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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 October 2013 (03.10.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/148167 A2

(51) International Patent Classification:
C12P 19/04 (2006.01)

(21) International Application Number:
PCT/US2013/030571

(22) International Filing Date:
12 March 2013 (12.03.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/618,610 30 March 2012 (30.03.2012) US
61/660,556 15 June 2012 (15.06.2012) US

(71) Applicant: **VERENIUM CORPORATION** [US/US]; 3550 John Hopkins Ct., San Diego, CA 92121 (US).

(72) Inventors: **TAN, Xuqiu**; 3550 John Hopkins Ct., San Diego, CA 92121 (US). **BARRETT, Kenneth, E.**; 3550 John Hopkins Ct., San Diego, CA 92121 (US). **DAVENPORT, Adrienne, Huston**; 3550 John Hopkins Ct., San Diego, CA 92121 (US). **WHIPPLE, Lawrence, E.**; 3550 John Hopkins Ct., San Diego, CA 92121 (US). **URBINA, Hugo, D.**; 3550 John Hopkins Ct., San Diego, CA 92121 (US). **ZHANG, Bin**; 3550 John Hopkins Ct., San Diego, CA 92121 (US). **WALL, Mark, A.**; 3550 John Hopkins Ct., San Diego, CA 92121 (US).

(74) Agents: **SIDDONS, Brian W.** et al.; Verenium Corporation, 3550 John Hopkins Ct., San Diego, CA 92121 (US).

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

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

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

[Continued on next page]

(54) Title: GENES ENCODING CELLULASE FOR HYDROLYZING GUAR FRACTURING FLUIDS UNDER EXTREME WELL CONDITIONS

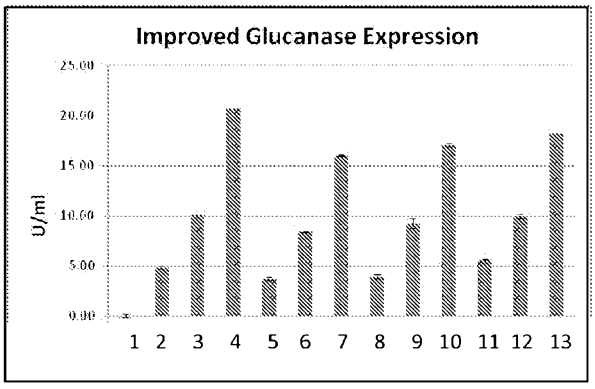


Figure 2

Legend

Lanes 1 = negative control
Lanes 2, 5, 8, 11 = SEQ ID NO:3
Lane 3, 6, 9, 12 = SEQ ID NO:4
Lane 4, 7, 10, 13 = SEQ ID NO:1

(57) Abstract: Polynucleotide sequences encoding a thermostable cellulase and directing its increased expression are provided, and hydraulic fracturing compositions comprising such thermostable cellulase.

WO 2013/148167 A2

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

— *of inventorship (Rule 4.17(iv))*

Published:

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

— *with sequence listing part of description (Rule 5.2(a))*

GENES ENCODING CELLULASE FOR HYDROLYZING GUAR FRACTURING FLUIDS UNDER EXTREME WELL CONDITIONS

RELATED APPLICATIONS

[0001] This application claims the benefit of priority under 35 U.S.C. §119(e) to U.S. Provisional Applications 61/618,610, filed March 30, 2012 and 61/660,556, filed June 15, 2012 each of which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] Polynucleotide sequences encoding a cellulase are provided. In particular, the provided sequences may provide increased expression of a specific, thermostable, thermotolerant, pressure stable cellulase, such as a cellulase for hydrolyzing guar fracturing fluids under extreme well conditions.

SEQUENCE LISTING

[0003] This application is being filed electronically via the USPTO EFS-WEB server, as authorized and set forth in MPEP § 502.05 and this electronic filing includes an electronically submitted sequence listing; the entire content of this sequence listing is hereby incorporated by reference into the specification of this application. The sequence listing is identified on the electronically filed ASCII (.txt) text file as follows:

File Name	Date of Creation	Size
D2570_SEQLISTING	March 9, 2013	40.4 kb

BACKGROUND

[0004] O-Glycosyl hydrolases (EC 3.2.1.-) are a widespread group of naturally-occurring enzymes that hydrolyze the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. The International Union of Biochemistry and Molecular Biology (IUBMB) enzyme nomenclature of glycosyl hydrolases (or glycosylases) is based principally on their substrate specificity and occasionally on their molecular mechanism (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), Accessed 10/24/2011).

[0005] IUBMB Enzyme Nomenclature EC 3.2.1.4 has been designated for a subgroup group of glycosylase-type enzymes termed "cellulases." Other names used for enzymes belonging to this group include: endoglucanase, endo-1,4-beta-glucanase, carboxymethyl cellulase, and beta-1,4-glucanase. The reaction catalyzed by enzymes belonging to this group is the endo-hydrolysis of 1,4-beta-D-glycosidic linkages in cellulose, lichenin, and cereal beta-D-glucans (such as barley beta-glucan). Since the predominant activities of the disclosed cellulase of the present invention are the endo-hydrolysis of barley beta-glucan and carboxymethyl cellulose, it is appropriately ascribed the IUBMB Enzyme Nomenclature EC 3.2.1.4.

[0006] An alternative classification of glycosyl hydrolases is based on amino acid sequence similarities (Henrissat, B. Accessed at UniProt 10/26/2011). According to this classification scheme, glycosyl hydrolases can be divided into more than 70 families. Based on a comparison of the primary amino acid sequence of the disclosed cellulase of the present invention with the sequences of other glycosyl hydrolases contained in public databases, the disclosed cellulase of the present invention may be assigned to glycosyl hydrolase Family 5. This family contains more than 20 endoglucanases (IUBMB Enzyme Nomenclature EC 3.2.1.4) whose predominant catalytic activity is the endo-hydrolysis of beta-1,4-glycosidic linkages in cellulosic substrates. Using this second way of classifying enzymes provides further support for the conclusion that the disclosed cellulase of the present invention should be ascribed the IUBMB Enzyme Nomenclature EC 3.2.1.4.

[0007] Cellulases are used for a variety of industrial and commercial purposes including but not limited to, oil and gas exploration, food and beverage, alcohol production potable or fuel, e.g. brewing, ethanol, wine, flavor, fragrance, textile, detergents, paper, pulp, environmental, and agriculture, as well as in research purposes (Rebecca S. Bryant, Erle C. Donaldson, Teh Fu Yen, George V. Chilingarian, Chapter 14 Microbial Enhanced Oil Recovery, In: Erle C. Donaldson, George V. Chilingarian and Teh Fu Yen, Editor(s), *Developments in Petroleum Science*, Elsevier, 1989, Volume 17, Part B, Pages 423-450; M. Karmakar and R.R. Ray, 2011. Current Trends in Research and Application of Microbial Cellulases. *Research Journal of Microbiology,* 6: 41-53).

[0008] Enzyme breakers have been successfully used in water based fracturing fluids since the early 1990s to hydrolyze polymeric viscosifiers (Brannon HD, Tjon-Joe-Pin RM, Carman PS, Wood WD, "*Enzyme breaker technology: a decade of improved well stimulation*",

paper SPE 84213 presented at the SPE Annual Technical Conference and Exhibition in Denver, Colorado, USA, 5-8 October 2003.

SUMMARY

[0009] Enzymes are proteins with 3-dimensional structures that are sensitive to higher temperatures and higher alkaline conditions present in many fracturing operations. The identification and/or engineering of enzymes that retain catalytic activity under challenging down-hole conditions (e.g., high temperature and pH) would broaden enzyme breaker applications in water based fracturing.

[0010] A composition is disclosed comprising a polymeric viscosifier, a surfactant, a thermostabilizer, and an enzyme breaker comprising a wild-type cellulase derived from a hyperthermophilic bacterium or a mutated variant thereof.

[0011] In some embodiments, the viscosifier is a guar gel comprising a linear guar, a crosslinked guar, or mixtures thereof. In some embodiments, the enzyme breaker specifically hydrolyzes ß-1,4 glycosidic bonds in the guar gel. In some embodiments, the enzyme breaker does not specifically hydrolyze α-1,6 glycosidic bonds in the guar gel. In some embodiments, the enzyme breaker retains its ability to hydrolyze ß-1,4 glycosidic bonds in the guar gel at temperatures up to about 275° F. In some embodiments, the enzyme breaker retains its ability to hydrolyze ß-1,4 glycosidic bonds in the guar gel at a pH of up to about 11.

[0012] In some embodiments, the enzyme breaker in the composition is a mutated variant comprising 12 mutations relative to the wild type cellulase. The 12 mutations may be selected from the group consisting of F38Y, Y61Q, M69E, D70P, R71S, I94Q, I166V, S183R, S191A, E212P, L231V, M276A, E277S, R280G, T297P, and T301Q. In some embodiments, the enzyme breaker has SEQ ID. NO. 2. In some embodiments, the enzyme breaker is encoded by a polynucleotide having SEQ ID. NO. 1. In some embodiments, the enzyme breaker is a mutated variant of the wild-type cellulase, and has a melting temperature that is at least 20° F greater than the melting temperature of the wild type cellulase at about pH 6.5, and at least 10° F greater than the melting temperature of the wild type cellulase at about pH 10.5.

[0013] A method is disclosed for reducing the viscosity of a polysaccharide gel that includes ß-1,4 glycosidic bonds. The method includes contacting the polysaccharide gel with a cellulase variant under permissive conditions and for a period of time sufficient to allow the

cellulase to hydrolyze the ß-1,4 glycosidic bonds in the polysaccharide gel thereby reducing the viscosity of the guar gel. The cellulase variant comprises at least 12 mutations compared to a wild-type cellulase derived from a hyperthermophilic bacterium, and the cellulase variant exhibits increased temperature and pH tolerance compared to the wild-type cellulase.

[0014] In some embodiments of the method, the at least 12 mutations in the cellulase variant were generated by Gene Site Saturation Mutagenesis comprising repeated cycles of reductive reassortment, recombination, and selection.

[0015] In some embodiments of the method, the polysaccharide gel comprises a linear guar, a crosslinked guar, or mixtures thereof.

[0016] In some embodiments of the method, the cellulase variant has SEQ ID NO:2. In some embodiments of the method, the cellulase variant is encoded by a polynucleotide having SEQ ID. NO. 1.

[0017] In some embodiments of the method, permissive conditions comprise a temperature range of from about 180° F to about 275° F. In some embodiments of the method, permissive conditions comprise a pH range of from about 9 to about 11.

[0018] A method of treating a subterranean formation is also disclosed. The method comprises: introducing into the subterranean formation a fracturing fluid that comprises a polysaccharide gel and a cellulase variant; and allowing the gel and the cellulase variant to react under permissive conditions and for a period of time sufficient to allow the cellulase variant to hydrolyze at least some of the ß-1,4 glycosidic bonds, but not the α-1,6 glycosidic bonds, in the gel, thereby generating a reduced viscosity fracturing fluid. The cellulase variant may comprise at least 12 mutations compared to a parental wild-type cellulase derived from a hyperthermophilic bacterium, and the cellulase variant may exhibit increased temperature and pH tolerance compared to the wild-type cellulase.

[0019] In some embodiments of the method, the reduced viscosity fracturing fluid generated by the cellulase variant comprises less residue that a comparable reduced viscosity fracturing fluid generated by a chemical breaker, wherein the dosage of cellulase variant and chemical breaker provide substantially the same reduction in viscosity.

[0020] In some embodiments of the method, the reduced viscosity fracturing fluid generated by the cellulase variant retains greater conductivity than a comparable reduced

viscosity fracturing fluid generated by a chemical breaker, wherein the dosage of cellulase variant and chemical breaker provide substantially the same reduction in viscosity.

[0021] In another embodiment the invention comprises SEQ ID NO:1, wherein said sequence encodes a protein. In a further embodiment the invention comprises a nucleotide sequence encoding a cellulase derived from *Thermotoga maritima,* or SEQ ID NO:3, comprising at least one mutation selected from T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, A6C, G6C, A9C, G9C, A15G, C15G, T22C, G22C, A24T, C24T, T33C, G33C, T39C, G39C, T40C, G40C, T42C, G42C, T54C, G54C, T57C, G57C, A66C, G66C, C81A, T81A, T84C, G84C, or any combination thereof, wherein optionally, any such mutations are silent. In a further embodiment of the invention, a least one such silent mutation results in expression of said cellulase at a higher level than a nucleotide sequence lacking at least one such mutation.

[0022] In another embodiment of the present invention, the invention comprises a nucleotide sequence from *Thermotoga maritima* having at least one mutation and having an increased expression level of a protein encoded by said nucleotide sequence compared to a *Thermotoga maritima* wild-type genomic sequence, wherein optionally, said mutation/s is silent.

[0023] In another embodiment of the present invention, the invention comprises a first nucleotide sequence encoding the polypeptide of SEQ ID NO:2 wherein said nucleotide sequence has been mutated with respect to a second sequence encoding SEQ ID NO:2 such that the expression level of said protein is increased relative to that of said protein encoded by said second nucleotide sequence.

[0024] A first nucleotide sequence encoding the polypeptide of SEQ ID NO:2 wherein said nucleotide sequence has been mutated with respect to a second sequence encoding SEQ ID NO:2 such that the expression level of said protein is increased relative to that of said protein encoded by said second nucleotide sequence.

[0025] In another embodiment of the present invention, the invention comprises a nucleotide sequence encoding a protein at least 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or 100% identical to SEQ ID NO:2, or a fragment thereof, wherein said nucleotide sequence comprises at least one mutation selected from T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, A6C, G6C, A9C, G9C, A15G, C15G,

T22C, G22C, A24T, C24T, T33C, G33C, T39C, G39C, T40C, G40C, T42C, G42C, T54C, G54C, T57C, G57C, A66C, G66C, C81A, T81A, T84C, G84C, or any combination thereof.

[0026] In another embodiment of the present invention, any of the proteins of the invention are expressed in bacterial expression systems, wherein the bacteria expression system is a gram-negative bacteria expression system, e.g., *Pseudomonas*, *E. coli*, *Ralstonia*, or *Caulobacter* expression system.

[0027] In another embodiment of the present invention, expression of the cellulase of the invention is produced at least 1.0 g/L, 2.0 g/L, 3.0 g/L, 4.0 g/L, 5.0 g/L, 6.0 g/L, 7.0 g/L, 8.0 g/L, 9.0 g/L, 10.0 g/L, 11.0 g/L, 12.0 g/L, 13.0 g/L, 14.0 g/L, 15.0 g/L, 16.0 g/L, 17.0 g/L, 18.0 g/L, 19.0 g/L, 20.0 g/L, 21.0 g/L, 22.0 g/L, 23.0 g/L, 24.0 g/L, 25.0, g/L, 26.0 g/L, 27.0 g/L, 28.0 g/L, 29.0 g/L, 30.0 g/L, 31.0 g/L, 32.0 g/L, 33.0 g/L, 34.0 g/L, or 35.0 g/L.

[0028] In another embodiment of the present invention, a cellulase of the present invention is combined with a second enzyme wherein the second enzyme is selected from the group consisting of: a lactase, a lipase, a protease, a catalase, a xylanase, a cellulase, a glucanase, a mannanase, an amylase, an amidase, an epoxide hydrolase, an esterase, phospholipase, transaminase, an amine oxidase, cellobiohydrolase, an ammonia lyase, or any combination thereof.

[0029] In another embodiment of the present invention, the invention comprises an isolated, recombinant, or synthetic nucleotide, having a nucleic acid sequence comprising SEQ ID NO:1, wherein the nucleic acid sequence encodes a polypeptide having a cellulase activity.

[0030] In another embodiment of the present invention, the invention comprises an isolated, recombinant, or synthetic nucleotide, comprising a nucleic acid sequence of SEQ ID NO:1, wherein the nucleic acid sequence encodes a polypeptide having a cellulase activity and the polypeptide comprises an amino acid sequence of SEQ ID NO:2, or an enzymatically active fragment thereof.

[0031] In another embodiment of the present invention, the invention comprises, an isolated, recombinant, or synthetic nucleic acid sequence comprising SEQ ID NO:1 that encodes a polypeptide having a cellulase activity, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO:2 and the polypeptide is produced in a recombinant *Pseudomonas fluorescens* expression system.

[0032] In another embodiment of the present invention, the present invention comprises a nucleotide sequence which encodes a polypeptide having cellulase activity, wherein the polypeptide is produced in a recombinant bacterial expression system.

[0033] In another embodiment of the present invention, the invention comprises a composition comprising a polypeptide encoded by SEQ ID NO:1, wherein optionally, the composition further comprises at least a second enzyme, and wherein optionally, the second enzyme is selected from the group consisting of: a lactase, a lipase, a protease, a catalase, a xylanase, a cellulase, a glucanase, a mannanase, an amylase, an amidase, an epoxide hydrolase, an esterase, phospholipase, transaminase, an amine oxidase, cellobiohydrolase, an ammonia lyase, or any combination thereof.

[0034] In another embodiment of the present invention, the invention comprises an isolated, recombinant, or synthetic nucleotide having a nucleic acid sequence comprising SEQ ID NO:1, wherein the nucleic acid sequence encodes a polypeptide having a cellulase activity.

[0035] In another embodiment of the present invention, the invention comprises an isolated, recombinant, or synthetic nucleotide comprising a nucleic acid sequence of SEQ ID NO:1, wherein the nucleic acid sequence encodes a polypeptide having a cellulase activity and the polypeptide comprises an amino acid sequence of SEQ ID NO:2, or an enzymatically active fragment thereof. In a further embodiment the isolated, recombinant, or synthetic nucleic acid sequence comprising SEQ ID NO:1 encodes a polypeptide having a cellulase activity, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO:2 and the polypeptide is produced in a recombinant *Pseudomonas fluorescens* expression system.

[0036] In another embodiment of the present invention, the invention comprises, a composition comprising a polymeric viscosifier, a surfactant, a thermostabilizer, and an enzyme breaker comprising a wild-type cellulase derived from a hyperthermophilic bacterium or a mutated variant thereof. In further embodiment of the composition, the viscosifier is a guar gel comprising a linear guar, a crosslinked guar, or mixtures thereof. In further embodiment of the composition the enzyme breaker specifically hydrolyzes ß-1,4 glycosidic bonds in the guar gel. In further embodiment of the composition the enzyme breaker does not specifically hydrolyze α-1,6 glycosidic bonds in the guar gel. In further embodiment of the composition the enzyme breaker retains its ability to hydrolyze ß-1,4 glycosidic bonds in the guar gel at temperatures up to about 275° F. In further embodiment of the composition wherein the enzyme breaker retains

its ability to hydrolyze ß-1,4 glycosidic bonds in the guar gel at a pH of up to about 11. In further embodiment of the composition, the enzyme breaker has SEQ ID. NO. 2. In further embodiment of the composition, the enzyme breaker is encoded by a polynucleotide having SEQ ID. NO. 1. In further embodiment of any of the above compositions, the enzyme breaker is a mutated variant of the wild-type cellulase, and has a melting temperature that is at least 20° F greater than the melting temperature of the wild type cellulase at about pH 6.5 and at least 10° F greater than the melting temperature of the wild type cellulase at about pH 10.5. In a further embodiment, the enzyme breaker is a mutated variant comprising 12 mutations relative to the wild type cellulase. In a further embodiment, the variant cellulase comprises at least one of the following mutations F38Y, Y61Q, M69E, D70P, R71S, I94Q, I166V, S183R, S191A, E212P, L231V, M276A, E277S, R280G, T297P and T301Q.

[0036] In a further embodiment, said enzyme breaker is a mutant variant of a cellulase encoded by SEQ ID NO:3.

[0037] In a further embodiment, the invention comprises a method of reducing the viscosity of a polysaccharide gel comprising ß-1,4 glycosidic bonds, the method comprising contacting the polysaccharide gel with a cellulase variant under permissive conditions and for a period of time sufficient to allow the cellulase to hydrolyze the ß-1,4 glycosidic bonds in the polysaccharide gel thereby reducing the viscosity of the guar gel, wherein the cellulase variant comprises at least 12 mutations compared to a wild-type cellulase derived from a hyperthermophilic bacterium, and wherein the cellulase variant exhibits increased temperature and pH tolerance compared to the wild-type cellulase. In a further embodiment, said 12 mutations were generated by Gene Site Saturation Mutagenesis comprising repeated cycles of reductive reassortment, recombination, and selection. In an alternative embodiment, said polysaccharide gel comprises a linear guar, a crosslinked guar, or mixtures thereof. In an alternative embodiment said cellulase variant is encoded by a polynucleotide having SEQ ID. NO. 1. In an alternative embodiment, wherein permissive conditions comprise a temperature range of from about 180° F to about 275° F.

In an alternative embodiment, wherein the permissive conditions comprise a pH range of from about 9 to about 11.

[0038] In a embodiment of the present invention, the invention comprises a method of treating a subterranean formation, comprising: introducing into the subterranean formation a

fracturing fluid that comprises a polysaccharide gel and a cellulase variant; and allowing the gel and the cellulase variant to react under permissive conditions and for a period of time sufficient to allow the cellulase variant to hydrolyze at least some of the ß-1,4 glycosidic bonds, but not the α-1,6 glycosidic bonds, in the gel, thereby generating a reduced viscosity fracturing fluid; wherein the cellulase variant comprises at least 12 mutations compared to a parental wild-type cellulase derived from a hyperthermophilic bacterium, and wherein the cellulase variant exhibits increased temperature and pH tolerance compared to the wild-type cellulase.

[0039] In a further embodiment of the above methods the hydrolysis with said cellulase variant comprises less residue that a comparable reduced viscosity fracturing fluid generated by a chemical breaker, wherein the dosage of cellulase variant and chemical breaker provide substantially the same reduction in viscosity.

[0040] In a further embodiment of the above methods wherein the reduced viscosity fracturing fluid generated by the cellulase variant retains greater conductivity than a comparable reduced viscosity fracturing fluid generated by a chemical breaker, wherein the dosage of cellulase variant and chemical breaker provide substantially the same reduction in viscosity.

DESCRIPTION OF THE INVENTION

[0041] Enzymes are proteins that act as catalysts. Proteins are polymers of amino acids linked in dehydration reactions by peptide bonds. The identity of the amino acids and the order in which they are linked to form proteins determines a given protein's activity. This order in which amino acids are assembled into proteins (the protein "sequence") is ultimately determined by the sequence of a DNA strand which "encodes" the protein.

[0042] The three-nucleotide sequence that specifies a given amino acid to be assembled into a protein is called a "codon." The 20 amino acids built into proteins are collectively encoded by 64 tri-nucleotide codon sequences. The series of codons which specifies a protein is called an "Open Reading Frame." An amino acid may be specified by as few as one or as many as six distinct codons. A change (or mutation) in the trinucleotide sequence of a codon that does not affect the amino acid specified is called a "silent" mutation.

[0043] As a result, there are many DNA sequences capable of encoding the same protein, because the DNA sequences differ from one another only through "silent" mutations. By altering one or more of the codons which encode a given protein, it may be possible to

greatly increase the amount of protein which a gene produces without affecting the sequence of the protein that is encoded.

[0044] In some embodiments, the invention comprises SEQ ID NO:1. In some embodiments, the invention comprises the polynucleotide sequence of SEQ ID NO:1. In some embodiments, this sequence encodes a protein. In some embodiments, this protein is an enzyme having cellulase activity.

[0045] The improved nucleotide sequence disclosed herein is given as SEQ ID NO:1 and encodes a previously disclosed cellulase enzyme (SEQ ID NO:2) that was evolved from a parent cellulase enzyme isolated from a DNA library originating from *Thermotoga maritima* strain MSB8. The disclosed cellulase is described in PCT Publication No. WO 2009/020459, as *SEQ ID NO:9 therein* (encoded by the polynucleotide *SEQ ID NO:8 therein*, described herein as SEQ ID NO:3). In some embodiments, the invention comprises the polynucleotide sequence of SEQ ID NO:1, or fragments thereof. In some embodiments, these sequences encode a protein. In some embodiments, the protein is an enzyme having cellulase activity.

[0046] The invention comprises multiple nucleotide base changes with respect to SEQ ID NO:3. These changes are silent as to the encoded protein. The 14 base changes are set forth below. "Position" indicates the number of the nucleotide within the Open Reading Frame, with the first nucleotide of the first codon numbered as 1. In the event that the Open Reading Frame of SEQ ID NO:1 is joined to another nucleic acid sequence at its 5' end so that the Open Reading Frame extends beyond the 5' end of SEQ ID NO:1, the "Position" will continue to refer to the bases as numbered from the 5' end of the Open Reading Frame of SEQ ID NO:1. Similarly, if the Open Reading Frame of SEQ ID NO:1 is truncated so that the Open Reading Frame does not begin at the 5' end of a sequence related to SEQ ID NO:1, the numbering system will continue to originate from the 5' end of said sequence corresponding to the 5' end of SEQ ID NO:1.

[0047] The nucleotide base changes, or mutations, are specified using the notation (old nucleotide) (position) (new nucleotide). The mutations are as follows: T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, or any combination thereof.

[0048] The base changes which distinguish SEQ ID NO:1 from prior reported sequences encoding the disclosed cellulase, collectively and individually, result in an Open

Reading Frame which leads to a higher level of protein expression than previously employed nucleotide sequences encoding the same protein.

[0049] In some embodiments, a nucleotide sequence encoding a cellulase derived from *Thermotoga maritima* is disclosed, wherein the nucleotide sequence comprises at least one mutation selected from T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, A6C, G6C, A9C, G9C, A15G, C15G, T22C, G22C, A24T, C24T, T33C, G33C, T39C, G39C, T40C, G40C, T42C, G42C, T54C, G54C, T57C, G57C, A66C, G66C, C81A, T81A, T84C, G84C, or any combination thereof. In some aspects of these embodiments, at least one mutation is silent as to the sequence of the encoded protein. In other aspects, at least one mutation results in the nucleotide sequence harboring at least one mutation directing expression of the cellulase at a higher level than a nucleotide sequence lacking the at least one mutation and not otherwise differing from the nucleotide sequence of the above.

[0050] In some embodiments, a nucleotide sequence encoding a cellulase is disclosed, wherein the nucleotide sequence comprises SEQ ID NO:3 and having at least one mutation selected from T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, A6C, G6C, A9C, G9C, A15G, C15G, T22C, G22C, A24T, C24T, T33C, G33C, T39C, G39C, T40C, G40C, T42C, G42C, T54C, G54C, T57C, G57C, A66C, G66C, C81A, T81A, T84C, G84C, or any combination thereof.

[0051] In some embodiments, a nucleotide sequence from *Thermotoga maritima* is disclosed having at least one mutation which increases the expression level of a protein encoded by said nucleotide sequence compared to a *Thermotoga maritima* genomic sequence. In some aspects, at least one mutation is silent.

[0052] In some embodiments, a first nucleotide sequence encoding the polypeptide of **SEQ ID NO:2** is disclosed wherein the nucleotide sequence has been mutated with respect to a second sequence encoding the polypeptide of SEQ ID NO:2 such that the expression level of the protein is increased relative to that of the protein encoded by the second nucleotide sequence.

[0053] *Thermotoga maritima* is a thermophilic eubacteria characterized by its ability to grow in extreme salt concentrations (i.e., from 0.25% NaCl to 6.00% NaCl). *Thermotoga maritima* belongs to the order Thermotogales whose members are thermophilic, rod-shaped, anaerobic and gram-negative. The minimum temperature for growth is around 55°C, optimum is 80°-85°C, and maximum is about 90°C. In some embodiments, the minimum temperature is less

than 55°C and the maximum temperature is greater than 90°C. These bacteria have slowly evolved from one of the deepest branches in the kingdom of eubacteria. Members of Thermotogales have been described as "wide-spread and cosmopolitan" (Huber, R. *et al.*, 2006), thriving in active geothermal areas. *Thermotoga maritima* is closely related to the species *Thermotoga neapolitana, Thermotoga petrophila,* and *Thermotoga naphthophila.* Specimens of *Thermotoga maritima* have been obtained from sea floors in Vulcano, Italy; Riberia Quente and Sao Miguel Island, Azores; Sangeang Island, Indonesia; and Fiji Island. (Huber, R. *et al.*, 2006).

[0054] Strain MSB8 was isolated from a geothermally heated marine sediment at Vulcano, Italy (Huber, 1986). The temperature at the collection site ranged from 70-100°C, with a pH of 6.5-7.0. The strain has been deposited at the Deutsche Sammlung von Mikroorganismen as DSM 3109 and at ATCC as ATCC43589 (Huber, R. *et al.*, 2006).

[0055] *Thermotoga maritima* strain MSB8 has been studied for its enzyme encoding genes due to the exceptional thermostability of the enzymes it produces. Liebl (Liebl, W. *et al.*, 1996) has published an "Analysis of a *Thermotoga maritima* DNA fragment encoding two similar thermostable cellulases, CelA and CelB, and characterization of the recombinant enzymes." Additionally, genes for amylolytic enzymes (Bibel, M. *et al.*, 1998), reverse gyrase (Bouthier de la Tour, C. *et al.*, 1998), alpha-amylase (Liebl, W. *et al.*, 1997), alpha-glucuronidase (Ruile, P. *et al.*, 1997), xylanase (Winterhalter, C. *et al.*, 1995), beta-glucosidase (Liebl, W. *et al.*, 1994), glucanotransferase (Liebl, W. *et al.*, 1992) have been isolated and analyzed. A study by Bronnenmeier (Bronnenmeier, K. *et al.*, 1995), "Purification of *Thermotoga maritima* enzymes for the degradation of cellulosic materials" has shown that these enzymes are of value for degrading cellulose and xylan.

Expression Systems

[0056] In some embodiments, the DNA encoding the cellulase of the present invention may be introduced, either on a plasmid or stably transformed into the genome of, for example, any number of gram negative bacterial systems such as E. coli, Pseudomonas species such as fluorescens, Pseudomonas putida, Pseudomonas aeruginosa, Ralstonia species, or Caulobacter species. Similarly, the cellulase may be introduced into any number of gram positive bacterial expression systems such as Bacillus species such as Bacillus subtilis, Bacillus megaterium, Bacillus brevis, Lactococcus species such as Lactococcus lactis, Lactobacillus

species, Streptomyces species such as Streptomyces lividans. Other gram negative, gram positive or unrelated eubacterial or archaeal expression systems may be used to express the cellulase.

[0057] In some embodiments, SEQ ID NO:1 is used to direct an increased level of expression in a number of systems in which the disclosed cellulase protein may be expressed. SEQ ID NO:1 may be introduced into any number of expression systems to express the disclosed cellulase at an improved accumulation level. For example, SEQ ID NO:1 may be introduced, either on a plasmid or stably transformed into the genome of, for example, any number of gram negative bacterial systems such as *E. coli*, *Pseudomonas* species such as *fluorescens*, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Ralstonia* species, or *Caulobacter* species. Similarly, SEQ ID NO:1 may be introduced into any number of gram positive bacterial expression systems such as *Bacillus* species such as *Bacillus subtilis*, *Bacillus megaterium*, *Bacillus brevis*, *Lactococcus* species such as *Lactococcus lactis*, *Lactobacillus* species, *Streptomyces* species such as *Streptomyces lividans*. Other gram negative, gram positive or unrelated eubacterial or archaeal expression systems may be used to express SEQ ID NO:1. In a further embodiment, SEQ ID NO:1 may be introduced into any number of eukaryotic expression systems such as *Saccharomyces*, *Schizosaccharomyces pombe*, *Pichia pastoris*, and *Hansanuela polymorpha*.

[0058] More specifically, SEQ ID NO:1 may be introduced into a plasmid to direct its expression. Plasmids to which SEQ ID NO:1 may be introduced include, for example, *E. coli* expression vectors of the families pQE, pET, and pASK; *Pseudomonas* expression vectors of the families pCN51 LT8, RSF1010, pWZ112T, and pMYC; *Bacillus* expression vectors of the families pBAX, pHT01, and pHIS1525; *Streptomyces* expression vectors of the families pIJ6021 and pIJ2460; and *Lactococcus*: expression vectors of the families pNZ9530 and pNZ8148, for example. These examples are for demonstrative purposes and do not represent a complete set of vectors in which the polynucleotide sequence of SEQ ID NO:1 can be expressed.

[0059] In some embodiments, the expression system could be any *Pseudomonas fluorescens* expression system known in the art, for example, the *Pseudomonas fluorescens* expression system that is commercially available from Dow Global Technologies Inc., strain DC454 (US Patent PUB. APP. NO. 20050130160 and US Patent PUB. APP. NO. 20050186666). A nucleic acid sequence encoding the cellulase enzyme or polypeptide is inserted either in the pMYC vector (Dow Global Technologies Inc., US Patent PUB. APP. NO.

20050130160) or in the pDOW1169 vector (Dow Global Technologies Inc., US Patent PUB. APP. NO. 20080058262) and then introduced into the *Pseudomonas fluorescens* host by electroporation. Those skilled in the art will know alternative vectors that can be used as embodiments of this invention.

[0060] In some embodiments, the cellulase will be expressed at least at the following expression levels: 1.0 g/L, 2.0 g/L, 3.0 g/L, 4.0 g/L, 5.0 g/L, 6.0 g/L, 7.0 g/L, 8.0 g/L, 9.0 g/L, 10.0 g/L, 11.0 g/L, 12.0 g/L, 13.0 g/L, 14.0 g/L, 15.0 g/L, 16.0 g/L, 17.0 g/L, 18.0 g/L, 19.0 g/L, 20.0 g/L, 21.0 g/L, 22.0 g/L, 23.0 g/L, 24.0 g/L, 25.0, g/L, 26.0 g/L, 27.0 g/L, 28.0 g/L, 29.0 g/L, 30.0 g/L, 31.0 g/L, 32.0 g/L, 33.0 g/L, 34.0 g/L, 35.0 g/L, or more.

Nucleic Acid

[0061] The invention provides isolated, synthetic, or recombinant nucleic acids comprising sequences completely complementary to the nucleic acid sequences of the invention (complementary (non-coding) and coding sequences also hereinafter collectively referred to as nucleic acid sequences of the invention).

[0062] The invention provides isolated, synthetic, or recombinant nucleic acids comprising a nucleic acid encoding at least one polypeptide having a cellulolytic activity, wherein the nucleic acid comprises a sequence having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or complete (100%) sequence identity (homology) to an exemplary nucleic acid of the invention, including the sequence of SEQ ID NO:1. For example, the invention provides isolated, synthetic, or recombinant nucleic acids comprising a nucleic acid sequence SEQ ID NO:1 (the exemplary polynucleotide sequence of this invention). The invention provides isolated, synthetic, or recombinant nucleic acids encoding a polypeptide comprising a sequences as set forth in SEQ ID NO:2 (the exemplary polypeptide sequences of this invention), and enzymatically active fragments thereof.

Polypeptide

[0063] Polypeptides and peptides of the invention are isolated, synthetic, or recombinant polypeptides. Peptides and proteins can be recombinantly expressed in vitro or in vivo. The peptides and polypeptides of the invention can be made and isolated using any method known in the art. Polypeptides and peptides of the invention can also be synthesized, whole or in part, using methods well known in the art. For example, cellulase polypeptides can be produced in a standard recombinant expression system (as described herein), chemically synthesized, or purified from organisms in which they are naturally expressed.

[0064] The invention provides isolated, synthetic, or recombinant polypeptides having cellulolytic activity comprising an amino acid sequence having at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more, or has 100% (complete) sequence identity to an exemplary amino acid sequence of the invention (e.g., SEQ ID NO:2), or an enzymatically active fragment thereof.

[0065] The invention provides isolated, synthetic, or recombinant polypeptides comprising a sequence as set forth in SEQ ID NO:2, and enzymatically active fragments thereof, and variants thereof.

[0066] In alternative embodiments, the invention provides polypeptides (and the nucleic acids that encode them) having cellulolytic activity but lacking a signal sequence, a prepro domain, a dockerin domain, and/or a carbohydrate binding module (CBM); and in one aspect, the carbohydrate binding module (CBM) comprises, or consists of, a cellulose binding module, a lignin binding module, a xylan binding module, a xylose binding module, a mannose binding module, a xyloglucan-specific module, and/or a arabinofuranoside binding module.

[0067] In alternative embodiments, the invention provides polypeptides (and the nucleic acids that encode them) having a cellulolytic activity further comprising a heterologous sequence; and in one aspect, the heterologous sequence comprises, or consists of a sequence encoding: (i) a heterologous signal sequence, a heterologous carbohydrate binding module, a heterologous dockerin domain, a heterologous catalytic domain (CD), or a combination thereof; (ii) the sequence of (i), wherein the heterologous signal sequence, carbohydrate binding module or catalytic domain (CD) is derived from a heterologous enzyme; or, (iii) a tag, an epitope, a

targeting peptide, a cleavable sequence, a detectable moiety or an enzyme; and in one aspect, the heterologous carbohydrate binding module (CBM) comprises, or consists of, cellulose binding module, a lignin binding module, a xylan binding module, a xylose binding module, a mannose binding module, a xyloglucan-specific module and/or a arabinofuranoside binding module; and in one aspect, the heterologous signal sequence targets the encoded protein to a vacuole, the endoplasmic reticulum, a chloroplast or a starch granule.

Enzymatic Activity

[0068] The enzymatic hydrolysis of pNP-β-D-lactopyranoside by the disclosed cellulase can be used as a measure of activity of the enzyme. The liberation of p-nitrophenol can be followed spectrophotometrically at 405nm. The increase in absorbance at 405nm can be converted to μmoles of p-nitrophenol by using a standard absorbance at those defined conditions. One unit of activity is defined as the quantity of enzyme required to liberate 0.42 μmole of p-nitrophenol from 2 mM pNP-β-D-lactopyranoside during one minute at pH 7.00 and 80°C. (Advances in Carbohydrate Chemistry and Biochemistry, Academic Press, 1999)

Thermostability

[0069] In some aspects, the recombinant nucleic acid of the present invention encodes a polypeptide having a cellulolytic activity that is thermostable. For example, a polypeptide of the invention, SEQ ID NO:2, or the variant evolved enzymes of the invention can be thermostable. The thermostable polypeptide according to the invention can retain binding and/or enzymatic activity, e.g., cellulolytic activity, a under conditions comprising a temperature in the range from greater than 37°C to about 95°C, or between about 55°C to about 85°C, or between about 70°C to about 75°C, or between about 70°C to about 95°C, between about 90°C to about 95°C, between about 95°C to about 105°C, or between about 95°C to about 110°C. In some aspects, wherein the polypeptide can retain binding and/or enzymatic activity, e.g., cellulolytic activity, under conditions comprising 1°C to about 5°C, between about 5°C to about 15°C, between about 15°C to about 25°C, between about 25°C to about 37°C. In some aspects polypeptides of the invention can retain binding and/or enzymatic activity, e.g., cellulolytic activity, under conditions comprising 90°C, 91°C, 92°C, 93°C, 94°C, 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 103.5°C, 104°C, 105°C, 107°C, 108°C, 109°C or 110°C, or

more. In some embodiments, the thermostable polypeptides according to the invention retains activity, e.g., a cellulolytic activity at a temperature in the ranges described above, under acidic conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5 or pH 4 or less (more acidic), or, retain a cellulolytic activity after exposure to acidic conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5 or pH 4 or less (more acidic); or, retain activity under basic conditions comprising about pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11, pH 11.5, pH 12, pH 12.5 or more (more basic) or, retain a cellulolytic activity after exposure to basic conditions comprising about pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11, pH 11.5, pH 12, pH 12.5 or more (more basic).

Thermotolerance

[0070] In some aspects, the recombinant nucleic acid of the present invention encodes a polypeptide having a cellulolytic activity that is thermotolerant. For example, a polypeptide of the invention, SEQ ID NO:2, or the variant evolved enzymes of the invention can be thermotolerant. In some aspects, the cellulolytic activity is thermotolerant, e.g., wherein the polypeptide retains cellulolytic activity after exposure to a temperature in the range from greater than 37°C to about 95°C, or between about 55°C to about 85°C, or between about 70°C to about 75°C, or between about 70°C to about 95°C, between about 90°C to about 95°C, between about 95°C to about 105°C, or between about 95°C to about 110°C. In some aspects, wherein the polypeptide retain a cellulolytic activity after exposure to conditions comprising a temperature range of between about 1°C to about 5°C, between about 5°C to about 15°C, between about 15°C to about 25°C, between about 25°C to about 37°C. In some aspects polypeptides of the invention can retain a cellulolytic activity after exposure to a temperature up to 90°C, 91°C, 92°C, 93°C, 94°C, 95°C, 96°C, 97°C, 98°C, 99°C, 100°C, 101°C, 102°C, 103°C, 104°C, 105°C, 107°C, 108°C, 109°C or 110°C, or more. In some aspects, the polypeptides encoded by nucleic acids of the invention retain cellulolytic activity under acidic conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5 or pH 4 or less (more acidic), or, retain a cellulolytic activity after exposure to acidic conditions comprising about pH 6.5, pH 6, pH 5.5, pH 5, pH 4.5 or pH 4 or less (more acidic); or, retain a cellulolytic activity under basic conditions comprising about pH 7, pH 7.5 pH 8.0, pH 8.5, pH 9, pH 9.5, pH 10, pH 10.5, pH 11, pH 11.5, pH 12, pH 12.5 or more (more basic).

Cellulosic digestion

[0071] In some aspects, the compositions and methods of the invention are used in the enzymatic digestion of biomass and can comprise use of many different enzymes, including the cellulases and hemicellulases. Cellulases used to practice the invention can digest cellulose to glucose. In some aspects, compositions used to practice the invention can include mixtures of enzymes, e.g., xylanases, xylosidases (e.g., β-xylosidases), cellobiohydrolases, and/or arabinofuranosidases, or other enzymes that can digest hemicelluloses, cellulose, and lignocellulosic material, to fermentable sugars and/or to monomer sugars.

[0072] Enzymes, e.g., endoglucanases, of the invention are used to digest cellulose or any beta-1,4-linked glucan-comprising synthetic or natural material, including those found in any plant material. Enzymes, e.g., endoglucanases, of the invention are used as commercial enzymes to digest cellulose from any source, including all biological sources, such as plant biomasses, e.g., corn, grains, grasses (e.g., Indian grass, such as Sorghastrum nutans; or, switch grass, e.g., Panicum species, such as Panicum virgatum), or, woods or wood processing byproducts, e.g., in the wood processing, pulp and/or paper industry, in textile manufacture and in household and industrial cleaning agents, and/or in biomass waste processing.

Dietary

[0073] In some embodiments, the cellulase of the present invention may be used to pre-treat, modify, digest, or enhance the digestion of, a food, food additive, or dietary supplement for animals or human beings. In some embodiments, the cellulase of the present invention may be used as a food, food additive, or dietary supplement for animals or human beings. In some aspects the cellulase will treat or will act as a prophylaxis for digestive disorders. In another aspect of the present invention the cellulase will alter or enhance digestion. In another aspect of the present invention the cellulase will enhance, alter, or aid in the digestion of foodstuffs. In a further aspect of the invention the cellulase will enhance, aid, or alter the nutrient value of foodstuffs. In a further aspect, the cellulase is active in the digestive tract, e.g., in a stomach and/or intestine.

[0074] In some embodiments, the cellulase of the invention may be used as an animal feed or an animal feed additive. In some embodiments the thermostability and or

thermotolerance of the cellulase allows for the formation of pellets without the need for a secondary agent such as salt or wax. An animal feed comprising a cellulase can be provided to an animal in any formulation known to those skilled in the art. Examples of animal feed formulations include, but are not limited to: a delivery matrix, a pellet, a tablet, a gel, a liquid, a spray, ground grain, or a powder.

[0075] The invention provides edible enzyme delivery matrix comprising a thermostable recombinant cellulase enzyme, e.g., a polypeptide of the invention. The invention provides methods for delivering a cellulase supplement to an animal, the method comprising: preparing an edible enzyme delivery matrix in the form of pellets comprising a granulated edible carrier and a thermostable recombinant cellulase enzyme, wherein the pellets readily disperse the cellulase enzyme contained therein into aqueous media, and administering the edible enzyme delivery matrix to the animal. The recombinant cellulase enzyme can comprise a polypeptide of the invention. The granulate edible carrier can comprise a carrier selected from the group consisting of a grain germ, a grain germ that is spent of oil, a hay, an alfalfa, a timothy, a soy hull, a sunflower seed meal and a wheat midd. The edible carrier can comprise grain germ that is spent of oil. The cellulase enzyme can be glycosylated to provide thermostability at pelletizing conditions. The delivery matrix can be formed by pelletizing a mixture comprising a grain germ and a cellulase. The pelletizing conditions can include application of steam. The pelletizing conditions can comprise application of a temperature in excess of about 80°C for about 5 minutes and the enzyme retains a specific activity of at least 350 to about 900 units per milligram of enzyme.

Methods of making ethanol

[0076] The invention provides methods for making ethanol comprising contacting a starch- comprising composition with a polypeptide having a cellulolytic activity, wherein the polypeptide has a sequence of the invention, or the polypeptide is encoded by a nucleic acid comprising a sequence of the invention, or an enzymatically active fragment thereof. The invention provides compositions comprising a starch and a polypeptide having a cellulolytic activity, wherein the polypeptide has a sequence of the invention, or the polypeptide is encoded by a nucleic acid comprising a sequence of the invention, or an enzymatically active fragment thereof.

Brewing and fermenting

[0077] The invention provides methods of brewing (e.g., fermenting) beer comprising the cellulase of the invention. In one exemplary process, starch-containing raw materials are disintegrated and processed to form a malt. An enzyme of the invention is used at any point in the fermentation process. The cellulase of the invention can be used in the brewing industry for the degradation of beta-glucans. In some aspects, the cellulases of the invention are used in the brewing industry for the clarification of the beverage. Enzymes of the invention can be used in the beverage industry in improving filterability of wort or beer, as described, e.g., in U.S. Pat. No. 4,746,517.

[0078] In some aspects, the cellulase of the invention is used in mashing and conversion processes. In the brewing and fermentation industries, mashing and conversion processes are performed at temperatures that are too low to promote adequate degradation of water-soluble glucans, mannans, arabinoxylans or xylans, or other polysaccharides. These polymers form gummy substrates that can cause increased viscosity in the mashing wort, resulting in longer mash run-off, residual haze and precipitates in the final beer product due to inefficient filtration and low extraction yield.

[0079] In some aspects, the cellulase of the invention are used in malthouse operations, e.g., glucanase is added to the process water, to shorten germination times and/or to encourage conversion of poor quality barley to acceptable malts. In some aspects, enzymes of the invention are used for mashing, e.g., they are added to increase wort filterability and/or improve lautering (separating the wort from the mash). In some aspects, enzymes of the invention are used in the fermentor and/or settling tank to, e.g., assist in haze clearing and/or to improve filtration. In some aspects, enzymes of the invention are used in adjunct brewing, e.g., a glucanase of the invention is added to breakdown glucans, mannans, arabinoxylans, or xylans, or other polysaccharides from barley, wheat, and/or other cereals, including glycans in malt. In some aspects, enzymes of the invention are used in malt brewing, e.g., a glucanase of the invention is added to modify poor malts with high glucan content.

[0080] The cellulase of the invention can be used in any beer or alcoholic beverage producing process, as described, e.g., in U.S. Patent No. 5,762,991; 5,536,650; 5,405,624; 5,021,246; 4,788,066.

Treating foods and food processing

[0081] The cellulases of the invention have numerous applications in food processing industry. For example, in one aspect, the enzymes of the invention are used to improve the extraction of oil from oil-rich plant material, e.g., oil-rich seeds, for example, soybean oil from soybeans, olive oil from olives, rapeseed oil from rapeseed and/or sunflower oil from sunflower seeds.

[0082] The cellulase of the invention can be used for separation of components of plant cell materials. For example, enzymes of the invention can be used in the separation of glucan- rich material (e.g., plant cells) into components. In some aspects, enzymes of the invention can be used to separate glucan- rich or oil-rich crops into valuable protein and oil and hull fractions. The separation process may be performed by use of methods known in the art.

[0083] The cellulase of the invention can be used in the preparation of fruit or vegetable juices, syrups, extracts and the like to increase yield. The enzymes of the invention can be used in the enzymatic treatment (e.g., hydrolysis of glucan- comprising plant materials) of various plant cell wall-derived materials or waste materials, e.g. from cereals, grains, wine or juice production, or agricultural residues such as vegetable hulls, bean hulls, sugar beet pulp, olive pulp, potato pulp, and the like. The enzymes of the invention can be used to modify the consistency and appearance of processed fruit or vegetables. The enzymes of the invention can be used to treat plant material to facilitate processing of plant material, including foods, facilitate purification or extraction of plant components. The cellulase of the invention can be used to improve feed value, decrease the water binding capacity, improve the degradability in waste water plants and/or improve the conversion of plant material to ensilage, and the like. The cellulase of the invention can also be used in the fruit and brewing industry for equipment cleaning and maintenance.

Detergent Compositions

[0084] The invention provides detergent compositions comprising one or more polypeptides of the invention and methods of making and using these compositions. The invention incorporates all methods of making and using detergent compositions, see, e.g., U.S. Patent No. 6,413,928; 6,399,561; 6,365,561; 6,380,147. The detergent compositions can be a

one and two part aqueous composition, a non-aqueous liquid composition, a cast solid, a granular form, a particulate form, a compressed tablet, a gel and/or a paste and a slurry form. The invention also provides methods capable of a rapid removal of gross food soils, films of food residue and other minor food compositions using these detergent compositions. Enzymes of the invention can facilitate the removal of starchy stains by means of catalytic hydrolysis of the starch polysaccharide. Enzymes of the invention can be used in dishwashing detergents in textile laundering detergents. The actual active enzyme content depends upon the method of manufacture of a detergent composition and is not critical, assuming the detergent solution has the desired enzymatic activity. In some aspects, the amount of glucosidase present in the final solution ranges from about 0.001 mg to 0.5 mg per gram of the detergent composition. The particular enzyme chosen for use in the process and products of this invention depends upon the conditions of final utility, including the physical product form, use pH, use temperature, and soil types to be degraded or altered. The enzyme can be chosen to provide optimum activity and stability for any given set of utility conditions. The detergents of the invention can comprise cationic, semi-polar nonionic, or zwitterionic surfactants; or, mixtures thereof.

[0085] The present invention provides cleaning compositions including detergent compositions for cleaning hard surfaces, detergent compositions for cleaning fabrics, dishwashing compositions, oral cleaning compositions, denture cleaning compositions, and contact lens cleaning solutions. In some aspects, the invention provides a method for washing an object comprising contacting the object with a polypeptide of the invention under conditions sufficient for washing. A polypeptide of the invention may be included as a detergent additive. The detergent composition of the invention may, for example, be formulated as a hand or machine laundry detergent composition comprising a polypeptide of the invention. A laundry additive suitable for pre-treatment of stained fabrics can comprise a polypeptide of the invention. A fabric softener composition can comprise a polypeptide of the invention. Alternatively, a polypeptide of the invention can be formulated as a detergent composition for use in general household hard surface cleaning operations.

Oil and Gas Exploration and Clean-up

[0086] To increase the productivity of oil and gas wells and shale gas reservoirs, a highly specialized technique called "hydraulic fracturing" is being increasingly utilized. In a

typical hydraulic fracturing operation, large volumes of guar-based fluid (in gel form and referred to as "fracturing fluid") are pumped into the wellbore under very high hydrostatic pressure. The pressurized fluid creates new fissures and fractures in the formation surrounding the wellbore. The sand particles contained in the fracturing fluid move and settle into the newly-created fractures and function to prop these channels open thus increasing oil and gas flow. Once the sand is deposited into the fractures, the gel has to be degraded (i.e., broken down) and brought back up to the surface so as to remove any blockage to the flow of oil or gas. Industry uses viscosity breakers (such as oxidizers, acids, or enzymes) to degrade the fracturing fluid and to remove any solid gel residue from the fissures and fractures.

[0087] In some embodiments, the disclosed cellulase will be used as a high temperature viscosity breaker to enhance oil and gas operations. More specifically, the disclosed cellulase of the present invention will be applied to a fracturing fluid when hydraulic fracturing is performed in oil or gas wells.

[0088] The enzyme encoded by SEQ ID NO:1, as well as cellulases encoded by other polynucleotides disclosed herein, or obtainable by methods disclosed herein, may potentially be used to hydrolyze a broad spectrum of polysaccharides – many of which are useful in oil and gas drilling, fracturing and well clean-up operations. The disclosed cellulases exhibit broad spectrum β-glycosidase activity, e.g., against guar, hydroxypropyl guar, carboxymethyl guar, carboxymethyl hydroxypropyl guar, carboxymethyl cellulose, barley β-glucan, and locust bean gum. The enzyme activity pattern is preferably both endo and exo, allowing effective reduction in the viscosity of polysaccharides, e.g., guar and derivatized guar solutions, by cleaving within long polysaccharide chains and also by cleaving disaccharide units from the ends of the polymers. Besides the aforementioned polysaccharides, other substrates of the disclosed enzymes include those capable of forming linear or cross-linked gels. Examples of suitable polysaccharide substrates include glactomannan gums, guars, derivatized guars, cellulose and cellulose derivatives, starch, starch derivatives, xanthan, derivatized xanthan and mixtures thereof. Specific examples also include, but are not limited to, guar gum, guar gum derivative, locust bean gum, karaya gum, xanthan gum, cellulose and cellulose derivatives, etc. Typical polymers or gelling agents to which the disclosed enzymes may be directed include guar gum, hydroxypropyl guar, carboxymethyl hydroxypropyl guar, hydroxyethyl cellulose, carboxymethyl hydroxyethyl cellulose, carboxymethyl cellulose, dialkyl carboxymethyl cellulose, etc. Other

examples of polymers include, but are not limited to, phosphomannans, scleroglucans, dextrans and other types of polymers. In some embodiments, a polymer substrate is carboxymethyl hydroxypropyl guar. In some embodiments, a disclosed enzyme may also be effective in hydrolyzing biogums (e.g., succinoglycan biogums made from date syrup or sucrose). In some embodiments, a disclosed enzyme may be used to hydrolyze cellulose-containing or derivatized cellulose-containing polymers – typically, the enzymes attack glucosidic linkages of the cellulose backbone. The disclosed enzymes may be suitable for degrading the polymer into mostly monosaccharide units, in some cases, by specifically hydrolyzing the exo(1,4)-β-D-glucosidic and endo(1,4)-β-D-glucosidic linkages between monosaccharide units and the cellulose backbone in the (1,4)-β-D-glucosidic linkages of any cellobiose fragments.

[0089] In each fracturing job that uses the disclosed cellulases, field operators will generally first perform an enzyme dose optimization study in an industrial lab. Such studies may include dilution of the cellulase to a concentration of 10~400 ppm and mixed with linear or cross-linked guar gum (25–60 lb/1,000 gal). Depending on the application conditions, guar gum maybe cross-linked using a cross-linker, especially for wells where higher temperature, pressure, and pH conditions are present. The enzyme dose information resulting from such optimization studies is then used in the actual fracturing job.

[0090] The unique activity of the disclosed cellulase allows for the hydrolysis of guar-based fracturing fluids in a smooth and controlled manner in deep wells, where high temperature and high pH conditions are present. Compared to chemical breakers, the disclosed cellulase of the present invention provides a non-corrosive and environmentally benign alternative to the harsh and non-selective chemical breakers.

[0091] In some embodiments of the present invention the cellulase may be used to treat, clean, or alter fluids used in oil and gas exploration activities. In a further aspect the cellulase of this invention will treat or alter the fluids, in part, or completely, so that the fluids may be used again, or recycled, for use in additional oil and gas exploration activities or to be disposed of in an environmentally friendly way.

[0092] Described herein are compositions for uses including oil and gas fracturing procedures. In some embodiments, the compositions include one or more cellulase enzymes derived from hyperthermophilic bacteria and/or non-naturally occurring variants thereof, in combination with a polymeric viscosifier. In some embodiments, the compositions may

optionally include additional ingredients including agents to control fluid loss, reduce formation damage, adjust pH, control microbial growth and improve temperature stability. Some typical additional ingredients include acids, anti-bacterial agents, clay stabilizer, corrosion inhibitor, crosslinker, friction reducer, iron control, scale inhibitor, surfactants and thermal stabilizers (e.g., sodium thiosulfate).

[0093] In some embodiments, the cellulase enzymes described herein possess glucanase, e.g., endoglucanase, mannanase, xylanase activity or a combination of these activities. In some aspects, the glucanase activity is an endoglucanase activity (e.g., endo-1,4-beta-D-glucan 4-glucano hydrolase activity) and comprises hydrolysis of 1,4-beta-D-glycosidic linkages in cellulose, cellulose derivatives (e.g., carboxy methyl cellulose and hydroxy ethyl cellulose) lichenin, beta- 1,4 bonds in mixed beta- 1,3 glucans, such as cereal beta-D-glucans or xyloglucans and other plant material containing cellulosic parts. In alternative aspects, these glucanases e.g., endoglucanases, mannanases, xylanases have increased activity and stability, including thermotolerance or thermostability, at increased or decreased pHs and temperatures.

[0094] In some embodiments, the enzyme breakers described may be encapsulated to stabilize the enzyme, improve thermostability and alkaline pH tolerance, and provide controlled release. An encapsulated breaker having a coating or membrane that hydrolytically degrades may be superior to prior art enzyme breaker systems, because it could allow better control of release time and ease of handling not previously afforded. For example, because the breaker is encapsulated in a material that reacts with water, rather than simply dissolves or dissipates in water, the release can be controlled through the reaction rate of the coating with water. Likewise, by insulating the enzyme from the harsh down-hole conditions (high temperature and pH) for some period of time, can provide delayed and complete viscosity breaks. Those skilled in the art will appreciate that the reaction rate of the coating (and therefore the breaker release profile) can be varied broadly depending on the encapsulating polymer chemistry employed. Examples of breaker encapsulation compositions and methods are provided in U.S. Patent Nos. 5,164,099, 6,163,766, 5,373,901, 5,437,331, and 6,357,527, the disclosures of each of which are incorporated herein by reference thereto.

DNA Sequences and Encoded Cellulase Sequences

[0095] Some cellulases derived from hyperthermophilic bacteria and/or non-naturally occurring variants thereof are described in PCT publication WO 2009/02049; the entire disclosure of which is incorporated herein by reference thereto. Included within the entire specification of the WO 2009/02049 publication, the entirety of which is hereby incorporated by reference are the below-listed DNA and amino acid SEQ ID NOS. These include:

- WO 2009/02049 *SEQ ID NOS: 1, 2* (wild-type 'parent' *T. maritima* cellulase), disclosed herein as SEQ ID NOs:5 and 6.
- WO 2009/02049 *SEQ ID NOS: 3* (wild-type DNA, altered to remove alternate starts) disclosed herein as SEQ ID NO:7.
- WO 2009/02049 *SEQ ID NOS: 6, 7* ("7X" combined Gene Site Saturation Mutagenesis ("GSSM") mutations) disclosed herein as SEQ ID NOs:8 and 9.
- WO 2009/02049 *SEQ ID NOS: 8,* 9 ("12X-6" combined GSSM mutations), disclosed herein as SEQ ID NOs:3 and 2.
- WO 2009/02049 *SEQ ID NOS: 10, 11* ("13X-1" combined GSSM mutations) disclosed herein as SEQ ID NOs:10 and 11.
- WO 2009/02049 *SEQ ID NOS: 12, 13* ("12X-1" combined GSSM mutations) disclosed herein as SEQ ID NOs:12 and 13.
- WO 2009/02049 *SEQ ID NOS: 16, 17* (alternative cellulase breaker from *Thermotoga sp.*) disclosed herein as SEQ ID NOs:14 and 15.
- WO 2009/02049 *SEQ ID NOS: 18, 19* ("7X" codon-optimized version of *T. maritima* cellulase for maize expression) disclosed herein as SEQ ID NOs:16 and 17.
- WO 2009/02049 *SEQ ID NOS: 20, 21* ("12X-6" codon-optimized version of *T. maritima* cellulase for maize expression) disclosed herein as SEQ ID NOs:18 and 19.

[0096] Besides the above-listed nucleotide and amino acid sequences related to wild-type and evolved variants of the cellulase from *Thermotoga maritima* strain MSB8, the additional mutants listed in Table 2 and Example 5 (from WO 2009/02049, and excerpted or reproduced below) are also deemed useful as components of the compositions described herein and/or in the methods of making these compositions.

Enzyme Activity Assays (Example 5 from WO 2009/02049)

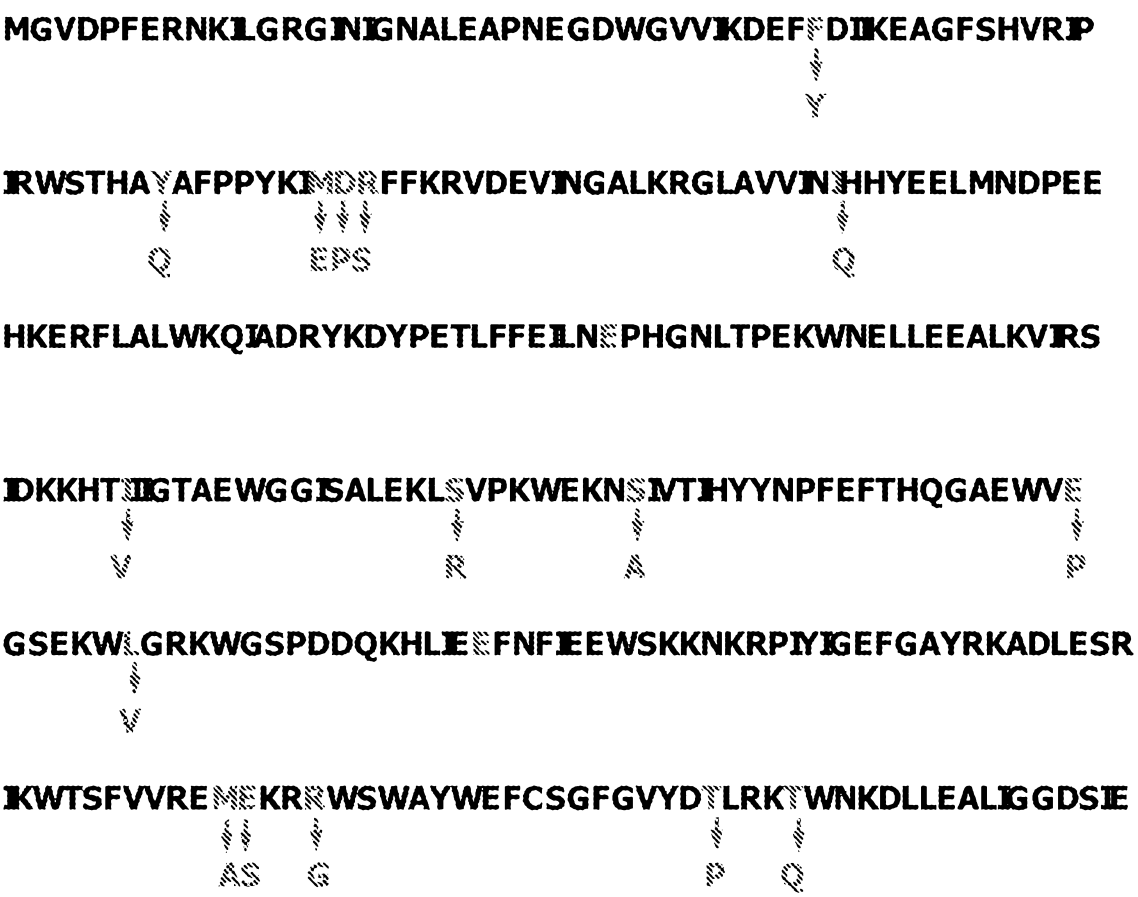
[0097] The following example describes exemplary enzymes, variants of the "parental" or "wild type" protein identified *in WO 2009/02049 as SEQ ID NO:2*, and data demonstrating their activity; the Figures and data are incorporated by reference.

[0098] Sequences are provided having specific residue changes to the "parent" (or "wild type") SEQ ID NO:6 (corresponding to *WO 2009/02049 SEQ ID NO:2*) encoded, e.g., by SEQ ID NO:5 (corresponding to *WO 2009/02049 SEQ ID NO:1*), as summarized (in part) in Table 1, above, and Tables 2 and 3, below.

Table 2

Position:	38	61	69	70	71	94	166	183	191	212	231	276	277	280	297	301
Mutation:	Y	Q	E	P	S	Q	V	R	A	P	V	A	S	G	P	Q
7X	Y	Q	E			Q		R	A			A				
10X-1	Y	Q	E			Q	V	R	A	P		A		G		
10X-2	Y	Q	E			Q	V		A	P		A		G	P	
11X-1	Y	Q	E			Q	V	R	A	P		A		G	P	
11X-2		Q	E			Q	V	R	A	P		A		G	P	Q
12X-1		Q	E		S	Q	V	R	A	P		A		G	P	Q
12X-2	Y	Q	E		S	Q		R	A	P		A		G	P	Q
12X-3		Q		P	S	Q	V	R	A	P		A		G	P	Q
12X-4	Y	Q			S	Q	V	R	A	P		A		G	P	Q
12X-5		Q	E	P		Q	V	R	A	P		A		G	P	Q
12X-6		Q	E	P	S		V	R	A	P		A		G	P	Q
12X-7		Q	E			Q	V	R	A	P	V	A		G	P	Q
13X-1	Y	Q	E		S		V	R	A	P	V	A		G	P	Q
13X-2	Y	Q	E	P	S	Q	V		A	P		A		G	P	Q
13X-3	Y	Q		P	S	Q	V	R	A	P		A		G	P	Q
13X-4	Y	Q	E	P		Q	V	R	A	P	V	A		G	P	
13X-5	Y	Q	E	P	S	Q	V	R	A			A		G	P	Q
13X-6	Y	Q	E	P		Q	V	R	A	P		A	S	G	P	
13X-7	Y	Q	E	P	S	Q		R	A	P		A		G	P	Q
14X	Y	Q	E	P	S	Q	V	R	A	P	V	A		G	P	

[0099] Table 3



[0100] Thermal tolerance of variants was measured using purified enzyme compared to the parental "wild-type" SEQ ID NO:6 (*WO 2009/02049 SEQ ID NO:2*), and a subset of the enzyme variants of Table 2 (the so called "7X variants"), as illustrated in Figure 9 and Figure 10; where the data illustrated therein demonstrate the thermal tolerance of the tested exemplary polypeptides (variants of the "parental" or "wild type" " SEQ ID NO:6, *WO 2009/02049 SEQ ID NO:2*) at 96° C through 100° C. In these figures, purified enzyme was heated for 30 minutes at the temperature indicated in the figures, and residual (thermotolerant) activity was measured at 37° C.

[0101] Heating temperatures between 84° C and 95° C activates the "thermal tolerant" variants (variants of the "parental" or "wild type" SEQ ID NO:2) slightly, resulting in having a residual (thermotolerant) activity of greater than the initial activity level (i.e., greater than 100%) (possibly due to improved folding upon cooling). As such, residual activity was normalized to 100%. Figure 9 illustrates a graphic summary of data from these thermal tolerance

studies for the enzymes of the invention identified as "10X-1", "12X- 1", "13X-1", "12X-6", "11X-1", "11X-2" and "7X", in addition to wild type; and Figure 10 is a "close-up" of part of Figure 9.

[0102] Thus, in one aspect, the disclosed cellulases are thermotolerant and/or thermostable; for example, the enzyme can retain at least 75 % residual activity (e.g., glucanase activity) after 2 minutes at 95° C; and in another aspect, retains 100 % activity after heating for 30 minutes at 95° C. In yet another aspect, the enzyme retains 100% activity after heating for 30 minutes at 96° C, 97° C, 98° C or 99° C. In yet another aspect, the disclosed cellulases retain at least 90% activity after heating for 30 minutes at 100° C.

[0103] In some embodiments, the enzymatic hydrolysis of pNP-β-D-lactopyranoside by the disclosed cellulase can be used as a measure of activity of the enzyme. The liberation of p-nitrophenol can be followed spectrophotometrically at 405nm. The increase in absorbance at 405nm can be converted to μmoles of p-nitrophenol by using a standard absorbance at those defined conditions. One unit of activity is defined as the quantity of enzyme required to liberate 0.42 μmole of p-nitrophenol from 2 mM pNP-β-D-lactopyranoside during one minute at pH 7.00 and 80°C.

[0104] An improved nucleotide sequence encoding the cellulase of SEQ ID NO:2 (corresponding to WO 2009/02049 SEQ ID NO:9) was disclosed in U.S. Provisional Application No. 61/618,610 filed on March 30, 2012; the entire disclosure of which is incorporated herein by reference thereto. This cellulase of SEQ ID NO:2 was evolved from the parent cellulase enzyme isolated from a DNA library originating from *Thermotoga maritima* strain MSB8. Enhancing expression of the disclosed cellulase involved 14 base changes with respect to the previously reported Open Reading Frame to generate SEQ ID NO:1 (below). These changes are silent as to the encoded protein. The 14 base changes are set forth below. "Position" indicates the number of the nucleotide within the Open Reading Frame, with the first nucleotide of the first codon numbered as 1. The mutation is specified using the notation (old nucleotide) (position) (new nucleotide). The mutations are as follows: T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C.

[0105] The base changes which distinguish SEQ ID NO:1 from prior reported sequences encoding the disclosed cellulase, collectively and individually, result in an Open

Reading Frame which leads to a higher level of protein expression than previously employed nucleotide sequences encoding the same protein.

[0106] SEQ ID NO:1 is used to direct an increased level of expression in a number of systems in which the disclosed cellulase protein (SEQ ID NO:2) may be expressed. SEQ ID NO:1 may be introduced into any number of expression systems to express the disclosed cellulase at an improved accumulation level. For example, SEQ ID NO:1 may be introduced, either on a plasmid or stably transformed into the genome of, for example, any number of gram negative bacterial systems such as *E. coli*, *Pseudomonas* species such as *fluorescens*, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Ralstonia* species, or *Caulobacter* species. Similarly, SEQ ID NO:1 may be introduced into any number of gram positive bacterial expression systems such as *Bacillus* species such as *Bacillus subtilis*, *Bacillus megaterium*, *Bacillus brevis*, Lactococcus specis such as *Lactococcus lactis*, *Lactobacillus* species, *Streptomyces* species such as *Streptomyces lividans*. Other gram negative, gram positive or unrelated eubacterial or archaeal expression systems may be used to express SEQ ID NO:1.

[0107] More specifically, SEQ ID NO:1 may be introduced into a plasmid to direct its expression. Plasmids which SEQ ID NO:1 may be introduced include, for example, E. coli expression vectors of the families pQE, pET, and pASK; *Pseudomonas* expression vectors of the families pCN51 LT8, RSF1010, pWZ112T, and pMYC; *Bacillus* expression vectors of the families pBAX, pHT01, and pHIS1525; *Streptomyces* expression vectors of the families pIJ6021 and pIJ2460; and *Lactococcus*: expression vectors of the families pNZ9530 and pNZ8148, for example. These examples are for demonstrative purposes and do not represent a complete set of vectors in which the polynucleotide sequence of SEQ ID NO:1 can be expressed.

Viscosifiers and other additives

[0108] In some embodiments, the polymeric viscosifier may be selected from one or more of guar, crosslinked guar, hydroxypropyl guar, carboxymethyl guar, carboxymethyl hydroxypropyl guar, carboxymethyl cellulose, barley β-glucan, and locust bean gum. In some embodiments, the hydrolytic pattern of the disclosed cellulases is both endo and exo, allowing effective reduction in the viscosity of polysaccharides, e.g., guar and derivatized guar solutions, by cleaving within long polysaccharide chains and also by cleaving disaccharide units from the ends of the polymers. Besides the aforementioned polysaccharides, other substrates of the

disclosed enzymes include those capable of forming linear or crosslinked gels (e.g., borate crosslinked guar). Examples of suitable polysaccharide substrates include galactomannan gums, guars, derivatized guars, cellulose and cellulose derivatives, starch, starch derivatizes, xanthan, derivatized xanthan and mixtures thereof. Specific examples also include, but are not limited to, guar gum, guar gum derivative, locust bean gum, karaya gum, xanthan gum, cellulose and cellulose derivatives, etc. Typical polymeric viscosifiers or gelling agents to which the disclosed enzymes may be directed include guar gum, hydroxypropyl guar, carboxymethyl hydroxypropyl guar, hydroxyethyl cellulose, carboxymethyl hydroxyethyl cellulose, carboxymethyl cellulose, dialkyl carboxymethyl cellulose, etc. Other examples of polymers include, but are not limited to, phosphomannons, scerolglucons, dextrans and other types of polymers. In some embodiments, a polymer substrate is carboxymethyl hydroxypropyl guar. In some embodiments, a disclosed enzyme may also be effective in hydrolyzing biogums (e.g., succinoglycan biogums made from date syrup or sucrose). In some embodiments, a disclosed enzyme may be used to hydrolyze cellulose-containing or derivatized cellulose-containing polymers – typically, the enzymes attack glucosidic linkages of the cellulose backbone. The disclosed enzymes may be suitable for degrading the polymer into mostly monosaccharide units, in some cases, by specifically hydrolyzing the exo(1,4)-β-D-glucosidic and endo(l,4)-β-D-glucosidic linkages between monosaccharide units and the cellulose backbone in the (1,4)-β-D-glucosidic linkages of any cellobiose fragments.

[0109] The fracturing fluid composition described herein may also comprise additional components, including agents to control fluid loss, reduce formation damage, adjust and/or buffer pH, control microbial growth and improve temperature stability. Crosslinkers may include borate, titanate, zirconate, and aluminum compounds. Typical hydraulic fracturing mixtures include about 98% water and sand and about 2% other additives, selected from acid, anti-bacterial agents, breaker (enzyme or chemical/oxidative), clay stabilizer, corrosion inhibitor, crosslinker, friction reducer, gelling agent (viscosifier), iron control, pH adjusting agent, scale inhibitor and surfactant.

Performance of disclosed cellulases in fracturing fluids

[0110] In each fracturing job that uses the disclosed cellulase, field operators will first perform an enzyme dose optimization study in an industrial lab. The disclosed cellulase will

be diluted to a concentration of 10~400 ppm and mixed with the polysaccharide gel forming compound, e.g., linear or cross-linked guar gum (25–60 lb/1,000 gal). Depending on the application conditions, guar gum or deriviatized guar is often cross-linked using a borate cross-linker, especially for wells where higher temperature, pressure, and pH conditions are present. The enzyme dose information resulting from such optimization study will be used in the actual fracturing job.

[0111] The unique activity profile of the disclosed cellulases allow for complete hydrolysis of guar-based fracturing fluids in a smooth and controlled manner in deep wells, where high temperature and high pH conditions are present. Compared to chemical breakers, the disclosed cellulase of the present invention provides a non-corrosive and environmentally benign alternative to the harsh and non-selective chemical breakers.

[0112] Enzyme breakers have been used for hydrolyzing polysaccharide viscosifiers, including e.g., guar gels at temperatures below 150° F, for many years. There is, however, an industry-wide demand for enzyme breakers that can function under higher temperature (e.g., 200-250° F) and extreme pH (≥10.5) conditions. To address this demand, the inventors have developed fracturing fluids comprising an exceptionally thermo-stable cellulase. See also PCT publication WO 2009/02049, which describes *inter alia* the protein and the DNA encoding the protein. The disclosed cellulases evolved from the parent enzyme have been shown to exhibit well differentiated performance under the extreme down-hole conditions encountered in gas shales and deeper oil/gas wells.

[0113] The disclosed cellulases can effectively break linear and borate cross-linked guar under broad ranges of temperature (80° F up to 250° F) and pH (3.0 up to 10.5). The results of rheological tests show that only a small dose is required (100 ppm or less) to achieve the complete break. The enzymatic reaction can be triggered by the changes of temperature and pH during fracturing operations. The disclosed cellulases also exhibit an excellent dose response that allows the operator to generate an ideal viscosity-time profile by adjusting enzyme dosage. The dose-dependent behavior will prove highly beneficial as it avoids premature viscosity loss and undesirable proppant screen-outs. Even in the presence of fluid additives, such as buffers, salts, stabilizers, crosslinkers, etc., the disclosed cellulases remain active for effective viscosity reduction.

[0114] The disclosed cellulase breakers reduce gel viscosity by specifically targeting β-1,4 glycosidic bonds between the mannose units in guar. Carbohydrate profiling tests demonstrated that this enzyme effectively and efficiently breaks the long guar polymers into small, soluble fragments that will eliminate gel re-healing. The conductivity tests demonstrate the complete hydrolysis of guar and removal of polymer residues that cause formation damage and reduce well conductivity.

[0115] As a biocatalyst for a guar breaking reaction, the disclosed cellulases, which exhibit mannanase activity, have several significant advantages over traditional oxidative chemical breakers. First, the enzyme specifically breaks long chain guar gum polymer through endo or exo –beta mannanase activity without undesirable reactions to the wellbore, formation or fracturing equipment. Second, the enzyme is not consumed in the reaction and can continue working on other guar polymers during their life time, thus providing an extended and controlled breaking profile. Third, the enzyme can break guar polymers into much smaller fragments, thus providing a more complete guar break with less residue. Consequently, efforts have been made by scientists to discover and characterize hyper-thermostable cellulases from extreme natural environments. *Thermotoga maritima sp.*, a hyperthermophilic bacterium identified from a hydrothermal vent sample, has a cellulase with endo-mannanase activity that can specifically cleave β-1,4 bond linked polysaccharides of long chain guar gum. Mathur EJ, Lam DE, "Carboxymethyl cellulase from *Thermotoga maritima*", U.S. Patent Nos. 5,962,258 (1999), 6,008,032 (1999) and 6,245,547 (2001). Specifically, this enzyme demonstrates a temperature optimum of 180° F, which is significantly higher than other known endo-mannnases in the glycoside hydrolase family. Pereira JH, Chen ZW, McAndrew RP, Sapra R, Chhabra SR, Sale KL, Simmons BA, Adams PD, "Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima*", Journal of Structure Biology, 2010, 172; 372-379; Wu TH, Huang CH, Ko TP, Lai HL, Ma YH, Chen CC, Cheng YS, Liu JR, Gup RT, "Diverse substrate recognition mechanism revealed by *Thermotoga maritima* Cel5A structures in complex with cellotetraose, cellobiose and mannotriose", Biochimica et Biophysica Acta, 2011, 1814; 1832-1840.

Thermostability of cellulase evaluated with Differential Scanning Calorimetry (DSC)

[0116] Enzymes have 3-dimensional structure which is critical for its catalytic function. Under certain conditions (heat, denaturant, extreme pH), enzymes can become unfolded and lose their catalytically active 3-dimensional structure. The melting temperature (Tm) is the temperature at which 50% of a protein becomes unfolded. This parameter is widely used for evaluation of thermostability of enzymes or proteins. Tm can be measured by differential scanning calorimetry (DSC).

[0117] DSC tests were carried out with cellulase wild-type protein and the selected cellulase variant. All protein samples were analyzed at a concentration of 1 mg/ml and a scan rate of 1 °C/min. The temperature range of each scan was 160~250° F. A constant pressure of 4.6 atm was maintained during all DSC experiments to prevent possible degassing of the solution on heating. Tm was calculated with the available software package. Figure 7 show the Tm results of wild-type and a selected cellulase variant. There is a 24° F increase of Tm value observed with the final cellulase candidate as compared to the wild-type enzyme. The Tm is pH dependent. The Tm difference between variant and wild-type is 24° F at pH 6.5 while the difference is 13° F at pH 10.5.

Temperature profiling of the cellulase variant using a chemical surrogate

[0118] The cellulase variant was further evaluated at various temperatures for its 1-4 linkage cleavage activity. The molecule pNP- β-D-lactopyranoside (Figure 8A) was used as a surrogate substrate. The enzymatic hydrolysis of pNP-β-D-Lactopyranoside by cellulase can produce free p-Nitrophenol, which can be measured spectrophotometrically at 405nm. Figure 8B exhibits the temperature profile of the cellulase variant up to 194° F, which is the temperature limit for the spectrophotometer. As shown in Figure 8B, this cellulase variant has an optimal temperature at 180° F.

[0119] In order to evaluate cellulase activity at above boiling temperature (>212° F), an assay was developed to quantitatively measure the residual activity of the cellulase variant. The enzyme was mixed with 25 lb guar per 1000 gal (ppt) at 1:1 ratio before heat challenged at 225° F, 250° F, and 275° F for 10, 20, and 30 minutes, respectively. Subsequently, activity of the heat challenged enzyme was measured using the pNP absorbance assay by the addition of 30 uL of heat treated enzyme to 3000 uL of 2 mM pNP-B-D-lactopyranoside, followed by measuring

405 absorbance for 10 minutes. Figure 8C shows the residual activity of the cellulase variant at 225° F, 250° F and 275° F. Robust cellulase activity was observed at 225° F when the enzyme was heat-treated for 10-30 minutes. When heat-treated cellulase was applied to borate crosslinked guar (25 lb per 1000 gal), complete gel breaks were observed at a small dose (20 gpt). The cellulase activity becomes compromised at 250° F and 275° F with relative residual activity at 11-14% of that at 225° F (Figure 8C), suggesting a small percentage of cellulase can still perform measurable catalytic reaction even after high temperature (>250° F) treatment.

pH profiling of the cellulase variant using a fluorescent surrogate

[0120] The molecule 4-Methylumbelliferyl β-D-cellobioside is a surrogate substrate that can be cleaved by cellulase at higher pH conditions (Figure 8A). The cleavage product, 4-methylumbelliferone can be quantified fluorometrically using excitation and emission wavelengths of 365 nm and 455 nm. Two millimoles of 4-Methylumbelliferyl β-D-cellobioside were dissolved in a buffer system with pH values ranging from 9.5 to 10.5. Cellulase activity at different pH conditions was measured kinetically in a spectrophotometer by addition of 30 uL cellulase into 3000 uL of 2 mM 4-Methylumbelliferyl β-D-cellobioside. The slope of each reaction curve was obtained by linear curve fitting. The relative activity was calculated according to results obtained at pH 9.5. At pH 10.5, a pH condition that is highly relevant to borate crosslinked guar fluids, cellulase exhibits reasonable activity (Figure 9).

Confirmation of Enzymatic breaking of β-1,4-linkage but not α-1,6-linkage polysaccharides by HPLC

[0121] Guar, a long-chain polysaccharide composed of mannose and galactose sugars, is the major viscosifier in water based fracturing operations. Polymannose forms a long chain backbone through β-1,4-linkage while galactose unit is attached to the mannose unit through α-1,6 linkage. The ratio of mannose to galactose sugars may be ranging from 1.6:1 to 1.8:1. Importantly, the polymannose backbone of guar is not soluble in water and galactose branches significantly increase water solubility. However, it has been reported that as few as 6 contiguous un-branched mannose units can form a local helical structure which is completely insoluble (11). Therefore, if cellulase breaks at the α-1,6 linkage of guar molecule, significant

amount of insoluble residue can be produced. As a result, the conductivity of the proppant pack can be impaired.

[0122] In order to confirm that enzymatic activity of the cellulase variant targets specifically at β-1, 4 glycosidic bonds of polysaccharide, a normal phase HPLC method was adopted for analysis of two surrogate substrates, 1,4-β-D-Mannopentaose and 63,64-α-D-Galactosyl-mannopentaose (Figure 10A). The cellulase variant at 3 units/mL was incubated with 1 mg/mL substrates for one hour at 80°C at pH 7.0. Using an amide column, actonitrile/H2O/triethylamine as a mobile phase, and an ELSD detector, the baseline separation for mannan oligosaccharides by HPLC was obtained. Figure 10B illustrates HPLC elution profiles of different mannan oligosaccarides including M1, M2, M3, M4, M6, and galactose standards (M1-M6 indicates mannan oligosaccharide chain length). Figure 4C shows an example of how β-1, 4-D-Mannopentaose, a mannan oligosaccharide with only β-1, 4 linkage in its structure, was digested by the cellulase variant. HPLC profiles demonstrate β-1,4-linkage cleavage with preferable production of M3 oligosaccarides along with production of M1, M2 and M4, confirming that the cellulase variant specifically attacks β-1, 4 linkage when breaking down the mannan oligosaccharides (Figure 10C(a)-(b)). On the other hand, for 63,64-α-D- Galactosyl-mannopentaose, a compound containing both a β-1,4-linkage and α-1,6-linkage, the HPLC profiles show that cellulase variant only breaks the β-1-4 linkages and produces an OGGM4 peak and a single unit mannose peak (Figure 10C(a)). Using galactose as a reference, no peak was observed for the single unit galactose from enzyme treated 6^3,6^4-α-D-Galactosyl-mannopentaose, confirming that the α-1, 6 linkages remain intact after mananase treatment (Figure 10C(b)).

Preparation of crosslinked guar fluids with breakers

[0123] Linear guar fluids were prepared by adding the dry polymer to water that was stirred in a blender sufficiently to generate a vortex. Stirring was continued for about 30 minutes for complete hydration. The pH was adjusted with a solution of sodium hydroxide. For crosslinked fluids, clay stabilizer and cleanup surfactant, when used, were added after full hydration of the guar. Next, caustic was added, followed by the enzyme addition. The crosslinkers were added last. The borate crosslinker was delayed in action by slow dissolution of

the active boron species. The zirconium crosslinker was also delayed due to the presence of sesquicarbonate buffer used at 6 ppt.

Bottle testing

[0124] Jars of guar or crosslinked guar were prepared with either ammonium persulfate (APS) or cellulase breakers. The jars were aged in a water bath and the level of breaking was visually assessed by tilting the bottles and observing the fluid movement. The bottle tests allowed a quick method for evaluating the effect of concentration, pH and temperature on breaking of a given fracturing fluid. Because the tests are qualitative in nature, the results are not shown.

Rheology

[0125] The viscometer measures viscosity using a cup and bob technique (Rotor 1 and Bob 5) with model 50 specifications under a nitrogen pressure of 400 psi. Fluid temperature was increased to the desired value by an oil bath. Heating takes about 20 minutes to reach the desired temperature and that temperature was maintained for the test duration. Continuous measurement of viscosity occurs at a rotational speed that delivers a wall shear rate of 100 s-1 with periodic shear ramps at 100, 75, 50, 25, 50, 75 and 100 s-1. The shear ramp data is used to calculate power law parameters (n' and K') for prediction of viscosity at lower shear rates that are commonly found in a fracture.

[0126] **Figure 11** demonstrates a range of activity in breaking 80 lb per 1000 gallon linear guar solution at 180°F. The concentration of cellulase was kept constant at 200 ppm. As pH increased from 9 to 11, a noticeable delay in breaking was observed, consistent with a lower activity for the enzyme at elevated levels of pH. Even at a pH of 11, the solution shows a slow breaking effect compared to the control that lacks enzyme.

[0127] **Figure 12** shows the effect of the enzyme concentration on breaking for a borate-crosslinked guar fluid at 200° F and pH of 10.5. Increasing dosages of cellulase result in more significant breaking of viscosity. The rheology results were run for several hours to show the continual loss of viscosity with time. Clearly for crosslinked gels, more enzyme is needed versus the linear gels containing a higher concentration of guar. The viscosity of crosslinked

fluid with a level of 50 ppm cellulase was indistinguishable from that of the crosslinked fluid control whereas higher levels of enzyme did show breaking activity.

[0128] Importantly, the enzyme is still active at 225° F and pH of 10 when thermal stabilizer is included at a level of 10 ppt (**Figure 13**). The control shows a gradual thermal deterioration but the enzyme clearly accelerates the degradation of viscosity. **Figure 14** presents comparison data of cellulase and encapsulated ammonium persulfate at 225° F and pH of 10 for a zirconium-crosslinked fluid using carboxymethylhydroxyethylcellulose (CMHPG). A concentration of 100 ppm provides a similar break profile to ammonium persulfate while 200 ppm results in immediate loss of viscosity. Note that the encapsulated ammonium persulfate releases some of the active material via thermal degradation of and/or diffusion through the encapsulating material which results in delayed breaking.

Residue analysis

[0129] Jars of guar or crosslinked guar were prepared with either ammonium persulfate or cellulase breakers. The jars were aged in a water bath overnight at 180°F. The contents were centrifuged at 3000 rpm for 5 minutes and vacuum filtered through a nominal 5 micron, weighed paper. The paper was dried overnight at 110° F and reweighed to calculate the amount of residue recovered. This value is expressed as a percentage of the original weight of polymer in the solution. The filter paper plus residue was then dried for another 16 hours and reweighed to ensure the moisture had been removed. However, the data suggest that the additional 16 hours of drying time is seen to be unnecessary. The fluids tested included 80 ppt of linear guar at pH of 10. Cellulase was tested at 200 and 400 ppm while ammonium persulfate was used at 1 and 5 ppt.

[0130] The results show that surprisingly the solutions containing cellulase had lower amounts of residue than the corresponding solutions with ammonium persulfate (**Figure 15**). In contrast, the oxidative breakers appear to cleave more randomly and result in a higher level of insoluble fragments that are captured as residue. If the jars were left for a longer time period at temperature (i.e., 180° F), it is possible that the residue will decrease even further for the enzyme treated fluids but not for the fluids containing oxidative breaker. For ammonium persulfate, the residue is sensitive to breaker concentration, whereas the sensitivity to enzyme concentration is much lower.

Conductivity study

[0131] Fracture conductivity was measured using a modified API cell with a fixed bottom piston. Conditions were 180° F, 3000 psi closure stress, 2 lbm/ft2 proppant loading (20/40 mesh Ottawa sand) and about a 5X concentration of fracturing fluid following leakoff. The fluid was left at temperature overnight before four hours of cleanup followed by a minimum of one hour of pack flow. The cleanup alternates flow through the pack and through the core each hour using 2 wt% KCl solution. Measurement of final conductivity, stable within 4% for one hour, was performed using 2 wt% KCl at a rate of 3 mL/min. Retained conductivity was calculated by comparison with the conductivity measured for the same conditions using 2 wt% KCl in place of the fracturing fluid.

[0132] Linear guar and boron-crosslinked guar with ammonium persulfate or cellulase were run as well as one test with crosslinked guar using both breakers. For linear guar, higher levels of retained conductivity were achieved with cellulase than with ammonium persulfate. Similarly, encapsulated ammonium persulfate provided a lower retained conductivity than did 200 ppm of cellulase for crosslinked-borate fluid (**Figure 16**). The combination of cellulase and encapsulated persulfate had even higher retained conductivity than either one of the breakers alone. However, the highest result was obtained with a 400 ppm dosage of the enzyme. Since the conductivity test is shut-in overnight, the cellulase is allowed extra time for reducing molecular weight of the guar. This will directly result in enhanced conductivity for the proppant pack.

[0133] In summary, the evolved cellulase variant is compatible with various additives in borate crosslinked guar gel fluids. Particularly, this cellulase variant shows robust guar gel breaking profiles in the temperature range of 180-225° F and a pH range up to 10.5. The pH limitation prevents testing this enzyme with borate crosslinked guar fluids in higher temperature range. However, positive guar breaking profiles were observed with zirconium crosslinked CMHPG at temperature of 225° F. Notably, residue analysis demonstrates that mannase treated fluids produce significantly less residue than ammonium persulfate treated fluids. More importantly, cellulase treatment achieved significantly higher retained conductivity than did ammonium persulfate treatment, confirming the notion that enzymes can provide a more complete guar break and improved well conditions relative to chemical oxidizers. Moreover, use

of cellulase is compatible with addition of ammonium persulfate should that be useful in certain oilfield applications.

DESCRIPTION OF THE FIGURES

[0134] Figure 1 is an image of SDS PAGE gel electrophoresis displaying various level of protein expression, as described in Example 2.

[0135] Figure 2 is a bar graph showing the level of activity of protein preparations, as described in Example 3.

[0136] Figure 3 is **SEQ ID NO:1**, the polynucleotide with 14 silent mutations: T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, as compared to SEQ ID NO:3.

[0137] Figure 4 is **SEQ ID NO:2**, the polypeptide encoded by **SEQ ID NO:1, 3, and 4**.

[0138] Figure 5 is **SEQ ID NO:3**, the unmodified parent polynucleotide sequence of SEQ ID NO:1 and 4.

[0139] Figure 6 is **SEQ ID NO:4**, the polynucleotide with 14 silent mutations: T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, as compared to SEQ ID NO:3, plus one additional point mutation upstream from the start codon (additional upstream sequence shown).

[0140] **Figure 7.** DSC tests of the wild type cellulase and the selected cellulase variant were performed with calculated Tm value labeled. A 24 °F increase of Tm was observed with the cellulase variant as compared to the wild type cellulase.

[0141] **Figure 8A.** Chemical structure of pNP-ß-D-lactopyranoside.

[0142] **Figure 8B**. Temperature profiling of cellulase measured by an pNP absorbance assay (405 nm) at different temperature (from 70° F to 200° F). The activity was calculated as percentage of activity at 180° F. The activity at above boiling temperature (>212° F) can be further evaluated with residual activity assay (**Figure 8C**).

[0143] **Figure 8C**. Residual activities of the cellulase variant were evaluated using heat challenged samples. The relative activity was calculated as percentage of activity obtained at 225° F.

[0144] **Figure 9**. pH profiling of the cellulase variant using a fluorescent assay. Relative activity was calculated by the percentage of activity at pH 9.5.

[0145] **Figure 10A**. Structures of two compounds for HPLC study. 1,4-ß-D-Mannopentaose is a compound with 5 units of mannose connected through 1,4-ß linkage. 6^3,6^4-α-D-Galactosyl-Mannopentaose is a compound mimic for guar gum structure with two galactose units connected through α-1,6 linkages to 1,4-ß-D-Mannopentanose, which allows analysis of hydrolytic products from cellulase treatment.

[0146] **Figure 10B**. HPLC analysis of oligosaccharides standards (M1-M6, and galactose). Note the mannose single unit (M1, red peak) is well separated from single unit galactose (blue peak) in the elution profile of HPLC.

[0146] **Figure 10C**. HPLC analysis of hydrolytic products from 1,4-ß-D-Mannopentaose and 6^3,6^4-α-D-Galactosyl-Mannopentaose after mannose treatment.

a) enzymatic digestion of 1,4-ß-D-Mannopentaose. HPLC elution profile of control sample with no enzyme treatment is shown in red while cellulase variant treated 1,4-ß-D-Mannopentaose is shown in blue.

b) enzymatic digestion of 63,64-ÿ-D-Galactosyl-Mannopentaose. O-GGM5 represents 6^3,6^4-α-D-Galactosyl-Mannopentaose with cellulase treated sample shown in red and no enzyme sample shown in blue. Single unit mannose (shown in pink) and galactose (shown in green) are included as references. Note there is not a galactose peak but only a mannose peak in the HPLC elution profile of cellulase treated O-GGM5, confirming that the cellulase variant specifically cleaves ß-1,4 glycosidic bonds between the mannose units.

[0147] **Figure 11**. Effects of pH on viscosity of 80 lbm guar/1000 gal US at 180° F. Shear ramps have been removed for clarity.

[0148] **Figure 12**. Rheology study of the cellulase variant on borate crosslinked guar at 200° F with pH = 10.5. All fluids contained a clay stabilizer and surfactant and used a delayed crosslinker.

[0149] **Figure 13**. Rheology study of the cellulase variant on borate crosslinked guar at 225° F with pH = 10. All fluids contained 10 lbm thermal stabilizer/1000 gal US and used a delayed crosslinker.

[0150] **Figure 14**. Rheology study of the cellulase variant with a zirconium crosslinked CMHPG at 225° F and pH = 9.9 with 10 lbm thermal stabilizer/1000 gal US.

[0151] **Figure 15**. Residue analysis comparing ammonium, persulfate with cellulase. Higher breaker levels reduce the residue in both cases, but the oxidative breaker shows higher amounts of residue.

[0152] **Figure 16**. Comparison of conductivity results for cellulase versus ammonium persulfate for both linear and boron-crosslinked guar formulations. Linear: 80 lbm guar/1000 gal US with cellulase or live ammonium persulfate; Crosslinked: 30 lbm guar/1000 gal US with cellulase or encapsulated ammonium persulfate.

Definition of Terms:

[0153] "cellulase" are enzymes having cellulase, endoglucanase, cellobiohydrolase, beta- glucosidase, xylanase, mannanase, β-xylosidase, arabinofuranosidase, and/or oligomerase activity.

[0154] "cellulolytic activity" is an enzyme having cellulase, endoglucanase, cellobiohydrolase, beta- glucosidase, xylanase, mannanase, β-xylosidase, arabinofuranosidase, and/or oligomerase activity.

[0155] A "codon" is a three polynucleotide sequence that specifies the identity of an amino acid to be added to a protein.

[0156] A "silent mutation" is a mutation in a codon that does not result in the specification of a different amino acid.

[0157] An "Open Reading Frame" is a series of codons that specifies the sequence of amino acids in a protein.

[0158] A base "position" is the numerical location of a base in a polynucleotide sequence, counted consecutively from the start of the open reading frame or from some other reference marker.

[0159] To "encode" a protein means to specify the amino acid sequence of that protein.

[0160] A "mutation" is a change in a nucleotide sequence or an amino acid sequence compared to a reference.

[0161] A "nucleotide" refers to one of the four bases which comprise DNA sequence – Adenine (A), Thymidine (T), Guanidine (G), and Cytosine (C).

[0162] *Thermotoga maritima* genomic sequence" refers to the *Thermotoga maritima* strain MSB8 genomic sequence specified by GenBank Accession No. AE000512.

[0163] An "Expression level" for a given protein is the amount of protein generated by an expression system, such as a transformed cell culture as measured per unit volume of cell culture.

[0164] An "Expression level" for a given enzyme is the amount of enzyme activity generated by an expression system, such as a transformed cell culture as measured per unit volume of cell culture.

[0165] "Wild-type" refers to a protein or nucleic acid sequence that can be obtained in nature.

Example 1-Method of making enhanced expression variants.

[0166] Two variants (SEQ ID NO:1 and NO: 4) were designed based on SEQ ID NO:3 to mutate at the DNA level to improve the gene expression. The design takes into account of many factors that may influence gene expression. The mutations were introduced on the PCR primers using PCR techniques known ot those of skill in the art. Both genes were PCR-amplified and cloned into the *Pseudomonas* vector pDOW1169 (DOW AgroSciences, IN) using standard molecular cloning techniques. The resulting expression constructs were transformed into *Pseudomonas fluorescens* DC454 (DOW AgroSciences, IN). A transformant with the SEQ ID NO:1 was designated as the lead as it showed the most enhanced expression.

Example 2 - Using SDS-PAGE gel electrophoresis and nonspecific protein staining to visualize expression levels of the SEQ ID NO:2 polypeptide expressed by constructs comprising SEQ ID NOs: 1, 3, and 4.

[0167] Criterion™ precast Tris-HCl polyacrylamide gel (Bio-rad Laboratories, Inc.) was used to separate proteins. The gel was run at 150V using Tris-glycine buffer (see Figure 1). Protein loading was normalized to load proteins from 0.33 OD_{600} cells for each lane. SeeBlue® pre-stained protein standard was used (Life Technologies). The gel was stained with a nonspecific dye, and each lane was visually inspected for the presence of a band at the size of SEQ ID NO:2, about 37 kilodaltons.

[0168] The results indicate that there is a single band having an accumulation level which varies across samples and which is absent from the negative control. This band has a size expected for SEQ ID NO:2.

[0169] The accumulation level of this band is significantly higher in lanes corresponding to protein extracts from cells harboring constructs comprising SEQ ID NO:1, and to a lesser extend SEQ ID NO:4, that SEQ ID NO:3 or the negative control.

Example 3 - Method of determining relative expression levels for variants.

[0170] Nucleic acid sequence comprising SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:4 gene were transformed into asuitable host cell for expression of the protein of SEQ ID NO:2. The cells were cultured in flasks so that the encoded protein would be expressed. The cultures were grown at 30°C and 220 rpm to an OD600 of ~0.9 in a designed complex medium, and induced with 0.3mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) for 24 hours. Cells were harvested and lysed either by sonication or heat-treatment at 80°C for 1 hour. Cellulase activity was measured by a p-Nitrophenyl (pNP) based assay using pNP-β-D-lactopyroanoside as substrate. (Advances in Carbohydrate Chemistry and Biochemistry, Academic Press, 1999). Activity levels were measured in U/ml as shown in Figure 2 to determine relative expression levels from each culture.

[0171] The results indicate that cells harboring the construct comprising SEQ ID NO:1 demonstrated significantly more SEQ ID NO:2 activity than those harboring SEQ ID NO:4, and that both SEQ ID NOs:1 and 4 yeilded a greater amount of activity of the expressed protein than the cells harboring SEQ ID NO:3.

WHAT IS CLAIMED IS:

1. A nucleotide sequence of SEQ ID NO:1.

2. The nucleotide sequence of claim 1, wherein said sequence encodes a protein.

3. A nucleotide sequence encoding a cellulase derived from *Thermotoga maritima* comprising at least one mutation selected from T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, A6C, G6C, A9C, G9C, A15G, C15G, T22C, G22C, A24T, C24T, T33C, G33C, T39C, G39C, T40C, G40C, T42C, G42C, T54C, G54C, T57C, G57C, A66C, G66C, C81A, T81A, T84C, G84C, or any combination thereof.

4. A nucleotide sequence of SEQ ID NO:3 comprising at least one mutation selected from T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, A6C, G6C, A9C, G9C, A15G, C15G, T22C, G22C, A24T, C24T, T33C, G33C, T39C, G39C, T40C, G40C, T42C, G42C, T54C, G54C, T57C, G57C, A66C, G66C, C81A, T81A, T84C, G84C, or any combination thereof.

5. The nucleotide sequence of claim 3 or 4 wherein at least one mutation is silent.

6. The nucleotide sequence of claim 3 or 4 wherein at least one mutation results in the nucleotide sequence harboring at least one mutation directing expression of said cellulase at a higher level than a nucleotide sequence lacking at least one mutation and not otherwise differing from the nucleotide sequence of claim 3 or 4.

7. A nucleotide sequence from *Thermotoga maritima* having at least one mutation which increases the expression level of a protein encoded by said nucleotide sequence compared to a *Thermotoga maritima* genomic sequence.

8. The nucleotide sequence of claim 7, wherein said mutation is silent.

9. A first nucleotide sequence encoding the polypeptide of SEQ ID NO:2 wherein said nucleotide sequence has been mutated with respect to a second sequence encoding SEQ ID NO:2 such that the expression level of said protein is increased relative to that of said protein encoded by said second nucleotide sequence.

10. A nucleotide sequence encoding a protein at least 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or 100% identical to SEQ ID NO:2, or a fragment thereof, wherein said nucleotide sequence comprises at least one mutation selected from T6C, T9C, T15G, A22C, G24T, A33C, A39C, A40C, A42C, A54C, A57C, T66C, G81A, A84C, A6C, G6C, A9C, G9C, A15G, C15G, T22C, G22C, A24T, C24T, T33C, G33C, T39C, G39C, T40C, G40C, T42C,

G42C, T54C, G54C, T57C, G57C, A66C, G66C, C81A, T81A, T84C, G84C, or any combination thereof.

11. The nucleotide sequence of claim 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, that encodes a polypeptide having a cellulase activity, wherein the polypeptide is produced in a recombinant bacteria expression system.

12. The nucleotide sequence of claim 11, wherein the bacteria expression system is a gram-negative bacteria expression system.

13. The gram-negative bacteria expression system of claim 12, wherein the gram-negative bacteria is a *Pseudomonas*, *E. coli*, *Ralstonia*, or *Caulobacter* expression system.

14. The gram-negative bacteria expression system of claim 12, wherein the *Pseudomonas* expression system is a *Pseudomonas fluorescens* expression system.

15. The expression system of claim 12 wherein the cellulase is produced at least 1.0 g/L, 2.0 g/L, 3.0 g/L, 4.0 g/L, 5.0 g/L, 6.0 g/L, 7.0 g/L, 8.0 g/L, 9.0 g/L, 10.0 g/L, 11.0 g/L, 12.0 g/L, 13.0 g/L, 14.0 g/L, 15.0 g/L, 16.0 g/L, 17.0 g/L, 18.0 g/L, 19.0 g/L, 20.0 g/L, 21.0 g/L, 22.0 g/L, 23.0 g/L, 24.0 g/L, 25.0, g/L, 26.0 g/L, 27.0 g/L, 28.0 g/L, 29.0 g/L, 30.0 g/L, 31.0 g/L, 32.0 g/L, 33.0 g/L, 34.0 g/L, or 35.0 g/L.

16. The polypeptide of SEQ ID NO:2,, wherein the amino acid sequence does not comprise a signal sequence, a proprotein sequence, a promoter sequence, or any combination thereof.

17. The polypeptide of SEQ ID NO:2 , wherein the sequence further comprises a heterologous sequence.

18. The polypeptide of claim 17, wherein said heterologous sequence is selected from the group consisting of: a signal sequence, a tag, an epitope, a promoter sequence, an N-terminal extension, a C-terminal extension, and any combination thereof.

19. A composition comprising a polypeptide of SEQ ID NO:2 and further comprising a second enzyme.

20. The composition of claim 19, wherein the second enzyme is selected from the group consisting of: a lactase, a lipase, a protease, a catalase, a xylanase, a cellulase, a glucanase, a mannanase, an amylase, an amidase, an epoxide hydrolase, an esterase, phospholipase, transaminase, an amine oxidase, cellobiohydrolase, an ammonia lyase, or any combination thereof.

21. An isolated, recombinant, or synthetic nucleotide having a nucleic acid sequence comprising SEQ ID NO:1, wherein the nucleic acid sequence encodes a polypeptide having a cellulase activity.

22. An isolated, recombinant, or synthetic nucleotide comprising a nucleic acid sequence of SEQ ID NO:1, wherein the nucleic acid sequence encodes a polypeptide having a cellulase activity and the polypeptide comprises an amino acid sequence of SEQ ID NO:2, or an enzymatically active fragment thereof.

23. An isolated, recombinant, or synthetic nucleic acid sequence comprising SEQ ID NO:1 that encodes a polypeptide having a cellulase activity, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO:2 and the polypeptide is produced in a recombinant *Pseudomonas fluorescens* expression system.

24. A composition comprising a polymeric viscosifier, a surfactant, a thermostabilizer, and an enzyme breaker comprising a wild-type cellulase derived from a hyperthermophilic bacterium or a mutated variant thereof.

25. The composition of Claim 24, wherein the viscosifier is a guar gel comprising a linear guar, a crosslinked guar, or mixtures thereof.

26. The composition of Claim 25, wherein the enzyme breaker specifically hydrolyzes ß-1,4 glycosidic bonds in the guar gel.

27. The composition of Claim 25, wherein the enzyme breaker does not specifically hydrolyze α-1,6 glycosidic bonds in the guar gel.

28. The composition of any of Claims 24-27, wherein the enzyme breaker retains its ability to hydrolyze ß-1,4 glycosidic bonds in the guar gel at temperatures up to about 275° F.

29. The composition of any of Claims 24-28, wherein the enzyme breaker retains its ability to hydrolyze ß-1,4 glycosidic bonds in the guar gel at a pH of up to about 11.

30. The composition of any of Claims 24-29, wherein the enzyme breaker is a mutated variant comprising 12 mutations relative to the wild type cellulase.

31. The composition of Claim 30, wherein the 12 mutations are selected from the group consisting of F38Y, Y61Q, M69E, D70P, R71S, I94Q, I166V, S183R, S191A, E212P, L231V, M276A, E277S, R280G, T297P and T301Q.

32. The composition of Claim 24, wherein the enzyme breaker has SEQ ID. NO. 2.

33. The composition of Claim 24, wherein the enzyme breaker is encoded by a polynucleotide having SEQ ID. NO. 1.

34. The composition of any of Claims 24-29, wherein the enzyme breaker is a mutated variant of the wild-type cellulase, and has a melting temperature that is at least 20° F greater than the melting temperature of the wild type cellulase at about pH 6.5 and at least 10° F greater than the melting temperature of the wild type cellulase at about pH 10.5.

35. The composition of Claim 24, wherein said enzyme breaker is a mutant variant of a cellulase encoded by SEQ ID NO:3.

36. A method of reducing the viscosity of a polysaccharide gel comprising ß-1,4 glycosidic bonds, the method comprising contacting the polysaccharide gel with a cellulase variant under permissive conditions and for a period of time sufficient to allow the cellulase to hydrolyze the ß-1,4 glycosidic bonds in the polysaccharide gel thereby reducing the viscosity of the guar gel, wherein the cellulase variant comprises at least 12 mutations compared to a wild-type cellulase derived from a hyperthermophilic bacterium, and wherein the cellulase variant exhibits increased temperature and pH tolerance compared to the wild-type cellulase.

37. The method of Claim 36, wherein at least 12 mutations were generated by Gene Site Saturation Mutagenesis comprising repeated cycles of reductive reassortment, recombination, and selection.

38. The method of Claim 36, wherein the polysaccharide gel comprises a linear guar, a crosslinked guar, or mixtures thereof.

39. The method of Claim 36, wherein the cellulase variant has SEQ ID NO:2.

40. The method of Claim 36, wherein the cellulase variant is encoded by a polynucleotide having SEQ ID. NO. 1.

41. The method of Claim 36, wherein permissive conditions comprise a temperature range of from about 180° F to about 275° F.

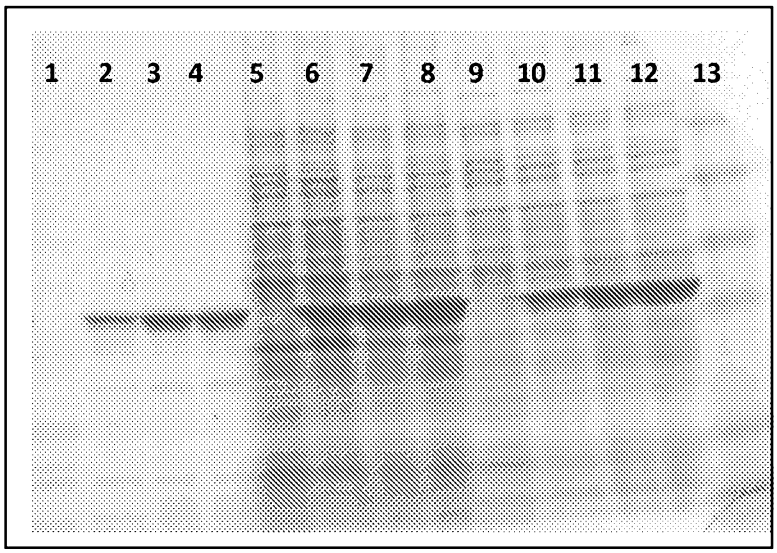
42. The method of Claim 36, wherein permissive conditions comprise a pH range of from about 9 to about 11.

43. A method of treating a subterranean formation, comprising:
introducing into the subterranean formation a fracturing fluid that comprises a polysaccharide gel and a cellulase variant; and allowing the gel and the cellulase variant to react under permissive conditions and for a period of time sufficient to allow the cellulase variant to hydrolyze at least

some of the ß-1,4 glycosidic bonds, but not the α-1,6 glycosidic bonds, in the gel, thereby generating a reduced viscosity fracturing fluid; wherein the cellulase variant comprises at least 12 mutations compared to a parental wild-type cellulase derived from a hyperthermophilic bacterium, and wherein the cellulase variant exhibits increased temperature and pH tolerance compared to the wild-type cellulase.

44. The method of Claim 36, wherein the reduced viscosity fracturing fluid generated by the cellulase variant comprises less residue that a comparable reduced viscosity fracturing fluid generated by a chemical breaker, wherein the dosage of cellulase variant and chemical breaker provide substantially the same reduction in viscosity.

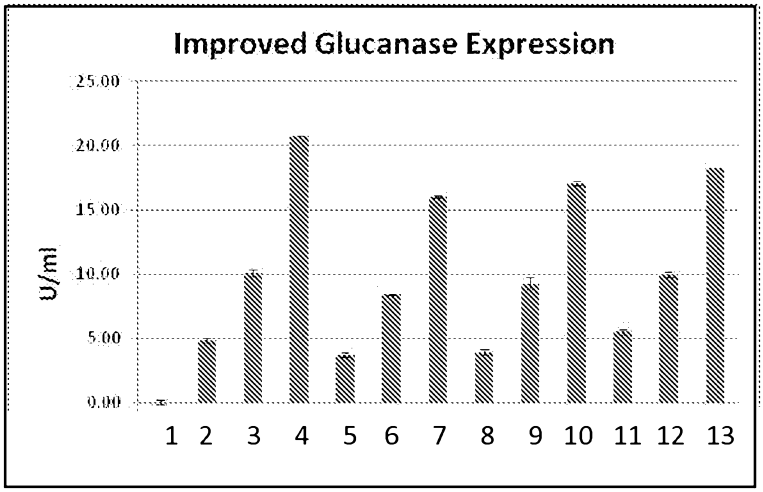
Figure 1



Legend

Lanes 1, 5, 9 = negative control
Lanes 2, 6, 10 = SEQ ID NO:3
Lane 3, 7, 11 = SEQ ID NO:4
Lane 4, 8, 12 = SEQ ID NO:1
Lane 13 = Molecular Weight Ladder

Figure 2



Legend

Lanes 1 = negative control
Lanes 2, 5, 8, 11 = SEQ ID NO:3
Lane 3, 6, 9, 12 = SEQ ID NO:4
Lane 4, 7, 10, 13 = SEQ ID NO:1

Figure 3

SEQ ID NO:1

ATGGGCGTCGATCCGTTTGAACGTAACAAAATCTTGGGCCGCGGCATTAATATCGGC
AATGCGCTCGAAGCACCAAATGAAGGCGACTGGGGAGTGGTGATAAAAGATGAGTTC
TTCGACATTATAAAAGAAGCCGGTTTCTCTCATGTTCGAATTCCAATAAGATGGAGT
ACGCACGCTCAGGCGTTTCCTCCTTATAAAATCGAGCCTTCTTTCTTCAAAAGAGTG
GATGAAGTGATAAACGGAGCCCTGAAAAGAGGACTGGCTGTTGTTATAAATATTCAT
CACTACGAGGAGTTAATGAATGATCCAGAAGAACACAAGGAAAGATTTCTTGCTCTT
TGGAAACAAATTGCTGATCGTTATAAAGACTATCCCGAAACTCTATTTTTTGAAATT
CTGAATGAACCTCACGGAAATCTTACTCCGGAAAAATGGAATGAACTGCTTGAGGAA
GCTCTAAAAGTTATAAGATCAATTGACAAAAAGCACACTGTGATTATAGGCACAGCT
GAATGGGGGGGTATATCTGCCCTTGAAAAACTGAGGGTCCCAAAATGGGAAAAAAAT
GCGATAGTTACAATTCACTACTACAATCCTTTCGAATTTACCCATCAAGGAGCTGAG
TGGGTGCCTGGATCTGAGAAATGGTTGGGAAGAAAGTGGGGATCTCCAGATGATCAG
AAACATTTGATAGAAGAATTCAATTTTATAGAAGAATGGTCAAAAAAGAACAAAAGA
CCAATTTACATAGGTGAGTTTGGTGCCTACAGAAAAGCTGACCTTGAATCAAGAATA
AAATGGACCTCCTTTGTCGTTCGCGAAGCCGAGAAAAGGGGGTGGAGCTGGGCATAC
TGGGAATTTTGTTCCGGTTTTGGTGTTTATGATCCTCTGAGAAAACAGTGGAATAAA
GATCTTTTAGAAGCTTTAATAGGAGGAGATAGCATTGAATGA

Figure 4

SEQ ID NO:2

MGVDPFERNKILGRGINIGNALEAPNEGDWGVVIKDEFFDIIKEAGFSHVRIPIRWS
THAQAFPPYKIEPSFFKRVDEVINGALKRGLAVVINIHHYEELMNDPEEHKERFLAL
WKQIADRYKDYPETLFFEILNEPHGNLTPEKWNELLEEALKVIRSIDKKHTVIIGTA
EWGGISALEKLRVPKWEKNAIVTIHYYNPFEFTHQGAEWVPGSEKWLGRKWGSPDDQ
KHLIEEFNFIEEWSKKNKRPIYIGEFGAYRKADLESRIKWTSFVVREAEKRGWSWAY
WEFCSGFGVYDPLRKQWNKDLLEALIGGDSIE

Figure 5

SEQ ID NO:3

ATGGGTGTTGATCCTTTTGAAAGGAACAAAATATTGGGAAGAGGCATTAATATAGGAAATGCGC
TTGAAGCACCAAATGAGGGAGACTGGGGAGTGGTGATAAAAGATGAGTTCTTCGACATTATAAA
AGAAGCCGGTTTCTCTCATGTTCGAATTCCAATAAGATGGAGTACGCACGCTCAGGCGTTTCCT
CCTTATAAAATCGAGCCTTCTTTCTTCAAAAGAGTGGATGAAGTGATAAACGGAGCCCTGAAAA
GAGGACTGGCTGTTGTTATAAATATTCATCACTACGAGGAGTTAATGAATGATCCAGAAGAACA
CAAGGAAAGATTTCTTGCTCTTTGGAAACAAATTGCTGATCGTTATAAAGACTATCCCGAAACT
CTATTTTTTGAAATTCTGAATGAACCTCACGGAAATCTTACTCCGGAAAAATGGAATGAACTGC
TTGAGGAAGCTCTAAAAGTTATAAGATCAATTGACAAAAAGCACACTGTGATTATAGGCACAGC
TGAATGGGGGGGTATATCTGCCCTTGAAAAACTGAGGGTCCCAAAATGGGAAAAAAATGCGATA
GTTACAATTCACTACTACAATCCTTTCGAATTTACCCATCAAGGAGCTGAGTGGGTGCCTGGAT
CTGAGAAATGGTTGGGAAGAAAGTGGGGATCTCCAGATGATCAGAAACATTTGATAGAAGAATT
CAATTTTATAGAAGAATGGTCAAAAAAGAACAAAAGACCAATTTACATAGGTGAGTTTGGTGCC
TACAGAAAAGCTGACCTTGAATCAAGAATAAAATGGACCTCCTTTGTCGTTCGCGAAGCCGAGA
AAAGGGGGTGGAGCTGGGCATACTGGGAATTTTGTTCCGGTTTTGGTGTTTATGATCCTCTGAG
AAAACAGTGGAATAAAGATCTTTTAGAAGCTTTAATAGGAGGAGATAGCATTGAATAA

Figure 6

SEQ ID NO:4

TCTACTAGTTAGGAGGTAACTTATGGGCGTCGATCCGTTTGAACGTAACAAAATCTTGGGCCGC
GGCATTAATATCGGCAATGCGCTCGAAGCACCAAATGAAGGCGACTGGGGAGTGGTGATAAAAG
ATGAGTTCTTCGACATTATAAAAGAAGCCGGTTTCTCTCATGTTCGAATTCCAATAAGATGGAG
TACGCACGCTCAGGCGTTTCCTCCTTATAAAATCGAGCCTTCTTTCTTCAAAAGAGTGGATGAA
GTGATAAACGGAGCCCTGAAAAGAGGACTGGCTGTTGTTATAAATATTCATCACTACGAGGAGT
TAATGAATGATCCAGAAGAACACAAGGAAAGATTTCTTGCTCTTTGGAAACAAATTGCTGATCG
TTATAAAGACTATCCCGAAACTCTATTTTTTGAAATTCTGAATGAACCTCACGGAAATCTTACT
CCGGAAAAATGGAATGAACTGCTTGAGGAAGCTCTAAAAGTTATAAGATCAATTGACAAAAAGC
ACACTGTGATTATAGGCACAGCTGAATGGGGGGGTATATCTGCCCTTGAAAAACTGAGGGTCCC
AAAATGGGAAAAAAATGCGATAGTTACAATTCACTACTACAATCCTTTCGAATTTACCCATCAA
GGAGCTGAGTGGGTGCCTGGATCTGAGAAATGGTTGGGAAGAAAGTGGGGATCTCCAGATGATC
AGAAACATTTGATAGAAGAATTCAATTTTATAGAAGAATGGTCAAAAAAGAACAAAAGACCAAT
TTACATAGGTGAGTTTGGTGCCTACAGAAAAGCTGACCTTGAATCAAGAATAAAATGGACCTCC
TTTGTCGTTCGCGAAGCCGAGAAAAGGGGGTGGAGCTGGGCATACTGGGAATTTTGTTCCGGTT
TTGGTGTTTATGATCCTCTGAGAAAACAGTGGAATAAAGATCTTTTAGAAGCTTTAATAGGAGG
AGATAGCATTGAATGA

Figure 7

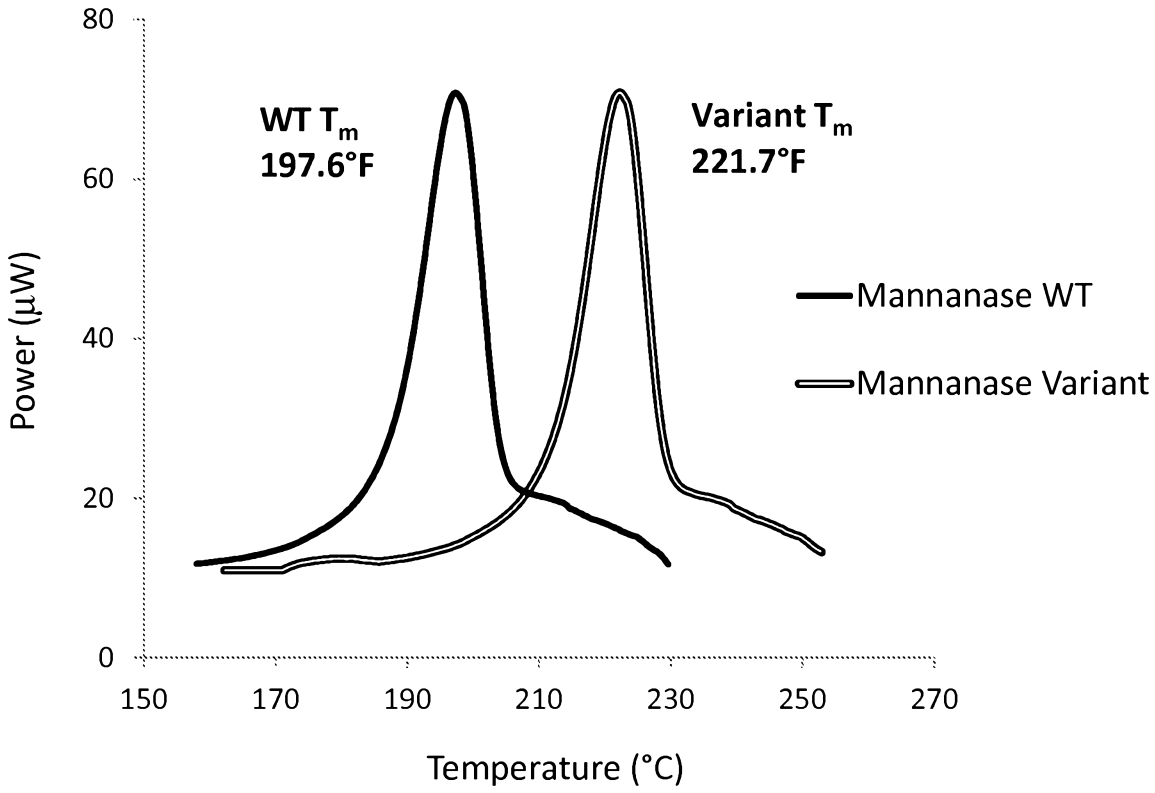

80
60
40
20
0
Power (µW)
150
170
190
210
230
250
270
Temperature (°C)
WT T_m
197.6°F
Variant T_m
221.7°F
Mannanase WT
Mannanase Variant

Figure 8A

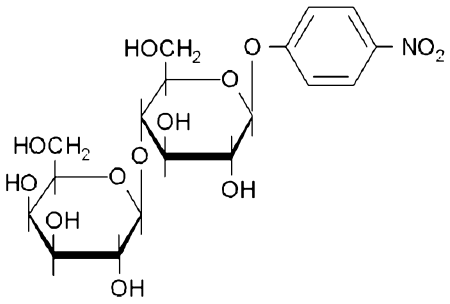


Figure 8B

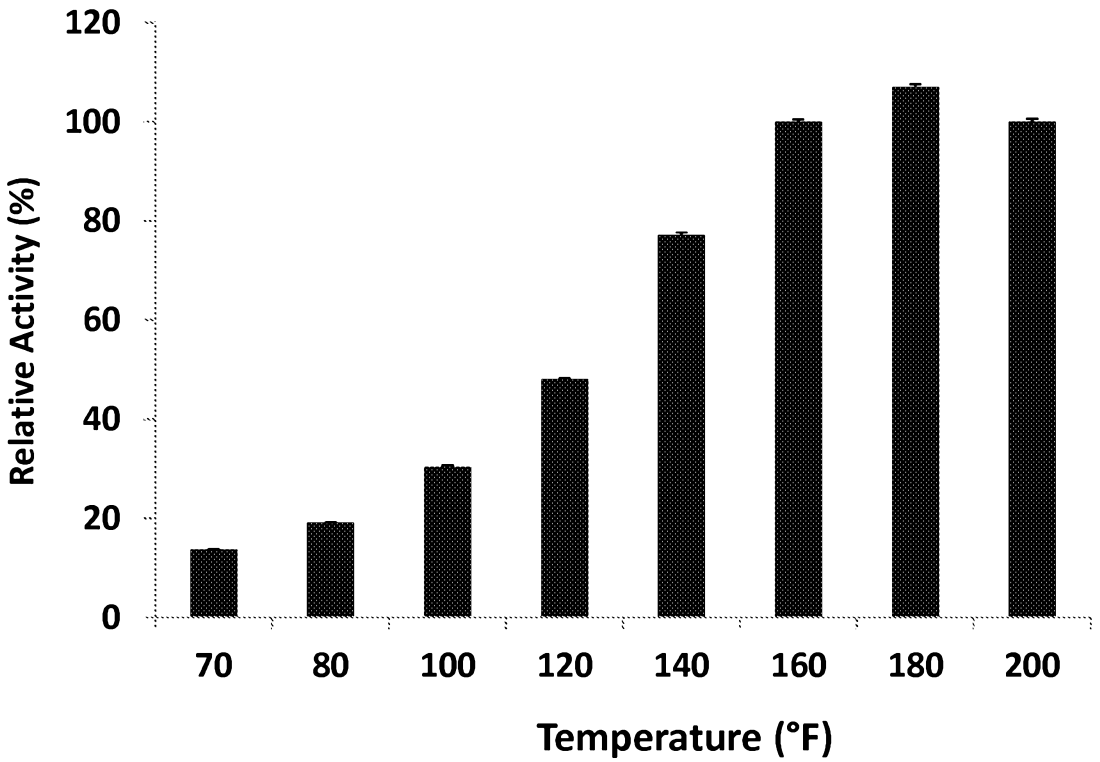
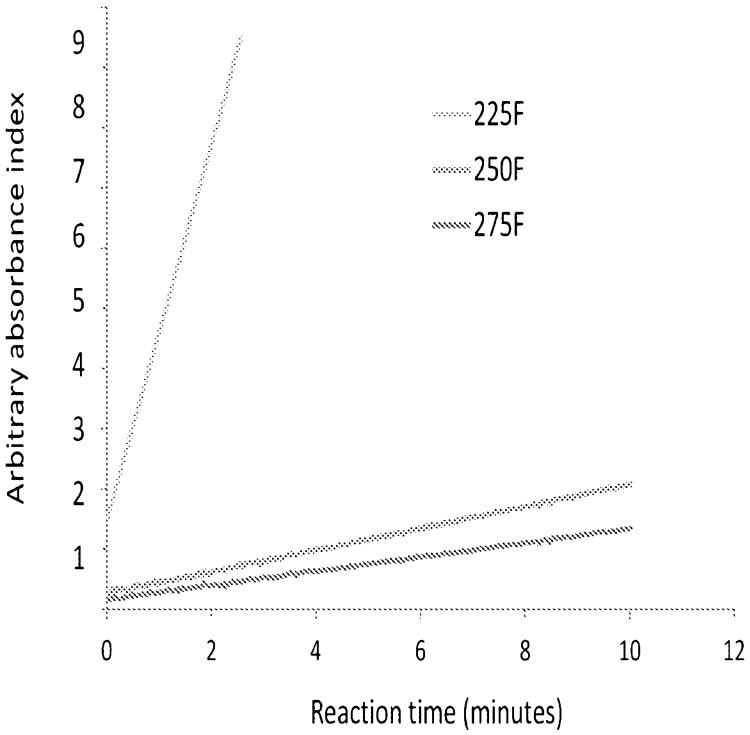


Figure 8C



Temperature (°F)	Relative activity (%)
225	100%
250	14%
275	11%

Figure 9

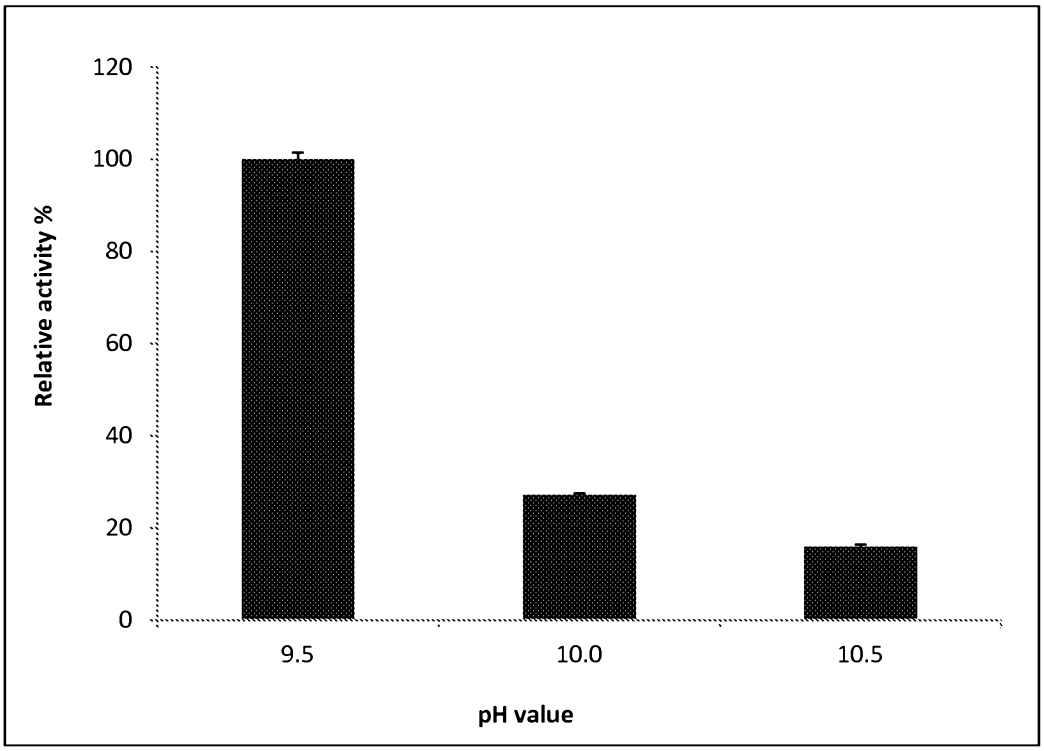


Figure 10A

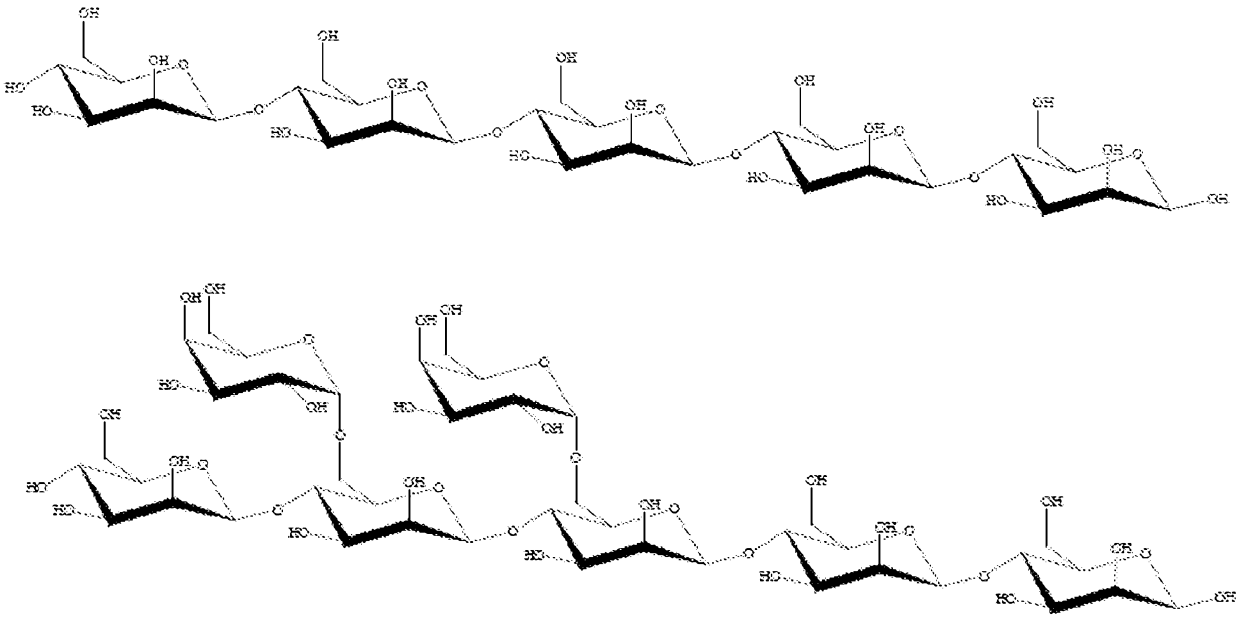


Figure 10B

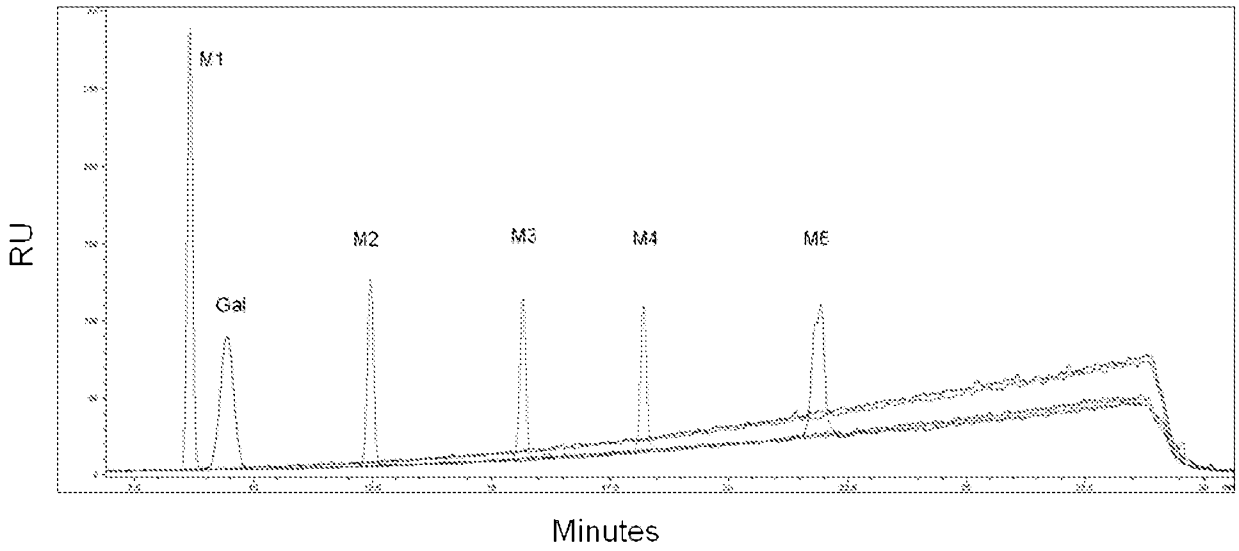


Figure 10C(a)

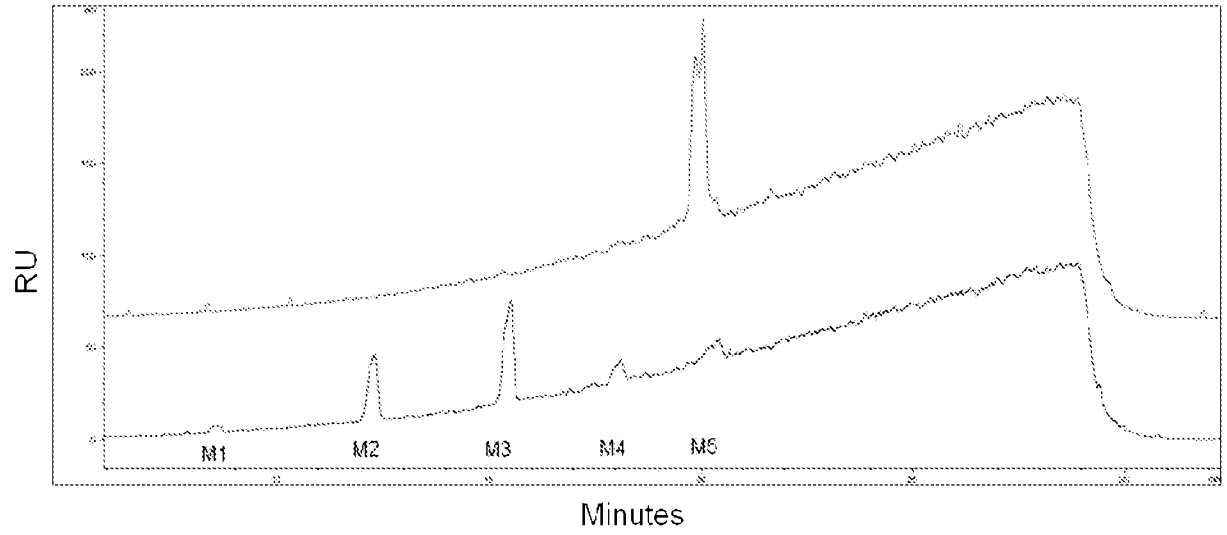


Figure 10C(b)

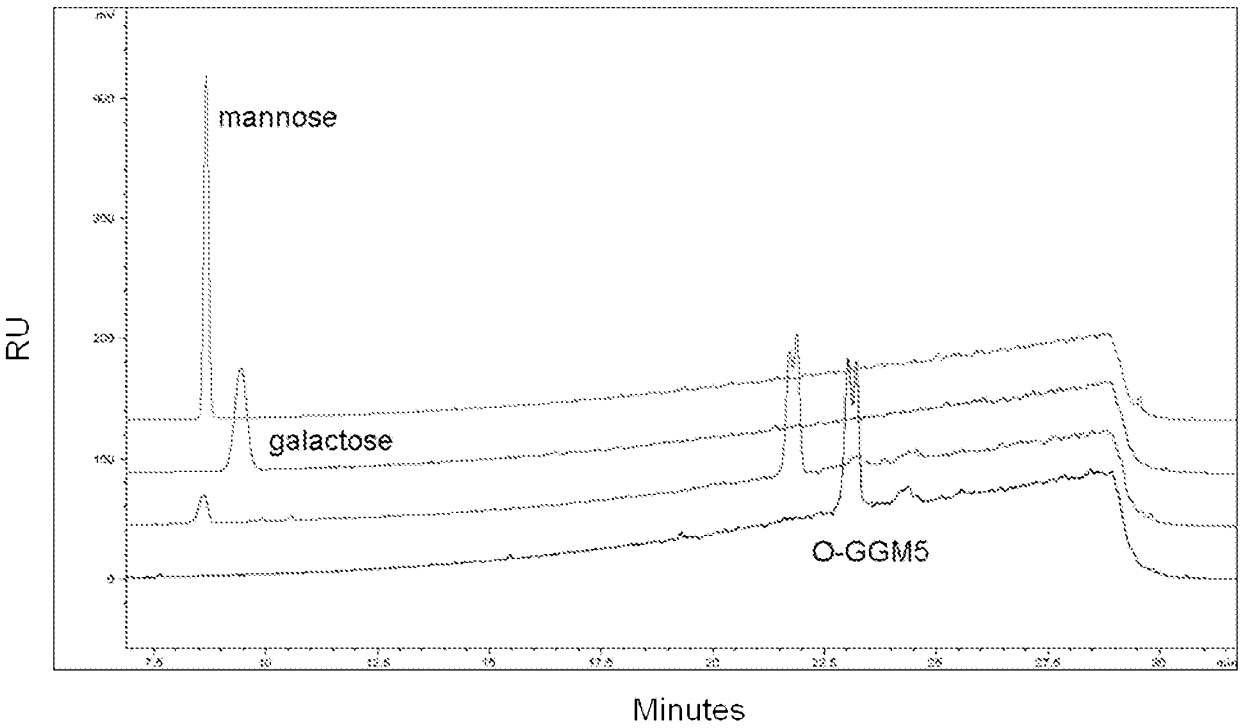


Figure 11

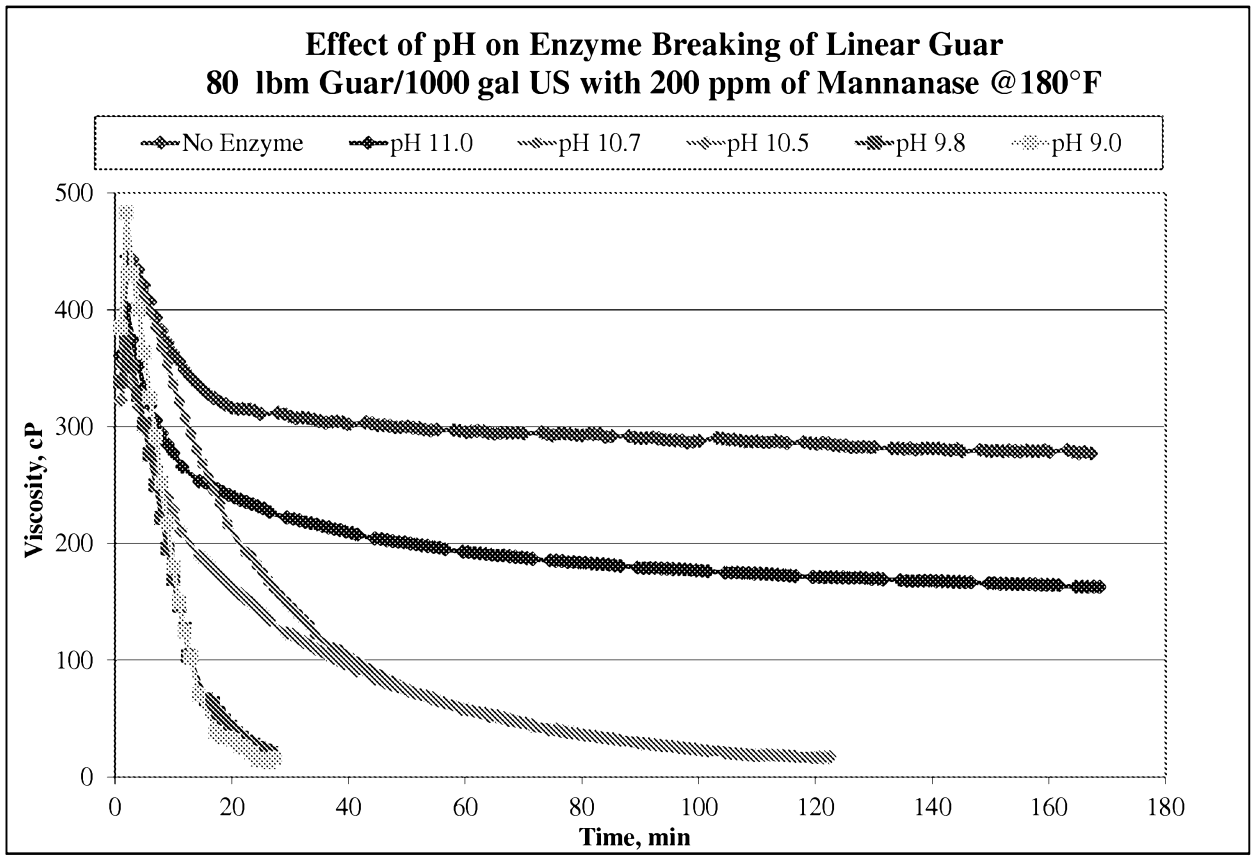

Effect of pH on Enzyme Breaking of Linear Guar
80 lbm Guar/1000 gal US with 200 ppm of Mannanase @180°F
No Enzyme
pH 11.0
pH 10.7
pH 10.5
pH 9.8
pH 9.0
Viscosity, cP
500
400
300
200
100
0
0
20
40
60
80
100
120
140
160
180
Time, min

Figure 12

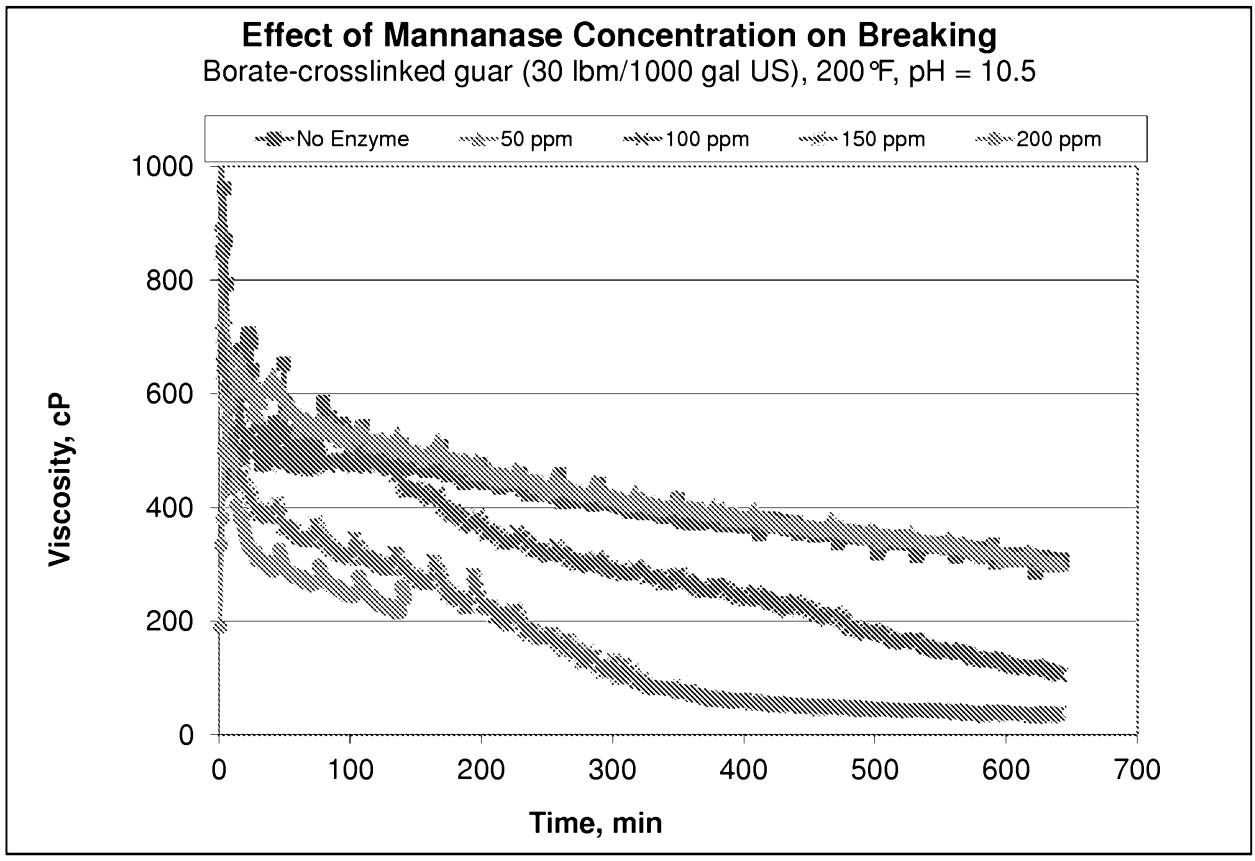

Effect of Mannanase Concentration on Breaking
Borate-crosslinked guar (30 lbm/1000 gal US), 200°F, pH = 10.5
No Enzyme
50 ppm
100 ppm
150 ppm
200 ppm
Viscosity, cP
1000
800
600
400
200
0
0
100
200
300
400
500
600
700
Time, min

Figure 13

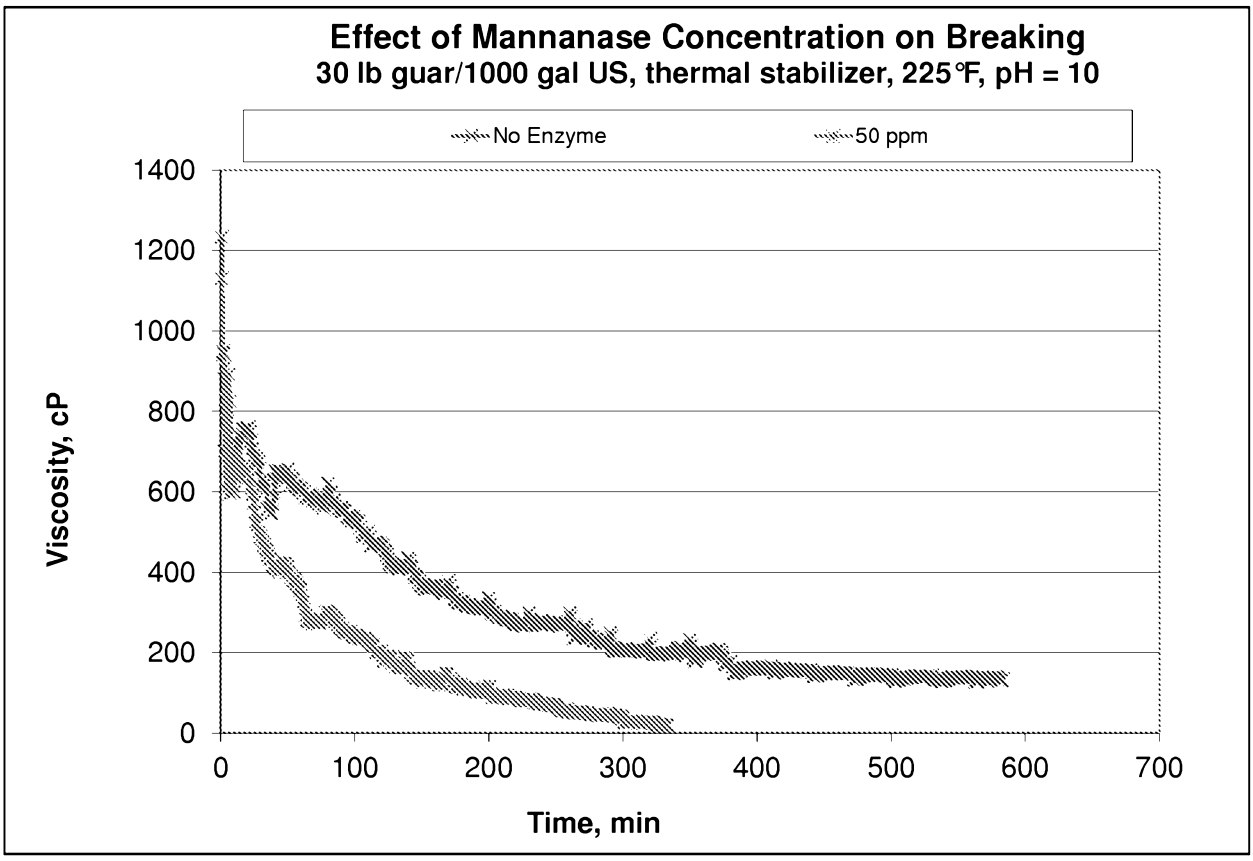

Effect of Mannanase Concentration on Breaking
30 lb guar/1000 gal US, thermal stabilizer, 225°F, pH = 10
No Enzyme
50 ppm
Viscosity, cP
1400
1200
1000
800
600
400
200
0
0
100
200
300
400
500
600
700
Time, min

Figure 14

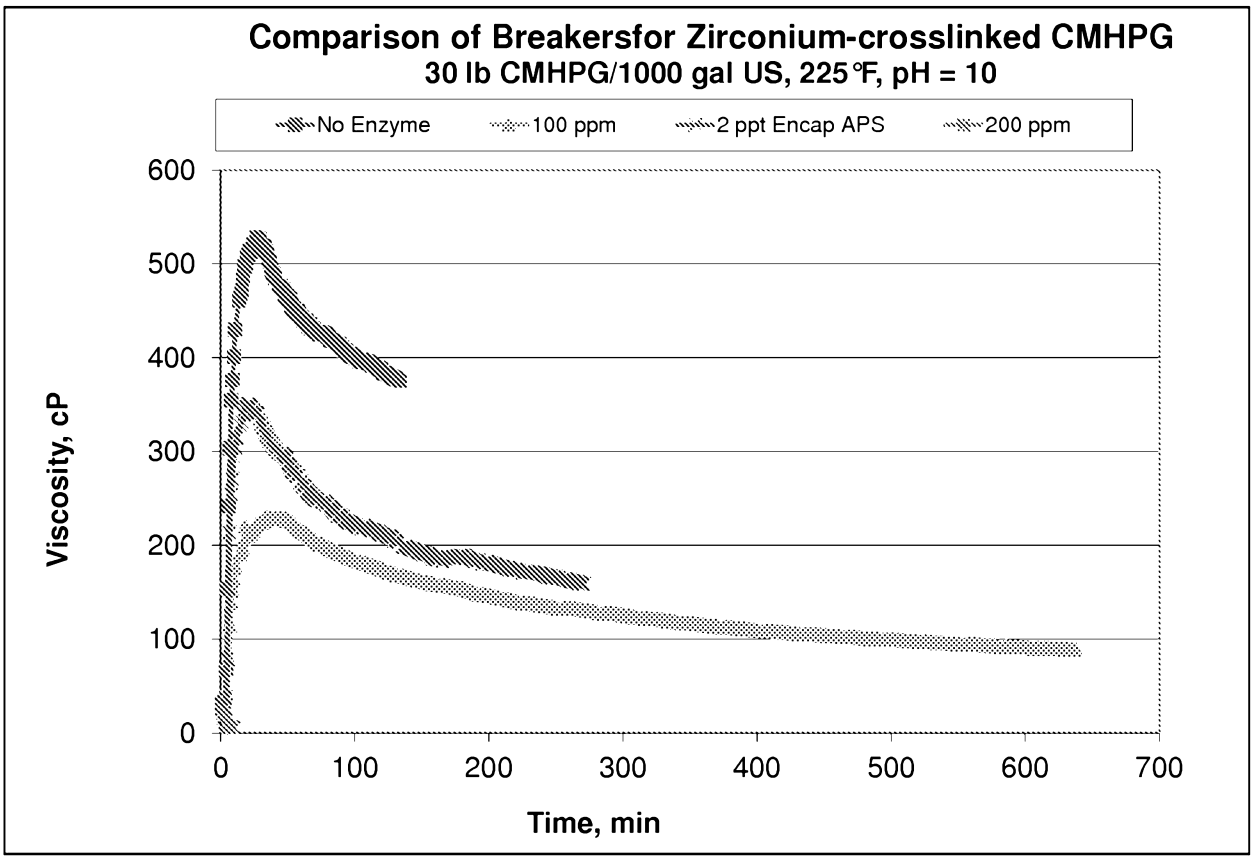

Comparison of Breakersfor Zirconium-crosslinked CMHPG
30 lb CMHPG/1000 gal US, 225°F, pH = 10
No Enzyme
100 ppm
2 ppt Encap APS
200 ppm
Viscosity, cP
600
500
400
300
200
100
0
0
100
200
300
400
500
600
700
Time, min

Figure 15

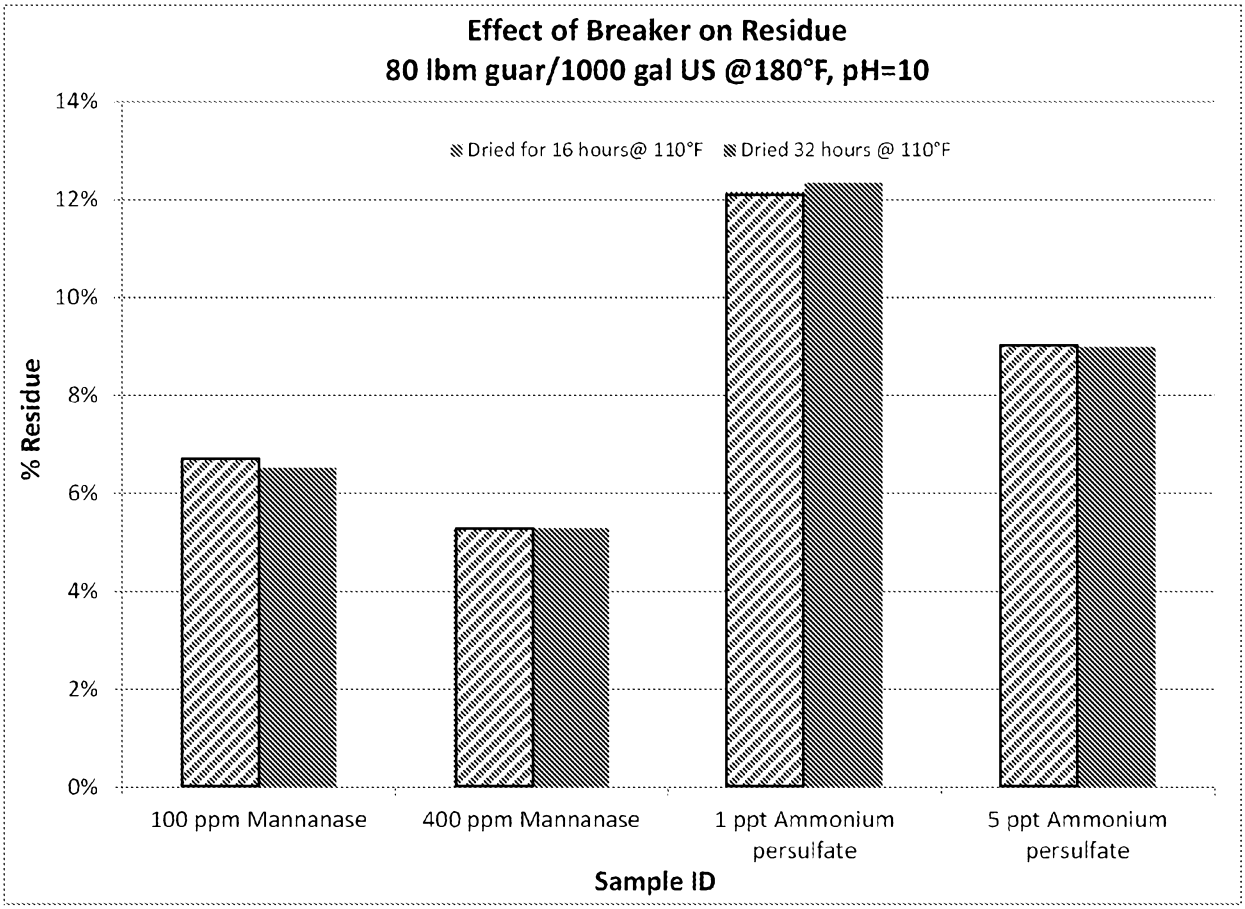
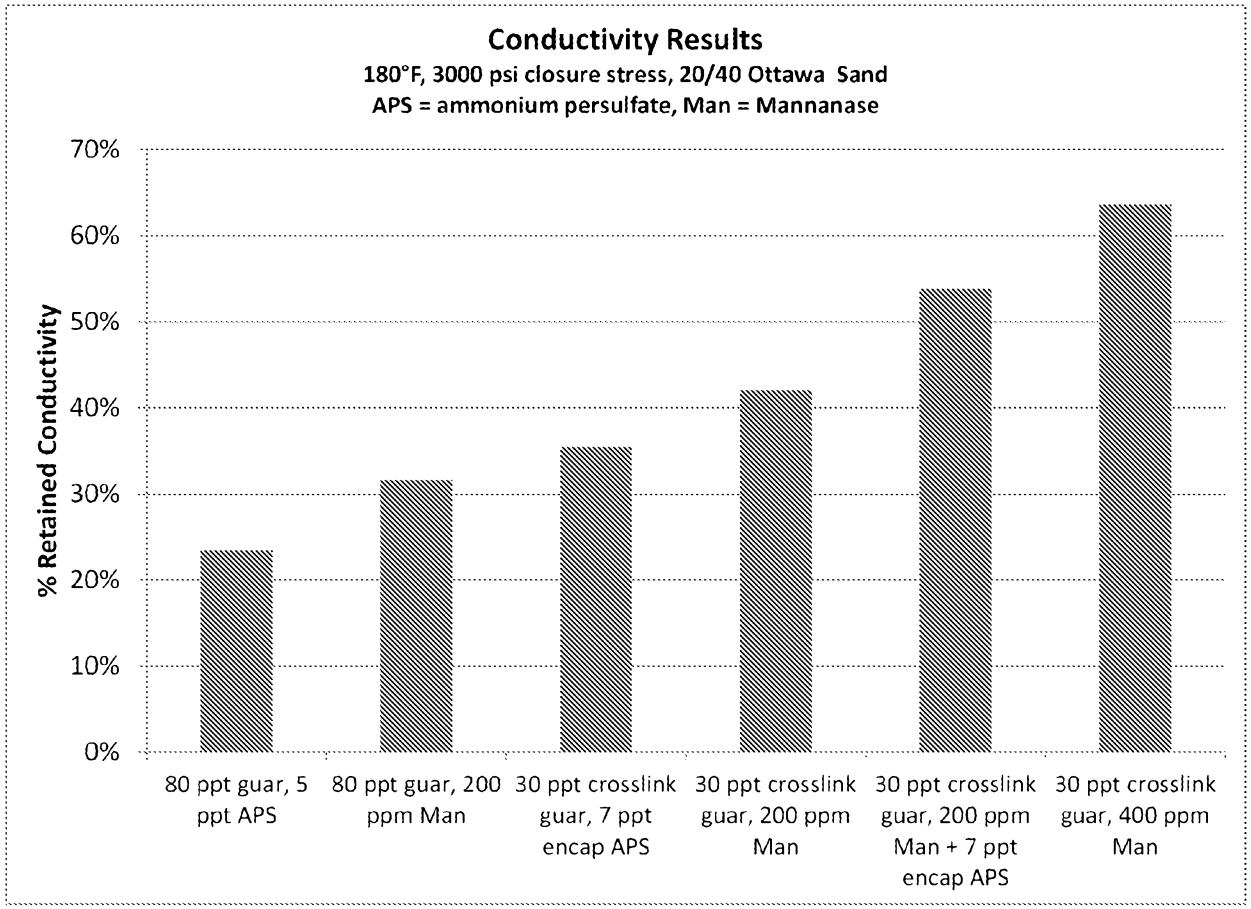

Effect of Breaker on Residue
80 lbm guar/1000 gal US @180°F, pH=10
Dried for 16 hours@ 110°F
Dried 32 hours @ 110°F
% Residue
14%
12%
10%
8%
6%
4%
2%
0%
100 ppm Mannanase
400 ppm Mannanase
1 ppt Ammonium persulfate
5 ppt Ammonium persulfate
Sample ID

Figure 16


Conductivity Results
180°F, 3000 psi closure stress, 20/40 Ottawa Sand
APS = ammonium persulfate, Man = Mannanase
% Retained Conductivity
70%
60%
50%
40%
30%
20%
10%
0%
80 ppt guar, 5 ppt APS
80 ppt guar, 200 ppm Man
30 ppt crosslink guar, 7 ppt encap APS
30 ppt crosslink guar, 200 ppm Man
30 ppt crosslink guar, 200 ppm Man + 7 ppt encap APS
30 ppt crosslink guar, 400 ppm Man

SEQUENCE LISTING

<110> Verenium Corporation
Zhang, Bin
Davenport, Adrienne Houston
Whipple, Lawrence A.
Urbina, Hugo D.
Barrett, Ken E.
Wall, Mark A.
Xuqiu, Tan

<120> GENES ENCODING CELLULASE FOR HYDROLYZING GUAR FRACTURING FLUIDS UNDER EXTREME WELL CONDITIONS

<130> D2570-1

<150> 61/618610
<151> 2012-03-30

<150> 61/660556
<151> 2012-06-15

<160> 21

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 954
<212> DNA
<213> Artificial Sequence

<220>
<223> Open reading frame encoding SEQ ID NO:2 conveying enhanced protein expression

<400> 1
atgggcgtcg atccgtttga acgtaacaaa atcttgggcc gcggcattaa tatcggcaat 60
gcgctcgaag caccaaatga aggcgactgg ggagtggtga taaaagatga gttcttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
caggcgtttc ctccttataa aatcgagcct tctttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaata ttcatcacta cgaggagtta 300
atgaatgatc cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctattt tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc acactgtgat tataggcaca gctgaatggg ggggtatatc tgcccttgaa 540
aaactgaggg tcccaaaatg ggaaaaaaat gcgatagtta caattcacta ctacaatcct 600
ttcgaattta cccatcaagg agctgagtgg gtgcctggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac ataggtgagt ttggtgccta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaagccga gaaaaggggg 840
tggagctggg catactggga attttgttcc ggttttggtg tttatgatcc tctgagaaaa 900
cagtggaata aagatctttt agaagcttta ataggaggag atagcattga atga 954

<210> 2
<211> 317
<212> PRT
<213> Thermotoga maritima

<400> 2
Met Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg Gly Ile
 1   5   10   15
Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp Gly Val
            20   25   30
Val Ile Lys Asp Glu Phe Phe Asp Ile Ile Lys Glu Ala Gly Phe Ser
        35   40   45
His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Gln Ala Phe Pro
    50   55   60
Pro Tyr Lys Ile Glu Pro Ser Phe Phe Lys Arg Val Asp Glu Val Ile
65   70   75   80
Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Val Ile Asn Ile His His
                85   90   95
Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg Phe Leu
            100   105   110
Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro Glu Thr
        115   120   125
Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr Pro Glu
    130   135   140

```
Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg Ser Ile
145                 150                 155                 160
Asp Lys Lys His Thr Val Ile Ile Gly Thr Ala Glu Trp Gly Gly Ile
                165                 170                 175
Ser Ala Leu Glu Lys Leu Arg Val Pro Lys Trp Glu Lys Asn Ala Ile
            180                 185                 190
Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln Gly Ala
        195                 200                 205
Glu Trp Val Pro Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp Gly Ser
    210                 215                 220
Pro Asp Asp Gln Lys His Leu Ile Glu Glu Phe Asn Phe Ile Glu Glu
225                 230                 235                 240
Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe Gly Ala
                245                 250                 255
Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser Phe Val
            260                 265                 270
Val Arg Glu Ala Glu Lys Arg Gly Trp Ser Trp Ala Tyr Trp Glu Phe
        275                 280                 285
Cys Ser Gly Phe Gly Val Tyr Asp Pro Leu Arg Lys Gln Trp Asn Lys
    290                 295                 300
Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu
305                 310                 315
```

<210> 3
<211> 954
<212> DNA
<213> Artificial Sequence

<220>
<223> Open reading frame encoding SEQ ID NO:2

<400> 3
```
atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtggtga taaaagatga gttcttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
caggcgtttc ctccttataa aatcgagcct tctttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaata ttcatcacta cgaggagtta 300
atgaatgatc cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctattt tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc acactgtgat tataggcaca gctgaatggg ggggtatatc tgcccttgaa 540
aaactgaggg tcccaaaatg ggaaaaaaat gcgatagtta caattcacta ctacaatcct 600
ttcgaattta cccatcaagg agctgagtgg gtgcctggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac ataggtgagt ttggtgccta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaagccga gaaaaggggg 840
tggagctggg catactggga attttgttcc ggttttggtg tttatgatcc tctgagaaaa 900
cagtggaata aagatctttt agaagcttta ataggaggag atagcattga ataa       954
```

<210> 4
<211> 976
<212> DNA
<213> Artificial Sequence

<220>
<223> 5' region and open reading frame encoding SEQ ID
NO:2 having modest increased expression

<400> 4
```
tctactagtt aggaggtaac ttatgggcgt cgatccgttt gaacgtaaca aaatcttggg 60
ccgcggcatt aatatcggca atgcgctcga agcaccaaat gaaggcgact ggggagtggt 120
gataaaagat gagttcttcg acattataaa agaagccggt ttctctcatg ttcgaattcc 180
aataagatgg agtacgcacg ctcaggcgtt tcctccttat aaaatcgagc cttctttctt 240
caaaagagtg gatgaagtga taaacggagc cctgaaaaga ggactggctg ttgttataaa 300
tattcatcac tacgaggagt taatgaatga tccagaagaa cacaaggaaa gatttcttgc 360
tctttggaaa caaattgctg atcgttataa agactatccc gaaactctat tttttgaaat 420
tctgaatgaa cctcacggaa atcttactcc ggaaaaatgg aatgaactgc ttgaggaagc 480
tctaaaagtt ataagatcaa ttgacaaaaa gcacactgtg attataggca cagctgaatg 540
gggggggtata tctgcccttg aaaaactgag ggtcccaaaa tgggaaaaaa atgcgatagt 600
tacaattcac tactacaatc ctttcgaatt tacccatcaa ggagctgagt gggtgcctgg 660
atctgagaaa tggttgggaa gaaagtgggg atctccagat gatcagaaac atttgataga 720
agaattcaat tttatagaag aatggtcaaa aaagaacaaa agaccaattt acataggtga 780
gtttggtgcc tacagaaaag ctgaccttga atcaagaata aaatggacct cctttgtcgt 840
tcgcgaagcc gagaaaaggg ggtggagctg ggcatactgg gaattttgtt ccggttttgg 900
tgtttatgat cctctgagaa aacagtggaa taaagatctt ttagaagctt taataggagg 960
agatagcatt gaatga                                                976
```

<210> 5
<211> 954

<212> DNA
<213> Thermotoga maritima

<400> 5
atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtggtga taaaagatga gttcttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
tacgcgtttc ctccttataa aatcatggat cgcttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaata ttcatcacta cgaggagtta 300
atgaatgatc cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctattt tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc acactataat tataggcaca gctgaatggg ggggtatatc tgcccttgaa 540
aaactgtctg tcccaaaatg ggaaaaaaat tctatagtta caattcacta ctacaatcct 600
ttcgaattta cccatcaagg agctgagtgg gtggaaggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac ataggtgagt ttggtgccta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaaatgga gaaaaggaga 840
tggagctggg catactggga attttgttcc ggttttggtg tttatgatac tctgagaaaa 900
acctggaata aagatctttt agaagcttta ataggaggag atagcattga ataa 954

<210> 6
<211> 317
<212> PRT
<213> Thermotoga maritima

<220>
<221> DOMAIN
<222> (19)...(296)
<223> Cellulase (glycosyl hydrolase family 5)

<400> 6
Met Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg Gly Ile
 1   5   10   15
Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp Gly Val
            20   25   30
Val Ile Lys Asp Glu Phe Phe Asp Ile Ile Lys Glu Ala Gly Phe Ser
        35   40   45
His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Tyr Ala Phe Pro
    50   55   60
Pro Tyr Lys Ile Met Asp Arg Phe Phe Lys Arg Val Asp Glu Val Ile
65   70   75   80
Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Val Ile Asn Ile His His
                85   90   95
Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg Phe Leu
            100   105   110
Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro Glu Thr
        115   120   125
Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr Pro Glu
    130   135   140
Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg Ser Ile
145   150   155   160
Asp Lys Lys His Thr Ile Ile Ile Gly Thr Ala Glu Trp Gly Gly Ile
                165   170   175
Ser Ala Leu Glu Lys Leu Ser Val Pro Lys Trp Glu Lys Asn Ser Ile
            180   185   190
Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln Gly Ala
        195   200   205
Glu Trp Val Glu Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp Gly Ser
    210   215   220
Pro Asp Asp Gln Lys His Leu Ile Glu Glu Phe Asn Phe Ile Glu Glu
225   230   235   240
Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe Gly Ala
                245   250   255
Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser Phe Val
            260   265   270
Val Arg Glu Met Glu Lys Arg Arg Trp Ser Trp Ala Tyr Trp Glu Phe
        275   280   285
Cys Ser Gly Phe Gly Val Tyr Asp Thr Leu Arg Lys Thr Trp Asn Lys
    290   295   300
Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu
305   310   315

<210> 7
<211> 954
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetically generated polynucleotide

<400> 7
atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggcgactgg ggagtcgtga taaaagatga gttcttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
tacgcgtttc ctccttataa aatcatggat cgcttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaata ttcatcacta cgaggagtta 300
atgaatgatc cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctattt tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc acactataat tataggcaca gctgaatggg ggggtatatc tgcccttgaa 540
aaactgtctg tcccaaaatg ggaaaaaaat tctatagtta caattcacta ctacaatcct 600
ttcgaattta cccatcaagg agctgagtgg gtggaaggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac ataggtgagt ttggtgccta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaaatgga gaaaaggaga 840
tggagctggg catactggga attttgttcc ggttttggtg tttatgatac tctgagaaaa 900
acctggaata aagatctttt agaagcttta ataggaggag atagcattga ataa 954

<210> 8
<211> 974
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetically generated polynucleotide

<400> 8
atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtggtga taaaagatga gtatttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
caggcgtttc ctccttataa aatcgaggat cgcttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaatc agcatcacta cgaggagtta 300
atgaatgatc cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctattt tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc acactataat tataggcaca gctgaatggg ggggtatatc tgcccttgaa 540
aaactgaggg tcccaaaatg ggaaaaaaat gcgatagtta caattcacta ctacaatcct 600
ttcgaattta cccatcaagg agctgagtgg gtggaaggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac ataggtgagt ttggtgccta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaagctga gaaaaggaga 840
tggagctggg catactggga attttgttcc ggttttggtg tttatgatac tctgagaaaa 900
acctggaata aagatctttt agaagcttta ataggaggag atagcattga ataacaccat 960
tccaagatgg cgtg 974

<210> 9
<211> 320
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetically generated polypeptide

<400> 9
Met Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg Gly Ile
 1   5   10   15
Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp Gly Val
            20   25   30
Val Ile Lys Asp Glu Tyr Phe Asp Ile Ile Lys Glu Ala Gly Phe Ser
        35   40   45
His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Gln Ala Phe Pro
    50   55   60
Pro Tyr Lys Ile Glu Asp Arg Phe Phe Lys Arg Val Asp Glu Val Ile
65   70   75   80
Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Val Ile Asn Gln His His
                85   90   95
Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg Phe Leu
            100   105   110
Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro Glu Thr
        115   120   125
Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr Pro Glu
    130   135   140
Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg Ser Ile
145   150   155   160
Asp Lys Lys His Thr Ile Ile Ile Gly Thr Ala Glu Trp Gly Gly Ile
                165   170   175
Ser Ala Leu Glu Lys Leu Arg Val Pro Lys Trp Glu Lys Asn Ala Ile
            180   185   190

Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln Gly Ala
195 200 205
Glu Trp Val Glu Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp Gly Ser
210 215 220
Pro Asp Asp Gln Lys His Leu Ile Glu Glu Phe Asn Phe Ile Glu Glu
225 230 235 240
Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe Gly Ala
245 250 255
Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser Phe Val
260 265 270
Val Arg Glu Ala Glu Lys Arg Arg Trp Ser Trp Ala Tyr Trp Glu Phe
275 280 285
Cys Ser Gly Phe Gly Val Tyr Asp Thr Leu Arg Lys Thr Trp Asn Lys
290 295 300
Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu His His Ser
305 310 315 320

<210> 10
<211> 954
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetically generated polynucleotide

<400> 10
atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtggtga taaaagatga gtatttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
caggcgtttc ctccttataa aatcgaggat tctttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaata ttcatcacta cgaggagtta 300
atgaatgatc cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctattt tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc acactgtgat tataggcaca gctgaatggg ggggtatatc tgcccttgaa 540
aaactgaggg tcccaaaatg ggaaaaaaat gcgatagtta caattcacta ctacaatcct 600
ttcgaattta cccatcaagg agctgagtgg gtgcctggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat gtgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac ataggtgagt ttggtgccta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaagccga gaaaaggggg 840
tggagctggg catactggga attttgttcc ggttttggtg tttatgatcc tctgagaaaa 900
cagtggaata aagatctttt agaagctcta ataggaggag atagcattga ataa 954

<210> 11
<211> 317
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetically generated polypeptide

<400> 11
Met Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg Gly Ile
1 5 10 15
Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp Gly Val
20 25 30
Val Ile Lys Asp Glu Tyr Phe Asp Ile Ile Lys Glu Ala Gly Phe Ser
35 40 45
His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Gln Ala Phe Pro
50 55 60
Pro Tyr Lys Ile Glu Asp Ser Phe Phe Lys Arg Val Asp Glu Val Ile
65 70 75 80
Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Val Ile Asn Ile His His
85 90 95
Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg Phe Leu
100 105 110
Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro Glu Thr
115 120 125
Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr Pro Glu
130 135 140
Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg Ser Ile
145 150 155 160
Asp Lys Lys His Thr Val Ile Ile Gly Thr Ala Glu Trp Gly Gly Ile
165 170 175
Ser Ala Leu Glu Lys Leu Arg Val Pro Lys Trp Glu Lys Asn Ala Ile
180 185 190
Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln Gly Ala
195 200 205
Glu Trp Val Pro Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp Gly Ser

210 215 220
Pro Asp Asp Gln Lys His Val Ile Glu Glu Phe Asn Phe Ile Glu Glu
225 230 235 240
Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe Gly Ala
245 250 255
Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser Phe Val
260 265 270
Val Arg Glu Ala Glu Lys Arg Gly Trp Ser Trp Ala Tyr Trp Glu Phe
275 280 285
Cys Ser Gly Phe Gly Val Tyr Asp Pro Leu Arg Lys Gln Trp Asn Lys
290 295 300
Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu
305 310 315

<210> 12
<211> 954
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetically generated polynucleotide

<400> 12
atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtggtga taaaagatga gttcttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
caggcgtttc ctccttataa aatcgaggat tctttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaatc agcatcacta cgaggagtta 300
atgaatgatc cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctattt tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc acactgtgat tataggcaca gctgaatggg ggggtatatc tgcccttgaa 540
aaactgaggg tcccaaaatg ggaaaaaaat gcgatagtta caattcacta ctacaatcct 600
ttcgaattta cccatcaagg agctgagtgg gtgcctggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac ataggtgagt ttggtgccta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaagccga gaaaaggggg 840
tggagctggg catactggga attttgttcc ggttttggtg tttatgatcc tctgagaaaa 900
cagtggaata aagatctttt agaagcttta ataggaggag atagcattga ataa 954

<210> 13
<211> 317
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetically generated polypeptide

<400> 13
Met Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg Gly Ile
1 5 10 15
Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp Gly Val
20 25 30
Val Ile Lys Asp Glu Phe Phe Asp Ile Ile Lys Glu Ala Gly Phe Ser
35 40 45
His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Gln Ala Phe Pro
50 55 60
Pro Tyr Lys Ile Glu Asp Ser Phe Phe Lys Arg Val Asp Glu Val Ile
65 70 75 80
Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Val Ile Asn Gln His His
85 90 95
Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg Phe Leu
100 105 110
Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro Glu Thr
115 120 125
Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr Pro Glu
130 135 140
Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg Ser Ile
145 150 155 160
Asp Lys Lys His Thr Val Ile Ile Gly Thr Ala Glu Trp Gly Gly Ile
165 170 175
Ser Ala Leu Glu Lys Leu Arg Val Pro Lys Trp Glu Lys Asn Ala Ile
180 185 190
Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln Gly Ala
195 200 205
Glu Trp Val Pro Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp Gly Ser
210 215 220
Pro Asp Asp Gln Lys His Leu Ile Glu Glu Phe Asn Phe Ile Glu Glu
225 230 235 240

Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe Gly Ala
245 250 255
Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser Phe Val
260 265 270
Val Arg Glu Ala Glu Lys Arg Gly Trp Ser Trp Ala Tyr Trp Glu Phe
275 280 285
Cys Ser Gly Phe Gly Val Tyr Asp Pro Leu Arg Lys Gln Trp Asn Lys
290 295 300
Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu
305 310 315

<210> 14
<211> 972
<212> DNA
<213> Thermotoga sp.

<400> 14
atggaacagt cagttgctga aagtgatagc aactcagcat ttgaatacaa caaaatggta 60
ggtaaaggag taaatattgg aaatgcttta gaagctcctt tcgaaggagc ttggggagta 120
agaattgagg atgaatattt tgagataata aagaaaaggg gatttgattc tgttaggatt 180
cccataagat ggtcagcaca tatatccgaa aagccaccat atgatattga caggaatttc 240
ctcgaaagag ttaaccatgt tgtcgatagg gctcttgaga ataatttaac agtaatcatc 300
aatacgcacc attttgaaga actctatcaa gaaccggata aatacggcga tgttttggtg 360
gaaatttgga gacagattgc aaaattcttt aaagattacc cggaaaatct gttctttgaa 420
atctacaacg agcctgctca gaacttgaca gctgaaaaat ggaacgcact ttatccaaaa 480
gtgctcaaag ttatcaggga gagcaatcca acccggattg tcattatcga tgctccaaac 540
tgggcacact atagcgcagt gagaagtcta aaattagtca acgacaaacg catcattgtt 600
tccttccatt actacgaacc tttcaaattc acacatcagg gtgccgaatg ggttaatccc 660
atcccacctg ttagggttaa gtggaatggc gaggaatggg aaattaacca aatcagaagt 720
catttcaaat acgtgagtga ctgggcaaag caaaataacg taccaatctt tcttggtgaa 780
ttcggtgctt attcaaaagc agacatggac tcaagggtta agtggaccga aagtgtgaga 840
aaaatggcgg aagaatttgg attttcatac gcgtattggg aattttgtgc aggatttggc 900
atatacgata gatggtctca aaactggatc gaaccattgg caacagctgt ggttggcaca 960
ggcaaagagt aa 972

<210> 15
<211> 323
<212> PRT
<213> Thermotoga sp.

<400> 15
Met Glu Gln Ser Val Ala Glu Ser Asp Ser Asn Ser Ala Phe Glu Tyr
1 5 10 15
Asn Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala
20 25 30
Pro Phe Glu Gly Ala Trp Gly Val Arg Ile Glu Asp Glu Tyr Phe Glu
35 40 45
Ile Ile Lys Lys Arg Gly Phe Asp Ser Val Arg Ile Pro Ile Arg Trp
50 55 60
Ser Ala His Ile Ser Glu Lys Pro Pro Tyr Asp Ile Asp Arg Asn Phe
65 70 75 80
Leu Glu Arg Val Asn His Val Val Asp Arg Ala Leu Glu Asn Asn Leu
85 90 95
Thr Val Ile Ile Asn Thr His His Phe Glu Glu Leu Tyr Gln Glu Pro
100 105 110
Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln Ile Ala Lys
115 120 125
Phe Phe Lys Asp Tyr Pro Glu Asn Leu Phe Phe Glu Ile Tyr Asn Glu
130 135 140
Pro Ala Gln Asn Leu Thr Ala Glu Lys Trp Asn Ala Leu Tyr Pro Lys
145 150 155 160
Val Leu Lys Val Ile Arg Glu Ser Asn Pro Thr Arg Ile Val Ile Ile
165 170 175
Asp Ala Pro Asn Trp Ala His Tyr Ser Ala Val Arg Ser Leu Lys Leu
180 185 190
Val Asn Asp Lys Arg Ile Ile Val Ser Phe His Tyr Tyr Glu Pro Phe
195 200 205
Lys Phe Thr His Gln Gly Ala Glu Trp Val Asn Pro Ile Pro Pro Val
210 215 220
Arg Val Lys Trp Asn Gly Glu Glu Trp Glu Ile Asn Gln Ile Arg Ser
225 230 235 240
His Phe Lys Tyr Val Ser Asp Trp Ala Lys Gln Asn Asn Val Pro Ile
245 250 255
Phe Leu Gly Glu Phe Gly Ala Tyr Ser Lys Ala Asp Met Asp Ser Arg
260 265 270
Val Lys Trp Thr Glu Ser Val Arg Lys Met Ala Glu Glu Phe Gly Phe
275 280 285
Ser Tyr Ala Tyr Trp Glu Phe Cys Ala Gly Phe Gly Ile Tyr Asp Arg
290 295 300

Trp Ser Gln Asn Trp Ile Glu Pro Leu Ala Thr Ala Val Val Gly Thr
305 310 315 320
Gly Lys Glu

<210> 16
<211> 1042
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetically generated polynucleotide

<400> 16
ggatccacca tgagggtgtt gctcgttgcc ctcgctctcc tggctctcgc tgcgagcgcc 60
accagcggcg tggacccgtt cgagaggaac aagatcctgg gcaggggcat caacatcggc 120
aacgccctgg aggccccgaa cgagggcgac tggggcgtgg tgatcaagga cgagtacttc 180
gacatcatca aggaggccgg cttcagccac gtgagaatcc cgatcaggtg gagcacccac 240
gcccaggcct tcccgccgta caagatcgag gacaggttct tcaagagggt ggacgaggtg 300
atcaacggcg ccctgaagag gggcctggcc gtggtgatca accagcacca ctacgaggag 360
ctgatgaacg acccggagga gcacaaggag aggttcctgg ccctgtggaa gcagatcgcc 420
gacaggtaca aggactaccc ggagaccctg ttcttcgaga tcctgaacga gccgcacggc 480
aacctgaccc cggagaagtg gaacgagctg ctggaggagg ccctgaaggt gatcaggagc 540
atcgacaaga agcacaccat catcatcggc accgccgagt ggggcggcat cagcgccctg 600
gagaagctga gggtgccgaa gtgggagaag aacgccatcg tgaccatcca ctactacaac 660
ccgttcgagt tcacccacca gggcgccgag tgggtggagg gcagcgagaa gtggctgggc 720
aggaagtggg gcagcccgga cgaccagaag cacctgatcg aggagttcaa cttcatcgag 780
gagtggagca agaagaacaa gaggccgatc tacatcggcg agttcggcgc ctacaggaag 840
gccgacctgg agagcaggat caagtggacc agcttcgtgg tgagggaggc cgagaagagg 900
aggtggagct gggcctactg ggagttctgc agcggcttcg gcgtgtacga caccctgagg 960
aagacctgga acaaggacct gctggaggcc ctgatcggcg gcgacagcat cgagagcgag 1020
aaggacgagc tgtgagagct ca 1042

<210> 17
<211> 341
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetically generated polypeptide

<400> 17
Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
 1 5 10 15
Ala Thr Ser Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg
 20 25 30
Gly Ile Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp
 35 40 45
Gly Val Val Ile Lys Asp Glu Tyr Phe Asp Ile Ile Lys Glu Ala Gly
 50 55 60
Phe Ser His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Gln Ala
65 70 75 80
Phe Pro Pro Tyr Lys Ile Glu Asp Arg Phe Phe Lys Arg Val Asp Glu
 85 90 95
Val Ile Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Val Ile Asn Gln
 100 105 110
His His Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg
 115 120 125
Phe Leu Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro
 130 135 140
Glu Thr Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr
145 150 155 160
Pro Glu Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg
 165 170 175
Ser Ile Asp Lys Lys His Thr Ile Ile Ile Gly Thr Ala Glu Trp Gly
 180 185 190
Gly Ile Ser Ala Leu Glu Lys Leu Arg Val Pro Lys Trp Glu Lys Asn
 195 200 205
Ala Ile Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln
 210 215 220
Gly Ala Glu Trp Val Glu Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp
225 230 235 240
Gly Ser Pro Asp Asp Gln Lys His Leu Ile Glu Glu Phe Asn Phe Ile
 245 250 255
Glu Glu Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe
 260 265 270
Gly Ala Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser
 275 280 285
Phe Val Val Arg Glu Ala Glu Lys Arg Arg Trp Ser Trp Ala Tyr Trp

290 295 300
Glu Phe Cys Ser Gly Phe Gly Val Tyr Asp Thr Leu Arg Lys Thr Trp
305 310 315 320
Asn Lys Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu Ser
325 330 335
Glu Lys Asp Glu Leu
340

<210> 18
<211> 1042
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetically generated polynucleotide

<400> 18
ggatccacca tgagggtgtt gctcgttgcc ctcgctctcc tggctctcgc tgcgagcgcc 60
accagcggcg tggacccgtt cgagaggaac aagatcctgg gcaggggcat caacatcggc 120
aacgccctgg aggccccgaa cgagggcgac tggggcgtgg tgatcaagga cgagttcttc 180
gacatcatca aggaggccgg cttcagccac gtgagaatcc cgatcaggtg gagcacccac 240
gcccaggcct tcccgccgta caagatcgag ccgagcttct tcaagagggt ggacgaggtg 300
atcaacggcg ccctgaagag gggcctggcc gtggtgatca acatccacca ctacgaggag 360
ctgatgaacg acccggagga gcacaaggag aggttcctgg ccctgtggaa gcagatcgcc 420
gacaggtaca aggactaccc ggagaccctg ttcttcgaga tcctgaacga gccgcacggc 480
aacctgaccc cggagaagtg gaacgagctg ctggaggagg ccctgaaggt gatcaggagc 540
atcgacaaga agcacaccgt gatcatcggc accgccgagt ggggcggcat cagcgccctg 600
gagaagctga gggtgccgaa gtgggagaag aacgccatcg tgaccatcca ctactacaac 660
ccgttcgagt tcacccacca gggcgccgag tgggtgccgg gcagcgagaa gtggctgggc 720
aggaagtggg gcagcccgga cgaccagaag cacctgatcg aggagttcaa cttcatcgag 780
gagtggagca agaagaacaa gaggccgatc tacatcggcg agttcggcgc ctacaggaag 840
gccgacctgg agagcaggat caagtggacc agcttcgtgg tgagggaggc cgagaagagg 900
ggctggagct gggcctactg ggagttctgc agcggcttcg gcgtgtacga cccgctgagg 960
aagcagtgga acaaggacct gctggaggcc ctgatcggcg gcgacagcat cgagagcgag 1020
aaggacgagc tgtgagagct ca 1042

<210> 19
<211> 341
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetically generated polypeptide

<400> 19
Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
1 5 10 15
Ala Thr Ser Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg
20 25 30
Gly Ile Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp
35 40 45
Gly Val Val Ile Lys Asp Glu Phe Phe Asp Ile Ile Lys Glu Ala Gly
50 55 60
Phe Ser His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Gln Ala
65 70 75 80
Phe Pro Pro Tyr Lys Ile Glu Pro Ser Phe Phe Lys Arg Val Asp Glu
85 90 95
Val Ile Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Val Ile Asn Ile
100 105 110
His His Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg
115 120 125
Phe Leu Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro
130 135 140
Glu Thr Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr
145 150 155 160
Pro Glu Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg
165 170 175
Ser Ile Asp Lys Lys His Thr Val Ile Ile Gly Thr Ala Glu Trp Gly
180 185 190
Gly Ile Ser Ala Leu Glu Lys Leu Arg Val Pro Lys Trp Glu Lys Asn
195 200 205
Ala Ile Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln
210 215 220
Gly Ala Glu Trp Val Pro Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp
225 230 235 240
Gly Ser Pro Asp Asp Gln Lys His Leu Ile Glu Glu Phe Asn Phe Ile
245 250 255
Glu Glu Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe
260 265 270

Gly Ala Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser
275 280 285
Phe Val Val Arg Glu Ala Glu Lys Arg Gly Trp Ser Trp Ala Tyr Trp
290 295 300
Glu Phe Cys Ser Gly Phe Gly Val Tyr Asp Pro Leu Arg Lys Gln Trp
305 310 315 320
Asn Lys Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu Ser
325 330 335
Glu Lys Asp Glu Leu
340

<210> 20
<211> 1042
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetically generated polynucleotide

<400> 20
ggatccacca tgagggtgtt gctcgttgcc ctcgctctcc tggctctcgc tgcgagcgcc 60
accagcggcg tggacccgtt cgagaggaac aagatcctgg gcaggggcat caacatcggc 120
aacgccctgg aggccccgaa cgagggcgac tggggcgtgg tgatcaagga cgagtacttc 180
gacatcatca aggaggccgg cttcagccac gtgagaatcc cgatcaggtg gagcacccac 240
gcccaggcct tcccgccgta caagatcgag gacagcttct tcaagagggt ggacgaggtg 300
atcaacggcg ccctgaagag gggcctggcc gtggtgatca acatccacca ctacgaggag 360
ctgatgaacg acccggagga gcacaaggag aggttcctgg ccctgtggaa gcagatcgcc 420
gacaggtaca aggactaccc ggagaccctg ttcttcgaga tcctgaacga gccgcacggc 480
aacctgaccc cggagaagtg gaacgagctg ctggaggagg ccctgaaggt gatcaggagc 540
atcgacaaga agcacaccgt gatcatcggc accgccgagt ggggcggcat cagcgccctg 600
gagaagctga gggtgccgaa gtgggagaag aacgccatcg tgaccatcca ctactacaac 660
ccgttcgagt tcacccacca gggcgccgag tgggtgccgg gcagcgagaa gtggctgggc 720
aggaagtggg gcagcccgga cgaccagaag cacgtgatcg aggagttcaa cttcatcgag 780
gagtggagca agaagaacaa gaggccgatc tacatcggcg agttcggcgc ctacaggaag 840
gccgacctgg agagcaggat caagtggacc agcttcgtgg tgagggaggc cgagaagagg 900
ggctggagct gggcctactg ggagttctgc agcggcttcg gcgtgtacga cccgctgagg 960
aagcagtgga acaaggacct gctggaggcc ctgatcggcg gcgacagcat cgagagcgag 1020
aaggacgagc tgtgagagct ca 1042

<210> 21
<211> 341
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetically generated polypeptide

<400> 21
Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
1 5 10 15
Ala Thr Ser Gly Val Asp Pro Phe Glu Arg Asn Lys Ile Leu Gly Arg
20 25 30
Gly Ile Asn Ile Gly Asn Ala Leu Glu Ala Pro Asn Glu Gly Asp Trp
35 40 45
Gly Val Val Ile Lys Asp Glu Tyr Phe Asp Ile Ile Lys Glu Ala Gly
50 55 60
Phe Ser His Val Arg Ile Pro Ile Arg Trp Ser Thr His Ala Gln Ala
65 70 75 80
Phe Pro Pro Tyr Lys Ile Glu Asp Ser Phe Phe Lys Arg Val Asp Glu
85 90 95
Val Ile Asn Gly Ala Leu Lys Arg Gly Leu Ala Val Val Ile Asn Ile
100 105 110
His His Tyr Glu Glu Leu Met Asn Asp Pro Glu Glu His Lys Glu Arg
115 120 125
Phe Leu Ala Leu Trp Lys Gln Ile Ala Asp Arg Tyr Lys Asp Tyr Pro
130 135 140
Glu Thr Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr
145 150 155 160
Pro Glu Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg
165 170 175
Ser Ile Asp Lys Lys His Thr Val Ile Ile Gly Thr Ala Glu Trp Gly
180 185 190
Gly Ile Ser Ala Leu Glu Lys Leu Arg Val Pro Lys Trp Glu Lys Asn
195 200 205
Ala Ile Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln
210 215 220
Gly Ala Glu Trp Val Pro Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp
225 230 235 240
Gly Ser Pro Asp Asp Gln Lys His Val Ile Glu Glu Phe Asn Phe Ile

```
                245                 250                 255
Glu Glu Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe
            260                 265                 270
Gly Ala Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser
        275                 280                 285
Phe Val Val Arg Glu Ala Glu Lys Arg Gly Trp Ser Trp Ala Tyr Trp
    290                 295                 300
Glu Phe Cys Ser Gly Phe Gly Val Tyr Asp Pro Leu Arg Lys Gln Trp
305                 310                 315                 320
Asn Lys Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu Ser
                325                 330                 335
Glu Lys Asp Glu Leu
            340
```