



(12) 发明专利申请

(10) 申请公布号 CN 101883846 A

(43) 申请公布日 2010. 11. 10

(21) 申请号 200880118571. 8

(51) Int. Cl.

(22) 申请日 2008. 10. 01

C12N 9/04 (2006. 01)

(30) 优先权数据

C12P 13/00 (2006. 01)

60/976, 555 2007. 10. 01 US

C12P 17/00 (2006. 01)

(85) PCT申请进入国家阶段日

2010. 05. 31

(86) PCT申请的申请数据

PCT/US2008/078513 2008. 10. 01

(87) PCT申请的公布数据

W02009/046153 EN 2009. 04. 09

(71) 申请人 科德克希思公司

地址 美国加利福尼亚州

(72) 发明人 奥诺拉托·坎帕皮阿诺

艾米丽·穆德弗 波蒂·博勒普

拉玛·沃兰德里

(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 8 页 说明书 76 页 序列表 82 页

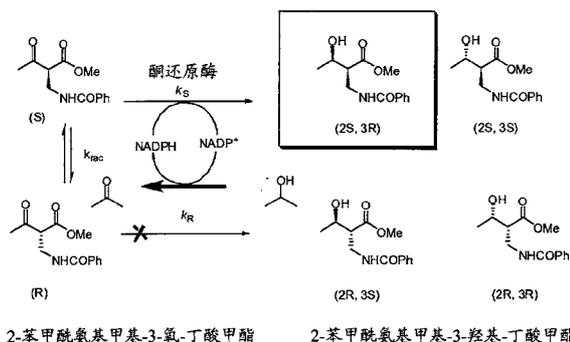
附图 1 页 按照条约第19条的修改 8 页

(54) 发明名称

用于生成氮杂环丁酮的还原酶多肽

(57) 摘要

本公开提供具有与自然存在的野生型酮还原酶相比具有改善的性质的工程酮还原酶。本文还提供了编码工程酮还原酶的多核苷酸、能够表达工程酮还原酶的宿主细胞以及应用工程酮还原酶合成各种手性化合物的方法。



1. 一种酮还原酶多肽,能够以至少约 60% 的立体异构体过量百分比将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。

2. 如权利要求 1 所述的多肽,其包含与基于 SEQ ID NO :2 或 4 或 86 的参考序列具有至少约 85% 同一性的氨基酸序列,所述参考序列具有以下特征:相应于 X94 的残基为苏氨酸;相应于 X199 的残基为组氨酸以及相应于 X202 的残基为缬氨酸或亮氨酸;条件是所述酮还原酶多肽具有其中相应于 X94 的残基为脂肪族或极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。

3. 如权利要求 2 所述的多肽,所述多肽中相应于 X94 的残基为苏氨酸。

4. 如权利要求 2 所述的多肽,所述多肽中相应于 X199 的残基为丙氨酸、组氨酸或天冬酰胺。

5. 如权利要求 2 所述的多肽,所述多肽中相应于 X94 的残基为极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;并且相应于 X202 的残基为缬氨酸或亮氨酸。

6. 如权利要求 2 所述的多肽,所述多肽中相应于 X94 的残基为苏氨酸;相应于 X199 的残基为丙氨酸、组氨酸或天冬酰胺;并且相应于 X202 的残基为缬氨酸或亮氨酸。

7. 如权利要求 2 至 6 中任一项所述的多肽,其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个:

相应于 X2 的残基为极性、非极性或非脂肪族残基;

相应于 X4 的残基为碱性残基或半胱氨酸;

相应于 X11 的残基为非极性、脂肪族、或芳香族残基;

相应于 X40 的残基为受限或碱性残基;

相应于 X80 的残基为非极性、脂肪族或极性残基;

相应于 X86 的残基为非极性、脂肪族或极性残基;

相应于 X96 的残基为极性、芳香族、非极性或非脂肪族残基;

相应于 X105 的残基为非极性、脂肪族、碱性或酸性残基;

相应于 X129 的残基为非极性或非极性残基;

相应于 X147 的残基为芳香族、非极性或非脂肪族残基;

相应于 X153 的残基为极性、非极性或非脂肪族残基;

相应于 X190 的残基为芳香族或受限残基;

相应于 X195 的残基为非极性或非脂肪族残基;

相应于 X196 的残基为非极性或非脂肪族残基;

相应于 X206 的残基为非极性或非芳香族残基;

相应于 X226 的残基为非极性或非脂肪族残基;

相应于 X248 为非极性或非碱性残基;

相应于 X249 为芳香族残基;并且

其中与上述参考序列相比,任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

8. 如权利要求 2 至 6 中任一项所述的多肽,其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个:

相应于 X2 的残基为丙氨酸;

相应于 X4 的残基为半胱氨酸；  
相应于 X11 的残基为苯丙氨酸；  
相应于 X40 的残基为精氨酸；  
相应于 X80 的残基为苏氨酸；  
相应于 X86 的残基为异亮氨酸；  
相应于 X96 的残基为缬氨酸或苯丙氨酸；  
相应于 X105 的残基为甘氨酸；  
相应于 X129 的残基为苏氨酸；  
相应于 X147 的残基为蛋氨酸或亮氨酸；  
相应于 X153 的残基为丙氨酸或丝氨酸；  
相应于 X190 的残基为组氨酸或脯氨酸；  
相应于 X195 的残基为缬氨酸；  
相应于 X196 的残基为亮氨酸；  
相应于 X206 的残基为苯丙氨酸；  
相应于 X226 的残基为缬氨酸；  
相应于 X248 的残基为赖氨酸或精氨酸；  
相应于 X249 的残基为色氨酸；

其中与所述参考序列相比，任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

9. 如权利要求 7 所述的多肽，其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个：

相应于 X40 的残基为受限或碱性残基；并且  
相应于 X147 的残基为芳香族、非极性或非脂肪族残基。

10. 如权利要求 9 所述的多肽，其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个：

相应于 X96 的残基为极性、芳香族、非极性或非脂肪族残基；  
相应于 X195 的残基为非极性或非脂肪族残基；  
相应于 X196 的残基为非极性或非脂肪族残基；  
相应于 X226 的残基为非极性或非脂肪族残基；  
相应于 X248 的残基为非极性或非碱性残基；  
相应于 X249 的残基为芳香族残基。

11. 如权利要求 9 或 10 所述的多肽，其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个：

相应于 X2 的残基为极性、非极性或非脂肪族残基；  
相应于 X4 的残基为碱性残基或半胱氨酸；  
相应于 X11 的残基为非极性、脂肪族或芳香族残基；  
相应于 X80 的残基为非极性、脂肪族或极性残基；  
相应于 X86 的残基为非极性、脂肪族或极性残基；  
相应于 X105 的残基为非极性、脂肪族、碱性或酸性残基；

相应于 X129 的残基为非极性或极性残基；  
相应于 X153 的残基为极性、非极性或脂肪族残基；  
相应于 X190 的残基为芳香族或受限残基；  
相应于 X206 的残基为非极性或芳香族残基。

12. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X40 的残基为受限或碱性残基；并且  
相应于 X147 的残基为芳香族、非极性或脂肪族残基；并且  
其中与所述参考序列相比,任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

13. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X40 的残基为受限或碱性残基；  
相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；  
相应于 X 147 的残基为芳香族、非极性或脂肪族残基；  
其中与所述参考序列相比,任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

14. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；  
相应于 X147 的残基为芳香族、非极性或脂肪族残基；  
相应于 X195 的残基为非极性或脂肪族残基；  
相应于 X196 的残基为非极性或脂肪族残基；并且  
其中与所述参考序列相比,任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

15. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X40 的残基为精氨酸。

16. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X147 的残基为蛋氨酸或亮氨酸。

17. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X40 的残基为精氨酸；并且相应于 X147 的残基为蛋氨酸或亮氨酸。

18. 如权利要求 1 所述的多肽,其中所述酮还原酶多肽包含具有以下氨基酸序列的区域或结构域:所述氨基酸序列与基于 SEQ ID NO:2、4 或 86 的参考序列的残基 90 至 211 具有至少 85% 同一性,所述参考序列具有以下特征:相应于 X94 的残基为苏氨酸;相应于 X199 的残基为组氨酸;并且相应于 X202 的残基为缬氨酸或亮氨酸;条件是所述酮还原酶的结构域或区域具有其中相应于 X94 的残基为脂肪族或极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;并且相应于 X202 的残基为亮氨酸或缬氨酸的氨基酸序列。

19. 如权利要求 18 所述的多肽,其中所述酮还原酶的结构域或区域包含其中相应于 X94 的残基为苏氨酸的氨基酸序列。

20. 如权利要求 18 所述的多肽,其中所述酮还原酶的结构域或区域包含其中相应于 X199 的残基为丙氨酸、组氨酸或天冬酰胺的氨基酸序列。

21. 如权利要求 18 所述的多肽,其中所述酮还原酶的结构域或区域包含其中相应于

X94 的残基为极性残基；相应于 X199 的残基为脂肪族、受限或极性残基；并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。

22. 如权利要求 18 所述的多肽，其中所述酮还原酶的结构域或区域包含其中相应于 X94 的残基为苏氨酸；相应于 X199 的残基为丙氨酸、组氨酸或天冬酰胺；并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。

23. 如权利要求 18 至 21 中任一项所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征中的一个或多个：

相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；

相应于 X105 的残基为非极性、脂肪族、碱性或酸性残基；

相应于 X129 的残基为非极性或极性残基；

相应于 X147 的残基为芳香族、非极性或脂肪族残基；

相应于 X153 的残基为极性、非极性或脂肪族残基；

相应于 X190 的残基为芳香族或受限残基；

相应于 X195 的残基为非极性或脂肪族残基；

相应于 X196 的残基为非极性或脂肪族残基；

相应于 X206 的残基为非极性或芳香族残基；

其中与所述参考序列相比，所述氨基酸序列可在相应于残基 90-211 的结构域中的其他氨基酸残基位置处任选地具有一个或多个残基差异。

24. 如权利要求 18 至 21 中任一项所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征中的一个或多个：

相应于 X96 的残基为缬氨酸或苯丙氨酸；

相应于 X105 的残基为甘氨酸；

相应于 X129 的残基为苏氨酸；

相应于 X147 的残基为蛋氨酸或亮氨酸；

相应于 X153 的残基为丙氨酸或丝氨酸；

相应于 X190 的残基为组氨酸或脯氨酸；

相应于 X195 的残基为缬氨酸；

相应于 X196 的残基为亮氨酸；

相应于 X206 的残基为苯丙氨酸；

其中与所述参考序列相比，所述氨基酸序列可在相应于残基 90-211 的结构域中的其他氨基酸残基位置处任选地具有一个或多个残基差异。

25. 如权利要求 18 所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征：

相应于 X147 的残基为芳香族、非极性或脂肪族残基；并且

其中与所述参考序列相比，任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

26. 如权利要求 18 所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征：

相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；

相应于 X147 的残基为芳香族、非极性或脂肪族残基；并且

其中与所述参考序列相比，任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

27. 如权利要求 18 所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征：

相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；

相应于 X147 的残基为芳香族、非极性或脂肪族残基；

相应于 X195 的残基为非极性或脂肪族残基；

相应于 X196 的残基为非极性或脂肪族残基；

其中与所述参考序列相比，任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

28. 如权利要求 1 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、46、50、52、54、56、58、60 和 62 的氨基酸序列。

29. 如权利要求 1 所述的多肽，其能够以至少约 90% 的立体异构体过量百分比将所述底物转化为所述产物。

30. 如权利要求 29 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、40、42、50、52、56、58、60 或 62 的氨基酸序列。

31. 如权利要求 1 所述的多肽，其能够以至少约 95% 的立体异构体过量百分比将所述底物转化为所述产物。

32. 如权利要求 31 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、42、50、52、56、58、60 或 62 的氨基酸序列。

33. 如权利要求 1 所述的多肽，其能够以至少约 99% 的立体异构体过量百分比将所述底物转化为所述产物。

34. 如权利要求 33 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、30、32、34、60 和 62 的氨基酸序列。

35. 如权利要求 1 所述的多肽，其能够以比参考多肽 SEQ ID NO :48 高至少 15 倍的速率将所述底物转化为所述产物。

36. 如权利要求 35 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、26、28、30、32、34、50、60 和 62 的氨基酸序列。

37. 如权利要求 1 所述的多肽，其能够以比参考多肽 SEQ ID NO :48 高至少 30 倍的速率将所述底物转化为所述产物。

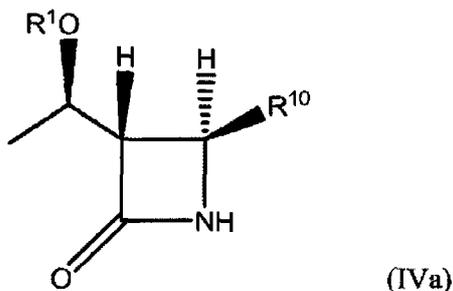
38. 如权利要求 37 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、26、30、34、60 和 62 的氨基酸序列。

39. 如权利要求 1 所述的多肽，其能够以比参考多肽 SEQ ID NO :48 高至少 40 倍的速率将所述底物转化为所述产物。

40. 如权利要求 39 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、22 和 60 的氨基酸序列。

41. 如权利要求 1 所述的多肽，其能够以比参考多肽 SEQ ID NO :48 高至少 50 倍的速率将所述底物转化为所述产物。

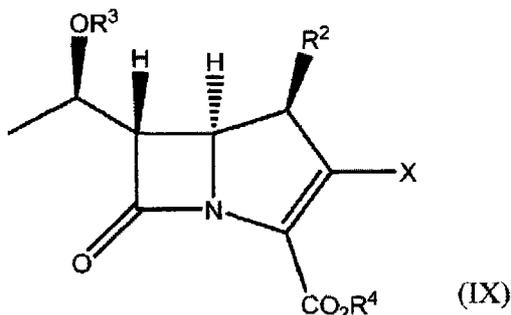
42. 如权利要求 42 所述的多肽,其包含相应于 SEQ ID NO :6、8、10 和 12 的氨基酸序列。
43. 一种酮还原酶多肽,能够以至少约 85% 的立体异构体过量百分比将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。
44. 如权利要求 43 所述的多肽,其包含相应于 SEQ ID NO :68、72、74、76、78 和 82 的氨基酸序列。
45. 一种多核苷酸,编码根据权利要求 1-44 中任一项所述的多肽。
46. 如权利要求 45 所述的多核苷酸,其为相应于 SEQ ID NO :5、7、9、11、13、15、17、19、21、23、25、27、29、31、33、35、37、39、41、45、49、51、53、55、57、59 或 61 的序列。
47. 一种多核苷酸,编码根据权利要求 43 或 44 所述的多肽。
48. 如权利要求 47 所述的多核苷酸,其为相应于 SEQ ID NO :67、71、73、75、77 或 81 的序列。
49. 一种表达载体,包含与适用于引导宿主细胞中的表达的控制序列可操作地连接的如权利要求 44 或 47 所述的多核苷酸。
50. 如权利要求 49 所述的表达载体,其中所述控制序列包含启动子。
51. 如权利要求 50 所述的表达载体,其中所述启动子包含大肠埃希氏菌启动子。
52. 如权利要求 50 所述的表达载体,其中所述控制序列包含分泌信号。
53. 一种宿主细胞,包含如权利要求 49 所述的表达载体。
54. 如权利要求 53 所述的宿主细胞,其为大肠埃希氏菌。
55. 一种组合物,包含如权利要求 1-42 中任一项所述的酮还原酶和具式 (I) 的化合物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯或具式 (II) 的化合物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。
56. 如权利要求 55 所述的组合物,其中底物为具式 (I) 的化合物和具式 (II) 的化合物。
57. 如权利要求 55 所述的组合物,还包含辅因子再生系统。
58. 如权利要求 57 所述的组合物,其中所述辅因子再生系统包括葡萄糖脱氢酶和葡萄糖;甲酸脱氢酶和甲酸;或异丙醇和仲醇脱氢酶。
59. 一种将具式 (I) 的底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯还原为具式 (II) 的产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的方法,所述方法包括在适合将所述底物还原为具式 (II) 的产物的反应条件下,将所述底物与如权利要求 1-44 中任一项所述的酮还原酶多肽接触或孵育。
60. 如权利要求 59 所述的方法,其中所述产物以高于约 99% 的立体异构体过量存在。
61. 如权利要求 59 所述的方法,其使用表达所述酮还原酶的完整细胞、或所述细胞的提取物或裂解物进行。
62. 如权利要求 59 所述的方法,其中所述酮还原酶是分离和 / 或纯化的,并且还原反应在所述酮还原酶的辅因子和任选地所述辅因子的再生系统的存在下进行。
63. 如权利要求 59 所述的方法,其中所述辅因子再生系统包括葡萄糖脱氢酶和葡萄糖;甲酸脱氢酶和甲酸;或异丙醇和仲醇脱氢酶。
64. 如权利要求 63 所述的方法,其中所述仲醇脱氢酶为所述酮还原酶。
65. 一种合成具式 (IVa) 的中间体的方法,



其中  $R^1$  为 H 或羟基保护基团, 并且  $R^{10}$  为卤素或  $-OAc$ , 其中 Ac 为乙酸,

所述方法中的步骤包括在适合于将具式 (I) 的底物还原或转化为具式 (II) 的产物的反应条件下, 将所述底物与本公开的酮还原酶接触或反应。

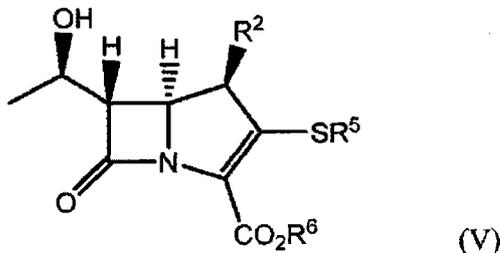
66. 一种合成具结构式 (IX) 的中间体的方法,



其中  $R^2$  为 H 或 C1-C4 烷基 (例如,  $-CH_3$ );  $R^3$  为 H 或羟基保护基团;  $R^4$  为 H、羧基保护基团、氨基、碱金属或碱土金属; 并且 X 为 OH 或离去基团,

所述方法中的步骤包括所述方法中的步骤包括在适合于将具式 (I) 的底物还原或转化为具式 (II) 的产物的反应条件下, 将所述底物与本公开的酮还原酶接触或反应。

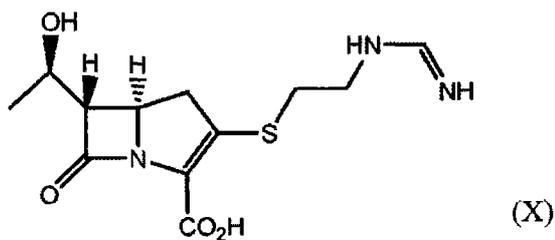
67. 一种合成具结构式 (V) 的碳杂青霉烯或其溶剂化物、水合物、盐和前药的方法:



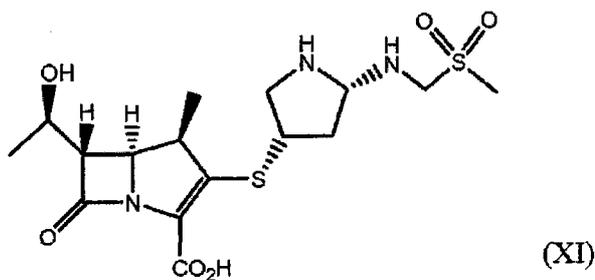
其中  $R^2$  为 H 或  $-CH_3$ ;  $R^5$  选自取代的或未取代的烷基、取代的或未取代的芳香基、取代的或未取代的杂烷基、取代的或未取代的杂环烷基、以及取代的或未取代的杂芳香基烷基; 并且  $R^6$  为 H 或前体基团,

所述方法中的步骤包括所述方法中的步骤包括在适合于将具式 (I) 的底物还原或转化为具式 (II) 的产物的反应条件下, 将所述底物与本公开的酮还原酶接触或反应。

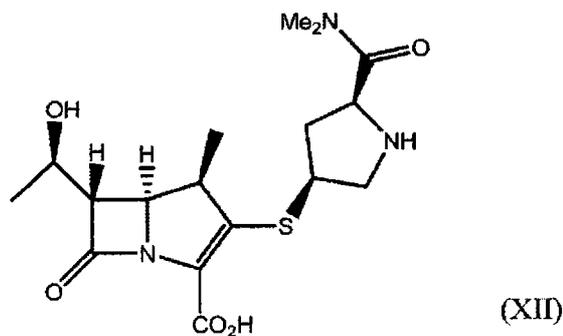
68. 如权利要求 67 所述的方法, 其中所述碳杂青霉烯具有结构式 (X):



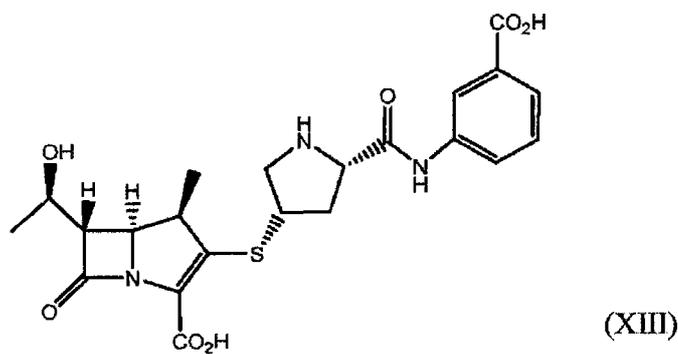
69. 如权利要求 67 所述的方法,其中所述碳杂青霉烯具有结构式 (XI) :



70. 如权利要求 67 所述的方法,其中所述碳杂青霉烯具有结构式 (XII) :



71. 如权利要求 67 所述的方法,其中所述碳杂青霉烯具有结构式 (XIII) :



## 用于生成氮杂环丁酮的还原酶多肽

[0001] 相关申请的交叉引用

[0002] 本申请根据 35U. S. C. § 119(e) 要求于 2007 年 10 月 1 日提交的申请序列第 60/976, 555 号的利益, 该申请的内容以引用方式并入本文。

[0003] 序列表、表格或计算机程序的引用

[0004] 以文件名 376247-022.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以引用方式整体并入本文。序列表的电子备份于 2008 年 10 月 1 日生成, 文件大小为 150 千字节。

[0005] 背景

[0006] 属于酮还原酶 (KRED) 或羰基还原酶类 (EC 1. 1. 1. 184) 的酶对于从相应的前体立体异构酮底物或相应的外消旋的醛底物中合成旋光的醇有用。典型的, KRED 将酮和醛底物转化为相应的醇产物, 但是也可催化逆反应, 将醇底物氧化为相应的酮/醛产物。用如 KRED 的酶对酮和醛的还原以及醇的氧化需要辅因子, 最常见的为还原的烟酰胺腺嘌呤二核苷酸 (NADH) 或还原的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。经常观察到酮还原酶和醇脱氢酶接受磷酸化的或非磷酸化的辅因子 (以其氧化的或还原的状态)。

[0007] KRED 酶可在广泛范围内的细菌和酵母中发现 (综述参见: Kraus 和 Waldman, *Enzyme catalysis in organic synthesis* (有机合成中的酶催化) 第 1&2 卷, VCH Weinheim 1995; Faber, K., *Biotransformations in organic chemistry* (有机化学中的生物转化), 第 4 版, Springer, Berlin Heidelberg New York, 2000; Hummel 和 KuIa, 1989, *Eur. J. Biochem.* 184 :1-13)。报道了一些 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)。

[0008] 为了回避生成关键化合物的许多化学合成的步骤, 酮还原酶增加地应用于不同酮和醛底物至手性醇产物的酶促转化。这些应用可将表达酮还原酶的完整细胞应用于生物催化的酮还原, 或者在完整细胞中多种酮还原酶的存在不利地影响所需产物的立体纯度和产量的情况下, 采用纯化的酶。对于体外应用, 将诸如葡萄糖脱氢酶 (GDH)、甲酸脱氢酶等的辅因子 (NADH 或 NADPH) 再生酶与酮还原酶联合使用。使用酮还原酶产生有用的化合物的例子包括 4- 氯乙酰乙酸酯类的不对称还原 (Zhou, *J. Am. Chem. Soc.* 1983:105 :5925-5926 ;Santaniello, *J. Chem. Res. (S)* 1984 :132-133 ;美国专利第 5, 559, 03 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 -5- 羟基 -3- 氧己酸叔丁酯的还原 (例如, 美国专利第 6, 645, 746 号和 WO 01/40450)、基于吡咯并三嗪的化合物的还原 (例如, 美国专利申请第 2006/0286646 号)、取代的苯乙酮的还原 (例如, 美国专利第 6, 800, 477 号); 和四氢噻吩酮 (ketothiolane) 的还原 (WO 2005/054491)。

[0009] 鉴别其它可用于将不同的酮底物转化至相应的手性醇化合物的酮还原酶是需要的。

[0010] 概述

[0011] 本公开提供具有还原 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯的外消旋混合物 (" 底物 ") 至 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯 (" 产物 ") 的能力的酮还原酶多肽、编码该多肽的多核苷酸和该多肽的使用方法。化合物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯为合成 (2R,3R)-3-((R)-1-(叔丁基二甲基甲硅烷氧基)乙基)-4-氧氮杂环丁-2-基乙酸酯 (" 氮杂环丁酮 (azetidinone); 乙酰基氧氮杂环丁酮 "; CAS 登记号 76855-69-1) 的中间体, 其为用于生产不同的碳杂青霉烯类抗生素的中间体 ( 倒数第二个中间体)。可从 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯合成的碳杂青霉烯类抗生素包括但不限于亚胺培南、美洛培南、多利培南、厄他培南、比阿培南、帕尼培南和与沙纳霉素类似的其它化合物。与自然存在的从克菲尔乳杆菌 (*Lactobacillus kefir*, " L. kefir "; SEQ ID NO:4)、短乳杆菌 (*Lactobacillus brevis*, " L. brevis "; SEQ ID NO:2) 或小乳杆菌 (*Lactobacillus minor*, " L. minor "; SEQ ID NO:86) 中得到的野生型酮还原酶相比, 本公开的工程酮还原酶多肽在还原或转化特定的底物至相应的手性醇产物上具有改善的性质。在一些实施方式中, 工程酮还原酶多肽与诸如 SEQ ID NO:48 的另外的工程酮还原酶多肽相比具有改善的性质。

[0012] 在一些实施方式中, 与序列为 SEQ ID NO:4、2 和 86 的野生型克菲尔乳杆菌 KRED、短乳杆菌 KRED 或小乳杆菌 KRED 相对比, 所述酮还原酶多肽至少具有以下特征: 残基 202 为缬氨酸或亮氨酸。在一些实施方式中, 与序列 SEQ ID NO:4、2 或 86 相对比, 本公开的酮还原酶具有以下特征中的至少两个: (1) 对应于 94 位 (也就是 X94) 的残基为脂肪族或极性残基, (2) 对应于 199 位 (也就是 X199) 的残基为脂肪族, 极性 or 受限残基, 和 (3) 对应于 202 位 (也就是 X202) 的残基为缬氨酸或亮氨酸。在一些实施方式中, 与序列 SEQ ID NO:4、2 或 86 相对比, 所述多肽具有至少以下特征: (1) 对应于 X94 位的残基为极性残基, (2) 对应于 X199 位的残基为脂肪族、受限或极性残基, 和 (3) 对应于 X202 位的残基为缬氨酸或亮氨酸。

[0013] 除以上所述特征外, 所述酮还原酶与序列 SEQ ID NO:2、4 或 86 相比可在其它残基位置处含有一个或多个残基差异。在一些实施方式中, 本文所述的酮还原酶多肽包含与基于 SEQ ID NO:2、4 或 86 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% 或更高的同一性的氨基酸序列, 所述参考序列具有以下特征: 对应于 X94 的残基为脂肪族或极性残基, 尤其是丙氨酸或苏氨酸; 对应于 X199 的残基为脂肪族、受限或极性残基, 尤其是丙氨酸、组氨酸或天冬酰胺; 并且对应于 X202 的残基为缬氨酸或亮氨酸; 条件是所述酮还原酶的氨基酸序列具有至少前面的特征, 也就是说, 对应于 X94 的残基为脂肪族或极性残基; 对应于 X199 的残基为脂肪族、受限或极性残基; 并且对应于 X202 的残基为缬氨酸或亮氨酸。

[0014] 在一些实施方式中, 例如当改善的性质来自单个残基差异或特定的残基差异的组合时, 与参考序列相比, 所述工程酮还原酶可任选地包括在多肽其它位置处的一个或多个残基差异。在一些实施方式中, 残基差异包含保守突变。在一些实施方式中, 可引入在其它残基位置处的另外的残基差异以产生在酶性质上的进一步的改善。这些改善可进一步提高

对于确定的底物的酶促活性,但也可包括在立体选择性、热稳定性、溶剂稳定性上的提高和/或降低的产物抑制。详述中提供了可产生一种或多种改善的酶性质的不同的残基差异。在一些实施方式中,改善的酮还原酶多肽包含对应于列于 SEQ ID NO :83、SEQ ID NO :84 或 SEQ ID NO :87 的序列式(或其区域,例如残基 90-211)的氨基酸序列。SEQ ID NO :84 为基于野生型克菲尔乳杆菌酮还原酶(SEQ ID NO :4)的氨基酸序列;SEQ ID NO :83 为基于野生型短乳杆菌酮还原酶(SEQ ID NO :2)的氨基酸序列;以及 SEQ ID NO :87 为基于野生型小乳杆菌酮还原酶(SEQ ID NO :86)的氨基酸序列。SEQ ID NO :83、84 和 87 的序列式详细说明对应于 X94 的残基为脂肪族或极性残基;对应于 X199 的残基为脂肪族、受限或极性残基;并且对应于 X202 的残基为缬氨酸或亮氨酸。如在详述中提供的,所述序列式进一步详细说明在其它残基位置的特征。

[0015] 在一些实施方式中,所述工程酮还原酶多肽可具有与野生型酮还原酶相比提高的将底物还原至产物的酶促活性。酶促活性的改善可通过采用标准的酶测定法比较所述酮还原酶与野生型酮还原酶的比活而测定。改善的量可在相应的野生型或参考酮还原酶的酶促活性的 1.5 倍(times)(或倍(fold))至多达 2 倍、5 倍、10 倍、20 倍、25 倍、50 倍、75 倍、100 倍或更高的范围内。在特定的实施方式中,工程酮还原酶显示比野生型或参考酮还原酶的酶促活性至少高 1.5 倍、2 倍、3 倍、4 倍、5 倍、10 倍、25 倍、50 倍、100 倍、500 倍或 1000 倍的改善的酶促活性。酶促活性的改善也包括在立体选择性、立体专一性、热稳定性、溶剂稳定性上的提高或降低的产物抑制。

[0016] 在一些实施方式中,本发明的酮还原酶多肽与 SEQ ID NO :4、SEQ ID NO :48 和 / 或 SEQ ID NO :66 相比,在酶促活性速率上,也就是其将底物转化为产物的速率上有改善。在一些实施方式中,所述酮还原酶多肽能够以比 SEQ ID NO :4 或 SEQ ID NO :90 的速率高至少 5 倍、10 倍、25 倍、50 倍、75 倍、100 倍、150 倍、200 倍、250 倍、300 倍、500 倍或 1000 倍的速率将底物转化为产物。

[0017] 在一些实施方式中,所述酮还原酶多肽能够以至少约 85% 的立体异构体过量百分比或比野生型克菲尔乳杆菌 KRED(SEQ ID NO :4) 高的立体异构体过量百分比,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 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 和 62 的氨基酸序列的多肽。

[0018] 在一些实施方式中,所述酮还原酶多肽能够以至少约 60-89% 的立体异构体过量百分比以及以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 1-15 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、46、50、52、54、56、58、60 和 62 的氨基酸序列的多肽。因为具有 SEQ ID NO :48 的氨基酸序列的参考多肽能够以一定速率(例如,在 50% IPA, pH 8 中,用约 10g/L KRED 在 20 小时内将 1g/L 底物 100% 转化)和相比于野生型(SEQ ID NO :4) 的改善的立体选择性将底物转化为产物,与 SEQ ID NO :48 相比改善的本文所述的多肽与野生型相比也有改善。

[0019] 在一些实施方式中,所述酮还原酶多肽能够以至少约 90-94% 的立体异构体过量

百分比,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO : 6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、40、42、50、52、56、58、60 和 62 的氨基酸序列的多肽。

[0020] 在一些实施方式中,所述酮还原酶多肽能够以至少约 95-99% 的立体异构体过量百分比,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO : 6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、42、50、52、56、58、60 和 62 的氨基酸序列的多肽。

[0021] 在一些实施方式中,所述酮还原酶多肽能够以至少约 99% 的立体异构体过量百分比,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、30、32、34、60 和 62 的氨基酸序列的多肽。

[0022] 在一些实施方式中,所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 15-30 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ IDNO :6、8、10、12、14、20、22、24、26、28、30、32、34、50、60 和 62 的氨基酸序列的多肽。

[0023] 在一些实施方式中,所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 30-40 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ IDNO :6、8、10、12、14、20、22、24、26、30、34、60 和 62 的氨基酸序列的多肽。

[0024] 在一些实施方式中,所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 40-50 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ IDNO :6、8、10、12、14、22 和 60 的氨基酸序列的多肽。

[0025] 在一些实施方式中,所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 50 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ IDNO :6、8、10 和 12 的氨基酸序列的多肽。

[0026] 在一些实施方式中,所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 50 倍的速率以及至少 99% 的立体异构体过量,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有所述性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10 和 12 的氨基酸序列的多肽。

[0027] 在一些实施方式中,本发明提供能够在 40°C 热处理 21 小时后保持其将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的能力的酮还原酶多肽。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO :

6、10 和 44 的氨基酸序列的多肽。

[0028] 在另一方面,本公开进一步提供可将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的酮还原酶多肽。在一些实施方式中,这些酮还原酶能够以至少约 85% 的立体异构体过量百分比将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。示例性的多肽包括但不限于包含相应于 SEQ ID NO :68、72、74、76、78 和 82 的氨基酸序列的多肽。

[0029] 在一些实施方式中,2R,3R 选择性酮还原酶多肽能够以比具有 SEQ ID NO :66 的氨基酸序列的多肽可达到的速率高至少约 1 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO :64、68、70、72、74、76、78、80 和 82 的氨基酸序列的多肽。因为具有 SEQ ID NO :66 的氨基酸序列的多肽能够以一定的立体异构体过量百分比和高于野生型克菲尔乳杆菌 KRED (SEQ ID NO :4) 的速率将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯,与 SEQ ID NO :66 相比有改善的任何多肽也相对于野生型克菲尔乳杆菌 KRED 有改善。

[0030] 在一些实施方式中,所述 2R,2R 选择性酮还原酶多肽能够以比具有 SEQ ID NO :66 的氨基酸序列的多肽可达到的速率高至少约 1-2 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯,转化为产物,2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO :64、68、70、72、74、76、78 和 80 的氨基酸序列的多肽。

[0031] 在一些实施方式中,所述 2R,2R 选择性酮还原酶多肽能够以比具有 SEQ ID NO :66 的氨基酸序列的多肽可达到的速率高至少约 5 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO :64、70、72、76、78 和 80 的氨基酸序列的多肽。

[0032] 在一些实施方式中,所述 2R,2R 选择性酮还原酶多肽能够以高于具有 SEQ ID NO :66 的氨基酸序列的多肽可达到的速率高至少约 5 倍的速率以及至少 85% 的立体异构体过量,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。具有此性质的示例性的多肽包括但不限于包含相应于 SEQ ID NO :72 和 78 的氨基酸序列的多肽。

[0033] 另一方面,本公开提供编码本文所述的工程酮还原酶的多核苷酸或在高度严格条件下与此多核苷酸杂交的多核苷酸。所述多核苷酸可包括可用于编码的工程酮还原酶表达的启动子或其它调节元件,并且可应用为特定需要的表达系统而优化的密码子。示例性的多核苷酸包括但不限于 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 和 81。示例性的多核苷酸也包括编码相应于 SEQ ID NO :83、84 和 87 的序列式的多肽的多核苷酸。

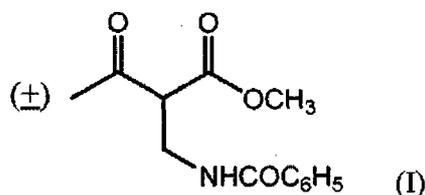
[0034] 在另一方面,本公开提供包含本文所述的多核苷酸和 / 或表达载体的宿主细胞。宿主细胞可为克菲尔乳杆菌或短乳杆菌或小乳杆菌,或者它们可为不同的生物体。所述宿主细胞可用于本文所述的工程酮还原酶的表达和分离,或者可选择地,它们可直接用于底物至立体异构产物的转化。

[0035] 无论用完整细胞、细胞提取物还是纯化的酮还原酶进行所述方法,可使用单一的

酮还原酶,或者可选择地,可使用两种或更多种酮还原酶的混合物。

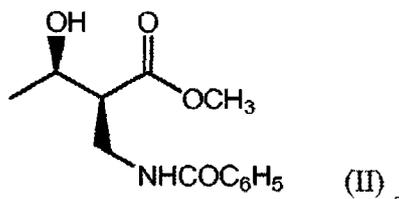
[0036] 如上所示,在一些实施方式中,本文所述的酮还原酶能够催化具结构式(I)的化合物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯 (“底物”) 中的酮基:

[0037]



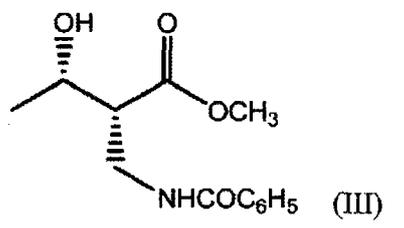
[0038] 至相应的具结构式(II)的立体异构醇产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯 (“产物”) 的还原反应:

[0039]



[0040] 在一些实施方式中,本文所述的酮还原酶能够催化具结构式(I)的化合物中的酮基至相应的具结构式(III)的立体异构醇产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯 (也可称为“产物”) 的还原反应:

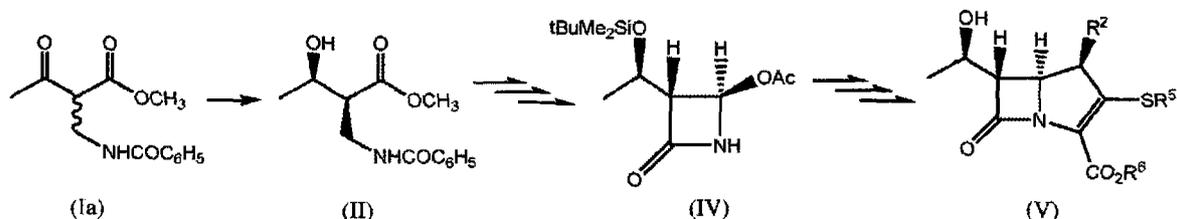
[0041]



[0042] 相应地,在一些实施方式中,本文提供的是还原具结构式(I)的底物至具结构式(II)或结构式(III)的醇产物的方法,所述方法包括在适合于还原或转化底物至具结构式(II)或结构式(III)的产物的反应条件下,将底物与本公开的酮还原酶多肽接触或孵育。

[0043] 在一些实施方式中,具结构式(II)的产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯可用于合成中间体或碳杂青霉烯类化合物,如方案 3 所示:

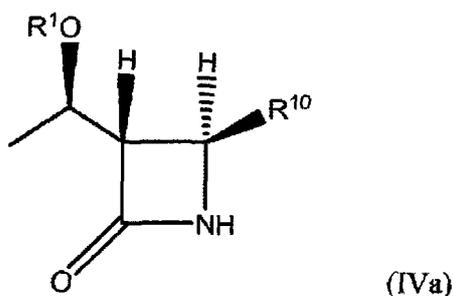
[0044]



[0045] 方案 3。

[0046] 相应地,在一些实施方式中,本公开的酮还原酶可用于合成具结构式(IVa)的中间体的方法,

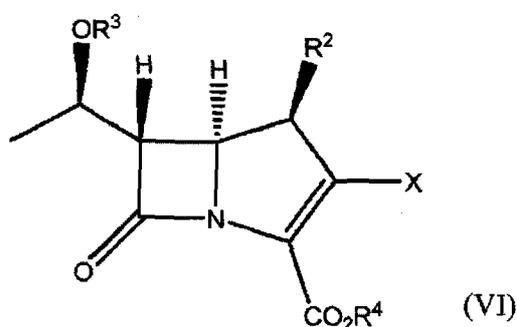
[0047]



[0048] 其中  $R^1$  为 H 或羟基保护基团, 并且  $R^{10}$  为卤素 (如, Cl) 或  $-OAc$  (Ac 为乙酸 (acetate))。相应地, 在合成具结构式 (IVa) 的中间体的方法中, 方法中的步骤包括在适合于还原或转化底物至具式 (II) 的产物的反应条件下, 将具式 (I) 的底物与本公开的酮还原酶接触或反应。

[0049] 在一些实施方式中, 本公开的酮还原酶可用于具结构式 (VI) 的中间体的合成:

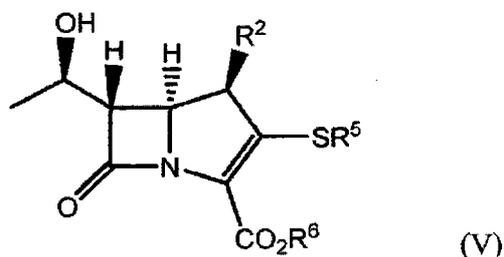
[0050]



[0051] 其中  $R^2$  为 H 或 C1-C4 烷基 (alkyl) (如,  $-CH_3$ );  $R^3$  为 H 或羟基保护基团;  $R^4$  为 H、羧基保护基团、氨基 (ammonia group)、碱金属或碱土金属; 并且 X 为 OH 或离去基团。示例性的离去基团包括但不限于  $-OP(O)(OR')$  或  $OS(O_2)R''$ , 其中  $R'$  和  $R''$  可为 C1-C6 烷基、C1-C6 烷芳基、芳香基、全氟代 C1-C6 烷基。相应地, 在一些实施方式中, 在合成具式 (VI) 的中间体的方法中, 方法中的步骤包括在适合于还原或转化底物至具式 (II) 的产物的反应条件下, 将具式 (I) 的底物与本公开的酮还原酶接触或反应。

[0052] 在一些实施方式中, 本公开的酮还原酶可用于合成具结构式 (V) 的基于碳杂青霉烯的治疗化合物或其溶剂化物、水合物、盐和前药的过程:

[0053]



[0054] 其中  $R^2$  为 H 或  $-CH_3$ ;  $R^5$  可为不同的取代基, 取代基包括但不限于, 取代或未取代的烷基、取代或未取代的芳香基、取代或未取代的杂烷基 (heteroalkyl)、取代或未取代的杂环烷基以及取代或未取代的杂芳香基烷基; 并且  $R^6$  为 H、或诸如可水解的酯基团的前体基团 (progroup)。相应地, 在合成具结构式 (V) 的化合物的方法中, 方法中的步骤可包括在适合于还原或转化底物至具式 (II) 的产物的反应条件下, 将具式 (I) 的底物与本公开的酮

还原酶接触或反应。具结构式 (V) 的示例性碳杂青霉烯包括但不限于亚胺培南、美洛培南、多利培南、厄他培南、比阿培南和帕尼培南。

[0055] 在一些实施方式中,本公开进一步提供酮还原酶在合成硫培南化合物的方法和用于硫培南合成的中间体上的应用。

[0056] 附图简述

[0057] 图 1 说明酮还原酶 (KRED) 在将具式 (I) 的底物化合物转化至相应的具式 (II) 的产物中的作用。该还原应用本发明的 KRED 和诸如 NADPH 的辅因子。使用异丙醇 (IPA) 将  $\text{NADP}^+$  转化 / 再循环至 NADPH。

[0058] 详述

[0059] 1.1 定义

[0060] 如本文所用,以下术语意在具有如下含义。

[0061] “酮还原酶”和“KRED”在本文可互换使用,指具有将羰基还原为其相应的醇的酶促能力的多肽。更确切地,本发明的酮还原酶多肽能够立体选择性地还原具上述分子式 (I) 的化合物至相应的具上述分子式 (II) 的产物。所述多肽典型地使用辅因子还原的烟酰胺腺嘌呤二核苷酸 (NADH) 或还原的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自然存在的 (野生型) 酮还原酶以及通过人工操作产生的非自然存在的工程多肽。

[0062] “编码序列”指核酸 (例如,基因) 中编码蛋白质氨基酸序列的部分。

[0063] “自然存在的”或“野生型”指在自然中发现的形式。例如,自然存在的或野生型多肽或多核苷酸序列为存在于生物体中的序列,该序列可从自然的来源中分离,并且没有通过人工操作被特意地修饰。

[0064] 当用于涉及如细胞、核酸或多肽时,“重组”指已经以不会存在于自然中的方式被修饰、或和自然形式相同但是从合成材料和 / 或通过使用重组技术操作的方式产生或衍生的材料,或相应于该材料的自然的或天然形式的材料。非限制性例子包括但不限于,表达在细胞的天然 (非重组的) 形式中未发现的基因或表达以不同的水平另外表达的天然基因的重组细胞。

[0065] “序列同一性百分比”和“同源性百分比”在此可互换使用,指多核苷酸之间和多肽之间的比较,并且通过在比较窗口中比较两个最优地比对的序列确定,其中比较窗口中的多核苷酸和多肽的部分可包含相对于参考序列 (其不包含添加或缺失) 的添加或缺失 (也就是间隔) 以最优地比对两个序列。所述百分比可通过确定两序列中发生的相同的核苷酸碱基或氨基酸残基的位置的数目以得到匹配位置的数目,将匹配位置的数目除以比较窗口中位置的总数,并且将结果乘以 100 以得到序列同一性百分比而计算。可选择地,所述百分比可通过确定两序列中发生的相同的核苷酸碱基或氨基酸残基或与缺口比对的核苷酸碱基或氨基酸残基的位置的数目以得到匹配位置的数目,将匹配位置的数目除以比较窗口中位置的总数,并且将结果乘以 100 以得到序列同一性百分比而计算。本领域中技术人员明白有许多可用的已确立的算法来比对两序列。用于比较的序列的最优比对可,比如,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实施 (GCG

Wisconsin 软件包中 GAP、BESTFIT、FASTA 和 TFASTA) 或通过目测检查来进行 (通常参见, Current Protocols in Molecular Biology (分子生物学最新实验设计), F. M. Ausubel 等编辑, Current Protocols, Greene Publishing Associates, Inc. 和 John Wiley & Sons, Inc. 联合出版, (1995 增刊) (Ausubel))。适合确定序列同一性和序列相似性百分比的算法的例子为 BLAST 和 BLAST 2.0 算法, 分别阐述于 Altschul 等, 1990, J. Mol. Biol. 215 :403-410 和 Altschul 等, 1977, Nucleic Acids Res. 3389-3402。进行 BLAST 分析的软件在美国国家生物技术信息中心 (National Center for Biotechnology Information) 网站对公众开放。该算法包括首先通过鉴别查询序列中长度 W 的短字 (short word) 来鉴别高分序列对 (HSP), 当与数据库序列中相同长度的字比对时, 所述短字匹配或符合一些正值阈值得分 T。T 指邻近字得分阈值 (Altschul 等, 如上)。这些起始相邻字匹配字串 (word hits) 作为起始搜索的种子以发现包含它们的更长的 HSP。所述字匹配字串继而沿每个序列的两个方向延伸, 直至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 (匹配的残基对的奖励得分; 总是 > 0) 和 N (不匹配的残基的惩罚得分; 总是 < 0) 计算。对氨基酸序列, 使用得分矩阵计算累积得分。当出现以下情况时, 每个方向上的字匹配字串的延伸被终止: 累计比对得分从其最大到达值降低量 X; 累积得分由于一个或多个负得分残基比对的累积而变至零或更低; 或者到达任一序列的末端。BLAST 算法参数 W、T 和 X 确定比对的敏感性和速度。BLASTN 程序 (用于核苷酸序列) 使用以下作为缺省值: 字长 (W) 为 11, 期望值 (E) 为 10, M = 5, N = -4, 以及两个链的比较。对于氨基酸序列, BLASTP 程序使用以下作为缺省值: 字长 (W) 为 3, 期望值 (E) 为 10 和 BLOSUM62 得分矩阵 (见 Henikoff 和 Henikoff, 1989, Proc Natl Acad Sci USA 89 :10915)。示例性的序列比对和 % 序列同一性的确定可应用 GCG Wisconsin 软件包 (Accelrys, Madison WI) 中 BESTFIT 或 GAP 程序, 使用提供的缺省参数。

[0066] “参考序列”指用于作为序列比较的基础的确定序列。参考序列可为更长序列的子集, 例如, 全长基因或多肽序列的区段。通常, 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 50 个残基长度或核酸或多肽的全长。因为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 (也就是完整序列的一部分), 和 (2) 还可包含两个序列之间不同的序列, 所以两个 (或更多个) 多核苷酸或多肽之间的比较典型地通过在 “比较窗口” 中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而完成。

[0067] 在一些实施方式中, “参考序列”可基于一级氨基酸序列, 其中参考序列为在一级序列上有一个或多个改变的序列。例如, “基于 SEQ ID NO :4 的在相应于 X202 的残基具有亮氨酸或缬氨酸的参考序列”指其中在 SEQ ID NO :4 的 X202 处的相应残基已变化为亮氨酸或缬氨酸的参考序列。

[0068] “比较窗口”指至少约 20 个连续的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 个连续的核苷酸或氨基酸的参考序列相比较, 并且其中比较窗口中的序列部分可包含与参考序列 (其不包含添加或缺失) 相比百分之二十或更少的添加或缺失 (也就是缺口) 以用于两序列的最优比对。比较窗口可长于 20 个连续的残基, 并且包括任性地 30、40、50、100 或更长的窗口。

[0069] “大体同一性”指与至少 20 个残基位置的比较窗口、经常在至少 30-50 个残基的

窗口中与参考序列比较,具有至少 80%序列同一性、至少 85%序列同一性和 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,其中序列同一性百分比通过在比较窗口将参考序列与包括参考序列的总计 20%或更少的缺失或添加的序列进行比较而计算。在应用于多肽的特定的实施方式中,术语“大体同一性”意思是当例如通过程序 GAP 或 BESTFIT 使用缺省的缺口权重进行最优比对时,两个多肽序列共享至少 80%序列同一性、优选地至少 89%序列同一性、至少 95%序列同一性或更高(例如,99%序列同一性)。优选地,不相同的残基位置具有保守的氨基酸取代。

[0070] 在用于指定的氨基酸或多核苷酸序列的编号的情况下,“相应于”、“参考于”或“相对于”指当指定的氨基酸或多核苷酸序列与参考序列相比较时特定的参考序列残基的编号。换句话说,指定聚合物的残基编号或残基位置为对于参考序列而不是对于指定氨基酸或多核苷酸序列中残基的实际的编号位置而指定的。例如,指定的氨基酸序列,如工程酮还原酶的氨基酸序列,可通过引入缺口以使两序列之间的残基匹配最优而与参考序列比对。在这些情况下,尽管存在缺口,指定氨基酸或多核苷酸序列的编号是相对于与之比对的参考序列而指定的。

[0071] “立体选择性”指在化学或酶促反应中一种立体异构体相对于另一种的优先形成。立体选择性可为部分的,其中一种立体异构体相对于另一种为有利的,或者可为完全的,其中只形成一种立体异构体。当立体异构体为对映异构体时,立体选择性称为对映选择性,即一种对应异构体在两者总和中的分数(典型地报道为百分比)。在本领域内该分数(典型地为百分比)通常可选择地报道为根据下式从中计算的对映异构体过量(e. e.):  
[主要对映异构体 - 次要对映异构体] / [主要对映异构体 + 次要对映异构体]。当立体异构体为非对映异构体时,立体选择性称为非对映选择性,即一种非对映异构体在两种非对映异构体混合物中的分数(典型地报道为百分比),通常可选择地报道为非对映异构体过量(d. e.)。对映异构体过量和非对映异构体过量为立体异构体过量的类型。

[0072] “高度立体选择性”指能够以至少约 85%立体异构体过量将底物转化或还原为相应的具有化学式(II)或(III)的产物的酮还原酶多肽。

[0073] “立体专一性”指在化学或酶促反应中一种立体异构体相对另一种的优先转化。立体专一性可为部分的,其中一种立体异构体的转化相对于另一种为有利的,或者可为完全的,其中只转化成一种立体异构体。

[0074] “化学选择性”指在化学或酶促反应中一种产物相对于另一种的优先形成。

[0075] “改善的酶性质”指与参考酮还原酶相比展现出在任何酶性质上的改善的酮还原酶多肽。对于本文所述的工程酮还原酶多肽,比较通常是针对野生型酮还原酶,尽管在一些实施方式中,参考酮还原酶可为另外的改善的工程酮还原酶。期望改善的酶性质包括但不限于酶促活性(其可以底物的转化百分比形式表达)、热稳定性、溶剂稳定性、pH 活性谱(profile)、辅因子需求、对抑制剂(例如,产物抑制)的不应性、立体专一性和立体选择性(包括对映选择性)。

[0076] “增强的酶促活性”指工程酮还原酶多肽的改善的性质,其可用与参考酮还原酶相比,比活(例如,所产生的产物/时间/蛋白重量)的提高或底物至产物的转化百分率(例如,使用一定量的 KRED,在一定时间段内起始量的底物至产物的转化百分比)的提高来表示。实施例提供示例性的确定酶活性的方法。包括经典的酶性质  $K_m$ 、 $V_{max}$  或  $k_{cat}$  的关

于酶活性的任何性质可被影响,其变化可导致增强的酶促活性。酶活性的改善可为从相应野生型酮还原酶的酶促活性的约 1.5 倍,至超过自然存在的酮还原酶或酮还原酶多肽衍生自的另外的工程酮还原酶的酶促活性达 2 倍、5 倍、10 倍、20 倍、25 倍、50 倍、75 倍、100 倍或更高。在特定的实施方式中,工程酮还原酶表现高于母酮还原酶活性的在 1.5 至 50 倍、1.5 至 100 倍范围内的改善的酶促活性。技术人员理解,任何酶的活性为扩散限制性的,使得催化转换速率不能超过包括任何需要的辅因子的底物的扩散速率。扩散极限的理论最大值,或  $k_{cat}/K_m$ ,通常为约  $10^8$  至  $10^9$  ( $M^{-1}s^{-1}$ )。因此,酮还原酶的酶活性的任何改善会存在与酮还原酶作用的底物的扩散速率相关的上限。酮还原酶的活性可通过用于测量酮还原酶的任一种标准的测定法来测定,如伴随酮至醇的还原由 NADPH 的氧化产生的 NADPH 的吸光度或荧光的降低,或通过偶联的测定法中所产生的产物。酶活性的比较通过确定的酶制品、在设定条件下的确定的测定法和一种或多种确定的底物而完成的,如在此进一步详细阐述。通常,当比较裂解物时,确定细胞数目和被测定的蛋白的量,同时使用相同的表达系统和相同的宿主细胞以使宿主细胞产生的以及裂解物中存在的酶量上的变化最小。

[0077] “转化”指底物至相应产物的酶促还原。“转化百分比”指一定时间段特定的条件下被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”可表示为底物至产物的“转化百分比”。

[0078] “热稳定”指与未处理的酶相比,在暴露于高温(例如 40–80°C)一段时间(例如 0.5–24 小时)后保持相似的活性(例如大于 60%至 80%)的酮还原酶多肽。

[0079] “溶剂稳定”指与未处理的酶相比,在暴露于不同浓度(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、甲基叔丁醚等)中一段时间(例如 0.5–24 小时)后保持相似的活性(大于例如 60%至 80%)的酮还原酶多肽。

[0080] “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 至 12)一段时间(例如 0.5–24 小时)后保持相似的活性(大于例如 60%至 80%)的酮还原酶多肽。

[0081] “热和溶剂稳定”指既热稳定也溶剂稳定的酮还原酶多肽。

[0082] 在工程酮还原酶的上下文中,如本文所用的“衍生自”鉴定工程化所基于的起始的酮还原酶和/或编码所述酮还原酶的基因。例如,SEQ ID NO:60 的工程酮还原酶是通过在多代期间人工进化编码 SEQ ID NO:4 的克菲尔乳杆菌酮还原酶的基因得到的。因此,所述工程酮还原酶“衍生自”SEQ ID NO:4 的野生型酮还原酶。

[0083] “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 的标准化的一致疏水性等级小于 0 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)。

[0084] “酸性氨基酸或残基”指当氨基酸包含在肽或多肽中时,含有表现出小于约 6 的 pK 值的侧链的亲水性氨基酸或残基。酸性氨基酸典型地在生理 pH 下由于失去氢离子而具有带负电荷的侧链。遗传编码的酸性氨基酸包括 L-Glu(E) 和 L-Asp(D)。

[0085] “碱性氨基酸或残基”指当氨基酸包含在肽或多肽中时,含有表现出大于约 6 的 pK 值的侧链的亲水性氨基酸或残基。碱性氨基酸典型地在生理 pH 下由于缔合水合氢离子而具有带正电荷的侧链。遗传编码的碱性氨基酸包括 L-Arg(R) 和 L-Lys(K)。

[0086] “极性氨基酸或残基”指含在生理 pH 下不带电荷的侧链、但是该侧链含有其中

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)。

[0087] “疏水性氨基酸或残基”指含有表现出根据 Eisenberg 等,1984, J. Mol. Biol. 179 :125-142 的标准化的一致疏水性等级的大于 0 的疏水性的侧链的氨基酸或残基。遗传编码的疏水性氨基酸包括 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)。

[0088] “芳香族氨基酸或残基”指具有包括至少一个芳香环或杂芳香环的侧链的亲水性或疏水性氨基酸或残基。遗传编码的芳香族氨基酸包括 L-Phe(F)、L-Tyr(Y) 和 L-Trp(W)。尽管由于 L-His(H) 杂芳香氮原子的 pKa, L-His(H) 有时被划分为碱性残基或由于侧链中包括杂芳香环而划分为芳香族残基,在此组氨酸被划分为亲水性残基或作为“受限残基”(见下)。

[0089] “受限氨基酸或残基”指具有受限几何学的氨基酸或残基。在此,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的几何学。脯氨酸因为它也含有五元环而具有受限的几何学。

[0090] “非极性氨基酸或残基”指含有在生理 pH 下不带电荷的侧链、并且该侧链含有其中两个原子共同共享的电子对通常为两个原子中的每一个等同维持的键(也就是说,侧链不是极性的)的疏水性氨基酸或残基。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)。

[0091] “脂肪族氨基酸或残基”指具有脂肪族烃侧链的疏水性氨基酸或残基。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)。

[0092] “半胱氨酸”或 L-Cys(C) 为罕见的,原因是其可与其它 L-Cys(C) 氨基酸或其它含硫烷基或含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它氨基酸。L-Cys(C) (以及具有含 -SH 的侧链的其它氨基酸)以还原的游离 -SH 或氧化的二硫键形式存在于肽中的能力影响 L-Cys(C) 是否对肽贡献净疏水性或亲水性特征。虽然 L-Cys(C) 表现出根据 Eisenberg(Eisenberg 等人,1984,如上)的标准化的一致性等级为 0.29 的疏水性,但应理解的是,对于本公开的目的 L-Cys(C) 被划分在其自身特有的组中。

[0093] “小氨基酸或残基”指具有包括其三个或更少碳原子和 / 或杂原子(不包括  $\alpha$ -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划分为脂肪族、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)。

[0094] “含羟基的氨基酸或残基”指含有羟基(-OH)部分的氨基酸。遗传编码的含羟基的氨基酸包括 L-Ser(S)、L-Thr(T) 和 L-Tyr(Y)。

[0095] “保守”氨基酸取代或突变指含有相似侧链的残基的可交换性,并且因此典型地包括多肽中的氨基酸被存在于相同或相似的氨基酸定义类别中的氨基酸所取代。然而如本文所用,在一些实施方式中,如果保守突变可代替地为从脂肪族残基至脂肪族残基、非极性残基至非极性残基、极性残基至极性残基、酸性残基至酸性残基、碱性残基至碱性残基、芳香族残基至芳香族残基或受限残基至受限残基的取代,则保守突变不包括从亲水性残基至亲水性残基、疏水性残基至疏水性残基、含羟基残基至含羟基残基、或小残基至小残基的取

代。此外,如本文所用,A、V、L或I可被保守的突变为另一脂肪族残基或者另一非极性残基。下表显示示例性的保守取代。

[0096] 表 1

[0097]

残基	可能的保守突变
A、L、V、I	其它脂肪族残基 (A、L、V、I) 其它非极性残基 (A、L、V、I、G、M)
G、M	其它非极性残基 (A、L、V、I、G、M)
D、E	其它酸性残基 (D、E)
K、R	其它碱性残基 (K、R)
P、H	其它受限残基 (P、H)
N、Q、S、T	其它极性残基
Y、W、F	其它芳香族 (Y、W、F) 残基
C	无

[0098] “非保守取代”指多肽中的氨基酸被具有明显不同的侧链性质的氨基酸取代或突变。非保守取代可使用在以上所列定义组之间而不是之内的氨基酸。在一个实施方式中,非保守突变影响 (a) 取代区域内的肽主链的结构 (例如,脯氨酸取代甘氨酸) (b) 电荷或疏水性,或 (c) 侧链的体积 (bulk)。

[0099] “缺失”指通过从参考多肽中移除一个或多个氨基酸而对多肽的修饰。缺失可包含移除 1 个或多个氨基酸、2 个或更多氨基酸、5 个或更多氨基酸、10 个或更多氨基酸、15 个或更多氨基酸或 20 个或更多氨基酸、多达构成所述多肽的氨基酸总数的 10%、多达构成所述多肽的氨基酸总数的 20%、或多达构成所述多肽的氨基酸总数的 30%,同时保留酶促活性和 / 或保留工程酮还原酶的改善的性质。缺失可针对多肽的内部部分和 / 或末端部分。在不同的实施方式中,缺失可包含连续的区段或可为不连续的。

[0100] “插入”指通过从参考多肽添加一个或多个氨基酸的多肽的修饰。在一些实施方式中,改善的工程酮还原酶包含插入一个或多个氨基酸至自然存在的酮还原酶多肽以及插入一个或多个氨基酸至其它的改善的酮还原酶多肽中。插入可为在多肽内部部分,或至羧基或氨基末端。如本文所用的插入包括本领域内已知的融合蛋白。插入可为连续的氨基酸区段或被自然存在的多肽中的一个或多个氨基酸隔开。

[0101] 如本文所用的“片段”指具有氨基末端和 / 或羧基末端缺失、但是其中保留的氨基酸序列与序列相应的位置相同的多肽。片段可为至少 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,并且多达全长酮还原酶多肽的 70%、80%、90%、95%、98%

和 99%，例如 SEQ ID NO :2、4 或 86 的多肽。

[0102] “分离的多肽”指与自然伴随其的其它杂质大体分离的多肽，所述杂质例如蛋白质、脂类和多核苷酸。术语包括从其自然存在的环境或表达系统（例如，宿主细胞或体外合成）中移出或纯化的多肽。改善的酮还原酶可在细胞内存在，在细胞培养基中存在，或以不同的诸如裂解物或分离的制品形式制备。如此，在一些实施方式中，改善的酮还原酶可为分离的多肽。

[0103] “大体纯的多肽”指其中多肽种类为存在的主导的种类（也就是说，基于摩尔或质量，其比组合中任何其它单独的大分子种类更丰富）的组合物，并且当目标种类以摩尔或%质量表示构成存在的大分子种类的至少约 50%时，该组合物通常为大体纯化的组合物。通常，大体纯的酮还原酶组合物将包含以摩尔或%质量表示，存在于组合物中的所有大分子种类的约 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。在一些实施方式中，目标种类被纯化至基本同质（也就是说，组合物中的杂质种类通过常规的检测方法无法检测出），其中该组合物基本上由单一的大分子种类组成。溶剂种类、小分子（< 500 道尔顿）以及基本离子种类不认为是大分子种类。在一些实施方式中，分离的改善的酮还原酶多肽是大体纯的多肽组合物。

[0104] 如本文所用的“严格杂交”指核酸杂交体 (hybrid) 稳定的条件。如本领域内技术人员已知的，杂交体的稳定性反映在杂交体的解链温度 ( $T_m$ )。通常，杂交体的稳定性为离子强度、温度、G/C 含量和促溶剂 (chaotropic agent) 的存在的函数。多核苷酸的  $T_m$  值可用预测解链温度的已知方法来计算（见如，Baldino 等人，Methods Enzymology 168 : 761-777 ;Bolton 等人，1962, Proc. Natl. Acad. Sci. USA 48 :1390 ;Bresslauer 等人，1986, Proc. Natl. Acad. Sci USA 83 :8893-8897 ;Freier 等人，1986, Proc. Natl. Acad. Sci USA 83 : 9373-9377 ;Kierzek 等人，Biochemistry 25 :7840-7846 ;Rychlik 等人，1990, Nucleic Acids Res 18 :6409-6412 (勘误，1991, Nucleic Acids Res 19 :698) ;Sambrook 等人，如上) ;Suggs 等人，1981, 于 Developmental Biology Using Purified Genes (应用纯化的基因的发育生物学) (Brown 等人编辑)，683-693 页，Academic Press ;以及 Wetmur，1991, Crit Rev Biochem Mol Biol 26 :227-259。所有出版物以引用方式并入本文)。在一些实施方式中，所述多核苷酸编码本文公开的多肽，并且在确定的条件下，如适度严格或高度严格的条件下，与编码本公开的工程酮还原酶的序列的互补序列杂交。

[0105] “杂交严格性”指在核酸杂交中的杂交条件，诸如清洗条件。通常，杂交反应在较低严格性条件下进行，之后进行变化的但是较高的严格性的清洗。术语“适度严格的杂交”指允许目标 -DNA 结合与该目标 DNA 具有约 60%同一性、优选地约 75%同一性、约 85%同一性、与靶标多核苷酸具有高于约 90%同一性的互补核酸的条件。示例性的适度严格条件为等同于在 42°C 下 50% 甲酰胺、5x Denhart' s 溶液、5x SSPE、0.2% SDS 中杂交，之后在 42°C 下 0.2x SSPE、0.2% SDS 中清洗的条件。“高度严格性杂交”通常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  $T_m$  低约 10°C 或更少的条件。在一些实施方式中，高度严格性状态指只允许在 65°C 下 0.018M NaCl 中形成稳定杂交体的那些核酸序列杂交的条件（也就是说，如果杂交在 65°C 下 0.018M NaCl 中不稳定，其将在高度严格性状态下不稳定，如本文所述的）。可以提供高度严格性条件，例如通过在等同于在 42°C 下 50% 甲酰胺、5x Denhart' s 溶液、5x SSPE、0.2% SDS 中杂交，之后在 65°C 下 0.1x SSPE 和 0.1% SDS 中

清洗的条件。另一个高度严格性条件为在等同于在 65°C 下含 0.1% (w : v) SDS 的 5X SSC 中杂交并且在 65°C 下含 0.1% SDS 的 0.1x SSC 中清洗。其它的高度严格性杂交的条件,以及适度严格性条件,在以上引用的参考文献中阐述。

[0106] “异源”多核苷酸指通过实验室技术引入宿主细胞的任何多核苷酸,并且包括通过实验室操作从宿主细胞中移出、经受实验室操作然后重新引入至宿主细胞中的多核苷酸。

[0107] “密码子优化”指编码蛋白质的多核苷酸的密码子至特定的生物体中优先使用的那些密码子的变化,以使得编码的蛋白质在感兴趣的生物体中有效率地表达。尽管基因遗传密码具简并性,即大多数氨基酸由称为“同义”(“synonyms”)或“同义”(“synonymous 同义”)密码子的几个密码子代表,但是众所周知,特定生物体密码子的使用是非随机的和对于特定的密码子三联体有偏好的。这种密码子使用偏倚性可在特定的基因、共同功能或祖先起源的基因、相对于低拷贝数蛋白质高度表达的蛋白质以及生物体基因组的聚集蛋白质编码区中更高。在一些实施方式中,编码酮还原酶的多核苷酸可为从选择用于表达的宿主生物体中的最佳化生产而进行密码子优化。

[0108] “优选、最佳的、高密码子使用偏倚性密码子”可交换地指在蛋白编码区中的使用频率高于编码相同的氨基酸的其他密码子的密码子。优选的密码子可根据在单个基因、具有共同功能或起源的一组基因、高度表达的基因中的密码子使用,在整个生物体的聚集蛋白质编码区域的密码子频率,在相关生物体的聚集蛋白质编码区域的密码子频率,或其组合来确定。随基因表达水平而频率增加的密码子通常为用于表达的最佳密码子。确定特定的生物体中的密码子频率(例如,密码子使用,相对的同义密码子使用)以及在密码子偏爱性的方法为公知的,包括多变量分析,例如,使用聚类分析或对应分析,以及用于基因中的密码子的有效数目(参见 GCGCodonPreference, Genetics Computer Group Wisconsin Package ;CodonW, John Peden, University of Nottingham ;McInerney, J. O, 1998, Bioinformatics14 :372-73 ;Stenico 等人, 1994, Nucleic Acids Res. 22:437-46 ;Wright, F., 1990, Gene 87 :23-29)。密码子使用表对越来越多的生物体可用(参见例如, Wada 等人, 1992, Nucleic Acids Res. 20 :2111-2118 ;Nakamura 等人, 2000, Nucl. Acids Res. 28 :292 ;Duret, 等人, 如上 ;Henaut 和 Danchin, “Escherichia coli and Salmonella, ” (“大肠埃希氏菌和沙门氏菌”) 1996, Neidhardt 等人编辑, ASM Press, Washington D. C., 2047-2066 页。获得密码子使用的数据源可依赖于可以编码蛋白质的任何可用的核苷酸序列。这些数据集包括实际上已知为编码表达的蛋白质(例如,完整蛋白质编码序列 -CDS)、表达序列标签 (ESTS) 或基因组序列的预测编码区域的核酸序列(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, Methods Enzymol. 266 :259-281 ;Tiwari 等人, 1997, Comput. Appl. Biosci. 13 :263-270)。

[0109] “控制序列”在此定义为包括对本公开的多肽的表达为必需的或有利的的所有组分。每种控制序列可为编码多肽的核苷酸序列的天然或外来的序列。所述控制序列包括但不限于前导序列、聚腺苷酸序列、前肽序列、启动子、信号肽序列和转录终止子。

[0110] “可操作地连接的”在此定义为一种构型,控制序列被以该构型置于相对于感兴

趣的多核苷酸合适的位置（也就是，处于功能性关系）使得控制序列指向或调节感兴趣的多核苷酸和 / 或多肽的表达。

[0111] “启动子序列”为可被用于表达感兴趣的多核苷酸的宿主细胞识别的核苷酸序列，例如编码序列。控制序列可包含合适的启动子序列。启动子序列包括转录控制序列，该序列介导目标多核苷酸的表达。启动子可为在选择的宿主细胞中表现转录活性的任何核苷酸序列，包括突变的、截断的和杂合 (hybrid) 的启动子，并且可从对宿主细胞同源的或异源的编码细胞外或细胞内的多肽的基因中获得。

#### [0112] 1.2 酮还原酶

[0113] 在一个方面，本公开提供可以将 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯的外消旋混合物（“底物”）立体选择性地还原或转化至 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯（“2S,3R 产物”）的工程酮还原酶（“KRED”）。当与从克菲尔乳杆菌 (SEQ ID NO :4)、短乳杆菌 (SEQ ID NO :2) 或小乳杆菌 (SEQ IDNO :86) 中得到的自然存在的野生型 KRED 酶或其它工程酮还原酶相比时，这些酮还原酶多肽（也在此描述为“2S,3R 选择性酮还原酶”）具有将 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯还原或转化至 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的改善的性质。

[0114] 在一些实施方式中，与野生型或另外的工程多肽如 SEQ ID NO :64 相比，改善的性质为就将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯还原或转化为 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的立体选择性的增加而言，也就是说，在此，产物的立体异构体过量的增加。在一些实施方式中，酮还原酶多肽的改善的性质为就其将更多百分比的底物转化或还原至产物的能力而言。在一些实施方式中，酮还原酶多肽的改善的性质为就其转化底物至产物的速率而言，这可通过其可使用相比于野生型或其它参考序列更少的改善的多肽以还原或转化相同量的产物的能力而证明。在一些实施方式中，酮还原酶多肽的改善的性质为就其稳定性或热稳定性而言。在一些实施方式中，酮还原酶多肽具有多于一种改善的性质，如增强的立体选择性和改善的酶促活性。

[0115] 本公开进一步提供能够将 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯的外消旋混合物（“底物”）立体选择性地还原或转化为 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯（“2R,3R 产物”）的工程酮还原酶。与产生 2S,3R 产物的工程多肽相似，能够产生 2R,3R 产物的工程酮还原酶（也在此描述为 2R,3R 选择性酮还原酶）与野生型多肽或另外的工程多肽如 SEQ ID NO :66 相比，具有改善的性质。在一些实施方式中，改善的性质为就将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯还原或转化为 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的立体选择性而言。在一些实施方式中，改善的性质为就其将底物还原至 2R,3R 产物的酶促活性而言。在一些实施方式中，酮还原酶多肽的改善的性质为就其稳定性或热稳定性而言。在一些实施方式中，酮还原酶多肽具有多于一种改善的性质，如增强的立体选择性和改善的酶促活性。

[0116] 如下所详述，可转化底物至 2S,3R 产物的酮还原酶多肽包含其中相应于 SEQ ID NO :2,4 或 86 中的 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。在一些实施方式中，酮还原酶多肽包含其中相应于 SEQ ID NO :2,4 或 86 中的 X94 的残基为脂肪族或极性残基，尤其是丙氨酸或苏氨酸；并且相应于 SEQ ID NO :2,4 或 86 中的 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。在一些实施方式中，2S,3R 选择性酮还原酶多肽包含相应于 SEQ ID NO :2,4

或 86 中的 X94 的残基为脂肪族或极性残基,尤其是丙氨酸或苏氨酸;相应于 SEQ ID NO :2、4 或 86 中的 X199 的残基为受限、极性或脂肪族残基,尤其是组氨酸、天冬酰胺或丙氨酸;并且相应于 SEQ ID NO :2、4 或 86 中的 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。

[0117] 如上所见,本公开中酮还原酶可参考克菲尔乳杆菌、短乳杆菌或小乳杆菌中自然存在的酮还原酶(也称为“ADH”或“醇脱氢酶”)或另外的工程酮还原酶的氨基酸序列而描述。照此,氨基酸残基位置是从开始于起始的蛋氨酸(M)残基(也就是,M代表残基位置1)的酮还原酶决定,尽管本领域内技术人员理解,该起始蛋氨酸残基可通过生物加工机制移除,例如在宿主细胞中或体外翻译系统中,以产生缺乏起始蛋氨酸残基的成熟蛋白质。氨基酸序列中存在的特定的氨基酸或氨基酸变化的氨基酸残基位置在此用术语“Xn”或“位置n”描述,其中n指残基位置。在相同残基位置的氨基酸残基在酮还原酶中不同时,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”。参考序列例如 SEQ ID NO :2 和 SEQ ID NO :4 和 SEQ ID NO :86 的野生型酮还原酶中的氨基酸残基被不同的氨基酸残基替换的取代突变,可由符号“→”表示。在此,在一些实施方式中,突变有时可表达为突变“至”氨基酸类型。例如,SEQ ID NO :4 中残基 199 可突变“至”极性残基。但是短语“至”的使用不排除从一个类型中的一个氨基酸至此类型中另外的氨基酸的突变。例如,SEQ ID NO :4 中残基 199 为脂肪族残基亮氨酸,但是其可突变为不同的脂肪族残基,例如,突变可为“L199A”(199 → A) 突变。克菲尔乳杆菌、短乳杆菌或小乳杆菌中自然存在酮还原酶(也称为“ADH”或“醇脱氢酶”)的氨基酸序列可从已知的编码酮还原酶活性的多核苷酸(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 :86) 中获得。

[0118] 在一些实施方式中,本文的酮还原酶多肽可具有对于参考序列(例如,自然存在的多肽或工程多肽)的许多修饰以产生改善的酮还原酶性质。如本文所用,“修饰”包括氨基酸取代、缺失和插入。可将任何一种修饰或修饰的组合的引入至自然存在的或工程多肽中以产生工程酶。在这样的实施方式中,氨基酸序列的修饰数目可包含一个或多个氨基酸、2个或更多氨基酸、3个或更多氨基酸、4个或更多氨基酸、5个或更多氨基酸、6个或更多氨基酸、8个或更多氨基酸、10个或更多氨基酸、15个或更多氨基酸、或20个或更多氨基酸、多达氨基酸总数的10%、多达参考多肽序列氨基酸总数的15%、多达参考多肽序列氨基酸总数的20%或多达参考多肽序列氨基酸总数的30%。在一些实施方式中,产生改善的酮还原酶性质的自然存在的多肽或工程多肽的修饰的数目可包含从约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个对参考序列的修饰。在一些实施方式中,修饰的数目可为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。

[0119] 在一些实施方式中,修饰包含对参考序列的氨基酸取代。可产生改善的酮还原酶性质的取代可在一个或多个氨基酸、2个或更多氨基酸、3个或更多氨基酸、4个或更多氨基酸、5个或更多氨基酸、6个或更多氨基酸、8个或更多氨基酸、10个或更多氨基酸、15个或更多氨基酸或20个或更多氨基酸、多达参考酶氨基酸总数的10%、多达参考酶氨基酸总数的15%、多达参考酶氨基酸总数的20%或多达参考酶氨基酸总数的30%。在一些实施

方式中,产生改善的酮还原酶性质的自然存在的多肽或工程多肽的取代的数目可包含从约 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 个对参考序列的取代。在一些实施方式中,修饰的数目可为 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。

[0120] 在一些实施方式中,改善的酮还原酶多肽包含与基于 SEQ ID NO:2、4 或 86 的参考序列具有至少约 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 或更高的同一性的氨基酸序列,所述参考序列在相应于 X202 的残基为亮氨酸或缬氨酸,条件是酮还原酶多肽具有其中相应于 X202 的残基为亮氨酸或缬氨酸的氨基酸序列。在一些实施方式中,相应于 X202 的残基为亮氨酸。在一些实施方式中,相比较于参考氨基酸序列,所述酮还原酶多肽可含有在其它残基位置上的一个或多个残基差异。所述差异包括不同的修饰,例如取代、缺失和插入。所述取代可为非保守取代、保守取代或非保守和保守取代的组合。在一些实施方式中,与参考序列相比,酮还原酶可任选地含有在其它氨基酸残基上的约 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。在一些实施方式中,与参考序列相比,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,参考序列为 SEQ ID NO:48。

[0121] 在一些实施方式中,改善的酮还原酶多肽包含与基于 SEQ ID NO:2、4 或 86 的参考序列具有至少约 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 或更高的同一性的氨基酸序列,所述参考序列具有以下特征:相应于 X94 的残基为脂肪族或极性残基,尤其是丙氨酸或苏氨酸;并且相应于 X202 的残基为缬氨酸或亮氨酸;条件是酮还原酶多肽具有包含至少先前的特征的氨基酸序列(也就是说,相应于 X94 的残基为脂肪族或极性残基;并且相应于 X202 的残基为缬氨酸或亮氨酸)。在一些实施方式中,酮还原酶具有其中相应于 X94 的残基为极性残基;并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。在一些实施方式中,酮还原酶具有其中相应于 X94 的残基为苏氨酸并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。在一些实施方式中,与参考序列相比,这些酮还原酶多肽可任选地含有在其它氨基酸残基处的约 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。在一些实施方式中,与参考序列相比,差异的数目可为在其它氨基酸残基处的 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。在一些实施方式中,参考序列为 SEQ ID NO:26 或 28。

[0122] 在一些实施方式中,改善的酮还原酶多肽包含与基于 SEQ ID NO:2、4 或 86 的参考序列具有至少约 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 或更高的同一性的氨基酸序列,所述参考序列具有以下特征:相应于 X94 的残基为脂肪族或极性残基,尤其是丙氨酸或苏氨酸;相应于 X199 的残基为脂肪族、受限或极性残基,尤其是丙氨酸、天冬酰胺或组氨酸;并且相应于 X202 的残基为缬氨酸或亮氨酸;条件是酮还原酶多肽具有包含至少先前的特征的氨基酸序列(也就是说,相应于 X94 的残基为脂肪族或极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;并且相应于

X202 的残基为缬氨酸或亮氨酸)。在一些实施方式中,相应于 X94 的残基为极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;并且相应于 X202 的残基为缬氨酸或亮氨酸。在一些实施方式中,酮还原酶具有以下氨基酸序列,其中相应于 X94 的残基为苏氨酸;相应于 X199 的残基为丙氨酸、天冬酰胺或组氨酸;并且相应于 X202 的残基为缬氨酸或亮氨酸。在一些实施方式中,与参考氨基酸序列相比,所述酮还原酶多肽可含有在其它残基位置上的一个或多个残基差异。所述差异包括不同的修饰,例如取代、缺失和插入。所述取代可为非保守取代、保守取代或非保守和保守取代的组合。在一些实施方式中,与参考序列相比,酮还原酶可任选地具有在其它氨基酸残基处的约 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。在一些实施方式中,与参考序列相比,差异的数目可为在其它氨基酸残基处的 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。在一些实施方式中,参考序列为 SEQ ID NO :22、24 或 30。

[0123] 鉴于以上,所述 2S,3R 选择性酮还原酶可用关于相应于 X94、X199 和 X202 的残基的不同的组合的特征而描述。例如,以相应于 X94 和 X202 的残基的特征为特征的 2S,3R 选择性多肽涉及在此所提供的对特定残基位置的组合的描述。类似地,以相应于 X94、X199 和 X202 的残基的特征为特征的 2S,3R 选择性多肽涉及在此所提供的对特定残基位置的组合的描述。如下文进一步描述的,与参考序列相比,这些酮还原酶可含有在氨基酸序列中的一个或多个另外的特征。

[0124] 在一些实施方式中,2S,3R 选择性酮还原酶多肽包含基于如列于 SEQ ID NO :83、SEQ ID NO :84 或 SEQ ID NO :87 的序列式(或其区域,如残基 90-211)的氨基酸序列。SEQ ID NO :84 为基于克菲尔乳杆菌酮还原酶(SEQ ID NO :4)的野生型氨基酸序列,SEQ ID NO :83 为基于短乳杆菌酮还原酶(SEQ ID NO :2)的野生型氨基酸序列,并且 SEQ ID NO :87 为基于小乳杆菌酮还原酶(SEQ ID NO :86)的野生型氨基酸序列。SEQ ID NO :83、84 或 87 详细说明相应于 X94 的残基为脂肪族或极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;并且相应于 X202 的残基为缬氨酸或亮氨酸。所述序列式进一步详细说明不同的其它残基位置的特征,如下所述。

[0125] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,可含有选自如下的一个或多个另外的特征:相应于 X2 的残基为极性、非极性或脂肪族残基;相应于 X4 的残基为碱性残基或半胱氨酸;相应于 X11 的残基为非极性的、脂肪族、或芳香族残基;相应于 X40 的残基为受限或碱性残基;相应于 X80 的残基为非极性、脂肪族或极性残基;相应于 X86 的残基为非极性、脂肪族或极性残基;相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基;相应于 X105 的残基为非极性的、脂肪族、碱性或酸性残基;相应于 X129 的残基为非极性或极性残基;相应于 X147 的残基为芳香族、非极性或脂肪族残基;相应于 X153 的残基为极性、非极性或脂肪族残基;相应于 X190 的残基为芳香族或受限残基;相应于 X195 的残基为非极性或脂肪族残基;相应于 X196 的残基为非极性或脂肪族残基;相应于 X206 的残基为非极性或芳香族残基;相应于 X226 的残基为非极性或脂肪族残基;相应于 X248 为非极性或碱性残基;相应于 X249 为芳香族残基。在一些实施方式中,氨基酸序列可含有至少所述特征中的两个、三个、四个、五个、六个

或更多。在一些实施方式中,相比较于 SEQ ID NO :2、4 或 86 的参考序列,包含基于 SEQ ID NO :83、84 或 87 的序列式(或其区域)的氨基酸序列的多肽可另外地含有在未被以上 X 标明的残基位置处的一个或多个残基差异。在一些实施方式中,差异可为在未被以上 X 限定的其它氨基酸残基位置的 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。在一些实施方式中,差异的数目可为其它氨基酸残基位置上的 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。在一些实施方式中,所述差异包含保守突变。

[0126] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列可具有相比较于 SEQ ID NO :2、4 或 86 的氨基酸序列的一个或多个保守突变。示例性的保守突变包括例如但是不局限于以下的氨基酸置换:将相应于 X21 的残基缬氨酸(V)置换为另一个脂肪族残基,例如,异亮氨酸;将相应于 X78 的残基谷氨酸(E)置换为另一个酸性残基,例如,天冬氨酸;将相应于 X145 的残基谷氨酸(E)置换为另一个酸性残基,例如,天冬氨酸;将相应于 X153 的残基亮氨酸(L)置换为另一个脂肪族残基,例如,丙氨酸;将相应于 X195 的残基亮氨酸(L)置换为另一个脂肪族残基,例如,缬氨酸;将相应于 X196 的残基置换为另一个脂肪族残基,例如,亮氨酸;将相应于 X226 的残基异亮氨酸(I)置换为另一个脂肪族残基,例如,缬氨酸。

[0127] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,可含有选自如下的一个或多个另外的特征:相应于 X2 的残基为丝氨酸、苏氨酸、谷氨酰胺或天冬酰胺、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,尤其为丝氨酸、苏氨酸或丙氨酸;相应于 X4 的残基为精氨酸、赖氨酸或半胱氨酸,尤其为精氨酸或半胱氨酸;相应于 X11 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸或色氨酸,尤其为异亮氨酸、亮氨酸或苯丙氨酸;相应于 X40 的残基为脯氨酸、组氨酸、精氨酸或赖氨酸,尤其为组氨酸或精氨酸;相应于 X80 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、谷氨酰胺或天冬酰胺,尤其为丙氨酸或苏氨酸;相应于 X86 的残基为丝氨酸、苏氨酸、谷氨酰胺、天冬酰胺、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,尤其为苏氨酸或异亮氨酸;相应于 X96 的残基为丝氨酸、苏氨酸、谷氨酰胺、天冬酰胺、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸或色氨酸,尤其为丝氨酸、天冬酰胺、缬氨酸或苯丙氨酸;相应于 X105 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、精氨酸、赖氨酸、天冬氨酸或谷氨酸,尤其为谷氨酸、赖氨酸、丙氨酸或甘氨酸;相应于 X129 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、谷氨酰胺或天冬酰胺,尤其为蛋氨酸或苏氨酸;相应于 X147 的残基为酪氨酸、苯丙氨酸、色氨酸、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,尤其为苯丙氨酸、蛋氨酸或亮氨酸;相应于 X153 的残基为丝氨酸、苏氨酸、谷氨酰胺、天冬酰胺、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,尤其为亮氨酸、丙氨酸或丝氨酸;相应于 X190 的残基为酪氨酸、苯丙氨酸、色氨酸、组氨酸或脯氨酸,尤其为酪氨酸、组氨酸或脯氨酸;相应于 X195 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,尤其为亮氨酸或缬氨酸;相应于 X196 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,尤其为亮氨酸;相应于 X206 的残基为甘氨酸、蛋氨酸、丙氨酸、

缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸或色氨酸,尤其为蛋氨酸或苯丙氨酸;相应于 X226 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,尤其为异亮氨酸或缬氨酸;相应于 X248 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、赖氨酸或精氨酸,尤其为甘氨酸、赖氨酸或精氨酸;并且相应于 X249 的残基为酪氨酸、苯丙氨酸或色氨酸,尤其为酪氨酸或色氨酸。在一些实施方式中,氨基酸序列可含有所述特征中的至少两个、三个、四个、五个、六个或更多。在一些实施方式中,相比较于 SEQ ID NO:2、4 或 86 的参考序列,包含基于 SEQ ID NO:83、84 或 87 的序列式(或其区域)的氨基酸序列的所述多肽可另外地含有在未被以上 X 标明的残基位置的一个或多个残基差异。在一些实施方式中,差异可为在未被以上 X 限定的其它氨基酸残基位置的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基位置上的 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。在一些实施方式中,所述差异包含保守突变。

[0128] 在一些实施方式中,包含基于 SEQ ID NO:83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,可另外地含有以下特征的一个或多个或至少全部:相应于 X40 的残基为受限或碱性残基,尤其为精氨酸;相应于 X147 的残基为芳香族、非极性或脂肪族残基,尤其为蛋氨酸或亮氨酸。在一些实施方式中,酮还原酶多肽除以上特征外还可含有以下特征中的一个或多个:相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基,尤其为苯丙氨酸或缬氨酸;相应于 X195 的残基为非极性或脂肪族残基,尤其为缬氨酸;相应于 X196 的残基为非极性或脂肪族残基,尤其为亮氨酸;相应于 X226 的残基为非极性或脂肪族残基,尤其为缬氨酸;相应于 X248 为非极性或碱性残基,尤其为精氨酸或赖氨酸;而且相应于 X249 为芳香族残基,尤其为色氨酸。在一些实施方式中,酮还原酶多肽除以上特征外还可含有以下的特征中的一个或多个:相应于 X2 的残基为极性、非极性或脂肪族残基,尤其为丙氨酸;相应于 X4 的残基为碱性残基或半胱氨酸,尤其为半胱氨酸;相应于 X11 的残基为非极性、脂肪族、或芳香族残基,尤其为苯丙氨酸;相应于 X80 的残基为非极性、脂肪族或极性残基,尤其为苏氨酸;相应于 X86 的残基为非极性、脂肪族或极性残基,尤其为异亮氨酸;相应于 X105 的残基为非极性的、脂肪族、碱性或酸性的残基,尤其为甘氨酸;相应于 X129 的残基为非极性或极性残基,尤其为苏氨酸;相应于 X153 的残基为极性、非极性或脂肪族残基,尤其为丙氨酸;相应于 X190 的残基为芳香族或受限残基,尤其为组氨酸或脯氨酸;以及相应于 X206 的残基为非极性或芳香族残基,尤其为苯丙氨酸。如对技术人员明显的,前述的特征的不同组合可用于形成本公开的酮还原酶。

[0129] 在一些实施方式中,包含基于 SEQ ID NO:83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,可另外地含有如下的特征中的一个或多个或至少全部:相应于 X40 的残基为受限或碱性残基,尤其为精氨酸;并且相应于 X147 的残基为芳香族、非极性或脂肪族残基,尤其为蛋氨酸或亮氨酸。在一些实施方式中,与参考序列 SEQ IDNO:2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中, 差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中, 差异包含保守突变。在一些实施方式中, 酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列, 条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0130] 在一些实施方式中, 包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列, 具有相应于 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶, 可另外地含有一个或多个或至少所有如下的特征: 相应于 X96 的残基为极性、芳香族、非极性或非脂肪族残基, 尤其为缬氨酸或苯丙氨酸; 并且相应于 X147 的残基为芳香族、非极性或非脂肪族残基, 尤其为蛋氨酸或亮氨酸。在一些实施方式中, 与参考序列 SEQ ID NO :2、4 或 86 相比, 酮还原酶多肽可另外地含有在其它氨基酸残基位置处的 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。在一些实施方式中, 差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中, 差异包含保守突变。在一些实施方式中, 酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列, 条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0131] 在一些实施方式中, 包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列, 具有相应于 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶, 可另外地含有一个或多个或至少所有如下的特征: 相应于 X40 的残基为受限或碱性残基, 尤其为精氨酸; 相应于 X96 的残基为极性、芳香族、非极性或非脂肪族残基, 尤其为缬氨酸或苯丙氨酸; 并且相应于 X147 的残基为芳香族、非极性或非脂肪族残基, 尤其为蛋氨酸或亮氨酸。在一些实施方式中, 与 SEQ ID NO :2、4 或 86 的参考序列相比较, 酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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。在一些实施方式中, 差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中, 差异包含保守突变。在一些实施方式中, 酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列, 条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0132] 在一些实施方式中, 包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列, 具有相应于 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶, 可另外地含有一个或多个或至少所有如下的特征: 相应于 X96 的残基为极性、芳香族、非极性或非脂肪族残基, 尤其为缬氨酸或苯丙氨酸; 相应于 X147 的残基为芳香族、非极性或非脂肪族残基, 尤其为蛋氨酸或亮氨酸; 相应于 X195 的残基为非极性或非脂肪族残基, 尤其为缬氨酸; 并且相应于 X196 的残基为非极性或非脂肪族残基, 尤其为亮氨酸。在

一些实施方式中,与参考序列 SEQ ID NO :2,4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2,4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0133] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X2 的残基为极性、非极性或脂肪族残基,尤其为丙氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2,4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2,4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0134] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X4 的残基为半胱氨酸或碱性残基,尤其为半胱氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2,4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的残基差异 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2,4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0135] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211,的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X11 的残基为非极性的、脂肪族、或芳香族残基,尤其为苯丙氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2,4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的

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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0136] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X40 的残基为受限或碱性残基,尤其为精氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0137] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X80 的残基为非极性、脂肪族或极性残基,尤其为苏氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0138] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式中或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X86 的残基为非极性、脂肪族或极性残基,尤其为异亮氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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%的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0139] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基,尤其为苯丙氨酸或缬氨酸。在一些实施方式中,与参考序列 SEQ IDNO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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%的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0140] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X105 的残基为非极性的、脂肪族、碱性或酸性的残基,尤其为甘氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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%的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0141] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X129 的残基为非极性或极性残基,尤其为苏氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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%的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0142] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基

90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X147 的残基为芳香族、非极性或脂肪族残基,尤其为蛋氨酸或亮氨酸。在一些实施方式中,与参考序列 SEQ ID NO:2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO:2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0143] 在一些实施方式中,包含基于 SEQ ID NO:83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X153 的残基为极性、非极性或脂肪族残基,尤其为丙氨酸或丝氨酸。在一些实施方式中,与参考序列 SEQ ID NO:2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO:2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0144] 在一些实施方式中,包含基于 SEQ ID NO:83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X190 的残基为芳香族或受限残基,尤其为组氨酸或脯氨酸。在一些实施方式中,与参考序列 SEQ ID NO:2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO:2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0145] 在一些实施方式中,包含基于 SEQ ID NO:83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X195 的残基为非极性或脂肪族残基,尤其为缬氨酸。在一些实施方式中,与参考序列 SEQ ID NO:2、4 或 86

相比, 酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中, 差异的数目可为其它氨基酸残基上的 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。在一些实施方式中, 差异包含保守突变。在一些实施方式中, 酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列, 条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0146] 在一些实施方式中, 包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列, 具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶, 进一步包括至少以下的另外的特征: 相应于 X196 的残基为非极性或非脂肪族残基, 尤其为亮氨酸。在一些实施方式中, 与参考序列 SEQ ID NO :2、4 或 86 相比, 酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中, 差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中, 差异包含保守突变。在一些实施方式中, 酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列, 条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0147] 在一些实施方式中, 包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列, 具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶, 进一步包括至少以下的另外的特征: 相应于 X206 的残基为非极性或非芳香族残基, 尤其为苯丙氨酸。在一些实施方式中, 与参考序列 SEQ ID NO :2、4 或 86 相比, 酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中, 差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中, 差异包含保守突变。在一些实施方式中, 酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列, 条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0148] 在一些实施方式中, 包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列, 具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶, 进一步包括至少以下的另外的特征: 相应于 X226 的残基为非极性或非脂肪族残基, 尤其为缬氨酸。在一些实施方式中, 与参考序列 SEQ ID NO :2、4 或 86 相比, 酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中, 差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0149] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X248 为非极性或碱性残基,尤其为赖氨酸或精氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0150] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X249 为芳香族残基,尤其为色氨酸。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 的同一性的氨基酸序列,条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0151] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的不同组合的特定的特征的改善的酮还原酶,可含有另外地一个或多个或至少所有如下的特征:相应于 X40 的残基为受限或碱性残基,尤其为精氨酸;相应于 X147 的残基为芳香族、非极性或脂肪族残基,尤其为蛋氨酸或亮氨酸。在一些实施方式中,酮还原酶多肽除以上特征外还可含有一个或多个以下的特征:相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基,尤其为缬氨酸;相应于 X195 的残基为非极性或脂肪族残基,尤其为缬氨酸;相应于 X196 的残基为非极性或脂肪族残基,尤其为亮氨酸;相应于 X226 的残基为非极性或脂肪族残基,尤其为缬氨酸;相应于 X248 为非极性或碱性残基,尤其为精氨酸或赖氨酸;而且相应于 X249 为芳香族残基,尤其为色氨酸。在一些实施方式中,酮还原酶多肽除以上特征外还可含有一个或

多个以下的特征：相应于 X2 的残基为极性、非极性或脂肪族残基；相应于 X4 的残基为碱性残基或半胱氨酸；相应于 X11 的残基为非极性的、脂肪族、或芳香族残基；相应于 X80 的残基为非极性、脂肪族或极性残基；相应于 X86 的残基为非极性、脂肪族或极性残基；相应于 X105 的残基为非极性的、脂肪族、碱性或酸性的残基；相应于 X129 的残基为非极性或非极性残基；相应于 X153 的残基为极性、非极性或脂肪族残基；相应于 X190 的残基为芳香族或受限残基；以及相应于 X206 的残基为非极性或非极性残基。如技术人员理解的，前述的特征的不同组合可用于形成本公开中酮还原酶。

[0152] 在一些实施方式中，包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列，具有相应于如此所述的 X202 的残基（也就是，缬氨酸或亮氨酸）的特定的特征的改善的酮还原酶，进一步包括至少以下的另外的特征：相应于 X153 的残基为极性、非极性或脂肪族残基，尤其为丙氨酸或丝氨酸。在一些实施方式中，与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比，酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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（例如，SEQ ID NO :46、52 或 54）。在一些实施方式中，差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中，酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 同一性的氨基酸序列（例如，SEQ ID NO :46、52 或 54），条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0153] 在一些实施方式中，包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列，具有相应于如此所述的 X202 的残基的特定的特征的改善的酮还原酶，进一步包括至少以下的另外的特征：相应于 X147 的残基为芳香族、非极性或非极性残基，尤其为蛋氨酸。在一些实施方式中，与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比，酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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（例如，SEQ ID NO :44）。在一些实施方式中，差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中，酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 同一性的氨基酸序列（例如，SEQ ID NO :44），条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0154] 在一些实施方式中，包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列，具有相应于如此所述的 X202 的残基的特定的特征的改善的酮还原酶，进一步包括至少以下的另外的特征：相应于 X80 的残基为非极性、脂肪族或极性残基，尤其为苏氨酸；并且相应于 X153 的残基为极性、非极性或非极性残基，尤其为丙氨酸或丝氨酸。在一些实施方式中，与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比，酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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（例如，SEQ ID NO :18）。在一些实施方式中，差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 同一性的氨基酸序列(例如,SEQ ID NO :18),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0155] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94 和 X202 的残基的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基,尤其为苯丙氨酸。在一些实施方式中,与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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(例如,SEQ IDNO :16)。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 同一性的氨基酸序列(例如,SEQ ID NO :16),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0156] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X199 和 X202 的残基的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X153 的残基为极性、非极性或脂肪族残基,尤其为丙氨酸或丝氨酸。在一些实施方式中,与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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(例如,SEQ IDNO :42、50、56)。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 同一性的氨基酸序列(例如,SEQ ID NO :42, 50, 56),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0157] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X40 的残基为受限或碱性残基,尤其为精氨酸。在一些实施方式中,与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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(例如,SEQ ID NO :12)。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,酮还原酶多肽包含与基于 SEQ IDNO :2、4 或 86 的、具有以上特征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%同一性的氨基酸序列(例如,SEQ ID NO:12),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0158] 在一些实施方式中,包含基于SEQ ID NO:83、84或87的序列式或其区域如残基90-211的氨基酸序列,具有相应于如此所述的X94、X199和X202的残基的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于X147的残基为芳香族、非极性或脂肪族残基,尤其为蛋氨酸或亮氨酸。在一些实施方式中,与具有以上特征的参考序列SEQ IDNO:2、4或86相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的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(例如,SEQ ID NO:6)。在一些实施方式中,差异的数目可为在其它氨基酸残基上的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。在一些实施方式中,酮还原酶多肽包含与基于SEQ ID NO:2、4或86的、具有以上特征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%同一性的氨基酸序列(例如,SEQ ID NO:6),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0159] 在一些实施方式中,包含基于SEQ ID NO:83、84或87的序列式或其区域如残基90-211的氨基酸序列,具有相应于如此所述的X94、X199和X202的残基的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于X153的残基为极性、非极性或脂肪族残基,尤其为丙氨酸或丝氨酸。在一些实施方式中,与具有以上特征的参考序列SEQ IDNO:2、4或86相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的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(例如,SEQ ID NO:34或36)。在一些实施方式中,差异的数目可为在其它氨基酸残基上的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。在一些实施方式中,酮还原酶多肽包含与基于SEQ ID NO:2、4或86的、具有以上特征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%同一性的氨基酸序列(例如,SEQ IDNO:34或36),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0160] 在一些实施方式中,包含基于SEQ ID NO:83、84或87的序列式或其区域如残基90-211的氨基酸序列,具有相应于如此所述的X94、X199和X202的残基的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于X40的残基为受限或碱性残基,尤其为精氨酸;相应于X147的残基为芳香族、非极性或脂肪族残基,尤其为蛋氨酸或亮氨酸。在一些实施方式中,与具有以上特征的参考序列SEQ ID NO:2、4或86相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的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(例如,SEQ ID NO:10)。在一些实施方式中,差异的数目可为在其它氨基酸残基上的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。在一些实施方式中,酮还原酶多肽包含与基于SEQ ID NO:2、4或86的、具有以上特征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%同一性的氨基酸序列(例如,SEQ ID NO:10),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0161] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X105 的残基为非极性的、脂肪族、碱性或酸性的残基,尤其为甘氨酸;相应于 X153 的残基为极性、非极性或脂肪族残基,尤其为丙氨酸;以及相应于 X206 的残基为非极性或芳香族残基,尤其为苯丙氨酸。在一些实施方式中,与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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(例如,SEQ IDNO :38)。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 同一性的氨基酸序列(例如,SEQ ID NO :38),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0162] 在一些实施方式中,包含基于 SEQ ID NO :83、84 或 87 的序列式或其区域如残基 90-211 的氨基酸序列,具有相应于如此所述的 X94、X199 和 X202 的残基的特定的特征的改善的酮还原酶,进一步包括至少以下的另外的特征:相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基,尤其为丝氨酸、缬氨酸或苯丙氨酸;相应于 X129 的残基为非极性或非极性残基,尤其为苏氨酸;以及相应于 X206 的残基为非极性或芳香族残基,尤其为苯丙氨酸。在一些实施方式中,与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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(例如,SEQ ID NO :40)。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,酮还原酶多肽包含与基于 SEQ ID NO :2、4 或 86 的、具有以上特征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% 同一性的氨基酸序列(例如,SEQ ID NO :40),条件是所述酮还原酶含有具有至少以上特征的氨基酸序列。

[0163] 在一些实施方式中,本公开的改善的酮还原酶包含具有相应于 SEQ IDNO :83、84 或 87 的序列式中的残基 90-211 的区域或结构域的氨基酸序列,其中相应于 X202 的残基为缬氨酸或亮氨酸。在一些实施方式中,与基于 SEQ ID NO :2、4 或 86 的参考序列的相应的结构域相比,相应于残基 90-211 的区域或结构域可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为 1、2、3、4、5、6、7、8、9、10、11、12、14、15、16、18 或约 20 个残基差异。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含具有与相应于基于 SEQ ID NO :2、4 或 86 的在 X202 的残基上具有以上特征的参考序列的残基 90-211 的结构域或区域具有至少 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 的同一性的结构域或区域的氨基酸序列(例如,SEQ ID NO :40),条件是所述酮还原酶的结构域或区域包含具有至少以上特征的

氨基酸序列。

[0164] 在一些实施方式中,本公开的改善的酮还原酶包含相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 90-211 的区域或结构域的氨基酸序列,其中的区域或结构域具有以下特征:相应于 X94 的残基为脂肪族或极性残基,尤其为丙氨酸或苏氨酸;并且相应于 X202 的残基为缬氨酸或亮氨酸。在一些实施方式中,酮还原酶具有相应于 X94 的残基为极性残基并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。在一些实施方式中,酮还原酶具有相应于 X94 的残基为苏氨酸并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。在一些实施方式中,与基于 SEQ ID NO :2、4 或 86 的参考序列的相应的结构域相比,相应于残基 90-211 的区域或结构域可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为 1、2、3、4、5、6、7、8、9、10、11、12、14、15、16、18 或约 20 个残基差异。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含具有与相应于基于 SEQ ID NO :2、4 或 86 的具有以上特征的参考序列的残基 90-211 的结构域或区域具有至少 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 的同一性的结构域或区域的氨基酸序列(例如,SEQ ID NO :40),条件是所述酮还原酶的结构域或区域包含具有至少以上特征的氨基酸序列。

[0165] 在一些实施方式中,本公开的改善的酮还原酶包含相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 90-211 的区域或结构域的氨基酸序列,其具有以下特征:相应于 X94 的残基为脂肪族或极性残基,尤其为丙氨酸或苏氨酸;相应于 X199 的残基为脂肪族、受限或极性残基,尤其为丙氨酸、天冬酰胺或组氨酸;以及相应于 X202 的残基为缬氨酸或亮氨酸。在一些实施方式中,相应于 X94 的残基为极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;以及相应于 X202 的残基为缬氨酸或亮氨酸。在一些实施方式中,酮还原酶具有其中相应于 X94 的残基为苏氨酸;相应于 X199 的残基为丙氨酸、天冬酰胺或组氨酸;以及相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。在一些实施方式中,与基于 SEQ ID NO :2、4 或 86 的参考序列的相应的结构域相比,相应于残基 90-211 的区域或结构域可另外地含有在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为 1、2、3、4、5、6、7、8、9、10、11、12、14、15、16、18 或约 20 个残基差异。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含具有与相应于基于 SEQ ID NO :2、4 或 86 的、具有以上特征的参考序列的残基 90-211 的结构域或区域具有至少 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 的同一性的结构域或区域的氨基酸序列,条件是所述酮还原酶的结构域或区域包含具有至少以上特征的氨基酸序列。

[0166] 在一些实施方式中,包含相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 90-211 的结构域或区域,并具有如此所述的相应于残基 X94、X199 和 X202 的各种残基组合的特定的特征的酮还原酶多肽,可在结构域或区域中进一步包括选自以下的一个或多个特征:相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基;相应于 X105 的残基为非极性、脂肪族、碱性或酸性残基;相应于 X129 的残基为非极性或极性残基;相应于 X147 的残基为芳香族、非极性或脂肪族残基;相应于 X153 的残基为极性、非极性或脂肪族残基;相应于 X190

的残基为芳香族或受限残基；相应于 X195 的残基为非极性或非脂肪族残基；相应于 X196 的残基为非极性或非脂肪族残基；以及相应于 X206 的残基为非极性或非芳香族残基。在一些实施方式中，与基于 SEQ ID NO :2、4 或 86 的参考序列的相应的结构域相比，相应于残基 90-211 的区域或结构域可另外地含有在未被以上 X 标明的其它氨基酸残基位置的约 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。在一些实施方式中，差异的数目可为在结构域内的其它氨基酸残基上的 1、2、3、4、5、6、7、8、9、10、11、12、14、15、16、18 或约 20 个残基差异。在一些实施方式中，差异包含保守突变。

[0167] 在一些实施方式中，包含具有如上所述的相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 90-211 的结构域或区域的酮还原酶多肽，可具有与 SEQ ID NO :2、4 或 86 的相应结构域的氨基酸序列相比在结构域或区域中的一个或多个保守突变。所述保守突变的例子包括氨基酸置换例如但是不局限于：将相应于 X145 的残基谷氨酸 (E) 置换为另一个酸性残基，例如，天冬氨酸；将相应于 X153 的残基亮氨酸 (L) 置换为另一个脂肪族残基，例如，丙氨酸；将相应于 X195 的残基亮氨酸 (L) 的残基置换为另一个脂肪族残基，例如，缬氨酸；以及将相应于 X196 的残基置换为另一个脂肪族残基，例如，亮氨酸。

[0168] 在一些实施方式中，包含相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 90-211 的结构域或区域，并具有如此所述的相应于残基 X94、X199 和 X202 的各种残基组合的特定的特征的酮还原酶多肽，可在结构域或区域中进一步包括选自以下的一个或多个特征：相应于 X96 的残基为丝氨酸、苏氨酸、谷氨酰胺、天冬酰胺、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸或色氨酸，尤其为天冬酰胺、丝氨酸、缬氨酸或苯丙氨酸；相应于 X105 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、精氨酸、赖氨酸、天冬氨酸或谷氨酸，尤其为谷氨酸、赖氨酸、丙氨酸或甘氨酸；相应于 X129 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、谷氨酰胺或天冬酰胺，尤其为蛋氨酸或苏氨酸；相应于 X147 的残基为酪氨酸、苯丙氨酸、色氨酸、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸，尤其为苯丙氨酸、蛋氨酸或亮氨酸；相应于 X153 的残基为丝氨酸、苏氨酸、谷氨酰胺、天冬酰胺、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸，尤其为亮氨酸、丙氨酸或丝氨酸；相应于 X190 的残基为酪氨酸、苯丙氨酸、色氨酸、组氨酸或脯氨酸，尤其为酪氨酸、组氨酸或脯氨酸；相应于 X195 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸，尤其为亮氨酸或缬氨酸；相应于 X196 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸，尤其为亮氨酸；相应于 X206 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸或色氨酸，尤其为蛋氨酸或苯丙氨酸。在一些实施方式中，与基于 SEQ ID NO :2、4 或 86 的参考序列的相应的结构域相比，相应于残基 90-211 的区域或结构域可另外地含有在未被以上 X 标明的其它氨基酸残基位置的约 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。在一些实施方式中，差异的数目可为在结构域内的其它氨基酸残基上的 1、2、3、4、5、6、7、8、9、10、11、12、14、15、16、18 或约 20 个残基差异。在一些实施方式中，差异包含保守突变。

[0169] 在一些实施方式中，包含相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 90-211 的结构域或区域，并具有如此所述的相应于残基 X94、X199 和 X202 的各种残基组合的特定的特征的酮还原酶多肽，可在结构域或区域中进一步包括以下特征：相应于 X147 的残基为

芳香族、非极性或脂肪族残基,尤其为蛋氨酸或亮氨酸。在一些实施方式中,与基于 SEQ ID NO:2、4 或 86 的参考序列的相应的结构域相比,相应于残基 90-211 的区域或结构域可另外地含有在其它氨基酸残基位置上的约 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。在一些实施方式中,差异的数目可为 1、2、3、4、5、6、7、8、9、10、11、12、14、15、16、18 或约 20 个残基差异。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含具有与相应于基于 SEQ ID NO:2、4 或 86 的具有以上特征的参考序列的残基 90-211 的结构域或区域具有至少 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 的同一性的结构域或区域的氨基酸序列,条件是所述酮还原酶的结构域或区域包含具有至少以上特征的氨基酸序列。

[0170] 在一些实施方式中,包含相应于 SEQ ID NO:83、84 或 87 的序列式中的残基 90-211 的结构域或区域,并具有如此所述的相应于残基 X94、X199 和 X202 的各种残基组合的特定的特征的酮还原酶多肽,可在结构域或区域中进一步包括以下特征:相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基,尤其为缬氨酸或苯丙氨酸;并且相应于 X147 的残基为芳香族、非极性或脂肪族残基,尤其为蛋氨酸或亮氨酸。在一些实施方式中,与基于 SEQ ID NO:2、4 或 86 的参考序列的相应的结构域相比,相应于残基 90-211 的区域或结构域可另外地含在其它氨基酸残基位置上的 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。在一些实施方式中,差异的数目可为 1、2、3、4、5、6、7、8、9、10、11、12、14、15、16、18 或约 20 个残基差异。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶多肽包含具有与相应于基于 SEQ ID NO:2、4 或 86 的、具有以上特征的参考序列的残基 90-211 的结构域或区域具有至少 85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 的同一性的结构域或区域的氨基酸序列,条件是所述酮还原酶的结构域或区域包含至少具有以上特征的氨基酸序列。

[0171] 在一些实施方式中,酮还原酶多肽可进一步包含相应于 SEQ ID NO:83、84 或 87 的序列式中的残基 1-89 的结构域或区域,其中所述结构域或区域可具有以下特征中的一个或多个:相应于 X2 的残基为极性、非极性或脂肪族残基;相应于 X4 的残基为碱性残基或半胱氨酸;相应于 X11 的残基为非极性、脂肪族、或芳香族残基;相应于 X40 的残基为受限或碱性残基;相应于 X80 的残基为非极性、脂肪族或极性残基;并且相应于 X86 的残基为非极性、脂肪族或极性残基。在一些实施方式中,包含相应于 SEQ ID NO:83、84 或 87 的序列式的残基 1-89 的区域或结构域的多肽可另外地具有与参考序列 SEQ ID NO:2、4 或 86 相比较在未用以上 X 标明的残基位置上的一个或多个残基差异。在一些实施方式中,与参考序列 SEQ ID NO:2、4 或 86 相比,差异可为在未被以上 X 标明的其它氨基酸残基位置的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基位置上的 1、2、3、4、5、6、7、8、9、10、11、12、14、15 或 16 个残基差异。在一些实施方式中,差异包含保守突变。

[0172] 在一些实施方式中,酮还原酶多肽可进一步包含相应于 SEQ ID NO:83、84 或 87 的序列式中的残基 1-89 的结构域或区域,其中所述结构域或区域可具有以下特征中的一个或多个:相应于 X2 的残基为丝氨酸、苏氨酸、谷氨酰胺或天冬酰胺、甘氨酸、蛋氨酸、丙氨

酸、缬氨酸、亮氨酸或异亮氨酸,尤其为丝氨酸、苏氨酸或丙氨酸;相应于 X4 的残基为精氨酸或赖氨酸或半胱氨酸,尤其为精氨酸或半胱氨酸;相应于 X11 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、酪氨酸、苯丙氨酸或色氨酸,尤其为异亮氨酸、亮氨酸或苯丙氨酸;相应于 X40 的残基为脯氨酸、组氨酸、精氨酸或赖氨酸,尤其为组氨酸或精氨酸;相应于 X80 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、谷氨酰胺或天冬酰胺,尤其为丙氨酸或苏氨酸;并且相应于 X86 的残基为丝氨酸、苏氨酸、谷氨酰胺、天冬酰胺、甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸,尤其为苏氨酸或异亮氨酸。在一些实施方式中,包含相应于 SEQ ID NO :83、84 或 87 的序列式的残基 1-89 的区域或结构域的多肽可另外地具有与参考序列 SEQ ID NO :2、4 或 86 相比较在未用以上 X 标明的残基位置上的一个或多个残基差异。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,差异可为在未被以上 X 标明的其它氨基酸残基位置的 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。在一些实施方式中,差异的数目可为在其它氨基酸残基位置上的 1、2、3、4、5、6、7、8、9、10、11、12、14、15 或 16 个残基差异。在一些实施方式中,差异包含保守突变。

[0173] 在一些实施方式中,酮还原酶多肽可进一步包含相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 1-89 的结构域或区域,其中所述结构域或区域可具有至少以下特征:相应于 X40 的残基为受限或碱性残基,尤其为精氨酸。在一些实施方式中,包含相应于 SEQ ID NO :111、112 或 139 的序列式的残基 1-89 的区域或结构域的多肽可另外地具有与参考序列 SEQ ID NO :2、4 或 114 相比较在其它残基位置上的一个或多个残基差异。在一些实施方式中,与基于 SEQ ID NO :2、4 或 86 的参考序列的相应区域或结构域相比,相应于残基 1-89 的区域或结构域可另外地含有在其它氨基酸残基位置的 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。在一些实施方式中,差异的数目可为在结构域中其它氨基酸残基的 1、2、3、4、5、6、7、8、9、10、11、12、14、15 或 16 个残基差异。在一些实施方式中,差异包含保守突变。在一些实施方式中,酮还原酶区域或结构域包含具有至少以上特征的氨基酸序列,并且其中所述氨基酸序列与相应于基于 SEQ ID NO :2、4 或 86 的具有以上特征的参考序列的残基 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% 的同源性。

[0174] 在一些实施方式中,酮还原酶多肽可进一步包含相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 212-252 的结构域或区域,其中结构域或区域可具有以下特征中的一个或多个:相应于 X226 的残基为非极性或非脂肪族残基;相应于 X248 为非极性或非碱性残基;并且相应于 X249 为芳香族残基。在一些实施方式中,包含相应于 SEQ ID NO :83、84 或 87 的序列式的残基 212-252 的区域或结构域的多肽可另外地具有与参考序列 SEQ ID NO :2、4 或 86 相比较在未用以上 X 标明的残基位置上的一个或多个残基差异。在一些实施方式中,与参考序列 SEQ ID NO :2、4 或 86 相比,差异可为在未被以上 X 标明的其它氨基酸残基位置的 1-2、1-3、1-4、1-5、1-6、1-7、1-8、1-9 或 1-10 个残基差异。在一些实施方式中,差异的数目可为在其它氨基酸残基位置的 1、2、3、4、5、6、7、8、9 或 10 个残基差异。在一些实施方式中,差异包含保守突变。

[0175] 在一些实施方式中,酮还原酶多肽可进一步包含相应于 SEQ ID NO :83、84 或 87 的序列式中残基 212-252 的结构域或区域,其中所述结构域或区域可具有以下特征中的一个

或多个；相应于 X226 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸或异亮氨酸，尤其为缬氨酸；相应于 X248 的残基为甘氨酸、蛋氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、赖氨酸或精氨酸，尤其为甘氨酸、赖氨酸或精氨酸；并且相应于 X249 的残基为酪氨酸、苯丙氨酸或色氨酸，尤其为酪氨酸或色氨酸。在一些实施方式中，包含相应于 SEQ ID NO :83、84 或 87 的序列式的残基 212-252 的区域或结构域的多肽可另外地具有与参考序列 SEQ ID NO :2、4 或 86 相比较在未用以上 X 标明的残基位置上一个或多个残基差异。在一些实施方式中，与参考序列 SEQ ID NO :2、4 或 86 相比，差异可为在未被以上 X 标明的其它氨基酸残基位置的 1-2、1-3、1-4、1-5、1-6、1-7、1-8、1-9 或 1-10 个残基差异。在一些实施方式中，差异的数目可为在其它氨基酸残基位置的 1、2、3、4、5、6、7、8、9 或 10 个残基差异。在一些实施方式中，差异包含保守突变。

[0176] 在一些实施方式中，酮还原酶多肽可进一步包含相应于 SEQ ID NO :83、84 或 87 的序列式中的残基 212-252 的结构域或区域，其中结构域或区域具有至少以下特征：相应于 X226 的残基为非极性或非脂肪族残基，尤其为缬氨酸。在一些实施方式中，包含相应于 SEQ ID NO :83、84 或 87 的序列式的残基 212-252 的区域或结构域的多肽可另外地具有与具有以上特征的参考序列 SEQ ID NO :2、4 或 86 相比较在其它残基位置上的一个或多个残基差异。在一些实施方式中，与基于 SEQ ID NO :2、4 或 86 的参考序列的相应结构域相比，相应于残基 212-252 的区域或结构域可另外地含有在其它氨基酸残基位置的 1-2、1-3、1-4、1-5、1-6、1-7、1-8、1-9 或 1-10 个残基差异。在一些实施方式中，差异的数目可为在结构域中其它氨基酸残基的 1、2、3、4、5、6、7、8、9 或 10 个残基差异。在一些实施方式中，差异包含保守突变。在一些实施方式中，酮还原酶的区域或结构域包含具有至少以上特征的氨基酸序列，并且其中所述氨基酸序列与相应于基于 SEQ ID NO :2、4 或 86 的具有以上特征的参考序列的残基 212-252 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。

[0177] 下表 2 提供示例性的 2S, 3R 选择性酮还原酶，SEQ ID NO :1-62，以及其相应的活性。除非另行标明，以下序列为衍生自野生型克菲尔乳杆菌酮还原酶序列 (SEQ ID NO :3 和 4)。在下表 2 中，每行列出两个 SEQ ID NO，其中奇数指编码由偶数提供的氨基酸序列的核苷酸序列。列出突变数的列为与 SEQ ID NO :4 的克菲尔乳杆菌 KRED 氨基酸序列相比较的氨基酸取代的数目，并且具体的取代列于“自克菲尔乳杆菌的突变”列中。在活性列中，一个“+”相当于与具有 SEQ ID NO :48 的氨基酸序列的多肽的转化底物至具式 (II) 的产物的能力相比 1-15 倍的提高。两个加号“++”表明多肽与 SEQ ID NO :48 相比约为 15 至 30 倍的改善。三个加号“+++”表明多肽与 SEQ ID NO :48 相比约为 30 至 40 倍的改善。四个加号“++++”表明多肽与 SEQ ID NO :48 相比约为 40 至 50 倍的改善，而且五个加号“+++++”表明多肽与 SEQ ID NO :48 相比约为高于 50 倍的改善。稳定性列中“+”号表明多肽能够在 40°C 热处理 21 小时后保持转化底物至具式 (II) 的产物的酶促活性。对于选择性列，单个加号“+”表明多肽能够以约 60-89% 的立体异构体过量转化底物至具式 (II) 的产物；两个加号“++”表明多肽能够以约 90-94% 的立体异构体过量转化底物至具式 (II) 的产物；三个加号“+++”表明多肽能够以约 95-99% 的立体异构体过量转化底物至具式 (II) 的产物；而且四个加号标记“++++”表明多肽能够以高于约 99% 的立体异构体过量转化底物至具式 (II) 的产物。相应地，在一些实施方式中，2S, 3R 选择性酮还原酶可包含相应于

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 或 62 的序列。

[0178] 表 2

[0179]

SEQ ID NO	与SEQ ID NO:2的残基差异	从克菲尔乳杆菌的突变数	活性	稳定性	选择性
1/2	短乳杆菌				
3/4	克菲尔乳杆菌				
47/48	A202V	1	比野生型有改善		+
37/38	A94T; E105G; L153A; L199A; A202L; M206F	6	+		+
15/16	A94T; S96F; A202V	3	+		+++
55/56	L153A; L199A; A202L	3	+		+++
	T86I; L199N; A202L	3	+		+++

[0180]

表2: 序列和相应的活性改善的列表					
SEQ ID NO	与SEQ ID NO:2的残基差异	从克菲尔乳杆菌的突变数	活性	稳定性	选择性
57/58					
51/52	L153A; A202L	2	+		+++
53/54	L153A; A202V	2	+		+
31/32	A94T; L199A; A202V	3	++		++++
33/34	A94T; L153A; L199H; A202L	4	+++		++++
49/50	L153A; L199H; A202L	3	++		+++
19/20	A94T; L199N; A202V	3	+++		++++
45/46	L153S; A202L	2	+		+
35/36	A94T; L153A; L199A; A202V	4	+		++
25/26	A94T; A202L	2	+++		+++
27/28	A94T; A202V	2	++		+++
29/30	A94T; L199A; A202L	3	+++		++++
21/22	A94T; L199H; A202L	3	++++		++++
23/24	A94T; L199H; A202V	3	+++		++++
41/42	L153A; L199N; A202L	3	+		+++
39/40	A94T; S96F; M129T; A202V; M206F	5	+		++
17/18	A80T; L153A; A202V;	3	+		+++
43/44	F147M; A202V	2	+	+	
9/10	H40R; A94T; F147L; L199H; A202L	5	+++++	+	++++
11/12	H40R; A94T; L199H; A202L	4	+++++		++++
5/6	A94T; F147L; L199H; A202L	4	+++++	+	++++
7/8	A94T; L199H; A202L	3	+++++		++++
59/60	I11F; H40R; A94F; S96V; F147M; L195V; V196L; L199W; I226V; G248K; Y249W	11	++++		++++

[0181]

SEQ ID NO	与SEQ ID NO:2的残基差异	从克菲尔乳杆菌的突变数	活性	稳定性	选择性
61/62	T2A; R4C; H40R; A94G; S96V; F147M; V196L; L199W; I226V; G248K; Y249W	11	+++		++++
13/14	H40R; A94F; S96V; F147M; L195V; V196L; L199W; I226V; Y249W	9	++++		++++

[0182] 在一些实施方式中,改善的 2S,3R 选择性酮还原酶包含与基于 SEQ IDNO :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 或 62 的参考序列具有至少约 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,条件是所述酮还原酶的氨基酸序列具有至少以下特征:相应于 X94 的残基为脂肪族或极性残基,尤其为丙氨酸或苏氨酸;对应于 X199 的残基为脂肪族、受限或极性残基,尤其为丙氨酸、组氨酸或天冬酰胺;并且相应于 X202 的残基为缬氨酸或亮氨酸。在一些实施方式中,与参考序列相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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、1-25、1-30、1-35 或约 1-40 个残基差异。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。

[0183] 在一些实施方式中,改善的 2S,3R 选择性酮还原酶包含与基于 SEQ IDNO :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 或 62 的参考序列具有至少约 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,条件是所述酮还原酶的氨基酸序列与 SEQ ID NO :2 或 4 或 86 相比包含表 2 中所列的任一多肽序列含有的任何一组突变。在一些实施方式中,与参考序列相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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、1-25、1-30、1-35 或约 1-40 个残基差异。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。

[0184] 在一些实施方式中,酮还原酶多肽能够以至少约 85% 的立体异构体过量百分比或以高于野生型克菲尔乳杆菌 KRED (SEQ ID NO :4) 的立体异构体过量百分比,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯,转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 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 或 62 的氨基酸序列的多肽。

[0185] 在一些实施方式中,酮还原酶多肽能够以至少约 60-89% 的立体异构体过量百分比并以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 1-15 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S,3R-2- 苯甲酰氨基甲基 -3- 羟

基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、46、50、52、54、56、58、60 或 62 的氨基酸序列的多肽。因为具有 SEQ ID NO :48 的氨基酸序列的多肽能够以一定速率（例如，在 50% IPA, pH8 中, 用约 10g/L KRED 在 20 小时内将 1g/L 底物 100% 转化）和与野生型 (SEQ ID NO :4) 相比较改善的立体选择性将底物转化为产物, 所以与 SEQ ID NO :48 相比改善的本文所述的多肽也比野生型改善。

[0186] 在一些实施方式中, 所述酮还原酶多肽能够以至少约 90-94% 的立体异构体过量百分比, 将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、40、42、50、52、56、58、60 或 62 的氨基酸序列的多肽。

[0187] 在一些实施方式中, 所述酮还原酶多肽能够以至少约 95-99% 的立体异构体过量百分比, 将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯, 转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、42、50、52、56、58、60 或 62 的氨基酸序列的多肽。

[0188] 在一些实施方式中, 所述酮还原酶多肽能够以至少约 99% 的立体异构体过量百分比将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、30、32、34、60 或 62 的氨基酸序列的多肽。

[0189] 在一些实施方式中, 所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 15-30 倍的速率, 将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、26、28、30、32、34、50、60 或 62 的氨基酸序列的多肽。

[0190] 在一些实施方式中, 所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 30-40 倍的速率, 将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、26、30、34、60 或 62 的氨基酸序列的多肽。

[0191] 在一些实施方式中, 所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 40-50 倍的速率, 将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、22 或 60 的氨基酸序列的多肽。

[0192] 在一些实施方式中, 所述酮还原酶多肽能够以比具有 SEQ ID NO :48 的氨基酸序列的多肽可达到的速率高至少约 50 倍的速率, 将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10 或 12 的氨基酸序列的多肽。

[0193] 在一些实施方式中, 所述酮还原酶多肽能够相比具有 SEQ ID NO :48 的氨基酸序列

的多肽可达到的速率高至少约 50 倍的速率并且以至少 99% 的立体异构体过量百分比, 将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10 和 12 的氨基酸序列的多肽。

[0194] 在一些实施方式中, 所述酮还原酶多肽能够在 40°C 热处理 21 小时后将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、10 或 44 的氨基酸序列的多肽。

[0195] 如上所示, 在一些实施方式中, 所述酮还原酶能够立体选择性地 将具式 (I) 的底物还原或转化为具式 (III) 的产物。下表 3 提供示例性的 2R, 3R 特异性酮还原酶 SEQ ID NO 63-82, 以及其相应的活性。除非另行标明, 以下序列衍生自野生型克菲尔乳杆菌酮还原酶序列 (SEQ ID NO :3 和 4)。在下表 3 中, 每行列出两个 SEQ ID NO, 其中奇数指编码由偶数提供的氨基酸序列的核苷酸序列。列出突变数的列为与 SEQ ID NO :4 的克菲尔乳杆菌 KRED 氨基酸序列相比较的氨基酸取代的数目, 并且列 " 序列 - 编码突变 " 列出与 SEQ ID NO :4 相比较的取代。在活性列中, 一个 " + " 相当于与具有 SEQ ID NO :66 的氨基酸序列的多肽的转化底物至具式 (III) 的产物的能力相比约 1 倍的提高。两个加号 " ++ " 表明多肽与 SEQ ID NO :66 相比约为 1 至 2 倍的改善。三个加号 " +++ " 表明多肽与 SEQ ID NO :66 相比约为 5 倍的改善。对于选择性列, 单个加号 " + " 表明多肽能够以小于约 85% 的立体异构体过量转化底物至具式 (III) 的产物 ; 两个加号 " ++ " 表明多肽能够以大于约 85% 的立体异构体过量转化底物至具式 (III) 的产物。相应地, 在一些实施方式中, 2S, 3R 选择性酮还原酶可包含相应于 SEQ ID NO :64、66、68、70、72、74、76、78、80 或 82 的序列。

[0196] 表 3 : 序列和相应活性改善的列表

[0197]

SEQ ID NO	与 SEQ ID NO :4 的残基差异	从克菲尔乳杆菌的突变数	活性	稳定性
65/66	H40R ; A94G ; S96V ; E145F ; F147M ; Y190P ; V196L ; L199W ; I225V ; Y249W	10	+	+
73/74	I11F ; H40R ; A94E ; S96V ; E145F ; F147M ; Y190P ; L195V ; V196L ; L199W ; I226V ; Y249H	12	++	++
81/82	D3V ; A10T ; H40R ; A94G ; S96V ; F147M ; Y190P ; V196L ; L199W ; I226V ; G248K ; Y249H	12	+	++
67/68	H40R ; A94F ; S96V ; E145F ; F147M ; Y190P ; L195V ; V196L ; L199W ; I226V ; G248R ; Y249W	12	++	++

SEQ ID NO	与 SEQ ID NO :4 的残基差异	从克菲尔乳杆菌的突变数	活性	稳定性
77/78	I11L ;H40R ;A94E ;S96V ;F147M ;Y190H ;V196L ;I226V ;G248K ;Y249H	10	+++	++
71/72	H40R ;T54A ;A94F ;S96V ;E105K ;E145D ;F147M ;V196L ;L199W ;I226V ;Y249W	11	+++	++
75/76	I11F ;H40R ;A94G ;S96V ;E145F ;F147M ;Y190H ;L195V ;V196L ;L199W ;A202V ;I226V ;Y249H ;A251T	14	+++	++
69/70	H40R ;E78D ;A94E ;S96V ;F147M ;Y190H ;L195V ;V196L ;I226V ;Y249H ;T250Y	11	+++	+
79/80	K8N ;V9G ;I11F ;H40R ;A94G ;S96V ;E145F ;F147M ;Y190P ;V196L ;I226V ;G248K ;Y249R	13	+++	+
63/64	V12I ;H40R ;A94E ;S96V ;F147M ;Y190P ;L195V ;V196L ;L199W ;I226V ;G248R ;Y249W	12	+++	+

[0198] 在一些实施方式中,改善的 2R,3R 选择性酮还原酶包含与基于 SEQ ID NO :64、66、68、70、72、74、76、78、80 或 82 的参考序列具有至少约 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,条件是所述酮还原酶的氨基酸序列与 SEQ ID NO :2 或 4 或 86 相比包含表 3 中所列的任一多肽序列含有的任何一组突变。在一些实施方式中,与参考序列相比,酮还原酶多肽可另外地含有在其它氨基酸残基位置的 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、1-25、1-30、1-35 或约 1-40 个残基差异。在一些实施方式中,差异的数目可为在其它氨基酸残基上的 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。在一些实施方式中,差异包含保守突变。

[0199] 在一些实施方式中,所述 2R,3R 选择性酮还原酶多肽能够以至少约 85% 的立体异构体过量百分比,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :68、72、74、76、78 或 82 的氨基酸序列的多肽。

[0200] 在一些实施方式中,所述 2R,3R 选择性酮还原酶多肽能够以比具有 SEQ ID NO :66 的氨基酸序列的多肽可达到的速率高至少约 1 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :64、68、70、72、74、76、78、80 或 82 的氨基酸序列的多肽。因为具有 SEQ ID NO :66 的氨基酸序列的多肽能够以高于野生型克菲尔乳杆菌 KRED (SEQ

ID NO :4) 的立体异构体过量以及速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯,所以比 SEQ ID NO :66 改善的任何多肽也比野生型克菲尔乳杆菌 KRED 改善。

[0201] 在一些实施方式中,所述 2R,3R 选择性酮还原酶多肽能够以比具有 SEQ ID NO :66 的氨基酸序列的多肽可达到的速率高至少约 1-2 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :64、68、70、72、74、76、78 或 80 的氨基酸序列的多肽。

[0202] 在一些实施方式中,所述 2R,3R 选择性酮还原酶多肽能够以比具有 SEQ ID NO :66 的氨基酸序列的多肽可达到的速率高至少约 5 倍的速率,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :64、70、72、76、78 或 80 的氨基酸序列的多肽。

[0203] 在一些实施方式中,所述 2R,3R 选择性酮还原酶多肽能够以比具有 SEQ ID NO :66 的氨基酸序列的多肽可达到的速率高至少约 5 倍的速率以及至少约 85% 的立体异构体过量百分比,将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。可用的示例性的多肽包括但不限于包含相应于 SEQ ID NO :72 或 78 的氨基酸序列的多肽。

[0204] 在一些实施方式中,本公开的 2S,3R 选择性酮还原酶多肽可包含与 SEQ ID NO :2 或 SEQ ID NO :4 或 SEQ ID NO :86 (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,条件是相应于 SEQ ID NO :2 或 SEQ ID NO :4 或 SEQ ID NO :86 的残基 X202 的残基为缬氨酸或亮氨酸,相应于 SEQ ID NO :2 或 4 或 86 的残基 94 的残基为苏氨酸,并且相应于 SEQ ID NO :2 或 4 或 86 的残基 199 的残基为组氨酸,并且所述 2S,3R 选择性酮还原酶多肽另外地具有以下取代中的一个或多个,以使得所述多肽相比于野生型克菲尔乳杆菌酮还原酶或另外的工程酮还原酶(例如 SEQ ID NO :48) 为进一步改善的(就立体选择性、酶促活性和 / 或热稳定性而言): 2 → A ; 4 → C ; 11 → F ; 40 → H ; 80 → T ; 86 → I ; 96 → F, V ; 105 → G ; 129 → T ; 147 → M, L ; 153 → A, S ; 195 → V ; 196 → L ; 206 → F ; 226 → V ; 248 → K ; 和 249 → W。

[0205] 在一些实施方式中,本文所述的 2S,3R 酮还原酶多肽可包含与 SEQ ID NO :2、4 或 86 (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,条件是相应于 SEQ ID NO :2、4 或 86 的残基 202 的残基为缬氨酸或亮氨酸,相应于 SEQ ID NO :2、4 或 86 的残基 94 的残基为苏氨酸,以及相应于 SEQ ID NO :2、4 或 86 的残基 199 的残基为组氨酸,并且所述 2S,3R 酮还原酶多肽另外地具有以下取代中的一个或多个,以使得所述多肽相比于野生型克菲尔乳杆菌酮还原酶或另外的工程酮还原酶(例如 SEQ ID NO :48) 为进一步改善的(就立体选择性、酶促活性和 / 或热稳定性而言): 40 → H ; 和 147 → L, M。

[0206] 本领域技术人员会明白,除非另行标明,以上确定的氨基酸类别中的一些类别并非互相排斥的。因此,具有表现两种或更多的物理化学性质的侧链的氨基酸可包含在多个类别中。任何氨基酸或残基的合适的分类对于本领域专业人员是明显的,尤其是按照在此

提供的详细公开。

[0207] 在一些实施方式中,改善的工程酮还原酶包括自然存在的酮还原酶多肽的缺失或其它工程酮还原酶多肽的缺失。在一些实施方式中,本文所述的每个改善的工程酮还原酶可包含本文所述的多肽的缺失。因此,对于本公开中酮还原酶多肽的每个实施方式,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、多达酮还原酶多肽氨基酸总数的10%、多达酮还原酶多肽氨基酸总数的10%、多达酮还原酶多肽氨基酸总数的20%、或多达酮还原酶多肽氨基酸总数的30%,只要保持酮还原酶活性的功能活性。在一些实施方式中,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。在一些实施方式中,缺失的数目可为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。在一些实施方式中,缺失可包括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。

[0208] 如本文所述的,本公开的酮还原酶多肽可为融合多肽的形式,其中酮还原酶多肽融合至其它多肽,例如,诸如但不局限于,抗体标签(例如,myc表位)、纯化序列(例如,组氨酸标签)和细胞定位信号(例如,分泌信号)。因此,酮还原酶多肽可以与其它多肽融合或不融合而使用。

[0209] 本文所述的多肽不局限于遗传编码的氨基酸。除了遗传编码的氨基酸,本文所述的多肽可以整体或部分包括自然存在的和/或合成的非编码氨基酸。本文所述的多肽可包含的某些通常遇见的非编码氨基酸包括但不限于:遗传编码氨基酸的D-立体异构体;2,3-二氨基丙酸(Dpr); $\alpha$ -氨基异丁酸(Aib); $\epsilon$ -氨基己酸(Aha); $\delta$ -氨基戊酸(Ava);N-甲基甘氨酸或肌氨酸(MeGly或Sar);鸟氨酸(Orn);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(Nle);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(Opef);3,4-二氯苯丙氨酸(Mpcf);2,4-二氟苯丙氨酸(Opff);3,4-二氯苯丙氨酸(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(InAla);萘-2-基丙氨酸(2nAla);噻唑基丙氨酸(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);苯乙烯丙氨酸(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(Afp);青霉素(Pen);1,2,3,4-四氢异喹啉-3-羧酸(Tic); $\beta$ -2-噻吩基丙氨酸(Thi);蛋氨酸亚砷(Mso);N(w)-硝基精氨酸(nArg);高赖

氨酸 (hLys) ; 膦酰甲基苯丙氨酸 (pmPhe) ; 磷酸丝氨酸 (pSer) ; 磷酸苏氨酸 (pThr) ; 高天冬氨酸 (hAsp) ; 高谷氨酸 (hGlu) ; 1- 氨基环戊-(2 或 3)- 烯-4 羧酸 ; 2- 哌啶酸 (PA) , 氮杂环丁基-3- 羧酸 (ACA) ; 1- 氨基环戊基-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- 乙酰基赖氨酸 (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) ; 羟基脯氨酸 (Hyp) 和高脯氨酸 (hPro) 。可包含于本文所述的多肽中的另外的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, 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 页以及其中所引用的文献中提供的各种氨基酸, 其全部以引用的形式并入本文) 。这些氨基酸可为 L- 或 D- 构型的。

[0210] 本领域技术人员将理解, 含有侧链保护基团的氨基酸或残基也可构成本文所述的多肽。在此情况下属于芳香族类别的此类保护的氨基酸的非限制性的例子包括但不限于 ( 保护基团列于括号中) : Arg (tos) 、Cys ( 苄基) 、Cys ( 硝基吡啶亚磺酰基) 、Glu (  $\delta$ - 苄基酯) 、Gln ( 占吨基) 、Asn (N-  $\delta$ - 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) 。

[0211] 可包含于本文所述的多肽中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 ; 吡啶-3- 羧酸 ; 高脯氨酸 (hPro) ; 和 1- 氨基环戊烷-3- 羧酸。

[0212] 如上文所述, 引入至自然存在的多肽以产生工程酮还原酶的各种修饰可针对酶的性质。

### [0213] 1.3 编码工程酮还原酶的多核苷酸

[0214] 在另一方面, 本公开提供编码工程酮还原酶的多核苷酸。多核苷酸可与控制基因表达的一种或多种异源调控序列可操作地连接以产生能够表达多肽的重组多核苷酸。包含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合适的宿主细胞以表达相应的酮还原酶多肽。

[0215] 因为对相应于不同氨基酸的密码子的认识, 蛋白质序列的可用性提供对于能够编码所述蛋白序列的所有多核苷酸的描述。其中相同氨基酸由可选择的或同义密码子编码的遗传密码的简并性允许产生极大数目的核酸, 所有这些核酸编码本发明的改善的酮还原酶。因此, 确定特殊的氨基酸序列后, 本领域的技术人员可以以一种不改变蛋白质的氨基酸序列的方式通过仅仅修饰一个或多个密码子的序列来产生任何数目的不同的核酸。在这点上, 本公开尤其涵盖可通过基于可能的密码子选择而选取组合来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 1 中呈现的氨基酸序列, 所有这些改变被认为具体地公开。

[0216] 在不同的实施方式中, 密码子被优选地选择以适应在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于在细菌中表达基因 ; 酵母中使用的优选密码子被用于在酵母中的表达 ; 而且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。当作例子, SEQ ID NO :1 的多核苷酸已经为在大肠埃希氏菌 (E. coli) 中的表达而进行密码子优化, 但仍编码克菲尔乳杆菌的自然存在的酮还原酶。

[0217] 在某些实施方式中,不是所有的密码子需要被取代以优化酮还原酶的密码子使用,因为自然序列将包含优选的密码子而且因为优选的密码子的使用可能不被所有的氨基酸残基需要。因此,编码酮还原酶的密码子优化的多核苷酸可包含位于全长编码区的约40%、50%、60%、70%、80%或高于90%的密码子位置的优选的密码子。

[0218] 在一些实施方式中,所述多核苷酸包含编码酮还原酶多肽的核苷酸序列,所述酮还原酶多肽具有与本文所述的任何参考工程酮还原酶具有至少约80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%或更多的序列同一性的氨基酸序列。相应地,在一些实施方式中,所述多核苷酸编码与基于SEQ ID NO:2、4或86的参考序列具有至少约80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多同一性的氨基酸序列,所述参考序列具有以下特征:相应于位置X94的残基为脂肪族或极性残基,尤其是丙氨酸或苏氨酸;相应于X199的残基为脂肪族、受限或极性残基,尤其是丙氨酸、组氨酸或天冬酰胺;并且相应于X202的残基为缬氨酸或亮氨酸,条件是所编码的酮还原酶多肽含有具有前述特征的氨基酸序列,也就是说,相应于位置X94的残基为脂肪族或极性残基;相应于X199的残基为脂肪族、受限或极性残基;并且相应于X202的残基为缬氨酸或亮氨酸。在一些实施方式中,所述多核苷酸编码具有以下氨基酸序列的酮还原酶:其中相应于X94的残基为苏氨酸;相应于X199的残基为丙氨酸、组氨酸或天冬酰胺;并且相应于X202的残基为缬氨酸或亮氨酸。在一些实施方式中,所述多核苷酸编码相应于SEQ ID NO:6、8、10、12、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48、50、52、54、56或58的氨基酸序列。

[0219] 在一些实施方式中,所述多核苷酸包含编码酮还原酶多肽的核苷酸序列,所述酮还原酶多肽具有与包含相应于SEQ ID NO:14、60或62的氨基酸的多肽具有至少约80%、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的氨基酸序列。

[0220] 在一些实施方式中,所述多核苷酸包含编码酮还原酶多肽的核苷酸序列,所述酮还原酶多肽具有与包含相应于SEQ ID NO:64、66、68、70、72、74、76、78、80或82的氨基酸序列的多肽具有至少约80%、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的氨基酸序列。

[0221] 在一些实施方式中,编码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或81。在一些实施方式中,所述多核苷酸能够在高度严格的条件下与包含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或81的多核苷酸杂交,其中高度严格性杂交的多核苷酸编码能够将具式(I)的底物立体选择性地还原或转化为具式(II)的产物,或者将具式(I)的底物立体选择性地还原或转化为具式(III)的产物的酮还原酶。

[0222] 在一些实施方式中,所述多核苷酸编码本文所述的多核苷酸,但是在核苷酸水平上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。在一些实施方式中,所述参考多核苷酸选自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 或 81。

[0223] 可采用各种方法处理编码改善的酮还原酶多肽的分离的多核苷酸以提供所述多肽的表达。取决于表达载体,在插入载体前对分离的多核苷酸的操作可为需要的或必需的。使用重组 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。在 Sambrook 等人, 2001, *Molecular Cloning: A Laboratory Manual* (分子克隆: 实验室手册), 第 3 版, Cold Spring Harbor Laboratory Press; 和 *Current Protocols in Molecular Biology* (分子生物学最新实验设计), Ausubel, F. 编辑, Greene Pub. Associates, 1998, 更新至 2006, 中提供指导。

[0224] 对于细菌宿主细胞,用于引导本公开的核酸构建体的转录的合适启动子包括从大肠埃希氏菌 lac 操纵子、大肠埃希氏菌 trp 操纵子、噬菌体  $\lambda$ 、天蓝色链霉菌 (*Streptomyces coelicolor*) 琼脂水解酶基因 (*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*)  $\alpha$ -淀粉酶基因 (*amyL*)、嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*) 麦芽淀粉酶基因 (*amyM*)、解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*)  $\alpha$ -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 *xy1A* 和 *xy1B* 基因和原核的  $\beta$ -内酰胺酶基因中获得的启动子 (Villa-Kamaroff 等人, 1978, *Proc. Natl Acad. Sci. USA* 75:3727-3731)、以及 *tac* 启动子 (DeBoer 等人, 1983, *Proc. Natl Acad. Sci. USA* 80:21-25)。另外的启动子在 "Useful proteins from recombinant bacteria (来自重组细菌的有用蛋白质)", *Scientific American*, 1980, 242:74-94; 以及在 Sambrook 等人, 如上, 中阐述。

[0225] 对于丝状真菌宿主细胞,用于引导本公开的核酸构建体的转录的合适启动子包括从米曲霉 (*Aspergillus oryzae*) TAKA 淀粉酶、米赫根毛霉 (*Rhizomucor miehei*) 天冬氨酸蛋白酶、黑曲霉 (*Aspergillus niger*) 中性  $\alpha$ -淀粉酶、黑曲霉酸稳定  $\alpha$ -淀粉酶、黑曲霉或泡盛曲霉 (*Aspergillus awamori*) 葡萄糖淀粉酶 (*glaA*)、米赫根毛霉脂肪酶、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*) 胰蛋白酶样蛋白酶 (W096/00787) 的基因中获得的启动子, 以及 NA2-tpi 启动子 (来自黑曲霉中性  $\alpha$ -淀粉酶和米曲霉磷酸丙糖异构酶基因的启动子的杂合体), 以及其突变的、截短的和杂合的启动子。

[0226] 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-磷酸甘油酸激酶的基因中。用于酵母宿主细胞的其他有用的启动子在 Romanos 等人, 1992, *Yeast* 8:423-488 中阐述。

[0227] 控制序列也可是合适的转录终止子序列,转录终止子序列是被宿主细胞识别以终止转录的序列。终止子序列可操作地连接在编码多肽的核酸序列的 3' 端。在所选宿主细胞中起作用的任何终止子可用于本发明中。

[0228] 例如,用于丝状真菌宿主细胞的示例性的转录终止子可从米曲霉 TAKA 淀粉酶、黑曲霉葡萄糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、黑曲霉  $\alpha$ -葡糖苷酶和尖孢镰刀菌胰蛋白酶样蛋白酶的基因中获得。

[0229] 用于酵母宿主细胞的示例性的终止子可从酿酒酵母烯醇化酶,酿酒酵母细胞色素

C(CYC1) 和酿酒酵母甘油醛-3-磷酸脱氢酶的基因中获得。用于酵母宿主细胞的其他有用的终止子在 Romanos 等人, 1992, 如上, 中阐述。

[0230] 控制序列也可作为合适的前导序列, 前导序列是对宿主细胞的翻译重要的 mRNA 的非翻译区。前导序列可操作地连接在编码多肽的核酸序列的 5' 端。可使用在所选宿主细胞中起作用的任何前导序列。用于丝状真菌宿主细胞的示例性的前导序列可从米曲霉 TAKA 淀粉酶和构巢曲霉磷酸丙糖异构酶的基因中获得。用于酵母宿主细胞的合适的前导序列可从酿酒酵母烯醇化酶 (ENO-1), 酿酒酵母 3-磷酸甘油酸激酶, 酿酒酵母  $\alpha$ -因子以及酿酒酵母醇脱氢酶 / 甘油醛-3-磷酸脱氢酶 (ADH2/GAP) 的基因中获得。

[0231] 控制序列也可能是聚腺苷酸化序列, 聚腺苷酸化序列是可操作地连接于核酸序列的 3' 端的序列, 当转录时, 其被宿主细胞识别为将多聚腺苷酸残基添加到转录的 mRNA 的信号。在所选宿主细胞中起作用的任何聚腺苷酸化序列可用于本发明中。用于丝状真菌宿主细胞的示例性的聚腺苷酸化序列可从米曲霉 TAKA 淀粉酶、黑曲霉葡萄糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、尖孢镰刀菌胰蛋白酶样蛋白酶和黑曲霉  $\alpha$ -葡糖苷酶的基因中取得。用于酵母宿主细胞的有用的聚腺苷酸化序列在 Guo 和 Sherman, 1995, MolCell Bio 15 : 5983-5990 中阐述。

[0232] 控制序列也可作为编码连接于多肽氨基端的氨基酸序列并且引导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' 端可固地包含以翻译阅读框与编码分泌性多肽的编码区的区段自然连接的信号肽编码区。可选择地, 编码序列的 5' 端可包含对编码序列为外源的信号肽编码区。在编码序列并非自然地包含信号肽编码区的情况下, 外源的信号肽编码区可为需要的。

[0233] 可选择地, 外源的信号肽编码区可简单地置换自然的信号肽编码区以提高多肽的分泌。然而, 将表达的多肽引导至选择的宿主细胞的分泌途径的任何信号肽编码区可用于本发明中。

[0234] 用于细菌宿主细胞的有效信号肽编码区为从芽孢杆菌 (*Bacillus*) NC1B11837 麦芽淀粉酶、嗜热脂肪芽孢杆菌  $\alpha$ -淀粉酶, 地衣芽孢杆菌枯草杆菌蛋白酶、地衣芽孢杆菌  $\beta$ -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 的基因中获得的信号肽编码区。另外的信号肽在 Simonen 和 Palva, 1993, Microbiol Rev 57 : 109-137 中描述。

[0235] 用于丝状真菌宿主细胞的有效信号肽编码区可为从米曲霉 TAKA 淀粉酶、黑曲霉中性淀粉酶、黑曲霉葡萄糖淀粉酶、米赫根毛霉天冬氨酸蛋白酶、特异质腐霉 (*Humicola insolens*) 纤维素酶和柔毛腐质霉 (*Humicolalanuginosa*) 脂肪酶的基因中获得的信号肽编码区

[0236] 用于酵母宿主细胞的有用的信号肽可从酿酒酵母  $\alpha$ -因子和酿酒酵母转化酶的基因中获得。其他有用的信号肽编码区在 Romanos 等人, 1992, 如上, 中阐述。

[0237] 控制序列也可作为编码位于多肽氨基端的氨基酸序列的前肽编码区。生成的多肽称为原酶 (proenzyme) 或多肽原 (或在一些情况下, 酶原)。多肽原通常无活性并且可通过从多肽原中催化裂解或自催化裂解前肽而转化为成熟的活性多肽。前肽编码区可从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  $\alpha$ -因子、米赫根毛霉天冬氨酸蛋白酶和嗜热毁丝霉 (*Myceliophthora thermophila*) 乳糖酶 (WO 95/33836) 的

基因中获得。

[0238] 当信号肽和前肽区域二者都存在于多肽氨基端时,前肽区域位于与多肽的氨基端相邻的位置,并且信号肽区域位于与前肽区域的氨基末端相邻的位置。

[0239] 可能还需要增加调控序列,其允许相对于宿主细胞的生长来调控多肽的表达。调控系统的例子是导致基因表达因应答化学的或物理的刺激而开和关的那些系统,所述刺激包括调控化合物的存在。在原核宿主细胞中,合适的调控序列包括 lac、tac 和 trp 操纵子系统。在酵母宿主细胞中,合适的调控系统包括,例如 ADH2 系统或 GAL1 系统。在丝状真菌中,合适的调控序列包括 TAKA  $\alpha$ -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。

[0240] 调控序列的其他的例子为允许基因扩增的那些序列。在真核系统中,调控序列包括在甲氨喋呤存在下扩增的二氢叶酸还原酶基因,以及使用重金属扩增的金属硫蛋白基因。在这些例子中,编码本发明的 KRED 多肽的核酸序列被可操作地连接在调控序列上。

[0241] 因此,在另一个实施方式中,本公开也涉及重组的表达载体,该重组的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,以及取决于他们将要引入的宿主的类型,一个或多个表达调控区域,诸如启动子和终止子,复制起点等。以上所述的不同的核苷酸和控制序列可被连接在一起以产生表达载体,该表达载体可包括一个或多个方便的限制性位点以允许编码多肽的氨基酸序列在此位点的插入或取代。可选择地,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至合适的表达载体中表达。在产生表达载体中,编码序列位于载体中以使编码序列被可操作地连接于用于表达的合适的控制序列上。

[0242] 重组表达载体可为任何载体(例如,质粒或病毒),其可方便地应用于重组 DNA 步骤中并且可带来多核苷酸序列的表达。载体的选择典型地由载体和将要引入载体的宿主细胞的相容性决定。载体可为线性的或闭合环状质粒。

[0243] 表达载体可为自主复制的载体,也就是说,作为染色体外实体存在,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。载体可包含保证自我复制的任何部件(means)。可选择地,载体可为当引入宿主细胞时并入基因组并且与其所整合的染色体一起复制的载体。此外,可应用单独载体或质粒,或一起包括要被引入宿主细胞基因组的总 DNA 的两个或更多的载体或质粒,或转座子。

[0244] 本发明的表达载体优选地包含一个或多个可选标记物,该标记物允许容易地选择转化细胞。可选标记物为其产物提供杀虫剂耐受性或病毒耐受性、重金属耐受性、营养缺陷型的原养型以及类似性质的基因。细菌可选标记物的例子为来自枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因,或带来诸如氨苄西林、卡那霉素、氯霉素(实施例 1)或四环素耐受性的抗生素耐受性的标记物。用于酵母宿主细胞的合适的标记物为 ADE2、HIS3、LEU2、LYS2、MET3、TRP1 和 URA3。

[0245] 用于丝状真菌宿主细胞的可选标记物包括但不局限于 *amdS*(乙酰胺酶)、*argB*(鸟氨酸氨甲酰基转移酶)、*bar*(草胺磷(phosphinothricin)乙酰胺基转移酶)、*hph*(潮霉素磷酸转移酶)、*niaD*(硝酸还原酶),*pyrG*(乳清酸核苷-5'-磷酸脱羧酶)、*sC*(硫酸腺苷转移酶)和 *trpC*(邻氨基苯甲酸合酶),以及其等同物。应用于曲霉属的细胞的实施方式包括构巢曲霉或米曲霉的 *amdS* 和 *pyrG* 基因以及吸水链霉菌(*Streptomyces hygroscopicus*)

的 bar 基因。

[0246] 本发明的表达载体优选地包含允许将载体整合至宿主细胞基因组中或载体在细胞中不依赖于基因组而自主复制的元件。对于整合到宿主细胞基因组,载体可依赖编码多肽的核酸序列或载体的任何其他元件以通过同源的或非同源的重组将载体整合入基因组。

[0247] 可选择地,表达载体可包含用于引导通过同源的重组整合至宿主细胞的基因组的另外的核酸序列。所述另外的核酸序列使载体在染色体的精确位置被整合至宿主细胞基因组成为可能。为提高整合至精确的位置的可能性,整合的元件可优选地含有足够数目的核酸,例如 100 至 10,000 个碱基对、优选地 400 至 10,000 个碱基对、以及最优选地 800 至 10,000 个碱基对,所述碱基对与相应的靶标序列高度同源以提高同源重组的概率。整合元件可为与宿主细胞基因组中靶标序列同源的任何序列。此外,整合元件可为非编码的或编码的核酸序列。另一方面,载体可通过非同源的重组整合至宿主细胞基因组中。

[0248] 对于自主复制,载体还可包含允许载体在相关的宿主细胞内自主复制的复制起点。细菌的复制起点的例子是 P15A ori (如图 5 中所示质粒) 或允许在大肠希埃氏菌中复制的质粒 pBR322、pUC19、pACYC177 (该质粒具有 P15A ori) 或 pACYC184 的复制起点,以及允许在芽胞杆菌中复制的 pUB110、pE194、pTA1060 或 pAM $\beta$ 1 的复制起点。用于酵母宿主细胞中的复制起点的例子为 2 微米的复制起点、ARS1、ARS4、ARS1 和 CEN3 的组合,以及 ARS4 和 CEN6 的组合。复制起点可为具有使其在宿主细胞内温度敏感地起作用的突变的复制起点 (参见例如, Ehrlich, 1978, Proc Natl AcadSci USA 75 :1433)。

[0249] 本发明的核酸序列的多于一个拷贝可插入宿主细胞中以提高基因产物的产量。核酸序列的拷贝数的增加可通过将序列的至少一个额外拷贝整合至宿主细胞基因组中或通过随核酸序列包括可扩增的可选标记物基因获得,其中包含可选标记物基因的扩增的拷贝并因此包含核酸序列的另外的拷贝的细胞,可被选择用于在合适的可选试剂的存在下培养细胞。

[0250] 用于本发明的许多表达载体为市售的。合适的市售表达载体包括来自 Sigma-Aldrich Chemicals, St. Louis MO. 的 p3xFLAG<sup>TM</sup>, 其包括 CMV 启动子和用于在哺乳动物宿主细胞中表达的 hGH 聚腺苷酸化位点以及 pBR322 复制起点和用于在大肠希埃氏菌中扩增的氨苄西林耐受性标记物。其他的合适的表达载体为 pBluescript11 SK(-) 和 pBK-CMV, 由 Stratagene, LaJolla CA 出售, 以及衍生自 pBR322 (Gibco BRL)、pUC (Gibco BRL)、pREP4、pCEP4 (Invitrogen) 或 pPoly (Lathé 等人, 1987, Gene 57 :193-201) 的质粒。

[0251] 1.4 表达酮还原酶多肽的宿主细胞

[0252] 在另一方面,本公开提供包含编码本公开的改善的酮还原酶多肽的多核苷酸的宿主细胞,所述多核苷酸被可操作地连接至一个或多个控制序列以在宿主细胞中表达酮还原酶。用于表达由本发明的表达载体编码的 KRED 多肽的宿主细胞在本领域是公知的,并且包括但不限于诸如大肠埃希氏菌、克菲尔乳杆菌、短乳杆菌、链霉菌属和鼠伤寒沙门氏菌 (*Salmonella typhimurium*) 细胞的细菌细胞;诸如酵母细胞 {例如,酿酒酵母或巴斯德毕赤氏酵母 (*Pichia pastoris*) (ATCC 登录号 201178)} 的真菌细胞;诸如果蝇 S2 以及灰翅夜蛾 (*Spodoptera*) Sf9 细胞的昆虫细胞;诸如 CHO、COS、BHK、293 和 Bowes 黑色素瘤细胞的动物细胞;以及植物细胞。上述宿主细胞的合适的培养基以及生长条件在本领域内是公知的。

[0253] 用于表达酮还原酶的多核苷酸可通过在本领域内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、氯化钙转染以及原生质体融合。将多核苷酸引入细胞的不同方法对于本领域技术人员是明显的。

[0254] 示例性的宿主细胞为大肠希埃氏菌 W3110 (*Escherichia coli* W3110)。表达载体通过将编码改善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而产生。表达载体也包含 P15a 复制起点以及氯霉素耐受性基因。大肠希埃氏菌 W3110 中包含所述多核苷酸的细胞可通过使所述细胞经受氯霉素的选择而分离。

[0255] 1.5 生成工程酮还原酶多肽的方法

[0256] 在一些实施方式中,为生成本发明的改善的 KRED 多核苷酸和多肽,自然存在的催化还原反应的酮还原酶为从克菲尔乳杆菌、短乳杆菌或小乳杆菌中取得(或衍生)。在一些实施方式中,母多核苷酸序列被密码子优化以提高酮还原酶在特定的宿主细胞中表达的密码子。作为例证,编码克菲尔乳杆菌野生型 KRED 多肽的母多核苷酸序列为从基于 Genbank 数据库中可获得的克菲尔乳杆菌 KRED 的已知多肽序列(Genbank 登录号 AAP94029 GI:33112056)制备的寡核苷酸而构建的。母多核苷酸序列(指定为 SEQ ID NO:1)为在大肠埃希氏菌中表达的密码子优化,并且将所述密码子-优化的多核苷酸克隆至表达载体,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。鉴定在大肠埃希氏菌中表达活性的酮还原酶的克隆并且测序基因以证实其身份(identity)。指定的序列(SEQ IDNO:1)为用于作为从克菲尔乳杆菌酮还原酶进化而来的工程酮还原酶的大多数实验和文库构建的起始点。

[0257] 工程酮还原酶可通过使编码自然存在的酮还原酶的多核苷酸经受如上讨论的诱变和/或定向进化方法而得到。示例性的定向进化技术为诱变和/或 DNA 改组,如在 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 和美国专利第 6,537,746 号中所描述。可用的其他定向进化程序包括但不限于交错延伸过程(staggered extension process, StEP),体外重组(Zhao 等人,1998, Nat. Biotechnol. 16:258-261),诱变 PCR(Caldwell 等人,1994, PCR Methods Appl. 3:S136-S140),以及盒式突变(Black 等人,1996, Proc Natl Acad Sci USA 93:3525-3529)。

[0258] 筛选诱变处理后获得的克隆中具有所需要的改善的酶性质的工程酮还原酶。从表达库中测量酶促活性可应用监控当 NADH 或 NADPH 转化至 NAD<sup>+</sup> 或 NADP<sup>+</sup> 时 NADH 或 NADPH 浓度的降低的速率(通过吸光度或荧光的降低)的标准生物化学技术来进行。(例如参见实施例 7。)在这个反应中,因为酮还原酶将酮底物还原为相应的羟基基团,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 浓度的降低的速率表明在固定量的裂解物(或从中产生的冻干的粉末)中的 KRED 多肽的相对的(酶促)活性。在需要的改善的酶性质为热稳定性时,酶活性可在将酶制品经受确定的温度和测量在热处理后余下的酶促活性的量之后测量到。然后包含编码酮还原酶的多核苷酸的克隆被分离,被测序以确定核苷酸序列的改变(如果有),并且用于在宿主细胞中表达酶。

[0259] 当工程多肽的序列为已知时,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。在一些实施方式中,高达 100 个碱基的片段可单独地合成,然后连接(例

如,通过酶促或化学的连接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聚核苷酸可通过化学合成制备,使用,例如,描述于 Beaucage 等人,1981,TetLett 22:1859-69 的经典的亚磷酰胺方法,或描述于 Matthes 等人,1984,EMBO J. 3:801-05 的方法,例如,如在自动化合成方法中典型地实践的。根据亚磷酰胺方法,寡核苷酸在例如,在自动化 DNA 合成仪中合成、纯化、退火、连接以及克隆至合适的载体中。另外,基本上任何核酸可从任何各种市售来源获得,例如 The Midland Certified Reagent Company, Midland, TX、The Great American Gene Company, Ramona, CA、ExpressGen Inc. Chicago, IL、Operon Technologies Inc., Alameda, CA 和许多其他公司。

[0260] 表达在宿主细胞中的工程酮还原酶可从细胞或培养基中回收,使用蛋白质纯化的公知技术的任一种或多种,包括但不限于溶菌酶处理、超声、过滤、盐析、超速离心和色谱。用于从诸如大肠埃希氏菌的细菌中裂解和高效提取蛋白质的适合溶液是以 St. Louis MO 的 Sigma-Aldrich 的商标名 CelLytic B™ 可商业地获得的。

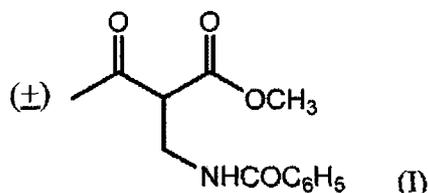
[0261] 分离酮还原酶多肽的色谱技术,包括但不限于反相色谱、高效液相色谱、离子交换色谱、凝胶电泳和亲和色谱。纯化特定的酶的条件会部分依赖于诸如净电荷、疏水性、亲水性、分子量、分子形状等因素,而且对本领域的技术人员将是显而易见的。

[0262] 在一些实施方式中,亲和技术可用于分离改善的酮还原酶。对于亲和色谱纯化,可应用特异性结合酮还原酶多肽的任何抗体。对于抗体的产生,不同宿主动物,包括但不限于兔、小鼠、大鼠等,可通过注射工程多肽免疫。通过侧链功能基团或附加在侧链功能基团上的接头的方法,多肽可附加在合适的载体诸如 BSA 上。依赖于宿主物种,不同的佐剂可用于提高免疫应答,包括但不限于弗氏(完全和不完全),诸如氢氧化铝的矿物质胶,诸如溶血卵磷脂、普朗尼克多元醇、多聚阴离子、肽、油乳剂、匙孔血蓝蛋白,二硝基酚的表面活性物质,以及诸如 BCG(bacilli CalmetteGuerin, 卡介苗)和短小棒状杆菌(*Corynebacterium parvum*)的可能有作用的人类佐剂。

[0263] 1.6 使用工程酮还原酶的方法以及用其制备的化合物

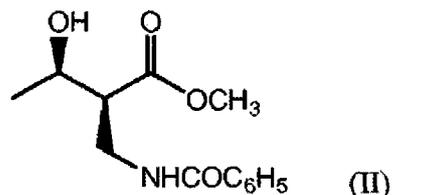
[0264] 在一些实施方式中,本文所述的酮还原酶能够催化具结构式(I)的化合物 2-苯甲酰氨基甲基-3-氧丁酸甲酯(“底物”)的酮基:

[0265]



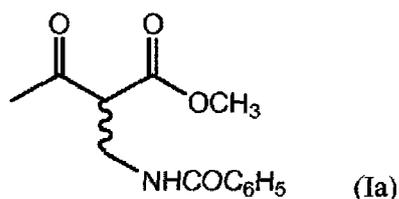
[0266] 至相应的具结构式(II)的立体异构醇产物,2S,3R-2-苯甲酰氨基甲基-3-羟基丁酸甲酯(“产物”)的还原反应:

[0267]



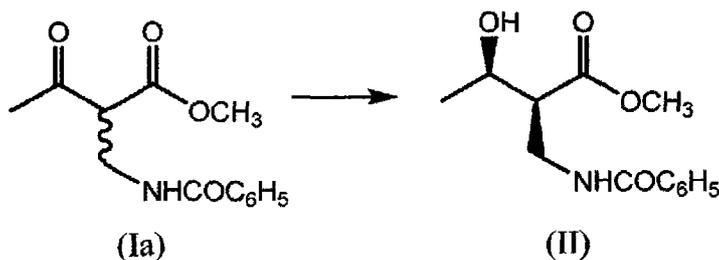
[0268] 在一些实施方式中,具式 (I) 的底物为外消旋混合物,如式 (Ia) 所示,

[0269]



[0270] 并且 2S,3R 选择性酮还原酶可用于还原或转化下面的方案 1 所示的反应中的外消旋底物以制备具式 (II) 的产物:

[0271]



[0272] 方案 1

[0273] 相应地,在一些实施方式中,本公开中酮还原酶可用于以至少约 60% 的立体异构体过量将具结构式 (I) 的底物立体选择性地还原为具结构式 (II) 的相应产物的方法,该方法包括在适合于将所述底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 或式 (Ia) 的底物与本公开的 2S,3R 选择性酮还原酶多肽接触或培养。在本方法的一些实施方式中,式 (II) 的产物以至少约 85% 的立体异构体过量生成。在本方法的一些实施方式中,2S,3R 选择性酮还原酶多肽与野生型克菲尔乳杆菌、短乳杆菌或小乳杆菌 KRED 序列 SEQ ID No :4,2 和 86 相比,具有至少以下特征:残基 202 为缬氨酸或亮氨酸。在一些实施方式中,2S,3R 选择性酮还原酶多肽与野生型克菲尔乳杆菌、短乳杆菌或小乳杆菌 KRED 序列 SEQ ID No :4,2 和 86 相比,具有至少以下特征:(1) 相应于 X94 的残基为脂肪族或极性残基;(2) 相应于 X199 的残基为脂肪族、受限或极性残基;并且(3) 相应于 X202 的残基为缬氨酸或亮氨酸。在一些实施方式中,2S,3R 选择性酮还原酶多肽与野生型克菲尔乳杆菌、短乳杆菌或小乳杆菌 KRED 序列 SEQ ID No :4,2 和 86 相比,具有至少以下特征:(1) 相应于 94 的残基为极性残基,(2) 相应于 199 的残基为受限残基,并且(3) 相应于 X202 的残基为缬氨酸或亮氨酸。

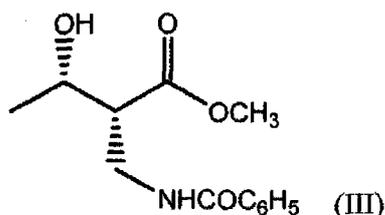
[0274] 在将底物转化为具结构式 (II) 的产物的方法的一些实施方式中,以至少约 60-89% 的立体异构体过量百分比和比具有 SEQ ID NO :48 的氨基酸序列的多肽能够达到的速率高至少约 1-15 倍的速率将底物转化为产物。可用在所述方法中的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,46,50,52,54,56,58,60 和 62 的氨基酸序列的多肽。

[0275] 在将底物转化为具结构式 (II) 的产物的方法的一些实施方式中,以至少约 90-94% 的立体异构体过量百分比将底物转化为产物。可用在所述方法中的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,40,42,50,52,56,58,60 和 62 的氨基酸序列的多肽。

[0276] 在将底物转化为具结构式 (II) 的产物的方法的一些实施方式中, 以至少约 95-99% 的立体异构体过量百分比将底物转化为产物。可用在所述方法中的示例性的多肽包括但不限于包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、42、50、52、56、58、60 和 62 的氨基酸序列的多肽。

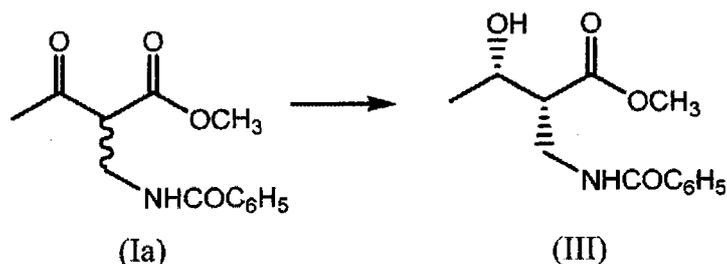
[0277] 在一些实施方式中, 本文所述的酮还原酶能够催化具结构式 (I) 的化合物的酮基至具结构式 (III) 的相应的立体异构醇产物 2R,3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯, (其也可称为“产物”) 的还原反应:

[0278]



[0279] 在一些实施方式中, 具式 (I) 的底物为外消旋的混合物, 如结构式 (Ia) 中所示, 并且 2R,3R 选择性酮还原酶可用于还原或转化如下面的方案 3 中所示的外消旋的底物以制备具式 (III) 的产物:

[0280]



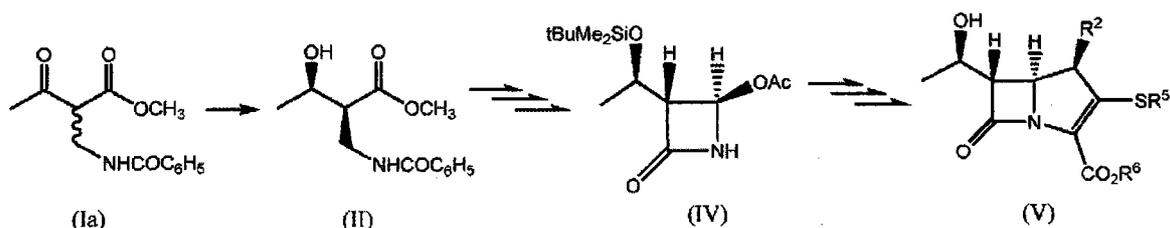
[0281] 方案 2

[0282] 在将底物转化为具结构式 (III) 的产物的方法的一些实施方式中, 以至少约 85% 的立体异构体过量百分比将底物转化为产物。可用于本方法中的示例性的多肽包括但不限于包含相应于 SEQ ID NO :68、72、74、76、78 和 82 的氨基酸序列的多肽。

[0283] 在将底物转化为具结构式 (III) 的产物的方法的一些实施方式中, 以比具有 SEQ ID NO :66 的氨基酸序列的多肽能够达到的速率高至少约 1 倍的速率将底物转化为产物。可用于本方法中的示例性的多肽包括但不限于包含相应于 SEQ ID NO :64、68、70、72、74、76、78、80 和 82 的氨基酸序列的多肽。

[0284] 在一些实施方式中, 本文所述的酮还原酶可用于合成碳杂青霉烯的方法, 该方法包括如下所示 (方案 3) 的一般过程,

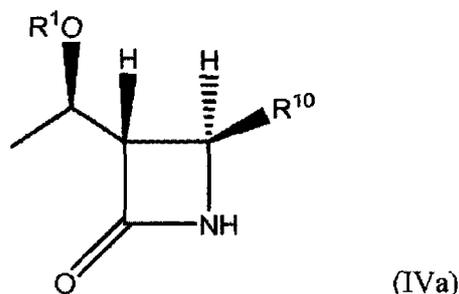
[0285]



[0286] 方案 3

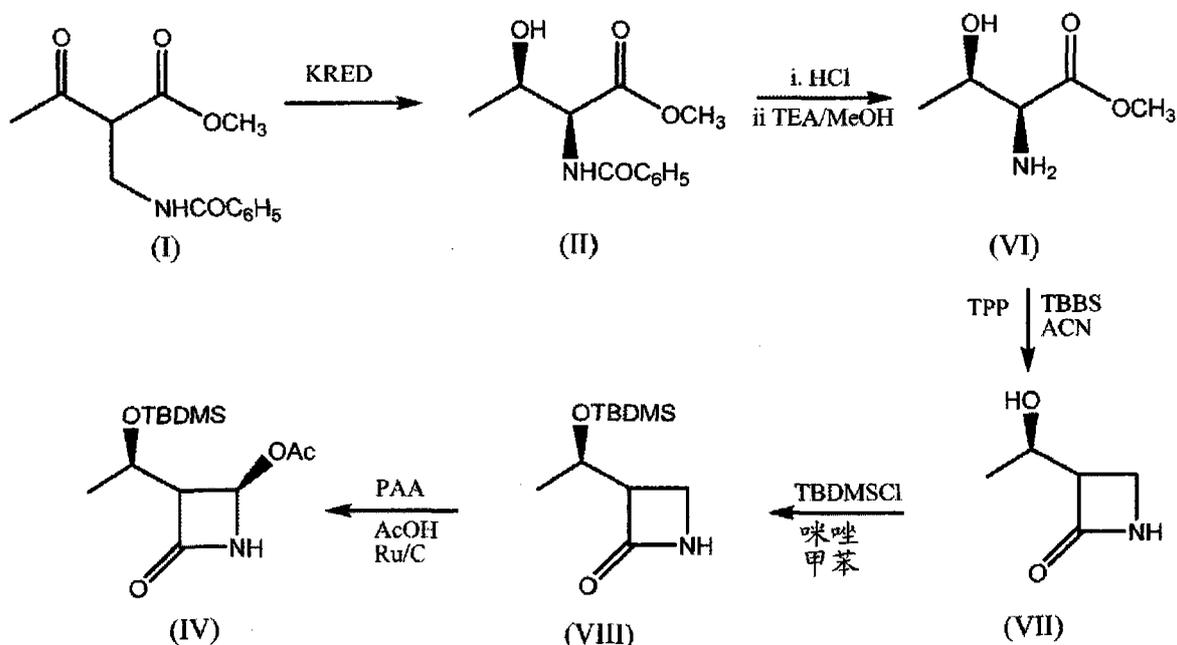
[0287] 其中 KRED 可为本发明的酮还原酶多肽中的任何一种。具通式 (V) 的碳杂青霉烯在下面进一步详细描述。对于不同基于碳杂青霉烯的治疗剂的合成, 一种重要的中间体为具结构式 (IVa) 的化合物,

[0288]



[0289] 其中  $R^1$  为 H 或羟基保护基团, 而且  $R^{10}$  为卤素 (例如, Cl) 或  $-OAc$  (Ac 为乙酸)。不同的羟基保护基团在本领域内为公知, 并且包括, 举例但不局限于, 甲硅烷基醚 (例如, 叔丁基二甲基甲硅烷基) 以及取代的二甲苯醚 (参见, Green 和 Wuts, *Protective Groups in Organic Synthesis* (有机合成中的保护基团), Wiley-Interscience, New York, 1999)。相应地, 在合成具结构式 (IVa) 的中间体的方法中, 方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。其中  $R^{10}$  为  $-OAc$  的具式 (IVa) 的中间体的合成描述于 *Tetrahedron Lett.* 23 : 2293 (1982); *Tetrahedron Lett.* 39 : 2399 (1983); *Tetrahedron Lett.* 39 : 2505 (1983); EP0290385; 的 EP0369691; 所有参考文献以引用方式并入本文。在一些实施方式中, 其中  $R^{10}$  为  $-OAc$  的具式 (IVa) 的中间体可如方案 4 中所示合成:

[0290]



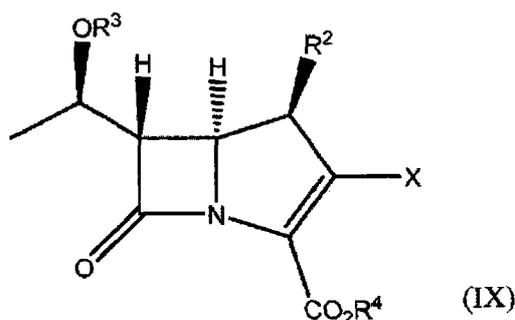
[0291] 方案 4

[0292] 在方案 4 中, 方法包括: (a) 在适合于将底物还原为具式 (II) 的产物的反应条件下将底物与本公开的 2S, 3R 选择性酮还原酶多肽接触或反应以将具式 (I) 的底物还原为具式 (II) 的产物; (b) 移除基团  $-C(O)C_6H_5$  以形成化合物 (VI); (c) 通过与三苯基膦和

N-叔丁基-2-苯并噻唑基亚磺酰胺反应将具式 (VI) 的化合物转化为具式 (VII) 的  $\beta$ -内酰胺 (参见例如, Tetrahedron Lett. 1995, 36(21) :3703); (d) 与叔丁基二甲基氯硅烷 (TBDMS-Cl) 反应以形成具式 (VIII) 的化合物; 以及 (e) 在过乙酸和钨存在下乙酰氧基化具式 (VIII) 的化合物以形成具式 (IV) 的中间体 (参见例如, Murahashi 等人, 1990, J. Am. Chem. Soc. 112(21) :7820-7822)。

[0293] 在一些实施方式中, 在基于碳杂青霉烯的治疗剂的合成中的另外一个中间体为具结构式 (IX) 的中间体:

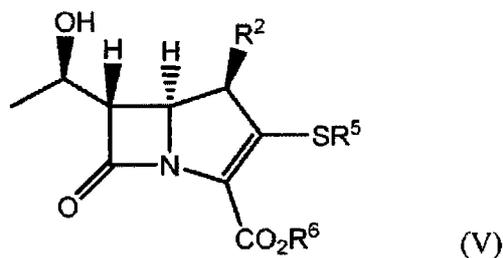
[0294]



[0295] 其中  $R^2$  为 H 或 C1-C4 烷基 (例如,  $-\text{CH}_3$ );  $R^3$  为 H 或羟基保护基团;  $R^4$  为 H、羧基保护基团、氨基、碱金属或碱土金属; 而且 X 为 OH 或离去基团。示例性的离去基团包括但不限于  $-\text{OP}(\text{O})(\text{OR}')$  或  $\text{OS}(\text{O})\text{R}''$ , 其中  $\text{R}'$  和  $\text{R}''$  可为 C1-C6 烷基、C1-C6 烷芳基、芳香基、全氟代 C1-C6 烷基。 $R^4$  的保护基团可为但不限于苄基、对硝基甲苯基以及甲氧基甲基。具结构式 (IX) 的中间体和其合成的描述提供于不同的参考文献中, 例如, 美国专利第 5, 317, 016 号; 美国专利第 4, 933, 333 号; 和 WO 0236594; 其公开以引用的形式并入本文。相应地, 在一些实施方式中, 在合成具式 (IX) 的中间体的方法中, 方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。

[0296] 在一些实施方式中, 本公开中酮还原酶可用于基于碳杂青霉烯的具结构式 (V) 的治疗剂化合物或其溶剂化物、水合物、盐和前药的合成过程:

[0297]

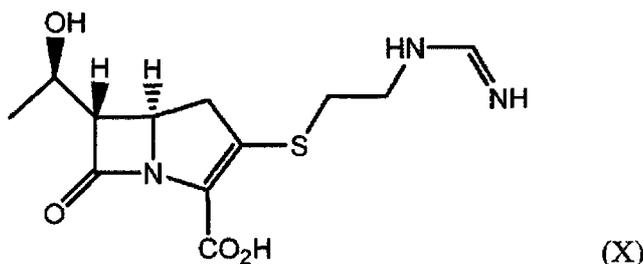


[0298] 其中  $R^2$  为 H 或  $-\text{CH}_3$ ;  $R^5$  可为不同的取代基, 包括但不限于, 取代的或未取代的烷基、取代的或未取代的芳香基、取代的或未取代的杂烷基、取代的或未取代的杂环烷基、以及取代的或未取代的杂芳香基烷基; 而且  $R^6$  为 H 或诸如可水解的酯基团的前体基团。如本文所用, 前体基团指当用于屏蔽活性药物的功能基团以形成前体部分时将药物转化为前药的保护基团的类型。前体基团典型地通过在特定的使用条件下可切断的键而附加在药物的功能基团上。因此, 前体基团为在特定的使用条件下切断以释放功能基团的前体部分的一部分。羧基基团可被屏蔽为酯 (包括甲硅烷基酯和硫酯)、酰胺或酰肼前体部分, 其可在体

内水解以提供羧基基团。合适的前体基团以及其相应的前体部分的其他特定的例子对于本领域内技术人员是明显的。相应地,在合成具结构式 (V) 的化合物的方法中,方法中的步骤可包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。基于碳杂青霉烯的不同治疗剂化合物的合成描述于包括但不限于美国专利第 4,925,836 ;5,317,016 号 ;WO 02/036594 ;WO2004/067532 ;和 WO 2007/104221 ;其公开以引用方式并入本文。

[0299] 在一些实施方式中,本公开中酮还原酶可用于合成具有下列结构 (X) 的亚胺培南或其溶剂化物、水合物、前药和盐的合成 :

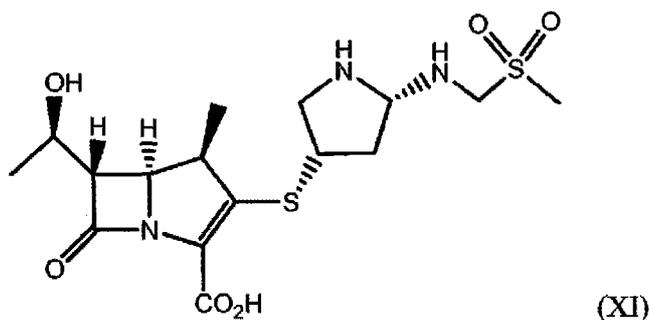
[0300]



[0301] 亚胺培南具有广谱的抗需氧和厌氧革兰氏阳性或阴性细菌的活性。亚胺培南对于铜绿假单胞菌 (*Pseudomonas aeruginosa*) 和肠球菌种 (*Enterococcus* species) 尤其有效。相应地,在合成具结构式 (X) 的亚胺培南的方法中,方法中的步骤可包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。从具结构式 (VI) 的中间体制备亚胺培南的程序描述于 WO 0236594,在此以引用的方式并入本文。

[0302] 在一些实施方式中,本公开的酮还原酶可用于具有结构式 (XI) 的多利培南或其溶剂化物、水合物或盐的合成 :

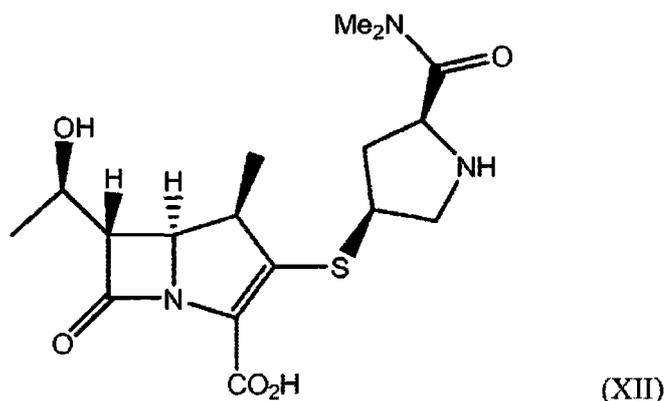
[0303]



[0304] 多利培南具有与亚胺培南或厄他培南相似的抗革兰氏阳性球菌的谱系和能力,并且具有与美洛培南相似的抗革兰氏阴性菌的活性。相应地,在合成具结构式 (XI) 的多利培南的方法中,方法中的步骤可包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。从具式 (VI) 的中间体制备多利培南的过程描述于美国专利第 5,317,016 号,在此以引用的方式并入本文。

[0305] 在一些实施方式中,本公开的酮还原酶可用于具有结构式 (XII) 的美洛培南或其溶剂化物、水合物、前药和盐的合成 :

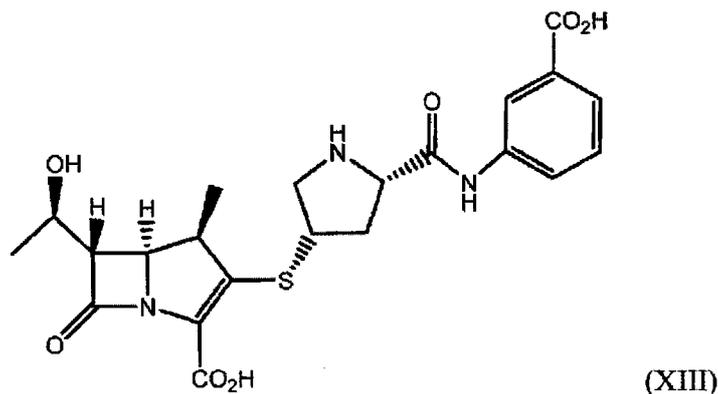
[0306]



[0307] 美洛培南被确认用于诸如阑尾炎和腹膜炎的腹内感染的治疗；细菌性脑膜炎的治疗；以及由可疑生物体导致的皮肤和皮肤结构感染的治疗。相应地，在合成具结构式 (XII) 的美洛培南的方法中，方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。制备美洛培南的过程描述于 WO 2007/104221，在此以引用的方式并入本文。

[0308] 在一些实施方式中，本公开的酮还原酶可用于具有结构式 (XIII) 的厄他培南或其溶剂化物、水合物、前药和盐的合成：

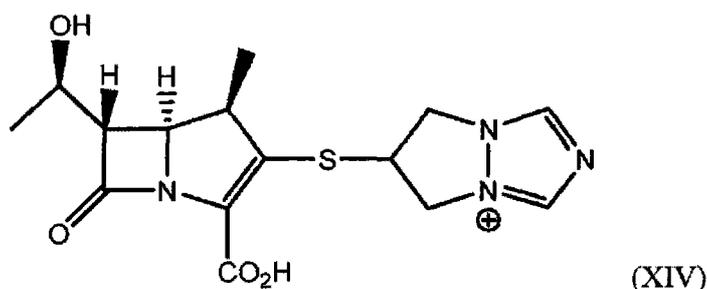
[0309]



[0310] 厄他培南可有效地抗革兰氏阴性细菌，但是对抗 MRSA、氨苄西林 - 耐受性肠球菌，铜绿假单胞菌或不动杆菌种 (*Acinetobacter species*) 无活性。厄他培南还对抗厌氧细菌有临床有效活性。厄他培南主要用于抗超广谱的产生  $\beta$ -内酰胺酶 (ESBL) 和产生高水平的 AmpC 的革兰氏阴性细菌。相应地，在合成具式 (XIII) 的厄他培南的方法中，方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。制备厄他培南的过程描述于美国专利第 5,478,820 号和美国专利第 7,342,005 号，在此以引用的方式并入本文。

[0311] 在一些实施方式中，本公开的酮还原酶可用于具有结构式 (XIV) 的比阿培南或其溶剂化物、水合物、前药或盐的合成：

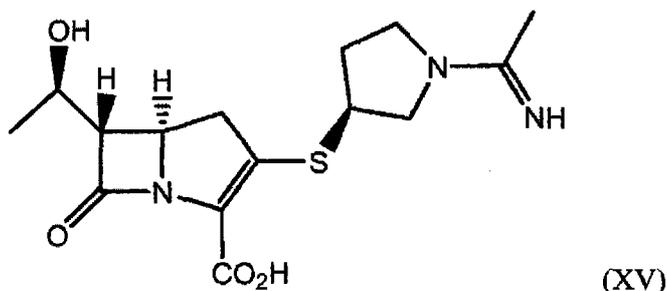
[0312]



[0313] 比阿培南为对于广泛范围的革兰氏阳性和革兰氏阴性细菌具有抗细菌活性的非肠道的碳杂青霉烯，并且能够抗人类肾脏脱氢肽酶-I (DHP-I)。比阿培南也显示抗厌氧细菌的体外活性。相应地，在合成具式 (XIV) 的比阿培南的方法中，方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。从中间体 (IVa) 制备比阿培南的过程描述于美国专利第 4, 925, 836 号，在此以引用的方式并入本文。

[0314] 在一些实施方式中，本公开的酮还原酶可用于具有结构式 (XII) 的帕尼培南或其溶剂化物、水合物和盐的合成：

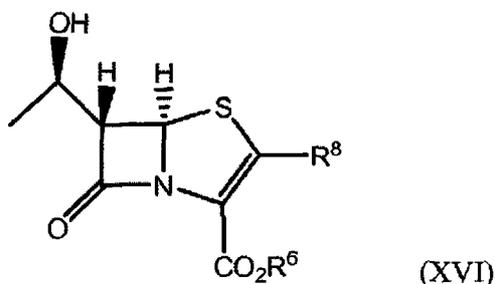
[0315]



[0316] 相应地，在合成具式 (XV) 的帕尼培南的方法中，方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。从具式 (II) 的中间体得到的帕尼培南描述于 Miyadera 等人, 1983, J Antibiotic (Tokyo) 36 (8) :1034-1039 ;以及 Hirai 等人, 1999, J. Synth Org. Chem. 57 (5) : 382-386, 在此以引用的方式并入本文。

[0317] 在一些实施方式中，本公开的酮还原酶可用于具式 (XVI) 的硫培南化合物的合成方法：

[0318]

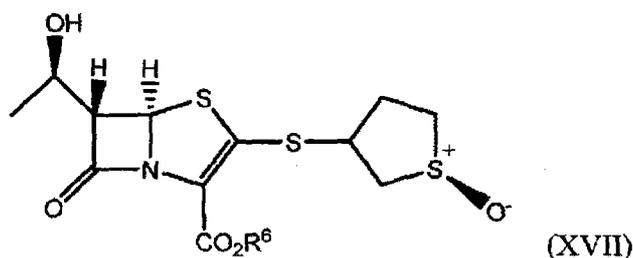


[0319] 其中  $R^6$  为 H 或前体基团，而且  $R^8$  可为取代的或未取代的烃基；取代的或未取代的芳香基，取代的或未取代的杂烃基，取代的或未取代的杂环烃基，以及取代的或未取代的杂芳香基烃基，或  $-SR^9$ ，其中  $R^9$  可为如  $R^8$  所述的取代基。他用于生物活性硫培南治疗剂化合物中的  $R^8$  的其取代基在本领域是公知的，其中一些进一步描述如下。硫培南的描述可在不

限于以下的文献中发现：美国专利第 3,951,954 号；美国专利第 4,234,579 号；美国专利第 4,287,181 号；美国专利第 4,452,796 号；美国专利第 4,343,693 号；美国专利第 4,348,264 号；美国专利第 4,416,891 号；美国专利第 4,457,924 号；美国专利第 5,013,729 号；美国专利第 5,506,225；Volkman 等人,1992, J. Org. Chem. 57 :4352-4361；以及 Volkman 和 O' Neill,1991, Strategies and Tactics in Organic Synthesis (有机合成中的手段和策略), 第 13 卷, 485-534 页；其公开以引用的方式并入本文。相应地, 在合成具结构式 (XVI) 的硫培南的方法中, 方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将式 (I) 的底物与本公开的酮还原酶接触或反应。

[0320] 在一些实施方式中, 本公开的酮还原酶可用于具式 (XVII) 的硫培南或其溶剂化物、水合物、前药和盐的合成方法：

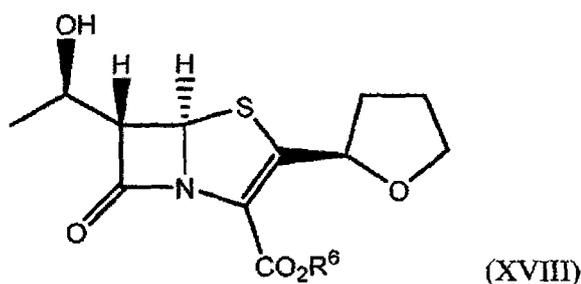
[0321]



[0322] 其中  $R^6$  为 H 或前体基团。相应地, 在合成具式 (XVII) 的化合物的方法中, 方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。具式 (XVII) 的硫培南的合成过程描述于美国专利第 5,013,729 号。 $R^6$  的不同的前体基团描述于美国申请公布 2008/0009474, 美国申请公布 2008/0125408；以及 Wujcik 等人, 2008, Rapid Commun. Mass Spectrom. 22 :3195-3206；其公开内容在此以引用的方式并入本文。

[0323] 在一些实施方式中, 本公开的酮还原酶可用于具有结构式 (XVIII) 的硫培南或其溶剂化物、水合物、前药和盐的合成方法：

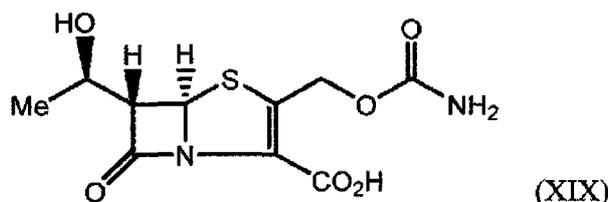
[0324]



[0325] 其中  $R^6$  为 H 或前体基团。相应地, 在合成具式 (XVIII) 的帕尼培南的方法中, 方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。具式 (XVIII) 的硫培南的合成过程描述于美国专利第 5,506,225 号和 WO 2007/039885。

[0326] 可用在此公开的酮还原酶合成的另一硫培南化合物为具式 (XIX) 的硫培南：

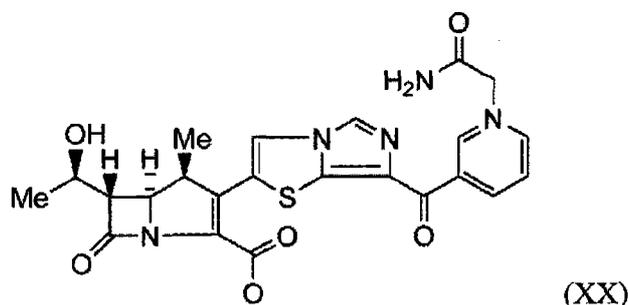
[0327]



[0328] 相应地,在合成具式 (XIX) 的化合物或其溶剂化物、水合物、前药和盐的方法中,方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。具结构式 (XIX) 的利替培南描述于美国专利第 4,482,565 号。

[0329] 在一些实施方式中,本公开的酮还原酶可用于具有结构式 (XX) 的硫培南或其溶剂化物、水合物、前药和盐的合成方法:

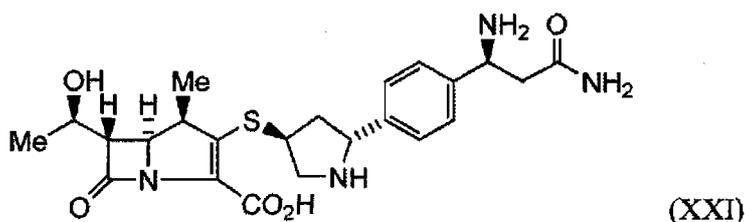
[0330]



[0331] 相应地,在合成具式 (XX) 的化合物的方法中,方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。具结构式 (XX) 的化合物描述于 EP1336612;在此以引用的方式并入本文。

[0332] 在一些实施方式中,本公开的酮还原酶可用于具有结构式 (XXI) 的化合物或其溶剂化物、水合物、前药和盐的合成方法:

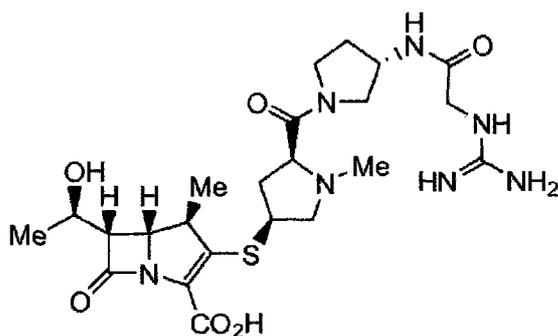
[0333]



[0334] 相应地,在合成具式 (XXI) 的化合物的方法中,方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。具结构式 (XXI) 的化合物描述于 JP2002114788;在此以引用的方式并入本文。

[0335] 在一些实施方式中,本公开的酮还原酶可用于具有结构式 (XXII) 的硫培南或其溶剂化物、水合物、前药和盐的合成方法:

[0336]

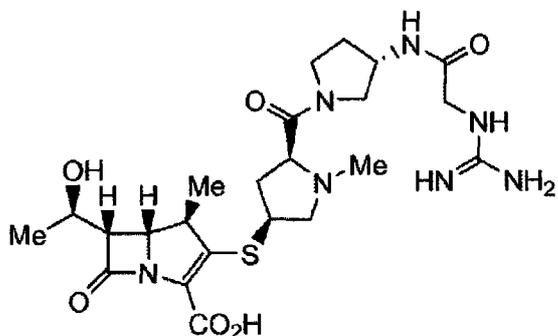


(XXII)

[0337] 相应地,在合成具式 (XXII) 的化合物的方法中,方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。具结构式 (XXII) 的化合物描述于美国专利第 6,924,279 号和美国专利第 7,034,150 号,其公开内容在此以引用的方式并入本文。

[0338] 在一些实施方式中,本公开的酮还原酶可用于具有结构式 (XXIII) 的硫培南或其溶剂化物、水合物、前药和盐的合成方法:

[0339]



(XXIII)

[0340] 相应地,在合成具式 (XXIII) 的化合物的方法中,方法中的步骤包括在适合于将底物还原或转化为具式 (II) 的产物的反应条件下将具式 (I) 的底物与本公开的酮还原酶接触或反应。具结构式 (XXIII) 的化合物描述于 W02008/047909 和美国专利第 5,659,043 号,其公开内容在此以引用的方式并入本文。

[0341] 在一些实施方式中,本公开的酮还原酶可作为与酮还原酶作用的底物和 / 或与由酮还原酶反应产生的产物的组合物存在,例如反应组合物。相应地,在一些实施方式中,组合物可含有本公开的 2S,3R 选择性酮还原酶和具结构式 (I) 的底物。在一些实施方式中,组合物可含有本公开的 2S,3R 选择性酮还原酶和具结构式 (II) 的产物。在一些实施方式中,组合物可含有本公开的 2S,3R 选择性酮还原酶、具式 (I) 的底物以及具式 (II) 的产物。

[0342] 在一些实施方式中,组合物可含有本公开的 2R,3R 选择性酮还原酶和具结构式 (I) 的底物。在一些实施方式中,组合物可含有本发明的 2R,3R 选择性酮还原酶和具结构式 (III) 的产物。在一些实施方式中,组合物可含有本公开的 2R,3R 选择性酮还原酶、式 (I) 的底物以及具式 (III) 的产物。

[0343] 因为酮还原酶反应可在存在辅因子 (NADH 或 NADPH) 再生系统下进行,反应条件可进一步包括辅因子再生系统的要素,其在以下进一步具体阐述。相应地,在一些实施方式中,前述的酮还原酶组合物还可包括含有葡萄糖脱氢酶和葡萄糖;甲酸脱氢酶和甲酸;或异丙醇和仲醇脱氢酶的辅因子再生系统。在一些实施方式中,所述仲醇脱氢酶为工程酮还原酶。可用于辅因子再循环的其他酶和底物在本领域内技术人员中是公知的。

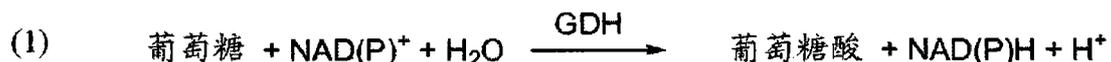
[0344] 如本领域技术人员所知的, 酮还原酶催化的还原反应典型地需要辅因子。由本文所述的工程酮还原酶催化的还原反应还典型地需要辅因子, 尽管工程酮还原酶的许多实施方式需要比由野生型酮还原酶催化的反应少得多的辅因子。如本文所用, 术语“辅因子”指与酮还原酶联合操作的非蛋白质化合物。适用于本文所述的工程酮还原酶的辅因子包括但不局限于  $\text{NADP}^+$  (烟酰胺腺嘌呤二核苷酸磷酸)、 $\text{NADPH}$  ( $\text{NADP}^+$  的还原形式)、 $\text{NAD}^+$  (烟酰胺腺嘌呤二核苷酸) 和  $\text{NADH}$  ( $\text{NAD}^+$  的还原形式)。通常, 还原形式的辅因子被加入反应混合物中。还原的  $\text{NAD(P)H}$  形式可任选地使用辅因子再生系统从氧化的  $\text{NAD(P)}^+$  形式中再生。

[0345] 术语“辅因子再生系统”指参与将氧化形式辅因子还原 (例如,  $\text{NADP}^+$  至  $\text{NADPH}$ ) 的反应的一组反应物。由酮还原酶催化酮底物还原而氧化的辅因子可被辅因子再生系统再生为还原形式。辅因子再生系统包含为还原氢等同物的来源和能够还原辅因子的氧化形式的化学计量的还原物。辅因子再生系统也可包含催化还原剂对氧化形式的辅因子的还原的催化剂, 例如酶催化剂。从  $\text{NAD}^+$  或  $\text{NADP}^+$  中分别再生  $\text{NADH}$  或  $\text{NADPH}$  的辅因子再生系统在本领域内是公知的, 并且可用于本文所述的方法。

[0346] 可使用的合适的示例性的辅因子再生系统包括但不局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, 仲醇 (例如, 异丙醇) 和仲醇脱氢酶, 亚磷酸盐和亚磷酸盐脱氢酶, 分子氢和氢化酶, 以及类似系统。这些系统可与作为辅因子的  $\text{NADP}^+/\text{NADPH}$  或  $\text{NAD}^+/\text{NADH}$  联合使用。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, 二者以引用形式并入本文。包含金属催化剂和还原试剂 (例如, 分子氢或甲酸) 的化学辅因子再生系统也适用。参见例如, PCT 公布 WO 2000/053731, 在此以引用的方式并入本文。

[0347] 术语“葡萄糖脱氢酶”和“GDH”在此可交换使用, 指催化 D-葡萄糖和  $\text{NAD}^+$  或  $\text{NADP}^+$  分别至葡萄糖酸和  $\text{NADH}$  或  $\text{NADPH}$  的  $\text{NAD}^+$  或  $\text{NADP}^+$ -依赖酶。如下反应式 (1), 描述葡萄糖脱氢酶催化的  $\text{NAD}^+$  或  $\text{NADP}^+$  被葡萄糖还原。

[0348]



[0349] 适用于本文所述的方法的实践的葡萄糖脱氢酶包括自然存在的葡萄糖脱氢酶以及非自然存在的葡萄糖脱氢酶。自然存在的葡萄糖脱氢酶编码基因在文献中有报道。例如, 枯草芽孢杆菌 61297 GDH 基因在大肠埃希氏菌中表达, 并且据报道显示与在天然宿主中产生的酶相同的物理化学的性质 (Vasanth 等人, 1983, Proc. Natl. Acad. Sci. USA 80 :785)。相应于 Genbank 登录号 M12276 的枯草芽孢杆菌 GDH 基因的基因序列, 报道于 Lampel 等人, 1986, J. Bacteriol. 166 :238-243, 并且作为 Genbank 登录号 D50453 以校正的形式报道于 Yamane 等人, 1996, Microbiology 142 :3047-3056。自然存在的 GDH 基因也包括编码来自以下的那些基因: 蜡状芽孢杆菌 (*B. cereus*) ATCC 14579 (Nature, 2003, 423 :87-91 ; Genbank 登录号 AE017013) 以及来自巨大芽孢杆菌 (*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)。来自芽孢杆菌种的葡萄糖脱氢酶在 PCT 公布 WO2005/018579 中提供为 SEQ ID NO :10 和 12 (分别由 PCT 公布中相应于 SEQ ID NO :9 和 11 的多核苷酸序列编码), 其公开内容在此以引用的方式并入本文。

[0350] 非自然存在的葡萄糖脱氢酶可通过已知的方法产生, 诸如, 例如诱变、定向进化

和类似方法。具有适当活性的 GDH 酶,无论是自然存在的还是非自然存在的,可使用描述于 PCT 公布 WO 2005/018579 的实施例 4 中的测定法容易地确定,其公开内容在此以引用的方式并入本文。示例性的非自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。编码他们的多核苷酸序列在 PCT 公布 WO 2005/018579 中分别提供为 SEQ ID NO :61、63、65、67、121、123 和 125。所有这些序列在此以引用的形式并入本文。适用于本文所述的酮还原酶催化还原反应的另外的非自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,其公开内容在此以引用的方式并入本文。

[0351] 应用于本文所述的酮还原酶催化的还原反应的葡萄糖脱氢酶可在描述于 PCT 公布 WO 2005/018579 的实施例 4 中的测定法中显示至少约  $10 \mu\text{mol}/\text{min}/\text{mg}$  和有时至少约  $10^2 \mu\text{mol}/\text{min}/\text{mg}$  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。

[0352] 本文所述的酮还原酶催化的还原反应通常发生于溶剂中。合适的溶剂包括水、有机溶剂(例如,乙酸乙酯、乙酸丁酯、1-辛醇、庚烷、辛烷、甲基叔丁基醚(MTBE)、甲苯及类似有机溶剂)、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-甲基咪唑六氟磷酸及类似离子液体)。在一些实施方式中,使用包括水和水性共溶剂系统的水性溶剂。

[0353] 示例性的水性共溶剂系统包含水和一种或多种有机溶剂。通常,选择水性共溶剂系统的有机溶剂组分以便其不完全使酮还原酶失活。通过用在候选溶剂系统中的感兴趣的确定的底物并应用诸如本文所述的酶活性测定法而测量工程酮还原酶的活性,合适的共溶剂系统可被容易地确定。

[0354] 水性共溶剂系统中的有机溶剂可与水性组分混溶,提供单一的液相,或可与水性组部分混溶或不混溶,提供双液相。通常,当应用水性共溶剂系统时,其选用为水分散于有机溶剂中,或者相反的两相。通常,当使用水性共溶剂系统时,需要选择可易于与水相分离的有机溶剂。通常,共溶剂系统中水和有机溶剂的比率典型地在从约 90 : 10 到约 10 : 90(v/v) 有机溶剂比水的范围内,以及在 80 : 20 和 20 : 80(v/v) 有机溶剂比水之间。共溶剂系统可在加入反应混合物之前提前形成,或可在反应容器中原位形成。

[0355] 水性溶液(水或水性共溶剂系统)可为 pH- 缓冲的或非缓冲的。通常,还原可在 pH 约 10 或更低、通常在从约 5 至约 10 的范围内发生。在一些实施方式中,还原在 pH 约 9 或更低、通常在从约 5 至约 9 的范围内发生。在一些实施方式中,还原可在 pH 约 8 或更低、经常在从约 5 至约 8 的范围内发生、以及通常在从约 6 至约 8 的范围内发生。还原可在 pH 约 7.8 或更低,或 7.5 或更低下发生。可选择地,还原可在中性 pH,也就是约 7 下发生。

[0356] 在还原反应的过程中,反应混合物的 pH 可改变。反应混合物的 pH 可通过在反应过程中加入酸或碱保持在需要的 pH 或在需要的 pH 范围内。可选择地,pH 可使用包含缓冲液的水性溶液控制。保持需要的 pH 范围的合适的缓冲液在本领域内是公知的,包括,例如,磷酸缓冲液、三乙醇胺缓冲液和类似缓冲液。也可使用缓冲液和酸或碱的组合。

[0357] 当应用葡萄糖 / 葡萄糖脱氢酶辅因子再生系统时,如果产生的葡萄糖酸不以别的方式中和,如在反应式 (3) 中所示的葡萄糖酸 ( $\text{pK}_a = 3.6$ ) 的共同产生引起反应混合物的 pH 下降。反应混合物的 pH 可通过标准的缓冲技术保持在需要的水平,其中缓冲液中和葡萄糖酸直至所提供的缓冲容量,或通过转化过程的同时加入碱。也可用加入缓冲液和加入

碱的组合。保持所需 pH 范围的适当的缓冲液如上所述。中和葡萄糖酸的合适的碱为有机碱例如胺、醇盐和类似有机碱,以及无机碱,例如氢氧化物盐(例如, NaOH)、碳酸盐(例如, NaHCO<sub>3</sub>)、碳酸氢盐(例如, K<sub>2</sub>CO<sub>3</sub>)、碱性磷酸盐(例如, K<sub>2</sub>HPO<sub>4</sub>, Na<sub>3</sub>PO<sub>4</sub>)及类似无机碱。在转化过程中同时加入碱可在监测反应混合物 pH 的同时手工完成,或更方便地通过使用自动滴定器作为 pH 固定器(pHstat)。部分缓冲容量和碱加入的组合可用于过程控制。

[0358] 当碱加入用于中和在酮还原酶催化的还原反应中释放的葡萄糖酸时,转化的发展可通过保持 pH 而加入的碱的量来监控。典型地,在还原过程中加入非缓冲的或部分缓冲的反应混合物的碱以水性溶液加入。

[0359] 在一些实施方式中,辅因子再生系统可包含甲酸脱氢酶。术语“甲酸脱氢酶”和“FDH”在此可交换使用,指分别催化甲酸和 NAD<sup>+</sup> 或 NADP<sup>+</sup> 至二氧化碳和 NADH 或 NADPH 的转化的 NAD<sup>+</sup> 或 NADP<sup>+</sup>- 依赖酶。适用于本文所述的酮还原酶催化的还原反应的辅因子再生系统包括自然存在的甲酸脱氢酶以及非自然存在的甲酸脱氢酶。甲酸脱氢酶包括那些 PCT 公布 W02005/018579 中相应于 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,分别由相应于 PCT 公布 W0 2005/018579 中 SEQ ID NO :69 和 71 的多核苷酸序列编码,其公开内容在此以引用的方式并入本文。应用于本文所述的方法的甲酸脱氢酶,无论是自然存在的或非自然存在的,可在描述于 PCT 公布 W02005/018579 的实施例 4 中的测定法中显示至少约 1 μ mol/min/mg,有时至少约 10 μ mol/min/mg,或至少约 10<sup>2</sup> μ mol/min/mg,多达约 10<sup>3</sup> μ mol/min/mg 或更高的活性。

[0360] 如本文所用,术语“甲酸”(formate)指甲酸阴离子(HCO<sub>2</sub><sup>-</sup>)、甲酸(HCO<sub>2</sub>H)及其混合物。甲酸可以盐的形式提供,典型地为碱金属盐或铵盐(例如, HCO<sub>2</sub>Na、KHCO<sub>2</sub>NH<sub>4</sub> 及类似物),以甲酸的形式提供,典型地为水性甲酸,或其混合物。甲酸为中等酸。在其 pKa (在水中 pKa = 3.7) 的几个 pH 单位之间的水性溶液中,甲酸以 HCO<sub>2</sub><sup>-</sup> 和 HCO<sub>2</sub>H 在平衡浓度下存在。在 pH 值高于约 pH4 时,甲酸以 HCO<sub>2</sub><sup>-</sup> 优势存在。当甲酸提供为甲酸时,反应混合物典型地被缓冲或通过加入碱以提供需要的 pH,典型地为约 pH5 或更高减少酸化。中和甲酸的合适的碱包括但不局限于诸如胺、醇盐和类似有机碱的有机碱,以及无机碱,例如氢氧化物盐(例如, NaOH)、碳酸盐(例如, NaHCO<sub>3</sub>)、碳酸氢盐(例如, K<sub>2</sub>CO<sub>3</sub>)、碱性磷酸盐(例如, K<sub>2</sub>HPO<sub>4</sub>, Na<sub>3</sub>PO<sub>4</sub>), 以及类似物。

[0361] 对于大于约 pH5 的 pH 值,其中甲酸以 HCO<sub>2</sub><sup>-</sup> 优势存在,如下反应式 (2),描述甲酸脱氢酶 - 催化的 NAD<sup>+</sup> 或 NADP<sup>+</sup> 被甲酸的还原。

[0362]



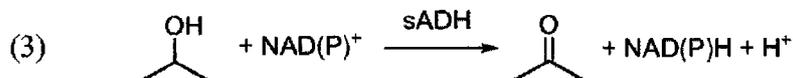
[0363] 当甲酸和甲酸脱氢酶用做辅因子再生系统时,反应混合物的 pH 可通过标准的缓冲技术保持在需要的水平,其中缓冲液释放质子直至所提供的缓冲容量,或者通过在转化的过程的同时加入酸。在反应过程中加入以保持 pH 的合适的酸包括诸如羧酸、磺酸、磷酸和类似物的有机酸,诸如氢卤酸(例如氢氯酸)、硫酸、磷酸及类似酸的矿物质酸,酸式盐,例如二氢磷酸盐(例如, KH<sub>2</sub>PO<sub>4</sub>)、硫酸氢盐(例如, NaHSO<sub>4</sub>) 和类似酸式盐。一些实施方式应用甲酸,其中维持溶液的甲酸浓度和 pH。

[0364] 当加入酸应用于在使用甲酸 / 甲酸脱氢酶辅因子再生系统的还原反应中以保持 pH 时,转化的进展可通过保持 pH 而加入的酸的量来监测。典型地,在转化过程中加入未缓

冲的或部分缓冲的反应混合物的酸以水性溶液加入。

[0365] 术语“仲醇脱氢酶”和“sADH”在此可交换使用,指分别催化仲醇和  $\text{NAD}^+$  或  $\text{NADP}^+$  至酮和  $\text{NADH}$  或  $\text{NADPH}$  的还原的  $\text{NAD}^+$  或  $\text{NADP}^+$ - 依赖酶。如下反应式 (3),描述  $\text{NAD}^+$  或  $\text{NADP}^+$  被仲醇还原,图示为异丙醇。

[0366]



[0367] 适用于本文所述的酮还原酶催化的还原反应的辅因子再生系统的仲醇脱氢酶包括自然存在的仲醇脱氢酶以及非自然存在的仲醇脱氢酶。自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的醇脱氢酶,而非自然存在的仲醇脱氢酶包括衍生自其中的工程醇脱氢酶。应用于本文所述的方法的仲醇脱氢酶,无论是自然存在的或非自然存在的,可显示至少约  $1 \mu\text{mol}/\text{min}/\text{mg}$ , 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。

[0368] 合适的仲醇包括低级仲烷醇和芳香基烷基甲醇。低级仲醇的例子包括异丙醇、2-丁醇、3-甲基-2-丁醇、3-戊醇、3,3-二甲基-2-丁醇和类似低级仲醇。在一个实施方式中,仲醇为异丙醇。合适的芳香基烷基甲醇包括未取代的和取代的1-芳香基乙醇。

[0369] 当仲醇和仲醇脱氢酶应用为辅因子再生系统时,产生的  $\text{NAD}^+$  或  $\text{NADP}^+$  被采用仲醇脱氢酶的仲醇至酮的偶联氧化所还原。一些工程酮还原酶也具有将仲醇还原剂脱氢的活性。在一些使用仲醇作为还原剂的实施方式中,工程酮还原酶和仲醇脱氢酶为相同的酶。

[0370] 在应用辅因子再生系统实现本文所述的酮还原酶催化的还原反应的实施方式中,可最初提供氧化的或还原形式的辅因子。如上所述,辅因子再生系统将氧化的辅因子转化为其还原形式,然后其还原形式用于酮还原酶底物的还原

[0371] 在一些实施方式中,不使用辅因子再生系统。对于不使用辅因子再生系统进行的还原反应,辅因子以还原形式加入反应混合物。

[0372] 在一些实施方式中,当所述过程使用宿主生物体的完整细胞进行时,完整细胞可自然地提供辅因子。可选择地或组合地,细胞可自然地或重组地提供葡萄糖脱氢酶。

[0373] 在进行本文所述的立体选择性还原反应时,工程酮还原酶,以及包含任选的辅因子再生系统的任何酶,可以纯化的酶、用编码所述酶的基因转化的完整细胞、和/或细胞提取物和/或所述细胞的裂解物的形式加入反应混合物中。编码工程酮还原酶以及任选地辅因子再生酶的基因可分别转化入宿主细胞或一起转化入相同的宿主细胞。例如,在一些实施方式中,可转化编码工程酮还原酶的基因进入一组宿主细胞,并且可转化编码辅因子再生酶的基因进入另外一组宿主细胞。两组转化的宿主细胞可以完整细胞的形式,或其衍生的裂解物或提取物的形式一起用于反应混合物。在其他实施方式中,宿主细胞可被用编码工程酮还原酶和辅因子再生酶的基因转化。

[0374] 转化了编码工程酮还原酶和/或任选的辅因子再生酶的完整细胞,或其细胞提取物和/或裂解物可以各种不同的形式应用,包括固体(例如,冻干的、喷雾干燥的以及类似固体)或半固体(例如,粗糊状物)。

[0375] 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)

部分纯化,之后在低压冻干前进行除盐程序(例如,超滤、透析及类似程序)。任何细胞制品可通过使用已知的交联剂如,例如戊二醛交联或固定到固相(例如,Eupergit C及类似固相)而被稳定。

[0376] 固体反应物(例如,酶、盐等)可以各种不同的形式提供给反应,包括粉末(例如,低压冻干的、喷雾干燥的以及类似的粉末)、溶液、乳剂、悬浮液及类似形式。反应物可使用为本领域普通技术人员共知的方法和仪器容易地冻干或喷雾干燥。例如,蛋白质溶液可以小量冷冻于 $-80^{\circ}\text{C}$ ,然后加入至预冷却的冻干舱内,之后应用真空。在从样品中移除水之后,在释放真空和回收冻干的样品之前,典型地将温度升高至 $4^{\circ}\text{C}$ 持续2小时。

[0377] 用于还原反应的反应物的量会根据所需产物的量以及伴随着所应用的酮还原酶底物的量而改变。以下指南可用于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。通常,酮底物可采用约20至300克/升的浓度,使用从约50mg至约5g酮还原酶以及约10mg至约150mg辅因子。本领域的普通技术人员会容易地理解如何改变这些量以使其适用于所需要的产率水平和生产规模。任选的辅因子再生系统的合适的量可基于应用辅因子和/或酮还原酶的量通过常规实验而容易地确定。通常,使用超过酮还原酶底物等摩尔水平的反应物(例如,葡萄糖、甲酸、异丙醇)以达到本质上完全或接近完全的对酮还原酶底物的转化。

[0378] 加入反应物的顺序不是严格的。反应物可同时一起加入至溶剂中(例如,单相溶剂、双相水性共溶剂系统和类似溶剂),或者可选择地,一些反应物可分别加入,以及在不同的时间点将一些反应物一起加入。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溶剂中。

[0379] 为了当使用水性共溶剂系统时的改善的混合效率,辅因子再生系统、酮还原酶和辅因子可首先加入和混合入水相。然后可加入和混合有机相,然后加入酮还原酶底物。可选择地,酮还原酶底物可在加入水相之前在有机相中预混合。

[0380] 进行本文所述的酮还原酶催化的还原反应的合适的条件包括可易于通过常规的实验优化的许多条件,所述常规实验包括但不限于,在实验pH和温度下将工程酮还原酶与底物接触,使用在本文所提供的实施例所述的方法检测产物。

[0381] 酮还原酶催化的反应典型地在从约 $15^{\circ}\text{C}$ 至约 $75^{\circ}\text{C}$ 范围的温度下进行。对于一些实施方式,反应在约 $20^{\circ}\text{C}$ 至约 $55^{\circ}\text{C}$ 范围的温度下进行。在另外的实施方式中,反应在约 $20^{\circ}\text{C}$ 至约 $45^{\circ}\text{C}$ 范围的温度下进行。反应也可在环境(ambient)条件下进行。

[0382] 通常允许还原反应进行至获得基本完全或接近完全的底物还原。底物至产物的还原可使用已知的方法通过检测底物和/或产物监控。合适的方法包括气相色谱、HPLC及类似方法。反应混合物中产生的醇还原产物的转化产率通常高于约50%,也可高于约60%,也可高于约70%,也可高于约80%,也可高于约90%,以及经常高于约97%。

## 实施例

[0383] 本公开的不同的特征和实施方式示例于以下有代表性的实施例中,其意为例证性的而不是限制性的。

[0384] 实施例1:酮还原酶粉末的生产;摇瓶程序

[0385] 含有具有感兴趣的酮还原酶基因的质粒的大肠埃希氏菌的单个微生物菌落接种

于盛有包含 30  $\mu$ g/ml 氯霉素和 1% 葡萄糖的 50ml 胰蛋白酶肉汤 (Tryptic broth) (12g/L 细菌用胰蛋白胨、24g/L 酵母提取物、4ml/L 甘油、65mM 磷酸钾, pH 7.0) 的 250ml 锥形瓶 (Erlenmeyer flask) 中。在孵育器中培养细胞过夜 (至少 16hrs), 30°C, 250rpm 振摇。培养物在盛有包含 1mM  $MgSO_4$ 、30  $\mu$ g/ml 氯霉素的 250ml Terrific 肉汤的 1 升烧瓶中稀释至在 600nm 下光密度 (OD600) 为 0.2, 并且允许其在 30°C 下培养。酮还原酶基因的表达在培养物的 OD600 为 0.6 至 0.8 时由 1mM IPTG 诱导, 并且孵育过夜 (至少 16 小时)。通过离心收集细胞 (5000rpm, 15 分钟, 4°C) 并且弃去上清液。细胞沉淀物在等体积的冷 (4°C) 100mM 三乙醇胺 (氯化物) 缓冲液, pH 7.0 (在 ADH-LK 和 ADH-LB 和从其中衍生的工程酮还原酶的情况下包括 1mM  $MgSO_4$ ) 中重悬浮, 通过如上离心收集。清洗的细胞在 12ml 冷三乙醇胺 (氯化物) 缓冲液中重悬浮, 并且在保持 4°C 下通过 12000psi 下的弗氏细胞压碎器两次。细胞碎片通过离心移除 (9000rpm, 45 分钟, 4°C)。收集澄清的裂解物上清液并储存于 -20°C。冻干冷冻的澄清的裂解物产生粗酮还原酶的干燥粉末。

[0386] 实施例 2: 酮还原酶的生产; 发酵程序。

[0387] 在充气搅动的 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 的微量元素溶液的 6.0L 生长培养基达到 30°C 的温度, 所述微量元素溶液包含 2g/L 二水氯化钙、2.2g/L 七水硫酸锌、0.5g/L 一水硫酸锰、1g/L 七水硫酸亚酮、0.1g/L 七水钼酸铵和 0.02g/L 十水四硼酸钠。在如实施例 3 所述的摇瓶中培养至起始 OD600 为 0.5 至 2.0 的包含感兴趣的酮还原酶基因的质粒的指数生长后期的大肠埃希氏菌 W3110 培养物接种至发酵罐。发酵罐在 500-1500rpm 下搅动, 并且空气以 1.0-15.0L/min 提供至发酵容器以保持溶解氧水平在 30% 饱和或更高。培养物的 pH 通过加入 20% v/v 氢氧化铵控制在 7.0。通过加入含 500g/L 结晶葡萄糖 (cerelose)、12g/L 氯化铵和 10.4g/L 七水硫酸镁的料液维持培养物的生长。在培养物达到 OD600 为 50 后, 通过加入异丙基 - $\beta$ -D- 硫代半乳糖苷 (IPTG) 至终浓度 1mM 以诱导酮还原酶的表达。培养物继续生长另外 14 小时。然后将培养物冷却至 4°C 并保持在 4°C 直至收集。用 Sorval RC 12BP 离心机在 4°C 以 5000G 离心 40 分钟收集细胞。收集的细胞直接用于下述下游的回收步骤, 或储存于 4°C 至此类使用。

[0388] 以每个体积的湿细胞糊状物重悬于 2 体积的 100mM 三乙醇胺 (氯化物) 缓冲液 pH 6.8 中, 在 4°C 下重悬浮细胞沉淀物。通过使用 12000psig 的压力, 使悬浮液通过装有两级均化阀 (two-stage homogenizing valve) 装置的匀浆器, 从细胞中释放细胞内酮还原酶。细胞匀浆在破坏后被迅速冷却至 4°C。10% w/v 聚乙烯亚胺溶液, pH 7.2, 加入裂解物中至终浓度 0.5% w/v 并且搅拌 30 分钟。通过用标准实验室离心机在 5000G 下离心 30 分钟使产生的悬浮液澄清。倾出澄清的上清液并使用具有截留分子量为 30Kd 的纤维素超滤膜浓缩 10 倍。将最终浓缩物分散于浅容器中, 在 -20°C 下冷冻并且低冻干为粉末。酮还原酶粉末储存于 -80°C。

[0389] 实施例 3: 检测 2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯立体异构体的分析方法

[0390] 对于常规分析, 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯的 2 个立体异构体以及 2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的 4 个立体异构体通过正相手性 HPLC 分离, 使用 Chiralpak IA 柱 (4.6 $\times$ 150mm (Chiral technologies, cat#80324); 等度 (92% 庚烷, 8% 乙醇); 40°C;

1.5mL/min 流速;样品量:10  $\mu$  L;检测:254nm 下的 UV 吸光度)。保留时间:2S,3R- 立体异构体:6.8min.,2R,3S- 非对映异构体:8.1min,2R,3R- 立体异构体:9.3min,2S,3S- 非对映异构体:10.1min.,底物立体异构体:11.6 和 13.5min。

[0391] 实施例 4:鉴定对于异丙醇 (IPA) 有活性的变体的高通量 NADPH 荧光初筛。

[0392] 通过定向进化获得并且包含进化的酮还原酶基因的质粒文库被转化入大肠埃希氏菌,并在包含 1%葡萄糖和 30  $\mu$  g/mL 氯霉素的 Luria-Bertani (LB) 肉汤培养基上涂板。在 30°C 孵育至少 16 小时后,使用 Q-bot<sup>®</sup> 自动菌落挑选仪 (Genetix USA, Inc., Beaverton, OR) 将菌落挑选至包含 180  $\mu$  L Luria-Bertani (LB)、1%葡萄糖、30  $\mu$  g/mL 氯霉素 (CAM) 和 2mM MgSO<sub>4</sub> 的 96-孔浅孔微量滴定板。在 30°C 下以 200rpm 摇动培养细胞。然后将 20  $\mu$  L 该培养物转移至含有 380  $\mu$  L Terrific 肉汤 (TB)、2mM MgSO<sub>4</sub> 和 30  $\mu$  g/mL CAM 的 96-深孔板中。在 30°C 以 250rpm 摇动孵育深孔板 2.5 至 3 小时后 (OD<sub>600</sub>0.6-0.8),通过终浓度为 1mM 的异丙基硫代半乳糖苷 (IPTG) 诱导细胞培养物的重组基因表达。然后在 30°C 以 250rpm 摇动培养平板 15-23 小时。

[0393] 通过离心沉淀细胞,重悬浮于 400  $\mu$  L 裂解缓冲液中并且通过在室温至少摇动 1 小时裂解细胞。该裂解缓冲液含有 100mM 三乙醇胺 (硫酸盐) 缓冲液,pH 7.0-7.2,1mg/mL 溶菌酶和 200  $\mu$  g/mL 硫酸多粘菌素 B 和 2mM MgSO<sub>4</sub>。然后平板在 4°C 以 4000RPM 用离心机旋转 10 分钟,澄澈的上清液 (澄清的上清液) 用于荧光测定法。

[0394] 在 96-孔黑色微量滴定板中,20  $\mu$  l 的裂解物 (在 100mM 三乙醇胺 (硫酸盐) 缓冲液,pH7.0 中 10 倍稀释) 加入至由 100mM 三乙醇胺 (硫酸盐) 缓冲液,pH 7.0,1mM MgSO<sub>4</sub>,1g/L NADP 和 50%异丙醇组成的 180  $\mu$  l 测定混合物中,并通过在 Flexstation (Molecular Devices, USA) 中 330nm 激发后 445nm 处的 NADPH 的荧光的增加测量反应进程。

[0395] 实施例 5:使用用于辅因子再循环的异丙醇进行的对 2-苯甲酰氨基甲基-3-氧丁酸甲酯的酮还原酶活性的高通量 HPLC 测定法。

[0396] 如实施例 4 中所述制备裂解物。通过转移检测量的细胞裂解物至 Costar 深孔板 (cat#3961) 的孔中测量酮还原酶活性。向 50  $\mu$  L 溶于 100mM 三乙醇胺 (硫酸盐) 缓冲液 pH7.0 的 3mg/ml Na-NADP、150  $\mu$  L 溶于 IPA 中的 3.3mg/ml 2-苯甲酰氨基甲基-3-氧丁酸甲酯中,加入 100  $\mu$  L 澄清的裂解物。将平板用铝/聚丙烯薄片热封口带 (Velocity 11 (Menlo Park, CA), Cat#06643-001) 热封口,在室温下摇动培养 24 小时。反应结束时,向每个孔加入 990  $\mu$  L MTBE,重新将平板封口并摇动 20 分钟。通过离心 (4000rpm,5 分钟,4°C) 分离有机相和水相,并且将 200  $\mu$  L 有机层转移至新的浅孔 96-孔板中以使用实施例 3 中所述方法分析。

[0397] 实施例 6:较高底物浓度下使用用于辅因子再循环的异丙醇进行的对 2-苯甲酰氨基甲基-3-氧丁酸甲酯的酮还原酶活性的高通量 HPLC 测定法。

[0398] 在 96-孔微量滴定板的孔中加入 10  $\mu$  L 溶于 100mM 三乙醇胺 (硫酸盐) 缓冲液 pH7.0 的 0.6mg/ml Na-NADP、150  $\mu$  L 溶于 IPA 的 100mg/ml 2-苯甲酰氨基甲基-3-氧丁酸乙酯、10  $\mu$  L 澄清的裂解物和 130  $\mu$  l 100mM TEA。将平板热封口,在室温下摇动培养 24 小时。反应结束时,向每个孔加入 990  $\mu$  L MTBE,重新将平板封口并摇动 20 分钟。通过离心 (4000rpm,5 分钟,4°C) 将分离有机相和水相中,并且将 200  $\mu$  L 有机层转移至新的浅孔 96-孔板中以使用实施例 3 中所述方法分析。

[0399] 该实施例描述用于鉴定对于 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯的立体选择性还原有改善的 KRED 变体的方法。

[0400] 实施例 7 :ADH-LK 变体对于 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯的还原的评估

[0401] 如实施例 5 所描述地评估几个 ADH-LK 变体对 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯的立体选择性还原。样品如在实施例 3 中描述地进行分析。

[0402] 表 4 列出相应于酮还原酶的 SEQ ID NO, 从 ADH-LK 的氨基酸突变的数目、转化百分比和所需要的 (2S, 3R) - 立体异构体百分比。

[0403] 表 4 :

[0404]

表 4			
SEQ ID NO	从 ADH-LK 的突变 数目	活性 <sup>A</sup>	立体选择性 <sup>B</sup>
90	1	++	+
92	4	++	+
96	1	+	+
98	1	++	+

[0405] A :+ :> 30%转化, ++ :> 60%转化 ;B+ :> 95% (2S, 3R) 异构体。

[0406] 具有 SEQ ID No. 90 的 ADH-LK 变体用于进一步进化用于 (2S, 3R)-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的生成的有效的 KRED。相似地, 主要生成 (2R, 3R) 立体异构体、与 ADH-LK 相比包含 10 个突变、具有 SEQ ID No. 94 的 ADH-LK 变体用于进一步进化用于 (2R, 3R)-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的生成的有效的 KRED。

[0407] 本实施例显示 ADH-LK 的变体可用于 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯的立体选择性还原。产生了 SEQ ID No. 90 和 94 的新变体, 其中移除了 BglI 位点。相应的新序列分别为 SEQ ID No. 48 和 SEQ ID No. 66。

[0408] 实施例 8 :衍生自 ADH-LK 的进一步的工程酮还原酶对 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯至 (2S, 3R)-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的还原。

[0409] 以实施例 6 中所述方法评估衍生自具有 SEQ-ID No. 48 的 ADH-LK 变体的改善的酮还原酶的增加的活性和 (2S, 3R) - 立体选择性。

[0410]

表5: ADH-LK变体将2-苯甲酰氨基甲基-3-氧丁酸甲酯还原为(2S,3R)-2-苯甲酰氨基甲基-3-羟基丁酸甲酯的活性					
SEQ ID No.	从克菲尔乳杆菌的突变	从ADH-LK (SEQ ID:4) 的突变数	活性	稳定性	选择性
2	短乳杆菌				
4	克菲尔乳杆菌				
48	A202V;	1	+		+
38	A94T; E105G; L153A; L199A; A202L; M206F;	6	+		+
16	A94T; S96F; A202V;	3	+		+++
56	L153A; L199A; A202L;	3	+		+++
58	T86I; L199N; A202L;	3	+		+++
52	L153A; A202L;	2	+		+++
54	L153A; A202V;	2	+		+
32	A94T; L199A; A202V;	3	++		++++
34	A94T; L153A; L199H; A202L;	4	+++		++++
50	L153A; L199H; A202L;	3	++		+++
20	A94T; L199N; A202V;	3	+++		++++
46	L153S; A202L;	2	+		+
36	A94T; L153A; L199A; A202V;	4	+		++
26	A94T; A202L;	2	+++		+++
28	A94T; A202V;	2	++		+++
30	A94T; L199A; A202L;	3	+++		++++
22	A94T; L199H; A202L;	3	++++		++++

[0411]

表5: ADH-LK变体将2-苯甲酰氨基甲基-3-氧丁酸甲酯还原为(2S,3R)-2-苯甲酰氨基甲基-3-羟基丁酸甲酯的活性					
SEQ ID No.	从克菲尔乳杆菌的突变	从ADH-LK (SEQ ID:4) 的突变数	活性 <sup>a</sup>	稳定性 <sup>b</sup>	选择性 <sup>c</sup>
24	A94T; L199H; A202V;	3	+++		++++
42	L153A; L199N; A202L;	3	+		+++
40	A94T; S96F; M129T; A202V; M206F;	5	+		++
18	A80T; L153A; A202V;	3	+		+++
44	F147M; A202V;	2	+	+	
10	H40R; A94T; F147L; L199H; A202L;	5	+++++	+	++++
12	H40R; A94T; L199H; A202L;	4	+++++		++++
6	A94T; F147L; L199H; A202L;	4	+++++	+	++++
8	A94T; L199H; A202L;	3	+++++		++++
60	I11F; H40R; A94F; S96V; F147M; L195V; V196L; L199W; I226V; G248K; Y249W;	11	++++		++++
62	T2A; R4C; H40R; A94G; S96V; F147M; V196L; L199W; I226V; G248K; Y249W;	11	+++		++++
14	H40R; A94F; S96V; F147M; L195V; V196L; L199W; I226V; Y249W;	9	++++		++++

[0412] <sup>a</sup>+: 高于具 SEQ ID No. 48 的 KRED 的活性 1-15 倍; ++: 高于具 SEQ ID No. 48 的 KRED 的活性 15-30 倍; +++: 高于具 SEQ ID No. 48 的 KRED 的活性 30-40 倍; ++++: 高于具 SEQ ID No. 48 的 KRED 的活性 40-50 倍; +++++: 高于具 SEQ ID No. 48 的 KRED 的活性 > 50 倍,

[0413] <sup>b</sup>+: 在 40℃ 预孵育 21 小时后保持活性,

[0414]  $^{\circ}$ +:60-89% (2S,3R)-产物; ++:90-94% (2S,3R)-产物; +++:95-99% (2S,3R)-产物; ++++:>99% (2S,3R)-产物.

[0415] 实施例 9:衍生自 ADH-LK 的工程酮还原酶对 2-苯甲酰氨基甲基-3-氧丁酸甲酯至 (2R,3R)-2-苯甲酰氨基甲基-3-羟基丁酸甲酯的还原。

[0416] 以实施例 6 中所述方法评估衍生自具有 SEQ-ID No. 66 的 ADH-LK 变体的改善的酮还原酶的增加的活性和 (2R,3R)-立体选择性。

[0417]

表6: ADH-LK变体将2-苯甲酰氨基甲基-3-氧丁酸甲酯还原为(2R,3R)-2-苯甲酰氨基甲基-3-羟基丁酸甲酯的活性				
SEQ ID NO	序列 - 编码突变	从克菲尔乳杆菌的突变数	活性 <sup>a</sup>	选择性 <sup>b</sup>
66	H40R; A94G; S96V; E145F; F147M; Y190P; V196L; L199W; I225V; Y249W	10	+	+
74	I11F; H40R; A94E; S96V; E145F; F147M; Y190P; L195V; V196L; L199W; I226V; Y249H;	12	++	++
82	D3V; A10T; H40R; A94G; S96V; F147M; Y190P; V196L; L199W; I226V; G248K; Y249H;	12	+	++
68	H40R; A94F; S96V; E145F; F147M; Y190P; L195V; V196L; L199W; I226V; G248R; Y249W;	12	++	++
76	I11L; H40R; A94E; S96V; F147M; Y190H; V196L; I226V; G248K; Y249H;	10	+++	++
72	H40R; T54A; A94F; S96V; E105K; E145D; F147M; V196L; L199W; I226V; Y249W;	11	+++	++
78	I11F; H40R; A94G; S96V; E145F; F147M; Y190H; L195V; V196L; L199W; A202V; I226V; Y249H; A251T;	14	+++	++
70	H40R; E78D; A94E; S96V; F147M; Y190H; L195V; V196L; I226V; Y249H; T250Y;	11	+++	+
80	K8N; V9G; I11F; H40R; A94G; S96V; E145F; F147M; Y190P; V196L; I226V; G248K; Y249R;	13	+++	+
64	V12I; H40R; A94E; S96V; F147M; Y190P; L195V; V196L; L199W; I226V; G248R; Y249W;	12	+++	+

[0418] <sup>a</sup>+: 高于具 SEQ ID No. 66 的 KRED 的活性 1 倍; ++: 高于具 SEQ ID No. 66 的 KRED 的活性 1-2 倍; +++: 高于具 SEQ ID No. 66 的 KRED 的活性 > 2 倍;

[0419] <sup>b+</sup> : < 85% (2R, 3R)-产物 ; ++ : > 85% (2R, 3R)-产物 .

[0420] 实施例 10 : (2S, 3R)-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的制备规模的生产。

[0421] 将 2- 苯甲酰氨基甲基 -3- 酮基丁酸甲酯 (25g)、异丙醇 (37.5ml) 和 0.1M 三乙醇胺 (氯化物)/0.04M MgSO<sub>4</sub> 缓冲液 pH7.2 (30ml) 装入 250ml 的具顶部搅拌器 (overhead stirrer) 的 3- 颈烧瓶。搅拌反应混合物并使用油浴调节温度至 37°C。加入 0.5ml 19g/L NADP-Na 之后加入 2.5ml 30g/L SEQ ID No. 10KRED 使反应开始 ; 两种都为在 0.1M 三乙醇胺 (氯化物)/0.04M MgSO<sub>4</sub> 缓冲液 pH7.2 中的溶液。通过在反应过程中取出 5 μ l 等份, 用 1ml 乙腈稀释, 用 0.25 μ m 注射器过滤器过滤并且按实施例 3 中所述分析以监控反应进程。当转化超过 96% 时, 加饱和的氯化钠水溶液 (12.5ml) 至反应混合物中, 之后加入 40ml 乙酸乙酯。反应混合物继续搅拌 15 分钟, 然后在真空下过滤通过 Celite 垫 (5g 置于多孔玻璃过滤器中)。用 20ml 乙酸乙酯清洗滤饼, 分离滤液的两相。有机层用 20ml 水清洗三次并用旋转蒸发仪浓缩至恒定重量的油状物。移除乙酸乙酯后, 加入甲苯 (10ml) 并真空蒸馏。得到为含 ~ 10% 甲苯的油状物的 (2S, 3R)-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯 (26.47g), 由 <sup>1</sup>H-NMR-CDCl<sub>3</sub> 确定。

[0422] 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整体并入本文, 如同对于所有的目的, 单独地指明每一个单独的出版物、专利、专利申请或其他文件通过引用方式并入本文。

## 序列表

<110> 奥诺拉托·坎帕皮阿诺  
 艾米丽·穆德弗  
 波蒂·博勒普  
 拉玛·沃兰德里

<120> 用于生成氮杂环丁酮的还原酶多肽

<130>376247-022

<141>2008-10-01

<150>60/976, 555

<151>2007-10-01

<160>87

<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 gggatcgggt 60
ttggcaatcg ccacgaaatt tgtagaggag ggtgcgaaag taatgattac tggtcgtcac 120
tccgatgtag gtgaaaaggc cgccaaatca gtaggcactc cggatcagat tcagtttttt 180
cagcacgatt catccgatga agatggctgg acgaaactgt tcgacgccac cgagaaagca 240
ttcggccccg ttagcacctt agtgaacaat gcagggattg cagttaacaa aagcgttgaa 300
gaaactacca cggccgaatg gcgtaaactg ctggccgtta atctggatgg tgtttttttc 360
ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtagcgat ccgagcctgg gggcatacaa cgttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggtcga tgatctgccg 600
ggtgctgagg aagcgatgtc acagcgtacg aaaacccta tgggccacat tggcgaaccg 660
aatgacatcg catatatctg tgtgtacctg gcatctaatt aatcgaaatt tgcgacgggt 720

```







atgaccgatc gtctgaaggc caaagtagcc atcgtaaccg gcgggacact gggatatcgg 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggctcgcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggct ggtaggcgat ccgacgctgg gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatcatgaa 600  
 ggtctggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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	Thr	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 Tyr Ile Lys  
 180 185 190  
 Thr Pro Leu Val Asp Asp His Glu Gly Leu 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

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggacact gggatcgggt 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccggt ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatcatgaa 600  
 ggtctggagg aatgatgctc acagcgtacg aaaacccta tggccacat tggtaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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 Thr 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 His Glu Gly Leu 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 GluPhe 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

```

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggacact gggatcgggt 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcgc 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggct ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggtcga tgatcatgaa 600
ggtctggagg aatgatgtc acagcgtacg aaaacccta tggccacat tggtaaccg 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 Arg 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
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Thr 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
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 His Glu Gly Leu 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
Ala Glu Phe Val Val Asp Gly Gly Tyr Thr Ala Gln		
245	250	

&lt;210&gt;11

&lt;211&gt;759

&lt;212&gt;DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 克菲尔乳杆菌 (L. Kefir) 的变体

&lt;400&gt;11

atgaccgatc gtctgaaggc caaagtagcc atcgtaaccg gcgggacact gggatatcgg 60  
 ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggctcgcgc 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatcatgaa 600  
 ggtctggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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 Arg 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 Thr 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 His Glu Gly Leu 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

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggacact gggatcgggt 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcgt 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccggt ttacgaccgt cgtgaacaat gcagggattt ttgttgtaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaagggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cggtgctcga tgattgggaa 600  
 ggtgctgagg aatgatgctc acagcgtacg aaaacccta tggccacat tgggaaccg 660  
 aatgacatcg catgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggtcgacgg cgggtggacc 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 Arg 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 Phe Val Val
           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 Met 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 Val Leu Asp Asp Trp 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 Val 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 Trp 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

```

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120
gcggatgtag gcgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgtttttta aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgagggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagacce cgctggtcga tgatctggaa 600
ggtgtggagg aatgatgtc acagcgtacg aaaaccctta tggccacat tggtaaccg 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			



atgaccgatc gtctgaaggc caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60  
 ttggcaatcg cegataaatt tgtagaggag ggtgcgaaag tagttattac tggctcgcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggagaca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacggcag gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatctggaa 600  
 ggtgtggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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 Thr  
 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 Ala 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 Val 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>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

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 60  
 ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggcgtcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccggt ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgagggggt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgataatgaa 600  
 ggtgtggagg aatgatgtc acagcgtacg aaaacccta tggccacat tgggaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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 Thr 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 Asn Glu Gly Val 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

```



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
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Thr 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
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 His Glu Gly Leu 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
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> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>23

atgaccgatc gtctgaaggc 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 gcagggatta ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccctgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatcatgaa 600  
 ggtgtggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggtcgacgg cgggtatacc gcacagtga 759

<210>24

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>24

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	Thr	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 His Glu Gly Val 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>25

<211>759

<212>DNA

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>25

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 60  
 ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggcgtcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcgggccgg ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgtttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgagggggt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatctggaa 600  
 ggtctggagg aatgatgtc acagcgtacg aaaacccta tggccacat tgggaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg cgggtatacc gcacagtga 759

<210>26

<211>252

<212>PRT

<213> 人工序列

<220>

<223> 克菲尔乳杆菌 (L. Kefir) 的变体

<400>26

```

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 Thr 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 Leu 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>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

```

atgaccgatac gtctgaaggg caaagtagcc 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
ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgagggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggtcga tgatctggaa 600
ggtgtggagg aatgatgtc acagcgtacg aaaacccta tggccacat tggtaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 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
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Thr 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
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 Val 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
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 gtctgaaggc caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60  
 ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggctcgcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatgcagaa 600  
 ggtctggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtstacctg gcactctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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	Thr	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 Ala Glu Gly Leu 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

```

atgaccgatac 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
ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgttagcaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgagggggt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatgcagaa 600
ggtgtggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggctgacgg 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 Thr 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 Ala Glu Gly Val 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

```

atgaccgatac 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
ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgtagcaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaggggtt cgtaggcgat ccgacggcag gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggtcga tgatcatgaa 600
ggtctggagg aatgatgtc acagegtacg aaaacccta tggccacat tggtaaccg 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			



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 gcagggatta ccgtagcaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacggcag gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatgcagaa 600  
 ggtgtagagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcactctgacg 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 Thr 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 ValPhe 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	Ala	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	Ala	Glu	Gly	Val	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>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

```

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
ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgttagcaa aagcgttgaa 300
gacactacca cgggggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgagggggt cgtaggcgat ccgacggcag gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatgcagaa 600
ggctctggagg aatgttttc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660
aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720
gcagaatttg tggctgacgg 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 Thr Val Ser
           85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Gly 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 Ala 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 Ala Glu Gly Leu Glu Glu Met Phe 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 gtctgaaggg caaagtagcc 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
ttcggcccgg ttacgaccgt cgtgaacaat gcagggatta ccgtttttaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcacgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgagggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggtcga tgatctggaa 600
ggtgtggagg aatgttttc acagegtacg aaaaccctta tggccacat tggtaaccg 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 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	Thr	Val	Phe
			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										
Thr	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	Val	Glu	Glu	Met	Phe	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										

&lt;210&gt;41

&lt;211&gt;759

&lt;212&gt;DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 克菲尔乳杆菌 (L. Kefir) 的变体

&lt;400&gt;41

atgaccgatc gtctgaaggc caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60  
 ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggea ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcatcca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacggcag gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgataatgaa 600  
 ggtctggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcactctgacg 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	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 Ala 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 Asn Glu Gly Leu 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

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcac 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  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaagggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatctggaa 600  
 ggtgtagagg aatgatgtc acagcgtacg aaaacccta tggccacat tggtaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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 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 Met 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 Val 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 gtctgaaggg caaagtagcc atcgtaaccg gcgggactct 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 cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgagggggtt cgtaggcgat ccgacgagcg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagacce cgctggtcga tgatctggaa 600
ggtctggagg aatgatgtc acagegtacg aaaacccta tggccacat tggtaaccg 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 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			



atgaccgatc gtctgaaggc caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60  
 ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggctcgcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatctggaa 600  
 ggtgtagagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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 Tyr Ile Lys  
 180 185 190  
 Thr Pro Leu Val Asp Asp Leu Glu Gly Val 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

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 60  
 ttggcaatcg ccgataaatt tntagaggag 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 gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacggcag gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctgggtcga tgatcatgaa 600  
 ggtctggagg aatgatgtc acagcgtacg aaaacccta tgggccacat tgggaaccg 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 Ala 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 His Glu Gly Leu 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

```

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 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 cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgagggggt cgtaggcgat ccgacggcag gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatctggaa 600
ggtctggagg aatgatgtc acagcgtacg aaaaccctta tggccacat tggtaaccg 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
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 Ala 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 Tyr Ile Lys		
180	185	190
Thr Pro Leu Val Asp Asp Leu Glu Gly Leu 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
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 gtctgaaggc caaagtagcc atcgtaaccg gcgggactct gggatatcgg 60  
 ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggctcgcac 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg catccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcaccctgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacggcag gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatctggaa 600  
 ggtgtagagg aatgatgtc acagcgtacg aaaaccctta tgggccacat tggatgaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcactctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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 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 Ala 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 Val 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

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggactct gggatcgggt 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcac 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  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaggggtt cgtaggcgat ccgacggcag gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctggctga tgatgcagaa 600  
 ggtctggagg aatgatgctc acagcgtacg aaaacccta tggccacat tggtaaccg 660  
 aatgacatcg catggatctg tgtgtacctg gcatctgacg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg 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 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 Ala 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 Ala Glu Gly Leu 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 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 ttacgatcgt cgtgaacaat gcagggattg cagtttccaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgagggggtt cgtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagacce cgctggtcga tgataatgaa 600
ggtctggagg aatgatgtc acagegtacg aaaaccctta tgggccacat tggtaaccg 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 Thr			
1	5	10	15
Leu Gly Ile Gly Leu Ala Ile Ala Asp Lys Phe Val Glu Glu Gly Ala			



atgaccgatc gtctgaaggc caaagtagcc ttcgtaaccg gcgggacact gggatatcgg 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggctcgcgt 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattt ttgttgtaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaagggat ggtaggcgat ccgacgctgg gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cggtgctcga tgattgggaa 600  
 ggtgctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catgggtctg tgtgtacctg gcactctgatg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg caagtggacc 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	Phe	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	Arg	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	Phe	Val	Val
				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 Met 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 Val Leu Asp Asp Trp 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 Val 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 Lys Trp 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

```

atggccgatt gtctgaaggg caaagtagcc atcgtaaccg gcgggacact gggatcgggt 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggtcgctcg 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgct 180
cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg gggttgttaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaagggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctgctcga tgattgggaa 600
ggtgctgagg aatgatgctc acagcgtacg aaaacccta tgggccacat tggngaaccg 660
aatgacatcg catgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg caagtggacc 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 Ala Asp Cys 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 Arg 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 Gly Val Val
           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 Met 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 Leu Asp Asp Trp 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 Val 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 Lys Trp 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

```

atgaccgatac gtctgaaggg caaagtagcc atcataaccg gcgggacact gggatcgggt 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggtcgtcgt 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg aagttgttaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaagggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccg atcaagaccc cggtgctcga tgattgggaa 600
ggtgctgagg aatgatgtc acagegtacg aaaaccctta tgggccacat tggtaaccg 660
aatgacatcg catgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg caggtggacc 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 Ile 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 Arg 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
Phe Gly Pro Val Thr Thr Val Val Asn Asn Ala Gly Ile Glu Val Val		
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 Met Val Gly Asp Pro Thr 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 Pro Ile Lys		
180	185	190
Thr Pro Val Leu Asp Asp Trp 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 Val Cys Val Tyr Leu Ala Ser Asp Glu Ser Lys Phe Ala Thr Gly		
225	230	235
Ala Glu Phe Val Val Asp Gly Arg Trp 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 gtctgaaggc caaagtagcc atcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggtcgctcg 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg gggttgttaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360
ggcaccctgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttttcgggat ggtaggcgat ccgacgctgg gggcatacaa cgctccaag 480
ggggcggtag gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgctgtgca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgattgggaa 600
ggtgctgagg aatgatgctc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660
aatgacatcg catgggtctg tgtgtacctg gcactctgatg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggtggacc 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 Arg 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 Gly Val Val
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  
 Phe Gly Met 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 Trp 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 Val 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 Trp 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

atgaccgatac gtctgaaggg caaagtagcc atcgtaaccg gcgggacact gggatcgggt 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcgt 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccggt ttacgaccgt cgtgaacaat gcagggattt ttgttgtaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttttcgggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cggtgctcga tgattgggaa 600  
 ggtgctgagg aatgatgctc acagcgtacg aaaacccta tggccacat tgggaaccg 660  
 aatgacatcg catgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggtcgacgg cagggtggacc 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 Arg 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 Phe Val Val
           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
Phe Gly Met 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 Val Leu Asp Asp Trp 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 Val 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 Arg Trp 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 gcgggacact gggatcgggt 60
ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcgt 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgatgaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg aagttgttaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaagggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcat atcaagaccc cggtgctcga tgatctggaa 600
ggtgctgagg aatgatgtc acagcgtacg aaaaccctta tgggccacat tggtaaccg 660
aatgacatcg catgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggcattac 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	Arg	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	Asp	Glu	Ala
65			70		75									80	
Phe	Gly	Pro	Val	Thr	Thr	Val	Val	Asn	Asn	Ala	Gly	Ile	Glu	Val	Val
			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	Met	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	His	Ile	Lys
	180		185		190										
Thr	Pro	Val	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	Val	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	His	Tyr	Ala	Gln				
			245		250										

&lt;210&gt;71

&lt;211&gt;759

&lt;212&gt;DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 克菲尔乳杆菌 (L. Kefir) 的变体

&lt;400&gt;71

atgaccgatc gtctgaaggc caaagtagcc atcgtaaccg gcgggacact gggatatcgg 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggctcgcgt 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggcg ctgatgttat tcgctttgtc 180  
 cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattt ttgttgtaa aagcgttgaa 300  
 gacactacca cgaaggaatg gcgtaactg ctgtccgta atctggatgg tgttttttc 360  
 ggcaccgcgc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgatgggat ggtaggcgat ccgacgctgg gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggctat atcaagaccc cgctgctcga tgattgggaa 600  
 ggtgctgagg aatgatgtc acagcgtacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catgggtctg tgtgtacctg gcactctgatg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg cgggtggacc 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 Arg Ala Asp Val Gly Glu Lys Ala Ala  
                   35                   40                   45  
 Lys Ser Ile Gly Gly Ala 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 Phe Val Val  
                   85                   90                   95  
 Lys Ser Val Glu Asp Thr Thr Thr Lys 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  
 Asp Gly Met 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 Leu Asp Asp Trp 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 Val 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 Trp 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

atgaccgatac gtctgaaggg caaagtagcc ttcgtaaccg gcgggacact gggatatcggc 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcgt 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgct 180  
 cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg aagttgttaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttttcgggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cggtgctcga tgattgggaa 600  
 ggtgctgagg aatgatgctc acagcgtacg aaaacccta tgggccacat tggngaaccg 660  
 aatgacatcg catgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggtcgacgg cgggcatacc 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 Phe 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 Arg 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 Glu Val Val
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
Phe Gly Met 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 Val Leu Asp Asp Trp 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 Val 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 His 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 gtctgaaggg caaagtagcc ttcgtaaccg gcgggacact gggatatcgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcgt 120
gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg cgtecgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg gggttgttaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttttcgggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcat atcaagaccc cggtgctcga tgattgggaa 600
ggtgttgagg aatgatgtc acagegtacg aaaacccta tggccacat tggtaaccg 660
aatgacatcg catgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg cgggcatacc acacagtga 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 Gly Lys Val Ala Phe 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	Arg	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	Gly	Val	Val
			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										
Phe	Gly	Met	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	His	Ile	Lys
	180		185		190										
Thr	Pro	Val	Leu	Asp	Asp	Trp	Glu	Gly	Val	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	Val	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	His	Thr	Thr	Gln				
			245		250										

&lt;210&gt;77

&lt;211&gt;759

&lt;212&gt;DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 克菲尔乳杆菌 (L. Kefir) 的变体

&lt;400&gt;77

atgaccgatc gtctgaaggc caaagtagcc ttggtaacgg gcgggacact gggatatcgg 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggctcgcgt 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg aagttgttaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcacccgtc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttgaagggat ggtaggcgat ccgacgctgg gggcatacaa cgctccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggccat atcaagaccc cgctgctcga tgatctggaa 600  
 ggtgctgagg aatgatgtc acagcgaacg aaaacccta tgggccacat tggatgaaccg 660  
 aatgacatcg catgggtctg tgtgtacctg gcactctgat aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggctgacgg aaagcatacc 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 Gly Lys Val Ala Leu 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 Arg 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 Glu Val Val  
                   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 Met 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 His 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 Val 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 Lys His 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

atgaccgatac gtctgaaggg caatggagcc ttcgtaaccg gcgggacact gggatcgggt 60  
 ttggcaatcg ccgataaatt tntagaggag ggtgcgaaag tagttattac tggcgtcgt 120  
 gcggatgtag gtgaaaaggc cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180  
 cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240  
 ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg gggttgttaa aagcgttgaa 300  
 gacactacca cggaggaatg gcgtaactg ctgtccgtta atctggatgg tgttttttc 360  
 ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420  
 atgagcagta ttttcgggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480  
 ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540  
 gtgcgtgtca acacagtaca tccgggcccc atcaagaccc cgctgctcga tgatctggaa 600  
 ggtgctgagg aatgatgctc acagcgtacg aaaacccta tgggccacat tggtaaccg 660  
 aatgacatcg catgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720  
 gcagaatttg tggtcgacgg aaagaggacc 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 Gly Asn Gly Ala Phe 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 Arg 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 Gly Val Val
           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
Phe Gly Met 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 Val 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 Lys Arg 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

```

atgaccgttc gtttgaaggg caaagtaacc atcgtaacgg gcgggacact gggatttgg 60
ttggcaatcg ccgataaatt tgtagaggag ggtgcgaaag tagttattac tggtcgtcgt 120
gcggatgtag gtgaaaagge cgccaaatca atcggcggca ctgatgttat tcgctttgtc 180
cagcacgatg cgtccgatga agcaggctgg acgaaactgt tcgacaccac cgaggaggca 240
ttcggcccgg ttacgaccgt cgtgaacaat gcagggattg gggttgttaa aagcgttgaa 300
gacactacca cggaggaatg gcgtaaaactg ctgtccgtta atctggatgg tgttttttc 360
ggcaccgctc tgggcattca gcgcatgaaa aataaaggct tgggcgctag catcatcaat 420
atgagcagta ttgaagggat ggtaggcgat ccgacgctgg gggcatacaa cgcttccaag 480
ggggcggtac gtatcatgtc gaaaagcgca gcgctggatt gcgcaactgaa ggactacgat 540
gtgcgtgtca acacagtaca tccgggcccc atcaagacce cgctgctcga tgattgggaa 600
ggtgctgagg aatgatgtc acagcgtacg aaaaccctta tgggccacat tggtaaccg 660
aatgacatcg cgtgggtctg tgtgtacctg gcatctgatg aatcgaaatt tgcgacgggt 720
gcagaatttg tggtcgacgg caagcatacc 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 Val Arg Leu Lys Gly Lys Val Thr 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			



<220>

<221>MISC\_FEATURE

<222>(2)..(2)

<223>Xaa 是极性、非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(4)..(4)

<223>Xaa 是碱性或半胱氨酸残基

<220>

<221>MISC\_FEATURE

<222>(11)..(11)

<223>Xaa 是脂肪族、非极性或芳香族残基

<220>

<221>MISC\_FEATURE

<222>(40)..(40)

<223>Xaa 是受限或碱性残基

<220>

<221>MISC\_FEATURE

<222>(80)..(80)

<223>Xaa 是脂肪族、非极性或极性残基

<220>

<221>MISC\_FEATURE

<222>(86)..(86)

<223>Xaa 是极性、脂肪族或非极性残基

<220>

<221>MISC\_FEATURE

<222>(94)..(94)

<223>Xaa 是脂肪族或极性残基

<220>

<221>MISC\_FEATURE

<222>(96)..(96)

<223>Xaa 是极性、芳香族、脂肪族或非极性残基

<220>

<221>MISC\_FEATURE

<222>(105)..(105)

<223>Xaa 是脂肪族、非极性、碱性或酸性残基

<220>

<221>MISC\_FEATURE

<222>(129)..(129)

<223>Xaa 是非极性或极性残基

<220>

<221>MISC\_FEATURE

<222>(147)..(147)

<223>Xaa 是芳香族、非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(153)..(153)

<223>Xaa 是脂肪族、非极性或极性残基

<220>

<221>MISC\_FEATURE

<222>(190)..(190)

<223>Xaa 是芳香族或受限残基

<220>

<221>MISC\_FEATURE

<222>(195)..(195)

<223>Xaa 是非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(196)..(196)

<223>Xaa 是非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(199)..(199)

<223>Xaa 是脂肪族、极性或受限残基

<220>

<221>MISC\_FEATURE

<222>(202)..(202)

<223>Xaa 是缬氨酸或亮氨酸

<220>

<221>MISC\_FEATURE

<222>(206)..(206)

<223>Xaa 是非极性或芳香族残基

<220>

<221>MISC\_FEATURE

<222>(226)..(226)

<223>Xaa 是非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(248)..(248)

<223>Xaa 是非极性或碱性残基

<220>

<221>MISC\_FEATURE

<222>(249)..(249)

<223>Xaa 是芳香族残基

<400>83

Met	Xaa	Asn	Xaa	Leu	Asp	Gly	Lys	Val	Ala	Xaa	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	Xaa	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	Xaa
65				70					75					80	
Phe	Gly	Pro	Val	Ser	Xaa	Leu	Val	Asn	Asn	Ala	Gly	Ile	Xaa	Val	Xaa

	85	90	95
Lys Ser Val Glu Glu Thr Thr Thr Xaa 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
Xaa Lys Asn Lys Gly Leu Gly Ala Ser Ile Ile Asn Met Ser Ser Ile			
	130	135	140
Glu Gly Xaa Val Gly Asp Pro Ser Xaa 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 Xaa Xaa Asp Asp Xaa Pro Gly Xaa Glu Glu Ala 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 Xaa 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 Xaa Xaa Thr Ala Gln			
	245	250	

<210>84

<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>(2)..(2)

<223>Xaa 是极性、非极性或非脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(4)..(4)

<223>Xaa 是碱性或半胱氨酸残基

<220>

<221>MISC\_FEATURE

<222>(11)..(11)

<223>Xaa 是脂肪族、非极性或芳香族残基

<220>

<221>MISC\_FEATURE

<222>(40)..(40)

<223>Xaa 是受限或碱性残基

<220>

<221>MISC\_FEATURE

<222>(80)..(80)

<223>Xaa 是脂肪族、非极性或极性残基

<220>

<221>MISC\_FEATURE

<222>(86)..(86)

<223>Xaa 是极性、脂肪族或非极性残基

<220>

<221>MISC\_FEATURE

<222>(94)..(94)

<223>Xaa 是脂肪族或极性残基

<220>

<221>MISC\_FEATURE

<222>(96)..(96)

<223>Xaa 是极性、芳香族、脂肪族或非极性残基

<220>

<221>MISC\_FEATURE

<222>(105)..(105)

<223>Xaa 是脂肪族、非极性、碱性或酸性残基

<220>

<221>MISC\_FEATURE

<222>(129)..(129)

<223>Xaa 是非极性或极性残基

<220>

<221>MISC\_FEATURE

<222>(147)..(147)

<223>Xaa 是芳香族、非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(153)..(153)

<223>Xaa 是脂肪族、非极性或极性残基

<220>

<221>MISC\_FEATURE

<222>(190)..(190)

<223>Xaa 是芳香族或受限残基

<220>

<221>MISC\_FEATURE

<222>(195)..(195)

<223>Xaa 是非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(196)..(196)

<223>Xaa 是非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(199)..(199)

<223>Xaa 是脂肪族、极性或受限残基

<220>

<221>MISC\_FEATURE

<222>(202)..(202)

<223>Xaa 是缬氨酸或亮氨酸

<220>

<221>MISC\_FEATURE

<222>(206)..(206)

<223>Xaa 是非极性或芳香族残基

<220>

<221>MISC\_FEATURE

<222>(226)..(226)

<223>Xaa 是非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(248)..(248)

<223>Xaa 是非极性或碱性残基

<220>

<221>MISC\_FEATURE

<222>(249)..(249)

<223>Xaa 是芳香族残基

<400>84

```

Met Xaa Asp Xaa Leu Lys Gly Lys Val Ala Xaa 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 Xaa 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 Xaa
65           70           75           80
Phe Gly Pro Val Thr Xaa Val Val Asn Asn Ala Gly Ile Xaa Val Xaa
           85           90           95
Lys Ser Val Glu Asp Thr Thr Thr Xaa 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
Xaa Lys Asn Lys Gly Leu Gly Ala Ser Ile Ile Asn Met Ser Ser Ile
           130          135          140
Glu Gly Xaa Val Gly Asp Pro Thr Xaa 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 Xaa Xaa Asp Asp Xaa Glu Gly Xaa Glu Glu Met 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 Xaa 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 Xaa Xaa Thr Ala Gln			
	245	250	

<210>85

<211>759

<212>DNA

<213> 小乳杆菌 (Lactobacillus Minor)

<400>85

atgaccgatc ggttgaaggg gaaagtagca attgtaactg gcggtacctt gggaattggc	60
ttggcaatcg ctgataagtt tgttgaagaa ggcgcaaagg ttgttattac cggccgtcac	120
gctgatgtag gtgaaaaagc tgccagatca atcggcggca cagacgttat ccgttttgc	180
caacacgatg cttctgatga aaccggctgg actaagttgt ttgatacgac tgaagaagca	240
tttggcccag ttaccacggt tgtcaacaat gccggaattg cggtcagcaa gagtgttgaa	300
gataccacaa ctgaagaatg gcgcaagctg ctctcagtta acttggatgg tgccttcttc	360
ggtaccgctc ttggaatcca acgtatgaag aataaaggac tcggagcatc aatcatcaat	420
atgtcatcta tcgaaggttt tgttggatgat ccagctctgg gtgcatacaa cgcttcaaaa	480
ggtgctgtca gaattatgtc taaatcagct gccttggatt gcgctttgaa ggactacgat	540
gttcgggtta aactgttca tccaggttat atcaagacac catttggttga cgatctttaa	600
ggggcagaag aatgatgtc acageggacc aagacaccaa tgggtcatat cggtgaacct	660
aacgatatcg cttggatctg tgtttacctg gcactctgacg aatctaaatt tgccactggt	720
gcagaattcg ttgtcgacgg agggtagacc gcccaatag	759

<210>86

<211>252

<212>PRT

<213> 小乳杆菌 (Lactobacillus Minor)

<400>86

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>87

<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>(2)..(2)

<223>Xaa 是极性、非极性或非脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(4)..(4)

<223>Xaa 是碱性或半胱氨酸残基

<220>

<221>MISC\_FEATURE

<222>(11)..(11)

<223>Xaa 是脂肪族、非极性或非芳香族残基

<220>

<221>MISC\_FEATURE

<222>(40)..(40)

<223>Xaa 是受限或碱性残基

<220>

<221>MISC\_FEATURE

<222>(80)..(80)

<223>Xaa 是脂肪族、非极性或极性残基

<220>

<221>MISC\_FEATURE

<222>(86)..(86)

<223>Xaa 是极性、脂肪族或非极性残基

<220>

<221>MISC\_FEATURE

<222>(94)..(94)

<223>Xaa 是脂肪族或极性残基

<220>

<221>MISC\_FEATURE

<222>(96)..(96)  
<223>Xaa 是极性、芳香族、脂肪族或非极性残基

<220>  
<221>MISC\_FEATURE  
<222>(105)..(105)  
<223>Xaa 是脂肪族、非极性、碱性或酸性残基

<220>  
<221>MISC\_FEATURE  
<222>(129)..(129)  
<223>Xaa 是非极性或极性残基

<220>  
<221>MISC\_FEATURE  
<222>(147)..(147)  
<223>Xaa 是芳香族、非极性或脂肪族残基

<220>  
<221>MISC\_FEATURE  
<222>(153)..(153)  
<223>Xaa 是脂肪族、非极性或极性残基

<220>  
<221>MISC\_FEATURE  
<222>(190)..(190)  
<223>Xaa 是芳香族或受限残基

<220>  
<221>MISC\_FEATURE  
<222>(195)..(195)  
<223>Xaa 是非极性或脂肪族残基

<220>  
<221>MISC\_FEATURE  
<222>(196)..(196)  
<223>Xaa 是非极性或脂肪族残基

<220>  
<221>MISC\_FEATURE

<222>(199).. (199)

<223>Xaa 是脂肪族、极性或受限残基

<220>

<221>MISC\_FEATURE

<222>(202).. (202)

<223>Xaa 是缬氨酸或亮氨酸

<220>

<221>MISC\_FEATURE

<222>(206).. (206)

<223>Xaa 是非极性或芳香族残基

<220>

<221>MISC\_FEATURE

<222>(226).. (226)

<223>Xaa 是非极性或脂肪族残基

<220>

<221>MISC\_FEATURE

<222>(248).. (248)

<223>Xaa 是非极性或碱性残基

<220>

<221>MISC\_FEATURE

<222>(249).. (249)

<223>Xaa 是芳香族残基

<400>87

Met	Xaa	Asp	Xaa	Leu	Lys	Gly	Lys	Val	Ala	Xaa	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	Xaa	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	Xaa
65						70								75	80

Phe Gly Pro Val Thr Xaa Val Val Asn Asn Ala Gly Ile Xaa Val Xaa			
	85	90	95
Lys Ser Val Glu Asp Thr Thr Thr Xaa 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
Xaa Lys Asn Lys Gly Leu Gly Ala Ser Ile Ile Asn Met Ser Ser Ile			
	130	135	140
Glu Gly Xaa Val Gly Asp Pro Ala Xaa 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 Xaa Xaa Asp Asp Xaa Glu Gly Xaa Glu Glu Met 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 Xaa 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 Xaa Xaa Thr Ala Gln			
	245	250	

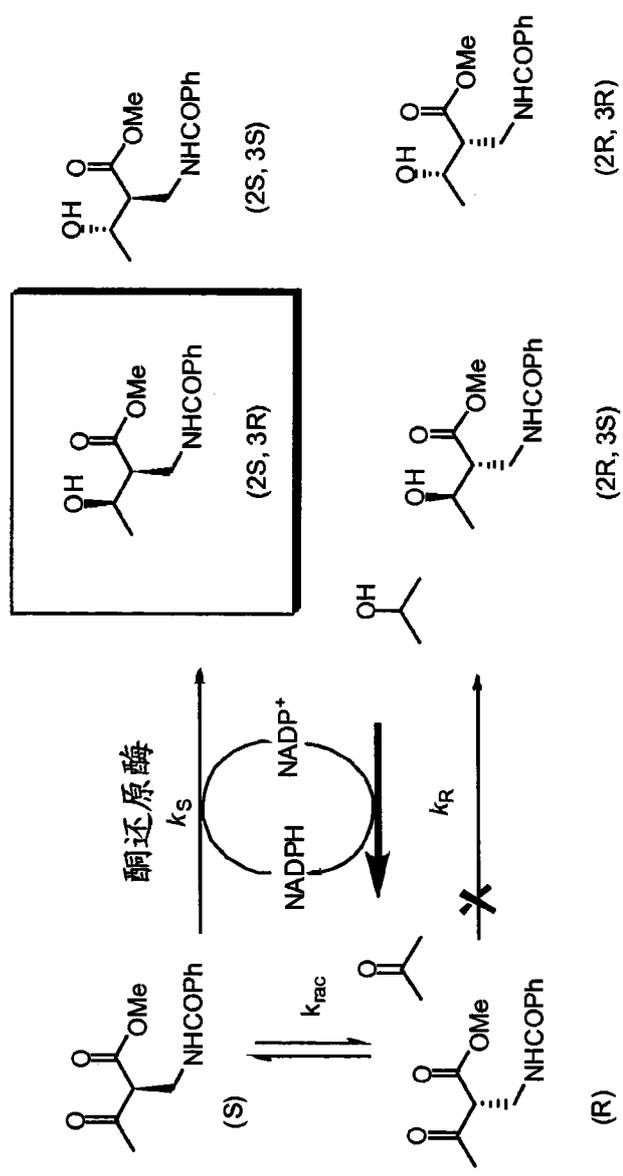


图 1

2-苯甲酰氨基甲基-3-氧-丁酸甲酯      2-苯甲酰氨基甲基-3-羟基-丁酸甲酯

1. 一种酮还原酶多肽,能够以至少约 60% 的立体异构体过量百分比将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。

2. 如权利要求 1 所述的多肽,其包含与基于 SEQ ID NO :2 或 4 或 86 的参考序列具有至少约 85% 同一性的氨基酸序列,所述参考序列具有以下特征:相应于 X94 的残基为苏氨酸;相应于 X199 的残基为组氨酸以及相应于 X202 的残基为缬氨酸或亮氨酸;条件是所述酮还原酶多肽具有其中相应于 X94 的残基为脂肪族或极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。

3. 如权利要求 2 所述的多肽,所述多肽中相应于 X94 的残基为苏氨酸。

4. 如权利要求 2 所述的多肽,所述多肽中相应于 X199 的残基为丙氨酸、组氨酸或天冬酰胺。

5. 如权利要求 2 所述的多肽,所述多肽中相应于 X94 的残基为极性残基;相应于 X199 的残基为脂肪族、受限或极性残基;并且相应于 X202 的残基为缬氨酸或亮氨酸。

6. 如权利要求 2 所述的多肽,所述多肽中相应于 X94 的残基为苏氨酸;相应于 X199 的残基为丙氨酸、组氨酸或天冬酰胺;并且相应于 X202 的残基为缬氨酸或亮氨酸。

7. 如权利要求 2 至 6 中任一项所述的多肽,其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个:

相应于 X2 的残基为极性、非极性或脂肪族残基;

相应于 X4 的残基为碱性残基或半胱氨酸;

相应于 X11 的残基为非极性、脂肪族、或芳香族残基;

相应于 X40 的残基为受限或碱性残基;

相应于 X80 的残基为非极性、脂肪族或极性残基;

相应于 X86 的残基为非极性、脂肪族或极性残基;

相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基;

相应于 X105 的残基为非极性、脂肪族、碱性或酸性残基;

相应于 X129 的残基为非极性或极性残基;

相应于 X147 的残基为芳香族、非极性或脂肪族残基;

相应于 X153 的残基为极性、非极性或脂肪族残基;

相应于 X190 的残基为芳香族或受限残基;

相应于 X195 的残基为非极性或脂肪族残基;

相应于 X196 的残基为非极性或脂肪族残基;

相应于 X206 的残基为非极性或芳香族残基;

相应于 X226 的残基为非极性或脂肪族残基;

相应于 X248 为非极性或碱性残基;

相应于 X249 为芳香族残基;并且

其中与上述参考序列相比,任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

8. 如权利要求 2 至 6 中任一项所述的多肽,其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个:

相应于 X2 的残基为丙氨酸;

相应于 X4 的残基为半胱氨酸；  
相应于 X11 的残基为苯丙氨酸；  
相应于 X40 的残基为精氨酸；  
相应于 X80 的残基为苏氨酸；  
相应于 X86 的残基为异亮氨酸；  
相应于 X96 的残基为缬氨酸或苯丙氨酸；  
相应于 X105 的残基为甘氨酸；  
相应于 X129 的残基为苏氨酸；  
相应于 X147 的残基为蛋氨酸或亮氨酸；  
相应于 X153 的残基为丙氨酸或丝氨酸；  
相应于 X190 的残基为组氨酸或脯氨酸；  
相应于 X195 的残基为缬氨酸；  
相应于 X196 的残基为亮氨酸；  
相应于 X206 的残基为苯丙氨酸；  
相应于 X226 的残基为缬氨酸；  
相应于 X248 的残基为赖氨酸或精氨酸；  
相应于 X249 的残基为色氨酸；

其中与所述参考序列相比，任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

9. 如权利要求 7 所述的多肽，其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个：

相应于 X40 的残基为受限或碱性残基；并且  
相应于 X147 的残基为芳香族、非极性或非脂肪族残基。

10. 如权利要求 9 所述的多肽，其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个：

相应于 X96 的残基为极性、芳香族、非极性或非脂肪族残基；  
相应于 X195 的残基为非极性或非脂肪族残基；  
相应于 X196 的残基为非极性或非脂肪族残基；  
相应于 X226 的残基为非极性或非脂肪族残基；  
相应于 X248 的残基为非极性或非碱性残基；  
相应于 X249 的残基为芳香族残基。

11. 如权利要求 9 或 10 所述的多肽，其中所述酮还原酶的氨基酸序列另外具有以下特征中的一个或多个：

相应于 X2 的残基为极性、非极性或非脂肪族残基；  
相应于 X4 的残基为碱性残基或半胱氨酸；  
相应于 X11 的残基为非极性、脂肪族或芳香族残基；  
相应于 X80 的残基为非极性、脂肪族或极性残基；  
相应于 X86 的残基为非极性、脂肪族或极性残基；  
相应于 X105 的残基为非极性、脂肪族、碱性或酸性残基；

相应于 X129 的残基为非极性或极性残基；  
相应于 X153 的残基为极性、非极性或脂肪族残基；  
相应于 X190 的残基为芳香族或受限残基；  
相应于 X206 的残基为非极性或芳香族残基。

12. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X40 的残基为受限或碱性残基；并且  
相应于 X147 的残基为芳香族、非极性或脂肪族残基；并且  
其中与所述参考序列相比,任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

13. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X40 的残基为受限或碱性残基；  
相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；  
相应于 X147 的残基为芳香族、非极性或脂肪族残基；  
其中与所述参考序列相比,任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

14. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；  
相应于 X147 的残基为芳香族、非极性或脂肪族残基；  
相应于 X195 的残基为非极性或脂肪族残基；  
相应于 X196 的残基为非极性或脂肪族残基；并且  
其中与所述参考序列相比,任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

15. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X40 的残基为精氨酸。

16. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X147 的残基为蛋氨酸或亮氨酸。

17. 如权利要求 7 所述的多肽,其中所述酮还原酶的氨基酸序列具有以下另外的特征：  
相应于 X40 的残基为精氨酸；并且相应于 X147 的残基为蛋氨酸或亮氨酸。

18. 如权利要求 1 所述的多肽,其中所述酮还原酶多肽包含具有以下氨基酸序列的区域或结构域：所述氨基酸序列与基于 SEQ ID NO :2、4 或 86 的参考序列的残基 90 至 211 具有至少 85% 同一性,所述参考序列具有以下特征：相应于 X94 的残基为苏氨酸；相应于 X199 的残基为组氨酸；并且相应于 X202 的残基为缬氨酸或亮氨酸；条件是所述酮还原酶的结构域或区域具有其中相应于 X94 的残基为脂肪族或极性残基；相应于 X199 的残基为脂肪族、受限或极性残基；并且相应于 X202 的残基为亮氨酸或缬氨酸的氨基酸序列。

19. 如权利要求 18 所述的多肽,其中所述酮还原酶的结构域或区域包含其中相应于 X94 的残基为苏氨酸的氨基酸序列。

20. 如权利要求 18 所述的多肽,其中所述酮还原酶的结构域或区域包含其中相应于 X199 的残基为丙氨酸、组氨酸或天冬酰胺的氨基酸序列。

21. 如权利要求 18 所述的多肽,其中所述酮还原酶的结构域或区域包含其中相应于

X94 的残基为极性残基；相应于 X199 的残基为脂肪族、受限或极性残基；并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。

22. 如权利要求 18 所述的多肽，其中所述酮还原酶的结构域或区域包含其中相应于 X94 的残基为苏氨酸；相应于 X199 的残基为丙氨酸、组氨酸或天冬酰胺；并且相应于 X202 的残基为缬氨酸或亮氨酸的氨基酸序列。

23. 如权利要求 18 至 21 中任一项所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征中的一个或多个：

相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；

相应于 X105 的残基为非极性、脂肪族、碱性或酸性残基；

相应于 X129 的残基为非极性或极性残基；

相应于 X147 的残基为芳香族、非极性或脂肪族残基；

相应于 X153 的残基为极性、非极性或脂肪族残基；

相应于 X190 的残基为芳香族或受限残基；

相应于 X195 的残基为非极性或脂肪族残基；

相应于 X196 的残基为非极性或脂肪族残基；

相应于 X206 的残基为非极性或芳香族残基；

其中与所述参考序列相比，所述氨基酸序列可在相应于残基 90-211 的结构域中的其他氨基酸残基位置处任选地具有一个或多个残基差异。

24. 如权利要求 18 至 21 中任一项所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征中的一个或多个：

相应于 X96 的残基为缬氨酸或苯丙氨酸；

相应于 X105 的残基为甘氨酸；

相应于 X129 的残基为苏氨酸；

相应于 X147 的残基为蛋氨酸或亮氨酸；

相应于 X153 的残基为丙氨酸或丝氨酸；

相应于 X190 的残基为组氨酸或脯氨酸；

相应于 X195 的残基为缬氨酸；

相应于 X196 的残基为亮氨酸；

相应于 X206 的残基为苯丙氨酸；

其中与所述参考序列相比，所述氨基酸序列可在相应于残基 90-211 的结构域中的其他氨基酸残基位置处任选地具有一个或多个残基差异。

25. 如权利要求 18 所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征：

相应于 X147 的残基为芳香族、非极性或脂肪族残基；并且

其中与所述参考序列相比，任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

26. 如权利要求 18 所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征：

相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；

相应于 X147 的残基为芳香族、非极性或脂肪族残基；并且

其中与所述参考序列相比，任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

27. 如权利要求 18 所述的多肽，其中相应于残基 90-211 的所述酮还原酶的结构域或区域另外具有以下特征：

相应于 X96 的残基为极性、芳香族、非极性或脂肪族残基；

相应于 X147 的残基为芳香族、非极性或脂肪族残基；

相应于 X195 的残基为非极性或脂肪族残基；

相应于 X196 的残基为非极性或脂肪族残基；

其中与所述参考序列相比，任选地所述氨基酸序列在其他氨基酸残基位置处具有一个或多个残基差异。

28. 如权利要求 1 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、46、50、52、54、56、58、60 和 62 的氨基酸序列。

29. 如权利要求 1 所述的多肽，其能够以至少约 90% 的立体异构体过量百分比将所述底物转化为所述产物。

30. 如权利要求 29 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、40、42、50、52、56、58、60 或 62 的氨基酸序列。

31. 如权利要求 1 所述的多肽，其能够以至少约 95% 的立体异构体过量百分比将所述底物转化为所述产物。

32. 如权利要求 31 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、16、18、20、22、24、26、28、30、32、34、42、50、52、56、58、60 或 62 的氨基酸序列。

33. 如权利要求 1 所述的多肽，其能够以至少约 99% 的立体异构体过量百分比将所述底物转化为所述产物。

34. 如权利要求 33 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、30、32、34、60 和 62 的氨基酸序列。

35. 如权利要求 1 所述的多肽，其能够以比参考多肽 SEQ ID NO :48 高至少 15 倍的速率将所述底物转化为所述产物。

36. 如权利要求 35 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、26、28、30、32、34、50、60 和 62 的氨基酸序列。

37. 如权利要求 1 所述的多肽，其能够以比参考多肽 SEQ ID NO :48 高至少 30 倍的速率将所述底物转化为所述产物。

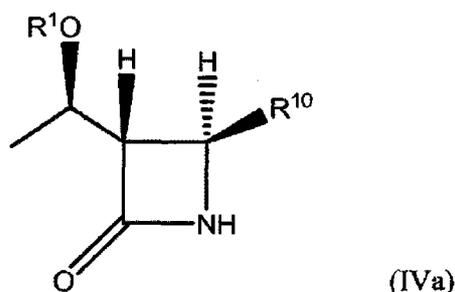
38. 如权利要求 37 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、20、22、24、26、30、34、60 和 62 的氨基酸序列。

39. 如权利要求 1 所述的多肽，其能够以比参考多肽 SEQ ID NO :48 高至少 40 倍的速率将所述底物转化为所述产物。

40. 如权利要求 39 所述的多肽，其包含相应于 SEQ ID NO :6、8、10、12、14、22 和 60 的氨基酸序列。

41. 如权利要求 1 所述的多肽，其能够以比参考多肽 SEQ ID NO :48 高至少 50 倍的速率将所述底物转化为所述产物。

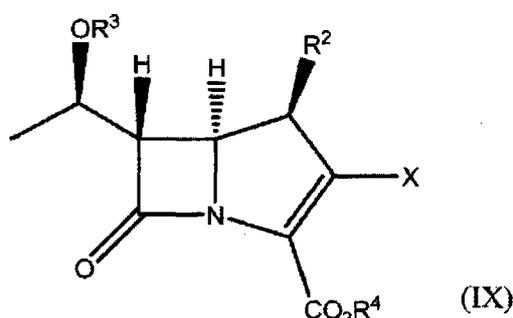
42. 如权利要求 41 所述的多肽,其包含相应于 SEQ ID NO :6、8、10 和 12 的氨基酸序列。
43. 一种酮还原酶多肽,能够以至少约 85% 的立体异构体过量百分比将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯转化为产物 2R, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。
44. 如权利要求 43 所述的多肽,其包含相应于 SEQ ID NO :68、72、74、76、78 和 82 的氨基酸序列。
45. 一种多核苷酸,编码根据权利要求 1-42 中任一项所述的多肽。
46. 如权利要求 45 所述的多核苷酸,其为相应于 SEQ ID NO :5、7、9、11、13、15、17、19、21、23、25、27、29、31、33、35、37、39、41、45、49、51、53、55、57、59 或 61 的序列。
47. 一种多核苷酸,编码根据权利要求 43 或 44 所述的多肽。
48. 如权利要求 47 所述的多核苷酸,其为相应于 SEQ ID NO :67、71、73、75、77 或 81 的序列。
49. 一种表达载体,包含与适用于引导宿主细胞中的表达的控制序列可操作地连接的如权利要求 45 或 47 所述的多核苷酸。
50. 如权利要求 49 所述的表达载体,其中所述控制序列包含启动子。
51. 如权利要求 50 所述的表达载体,其中所述启动子包含大肠埃希氏菌启动子。
52. 如权利要求 50 所述的表达载体,其中所述控制序列包含分泌信号。
53. 一种宿主细胞,包含如权利要求 49 所述的表达载体。
54. 如权利要求 53 所述的宿主细胞,其为大肠埃希氏菌。
55. 一种组合物,包含如权利要求 1-42 中任一项所述的酮还原酶和化合物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯或化合物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯。
56. 如权利要求 55 所述的组合物,其中所述化合物为 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯。
57. 如权利要求 55 所述的组合物,还包含辅因子再生系统。
58. 如权利要求 57 所述的组合物,其中所述辅因子再生系统包括葡萄糖脱氢酶和葡萄糖;甲酸脱氢酶和甲酸;或异丙醇和仲醇脱氢酶。
59. 一种将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯还原为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的方法,所述方法包括在适合将所述底物还原为所述产物的反应条件下,将所述底物与如权利要求 1-42 中任一项所述的酮还原酶多肽接触或孵育。
60. 如权利要求 59 所述的方法,其中所述产物以高于约 99% 的立体异构体过量存在。
61. 如权利要求 59 所述的方法,其使用表达所述酮还原酶的完整细胞、或所述细胞的提取物或裂解物进行。
62. 如权利要求 59 所述的方法,其中所述酮还原酶是分离和 / 或纯化的,并且还原反应在所述酮还原酶的辅因子和任选地所述辅因子的再生系统的存在下进行。
63. 如权利要求 59 所述的方法,其中所述辅因子再生系统包括葡萄糖脱氢酶和葡萄糖;甲酸脱氢酶和甲酸;或异丙醇和仲醇脱氢酶。
64. 如权利要求 63 所述的方法,其中所述仲醇脱氢酶为所述酮还原酶。
65. 一种合成具式 (IVa) 的中间体的方法,



其中  $R^1$  为 H 或羟基保护基团, 并且  $R^{10}$  为卤素或  $-OAc$ , 其中  $Ac$  为乙酸,

所述方法中的步骤包括在适合于将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯还原或转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的反应条件下, 将所述底物与如权利要求 1-42 中任一项所述的酮还原酶接触或反应。

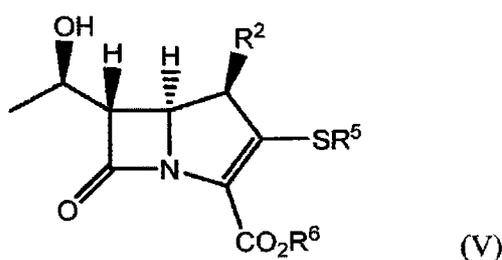
66. 一种合成具结构式 (IX) 的中间体的方法,



其中  $R^2$  为 H 或 C1-C4 烷基;  $R^3$  为 H 或羟基保护基团;  $R^4$  为 H、羧基保护基团、氨基、碱金属或碱土金属; 并且  $X$  为 OH 或离去基团,

所述方法中的步骤包括在适合于将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯还原或转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的反应条件下, 将所述底物与如权利要求 1-42 中任一项所述的酮还原酶接触或反应。

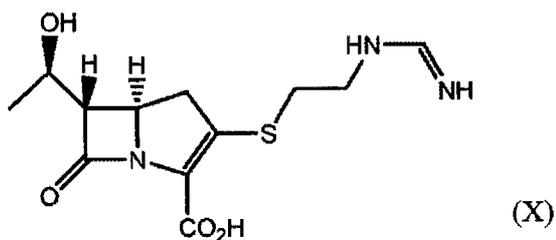
67. 一种合成具结构式 (V) 的碳杂青霉烯或其溶剂化物、水合物、盐和前药的方法:



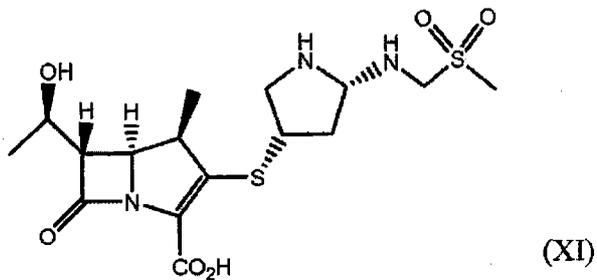
其中  $R^2$  为 H 或  $-CH_3$ ;  $R^5$  选自取代的或未取代的烷基、取代的或未取代的芳香基、取代的或未取代的杂烷基、取代的或未取代的杂环烷基、以及取代的或未取代的杂芳香基烷基; 并且  $R^6$  为 H 或前体基团,

所述方法中的步骤包括在适合于将底物 2- 苯甲酰氨基甲基 -3- 氧丁酸甲酯还原或转化为产物 2S, 3R-2- 苯甲酰氨基甲基 -3- 羟基丁酸甲酯的反应条件下, 将所述底物与如权利要求 1-42 中任一项所述的酮还原酶接触或反应。

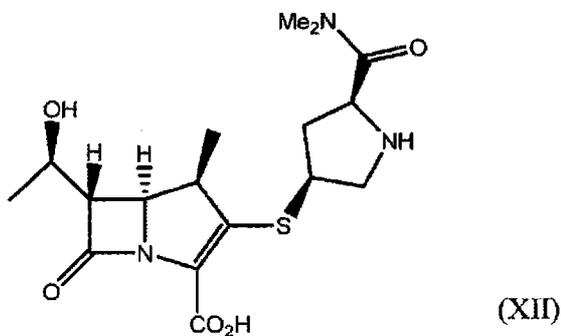
68. 如权利要求 67 所述的方法, 其中所述碳杂青霉烯具有结构式 (X):



69. 如权利要求 67 所述的方法, 其中所述碳杂青霉烯具有结构式 (XI) :



70. 如权利要求 67 所述的方法, 其中所述碳杂青霉烯具有结构式 (XII) :



71. 如权利要求 67 所述的方法, 其中所述碳杂青霉烯具有结构式 (XIII) :

