



(12) 发明专利

(10) 授权公告号 CN 101855342 B

(45) 授权公告日 2013. 07. 10

(21) 申请号 200880115770. 3

(51) Int. Cl.

(22) 申请日 2008. 09. 13

C12N 9/04 (2006. 01)

(30) 优先权数据

60/972, 058 2007. 09. 13 US

(56) 对比文件

WO 02086126 A1, 2002. 10. 31, 全文.

WO 2005017135 A1, 2005. 02. 24, 全文.

(85) PCT申请进入国家阶段日

2010. 05. 12

审查员 罗霄

(86) PCT申请的申请数据

PCT/US2008/076333 2008. 09. 13

(87) PCT申请的公布数据

W02009/036404 EN 2009. 03. 19

(73) 专利权人 科德克希思公司

地址 美国加利福尼亚州

(72) 发明人 杰克·梁 史提芬尼·J·珍妮

艾米丽·穆德弗 沙琳·奇恩

约翰·M·格鲁伯 安克·克莱伯

古伽特·W·哈思曼

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

代理人 杨淑媛 郑霞

权利要求书6页 说明书69页

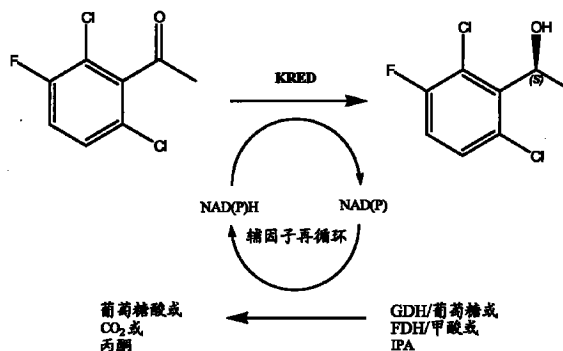
序列表112页 附图1页

(54) 发明名称

用于还原苯乙酮的酮还原酶多肽

(57) 摘要

本公开提供了与天然存在的野生型酮还原酶相比具有改进特性的工程酮还原酶。还提供了编码工程酮还原酶的多核苷酸、能表达工程酮还原酶的宿主细胞和使用工程酮还原酶来合成多种手性化合物的方法。



1. 一种工程酮还原酶多肽,该多肽衍生自野生型乳杆菌酮还原酶,能将苯乙酮立体选择性地还原为(S)-1-苯乙醇,其中所述多肽在对应于SEQ ID NO:2、4或98的X190的残基处具有脯氨酸残基。

2. 如权利要求1所述的多肽,其中所述多肽能以高于90%的百分比立体异构过量将苯乙酮立体选择性地还原为(S)-1-苯乙醇。

3. 如权利要求1所述的多肽,其中所述多肽能以高于99%的百分比立体异构过量将苯乙酮立体选择性地还原为(S)-1-苯乙醇。

4. 如权利要求1所述的多肽,该多肽为能以高于85%的百分比立体异构过量将2',6'-二氯-3'-氟苯乙酮底物立体选择性地还原为产物(S)-1-(2,6-二氯-3-氟苯基)乙醇的多肽,该多肽包含基于SEQ ID NO:2、4或98的氨基酸序列,该多肽在对应于X190的残基处具有脯氨酸,并且

该多肽包含另外具有一种或多种以下特点的氨基酸序列:

对应于X7的残基是苏氨酸、脯氨酸、色氨酸、精氨酸、组氨酸或天冬酰胺;

对应于X16的残基是丝氨酸;

对应于X43的残基是异亮氨酸;

对应于X60的残基是丙氨酸;

对应于X94的残基是丙氨酸、缬氨酸或半胱氨酸;

对应于X95的残基是异亮氨酸或亮氨酸;

对应于X96的残基是丝氨酸、天冬酰胺、苏氨酸或谷氨酸;

对应于X97的残基是赖氨酸、苏氨酸、缬氨酸、精氨酸、甲硫氨酸或异亮氨酸;

对应于X120的残基是苯丙氨酸或缬氨酸;

对应于X125的残基是甘氨酸或丝氨酸;

对应于X142的残基是天冬酰胺;

对应于X147的残基是苯丙氨酸、亮氨酸、异亮氨酸、缬氨酸或谷氨酰胺;

对应于X149的残基是甘氨酸或苯丙氨酸;

对应于X150的残基是天冬氨酸或组氨酸;

对应于X152的残基是丝氨酸、苏氨酸或甲硫氨酸;

对应于X196的残基是缬氨酸、亮氨酸、甲硫氨酸、苯丙氨酸或异亮氨酸;

对应于X202的残基是丙氨酸、色氨酸、酪氨酸或甲硫氨酸;

对应于X205的残基是精氨酸;并且

对应于X206的残基是甲硫氨酸或酪氨酸。

5. 如权利要求4所述的多肽,其中所述百分比立体异构过量为高于90%。

6. 如权利要求4所述的多肽,其中所述百分比立体异构过量为高于99%。

7. 如权利要求5或6所述的多肽,其中该多肽包含选自SEQ ID NO:6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,78,80,82,84,86,88,90,92和94的氨基酸序列。

8. 如权利要求4所述的多肽,其中所述多肽还能以比具有SEQ ID NO:6的序列的酮还原酶多肽所能达到的速率更高的速率将所述底物还原为所述产物。

9. 如权利要求8所述的多肽,其中该多肽包含选自SEQ ID NO:8,10,12,14,16,18,20,

22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 和 94 的氨基酸序列。

10. 如权利要求 4 所述的多肽, 其中所述多肽还能以比具有 SEQ ID NO:6 的序列的酮还原酶多肽所能达到的速率高至少 450% 的速率将所述底物还原为所述产物。

11. 如权利要求 10 所述的多肽, 所述多肽包含选自 SEQ ID NO:8, 10, 14, 16, 18, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 和 94 的氨基酸序列。

12. 如权利要求 4 所述的多肽, 其中所述多肽还能以比具有 SEQ ID NO:6 的序列的酮还原酶多肽所能达到的速率高至少 1500% 的速率将所述底物还原为所述产物。

13. 如权利要求 12 所述的多肽, 该多肽包含选自 SEQ ID NO:18, 32, 34, 36, 38, 40, 42, 44, 46, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 和 94 的氨基酸序列。

14. 如权利要求 4 所述的多肽, 当使用相对 2', 6' - 二氯 -3' - 氟苯乙酮底物的量按重量计小于 1% 的量的所述多肽进行时, 该多肽能在不到 24 小时内将至少 95% 的 2', 6' - 二氯 -3' - 氟苯乙酮底物转化为 (S)-1-(2,6-二氯-3-氟苯基) 乙醇。

15. 如权利要求 14 所述的多肽, 该多肽包含选自 SEQ ID NO:18, 32, 34, 36, 38, 40, 42, 44, 46, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 和 94 的氨基酸序列。

16. 一种多核苷酸, 该多核苷酸编码如权利要求 1 至 15 任一项所述的多肽。

17. 如权利要求 16 所述的多核苷酸, 该多核苷酸选自由 SEQ ID NO:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91 和 93 组成的组。

18. 一种表达载体, 该表达载体包含与适合于指引在宿主细胞中表达的一个或多个控制序列可操作地连接的如权利要求 16 所述的多核苷酸。

19. 一种宿主细胞, 该宿主细胞包含如权利要求 18 所述的表达载体。

20. 一种用于将 2', 6' - 取代苯乙酮底物立体选择性地还原为对应的取代 (S)-1-苯乙醇的方法, 所述 2', 6' - 取代苯乙酮底物任选地在 3', 4' 和 5' 位置的一个或多个被取代, 该方法包括在适合于将底物立体选择性地还原或转化为对应的 (S)-醇产物的反应条件下, 使所述底物与如权利要求 1-15 任一项所述的酮还原酶多肽接触。

21. 如权利要求 20 所述的方法, 其中:

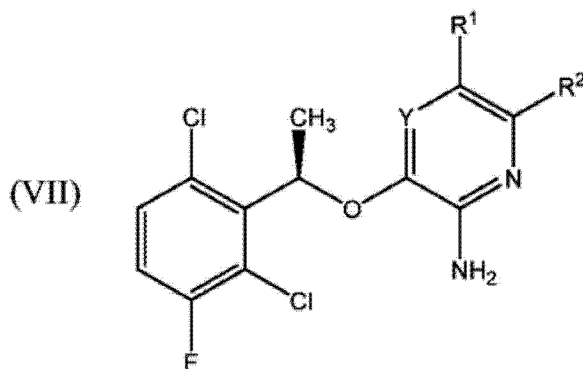
(i) 所述底物为 2', 6' - 二氯 -3' - 氟苯乙酮, 并且所述对应的 (S)-醇产物为 (S)-1-(2,6-二氯-3-氟苯基) 乙醇;

(ii) 所述酮还原酶多肽包含选自 SEQ ID NO:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92 和 94 的氨基酸序列; 或

(iii) 当所述方法用至少 200g/L 的底物和用小于 1g/L 的所述多肽进行时, 至少 95% 的所述底物在不到 24 小时内被还原为所述产物。

22. 如权利要求 21 所述的方法, 其中所述 (S)-1-(2,6-二氯-3-氟苯基) 乙醇以大于 99% 的立体异构过量形成。

23. 一种生产蛋白酪氨酸激酶抑制化合物的方法, 所述化合物具有结构式 (VII)



包括其盐、水合物及溶剂合物，其中：

Y 是 N 或 CR¹²；

R¹ 选自氢、卤素、C₆₋₁₂ 芳基、5-12 元杂芳基、C₃₋₁₂ 环烷基、3-12 元杂脂环、-O(CR⁶R⁷)_nR⁴、-C(O)R⁴、-C(O)OR⁴、-CN、-NO₂、-S(O)_mR⁴、-SO₂NR⁴R⁵、-C(O)NR⁴R⁵、-NR⁴C(O)R⁵、-C(=NR⁶)NR⁴R⁵、C₁₋₈ 烷基、C₂₋₈ 烯基和 C₂₋₈ 炔基；并且 R¹ 中的每个氢任选地由一个或多个 R³ 基团取代；

R² 是氢、卤素、C₁₋₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-S(O)_mR⁴、-SO₂NR⁴R⁵、-S(O)₂OR⁴、-NO₂、-NR⁴R⁵、-(CR⁶R⁷)_nOR⁴、-CN、-C(O)R⁴、-OC(O)R⁴、-O(CR⁶R⁷)_nR⁴、-NR⁴C(O)R⁵、-(CR⁶R⁷)_nC(O)OR⁴、-(CR⁶R⁷)_nNCR⁴R⁵、-C(=NR⁶)NR⁴R⁵、-NR⁴C(O)NR⁵R⁶、-NR⁴S(O)_pR⁵ 或 -C(O)NR⁴R⁵，并且 R² 中每个氢任选地由 R⁸ 取代；

每个 R³ 独立地为卤素、C₁₋₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-S(O)_mR⁴、-SO₂NR⁴R⁵、-S(O)₂OR⁴、-NO₂、-NR⁴R⁵、-(CR⁶R⁷)_nOR⁴、-CN、-C(O)R⁴、-OC(O)R⁴、-O(CR⁶R⁷)_nR⁴、-NR⁴C(O)R⁵、-(CR⁶R⁷)_nC(O)OR⁴、-(CR⁶R⁷)_nOR⁴、-(CR⁶R⁷)_nC(O)NR⁴R⁵、-(CR⁶R⁷)_nNCR⁴R⁵、-C(=NR⁶)NR⁴R⁵、-NR⁴C(O)NR⁵R⁶、-NR⁴S(O)_pR⁵ 或 -C(O)NR⁴R⁵，R³ 中每个氢任选地由 R⁸ 取代，并且邻近原子上的 R³ 基团可组合以形成 C₆₋₁₂ 芳基、5-12 元杂芳基、C₃₋₁₂ 环烷基或 3-12 元杂脂环基团；

每个 R⁴、R⁵、R⁶ 和 R⁷ 独立地为氢、卤素、C₁₋₁₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基；或者与相同氮原子结合的 R⁴、R⁵、R⁶ 和 R⁷ 的任两个可与其结合的氮原子一起组合以形成 3 至 12 元杂脂环或 5-12 元杂芳基基团，该基团任选地含有 1 至 3 个选自 N、O 和 S 的其他杂原子；或者与相同碳原子结合的 R⁴、R⁵、R⁶ 和 R⁷ 的任两个可组合以形成 C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环或 5-12 元杂芳基基团；并且 R⁴、R⁵、R⁶ 和 R⁷ 中每个氢任选地被 R⁸ 取代；

每个 R⁸ 独立地为卤素、C₁₋₁₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-NH₂、-CN、-OH、-O-C₁₋₁₂ 烷基、-O-(CH₂)_nC₃₋₁₂ 环烷基、-O-(CH₂)_nC₆₋₁₂ 芳基、-O-(CH₂)_n(3-12 元杂脂环) 或 -O-(CH₂)_n(5-12 元杂芳基)；并且 R⁸ 中每个氢任选地被 R¹¹ 取代；

每个 R⁹ 和 R¹⁰ 独立地为氢、卤素、C₁₋₁₂ 烷基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-S(O)_mR⁴、-SO₂NR⁴R⁵、-S(O)₂OR⁴、-NO₂、-NR⁴R⁵、-(CR⁶R⁷)_nOR⁴、-CN、-C(O)R⁴、-OC(O)R⁴、-NR⁴C(O)R⁵、-(CR⁶R⁷)_nC(O)OR⁴、-(CR⁶R⁷)_nNCR⁴R⁵、-NR⁴C(O)NR⁵R⁶、-NR⁴S(O)_pR⁵ 或 -C(O)NR⁴R⁵；R⁹ 或 R¹⁰ 可与 A 的环原子或 A 的取代基组合以形成 C₃₋₁₂ 环烷基、3-12 元杂脂环、C₆₋₁₂ 芳基或与 A 耦合的 5-12 元杂芳基；并且 R⁹ 和 R¹⁰ 中每个氢任选地被 R³ 取代；

每个 R¹¹ 独立地为卤素、C₁₋₁₂ 烷基、C₁₋₁₂ 烷氧基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、

5-12 元杂芳基、 $-O-C_{1-12}$ 烷基、 $-O-(CH_2)_n C_{3-12}$ 环烷基、 $-O-(CH_2)_n C_{6-12}$ 芳基、 $-O-(CH_2)_n$ (3-12 元杂脂环)、 $-O-(CH_2)_n$ (5-12 元杂芳基) 或 $-CN$, 并且 R^{11} 中每个氢任选地被卤素、 $-OH$ 、 $-CN$ 、可部分或完全卤素化的 $-C_{1-12}$ 烷基、可部分或完全卤素化的 $-O-C_{1-12}$ 烷基、 $-CO$ 、 $-SO$ 或 $-SO_2$ 取代;

R^{12} 是氢、卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-S(O)_m R^4$ 、 $-SO_2 NR^4 R^5$ 、 $-S(O)_2 OR^4$ 、 $-NO_2$ 、 $-NR^4 R^5$ 、 $-(CR^6 R^7)_n OR^4$ 、 $-CN$ 、 $-C(O) R^4$ 、 $-OC(O) R^4$ 、 $-O(CR^6 R^7)_n R^4$ 、 $-NR^4 C(O) R^5$ 、 $-(CR^6 R^7)_n C(O) OR^4$ 、 $-(CR^6 R^7)_n NCR^4 R^5$ 、 $-C(=NR^6) NR^4 R^5$ 、 $-NR^4 C(O) NR^5 R^6$ 、 $-NR^4 S(O)_p R^5$ 或 $-C(O) NR^4 R^5$, 并且 R^{12} 中每个氢任选地被 R^3 取代;

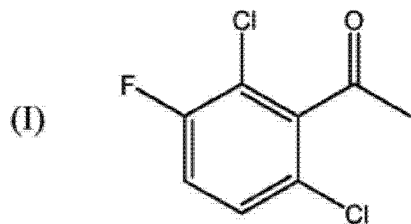
每个 R^{13} 独立地为卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-S(O)_m R^4$ 、 $-SO_2 NR^4 R^5$ 、 $-S(O)_2 OR^4$ 、 $-NO_2$ 、 $-NR^4 R^5$ 、 $-(CR^6 R^7)_n OR^4$ 、 $-CN$ 、 $-C(O) R^4$ 、 $-OC(O) R^4$ 、 $-O(CR^6 R^7)_n R^4$ 、 $-NR^4 C(O) R^5$ 、 $-(CR^6 R^7)_n C(O) OR^4$ 、 $-(CR^6 R^7)_n OR^4$ 、 $-(CR^6 R^7)_n C(O) NR^4 R^5$ 、 $-(CR^6 R^7)_n NCR^4 R^5$ 、 $-C(=NR^6) NR^4 R^5$ 、 $-NR^4 C(O) NR^5 R^6$ 、 $-NR^4 S(O)_p R^5$ 、 $-C(O) NR^4 R^5$ 、 $-(CR^6 R^7)_n$ (3-12 元杂脂环)、 $-(CR^6 R^7)_n$ (C_{3-12} 环烷基)、 $-(CR^6 R^7)_n$ (C_{6-12} 芳基)、 $-(CR^6 R^7)_n$ (5-12 元杂芳基)、 $-(CR^6 R^7)_n C(O) NR^4 R^5$ 或 $-(CR^6 R^7)_n C(O) R^4$, 邻近原子上的 R^{13} 基团可组合以形成 C_{6-12} 芳基、5-12 元杂芳基、 C_{3-12} 环烷基或 3-12 元杂脂环基团, 并且 R^{13} 中每个氢任选地被 R^3 取代;

每个 m 独立地为 0、1 或 2;

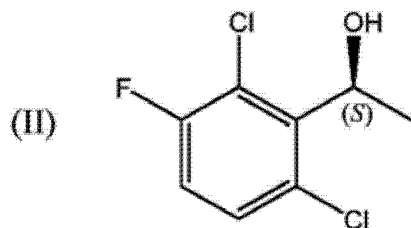
每个 n 独立地为 0、1、2、3 或 4; 并且

每个 p 独立地为 1 或 2,

所述方法包括将在合适的反应条件下用权利要求 1-15 任一项所述的酮还原酶多肽将式 (I) 的化合物

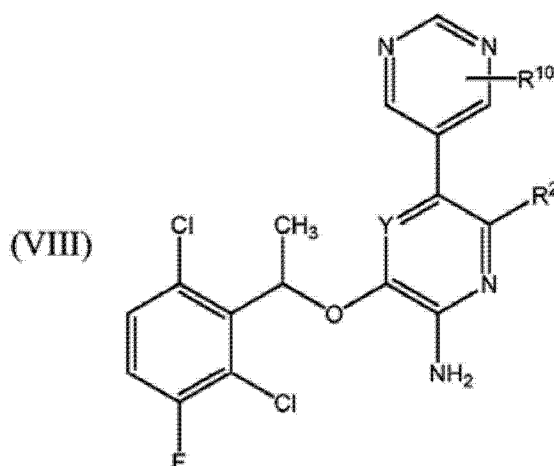


还原或转化为式 (II) 的化合物



的步骤。

24. 一种生产蛋白酪氨酸激酶抑制化合物的方法, 所述化合物具有结构式 (VIII)



包括其盐、水合物及溶剂合物,其中:

Y 是 N 或 CR¹;

R¹ 是氢、卤素、C₁₋₁₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-S(O)_mR⁴、-SO₂NR⁴R⁵、-S(O)₂OR⁴、-NO₂、-NR⁴R⁵、-(CR⁶R⁷)_nOR⁴、-CN、-C(O)R⁴、-OC(O)R⁴、-O(CR⁶R⁷)_nR⁴、-NR⁴C(O)R⁵、-(CR⁶R⁷)_nC(O)OR⁴、-(CR⁶R⁷)_nNCR⁴R⁵、-C(=NR⁶)NR⁴R⁵、-NR⁴C(O)NR⁵R⁶、-NR⁴S(O)_pR⁵ 或 -C(O)NR⁴R⁵, 并且 R¹ 中每个氢任选地被 R³ 取代;

R² 是氢、卤素、C₁₋₁₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-S(O)_mR⁴、-SO₂NR⁴R⁵、-S(O)₂OR⁴、-NO₂、-NR⁴R⁵、-(CR⁶R⁷)_nOR⁴、-CN、-C(O)R⁴、-OC(O)R⁴、-O(CR⁶R⁷)_nR⁴、-NR⁴C(O)R⁵、-(CR⁶R⁷)_nC(O)OR⁴、-(CR⁶R⁷)_nNCR⁴R⁵、-C(=NR⁶)NR⁴R⁵、-NR⁴C(O)NR⁵R⁶、-NR⁴S(O)_pR⁵ 或 -C(O)NR⁴R⁵, 并且 R² 中每个氢任选地被 R⁸ 取代;

每个 R³ 独立地为卤素、C₁₋₁₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-S(O)_mR⁴、-SO₂NR⁴R⁵、-S(O)₂OR⁴、-NO₂、-NR⁴R⁵、-(CR⁶R⁷)_nOR⁴、-CN、-C(O)R⁴、-OC(O)R⁴、-O(CR⁶R⁷)_nR⁴、-NR⁴C(O)R⁵、-(CR⁶R⁷)_nC(O)OR⁴、-(CR⁶R⁷)_nOR⁴、-(CR⁶R⁷)_nC(O)NR⁴R⁵、-(CR⁶R⁷)_nNCR⁴R⁵、-C(=NR⁶)NR⁴R⁵、-NR⁴C(O)NR⁵R⁶、-NR⁴S(O)_pR⁵ 或 -C(O)NR⁴R⁵, R³ 中每个氢任选地被 R⁸ 取代, 邻近原子上的 R³ 基团可组合以形成 C₆₋₁₂ 芳基、5-12 元杂芳基、C₃₋₁₂ 环烷基或 3-12 元杂脂环基团;

每个 R⁴、R⁵、R⁶ 和 R⁷ 独立地为氢、卤素、C₁₋₁₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基;或者与相同氮原子结合的 R⁴、R⁵、R⁶ 和 R⁷ 的任两个可与其结合的氮原子一起组合以形成 3 至 12 元杂脂环或 5-12 元杂芳基基团, 该基团任选地含有 1 至 3 个选自 N、O 和 S 的其他杂原子;或者与相同碳原子结合的 R⁴、R⁵、R⁶ 和 R⁷ 的任两个可组合以形成 C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环或 5-12 元杂芳基基团;并且 R⁴、R⁵、R⁶ 和 R⁷ 中每个氢任选地被 R⁸ 取代;

每个 R⁸ 独立地为卤素、C₁₋₁₂ 烷基、C₂₋₁₂ 烯基、C₂₋₁₂ 炔基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-NH₂、-CN、-OH、-O-C₁₋₁₂ 烷基、-O-(CH₂)_nC₃₋₁₂ 环烷基、-O-(CH₂)_nC₆₋₁₂ 芳基、-O-(CH₂)_n(3-12 元杂脂环) 或 -O-(CH₂)_n(5-12 元杂芳基);并且 R⁸ 中每个氢任选地被 R⁹ 取代;

每个 R⁹ 独立地为卤素、C₁₋₁₂ 烷基、C₁₋₁₂ 烷氧基、C₃₋₁₂ 环烷基、C₆₋₁₂ 芳基、3-12 元杂脂环、5-12 元杂芳基、-O-C₁₋₁₂ 烷基、-O-(CH₂)_nC₃₋₁₂ 环烷基、-O-(CH₂)_nC₆₋₁₂ 芳基、-O-(CH₂)_n(3-12 元杂脂环)、-O-(CH₂)_n(5-12 元杂芳基) 或 -CN, 并且 R⁹ 中每个氢任选地被卤素、-OH、-CN、可

部分或完全卤素化的 $-C_{1-12}$ 烃基、可部分或完全卤素化的 $-O-C_{1-12}$ 烃基、 $-CO$ 、 $-SO$ 或 $-SO_2$ 取代；

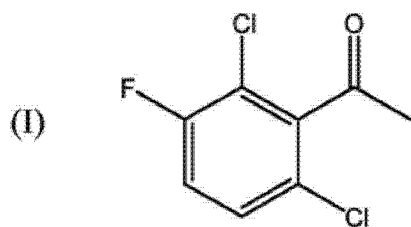
R^{10} 代表一个、两个或三个任选的取代基，其独立地为卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-S(O)_m R^4$ 、 $-SO_2 NR^4 R^5$ 、 $-S(O)_2 O R^4$ 、 $-NO_2$ 、 $-NR^4 R^5$ 、 $-(CR^6 R^7)_n OR^4$ 、 $-CN$ 、 $-C(O) R^4$ 、 $-OC(O) R^4$ 、 $-O(CR^6 R^7)_n R^4$ 、 $-NR^4 C(O) R^5$ 、 $-(CR^6 R^7)_n C(O) OR^4$ 、 $-(CR^6 R^7)_n OR^4$ 、 $-(CR^6 R^7)_n C(O) NR^4 R^5$ 、 $-(CR^6 R^7)_n NCR^4 R^5$ 、 $-C(=NR^6) NR^4 R^5$ 、 $-NR^4 C(O) NR^5 R^6$ 、 $-NR^4 S(O)_p R^5$ 、 $-C(O) NR^4 R^5$ 、 $-(CR^6 R^7)_n$ (3-12 元杂脂环)、 $-(CR^6 R^7)_n$ (C_{3-12} 环烷基)、 $-(CR^6 R^7)_n$ (C_{6-12} 芳基)、 $-(CR^6 R^7)_n$ (5-12 元杂芳基) 或 $-(CR^6 R^7)_n C(O) NR^4 R^5$ ，并且 R^{10} 中每个氢任选地被 R^3 取代；

每个 m 独立地为 0、1 或 2；

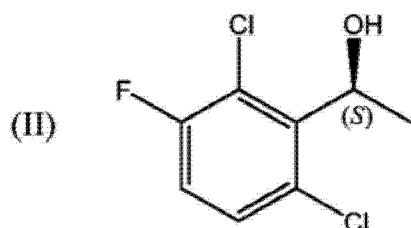
每个 n 独立地为 0、1、2、3 或 4；并且

每个 p 独立地为 1 或 2，

所述方法包括在合适的反应条件下用权利要求 1-15 任一项所述的酮还原酶多肽将式 (I) 的化合物



还原或转化为式 (II) 的化合物



的步骤。

25. 一种组合物，该组合物包含 2', 6' - 取代苯乙酮和 / 或对应的取代 (S)-1- 苯乙醇和如权利要求 1-15 中任一项所述的酮还原酶，所述 2', 6' - 取代苯乙酮任选地在 3', 4' 和 5' 位置的一个或多个被取代。

用于还原苯乙酮的酮还原酶多肽

[0001] 1. 相关申请的交叉引用

[0002] 本申请根据 35U.S.C. § 119(e) 要求 2007 年 9 月 13 日提交的申请序列第 60/972,058 号的利益,其内容通过引用并入本文。

[0003] 2. 序列表、表格或计算机程序的引用

[0004] 作为文件名 376247-017.txt 通过 EFS-Web 以计算机可读形式 (CRF) 根据 37C.F.R. § 1.821 随本申请同时提交的序列表通过引用并入本文。序列表的电子拷贝产生在 2008 年 9 月 13 日,文件大小为 213 千字节。

[0005] 发明领域

[0006] 本发明涉及用于还原苯乙酮的酮还原酶多肽。

[0007] 3. 背景

[0008] 属于酮还原酶 (KRED) 或羰基还原酶类 (EC 1.1.1.184) 的酶可用于从对应的前体立体异构酮底物和通过对应的消旋醛底物的立体专一性还原来合成光学上有活性的醇类。KRED 一般将酮和醛底物转化为对应的醇产物,但是还可催化相反的反应,将醇底物氧化为对应的酮/醛产物。通过诸如 KRED 的酶来还原酮和醛以及氧化醇需要辅因子,最常见的辅因子是还原型烟酰胺腺嘌呤二核苷酸 (NADH) 或还原型烟酰胺腺嘌呤二核苷酸磷酸 (NADPH),以及用于氧化反应的烟酰胺腺嘌呤二核苷酸 (NAD) 或烟酰胺腺嘌呤二核苷酸磷酸 (NADP)。NADH 和 NADPH 作为电子供体,而 NAD 和 NADP 作为电子受体。常常观察到酮还原酶和醇脱氢酶接受磷酸化或非磷酸化辅因子(以其氧化和还原状态),但不是两者。

[0009] KRED 酶可发现于多种细菌和酵母中(综述参见 Kraus 和 Waldman,1995, *Enzyme catalysis in organic synthesis*(有机合成中的酶催化),第 1&2 卷.VCH Weinheim;Faber, K.,2000, *Biotransformations in organic chemistry*(有机化学中的生物转化),第 4 版, Springer, Berlin Heidelberg New York;和 Hummel 和 KuLa, 1989, *Eur. J. Biochem.* 184:1-13)。已报道几种 KRED 基因和酶序列,例如,木兰假丝酵母 (*Candida magnoliae*) (Genbank 登录号 JC7338;GI:11360538)、近平滑假丝酵母 (*Candida parapsilosis*) (Genbank 登录号 BAA24528.1;GI:2815409)、赭色掷孢酵母 (*Sporobolomyces salmonicolor*) (Genbank 登录号 AF160799;GL6539734)。

[0010] 为了避免用于产生关键化合物的许多化学合成过程,增加地采用酮还原酶以将不同的酮和醛底物酶促转化为手性醇产物。这些应用可采用表达酮还原酶的全细胞用于生物催化酮和醛的还原,或在其中全细胞中多种酮还原酶的存在将相反地影响需要产物的立体纯度和产量的那些实例中通过使用纯化的酶。对体外应用,使用诸如葡萄糖脱氢酶 (GDH)、甲酸脱氢酶等的辅因子 (NADH 或 NADPH) 再生酶与酮还原酶的组合。使用酮还原酶来产生有用的化学化合物的实例包括 4-氯乙酰乙酸酯类的不对称还原 (Zhou,1983, *J. Am. Chem. Soc.* 105:5925-5926;Santaniello, *J. Chem. Res. (S)* 1984:132-133;美国专利第 5,559,030 号;美国专利第 5,700,670 号和美国专利第 5,891,685 号)、二羧酸类的还原(例如,美国专利第 6,399,339 号)、(S) 氯代-5-羟基-3-氧代己酸叔丁酯的还原(例如,美国专利第 6,645,746 和 WO 01/40450 号)、基于吡咯并三嗪的化合物的还原(例如,美国申请第

2006/0286646 号);取代苯乙酮的还原(例如,美国专利第 6,800,477 号);和酮四氢噻吩(ketothiolane)的还原(WO 2005/054491)。

[0011] 需要鉴定可用来将各种酮底物转化为其对应的手性醇产物的其他酮还原酶。

[0012] 4. 概述

[0013] 本公开提供了能将确定的酮底物立体选择性地还原为其对应醇产物、并且当与获自克菲尔乳杆菌(*L. kefir*)(SEQ ID NO:4)或短乳杆菌(*L. brevis*)(SEQ ID NO:2)或小乳杆菌(*L. minor*)(SEQ ID NO:98)的天然存在的野生型 KRED 酶相比时或当与其他工程酮还原酶相比时具有改进特性的工程酮还原酶(“KRED”)。本公开中显示来自乳杆菌种(*Lactobacillus species*)的天然存在的酮还原酶将化合物苯乙酮还原为(R)-1-苯乙醇。因为野生型酶一般选择性地用于将苯乙酮还原为其对应的(R)-醇,这些天然存在的酶是(R)-选择性酮还原酶或(R)-酮还原酶。对取代苯乙酮(诸如 2',6'-二氯-3'-氟苯乙酮),这些野生型克菲尔乳杆菌或短乳杆菌或小乳杆菌酮还原酶表现出对取代苯乙酮底物的不明显的活性,如果有活性的话。然而,衍生自野生型乳杆菌种的酮还原酶的本公开的工程酮还原酶能将苯乙酮还原为(S)-1-苯乙醇。因此本文所述的酮还原酶通过与用于还原苯乙酮的野生型克菲尔乳杆菌或短乳杆菌或小乳杆菌酮还原酶相比反向的对映选择性来表征。因此本公开的这些多肽称为(S)-选择性酮还原酶或(S)-酮还原酶。反向对映选择性是基于将野生型酮还原酶 190 位置的残基(即 X190)突变为不是酪氨酸的残基,优选地突变为非芳族残基,特别地突变为脯氨酸残基。

[0014] 此外,除了改变的立体选择性外,本文所述的工程酶可具有一种或多种改进的特性。例如,与野生型酮还原酶相比,工程酮还原酶多肽可具有改进的酶促活性用于将底物还原为产物和/或进一步增强对(S)对映体的立体选择性。酶特性的改进还可包括但不限于热稳定性、溶剂稳定性的增强或降低的产物抑制。如本文进一步公开的,尽管野生型酮还原酶在还原取代苯乙酮中显示不明显的活性,本公开提供了能将取代苯乙酮 2',6'-二氯-3'-氟苯乙酮还原或转化为(S)-1-[2,6-二氯-3-氟苯基]-乙醇的酮还原酶。

[0015] 相应地,在一些实施方案中,本公开涉及在对应于 SEQ ID NO:2,4 或 98 的 X190 的残基处具有不是酪氨酸的残基的酮还原酶多肽。在一些实施方案中,该残基是非芳族残基,诸如,例如脂肪族、受限、非极性或非胱氨酸残基。在一些实施方案中,该残基是脯氨酸。

[0016] 除了在对应于 X190 的残基处的特点,与 SEQ ID NO:2,4 或 98 的序列相比,该酮还原酶可在其他残基位置具有一个或多个残基的差异。在一些实施方案中,本文的酮还原酶多肽包含与基于 SEQ ID NO:2,4 或 98 的参考序列相比具有至少约 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高同一性的氨基酸序列,其在对应于 X190 的残基处具有非芳族残基,包括脂肪族、受限、非极性或非胱氨酸残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,条件是 该酮还原酶多肽在对应于 X190 的残基处具有不是酪氨酸的残基,特别是非芳族残基。在一些实施方案中,该酮还原酶多肽具有其中对应于 X190 的残基是脂肪族、受限、非极性或非胱氨酸残基的氨基酸序列。在一些实施方案中,该酮还原酶多肽具有其中对应于 X190 的残基是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸、特别是脯氨酸的氨基酸序列。在一些实施方案中,这些残基差异导致改进的特性,诸如改进的对底物的酶促活性。改进的特性可以是参考野生型酮还原酶或参考工程酮还原酶。例如,在一些实施方案中,酮还原酶的改进是与具有对应于 SEQ ID NO:6 的氨基酸

序列的工程酶的特性相比,其能以可测量活性以高于 99% 的立体异构过量将底物转化为产物,因此与野生型克菲尔乳杆菌或短乳杆菌或小乳杆菌酮还原酶相比有改进。可导致一种或多种改进酶特性的各种残基差异在详述中提供。在一些实施方案中,这些工程酮还原酶多肽是基于在 SEQ ID NO :95、96 和 119 中列出的序列式(或其区域,诸如残基 90-211)。

[0017] 在一些实施方案中,本公开的酮还原酶多肽能以高于 99% 的立体异构过量和与具有 SEQ ID NO :6 的序列的酮还原酶多肽相比有改进的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇。与 SEQ ID NO :6 相比关于酶促活性方面有改进的示例性多肽包括但不限于包含对应于 SEQ ID NO :8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48、50、52、54、56、58、60、62、64、66、68、70、72、74、76、78、80、82、84、86、88、90、92 和 94 的氨基酸序列的多肽。

[0018] 在一些实施方案中,该酮还原酶多肽能以高于 99% 的立体异构过量和与具有 SEQ ID NO :6 的序列的酮还原酶多肽相比有改进的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇,其中该多肽与具有 SEQ ID NO :6 的序列的多肽相比还具有改进的热稳定性。具有此类改进的示例性多肽包括但不限于包含对应于 SEQ ID NO :8、16、18、20、22、26、28、30、32、34、38、40、42、44、46、54、66、68、70、72、74、76、78、80、82、84、86、88、90、92 和 94 的氨基酸序列的多肽。

[0019] 在一些实施方案中,该酮还原酶多肽能以高于 99% 的立体异构过量和比具有 SEQ ID NO :6 的序列的酮还原酶多肽高至少约 450% 的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇。能具有此类改进的示例性多肽包括但不限于包含对应于 SEQ ID NO :8、10、14、16、18、22、24、26、28、30、32、34、36、38、40、42、44、46、54、56、58、60、62、64、66、68、70、72、74、76、78、80、82、84、86、88、90、92 和 94 的氨基酸序列的多肽。

[0020] 在一些实施方案中,该酮还原酶多肽能以高于 99% 的立体异构过量和比具有 SEQ ID NO :6 的序列的酮还原酶多肽高至少约 450% 的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇,其中该多肽与 SEQ ID NO :6 的多肽相比还具有改进的热稳定性。具有此类特性的示例性多肽包括但不限于包含对应于 SEQ ID NO :8、16、18、22、26、28、30、32、34、38、40、42、44、46、54、66、68、70、72、74、76、78、80、82、84、86、88、90、92 和 94 的氨基酸序列的多肽。

[0021] 在一些实施方案中,该酮还原酶多肽能以高于 99% 的立体异构过量和比具有 SEQ ID NO :6 的序列的酮还原酶多肽高至少约 1500% 的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇。能具有此类改进的示例性多肽包括但不限于包含对应于 SEQ ID NO :18、32、34、36、38、40、42、44、46、74、76、78、80、82、84、86、88、90、92 和 94 的氨基酸序列的多肽。

[0022] 在一些实施方案中,该酮还原酶多肽能以高于 99% 的立体异构过量和比具有 SEQ ID NO :6 的序列的酮还原酶多肽高至少约 1500% 的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇,其中该多肽与 SEQ ID NO :6 的多肽相比还具有改进的热稳定性。具有此类特性的示例性多肽包括但不限于包含对应于 SEQ ID NO :18、32、34、36、38、40、42、44、46、74、76、78、80、82、84、86、88、90、92 和 94 的氨基酸序列的多肽。

[0023] 在一些实施方案中,当用相对 2',6'-二氯-3'-氟苯乙酮底物的量按重量

计少于约 1% 的量的多肽进行时,该酮还原酶多肽能在不到约 24 小时内将 2',6' - 二氯 -3' - 氟苯乙酮底物的至少约 95% 以至约 99% 的立体异构过量转化为 (S)-1-[2,6- 二氯 -3- 氟苯基]-乙醇。具有该能力的示例性多肽包括但不限于包含对应于 SEQ ID NO :18、32、34、36、38、40、42、44、46、74、76、78、80、82、84、86、88、90、92 和 94 的氨基酸序列的多肽。

[0024] 在一些实施方案中,该酮还原酶多肽能以高于 99% 的立体异构过量和比具有 SEQ ID NO :6 的序列的酮还原酶多肽高至少约 450% 的速率将 2',6' - 二氯 -3' - 氟苯乙酮转化为 (S)-1-[2,6- 二氯 -3- 氟苯基]-乙醇,其中在 50°C 热处理 2 小时后,该多肽还能以比具有 SEQ ID NO :16 的序列的多肽高至少约 400% 的速率将底物转化为产物(其中 SEQ ID NO :16 的多肽也用相同的热处理来处理)。具有此类特性的示例性多肽包括但不限于包含对应于 SEQ ID NO :18、32、34、36、38、40、42、44、46、74、76、78、80、82、84、86、88、90、92 和 94 的氨基酸序列的多肽。

[0025] 在一些实施方案中,能将 2',6' - 二氯 -3' - 氟苯乙酮转化为 (S)-1-[2,6- 二氯 -3- 氟苯基]-乙醇的改进的酮还原酶多肽包含具有与对应于基于 SEQ ID NO :2、4 或 98 的参考序列的残基 90-211 的区域或结构域具有至少约 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 同一性的氨基酸序列的区域或结构域,条件是该酮还原酶多肽区域或结构域在对应于 X190 的残基处具有酪氨酸之外的残基,所述参考序列在对应于 X190 的残基处具有非芳族残基,包括脂肪族、受限、非极性或半胱氨酸残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸。在一些实施方案中,该酮还原酶多肽具有对应于参考序列的残基 90-211 的区域或结构域,其中对应于 X190 的残基是非芳族残基。在一些实施方案中,对应于 X190 的该残基可以是脂肪族、受限、非极性或半胱氨酸残基。在一些实施方案中,对应于 X190 的残基可以是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,特别是脯氨酸。在一些实施方案中,与参考序列相比,该酮还原酶多肽在结构域或区域可具有一个或多个残基差异。可不同于参考序列的各种残基的位置在详述中提供。

[0026] 在另一个方面,本公开提供了编码本文描述的工程酮还原酶的多核苷酸或在高度严谨条件下与此类多核苷酸杂交的多核苷酸。该多核苷酸可包括启动子和用于编码的工程酮还原酶的表达的其他调节元件,并且可利用为特定的希望的表达系统而优化的密码子。编码该工程酮还原酶的示例性多核苷酸包括但不限于包含对应于 SEQ ID NO :5、7、9、11、13、15、17、19、21、23、25、27、29、31、33、35、37、39、41、43、45、47、49、51、53、55、57、59、61、63、65、67、69、71、73、75、77、79、81、83、85、87、89、91 和 93 的序列的多核苷酸。

[0027] 在另一个方面,本公开提供了包含本文描述的多核苷酸和/或表达载体的宿主细胞。该宿主细胞可以是克菲尔乳杆菌或短乳杆菌,或者其可以是不同的有机体。该宿主细胞可用于本文描述的工程酮还原酶的表达和分离,或者可选地该宿主细胞可直接用于将式 (I) 或 (III) 的取代苯乙酮底物分别转化为对应的式 (II) 或 (IV) 的 (S)-醇产物。

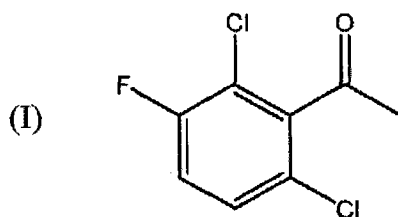
[0028] 无论使用全细胞、细胞提取物或纯化的酮还原酶来进行该方法,可使用单一酮还原酶,或者可选地可使用两种或多种酮还原酶的混合物。

[0029] 如上文所提,本文描述的酮还原酶能催化 2',6' - 取代苯乙酮,在 3'、4' 和 5' 位置的一个或多个被任选地取代,向对应的 (S)-醇产物的还原反应。

[0030] 在一些实施方案中,该酮还原酶能将结构式 (I) 的酮 2',6' - 二氯 -3' - 氟苯

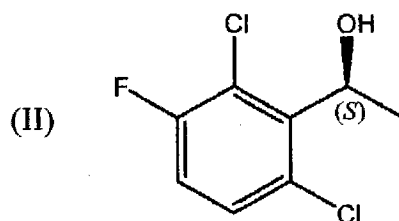
乙酮：

[0031]



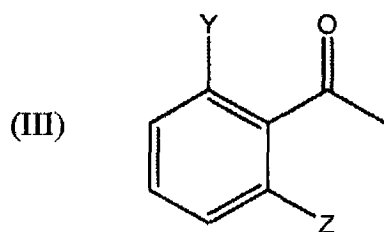
[0032] 还原或转化为结构式 (II) 的对应的手性醇产物 (S)-1-[2,6-二氯-3-氟苯基]-乙醇：

[0033]



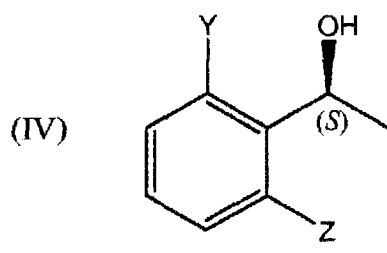
[0034] 在一些实施方案中,本文描述的酮还原酶能催化结构式 (III) 的 2',6'-取代苯乙酮化合物：

[0035]



[0036] 其在 3'、4' 和 5' 位置的一个或多个被任选地取代,其中 Y 和 Z 独立地选自 CH₃、CF₃、NH₂、OH、OCH₃、Cl 和 Br,还原为结构式 (IV) 的对应的手性醇产物：

[0037]



[0038] 相应地,在一些实施方案中,本公开提供了用于将 2',6' 取代苯乙酮底物,任选地在 3'、4' 和 5' 位置的一个或多个被取代,还原为对应的取代 (S)-苯乙醇的方法,其中该方法包括在适合于将底物还原或转化为对应的取代 (S)-苯乙醇的反应条件下使底物与本文描述的酮还原酶接触。在本方法的一些实施方案中,底物以高于约 25%、50%、75%、80%、85%、90%、95%、99% 或 99.9% 的立体异构过量被还原为产物。

[0039] 在一些实施方案中,本公开提供了用于将式 (III) 的 2',6' 取代苯乙酮还原为式 (IV) 的对应的取代 (S)-苯乙醇的方法,其中该方法包括在适合于将式 (III) 的底物还原或转化为式 (IV) 的对应的取代 (S)-苯乙醇产物的反应条件下使底物与本文描述的酮还原酶接触。在本方法的一些实施方案中,底物以高于约 25%、50%、75%、80%、85%、90%、

95%、99%或99.9%的立体异构过量被还原为产物。

[0040] 在一些实施方案中,本公开提供了用于将式(I)的2',6'-二氯-3'-氟苯乙酮底物还原为式(II)的其对应的(S)-醇产物(S)-1-[2,6-二氯-3-氟苯基]-乙醇的方法,其中该方法包括在适合于将2',6'-二氯-3'-氟苯乙酮还原或转化为(S)-1-[2,6-二氯-3-氟苯基]-乙醇的反应条件下使2',6'-二氯-3'-氟苯乙酮与本文描述的酮还原酶接触或育。在本方法的一些实施方案中,底物以高于约85%、90%、95%、99%或99.9%的立体异构过量被还原为产物。在一些实施方案中,底物以高于约85%的立体异构过量被还原为产物,其中酮还原酶多肽包含基于SEQ ID NO:95、96或119的序列式的氨基酸序列。在一些实施方案中,底物以高于约99%的立体异构过量被还原为产物,其中本方法中使用的酮还原酶多肽包含对应于SEQ ID NO:6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48、50、52、54、56、58、60、62、64、66、68、70、72、74、76、78、80、82、84、86、88、90、92和94的氨基酸序列。

[0041] 在本方法的一些实施方案中,当使用相对2',6'-二氯-3'-氟苯乙酮底物的量按重量计低于约1%的量的酮还原酶多肽进行该方法时,底物的至少约95%以高于约99%的立体异构过量、在不到24小时内被还原为产物,其中酮还原酶多肽包含对应于SEQ ID NO:18、32、34、36、38、40、42、44、46、74、76、78、80、82、84、86、88、90、92和94的氨基酸序列。

[0042] 在一些实施方案中,本公开提供了本文描述的酮还原酶和2,6取代苯乙酮,任选地在3'、4'或5'位置的一个或多个被取代,和/或对应的取代(S)-苯乙醇的组合物。在一些实施方案中,组合物包含本文描述的酮还原酶和式(I)的化合物和/或式(II)的化合物。在一些实施方案中,组合物包含本文描述的酮还原酶和式(III)的化合物和/或式(IV)的化合物。在一些实施方案中,组合物包含本文描述的酮还原酶和式(V)的化合物和/或式(VI)的化合物。在一些实施方案中,组合物还可包含辅因子再生系统。

[0043] 在一些实施方案中,本公开涉及工程酮还原酶在合成描述于W006021886、W006021884、W006021881和W004076412中的蛋白激酶抑制剂中的应用。在一些实施方案中,在用于合成这些蛋白激酶抑制剂的方法中,方法的步骤可包含使用本公开的酮还原酶将式(I)的底物2',6'-二氯-3'-氟苯乙酮还原或转化为其对应的式(II)的(S)-醇产物(S)-1-[2,6-二氯-3-氟苯基]-乙醇。

[0044] 5. 附图简述

[0045] 图1说明了酮还原酶(KRED)在将式(I)的底物化合物2',6'-二氯-3-氟苯乙酮转化为式(II)的对应的手性醇产物(S)-1-[2,6-二氯-3-氟苯基]-乙醇中的作用。该反应中,底物被生物催化地还原为对应的(S)-醇。该还原使用本文描述的KRED和诸如NADPH的辅因子。使用葡萄糖脱氢酶(GDH)来将NADP⁺转化/再循环为NADPH。葡萄糖被转化为葡萄糖酸,其然后通过加入氢氧化钠被转化为其钠盐(葡萄糖酸钠)。

[0046] 6. 详述

[0047] 6.1 定义

[0048] 如本文所用,以下术语预期具有以下含义。

[0049] “酮还原酶”和“KRED”在本文可交换使用,指具有将羰基基团还原为其对应的醇的酶促能力的多肽。更具体地,本发明的酮还原酶多肽能将上文式(I)的化合物立体选择

性地还原为上文式 (II) 的对应的产物。该多肽一般使用辅因子还原型烟酰胺腺嘌呤二核苷酸 (NADH) 或还原型烟酰胺腺嘌呤二核苷酸磷酸 (NADPH) 作为还原剂。如本文所用的酮还原酶包括天然存在的 (野生型) 酮还原酶以及通过人工操作产生的非天然存在的工程多肽。

[0050] “编码序列”指编码蛋白的氨基酸序列的那部分核酸 (例如, 基因)。“天然存在的”或“野生型”指自然中发现的形式。例如, 天然存在的或野生型多肽或多核苷酸序列是存在于有机体中的序列, 其可从自然中的来源分离, 并且没有有意地通过人工操作来修饰。

[0051] 当参考例如细胞、核酸或多肽使用时,“重组”指已经以不在自然中存在的方式被修饰, 或者与自然形式相同但从合成材料和 / 或通过使用重组技术的操作产生或衍生的材料, 或对应于材料的天然或自然形式的材料。非限制性实例包括但不限于表达在细胞的天然 (非重组) 形式内未发现的基因或者表达以不同水平另外表达的天然基因的重组细胞。

[0052] “序列同一性百分比”和“百分比同源性”在本文可交换使用, 指多核苷酸和多肽之间的比较, 并通过在比较窗口 (comparison window) 比较 两个最佳比对的序列来确定, 其中与用于两个序列的最佳比对的参考序列 (其不包含添加或缺失) 相比, 比较窗口中的多核苷酸或多肽序列的部分可包含添加或缺失 (即缺口)。百分比可通过以下计算: 确定两个序列中相同核酸碱基或氨基酸残基出现的位置的数目以产生匹配位置的数目, 用匹配位置的数目除以比较窗口中的位置总数, 结果乘以 100 以产生序列同一性的百分比。可选地, 百分比可通过以下计算: 确定两个序列中相同核酸碱基或氨基酸残基出现的位置的数目或者核酸碱基或氨基酸残基与缺口来比对的位置的数目以产生匹配位置的数目, 用匹配位置的数目除以比较窗口中位置的总数, 结果乘以 100 以产生序列同一性的百分比。本领域技术人员理解, 存在许多建立的算法可用来比对两个序列。用于比较的序列的最佳比对可例如通过 Smith 和 Waterman, 1981, Adv. Appl. Math. 2 :482 的局部同源性算法、通过 Needleman 和 Wunsch, 1970, J. Mol. Biol. 48 :443 的同源性比对算法、通过 Pearson 和 Lipman, 1988, Proc. Natl. Acad. Sci. USA 85 :2444 的搜索相似性方法、通过这些算法的计算机化执行 (GCG Wisconsin 软件包中的 GAP、BESTFIT、FASTA 和 TFASTA)、或通过视觉检查来进行 (一般参见 Current Protocols in Molecular Biology (分子生物学最新实验方法汇编), F. M. Ausubel 等, 编辑, Current Protocols, Greene Publishing Associates, Inc. 和 John Wiley & Sons, Inc. 的合资企业, (1995 增刊) (Ausubel))。适合于确定百分比序列同一性和序列相似性的算法的实例为 BLAST 和 BLAST 2.0 算法, 其分别描述于 Altschul 等, 1990, J. Mol. Biol. 215 :403-410 和 Altschul 等, 1977, Nucleic Acids Res. 3389-3402。用于执行 BLAST 分析的软件可通过美国国家生物技术信息中心的网站公开获得。该算法涉及首先通过鉴定查询序列中长度 W 的短字来鉴定高得分序列对 (HSP), 所述短字当与数据库序列中相同长度的字比对时, 匹配或者满足一些正值阈值得分 T。T 指相邻字得分阈值 (Altschul 等, 上文)。这些最初的相邻字匹配字串 (word hits) 作为起始搜索的种子来发现含有这些字的更长的 HSP。然后字匹配字串沿每个序列以两个方向延伸, 直到累积比对得分不能被增加。对核苷酸序列, 累积得分使用参数 M (匹配残基对的奖励分, 总是 > 0) 和 N (不匹配残基的惩罚分, 总是 < 0) 来计算。对氨基酸序列, 使用得分矩阵来计算累积得分。当出现以下情况时, 每个方向的字 匹配字串的延伸被暂停: 累积比对分从其最大获得值减少量 X 时; 由于一个或多个负得分残基比对的累积, 累积得分变为零或小于零; 或者达到任一序列的末端。

BLAST 算法参数 W、T 和 X 确定了比对的敏感性和速度。BLASTN 程序（用于核苷酸序列）使用以下作为默认参数：字长 (W) 为 11、期望值 (E) 为 10、 $M = 5$ 、 $N = -4$ 和双链比较。对氨基酸序列，BLASTP 程序使用以下作为默认参数：字长 (W) 为 3、期望值 (E) 为 10 和 BLOSUM62 得分矩阵（参见 Henikoff 和 Henikoff, 1989, Proc Natl Acad Sci USA 89:10915）。序列比对和 % 序列同一性的示例性确定可采用 GCG Wisconsin 软件包中的 BESTFIT 或 GAP 程序 (Accelrys, Madison WI)，使用提供的默认参数。

[0053] “参考序列”指用作序列比较基础的确定序列。参考序列可以是更大的序列的子集，例如，全长基因或多肽序列的区段。一般而言，参考序列为至少 20 个核苷酸或氨基酸残基长度、至少 25 个残基长度、至少为 50 个残基长度或核酸或多肽的全长。由于两个多核苷酸或多肽可各自 (1) 包含两个序列之间相似的序列（即完整序列的一部分），和 (2) 还可包含两个序列之间不同的序列，因此两个（或多个）多核苷酸或多肽之间的序列比较一般通过在“比较窗口”比较两个多核苷酸的序列以鉴定和比较序列相似性的局部区域来进行。

[0054] 在一些实施方案中，“参考序列”可基于一级氨基酸序列，其中参考序列是可在一级序列中具有一个或多个改变的序列。例如，“基于 SEQ IDNO:4 的在对应于 X190 的残基处具有脯氨酸的”参考序列指其中在 SEQ IDNO:4 的 X190 处的对应残基已改变为脯氨酸的参考序列。

[0055] “比较窗口”指至少约 20 个连续核苷酸位置或氨基酸残基的概念区段，其中序列可与至少 20 个连续核苷酸或氨基酸的参考序列比较，并且其中与用于两个序列最佳比对的参考序列（其不包含添加或缺失）相比，比较窗口中序列部分可包含 20% 或更少的添加或缺失（即缺口）。比较窗口可长于 20 个连续残基，并且包括任选地 30 个、40 个、50 个、100 个或更长的窗口。

[0056] “大体同一性”指在至少 20 个残基位置的比较窗口中、通常在至少 30-50 个残基的窗口中与参考序列相比，具有至少 80% 序列同一性、至少 85% 同一性和 89% 至 95% 序列同一性、更通常地至少 99% 序列同一性的多核苷酸或多肽序列，其中序列同一性的百分比通过在比较窗口中比较参考序列与包括总计为参考序列的 20% 或更少缺失或添加的序列来计算。在应用于多肽的具体实施方案中，术语“大体同一性”意指当诸如通过程序 GAP 或 BESTFIT 使用默认缺口权重进行最佳比对时，两个多肽序列共享至少 80% 的序列同一性，优选地至少 89% 的序列同一性、至少 95% 的序列同一性或更高（例如，99% 的序列同一性）。优选地，不相同的残基位置通过保守氨基酸置换而不同。

[0057] 当在特定氨基酸或多核苷酸序列的编号的上下文中使用时，“对应于”、“参考”或“相对于”指当给定氨基酸或多核苷酸序列与参考序列相比时，特定参考序列的残基的编号。换句话说，给定聚合体的残基数目或残基位置针对参考序列指定，而不是通过给定氨基酸或多核苷酸序列内残基的实际编号位置指定。例如，可通过将缺口引入以优化两个序列之间的残基匹配来比对给定氨基酸序列诸如工程酮还原酶的氨基酸序列和参考序列。在这些实例中，尽管存在缺口，给定氨基酸或多核苷酸序列中残基的编号针对其比对的参考序列产生。

[0058] “立体选择性”指化学或酶促反应中一种立体异构体针对另一种立体异构体的优先形成。立体选择性可以是部分的，其中一种立体异构体的形成优于另一种，或者其可以是完全的，其中只形成一种立体异构体。当立体异构体是对映体时，立体选择性称为对映选择

性,两者总和中一种对映体的分数(一般报道为百分比)。本领域通常可选地报道为根据下式从中计算的(一般作为百分比)对映体过量(e. e.):[主要对映体-次要对映体]/[主要对映体+次要对映体]。这还可称为立体异构过量(s. e)。其中立体异构体是非对映异构体时,立体选择性称为非对映选择性,两种非对映异构体混合物中一种非对映异构体的分数(一般报道为百分比)。

[0059] “高度立体选择性”指能以至少约85%的立体异构过量将2',6'-二氯-3'-氟苯乙酮(式(I))转化或还原为对应的(S)-醇产物(S)-1-[2,6-二氯-3-氟苯基]-乙醇(式(II))的酮还原酶多肽。

[0060] “改进的酶特性”指与参考酮还原酶相比表现出任何酶特性的改进的酮还原酶多肽。对本文描述的工程酮还原酶多肽,比较一般针对野生型酮还原酶,尽管在一些实施方案中,参考酮还原酶可以是另一种改进的工程酮还原酶。需要改进的酶特性包括但不限于酶促活性(其可以底物转化的百分比表示)、热稳定性、pH活性谱(profile)、辅因子需求、对抑制剂(例如产物抑制)的耐性(refractoriness)、立体专一性和立体选择性(包括立体选择性)。

[0061] “增强的酶促活性”指工程酮还原酶多肽的改进的特性,其可由与参考酮还原酶相比比活性(例如,产生的产物/时间/重量蛋白)的增强或者底物转化为产物的百分比(例如使用特定量的KRED在特定时间内起始量底物转化为产物的百分比)的增加来代表。确定酶活性的示例性方法提供于实施例。涉及酶活性的任何特性可被影响,包括经典的酶特性 K_m 、 V_{max} 或 k_{cat} ,其改变可导致增强的酶促活性。酶活性的改进可为从对应的野生型酮还原酶的酶促活性的约1.5倍,至超过天然存在的酮还原酶或酮还原酶多肽衍生自其的另一种工程酮还原酶的酶促活性多达2倍、5倍、10倍、20倍、25倍、50倍、75倍、100倍或更高。在特定实施方案中,工程酮还原酶表现出高于母酮还原酶活性1.5至50倍、1.5至100倍范围内的改进的酶促活性。熟练的技术人员应理解,任何酶的活性是扩散限制的,以致催化周转速率不能超过底物的扩散速率,包括任何需要的辅因子。扩散限制或 k_{cat}/K_m 的理论最大值一般为约 10^8 至 10^9 ($M^{-1}s^{-1}$)。因此,酮还原酶的酶活性的任何改进具有与酮还原酶所作用的底物的扩散速率相关的上限。酮还原酶活性可通过用于测量酮还原酶的任一种标准测定来测量,诸如由于伴随酮还原成醇的NADPH氧化导致的NADPH的吸光度或荧光的减弱(参见实施例5),或通过偶联测定中产生的产物来测量。使用确定的酶制品、在设定条件下的确定测定和一种或多种确定的底物进行酶活性的比较,进一步详细描述于本文。一般而言,当比较裂解物时,确定细胞的数目和测定的蛋白的量,以及使用相同的表达系统和相同的宿主细胞来使宿主细胞产生的和裂解物中存在的酶量的变化最小。

[0062] “转化”指将底物酶促还原为对应的产物。“百分比转化”指在特定条件下在一定时间内还原为产物的底物的百分比。因此,酮还原酶多肽的“酶促活性”或“活性”可表示为底物变为产物的“百分比转化”。

[0063] “热稳定”指与未处理的酶相比,暴露于高温(例如40-80℃)一段时间(例如0.5-24小时后)保持相似活性(超过例如60%至80%)的酮还原酶多肽。

[0064] “溶剂稳定”指与未处理的酶相比,在暴露于可变浓度(例如5-99%)的溶剂(异丙醇、四氢呋喃、2-甲基四氢呋喃、丙酮、甲苯、乙酸丁酯、甲基叔丁醚)一段时间(例如0.5-24小时)后保持相似活性(超过例如60%至80%)的酮还原酶多肽。

[0065] “pH 稳定”指与未处理的酶相比,在暴露于高或低 pH(例如 4.5-6 或 8-12)一段时间(例如 0.5-24 小时)后保持相似活性(超过例如 60%至 80%)的酮还原酶多肽。

[0066] “热和溶剂稳定性”指既热稳定又溶剂稳定的酮还原酶多肽。

[0067] 在工程酮还原酶上下文中,如本文所用的“衍生自”鉴定工程化所基于的初始酮还原酶和/或编码此类酮还原酶的基因。例如,SEQ ID NO:38 的工程酮还原酶是通过在多代期间人工演化编码 SEQ ID NO:4 的克菲尔乳杆菌酮还原酶的基因获得的。因此,该工程酮还原酶“衍生自”SEQ ID NO:4 的野生型酮还原酶。

[0068] “亲水性氨基酸或残基”指具有表现出根据 Eisenberg 等,1984, J. Mol. Biol. 179: 125-142 的归一化一致疏水性等级,小于零的疏水性的侧链的氨基酸或残基。遗传编码的亲水性氨基酸包括 L-Thr (T), L-Ser (S), L-His (H), L-Glu (E), L-Asn (N), L-Gln (Q), L-Asp (D), L-Lys (K) 和 L-Arg (R)。

[0069] “酸性氨基酸或残基”指当氨基酸包括于肽或多肽内时,具有表现出小于约 6 的 pK 值的侧链的亲水性氨基酸或残基。在生理 pH 下,酸性氨基酸由于缺少氢离子而一般具有带负电的侧链。遗传编码的酸性氨基酸包括 L-Glu (E) 和 L-Asp (D)。

[0070] “碱性氨基酸或残基”指当氨基酸包括于肽或多肽内时,具有表现出大于约 6 的 pK 值的侧链的亲水性氨基酸或残基。在生理 pH 下,碱性氨基酸由于与水合氢离子缔合而一般具有带正电的侧链。遗传编码的碱性氨基酸包括 L-Arg (R) 和 L-Lys (K)。

[0071] “极性氨基酸或残基”指具有在生理 pH 下不带电荷的侧链、但该侧链具有其中两个原子共同共享的电子对更靠近维持原子之一的至少一个键的侧链的亲水性氨基酸或残基。遗传编码的极性氨基酸包括 L-Asn (N), L-Gln (Q), L-Ser (S) 和 L-Thr (T)。

[0072] “疏水氨基酸或残基”指具有表现出根据 Eisenberg 等,1984, J. Mol. Biol. 179: 125-142 的归一化一致疏水性等级大于零的疏水性的侧链的氨基酸或残基。遗传编码的疏水氨基酸包括 L-Pro (P), L-Ile (I), L-Phe (F), L-Val (V), L-Leu (L), L-Trp (W), L-Met (M), L-Ala (A) 和 L-Tyr (Y)。

[0073] “芳族氨基酸或残基”指具有包括至少一个芳环或杂芳环的侧链的亲水或疏水氨基酸或残基。遗传编码的芳族氨基酸包括 L-Phe (F), L-Tyr (Y) 和 L-Trp (W)。尽管由于 L-His (H) 的杂芳氮原子的 pKa,其有时分类为碱性残基,或因为其侧链包括杂芳环而分类为芳族残基,本文中组氨酸分类为亲水性残基或作为“受限残基”(参见下文)。

[0074] “受限氨基酸或残基”指具有受限 (constrained) 几何学的氨基酸或残基。本文中,受限残基包括 L-pro (P) 和 L-his (H)。组氨酸因为其具有相对小的咪唑环而具有受限几何学。脯氨酸因为其还具有五元环而具有受限几何学。

[0075] “非极性氨基酸或残基”指具有在生理 pH 下不带电荷的侧链并且该侧链具有其中两个原子共同共享的电子对一般由两个原子中的每一个平等地维持的键(即侧链不是极性的)的疏水氨基酸或残基。遗传编码的非极性氨基酸包括 L-Gly (G), L-Leu (L), L-Val (V), L-Ile (I), L-Met (M) 和 L-Ala (A)。

[0076] “脂肪族氨基酸或残基”指具有脂肪族烃侧链的疏水氨基酸或残基。遗传编码的脂肪族氨基酸包括 L-Ala (A), L-Val (V), L-Leu (L) 和 L-Ile (I)。

[0077] “半胱氨酸”氨基酸 L-Cys (C) 是不常见的,因为其可与其他 L-Cys (C) 氨基酸或其他含有硫烷基或巯基的氨基酸形成二硫键。“半胱氨酸样残基”包括半胱氨酸和含有可用

于形成二硫键的巯基部分的其他氨基酸。L-Cys(C) (和具有含有 -SH 侧链的其他氨基酸) 在肽中以还原型游离 -SH 或氧化型二硫键合形式存在的能力影响 L-Cys(C) 是否有助于肽的净疏水或亲水性特征。尽管根据 Eisenberg 的归一化一致等级 (Eisenberg 等, 1984, 上文) L-Cys(C) 表现出 0.29 的疏水性, 应理解, 为了本公开的目的, L-Cys(C) 被分类为其自身独特的组。

[0078] “小氨基酸或残基”指具有包括总计 3 个或更少的碳和 / 或杂原子 (不包括 α -碳和氢) 的侧链的氨基酸或残基。根据以上定义, 小氨基酸或残基可进一步分为脂肪族、非极性、极性或酸性小氨基酸或残基。遗传编码的小氨基酸包括 L-Ala(A), L-Val(V), L-Cys(C), L-Asn(N), L-Ser(S), L-Thr(T) 和 L-Asp(D)。

[0079] “含有羟基的氨基酸或残基”指含有羟基 (-OH) 部分的氨基酸。遗传编码的含有羟基的氨基酸包括 L-Ser(S), L-Thr(T) 和 L-Tyr(Y)。

[0080] “保守”氨基酸取代或突变指具有相似侧链的残基的可交换性, 因此一般涉及用处于氨基酸的相同或相似定义类别的氨基酸取代多肽中的氨基酸。在一些实施方案中, 如本文所用的保守突变不包括从亲水性残基到亲水性残基、疏水残基到疏水残基、含有羟基的残基到含有羟基的残基或小残基到小残基的取代, 如果保守突变可相反是从脂肪族残基到脂肪族残基、非极性残基到非极性残基、极性残基到极性残基、酸性残基到酸性残基、碱性残基到碱性残基、芳族残基至芳族残基、或受限残基至受限残基的取代。进一步, 如本文所用, A、V、L 或 I 可被保守地突变为另一个脂肪族残基或突变为另一个非极性残基。下文表 1 显示了示例性的保守取代。

[0081] 表 1: 保守取代

[0082]

残基	可能的保守突变
A, L, V, I	其他脂肪族残基 (A, L, V, I) 其他非极性残基 (A, L, V, I, G, M)
G, M	其他非极性残基 (A, L, V, I, G, M)
D, E	其他酸性残基 (D, E)
K, R	其他碱性残基 (K, R)
P, H	其他受限残基 (P, H)
N, Q, S, T	其他极性残基
Y, W, F	其他芳族残基 (Y, W, F)
C	无

[0083] “非保守取代”指用具有显著不同侧链特性的氨基酸取代或突变多肽中的氨基酸。非保守取代可使用上文所列的定义组之间而非之内的氨基酸。在一个实施方案中, 非保守突变影响 (a) 取代区域内肽主链的结构 (例如脯氨酸取代甘氨酸) (b) 电荷或疏水性、或 (c) 侧链的体积。

[0084] “缺失”指通过从参考多肽去除一个或多个氨基酸来修饰多肽。氨基酸序列中的缺失可包含去除 1 个或多个氨基酸、2 个或更多个氨基酸、3 个或更多个氨基酸、4 个或更多个氨基酸、5 个或更多个氨基酸、6 个或更多个氨基酸、8 个或更多个氨基酸、10 个或更多个氨基酸、15 个或更多个氨基酸、或 20 个或更多个氨基酸、多达构成参考酶的氨基酸总数的 10%、多达构成参考酶的氨基酸总数的 15%、或多达构成参考酶的氨基酸总数的 20%, 同时保持酶促活性和 / 或保持工程酮还原酶的改进特性。缺失可涉及多肽的内部部分和 / 或末端部分。在不同实施方案中, 缺失可包含连续的区段, 或者可以是不连续的。

[0085] “插入”指通过从参考多肽添加一个或多个氨基酸来修饰多肽。在一些实施方案中,改进的工程酮还原酶包含向天然存在的酮还原酶多肽插入一个或多个氨基酸以及向其他改进的酮还原酶多肽插入一个或多个氨基酸。插入可以在多肽的内部部分,或者插入羧基或氨基末端。如本文所用的插入包括如本领域已知的融合蛋白。插入可以是氨基酸的连续区段,或者由天然存在的多肽中的一个或多个氨基酸隔开。

[0086] 如本文所用的“片段”指具有氨基末端和/或羧基末端缺失但其中剩余的氨基酸序列与序列中对应的位置相同的多肽。片段可以是至少 14 个氨基酸长、至少 20 个氨基酸长、至少 50 个氨基酸长或更长和多达 SEQ ID NO:2 或 SEQ ID NO:4 的全长天然存在的酮还原酶多肽的 70%、80%、90%、95%、98% 和 99%。

[0087] “分离的多肽”指与天然伴随其的其他污染物大体上分开的多肽,例如蛋白质、脂类和多核苷酸。该术语包含已被从其天然存在环境或表达系统(例如宿主细胞或体外合成)中移出或纯化的多肽。改进的酮还原酶可存在于细胞内、存在于细胞培养基中或以各种形式制备,诸如裂解物或分离的制品。这样,在一些实施方案中,改进的酮还原酶可以是分离的多肽。

[0088] “大体纯的多肽”指其中多肽物类(species)是存在的优势物类(即基于摩尔或重量基础,其比组合物中任何其他单个大分子物类更丰富)的组合物,并且当目标物类构成通过摩尔或%重量计存在的大分子种的至少约 50%时,其一般为大体纯化的组合物。一般而言,大体纯的酮还原酶组合物将包含组合物中存在的所有大分子物类的按摩尔或%重量计的约 60%或更多、约 70%或更多、约 80%或更多、约 90%或更多、约 95%或更多以及约 98%或更多。在一些实施方案中,目标物类被纯化为基本均质(即组合物中的污染物类不能通过传统的检测方法检测),其中组合物主要由单个大分子物类组成。溶剂物类、小分子(< 500 道尔顿)和元素离子物类不被认为是大分子物类。在一些实施方案中,分离的改进酮还原酶多肽是大体纯的多肽组合物。

[0089] 本文使用的“严紧杂交”指核酸杂交体(hybrid)稳定的条件。如本领域技术人员所知,杂交体的稳定性反映在杂交体的解链温度(T_m)。一般而言,杂交体的稳定性是离子强度、温度、G/C 含量和促溶剂(chaotropic agent)的存在的函数。多核苷酸的 T_m 值可使用用于预测解链温度的已知方法来计算(参见,例如, Baldino 等, Methods Enzymology 168:761-777; Bolton 等, 1962, Proc. Natl. Acad. Sci. USA 48:1390; Bresslauer 等, 1986, Proc. Natl. Acad. Sci. USA 83:8893-8897; Freier 等, 1986, Proc. Natl. Acad. Sci. USA 83:9373-9377; Kierzek 等, Biochemistry 25:7840-7846; Rychlik 等, 1990, Nucleic Acids Res 18:6409-6412(勘误, 1991, Nucleic Acids Res 19:698); Sambrook 等, 上文); Suggs 等, 1981, 于 Developmental Biology Using Purified Genes(使用纯化基因的发育生物学)(Brown 等, 编辑), 第 683-693 页, Academic Press; 和 Wetmur, 1991, Crit Rev Biochem Mol Biol 26:227-259。所有出版物通过引用并入本文)。在一些实施方案中,多核苷酸编码本文公开的多肽,并且在定义条件下诸如中等严紧或高严紧条件与编码本公开的工程酮还原酶的序列的互补序列杂交。

[0090] “杂交严紧性”涉及核酸杂交中诸如洗涤条件的杂交条件。一般而言,杂交反应在较低严紧性条件下进行,然后进行可变但更高严紧性的洗涤。术语“中等严紧性杂交”指允许靶 DNA 结合于与靶 DNA 具有约 60%同一性、优选地约 75%同一性、约 85%同一性、与

靶多核苷酸具有高于约 90% 同一性的互补核酸的条件。示例性中等严谨性条件是等同于在 42℃ 下于 50% 甲酰胺、5x Denhart 溶液、5xSSPE、0.2% SDS 中杂交,然后在 42℃ 下于 0.2xSSPE、0.2% SDS 中洗涤的条件。“高严谨性杂交”一般指比在确定的多核苷酸序列的溶液条件下确定的热解链温度低约 10℃ 或更少的条件。在一些实施方案中,高严谨性条件指仅允许 65℃ 下于 0.018M NaCl 中形成稳定杂交体的那些核酸序列的杂交的条件(即如果杂交体在 65℃ 下于 0.018M NaCl 中不稳定,其在高严谨性条件下将不稳定,如本文所涵盖)。可例如通过在等同于 42℃ 下 50% 甲酰胺、5x Denhart 溶液、5x SSPE、0.2% SDS 中杂交、然后通过 65℃ 下于 0.1x SSPE 和 0.1% SDS 中洗涤的条件下杂交来提供高严谨性条件。另一种高严谨性条件是在等同于 65℃ 下在含有 0.1% (w : v) SDS 的 5X SSC 中杂交并于 65℃ 下在含有 0.1% SDS 的 0.1x SSC 中洗涤的杂交条件下杂交。其他高严谨性杂交条件以及中等严谨性条件描述于上文引用的参考文献中。

[0091] “异源”多核苷酸指通过实验室技术被引入宿主细胞的任何多核苷酸,包括从宿主细胞中移出、经受实验室操作然后重新引入宿主细胞的多核苷酸。

[0092] “密码子优化”指编码蛋白的多核苷酸的密码子改变为在特定有机体中优先使用的那些密码子,以便编码的蛋白在感兴趣的有机体中有效地表达。尽管遗传密码是简并的,即大多数氨基酸由几个称为“同义”(synonyms)或“同义”(synonymous)密码子的密码子代表,公知的是特定有机体的密码子使用是非随机的,并且偏向于特定的密码子三联体。这种密码子使用偏好在特定基因、共同功能或祖先来源的基因、对比低拷贝数蛋白高表达的蛋白和有机体基因组的聚集蛋白编码区中可能更高。在一些实施方案中,编码酮还原酶的多核苷酸可进行密码子优化以便从选择用于表达的宿主有机体中最佳产生。

[0093] “优选、最佳、高密码子使用偏向密码子”可交换地指在蛋白编码区中的使用频率高于编码相同氨基酸的其他密码子的密码子。优选密码子可根据单个基因、共同功能或来源的一套基因、高表达基因中的密码子使用、整个有机体的聚集蛋白编码区中的密码子频率、相关有机体的聚集蛋白编码区中的密码子频率或其组合来确定。其频率随基因表达水平增加的密码子一般是用于表达的最佳密码子。用来确定特定有机体中密码子频率(例如,密码子使用、相对同义密码子使用)和密码子偏爱性的多种方法是已知的,包括多变量分析,例如,使用聚类分析或相关性分析和基因中使用的密码子的有效数目(参见 GCG CodonPreference, Genetics Computer Group Wisconsin Package ;CodonW, John Peden, University of Nottingham ;McInerney, J. O, 1998, Bioinformatics 14 : 372-73 ;Stenico 等, 1994, Nucleic Acids Res. 22:2437-46 ;Wright, F., 1990, Gene 87 : 23-29)。可获得不断增加的有机体清单的密码子使用表(参见例如, Wada 等, 1992, Nucleic Acids Res. 20 :2111-2118 ;Nakamura 等, 2000, Nucl. Acids Res. 28 :292 ;Duret 等, 上文 ;Henaut 和 Danchin, “ Escherichia coli and Salmonella (大肠杆菌和沙门氏菌),” 1996, Neidhardt, 等编辑, ASM Press, Washington D. C., 第 2047-2066 页。用于获得密码子使用的数据源可依赖于能编码蛋白的任何可获得的核苷酸序列。这些数据集包括实际上已知来编码表达蛋白(例如,完整蛋白编码序列 -CDS)、表达序列标签 (EST) 或基因组序列的预测编码区的核酸序列(参见例如, Mount, D., Bioinformatics : Sequence and Genome Analysis (生物信息学 : 序列和基因组分析), 第 8 章, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 2001 ;Uberbacher, E. C, 1996,

MethodsEnzymol. 266 :259-281 ;Tiwari 等,1997, Comput. Appl. Biosci. 13 :263-270)。

[0094] “控制序列”在本文定义为包括对感兴趣的多核苷酸和 / 或多肽的表达必需或有利的的所有组分。每种控制序列可以是编码多肽的核酸序列的天然或外源序列。此类控制序列包括但不限于前导序列、多聚腺苷化序列、前体肽序列、启动子、信号肽序列和转录终止子。控制序列至少包括启动子和转录和翻译终止信号。为了引入特定限制性酶切位点以促进控制序列与 感兴趣的多核苷酸例如编码多肽的核酸序列的编码区连接的目的,控制序列可提供有接头。

[0095] “可操作地连接”在本文定义为其中控制序列被适当地置于相对于多核苷酸序列的某一位置(即处于功能性关系)以便控制序列指引多核苷酸和 / 或多核苷酸编码的多肽的表达的构型。

[0096] “启动子库列”是由宿主细胞识别用于多核苷酸表达的核酸序列。控制序列可包含适当的启动子序列。启动子序列含有介导多核苷酸表达的转录控制序列。启动子可以是在选择的宿主细胞中显示转录活性的任何核酸序列,包括突变、截短和杂合(hybrid)的启动子,并且可获自编码与宿主细胞同源或异源的胞外或胞内多肽的基因。

[0097] 6.2 酮还原酶

[0098] 本公开提供了能将确定的酮底物立体选择性地还原为其对应的醇产物、并且当与获自克菲尔乳杆菌(SEQ ID NO :2)或短乳杆菌(SEQ ID NO :4)或小乳杆菌(SEQ ID NO :98)的天然存在的野生型 KRED 酶相比时或者当与其他工程酮还原酶相比时具有改进特性的工程酮还原酶(“KRED”)。如本公开所示,野生型克菲尔乳杆菌或短乳杆菌或小乳杆菌酮还原酶对 2',6' - 二氯 -3' - 氟苯乙酮的还原具有非常低的活性,如果有任何活性的话(参见实施例)。野生型酶对更少的取代苯乙酮底物具有更高的活性,野生型酶一般选择性地将苯乙酮还原为其对应的(R)-醇。野生型乳杆菌种酮还原酶将典型参考化合物苯乙酮还原为(R)-1-苯乙醇,并且因此称为(R)-选择性酮还原酶或(R)-酮还原酶。然而,衍生自野生型乳杆菌种酮还原酶的本公开的工程酮还原酶将苯乙酮还原为(S)-1-苯乙醇,并且因此称为(S)-选择性酮还原酶或(S)-酮还原酶。因此,本公开的改进的酮还原酶多肽与野生型克菲尔乳杆菌或短乳杆菌或小乳杆菌酮还原酶(ketoreductases)相比能反向对映选择性地还原苯乙酮。这种反向对映选择性是基于将野生型酶位置 190 的残基优选地突变为非芳族残基,特别是突变为脯氨酸残基。不限于理论,位于位置 190 的野生型酪氨酸残基似乎与处于 pro-S 构象的底物不协调。因此,在一些实施方案中,本公开的酮还原酶多肽在对应于 SEQ ID NO :2 或 4 或 98 的位置 190 的残基处具有不是酪氨酸的残基。优选地,该残基是非芳族残基,诸如,例如脂肪族、受限、非极性或非胱氨酸残基。在一些实施方案中,该残基是脯氨酸。

[0099] 在一些实施方案中,如上文提到,具有改进酶特性的工程酮还原酶参考 SEQ ID NO :4 的克菲尔乳杆菌酮还原酶或 SEQ ID NO :2 的短乳杆菌或 SEQ ID NO :98 的小乳杆菌酮还原酶描述。从起始甲硫氨酸(M)残基(即 M 代表残基位置 1)开始确定酮还原酶中氨基酸残基的位置,尽管熟练的技术人员将理解该起始甲硫氨酸残基可通过生物加工机制去除,诸如在宿主细胞或体外翻译系统以产生缺少起始甲硫氨酸残基的成熟蛋白。特定氨基酸或氨基酸改变位置处的氨基酸残基存在于氨基酸序列中,有时在本文描述为“Xn”或“位置 n”,其中 n 指残基位置。其中位于相同残基位置的氨基酸残基在酮还原酶之间不同时,不

同的残基可由“/”表示,排列为“克菲尔乳杆菌残基/短乳杆菌残基/小乳杆菌”。为参考序列中氨基酸残基的替换的取代突变,例如具有不同氨基酸残基的 SEQ ID NO:2 和 SEQ ID NO:4 和 SEQ ID NO:98 的野生型酮还原酶可由符号“→”表示。本文中突变优势描述为突变“为”一类氨基酸。例如,SEQ ID NO:2 的残基 16 可被突变“为”极性残基。但是用短语“为”不排除从一类氨基酸突变为相同类的另一个氨基酸。例如,SEQ ID NO:2 的残基 16 是极性残基苏氨酸,但是其可被突变为不同的极性残基,例如,突变可以是“T16S”(16 → S) 的突变。

[0100] 编码克菲尔乳杆菌、短乳杆菌或小乳杆菌的天然存在酮还原酶(也称为“ADH”或“醇脱氢酶”)的天然存在多核苷酸可获自己已知编码酮还原酶活性的分离的多核苷酸(例如,克菲尔乳杆菌:Genbank 登录号 AAP94029GI:33112056 或 SEQ ID NO:3;短乳杆菌:Genbank 登录号 CAD66648GL28400789 或 SEQ ID NO:1;和小乳杆菌:SEQ ID NO:97)。

[0101] 在一些实施方案中,酮还原酶多肽的改进特性(与野生型或另一种工程多肽相比)是关于其将式(III)的取代苯乙酮底物还原或转化为式(IV)的其对应的(S)-醇产物的其立体选择性的增强。在一些实施方案中,酮还原酶特性的改进特性是关于其将 2'-6'-二氯-3-氟苯乙酮还原为(S)-1-(2,6-二氯-3-氟苯基)乙醇的立体选择性的增加。在一些实施方案中,酮还原酶特性的改进特性是关于其将底物转化为产物的速率的增加。在一些实施方案中,酮还原酶特性的改进特性是关于其稳定性或热稳定性。在一些实施方案中,酮还原酶多肽具有超过一种改进的特性。

[0102] 在一些实施方案中,本文的酮还原酶多肽可具有对参考序列(例如天然存在的多肽或工程多肽)的多种修饰,以导致改进的酮还原酶特性。如本文所用,“修饰”包括氨基酸取代、缺失和插入。修饰的任一种或组合可被引入天然存在的或工程多肽,以产生工程酶。在此类实施方案中,氨基酸序列的修饰数目可包括 1 个或多个氨基酸、2 个或多个氨基酸、3 个或多个氨基酸、4 个或多个氨基酸、5 个或多个氨基酸、6 个或多个氨基酸、8 个或多个氨基酸、10 个或多个氨基酸、15 个或多个氨基酸或 20 个或多个氨基酸、多达参考多肽序列的氨基酸总数的 10%、多达参考多肽序列的氨基酸总数的 10%、多达参考多肽序列的氨基酸总数的 15%、多达参考多肽序列的氨基酸总数的 20%或多达参考多肽序列的氨基酸总数的 30%。在一些实施方案中,产生改进的酮还原酶特性的天然存在多肽或工程多肽的修饰数目可包含从约 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个参考序列的修饰。在一些实施方案中,修饰数目可以是 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个氨基酸残基。修饰可包含插入、缺失、取代或其组合。

[0103] 在一些实施方案中,修饰包含对参考序列的氨基酸取代。可产生改进的酮还原酶特性的取代可以位于 1 个或多个氨基酸、2 个或多个氨基酸、3 个或多个氨基酸、4 个或多个氨基酸、5 个或多个氨基酸、6 个或多个氨基酸、8 个或多个氨基酸、10 个或多个氨基酸、15 个或多个氨基酸或 20 个或多个氨基酸、多达参考酶序列的氨基酸总数的 10%、多达参考酶序列的氨基酸总数的 10%、多达参考酶序列的氨基酸总数的 20%或多达参考酶序列的氨基酸总数的 30%。在一些实施方案中,产生改进的酮还原酶特性的天然存在多肽或工程多肽的取代数目可包含约 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个参考序

列的氨基酸取代。在一些实施方案中,取代的数目可以是 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18, 20,22,24,26,30,35 或约 40 个氨基酸残基。

[0104] 在一些实施方案中,本文的酮还原酶多肽包含与基于 SEQ ID NO:2、4 或 98 的参考序列具有至少约 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99% 或更高同一性的氨基酸序列,条件是该酮还原酶多肽在对应于 X190 的残基处是酪氨酸之外的残基,特别是非芳族残基,所述参考序列在对应于 X190 的残基处具有非芳族残基(例如,脂肪族、受限、非极性或非极性氨基酸残基),优选地丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,特别是脯氨酸。在一些实施方案中,酮还原酶多肽具有其中对应于 X190 的残基是脂肪族、受限、非极性或非极性氨基酸残基的氨基酸序列。在一些实施方案中,酮还原酶具有其中对应于 X190 的残基是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸特别是脯氨酸的氨基酸序列。在一些实施方案中,这些酮还原酶多肽与参考氨基酸序列相比在其他残基位置可具有 1 个或多个残基差异。差异包括各种修饰,诸如取代、缺失和插入。取代可以是非保守取代、保守取代或非保守和保守取代的组合。在一些实施方案中,这些酮还原酶多肽与参考序列相比可具有任选地在其他氨基酸残基处的约 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,与参考序列相比,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。

[0105] 在一些实施方案中,这些立体选择性或高度立体选择性(本文中能以至少约 85% 的 e. e. 将底物还原为产物)的酮还原酶多肽包含对应于如 SEQ ID NO:95,96 和 119 所列的序列式(或其区域或结构域,诸如残基 90-211)的氨基酸序列。SEQ ID NO:95 是基于短乳杆菌酮还原酶的野生型氨基酸序列(SEQ ID NO:2);SEQ ID NO:96 是基于克菲尔乳杆菌酮还原酶的野生型氨基酸序列(SEQ ID NO:4);SEQ ID NO:119 是基于小乳杆菌酮还原酶的野生型氨基酸序列(SEQ ID NO:98)。基于 SEQ ID NO:95,96 或 119 的序列式的酮还原酶指定对应于 X190 的残基是非芳族残基。在一些实施方案中,酮还原酶多肽具有其中对应于 X190 的残基是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸的氨基酸序列。在一些实施方案中,酮还原酶多肽具有其中对应于 X190 的残基是脯氨酸的氨基酸序列。

[0106] 在一些实施方案中,包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列、具有如本文所述的残基 X190 的指定特点的酮还原酶多肽还可包括选自以下的一个或多个特点:对应于 X7 的残基是芳族、非极性、极性、受限、酸性或碱性残基;对应于 X16 的残基是极性残基;对应于 X43 的残基是非极性或非极性残基;对应于 X60 的残基是芳族、非极性或非极性脂肪族残基;对应于 X94 的残基是半胱氨酸、非极性或非极性脂肪族残基;对应于 X95 的残基是非极性或非极性脂肪族残基;对应于 X96 的残基是极性或非极性残基;对应于 X97 的残基是极性、非极性、脂肪族或非极性残基;对应于 X120 的残基是芳族、非极性或非极性脂肪族残基;对应于 X125 的残基是极性或非极性残基;对应于 X142 的残基是极性残基;对应于 X147 的残基是芳族、极性、非极性或非极性脂肪族残基;对应于 X149 的残基是非极性或非极性芳族残基;对应于 X150 的残基是受限或非极性残基;对应于 X152 的残基是非极性或非极性残基;对应于 X196 的残基是脂肪族、非极性或非极性芳族残基;对应于 X202 的残基是脂肪族、芳族或非极性残基;对应于 X205 的残基是碱性、非极性或非极性脂肪族残基;和对应于 X206 的残基是非极性或非极性芳族残基。在一些实施方案中,氨基酸序列可具有所述特点中的 2 个、3 个、4 个、5 个或 6 个

或更多。在一些实施方案中,包含对应于 SEQ ID NO:95,96 或 119 提供的序列式(或其区域)的氨基酸序列的多肽与 SEQ ID NO:2,4 或 98 的参考序列相比可另外具有不由 X 指定的突变的一个或多个残基。在一些实施方案中,突变可以是在不由上文 X 定义的其他氨基酸残基处的约 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个突变。在一些实施方案中,突变的数目可以是 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个其他氨基酸残基。在一些实施方案中,突变包含保守突变。

[0107] 在一些实施方案中,包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列的多肽与 SEQ ID NO:2,4 或 98 的氨基酸序列相比可具有一个或多个保守突变。示例性保守突变包括氨基酸置换,诸如但不限于:用另一个极性残基例如天冬酰胺、谷氨酰胺或丝氨酸置换对应于 X16 的残基苏氨酸(T);用另一个非极性或非脂肪族残基例如异亮氨酸置换对应于 X43 的残基缬氨酸;用脂肪族或芳香族残基例如丙氨酸置换对应于 X60 的残基;用另一个非极性或非脂肪族残基例如缬氨酸、亮氨酸或异亮氨酸置换对应于 X94 的残基丙氨酸(A);用另一个非极性或非脂肪族残基例如丙氨酸、亮氨酸或异亮氨酸置换对应于 X95 的残基缬氨酸(V);用另一个极性残基例如天冬酰胺、谷氨酰胺或苏氨酸置换对应于 X96 的残基丝氨酸(S);用另一个极性残基例如丝氨酸或天冬酰胺置换对应于 X142 的残基丝氨酸(S);用另一个非极性或非脂肪族残基例如丙氨酸、亮氨酸或异亮氨酸置换对应于 X196 的残基缬氨酸(V);和用另一个非极性或非脂肪族残基例如缬氨酸、亮氨酸或异亮氨酸置换对应于 X205 的残基丙氨酸(A)。

[0108] 在一些实施方案中,包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211)、具有如本文所述的残基 X190 的指定特点的氨基酸序列的酮还原酶多肽还可包括选自以下的一个或多个特点:对应于 X7 的残基是色氨酸、酪氨酸、苯丙氨酸、脯氨酸、组氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、谷氨酰胺、天冬酰胺、精氨酸或赖氨酸,特别是甘氨酸、组氨酸、苏氨酸、脯氨酸、色氨酸、精氨酸、组氨酸或天冬酰胺;对应于 X16 的残基是丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺,特别是丝氨酸;对应于 X43 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是异亮氨酸;对应于 X60 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是丙氨酸;对应于 X94 的残基是半胱氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是丙氨酸、缬氨酸或半胱氨酸;对应于 X95 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸,特别是异亮氨酸或亮氨酸;对应于 X96 的残基是天冬氨酸、谷氨酸、丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺,特别是丝氨酸、天冬酰胺、苏氨酸或谷氨酸;对应于 X97 的残基是丝氨酸、苏氨酸、天冬酰胺、谷氨酰胺、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、赖氨酸或精氨酸,特别是赖氨酸、苏氨酸、缬氨酸、精氨酸、甲硫氨酸或异亮氨酸;对应于 X120 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是苯丙氨酸或缬氨酸;对应于 X125 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是甘氨酸或丝氨酸;对应于 X142 的残基是丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺残基,特别是天冬酰胺;对应于 X147 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺,特别是苯丙氨酸、亮氨酸、异亮氨酸、缬氨酸

酸或谷氨酰胺;对应于 X149 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸或色氨酸,特别是甘氨酸或苯丙氨酸;对应于 X150 的残基是脯氨酸、组氨酸、天冬氨酸或谷氨酸,特别是天冬氨酸或组氨酸;对应于 X152 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺,特别是丝氨酸、苏氨酸或甲硫氨酸;对应于 X196 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是缬氨酸、亮氨酸、甲硫氨酸、苯丙氨酸或异亮氨酸;对应于 X202 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是丙氨酸、色氨酸、酪氨酸或甲硫氨酸;对应于 X205 的残基是赖氨酸、精氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是精氨酸;和对应于 X206 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸、色氨酸,特别是甲硫氨酸或酪氨酸。在一些实施方案中,氨基酸序列可具有所述特点中的 2 个、3 个、4 个、5 个、6 个或更多个。在一些实施方案中,包含对应于 SEQ ID NO:95,96 或 119 提供的序列式(或其区域)的氨基酸序列的多肽与 SEQ ID NO:2,4 或 98 的参考序列相比可另外具有不由 X 指定的突变的一个或多个残基。在一些实施方案中,突变可以是在不由上文 X 定义的其他氨基酸残基处的约 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个突变。在一些实施方案中,突变的数目可以是 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个其他氨基酸残基。在一些实施方案中,突变包含保守突变。

[0109] 在一些实施方案中,包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211、具有如本文所述对应于 X190 的残基处的特点的氨基酸序列的立体选择性酮还原酶多肽可另外具有以下特点的一个或多个或至少所有特点:对应于 X147 的残基是芳族、极性、非极性或非脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸,和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是色氨酸、甲硫氨酸或酪氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0110] 在一些实施方案中,包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211、具有如本文所述对应于 X190 的残基处的特点的氨基酸序列的立体选择性酮还原酶多肽可另外具有以下特点的一个或多个或至少所有特点:对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、色氨酸、脯氨酸、苏氨酸或精氨酸;对应于 X147 的残基是芳族、极性、非极性或非脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是色氨酸、甲硫氨酸或酪氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,

1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16, 18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0111] 在一些实施方案中,包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域(诸如残基 90-211)、具有如本文所述对应于 X190 的残基处的特点的氨基酸序列的立体选择性酮还原酶多肽可另外具有以下特点的一个或多个或至少所有特点:对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、色氨酸、脯氨酸、苏氨酸或精氨酸;对应于 X97 的残基是极性、非极性、脂肪族或碱性残基,特别是甲硫氨酸、缬氨酸、异亮氨酸、苏氨酸或精氨酸;对应于 X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是色氨酸、甲硫氨酸或酪氨酸。在一些实施方案中,与 SEQ ID NO :2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0112] 在一些实施方案中,包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域诸如残基 90-211、具有如本文所述对应于 X190 的残基处的特点的氨基酸序列的立体选择性酮还原酶多肽可另外具有以下特点的一个或多个或至少所有特点:对应于 X94 的残基是半胱氨酸、非极性或脂肪族残基,特别是半胱氨酸或缬氨酸;对应于 X96 的残基是极性 or 酸性残基,特别是 苏氨酸;和对应于 X147 的残基是芳族、极性、非极性 or 脂肪族残基,特别是谷氨酰胺、异亮氨酸 or 亮氨酸。在一些实施方案中,与 SEQ ID NO :2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0113] 在一些实施方案中,包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域诸如残基 90-211、具有如本文所述对应于 X190 的残基处的特点的氨基酸序列的立体选择性酮还原酶多肽可另外具有以下特点的一个或多个或至少所有特点:对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、色氨酸、脯氨酸、苏氨酸或精氨酸;对应于

X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是缬氨酸、亮氨酸、甲硫氨酸、苯丙氨酸或异亮氨酸;和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是色氨酸、甲硫氨酸或酪氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0114] 在一些实施方案中,包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211、具有如本文所述对应于 X190 的残基处的特点的氨基酸序列的立体选择性酮还原酶多肽可另外具有以下特点的一个或多个或至少所有特点:对应于 X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是缬氨酸、亮氨酸、甲硫氨酸、苯丙氨酸或异亮氨酸;和对应于 X202 的残基是脂肪族、芳族或非极性残基。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0115] 在一些实施方案中,包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211、具有如本文所述对应于 X190 的残基处的特点的氨基酸序列的立体选择性酮还原酶多肽可另外具有以下特点的一个或多个或至少所有特点:对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、色氨酸、脯氨酸、苏氨酸或精氨酸;对应于 X96 的残基是极性 or 酸性残基,特别是苏氨酸;对应于 X147 的残基是芳族、极性、非极性 or 脂肪族残基,特别是谷氨酰胺、异亮氨酸 or 亮氨酸;对应于 X196 的残基是脂肪族、非极性 or 芳族残基,特别是缬氨酸、亮氨酸、甲硫氨酸、苯丙氨酸 or 异亮氨酸;和对应于 X202 的残基是脂肪族、芳族 or 非极性残基,特别是色氨酸、甲硫氨酸 or 酪氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:

2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0116] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;和对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、色氨酸、脯氨酸、苏氨酸或精氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0117] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X16 的残基是极性残基,特别是丝氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0118] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;和对应于 X43 的残基是非极性或非脂肪族残基,特别是异亮氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0119] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式

或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X60 的残基是芳族、非极性或脂肪族残基,特别是丙氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0120] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X94 的残基是半胱氨酸、非极性或脂肪族残基,特别是半胱氨酸或缬氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0121] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X95 的残基是非极性或脂肪族残基,特别是亮氨酸或异亮氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0122] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X96 的残基是极性 or 酸性残基,特别是苏氨酸或谷氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或

98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0123] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X97 的残基是极性、非极性、脂肪族或碱性残基,特别是甲硫氨酸、缬氨酸、异亮氨酸、苏氨酸或精氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0124] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X120 的残基是芳族、非极性或非脂肪族残基,特别是缬氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0125] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X125 的残基是极性或非极性残基,特别是丝氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,

5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0126] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X142 的残基是极性残基,特别是天冬酰胺。在一些实施方案中,与 SEQ ID NO :2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0127] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X147 的残基是芳族、极性、非极性、或脂肪族残基,特别是谷氨酰胺、亮氨酸或异亮氨酸。在一些实施方案中,与 SEQ ID NO :2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0128] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X149 的残基是非极性或非芳族残基,特别是苯丙氨酸。在一些实施方案中,与 SEQ ID NO :2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考

序列具有至少 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 或 99% 同一性。

[0129] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X150 的残基是受限或酸性残基,特别是组氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0130] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X152 的残基是非极性或极性残基,特别是甲硫氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0131] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的序列式的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0132] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式

或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是甲硫氨酸、酪氨酸或色氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0133] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;和对应于 X205 的残基是碱性、非极性或非脂肪族残基,特别是精氨酸或缬氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0134] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于 X206 的残基是非极性或非芳族残基,特别是酪氨酸。在一些实施方案中,与 SEQ ID NO:2,4 或 98 的参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0135] 在一些实施方案中,本公开的改进的酮还原酶包含具有列于下文表 2 中突变集的任一种的氨基酸序列。在一些实施方案中,该多肽与包含选自 SEQ DyF NO:6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的多肽具有至少 85%,

86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或99%序列同一性,其中该酮还原酶多肽的氨基酸序列包括列于表2中的取代组合的任一套。在一些实施方案中,与参考序列相比,酮还原酶多肽可另外具有在其他氨基酸残基处的1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35或约1-40个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35或约40个残基差异。在一些实施方案中,差异包含保守突变。

[0136] 在一些实施方案中,改进的酮还原酶多肽包含选自SEQ ID NO:6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,78,80,82,84,86,88,90,92和94的氨基酸序列。

[0137] 在一些实施方案中,改进的酮还原酶包含基于SEQ ID NO:95,96或119的序列式或其区域诸如残基90-211的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于X190的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸,和对应于X196的残基是脂肪族、非极性或非极性芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸。在一些实施方案中,与具有前述特点的SEQ ID NO:2,4或98的参考序列诸如SEQ ID NO:8,10,14,16,24,26或48相比,酮还原酶多肽可另外具有在其他氨基酸残基处的1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35或约1-40个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35或约40个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于SEQ ID NO:2,4或98的参考序列(例如SEQ ID NO:8,10,14,16,24,26或48)具有至少85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或99%同一性。

[0138] 在一些实施方案中,改进的酮还原酶包含基于SEQ ID NO:95,96或119的序列式或其区域诸如残基90-211的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于X190的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于X125的残基是极性或非极性残基,特别是丝氨酸;和对应于X196的残基是脂肪族、非极性或非极性芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸。在一些实施方案中,与具有前述特点的SEQ ID NO:2,4或98的参考序列诸如SEQ ID NO:52相比,酮还原酶多肽可另外具有在其他氨基酸残基处的1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35或约1-40个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35或约40个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于SEQ ID NO:2,4或98的参考序列(例如SEQ ID NO:52)具有至少85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或99%同一性。

[0139] 在一些实施方案中,改进的酮还原酶包含基于SEQ ID NO:95,96或119的序列式或其区域诸如残基90-211的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于X190的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于X95的残

基是非极性或脂肪族残基,特别是亮氨酸或异亮氨酸;和对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ IDNO:20,62 或 64 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:20,62 或 64)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0140] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X206 的残基是非极性或芳族残基,特别是酪氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:36 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:36)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0141] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、色氨酸、脯氨酸、苏氨酸或精氨酸;和对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:54 或 56 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ IDNO:54 或 56)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0142] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;和对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:22,66,68 或 72 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:22,66,68 或 72)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0143] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是甲硫氨酸、酪氨酸或色氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:28,30 或 32 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:28,30 或 32)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0144] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X152 的残基是非极性或非极性残基,特别是甲硫氨酸;对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X205 的残基是碱性、非极性或非极性残基,特别是精氨酸或缬氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:20 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原

酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列(例如 SEQ ID NO :20)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0145] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X43 的残基是非极性或非脂肪族残基,特别是异亮氨酸;对应于 X147 的残基是芳族、极性、非极性或非脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;对应于 X196 的残基是脂肪族、非极性或非芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO :2,4 或 98 的参考序列诸如 SEQ ID NO :70 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列(例如 SEQ ID NO :70)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0146] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X94 的残基是半胱氨酸、非极性或非脂肪族残基,特别是半胱氨酸或缬氨酸;对应于 X196 的残基是脂肪族、非极性或非芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X205 的残基是芳族、极性、非极性或非脂肪族残基,特别是谷氨酰胺、亮氨酸或异亮氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO :2,4 或 98 的参考序列诸如 SEQ ID NO :34 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO :2,4 或 98 的参考序列(例如 SEQ ID NO :34)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0147] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO :95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X97 的残基是极性、非极性、脂肪族或碱性残基,特别是精氨酸、缬氨酸、甲硫氨酸、苏氨酸或异亮氨酸;对应于 X147 的残基是芳族、极性、非极性或非脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;和对应于 X196 的残基是脂肪族、非极性或非芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO :2,4 或 98 的参考

序列诸如 SEQ ID NO:74 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:74)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0148] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、苏氨酸、脯氨酸、色氨酸或精氨酸;对应于 X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是甲硫氨酸、酪氨酸或色氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:40,76,78,80 或 82 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:40,76,78,80 或 82)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0149] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、苏氨酸、脯氨酸、色氨酸或精氨酸;对应于 X94 的残基是半胱氨酸、非极性或脂肪族残基,特别是半胱氨酸或缬氨酸;对应于 X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、亮氨酸或异亮氨酸;对应于 X196 的残基是脂肪族、非极性或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是甲硫氨酸、酪氨酸或色氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:42 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如

SEQ ID NO:42) 具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0150] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、苏氨酸、脯氨酸、色氨酸或精氨酸;对应于 X94 的残基是半胱氨酸、非极性或脂肪族残基,特别是半胱氨酸或缬氨酸;对应于 X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、亮氨酸或异亮氨酸;对应于 X149 的残基是非极性或非芳族残基,特别是苯丙氨酸;对应于 X150 的残基是受限或酸性残基,特别是组氨酸;对应于 X196 的残基是脂肪族、非极性或非芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是甲硫氨酸、酪氨酸或色氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:84 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:84)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0151] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、苏氨酸、脯氨酸、色氨酸或精氨酸;对应于 X96 的残基是极性或非极性残基,特别是苏氨酸或谷氨酸;对应于 X147 的残基是芳族、极性、非极性或非芳族残基,特别是谷氨酰胺、亮氨酸或异亮氨酸;对应于 X196 的残基是脂肪族、非极性或非芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是甲硫氨酸、酪氨酸或色氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:44 或 46 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:44 或 46)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0152] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于

X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、苏氨酸、脯氨酸、色氨酸或精氨酸;对应于 X96 的残基是极性、或酸性残基,特别是苏氨酸或谷氨酸;对应于 X120 的残基是芳族、非极性、或脂肪族残基,特别是缬氨酸;对应于 X147 的残基是芳族、极性、非极性、或脂肪族残基,特别是谷氨酰胺、亮氨酸或异亮氨酸;对应于 X196 的残基是脂肪族、非极性、或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X202 的残基是脂肪族、芳族、或非极性残基,特别是甲硫氨酸、酪氨酸或色氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:86 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:86)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0153] 在一些实施方案中,改进的酮还原酶包含基于 SEQ ID NO:95,96 或 119 的序列式或其区域诸如残基 90-211 的氨基酸序列,其中该氨基酸序列具有至少以下特点:对应于 X190 的残基是非芳族残基,特别是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸;对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基,特别是组氨酸、苏氨酸、脯氨酸、色氨酸或精氨酸;对应于 X97 的残基是极性、非极性、脂肪族或碱性残基,特别是缬氨酸、甲硫氨酸、苏氨酸或异亮氨酸;对应于 X147 的残基是芳族、极性、非极性、或脂肪族残基,特别是谷氨酰胺、亮氨酸或异亮氨酸;对应于 X196 的残基是脂肪族、非极性、或芳族残基,特别是甲硫氨酸、异亮氨酸、亮氨酸或苯丙氨酸;和对应于 X202 的残基是脂肪族、芳族、或非极性残基,特别是甲硫氨酸、酪氨酸或色氨酸。在一些实施方案中,与具有前述特点的 SEQ ID NO:2,4 或 98 的参考序列诸如 SEQ ID NO:88,90,92 或 94 相比,酮还原酶多肽可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18,1-20,1-22,1-24,1-26,1-30,1-35 或约 1-40 个残基差异。在一些实施方案中,差异的数目可以是在其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18,20,22,24,26,30,35 或约 40 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,酮还原酶多肽包含具有前述特点的氨基酸序列,并且其中氨基酸序列与具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列(例如 SEQ ID NO:88,90,92 或 94)具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0154] 在一些实施方案中,本公开的改进的酮还原酶包含具有对应于 SEQ ID NO:95,96 或 119 的序列式的残基 90-211 的区域或结构域的氨基酸序列,其中对应于 X190 的残基不是酪氨酸。在一些实施方案中,对应于残基 90-211 的结构域或区域包含其中对应于 X190 的残基是非芳族残基诸如脂肪族、受限、非极性、或半胱氨酸残基的氨基酸序列。在一些实施方案中,对应于残基 90-211 的结构域或区域包含其中对应于 X190 的残基是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸特别是脯氨酸的氨基酸序列。在一些实施方案中,与基于 SEQ ID NO:2,4 或 98 的参考序列的对应结构域相比,对应于残基 90-211 的区域或结构域可另外具有

在其他氨基酸残基处的 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18 或 1-20 个残基差异。在一些实施方案中, 差异的数目可以是在结构域的其他氨基酸残基处的 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18 或 20 个残基差异。在一些实施方案中, 差异包含保守突变。在一些实施方案中, 酮还原酶多肽包含具有对应于 SEQ ID NO :95, 96 或 119 的序列式的残基 90-211 的结构域或区域的氨基酸序列, 其中对应于 X190 的残基具有至少前述特点, 并且其中该结构域或区域的氨基酸序列与对应于具有前述特点的基于 SEQ ID NO :2, 4 或 98 的参考序列的残基 90-211 的氨基酸序列具有至少 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 或 99% 同一性。

[0155] 在一些实施方案中, 具有对应于 SEQ ID NO :95, 96 或 119 的序列式的残基 90-211 结构域或区域、并且在对应于 X190 的残基处具有如本文所述的指定特点的酮还原酶多肽的结构域或区域还可包括选自以下的一个或多个特点: 对应于 X94 的残基是半胱氨酸、非极性或非脂肪族残基; 对应于 X95 的残基是非极性或非脂肪族残基; 对应于 X96 的残基是极性或非脂肪族残基; 对应于 X97 的残基是极性、非极性、脂肪族或碱性残基; 对应于 X120 的残基是芳香族、非极性或非脂肪族残基; 对应于 X125 的残基是极性或非极性残基; 对应于 X147 的残基是芳香族、极性、非极性或非脂肪族残基; 对应于 X149 的残基是非极性或非脂肪族残基; 对应于 X150 的残基是受限或非脂肪族残基; 对应于 X152 的残基是非极性或非脂肪族残基; 对应于 X196 的残基是脂肪族、非极性或非脂肪族残基; 对应于 X202 的残基是脂肪族、芳香族或非极性残基; 对应于 X205 的残基是碱性、非极性或非脂肪族残基; 和对应于 X206 的残基是非极性或非脂肪族残基。在一些实施方案中, 与基于 SEQ ID NO :2, 4 或 98 的参考序列的对应结构域相比, 对应于残基 90-211 的区域或结构域可另外具有在其他氨基酸残基处的 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18 或 1-20 个残基差异。在一些实施方案中, 差异的数目可以是在结构域的其他氨基酸残基处的 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18 或约 20 个残基差异。在一些实施方案中, 差异包含保守突变。

[0156] 在一些实施方案中, 与 SEQ ID NO :2, 4 或 98 的对应结构域的氨基酸序列相比, 具有对应于如上文所述 SEQ ID NO :95, 96 或 119 的序列式的残基 90-211 的氨基酸序列的结构域或区域的酮还原酶多肽可具有在该结构域或区域中的一个或多个保守突变。此类保守突变的实例包括氨基酸置换, 诸如但不限于: 用另一个非极性或非脂肪族残基例如缬氨酸、亮氨酸或异亮氨酸置换对应于 X94 的残基丙氨酸 (A); 用另一个非极性或非脂肪族残基例如丙氨酸、亮氨酸或异亮氨酸置换对应于 X95 的残基缬氨酸 (V); 用另一个极性残基例如天冬酰胺、谷氨酰胺或苏氨酸置换对应于 X96 的残基丝氨酸 (S); 用另一个非极性或非脂肪族残基例如丙氨酸、亮氨酸或异亮氨酸置换对应于 X196 的残基缬氨酸 (V); 和用另一个非极性或非脂肪族残基例如缬氨酸、亮氨酸或异亮氨酸置换对应于 X205 的残基丙氨酸 (A)。

[0157] 在一些实施方案中, 具有对应于 SEQ ID NO :95, 96 或 119 的序列式的残基 90-211 的结构域或区域、并且具有如本文所述对应于 X190 的残基处的指定特点的酮还原酶多肽的区域或结构域还可包括选自以下的特点的一个或多个: 对应于 X94 的残基是半胱氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸, 特别是丙氨酸、缬氨酸或半胱氨酸; 对应于 X95 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸, 特别是异亮氨酸或亮氨酸; 对应于 X96 的残基是天冬氨酸、谷氨酸、丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺, 特别是丝氨酸、天冬酰胺、苏氨酸或谷氨酸; 对应于 X97 的残基是丝氨酸、苏氨酸、天冬

酰胺、谷氨酰胺、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、赖氨酸或精氨酸,特别是赖氨酸、苏氨酸、缬氨酸、精氨酸、甲硫氨酸或异亮氨酸;对应于 X120 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是苯丙氨酸或缬氨酸;对应于 X125 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是甘氨酸或丝氨酸;对应于 X142 的残基是丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺残基,特别是天冬酰胺;对应于 X147 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺,特别是苯丙氨酸、亮氨酸、异亮氨酸、缬氨酸或谷氨酰胺;对应于 X149 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸或色氨酸,特别是甘氨酸或苯丙氨酸;对应于 X150 的残基是脯氨酸、组氨酸、天冬氨酸或谷氨酸,特别是天冬氨酸或组氨酸;对应于 X152 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺,特别是丝氨酸、苏氨酸或甲硫氨酸;对应于 X196 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是缬氨酸、亮氨酸、甲硫氨酸、苯丙氨酸或异亮氨酸;对应于 X202 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是丙氨酸、色氨酸、酪氨酸或甲硫氨酸;对应于 X205 的残基是赖氨酸、精氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是精氨酸;并且对应于 X206 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸、色氨酸,特别是甲硫氨酸或酪氨酸。在一些实施方案中,与基于 SEQ ID NO:2,4 或 98 的参考序列的对应结构域相比,对应于残基 90-211 的区域或结构域可另外具有在其他氨基酸残基处的约 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18 或 1-20 个残基差异。在一些实施方案中,差异的数目可以是在结构域的其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18 或约 20 个残基差异。在一些实施方案中,差异包含保守突变。

[0158] 在一些实施方案中,具有对应于 SEQ ID NO:95,96 或 119 的序列式的残基 90-211 的结构域或区域、并且具有如本文所述对应于 X190 的残基处的指定特点的酮还原酶多肽的区域或结构域内还可包括选自以下的一个或多个或所有特点:对应于 X147 的残基是芳族、极性、非极性 or 脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸,对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是色氨酸、甲硫氨酸或酪氨酸。在一些实施方案中,与基于 SEQ ID NO:2,4 或 98 的参考序列的结构域相比,对应于残基 90-211 的区域或结构域可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18 或 1-20 个残基差异。在一些实施方案中,差异的数目可以是在结构域的其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18 或约 20 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,该酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中该氨基酸序列与对应于具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列的残基 90-211 的氨基酸序列相比具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0159] 在一些实施方案中,具有对应于 SEQ ID NO:95,96 或 119 的序列式的残基 90-211 的结构域或区域、并且具有如本文所述对应于 X190 的残基处的指定特点的酮还原酶多肽的区域或结构域还可包括选自以下的一个或多个或所有特点:对应于 X97 的残基是极性、

非极性、脂肪族或碱性残基,特别是甲硫氨酸、缬氨酸、异亮氨酸、苏氨酸或精氨酸;对应于 X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸;对应于 X202 的残基是脂肪族、芳族或非极性残基,特别是色氨酸、甲硫氨酸或酪氨酸。在一些实施方案中,与基于 SEQ ID NO:2,4 或 98 的参考序列的结构域相比,对应于残基 90-211 的区域或结构域可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18 或 1-20 个残基差异。在一些实施方案中,差异的数目可以是在结构域的其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18 或约 20 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,该酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中该氨基酸序列与对应于具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列的残基 90-211 的氨基酸序列相比具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0160] 在一些实施方案中,具有对应于 SEQ ID NO:95,96 或 119 的序列式的残基 90-211 的结构域或区域、并且具有如本文所述对应于 X190 的残基处的指定特点的酮还原酶多肽的区域或结构域还可包括选自以下的一个或多个或所有特点:对应于 X94 的残基是半胱氨酸、非极性或脂肪族残基,特别是半胱氨酸或缬氨酸;对应于 X96 的残基是极性 or 酸性残基,特别是苏氨酸;对应于 X147 的残基是芳族、极性、非极性或脂肪族残基,特别是谷氨酰胺、异亮氨酸或亮氨酸。在一些实施方案中,与基于 SEQ ID NO:2,4 或 98 的参考序列的结构域相比,对应于残基 90-211 的区域或结构域可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18 或 1-20 个残基差异。在一些实施方案中,差异的数目可以是在结构域的其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,18 或约 20 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,该酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中该氨基酸序列与对应于具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列的残基 90-211 的氨基酸序列相比具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性。

[0161] 在一些实施方案中,该酮还原酶多肽还可包括对应于 SEQ ID NO:95,96 或 119 的序列式的残基 1-89 的区域或结构域。在一些实施方案中,对应于残基 1-89 的区域或结构域可具有以下特点的一种或多种:对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基;对应于 X16 的残基是极性残基;对应于 X43 的残基是非极性 or 极性残基;对应于 X60 的残基是芳族或非极性 or 脂肪族残基。

[0162] 在一些实施方案中,对应于残基 1-89 的结构域或区域与对应于基于 SEQ ID NO:2,4 或 98 的参考序列的残基 1-89 的氨基酸序列具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%同一性,条件是该酮还原酶多肽的区域或结构域具有其中对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基特别是组氨酸的氨基酸序列,所述参考序列在对应于 X7 的残基处具有芳族、非极性、极性、受限或碱性残基特别是组氨酸。在一些实施方案中,与基于 SEQ ID NO:2,4 或 98 的参考序列的结构域相比,对应于残基 1-89 的区域或结构域可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18 或 1-20 个残基差异。在一些实施方案中,差异的数目可以是在结构域的其他氨基酸残基处的 1,2,3,4,5,6,7,8,

9,10,11,12,14,15,16,18 或约 20 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,该酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中该氨基酸序列与对应于具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列的残基 1-89 的氨基酸序列相比具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0163] 在一些实施方案中,对应于残基 1-89 的区域或结构域可具有一种或多种或至少所有以下特点:对应于 X7 的残基是芳族、非极性、极性、受限或碱性残基;对应于 X16 的残基是极性残基;对应于 X43 的残基是非极性或极性残基;对应于 X60 的残基是芳族或非极性或非脂肪族残基。在一些实施方案中,与基于 SEQ ID NO:2,4 或 98 的参考序列的结构域相比,对应于残基 1-89 的区域或结构域可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15,1-16,1-18 或 1-20 个残基差异。在一些实施方案中,差异的数目可以是在结构域的其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15 或 16 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,该酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中该氨基酸序列与具有前述特点的对对应于基于 SEQ ID NO:2,4 或 98 的参考序列的残基 1-89 的氨基酸序列相比具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0164] 在一些实施方案中,对应于残基 1-89 的区域或结构域可具有一种或多种或至少所有以下特点:对应于 X7 的残基是色氨酸、酪氨酸、苯丙氨酸、脯氨酸、组氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、谷氨酰胺、天冬酰胺、精氨酸或赖氨酸,特别是甘氨酸、组氨酸、苏氨酸、脯氨酸、色氨酸、精氨酸、组氨酸或天冬酰胺;对应于 X16 的残基是丝氨酸、苏氨酸、天冬酰胺或谷氨酰胺,特别是丝氨酸;对应于 X43 的残基是甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是异亮氨酸;对应于 X60 的残基是酪氨酸、苯丙氨酸、色氨酸、甘氨酸、甲硫氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,特别是丙氨酸。在一些实施方案中,与基于 SEQ ID NO:2,4 或 98 的参考序列的结构域相比,对应于残基 1-89 的区域或结构域可另外具有在其他氨基酸残基处的 1-2,1-3,1-4,1-5,1-6,1-7,1-8,1-9,1-10,1-11,1-12,1-14,1-15 或 1-16 个残基差异。在一些实施方案中,差异的数目可以是在结构域的其他氨基酸残基处的 1,2,3,4,5,6,7,8,9,10,11,12,14,15 或 16 个残基差异。在一些实施方案中,差异包含保守突变。在一些实施方案中,该酮还原酶多肽包含具有至少前述特点的氨基酸序列,并且其中该氨基酸序列与对应于具有前述特点的基于 SEQ ID NO:2,4 或 98 的参考序列的残基 1-89 的氨基酸序列相比具有至少 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性。

[0165] 下文表 2 提供了本文公开的具有与取代苯乙酮的还原相关的活性水平的一些 SEQ ID NO 的列表。下文所有序列来自野生型克菲尔乳杆菌酮还原酶序列 (SEQ ID NO:3 和 4),除非另外指明。

[0166] 表 2:序列列表

[0167]

多核苷酸 SEQ ID NO.	多肽 SEQ ID NO.	与 SEQ ID NO:4 相比 多肽残基的取代	活性	稳定性
3	4	无 (野生型克菲尔乳杆菌序列)	-	-
5	6	Y190C	+	-
7	8	Y190A ;V196I ;	++	+
9	10	Y190P ;V196I ;	++	
11	12	K97R ;Y190C ;	+	
13	14	Y190P ;V196L	++	
15	16	Y190A ;V196L	++	+
47	48	Y190P ;V196M ;	+	
49	50	Y190P ;V196F ;	+	
51	52	G125S ;Y190P ;V196L ;	+	
19	20	V95I ;Y190P ;V196I ;	+	+
17	18	T152M ;Y190P ;V196I ;M205R ;	+++	+
23	24	Y190P ;V196L ; (与 SEQ ID NO:2 相比)	++	
25	26	Y190A ;V196L ; (与 SEQ ID NO:2 相比)	++	+
35	36	Y190P ;V196L ;M206Y ;	+++	
53	54	G7N ;Y190A ;V196L ;	++	+
55	56	G7H ;Y190P ;V196L ;	++	
61	62	V95L ;Y190P ;V196L ;	++	
63	64	V95I ;Y190P ;V196L ;	++	
57	58	T16S ;Y190p ;V196L ;	++	
59	60	Y190P ;V196L	++	
21	22	F147L ;Y190P ;V196L ;	++	++
65	66	F147Q ;Y190A ;V196L ;	++	+
67	68	F147I ;Y190P ;V196L ;	++	++
69	70	V43I ;F147L ;Y190P ;V196L ;	++	++
71	72	F147L ;Y190A ;V196L ;	++	++
31	32	Y190P ;V196L ;A202W ;	+++	+
27	28	Y190P ;V196L ;A202M ;	++	+
29	30	Y190A ;V196L ;A202Y ;	++	+
33	34	A94V ;Y190A ;V196L ;M205V ;	+++	+
73	74	K97R ;F147I ;Y190P ;V196L ;	+++	+
37	38	F147L ;Y190P ;V196L ;A202W ;	+++	++
39	40	G7H ;F147L ;Y190P ;V196L ;A202W ;	+++	++
75	76	G7T ;F147L ;Y190P ;V196L ;A202W ;	+++	++
77	78	G7P ;F147L ;Y190P ;V196L ;A202W ;	+++	++
79	80	G7W ;F147L ;Y190P ;V196L ;A202W ;	+++	++
81	82	G7R ;F147L ;Y190P ;V196L ;A202W ;	+++	++
83	84	G7H ;A94C ;F147L ;G149F ;D150H ;Y190P ;V196L ;A202W ;	+++	++
41	42	G7H ;A94V ;F147L ;Y190P ;V196L ;A202W ;	+++	++
43	44	G7H ;S96E ;F147L ;Y190P ;V196L ;A202W ;	+++	++
85	86	G7H ;S96T ;F120V ;F147L ;Y190P ;V196L ;A202W ;	+++	++
45	46	G7H ;S96T ;F147L ;Y190P ;V196L ;A202W ;	+++	++
87	88	G7H ;K97V ;F147L ;Y190P ;V196L ;A202W ;	+++	++
89	90	G7H ;K97M ;F147L ;Y190P ;V196L ;A202W ;	+++	++
91	92	G7H ;K97T ;F147L ;Y190P ;V196L ;A202W ;	+++	++
93	94	G7H ;K97I ;F147L ;Y190P ;V196L ;A202W ;	+++	++

[0168] 在上文表 2 的活性列中,单个加号“+”表示 SEQ ID NO :6 的活性的 100-450% 的活性改进,两个加号“++”表示 SEQ ID NO :6 的 450-1500% 的活性改进,三个加号“+++”表示 SEQ ID NO :6 的 大于 1500% 的活性改进。在稳定性列中,单个加号“+”表示在 50℃ 下热处理 2 小时之后该多肽显示可测量的活性,两个加号“++”表示当比较两个蛋白在 50℃ 下热处理 2 小时之后的活性时,与 SEQ ID NO :16 相比该多肽的活性具有大于 400% 的改进。

[0169] 在一些实施方案中,与具有 (S) 选择性的工程 KRED 酶例如 SEQ ID NO :6 相比,本公开的酮还原酶多肽在其酶促活性速率例如其将底物转化为产物的速率方面具有改进。本文使用具有 SEQ ID NO :6 的序列的多肽作为参考多肽,因为野生型克菲尔乳杆菌或短乳杆菌 KRED 没有表现出可估计的将 2',6' - 二氯 -3' - 氟苯乙酮转化为 (S)-1-[2,6- 二氯 -3- 氟苯基]-乙醇的活性。在一些实施方案中,该酮还原酶多肽能以超过 SEQ ID NO :6 的速率的至少 5 倍、10 倍、25 倍、50 倍、75 倍、100 倍、150 倍、200 倍、250 倍或 300 倍的速率将底物转化为产物。在一些实施方案中,该酮还原酶多肽能以 SEQ ID NO :6 的速率的至少 100%,150%,200%,250%,300%,400%,450%,500%,750%,1000%,1250% 或 1500% 将底物转化为产物。

[0170] 在一些实施方案中,本公开的酮还原酶多肽能以高于 99% 的立体异构过量并且以比具有 SEQ ID NO :6 的序列的酮还原酶多肽改进的速率将 2',6' - 二氯 -3' - 氟苯乙酮转化为 (S)-1-[2,6- 二氯 -3- 氟苯基]-乙醇。比 SEQ ID NO :6 的酶促活性改进的示例性多肽包括但不限于包含对应于 SEQ ID NO :8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的多肽。

[0171] 在一些实施方案中,本公开的酮还原酶多肽能以高于 99% 的立体异构过量并且以比具有 SEQ ID NO :6 的序列的酮还原酶多肽改进的速率将 2',6' - 二氯 -3' - 氟苯乙酮转化为 (S)-1-[2,6- 二氯 -3- 氟苯基]-乙醇,其中该多肽与具有 SEQ ID NO :6 的序列的多肽相比还具有改进的热稳定性。具有此类改进的示例性多肽包括但不限于包含对应于 SEQ ID NO :8,16,18,20,22,26,28,30,32,34,38,40,42,44,46,54,66,68,70,72,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的多肽。

[0172] 在一些实施方案中,本公开的酮还原酶多肽能以高于 99% 的立体异构过量并且以比具有 SEQ ID NO :6 的序列的酮还原酶多肽高至少约 450% 的速率将 2',6' - 二氯 -3' - 氟苯乙酮转化为 (S)-1-[2,6- 二氯 -3- 氟苯基]-乙醇。能具有此类改进的示例性多肽包括但不限于包含对应于 SEQ ID NO :8,10,14,16,18,22,24,26,28,30,32,34,36,38,40,42,44,46,54,56,58,60,62,64,66,68,70,72,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的多肽。

[0173] 在一些实施方案中,本公开的酮还原酶多肽能以高于 99% 的立体异构过量并且以比具有 SEQ ID NO :6 的序列的酮还原酶多肽高至少约 450% 的速率将 2',6' - 二氯 -3' - 氟苯乙酮转化为 (S)-1-[2,6- 二氯 -3- 氟苯基]-乙醇,其中该多肽与具有 SEQ ID NO :6 的序列的多肽相比还具有改进的热稳定性。具有此类特性的示例性多肽包括但不限于包含对应于 SEQ ID NO :8,16,18,22,26,28,30,32,34,38,40,42,44,46,54,66,68,70,72,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的多肽。

[0174] 在一些实施方案中,本公开的酮还原酶多肽能以高于 99% 的立体异构过量并且以比具有 SEQ ID NO:6 的序列的酮还原酶多肽高至少约 1500% 的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇。能具有此类改进的示例性多肽包括但不限于包含对应于 SEQ ID NO:18,32,34,36,38,40,42,44,46,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的多肽。

[0175] 在一些实施方案中,本公开的酮还原酶多肽能以高于 99% 的立体异构过量并且以比具有 SEQ ID NO:6 的序列的酮还原酶多肽高至少约 1500% 的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇,其中该多肽与 SEQ ID NO:6 的多肽相比还具有改进的热稳定性。具有此类特性的示例性多肽包括但不限于包含对应于 SEQ ID NO:18,32,34,36,38,40,42,44,46,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的多肽。

[0176] 在一些实施方案中,当使用相比 2',6'-二氯-3'-氟苯乙酮底物的量按重量计小于约 1% 的量的多肽进行时,本公开的酮还原酶多肽能在不到约 24 小时内以至少约 99% 的立体异构过量将至少约 95% 的 2',6'-二氯-3'-氟苯乙酮底物转化为 (S)-1-(2,6-二氯-3-氟苯基)乙醇。具有该能力的示例性多肽包括但不限于包含对应于 SEQ ID NO:18,32,34,36,38,40,42,44,46,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的多肽。

[0177] 在一些实施方案中,本公开的酮还原酶多肽能以高于 99% 的立体异构过量并且以比具有 SEQ ID NO:6 的序列的酮还原酶多肽高至少约 450% 的速率将 2',6'-二氯-3'-氟苯乙酮转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇,其中在 50°C 下热处理 2 小时后该多肽还能以比具有 SEQ ID NO:16 的序列的多肽高至少约 400% 的速率将底物转化为产物(其中 SEQ ID NO:16 的多肽也用相同的热处理来处理)。具有此类特性的示例性多肽包括但不限于包含对应于 SEQ ID NO:18,32,34,36,38,40,42,44,46,74,76,78,80,82,84,86,88,90,92 或 94 的氨基酸序列的多肽。

[0178] 在一些实施方案中,该酮还原酶多肽能以至少约 99% 的百分比 e. e. 将底物立体选择性地还原为产物,其中该多肽包含对应于 SEQ ID NO:6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,78,80,82,84,86,88,90,92 或 94 的氨基酸序列。

[0179] 在一些实施方案中,该酮还原酶多肽能以至少约 25%,50%,75%,80%,85%,85%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99%,99.9% 或 99.99% 的百分比立体异构过量将 2',6'-二氯-3'-氟苯乙酮立体选择性地还原为 (S)-1-(2,6-二氯-3-氟苯基)乙醇。

[0180] 在一些实施方案中,该酮还原酶多肽可包含与 SEQ ID NO:2,4 或 98 或其区域或结构域诸如残基 90-211 具有至少约 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98% 或 99% 同一性的氨基酸序列,条件是对应于残基 X190 的残基不是酪氨酸,特别地对应于残基 X190 的残基是非芳族残基,并且其中该多肽能以至少约 85% e. e 将底物还原为产物。在一些实施方案中,对应于 X190 的残基是脂肪族、受限、非极性或半胱氨酸残基。在一些实施方案中,对应于 X190 的残基是脯氨酸,并且另外具有以下取代的一种或多种,以便该多肽与野生型克菲尔乳杆菌酮还原酶或另一种工程酮还原酶相

比进一步改进（关于立体选择性、酶促活性和 / 或热稳定性）：7 → H, T, P, W, R, N（即对应于 SEQ ID NO :2,4 或 98 的残基 7 的残基被取代为组氨酸、苏氨酸、脯氨酸、色氨酸、精氨酸或天冬酰胺）；16 → S；43 → I；60 → A；94 → C, V；95 → I, L；96 → E, T；97 → R, V, M, T, I；120 → V；125 → S；142 → N；147 → L, Q, I, V；149 → F；150 → H；152 → H；196 → I, L, M, F；202 → W, M, F 和 206 → Y。在一些实施方案中，对应于 X190 的残基是脯氨酸，并且另外具有以下取代的一种或多种，以便该多肽与野生型克菲尔乳杆菌酮还原酶或另一种工程酮还原酶相比进一步改进：7 → H；94 → V；96 → T；147 → L；196 → L 和 202 → W。

[0181] 如本领域技术人员将理解，除非另外指明，上文定义的一些分类不是相互排斥的。因此，具有表现出两种或多种物理 - 化学特性的侧链的氨基酸可包括于多种分类中。任何氨基酸或残基的适当分类对本领域技术人员将是显而易见的，特别是根据本文提供的详细的公开内容。

[0182] 在一些实施方案中，改进的工程酮还原酶包含天然存在的酮还原酶多肽的缺失或其他工程酮还原酶多肽的缺失。在一些实施方案中，本文描述的每种改进的工程酮还原酶可包含本文描述的多肽的缺失。因此，对本公开的酮还原酶多肽的每一种实施方案，缺失可包含一个或多个氨基酸、2 个或更多个氨基酸、3 个或更多个氨基酸、4 个或更多个氨基酸、5 个或更多个氨基酸、6 个或更多个氨基酸、8 个或更多个氨基酸、10 个或更多个氨基酸、15 个或更多个氨基酸或 20 个或更多个氨基酸、多达酮还原酶多肽的氨基酸总数的 10%、多达酮还原酶多肽的氨基酸总数的 10%、多达酮还原酶多肽的氨基酸总数的 20% 或多达酮还原酶多肽的氨基酸总数的 30%，只要保持了酮还原酶活性的功能活性。在一些实施方案中，缺失可包含 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-25, 1-30, 1-35 或约 1-40 个氨基酸残基。在一些实施方案中，缺失的数目可以是 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 20, 22, 24, 26, 30, 35 或约 40 个氨基酸。在一些实施方案中，缺失可包含 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18 或 20 个氨基酸残基的缺失。

[0183] 如本文所述，本公开的酮还原酶多肽可以是以其中酮还原酶多肽与其他多肽融合的融合多肽的形式，所述其他多肽诸如抗体标签（例如 myc 表位）或纯化序列（例如 His 标签）。因此，可使用具有或没有与其他多肽融合的酮还原酶多肽。

[0184] 在一些实施方案中，本文描述的多肽不限于遗传编码的氨基酸。除了遗传编码的氨基酸外，本文描述的多肽可完全或部分地包含天然存在和 / 或合成的非编码氨基酸。本文所述的多肽可包含的一些经常遇到的非编码氨基酸，包括但不限于：遗传编码的氨基酸的 D- 立体异构体；2,3- 二氨基丙酸 (Dpr)； α - 氨基异丁酸 (Aib)； ϵ - 氨基己酸 (Aha)； δ - 氨基戊酸 (Ava)；N- 甲基甘氨酸或肌氨酸 (MeGly 或 Sar)；鸟氨酸 (Om)；瓜氨酸 (Cit)；叔丁基丙氨酸 (Bua)；叔丁基甘氨酸 (Bug)；N- 甲基异亮氨酸 (MeIle)；苯基甘氨酸 (Phg)；环己基丙氨酸 (Cha)；正亮氨酸 (NIle)；萘基丙氨酸 (Nal)；2- 氯苯丙氨酸 (Ocf)；3- 氯苯丙氨酸 (Mcf)；4- 氯苯丙氨酸 (Pcf)；2- 氟苯丙氨酸 (Off)；3- 氟苯丙氨酸 (Mff)；4- 氟苯丙氨酸 (Pff)；2- 溴苯丙氨酸 (Obf)；3- 溴苯丙氨酸 (Mbf)；4- 溴苯丙氨酸 (Pbf)；2- 甲基苯丙氨酸 (Omf)；3- 甲基苯丙氨酸 (Mmf)；4- 甲基苯丙氨酸 (Pmf)；2- 硝基苯丙氨酸 (Onf)；3- 硝基苯丙氨酸 (Mnf)；4- 硝基苯丙氨酸 (Pnf)；2- 氰基苯丙氨酸 (Ocf)；3- 氰基苯丙氨酸 (Mcf)；4- 氰基苯丙氨酸 (Pcf)；2- 三氟甲基苯丙氨酸 (Otf)；3- 三氟甲基苯丙氨酸 (Mtf)；

4- 三氟甲基苯丙氨酸 (Ptf) ; 4- 氨基苯丙氨酸 (Paf) ; 4- 碘苯丙氨酸 (Pif) ; 4- 氨基苯丙氨酸 (Pamf) ; 2, 4- 二氯苯丙氨酸 (Opaf) ; 3, 4- 二氯苯丙氨酸 (Mpcf) ; 2, 4- 二氟苯丙氨酸 (Opff) ; 3, 4- 二氟苯丙氨酸 (Mpff) ; 吡啶 -2- 基丙氨酸 (2pAla) ; 吡啶 -3- 基丙氨酸 (3pAla) ; 吡啶 -4- 基丙氨酸 (4pAla) ; 萘 -1- 基丙氨酸 (1nAla) ; 萘 -2- 基丙氨酸 (2nAla) ; 噻唑丙氨酸 (taAla) ; 苯并噻吩基丙氨酸 (bAla) ; 噻吩基丙氨酸 (tAla) ; 呋喃基丙氨酸 (fAla) ; 高苯丙氨酸 (hPhe) ; 高酪氨酸 (hTyr) ; 高色氨酸 (hTrp) ; 五氟苯丙氨酸 (5ff) ; 苯乙基丙氨酸 (styrylalanine) (sAla) ; 葱基丙氨酸 (authrylalanine, aAla) ; 3, 3- 二苯丙氨酸 (Dfa) ; 3- 氨基 -5- 苯基戊酸 (phenypentanoic acid) (Afp) ; 青霉素 (Pen) ; 1, 2, 3, 4- 四氢异喹啉 -3- 羧酸 (Tic) ; β -2- 噻吩基丙氨酸 (Thi) ; 甲硫氨酸亚砷 (Mso) ; N(w)- 硝基精氨酸 (nArg) ; 高赖氨酸 (hLys) ; 膦酰基甲基苯丙氨酸 (pmPhe) ; 磷酸丝氨酸 (pSer) ; 磷酸苏氨酸 (pThr) ; 高天冬氨酸 (hAsp) ; 高谷氨酸 (homoglutamic acid) (hGlu) ; 1- 氨基环戊 - (2 或 3)- 烯 -4 羧酸 ; 哌可酸 (PA) 、吡丁啉 -3- 羧酸 (ACA) ; 1- 氨基环戊烷 -3- 羧酸 ; 烯丙基甘氨酸 (aOly) ; 炔丙基甘氨酸 (pgGly) ; 高丙氨酸 (hAla) ; 正缬氨酸 (nVal) ; 高亮氨酸 (hLeu) 、高缬氨酸 (hVal) ; 高异亮氨酸 (homoisoleucine) (hIle) ; 高精氨酸 (hArg) ; N- 乙酰赖氨酸 (AcLys) ; 2, 4- 二氨基丁酸 (Dbu) ; 2, 3- 二氨基丁酸 (Dab) ; N- 甲基缬氨酸 (MeVal) ; 高半胱氨酸 (hCys) ; 高丝氨酸 (hSer) ; 羟脯氨酸 (Hyp) 和高脯氨酸 (hPro) 。可包含于本文描述的多肽的其他非编码氨基酸对本领域技术人员将是显而易见的 (参见, 例如, 提供于 Fasman, 1989, CRC Practical Handbook of Biochemistry and Molecular Biology (CRC 生物化学和分子生物学实践手册), CRC Press, Boca Raton, FL, 第 3-70 页中和其中引用的参考文献中的各种氨基酸, 其全部通过引用并入) 。这些氨基酸可以是 L- 或 D- 构型。

[0185] 本领域技术人员将承认, 具有侧链保护基团的氨基酸或残基也可构成本文描述的多肽。在本实例中属于芳族类别的此类保护氨基酸的非限制性实例包括 (保护基团列于括号内) 但不限于 : Arg(tos) 、 Cys(苄基) 、 Cys(硝基吡啶亚磺酰基) 、 Glu(δ - 苄基酯) 、 Gln(占吨基) 、 Asn(N- δ - 占吨基) 、 His(bom) 、 His(苄基) 、 His(tos) 、 Lys(fmoc) 、 Lys(tos) 、 Ser(O- 苄基) 、 Thr(O- 苄基) 和 Tyr(O- 苄基) 。

[0186] 可包括于本文描述的多肽的构象受限的非编码氨基酸包括但不限于, N- 甲基氨基酸 (L- 构型) ; 1- 氨基环戊 - (2 或 3)- 烯 -4 羧酸 ; 哌可酸 ; 吡丁啉 -3- 羧酸 ; 高脯氨酸 (hPro) 和 1- 氨基环戊烷 -3- 羧酸。

[0187] 如上文所述, 被引入天然存在的多肽以产生工程酮还原酶的各种修饰可被靶向酶的特定特性。

[0188] 6.3 编码工程酮还原酶的多核苷酸

[0189] 在另一个方面, 本公开提供了编码工程酮还原酶的多核苷酸。该多核苷酸可以与控制基因表达的一种或多种异源调节序列可操作地连接以产生能表达该多肽的重组多核苷酸。含有编码工程酮还原酶的异源多核苷酸的表达构建体可被引入适当的宿主细胞来表达对应的酮还原酶多肽。

[0190] 由于对应于各种氨基酸的密码子的知识, 蛋白序列的可用性提供了能编码该蛋白序列的所有多核苷酸的描述。其中相同的氨基酸由可选或同义密码子编码的遗传密码的简并性允许制备相当大量的核酸, 所有这些核酸编码本文公开的改进的酮还原酶。因此, 已鉴

定了特定的氨基酸序列,本领域技术人员可通过简单地以不改变蛋白质的氨基酸序列的方式修饰一个或多个密码子的序列来制备任何数目的不同的核酸。在这方面,本公开特别涵盖可通过选择基于可能的密码子选择的组合制备的多核苷酸的每一种可能的改变,并且,对于本文公开的任何多肽,包括表 2 中所示的氨基酸序列,所有此类改变被视为特别公开。在不同实施方案中,优选地选择密码子来适合其中产生蛋白的宿主细胞。例如,细菌中使用的优选密码子被用来在细菌中表达基因;酵母中使用的优选密码子被用于在酵母中的表达;哺乳动物中使用的优选密码子被用于在哺乳动物细胞中的表达。作为实例,SEQ ID NO: 3 的多核苷酸已被密码子优化用于在大肠杆菌 (*E. coli*) 中表达,但是仍编码克菲尔乳杆菌的天然存在的酮还原酶。

[0191] 在一些实施方案中,多核苷酸包含编码与本文所述的参考工程酮还原酶多肽具有至少约 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%或更高序列同一性的氨基酸序列的酮还原酶多肽的核苷酸序列,其中编码的酮还原酶多肽包含其中对应于 SEQ ID NO:2,4 或 98 的 X190 的残基不是酪氨酸的氨基酸序列。在一些实施方案中,多核苷酸编码包含其中对应于 X190 的残基是非芳族残基的氨基酸序列的酮还原酶多肽。在一些实施方案中,多核苷酸编码包含其中对应于 X190 的残基是丙氨酸、异亮氨酸、半胱氨酸或脯氨酸特别是脯氨酸的氨基酸序列的酮还原酶多肽。在一些实施方案中,多核苷酸编码包含选自 SEQ ID NO:6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列的工程酮还原酶多肽。

[0192] 在一些实施方案中,编码工程酮还原酶的多核苷酸选自 SEQ ID NO:5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39,41,43,45,47,49,51,53,55,57,59,61,63,65,67,69,71,73,75,77,79,81,83,85,87,89,91 和 93。在一些实施方案中,该多核苷酸能在高严谨条件下与包含 SEQ ID NO:5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39,41,43,45,47,49,51,53,55,57,59,61,63,65,67,69,71,73,75,77,79,81,83,85,87,89,91 和 93 的多核苷酸杂交,其中在高严谨条件下杂交的多核苷酸对取代苯乙酮底物具有 (S)-选择性,例如能将结构式 (I) 的底物还原或转化为结构式 (II) 的产物。在一些实施方案中,在高严谨条件下杂交的多核苷酸能将结构式 (III) 的底物还原或转化为结构式 (IV) 的产物。

[0193] 在一些实施方案中,多核苷酸编码本文所述的多肽,但在核苷酸水平与编码工程酮还原酶的参考多核苷酸具有约 80%或更高的序列同一性、约 85%,86%,87%,88%,89%,90%,91%,92%,93%,94%,95%,96%,97%,98%或 99%或更高的序列同一性。在一些实施方案中,参考多核苷酸选自对应于 SEQ ID NO:5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39,41,43,45,47,49,51,53,55,57,59,61,63,65,67,69,71,73,75,77,79,81,83,85,87,89,91 和 93 的多核苷酸序列。

[0194] 编码改进的酮还原酶多肽的分离的多核苷酸可以多种方式操作以提供多肽的表达。取决于表达载体,在插入载体之前对分离的多核苷酸的操作可以是需要的或必需的。利用重组 DNA 方法修饰多核苷酸和核酸序列的技术在本领域是公知的。指南提供于 Sambrook 等,2001, *Molecular Cloning: A Laboratory Manual* (分子克隆实验手册),第 3 版, Cold Spring Harbor Laboratory Press;和 *Current Protocols in Molecular Biology* (分子

生物学最新技术), Ausubel, F. 编辑, Greene Pub. Associates, 1998, 更新至 2006。

[0195] 对细菌宿主细胞, 用于指引本公开的核酸构建体的转录的适合启动子包括获自大肠杆菌 lac 操纵子、天蓝色链霉菌 (*Streptomyces coelicolor*) 琼脂糖酶基因 (dagA)、枯草芽孢杆菌 (*Bacillus subtilis*) 果聚糖蔗糖酶基因 (sacB)、地衣芽孢杆菌 (*Bacillus licheniformis*) α -淀粉酶基因 (amyL), 嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*) 麦芽糖淀粉酶基因 (amyM)、解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*) α -淀粉酶基因 (amyQ)、地衣芽孢杆菌青霉素酶基因 (penP)、枯草芽孢杆菌 (*Bacillus subtilis*) xylA 和 xylB 基因和原核 β -内酰胺酶基因的启动子 (Villa-Kamaroff 等, 1978, Proc. Natl Acad. Sci. USA 75 :3727-3731), 以及 tac 启动子 (DeBoer 等, 1983, Proc. Natl Acad. Sci. USA 80 :21-25)。其他启动子描述于 Sambrook 等, 上文。

[0196] 对丝状真菌宿主细胞, 用于指引本公开的核酸构建体的转录的适合启动子包括获自米曲霉 (*Aspergillus oryzae*) TAKA 淀粉酶、米黑根毛霉 (*Rhizomucor miehei*) 天冬氨酸蛋白酶、黑曲霉 (*Aspergillus niger*) 中性 α -淀粉酶、黑曲霉酸稳定型 α -淀粉酶、黑曲霉或泡盛曲霉 (*Aspergillus awamori*) 葡萄糖淀粉酶 (glaA)、米黑根毛霉脂肪酶、米曲霉碱性蛋白酶、米曲霉磷酸丙糖异构酶、构巢曲霉 (*Aspergillus nidulans*) 乙酰胺酶和尖孢镰刀菌 (*Fusarium oxysporum*) 胰酶样蛋白酶 (WO 96/00787) 基因的启动子, 以及 NA2-tpi 启动子 (来自黑曲霉中性 α -淀粉酶和米曲霉磷酸丙糖异构酶基因的启动子的杂合体) 及其突变、截短和杂合启动子。

[0197] 在酵母宿主中, 有用的启动子可来自酿酒酵母 (*Saccharomyces cerevisiae*) 烯醇化酶 (ENO-1)、酿酒酵母半乳糖激酶 (GAL1)、酿酒酵母醇脱氢酶 / 甘油醛 -3- 磷酸脱氢酶 (ADH2/GAP) 和酿酒酵母 3- 磷酸甘油激酶的基因。用于酵母宿主细胞的其他启动子描述于 Romanos 等, 1992, Yeast 8 :423-488。

[0198] 控制序列也可以是适合的转录终止子序列, 由宿主细胞识别来终止转录的序列。终止子序列与编码多肽的核酸序列的 3' 末端可操作地连接。在选择宿主细胞中有功能的任何终止子可用在本发明中。

[0199] 例如, 用于丝状真菌宿主细胞的示例性转录终止子可获自米曲霉 TAKA 淀粉酶、黑曲霉葡萄糖淀粉酶、构巢曲霉邻氨基苯甲酸合成酶、黑曲霉 α -葡萄糖苷酶和尖孢镰刀菌胰酶样蛋白酶的基因。

[0200] 用于酵母宿主细胞的示例性终止子可获自酿酒酵母烯醇化酶、酿酒酵母细胞色素 C (CYC1) 和酿酒酵母甘油醛 -3- 磷酸脱氢酶的基因。其他用于酵母宿主细胞的终止子描述于 Romanos 等, 1992, 上文。

[0201] 控制序列也可以是适合的前导序列, 前导序列是对宿主细胞的翻译重要的 mRNA 的非翻译区。前导序列与编码多肽的核酸序列的 5' 末端可操作地连接。可使用在选择宿主细胞中有功能的任何前导序列。用于丝状真菌宿主细胞的示例性前导序列获自米曲霉 TAKA 淀粉酶和构巢曲霉磷酸丙糖异构酶基因。适合于酵母宿主细胞的前导序列获自酿酒酵母烯醇化酶 (ENO-1)、酿酒酵母 3- 磷酸甘油激酶、酿酒酵母 α -因子和酿酒酵母醇脱氢酶 / 甘油醛 -3- 磷酸脱氢酶 (ADH2/GAP) 的基因。

[0202] 控制序列也可以是多聚腺苷化序列, 多聚腺苷化序列是与核酸序列的 3' 末端可

操作地连接并且当转录时由宿主细胞识别作为信号以便向转录的 mRNA 加入多腺苷残基的序列。在选择宿主细胞中有功能的任何多聚腺苷化序列可用在本发明中。用于丝状真菌宿主细胞的示例性多聚腺苷化序列可来自米曲霉 TAKA 淀粉酶、黑曲霉葡萄糖淀粉酶、构巢曲霉邻氨基苯甲酸合成酶、尖孢镰刀菌胰酶样蛋白酶和黑曲霉 α -葡萄糖苷酶的基因。用于酵母宿主细胞的多聚腺苷化序列描述于 Guo 和 Sherman, 1995, Mol CellBio 15: 5983-5990。

[0203] 控制序列还可以是编码与多肽的氨基末端连接的氨基酸序列并且指引编码的多肽进入细胞的分泌途径的信号肽编码区。核酸序列的编码序列的 5' 末端可内在地含有以翻译阅读框与编码分泌多肽的编码区的区段天然地连接的信号肽编码区。可选地, 编码序列的 5' 末端可含有编码序列外源的信号肽编码区。在编码序列不天然地含有信号肽编码区的情况中, 可能需要外源信号肽编码区。

[0204] 可选地, 外源信号肽编码区可简单地替换天然信号肽编码区, 以便增强多肽的分泌。然而, 指引表达的多肽进入选择的宿主细胞的分泌途径的任何信号肽编码区可用在本发明中。

[0205] 用于细菌宿主细胞的有效信号肽编码区是获自芽孢杆菌 (Bacillus) NC1B 11837 麦芽糖淀粉酶、嗜热脂肪芽孢杆菌 α -淀粉酶、地衣芽孢杆菌枯草杆菌蛋白酶、地衣芽孢杆菌 β -内酰胺酶、嗜热脂肪芽孢杆菌中性蛋白酶 (nprT, nprS, nprM) 和枯草芽孢杆菌 prsA 基因的信号肽编码区。其他信号肽 描述于 Simonen 和 Palva, 1993, Microbiol Rev 57: 109-137。

[0206] 用于丝状真菌宿主细胞的有效信号肽编码区可以是获自米曲霉 TAKA 淀粉酶、黑曲霉中性淀粉酶、黑曲霉葡萄糖淀粉酶、米黑根毛霉天冬氨酸蛋白酶、特异腐质霉 (Humicola insolens) 纤维素酶和柔毛腐质霉 (Humicolalanuginosa) 脂肪酶基因的信号肽编码区。

[0207] 用于酵母宿主细胞的信号肽可来自酿酒酵母 α -因子和酿酒酵母转化酶基因。其他有用的信号肽编码区描述于 Romanos 等, 1992, 上文。

[0208] 控制序列也可以是编码位于多肽氨基末端的氨基酸序列的前肽编码区。得到的多肽称为原酶 (proenzyme) 或多肽原 (或酶原, 在一些实例中)。一般而言, 多肽原是没有活性的, 并且可通过从多肽原催化或自催化切割前肽转化为成熟有活性的多肽。前肽编码区可获自枯草芽孢杆菌碱性蛋白酶 (aprE)、枯草芽孢杆菌中性蛋白酶 (nprT)、酿酒酵母 α -因子、米黑根毛霉天冬氨酸蛋白酶和嗜热毁丝霉 (Myceliophthora thermophila) 乳糖酶 (W095/33836) 基因。

[0209] 在信号肽和前肽区域都存在于多肽氨基末端的情况中, 前肽区域位于多肽氨基末端的相邻位置, 而信号肽区域位于前肽区域的氨基末端的相邻位置。

[0210] 可能还需要加入调节序列, 这允许相对于宿主细胞的生长来调节多肽的表达。调节系统的实例是响应于化学或物理刺激使基因表达启动或关闭的系统, 所述刺激包括调节化合物的存在。原核宿主细胞中, 适合的调节序列包括 lac、tac 和 trp 操纵子系统。酵母宿主细胞中, 作为实例, 适合的调节系统包括 ADH2 系统或 GAL1 系统。丝状真菌中, 适合的调节序列包括 TAKA α -淀粉酶启动子、黑曲霉葡萄糖淀粉酶启动子和米曲霉葡萄糖淀粉酶启动子。

[0211] 调节序列的其他实例是允许基因扩增的序列。真核系统中,这些序列包括在氨甲蝶呤存在下扩增的二氢叶酸还原酶基因和使用重金属扩增的金属硫蛋白基因。在这些实例中,编码本发明的 KRED 多肽的核酸序列将与调节序列可操作地连接。

[0212] 因此,在一些实施方案中,本公开还涉及重组表达载体,该重组表达载体包含编码工程酮还原酶多肽或其变体的多核苷酸,和一个或多个表达调节区域,取决于其将被引入的宿主类型,诸如启动子和终止子、复制起点等。上文所述的各种核酸和控制序列可连接到一起以产生可包括一个或多个方便的限制性酶切位点以便允许编码多肽的核酸序列在此位点插入或取代的重组表达载体。可选地,本公开的核酸序列可通过将核酸序列或包含该序列的核酸构建体插入用于表达的适当载体来表达。产生表达载体时,编码序列位于载体内,以便编码序列可操作地与用于表达的适当的控制序列连接。

[0213] 重组表达载体可以是任何载体(例如质粒或病毒),其可方便地经受重组 DNA 程序,并且可引起多核苷酸序列的表达。载体的选择一般取决于载体与该载体将被引入的宿主细胞的相容性。载体可以是线性质粒或闭合的环状质粒。

[0214] 表达载体可以是自主复制载体,即作为染色体外实体存在的载体,其复制不依赖于染色体的复制,例如,质粒、染色体外元件、微型染色体或人工染色体。载体可含有用于确保自我复制的任何部件(means)。可选地,载体可以是当引入宿主细胞时被整合到基因组中并与其所整合的染色体一起复制的载体。此外,可使用单个载体或质粒、或一起含有将引入宿主细胞基因组的总 DNA 的两个或多个载体或质粒、或转座子。

[0215] 本发明的表达载体优选地含有一个或多个可选择标记,这允许容易地检测转化细胞。可选择标记可以是其产物提供杀生物剂或病毒抗性、对重金属的抗性、营养缺陷型的原养型及类似性质的基因。细菌可选择标记的实例是来自枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因,或赋予抗生素抗性的标记,所述抗生素抗性诸如氨苄青霉素、卡那霉素、氯霉素(实施例 1)或四环素抗性。适合于酵母宿主细胞的标记是 ADE2, HIS3, LEU2, LYS2, MET3, TRP1 和 URA3。

[0216] 用在丝状真菌宿主细胞的可选择标记包括但不限于, *amdS*(乙酰胺酶)、*argB*(鸟氨酸氨甲酰基转移酶)、*bar*(草丁膦(phosphinothricin)乙酰转移酶)、*hph*(潮霉素磷酸转移酶)、*niaD*(硝酸还原酶)、*pyrG*(乳清苷-5'-磷酸脱羧酶)、*sC*(硫酸腺苷酰转移酶(sulfate adenyltransferase))和 *trpC*(邻氨基苯甲酸合成酶),以及其等效物。用在曲霉细胞中的实施方案包括构巢曲霉或米曲霉的 *amdS* 和 *pyrG* 基因和吸水链霉菌(*Streptomyces hygroscopicus*)的 *bar* 基因。

[0217] 本发明的表达载体可含有允许载体整合到宿主细胞的基因组或载体独立于基因组在细胞中自主复制的元件。对整合到宿主细胞的基因组,载体可依赖于编码多肽的核酸序列或载体的任何其他元件来通过同源或非同源重组将载体整合到基因组中。

[0218] 可选地,表达载体可含有用于指引通过同源重组整合到宿主细胞基因组的其他核酸序列。其他核酸序列使载体能被整合到宿主细胞基因组染色体的精确位置上。为了增加在精确位置整合的可能性,整合元件应优选地含有与对应的靶序列高度同源的足够数目的核酸,诸如 100 至 10,000 个碱基对,优选地 400 至 10,000 个碱基对,最优选地 800 至 10,000 个碱基对,以增强同源重组的概率。整合元件可以是与宿主细胞基因组内靶序列同源的任何序列。此外,整合元件可以是非编码或编码的核酸序列。另一方面,载体可通过非同源重

组被整合到宿主细胞的基因组中。

[0219] 对自主复制,载体还可包含使载体能在相关的宿主细胞中自主复制的复制起点。细菌复制起点的实例是 P15A ori 或允许在大肠杆菌中复制的质粒 pBR322、pUC19、pACYC177(该质粒具有 P15A ori) 或者 pACYC184 的复制起点,和允许在芽孢杆菌中复制的 pUB110、pE194、pTA1060 或 pAM β 1 的复制起点。用于在酵母宿主细胞中使用的复制起点的实例是 2 微米复制起点、ARS1、ARS4、ARS1 和 CEN3 的组合和 ARS4 和 CEN6 的组合。复制起点可以是具有使其在宿主细胞中温度敏感地起作用的突变的复制起点(参见,例如, Ehrlich,1978,Proc Natl Acad Sci USA75:1433)。

[0220] 本发明的核酸序列的不止一个拷贝可插入宿主细胞以便增加基因产物的产生。核酸序列拷贝数的增加可通过将序列的至少一个其他拷贝整合到宿主细胞基因组或通过随核酸序列中包括可扩增的选择标记基因来获得,其中含有可选择标记基因的扩增拷贝并由此含有核酸序列的另外拷贝的细胞可通过将细胞在适当的可选择物质存在下培养来选择。

[0221] 本公开中使用的许多表达载体是商业可获得的。适合的商业化表达载体包括来自 Sigma-Aldrich Chemicals, St. Louis MO. 的 p3xFLAGTM 表达载体,其包括用于在哺乳动物宿主细胞中表达的 CMV 启动子和 hGH 多聚腺苷化位点和 pBR322 复制起点和用于在大肠杆菌中扩增的氨苄青霉素抗性标记。其他适合的表达载体是可商业获自 Stratagene, LaJolla CA 的 pBluescriptII SK(-) 和 pBK-CMV,和衍生自 pBR322(Gibco BRL)、pUC(Gibco BRL)、pREP4、pCEP4(Invitrogen) 或 pPoly(Lathe 等,1987, Gene 57:193-201) 的质粒。

[0222] 6.4 用于表达酮还原酶多肽的宿主细胞

[0223] 在另一个方面,本公开提供了包含编码本公开的改进酮还原酶多肽的多核苷酸的宿主细胞,该多核苷酸与用于在宿主细胞中表达酮还原酶的一个或多个控制序列可操作地连接。用于表达由本发明的表达载体编码的 KRED 多肽的宿主细胞是本领域公知的,并且包括但不限于细菌细胞,诸如大肠杆菌、克菲尔乳杆菌、短乳杆菌、小乳杆菌、链霉菌(*Streptomyces*) 和鼠伤寒沙门氏菌(*Salmonella typhimurium*) 细胞;真菌细胞,诸如酵母细胞(例如酿酒酵母或巴氏毕赤酵母(*Pichia pastoris*)(ATCC 登录号 201178));昆虫细胞,诸如果蝇 S2 和灰翅夜蛾(*Spodoptera*) Sf9 细胞;动物细胞,诸如 CHO、COS、BHK、293 和 Bowes 黑色素瘤细胞;和植物细胞。用于上述宿主细胞的适当的培养基和生长条件是本领域公知的。

[0224] 用于表达酮还原酶的多核苷酸可通过本领域已知的各种方法引入细胞。技术包括但不限于电穿孔、生物射弹粒子轰击、脂质体介导的转染、氯化钙转染和原生质体融合。用于将多核苷酸引入细胞的各种方法对熟练的技术人员将是显而易见的。

[0225] 示例性宿主细胞是大肠杆菌 W3110。表达载体通过将编码改进的酮还原酶的多核苷酸可操作地连接到质粒 pCK110900 中以与在 lacI 阻抑物控制下的 lac 启动子可操作地连接来产生。表达载体还含有 P15a 复制起点和氯霉素抗性基因。大肠杆菌 W3110 中含有主题多核苷酸的细胞可通过使细胞经受氯霉素选择来分离。

[0226] 6.5 产生工程酮还原酶多肽的方法

[0227] 在一些实施方案中,为了制备本公开的改进 KRED 多核苷酸和多肽,催化还原反应的天然存在的酮还原酶获自(或衍生自)克菲尔乳杆菌或短乳杆菌或小乳杆菌。在一些实施方案中,母体多核苷酸序列被密码子优化以增强酮还原酶在特定宿主细胞中的表达。作

为例证,编码克菲尔乳杆菌的野生型 KRED 多肽的母体多核苷酸序列是从基于 Genbank 数据库中可获得的克菲尔乳杆菌 KRED 序列 (Genbank 登录号 AAP94029GI :33112056) 的已知多肽序列制备的寡核苷酸构建的。命名为 SEQ ID NO :3 的母体多核苷酸序列被密码子优化以在大肠杆菌中表达,并将密码子优化的多核苷酸克隆到表达载体中,将酮还原酶基因的表达置于 lac 启动子和 lacI 阻抑物基因的控制下。鉴定在大肠杆菌中表达活性酮还原酶的克隆,并测序该基因以确认其同一性。命名的序列 (SEQ ID NO :3) 是用作自克菲尔乳杆菌酮还原酶演化的工程酮还原酶的大多数实验和文库构建的起始点的母体序列。

[0228] 工程酮还原酶可通过使编码天然存在的酮还原酶的多核苷酸经受如上文讨论的诱变和 / 或定向演化方法来获得。示例性定向演化技术是诱变和 / 或 DNA 改组,如描述于 Stemmer,1994, Proc Natl Acad Sci USA 91:10747-10751;WO 95/22625;WO 97/0078;WO 97/35966;WO 98/27230;WO 00/42651;WO 01/75767 和美国专利 6,537,746。可使用的其他定向演化程序包括但不限于交错延伸过程 (staggered extension process, StEP)、体外重组 (Zhao 等,1998, Nat. Biotechnol. 16 :258-261)、诱变 PCR (Caldwell 等,1994, PCR Methods Appl. 3 :S136-S140) 和盒式诱变 (Black 等,1996, Proc Natl Acad Sci USA 93 :3525-3529)。用于本文目的的其他诱变和定向演化技术可发现于以下参考文献:Ling, 等, 1997, " Approaches to DNA Mutagenesis: an overview (DNA 诱变方法综述), " Anal. Biochem. 254 (2) :157-78;Dale 等,1996, " Oligonucleotide-directed random Mutagenesis using the phosphorothioate Method (使用硫代磷酸方法的寡核苷酸定向随机诱变), " Methods Mol. Biol. 57 :369-74;Smith, 1985, " In vitro Mutagenesis (体外诱变), " Ann. Rev. Genet. 19 :423-462;Botstein 等,1985, " Strategies and applications of in vitro Mutagenesis (体外诱变的策略和应用), " Science 229 :1193-1201;Carter, 1986, " Site-directed Mutagenesis (位点定向诱变), " Biochem. J. 237 :1-7;Kramer 等,1984, " Point Mismatch Repair (点错配修复), " Cell, 38 :879-887;Wells 等,1985, " Cassette Mutagenesis: an efficient Method for generation of multiple mutations at defined sites (盒式诱变:用于在特定位点产生多个突变的有效方法), " Gene 34 :315-323;Minshull 等,1999, " Protein evolution by molecular breeding (通过分子育种的蛋白演化), " Curr Opin Chem Biol 3 :284-290;Christians 等,1999, " Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling (使用 DNA 家族改组的用于 AZT 磷酸化的胸苷激酶的定向演化), " Nature Biotech 17 :259-264;Cramer 等,1998, " DNA shuffling of a family of genes from diverse species accelerates directed evolution (来自多样物种的基因家族的 DNA 改组加速定向演化), " Nature 391 :288-291;Cramer 等,1997, " Molecular evolution of an arsenate detoxification pathway by DNA shuffling (通过 DNA 改组的砷酸盐解毒途径的分子演化), " Nature Biotech 15 :436-438;Zhang 等, 1997, " Directed evolution of an effective fructosidase from a galactosidase by DNA shuffling and screening (通过 DNA 改组和筛选的有效果糖苷酶从半乳糖苷酶的定向演化), " Proc Natl Acad Sci USA 94 :4504-4509;Cramer 等,1996, " Improved green fluorescent protein by molecular evolution using DNA shuffling (使用 DNA 改组通过分子演化的改进的绿色荧光蛋白), " Nature Biotech 14 :315-319;和 Stemmer,

1994, " Rapidevolution of a protein in vitro by DNA shuffling(通过 DNA 改组的蛋白质的体外迅速演化), " Nature 370 :389-391。所有参考文献通过引用并入本文。

[0229] 筛选诱变处理后获得的克隆中具有需要的改进酶特性的工程酮还原酶。测量表达文库的酶活性可使用监测当 NADH 或 NADPH 被转化为 NAD^+ 或 NADP^+ 时 NADH 或 NADPH 浓度减少的速率(通过吸光度或荧光的降低)的标准生物化学技术来实现。(例如,参见实施例 7)。该反应中,当酮还原酶将酮底物还原为对应的羟基基团时,酮还原酶消耗(氧化)NADH 或 NADPH。如通过吸光度或荧光的降低测量的每单位时间 NADH 或 NADPH 浓度下降的速率指示固定量的裂解物(或从其制备的冻干粉末)中 KRED 多肽的相对(酶促)活性。产物的立体化学可通过各种已知技术确定,如实施例中提供。在需要的改进酶特性是热稳定性的情况中,酶活性可在酶制品经受特定温度以及测量热处理之后余下的酶促活性量之后测量。然后分离含有编码酮还原酶的多核苷酸的克隆,并测序以鉴定核苷酸序列的改变(如果有改变),并用来在宿主细胞中表达酶。

[0230] 在工程多肽的序列已知的情况中,编码酶的多核苷酸可根据已知的合成方法通过标准的固相方法制备。在一些实施方案中,可单独合成多达约 100 个碱基的片段,然后连接(例如通过酶促或化学连接(ligation)方法或聚合酶介导的方法)以形成任何需要的连续序列。例如,本发明的多核苷酸和寡核苷酸可通过化学合成来制备,使用例如描述于 Beaucage 等,1981, Tet Lett 22 :1859-69 的经典亚磷酰胺方法,或描述于 Matthes 等, 1984, EMBO J. 3 :801-05 的方法,例如其一般以自动合成方法实践。根据亚磷酰胺方法,寡核苷酸例如在自动 DNA 合成仪中合成、纯化、退火、连接并克隆到适当的载体。另外,基本上任何核酸可获自各种商业来源的任一种,诸如 The Midland Certified Reagent Company, Midland, TX ;The GreatAmerican Gene Company, Ramona, CA ;ExpressGen Inc.Chicago, IL ;Operon Technologies Inc., Alameda, CA 和许多其他公司。

[0231] 使用用于蛋白纯化的公知技术的任一种或多种,可从细胞和/或培养基中回收宿主细胞中表达的工程酮还原酶,所述技术包括但不限于溶菌酶处理、超声处理、过滤、盐析法、超速离心和层析。用于从细菌诸如大肠杆菌中裂解和高效提取蛋白的适合溶液是以 St.Louis MO 的 Sigma-Aldrich 的商标名 CelLytic B™ 可商业地获得的。

[0232] 用于分离酮还原酶多肽的层析技术包括但不限于反相层析、高效液相层析、离子交换层析、凝胶电泳和亲和层析。用于纯化特定酶的条件将部分取决于诸如净电荷、疏水性、亲水性、分子量、分子形状等因素,并且对熟悉本领域的技术人员将是显而易见的。

[0233] 在一些实施方案中,亲和技术可用来分离改进的酮还原酶。对亲和层析纯化,可使用特异结合酮还原酶多肽的任何抗体。为了产生抗体,包括但不限于兔、小鼠、大鼠等的各种宿主动物可通过用酮还原酶注射来免疫接种。通过侧链功能基团或附着于侧链功能基团的接头,酮还原酶多肽可附着于适合的载体,诸如 BSA。取决于宿主物种,可使用各种佐剂来增强免疫应答,包括但不限于弗氏(完全和不完全)、诸如氢氧化铝的矿物胶、诸如溶血卵磷脂的表面活性物质、普朗尼克多元醇、聚阴离子、肽、油乳剂、匙孔血蓝蛋白、二硝基苯酚和潜在有用的人类佐剂,诸如 BCG(卡介苗,bacilli Calmette Guerin)和短小棒状杆菌(Corynebacterium parvum)。

[0234] 酮还原酶可以表达酶的细胞的形式、作为粗提物或作为分离或纯化的制品被制备和使用。酮还原酶可作为粉末形式(例如丙酮粉末)的冻干产物制备或作为酶溶液制备。

在一些实施方案中,酮还原酶可以是以大体纯的制品的形式。

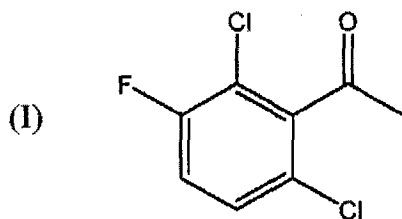
[0235] 在一些实施方案中,酮还原酶多肽可附着于固相基质。基质可以是固相、表面和/或膜。固体支持物可主要包含有机聚合体,诸如聚苯乙烯、聚乙烯、聚丙烯、聚氟乙烯、聚氧乙烯和聚丙烯酰胺,以及其共聚物及接枝物(graft)。固体支持物也可以是无机的,诸如玻璃、硅胶、可控孔度玻璃(controlled pore glass, CPG)、反相硅胶或金属,诸如金或铂。基质的结构可以是以珠、球、微粒(particle)、颗粒(granule)、凝胶、膜或表面的形式。表面可以是平面的、大体平面的或非平面的。固体支持物可以是多孔或无孔的,并且可具有膨胀或非膨胀特性。固体支持物可配置为孔、凹陷、或其他容器、导管、要素或位置的形式。多种支持物可配置为试剂的机器人递送可访问的或通过检测方法和/或仪器可访问的阵列上的各种位置。

[0236] 6.6 使用工程酮还原酶的方法和用该工程酮还原酶制备的化合物

[0237] 本文描述的酮还原酶能催化 2', 6' 取代苯乙酮底物,任选地在 3', 4' 或 5' 位置的一个或多个被取代,中的酮基团转化为对应的取代(S)-苯乙醇的还原反应。

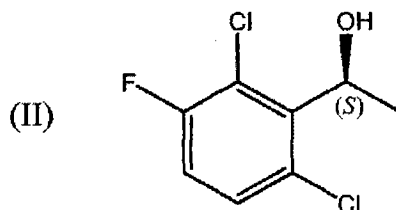
[0238] 在一些实施方案中,酮还原酶能将结构式(I)的底物化合物(2', 6' - 二氯 - 3' - 氟苯乙酮):

[0239]



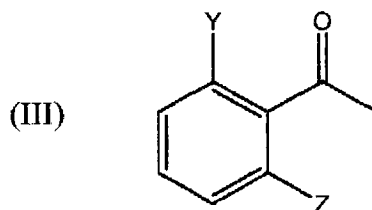
[0240] 还原或转化为结构式(II)的对应的手性醇产物(S)-1-[2,6-二氯-3-氟苯基]-乙醇:

[0241]



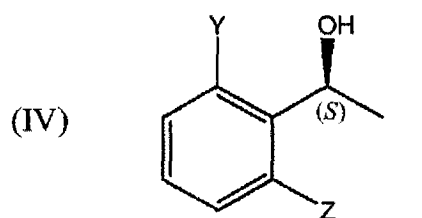
[0242] 在一些实施方案中,本文所述的酮还原酶能将结构式(III)的 2', 6' - 取代苯乙酮化合物:

[0243]



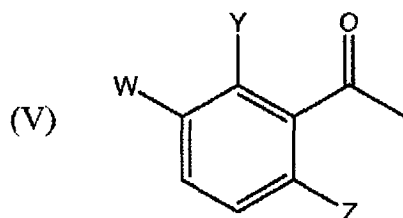
[0244] 其中 Y 和 Z 独立地选自 CH₃、CF₃、NH₂、OH、OCH₃、Cl 和 Br, 还原或转化为结构式(IV)的对应的手性醇产物:

[0245]



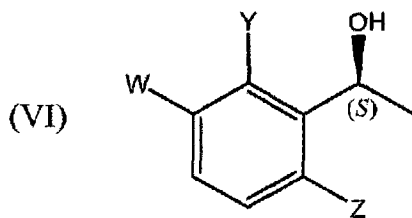
[0246] 在一些实施方案中,本文所述的酮还原酶能催化结构式 (III) 的 2',6' - 取代苯乙酮化合物,其可在 3',4' 和 5' 位置的一个或多个位置被相似地取代,转化为对应的 (S)- 醇产物的还原反应。本文所述的酮还原酶催化除了苯乙酮之外特定的其他取代的 2',6' - 取代苯乙酮化合物的还原反应的能力可通过常规实验法来确定,例如通过诸如实施例中所述的方法。结构式 (I) 的化合物 2',6' - 二氯 -3' - 氟苯乙酮是其他取代的 2',6' - 取代苯乙酮化合物的实例。因此,在一些实施方案中,本文公开的酮还原酶能催化结构式 (V) 的化合物:

[0247]



[0248] 其中 Y 和 Z 独立地选自 CH_3 、 CF_3 、 NH_2 、 OH 、 OCH_3 、 Cl 和 Br , W 选自 H 或 F、 Cl 或 Br , 转化为结构式 (VI) 的对应的 (S)- 醇产物的还原反应:

[0249]



[0250] 因此,在一些实施方案中,本文描述的酮还原酶可用在用于将 2',6' - 取代苯乙酮底物,任选地在 3',4' 或 5' 位置的一个或多个位置被取代,还原为对应的取代 (S)- 苯乙醇的方法中,其中该方法包括在适合于将取代苯乙酮还原或转化为对应的取代 (S)- 苯乙醇的反应条件下使取代苯乙酮底物与本文描述的酮还原酶接触。在该方法的一些实施方案中,底物以大于约 25%,50%,75%,80%,85%,90%,95%,99% 或 99.9% 的立体异构过量被还原为产物。

[0251] 在一些实施方案中,本文描述的酮还原酶可用在用于将式 (III) 的 2',6' - 取代苯乙酮底物还原为式 (IV) 的对应的取代 (S)- 苯乙醇化合物的方法中,其中该方法包括在适合于将式 (III) 的化合物还原或转化为式 (IV) 的对应的取代 (S)- 苯乙醇化合物的反应条件下使式 (III) 的化合物与本文所述的酮还原酶多肽接触或孵育。在该方法的一些实施方案中,底物以大于约 25%,50%,75%,80%,85%,90%,95%,99% 或 99.9% 的立体异构过量被还原为产物。

[0252] 在一些实施方案中,本文所述的酮还原酶可用在用于将式 (I) 的 2',6' - 二氯 -3' - 氟苯乙酮底物还原为其对应的 (S)- 醇产物,式 (II) 的 (S)-1-[2,6- 二氯 -3- 氟

苯基]-乙醇的方法中,其中该方法包括在适合于将 2',6'-二氯-3'-氟苯乙酮还原或转化为 (S)-1-[2,6-二氯-3-氟苯基]-乙醇的反应条件下使 2',6'-二氯-3'-氟苯乙酮与本文所述的酮还原酶多肽接触或孵育。在该方法的一些实施方案中,底物以大于约 85%,90%,95%,99%或 99.9%的立体异构过量被还原为产物。在一些实施方案中,底物以大于约 85%的立体异构过量被还原为产物,其中该酮还原酶多肽包含基于 SEQ ID NO:95,96 或 119 的序列式的氨基酸序列。

[0253] 在一些实施方案中,衍生自野生型乳杆菌属酮还原酶的工程 (S)-选择性酮还原酶可用在以大于约 85%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99%或 99.9%或更高的立体异构过量将苯乙酮还原为 (S)-1-苯乙醇的方法中。

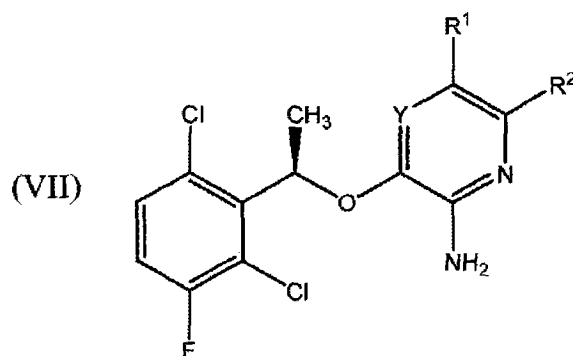
[0254] 在一些实施方案中,底物以大于约 99%的立体异构过量被还原为产物,其中该方法中使用的酮还原酶多肽包含选自 SEQ ID NO:6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列。在该方法的一些实施方案中,当方法使用相对 2',6'-二氯-3'-氟苯乙酮底物的量的按重量计低于约 1%的量的酮还原酶多肽来实行时,至少约 95%的底物在不到 24 小时内以大于约 99%的立体异构过量被还原为产物。

[0255] 在该方法的一些实施方案中,当方法使用至少约 200g/L 的底物和少于约 1g/L 的酮还原酶多肽来进行时,至少约 95%的底物在不到 24 小时内以至少约 99%的立体异构过量被还原为产物,其中该方法中使用的酮还原酶多肽包含选自 SEQ ID NO:18,32,34,36,38,40,42,44,46,74,76,78,80,82,84,86,88,90,92 和 94 的氨基酸序列。

[0256] 在一些实施方案中,本公开的酮还原酶多肽和方法可用来合成描述于以下参考文献中的蛋白酪氨酸激酶抑制化合物:W02006021886(氨基杂芳基化合物)、W02006021884(对映体大体纯的氨基杂芳基化合物)、W02006021881(吡唑-取代的氨基杂芳基化合物)和 W02004076412(氨基杂芳基化合物)),其合成依赖于式 (II) 的化合物作为中间体。所有参考文献通过引用整体并入本文。

[0257] 相应地,在一些实施方案中,本文所述的酮还原酶多肽和方法可用来产生结构式 (VII) 的蛋白酪氨酸激酶抑制化合物,

[0258]



[0259] 包括其盐、水合物及溶剂合物,如参考文献 W004076412 和 W006021884 中所述,其中 R^1 、 R^2 、Y 和 N 描述于本文。在一些实施方案中,对式 (VII) 的化合物,

[0260] Y 是 N 或 CR^{12} ;

[0261] R^1 选自氢、卤素、 C_{6-12} 芳基、5-12 元杂芳基、 C_{3-12} 环烷基、3-12 元杂脂环、 $-O(CR^6R^7)_nR^4$ 、 $-C(O)R^4$ 、 $-C(O)OR^4$ 、 $-CN$ 、 $-NO_2$ 、 $-S(O)_mR^4$ 、 $-SO_2NR^4R^5$ 、 $-C(O)NR^4R^5$ 、 $-NR^4C(O)R^5$ 、 $-C(=NR^6)$

NR^4R^5 、 C_{1-8} 烷基、 C_{2-8} 烯基和 C_{2-8} 炔基；并且 R^1 中的每个氢任选地由一个或多个 R^3 基团取代；
[0262] R^2 是氢、卤素、 C_{1-2} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-\text{S}(\text{O})_{\text{m}}\text{R}^4$ 、 $-\text{SO}_2\text{NR}^4\text{R}^5$ 、 $-\text{S}(\text{O})_2\text{OR}^4$ 、 $-\text{NO}_2$ 、 $-\text{NR}^4\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{OR}^4$ 、 $-\text{CN}$ 、 $-\text{C}(\text{O})\text{R}^4$ 、 $-\text{OC}(\text{O})\text{R}^4$ 、 $-\text{O}(\text{CR}^6\text{R}^7)_n\text{R}^4$ 、 $-\text{NR}^4\text{C}(\text{O})\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{C}(\text{O})\text{OR}^4$ 、 $-(\text{CR}^6\text{R}^7)_n\text{NCR}^4\text{R}^5$ 、 $-\text{C}(=\text{NR}^6)\text{NR}^4\text{R}^5$ 、 $-\text{NR}^4\text{C}(\text{O})\text{NR}^5\text{R}^6$ 、 $-\text{NR}^4\text{S}(\text{O})_{\text{p}}\text{R}^5$ 或 $-\text{C}(\text{O})\text{NR}^4\text{R}^5$ ，并且 R^2 中每个氢任选地由 R^8 取代；

[0263] 每个 R^3 独立地为卤素、 C_{1-2} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-\text{S}(\text{O})_{\text{m}}\text{R}^4$ 、 $-\text{SO}_2\text{NR}^4\text{R}^5$ 、 $-\text{S}(\text{O})_2\text{OR}^4$ 、 $-\text{NO}_2$ 、 $-\text{NR}^4\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{OR}^4$ 、 $-\text{CN}$ 、 $-\text{C}(\text{O})\text{R}^4$ 、 $-\text{OC}(\text{O})\text{R}^4$ 、 $-\text{O}(\text{CR}^6\text{R}^7)_n\text{R}^4$ 、 $-\text{NR}^4\text{C}(\text{O})\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{C}(\text{O})\text{OR}^4$ 、 $-(\text{CR}^6\text{R}^7)_n\text{OR}^4$ 、 $-(\text{CR}^6\text{R}^7)_n\text{C}(\text{O})\text{NR}^4\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{NCR}^4\text{R}^5$ 、 $-\text{C}(=\text{NR}^6)\text{NR}^4\text{R}^5$ 、 $-\text{NR}^4\text{C}(\text{O})\text{NR}^5\text{R}^6$ 、 $-\text{NR}^4\text{S}(\text{O})_{\text{p}}\text{R}^5$ 或 $-\text{C}(\text{O})\text{NR}^4\text{R}^5$ ， R^3 中每个氢任选地由 R^8 取代，并且邻近原子上的 R^3 基团可组合以形成 C_{6-12} 芳基、5-12 元杂芳基、 C_{3-12} 环烷基或 3-12 元杂脂环基团；

[0264] 每个 R^4 、 R^5 、 R^6 和 R^7 独立地为氢、卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基；或者与相同氮原子结合的 R^4 、 R^5 、 R^6 和 R^7 的任两个可与其结合的氮原子一起组合以形成 3 至 12 元杂脂环或 5-12 元杂芳基基团，该基团任选地含有 1 至 3 个选自 N、O 和 S 的其他杂原子；或者与相同碳原子结合的 R^4 、 R^5 、 R^6 和 R^7 的任两个可组合以形成 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环或 5-12 元杂芳基基团；并且 R^4 、 R^5 、 R^6 和 R^7 中每个氢任选地被 R^8 取代；

[0265] 每个 R^8 独立地为卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-\text{NH}_2$ 、 $-\text{CN}$ 、 $-\text{OH}$ 、 $-\text{O}-\text{C}_{1-12}$ 烷基、 $-\text{O}-(\text{CH}_2)_n\text{C}_{3-12}$ 环烷基、 $-\text{O}-(\text{CH}_2)_n\text{C}_{6-12}$ 芳基、 $-\text{O}-(\text{CH}_2)_n$ (3-12 元杂脂环) 或 $-\text{O}-(\text{CH}_2)_n$ (5-12 元杂芳基)；并且 R^8 中每个氢任选地被 R^{11} 取代；

[0266] 每个 R^9 和 R^{10} 独立地为氢、卤素、 C_{1-12} 烷基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-\text{S}(\text{O})_{\text{m}}\text{R}^4$ 、 $-\text{SO}_2\text{NR}^4\text{R}^5$ 、 $-\text{S}(\text{O})_2\text{OR}^4$ 、 $-\text{NO}_2$ 、 $-\text{NR}^4\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{OR}^4$ 、 $-\text{CN}$ 、 $-\text{C}(\text{O})\text{R}^4$ 、 $-\text{OC}(\text{O})\text{R}^4$ 、 $-\text{NR}^4\text{C}(\text{O})\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{C}(\text{O})\text{OR}^4$ 、 $-(\text{CR}^6\text{R}^7)_n\text{NCR}^4\text{R}^5$ 、 $-\text{NR}^4\text{C}(\text{O})\text{NR}^5\text{R}^6$ 、 $-\text{NR}^4\text{S}(\text{O})_{\text{p}}\text{R}^5$ 或 $-\text{C}(\text{O})\text{NR}^4\text{R}^5$ ； R^9 或 R^{10} 可与 A 的环原子或 A 的取代基组合以形成 C_{3-12} 环烷基、3-12 元杂脂环、 C_{6-12} 芳基或与 A 稠合的 5-12 元杂芳基；并且 R^9 和 R^{10} 中每个氢任选地被 R^3 取代；

[0267] 每个 R^{11} 独立地为卤素、 C_{1-12} 烷基、 C_{1-12} 烯氧基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-\text{O}-\text{C}_{1-12}$ 烷基、 $-\text{O}-(\text{CH}_2)_n\text{C}_{3-12}$ 环烷基、 $-\text{O}-(\text{CH}_2)_n\text{C}_{6-12}$ 芳基、 $-\text{O}-(\text{CH}_2)_n$ (3-12 元杂脂环)、 $-\text{O}-(\text{CH}_2)_n$ (5-12 元杂芳基) 或 $-\text{CN}$ ，并且 R^{11} 中每个氢任选地被卤素、 $-\text{OH}$ 、 $-\text{CN}$ 、可部分或完全卤素化的 $-\text{C}_{1-12}$ 烷基、可部分或完全卤素化的 $-\text{O}-\text{C}_{1-12}$ 烷基、 $-\text{CO}$ 、 $-\text{SO}$ 或 $-\text{SO}_2$ 取代；

[0268] R^{12} 是氢、卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-\text{S}(\text{O})_{\text{m}}\text{R}^4$ 、 $-\text{SO}_2\text{NR}^4\text{R}^5$ 、 $-\text{S}(\text{O})_2\text{OR}^4$ 、 $-\text{NO}_2$ 、 $-\text{NR}^4\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{OR}^4$ 、 $-\text{CN}$ 、 $-\text{C}(\text{O})\text{R}^4$ 、 $-\text{OC}(\text{O})\text{R}^4$ 、 $-\text{O}(\text{CR}^6\text{R}^7)_n\text{R}^4$ 、 $-\text{NR}^4\text{C}(\text{O})\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{C}(\text{O})\text{OR}^4$ 、 $-(\text{CR}^6\text{R}^7)_n\text{NCR}^4\text{R}^5$ 、 $-\text{C}(=\text{NR}^6)\text{NR}^4\text{R}^5$ 、 $-\text{NR}^4\text{C}(\text{O})\text{NR}^5\text{R}^6$ 、 $-\text{NR}^4\text{S}(\text{O})_{\text{p}}\text{R}^5$ 或 $-\text{C}(\text{O})\text{NR}^4\text{R}^5$ ，并且 R^{12} 中每个氢任选地被 R^3 取代；

[0269] 每个 R^{13} 独立地为卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-\text{S}(\text{O})_{\text{m}}\text{R}^4$ 、 $-\text{SO}_2\text{NR}^4\text{R}^5$ 、 $-\text{S}(\text{O})_2\text{OR}^4$ 、 $-\text{NO}_2$ 、 $-\text{NR}^4\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{OR}^4$ 、 $-\text{CN}$ 、 $-\text{C}(\text{O})\text{R}^4$ 、 $-\text{OC}(\text{O})\text{R}^4$ 、 $-\text{O}(\text{CR}^6\text{R}^7)_n\text{R}^4$ 、 $-\text{NR}^4\text{C}(\text{O})\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{C}(\text{O})\text{OR}^4$ 、 $-(\text{CR}^6\text{R}^7)_n\text{OR}^4$ 、 $-(\text{CR}^6\text{R}^7)_n\text{C}(\text{O})\text{NR}^4\text{R}^5$ 、 $-(\text{CR}^6\text{R}^7)_n\text{NCR}^4\text{R}^5$ 、 $-\text{C}(=\text{NR}^6)\text{NR}^4\text{R}^5$ 、 $-\text{NR}^4\text{C}(\text{O})\text{NR}^5\text{R}^6$ 、 $-\text{NR}^4\text{S}(\text{O})_{\text{p}}\text{R}^5$ 或 $-\text{C}(\text{O})\text{NR}^4\text{R}^5$ ，并且 R^{13} 中每个氢任选地被 R^3 取代；

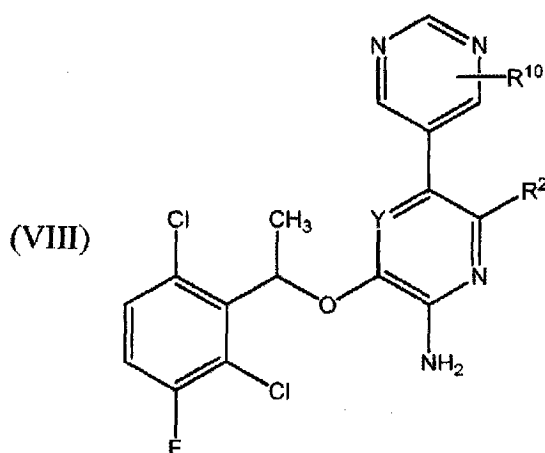
$_pR^5$ 、 $-C(O)NR^4R^5$ 、 $-(CR^6R^7)_n$ (3-12 元杂脂环)、 $-(CR^6R^7)_n$ (C_{3-12} 环烷基)、 $-(CR^6R^7)_n$ (C_{6-12} 芳基)、 $-(CR^6R^7)_n$ (5-12 元杂芳基)、 $-(CR^6R^7)_nC(O)NR^4R^5$ 或 $-(CR^6R^7)_nC(O)R^4$, 邻近原子上的 R^{13} 基团可组合以形成 C_{6-12} 芳基、5-12 元杂芳基、 C_{3-12} 环烷基或 3-12 元杂脂环基团, 并且 R^{13} 中每个氢任选地被 R^3 取代;

[0270] 其中, 每个 m 独立地为 0、1 或 2; 每个 n 独立地为 0、1、2、3 或 4; 并且每个 p 独立地为 1 或 2。各种取代基以及式 (VII) 包含的特定化合物的说明描述于 W004076412 和 W006021884。

[0271] 相应地, 在用于产生结构式 (VII) 的对映体纯的化合物的方法中, 方法的步骤可包括在适合于将式 (I) 的底物化合物还原或转化为式 (II) 的产物化合物的反应条件下使用本文所述的酮还原酶多肽将式 (I) 的化合物还原或转化为式 (II) 的化合物。从式 (II) 的化合物合成式 (VII) 的化合物描述于引用的参考文献中。

[0272] 在一些实施方案中, 本文所述的酮还原酶多肽和方法可用来产生结构式 (VIII) 的蛋白酪氨酸激酶抑制化合物,

[0273]



[0274] 包括其盐、水合物及溶剂合物, 如 W006021886 中所述, 其中 R^{10} 、 R^2 、Y 和 N 描述于本文。在一些实施方案中, 对式 (VIII) 的化合物,

[0275] Y 是 N 或 CR^1 ;

[0276] R^1 是氢、卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-S(O)_mR^4$ 、 $-SO_2NR^4R^5$ 、 $-S(O)_2OR^4$ 、 $-NO_2$ 、 $-NR^4R^5$ 、 $-(CR^6R^7)_nOR^4$ 、 $-CN$ 、 $-C(O)R^4$ 、 $-OC(O)R^4$ 、 $-O(CR^6R^7)_nR^4$ 、 $-NR^4C(O)R^5$ 、 $-(CR^6R^7)_nC(O)OR^4$ 、 $-(CR^6R^7)_nNCR^4R^5$ 、 $-C(=NR^6)NR^4R^5$ 、 $-NR^4C(O)NR^5R^6$ 、 $-NR^4S(O)_pR^5$ 或 $-C(O)NR^4R^5$, 并且 R^1 中每个氢任选地被 R^3 取代;

[0277] R^2 是氢、卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-S(O)_mR^4$ 、 $-SO_2NR^4R^5$ 、 $-S(O)_2OR^4$ 、 $-NO_2$ 、 $-NR^4R^5$ 、 $-(CR^6R^7)_nOR^4$ 、 $-CN$ 、 $-C(O)R^4$ 、 $-OC(O)R^4$ 、 $-O(CR^6R^7)_nR^4$ 、 $-NR^4C(O)R^5$ 、 $-(CR^6R^7)_nC(O)OR^4$ 、 $-(CR^6R^7)_nNCR^4R^5$ 、 $-C(=NR^6)NR^4R^5$ 、 $-NR^4C(O)NR^5R^6$ 、 $-NR^4S(O)_pR^5$ 或 $-C(O)NR^4R^5$, 并且 R^2 中每个氢任选地被 R^8 取代;

[0278] 每个 R^3 独立地为卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-S(O)_mR^4$ 、 $-SO_2NR^4R^5$ 、 $-S(O)_2OR^4$ 、 $-NO_2$ 、 $-NR^4R^5$ 、 $-(CR^6R^7)_nOR^4$ 、 $-CN$ 、 $-C(O)R^4$ 、 $-OC(O)R^4$ 、 $-O(CR^6R^7)_nR^4$ 、 $-NR^4C(O)R^5$ 、 $-(CR^6R^7)_nC(O)OR^4$ 、 $-(CR^6R^7)_nNCR^4R^5$ 、 $-C(=NR^6)NR^4R^5$ 、 $-NR^4C(O)NR^5R^6$ 、 $-NR^4S(O)_pR^5$ 或 $-C(O)NR^4R^5$, R^3 中每个氢任选地被 R^8 取代, 邻近原子上的 R^3 基团可组合以形成 C_{6-12} 芳基、

5-12 元杂芳基、 C_{3-12} 环烷基或 3-12 元杂脂环基团；

[0279] 每个 R^4 、 R^5 、 R^6 和 R^7 独立地为氢、卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基；或者与相同氮原子结合的 R^4 、 R^5 、 R^6 和 R^7 的任两个可与其结合的氮原子一起组合以形成 3 至 12 元杂脂环或 5-12 元杂芳基基团，该基团任选地含有 1 至 3 个选自 N、O 和 S 的其他杂原子；或者与相同碳原子结合的 R^4 、 R^5 、 R^6 和 R^7 的任两个可组合以形成 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环或 5-12 元杂芳基基团；并且 R^4 、 R^5 、 R^6 和 R^7 中每个氢任选地被 R^8 取代；

[0280] 每个 R^8 独立地为卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-NH_2$ 、 $-CN$ 、 $-OH$ 、 $-O-C_{1-12}$ 烷基、 $-O-(CH_2)_n C_{3-12}$ 环烷基、 $-O-(CH_2)_n C_{6-12}$ 芳基、 $-O-(CH_2)_n$ (3-12 元杂脂环) 或 $-O-(CH_2)_n$ (5-12 元杂芳基)；并且 R^8 中每个氢任选地被 R^9 取代；

[0281] 每个 R^9 独立地为卤素、 C_{1-12} 烷基、 C_{1-12} 烷氧基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-O-C_{1-12}$ 烷基、 $-O-(CH_2)_n C_{3-12}$ 环烷基、 $-O-(CH_2)_n C_{6-12}$ 芳基、 $-O-(CH_2)_n$ (3-12 元杂脂环)、 $-O-(CH_2)_n$ (5-12 元杂芳基) 或 $-CN$ ，并且 R^9 中每个氢任选地被卤素、 $-OH$ 、 $-CN$ 、可部分或完全卤素化的 $-C_{1-12}$ 烷基、可部分或完全卤素化的 $-O-C_{1-12}$ 烷基、 $-CO$ 、 $-SO$ 或 $-SO_2$ 取代；

[0282] R^{10} 代表一个、两个或三个任选的取代基，其独立地为卤素、 C_{1-12} 烷基、 C_{2-12} 烯基、 C_{2-12} 炔基、 C_{3-12} 环烷基、 C_{6-12} 芳基、3-12 元杂脂环、5-12 元杂芳基、 $-S(O)_m R^4$ 、 $-SO_2 NR^4 R^5$ 、 $-S(O)_2 OR^4$ 、 $-NO_2$ 、 $-NR^4 R^5$ 、 $-(CR^6 R^7)_n OR^4$ 、 $-CN$ 、 $-C(O) R^4$ 、 $-OC(O) R^4$ 、 $-O(CR^6 R^7)_n R^4$ 、 $-NR^4 C(O) R^5$ 、 $-(CR^6 R^7)_n C(O) OR^4$ 、 $-(CR^6 R^7)_n OR^4$ 、 $-(CR^6 R^7)_n C(O) NR^4 R^5$ 、 $-(CR^6 R^7)_n NCR^4 R^5$ 、 $-C(=NR^6) NR^4 R^5$ 、 $-NR^4 C(O) NR^5 R^6$ 、 $-NR^4 S(O)_p R^5$ 、 $-C(O) NR^4 R^5$ 、 $-(CR^6 R^7)_n$ (3-12 元杂脂环)、 $-(CR^6 R^7)_n$ (C_{3-12} 环烷基)、 $-(CR^6 R^7)_n$ (C_{6-12} 芳基)、 $-(CR^6 R^7)_n$ (5-12 元杂芳基) 或 $-(CR^6 R^7)_n C(O) NR^4 R^5$ ，并且 R^{10} 中每个氢任选地被 R^3 取代；

[0283] 其中每个 m 独立地为 0、1 或 2；每个 n 独立地为 0、1、2、3 或 4；并且每个 p 独立地为 1 或 2。各种取代基以及式 (VIII) 包含的特定化合物的说明描述于 WO2006021886。

[0284] 相应地，在用于产生结构式 (VIII) 的化合物的方法中，方法的步骤可包括在适合于将式 (I) 的底物化合物还原或转化为式 (II) 的产物化合物的反应条件下使用本文所述的酮还原酶多肽将式 (I) 的化合物还原或转化为式 (II) 的化合物。从式 (II) 的化合物合成式 (VIII) 的化合物描述于引用的参考文献中。

[0285] 如本领域技术人员所知，酮还原酶催化的还原反应一般需要辅因子。本文所述的工程酮还原酶催化的还原反应一般也需要辅因子，尽管工程酮还原酶的许多实施方案需要远比野生型酮还原酶催化的反应更少的辅因子。如本文所用，术语“辅因子”指与酮还原酶组合作用的非蛋白化合物。适合与本文描述的工程酮还原酶一起使用的辅因子包括但不限于 $NADP^+$ (烟酰胺腺嘌呤二核苷酸磷酸)、 $NADPH$ ($NADP^+$ 的还原形式)、 NAD^+ (烟酰胺腺嘌呤二核苷酸) 和 $NADH$ (NAD^+ 的还原形式)。一般而言，将还原形式的辅因子加到反应混合物中。可任选地使用辅因子再生系统从氧化型 $NAD(P)^+$ 形式再生还原型 $NAD(P)H$ 形式。

[0286] 术语“辅因子再生系统”指参与还原氧化型辅因子 (例如 $NADP^+$ 到 $NADPH$) 的反应的一套反应物。酮还原酶催化的酮底物的还原氧化的辅因子通过辅因子再生系统以还原形式被再生。辅因子再生系统包含为还原氢等效物来源并且能还原氧化型辅因子的化学计

量的还原剂。辅因子再生系统还可包含催化剂,例如催化通过还原剂对氧化型辅因子的还原的酶催化剂。分别从 NAD^+ 或 NADP^+ 再生 NADH 或 NADPH 的辅因子再生系统在本领域是已知的,并且可用在本文所述的方法中。

[0287] 可采用的适合的示例性辅因子再生系统包括但不限于葡萄糖和葡萄糖脱氢酶、甲酸和甲酸脱氢酶、葡萄糖-6-磷酸和葡萄糖-6-磷酸脱氢酶、仲(例如异丙醇)醇和仲醇脱氢酶、亚磷酸和亚磷酸脱氢酶、分子氢和氢化酶和类似系统。这些系统可与作为辅因子的 $\text{NADP}^+/\text{NADPH}$ 或 NAD^+/NADH 组合使用。使用氢化酶的电化学再生也可用作辅因子再生系统。参见例如,美国专利第 5,538,867 和 6,495,023 号,两者都通过引用并入本文。包含金属催化剂和还原剂(例如分子氢或甲酸)的化学辅因子再生系统也是适合的。参见例如 PCT 公布 WO 2000/053731,其通过引用并入本文。

[0288] 术语“葡萄糖脱氢酶”和“GDH”在本文可交换使用,指分别催化 D-葡萄糖和 NAD^+ 或 NADP^+ 转化为葡萄糖酸和 NADH 或 NADPH 的依赖于 NAD^+ 或 NADP^+ 的酶。以下的反应式(1)描述了葡萄糖脱氢酶-催化的通过葡萄糖的 NAD^+ 或 NADP^+ 的还原。

[0289]



[0290] 适合用于本文所述的方法的实践的葡萄糖脱氢酶包括天然存在的葡萄糖脱氢酶以及非天然存在的葡萄糖脱氢酶。天然存在的葡萄糖脱氢酶编码基因已报道于文献资料中。例如,枯草芽孢杆菌 61297GDH 基因在大肠杆菌中表达,并报道为显示与其天然宿主产生的酶相同的物理化学特性(Vasantha 等,1983, Proc. Natl. Acad. Sci. USA 80:785)。对应于 Genbank 登录号 M12276 的枯草芽孢杆菌 GDH 基因的基因序列由 Lampel 等,1986, J. Bacteriol. 166:238-243 报道,并且 Yamane 等,1996, Microbiology 142:3047-3056 报道其修正形式为 Genbank 登录号 D50453。天然存在的 GDH 基因还包括编码来自以下的 GDH 的那些基因:蜡状芽孢杆菌 (*B. cereus*) ATCC 14579 (Nature, 2003, 423:87-91; Genbank 登录号 AE017013) 和巨大芽孢杆菌 (*B. megaterium*) (Eur. J. Biochem., 1988, 174:485-490, Genbank 登录号 X12370; J. Ferment. Bioeng, 1990, 70:363-369, Genbank 登录号 GI216270)。来自芽孢杆菌 (*Bacillus* 甲.) 的葡萄糖脱氢酶在 PCT 公布 WO 2005/018579 中提供为 SEQ ID NO:10 和 12(分别由对应于 PCT 公布的 SEQ ID NO:9 和 11 的多核苷酸序列编码),其公开内容通过引用并入本文。

[0291] 非天然存在的葡萄糖脱氢酶可使用已知的方法产生,诸如,例如,诱变、定向演化 and 类似方法。无论是天然存在或非天然存在的酶,具有适合的活性的 GDH 酶可使用 PCT 公布 WO 2005/018579 的实施例 4 中所述的测定容易地鉴定,其公开内容通过引用并入本文。示例性的非天然存在的葡萄糖脱氢酶在 PCT 公布 WO 2005/018579 中提供为 SEQ ID NO:62, 64, 66, 68, 122, 124 和 126。编码这些酶的多核苷酸序列在 PCT 公布 WO2005/018579 中分别提供为 SEQ ID NO:61, 63, 65, 67, 121, 123 和 125。所有这些序列通过引用并入本文。适合用在本文公开的酮还原酶-催化的还原反应中的其他非天然存在的葡萄糖脱氢酶提供于美国申请公布第 2005/0095619 和 2005/0153417 号,其公开内容通过引用并入本文。

[0292] 本文所述的酮还原酶-催化的还原反应中采用的葡萄糖脱氢酶在 PCT 公布 WO 2005/018579 的实施例 4 所述的测定中可表现出至少约 $10 \mu\text{mol}/\text{min}/\text{mg}$ 的活性,有时至少约 $10^2 \mu\text{mol}/\text{min}/\text{mg}$ 或约 $10^3 \mu\text{mol}/\text{min}/\text{mg}$ 、多达约 $10^4 \mu\text{mol}/\text{min}/\text{mg}$ 或更高的活性。

[0293] 本文所述的酮还原酶 – 催化的还原反应一般在溶剂中进行。适合的溶剂包括水、有机溶剂（例如乙酸乙酯、乙酸丁酯、1- 辛醇、庚烷、辛烷、甲基叔丁醚 (MTBE)、甲苯和类似溶剂）、离子液体（例如，1- 乙基 4- 甲基咪唑鎓四氟硼酸、1- 丁基 -3- 甲基咪唑鎓四氟硼酸、1- 丁基 -3- 甲基咪唑六氟磷酸和类似离子液体）。在一些实施方案中，使用包括水和水性共溶剂系统的水性溶剂 (aqueous solvent)。

[0294] 示例性水性共溶剂系统具有水和一种或多种有机溶剂。一般而言，选择水性共溶剂系统的有机溶剂组分，以便其不完全失活酮还原酶。适当的共溶剂系统可利用诸如描述于本文的酶活性测定、通过在候选溶剂系统中用感兴趣的特定底物来测量特定工程酮还原酶的酶促活性来容易地鉴定。

[0295] 水性共溶剂系统的有机溶剂组分可与水性组分混溶，提供了单一液相，或者可与水性组分部分混溶或不混溶，提供两种液相。一般而言，当采用水性共溶剂系统时，选择为双相的，其中水分散于有机溶剂中，或反之亦然。一般而言，当使用水性共溶剂系统时，需要选择可容易地与水性分离的有机溶剂。一般而言，共溶剂系统中水和有机溶剂的比率一般在从约 90 : 10 至约 10 : 90 (v/v) 有机溶剂比水、介于 80 : 20 和 20 : 80 (v/v) 之间有机溶剂比水的范围内。共溶剂系统可在添加到反应混合物之前预先形成，或者其可在反应容器中原位形成。

[0296] 水性溶剂（水或水性共溶剂系统）可以是 pH- 缓冲或不缓冲的。一般而言，还原可在 pH 约 10 或更低、通常在从约 5 至约 10 的范围内进行。在一些实施方案中，还原在 pH 约 9 或更低、通常在从约 5 至约 9 的范围内进行。在一些实施方案中，还原在 pH 约 8 或更低、通常在从约 5 至约 8 的范围内和通常在从约 6 至约 8 的范围内进行。还原还可在 pH 约 7.8 或更低、或 7.5 或更低下进行。可选地，还原可在中性 pH (即约 7) 下进行。

[0297] 在还原反应过程中，反应混合物的 pH 可以改变。通过在反应期间加入酸或碱，反应混合物的 pH 可维持在需要的 pH 下或在需要的 pH 范围内。可选地，可通过使用包含缓冲剂的水性溶剂来控制 pH。保持需要的 pH 范围的适合缓冲液在本领域是已知的，并且包括例如，磷酸缓冲液、三乙醇胺缓冲液和类似缓冲液。还可使用缓冲液和加入酸或碱的组合。

[0298] 当采用葡萄糖 / 葡萄糖脱氢酶的辅因子再生系统时，如果得到的水性葡萄糖酸不被另外中和，如反应式 (1) 中代表的葡萄糖酸 ($pK_a = 3.6$) 的共同产生导致反应混合物 pH 的下降。通过标准的缓冲技术或在转化过程的同时加入碱，反应混合物的 pH 可保持在需要的水平，在所述标准的缓冲技术中缓冲液中和葡萄糖酸可达提供的缓冲能力。还可使用缓冲和加入碱的组合。保持需要的 pH 范围的适合缓冲液描述于上文。用于中和葡萄糖酸的适合碱为有机碱（例如胺、醇盐和类似有机碱）和无机碱例如氢氧化物盐（例如 NaOH）、碳酸盐（例如 $NaHCO_3$ ）、碳酸氢盐（例如 K_2CO_3 ）、碱性磷酸盐（例如 K_2HPO_4 、 Na_3PO_4 ）和类似无机碱。转化过程的同时加入碱可在监测反应混合物的 pH 的同时人工地进行，或者更方便地，通过使用自动滴定仪作为 pH 稳态 (pH stat)。部分缓冲能力和加入碱的组合还可用于过程控制。

[0299] 当采用加入碱来中和酮还原酶 – 催化的还原反应期间释放的葡萄糖酸时，转化过程可通过保持 pH 所加入的碱的量来监测。在还原期间加入到未缓冲或部分缓冲的反应混合物中的碱一般以水性溶液加入。

[0300] 在一些实施方案中，辅因子再生系统可包含甲酸脱氢酶。术语“甲酸脱氢酶”和

“FDH”在本文可交换使用,指分别催化甲酸和 NAD^+ 或 NADP^+ 转化为二氧化碳和 NADH 或 NADPH 的依赖于 NAD^+ 或 NADP^+ 的酶。适合用作本文所述的酮还原酶 - 催化的还原反应中的辅因子再生系统的甲酸脱氢酶包括天然存在的甲酸脱氢酶以及非天然存在的甲酸脱氢酶。甲酸脱氢酶包括对应于 PCT 公布 WO 2005/018579 的 SEQ ID NO:70 (假单胞菌种 (*Pseudomonas* sp.)) 和 72 (博伊丁假丝酵母 (*Candida boidinii*)) 的甲酸脱氢酶,其分别由对应于 PCT 公布 2005/018579 的 SEQ ID NO:69 和 71 的多核苷酸序列编码,该 PCT 公布的公开内容通过引用并入本文。无论是天然存在或非天然存在的,本文所述的方法中采用的甲酸脱氢酶可表现出至少约 $1 \mu\text{mol}/\text{min}/\text{mg}$ 、有时至少约 $10 \mu\text{mol}/\text{min}/\text{mg}$ 或至少约 $10^2 \mu\text{mol}/\text{min}/\text{mg}$ 、多达约 $10^3 \mu\text{mol}/\text{min}/\text{mg}$ 或更高的活性,并且在 PCT 公布 WO 2005/018579 的实施例 4 所述的测定中可容易地筛选其活性。

[0301] 如本文所用,术语“甲酸”(formate)指甲酸阴离子 (HCO_2^-)、甲酸 (HCO_2H) 及其混合物。甲酸可以盐的形式提供,一般为碱金属盐或铵盐 (例如, HCO_2Na 、 KHCO_2NH_4 和类似盐),以甲酸的形式提供,一般为水性甲酸或其混合物。甲酸是中强酸。在其 pK_a (于水中 $\text{pK}_a = 3.7$) 的几个 pH 单位内的水溶液中,甲酸作为平衡浓度的 HCO_2^- 和 HCO_2H 存在。在 pH 值高于约 pH 4 时,甲酸主要作为 HCO_2^- 存在。当甲酸作为甲酸提供时,反应混合物一般被缓冲或者通过加入碱变得酸性更弱以提供需要的 pH,一般为约 pH 5 或更高。用于中和甲酸的适合的碱包括但不限于有机碱例如胺、醇盐和类似有机碱,和无机碱例如氢氧化物盐 (例如 NaOH)、碳酸盐 (例如 NaHCO_3)、碳酸氢盐 (例如 K_2CO_3)、碱性磷酸盐 (例如 K_2HPO_4 、 Na_3PO_4) 和类似无机碱。

[0302] 对高于约 pH 5 的 pH 值,其中甲酸主要作为 HCO_2^- 存在,下文反应式 (2) 描述了甲酸脱氢酶 - 催化的甲酸对 NAD^+ 或 NADP^+ 的还原。

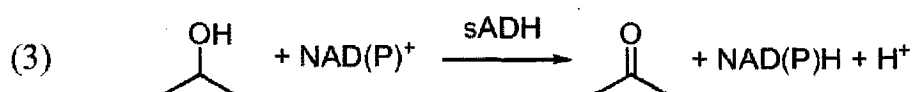


[0304] 当采用甲酸和甲酸脱氢酶作为辅因子再生系统时,可通过标准的缓冲技术或者通过在转化过程的同时加入酸将反应混合物的 pH 保持在需要的水平,在所述标准的缓冲技术中缓冲液释放质子以便达到提供的缓冲能力。在反应过程中加入以保持 pH 的适合的酸包括有机酸例如羧酸、磺酸、磷酸和类似有机酸,无机酸例如氢卤酸 (诸如盐酸)、硫酸、磷酸和类似无机酸)、酸式盐 (例如二氢磷酸盐 (例如 KH_2PO_4)、亚硫酸盐 (例如 NaHSO_4) 和类似酸式盐。一些实施方案利用甲酸,其中保持甲酸的浓度和溶液的 pH。

[0305] 在使用甲酸 / 甲酸脱氢酶辅因子再生系统的还原反应期间,当采用加入酸来保持 pH 时,转化的过程可通过保持 pH 所加入的酸的量来监测。一般而言,在转化过程中加入到未缓冲或部分缓冲的反应混合物中的酸以水性溶液加入。

[0306] 术语“仲醇脱氢酶”和“sADH”在本文可交换使用,指分别催化仲醇和 NAD^+ 或 NADP^+ 转化为酮和 NADH 或 NADPH 的依赖于 NAD^+ 或 NADP^+ 的酶。下文的反应式 (3) 描述了由异丙醇示例的仲醇对 NAD^+ 或 NADP^+ 的还原。

[0307]



[0308] 适合在本文所述的酮还原酶 - 催化的还原反应中用作辅因子再生系统的仲醇

脱氢酶包括天然存在的仲醇脱氢酶以及非天然存在的仲醇脱氢酶。天然存在的仲醇脱氢酶包括来自布氏嗜热厌氧菌 (*Thermoanaerobium brockii*)、红平红球菌 (*Rhodococcus erythropolis*)、克菲尔乳杆菌和短乳杆菌的已知醇脱氢酶,非天然存在的仲醇脱氢酶包括从其衍生的工程醇脱氢酶。无论是天然存在或非天然存在的,本文所述的方法中采用的仲醇脱氢酶可表现出至少约 $1 \mu\text{mol}/\text{min}/\text{mg}$ 、有时至少约 $10 \mu\text{mol}/\text{min}/\text{mg}$ 或至少约 $10^2 \mu\text{mol}/\text{min}/\text{mg}$ 、多达约 $10^3 \mu\text{mol}/\text{min}/\text{mg}$ 或更高的活性。

[0309] 适合的仲醇包括低级仲烷醇 (alkanol) 和芳基-羟基甲醇。低级仲醇的实例包括异丙醇、2-丁醇、3-甲基-2-丁醇、2-戊醇、3-戊醇、3,3-二甲基-2-丁醇和类似低级仲醇。在一个实施方案中,仲醇是异丙醇。适合的芳基-羟基甲醇包括未取代和取代的 1-芳基乙醇。

[0310] 当采用仲醇和仲醇脱氢酶作为辅因子再生系统时,得到的 NAD^+ 或 NADP^+ 通过仲醇脱氢酶将仲醇偶联氧化为酮而被还原。一些工程酮还原酶还具有使仲醇还原剂脱氢的活性。在使用仲醇作为还原剂的一些实施方案中,工程酮还原酶和仲醇脱氢酶是相同的酶。

[0311] 在采用辅因子再生系统执行本文所述的酮还原酶-催化的还原反应的实施方案中,辅因子的氧化或还原形式可最初提供。如上文所述,辅因子再生系统将氧化型辅因子转化为其还原型,然后还原型辅因子在酮还原酶底物的还原中被利用。

[0312] 在一些实施方案中,不使用辅因子再生系统。对不使用辅因子再生系统执行的还原反应,将辅因子以还原型加入到反应混合物中。

[0313] 在一些实施方案中,当使用宿主有机体的全细胞执行该过程时,全细胞可天然地提供辅因子。可选地或组合地,细胞可天然地或重组地提供葡萄糖脱氢酶。

[0314] 在执行本文所述的立体选择性还原反应中,工程酮还原酶和包含任选辅因子再生系统的任何酶可以纯化的酶、用编码酶的基因转化的全细胞和 / 或细胞提取物和 / 或此类细胞的裂解物的形式加入到反应混合物中。编码工程酮还原酶和任选地辅因子再生酶的基因可被单独转化到宿主细胞,或者一起转化到相同的宿主细胞中。例如,在一些实施方案中,一套宿主细胞可用编码工程酮还原酶的基因转化,另一套宿主细胞可用编码辅因子再生酶的基因转化。两套转化细胞可以全细胞的形式或以衍生自全细胞的裂解物或提取物的形式在反应混合物中一起使用。在其他实施方案中,宿主细胞可用编码工程酮还原酶和辅因子再生酶的基因转化。

[0315] 用编码工程酮还原酶和 / 或任选的辅因子再生酶的基因转化的全细胞或其细胞提取物和 / 或其裂解物可以多种不同形式采用,包括固体 (例如冻干的固体、喷雾干燥的固体和类似固体) 或半固体 (例如粗糊剂)。

[0316] 细胞提取物或细胞裂解物可通过沉淀 (硫酸铵、聚乙烯亚胺、热处理或类似处理),然后在冻干之前进行脱盐程序 (例如超滤、透析和类似过程) 被部分纯化。任何细胞制品可通过使用已知交联剂 (诸如,例如,戊二醛) 交联或固定于固相上 (例如 Eupergit C 和类似固相) 被稳定。

[0317] 固体反应物 (例如酶、盐等) 可以多种不同形式提供给反应,包括粉末 (例如冻干粉末、喷雾干燥的粉末和类似粉末)、溶液、乳剂、悬浮液和类似形式。使用本领域普通技术人员已知的方法和设备可容易地冻干或喷雾干燥反应物。例如,蛋白溶液可以小等份冷冻于 -80°C , 然后加入到预先冷却的冻干箱,随后应用真空。从样品去除水之后,在释放真空和

回收冻干样品之前,温度一般升高到 4℃持续 2 小时。

[0318] 还原反应中使用的反应物的量将一般取决于需要的产物的量和伴随的采用的酮还原酶底物的量而变化。以下指南可用来确定使用的酮还原酶、辅因子和任选的辅因子再生系统的量。一般而言,采用的酮底物的浓度可为约 20 至 300 克 / 升,使用从约 50mg 至约 5g 的酮还原酶和约 10mg 至约 150mg 的辅因子。本领域普通技术人员将容易理解如何改变这些量以调整这些量使其适应需要的生产率水平和产物规模。任选的辅因子再生系统的适当的量可基于使用的辅因子和 / 或酮还原酶的量容易地通过常规实验来确定。一般而言,还原剂(例如葡萄糖、甲酸、异丙醇)以高于酮还原酶底物的等摩尔水平的水平使用以获得酮还原酶底物基本上完全或接近完全的转化。

[0319] 加入反应物的顺序不是关键的。反应物可同时一起加入到溶剂中(例如单相溶剂、双相水性共溶剂系统和类似溶剂),或可选地,可单独加入一些反应物,并且一些反应物可在不同的时间点一起加入。例如,辅因子再生系统、辅因子、酮还原酶和酮还原酶底物可首先加入到溶剂中。

[0320] 当使用水性共溶剂系统时,为了改进混合效率,可首先加入辅因子再生系统、酮还原酶和辅因子并缓和到水性中。然后可加入有机相并混合,随后加入酮还原酶底物。可选地,在加入到水相之前,酮还原酶底物可在有机相中预先混合。

[0321] 用于执行本文所述的酮还原酶-催化的还原反应的适合条件包括可容易地通过常规实验被优化的多种条件,这些实验包括但不限于使工程酮还原酶和底物在实验 pH 和温度下接触并检测产物,例如使用本文提供的实施例中描述的方法。

[0322] 酮还原酶催化的还原一般在从约 15℃至约 75℃范围的温度下进行。对一些实施方案,反应在从约 20℃至约 55℃范围的温度下进行。在其他实施方案中,反应在从约 20℃至约 45℃范围的温度下进行。反应还可在常温条件下进行。

[0323] 一般允许还原反应进行,直至获得底物的基本完全或接近完全的还原。底物还原为产物可通过检测底物和 / 或产物使用已知方法监测。适合的方法包括气相层析、HPLC 和类似方法。反应混合物中产生的醇还原产物的转化产量一般大于约 50%,还可能大于约 60%,还可能大于约 70%,还可能大于约 80%,还可能大于 90%,并且常常大于约 97%。

[0324] 7. 实施例

[0325] 本公开的各种特点和实施方案例证于以下代表性实施例中,这些实施例预期为例证性的而不是限制性的。

[0326] 在以下描述中,使用葡萄糖脱氢酶(GDH)的地方,其为 GDH CDX901,获自 Julich Chiral Solutions, Julich, Germany。

[0327] 7.1 实施例 1:野生型酮还原酶基因获得和表达载体的构建。

[0328] 基于酮还原酶的报道的氨基酸序列和美国临时申请序列号 60/848,950 和 W02008042876 中所述的密码子优化算法,其通过引用并入本文,设计酮还原酶(KRED)编码基因用于在大肠杆菌中表达。使用包括 42 个核苷酸的寡核苷酸合成基因,并克隆到在 lac 启动子控制下的表达载体 pCK110900 中(描述于美国专利申请公布 20060195947 的图 3)。该表达载体还含有 P15a 复制起点和氯霉素抗性基因。使用标准方法将得到的质粒转化到大肠杆菌 W3110 中。密码子优化的基因以及编码的多肽列于表 3。如美国临时申请序列号 60/848,950 中所述证实了野生型酮还原酶的活性。

[0329]

表3: 使用的酮还原酶的缩写、来源和参考文献					
酮还原酶缩写	从中最初鉴定酶的微生物	Genebank 登录号	GI号	多核苷酸 SEQ ID NO.	多肽SEQ ID NO.或来源
ADH-CM	木兰假丝酵母	AB036927.1	12657576	SEQ ID NO 1, 于美国申请公布20060195947	SEQ ID NO 2, 于美国申请公布20060195947
YDL	酿酒酵母	NP_010159.1	6320079	SEQ ID NO:109	SEQ ID NO:110
ADH-LB	短乳杆菌	1NXQ_A	30749782	SEQ ID NO:1	SEQ ID NO:2
ADH-RE	红平红杆菌	AAN73270.1	34776951	SEQ ID NO:111	SEQ ID NO:112
YGL	酿酒酵母	NP_011476	6321399	SEQ ID NO:113	SEQ ID NO:114
YPR	酿酒酵母	NP_010656.1	6320576	SEQ ID NO:115	SEQ ID NO:116
GRE	酿酒酵母	NP_014490.1	6324421	SEQ ID NO:117	SEQ ID NO:118
ADH-LK	克菲尔乳杆菌	AAP94029.1	33112056	SEQ ID N:3	SEQ ID NO:4
ADH-SB	赭色掷孢酵母	Q9UUN9	30315955	SEQ ID No 103	SEQ ID No 104
ADH-SC	天蓝色链霉菌	NP_631415.1	21225636	SEQ ID No 101	SEQ ID No 102
ADH-TB	布氏热厌氧菌	X64841.1	1771790	SEQ ID No 107	SEQ ID No 108
ADH-HE	马肝	DEHOAL	625197	SEQ ID No 105	SEQ ID NO 106
ADH-CP	近平滑假丝酵母	BAA24528	2815409		Julich Chiral Solutions 目录号 03.11
DR-LB	短乳杆菌双乙酰还原酶	ABJ63353.1	116098204		Julich Chiral Solutions 目录号 8.1
ADH-CB	博伊丁假丝酵母	CAD66648	28400789		Julich Chiral Solutions 目录号 02.10
LDH-LL	莱希曼氏乳杆菌				Fluka 目录号 61306

[0330] 编码本公开的工程酮还原酶的多核苷酸被同样地克隆到载体 pCK110900 中用于在大肠杆菌 W3110 中表达。

[0331] 7.2 实施例 2: 酮还原酶粉末的产生; 摇瓶程序。

[0332] 含有具有感兴趣的酮还原酶基因的质粒的大肠杆菌的单个微生物菌落被接种到含有 30 μ g/ml 氯霉素和 1% 葡萄糖的 50ml Luria Bertani 肉汤中。细胞在 30°C 下伴有

250rpm的振荡下在培养箱中生长过夜(至少16小时)。培养物稀释到250ml Terrific 肉汤中(12g/L 细菌胰蛋白胨、24g/L 酵母提取物、4ml/L 甘油、65mM 磷酸钾, pH 7.0、1mM MgSO_4 、30 $\mu\text{g/ml}$ 氯霉素)于1升瓶中)至600nm的光密度(OD600)为0.2,并允许其在30℃下生长。当培养物的OD600为0.6至0.8时,用1mM IPTG诱导酮还原酶基因的表达,并孵育过夜(至少16小时)。通过离心(5000rpm,15min,4℃)收获细胞,并弃掉上清液。细胞沉淀用等体积的冷(4℃)100mM 三乙醇胺(氯化物)缓冲液, pH 7.0(在ADH-LK和ADH-LB和从其衍生的工程酮还原酶的实例中包括2mM MgSO_4)重悬,并通过如上文的离心收获。洗涤的细胞在两倍体积的冷三乙醇胺(氯化物)缓冲液中重悬,并以12000psi通过French Press两次,同时保持在4℃。通过离心(9000rpm,45min.,4℃)去除细胞碎片。收集透明的裂解物上清液并储存于-20℃。冷冻的透明裂解物的冻干提供了粗酮还原酶的干粉。

[0333] 7.3 实施例3:酮还原酶的产生;发酵程序。

[0334] 在通气搅拌的15L发酵罐中,使含有0.88g/L 硫酸铵、0.98g/L 柠檬酸钠;12.5g/L 三水合磷酸氢二钾、6.25g/L 磷酸二氢钾、6.2g/L Tastone-154 酵母提取物、0.083g/L 柠檬酸铁铵和8.3ml/L 微量元素溶液的6.0L生长培养基达到30℃的温度,所述微量元素溶液含有2g/L 二水合氯化钙、2.2g/L 七水合硫酸锌、0.5g/L 一水合硫酸锰、1g/L 七水合硫酸铜、0.1g/L 四水合钼酸铵和0.02g/L 十水合四硼酸钠。发酵罐用含有感兴趣的酮还原酶基因的质粒的指数生长末期的大肠杆菌W3110培养物接种,所述培养物如实施例3中所述在摇瓶中生长至起始OD600为0.5至2.0。发酵罐以500-1500rpm搅拌,并以1.0-15.0L/min向发酵容器提供空气以保持30%饱和或更高的溶解氧水平。通过加入20% v/v 氢氧化铵控制培养物的pH为7.0。通过加入含有500g/L 工业葡萄糖(cerelease)、12g/L 氯化铵和10.4 g/L 七水合硫酸镁的料液保持培养物的生长。在培养物达到OD600为50后,通过加入终浓度为1mM的异丙基- β -D-硫代半乳糖苷(IPTG)诱导酮还原酶的表达。培养物生长另外14小时。然后冷却培养物至4℃并保持在4℃直至收获。通过在Sorval RC 12BP离心机中于4℃ 5000G下离心40分钟收获细胞。收获的细胞直接用在以下的下游回收过程或储存于4℃直至此类使用。

[0335] 细胞沉淀于4℃下重悬,每体积的湿细胞糊剂(wet cell paste)重悬于2倍体积的100mM 三乙醇胺(氯化物)缓冲液, pH 6.8。使用12000psig的压强通过使悬浮液通过装有两相匀浆器部件的匀浆器,使胞内酮还原酶从细胞中释放。破坏后立即将细胞匀浆冷却至4℃。将10% w/v 聚乙烯亚胺, pH 7.2的溶液加入到裂解物中至终浓度为0.5% w/v并搅拌30分钟。得到的悬浮液通过在标准实验室离心机内以5000G离心30分钟变得澄清。轻轻倒出透明的上清液,并使用具有30Kd分子量截留的纤维素超滤膜浓缩10倍。最终的浓缩液分散到浅容器中,在-20℃下冷冻并冻干为粉末。酮还原酶粉末储存于-80℃。

[0336] 7.4 实施例4:确定2',6'-二氯-3'-氟苯乙酮向(S)-1-[2',6'-二氯-3'-氟苯基]-乙醇的转化和对映体过量的分析方法。

[0337] 2',6'-二氯-3'-氟苯乙酮的还原和醇产物的手性纯度通过反相手性HPLC(4.6x150mm Chiralpak AD-RH柱(没有保护柱(guard cartridge));50:50ACN/ H_2O , 0.8mL/min;25℃;使用以下保留时间于254nm下检测:(S)-醇5.77min;(R)-醇6.19min;酮7.49min)或正相手性HPLC(4.6x250mm ChiralpakAD柱(没有保护柱);2:98IPA/己烷以2.5mL/min于室温下(未调节);使用以下保留时间于220nm下检测:(S)-醇4.72min;

(R)-醇 5.30min;酮 2.03min) 来确定。

[0338] 可选地,使用以下气相层析分析方法:使用 HP-5 柱 (30mx0.25mm)、温度程序为以 50℃ /min 的 100℃ (1min) 至 200℃ (4min) (酮的保留时间是 4.33min,醇的保留时间是 4.70min) 的非手性方法和使用 Beta Cyclodextrin (DM) 柱 (30mx0.25mm) 以 165℃ 恒温 (酮的保留时间是 3.42min, R-异构体的保留时间是 5.92min, R-异构体的保留时间是 6.25min) 的手性方法。

[0339] 7.5 实施例 5:评估野生型酮还原酶对 2',6'-二氯-3'-氟苯乙酮的还原

[0340] 描述于实施例 1 的表 3 的 KRED 使用化学计量的 NADH 或 NADPH 作为辅因子来筛选。向 96 深孔平板的每个孔加入 5-10mg KRED、于 500 μL 100mM pH 7.0 三乙醇胺 (氯化物) 缓冲液的 20mg NAD(P)H 和 15 μL 底物 (~40g/L 底物;通过辅因子的转化限制为 ~25)。将平板封口并振荡 6 小时。通过加入 1mL EtOAc 猝灭反应。产物的转化和立体纯度如实施例 4 所述测定。

[0341] 在这些条件下,使用 NADPH 或 NADH, YDL、YGL、GRE、ADH-RE、ADH-SB、ADH-SC、ADH-HL、LDH-LL、ADH-CP、ADH-CB 和 DR-LB 不产生可检测的转化,而 ADH-LB 和 YPR 产生 < 0.5% 的转化。ADH-LK 将 < 1% 的底物转化为手性醇。

[0342] 本实施例证明野生型酮还原酶对 2',6'-二氯-3'-氟苯乙酮具有非常低的活性,如果有活性的话。

[0343] 7.6 实施例 6:评估 ADH-LK 变体对 2',6'-二氯-3'-氟苯乙酮的还原

[0344] 当在实施例 5 中所述和如表 4 中所列的条件下评估时,已如 2007 年 8 月 24 日提交的美国申请序列号 60/957,974 和 2008 年 8 月 24 日提交的美国申请序列号 12/197,286 中所述产生的几种 ADH-LK 变体将 > 0.5% 的底物转化为手性醇。

[0345]

表4		
SEQ ID NO. ^c	转化 ^a	对映体过量 ^b
4	0	0
115	+	+
117	+	++
111	+	++
119	++	++
121	++	++
113	++	++
6	++	++

[0346] ^a0:0.5-1% 的转化;+:1-20% 的转化;++ > 20% 的转化

[0347] ^b0:< 90% e. e. (S-对映体);+:90-99% e. e. (S-对映体);++ > 99% e. e. (S-对映体)。

[0348] ^c指美国申请序列号 60/957,974 和 12/197,286 中的 SEQ ID NO

[0349] 本实施例说明其中位置 190 的酪氨酸残基被改变为苯丙氨酸、脯氨酸、半胱氨酸或丙氨酸的 ADH-LK 变体将 2',6' - 二氯 -3' - 氟苯乙酮还原为对应的 S- 醇。

[0350] 7.7 实施例 7: 鉴定还原 2',6' - 二氯 -3' - 氟苯乙酮的酶的高通量 NADPH 荧光预筛选。

[0351] 通过定向演化获得并含有演化的酮还原酶基因的质粒文库被转化到大肠杆菌 W3110, 并铺板到含有 1% 葡萄糖和 30 μ g/mL 氯霉素 (CAM) 的 Luria-Bertani (LB) 琼脂培养基上。在 30°C 下孵育至少 16 小时后, 使用 Q-bot® 机器人菌落挑取器 (Genetix USA, Inc., Beaverton, OR) 将菌落挑取到含有 180 μ L Terrific 肉汤 (TB)、1% 葡萄糖和 30 μ g/mL 氯霉素 (CAM) 的 96 孔浅孔微量滴定板中。细胞在 30°C 下、200rpm 振荡生长过夜。然后将 5 μ L 该培养物转移到含有 380 μ L Terrific 肉汤 (TB)、1mM $MgSO_4$ 和 30 μ g/mL CAM 的 96 深孔平板中。在深孔平板在 30°C 下、250rpm 振荡孵育 2.5 至 3 小时后 (OD_{600} 0.6-0.8), 细胞培养物的重组基因表达通过终浓度为 1mM 的异丙基硫代半乳糖苷 (IPTG) 诱导。然后平板在 30°C 下、250rpm 振荡孵育 15-17 小时。

[0352] 通过离心沉淀细胞, 并重悬于 300 μ L 裂解缓冲液中, 通过室温下振荡至少 1 小时裂解。裂解缓冲液含有 100mM 三乙醇胺 (氯化物) 缓冲液, pH7.0-7.2、1mg/mL 溶菌酶和 750 μ g/mL 多粘菌素 B 硫酸盐。然后平板在离心机中以 4000RPM、4°C 下旋转 20 分钟, 并在荧光测定中分析透明的上清液 (裂解物)。

[0353] 在 96 孔黑色微量滴定板中, 将 20 μ L 每种裂解物 (于 40-50°C 下预先处理 0-24hr, 如果需要, 于 100mM 三乙醇胺 (氯化物) 缓冲液, pH 7.0、1mM $MgSO_4$ 中稀释) 加入到由 100mM 三乙醇胺 (氯化物) 缓冲液, pH7.0、1mM $MgSO_4$ 、0.2g/L NADPH、100-600mM 葡萄糖、600-900mM 葡萄糖酸钠和 0.2g/L 2',6' - 二氯 -3' - 氟苯乙酮组成的 180 μ L 测定混合物中, 通过在 Flexstation (Molecular Devices, USA) 中 330nm 激发后观察 445nm 处 NADPH 荧光的降低来测量反应的进程。

[0354] 本实施例描述了用来鉴定改进了 2',6' - 二氯 -3' - 氟苯乙酮的还原速率的 KRED 变体的方法。

[0355] 7.8 实施例 8: 衍生自 ADH-LK 的工程酮还原酶对 2',6' - 二氯 -3' - 氟苯乙酮的还原。

[0356] 用于将 2',6' - 二氯 -3' - 氟苯乙酮还原为 (S)-1-[2',6' - 二氯 -3' - 氟苯基]-乙醇的改进的 ADH-LK 变体在小规模化学反应中分析。于 25°C 下向 100mL 三颈容器加入 30mL 100mM 三乙醇胺 (氯化物) 缓冲液 (pH 7)、2mM $MgSO_4$ 、200mg 具有如下表所述的 SEQ ID NO 的 KRED、50mg GDH、15mg NADP-Na、3.13g 葡萄糖、6g 2',6' - 二氯 -3' - 氟苯乙酮 (200g/L), 所述三颈容器装有 PTFE- 包被的磁性搅拌杆和 pH 电极, 该 pH 电极与自动滴定仪连接以通过通到容器内的饲管 pH- 控制地加入需要的碱。自动滴定仪通过加入 4N NaOH 保持 pH 为 7, 其被连续记录。通过碱的加入速率和累积加入以及反应混合物的定期取样来监测反应进程, 定期取样用于使用乙酸乙酯来提取和通过实施例 4 的方法分析。

[0357] 表 5 给出了对应于酮还原酶的 SEQ ID NO、从野生型 ADH-LK 的氨基酸突变的数目和 2',6' - 二氯 -3' - 氟苯乙酮向 (S)-1-[2',6' - 二氯 -3' - 氟苯基]-乙醇的转化。S- 醇的立体纯度总是 > 99.9%。

[0358] 表 5:

[0359]

表5: ADH-LK变体的活性和稳定性			
SEQ ID NO.	从ADH-LK 的突变数	活性	稳定性
4	0	-	
6	1	+	
8	2	++	+
10	2	++	
12	2	+	
14	2	++	
16	2	++	+
48	2	+	

[0360]

50	2	+	
52	3	+	
20	3	+	+
18	4	+++	+
36	3	+++	
54	3	++	+
56	3	++	
62	3	++	
64	3	++	
58	3	++	
60	2	++	
22	3	++	++
66	3	++	+
68	3	++	++
70	4	++	++
72	3	++	++
32	3	+++	+
28	3	++	+
30	3	++	+
34	4	+++	+
74	4	+++	+
38	4	+++	++
40	5	+++	++
76	5	+++	++
78	5	+++	++
80	5	+++	++
82	5	+++	++
84	8	+++	++
42	6	+++	++
44	6	+++	++
86	7	+++	++
46	6	+++	++
88	6	+++	++
90	6	+++	++
92	6	+++	++
94	6	+++	++

[0361] a. - :没有活性 ;+ :100-450%的 SEQ ID No. 6 的活性 ;++ :450-1500%的 SEQ ID No. 6 的活性 ;+++ :> 1500%的 SEQ ID No. 6 的活性。

[0362] b. + :50℃下 2 小时后可测量的活性 ;++ :50℃下 2 小时后 > 400%的 SEQ ID No. 16

的活性。

[0363] 本实施例说明衍生自野生型酮还原酶 ADH-LK 的工程酮还原酶与酮还原酶 ADH-LK 相比提供了改进的活性。

[0364] 7.9 实施例 9:衍生自 ADH-LB 的工程酮还原酶对 2',6' - 二氯 -3' - 氟苯乙酮的还原。

[0365] 用于将 2',6' - 二氯 -3' - 氟苯乙酮还原为 (S)-1-[2',6' - 二氯 -3' - 氟苯基]-乙醇的改进的 ADH-LB 变体如实施例 8 中的 ADH-LK 变体所述在小规模化学反应中分析。表 6 给出了对应于酮还原酶的 SEQ ID NO、从野生型 ADH-LK 的氨基酸突变的数目和 2',6' - 二氯 -3' - 氟苯乙酮向 (S)-1-[2',6' - 二氯 -3' - 氟苯基]-乙醇的转化。S-醇的立体纯度总是 > 99.9%。

[0366] 表 6:

[0367]

表6: ADH-LK变体的活性和稳定性			
SEQ ID NO.	从ADH-LK的突变数目	活性	稳定性
4	0	-	
6	1	+	
8	2	++	+
10	2	++	
12	2	+	
14	2	++	
16	2	++	+
48	2	+	
50	2	+	
52	3	+	
20	3	+	+
18	4	+++	+

[0368] a. -:没有活性;+:100-450%的 SEQ ID No. 6 的活性;+:450-1500%的 SEQ ID No. 6 的活性;+++ :> 1500%的 SEQ ID No. 6 的活性。

[0369] b. + :50℃下 2 小时后可测量的活性;+:50℃下 2 小时后 > 400%的 SEQ ID No. 16 的活性。

[0370] 本实施例说明衍生自野生型酮还原酶 ADH-LB 的工程酮还原酶与酮还原酶 ADH-LB 相比也提供了改进的活性。

[0371] 7.10 实施例 10:制备规模产生 (S)-1-[2',6' - 二氯 -3' - 氟苯基]-乙醇

[0372] 向 500mL 套层 (jacketed) 三颈圆底瓶加入水 (120mL)、三乙醇胺 (1.8g), 然后加入氢氯酸来调节 pH 至 7.0, 所述套层三颈圆底瓶装有 Ace Glass 机械搅拌器 (75mm 直径的特氟龙搅拌刀片) 和 pH 电极, 该 pH 电极与自动滴定仪连接以通过通到容器内的输送管 pH- 控制地加入需要的碱。加入 1M 的硫酸镁溶液 (120 μ L, 0.12mmol, 14.4mg MgSO₄)。

通过使加热液循环通过瓶的套层将溶液加热至 30℃。加入葡萄糖 (20g), 然后加入 Na-NADP(120mg)、GDH(0.50g) 和具有 SEQ ID No. 38 的 KRED(0.50g)。pH 稳态被设定为通过输送管加入 4N NaOH 保持 pH 为 7.0±0.1。加入 2',6'-二氯-3'-氟苯乙酮 (50g) 起始反应。需要定期润洗电极以去除酶衍生的材料。随着反应的进行, 分部加入另外的葡萄糖: 10g 于 104min(在已加入 17.5mL 4N NaOH 之后)、5g 于 275min(在已加入 35.2mL 4N NaOH 之后)、5g 于 379min(在已加入 42mL 4N NaOH 之后) 和 8g 于 488min(在已加入 47mL 4N NaOH 之后)。24 小时后停止反应。然后加入庚烷 (150mL), 并将混合物加热至 40℃ 持续 45 分钟。在冷却至 30℃ 后, 将得到的混合物倒入分液漏斗, 排干大多数底层水性。上层的庚烷乳剂在真空下通过硅藻土垫过滤 (350mL, 85mm 直径的粗滤器)。用庚烷 (150mL) 洗涤滤器, 将滤液转移到分液漏斗并分离两相。庚烷相在旋转真空蒸发器上浓缩 (~50℃, ~150mmHg 增加到 40mmHg) 以产生作为油的 (S)-1-[2',6'-二氯-3'-氟苯基]-乙醇 (47.8g, 94%), 其一旦静置就结晶。

[0373] 7.11 实施例 11: 衍生自 ADH-LK 的工程酮还原酶对 2',6'-取代苯乙酮的还原。

[0374] 测试野生型 ADH-LK 和为了将 2',6'-二氯-3'-氟苯乙酮还原为 (S)-1-[2',6'-二氯-3'-氟苯基]-乙醇而改进的 ADH-LK 变体对两种其他 2',6'-取代苯乙酮的活性。制备 5mL 100mM 三乙醇胺 (氯化物) 缓冲液 (pH 7, 2mM MgSO₄)、33mg 具有 SEQ ID No. 10 的 KRED、8mg GDH、3mg NADP-Na 和 330mg 葡萄糖的溶液。1ml 该溶液于 25℃ 下用 0.3ml 1M(钠) 磷酸缓冲液 pH 7 和 20mg 2',6'-取代苯乙酮处理。反应样品 (24hr) 通过实施例 5 的方法分析。

[0375] 表 7 给出了使用 ADH-LK 和具有 SEQ ID NO 10 的 ADH-K 变体的两种 2',6'-取代苯乙酮的转化和得到的手性醇的对映体纯度。

[0376] 表 7:

[0377]

表 7: ADH-LK 和 ADH-LK 变体对两种 2',6'-取代苯乙酮的活性				
2',6'-取代苯乙酮	野生型 ADH-LK (SEQ ID NO. 4)		工程 KRED SEQ ID NO. 10	
	转化 (%)	立体纯度 (%)	转化 (%)	立体纯度 (%)
2',6'-二氯苯乙酮	0	-	88	>99 S
2',6'-二甲氧基苯乙酮	0	-	72	>99 S

[0378] 本实施例显示含有 Y190P 突变的 ADH-LK 变体提供了对 2',6'-取代苯乙酮的改进的活性, 并且提供了对应的 2',6'-取代 (S)-1-苯乙醇。

[0379] 7.12 实施例 12: ADH-LK 和衍生自 ADH-LK 的含有 Y190 突变的工程酮还原酶对未取代的苯乙酮的还原。

[0380] 向如实施例 7 中制备的每孔含有 100 μL 细胞裂解物的 96 孔板的每个孔加入: 50 μL 于 100mM 三乙醇胺 (氯化物) 缓冲液 pH 7.0 的 7mM Na-NADP⁺、300 μL 异丙醇和 50 μL 于 THF 的 100mg/ml 苯乙酮。封口平板, 在轨道振荡器上以 850rpm 室温下搅拌 24h。向每

孔加入 1mL 甲基叔丁基醚 (MTBE), 并封口平板, 然后以 850rpm 室温下振荡 10 分钟。平板在 4,000rpm (3220xg) 下离心 2 分钟以分离相, 从每孔将 50 μ L 有机相转移到含有 150 μ L MTBE 的浅孔平板的孔中。封口平板并通过正相 HPLC 分析 (配有 OD-H 保护柱的 Daicel Chiralcel OD-H 柱 (4.6x250mm); 2.5 μ L 注射; 流动相: 95 : 5v/v 庚烷-IPA; 流速: 1.5mL min⁻¹; 柱温: 40°C; 波长: 215nm)。保留时间: 苯乙酮: 3.5min; (R)-1-苯基乙醇: 5.3min; (S)-1-苯基乙醇: 5.8min。

[0381] 表 8 显示了 ADH-LK 和 ADH-LK 变体对苯乙酮的转化和得到的手性醇的立体异构纯度。

[0382] 表 8:

[0383]

表8: ADH-LK对苯乙酮的活性和立体选择性			
SEQ ID NO.	转化 (%)	对映体过量 (%)	R或S
4 (野生型)	94	100	R
14	79	46	S
22	80	35	S
28	80	48	S
32	77	50	S
36	89	86	S
42	90	57	S
44	90	41	S

[0384] 本实施例证明野生型乳杆菌酮还原酶对苯乙酮是 R- 选择性, 而从其衍生的本发明的工程酮还原酶对苯乙酮是 S- 选择性。

[0385] 尽管已例证和描述了各种具体的实施方案, 应理解可进行各种改变而不偏离本发明的精神和范围。

[0386] 对于所有的目的, 本申请中引用的所有出版物、专利、专利申请和其他文件通过引用以其整体并入本文, 如同对于所有的目的, 单独地指明每个个体出版物、专利、专利申请或其他文件通过引用并入。

序列表

<110> 杰克·梁

史提芬尼·珍妮

艾米丽·穆德弗

沙琳·奇恩

约翰·格鲁伯

安克·克莱伯

吉伽特·哈思曼

<120> 用于还原苯乙酮的酮还原酶多肽

<130>376247-017

<141>2008-09-13

<150>60/972, 058

<151>2007-09-13

<160>119

<210>1

<211>762

<212>DNA

<213> 人工序列

<220>

<223> 密码子优化的短乳杆菌 (L. Brevis) 序列

<400>1

```
atgtctaacc gtctggatgg caaagtagcc atcattaccg gcgggactct gggtatcggt 60
ttggcaatcg ccacgaaatt tgtagaggag ggtgcgaaag taatgattac tggtcgtcac 120
tccgatgtag gtgaaaaggc cgccaaatca gtaggcactc cggatcagat tcagtttttt 180
cagcacgatt catccgatga agatggctgg acgaaactgt tcgacgccac cgagaaaagca 240
ttcggecccg ttagcacctt agtgaacaat gcagggattg cagttaacaa aagcgttgaa 300
gaaactacca cggccgaatg gcgtaaaactg ctggccgtta atctggatgg tgtttttt 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgagcctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
```

gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctgggtcga tgatctgccg 600
 ggtgctgagg aagcgatgtc acagegtacg aaaaccacct tgggccacat tggcgaaccg 660
 aatgacatcg catatatctg tgtgtacctg gcactaatg aatcgaaatt tgcgacgggt 720
 tccgaatttg tggtegacgg cgggtatacc gcacagtaat ga 762

<210>2

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 密码子优化的短乳杆菌 (L. Brevis) 序列的翻译

<400>2

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

Arg Thr Lys Thr Pro Met Gly His Ile Gly Glu Pro Asn Asp Ile Ala
 210 215 220
 Tyr Ile Cys Val Tyr Leu Ala Ser Asn Glu Ser Lys Phe Ala Thr Gly
 225 230 235 240
 Ser Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln
 245 250

<210>3

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 密码子优化的克菲尔乳杆菌 (L. Kefir) 序列

<400>3

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
 ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
 ttcgggcccg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
 gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
 ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
 atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctgggtcga tgatctggaa 600
 ggtgctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
 aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
 gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>4

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 密码子优化的克菲尔乳杆菌 (L. Kefir) 序列的翻译

<400>4

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>5

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>5

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctgc atcaagaccc cgctgggtcga tgatctggaa 600
gggtctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>6

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>6

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Cys	Ile	Lys
180				185				190							
Thr	Pro	Leu	Val	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>7

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>7

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcgcg atcaagaccc cgctgatcga tgatctggaa 600
gggtctgagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>8

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>8

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>9

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>9

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggccccg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgatcga tgatctggaa 600
ggtgctgagg aaatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>10

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>10

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>11

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>11

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg  60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccag aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctgt atcaagaccc cgctgggtcga tgatctggaa 600
gggtctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga                               759

```

<210>12

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>12

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Arg Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Cys	Ile	Lys
180				185				190							
Thr	Pro	Leu	Val	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>13

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>13

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 60
ttggcaatcg ccgataaatt ttagaggag ggtgcgaaag tagttattac tggcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggctgtgagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>14

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>14

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>15

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>15

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcgcc atcaagaccc cgctgctaga tgatctggaa 600
ggtgctgagg aaatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>16

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>16

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>17

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>17

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgatgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgatcga tgatctggaa 600
gggtctgagg aaaggatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>18

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>18

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Met	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Pro	Ile	Lys
180				185				190							
Thr	Pro	Leu	Ile	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Arg	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>19

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>19

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg caatttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgatcga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```


aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>20

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>20

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>21

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>21

```

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggccccg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aaatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>22

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>22

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>23

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 短乳杆菌 (*L. brevis*) 的变体

<400>23

```

atgtctaacc gtctggatgg caaagtagcc atcattaccg gcgggactct gggatatcgg 60
ttggcaatcg ccacgaaatt tgtagaggag ggtgcgaaag taatgattac tggtcgtcac 120
tccgatgtag gtgaaaagge cgccaaatca gtaggcactc cggatcagat tcagtttttt 180
cagcacgatt catccgatga agatggctgg acgaaactgt tcgacgccac cgagaaagca 240
ttcggcccggt ttagcacctt agtgaacaat gcagggattg cagttaacaa aagcgttgaa 300
gaaactacca cggccgaatg gcgtaaactg ctggccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgagcctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctgga tgatctgccg 600
ggtgctgagg aagcgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catatatctg tgtgtacctg gcactaatg aatcgaaatt tgcgacgggt 720
tccgaatttg tggtcgacgg cgggtatacc gcacagtaa 759

```

<210>24

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 短乳杆菌 (*L. brevis*) 的变体

<400>24

```

Met Ser Asn Arg Leu Asp Gly Lys Val Ala Ile Ile Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Thr Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Met Ile Thr Gly Arg His Ser Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Val Gly Thr Pro Asp Gln Ile Gln Phe Phe Gln His Asp Ser
50           55           60
Ser Asp Glu Asp Gly Trp Thr Lys Leu Phe Asp Ala Thr Glu Lys Ala
65           70           75           80
Phe Gly Pro Val Ser Thr Leu Val Asn Asn Ala Gly Ile Ala Val Asn
85           90           95
Lys Ser Val Glu Glu Thr Thr Thr Ala Glu Trp Arg Lys Leu Leu Ala
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Ser	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Pro	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Pro	Gly	Ala	Glu	Glu	Ala	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Tyr	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asn	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ser	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>25

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 短乳杆菌 (*L. brevis*) 的变体

<400>25

```

atgtctaacc gtctggatgg caaagtagcc atcattaccg gcgggactct gggatatcgg 60
ttggcaatcg ccacgaaatt tntagaggag ggtgcgaaag taatgattac tggtcgtcac 120
tccgatgtag gtgaaaaggc cgccaaatca gtaggcactc cggatcagat tcagtttttt 180
cagcacgatt catccgatga agatggctgg acgaaactgt tcgacgccac cgagaaagca 240
ttcgccccgg ttagcacctt agtgaacaat gcagggattg cagttaacaa aagcgttgaa 300
gaaactacca cggccgaatg gcgtaactg ctggccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgagcctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcgcg atcaagaccc cgctgctgga tgatctgccg 600
gggtctgagg aagcgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catatatctg tgtgtacctg gcatctaatt aatcgaaatt tgcgacgggt 720
 tccgaatttg tggtegacgg cgggtatacc gcacagtaa 759

<210>26

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 短乳杆菌 (*L. brevis*) 的变体

<400>26

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

210	215	220	
Tyr Ile Cys Val Tyr Leu Ala Ser Asn Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ser Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>27

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>27

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtatggagg aaatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggc 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>28

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>28

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>29

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>29

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcgcc atcaagaccc cgctgctcga tgatctggaa 600
ggttacgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>30

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>30

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Ala	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Tyr	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>31

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>31

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>32

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>32

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>33

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>33

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg tagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcgcc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aagtgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>34

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>34

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>35

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>35

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aatgtattc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>36

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>36

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Pro	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Met	Tyr	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>37

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>37

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcggt 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccggt ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>38

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>38

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>39

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>39

```

atgaccgatac gtctgaagca taaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>40

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>40

Met Thr Asp Arg Leu Lys His Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>41

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>41

```

atgaccgatac gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg tagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>42

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>42

```

Met Thr Asp Arg Leu Lys His Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Val Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Leu	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Pro	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Trp	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>43

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>43

```

atgaccgatc gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagttgaaaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>44

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>44

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>45

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>45

```

atgaccgatac gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccggt ttacgaccgt cgtgaacaat gcagggattg cagttacaaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>46

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>46

Met Thr Asp Arg Leu Lys His Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>47

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>47

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgatgga tgatctggaa 600
gggtctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>48

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>48

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```


Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Pro	Ile	Lys
180				185				190							
Thr	Pro	Leu	Met	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>49

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>49

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgttcga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>50

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>50

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>51

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>51

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgctc tgagcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aaatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>52

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>52

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>53

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>53

```

atgaccgatc gtctgaagaa caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcgcc atcaagaccc cgctgctaga tgatctggaa 600
gggtctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>54

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>54

```

Met Thr Asp Arg Leu Lys Asn Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Ala	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>55

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>55

```

atgaccgatc gtctgaagca caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctaga tgatctggaa 600
gggtctgagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>56

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>56

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>57

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>57

```

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggagtct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggccccg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>58

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>58

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Ser

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

<210>59

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>59

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
gggtctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>60

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>60

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Phe	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Pro	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>61

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>61

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt ttagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cactgtccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctaga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>62

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>62

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>63

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>63

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggccccg ttacgaccgt cgtgaacaat gcagggattg caatttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>64

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>64

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>65

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>65

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggca ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcgcc atcaagaccc cgctgctaga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>66

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>66

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Gln	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Ala	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>67

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>67

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggat agtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```


aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>68

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>68

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>69

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>69

```

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatatag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggccccg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>70

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>70

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>71

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>71

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgtcgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccctgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct tgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcgcc atcaagaccc cgctgctaga tgatctggaa 600
gggtctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>72

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>72

```

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Leu	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Ala	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Ala	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>73

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>73

```

atgaccgatc gtctgaaggg caaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccag aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggat agtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>74

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>74

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>75

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>75

```

atgaccgatac gtctgaagac caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aaatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>76

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>76

Met Thr Asp Arg Leu Lys Thr Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>77

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>77

```

atgaccgatac gtctgaagcc caaagtagcc atcgtaaccg gcgggactct gggatatcggt 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>78

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>78

```

Met Thr Asp Arg Leu Lys Pro Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Leu	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Pro	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Trp	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>79

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>79

```

atgaccgatc gtctgaagtg gaaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>80

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>80

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>81

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>81

```

atgaccgatac gtctgaagcg taaagtagcc atcgtaaccg gcgggactct gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggccccg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>82

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>82

Met Thr Asp Arg Leu Lys Arg Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>83

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>83

```

atgaccgatac gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattt gcgtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtattccat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>84

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>84

```

Met Thr Asp Arg Leu Lys His Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Cys Val Ser
85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val Asn Leu Asp Gly Val Phe Phe Gly Thr Arg Leu Gly Ile Gln Arg			
115	120	125	
Met Lys Asn Lys Gly Leu Gly Ala Ser Ile Ile Asn Met Ser Ser Ile			
130	135	140	
Glu Gly Leu Val Phe His Pro Thr Leu Gly Ala Tyr Asn Ala Ser Lys			
145	150	155	160
Gly Ala Val Arg Ile Met Ser Lys Ser Ala Ala Leu Asp Cys Ala Leu			
165	170	175	
Lys Asp Tyr Asp Val Arg Val Asn Thr Val His Pro Gly Pro Ile Lys			
180	185	190	
Thr Pro Leu Leu Asp Asp Leu Glu Gly Trp Glu Glu Met Met Ser Gln			
195	200	205	
Arg Thr Lys Thr Pro Met Gly His Ile Gly Glu Pro Asn Asp Ile Ala			
210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>85

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>85

```

atgaccgatc gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagttactaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgtttttgtc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>86

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>86

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>87

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>87

```

atgaccgatac gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggccccg ttacgaccgt cgtgaacaat gcagggattg cagtttccgt gagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttcttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>88

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>88

Met Thr Asp Arg Leu Lys His Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>89

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>89

```

atgaccgatc gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccat gagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcaccgcgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcacttgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>90

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>90

```

Met Thr Asp Arg Leu Lys His Lys Val Ala Ile Val Thr Gly Gly Thr
1           5           10           15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
20           25           30
Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
35           40           45
Lys Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
50           55           60
Ser Asp Glu Ala Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala
65           70           75           80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser
85           90           95
Met Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser
100          105          110

```

Val	Asn	Leu	Asp	Gly	Val	Phe	Phe	Gly	Thr	Arg	Leu	Gly	Ile	Gln	Arg
115				120				125							
Met	Lys	Asn	Lys	Gly	Leu	Gly	Ala	Ser	Ile	Ile	Asn	Met	Ser	Ser	Ile
130				135				140							
Glu	Gly	Leu	Val	Gly	Asp	Pro	Thr	Leu	Gly	Ala	Tyr	Asn	Ala	Ser	Lys
145				150				155				160			
Gly	Ala	Val	Arg	Ile	Met	Ser	Lys	Ser	Ala	Ala	Leu	Asp	Cys	Ala	Leu
165				170				175							
Lys	Asp	Tyr	Asp	Val	Arg	Val	Asn	Thr	Val	His	Pro	Gly	Pro	Ile	Lys
180				185				190							
Thr	Pro	Leu	Leu	Asp	Asp	Leu	Glu	Gly	Trp	Glu	Glu	Met	Met	Ser	Gln
195				200				205							
Arg	Thr	Lys	Thr	Pro	Met	Gly	His	Ile	Gly	Glu	Pro	Asn	Asp	Ile	Ala
210				215				220							
Trp	Ile	Cys	Val	Tyr	Leu	Ala	Ser	Asp	Glu	Ser	Lys	Phe	Ala	Thr	Gly
225				230				235				240			
Ala	Glu	Phe	Val	Val	Asp	Gly	Gly	Tyr	Thr	Ala	Gln				
245				250											

<210>91

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>91

```

atgaccgatc gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcgccccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccac gagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtagc aaaacccta tgggccacat tggcgaaccg 660

```

aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>92

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L.Kefir) 的变体

<400>92

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

210	215	220	
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
245	250		

<210>93

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>93

```

atgaccgatac gtctgaagca taaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggccccg ttacgaccgt cgtgaacaat gcagggattg cagtttccat cagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600
ggttgggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

```

<210>94

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>94

Met Thr Asp Arg Leu Lys His Lys Val Ala Ile Val Thr Gly Gly Thr

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

<210>95

<211>253

<212>PRT

<213> 人工序列

<220>

<223> 具有短乳杆菌 (*L. brevis*) 主链的工程酮还原酶序列式

<223> 合成构建体

<220>

<221>MISC_FEATURE

<222> (7).. (7)

<223>Xaa 是极性、非极性、酸性、芳族、受限或碱性残基

<220>

<221>MISC_FEATURE

<222> (16).. (16)

<223>Xaa 是极性残基

<220>

<221>MISC_FEATURE

<222> (43).. (43)

<223>Xaa 是非极性或极性残基

<220>

<221>MISC_FEATURE

<222> (60).. (60)

<223>Xaa 是芳族、非极性或脂肪族残基

<220>

<221>MISC_FEATURE

<222> (94).. (94)

<223>Xaa 是脂肪族、非极性或非胱氨酸残基

<220>

<221>MISC_FEATURE

<222> (95).. (95)

<223>Xaa 是非极性或非脂肪族残基

<220>

<221>MISC_FEATURE

<222> (96).. (96)

<223>Xaa 是酸性或极性残基

<220>

<221>MISC_FEATURE

<222>(97)..(97)
<223>Xaa 是碱性、脂肪族、非极性或极性残基

<220>
<221>MISC_FEATURE
<222>(120)..(120)
<223>Xaa 是芳族、脂肪族或非极性残基

<220>
<221>MISC_FEATURE
<222>(125)..(125)
<223>Xaa 是极性或非极性残基

<220>
<221>MISC_FEATURE
<222>(142)..(142)
<223>Xaa 是极性残基

<220>
<221>MISC_FEATURE
<222>(147)..(147)
<223>Xaa 是芳族、脂肪族、极性或非极性残基

<220>
<221>MISC_FEATURE
<222>(149)..(149)
<223>Xaa 是非极性或芳族残基

<220>
<221>MISC_FEATURE
<222>(150)..(150)
<223>Xaa 是酸性或受限残基

<220>
<221>MISC_FEATURE
<222>(152)..(152)
<223>Xaa 是极性或非极性残基

<220>
<221>MISC_FEATURE

<222>(190)..(190)

<223>Xaa 是非芳族残基

<220>

<221>MISC_FEATURE

<222>(196)..(196)

<223>Xaa 是脂肪族、非极性或芳族残基

<220>

<221>MISC_FEATURE

<222>(202)..(202)

<223>Xaa 是芳族、非极性或脂肪族残基

<220>

<221>MISC_FEATURE

<222>(205)..(205)

<223>Xaa 是碱性、非极性或脂肪族残基

<220>

<221>MISC_FEATURE

<222>(206)..(206)

<223>Xaa 是芳族或非极性残基

<400>95

Met	Ser	Asn	Arg	Leu	Asp	Xaa	Lys	Val	Ala	Ile	Ile	Thr	Gly	Gly	Xaa
1				5					10					15	
Leu	Gly	Ile	Gly	Leu	Ala	Ile	Ala	Thr	Lys	Phe	Val	Glu	Glu	Gly	Ala
				20				25						30	
Lys	Val	Met	Ile	Thr	Gly	Arg	His	Ser	Asp	Xaa	Gly	Glu	Lys	Ala	Ala
		35					40					45			
Lys	Ser	Val	Gly	Thr	Pro	Asp	Gln	Ile	Gln	Phe	Xaa	Gln	His	Asp	Ser
		50				55					60				
Ser	Asp	Glu	Asp	Gly	Trp	Thr	Lys	Leu	Phe	Asp	Ala	Thr	Glu	Lys	Ala
65				70					75					80	
Phe	Gly	Pro	Val	Ser	Thr	Leu	Val	Asn	Asn	Ala	Gly	Ile	Xaa	Xaa	Xaa
				85					90					95	
Xaa	Ser	Val	Glu	Glu	Thr	Thr	Thr	Ala	Glu	Trp	Arg	Lys	Leu	Leu	Ala
			100					105						110	
Val	Asn	Leu	Asp	Gly	Val	Phe	Xaa	Gly	Thr	Arg	Leu	Xaa	Ile	Gln	Arg

115	120	125
Met Lys Asn Lys Gly Leu Gly Ala Ser Ile Ile Asn Met Xaa Ser Ile		
130	135	140
Glu Gly Xaa Val Xaa Xaa Pro Xaa Leu Gly Ala Tyr Asn Ala Ser Lys		
145	150	155
Gly Ala Val Arg Ile Met Ser Lys Ser Ala Ala Leu Asp Cys Ala Leu		
165	170	175
Lys Asp Tyr Asp Val Arg Val Asn Thr Val His Pro Gly Xaa Ile Lys		
180	185	190
Thr Pro Leu Xaa Asp Asp Leu Pro Gly Xaa Glu Glu Xaa Xaa Ser Gln		
195	200	205
Arg Thr Lys Thr Pro Met Gly His Ile Gly Glu Pro Asn Asp Ile Ala		
210	215	220
Tyr Ile Cys Val Tyr Leu Ala Ser Asn Glu Ser Lys Phe Ala Thr Gly		
225	230	235
Ser Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln Ter		
245	250	

<210>96

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 具有克菲尔乳杆菌 (L. kefir) 主链的工程酮还原酶序列式

<223> 合成构建体

<220>

<221>MISC_FEATURE

<222> (7).. (7)

<223>Xaa 是极性、非极性、酸性、芳族、受限或碱性残基

<220>

<221>MISC_FEATURE

<222> (16).. (16)

<223>Xaa 是极性残基

<220>

<221>MISC_FEATURE

<222> (43).. (43)

<223>Xaa 是非极性或极性残基

<220>

<221>MISC_FEATURE

<222> (60).. (60)

<223>Xaa 是芳族、非极性或脂肪族残基

<220>

<221>MISC_FEATURE

<222> (94).. (94)

<223>Xaa 是脂肪族、非极性或半胱氨酸残基

<220>

<221>MISC_FEATURE

<222> (95).. (95)

<223>Xaa 是非极性或脂肪族残基

<220>

<221>MISC_FEATURE

<222> (96).. (96)

<223>Xaa 是酸性或极性残基

<220>

<221>MISC_FEATURE

<222> (97).. (97)

<223>Xaa 是碱性、脂肪族、非极性或极性残基

<220>

<221>MISC_FEATURE

<222> (120).. (120)

<223>Xaa 是芳族、脂肪族或非极性残基

<220>

<221>MISC_FEATURE

<222> (125).. (125)

<223>Xaa 是极性或非极性残基

<220>

<221>MISC_FEATURE

<222>(142)..(142)

<223>Xaa 是极性残基

<220>

<221>MISC_FEATURE

<222>(147)..(147)

<223>Xaa 是芳族、脂肪族、极性或非极性残基

<220>

<221>MISC_FEATURE

<222>(149)..(149)

<223>Xaa 是非极性或非芳族残基

<220>

<221>MISC_FEATURE

<222>(150)..(150)

<223>Xaa 是酸性或受限残基

<220>

<221>MISC_FEATURE

<222>(152)..(152)

<223>Xaa 是极性或非极性残基

<220>

<221>MISC_FEATURE

<222>(190)..(190)

<223>Xaa 是非芳族残基

<220>

<221>MISC_FEATURE

<222>(196)..(196)

<223>Xaa 是脂肪族、非极性或非芳族残基

<220>

<221>MISC_FEATURE

<222>(202)..(202)

<223>Xaa 是芳族、非极性或非脂肪族残基

<220>

<221>MISC_FEATURE

<222>(205)..(205)

<223>Xaa 是碱性、非极性或脂肪族残基

<220>

<221>MISC_FEATURE

<222>(206)..(206)

<223>Xaa 是芳族或非极性残基

<400>96

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

Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly
 225 230 235 240
 Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln
 245 250

<210>97

<211>759

<212>DNA

<213> 小乳杆菌 (Lactobacillus Minor)

<400>97

atgaccgatc ggttgaaggg gaaagtagca attgtaactg gcggtacctt gggaattggc 60
 ttggcaatcg ctgataagtt tgttgaagaa ggcgcaaagg ttgttattac cggccgtcac 120
 gctgatgtag gtgaaaaagc tgccagatca atcggcggca cagacgttat ccgttttgtc 180
 caacacgatg cttctgatga aaccggctgg actaagttgt ttgatacgac tgaagaagca 240
 tttggcccag ttaccacggt tgtcaacaat gccggaattg cggtcagcaa gagtgttgaa 300
 gataccacaa ctgaagaatg gcgcaagctg ctctcagtta acttggatgg tgtcttcttc 360
 ggtacccgtc ttggaatcca acgtatgaag aataaaggac tcggagcatc aatcatcaat 420
 atgtcatcta tcgaagggtt tgttgggtgat ccagctctgg gtgcatacaa cgcttcaaaa 480
 ggtgctgtca gaattatgtc taaatcagct gccttggatt gcgctttgaa ggactacgat 540
 gttcgggtta acactgttca tccaggttat atcaagacac cattgggttga cgatcttgaa 600
 ggggcagaag aaatgatgtc acagcggacc aagacaccaa tgggtcatat cggatgaacct 660
 aacgatatcg cttggatctg tgtttacctg gcactctgacg aatctaaatt tgccactggt 720
 gcagaattcg ttgtcgacgg aggggtacacc gcccaatag 759

<210>98

<211>252

<212>PRT

<213> 小乳杆菌 (Lactobacillus Minor)

<400>98

Met Thr Asp Arg Leu Lys Gly Lys Val Ala Ile Val Thr Gly Gly Thr
 1 5 10 15
 Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala
 20 25 30
 Lys Val Val Ile Thr Gly Arg His Ala Asp Val Gly Glu Lys Ala Ala
 35 40 45
 Arg Ser Ile Gly Gly Thr Asp Val Ile Arg Phe Val Gln His Asp Ala
 50 55 60

Ser Asp Glu Thr Gly Trp Thr Lys Leu Phe Asp Thr Thr Glu Glu Ala			
65	70	75	80
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Ala Val Ser			
	85	90	95
Lys Ser Val Glu Asp Thr Thr Thr Glu Glu Trp Arg Lys Leu Leu Ser			
	100	105	110
Val Asn Leu Asp Gly Val Phe Phe Gly Thr Arg Leu Gly Ile Gln Arg			
	115	120	125
Met Lys Asn Lys Gly Leu Gly Ala Ser Ile Ile Asn Met Ser Ser Ile			
	130	135	140
Glu Gly Phe Val Gly Asp Pro Ala Leu Gly Ala Tyr Asn Ala Ser Lys			
145	150	155	160
Gly Ala Val Arg Ile Met Ser Lys Ser Ala Ala Leu Asp Cys Ala Leu			
	165	170	175
Lys Asp Tyr Asp Val Arg Val Asn Thr Val His Pro Gly Tyr Ile Lys			
	180	185	190
Thr Pro Leu Val Asp Asp Leu Glu Gly Ala Glu Glu Met Met Ser Gln			
	195	200	205
Arg Thr Lys Thr Pro Met Gly His Ile Gly Glu Pro Asn Asp Ile Ala			
	210	215	220
Trp Ile Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly			
225	230	235	240
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln			
	245	250	

<210>99

<211>1011

<212>DNA

<213> 人工序列

<220>

<223> 来自近平滑假丝酵母 (*C. parapsilosis*) 的 SADH

<400>99

```

atgtcaattc catcaagcca gtacggattc gtattcaata agcaatcagg acttaatttg 60
agaaatgatt tgcctgtcca caagcccaaa gcggggtcaat tgttggtgaa agttgatgct 120
gttggattgt gtcattctga tttacatgtc atttacgaag ggttggattg tggtgataat 180
tatgtcatgg gacatgaaat tgctggaact gttgctgctg tgggtgatga tgtcattaac 240
tacaaggttg gtgatcgtgt tgcctgtgtc ggacccaatg gatgtggtgg gtgcaagtat 300
tgtcgtgggtg ccattgacaa tgtatgtaaa aacgcatttg gtgattgggt cggattgggg 360

```



```

tacgatggtg ggtatcaaca gtacttgttg gttactagac cacgtaactt gtctcgtatc 420
ccagataacg tatctgcaga cgtggctgcg gcttcaactg atgctgtatt gacaccatat 480
cacgcaatca agatggctca agtgtcacca acttcgaata tcttgcttat tggtgctggt 540
ggattgggtg gaaatgcaat tcaagttgcc aaggcatttg gtgcgaaagt tactgttttg 600
gacaaaaaaaa aggaggctcg tgaccaagca aagaagttgg gtgctgatgc agtttatgaa 660
acattgccag aatccatttc tctggtctct ttttcagcat gttttgattt tgtttcagtg 720
caagctacat ttgatgtatg tcaaaagtat gttgaaccaa aggggtgtaat tatgcccgtg 780
ggactcgggtg ctcttaattt atcgtttaat ttgggagatt tggcattgag agaaattcga 840
atcttgggta gtttttgggg aactactaat gatttggatg atgttttgaa attggttagt 900
gaaggtaaag ttaaaccctg tgtgagaagt gccaaattga aggaattgcc agagtatatt 960
gaaaaattga gaaacaatgc ttatgaaggt agagttgttt ttaatccata g          1011

```

<210>100

<211>336

<212>PRT

<213> 人工序列

<220>

<223> 来自近平滑假丝酵母 (*C. parapsilosis*) 的 SADH

<400>100

```

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

```

atgaaggcac tgcagtaccg caccatcggc gccccgcccc aggtcgtcac cgtcccggac 60
ccggagccgg gccccggcca ggtgctgttg aaggtgaccg cggccggagt ctgccactcc 120
gacatcgcgg tgatgagctg gcccgccgag ggcttcccg t acgagctgcc gctcacccctc 180
ggccacgagg gcgtcggcac cgtggccgcg ctccggcgccg gggtgacggg gctcgccgag 240
ggcgacgcgg tcgccgtgta cgggccttgg ggctgcggca cctgcgccaa gtgcgcggag 300

```

ggcaaggaga actactgcct gcgcgccgac gagctgggca tccgtccgcc ggggctcggg 360
cgctccgggt ccatggccga gtacctgtctg atcgacgacc cccggcacct ggtcccgtg 420
gacgggctcg acccggtcgc ggcggtgccg ctcaccgacg ccggactgac gccgtaccac 480
gcatcaage ggtegttgc caagctggtc cccggtcca ccgcggtggt catcggcacc 540
ggtggtctcg gccacgtcgc catccagctg ctgcgcgcc tgacgtccgc ccgggtggtc 600
gccctggacg tcagegagga gaagctgcgc ctgcccgtg cggtgggcgc gcacgaggcg 660
gtgctgtcgg acgcgaaggc cgcggacgcg gtgcgcgaga tcaccggcgg tctcgtgcc 720
gaggccgtgt tcgacttcgt cggcgtggcg cccaccgtgc agaccgccgg agccgtcgcg 780
gccgtcgagg gcgatgtcac cctggtcggc atcggcggcg gatcgtgcc cgtcggcttc 840
ggcatgctgc cgttcgaggt gtcggtcaac gccccctact ggggcagccg cagcgagctg 900
accgaggtgc tgaacctggc ccgctccggt gccgtgtcgg tgcacaccga gacgtactcc 960
ctggacgacg ccccgctcgc ctacgagcgg ctgcacgagg gcagggtcaa cggccgcgcg 1020
gtgacctgc cccacggctg a 1041

```

<210>102

<211>346

<212>PRT

<213> 人工序列

<220>

<223> 来自天蓝色链霉菌 (S. coelicolor) 的 ADH

<400>102

```

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

```

Leu	Leu	Ile	Asp	Asp	Pro	Arg	His	Leu	Val	Pro	Leu	Asp	Gly	Leu	Asp
130						135					140				
Pro	Val	Ala	Ala	Val	Pro	Leu	Thr	Asp	Ala	Gly	Leu	Thr	Pro	Tyr	His
145					150					155				160	
Ala	Ile	Lys	Arg	Ser	Leu	Pro	Lys	Leu	Val	Pro	Gly	Ser	Thr	Ala	Val
			165					170					175		
Val	Ile	Gly	Thr	Gly	Gly	Leu	Gly	His	Val	Ala	Ile	Gln	Leu	Leu	Arg
		180						185				190			
Ala	Leu	Thr	Ser	Ala	Arg	Val	Val	Ala	Leu	Asp	Val	Ser	Glu	Glu	Lys
	195					200					205				
Leu	Arg	Leu	Ala	Arg	Ala	Val	Gly	Ala	His	Glu	Ala	Val	Leu	Ser	Asp
	210					215					220				
Ala	Lys	Ala	Ala	Asp	Ala	Val	Arg	Glu	Ile	Thr	Gly	Gly	Leu	Gly	Ala
225				230					235					240	
Glu	Ala	Val	Phe	Asp	Phe	Val	Gly	Val	Ala	Pro	Thr	Val	Gln	Thr	Ala
			245						250				255		
Gly	Ala	Val	Ala	Ala	Val	Glu	Gly	Asp	Val	Thr	Leu	Val	Gly	Ile	Gly
		260						265					270		
Gly	Gly	Ser	Leu	Pro	Val	Gly	Phe	Gly	Met	Leu	Pro	Phe	Glu	Val	Ser
	275					280					285				
Val	Asn	Ala	Pro	Tyr	Trp	Gly	Ser	Arg	Ser	Glu	Leu	Thr	Glu	Val	Leu
	290					295					300				
Asn	Leu	Ala	Arg	Ser	Gly	Ala	Val	Ser	Val	His	Thr	Glu	Thr	Tyr	Ser
305				310						315				320	
Leu	Asp	Asp	Ala	Pro	Leu	Ala	Tyr	Glu	Arg	Leu	His	Glu	Gly	Arg	Val
			325						330				335		
Asn	Gly	Arg	Ala	Val	Ile	Leu	Pro	His	Gly						
	340							345							

<210>103

<211>1032

<212>DNA

<213> 人工序列

<220>

<223> 密码子优化的来自赭色掷孢酵母 (*S. salmonicolor*) 的 ADH

<400>103

atggcaaaga tcgacaacgc agttctgccg gagggttctc tgggtgctggt caccggcgcg 60

```

aacggctttg tcgctagcca tgtggtegaa caactgctgg aacacggcta taaggtgcgc 120
ggcactgctc gctctgcctc caaactggcg aacctgcaga aacgttggga cgccaaatac 180
cctggtcgtt tcgagactgc cgttgttgaa gacatgctga agcaggggtgc atatgatgaa 240
gttattaaag gcgcggcagg tgtegeccac atcgcgccg tggtcagctt ttctaacaaa 300
tatgatgagg tggttaactc tgcgateggt ggcacgctga atgccctgcg tgccgcagct 360
gctacgcctt ccgtgaaacg ttttgtgctg accagcagca ctgtttctgc actgattcca 420
aaacctaacg tcgaaggtat ttatctggat gagaagagct ggaacctgga aagcattgat 480
aaggctaaaa ccctgcctga atctgatccg cagaaaagcc tgtgggtcta cgccgcaagc 540
aaaacggaag cggaactggc tgcctggaaa ttcattggacg aaaacaaacc gcactttact 600
ctgaatgccg ttctgccaaa ctacactatc ggtaccatit ttgaccaga aaccaatcc 660
ggttccactt ccggctggat gatgtctctg ttcaatggcg aagtatctcc ggcactggcg 720
ctgatgccgc cgcagtacta tgtctctgca gttgatatcg gtctgctgca cctgggttgt 780
ctggttctgc cgcaaatcga acgcccgtcg gtttacggca ccgcaggcac ctttgattgg 840
aacaccgttc tggcgacctt ccgtaaactg tatccgtcca agacgttccc ggctgacttt 900
ccggatcagg gccaggatct gtccaaatit gataccgcc cgagcctgga gattctgaaa 960
tccttgggcc gccctggctg gcgtagcatc gaggaatcta tcaaagatct ggtgggttcc 1020
gagaccgcct aa 1032

```

<210>104

<211>343

<212>PRT

<213> 人工序列

<220>

<223> 来自赭色掷孢酵母 (*S. salmonicolor*) 的 ADH

<400>104

```

Met Ala Lys Ile Asp Asn Ala Val Leu Pro Glu Gly Ser Leu Val Leu
1           5           10           15
Val Thr Gly Ala Asn Gly Phe Val Ala Ser His Val Val Glu Gln Leu
20           25           30
Leu Glu His Gly Tyr Lys Val Arg Gly Thr Ala Arg Ser Ala Ser Lys
35           40           45
Leu Ala Asn Leu Gln Lys Arg Trp Asp Ala Lys Tyr Pro Gly Arg Phe
50           55           60
Glu Thr Ala Val Val Glu Asp Met Leu Lys Gln Gly Ala Tyr Asp Glu
65           70           75           80
Val Ile Lys Gly Ala Ala Gly Val Ala His Ile Ala Ser Val Val Ser
85           90           95

```

Phe Ser Asn Lys Tyr Asp Glu Val Val Thr Pro Ala Ile Gly Gly Thr		
100	105	110
Leu Asn Ala Leu Arg Ala Ala Ala Thr Pro Ser Val Lys Arg Phe		
115	120	125
Val Leu Thr Ser Ser Thr Val Ser Ala Leu Ile Pro Lys Pro Asn Val		
130	135	140
Glu Gly Ile Tyr Leu Asp Glu Lys Ser Trp Asn Leu Glu Ser Ile Asp		
145	150	155
Lys Ala Lys Thr Leu Pro Glu Ser Asp Pro Gln Lys Ser Leu Trp Val		
165	170	175
Tyr Ala Ala Ser Lys Thr Glu Ala Glu Leu Ala Ala Trp Lys Phe Met		
180	185	190
Asp Glu Asn Lys Pro His Phe Thr Leu Asn Ala Val Leu Pro Asn Tyr		
195	200	205
Thr Ile Gly Thr Ile Phe Asp Pro Glu Thr Gln Ser Gly Ser Thr Ser		
210	215	220
Gly Trp Met Met Ser Leu Phe Asn Gly Glu Val Ser Pro Ala Leu Ala		
225	230	235
Leu Met Pro Pro Gln Tyr Tyr Val Ser Ala Val Asp Ile Gly Leu Leu		
245	250	255
His Leu Gly Cys Leu Val Leu Pro Gln Ile Glu Arg Arg Arg Val Tyr		
260	265	270
Gly Thr Ala Gly Thr Phe Asp Trp Asn Thr Val Leu Ala Thr Phe Arg		
275	280	285
Lys Leu Tyr Pro Ser Lys Thr Phe Pro Ala Asp Phe Pro Asp Gln Gly		
290	295	300
Gln Asp Leu Ser Lys Phe Asp Thr Ala Pro Ser Leu Glu Ile Leu Lys		
305	310	315
Ser Leu Gly Arg Pro Gly Trp Arg Ser Ile Glu Glu Ser Ile Lys Asp		
325	330	335
Leu Val Gly Ser Glu Thr Ala		
340		

<210>105

<211>1125

<212>DNA

<213> 人工序列

<220>

<223> 密码子优化的来自马肝的 ADH, E 型

<400>105

```

atgagcacag cagggaaagt aattaaatgc aaagcagctg tgctgtggga ggaaaagaaa 60
ccattttcca ttgaggaggt ggaggtegca cctccgaagg cccatgaagt ccgtattaag 120
atgggtggcca cagggatctg tcgctcagat gaccacgtgg ttagtgggac cctggtcaca 180
cctctgcctg tgategcagg ccatgaggca gctggcattg tggagagcat cggggaaggc 240
gtcactacag tacgtccagg tgataaagtc atcccactgt ttactcctca gtgtgggaaa 300
tgccgtgttt gtaaacaccc tgaaggcaac ttctgcttga aaaatgatct gagcatgcct 360
cgtgggacca tgcaggatgg taccagccgt ttcacctgcc gtgggaagcc tatccaccac 420
ttcctgggca ccagcacctt ctcccagtac accgtgggtg acgagatctc agtggccaag 480
atcgatgcgg cctcaccgct ggagaaagtc tgtctgattg gctgtgggtt ttctactggt 540
tatgggtctg cagtcaaggt tgccaaggtc acccagggtt ccacctgtgc cgtgtttggc 600
ctgggggggg tgggcctgtc tgttatcatg ggctgtaaag cagccggggc ggcccgtatc 660
attggggtgg acatcaacaa agacaagttt gcaaaggcca aagaagtggg tgccactgag 720
tgtgtcaacc ctcaggacta caagaaacct atccaggagg tgctgacaga aatgagcaat 780
gggggcgtgg atttttcatt tgaagtcatt ggtcgtctgg acaccatggt gactgccttg 840
tcatgtctgc aagaagcata tgggtgtgag gtcacgtag gggtagctcc tgattcccaa 900
aatctgtcta tgaatcctat gttgtgtctg agtgggcgta cctggaaagg ggctatTTTT 960
ggtggtttta agagtaaaga ttctgtccct aaactgggtg ccgattttat ggctaaaaag 1020
tttgcactgg atcctttaat caccatggtt ttaccttttg aaaaaattaa tgaagggttt 1080
gacctgtgc gctctgggga gagtatccgt accatcctga cgttt 1125

```

<210>106

<211>375

<212>PRT

<213> 人工序列

<220>

<223> 来自马肝的 ADH, E 型

<400>106

```

Met Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
1           5           10          15
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
          20          25          30
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
          35          40          45
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val

```

50	55	60
Ile Ala Gly His Glu	Ala Ala Gly Ile Val	Glu Ser Ile Gly Glu Gly
65	70	75
Val Thr Thr Val Arg	Pro Gly Asp Lys Val	Ile Pro Leu Phe Thr Pro
85	90	95
Gln Cys Gly Lys Cys	Arg Val Cys Lys His	Pro Glu Gly Asn Phe Cys
100	105	110
Leu Lys Asn Asp Leu	Ser Met Pro Arg Gly	Thr Met Gln Asp Gly Thr
115	120	125
Ser Arg Phe Thr Cys	Arg Gly Lys Pro Ile	His His Phe Leu Gly Thr
130	135	140
Ser Thr Phe Ser Gln	Tyr Thr Val Val Asp	Glu Ile Ser Val Ala Lys
145	150	155
Ile Asp Ala Ala Ser	Pro Leu Glu Lys Val	Cys Leu Ile Gly Cys Gly
165	170	175
Phe Ser Thr Gly Tyr	Gly Ser Ala Val Lys	Val Ala Lys Val Thr Gln
180	185	190
Gly Ser Thr Cys Ala	Val Phe Gly Leu Gly	Gly Val Gly Leu Ser Val
195	200	205
Ile Met Gly Cys Lys	Ala Ala Gly Ala Ala	Arg Ile Ile Gly Val Asp
210	215	220
Ile Asn Lys Asp Lys	Phe Ala Lys Ala Lys	Glu Val Gly Ala Thr Glu
225	230	235
Cys Val Asn Pro Gln	Asp Tyr Lys Lys Pro	Ile Gln Glu Val Leu Thr
245	250	255
Glu Met Ser Asn Gly	Gly Val Asp Phe Ser	Phe Glu Val Ile Gly Arg
260	265	270
Leu Asp Thr Met Val	Thr Ala Leu Ser Cys	Cys Gln Glu Ala Tyr Gly
275	280	285
Val Ser Val Ile Val	Gly Val Pro Pro Asp	Ser Gln Asn Leu Ser Met
290	295	300
Asn Pro Met Leu Leu	Leu Ser Gly Arg Thr	Trp Lys Gly Ala Ile Phe
305	310	315
Gly Gly Phe Lys Ser	Lys Asp Ser Val Pro	Lys Leu Val Ala Asp Phe
325	330	335
Met Ala Lys Lys Phe	Ala Leu Asp Pro Leu	Ile Thr His Val Leu Pro
340	345	350
Phe Glu Lys Ile Asn	Glu Gly Phe Asp Leu	Leu Arg Ser Gly Glu Ser
355	360	365

Ile Arg Thr Ile Leu Thr Phe
370 375

<210>107

<211>1056

<212>DNA

<213> 人工序列

<220>

<223> 密码子优化的来自布氏嗜热厌氧菌 (*T. Brockii*) 的 ADH-TB

<400>107

```
atgaaaggct tgcgatgct gagcatcggc aaagtgggtt ggattgaaaa agaaaaaccg 60
gcgccaggcc cgttcgatgc aattgtgcgc cctctggcag tagcgccgtg taccagcgat 120
attcatactg tgtttgaagg tgccattggc gagcgtcaca atatgattct gggccatgaa 180
gccgttggtg aagtgtgtga ggttggcagc gaagtgaagg atttcaaacc gggcgatcgc 240
gttgtcgttc cagcgattac cccggattgg cgcaccagcg aagtcagcg cggctacat 300
cagcactctg gcggcatgct ggccggctgg aaattcagca atgtaaagga tgggtgtgtc 360
ggtgaatttt ttacgttaa cgacgcagac atgaatctgg cgcacctgcc gaaagaaatc 420
ccgctggaag cagcggttat gattccggat atgatgacca cgggttttca cggcgcagag 480
ctggcggaca ttgaactggg cgctacggta gccgtactgg gcatcggtcc ggtgggcctg 540
atggcagttg caggcgctaa gctgcgcggc gcaggctgta ttattgccgt tggttctcgc 600
ccggtgtgtg tggacgccgc taagtattat ggtgcaacgg acattgtcaa ttacaaggac 660
ggcccaattg aatctcagat catgaacctg acggaaggta aaggcgttga cgccgcgatt 720
atcgctggcg gcaacgccga catcatggcg accgcagtta aaatcgtcaa gccaggtggt 780
actattgcta acgtcaacta cttcggcgaa ggtgaggctc tgcctgtccc acgtctggaa 840
tggggttgcg gtatggcaca taaaaccatt aaaggtggcc tgtgcccagg cggccgtctg 900
cgtatggaac gcctgatcga tctggtcttc tacaaacgcg tggatcctag caaactggtg 960
actcacgttt tccgcggctt tgataacatc gaaaaagctt ttatgctgat gaaagataaa 1020
ccgaaagatc tgattaaacc ggttgtcatc ctggct 1056
```

<210>108

<211>352

<212>PRT

<213> 人工序列

<220>

<223> 来自布氏嗜热厌氧菌 (*T. Brockii*) 的 ADH-TB

<400>108

```

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

```


<223> 来自酿酒酵母 (*Saccharomyces cerevisiae*) 的 YDL ADH

<400>110

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

Leu Glu His Glu Pro Leu Arg Leu Tyr Trp Asn Lys Leu Tyr Gly Lys
 290 295 300
 Tyr Asn Tyr Ala Ala Gln Lys Val
 305 310

<210>111

<211>1047

<212>DNA

<213> 人工序列

<220>

<223> 来自红平红球菌 (*Rhodococcus erythropolis*) 的 ADH

<400>111

```
atgaaagcca ttcagtacac tcgtatcggg gcggaaccag aactgactga aatccccgaag 60
ccggaaccgg gcccgggcga agtactgctg gaagtcacgg cagctggcgt gtgccattcc 120
gatgatttca ttatgtctct gccggaagaa cagtacacct acggcctgcc gctgaccctg 180
ggatcatgaag gtgctggtaa agttgccgca gttggcgaag gtgttgaagg gttggatatt 240
ggcaccaatg tggttgtgta cggcccatgg ggttgtggca actgttggca ttgcagtcag 300
ggcctggaga actattgctc ccgtgcgcag gaactgggta ttaacccgcc tggctcgggt 360
gctccggggg ctttggcaga atttatgatt gtgcactcac cagtcattt ggtcccgatt 420
ggcgatttag accctgttaa aactgttccg ttgactgatg cgggcctgac ccataccat 480
gcaatcaaac gtcacctgcc gaaactgcgc ggcggtcttt atgcagtagt gatcggtagc 540
ggtggccttg gccacgtggc tatccaactg ctgcgtcatt tatctgctgc aacggtgac 600
gccttggacg tttctgccga taaactggaa ctggctacca aagtcggcgc acatgaagta 660
gtcctgtctg ataaagatgc agcggagaat gtgcgtaaaa ttactggtag ccaagtgca 720
gctttggtgt tggattttgt gggctatcag cctaccattg acaccgcat ggcagtggcg 780
ggcgtgggct ctgacgtcac cattgttggg atcggtgatg gccaggcaca tgcgaaagtt 840
ggtttcttcc agagtcctta tgaggcatcg gttacggtag cttattgggg cgctcgtaat 900
gaactgatcg aattgatcga tctggcgcat gctgggtattt tcgacattgc cgttgagacc 960
ttctcttttg ataatggtgc agaggcctat cgtcgcctgg ctgcgggcac actgtcaggc 1020
cgtgcggtag tcgtcccggg cctgtaa 1047
```

<210>112

<211>348

<212>PRT

<213> 人工序列

<220>

<223> 来自红平红球菌 (*Rhodococcus erythropolis*) 的 ADH

<400>112

```

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

```

290	295	300	
Leu Ile Asp Leu Ala His Ala Gly Ile Phe Asp Ile Ala Val Glu Thr			
305	310	315	320
Phe Ser Leu Asp Asn Gly Ala Glu Ala Tyr Arg Arg Leu Ala Ala Gly			
	325	330	335
Thr Leu Ser Gly Arg Ala Val Val Val Pro Gly Leu			
340	345		

<210>113

<211>1047

<212>DNA

<213> 人工序列

<220>

<223> 来自酿酒酵母 (*Saccharomyces cerevisiae*) 的 YGL ADH

<400>113

```

atgaccacgg aaaaaaccgt agtgttcgtg tcaggcgcca ccggttttat tgctctgcac 60
gtggtagatg acctgctgaa aactgggttac aaagtaattg gttccggctg ttctcaggaa 120
aaaaatgacg gtttgctgaa gaagttcaag tccaaccgca atctgagcat ggaaattgtg 180
gaagatatgg cggcaccaaa cgccttcgat aaagtattcc agaaacatgg taaagaaatt 240
aaagtgggtc tgcataatgc gtccccgggc catttcaaca ctaccgattt cgaaaaagac 300
ttactgatcc cggcggtaaa cggtaccaaa tctattttgg aagcaattaa gaactatgcc 360
gcagacaccg tggaaaaagt gggtattact tcatctgttg ccgcgttggc ctctccgggt 420
gatatgaaag ataccagctt cgtgggttac gaagaatcct ggaataaaga cacctgggaa 480
tcgtgtcagg cgaatgctgt gtccgcttat tgcggttcta aaaaattcgc agagaaaacg 540
gcgtgggact tcttgaaga aaaccagagc agcattaaat ttactctgtc cacgattaac 600
ccaggcttcg tttttgggtc gcagctgttc gccgactcct tgcgcaatgg tattaactct 660
agcagtgcga ttattgcgaa cctgggtgtc tataaattag gggataactt ctacaattat 720
agcggccccg ttatcgacgt ccgtgacgtt tccaaagctc atctgctggc atttgagaaa 780
cctgaatgcg ccggtcagcg cctgtttctg tgcgaggata tggtctgttc ccaggaagcc 840
ctggacattc tgaacgaaga atttccacag ctgaagggca agatcgcaac gggcgaacct 900
ggcagcggct cgaccttcct gactaaaaat tgttgcaaat gcgacaatcg taaaactaaa 960
aacttgctgg gcttccagtt caacaaattc cgcgactgca ttgtcgatac tgcgtcccag 1020
ttgctggaag tgcaaagcaa aagctaa                                     1047

```

<210>114

<211>348

<212>PRT

<213> 人工序列

<220>

<223> 来自酿酒酵母 (*saccharomyces cervisiae*) 的 YGL ADH

<400>114

```

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

```


260	265	270
Asp Met Phe Cys Ser Gln Glu Ala Leu Asp Ile Leu Asn Glu Glu Phe		
275	280	285
Pro Gln Leu Lys Gly Lys Ile Ala Thr Gly Glu Pro Gly Ser Gly Ser		
290	295	300
Thr Phe Leu Thr Lys Asn Cys Cys Lys Cys Asp Asn Arg Lys Thr Lys		
305	310	315
Asn Leu Leu Gly Phe Gln Phe Asn Lys Phe Arg Asp Cys Ile Val Asp		
325	330	335
Thr Ala Ser Gln Leu Leu Glu Val Gln Ser Lys Ser		
340	345	

<210>115

<211>939

<212>DNA

<213> 人工序列

<220>

<223> 来自酿酒酵母 (*Saccharomyces cerevisiae*) 的 YPR ADH

<400>115

```

atgccggcaa cgtaaataaa cagcagtgtc accttaaaat taaacacagg tgcgagcatt 60
cctgtcctgg ggttcggcac ctggcgctct gtcgataaca acggctatca tagtgtaatt 120
gcggcgctga aagcggggta ccgtcatatc gatgctgcgg ccatctatct gaatgaagaa 180
gaagtcggcc gtgcgatcaa ggactccggt gttcctcgtg aagaaatfff tattaccacc 240
aaactgtggg gcaccgaaca acgcgatcca gaagcagccc tgaacaaatc tctgaaacgt 300
ctgggtctgg actatgtgga cctgtatctg atgcactggc cggccctctt gaaaacagac 360
cgtgtaactg acggtaacgt cctgtgcatc ccgaccctgg aagatggcac cgtggacatc 420
gataccaaag agtggaaatf tattaaaacc tgggaactga tgcaggaatt gccgaaaact 480
ggtaagacca aagccgtcgg tgtgtccaat tttccatca acaatatcaa agaactgtg 540
gaatcgccaa ataacaaggt cgttccagca accaatcaga tcgagattca tccgttgtg 600
ccgcaggatg aattaatcgc cttttgtaaa gaaaaaggca ttgtggtcga agcatatagc 660
ccattcggct ccgctaacgc cccgtgtctg aaagaacagg cgattatcga tatggccaaa 720
aagcacggcg tcgaaccggc gcaactgatt atcagctggc cgattcagcg cggttatgtg 780
gtattggcca agtccgtaaa tccggagcgt atcgtgtcga actttaagat ttttaccctg 840
ccagaggatg atttcaaaac catctctaac ctgagcaaag tgcacggtag caaacgtgtc 900
gttgacatga aatggggctc atttccgatt tttcaataa 939

```

<210>116

<211>312

<212>PRT

<213> 人工序列

<220>

<223> 来自酿酒酵母 (*Saccharomyces cerevisiae*) 的 YPR ADH

<400>116

```

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

```

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

<210>117

<211>1029

<212>DNA

<213> 人工序列

<220>

<223> 来自酿酒酵母 (*Saccharomyces cerevisiae*) 的 GRE ADH

<400>117

```

atgtctgtgt tcgtgtcagg cgcgaaatggt tttattgctc agcacatcgt agatctgctg 60
ctgaaagaag attacaaagt aattgggttcc gcacgttctc aggaaaaagc tgaaaatttg 120
accgaagcct tcggtaacaa cccgaaatit agcatggaag tggcgcctga tattagcaaa 180
ctggatgcct tcgatcatgt attccagaaa catggtaaag atattaaaaat cgtcctgcat 240
accgcgtccc cgtttttgttt cgatattacc gattccgaac gtgacttact gatccccggcg 300
gtaaacgggtg tcaaaggtat tttgcacagt attaagaaat atgccgcaga cagcgtggaa 360
cgcggtggttc tgacttcate ttacgccgcg gtatttgata tggcgaagga aaacgataag 420
agcctgacct tcaacgaaga atcctggaat ccggcgacct gggaatcgtg tcagagtgat 480
ccggtgaacg cttattgcgg ttctaaaaaa ttgcagaga aagcagcgtg ggaattcttg 540
gaagaaaacc gtgatagcgt gaaatttgag ctgacagcgg tcaaccaggt ttacgttttt 600
ggtccgcaga tgttcgataa agatgttaaa aaacacttga acaccagctg cgaactgggtg 660
aactctctga tgcatctgag ccctgaagat aaaattccgg aactgttttg cggttacatc 720
gacgtccgtg acgttgcgaa agctcatctg gttgcatttc agaaacgtga aacaatcggt 780
cagcgcctga tcgtgtcgga ggcacgtttc acgatgcagg acgttctgga cattctgaac 840
gaagactttc cagtactgaa gggcaatata ccggtcggca agcctggcag cggcgccacc 900
cataatactc tgggcgccac cctggacaat aaaaaaagca aaaaattgct gggttcaaaa 960
ttccgtaate tgaaagagac tattgacgat actgcgtccc agatcctgaa attcgaaggt 1020
cgcattttaa                                     1029

```

<210>118

<211>342

<212>PRT

<213> 人工序列

<220>

<223> 来自酿酒酵母 (*Saccharomyces cerevisiae*) 的 GRE ADH

<400>118

```

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

```

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

<210>119

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 具有小乳杆菌 (*L. minor*) 主链的工程酮还原酶序列式

<223> 合成构建体

<220>

<221>MISC_FEATURE

<222> (7).. (7)

<223>Xaa 是极性、非极性、酸性、芳族、受限或碱性残基

<220>

<221>MISC_FEATURE

<222> (16).. (16)

<223>Xaa 是极性残基

<220>

<221>MISC_FEATURE

<222> (43).. (43)

<223>Xaa 是非极性或非极性残基

<220>

<221>MISC_FEATURE

<222> (60).. (60)

<223>Xaa 是芳族、非极性或脂肪族残基

<220>

<221>MISC_FEATURE

<222> (94).. (94)

<223>Xaa 是脂肪族、非极性或非胱氨酸残基

<220>

<221>MISC_FEATURE

<222> (95).. (95)

<223>Xaa 是非极性或非脂肪族残基

<220>

<221>MISC_FEATURE

<222> (96).. (96)

<223>Xaa 是酸性或极性残基

<220>

<221>MISC_FEATURE

<222> (97).. (97)

<223>Xaa 是碱性、脂肪族、非极性或非极性残基

<220>

<221>MISC_FEATURE

<222> (120).. (120)

<223>Xaa 是芳族、脂肪族或非极性残基

<220>

<221>MISC_FEATURE

<222> (125).. (125)

<223>Xaa 是极性或非极性残基

<220>

<221>MISC_FEATURE

<222> (142).. (142)

<223>Xaa 是极性残基

<220>
<221>MISC_FEATURE
<222>(147)..(147)
<223>Xaa 是芳族、脂肪族、极性或非极性残基

<220>
<221>MISC_FEATURE
<222>(149)..(149)
<223>Xaa 是非极性或非芳族残基

<220>
<221>MISC_FEATURE
<222>(150)..(150)
<223>Xaa 是酸性或受限残基

<220>
<221>MISC_FEATURE
<222>(152)..(152)
<223>Xaa 是极性或非极性残基

<220>
<221>MISC_FEATURE
<222>(190)..(190)
<223>Xaa 是非芳族残基

<220>
<221>MISC_FEATURE
<222>(196)..(196)
<223>Xaa 是脂肪族、非极性或非芳族残基

<220>
<221>MISC_FEATURE
<222>(202)..(202)
<223>Xaa 是芳族、非极性或非脂肪族残基

<220>
<221>MISC_FEATURE
<222>(205)..(205)
<223>Xaa 是碱性、非极性或非脂肪族残基

<220>

<221>MISC_FEATURE

<222>(206).. (206)

<223>Xaa 是芳族或非极性残基

<400>119

```

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

```

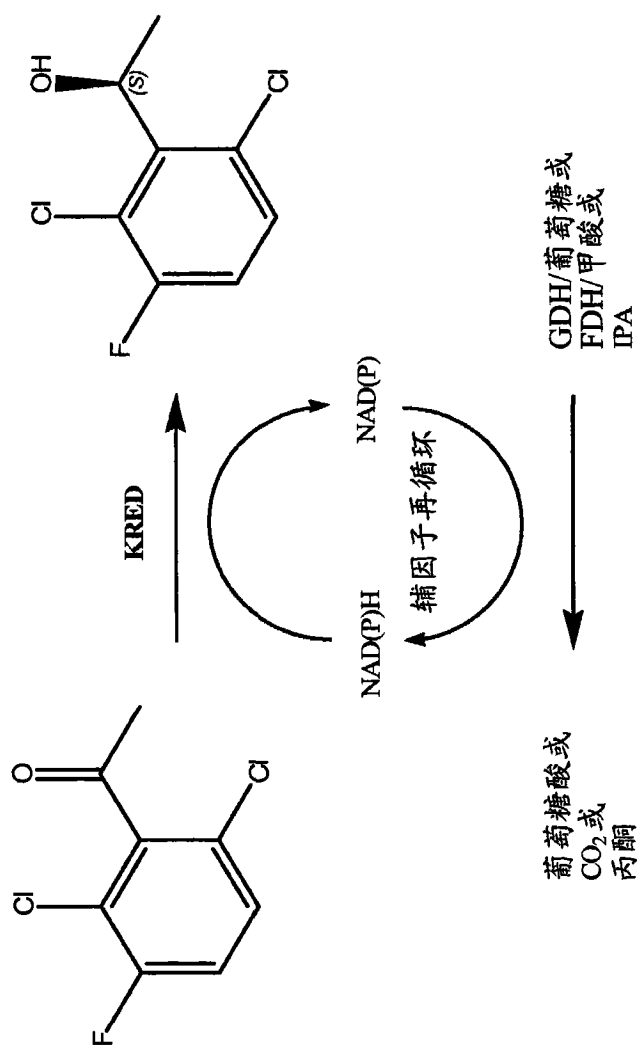



图 1