



(12)发明专利申请

(10)申请公布号 CN 106170546 A

(43)申请公布日 2016. 11. 30

(21)申请号 201580015020.9

(22)申请日 2015.03.20

(30)优先权数据

61/968,853 2014.03.21 US

(85)PCT国际申请进入国家阶段日

2016.09.20

(86)PCT国际申请的申请数据

PCT/US2015/021813 2015.03.20

(87)PCT国际申请的公布数据

W02015/143360 EN 2015.09.24

(71)申请人 丹尼斯科美国公司

地址 美国加利福尼亚州

(72)发明人 M·科尔克曼 A·赫明森

R·梅耶达尔 R·博特 L·巴贝

(74)专利代理机构 北京市中咨律师事务所

11247

代理人 张莉 黄革生

(51)Int.Cl.

C12N 9/54(2006.01)

C12N 15/57(2006.01)

C11D 3/386(2006.01)

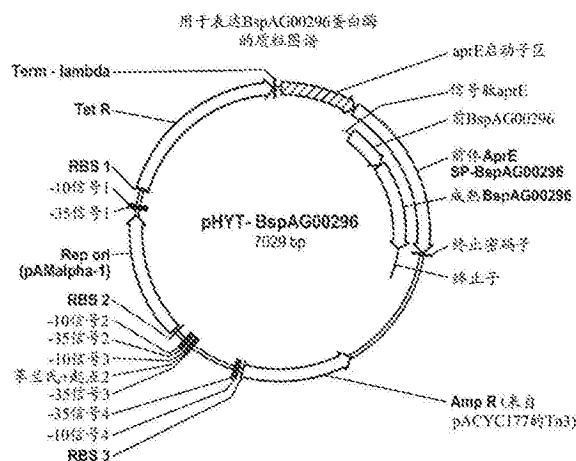
权利要求书5页 说明书77页 附图19页

(54)发明名称

芽孢杆菌属的丝氨酸蛋白酶

(57)摘要

本公开涉及从芽孢杆菌属物种(*Bacillus* spp.)克隆的丝氨酸蛋白酶及其变体。包含丝氨酸蛋白酶的组合物适用于清洁织物和硬质表面,以及多种工业应用。



1. 一种在WHY-进化枝中的重组多肽或其活性片段。
2. 一种重组多肽或其活性片段,所述重组多肽或其活性片段包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少70%的同一性。
3. 根据权利要求1或2所述的重组多肽或其活性片段,其中所述重组多肽或其活性片段具有蛋白水解活性。
4. 根据上述权利要求中任一项所述的重组多肽或其活性片段,其中所述多肽或其活性片段包含DTGIDXXHXXLXNLVXTSLGXSVGGXXXDVXGH基序,其中起始D为活性位点天冬氨酸,末端H为活性位点组氨酸,并且X为任意氨基酸。
5. 根据上述权利要求中任一项所述的重组多肽或其活性片段,前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP010283106、或WP006679321的氨基酸序列。
6. 根据权利要求1-或5中任一项所述的重组多肽或其活性片段,其中所述多肽或其活性片段包含DTGIDXXHXXLX_aNLVXTSLGXSVGGX_bXX_cDVXGH基序,其中起始D为活性位点天冬氨酸,末端H为活性位点组氨酸,并且X、X_a、X_b和X_c为任意氨基酸,前提条件是当X_a为精氨酸时,X_b和X_c不为甘氨酸。
7. 根据权利要求4-6中任一项所述的重组多肽或其活性片段,其中所述基序的VXG序列为VQG。
8. 根据权利要求7所述的重组多肽或其活性片段,其中所述VQG序列在残基位置63-65处,其中所述多肽或其活性片段的氨基酸位置通过与以SEQ ID NO:7示出的氨基酸序列对应来编号。
9. 根据上述权利要求中任一项所述的重组多肽或其活性片段,其中所述多肽或其活性片段包含在残基位置80-82处的VSG序列,其中所述多肽或其活性片段的氨基酸位置通过与以SEQ ID NO:7示出的氨基酸序列对应来编号。
10. 根据上述权利要求中任一项所述的重组多肽或其活性片段,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的插入,其中所述插入介于残基位置39-47之间,其中所述残基位置通过与以SEQ ID NO:18示出的氨基酸序列对应来编号。
11. 根据权利要求10所述的重组多肽或其活性片段,其中所述残基位置39-47用HQSLANLVNTSLG替换。
12. 根据上述权利要求中任一项所述的重组多肽或其活性片段,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的缺失,其中所述缺失介于残基位置51-64之间,其中所述残基位置通过与以SEQ ID NO:18示出的氨基酸序列对应来编号。
13. 根据权利要求12所述的重组多肽或其活性片段,其中所述残基位置51-64用VGGSTMDVQGH、VGGSA/PEDVQGH、VGGNPEDRQ GH、或VGGTPADVHGH替换。
14. 根据上述权利要求中任一项所述的重组多肽或其活性片段,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的缺失,其中所述缺失介于残基位置68-95之间,其中所述残基位置通过与以SEQ ID NO:18示出的氨基酸序列对应来编号。

15. 根据权利要求14所述的重组多肽或其活性片段, 其中所述残基位置68-95用VAGTIASVSGVSGVMHNATLVPVKV替换。

16. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽或其活性片段在SWT77-进化枝中。

17. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽或其活性片段在SWT22-进化枝中。

18. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽或其活性片段在WP026675114-进化枝中。

19. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽或其活性片段在BspAG00296-进化枝中。

20. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中蛋白酶活性包括酪蛋白水解。

21. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述蛋白酶活性包括二甲基酪蛋白水解。

22. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽在表面活性剂存在下具有蛋白酶活性。

23. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽在5至12的pH范围下保持其最大蛋白酶活性的至少50%。

24. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽在7至11的pH范围下保持其最大蛋白酶活性的至少50%。

25. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽在55℃至80℃的温度范围下保持其最大蛋白酶活性的至少50%。

26. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽在45℃至75℃的温度范围下保持其最大蛋白酶活性的至少50%。

27. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽在胁迫条件下于50℃下20分钟之后保持至少80%活性。

28. 根据权利要求27所述的重组多肽或其活性片段, 其中所述胁迫条件是在LAS/EDTA测定中。

29. 根据权利要求27所述的重组多肽或其活性片段, 其中所述胁迫条件是在Tris/EDTA测定中。

30. 根据权利要求27所述的重组多肽或其活性片段, 其中所述胁迫条件是在HDL测定中。

31. 根据上述权利要求中任一项所述的重组多肽或其活性片段, 其中所述多肽具有在洗涤剂组合物中的清洁活性。

32. 根据权利要求31所述的重组多肽, 其中所述清洁活性包括水解选自下列的底物: 卵黄、血液、奶、墨以及它们的组合。

33. 根据权利要求31或32所述的重组多肽或其活性片段, 其中所述洗涤剂组合物选自: 衣物洗涤剂、织物软化洗涤剂、自动盘碟洗涤剂、手洗盘碟洗涤剂和硬质表面清洁洗涤剂。

34. 根据权利要求31-33中任一项所述的重组多肽或其活性片段, 其中所述洗涤剂组合

物选自：液体、粉末、颗粒、固体、单个单位剂量、片剂、凝胶、糊剂、棒和片材。

35. 根据上述权利要求中任一项所述的重组多肽或其活性片段，其中所述重组多肽或其活性片段包含选自下列的至少一个置换：X003N、X006R、X010E、X020I、X026N、X028R、X029I、X038A、X041P、X042N、X044R、X048D、X053R、X059G、X061G、X085Q、X088R、X090I、X096G、X098N、X103M、X104Y、X107Q、X113A、X115S、X117N、X131D、X132S、X133D、X136N、X137N、X138I、X139N、X143S、X144S、X146T、X147L、X157R、X168N、X169A、X178N、X179R、X180T、X204Y、X207G、X208Q、X209F、X210R、X212L、X219T、X222V、X229I、X230K、X231S、X231A、X239T、X240Q、X241V、X243N、X245L、X246R、X247D、X255L、X256N、X257Q、X264N、X266Y、X271A和X273G。

36. 根据上述权利要求中任一项所述的重组多肽或其活性片段，其中所述重组多肽或其活性片段包含选自下列的至少一个置换：P003N、Q006R、N010E、T020I、S026N、I028R、Q029I、H038A、Q041P、S042N、A044R、N048D、Q053R、S059G、M061G、H085Q、T088R、V090I、N096G、S098N、L103M、F104Y、T107Q、S113A、D115S、G117N、N131D、Q132S、S133D、A136N、A137N、A138I、Q139N、N143S、A144S、S146T、I147L、A157R、S168N、V169A、T178N、G179R、A180T、V204Y、N207G、G208Q、Y209F、A210R、F212L、S219T、A222V、N229I、R230K、A231S、V231A、S239T、N240Q、A241V、S243N、M245L、Q246R、N247D、P255L、T256N、F257Q、D264N、N266Y、Q271A和S273G。

37. 一种组合物，其包含表面活性剂和根据权利要求1-36中任一项所述的重组多肽。

38. 根据权利要求37所述的组合物，其中所述表面活性剂选自：阴离子表面活性剂、阳离子表面活性剂、两性离子表面活性剂、两性表面活性剂、半极性非离子表面活性剂以及它们的组合。

39. 根据权利要求37所述的组合物，其中所述表面活性剂为阴离子表面活性剂或阳离子表面活性剂。

40. 根据权利要求37所述的组合物，其中所述表面活性剂为非离子表面活性剂。

41. 根据权利要求37-40中任一项所述的组合物，其中所述组合物为洗涤剂组合物。

42. 根据权利要求41所述的组合物，其中所述洗涤剂组合物选自：衣物洗涤剂、织物软化洗涤剂、自动盘碟洗涤剂、手洗盘碟洗涤剂和硬质表面清洁洗涤剂。

43. 根据权利要求41或42所述的组合物，其中所述组合物选自：液体、粉末、颗粒、固体、单个单位剂量、片剂、凝胶、糊剂、棒和片材。

44. 根据权利要求37-43中任一项所述的组合物，其中所述组合物还包含至少一种钙离子和/或锌离子。

45. 根据权利要求37-44中任一项所述的组合物，其中所述组合物还包含至少一种稳定剂。

46. 根据权利要求37-45中任一项所述的组合物，其中所述组合物包含约0.001重量%至约1.0重量%的根据权利要求1-36中任一项所述的重组多肽。

47. 根据权利要求37-46中任一项所述的组合物，其还包含至少一种漂白剂。

48. 根据权利要求37-47中任一项所述的组合物，其中所述组合物不含磷酸盐。

49. 根据权利要求37-47中任一项所述的组合物，其中所述组合物包含磷酸盐。

50. 根据权利要求37-49中任一项所述的组合物，其中所述组合物不含硼。

51. 根据权利要求37-49中任一项所述的组合物,其中所述组合物包含硼。

52. 根据权利要求37-51中任一项所述的组合物,其还包含至少一种助剂成分。

53. 根据权利要求37-52中任一项所述的组合物,其还包含选自下列的一种或多种附加的酶或酶衍生物:酰基转移酶、 α -淀粉酶、 β -淀粉酶、 α -半乳糖苷酶、阿拉伯糖苷酶、芳基酯酶、 β -半乳糖苷酶、卡拉胶酶、过氧化氢酶、纤维二糖水解酶、纤维素酶、软骨素酶、角质酶、内切- β -1,4葡聚糖酶、内切- β -甘露聚糖酶、酯酶、外切-甘露聚糖酶、半乳聚糖酶、葡糖淀粉酶、半纤维素酶、透明质酸酶、角蛋白酶、漆酶、乳糖酶、木质素酶、脂肪酶、脂氧合酶、甘露聚糖酶、氧化酶、果胶酸裂解酶、果胶乙酰酯酶、果胶酶、戊聚糖酶、过氧化物酶、酚氧化酶、磷酸酶、磷脂酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、普鲁兰酶、还原酶、聚鼠李糖半乳糖醛酸酶(rhamnogalacturonases)、 β -葡聚糖酶、鞣酸酶、转谷氨酰胺酶、木聚糖乙酰酯酶、木聚糖酶、木葡聚糖酶、木糖苷酶、金属蛋白酶、附加的丝氨酸蛋白酶以及它们的组合。

54. 根据权利要求37-53中任一项所述的组合物,其中所述组合物在约7至约12的pH下配制。

55. 一种清洁的方法,所述方法包括使表面或物品与包含下列物质的组合物接触:(i)缓冲液和根据权利要求1-36中任一项所述的重组多肽,或(ii)根据权利要求37-54中任一项所述的组合物。

56. 根据权利要求55所述的方法,其中所述物品为盘碟或织物。

57. 根据权利要求55-56所述的方法,所述方法还包括在将所述表面或物品与所述组合物接触之后冲洗所述表面或物品的步骤。

58. 根据权利要求57所述的方法,所述方法还包括在所述冲洗所述表面或物品之后干燥所述表面或物品的步骤。

59. 一种清洁表面或物品的方法,所述方法包括:提供根据权利要求37-54中任一项所述的组合物和需要清洁的表面或物品;以及使所述组合物与所述需要清洁的表面或物品在适于清洁所述表面或物品的条件下接触以产生经清洁的表面或物品。

60. 根据权利要求59所述的方法,所述方法还包括冲洗所述经清洁的表面或物品以产生经冲洗的表面或物品的步骤。

61. 根据权利要求60所述的方法,所述方法还包括干燥所述经冲洗的表面或物品的步骤。

62. 一种用于产生重组多肽的方法,所述方法包括:

(a)用包含编码根据权利要求1-36中任一项所述的多肽的多核苷酸的表达载体稳定地转化宿主细胞;

(b)在适用于所述宿主细胞产生所述多肽的条件下培养所述经转化的宿主细胞;以及

(c)回收所述多肽。

63. 根据权利要求62所述的方法,其中所述宿主细胞为细菌细胞或丝状真菌。

64. 根据权利要求62或63所述的方法,其中所述宿主细胞为芽孢杆菌属物种(*Bacillus* spp.)、链霉菌属物种(*Streptomyces* spp.)、埃希氏菌属物种(*Escherichia* spp.)、曲霉属物种(*Aspergillus* spp.)、木霉属物种(*Trichoderma* spp.)、假单胞菌属物种(*Pseudomonas* spp.)、棒状杆菌属物种(*Corynebacterium* spp.)、酵母属物种(*Saccharomyces* spp.)、或毕赤酵母属物种(*Pichia* spp.)。

65. 根据权利要求62-64中任一项所述的方法, 其中所述表达载体包含编码异源前肽的异源多核苷酸序列。

66. 根据权利要求62-65中任一项所述的方法, 其中所述表达载体包含异源启动子和编码异源信号肽的多核苷酸序列中的一者或两者。

67. 根据权利要求62-66中任一项所述的方法, 其中所述宿主细胞在培养基或发酵液中培养。

68. 一种多核苷酸, 其包含编码选自下列的氨基酸序列的核酸序列: SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、和SEQ ID NO:44。

69. 一种表达载体, 其包含根据权利要求68所述的多核苷酸。

70. 一种宿主细胞, 其用根据权利要求69所述的载体转化。

71. 根据权利要求70所述的宿主细胞, 其中所述宿主细胞为选自下列的菌种: 芽孢杆菌属物种(*Bacillus* spp.)、链霉菌属物种(*Streptomyces* spp.)、埃希氏菌属物种(*Escherichia* spp.)、曲霉属物种(*Aspergillus* spp.)、木霉属物种(*Trichoderma* spp.)、假单胞菌属物种(*Pseudomonas* spp.)、棒状杆菌属物种(*Corynebacterium* spp.)、酵母属物种(*Saccharomyces* spp.)、或毕赤酵母属物种(*Pichia* spp.)。

72. 根据权利要求71所述的宿主细胞, 其中所述芽孢杆菌属物种(*Bacillus* spp.)为枯草芽孢杆菌(*Bacillus subtilis*)。

73. 一种纺织物或皮革处理组合物, 其包含根据权利要求1-36中任一项所述的多肽。

74. 一种毛加工、动物饲料、接触镜片清洁、伤口清洁组合物, 其包含根据权利要求1-36中任一项所述的多肽。

芽孢杆菌属的丝氨酸蛋白酶

[0001] 相关申请的交叉引用

[0002] 本申请要求2014年3月21日提交的美国临时专利申请号61/968853的权益,其内容全文以引用方式并入本文。

[0003] 发明背景

[0004] 本公开涉及从芽孢杆菌属物种(*Bacillus* spp.)克隆的丝氨酸蛋白酶及其变体。包含丝氨酸蛋白酶的组合物适用于清洁织物和硬质表面,以及多种工业应用。

[0005] 丝氨酸蛋白酶是具有引发蛋白质肽键的水解的活性位点丝氨酸的酶(EC 3.4.21)。丝氨酸蛋白酶基于其结构有两大类别:胰凝乳蛋白酶样(胰蛋白酶样)和枯草杆菌蛋白酶样。原型枯草杆菌蛋白酶(EC 3.4.21.62)初始从枯草芽孢杆菌(*B. subtilis*)获得。枯草杆菌蛋白酶及其同源物是MEROPS分类方案的S8肽酶家族的成员。S8家族的成员在其氨基酸序列中具有顺序为Asp、His和Ser的催化三联体。

[0006] 虽然丝氨酸蛋白酶很早就工业酶领域中已知,但仍然需要适用于特定的条件和应用的工程化蛋白酶。

发明内容

[0007] 本发明的组合物和方法涉及从芽孢杆菌属物种(*Bacillus* spp.)克隆的重组丝氨酸蛋白酶及其变体。包含丝氨酸蛋白酶的组合物适用于清洁织物和硬质表面,以及多种工业应用。

[0008] 在一些实施方案中,本发明的多肽或其活性片段可以为新型的多肽或上述本发明多肽中的任一种,其中重组多肽或其活性片段包含DTGIDXXHXXLXNLVXTSLGXSVGGXXXDVXGH基序,其中起始D为活性位点天冬氨酸,末端H为活性位点组氨酸,并且X为任意氨基酸。在一些实施方案中,本发明的多肽或其活性片段可以为新型的多肽或上述本发明多肽中的任一种,其中重组多肽或其活性片段包含DTGIDXXHXXLXNLVXTSLGXSVGGXXXDVXGH基序,其中起始D为活性位点天冬氨酸,末端H为活性位点组氨酸,并且X为任意氨基酸,前体条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP010283106或WP006679321的氨基酸序列。在一些实施方案中,本发明是重组多肽或其活性片段,其中所述重组多肽或其活性片段包含DTGIDXXHXXLXaNLVXTSLGXSVGGXbXXcDVXGH基序,其中起始D为活性位点天冬氨酸,末端H为活性位点组氨酸,并且X、Xa、Xb和Xc为任意氨基酸,前提条件是当Xa为精氨酸时,Xb和Xc不为甘氨酸。在一些实施方案中,基序的VXG序列为VQG。在一些实施方案中,VQG序列在残基位置63-65处,其中所述多肽或其活性片段的氨基酸位置通过与SEQ ID NO:7示出的氨基酸序列对应来编号。在一些实施方案中,所述多肽或其活性片段包含在残基位置80-82处的VSG序列,其中所述多肽或其活性片段的氨基酸位置通过与SEQ ID NO:7示出的氨基酸序列对应来编号。

[0009] 在一些实施方案中,本发明是重组多肽或其活性片段,相比于SEQ ID NO:18,所述多肽或其活性片段具有至少一个氨基酸残基的插入,其中所述插入介于残基位置39-47之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号。在一些实施

方案中,残基位置39-47用HQSLANLVNTSLG替换,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号。在一些实施方案中,本发明是重组多肽或其活性片段,相比于SEQ ID NO:18,所述多肽或其活性片段具有至少一个氨基酸残基的缺失,其中所述缺失介于残基位置51-64之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号。在一些实施方案中,残基位置51-64用VGGSTMDVQGH替换,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号。在一些实施方案中,本发明是重组多肽或其活性片段,相比于SEQ ID NO:18,所述多肽或其活性片段具有至少一个氨基酸残基的缺失,其中所述缺失介于残基位置68-95之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号。在一些实施方案中,残基位置68-95用VAGTIASYGSVSGVMHNATLVPVKV替换,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号。

[0010] 在一些实施方案中,本发明是在WHY-进化枝中的重组多肽或其活性片段。在一些实施方案中,本发明是在SWT77-进化枝中的重组多肽或其活性片段。在一些实施方案中,本发明是在SWT22-进化枝中的重组多肽或其活性片段。在一些实施方案中,本发明是在WP026675114-进化枝中的重组多肽或其活性片段。在一些实施方案中,本发明是在BspAG00296-进化枝中的重组多肽或其活性片段。

[0011] 在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少70%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少70%的同一性,前体条件是所述氨基酸序列不包含WP010283106、W02012175708-0002、W02012175708-0004、或W02012175708-0006。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列具有与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列的至少75%同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少75%的同一性,前体条件是所述氨基酸序列不包含WP010283106、W02012175708-0002、W02012175708-0004、或W02012175708-0006。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少75%的同一性,前体条件是所述氨基酸序列不包含WP010283106、W02012175708-0002、W02012175708-0004、或W02012175708-0006。

NO:44的氨基酸序列具有至少80%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少80%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002或W02012175708-0004。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少85%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少85%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少85%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002或W02012175708-0004。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少90%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少90%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少90%的同一性,前体条件是氨基酸序列不包含W02012175708-0004。在一些实施方案中,本发明是重组多肽或其活性片段,其包含SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列。

[0012] 在一些实施方案中,所述重组多肽具有蛋白酶活性,具体地讲酪蛋白水解。在一些实施方案中,所述重组多肽在8至12的pH范围下保持其最大蛋白酶活性的至少50%。在一些实施方案中,所述重组多肽在50℃至75℃的温度范围下保持其最大蛋白酶活性的至少50%。在一些实施方案中,所述重组多肽具有在包括自动盘碟洗涤剂 and 衣物洗涤剂在内的

洗涤剂组合物中的清洁活性。

[0013] 在一些实施方案中,本发明是包含表面活性剂和上述重组多肽的组合物。在一些实施方案中,表面活性剂选自非离子表面活性剂、阴离子表面活性剂、阳离子表面活性剂、两性离子表面活性剂、两性表面活性剂、半极性非离子表面活性剂以及它们的组合。在一些实施方案中,所述组合物是洗涤剂组合物,诸如衣物洗涤剂、织物软化洗涤剂、餐具洗涤剂和硬质表面清洁洗涤剂。在一些实施方案中,所述组合物还包含至少一种钙离子和/或锌离子、至少一种稳定剂、至少一种漂白剂、磷酸盐或硼酸盐。在一些实施方案中,所述组合物不含磷酸盐和/或不含硼酸盐。在一些实施方案中,所述组合物是颗粒、粉末、固体、棒、液体、片剂、凝胶、糊剂或单位剂量组合物。在一些实施方案中,所述组合物还包含选自下列的一种或多种附加的酶或酶衍生物:酰基转移酶、 α -淀粉酶、 β -淀粉酶、 α -半乳糖苷酶、阿拉伯糖苷酶、芳基酯酶、 β -半乳糖苷酶、卡拉胶酶、过氧化氢酶、纤维二糖水解酶、纤维素酶、软骨素酶、角质酶、内切- β -1,4葡聚糖酶、内切- β -甘露聚糖酶、酯酶、外切-甘露聚糖酶、半乳聚糖酶、葡糖淀粉酶、半纤维素酶、透明质酸酶、角蛋白酶、漆酶、乳糖酶、木质素酶、脂肪酶、脂氧合酶、甘露聚糖酶、氧化酶、果胶酸裂解酶、果胶乙酰酯酶、果胶酶、戊聚糖酶、过氧化物酶、酚氧化酶、磷酸酶、磷脂酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、普鲁兰酶、还原酶、聚鼠李糖半乳糖醛酸酶(rhamnogalacturonases)、 β -葡聚糖酶、鞣酸酶、转谷氨酰胺酶、木聚糖乙酰酯酶、木聚糖酶、木葡聚糖酶、木糖苷酶、金属蛋白酶、附加的丝氨酸蛋白酶以及它们的组合。

[0014] 在一些实施方案中,本发明是清洁方法,该方法包括使表面或物品与上文所列的组合物接触。在一些实施方案中,本发明是用于产生重组多肽的方法,该方法包括用包含编码上述重组多肽的多核苷酸的表达载体稳定转化宿主细胞。

[0015] 附图简述

[0016] 图1提供用于表达BspAG00296丝氨酸蛋白酶的pHYT-BspAG00296的质粒图谱。

[0017] 图2提供用于表达BspM04033丝氨酸蛋白酶的pBN-BspM04033的质粒图谱。

[0018] 图3提供BspAG00296对DMC底物的蛋白酶活性的曲线图。

[0019] 图4提供BspM04033对DMC底物的蛋白酶活性的曲线图。

[0020] 图5提供重垢型液体(HDL)衣物洗涤剂中BspAG00296的清洁效率曲线。

[0021] 图6提供重垢型干燥(HDD)衣物洗涤剂中BspAG00296的清洁效率曲线。

[0022] 图7提供自动盘碟洗涤(ADW)洗涤剂中BspAG00296的清洁效率曲线。

[0023] 图8提供重垢型液体(HDL)衣物洗涤剂中BspM04033的清洁效率曲线。

[0024] 图9提供重垢型干燥(HDD)衣物洗涤剂中BspM04033的清洁效率曲线。

[0025] 图10提供自动盘碟洗涤(ADW)洗涤剂中BspM04033的清洁效率曲线。

[0026] 图11提供WHY-进化枝、SWT77-进化枝、BspAG00296-进化枝、WP026675114-进化枝、和SWT22-进化枝枯草杆菌蛋白酶、以及各种其它细菌丝氨酸蛋白酶的系統发育树。

[0027] 图12A-1-图12E提供枯草杆菌蛋白酶BspAG00296、BspM04033、BspW01765、BspAA02831、SWT4、SWT22、SWT32、SWT40、SWT41、SWT77、SWT123的氨基酸序列与多种其它细菌丝氨酸蛋白酶的序列的CLUSTAL W比对。1JEA和1CSE结构中的残基编号相对于枯草杆菌蛋白酶BPN';然而BspM04033和所示的所有其它蛋白酶的残基的编号是连续线性序列。

[0028] 图13A-图13B提供WHY-进化枝氨基酸序列的区域的基于结构的比对,所述区域包

含被催化残基D33和H66(根据BspM04033线性序列编号的残基)包括的基序。

[0029] 图14提供存在于WHY-进化枝枯草杆菌蛋白酶中的序列基序变化的结构图像。

[0030] 图15A提供示出来自BspAG00296和SWT77-tr的晶体结构的单体的叠加的示意图。

[0031] 图15B提供当SWT77-tr的结构与迟缓芽孢杆菌(*B. lentus*)枯草杆菌蛋白酶相比时,发现的序列基序变化的结构图像。

[0032] 详细描述

[0033] 本发明描述了涉及来自多种芽孢杆菌属菌种的重组丝氨酸蛋白酶的组合物和方法。所述组合物和方法部分地基于重组BspAG00296和BspM04033等在表面活性剂存在下、在碱性反应条件下、和在升高的温度下具有蛋白酶活性的观察结果。预期由SWT77、BspW01765、BspAA02831、SWT4、SWT22、SWT32、SWT40、SWT41和SWT123共享的BspAG00296、BspM04033的这些特征结构使得这些蛋白酶良好地适用于清洁织物和硬质表面,以及纺织物、皮革和毛加工。新蛋白酶也良好地适用于包含在用于蛋白质降解的组合物中,其包括但不限于衣物洗涤剂 and 盘碟洗涤剂。

[0034] I. 定义

[0035] 在详细描述本发明的组合物和方法之前,为了清楚起见对以下术语进行定义。未定义的术语和缩写应该与它们在本领域中所用的普通含义一致。除非在本文中另有定义,否则本文所用的全部技术和科学术语都具有本领域普通技术人员通常理解的相同的含义。除非另有指示,本公开的实施例涉及常用于分子生物学、蛋白质工程、微生物学、和重组DNA的常规技术,其均在本领域的技术范围内。此类技术对本领域技术人员是已知的并且描述于本领域技术人员熟知的多种教材和工具书中。尽管与本文所述的那些类似或等效的任何方法和材料均可用于本公开的实施例,但在本文中对一些合适的方法和材料进行了描述。紧接下文定义的术语通过参考说明书整体而得到更充分的描述。

[0036] 如本文所用,除非上下文明确另外指出,否则单数冠词“一个”、“一种”和“所述”涵盖多个指代物。除非另外指出,否则核酸序列以5'至3'取向从左至右写出;并且氨基酸序列以氨基至羧基取向从左至右写出。应当理解,除非有相反指示,否则本公开内容不局限于本文中所述的具体方法、方案和试剂。

[0037] 预期,本说明书通篇给出的每一最大数值限度包括每一较低数值限度,如同这样的较低数值限度在本文中明确地写出一样。本说明书通篇给出的每一最小数值限度将包括每一较高数值限度,如同这样的较高数值限度在本文中明确地写出一样。本说明书通篇给出的每一数值范围将包括落入这样的较宽数值范围内的每一较窄数值范围,如同这样的较窄数值范围在本文中明确地写出一样。

[0038] 本文中结合数值使用的术语“约/大约”指该数值的 ± 0.5 的范围,除非在上下文对该术语另外特别限定。例如,短语“约6的pH值”指5.5至6.5的pH值,除非对该pH值另外特别限定。

[0039] 如本文所用,术语“蛋白酶(“protease”和“proteinase”)”指能够分解蛋白质和肽的酶。蛋白酶能够通过水解肽键来进行“蛋白水解”,所述肽键将形成蛋白质的肽链或多肽链中的氨基酸连接在一起。将蛋白酶作为蛋白质消化酶的这种活性称之为“蛋白水解活性”。存在很多用于测量蛋白水解活性的已知过程。例如,可通过分析相应蛋白酶水解合适底物的能力的比较测定来确定蛋白水解活性。可用于分析蛋白酶或蛋白水解活性的示例性

底物包括但不限于二甲基酪蛋白(Sigma C-9801)、牛胶原(Sigma C-9879)、牛弹性蛋白(Sigma E-1625)和牛角蛋白(ICN Biomedical 902111)。利用这些底物的比色测定在本领域中是熟知的(参见,例如WO 99/34011和美国专利6,376,450)。pNA肽基测定(参见,例如Del Mar等人,Anal Biochem,99:316-320,1979)也用于确定活性酶浓度。该测定测量当酶水解可溶性合成底物诸如琥珀酰基-丙氨酸-丙氨酸-脯氨酸-苯丙氨酸-对硝基苯胺(suc-AAPF-pNA)时释放对硝基苯胺的速率。在分光光度计上于410nm下测量由水解反应产生黄色的速率,并且其与活性酶浓度成比例。此外,可使用在280纳米(nm)下的吸光度测量结果来确定经纯化蛋白质的样品中的总蛋白质浓度。对底物的活性/蛋白质浓度得到酶比活性。

[0040] 如本文结合多肽诸如蛋白酶所用,术语“变体”是指包含至少一个氨基酸残基与亲本多肽或参考多肽的氨基酸序列不同的氨基酸序列的多肽(包括但不限于野生型多肽)。所述不同可以是插入、缺失或置换的改变。在一些实施方案中,与亲本多肽或参考多肽的氨基酸序列不同的多肽变体包含一种或多种天然存在的或人工制造的氨基酸的置换、插入或缺失。在其它实施方案中,与亲本多肽或参考多肽的氨基酸序列不同的多肽变体包含一种或多种天然存在的氨基酸的置换、插入或缺失。在其它实施方案中,与亲本多肽或参考多肽的氨基酸序列不同的多肽变体包含一种或多种人工制造的氨基酸的置换、插入或缺失。

[0041] 如本文所用,“芽孢杆菌属”包括“芽孢杆菌”属内的所有物种,如本领域技术人员所已知的,包括但不限于,枯草芽孢杆菌(*B. subtilis*)、地衣芽孢杆菌(*B. licheniformis*)、迟缓芽孢杆菌(*B. lentus*)、短芽孢杆菌(*B. brevis*)、嗜热脂肪芽孢杆菌(*B. stearothermophilus*)、嗜碱芽孢杆菌(*B. alkalophilus*)、解淀粉芽孢杆菌(*B. amyloliquefaciens*)、克劳氏芽孢杆菌(*B. clausii*)、索诺拉沙漠芽孢杆菌(*B. sonorensis*)、耐盐芽孢杆菌(*B. halodurans*)、短小芽孢杆菌(*B. pumilus*)、灿烂芽孢杆菌(*B. lautus*)、饲料芽孢杆菌(*B. pabuli*)、蜡状芽孢杆菌(*B. cereus*)、黏琼脂芽孢杆菌(*B. agaradhaerens*)、秋叶氏芽孢杆菌(*B. akibai*)、克氏芽孢杆菌(*B. clarkii*)、假坚强芽孢杆菌(*B. pseudofirmus*)、列城芽孢杆菌(*B. lehensis*)、巨大芽孢杆菌(*B. megaterium*)、凝结芽孢杆菌(*B. coagulans*)、环状芽孢杆菌(*B. circulans*)、吉氏芽孢杆菌(*B. gibsonii*)和苏云金芽孢杆菌(*B. thuringiensis*)。认识到,芽孢杆菌属持续经历着分类学上的重构。因此,预期的是该属包括已被重新归类的种,包括但不限于诸如嗜热脂肪芽孢杆菌(其现在被命名为“嗜热脂肪地芽孢杆菌(*Geobacillus tearothermophilus*)”)、或多粘芽孢杆菌(*B. polymyxa*)(其现在为“多粘类芽孢杆菌”)。认为在应激环境条件下产生抗性内生孢子是芽孢杆菌属的限定特征,但是该特征还适用于最近命名的脂环酸芽孢杆菌属(*Alicyclobacillus*)、兼性芽孢杆菌属(*Amphibacillus*)、硫胺素芽孢杆菌属(*Aneurinibacillus*)、厌氧芽孢杆菌属(*Anoxybacillus*)、短芽孢杆菌属(*Brevibacillus*)、线状芽孢杆菌属(*Filobacillus*)、薄壁芽孢杆菌属(*Gracilibacillus*)、喜盐芽孢杆菌属(*Halobacillus*)、类芽孢杆菌属(*Paenibacillus*)、需盐芽孢杆菌属(*Salibacillus*)、耐热芽孢杆菌属(*Thermobacillus*)、*Ureibacillus*和枝芽孢杆菌属(*Virgibacillus*)。

[0042] 在本文中互换使用的术语“多核苷酸”和“核酸”是指具有任何长度的在链中共价键合的核苷酸单体的聚合物。DNA(脱氧核糖核苷酸)、包含脱氧核糖核苷酸的多核苷酸、和RNA(核糖核苷酸)、核糖核苷酸的聚合物是具有不同生物功能的多核苷酸或核酸的示例。多核苷酸或核酸包括但不限于单链、双链或三链DNA、基因组DNA、cDNA、RNA、DNA-RNA杂合体或

者包含嘌呤和嘧啶碱基或其它天然的、经化学修饰的、经生物化学修饰的、非天然或衍生的核苷酸碱基的聚合物。以下为多核苷酸的非限制性示例：基因、基因片段、染色体片段、一个或多个经表达的序列标签(EST)、外显子、内含子、信使RNA(mRNA)、转录RNA(tRNA)、核糖体RNA(rRNA)、核酶、互补DNA(cDNA)、重组多核苷酸、分支多核苷酸、质粒、载体、任意序列的分离DNA、任意序列的分离RNA、核酸探针和引物。

[0043] 如本文所用，术语“突变”是指相对于参考氨基酸或核酸序列进行的改变。预期，该术语涵盖置换、插入和缺失。

[0044] 如本文所用，术语“载体”是指用于将核酸引入或转移到靶细胞或组织中的核酸构建体。载体通常用于将外来DNA引入细胞或组织中。载体包括质粒、克隆载体、噬菌体、病毒(例如，病毒载体)、黏粒、表达载体、穿梭载体等。载体通常包含复制起点、多克隆位点和选择性标记。通常将载体插入靶细胞中的方法称为转化。在一些实施方案中，本发明包括含有与合适原序列(例如，分泌序列、信号肽序列等)可操作地连接的编码丝氨酸蛋白酶多肽(例如，前体或成熟丝氨酸蛋白酶多肽)的DNA序列的载体，所述原序列能够实现该DNA序列在合适宿主中的表达以及重组多肽链的折叠和转位。

[0045] 如本文所用，术语“表达盒”、“表达质粒”或“表达载体”是指用于在靶细胞中表达目的核酸的重组或合成产生的核酸构建体或载体。表达载体或表达盒通常包含驱动外来核酸表达的启动子核苷酸序列。表达载体或表达盒通常还包含允许特定核酸在靶细胞中转录的任何其它指定的核酸元件。可将重组表达盒并入质粒、染色体、线粒体DNA、质体DNA、病毒或核酸片段中。很多原核和真核表达载体均可商购获得。

[0046] 如本文所用，“质粒”是指能够独立于染色体DNA复制的染色体外DNA分子。质粒是双链的且可以是环状的并且通常用作克隆载体。

[0047] 如本文在将核酸序列引入细胞中的语境中所用，术语“引入”是指任何适于将核酸序列转移到细胞中的方法。用于引入的此类方法包括但不限于原生质体融合、转染、转化、电穿孔、缀合和转导。转化是指由摄取、任选的基因组并入和遗传物质(例如，DNA)表达引起的细胞的遗传改变。

[0048] 如本文所用，当将核酸置于与另一核酸序列的功能性关系中时，其与另一核酸序列“可操作地连接”。例如，如果启动子影响编码序列的转录，则该启动子或增强子与核苷酸编码序列可操作地连接。如果将核糖体结合位点定位成有利于编码序列的翻译，则其可与编码序列可操作地连接。通常来说，“可操作连接的”DNA序列是连续的。然而，增强子无需是连续的。连接通过在方便的限制性位点进行连接来实现。如果这样的位点不存在，则可根据常规实践使用合成的寡核苷酸衔接子或接头。

[0049] 如本文所用，术语“基因”是指编码多肽并且包含编码区之前和之后的区域的多核苷酸(例如，DNA区段)。在一些情况下，基因包含在各个编码区段(外显子)之间的间插序列(内含子)。

[0050] 如本文所用，当提及细胞使用时，“重组的”通常指已通过引入外来核酸序列而对该细胞进行了修饰或者指该细胞来源于经如此修饰的细胞。例如，重组细胞可包含在天然形式(非重组)细胞中不以相同形式存在的基因，或者重组细胞可包含已进行修饰并且重新引入细胞中的天然基因(存在于天然形式细胞中)。重组细胞可包含细胞的同源核酸，所述细胞已经进行修饰但未从该细胞中除去核酸；这样的修饰包括通过基因替换、位点特异性

突变以及本领域普通技术人员已知的相关技术获得的那些。重组DNA技术包括用于体外产生重组DNA并将该重组DNA转移到在此其可表达或增殖的细胞中,从而产生重组多肽的技术。多核苷酸或核酸的“重组”一般是指装配或组合两个或更多个核酸或多核苷酸链或片段以产生新的多核苷酸或核酸。

[0051] 如果核酸或多核苷酸处于其天然状态或者当通过本领域技术人员已知的方法操作时可转录和/或翻译以产生多肽或其片段,可认为该核酸或多核苷酸“编码”多肽。还认为这样的核酸的反义链编码序列。

[0052] 术语“宿主菌株”和“宿主细胞”是指用于表达含有目的DNA序列的载体的合适宿主。

[0053] “蛋白质”或“多肽”包含氨基酸残基的聚合物序列。术语“蛋白质”和“多肽”在本文中可互换使用。本公开内容通篇使用遵照IUPAC-IUB生物化学命名联合委员会(Joint Commission on Biochemical Nomenclature(JCBN))定义的氨基酸的单字母和三字母代码。单字母X是指二十种氨基酸中的任一种。还应理解,由于遗传密码的简并性,多肽可由一种以上核苷酸序列编码。突变可如下命名:亲本氨基酸的单字母代码,接着是位置编号,之后是变体氨基酸的单字母代码。例如,对于甘氨酸(G)在位置87处突变为丝氨酸(S)表示为“G087S”或“G87S”。突变还可通过如下命名:使用氨基酸的三字母代码,接着是按从N端计其在多肽链中的位置;例如,A1a10为在位置10处的丙氨酸。多个突变通过在突变之间插入“-”、“+”、“/”或“;”来指示。在位置87和90处的突变表示为“G087S-A090Y”或“G87S-A90Y”或“G87S+A90Y”或“G087S+A090Y”。就缺失而言,使用一个字母代码“Z”。就相对于亲本序列的插入,一个字母代码“Z”在位置编号的左侧。就缺失而言,一个字母代码“Z”在位置编号的右侧。就插入而言,位置编号是在一个或多个被插入氨基酸之前的位置编号,对于每个氨基酸加上0.01。例如,在位置87和88之间的三个氨基酸丙氨酸(A)、丝氨酸(S)和酪氨酸(Y)的插入示为“Z087.01A-Z087.02S-Z087.03Y”。因此,上述所有突变加上在位置100处的缺失为:“G087S-Z087.01A-Z087.02S-Z087.03Y-A090Y-A100Z”。当描述修饰时,位置其后是列于括号中的氨基酸表示通过所列的任一种氨基酸在该位置处进行置换的列表。例如,6(L,I)意指第6位可置换为亮氨酸或异亮氨酸。

[0054] “原序列”(prosequence)或“前肽序列”是指介于信号肽序列和成熟蛋白酶序列之间的氨基酸序列,其是蛋白酶的适当折叠和分泌所必需的;有时,将其称为分子内伴侣。原序列或前肽序列的切割产生成熟的活性蛋白酶。细菌丝氨酸蛋白酶通常作为酶原表达。

[0055] 术语“信号序列”和“信号肽”是指可参与蛋白质之成熟或前体形式的分泌或直接转运的氨基酸残基的序列。信号序列通常位于前体或成熟蛋白质序列的N端。信号序列可以是内源性的或外源性的。信号序列通常不存在于成熟蛋白质中。信号序列通常在转运蛋白质之后通过信号肽酶从蛋白质中切割。

[0056] 术语蛋白质、多肽或肽的“成熟”形式是指蛋白质、多肽或肽的无信号肽序列和前肽序列的功能性形式。

[0057] 术语蛋白质或肽的“前体”形式是指具有与该蛋白质的氨基或羧基末端可操作地连接的原序列的成熟蛋白质形式。前体也可具有与原序列的氨基末端可操作地连接的“信号”序列。前体还可具有参与翻译后活动的附加多肽(例如,从其切割以留下成熟蛋白质或肽形式的多肽)。

[0058] 术语“野生型”在提及氨基酸序列或核酸序列时指该氨基酸序列或核酸序列是天然序列或天然存在序列。如本文所用,术语“天然存在的”是指见于自然界中的任何物质(例如,蛋白质、氨基酸或核酸序列)。反之,术语“非天然存在的”是指不存在于自然界中的任何物质(例如,在实验室中产生的重组核酸和蛋白质序列以及野生型序列的变化形式)。

[0059] 如本文中关于氨基酸残基位置所用,“对应于”是指蛋白质或肽中所列举位置处的氨基酸残基,或者蛋白质或肽中与所列举残基类似、同源或等同的氨基酸残基。如本文所用,“对应区域”一般是指相关蛋白质或参考蛋白质中的类似位置。

[0060] 术语“来源于”和“获自”不仅是指由所考虑生物的菌株产生或可由其产生的蛋白质,而且还是指由从此菌株分离的DNA序列编码并且在含有此DNA序列的宿主生物体中产生的蛋白质。另外,该术语是指由合成来源和/或cDNA来源的DNA序列编码并且具有所考虑蛋白质的鉴别特征的蛋白质。例如,“来源于芽孢杆菌的蛋白酶”是指由芽孢杆菌天然产生的具有蛋白水解活性的那些酶以及与由芽孢杆菌来源产生的那些类似但通过使用遗传工程基技术由转化有编码丝氨酸蛋白酶的核酸的其它宿主细胞产生的丝氨酸蛋白酶。

[0061] 在两个多核苷酸或多肽序列语境中的术语“相同”是指如使用下文所述和本领域已知的序列比较或分析算法所测量的,两个序列中的核酸或氨基酸在比对最大一致性时相同。

[0062] 如本文所用,“同源基因”或“同源蛋白质”是指彼此相同或非常类似并且据信来源于共同祖先的一对基因或蛋白质。所述术语包括通过物种形成(即新物种的演变)分离的基因或蛋白质(例如,直系同源基因或直系同源蛋白质),以及通过遗传复制分离的基因或蛋白质(例如旁系同源基因或旁系同源蛋白质)。

[0063] 如本文所用,“同一性%”或“百分比同一性”或“PID”是指蛋白质序列的同一性。百分比同一性可使用本领域中已知的标准技术来确定。可用的算法包括BLAST算法(参见,Altschul等人,J Mol Biol,215:403-410,1990;以及Karlin和Altschul,Proc Natl Acad Sci USA,90:5873-5787,1993)。BLAST程序使用数个检索参数,其中大多数被设置为默认值。NCBI BLAST算法发现在生物学相似性方面最相关的序列,但是不建议用于具有少于20个残基的查询序列(Altschul等人,Nucleic Acids Res,25:3389-3402,1997;和Schaffer等人,Nucleic Acids Res,29:2994-3005,2001)。用于核酸序列检索的示例性默认BLAST参数包括:相邻字长阈值=11;E-值截止值=10;计分矩阵=NUC.3.1(匹配=1,失配=-3);空位开放=5;和空位延伸=2。用于氨基酸序列检索的示例性默认BLAST参数包括:字长=3;E-值截止值=10;评分矩阵=BLOSUM62;空位开放=11;和空位延伸=1。百分比(%)氨基酸序列同一性值通过将匹配相同残基的数目除以“参考”序列的残基的总数目(包括程序为了进行最佳/最大比对而产生的任何空位)来确定。BLAST算法将“参考”序列作为“查询序列”。

[0064] 如本文所用,“同源蛋白质”或“同源蛋白酶”指在一级、二级和/或三级结构方面具有明显相似性的蛋白质。当对蛋白质进行比对时,蛋白质源性可指线性氨基酸序列的相似性。蛋白质序列的同源检索可使用来自NCBI BLAST的BLASTP和PSI-BLAST以0.001的阈值(E值截止值)来进行。(Altschul SF, Madde TL, Shaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI BLAST a new generation of protein database search programs. Nucleic Acids Res 1997 Set 1; 25(17):3389-402)。利用这一信息,可对蛋白质序列进行分组。可使用氨基酸序列来建立系统发育树。可将氨基酸序列输入诸如

Vector NTI Advance工具包的程序中并使用邻接(Neighbor Joining,NJ)法来生成指南树(Guide Tree)(Saitou和Nei,Mol Biol Evol,4:406-425,1987)。可使用Kimura序列距离校正并忽略具有空位的位置来计算树构建。诸如AlignX的程序可将计算的距离值在括号中显示在系统发育树上所显示的分子名称之后。

[0065] 了解分子之间的同源性可揭示分子的进化史以及有关其功能的信息。如果新测序的蛋白质与已表征的蛋白质同源,则强烈指示新蛋白质的生物化学功能。两个实体之间的最基本关系是同源性;如果两个分子来源于同一祖先,则认为其是同源的。同源分子或同源物可分为两类:旁系同源物和直系同源物。旁系同源物是存在于一个物种中的同源物。旁系同源物通常在其具体的生物化学功能方面有所不同。直系同源物是存在于不同物种但具有非常相似或相同功能的同源物。蛋白质超家族是可推断出具有共同系谱的最大蛋白质组群(进化枝)。通常来说,该共同系谱基于序列比对和机械相似性。超家族通常包含数个在家族内表现出序列相似性的蛋白质家族。基于MEROPS蛋白酶分类系统,术语“蛋白质宗族”通常用于蛋白酶超家族。

[0066] CLUSTAL W算法是序列比对算法的另一示例(参见,Thompson等人,Nucleic Acids Res,22:4673-4680,1994)。用于CLUSTAL W算法的默认参数包括:空位开放罚分=10.0;空位延伸罚分=0.05;蛋白质加权矩阵=BLOSUM系列;DNA加权矩阵=IUB;延迟发散序列%=40;空位分离距离=8;DNA转换加权=0.50;亲水性残基列表=GPSNDQEK;使用负矩阵=关;Toggle残基特异性罚分=开;Toggle亲水性罚分=开;和Toggle末端空位分离罚分=关。在CLUSTAL算法中,包括发生任一末端的缺失。例如,在具有500个氨基酸的多肽的任一末端(或在多肽内)具有5个氨基酸缺失的变体相对于“参考多肽”具有99%的序列同一性百分比($495/500 \times 100$)。这样的变体涵盖在相对于该多肽具有“至少99%序列同一性”的变体内。

[0067] 当核酸或多核苷酸至少部分或完全与其它组分分离时,其是“分离的”,所述其它组分包括但不限于例如其它蛋白质、核酸、细胞等。类似地,当多肽、蛋白质或肽至少部分或完全与其它组分分离时,其是“分离的”,所述其它组分包括但不限于例如其它蛋白质、核酸、细胞等。按摩尔计,分离的物类比组合物中的其它物类的丰度高。例如,分离的物类可占存在的所有大分子物类的至少约60%、约65%、约70%、约75%、约80%、约85%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%或约100%(按摩尔计)。优选地,感兴趣的物类被纯化至基本同质(即,通过常规的检测方法在组合物中检测不到污染物物类)。纯度和同质性可使用本领域中熟知的多种技术来确定,诸如对核酸或蛋白质样品分别进行琼脂或多聚酰胺凝胶电泳,之后通过染色可视化。如果期望的话,可利用高分辨率技术,诸如高效液相色谱(HPLC)或类似方式来纯化物质。

[0068] 术语“经纯化的”当应用于核酸或多肽时一般指代基本不含其它组分的核酸或多肽,如通过本领域中熟知的分离技术所确定的(例如,经纯化的多肽或多核苷酸在电泳凝胶、色谱洗脱液和/或经受密度梯度离心的介质中形成离散带)。例如,在电泳凝胶中产生基本一条带的核酸或多肽是“经纯化的”。经纯化的核酸或多肽的纯度为至少约50%,通常纯度为至少约60%、约65%、约70%、约75%、约80%、约85%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%、约99.5%、约99.6%、约99.7%、约99.8%或更高(例如,按摩尔计的重量百分比)。在相关方面,当在应用纯化或富集技术之后

分子的浓度显著增加时,组合物富含该分子。术语“富含”指以高于起始组合物的相对或绝对浓度存在于组合物中的化合物、多肽、细胞、核酸、氨基酸或者其它指定材料或组分。

[0069] 如本文所用,术语“功能性测定”是指指示蛋白质活性的测定。在一些实施方案中,该术语是指其中分析蛋白质以其一般能力发挥功能的能力的测定系统。例如,在蛋白酶的情况下,功能性测定涉及确定蛋白酶水解蛋白质性底物的有效性。

[0070] 术语“清洁活性”指丝氨酸蛋白酶多肽或参考蛋白酶在本公开的蛋白水解、水解、清洁或其它过程中普遍的条件下获得的清洁性能。在一些实施方案中,丝氨酸蛋白酶多肽或参考蛋白酶的清洁性能可通过使用用于清洁物品或表面上的一种或多种不同酶敏感性污渍(例如,由食物、草、血液、墨、乳、油和/或卵蛋白产生的污渍)的多种测定来确定。可如下确定变体或参考蛋白酶的清洁性能:使物品或表面上的污渍经受标准的洗涤条件并使用多种色谱、分光光度或其它定量方法来评估除去污渍的程度。示例性清洁测定和方法是本领域已知的,并且包括但不限于描述于W099/34011和美国专利6,605,458中的那些,所述专利文献均以引用方式并入本文,以及在下文提供的实施例中所包括的那些清洁测定和方法。

[0071] 术语丝氨酸蛋白酶多肽或参考蛋白酶的“清洁有效量”是指在特定清洁组合物中实现期望酶促活性水平的蛋白酶的量。本领域普通技术人员可容易地确定这样的有效量,并且其基于很多因素,诸如所使用的具体蛋白酶、清洁应用、清洁组合物的特定组成以及需要的是液体组合物还是干(例如,颗粒、片剂、棒)组合物等。

[0072] 术语“清洁助剂材料”指清洁组合物中包含的除本公开的丝氨酸蛋白酶多肽之外的任何液体、固体或气体材料。在一些实施方案中,本公开内容的清洁组合物包含一种或多种清洁助剂材料。通常基于清洁组合物的具体类型和形式(例如,液体、颗粒、粉末、棒、糊剂、喷雾、片剂、凝胶、泡沫或其它组合物)来选择每种清洁辅助性材料。优选地,每种清洁助剂材料均与组合物中使用的蛋白酶相容。

[0073] 清洁组合物和清洁制剂包括适于清洁、漂白、消毒和/或灭菌任何物体、物品和/或表面的任何组合物。这样的组合物和制剂包括但不限于例如液体和/或固体组合物,包括清洁或洗涤剂组合物(例如,液体、片剂、凝胶、棒、颗粒和/或固体衣物洗涤清洁或洗涤剂组合物以及精细织物洗涤剂组合物;硬质表面清洁组合物和制剂,诸如用于玻璃、木材、陶瓷以及金属台面和橱窗的清洁组合物和制剂;地毯清洁剂;烘箱清洁剂;织物清新剂;织物软化剂;以及纺织物、衣物洗涤增效剂清洁组合物或洗涤剂组合物、衣物洗涤添加剂清洁组合物和衣物洗涤预去斑剂清洁组合物;盘碟洗涤组合物,包括手洗或人工盘碟洗涤组合物(例如,“手洗”或“人工”盘碟洗涤洗涤剂)和自动盘碟洗涤组合物(例如“自动盘碟洗涤洗涤剂”)。单个剂量单位形式还可与本发明一起使用,其包括但不限于丸剂、片剂、胶囊锭、或其它单个剂量单位诸如预先测定的粉末、悬浮液或液体。

[0074] 除非另外指出,否则本文中使用的清洁组合物或清洁制剂包括颗粒或粉末形式的通用或重垢型洗涤剂,尤其是清洁洗涤剂;液体、颗粒、凝胶、固体、片剂、糊剂或单位剂量形式多功能洗涤剂,具体地所谓的重垢型液体(HDL)或重垢型干燥(HDD)洗涤剂类型;液体精细织物洗涤剂;手洗或人工盘碟洗涤剂,包括高度起泡类型的那些;手洗或人工餐具洗涤、自动盘碟洗涤或者盘碟或餐具洗涤剂,包括用于家用和工业用途的多种片剂、粉末、固体、颗粒、液体、凝胶和漂洗助剂类型;液体清洁剂和消毒剂,包括抗菌手洗类型、清洁棒、漱口

剂、义齿清洁剂、洗车液、地毯清洗剂、浴室清洁剂；人以及其它动物用的洗发剂和/或毛发冲洗剂；沐浴凝胶以及泡沫浴和金属清洁剂；以及清洁辅剂，诸如漂白添加剂和“去污棒”或预处理类型。在一些实施方案中，颗粒组合物为“紧凑”形式。在一些实施方案中，液体组合物为“浓缩”形式。

[0075] 如本文所用，术语“织物清洁组合物”包括手洗和机洗衣物洗涤剂组合物，包括衣物洗涤剂添加剂组合物和适用于浸泡和/或预处理沾污织物（例如，衣物、亚麻制品以及其它纺织物材料）的组合物。

[0076] 如本文所用，“非织物清洁组合物”包括非纺织物（即，非织物）表面清洁组合物，包括但不限于例如手洗或人工或自动盘碟洗涤剂组合物、口腔清洁组合物、义齿清洁组合物、接触镜片清洁组合物、伤口清创组合物和个人清洁组合物。

[0077] 如本文所用，术语“洗涤剂组合物”或“洗涤剂制剂”用于指在用于清洁沾污或脏物体（包括特定织物和/或非织物物体或物品）的洗涤介质的组合物。本公开的此类组合物不限于任何特定的洗涤剂组合物或制剂。实际上，在一些实施方案中，本公开的洗涤剂包含本公开的至少一种丝氨酸蛋白酶多肽，并且另外包含一种或多种表面活性剂、转移酶、水解酶、氧化还原酶、助洗剂（助洗剂盐）、漂白剂、漂白活性剂、荧光染料、结块抑制剂、掩蔽剂、酶活化剂、抗氧化剂和/或增溶剂。在一些情况下，助洗剂盐是硅酸盐和磷酸盐的混合物，优选地其中硅酸盐（例如，偏硅酸钠）多于磷酸盐（例如，三聚磷酸钠）。本公开的一些组合物诸如但不限于清洁组合物或洗涤剂组合物不含任何磷酸盐（例如，磷酸盐或磷酸盐助洗剂）。

[0078] 如本文所用，术语“漂白”指在足够的时间长度和/或合适的pH和/或温度条件下处理材料（例如，织物、衣物、纸浆等）以实现增亮（即，美白）和/或清洁该材料。适于漂白的化学物质的示例包括但不限于例如 ClO_2 、 H_2O_2 、过酸类、 NO_2 等。

[0079] 如本文所用，蛋白酶（例如，本公开的丝氨酸蛋白酶多肽）的“洗涤性能”是指丝氨酸蛋白酶多肽对洗涤的贡献，与未向组合物添加丝氨酸蛋白酶多肽的洗涤剂相比，其赋予该洗涤剂附加的清洁性能。洗涤性能在相关的洗涤条件下进行比较。在一些测试系统中，可以以模拟典型的用于特定细分市场（例如，手洗或手动盘碟洗涤、自动盘碟洗涤、盘碟清洁、餐具清洁、织物清洁等）中家庭应用的一个或多个条件的方式来控制其它相关因数，诸如洗涤剂组成、SUD浓度、水硬度、洗涤机器、时间、pH、和/或温度。

[0080] 术语“相关洗涤条件”用于本文以指示条件，具体地讲，实际用于家庭应用中的手洗盘碟洗涤、自动盘碟洗涤、或衣物洗涤剂细分市场中的洗涤温度、时间、洗涤机器、SUD浓度、洗涤剂类型和水硬度。

[0081] 术语“改善的洗涤性能”用于指示在相关洗涤条件下在去污效果方面获得的良好最终结果，或相对于相应的野生型或起始亲本蛋白酶，获得相同最终结果需要按重量计，更少的本公开的丝氨酸蛋白酶多肽。

[0082] 如本文所用，术语“消毒”指从表面去除污染物，以及抑制或杀死物品表面上的微生物。并非旨在使本公开局限于任何具体表面、物品或者所需除去的污染物或微生物。

[0083] 本文中清洁组合物的“紧凑”形式最佳反映在密度，并且就组合物而言，最佳反映在无机填料盐的量。无机填料盐是粉末形式的洗涤剂组合物的常规成分。在常规的洗涤剂组合物中，填料盐以相当量，通常为按总组合物的重量计约17%至约35%存在。相比之下，在紧凑型组合物中，填料盐以不超过总组合物的约15%的量存在。在一些实施方案中，填料

盐以不超过按组合物的重量计约10%，或更优选约5%的量存在。在一些实施方案中，无机填料盐选自碱金属和碱土金属的硫酸盐和盐酸盐。在一些实施方案中，填料盐是硫酸钠。

[0084] II. 丝氨酸蛋白酶多肽

[0085] 本公开提供了新的丝氨酸蛋白酶。本公开的丝氨酸蛋白酶多肽包括分离的、重组的、基本纯的或非天然存在的多肽。在一些实施方案中，所述多肽可用于清洁应用，并且可掺入到可用于清洁有此需要的物品或表面的清洁组合物中。

[0086] 在一些实施方案中，本发明的多肽，或其活性片段可以为WHY-进化枝多肽。WHY-进化枝来源于N端附近的完全保守残基WHY(BspAG00296、BspM04033和该进化枝的其它成员中的W残基位置7)。在一些实施方案中，本发明的多肽，或其活性片段可以为WHY-进化枝多肽，前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP010283106或WP006679321。

[0087] 在一些实施方案中，本发明的多肽或其活性片段可以为新型的多肽或上述本发明多肽中的任一种，其中重组多肽或其活性片段包含DTGIDXXHXXLXNLVXTSLGXSVGGXXXDVXGH基序，其中起始D为活性位点天冬氨酸，末端H为活性位点组氨酸，并且X为任意氨基酸。在一些实施方案中，本发明的多肽或其活性片段可以为新型的多肽或上述本发明多肽中的任一种，其中重组多肽或其活性片段包含DTGIDXXHXXLXNLVXTSLGXSVGGXXXDVXGH基序，其中起始D为活性位点天冬氨酸，末端H为活性位点组氨酸，并且X为任意氨基酸，前体条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP010283106或WP006679321的氨基酸序列。在一些实施方案中，本发明的多肽或其活性片段可以为新型的多肽或上述本发明多肽中的任一种，其中重组多肽或其活性片段包含DTGIDXXHXXLXNLVXTSLGXSVGGXXXDVXGH基序，其中起始D为活性位点天冬氨酸，末端H为活性位点组氨酸，并且X为任意氨基酸，前体条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP026675114、WP025025887、WP010283106或WP006679321的氨基酸序列。

[0088] 在一些实施方案中，本发明的多肽或其活性片段可以为新型的多肽或上述本发明多肽中的任一种，其中重组多肽或其活性片段包含DTGIDXXHXXLXaNLVXTSLGXSVGGXbXXcDVXGH基序，其中起始D为活性位点天冬氨酸，末端H为活性位点组氨酸，并且X、Xa、Xb和Xc为任意氨基酸，前提条件是当Xa为精氨酸时，Xb和Xc不为甘氨酸。在一些实施方案中，基序的VXG序列为VQG。在一些实施方案中，VQG序列在残基位置63-65处，其中所述多肽或其活性片段的氨基酸位置通过与以SEQ ID NO:7示出的氨基酸序列对应来编号。

[0089] 在一些实施方案中，本发明的多肽或其活性片段可以为新型的多肽或上述本发明多肽中任一种，其中所述多肽或其活性片段包含在残基位置80-82处的VSG序列，其中所述多肽或其活性片段的氨基酸位置通过与以SEQ ID NO:7示出的氨基酸序列对应来编号。在一些实施方案中，本发明的多肽或其活性片段可以为新型的多肽或上述本发明多肽中任一种，其中所述多肽或其活性片段包含在残基位置80-82处的VSG序列，其中所述多肽或其活性片段的氨基酸位置通过与以SEQ ID NO:7示出的氨基酸序列对应来编号，并且前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP010283106或WP006679321的氨基酸序列。在一些实施方案中，本发明的多肽或其活性片

段可以为新型多肽或上述本发明多肽中任一种,其中所述多肽或其活性片段包含在残基位置80-82处的VSG序列,其中所述多肽或其活性片段的氨基酸位置通过与SEQ ID NO:7示出的氨基酸序列对应来编号,并且前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP026675114、WP025025887、WP010283106或WP006679321的氨基酸序列。

[0090] 在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的插入,其中所述插入介于残基位置39-47之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号。在其它实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的插入,其中所述插入介于残基位置39-47之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号,并且前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP010283106或WP006679321的氨基酸序列。在其它实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的插入,其中所述插入介于残基位置39-47之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号,并且前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP026675114、WP025025887、WP010283106或WP006679321的氨基酸序列。在一些实施方案中,残基位置39-47用HQSLANLVNTSLG替换。

[0091] 在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的缺失,其中所述缺失介于残基位置51-64之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号。在其它实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的缺失,其中所述缺失介于残基位置51-64之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号,并且前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP010283106或WP006679321的氨基酸序列。在其它实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的缺失,其中所述缺失介于残基位置51-64之间,其中所述残基位置通过与SEQ ID NO:18示出的氨基酸序列对应来编号,并且前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP026675114、WP025025887、WP010283106或WP006679321的氨基酸序列。在一些实施方案中,残基位置51-64用VGGSTMDVQGH、VGGSA/PEDVQGH、VGGNPEDRQGH、或VGGTPADVHGH替换。在一些实施方案中,残基位置51-64用VGGSTMDVQGH替换。在一些实施方案中,残基位置51-64用VGGSA/PEDVQGH替换。在一些实施方案中,残基位置51-64用VGGSAEDVQGH替换。在一些实施方案中,残基位置51-64用VGGSPEDVQGH替换。在一些实施方案中,残基位置51-64用VGGNPEDRQGH替换。在一些实施方案中,残基位置51-64用VGGTPADVHGH替换。

[0092] 在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的缺失,其中所述缺失介于残基位置68-95之间,其中所述残基位置通过与以SEQ ID NO:18示出的氨基酸序列对应来编号,并且前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP010283106或WP006679321的氨基酸序列。在其它实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中相比于SEQ ID NO:18,所述多肽或其活性片段包含至少一个氨基酸残基的缺失,其中所述缺失介于残基位置68-95之间,其中所述残基位置通过与以SEQ ID NO:18示出的氨基酸序列对应来编号,并且前提条件是所述多肽不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP026675114、WP025025887、WP010283106或WP006679321的氨基酸序列。在一些实施方案中,残基位置68-95用vagtiasygsvsqvmhnatlvpvkv替换。

[0093] 在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中所述多肽或其活性片段在SWT77-进化枝中。SWT77-进化枝可如实施例13中所述通过形成系统发育树(诸如通过使用邻接法)来测定。在一些实施方案中,任何SWT77-进化枝成员的距离值在SWT77序列的直系祖先节点范围内。

[0094] 在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中所述多肽或其活性片段在SWT22-进化枝中。SWT22-进化枝可如实施例13中所述通过形成系统发育树(诸如通过使用邻接法)来测定。在一些实施方案中,任何SWT22-进化枝成员的距离值在SWT22序列的直系祖先节点范围内。

[0095] 在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中所述多肽或其活性片段在WP026675114-进化枝中。在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中的任一种,其中所述多肽或其活性片段在WP026675114-进化枝中,前提条件是所述多肽不是WP026675114。WP026675114-进化枝可如实施例13中所述通过形成系统发育树(诸如通过使用邻接法)来测定。在一些实施方案中,任何WP026675114-进化枝成员的距离值在WP026675114序列的直系祖先节点范围内。

[0096] 在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明多肽中任一种,其中所述多肽或其活性片段在BspAG00296-进化枝中。BspAG00296-进化枝可如实施例13中所述通过形成系统发育树(诸如通过使用邻接法)来测定。在一些实施方案中,任何BspAG00296-进化枝成员的距离值在BspAG00296序列的直系祖先节点范围内。

[0097] 在一些实施方案中,本发明的多肽为对示例性多肽具有指定程度的氨基酸序列同源性的多肽,例如与选自SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、和SEQ ID NO:44的氨基酸序列具有70%、72%、74%、76%、78%、80%、85%、90%、95%、96%、97%、98%、99%或100%的序列同一性。在其它实施方案中,本发明的多肽为对示例性多肽具有指定程度的氨基酸序列同源性的多肽,例如与选自SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、和SEQ ID NO:43的氨

氨基酸序列具有70%、72%、74%、76%、78%、80%、85%、90%、95%、96%、97%、98%、99%或100%的序列同一性。同源性可通过氨基酸序列比对来测定,例如使用程序诸如BLAST、ALIGN或CLUSTAL,如本文所述。在一些实施方案中,多肽是分离的、重组的、基本上纯的、或非天然存在的酶,其具有蛋白酶活性(例如,二甲基酪蛋白水解活性)。

[0098] 还提供了具有蛋白酶活性,诸如碱性蛋白酶活性的本发明的多肽酶,所述酶包含氨基酸序列,当使用前述比对方法中任一种进行比对时,所述氨基酸序列的不多于50、不多于40、不多于30、不多于25、不多于20、不多于15、不多于10、不多于9、不多于8、不多于7、不多于6、不多于5、不多于4、不多于3、不多于2、或不多于1个氨基酸残基与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43或SEQ ID NO:44的氨基酸序列不同。另外,还提供了具有蛋白酶活性,诸如碱性蛋白酶活性的本发明的多肽酶,所述酶包含氨基酸序列,当使用前述比对方法中任一种进行比对时,所述氨基酸序列的不多于50、不多于40、不多于30、不多于25、不多于20、不多于15、不多于10、不多于9、不多于8、不多于7、不多于6、不多于5、不多于4、不多于3、不多于2、或不多于1个氨基酸残基与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、或SEQ ID NO:43的氨基酸序列不同。

[0099] 在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明的多肽中任一种,其中所述多肽或其活性片段包含氨基酸序列,所述氨基酸序列与选自SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、和SEQ ID NO:44的氨基酸序列具有至少70%的同一性,其中所述重组多肽或其活性片段包含选自下列的至少一个置换:X003N、X006R、X010E、X020I、X026N、X028R、X029I、X038A、X041P、X042N、X044R、X048D、X053R、X059G、X061G、X085Q、X088R、X090I、X096G、X098N、X103M、X104Y、X107Q、X113A、X115S、X117N、X131D、X132S、X133D、X136N、X137N、X138I、X139N、X143S、X144S、X146T、X147L、X157R、X168N、X169A、X178N、X179R、X180T、X204Y、X207G、X208Q、X209F、X210R、X212L、X219T、X222V、X229I、X230K、X231S、X231A、X239T、X240Q、X241V、X243N、X245L、X246R、X247D、X255L、X256N、X257Q、X264N、X266Y、X271A和X273G。在一些实施方案中,本发明的多肽或其活性片段可以为新型多肽或上述本发明的多肽中任一种,其中所述多肽或其活性片段包含氨基酸序列,所述氨基酸序列与选自SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、和SEQ ID NO:43的氨基酸序列具有至少70%的同一性,其中所述重组多肽或其活性片段包含选自下列的至少一个置换:X003N、X006R、X010E、X020I、X026N、X028R、X029I、X038A、X041P、X042N、X044R、X048D、X053R、X059G、X061G、X085Q、X088R、X090I、X096G、X098N、X103M、X104Y、X107Q、X113A、X115S、X117N、X131D、X132S、X133D、X136N、X137N、X138I、X139N、X143S、X144S、X146T、X147L、X157R、X168N、X169A、X178N、X179R、X180T、X204Y、X207G、X208Q、X209F、X210R、X212L、

X219T、X222V、X229I、X230K、X231S、X231A、X239T、X240Q、X241V、X243N、X245L、X246R、X247D、X255L、X256N、X257Q、X264N、X266Y、X271A和X273G。在一些实施方案中,所述置换选自:P003N、Q006R、N010E、T020I、S026N、I028R、Q029I、H038A、Q041P、S042N、A044R、N048D、Q053R、S059G、M061G、H085Q、T088R、V090I、N096G、S098N、L103M、F104Y、T107Q、S113A、D115S、G117N、N131D、Q132S、S133D、A136N、A137N、A138I、Q139N、N143S、A144S、S146T、I147L、A157R、S168N、V169A、T178N、G179R、A180T、V204Y、N207G、G208Q、Y209F、A210R、F212L、S219T、A222V、N229I、R230K、A231S、V231A、S239T、N240Q、A241V、S243N、M245L、Q246R、N247D、P255L、T256N、F257Q、D264N、N266Y、Q271A和S273G。

[0100] 在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少70%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少70%的同一性,前体条件是所述氨基酸序列不包含WP010283106、WP006679321、W02012175708-0002、W02012175708-0004、或W02012175708-0006。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少70%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、WP026675114、WP025025887、WP010283106或WP006679321。

[0101] 在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少75%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少75%的同一性,前体条件是所述氨基酸序列不包含WP010283106、W02012175708-0002、W02012175708-0004、或W02012175708-0006。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少75%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002、W02012175708-0004、W02012175708-0006、

WP026675114、WP025025887、WP010283106或WP006679321。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少75%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002、W02012175708-0004、WP026675114或WP010283106。

[0102] 在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少80%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少80%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002或W02012175708-0004。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少80%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002、W02012175708-0004、WP026675114或WP025025887。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少80%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002。

[0103] 在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少85%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少85%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0002或W02012175708-0004。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少85%的同一性,前体条件是所述氨基酸序列不包含W02012175708-

0002、W02012175708-0004、或WP026675114。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、EQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少85%的同一性。

[0104] 在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少90%的同一性。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少90%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0004。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少90%的同一性,前体条件是所述氨基酸序列不包含W02012175708-0004或WP026675114。在一些实施方案中,本发明是重组多肽或其活性片段,其包含氨基酸序列,所述氨基酸序列与SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、EQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列具有至少90%的同一性。

[0105] 在一些实施方案中,本发明是重组多肽或其活性片段,其包含SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、EQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43、或SEQ ID NO:44的氨基酸序列。

[0106] 如上所述,本发明的变体酶多肽具有酶活性(例如,蛋白酶活性)并且因此可用于清洁应用,包括但不限于用于清洁盘碟物品、餐具物品、织物和具有硬质表面的物品(例如,桌子、桌面、墙壁、家具物品、地面、天花板等的硬质表面)的方法。在下文对包含本发明的一种或多种变体丝氨酸蛋白酶多肽的示例性清洁组合物进行描述。本发明的酶多肽的酶活性(例如,蛋白酶活性)可使用本领域普通技术人员熟知的操作来容易地确定。下文所示的实施例对用于评估酶活性和清洁性能的方法进行了描述。本发明的多肽酶在去除污渍(例如,蛋白质污渍,诸如血液/乳/墨或卵黄)、清洁硬质表面或者清洁衣物、盘碟或餐具物品方面的性能可使用本领域中熟知的操作和/或通过使用实施例中的所述的操作来容易地确定。

[0107] 本发明的丝氨酸蛋白酶多肽可在广泛范围的pH条件下具有蛋白酶活性。在一些实施方案中,所述丝氨酸蛋白酶多肽具有对作为底物的二甲基酪蛋白的蛋白酶活性,如实施例7中所展示的。在一些实施方案中,所述丝氨酸蛋白酶多肽在约4.0至约12.0的pH下具有蛋白酶活性。在一些实施方案中,所述丝氨酸蛋白酶多肽在约6.0至约12.0的pH下具有蛋白

酶活性。在一些实施方案中,所述丝氨酸蛋白酶多肽在约6.0至约12.0,或约7.0至约12.0的pH下具有至少50%、60%、70%、80%或90%的最大蛋白酶活性。在一些实施方案中,所述丝氨酸蛋白酶多肽在高于约6.0、6.5、7.0、7.5、8.0、8.5、9.0、9.5、10.0、10.5、11.0或11.5的pH下具有蛋白酶活性。在一些实施方案中,所述丝氨酸蛋白酶多肽在低于约12.0、11.5、11.0、10.5、10.0、9.5、9.0、8.5、8.0、7.5、7.0或6.5的pH下具有蛋白酶活性。

[0108] 在一些实施方案中,本发明的丝氨酸蛋白酶多肽在约10℃至90℃,或约30℃至约80℃的温度范围下具有蛋白酶活性。在一些实施方案中,本发明的丝氨酸蛋白酶多肽在约55℃至约75℃的温度范围下具有蛋白酶活性。在一些实施方案中,丝氨酸蛋白酶多肽在约55℃至约75℃的温度下具有至少50%、60%、70%、80%或90%的最大蛋白酶活性。在一些实施方案中,丝氨酸蛋白酶在高于50℃、55℃、60℃、65℃或70℃的温度下具有活性。在一些实施方案中,丝氨酸蛋白酶在低于75℃、80℃、70℃、65℃、60℃或55℃的温度下具有活性。

[0109] 在一些实施方案中,本发明的丝氨酸蛋白酶多肽在胁迫条件下于50℃下20分钟之后具有至少80%活性。胁迫条件可例如为实施例11中所示的那些。在一些实施方案中,胁迫条件是在LAS/EDTA测定、Tris/EDTA测定、或OMO HDL测定中。

[0110] 在一些实施方案中,本发明的丝氨酸蛋白酶多肽证明在清洁组合物中具有清洁性能。清洁组合物通常包含对酶的稳定性和性能有害的成分,导致清洁组合物成为酶(例如丝氨酸蛋白酶)保留功能的苛刻环境。因此,将酶放入清洁组合物中并且期望酶功能(例如丝氨酸蛋白酶活性,诸如通过清洁性能证明的)并非微不足道的。在一些实施方案中,本发明的丝氨酸蛋白酶多肽证明在自动盘碟洗涤(ADW)洗涤剂组合物中具有清洁性能。在一些实施方案中,在自动盘碟洗涤(ADW)洗涤剂组合物中的清洁性能包括清洁卵黄污渍。在一些实施方案中,本发明的丝氨酸蛋白酶多肽证明在衣物洗涤剂组合物中具有清洁性能。在一些实施方案中,在衣物洗涤剂组合物中的清洁性能包括清洁血液/乳/墨污渍。在每种清洁组合物中,本发明的丝氨酸蛋白酶多肽均证明在具有或不具有漂白剂组分下具有清洁性能。

[0111] 可对本发明的多肽进行多种改变,诸如一个或多个保守或非保守的氨基酸插入、缺失和/或置换,包括这样的变化基本不改变该多肽的酶活性的情况。类似地,还可对本发明的核酸进行多种改变,诸如一个或多个密码子中一个或多个核苷酸的一种或多种置换,使得特定的密码子编码相同或不同的氨基酸,从而产生沉默变化(例如,当所编码的氨基酸不因核苷酸突变而改变时)或非沉默变化;对该序列中一个或多个核酸(或密码子)的一种或多种缺失;该序列中一个或多个核酸(或密码子)的一种或多种添加或插入,和/或该序列中一个或多个核酸(或密码子)的切割或者一种或多种截短。与起始核酸序列所编码的多肽酶相比,核酸序列中的很多此类变化可基本不改变得到的编码多肽酶的酶活性。还可对本发明的核酸序列进行修饰以包含允许在表达系统(例如,细菌表达系统)中最佳表达的一个或多个密码子,同时如果期望的话所述一个或多个密码仍编码相同的氨基酸。

[0112] 在一些实施方案中,本发明提供具有期望的酶活性(例如,蛋白酶活性或清洁性能活性)的酶多肽的属,其包含具有本文所述的氨基酸置换的序列并且还包含一个或多个附加的氨基酸置换,诸如保守和非保守置换,其中所述多肽表现出、保持或近似保持期望的酶活性(例如,蛋白水解活性,如体现在SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:7、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:22、SEQ ID NO:25、SEQ ID NO:28、SEQ ID NO:31、SEQ ID NO:34、SEQ ID NO:37、SEQ ID NO:40、SEQ ID NO:43或SEQ

ID NO:44的多肽酶的清洁活性或性能中的)。根据本发明的氨基酸置换可包括但不限于一种或多种非保守置换和/或一种或多种保守氨基酸置换。保守氨基酸残基置换通常涉及将一个功能类的氨基酸残基中的成员替换成属于相同功能类的残基(认为保守的氨基端残基在功能上是同源的或者在计算功能同源性百分比时是保守的)。保守氨基酸置换通常涉及用功能上类似的氨基酸来置换氨基酸序列中的氨基酸。例如,丙氨酸、甘氨酸、丝氨酸和苏氨酸在功能上类似,并且因此可充当彼此的保守氨基酸置换。天冬氨酸和谷氨酸可充当彼此的保守置换。天冬酰胺和谷氨酰胺可充当彼此的保守置换。精氨酸、赖氨酸和组氨酸可充当彼此的保守置换。异亮氨酸、亮氨酸、甲硫氨酸和缬氨酸可充当彼此的保守置换。苯丙氨酸、酪氨酸和色氨酸可充当彼此的保守置换。

[0113] 还可设想其它的保守氨基酸置换组。例如,可通过类似的功能或化学结构或组成来对氨基酸进行分组(例如,酸性、碱性、脂族、芳族、含硫)。例如,脂族分组可包括:甘氨酸(G)、丙氨酸(A)、缬氨酸(V)、亮氨酸(L)、异亮氨酸(I)。包含认为是彼此之保守置换的氨基酸的其它组包括:芳族:苯丙氨酸(F)、酪氨酸(Y)、色氨酸(W);含硫:甲硫氨酸(M)、半胱氨酸(C);碱性:精氨酸(R)、赖氨酸(K)、组氨酸(H);酸性:天冬氨酸(D)、谷氨酸(E);非极性的不带电荷残基:半胱氨酸(C)、甲硫氨酸(M)和脯氨酸(P);亲水性的不带电荷残基:丝氨酸(S)、苏氨酸(T)、天冬酰胺(N)和谷氨酰胺(Q)。另外的氨基酸分组是本领域技术人员熟知的并且在多种标准教科书中进行了描述。本文多肽序列结合上述置换组的列表提供了所有经保守置换的多肽序列的明确列表。

[0114] 上述氨基酸残基类别中存在较为保守的置换,其也或者作为替代可以是合适的。较为保守的保守置换组包括:缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸和天冬酰胺-谷氨酰胺。

[0115] 本发明多肽序列的保守置换的变型(例如,本发明的变体丝氨酸蛋白酶)包括用同一保守置换组中保守选择的氨基酸对该多肽序列的氨基酸进行小百分比,有时小于5%、4%、3%、2%或1%的置换,或者少于10、9、8、7、6、5、4、3、2或1个氨基酸置换。

[0116] III. 编码丝氨酸蛋白酶的核酸

[0117] 本发明提供分离的、非天然存在的或重组的核酸(本文中也称为“多核苷酸”),其可被统称为“本发明的核酸”或“本发明的多核苷酸”,所述核酸编码本发明的多肽。本发明的核酸(包括下文所述的全部核酸)可用于重组产生(例如,表达)本发明的多肽,其通常通过表达包含编码目的多肽或其片段的序列的质粒表达载体来进行。如上文所讨论的,多肽包括具有酶活性(蛋白水解活性)的丝氨酸蛋白酶多肽,其可用于清洁应用和用于清洁需要清洁的物品或表面(例如,物品的表面)的清洁组合物。

[0118] 在一些实施方案中,本发明的多核苷酸是对示例性多肽具有指定程度的核酸同源性的多核苷酸。在一些实施方案中,多核苷酸包含选自下列的核酸序列:SEQ ID NO:1、SEQ ID NO:5、SEQ ID NO:8和SEQ ID NO:12。在其它实施方案中,本发明的多核苷酸还可具有与选自SEQ ID NO:1、SEQ ID NO:5、SEQ ID NO:8和SEQ ID NO:12的核酸序列互补的核酸序列。同源性可通过氨基酸序列比对来测定,例如使用程序诸如BLAST、ALIGN或CLUSTAL,如本文所述。

[0119] 在一些实施方案中,本发明提供了分离的、重组的、基本纯的或非天然存在的核酸,其包含编码本发明的任意多肽(包括任意融合蛋白等)的核苷酸序列,所述多肽在上文

标题为“本发明的多肽”的部分或本文的其它部分中进行了描述。本发明还提供了分离的、重组的、基本纯的或非天然存在的核酸，其包含编码上文以及本文中其它部分所述的本发明的两种或更多种的任意多肽的組合的核苷酸序列。本发明提供了编码本发明的丝氨酸蛋白酶多肽的核酸，其中该丝氨酸蛋白酶多肽是具有蛋白水解活性的成熟形式。在一些实施方案中，丝氨酸蛋白酶(例如，BspAG00296)与同源前肽序列(例如，BspAG00296前肽)重组表达。在其它实施方案中，丝氨酸蛋白酶与异源前肽序列(例如，来自另一种枯草杆菌蛋白酶的前肽序列)重组表达。

[0120] 本发明的核酸可通过使用任何合适的合成、操纵和/或分离技术或者它们的组合来产生。例如，可使用标准的核酸合成技术，诸如本领域技术人员熟知的固相合成技术来产生本发明的多核苷酸。在此类技术中，通常合成至多50或更多个核苷酸碱基的片段，然后(例如，通过酶连接或化学连接方法，或聚合酶介导的重组方法)接合以形成基本上任何期望的连续核酸序列。本发明的核酸的合成还可通过本领域已知的任何合适的方法来促进，包括但不限于使用经典的亚磷酰胺方法化学合成(参见例如，Beaucage等人，Tetrahedron Letters 22:1859-69[1981])；或由Matthes等人描述的方法(参见，Matthes等人，EMBO J. 3:801-805[1984]，如通常在自动合成方法中实施。还可通过使用自动化DNA合成仪来产生本发明的核酸。定制核酸可从多种商业来源(例如，The Midland Certified Reagent Company、Great American Gene Company、Operon Technologies Inc. 和DNA2.0)订购。用于合成核酸的其它技术和相关的原理是本领域已知的(参见例如，Itakura等人，Ann.Rev.Biochem. 53:323[1984]；和Itakura等人，Science 198:1056[1984])。

[0121] 如上所述，可用于修饰核酸的重组DNA技术在本领域中是熟知的。例如，本领域技术人员知晓并且可容易地采用诸如限制性内切核酸酶消化、连接、逆转录和cDNA产生以及聚合酶链反应(例如，PCR)的技术。还可通过使用一种或多种寡核苷酸探针筛选cDNA文库来获得本发明的核苷酸，所述探针可与编码本发明的丝氨酸蛋白酶多肽的多核苷酸杂交或者对其进行PCR扩增。用于筛选并分离cDNA克隆的过程以及PCR扩增过程是本领域技术人员熟知的，并且在本领域技术人员已知的标准参考文献中进行了描述。本发明的一些核酸可借助例如已知的诱变操作(例如，定点诱变、位点饱和诱变和体外重组)通过改变天然存在的多核苷酸主链(例如，编码酶或亲本蛋白酶的天然存在多核苷酸主链)来获得。适于产生编码本发明丝氨酸蛋白酶多肽的本发明的经修饰多核苷酸的多种方法在本领域中是已知的，包括但不限于例如位点饱和诱变、扫描诱变、插入诱变、缺失诱变、随机诱变、定点诱变和定向进化以及多种其它重组方法。

[0122] IV. 用于产生丝氨酸蛋白酶的载体、宿主细胞和方法

[0123] 本发明提供了分离的或重组的载体，其包含至少一种本文所述的本发明的丝氨酸蛋白酶多核苷酸(例如，编码本文所述的本发明的丝氨酸蛋白酶多肽的多核苷酸)，包含至少一种本发明的核酸或多核苷酸的分离的或重组的表达载体或表达盒，包含至少一种本发明的核酸或多核苷酸的分离的、基本上纯的、或重组DNA构建体，包含至少一种本发明的多核苷酸的分离的或重组的细胞，包含含有至少一种本发明多核苷酸的细胞的细胞培养物，包含至少一种本发明的核苷酸或多核苷酸的细胞培养物，以及包含一种或多种此类载体、核酸、表达载体、表达盒、DNA构建体、细胞、细胞培养物或它们的任何组合或混合物的组合物。

[0124] 在一些实施方案中,本发明提供了包含本发明的至少一种载体(例如,表达载体或DNA构建体)的重组细胞,所述载体包含本发明的至少一种核酸或多核苷酸。一些此类重组细胞用此类至少一种载体转化或转染。通常将这样的细胞称为宿主细胞。一些这样的细胞包含细菌细胞,包括但不限于芽孢杆菌属(*Bacillus* sp.)细胞,诸如枯草芽孢杆菌细胞。本发明还提供了包含本发明的至少一种丝氨酸蛋白酶多肽的重组细胞(例如,重组宿主细胞)。

[0125] 在一些实施方案中,本发明提供了包含本发明的核酸或多核苷酸的载体。在一些实施方案中,所述载体是表达载体或表达盒,其中编码本发明的丝氨酸蛋白酶多肽的本发明的多核苷酸序列与有效基因表达所需的一个或另外的核酸区段可操作地连接(例如,启动子与编码本发明的丝氨酸蛋白酶多肽的本发明的多核苷酸可操作地连接)。载体可包含转录终止子和/或选择基因,诸如抗生素抗性基因,其能够通过含抗微生物剂培养基中生长来实现经质粒感染的宿主细胞的连续培养维持。

[0126] 表达载体可衍生自质粒或病毒DNA,或者在一些另选的实施方案中,包含两者的元件。示例性载体包括但不限于pC194、pJH101、pE194、pHP13(参见,Harwood and Cutting[编辑],第3章,Molecular Biological Methods for *Bacillus*,John Wiley&Sons[1990];枯草芽孢杆菌的合适的复制质粒包括第92页中列出的那些),还可参见,Perego, Integrational Vectors for Genetic Manipulations in *B.subtilis*,Sonenshein等人[编辑]*B.subtilis* and Other Gram-Positive Bacteria:Biochemistry,Physiology and Molecular Genetics,American Society for Microbiology,Washington,D.C.[1993],第615-624页);以及p2JM103BBI。

[0127] 对于在细胞中表达和产生目的蛋白(例如,丝氨酸蛋白酶多肽),在适于表达丝氨酸蛋白酶的条件下,将包含至少一个拷贝的编码丝氨酸蛋白酶多肽的多核苷酸(在一些情况下,包含多个拷贝)的至少一种表达载体转化到细胞中。在本发明的一些实施方案中,编码丝氨酸蛋白酶多肽的多核苷酸序列(以及载体中包括的其它序列)整合到宿主细胞的基因组中,而在其它实施例中,包含编码丝氨酸蛋白酶多肽的多核苷酸序列的质粒载体在细胞中保持为自主染色体外元件。本发明提供了染色体外核酸元件以及整合到宿主细胞基因组中的引入核苷酸序列二者。本文中所述的载体可用于产生本发明的丝氨酸蛋白酶多肽。在一些实施方案中,编码丝氨酸蛋白酶多肽的多核苷酸构建体存在于能够将编码丝氨酸蛋白酶多肽的多核苷酸整合到宿主染色体中并任选地对其进行扩增的整合型载体上。整合位点的示例是本领域技术人员熟知的。在一些实施方案中,编码本发明丝氨酸蛋白酶多肽的多核苷酸的转录通过为所选前体蛋白酶的野生型启动子的启动子来实现。在一些其它实施方案中,启动子与前体蛋白酶异源,但是在宿主细胞中具有功能。具体地讲,适用于细菌宿主细胞的启动子的示例包括但不限于,例如,amyE、amyQ、amyL、pstS、sacB、SPAC、AprE、Veg、HpaII启动子,嗜热脂肪芽孢杆菌麦芽糖淀粉酶基因、解淀粉芽孢杆菌(BAN)淀粉酶基因、枯草芽孢杆菌碱性蛋白酶基因、克劳氏芽孢杆菌碱性蛋白酶基因、短小杆菌木糖苷酶基因、苏云金芽孢杆菌cryIIIA、和地衣芽孢杆菌 α -淀粉酶基因的启动子。另外的启动子包括但不限于A4启动子以及噬菌体 λ PR或PL启动子,和大肠杆菌lac、trp或tac启动子。

[0128] 本发明的丝氨酸蛋白酶多肽可在任何合适的微生物(包括细菌和真菌)的宿主细胞中产生。例如,在一些实施方案中,本发明的丝氨酸蛋白酶多肽可在革兰氏阳性菌中产

生。在一些实施方案中,宿主细胞是芽孢杆菌属物种(*Bacillus* spp.)、链霉菌属物种(*Streptomyces* spp.)、埃希氏菌属物种(*Escherichia* spp.)、曲霉属物种(*Aspergillus* spp.)、木霉属物种(*Trichoderma* spp.)、假单胞菌属物种(*Pseudomonas* spp.)、棒状杆菌属物种(*Corynebacterium* spp.)、酵母属物种(*Saccharomyces* spp.)或毕赤酵母属物种(*Pichia* spp.)。在一些实施方案中,通过芽孢杆菌属(*Bacillus* sp.)宿主细胞来产生丝氨酸蛋白酶多肽。可用于制备本发明的丝氨酸蛋白酶多肽的芽孢杆菌属宿主细胞的示例包括但不限于地衣芽孢杆菌、迟缓芽孢杆菌、枯草芽孢杆菌、解淀粉芽孢杆菌、迟缓芽孢杆菌、索诺拉沙漠芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、嗜碱芽孢杆菌、凝结芽孢杆菌、环状芽孢杆菌、短小芽孢杆菌、苏云金芽孢杆菌、克劳氏芽孢杆菌、和巨大芽孢杆菌、以及在芽孢杆菌属范围内的其它生物体。在一些实施方案中,使用枯草芽孢杆菌宿主细胞来产生丝氨酸蛋白酶多肽。美国专利5,264,366和4,760,025(RE 34,606)描述了可用于产生本发明的丝氨酸蛋白酶多肽的多种芽孢杆菌宿主菌株,但是也可使用其它合适的菌株。

[0129] 可用于产生本发明的丝氨酸蛋白酶多肽的多种工业细菌菌株包括非重组(即野生型)芽孢杆菌属(*Bacillus* sp.)菌株,以及天然存在的菌株和/或重组菌株的变体。在一些实施方案中,宿主菌株是其中已将编码目的多肽的多核苷酸导入宿主中的重组菌株。在一些实施方案中,宿主菌株是枯草芽孢杆菌宿主菌株并且具体地讲重组枯草芽孢杆菌宿主菌株。多种枯草芽孢杆菌菌株是已知的,包括但不限于例如1A6(ATCC 39085)、168(1A01)、SB19、W23、Ts85、B637、PB1753至PB1758、PB3360、JH642、1A243(ATCC 39,087)、ATCC 21332、ATCC 6051、MI113、DE100(ATCC 39,094)、GX4931、PBT 110和PEP 211菌株(参见,例如Hoch等人,Genetics 73:215-228[1973];还参见,美国专利4,450,235和4,302,544,以及EP 0134048,其全文以引用方式并入本文)。枯草芽孢杆菌作为表达宿主细胞的用途在本领域中是熟知的(参见,例如Palva等人,Gene 19:81-87[1982];Fahnestock和Fischer, J.Bacteriol.,165:796-804[1986];和Wang等人,Gene 69:39-47[1988])。

[0130] 在一些实施方案中,芽孢杆菌宿主细胞是在至少一个以下基因中包含突变或缺失的芽孢杆菌属(*Bacillus* sp.):degU、degS、degR和degQ。在一些实施方案中,突变在degU基因中,并且在一些实施方案中,突变是degU(Hy)32(参见,例如Msadek等人, J.Bacteriol.172:824-834[1990];和Olmos等人,Mol.Gen.Genet.253:562-567[1997])。在一些实施方案中,芽孢杆菌宿主包含scoC4(参见例如,Caldwell等人, J.Bacteriol.183:7329-7340(2001);spoIIE(参见例如Arigoni等人,Mol.Microbiol.31:1407-1415[1999]);和/或oppA或opp操纵子的其它基因(参见例如,Perego等人,Mol.Microbiol.5:173-185[1991])中的突变或缺失。实际上,预期与oppA基因中的突变产生相同表型的opp操纵子中的任何突变均可用于本发明的经改变的芽孢杆菌菌株的一些实施方案。在一些实施方案中,这些突变单独发生,而在另一些实施方案中,存在突变的组合。在一些实施方案中,可用于产生本发明的丝氨酸蛋白酶多肽的经改变的芽孢杆菌宿主细胞菌株是已在上述基因中的一个或多个中包含突变的芽孢杆菌宿主菌株。另外,可使用包含内源性蛋白酶基因的突变和/或缺失的芽孢杆菌属宿主细胞。在一些实施方案中,芽孢杆菌宿主细胞包含aprE和nprE基因的缺失。在其它实施方案中,芽孢杆菌属宿主细胞包含5个蛋白酶基因的缺失,然而在其它实施方案中,芽孢杆菌属宿主细胞包含9个蛋白酶基因的缺失(参见例如,美国专利申请公布号2005/0202535,其以引用方式并入本文)。

[0131] 使用本领域中已知的任何合适方法用编码本发明的至少一种丝氨酸蛋白酶多肽的至少一种核酸来转化宿主细胞。利用质粒DNA构建体或载体来将核酸(例如,DNA)引入芽孢杆菌细胞或大肠杆菌细胞中以及将这样的质粒DNA构建体或载体转化到此类细胞中的方法是熟知的。在一些实施方案中,随后从大肠杆菌细胞分离质粒,并将其转化到芽孢杆菌细胞中。然而,并非必须使用诸如大肠杆菌的中间微生物,并且在一些实施方案中,将DNA构建体或载体直接引入芽孢杆菌属宿主中。

[0132] 本领域技术人员充分意识到适用于将本发明的核酸或多核苷酸序列引入芽孢杆菌细胞中的方法(参见例如,Ferrari等人,“Genetics”,Harwood等人[编辑],Bacillus, Plenum Publishing Corp.[1989],第57-72页;Saunders等人,J.Bacteriol.157:718-726 [1984];Hoch等人,J.Bacteriol.93:1925-1937[1967];Mann等人,Current Microbiol.13: 131-135[1986];Holubova,Folia Microbiol.30:97[1985];Chang等人, Mol.Gen.Genet.168:11-115[1979];Vorobjeva等人,FEMS Microbiol.Lett.7:261-263 [1980];Smith等人,Appl.Env.Microbiol.51:634[1986];Fisher等人, Arch.Microbiol.139:213-217[1981];以及McDonald,J.Gen.Microbiol.130:203[1984])。实际上,包括原生质体转化和集合、转导、和原生质体融合在内的此类转化方法为人们所熟知并且适用于本发明。转化的方法用于将包含编码本发明丝氨酸蛋白酶多肽的核酸的DNA构建体或载体引入宿主细胞中。转化芽孢杆菌细胞的本领域已知方法包括这样的方法,例如质粒标记拯救转化,其涉及通过携带部分同源的内

[0133] 生质粒(resident plasmid)的感受态细胞摄入供体质粒(参见,Contente等人, Plasmid 2:555-571[1979];Haima等人,Mol.Gen.Genet.223:185-191[1990];Weinrauch等人,J.Bacteriol.154:1077-1087[1983];和Weinrauch等人,J.Bacteriol.169:1205-1211 [1987])。在该方法中,引入的供体质粒与内生“辅助性”质粒的同源区在模拟染色体转化的过程中发生重组。

[0134] 除常用的方法之外,在一些实施方案中,用包含编码本发明丝氨酸蛋白酶多肽的核酸的DNA构建体或载体直接转化宿主细胞(即,在引入宿主细胞中之前,不使用中间细胞来扩增或者以别的方式处理该DNA构建体或载体)。将本发明的DNA构建体或载体引入宿主细胞中包括本领域已知的将核酸序列(例如DNA序列)引入宿主细胞但不插入质粒或载体的那些物理和化学方法。这样的方法包括但不限于氯化钙沉淀、电穿孔、裸DNA、脂质体等。在另外的实施方案中,将DNA构建体或载体与质粒共转化,而不插入质粒中。在另一个实施方案中,通过本领域中已知的方法从经改变的芽孢杆菌菌株缺失选择性标记(参见,Stahl等人,J.Bacteriol.158:411-418[1984];和Palmeros等人,Gene 247:255-264[2000])。

[0135] 在一些实施方案中,将本发明的转化细胞在常规的营养培养基中进行培养。合适的培养条件(诸如温度、pH等)是本领域技术人员已知的,并且在科学文献中进行了充分描述。在一些实施方案中,本发明提供了包含本发明的至少一种丝氨酸蛋白酶多肽或至少一种核酸的培养物(例如,细胞培养物)。还提供组合物,所述组合物包含至少一种本发明的核酸、载体或DNA构建体。

[0136] 在一些实施方案中,在允许表达本发明蛋白酶的条件下在合适的营养培养基中培养经编码本发明的至少一种丝氨酸蛋白酶多肽的至少一种多核苷酸序列转化的宿主细胞,在此之后从培养物中回收所得的蛋白酶。用于培养细胞的培养基包括适用于使宿主细胞生

长的任何常规培养基,诸如基本培养基或包含适当的补充剂的复合培养基。合适的培养基可得自商业供应商或可根据公布的配方(例如,美国典型培养物保藏中心的目录)制备。在一些实施方案中,通过常规操作来从培养基中回收细胞产生的蛋白酶,包括但不限于例如,通过离心或过滤从培养基中分离宿主细胞、借助于盐(硫酸铵)沉淀出上清液或滤液的蛋白质性组分、色谱纯化(例如,离子交换、凝胶过滤、亲和等)。适用于回收或纯化变体蛋白酶的任何方法可用于本发明。

[0137] 在一些实施方案中,通过重组宿主细胞产生的丝氨酸蛋白酶多肽分泌到培养基中。编码纯化促进结构域的核酸序列可用于促进蛋白质的纯化。包含编码丝氨酸蛋白酶多肽的多核苷酸序列的载体或DNA构建体可还

[0138] 包含编码促进该丝氨酸蛋白酶多肽纯化的纯化促进结构域的核酸序列(参见,例如Kroll等人,DNA Cell Biol.12:441-53[1993])。这样的纯化促进结构域包括但不限于例如允许在固定化金属上进行纯化的金属螯合肽,诸如组氨酸-色氨酸模块(参见,Porath, Protein Expr.Purif.3:263-281[1992]);允许在固定化免疫球蛋白上进行纯化的蛋白A结构域;和用于LAGS延伸/亲和纯化系统的结构域。还可利用在纯化结构域和异源蛋白质之间包括可切割的接头序列诸如因子XA或肠激酶(例如,可获自Invitrogen(San Diego,CA)的序列)来促进纯化。

[0139] 用于检测和测量酶(例如,本发明的丝氨酸蛋白酶多肽)的酶活性的测定是熟知的。用于检测和测量蛋白酶(例如,本发明的丝氨酸蛋白酶多肽)的活性的多种测定也是本领域普通技术人员已知的。具体地讲,测定适用于测量蛋白酶活性,所述蛋白酶活性是基于酸可溶性肽从酪蛋白或血红蛋白的释放,其作为在280nm下的吸光度进行测量或使用Folin法比色测量,所述方法是本领域技术人员所熟知的。一种示例性的测定涉及溶解生色底物(参见,例如Ward, "Proteinases," Fogarty(编辑), Microbial Enzymes and Biotechnology, Applied Science, London, [1983], 第251-317页)。其它的示例性测定包括但不限于琥珀酰-Ala-Ala-Pro-Phe-对硝基苯胺测定(suc-AAPF-pNA)和2,4,6-三硝基苯磺酸钠盐测定(TNBS测定)。本领域人员已知的很多其它参考文献提供了合适的方法(参见,例如Wells等人, Nucleic Acids Res.11:7911-7925[1983]; Christianson等人, Anal.Biochem.223:119-129[1994]; 和Hsia等人, Anal Biochem.242:221-227[1999])。

[0140] 多种方法可用于确定成熟蛋白酶(例如,本发明的成熟丝氨酸蛋白酶多肽)在宿主细胞中的产生水平。这样的方法包括但不限于例如利用对该蛋白酶具有特异性的多克隆或单克隆抗体的方法。示例性的方法包括但不限于酶联免疫吸附测定(ELISA)、放射免疫测定(RIA)、荧光免疫测定(FIA)和荧光激活细胞分选(FACS)。这些及其它测定在本领域中是熟知的(参见例如, Maddox等人, J.Exp.Med.158:1211[1983])。

[0141] 在一些其它实施方案中,本发明提供了用于制备或产生本发明的成熟丝氨酸蛋白酶多肽的方法。成熟丝氨酸蛋白酶多肽不包含信号肽或前肽序列。一些方法包括在重组细菌宿主细胞诸如例如芽孢杆菌属细胞(例如枯草芽孢杆菌细胞)中制备或产生本发明的丝氨酸蛋白酶多肽。在一些实施方案中,本发明提供了产生本发明的丝氨酸蛋白酶多肽的方法,该方法包括在有助于产生该丝氨酸蛋白酶多肽的条件下培养包含重组表达载体的重组宿主细胞,所述重组表达载体包含编码本发明的丝氨酸蛋白酶多肽的核酸。一些这样的方法还包括从培养物中回收丝氨酸蛋白酶多肽。

[0142] 在一些实施方案中,本发明提供了产生本发明的丝氨酸蛋白酶多肽的方法,所述方法包括:(a)将包含编码本发明的丝氨酸蛋白酶多肽的核酸的重组表达载体引入细胞(例如细菌细胞,诸如枯草芽孢杆菌细胞)群中;以及(b)在有助于产生由该表达载体编码的丝氨酸蛋白酶多肽的条件下在培养基中培养细胞。一些这样的方法还包括:(c)从细胞或者从培养基中分离丝氨酸蛋白酶多肽。

[0143] V. 包含丝氨酸蛋白酶的组合物

[0144] A. 织物和家庭护理产品

[0145] 除非另外指出,否则本文中提供的所有组分或组合物水平均是指该组分或组合物的活性水平,并且不包括可商购获得来源中可能存在的杂质,例如残余溶剂或副产物。酶组分重量基于总活性蛋白质。除非另外指出,否则所有的百分比和比率均为按重量计。除非另外指出,否则所有的百分比和比率均基于总组合物来计算。本发明的组合物包括清洁组合物,诸如洗涤剂组合物。在所例示的洗涤剂组合物中,酶水平通过按总组合物的重量计的纯酶来表示,并且除非另外规定,否则洗涤剂成分表示为按总组合物的重量计。

[0146] 尽管不是本发明的目的所必需的,但是下文所举例说明的助剂的非限制性列表适用于本发明的清洁组合物。在一些实施方案中,引入这些助剂例如是为了有助于或增强清洁性能(对于处理需清洁的基质),或者为了修饰清洁组合物的美观,如在香料、着色剂、染料等的情况下。应当理解,这样的助剂在本发明的丝氨酸蛋白酶多肽之外。这些附加组分的准确性质及其引入水平将取决于组合物的物理形式以及其所用于的清洁操作的性质。合适的助剂材料包括但不限于漂白催化剂、其它酶、酶稳定体系、螯合剂、光学增白剂、去垢性聚合物、染料转移剂、分散剂、抑泡剂、染料、香料、着色剂、填料盐、光活化剂、荧光剂、织物调理剂、可水解表面活性剂、防腐剂、抗氧化剂、抗收缩剂、抗皱剂、杀菌剂、杀真菌剂、色粒、银护理剂、抗晦暗剂和/或抗腐蚀剂、碱度源、增溶剂、载体、加工助剂、颜料和pH控制剂、表面活性剂、助洗剂、螯合剂、染料转移抑制剂、沉积助剂、分散剂、另外的酶和酶稳定剂、催化材料、漂白活性剂、漂白增效剂、过氧化氢、过氧化氢源、预成形的过酸、聚合物分散剂、粘土污垢去除/抗再沉积剂、增白剂、抑泡剂、染料、香料、结构增弹剂、织物软化剂、载体、水溶助长剂、加工助剂和/或颜料。除下文的公开内容之外,其它这样的助剂的合适示例以及使用水平见于美国专利5,576,282、6,306,812、6,326,348、6,610,642、6,605,458、5,705,464、5,710,115、5,698,504、5,695,679、5,686,014和5,646,101,其全部通过引用并入本文。在其中清洁助剂材料与本发明的丝氨酸蛋白酶多肽在清洁组合物中不相容的实施方案中,则可使用使清洁助剂材料和蛋白酶保持分离(即,彼此不接触)直至适宜组合这两种组分的合适方法。这样的分离方法包括本领域中已知的任何合适方法(例如胶囊锭(gelcap)、包封、片剂、物理分离等)。前述助剂成分可构成本发明的清洁组合物的其余部分。

[0147] 本发明的清洁组合物有利地用于例如衣物洗涤应用、硬质表面清洁应用、盘碟洗涤应用(包括自动盘碟洗涤和手洗盘碟洗涤),以及化妆品应用,诸如义齿、牙齿、毛发和皮肤清洁。本发明的酶还适用于接触镜片清洁和伤口清创应用。另外,由于在温度较低的溶液中效能增加的独特优势,本发明的酶理想地适于衣物洗涤应用。此外,本发明的酶可用于颗粒组合物和液体组合物。

[0148] 本发明的丝氨酸蛋白酶多肽还可用于清洁添加剂产品。在一些实施方案中,可用

于低温溶液清洁应用。在一些实施方案中,本发明提供了包含本发明的至少一种酶的清洁添加剂产品,当期望附加的漂白效果时,其理想地适于包括在洗涤过程中。这样的实例包括但不限于低温溶液清洁应用。在一些实施方案中,添加剂产品为其最简单的形式,即一种或多种蛋白酶。在一些实施方案中,将添加剂包装成用于添加至清洁过程的剂型。在一些实施方案中,将添加剂包装成用于添加至采用过氧源并且期望漂白效果增加的清洁过程的剂型。任何合适的单剂量单位形式均可用于本发明,包括但不限于丸剂、片剂、胶囊锭剂或其它单剂量单位,诸如预先测量的散剂或液体。在一些实施方案中,包含填料或载体材料以增加这样的组合物的体积。合适的填料或载体材料包括但不限于多种硫酸盐、碳酸盐和硅酸盐以及滑石、粘土等。用于液体组合物的合适填料或载体材料包括但不限于水或低分子量伯醇和仲醇,包括多元醇和二醇。这些醇的示例包括但不限于甲醇、乙醇、丙醇和异丙醇。在一些实施方案中,组合物包含约5%至约90%的此类材料。酸性填料可用于降低pH。或者,在一些实施方案中,清洁添加剂包含助剂成分,如下文更充分描述的。

[0149] 本发明的清洁组合物和清洁添加剂需要有效量的单独或者与其它蛋白酶和/或另外的酶组合的本文中提供的丝氨酸蛋白酶多肽中的至少一种。所需的酶水平通过添加本发明的一种或多种丝氨酸蛋白酶多肽来实现。通常来说,本发明的清洁组合物包含至少约0.0001重量百分比、约0.0001重量百分比至约10重量百分比、约0.001重量百分比至约1重量百分比或者约0.01重量百分比至约0.1重量百分比的本发明的丝氨酸蛋白酶多肽中的至少一种。

[0150] 通常将本文中的清洁组合物配制成使得在用于水性清洁操作期间洗涤水将具有约4.0至约11.5、或者甚至约5.0至约11.5、或者甚至约5.0至约8.0、或者甚至约7.5至约10.5的pH。通常将液体产品制剂配制成具有约3.0至约9.0或者甚至约3至约5的pH。通常将颗粒洗涤剂产品配制成具有约9至约11的pH。在一些实施方案中,可将本发明的清洁组合物配制成在洗涤条件下具有碱性pH,诸如约8.0至约12.0、或者约8.5至约11.0、或者约9.0至约11.0的pH。在一些实施方案中,可将本发明的清洁组合物配制成在洗涤条件下具有中性pH,诸如约5.0至约8.0、或者约5.5至约8.0、或者约6.0至约8.0、或者约6.0至约7.5的pH。在一些实施方案中,当在20℃下将清洁组合物以1:100(重量:重量)溶于去离子水中时测量中性pH条件,其使用常规pH计来测量。用于将pH控制在建议使用水平的技术包括使用缓冲液、碱、酸等,并且是本领域技术人员熟知的。

[0151] 在一些实施方案中,当在颗粒组合物或液体中使用丝氨酸蛋白酶多肽时,期望该丝氨酸蛋白酶多肽为经包封颗粒的形式以保护该丝氨酸蛋白酶多肽在储存期间免受颗粒组合物的其它组分的影响。另外,包封也是控制丝氨酸蛋白酶多肽在清洁过程期间的可用性的一种方式。在一些实施方案中,包封增强丝氨酸蛋白酶多肽和/或另外的酶的性能。就此而言,将本发明的丝氨酸蛋白酶多肽用本领域中已知的任何合适的包封材料进行包封。在一些实施方案中,包封材料通常包封本发明的丝氨酸蛋白酶多肽的至少一部分。通常来说,包封材料是水溶性的和/或水分散性的。在一些实施方案中,包封材料具有0℃或更高的玻璃化转变温度(T_g)。玻璃化转变温度在W097/11151中进行了更详细的描述。包封材料通常选自碳水化合物、天然或合成树胶、甲壳质、脱乙酰壳多糖、纤维素和纤维素衍生物、硅酸盐、磷酸盐、硼酸盐、聚乙烯醇、聚乙二醇、固体石蜡以及它们的组合。当包封材料是碳水化合物时,其通常选自单糖、低聚糖、多糖以及它们的组合。在一些典型的实施方案中,包封材

料是淀粉(参见例如,EP 0 922 499;US 4,977,252;US 5,354,559和US 5,935,826)。在一些实施方案中,包封材料是由塑料诸如热塑性塑料、丙烯腈、甲基丙烯腈、聚丙烯腈、聚甲基丙烯腈及它们的混合物制成的微球体;可使用的可商购获得微球体包括但不限于由EXPANCEL[®](Stockviksverken,Sweden)以及PM 6545、PM 6550、PM 7220、PM 7228、EXTENDOSPHERES[®]、LUXSIL[®]、Q-CEL[®]和SPHERICEL[®](PQ Corp.,Valley Forge,PA)供应的那些。

[0152] 如本文所用,本发明的变体蛋白酶可特别用于清洁工业中,其包括但不限于衣物洗涤剂 and 盘碟洗涤剂。这些应用将酶置于各种环境胁迫下。本发明的变体蛋白酶提供超过许多当前使用的酶的优点,这是由于其在各种条件下的稳定性。

[0153] 实际上,存在洗涤时涉及的蛋白酶暴露于其中的多种洗涤条件,包括不同的洗涤剂制剂、洗涤水体积、洗涤水温度、和洗涤时间长度。此外,用于不同地理区域中的洗涤剂制剂具有存在于洗涤水中的其相关组分的不同浓度。例如,欧洲洗涤剂通常具有约4500-5000ppm的洗涤水中的洗涤剂组分,然而日本洗涤剂通常具有约667ppm的洗涤水中的洗涤剂组分。在北美,具体地讲美国,洗涤剂通常具有约975ppm的存在于洗涤水中的洗涤剂组分。

[0154] 低洗涤剂浓度体系包括洗涤水中存在少于约800ppm的洗涤剂组分的洗涤剂。通常认为日本洗涤剂是低洗涤剂浓度体系,因为其具有约667ppm的存在于洗涤水中的洗涤剂组分。

[0155] 中等洗涤剂浓度包括洗涤水中存在介于约800ppm和约2000ppm的洗涤剂组分的洗涤剂。一般认为北美洗涤剂是中等洗涤剂浓度体系,因为其具有约975ppm的存在于洗涤水中的洗涤剂组分。巴西通常具有约1500ppm的存在于洗涤水中的洗涤剂组分。

[0156] 高洗涤剂浓度体系包括洗涤水中存在大于约2000ppm的洗涤剂组分的洗涤剂。一般认为欧洲洗涤剂是高洗涤剂浓度体系,因为其具有约4500-5000ppm的存在于洗涤水中的洗涤剂组分。

[0157] 拉丁美洲洗涤剂一般是高泡磷酸盐助洗剂洗涤剂并且用于拉丁美洲的洗涤剂范围落入中等洗涤剂浓度和高洗涤剂浓度中,因为其范围是从1500ppm到6000ppm的洗涤水中的洗涤剂组分。如上所述,巴西通常具有约1500ppm的存在于洗涤水中的洗涤剂组分。然而,其它高泡磷酸盐助洗剂洗涤剂地区,不限于其它拉丁美洲国家,可具有至多约6000ppm的存在于洗涤水中的洗涤剂组分的高洗涤浓度体系。

[0158] 如前所述,显而易见的是在全世界在典型的洗涤溶液中洗涤剂组合物的浓度从小于约800ppm的洗涤剂组合物(“低洗涤剂浓度地区”),例如在日本约667ppm,至介于约800ppm至约2000ppm(“中等洗涤剂浓度地区”),例如在美国约975ppm和在巴西约1500ppm,至大于约2000ppm(“高洗涤剂浓度地区”),例如在欧洲约4500ppm至约5000ppm和在高泡磷酸盐助洗剂地区约6000ppm而变化。

[0159] 根据经验确定典型的洗涤溶液的浓度。例如,在美国,典型的洗衣机容纳约64.4L的洗涤溶液的体积。因此,为了获得约975ppm的在洗涤溶液中的洗涤剂浓度,必须将约62.79g的洗涤剂组合物加入64.4L洗涤溶液中。该量是通过消费者使用与洗涤剂一起提供的量杯测量到洗涤水中的典型的量。

[0160] 又如,不同地区使用不同的洗涤温度。日本的洗涤水温度通常低于欧洲使用的。例如,北美和日本的洗涤水温度通常介于约10℃和约30℃之间(例如,约20℃),然而欧洲的洗涤水的温度通常介于约30℃和约60℃之间(例如,约40℃)。然而,对于节能的兴趣,许多消费者转而使用冷水洗涤。此外,在一些其它区域中,冷水通常用于衣物洗涤,以及盘碟洗涤应用。在一些实施方案中,本发明的“冷水洗涤”利用适用于在约10℃至约40℃,或约20℃至约30℃,或约15℃至约25℃,以及在约15℃至约35℃的范围内的所有其它组合,和在10℃至40℃内的所有范围的温度下洗涤的“冷水洗涤剂”。

[0161] 又如,不同地区通常具有不同的水硬度。水硬度通常按每加仑混合 $\text{Ca}^{2+}/\text{Mg}^{2+}$ 的格令数(grain)来描述。硬度是水中钙(Ca^{2+})和镁(Mg^{2+})的量的量度。在美国,大多数水都较硬,但硬度不同。中等硬(60-120ppm)至硬(121-181ppm)水具有60至181份每百万份(份每百万份转化成格令每US加仑为ppm#除以17.1等于格令每加仑)的硬矿物质。

[0162] 表I:水硬度

[0163]

水	格令每加仑	份每百万份
软	小于1.0	小于17
略硬	1.0至3.5	17至60
中等硬	3.5至7.0	60至120
硬	7.0至10.5	120至180
非常硬	大于10.5	大于180

[0164] 欧洲水硬度通常大于约10.5(例如约10.5至约20.0)格令每加仑混合的 $\text{Ca}^{2+}/\text{Mg}^{2+}$ (例如约15格令每加仑混合的 $\text{Ca}^{2+}/\text{Mg}^{2+}$)。北美水硬度通常大于日本水硬度,但小于欧洲水硬度。例如,北美水硬度可以介于约3至约10格令、约3至约8格令或约6格令。日本水硬度通常低于北美水硬度,通常小于约4,例如约3格令每加仑混合的 $\text{Ca}^{2+}/\text{Mg}^{2+}$ 。

[0165] 因此,在一些实施方案中,本发明提供了在至少一组洗涤条件(例如水温、水硬度和/或洗涤剂浓度)下显示出惊人洗涤性能的丝氨酸蛋白酶多肽。在一些实施方案中,本发明的丝氨酸蛋白酶多肽与其它丝氨酸蛋白酶多肽蛋白酶在洗涤性能方面相当。在本发明的一些实施方案中,本文中提供的丝氨酸蛋白酶多肽显示出增强的氧化稳定性、增强的热稳定性、增强的在多种条件下的清洁能力和/或增强的螯合剂稳定性。另外,本发明的丝氨酸蛋白酶多肽仍可单独或者与助洗剂和稳定剂组合用于不包含洗涤剂的清洁组合物。

[0166] 在本发明的一些实施方案中,清洁组合物包含按该组合物的重量计约0.00001%至约10%水平的本发明的至少一种丝氨酸蛋白酶多肽,并且按该组合物的重量计的余量(例如,约99.999%至约90.0%)包含清洁助剂材料。在本发明的一些其它实施方案中,本发明的清洁组合物包含按该组合物的重量计约0.0001%至约10%、约0.001%至约5%、约0.001%至约2%、约0.005%至约0.5%水平的至少一种丝氨酸蛋白酶多肽,并且该清洁组合物的余量(例如,按重量计约99.9999%至约90.0%、约99.999%至约98%、约99.995%至约99.5%)包含清洁助剂材料。

[0167] 在一些实施方案中,本发明的清洁组合物包含一种或多种另外的提供清洁性能和/或织物护理和/或盘碟洗涤益处的洗涤剂酶。

[0168] 合适的酶的示例包括但不限于,附加的丝氨酸蛋白酶、酰基转移酶、 α -淀粉酶、 β -

淀粉酶、 α -半乳糖苷酶、阿拉伯糖苷酶、芳基酯酶、 β -半乳糖苷酶、卡拉胶酶、过氧化氢酶、纤维二糖水解酶、纤维素酶、软骨素酶、角质酶、内切- β -1,4-葡聚糖酶、内切- β -甘露聚糖酶、酯酶、外切甘露聚糖酶、半乳聚糖酶、葡糖淀粉酶、半纤维素酶、透明质酸酶、角蛋白酶、漆酶、乳糖酶、木质素酶、脂肪酶、脂氧合酶、甘露聚糖酶、金属蛋白酶、非丝氨酸蛋白酶、氧化酶、果胶酸裂解酶、果胶乙酰酯酶、果胶酶、戊聚糖酶、过氧化物酶、过水解酶、酚氧化酶、磷酸酶、磷脂酶、植酸酶、聚半乳糖醛酸酶、蛋白酶、普鲁兰酶、还原酶、聚鼠李糖半乳糖醛酸酶、 β -葡聚糖酶、鞣酸酶、转谷氨酰胺酶、木聚糖乙酰酯酶、木聚糖酶、木葡聚糖酶、和木糖苷酶、或它们的任何组合或混合物。在一些实施方案中,使用酶的组合(即,“cocktail”),其包含常规可用的酶如淀粉酶、脂肪酶、角质酶和/或纤维素酶与所用蛋白酶结合。

[0169] 除本文中提供的丝氨酸蛋白酶多肽之外,任何其它合适的蛋白酶均可用于本发明的组合物。合适的蛋白酶包括动物、植物或微生物来源的那些。在一些实施方案中,使用微生物蛋白酶。在一些实施方案中,包括经化学或遗传修饰的突变体。在一些实施方案中,蛋白酶是丝氨酸蛋白酶,优选碱性微生物蛋白酶或胰蛋白酶样蛋白酶。碱性蛋白酶的示例包括包括枯草杆菌蛋白酶,尤其是来源于芽孢杆菌的那些(例如,枯草杆菌蛋白酶、迟缓芽孢杆菌枯草杆菌蛋白酶、解淀粉芽孢杆菌枯草杆菌蛋白酶、枯草菌溶素(subtilisin Carlsberg)、枯草杆菌蛋白酶309、枯草杆菌蛋白酶147和枯草杆菌蛋白酶168)。附加的示例包括描述于美国专利RE 34,606、5,955,340、5,700,676、6,312,936和6,482,628中的那些突变体蛋白酶,这些专利均以引用方式并入本文。另外的蛋白酶示例包括但不限于胰蛋白酶(例如,猪或牛来源的胰蛋白酶)和WO 89/06270中所述的镰孢菌(*Fusarium*)蛋白酶。在一些实施方案中,可用于本发明的可商购获得的蛋白酶包括但不限于MAXATASE[®]、MAXACAL[™]、MAXAPEM[™]、OPTICLEAN[®]、OPTIMASE[®]、PROPERASE[®]、PURAFFECT[®]、PURAFFECT[®] OXP、PURAMAX[™]、EXCELLASE[™]、PREFERENZ[™]蛋白酶(例如P100、P110、P280)、EFFECTENZ[™]蛋白酶(例如P1000、P1050、P2000)、EXCELLENZ[™]蛋白酶(例如P1000)、ULTIMASE[®]和PURAFAST[™](Genencor); ALCALASE[®]、SAVINASE[®]、PRIMASE[®]、DURAZYM[™]、POLARZYME[®]、OVOZYME[®]、KANNASE[®]、LIQUANASE[®]、NEUTRASE[®]、RELEASE[®]和 ESPERASE[®](Novozymes); BLAP[™]和BLAP[™]变体(Henkel Kommanditgesellschaft auf Aktien, Duesseldorf, Germany),以及KAP(嗜碱芽孢杆菌枯草杆菌蛋白酶;Kao Corp., Tokyo, Japan)。多种蛋白酶在以下专利中进行了描述:W095/23221、W0 92/21760、W0 09/149200、W0 09/149144、W0 09/149145、W0 11/072099、W0 10/056640、W0 10/056653、W0 11/140364、W0 12/151534、美国专利公开2008/0090747;和美国专利5,801,039、5,340,735、5,500,364、5,855,625、US RE 34,606、5,955,340、5,700,676、6,312,936、6,482,628、8,530,219,以及多篇其它专利。在一些另外的实施方案中,金属蛋白酶可用于本发明,其包括但不限于描述于W01999014341、W01999033960、W01999014342、W01999034003、W02007044993、W02009058303、W02009058661、W02014194032、W02014194034和W02014194054中的金属蛋白酶。示例性的金属蛋白酶包括nprE、在枯草芽孢杆菌中表达的中性金属蛋白酶的重组形式(参见例如,W0 07/044993)和PMN,其是来自解淀粉芽孢杆菌的经纯化的中性金属蛋白酶。

[0170] 另外,任何合适的脂肪酶均可用于本发明。合适的脂肪酶包括但不限于细菌或真

菌来源的那些。本发明涵盖经化学或遗传修饰的突变体。可用的脂肪酶的示例包括柔毛腐质霉(*Humicola lanuginosa*)脂肪酶(参见例如,EP 258 068和EP 305 216);米黑根毛霉(*Rhizomucor miehei*)脂肪酶(参见例如,EP 238 023);假丝酵母属(*Candida*)脂肪酶,诸如南极假丝酵母(*C. antarctica*)脂肪酶(例如,南极假丝酵母脂肪酶A或B;参见例如,EP 214 761),假单胞菌脂肪酶诸如产碱假单胞菌(*P. alcaligenes*)脂肪酶和类产碱假单胞菌(*P. pseudoalcaligenes*)脂肪酶(参见例如,EP 218 272)、洋葱假单胞菌(*P. cepacia*)脂肪酶(参见例如,EP 331 376)、施氏假单胞菌(*P. stutzeri*)脂肪酶(参见例如,GB 1,372,034)、荧光假单胞菌(*P. fluorescens*)脂肪酶、芽孢杆菌脂肪酶(例如,枯草芽孢杆菌脂肪酶[Dartois等人,Biochem. Biophys. Acta 1131:253-260[1993]]);嗜热脂肪芽孢杆菌(*B. stearothermophilus*)脂肪酶[参见例如,JP 64/744992];和短小芽孢杆菌(*B. pumilus*)脂肪酶[参见例如,W0 91/16422]]。

[0171] 此外,很多克隆的脂肪酶可用于本发明的一些实施方案,包括但不限于沙门柏干酪青霉(*Penicillium camembertii*)脂肪酶(参见,Yamaguchi等人,Gene 103:61-67[1991])、白地霉(*Geotricum candidum*)脂肪酶(参见,Schimada等人,J. Biochem., 106:383-388[1989])以及多种根霉(*Rhizopus*)脂肪酶,例如德氏根霉(*R. delemar*)脂肪酶(参见,Hass等人,Gene 109:117-113[1991])、雪白根霉(*R. niveus*)脂肪酶(Kugimiya等人,Biosci Biotech. Biochem. 56:716-719[1992])和米根霉(*R. oryzae*)脂肪酶。

[0172] 其它类型的脂肪酶多肽酶诸如角质酶也可用于本发明的一些实施方案,包括但不限于来源于门多萨假单胞菌(*Pseudomonas mendocina*)的角质酶(参见,W0 88/09367)和来源于腐皮镰孢菌(*Fusarium solani pisi*)的角质酶(参见,W0 90/09446)。

[0173] 另外合适的脂肪酶包括以下脂肪酶,诸如MILIPASE™、LUMA FAST™和LIPOMAX™(Genencor); LIPEX®、LIPOLASE®和LIPOLASE®ULTRA(Novozymes);以及LIPASE P™“Amano”(Amano Pharmaceutical Co. Ltd., Japan)。

[0174] 在本发明的一些实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.00001%至约10%的另外脂肪酶的水平脂肪酶和按该组合物的重量计余量的清洁助剂材料。在本发明的一些其它实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.0001%至约10%、约0.001%至约5%、约0.001%至约2%、约0.005%至约0.5%脂肪酶的水平脂肪酶。

[0175] 在本发明的一些实施方案中,任何合适的淀粉酶均可用于本发明。在一些实施方案中,还可使用任何适用于碱性溶液的淀粉酶(例如, α 和/或 β 淀粉酶)。合适的淀粉酶包括但不限于细菌或真菌来源的那些。在一些实施方案中包括经化学或经遗传修饰的突变体。可用于本发明的淀粉酶包括但不限于从地衣芽孢杆菌获得的 α -淀粉酶(参见例如,GB 1,296,839)。另外的合适淀粉酶包括见于以下专利中的那些:W09510603、W09526397、W09623874、W09623873、W09741213、W09919467、W00060060、W00029560、W09923211、W09946399、W00060058、W00060059、W09942567、W00114532、W002092797、W00166712、W00188107、W00196537、W00210355、W09402597、W00231124、W09943793、W09943794、W02004113551、W02005001064、W02005003311、W00164852、W02006063594、W02006066594、W02006066596、W02006012899、W02008092919、W02008000825、W02005018336、W02005066338、W02009140504、W02005019443、W02010091221、W02010088447、W00134784、

W02006012902、W02006031554、W02006136161、W02008101894、W02010059413、W02011098531、W02011080352、W02011080353、W02011080354、W02011082425、W02011082429、W02011076123、W02011087836、W02011076897、W094183314、W09535382、W09909183、W09826078、W09902702、W09743424、W09929876、W09100353、W09605295、W09630481、W09710342、W02008088493、W02009149419、W02009061381、W02009100102、W02010104675、W02010117511和W02010115021。可用于本发明的可商购获得的淀粉酶包括但不限于DURAMYL[®]、TERMAMYL[®]、FUNGAMYL[®]、STAINZYME[®]、STAINZYME PLUS[®]、STAINZYME ULTRA[®]和BAN[™](Novozymes),以及POWERASE[™]、RAPIDASE[®]和MAXAMYL[®]P(Genencor)。

[0176] 在本发明的一些实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.00001%至约10%的另外淀粉酶的水平淀粉酶和按该组合物的重量计余量的清洁助剂材料。在本发明的一些其它实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.0001%至约10%、约0.001%至约5%、约0.001%至约2%、约0.005%至约0.5%的淀粉酶的水平淀粉酶。

[0177] 在一些另外的实施方案中,任何合适的纤维素酶均可用于本发明的清洁组合物。合适的纤维素酶包括但不限于细菌或真菌来源的那些。在一些实施方案中包括经化学或经遗传修饰的突变体。合适的纤维素酶包括但不限于特异腐质霉(*Humicola insolens*)纤维素酶(参见例如,美国专利4,435,307)。特别合适的纤维素酶是具有颜色护理有益效果的纤维素酶(参见例如,EP 0 495 257)。可用于本发明的可商购获得的纤维素酶包括但不限于CELLUZYME[®]、CAREZYME[®](Novozymes)、REVITALENZ[™]100(Danisco US Inc.)和KAC-500(B)[™](Kao Corporation)。在一些实施方案中,纤维素酶作为成熟野生型纤维素酶或变体纤维素酶的一部分或片段引入,其中N末端的一部分缺失(参见例如,美国专利5,874,276)。另外合适的纤维素酶包括见于W02005054475、W02005056787、美国专利7,449,318和美国专利7,833,773中的那些。在一些实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.00001%至约10%的另外纤维素酶的水平纤维素酶和按该组合物的重量计余量的清洁助剂材料。在本发明的一些其它实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.0001%至约10%、约0.001%至约5%、约0.001%至约2%、约0.005%至约0.5%的纤维素酶的水平纤维素酶。

[0178] 任何适用于洗涤剂组合物的甘露聚糖酶也可用于本发明。合适的甘露聚糖酶包括但不限于细菌或真菌来源的那些。在一些实施方案中包括经化学或经遗传修饰的突变体。可用于本发明的多种甘露聚糖酶是已知的(参见例如,美国专利6,566,114、美国专利6,602,842和美国专利6,440,991,其全部通过引用并入本文)。可用于本发明的可商购获得的甘露聚糖酶包括但不限于MANNASTAR[®]、PURABRITE[™]和MANNAWAY[®]。在一些实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.00001%至约10%的另外甘露聚糖酶的水平甘露聚糖酶和按该组合物的重量计余量的清洁助剂材料。在本发明的一些实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.0001%至约10%、约0.001%至约5%、约0.001%至约2%、约0.005%至约0.5%的甘露聚糖酶的水平甘露聚糖酶。

[0179] 在一些实施方案中,将过氧化物酶与过氧化氢或其来源(例如,过碳酸盐、过硼酸盐或过硫酸盐)组合用于本发明的组合物。在一些另选的实施方案中,将氧化酶与氧组合使用。两种类型的酶优选地与增强剂一起用于“溶液漂白”(即,当将织物在洗涤液中一起洗涤时防止纺织品染料从经染色的织物转移至另一织物)(参见例如,W094/12621和W095/01426)。合适的过氧化物酶/氧化酶包括但不限于植物、细菌或真菌来源的那些。在一些实施方案中包括经化学或经遗传修饰的突变体。在一些实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.00001%至约10%的另外过氧化物酶和/或氧化酶的水平。过氧化物酶和/或氧化酶和按该组合物的重量计余量的清洁助剂材料。在本发明的一些其它实施方案中,本发明的清洁组合物还包含按该组合物的重量计约0.0001%至约10%,约0.001%至约5%,约0.001%至约2%,约0.005%至约0.5%的过氧化物酶和/或氧化酶水平的过氧化物酶和/或氧化酶。

[0180] 在一些实施方案中,附加的酶可使用,其包括但不限于过水解酶(参见例如,W02005056782、W02007106293、W02008063400、W02008106214、和W02008106215)。另外,在一些实施方案中,本文中涵盖上述酶的混合物,特别是一种或多种另外的蛋白酶、淀粉酶、脂肪酶、甘露聚糖酶和/或至少一种纤维素酶的混合物。实际上,预期这些酶的多种混合物均可用于本发明。还预期,丝氨酸蛋白酶多肽和一种或多种另外的酶的不同水平的范围均可独立达到约10%,并且该清洁组合物的余量为清洁助剂材料。清洁助剂材料的具体选择可通过考虑所需清洁的表面、物品或织物以及该组合物在使用期间用于清洁条件的期望形式(例如,通过洗涤剂使用)来容易地进行。

[0181] 合适的清洁助剂材料的示例包括但不限于,表面活性剂、助洗剂、漂白剂、漂白活性剂、漂白催化剂、其它酶、酶稳定体系、螯合剂、荧光增白剂、去垢性聚合物、染料转移剂、染料转移抑制剂、催化材料、过氧化氢、过氧化氢源、预成形过酸、聚合物分散剂、粘土污垢去除剂、结构增弹剂、分散剂、抑泡剂、染料、香料、着色剂、填料盐、水溶助长剂、光活化剂、荧光剂、织物调理剂、织物软化剂、载体、水溶助长剂、加工助剂、溶剂、颜料、可水解的表面活性剂、防腐剂、抗氧化剂、抗收缩剂、抗皱剂、杀菌剂、杀真菌剂、色斑、银护理剂、防晦暗和/或抗腐蚀剂、碱度来源、增溶剂、载体、加工助剂、颜料和pH控制剂(参见例如,美国专利6,610,642;6,605,458;5,705,464;5,710,115;5,698,504;5,695,679;5,686,014和5,646,101)。下文详细地举例说明了具体清洁组合物材料的实施方案。在其中清洁助剂材料不与清洁组合物中的本发明变体蛋白酶相容的实施方案中,则使用保持清洁助剂材料和一种或多种蛋白酶分离(即不彼此接触)直至适宜组合这两种组分的合适方法。此类分离方法包括本领域已知的任何合适的方法(例如,胶囊锭、封装、片剂、物理分离等)。

[0182] 在一些实施方案中,可用于清洁需要去除蛋白质性污渍的多种表面的组合物中包含有效量的本文提供的一种或多种丝氨酸蛋白酶多肽。这样的清洁组合物包括用于诸如清洁硬质表面、织物和盘碟的应用的清洁组合物。实际上,在一些实施方案中,本发明提供了织物清洁组合物,而在另一些实施方案中,本发明提供了非织物清洁组合物。值得注意的是,本发明还提供了适于个人护理的清洁组合物,包括口腔护理组合物(包括洁牙剂、牙膏、漱口剂等,以及义齿清洁组合物)、皮肤清洁组合物和毛发清洁组合物。预期,本发明涵盖任何形式(即,液体、颗粒、棒、半固体、凝胶、乳剂、片剂、胶囊剂等)的洗涤剂组合物。

[0183] 以举例的方式,在下文对数种其中可使用本发明的丝氨酸蛋白酶多肽的清洁组合

物进行更详细描述。在其中将本发明的清洁组合物配制成适用于洗衣机洗涤方法的组合物的一些实施方案中,本发明的组合物优选地包含至少一种表面活性剂和至少一种助洗剂化合物,以及一种或多种清洁助剂材料,其优选地选自有机聚合物化合物、漂白剂、另外的酶、抑泡剂、分散剂、石灰皂分散剂、悬污剂和抗再沉积剂以及腐蚀抑制剂。在一些实施方案中,衣物洗涤组合物还包含软化剂(即,作为另外的清洁助剂材料)。本发明的组合物还可用于固体或液体形式的洗涤剂添加剂产品。这样的添加剂产品旨在补充和/或提高常规洗涤剂组合物的性能,并且可在洗涤过程的任意阶段添加。在一些实施方案中,本文的衣物洗涤剂组合物的密度的范围是约400g/L至约1200g/L,然而在其它实施方案中,其范围是在20℃下测量的约500g/L至约950g/L组合物。

[0184] 在配制为用于人工盘碟洗涤方法的组合物的实施方案中,本发明的组合物优选地包含至少一种表面活性剂,并且优选地包含至少一种选自以下的另外的清洁助剂材料:有机聚合物化合物、泡沫增强剂、II族金属离子、溶剂、水溶助长剂和另外的酶。

[0185] 在一些实施方案中,各种清洁组合物诸如美国专利6,605,458中提供的那些可与本发明的丝氨酸蛋白酶多肽一起使用。因此,在一些实施方案中,包含本发明的至少一种丝氨酸蛋白酶多肽的组合物是紧凑型颗粒织物清洁组合物,而在另一些实施方案中,该组合物是可用于洗涤着色织物的颗粒织物清洁组合物,在另外的实施方案中,该组合物是通过洗涤能力提供软化的颗粒织物清洁组合物,在另外的实施方案中,该组合物是重垢型液体织物清洁组合物。在一些实施方案中,包含本发明的至少一种丝氨酸蛋白酶多肽的组合物是织物清洁组合物,诸如美国专利6,610,642和6,376,450中所述的那些。此外,本发明的丝氨酸蛋白酶多肽可用于在欧洲或日本洗涤条件下具有特别实用性的颗粒衣物洗涤剂组合物(参见例如,美国专利6,610,642)。

[0186] 在一些另选的实施方案中,本发明提供了包含本文中提供的至少一种丝氨酸蛋白酶多肽的硬质表面清洁组合物。因此,在一些实施方案中,包含本发明的至少一种丝氨酸蛋白酶多肽的组合物是硬质表面清洁组合物,诸如美国专利6,610,642、6,376,450和6,376,450中所述的那些。

[0187] 在另一个实施方案中,本发明提供了包含本文中提供的至少一种丝氨酸蛋白酶多肽的盘碟洗涤组合物。因此,在一些实施方案中,包含本发明的至少一种丝氨酸蛋白酶多肽的组合物是硬质表面清洁组合物,诸如美国专利6,610,642和6,376,450中所述的那些。在另一些实施方案中,本发明提供了包含本文中提供的至少一种丝氨酸蛋白酶多肽的盘碟洗涤组合物。在一些另外的实施方案中,包含本发明的至少一种丝氨酸蛋白酶多肽的组合物包含口腔护理组合物,诸如美国专利6,376,450和6,376,450中所述的那些。前述美国专利6,376,450、6,605,458、6,605,458和6,610,642中包含的化合物和清洁助剂材料的配制和描述可用于本文提供的丝氨酸蛋白酶多肽。

[0188] 本发明的清洁组合物可配制成任何合适的形式并且通过由配制者选择的任何方法来制备(参见例如,美国专利5,879,584;5,691,297;5,574,005;5,569,645;5,565,422;5,516,448;5,489,392;和5,486,303。当期望低pH清洁组合物时,通过添加诸如单乙醇胺或酸性材料(诸如HCl)的材料来调节此类组合物的pH。

[0189] 在一些实施方案中,根据本发明的清洁组合物包含酸化颗粒或氨基羧基助洗剂。氨基羧基助洗剂的示例包括按氨基羧酸、其盐和衍生物、在一些实施方案中,氨基羧基

助洗剂是氨基多羧基助洗剂,例如甘氨酸-N,N-二乙酸或通式 $\text{MOOC-CHR-N}(\text{CH}_2\text{COOM})_2$ 的衍生物,其中R是 C_{1-12} 烷基并且M是碱金属。在一些实施方案中,氨基羧基助洗剂可为甲基甘氨酸二乙酸(MGDA)、GLDA(谷氨酸-N,N-二乙酸)、亚氨基二琥珀酸(IDS)、羧甲基菊粉及其盐和衍生物、天冬氨酸-N-单乙酸(ASMA)、天冬氨酸-N,N-二乙酸(ASDA)、天冬氨酸-N-单丙酸(ASMP)、亚氨基二琥珀酸(IDA)、N-(2-磺甲基)天冬氨酸(SMAS)、N-(2-磺乙基)天冬氨酸(SEAS)、N-(2-磺甲基)谷氨酸(SMGL)、N-(2-磺乙基)谷氨酸(SEGL)、IDS(亚氨基二乙酸)及其盐和衍生物(诸如N-甲基亚氨基二乙酸(MIDA))、 α -丙氨酸-N,N-二乙酸(α -ALDA)、丝氨酸-N,N-二乙酸(SED)、异丝氨酸-N,N-二乙酸(ISDA)、苯基丙氨酸-N,N-二乙酸(PHDA)、邻氨基苯甲酸-N,N-二乙酸(ANDA)、对氨基苯磺酸-N,N-二乙酸(SLDA)、牛磺酸-N,N-二乙酸(TUDA)和磺甲基-N,N-二乙酸(SMDA)以及它们的碱金属盐和衍生物。在一些实施方案中,酸化颗粒具有约400 μ 至约1200 μ 的加权几何平均粒度和至少550g/L的堆密度。在一些实施方案中,酸化颗粒包含至少约5%的助洗剂。

[0190] 在一些实施方案中,酸化颗粒可包含任意酸,包括有机酸和无机酸。有机酸可具有一个或两个羧基并且在一些情况下可具有至多15个碳,特别是至多10个碳,诸如甲酸、乙酸、丙酸、癸酸、草酸、琥珀酸、脂肪酸、马来酸、富马酸、癸二酸、苹果酸、乳酸、乙醇酸、酒石酸和水合乙醛酸。在一些实施方案中,所述酸是柠檬酸。无机酸包括盐酸和硫酸。在一些情况下,本发明的酸化颗粒是包含高水平氨基羧基助洗剂的高活性颗粒。已发现,硫酸还有助于最终颗粒的稳定性。

[0191] 虽然出于本发明的目的不是必要的,但下文所示的助剂的非限制性列表适用于即时清洁组合物。在一些实施方案中,引入这些助剂例如是为了有助于或增强清洁性能(对于处理需清洁的基质),或者为了修饰清洁组合物的美观,如在香料、着色剂、染料等的情况下。应当理解此类助剂是除本发明的变体蛋白质之外的。这些附加组分的精确性质,及其掺入量将取决于组合物的物理形式和其将用于的清洁操作的性质。合适的助剂材料包括但不限于,表面活性剂、助洗剂、螯合剂、染料转移抑制剂、沉积助剂、分散剂、附加的酶、和酶稳定剂、催化材料、漂白活性剂、漂白增强剂、过氧化氢、过氧化氢源、预成形过酸、聚合物分散剂、粘土污垢去除/抗再沉积剂、增白剂、抑泡剂、染料、香料、结构增弹剂、织物软化剂、载体、水溶助长剂、加工助剂和/或颜料。除下文公开内容之外,此类其它助剂的适宜示例和用量存在于美国专利5,576,282、6,306,812和6,326,348中,其以引用方式并入。前述助剂成分可构成本发明的清洁组合物的其余部分。

[0192] 在一些实施方案中,根据本发明的清洁组合物包含至少一种表面活性剂和/或表面活性剂体系,其中所述表面活性剂选自非离子表面活性剂、阴离子表面活性剂、阳离子表面活性剂、两性表面活性剂、两性离子表面活性剂、半极性非离子表面活性剂以及它们的混合物。在一些低pH清洁组合物实施方案中(例如,具有约3至约5的净pH的组合物),所述组合物通常不包含烷基烷氧基化硫酸盐,因为据信此类表面活性剂可被此类组合物水解。按清洁组合物的重量计,在一些实施方案中,表面活性剂以约0.1%至约60%的水平存在,而在一些另选的实施方案中,该水平为约1%至约50%,而在另一些实施方案中,该水平为约5%至约40%。

[0193] 在一些实施方案中,本发明的清洁组合物包含一种或多种洗涤剂助洗剂或助洗剂体系。在掺入至少一种助洗剂的一些实施方案中,清洁组合物包含按该清洁组合物的重量

计至少约1%、约3%至约60%或者甚至约5%至约40%的助洗剂。助洗剂包括但不限于多磷酸的碱金属盐、铵盐和链烷醇铵盐；碱金属硅酸盐；碱土金属和碱金属碳酸盐；铝硅酸盐；聚羧酸盐化合物；醚羟基聚羧酸盐；马来酸酐与乙烯或乙烯基甲基醚、1,3,5-三羟基苯-2,4,6-三磺酸、以及羧甲基氧琥珀酸的共聚物；聚乙酸诸如乙二胺四乙酸和次氨基三乙酸的各種碱金属盐、铵盐和取代的铵盐；以及聚羧酸诸如苯六甲酸、琥珀酸、柠檬酸、氧联二琥珀酸、聚马来酸、苯-1,3,5-三羧酸、羧甲基氧琥珀酸；以及它们的可溶性盐。实际上，预期任何合适的助洗剂都可用于本发明的各种实施方案。

[0194] 在一些实施方案中，助洗剂形成水溶性硬度离子络合物（例如，螯合助洗剂），诸如柠檬酸盐和聚磷酸盐（例如，三聚磷酸钠和六水合三聚磷酸钠、三聚磷酸钾以及混合的三聚磷酸钠与三聚磷酸钾等）。预期，任何合适的助洗剂都可用于本发明，包括本领域中已知的那些（参见例如，EP2 100 949）。

[0195] 在一些实施方案中，用于本文的助洗剂包括磷酸盐助洗剂和非磷酸盐助洗剂。在一些实施方案中，助洗剂是磷酸盐助洗剂。在一些实施方案中，助洗剂是非磷酸盐助洗剂。如果存在的话，以按组合物的重量计0.1%至80%、或者5%至60%、或者10%至50%的水平使用助洗剂。在一些实施方案中，产品包含磷酸盐助洗剂和非磷酸盐助洗剂的混合物。合适的磷酸盐助洗剂包括单磷酸盐、二磷酸盐、三聚磷酸盐或低聚多磷酸盐，包括这些化合物的碱金属盐，包括钠盐。在一些实施方案中，助洗剂可为三聚磷酸钠(STPP)。另外，组合物可包含碳酸盐和/或柠檬酸盐，优选柠檬酸盐，其有助于实现本发明的中性pH组合物。其它合适的非磷酸盐助洗剂包括多元羧酸的均聚物和共聚物以及它们部分或完全中和的盐，单体多元羧酸和羟基羧酸以及它们的盐。在一些实施方案中，上述化合物的盐包括铵盐和/或碱金属盐，即锂盐、钠盐和钾盐，包括钠盐。合适的多元羧酸包括无环羧酸、脂环族羧酸、杂环羧酸和芳族羧酸，其中在一些实施方案中，它们可包含至少两个羧基基团，在每种情况下，所述羧基均彼此隔开，在一些情况下，彼此相隔不超过两个碳原子。

[0196] 在一些实施方案中，本发明的清洁组合物包含至少一种螯合剂。合适的螯合剂包括但不限于铜、铁和/或锰螯合剂以及它们的混合物。在使用至少一种螯合剂的实施方案中，本发明的清洁组合物包含按主题清洁组合物的重量计约0.1%至约15%或者甚至约3.0%至约10%的螯合剂。

[0197] 在另一些实施方案中，本文中提供的清洁组合物含有至少一种沉积助剂。合适的沉积助剂包括但不限于聚乙二醇；聚丙二醇；聚羧酸盐；去垢性聚合物，诸如聚对苯二甲酸；粘土，诸如高岭土、蒙脱土、绿坡缕石(atapulgit)、伊利石、膨润土、多水高岭土以及它们的混合物。

[0198] 如本文所指出的，在一些实施方案中，抗再沉积剂可用于本发明的一些实施方案。在一些实施方案中，可使用非离子表面活性剂。例如，在自动盘碟洗涤实施方案中，非离子表面活性剂可用于表面修饰目的，特别是用于掩盖(sheeting)，以避免成膜和起斑以及用于改善光泽度。这些非离子表面活性剂还可用于防止污垢再沉积。在一些实施方案中，抗再沉积剂是本领域中已知的非离子表面活性剂（参见例如，EP2 100 949）。在一些实施方案中，非离子表面活性剂可以是乙氧基化的非离子表面活性剂、环氧封端的聚（烷氧基化）醇和氧化胺表面活性剂。

[0199] 在一些实施方案中，本发明的清洁组合物包含一种或多种染料转移抑制剂。合适

的聚合物染料转移抑制剂包括但不限于聚乙烯基吡咯烷酮聚合物、聚胺N-氧化物聚合物、N-乙烯基吡咯烷酮和N-乙烯基咪唑的共聚物、聚乙烯基噁唑烷酮和聚乙烯基咪唑或者它们的混合物。在其中使用至少一种染料转移抑制剂的实施方案中,本发明的清洁组合物包含按该组合物的重量计约0.0001%至约10%、约0.01%至约5%或者甚至约0.1%至约3%。

[0200] 在一些实施方案中,本发明的组合物中包含硅酸盐。在一些这样的实施方案中,可使用硅酸钠(例如,二硅酸钠、偏硅酸钠和结晶页硅酸盐)。在一些实施方案中,硅酸盐以约1%至约20%的水平存在。在一些实施方案中,硅酸盐以按组合物的重量计约5%至约15%的水平存在。

[0201] 在另一些实施方案中,本发明的清洁组合物还包含分散剂。合适的水溶性有机材料包括但不限于均聚酸或共聚酸或它们的盐,其中多元羧酸包含至少两个羧基,其彼此相隔不超过两个碳原子。

[0202] 在一些另外的实施方案中,通过任何合适的技术来稳定清洁组合物中使用的酶。在一些实施方案中,本文中所采用的酶通过在成品组合物中存在可向该酶提供钙和/或镁离子的钙和/或镁离子的水溶性源来稳定。在一些实施方案中,酶稳定剂包括低聚糖、多糖和无机二价金属盐,包括碱土金属,诸如钙盐,例如甲酸钠。预期,用于酶稳定的多种技术均可用于本发明。例如,在一些实施方案中,本文中所采用的酶通过在成品组合物中存在可向该酶提供锌(II)、钙(II)和/或镁(II)离子的此类离子的水溶性源来稳定,其它金属离子(例如钡(II)、钪(II)、铁(II)、锰(II)、铝(III)、锡(II)、钴(II)、铜(II)、镍(II)和氧钒(IV))同样如此。氯化物和硫酸盐也可用于本发明一些实施方案。合适的低聚糖和多糖(例如,糊精)的示例在本领域中是已知的(参见例如,WO 07/145964)。在一些实施方案中,如果期望的话,还可使用可逆性蛋白酶抑制剂,诸如含硼化合物(例如,硼酸盐、4-甲酰基-苯基硼酸)和/或使用三肽醛以进一步提高稳定性。

[0203] 在一些实施方案中,本发明的组合物中存在漂白剂、漂白活性剂和/或漂白催化剂。在一些实施方案中,本发明的清洁组合物包含无机和/或有机漂白化合物。无机漂白剂包括但不限于过氧化氢合物盐(例如过硼酸盐、过碳酸盐、过磷酸盐、过硫酸盐和过硅酸盐)。在一些实施方案中,无机过氧化氢合物盐是碱金属盐。在一些实施方案中,无机过氧化氢合物盐作为结晶固体包含在内而无需额外保护,但是在一些其它实施方案中,该盐是经包被的。本领域中已知的任何合适的盐均可用于本发明(参见例如,EP 2 100 949)。

[0204] 在一些实施方案中,在本发明的组合物中使用漂白活性剂。漂白活性剂通常是有机过酸前体,其增强在60°C和以下的温度下在清洁过程中的漂白作用。适用于本文的漂白活性剂包括这样的化合物,其在过水解条件下产生优选具有约1至约10个碳原子、特别是约2至约4个碳原子的脂族过氧羧酸,和/或任选经取代的过苯甲酸。附加的漂白活性剂在本领域中是已知的并且可用于本发明(参见例如,EP 2 100 949)。

[0205] 此外,在一些实施方案中并且如本文中进一步描述的,本发明的清洁组合物还包含至少一种漂白催化剂。在一些实施方案中,可使用三氮杂环壬烷锰及相关络合物,以及钴、铜、锰和铁络合物。附加的漂白催化剂可用于本发明中(参见例如,US 4,246,612;5,227,084;4,810,410;WO 99/06521;和EP 2 100 949)。

[0206] 在一些实施方案中,本发明的清洁组合物包含一种或多种催化金属络合物。在一

些实施方案中,可使用含金属的漂白催化剂。在一些实施方案中,金属漂白催化剂包含催化剂体系,该催化剂体系包含:具有确定的漂白催化活性的过渡金属阳离子(例如,铜、铁、钛、钪、钨、钼或锰阳离子)、具有微小漂白催化活性或无漂白催化活性的辅助性金属阳离子(例如,锌或铝阳离子),以及对催化性和辅助性金属阳离子具有确定的稳定性常数的螯合剂(sequesterant),特别地可使用乙二胺四乙酸、乙二胺四(亚甲基膦酸)以及它们的水溶性盐(参见例如,美国专利4,430,243)。在一些实施方案中,借助于锰化合物来催化本发明的清洁组合物。这样的化合物及其使用水平在本领域中是熟知的(参见例如,美国专利5,576,282)。在另外的实施方案中,钴漂白催化剂可用于本发明的清洁组合物。多种钴漂白催化剂在本领域中是已知的(参见例如,美国专利5,597,936和5,595,967)并且可通过已知的操作来容易地制备。

[0207] 在另一些实施方案中,本发明的清洁组合物包含大多环刚性配体(MRL)的过渡金属络合物。实际上而非限制,在一些实施方案中,对本发明提供的组合物和清洁方法进行调整以在水性洗涤介质中提供约至少亿分之一的活性MRL物质,而在另一些实施方案中,在洗涤液中提供约0.005ppm至约25ppm、更优选约0.05ppm至约10ppm并且最优选约0.1ppm至约5ppm的MRL。

[0208] 在一些实施方案中,在本发明的过渡金属漂白催化剂中的过渡金属包括但不限于锰、铁和铬。MRL还包括但不限于交叉桥接的特殊超刚性配体(例如,5,12-二乙基-1,5,8,12-四氮杂双环[6.6.2]十六烷)。合适的过渡金属MRL容易通过已知的过程制备(参见例如,WO 2000/32601和美国专利6,225,464)。

[0209] 在一些实施方案中,本发明的清洁组合物包含金属护理剂。金属护理剂可用于防止和/或降低金属(包括铝、不锈钢和非铁金属(例如银和铜))的锈蚀、腐蚀和/或氧化。合适的金属护理剂包括描述于EP 2 100 949、WO 94/26860和WO 94/26859中的那些。在一些实施方案中,金属护理剂是锌盐。在一些另外的实施方案中,本发明的清洁组合物包含按重量计约0.1%至约5%的一种或多种金属护理剂。

[0210] 在一些实施方案中,清洁组合物是具有变体丝氨酸蛋白酶多肽蛋白酶的高密度液体(HDL)组合物。HDL液体衣物洗涤剂可包含去污表面活性剂(10%-40%),所述去污表面活性剂包括阴离子去污表面活性剂(选自直链或支链或无规链的,取代或未取代的烷基硫酸盐、烷基磺酸盐、烷基烷氧基化硫酸盐、烷基磷酸盐、烷基膦酸盐、烷基羧酸盐、和/或它们的混合物);并且任选地非离子表面活性剂(选自直链或支链或无规链的、取代或未取代的烷基烷氧基化醇,例如C₈-C₁₈烷基乙氧基化醇和/或C₆-C₁₂烷基酚烷氧基化物),任选地其中阴离子去污表面活性剂(具有6.0至9的亲水指数(HLC))与非离子去污表面活性剂的重量比大于1:1。合适的去污表面活性剂还包括阳离子去污表面活性剂(选自烷基吡啶鎓化合物、烷基季铵化合物、烷基季磷鎓化合物、烷基叔铊鎓化合物和/或它们的混合物);两性离子去污表面活性剂和/或两性去污表面活性剂(选自链烷醇胺磺基-甜菜碱);两性表面活性剂;半极性非离子表面活性剂以及它们的混合物。

[0211] 所述组合物可任选地包含表面活性增强聚合物,其由下列组成:两亲性烷氧基化油脂清洁聚合物(选自具有支化亲水性和疏水性的烷氧基化聚合物,诸如在0.05重量%-10重量%范围内的烷氧基化聚亚烷基胺)和/或无规接枝聚合物(通常由包选自以下的单体的亲水性主链构成:不饱和C₁-C₆羧酸、醚、醇、醛、酮、酯、糖单元、烷氧基单元、马来酸酐、饱

和多元醇诸如甘油、以及它们的混合物；和选自以下的一个或多个疏水性侧链： C_4 - C_{25} 烷基基团、聚丙烯、聚丁烯、饱和 C_1 - C_6 一元羧酸的乙烯基酯、丙烯酸或甲基丙烯酸的 C_1 - C_6 烷基酯、以及它们的混合物。

[0212] 组合物可包含附加的聚合物，诸如去垢性聚合物（包括阴离子封端的聚酯，例如SRP1，包含至少一个选自糖类、二元羧酸、多元醇以及它们的组合的单体单元的呈无规或嵌段构型的聚合物，基于对苯二甲酸乙二酯的聚合物及其呈无规或嵌段构型的共聚物，例如Repel-o-tex SF、SF-2和SRP6、Texcare SRA100、SRA300、SRN100、SRN170、SRN240、SRN300和SRN325、Marloquest SL）、抗再沉积聚合物（0.1重量%至10重量%，包括羧酸酯聚合物，诸如包含至少一种选自下列的单体的聚合物：丙烯酸、马来酸（或马来酸酐）、富马酸、衣康酸、乌头酸、中康酸、柠康酸、亚甲基丙二酸、以及它们的任何混合物，乙烯基吡咯烷酮均聚物和/或聚乙二醇，分子量在500至100,000Da的范围内）；纤维素聚合物（包括选自烷基纤维素、烷基烷氧基烷基纤维素、羧基烷基纤维素、烷基羧基烷基纤维素的那些，其示例包括羧甲基纤维素、甲基纤维素、甲基羟乙基纤维素、甲基羧甲基纤维素以及它们的混合物）和聚合物羧酸酯（诸如，马来酸酯/丙烯酸酯无规共聚物或聚丙烯酸酯均聚物）。

[0213] 该组合物还可包含饱和或不饱和的脂肪酸，优选饱和或不饱和的 C_{12} - C_{24} 脂肪酸（0重量%至10重量%）；沉积助剂（其示例包括多糖（优选地纤维素聚合物）、聚二烯丙基二甲基卤化铵（DADMAC）和DADMAC与乙烯基吡咯烷酮的共聚物、丙烯酰胺、咪唑、咪唑啉鎓卤化物以及它们的混合物（其为无规或嵌段构型的）、阳离子瓜尔胶、阳离子纤维素（诸如阳离子羟乙基纤维素）、阳离子淀粉、阳离子聚丙烯酰胺以及它们的混合物）。

[0214] 该组合物还可包含以下物质：染料转移抑制剂，其示例包括酞菁锰、过氧化物酶、聚乙烯基吡咯烷酮聚合物、聚胺N-氧化物聚合物、N-乙烯基吡咯烷酮和N-乙烯咪唑的共聚物、聚乙烯基噁唑烷酮和聚乙烯咪唑和/或它们的混合物；螯合剂，其示例包括乙二胺四乙酸（EDTA）、二亚乙基三胺五亚甲基膦酸（DTPMP）、羟基-乙烷二膦酸（HEDP）、乙二胺N,N'-二琥珀酸（EDDS）、甲基甘氨酸二乙酸（MGDA）、二亚乙基三胺五乙酸（DTPA）、丙二胺四乙酸（PDTA）、2-羟基吡啶-N-氧化物（HPNO）、或甲基甘氨酸二乙酸（MGDA）、谷氨酸N,N'-二乙酸（N,N'-二羧基甲基谷氨酸四钠盐（GLDA））、次氨基三乙酸（NTA）、4,5-二羟基-间-苯二磺酸、柠檬酸及其任意盐、N-羟乙基乙二胺三乙酸（HEDTA）、三亚乙基四胺六乙酸（TTHA）、N-羟乙基亚氨基二乙酸（HEIDA）、二羟基乙基甘氨酸（DHEG）、乙二胺四丙酸（EDTP）以及它们的衍生物。

[0215] 所述组合物还可包含选自下列的酶（一般约0.01重量%活性酶至0.5重量%活性酶）：蛋白酶；淀粉酶；脂肪酶；纤维素酶；胆碱氧化酶；过氧化物酶/氧化酶；果胶酸裂解酶；甘露聚糖酶；角质酶；漆酶；磷脂酶；溶血磷脂酶；酰基转移酶；过水解酶；芳基酯酶以及它们的任何混合物。该组合物可包含酶稳定剂（其示例包括多元醇（诸如丙二醇或甘油）、糖或糖醇、乳酸、可逆性蛋白酶抑制剂、硼酸或硼酸衍生物，例如芳族硼酸酯或苯基硼酸衍生物，诸如4-甲酰基苯基硼酸）。

[0216] 该组合物还可包含硅氧烷或基于脂肪酸的抑泡剂、调色染料、钙和镁阳离子、可视信号成分、消泡剂（0.001重量%至约4.0重量%）、和/或结构剂/增稠剂（0.01重量%至5重量%，选自甘油二酯和甘油三酯、乙二醇二硬脂酸酯、微晶纤维素、基于纤维素的材料、微纤维纤维素、生物聚合物、黄原胶、结冷胶、以及它们的混合物）。

[0217] 该组合物可为任何液体形式如液体或凝胶形式，或者它们的任意组合。

[0218] 在一些实施方案中,以单位剂型提供本发明的清洁组合物,包括片剂、胶囊剂、扁囊剂(sachet)、袋剂(pouch)和多隔室袋剂。在一些实施方案中,将单位剂型设计成提供成分在多隔室袋剂(或其它单位剂量形式)内的受控释放。合适的单位剂型和受控释放形式在本领域中是已知的(有关适用于单位剂型和受控释放形式的材料,参见例如EP 2 100 949, WO 02/102955, 美国专利4,765,916和4,972,017,以及W004/111178)。在一些实施方案中,通过经水溶性膜或水溶性小袋包装的片剂来提供单位剂型。多种单位剂型形式提供于EP 2 100 947和W02013/165725中(其在此通过引用并入)并且在本领域中是已知的。

[0219] 在一些实施方案中,清洁组合物是具有变体丝氨酸蛋白酶多肽蛋白酶的高密度粉末(HDD)组合物。HDD粉末衣物洗涤剂可包含去污表面活性剂,其包括阴离子去污表面活性剂(例如,直链或支链或无规链的、取代或未取代的烷基硫酸盐、烷基磺酸盐、烷基烷氧基化硫酸盐、烷基磷酸盐、烷基膦酸盐、烷基羧酸盐和/或它们的混合物),非离子去污表面活性剂(例如,直链或支链或无规链的、取代或未取代的C8-C18烷基乙氧基化物、和/或C6-C12烷基酚烷氧基化物)、阳离子去污表面活性剂(例如,烷基吡啶鎓化合物、烷基季铵化合物、烷基季磷化合物、烷基三元磷化合物、以及它们的混合物)、两性离子和/或两性去污表面活性剂(例如,链烷醇胺磺基甜菜碱);两性表面活性剂;半极性非离子表面活性剂以及它们的混合物;助洗剂(不含磷酸盐助洗剂(例如,沸石助洗剂,其示例包括沸石A、沸石X、沸石P、和沸石MAP,其在0重量%至小于10重量%的范围内);磷酸盐助洗剂(例如,三聚磷酸钠,其在0重量%至小于10重量%范围内);柠檬酸、柠檬酸盐和次氨基三乙酸或其盐,其在小于15重量%范围内;硅酸盐(硅酸钠或硅酸钾或偏硅酸钠,其在0重量%至小于10重量%范围内,或层状硅酸盐(SKS-6));碳酸盐(碳酸钠或碳酸氢钠,其在0重量%至小于10重量%范围内);以及漂白剂(包括光漂白剂,例如磺化酞菁锌、磺化酞菁铝、咕吨类染料、以及它们的混合物);疏水性或亲水性漂白活性剂(例如,十二烷基氧苯磺酸盐、癸酰基氧苯磺酸盐、癸酰基氧苯甲酸或其盐、3,5,5-三甲基己酰基氧苯磺酸盐、四乙酰基乙二胺-TAED、壬酰基氧苯磺酸盐-NOBS、腈季铵化合物、以及它们的混合物);过氧化氢;过氧化氢源(例如,无机过氧化氢化合物盐,其示例包括过硼酸盐、过碳酸盐、过硫酸盐、过磷酸盐或过硅酸盐的一水合钠盐或四水合钠盐);预成形的亲水性和/或疏水性过酸(例如,过羧酸和盐、过碳酸和盐、过亚胺酸和盐、以及过一硫酸和盐)以及它们的混合物和/或漂白催化剂(例如,亚胺漂白增效剂,其示例包括亚胺阳离子和聚离子;亚胺鎓两性离子;经修饰的胺;经修饰的氧化胺;N-磺酰基亚胺;N-膦酰基亚胺;N-酰基亚胺;噻二唑二氧化物;全氟亚胺;环状糖酮以及它们的混合物;含金属的漂白催化剂(例如,铜、铁、钛、钆、钼、或锰阳离子连同辅助性金属阳离子,诸如锌或铝和多价螯合剂诸如乙二胺四乙酸、乙二胺四(亚甲基膦酸)及其水溶性盐)。

[0220] 该组合物还可包含另外的洗涤剂成分,包括香料微胶囊、经淀粉包封的香料香精剂、调色剂、另外的聚合物(包括织物完整剂和阳离子聚合物)、锁色成分(dye lock ingredient)、织物软化剂、增白剂(例如C.I. 荧光增白剂)、絮凝剂、螯合剂、烷氧基化多胺、织物沉积助剂和/或环糊精。

[0221] 在一些实施方案中,清洁组合物是具有本发明的丝氨酸蛋白酶的自动盘碟洗涤(ADW)洗涤剂组合物。ADW洗涤剂组合物可包含以按重量计0%至10%的量存在的选自以下的两种或更多种非离子表面活性剂:乙氧基化非离子表面活性剂、醇烷氧基化表面活性剂、环氧封端的聚(烷氧基化的)醇或氧化胺表面活性剂;在5%-60%范围内的助洗剂,包含磷

酸盐(单磷酸盐、二磷酸盐、三聚磷酸盐或低聚多磷酸盐,优选三聚磷酸钠-STPP)或不含磷酸盐的助洗剂(基于氨基酸的化合物,其示例包括MGDA(甲基-甘氨酸-二乙酸)及其盐和衍生物、GLDA(谷氨酸-N,N二乙酸)及其盐和衍生物、IDS(亚氨基二琥珀酸)及其盐和衍生物、羧甲基菊粉及其盐和衍生物以及它们的混合物、次氨基三乙酸(NTA)、二亚乙基三胺五乙酸(DTPA)、B-丙氨酸二乙酸(B-ADA)以及它们的盐]、多元羧酸的均聚物和共聚物以及它们的部分或完全中和的盐、单体多元羧酸和羟基羧酸以及它们的盐,在按重量计0.5%至50%的范围内;在按重量计约0.1%至约50%范围内的磺化/羧基化聚合物(赋予产品尺寸稳定性);在按重量计约0.1%至约10%范围内的干燥助剂(选自聚酯,尤其是任选地与具有3至6个有助于多聚的官能团的其它单体在一起的阴离子聚酯,特别是酸、醇或酯官能团,反应性环状碳酸酯和脲类型的聚碳酸酯-、聚脲-和/或聚脲-聚有机硅氧烷化合物或它们的前体化合物);在按重量计约1%至约20%范围内的硅酸盐(硅酸钠或硅酸钾,例如二硅酸钠、偏硅酸钠和结晶页硅酸盐);漂白无机物(例如过氧化氢合物盐,诸如过硼酸盐、过碳酸盐、过磷酸盐、过硫酸盐和过硅酸盐)和有机物(例如,有机过氧酸,包括二酰基和四酰基过氧化物,特别是二过氧十二烷二酸、二过氧十四烷二酸和二过氧十六烷二酸);在按重量计约0.1%至约10%范围内的漂白活性剂-有机过酸前体;漂白催化剂(选自三氮杂环壬烷合锰及相关络合物,联吡啶胺合Co、Cu、Mn和Fe及相关络合物,以及五胺乙酸酯合钴(III)及相关络合物);在按重量计约0.1%至5%范围内的金属护理剂(选自苯并三唑(benzotriazole)、金属盐及络合物,和/或硅酸盐);在约0.01至5.0mg活性酶/g自动盘碟洗涤剂组合物范围内的酶(酰基转移酶、 α -淀粉酶、 β -淀粉酶、 α -半乳糖苷酶、阿拉伯糖苷酶、芳基酯酶、 β -半乳糖苷酶、卡拉胶酶、过氧化氢酶、纤维二糖水解酶、纤维素酶、软骨素酶、角质酶、内切- β -1,4-葡聚糖酶、内切- β -甘露聚糖酶、酯酶、外切-甘露聚糖酶、半乳聚糖酶、葡糖淀粉酶、半纤维素酶、透明质酸酶、角蛋白酶、漆酶、乳糖酶、木质素酶、脂肪酶、脂氧合酶、甘露聚糖酶、氧化酶、果胶酸裂解酶、果胶乙酰酯酶、果胶酶、戊聚糖酶、过氧化物酶、酚氧化酶、磷酸酶、磷脂酶、植酸酶、多聚半乳糖醛酸酶、蛋白酶、普鲁兰酶、还原酶、聚鼠李糖半乳糖醛酸酶、 β -葡聚糖酶、鞣酸酶、转谷氨酰胺酶、木聚糖乙酰酯酶、木聚糖酶、木葡聚糖酶和木糖苷酶,以及它们的任意混合物);以及酶稳定剂组分(选自低聚糖、多糖和无机二价金属盐)。

[0222] 如上所述,通过由配制人员选择的任何方法将本发明的清洁组合物配制成任何合适的形式并制备本发明的清洁组合物,其非限制性示例描述于美国专利5,879,584、5,691,297、5,574,005、5,569,645、5,565,422、5,516,448、5,489,392、5,486,303、4,515,705、4,537,706、4,515,707、4,550,862、4,561,998、4,597,898、4,968,451、5,565,145、5,929,022、6,294,514和6,376,445中。在其中期望低pH清洁组合物的一些实施方案中,此类组合物的pH经由添加酸性材料诸如HCl来调节。

[0223] 本文所公开的清洁组合物可用于清洁部位(例如表面、物品、盘碟或织物)。通常,所述部位的至少一部分与以纯态或在洗涤液体中稀释的本发明清洁组合物的实施方案接触,并且然后任选地洗涤和/或冲洗所述部位。出于本发明的目的,“洗涤”包括但不限于洗涤和机械搅拌。在一些实施方案中,清洁组合物通常以在溶液中约500ppm至15,000ppm的浓度使用。当洗涤溶剂为水,水温的范围通常是从约5℃至约90℃,并且当部位包括织物时,水与织物的质量比通常为约1:1至约30:1。

[0224] 有利地包含本发明的丝氨酸蛋白酶多肽的代表性洗涤剂制剂包括见于

W02013063460,第78-152页中的洗涤剂制剂,并且特别地第94至152页的表格在此通过引用并入。通常将丝氨酸蛋白酶以按组合物的重量计0.00001%至10%酶蛋白的水平掺入到洗涤剂组合物中。在一些实施方案中,洗涤剂组合物包含按该组合物的重量计超过0.0001%、0.001%、0.01%或0.1%的丝氨酸蛋白酶。在一些实施方案中,洗涤剂组合物包含按该组合物的重量计少于1%、0.1%、0.01%或0.001%的丝氨酸蛋白酶。

[0225] 本发明提供用于清洁或洗涤需要清洁的物品或表面(例如,硬质表面)的方法,其包括但不限于,用于清洁或洗涤盘碟物品、餐具物品、织物物品、衣物洗涤物品、个人护理物品等的方法,以及用于清洁或洗涤硬质表面或软表面(例如,物品的硬质表面)的方法。

[0226] 在一些实施方案中,本发明提供用于清洁需要清洁的物品、物体或表面的方法,所述方法包括使物品或表面(或期望清洁的物品或表面的一部分)与至少一种本发明的丝氨酸蛋白酶多肽或本发明的组合物接触足够的时间和/或在适于和/或有效清洁物品、物体或表面到期望的程度的条件下接触。一些此类方法还包括用水冲洗所述物品、物体或表面。就一些此类方法而言,清洁组合物为盘碟洗涤剂组合物并且待清洁的物品或物体为盘碟物品或餐具物品。如本文所用,“盘碟物品”是一般用于提供或食用食品的物品。盘碟物品可以为但不限于例如碟、盘、杯、碗等等。如本文所用,“餐具”是更广义的术语,其包括但不限于,例如碟、餐具、刀、叉、勺、筷子、玻璃器具、罐、调味汁碟、酒器、服务用品等。“餐具物品”旨在包括这些或用于提供或食用食品的类似物品中任一种。就一些此类方法而言,清洁组合物为自动盘碟洗涤剂组合物或手洗盘碟洗涤剂组合物并且待清洁的物品或物体为盘碟物品或餐具物品。就一些此类方法而言,清洁组合物为衣物洗涤剂组合物(例如,粉末衣物洗涤剂组合物或液体衣物洗涤剂组合物),并且待清洁的物品为织物物品。在一些其它实施方案中,清洁组合物为衣物洗涤预处理组合物。

[0227] 在一些实施方案中,本发明提供用于分别清洁或洗涤任选地需要清洁或洗涤的织物物品的方法。在一些实施方案中,所述方法包括提供含有变体蛋白酶的组合物,其包括但不限于,织物或衣物洗涤清洁组合物和需要清洁的织物物品或衣物洗涤物品,并且在足够或有效清洁或洗涤织物物品或衣物洗涤物品到期望的程度的条件下,使所述织物物品或衣物洗涤物品(或期望清洁的物品的一部分)与组合物接触。

[0228] 在一些实施方案中,本发明提供用于清洁或洗涤任选地需要清洁的物品或表面(例如硬质表面)的方法,所述方法包括提供待清洁或洗涤的物品或表面并使所述物品或表面(或期望清洁或洗涤的物品或表面的一部分)与至少一种本发明的丝氨酸蛋白酶多肽或包含至少一种此类丝氨酸蛋白酶多肽的本发明组合物接触足够的时间和/或在足够或有效清洁或洗涤物品或表面至期望的程度的条件下接触。此类组合物包括但不限于,例如,本发明的清洁组合物或洗涤剂组合物(例如,手洗盘碟洗涤剂组合物、手洗盘碟洗涤清洁组合物、衣物洗涤剂或织物洗涤剂或衣物洗涤清洁组合物或织物清洁组合物、液体衣物洗涤剂、液体衣物洗涤清洁组合物、粉末衣物洗涤剂组合物、粉末衣物洗涤清洁组合物、自动盘碟洗涤剂组合物、衣物洗涤增效剂清洁组合物或洗涤剂组合物、衣物洗涤清洁添加剂、和衣物洗涤预去斑剂组合物等)。在一些实施方案中,将所述方法重复一次或多次,具体地讲在期望附加的清洁或洗涤时。例如,在一些情况下,所述方法任选地还包括使物品或表面保持与至少一种变体蛋白酶或组合物接触足够或有效清洁或洗涤物品或表面到期望程度的一段时间。在一些实施方案中,所述方法还包括用水和/或其它液体冲洗物品或表面。在一些实施

方案中,所述方法还包括使物品或表面与至少一种本发明的变体蛋白酶或本发明的组合物再次接触并使得物品或表面保持与至少一种变体蛋白酶或组合物接触足以清洁或洗涤物品或表面到期望程度的一段时间。在一些实施方案中,清洁组合物为盘碟洗涤剂组合物并且待清洁的物品为盘碟物品或餐具物品。在本发明方法的一些实施方案中,清洁组合物为自动盘碟洗涤剂组合物或手洗盘碟洗涤剂组合物并且待清洁的物品为盘碟物品或餐具物品。在所述方法的一些实施方案中,清洁组合物为衣物洗涤剂组合物并且待清洁的物品为织物物品。

[0229] 本发明还提供在自动洗碗机中清洁餐具或盘碟物品的方法,所述方法包括提供自动洗碗机,将足以清洁餐具或盘碟物品的量的包含至少一种本发明的丝氨酸蛋白酶多肽或本发明的组合物的自动盘碟洗涤组合物放入洗碗机中(例如,通过将组合物放入洗碗机中的适当的或提供的洗涤剂隔室或分配器中),将盘碟或餐具物品放入洗碗机中,并且操作洗碗机以便清洁餐具或盘碟物品(例如,按照制造商的说明)。在一些实施方案中,所述方法包括本文所述的任何自动盘碟洗涤组合物,其包含但不限于至少一种本文提供的丝氨酸蛋白酶多肽。自动盘碟洗涤组合物的用量可容易地根据制造商的说明或建议来确定,并且可采用包含至少一种本发明的变体蛋白酶的自动盘碟洗涤组合物的任何形式(例如液体、粉末、固体、凝胶、片剂等),包括本文所述的任一种。

[0230] 本发明还提供用于清洁任选地需要清洁的表面、物品或物体的方法,所述方法包括使物品或表面(或期望清洁的物品或表面的一部分)与以纯态或在洗涤液体中稀释的至少一种本发明的丝氨酸蛋白酶多肽或本发明的清洁组合物接触足够的时间和/或在足够或有效清洁或洗涤物品或表面到期望程度的条件下接触。然后根据需要,可(任选地)洗涤和/或冲洗表面、物品或物体。

[0231] 本发明还提供在洗衣机中清洁衣物或织物物品的方法,所述方法包括提供洗衣机,将足以清洁衣物或织物物品的量的包含至少一种本发明的丝氨酸蛋白酶多肽酶的衣物洗涤剂组合物放入洗衣机中(例如,通过将组合物放入洗衣机中的适当的或提供的洗涤剂隔室或分配器中),将衣物或织物物品放入洗衣机中,并且操作洗衣机以便清洁衣物或织物物品(例如,按照制造商的说明)。本发明的方法包括本文所述的任何衣物洗涤剂组合物,其包含但不限于本文所提供的任何丝氨酸蛋白酶多肽酶中的任一种。衣物洗涤剂组合物的用量可容易地根据制造商的说明或建议来确定,并且可采用包含本发明的至少一种变体蛋白酶的衣物洗涤剂组合物的任何形式(例如固体、粉末、液体、片剂、凝胶等),包括本文所述的任一种。

[0232] B. 纺织物处理

[0233] 还提供了使用本发明的丝氨酸蛋白酶多肽来处理织物(例如,以使纺织物退浆)的组合物和方法。织物处理方法在本领域中是熟知的(参见,例如美国专利6,077,316)。例如,可通过包括使织物与在溶液中的丝氨酸蛋白酶接触的方法来改善织物的触感和外观。可用该溶液在压力下处理织物。

[0234] 本发明的丝氨酸蛋白酶可在编织纺织物期间或之后施用,或者在退浆阶段期间施用,或者是一个或多个另外的织物处理步骤。在编织纺织物期间,使线暴露于相当大的机械应变。在于机械织机上进行编织之前,经常用施胶淀粉或淀粉衍生物包被经纱以增大其拉伸强度并防止断裂。本发明的丝氨酸蛋白酶可在编织期间或之后施用以除去这些施胶淀粉

或淀粉衍生物。在编织之后,可使用丝氨酸蛋白酶来除去浆料涂层(size coating),之后进一步处理织物以确保均相且耐洗的结果。

[0235] 本发明的丝氨酸蛋白酶可单独使用或者可作为洗涤剂添加剂与其它使织物(包括含棉织物)退浆的退浆化学试剂和/或退浆酶一起在例如含水组合物中使用。在用于对靛蓝染色的斜纹粗棉布织物和服装产生石磨洗外观的组合物和方法中还可使用淀粉酶。对于衣服的制作,可将织物剪切并缝制成衣服或服装,之后对其进行修整(finish)。特别地,对于粗斜纹棉布裤的制作,已开发了不同的酶促修整方法。斜纹粗棉布服装的修整通常以酶促退浆步骤起始,在此期间使服装经受蛋白水解酶的作用以赋予织物柔软性并且使棉织物更容易进行随后的酶促修整步骤。丝氨酸蛋白酶可用于修整斜纹粗棉布服装(例如,“生物石磨洗工艺”)、酶促退浆以及赋予织物柔软性和/或修整工艺的方法。

[0236] C. 皮革和毛处理

[0237] 本文中所述的丝氨酸蛋白酶多肽还可用于蛋白质从动物的酶辅助性去除及其后续降解或处理,诸如毛、皮肤、毛发、皮革等。在一些情况下,与传统地浸入烫水中或脱毛处理相比,将动物屠体浸入包含本发明丝氨酸蛋白酶多肽的溶液中可用于保护皮肤免受损害。在一些实施方案中,可在适用于消化或引发鸟类羽毛的降解的条件下,用本发明的分离的丝氨酸蛋白酶多肽对毛进行喷雾。在一些实施方案中,可如上所述将本发明的丝氨酸蛋白酶与氧化剂组合使用。

[0238] 在一些实施方案中,通过使用本发明的丝氨酸蛋白酶多肽来辅助除去与原料毛缔合的油或脂肪。在一些实施方案中,丝氨酸蛋白酶多肽用于清洁毛以及消毒和部分脱水纤维的组合物。在另一些实施方案中,所公开的丝氨酸蛋白酶多肽可用于从鸟类羽毛回收蛋白质。在一些其它实施方案中,将丝氨酸蛋白酶多肽与95%乙醇或其它极性有机溶剂在约0.5%(v/v)表面活性剂存在或不存在下组合施用于洗涤溶液。

[0239] D. 动物饲料

[0240] 在本发明的另一个实施方案中,本发明的丝氨酸蛋白酶多肽可用作包含丝氨酸蛋白酶及其变体的动物饲料组合物、动物饲料添加剂和/或宠物食品的组分。本发明还涉及用于制备这样的动物饲料组合物、动物饲料添加剂组合物和/或宠物食品的方法,该方法包括将丝氨酸蛋白酶多肽与一种或多种动物饲料成分和/或动物饲料添加剂成分和/或宠物食品成分混合。此外,本发明涉及丝氨酸蛋白酶多肽在制备动物饲料组合物和/或动物饲料添加剂组合物和/或宠物食品中的用途。

[0241] 术语“动物”包括所有的非反刍动物和反刍动物。在一个具体实施方案中,所述动物是非反刍动物,诸如马和单胃动物。单胃动物的示例包括但不限于猪和野猪,包括但不限于小猪、生长猪、母猪;家禽,诸如火鸡、鸭、鸡、肉鸡、蛋鸡(layer);鱼,诸如大麻哈鱼、鲑鱼、罗非鱼、鲶鱼和鲤鱼;以及甲壳动物,诸如虾和对虾。在另一个实施方案中,所述动物是反刍动物,包括但不限于牛、幼牛、山羊、绵羊、长颈鹿、野牛、驼鹿、麋鹿、牦牛、水牛、鹿、骆驼、羊驼、美洲驼、羚羊、叉角羚和蓝牛。

[0242] 在本发明的情形下,预期术语“宠物食品”应理解为意指用于家养动物的食品,所述家养动物诸如但不限于狗、猫、沙鼠、仓鼠、南美栗鼠、观赏型大鼠、豚鼠;禽类宠物,诸如金丝雀、长尾鹦鹉和鹦鹉;爬行动物类宠物,诸如龟、蜥蜴和蛇;以及水生类宠物,诸如热带鱼和青蛙。

[0243] 术语“动物饲料组合物”、“饲料”和“草料”可互换使用,并且可包含选自以下的一种或多种饲料材料:a)谷类,诸如小粒谷物(例如小麦、大麦、黑麦、燕麦以及它们的组合)和/或大粒谷物诸如玉蜀黍或高粱;b)来自谷类的副产物,诸如玉米谷蛋白粉、可溶性谷物干酒糟(DDGS)(特别是基于玉米的可溶性谷物干酒糟(cDDG))、麦麸、小麦粗粉、小麦次粉、米糠、稻壳、燕麦壳、棕榈仁和柑橘渣;c)由以下来源获得的蛋白质:诸如大豆、向日葵、花生、羽扇豆、豌豆、蚕豆、棉花、卡诺拉、鱼粉、干血浆蛋白质、肉和骨粉、马铃薯蛋白、乳清、干椰肉、芝麻;d)从植物和动物来源获得的油和脂肪;e)矿物质和维生素。

实施例

[0244] 提供了以下实施例以证明并举例说明本公开内容的某些优选实施方案和方面,但不应解释为限制性的。

[0245] 在下文的实验公开内容中,应用了以下缩写:ADW(自动盘碟洗涤);BMI(血液/乳/墨);BSA(牛血清白蛋白);CAPS(N-环己基-3-氨基丙磺酸);CHES(N-环己基-2-氨基乙磺酸);DMC(二甲基酪蛋白);HDD(重垢型干/粉末);HDL(重垢型液体);HEPES(4-(2-羟乙基)-1-哌嗪乙磺酸);MTP(微量滴定板);ND(未进行);OD(光密度);PCR(聚合酶链反应);ppm(百万分之一);QS(足量);rpm(转数/分钟);AAPF(琥珀酰基-Ala-Ala-Pro-Phe-对硝基苯胺);TNBSA(2,4,6-三硝基苯磺酸);v/v(体积/体积);w/v(重量/体积)。

[0246] 实施例1

[0247] 蛋白质测定方法

[0248] 通过免染成像仪Criterion进行蛋白质测定

[0249] 通过免染成像仪Criterion方法来量化蛋白质。该方法利用免染预制PAGE凝胶,其中每个带的强度取决于存在于感兴趣的蛋白中的色氨酸残基的量。用于PAGE的Criterion™TGX(Tris-甘氨酸延伸型)Stain-Free™预制凝胶包括独特的三卤代化合物。这允许利用Gel Doc™EZ成像系统对蛋白质进行快速荧光检测。在紫外线诱导的反应中三卤代化合物与色氨酸残基反应以产生荧光,其可通过Gel Doc EZ成像仪在凝胶内容易地检测到。用于测定的试剂:浓缩的(10x)Laemmli样品缓冲液(Kem-En-Tec,产品目录号42556);18或26孔Criterion TGX免染预制凝胶(Bio-Rad,产品目录号分别为567-8124和567-8125);以及蛋白标记物“Precision Plus Protein Standards”(Bio-Rad,产品目录号161-0363)。如下进行测定:将25μl蛋白质样品和25μl 0.5M HCL添加到冰上的96孔-PCR板中并持续10分钟以使蛋白酶失活并防止自水解。然后,将50μl的酸性蛋白质混合物添加到96孔-PCR板中的包含0.385mg DTT的50μL样品缓冲液中。所述板通过得自Bio-Rad的Microseal “B”膜密封并置于PCR机中并且加热至70℃并持续10分钟。此后,通过运行缓冲液来填充腔室并设置凝胶盒。然后,将10μL的每种样品连同标记物一起置于每个袋中。此后,在200V下开始电泳并持续35min。在电泳之后,将凝胶转移到成像仪中。Image Lab软件用于计算每个带的强度。通过已知蛋白质量和标准样品的色氨酸含量,可制得校准曲线。可通过带强度和色氨酸数到蛋白质浓度的外推来确定实验样品的量。将该蛋白质量化方法用于制备用于实施例中所示的测定的样品BspAG00296和BspM04033蛋白酶。

[0250] N端和C端氨基酸测定

[0251] 在准备序列确认时,可使分离蛋白酶的样品在10kDa旋转过滤器中经受一系列化

学处理。样品通过脲和DTT处理变性和还原。进行胍基化步骤以将赖氨酸转化成高精氨酸以避免赖氨酸侧链乙酰化。使用碘代乙酰胺的乙酰化反应然后仅改性蛋白质的N端残基。然后将样品与包含 ^{18}O 水的缓冲液混合,并且加入胰蛋白酶和胰凝乳蛋白酶用于消化。所得的肽将包含 ^{18}O 和 ^{16}O 的混合物,除了羧基末端将保留原生的 ^{16}O 之外。使用纳米-LC系统之后使用LTQ Orbitrap(Thermo Fisher)高分辨质谱仪分离消化产物并分析,然后从多肽的MS/MS片段谱和多肽的同位素模式推导氨基酸序列。

[0252] 在一些情况下,在通过LC-MS/MS分析之前,如下所述,在SDS-PAGE凝胶上运行蛋白质样品。在准备序列确认时,包括N端确定和C端确定时,如下所述使来自SDS-PAGE凝胶的蛋白质带经历一系列化学处理。在单独的化学处理之间,使用蒸馏水和乙醇洗涤凝胶片。通过DTT/碘代乙酰胺处理来还原/烷基化蛋白质。进行胍基化步骤以将赖氨酸转化成高精氨酸以避免赖氨酸侧链乙酰化。使用磺基-NHS-乙酯的乙酰化反应仅改性N端残基。用包含 ^{18}O : ^{16}O 水和用于蛋白质消化的胰凝乳蛋白酶或胰蛋白酶的缓冲液溶胀凝胶片。之后进行上文对于不需要胶内处理的样品所述的后续步骤。

[0253] 实施例2

[0254] 发现并识别丝氨酸蛋白酶BspAG00296

[0255] 将芽孢杆菌属(*Bacillus* sp.)1M5(Culture Collection Dupont)选择成可用于工业应用的酶的潜在来源。为识别由芽孢杆菌属1M5产生的酶和编码这些酶的基因,通过合成(SBS)技术使用Illumina[®]测序对芽孢杆菌属1M5的整个基因组进行测序。基因组测序和序列数据的组合通过BaseClear(Leiden, The Netherlands)进行。通过BioXpr(Namur, Belgium)来标注重叠群。以这种方法识别的菌株芽孢杆菌属1M5中的基因中的一个编码示出与各种其它细菌的丝氨酸蛋白酶同源的蛋白质。该基因的序列BspAG00296.n以SEQ ID NO:1示出。

[0256] SEQ ID NO:1示出BspAG00296.n基因的核苷酸序列:ATGAAGAAGTTCTTATGTCTGTCGGTGTGATGTTGGTTTTATCTGTGTTTTCTGGCAATGTGTTGGCGAATGATGAGGTCAAAAAGGAAGATTATGTTGACGGGCAGTTGATTGTTTCAGTGACGCAAGCTTTGACTCAAAAAGGGAAGCCGATGCTTCAAGCATTGACAAGCACCTCGAAGCTGTTGAATGCAGAATTGAAGAAAAACGGTTTTGAAGTAGCGGATTCGCTGCTGGAAGTGAAGGGAAATGATTCCGTCGATATTTTCAGCGACAGCTTTAAAGAGGAGGCAGCAAAAAATACCGGATTTGTTTACCTTGTAGAATATTCTACAGATGCTTATGCTTCCATCGATGATGCGAAGAAGGCGCTCGAAAAACAGTTAACGGACATCGGCTTAAAAGTAAATATGTCTCTGAAAACCTTTACAGTCGAGCTGTGCGCCGAAGCGGCTGAAGAGGTAATACAGCCGGCAATGCATGCTAATCAGCGCTGGCATTATGAAATGATTGCGGCGCCGCAAGCTTGGAATATTACGACCGGCAGCAGGAATGTTTCAATGGCGGTGCTTGATACAGGAATTGATTCATCACATCCGAACCTTAGCAAACCTTGTGAATACAAGCTTGGGGAGGAGCTTTGTCGGCGGAACGCCTGCTGATGTACACGGACATGGGACTCATGTTGCCGGTACGATTGCCAGCTACGGCTCCGTATCAGGTGTTATGCAAAACGCTACGCTTATTTCCGTAAAAGTATTGGATAACAGCGGCAGCGGCACAATTTATGGCATCCAGCAAGGCATTCTGTATGCCGCGAGCATTAAACGCCGATGTAATCAACATGTCCTTGGGAGGCGGCAGCTACAATCAAGGAATGAATGATGCGATTGAGACAGCCGTTAATTCGGAACAGTTGTCGTGGCTGCGTCAGGAAACAACGGGGCATCAAGCATTTCTACCCTGCCGCTTACAGCGGAGCGATTGCTGTGCGTTCCGTGACATCCAGCCGGACAAGATCAAGCTTCTCCAATATGGATCAGGCTTAGAGTTAATGGCTCCTGGCTCCAATATTTACAGCACATATCCAAACAGCCGGTAGTCCACGCTATCCGGAACATCAATGGCAACGCCGATGTTGCCGGGGTCGCCGGGTTAATCCGCTCGGTCAATCCTAATCTTTCCGCGGCGCAAGTAAGAACGATTTTGCGGAATACGGCTCAATACGCAGGCAGCTCCACGCAGTACGGCTAT

GGAATCGTCGATGCGTATGCTGCGGTACTCTCAGCCCGC.

[0257] 由BspAG00296.n基因编码的前酶原以SEQ ID NO:2示出。在N端处,蛋白质具有长度为23个氨基酸的信号肽,如由SignalP-NN所预测的(Emanuelsson等人,Nature Protocols(2007)2:953-971)。该信号肽序列在SEQ ID NO:2中加下划线并加粗。信号肽的存在指示该丝氨酸蛋白酶是分泌酶。所述酶具有预测为135个氨基酸的原序列。预测的全加工成熟链的序列(BspAG00296,274个氨基酸)以SEQ ID NO:3示出:

[0258] SEQ ID NO:2示出丝氨酸蛋白酶前体BspAG00296的氨基酸序列:
MKKFLCLSVLMLVLSVFSGNVLA***NDEVKKEDYVDGQLIVSVDA******SFDSKG KPMLQAL***
TSTSKLLNAELKKNGFEVADSLLEVKGNDSDIFSDSFKEEAAKNT
GFVYLVEYSTDAYASIDDAKKALEKQLTDIGLKVKYVSENFTVELSAEAAEEVI QP

AMHANQRWHYEMIRAPQAWNITTGSRNVRMAVLDTGIDSSHPNLANLVNTSLGRSFVGGTPADVHGHGTHVAGTIAS
YGSVSGVMQNATLISVKVLDNSGSGTIYGIQQGILYAASINADVIMSLGGGSYNQGMNDAIQTAVNSGTVVVAASG
NNGASSISYPAAYSGAIAVGSVTSSRTRSSFSNYSGLELMAPGSNIYSTYPNSRYATLSGTSMATPHVAGVAGLIR
SVNPNLSAAQVRTILRNATAQYAGSSTQYGYGIVDAYAAVLSAR

[0259] SEQ ID NO:3示出预测的成熟蛋白酶BspAG00296(274个氨基酸)的氨基酸序列:
AMHANQRWHYEMIRAPQAWNITTGSRNVRMAVLDTGIDSSHPNLANLVNTSLGRSFVGGTPADVHGHGTHVAGTIAS
YGSVSGVMQNATLISVKVLDNSGSGTIYGIQQGILYAASINADVIMSLGGGSYNQGMNDAIQTAVNSGTVVVAASG
NNGASSISYPAAYSGAIAVGSVTSSRTRSSFSNYSGLELMAPGSNIYSTYPNSRYATLSGTSMATPHVAGVAGLIR
SVNPNLSAAQVRTILRNATAQYAGSSTQYGYGIVDAYAAVLSAR。

[0260] 实施例3

[0261] BspAG00296的异源表达

[0262] 使用由下列组成的表达盒在枯草芽孢杆菌中产生BspAG00296蛋白酶:枯草芽孢杆菌aprE启动子、枯草芽孢杆菌aprE信号肽序列、原生BspAG00296蛋白酶前肽、成熟BspAG00296蛋白酶和BPN'终止子。该盒克隆到pHYT复制穿梭载体中并转化成合适的枯草芽孢杆菌菌株。pHYT载体通过使用BstEII和EcoRI位点在抗四环素基因之后添加终止子而衍生自pHY300PLK(Takara)(终止子序列,GGTTACCTTGAATGTATATAAACATTCTCAAAGGATTCTAATAAAAAACGCTCGGTTGCCGCCGGCGTTTTTATGCATCGATGGAATTC)(SEQ ID NO:45)。pHY300PLK中的HindIII位点也使用克隆到BamHI和HindIII位点中的接头除去(新的接头序列,GGATCCTGACTGCCTGAGCTT)(SEQ ID NO:46)。

[0263] 包含BspAG00296基因的pHYT载体(pHYT-BspAG00296)的图谱BspAG00296示于图1中。

[0264] 为制备BspAG00296,在富集半组合培养基中培养包含pHYT-BspAG00296的枯草芽孢杆菌转化体,所述培养基基于MOP缓冲液,其中脲作为主要氮源,葡萄糖作为主碳源,并且补充有1%大豆胨用于稳健的细胞生长。该培养基补充有25ppm四环素。在温育之后(32℃下2天),在生长介质中检测到BspAG00296蛋白酶。在离心和过滤之后,将具有BspAG00296蛋白酶的培养上层清液用于测定和纯化。

[0265] 如实施例1所述分析BspAG00296蛋白质的样品。最显著的蛋白质的序列(约28kDa)测定为对应于SEQ ID NO:4中列出的序列。

[0266] SEQ ID NO:4示出成熟蛋白酶BspAG00296(273个残基)的主要形式的氨基酸序列:

MHANQRWHYEMIRAPQAWNITTGSRNVRMAVLDTGIDSSHPNLANLVNTSLGRSFVGGTPADVHGHGTHVAGTIASY
GSVSGVMQNATLISVKVLDNSGSGTIYGIQQGILYAASINADVINMSLGGGSYNQGMNDAIQTAVNSGTVVVAASGN
NGASSISYPAAYSGAIAVGSVTSRTRSSFSNYGSGLELMAPGSNIYSTYPNSRYATLSGTSMATPHVAGVAGLIRS
VNPNLSAAQVRTILRNTAQYAGSSTQYGYGIVDAYAAVLSAR。

[0267] 实施例4

[0268] 发现并识别丝氨酸蛋白酶BspM04033

[0269] 将芽孢杆菌属(*Bacillus* sp.)WDG290(Culture Collection Dupont)选择成可用于工业应用的酶的潜在来源。为识别由芽孢杆菌属WDG290产生的酶和编码这些酶的基因,通过合成(SBS)技术使用Illumina®测序对芽孢杆菌属WDG290的整个基因组进行测序。基因组测序和序列数据的组合通过BaseClear(Leiden, The Netherlands)进行。通过BioXpr(Namur, Belgium)来标注重叠群。以这种方法识别的菌株芽孢杆菌属WDG290中的基因中的一个编码示出与各种其它细菌的丝氨酸蛋白酶同源的蛋白质。该基因的序列BspM04033.n以SEQ ID NO:5示出。

[0270] SEQ ID NO:5示出BspM04033.n基因的核苷酸序列:ATGGAGGAGAAAAATGTGAAAAAAG
TGCAGTTTGGGTCCTTATGACGGTGTGGTTTTTCAGTCTGTTTTTAAATCCTGCCGAATTGGCGCGCAGGCCTCTG
ATGCAGCTTCAGAAAAAGATGACACTGCCTACATAGAGGGGCAGTTGATTGTATCGGTAAAGAGCAGTGACGTTTCA
GTGAAGGGAATCGAAGGGGTAAACAAGAAGATCATGGGCGATGTCCTGAGAGAACGGGGATTCCGCATAACGGATTCC
TATTATGGGACTCGGCGATCCTGCTGAAGTGAATGCCTTTACGAACCAGGAGTTCAGTGAATCCGTCGTGAAGAATA
TGGGGCTCGTTTACCTTGCAGAATACGATGTGTCTGTTTATGCATCAGTAGAAGAAGCGAAACGGGAGCTGGCCGAA
GCGCTCAAAGAGAACGGAATGGAAATCAGACACATCTCGAAGAACTATGAAATGCACGCGATCGGGGAACCTGCCGA
TGTCTCTCCCCAGATGCACCCGAACAGCAGTGGCATTACAACATGATTAATGCACCGCAGGCGTGGGGGACAACGA
CAGGCTCCTCAAGTGTCATTACAGGCTGTGCTTGATACGGGGATTGACCACAATCATCAGAGTCTCGAAACTTAGTA
AACACAAGTCTCGGACAGAGCTTTGTGGGCGGAAGTACGATGGATGTTCAAGGGCACGGAACGCACGTTGCCGGTAC
GATTGCAAGCTACGGTTCTGTGTCCGGCGTGATGCACAATGCTACGCTCGTACCGGTTAAAGTGCTGAATGACAGTG
GATCAGGGTCACTTTTCGGCATTACGCAGGGAATCCTGTATTACAGCTGATATCGGGGCCGACGTGATCAACATGTCT
CTTGGCGGCGGCGGTTACAACCAGAGTATGGCAGAAGCTGCACAGACAGCGGTAAATGCCGGTTCGATTGTAATTGC
GGCAAGCGGAAATGACGGAGCGGGCAGTATTTTCGTATCCGGCAGCGTACAGCAGCGTCATTGCGGTTGGGTCTGTAA
CCTCGACAGGTGCCGTTCCAACCTCTCAAACTACGGCAGCGGACTTGAAGTATGGCACCTGGTTCAAATATTTAC
AGCACCGTACCGAATAACGGCTATGCCACATTCTCGGTACGTCGATGGCATCCCCGCATGCAGCAGGTGTTGCCGG
TCTGATGAGAGCGGTCAATCCGAATCTATCGGTATCGAATGCCAGATCGATTATGCAGAACACGGCTCAGTATGCCG
GAAGCCCAGCTTTCTACGGGTACGGGATCGTTGACGCGAACGCAGCGGTTACAGCAGGCATCAGGGGGAAGCGGCGGT
CCTTCCAATATTACTGAAACGAGTATATCCACTGACCGTTTCTATGTGCAGCGAGGTCAGAACGTGACGTCAACTGC
TCAGGTTACGAATGAAAACGGACAGGGTCTTGCCAACGCGACGGTGACCTTACCATACCCGTCCTCAACGGATCAA
CGCTTACGAATACAGCAACGACCAACAGTTCCGGTTTTCGCCTCATGGACGGTCGGCACATCCGGTGCCACCGCAACA
GGCACCTATTACGTAGAAGCATCATCTTCTTTCAGGGGTATCAGGGAAGTCCGCTTCAACGAGTTTCTTTGTTTA
C.

[0271] 由BspM04033.n基因编码的前酶原以SEQ ID NO:6示出。在N端处,蛋白质具有长度为33个氨基酸的信号肽,如由SignalP-NN所预测的(Emanuelsson等人, Nature Protocols (2007)2:953-971)。该信号肽序列在SEQ ID NO:6中加下划线并加粗。信号肽的存在指示该

丝氨酸蛋白酶是分泌酶。所述酶具有预测为133个氨基酸的原序列(该预测基于类芽孢杆菌枯草杆菌蛋白酶中的预成熟接合:W02012175708,SEQ ID NO:6)。预测的经加工成熟链的序列(BspM04033,382个氨基酸)以SEQ ID NO:7示出。

[0272] SEQ ID NO:6示出丝氨酸蛋白酶前体BspM04033的氨基酸序列:

*MEEKNVKKS AVVVLMTVLVFSLFLNPAGIGAQA SDAASEKDDTAYIEG
QLIVSVKSSDVS VKGIEGVNKKIMGDVLRERGF AITDSIMGLGDP AEVNAFTN
QEFSES VVKNMGLVYLA EYDVS VYASVEEAKRELAEALKENGMEIRHISKNYE
MHAIGEPADVSP*

QMHPNQWHYMNINAPQAWGTTTGSSSVIQA VLDTGIDHNHQLANLVNTSLGQSFVGGSTMDVQGHGTHVAGT IAS
YGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSM AEAQTAVNAGSIVIAASG
NDGAGSISYP AAYSSVIAVGSVTSTGARSNFSNYGSGLEL MAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMR
AVNP NLSVSNARSIMQNTAQYAGSPTFYGYGIVDANA AVQQASGGSGGPSNITETSISTDRFYVQRGQNV TSTAQVT
NENGQGLANATVTFTITRPN GSTLTNTATTNSSGFASWTVGTSGATATGTYSVEASSSLQGYQGSSASTSFFVY.

[0273] SEQ ID NO:7示出预测的成熟蛋白酶BspM04033(382个氨基酸)的氨基酸序列:

QMHPNQWHYMNINAPQAWGTTTGSSSVIQA VLDTGIDHNHQLANLVNTSLGQSFVGGSTMDVQGHGTHVAGT IAS
YGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSM AEAQTAVNAGSIVIAASG
NDGAGSISYP AAYSSVIAVGSVTSTGARSNFSNYGSGLEL MAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMR
AVNP NLSVSNARSIMQNTAQYAGSPTFYGYGIVDANA AVQQASGGSGGPSNITETSISTDRFYVQRGQNV TSTAQVT
NENGQGLANATVTFTITRPN GSTLTNTATTNSSGFASWTVGTSGATATGTYSVEASSSLQGYQGSSASTSFFVY.

[0274] 实施例5

[0275] BspM04033的异源表达

[0276] 使用由下列组成的表达盒在枯草芽孢杆菌中产生BspM04033蛋白酶:枯草芽孢杆菌aprE启动子、枯草芽孢杆菌aprE信号肽序列、原生BspM04033蛋白酶前肽、成熟BspM04033蛋白酶和BPN'终止子。将该盒克隆到基于pBN的复制穿梭载体中(Babe等人,(1998), Biotechnol. Appl. Biochem. 27:117-124)并且使用质粒转化合适的枯草芽孢杆菌菌株。

[0277] 包含BspM04033基因的pBN载体(pBN-BspM04033)的谱图示于图2中。

[0278] 质粒pBN-BspM04033中的BspM04033基因的核苷酸预成熟序列以SEQ ID NO:8示出:TCTGATGCAGCTTCAGAAAAAGATGACACTGCCTACATAGAGGGGCAGTTGATTGTATCGGTAAAGAGCAGTGA
CGTTTCAGTGAAGGGAATCGAAGGGGTAAACAAGAAGATCATGGGCGATGTCCTGAGAGAACGGGGATTCCGCATAA
CGGATTCTATTATGGGACTCGGCGATCCTGCTGAAGTGAATGCCTTTACGAACCAGGAGTTCAGTGAATCCGTCTGTG
AAGAATATGGGGCTCGTTTACCTTGCAGAATACGATGTGTCTGTTTATGCATCAGTAGAAGAAGCGAAACGGGAGCT
GGCCGAAGCGCTCAAAGAGAACGGAATGGAAATCAGACACATCTCGAAGAACTATGAAATGCACGCGATCGGGGAAC
CTGCCGATGTCTCTCCCCAGATGCACCCGAACCAGCAGTGGCATTACAACATGATTAATGCACCGCAGGCGTGGGGG
ACAACGACAGGCTCCTCAAGTGTCAATTCAGGCTGTGCTTGATACGGGGATTGACCACAATCATCAGAGTCTCGAAA
CTTAGTAAACACAAGTCTCGGACAGAGCTTTGTGGGCGGAAGTACGATGGATGTTCAAGGGCACGGAACGCACGTTG
CCGGTACGATTGCAAGCTACGGTTCTGTGTCCGGCGTGATGCACAATGCTACGCTCGTACCGGTTAAAGTGCTGAAT
GACAGTGGATCAGGGTCACTTTTCGGCATTACGCAGGGAATCCTGTATTCAGCTGATATCGGGGCCGACGTGATCAA
CATGTCTCTTGGCGGCGGCGGTTACAACCAGAGTATGGCAGAAGCTGCACAGACAGCGGTAAATGCCGGTTCGATTG

TAATTGCGGCAAGCGGAAATGACGGAGCGGGCAGTATTTTCGTATCCGGCAGCGTACAGCAGCGTCATTGCGGTTGGG
TCTGTAACCTCGACAGGTGCCCGTTCCAACCTCTCAAACCTACGGCAGCGGACTTGAAGTATGGCACCTGGTTCAAA
TATTTACAGCACCGTACCGAATAACGGCTATGCCACATTCTCGGGTACGTCGATGGCATCCCCGCATGCAGCAGGTG
TTGCCGGTCTGATGAGAGCGGTCAATCCGAATCTATCGGTATCGAATGCCAGATCGATTATGCAGAACACGGCTCAG
TATGCCGGAAGCCCGACTTTCTACGGGTACGGGATCGTTGACGCGAACGCAGCGGTTTACAGCAGGCATCAGGGGGAAG
CGGCGGTCTTCCAATATTACTGAAACGAGTATATCCACTGACCGTTTCTATGTGCAGCGAGGTCAGAACGTGACGT
CAACTGCTCAGGTTACGAATGAAAACGGACAGGGTCTTGCCAAACGCGACGGTGACCTTACCATCACCCGTCCAAAC
GGATCAACGCTTACGAATACAGCAACGACCAACAGTTCCGGTTTCGCCTCATGGACGGTCGGCACATCCGGTGCCAC
CGCAACAGGCACCTATTAGTAGAAGCATCATCTTCTCTTCAGGGGTATCAGGGAAGTTCCGCTTCAACGAGTTTCT
TTGTTTAC.

[0279] 由质粒pBN-BspM04033表达的BspM04033前体蛋白质的氨基酸序列以SEQ ID NO:9示出,其中预测的前肽以下划线文本示出:SDAASEKDDTAYIEGQLIVSVKSSDVS VKGIEGVNKKIMGD
VLRRERGAITDSIMGLGDAEVNAFTNQEFSES VVKMGLVYLA EYDVS VYASVEEAKRELAELKENGMEIRHISK
NYEMHAIGEPADVSPQMHPNQWHYNNMINAPQAWGTTTGSSSVIQA VLDTGIDHNHQLANLVNTSLGQSFVGGSTM
DVQGHGTHVAGTIASYGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSMAEAA
QTAVNAGSIVIAASGNDGAGSISYPAAYSSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGT
SMASPHAAGVAGLMRAVNP NLSVSNARSIMQNTAQYAGSPTFYGYGIVDANA AAVQQASGGSGGPSNITETSISTDRF
YVQRGQNVSTAQVTNENGQGLANATVTFTITRPNGSTLTNTATTNSSGFASWTVGTSGATATGTYSVEASSSLQGY
QGSSASTSFFVY.

[0280] 为制备BspM04033,将包含pBN-BspM04033的枯草芽孢杆菌转化体在具有10ppm新霉素的TSB(液体培养基)中在15ml Falcon管中培养16小时,并且将300μl该预培养物添加到填充有30mL培养基(下文所述的)的500mL烧瓶中,所述培养基补充有10ppm新霉素。在以180rpm恒定旋转混合的情况下,将烧瓶在32℃下温育48小时。通过在锥形管中在14500rpm下离心20分钟收获培养物。将培养上层清液用于测定。所述培养基为基于MOP缓冲液的富集半组合培养基,其中脲作为主要氮源,葡萄糖作为主碳源,并且补充有1%大豆脲用于稳健的细胞生长。

[0281] 如实施例1所述分析BspM04033蛋白质的样品。所述样品包含两种主要形式的酶,一种为约44kDa并且另一种为约28kDa。较大的蛋白质对应于如以SEQ ID NO:10示出的缺乏原序列和后续残基Q1(Gln)的全长经处理蛋白酶区域。SEQ ID NO:10示出观察到的全长BspM04033蛋白质,其由质粒pBN-BspM04033(381个氨基酸)表达:MHPNQWHYNNMINAPQAWGTTTGSSSVIQA VLDTGIDHNHQLANLVNTSLGQSFVGGSTM DVQGHGTHVAGTIASYGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSMAEAAQTAVNAGSIVIAASGNDGAGSISYPAAYSSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMRAVNP NLSVSNARSIMQNTAQYAGSPTFYGYGIVDANA AAVQQASGGSGGPSNITETSISTDRFYVQRGQNVSTAQVTNENGQGLANATVTFTITRPNGSTLTNTATTNSSGFASWTVGTSGATATGTYSVEASSSLQGYQGSSASTSFFVY.

[0282] 在样品储存时最显著的蛋白质样品的序列(约28kDa)测定为对应于C端截短形式的以SEQ ID NO:11列出的序列。该多肽缺乏在预测的前区域之后的Gln1,并且在长度上与肽酶S8家族结构域一致。

[0283] SEQ ID NO:11示出观察到的蛋白酶BspM04033(276个氨基酸)的主要形式的氨基

酸序列:MHPNQWHYNMINAPQAWGTTTSSSVIQAVLDTGIDHNHQLANLVNTSLGQSFVGGSTMDVQGHGTHV
AGTIASYGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADVINMSLGGGGYNQSMASMAEAAQTAVNAGSI
VIAASGNDGAGSISYPAAYSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAG
VAGLMRAVNPNSVSNARSIMQNTAQYAGSPTFYGYGIVDANAQVQASGGS。

[0284] 实施例6

[0285] 发现并识别丝氨酸蛋白酶BspW01765

[0286] 将芽孢杆菌属(*Bacillus* sp.)SWT211(Dupont Culture Collection)选择成可用于工业应用的酶的潜在来源。为识别由芽孢杆菌属SWT211产生的酶和编码这些酶的基因,通过合成(SBS)技术使用Illumina®测序对芽孢杆菌属SWT211的整个基因组进行测序。基因组测序、序列数据的组合和注释通过BaseClear(Leiden, The Netherlands)进行。以这种方法识别的SWT211中的基因中的一个编码示出与各种其它细菌的丝氨酸蛋白酶同源的蛋白质。该基因的序列BspW01765.n以SEQ ID NO:12示出。

[0287] SEQ ID NO:12示出BspW01765.n基因的核苷酸序列:ATGAAGAAGTTATTTACCTTGTTTT
TATTGACACTTGTAATGCTTGTTGGGGTTATTTTCTGTAAATGTCATGGCAGATAATGAGGAAGAAAAAGAAGACCAT
AAGTACATTGAAGGTCAATTAATCGTATCGGTAGAACCGGATGCAATGATAACTCAATAGGACAAATGAATATCAC
CTCAGATAAATTACAAAATAACTCCTCTCTAAAGAATAAAGGATTTAAATAGCAGATTCTTTATTGGAACAATA
CTCCTGGTGTTCAAAGTATATTGAGCAGTAGCTTTGTACAAGATGCTGCGAAAAGAACAGGGCTCGTTTACCTCATA
GAATATTCCCAGAAAAATTTGAATCCATTGAGGCAGCAAAAAAGACCTTGAAAAACCTTAACAGAACTTGGATT
TAATGTGAGATATGTTTCAGAAAACTTTGTTGTTGAGCTTTTAGAGACAGAAGCTACCTCAGATACTGGTGAAGATA
TCATCACGCCATTTATGCACAGTAATCAAGAATGGCATTACGGCATGATTAATGCCCTGATGCTTGGGGTATTACT
ACAGGTGACAGTAATGTAACAATAGCAGTATTGGATACTGGAATAGATTCTAGCCATTCAAGTTTAAGTAACTTAGT
AGATACTAGTCTTGGAAGAAGCTATGTTGGTGGTTCTCCAGAGGATGTTCAAGGTCATGGAACGCACGTAGCAGGTA
CGATAGCAAGCTATGGTGCAGTATCGGGTGTCATGCAGGATGCAACACTCATTCTGTCAAAGTTTTAGGTGATGAT
GGAAGTGGGTCAATGTATGGCATAACAAGGAGTTTTATATGCTACAAGTATTGGTGCAGACGTCATTAATATGTC
TTTAGGCGGAGGCGGTTATAATCAAGGTTTCAATGATGCTATTGATACAGCAGTTGCGAATGGATCAGTTGTAATTG
CTGCTTCTGGTAATGATGGTAGAGCTTCTATTTCTATCCAGCAGCTTATGATGGAGCAATTGCAGTTGGGTGAGTA
ACTTCTAGTGGTAATCGCTCAAACCTTCTAACTATGGAAGTGGTCTTGAGTTAATGGCACCAGGATCAAGTATCTA
CAGCACCTATCCTAATGGTCAGTACAGAACGTTATCAGGTACATCTATGGCAGCTCCACATGCTGCAGGTGTTGCAG
GACTAGTACGGGCAGTAAATCCGAACCTTGTCAGTAGCAGAAAGTGAGAAACATATTAGCGGATACAGCACAATATGCA
GGTAGTTCTCATCAGTATGGAACGGTATTGTAGATGCTTTTGCAGCGGTTCAAGCAGCAGGTGGATCTGGTGGAAC
ACCATCACCTGGTGTACGAATACAGTTGTTTCAACAGATAAAAGTGTTTATGAGCGTGGTGAGCAAGTAACGATGA
CAACAACCTGTTACAGATGAAGGCGGTAATGCTCTTCAAGACGCTACAGTTAATTACACAATTACACGTCCAAATGGA
TCTACTGTAACAAATACAACAACCTACAAATTCAAATGGAATTGCAACGTGGATAATTGGATCTAATTCACAACTGC
TTTAGGGACTTACGATGTGACGGCAGAACTAGTCTATCAGGCTATCAAACCTAGCTCTGATACTACTTCCTTTAGCT
TCTCTGATCAAGCACAGACCCAAACAAACAGTAACGGATGTTTCAACGAATAGTAGCTATTATGCACGTGGTCAGAAT
GTAACCATATCAGCTGAAGTGAAGGATCAAGATGGAGAGGCCCTATCAAATGCTACGGTTTCTTTTACAATTATCAG
ACCAAATGGAAGTACGTTGACGAATACAGCTACAACCTAATAGCGCAGGTGTGGCCACTTGACTGTATCAACGAGTA
GTGGAAGTGAAGAGGGACATATGAAGTAACTGCAGAGTCTTCTTACTCTACTTATGATGGAAGTTCAGATACCACA
ATCTTTTATGTTTAT.

[0288] 由BspW01765.n基因编码的前酶原以SEQ ID NO:13示出。在N端处,蛋白质具有长度为25个氨基酸的信号肽,如由SignalP-NN所预测的(Emanuelsson等人,Nature Protocols(2007)2:953-971)。该信号肽序列在SEQ ID NO:13中加下划线并加粗。信号肽的存在指示该丝氨酸蛋白酶是分泌酶。所述酶具有预测(基于博戈里亚芽孢杆菌(*Bacillus bogoriensis*)蛋白酶的预成熟接头:参见W02012175708,SEQ ID NO:4)为142个氨基酸(在SEQ ID NO:13中呈斜体形式)的原序列。预测的经加工成熟链的序列(BspW01765,488个氨基酸)以SEQ ID NO:14示出。BspW01765的成熟链由N端处的肽S8家族结构域和在C端处具有未知功能的结构域组成。BspW01765蛋白酶的催化的、肽酶结构域的序列预测为270个氨基酸并且以SEQ ID NO:15示出。

[0289] SEQ ID NO:13示出预测的丝氨酸蛋白酶前体BspW01765的氨基酸序列:

MKKLFTLFLLLTMLVGLFSVNVMA***DNEEEKEDHKYIEGQLIV***
SVEPDANDNSIGQMNITSDKLQNNSSLKNKGFKIADSLLENTPGVQSIFSSSF
VQDAAKRTGLVYLIEYSPEKFE***SIQAAKKDLEKTLTELGFNVRYVSENFVVELL***
ETEATSDTGEDIITPFMHSNQEWHYGMINAPDAWGITTGDSNVTIAVLDTGIDSSHSSLSNLVDTSLGRSY
VGGSPEDVQGHGTHVAGTIASYGAVSGVMQDATLISVKVLGDDGSGSMYGIQQGVLYATSIGADV INMSLGGGGYNQ
GFNDAIDTAVANGSVVIAASGNDGRASISYPAAYDGAIAVGSVTSSGNRSNFSNYGSGLELMAPGSSISTYPNGQY
RTLSGTSMAPHAAGVAGLVRVNPNSVAEVRN ILADTAQYAGSSHQYGN GIVDAFAAVQAAGSGGTPSPGVTNT
VVSTDKSVYERGEQVTMTT TTTVTDEGGNALQDATVNYTITRPNGSTVTNTTTTNSNGIATWII GSNSTALGTYDVTA
ETSLSGYQTSSD TT SFSDQAQTQQTVTDVSTNSSYYARGQNV TISAEVKDQDGEALSNATVSFTIIRPNGSTLTN
TATTNSAGVATWTVSTSSGTARGTYEVTAESSYSTYDGSDDTTIFVYVY.

[0290] SEQ ID NO:14示出预测的成熟蛋白酶BspW01765(382个氨基酸)的氨基酸序列:

MHSNQEWHYGMINAPDAWGITTGDSNVTIAVLDTGIDSSHSSLSNLVDTSLGRSYVGGSPEDVQGHGTHVAGTIASY
GAVSGVMQDATLISVKVLGDDGSGSMYGIQQGVLYATSIGADV INMSLGGGGYNQGFNDAIDTAVANGSVVIAASGN
DGRASISYPAAYDGAIAVGSVTSSGNRSNFSNYGSGLELMAPGSSISTYPNGQYRTLSGTSMAPHAAGVAGLVR
VNPNSVAEVRN ILADTAQYAGSSHQYGN GIVDAFAAVQAAGSGGTPSPGVTNTVVSTDKSVYERGEQVTMTT TTTVT
DEGGNALQDATVNYTITRPNGSTVTNTTTTNSNGIATWII GSNSTALGTYDVTAETSLSGYQTSSD TT SFSDQA
QTQQTVTDVSTNSSYYARGQNV TISAEVKDQDGEALSNATVSFTIIRPNGSTLTNTATTNSAGVATWTVSTSSGTAR
GTYEVTAESSYSTYDGSDDTTIFVYVY.

[0291] SEQ ID NO:15示出BspW01765蛋白酶(270个氨基酸)的肽酶S8家族结构域的预测

的氨基酸序列:MHSNQEWHYGMINAPDAWGITTGDSNVTIAVLDTGIDSSHSSLSNLVDTSLGRSYVGGSPEDVQ
GHGTHVAGTIASYGAVSGVMQDATLISVKVLGDDGSGSMYGIQQGVLYATSIGADV INMSLGGGGYNQGFNDAIDTA
VANGSVVIAASGNDGRASISYPAAYDGAIAVGSVTSSGNRSNFSNYGSGLELMAPGSSISTYPNGQYRTLSGTSM
APHAAGVAGLVRVNPNSVAEVRN ILADTAQYAGSSHQYGN GIVDAFAAVQ.

[0292] 实施例7

[0293] BspAG00296和BspM04033的蛋白酶活性

[0294] BspAG00296、BspM04033蛋白酶的蛋白酶活性通过测量二甲基酪蛋白(DMC)底物的水解来测试。用于DMC测定的试剂溶液是:在100mM碳酸钠pH 9.5中的2.5%二甲基酪蛋白(DMC,Sigma)、在试剂A中的0.075% TNBSA(2,4,6-三硝基苯磺酸,Thermo Scientific)。试剂A:在15mL 4N NaOH中的45.4g Na₂B₄O₇·10H₂O(Merck),在MQ水中达到1000mL的终体积,稀

释溶液:10mM NaCl、0.1mM CaCl₂、0.005% Tween-80。将蛋白酶上层清液在稀释溶液中稀释至用于测定的适当浓度。96孔微量滴定板(MTP)用95μl DMC底物填充,之后添加5μl经稀释的蛋白酶上层清液。然后在缓慢搅拌下添加100μL的TNBSA在试剂A中的溶液。使用SpectraMax读板仪在RT下以动力学模式于405nm处经5分钟测量活性。从数值中减去不包含蛋白酶的空白的吸光度。活性表达为mOD/min。BspAG00296的蛋白酶活性曲线示于图3中并且BspM04033的蛋白酶活性曲线示于图4中。使用DMC测定,发现BspAG00296蛋白酶的比活性为56mOD/min/ppm,并且发现BspM04033蛋白酶的比活性为71mOD/min/ppm。在相同测定条件下,发现GG36和BPN'蛋白酶的比活性分别为54mOD/min/ppm和23mOD/min/ppm。

[0295] 实施例8

[0296] BspAG00296和BspM04033蛋白酶的pH曲线

[0297] 使用偶氮酪蛋白作为底物,在包含50mM CaCl₂的50mM乙酸酯/Bis-Tris/HEPES/CHES缓冲液中,研究BspAG00296和BspM04033蛋白酶的蛋白水解活性的pH依赖性。以1pH单位增量在介于4至12之间的pH下测量活性。将一个Protaxyme AK片剂(Megazyme, Ireland)连同1.9mL的适当缓冲液和磁体一起加入玻璃测试管中,之后在安装有磁力搅拌器的温控水浴中在40℃下温和水合5分钟。将新鲜制备的蛋白酶的100微升样品(在去离子水中稀释至用于测定的适当浓度)加入预水合的底物中并且使反应在40℃下进行10分钟。为终止反应,加入10mL pH 12的2%w/v Tris缓冲液,将溶液混合,并且将样品立即通过Whatman 1号过滤器过滤。收集上层清液,并且测量上层清液在590nm下的吸光度以量化反应的产物。减去只有缓冲液的对照的吸光度,并通过将最佳pH下的活性限定为100%来将所得值转化为相对活性的百分比。在该测定的条件下,确定BspAG00296在6-12的pH范围内保持≥50%的活性,并且确定BspM04033在7-12的pH范围内保持≥50%的活性。

[0298] 实施例9

[0299] BspAG00296和BspM04033蛋白酶的溫度曲线

[0300] 使用偶氮酪蛋白作为底物,在包含50mM CaCl₂的50mM乙酸酯/Bis-Tris/HEPES/CHES缓冲液中在pH 9下,研究BspM04033蛋白酶的蛋白水解活性的溫度依赖性。以10℃增量在介于30℃和80℃之间的溫度下测量活性。将一个Protaxyme AK片剂(Megazyme, Ireland)连同1.9mL的适当缓冲液和磁体一起加入玻璃测试管中,之后在安装有磁力搅拌器的温控水浴中在设定溫度下温和水合5分钟。将新鲜制备的蛋白酶的100μl样品(在去离子水中稀释至用于测定的适当浓度)加入预水合的底物中并且使反应在介于30℃和80℃之间的溫度下进行10分钟。为终止反应,加入10mL pH 12的2%w/v Tris缓冲液,并将溶液混合,并且立即通过Whatman 1号过滤器过滤。收集上层清液,并且测量上层清液在590nm下的吸光度以量化反应的产物。从每个样品读数中减去只有缓冲液的对照的吸光度,并通过将最佳溫度下的活性限定为100%来将所得数值转化为相对活性的百分比。在该测定的条件下,确定BspAG00296在55-75℃的范围内保持≥50%的活性,并且确定BspM04033在55-80℃的范围内保持≥50%的活性。

[0301] 实施例10

[0302] BspAG00296和BspM04033的清洁性能

[0303] BspAG00296和BspM04033蛋白酶的清洁性能,就基于衣物洗涤的应用而言,对BMI(棉上的血液/奶/墨)微样本(EMPA-116, Center for Testmaterials, The Netherlands)进

进行测试,并且就基于盘碟的应用而言,对卵黄(老化并用炭黑染料着色的聚丙烯酰织物上卵黄)微样本(PAS-38,Center for Testmaterials,The Netherlands)进行测试。在酶添加之前,包含预穿孔(以安装在MTP上)和预冲洗样本的MTP(Corning 3641)用洗涤剂填充。将商业洗涤剂加热灭活以除去酶并如表1所述投料。

[0304] 通过在水浴中加热至95℃并持续4小时使重垢型液体(HDL)衣物洗涤剂失活。通过制备10%w/v溶液并且在95℃下加热4小时使重垢型干燥(HDD)衣物洗涤剂失活。在将HDD和HDL洗涤剂加热4小时之后,蛋白酶活性不存在。在灭活处理之后,使用N-suc-AAPF-pNA底物测定蛋白酶活性。用于AAPF水解测定的试剂溶液为:100mM Tris/HCl pH 8.6,其包含0.005% **TWEEN**[®]-80(Tris稀释缓冲液);100mM pH 8.6的Tris缓冲液,其包含10mM CaCl₂和0.005% **TWEEN**[®]-80(Tris/Ca缓冲液);和160mM suc-AAPF-pNA的DMSO溶液(suc-AAPF-pNA原液)(Sigma:S-7388)。为制备底物工作溶液,将1ml suc-AAPF-pNA原液加入100ml Tris/Ca缓冲液中并充分混合。将酶样品加入包含1mg/suc-AAPF-pNA工作溶液的MTP板(Costar 9017)中并使用SpectraMax读板仪在RT下以动力学模式于405nm处经3分钟测定活性。蛋白酶活性表达为mOD • min⁻¹。

[0305] 构成具有表1中所述最终洗涤剂洗涤浓度(g/L)的洗涤溶液并用于清洁性能测定中。

[0306] 表1:用于性能测定的洗涤剂条件的列表

洗涤剂*	类型	最终洗涤剂洗涤浓度, (g/L)	硬度浓度 (ppm)	缓冲液	设定 pH
OMO Color	HDD	5.3	250	2mM NaCO ₃	10.6
Kirkland Ultra	HDD	1.09	150	2mM NaCO ₃	10.6
OMO Klein & Krachtig	HDL	2.8	250	5mM 钠 HEPES	8.2
Kirkland Ultra	HDL	0.71	150	5mM 钠 HEPES	8.2
GSM-B 10.5	ADW	3	374	未缓冲的	按原样 ~10.5
GSM-B 9	ADW	3	374	未缓冲的, 加入 1M 柠檬酸盐来调节 pH	9

[0308] *洗涤剂来源:Kirkland Ultra HDD和HDL(Sun Products)于2012年购自美国的当地超市。OMO color HDD和OMO Klein&Krachtig(Unilever)于2013年购买自荷兰的当地超市。GSM-B购自WFK Testgewebe GmbH,Germany,www.testgewebe.de,组成在表2中给出。

[0309] 表2:GSM-B pH10.5ADW洗涤剂的组成

GSM-B 无磷酸盐洗涤剂	
组分	重量%
脱水柠檬酸钠	30.0
马来酸/丙烯酸共聚物钠盐 (SOKALAN [®] CP5; BASF)	12.0
过硼酸钠一水合物	5.0

[0311]	TAED	2.0
	二硅酸钠: Prottil A (Cognis)	25.0
	线性脂肪醇乙氧基化物	2.0
	无水碳酸钠	添加至 100

[0312] 就清洁测定而言,将在稀释缓冲液:10mM NaCl、0.1mM CaCl₂、0.005% Tween-80中稀释的10uL蛋白酶加入经洗涤剂填充的微样本板中以达到200uL的最终体积,其具有0.04ppm至10ppm的最终酶浓度。利用HDL或HDD制剂的衣物洗涤清洁测定在25℃下进行15分钟,然而自动盘碟(ADW)测定在40℃下进行30分钟。

[0313] 在温育之后,将100uL的上层清液转移到新鲜的MTP(Kisker G080-F)中,并且使用SpectraMax读板仪在600nm下读取EMPA-116样本的吸光度,或在405nm下读取PAS-38样本的吸光度。减去只有缓冲液的对照的吸光度并且所得的在600nm下的OD值(对于HDL和HDD洗涤剂)和在405nm下的OD值(对于ADW洗涤剂)作为蛋白酶浓度的函数作图。数据拟合成Langmuir公式。BspAG00296和BspM04033的清洁性能分别图示于图5-7和图8-10中。

[0314] 实施例11

[0315] 蛋白酶的稳定性评估

[0316] 在各种条件下测定BspAG00296、BspM04033、芽孢杆菌属(*B. sp.*) NN018132(WO2012175708—002)全长序列(SEQ ID NO:16)和截短形式(SEQ ID NO:17)、GG36(SEQ ID NO:18)、和FNA(SEQ ID NO:19)蛋白酶的稳定性。

[0317] SEQ ID NO:16示出全长芽孢杆菌属(*B. sp.*) NN018132蛋白酶的序列:
LMHNNQRWHYEMINAPQAWGITTGSSNVRIAVLDTGIDANHPNLRNLVDTSLGRSFVGGGTGDVQGHGTHVAGTIAS
YGSVSGVMQNARLIPVKVLGDNGSGSMYGIQQGILYAASINADVINMSLGGGGYDSGMNNAINTAVSSGTLVIAASG
NDGRGSISYPAAYSNAIAVGSVTSNRTRS NFSNYGSGLELMAPGSNIYSTYPNGQFRTLSGTSMATPHVAGVAGLIK
SANPNLSVTQVRNILRDTAQYAGSSNQYGYGIVNAYA AVQAAGGAVSYETNTSVSTNQSTYYRGNNVTMTAIVTDQ
NNSRLQGATVNFTITRPNGT TVTNATTTNSSGVATWTIGSNSSTAVGTYQVRAQTTPNYQSSSATT SFRLQ.

[0318] SEQ ID NO:17示出截短的芽孢杆菌属(*B. sp.*) NN018132蛋白酶的序列:
MHNNQRWHYEMINAPQAWGITTGSSNVRIAVLDTGIDANHPNLRNLVDTSLGRSFVGGGTGDVQGHGTHVAGTIAS
GSVSGVMQNARLIPVKVLGDNGSGSMYGIQQGILYAASINADVINMSLGGGGYDSGMNNAINTAVSSGTLVIAASG
DGRGSISYPAAYSNAIAVGSVTSNRTRS NFSNYGSGLELMAPGSNIYSTYPNGQFRTLSGTSMATPHVAGVAGLIK
ANPNLSVTQVRNILRDTAQYAGSSNQYGYGIVNAYA AVQAAGG.

[0319] SEQ ID NO:18示出GG36蛋白酶的序列:AQSVPWGISRVQAPAAHNRGLTGSGVKVAVLDTGIS
THPDLNIRGGASFVPGEPTQDGNGHGHVAGTIAALNNSIGVLGVAPSAELYAVKVLGASGSGSVSSIAQGLEWAG
NNGMHVANLSLGSPSPSATLEQAVNSATSRGVLVVAASGNSGAGSISYPARYANAMAVGATDQNNNRASFSQYGAGL
DIVAPGVNVQSTYPGSTYASLNGTSMATPHVAGAAALVKQKNPSWSNVQIRNHLKNTATSLGSTNLYGSLVNAEAA
TR.

[0320] SEQ ID NO:19示出FNA蛋白酶(BPN'Y217L)的序列:AQSVPYGVSQIKAPALHSQGYTG
SNV KVAVIDSGIDSSHPDLKVAGGASMVPSETNPFQDNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAVKVLGADGSGQ
YSWIINGIEWAIANNMDVINMSLGGPSGSAALKA AVDKAVASGVVVVAAAGNEGTSGSSSTVGYPGKYPSVIAVGAV
DSSNQRAFSSVGPPELDVMAPGVSIQSTLPGNKYGALNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENTTTKL
GDSFYYGKGLINVQAAAQ.

[0321] 通过在设定温度下温育之后测量残余的蛋白水解活性,在下文所示的三种胁迫条件下来测试稳定性。

[0322] 1.LAS/EDTA:0.02%LAS,2.1mM EDTA在50mM HEPES pH8中的溶液,0.005%Tween 80

[0323] 2.Tris/EDTA:50mM Tris,1mM EDTA,pH 9,0.005%Tween 80

[0324] 3.OMO HDL:10%OMO Klein&Krachtig(在使用之前使蛋白酶失活)

[0325] 就胁迫条件而言,将稀释的酶样品在96孔PCR板中在胁迫缓冲液/洗涤剂中混合并使用Tetrad2热循环仪在30℃、40℃、50℃、60℃和75℃下温育20分钟。就无胁迫条件而言,在与胁迫介质混合之后立即测定酶以构建基线(初始活性)。通过前述AAPF-pNA(对于OMO HDL)或DMC(对于LAS/EDTA和Tris/EDTA)底物测定的水解来测量胁迫条件和无胁迫条件下的蛋白酶活性。通过获取在每个温度下胁迫活性对无胁迫活性的比率并将其乘以100来计算残余活性百分比。对于在各种温度下运行的每个条件,每种蛋白酶的残余活性百分比示于表3-5中。

表 3: 蛋白酶在 LAS/EDTA 中的稳定性

酶	无胁迫的	温育温度				
		30℃	40℃	50℃	60℃	75℃
BspAG00296	100	>90	54	11	11	9
BspM04033	100	>90	>90	>90	66	23
芽孢杆菌属 (<i>B. sp.</i>) NN018132	100	64	10	8	8	9
GG36	100	78	16	1	1	1
FNA	100	>90	84	7	4	1

表 4: 蛋白酶在 Tris/EDTA 中的稳定性

酶	无胁迫的	温育温度				
		30℃	40℃	50℃	60℃	75℃
BspAG00296	100	>90	>90	>90	81	17
BspM04033	100	>90	>90	>90	42	26
芽孢杆菌属 (<i>B. sp.</i>) NN018132	100	>90	>90	>90	20	21
GG36	100	>90	87	8	6	6
FNA	100	>90	>90	24	19	20

表 5: 蛋白酶在 OMO HDL 中的稳定性

酶	无胁迫的	温育温度				
		30℃	40℃	50℃	60℃	75℃
BspAG00296	100	89	85	10	2	1
BspM04033	100	>90	>90	>90	15	0
芽孢杆菌属 (<i>B. sp.</i>) NN018132	100	75	45	0	0	0
GG36	100	85	81	35	3	2
FNA	100	>90	>90	71	1	1

[0329] 实施例12

[0330] 识别另外的芽孢杆菌属丝氨酸蛋白酶

[0331] 通过对购自DuPont Culture Collection的其它芽孢杆菌属物种芽孢杆菌属(B.sp.)SWT81(BspAA02831)、芽孢杆菌属(B.sp.)SWT4(SWT4_1110112)、芽孢杆菌属(B.sp.)SWT22(SWT22_1181566)、芽孢杆菌属(B.sp.)SWT32(SWT32_1214607)、芽孢杆菌属(B.sp.)SWT40(SWT40_1237842)、芽孢杆菌属(B.sp.)SWT41(SWT41_1431481)、芽孢杆菌属(B.sp.)SWT77(SWT77_1339394)、和芽孢杆菌属(B.sp.)SWT123(SWT123_1418561)的基因组测序来识别另外的枯草杆菌蛋白酶。基因组测序、组合和注释基本上如实施例2中所述。所有编码蛋白质的基因组与BspAG00296和BspM04033同源。

[0332] BspAA02831的前酶原形式的氨基酸序列以SEQ ID NO:20示出。预测的信号肽序列加下划线并且原序列呈斜体。预测的全加工成熟链的序列(BspAA02831,381个氨基酸)加粗。

[0333] SEQ ID NO:20示出BspAA02831的前酶原形式的氨基酸序列:MKKWLGMSAVVLMVLSL
FTGSGFANESK GKNNGDYIEGQLVISIEDQSE

FSIQSTNNIINKDQVLENKGF EIVDSLLGQSDPNEIQAFNHDF TATVVNEMGM
VYLVEYDVKKYKSIDKAKKELEKTMKDLGLEVRVYVSENFVMHAMEEVTAEDV
SIAMHNNQRWHYEMINAPQAWNITTGSRNVRIAVLDTGIDANHPNLR
NLVNTSLGRSFVGGGTGDVQGHGTHVAGTIASYGSVSGVMQNATLIP
VKVLGDNGSGSMYGIQQGILYAASVNSDVINMSLGGGGYSQGMDDA
IRTAVSSGTIVVAATGNDSRGSISYPAAYSGAIAVGSVTSNRTRSSFSNY
GQGLELMAPGSNIYSTYPNGQFRTLSTMATPHVAGVAGLIRAANP
NISVSEARSILQNTAQYAGSFNQYGYGIVDANA AVRAARGQSQQPSYE
TNTTVSTNASSYRRGQSVTVRADVVDQDGRALANSTVQFTITRPNGT
TVTNTATTNNSGVATWTIATSSSTARGTYGVQAATSLSGYEGSTATTS
FSVN.

[0334] BspAA02831的酶原形式的氨基酸序列以SEQ ID NO:21示出。原序列呈斜体。预测的全加工成熟链的序列(BspAA02831,381个氨基酸)加粗。

[0335] SEQ ID NO:21示出BspAA02831的酶原形式的氨基酸序列:

NESKGKNNGDYIEGQLVISIEDQSEFSIQSTNNIINKDQVLENKGFEIVDSL

Q

SDPNEIQAFNHDFATVNVNEMGMVYLVEYDVKKYKSIDKAKKELEKTMKDLG
LEVRYVSENFVMHAMEEVTAEDVSIAMHNNQRWHYEMINAPQAWNITT
GSRNVRIAFLDTGIDANHPNLRNLVNTSLGRSFVGGGTGDVQGHGT
HVAGTIASYGSVSGVMQNATLIPVKVLGDNGSGSMYGIQQGILYAAS
VNSDVINMSLGGGGYSQGMDDAIRTAVSSGTIVVAATGNDSRGSISYP
AAYSGAIAVGSVTSNRTRSSFSNYGQGLELMAPGSNIYSTYPNGQFRT
LSGTSMATPHVAGVAGLIRAANPNISVSEARSILQNTAQYAGSFNQYG
YGIVDANA AVRAARGQSQQPSYETNTTVSTNASSYRRGQSVTVRADV
VDQDGRALANSTVQFTITRPNGTTVTNTATTNNSGVATWTIATSSSTA
RGTYGVQAATSLSGYEGSTATTSFSVN.

[0336] 预测的全加工成熟链的序列(BspAA02831, 381个氨基酸)示出于SEQ ID NO:22中:
 MHNNQRWHYEMINAPQAWNITTGSRNVRIAFLDTGIDANHPNLRNLVNTSLGRSFVGGGTGDVQGHGTHVAGTIASY
 GSVSGVMQNATLIPVKVLGDNGSGSMYGIQQGILYAASVNSDVINMSLGGGGYSQGMDDAIRTAVSSGTIVVAATGN
 DSRGSISYP AAYSGAIAVGSVTSNRTRSSFSNYGQGLELMAPGSNIYSTYPNGQFRTLSGTSMATPHVAGVAGLIRA
 ANPNISVSEARSILQNTAQYAGSFNQYGYGIVDANA AVRAARGQSQQPSYETNTTVSTNASSYRRGQSVTVRADVD
 QDGRALANSTVQFTITRPNGTTVTNTATTNNSGVATWTIATSSSTARGTYGVQAATSLSGYEGSTATTSFSVN.

[0337] SWT4的前酶原形式的氨基酸序列以SEQ ID NO:23示出。预测的信号肽序列加下划线并且原序列呈斜体。预测的全加工成熟链的序列(SWT4, 381个氨基酸)加粗。

[0338] SEQ ID NO:23示出SWT4的前酶原形式的氨基酸序列:VKKSAVWVLMTVLVFSLFLNPAGI
GAQA*SDAASEKDDTAYIEGQLIVSVKS*

SDVSV

KGIEGVNKKIMGDVLRERGF AITDSIMGLGDPGEVNAFTNQEFSES VVKNMG
LVYLA EYDVSVYASVEEAKRALAEALKENGMEIRHISKNYEMHAIGELADVSPQ
MHPNQQWHYNMINAPQAWGTTTGSSSVIQAVLDTGIDHNHQSLANL
VNTSLGQS FVGGSTMDVQGHGTHVAGTIASYGSVSGVMH NATLVPV
KVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSM AEEAQ
TAVNAGSIVIAASGNDGAGSVSYP AAYSSVIAVGSVTSTGARSNFSNYG
SGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMRAVNPN
LSVSNARSIMQNTAQYAGSPTFYGYGIVDANA AVQQASGGSGDPSNIT
ETSISTDRFYVQRGQNVTSTAQVTNENGQGLANATVTFTITRPNGSTL
TNTATTNSSGFASWTVGTSGATATGTYSVEASSLQGYQGSSASTSFF
VY.

[0339] SWT4的酶原形式的氨基酸序列以SEQ ID NO:24示出。原序列呈斜体。预测的全加

工成熟链的序列(SWT4,381个氨基酸)加粗。

[0340] SEQ ID NO:24示出SWT4的酶原形式的氨基酸序列:

SDAASEKDDTAYIEGQLIVSVKSSDVSVKGIEGVNKKIMGDVLRERGF AITDSI
MGLGDPGEVN
AFTNQEFSESVVKNMGLVYLA EYDVS VYASVEEAKRALAEALKENGMEIRHIS
KNYEMHAI GELADVSPQMHPNQQWHYNMINAPQAWGTTTGSSSVIQA
VLDTGIDHNHQSLANLVNTSLGQSFVGGSTMDVQGHGTHVAGTIAS
YGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMS
LGGGGYNQSM AEEAAQTAVNAGSIVIAASGNDGAGSVSYPAAYSSVIA
VGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMAS
PHAAGVAGLMRAVNPNLSVSNARSIMQNTAQYAGSPTFYGYGIVDA
NAAVQQASGGSGDPSNITETSISTDRFYVQRGQNV TSTAQVTNENGQ
GLANATVTFTITRPNGSTLTNTATTNSSGFASWTVGTSGATATGTYSV
EASSSLQGYQGSSASTSFFVY.

[0341] 预测的全加工成熟链的序列(SWT4,381个氨基酸)以SEQ ID NO:25示出:

MHPNQQWHYNMINAPQAWGTTTGSSSVIQA VLDTGIDHNHQSLANLVNTSLGQSFVGGSTMDVQGHGTHVAGTIASY
GSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSM AEEAAQTAVNAGSIVIAASGN
DGAGSVSYPAAYSSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMRA
VNPNSVSNARSIMQNTAQYAGSPTFYGYGIVDANA AAVQQASGGSGDPSNITETSISTDRFYVQRGQNV TSTAQVTN
ENGQGLANATVTFTITRPNGSTLTNTATTNSSGFASWTVGTSGATATGTYSVEASSSLQGYQGSSASTSFFVY.

[0342] SWT22的前酶原形式的氨基酸序列以SEQ ID NO:26示出。预测的信号肽序列加下划线并且原序列呈斜体。预测的全加工成熟链的序列(SWT22,488个氨基酸)加粗。

[0343] SEQ ID NO:26示出SWT22的前酶原形式的氨基酸序列:MKKLLTLSILTLAMLVGFFSVNV
FADNEVQKKEDHKYIDGQLIVSVEMDG

KENSL

*KGQLNSTTELLQDNAELKKKGFAVSDSLLEEK*TADSQSVFSDSFVEKAAKKTG
FVYLM~~EYSTDEYDSIKTAMKELEKTLNELGLKVRYVSENFVVELLE~~TDVAEAE
DENKIAPLMHRNQEW~~HYGMINAPDAWGITTGSSNVRMAVLD~~TGIDSS
HPSLRNLVD~~TS~~LGRSYVGGNPEDRQGHGTHVAGT~~IASYGNVSGVMQ~~
NASLISVKVLGDDGSGSTYGIQQGVLYAASINSDVINMSLGGGGYSQG
FSDAIDTAVANGTVVIAASGNDGRASISYPAA~~YDGAIAVGSVTSSG~~SRS
NFSNYGNGLELMA~~PGSSIYSTYPNGQYRTL~~SGTSMAAPHAAGVAGLV
RAVDPSLSVSQVRGILADTAQYAGSSHQYGN~~GIVDAYAAVQAAGGSG~~
GAPAPSETNTSVSTNGSVFERGDDVTMTASVTDDNGNGLQGA~~AVNFT~~
ITRPNGSTVTNTATTNSSGNATWTIGSNSQTALGTYE~~VTAETTL~~SGYE
SSSDTTSFSFSNQAQTHQTVTDVSTNSNYARGQNVTVSAEVRDQDG
AVLSNATVSFTITRPNGSTVTNTGATNSAGVATWTVSTSGATATGTY
QVTAETTLTNYDGSSDSTS~~FFVY~~.

[0344] SWT22的酶原形式的氨基酸序列以SEQ ID NO:27示出。原序列呈斜体。预测的全加工成熟链的序列(SWT22,488个氨基酸)加粗。

[0345] SEQ ID NO:27示出SWT22的酶原形式的氨基酸序列:

*DNEVQKKEDHKYIDGQLIVSVEMDGKENS*SLKGQLNSTTELLQDNAELKKKG
FAVSDSLLEEK
TADSQSVFSDSFVEKAAKKTGFVYLM~~EYSTDEYDSIKTAMKELEKTLNELGLK~~
VRYVSENFVVELLETDVAEADENKIAPLMHRNQEW~~HYGMINAPDAWGI~~
TTGSSNVRMAVLD~~TGIDSS~~HPSLRNLVD~~TS~~LGRSYVGGNPEDRQGHG
THVAGT~~IASYGNVSGVMQ~~NASLISVKVLGDDGSGSTYGIQQGVLYAA
SINSDVINMSLGGGGYSQGFSDAIDTAVANGTVVIAASGNDGRASISYP
AA~~YDGAIAVGSVTSSG~~SRSNFSNYGNGLELMA~~PGSSIYSTYPNGQYRT~~
LSGT~~SMAAPHAAGVAGLVRAVDPSLSVSQVRGILADTAQYAGSSHQY~~
GNGIVDAYAAVQAAGGSGGAPAPSETNTSVSTNGSVFERGDDVTMTA
SVTDDNGNGLQGA~~AVNFTITRPNGSTVTNTATTNSSGNATWTIGSNS~~
QTALGTYE~~VTAETTL~~SGYESSSDTTSFSFSNQAQTHQTVTDVSTNSNY
YARGQNVTVSAEVRDQDGAVLSNATVSFTITRPNGSTVTNTGATNSA
GVATWTVSTSGATATGTYQVTAETTLTNYDGSSDSTS~~FFVY~~.

[0346] 预测的全加工成熟链的序列(SWT22,488个氨基酸)以SEQ ID NO:28示出:
MHRNQEW~~HYGMINAPDAWGITTGSSNVRMAVLD~~TGIDSSHPSLRNLVD~~TS~~LGRSYVGGNPEDRQGHGTHVAGT~~IASY~~
GNVSGVMQNASLISVKVLGDDGSGSTYGIQQGVLYAASINSDVINMSLGGGGYSQGFSDAIDTAVANGTVVIAASGN

DGRASISYPAAYDGAIAVGSVTSSGSRNFSNYGNGLLEMAPGSSIYSTYPNGQYRTLSTGTSMAAPHAAGVAGLVRA
VDPSLSVSQVRGILADTAQYAGSSHQYGNGLVDAYAAVQAAGGSGGAPAPSETNTSVSTNGSVFERGDDVTMTASVT
DDNGNGLQGA AVNFTITRPNGSTVTNTATTNSSGNATWTIGSNSQTALGTYEVTAEETLSGYESSSDTTSFSFSNQA
QTHQTVTDVSTNSNYARGQNVTVSAEVRDQDGAVLSNATVSFTITRPNGSTVTNTGATNSAGVATWTVSTSGATAT
GTYQVTAETTLTNYDGSSDSTSFYVY.

[0347] SWT32的前酶原形式的氨基酸序列以SEQ ID NO:29示出。预测的信号肽序列加下划线并且原序列呈斜体。预测的全加工成熟链的序列(SWT32,381个氨基酸)加粗。

[0348] SEQ ID NO:29示出SWT32的前酶原形式的氨基酸序列:VKKSAVWVLMTVLVFSLFLNP
AGIGAQA***SDAASEKDDTAYIEGQLIVSVKS***

SDVSV

KGIEGLNKKIMGNVLRERGF AITDSIMGLGDP AEVNAFTNQEFSES VVKNMGLVYLA EYDVSVYASVEEAKRALAEAL KENGMEIRHISKNYEMHAIGEPADVSPQ
MHPNQ QWHYNMINAPQAWGTTTGSSSVIQAVLDTGIDHNHQSLANLVNTSLGQS FVGGSTMDVQGHGTHVAGT IASVSGVSGVMHNATLVPV
KVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSM AEA AQTAVNAGSIVIAASGNDGAGSISYP AAYSSVIAVGSVTSTGARSNFSNYG
SGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMRAVNPNLSVSDARSIMQNTAQYAGSPTFYGYGIVDANA AAVQQASGGSGGPSNIT
ETSISTDRFYVQRGQNVTSTAQVTNENGQGLANATVTFITRPNGSTLTNTATTNGSGFASWT VGTSGATATGTYSVEASSLQGYQGSSASTSFF
VY.

[0349] SWT32的酶原形式的氨基酸序列以SEQ ID NO:30示出。原序列呈斜体。预测的全加工成熟链的序列(SWT32,381个氨基酸)加粗。

[0350] SEQ ID NO:30示出SWT32的酶原形式的氨基酸序列:

SDAASEKDDTAYIEGQLIVSVKSSDVSVKGIEGLNKKIMGNVLRERGF AITDSIMGLGDP AEVNAFTNQEFSES VVKNMGLVYLA EYDVSVYASVEEAKRALAEAL
KENGMEIRHISKNYEMHAIGEPADVSPQMHPNQ QWHYNMINAPQAWGTTTGSSSVIQAVLDTGIDHNHQSLANLVNTSLGQS FVGGSTMDVQGHG
THVAGT IASVSGVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSM AEA AQTAVNAGSIVIAASGNDGAGSISYP
AAYSSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMRAVNPNLSVSDARSIMQNTAQYAGSPTFY
GYGIVDANA AAVQQASGGSGGPSNITETSISTDRFYVQRGQNVTSTAQVTNENGQGLANATVTFITRPNGSTLTNTATTNGSGFASWT VGTSGAT
ATGTYSVEASSLQGYQGSSASTSFFVY.

[0351] 预测的全加工成熟链的序列(SWT32,381个氨基酸)以SEQ ID NO:31示出:
MHPNQWHYNMINAPQAWGTTTSSSVIQAVLDTGIDHNHQLANLVNTSLGQSFVGGSTMDVQGHGTHVAGTIASY
GSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADVINMSLGGGGYNQSMAEAAQTAVNAGSIVIAASGN
DGAGSISYPAAYSSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMRA
VNPNSVSDARSIMQNTAQYAGSPTFYGYGIVDANAQVQASGGSGGPSNITETSIISTDRFYVQRGQNVSTAQVTN
ENGQGLANATVTFTITRPNGSTLTNTATTNGSGFASWTVGTSGATATGTYSVEASSSLQGYQGSSASTSFFVY.

[0352] SWT40的前酶原形式的氨基酸序列以SEQ ID NO:32示出。预测的信号肽序列加下划线并且原序列呈斜体。预测的全加工成熟链的序列(SWT40,381个氨基酸)加粗。

[0353] SEQ ID NO:32示出SWT40的前酶原形式的氨基酸序列:MKKWLGMSAVVLMVFSMFTGAG
FANESKKGKNNGDYIEGQLVISIEDQS

QFSIQ

*ATNNIINKDEVLENNGFEIVDSLLGQNDPNEIQAYNHDF*TATVVNEMGLVYLV
EYDVKKYKSIDKAKKELEKTMKDLGLEVRYVSENFVMHAMEEVTAEVSIAM
HNNQRWHYEMINAPQAWNVTTGSRNVRIAVLDTGIDANHPNLRNLV
NTSLGRSFVGGGTGDVQGHGTHVAGTIASYGSVSGVMQNATLIPVK
VLGDNGSGSMYGIQQGILYAASVNSDVINMSLGGGGYSQGMDDAIRT
AVSSGTIVVAATGNDSRGSISYPAAYSGAIAVGSVTSNRTRSSFSNYGQ
GLELMAPGSNIYSTYPNGQFRTLSGTSMATPHVAGVAGLIRAANPNIS
VAEARSILQNTAQYAGSFNQYGYGIVDANAAVRAARGQTEQPRYET
NTTVSTNASTYRRGQSVTVRADVVDQDGRALANSTVQFTITRPNGTT
VTNTATTNSSGVATWTIGTSSSTARGTYGVQAATSLSGYEGSTATTSF
VVN.

[0354] SWT40的酶原形式的氨基酸序列以SEQ ID NO:33示出。原序列呈斜体。预测的全加工成熟链的序列(SWT40,381个氨基酸)加粗。

[0355] SEQ ID NO:33示出SWT40的酶原形式的氨基酸序列:

*NESKGKNNGDYIEGQLVISIEDQSQFSIQATNNIINKDEVLENNGFEIVDSLIG
QNDPNEIQAY*
*NHDF*TATVVNEMGLVYLVEYDVKKYKSIDKAKKELEKTMKDLGLEVRYVSEN
*FVMHAMEEVTAEVSLAMHNNQRWHYEMINAPQAWNVT*TGSRNVRIA
VLDTGIDANHPNLRNLVNTSLGRSFVGGGTGDVQGHGTHVAGTIASY
GSVSGVMQNATLIPVKVLGDNGSGSMYGIQQGILYAASVNSDVINMS
LGGGGYSQGMDDAIRTAVSSGTIVVAATGNDSRGSISYPAAYSGAIAV
GSVTSNRTRSSFSNYGQGLELMAPGSNIYSTYPNGQFRTLSTGTSMATP
HVAGVAGLIRAANPNISVAEARSILQNTAQYAGSFNQYGYGIVDANA
AVRAARGQTEQPRYETNTTVSTNASTYRRGQSVTVRADVVDDQDGRA
LANSTVQFTITRPNGTTVTNTATTNSSGVATWTIGTSSSTARGTYGVQ
AATSLSGYEGSTATTSFVVN.

[0356] 预测的全加工成熟链的序列(SWT40,381个氨基酸)以SEQ ID NO:34示出:
MHNNQRWHYEMINAPQAWNVTGSRNVRIA VLDTGIDANHPNLRNLVNTSLGRSFVGGGTGDVQGHGTHVAGTIASY
GSVSGVMQNATLIPVKVLGDNGSGSMYGIQQGILYAASVNSDVINMSLGGGGYSQGMDDAIRTAVSSGTIVVAATGN
DSRGSISYPAAYSGAIAVGSVTSNRTRSSFSNYGQGLELMAPGSNIYSTYPNGQFRTLSTGTSMATPHVAGVAGLIRA
ANPNISVAEARSILQNTAQYAGSFNQYGYGIVDANA AVRAARGQTEQPRYETNTTVSTNASTYRRGQSVTVRADVVDD
QDGRALANSTVQFTITRPNGTTVTNTATTNSSGVATWTIGTSSSTARGTYGVQAATSLSGYEGSTATTSFVVN.

[0357] SWT41的前酶原形式的氨基酸序列以SEQ ID NO:35示出。预测的信号肽序列加下
划线并且原序列呈斜体。预测的全加工成熟链的序列(SWT41,381个氨基酸)加粗。

[0358] SEQ ID NO:35示出SWT41的前酶原形式的氨基酸序列:VKKSAVWVLMTVLVFSLFLNPAG
IGAQA*SDAASEKDDTAYIEGQLIVSVKS*

SDVS

*VKGIEGLNKKIMGNVLRERGF*AITDSIMGLGDPAEVNAFTNQEFSSESVVKNM
*GLVYLA*EYDVSVYASVEEAKRALAEALKENGMEIRHISKNYEMHAIGEPADVSP
QMHPNQQWHYNMINAPQAWGTTTGSSSVIQAVLDTGIDHNHQLAN
LVNTSLGQSFVGGSTMDVQGHGTHVAGTIASYGSVSGVMHNATLVP
VKVLNDSGSGSLFGITQGILYSADIGADVINMSLGGGGYNQSMAEAA
QTAVNAGSIVIAASGNDGAGSISYPAAYSSVIAVGSVTSTGARSNFSNY
GSGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMRAVNP
NLSVSDARSIMQNTAQYAGSPTFYGYGIVDANA AVQQASGGSGGPSN
ITETSISTDRFYVQRGQNVTSTAQVTNENGQGLANATVTFTITRPNGS
TLTNTATTNGSGFASWTVGTSATATGTYSVEASSSLQGYQGSSASTS
FFVY.

[0359] SWT41的酶原形式的氨基酸序列以SEQ ID NO:36示出。原序列呈斜体。预测的全加

工成熟链的序列(SWT41,381个氨基酸)加粗。

[0360] SEQ ID NO:36示出SWT41的酶原形式的氨基酸序列:

SDAASEKDDTAYIEGQLIVSVKSSDVS VKGIEGLNKKIMGNVLRERGF AITDSI
MGLGDPAEVN
AFTNQEFSESVVKNMGLVYLA EYDVS VYASVEEAKRALAEALKENGMEIRHIS
KNYEMHAIGEPADVSPQMHPNQQWHY NMINAPQAWGTTTGSSSVIQA
VLDTGIDH NHQSLANLVNTSLGQS FVGGSTMDVQGHGTHVAGTIAS
YGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMS
LGGGGYNQSMAEAAQTAVNAGSIVIAASGNDGAGSISYPAA YSSVIAV
GSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMASP
HAAGVAGLMRAVNP NLSVSDARSIMQNTAQYAGSPTFYGYGIVDAN
AAVQQASGGSGGPSNITETSISTDRFYVQRGQNVTSTAQVTNENGQG
LANATVTFTITRPNGSTLTNTATTNGSGFASWTVGTSGATATGTYSV
EASSSLQGYQGSSASTSFFVY.

[0361] 预测的全加工成熟链的序列(SWT41,381个氨基酸)以SEQ ID NO:37示出:
MHPNQQWHY NMINAPQAWGTTTGSSSVIQA VLDTGIDH NHQSLANLVNTSLGQS FVGGSTMDVQGHGTHVAGTIAS
YGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADV INMSLGGGGYNQSMAEAAQTAVNAGSIVIAASGN
DGAGSISYPAA YSSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMASPHAAGVAGLMRA
VNP NLSVSDARSIMQNTAQYAGSPTFYGYGIVDANA AAVQQASGGSGGPSNITETSISTDRFYVQRGQNVTSTAQVTN
ENGQGLANATVTFTITRPNGSTLTNTATTNGSGFASWTVGTSGATATGTYSVEASSSLQGYQGSSASTSFFVY.

[0362] SWT77的前酶原形式的氨基酸序列以SEQ ID NO:38示出。预测的信号肽序列加下划线并且原序列呈斜体。预测的全加工成熟链的序列(SWT77,381个氨基酸)加粗。

[0363] SEQ ID NO:38示出SWT77的前酶原形式的氨基酸序列:LKKSAVWVLMTVLVFSLFLNPAG
IGAQA*SDAASGKEEAAYIEGQLIVSVKA*

SDASV

KGIEGV

NQKVMGNELRERGF AITDSIMGLGDPAEVNAFTNQEFSESVVRNMGLVYLA E
YDVS

VYKSSDEAKRSLAEALKENGMEIRHISENYEMHAIGEPADVSPQMHPNQQW
HYNMINAPQAWETTTGSSSVIQA VLDTGIDH NHQSLANLVNTSLGQS
FVGGSTMDVQGHGTHVAGTIAS YGSVSGVMHNATLVPVKVLNDSGS
GSLFGITQGILYSADIGADV INMSLGGGGYNQSMAEAAQTAVDAGSI
VIAASGNDGAGSISYPAA YSSVIAVGSVTSTGARSNFSNYGSGLELMAP

**GSNIYSTVPNNGYATFSGTSMAPHAAGVAGLMRAVNSNLSVSDARS
IMQNTAQYAGSPTFYGYGIVDANA AVQQASGGSGGPSNITETSISTDR
YYVQRGQNVNSTAQVTNENGQALANATVTFTITRPNGSTLTNTATTN
SSGVASWTVGTSGGTATGTYSVEASSSLQGYQGSSASTSFFVY.**

[0364] SWT77的酶原形式的氨基酸序列以SEQ ID NO:39示出。原序列呈斜体。预测的全加工成熟链的序列(SWT77,381个氨基酸)加粗。

[0365] SEQ ID NO:39示出SWT77的酶原形式的氨基酸序列:

*SDAASGKEEAAYIEGQLIVSVKASDASVKGIEGVNQKVMGNELRERGFITDSI
MGLGDPAEV
NAFTNQEFSESVVRNMGLVYLAEYDVSVKSSDEAKRSLAEALKENGMEIRHIS
ENYEMHAIGEPADVSPQMHPNQQWHYNMINAPQAWETTTGSSSVIQA
VLDTGIDHNNHQLANLVNTSLGQSFVGGSTMDVQGHGTHVAGTIAS
YGSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADVINMS
LGGGGYNQSMAEAAQTAVDAGSIVIAASGNDGAGSISYPAAYSSVIAV
GSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMAP
HAAGVAGLMRAVNSNLSVSDARSIMQNTAQYAGSPTFYGYGIVDAN
AAVQQASGGSGGPSNITETSISTDRYYVQRGQNVNSTAQVTNENGQA
LANATVTFTITRPNGSTLTNTATTNSSGVASWTVGTSGGTATGTYSVE
ASSSLQGYQGSSASTSFFVY.*

[0366] 预测的全加工成熟链的序列(SWT77,381个氨基酸)以SEQ ID NO:40示出:
MHPNQQWHYNMINAPQAWETTTGSSSVIQA VLDTGIDHNNHQLANLVNTSLGQSFVGGSTMDVQGHGTHVAGTIASY
GSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADVINMSLGGGGYNQSMAEAAQTAVDAGSIVIAASGN
DGAGSISYPAAYSSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMAPHAAGVAGLMRA
VNSNLSVSDARSIMQNTAQYAGSPTFYGYGIVDANA AVQQASGGSGGPSNITETSISTDRYYVQRGQNVNSTAQVTN
ENGQALANATVTFTITRPNGSTLTNTATTNSSGVASWTVGTSGGTATGTYSVEASSSLQGYQGSSASTSFFVY.

[0367] SWT123的前酶原形式的氨基酸序列以SEQ ID NO:41示出。预测的信号肽序列加下划线并且原序列呈斜体。预测的全加工成熟链的序列(SWT123,488个氨基酸)加粗。

[0368] SEQ ID NO:41示出SWT123的前酶原形式的氨基酸序列:MKKLLTLFLLTLVMLVGLFSVN
VMA*DNEDQKYIEGQLIVSVETNVGGYSIT*

GLM

NNTSEILQDNATLRNKGFHVADTLLENNAAGVQSVFSSNFVEETAKRTGLVYL
MEYSPEDYESIQEAKNDLENTLKEGLKVRYVSENFVVELFETETPSNTDEENI
ISPFMHSNQEWYHGMINAPDAWGITTGSSNVRIAILDTGIDSSHPSLRN
LVDTGLGRSYVGGSPEDVQGHGTHVAGTIASYGAVSGVMQDATLIS
VKVLGDDGSGSMYGIQQGVLYAASVGADVINMSLGGGGYNQGFSDA
IDTAVANGTVVIAASGNDGRASISYPAAVDGAIAVGSVTSSGNRSNFSN
YGSGLLEMAPGSSIYSTYPNGQYRTLSTGTSMAAPHAAGVAGLVRAVN
PNLSVAEVRISILADTAQYAGSTYQYGNGIVDAFAAVQAAGGSGGTPS
PGVTNTVVSTDKSVYERGDQVTMTATVTDEDGNALQGASVNYTITR
PNGSDVTNTATTNTNGIATWTIGSNSQTAIGTYDVTAESSLSGYESSTD
TTSFRFSDQAQSQQTVDVSTNSSYARGQNVTISAEVTDQDGAALSN
ATVSFTITRPNGSTLTNTATTNSAGVASWTVSTSSGTARGTYEVTAES
TYSTYEGSSDTTTSFYVY.

[0369] SWT123的酶原形式的氨基酸序列以SEQ ID NO:42示出。原序列呈斜体。预测的全加工成熟链的序列(SWT123,488个氨基酸)加粗。

[0370] SEQ ID NO:42示出SWT123的酶原形式的氨基酸序列:

DNEDQKYIEGQLIVSVETNVGGYSITGLMNNTSEILQDNATLRNKGFHVADTL
LEN
NAAGVQSVFSSNFVEETAKRTGLVYLMEYSPEDYESIQEAKNDLENTLKEGL
KVRYVSENFVVELFETETPSNTDEENIISPFMHSNQEWYHGMINAPDAWG
ITTGSSNVRIAILDTGIDSSHPSLRNLVDTGLGRSYVGGSPEDVQGHGT
HVAGTIASYGAVSGVMQDATLISVKVLGDDGSGSMYGIQQGVLYAA
SVGADVINMSLGGGGYNQGFSDAIDTAVANGTVVIAASGNDGRASIS
YPAAVDGAIAVGSVTSSGNRSNFSNYGSGLLEMAPGSSIYSTYPNGQY
RTLSTGTSMAAPHAAGVAGLVRAVNPNLSVAEVRISILADTAQYAGSTY
QYGNGIVDAFAAVQAAGGSGGTPSPGVTNTVVSTDKSVYERGDQVT
MTATVTDEDGNALQGASVNYTITRPNGSDVTNTATTNTNGIATWTIG
SNSQTAIGTYDVTAESSLSGYESSTDTTTSFRFSDQAQSQQTVDVSTNS
SYARGQNVTISAEVTDQDGAALSNATVSFTITRPNGSTLTNTATTNS
AGVASWTVSTSSGTARGTYEVTAE**STYSTYEGSSDTTTSFYVY.**

[0371] 预测的全加工成熟链的序列(SWT123,488个氨基酸)以SEQ ID NO:43示出:
MHSNQEWYHGMINAPDAWGITTGSSNVRIAILDTGIDSSHPSLRNLVDTGLGRSYVGGSPEDVQGHGTHVAGTIASY
GAVSGVMQDATLISVKVLGDDGSGSMYGIQQGVLYAASVGADVINMSLGGGGYNQGFSDAIDTAVANGTVVIAASGN

DGRASISYPAAAYDGAIAVGSVTSSGNRSNFSNYGSGLELMAPGSSIYSTYPNGQYRTLSGTSMAAPHAAGVAGLVRA
VNPNLSVAEVRSLADTAQYAGSTYQYGNIGVDAFAAVQAAGGSGGTPSPGVTNTVVSTDKSVYERGDQVTMTATVT
DEDGNALQGASVNYTITRPNNGSDVTNTATTNTNGIATWTIGSNSQTAIGTYDVTAESSLSGYESSTDTTSFRRSDQA
QSQQTVTDVSTNSSYYARGQNVITISAEVTDQDGAALSNATVSFTITRPNGSTLTNTATTNSAGVASWTVSTSSGTAR
GTYEVTAEYSTYTYEGSSDTSFYVY.

[0372] 实施例13

[0373] 同源蛋白酶的识别

[0374] 对BspAG00296(SEQ ID NO:3,274个氨基酸)、BspM04033(SEQ ID NO:11,276个氨基酸)和SWT77(SEQ ID NO:40,381个氨基酸)的预测成熟形式的氨基酸序列进行针对NCBI非冗余蛋白质数据库的BLAST检索(Altschul等人,Nucleic Acids Res,25:3389-402,1997)。类似的检索针对基因组查询专利数据库再次运行,其中分别使用SEQ ID NO:3、SEQ ID NO:7、和SEQ ID NO:40作为查询序列将检索参数设为默认值。对于BspAG00296的检索结果的子集示于表6和表7中;对于BspM04033示于表8和表9中;并且对于SWT77示于表10和表11中。将两个检索子集的百分比同一性(PID)限定为相同残基的数目除以逐对比对中所比对残基的数目。标记“序列长度”的列是指与所列登录号相关联的蛋白质序列的长度(以氨基酸计),然而标记“对比长度”的列是指用于PID计算的比对蛋白质序列的长度(以氨基酸计)。

[0375]

表 6: 具有与从 NCBI 非冗余蛋白质数据库识别的 BspAG00296 (SEQ ID NO:3) 蛋白质的百分比同一性的序列的列表

登录号	PID	生物	序列长度	比对长度
WP_026675114.1	82	博戈里亚芽孢杆菌 (<i>B. bogoriensis</i>)	539	273
WP_010283106	77	<i>B. timonensis</i>	544	273
WP_006679321	77	<i>P. dendritiformis</i>	578	273
WP_025025887.1	75	解甘露糖醇芽孢杆菌 (<i>B. mannanyticus</i>)	550	272
WP_026080796.1	54	地衣芽孢杆菌 (<i>B. licheniformis</i>)	378	252
CAJ70731	53	地衣芽孢杆菌 (<i>B. licheniformis</i>)	379	252
BAA06157	50	芽孢杆菌属 (<i>B. sp.</i>) <i>Sendai</i>	382	267
AAA22212	50	嗜碱芽孢杆菌 (<i>B. alcalophilus</i>)	380	267
WP_006636716	50	索诺拉沙漠芽孢杆菌 (<i>B. sonorensis</i>)	378	252
AAC43581	50	芽孢杆菌属 (<i>B. sp.</i>) <i>SprD</i>	379	259
P29599	50	迟缓芽孢杆菌 (<i>B. lentus</i>)	269	267
ABI26631.1	46	克劳氏芽孢杆菌 (<i>B. clausii</i>)	361	264
WP_010333625	49	莫哈韦芽孢杆菌 (<i>B. mojavenensis</i>)	381	267
BAN09118	49	枯草芽孢杆菌 (<i>B. subtilis</i>)	381	267
WP_012957236.1	48	假坚强芽孢杆菌 (<i>B. pseudofirmus</i>)	374	253
CAA74536	48	枯草芽孢杆菌 (<i>B. subtilis</i>) <i>str168</i>	381	267
WP_010329279	48	死谷芽孢杆菌 (<i>B. vallismortis</i>)	381	267
BAD21128	48	芽孢杆菌属 (<i>B. sp.</i>) <i>KSM-LDI</i>	377	269
AAC43580	47	芽孢杆菌属 (<i>B. sp.</i>) <i>SprC</i>	378	273
BAD11988	47	芽孢杆菌属 (<i>B. sp.</i>) <i>KSM-LDI</i>	376	273
WP_003327717.1	47	深褐芽孢杆菌 (<i>B. atrophaeus</i>)	382	267
ADN04910	48	环状芽孢杆菌 (<i>B. circulans</i>)	275	267
AFP23380.1	47	列城芽孢杆菌 (<i>B. lehensis</i>)	276	267
WP_007497196	48	平流层芽孢杆菌 (<i>B.</i>	383	267

[0376]

		<i>stratosphericus</i>)		
ADK11996	48	短小芽孢杆菌 (<i>B. pumilus</i>)	383	267
CAA24990	46	解淀粉芽孢杆菌 (<i>B. amyloliquefaciens</i>)	376	264
WP_022553591.1	46	甲基营养芽孢杆菌 (<i>B. methylotrophicus</i>)	382	264
ABY25856	46	<i>G. stearothermophilus</i>	382	264

[0377]

表 7: 具有与从基因组查询数据库识别的 BspAG00296 (SEQ ID NO:3) 蛋白质的百分比同一性的序列的列表

专利号#	PID	生物	序列长度	比对长度
WO2012175708-0004	82.85	博戈里亚芽孢杆菌 (<i>B. bogoriensis</i>)	541	274
WO2012175708-0002	82.35	芽孢杆菌属 (<i>B. sp.</i>) NN018132	548	274
WO2012175708-0006	77.66	<i>P. dendritiformis</i>	578	273
DE10260903	55.16	地衣芽孢杆菌 (<i>B. licheniformis</i>) 合成	379	252
JP2002533080-0001	54.55	地衣芽孢杆菌 (<i>B. licheniformis</i>)	275	253
JP1991072876-0004	54.37	地衣芽孢杆菌 (<i>B. licheniformis</i>)	274	252
US5719021-0004	54.37	地衣芽孢杆菌 (<i>B. licheniformis</i>)	350	252
WO2011014278-0109	54.37	地衣芽孢杆菌 (<i>B. licheniformis</i>)	274	252
US6274365-0007	54.37	地衣芽孢杆菌 (<i>B. licheniformis</i>)	274	252
US6908991-0004	54.37	芽孢杆菌属 (<i>B. sp.</i>)	274	252
EP0405901-0008	54.37	芽孢杆菌属 (<i>B. sp.</i>)	274	252
US7449187-0007	50.19	芽孢杆菌属 (<i>B. sp.</i>)	268	267
WO2010123754-0051	50.19	芽孢杆菌属 (<i>B. sp.</i>)	269	267
US20110045572-0034	50.19	芽孢杆菌属 (<i>B. sp.</i>)	269	267

[0378]

表 8: 具有与从 NCBI 非冗余蛋白质数据库识别的 BspM04033 (SEQ ID NO:11) 蛋白质的百分比同一性的序列的列表

登录号	PID	生物	序列长度	比对长度
WP_010283106	77	<i>B. timonensis</i>	544	275
WP_026675114.1	75	博戈里亚芽孢杆菌 (<i>B. bogoriensis</i>)	539	276
WP_025025887.1	73	解甘露糖醇芽孢杆菌 (<i>B. mannanilyticus</i>)	550	276

[0379]

表 8: 具有与从 NCBI 非冗余蛋白质数据库识别的 BspM04033 (SEQ ID NO:11) 蛋白质的百分比同一性的序列的列表				
登录号	PID	生物	序列长度	比对长度
WP_006679321	71	<i>P. dendritiformis</i>	578	275
AEU12640.1	48	地衣芽孢杆菌 (<i>B. licheniformis</i>)	379	255
WP_024712963.1	48	特基拉芽孢杆菌 (<i>B. tequilensis</i>)	894	226
WP_014113314.1	48	枯草芽孢杆菌亚种 <i>spizizenii</i> TU-B-10	894	226
WP_014730854.1	48	<i>Mesotoga prima</i> MesG1.Ag.4.2	503	267
WP_010329314.1	48	死谷芽孢杆菌 (<i>B. vallismortis</i>)	893	226
AAC43580	47	芽孢杆菌属 (<i>B. sp.</i>) <i>SprC</i>	378	274
CAJ70731	47	地衣芽孢杆菌 (<i>B. licheniformis</i>)	379	269
WP_012957236.1	47	假坚强芽孢杆菌 (<i>B. pseudofirmus</i>)	374	251
AAC43581	46	芽孢杆菌属 (<i>B. sp.</i>) <i>SprD</i>	379	270
BAD11988	46	芽孢杆菌属 (<i>B. sp.</i>) <i>KSM-LDI SA</i> 蛋白酶	376	274
P27693.1	46	嗜碱芽孢杆菌 (<i>B. alcalophilus</i>)	380	269
BAD21128	46	芽孢杆菌属 (<i>B. sp.</i>) <i>KSM_LDI</i>	377	270
P29599	44	迟缓芽孢杆菌 (<i>B. lentus</i>) <i>Savinase</i>	269	269
P41362.1	45	克劳氏芽孢杆菌 (<i>B. clausii</i>)	380	269
BAA06157	45	芽孢杆菌属 (<i>B. sp.</i>) <i>Sendai</i>	382	269
ADN04910	44	环状芽孢杆菌 (<i>B. circulans</i>)	275	267
AFP23380.1	43	列城芽孢杆菌 (<i>B. lehensis</i>)	276	267
WP_007497196	43	平流层芽孢杆菌 (<i>B. stratosphericus</i>)	383	267
ADK11996	43	短小芽孢杆菌 (<i>B. pumilus</i>)	383	267
WP_006636716	43	索诺拉沙漠芽孢杆菌 (<i>B. sonorensis</i>)	378	269
WP_003327717.1	42	深褐芽孢杆菌 (<i>B. atrophaeus</i>)	382	271
WP_017417394.1	41	解淀粉芽孢杆菌 (<i>B. amyloliquefaciens</i>)	382	271
WP_010333625	41	莫哈韦芽孢杆菌 (<i>B. mojavensis</i>)	381	271
WP_032721270.1	41	枯草芽孢杆菌 (<i>B. subtilis</i>)	381	271
WP_015252429.1	41	枯草芽孢杆菌亚种 <i>subtilis</i> str. <i>BSPI</i>	381	271
WP_022553591.1	41	甲基营养芽孢杆菌 (<i>B. methylotrophicus</i>)	382	271
ABY25856	41	<i>G. stearothermophilus</i>	382	271

[0380]

表 9: 具有与从基因组查询数据库识别的 BspM04033 (SEQ ID NO:11) 蛋白质的百分

[0381]

比同一性的序列的列表				
专利号#	PID	生物	序列长度	比对长度
WO2012175708-0004	75.2	博戈里亚芽孢杆菌 (<i>B. bogoriensis</i>)	541	274
WO2012175708-0002	74.6	芽孢杆菌属 (<i>B. sp.</i>) NN018132	548	276
WO2012175708-0006	71.6	<i>P. dendritiformis</i>	578	275
WO9628566	47.9	合成	274	269
WO9406915-0001	47.8	空	275	274
CN102676561-0002	47.2	地衣芽孢杆菌 (<i>B. licheniformis</i>)	350	269
US20090011489-0005	47.2	地衣芽孢杆菌 (<i>B. licheniformis</i>)	274	269
CN101215534-0002	47.2	地衣芽孢杆菌 (<i>B. licheniformis</i>) YPIA	379	269
WO2011014278-0116	47.1	地衣芽孢杆菌 (<i>B. licheniformis</i>)	269	263
WO2009005647-0008	47.1	地衣芽孢杆菌 (<i>B. licheniformis</i>)	374	263
US5275945-0002	47.1	空	377	274
WO2011014278-0109	46.8	地衣芽孢杆菌 (<i>B. licheniformis</i>)	274	269
US8168417-5227	46.8	地衣芽孢杆菌 (<i>B. licheniformis</i>)	379	269
US20030049619-0011	46.8	地衣芽孢杆菌 (<i>B. licheniformis</i>)	379	269
EP2166076-0002	46.5	地衣芽孢杆菌 (<i>B. licheniformis</i>)	274	269
EP1921148-0026	46.2	地衣芽孢杆菌 (<i>B. licheniformis</i>)	378	273

[0382]

表 10: 具有与从 NCBI 非冗余蛋白质数据库识别的 SWT77 (SEQ ID NO:40) 蛋白质的百分比同一性的序列的列表				
登录号	PID	生物	序列长度	比对长度
WP_026675114.1	70.8	博戈里亚芽孢杆菌 (<i>B. bogoriensis</i>)	539	380
WP_010283106.1	69.8	<i>B. timonensis</i>	544	387
WP_025025887.1	69.8	解甘露糖醇芽孢杆菌 (<i>B. mannilyticus</i>)	550	380
WP_006679321.1	58.5	<i>P. dendritiformis</i> C454	578	381
WP_014113314.1	48.2	枯草芽孢杆菌亚种 <i>spizizenii</i> TU-B-10	894	226
AAT75303.1	48.2	莫哈韦芽孢杆菌 (<i>B. mojavensis</i>)	379	226

[0383]

WP_024712963.1	48.2	特基拉芽孢杆菌 (<i>B. tequilensis</i>)	894	226
WP_010329314.1	47.8	死谷芽孢杆菌 (<i>B. vallismortis</i>)	893	226
AEU12640.1	47.8	地衣芽孢杆菌 (<i>B. licheniformis</i>)	379	255
AAC43580.1	47.4	芽孢杆菌属 (<i>B. sp.</i>) <i>SprC</i>	378	274
WP_012957236.1	46.2	假坚强芽孢杆菌 (<i>B. pseudofirmus</i>) <i>OF4</i>	374	251
AAC43581.1	45.6	芽孢杆菌属 (<i>B. sp.</i>) <i>SprD</i>	379	270
BAD11988.2	45.6	芽孢杆菌属 (<i>B. sp.</i>) <i>KSM-LD1 SA</i>	376	274
AIC95824.1	45.4	列城芽孢杆菌 (<i>B. lehensis</i>)	378	269
BAD21128.1	45.2	芽孢杆菌属 (<i>B. sp.</i>) <i>KSM-LD1 SB</i>	377	270
BAA06157.1	44.2	嗜碱芽孢杆菌 (<i>B. alcalophilus</i>)	382	269
P29599.1	43.9	迟缓芽孢杆菌 (<i>B. lentus</i>)	269	269
WP_006636716.1	43.5	索诺拉沙漠芽孢杆菌 (<i>B. sonorensis</i>)	378	255
AAX14553.1	43.1	短小芽孢杆菌 (<i>B. pumilus</i>)	381	267
WP_033016381.1	42.1	<i>G. stearothermophilus</i>	351	273

[0384]

表 11: 具有与从基因组查询数据库识别的 SWT77 (SEQ ID NO:40) 蛋白质的百分比同一性的序列的列表

专利号#	PID	生物	序列长度	比对长度
WO2012175708-0004	69.74	博戈里亚芽孢杆菌 (<i>B. bogoriensis</i>)	541	380
WO2012175708-0002	69.05	芽孢杆菌属 (<i>B. sp.</i>) <i>NN018132</i>	548	378
WO2012175708-0006	58.53	<i>P. dendritiformis</i>	578	381
CN102703482-0002	49.2	地衣芽孢杆菌 (<i>B. licheniformis</i>) ; <i>YPIA CCTCC M207021</i> 合成	379	250
WO9628566	49.02	合成	274	255
WO2005124012-0020	48.24	地衣芽孢杆菌 (<i>B. licheniformis</i>) 合成	274	255
WO03062380-0005	48.03	地衣芽孢杆菌 (<i>B. licheniformis</i>)	273	254
WO2005124012-0014	47.84	地衣芽孢杆菌 (<i>B. licheniformis</i>) 合成	274	255
WO9406915-0001	47.45	芽孢杆菌属 (<i>B. sp.</i>)	275	274
US5275945-0002	46.35	芽孢杆菌属 (<i>B. sp.</i>)	377	274

[0385] 构建以下枯草杆菌蛋白酶的氨基酸序列的系统发育树: BspAG00296 (SEQ ID NO: 4)、BspM04033 (SEQ ID NO: 11)、BspW01765 (SEQ ID NO: 15)、BspAA02831 (SEQ ID NO: 22)、SWT4 (SEQ ID NO: 25)、SWT22 (SEQ ID NO: 28)、SWT32 (SEQ ID NO: 31)、SWT40 (SEQ ID NO: 34)、SWT41 (SEQ ID NO: 37)、SWT77 (SEQ ID NO: 40)、SWT123 (SEQ ID NO: 43)、解淀粉芽孢杆菌 (NCBI 登陆号: CAA24990)、迟缓芽孢杆菌 (NCBI 登陆号: P29599)、芽孢杆菌属 (*B. sp.*) *SprC* (NCBI 登陆号: AAC43580)、地衣芽孢杆菌 (NCBI 登陆号: CAJ70731)、芽孢杆菌属 (*B. sp.*)

NN018132(SEQ ID NO:17)和博戈里亚芽孢杆菌(*B.bogoriensis*)(W02012175708A2的SEQ ID NO:4)、博戈里亚芽孢杆菌(*B.bogoriensis*)(NCBI登陆号:WP026675114.1)、*B.timonensis*(NCBI登陆号:WP010283106)、以及*P.dendritiformis*(NCBI登陆号:WP006679321)。将序列输入Vector NTI Advance工具包中并且使用邻接法(NJ)形成指南树(Saitou和Nei, Mol Biol Evol, 4:406-425, 1987)。使用以下参数计算树构建:Kimura序列距离校正并忽略具有空位的位置。AlignX将计算的距离值在括号中显示在图11所示的系统发育树上所显示的分子名称之后。BspAG00296、BspM04033、BspW01765、BspAA02831、SWT4、SWT22、SWT32、SWT40、SWT41、SWT77、SWT123、芽孢杆菌属(*B.sp.*)NN018132(SEQ ID NO:17)和博戈里亚芽孢杆菌(*B.bogoriensis*)(W02012175708A2的SEQ ID NO:4)、博戈里亚芽孢杆菌(*B.bogoriensis*)(NCBI登陆号:WP026675114.1)、*B.timonensis*(NCBI登陆号:WP010283106)、以及*P.dendritiformis*(NCBI登陆号:WP006679321)枯草杆菌蛋白酶全部聚集在相同区域(如图11所示)中以形成WHY-进化枝。BspM04033、SWT4、SWT32和SWT77枯草杆菌蛋白酶全部聚集在相同子区域(如图11所示)中以形成SWT77-进化枝。BspAG00296枯草杆菌蛋白酶聚集在子区域(如图11所示)中以形成BspAG00296-进化枝。BspAA02831、SWT40和WP026675114枯草杆菌蛋白酶全部聚集在相同子区域(如图11所示)中以形成WP026675114-进化枝。BspW01765、SWT41、SWT123和SWT22枯草杆菌蛋白酶全部聚集在相同子区域(如图11所示)中以形成SWT22-进化枝。

[0386] 实施例14

[0387] WHY-进化枝枯草杆菌蛋白酶的独特特征结构

[0388] 进行以下蛋白酶的基于结构的比对:BspAG00296(SEQ ID NO:4)、BspM04033(SEQ ID NO:11)、BspW01765(SEQ ID NO:15)、BspAA02831(SEQ ID NO:22)、SWT4(SEQ ID NO:25)、SWT22(SEQ ID NO:28)、SWT32(SEQ ID NO:31)、SWT40(SEQ ID NO:34)、SWT41(SEQ ID NO:37)、SWT77(SEQ ID NO:40)、SWT123(SEQ ID NO:43)、得自解淀粉芽孢杆菌的BPN'枯草杆菌蛋白酶(pdb登录号2STI)、得自地衣芽孢杆菌的Carlsberg(pdb登录号1CSE)、迟缓芽孢杆菌枯草杆菌蛋白酶(pdb登录号1JEA)、芽孢杆菌属(*B.sp.*)NN018132(SEQ ID NO:17)、博戈里亚芽孢杆菌(*B.bogoriensis*)(W02012175708-004)、博戈里亚芽孢杆菌(*B.bogoriensis*)(NCBI登录号:WP026675114)、解甘露糖醇芽孢杆菌(*B.mannanilyticus*)(NCBI登陆号:WP025025887)、*B.timonensis*(NCBI登陆号:WP010283106)、以及*P.dendritiformis*(NCBI登陆号:WP006679321),所述比对使用Molecular Operating Environment(MOE)软件(Chemical Computing Group, Montreal, Quebec, Canada)中的“比对”操作以寻找结构相似性,并且示于图12中。该比对向常规的序列比对应用保守结构基序作为另一向导。使用2012年10月发行的MOE中存在的标准程序默认值来进行该比对。

[0389] 如图12所示,枯草杆菌蛋白酶BspAG00296、BspM04033、BspW01765、BspAA02831、SWT4、SWT22、SWT32、SWT40、SWT41、SWT77、SWT123、芽孢杆菌属(*B.sp.*)NN018132(SEQ ID NO:17)、博戈里亚芽孢杆菌(*B.bogoriensis*)(W02012175708-004)、博戈里亚芽孢杆菌(*B.bogoriensis*)(NCBI登陆号:WP026675114)、解甘露糖醇芽孢杆菌(*B.mannanilyticus*)(NCBI登陆号:WP025025887)、*B.timonensis*(NCBI登陆号:WP010283106)、以及*P.dendritiformis*(NCBI登陆号:WP006679321)序列的结构比对示出相对于下列枯草杆菌蛋白酶的序列的一个插入和两个缺失的共同模式:得自解淀粉芽孢杆菌的BPN'、得自地衣

芽孢杆菌的Carlsberg和得自迟缓芽孢杆菌的枯草杆菌蛋白酶,对此可获得三维结构(pdb登录号分别为2ST1、1CSE和1JEA)。1JEA和1CSE结构中的残基编号相对于枯草杆菌蛋白酶BPN';然而BspM04033和所示的所有其它蛋白酶的残基的编号是连续线性序列。

[0390] 这些WHY-进化枝枯草杆菌蛋白酶共享足够的特征结构以形成进化枝,随后被称为WHY-进化枝,其中术语WHY来源于N端附近的完全保守残基WHY(BspM04033和该进化枝的其它成员中的W残基位置7)。此外,WHY-进化枝枯草杆菌蛋白酶与迟缓芽孢杆菌枯草杆菌蛋白酶共享共同的缺失,并因此其结构将作为参考用于理解WHY-进化枝枯草杆菌蛋白酶的不同特性的可能结果。除了*P.dendritiformis*、*B.manannilyticus*、*B.timonensis*枯草杆菌蛋白酶之外,该进化枝的所有其它成员均具有在对应于图13上所示基序内的BspM04033中45-47的位置处的保守残基NLV。这些WHY-进化枝枯草杆菌蛋白酶的其它显著的共享特征结构为:在基序内缺失1之后的序列VQG(BspM04033中的残基63-65)(图13)和缺失2之后的序列VSG(BspM04033中的残基80-82)。这种独特序列区域的汇集赋予WHY-进化枝与肽酶S8枯草杆菌蛋白酶家族的其它商业酶的显著不同。

[0391] 在图13中,WHY-进化枝基序被催化残基D33和H66包括(残基根据BspM04033线性序列编号)。所有丝氨酸蛋白酶共同的催化三联体由Asp(D)33、His(H)66和Ser(S)216组成。

[0392] 当与迟缓芽孢杆菌枯草杆菌蛋白酶和其它商业枯草杆菌蛋白酶相比时,D33-H66基序结合了存在于所有WHY-进化枝序列中的共同插入(插入1)和缺失(缺失1)。插入1导致用BspM04033中的HQSLANLVNTSLG(40-52)替换迟缓芽孢杆菌枯草杆菌蛋白酶中的残基HPDLNIRGG(39-47)。缺失1导致用BspM04033中的残基VGGSTMDVQGH(56-66)替换迟缓芽孢杆菌枯草杆菌蛋白酶中的残基VPGEPTQDGNH(51-64)。在pdb登录号1JEA中,编号相对于枯草杆菌蛋白酶BPN'。相对于枯草杆菌蛋白酶BPN',迟缓芽孢杆菌和Carlsberg枯草杆菌蛋白酶具有在不同序列位置处出现的单个残基缺失(参见图13)。

[0393] 在上述WHY-进化枝基序外部,发现WHY-进化枝酶中的第二共同缺失(缺失2)。在该情况下,迟缓芽孢杆菌枯草杆菌蛋白酶的残基VAGTIAALNNSIGVLGVA PSAELYAVKV(68-95)被BspM04033中的VAGTIASVSGVSMHNATL VPVKV(70-94)替换。在枯草杆菌蛋白酶中,诸如迟缓芽孢杆菌以及在BPN'和Carlsberg中,该区域形成保守钙结合位点。存在于BspM04033序列的缺失2中的残基改性可导致相应钙结合位点的损失。

[0394] 图14使用迟缓芽孢杆菌枯草杆菌蛋白酶结构示出WHY-进化枝的成员的结构的模型。WHY-进化枝基序片段使用迟缓芽孢杆菌结构作为参考(呈浅灰色)用黑色高亮。所有丝氨酸蛋白酶共同的催化三联体的Asp(D)33和His(H)66残基侧链以棒状物示出。通过箭头还示出环的并置,其中预计出现这两个缺失和一个插入。

[0395] 在迟缓芽孢杆菌枯草杆菌蛋白酶中,连同BPN'和Carlsberg枯草杆菌蛋白酶,包括残基70-94的片段形成扩展环以连同存在于这些枯草杆菌蛋白酶的N端处的残基Asp 41和残基Gln2一起形成紧密的钙结合位点。该钙结合位点是这些枯草杆菌蛋白酶的整体部分。通过突变除去该钙结合位点显著降低枯草杆菌蛋白酶BPN'中的稳定性(Bryan等人,1991Biochemistry 31 4937-4945)。由于缺失2,预计包含BspM04033枯草杆菌蛋白酶中的残基TIASVSGV(73-83)的减少的环区域将不再结合钙。值得注意的是Asp41在插入1的区域中出现,并且在WHY-进化枝枯草杆菌蛋白酶的N端处的序列与成熟蛋白质的N端处的其它枯草杆菌蛋白酶是基本上趋异的(图13)。在此假定插入1连同基本上不同的WHY-进化枝蛋

白酶的N端序列(图13)将赋予蛋白质稳定性以补偿前述钙结合环的去除。已知BPN'、Carlsberg和迟缓芽孢杆菌枯草杆菌蛋白酶通过由缺失2消除的环中的钙结合而强稳定化。因为发现在包含螯合剂诸如EDTA的洗涤剂存在下,WHY-进化枝枯草杆菌蛋白酶相对于这些其它枯草杆菌蛋白酶稳定,所以该非常具有吸引力的特征结构可能是存在于WHY-进化枝枯草杆菌蛋白酶中的独特序列基序与缺失2和改变的N端组合的结果。最后,发现其它共同缺失,缺失1在另一个相邻环中出现(图14),并且可假定调节蛋白质主链折叠以补偿由插入1和缺失2赋予的变化。从图14明显的是,插入1在将由于缺失2而减少的另一个环附近的环中出现。假定在WHY-进化枝的三维结构中,膨胀的环将补偿由缺失2形成的腔。

[0396] 下文列出了本文所报道的WHY-进化枝酶的最稳定成员BspM04033,和其它前述成员芽孢杆菌属(*B. sp.*)NN018132之间的残基差异:P3N、Q6R、N10E、T20I、S26N、I28R、Q29I、H38A、Q41P、S42N、A44R、N48D、Q53R、S59G、M61G、H85Q、T88R、V90I、N96G、S98N、L103M、F104Y、T107Q、S113A、D115S、G117N、N131D、Q132S、S133D、A136N、A137N、A138I、Q139N、N143S、A144S、S146T、I147L、A157R、S168N、V169A、T178N、G179R、A180T、V204Y、N207G、G208Q、Y209F、A210R、F212L、S219T、A222V、N229I、R230K、A231S、V231A、S239T、N240Q、A241V、S243N、M245L、Q246R、N247D、P255L、T256N、F257Q、D264N、N266Y、Q271A和S273G。

[0397] 实施例15

[0398] WHY-进化枝枯草杆菌蛋白酶的晶体结构

[0399] 使用X-射线晶体学测定两个截短的WHY-进化枝结构的三维结构。解释了纯化的BspAG00296(SEQ ID NO:4,273个氨基酸)和由截短形式的SWT77-tr(SEQ ID NO:44,273个氨基酸)蛋白质组成的纯化SWT77的结构。两种蛋白质共享WHY-进化枝基序,但具有仅74.4%相同的线性氨基酸序列。

[0400] 分离并且结晶的截短的SWT77蛋白酶(SWT77-tr,273个氨基酸)的序列以SEQ ID NO:44示出:MHPNQWHYMNINAPQAWETTTGSSSVIQAVALDTGIDHNHQLANLVNTSLGQSFVGGSTMDVQGHGTHVAGTIASYSVSGVMHNATLVPVKVLNDSGSGSLFGITQGILYSADIGADVINMSLGGGGYNQSMAEAAQTAVDAGSIVIAASGNDGAGSISYPAAYSVIAVGSVTSTGARSNFSNYGSGLELMAPGSNIYSTVPNNGYATFSGTSMAPHAAGVAGLMRAVNSNLSVSDARSIMQNTAQYAGSPTFYGYGIVDANAAVQQAS。

[0401] 确定BspAG00296的结构呈空间群C2形式,其具有呈不对称单元形式的两个分子,其中晶胞尺寸 $a=111.1$, $b=63.3$ 且 $c=72.7\text{\AA}$,并且 $\beta=90.02^\circ$,精确至分辨率 $1.5\text{\AA}$ 。通过悬滴法获得结晶,所述方法由在pH 5.5的20mM乙酸钠缓冲液和0.15M NaCl中的1%蛋白质溶液起始。储液包含0.8M NH_4SO_4 、200mM MgCl_2 和0.1M Bis-Tris丙烷,pH 6.5。在Bruker X8Proteum系统(Bruker Axis Inc.,Madison,WI,USA)上收集数据。使用分子置换测定结构,其中迟缓芽孢杆菌枯草杆菌蛋白酶(pdb登录号1JEA)的坐标作为起始模型。使用程序COOT将BspAG00296的坐标拟合成所得的电密度(Emsley,P等人,Acta Cryst.D66 486-501(2010))。在拟合和再拟合调节之后,使用REFMAC程序利用CCP4软件工具包中的标准默认值对坐标进行精修。最终模型具有良好的立体化学,并且R-work为0.14且R-free为0.15,所有数据精确至 $1.5\text{\AA}$ 。

[0402] 确定SWT77-tr的结构呈空间群P21212形式,其具有呈不对称单元四个分子,其中晶胞尺寸 $a=149.3$, $b=80.1$ 且 $c=82.4\text{\AA}$,精确至分辨率 $0.188\text{\AA}$ 。通过悬滴法获得结晶,所述方法由pH 5.5的50mM乙酸钠和0.10M氯化钠和1mM PMSF中的20mg/mL SWT-77-tr蛋白

质原液起始。储液包含3.5M甲酸钠+0.10M Bicine,pH 9.0。在Bruker X8 Proteum上收集数据。使用分子置换确定结构,其使用BspAG00296结构的单体作为起始模型。使用Coot软件包拟合SWT77-tr的坐标,并且使用CCP4软件包系统中的REFMAC程序对模型进行精修。最终模型具有良好的立体化学,并且R-work为0.17且R-free为0.22,所有数据精确至1.88Å。

[0403] 将BspAG00296和SWT77-tr的单体的坐标与1567个共同原子的0.342Å的总体RMS叠加。虽然确定这两个结构呈包含不对称单元形式的两个或四个分子的不同空间群,但两个结构的总体折叠在实验误差范围内是相同的。这示于图15A中。

[0404] 这确认了WHY-进化枝中的变化消除了存在于其它已知商业枯草杆菌蛋白酶中的紧密钙位点。如从图15B中可见,当将SWT77-tr的主链折叠的示意图(黑色)与迟缓芽孢杆菌枯草杆菌蛋白酶(浅灰色)在包括缺失1和2的片段的区域中进行比较时,没有发现由SWT77-tr的残基76-81形成的环中的钙的电子密度,即在一定程度上也由插入1和N端片段中的改变补偿。

[0405] 当如图15B所示将结构叠加时,可清楚地看到由缺失2和N端引起的变化,所述变化消除了可见于其它蛋白酶中的钙结合,以及由插入1和缺失1引起的变化。

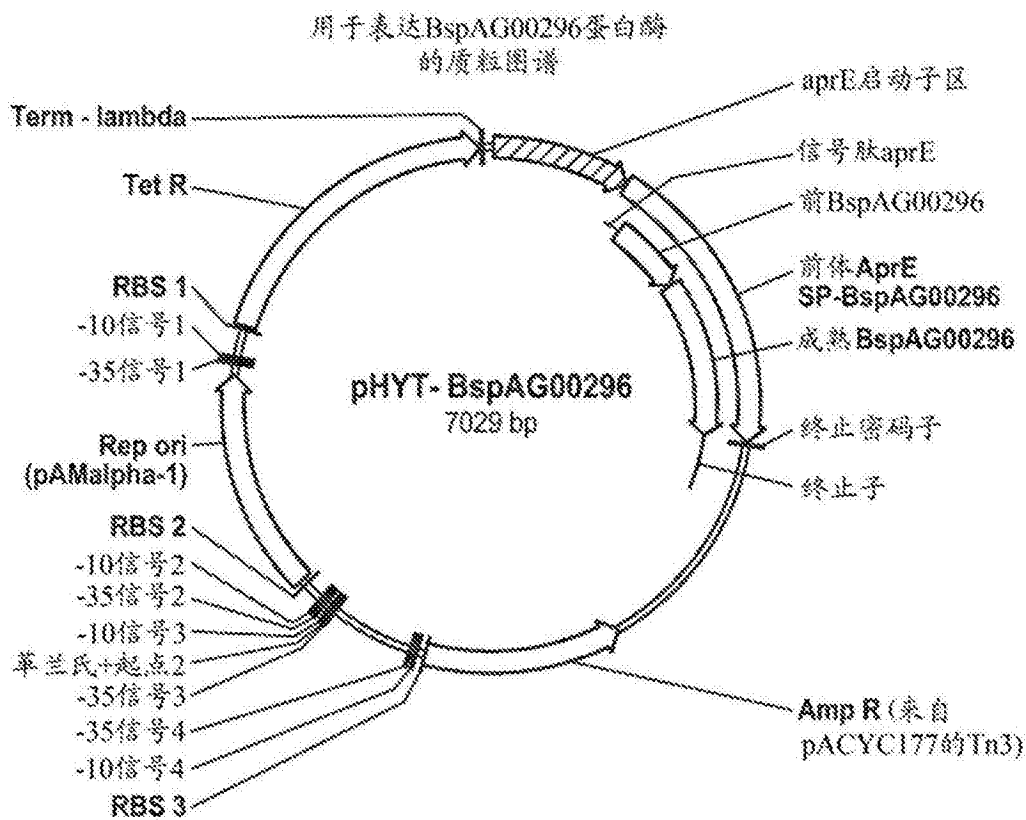


图1

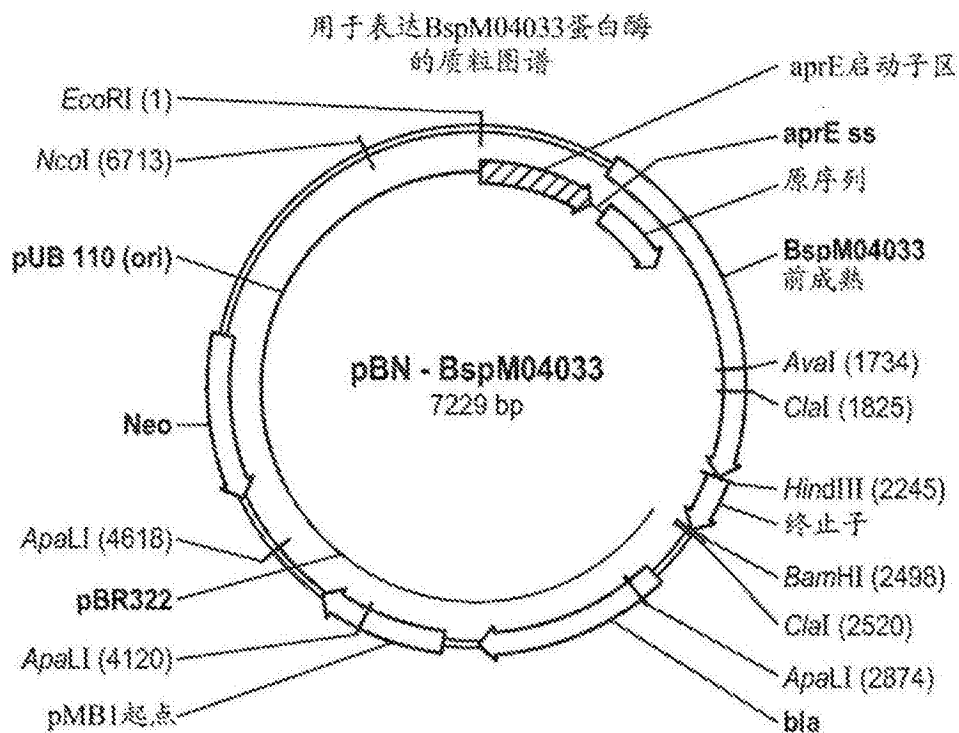


图2

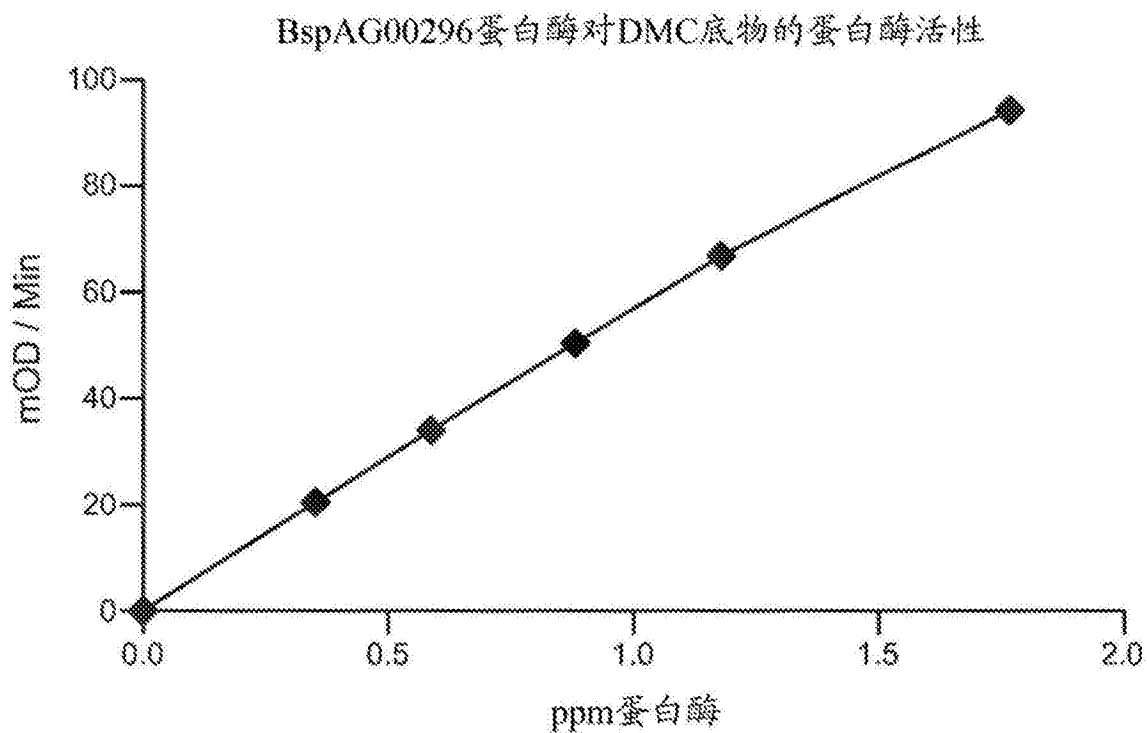


图3

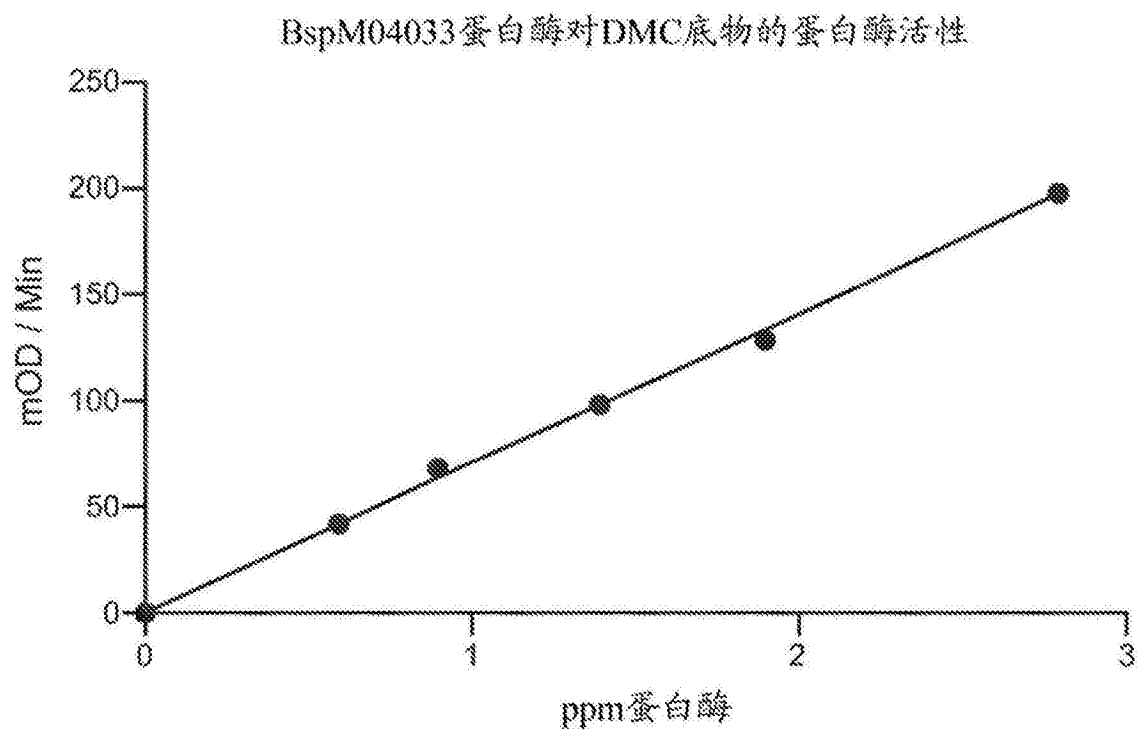


图4

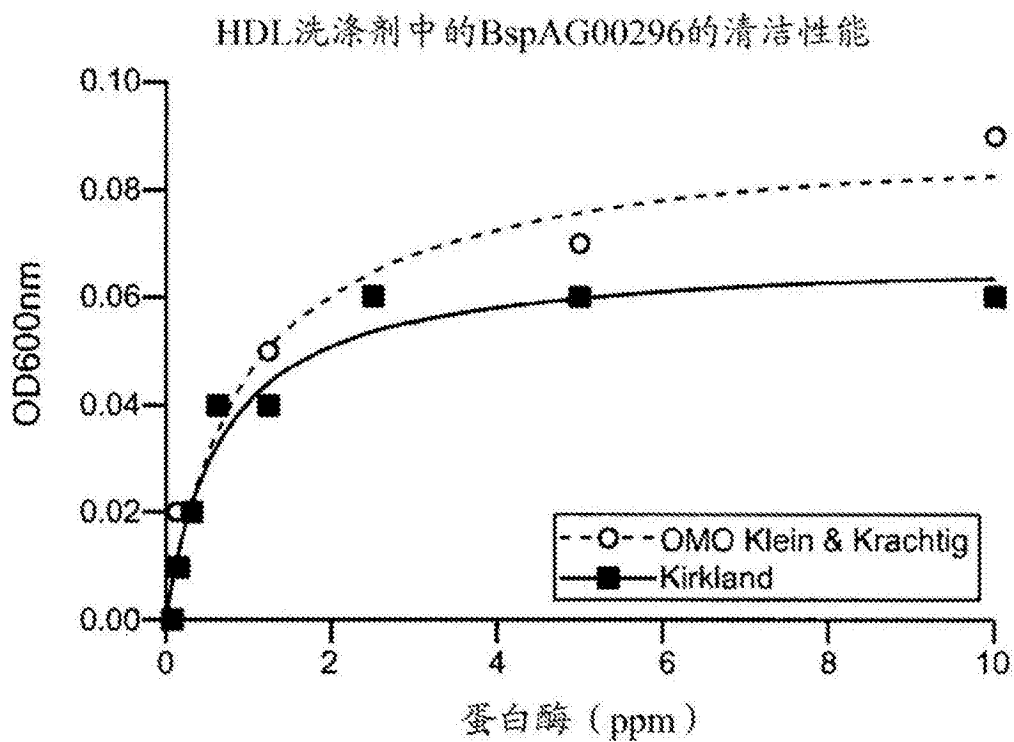


图5

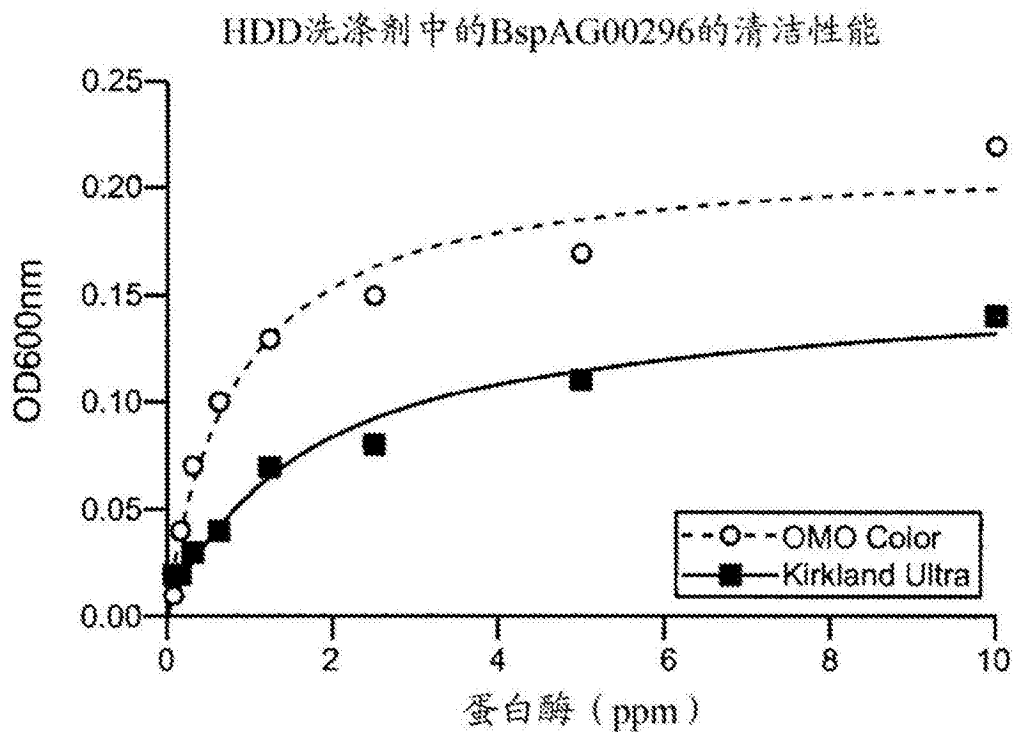


图6

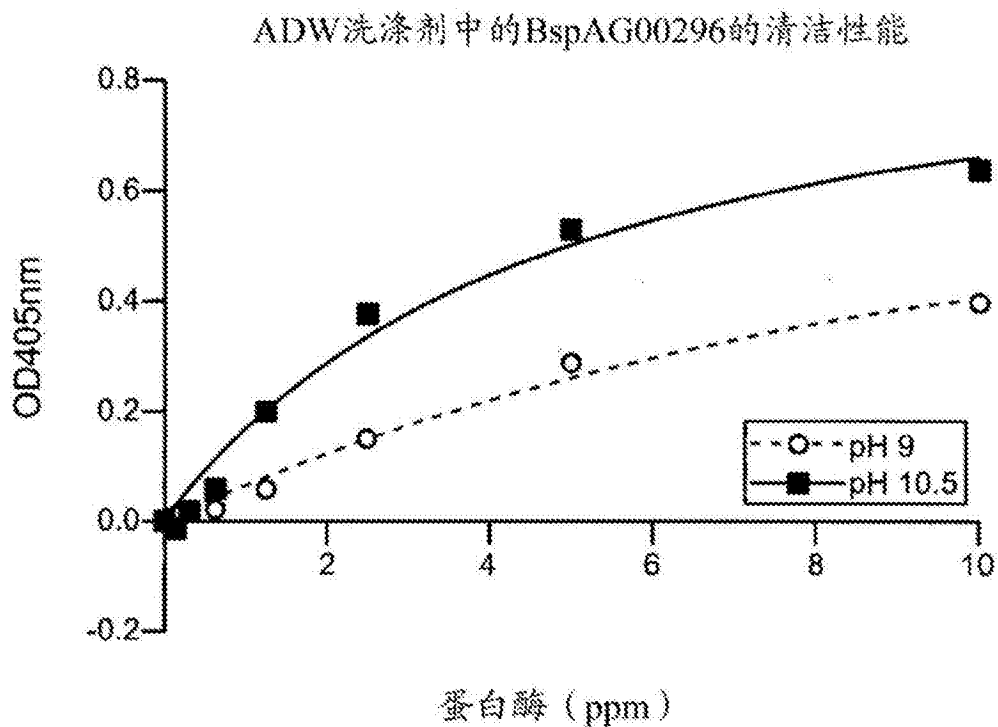


图7

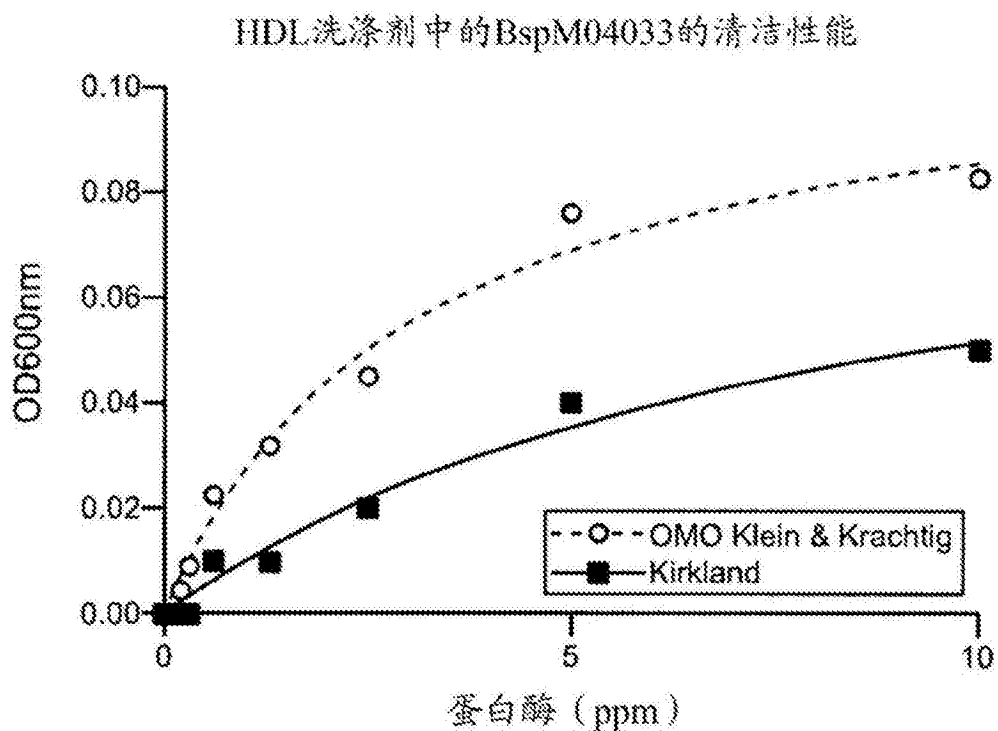


图8

HDD洗涤剂中的BspM04033的清洁性能

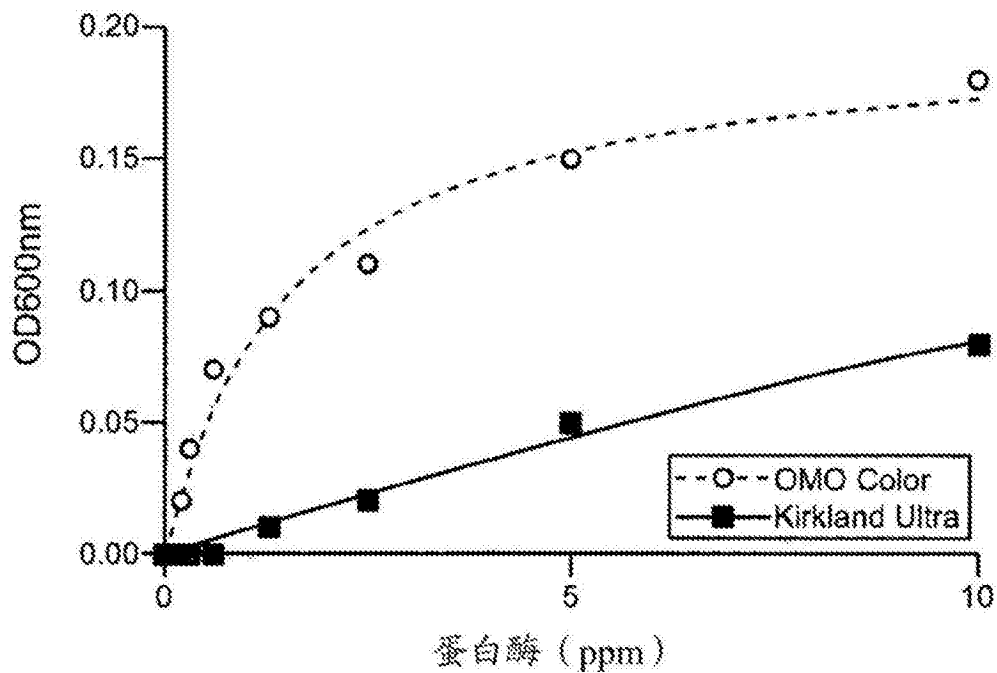


图9

ADW洗涤剂中的BspM04033的清洁性能

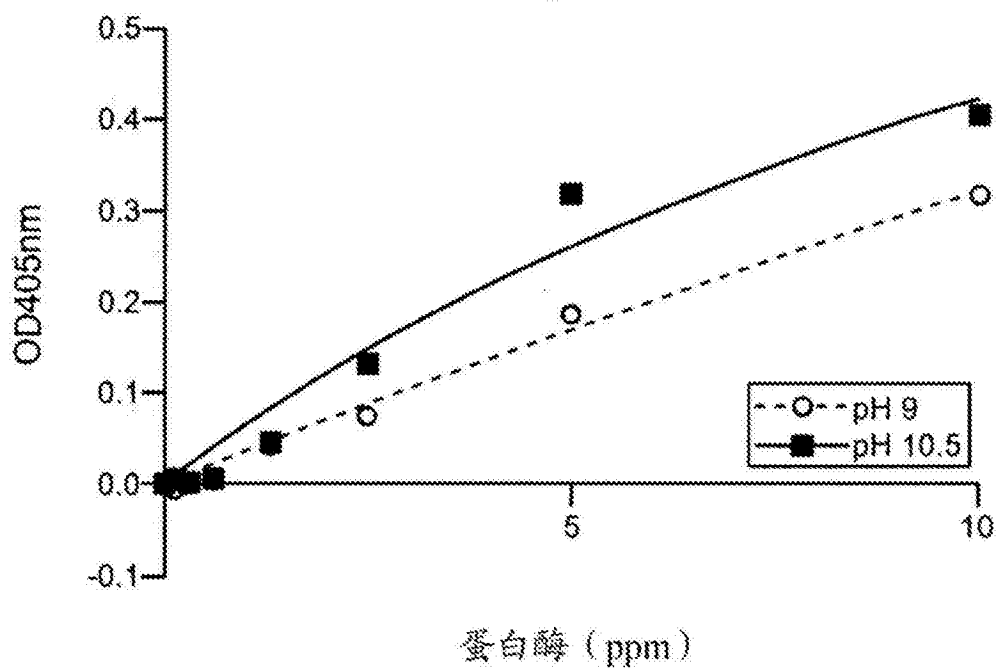


图10

BspMD4033	(51)	EGQSEVGG	SPNDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	100
BspAG00296	(51)	EGRSEVGG	TPADVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
BspAA02831	(51)	EGRSEVGG	CICDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
BspW01765	(51)	EGRSYVGG	SPEDVQQGCHGTHVAGTHIASYGA	VSCVMHQNATLIPVKV	
SWT40	(51)	EGRSEVGG	CICDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
SWT123	(51)	EGRSYVGG	SPEDVQQGCHGTHVAGTHIASYGA	VSCVMHQNATLIPVKV	
SWT22	(51)	EGRSYVGG	NPEDVQQGCHGTHVAGTHIASYGN	VSCVMHQNATLIPVKV	
SWT32	(51)	EGQSEVGG	SPNDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
SWT4	(51)	EGQSEVGG	SPNDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
SWT41	(51)	EGRSYVGG	SAEDVQQGCHGTHVAGTHIASYGA	VSCVMHQNATLIPVKV	
SWT77	(51)	EGQSEVGG	SPNDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
B. sp. NN018132	(51)	EGRSEVGG	CICDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
WO2012175708-0004	(51)	EGRSEVGG	CICDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
B. bogoriensis_WP_026675114.1	(51)	EGRSEVGG	CICDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
P_dendritiformis_WP006679321	(50)	EGSSEVGG	TINDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
B_mannanilyticus_WP_025025887.1	(51)	EGRSEVGG	TINDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	
B_timonensis_WP_010283106.1	(51)	EGKSEVGG	TINDVQQGCHGTHVAGTHIASYGT	VSCVMHQNATLIPVKV	
B. lentus_P29599	(45)	CGASEVPCPEP	STQPCNGCHGTHVAGTHIASYGL	VSCVMHQNATLIPVKV	
B_amyloliquefaciens_CAA24990	(50)	WVPSSTNP	F-QDNNSHGCHGTHVAGTHIASYGL	VSCVMHQNATLIPVKV	
Bacillus_sp_SprC_AAC43580	(46)	CGASEVSGEPNALQ	CNGCHGTHVAGTHIASYGL	VSCVMHQNATLIPVKV	
B_licheniformis_CAJ70731	(46)	CGASEVAGEA	YNTDCNGCHGTHVAGTHIASYGL	VSCVMHQNATLIPVKV	
共有序列	(51)	LGRSEVGG	STDVQQGCHGTHVAGTHIASYGS	VSCVMHQNATLIPVKV	

图12A-2

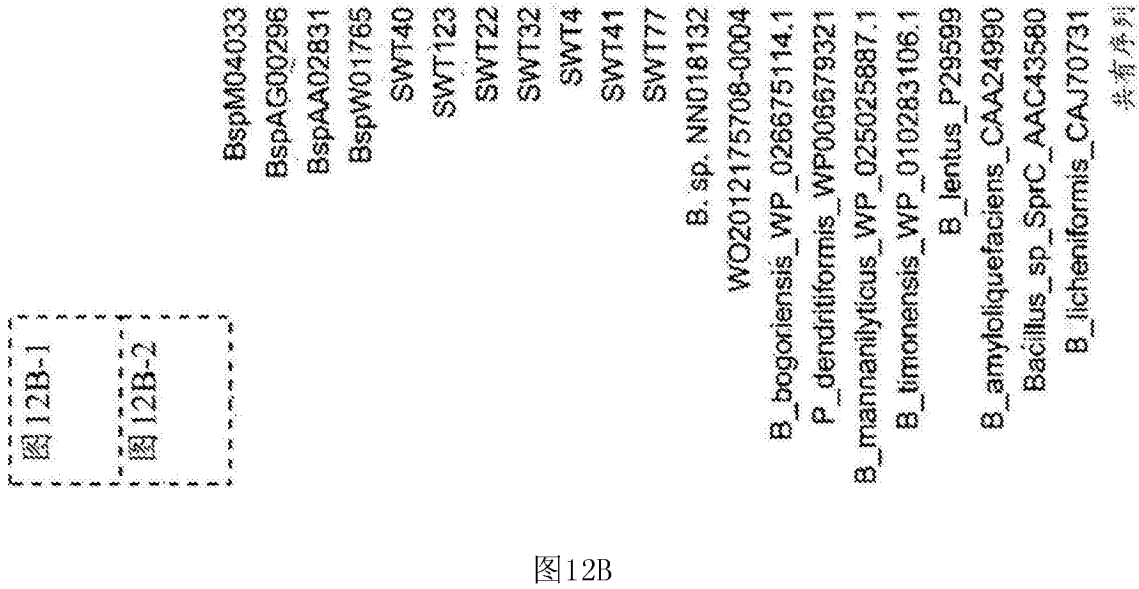


图12B

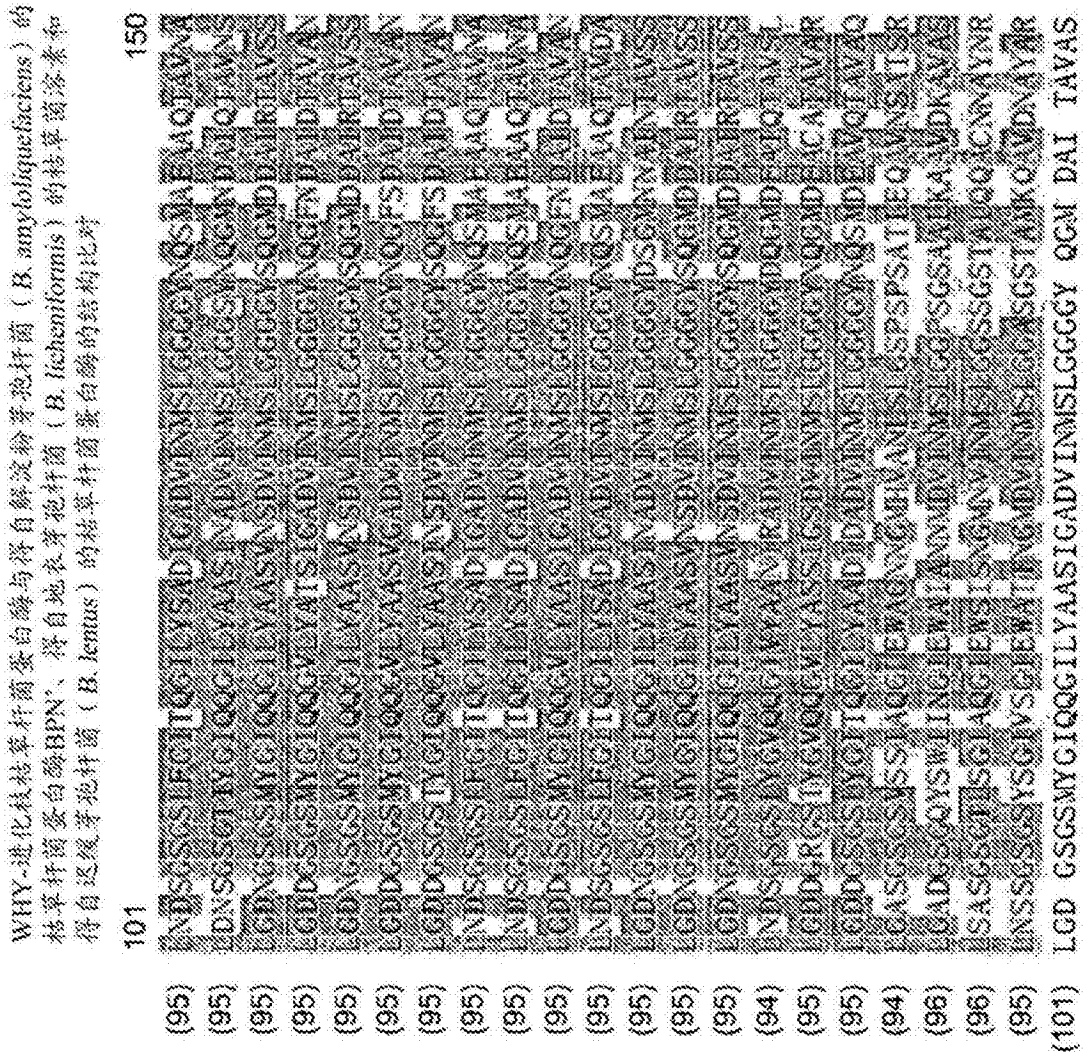


图12B-1

BspM04033	(145)	CSIVFAASGNDG	AGSISYPAAVSSVIAVGSVTSICARSNF	SVNYGSG	200
BspAG00296	(145)	CTIVFAASGNG	ASSISNPAAVSCAIVGCVTSIRIRSSF	SVNYGSG	
BspAA02831	(145)	CTIVFAASGNDG	RGSSISYPAAVSCAIVGCVTSIRIRSSF	SVNYGSG	
BspW01765	(145)	CSWVFAASGNDG	RASISYPAAVDCAIVGCVTSICARSNF	SVNYGSG	
SWT40	(145)	CTIVFAASGNDG	RGSSISYPAAVSCAIVGCVTSIRIRSSF	SVNYGSG	
SWT123	(145)	CTIVFAASGNDG	RASISYPAAVDCAIVGCVTSICARSNF	SVNYGSG	
SWT22	(145)	CTIVFAASGNDG	RASISYPAAVDCAIVGCVTSICARSNF	SVNYGSG	
SWT32	(145)	CSIVFAASGNDG	AGSISYPAAVSSVIAVGSVTSICARSNF	SVNYGSG	
SWT4	(145)	CSIVFAASGNDG	AGSISYPAAVSSVIAVGSVTSICARSNF	SVNYGSG	
SWT41	(145)	CSWVFAASGNDG	RASISYPAAVDCAIVGCVTSICARSNF	SVNYGSG	
SWT77	(145)	CSIVFAASGNDG	AGSISYPAAVSSVIAVGSVTSICARSNF	SVNYGSG	
B. sp. NN018132	(145)	CTIVFAASGNDG	RGSSISYPAAVSCAIVGCVTSIRIRSSF	SVNYGSG	
WO2012175708-0004	(145)	CSWVFAASGNDG	RGSSISYPAAVSCAIVGCVTSIRIRSSF	SVNYGSG	
B_bogoriensis_WP_026675114.1	(145)	CTIVFAASGNDG	RGSSISYPAAVSCAIVGCVTSIRIRSSF	SVNYGSG	
P_dendritiformis_WP006679321	(144)	CTIVFAASGNDG	RPSISYPAAVSCAIVGCVTSIRIRSSF	SVNYGSG	
B_mannanilyticus_WP_025025887.1	(145)	CTIVFAASGNDG	RGSSISYPAAVSCAIVGCVTSIRIRSSF	SVNYGSG	
B_timonensis_WP_010283106.1	(145)	CTIVFAASGNDG	ASSISYPAAVSSVIAVGSVTSIRIRSSF	SVNYGSG	
B_lentus_P29599	(144)	GVWVFAASGNSG	AGSISYPAAVNAHVCAIDQNNRASF	SVNYGAG	
B_amyloliquefaciens_CAA24990	(146)	GVWVFAASGNEGTSGSSSTVGVCKPSPVIVGAVDSSNCRASF	SVNYGPE		
Bacillus_sp_SprC_AAC43580	(146)	CTIVFAASGNSGNDG	TMCPARYSSVIAVCAVSSNNIRASF	SVNYGSE	
B_licheniformis_CAJ70731	(145)	GVWVFAASGNSGNDG	TMCPARYSSVIAVCAVSSNNIRASF	SVNYGSE	
共有序列	(151)	CTIVFAASGNDG	RGSSISYPAAVSSVIAVGSVTSIRIRSSF	SVNYGSG	

图12B-2

		251			300
BspM04033	(241)	ARSHMQNEAQVAGCSPTFYQYCGIVDA	AAVQQA	SGCG	
BspAG00296	(241)	VRHILRNTAQVAGCSPTQYCGIVDA	AAVLSAR		
BspAA02831	(241)	ARSHQNTAQVAGCSFNQYCGIVDA	AAVRAR		
BspW01765	(241)	VRNHLADIQAQVAGCSHQYQYCGIVDA	FAVQAG	SGCG	
SWT40	(241)	ARSHQNTAQVAGCSFNQYCGIVDA	AAVRAR	QTE	
SWT123	(241)	VRSHELADFAQVAGCSFYQYCGIVDA	FAVQAG	SGCG	
SWT22	(241)	VRCHLADFAQVAGCSHQYQYCGIVDA	AAVQAG	SGCG	
SWT32	(241)	ARSHMQNTAQVAGCSPTFYQYCGIVDA	AAVQQA	SGCG	
SWT4	(241)	ARSHMQNTAQVAGCSPTFYQYCGIVDA	AAVQQA	SGCG	
SWT41	(241)	VRNHLADIQAQVAGCSHQYQYCGIVDA	AAVQAG	SGCG	
SWT77	(241)	ARSHMQNTAQVAGCSPTFYQYCGIVDA	AAVQQA	SGCG	
B. sp. NN018132	(241)	VRNHLRDTAQVAGCSNQYQYCGIVDA	AAVQAG		
WO2012175708-0004	(241)	ARTHEL RNTAQVAGCSFNQYQYCGIVDA	AAVRAR		
B. bogoriensis_WP_026675114.1	(241)	ARSHQNTAQVAGCSFNQYQYCGIVDA	AAVRAR	SGSQQPSYE	---TN
P. dendritiformis_WP006679321	(240)	AGDIURNTAQPNAGSSDQYQYCHGIVDA	HAAMLAA	AGCDTPAPS	---AP
B. mannaniyiticus_WP_025025887.1	(241)	ARNHLRNTAQVAGCSFNQYQYCGIVDA	HAAMLAA	AGCDNEDPNT	---TT
B. timonensis_WP_010283106.1	(241)	ARQHRLRDTAQEAGCSFTQYQYCGIVDA	HAAMVAA	SGCGGCTPPPPISDIP	
B. lentus_P29599	(240)	IRNHEKNTATSLGSLNAGSGLNAGNATR			
B. amyloliquefaciens_CAA24990	(246)	VRSSLENITIKLGDSFYNGKGLINVQAAQ			
Bacillus_sp_SprC_AAC43580	(246)	IRERKNTATINLGDPPFYQYKGVNVEAALQ			
B. licheniformis_CAJ70731	(245)	VRNRESSTATINLGSFYQYKGLNVEAAQ			
共有序列	(251)	ARSILNTAQVAGCSQYQYCGIVDA	AAVQAA	G	

图12C-2

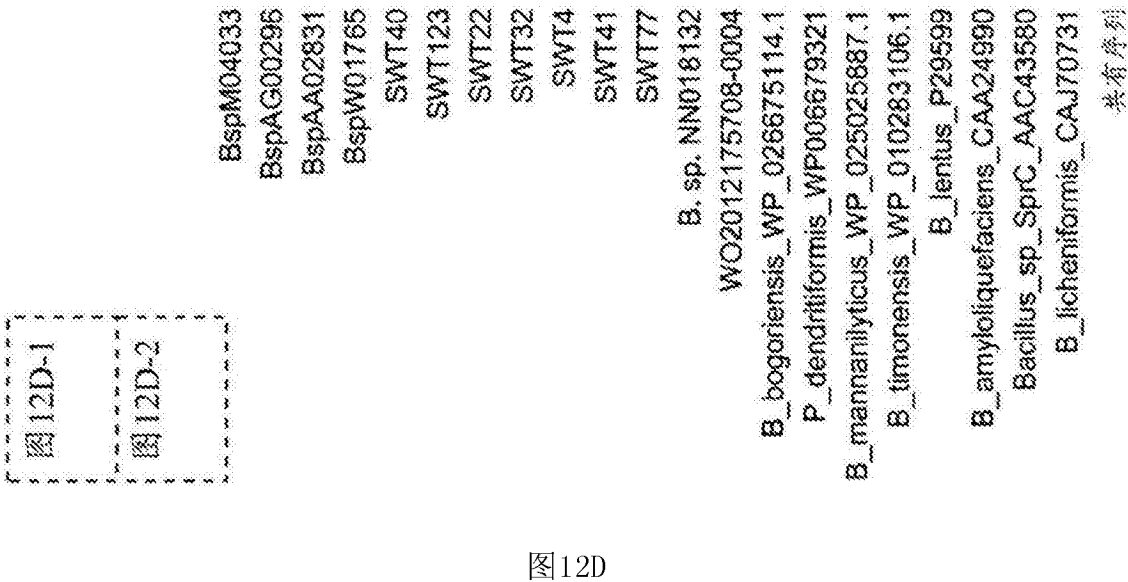


图12D

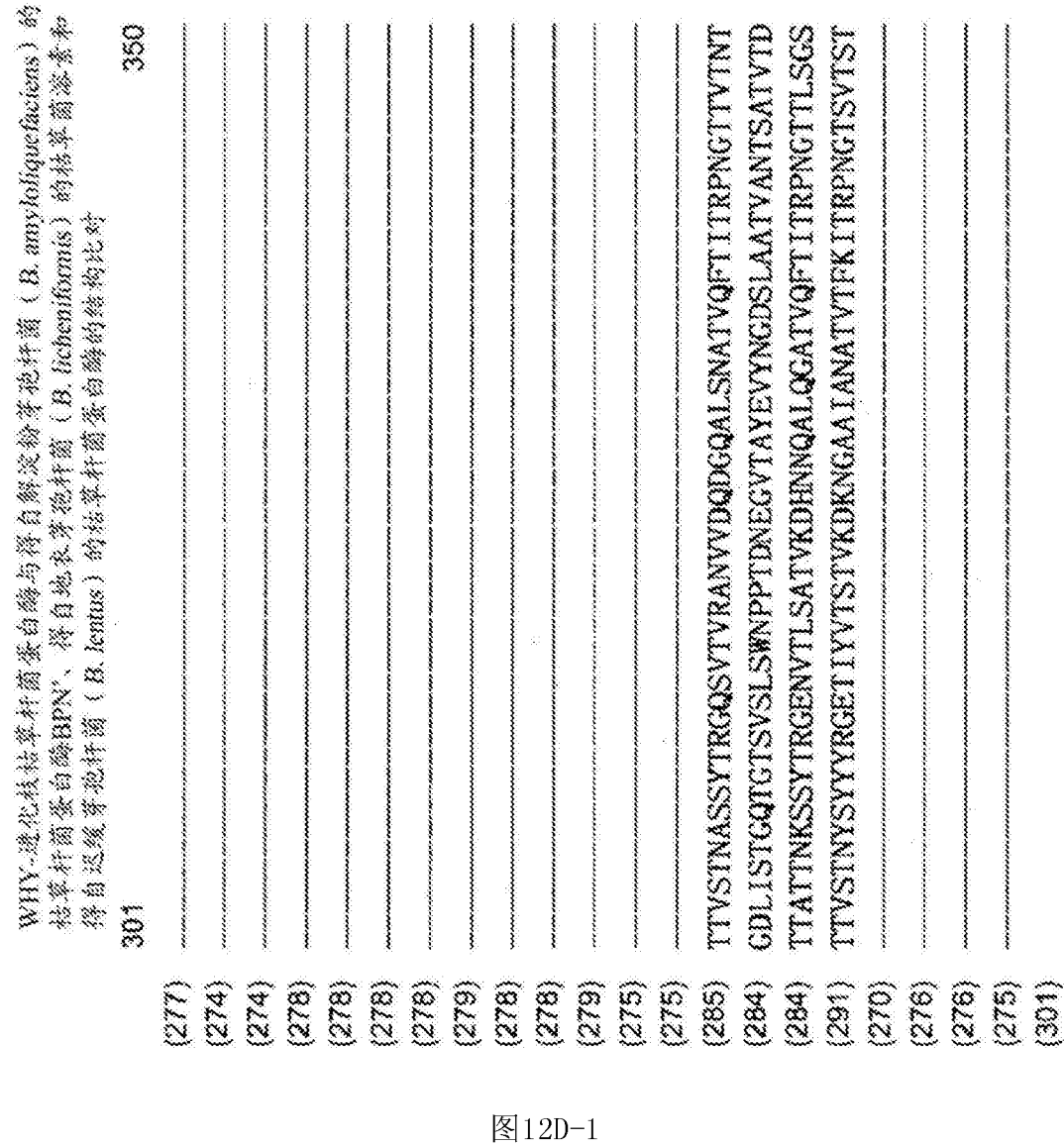


图12D-1

	351	400
BspM04033	(277)	
BspAG00296	(274)	
BspAA02831	(274)	
BspW01765	(278)	
SWT40	(278)	
SWT123	(278)	
SWT22	(278)	
SWT32	(279)	
SWT4	(278)	
SWT41	(278)	
SWT77	(279)	
B. sp. NN018132	(275)	
WO2012175708-0004	(275)	
B_bogoriensis_WP_026675114.1	(335)	ATTNNSGVATWTIATSSSTARGTYGVQAATSLSGYEGSTATTRFSVN
P_dendritiformis_WP006679321	(334)	LTADTTYTFTVRAVDASGNRSEASNAVTVTITDSDSSQPSPTWAPGISYKI
B_mannanilyticus_WP_025025887.1	(334)	ATTNASGVASWTIVSTSEYTAIGTYQVQATSSKSGYSGSGTTSFSVR
B_timonensis_WP_010283106.1	(341)	GTNSSGVATWSIGTNYTATGTYQVQDATAASKSGYTTSTASTFKMY
B_lentus_P29599	(270)	
B_amyloliquefaciens_CAA24990	(276)	
Bacillus_sp_SprC_AAC43580	(276)	
B_licheniformis_CAJ70731	(275)	
共有序列		(351)

图12D-2

WHY-进化枝枯草杆菌蛋白酶与得自解淀粉芽孢杆菌 (*B. amyloliquefaciens*) 的
枯草杆菌蛋白酶BPN[®]、得自地衣芽孢杆菌 (*B. licheniformis*) 的枯草菌溶素和
得自迟缓芽孢杆菌 (*B. lentus*) 的枯草杆菌蛋白酶的结构比较

	401	435
BspM04033	(277)	SEQ ID NO:47
BspAG00296	(274)	SEQ ID NO:48
BspA002831	(274)	SEQ ID NO:49
BspW01765	(278)	SEQ ID NO:50
SWT40	(278)	SEQ ID NO:51
SWT123	(278)	SEQ ID NO:52
SWT22	(278)	SEQ ID NO:53
SWT32	(279)	SEQ ID NO:54
SWT4	(278)	SEQ ID NO:55
SWT41	(278)	SEQ ID NO:56
SWT77	(279)	SEQ ID NO:57
B. sp. NN018132	(275)	SEQ ID NO:58
WO2012175708-0004	(275)	SEQ ID NO:59
B_bogoriensis_WP_026675114.1	(382)	SEQ ID NO:60
P_dendritiformis_WP006679321	(384)	SEQ ID NO:61
B_mannanlyticus_WP_025025887.1	(381)	SEQ ID NO:62
B_timonensis_WP_010283106.1	(388)	SEQ ID NO:63
B_lentus_P29599	(270)	SEQ ID NO:64
B_amyloliquefaciens_CAA24990	(276)	SEQ ID NO:65
Bacillus_sp_SprC_AAC43580	(276)	SEQ ID NO:66
B_licheniformis_CAJ70731	(275)	SEQ ID NO:67
共有序列	(401)	SEQ ID NO:68
GEEVYGEATYQCLQEHISMAGWEPLNVPALWLEK		

图12E

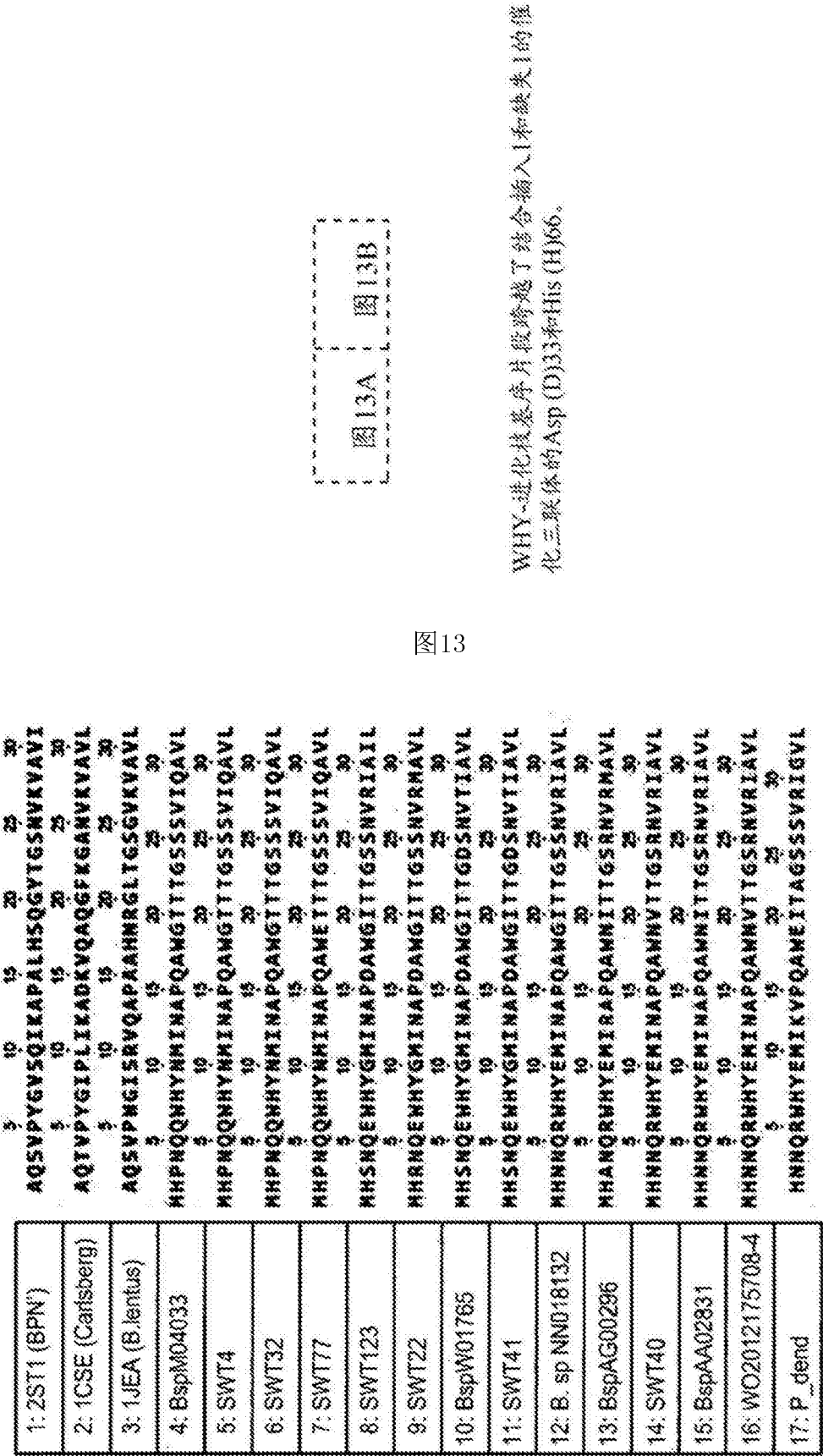


图13A



WHY-进化枝氨基酸序列跨越了结合插入1和缺失1的进化三联体的Asp (D)33和His (H)66。

图13A

存在于WHY-进化枝枯草杆菌蛋白酶中的序列基序变化的潜在结构性后果。使用最接近的已知枯草杆菌蛋白酶结构，使用迟缓芽孢杆菌枯草杆菌蛋白酶结构作为参考（呈浅灰色）将基序片段以黑色高亮。所有丝氨酸蛋白酶共同的催化三联体的Asp (D)33和His (H)66残基侧链以棒状物示出。还由箭头指示了提出进行两个缺失和一个插入的位置。插入1发生在邻近预期发生缺失2的环的环中。预期缺失1改变相邻的另一个环使得该环具有插入1。因为与缺失2相关联的环连同在N-端片段上的残基2和被插入环中的残基一起形成钙结合位点，所以其容易消除钙环。

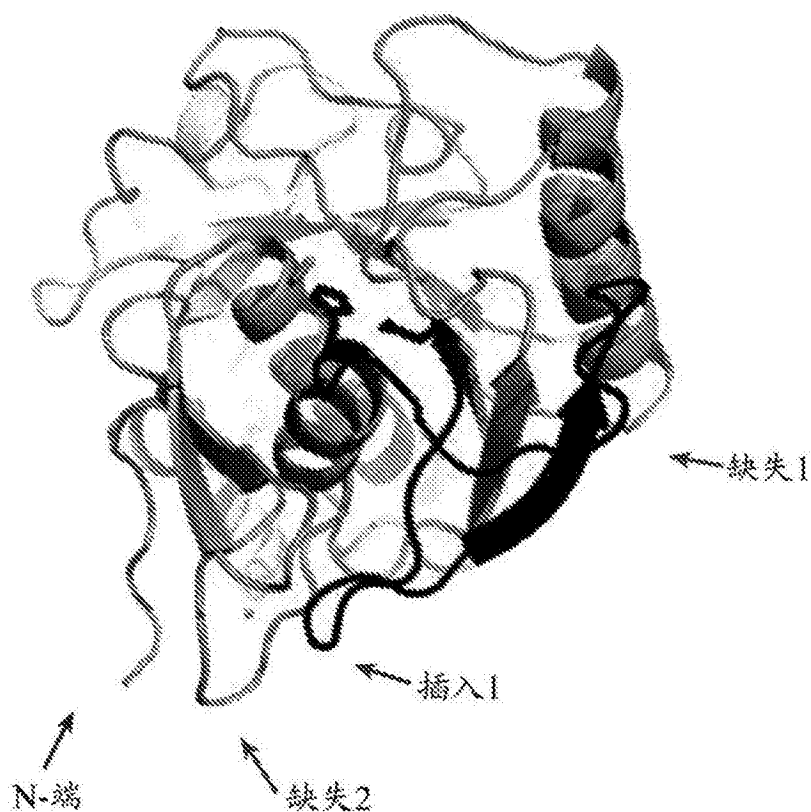


图14

示意性示出来自BspAG00296（深灰）和SWT77-tr（黑）的晶体结构的单体的叠加。



图15A

当SWT77-tr（呈黑色）的结构与迟缓芽孢杆菌枯草杆菌蛋白酶（呈浅灰色）相比时，发现的序列基序变化的结构图像。

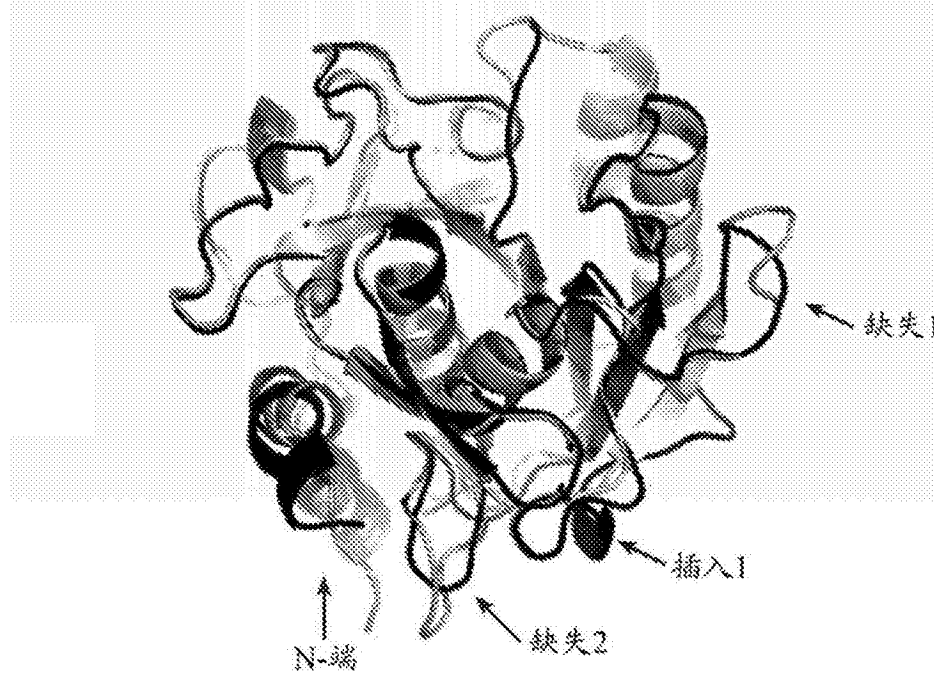


图15B