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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 June 2003 (26.06.2003)

PCT

(10) International Publication Number
WO 03/052057 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/34882

(22) International Filing Date: 30 October 2002 (30.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/026,994 18 December 2001 (18.12.2001) US

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

(71) Applicants and
(72) Inventors: **DUNN-COLEMAN, Nigel** [US/US]; 142 Johnson Avenue, Los Gatos, CA 95032 (US). **WARD, Michael** [GB/US]; 4372 24th Street, San Francisco, CA 94114 (US). **YAO, Jian** [US/US]; 1761 Wright Avenue, Sunnyvale, CA 94087 (US). **GOEDEGEBUUR, Frits** [NL/NL]; Rozenlaan 128, NL-3135 XV Vlaardingen (NL).

(74) Agent: **BOYD, Victoria, L.**; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** EGVI ENDOGLUCANASE AND NUCLEIC ACIDS ENCODING THE SAME

CCACGCGTCCGAGCAGTGTCCTCCTCCTCACTGCTTCGTCATGAAGGTCTCTCGAGTCCTTGCCCTTGTCCTGGGGGCCGT
CATCCCTGCCCATGCTGCCTTTTCATGGAAGAACGTCAAGCTCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCA
TCCCAAGACAAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGC
CGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCA
AAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGG
CGCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGCTGT
CGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGAC
CTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAA
GCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGGCACGGC
TGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATA
CTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGG
CACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAGATCTATA
CTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGA
TGCTCAGCTGTTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGACCTA
TTACTACAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGG
TCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGG
AATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGACGGCAT
CGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTT
CACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGT
CGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGA
CGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGA
CACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC
CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAA
GGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCAC
CGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCAC
CGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCC

(57) **Abstract:** The present invention provides a novel endoglucanase nucleic acid sequence, designated egl6, and the corresponding EGVI amino acid sequence. The invention also provides expression vectors and host cells comprising a nucleic acid sequence encoding EGVI, recombinant EGVI proteins and methods for producing the same.

WO 03/052057 A2

EGVI Endoglucanase and Nucleic Acids Encoding the Same

Government Support

Portions of this work were funded by Subcontract No. ZCO-30017-01 with the National Renewable Energy Laboratory under Prime Contract No. DE-AC36-99GO10337 with the U.S. Department of Energy. Accordingly, the United States Government may have certain rights in this invention.

Field of the Invention

The present invention relates to isolated *egl6* nucleic acid sequences which encode polypeptides having endoglucanase activity. The invention also relates to nucleic acid constructs, vectors, and host cells comprising the nucleic acid sequences as well as methods for producing recombinant EGVI polypeptides.

References

Altschul, S. F., *et al.*, J. Mol. Biol. 215:403-410, 1990.

Altschul, S. F., *et al.*, Nucleic Acids Res. 25:3389-3402, 1997.

Aro, N., *et al.*, J. Biol. Chem., 10.1074/ M003624200, April 13, 2001.

Aubert, *et al.*, Ed., p11 et seq., Academic Press, 1988.

Ausubel G. M., *et al.* CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y., 1993.

Baldwin, D., *et al.*, Curr. Opin. Plant Biol. 2(2):96-103, 1999.

Baulcombe, D., Arch. Virol. Suppl. 15:189-201, 1999.

Bhikhabhai, R. *et al.*, J. Appl. Biochem. 6:336, 1984.

Brumbauer, A. *et al.*, Bioseparation 7:287-295, 1999.

Carter *et al.*, Nucl. Acids Res. 13:4331, 1986.

Chen *et al.*, Biochem. Biophys. Acta. 1121:54-60, 1992.

Coligan, J. E. *et al.*, eds., CURRENT PROTOCOLS IN IMMUNOLOGY, 1991.

Collen, A., *et al.*, Journal of Chromatography A 910:275-284, 2001.

Coughlan, *et al.*, BIOCHEMISTRY AND GENETICS OF CELLULOSE DEGRADATION.

Cummings and Fowler, Curr. Genet. 29:227-233, 1996.

Dayhoff *et al.* in Atlas of Protein Sequence and Structure, Volume 5, Supplement 3, Chapter 22, pp. 345-352, 1978.

Deutscher, M.P., Methods Enzymol. 182:779-80, 1990.

Doolittle, R. F., OF URFS AND ORFS, University Science Books, CA, 1986.

Ellouz, S. *et al.*, J. Chromatography 396:307, 1987.

Fields and Song, *Nature* 340:245-246, 1989.

Filho, *et al.* Can. J. Microbiol. 42:1-5, 1996.

Fliess, A., *et al.*, Eur. J. Appl. Microbiol. Biotechnol. 17:314, 1983.

Freer, *et al.* J. Biol. Chem. 268:9337-9342, 1993.

Freshney, R. I., ed., ANIMAL CELL CULTURE, 1987.

Goyal, A. *et al.* Bioresource Technol. 36:37, 1991.

Halldorsdottir, S *et al.*, Appl Microbiol Biotechnol. 49(3):277-84, 1998.

Hu *et al.*, Mol Cell Biol. 11:5792-9, 1991.

Hemmpel, W.H. ITB Dyeing/Printing/Finishing 3:5-14, 1991.

Herr *et al.*, Appl. Microbiol. Biotechnol. 5:29-36, 1978.

Jakobovits, A, *et al.*, Ann N Y Acad Sci 764:525-35, 1995.

Jakobovits, A, Curr Opin Biotechnol 6(5):561-6, 1995.

Jones *et al.*, Nature 321:522-525, 1986.

Kawaguchi, T *et al.*, Gene 173(2):287-8, 1996.

Knowles, J. *et al.*, TIBTECH 5, 255-261, 1987.

Kohler and Milstein, Nature 256:495, 1975.

Krishna, S. *et al.*, Bioresource Tech. 77:193-196, 2001.

Kumar, A., *et al.*, Textile Chemist and Colorist 29:37-42, 1997.

Lehtio, J. *et al.*, FEMS Microbiology Letters 195:197-204, 2001.

Li and Ljungdahl Appl. Environ. Microbiol. 62:209-213, 1996.

Linder, M. and Teeri, T.T., Biotechnol. 57:15-28, 1997.

Medve, J. *et al.*, J. Chromatography A 808:153, 1998.

Ohmiya *et al.*, Biotechnol. Gen. Engineer. Rev. 14:365-414, 1997.

Ooi *et al.*, Nucleic Acids Res. 18(19):5884, 1990.

Ortega *et al.*, International Biodeterioration and Biodegradation 47:7-14, 2001.

Penttila *et al.*, Yeast 3:175-185, 1987.

Penttila *et al.*, Gene 63: 103-112, 1988.

Pere, J., *et al.*, In Proc. Tappi Pulping Conf., Nashville, TN, 27-31, pp. 693-696, 1996.

Riechmann *et al.*, Nature 332:323-327, 1988.

Rothstein *et al.*, Gene 55:353-356, 1987.

Saarilahti *et al.*, Gene 90:9-14, 1990.

Sakamoto *et al.*, Curr. Genet. 27:435-439, 1995.

Saloheimo M, *et al.*, Gene 63:11-22, 1988.

Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989.

Schulein, Methods Enzymol., 160, 25, pages 234 et seq, 1988.

Scopes, Methods Enzymol. 90 Pt E:479-90, 1982.

Spilliaert R, *et al.*, Eur J Biochem. 224(3):923-30, 1994.

Stahlberg, J. *et al.*, Bio/Technol. 9:286-290, 1991.

Strathern *et al.*, eds. (1981) The Molecular Biology of the Yeast Saccharomyces.

Suurnakki, A. *et al.*, Cellulose 7:189-209, 2000.

Te'o, J. *et al.*, FEMS Microbiology Letters 190:13-19, 2000.

Tilbeurgh, H. *et al.*, FEBS Lett. 16:215, 1984.

Timberlake *et al.*, *Cell* 1:29-37, 1981.

Tomaz, C. and Queiroz, J., J. Chromatography A 865:123-128, 1999.

Tomme, P. *et al.*, Eur. J. Biochem. 170:575-581, 1988.

Tormo, J. *et al.*, EMBO J. 15:5739-5751, 1996.

Tyndall, R.M., Textile Chemist and Colorist 24:23-26, 1992.

Van Rensburg *et al.*, Yeast 14:67-76, 1998.

Van Tilbeurgh, H. *et al.*, FEBS Lett. 204:223-227, 1986.

Verhoeyen *et al.*, Science 239:1534-1536, 1988.

Warrington, *et al.*, *Genomics* 13:803-808, 1992.

Wells *et al.*, Gene 34:315, 1985.

Wells *et al.*, Philos. Trans. R. Soc. London SerA 317:415, 1986.

Wood, Biochem. Soc. Trans., 13, pp. 407-410, 1985.

Wood *et al.*, METHODS IN ENZYMOLOGY, 160, 25, p. 87 et seq., Academic Press, New York, 1988.

Zoller *et al.*, Nucl. Acids Res. 10:6487, 1987.

Background of the Invention

Cellulose and hemicellulose are the most abundant plant materials produced by photosynthesis. They can be degraded and used as an energy source by numerous microorganisms, including bacteria, yeast and fungi, that produce extracellular enzymes capable of hydrolysis of the polymeric substrates to monomeric sugars (Aro *et al.*, 2001).

As the limits of non-renewable resources approach, the potential of cellulose to become a major renewable energy resource is enormous (Krishna *et al.*, 2001). The effective utilization of cellulose through biological processes is one approach to overcoming the shortage of foods, feeds, and fuels (Ohmiya *et al.*, 1997).

Cellulases are enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellooligosaccharides, and the like. Cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases ([beta] -D-glucoside glucohydrolase; EC 3.2.1.21) ("BG"). (Knowles *et al.*, 1987; Shulein, 1988). Endoglucanases act mainly on the amorphous parts of the cellulose fibre, whereas cellobiohydrolases are also able to degrade crystalline cellulose (Nevalainen and Penttila, 1995). Thus, the presence of a cellobiohydrolase in a cellulase system is required for efficient solubilization of crystalline cellulose (Suurnakki, *et al.* 2000). Beta-glucosidase acts to liberate D-glucose units from cellobiose, cello-oligosaccharides, and other glucosides (Freer, 1993).

Cellulases are known to be produced by a large number of bacteria, yeast and fungi. Certain fungi produce a complete cellulase system capable of degrading crystalline forms of cellulose, such that the cellulases are readily produced in large quantities via fermentation. Filamentous fungi play a special role since many yeast, such as *Saccharomyces cerevisiae*, lack the ability to hydrolyze cellulose. See, *e.g.*, Aro *et al.*, 2001; Aubert *et al.*, 1988; Wood *et al.*, 1988, and Coughlan, *et al.*.

The fungal cellulase classifications of CBH, EG and BG can be further expanded to include multiple components within each classification. For example, multiple CBHs, EGs and BGs have been isolated from a variety of fungal sources including *Trichoderma* reesei which contains known genes for 2 CBHs, *i.e.*, CBH I and CBH II, at least 5 EGs, *i.e.*, EG I, EG II , EG III, EGIV and EGV, and at least 2 BGs, i.e., BG1 and BG2.

In order to efficiently convert crystalline cellulose to glucose the complete cellulase system comprising components from each of the CBH, EG and BG classifications is required, with isolated components less effective in hydrolyzing crystalline cellulose (Filho *et al.*, 1996). A synergistic relationship has been observed between cellulase components from different classifications. In particular, the EG-type cellulases and CBH- type cellulases synergistically interact to more efficiently degrade cellulose. See, *e.g.*, Wood, 1985.

Cellulases are known in the art to be useful in the treatment of textiles for the purposes of enhancing the cleaning ability of detergent compositions, for use as a

softening agent, for improving the feel and appearance of cotton fabrics, and the like (Kumar *et al.*, 1997).

Cellulase-containing detergent compositions with improved cleaning performance (US Pat. No. 4,435,307; GB App. Nos. 2,095,275 and 2,094,826) and for use in the treatment of fabric to improve the feel and appearance of the textile (US Pat. Nos. 5,648,263, 5,691,178, and 5,776,757; GB App. No. 1,358,599; The Shizuoka Prefectural Hamamatsu Textile Industrial Research Institute Report, Vol. 24, pp. 54-61, 1986), have been described.

Hence, cellulases produced in fungi and bacteria have received significant attention. In particular, fermentation of *Trichoderma spp.* (*e.g.*, *Trichoderma longibrachiatum* or *Trichoderma reesei*) has been shown to produce a complete cellulase system capable of degrading crystalline forms of cellulose. U.S. Pat. No. 5,475,101 discloses the purification and molecular cloning of one particularly useful enzyme designated EGIII which is derived from *Trichoderma longibrachiatum*.

Although cellulase compositions have been previously described, there remains a need for new and improved cellulase compositions for use in household detergents, stonewashing compositions or laundry detergents, etc. Cellulases that exhibit resistance to surfactants (*e.g.*, linear alkyl sulfonates, LAS), improved performance under conditions of thermal stress, increased or decreased cellulolytic capacity, and/or high level expression in vitro, are of particular interest.

Summary of the Invention

The invention provides an isolated cellulase protein, identified herein as EGVI, and nucleic acids which encode EGVI.

In one aspect, EGVI polypeptides or proteins comprise a sequence having at least 80%, 85%, 90%, 95%, 98% or more sequence identity to the sequence presented as SEQ ID NO:2.

In a related aspect, the invention includes (i) fragments of EGVI, preferably at least about 20-100 amino acids in length, more preferably about 100-200 amino acids in length, and (ii) a pharmaceutical composition comprising EGVI. In various embodiments, the fragment corresponds to the N-terminal domain of EGVI or the C-terminal domain of EGVI.

In another aspect the invention includes an isolated polynucleotide having a sequence which encodes EGVI, a sequence complementary to the *egl6* coding sequence, and a composition comprising the polynucleotide. The polynucleotide may be mRNA, DNA, cDNA, genomic DNA, or an antisense analog thereof.

An *egl6* polynucleotide may comprise an isolated nucleic acid molecule which hybridizes to the complement of the nucleic acid presented as SEQ ID NO: 1 under moderate to high stringency conditions, where the nucleic acid molecule encodes an EGVI polypeptide that exhibits endoglucanase activity.

The polynucleotide may encode an EGVI protein having at least 80%, 85%, 90%, 95%, 98% or more sequence identity to the sequence presented as SEQ ID NO:1. In a specific embodiment, the polynucleotide comprises a sequence substantially identical to SEQ ID NO:1. The invention also contemplates fragments of the polynucleotide, preferably at least about 15-30 nucleotides in length.

The invention further provides recombinant expression vectors containing a nucleic acid sequence encoding EGVI or a fragment or splice variant thereof, operably linked to regulatory elements effective for expression of the protein in a selected host. In a related aspect, the invention includes a host cell containing the vector.

The invention further includes a method for producing EGVI by recombinant techniques, by culturing recombinant prokaryotic or eukaryotic host cells comprising nucleic acid sequence encoding EGVI under conditions effective to promote expression of the protein, and subsequent recovery of the protein from the host cell or the cell culture medium.

In yet another aspect, the invention includes an antibody specifically immunoreactive with EGVI.

Analytical methods for detecting *egl6* nucleic acids and EGVI proteins also form part of the invention.

In another aspect the invention provides for an enzymatic composition useful in the conversion of cellulose to sugars and/or ethanol. In a preferred embodiment the enzymatic composition comprises EGVI. The composition may further comprise additional cellulase enzymes such as other endoglucanases, beta-glucosidases and/or cellbiohydrolases. The composition may be enriched in EGVI.

Brief Description of the Figures

Figure 1 is a single stranded depiction of the nucleic acid sequence (SEQ ID NO:1), of the *T. reesei egl6* cDNA, wherein the non-coding sequence is indicated as bolded.

Figure 2 shows the predicted amino acid sequence (SEQ ID NO:2) and signal sequence (SEQ ID NO:3) based on the nucleotide sequence provided in Figure 1, wherein the signal sequence is indicated as bolded.

Detailed Description Of The Invention

I. Definitions.

Unless otherwise indicated, all technical and scientific terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook *et al.,* 1989, and Ausubel FM *et al.,* 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

All publications cited herein are expressly incorporated herein by reference for the purpose of describing and disclosing compositions and methodologies which might be used in connection with the invention.

The term "polypeptide" as used herein refers to a compound made up of a single chain of amino acid residues linked by peptide bonds. The term "protein" as used herein may be synonymous with the term "polypeptide" or may refer, in addition, to a complex of two or more polypeptides.

The term "nucleic acid molecule" includes RNA, DNA and cDNA molecules. It will be understood that, as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding a given protein such as EGVI may be produced. The present invention contemplates every possible variant nucleotide sequence, encoding EGVI, all of which are possible given the degeneracy of the genetic code.

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

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

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

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

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

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

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

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

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

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

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

conditions are used to identify sequences having about 80% or more sequence identity with the probe.

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

As used herein, "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention.

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

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

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

It follows that the term "EGVI expression" refers to transcription and translation of the *egl6* gene, the products of which include precursor RNA, mRNA, polypeptide, post-translationally processed polypeptides, and derivatives thereof, including EGVI from related species such as *Trichoderma longibrachiatum* (*reesei*), *Trichoderma viride*, *Trichoderma koningii, Hypocrea jecorina and Hypocrea schweinitzii.* By way of example, assays for EGVI expression include Western blot for EGVI protein, Northern blot analysis and reverse transcriptase polymerase chain reaction (RT-PCR) assays for EGVI mRNA, and

*en*doglucanase activity assays as described inShoemaker S.P. and Brown R.D.Jr. (Biochim. Biophys. Acta, 1978, 523:133-146) and Schulein (1988).

The term "alternative splicing" refers to the process whereby multiple polypeptide isoforms are generated from a single gene, and involves the splicing together of nonconsecutive exons during the processing of some, but not all, transcripts of the gene. Thus a particular exon may be connected to any one of several alternative exons to form messenger RNAs. The alternatively-spliced mRNAs produce polypeptides ("splice variants") in which some parts are common while other parts are different.

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

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

The term "filamentous fungi" means any and all filamentous fungi recognized by those of skill in the art. A preferred fungus is selected from the group consisting of Aspergillus, Trichoderma, Fusarium, Chrysosporium, Penicillium, Humicola, Neurospora, or alternative sexual forms thereof such as Emericella, Hypocrea.

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

The term "cellulase" refers to a category of enzymes capable of hydrolyzing cellulose polymers to shorter cello-oligosaccharide oligomers, cellobiose and/or glucose. Numerous examples of cellulases, such as exoglucanases, exocellobiohydrolases, endoglucanases, and glucosidases have been obtained from cellulolytic organisms, particularly including fungi, plants and bacteria.

The term "cellulose binding domain" as used herein refers to portion of the amino acid sequence of a cellulase or a region of the enzyme that is involved in the cellulose binding activity of a cellulase or derivative thereof. Cellulose binding domains generally function by non-covalently binding the cellulase to cellulose, a cellulose derivative or other polysaccharide equivalent thereof. Cellulose binding domains permit or facilitate hydrolysis of cellulose fibers by the structurally distinct catalytic core region, and typically function independent of the catalytic core. Thus, a cellulose binding domain will not possess the

significant hydrolytic activity attributable to a catalytic core. In other words, a cellulose binding domain is a structural element of the cellulase enzyme protein tertiary structure that is distinct from the structural element which possesses catalytic activity.

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

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

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

As used herein, the term "stonewashing composition" refers to a formulation for use in stonewashing cellulose containing fabrics. Stonewashing compositions are used to modify cellulose containing fabrics prior to sale, *i.e.*, during the manufacturing process. In contrast, detergent compositions are intended for the cleaning of soiled garments and are not used during the manufacturing process.

As used herein, the term "detergent composition" refers to a mixture which is intended for use in a wash medium for the laundering of soiled cellulose containing fabrics. In the context of the present invention, such compositions may include, in addition to cellulases and surfactants, additional hydrolytic enzymes, builders, bleaching agents, bleach activators, bluing agents and fluorescent dyes, caking inhibitors, masking agents, cellulase activators, antioxidants, and solubilizers.

As used herein, the term "decrease or elimination in expression of the *egl6* gene" means that either that the *egl6* gene has been deleted from the genome and therefore cannot be expressed by the recombinant host microorganism; or that the *egl6* gene has

been modified such that a functional EGVI enzyme is not produced by the recombinant host microorganism.

The term "altered *egl6*" or "altered *egl6* gene" means that the nucleic acid sequence of the gene has been altered by removing, adding, and/or manipulating the coding sequence or the amino acid sequence of the expressed protein has been modified.

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

As used herein, the terms "active" and "biologically active" refer to a biological activity associated with a particular protein, such as the enzymatic activity associated with a protease. It follows that the biological activity of a given protein refers to any biological activity typically attributed to that protein by those of skill in the art.

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

A wild type fungal cellulase composition is one produced by a naturally occurring fungal source and which comprises one or more BGL, CBH and EG components wherein each of these components is found at the ratio produced by the fungal source. Thus, an enriched EGVI composition would have EGVI at an altered ratio wherein the ratio of EGVI to other cellulase components (i.e., CBHs, beta-glucosidases and other endoglucanases) is elevated. This ratio may be increased by either increasing EGVI or decreasing (or eliminating) at least one other component by any means known in the art.

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

Fungal cellulases may contain more than one EG component. The different components generally have different isoelectric points which allow for their separation via

ion exchange chromatography and the like. Either a single EG component or a combination of EG components may be employed in an enzymatic solution.

When employed in enzymatic solutions, the EG component is generally added in an amount sufficient to allow the highest rate of release of soluble sugars from the biomass. The amount of EG component added depends upon the type of biomass to be saccharified, which can be readily determined by the skilled artisan. However, when employed, the weight percent of the EGVIII component relative to any CBH type components present in the cellulase composition is from preferably about 1, preferably about 5, preferably about 10, preferably about 15, or preferably about 20 weight percent to preferably about 25, preferably about 30, preferably about 35, preferably about 40, preferably about 45 or preferably about 50 weight percent. Furthermore, preferred ranges may be about 0.5 to about 15 weight percent, about 0.5 to about 20 weight percent, from about 1 to about 10 weight percent, from about 1 to about 15 weight percent, from about 1 to about 20 weight percent, from about 1 to about 25 weight percent, from about 5 to about 20 weight percent, from about 5 to about 25 weight percent, from about 5 to about 30 weight percent, from about 5 to about 35 weight percent, from about 5 to about 40 weight percent, from about 5 to about 45 weight percent, from about 5 to about 50 weight percent, from about 10 to about 20 weight percent, from about 10 to about 25 weight percent, from about 10 to about 30 weight percent, from about 10 to about 35 weight percent, from about 10 to about 40 weight percent, from about 10 to about 45 weight percent, from about 10 to about 50 weight percent, from about 15 to about 20 weight percent, from about 15 to about 25 weight percent, from about 15 to about 30 weight percent, from about 15 to about 35 weight percent, from about 15 to about 30 weight percent, from about 15 to about 45 weight percent, from about 15 to about 50 weight percent.

II. Target Organisms

A. Filamentous fungi

Filamentous fungi include all filamentous forms of the subdivision Eumycota and Oomycota. The filamentous fungi are characterized by vegetative mycelium having a cell wall composed of chitin, glucan, chitosan, mannan, and other complex polysaccharides, with vegetative growth by hyphal elongation and carbon catabolism that is obligately aerobic.

In the present invention, the filamentous fungal parent cell may be a cell of a species of, but not limited to, *Trichoderma, e.g., Trichoderma longibrachiatum (reesei), Trichoderma viride, Trichoderma koningii, Trichoderma harzianum; Penicillium sp.;*

Humicola sp., including *Humicola insolens; Chrysosporium sp.*, including C. lucknowense; *Gliocladium sp.; Aspergillus sp.*; *Fusarium sp., Neurospora sp., Hypocrea sp., and Emericella sp.* As used herein, the term "*Trichoderma*" or "*Trichoderma sp.*" refers to any fungal strains which have previously been classified as *Trichoderma* or are currently classified as *Trichoderma*.

In one preferred embodiment, the filamentous fungal parent cell is an *Aspergillus niger*, *Aspergillus awamori, Aspergillus aculeatus,* or *Aspergillus nidulans* cell.

In another preferred embodiment, the filamentous fungal parent cell is a *Trichoderma reesei* cell.

III. Cellulases

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. As set forth above, cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases (EC 3.2.1.21) ("BG"). (Knowles, *et al.*, 1987; Schulein, 1988).

Certain fungi produce complete cellulase systems which include exo-cellobiohydrolases or CBH-type cellulases, endoglucanases or EG-type cellulases and beta-glucosidases or BG-type cellulases (Schulein, 1988). 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. The different components, *i.e.*, the various endoglucanases and exocellobiohydrolases in a multi-component or complete cellulase system, generally have different properties, such as isoelectric point, molecular weight, degree of glycosylation, substrate specificity and enzymatic action patterns.

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

carbohydrate residues, such as cellobiose. This yields glucose as the sole product for the microorganism and reduces or eliminates cellobiose which inhibits cellobiohydrolases and endoglucanases.

Accordingly, β-glucosidase-type cellulases are considered to be an integral part of the cellulase system because they drive the overall reaction to glucose. Increased expression of BG in T. reesei has been shown to improve degradation of cellulose to glucose. See EP0562003, which is hereby incorporated by reference. In addition, β-glucosidases can catalyze the hydrolysis of a number of different substrates, and therefore they find utility in a variety of different applications. Some β-glucosidases can be added to grapes during wine making to enhance the potential aroma of the finished wine product. Yet another application can be to use β-glucosidase in fruit to enhance the aroma thereof. Alternatively, β-glucosidase can be used directly in food additives or wine processing to enhance the flavor and aroma.

Cellulases also find a number of uses in detergent compositions including to enhance cleaning ability, as a softening agent and to improve the feel of cotton fabrics (Hemmpel, 1991; Tyndall, 1992; Kumar *et al.*, 1997). While the mechanism is not part of the invention, softening and color restoration properties of cellulase have been attributed to the alkaline endoglucanase components in cellulase compositions, as exemplified by U.S. Patent Nos. 5,648,263, 5,691,178, and 5,776,757, which disclose that detergent compositions containing a cellulase composition enriched in a specified alkaline endoglucanase component impart color restoration and improved softening to treated garments as compared to cellulase compositions not enriched in such a component. In addition, the use of such alkaline endoglucanase components in detergent compositions has been shown to complement the pH requirements of the detergent composition (*e.g.*, by exhibiting maximal activity at an alkaline pH of 7.5 to 10, as described in U.S. Patent Nos. 5,648,263, 5,691,178, and 5,776,757).

Cellulase compositions have also been shown to degrade cotton-containing fabrics, resulting in reduced strength loss in the fabric (U.S. Patent No. 4,822,516), contributing to reluctance to use cellulase compositions in commercial detergent applications. Cellulase compositions comprising endoglucanase components have been suggested to exhibit reduced strength loss for cotton-containing fabrics as compared to compositions comprising a complete cellulase system.

Cellulases have also been shown to be useful in degradation of cellulase biomass to ethanol (wherein the cellulase degrades cellulose to glucose and yeast or other

microbes further ferment the glucose into ethanol), in the treatment of mechanical pulp (Pere *et al.*, 1996), for use as a feed additive (WO 91/04673) and in grain wet milling.

Most CBHs and EGs have a multidomain structure consisting of a core domain separated from a cellulose binding domain (CBD) by a linker peptide (Suurnakki *et al.*, 2000). The core domain contains the active site whereas the CBD interacts with cellulose by binding the enzyme to it (van Tilbeurgh *et al.*, 1986; Tomme *et al.*, 1988). The CBDs are particularly important in the hydrolysis of crystalline cellulose. It has been shown that the ability of cellobiohydrolases to degrade crystalline cellulose clearly decreases when the CBD is absent (Linder and Teeri, 1997). However, the exact role and action mechanism of CBDs is still a matter of speculation. It has been suggested that the CBD enhances the enzymatic activity merely by increasing the effective enzyme concentration at the surface of cellulose (Stahlberg *et al.*, 1991), and/or by loosening single cellulose chains from the cellulose surface (Tormo *et al.*, 1996). Most studies concerning the effects of cellulase domains on different substrates have been carried out with core proteins of cellobiohydrolases, as their core proteins can easily be produced by limited proteolysis with papain (Tomme *et al.*, 1988). Numerous cellulases have been described in the scientific literature, examples of which include: from *Trichoderma reesei*: Shoemaker, S. et al., Bio/Technology, 1:691-696, 1983, which discloses *CBHI*; Teeri, T. et al., Gene, 51:43-52, 1987, which discloses *CBHII*; Penttila, M. et al., Gene, 45:253-263, 1986, which discloses *EGI*; Saloheimo, M. et al., Gene, 63:11-22, 1988, which discloses *EGII*; Okada, M. et al., Appl. Environ. Microbiol., 64:555-563, 1988, which discloses *EGIII*; Saloheimo, M. et al., Eur. J. Biochem., 249:584-591, 1997, which discloses *EGIV*; Saloheimo, A. et al., Molecular Microbiology, 13:219-228, 1994, which discloses *EGV*; Barnett, C. C., et al., Bio/Technology, 9:562-567, 1991, which discloses *BGL1*, and Takashima, S. et al., J. Biochem., 125:728-736, 1999, which discloses *BGL2*. Cellulases from species other than Trichoderma have also been described e.g., Ooi *et al.*, 1990, which discloses the cDNA sequence coding for endoglucanase F1-CMC produced by *Aspergillus aculeatus*; Kawaguchi T *et al.*, 1996, which discloses the cloning and sequencing of the cDNA encoding beta-glucosidase 1 from *Aspergillus aculeatus*; Sakamoto *et al.*, 1995, which discloses the cDNA sequence encoding the endoglucanase CMCase-1 from *Aspergillus kawachii* IFO 4308; Saarilahti *et al.*, 1990 which discloses an endoglucanase from *Erwinia carotovara*; Spilliaert R, *et al.*, 1994, which discloses the cloning and sequencing of bglA, coding for a thermostable beta-glucanase from *Rhodothermus marinu*; and Halldorsdottir S *et al.*, 1998, which discloses the cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase

family 12. However, there remains a need for identification and characterization of novel cellulases, with improved properties, such as improved performance under conditions of thermal stress or in the presence of surfactants, increased specific activity, altered substrate cleavage pattern, and/or high level expression in vitro.

The development of new and improved cellulase compositions that comprise varying amounts CBH-type, EG-type and BG-type cellulases is of interest for use: (1) in 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"); (2) in compositions for degrading wood pulp or other biomass into sugars (*e.g.*, for bio-ethanol production); and/or (3) in feed compositions.

IV. Methods of Identifying Novel Sequences

Open reading frames (ORFs) are analyzed following full or partial sequencing of the *T. reesei* genome or of clones of cDNA libraries derived from T. reesei mRNA and are further analyzed using sequence analysis software, and by determining homology to known sequences in databases (public/private).

V. *egl6* Nucleic Acids And EGVI Polypeptides.

A. *egl6* Nucleic acids

The nucleic acid molecules of the present invention include the native coding sequence for *egl6*. In one embodiment the sequence is the cDNA sequence for *egl6* presented herein as SEQ. ID. NO:1 or SEQ. ID. NO:4, and homologues thereof in other species, naturally occurring allelic and splice variants, nucleic acid fragments, and biologically active (functional) derivatives thereof, such as, amino acid sequence variants of the native molecule and sequences which encode fusion proteins. The sequences are collectively referred to herein as "EGVI-encoding nucleic acid sequences".

A Basic BLASTN search (http://www.ncbi.nlm.nih.gov/BLAST) of the non-redundant nucleic acid sequence database was conducted on September 12, 2001, with the *egl6* gene sequence presented in Figure 1 (SEQ ID NO:1), indicated no sequences producing significant alignments (i.e. with an E value of less than 10^{-5}).

Part of the *egl6* sequence presented in Figure 1 (SEQ ID NO:1) is identical to parts of the sequences of *Trichoderma reesei* ESTs disclosed as SEQ ID NO:7511 and SEQ ID NO:7641 and annotated as a endoglucanases in patent application WO 0056762.

An *egl6* nucleic acid sequence of this invention may be a DNA or RNA sequence, derived from genomic DNA, cDNA, mRNA, or may be synthesized in whole or in part. The

DNA may be double-stranded or single-stranded and if single-stranded may be the coding strand or the non-coding (antisense, complementary) strand. The nucleic acid sequence may be cloned, for example, by isolating genomic DNA from an appropriate source, and amplifying and cloning the sequence of interest using a polymerase chain reaction (PCR). Alternatively, nucleic acid sequence may be synthesized, either completely or in part, especially where it is desirable to provide host-preferred sequences for optimal expression. Thus, all or a portion of the desired structural gene (that portion of the gene which encodes a polypeptide or protein) may be synthesized using codons preferred by a selected host.

Due to the inherent degeneracy of the genetic code, nucleic acid sequences other than the native form which encode substantially the same or a functionally equivalent amino acid sequence may be used to clone and/or express EGVI-encoding nucleic acid sequences. Thus, for a given EGVI-encoding nucleic acid sequence, it is appreciated that as a result of the degeneracy of the genetic code, a number of coding sequences can be produced that encode a protein having the same amino acid sequence. For example, the triplet CGT encodes the amino acid arginine. Arginine is alternatively encoded by CGA, CGC, CGG, AGA, and AGG. Therefore it is appreciated that such substitutions in the coding region fall within the nucleic acid sequence variants covered by the present invention. Any and all of these sequence variants can be utilized in the same way as described herein for the native form of an EGVI-encoding nucleic acid sequence.

A "variant" EGVI-encoding nucleic acid sequence may encode a "variant" EGVI amino acid sequence which is altered by one or more amino acids from the native polypeptide sequence or may be truncated by removal of one or more amino acids from either end of the polypeptide sequence, both of which are included within the scope of the invention. Similarly, the term "modified form of", relative to EGVI, means a derivative or variant form of the native EGVI protein-encoding nucleic acid sequence or the native EGVI amino acid sequence.

Similarly, the polynucleotides for use in practicing the invention include sequences, which encode native EGVI proteins and splice variants thereof, sequences complementary to the native protein coding sequence, and novel fragments of EGVI encoding polynucleotides. An EGVI encoding nucleic acid sequence may contain one or more intron sequences if it is a genomic DNA sequence.

In one general embodiment, an EGVI-encoding nucleotide sequence has at least 70%, preferably 80%, 85%, 90%, 95%, 98%, or more sequence identity to the *egl6* coding sequence presented herein as SEQ ID NO:1.

In another embodiment, an EGVI-encoding nucleotide sequence will hybridize under moderate to high stringency conditions to a nucleotide sequence that encodes an EGVI protein. In a related embodiment, an EGVI-encoding nucleotide sequence will hybridize under moderate to high stringency conditions to the nucleotide sequence presented as SEQ ID NO:1.

It is appreciated that some nucleic acid sequence variants that encode EGVI may or may not selectively hybridize to the parent sequence. By way of example, in situations where the coding sequence has been optimized based on the degeneracy of the genetic code, a variant coding sequence may be produced that encodes an EGVI protein, but does not hybridize to a native EGVI-encoding nucleic acid sequence under moderate to high stringency conditions. This would occur, for example, when the sequence variant includes a different codon for each of the amino acids encoded by the parent nucleotide.

As will be further understood by those of skill in the art, in some cases it may be advantageous to produce nucleotide sequences possessing non-naturally occurring codons e.g., inosine or other non-naturally occurring nucleotide analog. Codons preferred by a particular eukaryotic host can be selected, for example, to increase the rate of EGVI protein expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from the naturally occurring sequence. Hence, a native EGVI-encoding nucleotide sequence may be engineered in order to alter the coding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the EGVI protein by a cell.

Particularly preferred are nucleic acid substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the native polynucleotide or polypeptide.

The variations can be made using methods known in the art such as oligonucleotide-mediated (site-directed) mutagenesis, and PCR mutagenesis. Site-directed mutagenesis (Carter *et al.*, 1986; Zoller *et al.*, 1987), cassette mutagenesis (Wells *et al.*, 1985), restriction selection mutagenesis (Wells *et al.*, 1986) or other known techniques can be performed on the cloned DNA to produce the EGVI polypeptide-encoding variant DNA.

However, in some cases it may be advantageous to express variants of *egl6* which lack the properties or activities of the native *egl6* polynucleotide or EGVI polypeptide. In such cases, mutant or modified forms of the native EGVI-encoding nucleic acid sequence may be generated using techniques routinely employed by those of skill in the art.

B. EGVI Polypeptides

In one preferred embodiment, the invention provides an EGVI polypeptide, having a native mature or full-length EGVI polypeptide sequence comprising the sequence presented in Figure 2 (SEQ ID NO:2). An EGVI polypeptide of the invention can be the mature EGVI polypeptide, part of a fusion protein or a fragment or variant of the EGVI polypeptide sequence presented in Figure 2 (SEQ ID NO:2).

Ordinarily, an EGVI polypeptide of the invention has at least 80% identity to an EGVI amino acid sequence over its entire length. More preferable are EGVI polypeptide sequences that comprise a region having at least 80, 85, 90, 95, 98% or more sequence identity to the EGVI polypeptide sequence of Figure 2 (SEQ ID NO:2), using a sequence alignment program, as detailed herein.

Typically, a "modified form of" a native EGVI protein or a "variant" EGVI protein has a derivative sequence containing at least one amino acid substitution, addition, deletion or insertion, respectively.

It is well known in the art that certain amino acid substitutions may be made in protein sequences without affecting the function of the protein. Generally, conservative amino acid substitutions or substitutions of similar amino acids are tolerated without affecting protein function. Similar amino acids can be those that are similar in size and/or charge properties, for example, aspartate and glutamate, and isoleucine and valine, are both pairs of similar amino acids. Similarity between amino acid pairs has been assessed in the art in a number of ways. For example, Dayhoff *et al.* (1978), which is incorporated by reference herein provides frequency tables for amino acid substitutions which can be employed as a measure of amino acid similarity. Dayhoff *et al.*'s frequency tables are based on comparisons of amino acid sequences for proteins having the same function from a variety of evolutionarily different sources.

Fragments and variants of the EGVI polypeptide sequence of Figure 2 (SEQ ID NO:2), are considered to be a part of the invention. A fragment is a variant polypeptide which has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides. The fragments can be "free-standing" or comprised within a larger polypeptide of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically active fragments which are those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that are antigenic or immunogenic in an animal, particularly a human. In this aspect, the invention includes (i) fragments of EGVI,

preferably at least about 20-100 amino acids in length, more preferably about 100-200 amino acids in length, and (ii) a pharmaceutical composition comprising EGVI. In various embodiments, the fragment corresponds to the N-terminal domain of EGVI or the C-terminal domain of EGVI.

EGVI polypeptides of the invention also include polypeptides that vary from the EGVI polypeptide sequence of Figure 2 (SEQ ID NO:2). These variants may be substitutional, insertional or deletional variants. The variants typically exhibit the same qualitative biological activity as the naturally occurring analogue, although variants can also be selected which have modified characteristics as further described below.

A "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.

An "insertion" or "addition" is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring sequence.

A "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1 to 20 amino acids, although considerably larger insertions may be tolerated. Deletions range from about 1 to about 20 residues, although in some cases deletions may be much larger.

Substitutions, deletions, insertions or any combination thereof may be used to arrive at a final derivative. Generally these changes are done on a few amino acids to minimize the alteration of the molecule. However, larger changes may be tolerated in certain circumstances.

Amino acid substitutions can be the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of an isoleucine with a valine, *i.e.*, conservative amino acid replacements. Insertions or deletions may optionally be in the range of 1 to 5 amino acids.

Substitutions are generally made in accordance with known "conservative substitutions". A "conservative substitution" refers to the substitution of an amino acid in one class by an amino acid in the same class, where a class is defined by common physicochemical amino acid side chain properties and high substitution frequencies in homologous proteins found in nature (as determined, *e.g.*, by a standard Dayhoff frequency exchange matrix or BLOSUM matrix). (*See* generally, Doolittle, R.F., 1986.)

A "non-conservative substitution" refers to the substitution of an amino acid in one class with an amino acid from another class.

EGVI polypeptide variants typically exhibit the same qualitative biological activity as the naturally-occurring analogue, although variants also are selected to modify the characteristics of the EGVI polypeptide, as needed. For example, glycosylation sites, and more particularly one or more O-linked or N-linked glycosylation sites may be altered or removed. Those skilled in the art will appreciate that amino acid changes may alter post-translational processes of the EGVI polypeptide, such as changing the number or position of glycosylation sites or altering the membrane anchoring characteristics or secretion characteristics or other cellular localization characteristics.

Also included within the definition of EGVI polypeptides are other related EGVI polypeptides. Thus, probe or degenerate polymerase chain reaction (PCR) primer sequences may be used to find other related polypeptides. Useful probe or primer sequences may be designed to: all or part of the EGVI polypeptide sequence, or sequences outside the coding region. As is generally known in the art, preferred PCR primers are from about 15 to about 35 nucleotides in length, with from about 20 to about 30 being preferred, and may contain inosine as needed. The conditions for the PCR reaction are generally known in the art.

Covalent modifications of EGVI polypeptides are also included within the scope of this invention. For example, the invention provides EGVI polypeptides that are a mature protein and may comprise additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one polypeptide chain). Such sequences can, for example, play a role in the processing of the protein from a precursor to a mature form, allow protein transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production.

Also contemplated are modifications directed to alteration of an active site, alteration of the pH optima, temperature optima, and/or substrate affinity of the EGVI enzyme.

Figure 2 shows the predicted amino acid sequence (SEQ ID NO:2) of an exemplary EGVI polypeptide based on the nucleotide sequence provided in Figure 1.
The predicted molecular weight of the encoded EGVI polypeptide is 87.1kDa. A predicted signal peptide of 19 amino acids precedes the mature amino terminus of EGVI as provided in the figure suggesting that the EGVI polypeptide is secreted (Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., Protein Engineering, 10:1-6, 1997). The last 35 amino acids of

the sequence (SEQ ID NO:2) has up to 65% identity with the cellulose binding domains present on many fungal secreted cellulases and hemicellulases (Tomme, P., Warren, R.A., Miller, R.C., Jr., Kilburn, D.G. & Gilkes, N.R. (1995) in Enzymatic Degradation of Insoluble Polysaccharides (Saddler, J.N. & Penner, M., eds.), Cellulose-binding domains: classification and properties. pp. 142-163, American Chemical Society, Washington). The amino acids from approximately residue number 744 to approximately residue number 801 have the characteristics of a serine- and threonine-rich linker region which is generally found between the cellulose binding domain and the catalytic region of fungal enzymes having a cellulose binding domain.

A Basic BLASTP search (http://www.ncbi.nlm.nih.gov/BLAST) of the non-redundant protein database, conducted on September 12, 2001 with the EGVI amino acid sequence indicated 51% identity with GenBank Accession Number AB015511 (avicellase III of *Aspergillus aculeatus*), 49% sequence identity to GenBank Accession Number AJ292929 (CEL6 protein of *Agaricus bisporus*), 39% sequence identity to GenBank Accession Number AE007608 (probable secreted sialidase of *Clostridium acetabutylicum*), and 40% sequence identity to GenBank Accession Number AL031515 (probable secreted cellulase of *Streptomyces coelicolor*). These sequence similarities indicate that EGVI is a member of glycosyl hydrolase family 74 (Henrissat, B. and Bairoch, A. (1993) Biochem. J. 293:781-788).

C. Anti-EGVI Antibodies.

The present invention further provides anti-EGVI antibodies. The antibodies may be polyclonal, monoclonal, humanized, bispecific or heteroconjugate antibodies.

Methods of preparing polyclonal antibodies are known to the skilled artisan. The immunizing agent may be an EGVI polypeptide or a fusion protein thereof. It may be useful to conjugate the antigen to a protein known to be immunogenic in the mammal being immunized. The immunization protocol may be determined by one skilled in the art based on standard protocols or routine experimentation.

Alternatively, the anti-EGVI antibodies may be monoclonal antibodies. Monoclonal antibodies may be produced by cells immunized in an animal or using recombinant DNA methods. (See, *e.g.*, Kohler *et al.*, 1975; U.S. Patent No. 4,816,567).

An anti-EGVI antibody of the invention may further comprise a humanized or human antibody. The term "humanized antibody" refers to humanized forms of non-human (*e.g.*, murine) antibodies that are chimeric antibodies, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding partial sequences of antibodies)

which contain some portion of the sequence derived from non-human antibody. Methods for humanizing non-human antibodies are well known in the art, as further detailed in Jones *et al.*, 1986; Riechmann *et al.*, 1988; and Verhoeyen *et al.*, 1988. Methods for producing human antibodies are also known in the art. See, *e.g.*, Jakobovits, A, *et al.*, 1995 and Jakobovits, A, 1995.

VI. Expression Of Recombinant EGVI

The methods of the invention rely on the use cells to express EGVI, with no particular method of EGVI expression required.

The invention provides host cells which have been transduced, transformed or transfected with an expression vector comprising an EGVI-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the parental host cell prior to transduction, transformation or transfection and will be apparent to those skilled in the art.

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

A. Nucleic Acid Constructs/Expression Vectors.

Natural or synthetic polynucleotide fragments encoding EGVI ("EGVI-encoding nucleic acid sequences") may be incorporated into heterologous nucleic acid constructs or vectors, capable of introduction into, and replication in, a filamentous fungal or yeast cell. The vectors and methods disclosed herein are suitable for use in host cells for the expression of EGVI. Any vector may be used as long as it is replicable and viable in the cells into which it is introduced. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Cloning and expression vectors are also described in Sambrook *et al.,* 1989, Ausubel FM *et al.,* 1989, and Strathern *et al.*, 1981, each of which is expressly incorporated by reference herein. Appropriate expression vectors for fungi are described in van den Hondel, C.A.M.J.J. et al. (1991) In: Bennett, J.W. and Lasure, L.L. (eds.) More Gene Manipulations in Fungi. Academic Press, pp. 396-428. The appropriate DNA sequence may be inserted into a plasmid or vector (collectively referred to herein as "vectors") by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by standard

procedures. Such procedures and related sub-cloning procedures are deemed to be within the scope of knowledge of those skilled in the art.

Recombinant filamentous fungi comprising the coding sequence for EGVI may be produced by introducing a heterologous nucleic acid construct comprising the EGVI coding sequence into the cells of a selected strain of the filamentous fungi.

Once the desired form of an *egl6* nucleic acid sequence, homologue, variant or fragment thereof, is obtained, it may be modified in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence.

A selected *egl6* coding sequence may be inserted into a suitable vector according to well-known recombinant techniques and used to transform filamentous fungi capable of EGVI expression. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used to clone and express EGVI. Therefore it is appreciated that such substitutions in the coding region fall within the sequence variants covered by the present invention. Any and all of these sequence variants can be utilized in the same way as described herein for a parent EGVI-encoding nucleic acid sequence.

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

Heterologous nucleic acid constructs may include the coding sequence for *egl6*, or a variant, fragment or splice variant thereof: (i) in isolation; (ii) in combination with additional coding sequences; such as fusion protein or signal peptide coding sequences, where the *egl6* coding sequence is the dominant coding sequence; (iii) in combination with non-coding sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host; and/or (iv) in a vector or host environment in which the *egl6* coding sequence is a heterologous gene.

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

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

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

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

Such selectable markers confer to transformants the ability to utilize a metabolite that is usually not metabolized by the filamentous fungi. For example, the amdS gene from

T. reesei which encodes the enzyme acetamidase that allows transformant cells to grow on acetamide as a nitrogen source. The selectable marker (e.g. pyrG) may restore the ability of an auxotrophic mutant strain to grow on a selective minimal medium or the selectable marker (e.g. olic31) may confer to transformants the ability to grow in the presence of an inhibitory drug or antibiotic.

The selectable marker coding sequence is cloned into any suitable plasmid using methods generally employed in the art. Exemplary plasmids include pUC18, pBR322, and pUC100.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Sambrook *et al.*, 1989; Freshney, 1987; Ausubel, *et al.*, 1993; and Coligan *et al.*, 1991. All patents, patent applications, articles and publications mentioned herein, are hereby expressly incorporated herein by reference.

B. Host Cells and Culture Conditions For Enhanced EGVI Production

(i) Filamentous Fungi

Thus, the present invention provides filamentous fungi comprising cells which have been modified, selected and cultured in a manner effective to result in enhanced EGVI production or expression relative to the corresponding non-transformed parental fungi.

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

EGVI expressing cells are cultured under conditions typically employed to culture the parental fungal line. Generally, cells are cultured in a standard medium containing physiological salts and nutrients, such as described in Pourquie, J. et al., Biochemistry and Genetics of Cellulose Degradation, eds. Aubert, J. P. et al., Academic Press, pp. 71-86, 1988 and Ilmen, M. et al., Appl. Environ. Microbiol. 63:1298-1306, 1997. Culture conditions are also standard, *e.g.*, cultures are incubated at 28°C in shaker cultures or fermenters until desired levels of EGVI 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; "http://www.atcc.org/"). After fungal growth has been established, the cells are exposed to conditions effective to cause or permit the over expression of EGVI.

In cases where an EGVI 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 high-level EGVI expression.

(ii) Yeast

The present invention also contemplates the use of yeast as a host cell for EGVI production. Several other genes encoding hydrolytic enzymes have been expressed in various strains of the yeast *S. cerevisiae*. These include sequences encoding for two endoglucanases (Penttila *et al.*, 1987), two cellobiohydrolases (Penttila *et al.*, 1988) and one beta-glucosidase from *Trichoderma reesei* (Cummings and Fowler, 1996), a xylanase from *Aureobasidlium pullulans* (Li and Ljungdahl, 1996), an alpha-amylase from wheat (Rothstein *et al.*, 1987), etc. In addition, a cellulase gene cassette encoding the *Butyrivibrio fibrisolvens* endo- [beta] -1,4-glucanase (END1), *Phanerochaete chrysosporium* cellobiohydrolase (CBH1), the *Ruminococcus flavefaciens* cellodextrinase (CEL1) and the *Endomyces fibrilizer* cellobiase (Bgl1) was successfully expressed in a laboratory strain of *S. cerevisiae* (Van Rensburg *et al.*, 1998).

C. Introduction of an EGVI-Encoding Nucleic Acid Sequence into Host Cells.

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

Various methods may be employed for delivering an expression vector into cells *in vitro*. After a suitable vector is constructed, it is used to transform strains of fungi or yeast. General methods of introducing nucleic acids into cells for expression of heterologous nucleic acid sequences are known to the ordinarily skilled artisan. Such methods include, but not limited to, electroporation; nuclear microinjection or direct microinjection into single cells; bacterial protoplast fusion with intact cells; use of polycations, *e.g.*, polybrene or polyornithine; membrane fusion with liposomes, lipofectamine or lipofection-mediated transfection; high velocity bombardment with DNA-coated microprojectiles; incubation with

calcium phosphate-DNA precipitate; DEAE-Dextran mediated transfection; infection with modified viral nucleic acids; and the like.

Preferred methods for introducing a heterologous nucleic acid construct (expression vector) into filamentous fungi (*e.g.*, *T. reesei*) include, but are not limited to the use of a particle or gene gun, permeabilization of filamentous fungi cells walls prior to the transformation process (*e.g.*, by use of high concentrations of alkali, *e.g.*, 0.05 M to 0.4 M $CaC1_2$ or lithium acetate), protoplast fusion or agrobacterium mediated transformation. An exemplary method for transformation of filamentous fungi by treatment of protoplasts or spheroplasts with polyethylene glycol and $CaCl_2$ is described in Campbell, E.I. et al., Curr. Genet. 16:53-56, 1989 and Penttila, M. et al., Gene, 63:11-22, 1988.

In addition, heterologous nucleic acid constructs comprising an EGVI-encoding nucleic acid sequence can be transcribed *in vitro*, and the resulting RNA introduced into the host cell by well-known methods, *e.g.*, by injection.

Following introduction of a heterologous nucleic acid construct comprising the coding sequence for *egl6*, the genetically modified cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying expression of an EGVI-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the host cell selected for expression, and will be apparent to those skilled in the art.

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

The invention further includes novel and useful transformants of filamentous fungi such as *Trichoderma reesei* for use in producing fungal cellulase compositions. The invention includes transformants of filamentous fungi especially fungi comprising the *egl6* coding sequence, comprising a modified form of the *egl6* coding sequence or deletion of the *egl6* coding sequence.

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

VII. Analysis For EGVI Nucleic Acid Coding Sequences and/or Protein Expression.

In order to evaluate the expression of EGVI by a cell line that has been transformed with an EGVI-encoding nucleic acid construct, assays can be carried out at the protein level, the RNA level or by use of functional bioassays particular to *en*doglucanase activity and/or production.

In one exemplary application of the *egl6* nucleic acid and protein sequences described herein, a genetically modified strain of filamentous fungi, *e.g.*, *Trichoderma reesei*, is engineered to produce an increased amount of EGVI. Such genetically modified filamentous fungi would be useful to produce a cellulase product with greater increased cellulolytic capacity. In one approach, this is accomplished by introducing the coding sequence for *egl6* into a suitable host, *e.g.*, a filamentous fungi such as *Trichoderma reesei*.

Accordingly, the invention includes methods for expressing EGVI in a filamentous fungus or other suitable host by introducing an expression vector containing the DNA sequence encoding EGVI into cells of the filamentous fungus or other suitable host.

In another aspect, the invention includes methods for modifying the expression of EGVI in a filamentous fungus or other suitable host. Such modification includes a decrease or elimination in expression, or expression of an altered form of EGVI. An altered form of EGVI may have an altered amino acid sequence or an altered nucleic acid sequence.

In general, assays employed to analyze the expression of EGVI include, Northern blotting, dot blotting (DNA or RNA analysis), RT-PCR (reverse transcriptase polymerase chain reaction), or *in situ* hybridization, using an appropriately labeled probe (based on the nucleic acid coding sequence) and conventional Southern blotting and autoradiography.

In addition, the production and/or expression of EGVI may be measured in a sample directly, for example, by assays for *en*doglucanase activity, expression and/or production. Such assays are described, for example, in Shoemaker, S.P. and Brown, R.D.Jr. (Biochim. Biophys. Acta, 1978, 523:133-146; Schulein (1988) and U.S. Patent Nos. 5,246,853 and 5,475,101 each of which is expressly incorporated by reference herein. The ability of EGVI to hydrolyze isolated soluble and insoluble substrates can be measured using assays described in Suurnakki *et al.* (2000) and Ortega *et al.* (2001). Substrates useful for assaying cellobiohydrolase, endoglucanase or β-glucosidase activities include crystalline cellulose, filter paper, phosphoric acid swollen cellulose, hydroxyethyl cellulose,

carboxymethyl cellulose, cellooligosaccharides, methylumbelliferyl lactoside, methylumbelliferyl cellobioside, orthonitrophenyl lactoside, paranitrophenyl lactoside, orthonitrophenyl cellobioside, paranitrophenyl cellobioside, orthonitrophenyl glucoside, paranitrophenyl glucoside, methylumbelliferyl glycoside

In addition, protein expression, may be evaluated by immunological methods, such as immunohistochemical staining of cells, tissue sections or immunoassay of tissue culture medium, *e.g.*, by Western blot or ELISA. Such immunoassays can be used to qualitatively and quantitatively evaluate expression of EGVI. The details of such methods are known to those of skill in the art and many reagents for practicing such methods are commercially available.

A purified form of EGVI may be used to produce either monoclonal or polyclonal antibodies specific to the expressed protein for use in various immunoassays. (See, *e.g.*, Hu *et al.*, 1991). Exemplary assays include ELISA, competitive immunoassays, radioimmunoassays, Western blot, indirect immunofluorescent assays and the like. In general, commercially available antibodies and/or kits may be used for the quantitative immunoassay of the expression level of *en*doglucanase proteins.

VIII. Isolation And Purification Of Recombinant EGVI Protein.

In general, an EGVI protein produced in cell culture is secreted into the medium and may be purified or isolated, *e.g.*, by removing unwanted components from the cell culture medium. However, in some cases, an EGVI protein may be produced in a cellular form necessitating recovery from a cell lysate. In such cases the EGVI protein is purified from the cells in which it was produced using techniques routinely employed by those of skill in the art. Examples include, but are not limited to, affinity chromatography (Tilbeurgh *et al.*, 1984), ion-exchange chromatographic methods (Goyal *et al.*, 1991; Fliess *et al.*, 1983; Bhikhabhai *et al.*, 1984; Ellouz *et al.*, 1987), including ion-exchange using materials with high resolution power (Medve *et al.*, 1998), hydrophobic interaction chromatography (Tomaz and Queiroz, 1999), and two-phase partitioning (Brumbauer, *et al.*, 1999).

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

Once expression of a given EGVI protein is achieved, the EGVI protein thereby produced is purified from the cells or cell culture. Exemplary procedures suitable for such purification include the following: antibody-affinity column chromatography, ion exchange chromatography; ethanol precipitation; reverse phase HPLC; chromatography on silica or

on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; and gel filtration using, *e.g.*, Sephadex G-75. Various methods of protein purification may be employed and such methods are known in the art and described *e.g.* in Deutscher, 1990; Scopes, 1982. The purification step(s) selected will depend, *e.g.*, on the nature of the production process used and the particular protein produced.

IX. Utility Of *egl6* and EGVI

It can be appreciated that the *egl6* nucleotide, the EGVI protein and compositions comprising EGVI protein activity find utility in a wide variety applications, some of which are described below.

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

In one preferred approach, the cellulase of the invention finds utility in detergent compositions or in the treatment of fabrics to improve the feel and appearance.

Since the rate of hydrolysis of cellulosic products may be increased by using a transformant having at least one additional copy of the egl6 gene inserted into the genome, products that contain cellulose or heteroglycans can be degraded at a faster rate and to a greater extent. Products made from cellulose such as paper, cotton, cellulosic diapers and the like can be degraded more efficiently in a landfill. Thus, the fermentation product obtainable from the transformants or the transformants alone may be used in compositions to help degrade by liquefaction a variety of cellulose products that add to the overcrowded landfills.

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

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

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

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

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 occuring cellulase mixture produced by a microbe may not be the most efficient for rapid conversion of cellulose in biomass to glucose. It is known that endoglucanases act to produce new cellulose chain ends which themselves are substrates for the action of cellobiohydrolases and thereby improve the efficiency of hydrolysis of the entire cellulase system. Therefore, the use of increased or optimized endoglucanase activity would greatly enhance the production of ethanol.

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

In another embodiment the cellulosic feedstock may be pretreated. Pretreatment may be by elevated temperature and the addition of either of dilute acid, concentrated acid

or dilute alkali solution. The pretreatment solution is added for a time sufficient to at least partially hydrolyze the hemicellulose components and then neutralized.

In an alternative approach, a cellulase composition, which is deficient in or free of endoglucanase, is preferred. The deletion of the endoglucanase gene of this invention would be particularly useful in preparing cellulase compositions for use in detergents. Additionally, such compositions are useful for the production of cellooligosaccharides. The deletion of the egl6 gene from T. reesei strains would be particularly useful in preparing cellulase compositions for use in the detergents and in isolating cellooligosaccharides. The cellulase enzymes have been used in a variety of detergent compositions to enzymatically clean clothes. However, it is known in this art that use of cellulase enzymes can impart degradation of the cellulose fibers in clothes. One possibility to decrease the degradation effect is to produce a detergent that does not contain endoglucanase . Thus, the deletion of this protein would effect the cellulase system to inhibit the other components via accumulation of cellobiose. The modified microorganisms of this invention are particularly suitable for preparing such compositions because the egl6 gene can be deleted leaving the remaining CBH and EG components resulting in improved cleaning and softening benefits in the composition without degradative effects.

The detergent compositions of this invention may employ besides the cellulase composition (irrespective of the endoglucanase content, i.e., endoglucanase-free, substantially endoglucanase-free, or endoglucanase enhanced), a surfactant, including anionic, non-ionic and ampholytic surfactants, a hydrolase, building agents, bleaching agents, bluing agents and fluorescent dyes, caking inhibitors, solubilizers, cationic surfactants and the like. All of these components are known in the detergent art. The cellulase composition as described above can be added to the detergent composition either in a liquid diluent, in granules, in emulsions, in gels, in pastes, and the like. Such forms are well known to the skilled artisan. When a solid detergent composition is employed, the cellulase composition is preferably formulated as granules. Preferably, the granules can be formulated so as to contain a cellulase protecting agent. For a more thorough discussion, see US Patent Number 6,162,782 entitled "Detergent compositions containing cellulase compositions deficient in CBH I type components," which is incorporated herein by reference.

In yet another embodiment, the detergent compositions can also contain enhanced levels of endoglucanase or altered endoglucanase. In this regard, it really depends upon the type of product one desires to use in detergent compositions to give the appropriate effects.

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

Portions of the *egl6* nucleic acid sequence that are capable of binding to cellulose can be used to generate bacterial chimeric surface proteins, allowing whole-cell immobilization onto cellulose filters or other fibrous solid supports as described in Lehtio *et al.*, 2001.

In addition the *egl6* nucleic acid sequence finds utility in the identification and characterization of related nucleic acid sequences. A number of techniques useful for determining (predicting or confirming) the function of related genes or gene products include, but are not limited to, (A) DNA/RNA analysis, such as (1) overexpression, ectopic expression, and expression in other species; (2) gene knock-out (reverse genetics, targeted knock-out, viral induced gene silencing (VIGS, see Baulcombe, 1999); (3) analysis of the methylation status of the gene, especially flanking regulatory regions; and (4) in situ hybridization; (B) gene product analysis such as (1) recombinant protein expression; (2) antisera production, (3) immunolocalization; (4) biochemical assays for catalytic or other activity; (5) phosphorylation status; and (6) interaction with other proteins via yeast two-hybrid analysis; (C) pathway analysis, such as placing a gene or gene product within a particular biochemical or signaling pathway based on its overexpression phenotype or by sequence homology with related genes; and (D) other analyses which may also be performed to determine or confirm the participation of the isolated gene and its product in a particular metabolic or signaling pathway, and help determine gene function.

Endoglucanases and beta-glucosidases may be responsible for the production of disaccharides, such as sophorose, from cellooligosaccharides and glucose by transglycosylation reactions. Sophorose is known to be a very potent inducer of cellulase gene expression (Ilmen, M. et al., 1997, Appl. Environ. Microbiol. 63:1298-1306 and references therein). In this way EGs and BGLs may play an important role in the process of induction of cellulase gene expression. Over-expression of certain EGs or BGLs in a fungal strain may lead to higher overall cellulase productivity by that strain.

A. Homology To Known Sequences

The function of a related EGVI-encoding nucleic acid sequence may be determined by homology to known genes having a particular function. For example, a comparison of the coding sequence of an identified nucleic acid molecule to public nucleic acid sequence

databases is used to confirm function by homology to known genes or by extension of the identified nucleic acid sequence.

The term "% homology" is used interchangeably herein with the term "% identity" herein and refers to the level of nucleic acid or amino acid sequence identity between the nucleic acid sequence that encodes EGVI or the EGVI amino acid sequence, when aligned using a sequence alignment program.

For example, as used herein, 80% homology means the same thing as 80% sequence identity determined by a defined algorithm, and accordingly a homologue of a given sequence has greater than 80% sequence identity over a length of the given sequence. Exemplary levels of sequence identity include, but are not limited to, 80, 85, 90, 95, 98% or more sequence identity to a given sequence, *e.g.*, the coding sequence for *egl6*, as described herein.

Exemplary computer programs which can be used to determine identity between two sequences include, but are not limited to, the suite of BLAST programs, *e.g.*, BLASTN, BLASTX, and TBLASTX, BLASTP and TBLASTN, publicly available on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. See also, Altschul, *et al.*, 1990 and Altschul, *et al.*, 1997.

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

A preferred alignment of selected sequences in order to determine "% identity" between two or more sequences, is performed using for example, the CLUSTAL-W program in MacVector version 6.5, operated with default parameters, including an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM 30 similarity matrix.

In one exemplary approach, sequence extension of a nucleic acid encoding *egl6* may be carried out using conventional primer extension procedures as described in Sambrook *et al.*, *supra*, to detect *egl6* precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA and/or to identify ORFs that encode a full length protein.

In yet another aspect, the present invention includes the entire or partial nucleotide sequence of the nucleic acid sequence of *egl6* for use as a probe. Such a probe may be used to identify and clone out homologous nucleic acid sequences from related organisms.

Screening of a cDNA or genomic library with the selected probe may be conducted using standard procedures, such as described in Sambrook *et al.*, (1989). Hybridization conditions, including moderate stringency and high stringency, are provided in Sambrook *et al.*, *supra*.

The probes or portions thereof may also be employed in PCR techniques to generate a pool of sequences for identification of closely related *egl6* sequences. When *egl6* sequences are intended for use as probes, a particular portion of an EGVI encoding sequence, for example a highly conserved portion of the coding sequence may be used.

For example, an *egl6* nucleotide sequence may be used as a hybridization probe for a cDNA library to isolate genes, for example, those encoding naturally-occurring variants of EGVI from other fungal, bacterial or plant species, which have a desired level of sequence identity to the *egl6* nucleotide sequence disclosed in Figure 1 (SEQ ID NO:1). Exemplary probes have a length of about 20 to about 50 bases.

B. Two Hybrid Analysis

Proteins identified by the present invention can be used in the yeast two-hybrid system to "capture" protein binding proteins which are putative signal pathway proteins. The yeast two hybrid system is described in Fields and Song, *Nature* 340:245-246 (1989). Briefly, in a two-hybrid system, a fusion of a DNA-binding domain-*egl6* (*e.g.*, GAL4-*egl6* fusion) is constructed and transfected into yeast cells. The whole *egl6* gene, or subregions of the *egl6* gene, may be used. A second construct containing the library of potential binding partners fused to the DNA activation domain is co-transfected. Yeast co-transformants harboring proteins that bind to the EGVI protein are identified by, for example, beta-galactosidase or luciferase production (a screen), or survival on plates lacking an essential nutrient (a selection), as appropriate for the vectors used.

C. Microarray Analysis

In addition, microarray analysis, also known as expression profiling or transcript profiling, may be used to simultaneously evaluate the presence or expression of given DNA sequences, or changes in the expression of many different genes. In one approach, a large set of DNA sequences (probes), usually a broad set of expressed sequence tags, cDNAs, cDNA fragments, or sequence-specific oligonucleotides, is arrayed on a solid

support such as a glass slide or nylon membrane. Labelled target for hybridization to the probes is generated by isolating mRNA from control and induced tissue, then labeling each mRNA pool either directly or via a cDNA or cRNA intermediate, with a distinct marker, usually a fluorescent dye. The microarray is hybridized with the complex probes, and the relative hybridization signal intensity associated with each location on the array can be quantitated for each marker dye. Differences in expression between the control and induced states can be measured as a ratio of the signal from the two marker dyes. (See Baldwin, D *et al.*, 1999.)

Microarray analysis of the source organism from which *egl6* was derived may be carried out, to facilitate the understanding of gene function by identifying other genes that are coordinately regulated as a consequence of the overexpression of *egl6*. The identity of coordinately regulated genes may help to place the *egl6* gene in a particular pathway. Alternatively, such analysis may be used to identify other genes involved in the same pathway using microarray analysis.

All publications, patents and patent applications are herein expressly incorporated by reference in their entirety.

While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention.

EXAMPLE 1

In one exemplary approach, a cDNA fragment for use as a probe is isolated by extracting total RNA from mycelia of a *T. reesei* strain grown under conditions known to induce cellulase production and obtaining the polyadenylated (polyA) fraction therefrom. The polyA RNA is used to produce a cDNA pool which is then amplified using specific primers based on the *egl6* nucleic acid sequence provided herein.

Total RNA is isolated from the mycelia using methods known in the art, for example as described in Timberlake *et al.*, 1981; Maniatis, *et al.*, 1989; Ausubel, *et al.*, 1993 and Sambrook *et al.*, 1989, each of which is expressly incorporated by reference herein. Once isolated, Northern blots are performed to confirm cellulase expression and select an optimal induction time for cellulase expression and corresponding RNA isolation.

Messenger RNA (mRNA), having a poly (A) tail at the 3' end, may be purified from total RNA using methods known in the art.

The T. reesei RNA is used as template for RT-PCR using methods known in the art (Loftus, J. et al., Science, 249:915-918, 1990). During this procedure the mRNA is reverse

transcribed to produce first strand cDNA. The cDNA subsequently serves as template for PCR amplification of egl6 cDNA sequences using specific olionucleotide primers designed in accordance with SEQ ID No. 1 or SEQ ID No. 4.

Table 1. Sequences Provided in Support of the Invention.

Description	SEQ. ID NO.
full lengthT. reesei *egl6* cDNA nucleic acid sequence **CCACGCGTCCGAGCAGTGTCCTCCTCCTCACTGCTTCGTC**ATGAAG GTCTCTCGAGTCCTTGCCCTTGTCCTGGGGGCCGTCATCCCTGCCCATG CTGCCTTTTCATGGAAGAACGTCAAGCTCGGCGGCGGCGGCGGCTTCG TCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATATGCAC GAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGA CCGCCGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGG GCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATG CCGCAGTCGGCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCA TCATTCGCTCGTCAGACCGCGGCGCAACGTGGTCCTTCACCAACTTGCC CTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTC TGGCTGTCGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTC AGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGACCTTTTCCAA GGTCTCGTCGTTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGA TTCCAACGGCTACAACAGCGACAAGCAAGGACTCATGTGGGTTACGTT CGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGT TGGCACGGCTGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGC CGGCTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATACTTTCC TCACAAGGCGAAACTGCAGCCAGCAGAGAAGGCCTTGTATCTGACCTA TTCCGATGGCACAGGGCCGTATGATGGCACACTTGGCTCAGTGTGGAG GTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGG ATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAG CCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGATGCTC AGCTGTTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGG CGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTC CCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGT CACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCT CGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAAT GACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAA TGTGTCAATCCAATCACTGGCAGACGGCATCGAGGAATTCTCCGTCCA GGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGG AGACGACAACGGCTTCACCTTTGCCAGCAGAAACGACCTCGGGACATC GCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGT CGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACAC CGCCGGCACGCAACAGGTGGCCATCTCGTCCGACGGCGGCGCGACGTG GAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGG CCTATTCGGCCGACGGCGACACGATCCTCTGGTCGACCGCCTCGTCCG GCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCC TGCCCGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCT TCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAAGGACACCGGCA GCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGG GATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGTCTCGACC GACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAA	1

GTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGC TCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGG GCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATC CAGGGCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGC AGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGG CGTCTTTTACGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTC CTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAG CACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGT GACTTCGTCGAGGACCAGCTCGGCCGCCGGTCCCACGGGGTCAGGGGT CGCCGGTCATTATGCTCAGTGCGGAGGGATTGGGTGGACGGGGCCGAC GCAGTGTGTGGCGCCGTATGTCTGCCAGAAGCAGAATGATTATTACTA CCAGTGTGTGTGA**TGCTTGAACTGCCAAGCTCACGAGGAGAGCTAC** **ATACCCCTAGGCTCGCAGTAAAGAGCTCAAGCATCCGAAGAAGCA** **CTAGTAGTAGAGATCCAGTCAGATAATTATCCATTTGTTTGAATTA** **AATGATCTTCTATTGAAAAAAAAAAAAAA**	
T. reesei EGVI predicted amino acid sequence AFSWKNVKLGGGGGFVPGIIFHPKTKGVAYARTDIGGLYRLNADDSWTAVTD GIADNAGWHNWGIDAVALDPQDDQKVYAAVGMYTNSWDPSNGAIIRSSDRG ATWSFTNLPFKVGGNMPGRGAGERLAVDPANSNIIYFGARSGNGLWKSTDGG VTFSKVSSFTATGTYIPDPSDSNGYNSDKQGLMWVTFDSTSSTTGGATSRIFVG TADNITASVYVSTNAGSTWSAVPGQPGKYFPHKAKLQPAEKALYLTYSDGTG PYDGTLGSVWRYDIAGGTWKDITPVSGSDLYFGFGGLGLDLQKPGTLVVASL NSWWPDAQLFRSTDSGTTWSPIWAWASYPTETYYYSISTPKAPWIKNNFIDVT SESPSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGGHDLTNWDTRHNVS IQSLADGIEEFSVQDLASAPGGSELLAAVGDDNGFTFASRNDLGTSPQTVWAT PTWATSTSVDYAGNSVKSVVRVGNTAGTQVAISSDGGATWSIDYAADTSMN GGTVAYSADGDTILWSTASSGVQRSQFQGSFASVSSLPAGAVIASDKKTNSVF YAGSGSTFYVSKDTGSSFTRGPKLGSAGTIRDIAAHPTTAGTLYVSTDVGIFRS TDSGTTFGQVSTALTNTYQIALGVGSGSNWNLYAFGTGPSGARLYASGDSGA SWTDIQGSQGFGSIDSTKVAGSGSTAGQVYVGTNGRGVFYAQGTVGGGTGGT SSSTKQSSSSTSSASSSTTLRSSVVSTTRASTVTSSRTSSAAGPTGSGVAGHYAQ CGGIGWTGPTQCVAPYVCQKQNDYYYQCV	2
T. reesei EGVI protein predicted signal sequence: **MKVSRVLALVLGAVIPAHA**	3
T. reesei *egl6* nucleic acid coding sequence ATGAAGGTCTCTCGAGTCCTTGCCCTTGTCCTGGGGGCCGTCATCCCTGCC CATGCTGCCTTTTCATGGAAGAACGTCAAGCTCGGCGGCGGCGGCGGCTTC GTCCCCGGCATCATCTTCCATCCCAAGACAAAAGGCGTAGCATATGCACGA ACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGC CGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCG ACGCTGTTGCGCTTGATCCGCAGGACGATCAAAAGGTGTATGCCGCAGTCG GCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGT CAGACCGCGGCGCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGG GTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGCTGTCGATCCGGCC AACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGGAAG TCTACGGACGGCGGCGTGACCTTTTCCAAGGTCTCGTCGTTCACGGCAACT GGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAA GCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGG GAGCCACGTCTCGTATCTTTGTTGGCACGGCTGATAACATCACTGCTTCAG TCTATGTGAGCACGAATGCCGGCTCCACGTGGAGTGCTGTACCGGGGCAG CCAGGGAAATACTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAGGC CTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGGCACACTTGG CTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCC	4

CTGTCTCTGGATCAGATCTATACTTTGGCTTTGGCGGCCTTGGCCTCGATTT
GCAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGA
TGCTCAGCTGTTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTG
GGCGTGGGCGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCC
CAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCAC
CGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGA
TTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGGAATGACAATC
TTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATC
CAATCACTGGCAGACGGCATCGAGGAATTCTCCGTCCAGGACCTGGCCTCT
GCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTT
CACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGG
CAACGCCCACATGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACTCG
GTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGC
CATCTCGTCCGACGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACA
CGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGACACGATC
CTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGC
TTTGCCTCCGTCTCGAGCCTGCCCGCGGGCGCCGTCATCGCCTCGGACAAG
AAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGC
AAGGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGG
GACGATCCGGGATATCGCTGCTCACCCGACCACCGCGGGCACGTTGTATGT
CTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGG
CCAAGTCTCCACCGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGG
CTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCCGTCAGGGGC
TCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGG
GCTCCCAGGGCTTCGGCTCCATCGACAGCACCAAGGTCGCCGGCAGCGGC
AGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTA
CGCTCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCA
AGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTCGAGCACCACGCTGAGG
TCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACC
AGCTCGGCCGCCGGTCCCACGGGGTCAGGGGTCGCCGGTCATTATGCTCAG
TGCGGAGGGATTGGGTGGACGGGGCCGACGCAGTGTGTGGCGCCGTATGT
CTGCCAGAAGCAGAATGATTATTACTACCAGTGTGTGTGA

CLAIMS

1. An isolated polynucleotide derived from a fungal source, which polynucleotide comprises a nucleotide sequence encoding an enzyme having endoglucanase activity.

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

(a) a nucleic acid sequence which encodes or is complementary to a sequence which encodes an EGVI polypeptide having at least 85% sequence identity to the amino acid sequence presented in Figure 2 (SEQ ID NO:2);

(b) a nucleic acid sequence which encodes or is complementary to a sequence which encodes an EGVI polypeptide having at least 90% sequence identity to the amino acid sequence presented in Figure 2 (SEQ ID NO:2);

(c) a nucleic acid sequence which encodes or is complementary to a sequence which encodes an EGVI polypeptide having at least 95% sequence identity to the amino acid sequence presented in Figure 2;

(d) a nucleic acid sequence which encodes or is complementary to a sequence which encodes an EGVI polypeptide having the amino acid sequence presented in Figure 2;

(e) a nucleic acid sequence which encodes or is complementary to a sequence which encodes an EGVI polypeptide having at least 95% sequence identity to the amino acid sequence presented as SEQ ID NO:2;

(f) a nucleic acid sequence which encodes or is complementary to a sequence which encodes an EGVI polypeptide having the amino acid sequence presented as SEQ ID NO:2;

(g) a nucleic acid sequence presented as SEQ ID NO:4, or the complement thereof; and

(h) a nucleic acid sequence that hybridizes, under high stringency conditions to the sequence presented as SEQ ID NO:4, or the complement or a fragment thereof, wherein said isolated polynucleotide encodes a polypeptide having the biological activity of an endoglucanase.

3. The isolated polynucleotide of Claim 2, wherein % identity is calculated using the CLUSTAL-W program in MacVector version 6.5, operated with default parameters, including an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM 30 similarity matrix.

4. The isolated polynucleotide of Claim 2, wherein hybridization is conducted at 42°C in 50% formamide, 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/ml

denatured carrier DNA followed by washing two times in 2X SSPE and 0.5% SDS at room temperature and two additional times in 0.1 SSPE and 0.5% SDS at 42°C.

5. The isolated polynucleotide of Claim 2, wherein said polynucleotide is an RNA molecule.

6. The isolated polynucleotide encoding an enzyme having endoglucanase activity, wherein the enzyme is derived from a Trichoderma source.

7. The isolated polynucleotide of Claim 6, wherein the enzyme is derived from Trichoderma reesei.

8. An expression construct including a polynucleotide sequence (i) having at least 85% sequence identity to the amino acid sequence presented in Figure 2 (SEQ ID NO:2), or (ii) being capable of hybridizing to a probe derived from the nucleotide sequence disclosed in Figure 2 under conditions of intermediate to high stringency, or (iii) being complementary to a nucleotide sequence having at least 85% sequence identity to the amino acid sequence presented in Figure 2 (SEQ ID NO:2).

9. A vector including the expression construct of Claim 8.

10. A vector comprising an isolated polynucleotide of Claim 2, operably linked to control sequences recognized by a host cell transformed with the vector.

11. A host cell transformed with the vector of Claim 9.

12. A host cell transformed with the vector of Claim 10.

13. The host cell of Claim 12, which is a prokaryotic cell.

14. The host cell of Claim 12, which is a eukaryotic cell.

15. A recombinant host cell comprising a polynucleotide of Claim 2.

16. The recombinant host cell of Claim 15, which is a prokaryotic cell.

17. The recombinant host cell of Claim 15, which is a eukaryotic cell.

18. A substantially purified EGVI polypeptide with the biological activity of an endoglucanase, comprising a sequence selected from the group consisting of:

(a) an amino acid sequence having at least 85% sequence identity to the amino acid sequence presented in Figure 2 (SEQ ID NO:2);

(b) an amino acid sequence having at least 90% sequence identity to the amino acid sequence presented in Figure 2 (SEQ ID NO:2);

(c) an amino acid sequence having at least 95% sequence identity to the amino acid sequence presented in Figure 2;

(d) an amino acid sequence presented in Figure 2;

(e) an amino acid sequence having at least 95% sequence identity to the amino acid sequence presented as SEQ ID NO:2;

(f) an amino acid sequence presented as SEQ ID NO:2;

(g) a substantially purified biologically active fragment of the amino acid sequence presented as SEQ ID NO:2.

19. A method of producing an enzyme having endoglucanase activity, comprising:

(a) stably transforming a host cell with an expression vector comprising a polynucleotide as defined in Claim2;

(b) cultivating said transformed host cell under condition suitable for said host cell to produce said endoglucanase; and

(c) recovering said endoglucanase.

20. The method of Claim 19 wherein the host cell is a filamentous fungi or yeast cell.

21. A purified enzyme having endoglucanase activity prepared by the method of Claim 19.

22. A recombinant host cell comprising a deletion or insertion or other alteration in the egl6 gene which inactivates the gene and prevents EGVI polypeptide production.

23. An antisense oligonucleotide complementary to a messenger RNA that encodes an EGVI polypeptide having the sequence presented as SEQ ID NO:2, wherein upon exposure to a endoglucanase-producing host cell, said oligonucleotide decreases or inhibits the production of endoglucanase by said host cell.

24. The antisense oligonucleotide of Claim 23, wherein the host cell is a filamentous fungi.

25. A detergent composition, said composition comprising a polypeptide selected from the group consisting of:

(a) an amino acid sequence having at least 85% sequence identity to the amino acid sequence presented in Figure 2 (SEQ ID NO:2);

(b) an amino acid sequence having at least 90% sequence identity to the amino acid sequence presented in Figure 2 (SEQ ID NO:2);

(c) an amino acid sequence having at least 95% sequence identity to the amino acid sequence presented in Figure 2;

(d) an amino acid sequence presented in Figure 2;

(e) an amino acid sequence having at least 95% sequence identity to the amino acid sequence presented as SEQ ID NO:2;

(f) an amino acid sequence presented as SEQ ID NO:2;

(g) a substantially purified biologically active fragment of the amino acid sequence presented as SEQ ID NO:2.

26. A method of expressing a heterologous polypeptide having β-glucosidase activity in an Aspergillus species, comprising:

(a) Providing a host Aspergillus with an expression vector comprising a polynucleotide encoding a signal sequence linked to a polynucleotide encoding a heterologous β-glucosidase, thereby encoding a chimeric polypeptide;

(b) Cultivating said host Aspergillus under conditions suitable for said Aspergillus to produce said chimeric polypeptide, wherein said chimeric polypeptide is produced.

27. A method of producing ethanol, said method comprising the steps of:

a) contacting a biomass composition with an enzymatic composition comprising β-glucosidase 4 to yield a sugar solution;

b) adding to the sugar solution a fermentative microorganism; and

c) culturing the fermentative microorganism under conditions sufficient to produce ethanol,

wherein the biomass composition may be optionally pretreated.

28. The method of claim 27 wherein step (a) further comprises the addition of at least one endoglucanase.

29. The method of claim 27 wherein step (a) further comprises the addition of at least one cellbiohydrolase.

30. The method of claim 28 wherein step (a) further comprises the addition of at least one cellbiohydrolase.

31. The method of claim 27 wherein the pretreatment is with a dilute acid.

32. A method of producing ethanol, said method comprising the steps of:

a) contacting a biomass composition with an enzymatic composition comprising a β-glucosidase 4 and a fermentative microorganism; and

b) culturing the fermentative microorganism under conditions sufficient to produce ethanol,

wherein the biomass composition may be optionally pretreated.

33. The method of claim 32 wherein step (a) further comprises the addition of at least one endoglucanase.

34. The method of claim 32 wherein step (a) further comprises the addition of at least one cellbiohydrolase.

35. The method of claim 33 wherein step (a) further comprises the addition of at least one cellbiohydrolase.

36. The method of claim 32 wherein the pretreatment is with a dilute acid.

Figure 1-A

T. reesei egl6 nucleotide sequence (coding and non-coding)

CCACGCGTCCGAGCAGTGTCCTCCTCCTCACTGCTTCGTCATGAAGGTCTCTCGAGTCCTTGCCCTTGTCCTGGGGGCCGT
CATCCCTGCCCATGCTGCCTTTTCATGGAAGAACGTCAAGCTCGGCGGCGGCGGCGGCTTCGTCCCCGGCATCATCTTCCA
TCCCAAGACAAAAGGCGTAGCATATGCACGAACAGATATTGGCGGGCTGTACCGCCTCAACGCCGACGACTCATGGACCGC
CGTCACGGATGGGATTGCTGATAATGCCGGCTGGCACAACTGGGGCATCGACGCTGTTGCGCTTGATCCGCAGGACGATCA
AAAGGTGTATGCCGCAGTCGGCATGTATACGAACAGCTGGGATCCGAGTAATGGAGCCATCATTCGCTCGTCAGACCGCGG
CGCAACGTGGTCCTTCACCAACTTGCCCTTCAAAGTCGGGGGTAACATGCCAGGACGCGGAGCCGGAGAGCGTCTGGCTGT
CGATCCGGCCAACTCCAACATCATCTACTTTGGTGCTCGCTCAGGAAACGGCCTCTGGAAGTCTACGGACGGCGGCGTGAC
CTTTTCCAAGGTCTCGTCGTTCACGGCAACTGGGACGTACATCCCAGACCCGAGTGATTCCAACGGCTACAACAGCGACAA
GCAAGGACTCATGTGGGTTACGTTCGACTCAACCAGCAGCACGACCGGGGGAGCCACGTCTCGTATCTTTGTTGGCACGGC
TGATAACATCACTGCTTCAGTCTATGTGAGCACGAATGCCGGCTCCACGTGGAGTGCTGTACCGGGGCAGCCAGGGAAATA
CTTTCCTCACAAGGCGAAACTGCAGCCAGCAGAGAAGGCCTTGTATCTGACCTATTCCGATGGCACAGGGCCGTATGATGG
CACACTTGGCTCAGTGTGGAGGTACGACATTGCAGGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAGATCTATA
CTTTGGCTTTGGCGGCCTTGGCCTCGATTTGCAAAAGCCAGGAACCCTTGTTGTTGCTTCTTTGAACTCTTGGTGGCCAGA
TGCTCAGCTGTTTCGGTCGACCGACTCTGGGACAACATGGAGCCCGATCTGGGCGTGGGCGAGCTATCCGACTGAGACCTA
TTACTACAGCATCTCAACTCCCAAAGCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACCGTCCGATGG
TCTCATCAAGCGCCTCGGCTGGATGATTGAGTCTCTCGAGATTGACCCAACCGACAGCAACCACTGGCTCTACGGCACCGG
AATGACAATCTTTGGCGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAATCACTGGCAGACGGCAT
CGAGGAATTCTCCGTCCAGGACCTGGCCTCTGCACCCGGCGGAAGCGAGCTATTGGCCGCAGTCGGAGACGACAACGGCTT
CACCTTTGCCAGCAGAAACGACCTCGGGACATCGCCGCAGACGGTCTGGGCAACGCCCACATGGGCCACCTCGACGAGCGT
CGACTACGCCGGGAACTCGGTCAAGAGCGTCGTCCGCGTCGGCAACACCGCCGGCACGCAACAGGTGGCCATCTCGTCCGA
CGGCGGCGCGACGTGGAGCATCGACTACGCGGCCGACACGTCCATGAACGGCGGCACGGTGGCCTATTCGGCCGACGGCGA
CACGATCCTCTGGTCGACCGCCTCGTCCGGCGTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCCTCCGTCTCGAGCCTGCC
CGCGGGCGCCGTCATCGCCTCGGACAAGAAGACCAACAGCGTCTTCTACGCCGGCTCCGGATCGACCTTTTACGTCAGCAA
GGACACCGGCAGCAGCTTCACGCGCGGGCCCAAGCTGGGCAGCGCAGGGACGATCCGGGATATCGCTGCTCACCCGACCAC
CGCGGGCACGTTGTATGTCTCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAGTCTCCAC
CGCCCTGACCAACACCTACCAGATCGCCCTGGGTGTGGGCTCAGGCTCGAACTGGAACCTGTATGCCTTCGGCACCGGCCC

Figure 1-B
T. reesei egl6 nucleotide sequence

GTCAGGGGCTCGCCTCTACGCCAGTGGAGACAGCGGCGCCTCCTGGACGGACATCCAGGGCTCCCAGGGCTTCGGCTCCAT
CGACAGCACCAAGGTCGCCGGCAGCGGCAGCACCGCCGGGCAAGTCTACGTGGGCACCAACGGCCGGGGCGTCTTTTACGC
TCAGGGAACCGTCGGCGGCGGCACGGGCGGGACTTCCTCGTCGACCAAGCAGAGCAGCAGCAGTACCTCTTCCGCCAGCTC
GAGCACCACGCTGAGGTCGAGCGTTGTATCCACGACCCGGGCTTCGACGGTGACTTCGTCGAGGACCAGCTCGGCCGCCGG
TCCCACGGGGTCAGGGGTCGCCGGTCATTATGCTCAGTGCGGAGGGATTGGGTGGACGGGGCCGACGCAGTGTGTGGCGCC
GTATGTCTGCCAGAAGCAGAATGATTATTACTACCAGTGTGTGTGA**TGCTTGAACTGCCAAGCTCACGAGGAGAGCTACAT**
ACCCCTAGGCTCGCAGTAAAGAGCTCAAGCATCCGAAGAAGCACTAGTAGTAGAGATCCAGTCAGATAATTATCCATTTGT
TTGAATTAAATGATCTTCTATTGAAAAAAAAAAAAA

Figure 2-A
Predicted *T. reesei egl6* amino acid sequence

MKVSRVLALV LGAVIPAHAA FSWKNVKLGG GGGFVPGIIF HPKTKGVAYA RTDIGGLYRL NADDSWTAVT
10 20 30 40 50 60 70

DGIADNAGWH NWGIDAVALD PQDDQKVYAA VGMYTNSWDP SNGAIIRSSD RGATWSFTNL PFKVGGNMPG
80 90 100 110 120 130 140

RGAGERLAVD PANSNIIYFG ARSGNGLWKS TDGGVTFSKV SSFTATGTYI PDPSDSNGYN SDKQGLMWVT
150 160 170 180 190 200 210

FDSTSSTTGG ATSRIFVGTA DNITASVYVS TNAGSTWSAV PGQPGKYFPH KAKLQPAEKA LYLTYSDGTG
220 230 240 250 260 270 280

PYDGTLGSVW RYDIAGGTWK DITPVSGSDL YFGFGGLGLD LQKPGTLVVA SLNSWWPDAQ LFRSTDSGTT
290 300 310 320 330 340 350

WSPIWAWASY PTETYYYSIS TPKAPWIKNN FIDVTSESPS DGLIKRLGWM IESLEIDPTD SNHWLYGTGM
360 370 380 390 400 410 420

TIFGGHDLTN WDTRHNVSIQ SLADGIEEFS VQDLASAPGG SELLAAVGDD NGFTFASRND LGTSPQTVWA
430 440 450 460 470 480 490

TPTWATSTSV DYAGNSVKSV VRVGNTAGTQ VAISSDGGAT WSIDYAADTS MNGGTVAYSA DGDTILWSTA
500 510 520 530 540 550 560

SSGVQRSQFQ GSFASVSSLP AGAVIASDKK TNSVFYAGSG STFYVSKDTG SSFTRGPKLG SAGTIRDIAA
570 580 590 600 610 620 630

Figure 2-B
Predicted *T. reesei egl6* amino acid sequence

```
HPTTAGTLYV STDVGIFRST DSGTTFGQVS TALTNTYQIA LGVGSGSNWN LYAFGTGPSG ARLYASGDSG
        640         650         660         670         680         690         700

ASWTDIQGSQ GFGSIDSTKV AGSGSTAGQV YVGTNGRGVF YAQGTVGGGT GGTSSSTKQS SSSTSSASSS
        710         720         730         740         750         760         770

TTLRSSVVST TRASTVTSSR TSSAAGPTGS GVAGHYAQCG GIGWTGPTQC VAPYVCQKQN DYYYQCV
        780         790         800         810         820         830      837
```