



(12) 发明专利申请

(10) 申请公布号 CN 104736701 A

(43) 申请公布日 2015. 06. 24

(21) 申请号 201380028620. X

H · D · 厄比纳 B · 张 M · A · 沃尔

(22) 申请日 2013. 03. 12

(74) 专利代理机构 北京纪凯知识产权代理有限公司

公司 11245

(30) 优先权数据

代理人 赵蓉民 陆惠中

61/618, 610 2012. 03. 30 US

61/660, 556 2012. 06. 15 US

(51) Int. Cl.

(85) PCT国际申请进入国家阶段日

C12N 9/42(2006. 01)

2014. 11. 28

(86) PCT国际申请的申请数据

PCT/US2013/030571 2013. 03. 12

(87) PCT国际申请的公布数据

W02013/148167 EN 2013. 10. 03

(71) 申请人 巴斯夫酶有限责任公司

地址 美国加利福尼亚州

(72) 发明人 X · 谭 K · E · 巴雷特

A · H · 达文波特 L · E · 惠普尔

权利要求书3页 说明书26页

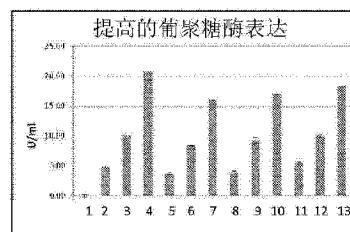
序列表17页 附图14页

(54) 发明名称

在极端井况下水解瓜尔胶压裂流体的纤维素酶的编码基因

(57) 摘要

提供编码热稳定的纤维素酶并导致其表达增加的多核苷酸序列, 和包括这种热稳定的纤维素酶的水力压裂组合物。公开的组合物包括聚合物增粘剂、表面活性剂、热稳定剂和酶分解剂, 所述酶分解剂包括衍生自嗜热细菌的野生型纤维素酶或其突变变体。在一些实施方式中, 增粘剂是瓜尔胶, 包括线性瓜尔胶、交联瓜尔胶或其混合物。在一些实施方式中, 酶分解剂特异性地水解所述瓜尔胶中的 β -1, 4 糖苷键。在一些实施方式中, 酶分解剂不特异性地水解所述瓜尔胶中的 α -1, 6 糖苷键。在一些实施方式中, 酶分解剂在上至约 275 °F 温度下保持其水解所述瓜尔胶中的 β -1, 4 糖苷键的能力。



图例

泳道1 = 阴性对照
泳道2、5、8、11 = SEQ ID NO:3
泳道3、6、9、12 = SEQ ID NO:4
泳道4、7、10、13 = SEQ ID NO:1

1. SEQ ID NO:1 的核苷酸序列。
2. 权利要求 1 所述的核苷酸序列,其中所述序列编码蛋白质。
3. 核苷酸序列,编码衍生自海栖热袍菌 (*Thermotoga maritima*) 的纤维素酶,其包括选自下列的至少 1 个突变: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、或其任意组合。
4. SEQ ID NO:3 的核苷酸序列,包括选自下列的至少 1 个突变: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、或其任意组合。
5. 权利要求 3 或 4 所述的核苷酸序列,其中至少 1 个突变是沉默的。
6. 权利要求 3 或 4 所述的核苷酸序列,其中至少 1 个突变导致具有至少 1 个突变的所述核苷酸序列指导所述纤维素酶以高于缺少至少 1 个突变并且不以其他方式区别于权利要求 3 或 4 所述的核苷酸序列的核苷酸序列的水平表达。
7. 来自海栖热袍菌的核苷酸序列,具有至少 1 个突变,所述至少 1 个突变使由所述核苷酸序列编码的蛋白质的表达水平与海栖热袍菌基因组序列相比增加。
8. 权利要求 7 所述的核苷酸序列,其中所述突变是沉默的。
9. 第一核苷酸序列,编码 SEQ ID NO:2 的多肽,其中所述核苷酸序列已相对于编码 SEQ ID NO:2 的第二序列被突变,使得所述蛋白质的表达水平相对于由所述第二核苷酸序列编码的所述蛋白质增加。
10. 核苷酸序列,编码与 SEQ ID NO:2 具有至少 50%,60%,70%,80%,90%,95%,97%,98%,99%或 100%同一性的蛋白质,或其片段,其中所述核苷酸序列包括选自下列的至少 1 个突变: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、或其任意组合。
11. 权利要求 1、2、3、4、5、6、7、8、9 或 10 所述的核苷酸序列,其编码具有纤维素酶活性的多肽,其中所述多肽在重组的细菌表达系统中生成。
12. 权利要求 11 所述的核苷酸序列,其中所述细菌表达系统是革兰氏阴性细菌表达系统。
13. 权利要求 12 所述的革兰氏阴性细菌表达系统,其中所述革兰氏阴性细菌是假单胞菌 (*Pseudomonas*)、大肠杆菌 (*E. coli*)、青枯菌 (*Ralstonia*)、或柄杆菌 (*Caulobacter*) 表达系统。
14. 权利要求 12 所述的革兰氏阴性细菌表达系统,其中所述假单胞菌表达系统是荧光假单胞菌 (*Pseudomonas fluorescens*) 表达系统。
15. 权利要求 12 所述的表达系统,其中所述纤维素酶生成至少 1.0g/L、2.0g/L、3.0g/L、4.0g/L、5.0g/L、6.0g/L、7.0g/L、8.0g/L、9.0g/L、10.0g/L、11.0g/L、12.0g/L、13.0g/L、14.0g/L、15.0g/L、16.0g/L、17.0g/L、18.0g/L、19.0g/L、20.0g/L、21.0g/L、22.0g/L、

23. 0g/L、24. 0g/L、25. 0、g/L、26. 0g/L、27. 0g/L、28. 0g/L、29. 0g/L、30. 0g/L、31. 0g/L、32. 0g/L、33. 0g/L、34. 0g/L、或 35. 0g/L。

16. SEQ ID NO:2 的多肽,其中所述氨基酸序列不包含信号序列、前蛋白序列、启动子序列、或其任意组合。

17. SEQ ID NO:2 的多肽,其中所述序列进一步包括异源序列。

18. 权利要求 17 所述的多肽,其中所述异源序列选自:信号序列、标签、表位、启动子序列、N 端延伸、C 端延伸、和其任意组合。

19. 组合物,包含 SEQ ID NO:2 的多肽,并且进一步包括第二酶。

20. 权利要求所述 19 的组合物,其中所述第二酶选自:乳糖酶、脂酶、蛋白酶、过氧化氢酶、木聚糖酶、纤维素酶、葡聚糖酶、甘露聚糖酶、淀粉酶、酰胺酶、环氧化物水解酶、酯酶、磷脂酶、转氨酶、胺氧化酶、纤维二糖水解酶、氨裂解酶、或其任意组合。

21. 分离的、重组的或合成的核苷酸,具有包含 SEQ ID NO:1 的核酸序列,其中所述核酸序列编码具有纤维素酶活性的多肽。

22. 分离的、重组的或合成的核苷酸,包含 SEQ ID NO:1 的核酸序列,其中所述核酸序列编码具有纤维素酶活性的多肽,并且所述多肽包含 SEQ ID NO:2 的氨基酸序列、或其酶活性片段。

23. 分离的、重组的或合成的核酸序列,包含 SEQ ID NO:1,其编码具有纤维素酶活性的多肽,其中所述多肽包含 SEQ ID NO:2 的氨基酸序列,并且所述多肽在重组荧光假单胞菌表达系统中生成。

24. 组合物,包括聚合物增粘剂、表面活性剂、热稳定剂和酶分解剂,所述酶分解剂包括衍生自嗜嗜热细菌的野生型纤维素酶或其突变变体。

25. 权利要求 24 所述的组合物,其中所述增粘剂是瓜尔胶,包括线性瓜尔胶、交联瓜尔胶、或其混合物。

26. 权利要求 25 所述的组合物,其中所述酶分解剂特异性地水解所述瓜尔胶中的 β -1,4 糖苷键。

27. 权利要求 25 所述的组合物,其中所述酶分解剂不特异性地水解所述瓜尔胶中的 α -1,6 糖苷键。

28. 权利要求 24-27 中任一项所述的组合物,其中所述酶分解剂在上至约 275 °F 温度下保持其水解所述瓜尔胶中的 β -1,4 糖苷键的能力。

29. 权利要求 24-28 中任一项所述的组合物,其中所述酶分解剂在上至约 11 的 pH 下保持其水解所述瓜尔胶中的 β -1,4 糖苷键的能力。

30. 权利要求 24-29 中任一项所述的组合物,其中所述酶分解剂是包括相对于所述野生型纤维素酶的 12 个突变的突变变体。

31. 权利要求 30 所述的组合物,其中所述 12 个突变选自 F38Y、Y61Q、M69E、D70P、R71S、I94Q、I166V、S183R、S191A、E212P、L231V、M276A、E277S、R280G、T297P 和 T301Q。

32. 权利要求 24 所述的组合物,其中所述酶分解剂具有 SEQ ID. NO. 2。

33. 权利要求 24 所述的组合物,其中所述酶分解剂由具有 SEQ ID. NO. 1 的多核苷酸编码。

34. 权利要求 24-29 中任一项所述的组合物,其中所述酶分解剂是所述野生型纤维素

酶的突变变体,并且具有如下熔融温度:在约 pH 6.5 下高于所述野生型纤维素酶的熔融温度至少 20 °F,和在约 pH 10.5 下高于所述野生型纤维素酶的熔融温度至少 10 °F。

35. 权利要求 24 所述的组合物,其中所述酶分解剂是由 SEQ ID NO:3 编码的纤维素酶的突变变体。

36. 降低包含 β -1,4 糖苷键的多糖凝胶的粘度的方法,所述方法包括使所述多糖凝胶在允许条件下接触纤维素酶变体,并且接触时间足以使所述纤维素酶水解所述多糖凝胶中的 β -1,4 糖苷键,从而降低所述瓜尔胶的粘度,其中所述纤维素酶变体与衍生自嗜热细菌的野生型纤维素酶相比包含至少 12 个突变,并且其中所述纤维素酶变体呈现与所述野生型纤维素酶相比增加的温度和 pH 耐受性。

37. 权利要求 36 所述的方法,其中至少 12 个突变通过基因位点饱和诱变生成,所述基因位点饱和诱变包括还原性重配、重组和选择的重复循环。

38. 权利要求 36 所述的方法,其中所述多糖凝胶包括线性瓜尔胶、交联瓜尔胶、或其混合物。

39. 权利要求 36 所述的方法,其中所述纤维素酶变体具有 SEQ ID NO:2。

40. 权利要求 36 所述的方法,其中所述纤维素酶变体由具有 SEQ ID. NO. 1 的多核苷酸编码。

41. 权利要求 36 所述的方法,其中允许条件包括约 180 °F 至约 275 °F 的温度范围。

42. 权利要求 36 所述的方法,其中允许条件包括约 9 至约 11 的 pH 范围。

43. 处理地下地层的方法,包括:

向所述地下地层引入压裂流体,所述压裂流体包括多糖凝胶和纤维素酶变体;和使所述凝胶和所述纤维素酶变体在允许条件下反应,并且反应时间足以使所述纤维素酶变体水解所述凝胶中的至少一些 β -1,4 糖苷键,而不水解 α -1,6 糖苷键,从而生成粘度降低的压裂流体;其中所述纤维素酶变体与衍生自嗜热细菌的亲本野生型纤维素酶相比包括至少 12 个突变,并且其中所述纤维素酶变体呈现与所述野生型纤维素酶相比增加的温度和 pH 耐受性。

44. 权利要求 36 所述的方法,其中通过所述纤维素酶变体生成的所述粘度降低的压裂流体,相对于通过化学分解剂生成的相当的粘度降低压裂流体,包括较少残留物,其中纤维素酶变体和化学分解剂的剂量提供基本上相同的粘度降低。

在极端井况下水解瓜尔胶压裂流体的纤维素酶的编码基因

[0001] 相关申请

[0002] 本申请根据 35U. S. C. § 119(e) 规定要求 2012 年 3 月 30 日提交的美国临时申请 61/618, 610 和 2012 年 6 月 15 日提交的美国临时申请 61/660, 556 的优先权, 其全部内容均被引入本文作为参考。

发明领域

[0003] 提供编码纤维素酶的多核苷酸序列。具体地, 提供的序列可提供特异性的、热稳定的、耐热的、压力稳定的纤维素酶, 如用于在极端井况下水解瓜尔胶压裂流体的纤维素酶的表达增加。

[0004] 序列表

[0005] 本申请通过在 MPEP § 502. 05 中授权和提出的 USPTO EFS-WEB 服务器进行电子提交, 并且该电子提交包括电子提交的序列表; 该序列表的全部内容被引入本申请的说明书作为参考。序列表在电子提交的 ASCII (.txt) 文档上被确定, 如下:

[0006]

文件名称	建立日期	大小
D2490_SEQLISTING	2013 年 3 月 9 日	40. 4kb

[0007] 发明背景

[0008] O-糖基水解酶 (EC 3. 2. 1. -) 是一大组天然存在的酶, 其水解两个或更多个糖之间或糖和非糖部分之间的糖苷键。糖基水解酶 (或转葡糖基酶 (glycosylase)) 的国际生物化学与分子生物学联盟 (International Union of Biochemistry 和 Molecular Biology, IUBMB) 酶命名主要基于其底物特异性, 并且有时基于其分子机制 (国际生物化学与分子生物学联盟命名委员会 (NC-IUBMB), 10/24/2011 访问)。

[0009] IUBMB 酶命名 EC 3. 2. 1. 4 已被指定为被称为“纤维素酶”的转葡糖基酶型酶子组。用于属于该组的酶其他名称包括: 内切葡聚糖酶、内切 -1, 4- β -葡聚糖酶、羧甲基纤维素酶、和 β -1, 4-葡聚糖酶。通过属于该组的酶催化的反应是纤维素、地衣多糖和谷类 β -D-葡聚糖 (如大麦 β -葡聚糖) 中的 1, 4- β -D-糖苷键的内切水解。由于本发明公开的纤维素酶的主要活性是大麦 β -葡聚糖和羧甲基纤维素的内切水解, 其被适当地归类于 IUBMB 酶命名 EC 3. 2. 1. 4。

[0010] 糖基水解酶的可选分类是基于氨基酸序列相似性 (Henrissat, B. 在 UniProt10/26/2011 访问)。根据该分类表, 糖基水解酶可被分成 70 多个家族。基于本发明公开的纤维素酶的主要氨基酸序列与公共数据库包含的其他糖基水解酶的序列的比较, 本发明公开的纤维素酶可被归属于糖基水解酶家族 5。该家族包含 20 多种内切葡聚糖酶 (IUBMB 酶命名 EC 3. 2. 1. 4), 其主要催化活性是纤维质底物中的 β -1, 4-糖苷键的内切水解。应用第二种酶分类方式为本发明公开的纤维素酶应被归类为 IUBMB 酶命名 EC 3. 2. 1. 4 的结论提供了进一步支持。

[0011] 纤维素酶用于多种工业和商业目的,包括但不限于,油和天然气勘探、食品和饮料、饮用醇或燃料醇生产,例如,酿造、乙醇、酒、调味剂、香味剂、织物、清洁剂、纸、浆、环境和农业以及研究目的(Rebecca S.Bryant,Erle C.Donaldson,Teh Fu Yen,George V.Chilingarian,Chapter 14Microbial 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,Pages423-450 ;M.Karmakar and R.R.Ray,2011.Current Trends in Research and Application of Microbial Cellulases.Research Journal of Microbiology,6:41-53)。

[0012] 自1990年代早期,酶分解剂已被成功用于水基压裂流体,以水解聚合物增粘剂(Brannon HD,Tjon-Joe-Pin RM,Carman PS,Wood WD,“Enzyme breaker technology:a decade of improved well stimulation”,文件SPE 84213,显示在SPE Annual Technical Conference and Exhibition in Denver,Colorado,USA,5-8October2003。

[0013] 发明概述

[0014] 酶是具有对存在于很多压裂操作中的较高温度和较高碱性条件敏感的三维结构的蛋白质。在复杂的井下条件下(例如,高温和pH)仍保持催化活性的酶的鉴定和/或改造将拓宽酶分解剂在水基压裂中的应用。

[0015] 公开了组合物,其包括聚合物增粘剂、表面活性剂、热稳定剂、和酶分解剂,该酶分解剂包括衍生自超嗜热细菌的野生型纤维素酶或其突变变体。

[0016] 在一些实施方式中,增粘剂是瓜尔胶,其包括线性瓜尔胶、交联瓜尔胶或其混合物。在一些实施方式中,酶分解剂特异性地水解瓜尔胶中的 β -1,4糖苷键。在一些实施方式中,酶分解剂不特异性地水解瓜尔胶中的 α -1,6糖苷键。在一些实施方式中,酶分解剂在上至约275°F的温度下保持其水解瓜尔胶中的 β -1,4糖苷键的能力。在一些实施方式中,酶分解剂在上至约11的pH下保持其水解瓜尔胶中的 β -1,4糖苷键的能力。

[0017] 在一些实施方式中,组合物中的酶分解剂是突变变体,其相对于野生型纤维素酶包括12个突变。12个突变可选自F38Y、Y61Q、M69E、D70P、R71S、I94Q、I166V、S183R、S191A、E212P、L231V、M276A、E277S、R280G、T297P和T301Q。在一些实施方式中,酶分解剂具有SEQ ID. NO. 2。在一些实施方式中,酶分解剂由具有SEQ ID. NO. 1的多核苷酸编码。在一些实施方式中,酶分解剂是野生型纤维素酶的突变变体,并且具有这样的熔融温度:在约pH 6.5下比野生型纤维素酶的熔融温度高至少20°F,和在约pH 10.5下比野生型纤维素酶的熔融温度高至少10°F。

[0018] 公开了降低包括 β -1,4糖苷键的多糖凝胶的粘度的方法。方法包括使多糖凝胶在允许条件下接触纤维素酶变体,并且接触时间足以使纤维素酶水解多糖凝胶中的 β -1,4糖苷键,从而降低瓜尔胶的粘度。纤维素酶变体,与衍生自超嗜热细菌的野生型纤维素酶相比,包括至少12个突变,并且纤维素酶变体呈现与野生型纤维素酶相比增加的温度和pH耐受性。

[0019] 在该方法的一些实施方式中,纤维素酶变体中的至少12个突变通过基因位点饱和和诱变生成,该基因位点饱和和诱变包括还原性重配、重组和选择重复循环。

[0020] 在该方法的一些实施方式中,多糖凝胶包括线性瓜尔胶、交联瓜尔胶或其混合物。

[0021] 在该方法的一些实施方式中,纤维素酶变体具有SEQ ID NO:2。在该方法的一些实

施方式中,纤维素酶变体由具有 SEQ ID NO. 1 的多核苷酸编码。

[0022] 在该方法的一些实施方式中,允许条件包括约 180 °F 至约 275 °F 的温度范围。在该方法的一些实施方式中,允许条件包括约 9 至约 11 的 pH 范围。

[0023] 还公开了处理地下地层的方法。方法包括:向地下地层引入包括多糖凝胶和纤维素酶变体的压裂流体;和使凝胶和纤维素酶变体在允许条件下反应并且反应时间足以使纤维素酶变体水解凝胶中的至少一些 β -1,4 糖苷键,而不水解 α -1,6 糖苷键,从而生成粘度降低的压裂流体。纤维素酶变体与衍生自嗜热细菌的亲本野生型纤维素酶相比可包括至少 12 个突变,并且纤维素酶变体可呈现与野生型纤维素酶相比增加的温度和 pH 耐受性。

[0024] 在该方法的一些实施方式中,通过纤维素酶变体生成的粘度降低压裂流体,相对于通过化学分解剂生成的相当的粘度降低压裂流体,包括较少的残留物,其中纤维素酶变体和化学分解剂的剂量提供基本上相同的粘度降低。

[0025] 在该方法的一些实施方式中,通过纤维素酶变体生成的粘度降低压裂流体保持的传导性大于通过化学分解剂生成的相当的粘度降低压裂流体,其中纤维素酶变体和化学分解剂的剂量提供基本上相同的粘度降低。

[0026] 在另一实施方式中,本发明包含 SEQ ID NO:1,其中所述序列编码蛋白质。在进一步实施方式中,本发明包括编码衍生自海栖热袍菌 (*Thermotoga maritima*) 的纤维素酶的核苷酸序列、或 SEQ ID NO:3,该序列包括选自下列的至少 1 个突变: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、或其任意组合,其中任选地,任何这样的突变是沉默的。本发明的进一步实施方式中,至少一个这种沉默突变导致所述纤维素酶的表达水平高于缺少至少一个这种突变的核苷酸序列。

[0027] 在本发明的另一实施方式中,本发明包括这样的来自海栖热袍菌的核苷酸序列:具有至少 1 个突变,并且与海栖热袍菌野生型基因组序列相比,具有增加的由所述核苷酸序列编码的蛋白质表达水平,其中任选地,所述突变(一个或多个)是沉默的。

[0028] 在本发明的另一实施方式中,本发明包括第一核苷酸序列,其编码 SEQ ID NO:2 的多肽,其中所述核苷酸序列已相对于编码 SEQ ID NO:2 的第二序列被突变,使得所述蛋白质的表达水平相对于由所述第二核苷酸序列编码的所述蛋白质增加。

[0029] 编码 SEQ ID NO:2 的多肽的第一核苷酸序列——其中所述核苷酸序列已相对于编码 SEQ ID NO:2 的第二序列被突变,使得所述蛋白质的表达水平相对于由所述第二核苷酸序列编码的所述蛋白质增加。

[0030] 在本发明的另一实施方式中,本发明包括编码与 SEQ ID NO:2 具有至少 50%、60%、70%、80%、90%、95%、97%、98%、99% 或 100% 同一性的蛋白质的核苷酸序列或其片段,其中所述核苷酸序列包括至少 1 个突变,该突变选自 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、或其任意组合。

[0031] 在本发明的另一实施方式中,本发明的蛋白质中的任一种在细菌表达系统中表达,其中细菌表达系统是革兰氏阴性细菌表达系统,例如,假单胞菌 (*Pseudomonas*)、大肠杆

菌 (*E. coli*)、青枯菌 (*Ralstonia*)、或柄杆菌 (*Caulobacter*) 表达系统。

[0032] 在本发明的另一实施方式中,本发明的纤维素酶的表达以至少 1.0g/L、2.0g/L、3.0g/L、4.0g/L、5.0g/L、6.0g/L、7.0g/L、8.0g/L、9.0g/L、10.0g/L、11.0g/L、12.0g/L、13.0g/L、14.0g/L、15.0g/L、16.0g/L、17.0g/L、18.0g/L、19.0g/L、20.0g/L、21.0g/L、22.0g/L、23.0g/L、24.0g/L、25.0g/L、26.0g/L、27.0g/L、28.0g/L、29.0g/L、30.0g/L、31.0g/L、32.0g/L、33.0g/L、34.0g/L、或 35.0g/L 生成。

[0033] 在本发明的另一实施方式中,本发明的纤维素酶组合第二酶,其中第二酶选自:乳糖酶、脂酶、蛋白酶、过氧化氢酶、木聚糖酶、纤维素酶、葡聚糖酶、甘露聚糖酶、淀粉酶、酰胺酶、环氧化物水解酶、酯酶、磷脂酶、转氨酶、胺氧化酶、纤维二糖水解酶、氨裂解酶或其任意组合。

[0034] 在本发明的另一实施方式中,本发明包括分离的、重组的或合成的核苷酸,其具有包含 SEQ ID NO:1 的核酸序列,其中该核酸序列编码具有纤维素酶活性的多肽。

[0035] 在本发明的另一实施方式中,本发明包括分离的、重组的或合成的核苷酸,其包含 SEQ ID NO:1 的核酸序列,其中该核酸序列编码具有纤维素酶活性的多肽和包含 SEQ ID NO:2 的氨基酸序列的多肽、或其酶活性片段。

[0036] 在本发明的另一实施方式中,本发明包括分离的、重组的或合成的核酸序列,其包括编码具有纤维素酶活性的多肽的 SEQ ID NO:1,其中该多肽包含 SEQ ID NO:2 的氨基酸序列,并且该多肽在重组荧光假单胞菌 (*Pseudomonas fluorescens*) 表达系统中生成。

[0037] 在本发明的另一实施方式中,本发明包括编码具有纤维素酶活性的多肽的核苷酸序列,其中该多肽在重组细菌表达系统中生成。

[0038] 在本发明的另一实施方式中,本发明包括组合物,其包括由 SEQ ID NO:1 编码的多肽,其中任选地,组合物进一步包括至少第二酶,和其中任选地,第二酶选自:乳糖酶、脂酶、蛋白酶、过氧化氢酶、木聚糖酶、纤维素酶、葡聚糖酶、甘露聚糖酶、淀粉酶、酰胺酶、环氧化物水解酶、酯酶、磷脂酶、转氨酶、胺氧化酶、纤维二糖水解酶、氨裂解酶或其任意组合。

[0039] 在本发明的另一实施方式中,本发明包括分离的、重组的或合成的核苷酸,其具有包含 SEQ ID NO:1 的核酸序列,其中该核酸序列编码具有纤维素酶活性的多肽。

[0040] 在本发明的另一实施方式中,本发明包括分离的、重组的或合成的核苷酸,其包含 SEQ ID NO:1 的核酸序列,其中该核酸序列编码具有纤维素酶活性的多肽,并且该多肽包含 SEQ ID NO:2 的氨基酸序列、或其酶活性片段。在进一步实施方式中,分离的、重组的或合成的包含 SEQ ID NO:1 的核酸序列编码具有纤维素酶活性的多肽,其中该多肽包含 SEQ ID NO:2 的氨基酸序列,并且该多肽在重组荧光假单胞菌表达系统中生成。

[0041] 在本发明的另一实施方式中,本发明包括聚合物,其包括聚合物增粘剂、表面活性剂、热稳定剂和酶分解剂,该酶分解剂包括衍生自嗜热细菌的野生型纤维素酶或其突变变体。在组合物的进一步实施方式中,增粘剂是瓜尔胶,包括线性瓜尔胶、交联瓜尔胶或其混合物。在组合物的进一步实施方式中,酶分解剂特异性地水解瓜尔胶中的 β -1,4 糖苷键。在组合物的进一步实施方式中,酶分解剂不特异性地水解瓜尔胶中的 α -1,6 糖苷键。在组合物的进一步实施方式中,酶分解剂在上至约 275 °F 的温度下保持其水解瓜尔胶中的 β -1,4 糖苷键的能力。在组合物的进一步实施方式中,其中酶分解剂在上至约 11 的 pH 下保持其水解瓜尔胶中的 β -1,4 糖苷键的能力。在组合物的进一步实施方式中,酶分解剂具

有 SEQ ID. NO. 2。在组合物的进一步实施方式中,酶分解剂由具有 SEQ ID. NO. 1 的多核苷酸编码。在上述组合物中任一种的进一步实施方式中,酶分解剂是野生型纤维素酶的突变变体,并且具有这样的熔融温度:在约 pH 6.5 下大于野生型纤维素酶的熔融温度至少 20 °F,和在约 pH 10.5 下大于野生型纤维素酶的熔融温度至少 10 °F。在进一步实施方式中,酶分解剂是突变变体,相对于野生型纤维素酶包括 12 个突变。在进一步实施方式中,变体纤维素酶包括下列突变中的至少一个:F38Y、Y61Q、M69E、D70P、R71S、I94Q、I166V、S183R、S191A、E212P、L231V、M276A、E277S、R280G、T297P 和 T301Q。

[0042] 在进一步实施方式中,所述酶分解剂是由 SEQ ID NO:3 编码的纤维素酶的突变变体。

[0043] 在进一步实施方式中,本发明包括降低包括 β -1,4 糖苷键的多糖凝胶的粘度的方法,方法包括使多糖凝胶在允许条件下接触纤维素酶变体并且接触时间足以允许纤维素酶水解多糖凝胶中的 β -1,4 糖苷键,从而降低瓜尔胶的粘度,其中纤维素酶变体与衍生自嗜嗜热细菌的野生型纤维素酶相比包括至少 12 个突变,并且其中纤维素酶变体呈现与野生型纤维素酶相比增加的温度和 pH 耐受性。在进一步实施方式中,所述 12 个突变通过基因位点饱和诱变生成,该基因位点饱和诱变包括还原性重配、重组和选择的重复循环。在可选的实施方式中,所述多糖凝胶包括线性瓜尔胶、交联瓜尔胶或其混合物。在可选的实施方式中,所述纤维素酶变体由具有 SEQ ID. NO. 1 的多核苷酸编码。在可选的实施方式中,其中允许条件包括约 180 °F 至约 275 °F 的温度范围。

[0044] 在可选的实施方式中,其中允许条件包括约 9 至约 11 的 pH 范围。

[0045] 在本发明的实施方式中,本发明包括处理地下地层的方法,包括:向地下地层引入压裂流体,该压裂流体包括多糖凝胶和纤维素酶变体;和使凝胶和纤维素酶变体在允许条件下反应,并且反应时间足以允许纤维素酶变体水解凝胶中的至少一些 β -1,4 糖苷键,而不水解 α -1,6 糖苷键,从而生成粘度降低的压裂流体;其中纤维素酶变体与衍生自嗜嗜热细菌的亲本野生型纤维素酶相比包括至少 12 个突变,并且其中纤维素酶变体呈现与野生型纤维素酶相比增加的温度和 pH 耐受性。

[0046] 在上述方法的进一步实施方式中,通过所述纤维素酶变体的水解包括少于通过化学分解剂生成的相当的粘度降低压裂流体的残留物,其中纤维素酶变体和化学分解剂的剂量提供基本上相同的粘度降低。

[0047] 在上述方法的进一步实施方式中,其中通过纤维素酶变体生成的粘度降低压裂流体保持的传导性大于通过化学分解剂生成的相当的粘度降低压裂流体,其中纤维素酶变体和化学分解剂的剂量提供基本上相同的粘度降低。

[0048] 发明描述

[0049] 酶是充当催化剂的蛋白质。蛋白质是在脱水反应中通过肽键连接的氨基酸聚合物。氨基酸的同一性及其连接形成蛋白质的次序决定给定蛋白质的活性。氨基酸组装成蛋白质的这种次序(蛋白质“序列”)最终由“编码”蛋白质的 DNA 链的序列决定。

[0050] 确定将要组装成蛋白质的给定氨基酸的三核苷酸序列被称为“密码子”。建成蛋白质的 20 种氨基酸由 64 个三核苷酸密码子序列共同编码。确定蛋白质的密码子系列被称为“开放阅读框”。氨基酸可由少如 1 个或多如 6 个不同的密码子确定。不影响被确定的氨基酸的密码子三核苷酸序列的变化(或突变)被称为“沉默的”突变。

[0051] 因此,存在多种能够编码相同蛋白质的 DNA 序列,因为 DNA 序列仅通过“沉默的”突变彼此区别。通过改变编码给定蛋白质的密码子中的一个或多个,可以在不影响被编码蛋白质的序列的情况下大幅增加基因生成的蛋白质量。

[0052] 在一些实施方式中,本发明包含 SEQ ID NO:1。在一些实施方式中,本发明包含 SEQ ID NO:1 的多核苷酸序列。在一些实施方式中,该序列编码蛋白质。在一些实施方式中,该蛋白质是具有纤维素酶活性的酶。

[0053] 本文公开的改进的核苷酸序列以 SEQ ID NO:1 给出,并且编码前文公开的演变自亲本纤维素酶的纤维素酶 (SEQ ID NO:2),该亲本纤维素酶分离自 DNA 库,源自海栖热袍菌菌株 MSB8。本公开的纤维素酶被描述于 PCT 公开号 W02009/020459,作为其中的 SEQ ID NO:9(由其中的多核苷酸 SEQ ID NO:8 编码,其在本文中被描述为 SEQ ID NO:3)。在一些实施方式中,本发明包含 SEQ ID NO:1 的多核苷酸序列、或其片段。在一些实施方式中,这些序列编码蛋白质。在一些实施方式中,蛋白质是具有纤维素酶活性的酶。

[0054] 本发明包括关于 SEQ ID NO:3 的多个核苷酸碱基变化。这些变化对于被编码蛋白质是沉默的。14 个碱基变化在下文中被提出。“位”表示开放阅读框中核苷酸的编号,其中第一密码子的第一核苷酸编号为 1。在 SEQ ID NO:1 的开放阅读框在 5' 端接合另一核酸序列以使开放阅读框延伸越过的 SEQ ID NO:1 的 5' 端的情况下,“位”将继续指代从 SEQ ID NO:1 的开放阅读框 5' 端开始编码的碱基。类似地,如果 SEQ ID NO:1 的开放阅读框被截短以使开放阅读框不起始于 SEQ ID NO:1 关联序列的 5' 端,则编码系统将继续从所述序列的 5' 端——相应于 SEQ ID NO:1 的 5' 端开始。

[0055] 核苷酸碱基变化或突变用标记(旧核苷酸)(位)(新核苷酸)指示。突变如下:T6C、T9C、T15G、A22C、G24T、A33C、A39C、A40C、A42C、A54C、A57C、T66C、G81A、A84C 或其任意组合。

[0056] 使 SEQ ID NO:1 区别于之前报道的编码本公开的纤维素酶的序列的碱基变化共同或单独生成开放阅读框,该开放阅读框导致蛋白质表达水平高于此前使用的编码相同蛋白质的核苷酸序列。

[0057] 在一些实施方式中,公开了编码衍生自海栖热袍菌的纤维素酶的核苷酸序列,其中核苷酸序列包括选自下列的至少 1 个突变: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 或其任意组合。在这些实施方式的一些方面中,至少 1 个突变对于被编码蛋白质的序列是沉默的。在其他方面中,至少 1 个突变造成核苷酸序列包括至少 1 个突变,其导致纤维素酶表达水平高于缺少该至少 1 个突变并且不以其他形式区别于上述核苷酸序列的核苷酸序列。

[0058] 在一些实施方式中,公开了编码纤维素酶的核苷酸序列,其中核苷酸序列包括 SEQ ID NO:3 并且具有选自下列的至少 1 个突变: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 或其任意组合。

[0059] 在一些实施方式中,公开了来自海栖热袍菌的核苷酸序列,其具有至少 1 个突变,

该至少 1 个突变使由所述核苷酸序列编码的蛋白质的表达水平与海栖热袍菌基因组序列相比增加。在一些方面中,至少 1 个突变是沉默的。

[0060] 在一些实施方式中,公开了编码 SEQ ID NO:2 的多肽的第一核苷酸序列,其中核苷酸序列已相对于编码 SEQ ID NO:2 的多肽的第二序列被突变,使得蛋白质表达水平相对于由第二核苷酸序列编码的蛋白质增加。

[0061] 海栖热袍菌是嗜热性真细菌,特征在于其在极端盐浓度(即,0.25% NaCl 至 6.00% NaCl)中生长的能力。海栖热袍菌属于栖热袍菌(Thermotogales)纲,其成员嗜热、棒状、厌氧和革兰氏阴性。最低生长温度在 55℃ 左右,最佳为 80°-85℃,和最高为约 90℃。在一些实施方式中,最低温度小于 55℃ 并且最高温度大于 90℃。这些细菌已从真细菌领域最深分支中的一种缓慢演变。栖热袍菌的成员已被描述为“广泛传播的和世界性的”(Huber, R. et al., 2006),在活跃的地热区中盛兴。海栖热袍菌与种新阿波罗栖热袍菌(*Thermotoga neapolitana*)、*Thermotoga petrophila*、和 *Thermotoga naphthophila* 紧密关联。海栖热袍菌的样本已得自下列中的海底:武尔卡诺,意大利;Riberia Quente 和 Sao Miguel 岛,亚速尔群岛;Sangeang 岛,印度尼西亚;和斐济岛。(Huber, R. et al., 2006)。

[0062] 菌株 MSB8 分离自处于意大利武尔卡诺的地热加热的海洋沉积物(Huber, 1986)。采集位置的温度范围为 70-100℃,并且 pH 为 6.5-7.0。该菌株已作为 DSM3109 被保藏在 Deutsche Sammlung von Mikroorganismen,和作为 ATCC43589 保藏在 ATCC(Huber, R. et al., 2006)。

[0063] 由于其生成的酶的优越热稳定性,已经研究了海栖热袍菌菌株 MSB8 的酶编码基因。Liebl(Liebl, W. et al., 1996)已经公开了“Analysis of a *Thermotoga maritima* DNA fragment encoding two similar thermostable cellulases, CelA and CelB, and characterization of the recombinant enzymes”。另外,淀粉分解酶(Bibel, M. et al., 1998)、逆旋转酶(Bouthier de la Tour, C. et al., 1998)、 α -淀粉酶(Liebl, W. et al., 1997)、 α -葡萄糖苷酸酶(Ruile, P. et al., 1997)、木聚糖酶(Winterh 改变, C. et al., 1995)、 β -葡萄糖苷酶(Liebl, W. et al., 1994)、葡聚糖转移酶(Liebl, W. et al., 1992)的基因已被分离和分析。Bronnenmeier 的研究(Bronnenmeier, K. et al., 1995),“Purification of *Thermotoga maritima* enzymes for the degradation of cellulosic materials”,已显示这些酶具有降解纤维素和木聚糖的价值。

[0064] 表达系统

[0065] 在一些实施方式中,编码本发明的纤维素酶的 DNA 可被引入——在质粒上或稳定地转化到——下列的基因组中:例如,任意数量的革兰氏阴性细菌系统,如大肠杆菌、假单胞菌属如荧光假单胞菌、恶臭假单胞菌(*Pseudomonas putida*)、铜绿假单胞菌(*Pseudomonas aeruginosa*)、青枯菌属(*Ralstonia*)、或柄杆菌属(*Caulobacter*)。类似地,纤维素酶可被引入任意数量的革兰氏阳性细菌表达系统中,如芽孢杆菌属(*Bacillus*)种如枯草芽孢杆菌(*Bacillus subtilis*)、巨大芽孢杆菌(*Bacillus megaterium*)、短芽孢杆菌(*Bacillus brevis*)、乳酸菌属(*Lactococcus*)种如乳酸乳球菌(*Lactococcus lactis*)、乳酸杆菌属(*Lactobacillus*)种、链霉菌属(*Streptomyces*)种如变铅青链霉菌(*Streptomyces lividans*)。可用其他革兰氏阴性、革兰氏阳性或不相关的真细菌或古细菌表达系统表达纤维素酶。

[0066] 在一些实施方式中, SEQ ID NO:1 被用于指导在多种可表达本公开的纤维素酶蛋白质的系统中增加表达水平。SEQ ID NO:1 可被引入任意数量的表达系统, 以在提高的累积水平上表达本公开的纤维素酶。例如, SEQ ID NO:1 可被引入——在质粒上或稳定地转化到——下列的基因组中: 例如, 任意数量的革兰氏阴性细菌系统如大肠杆菌、假单胞菌属如荧光假单胞菌、恶臭假单胞菌、铜绿假单胞菌、青枯菌属种、或柄杆菌属种。类似地, SEQ ID NO:1 可被引入任意数量的革兰氏阳性细菌表达系统, 如芽孢杆菌属种如枯草芽孢杆菌、巨大芽孢杆菌、短芽孢杆菌、乳酸菌属种如乳酸乳球菌、乳酸杆菌属种、链霉菌属种如变铅青链霉菌。可用其他革兰氏阴性、革兰氏阳性或不相关的真细菌或古细菌表达系统表达 SEQ ID NO:1。在进一步实施方式中, SEQ ID NO:1 可被引入任意数量的真核表达系统, 如酵母属 (*Saccharomyces*)、粟酒裂殖酵母 (*Schizosaccharomyces pombe*)、毕赤酵母 (*Pichia pastoris*) 和多形汉森酵母 (*Hansenula polymorpha*)。

[0067] 更具体地, SEQ ID NO:1 可被引入质粒, 以指导其表达。可引入 SEQ ID NO:1 的质粒包括, 例如, pQE、pET、和 pASK 家族的大肠杆菌表达载体; pCN51LT8、RSF1010、pWZ112T、和 pMYC 家族的假单胞菌表达载体; pBAX、pHT01 和 pHIS1525 家族的芽孢杆菌表达载体; pIJ6021 和 pIJ2460 家族的链霉菌表达载体; 和 pNZ9530 和 pNZ8148 家族的乳球菌表达载体, 例如。这些实例的目的是为了证明, 而不代表可表达 SEQ ID NO:1 的多核苷酸序列的整套载体。

[0068] 在一些实施方式中, 表达系统可以是本领域已知的任何荧光假单胞菌表达系统, 例如, 可购自 Dow Global Technologies Inc. 的荧光假单胞菌表达系统, 菌株 DC454 (美国专利公开申请号 20050130160 和美国专利公开申请号 20050186666)。编码纤维素酶或多肽的核酸序列被插入 pMYC 载体 (Dow Global Technologies Inc., 美国专利公开申请号 20050130160) 或 pDOW1169 载体 (Dow Global Technologies Inc., 美国专利公开申请号 20080058262), 然后通过电穿孔被引入荧光假单胞菌宿主。本领域技术人员将知道可用作本发明实施方式的可选载体。

[0069] 在一些实施方式中, 纤维素酶将至少以下列表达水平表达: 1.0g/L、2.0g/L、3.0g/L、4.0g/L、5.0g/L、6.0g/L、7.0g/L、8.0g/L、9.0g/L、10.0g/L、11.0g/L、12.0g/L、13.0g/L、14.0g/L、15.0g/L、16.0g/L、17.0g/L、18.0g/L、19.0g/L、20.0g/L、21.0g/L、22.0g/L、23.0g/L、24.0g/L、25.0g/L、26.0g/L、27.0g/L、28.0g/L、29.0g/L、30.0g/L、31.0g/L、32.0g/L、33.0g/L、34.0g/L、35.0g/L、或更多。

[0070] 核酸

[0071] 本发明提供分离的、合成的、或重组的核酸, 其包括完全与本发明的核酸序列互补的序列 (互补 (非编码) 和编码序列在下文中也被统称为本发明的核酸序列)。

[0072] 本发明提供分离的、合成的、或重组的核酸, 其包括编码至少一种具有纤维素分解活性的多肽的核酸, 其中该核酸包括这样的序列: 与包含 SEQ ID NO:1 的序列的本发明示例性核酸具有至少约 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%、或更多、或完全 (100%) 序列同一性 (同源性)。例如, 本发明提供分离的、合成的、或重组的核酸, 其包括核酸序列 SEQ ID NO:1 (本

发明的示例性多核苷酸序列)。本发明提供分离的、合成的、或重组的核酸,其编码包含 SEQ ID NO:2 所示序列(本发明的示例性多肽序列)的多肽、及其酶活性片段。

[0073] 多肽

[0074] 本发明的多肽和肽是分离的、合成的、或重组的多肽。肽和蛋白质可在体外或体内重组地表达。本发明的肽和多肽可利用本领域已知的任何方法制备和分离。本发明的多肽和肽也可全部或部分利用本领域公知的方法被合成。例如,纤维素酶多肽可在标准重组表达系统(如本文所述)中生成,被化学合成,或将其从自然表达的生物体中纯化。

[0075] 本发明提供分离的、合成的、或重组的具有纤维素分解活性的多肽,其包括与本发明的示例性氨基酸序列(例如,SEQ ID NO:2)具有至少约 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%、或更多、或具有 100%(完全)序列同一性的氨基酸序列、或其酶活性片段。

[0076] 本发明提供分离的、合成的、或重组的包含 SEQ ID NO:2 所示序列的多肽、及其酶活性片段、及其变体。

[0077] 在可选的实施方式中,本发明提供具有纤维素分解活性但缺少信号序列、前原结构域、锚定结构域和/或糖结合模块(CBM)的多肽(及其编码核酸);并且一方面,糖结合模块(CBM)包括下列或由下列组成:纤维素结合模块、木质素结合模块、木聚糖结合模块、木糖结合模块、甘露糖结合模块、木葡聚糖特异性模块和/或阿拉伯呋喃糖酶结合模块。

[0078] 在可选的实施方式中,本发明提供具有纤维素分解活性的多肽(及其编码核酸),其进一步包括异源序列;并且一方面,异源序列包括如下序列或由如下序列组成:所述序列编码:(i) 异源信号序列、异源糖结合模块、异源锚定结构域、异源催化结构域(CD)、或其组合;(ii) (i) 的序列,其中异源信号序列、糖结合模块或催化结构域(CD)衍生自异源酶;或(iii) 标签、表位、靶肽、可切割序列、可检测部分或酶;并且一方面,异源糖结合模块(CBM)包括下列或由下列组成:纤维素结合模块、木质素结合模块、木聚糖结合模块、木糖结合模块、甘露糖结合模块、木葡聚糖特异性模块和/或阿拉伯呋喃糖酶结合模块;并且一方面,异源信号序列使编码蛋白质靶向液泡、内质网、叶绿体或淀粉颗粒。

[0079] 酶活性

[0080] 通过本公开的纤维素酶进行的 pNP- β -D-吡喃乳糖苷(lactopyranoside)的酶水解可用作酶活性的度量。p-硝基苯酚的释放可在 405nm 下通过分光光度法跟踪。在 405nm 处吸光度的增加,通过利用在那些限定条件下的标准品吸光度,可转换成 p-硝基苯酚的 μ 摩尔数。1 单位活性被定义为在 pH 7.00 和 80°C 下 1 分钟期间从 2mM pNP- β -D-吡喃乳糖苷释放 0.42 μ 摩尔 p-硝基苯酚所需的酶量。(Advances in Carbohydrate Chemistry and Biochemistry, Academic Press, 1999)

[0081] 热稳定性

[0082] 在一些方面中,本发明的重组核酸编码热稳定的、具有纤维素分解活性的多肽。例如,本发明的多肽、SEQ ID NO:2、或本发明的变体演变的酶可以是热稳定的。根据本发明的热稳定多肽可在包括下列范围温度的条件下保持结合和/或酶活性,例如,纤维素分解活性:大于 37°C 至约 95°C、或约 55°C 至约 85°C 之间、或约 70°C 至约 75°C 之间、或约 70°C 至约

95℃之间、约 90℃至约 95℃之间、约 95℃至约 105℃之间、或约 95℃至约 110℃之间。在一些方面中,其中多肽可在下列条件下保持结合和 / 或酶活性,例如,纤维素分解活性:包括 1℃至约 5℃、约 5℃至约 15℃之间、约 15℃至约 25℃之间、约 25℃至约 37℃之间。在一些方面中,本发明的多肽可在下列条件下保持结合和 / 或酶活性,例如,纤维素分解活性:包括 90℃、91℃、92℃、93℃、94℃、95℃、96℃、97℃、98℃、99℃、100℃、101℃、102℃、103℃、103.5℃、104℃、105℃、107℃、108℃、109℃或 110℃、或更多。在一些实施方式中,根据本发明的热稳定多肽在上述范围的温度下、在下列酸性条件下保持活性,例如,纤维素分解活性:包括约 pH 6.5、pH 6、pH 5.5、pH 5、pH 4.5 或 pH 4 或更小(更酸性),或者,在下列暴露于酸性条件后保持纤维素分解活性:包括约 pH 6.5、pH 6、pH 5.5、pH 5、pH 4.5 或 pH 4 或更小(更酸性);或者,在下列碱性条件下保持活性:包括约 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 或更大(更碱性),或者在暴露于下列碱性条件后保持纤维素分解活性:包括约 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 或更多(更碱性)。

[0083] 耐热性

[0084] 在一些方面中,本发明的重组核酸编码耐热的、具有纤维素分解活性的多肽。例如,本发明的多肽、SEQ ID NO:2、或本发明的变体演变酶可以是耐热的。在一些方面中,纤维素分解活性是耐热的,例如,其中多肽在暴露于下列范围的温度后保持纤维素分解活性:大于 37℃至约 95℃、或约 55℃至约 85℃之间、或约 70℃至约 75℃之间、或约 70℃至约 95℃之间、约 90℃至约 95℃之间、约 95℃至约 105℃之间、或约 95℃至约 110℃之间。在一些方面中,其中多肽在暴露于包括下列温度范围的条件后保持纤维素分解活性:包括约 1℃至约 5℃之间、约 5℃至约 15℃之间、约 15℃至约 25℃之间、约 25℃至约 37℃之间。在一些方面中,本发明的多肽可在暴露于上至 90℃、91℃、92℃、93℃、94℃、95℃、96℃、97℃、98℃、99℃、100℃、101℃、102℃、103℃、104℃、105℃、107℃、108℃、109℃或 110℃、或更高的温度后保持纤维素分解活性。在一些方面中,由本发明的核酸编码的多肽在下列酸性条件下保持纤维素分解活性:包括约 pH 6.5、pH 6、pH 5.5、pH 5、pH 4.5 或 pH 4 或更小(更酸性),或在下列暴露于酸性条件后保持纤维素分解活性:包括约 pH 6.5、pH 6、pH 5.5、pH 5、pH 4.5 或 pH 4 或更小(更酸性);或在下列碱性条件下保持纤维素分解活性:包括约 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 或更多(更碱性)。

[0085] 纤维质消化

[0086] 在一些方面中,本发明的组合物和方法用于生物质的酶消化,并且可包括使用多种不同的酶,包括纤维素酶和半纤维素酶。用于实践本发明的纤维素酶可将纤维素消化成葡萄糖。在一些方面中,用于实践本发明的组合物可包括酶混合物,该酶例如是,木聚糖酶、木糖苷酶(例如,β-木糖苷酶)、纤维二糖水解酶、和 / 或阿拉伯呋喃糖苷酶、或其他可将半纤维、纤维素和木质纤维素材料消化以发酵糖和 / 或成单体糖的酶。

[0087] 本发明的酶,例如,内切葡聚糖酶,被用于消化纤维素或任何 β-1,4-连接的葡聚糖——包括合成或天然物质,包括在任何植物材料中发现的那些物质。本发明的酶,例如,内切葡聚糖酶,被用作商品酶,以消化来自任何来源的纤维素,包括所有生物来源,如植物生物质,例如,玉米、谷物、草类(例如,印度草,如蓝滨麦 (*Sorghastrum nutans*));或,柳枝

稷 (switch grass), 例如, 黍属 (*Panicum species*), 如柳枝稷 (*Panicum virgatum*)), 或, 木材或木材处理副产物, 例如, 在木材处理中、纸浆和 / 或纸业、在织物制造和家庭和工业清洗剂中和 / 或在生物质废料处理中。

[0088] 饮食

[0089] 在一些实施方式中, 本发明的纤维素酶可用于预处理、改变、消化动物或人类食物、食物添加剂、或饮食补充剂、或增强动物或人类食物、食物添加剂、或饮食补充剂的消化。在一些实施方式中, 本发明的纤维素酶可用作动物或人类的食物、食物添加剂或饮食补充剂。在一些方面中, 纤维素酶将治疗消化性疾病或充当消化性疾病预防手段。在本发明的另一方面, 纤维素酶将改变或增强消化。在本发明的另一方面, 纤维素酶将增强、改变或辅助食物消化。在本发明的进一步方面, 纤维素酶将增强、辅助或改变食物的营养价值。在进一步方面, 纤维素酶活跃在消化道中, 例如, 胃和 / 或肠中。

[0090] 在一些实施方式中, 本发明的纤维素酶可用作动物饲料或动物饲料添加剂。在一些实施方式中, 纤维素酶的热稳定性和或耐热性允许丸粒形成, 而无需第二试剂如盐或蜡。包括纤维素酶的动物饲料可以本领域技术人员已知的任何制剂被提供给动物。动物饲料制剂的实例包括但不限于: 递送基质、丸粒、片剂、凝胶、液体、喷雾、磨碎谷粒、或粉末。

[0091] 本发明提供可食用的酶递送基质, 其包括热稳定的重组纤维素酶, 例如, 本发明的多肽。本发明提供递送纤维素酶补充剂至动物的方法, 方法包括: 制备丸粒形式的可食用酶递送基质, 其包括颗粒状可食用载体和热稳定重组纤维素酶, 其中丸粒容易将包含在其中的纤维素酶分散到水介质中; 和向动物给予可食用酶递送基质。重组纤维素酶可包括本发明的多肽。颗粒状可食用载体可包括选自下列的载体: 谷类胚芽、无油谷类胚芽、干草、苜蓿、蒂莫西 (timothy)、大豆皮、葵花籽粕和小麦次粉 (wheat midd)。可食用的载体可包括无油谷类胚芽。纤维素酶可以被糖基化, 以提供在造粒条件下的热稳定性。递送基质可通过使包括谷类胚芽和纤维素酶的混合物成粒而形成。造粒条件可包括施加蒸汽。造粒条件可包括施加超过约 80°C 的温度约 5 分钟, 并且酶保持每毫克酶至少 350 至约 900 单位的比活性。

[0092] 制备乙醇的方法

[0093] 本发明提供制备乙醇的方法, 包括使含淀粉组合物接触具有纤维素分解活性的多肽——其中多肽具有本发明的序列, 或多肽由包括本发明的序列的核酸编码——或其酶活性片段。本发明提供组合物, 其包括淀粉和具有纤维素分解活性的多肽——其中多肽具有本发明的序列, 或多肽由包括本发明的序列的核酸编码——或其酶活性片段。

[0094] 酿造和发酵

[0095] 本发明提供酿造 (例如, 发酵) 包括本发明的纤维素酶的啤酒的方法。在一个示例性方法中, 将含淀粉原材料崩解和处理以形成麦芽。本发明的酶在发酵过程中的任意点施用。本发明的纤维素酶可以用于酿造业, 以降解 β -葡聚糖。在一些方面中, 本发明的纤维素酶用于酿造业, 以净化饮料。本发明的酶可由于提高麦芽汁或啤酒过滤性而用于饮料业, 如例如美国专利号 4,746,517 描述。

[0096] 在一些方面中, 本发明的纤维素酶用于淀粉糖化和转化过程。在酿造和发酵业中, 淀粉糖化和转化过程在过低温度下进行, 从而不能促进水溶性葡聚糖、甘露聚糖、阿拉伯木聚糖或木聚糖、或其他多糖充分降解。这些聚合物形成胶质底物, 其可由于过滤无效和提取

率低而导致淀粉糖化麦芽汁的粘度增加,造成麦芽浆耗竭时间较长,最终啤酒产品中残留混浊和沉淀。

[0097] 在一些方面中,本发明的纤维素酶用于麦芽厂操作,例如,添加葡聚糖酶至处理水,以缩短发芽时间和 / 或促进差品质大麦向可接受麦芽转化。在一些方面中,本发明的酶用于淀粉糖化,例如,其被添加以增加麦芽汁过滤性和 / 或提高过滤(从麦芽浆分离麦芽汁)。在一些方面中,本发明的酶用于发酵罐和 / 或沉淀池,从而例如有助于混浊清除和 / 或改进过滤。在一些方面中,本发明的酶用于辅助酿造,例如,添加本发明的葡聚糖酶以分解葡聚糖、甘露聚糖、阿拉伯木聚糖或木聚糖、或其他来自大麦、小麦和 / 或其他谷类的多糖,包括麦芽中的聚糖。在一些方面中,本发明的酶用于麦芽酿造,例如,添加本发明的葡聚糖酶以通过高葡聚糖含量来改性不良麦芽。

[0098] 本发明的纤维素酶可用于任何啤酒或酒精饮料生产方法,如例如美国专利号 5,762,991 ;5,536,650 ;5,405,624 ;5,021,246 ;4,788,066 所述。

[0099] 处理食物和食物处理

[0100] 本发明的纤维素酶在食物处理工业中具有多种用途。例如,一方面,本发明的酶用于改进油从富油植物材料例如富油种子的提取,例如,大豆油从大豆,橄榄油从橄榄,油菜籽油从油菜籽和 / 或葵花籽油从葵花籽的提取。

[0101] 本发明的纤维素酶可用于分离植物细胞材料的组分。例如,本发明的酶可用于将富葡聚糖材料(例如,植物细胞)分离成各种组分。在一些方面中,本发明的酶可用于将富葡聚糖或富油作物分离成有价值的蛋白质和油和壳部分。分离过程可通过利用本领域已知的方法进行。

[0102] 本发明的纤维素酶可用于制备水果或蔬菜汁、糖浆、提取物及类似物,以增加产率。本发明的酶可用于酶处理(例如,水解含葡聚糖植物材料)不同植物细胞壁衍生材料或废料——例如,来自谷类、谷物、酒或果汁生产、或农业残留物如蔬菜皮、豆荚、甜菜浆、橄榄浆、马铃薯浆及类似物。本发明的酶可用于改变处理后的水果或蔬菜的一致性和表观。本发明的酶可用于处理植物材料,以有助于处理植物材料,包括食物,有助于纯化或提取植物组分。本发明的纤维素酶可用于提高饲料价值,降低水结合能力,提高废水工厂的降解性,和 / 或提高植物材料向青贮饲料的转化,及类似作用。本发明的纤维素酶也可用于水果和酿造工业的设备清洁和维护。

[0103] 清洁剂组合物

[0104] 本发明提供包括一种或多种本发明的多肽的清洁剂组合物以及制备和应用这些组合物的方法。本发明包括所有制备和应用清洁剂组合物的方法,参见,例如,美国专利号 6,413,928 ;6,399,561 ;6,365,561 ;6,380,147。清洁剂组合物可以是一部分和两部分含水组合物、不含水液体组合物、铸造固体、颗粒形式、微粒形式、压缩片剂、凝胶和 / 或糊剂和浆液形式。本发明还提供能够利用这些清洁剂组合物迅速去除食物粗土(gross food soils)、食物残留物膜和其他少量食物组成的方法。本发明的酶可通过催化水解淀粉多糖促进去除淀粉类污渍。本发明的酶可用于餐具清洗清洁剂和织物洗涤清洁剂。实际活性酶含量取决于清洁剂组合物的制备方法并且不是严格的——假设清洁剂溶液具有预期酶活性。在一些方面中,最终溶液中存在的葡萄糖苷酶量的范围为每克清洁剂组合物约 0.001mg 至 0.5mg。选用于本发明的方法和产品的具体酶取决于最终实用条件,包括物理产品形式、

应用 pH、应用温度、和待降解或改变的油类型。酶可被选择以提供对于任何给定组实用条件而言的最佳活性和稳定性。本发明的清洁剂可包括阳离子型、半极性非离子型或两性离子型表面活性剂；或其混合物。

[0105] 本发明提供清洁组合物，其包括用于清洁硬质表面的清洁剂组合物、用于清洁织物的清洁剂组合物、餐具清洗组合物、口腔清洁组合物、假牙清洁组合物、和隐形眼镜清洁液。在一些方面中，本发明提供清洗物体的方法，包括使物体在足以清洗的条件下接触本发明的多肽。本发明的多肽可被包括作为清洁剂添加剂。本发明的清洁剂组合物可例如被配制为包括本发明的多肽的手洗或机洗清洁剂组合物。适于有污织布的预处理的洗衣添加剂可包括本发明的多肽。织布软化剂组合物可包括本发明的多肽。可选地，本发明的多肽可被配制为用于一般家庭硬质表面清洁操作的清洁剂组合物。

[0106] 油和气的勘探和清理

[0107] 为增加油和气和页岩气储层的生产力，被称为“水力压裂”的高度专业化技术被渐多地应用。在一般的水力压裂操作中，大体积的瓜尔胶基流体（凝胶形式，并且被称为“压裂流体”）在极高的静水压力下被泵送到井眼中。加压流体在井眼周围的地层中产生新裂纹和裂缝。压裂流体中包含的沙粒移动并沉降到新产生的裂缝中，并起作用以使这些通道打开，因此增加油和气流。在沙子沉积到裂缝中后，必须使凝胶降解（即，分解）并向上返回至地表，从而去除油或气流动的任何阻碍。工业利用降粘剂（如氧化剂、酸、或酶）降解压裂流体和从裂纹并且裂缝去除任何固体凝胶残留物。

[0108] 在一些实施方式中，本公开的纤维素酶将用作高温降粘剂，以增强油和气操作。更具体地，本发明公开的纤维素酶将在油或气井中进行水力压裂时被施用于压裂流体。

[0109] 由 SEQ ID NO:1 编码的酶，以及由本文公开的其他多核苷酸编码或通过本文公开的方法可获得的纤维素酶，可潜在地用于水解宽范围多糖——其中多种可用于油和气的勘探、压裂和井清理操作。本公开的纤维素酶呈现宽范围的 β -糖苷酶活性，例如，针对瓜尔胶、羟丙基瓜尔胶、羧甲基瓜尔胶、羧甲基羟丙基瓜尔胶、羧甲基纤维素、大麦 β -葡聚糖和刺槐豆胶。酶活性模式优选内切和外切，这允许通过从多糖长链中切割和通过从聚合物末端切割二糖单位，有效降低多糖，例如，瓜尔胶和衍生化的瓜尔胶溶液的粘度。在上述多糖之外，本公开的酶的其他底物包括能够形成线性或交联凝胶的那些。适当的多糖底物的实例包括半乳甘露聚糖胶 (glactomannan gums)、瓜尔胶、衍生化的瓜尔胶、纤维素和纤维素衍生物、淀粉、淀粉衍生物、黄原胶、衍生化的黄原胶和其混合物。具体实例还包括但不限于，瓜尔胶、瓜尔胶衍生物、刺槐豆胶、梧桐胶、黄原胶、纤维素和纤维素衍生物等。本公开的酶可涉及的一般的聚合物或胶凝剂包括瓜尔胶、羟丙基瓜尔胶、羧甲基羟丙基瓜尔胶、羟乙基纤维素、羧甲基羟乙基纤维素、羧甲基纤维素、二烷基羧甲基纤维素等。聚合物的其他实例包括但不限于，磷酸甘露聚糖、硬葡聚糖、右旋糖酐和其他聚合物类型。在一些实施方式中，聚合物底物是羧甲基羟丙基瓜尔胶。在一些实施方式中，公开的酶还可有效水解生物胶（例如，由枣浆或蔗糖制成的琥珀酰聚糖生物胶）。在一些实施方式中，公开的酶可用于水解含纤维素的或衍生化的含纤维素的聚合物，一般地，酶攻击纤维素骨架的糖苷键。本公开的酶可适于使聚合物降解成大部分单糖单位，在一些情况下，通过特异性地水解任何纤维二糖片段的 (1, 4)- β -D-糖苷键中的单糖单位和纤维素骨架之间的外 (1, 4)- β -D-糖苷和内 (1, 4)- β -D-糖苷键。

[0110] 在使用本公开的纤维素酶的每个压裂工作中,现场操作人员一般首先在工业实验室中进行酶剂量优化研究。这种研究可包括稀释纤维素酶至 10 ~ 400ppm 的浓度并与线性或交联瓜尔胶 (25 - 60lb/1,000gal) 混合。根据施用条件,可利用交联剂交联瓜尔胶,尤其对于存在较高温度、压力和 pH 条件的井而言。然后这种优化研究产生的酶剂量信息被用于实际压裂工作。

[0111] 本公开的纤维素酶的独特活性允许瓜尔胶基压裂流体以平滑和可控的方式在存在高温和高 pH 条件的深井中水解。与化学分解剂相比,本发明公开的纤维素酶提供替代苛刻的并且非选择性的化学分解剂的、无腐蚀性和环境友好的选择。

[0112] 在本发明一些实施方式中,纤维素酶可用于处理、清洁或改变用于油和气勘探活性的流体。在进一步方面,本发明的纤维素酶将部分或全部处理或改变流体,使得流体可被再次使用或再循环,用于另外的油和气勘探活性,或以环境友好的方式被处置。

[0113] 本文描述了用于包括油和气压裂程序的应用的组合物。在一些实施方式中,组合物包括与聚合物增粘剂组合的、衍生自超嗜热细菌和 / 或其非天然存在变体的一种或多种纤维素酶。在一些实施方式中,组合物可任选地包括另外的成分,包括控制流体损失、降低地层损害、调节 pH、控制微生物生长和提高温度稳定性的试剂。一些典型的、另外的成分包括酸、抗菌剂、粘土稳定剂、腐蚀抑制剂、交联剂、减摩剂、铁质控制、阻垢剂、表面活性剂和热稳定剂 (例如,硫代硫酸钠)。

[0114] 在一些实施方式中,本文描述的纤维素酶具有葡聚糖酶——例如,内切葡聚糖酶、甘露聚糖酶、木聚糖酶——活性或这些活性的组合。在一些方面中,葡聚糖酶活性是内切葡聚糖酶活性 (例如,内切-1, 4- β -D- 葡聚糖 4- 葡聚糖水解酶活性),并且包括水解纤维素、纤维素衍生物 (例如,羧甲基纤维素和羟乙基纤维素) 地衣多糖中的 1, 4- β -D- 糖苷键、混合 β -1, 3 葡聚糖如谷类 β -D- 葡聚糖或木葡聚糖和其他包含纤维质部分的植物材料中的 β -1, 4 键。在可选的方面,这些葡聚糖酶,例如,内切葡聚糖酶、甘露聚糖酶、木聚糖酶,具有增加的活性和稳定性,包括在增加或减少的 pH 和温度下的耐热性或热稳定性。

[0115] 在一些实施方式中,所述酶分解剂可被封装,以稳定酶,提高热稳定性和碱性 pH 耐受性,和提供受控释放。被封装的分解剂——具有水解性降解的涂层或膜,可优于现有技术酶分解剂系统,因为其可允许更好地控制释放时间和前所未有地便于处理。例如,由于分解剂被封装在与水发生反应而非仅仅溶解或消散在水中的材料中,释放可通过涂层与水的反应速率被控制。同样,通过隔离酶与苛刻的井下条件 (高温和 pH) 一定时间段,可提供延迟和完全的粘度下降。本领域技术人员将理解,涂层反应速率 (进而,分解剂释放特征) 可根据所用封装聚合物的化学而宽范围地改变。分解剂封装组成和方法的实例被提供于美国专利号 5,164,099、6,163,766、5,373,901、5,437,331 和 6,357,527,其公开内容均被引入本文作为参考。

[0116] DNA 序列和被编码的纤维素酶序列

[0117] 衍生自超嗜热细菌和 / 或其非天然存在变体的一些纤维素酶被描述于 PCT 公开 WO 2009/02049 ;其全部公开内容被引入本文作为参考。下列 DNA 和氨基酸 SEQ ID NOS 被包括 WO 2009/02049 公开的完整说明书内,其全部内容被引入本文作为参考。这些包括:

[0118] • WO 2009/02049SEQ ID NOS :1、2 (野生型 ‘亲本’ 海栖热袍菌纤维素酶),在本文中公开为 SEQ ID NOS:5 和 6。

[0119] • WO 2009/02049SEQ ID NOS :3(野生型 DNA,改变以去除候补起点),在本文中公开为 SEQ ID NO:7。

[0120] •WO 2009/02049SEQ ID NOS :6、7(“7X”组合基因位点饱和诱变(“GSSM”)突变)在本文中公开为 SEQ ID NOs:8 和 9。

[0121] •WO 2009/02049SEQ ID NOS :8、9(“12X-6”组合 GSSM 突变),在本文中公开为 SEQ ID NOs:3 和 2。

[0122] •WO 2009/02049SEQ ID NOS :10、11(“13X-1”组合 GSSM 突变),在本文中公开为 SEQ ID NOs:10 和 11。

[0123] •WO 2009/02049SEQ ID NOS :12、13(“12X-1”组合 GSSM 突变),在本文中公开为 SEQ ID NOs:12 和 13。

[0124] • WO 2009/02049SEQ ID NOS :16、17(可选的纤维素酶分解剂,来自栖热袍菌属(*Thermotoga* sp.)),在本文中公开为 SEQ ID NOs:14 和 15。

[0125] •WO 2009/02049SEQ ID NOS :18、19(用于玉米表达的海栖热袍菌纤维素酶的“7X”密码子优化版本),在本文中公开为 SEQ ID NOs:16 和 17。

[0126] • WO 2009/02049SEQ ID NOS :20、21(用于玉米表达的海栖热袍菌纤维素酶的“12X-6”密码子优化版本),在本文中公开为 SEQ ID NOs:18 和 19。

[0127] 除上列来自海栖热袍菌菌株 MSB8 的纤维素酶的野生型和演变变体的相关核苷酸和氨基酸序列外,表 2 和实施例 5(WO 2009/02049,并在下文中引用或再现)列举的另外的突变体也被认为可用作本文所述组合物和/或制备这些组合物的方法中的组分。

[0128] 酶活性分析(WO 2009/02049 的实施例 5)

[0129] 以下实施例描述示例性酶,WO 2009/02049 中鉴定为 SEQ ID NO:2 的“亲本”或“野生型”蛋白质的变体,和证明其活性的数据;附图和数据被引入作为参考。

[0130] 提供这样的序列:其相对于“亲本”(或“野生型”)SEQ ID NO:6(相应于 WO 2009/02049SEQ ID NO:2)——例如由 SEQ ID NO:5(相应于 WO 2009/02049SEQ ID NO:1)编码——具有特定的残基变化,如上文表 1(部分)和下文表 2 和 3 概述。

[0131] 表 2

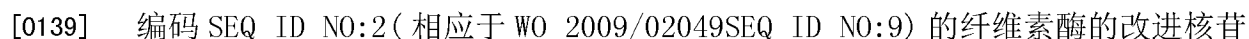
[0132]

位置	38	61	69	70	71	94	166	183	191	212	231	276	277	280	297	301
突变	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-	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	

[0133] 表 3

[0134]



酸序列被公开于 2012 年 3 月 30 日提交的美国临时申请号 61/618,610 ;其全部公开内容被引入本文作为参考。SEQ ID NO:2 的这种纤维素酶演变自从源自海栖热袍菌菌株 MSB8 的 DNA 库分离的亲本纤维素酶。本公开的纤维素酶的表达增强涉及相对于此前报道的开放阅读框的 14 个碱基变化,从而生成 SEQ ID NO:1(下文)。这些变化对于被编码蛋白质是沉默的。14 个碱基变化在下文中被提出。“位”表示开放阅读框中的核苷酸编号,其中第一密码子的第一核苷酸编号为 1。突变利用标记(旧核苷酸)(位)(新核苷酸)指示。突变如下:T6C、T9C、T15G、A22C、G24T、A33C、A39C、A40C、A42C、A54C、A57C、T66C、G81A、A84C。

[0140] 共同和单独地使 SEQ ID NO:1 区别于现有技术报道的编码本公开的纤维素酶的序列的碱基变化产生这样的开放阅读框:其导致蛋白质表达水平高于此前使用的编码相同蛋白质的核苷酸序列。

[0141] SEQ ID NO:1 用于使多种可表达本公开的纤维素酶蛋白质(SEQ ID NO:2)的系统中的表达水平增加。SEQ ID NO:1 可被引入任意数量的表达系统,以在提高的累积水平上表达本公开的纤维素酶。例如,SEQ ID NO:1 可被引入——在质粒上或稳定地转化到——下列的基因组中:例如,任意数量的革兰氏阴性细菌系统,如大肠杆菌、假单胞菌属种如荧光假单胞菌、恶臭假单胞菌、铜绿假单胞菌、青枯菌属种、或柄杆菌属种。类似地,SEQ ID NO:1 可被引入任意数量的革兰氏阳性细菌表达系统,如芽孢杆菌属种如枯草芽孢杆菌、巨大芽孢杆菌、短芽孢杆菌、乳球菌属种如乳酸乳球菌、乳酸杆菌属种、链霉菌属种如变铅青链霉菌。其他革兰氏阴性、革兰氏阳性或不相关的真细菌或古细菌表达系统可用于表达 SEQ ID NO:1。

[0142] 更具体地,SEQ ID NO:1 可被引入质粒,以指导其表达。可引入 SEQ ID NO:1 的质粒包括,例如,pQE、pET、和 pASK 家族的大肠杆菌表达载体;pCN51LT8、RSF1010、pWZ112T、和 pMYC 家族的假单胞菌表达载体;pBAX、pHT01、和 pHIS1525 家族的芽孢杆菌表达载体;pIJ6021 和 pIJ2460 家族的链霉菌表达载体;和乳球菌:例如 pNZ9530 和 pNZ8148 家族的表达载体。这些实例以证明为目的,而不代表可表达 SEQ ID NO:1 的多核苷酸序列的整套载体。

[0143] 增粘剂和其他添加剂

[0144] 在一些实施方式中,聚合物增粘剂可选自下列中一种或多种:瓜尔胶、交联瓜尔胶、羟丙基瓜尔胶、羧甲基瓜尔胶、羧甲基羟丙基瓜尔胶、羧甲基纤维素、大麦 β -葡聚糖和刺槐豆胶。在一些实施方式中,本公开的纤维素酶水解模式是内切和外切,这允许通过多糖长链中切割和通过切割聚合物末端的二糖单位而有效降低多糖例如瓜尔胶和衍生化的瓜尔胶溶液的粘度。除上述多糖外,本公开的酶的其他底物还包括能够形成线性或交联凝胶(例如,硼酸盐(酯)交联瓜尔胶)的那些。适当的多糖底物的实例包括半乳甘露聚糖胶、瓜尔胶、衍生化的瓜尔胶、纤维素和纤维素衍生物、淀粉、淀粉衍生物、黄原胶、衍生化的黄原胶和其混合物。具体实例还包括但不限于,瓜尔胶、瓜尔胶衍生物、刺槐豆胶、梧桐胶、黄原胶、纤维素和纤维素衍生物等。本公开的酶可涉及的一般的聚合物增粘剂或胶凝剂包括瓜尔胶、羟丙基瓜尔胶、羧甲基羟丙基瓜尔胶、羟乙基纤维素、羧甲基羟乙基纤维素、羧甲基纤维素、二烷基羧甲基纤维素等。其他聚合物实例包括但不限于,磷酸甘露糖、硬葡聚糖、右旋糖酐和其他聚合物类型。在一些实施方式中,聚合物底物是羧甲基羟丙基瓜尔胶。在一些实施方式中,公开的酶还可有效水解生物胶(例如,由枣浆或蔗糖制成的琥珀酰聚糖

生物胶)。在一些实施方式中,公开的酶可用于水解含纤维素或衍生化的含纤维素聚合物,一般地,酶攻击纤维素骨架的糖苷键。本公开的酶可适于使聚合物降解成大部分单糖单位,在一些情况下,通过特异性地水解任何纤维二糖片段的(1,4)- β -D-糖苷键中的单糖单位和纤维素骨架之间的外(1,4)- β -D-糖苷和内(1,4)- β -D-糖苷键。

[0145] 本文描述的压裂流体组合物还可包括另外的组分,包括控制流体损失、降低地层损害、调节和/或缓冲pH、控制微生物生长和提高温度稳定性的试剂。交联剂可包括硼酸盐(酯)、钛酸盐(酯)、锆酸盐(酯)和铝化合物。一般的水力压裂混合物包括约98%水和沙和约2%其他添加剂,该添加剂选自酸、抗菌剂、分解剂(酶的或化学的/氧化性的)、粘土稳定剂、腐蚀抑制剂、交联剂、减摩剂、胶凝剂(增粘剂)、铁质控制、pH调节剂、阻垢剂和表面活性剂。

[0146] 压裂流体中本公开的纤维素酶的性能

[0147] 在使用本公开的纤维素酶的每个压裂工作中,现场操作人员一般首先在工业实验室中进行酶剂量优化研究。本公开的纤维素酶将被稀释至10~400ppm的浓度并与多糖凝胶形成化合物,例如线性或交联瓜尔胶(25-60lb/1,000gal)混合。根据施用条件,通常利用硼酸盐(酯)交联剂交联瓜尔胶或衍生化瓜尔胶,尤其对于存在较高温度、压力和pH条件的井而言。这种优化研究产生的酶剂量信息将被用于实际压裂工作。

[0148] 本公开的纤维素酶的独特活性特征允许瓜尔胶基压裂流体以平滑和可控的方式在存在高温和高pH条件的深井中完全水解。与化学分解剂相比,本发明公开的纤维素酶提供替代苛刻的并且非选择性的化学分解剂的、无腐蚀性和环境友好的选择。

[0149] 酶分解剂已多年被用于在150°F以下的温度下水解多糖增粘剂,包括例如,瓜尔胶。但是,关于可在较高温度(例如,200-250°F)和极端pH(≥ 10.5)条件下工作的酶分解剂存在行业范围要求。为解决这种要求,发明人已研发出包括异常热稳定的纤维素酶的压裂流体。还参见PCT公开WO 2009/02049,其描述蛋白质和编码蛋白质的DNA等等。已显示公开的演变自亲本酶的纤维素酶在气页岩和较深油/气井中遭遇的极端井下条件下呈现明确差异的性能。

[0150] 本公开的纤维素酶可在宽范围温度(80°F上至250°F)和pH(3.0上至10.5)下有效分解线性和硼酸盐(酯)交联的瓜尔胶。流变性测试的结果显示,仅需少量(100ppm或更小)实现完全分解。酶反应可通过压裂操作期间的温度和pH的变化触发。本公开的纤维素酶还呈现优异的剂量响应,其允许操作人员通过调节酶剂量建立理想的粘度-时间图谱。剂量依赖性表现将证实高度有益的,因为其避免过早粘度损失和不期望的支撑剂筛出。即使在流体添加剂如缓冲剂、盐、稳定剂、交联剂等存在的情况下,本公开的纤维素酶仍保持有效粘度降低的活性。

[0151] 本公开的纤维素酶分解剂通过特异性地靶向瓜尔胶中的甘露糖单元之间的 β -1,4糖苷键降低凝胶粘度。糖分析测试证明,这种酶有效和高效地使长链瓜尔胶聚合物分解成小的可溶的片段,其将消除凝胶再愈合(re-healing)。传导性测试证明了瓜尔胶的完全水解和导致地层损害并降低井传导性的聚合物残留物的去除。

[0152] 作为瓜尔胶分解反应的生物催化剂,本公开的纤维素酶呈现甘露聚糖酶活性,相对于传统氧化性化学分解剂具有若干显著优势。第一,酶通过内切或外切- β 甘露聚糖酶活性特异性地分解长链瓜尔胶聚合物,而对井眼、地层或压裂设备无不期望的反应。第

二,酶不在反应中被消耗,并且可在其寿命期间继续作用于其他瓜尔胶聚合物,因此提供延长的和受控的分解图谱。第三,酶可使瓜尔胶聚合物分解成小很多的片段,因此提供较充分的瓜尔胶分解和较少的残留物。因此,科学家已做出尝试以从极端自然环境下发现和表征超热稳定的纤维素酶。海栖热袍菌属,从热液喷口样品鉴定的嗜嗜热细菌,具有纤维素酶,其具有内切-甘露聚糖酶活性,可特异性地切割长链瓜尔胶的 β -1,4键连接的多糖。Mathur EJ, Lam DE, “Carboxymethyl cellulase from *Thermotoga maritima*”, 美国专利号 5,962,258(1999)、6,008,032(1999) 和 6,245,547(2001)。具体地,这种酶证明最佳温度为 180 °F,其显著高于其他已知的糖苷水解酶家族的内切甘露聚糖酶。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。

[0153] 通过差示扫描量热法 (DSC) 评价的纤维素酶热稳定性

[0154] 酶具有对于其催化功能关键的三维结构。在特定条件(热、变性剂、极端 pH)下,酶可展开并丧失其催化活性三维结构。熔融温度(T_m)是 50%蛋白质展开的温度。该参数广泛用于评价酶或蛋白质的热稳定性。 T_m 可通过差示扫描量热法(DSC)测量。

[0155] DSC 测试利用纤维素酶野生型蛋白质和选定的纤维素酶变体进行。所有蛋白质样品在 1mg/ml 的浓度和 1°C/min 的扫描速率下分析。每次扫描的温度范围为 160 ~ 250 °F。在所有 DSC 实验期间保持 4.6atm 的恒定压力,防止加热时可能的溶液脱气。通过可用软件包计算 T_m 。图 7 显示野生型和选定的纤维素酶变体的 T_m 结果。观察到与野生型酶相比最终纤维素酶候选的 T_m 值增加 24 °F。 T_m 与 pH 有关。变体和野生型之间的 T_m 差异在 pH 6.5 下为 24 °F,而该差异在 pH 10.5 下为 13 °F。

[0156] 利用化学替代进行的纤维素酶变体的温度特征

[0157] 在不同温度下进一步评价纤维素酶变体的 1-4 键切割活性。分子 pNP- β -D-吡喃乳糖苷(图 8A)用作替代底物。pNP- β -D-吡喃乳糖苷通过纤维素酶进行的酶水解可生成游离 p-硝基苯酚,其可通过分光光度法在 405nm 下测量。图 8B 呈现纤维素酶变体上至分光光度计温度极限 194 °F 的温度图谱。如图 8B 所示,这种纤维素酶变体具有 180 °F 的最佳温度。

[0158] 为评价在沸腾温度以上(>212 °F)的纤维素酶活性,建立分析以定量测量纤维素酶变体的剩余活性。将酶与 251b 瓜尔胶每 1000gal(ppt)以 1:1 的比混合,然后分别在 225 °F、250 °F 和 275 °F 下加热挑战 10、20 和 30 分钟。随后,通过如下利用 pNP 吸光度分析测量加热挑战的酶的活性:添加 30uL 加热处理的酶至 3000uL 2mM pNP-B-D-吡喃乳糖苷,然后测量 405 吸光度 10 分钟。图 8C 显示纤维素酶变体在 225 °F、250 °F 和 275 °F 下的剩余活性。酶被加热处理 10-30 分钟时,在 225 °F 下观察到稳健的纤维素酶活性。当被加热处理的纤维素酶被施加于硼酸盐(酯)交联瓜尔胶(251b/1000gal)时,在小剂量(20gpt)下观察到完全的凝胶分解。纤维素酶活性在 250 °F 和 275 °F 下受损,其中相对剩余活性为在

225 °F 下的剩余活性的 11-14% (图 8C), 表明小百分比的纤维素酶即使在高温 (>250 °F) 处理后仍可进行可测的催化反应。

[0159] 利用荧光替代进行的纤维素酶变体的 pH 特征

[0160] 分子 4- 甲基伞形基 (Methylumbelliferyl) β -D- 纤维二糖苷是可在较高 pH 条件下被纤维素酶切割的替代底物 (图 8A)。切割产物, 4- 甲基伞形酮, 可利用 365nm 和 455nm 的激发和辐射波长通过荧光测定定量。将 2 毫摩 4- 甲基伞形基 β -D- 纤维二糖苷溶解于缓冲系统中, 其中 pH 值范围为 9.5 至 10.5。通过添加 30uL 纤维素酶到 3000uL 的 2mM 4- 甲基伞形基 β -D- 纤维二糖苷中, 不同 pH 条件下的纤维素酶活性在分光光度计中进行动态测量。各反应曲线的斜率通过线性曲线拟合获得。根据在 pH 9.5 下获得的结果计算相对活性。在 pH 10.5 下——与硼酸盐 (酯) 交联瓜尔胶流体高度相关的 pH 条件, 纤维素酶呈现合理的活性 (图 9)。

[0161] 通过 HPLC 确定 β -1, 4- 键而非 α -1, 6- 键多糖的酶分解

[0162] 瓜尔胶, 由甘露糖和半乳糖类组成的长链多糖, 是水基压裂操作中的主要增粘剂。聚甘露糖通过 β -1, 4- 键形成长链骨架, 而半乳糖单元通过 α -1, 6 键连接于甘露糖单元。甘露糖与半乳糖的比可在 1.6:1 至 1.8:1 范围内。重要地, 瓜尔胶的聚甘露糖骨架不可溶于水, 而半乳糖分支显著增加水溶解度。但是, 已经报道少如 6 个连续未分支甘露糖单元可形成完全不可溶的局部螺旋结构 (11)。因此, 如果纤维素酶在瓜尔胶分子的 α -1, 6 键处分解, 可生成大量不可溶的残留物。因此, 支撑剂封填 (proppant pack) 的传导性可被削弱。

[0163] 为确定纤维素酶变体的酶活性特异性地靶向多糖的 β -1, 4 糖苷键, 将正常相 HPLC 法用于分析两种替代底物, 1, 4- β -D- 甘露五糖和 63, 64- α -D- 半乳糖基 - 甘露五糖 (图 10A)。将 3 单位 /mL 的纤维素酶变体与 1mg/mL 底物在 80°C、pH 7.0 下一起温育 1 小时。利用酰胺柱、乙腈 /H₂O/ 三乙胺作为流动相、和 ELSD 检测器, 获得通过 HPLC 的甘露寡糖基线分离。图 10B 示例不同甘露寡糖——包括 M1、M2、M3、M4、M6 和半乳糖标准品 (M1-M6 表示甘露寡糖链长度)——的 HPLC 洗脱图谱。图 4C 显示 β -1, 4-D- 甘露五糖——其结构中仅具有 β -1, 4 键的甘露寡糖——如何被纤维素酶变体消化的实例。HPLC 图谱证明 β -1, 4- 键切割, 并且优选生成 M3 寡糖, 并且生成 M1、M2 和 M4, 确定纤维素酶变体在分解甘露寡糖 (图 10C(a)-(b)) 时特异性地攻击 β -1, 4 键。另一方面, 关于 63, 64- α -D- 半乳糖基 - 甘露五糖——包含 β -1, 4- 键和 α -1, 6- 键的化合物, HPLC 图谱显示, 纤维素酶变体仅分解 β -1-4 键并生成 OGGM4 峰和单一的单位甘露糖峰 (图 10C(a))。利用半乳糖作为参考, 未观察到来自酶处理的 6³, 6⁴- α -D- 半乳糖基 - 甘露五糖的单一单位半乳糖的峰, 确定 α -1, 6 键在甘露聚糖酶处理后保持完整 (图 10C(b))。

[0164] 利用分解剂制备交联瓜尔胶流体

[0165] 线性瓜尔胶流体通过向在掺合器中充分搅拌生成涡流的水添加干燥聚合物制备。搅拌持续约 30 分钟以充分水合。利用氢氧化钠溶液调节 pH。对于交联流体, 粘土稳定剂和清理表面活性剂, 在使用时, 在瓜尔胶充分水合后被添加。然后添加苛性碱, 然后酶添加。最后添加交联剂。硼酸盐 (酯) 交联剂的作用通过活性硼物种的缓慢溶解被延迟。锆交联剂也由于存在以 6ppt 使用的倍半碳酸盐缓冲液被延迟。

[0166] 瓶测试

[0167] 利用过硫酸铵 (APS) 或纤维素酶分解剂制备瓜尔胶或交联瓜尔胶罐。该罐在水浴

中留置,并通过倾斜瓶和观察流体移动来视觉评估分解水平。瓶测试能够实现快速评价浓度、pH 和温度对给定压裂流体分解的影响的方法。由于测试实质上是定性的,结果未显示。

[0168] 流变性

[0169] 粘度计利用杯轮技术 (cup and bob technique) (转子 1 和轮 5) 通过模型 50 规格在 400psi 的氮气压力下测量粘度。流体温度通过油浴增加至预期值。加热用时约 20 分钟以达到预期温度并且该温度在测试期间保持。连续粘度测量在这样的旋转速度下进行: 传递 100s^{-1} 的壁面剪切速率,其中周期性剪切率为 100、75、50、25、50、75 和 100s^{-1} 。利用剪切率数据计算幂律参数 (n' 和 K'),用于处于常在裂缝中发现的较低剪切速率下的粘度预测。

[0170] 图 11 证明在 180°F 下分解 80lb/1000 加仑线性瓜尔胶溶液的活性范围。纤维素酶浓度保持恒定在 200ppm。随着 pH 从 9 增加至 11,观察到分解显著延迟,与酶在升高的 pH 水平下的较低活性一致。即使在 pH 11 下,与无酶的对照相比,溶液也显示缓慢分解效应。

[0171] 图 12 显示在 200°F 和 pH 10.5 下酶浓度对分解硼酸盐 (酯) 交联的瓜尔胶流体的影响。纤维素酶剂量增加导致粘性破坏更显著。流变性结果进行数小时以显示粘性随时间持续损失。对于交联凝胶显而易见的是,与包含较高浓度的瓜尔胶的线性凝胶相比,需要更多酶。具有 50ppm 纤维素酶水平的交联流体的粘度无法与交联流体对照的粘度区别,但较高水平的酶的确显示分解活性。

[0172] 重要地,当以 10ppt 的水平包括热稳定剂时,酶在 225°F 和 pH 10 下仍具有活性 (图 13)。对照显示渐进的热退化,但酶显著促进粘度下降。图 14 显示利用羧甲基羟乙基纤维素 (CMHPG) 的关于铅交联流体的纤维素酶和封装式过硫酸铵在 225°F 和 pH 10 下的比较数据。100ppm 的浓度提供与过硫酸铵类似的分解特征,而 200ppm 则造成即时粘度损失。注意,封装式过硫酸铵通过热降解和 / 或扩散穿过造成延迟分解的封装材料释放活性材料中的一些。

[0173] 残留物分析

[0174] 通过过硫酸铵或纤维素酶分解剂制备瓜尔胶或交联瓜尔胶罐。将罐在 180°F 下在水浴中留置过夜。将内含物在 3000rpm 下离心 5 分钟,并通过标称 5 微米的称重后的纸真空过滤。将纸在 110°F 下干燥过夜,并再称重,以计算残留物的回收量。该值以溶液中聚合物的原始重量的百分比表示。然后将滤纸加残留物再干燥 16 小时,并再称重,以确保水分已被去除。但是,数据表明,另外的 16 小时干燥时间看起来是不必要的。测试的流体包括在 pH 10 下的 80ppt 线性瓜尔胶。纤维素酶在 200 和 400ppm 下测试,同时过硫酸铵以 1 和 5ppt 应用。

[0175] 结果显示,令人惊讶地,包含纤维素酶的溶液具有的残留物量低于相应的具有过硫酸铵的溶液 (图 15)。相反,氧化性分解剂呈现较随机地切割和生成较高水平的不可溶片段——其被捕获为残留物。如果将罐在温度下 (即, 180°F) 留置较长时间段,则有可能酶处理流体的残留物甚至进一步减少,而包含氧化性分解剂的流体则不。对于过硫酸铵而言,残留物对分解剂浓度敏感,而对酶浓度的敏感性则低得多。

[0176] 传导性研究

[0177] 裂缝传导性利用具有固定底部活塞的改型 API 室来测量。条件为在漏失后 180°F , 3000psi 闭合应力, 2lbm/ft² 支撑剂负荷 (20/40 目渥太华沙) 和约 5X 浓度的压裂流体。流

体留置在温度下过夜,然后 4 小时清理,然后最少 1 小时封填流动。利用 2wt% KCl 溶液,每小时清理交替流动通过封填物和通过核心。最终传导性——在 4% 内稳定 1 小时——的测量利用处于 3mL/min 速率的 2wt% KCl 进行。通过与利用 2wt% KCl 代替压裂流体以相同条件测量的传导性进行比较,计算保持的传导性。

[0178] 通过过硫酸铵或纤维素酶运行线性瓜尔胶和硼交联瓜尔胶的测试以及关于利用这两种分解剂的交联瓜尔胶的一个测试。对于线性瓜尔胶,通过纤维素酶保持的传导性水平高于过硫酸铵。类似地,对于交联的硼酸盐(酯)流体,封装式过硫酸铵提供的保持传导性低于 200ppm 纤维素酶(图 16)。纤维素酶和封装式过硫酸盐的组合具有甚至高于单独任一种分解剂的保持传导性。但是,最高结果通过 400ppm 剂量的酶获得。由于传导性测试过夜进行(shut-in),允许纤维素酶额外时间降低瓜尔胶的分子量。这将直接导致支撑剂封填的传导性增强。

[0179] 总之,演变的纤维素酶变体与硼酸盐(酯)交联瓜尔胶流体中的不同添加剂相容。具体地,这种纤维素酶变体在 180-225 °F 的温度范围和上至 10.5 的 pH 范围内显示稳健的瓜尔胶分解特征。pH 限制防止在较高温度范围内通过硼酸盐(酯)交联的瓜尔胶流体测试这种酶。但是,在 225 °F 的温度下利用铅交联 CMHPG 观察到正向瓜尔胶分解特征。显而易见地,残留物分析证明,甘露聚糖酶处理的流体生成的残留物明显少于过硫酸铵处理的流体。更重要地,纤维素酶处理实现了保持传导性明显高于过硫酸铵处理,确定了如下观点:酶相对于化学氧化剂可提供更充分的瓜尔胶分解和改善的井况。此外,纤维素酶的应用与过硫酸铵的添加相容,如果其可用于某些油田应用。

[0180] 附图简述

[0181] 图 1 是 SDS PAGE 凝胶电泳图,显示不同水平的蛋白质表达,如实施例 2 所述。

[0182] 图 2 是显示蛋白质制剂活性水平的柱状图,如实施例 3 所述。

[0183] 图 3 是 SEQ ID NO:1,为与 SEQ ID NO:3 相比具有 14 个沉默突变:T6C、T9C、T15G、A22C、G24T、A33C、A39C、A40C、A42C、A54C、A57C、T66C、G81A、A84C 的多核苷酸。

[0184] 图 4 是 SEQ ID NO:2,由 SEQ ID NO:1、3 和 4 编码的多肽。

[0185] 图 5 是 SEQ ID NO:3,SEQ ID NO:1 和 4 的未修饰亲本多核苷酸序列。

[0186] 图 6 是 SEQ ID NO:4,为与 SEQ ID NO:3 相比具有 14 个沉默突变:T6C、T9C、T15G、A22C、G24T、A33C、A39C、A40C、A42C、A54C、A57C、T66C、G81A、A84C 和起始密码子上游(所示另外的上游序列)的一个另外的点突变的多核苷酸。

[0187] 图 7. 进行野生型纤维素酶和选定纤维素酶变体的 DSC 测试,其中标记计算的 T_m 值。与野生型纤维素酶相比,通过纤维素酶变体观察到 T_m 的 24 °F 增量。

[0188] 图 8A. pNP-β-D-吡喃乳糖苷的化学结构。

[0189] 图 8B. 在不同温度(70 °F 至 200 °F)下通过 pNP 吸光度分析(405nm)测量的纤维素酶的温度图谱。活性以在 180 °F 下的活性的百分比计算。在沸腾温度以上(>212 °F)活性可通过剩余活性分析进一步评价(图 8C)。

[0190] 图 8C. 利用热挑战样品来评价纤维素酶变体的剩余活性。相对活性计算为在 225 °F 下获得的活性的百分比。

[0191] 图 9. 利用荧光分析进行的纤维素酶变体的 pH 图谱。相对活性通过在 pH 9.5 下的活性的百分比计算。

[0192] 图 10A. HPLC 研究的两种化合物的结构。1, 4- β -D- 甘露五糖是具有通过 1, 4- β 键连接的 5 单位甘露糖的化合物。6³, 6⁴- α -D- 半乳糖基 - 甘露五糖是模拟瓜尔胶结构的化合物, 其中两个半乳糖单位通过 α -1, 6 键连接于 1, 4- β -D- 甘露五糖, 其允许进行纤维素酶处理的水解产物分析。

[0193] 图 10B. 寡糖标准品 (M1-M6 和半乳糖) 的 HPLC 分析。注意, HPLC 洗脱图谱中甘露糖 1 单位 (M1, 红色峰) 与 1 单位半乳糖 (蓝色峰) 明确分离。

[0194] 图 10C. 甘露糖处理后 1, 4- β -D- 甘露五糖和 6³, 6⁴- α -D- 半乳糖基 - 甘露五糖的水解产物的 HPLC 分析。

[0195] a) 1, 4- β -D- 甘露五糖的酶消化。无酶处理的对照样品的 HPLC 洗脱图谱以红色显示, 而纤维素酶变体处理的 1, 4- β -D- 甘露五糖以蓝色显示。

[0196] b) 6³, 6⁴- α -D- 半乳糖基 - 甘露五糖的酶消化。O-GGM5 表示 6³, 6⁴- α -D- 半乳糖基 - 甘露五糖, 其中纤维素酶处理样品以红色显示并且无酶样品以蓝色显示。1 单位甘露糖 (以粉色显示) 和半乳糖 (以绿色显示) 被包括作为参考。注意, 纤维素酶处理的 O-GGM5 的 HPLC 洗脱图谱中无半乳糖峰, 只有甘露糖峰, 证明纤维素酶变体特异性地切割甘露糖单元之间的 β -1, 4 糖苷键。

[0197] 图 11. 在 180 °F 下 pH 对 80 lbm 瓜尔胶 /1000 gal US 的粘度的影响。为清楚起见, 剪切率已被去除。

[0198] 图 12. 在 200 °F、pH = 10.5 下基于硼酸盐 (酯) 交联瓜尔胶的纤维素酶变体流变研究。所有流体均包含粘土稳定剂和表面活性剂并应用延迟交联剂。

[0199] 图 13. 在 225 °F、pH = 10 下基于硼酸盐 (酯) 交联瓜尔胶的纤维素酶变体流变研究。所有流体均包含 10 lbm 热稳定剂 /1000 gal US 并应用延迟交联剂。

[0200] 图 14. 在 225 °F 和 pH = 9.9 下、以 10 lbm 热稳定剂 /1000 gal US、利用锆交联 CMHPG 进行的纤维素酶变体流变研究。

[0201] 图 15. 比较过硫酸铵与纤维素酶的残留物分析。较高分解剂水平在两种情况下均使残留物减少, 但氧化分解剂显示残留物量较高。

[0202] 图 16. 针对线性和硼交联瓜尔胶制剂的纤维素酶与过硫酸铵的传导性结果比较。线性: 80 lbm 瓜尔胶 /1000 gal US, 利用纤维素酶或暴露式 (live) 过硫酸铵; 交联: 30 lbm 瓜尔胶 /1000 gal US, 利用纤维素酶或封装式过硫酸铵。

[0203] 术语定义:

[0204] “纤维素酶”是具有纤维素酶、内切葡聚糖酶、纤维二糖水解酶、 β - 葡萄糖苷酶、木聚糖酶、甘露聚糖酶、 β - 木糖苷酶、阿拉伯呋喃糖苷酶和 / 或寡聚酶活性的酶。

[0205] “纤维素分解活性”是具有纤维素酶、内切葡聚糖酶、纤维二糖水解酶、 β - 葡萄糖苷酶、木聚糖酶、甘露聚糖酶、 β - 木糖苷酶、阿拉伯呋喃糖苷酶和 / 或寡聚酶活性的酶。

[0206] “密码子”是三核苷酸序列, 其确定所要加入蛋白质的氨基酸的同一性。

[0207] “沉默突变”是密码子中不导致不同氨基酸确定的突变。

[0208] “开放阅读框”是确定蛋白质中的氨基酸序列的密码子系列。

[0209] 碱基“位”是多核苷酸序列中碱基的位置编号, 从开放阅读框起点或从一些其他参考标志连续计数。

[0210] “编码”蛋白质意为指定该蛋白质的氨基酸序列。

[0211] “突变”是与参考相比核苷酸序列或氨基酸序列的变化。

[0212] “核苷酸”指代四种碱基中的一种,包括 DNA 序列——腺嘌呤 (A)、胸腺嘧啶 (T)、鸟嘌呤 (G) 和胞嘧啶 (C)。

[0213] “海栖热袍菌基因组序列”指代由 GenBank 登录号 AE000512 确定的海栖热袍菌菌株 MSB8 基因组序列。

[0214] 给定蛋白质的“表达水平”是通过表达系统如转化细胞培养物生成的、以单位体积细胞培养物测量的蛋白质量。

[0215] 给定酶的“表达水平”是通过表达系统如转化细胞培养物生成的、以单位体积细胞培养物测量的酶活性量。

[0216] “野生型”指代可在自然中获得的蛋白质或核酸序列。

[0217] 实施例 1——制备表达增强变体的方法。

[0218] 基于 SEQ ID NO:3 设计两种变体 (SEQ ID NO:1 和 NO:4),在 DNA 水平上突变,从而提高基因表达。设计考虑可影响基因表达的多种因素。利用本领域技术人员已知的 PCR 技术将突变引入 PCR 引物。两种基因均进行 PCR 扩增,并利用标准分子克隆技术被克隆到假单胞菌载体 pDOW1169 (DOW AgroSciences, IN) 中。将所得表达构建体转化到荧光假单胞菌 DC454 (DOW AgroSciences, IN) 中。将带有 SEQ ID NO:1 的转化体认定为代表,因为其显示最为增强的表达。

[0219] 实施例 2——利用 SDS-PAGE 凝胶电泳和非特异性蛋白质染色视觉化由包括 SEQ ID NOs :1、3 和 4 的构建体表达的 SEQ ID NO:2 多肽的表达水平。

[0220] 利用 Criterion™预先浇注的 Tris-HCl 聚丙烯酰胺凝胶 (Bio-rad Laboratories, Inc.) 分离蛋白质。利用 Tris- 甘氨酸缓冲液将凝胶在 150V 下运行 (参见图 1)。将蛋白质负载标准化为在每个泳道负载来自 0.33OD₆₀₀ 细胞的蛋白质。使用 SeeBlue®预先染色的蛋白质标准品 (Life Technologies)。将凝胶用非特异性染料染色,并且每个泳道目视检查是否存在约 37 千道尔顿大小的 SEQ ID NO:2 的带。

[0221] 结果显示存在这样的单个带:所有样品具有不同累积水平并且阴性对照无累积水平。该带具有 SEQ ID NO:2 的预期大小。

[0222] 在相应于具有包含 SEQ ID NO:1 的构建物的细胞的蛋白质提取物的泳道中,该带的累积水平明显高于 SEQ ID NO:3 或阴性对照,并且 SEQ ID NO:4 较小程度高于 SEQ ID NO:3 或阴性对照。

[0223] 实施例 3——确定变体相对表达水平的方法。

[0224] 将包含 SEQ ID NO:1、SEQ ID NO:3 和 SEQ ID NO:4 基因的核酸序列转化到适当的宿主细胞中,以表达 SEQ ID NO:2 的蛋白质。在烧瓶中培养细胞,使得被编码蛋白质被表达。使培养物在设计的复合培养基中、在 30℃ 和 220rpm 下生长至 ~ 0.9 的 OD₆₀₀,并引入 0.3mM IPTG (异丙基 β-D-1- 硫代吡喃半乳糖苷) 24 小时。收集细胞,并将其通过超声或在 80℃ 下热处理 1 小时裂解。利用 pNP-β-D- 吡喃乳糖苷作为底物,通过基于 p- 硝基苯基 (pNP) 的分析测量纤维素酶活性。(Advances in Carbohydrate Chemistry and Biochemistry, Academic Press, 1999)。以 U/ml 为单位测量活性水平,如图 2 所示,以确定各培养物的相对表达水平。

[0225] 结果显示,具有包含 SEQ ID NO:1 的构建物的细胞显示了其 SEQ ID NO:2 活性显

著高于具有 SEQ ID NO:4 的细胞,并且 SEQ ID NOs:1 和 4 产生的表达蛋白质的活性量均大于具有 SEQ ID NO:3 的细胞。

[0001]

序列表

<110> 维莱尼姆公司

B. 张

A. H. 达文波特

L. E. 惠普尔

H. D. 厄比纳

K. E. 巴雷特

M. A. 沃尔

X. 谭

<120> 在极端井况下水解瓜尔胶压裂流体的纤维素酶的编码基因

<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> 人工序列

<220>

<223> 编码SEQ ID NO:2的开放阅读框, 输送增强的蛋白质表达

<400> 1

```
atgggcgteg atcggtttga acgtaacaaa atcttgggcc gcggcattaa tatcggcaat 60
gcgctcgaag caccaaatga aggcgactgg ggagtgggtga taaaagatga gttcttcgac 120
attataaaag aagccggttt ctctcatggt cgaattccaa taagatggag tacgcacgct 180
caggcggttt ctccttataa aatcgagcct tctttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctggt gttataaata ttcactacta cgaggagtta 300
atgaatgata cagaagaaca caaggaaaga tttcttgcgc ttggaaaca aattgctgat 360
cgttataaag actatccga aactctattt ttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc aactgtgat tataggcaca gctgaatggg ggggtatata tgcccttgaa 540
aaactgaggg tcccaaatg ggaaaaaat gcgatagtta caattcacta ctacaatcct 600
ttogaattta cccatcaagg agetgagtgg gtgcctggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaattttac ataggtgagt ttggtgccta cagaaaagct 780
```

[0002]

<210> 2
 <211> 317
 <212> PRT
 <213> 海栖热袍菌

[0003]

<210> 3
 <211> 954
 <212> DNA
 <213> 人工序列

<220>

<223> 编码SEQ ID NO:2的开放阅读框

<400> 3

```
atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtgggtga taaaagatga gttcttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
caggcgtttc ctctttataa aatcgagcct tctttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaata ttcateacta cgaggagtta 300
atgaatgata cagaagaaca caaggaaaga ttcttctgtc ttggaaaca aattgctgat 360
cgttataaag actateccga aactctatit ttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagtat aagatcaatt 480
gacaaaaagc acaactgtgat 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 atggacctec ttgtctgtc gcgaagccga gaaaaggggg 840
tgagctggg cactctggga atttgttcc ggttttggtg tttatgatcc tctgagaaaa 900
cagtggaaata aagatctttt agaagcttta ataggaggag atagcattga ataa 954
```

<210> 4
 <211> 976
 <212> DNA
 <213> 人工序列

<220>

<223> 5' 区域和编码SEQ ID NO:2的开放阅读框, 具有最适度增加的表达

<400> 4

```
tctactagtt aggaggtaac ttatgggcgt cgatccgttt gaacgtaaca aaatcttggg 60
ccgcggcatt aatatcggca atgcgctcga agcaccaaaat gaaggecgact ggggagtggg 120
gataaaagat gatttcttcg acattataaa agaagccggg ttctctcatg ttcgaattcc 180
aataagatgg agtaacgcag ctcaggcgtt tcctccttat aaaatcgagc cttctttctt 240
caaaagagtg gatgaagtga taaacggagc cctgaaaaga ggactggctg ttgttataaa 300
tattcatcac tacgaggagt taatgaatga tccagaagaa cacaaggaaa gatttcttgc 360
tcttttgaaa caaattgctg atcgttataa agactatccc gaaactctat tttttgaaat 420
tctgaatgaa cctcacggaa atcttactcc ggaaaaatgg aatgaactgc ttgaggaagc 480
tctaaaagtt ataagatcaa ttgacaaaaa gcacactgtg attataggca cagctgaatg 540
ggggggtata totgcccttg aaaaactgag ggtcccaaaa tgggaaaaaa atgcgatagt 600
tacaattcac tactacaatc etttcgaatt tacccatcaa ggagctgagt ggggtgcctg 660
atctgagaaa tggttgggaa gaaagtgggg atctccagat gatcagaaac atttgataga 720
agaattcaat tttatagaag aatgggtcaaa aaagaacaaa agaccaattt acataggtga 780
```

[0004]

```

gtttggtgcc tacagaaaag ctgacctga atcaagaata aaatggacct cctttgtcgt 840
tcgcgaagcc gagaaaaggg ggtggagctg ggcatactgg gaattttgtt cgggttttgg 900
tgtttatgat cctctgagaa aacagtggaa taaagatctt ttagaagctt taataggagg 960
agatagcatt gaatga 976

```

<210> 5

<211> 954

<212> DNA

<213> 海栖热袍菌

<400> 5

```

atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtgggtga taaaagatga gttcttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
tacgcgttcc ctctttataa aatcatggat cgttcttcca aaagagtggg tgaagtata 240
aacggagccc tgaaaagagg actggctgtt gttataaata ttcatcacta cgaggagtta 300
atgaatgate cagaagaaca caaggaaaga tttcttgcct tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctatit ttigaaatcc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagetc taaaagtatt aagatcaatt 480
gacaaaaagc acactataat tataggcaca gctgaatggg ggggtatata tgccttgaa 540
aaactgtctg tcccaaatg ggaaaaaat tctatagtta caattcacta ctacaatcct 600
ttcgaattta cccatcaagg agctgagtgg gtggaaggat ctgagaaatg gttgggaaga 660
aagtggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac atagggtgagt ttggtgecta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaaatgga gaaaaggaga 840
tgtagctggg catactggga atttgttccc gggtttgggtg tttatgatac tetgagaaaa 900
acctggaata aagatctttt agaagcttta ataggaggag atagcattga ataa 954

```

<210> 6

<211> 317

<212> PRT

<213> 海栖热袍菌

<220>

<221> 结构域

<222> (19)... (296)

<223> 纤维素酶 (糖基水解酶家族 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

```

[0005]

	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> 人工序列

<220>

<223> 合成产生的多核苷酸

<400> 7

```

atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggcgactgg ggagtcgtga taaaagatga gttcttcgac 120
attataaaag aagecggttt ctctcatggt cgaattccaa taagatggag tacgcacgct 180
tacgcgtttc ctccttataa aatcatggat cgtctcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctggt gttataaata ttcactacta cgaggagtta 300
atgaatgata cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccca aactctatit tttgaaatc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagtatt aagatcaatt 480
gacaaaaagc acactataat tataggcaca gctgaatggg ggggtatata tgccttgtaa 540
aaactgtctg tcccaaatg ggaaaaaat tctatagtta caattcacta ctacaatcct 600

```

[0006]

```

ttcgaatttta cccatcaagg agctgagtgg gtggaaggat ctgagaaatg gttgggaaga 660
aagtgggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac atagggtgagt 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> 人工序列

<220>

<223> 合成产生的多核苷酸

<400> 8

```

atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtgggtga taaaagatga gtatttcgac 120
attataaaag aagccggttt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
caggcgtttc ctctttataa aatcgaggat cgcttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggtgtgtt gttataaate agcatcacta cgaggagtta 300
atgaatgata cagaagaaca caaggaaaga tttcttgctc tttggaaaca aattgctgat 360
cgttataaag actatcccga aactctatit tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagctc taaaagtatt aagatcaatt 480
gacaaaaagc aactataaat tataggcaca gctgaatggg ggggtatate tgcccttgaa 540
aaactgaggg tcccaaaatg ggaaaaaaat gcgatagtta caattcacta ctacaatect 600
ttcgaatttta cccatcaagg agctgagtgg gtggaaggat ctgagaaatg gttgggaaga 660
aagtgggggat ctccagatga tcagaaacat ttgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac atagggtgagt 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> 人工序列

<220>

<223> 合成产生的多肽

<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

```

[0007]

50	55	60
Pro Tyr Lys Ile Glu Asp Arg Phe Phe Lys Arg Val Asp Glu Val Ile		
65	70	75
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
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
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> 人工序列

<220>

<223> 合成产生的多核苷酸

<400> 10

```

atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtgggtga taaaagatga gtatttcgac 120
attataaaag aagccgggtt ctctcatgtt cgaattccaa taagatggag tacgcacgct 180
caggcgtttc ctctttataa aatcgaggat tctttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaata ttcatcacta cgaggagtta 300
atgaatgate cagaagaaca caaggaaaga tttcttgcctc tttggaaaca aattgctgat 360

```

[0008]

```

cgttataaag actatcccga aactctattt tttgaaattc tgaatgaacc tcacggaaat 420
cttactccgg aaaaatggaa tgaactgctt gaggaagetc taaaagttat aagatcaatt 480
gacaaaaaagc acactgtgat tataggcaca gctgaatggg ggggtatate tgcccttgaa 540
aaactgagggg tcccaaaatg ggaaaaaaat gcgatagtta caattcacta ctacaatcct 600
ttcgaatttta cccatcaagg agctgagtgg gtgccctggat ctgagaaatg gttgggaaga 660
aagtgggggat ctccagatga tcagaaacat gtgatagaag aattcaattt tatagaagaa 720
tggtcaaaaa agaacaaaag accaatttac ataggtagt ttggtgccta cagaaaagct 780
gaccttgaat caagaataaa atggacctcc tttgtcgttc gcgaagccga gaaaaggggg 840
tgtagctggg catactggga attttgttcc ggttttgggtg tttatgatcc tctagaaaaa 900
cagtgggaata aagatctttt agaagctcta ataggaggag atagcattga ataa      954

```

<210> 11

<211> 317

<212> PRT

<213> 人工序列

<220>

<223> 合成产生的多肽

<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

```

[0009]

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> 人工序列

<220>

<223> 合成产生的多核苷酸

<400> 12

```

atgggtgttg atccttttga aaggaacaaa atattgggaa gaggcattaa tataggaaat 60
gcgcttgaag caccaaatga gggagactgg ggagtggatg taaaagatga gttcttcgac 120
attataaaag aagecggttt ctctcatgtt cgaattccaa taagatggag tacgcacget 180
caggcggttc ctctttataa aatcgaggat tctttcttca aaagagtgga tgaagtgata 240
aacggagccc tgaaaagagg actggctgtt gttataaate agcatcacta cgaggagtta 300
atgaatgate cagaagaaca caaggaaaga tttcttgetc ttgggaaaca aattgetgat 360
cgttataaag actatcccga aactctatit tttgaaatc tgaatgaace tcacggaaat 420
cttactecgg aaaaatggaa tgaactgett gaggaagctc taaaagttat aagatcaatt 480
gacaaaaagc aactgtgat tataggcaca gctgaatggg ggggtatata tgcccttgaa 540
aaactgaggg tcccaaatg ggaaaaaat 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
tgagagctgg catactggga attttgttcc ggttttgggt tttatgatcc tctgagaaaa 900
cagtggaaata aagatctttt agaagcttta ataggaggag atagcattga ataa 954

```

<210> 13

<211> 317

<212> PRT

<213> 人工序列

<220>

<223> 合成产生的多肽

<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

[0010]

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> 热袍菌属某种

<400> 14

```

atggaacagt cagttgctga aagtgatagc aactcagcat ttgaatacaa caaaatggta 60
ggtaaaggag taaatatitgg aaatgcttta gaagtcctt tcgaaggagc ttggggagta 120
agaattgagg atgaatatit tgagataata aagaaaaggg gatttgattc tgtaggatt 180
cccataagat ggtcagcaca tatatccgaa aagccaccat atgatattga caggaatttc 240
ctcgaaagag ttaaccatgt tgtagatagg gctcttgaga ataatttaac agtaatcatc 300

```

[0011]


```

aatacgcacc attttgaaga actctatcaa gaaccggata aatacggcga tgttttggtg 360
gaaatttgga gacagattgc aaaattcttt aaagattacc cggaaaatct gttctttgaa 420
atctacaacg agcctgctca gaacttgaca gctgaaaaat ggaacgcact ttatccaaaa 480
gtgctcaaaag ttatcaggga gagcaatcca acccggattg tcattatcga tgetccaaac 540
tgggcacact atagegcagt gagaagtcta aaattagtca acgacaaacg catcattgtt 600
tccttccatt actacgaacc tttaaattc acacatcagg gtgcgaatg ggtaatecc 660
atcccacctg ttaggggtta gtggaatggc gaggaatggg aaattaacca aatcagaagt 720
catttcaa atcgtgagtga ctgggcaaag caaaataacg tacciaatctt tcttggtgaa 780
ttcgggtgctt attcaaaagc agacatggac tcaagggtta agtggaccga aagtgtgaga 840
aaaatggcgg aagaatttgg attttcatac gcgtattggg aattttgtgc aggatttggc 900
atatacgata gatgggtctca aaactggatc gaaccattgg caacagctgt ggttggcaca 960
ggcaaagagt aa 972

```

<210> 15

<211> 323

<212> PRT

<213> 热袍菌属某种

<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

```

[0012]

	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> 人工序列

<220>

<223> 合成产生的多核苷酸

<400> 16

```

ggatccacca tgagggtgtt gctcgttgcc ctcgctctcc tggtctctgc tgcgagcgcc 60
accagcggcg tggaccctgt cgagaggaac aagatcctgg gcaggggcat caacatcggc 120
aacgccttgg agggcccgaa cgaggcgac tggggcgtgg tgatcaagga cgagtacttc 180
gacatcatea aggaggccgg cttcagccac gtgagaatcc cgatcagggt gagcaccac 240
gcccaggcct tcccgcctga caagategag gacaggttct tcaagagggt ggacgagggt 300
atcaacggcg ccctgaagag gggcctggcc gtggtgatca accagcacca ctacgaggag 360
ctgatgaacg acccggagga gcacaaggag aggttctctg ccctgtggaa gcagatcgcc 420
gacaggtaca aggaactacc ggagacctg ttcttcgaga tctgaacga gccgcacggc 480
aacctgaccc cggagaagtg gaacgagctg ctggaggagg ccctgaaggt gatcaggagc 540
atcgacaaga agcacaccat catcatcggc accgccgagt ggggcggcat cagcgccctg 600
gagaagctga ggggtgccga gtgggagaag aacgccatcg tgaccatcca ctactacaac 660
cgttcgagt tcaccaccca gggcgccgag tgggtggagg gcagcgagaa gtggctgggc 720
aggaagtggg gcagcccga cgaccagaag cacctgatcg aggagttaa cttcatcgag 780
gagtggagca agaagaacaa gaggcgate tacatcggcg agttcggcgc ctacaggaag 840
gccgacctgg agagcaggat caagtggacc agcttcgtgg tgagggaggc cgagaagagg 900
aggtggagct gggcctactg ggagttctgc agcggcttcg gegtgtacga caccctgagg 960
aagacctgga acaaggacct gctggaggcc ctgatcggcg gcgacagcat cgagagcgag 1020
aaggacgagc tgtgagagct ca 1042

```

<210> 17

<211> 341

<212> PRT

<213> 人工序列

<220>

<223> 合成产生的多肽

<400> 17

[0013]

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> 人工序列

[0014]

<220>

<223> 合成产生的多核苷酸

<400> 18

```

ggatccacca tgagggtgtt getegtgtt gctctctctc ttggtctctc tgcgagcgcc 60
accagcggcg tggaccggtt cgagaggaac aagatcctgg gcaggggcat caacatcggc 120
aacgccctgg aggccccgaa cgaggcgac tggggcgtgg tgateaagga cgagttcttc 180
gacatcatca aggaggccgg cttcagccac gtgagaatcc cgatcagggt gagcaccac 240
gcccaggcct tcccgcgta caagatcgag ccgagcttct tcaagagggt ggacgagggt 300
atcaacggcg ccctgaagag gggcctggcc gtggtgatca acatccacca ctacgaggag 360
ctgatgaacg acccggagga gcacaaggag aggttctctg ccctgtggaa gcagatcgcc 420
gacaggtaca aggactaccc ggagaccctg ttcttcgaga tctgaacga gccgcacggc 480
aacctgaccc cggagaagtg gaacgagctg ctggaggagg ccctgaaggt gatcaggagc 540
atcgacaaga agcacaccgt gatcatcggc accgcagat gggcgggcat cagcgccctg 600
gagaagctga ggggtgcgaa gtgggagaag aacgccatg tgaccatca ctactacaac 660
ccgttcgagt taccacacca gggcgccgag tgggtgcgag gcagcgagaa gtggtctggc 720
aggaaagtgg gcagcccgga cgaccagaag cactgatcg aggagttcaa cttcatcgag 780
gagtggagca agaagaacaa gaggccgate tacatcggcg agttcggcgc ctacaggaag 840
gccgacctgg agagcaggat caagtggacc agcttcgtgg tgaggaggc cgagaagagg 900
ggctggagct gggcctactg ggagttctgc agcggtctg gcgtgtacga cccgtgagg 960
aagcagtggg acaaggacct gctggaggcc ctgatcggcg gcgacagcat cgagagcgag 1020
aaggacgagc tgtgagagct ca 1042

```

<210> 19

<211> 341

<212> PRT

<213> 人工序列

<220>

<223> 合成产生的多肽

<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

```

[0015]

130	135	140
Glu Thr Leu Phe Phe Glu Ile Leu Asn Glu Pro His Gly Asn Leu Thr		
145	150	155
Pro Glu Lys Trp Asn Glu Leu Leu Glu Glu Ala Leu Lys Val Ile Arg		160
	165	170
Ser Ile Asp Lys Lys His Thr Val Ile Ile Gly Thr Ala Glu Trp Gly		175
	180	185
Gly Ile Ser Ala Leu Glu Lys Leu Arg Val Pro Lys Trp Glu Lys Asn		190
	195	200
Ala Ile Val Thr Ile His Tyr Tyr Asn Pro Phe Glu Phe Thr His Gln		205
	210	215
Gly Ala Glu Trp Val Pro Gly Ser Glu Lys Trp Leu Gly Arg Lys Trp		220
225	230	235
Gly Ser Pro Asp Asp Gln Lys His Leu Ile Glu Glu Phe Asn Phe Ile		240
	245	250
Glu Glu Trp Ser Lys Lys Asn Lys Arg Pro Ile Tyr Ile Gly Glu Phe		255
	260	265
Gly Ala Tyr Arg Lys Ala Asp Leu Glu Ser Arg Ile Lys Trp Thr Ser		270
	275	280
Phe Val Val Arg Glu Ala Glu Lys Arg Gly Trp Ser Trp Ala Tyr Trp		285
	290	295
Glu Phe Cys Ser Gly Phe Gly Val Tyr Asp Pro Leu Arg Lys Gln Trp		300
305	310	315
Asn Lys Asp Leu Leu Glu Ala Leu Ile Gly Gly Asp Ser Ile Glu Ser		320
	325	330
Glu Lys Asp Glu Leu		335
	340	

<210> 20

<211> 1042

<212> DNA

<213> 人工序列

<220>

<223> 合成产生的多核苷酸

<400> 20

```

ggatccacca tgagggtgtt gctcgttgcc ctcgctctcc tggctctcgc tgcgagcgcc 60
accagcggcg tggaccggtt cgagaggaac aagatcctgg gcaggggcat caacatcggc 120
aacgccctgg aggccecgaa cgagggcgac tggggcgtgg tgatcaagga cgagtacttc 180
gacatcatca aggaggccgg cttcagccac gtgagaatcc cgatcaggtg gacacccac 240
gcccaggcct tcccgcgta caagatcgag gacagcttct tcaagagggt ggacgaggtg 300
atcaacggcg ccctgaagag gggcctggcc gtggtgatca acatccacca ctacgaggag 360
ctgatgaacg acccgagga gcacaaggag aggttcctgg ccctgtggaa gcagatcgcc 420
gacaggtaca aggactacce ggagaccctg ttcttcgaga tctgaacga gccgcacggc 480
aacctgaccc cggagaagtg gaacgagctg ctggaggagg ccctgaaggt gatcaggagc 540
atcgacaaga agcacaccgt gateatcggc accgccgagt ggggcggcgt cagcgccctg 600
gagaagctga ggggtgccgaa gtgggagaag aacgccatcg tgaccatcca ctactacaac 660
ccgttcgagt tcaccacca gggcgccgag tgggtgccgg gcagcgagaa gtggctgggc 720

```

[0016]

```

aggaagtggg gcagcccgga cgaccagaag cacgtgatcg aggagttcaa cttcatcgag 780
gagtggagca agaagaacaa gaggccgatc tacatcggcg agttcggcgc ctacaggaag 840
gccgacctgg agagcaggat caagtggacc agcttcgtgg tgagggaggc cgagaagagg 900
ggctggagct gggcctaetg ggagttctgc agcggttcg gcgtgtacga cccgtgagg 960
aagcagtgga acaaggacct gctggaggcc ctgatcggcg gcgacagcat cgagagcgag 1020
aaggacgagc tgtgagagct ca 1042

```

<210> 21

<211> 341

<212> PRT

<213> 人工序列

<220>

<223> 合成产生的多肽

<400> 21

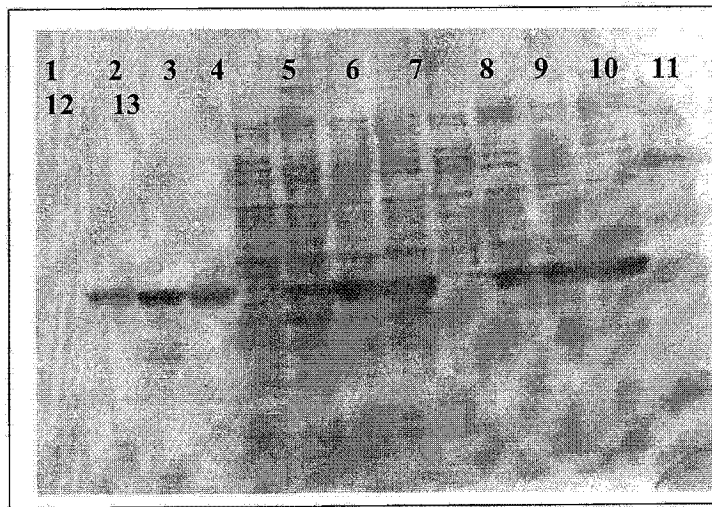
```

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

```

[0017]

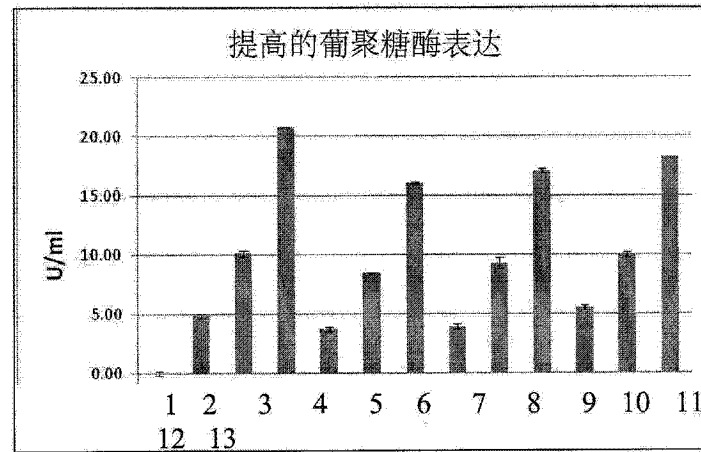
[illegible]



图例

泳道 1、5、9 = 阴性对照
泳道 2、6、10 = SEQ ID NO:3
泳道 3、7、11 = SEQ ID NO:4
泳道 4、8、12 = SEQ ID NO:1
通道 13 = 分子量梯

图 1



图例

泳道 1 = 阴性对照

泳道 2、5、8、11 = SEQ ID NO:3

泳道 3、6、9、12 = SEQ ID NO:4

泳道 4、7、10、13 = SEQ ID NO:1

图 2

SEQ ID NO:1

ATGGGCGTCGATCCGTTTGAACGTAACAAAATCTTGGGCGCGGCATTAATATCGGC
 AATGCGCTCGAAGCACCAAATGAAGGCGACTGGGGAGTGGTGATAAAAGATGAGTTC
 TTCGACATTATAAAAGAAGCCGGTTTCTCTCATGTTTGAATTCCAATAAGATGGAGT
 ACGCACGCTCAGGCGTTTCCCTCCTTATAAAATCGAGCCTTCTTTCTTCAAAGAGTG
 GATGAAGTGATAAACGGAGCCCTGAAAAGAGGACTGGCTGTTGTTATAAATATTCAT
 CACTACGAGGAGTTAATGAATGATCCAGAAGAACACAAGGAAAGATTTCTTGCTCTT
 TGGAAACAAATTGCTGATCGTTATAAAGACTATCCCGAACTCTATTTTTTGAAATT
 CTGAATGAACCTCACGGAAATCTTACTCCGGAAAAATGGAATGAACTGCTTGAGGAA
 GCTCTAAAAGTTATAAGATCAATTGACAAAAAGCACACTGTGATTATAGGCACAGCT
 GAATGGGGGGGTATATCTGCCCTTGAAAACTGAGGGTCCCAAAATGGGAAAAAAT
 GCGATAGTTACAATTCCTACTACTACAATCCTTTTGAATTTACCCATCAAGGAGCTGAG
 TGGGTGCCTGCATCTGAGAAATGGTTGGGAAGAAAGTGGGGATCTCCAGATGATCAG
 AAACATTTGATAGAAGAATTCAATTTTATAGAAGAATGGTCAAAAAAGAACAAAAGA
 CCAATTTACATAGGTGAGTTTGGTGCCTACAGAAAAGCTGACCTTGAATCAAGAATA
 AAATGGACCTCCTTTGTCGTTCCGGAAGCCGAGAAAAGGGGGTGGAGCTGGGCATAC
 TGGGAATTTTGTTCGGTTTGGTGTATGATCCTCTGAGAAAACAGTGAATAAA
 GATCTTTTAGAAGCTTTAATAGGAGGAGATAGCATTGAATGA

图 3

SEQ ID NO:2

MGVDPFERNKILGRGINIGNALEAPNEG DWGVV IKDEF FDI I KEAGFSHVRIPIRWS
THAQAFPPYKIEPSFFKRVDEVINGALKRGLAVVINIHHYEELMNDPEEHKERFLAL
WKQIADRYKDYPETLFFEILNEPHGNLTPEKWNELLEALKVIRSIDKKHTVIIGTA
EWGGISALEKL RVPKWEKNAIVTIHYYNPF EFTHQGAEWVPGSEKWLGRKWGSPDDQ
KHLIEEFNFIEEWSKKNKRPIYIGEFGAYRKADLESRIKWTSFVVREAEKRGWSWAY
WEFCSGFGVYDPLRKQWNKDLLEALIGGDSIE

图 4

SEQ ID NO:3

ATGGGTGTTGATCCTTTTGAAAGGAACAAAATATTGGGAAGAGGCATTAATATAGGAAATGCGC
TTGAAGCACCAAATGAGGGAGACTGGGGAGTGGTGATAAAAGATGAGTTCTTCGACATTATAAA
AGAAGCCGGTTTCTCTCATGTTCCAATTCCAATAAGATGGAGTACGCACGCTCAGGCGTTTCCT
CCTTATAAAATCGAGCCTTCTTTCTTCAAAAGAGTGGATGAAGTGATAAACGGAGCCCTGAAAA
GAGGACTGGCTGTTGTTATAAATATTCATCACTACGAGGAGTTAATGAATGATCCAGAAGAACA
CAAGGAAAGATTTCTTGCTCTTTGGAAACAAATTGCTGATCGTTATAAAGACTATCCCGAAACT
CTATTTTTTGAATTTCTGAATGAACCTCACGGAAATCTTACTCCGGAAAAATGGAATGAACTGC
TTGAGGAAGCTCTAAAAGTTATAAGATCAATTGACAAAAAGCACACTGTGATTATAGGCACAGC
TGAATGGGGGGGTATATCTGCCCTTGAAAACTGAGGGTCCCAAAATGGGAAAAAAATGCGATA
GTTACAATTCACTACTACAATCCTTTTGAATTTACCCATCAAGGAGCTGAGTGGGTGCCTGGAT
CTGAGAAATGTTGGGAAGAAAGTGGGGATCTCCAGATGATCAGAAACATTTGATAGAAGAATT
CAATTTTATAGAAGAATGGTCAAAAAAGAACAAAAGACCAATTTACATAGGTGAGTTTGGTGCC
TACAGAAAAGCTGACCTTGAATCAAGAATAAAATGGACCTCCTTTGTCGTTTCGCGAAGCCGAGA
AAAGGGGGTGGAGCTGGGCATACTGGGAATTTTGTTCGGTTTTTGGTGTTTATGATCCTCTGAG
AAAACAGTGGAATAAAGATCTTTTAGAAGCTTTAATAGGAGGAGATAGCATTGAATAA

图 5

SEQ ID NO:4

TCTACTAGTTAGGAGGTAACCTTATGGGCGTCGATCCGTTTGAACGTAACAAAATCTTGGGCCGC
GGCATTAAATATCGGCAATGCGCTCGAAGCACCAAATGAAGCGACTGGGGAGTGGTGATAAAAG
ATGAGTTCTTCGACATTATAAAAGAAGCCGGTTTCTCTCATGTTTCAATCCAATAAGATGGAG
TACGCACGCTCAGGCGTTTCTCCTTATAAAATCGAGCCTTCTTTCTTCAAAAGAGTGGATGAA
GTGATAAACGGAGCCCTGAAAAGAGGACTGGCTGTTGTTATAAATATTCATCACTACGAGGAGT
TAATGAATGATCCAGAAGAACAAGGAAAGATTTCTTGCTCTTTGGAAACAAATTGCTGATCG
TTATAAGACTATCCCGAAACTCTATTTTTTTGAAATTCTGAATGAACCTCACGGAAATCTTACT
CCGGAAAAATGGAATGAACTGCTTGAGGAAGCTCTAAAAGTTATAAGATCAATTGACAAAAAGC
ACACTGTGATTATAGGCACAGCTGAATGGGGGGGTATATCTGCCCTTGAAAACTGAGGGTCCC
AAAATGGGAAAAAAATGCGATAGTTACAATTCACTACTACAATCCTTTTCAATTTACCCATCAA
GGAGCTGAGTGGGTGCCTGGATCTGAGAAATGGTTGGGAAGAAAGTGGGGATCTCCAGATGATC
AGAAACATTTGATAGAAGAATTCAATTTTATAGAAGAATGGTCAAAAAAGAACAAAAGACCAAT
TTACATAGGTGAGTTTGGTGCCTACAGAAAAGCTGACCTTGAATCAAGAATAAAATGGACCTCC
TTTGTCTGTTTCGCGAAGCCGAGAAAAGGGGGTGGAGCTGGGCATACTGGGAATTTTGTTCGGGT
TTGGTGTTTATGATCCTCTGAGAAAACAGTGGAATAAAGATCTTTTAGAAGCTTTAATAGGAGG
AGATAGCATTGAATGA

图 6

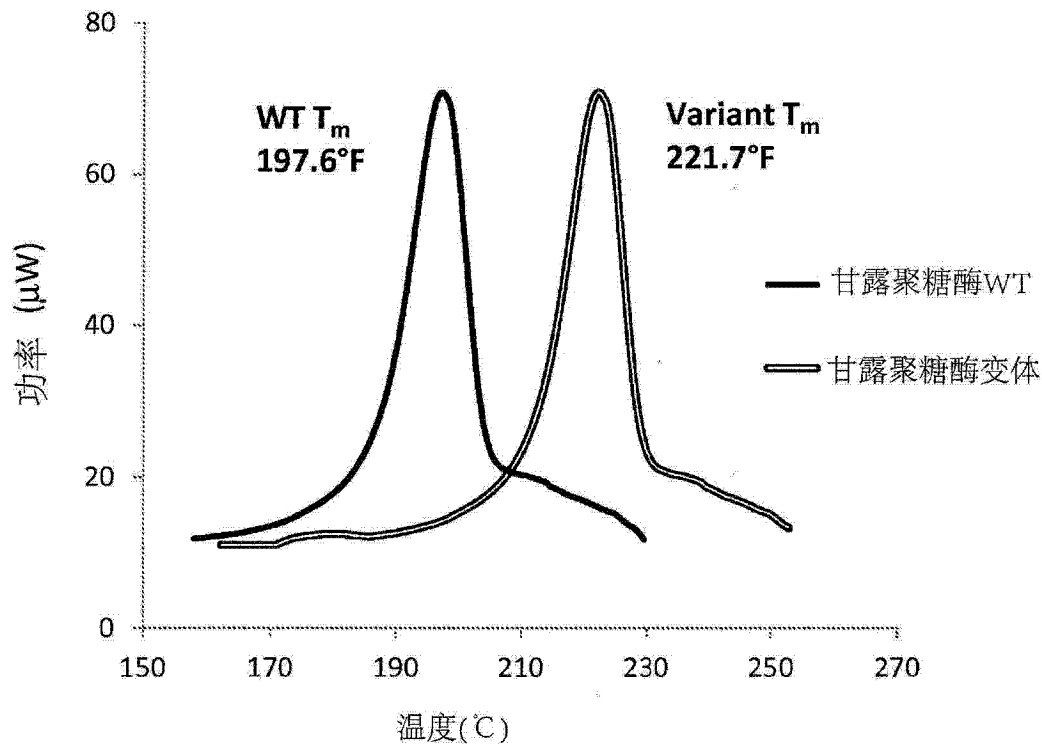


图 7

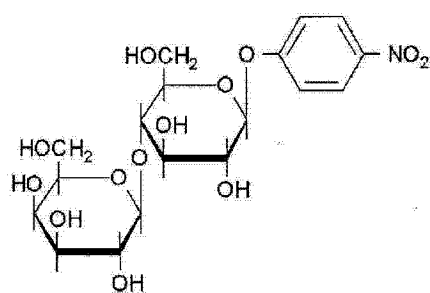


图 8A

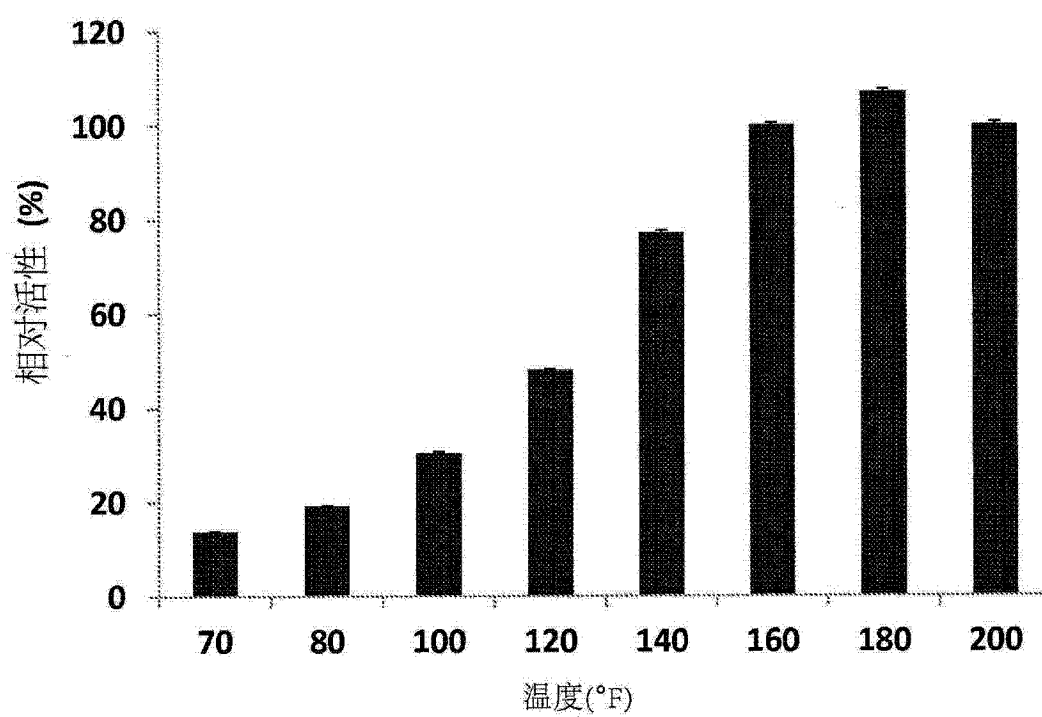


图 8B

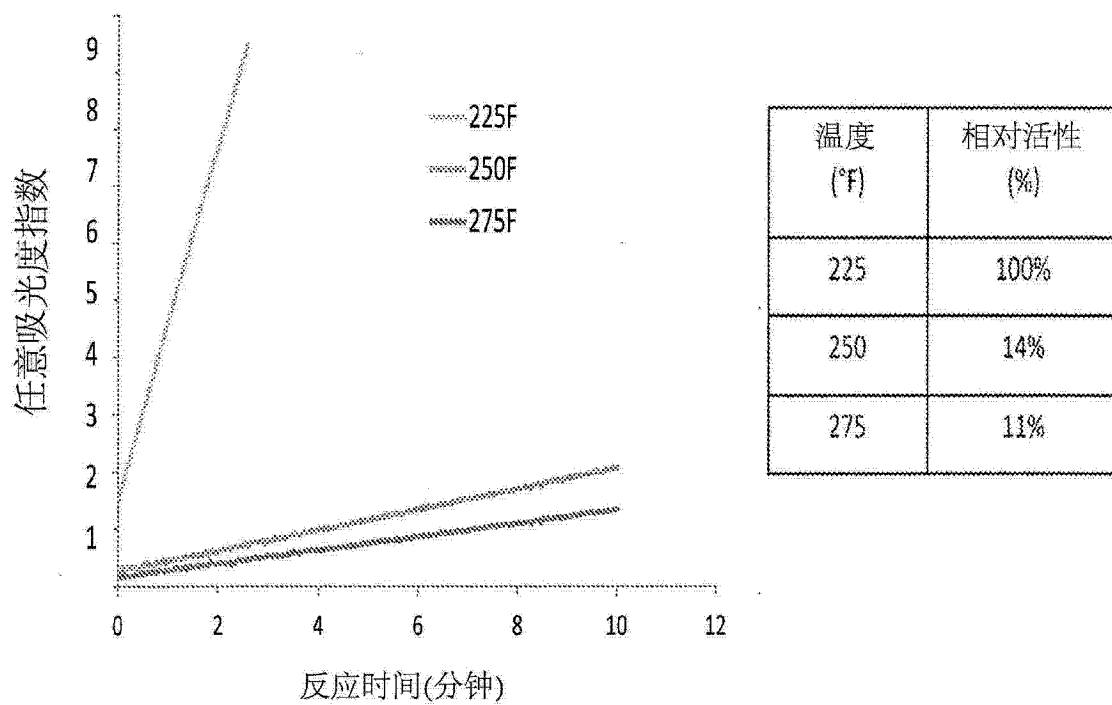


图 8C

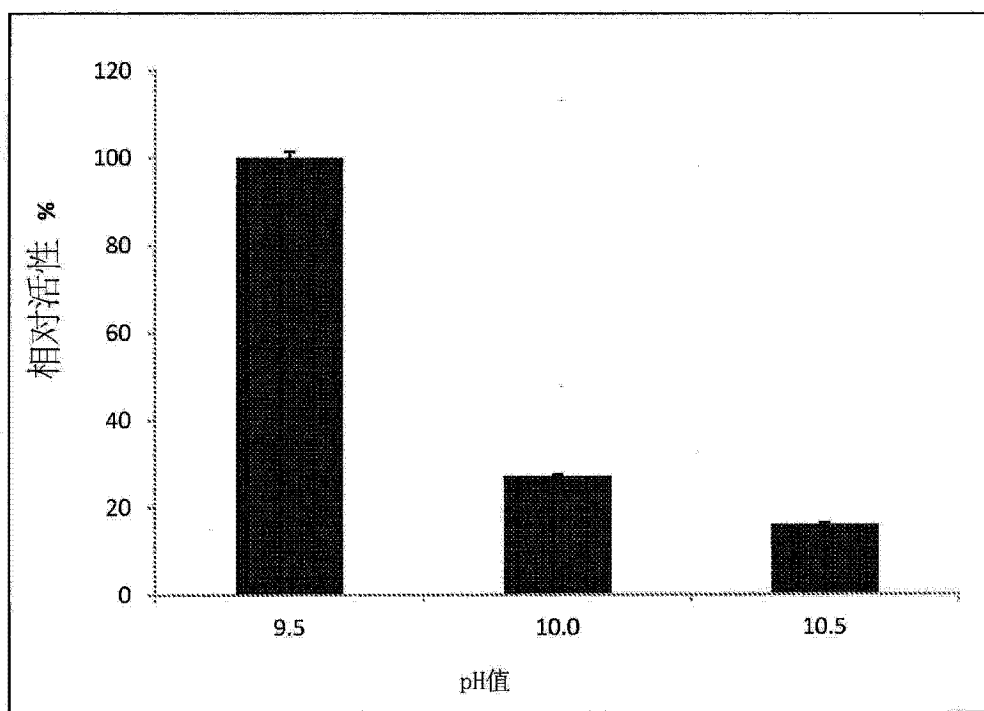


图 9

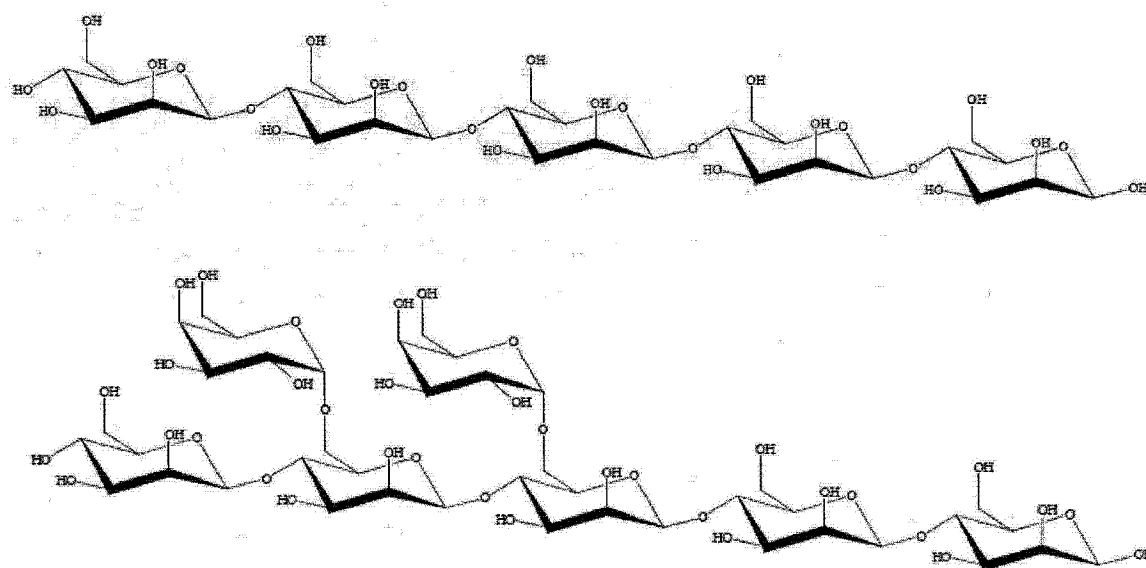


图 10A

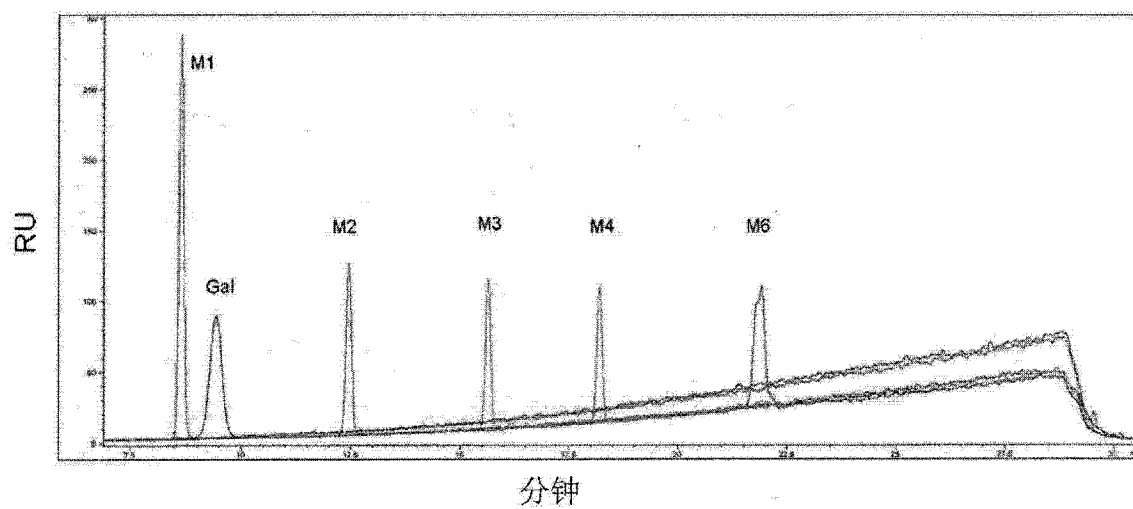


图 10B

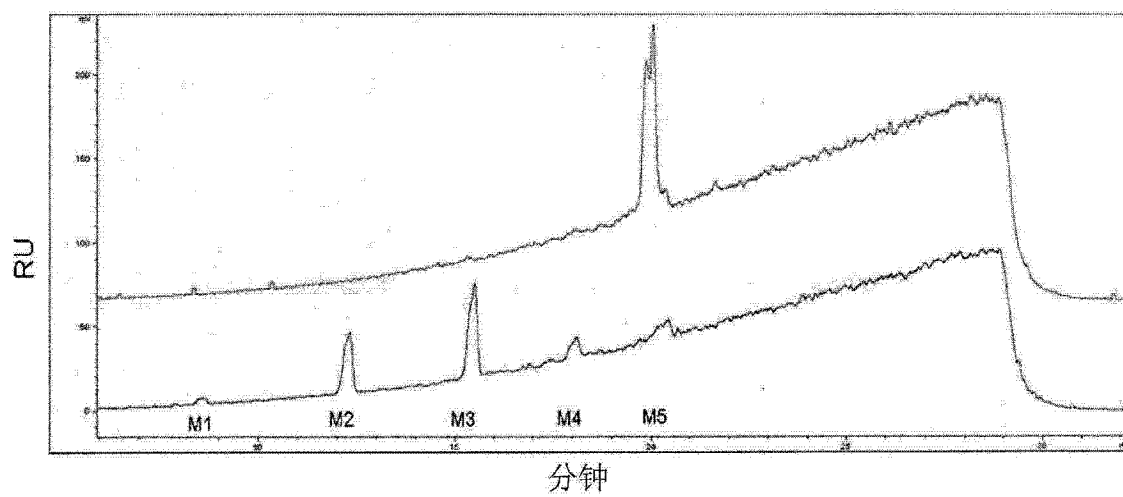


图 10C(a)

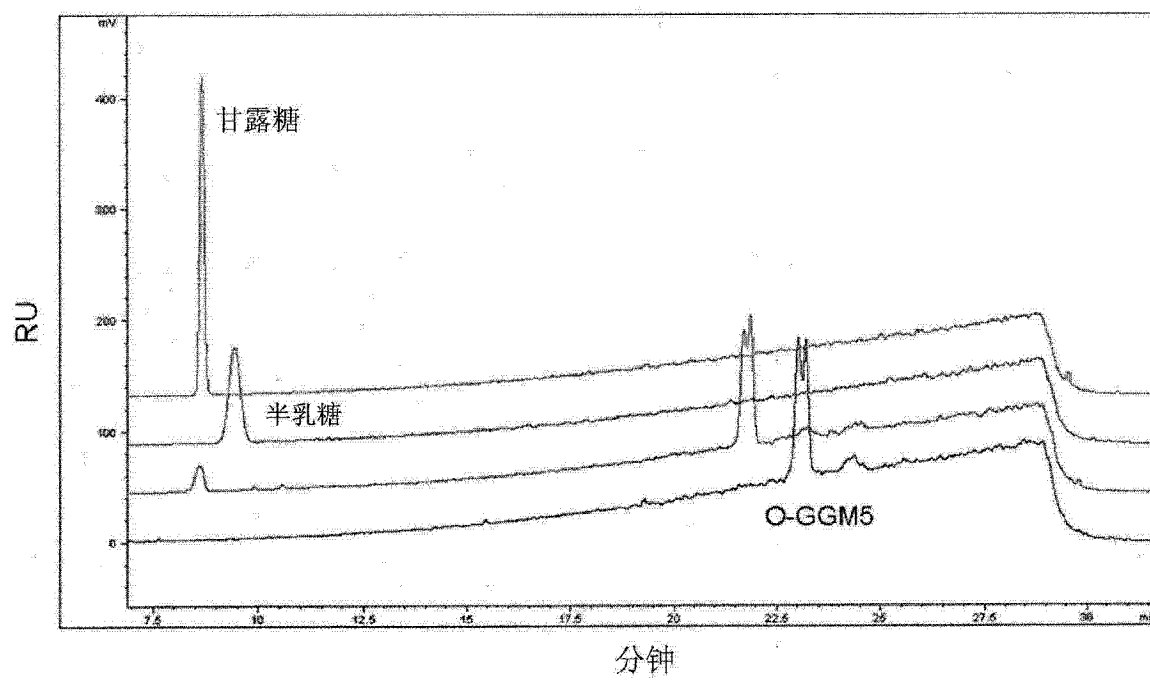


图 10C(b)

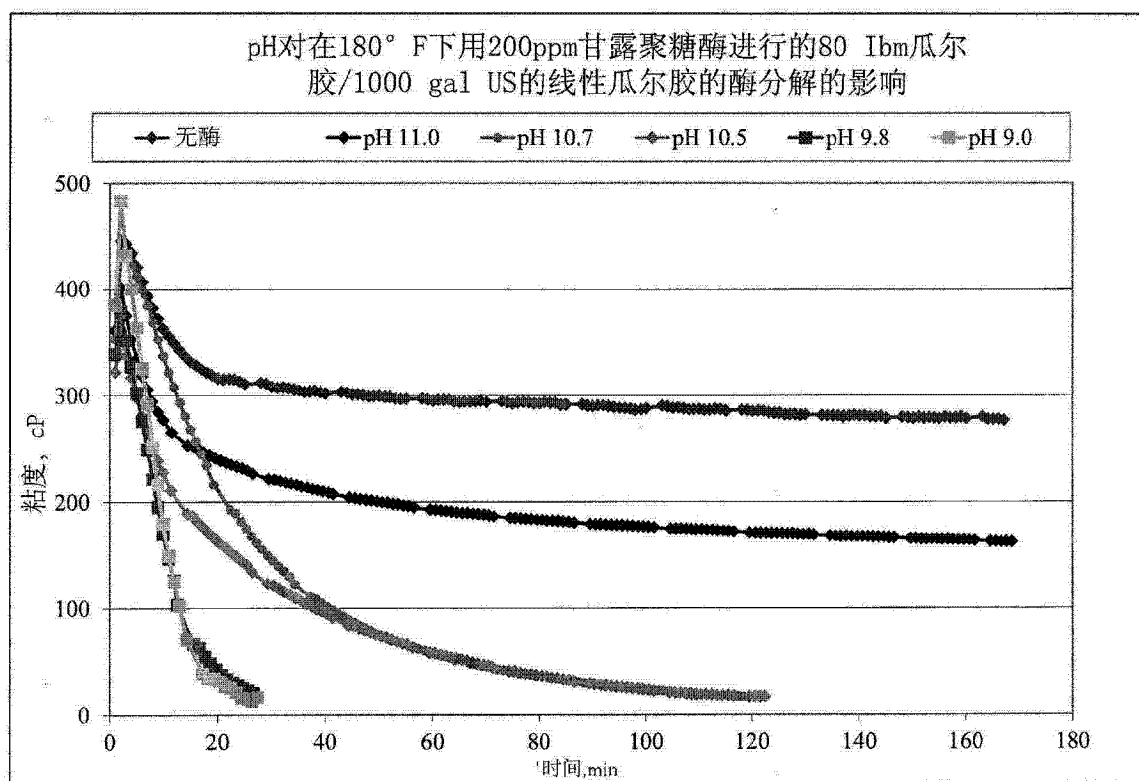


图 11

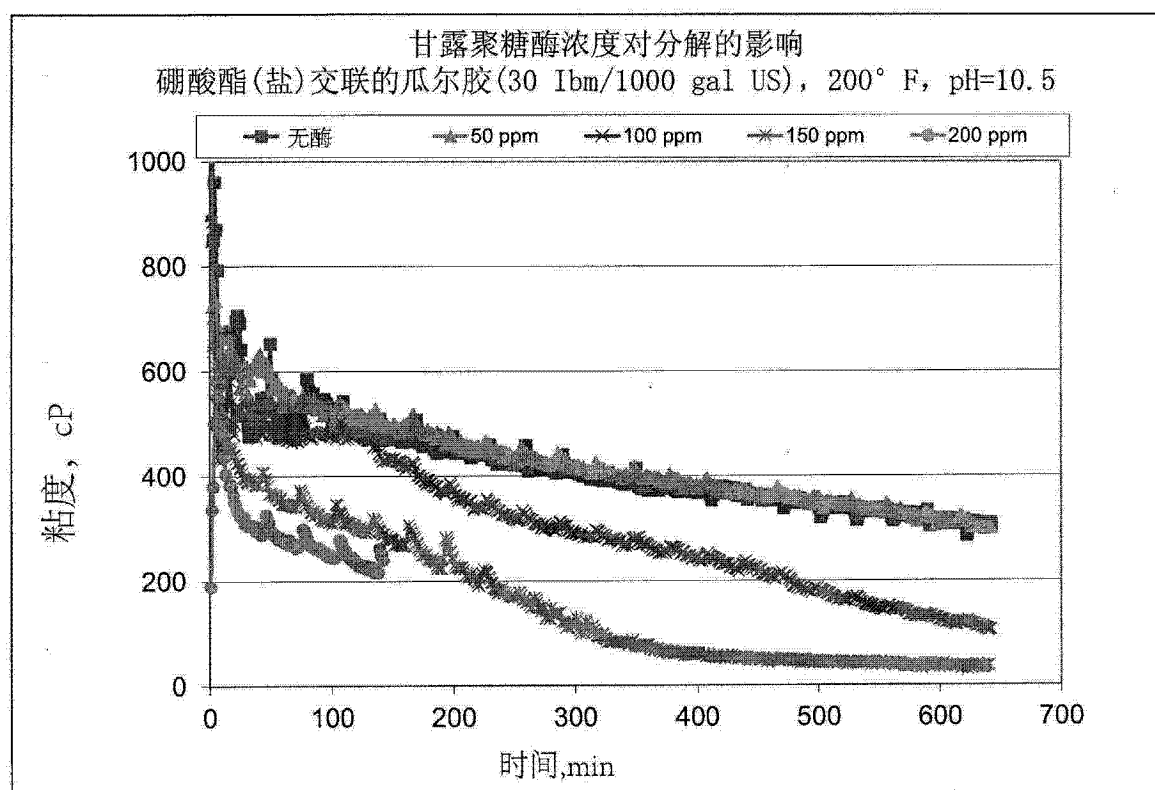


图 12

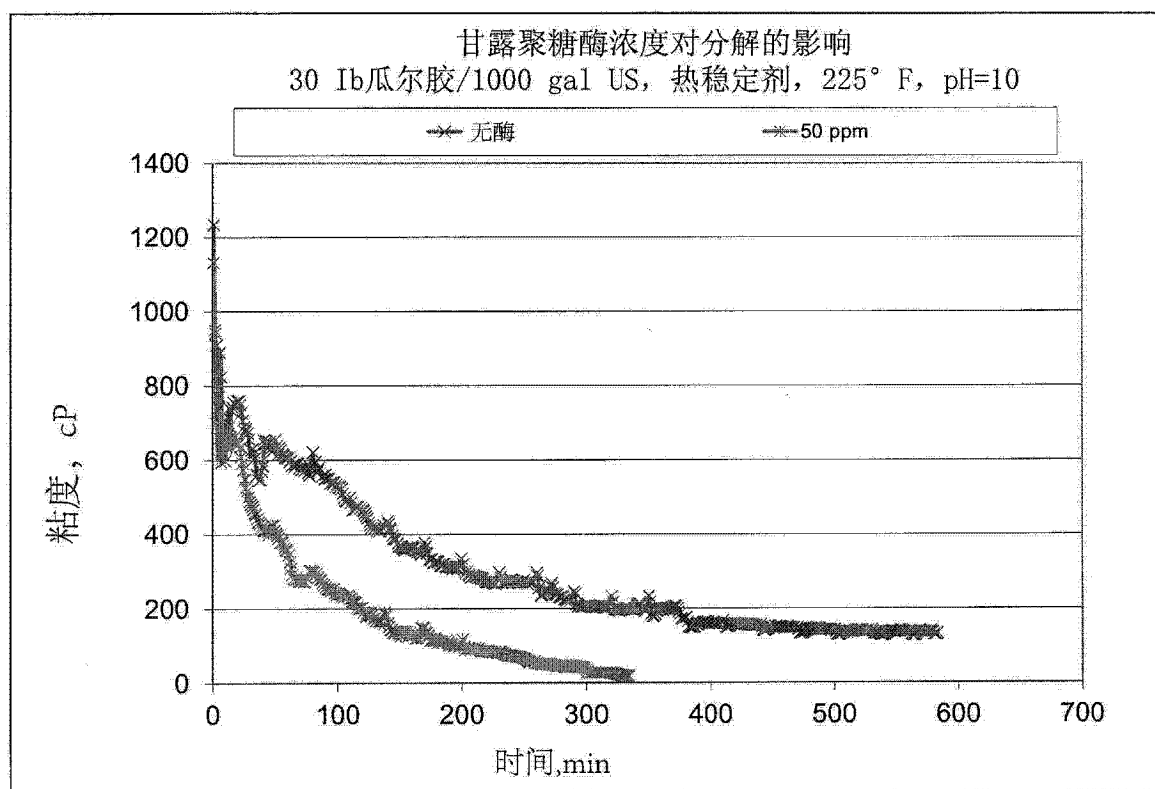


图 13

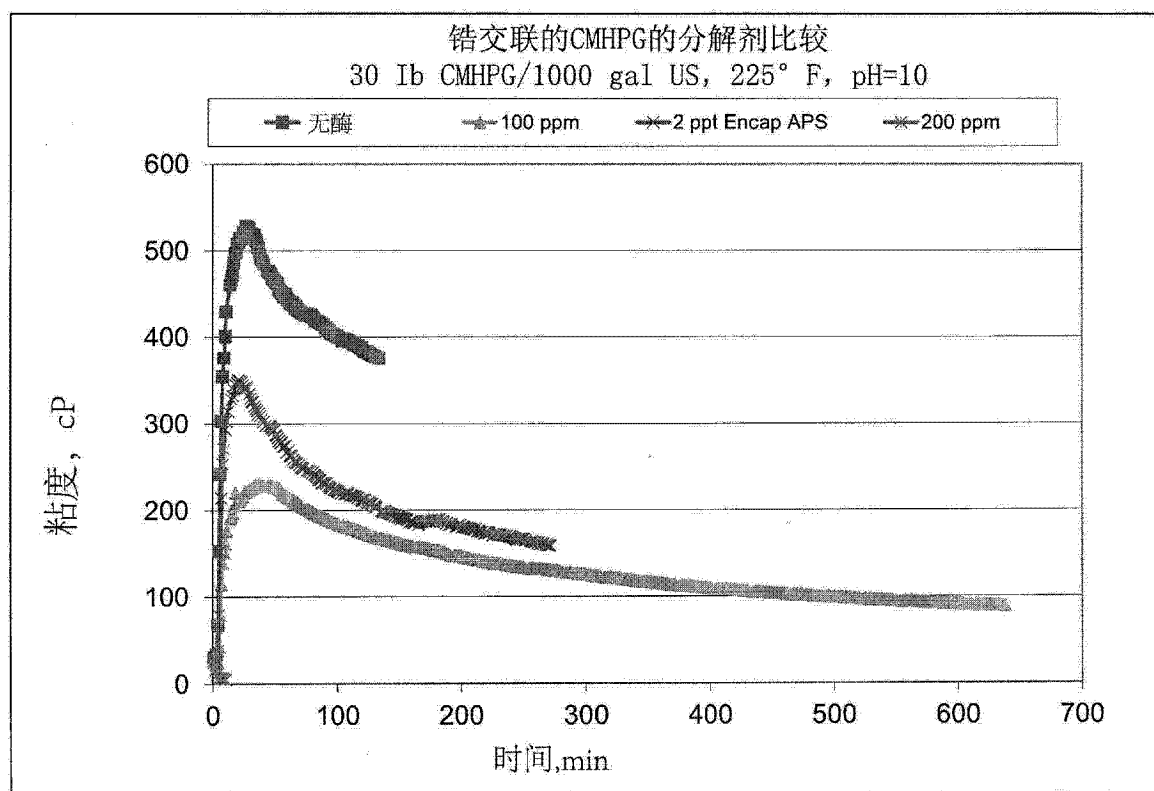


图 14

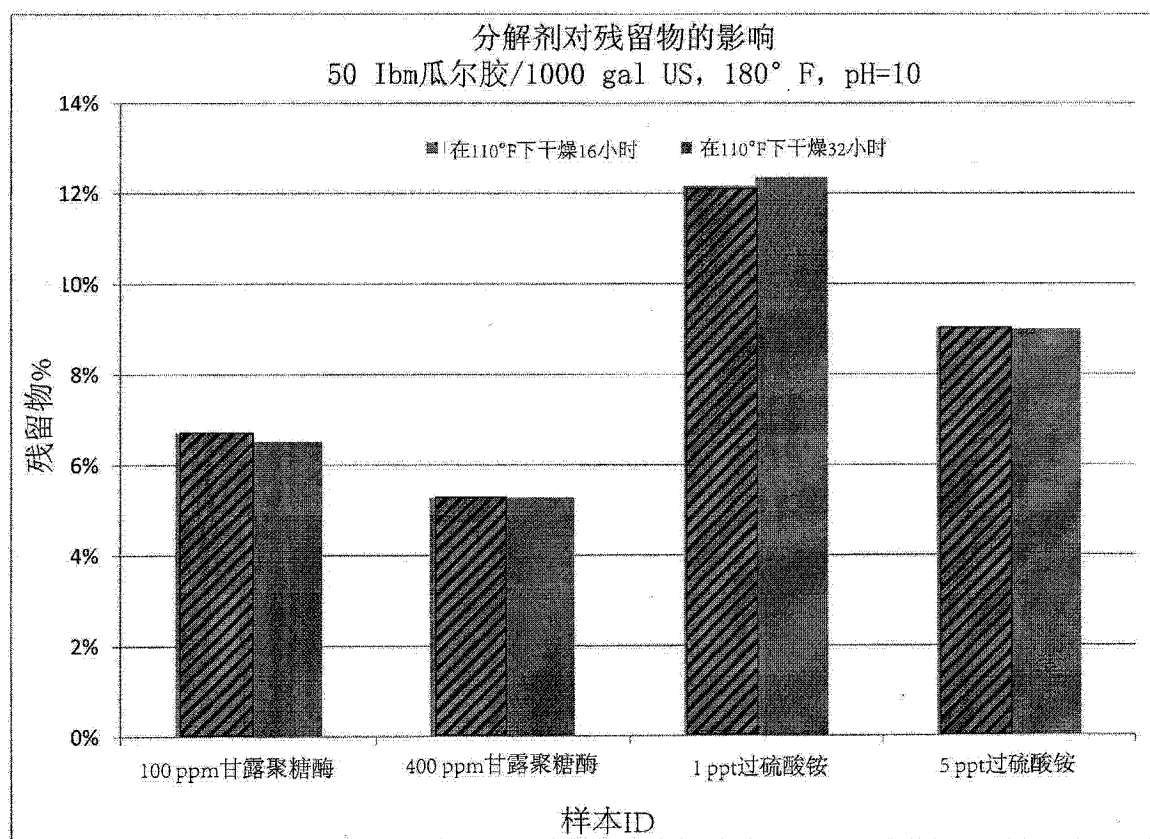


图 15

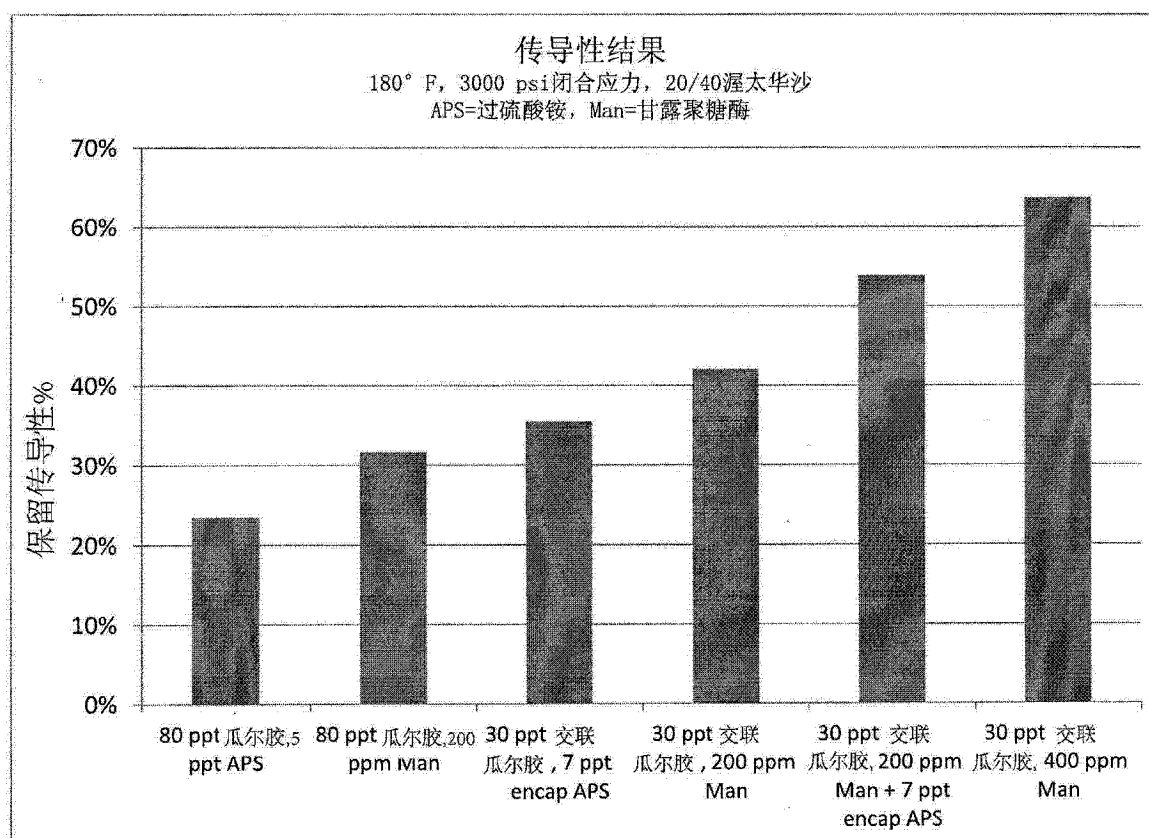


图 16