	TASGTTTTKA	SSTSTSSTST	GTGVAAHWGQ	CGGQGWITGPT	TCASGTTCTV	VNPYYSQCL												
SEQ ID NO:2	MYRKLAVISA	FLATARAQSA	CTLQSETHPP	LTWQKCSSGG	TCTQQTGSVV	IDANWRWTHA	TNSSTNCYDG	NTWSSTLCPD	NETCAKNCCCL											
	DGAAYASTYG	VTTSGNSLSI	GFVTQSAQKN	VGARLLYLMAS	DDTYQEFTLL	GNEFSFDVDV	SQLPCGLNGA	LYFVSMDADG	GVSKYPTNTA											
	GAKYGTGYCD	SQCPRDLKFI	NGQANVEGWE	PSSNNANTGI	GGHGSCCSEM	DIWEANSISE	ALTPHPCTTV	GQEICEGDGC	GGTYSDNRYG											
	GTCDPDGCDW	NPYRLGNTSF	YGPSSSFTLD	TTKKLTVVVQ	FETSGAINRY	YVQNGVTFQ	PNAELGSYSG	NELNDDYCTA	EEAEFGGSSF											
	SDKGGLTQFK	KATSGGMVLV	MSLWDDYYAN	MLWLDSTYPT	NETSSTPGAV	RGSCSTSSGV	PAQVESQSPN	AKVTFSNIKF	GPIGSTGNPS											
	GGNPPGGNPP	GTTTTTRRPAT	TTGSSPGPTQ	SHYGQCGGIG	YSGPTVCASG	TTCQVLNPHY	SQCL													
SEQ ID NO:3	MSALNSFNMY	KSALILGSL	ATAGAAQQIGT	YTAETHPSLS	WSTCKSGGSC	TTNSGAITLD	ANRWVHGVN	TSTNCYTGNT	WNSAICDTDA											
	SCAQDCALDG	ADYSGTYGIT	TSGNSLRINF	VTGSNVGSRT	YLMADNTHYQ	IFDLLNQEFT	FTVDVSHLPC	GLNGALYFVT	MDADGGVSKY											
	PNNKAGAQYG	VGYCDSQCPR	DLKFIAGQAN	VEGWTPSANN	ANTGIGNHGA	CCAELDIWEA	NSISEALTPH	PCDTPGLSVC	TTDACGGTYS											
	SDRYAGTCDP	DGCDENPYRL	GVTDFFYGSBK	TVDTTKPFTV	VTQFVTNDGT	STGSLSEIRR	YYVQNGVVIP	QPSSKISGIS	GNVINSDYCA											
	AEISTFGGTA	SFNKHGGLTN	MAAGMEAGMV	LVMSLWDDYA	VNMLWLDSTY	PTNATGTPGA	ARGTCATTSG	DPKTVESQSG	SSYVTFSDIR											
	VGPFNSTFSG	GSSTGGSTTT	TASRTTTTSA	SSTSTSSTST	GTGVAGHWGQ	CGGQGWITGPT	TCVSGTTCTV	VNPYYSQCL												
SEQ ID NO:4	ESACTLQSET	HPPLTWQKCS	SGGTCTQQTG	SVVIDANWRW	THATNSSTNC	YDGN	TWSSTL	CPDNETCAKN	CCLDGAAYAS	TYGVTTSGNS										
	LSIDFVTQSA	QKNVGARLYL	MASD	TTYQEF	TL	LGNEFSFD	VDSQLPCGL	NGALYFVSMD	ADGGVSKYPT	NTAGAKYGTG	YCD	SQCPRDL								
	KFINGQANVE	GWEPSSNNAN	TGIGGHGSCC	SEMDI	WQANS	I	SEALTPHPC	TTVGQEICEG	DGCGGTYS	SDN	RYGGTC	DPDG	CDWNPYRLGN							
	TSFYGP	GSSF	TLDTTKKLTV	VTQFETSGAI	NRYYVQNGVT	FQQPNAELGS	YSGNELNDDY	CTAE	EA	EEFGG	SSF	SDKGGLT	QFKKATSGGM							
	VLVMSLWDDY	YANMLWLDST	YPTNETSSTP	GA	VRGSCSTS	SGVPAQVESQ	SPNAKVTFSN	IKFGPIGSTG	NPSG											
SEQ ID NO:5	MASSFQLYKA	LLFFSSLLSA	VQAQKVGTQQ	AEVHPGLTWQ	TCTSSGSC	TT	VNGE	VTIDAN	WRWLHTVNGY	TNCYTGNEWD	TSICTSNEVC									
	AEQCAVDGAN	YASTYGITTS	GSSLRLNFVT	QSQQKNIGSR	VYLMDD	EDTY	TMFYLLNKEF	TFD	V	DV	SEL	P	CGLNGA	VYFV	SMDADGGKSR					
	YATNEAGAKY	GTGYCDSQCP	RDLK	FINGVA	NVEGWESSDT	NPNGGVGNHG	SCCAEMDIWE	ANSIS	T	AFT	P	HPCD	TPG	QTL	CTGDS	CGGT	Y			
	SNDRYGGTCD	PDGCD	FN	SYR	QGNK	TFYGP	G	LVTD	TNSP	V	VTQFL	TDDN	TDTG	TLSEIK	RFYVQNGVVI	PNSE	STYPAN	PGNSITTEFC		
	ESQKELFGDV	DVFS	AHGGMA	GMGAAL	EQGM	VLVLSLWDDN	YSNMLWLD	SN	YPTD	ADPTQ	P	GIARG	TCPTD	SGVP	SEVEAQ	YPNAYV	VVYSN			
	IKFGPIGSTF	GNGGSGPTT	TVTT	STATST	TSSAT	STATG	QAQHWEQCGG	NGWTGPTVCA	SPWACTVVNS	WYSQCL										
SEQ ID NO:6	MYRAIATASA	LIAAVRAQQV	CSLTQESKPS	LNWSKCTSSG	CSNVKGSVTI	DANWRWTHQV	SGSTNCYTGN	KWDT	SVCTSG	KVCAEKCCCLD										
	GADYASTYGI	TSSGDQLSLS	FVTKGPYSTN	IGSRTYLMED	ENTYQMFQ	LL	GNEFT	FDV	V	SNIGCG	LNGA	LYFVSMDADG	GKAKY	PGNKA						
	GAKYGTGYCD	AQCP	RDVKFI	NGQANS	DGWQ	PSDS	DVNGGI	GNLGT	CCPEM	DIWEANS	IST	AYTPHP	CTKL	TQHSCT	G	DSC	GGTYS	DNDRY	G	
	GTC	DADG	CDE	NSYRQ	GNKTF	YGP	GSG	FNVD	TTKKV	TVVTQ	FHKGS	NGRLS	EITRL	VQNG	KVIAN	SESKI	AGVP	GN	SLTA	DFCTKQKKVF
	NDPDDFTKKG	AWSGM	SDALE	APMV	LVMSLW	HDHHS	NMLWL	DSTYPT	DSTK	LGSQR	GSCST	SSGV	PADLEK	NVPNSK	VAFS	NIKFG	P	IGST		
	YKSDG	TTPTN	PTNP	SEPSNT	ANPN	PGTVDQ	WGQC	GGSNYS	GPTACK	SGFT	CKKIND	FYSQ	CQ							

TABLE 7

TABLE 7																				
SEQ ID NO.	Amino acid sequence																			
SEQ ID NO:7	MLASTFSYRM	YKTALILAAL	LGSGQAQQVG	TSQAEVHPSM	TWQSCTAGGS	CTTNNGKVVI	DANWRWVKV	GDYTNCYTGN	TWDTTICPDD											
	ATCASNCALE	GANYESTYGV	TASGNSLRIN	FVTTSQQKNI	GSRLYMMKDD	STYEMFKLLN	QEFTFDVDVS	NLPCGLNGAL	YFVAMDADGG											
	MSKYPTNKAG	AKYGTGYCDS	QCPRDLKFIN	GQANVEGWQP	SSNDANAGTG	NHGSCCAEMD	IWEANSISTA	FTPHPCDTPG	QVMCTGDACG											
	GTYSDDRYGG	TCDDPDGCDFN	SFRQGNKTFY	GPGMTVDTKS	KFTVVTQFIT	DDGTSSGTLK	EIKRFYVQNG	KVIPNSESTW	TGVSGNSITT											
	EYCTAQKSLF	QDQNVFEKHG	GLEGMGAALA	QGMVLVMSLW	DDHSANMLWL	DSNYPTTASS	TTPGVARGTC	DISSGVPADV	EANHPDAYVV											
	YSNIKVGPIG	STFNSGGSNP	GGGTTTTTTT	QPTTTTTITAG	NPGGTGVAQH	YGQCGGIGWT	GPTTCASPYT	CQKINDYYSQ	CL											
SEQ ID NO:8	MLPSTISYRI	YKNALFFAAL	FGAVQAQKVG	TSKAEVHPSM	AWQTCAADGT	CTTKNGKVVI	DANWRWHDV	KGYTNICYTGN	TWNAELCPDN											
	ESCAENCALE	GADYAATYGA	TTSGNALSLK	FVTQSQQKNI	GSRLYMMKDD	NTYETFKLLN	QEFTFDVDVS	NLPCGLNGAL	YFVSMADAGG											
	LSRYTGNEAG	AKYGTGYCDS	QCPRDLKFIN	GLANVEGWTP	SSSDANAGNG	GHGSCCAEMD	IWEANSISTA	YTPHPCDTPG	QAMCNGDSCG											
	GTYSDDRYGG	TCDDPDGCDFN	SYRQGNKSFY	GPGMTVDTKK	KMTVVVTQFLT	NDGTATGTLN	EIKRFYVQDG	KVIANSESTW	PNLGGNSLTN											
	DFCKAQKTVF	GMDTFSKHG	GMEGMGAALA	EGMVLVMSLW	DDHNSNMLWL	DSNSPTTGTS	TTPGVARGSC	DISSGDPKDL	EANHPDASVV											
	YSNIKVGPIG	STFNSGGSNP	GGSTTTTKPA	TSTTTTKATT	TATNTTGPT	GTGVAQPPWAQ	CGGIGYSGPT	QCAAPYTCTK	QNDYYSQCL											
SEQ ID NO:9	MHPSLQTILL	SALFTTAHAQ	QACSSKPETH	PPLSWSRCR	SGCRSVQGA	TVDANWLWTT	VDGSQNCYTG	NRWDTSICSS	EKTCSESCCI											
	DGADYAGTYG	VTTTGDALSL	KFVQQGPYSK	NVGSRLYLMK	DESRYEMFTL	LGNEFTFDVD	VSKLGCGLNG	ALYFVSMDED	GGMKRFFPMNK											
	AGAKFGTGYC	DSQCPRDVKF	INGMANSKDW	IPSKSDANAG	IGSLGACC	MDIWEANNIA	SAFTPHPCKN	SAYHSCTGDG	CGGTYSKNRY											
	SGDCDDPDGCD	FNSYRLGNTT	FYGP	GPKFTI	DTTRKISVVT	QFLKGRDGS	REIKRFYVQN	GKVIPNSVSR	VRGV	P	GN	SIT	QGFCNAQKKM							
	FGAHESFNAK	GMKGM	SAAV	SKPMVLVMSL	WDDHNSNMLW	LDSTYPTNSR	QRGSKRGSCP	ASSGRPTDVE	SSAPDSTVVF	SNIKFGPIGS	TFSR	GK								
SEQ ID NO:10	ESACTLQSET	HPPLTWQKCS	SGGTCTQQTG	SVVIDANWRW	THATNSSTNC	YDGN	TWSSTL	CPDNETCAKN	CCLDGAAYAS	TYGVT	TS	GNS								
	LSIDFVTQSA	QKNVGARLYL	MASD	TTYQEF	TLLGNEFSFD	VDVSQ	LPCGL	NGALYFVSM	D	ADGGVSKYPT	NTAGAKYGTG	Y	CDS	QCPRDL						
	KFINGQANVE	GWEPSSNNAN	TGIGGHGSCC	SEMDIWQANS	I	SEALTPHPC	TTVGQEICEG	DGCGGTYS	DN	RYGGTCDPDG	CDWNPYRLGN									
	TSFYGP	GSSE	TLDTTKKLT	V	TQFETSGAI	NRYYVQNGVT	FQQPNAEIGS	YSGNELNDDY	CTAEEAEFEG	SSFS	DKGGLT	Q	FKKATSGGM							
	VLVMSLWDDY	YANMLWLDST	YPTNETSSTP	GAVRGSCSTS	SGVPAQVESQ	SPNAKVTFESN	IKFGPIGSTG	NPSG												
SEQ ID NO:11	MHQRALLFSA	LAVAANAQQV	GTQKPETHPP	LTWQKCTAAG	SCSQQSGSVV	IDANWRWLHS	TKD	TTN	CYTG	NTWNT	ELCPD	NE	SCAQNC	AV						
	DGADYAGTYG	VTTSGSELKL	SFVTGANVGS	RLYLMQDDET	YQHFNLLNNE	FTFDVDVSNL	PCGLNGALYF	VAMDADG	GMS	KYPSN	KAGAK									
	YGTGYCDSQC	PRDLKFINGM	ANVEGWKPSS	NDKNAGVGGH	GSCCPEMDIW	EANSISTAVT	PHPCDDVSQT	MCSGDACGGT	YSATRYAGTC											
	DPDGCDFNPF	RMGNESFYGP	GKIVDTKSEM	TVVTQFITAD	GTD	TGALSEI	KRLYVQNGKV	IANSVSNVAD	VSGNSISSDF	CTAQK	KAFGD									
	EDIFAKHGG	L	SGMGKALSEM	VLIMSIWDDH	HSSMMWLDST	YPTDADPSKP	GVARGTCEHG	AGDPEKVESQ	HPDASVTFSN	IKFGPIGSTY	KA									
SEQ ID NO:12	MYRSLIFATS	LLSLAKGQLV	GNLYCKG	SC	T	GNLYCKG	SC	T	GNLYCKG	SC	T	GNLYCKG	SC	T	GNLYCKG	SC	T	GNLYCKG	SC	T
	RQLLGTEVHH	QGLYSTNIGS	RTYLMQDDST	YQLFKFTGSQ	EFTFDVDLSN	LPCGLNGALY	FVSM	DADGGL	KKYPTNKAGA	KYGTGYCDAQ										
	CPRDLKFING	EGNVEGWQPS	KNDQ	NAGVGG	HGSCCAEMDI	WEANSVSTAV	TPHSCSTIEQ	S	RCDGDGCGG	TYSADRYAGV	CDPDG	CDFNS								
	YRMGVKDFY	G	KGT	VDTSKK	FTVVTQFIGS	G	DAMEIKRFY	VQNGKTIPQP	DSTIPGVTGN	SIT	TF	FCDAQ	KKAFGD	KYTF	KDKGGM	ANMP				
	STCNGMVLVM	SLWDDHYSNM	LWLDSTYPTD	KNP	DTDAGSG	RGECAITSGV	PADVESQHPD	ASVIYSNIKF	GPINTTFG											

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:13	MLAKFAALAA LVASANAQAV CSLTAETHPS LNWSKCTSSG CTNVAGSITV DANWRWTHIT SGSTNCYSGN EWDTSLCSTN TDCATKCCVD GAEYSSTYGI QTSGNSLSLQ FVTKGSYSTN IGSRTYLMNG ADAYQGFELL GNEFTFDVDV SGTGCGLINGA LYFVSMDLDG GKAKYTNKKA GAKYGTGYCD AQCPRDLKYI NGIANVEGWT PSTNDANAGI GDHGTCCSEM DIWEANKVST AFTPHPCTTI EQHMCEDGSC GGTYSDDRYG GTCDADGCDF NSYRMGNITF YGEGKTVDT S YGEGKTVDT SKFTVVVTQFI KDSAGDLAEI KRFYVQNGKV IENSQSNVDG VSGNSITQSF CNAQKTAFGD IDDFNKKGGL KQMGKALAKP MVLVMSIWDD HAANMLWLDS TYPVEGGPGA YRGECPITSG VPAEVEANAP NSKVIFSNIK FGPIGSTFSG GSSGTPPSNP SSSVKPVTST AKPSSSTSTAS NPSGTGAAHW AQCGGIGFSG PTTCQSPYTC QKINDYYSQC V
SEQ ID NO:14	MFKKVALTAL CFLAVAQAQQ VGREVAENHP RLPWQRCTRN GGCQTVSNGQ VVLDAWNRWL HVTDGYTNCY TGNSWNSTVC SDPTTCAQRC ALEGANYQQT YGITTNGDAL TIKFLTRSQQ TNVGARVYLM ENENRYQMFN LLNKEFTFDV DVSKVPCGIN GALYFIQMDA DGMSKQPNN RAGAKYGTGY CDSQCPRDIK FIDGVANSAD WTPSETDPNA GRGRYGICCA EMDIWEANSI SNAYTPHPCR TQNDGGYQRC EGRDCNQPRY EGLCDPDGCD YNPFRMGNKD FYGPCKTVDT NRKMTVVVTQF ITHDNTDTGT LVDIRRLYVQ DGRVIANPPT NFPGLMPAHD SITEQFCTDQ KNLEFDYSSF ARDGGLAHMG RSLAKGHVLA LSIWNHDGAH MLWLDSNYPT DADPNKPGIA RGTCPTTGGT PRETEQNHPD AQVIFSNIK GDIGSTFSGY
SEQ ID NO:15	MYSAAVLATF SFLLGAGAQQ VGTSTAETHP ALTVQKCAAG GTCTDESDSI VLDANWRWLH STSGSTNCYT GNTWDTTLCP DAATCTTNCA LDGADYEGTY GITTSGDSLK LSFVTGSNVG SRTYLMDSER TYKEFALLGN EFTFTVDVSK LPCGLNGALY FVPMDADGGM SKYPTNKAGA KYGTGYCDAQ CPQDMKFVNG TANVEGWVPD SNSANSGETN IGSCSEFDV WEANSMSQAL TPHVCTVDSQ TACTGDDCAS NTGVCDGDGC DFNPYRMGNT TFYSGMTID TSKPFSVVTD FITDDGTETG TLTEIKRFYV QDDVVYEQPS SDISGVSGNS ITDDFCAAQK TAFGDTDYFT QNGGMAAMGK KMADGMVLVL SIWDDYNVNM LWLSDSYPTT KDASTPGVSR GSCATDSGVP ATVEAASGSA YVTFSSIKYG PIGSTFNAPA DSSSSVSASS SPAPIASSSS SASIAPVSSV VAAIVSSSAQ AISSAAPVVS SSAQAIISSAA PVVSSVVSSA APVATSTSKS KCSKVSSTLK TSVAAPATSA TSAAVVATSS AASSTGSVPL YGNCTGGKTC SEGTCCVVQND YYSQCVAAS
SEQ ID NO:16	MTWQRCGTG GSSCTNVNGE IVIDANRWI HATGGYTNCF DGNEWNKTAC PSNAACTKNC AIEGSDYRGT YGITTSGNSL TLKFITKGQY STNVGSRTYL MKDTNNYEMF NLIGNEFTFD VDLSQLPCGL NGALYFVSMP EKGQGTPGAK YGTGKLSQCS VHISKTLTDA CARDLKFEVGG EANADGWQAS TSDPNAGVGK KGACCAEMDV WEANSMTAL TPHSCQPEGY AVCEESNCGG TYSLDRYAGT CDANGCDFNP YRVGNKDFYG KGKTVDTSKK MTVVVTQFLGT GSDLTELKRF YVQDGKVISN PEPTIPGMTG NSITQKWCDD QKEVFKEEVY PFNQWGGMAS MGKGMAQGMV LVMSLWDDHY SNMLWLDSTY PTDRDPESPG AARGECAITS GAPAEVEANN PDASVMFSNI KFGPIGSTFQ QPA
SEQ ID NO:17	MQIKSYIQYL AAALPLLSSV AAQQAQTITA ENHPRMTWKR CSGPGNCQTV QGEVVIDANW RWLHNNGQNC YEGNKWTSQC SSATDCAQRC ALDGANYQST YGASTSGDSL TLKFVTKHEY GTNIGSRFYL MANQNKYQMF TLMNNEFAFD VDLKSKVECGI NSALYFVAME EDGGMASYPS NRAGAKYGTG YCDAQCARDL KFIGGKANIE GWRPSTNDPN AGVGPMGACC AEIDVWESNA YAYAFTHAC GSKNRYHICE TNNCGGTYS DRFAGYCDAN GCDYNPYRMG NKDFYKGKGT VDTNRKFTTV SRFERNRLSQ FFVQDGRKIE VPPPTWPGLP NSADITPELC DAQFRVFDNR NRFAETGGFD ALNEALTIPM VLVMSIWDDH HSNMLWLDSS YPPEKAGLPG GDRGPCPTTS GVPAEVEAQY PNAQVVWSNI RFGPIGSTVN V
SEQ ID NO:18	MRTAKFATLA ALVASAAAQQ ACSLTTERHP SLSWKKCTAG GQCQTVOASI TLDSNWRWTH QVSGSTNCYT GNKWDTSICT DAKSCAQNCC VDGADYTSTY GITTNGDLSL LKFVTKGQYS TNVGSRTYLM DGEDKYQTFE LLGNEFTFDV DVSNIGCGLN GALYFVSMDA DGGLSRYPGN KAGAKYGTGY CDAQCPRDIK FINGEANIEG WTGSTNDPNA GAGRYGTCCS EMDIWEANNM ATAFTPHPCT IIGQSRCEGD SCGGTYSNER YAGVCDPDGC DFNSYRQGNK TFYKGKMTVD TTKKITVVTD FLKDANGDLG EIKRFYVQDG KIIPNSESTI PGVEGNSITQ DWCDRQKVAE GDIDDFNRKG GMKQMGKALA GPMVLVMSIW DDHASNMLWL DSTFPVDAAG KPGAERGACP TTSGVPAEVE AEAPNSNVVF SNIRFGPIGS TVAGLPGAGN GGNNGGNPPP PTTTTSAPPA TTTTASAGPK AGRWQQCGGI GFTGPTQCEE PYTCTKLNDW YSQCL

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:25	MVDIQIATFL LLGVVGVAAQ QVGTYIPENH PLLATQSCTA SGGCTTSSSK IVLDANRRWI HSTLGTTSCL TANGWDPTLC PDGITCANYC ALDGVSYSSST YGITTSGSAL RLQFVTGTNI GSRVFLMADD THYRTFQLLN QELAFDVDVS KLPCGLNGAL YFVAMDADGG KSKYPGNRAG AKYGTGYCDS QCPRDVQFIN GQANVQGWNA TSATTGTGSY GSCCTELDIW EANSNAAALT PHTCTNNAQT RCSGSNCTSN TGFCDDADGCD FNSFRLGNTT FLGAGMSVDT TKTFTVVTFQ ITSDNTSTGN LTEIRRFYVQ NGNVIPNSVV NVTGIGAVNS ITDPFCSQQK KAFIETNYFA QHGGLAQLGQ ALRTGMVLAF SISDDPANHM LWLDSNFPPS ANPAVPGVAR GMC-SITSGNP ADVGILNPSP YVSFLNIKEG SIGTTRPA
SEQ ID NO:26	MHQRALLFSA LAVAANAQQV GTQTPETHPP LTWQKCTAAG SCSQQSGSVV IDANWRWLHS TKDTTNCYTG NTWNTELCPD NESCAQNCAL DGADYAGTYG VTTSGSELKL SFVTGANVGS RLYLMQDDET YQHFNLLNHE FTFDVDVSNL PCGLNGALYF VAMDADGMS KYPSNKAGAK YGTGYCDSQC PRDLKFINGM ANVEGWERSS SDKNAGVGGH GSCCPEMDIW EANSISTAVT PHPCDDVSQT MCSGDACGGT YSESRYAGTC DPDGCDFNPF RMGNESFYGP GKIVDTKSKM TVVTQFITAD GTDSGALSEI KRLYVQNGKV IANSVSNVAG VSGNSITSDF CTAQKKAFGD EDIFAKHGGL SGMGKALSEM VLIMSIWDDH HSSMMWLDST YPTDADPSKP GVARGTCEHG AGDPENVESQ HPDASVTFSN IKFGPIGSTY EG
SEQ ID NO:27	MFRATATLLAF TMAAMVFGQQ VGTNTAENHR TLTSQKCTKS GGCSNLNTKI VLDANWRWLH STSGYTNCYT GNQWDATLCP DGKTCAANCA LDGADYTGTY GITASGSSLK LQFVTGSNVG SRVYLMADDT HYQMFQLLNQ EFTFDVDMSN LPCGLNGALY LSAMDADGGM AKYPTNKAGA KYGTGYCDSQ CPRDIKFIN EANVEGWNAT SANAGTGNYG TCCTEMDIWE ANNDAAAYTP HPCTTNAQTR CSGSDCTRDT GLCDADGCD NSFRMGDQTF LGKGLTVDT S KPFTVVVTQFI TNDGTSAGTL TEIRRLYVQN GKVIQNSSVK IPGIDPVNSI TDNECSQQKT AFGDTNYFAQ HGGLKQVGEA LRTGMVLALS IWDDYAAANML WLDSNYPTNK DPSTPGVARG TCATTSGVPA QIEAQSPNAY VVFSNIKEFGD LNTTYTGTVS SSSVSSSHSS TSTSSSHSSS STPPTQPTGV TVPQWGQCGG IGYTGSTTCA SPYTCHVLNP YYSQCY
SEQ ID NO:28	MYQRALLFSF FLAAARAHEA GTVTAENHPS LTWQKCSSGG SCTTQNGKVV IDANWRWVHT TSGYTNCYTG NTWDTSICPD DVTCAQNCAL DGADYSGTYG VTTSGNALRL NFVTQSSGKN IGSRLYLLQD DTTYQIFKLL GQEFTFDVDV SNLPCGLNGA LYFVAMDADG NLSKYPGNKA GAKYGTGYCD SQCPRLKFI NGQANVEGWQ PSANDPNAGV GNHGSSCAEM DVWEANSIST AVTPHPCDTP GQTMCGDDC GGTYSSTRYA GTCDDPDGCD NPYQPGNHSF YGPGKIVDTS SKFTVVVTQFI TDDGTSPGTL TEIKRFRYVQN GKVIPQSEST ISGVTGNSIT TEYCTAQKAA FGDNTGFFTH GGLQKISQAL AQGMVLVMSL WDDHAANMLW LDSTYPTDAD PDTPGVARGT CPTTSGVPAD VESQNPNSYV IYSNIKVGP I NSTFTAN
SEQ ID NO:29	MQIKSYIQYL AAALPLLSSV AAQAGTITA ENHPRMTWKR CSGPGNCQTV QGEVVIDANW RWLHNNGQNC YEGNKWTSQC SSATDCAQRC ALDGANYQST YGASTSGDSL TLKFVTKHEY GTNIGSRFYL MANQNKYQMF TLMNNEFAFD VDLSKVECGI NSALYFVAME EDGGMASYPS NRAGAKYGTG YCDAQCARDL KFIGGKANIE GWRPSTNDPN AGVGPMGACC AEIDVWESNA YAYAFTPHAC GSKNRYHICE TNNCGGTYS DRFAGYCDAN GCDYNPYRMG NKDFYKGKGT VDTNRKFTVV SRFERNRLSQ FFVQDGRKIE VPPPTWPGLP NSADITPELC DAQFRVFDDR NRFAETGGFD ALNEALTIPM VLVMSIWDDH HSNMLWLDSS YPPEKAGLPG GDRGPCPTTS GVPAEVEAQY PDAQVVWSNI RFGPIGSTVN V
SEQ ID NO:30	MYRKLAVIS FLATARAQSA CTLQSETHPP LTWQKCSSGG TCTQQTGSVV IDANWRWTHA TNSSTNCYDG NTWSSTLCPD NETCAKNCC DGAAYASTYG VTTSGNSLSI GFVTQSAQKN VGARLYLMAS DTTYQEFTLL GNEFSFDVDV SQLPCCGLNGA LYFVSMDADG GVSKEYPTNTA GAKYGTGYCD SQCPRLKFI NGQANVEGWE PSSNNANTGI GGHGSCCSEM DIWEANSISE ALTPHPCTTV GQEICEGDGC GGTYSNRYG GTCDDPDGCDW DPYRLGNTSF YGPGSSFTLD TTKKLTVVVTQ FETSGAINRY YVQNGVTFQQ PNAELGSYSG NGLNDDYCTA EAEFEGGSF SDKGGLTQFK KATSGGMVLV MSLWDDYYAN MLWLDSTYPT NETSSTPGAV RGSCSTSSGV PAQVESQSPN AKVTFSNIKE GPIGSTGDPS GGNPPGGNPP GTTTTRRPAT TTGSSPGPTQ SHYGQCGGIG YSGPTVCASG TTCQVLNPYY SQCL

TABLE 7

TABLE 7			
SEQ ID NO.	Amino acid sequence		
SEQ ID NO:31	MYQRALLFSF FLAAARAQQA GTVTAENHPS LTWQQCSSGG SCTTQNGKV IDANWRWVHT TSGYTNCYTG NTWDTSI		
SEQ ID NO:32	MIRKITTLLA LVGVVRGQAA CSLTAETHPS LTWQKCSSGG SCTNVAGSVT IDANWRWTH TSGYTNCYTG NKWDT		
SEQ ID NO:33	MLRRALLSS SAILAVKAQQ AGTATAENHP PLTWQECTAP GSCTTQNGAV VLDANWRWVH DVNGYTNCYT GNTWDPT		
SEQ ID NO:34	MHQRALLFSA FWTAVQAQQA GTLTAETHPS LTWQKCAAGG TCTEQKGSVV LDSNWRWLHS VDGSTN		
SEQ ID NO:35	MASTLSFKIY KNALLLAAFL GAAQAQQVGT STAEVHPSLT WQKCTAGGSC TSQSGKVVID SNWRWVHNTG GYTNCYT		
SEQ ID NO:36	MVS		

TABLE 7

TABLE 7																				
SEQ ID NO.	Amino acid sequence																			
SEQ ID NO:37	MFPYIALVSF	SFLSVVLAQQ	VGTLTAETHP	QLTVQQCTRG	GSCTTQQRSV	VLDGNWRWLH	STSGSNNCYT	GNTWDTSLCP	DAATCSRNCA											
	LDGADYSGTY	GITSSGNALT	LKFVTHGPYS	TNIGSRVYLL	ADDSHYQMFN	LKNKEFTFDV	DVSQLPCGLN	GALYFSQMDA	DGGTGRFPNN											
	KAGAKYGTGY	CDSQCPHDIK	FINGEANVQG	WQSPNDSNA	GKGQYGSCCA	EMDIWEANSM	ASAYTPHPCT	VTTPTRCQGN	DCGDGDNRYG											
	GVCDKDG CDE	NSFRMGDKNF	LPGKTVNTN	SKFTVVVTQFL	TSDNNTTSGTL	SEIRRLYVQN	GRVIONSKVN	IPGMASTLDS	ITESFCSTQK											
	TVFGDTNSFA	SKGGLRAMGN	AFDKGMVLVL	SIWDDHEAKM	LWLDSNYPLD	KSASAPGVAR	GTCATTSGEP	KDVESQSPNA	QVIFSNIKYG	DIGSTYSN										
SEQ ID NO:38	MYRAIATASA	LIAAVRAQQV	CSLTQESKPS	LNWSKCTSSG	CSNVKGSVTI	DANWRWTHQV	SGSTNCYTGN	KWDTSVCTSG	KVCAERCCCLD											
	GADYASTYGI	TSSGDQLSLS	FVTKGPYSTN	IGSRTYLMED	ENTYQMFQLL	GNEFTFDVDV	SNIGCGLNGA	LYFVSMDADG	GKAKYPGNKA											
	GAKYGTGYCD	AQCPRDVKFI	NGQANSDBGWQ	PSDSDVNGGI	GNLGTCCPEM	DIWEANSIST	AYTPHPCTKL	TQHSCTGDSC	GGTYSNDRYG											
	GTCDDAGCDE	NSYRQGNKTF	YGP GSGFNVD	TTKKVTVVVTQ	FHKGSNGRLS	EITRLYVQNG	KVIANSESKI	AGVPGNSLTA	DFCTKQKKVF											
	NDPDDFTKKG	AWSGMSDALE	APMVIVMSLW	HDHHSNMLWL	DSTYPTDSTK	LGSQRGSCST	SSGVPADLEK	NVPNSKVAFS	NIKEFGPIGST											
	YKSDGTTPTN	PTNPSEPSNT	ANPNPGTVDQ	WGQC GGSNYS	GPTACKSGFT	CKKINDEFSQ	CQ													
SEQ ID NO:39	MYSAAVLATF	SFLLGAGAQQ	VGTLKTESHP	PLTIQKCAAG	GTCTDEADSV	VLDANWRWLH	STSGSTNCYT	GNTWDTTLCP	DAATCTANCA											
	FDGADYEGTY	GITSSGDSLK	LSFVTGSNVG	SRTYIMDSET	TYKEFALLIGN	EFTFTVDVSK	LPCGLNGALY	FVPMADAGGM	SKYPTNKAGA											
	KYGTGYCDAQ	CPQDMKFVSG	GANN EGWVPD	SNSANS GTGN	IGSCCSEFDV	WEANSM SQAL	TPHTCTVDGQ	TACTGDDCAG	NTGVCDDAGC											
	DFNPYRMGNT	TFYGS GKTID	TTKPF SVVTQ	FITDDGTETG	TLTEIKRFYV	QDDVVYEQPN	SDISGVSGNS	ITDDECTAQK	TAFGDTDYFS											
	QKGGMAAMGK	KMADGMVLVL	SIWDDYNVNM	LWLDSDYPTT	KDASTPGVSR	GSCATTSGVP	ATVEAASGSA	YVTFESSIKYG	PIGSTFKAPA											
	DSSSPV VASS	SPA AVAAVVS	TSSAQAVPSH	PAVSSSQAAV	STPEAVSSAP	EVPASSSAAQ	SVAPTSTKPK	CSKVSQSSTL	ATSVAAPATT											
	ATSAAVAATS	AASSSGSVPL	YG NCTGGKTC	SEGT CVVQNP	WYSQC VASS															
SEQ ID NO:40	MFRAAALLAF	TCLAMVSGQQ	AGNTAENHP	QLSQQQCTTS	GGCKPLSTKV	VLDSNWRWVH	STSGYTNCYT	GNEWDTSLCP	DGKTC AANCA											
	LDGADYSGTY	GITSTGTALT	LKFVTGSNVG	SRVYLMADDT	HYQLLKLINQ	EFTFDVDMSN	LPCGLNGALY	LSAMDADGGM	SKYPGNKAGA											
	KYGTGYCDSQ	CPKDIK FING	EANVG NWTET	GSNTGTGSYG	TCCSEMDIWE	ANNDAAAF TP	HPCTTTGQTR	CSGDDCARNT	GLCDGDGCDE											
	NSFRMGDKTF	LGKGMTVDTS	KPFTVVVTQFL	TNDNTSTGTL	SEIRRIYIQN	GKVIQNSVAN	IPGVDPVNSI	TDNFECAQQKT	AFGDTNWF AQ											
	KGGLKQMGEA	LGNGMVLALS	IWDDHAANML	WLDSDYPTDK	DPSAPGVARG	TCATTSGVPS	DVESQVPNSQ	VVFSNIKEGD	IGSTFSGTSS											
	PNPPGGSTTS	SPVTTSP TTP	PTGPTVPQWG	QCGGIGYSGS	TTCASPYTCH	VLNPIYSQCY														
SEQ ID NO:41	MYRKLAVIS A	FLATARAQSA	CTLQSETHPP	LTWQKCSSGG	TCTQQTG SVV	IDANWRWTHA	TNSSTNCYDG	NWSSSTLC PD	NETCAKNCC L											
	DGAAYASTYG	VTTSGNSLSI	GFVTQSAQKN	VGARLYLMAS	DTTYQEFTLL	GNEFSFDVDV	SQLPCGLNGA	LYFVSMDADG	GVS KYPTNTA											
	GAKYGTGYCD	SQCPRDLKFI	NGQANVEGWE	PSSNNANTGI	GGHGSCCSEM	DIWEANSISE	ALTPHPCTTV	GQEICEGDGC	GGTYS DNRYG											
	GTCDDPDGCDW	NPYRLGNTSF	YGP GSSFTLD	TTKKLTVVVTQ	FETSGAINRY	YVQNGVTFQQ	PNAELGSYSG	NELNDDDYCTA	EEAEFGGSSF											
	SDKGGLTQFK	KATSGGMVLV	MSLWDDYYAN	MLWLDSTYPT	NETSSTPGAV	RGSCSTSSGV	PAQVESQSPN	AKVTFSNIKF	GPIGSTGNPS											
	GGNPPGGNRG	TTTTRRPATT	TGSSPGPTQS	HYGQC G GIGY	SGPTVCASGT	TCQVLNPYYS	QCL													
SEQ ID NO:42	MPSTYDIYKK	LLLLASFLSA	SQAQQVGTSK	AEVHPSLTWQ	TCTSGGSC TT	VNGKVVVDAN	WRWVHNVDGY	NNCYTGNTWD	TTLCPDDETC											
	ASNCALEGAD	YSGTYGV TTS	GNSLRINFVT	QASQKNIGSR	LYLMEDDSTY	KMFKLLNQE F	TFDVDVSNLP	CGLNGAVYFV	SMDADGGMAK											
	YPANKAGAKY	GTGYCDSQCP	RDLK FINGMA	NVEGWEP SAN	DANAGTGNHG	SCCAEMDIWE	ANSISTAYTP	HPCDTPGQVM	CTGDS CGGTY											
	SSDRYGGTCD	PDGCDFNSYR	QGNKTFYGP G	MTVDTKSKIT	VVTQFLTNDG	TASGTLSEIK	RFYVQNGKVI	PNSESTWSGV	SGNSIT TAYC											
	NAQKTLFGDT	DVFTKHGGME	GMGAALAE GM	VLVLSLWDDH	NSNMLWLD SN	YPTDKPSTTP	GVARGSCDIS	SGDPKDVEAN	DANAYVVYSN											
	IKVGPIGSTF	SGSTGGGSSS	STTATSKTTT	TSATKTTT TTT	TKTTTTTTSAS	STSTGGAQHW	AQCGGIGWTG	PTTCVAPYTC	QKQNDYYSQC	L										

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:43	MISKVLAFS LLAARAQQA GTLTETETHP LSVSQCTASG CTTSAQSI VV DANRWLHST TGSTNCY TGN TWDKTLCPDG ATCAANCALD GADYSGVYGI TTSGNSIKLN FVTKGANTNV GSRTYILMAAG STTQYQMLKL LNQEFTFDVD VSNLPCGLNG ALYFAAMDAD GGLSRFPPTNK AGAKYGTGYC DAQCPQDIKF INGVANSVGW TPSSNDVNAG AGQYGSCCSE MDIWEANKIS AAYTPHPCSV DTQTRCTGTD CGIGARYSSL CDADGCDFNS YRQGN TSYFG AGLTVNTNKV FTVVTQFITN DGTASGTLKE IRRFYVQNGV VIPNSQSTIA GVPGN SITDS FCAAQKTAFG DTNEFATKGG LATMSKALAK GMVLVMSIWD DHTANMLWLD APYPATKSPS APGVTRGSCS ATSGNPVDVE ANSPGSSVTF SNIKWGPINS TYTGSGAAPS VPGTTTVSSA PASTATSGAG GVAKYAQCGG SGYSGATA CV SGSTCVALNP YYSQCQ
SEQ ID NO:44	MFPAAATLFAF SLFAAVYGOQ VGTQLAETHP RL TWQKCTRS GGCQTQ SNGA IVLDANRWV HNVGGYTNCY TGNTWNTSLC PDGATCAKNC ALDGANYQST YGITTSGNAL TLKFVTQSEQ KNIGSRVYLL ESDTKYQLFN PLNQEFTFDV DVSQLP CGLN GAVYFSAMDA DGM SKFPNN AAGAKYGTGY CDSQCPRDIK FINGEANVQG WQSPNDTNA GTGNYGACCN EMDVWEANSI STAYTPHPCT QQGLVRCSGT ACGGGSNRYG SICDPDGCDF NSFRMGDKSF YGPGLTVNTQ QKFTVVVTQFL TNNSSSGTL REIRRLYVQN GRVIQNSKVN IPGMPSTMDS VTTEFCNAQK TAFNDTFSFQ QKGGMANMSE ALRRGMVLVL SIWDDHAANM LWLDSNYPTD RPASQPGVAR GTCPTSSGKP SDVENSTANS QVIYSNIKFG DIGSTYSA
SEQ ID NO:45	MKGSISYQIY KGALLLSALL NSVSAQQVGT LTAETHPALT WSKCTAGXCS QVSGSVVIDA NWPXVHSTSG STNCYTGN TW DATLCPDDVT CAANCAVDGA RRQHLRV TTS GNSLRINFVT TASQKNIGSR LYLLENDTTY QKFNL LNQEF TFDVDVSNLP CGLNGALYFV DMDADGGMAK YPTNKAGAKY GTGYCDSQCP RDLKFINGQA NVDGWTPSKN DVNSGIGNHG SCCAEMDIWE ANSISNAVTP HPCDTPSQTM CTGQRCGGTY STDRYGGTCD PDGCDENPYR MGVTNFYGPG ETIDTKSPFT VVTQFLTNDG TSTGTLSEIK RFYVQGGKVI GNPQSTIVGV SGNSITDSWC NAQKSAFGDT NEFSKHGGMA GMGAGLADGM VLVMSLWDDH ASDMLWLDST YPTNATSTTP GAKRGTCDIS RRPNTVESTY PNAYVIYSNI KTGPLNSTFT GGTSSSSSTT TTTSKSTSTS SSSKTTTTVT TTTTSSGSSG TGARDWAQCG GNGWTGPTTC VSPYTCTKQN DWYSQCL
SEQ ID NO:46	MFRTAALTAF TLAAVVLGOQ VGTLTAENHP ALSIQOCTAS GCTTQQKS VV LDSNWRWTHS LPVHTN CYTG NAWDASLCPD PTTCATNCAI DGADYSGTYG ITTSGNALTL RFVTNGPYSK NIGSRVYLLD DADHYKMF DL KNQEFTFDVD MSGLPCGLNG ALYFSEMPAD GGKAAHTSNK AGAKYGTGYC DAQCPHDIKW INGEANILDW SASATDANAG NGRYGACCAE MDIWEANSEA TAYTPHVCRD EGLYRCSGTE CGDGDNR YGG VCDKDGCDFN SYRMGDKNFL GRGKTIDTTK KITVVVTQFIT DDNTSSGNLV EIRRVYVQDG VTYQNSFSTF PSLSQYNSIS DDFCVAQKTL FGDNQYYNTH GGTEKMGDAM ANGMVLIMSL WSDHAAHMLW LDSDYPLDKS PSEPGVSRGA CATTGDPDD VVANHPNASV TFSNIKYGPI GSTYGGSTPP VSSGNTSAPP VTSTTSSGPT TPTGPTGTVP KWGQC GGNGY SGPTTCVAGS TCTYSNDWYS QCL
SEQ ID NO:47	MYQRALLFSA LLSVSR AQQA GTAQEEVHPS LTWQRC EASG SCTEVAGSVV LDSNWRWTHS VDGYTN CYTG NEWDATLCPD NESCAQNCAV DGADYEATYG ITSNGDSLTL KFVTGSNVGS RVYLMEDDET YQMF DLNNE FTFDVDVSNL PCGLNGALYF TSMADAGGLS KYEGNTAGAK YGTGYCDSQC PRDIKFINGL GNVEGW EPSD SDANAGVGGM GTCCPEMDIW EANSISTAYT PHPCDSVEQT MCEGDSGGT YSDDRYGGTC DPDGCDFNSY RMGN TSYFGP GAIIDTSSKF TVVTQFIADG GSLSEIKRFY VQNGEVIPNS ESNISGVEGN SITSEFCTAQ KTAFGDEDIF AQHGGLS AMG DAASAMVLIL SIWDDHHSSM MWLDSSYPTD ADPSQPGVAR GTCEQAGADP DVVESEHADA SVTFSNIKFG PIGSTF
SEQ ID NO:48	MYRAIATASA LIAAVRAQQV CSLTTETKPA LTWSKCTSSG CSNVQGSVTI DANRWTHQV SGSTNCHTGN KWDTSVCTSG KVCAEKCCVD GADYASTYGI TSSGNQLSLS FVTKGSYGTN IGSRTYLMED ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSMDADG GKAKYPGNKA GAKYGTGYCD AQCPRDV KFI NGQANS DGWE PSKSDVNGGI GNLGTCCPEM DIWEANSIST AYTPHPCTKL TQHACTGDSC GGTYSNDRYG GTC DADGCDF NAYRQGNKTF YGP GSGFNVD TTKKVTVVTQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGNPGSSLTS DFCTTQKKVF GDIDDFAKG AWNGMSDALE APMVLVMSLW HDHHSNMLWL DSTYPTDSTA LGSQRGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YNKEGTQPQ TNPTNPNTN PTNPGTVDQW GQC GG TNYSG PTACKSPFTC K KINDFYSQC Q

TABLE 7

TABLE 7																				
SEQ ID NO.	Amino acid sequence																			
SEQ ID NO:49	MFRTAALTAF	TLAAVVLGQQ	VGTLAAENHP	ALSIQQCTAS	GCTTQQKSVV	LDSNWRWTHS	TAGATNCYTG	NAWDSSLCPN	PTTCATNCAI											
	DGADYSGTYG	ITTSGNSLTL	RFVTNGQYSE	NIGSRVYLLD	DADHYKLFNL	KNQEFTFDVD	MSGLP	PCGLNG	ALYFSEMAAD	GGKAAHTGNN										
	AGAKYGTGYC	DAQCPHDIKW	INGEANILDW	SGSATDPNAG	NGRYGACCAE	MDIWEANSEA	TAYTPHVCRD	EGLYRCSGTE	CGDGDNR	YGG										
	VCDKDGCDEN	SYRMGDKNFL	GRGKTIDTTK	KITVVTQFIT	DDNTPGTGNLV	EIRRVYVQDG	VTYQNSFSTF	PSLSQYNSIS	DDFCVAQKTL											
	FGDNQYYNTH	GGTEKMGDSL	ANGMVLIMSL	WSDHAAHMLW	LDSDYPLDKS	PSEPGVSRGA	CATTGDPDD	VVANHPNASV	TFSNIKYGPI											
	GSTYGGSTPP	VSSGNTSVPP	VTSTTSSGPT	TPTGPTGTVP	KWGQCGGIGY	SGPTSCVAGS	TCTYSNEWYS	QCL												
SEQ ID NO:50	MYQKLALISA	FLATARAQSA	CTLQAETHPP	LTWQKCSSGG	TCTQQTGSVV	IDANWRWTHA	TNSSTNCYDG	NTWSSTLCPD	NETCAKNCC	L										
	DGAAYASTYG	VTTSADSLSI	GFVTQSAQKN	VGARLLYLMAS	DDTYQEFTLL	GNEFSFDVDV	SQLPCGLNGA	LYFVSMADAG	GVTKYPTNTA											
	GAKYGTGYCD	SQCPRDLKFI	NGQANVEGWE	PSSNNANTGI	GGHGSCCSEM	DIWEANSISE	ALTPHPCTTV	GQEICEGDSC	GGTYS	GDRYG										
	GTCDDPDGCDW	NPYRLGNTSF	YGPSSFTLD	TTKKLTVVTQ	FETSGAINRY	YVQNGVTFQO	PNAELGDYSG	NSLDDDYCAA	EEAEFGG	SSF										
	SDKGGLTQFK	KATSGGMVLV	MSLWDDYYAN	MLWLDSTYPT	DETSSTPGAV	RGSSSTSSGV	PAQLESNSPN	AKVVYSNIKF	GPIGSTGNPS											
	GGNPPGGNPP	GTTTPRPATS	TGSSPGPTQT	HYGQCGGIGY	IGPTVCASGS	TCQVLNPIYS	QCL													
SEQ ID NO:51	MTWQSCTAKG	SCTNKNGKIV	IDANWRWLHK	KEYDNCYTG	NEWDATACPD	NKACAANCAV	DGADYSGTYG	ITAGSNSLKL	KFITKGSYST											
	NIGSRTYLMK	DDTYEMFKF	TGNQEFTFDV	DVSNLPCGFN	GALYFVSMDA	DGGLKKYSTN	KAGAKYGTGY	CDAQCPRDLK	FINGEGNVEG											
	WKPSSNDANA	GVGGHGSCCA	EMDIWEANSV	STAVTPHSCS	TIEQSRCDGD	GCGGTYSADR	YAGVCDPDGC	DFNSYRMGVK	DFYGKGKTVD											
	TSKKFTVVVTQ	FIGHTDAMEI	KRFYVQNGKT	IAQPASAVPG	VEGNSITTFE	CDQQKAVFGD	TYTFKDKGGM	ANMAKALANG	MVLVMSLWDD											
	HYSNMLWLDS	TYPTDKNPDT	DLGTGRGECE	TSSGVPADVE	SQHADATVVY	SNIKFGPLNS	TFG													
SEQ ID NO:52	MASAISEQVY	RSALILSAFL	PSITQAAQIG	TYTTETHPSM	TWETCTSGGS	CATNQGSVVM	DANWRWVHQV	GSTTNCYTGN	TWDTSICD	T										
	ETCATECAVD	GADYESTYGV	TTSGSQIRLN	FVTQNSNGAN	VGSRLYMMAD	NTHYQMFKLL	NQEFTFDVDV	SNLPCGLNGA	LYFVTMDEDG											
	GVSKYPNNKA	GAQYGVGYCD	SQCPRDLKFI	QGQANVEGWT	PSSNNENTGL	GNYGSCCAEL	DIWESNSISQ	ALTPHPCDTA	TNTMCTGDAC											
	GGTYSSDRYA	GTCDDPDGCD	NPYRMGNTTF	YGPCKTIDTN	SPFTVVVTQFI	TDDGTDGTGL	SEIRRYVQVN	GVTYAQPDSD	ISGITGNAIN											
	ADYCTAENTV	FDGPGTFAKH	GGFSAMSEAM	STGMVLVMSL	WDDYYADMLW	LDSTYPTNAS	SSTPGAVRGS	CSTD	SGVPAT	IESESPDSYV										
	TYSNIKVGPI	GSTFSSGSGS	GSSGSGSGSGS	ASTSTTSTKT	TAATSTSTAV	AQHYSQCGGQ	DWTGPTTCVS	PYTCQVQNAY	YSQCL											
SEQ ID NO:53	MKAYFEYIVA	ALPLLGLATA	QQVGKQTTET	HPKLSWKKCT	GKANCNTVNA	EVVIDSNWRW	LHDSSGKNCY	DGNKWTSA	S	ATDCASKCQ										
	LDGANYGTTY	GASTSGDAL	LKFVTKHEYG	TNIGSRFYLM	NGASKYQMFT	LMNNEFAFDV	DLSTVECGLN	AALYFVAMEE	DGGMASYSSN											
	KAGAKYGTGY	CDAQCARDLK	FVGGKANIEG	WTPSTNDANA	GVGPYGGCCA	EIDVWESNAH	SFAFTPHACK	TNKYHVCCER	NCGGTYSEDR											
	FAGLCDANGC	DYNPYRMGNT	DFYGKGKTVD	TSKKFTVVVSR	FEENKLTQFF	VQNGQKIEIP	GPKWDGIPSD	NANITPEFCS	AQFQAFGDRD											
	RFAEVGGFAQ	LNSALRMPMV	LVMSIWDDHY	ANMLWLDSVY	PPEKEGQPGA	ARGDCPQSSG	VPAEVESQYA	NSKVVIYSNIR	FGPVGSTVNV											
SEQ ID NO:54	MFSKFALTGS	LLAGAVNAQG	VGTTQQTETHP	QMTWQSC	TSP	SSCTTNQGEV	VIDSNWRWVH	DKDGYVNCYT	GNTWNTT	LCP										
	LDGADYSSTY	GITTSGNALS	LQFVTQSSGK	NIGSRTYLME	SSTKYHLFDL	IGNEFAFDVD	LSKLPCGLNG	ALYFVTMDAD	GGMAKYSTNT											
	AGAEYGTGYC	DSQCPRDLKF	INGQGNVEGW	TPSTNDANAG	VGGLGCC	SE	MDVWEANSMD	MAYTPHPCET	AAQHSCNADE	CGGTYS	SSRY									
	AGDCDDPGCD	WNPFMRGNKD	FYGS	GDTVD	T	SQKFTVVVTQF	HGSGSSLTEI	SQYIIQGGTK	IQQPNSTWPT	LTGYN	SITDD	FCKAQKVEFN								
	DTDVFSEKGG	LAQMGAGMAD	GMVLVMSLWD	DHYANMLWLD	STYPVDADAS	SPGKQ	RG	TCA	TTSGVPADVE	SSDASATVIY	SNIKFGPIGA	TY								

TABLE 7

TABLE 7																				
SEQ ID NO.	Amino acid sequence																			
SEQ ID NO:55	MFPAALLSF	TLLAVASAAQ	IGTNTAEVHP	SLTVSQCTTS	GGCTSSTQSI	VLDANWRWLH	STSGYTNCTY	GNQWNSDLCP	DPDTCATNCA											
	LDGASYESTY	GISTDGNVAV	LNFTVQGSQT	NVGSRVYLLS	DDTHYQTFSL	LNKEFSFDVD	ASNIGCGING	AVYFVQMDAD	GGLSKYSSNK											
	AGAQTGTGYC	DSQCPQDIKF	INGEANLLDW	NATSANSGTG	SYGSCCPEMD	IWEANKYAAA	YTPHPCSVSG	QTRCTGTSCG	AGSERYDGYC											
	DKDGCDFNSW	RMGNETFLGP	GMTIDTNKKF	TIVTQFITDD	NTANGTLSEI	RRLYVQGGTV	IQNSVANQPN	IPKVNSTITDS	FCTAQKTEFG											
	DQDYFGTIGG	LSQMKGAMSD	MVLVMSIWDD	YDAEMLWLDS	NYPTSGSAST	PGISRGPCSA	TSGLPATVES	QQASASVTYS	NIKWGDIGST											
	YSGSGSSGSS	SSSSSSAASA	STSTHTSAAA	TATSSAAAAT	GSPVPAYGQC	GGQSYTGSTT	CASPYVCKVS	NAYYSQCCLPA												
SEQ ID NO:56	MKRALCASLS	LLAAAVAQQV	GTNEPEVHPK	MTWKKCSSGG	SCSTVNGEVV	IDGNWRWIHN	IGGYENCYSG	NKWTSVCSTN	ADCATKCAME											
	GAKYQETYGV	STSGDALTLK	FVQQNSSGKN	VGSRMYLMNG	ANKYQMFTLK	NNEFAFDVDL	SSVECGMNSA	LYFVPMKEDG	GMSTEPNNKA											
	GAKYGTGYCD	AQCARDLKFI	GGKGNIEGWQ	PSSTDSSAGI	GAQGACCAEI	DIWESKNNAF	AFTPHPCENN	EYHVCTEPNC	GGTYADDDRYG											
	GGCDANGCDY	NPYRMGNPDF	YGPCKTIDTN	RKFTVISRFE	NNRNYQIIMQ	DGVAHRIPGP	KFDGLEGETG	ELNEQFCTDQ	FTVFDERNRF											
	NEVGWWSKLN	AAYEIPMVLV	MSIWSDDHFAN	MLWLDSTYPP	EKAGQPGSAR	GPCPADGGDP	NGVVNQYPNA	KVIWSNVREF	PIGSTYQVD											
SEQ ID NO:57	MQLTKAGVFL	GALMGGAAAQ	QVGTQTAENH	PKMTWKKCTG	KASCTTVNGE	VVIDANWRWL	HDASSKNCYD	GNRWTDSCRT	ASDCAAKCSL											
	EGADYAKTYG	ASTSGDALSL	KFVTRHDYGT	NIGSRFYLMN	GASKYQMFSL	LGNEFAFDVD	LSTIECGLNS	ALYFVAMEED	GGMKSYSSNK											
	AGAKYGTGYC	DAQCARDLKF	VGGKANIEGW	KPSSNDANAG	VGPYGACCAE	IDVWESNAHA	FAFTPHPCTD	NKYHVCQDSN	CGGTYSDDRF											
	AGKCDANGCD	INPYRLGNTD	FYGKGKTVDT	SKKFTVVTRF	ERDALTOFFV	QNNKRIDMPS	PALEGLPATG	AITAEYCTNV	FNVFGDNRNF											
	DEVGWSQLQ	QALSLPMVLV	MSIWDDHYSN	MLWLDSVYPP	DKEGSPGAAR	GDCPQDSGVP	SEVESQIPGA	TVVWSNIREG	PVGSTVNV											
SEQ ID NO:58	MYRIVATASA	LIAAARAQQV	CSLNTETKPA	LTWSKCTSSG	CSDVKGSVVI	DANWRWTHQT	SGSTNCYTGN	KWDTSICTDG	KTCAEKCCCLD											
	GADYSGTYGI	TSSGNQLSLG	FVTNGPYSKN	IGSRTYLMEN	ENTYQMFQLL	GNEFTFDVDV	SGIGCGLNGA	PHFVSMDEDG	GKAKYSGNKA											
	GAKYGTGYCD	AQCPRDVKFI	NGVANSEGWK	PSDSDVNAGV	GNLGTCCPEM	DIWEANSIST	AFTPHPCTKL	TQHSCTGDSC	GGTYSSDDRYG											
	GTCADADGCD	NAYRQGNKTF	YGPGSNFNID	TTKKMTVVVTQ	FHKGSNGRLS	EITRLYVQNG	KVIANSESKI	AGNPGSSLTS	DFCSKQKSVF											
	GDIDDFSKKG	GWNGMSDAL	APMVLVMSLW	HDHHSNMLWL	DSTYPTDSTK	VGSQRGSCAT	TSGKPSDLER	DVPNSKVVSFS	NIKFGPIGST											
	YKSDGTTPNP	PASSSTTGSS	TPTNPPAGSV	DQWGQCQGGQN	YSGPTTCKSP	FTCKKINDFY	SQCQ													
SEQ ID NO:59	MYQRALLFSA	LATAVSAQQV	GTQKAEVHPA	LTWQKCTAAG	SCTDQKGSVV	IDANWRWLHS	TEDTTNCTYG	NEWNAELCPD	NEACAKNCAL											
	DGADYSGTYG	VTADGSSLKL	NFVTSANVGS	RLYLMEDDET	YQMFENLLNNE	FTFDVDVSNL	PCGLNGALYF	VSMADAGGLS	KYPGNKAGAK											
	YGTGYCDSQC	PRDLKFINGE	ANVEGWKPSD	NDKNAGVGGY	GSCCPEMDIW	EANSISTAYT	PHPCDGMEQT	RCDGNDCCGT	YSSTRYAGTC											
	DPDGCDFNSF	RMGNESFYGP	GGLVDTKSPI	TVVTQFVTAG	GTDSGALKEI	RRVYVQGGKV	IGNSASNAG	VEGDSITSDF	CTAQKKAFGD											
	EDIFSKHGG	EGMGKALNKM	ALIVSIWDDH	ASSMMWLDST	YPVDADASTP	GVARGTCEHG	LGDPETVESQ	HPDASVTFSN	IKFGPIGSTY KSV											
SEQ ID NO:60	MSALNSFNMY	KSALLIGSL	ATAGAQQIGT	YTAETHPSLS	WSTCKSGGSC	TTNSGAITLD	ANWRVWHGVN	TSTNCYTGNT	WNTAICDTDA											
	SCAQDCALDG	ADYSGTYGIT	TSGNSLRINF	VTGSNVGSRT	YLMADNTHYQ	IFDLLNQEFT	FTVDVSNLPC	GLNGALYFVT	MDADGGVSKY											
	PNNKAGAQYG	VGYCDSQCPR	DLKFIAGQAN	VEGWTPTSTNN	SNTGIGNHGS	CCAELDIWEA	NSISEALTPH	PCDTPGLTVC	TADDCGGTYS											
	SNRYAGTCDP	DGCDFNPYRL	GVTDFYGS GK	TVDTTKPFTV	VTQFVTDDGT	SSGSLSEIRR	YYVQNGVVIP	QPSSKISGIS	GNVINSDFCA											
	AELSAFGETA	SFTNHGGLKN	MGSALEAGMV	LVMSLWDDYS	VNMLWLDSTY	PANETGTPGA	ARGSCPTTSG	NPKTVESQSG	SSYVVFSDIK											
	VGPENSTFSG	GTSTGGSTTT	TASGTTSTKA	STTSTSSTST	GTGVAAHWGQ	CGGQGWGTGPT	TCASGTTCTV	VNPYYSQCL												

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:61	MRTAKFATLA ALVASAAAQQ ACSLTTERHP SLSWNKCTAG GQCQTVQASI TLDSNWRWTH QVSGSTNCYT GNKWDTSICT DAKSCAQNCC VDGADYTSTY GITTNGDLSL LKFVTKGQHS TNVGSRTYLM DGEDKYQTFE LLGNEFTFDV DVSNIGCGLN GALYFVSMDA DGGLSRYPGN KAGAKYGTGY CDAQCPRDIK FINGEANIEG WTGSTNDPNA GAGRYGTCCS EMDIWEANNM ATAFTHPCT IIGQSRCEGD SCGGTYSNER YAGVCDPDGC DFNSYRQGNK TFYGGKMTVD TTKKITVVVTQ FLKDANGDLG EIKRFYVQDG KIIPNSESTI PGVEGNSITQ DWCDRQKVAF GDIDDFNRKG GMKQMGKALA GPMVLVMSIW DDHASNMLWL DSTFPVDAAG KPGAERGACP TTSGVPAEVE AEAPNSNVVF SNIRFGPIGS TVAGLPGAGN GGNNGGNPPP PTTTTSSAPA TTTTASAGPK AGRWQQCGGI GFTGPTQCEE PYICTKLNDW YSQCL
SEQ ID NO:62	MMYKKFAALA ALVAGASAAQQ ACSLTAENHP SLTWKRCTSG GSCSTVNGAV TIDANWRWTH TVSGSTNCYT GNQWDTSLCT DGKSCAQTC VDGADYSSTY GITTSGDSL N LKFVTKHQYG TNVGSRVYLM ENDTKYQMFEE LLGNEFTFDV DVSNLGCGLN GALYFVSMDA DGGMSKYSGN KAGAKYGTGY CDAQCPRDLK FINGEANVGN WTPSTNDANA GFGRYGSCCS EMDVWEANNM ATAFTHPCT TVGQSRCEAD TCGGTYSDDR YAGVCDPDGC DFNAYRQGD K TFYGGKMTVD TNKKMTVVVTQ FHKNSAGVLS EIKRFYVQDG KIIANAESKI PGNPGNSITQ EYCDAAQKVAF SNTDDFNKRG GMAQMSKALA GPMVLVMSVW DDHYANMLWL DSTYPIDQAG APGAERGACP TTSGVPAEIE AQVPNSNVIF SNIRFGPIGS TVPGLDGSNP GNPTTTVVPP ASTSTSRPTS STSSPVSTPT GQPGGCTTQK WGQCGGIGYT GCTNCVAGTT CTQLNPWYSQ CL
SEQ ID NO:63	MASLSLKIC RNALILSSVL STAQGGQVGT YQTEHPSMT WQTCGNGGSC STNQGSVVLD ANWRWVHQTG SSSNCYTGNK WDTSYCSTND ACAKKCALDG ADYSNTYGIT TSGSEVRINF VTSNSNGKNV GSRVYMMADD THYEVYKLLN QEFTFDVDVS KLPCGLNGAL YFVVMADAGG VSKYPNNKAG AKYGTGYCDS QCPRDLKFIQ GQANVEGWVS STNNANTGTG NHGSCCAELD IWESNISQA LTPHPCDTPT NTLCTGDACG GTYSSTDYSG TCDPDGCDEN PYRVGNITFY GPCKTIDTNK PITVVTQFIT DDGTSSGTL EIKRFYVQDG VTYPQPSADV SGLSGNTINS EYCTAENTLF EGSGSFAKHG GLAGMGEAMS TGMVLVMSLW DDYANMLWL DSNYPTNEST SKPGVARGTC STSSGVPSEV EASNPSAYVA YSNIKVGPIG STFKS
SEQ ID NO:64	MYRAIATASA LIAAVRAQQV CSLTPETKPA LWSKCTSSG CSNVQGSVTI DANWRWTHQL SGSTNCYTGN KWDTSICTSG KVCAEKCCID GAEYASTYGI TSSGNQLSLS FVTKGAYGTN IGSRTYLMED ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSMDADG GKAKYPGNKA GAKYGTGYCD AQCPRDVKFI NGQANSDGWQ PSKSDVNAGI GNMGTCCPEM DIWEANSIST AYTPHPCTKL TQHSCTGDSC GGTYSNDRYG GTCDDADGDF NAYRQGNKTF YGPGSGFNVD TTKKVTVVVTQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGVPGSSLTP EFCTAQKKVF GDTDDFAKKG AWSGMSDALE APMVLVMSLW HDHHSNMLWL DSTYPTDSTK LGAQRGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YKEGVPEPTN PTNPTNPTNP TNPGTVDQWA QCGGTNYSGP TACKSPFTCK KINDEFYSQCQ
SEQ ID NO:65	MFPKSSLLVL SFLATAYAQQ VGTQTAEVHP SLNWARCTSS GCTNVAGSVT LDANWRWLHT TSGYTNCYTG NSWNTTLCPD GATCAQNICAL DGANYQSTCG ITTSGNALT L KFVTQGEQKN IGSRVYLMAS ESRYEMFGLL NKEFTFDVDV SNLPCGLNGA LYFSSMDADG GMAKNPGNKA GAKYGTGYCD SQCPRDIKFI NGEANVAGWN GSPNDTNAGT GNWGACCNEM DIWEANSISA AYTPHPCTVQ GLSRCSGTAC GTNDRYGTVC DPDGCDFNSY RMGDKTYYGP GGTGVDTRSK FTVVTQFLTN NNSSSGTLSE IRRLYVQNGR VVQNSKVNIP GMSNTLDSIT TGFCDSQKTA FGDTRSFQNK GGMSAMGQAL GAGMVLVLSV WDDHAANMLW LDSNYPVDAD PSKPGIARGT CSTTSGKPTD VEQSAANSSV TFSNIKFEDI GTTYTGGSVT TTPGNPGTTT STAPGAVQTK WGQCQGGQWT GPTRCESGST CTVVNQWYSQ CI
SEQ ID NO:66	MFRKAALLAF SFLAIAHGQQ VGTNQAENHP SLPSQHCTAS GCTTSSTSVV LDANWRWVHT TTGYTNCYTG QTDASICPD GVTCAKACAL DGADYSGTYG ITTSGNALT L QFVKGTNVGS RVYLLQDASN YQLFKLINQE FTFDVDMSNL PCGLNGAVYL SQMDQDGGVS RFPTNTAGAK YGTGYCDSQC PRDIKFINGE ANVAGWTGSS SDPNSGTGNY GTCCSEMDIW EANSVAAAAYT PHPCSVNQQT RCTGADCGQD ANRYKGVCDP DGCDFNSFRM GDQTFLGKGL TVDTSRKFTI VTQFISDDGT SSGNLAEIRR FYVQDGKVIP NSKVNIAAGCD AVNSITDKFC TQKTAFGDT NRFADQGGGLK QMGAAALKSGM VLALSLWDDH AANMLWLDS YPTTADASKP GVARGTCPNT SGVPKDVESQ SGSATVTYSN IKWGDNLSTF SGTASNPTGP SSSPSGPSS SSSTAGSQPT QPSSGSVAQW GQCQGGIGYS ATGCVSPYTC HVVNPYYSQC Y

TABLE 7

SEQ ID NO.		Amino acid sequence	
SEQ ID NO:67		TETHPRLTWK RCTSGGNCST VNGAVTIDAN WRWHTVSGS TNCYTGNEWD TSICSDGKSC AQTCCVDGAD YSSTYGITTS GDSLNLKFVT KHQHGTVNGS RYVLMENDTK YQMFELLGNE FTFDVDVSNL GCGLNGALYF VSMADAGGMS KYSGNKAGAK YGTGYCDAQC PRDLKFINGE ANIENWTPST NDANAGFGRY GSCCSEMDIW EANNMATAFT PHPCTIIGQS RCEGNSCGGT YSSERYAGVC DPDGCDFNAY RQGDKTFYGK GMTVDLTKKM TVVTQFHKNS AGVLSEIKRF YVQDGKIIAN AESKIPGNPG NSITQEWCDQ KQVAFGDIDD FNRKGGMAQM SKALEGPMVL VMSVWDDHYA NMLWLDSTYP IDKAGTPGAE RGACPTTSGV PAEIEAQVPN SNVIFSNIRF GPIGSTVPGL DGSTPSNPTA TVAPPTSTTT SVRSSTTQIS TPTSQPGGCT TQKWGQCGGI GYTGCTNCVA GTTCTELNPW YSQCL	
SEQ ID NO:68		MFHKAVLVAE SLVTIVHGQQ AGQTAEENHP QLSSQKCTAG GSCTSASTSV VLDSNWRWVH TTSGYTNCYT GNTWDASICS DPVSCAQNCA LDGADYAGTY GITTSGDALT LKFVTGSNVG SRVYLMEDET NYQMFKLMNQ EFTFDVDVSN LPCGLNGAVY FVQMDQDGGT SKFPNNKAGA KFGTGCDSDQ CPQDIKFIN GAGDANSGTGS FGTCCQEMDI WEANSISAAY TPHPCTVTEQ TRCSGSDCGQ GSDFRFGICD PDGCDFNSFR MGNTEFYGKG LTVDTSQKFT IVTQFISDDG TADGNLAEIR RFYVQNGKVI PNSVVQITGI DPVNSITEDF CTQQKTVFGD TNNEFAAKGGL KOMGEAVKNG MVLALSLWDD YAAQMLWLD DYP TTTADPSQ PGVARGTCPT TSGVPSQVEG QEGSSSVIYS NIKFGDLNST FTGTLTNPSS PAGPPVTSSP SEPSQSTQPS QPAQPTQPAG TAAQWAQCGG MGFTGPTVCA SPFTCHVLNP YYSQCY	
SEQ ID NO:69		MFRAAALLAE TCLAMVSGQQ AGNTAENHP QLQSQQCCTS GGCKPLSTKV VLDSNWRWVH STSGYTNCYT GNEWNTSLCP DGKTCANCA LDGADYSGTY GITSTGTALT LKFVTGSNVG SRVYLMADDT HYQLLKLINQ EFTFDVDMSN LPCGLNGALY LSAMDADGGM SKYPGNKAGA KYGTGYCDSQ CPKDIKFIN EANVGNWTET GSNTGTGSYG TCCSEMDIWE ANNDAAFTP HPCTTTGQTR CSGDDCARNT GLCDHGDGCD FNSFRMGDKT FLGKGMTVDT SKPFTDVTQF LTNDNTSTGT LSEIRRIYIQ NGKVIQNSVA NIPGVDPVNS ITDNFCAQQK TAFGDTNWFA QKGGCLKMGE ALNGMVLAL SIWDDHAANM SIWDDSDYPTD KDPSAPGVAR GTCATTSGVP SDVESQVPNS QVVF SNIKFG DIGSTFSGTS SPNP PGGSTT SSPVTTSTPT PPTGPTVPQW GQCGGIGYSG STTCASPYTC HVINPYYSQC Y	
SEQ ID NO:70		MMMKQYLQYL AAALPLVGLA AGQRAGNETP ENHPPLTWQR CTAPGNCQTV NAEVVIDANW RWLHDDNMQN CYDGNQWTNA CSTATDCAEK CMIEGAGDYL GTYGASTSGD ALTLKFVTKH EYGTNVGSRF YLMNGPDKYQ MFNLMGNELA FDVDLSTVEC GINSALYFVA MEEDGGMASY PSNQAGARYG TGYCDAQCAR DLKFVGGKAN IEGWKSSSTD PNAGVGPIYS CCAEIDVWES NAYAFATPH ACTTNEYHVC ETTNCGGTYS EDRFAGKCD A NGCDYNPYRM GNPDFYGGK TLDTSRKFTV VSRFEENKLS QYFIQDGRKI EIPPTWEGM PNSS EITPEL CSTMFDVFND RNRFEVGGF EQLNNALRVP MVLVMSIWDD HYANMLWLD IYPPEKEGQP GAARGDCPTD SGVPAEVEAQ FPDAQVVWSN IREGPIGSTY DF	
SEQ ID NO:71		MYRSATFLT F ASLVLGQQVG TYTAERHPSM PIQVCTAPGQ CTRESTEVVL DANWRWTHIT NGYTNCYTGN EWNATACPDG ATCAKNCAVD GADYSGTYGI TTPSSGALRL QFVKKNDNGQ NVGSRVYLMA SSDKYKLFNL LNKEFTFDVD VSKLPCGLNG AVYFSEMLED GGLKSFSGNK AGAKYGTGYC DSQCPQDIKF INGEANVEGW GGADGNSGTG KYGICCAEMD IWEANS DATA YTPHVCSVNE QTRCEGVDCG AGSDRYNSIC DKDGCDFNSY RLGNREFYGP GKTVDTTTRPF TIVTQFVTDG GTDSGNLKSI HRYVVDGNV IPNSVTEVAG VDQTNFISEG FCEQQKSAFG DNNYFGQLGG MRAMGESLKK MVLVLSIWDD HAVNMNWLDS IFPNDADPEQ PGVARGRCDP ADGVPATIEA AHPDAYVIYS NIKFGAINST FTAN	
SEQ ID NO:72		MYRTLAFASL SLYGAARAQQ VGTSTAENHP KLTWQTCTGT GGTNCSNKS G SVVLDSNWRW AHNVGGYTNC YTGNSWSTQY CPDGDSCTKN CAIDGADYSG TYGITTSNNA LSLKFVTKGS FSSNIGSRTY LMETDTKYQM FNLINKETTF DVDVSKLPCG LINGALYFVEM AADGGIGKGN NKAGAKYGTG YCDSQCPHDI KFIN GKANVE GWNPSDADPN GGAGKIGACC PEMDIWEANS ISTAYTPHPC RGVGLQECSD AASC GDGSNR YDGQCDKDCG DFN SYRMGVK DFYGP GATLD TTKMTVITQ FLGSGSSLSE IKRFYVQNGK VYKNSQSAVA GVTGNSITES FCTAQKKAFG DTSSFAALGG L NEMGASLAR GHVLIMSLWG DHAVNMLWLD STYPTDADPS KPGAARGTCP TTSGKPEDVE KNSPDATVVF SNIKFGPIGS TFAQPA	

TABLE 7

TABLE 7																				
SEQ ID NO.	Amino acid sequence																			
SEQ ID NO:73	MYQKLALISA	FLATARAQSA	CTLQAETHPP	LTWQKCSSGG	TCTQQTGSVV	IDANWRWTHA	TNSSTNCYDG	NTWSSTLCPD	NETCAKNCCCL											
	DGAAYASTYG	VTTSAADSLSI	GFVTQSAQKN	VGARLYLMAS	DTTYQEFTLL	GNEFSFDVDV	SQLPCGLNGA	LYFVSMADAG	GVSKYPTNTA											
	GAKYGTGYCD	SQCPRDLKFI	NGQANVEGWE	PSSNNANTGI	GGHGSCCSEM	DIWEANSISE	ALTPHPCTTV	GQEICDGDSC	GGTYSGDRYG											
	GTCDDPDGCDW	NPYRLGNTSF	YGPSSFTLD	TTKKLTVVVTQ	FETSGAINRY	YVQNGVTFQQ	PNAELGDYSG	NSLDDDDYCAA	EAEAFGGSSF											
	SDKGGLTQFK	KATSGGMVLV	MSLWDDYYAN	MLWLDSTYPT	NETSSTPGAV	RGSCSTSSGV	PAQLESNSPN	AKVVYSNIKF	GPIGSTGNSS											
	GGNPPGGNPP	GTTTTRRPAT	STGSSPGPTQ	THYGQCGGIG	YSGPTVCASG	STCQVLNPYY	SQCL													
SEQ ID NO:74	MVDSFSIYKT	ALLLSMLATS	NAQQVGTYTA	ETHPSLTWQT	CSGSGSCTTT	SGSVVIDANW	RWVHEVGGYT	NCYSGNTWDS	SICSTDTTCA											
	SECALEGATY	ESTYGVTTSG	SSLRLNFEVTT	ASQKNIGSRL	YLLADDSTYE	TFKLFNREFT	FDVDVSNLPC	GLNGALYFVS	MDADGGVSRF											
	PTNKAGAKYG	TGYCDSQCPR	DLKFIDGQAN	IEGWEPSSTD	VNAGTGNHGS	CCPEMDIWEA	NSISSAFTAH	PCDSVQQTMC	TGDTCCGGTYS											
	DTTDRYSGTC	DPDGCDFNPY	RFGNTNFYGP	GKTVDNSKPF	TVVTQFITHD	GTDTGTLTEI	RRLYVQNGVV	IGNGPSTYTA	ASGNSITESF											
	CKAEKTLFGD	TNVFETHGGL	SAMGDALGDG	MVLVLSLWDD	HAADMLWLDS	DYPTTSCASS	PGVARGTCPT	TTGNATYVEA	NYPNSYVTYS											
	NIKEFTLNST	YSGTSSGGSS	SSSTTLTTKA	STSTTSSKTT	TTTSKTSTTS	SSSTNVAQLY	GQCGGQGTG	PTTCASGTCTKQNDYYSQCL												
SEQ ID NO:75	MYRILKSFIL	LSLVNMSLSQ	KIGKLTPEVH	PPMTFQKCSE	GGSCETIQGE	VVVDANWRWV	HSAQGQNCYT	GNTWNPTICP	DDETCACENCY											
	LDGANYESVY	GVTTSEDSVR	LNFVTQSQGK	NIGSRFLFMS	NESNYQLFHV	LGQEFTFDVD	VSNLDCGLNG	ALYLVMSDSD	GG SARFPTNE											
	AGAKYGTGYC	DAQCPRDLKF	ISGSANVDGW	IPSTNNPNTG	YGNLGSCCAE	MDLWEANNMA	TAVTPHPCDT	SSQSVCKSDS	CGGAASSNRY											
	GGICDPDGCD	YNPYRMGNTS	FFGPNKMIDT	NSVITVVVTQF	ITDDGSSD GK	LTSIKRLYVQ	DGNVISQSVS	TIDGVEGNEV	NEEFCTNQKK											
	VFGDEDSFTK	HGGLAKMGEA	LKDGMMVLVS	LWDDYQANML	WLDSSYPTTS	SPTDPGVARG	SCPTTSGVPS	KVEQNYPNAY	VVYSNIKVGP	IDSTYKK										
SEQ ID NO:76	MISRVLAISS	LLAAARAQOI	GTNTAEVHPA	LTSIVIDANW	RWLHTTSGYT	NCYTGNSWDA	TLCPDAVTCA	ANCALDGADY	SGTYGITTSG											
	NSLKLNFVTK	GANTNVGSRT	YLM AAGSKTQ	YQLLKL LGQE	FTFDVDVSNL	PCGLNGALYF	AEMDADGGVS	RFPTNKAGAQ	YGTGYCDAQC											
	PQDIKIFINGQ	ANSVGWTPSS	NDVNTGTGQY	GSCCSEMDIW	EANKISAAAY	PHPCSVDGQT	RCTGTD CGIG	ARYSSLCDDAD	GCDFNSYRMG											
	DTGFYGAGLT	VDTSKVFTVV	TQFITNDGTT	SGTLSEIRRF	YVQNGKVIPN	SQSKVTGVSG	NSITDSFCAA	QKTAFGDTNE	FATKGGLATM											
	SKALAKGMVL	VMSIWDDHSA	NMLWLDAPYP	ASKSPSAAGV	SRGSCSASSG	VPADVEANSP	GASVTYSNIK	WGPINSTYSA	GTGSNTGSGS											
	GSTTTLVSSV	PSSTPTSTTG	VPKYGQC GGS	GYTGPTNCIG	STCVSMGQYY	SQCQ														
SEQ ID NO:77	MYRQVATALS	FASLVLGQQV	GTLTAETHPS	LPIEVCTAPG	SCTKEDTTVV	LDANWRWTHV	TDGYTNCYTG	NAWNETACPD	GKTC AANCAI											
	DGA EYEKTYG	ITTP EEGALR	LNFVTESNVG	SRVYLMAGED	KYRLFNLLNK	EFTMDVDVSN	LPCGLNGAVY	FSEMDDEDGM	SRFEGNKAGA											
	KYGTGYCDSQ	CPRDIKFING	EANSEGW GGE	DGNSGTGKYG	TCCAEMDIWE	ANLDATA YTP	HPCKVTEQTR	CEDDTECGAG	DARYEGLCDR											
	DGCDFNSFRL	GNKEFYGPEK	TVDTSKPFTL	VTQFVTADGT	DTGALQSIRR	FYVQDGTVIP	NSETVVVEGVD	PTNEITDDFC	AQOKTAFGDN											
	NHFKTIIGGLP	AMGKSLEKMV	LVL SIWDDHA	VYMNWLDSNY	PTDADPTKPG	VARGRCDPEA	GVPETVEAAH	PDAYVIYSNI	KIGALNSTFA	AA										
SEQ ID NO:78	MSSEFQVYRAA	LLLSILATAN	AQQVGTYTTE	THPSLTWQTC	TSDGSC TTN D	GEVVIDANWR	WVHSTSSATN	CYTGNEWDTS	ICTDDDVTC AA											
	NCALDGATYE	ATYGVTTSGS	ELRLNFVTQG	SSKNIGSRLY	LMSDDSNYEL	FKLLGQEFTF	DVDVSNLPCG	LNGALYFVAM	DADGGTSEYS											
	GNKAGAKYGT	GYCDSQCPRD	LKFINGEANC	DGWEPSNNV	NTGVGDHGSC	CAEMDVWEAN	SISNAFTAHP	CDSVSQTMCD	GDSCGGT YSA											
	SGDRYSGTCD	PDGCDYNPYR	LGNTDFYGP G	LTVDTNSPFT	VVTQFITDDG	TSSGTLTEIK	RLYVQNGEVI	ANGASTYSSV	NGSSITS AFC											
	ESEKTLFGDE	NVFDKHG GLE	GMGEAMAKGM	VLVL SLWDDY	AADMLWLDSD	YPVNSSASTP	GVARGTCSTD	SGVPATVEAE	SPNAYVTYSN											
	IKFGPIGSTY	SSGSSSGSGS	SSSSSSTTTK	ATSTTLKTTS	TTSSGSSSTS	AAQAYQC G G	QGWGTPTTCV	SGYTCTYENA	YYSQCL											

TABLE 7

TABLE 7																				
SEQ ID NO.	Amino acid sequence																			
SEQ ID NO:79	MYRAIATASA	LLATARAQQV	CTLNTENKPA	LTWAKCTSSG	CSNVRGSVVV	DANWRWAHST	SSSTNCYTG	TWDKTLCPDG	KTCADKCCLD											
	GADYSGTYGV	TSSGNQLNLK	FVTVGPYSTN	VGSRLYLMED	ENNYQMFDLL	GNEFTFDVDV	NNIGCGLNGA	LYFVSMDKDG	GKSRSFSTNKA											
	GAKYGTGYCD	AQCPRDVKFI	NGVANSDEWK	PSDSCKNAGV	GKYGTCCPEM	DIWEANKIST	AYTPHPCKSL	TQQSCEGDAC	GGTYSATRYA											
	GTCDDPDGCD	NPYRQGNKTF	YGPGSGENVD	TTKKVTVVVTQ	FIKGS	DGKLS	EIKRLYVQNG	KVIGNPQSEI	ANNPGSSVTD	SFCKAQKVAF										
	NDPDDFNKKG	GWSGMSDALA	KPMVLVMSLW	HDHYANMLWL	DSTYPKGSKT	PGSARGSCPE	DSGDPDTLEK	EVPNSGVSES	NIKFGPIGST											
	YTG	TGGSNPD	PEEP	PEEP	PEEP	VGTV	PQYGQC	GGINYS	GPTA	CVSPYKCNKI	NDFYS	QCCQ								
SEQ ID NO:80	EQAGTATAEN	HPPLTWQECT	APGSC	TTQNG	AVVLDANWRW	VHDVNGYTNC	YTGNTWDP	TY	CPDDE	TCQ	AN	CALDGADYEG	TYGVTSSGSS							
	LKLN	FVTGSN	VGSR	LYLLQD	DSTYQIFKLL	NREFSFDVDV	SNLPCGLNGA	LYFVAMDADG	GVS	KYPNNKA	GAKYGTGYCD	SQCPRDLKFI								
	DGEAN	VEGWQ	PSSNNANTGI	GDHGSCCAEM	DVWEANSISN	AVTPHPCDTP	GQTMC	SGDDC	GGTYSNDRYA	GTCDDPDG	CDF	NPYRMGNTSF								
	YGP	KI	IDTT	KPFTVVVTQFL	TDDGTD	TGTL	SEIKRFYIQN	SNVIPQPN	SD	ISGVTGNSIT	TEFCTAQKQA	FGDTDDFSQH	GGLAKMGAAM							
	QQGMVLVMSL	WDDYAAQMLW	LDSDYPTDAD	PTTPGIARGT	CPTDSGVPSD	VESQSPNSYV	TYSNIKFGPI	NSTFTAS												
SEQ ID NO:81	MFPTLALVSL	SFLAIAYGQQ	VGTLTAETHP	KLSVSQC	TAG	GSCTTVQRSV	VLDSNWRWLH	DVG	STNCYT	GNTWDDSLCP	DPTTCAANCA									
	LDGADYSGTY	GITTS	GNALS	LKFVTQGPYS	TNIGSRVYLL	SEDDSTYEMF	NLKNQEFTFD	VDMSALPCGL	NGALYFVEMD	KDGGSGRFPT										
	NKAGSKYGTG	YCDTQC	PHDI	KFINGEANVL	DWAGSSNDPN	AGTGHYGTCC	NEMDIWEANS	MGA	AVTPHVC	TVQGQTRCEG	TDCGDGDERY									
	DGICDKDGCD	FNSWRMGDQT	FLPGKTVDT	SSKFTVVVTQF	ITADNTTSGD	LSEIRRLYVQ	NGKVIANSKT	QIAGMDAYDS	ITDDFCNAQK											
	TTFGDTNTFE	QMGGLATMGD	AFETGMVLVM	SIWDDHEAKM	LWLDSDYPTD	ADASAPGVSR	GPCPTTSGDP	TDVESQSPGA	TVIFSNIKTG	PIGSTFTS										
SEQ ID NO:82	MLSASKAAAI	LA	FCAHTASA	WVVGDQQTET	HPKLNWQRCT	GKGRSSCTNV	NGEVVIDANW	RWLAHRSGYT	NCYTGSEWNQ	SACP	NNEACT									
	KNCAIEGSDY	AGTYGITTSG	NQMN	IKFITK	RPYSTNIGAR	TYLMKDEQNY	EMFQLIGNEF	TFD	VDL	SQRC	GMNGALYFVS	MPQKGQGAPG								
	AKYGTGYCDA	QCARDLK	FVR	GSANAE	GWTK	SASDPNSGVG	KKGACCAQMD	VWEANS	SAATA	LTPHSCQPAG	YSVCEDTNCG	GTYSEDRYAG								
	TCDANGCDFN	PFRVGVKDFY	GKGKTVDTTK	KMTVVVTQFVG	SGNQLSEIKR	FYVQDGKVIA	NPEPTIPGME	WCNTQKKVFQ	EEAYPFNEFG											
	GMA	SMSEGMS	QGMVLVMSLW	DDHYANMLWL	DSNWPREADP	AKPGVARRDC	PTSGGKPSEV	EAANPNAQVM	FSNIKFGPIG	STFAHAA										
SEQ ID NO:83	MFR	TATLLAF	TMAAMVFGQQ	VGNTARSHP	ALTSQKCTKS	GGCSNLNTKI	VLDANWRWLH	STSGYTNCYT	GNQWDATLCP	DGKTCAANCA										
	LDGADYTGTY	GITASGSSLK	LQFVTGSNVG	SRVYLMADDT	HYQMFQLLNQ	EFTFDV	DMSN	LPCGLNGALY	LSAMDADGGM	AKYPTNKAGA										
	KYGTGYCDSQ	CPRDIKFING	EANVEGWNAT	SANAGTGNYG	TCCTEMDIWE	ANNDA	AAAYTP	HPCTTNAQTR	CSGSDCTRDT	GLCDADGCDF										
	NSFRMGDQTF	LGKGLTV	DTS	KPFTVVVTQFI	TNDGTSAGTL	TEIRRLYVQN	GKVIQNSSVK	IPGIDPVNSI	TDNFC	SQQT	AFGDTNYFAQ									
	HGGLKQVGEA	LRTGMVLALS	IWDDYAAANML	WLDSNYPTNK	DPSTPGVARG	TCATTSGVPA	QIEAQSPNAY	VVFSNIKFGD	LN	TTYTGTVS										
	SSSVSSSHSS	TSTSSSHSSS	STPPTQPTGV	TVPQWGQC	GG	IGYTGSTTCA	SPYTCHVLNP	YYSQCY												
SEQ ID NO:84	MYQRALLFSA	LMAGVSAQQV	GTQKPETHPP	LAWKECTSSG	CTSKDGSVVI	DANWRWVHSV	DGYKNCYTGN	EWDSTLCPDD	ATCATNCAVD											
	GADYAGTYGA	TTEGDSL	SIN	FVTGSNIGSR	FYLME	DENKY	QMFKLLNKEF	TFD	VDVSTLP	CGLNGALYFV	SMDADGGMSK	YETNKAGAKY								
	GTGYCDSQCP	RDLKFINGKG	NVEGWKPSAN	SCCAEMDIWE	ANSI	STALTP	HPCDTNGQTI	CEGDS	CGGT	Y	STTRYAGTCD									
	PDGCDENPFR	MGNESFYGPG	KMVDTKSKMT	VVTQFITSDG	TDTGSLKEIK	RVYVQNGKVI	ANSASDVSGI	TGNSITSDFC	TAQKKTFFGDE											
	DVF	NKHGGLS	GMGDALGEGM	VLVMSLWDDH	NSNMLWLDGE	KYPTDAAASK	AGVSRGTCST	DSGK	PSTVES	ESGSAKVVS	NIKVGSIGST	FSA								

TABLE 7

SEQ ID NO.		Amino acid sequence																			
SEQ ID NO:85		MTSKIALASL	FAAAYGQQIG	TYTTETHPSL	TWQSCTAKGS	CTTQSGSIVL	DGNWRWTHST	TSSTNCYTG	TWDATLCPDD	ATCAQNCA	L	GADYSGTYGI	TTSGDSLRLN	FVTQTANKV	GSRVYLLADN	THYKTFNLLN	QEFTFDVDVS	NLPCGLNGAV	YFANLPADGG	ISSTNKAGAAQ	
		YGTGYCDSQC	PRDGKFINGK	ANVDGWPSS	NNPNTGVGNY	GSCCAEMDIW	EANSISTAVT	PHSCDVTQT	VCTGDNCGGT	YSTTRYAGTC		DPDGCDFNPY	RQGNESFYGP	GKTVDTNSVF	TIVTQFLTTD	GTSSGTLNEI	KRFYVQNGKV	IPNSESTISG	VTGNSITTFP	CTAQKTA	AFGD
		PTSFSDHGGL	ASMSAAFEAG	MVLVLSLWDD	YYANMLWLDS	TYPTTKTGAG	GPRGTCSTSS	GVPASVEASS	PNAYVVYSNI	KVGAINSTFG											
SEQ ID NO:86		MYTKFAALAA	LVATVRGQAA	CSLTAETHPS	LQWQKCTAPG	SCTTVSGQVT	IDANWRWLHQ	TNSSTNCYTG	NEWDT	SICSS	DTDCA	T	DGADYTGTYG	VTASGNSLNL	KFVTQGPYSK	NIGSRMYLME	SESKYQGFTL	LGQEFTFDVD	ALYFVSM	D	L
		AGAKYGTGYC	DSQCPRDLKF	INGQANIDGW	QPSSNDANAG	LGNHGSCCSE	MDIWEANKVS	AAYTPHPCTT	IGQTMCTGDD	CGGTYS	SDRY	AGICDPDGCD	FNSYRMGDT	FYGP	GKTVDT	LTGSDGNLSE	IKRFYVQNGK	VIPNSESKIA	GVSGNSITTD	FCTAQKTA	AFG
		DTNVFEERGG	LAQMGKALAE	PMVLVLSVWD	DHAVNMLWLD	STYPTDSTKP	GAARGDCPIT	SGVPADVESQ	APNSNVIYSN	IRFGPIN	STY	TGTPSGGNPP	GGGTTTTTTT	TTSKPSGP	TT	HNPSGPGQQT	HWGQCGGQGW	TGPTVCQSPY	TKKYSNDWYS	QCL	
SEQ ID NO:87		MYQRALLFSA	LLSVSRAQQA	GTAQE	EEVHPS	LTWQRC	EASG	SCTEVAGSVV	LDSNWRWTHS	VDGYTN	CYTG	NEWDA	T	LC	PD	NESCAQ	NC	AV			
		DGADYEATYG	ITSNGDSLTL	KFVTGSNVGS	RVYLMEDDET	YQMEDLLNNE	FTFDVDVSNF	PCGLNGALYF	TSMDADGGLS	KYEGNTAGAK		YGTGYCDSQC	PRDIKFINGL	GNVEGWEP	SD	ANAGVGGM	GTCCPEMDIW	EANSISTAYT	PHPCDSVEQT	MCEGDS	CGGT
		DPDGCDFNSY	RMGNTRFYGP	GAIIDTSSKF	TVVTQFIADG	GSLSEIKRFY	VQNGEVI	PNS	ESN	ISGVEGN	SITSEFCTAQ	KTA	FGDE	D	I	F					
		AQHGGLSAMG	DAASAMVLIL	SIWDDHHSSM	MWLDSSYPTD	ADPSQPGVAR	GTCEQAGADP	DVVESEHADA	SVTFSNIKEG	PIGSTF											
SEQ ID NO:88		MMMKQYLQYL	AAGSLMTGLV	AGQGVGTQQT	ETHPRITWKR	CTGKANCTTV	QAEVV	IDS	NW	RWIHTSGGTN	CYDGN	AWNTA	AC	STAT	DCAS						
		KCLMEGAGNY	QQTYGASTSG	DSLTLKFVTK	HEYGTNVGSR	FYLMNGASKY	QMFTLMNNEF	TFD	VDLSTVE	CGLNSALYFV	AMEED	GGMRS									
		YPTNKAGAKY	GTGYCDAQCA	RDLKFVGGKA	NIEGWRESSN	DENAGVGPYG	GCCAEIDVWE	SNAHAYAFT	P	HACENNNYHV	CERD	TCGGTY									
		SEDRFAGGCD	ANGCDYNPYR	MGNPDFYGKG	KTVDTTKKFT	VVTRFQDDNL	EQFFVQNGQK	ILAPAPT	FDG	IPASP	NLTPE	FCSTQ	FDVFT								
		DRNRFREVGD	FPQLNAALRI	PMVLVMSIWA	DHYANMLWLD	SVYPPEKEGE	PGAARGPCAQ	DSGV	PSEVKA	NYPNAKV	VMS	NIRFG	PIGST	VNV							
SEQ ID NO:89		MYQRALLFSE	FLAAARAQQA	GTVTAENHPS	LTWQCCSSGG	SCTTQNGKV	IDANWRWVHT	TSGYTN	CYTG	NTWDT	SICPD	DV	TCAQ	NCAL							
		DGADYS	SGTYG	VTTSGNALRL	NFVTQSSGKN	IGSRLYL	LQD	DTTYQIFKLL	GQEFTFDVDV	SNLPCGLNGA	LYFVAM	DADG	GLSKY	PPGNKA							
		GAKYGTGYCD	SQCPRDLKFI	NGQANVEGWQ	PSANDPNAGV	GNHGS	CCAEM	DVWEANSIST	AVTPHPCDTP	GQ	TCQGD	DC	GGTY	SS	TRYA						
		GTCDDPDG	CDF	NPYRQGNHSF	YGP	GKIV	DT	S	KFTVVTQFI	TDDG	TPSGTL	TEIKR	FYVQN	GKVI	PQSEST	ISGVT	GN	SIT	TEYCTAQ	KAA	
		FGDNTGFFTH	GGLQKISQAL	AQGMVLVMSL	WDDHAANMLW	LDSTYPTDAD	PDTPGVARGT	CPTTSGVPAD	VESQN	PNSYV	IYSNI	KVGPI	NSTFTAN								
SEQ ID NO:90		MFAIVLLGLT	RSIGTG	TNQ	A	ENHPSLSWQN	CRSGG	SCTQT													
		GTYGITSSGN	SLKLVFVTNG	PYSTNIGSRV	YLLKDESHYQ	IFDLKNKEFT	FTVDDSNLDC	GLNGALYFVS	MDEDGGTSRF	SSNKAGAKY											
		TGYCDAQCPH	DIKFINGEAN	VENWKPPQTND	ENAGNGRYGA	CCTEMDIWEA	NKYATA	YTPH	ICTVNGEYRC	DGSECGD	TDS	GNRY	GGV	CDK							
		DGCDFNSYRM	GNTSFWGPGL	IIDTGKPVTV	VTQFVTKDGT	DNGQLSEIRR	KYVQGGK	VIE	NTVVNI	AGMS	SGNSIT	DDFC	NEQK	SA	F	G	D	T			
		NDFEKKGGLS	GLGKA	FDYGM	VLVLSLWDDH	QVNMLWLD	SI	YPTDQ	PASQP	GVKRGP	CATS	SGAP	SD	VESQ	HPD	SSVT	FSD	IRFG	PID	STY	

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:97	MKQYLQYLAA ALPLMSLVSA QGVGTSTSET HPKITWKKCS SGGSCSTVNA EVVIDANWRW LHNADSKNCY DGNEWTDACT SSDDCTSKCV LEGAEYGKTY GASTSGDLSL LKFLTKHEYG TNIGSRFYLM NGASKYQMFT LMNNEFAFDV DLSTVECGLN SALYFVAMEE DGMASYSTN KAGAKYGTGY CDAQCARDLK FVGGKANYDG WTPSSNDANA GVGALGGCCA EIDVWESNAH AFAFTPHACE NNNYHVCEDT TCGGTYSEDR FAGDCDANGC DYNPYRVGNT DFYGGKMTVD TSKKFTVVSQ FQENKLTQFF VQNGKKIEIP GPKHEGLPTE SSDITPELCS AMPEVFGDRD RFAEVGGFDA LNKALAVPMV LVMSIWDDHY ANMLWLDSSY PPEKAGTPGG DRGPCAQDSG VPSEVESQYP DATVVWSNIR FGPIGSTVQV
SEQ ID NO:98	MFPKASLIAL SFIAAVYGQQ VGTQMAEVHP KLPSQLCTKS GCTNQNTAVV LDANWRWLHT TSGYTNCYTG NSWDATLCPD ATTCAQNCAV DGADYSGTYG ITTSGNALT LKFKTGTNVGS RVYLMQTDTA YQMFQLLNQE FTFDVDMNL PCGLNGALYL SQMDQDGLS KFPTNKAGAK YGTGYCDSQC PHDIKIFINGM ANVAGWAGSA SDPNAGSGTL GTCCSEMEDIW EANNDAAAFT PHPCSVDGQT QCSGTQCDD DERYSGLCDK DGCDFNSFRM GDKSFLGKGM TVDTSRKFTV VTQFVTTDGT TNGDLHEIRR LYVQDGKVIQ NSVVSIPGID AVDSITDNFC AQQKSVMFGDT NYFATLGLLK KMGAAALKSGM VLAMSVWDDH AASMQLDSN YPADGDATKP GVARGTCSAD SGLPTNVESQ SASASVTFSN IKWGDINTTF TGTGSTSPSS PAGPVSSSTS VASQPTQPAQ GTVAQWQCQG GTGFTGPTVC ASPFTCHVVN PYYSQCY
SEQ ID NO:99	MFRTAALLSF AYLAVVYGQQ AGTSTAETHP PLTWEQCTSG GSCTTQSSSV VLDSNWRWTH VVGGYTNCYT GNEWNTTVCP DGTTCAANCA LDGADYEGTY GISTSGNALT LKFVTASAO LKFGKANVE GWTPSSTSPN AGTGGTGICC NEMDIWEANS ISEALTPHPC TAQGGTACTG DSCSSPNSTA NKAGAKYGTG YCDSQCPRDI KFIEGKANVE YGPGLTVDTT SKITVVVTQFI TSDNNTTGD LTAIRRIYVQN GQVIONSMNS IAGVTPTNEI TTDFCDQOKT GICDQAGCDF NSFRMGDTSF YGPGLTVDTT SKITVVVTQFI TSDNNTTGD LTAIRRIYVQN GQVIONSMNS IAGVTPTNEI TTDFCDQOKT AFGDTNNTFSE KGGLTGMGAA FSRGMVLVLS IWDDDAEAML WLDSTYPVGK TGPGAARGTC ATTSQGPQDV ETQSPNAQVV FSNIKFGAIG STFSSTGTGT GTGTGTGTGT GTTTSSAPAA TQTKYGQCQG QGWGTGATVCA SGSTCTSSGP YYSQCL
SEQ ID NO:100	MFRTAALTAF TFAAVVLGQQ VGTLTTENHP ALSIQQCTAT GCTTQQKSVV LDSNWRWTHS TAGATNCYTG NAWDPALCPD PATCATNCAI DGADYSGTYG ITTSGNALT LKFVTASAO LKFGKANVE GWTPSSTSPN AGTGGTGICC NEMDIWEANS ISEALTPHPC TAQGGTACTG DSCSSPNSTA AGAKYGTGYC DAQCPHDIKW INGEANVLDM SASATDDNAG NGRYGACCAE MDIWEANSEA TAYTPHVCRD EGLYRCSGTE CGDGNRNYGG VCDKDGCDFN SYRMGDKNFL GRGKTIDTTK KVTVVVTQFIT DNNTPTGNLV EIRRVYVQNG VVYQNSFSTF PSLSQYNSIS DEFCEVAQKTL FGDNQYYNTH GGTTKMGDAF DNGMVLIMSL WSDHAAHMLW LDSDYPLDKS PSEPGVSRGA CPTSSGDPDD VVANHPNASV TFSNIKYGPI GSTFGGSTPP VSSGGSSVPP VTSTTSSGTT TPTGPTGTVP KWGQCQGIGY SGPTACVAGS TCTYSNDWYS QCL
SEQ ID NO:101	MYRAIATASA LIAAVRAQQV CSLTPETKPA LSWSKCTSSG CSNVQGSVTI DANWRWTHQL SGSTNCYTGN KWDTSICTSG KVCAEKCCID GAEYASTYGI TSSGNQLSLS FVTKGTYGTN IGSRTYLMED ENTYQMFQLL GNEFTFDVDV SNIGCGLNGA LYFVSMDADG GKAKYPGNKA GAKYGTGYCD AQCPRDVKFI NGQANS DGWQ PSKSDVNGGI GNLGTCCPEM DIWEANSIST AHTPHPC TKL TQHSCTGDSC GGTYSIEDRYG GTCDDAGCDF NAYRQGNKTF YGPGS GFENV D TTKKVTVVVTQ FHKGSNGRLS EITRLYVQNG KVIANSESKI AGVPGSSLTP EFCTAQKKVF GDIDDFEKKG AWGMSDALE APMVLVMSLW HDHHSNMLWL DSTYPTDSTK LGAQRGSCST SSGVPADLEK NVPNSKVAFS NIKFGPIGST YKEGQPEPTN PTNPNTTPG GTVDQWQCQG GTNYSGPTAC KSPFTCKKIN DFYSQCCQ
SEQ ID NO:102	MFRTATLLAF TMAAMVFGQQ VGTNTAENHR TLTSQKCTKS GGCSNLNTKI VLDANWRWLH STSGYTNCYT GNQWDATLCP DGKTCANCA LDGADYTGTY GITASGSSLK LQFVTGSNVG SRVYLMADDT HYQMFQLLNQ EFTFDVDMNS LPCGLNGALY LSAMDADGGM AKYPTNKAGA KYGTGYCDSQ CPRDIK FING EANVEGWNAT SANAGTGNYG TCCTEMDIWE ANNDAAAYTP HPCTTNAQTR CSGSDCTRDT GLCDADG CDF NSFRMGDQTF LGKGLTV DTS KPFTVVVTQFI TNDGTSAGTL TEIRRLYVQN GKVIQNSSVK IPGIDL VNSI TDNFC SQKT AFGDTNYFAQ HGGLKQVGEA LRTGMVLALS IWDDYAAANML WLDSNYPTNK DPSTPGVARG TCATTSGVPA QIEAQSPNAY VVFSNIKFGD LNTTYTGTVS SSSVSSSHSS TSTSSSHSS STPPTQPTGV TVPQWQCQG IGYTGSTTCA SPYTCHVLNP YYSQCY

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:103	MYQTSLLASL SFLLATSQAQ QVGTQTAETH PKLTTQKCTT AGGCTDQSTS IVLDANWRWL HTVDGYTNCY TGQEWDTSIC TDGKTCAEKC ALDGADYEST YGISTSGNAL TMNFVTKSSQ TNIGGRVYLL AADSDDTYEL FKLKNQEFTF DVDVSNLPCG LINGALYFSEM DSDGGLSKYT TNKAGAKYGT GYCDTQC PHD IKFINGEANV QNWTASSTDK NAGTGHYGSC CNEMDIWEAN SQATAFTPHV CEAKVEGQYR CEGTECGDGD NRYGGVCDKD GCDFNSYRMG NETFYGSNGS TIDTTKKFTV VTQFITADNT ATGALTEIRR KYVQNDVVIE NSYADYETLS KFNSITDDFC AAQKTLSGDT NDFKTKGGIA RMGESFERGM VLVMSVWDDH AANALWLDSS YPTDADASKP GVKRGPCSTS SGVPSDVEAN DADSSVIYSN IRYGDIGSTF NKTA
SEQ ID NO:104	MFSKVALTAL CFLAVAQAQQ VGREVAENHP RLPWQRCTRN GGCQTVSNGQ VVLDANWRWL HVT DGYTNCY TGNAWNSSVC SDGATCAQRC ALEGANYQQT YGITTSGDAL TIKFLTRSEQ TNIGARVYLM ENEDRYQM FN LLNKEFTFDV DVSKVPCGIN GALYFIQMDA DGGLSSQPNN RAGAKYGTGY CDSQCPRDIK FINGEANSVG WEPSETDPNA GKGQYGICCA EMDIWEANSI SNAYTPHPCQ TVNDGGYQRC QGRDCNQPRY EGLCDPDGCD YNPFRMGNKD FYGP GKTVDT NRKMTVVTTQF ITHDNTDTGT LVDIRRLYVQ DGRVIANPPT NFPGLMPAHD SITQEFCDDA KRAFEDNDSF GRNGGLAHMG RSLAKGHVLA LSIWNDHTAH MLWLD SNYPT DADPNKPGIA RGTCPTTGS PRDTEQNHPD AQVIFSNIKF GDIGSTFSGN
SEQ ID NO:105	MYRKLAVISA FLAAARAQQV CTQQAETHPP LTWQKCTASG CTPQQGSVV L DANWRWTHDT KSTTNCYDGN TWSSTLC PDD ATCAKNCCLD GANYSGTYGV TTSGDALTLQ FVTASNVGSR LYLMANDSTY QEFTLSGNEF SFDVDVSQLP CGLNGALYFV SMDADGGQSK YPGNAAGAKY GTGYCDSQCP RDLKFINGQA NVEGWEPSSN NANTGVGGHG SCCSEMDIWE ANSISEALT P HPCETVGQTM CSGDSCGGTY SNDRYGGTCD PDGCDWNPYR LGNTSFYGP G SSFALDTTKK LTVVTQFATD GSISRYYVQN GVKFQQPNAQ VGSYSGNTIN TDYCAAEEQTA FGGTSFTDKG GLAQINKAFQ GGMVLVMSLW DDYAVNMLWL DSTYPTNATA STPGA KRGS C STSSGVPAQV EAQSPNSKVI YSNIRFGPIG STGGNTGSNP PGTSTTRAPP SSTGSSPTAT QTHYGQC CGT GW TGPTRCAS GYTCQVLNPF YSQCL
SEQ ID NO:106	MRASLLAFSL AAAVAGGQQA GTLTAKRHPS LTWQKCTRGG CPTLNTTMVL DANWRWTHAT SGSTKCYTGN KWQATLC PDG KSCAANCALD GADYTGTYGI TGSGWSLT LQ FVTDNVGARA YLMADDTQYQ MLELLNQELW FDVDM SNIPC GLNGALYLSA MDADGGMRKY PTNKAGAKYA TGYCDAQCPR DLKYINGIAN VEGWTPSTND ANGIGDHGSC CSEMDIWEAN KVSTAFTHP CTTIEQHMCE GDSCGGTYSD DRYGVLC DAD GCDFNSYRMG NTTFYGE GKT VDTSSKFTVV TQFIKDSAGD LAELKAFYVQ NGKVIENSQS NVDGVSGNSI TQSECKSQKT AFGDIDDFNK KGGLKQMGKA LAQAMVLVMS IWDDHAANML WLDSTYPVPK VPGA YRGS GP TTSGVPAEVD ANAPNSKVAF SNIKFGHLGI SPFSGGSSGT PPSNPSSSAS PTSSTAKPSS TSTASNPSGT GAAHWAQCGG IGFSGPTTCP EPYTCAKDHD IYSQCV
SEQ ID NO:107	MLASTFSYRM YKTALILAAL LGSGQAQQVG TSQAEVHPSM TWQSCTAGGS CTTNNGKVVI DANWRWVHKV GDYTNCYTGN TWDKTLC PDD ATCASNCALE GANYQSTYGA TTSGDSLRLN FVTTSQQKNI GSRLYMMKDD TTYEMFKLLN QEFTFDVDVS NLPCGLNGAL YFVAMDADGG MSKYPTNKAG AKYGTGYCDS QCPRDLKFIN GQANVEGWQP SSNDANAGTG NHGSCCAEMD IWEANSISTA FTPHPCDTPG QVMCTGDACG GTYSDDRYGG TCDDPDGCDFN SFRQGNKTFY GPGMTVDTKS KFTVVTQFIT DDGTASGTLK EIKRFFYVQNG KVIPNSESTW SGVGGNSITN DYCTAQKSLF KDQNVFAKHG GMEGMGAALA QGMVLVMSLW DDHAANMLWL DSNYPTTASS STPGVARGTC DISSGVPADV EANHPPDASVV YSNIKVGPIG STFNSGGSNP GGGTTTTAKP TTTTTTAGSP GGTGVAQHYG QCGGNGWQGP TTCASPYTCQ KLNDFYSQCL
SEQ ID NO:108	MQIKQYLQYL AAALPLVNMA AAQRAGTQQT ETHPRLSWKR CSSGGNCQTV NAEIVIDANW RWLHDSNYQN CYDGNRWTS A CSSATDCAQK CYLEGANYGS TYGVSTSGDA LTLKFVTKHE YGTNIGSRVY LMNGSDKYQM FTLMNNEFAF DVDLSKVECG LNSALYFVAM EEDGGMRSYS SNKAGAKYGT GYCDAQCARD LKFVG GKANI EGWRPSTNDA NAGVGPYGAC CAEIDVWESN AYAF AFTPHG CLNNNYHVCE TSNCGGTYSE DRFGGLCDAN GCDYNPYRMG NKDFY GKGT VDTSRKFTVV TRFEENKLTQ FFIQDGRKID IPPPTWPGLP NSSAITPELC TNLSKVFFDDR DRYEETGGER TINEALRIPM VLVMSIWDGH YASMLWLDSV YPPEKAGQPG AERGPCAPTS GVPAEVEAQF PNAQVIWSNI RFGPIGSTYQ V

TABLE 7

SEQ ID NO.		Amino acid sequence																			
SEQ ID NO:109		MTSR	IALV	SL	FAAV	GQQV	G	TYQT	ETHPSL	TWQS	TAKGS	CTTN	TGSIVL	DGNWR	WTHGV	GTST	NCYTGN	TWDAT	LC	PDD	ATCAQNCAL
		GADY	SGTYGI	TTSG	NSLRIN	FVTQ	SANKNI	FSRV	YIMADT	THYK	TFNLLN	QEFT	FDVDVS	NLPC	GLNGAV	YFANL	PADGG	ISST	NTAGAE		
		YGTG	YCD	SQC	PRDM	KFIKGQ	ANVD	GWVPSS	NNANT	GVGNH	GSCCA	EMDIW	EANSI	STAVT	PHSCD	TVTQT	VCTG	DDCCGGT	YSSSR	YAGTC	
		DPDG	CD	FN	SY	RMGDE	TFYGP	GKTVD	TNSVF	TVVTQ	FLTTD	GTASG	TLNEI	KRFYV	QDGKV	IPNSY	STISG	VSGNS	ITTPF	CDAQKTA	FGD
		PTSF	SDHGG	L	ASMSA	A	FEAG	MVLVL	SLWDD	YYANM	LWLD	TYPV	GKTSAG	GPRGT	CDTSS	GVPAS	VEASS	PNAYV	VYSNI	KVGAIN	STYG
SEQ ID NO:110		MFV	FVLL	WL	T	QSLG	TGTNQA	ENHPS	LSWQN	CRSG	G	SCTQT	SGSV	VLD	SNW	RWTHD	SSLTN	CYDGN	EWSSS	LCPDP	KTCSD
		GTYG	ITSSGN	SLKL	V	FTNG		PYSTN	IGSRV	YLLK	DESHYQ	IFDL	KNKEFT	FTVDD	SNLDC	GLNGAL	YFVS	MDED	GGTSRF	SSNK	AGAKYG
		TGYC	DAQCPH	DIKFI	NGEAN			VENWK	PQTND	ENAGN	G	RYGA	CCTEM	DIWEA	NKYATA	YTPH	ICTVN	GEYRC	DGSEC	G	DTDS
		DGCD	FN	SYRM	GNTS	FWGPGL		IIDTG	KPVT	VTQF	VTKDGT	DNGQL	SEIRR	KYVQG	GKVIE	NTVVN	NIAGMS	SGNSIT	DDFC	NEQK	SAFGDT
		NDFE	KKGGLS	GLGKA	FDYGM			VLVL	SLWDDH	QVNML	WLD	SI	YPTD	Q	PASQP	GVKRG	PCATS	SGAP	S	D	VESQ
																					HPDSSVTFSD
																					IRFGPIDSTY
SEQ ID NO:111		MFRKA	ALLAF	SFLA	IAHGQQ			VGTNQ	AENHP	SLPS	QKCTAS	GCTT	S	TS	SVV	LDANWR	WVHT	TTGY	TNCYTG	Q	TWDAS
		DGADY	SGTYG	ITTS	GNALTL			QFVK	G	TNVGS	RVYLL	QDASN	YQMF	QLINQE	FTFD	V	DMSNL	PCGL	NGAVYL	SQMD	Q
		YGTG	YCD	SQC	PRDI	K	FINGE	ANVE	G	WTGSS	TDSN	SGTGN	Y	GTCC	SEMDIW	EANSV	AAAYT	PHPC	SVNQQT	RCTG	AD
		DGCD	FN	SFRM	GDQT	FLGKGL		TVDT	SRKFTI	VTQF	ISDDGT	TSGN	LA	EIRR	FYVQD	GNVIP	NSKVS	IA	GID	AVNS	ITD
		NRFAA	Q	GGLK	QMGAA	LKSGM		VLAL	SLWDDH	AANML	WLDSD	YPTT	ADASNP	G	VARGT	CPPT	SGFPR	D	VESQ	SGSAT	V
		TGTL	TT	PSGS	SSPSS	PASTS		GSST	SASSA	SVPT	QSGTVA	QWAQ	CGGIGY	SGATT	CVSPY	TCHVV	V	N	AYYS	QC	Y
SEQ ID NO:112		MYRAI	ATASA	LIAA	ARAQQV			CTLTT	ETKPA	LTWSK	CTSSG	CTD	VKGSVGI	DANWR	WTHQT	SSST	NCYTGN	KWDT	SVCTSG	ETCAQ	KCCLD
		GADY	AGTYGI	TSSGN	QLSLG			FVTKG	SFSTN	IGSR	TYLMEN	ENTY	QMFQLL	GNEFT	FDVDV	SNIG	CGLNGA	LYFV	SMDADG	GKARY	PANKA
		GAKY	G	TYCD	AQCPR	DVKFI		NGKAN	SDGWK	PSDS	DINAGI	GNMG	TCCPEM	DIWEA	NSIST	AFTPH	PCTKL	TQHAC	TG	DSC	GGTYSNDRYG
		GTCDA	DG	CDF	NSYR	QGNKTF		YGRG	SDFNVD	TTKKV	T	VVTQ	FKKG	SNGRLS	EITRL	YVQNG	KVIAN	SESKI	PGNS	GSSSLTA	DFCSKQKSVF
		GDID	D	FSKKG	GWSGM	SDALE		SPPMV	L	VMSL	WHDH	HSNMLW	LDST	YPTDST	KLGA	Q	RGSCA	TTSG	V	PSDLE	RDVPNSKVSF
		TYSSG	TTNPP	PSST	D	TSTTP		TNPPT	G	GTG	TVG	QYGC	QGGQTY	TGPK	D	CKSPY	TCKK	INDEYS	QCQ		
SEQ ID NO:113		MSSE	FIYRAA	LLL	SILATAN	AQQV	GTYTTE	THPSL	TWQTC	TSDG	SCTTND	GEVVI	DANWR	WVHST	SSATN	CYTGN	EW	D	T	S	ICTDD
		NCA	LDGATYE	ATYG	VTTSGS	ELRL	NFVTQG	SSKN	IGSR	LY	IMSDD	SNYEL	EKLL	GQEFTF	DVDV	SNLPCG	INGAL	YFVAM	DADGG	TSEYS	
		GNKA	GAKYGT	GYCD	SQCPRD	LKFINGE	ANC	DGWE	PSSNNV	NTGV	DHGSC	CAEMD	VWEAN	SISNA	FTAHP	CDSV	SQ	TMCD	GDSC	GGTYS	A
		SGDR	YSGTCD	PDGCD	YNPYR	LGNTD	FYGP	LTVD	TNSPFT	VVTQ	FITDDG	TSSG	TLTEIK	RLYV	QNGEVI	ANGAS	T	YSSV	NGSSIT	SAFC	
		ESEK	TLFGDE	NVFD	KHGGLE	GMGEA	MAKGM	VLVL	SLWDDY	AADM	LWLDSD	YPVN	SSASTP	G	VARGT	CSTD	SGVP	ATVEAE	SPNAY	V	TYSN
		IKFG	PIGSTY	SSG	SSSGSGS	SSSSS	STTK	ATST	TLKTTS	TTSSG	SSSTS	AAQAY	GQCGG	QGW	TGPTTCV	SGYT	CTYENA	YYSQ	CL		
SEQ ID NO:114		MHQR	ALLFSA	LLTAV	RAQQA	GTLTEE	VHPS	LTWQ	KCTSEG	SCTEQ	SGSVV	IDSN	NRWTHS	VNDST	NCYTG	NTWD	AT	LC	PD	DE	TCAANCAL
		DGADY	ESTYG	VTTD	GDSLTL	KFVTG	SNVGS	RLYL	MDT	SDE	GYQT	FNLLDA	EFTF	D	V	VS	N	LPCGL	NGALY	FTAM	DADGGV
		KYGT	G	YCD	SQ	CPRDL	KFIDG	QANV	D	GWEP	S	SNND	TGIGN	HGSC	CPEMDI	WEANK	I	STAL	T	PHPC	D
		CDPD	G	CD	FN	YRMGN	D	SFYG	PGKT	ID	TGSK	MTVVT	QFITD	GSGS	LSEIKR	YVQNG	GNVIA	NADSN	I	SGVT	GNSIT
		I	FAEH	NGLAG	ISDAM	SSMVL	ILSL	WDDYYA	SMEWL	DSDYP	ENATA	TDPGV	ARGT	C	D	SESG	VPAT	VEGAHP	DSSV	TFSNIK	FGPIN
																					STFSA SA

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:115	MYAKFATLAA LVAGAAAQNA CTLTAENHPS LTWSKCTSGG SCTSVQGSIT IDANWRWTHR TDSATNCYEG NKWDTSYCSD GPSCASKCCI DGADYSSTYG ITTSGNSLNL KFTVTKGQYST NIGSRTYLME SDTKYQMFQL LGNEFTFDVD VSNLGCGLNG ALYFVSMDDAD GGMSKYSGNK AGAKYGTGYC DSQCPRDLKF INGEANVENW QSSTNDANAG TGKYGSCCSE MDVWEANNMA AAFTHPCXV IGQSRCEGDS CCGTYSTDRY AGICDPDGCD FNSYRQGNKT FYGKGMTVDT TKKITVVVTQF LKNSAGELSE IKRFYVQNGK VIPNSESTIP GVEGNSITQD WCDRQKAAFG DVTDXQDKGG MVQMGKALAG PMVLVMSIWD DHAVNMLWLD STWPIDGAGK PGAERGACPT TSGVPAEVEA EAPNSNVIFS NIRFGPIGST VSGLPDGGSG NPNPPVSSST PVPSSSTTSS GSSGPTGGTG VAKHYEQCGG IGFTGPTQCE SPYTCTKLND WYSQCL
SEQ ID NO:116	MYAKFATLAA LVAGASAAQAV CSLTAETHPS LTWQKCTAPG SCTNVAGSIT IDANWRWTHQ TSSATNCYSG SKWDSICTT GTDCASKCCI DGAEYSSTYG ITTSGNALNL KFTVTKGQYST NIGSRTYLME SDTKYQMFKL LGNEFTFDVD VSNLGCGLNG ALYFVSMDDAD GGMSKYSGNK AGAKYGTGYC DAQCPRDLKF INGEANVEGW ESSTNDANAG SGKYGSCCTE MDVWEANNMA TAFTPHPCIT IGQTRCEGDT CCGTYSDDRY AGVCDPDGCD FNSYRQGNKT FYGKGMTVDT TKKITVVVTQF LKNSAGELSE IKRFYAQDGK VIPNSESTIA GIPGNSITKA YCDAQKTVFQ NTDDFTAKGG LVQMGKALAG DMVLVMSVWD DHAVNMLWLD STYPTDQVGV AGAERGACPT TSGVPSDVEA NAPNSNVIFS NIRFGPIGST VQGLPSSGGT SSSSSAAPQS TSTKASTTS AVRTTSTATT KTSSAPAAQG TNTAKHWQQC GGNGWTGPTV CESPYKCTKQ NDWYSQCL
SEQ ID NO:117	MLTLVYFLLS LVVSLEIGTQ QSEDHPKLTW QNGSSSVSGS IVLDSNWRWV HDSGTTNCYD GNLWSKDLCP SSDTCSQKCY IEGADYSGTY GIQSSGSKLT LKFVTKGSYS TNIGSRVYLL KDENTYESFK LKNKEFTFTV DDSKLNCGLN GALYFVAMDA DGGKAKYSSF KPGAKYGMGY CDAQCPHDMK FISGKANVDD WKPDNDENS GNGKLGTCES EMDIWEGNMK SQAYTVHACT KSGQYECTGQ QCGDTSQDGR FKGTCDKDG DYASWRWGDQ SFYGEKTV DTKQPVTVVTQ FIGDPLTEIR RLYVQGGKTI NNSKTSNLAD TYDSITDKFC DATKEASGDT NDFKAKGAMS GFSTNLNNGQ VLVMSLWDDH TANMLWLDST YPTDSSDSTA QRGPCPTSSG VPKDVESQHG DATVVFSDIK FGAINSTFKY N
SEQ ID NO:118	MLAAALFTEA CSVGVGTTKTP ENHPKLNWQN CASKGSCSQV SGEVTMDSNW RWTHDGNNGN CYDGNWTWISS LCPDDKTCSD KCVLDGAEYQ ATYGIQSNGT ALTLKFVTHG SYSTNIGSRL YLLKDKSTYY VFKLNNKEFT FSVDVSKLPC GLNGALYFVE MDADGGKAKY AGAKPGAHEYG LGYCDAQCPD DLKFINGEAN SEGWKPKQSGD KNAGNGKYGS CCSEMDVWES NSQATALTPH VCKTTGQQRC SGKSECGGQD GQDRFAGLCD EDGCDFNNWR MGDKTFFGPG LIVDTKSPFV VVTQFYGSPV TEIRRKYVQN GKVIENSKSN IPGIDATAAI SDHFCEQKKK AFGDTNDFKN KGGFAKLQGV FDRGMVLVLS LWDDHQVAML WLDSTYPTNK DKSQPGVDRG PCPTSSGKPD DVESASADAT VVYGNIKFGA LDSTY
SEQ ID NO:119	MLTLVYFLLS LVVSLEIGTQ QSEDHPKLTW QNGSSSVSGS IVLDSNWRWV HDSGTTNCYD GNLWSKDLCP SSNTCSQKCY IEGADYSGTY GIQSSGSKLT LKFVTKGSYS TNIGSRVYLL KDENTYESFK LKNKEFTFTV DDSKLNCGLN GALYFVAMDA DGGKAKYSSF KPGAKYGMGY CDAQCPHDMK FISGKANVDD WKPDNDENS GNGKLGTCES EMDIWEGNMK SQAYTVHACT KSGQYECTGQ QCGDTSQDGR FKGTCDKDG DYASWRWGDQ SFYGEKTV DTKQPVTVVTQ FIGDPLTEIR RLYVQGGKTI NNSKTSNLAD TYDSITDKFC DATKEASGDT NDFKAKGAMS GFSTNLNNGQ VLVMSLWDDH TANMLWLDST YPTDSTKTGA SRGPCAVSSG VPKDVESQYG DATVIYSDIK FGAINSTFKW N
SEQ ID NO:120	MILALLSLAK SLGIATNQAE THPKLTWTRY QSKGSGQTVN GEIVLDSNWR WTHHSGTNCY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGI STSGNSL KLGFVTHGSY STNIGSRVYL LRDSKNYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDACPHDM KFINGEANVL DWKPQSDEN SGNGRYGACC TEMDIWEANS MATAYTPHVC TVTGLRRC EG TECGDTDANQ RYNGICDKDG CDFNSYRLGD KTFFFGVGKTV DSSKPVTVVT QFVTSNGQDS GTLSEIRRKY VQGGKVIENS KVNIAGITAG NSVTDTFCNE QKKAFGDNND FEKKGGLGAL SKQLDAGMVL VLSLWDDHSV NMLWLDSTYP TNAAAGALGT ERGACATSSG APSDVESQSP DATVTFSDIK FGPIDSTY

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:121	MLVIALILRG LSVGTTGTQQS ETHPSLSWQQ TSKGGSGQSV SGSVVLDSNW RWTHTTDGTT NCYDGNWSS DLCPDASTCS SNCVLEGADY SGTYGITGSG SSLKLGFTVK GSYSTNIGSR VYLLGDESHY KLFKLENNEF TFTVDDDSNLE CGLNGALYFV AMDEDDGGASK YSGAKPGAKY GMGYCDAQCP HDMKFFINGDA NVEGWKPSDN DENAGTGKWG ACCTEMDIWE ANKYATAYTP HICTKNGEYR CEGTDCGDTK DNNRYGGVCD KDGCDFNSWR MGNQSFWGPG LIIDTGKPVV VVTQFLADGG SLSEIRRKYV QGGKVIENTV TKISGMDEFD SITDEFNCQQ KKAFRDITNDF EKKGGLKGLG TAVDAGVVLV LSLWDDHDVN MLWLDIYPT DSGSKAGADR GPCATSSGVP KDVESNYASA SVTFSDIKFG PIDSTY
SEQ ID NO:122	MLLALFAFGK SLGIATNQAE NHPKLTWTRY QSKGSGQTVN GEIVLDSNWR WTHHSGTNCY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGISSSGNSL KLGFVTHGSY STNIGSRVYL LRDSKNYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPQSDEN SGNGRYGACC TEMDIWEANS MATAYTPHVC TVTGIRRCCEG TECGDTDANQ RYNGICDKDG CDFNSYRLGD KSFFGVGKTV DSSKPVTVVT QFVTSNGQDS GTLSEIRRKY VQGGKVIENS KVNIAGMAAG NSITDTFCNE QKKAFGDNND FEKKGGLGAL SKQLDSGMVL VLSLWDDHSV NMLWLDSTYP TNAAAGALGT ERGACATSSG APSDVESQSP DATVTFSDIK FGPIDSTY
SEQ ID NO:123	MLASVVYLV LVSLEIGTQ QSEEHPKLTW QNGSSSVSGS IVLDSNWRWL HDSGTTNCYD GNLWSDDLCP NADTCSSKCY IEGADYSGTY GITSSGSKVT LKFVTKGSYS TNIGSRIYLL KDENTYETFK LKNKEFTFTV DSKLDCGLN GALYFVAMDA DGGKAKYSSF KPGAKYGMGY CDAQCPHDMK FISGKANVDD WKPQDNDENS GDGKLGTCES EMDIWEGNAK SQAYTVHACS KSGQYECTGQ QCGDITDSGDR FKGTCDKDG DYASWRWGDQ SFYGEGETVD TKSPVTVVTQ FIGDPLTEIR RVYVQGGKTI NNSKTSNLAD TYDSITDKFC DATKDATGDT NDFKAKGAMA GFSTNLNTAQ VLVSVHCGMI IQPICCGLIR RIQRIQQKV QAVDRVLCRR VFQRM LKASM VMLQSRTRL SLELSTRPLV GISPAGRLFF F
SEQ ID NO:124	MILALLVLGK SLGIATNQAE THPKLTWTRY QSKGSGSTVN GEIVLDSNWR WTHHSGTNCY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGISTSGNSL KLGFVTHGSY STNIGSRVYL LKDTKSYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPQSDEN SGNGRYGACC TEMDIWEANS MATAYTPHVC TVTGLRRCEG TECGDTDNDQ RYNGICDKDG CDFNSYRLGD KSFFGVGKTV DSSKPVTVVT QFVTSNGQDS GTLSEIRRKY VQGGKVIENS KVN VAGITAG NSVTDTFCNE QKKAFGDNND FEKKGGLGAL SKQLDAGMVL VLSLWDDHSV NMLWLDSTYP TNAAAGALGT ERGACATSSG KPSDVESQSP DATVTFSDIK FGPIDSTY
SEQ ID NO:125	MLCIGLISFV YSLGVGTNTA ETHPKLTWKN GGQTVNGEVT VDSNWRWTHT KGSTKNCYDG NLWSKDLCPD AATCGKNCVL EGADYSGTYG VTSSGNALTL KFTVTHGSYST NVGSRLYLLK DEKTYQM FNL NGKEFTFTVD VSNLPCGLNG ALYHVNMDED GGTKRYPDNE AGAKYGTGYC DAQCPTDLKF INGIPNSDGM KPQSNDKNSG NGKYGSCCSE MDIWEANSIC SAVTPHVCDN LQQTRCQGT A CGENG GGSRF GSSCDPDGCD FNSWRMGNKT FYGPGLIVDT KSKFTVVTTQF VGNPVTEIKR KYVQNGKVIE NSYSNIEGMD KFNSVSDKFC TAQKKAFGDT DSFTKHGGFK QLGSALAKGM VLVLSLWDDH TVNMLWLDSV YPTNSKKAGS DRGPCPTTSG VPADVESKSA DANVIYSDIR FGAIDSTYK
SEQ ID NO:126	MLGALVALAS CIGVGTNTPE KHPDLKWTNG GSSVSGSI VV DSNWRWTHIK GETKNCYDGN LWSDKYCPDA ATCGKNCVLE GADYSGTYGV TTSGDAATLK FVTHGQYSTN VGSRLYLLKD EKTYQM FNLV GKEFTFTVDV SNLPCGLNGA LYFVQMDS DG GMAKYPDNQA GAKYGTGYCD AQCPTDLKFI NGIPNSDGWK PQKNDKNSGN GKYGSCCSEM DIWEANSMAT AYTPHVCDKL EQTRCSGSAC GQNGGGDRFS SSCDPDG CDF NSWRMGNKT F WGPGLIVDTK KPVQVVTTQFV GSGGSVTEIK RKYVQGGKVI DNSMTNIAAM SKQYNSVSDE FCQAQKKA FG DNDSFTKHGG FRQLGATLSK GHVLVLSLWD DHDVNMLWLD SVYPTNSNKP GADRGPC KTS SGVPSDVESQ NADSTVKYSD IRFGAIDSTY SK
SEQ ID NO:127	MLAAALFTFA CSVGVGTKTT ETHPKLNWQQ CACKGSCSQV SGEVTMDSNW RWTHDNGKN CYDGN TWISS LCPDDKTCS D KCVLDGA EYQ ATYGIQSNGT ALTPKFVTHG SYSTNIGSRL YLLKDKSTYY VFQLNNKEFT FSVDVSKLPC GLNGALYFVE MDADGGKSKY AGAKPGA EYG LGYCDAQCP DLKFINGEAN SEGWK PQSGD KNAGNGKYGS CCSEMDVWES NSMATALTPH VCKTTGQTRC SGKSECGGQD GQDRFAGNCD EDGCDENNWR MGDKTFFGPG LTVDTKSPFV VVTQFYGSPV TEIRRKYVQN GKVIENAKSN IPGIDATNAI SDTFCEQ QKK AFGDTNDFKN KGGFTKLGSV FSRGMVLVLS LWDDHQVAML WLDSTYPTNK DKSVPGVDRG PCPTSSGKPD DVESASGDAT VVYGNIKFGA LDSTY

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:128	MFGELLSLFA LQFALEIGTQ TSEHPSITW ELNGARQSGQ IVIDSNWRWL HDSGTTNCYD GNTWSSDLCP DPEKCSQNCY LEGADYSGTY GISASGSQLT LGFVTKGSYS TNIGSRVYLL KDENTYPMFK LKNKEFTFTV DVSNLPCGLN GALYFVAMPS DGGKAKYPLA KPGAKYGMGY CDAQCPHDMK FINGEANVLD WKPQSN DENA GTGRYGTCTCT EMDIWEANSQ ATAYTVHACS KNARCEGTEC GDDSASQRYN GICDKDGCDF NSWRWGNKTF FGPGLTVDSS KPVTVVTQFI GDPLTEIRRI WVQGGKVIQN SFTNVSGITS VDSITNTFCD ESKVATGDTN DFKAKGGMSG FSKALDTEVV LVLSLWDDHT ANMLWLDSTY PTDSTAIGAS RGPCATSSGD PKDVESASAN ASVKFSDIKF GALDSTY
SEQ ID NO:129	MLASLLPLSN SLGTASNQAE THPKLTWTQY TGKGAGQTVN GEIVLDSNWR WTHKDGTCNY DGNTWSSSLC PDPTTCSNNC NLDGADYPGT YGITTSGNQL KLGFVTHGSY STNIGSRVYL LRDSKNYQMF KLKNKEFTFT VDDSKLPCGL NGAVYFVAMD EDGGTAKHSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPQSN DEN SGNGRWGARC TEMDIWEANS RATAYTPHIC TKTGLYRCEG TECGDSDTNR YGGVCDKDG C DFNSYRMGDK SFFGQGKTVD SSKPVTVVTQ FITDNNQDSG KLTEIRRKYV QGGKVIDNSK VNIAGITAGN PITDTFCDEA KKAFGDNND F EKKGGLSALG TQLEAGFVLV LSLWDDHSVN MLWLDSTYPT NASPGALGVE RGDCAITSGV PADVESQSAD ASVTFSDIKF GPIDSTY
SEQ ID NO:130	MLCIGLISFV YSLGVGTNTA ETHPKLTWKN GGQTVNGEVT VDSNWRWTHT KGSTKNICYDG NLWSKDLCPD AATCGKNCVL EGADYSGTYG VTSSGNALTL KFVTHGSYST NVGSRLYLLK DEKTYQMFNL NGKEFTFTVD VSNLPCGLSG ALYHVNMD ED GGTKRYPDNE AGAKYGTGYC DAQCPTDLKF INGIPNSDGW KPQSN DKNSG NGKYGSCCSE MDIWEANSIC SAVTPHVCDN LQQTRCQGAA CGENGGSRF GSSCDPDGCD FNSWGMGNKT FYGPGLIVDT KSKFTVVVTQF VGNPVT EIKR KYVQNGKVIE NSYSNIEGMD KFNSVSDKFC TAQKKAFGDT DSFTKHGGFK QLGSALAKGM VLVLSLWDDH TVNMLWLDSV YPTNSKKAGS DRGPCPTTSG VPADVESKSA DANVIYSDIR FGAIDSTYK
SEQ ID NO:131	MILALLVLGK SLGIATNQAE THPKLTWTRY QSKGSGSTVN GEIVLDSNWR WTHHSGTCNY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGISTSGNSL KLGFVTHGSY STNIGSRVYL LRDSKNYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPQSN DEN SGNGRYGACC TEMDIWEANS MATAYTPHVC TTVTGLRRC EG TECGDTDNDQ RYNGICDKDG CDFNSYRLGD KSFFGVGKTV DSSKPVTVVT QFVTSNGQDS GILSETRRKY VQGGKVIENS KVN VAGITAG NSVTD TFCNE QKKAFGDNND FEKKGGLGAL SKQLDAGMVL VLSLWDDHSV NMLWLDSTYP TNAAAGALGT ERGACATSSG KPSDVESQSP DATVTFSDIK FGPIDSTY
SEQ ID NO:132	MIGIVLIQTV FGIGVGTQQS ESHPSLSWQQ CSKGGSCTSV SGSIVLDSNW RWTHIPDGTT NCYDGN EWSS DLCPDPTTCS NNCVLEGADY SGTYGISTSG SSAKLGFVTK GSYSTNIGSR VYLLGDESHY KIFDLKNKEF TFTVDDSNLE CGLNGALYFV AMDEDGGASR FTLAKPGAKY GTGYCDAQCP HDIKFINGEA NVQDWKP SDN DDNAGTGHYG ACCTEMDIWE ANKYATAYTP HICTENGEYR CEGKSCGDSS DDRYGGVCDK DGCDFNSWRL GNQSFWGPGL IIDTGKPVTV VTQFVTKDGT DSGALSEIRR KYVQGGKTIE NTVVKISGID EVDSITDEF C NQQKQAFGDT NDFEKKGGLS GLGKAFDYGV VLVLSLWDDH DVNMLWLDSV YPTNPAGKAG ADRGPCATSS GDPKEVEDKY ASASVTFSDI KFGPIDSTY
SEQ ID NO:133	MLVFGIVSFV YSIGVGTNTA ETHPKLTWKN GGSTTNGEVT VDSNWRWTHT KGSTKNICYDG NLWSKDLCPD AATCGKNCVL EGADYSGTYG VTSSGDALTL KFVTHGSYST NVGSRLYLLK DEKTYQMFNL NGKEFTFTVD VSQLPCGLNG ALYFVCM DQD GGMSRYPDNQ AGAKYGTGYC DAQCPTDLKF INGLPNSDGW KPQSN DKNSG NGKYGSCCSE MDIWEANSLA TAVTPHVCDQ VGQTRCEGRA CGENGGGDRF GSICDPDGCD FNSWRMGNKT FWGPGLIIDT KKPVTVVVTQF IGPVTEIKR EYVQGGKVIE NSYTNIEGMD KFNSISDKFC TAQKKAFGDN DSFTKHGGFS KLGQSFTKGQ VLVLSLWDDH TVNMLWLDSV YPTNSKKLGS DRGPCPTSSG VPADVESKNA DSSVKYSDIR FGSIDSTYK
SEQ ID NO:134	MLSFVFLIGF GVSLEIGTQQ SENHPTLSWQ QCTSSGSCTS QSGSIVLDSN WRWVHDSGTT NCYDGN EWSS DLCPDPETCS KNCYLDGADY SGTYGITSNG SSLKLGFVTE GSYSTNIGSR VYLKDDTNTY QIFKLKNHEF TFTVDVSNLP CGLNGALYFV EMEADGGKGK YPLAKPGAQY GMGYCDAQCP HDMKFINGNA NVLDWK PQET DENSGNGRYG TCCTEMDIWE ANSQATAYTP HICTKDGQYQ CEGTECGDSD ANQRYNGVCD KDGCDFNSYR LGNKTEFFGPG LIVDSKKPVT VVTQFITSNG QDSGDLTEIR RIYVQGGKTI QNSFTNIAGL TSVDSITEAF CDESKDLFGD TNDFEKA KGF TAMGKS LDTG VVLVL SLWDD HSVNMLWLDS TYPTDAAAGA LGTQRGPCAT SSGAPSDVES QSPDASVTFS DIKFGPLDST Y

TABLE 7

TABLE 7										
SEQ ID NO.	Amino acid sequence									
SEQ ID NO:135	MLTLVVYLLS	LVVSLEIGTQ	QSESHPALTW	QREGSSASGS	IVLDSNWRWV	HDSGTTNCYD	GNEWSTDLC	SSDTCTQKCY	IEGADYSGTY	
	GITTSGSKLT	LKFVTKGSYS	TNIGSRVYLL	KDENTYETFK	LKNKEFTFTV	DDSKLDCGLN	GALYFVAMDA	DGGKQKYSF	KPGAKYGMGY	
	CDAQCPHDMK	FISGKANVED	WKPQDNDENS	GNGKLGTC	EMDIWEGNAK	SQAYTVHACT	KSGQYECTGT	DCGSDSRYQ	GTCDKDGC	DY
	ASYRWGDHSE	YGEKKTVDTK	QPITVVVTQFI	GDPLTEIRRL	YIQGGKVINN	SKTONLASVY	DSITDAFCDA	TKAASGDTND	FKAKGAMAGF	
	SKNLDTPQVL	VLSLWDDHTA	NMLWLDSTYP	TDSRDATAER	GPCATSSGVP	KDVESNQADA	SVVFSDIKFG	AINSTYSYN		
SEQ ID NO:136	MFGELLSLFA	LQFALEIGTQ	TSESHPSITW	ELNGARQSGQ	IVIDSNWRWL	HDSGTTNCYD	GNTWSSDLCP	DPEKCSQNCY	LEGADYSGTY	
	GISASGSQLT	LGFVTKGSYS	TNIGSRVYLL	KDENTYQMFK	LKNKEFTFTV	DVSNLPCGLN	GALYFVAMPS	DGGKAKYPLA	KPGAKYGMGY	
	CDAQCPHDMK	FINGEANVLD	WKPQSNDENA	GTGRYGTCT	EMDIWEANSQ	ATAYTVHACS	KNARCEGTEC	GDDASASQRYN	GICDKDGC	DF
	NSWRWGNKTF	FGPGLTVDSS	KPVTVVVTQFI	GDPLTEIRRI	WVQGGKVIQN	SFTNVSGITS	VDSITNTFC	ESKVATGDTN	DFKAKGGMMSG	
	FSKALDTEVV	LVLSLWDDHT	ANMLWLDSTY	PSNSTAIGAT	RGPCATSSGD	PKNVESASAN	ASVKFSDIKF	GAFDSTY		
SEQ ID NO:137	MLALVYFLLS	LVVSLEIGTQ	QSEDHPPKLTW	QNGSSSVSGS	IVLDSNWRWV	HDSGTTNCYD	GNLWSTDLC	SSDTCTSKCY	IEGADYSGTY	
	GITSSGSKVT	LKFVTKGSYS	TNIGSRIYLL	KDENTYETFK	LKNKEFTFTV	DDSQLNCGLN	GALYFVAMDA	DGGKAKYSSF	KPGAKYGMGY	
	CDAQCPHDMK	FISGKANVDD	WKPQDNDENS	GNGKLGTC	EMDIWEGNAK	SQAYTVHACT	KSGQYECTGQ	QCGDTS	GDR	FKGTCDKDG
	DYASWRWGDQ	SFYGEKTV	TKQPVTVVTQ	FIGDPLTEIR	RLYVQGGKTI	NNSKTSNLAD	TYDSITDKFC	DATKEASGDT	NDFKAKGAMS	
	GFSTNLNTAQ	VLVLSLWDDH	TANMLWLDST	YPTDSTKTGA	SRGPCAVTSG	VPKDVESQYG	SAQVVYSDIK	FGAINSTY		
SEQ ID NO:138	MLALVYFLLS	FVVSLEIGTQ	QSEDHPPKLTW	QNGSSSVSGS	IVLDSNWRWV	HDSGTTNCYD	GNLWSTDLC	SSDTCSSKCY	IEGADYSGTY	
	GISASGSKLT	LKFVTKGSYS	TNIGSRVYLL	KDENTYETFK	LKGKEFTFTV	DDSKLDCGLN	GALYFVAMDA	DGGKAKYSSF	KPGAKYGMGY	
	CDAQCPHDMK	FISGKANVDD	WKPQDNDENS	GNGKLGTC	EMDIWEGNAK	SQAYTVHACT	KSGQYECTGQ	QCGDTS	GDR	FKGTCDKDG
	DYASWRWGDQ	SFYGEKTV	TKQPVTVVTQ	FIGDPLTEIR	RVYVQGGKVI	NNSKTSNLAN	VYDSITDKFC	DDTKDATGDT	NDFKAKGAMS	
	GFSTNLNTAQ	VLVMSLWDDH	TANMLWLDST	YPTDSTKTGA	SRGPCAVLSG	VPKNVESQHG	DATVIYSDIK	FGAINSTFSY	N	
SEQ ID NO:139	MFLALFVLGK	SLGIATNQAE	NHPKLTWTRY	QSKGSGQTVN	GEVVLDSNWR	WTHHSGTNCY	DGNTWSTSLC	PDPQTCSSNC	DLDGADYPGT	
	YGISSSGNSL	KLGFVTHGSY	STNIGSRVYL	LRDSKNYEMF	KLKNKEFTFT	VDDSKLPCGL	NGALYFVAME	EDGGVAKNSI	NKAGAQQYGTG	
	YCDAQCPHDM	KFINGEANVL	DKPQSNDEN	SGNGRYGACC	IEMDIWEANS	MATAYTPHVC	TVTGIHRCEG	TECGD	T	DANQ
	CDFNSYRMGD	KSFFGVGKTV	DSSKPVTVVT	QFVTSNGQDG	GTLSEIKRKY	VQGGKVIENS	KVNIAGITAV	NSITDTFCNE	QKKA	FGDNND
	FEKKGGLGAL	SKQLDLGMVL	VLSLWDDHVS	NMLWLDSTYP	TDAAGALGT	ERGACATSSG	KPSDVESQSP	DASVTFSDIK	FGPIDSTY	
SEQ ID NO:140	MLLCLLSIAN	SLGVGTNTAE	NHPKLSWKNG	GSSVSGSVTV	DANWRWTHIK	GETKNCYDGN	LWSDKYCPDA	ATCGKNCVIE	GADYQGT	YGV
	SSSGDGLTLT	FVTHGQYSTN	VGSRLYLMDK	EKTYQMFNLN	GKEFTFTVDV	SNLPCGLNGA	LYFVQMDSDG	GMAKYPDNQA	GAKYGTGYCD	
	AQCPTDLKFI	NGIPNSDGWK	PQKNDKNSGN	GKYGSCCSEM	DIWEANSQAT	AYTPHVCDKL	EQTRCSGSSC	GHTGGGERFS	SSCDPDG	CDF
	NSWRMGNKTF	WGPGLIVDTK	KPVQVVVTQFV	GSGNSCTEIK	RKYVQGGKVI	DNSMSNIAGM	SKQYNSVSD	FCQAQKKA	FG	DNDSFTKHGG
	FRQLGATLGK	GHVLVLSLWD	DHDVNMLWLD	SVYPTNSNKP	GSDRGPCCKTS	SGIPADVESQ	AASSSVKYSD	IRFGAIDSTY	K	
SEQ ID NO:141	MLCIGLISFV	YSLGVGTNTA	ETHPKLTWKN	GGQTVNGEVT	VDSNWRWTH	KGSTKNCYDG	NLWSKDLCPD	AATCGKNCVL	EGADYSGTYG	
	VTSSSGNALTL	KFVTHGSYST	NVGSRLYLMDK	DEKTYQMFNL	NGKEFTFTVD	VSNLPCGLNG	ALYHVNMDED	GGTKRYPDNE	AGAKYGTGYC	
	DAQCPTDLKF	INGIPNSDGW	KPQSNDKNSG	NGKYGSCCSE	MDIWEANSIC	SAVTPHVCDT	LQTRCQGT	CGENGGSRF	GSSCDPDG	C
	FNSWRMGNKT	FYGPGLIVDT	KSKFTVVVTQF	VGSPVTEIKR	KYVQNGKVIE	NSFSNIEGMD	KFNSISDKFC	TAQKKA	FGDT	DSFTKHGGFK
	QLGSALAKGM	VLVLSLWDDH	TVNMLWLDV	YPTNSKKAGS	DRGPCPTTSG	VPADVESKSA	NANVIYSDIR	FGAIDSTYK		

TABLE 7

TABLE 7	
SEQ ID NO.	Amino acid sequence
SEQ ID NO:142	MLLCLLGIAS SLDAGTNTAE NHPQLSWKNG GSSVSGSVTV DANWRWTHIK GETKNCYDGN LWSDKYCPDA ATCGQNCVIE GADYQGTYGV SASGNALTIT FVTHGQYSTN VGSRLYLLKD EKTYQIFNLI GKEFTFTVDV SNLPCGLNGA LYFVQMDADG GTAKYSDNKA GAKYGTGYCD AQCPDCLKFI NGIPNSDGWK PQKNDKNSGN GRYGSCCSEM DVWEANSLAT AYTPHVCDKL EQVRCDGRAC GQNGGGDRFS SSCDPDGCDF NSWRLGNKTF WGPGLIIVDTK QPVQVVVTQV GSGTSVTEIK RKYVQGGKVI DNSFTKLDSL TKQYNSVSDE FCVAQKKAFG DNDSFTKHGG FRQLGATLAK GHVLVLSLWD DHDVNMLWLD SVYPTNSNKP GADRGPCCKTS SGVPADVESQ AASSSVKYSD IRFGAIDSTY K
SEQ ID NO:143	MLGIGFVCIV YSLGVGTNTA ENHPKLTWKN SGSTTNGEVT VDSNWRWTHT KGTTKNCYDG NLWSKDLCPD AATCGKNCVL EGADYSGTYG VTSSGDALT L KFTVTHGSYST NVGSRLYLLK DEKTYQIFNL NGKEFTFTVD VSNLPCGLNG ALYFVNMDAD GGTGRYPDNQ AGAKYGTGYC DAQCPTDLKF INGIPNSDGW KPQSNDKNSG NGKYGSCCSE MDIWEANSLA TAVTPHVCDQ VGQTRCEGRA CGENGGGDRF GSSCDPDGCD FNSWRLGNKT FWGPGLIIVDT KKPVTVVTQF VGSPVTEIKR KYVQGGKVI NSYTNIEGLD KFNSISDKFC TAQKKAFGDN DSFIKHGGFR QLQSFTEKGQ VLVLSLWDDH TVNMLWLDV YPTNSKKPGA DRGPCPTSSG VPADVESKNA GSSVKYSDIR FGSIDSTYK
SEQ ID NO:144	MATLVGILVS LFALEVALEI GTQTSESHPS LSWELNGQRQ TGSIVIDSNW RWLHDSGTTN CYDGNESWSSD LCPDPEKCSQ NCYLEGADYS GTYGISSSGN SLQLGFVTKG SYSTNIGSRV YLLKDENTYA TFKLKNKEFT FTADVSNLPC GLNGALYFVA MPADGGKSKY PLAKPGAKYG MGYCDAQCPH DMKFINGEAN ILDWKPSSND ENAGAGRYGT CCTEMDIWEA NSQATAYTVH ACSKNARCEG TECGDDDDGRY NGICDKDGCD FNSWRWGNKT FFGPNLIVDS SKPVTVVTQF IGDPLTEIRR IYVQGGKVIQ NSFTNISGVA SVDSITDAFC NENKVATGDT NDFKAKGGMS GFSKALDTEV VLVLSLWDDH TANMLWLDST YPTDSSALGA SRGPCAITSG EPKDVESASA NASVKFSDIK FGAIDSTY
SEQ ID NO:145	MLTLVYFLLS LVVSLEIGTQ QSESHPQLSW QNGSSSVSGS IVLDSNWRWV HDSGTTNCYD GNLWSTDLCP SSdTCTSKCY IEGADYSGTY GITSSGSKLT LKFVTKGSYS TNIGSRVYLL KDENTYETFK LKNKEFTFTV DSKLDCGLN GALYFVAMDA DGGKAKYSSF KPGAKYGMGY CDAQCPHDMK FISGKANVDD WKPDNDENS GNGKLGTCSS EMDIWEGNAK SQAYTVHACT KSGQYECTGQ QCQGTDSGDR FKGTCDKDG C DYASWRWGDQ SFYGEKTV D TKQPLTVVTQ FVGDPLTEIR RVYVQGGKTI NNSKTSNLAD TYDSITDKFC DATKEASGDT NDFKAKGAMS GFSTNLNTAQ VLVMSLWDDH TANMLWLDST YPTDSTKTGA SRGPCAVSSG VPKDVESQHG DATVIYSDIK FGAINSTFKW N
SEQ ID NO:146	MLSLSVIFLV GLGFSLGVGT QQSESHPSLS WQNCSAKGSC QSVSGSIVLD SNWRWLHDSG TTNCYDGNESW STDLCPDAST CDKNCYIEGA DYSGTYGITS SGAQLKLGFV TKGSYSTNIG SRVYLLRDES HYQLFKLKNH EFTFTVDDSQ LPCGLNGALY FVEMAEDGGA KPGAQYGMGY CDAQCPHDMK FITGEANVKD WKPQETDENA GNGHYGACCT EMDIWEANSQ ATAYTPHICS KTGIYRCEGT ECGNDANQR YNGVCDKDG C DFNSYRLGNK TFWGPGLTVD SNKAMIVVTQ FTTSNNQDSG ELSEIRRIYV QGGKTIQNSD TNVQGITTTN KITQAFCDET KVTFGDTNDF KAKGGFSGLS KSLESGAVLV LSLWDDHSVN MLWLDSTYPT DSAGKPGADR GPCAITSGDP KDVESQSPNA SVTFSDIKFG PIDSTY
SEQ ID NO:147	MILALLVLGK SLGIATNQAE THPKLTWTRY QSKGSGSTVN GEIVLDSNWR WTHHSGTNCY DGNTWSTSLC PDPTTCSNNC DLDGADYPGT YGISTSGNSL KLGFVTHGSY STNIGSRVYL LKDTKSİYEMF KLKNKEFTFT VDDSKLPCGL NGALYFVAMD EDGGVSKNSI NKAGAQYGTG YCDAQCPHDM KFINGEANVL DWKPQSNDEN SGNRGYGACC TEMDIWEANS MATAYTPHVC TVTGLRRC EG TECGDTDNDQ RYNGICDKDG CDFNSYRLGD KSFFGVGKTV DSSKPVTVVT QFVTSNGQDS GTLSEIRRKY VQGGKVIENS KVNAGITAG NSVTDTFCNE QKKAFGDNND FEKKGGFGAL SKQLVAGMVL VLSLWDDHSV NMLWLDSTYP TNAAAGALGT ERGACATSSG KPSDVESQSP DATVTFSDIK FGPIDSTY
SEQ ID NO:148	MLCVGLFGLV YSIGVGTNTQ ETHPKLSWKQ CSSGGSCCTQ QGSVIDSNW RWTHSTKDLT NCYDGNLWDS TLCPDGTTCS KNCVLEGADY SGTYGITSSG DSLTLKFVTH GSYSTNVGSR LYLLKDDNNY QIFNLAGEEF TFTVDVSNLP CGLNGALYFV EMDQDGGKGK HKENEAGAKY GTGYCDAQCP TDCLKFIDGIA NSDGWKPDQN DENSGNGKYG SCCSEMDIWE ANSLATAYTP HVCDTKGQKR CQGTACGENG GGDRFGSECD PDGCDFNSWR QGNKSFWGPG LIIDTKKSVQ VVTQFIGSGS SVTEIRRKYV QNGKVIENSÝ STISGTEKYN SISDDYCNAQ KKAFGDTNSF ENHGGFKRFS QHIQDMVLVL SLWDDHTVNM LWLDSVYPTN SNKPGADRG CETS SGVPAD VESKSASASV KYSDIRFGPI DSTYK

TABLE 7

SEQ ID NO.	Amino acid sequence
SEQ ID NO:149	MLLCLWSIAY SLGVGTNTAE NHPKLSWKNG GSSVSGSVTV DANRWTHIK GETKNCYDGN LMSDKYCPDA ATCGKNCVIE GADYQGTYGV SASGDGLTLT FVTHGQYSTN VGSRLYLMDK EKTYQIFNLN GKEFTFTVDV SNLPCGLNGA LYFVQMDSDG GMAKYPDNQA GAKYGTGYCD AQCPTDLKEI NGIPNSDGMK PQKNDKNSGN GKYGCCSEM DIWEANSQAT AYTPHVCCKL EQTRCSGSAC GHTGGERFS SSCDPDGCDF NSWRMGNKTE WGPGLIVDTK KPVQVVTQFV GSGNSCTEIK RKYVQGGKVI DNSMSNIAGM TKQYNSVSDD FCQAQKKAFG DNDSFTKHGG FRQLGATLGK GHVLVLSLWD DHVNMMLWLD SVYPTNSNKP GSDRGPKTS SGIPADVESQ AASSSVKYSY IRFGAIDSTY K
SEQ ID NO:299	QSACTLQSET HPPLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC YDGNWSSSTL CPDNETCAKN CCLDGAAYAS TYGVTTSGNS LSIGFVTQSA QKNVGARLYL MASDTTYQEF TLLGNEFSD VDVSQLPCGL NGALYFVSMAD ADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KEINGQANVE GWEPSNNAN TGIGGHGSCC SEMDIWEANS ISEALTPHPC TTVGQEICEG DCGGTYSDN AYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTTKKLTV VTQFETSGAI NRYVQNGVT FQOPNAELGS YSGNELNDDY CTAEAEAFGG SSFSDKGGLT QFKKATSGGM VLVMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS SGVPAQVESQ SPNAKVTFNS IKFGPIGSTG NPSGGNPPGG NPPGTTTTRR PATTGSSPG PTQSHYGQCG GIGYSGPTVC ASGTTCCQLN PYYSQCL
SEQ ID NO:300	QSACTLQSET HPPLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC YDGNWSSSTL CPDNETCAKN CCLDGAAYAS TYGVTTSGNS LSIGFVTQSA QKNVGARLYL MASDTTYQEF TLLGNEFSD VDVSQLPCGL NGALYFVSMAD ADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KEINGQANVE GWEPSNNAN TGIGGHGSCC SEMDIWEANS ISEALTPHPC TTVGQEICEG DCGGTYSDN AYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTTKKLTV VTQFETSGAI NRYVQNGVT FQOPNAELGS YSGNELNDDY CTAEAEAFGG SSFSDKGGLT QFKKATSGGM VLVMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS SGVPAQVESQ SPNAKVTFNS IKFGPIGSTG NPSGGNPPGG NPPGTTTTRR PATTGSSPG PTQSHYGQCG GIGYSGPTVC ASGTTCCQLN PYYSQCL
SEQ ID NO:301	MSALNSFNMY KSALILGSL ATAGAQQIGT YTAETHPSLS WSTCKSGGSC TTNSGAILTD ANRWVHGVN TSTNCYTGNT WNTAICDTDA SCAQDCALDG ADYSGTYGIT TSGNSLRNLF VTGSNVGSR YLMADNTHYQ IFDLINQEFT FTVDVSHLPC GLNGALYFVT MDADGGVSKY PNNKAGAQY VGYCDSQCPR DLKFIAGQAN VEGWTPSSNN ANTGLGNHGA CCAELDIWEA NSISEALTPH PCDTPLGSLVC TTDACGGTYS SDKYAGTCDP DGCDFNPYRL GVTDFYGSCK TVDTTKPITV VTQFVTDGDT STGTLSEIRR YVQNGVVIP QPSSKISGVS GNVINSDFC AEISTFGETA SEFKHGGGLAK MGAGMEAGMV LVMSLWDDYS VNMLWLDSTY PTNATGTPGA AKGSCPTTSG DPKTVESQSG SSYVTFSDIR VGPFNSTFSG GSSTGGSSTT TASGTTTTKA SSTSTSTST GTGVAHWHGQ CGGQGWGTGPT TCASGTTCTV VNPYYSQCL
SEQ ID NO:302	QQIGTYTAET HPSLSWSTCK SGGSCCTNSG AITLDANWRW VHGVTSTNC YTGNTWNTAI CDTDASCAQD CALDGADYSY TYGITTSGNS LRLNFTGNS VGSRTYLMAD NTHYQIFDILL NQEFFFTVDV SHLPCGLNGA LYFVTMDADG GVSYPNNKA GAQYGVGYCD SQCPRLKFI AGQANVEGWT PSSNNANTGL GNHGACCAEL DIWEANSISE ALTPHPCDTP GLSVCTTDAC GGTYSDDKYA GTCDDPDGCDF NPYRLGVTDF YSGGKTVDTT KPITVVTQFV TDDGTSTGTL SEIRRYVQON GVVIPQPSK ISGVSGNVIN SDFCDAEIST FGETASFH KH GGLAKMGAGM EAGMVLVMSL WDDYSVNMLW LDSTYPTNAT GTPGAAGKSC PTTSGDPKTV ESQSGSSYVT FSDIRVGPEN STFGSGSSTG GSSTTTASGT TTTKASSTST SSTSTGTGVA AHWGQCGGQG WTGPTTCASG TTCTVNPYY SQCL

WHAT IS CLAIMED IS:

1. A polypeptide comprising a variant cellobiohydrolase I (“CBH I”) catalytic domain as compared to a reference CBH I catalytic domain, comprising:
 - (a) a substitution at the amino acid position corresponding to R268 of *T. reesei* CBH I (“R268 substitution”);
 - (b) a substitution at the amino acid position corresponding to R411 of *T. reesei* CBH I (“R411 substitution”); or
 - (c) both an R268 substitution and an R411 substitution,wherein substitution (a), (b) or (c) decreases product inhibition as compared to the reference CBH I catalytic domain.
2. The polypeptide of claim 1, which comprises an R268 substitution.
3. The polypeptide of claim 2, wherein the R268 substituent is a lysine.
4. The polypeptide of claim 2, wherein the R268 substituent is an alanine.
5. The polypeptide of any one of claims 1 to 4, which comprises an R411 substitution.
6. The polypeptide of claim 5, wherein the R411 substituent is a lysine.
7. The polypeptide of claim 5, wherein the R411 substituent is an alanine.
8. The polypeptide of any one of claims 1 to 7, wherein the variant CBH I catalytic domain comprises an amino acid sequence having at least 90% sequence identity to amino acids 18-444 of SEQ ID NO:2.
9. The polypeptide of claim 8, wherein the variant CBH I catalytic domain comprises an amino acid sequence having at least 95% sequence identity to amino acids 18-444 of SEQ ID NO:2.

10. The polypeptide of claim 9, wherein, other than said R268 and/or R411 substitutions, the variant CBH I catalytic domain comprises the sequence of amino acids 18-444 of SEQ ID NO:2.

11. The polypeptide of any one of claims 8 to 10, wherein the variant CBH I catalytic domain comprises one of the following amino acid substitutions or pairs of amino acid substitution as compared to a protein of SEQ ID NO:2:

- (a) R268K and R411K;
- (b) R268K and R411A;
- (c) R268A and R411K;
- (d) R268A and R411A;
- (e) R268A;
- (f) R268K;
- (g) R411A; and
- (h) R411K.

12. The polypeptide of any one of claims 8 to 11, wherein the variant CBH I catalytic domain does not comprise a R268A substitution.

13. The polypeptide of claim 12 whose amino acid sequence does not comprise SEQ ID NO:299.

14. The polypeptide of claim 12 whose amino acid sequence does not consist of SEQ ID NO:299.

15. The polypeptide of any one of claims 8 to 14 wherein the variant CBH I catalytic domain does not comprise a R411A substitution.

16. The polypeptide of claim 15 whose amino acid sequence does not comprise SEQ ID NO: 301 or SEQ ID NO:300.

17. The polypeptide of claim 15 whose amino acid sequence does not consist of SEQ ID NO: 301 or SEQ ID NO:300.

18. A polypeptide comprising an amino acid sequence having at least 95% sequence identity to the amino acid sequence corresponding to positions 18-444 of SEQ ID NO:2, which has an R268K substitution and an R411A substitution as compared to a protein of SEQ ID NO:2.

19. The polypeptide of claim 18 in which said amino acid sequence has at least 97% sequence identity to the amino acid sequence corresponding to positions 18-444 of SEQ ID NO:2.

20. The polypeptide of any one of claims 1 to 7, wherein the variant CBH I catalytic domain comprises an amino acid sequence having at least 90% sequence identity to amino acids 26-455 of SEQ ID NO:1.

21. The polypeptide of claim 20, wherein the variant CBH I catalytic domain comprises an amino acid sequence having at least 95% sequence identity to amino acids 26-455 of SEQ ID NO:1.

22. The polypeptide of claim 21, wherein, other than said R268 and/or R411 substitutions, the variant CBH I catalytic domain comprises the sequence of amino acids 26-455 of SEQ ID NO:1.

23. The polypeptide of any one of claims 20 to 22, wherein the variant CBH I catalytic domain comprises one of the following amino acid substitutions or pairs of amino acid substitutions as compared to a protein of SEQ ID NO:1:

- (a) R273K and R422K;
- (b) R273K and R422A;
- (c) R273A and R422K;
- (d) R273A and R422A;
- (e) R273A;
- (f) R273K;
- (g) R422A; and
- (h) R422K.

24. The polypeptide of any one of claims 20 to 23, wherein the variant CBH I catalytic domain comprises the amino acid substitutions R273K and R422K as compared to a protein of SEQ ID NO:1.
25. The polypeptide of any one of claims 20 to 23, wherein the variant CBH I catalytic domain does not comprise both R273K and R422K substitutions as compared to a protein of SEQ ID NO:1.
26. The polypeptide of claim 25 whose amino acid sequence does not comprise SEQ ID NO: 301 or SEQ ID NO:302.
27. The polypeptide of claim 25 whose amino acid sequence does not consist of SEQ ID NO: 301 or SEQ ID NO:302.
28. The polypeptide of any one of claims 1 to 7, wherein the variant CBH I catalytic domain comprises an amino acid sequence having at least 90% or at least 95% sequence identity of the amino acid sequence of the catalytic domain of any one of SEQ ID NOs:1-149.
29. The polypeptide of any one claims 1 to 28 in which the variant CBH I catalytic domain is operably linked to a cellulose binding domain.
30. The polypeptide of claim 29 in which the catalytic domain is operably linked to a cellulose binding domain via a linker.
31. The polypeptide of claim 29 or claim 30 in which the cellulose binding domain is C-terminal to the catalytic domain.
32. The polypeptide of claim 29 or claim 30 in which the cellulose binding domain is N-terminal to the catalytic domain.
33. The polypeptide of any one of claims 1 to 32 which is a mature polypeptide.
34. The polypeptide of any one of claims 1 to 32 which further comprises a signal sequence.

35. The polypeptide of any one of claims 1 to 34 towards which cellobiose has an IC_{50} that is at least 2-fold the IC_{50} of a reference CBH I lacking said R268 substitution and/or R411 substitution.

36. The polypeptide of any one of claims 1 to 35 which CBH I activity that is at least 50% the CBH I activity of a reference CBH I lacking said R268 substitution and/or R411 substitution.

37. A composition comprising a polypeptide according to any one of claims 1 to 36.

38. The composition of claim 37 in which said polypeptide represents at least 1% of all polypeptides in said composition.

39. The composition of claim 38 in which said polypeptide represents at least 5% of all polypeptide in said composition.

40. The composition of claim 39 in which said polypeptide represents at least 25% of all polypeptide in said composition.

41. The composition of any one of claims 37 to 40 which is a whole cellulase.

42. The composition of claim 41, wherein the whole cellulase is produced by a host cell that recombinantly expresses said polypeptide.

43. The composition of any one of claims 37 to claim 42 which is filamentous fungal whole cellulase.

44. A fermentation broth comprising a polypeptide according to any one of claims 1 to 36.

45. The fermentation broth of claim 44, which is a filamentous fungal fermentation broth.

46. The fermentation broth of claim 44 or claim 45 which is a cell-free fermentation broth.

47. A method for saccharifying biomass, comprising: treating biomass with a composition according to any one of claims 37 to 43 or with a fermentation broth according to any one of claims 44 to 46.

48. The method of claim 47, further comprising recovering monosaccharides.

49. A method for producing ethanol, comprising:

(a) treating biomass with a composition according to any one of claims 37 to 43 or with a fermentation broth according to any one of claims 44 to 46, thereby producing monosaccharides; and

(b) culturing a fermenting microorganism in the presence of the monosaccharides produced in step (a) under fermentation conditions, thereby producing ethanol.

50. The method of claim 49, further comprising, prior to step (a), pretreating the biomass.

51. The method of claim 49 or claim 50, wherein said fermenting microorganism is a bacterium or a yeast.

52. The method of claim 51, wherein said fermenting microorganism is a bacterium selected from *Zymomonas mobilis*, *Escherichia coli* and *Klebsiella oxytoca*.

53. The method of claim 51, wherein said fermenting microorganism is a yeast selected from *Saccharomyces cerevisiae*, *Saccharomyces uvarum*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Candida pseudotropicalis*, and *Pachysolen tannophilus*.

54. The method of any one of claims 49 to 53, wherein said biomass is corn stover, bagasses, sorghum, giant reed, elephant grass, miscanthus, Japanese cedar, wheat straw, switchgrass, hardwood pulp, softwood pulp, crushed sugar cane, energy cane, or Napier grass.

55. A nucleic acid comprising a nucleotide sequence encoding the polypeptide of any one of claims 1 to 36.

56. A vector comprising the nucleic acid of claim 55.

57. The vector of claim 56 which further comprises an origin of replication.
58. The vector of claim 56 or claim 57 which further comprises a promoter sequence operably linked to said nucleotide sequence.
59. The vector of claim 58, wherein the promoter sequence is operable in yeast.
60. The vector of claim 58, wherein the promoter sequence is operable in filamentous fungi.
61. A recombinant cell engineered to express the nucleic acid of claim 55.
62. The recombinant cell of claim 61 which is a eukaryotic cell.
63. The recombinant cell of claim 62 which is a filamentous fungal cell.
64. The recombinant cell of claim 63, wherein the filamentous fungal cell is of the genus *Aspergillus*, *Penicillium*, *Rhizopus*, *Chrysosporium*, *Myceliophthora*, *Trichoderma*, *Humicola*, *Acremonium* or *Fusarium*.
65. The recombinant cell of claim 63, wherein the filamentous fungal cell is of the species *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Penicillium chrysogenum*, *Myceliophthora thermophila*, or *Rhizopus oryzae*.
66. The recombinant cell of claim 62 which is a yeast cell.
67. The recombinant cell of claim 66 which is a yeast cell of the genus *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Klockera*, *Schwanniomyces* or *Yarrowia*.
68. The recombinant cell of claim 67, wherein the yeast cell is of the species *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus* or *K. fragilis*.
69. The recombinant cell of claim 68, which is a *S. cerevisiae* cell.

70. A host cell transformed with the vector of any one of claims 56 to 60.
71. The host cell of claim 70 which is a prokaryotic cell.
72. The host cell of claim 71 which is a bacterial cell.
73. The host cell of claim 70 which is a eukaryotic cell.
74. A method of producing a polypeptide according to any one of claims 1 to 36, comprising culturing the recombinant cell engineered to express said polypeptide under conditions in which the polypeptide is expressed.
75. The method of claim 74, wherein the polypeptide comprises a signal sequence and wherein the recombinant cell is cultured under conditions in which the polypeptide is secreted from the recombinant cell.
76. The method of claim 75, further comprising recovering the polypeptide from the cell culture.
77. The method of claim 76, wherein recovering the polypeptide comprises a step of centrifuging away cells and/or cellular debris.
78. The method of claim 76, wherein recovering the polypeptide comprises a step of filtering away cells and/or cellular debris.
79. A method for generating a product tolerant variant CBH I polypeptide, comprising
- (a) modifying the nucleotide sequence of a CBH I-encoding nucleic acid so that the nucleic acid encodes a variant CBH I polypeptide, wherein said variant CBH I polypeptide comprises:
 - (i) an R268 substitution;
 - (ii) an R411 substitution; or
 - (iii) both an R268 substitution and an R411 substitution; and
 - (b) expressing said variant CBH I polypeptide,

thereby generating a product tolerant variant CBH I polypeptide.

80. A method for generating a nucleic acid that encodes a product tolerant variant CBH I polypeptide, comprising modifying the nucleotide sequence of a CBH I-encoding nucleic acid so that the nucleic acid encodes a variant CBH I polypeptide, wherein said variant CBH I polypeptide comprises:

- (i) an R268 substitution;
- (ii) an R411 substitution; or
- (iii) both an R268 substitution and an R411 substitution,

thereby generating a nucleic acid that encodes a product tolerant variant CBH I polypeptide.

81. The method of claim 79 or claim 80, wherein the modification is by site directed mutagenesis.

82. The method of any one of claims 79 to 81, wherein variant CBH I polypeptide comprises an R268 substitution.

83. The method of claim 82, wherein the R268 substituent is a lysine.

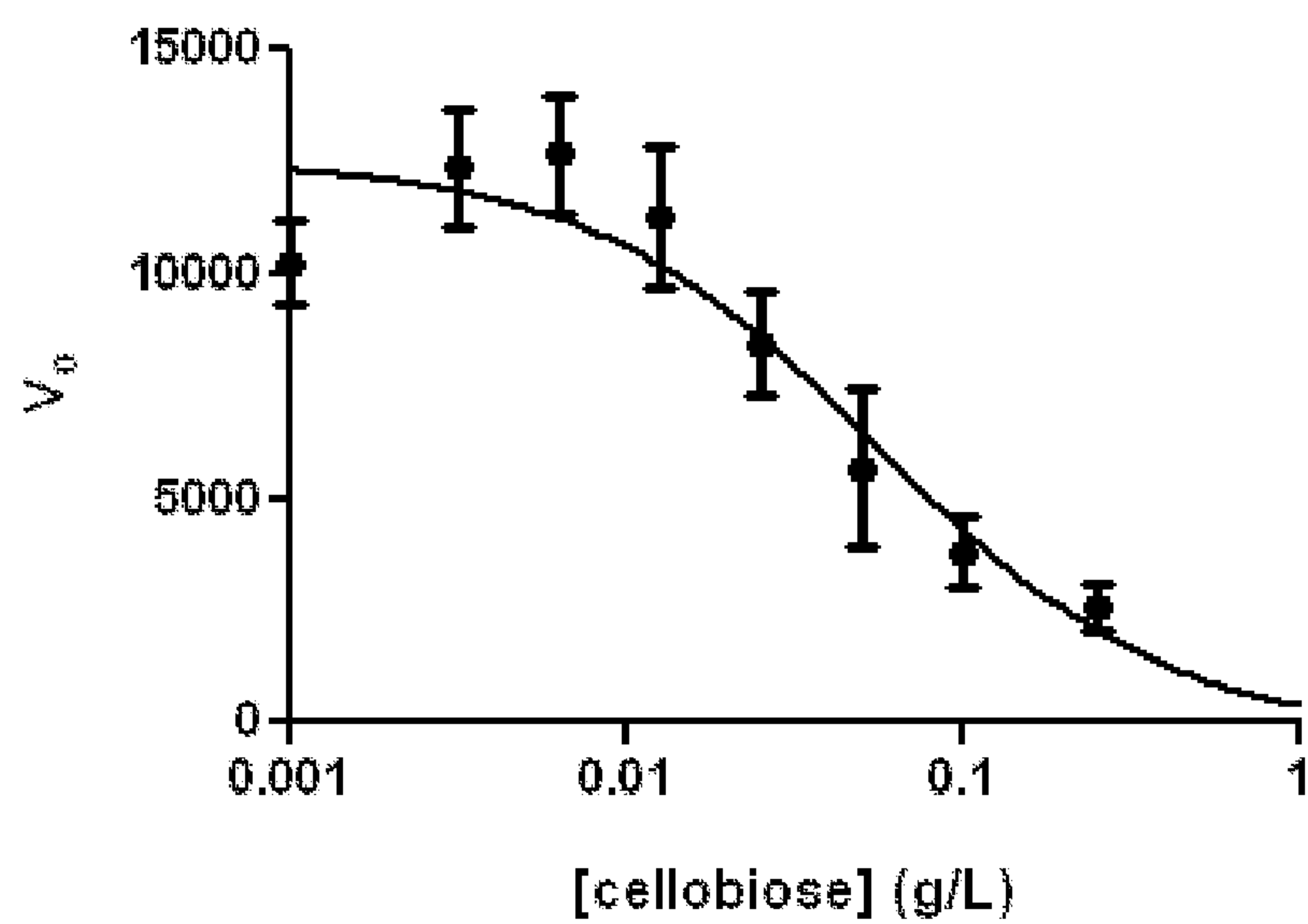
84. The method of claim 82, wherein the R268 substituent is an alanine.

85. The method of any one of claims 79 to 84, which comprises an R411 substitution.

86. The method of claim 85, wherein the R411 substituent is a lysine.

87. The method of claim 85, wherein the R411 substituent is an alanine.

A.



B.

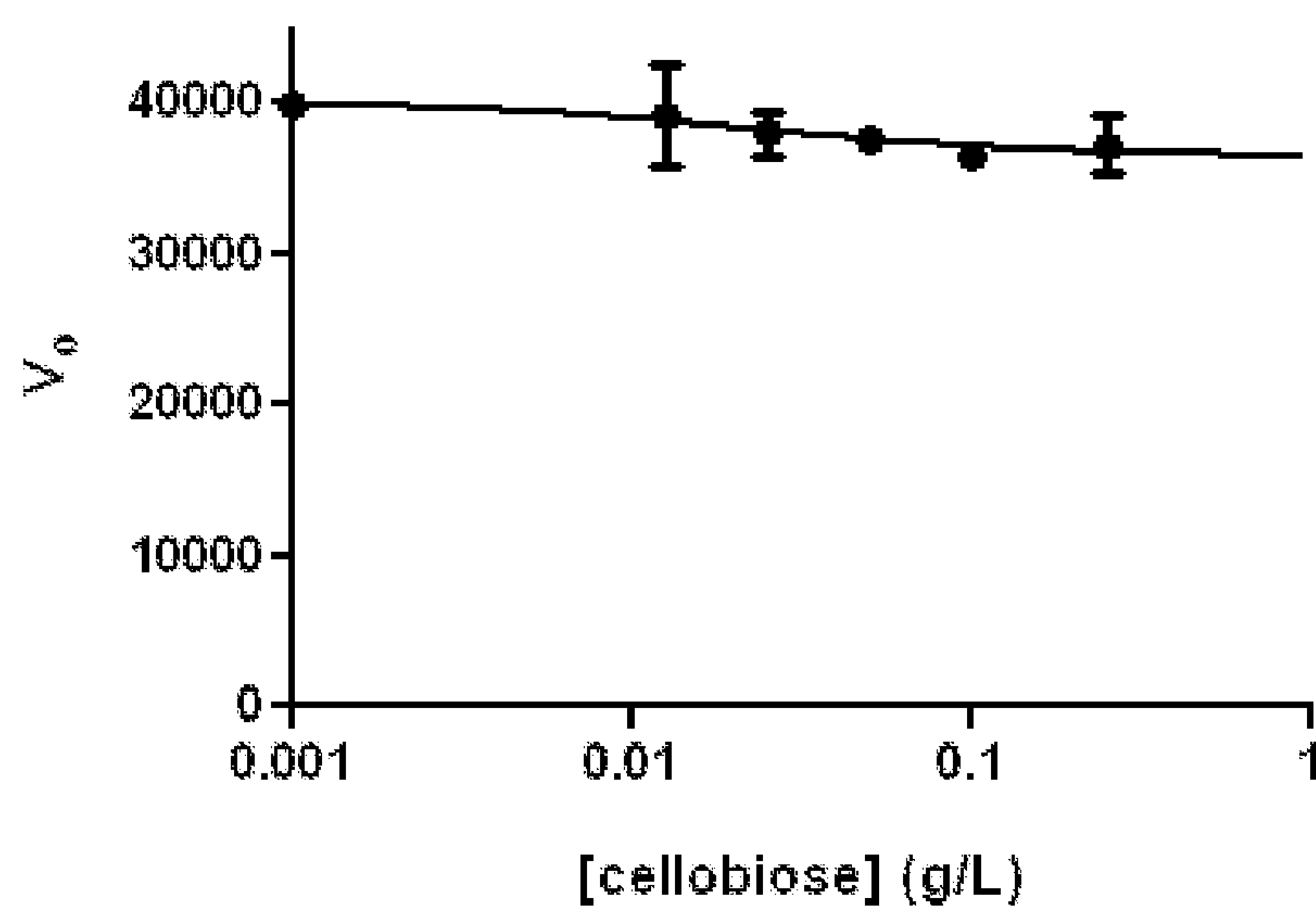
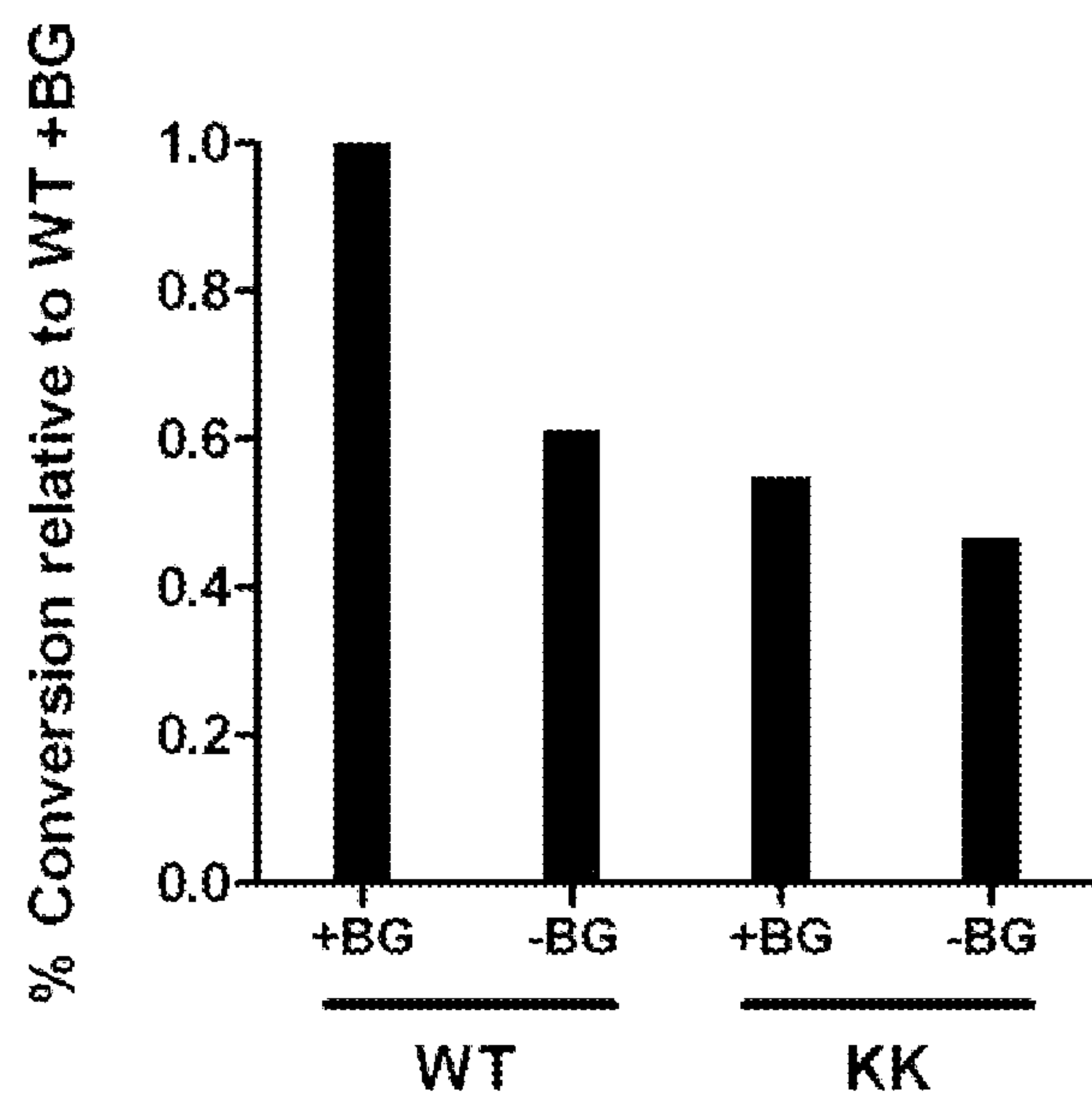


FIGURE 1A-1B

A.



B.

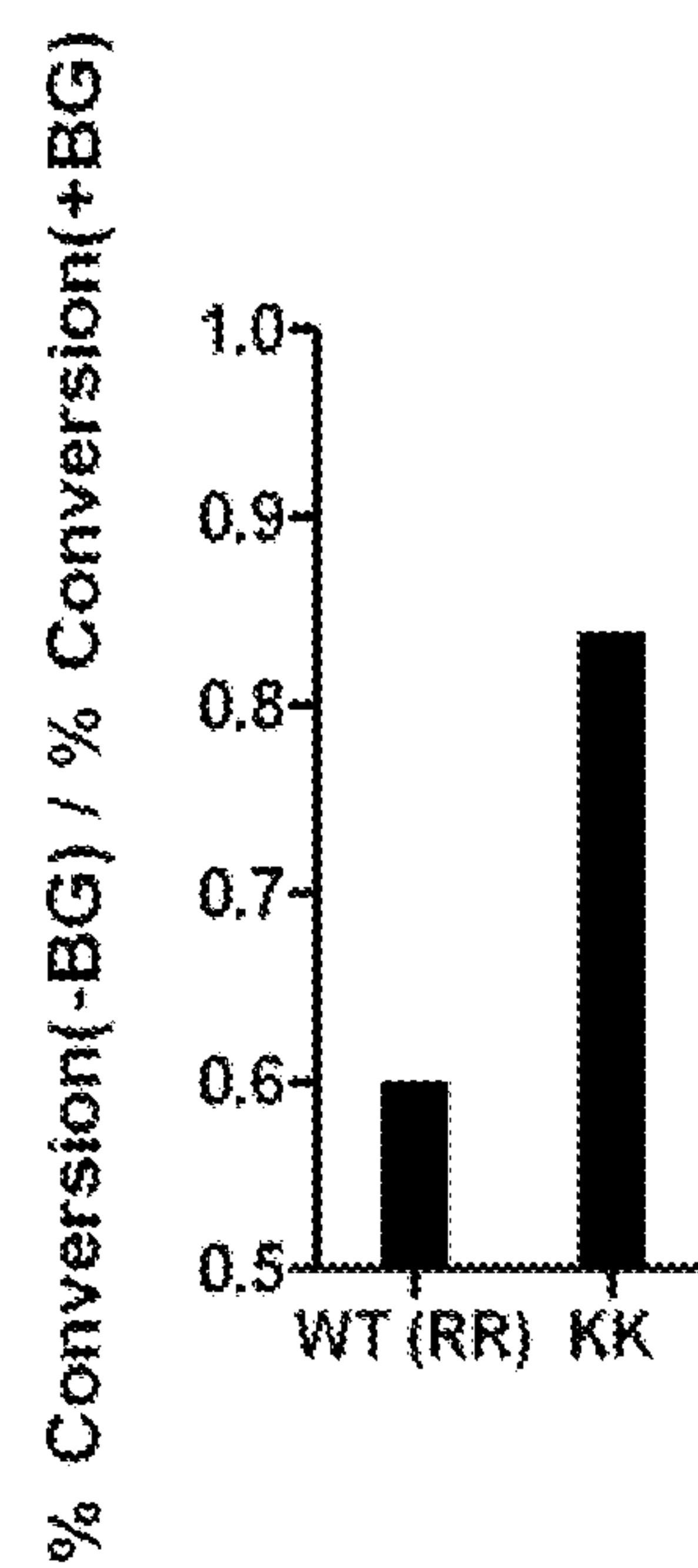


FIGURE 2A-2B

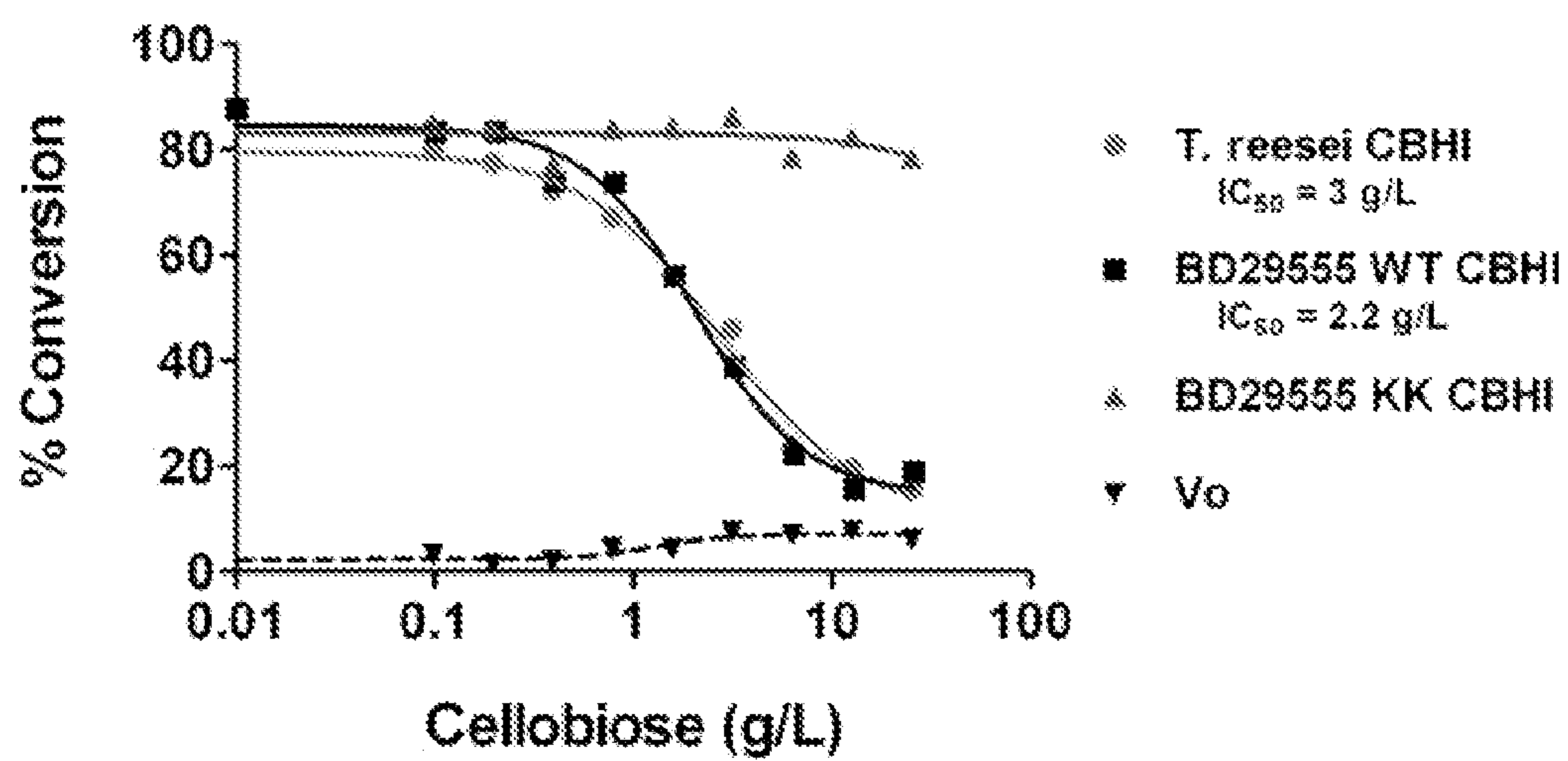


FIGURE 3

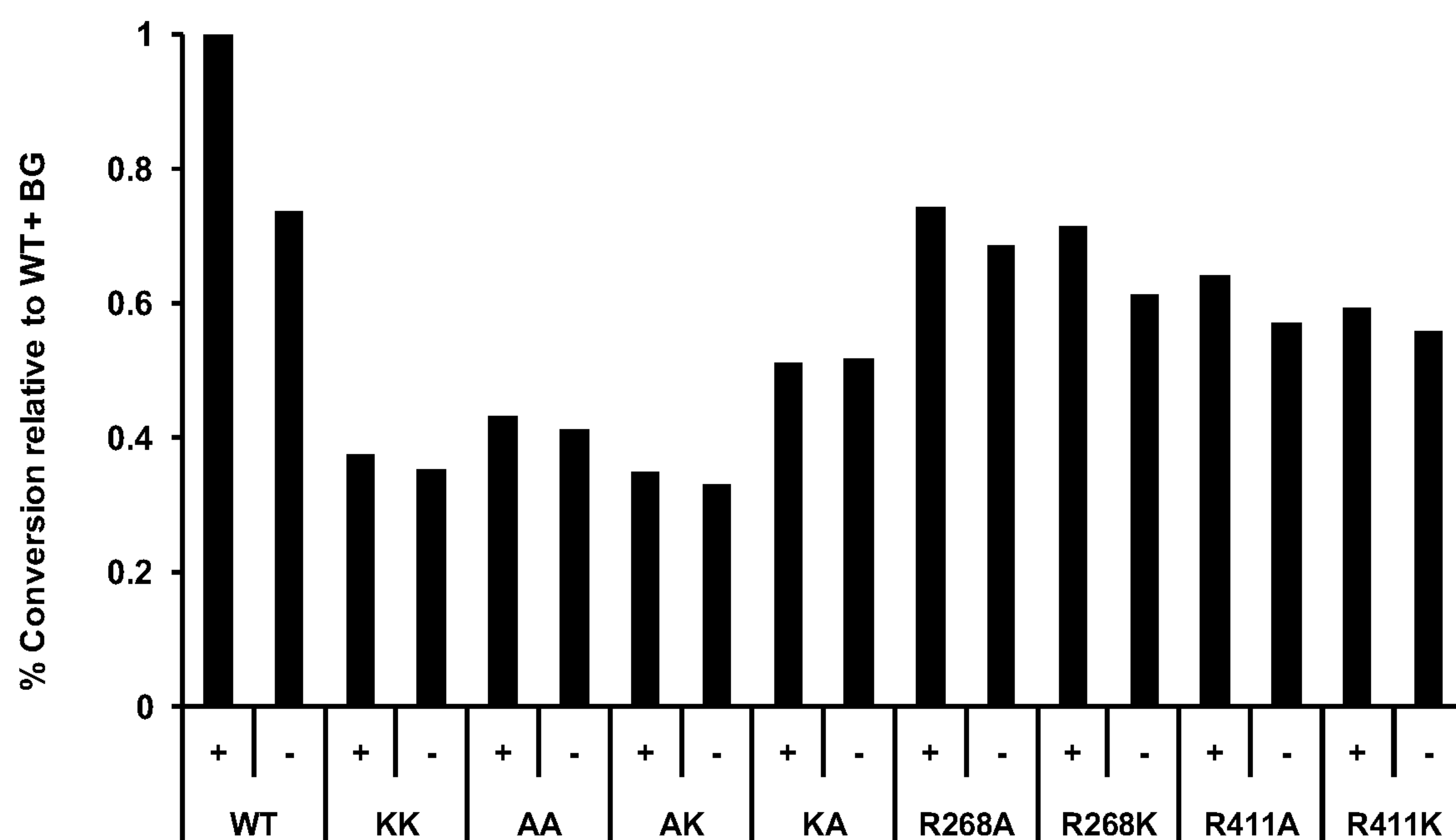


FIGURE 4

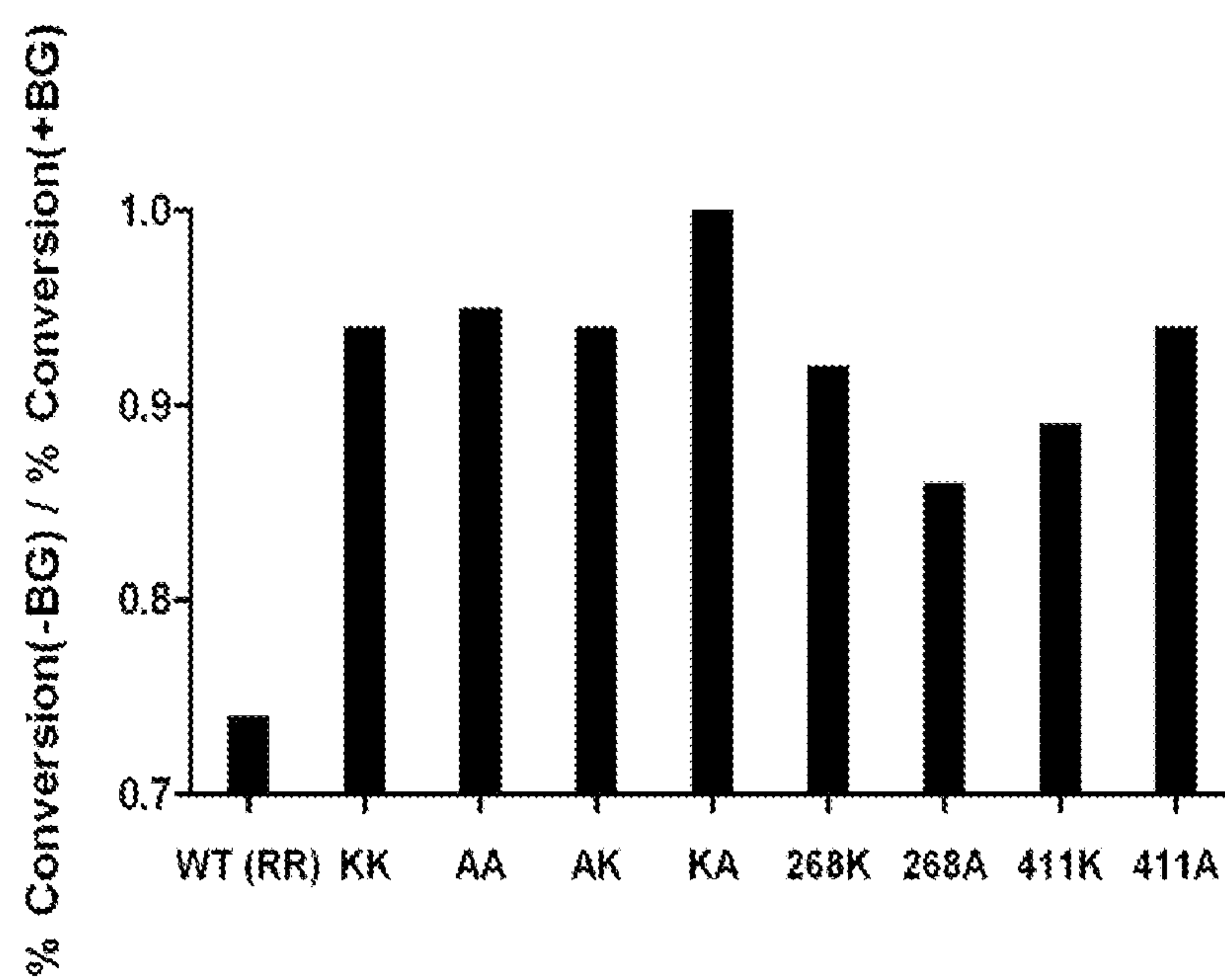


FIGURE 5

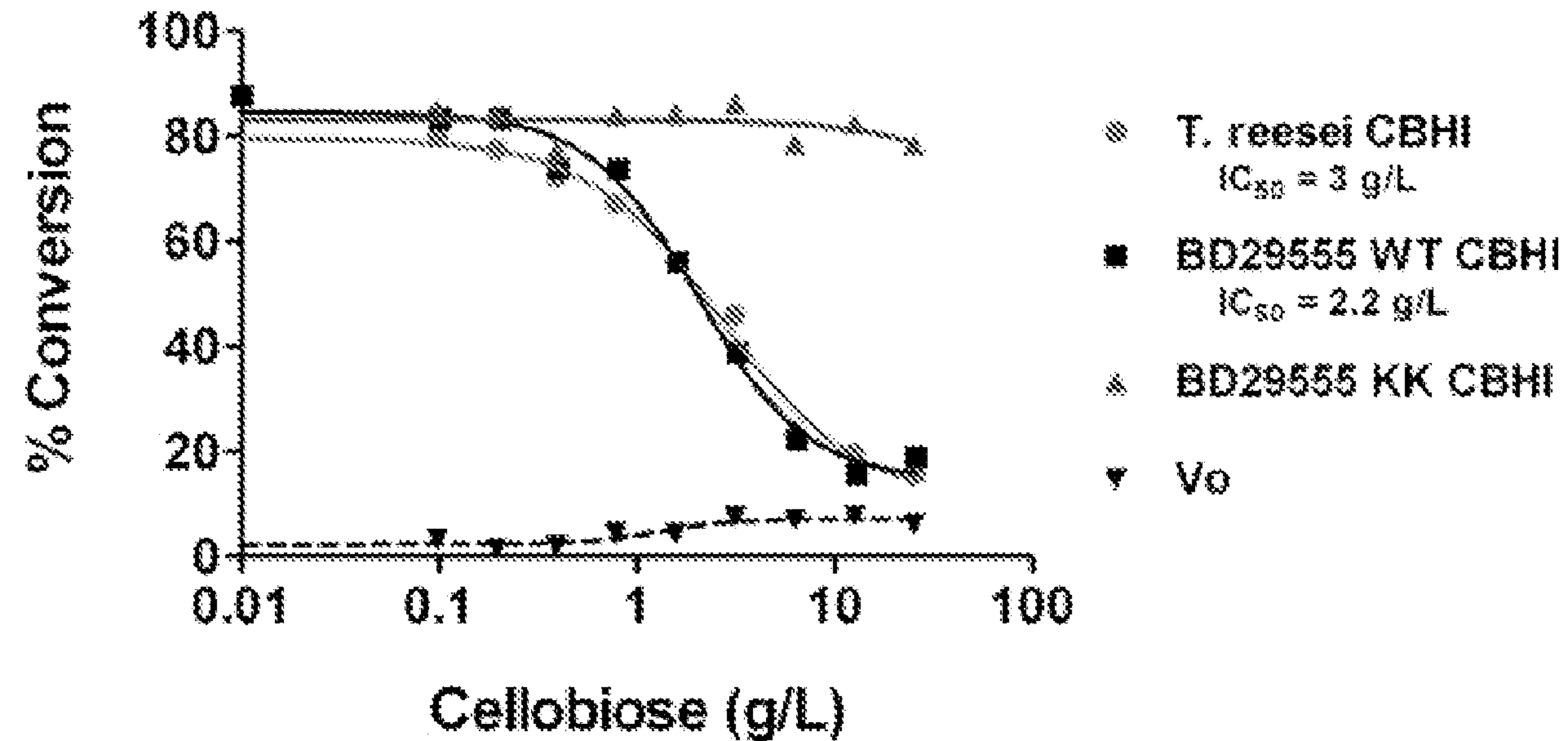


FIGURE 3