



(12)发明专利

(10)授权公告号 CN 103289908 B

(45)授权公告日 2018.05.11

(21)申请号 201210316338.9

(22)申请日 2007.10.01

(65)同一申请的已公布的文献号

申请公布号 CN 103289908 A

(43)申请公布日 2013.09.11

(30)优先权数据

06121633.9 2006.10.02 EP

60/848,357 2006.10.02 US

(62)分案原申请数据

200780036866.6 2007.10.01

(83)生物保藏信息

CBS 121879 2007.09.20

CBS 120327 2006.09.27

CBS 120328 2006.09.27

(73)专利权人 帝斯曼知识产权资产管理有限公司
地址 荷兰海尔伦

(72)发明人 安东尼斯·杰若恩·阿迪瑞安·马
里斯·范
雅各布斯·托马斯·普若克
汉迪瑞克·乌特尔·威塞林克
强汉斯·皮埃特·范迪肯
亚伦·阿迪瑞安·文科勒
强汉斯·汉迪瑞克·温迪·德

(74)专利代理机构 北京东方亿思知识产权代理
有限责任公司 11258

代理人 肖善强

(51)Int.Cl.

C12N 1/19(2006.01)

C12N 15/81(2006.01)

C12P 7/06(2006.01)

C12P 7/56(2006.01)

C12P 7/42(2006.01)

C12P 7/54(2006.01)

C12P 7/40(2006.01)

C12P 7/46(2006.01)

C12P 7/48(2006.01)

C12P 7/18(2006.01)

C12P 5/02(2006.01)

C12P 7/20(2006.01)

C12P 7/16(2006.01)

(续)

(56)对比文件

WO 2006009434 A1, 2006.01.26, (续)

审查员 王莉

权利要求书3页 说明书30页

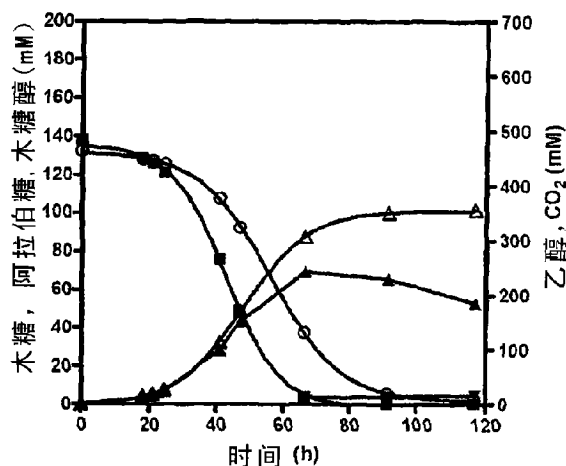
序列表44页 附图4页

(54)发明名称

发酵阿拉伯糖的真核细胞的代谢工程改造

(57)摘要

本发明涉及发酵阿拉伯糖的真核细胞的代谢工程改造。具体地,本发明涉及表达下述多核苷酸序列的真核细胞,所述序列编码ara A、ara B和ara D酶,从而这些核苷酸序列的表达赋予细胞使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物如乙醇的能力。任选地,该真核细胞还能够将木糖转化为乙醇。



[转续页]

[接上页]

(51)Int.Cl.

C12P 35/00(2006.01)

(56)对比文件

Miroslav Sedlak et al.Expression of E. coli araBAD operon encoding enzymes for metabolizing L-arabinose in Saccharomyces cerevisiae.《Enzyme and Microbial Technology》.2001,第28卷16-24.

Jessica Becker et al.A Modified

Saccharomyces cerevisiae Strain That Consumes L-Arabinose and Produces Ethanol.《APPLIED AND ENVIRONMENTAL MICROBIOLOGY》.2003,第69卷(第7期),4144-4150.

Michiel Kleerebezem et al.Complete genome sequence of Lactobacillus plantarum WCFS1.《PNAS》.2003,第100卷(第4期),1990-1995.

1.能够在L-阿拉伯糖上生长且表达以下核苷酸序列的酵母细胞,其中这些核苷酸序列的表达赋予细胞使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物的能力:

(a) 编码阿拉伯糖异构酶(araA)的核苷酸序列,其中所述核苷酸序列选自由以下组成的组:

- i. 编码araA的核苷酸序列,所述araA的氨基酸序列为SEQ ID NO:1的氨基酸序列,
- ii. 核苷酸序列,其为SEQ ID NO:2的核苷酸序列,

(b) 编码L-核酮糖激酶(araB)的核苷酸序列,其中所述核苷酸序列选自由以下组成的组:

- i. 编码araB的核苷酸序列,所述araB的氨基酸序列为SEQ ID NO:3的氨基酸序列,
- ii. 核苷酸序列,其为SEQ ID NO:4的核苷酸序列,

(c) 编码L-核酮糖-5-P-4-差向异构酶(araD)的核苷酸序列,其中所述核苷酸序列选自由以下组成的组:

- i. 编码araD的核苷酸序列,所述araD的氨基酸序列为SEQ ID NO:5的氨基酸序列,
- ii. 核苷酸序列,其为SEQ ID NO:6的核苷酸序列,

其中这些核苷酸序列的表达赋予所述细胞使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物的能力,以及其中所述araA、araB和araD核苷酸序列中的一个、两个或三个来源于*Lactobacillus plantarum*的种。

2.根据权利要求1的细胞,其中所述细胞属于以下属之一:*Saccharomyces*、*Kluyveromyces*、*Candida*、*Pichia*、*Schizosaccharomyces*、*Hansenula*、*Kloeckera*、*Schwanniomyces*或*Yarrowia*。

3.根据权利要求2的细胞,其中所述酵母细胞属于以下种之一:*S.cerevisiae*、*S.bulderi*、*S.barnetti*、*S.exiguus*、*S.uvarum*、*S.diastaticus*、*K.lactis*、*K.marxianus*或*K.fragilis*。

4.根据权利要求1的细胞,其中编码araA、araB和/或araD的核苷酸序列与下述启动子可操作地连接,所述启动子引起相应的核苷酸序列在所述细胞中的充分表达,从而赋予细胞使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或期望的发酵产物的能力。

5.根据权利要求1的细胞,其中所述细胞显示直接将木糖异构化为木酮糖的能力。

6.根据权利要求5的细胞,其中所述细胞包含遗传修饰,所述遗传修饰能提高戊糖磷酸途径的通量。

7.根据权利要求6的细胞,其中所述遗传修饰包括过表达戊糖磷酸途径非氧化部分的至少一个基因。

8.根据权利要求7的细胞,其中所述基因选自由以下组成的组:编码5-磷酸核酮糖异构酶、5-磷酸核酮糖差向异构酶、转酮酶和转醛酶的基因。

9.根据权利要求8的细胞,其中所述遗传修饰包括至少过表达编码转酮酶和转醛酶的基因。

10.根据权利要求9的细胞,其中该细胞还包括提高木酮糖激酶比活性的遗传修饰。

11.根据权利要求10的细胞,其中所述能提高戊糖磷酸途径的通量的遗传修饰包括过

表达编码木酮糖激酶的基因。

12. 根据权利要求9的细胞,其中所述被过表达的基因对所述细胞而言是内源的。

13. 根据权利要求1的细胞,其中所述细胞包含降低所述细胞中非特异性醛糖还原酶活性的遗传修饰。

14. 根据权利要求13的细胞,其中所述遗传修饰降低编码非特异性醛糖还原酶基因的表达,或使所述基因失活。

15. 根据权利要求14的细胞,其中通过缺失所述编码非特异性醛糖还原酶的基因的至少部分或通过破坏所述编码非特异性醛糖还原酶的基因,使该基因失活。

16. 根据权利要求15的细胞,其中编码非特异性醛糖还原酶的每个基因在所述细胞中的表达被降低或失活。

17. 根据权利要求1的细胞,其中所述发酵产物选自由以下组成的组:乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、 β -内酰胺抗生素和头孢菌素。

18. 核酸构建体,所述核酸构建体包含以下的编码araA的核酸序列、编码araB的核酸序列和/或编码araD的核酸序列:

(a) 编码阿拉伯糖异构酶(araA)的核苷酸序列,其中所述核苷酸序列选自由以下组成的组:

- i. 编码araA的核苷酸序列,所述araA的氨基酸序列为SEQ ID NO:1的氨基酸序列,
- ii. 核苷酸序列,其为SEQ ID NO:2的核苷酸序列,

(b) 编码L-核酮糖激酶(araB)的核苷酸序列,其中所述核苷酸序列选自由以下组成的组:

- i. 编码araB的核苷酸序列,所述araB的氨基酸序列为SEQ ID NO:3的氨基酸序列,
- ii. 核苷酸序列,其为SEQ ID NO:4的核苷酸序列,

(c) 编码L-核酮糖-5-P-4-差向异构酶(araD)的核苷酸序列,其中所述核苷酸序列选自由以下组成的组:

- i. 编码araD的核苷酸序列,所述araD的氨基酸序列为SEQ ID NO:5的氨基酸序列,
- ii. 核苷酸序列,其为SEQ ID NO:6的核苷酸序列,

其中这些核苷酸序列的表达赋予细胞使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物的能力,以及其中所述araA、araB和araD核苷酸序列中的一个、两个或三个来源于*Lactobacillus plantarum*的种。

19. 用于生产选自下组的发酵产物的工艺,所述组由乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、 β -内酰胺抗生素和头孢菌素组成,所述工艺包括:

(a) 用根据权利要求1到15中任一项的经修饰的细胞来发酵培养基,所述培养基含有阿拉伯糖来源和可选的木糖来源,从而所述细胞将阿拉伯糖和可选的木糖发酵为发酵产物;和可选地

(b) 回收所述发酵产物。

20. 用于生产选自下组的发酵产物的工艺,所述组由乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、 β -内酰胺

胺抗生素和头孢菌素组成,所述工艺包括:

(a) 用根据权利要求1到15中任一项细胞的和能够使用木糖和/或显示将木糖直接异构化为木酮糖的能力的细胞来发酵培养基,所述培养基至少含有阿拉伯糖来源和木糖来源,从而每种细胞都能将阿拉伯糖和/或木糖发酵为发酵产物;和可选地

(b) 回收所述发酵产物。

21. 根据权利要求19的工艺,其中所述培养基还含有葡萄糖来源。

22. 根据权利要求21的工艺,其中所述发酵产物是乙醇。

23. 根据权利要求22的工艺,其中所述乙醇体积生产力为至少每小时每升0.5g乙醇。

24. 根据权利要求22的工艺,其中所述乙醇产率为至少30%。

25. 根据权利要求20的工艺,其中所述工艺是厌氧的。

26. 根据权利要求20的工艺,其中所述工艺是需氧的。

27. 根据权利要求26的工艺,其中所述工艺在氧气受限的条件下进行。

发酵阿拉伯糖的真核细胞的代谢工程改造

[0001] 本申请是中国发明专利申请200780036866.6 (PCT/NL2007/000246) 的分案申请, 母案申请日为2007年10月1日, 题目为“发酵阿拉伯糖的真核细胞的代谢工程改造”。

发明领域

[0002] 本发明涉及下述真核细胞和使用该细胞生产发酵产物的工艺, 所述细胞具有使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物的能力。

[0003] 发明背景

[0004] 乙醇染料被认为是化石燃料的有价值的备选方案。从植物生物质的半纤维素级分到乙醇的经济学可行的生产需要以相当的速率和高产率同时发酵转化戊糖和己糖这二者。酵母(尤其是*Saccharomyces* spp.) 是该工艺的最适当的候选者, 因为在需氧和厌氧时它们均能够在己糖上快速生长和发酵。另外, 它们对木质纤维素水解产物的毒性环境比(经一般修饰的)细菌有抗性得多。

[0005] EP 1499708描述了制造下述*S.cerevisiae*菌株的工艺, 所述菌株能够从L-阿拉伯糖生产乙醇。通过引入*Bacillus subtilis*的araA (L-阿拉伯糖异构酶、来自*Escherichia coli*的araB (L-核酮糖激酶) 和araD (L-核酮糖-5-P₄-差向异构酶) 基因来修饰这些菌株。另外, 这些菌株基因组中带有额外的突变, 或者过表达TAL1 (转醛酶) 基因。然而, 这些基因具有若干缺点。它们在氧受限的条件下发酵阿拉伯糖。另外, 它们具有0.05g·g⁻¹·h⁻¹的低乙醇生产速率(Becker and Boles, 2003)。另外, 这些菌株不能够在厌氧条件下使用L-阿拉伯糖。最后, 这些*S.cerevisiae*菌株具有野生型背景, 因此它们不能被用于共同发酵若干种C5糖。

[0006] WO 03/062430和WO 06/009434公开了能够将木糖转化为乙醇的酵母菌株。这些酵母菌株能够直接将木糖异构化为木酮糖。

[0007] 仍然存在对生产乙醇的下述备选菌株的需要, 所述菌株运行更好, 并且更有活力和对相对苛刻的生产条件更具抗性。

附图说明

[0008] 图1: pRW231和pRW243的质粒图谱。

[0009] 图2: 在含0.5%半乳糖(A)和0.1%半乳糖+2% L-阿拉伯糖(B)的合成培养基中, 菌株RWB219(○)和IMS0001(●)摇瓶培养的生长模式。培养物在含半乳糖(A)的合成培养基中培养72小时, 然后被转移至含半乳糖和阿拉伯糖(B)的合成培养基。通过测量OD₆₆₀来测定生长。

[0010] 图3. *S.cerevisiae* IMS0001在摇瓶培养物中系列转移期间的生长速率, 所述培养物含有含2% (w/v) L-阿拉伯糖的合成培养基。每个数据点代表从(指数)生长期间测量的OD₆₆₀评价的生长速率。闭合和开放的圆圈代表一式两份的系列转移实验。

[0011] 图4. *S.cerevisiae* IMS0001在含2% (w/v) L-阿拉伯糖的合成培养基中厌氧SBR发

酵期间的生长速率。每个数据点代表从指数生长期间的CO₂谱(实线)评价的生长速率。

[0012] 图5. 菌株IMS0002厌氧分批发酵期间的糖消耗和产物形成。发酵在1升合成培养基中进行,所述培养基补充有20g l⁻¹阿拉伯糖(A);20g l⁻¹葡萄糖和20g l⁻¹阿拉伯糖(B);30g l⁻¹葡萄糖、15g l⁻¹木糖和15g l⁻¹阿拉伯糖(C)。用菌株IMS0002和RWB218的混合物进行的厌氧分批发酵期间的糖消耗和产物形成。发酵在1升合成培养基中进行,所述培养基补充有30g l⁻¹葡萄糖、15g l⁻¹木糖和15g l⁻¹阿拉伯糖(D)。符号:葡萄糖(●);木糖(○);阿拉伯糖(■);从累计的CO₂生产计算的乙醇(□);通过HPLC测量的乙醇(▲);累计的CO₂生产(△);木糖醇(▼)。

[0013] 图6. 选择用于在木糖上厌氧生长的菌株IMS0002细胞厌氧分批发酵期间的糖消耗和产物形成。发酵在1升合成培养基中进行,所述培养基补充有20g l⁻¹木糖和20g l⁻¹阿拉伯糖(B)。符号:木糖(○);阿拉伯糖(■);通过HPLC测量的乙醇(▲);累计的CO₂生产(△);木糖醇(▼)。

[0014] 图7. 菌株IMS0003厌氧分批发酵期间的糖消耗和产物形成。发酵在1升合成培养基中进行,所述培养基补充有:30g l⁻¹葡萄糖、15g l⁻¹木糖和15g l⁻¹阿拉伯糖。符号:葡萄糖(●);木糖(○);阿拉伯糖(■);从累计的CO₂生产计算的乙醇(□);通过HPLC测量的乙醇(▲);累计的CO₂生产(△)。

[0015] 发明详述

[0016] 真核细胞

[0017] 本发明第一方面涉及能够表达以下核苷酸序列的真核细胞,其中,这些核苷酸序列的表达赋予细胞使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物如乙醇的能力:

[0018] (a) 编码阿拉伯糖异构酶(araA)的核苷酸序列,其中所述核苷酸序列选自由以下组成的组:

[0019] (i) 编码araA的核苷酸序列,所述araA包含与SEQ ID NO:1的氨基酸序列具有至少55%序列同一性的氨基酸序列,

[0020] (ii) 核苷酸序列,其包含与SEQ ID NO:2的核苷酸序列具有至少60%序列同一性的核苷酸序列,

[0021] (iii) 核苷酸序列,其互补链与(i)或(ii)的序列的核酸分子杂交,

[0022] (iv) 核苷酸序列,其序列由于遗传密码子的简并性而与(iii)的核酸分子的序列不同;

[0023] (b) 编码L-核酮糖激酶(araB)的核苷酸序列,其中所述核苷酸序列选自由以下组成的组:

[0024] (i) 编码araB的核苷酸序列,所述araB包含与SEQ ID NO:3的氨基酸序列具有至少20%序列同一性的氨基酸序列,

[0025] (ii) 核苷酸序列,其包含与SEQ ID NO:4的核苷酸序列具有至少50%序列同一性的核苷酸序列,

[0026] (iii) 核苷酸序列,其互补链与(i)或(ii)的序列的核酸分子杂交,

[0027] (iv) 核苷酸序列,其序列由于遗传密码子的简并性而与(iii)的核酸分子的序列不同;

[0028] (c) 编码L-核酮糖-5-P-4-差向异构酶 (araD) 的核苷酸序列, 其中所述核苷酸序列选自自由以下组成的组:

[0029] (i) 编码araD的核苷酸序列, 所述araD包含与SEQ ID NO:5的氨基酸序列具有至少60%序列同一性的氨基酸序列,

[0030] (ii) 核苷酸序列, 其包含与SEQ ID NO:6的核苷酸序列具有至少60%序列同一性的核苷酸序列,

[0031] (iii) 核苷酸序列, 其互补链与(i)或(ii)的序列的核酸分子杂交,

[0032] (iv) 核苷酸序列, 其序列由于遗传密码子的简并性而与(iii)的核酸分子的序列不同。

[0033] 一个优选的实施方案涉及能够表达以下核苷酸序列的真核细胞, 从而这些核苷酸序列的表达赋予细胞使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物如乙醇的能力:

[0034] (a) 编码阿拉伯糖异构酶 (araA) 的核苷酸序列, 其中所述核苷酸序列选自自由以下组成的组:

[0035] (i). 核苷酸序列, 其包含与SEQ ID NO:2的核苷酸序列具有至少60%序列同一性的核苷酸序列,

[0036] (ii). 核苷酸序列, 其互补链与(i)的序列的核酸分子杂交,

[0037] (iii). 核苷酸序列, 其序列由于遗传密码子的简并性而与(ii)的核酸分子的序列不同;

[0038] (b) 编码L-核酮糖激酶 (araB) 的核苷酸序列, 其中所述核苷酸序列选自自由以下组成的组:

[0039] (i). 编码araB的核苷酸序列, 所述araB包含与SEQ ID NO:3的氨基酸序列具有至少20%序列同一性的氨基酸序列,

[0040] (ii). 核苷酸序列, 其包含与SEQ ID NO:4的核苷酸序列具有至少50%序列同一性的核苷酸序列,

[0041] (iii). 核苷酸序列, 其互补链与(i)或(ii)的序列的核酸分子杂交,

[0042] (iv). 核苷酸序列, 其序列由于遗传密码子的简并性而与(iii)的核酸分子的序列不同;

[0043] (c) 编码L-核酮糖-5-P-4-差向异构酶 (araD) 的核苷酸序列, 其中所述核苷酸序列选自自由以下组成的组:

[0044] (i). 编码araD的核苷酸序列, 所述araD包含与SEQ ID NO:5的氨基酸序列具有至少60%序列同一性的氨基酸序列,

[0045] (ii). 核苷酸序列, 其包含与SEQ ID NO:6的核苷酸序列具有至少60%序列同一性的核苷酸序列,

[0046] (iii). 核苷酸序列, 其互补链与(i)或(ii)的序列的核酸分子杂交,

[0047] (iv). 核苷酸序列, 其序列由于遗传密码子的简并性而与(iii)的核酸分子的序列不同。

[0048] 序列同一性和相似性

[0049] 序列同一性在本文中被定义为: 通过比较序列测定的, 两条或更多条氨基酸(多肽

或蛋白质)序列或两条或更多条核酸(多核苷酸)序列之间的关系。通常,在被比较的序列的整个长度上比较序列的同一性或相似性。在本领域内,“同一性”也表示氨基酸或核酸序列之间的序列相关性(relatedness)程度,所述相关性程度根据情况由这类序列的字符串之间的匹配测定。通过将一个多肽的氨基酸序列及其保守的氨基酸取代与另一多肽进行比较,测定两条氨基酸序列之间的“相似性”。“同一性”和“相似性”可通过本领域技术人员已知的多种方法容易地计算。

[0050] 测定同一性的优选方法被设计为在被测试的序列之间给出最大匹配。测定同一性和相似性的方法被编码于公众可获得的计算机程序中。测定两条序列之间同一性和相似性的优选的计算机程序方法包括例如BestFit、BLASTP、BLASTN和FASTA(Altschul, S.F. et al., J.Mol.Biol.215:403-410(1990), 公众可从NCBI和其它来源获得(BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894)。使用的一种最优选的算法是EMBOSS(<http://www.ebi.ac.uk/emboss/align>)。使用EMBOSS进行氨基酸序列比较的优选的参数为缺口开放10.0,缺口延伸0.5,Blosum 62矩阵。使用EMBOSS进行核酸序列比较的优选的参数为缺口开放10.0,缺口延伸0.5,DNA完全矩阵(DNA同一性矩阵)。

[0051] 可选地,在测定氨基酸相似性程度时,技术人员也可考虑所谓的“保守”氨基酸取代,这是技术人员应当明白的。保守氨基酸取代是指具有相似侧链的残基的互换性。例如,具有脂肪族侧链的氨基酸组为甘氨酸、丙氨酸、缬氨酸、亮氨酸和异亮氨酸;具有脂肪族-羟基测量的氨基酸组是丝氨酸和苏氨酸;具有含酰胺侧链的氨基酸组是天冬酰胺和谷氨酰胺;具有芳香族侧链的氨基酸组是苯丙氨酸、酪氨酸和色氨酸;具有碱性侧链的氨基酸组为赖氨酸、精氨酸和组氨酸;具有含硫侧链的氨基酸组为半胱氨酸和甲硫氨酸。优选的保守氨基酸取代基为:缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸和天冬酰胺-谷氨酰胺。本文公开的氨基酸序列的取代变体是下述变体,其中公开的序列中至少一个残基被去除,并在其位置中插入不同的残基。优选地,氨基酸改变是保守的。对各天然存在的氨基酸的优选的保守取代如下:Ala到ser;Arg到lys;Asn到gln或his;Asp到glu;Cys到ser或ala;Gln到asn;Glu到asp;Gly到pro;His到asn或gln;Ile到leu或val;Leu到ile或val;Lys到arg;gln或glu;Met到leu或ile;Phe到met、leu或tyr;Ser到thr;Thr到ser;Trp到tyr;Tyr到trp或phe;和Val到ile或leu。

[0052] 杂交核酸序列

[0053] 编码在本发明的细胞中表达的酶的核苷酸序列也可以通过它们分别与SEQ ID NO.2、4、6、8、16、18、20、22、24、26、28、30的核苷酸序列在中度杂交条件下(或优选地在严格杂交条件下)杂交的能力来定义。严格杂交条件在本文中定义为下述条件,其允许至少约25个、优选地约50个核苷酸、75或100个和最优选地约200或更多个核苷酸的核酸序列,在约65℃的温度下,在含约1M盐、优选地6xSSC或具有相当的离子强度的任何其它溶液中杂交,并在65℃下在含约0.1M盐或更少、优选地0.2xSSC或具有相当的离子强度的任何其它溶液中洗涤。优选地,杂交进行过夜,即至少进行10小时,并且优选地,洗涤进行至少一个小时,至少更换两次洗涤溶液。这些条件通常会允许具有约90%或更高序列同一性的序列特异杂交。

[0054] 中度条件在本文中定义为下述条件,所述条件允许至少50个核苷酸、优选地约200个或更多个核苷酸的核酸序列,在约45℃的温度下,在含约1M盐、优选地6xSSC或具有相当

的离子强度的任何其它溶液中杂交,并在室温下在含约0.1M盐、优选地6xSSC或具有相当的离子强度的任何其它溶液中洗涤。优选地,杂交过夜进行,即至少进行10小时,并且优选地将洗涤进行至少一个小时,至少两次更换洗涤溶液。这些条件通常会允许具有至多50%序列同一性的序列特异杂交。本领域技术人员将能改变这些杂交条件,从而特异地鉴定同一性在50%和90%之间变化的序列。

[0055] AraA

[0056] 编码在本发明的细胞中表达的阿拉伯糖异构酶(araA)的一个优选的核苷酸序列选自自由以下组成的组:

[0057] (a) 编码araA多肽的核苷酸序列,所述araA包含与SEQ ID NO.1的氨基酸序列至少具有55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、98%或99%序列同一性的氨基酸序列;

[0058] (b) 核苷酸序列,其包含与SEQ ID NO.2的核苷酸序列至少具有60%、70%、80%、90%、95%、97%、98%或99%序列同一性的核苷酸序列;

[0059] (c) 核苷酸序列,其互补链与(a)或(b)的核酸分子序列杂交;

[0060] (d) 核苷酸序列,其序列由于遗传密码子的简并性而与(c)的核酸分子的序列不同。

[0061] 编码araA的核苷酸序列可编码原核或真核araA,即具有下述氨基酸序列的araA,所述氨基酸序列与原核生物或真核生物中天然存在的araA的氨基酸序列相同。本发明人发现与araB和araD共表达时,具体araA的下述能力不太取决于该araA是原核或是真核来源,所述能力是赋予真核宿主细胞使用阿拉伯糖和/或将阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物如乙醇的能力。更确切地说,这取决于araA的氨基酸序列与SEQ ID NO.1序列的相关性。

[0062] AraB

[0063] 编码在本发明的细胞中表达的L-核酮糖激酶(AraB)的一个优选的核苷酸序列选自自由以下组成的组:

[0064] (a) 编码下述多肽的核苷酸序列,所述多肽包含与SEQ ID NO.3的氨基酸序列至少具有20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、98%或99%序列同一性的氨基酸序列;

[0065] (b) 核苷酸序列,其包含与SEQ ID NO.5的核苷酸序列至少具有50%、60%、70%、80%、90%、95%、97%、98%或99%序列同一性的核苷酸序列;

[0066] (c) 核苷酸序列,其互补链与(a)或(b)的核酸分子序列杂交;

[0067] (d) 核苷酸序列,其序列由于遗传密码子的简并性而与(c)的核酸分子的序列不同。

[0068] 编码araB的核苷酸序列可编码原核或真核araB,即具有下述氨基酸序列的araB,所述氨基酸序列与原核生物或真核生物中天然存在的araB的氨基酸序列相关。本发明人发现与araA和araD共表达时,具体araB的下述能力不太取决于该araB是原核或是真核来源,所述能力赋予真核宿主细胞使用阿拉伯糖和/或将阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物如乙醇的能力。更确切地说,这取决于araB的氨基酸序列与SEQ ID NO.3序列的相关性。

[0069] AraD

[0070] 编码在本发明的细胞中表达的L-核酮糖-5-P-4-差向异构酶(araD)的一个优选的核苷酸序列选自由以下组成的组:

[0071] (e) 编码下述多肽的核苷酸序列,所述多肽包含与SEQ ID NO.5的氨基酸序列至少具有60%、65%、70%、75%、80%、85%、90%、95%、97%、98%或99%序列同一性的氨基酸序列;

[0072] (f) 核苷酸序列,其包含与SEQ ID NO.6的核苷酸序列至少具有60%、65%、70%、75%、80%、85%、90%、95%、97%、98%或99%序列同一性的核苷酸序列;

[0073] (g) 核苷酸序列,其互补链与(a)或(b)的核酸分子序列杂交;

[0074] (h) 核苷酸序列,其序列由于遗传密码子的简并性而与(c)的核酸分子的序列不同。

[0075] 编码araD的核苷酸序列可编码原核或真核araD,即具有下述氨基酸序列的araD,所述氨基酸序列与原核生物或真核生物中天然存在的araD的氨基酸序列相同。本发明人发现与araA和araB共表达时,具体araD的下述能力不太取决于该araD是原核或是真核来源,所述能力赋予真核宿主细胞使用阿拉伯糖和/或将阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或转化为期望的发酵产物如乙醇的能力。更确切地说,这取决于araD的氨基酸序列与SEQ ID NO.5序列的相关性。

[0076] 惊人地,密码子偏好指数(codon bias index)表明:比起EP 1 499 708中所述的原核araA、araB和araD基因来说,Lactobacillus plantarum araA、araB和araD基因更适合在酵母中表达。

[0077] 应当注意L.plantarum是通常被认为安全(GRAS)的生物,其被食物注册机构(food registration authorities)认可为安全的。因此,优选的核苷酸序列分别编码具有分别与上述SEQ ID NO:1、3或5序列分别相关的氨基酸序列的araA、araB或araD。优选的核苷酸序列分别编码真菌araA、araB或araD(例如来自Basidiomycete),更优选地分别编码来自厌氧真菌(例如属于Neocallimastix、Caecomyces、Piromyces、Orpinomyces或Ruminomyces科的厌氧真菌)的araA、araB或araD。或者,优选的核苷酸序列分别编码细菌araA、araB或araD,其优选地来自革兰氏阳性细菌,更优选地来自Lactobacillus属,最优选地来自Lactobacillus plantarum种。优选地,araA、araB和araD核苷酸序列中的一条、两条或三条来源于Lactobacillus属,更优选地来源于Lactobacillus plantarum种。在本发明的细胞中表达的细菌araA不是EP 1 499 708中公开的Bacillus subtilis araA(作为SEQ ID NO:9给出)。SEQ ID NO:10表示编码SEQ ID NO:9的核苷酸序列。在本发明细胞中表达的细菌araB和araD不是EP 1 499 708中公开的Escherichia coli(E.coli)的araB和araD(作为SEQ ID NO:11和SEQ ID NO:13给出)。SEQ ID NO:12代表编码SEQ ID NO:11的核苷酸序列。SEQ ID NO:14代表编码SEQ ID NO:13的核苷酸序列。

[0078] 为了提高(细菌)araA、araB和araD酶分别以活性形式在本发明的真核宿主细胞如酵母中表达的可能性,可使相应的编码核苷酸序列适应选定的真核宿主细胞以优化其密码子使用。编码araA、araB和araD酶(或本发明的其它酶,见下文)的核苷酸序列对所选宿主细胞密码子使用的适应性可被表述为密码子适应指数(CAI)。密码子适应指数在本文中定义为基因的密码子使用朝向高度表达的基因的密码子使用的适应性。每种密码子的相对适应

性(w)是对相同氨基酸而言每种密码子的使用与最大量密码子使用的比例。CAI指数被定义为这些相对适应性值的几何平均值。非同义密码子和终止密码子(取决于遗传密码子)排除在外。CAI值范围从0到1,更高的值表示最大量密码子的更高比例(见Sharp and Li,1987, *Nucleic Acids Research* 15:1281-1295;还见Jansen et al.,2003, *Nucleic Acids Res.*31(8):2242-51)。已适应的核苷酸序列优选地具有至少0.2、0.3、0.4、0.5、0.6或0.7的CAI。

[0079] 在一个优选的实施方案中,编码本文之前定义的核苷酸araA、araB和araD的核苷酸序列的表达赋予细胞使用L-阿拉伯糖和/或将其转化为L-核酮糖和/或5-磷酸木酮糖的能力。不期望受任何理论束缚的情况下,预计L-阿拉伯糖首先被转化为L-核酮糖,所述L-核酮糖随后被转化为5-磷酸木酮糖,所述5-磷酸木酮糖是进入戊糖磷酸途径的主要分子。在本发明的上下文中,“使用L-阿拉伯糖”优选地表示在至少20天期间,存在至少0.5%L-阿拉伯糖时,在需氧或厌氧条件下培养的经转化的细胞于660nm处的光密度(OD₆₆₀)被提高了约0.5至1.0或更多。更优选地,OD₆₆₀被提高了0.5到1.5或更多。更优选地,在存在至少1%、至少1.5%或至少2%L-阿拉伯糖时培养细胞。最优选地,在存在约2%L-阿拉伯糖时培养细胞。

[0080] 在本发明的上下文中,在至少20天期间,存在L-阿拉伯糖(与前一段中相同的优选的浓度)的需氧或厌氧条件下培养的细胞中使用合适的测定法检测到可检测量的L-核酮糖时,该细胞能够“将L-阿拉伯糖转化为L-核酮糖”。优选地,该测定法是针对L-核酮糖的HPLC。

[0081] 在本发明的上下文中,在至少20天期间,在存在L-阿拉伯糖(与前一段中相同的优选的浓度)的需氧或厌氧条件下培养的细胞中使用合适的测定法检测到至少2%的5-磷酸木酮糖增加时,细胞“能够将L-阿拉伯糖转化为5-磷酸木酮糖”。优选地,针对5-磷酸木酮糖的基于HPLC的测定法被描述于Zaldivar J., et al.((2002), *Appl.Microbiol.Biotechnol.*,59:436-442)中。该测定法在实验部分简要描述。更优选地,所述增加为至少5%、10%、15%、20%、25%或更多。

[0082] 在另一个优选的实施方案中,在至少一个月到一年期间,在存在L-阿拉伯糖(与前一段中相同的优选的浓度)的需要或厌氧条件下培养时,本文较早定义的编码araA、araB和araD的核苷酸序列的表达赋予细胞将L-阿拉伯糖转化为期望的发酵产物的能力。更优选地,当细胞在前一句中给出的条件下培养时,使用合适的测定法检测到可检测量的期望的发酵产物,则细胞能够将L-阿拉伯糖转化为期望的发酵产物。进一步更优选地,该测定法为HPLC。进一步更优选地,该发酵产物为乙醇。

[0083] 用于用上文所述分别编码araA、araB和araD酶转化的细胞优选地是下述宿主细胞,其能够将木糖主动或被动转运进入细胞中并在细胞中进行木糖异构化。所述细胞优选地能够进行活性糖酵解。所述细胞还可含有内源戊糖磷酸途径,并可含有内源木酮糖激酶活性,从而从木糖异构化而来的木酮糖可被代谢为丙酮酸。细胞还优选地含有将丙酮酸转化为期望的发酵产物的酶,所述期望的发酵产物如乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、β-内酰胺抗生素或头孢菌素。可以通过引入W02007/041269中公开的丁醇途径的一个或多个基因使细胞能够生产丁醇。

[0084] 一种优选的细胞天然能够进行醇发酵,优选地进行厌氧的醇发酵。宿主细胞还优选地具有对乙醇的高耐受、对低pH的高耐受(能够在低于5、4、3或2.5的pH下生长)和对有机酸如乳酸、乙酸或甲酸和糖降解产物如糠醛和羟甲基糠醛的高耐受以及对提高的温度的高耐受。宿主细胞的任何这些特征或活性可天然存在于宿主细胞中,或可通过遗传选择或通过遗传修饰被引入。合适的宿主细胞是真核微生物例如真菌,然而,最合适作为宿主细胞的是酵母或丝状真菌。

[0085] 酵母在本文中定义为下述真核微生物,其包括真菌亚门的所有物种(Alexopoulos,C.J.,1962,In:Introductory Mycology,John Wiley & Sons,Inc.,New York),其主要以单细胞形式生长。酵母可通过单细胞原植体的出芽生长,或可通过生物的分裂生长。作为宿主细胞的优选的酵母属于*Saccharomyces*、*Kluyveromyces*、*Candida*、*Pichia*、*Schizosaccharomyces*、*Hansenula*、*Kloeckera*、*Schwanniomyces*或*Yarrowia*之一的属。优选地,酵母能够进行厌氧发酵,更优选地能够进行厌氧的醇发酵。

[0086] 丝状真菌在本文中定义为包含真菌亚门所有丝状形式的真核微生物。这些真菌的特征是由甲壳质、纤维素和其它复杂多糖组成的营养菌丝体。本发明的丝状真菌从形态上、生理上和遗传上与酵母不同。丝状真菌的植物性生长是通过菌丝伸长,且大部分丝状真菌的碳代谢是专性需氧的。作为宿主细胞的优选的丝状真菌属于*Aspergillus*、*Trichoderma*、*Humicola*、*Acremonium*、*Fusarium*或*Penicillium*之一的属。

[0087] 数年来,已提出引入多种生物用于从农作物糖生产生物乙醇的建议。然而在实践中,所有主要的生物乙醇生产工艺都继续使用*Saccharomyces*属的酵母作为乙醇生产者。这归因于*Saccharomyces*物种用于工业工艺的许多吸引人的特征,即对高酸、乙醇和渗透的耐受,厌氧生长的能力,当然还有其高的醇发酵能力。作为宿主细胞的优选的酵母物种包括*S.cerevisiae*、*S.bulderi*、*S.barnetti*、*S.exiguus*、*S.uvarum*、*S.diastaticus*、*K.lactis*、*K.marxianus*、*K.fragilis*。

[0088] 在一个优选的实施方案中,本发明的宿主细胞是已用核酸构建体转化的宿主细胞,所述核酸构建体包含编码如上文定义的araA、araB和araD酶的核苷酸序列。在一个更优选的实施方案中,用三种核酸构建体共转化宿主细胞,每种核酸构建体包含编码araA、araB或araD的核苷酸序列。包含araA、araB和/或araD编码序列的核酸构建体能够在宿主细胞中表达araA、araB和/或araD酶。为此,可如例如WO 03/0624430中所述构建核酸构建体。宿主细胞可包含单拷贝的每种核酸构建体,但优选地包含多拷贝。核酸构建体可维持为游离状态,并因此包含用于自主复制的序列,例如ARS序列。合适的游离核酸构建体可例如基于酵母2 μ 或pKD1(Fleer et al.,1991,Biotechnology 9:968-975)质粒。然而优选地,每种核酸构建体以一个或多个拷贝被整合进宿主细胞的基因组中。整合进宿主细胞的基因组可通过不符合惯例的重组随机发生,但是优选地核酸构建体通过同源重组被整合进宿主细胞的基因组中,如真菌分子遗传学领域所公知的(见例如WO 90/14423、EP-A-0 481 008、EP-A-0 635 574和US 6,265,186)。因此,在一个更优选的实施方案中,本发明的细胞包含含有araA、araB和/或araD编码序列的核酸构建体,并且能够表达araA、araB和/或araD酶。在一个进一步更优选的实施方案中,araA、araB和/或araD编码序列各自与启动子可操作地连接,所述启动子引起相应的核苷酸序列在细胞中的充分表达,以赋予细胞使用L-阿拉伯糖和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖的能力。优选地,该细胞为酵母细

胞。因此在又一方面中,本发明还包括如前文较早阐述的核酸构建体。优选地,该核酸构建体包含编码araA、araB和/或araD的核酸序列。编码araA、araB或araD的核酸序列均已在本文中较早定义。

[0089] 进一步更优选地,相应核苷酸序列在细胞中的表达赋予细胞将L-阿拉伯糖转化为本文稍后定义的期望的发酵产物的能力。在一个进一步更优选的实施方案中,该发酵产物为乙醇。进一步更优选地,该细胞为酵母细胞。

[0090] 本文使用术语“可操作地连接”是指多核苷酸元件(或编码序列或核酸序列)在功能性相互关系中连接。当核酸序列被置于与另一核酸序列的功能性相互关系中时,其是“可操作地连接”的。例如,如果启动子或增强子能影响编码序列的转录,则其与编码序列是可操作地连接的。可操作地连接的意味着被连接的核酸序列通常是连续的,并且在需要接合两条蛋白质编码区时是连续的和符合读码框的。

[0091] 本文使用术语“启动子”是指下述核酸片段,其能发挥功能以控制一个或多个基因的转录,相对于转录方向而言位于基因转录起点的上游,并且在结构上通过存在DNA-依赖型RNA聚合酶的结合位点、转录起点和任何其它DNA序列鉴定,所述任何其它DNA序列包括但不限于转录因子结合位点、阻抑和激活蛋白结合位点以及本领域技术人员已知直接或间接作用以调节来自该启动子的转录量的任何其它序列。“组成型”启动子是在大部分环境和发育条件下有活性的启动子。“诱导型”启动子是在环境或发育调节下有活性的启动子。

[0092] 能够用于达成编码araA、araB和/或araD的核苷酸序列表达的启动子对要被表达的编码酶的核苷酸序列可以不是天然的,即启动子对与其可操作地连接的核苷酸序列(编码序列)是异源的。尽管启动子优选地对与其可操作地连接的编码序列是异源的,但是也启动子是同源的也是优选,即对宿主细胞而言是内源的。优选地,与对编码序列是固有的启动子相比,优选地在可获得阿拉伯糖,或阿拉伯糖和葡萄糖,或木糖和阿拉伯糖,或木糖和阿拉伯糖和葡萄糖作为碳源,更优选地作为主要碳源(即多于50%的可获得的碳源由阿拉伯糖,或阿拉伯糖和葡萄糖,或木糖和阿拉伯糖,或木糖和阿拉伯糖和葡萄糖组成),最优选地作为唯一碳源条件下,(对核苷酸序列)异源的启动子能够生产高稳态水平的包含编码序列的转录本(或每单位时间能够生产更多转录本分子,即mRNA分子)。该上下文中合适的启动子包括组成型和诱导型天然启动子以及经改造的启动子。用于本发明中的一个优选的启动子除了会对分解代谢产物(葡萄糖)阻抑不灵敏外,和/或会优选地不需要阿拉伯糖和/或木糖用于诱导。

[0093] 具有这些特征的启动子是可广泛获得的,并是技术人员已知的。这类启动子的合适的例子包括例如来自糖酵解基因的启动子,如来自酵母或丝状真菌的果糖磷酸激酶(PPK)、磷酸丙糖异构酶(TPI)、甘油醛-3-磷酸脱氢酶(GPD、TDH3或GAPDH)、丙酮酸激酶(PYK)、磷酸甘油酸酯激酶(PGK)启动子;关于来自酵母的这类启动子的更多细节可见(W093/03159)。其它有用的启动子是核糖体蛋白质编码基因启动子、乳糖酶基因启动子(LAC4)、醇脱氢酶启动子(ADH1、ADH4等等)、烯醇化酶启动子(ENO)、葡萄糖-6-磷酸异构酶启动子(PGI1, Hauf et al, 2000)或己糖(葡萄糖)转运子启动子(HXT7)或甘油醛-3-磷酸脱氢酶(TDH3)。PGI1启动子的序列在SEQ ID NO:51中给出。HXT7启动子的序列在SEQ ID NO:52中给出。TDH3启动子的序列在SEQ ID NO:49中给出。其它组成型和诱导型启动子与增强子或上游激活序列会是本领域技术人员已知的。需要时可修饰本发明的宿主细胞中使用的

启动子,以影响其控制特征。

[0094] 本发明的一种优选的细胞是用*L.plantarum*的araA、araB和araD基因转化的真核细胞。更优选地,该真核细胞是酵母细胞,进一步更优选地是用*L.plantarum*的araA、araB和araD基因转化的*S.cerevisiae*菌株。最优选地,该细胞为CBS 120327或CBS 120328,其均在2006年9月27日保藏于CBS Institute(荷兰)。

[0095] 用于表明给定的(重组)核酸或多肽分子和给定的宿主生物或宿主细胞之间相互关系时,术语“同源的”被理解为表示该核酸或多肽分子天然地由相同物种的宿主细胞或生物产生,优选地由相同变种或菌株产生。典型地,如果与宿主细胞同源,则编码多肽的核酸序列会与并非其天然环境中的另一启动子序列或适用时另一分泌信号序列和/或终止子序列可操作地连接。当用于表明两条核酸序列的相关性时,术语“同源的”表示一条单链核酸序列可与互补的单链核酸序列杂交。杂交程度可取决于大量因素,包括如前文所述的序列之间的同一性量和杂交条件如温度和盐浓度。优选地,同一性区域大于约5bp,更优选地同一性区域大于约10bp。

[0096] 关于核酸(DNA或RNA)或蛋白质使用时,术语“异源的”表示下述核酸或蛋白质,其在存在的生物、细胞、基因组或DNA或RNA序列中不是天然存在的部分,或在基因组或DNA或RNA序列中存在的位置与其天然存在的位置不同。异源的核酸或蛋白质对其被引入的细胞不是内源的,而是得自合成或重组产生的另一细胞。一般地,尽管不是必须地,这类核酸编码下述蛋白质,其中引入或表达DNA的细胞通常不生产所述蛋白质。类似地,外源RNA编码存在该外源RNA的细胞中通常不表达的蛋白质。异源核酸和蛋白质也可表示外来核酸或蛋白质。术语异源核酸或蛋白质包括下述任何核酸或蛋白质,其会被本领域技术人员识别为对表达该核酸或蛋白质的细胞而言是异源或外来的。术语异源的还适用于核酸或氨基酸序列的非天然组合,即组合中至少两条组合的序列对彼此是外来的。

[0097] 能够使用和/或转化L-阿拉伯糖和木糖的优选的真核细胞

[0098] 在一个更优选的实施方案中,表达araA、araB和araD的本发明的细胞能够使用L-阿拉伯糖和/或将其转化为L-核酮糖和/或5-磷酸木酮糖和/或本文较早定义的期望的发酵产物,并且还显示出使用木糖和/或将木糖转化为木酮糖的能力。木糖到木酮糖的转化优选为一步异构化步骤(直接将木糖异构化为木酮糖)。因此这类细胞既能使用L-阿拉伯糖又能使用木糖。“使用”木糖优选地具有与本文较早定义的“使用”L-阿拉伯糖相同的含义。

[0099] 酶定义按照W0 06/009434中针对木糖异构酶(EC 5.3.1.5)、木酮糖激酶(EC 2.7.1.17)、5-磷酸核酮糖差向异构酶(5.1.3.1)、5-磷酸核酮糖异构酶(EC 5.3.1.6)、转酮酶(EC 2.2.1.1)、转醛酶(EC 2.2.1.2)和醛糖还原酶(EC1.1.1.21)使用的定义。

[0100] 在一个优选的实施方案中,如前文较早定义的表达araA、araB和araD的本发明的真核细胞具有将木糖异构化为木酮糖的能力,如例如W003/0624430或W0 06/009434中所述。通过用包含编码木糖异构酶的核苷酸序列的核酸构建体转化宿主细胞,赋予该宿主细胞将木糖异构化为木酮糖的能力。被转化的宿主细胞将木糖异构化为木酮糖的能力是木糖到木酮糖的直接异构化。这被理解为表示木糖在由木糖异构酶催化的单一反应中被异构化为木酮糖,与从木糖通过木糖醇中间产物到木酮糖的两步骤转化相反,所述两步骤转化分别由木糖还原酶和木糖醇脱氢酶催化。

[0101] 核苷酸序列编码木糖异构酶,其优选地在本发明的经转化的宿主细胞中以活性形

式表达。因此,核苷酸序列在宿主细胞中的表达产生木糖异构酶,该酶在30℃下至少具有每mg蛋白质10U木糖异构酶活性的比活性,优选地在30℃下至少具有每mg 20U、25U、30U、50U、100U、200U、300或500U的比活性。在经转化的宿主细胞中表达的木糖异构酶的比活性在本文中被定义为:每mg宿主细胞细胞游离裂解物(例如酵母细胞游离裂解物)的蛋白质的木糖异构酶活性单位的量。木糖异构酶活性的测定在前文中已描述。

[0102] 优选地,编码木糖异构酶的核苷酸序列在宿主细胞中的表达产生对木糖具有下述 K_m 的木糖异构酶,所述 K_m 少于50、40、30或25mM,更优选地,对木糖的 K_m 为约20mM或更少。

[0103] 编码木糖异构酶的一条优选的核苷酸序列可选自由以下组成的组:

[0104] (e) 编码下述多肽的核苷酸序列,所述多肽包含与SEQ ID NO.15的氨基酸序列至少具有60%、65%、70%、75%、80%、85%、90%、95%、97%、98%或99%序列同一性的氨基酸序列;

[0105] (f) 核苷酸序列,其包含与SEQ ID NO.16的核苷酸序列至少具有40%、50%、60%、70%、80%、90%、95%、97%、98%或99%序列同一性的核苷酸序列;

[0106] (g) 核苷酸序列,其互补链与(a)或(b)的核酸分子序列杂交;

[0107] (h) 核苷酸序列,其序列由于遗传密码子的简并性而与(c)的核酸分子的序列不同。

[0108] 编码木糖异构酶的核苷酸序列可编码原核或真核木糖异构酶,即具有下述氨基酸序列的木糖异构酶,所述氨基酸序列与原核生物或真核生物中天然存在的木糖异构酶的氨基酸序列相同。本发明人发现具体木糖异构酶的下述能力不太取决于该异构酶是原核或是真核来源,所述能力赋予真核宿主细胞将木糖异构化为木酮糖的能力。更确切地说,这取决于异构酶的氨基酸序列与*Piromyces*序列(SEQ ID NO.7)的相关性。惊人地,与其它已知的真核异构酶相比,真核*Piromyces*异构酶与原核异构酶更加相关。因此,优选的核苷酸序列编码具有下述氨基酸序列的木糖异构酶,所述氨基酸序列与上文定义的*Piromyces*序列相关。一种优选的核苷酸序列编码真菌木糖异构酶(例如来自Basidiomycete),更优选地编码来自另一厌氧真菌的木糖异构酶,例如来自属于*Neocallimastix*、*Caecomyces*、*Piromyces*、*Orpinomyces*或*Ruminomyces*科的厌氧真菌的木糖异构酶。或者,一种优选的核苷酸序列编码细菌木糖异构酶,优选地来自革兰氏阴性细菌,更优选地编码来自Bacteroides纲、或来自Bacteroides属的异构酶,最优选地来自*B. thetaiotaomicron* (SEQ ID NO.15)。

[0109] 为了提高木糖异构酶以活性形式在真核宿主细胞如酵母中表达的可能性,可使编码木糖异构酶的核苷酸序列适应前文定义的真核宿主细胞,以优化其密码子使用。

[0110] 用于用编码上文定义的木糖异构酶的核苷酸序列转化的宿主细胞优选地为能够将木糖主动或被动转运进入细胞的宿主细胞。该宿主细胞优选地含有活性糖酵解。该细胞还可含有内源戊糖磷酸途径,并可含有内源木酮糖激酶活性,从而从木糖异构化而来的木酮糖可被代谢为丙酮酸。该宿主还优选地含有将丙酮酸转化为期望的发酵产物的酶,所述期望的发酵产物如乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、 β -内酰胺抗生素或头孢菌素。一种优选的宿主细胞天然能够进行醇发酵,优选地进行厌氧的醇发酵。宿主细胞还优选地具有对乙醇的高耐受、对低pH的高耐受(即能够在低于5、4、3或2.5的pH下生长)和对有机酸如乳酸、乙酸或甲酸和糖降解产物如糠醛和羟甲基糠醛的高耐受和对提高的温度的高耐受。宿主细胞的

任何这些特征或活性可天然存在于宿主细胞中,或可通过遗传修饰被引入或修饰。合适的宿主细胞是真核微生物例如真菌,然而,最合适作为宿主细胞的是酵母或丝状真菌。优选的酵母和丝状真菌本文已定义。

[0111] 本文使用的措辞宿主细胞具有与细胞相同的含义。

[0112] 本发明的细胞优选地用下述核酸构建体转化,所述核酸构建体包含编码木糖异构酶的核苷酸序列。优选地使用的核酸构建体与使用的包含编码araA、araB或araD的核苷酸序列的核酸构建体相同。

[0113] 在本发明的另一个优选的实施方案中,本发明的细胞(其表达araA、araB和araD,并显示将木糖直接异构化为木酮糖的能力,如前文所定义)还包含下述遗传修饰,其提高戊糖磷酸途径的通量,如WO 06/009434中所述。具体地,该遗传修饰引起非氧化部分戊糖磷酸途径通量的提高。引起戊糖磷酸途径非氧化性部分通量提高的遗传修饰在本文中应被理解为表示下述修饰,与除了引起通量提高的遗传修饰外是遗传同一的菌株中的通量相比,所述修饰至少将通量提高1.1、1.2、1.5、2、5、10或20倍。可如下测量戊糖磷酸途径非氧化部分的通量:使经修饰的宿主在作为唯一碳源的木糖上生长,测定木糖比消耗速率(specific xylose consumption rate)开从木糖比消耗速率中减去木糖醇比生产速率,如果产生任何木糖醇的话。然而,戊糖磷酸途径的非氧化部分通量与在作为唯一碳源的木糖上的生长速率成比例,优选地与在作为唯一碳源的木糖上的厌氧生长速率成比例。在作为唯一碳源的木糖上的生长速率($\mu_{\max}$)与戊糖磷酸途径非氧化部分的通量之间存在线性关系。木糖比消耗速率(Q_s)等于生长速率(μ)除以生物质相对于糖的产率(Y_{xs}),因为生物质相对于糖的产率是恒定的(在给定的一组条件下:厌氧、生长培养基、pH、菌株的遗传背景等;即 $Q_s = \mu / Y_{xs}$)。因此,可从这些条件下的最大生长速率的提高推论出戊糖磷酸途径非氧化部分提高的通量。在一个优选的实施方案中,细胞包含下述遗传修饰,其提高戊糖磷酸途径的通量,并具有至少346mg木糖/g生物质/h的木糖比消耗速率。

[0114] 提高戊糖磷酸途径通量的遗传修饰可以多种途径被引入宿主细胞中。这些途径包括例如达到木酮糖激酶和/或戊糖磷酸途径非氧化部分的一种或多种酶的更高稳定状态活性水平,和/或降低的稳态水平的非特异性醛糖还原酶活性。可通过选择突变体(自发的或由化学品或辐射诱导)和/或通过重组DNA技术(例如分别通过过表达或失活基因,所述基因编码调节这些基因的酶或因素)影响稳态活性水平的这些改变。

[0115] 在一个更优选的宿主细胞中,遗传修饰包括过表达戊糖磷酸途径(非氧化部分)的至少一种酶。优选地,该酶选自由编码5-磷酸核酮糖异构酶、5-磷酸核酮糖差向异构酶、转酮酶和转醛酶的酶组成的组,如WO06/009434中所述。

[0116] 可过表达戊糖磷酸途径(非氧化部分)的酶的多种组合。例如,被过表达的酶可以至少是酶5-磷酸核酮糖异构酶和5-磷酸核酮糖差向异构酶;或至少是酶5-磷酸核酮糖异构酶和转酮酶;或至少是酶5-磷酸核酮糖异构酶和转醛酶;或至少是酶5-磷酸核酮糖差向异构酶和转酮酶;或至少是酶5-磷酸核酮糖差向异构酶和转醛酶;或至少是酶转酮酶和转醛酶;或至少是酶5-磷酸核酮糖差向异构酶、转酮酶和转醛酶;或至少是酶5-磷酸核酮糖异构酶、转酮酶和转醛酶;或至少是酶5-磷酸核酮糖异构酶、5-磷酸核酮糖差向异构酶和转醛酶;或至少是酶5-磷酸核酮糖异构酶、5-磷酸核酮糖差向异构酶和转酮酶。在本发明的一个实施方案中,在宿主细胞中过表达5-磷酸核酮糖异构酶、5-磷酸核酮糖差向异构酶、转酮酶

和转醛酶各种酶。更优选的是下述宿主细胞,其中遗传修饰包括至少过表达转酮酶和转醛酶两种酶,因为这样的宿主细胞已经能够在木糖上厌氧生长。事实上,在一些条件下,我们发现仅过表达转酮酶和转醛酶的宿主细胞在木糖上已经具有与过表达所有四种酶的宿主细胞相同的厌氧生长速率,所述四种酶即5-磷酸核酮糖异构酶、5-磷酸核酮糖差向异构酶、转酮酶和转醛酶。另外,过表达5-磷酸核酮糖异构酶和5-磷酸核酮糖差向异构酶两种酶的宿主细胞是超过仅过表达异构酶或仅过表达差向异构酶的宿主细胞而被优选的,因为仅过表达这些酶之一即可产生代谢失衡。

[0117] 本领域可获得多种手段用于在本发明的细胞中过表达酶。具体地,可通过提高宿主细胞中编码酶的基因的拷贝数来过表达酶,例如通过在宿主细胞的基因组中整合额外的基因拷贝,通过从附加体多拷贝表达载体中表达基因,或通过引入包含多拷贝基因的附加体表达载体来实现。

[0118] 或者,可通过使用下述启动子达成酶在本发明宿主细胞中的过表达,所述启动子不是编码要过表达的酶的序列所固有的,即对与其可操作地连接的编码序列是异源的启动子。对该目的而言合适的启动子在本文中已定义。

[0119] 用于过表达酶的编码序列优选地与本发明的宿主细胞是同源的。然而,与本发明宿主细胞异源的编码序列也可同样使用,如WO 06/009434中所述。

[0120] 用于在本发明的宿主细胞中过表达5-磷酸核酮糖异构酶的核苷酸序列是编码具有5-磷酸核酮糖异构酶活性的多肽的核苷酸序列,其中优选地该多肽具有与SEQ ID NO.17至少具有50%、60%、70%、80%、90%或95%同一性的氨基酸序列,或其中该核苷酸序列能够在中度条件下、优选地在严格条件下与SEQ ID NO.18的核苷酸序列杂交。

[0121] 用于在本发明的宿主细胞中过表达5-磷酸核酮糖差向异构酶的核苷酸序列是编码具有5-磷酸核酮糖差向异构酶活性的多肽的核苷酸序列,其中优选地该多肽具有与SEQ ID NO.19至少具有50%、60%、70%、80%、90%或95%同一性的氨基酸序列,或其中该核苷酸序列能够在中度条件下、优选地在严格条件下与SEQ ID NO.20的核苷酸序列杂交。

[0122] 用于在本发明的宿主细胞中过表达转酮酶的核苷酸序列是编码具有转酮酶活性的多肽的核苷酸序列,其中优选地该多肽具有与SEQ ID NO.21至少具有50%、60%、70%、80%、90%或95%同一性的氨基酸序列,或其中该核苷酸序列能够在中度条件下、优选地在严格条件下与SEQ ID NO.22的核苷酸序列杂交。

[0123] 用于在本发明的宿主细胞中过表达转醛酶的核苷酸序列是编码具有转醛酶活性的多肽的核苷酸序列,其中优选地该多肽具有与SEQ ID NO.23至少具有50%、60%、70%、80%、90%或95%同一性的氨基酸序列,或其中该核苷酸序列能够在中度条件下、优选地在严格条件下与SEQ ID NO.24的核苷酸序列杂交。

[0124] 当涉及在经遗传修饰的宿主细胞中生产酶时,酶的过表达表示:与相同条件下未经修饰的宿主细胞相比,酶以更高水平的酶比活性被生产。通常,这表示与相同条件下未经修饰的宿主细胞相比,具有酶活性的蛋白质(或在多亚基酶的情况下为多个蛋白质)以更大的量被生产,或者说以更高的稳态水平被生产。类似地,这通常表示与相同条件下未经修饰的宿主细胞相比,编码具有酶活性的蛋白质的mRNA以更大量被生产,或者说以更高的稳态水平被生产。因此,优选地通过使用如本文所述的适当酶测定法,通过测量宿主细胞中酶的比活性水平测定酶的过表达。或者,可例如使用对酶特异的抗体通过定量酶蛋白质的比稳

态水平,或通过定量编码该酶的mRNA的比稳态水平,间接测定酶的过表达。后者尤其适用于戊糖磷酸途径的酶,对该途径而言酶测定法不是容易可行的,因为酶的底物不是可商业获得的。优选地,在本发明的宿主细胞中,与除了引起过表达的遗传修饰外遗传上同一的菌株相比,要被过表达的酶以至少1.1、1.2、1.5、2.5、10或20的倍数被过表达。应当理解,这些过表达水平可适用于酶活性的稳态水平、酶的蛋白质的稳态水平,以及编码酶的转录本的稳态水平。

[0125] 在又一个优选的实施方案中,本发明的宿主细胞(其表达araA、araB和araD,并显示将木糖直接异构化为木酮糖的能力,并可选地包含下述遗传修饰,其提高戊糖磷酸途径的通量,如前文所述)还包含提高木酮糖激酶比活性的遗传修饰。优选地,该遗传修饰例如通过过表达编码木酮糖激酶的核苷酸序列,来引起木酮糖激酶的过表达。编码木酮糖激酶的基因对宿主细胞可以是内源的,或可以是对宿主细胞异源的木酮糖激酶。用于在本发明的宿主细胞中过表达木酮糖激酶的核苷酸序列是编码具有木酮糖激酶活性的多肽的核苷酸序列,其中优选地该多肽具有与SEQ ID NO.25至少具有50%、60%、70%、80%、90%或95%同一性的氨基酸序列,或其中该核苷酸序列能够在中度条件下、优选地在严格条件下与SEQ ID NO.26的核苷酸序列杂交。

[0126] 一种尤其优选的木酮糖激酶是与来自*Piromyces*的木酮糖激酶xylB相关的木糖激酶,如WO 03/0624430中所述。用于在本发明的宿主细胞中过表达木酮糖激酶的一种更优选的核苷酸序列是编码具有木酮糖激酶活性的多肽的核苷酸序列,其中优选地该多肽具有与SEQ ID NO.27至少具有50%、60%、70%、80%、90%或95%同一性的氨基酸序列,或其中该核苷酸序列能够在中度条件下、优选地在严格条件下与SEQ ID NO.28的核苷酸序列杂交。

[0127] 在本发明的宿主细胞中,提高木酮糖激酶比活性的遗传修饰可以与如上所述提高戊糖磷酸途径通量的任何修饰组合,但是该组合对本发明不是必需的。因此,除了表达本文定义的araA、araB和araD酶之外还包含提高木酮糖激酶比活性的遗传修饰的本发明宿主细胞特定地包含在本发明中。本领域中用于达成和分析木酮糖激酶在本发明宿主细胞中过表达的多种手段与上文针对戊糖磷酸途径的酶所述相同。优选地,在本发明的宿主细胞中,与除了引起过表达的遗传修饰外遗传上同一的菌株相比,要被过表达的木酮糖激酶以至少1.1、1.2、1.5、2.5、10或20的倍数被过表达。应当理解,这些过表达水平可应用于酶活性的稳态水平、酶的蛋白质的稳态水平,以及编码酶的转录本的稳态水平。

[0128] 在又一个优选的实施方案中,本发明的宿主细胞(其表达araA、araB和araD,并显示将木糖直接异构化为木酮糖的能力,并可选地包含提高戊糖磷酸途径通量的遗传修饰,和/或还包含提高木酮糖激酶比活性的遗传修饰,均如前所述)还包含降低宿主细胞中非特异性醛糖还原酶活性的遗传修饰。优选地,通过一种或多种遗传修饰降低宿主细胞中非特异性醛糖还原酶活性,所述遗传修饰降低非特异性醛糖还原酶的表达或使编码该酶的基因失活,如WO 06/009434中所述。优选地,遗传修饰降低或失活宿主细胞中非特异性醛糖还原酶各内源拷贝的表达。宿主细胞可由于二倍性、多倍性或非整倍性而包含多拷贝的编码非特异性醛糖还原酶的基因,和/或宿主细胞可含有具有醛糖还原酶活性的若干种不同的(同工)酶,其氨基酸序列不同并且各自由不同的基因编码。另外,在这类情况下编码非特异性醛糖还原酶的各基因的表达优选地被降低或失活。优选地,通过至少删除基因的部分或通过破坏基因将该基因失活或删除,其中在该语境中术语基因还包括编码序列上游或下游的

任何非编码序列,其(部分)删除或失活导致宿主细胞中非特异性醛糖还原酶活性表达的降低。编码要在本发明的宿主细胞中降低其活性的醛糖还原酶的核苷酸序列是编码具有醛糖还原酶活性的多肽的核苷酸序列,其中优选地该多肽具有与SEQ ID NO.29至少具有50%、60%、70%、80%、90%或95%同一性的氨基酸序列,或其中该核苷酸序列能够在中度条件下、优选地在严格条件下与SEQ ID NO.30的核苷酸序列杂交。

[0129] 在本发明的宿主细胞中,本文定义的araA、araB和araD酶的表达与降低非特异性醛糖还原酶活性的遗传修饰组合。导致非特异性醛糖还原酶活性降低的遗传修饰可以与如上所述在宿主细胞中提高戊糖磷酸途径通量的任何修饰和/或提高木酮糖激酶比活性的任何修饰组合,但是这些组合对本发明不是必需的。因此,包含降低非特异性醛糖还原酶活性的其它遗传修饰的、表达araA、araB和araD的宿主细胞特定地包含在本发明中。

[0130] 在一个优选的实施方案中,宿主细胞是于2006年9月27日保藏于CBS协会(荷兰)的CBS 120327。

[0131] 在又一个优选的实施方案中,本发明涉及还适应L-阿拉伯糖利用(使用L-阿拉伯糖和/或将其转化为L-核酮糖和/或5-磷酸木酮糖和/或期望的发酵产物)并可选地适应木糖利用的经修饰的宿主细胞,所述宿主细胞通过选择自发或(例如通过辐射或化学品)诱导的突变体获得,所述突变体在L-阿拉伯糖和可选的木糖上生长,优选地在作为唯一碳源的L-阿拉伯糖和可选的木糖上生长,更优选地在厌氧条件下生长。突变体的选择可以通过例如Kuyper et al. (2004, FEMS Yeast Res. 4:655-664)所述的培养物连续传代进行,和/或通过例如WO 06/009434实施例4中所述在恒化培养中于选择压力下培养进行。该选择过程可根据需要继续。该选择过程优选地在一周内直至一年内完成。然而如果需要的话,该选择过程可进行更长的时间段。在该选择过程中,优选地在存在约20g/l L-阿拉伯糖和/或约20g/l木糖时培养细胞。预期在该选择过程结束时获得的细胞关于其下述能力被改进,所述能力为使用L-阿拉伯糖和/或木糖的能力,和/或将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或期望的发酵产物如乙醇的能力。在该语境中,“改进的细胞”可表示获得的细胞与其来源的细胞相比,能够以更有效的方式使用L-阿拉伯糖和/或木糖。例如,预期在相同的条件下,获得的细胞比其来源的细胞更好地生长:比生长速率提高至少2%。优选地,该提高至少为4%、6%、8%、10%、15%、20%、25%或更多。如技术人员所已知的,比生长速率可以从OD₆₆₀计算。因此,通过监测OD₆₆₀,可以推论比生长速率。在该语境中,“改进的细胞”也可表示获得的细胞与其来源的细胞相比,以更有效的方式将L-阿拉伯糖转化为L-核酮糖和/或5-磷酸木酮糖和/或期望的发酵产物如乙醇。例如,预期获得的细胞生产更高量的L-核酮糖和/或5-磷酸木酮糖和/或期望的发酵产物如乙醇:相同条件下与其来源的细胞相比,至少一种这些化合物至少提高2%。优选地,该提高至少为4%、6%、8%、10%、15%、20%、25%或更多。在该语境中,“改进的细胞”也可表示获得的细胞与其来源的细胞相比,以更有效的方式将木糖转化为木酮糖和/或期望的发酵产物如乙醇。例如,预期获得的细胞生产更高量的木酮糖和/或期望的发酵产物如乙醇:相同条件下与其来源的细胞相比至少一种这些化合物至少提高2%。优选地,该提高至少为4%、6%、8%、10%、15%、20%、25%或更多。

[0132] 在本发明的一个优选的宿主细胞中,至少一种上述遗传修饰(包括通过选择突变体获得的修饰)赋予宿主细胞在作为碳源、优选地作为唯一碳源的L-阿拉伯糖和可选的木

糖上、优选地在厌氧条件下生长的能力优选地,经修饰的宿主细胞基本上不生产木糖醇,例如产生的木糖醇低于检测极限,或例如以摩尔为基础少于消耗的碳的5%、2%、1%、0.5%或0.3%。

[0133] 优选地,经修饰的宿主细胞具有在作为唯一碳源的L-阿拉伯糖和可选的木糖上在需氧条件下以至少0.001、0.005、0.01、0.03、0.05、0.1、0.2、0.25或0.3h⁻¹的速率生长的能力,或者如果可应用的话,在厌氧条件下以至少0.001、0.005、0.01、0.03、0.05、0.07、0.08、0.09、0.1、0.12、0.15或0.2h⁻¹的速率生长的能力。优选地,经修饰的宿主细胞具有在作为唯一碳源的葡萄糖和L-阿拉伯糖和可选的木糖(1:1重量比)的混合物上在需氧条件下以至少0.001、0.005、0.01、0.03、0.05、0.1、0.2、0.25或0.3h⁻¹的速率生长的能力,或者如果可应用的话,在厌氧条件下以至少0.001、0.005、0.01、0.03、0.05、0.1、0.12、0.15或0.2h⁻¹的速率生长的能力。

[0134] 优选地,经修饰的宿主细胞具有至少346、350、400、500、600、650、700、750、800、900或1000mg/g细胞/h的L-阿拉伯糖和可选的木糖的比消耗速率。优选地,经修饰的宿主细胞具有下述相对于L-阿拉伯糖和可选的木糖的发酵产物(如乙醇)产率,所述产率至少是宿主细胞的相对于葡萄糖的发酵产物(如乙醇)产率的20%、25%、30%、35%、40%、45%、50%、55%、60%、70%、80%、85%、90%、95%或98%。更优选地,经修饰的宿主细胞的相对于L-阿拉伯糖和可选的木糖的发酵产物(如乙醇)产率等于宿主细胞的相对于葡萄糖的发酵产物(如乙醇)产率。同样,经修饰的宿主细胞的相对于L-阿拉伯糖和可选的木糖的生物质产率优选地至少是宿主细胞的相对于葡萄糖的生物质产率的55%、60%、70%、80%、85%、90%、95%或98%。更优选地,经修饰的宿主细胞的相对于L-阿拉伯糖和可选的木糖的生物质产率等于宿主细胞的相对于葡萄糖的生物质产率。应当理解在相对于葡萄糖和L-阿拉伯糖和可选的木糖的产率的比较中,两种产率均在需氧条件下比较或均在厌氧条件下比较。

[0135] 在一个更优选的实施方案中,宿主细胞是于2006年9月27日保藏于CBS协会(荷兰)的CBS 120328,或于2007年9月20日保藏于CBS协会(荷兰)的CBS 121879。

[0136] 在一个优选的实施方案中,细胞表达一种或多种酶,所述酶赋予细胞生产至少一种选自下组的发酵产物的能力,该组由乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、β-内酰胺抗生素和头孢菌素组成。在一个更优选的实施方案中,本发明的宿主细胞是用于生产乙醇的宿主细胞。在另一个优选的实施方案中,本发明涉及用于生产除乙醇外的发酵产物的经转化的宿主细胞。这样的非乙醇发酵产物原则上包括可由真核微生物如酵母或丝状真菌生产的任何粗制化学品(bulk chemical)或精细化学品。这类发酵产物包括例如乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、β-内酰胺抗生素和头孢菌素。用于生产非乙醇发酵产物的一个优选的本发明宿主细胞是含有下述遗传修饰的宿主细胞,所述修饰导致降低的醇脱氢酶活性。

[0137] 方法

[0138] 另一方面,本发明涉及下述发酵工艺,其中本发明的宿主细胞被用于发酵碳源,所述碳源包含L-阿拉伯糖来源和任选的木糖来源。优选地,该L-阿拉伯糖来源和木糖来源为L-阿拉伯糖和木糖。另外,发酵培养基中的碳源也可以包含葡萄糖来源。L-阿拉伯糖、木糖

或葡萄糖来源可以同样是L-阿拉伯糖、木糖或葡萄糖,或可以是包含L-阿拉伯糖、木糖或葡萄糖单元的任何碳水化合物寡聚体或多聚体,例如木质纤维素、木聚糖、纤维素、淀粉、阿拉伯糖聚糖等等。为了从这类碳水化合物中释放木糖或葡萄糖。可向发酵培养基中添加或由经修饰的宿主细胞生产合适的碳水化合物酶(如木聚糖酶、葡聚糖酶、淀粉酶等等)。在后一情况下,经修饰的宿主细胞可被遗传工程改造为生产和分泌这类碳水化合物酶。使用寡聚体或多聚体的葡萄糖来源的另一个优点是其使得在发酵期间能够维持(更)低的游离葡萄糖浓度,所述发酵例如使用速率限制量的碳水化合物酶。因此,这进而会防止代谢和转运非葡萄糖的糖如木糖所需的体系阻抑。在一种优选的工艺中,经修饰的宿主细胞发酵L-阿拉伯糖(可选地为木糖)和葡萄糖二者,优选地同时发酵,在该情况下优选地使用经修饰的宿主细胞,所述宿主细胞对妨碍二次生长(diauxic growth)的葡萄糖阻抑不敏感。除了作为碳源的L-阿拉伯糖、任选的木糖(和葡萄糖)的来源外,发酵培养基还会包含经修饰的宿主细胞生长所需的适当成分。用于微生物如酵母和丝状真菌生长的发酵培养基的组成是本领域公知的。

[0139] 在一种优选的工艺中,提供了用于生产选自下组的发酵产物的工艺,该组由乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、 β -内酰胺抗生素和头孢菌素组成,其中该工艺包括步骤:

[0140] (a) 用本文定义的经修饰的宿主细胞发酵含有阿拉伯糖和可选的木糖来源的培养基;和可选地

[0141] (b) 回收所述发酵产物。

[0142] 发酵工艺是用于生产发酵产物如乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、 β -内酰胺抗生素(如青霉素G和青霉素V及其发酵衍生物)和头孢菌素的工艺。该发酵工艺可以是需氧或厌氧的发酵工艺。厌氧发酵工艺在本文中定义为不存在氧时进行的发酵工艺,或其中基本不消耗氧、优选地消耗少于5、2.5或1mmol/L/h、更优选地消耗0mmol/L/h(即氧消耗是不可检测的)的工艺,并且其中有机分子发挥电子供体和电子受体两种作用。不存在氧时,在糖酵解和生物物质形成中产生的NADH不能被氧化磷酸化氧化。为了解决该问题,许多微生物使用丙酮酸或其衍生物之一作为电子和氢受体,从而再生NAD⁺。因此,在一种优选的厌氧发酵工艺中,丙酮酸被用作电子(和氢受体),并被还原为发酵产物如乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、 β -内酰胺抗生素和头孢菌素。在一个优选的实施方案中,发酵工艺是厌氧的。厌氧的发酵工艺是有利的,因为其比需氧工艺更便宜:需要更少的专用设备。另外,预期厌氧过程比需氧过程给出更高的产物产率。在需氧条件下,通常生物物质产率比厌氧条件下更高。因此,通常在需氧条件下,预期的产物产率比厌氧条件下更低。根据本发明人,本发明的工艺是目前为止开发出的首个使用包含L-阿拉伯糖来源的培养基的厌氧发酵工艺。

[0143] 在另一个优选的实施方案中,发酵工艺在处于氧限制的条件下进行。更优选地,发酵工艺是需氧的并处于氧限制的条件下。氧限制的发酵工艺是下述工艺,其中氧消耗受从气体转移至液体的氧的限制。通过进入气流的量和组成以及使用的发酵设备的实际混合/质量转移特性测定氧限制的程度。优选地,在处于氧限制条件下的工艺中,氧消耗的速率至少为5.5mmol/L/h,更优选地至少为6mmol/L/h,进一步更优选地至少为7mmol/L/h。

[0144] 发酵工艺优选地在对经修饰的细胞最适的温度下进行。因此,对大部分酵母或真菌细胞而言,在少于42℃、优选地少于38℃的温度下进行发酵工艺。对酵母或丝状真菌宿主细胞而言,发酵工艺优选地在低于35、33、30或28℃的温度下且高于20、22或25℃的温度下进行。

[0145] 一种优选的工艺是用于生产乙醇的工艺,其中该工艺包括步骤:(a)用本文定义的经修饰的宿主细胞发酵含有阿拉伯糖和可选的木糖来源的培养基;和可选地(b)回收乙醇。发酵培养基也可包含也被发酵为乙醇的葡萄糖来源。在一个优选的实施方案中,用于生产乙醇的发酵工艺是厌氧的。厌氧的在前文中已被定义。在另一个优选的实施方案中,用于生产乙醇的发酵工艺是需氧的。在另一个优选的实施方案中,用于生产乙醇的发酵工艺处于氧限制的条件下,更优选地处于需氧和氧限制的条件下。氧限制的条件在前文中已定义。

[0146] 在该工艺中,乙醇的体积生产力(productivity)优选地为每小时每升至少0.5、1.0、1.5、2.0、2.5、3.0、5.0或10.0g乙醇。工艺中相对于L-阿拉伯糖和可选的木糖和/或葡萄糖的乙醇产率至少为20%、25%、30%、35%、40%、45%、50%、60%、70%、80%、90%、95%或98%。乙醇产率在本文中被定义为理论最大产率的百分比,所述理论最大产率对葡萄糖和L-阿拉伯糖和可选的木糖而言为每g葡萄糖或木糖0.51g乙醇。在另一个优选的实施方案中,本发明涉及用于生产选自下组的发酵产物的工艺,该组由乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、β-内酰胺抗生素和头孢菌素组成。该工艺优选地包括步骤:(a)用上文定义的经修饰的宿主细胞发酵含有L-阿拉伯糖和可选的木糖来源的培养基;和可选地(b)回收所述发酵产物。在一种优选的工艺中,培养基还含有葡萄糖来源。

[0147] 在导致乙醇生产的本发明的发酵工艺中,可通过与已知的乙醇发酵工艺比较引证若干优点:

[0148] -厌氧工艺是可能的。

[0149] -氧限制的条件也是可能的。

[0150] -可获得更高的乙醇产率和乙醇生产速率。

[0151] -使用的菌株可能能够使用L-阿拉伯糖和可选的木糖。

[0152] 除了上述发酵工艺外,提供了另一种发酵工艺作为本发明的另一方面,其中至少使用两种不同的细胞发酵包含至少两种选自下组的碳来源的碳源,该组由L-阿拉伯糖的来源、木糖来源和葡萄糖来源组成,但不仅限于此。在该发酵工艺中,“至少两种不同的细胞”表示该工艺优选地是共同发酵工艺。在一个优选的实施方案中,使用两种不同的细胞:一种是如前文定义的本发明细胞,其能够使用L-阿拉伯糖和/或将其转化为L-核酮糖和/或5-磷酸木酮糖和/或期望的发酵产物如乙醇,并可选的能够使用木糖;另一种是例如WO 03/062430和/或WO 06/009434中定义的菌株,其能够使用木糖和/或将其转化为期望的发酵产物如乙醇。能够使用木糖的细胞优选地是如前文定义的显示直接将木糖异构化为木酮糖(在一个步骤中)的能力的菌株。这两种不同的菌株优选地在存在L-阿拉伯糖来源、木糖来源和可选的葡萄糖来源时被培养。可共同培养三种或更多不同的细胞,和/或可使用三种或更多碳来源,只要至少一种细胞能够使用存在的至少一种碳来源和/或将其转化为期望的发酵产物如乙醇即可。表述“使用至少一种碳来源”具有与表述“使用L-阿拉伯糖”相同的含义。表述“将其(即碳来源)转化为期望的发酵产物”具有与表述“将L-阿拉伯糖转化为期望

的发酵产物”相同的含义。

[0153] 在一个优选的实施方案中,本发明涉及生产选自下组的发酵产物的工艺,该组由乙醇、乳酸、3-羟基-丙酸、丙烯酸、乙酸、琥珀酸、柠檬酸、苹果酸、延胡索酸、氨基酸、1,3-丙烷-二醇、乙烯、甘油、丁醇、 β -内酰胺抗生素和头孢菌素组成,其中该工艺包括下述步骤:

[0154] (a) 用前文定义的本发明的细胞和能够使用木糖和/或显示直接将木糖转化为木酮糖的能力的细胞发酵培养基,所述培养基至少含有L-阿拉伯糖来源和木糖来源,其中各细胞将L-阿拉伯糖和/或木糖发酵为发酵产物,和可选地

[0155] (b) 回收所述发酵产物。

[0156] 上述发酵工艺的所有优选的实施方案也是该进一步发酵工艺的优选的实施方案:鉴定发酵产物、鉴定L-阿拉伯糖来源和木糖来源、发酵条件(需氧或厌氧条件、氧限制的条件、进行该工艺的温度、乙醇生产力、乙醇产率)。

[0157] 遗传修饰

[0158] 为了在如上所述本发明的宿主细胞中过表达酶,以及为了宿主细胞(优选地为酵母)的其它遗传修饰,通过本领域公知的方法用本发明的多种核酸构建体转化宿主细胞。这类方法例如从标准手册已知,例如Sambrook and Russel (2001) "Molecular Cloning: A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, 或F. Ausubel et al., eds., "Current protocols in molecular biology", Green Publishing and Wiley Interscience, New York (1987)。用于真菌宿主细胞转化和遗传修饰的方法可从例如EP-A-0635574、WO 98/46772、WO 99/60102和WO 00/37671已知。

[0159] 上文已描述了用于在本发明宿主细胞中过表达酶的核酸构建体中使用的启动子。在用于过表达的核酸构建体中,编码酶的核苷酸序列的3'-端优选地与转录终止子序列可操作地连接。优选地,该转录终止子序列在选择宿主细胞例如选择的酵母细胞中可操作。在任何情况下,终止子的选择不是关键性的;其可例如来自任何酵母基因,尽管如果来自非酵母的真核基因时终止子有时也可工作。转录终止序列还优选地包含多聚腺苷酸化信号。优选的终止子序列是醇脱氢酶(ADH1)和PGI1终止子。更优选地,ADH1和PGI1终止子均来自*S. cerevisiae* (分别为SEQ ID NO:50和SEQ ID NO:53)。

[0160] 可选地,可选择的标记物可存在于核酸构建体中。本文使用术语“标记物”表示编码性状或表型的基因,所述性状或表型允许选择或筛选含有该标记物的宿主细胞。标记物基因可以是抗生素抗性基因,其中适当的抗生素可被用于从未被转化的细胞中选择经转化的细胞。然而,优选地使用非抗生素抗性标记物,如营养缺陷标记物(URA3, TRP1, LEU2)。在一个优选的实施方案中,用核酸构建体转化的宿主细胞是无标记基因的。用于构建无重组标记物基因宿主细胞的方法在EP-A-0635574中公开并且基于双向标记物的使用。或者,可筛选的标记物如绿色荧光蛋白、lacZ、萤光素酶、氯霉素乙酰转移酶、 β -葡萄糖苷酸酶可被掺入本发明的核酸构建体中,允许筛选经转化的细胞。

[0161] 可存在于本发明核酸构建体中的可选的其它元件包括但不限于,一条或多条前导序列、增强子、整合因子和/或受体基因、内含子序列、着丝粒、调聚物和/或基质附着(MAR)序列。本发明的核酸构建体还可包含用于自主复制的序列,如ARS序列。合适的附加体核酸构建体可例如基于酵母的2 μ 或pKD1 (Fleer et al., 1991, Biotechnology 9:968-975) 质

粒。或者,核酸构建体可包含用于整合的序列,优选地通过同源重组整合。这类序列因此可以是与宿主细胞基因组中用于整合的靶位点同源的序列。本发明的核酸构建体可以用本身已知的方式提供,所述方式通常涉及例如限制性酶切(restricting)和连接核酸/核酸序列的技术,针对该技术参考标准手册例如Sambrook and Russel(2001) "Molecular Cloning: A Laboratory Manual (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press.

[0162] 用于酵母或真菌中失活和基因破坏的方法是本领域公知的(见例如Fincham, 1989, Microbiol Rev. 53 (1):148-70或EP-A-0 635 574)。

[0163] 在本文件及其权利要求中,动词“包含”及其变化形式以其非限制性的含义使用,表示该词后面的项目是包含在内的,但是不排除未明确提到的项目。另外,不定冠词“一个/种”(“a”或“an”)涉及元素时不排除存在多于一个元素的可能性,除非上下文明确地要求存在且仅存在一个元素。因此,不定冠词“一个/种”(“a”或“an”)通常表示“至少一个/种”。

[0164] 本发明还通过以下的实施例描述,所述实施例不应被解释为限制本发明的范围。

实施例

[0165] 质粒和菌株构建

[0166] 菌株

[0167] 本工作中描述的消耗L-阿拉伯糖的*Sachharomyces cerevisiae*菌株基于菌株RWB220,其自身是RWB217的衍生物。RWB217是CEN.PK菌株,其中编码戊糖磷酸途径中酶表达的基因TAL1、TKL1、RPE1、RKI1已被过表达(Kuyper et al., 2005a)。另外,编码醛糖还原酶(GRE3)的基因已被缺失。菌株RWB217还含有两种质粒:单拷贝质粒,其带有用于过表达木酮糖激酶(XKS1)的LEU2标记物和附加体;多拷贝质粒,其带有URA3作为表达木酮糖异构酶Xyl1A的标记物。针对在木糖上改善的生长对RWB217进行选择步骤,这在Kuyper et al. (2005b)中描述。该步骤得到了两种纯净的菌株RWB218(Kuyper et al., 2005b)和RWB219。RWB218和RWB219之间的差异是在选择步骤后,通过在含葡萄糖作为碳源的矿物质培养基上涂布和重新划线获得RWB218,而对RWB219使用木糖作为碳源。

[0168] 将菌株RWB219在含葡萄糖作为碳源的YP (YPD) 上非选择性地培养,从而促进两种质粒的丢失。涂布在YPD上后,通过在尿嘧啶和亮氨酸辅源营养上观察,测试质粒丢失的单个菌落。用含有cre重组酶的pSH47转化两种质粒均丢失的菌株,从而去除整合RKI1过表达构建体后仍然存在的KanMX盒(Guldener et al., 1996)。将带有质粒的菌落重悬于含1%半乳糖的酵母蛋白胨培养基(YP) (10g/l酵母提取物和20g/l蛋白胨,均来自BD Difco Belgium)中,在30℃孵育1小时。将约200个细胞涂布在YPD上。针对KanMX标记物(G418抗性)和pSH47 (URA3) 的丢失检验得到的菌落。然后将KanMX标记物和pSH47质粒均丢失的菌株命名为RWB220。为了获得在本专利中测试的菌株,用pRW231和pRW243转化RWB220,得到菌株IMS0001。

[0169] 在构建期间,菌株在复合的YP:10g l⁻¹酵母提取物(BD Difco)、20g l⁻¹蛋白胨(BD Difco)上或在合成培养基(MY) (Verduyn et al., 1992)上维持,所述培养基补充有葡萄糖(2%)作为碳源(YPD或MYD)并且在平板的情况下补充有1.5%琼脂。用质粒转化后,将菌株涂布在MYD上。酵母的转化根据Gietz and Woods (2002)进行。质粒在*Escherichia coli*菌

株XL-1 blue (Stratagene, La Jolla, CA, 美国) 上扩增。转化根据Inoue et al. (1990) 进行。E.coli在LB (Luria-Bertani) 平板上或液体TB (Terrific Broth) 培养基中培养,用于分离质粒 (Sambrook et al, 1989)。

[0170] 质粒

[0171] 为了在L-阿拉伯糖上生长,酵母需要表达三种不同的基因:L-阿拉伯糖异构酶 (AraA)、L-核酮糖激酶 (AraB) 和L-核酮糖-5-P-4-差向异构酶 (AraD) (Becker and Boles, 2003)。在本工作中,我们选择在*S.cerevisiae*中表达来自乳酸细菌*Lactobacillus plantarum*的AraA、AraB和AraD。因为最终的目的是将L-阿拉伯糖与其它糖如D-木糖组合消耗,所以将编码细菌L-阿拉伯糖途径的基因与编码D-木糖消耗的基因组合在同一质粒上。

[0172] 为了得到高水平的表达,将*L.plantarum* AraA和AraD基因连接进带有2 μ XylA的质粒pAKX002中。

[0173] 通过用SpeI 5' Ptdh3和5' AraAPtdh3扩增截短版的TDH3启动子 (SEQ ID NO:49)、用Ptdh5' AraA和Tadh3' AraA扩增AraA基因和用3' AraATadh1和3' Tadh1-SpeI扩增ADH1终止子 (SEQ ID NO:50),构建AraA盒。从凝胶中提取三种片段并将其以大致等摩尔的量混合。使用SpeI-5' Ptdh3和3' Tadh1SpeI寡聚体对该混合物进行PCR。凝胶纯化得到的P_{TDH3}-AraA-T_{ADH1}盒,在5'和3'SpeI位点切割,然后连接进用NheI切割的pAKX002中,得到质粒pRW230。

[0174] 如下制造AraD构建体:首先用寡聚体SalI 5' Phxt7和5' AraDPhxt扩增截短版的HXT7启动子 (SEQ ID NO:52),用Phxt5' AraD和Tpgi3' AraD扩增AraD基因,和用3' AraDTpgi和3' TpgiSalI寡聚体扩增GPI1终止子 (SEQ ID NO:53) 区。从凝胶中提取得到的片段并将其以大致等摩尔的量混合,然后使用SalI 5' Phxt7和3' TpgiSalI寡聚体进行PCR。凝胶纯化得到的P_{HXT7}-AraD-T_{GPI1}盒,在5'和3'SalI位点切割,然后连接进用XhoI切割的pRW230中,得到质粒pRW231 (图1)。

[0175] 因为L-核酮糖激酶的过高表达对生长是有害的 (Becker and Boles, 2003),所以将AraB基因与编码木酮糖激酶的XKS1基因组合在整合质粒上。为此,首先如下将p415ADHXKS (Kuyper et al., 2005a) 改变为pRW229:用PvuI切割p415ADHXKS和pRS305二者,并将来自p415ADHXKS的含ADHXKS的PvuI片段与来自pRS305的载体主链连接,得到pRW229。

[0176] 通过用SacI 5' Ppgi1和5' AraBPpgi1寡聚体扩增PGI1启动子,用Ppgi5' AraB和Tadh3' AraB寡聚体扩增AraB基因,和用3' AraBTadh1和3' Tadh1SacI寡聚体扩增ADH1终止子,制造PGI1启动子 (SEQ ID NO:51) 和ADH1终止子 (SEQ ID NO:50) 之间含有*L.plantarum* AraB基因的盒。从凝胶中提取三种片段并将其以大致等摩尔的量混合。使用SacI-5' Ppgi1和3' Tadh1SacI寡聚体对该混合物进行PCR。凝胶纯化得到的P_{PGI1}-AraB-T_{ADH1}盒,在5'和3'SacI位点切割,然后连接进用SacI切割的pRW22中,得到质粒pRW243 (图1)。

[0177] 用pRW231和pRW243 (表2) 转化菌株RWB220,得到菌株IMS0001。

[0178] 限制性内切核酸酶 (New England Biolabs, Beverly, MA, 美国和Roche, Basel, 瑞士) 和DNA连接酶 (Roche) 根据制造商的说明使用。用Qiaprep离心小量制备试剂盒 (Qiagen, Hilden, Germany) 从E.coli分离质粒。在1 $\times$ TBE (Sambrook et al, 1989) 中于1%琼脂糖 (Sigma, St. Louis, MO, 美国) 凝胶上分离DNA片段。用Qiaquick凝胶提取试剂盒 (Qiagen) 从凝胶上分离片段。AraA、AraB和AraD盒 (的元件) 的扩增根据制造商的说明用VentR DNA聚合酶 (New England Biolabs) 完成。启动子和终止子的模板为*S.cerevisiae* CEN.PK113-7D的

染色体DNA, Ara基因的模板为*Lactobacillus plantarum* DSM20205的染色体DNA。聚合酶链式反应(PCR)在Biometra TGradient Thermocycler (Biometra, **Göttingen**, 德国)中进行, 使用一下的设置: 在55℃、60℃或65℃退火1分钟, 在75℃延伸1到3分钟(取决于期望的片段长度)和在94℃变性1分钟, 进行30个循环。

[0179] 培养和培养基

[0180] 在30℃于合成培养基(Verduyn et al., 1992)中进行摇瓶培养。灭菌前用2M KOH将培养基的pH调节至6.0。对固体合成培养基而言, 添加1.5%琼脂。

[0181] 通过用冷冻菌种培养物接种500-ml摇瓶中含适当糖的100ml培养基, 制备预培养物。在300℃于定轨摇床(200rpm)中孵育后, 使用该培养物接种摇瓶培养物或发酵罐培养物。用于厌氧培养的合成培养基补充有溶于乙醇中的 0.42g l^{-1} Tween 80和 0.01g l^{-1} 麦角固醇(Andreasen and Stier, 1953; Andreasen and Stier, 1954)。厌氧(程序化)分批培养在30℃下于工作体积为1升的2升实验室发酵罐(Applikon, Schiedam, 瑞士)中进行。通过自动添加2M KOH将培养物pH维持在pH 5.0。以800rpm搅拌培养物, 并用 0.5 l min^{-1} 的氮气鼓泡($<10\text{ppm}$ 氧)。为了最小化氧的扩散, 对发酵罐装配Norprene输送管(Cole Palmer Instrument company, Vernon Hills, USA)。用氧电极(Applisens, Schiedam, 荷兰)监测溶解氧。通过以约 0.05 l min^{-1} 顶空充气在相同的实验装置中达到氧限制的条件。

[0182] 干重的测定

[0183] 在预先称重的硝酸纤维素滤纸(孔径 $0.45\mu\text{m}$; Gelman laboratory, Ann Arbor, USA)上过滤培养物样品(10.0ml)。去除培养基后, 用去离子水洗涤滤纸并在微波炉(Bosch, Stuttgart, 德国)中于360W干燥20分钟并称重。一式两份的测定变化小于1%。

[0184] 气体分析

[0185] 在冷凝器(2℃)中冷却排出气, 并用MD-110-48P-4型Permapure干燥器(Permapure, Toms River, USA)干燥。用NGA 2000分析仪(Rosemount Analytical, Orrville, USA)测定 O_2 和 CO_2 浓度。如先前所述(Van Urk et al., 1988; Weusthuis et al., 1994)测定排出气体流速和氧的比消耗速率与二氧化碳的比生产速率。在计算这些生物-比速率时, 考虑由取出培养物样品引起的体积改变。

[0186] 代谢物分析

[0187] 使用Waters Alliance 2690 HPLC(Waters, Milford, USA)通过HPLC分析葡萄糖、木糖、阿拉伯糖、木糖醇、有机酸、甘油和乙醇, 所述设备装配有BioRad HPX 87H柱(BioRad, Hercules, USA)、Waters 2410趋光指数检测器和Waters 2487UV检测器。在60℃下用 0.5g l^{-1} 的硫酸以 0.6ml min^{-1} 的流速洗脱柱。

[0188] 对5-磷酸木酮糖的测定(Zaldivar J., et al, Appl. Microbiol. Biotechnol., (2002), 59:436-442)

[0189] 为了分析细胞内代谢产物如5-磷酸木酮糖, 在葡萄糖耗尽之前(培养的第22和26小时)和葡萄糖耗尽之后(培养的42、79和131小时)从反应器中一式两份手机5ml培养液。用于停止代谢、固相提取代谢产物和分析的步骤已由Smits H.P. et al. (Anal. Biochem., 261:36-42, (1998))详细描述。然而, 通过高压离子交换色谱的分析被轻微修饰, 所述高压离子交换色谱与用于分析细胞提取物的脉冲安培检测偶联。使用的溶液为洗脱液A、75mM NaOH和洗脱液B、500mM NaAc。为了防止洗脱液中碳酸盐的污染, 使用具有低碳酸盐浓度的

NaOH溶液 (Baker Analysed, Deventer, 荷兰) 代替NaOH块。用氦 (He) 将洗脱液脱气30分钟,, 然后保持在氦气压下。将梯度泵编程为产生以下的梯度:100% A和0% B (0分钟), A到70% 的线性降低和B到30%的线性提高 (0-30分钟), A到30%的线性降低和B到70%的线性提高 (30-70分钟), A到0的线性降低和B到100%的线性提高 (70-75分钟), 0% A和100% B (75-85分钟), A到100%的线性提高和B到0%的线性降低 (85-95分钟)。流动相以1ml/分钟的流速移动。其它条件根据Smits et al. (1998)。

[0190] 碳回收

[0191] 碳回收被计算为形成的产物中的碳除以消耗的糖碳总量, 并且基于48%的生物物质碳含量。为了针对发酵期间的乙醇蒸发进行校正, 推测产生的乙醇量等于测量的累计CO₂生产减去由于生物物质合成而发生的CO₂生产 (每克生物物质5.85mmol CO₂ (Verduyn et al., 1990)) 和与乙酸盐形成相关的CO₂生产。

[0192] 针对在L-阿拉伯糖上的生长进行选择

[0193] 根据上述步骤构建菌株IMS0001 (CBS 120327, 在27/09/06保藏于CBS), 其含有编码木糖 (XylA和XKS1) 和阿拉伯糖 (AraA、AraB、AraD) 二者代谢途径的基因。尽管能够在木糖上生长 (数据未显示), 但是菌株IMS0001似乎不能在补充有2%L-阿拉伯糖的固体合成培养基上生长。通过在摇瓶中的系列转移和通过在发酵罐中的程序化分批培养 (SBR) 选择能够利用L-阿拉伯糖作为碳源生长的IMS0001突变体。

[0194] 对系列转移实验而言, 用菌株IMS0001或参考菌株RWB219接种含有100ml合成培养基的500-ml摇瓶, 所述培养基含有0.5%半乳糖。72小时后, 在3.0的660nm吸光度下, 使用培养物接种新的摇瓶, 其含有0.1%半乳糖和2%阿拉伯糖。基于用D-核酮糖作为校准标准的HPLC测定, 测定了在菌株IMS0001的第一次培养中, 在含半乳糖/阿拉伯糖混合物的培养基上, 部分阿拉伯糖已被转化为核酮糖并随后分泌至上清液中。使用Waters Alliance 2690 HPLC (Waters, Milford, USA) 进行这些HPLC分析, 所述设备装配有BioRad HPX 87H柱 (BioRad, Hercules, USA)、Waters2410趋光指数检测器和Waters 2487UV检测器。在60℃下用0.5g l⁻¹的硫酸以0.6ml min⁻¹的流速洗脱柱。与参考菌株RWB219相反, 菌株IMS0001培养物的OD₆₆₀在耗尽半乳糖后提高。当观察到菌株IMS0001在阿拉伯糖上生长约850个小时后 (图2), 将该培养物以1.7的OD₆₆₀转移至含2%阿拉伯糖的摇瓶中。然后将培养物以2-3的OD₆₆₀先后顺序至含2%阿拉伯糖的信箱培养基中, 通过间断地用HPLC测量阿拉伯糖浓度证实阿拉伯糖的利用 (数据未显示)。在约3600小时内这些培养物的生长速率从0提高至0.15h⁻¹ (图3)。

[0195] 通过用100ml摇瓶培养物接种1升补充有2%阿拉伯糖的合成培养基, 开始在氧限制条件下的分批发酵, 所述摇瓶培养物是阿拉伯糖培养的IMS0001细胞的摇瓶培养物, 其以约0.12h⁻¹的最大生长速率在2% of L-阿拉伯糖上生长。当观察到在阿拉伯糖上生长时, 通过用氮气鼓泡对培养物加以厌氧条件。通过手动或自动地将90%的培养物替换为含20g l⁻¹阿拉伯糖的合成培养基, 开始厌氧分批培养的顺序循环。对SBR发酵期间的每个循环而言, 从CO₂谱估计指数生长速率 (图4)。在13个循环中, 指数生长速率从0.025h⁻¹提高至0.08h⁻¹。20个循环后取样, 涂布在补充有2% L-阿拉伯糖的固体合成培养基上, 并在30℃孵育数天。在含L-阿拉伯糖的固体合成培养基上将分离的菌落再划线两次。最后, 用单个菌落接种含有含2%L-阿拉伯糖的合成培养基的摇瓶, 并在30℃孵育5天。该培养物被命名为菌株

IMS0002 (CBS 120328, 在27/09/06保藏于Centraal Bureau voor Schimmelculturen (CBS))。采取培养物样品, 添加30%甘油并将样品储存于-80℃下。

[0196] 混合培养物发酵

[0197] 生物质水解产物是工业生物技术期望的原料, 其含有由多种糖组成的复杂混合物, 所述糖中葡萄糖、木糖和阿拉伯糖通常存在于显著的级分中。为了完成葡萄糖和阿拉伯糖以及木糖的乙醇发酵, 用阿拉伯糖发酵菌株IMS0002和木糖发酵菌株RWB218的混合培养物进行厌氧分批发酵。用100ml 菌株IMS0002的预培养物接种含800ml 合成培养基的厌氧分批发酵罐, 所述合成培养基含有 30g l^{-1} D-葡萄糖、 15g l^{-1} D-木糖和 15g l^{-1} L-阿拉伯糖。10小时后添加100ml RWB218接种物。与仅用菌株IMS0002进行混合糖发酵相反, 在葡萄糖耗尽后木糖和阿拉伯糖均被消耗 (图5D)。混合培养完全消耗了所有的糖, 并且在80小时内以 0.42g g^{-1} 糖的高总产率生产了 $564.0 \pm 6.3\text{mmol l}^{-1}$ 乙醇 (从 CO_2 生产计算)。木糖醇仅小量产生, 浓度为 4.7mmol l^{-1} 。

[0198] 对菌株IMS0002的表征

[0199] 在合成培养基上厌氧分批发酵期间测定菌株IMS0002的生长和产物形成, 所述培养基含有L-阿拉伯糖作为唯一的碳源, 或含有葡萄糖、木糖和L-阿拉伯糖的混合物。在含有100ml 含2% L-阿拉伯糖的合成培养基的摇瓶中, 通过用菌株IMS0002的-80℃冷冻菌种接种并在30℃下孵育48小时, 制备用于该厌氧分批发酵的预培养物。

[0200] 图5A显示在约70小时的厌氧分批发酵期间, 菌株IMS0002能够将 20g l^{-1} L-阿拉伯糖发酵为乙醇。使用L-阿拉伯糖作为唯一碳源时, 厌氧条件下的比生长速率为 $0.05 \pm 0.001\text{h}^{-1}$ 。考虑到分批发酵期间的乙醇蒸发, 来自 20g l^{-1} 阿拉伯糖的乙醇产率为 $0.43 \pm 0.003\text{g g}^{-1}$ 。不进行蒸发校正时, 乙醇产率为 $0.35 \pm 0.01\text{g g}^{-1}$ 阿拉伯糖。在阿拉伯糖上的厌氧生长期间未观察到阿拉伯糖醇的形成。

[0201] 在图5B中, 展示了菌株IMS0002对 20g l^{-1} 葡萄糖和 20g l^{-1} L-阿拉伯糖的混合物的乙醇发酵。葡萄糖耗尽后开始消耗L-阿拉伯糖。在70小时内, 葡萄糖和L-阿拉伯糖均被完全消耗。来自总糖的乙醇产率为 $0.42 \pm 0.003\text{g g}^{-1}$ 。

[0202] 在图5C中, 展示了菌株IMS0002对 30g l^{-1} 葡萄糖、 15g l^{-1} D-木糖和 15g l^{-1} L-阿拉伯糖的混合物的发酵模式。在80小时内, 葡萄糖和阿拉伯糖均被完全消耗。100mM木糖中仅20mM被菌株IMS0002消耗。另外, 观察到20mM木糖醇的形成。明显地, 木糖被菌株IMS0002转化为木糖醇。因此, 来自总糖的乙醇产率低于上述发酵: $0.38 \pm 0.001\text{g g}^{-1}$ 。来自总葡萄糖和阿拉伯糖的乙醇产率与其它发酵相似: $0.43 \pm 0.001\text{g g}^{-1}$ 。

[0203] 表1显示了针对菌株IMS0002的厌氧分批发酵观察的阿拉伯糖消耗速率和乙醇生产速率。阿拉伯糖以 $0.23\text{--}0.75\text{g h}^{-1}\text{g}^{-1}$ 生物质干重的速率被消耗。从阿拉伯糖生产乙醇的速率从 0.08 到 $0.31\text{g h}^{-1}\text{g}^{-1}$ 生物质干重变化。

[0204] 最初, 构建的菌株IMS0001能够发酵木糖 (数据未显示)。与我们的预期相反, 所选择的菌株IMS0002不能将木糖发酵为乙醇 (图5C)。为了重新获取发酵木糖的能力, 将菌株IMS0002的菌落转移至含2% D-木糖的固体合成培养基上, 并在厌氧罐中于30℃下孵育25天。随后再次将菌落转移至含2%阿拉伯糖的固体合成培养基上。在30℃孵育4天后, 将菌落转移至含合成培养基的摇瓶中, 所述合成培养基含2%阿拉伯糖。在30℃孵育6天后, 添加30%的甘油, 取样并储存于-80℃下。用这样的冷冻菌种接种摇瓶并用作在含 20g l^{-1} 木糖和

20g L⁻¹阿拉伯糖的合成培养基上厌氧分批发酵的预培养物,所述摇瓶含有100ml含2%阿拉伯糖的合成培养基。在图6中,显示了该分批发酵的发酵模式。木糖和阿拉伯糖同时被消耗。阿拉伯糖在70小时内被耗尽,而木糖在120小时内被耗尽。从总糖中至少生产了250mM乙醇,未计入乙醇的蒸发。假设3.2g L⁻¹的最终生物质干重(假设0.08g g⁻¹糖的生物质产率),则由累计的CO₂生产(355mmol L⁻¹)评价的最终乙醇浓度约为330mmol L⁻¹,对应于0.41g g⁻¹戊糖的乙醇产率。除了乙醇、甘油和有机酸外,产生了少量的木糖醇(约5mM)。

[0205] 菌株IMS0003的选择

[0206] 最初,构建的菌株IMS0001能够发酵木糖(数据未显示)。与我们的预期相反,所选择的菌株IMS0002不能将木糖发酵为乙醇(图5C)。为了重新获取发酵木糖的能力,将菌株IMS0002的菌落转移至含2%D-木糖的固体合成培养基上,并在厌氧罐中于30℃下孵育25天。随后再次将菌落转移至含2%阿拉伯糖的固体合成培养基上。在30℃孵育4天后,将菌落转移至含合成培养基的摇瓶中,所述合成培养基含2%阿拉伯糖。在30℃孵育6天后,添加30%的甘油,取样并储存于-80℃下。

[0207] 将来自于该冷冻菌种的样品在含2%L-阿拉伯糖的固体合成培养基上涂布,并在30℃孵育数天。在含L-阿拉伯糖的固体合成培养基上将分离的菌落再划线两次。最后,用单个菌落接种含有含2%L-阿拉伯糖的合成培养基的摇瓶,并在30℃孵育4天。该培养物被命名为菌株IMS0003(CBS 121879,在20/09/07保藏于CBS)。采取培养物样品,添加30%甘油并将样品储存于-80℃下。

[0208] 对菌株IMS0003的表征

[0209] 在合成培养基上厌氧分批发酵期间测定菌株IMS0003的生长和产物形成,所述培养基含有30g L⁻¹葡萄糖、15g L⁻¹ D-木糖和15g L⁻¹ L-阿拉伯糖的混合物。在含有100ml含2%L-阿拉伯糖的合成培养基的摇瓶中,通过用菌株IMS0003的-80℃冷冻菌种接种并在30℃下孵育48小时,制备用于该厌氧分批发酵的预培养物。

[0210] 在图7中,显示了菌株IMS0003对30g L⁻¹葡萄糖、15g L⁻¹ D-木糖和15g L⁻¹ L-阿拉伯糖混合物的发酵模式。葡萄糖耗尽后开始消耗阿拉伯糖。在70小时内,葡萄糖、木糖和阿拉伯糖均被完全消耗。木糖和阿拉伯糖被同时消耗。从总糖中至少产生406mM乙醇,未计入乙醇的蒸发。从累计的CO₂生产计算的最终乙醇浓度为572mmol L⁻¹,对应于0.46g g⁻¹总糖的乙醇产率。与菌株IMS0002(图5C)或菌株IMS0002和RWB218的混合培养物(图5D)对葡萄糖、木糖和阿拉伯糖混合物的发酵相反,菌株IMS0003不产生可检测量的木糖醇。

[0211] 表格

[0212] 表1:使用的*S.cerevisiae*菌株。

[0213]

菌株	特征	参考文献
RWB217	<i>MATA ura3-52 leu2-112 loxP-P_{TP1}::(-266,-1)TAL1 gre3::hphMX pUGP_{TP1}-TKL1 pUGP_{TP1}-RPE1 KanloxP-P_{TP1}::(-?,-1)RKII {p415ADHXKS, pAKX002}</i>	Kuyper et al. 2005a
RWB218	<i>MATA ura3-52 leu2-112 loxP-P_{TP1}::(-266,-1)TAL1 gre3::hphMX pUGP_{TP1}-TKL1 pUGP_{TP1}-RPE1 KanloxP-P_{TP1}::(-?,-1)RKII {p415ADHXKS1, pAKX002}</i>	Kuyper et al. 2005b
RWB219	<i>MATA ura3-52 leu2-112 loxP-P_{TP1}::(-266,-1)TAL1 gre3::hphMX pUGP_{TP1}-TKL1 pUGP_{TP1}-RPE1 KanloxP-P_{TP1}::(-?,-1)RKII {p415ADHXKS1, pAKX002}</i>	本工作
RWB220	<i>MATA ura3-52 leu2-112 loxP-P_{TP1}::(-266,-1)TAL1 gre3::hphMX pUGP_{TP1}-TKL1 pUGP_{TP1}-RPE1 loxP-P_{TP1}::(-?,-1)RKII</i>	本工作
IMS0001	<i>MATA ura3-52 leu2-112 loxP-P_{TP1}::(-266,-1)TAL1 gre3::hphMX pUGP_{TP1}-TKL1 pUGP_{TP1}-RPE1 loxP-P_{TP1}::(-?,-1)RKII {pRW231, PRW243}</i>	本工作
IMS0002	<i>MATA ura3-52 leu2-112 loxP-P_{TP1}::(-266,-1)TAL1 gre3::hphMX pUGP_{TP1}-TKL1 pUGP_{TP1}-RPE1 loxP-P_{TP1}::(-?,-1)RKII {pRW231, PRW243}</i> , 在 L-阿拉伯糖上 针对厌氧生长进行选择	本工作

[0214] 表2:使用的质粒

[0215]

质粒	特征	参考文献
pRS305	整合, LEU2	Gietz and Sugino, 1988
pAKX002	2 μ , URA3, P _{TPII} -Piromyces xylA	Kuyper et al. 2003
p415ADHXS1	CEN, LEU2, P _{ADH1} -S.cerXKS1	Kuyper et al., 2005a
pRW229	整合, LEU2, P _{ADH1} -S.cerXKS1	本工作
pRW230	带有 P _{TDH3} -AraA 的 pAKX002	本工作
pRW231	带有 P _{TDH3} -AraA 和 P _{HXT7} -AraD 的 pAKX002	本工作
pRW243	LEU2, 整合, P _{ADH1} -ScXKS1-T _{CYC} , P _{PGI1} -L.plantarumAraB-T _{ADH1}	本工作

[0216] 表3:本工作中使用的寡聚物

[0217]

寡聚物	DNA 序列
AraA 表达盒	
SpeI5'Ptdh3 SEQ ID NO:31	5'GACTAGTCGAGTTTATCATTATCAATACTGC3'
5'AraAPtdh SEQ ID NO:32	5'CTCATAATCAGGTACTGATAACATTTTGTGTTTATGTGTGTTTATTC3'
Ptdh5'AraA SEQ ID NO:33	5'GAATAAACACACATAAAACAAAATGTTATCAGTACCTGATTATGAG3
Tadh3'AraA SEQ ID NO:34	5'AATCATAAATCATAAGAAATTCGCTTACTTTAAGAATGCCTTAGTCAT3'
3'AraATadh1 SEQ ID NO:35	5'ATGACTAAGGCATTCTTAAAGTAAGCGAATTTCTTATGATTTATGATT3'
3'Tadh1SpeI SEQ ID NO:36	5'CACTAGTCTCGAGTGTGGAAGAACGATTACAACAGG3'
AraB 表达盒	
SacI5'Ppgi1 SEQ ID NO:37	5'CGAGCTCGTGGGTGTATTGGATTATAGGAAG3'
5'AraBPpgi1 SEQ ID NO:38	5'TTGGGCTGTTTCAACTAAATTCATTTTAGGCTGGTATCTTGATTCTA3'
Ppgi5'AraB SEQ ID NO:39	5'TAGAATCAAGATACCAGCCTAAAAATGAATTTAGTTGAAACAGCCCAA3'
Tadh3'AraB SEQ ID NO:40	5'AATCATAAATCATAAGAAATTCGCTCTAATATTTGATTGCTTGCCCAG3'
3'AraBTadh1 SEQ ID NO:41	5'CTGGGCAAGCAATCAAATATTAGAGCGAATTTCTTATGATTTATGATT3'
3'Tadh1SacI SEQ ID NO:42	5'TGAGCTCGTGTGGAAGAACGATTACAACAGG3'
AraD 表达盒	
Sall5'Phxt7	5'ACGCGTCGACTCGTAGGAACAATTCGG3'

[0218]

SEQ ID NO:43	
5'AraDPhxt SEQ ID NO:44	5'CTTCTTGTTTTAATGCTTCTAGCATTTTTTGATTAAAAATTAAAAAACTT3'
Phxt5'AraD SEQ ID NO:45	5'AAGTTTTTTTTAATTTTAATCAAAAAATGCTAGAAGCATTAAAAACAAGAAG3'
Tpgi3'AraD SEQ ID NO:46	5'GGTATATAITTAAGAGCGATTTGTTTACTTGCGAACTGCAATGATCC3'
3'AraDTpgi SEQ ID NO:47	5'GGATCATGCAGTTTCGCAAGTAAACAAATCGCTCTTAAATATATACC3'
3'TpgiSall SEQ ID NO:48	5'CGCAGTCGACCTTTTAAACAGTTGATGAGAACC3'

[0219] 表4:

[0220] 在*S.cerevisiae* IMS0002的厌氧分批发酵期间观察到的最大葡萄糖和阿拉伯糖比消耗速率和乙醇比生产速率。

[0221] q_{glu} : 葡萄糖比消耗速率

[0222] q_{ara} : 阿拉伯糖比消耗速率

[0223] $q_{eth, glu}$: 在葡萄糖上生长期间的乙醇比生产速率

[0224] $q_{eth, ara}$: 在阿拉伯糖上生长期间的乙醇比生产速率

[0225]

C-源	q_{glu} $g\ h^{-1}\ g^{-1}\ DW$	q_{ara} $g\ h^{-1}\ g^{-1}\ DW$	$q_{eth, glu}$ $g\ h^{-1}\ g^{-1}\ DW$	$q_{eth, ara}$ $g\ h^{-1}\ g^{-1}\ DW$
20 g l ⁻¹ 阿拉伯糖	-	0.75 ± 0.04	-	0.31 ± 0.02
20 g l ⁻¹ 葡萄糖 20 g l ⁻¹ 阿拉伯糖	2.08 ± 0.09	0.41 ± 0.01	0.69 ± 0.00	0.19 ± 0.00
30 g l ⁻¹ 葡萄糖 15 g l ⁻¹ 木糖 15 g l ⁻¹ 阿拉伯糖	1.84 ± 0.04	0.23 ± 0.01	0.64 ± 0.03	0.08 ± 0.01

[0226] 参考文献列表

[0227] Andreassen AA, Stier TJ (1954) Anaerobic nutrition of *Saccharomyces cerevisiae*. II. Unsaturated fatty acid requirement for growth in a defined medium. *J Cell Physiol* 43:271-281

[0228] Andreassen AA, Stier TJ (1953) Anaerobic nutrition of *Saccharomyces cerevisiae*. I. Ergosterol requirement for growth in a defined medium. *J Cell*

Physiol 41:23-36

[0229] Becker J,Boles E(2003)A modified *Saccharomyces cerevisiae* strain that consumes L-Arabinose and produces ethanol.Appl Environ Microbiol 69:4144-4150

[0230] Gietz R.D.,Sugino A.(1988).New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites.Gene 74:527-534.

[0231] Gietz,R.D.,and R.A.Woods.2002.Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.Methods Enzymol.350:87-96.

[0232] Guldener U,Heck S,Fielder T,Beinhauer J,Hegemann JH.(1996)A new efficient gene disruption cassette for repeated use in budding yeast.Nucleic Acids Res.1996 Jul 1;24(13):2519-24.

[0233] Hauf J,Zimmermann FK,Muller S.Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast *Saccharomyces cerevisiae*.Enzyme Microb Technol.2000 Jun 1;26(9-10):688-698.

[0234] Inoue H.,H.Nojima and H.Okayama,High efficiency transformation of *Escherichia coli* with plasmids.Gene 96(1990),pp.23-28

[0235] Kuyper M,Hartog MMP,Toirkens MJ,Almering MJH,Winkler AA,Van Dijken JP,Pronk JT(2005a)Metabolic engineering of a xylose-isomerase-expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation.Fems Yeast Research 5:399-409

[0236] Kuyper M,Toirkens MJ,Diderich JA,Winkler AA,Van Dijken JP,Pronk JT (2005b) Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting *Saccharomyces cerevisiae* strain.Fems Yeast Research 5:925-934

[0237] Sambrook,K.,Fritsch,E.F.and Maniatis,I.(1989)Molecular Cloning:A Laboratory Manual,2nd edn.Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY.

[0238] Van Urk H,Mak PR,Scheffers WA,Van Dijken JP(1988)Metabolic responses of *Saccharomyces cerevisiae* CBS 8066 and *Candida utilis* CBS 621 upon transition from glucose limitation to glucose excess.Yeast 4:283-291

[0239] Verduyn C,Postma E,Scheffers WA,Van Dijken JP(1990) Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures.J Gen Microbiol 136:395-403

[0240] Verduyn C,Postma E,Scheffers WA,Van Dijken JP(1992)Effect of benzoic acid on metabolic fluxes in yeasts:a continuous-culture study on the regulation of respiration and alcoholic fermentation.Yeast 8:501-517

[0241] Weusthuis RA,Visser W,Pronk JT,Scheffers WA,Van Dijken JP(1994) Effects of oxygen limitation on sugar metabolism in yeasts-a continuous-culture study of the Kluyver effect.Microbiology 140:703-715

序列表

<110> TU·德尔弗特
 AJA·范玛利斯
 JP·玻克
 HW·温斯林克
 JP·范迪加
 AA·温克勒
 JH·德温德
 <120> 发酵阿拉伯糖的真核细胞的代谢工程改造

<130> P6011342PCT

<150> EP06121633.9

<151> 2006-10-02

<150> US 60/848,357

<151> 2006-10-02

<160> 53

<170> PatentIn version 3.3

<210> 1

<211> 474

<212> PRT

<213> Lactobacillus plantarum

<400> 1

[0001]

Met Leu Ser Val Pro Asp Tyr Glu Phe Trp Phe Val Thr Gly Ser Gln
 1 5 10 15

His Leu Tyr Gly Glu Glu Gln Leu Lys Ser Val Ala Lys Asp Ala Gln
 20 25 30

Asp Ile Ala Asp Lys Leu Asn Ala Ser Gly Lys Leu Pro Tyr Lys Val
 35 40 45

Val Phe Lys Asp Val Met Thr Thr Ala Glu Ser Ile Thr Asn Phe Met
 50 55 60

Lys Glu Val Asn Tyr Asn Asp Lys Val Ala Gly Val Ile Thr Trp Met
 65 70 75 80

His Thr Phe Ser Pro Ala Lys Asn Trp Ile Arg Gly Thr Glu Leu Leu
 85 90 95

Gln Lys Pro Leu Leu His Leu Ala Thr Gln Tyr Leu Asn Asn Ile Pro
 100 105 110

Tyr Ala Asp Ile Asp Phe Asp Tyr Met Asn Leu Asn Gln Ser Ala His
 115 120 125

Gly Asp Arg Glu Tyr Ala Tyr Ile Asn Ala Arg Leu Gln Lys His Asn
 130 135 140

Lys Ile Val Tyr Gly Tyr Trp Gly Asp Glu Asp Val Gln Glu Gln Ile

145	150	155	160
Ala Arg Trp Glu	Asp Val Ala Val Ala	Tyr Asn Glu Ser Phe	Lys Val
	165	170	175
Lys Val Ala Arg	Phe Gly Asp Thr	Met Arg Asn Val Ala	Val Thr Glu
	180	185	190
Gly Asp Lys Val	Glu Ala Gln Ile	Lys Met Gly Trp	Thr Val Asp Tyr
	195	200	205
Tyr Gly Ile Gly	Asp Leu Val Glu	Glu Ile Asn Lys	Val Ser Asp Ala
	210	215	220
Asp Val Asp Lys	Glu Tyr Ala Asp	Leu Glu Ser Arg	Tyr Glu Met Val
	225	230	235
Gln Val Asp Asn	Asp Ala Asp Thr	Tyr Lys His Ser	Val Arg Val Gln
	245	250	255
Leu Ala Gln Tyr	Leu Gly Ile Lys	Arg Phe Leu Glu	Arg Gly Gly Tyr
	260	265	270
Thr Ala Phe Thr	Thr Asn Phe Glu	Asp Leu Trp Gly	Met Glu Gln Leu
	275	280	285
Pro Gly Leu Ala	Ser Gln Leu Leu	Ile Arg Asp Gly	Tyr Gly Phe Gly
	290	295	300
Ala Glu Gly Asp	Trp Lys Thr Ala	Ala Leu Gly	Arg Val Met Lys Ile
	305	310	315
Met Ser His Asn	Lys Gln Thr Ala	Phe Met Glu	Asp Tyr Thr Leu Asp
	325	330	335
Leu Arg His Gly	His Glu Ala Ile	Leu Gly Ser His	Met Leu Glu Val
	340	345	350
Asp Pro Ser Ile	Ala Ser Asp Lys	Pro Arg Val Glu	Val His Pro Leu
	355	360	365
Asp Ile Gly Gly	Lys Asp Asp	Pro Ala Arg Leu	Val Phe Thr Gly Ser
	370	375	380
Glu Gly Glu Ala	Ile Asp Val Thr	Val Ala Asp	Phe Arg Asp Gly Phe
	385	390	395
Lys Met Ile Ser	Tyr Ala Val Asp	Ala Asn Lys	Pro Glu Ala Glu Thr
	405	410	415
Pro Asn Leu Pro	Val Ala Lys Gln	Leu Trp Thr	Pro Lys Met Gly Leu

[0002]

	420	425	430
	Lys Lys Gly Ala Leu Glu Trp Met Gln Ala Gly Gly Gly His His Thr 435 440 445		
	Met Leu Ser Phe Ser Leu Thr Glu Glu Gln Met Glu Asp Tyr Ala Thr 450 455 460		
	Met Val Gly Met Thr Lys Ala Phe Leu Lys 465 470		
	<210> 2		
	<211> 1425		
	<212> DNA		
	<213> Lactobacillus plantarum		
	<400> 2		
	atgtttatcag tacctgatta tgagtttttg tttgttaccg gttcacaaca cttttatggt	60	
	gaagaacaat tgaagtctgt tgctaaggat gcgcaagata ttgcggataa attgaatgca	120	
	agcggcaagt taccttataa agtagtcttt aaggatgtta tgacgacggc tgaaagtatc	180	
	accaacttta tgaaagaagt taattacaat gataaggtag ccggtgttat tacttggtatg	240	
	cacacattct caccagctaa gaactggatt cgtggaactg aactgttaca aaaaccatta	300	
	ttacacttag caacgcaata tttgaataat attccatattg cagacattga ctttgattac	360	
[0003]	atgaacctta accaaagtgc ccatggcgac cgcgagtatg cctacattaa cgcccggttg	420	
	cagaaacata ataagattgt ttacggctat tggggcgatg aagatgtgca agagcagatt	480	
	gcacgttggg aagacgtcgc cgtagcgtac aatgagagct ttaaagttaa ggttgctcgc	540	
	tttggcgaca caatgcgtaa tgtggccgtt actgaagggtg acaagggtga agctcaaatt	600	
	aagatgggct ggacagttga ctattatggt atcgggtgact tagttgaaga gatcaataag	660	
	gtttcggatg ctgatgttga taaggaatac gctgacttgg agtctcggta tgaaatggtc	720	
	caagttgata acgatgcgga cacgtataaa cattcagttc gggttcaatt ggcacaatat	780	
	ctgggtatta agcggttctt agaaagaggc gggttacacag cctttaccac gaactttgaa	840	
	gatcttttgg ggatggagca attacctgggt ctagcttcac aattattaat tcgtgatggg	900	
	tatggttttg gtgctgaagg tgactggaag acggctgctt taggacgggt tatgaagatt	960	
	atgtctcaca acaagcaaac cgcttttatg gaagactaca cgtagactt gcgtcatggt	1020	
	catgaagcga tcttaggttc acacatgttg gaagttgatc cgtctatcgc aagtgataaa	1080	
	ccacgggtcg aagttcatcc attggatatt gggggtaaag atgattcctgc tcgcctagta	1140	
	tttactgggt cagaagggtga agcaattgat gtcaccgttg ccgattttccg tgatgggttc	1200	
	aagatgatta gctacgcggt agatgcgaat aagccagaag ccgaaacacc taattttacca	1260	
	gttgctaagc aattatggac cccaaagatg ggcttgaaga aggggtgcact agaattggatg	1320	
	caagctggtg gtggtcacca cacgatgctg tccttctcgt taactgaaga acaaatggaa	1380	
	gactatgcaa ccatgggttg catgactaag gcattcttaa agtaa	1425	

<210> 3
 <211> 533
 <212> PRT
 <213> Lactobacillus plantarum
 <400> 3
 Met Asn Leu Val Glu Thr Ala Gln Ala Ile Lys Thr Gly Lys Val Ser
 1 5 10 15
 Leu Gly Ile Glu Leu Gly Ser Thr Arg Ile Lys Ala Val Leu Ile Thr
 20 25 30
 Asp Asp Phe Asn Thr Ile Ala Ser Gly Ser Tyr Val Trp Glu Asn Gln
 35 40 45
 Phe Val Asp Gly Thr Trp Thr Tyr Ala Leu Glu Asp Val Trp Thr Gly
 50 55 60
 Ile Gln Gln Ser Tyr Thr Gln Leu Ala Ala Asp Val Arg Ser Lys Tyr
 65 70 75 80
 His Met Ser Leu Lys His Ile Asn Ala Ile Gly Ile Ser Ala Met Met
 85 90 95
 [0004] His Gly Tyr Leu Ala Phe Asp Gln Gln Ala Lys Leu Leu Val Pro Phe
 100 105 110
 Arg Thr Trp Arg Asn Asn Ile Thr Gly Gln Ala Ala Asp Glu Leu Thr
 115 120 125
 Glu Leu Phe Asp Phe Asn Ile Pro Gln Arg Trp Ser Ile Ala His Leu
 130 135 140
 Tyr Gln Ala Ile Leu Asn Asn Glu Ala His Val Lys Gln Val Asp Phe
 145 150 155 160
 Ile Thr Thr Leu Ala Gly Tyr Val Thr Trp Lys Leu Ser Gly Glu Lys
 165 170 175
 Val Leu Gly Ile Gly Asp Ala Ser Gly Val Phe Pro Ile Asp Glu Thr
 180 185 190
 Thr Asp Thr Tyr Asn Gln Thr Met Leu Thr Lys Phe Ser Gln Leu Asp
 195 200 205
 Lys Val Lys Pro Tyr Ser Trp Asp Ile Arg His Ile Leu Pro Arg Val
 210 215 220
 Leu Pro Ala Gly Ala Ile Ala Gly Lys Leu Thr Ala Ala Gly Ala Ser
 225 230 235 240

Leu Leu Asp Gln Ser Gly Thr Leu Asp Ala Gly Ser Val Ile Ala Pro
 245 250 255
 Pro Glu Gly Asp Ala Gly Thr Gly Met Val Gly Thr Asn Ser Val Arg
 260 265 270
 Lys Arg Thr Gly Asn Ile Ser Val Gly Thr Ser Ala Phe Ser Met Asn
 275 280 285
 Val Leu Asp Lys Pro Leu Ser Lys Val Tyr Arg Asp Ile Asp Ile Val
 290 295 300
 Met Thr Pro Asp Gly Ser Pro Val Ala Met Val His Val Asn Asn Cys
 305 310 315 320
 Ser Ser Asp Ile Asn Ala Trp Ala Thr Ile Phe Arg Glu Phe Ala Ala
 325 330 335
 Arg Leu Gly Met Glu Leu Lys Pro Asp Arg Leu Tyr Glu Thr Leu Phe
 340 345 350
 Leu Glu Ser Thr Arg Ala Asp Ala Asp Ala Gly Gly Leu Ala Asn Tyr
 355 360 365
 [0005] Ser Tyr Gln Ser Gly Glu Asn Ile Thr Lys Ile Gln Ala Gly Arg Pro
 370 375 380
 Leu Phe Val Arg Thr Pro Asn Ser Lys Phe Ser Leu Pro Asn Phe Met
 385 390 395 400
 Leu Thr Gln Leu Tyr Ala Ala Phe Ala Pro Leu Gln Leu Gly Met Asp
 405 410 415
 Ile Leu Val Asn Glu Glu His Val Gln Thr Asp Val Met Ile Ala Gln
 420 425 430
 Gly Gly Leu Phe Arg Thr Pro Val Ile Gly Gln Gln Val Leu Ala Asn
 435 440 445
 Ala Leu Asn Ile Pro Ile Thr Val Met Ser Thr Ala Gly Glu Gly Gly
 450 455 460
 Pro Trp Gly Met Ala Val Leu Ala Asn Phe Ala Cys Arg Gln Thr Ala
 465 470 475 480
 Met Asn Leu Glu Asp Phe Leu Asp Gln Glu Val Phe Lys Glu Pro Glu
 485 490 495
 Ser Met Thr Leu Ser Pro Glu Pro Glu Arg Val Ala Gly Tyr Arg Glu
 500 505 510

Phe Ile Gln Arg Tyr Gln Ala Gly Leu Pro Val Glu Ala Ala Ala Gly
 515 520 525

Gln Ala Ile Lys Tyr
 530

<210> 4
 <211> 1602
 <212> DNA
 <213> Lactobacillus plantarum

<400> 4
 atgaatttag ttgaaacagc ccaagcgatt aaaactggca aagtttcttt aggaattgag 60
 cttggctcaa ctcgaattaa agccgttttg atcacggacg attttaatac gattgcttcg 120
 ggaagttacg tttgggaaaa ccaatttggt gatggtactt ggacttacgc acttgaagat 180
 gtctggaccg gaattcaaca aagttatacg caattagcag cagatgtccg cagtaaatat 240
 cacatgagtt tgaagcatat caatgctatt ggcattagtg ccatgatgca cggataccta 300
 gcatttgatc aacaagcgaa attattagtt ccgtttcggc cttggcgtaa taacattacg 360
 gggcaagcag cagatgaatt gaccgaatta tttgatttca acattccaca acggtggagt 420
 atcgcgcact tataaccaggc aatcttaaat aatgaagcgc acgttaaaca ggtggacttc 480
 ataacaacgc tggttggtta tgtaacctgg aaattgtcgg gtgagaaagt tctaggaatc 540
 ggtgatgcgt ctggcgtttt cccaattgat gaaacgactg acacatacaa tcagacgatg 600
 ttaaccaagt ttagccaact tgacaaagtt aaaccgtatt catgggatat ccggcatatt 660
 ttaccgcggg ttttaccagc gggagccatt gctggaaagt taacggctgc cggggcgagc 720
 ttacttgatc agagcggcac gtcgacgct ggcagtgtta ttgcaccgcc agaaggggat 780
 gctggaacag gaatggctcg tacgaacagc gtccgtaaac gcacgggtaa catctcggtg 840
 ggaacctcag cattttcgat gaacgttcta gataaacat tgtctaaagt ctatcgcgat 900
 attgatattg ttatgacgcc agatgggtca ccagtgtcaa tgggtgatgt taataattgt 960
 tcatcagata ttaatgctg ggcaacgatt tttcgtgagt ttgcagcccg gttgggaatg 1020
 gaattgaaac cggatcgatt atatgaaacg ttattcttgg aatcaactcg cgctgatgcg 1080
 gatgctggag ggttggttaa ttatagttat caatccgggt agaatattac taagattcaa 1140
 gctggtcggc cgctatttgt acggacacca aacagtaaatt ttagtttacc gaactttatg 1200
 ttgacceaat tatatgcggc gttcgcaccc ctccaacttg gtatggatat tcttgttaac 1260
 gaagaacatg ttcaaacgga cgttatgatt gcacaggggt gattgttccg aacgccggta 1320
 attggccaac aagtattggc caacgcactg aacattccga ttactgtaat gagtactgct 1380
 ggtgaaggcg gcccatgggg gatggcagtg ttagccaact ttgcttgtcg gcaaactgca 1440
 atgaacctag aagatttctt agatcaagaa gtctttaaag agccagaaag tatgacgttg 1500
 agtccagaac cggaacgggt ggccggatat cgtgaattta ttcaacgtta tcaagctggc 1560
 ttaccagttg aagcagcggc tgggcaagca atcaaatatt ag 1602

[0006]

<210> 3
 <211> 533
 <212> PRT
 <213> Lactobacillus plantarum

<400> 3

Met Asn Leu Val Glu Thr Ala Gln Ala Ile Lys Thr Gly Lys Val Ser
 1 5 10 15

Leu Gly Ile Glu Leu Gly Ser Thr Arg Ile Lys Ala Val Leu Ile Thr
 20 25 30

Asp Asp Phe Asn Thr Ile Ala Ser Gly Ser Tyr Val Trp Glu Asn Gln
 35 40 45

Phe Val Asp Gly Thr Trp Thr Tyr Ala Leu Glu Asp Val Trp Thr Gly
 50 55 60

Ile Gln Gln Ser Tyr Thr Gln Leu Ala Ala Asp Val Arg Ser Lys Tyr
 65 70 75 80

His Met Ser Leu Lys His Ile Asn Ala Ile Gly Ile Ser Ala Met Met
 85 90 95

[0007]

His Gly Tyr Leu Ala Phe Asp Gln Gln Ala Lys Leu Leu Val Pro Phe
 100 105 110

Arg Thr Trp Arg Asn Asn Ile Thr Gly Gln Ala Ala Asp Glu Leu Thr
 115 120 125

Glu Leu Phe Asp Phe Asn Ile Pro Gln Arg Trp Ser Ile Ala His Leu
 130 135 140

Tyr Gln Ala Ile Leu Asn Asn Glu Ala His Val Lys Gln Val Asp Phe
 145 150 155 160

Ile Thr Thr Leu Ala Gly Tyr Val Thr Trp Lys Leu Ser Gly Glu Lys
 165 170 175

Val Leu Gly Ile Gly Asp Ala Ser Gly Val Phe Pro Ile Asp Glu Thr
 180 185 190

Thr Asp Thr Tyr Asn Gln Thr Met Leu Thr Lys Phe Ser Gln Leu Asp
 195 200 205

Lys Val Lys Pro Tyr Ser Trp Asp Ile Arg His Ile Leu Pro Arg Val
 210 215 220

Leu Pro Ala Gly Ala Ile Ala Gly Lys Leu Thr Ala Ala Gly Ala Ser
 225 230 235 240

Arg Lys

<210> 6
 <211> 729
 <212> DNA
 <213> Lactobacillus plantarum
 <400> 6
 atgctagaag cattaaaaca agaagtttat gaggctaaca tgcagcttcc aaagctgggc 60
 ctggttactt ttacctgggg caatgtctcg ggcattgacc gggaaaaagg cctattcgtg 120
 atcaagccat ctggtggtga ttatggtgaa ttaaaaccaa gcgatttagt cgttggttaac 180
 ttacaggggtg aagtgggtga aggtaaacta aatccgtcta gtgatacgcc gactcatacg 240
 gtgttatata acgcttttcc taatattggc ggaattgtcc atactcattc gccatgggca 300
 gttgcctatg cagctgctca aatggatgtg ccagctatga acacgaccca tgctgatacg 360
 ttctatggtg acgtgccggc cgcggatgcy ctgactaagg aagaaattga agcagattat 420
 gaaggcaaca cgggtaaaac cattgtgaag acgttccaag aacggggcct cgattatgaa 480
 gctgtaccag cctcattagt cagccagcac ggccatttg cttggggacc aacgccagct 540
 aaagccgttt acaatgctaa agtggttgaa gtggttgccg aagaagatta tcatactgcy 600
 caattgaccc gtgcaagtag cgaattacca caatatatat tagataagca ttatttacgt 660
 aagcatggtg caagtgccta ttatggtcaa aataatgcgc attctaagga tcatgcagtt 720
 cgcaagtaa 729

[0008]

<210> 7
 <211> 437
 <212> PRT
 <213> Piromyces的种
 <400> 7
 Met Ala Lys Glu Tyr Phe Pro Gln Ile Gln Lys Ile Lys Phe Glu Gly
 1 5 10 15
 Lys Asp Ser Lys Asn Pro Leu Ala Phe His Tyr Tyr Asp Ala Glu Lys
 20 25 30
 Glu Val Met Gly Lys Lys Met Lys Asp Trp Leu Arg Phe Ala Met Ala
 35 40 45
 Trp Trp His Thr Leu Cys Ala Glu Gly Ala Asp Gln Phe Gly Gly Gly
 50 55 60
 Thr Lys Ser Phe Pro Trp Asn Glu Gly Thr Asp Ala Ile Glu Ile Ala
 65 70 75 80
 Lys Gln Lys Val Asp Ala Gly Phe Glu Ile Met Gln Lys Leu Gly Ile
 85 90 95

Pro Tyr Tyr Cys Phe His Asp Val Asp Leu Val Ser Glu Gly Asn Ser
 100 105 110
 Ile Glu Glu Tyr Glu Ser Asn Leu Lys Ala Val Val Ala Tyr Leu Lys
 115 120 125
 Glu Lys Gln Lys Glu Thr Gly Ile Lys Leu Leu Trp Ser Thr Ala Asn
 130 135 140
 Val Phe Gly His Lys Arg Tyr Met Asn Gly Ala Ser Thr Asn Pro Asp
 145 150 155 160
 Phe Asp Val Val Ala Arg Ala Ile Val Gln Ile Lys Asn Ala Ile Asp
 165 170 175
 Ala Gly Ile Glu Leu Gly Ala Glu Asn Tyr Val Phe Trp Gly Gly Arg
 180 185 190
 Glu Gly Tyr Met Ser Leu Leu Asn Thr Asp Gln Lys Arg Glu Lys Glu
 195 200 205
 His Met Ala Thr Met Leu Thr Met Ala Arg Asp Tyr Ala Arg Ser Lys
 210 215 220
 [0009] Gly Phe Lys Gly Thr Phe Leu Ile Glu Pro Lys Pro Met Glu Pro Thr
 225 230 235 240
 Lys His Gln Tyr Asp Val Asp Thr Glu Thr Ala Ile Gly Phe Leu Lys
 245 250 255
 Ala His Asn Leu Asp Lys Asp Phe Lys Val Asn Ile Glu Val Asn His
 260 265 270
 Ala Thr Leu Ala Gly His Thr Phe Glu His Glu Leu Ala Cys Ala Val
 275 280 285
 Asp Ala Gly Met Leu Gly Ser Ile Asp Ala Asn Arg Gly Asp Tyr Gln
 290 295 300
 Asn Gly Trp Asp Thr Asp Gln Phe Pro Ile Asp Gln Tyr Glu Leu Val
 305 310 315 320
 Gln Ala Trp Met Glu Ile Ile Arg Gly Gly Gly Phe Val Thr Gly Gly
 325 330 335
 Thr Asn Phe Asp Ala Lys Thr Arg Arg Asn Ser Thr Asp Leu Glu Asp
 340 345 350
 Ile Ile Ile Ala His Val Ser Gly Met Asp Ala Met Ala Arg Ala Leu
 355 360 365

Glu Asn Ala Ala Lys Leu Leu Gln Glu Ser Pro Tyr Thr Lys Met Lys
370 375 380

Lys Glu Arg Tyr Ala Ser Phe Asp Ser Gly Ile Gly Lys Asp Phe Glu
385 390 395 400

Asp Gly Lys Leu Thr Leu Glu Gln Val Tyr Glu Tyr Gly Lys Lys Asn
405 410 415

Gly Glu Pro Lys Gln Thr Ser Gly Lys Gln Glu Leu Tyr Glu Ala Ile
420 425 430

Val Ala Met Tyr Gln
435

<210> 8
<211> 1669
<212> DNA
<213> *Piromyces*的种

[0010]

<400> 8
gtaaatggct aaggaatatt tcccacaaat tcaaaagatt aagttcgaag gtaaggattc 60
taagaatcca ttacgcttcc actactacga tgctgaaaag gaagtcattg gtaagaaaat 120
gaaggattgg ttacgcttcc ccatggcctg gtggcacact ctttgcgccg aagggtgctga 180
ccaattcggg ggaggtacaa agtctttccc atggaacgaa ggtactgatg ctattgaaat 240
tgccaagcaa aagggtgatg ctgggttcga aatcatgcaa aagcttggtg ttccatacta 300
ctgtttccac gatgttgatc ttgtttccga aggtaactct attgaagaat acgaatccaa 360
ccttaaggct gtcgttgctt acctcaagga aaagcaaaag gaaaccggta ttaagcttct 420
ctggagtact gctaacgtct tcggtcacaa gcgttacatg aacgggtgcct cactaacc 480
agactttgat gttgtcgccc gtgctattgt tcaaattaag aacgccatag acgccgggat 540
tgaacttggg gctgaaaact acgtcttctg gggtggtcgt gaagggtaca tgagtctcct 600
taacactgac caaaagcgtg aaaaggaaca catggccact atgcttacca tggctcgtga 660
ctacgctcgt tccaagggat tcaagggtac tttcctcatt gaaccaaagc caatggaacc 720
aaccaagcac caatacgatg ttgacactga aaccgctatt ggtttcctta aggccacaa 780
cttagacaag gacttcaagg tcaacattga agttaaccac gctactcttg ctggtcacac 840
tttcgaacac gaacttgctt gtgctgttga tgctggtatg ctcggttcca ttgatgctaa 900
ccgtgggtgac taccaaaacg gttgggatac tgatcaattc ccaattgatc aatacgaact 960
cgtccaagct tggatggaaa tcatccgtgg tgggtggttc gttactggtg gtaccaactt 1020
cgatgccaaag actcgtcgta actctactga cctcgaagac atcatcattg cccacgtttc 1080
tggtatggat gctatggctc gtgctcttga aaacgctgcc aagctcctcc aagaatctcc 1140
atacaccaag atgaagaagg aacgttacgc ttccttcgac agtggtattg gtaaggactt 1200
tgaagatggg aagctcacc cgaacaagt ttacgaatac ggtaagaaga acggtgaacc 1260

```

aaagcaaact tctggtaagc aagaactcta cgaagctatt gttgccatgt accaataagt 1320
taatcgtagt taaattggta aaataattgt aaaatcaata aacttgtcaa tcctccaatc 1380
aagtttataaa gatcctatct ctgtactaat taaatatagt acaaaaaaaaa atgtataaac 1440
aaaaaaaaagt ctaaaagacg gaagaattta atttagggaa aaaataaaaa taataataaa 1500
caatagataa atcctttata ttaggaaaat gtcccatgtg attattttca tttctactaa 1560
aaaagaaagt aaataaaaca caagaggaaa ttttcccttt tttttttttt tgtaataaat 1620
tttatgcaaa tataaatata aataaaataa taaaaaaaaa aaaaaaaaaa 1669

```

```

<210> 9
<211> 496
<212> PRT
<213> Bacillus subtilis

```

```
<400> 9
```

```
Met Leu Gln Thr Lys Asp Tyr Glu Phe Trp Phe Val Thr Gly Ser Gln
1          5          10          15

```

```
His Leu Tyr Gly Glu Glu Thr Leu Glu Leu Val Asp Gln His Ala Lys
          20          25          30

```

```
Ser Ile Cys Glu Gly Leu Ser Gly Ile Ser Ser Arg Tyr Lys Ile Thr
          35          40          45

```

[0011]

```
His Lys Pro Val Val Thr Ser Pro Glu Thr Ile Arg Glu Leu Leu Arg
          50          55          60

```

```
Glu Ala Glu Tyr Ser Glu Thr Cys Ala Gly Ile Ile Thr Trp Met His
65          70          75          80

```

```
Thr Phe Ser Pro Ala Lys Met Trp Ile Glu Gly Leu Ser Ser Tyr Gln
          85          90          95

```

```
Lys Pro Leu Met His Leu His Thr Gln Tyr Asn Arg Asp Ile Pro Trp
          100          105          110

```

```
Gly Thr Ile Asp Met Asp Phe Met Asn Ser Asn Gln Ser Ala His Gly
          115          120          125

```

```
Asp Arg Glu Tyr Gly Tyr Ile Asn Ser Arg Met Gly Leu Ser Arg Lys
          130          135          140

```

```
Val Ile Ala Gly Tyr Trp Asp Asp Glu Glu Val Lys Lys Glu Met Ser
          145          150          155          160

```

```
Gln Trp Met Asp Thr Ala Ala Ala Leu Asn Glu Ser Arg His Ile Lys
          165          170          175

```

```
Val Ala Arg Phe Gly Asp Asn Met Arg His Val Ala Val Thr Asp Gly

```

	180	185	190
	Asp Lys Val Gly Ala His Ile Gln Phe Gly Trp Gln Val Asp Gly Tyr 195 200 205		
	Gly Ile Gly Asp Leu Val Glu Val Met Asp Arg Ile Thr Asp Asp Glu 210 215 220		
	Val Asp Thr Leu Tyr Ala Glu Tyr Asp Arg Leu Tyr Val Ile Ser Glu 225 230 235 240		
	Glu Thr Lys Arg Asp Glu Ala Lys Val Ala Ser Ile Lys Glu Gln Ala 245 250 255		
	Lys Ile Glu Leu Gly Leu Thr Ala Phe Leu Glu Gln Gly Gly Tyr Thr 260 265 270		
	Ala Phe Thr Thr Ser Phe Glu Val Leu His Gly Met Lys Gln Leu Pro 275 280 285		
	Gly Leu Ala Val Gln Arg Leu Met Glu Lys Gly Tyr Gly Phe Ala Gly 290 295 300		
	Glu Gly Asp Trp Lys Thr Ala Ala Leu Val Arg Met Met Lys Ile Met 305 310 315 320		
[0012]	Ala Lys Gly Lys Arg Thr Ser Phe Met Glu Asp Tyr Thr Tyr His Phe 325 330 335		
	Glu Pro Gly Asn Glu Met Ile Leu Gly Ser His Met Leu Glu Val Cys 340 345 350		
	Pro Thr Val Ala Leu Asp Gln Pro Lys Ile Glu Val His Ser Leu Ser 355 360 365		
	Ile Gly Gly Lys Glu Asp Pro Ala Arg Leu Val Phe Asn Gly Ile Ser 370 375 380		
	Gly Ser Ala Ile Gln Ala Ser Ile Val Asp Ile Gly Gly Arg Phe Arg 385 390 395 400		
	Leu Val Leu Asn Glu Val Asn Gly Gln Glu Ile Glu Lys Asp Met Pro 405 410 415		
	Asn Leu Pro Val Ala Arg Val Leu Trp Lys Pro Glu Pro Ser Leu Lys 420 425 430		
	Thr Ala Ala Glu Ala Trp Ile Leu Ala Gly Gly Ala His His Thr Cys 435 440 445		
	Leu Ser Tyr Glu Leu Thr Ala Glu Gln Met Leu Asp Trp Ala Glu Met		

	450	455	460	
	Ala Gly Ile Glu Ser Val Leu Ile Ser Arg Asp Thr Thr Ile His Lys			
	465	470	475	480
	Leu Lys His Glu Leu Lys Trp Asn Glu Ala Leu Tyr Arg Leu Gln Lys			
		485	490	495
<210>	10			
<211>	1511			
<212>	DNA			
<213>	Bacillus subtilis			
<400>	10			
[0013]	atgagaaagg ggcagtttac atgcttcaga caaaggatta tgaattctgg tttgtgacag			60
	gaagccagca cctatacggg gaagagacgc tgggaactcgt agatcagcat gctaaaagca			120
	tttgtgaggg gctcagcggg atttcttcca gatataaaat cactcataag cccgtcgtca			180
	cttcaccgga aaccattaga gagctgttaa gagaagcggg gtacagttag acatgtgctg			240
	gcatcattac atggatgcac acattttccc ctgcaaaaat gtggatagaa ggcctttcct			300
	cttatcaaaa accgcttatg catttgcata cccaatataa tcgcatatc ccgtggggta			360
	cgattgacat ggattttatg aacagcaacc aatccgcgca tggcgatcga gagtacggtt			420
	acatcaactc gagaatgggg cttagccgaa aagtcattgc cggtattgg gatgatgaag			480
	aagtgaaaaa agaaatgtcc cagtggatgg atacggcggc tgcattaaat gaaagcagac			540
	atattaaggt tgccagatgt ggagataaca tgcgtcatgt cgcggaacg gacggagaca			600
	aggtgggagc gcatattcaa tttggctggc aggttgacgg atatggcatc ggggatctcg			660
	ttgaagtgat ggatcgcatt acggacgacg aggttgacac gctttatgcc gagtatgaca			720
	gactatatgt gatcagttag gaaacaaaac gtgacgaagc aaaggtagcg tccattaaag			780
	aacaggcgaa aattgaactt ggattaaccg cttttcttga gcaaggcggg tacacagcgt			840
	ttacgacatc gtttgaagtg ctgcacggaa tgaaacagct gccgggactt gccgttcagc			900
	gcctgatgga gaaaggctat gggtttgccg gtgaaggaga ttggaagaca gcggcccttg			960
	tacggatgat gaaaatcatg gctaaaggaa aaagaacttc cttcatggaa gattacacgt			1020
	accattttga accgggaaat gaaatgattc tgggctctca catgcttgaa gtgtgtccga			1080
	ctgtcgcttt ggatcagccg aaaatcgagg ttcatcgtct ttcgattggc ggcaagagg			1140
	accctgcgcg tttggtatgt aacggcatca gcggttctgc cattcaagct agcattgttg			1200
	atattggcgg gcgtttccgc cttgtgctga atgaagtcaa cggccaggaa attgaaaag			1260
	acatgccgaa ttaccgggtt gcccgtgttc tctggaagcc ggagccgtca ttgaaaacag			1320
	cagcggaggg atggatttta gccgcggtg cacaccatac ctgcctgtct tatgaactga			1380
	cagcggagca aatgcttgat tggcggaata tggcggaat cgaaagtgtt ctcatctccc			1440
	gtgatacgac aattcataaa ctgaaacacg agttaaaatg gaacgaggcg ctttaccggc			1500
	ttcaaaagta g			1511

<210> 11
 <211> 566
 <212> PRT
 <213> E. coli

<400> 11

Met Ala Ile Ala Ile Gly Leu Asp Phe Gly Ser Asp Ser Val Arg Ala
 1 5 10 15

Leu Ala Val Asp Cys Ala Ser Gly Glu Glu Ile Ala Thr Ser Val Glu
 20 25 30

Trp Tyr Pro Arg Trp Gln Lys Gly Gln Phe Cys Asp Ala Pro Asn Asn
 35 40 45

Gln Phe Arg His His Pro Arg Asp Tyr Ile Glu Ser Met Glu Ala Ala
 50 55 60

Leu Lys Thr Val Leu Ala Glu Leu Ser Val Glu Gln Arg Ala Ala Val
 65 70 75 80

Val Gly Ile Gly Val Asp Ser Thr Gly Ser Thr Pro Ala Pro Ile Asp
 85 90 95

[0014] Ala Asp Gly Asn Val Leu Ala Leu Arg Pro Glu Phe Ala Glu Asn Pro
 100 105 110

Asn Ala Met Phe Val Leu Trp Lys Asp His Thr Ala Val Glu Arg Ser
 115 120 125

Glu Glu Ile Thr Arg Leu Cys His Ala Pro Gly Asn Val Asp Tyr Ser
 130 135 140

Arg Tyr Ile Gly Gly Ile Tyr Ser Ser Glu Trp Phe Trp Ala Lys Ile
 145 150 155 160

Leu His Val Thr Arg Gln Asp Ser Ala Val Ala Gln Ser Ala Ala Ser
 165 170 175

Trp Ile Glu Leu Cys Asp Trp Val Pro Ala Leu Leu Ser Gly Thr Thr
 180 185 190

Arg Pro Gln Asp Ile Arg Arg Gly Arg Cys Ser Ala Gly His Lys Ser
 195 200 205

Leu Trp His Glu Ser Trp Gly Gly Leu Pro Pro Ala Ser Phe Phe Asp
 210 215 220

Glu Leu Asp Pro Ile Leu Asn Arg His Leu Pro Ser Pro Leu Phe Thr
 225 230 235 240

Asp Thr Trp Thr Ala Asp Ile Pro Val Gly Thr Leu Cys Pro Glu Trp
 245 250 255
 Ala Gln Arg Leu Gly Leu Pro Glu Ser Val Val Ile Ser Gly Gly Ala
 260 265 270
 Phe Asp Cys His Met Gly Ala Val Gly Ala Gly Ala Gln Pro Asn Ala
 275 280 285
 Leu Val Lys Val Ile Gly Thr Ser Thr Cys Asp Ile Leu Ile Ala Asp
 290 295 300
 Lys Gln Ser Val Gly Glu Arg Ala Val Lys Gly Ile Cys Gly Gln Val
 305 310 315 320
 Asp Gly Ser Val Val Pro Gly Phe Ile Gly Leu Glu Ala Gly Gln Ser
 325 330 335
 Ala Phe Gly Asp Ile Tyr Ala Trp Phe Gly Arg Val Leu Ser Trp Pro
 340 345 350
 Leu Glu Gln Leu Ala Ala Gln His Pro Glu Leu Lys Ala Gln Ile Asn
 355 360 365
 [0015] Ala Ser Gln Lys Gln Leu Leu Pro Ala Leu Thr Glu Ala Trp Ala Lys
 370 375 380
 Asn Pro Ser Leu Asp His Leu Pro Val Val Leu Asp Trp Phe Asn Gly
 385 390 395 400
 Arg Arg Ser Pro Asn Ala Asn Gln Arg Leu Lys Gly Val Ile Thr Asp
 405 410 415
 Leu Asn Leu Ala Thr Asp Ala Pro Leu Leu Phe Gly Gly Leu Ile Ala
 420 425 430
 Ala Thr Ala Phe Gly Ala Arg Ala Ile Met Glu Cys Phe Thr Asp Gln
 435 440 445
 Gly Ile Ala Val Asn Asn Val Met Ala Leu Gly Gly Ile Ala Arg Lys
 450 455 460
 Asn Gln Val Ile Met Gln Ala Cys Cys Asp Val Leu Asn Arg Pro Leu
 465 470 475 480
 Gln Ile Val Ala Ser Asp Gln Cys Cys Ala Leu Gly Ala Ala Ile Phe
 485 490 495
 Ala Ala Val Ala Ala Lys Val His Ala Asp Ile Pro Ser Ala Gln Gln
 500 505 510

Lys Met Ala Ser Ala Val Glu Lys Thr Leu Gln Pro Arg Ser Glu Gln
515 520 525

Ala Gln Arg Phe Glu Gln Leu Tyr Arg Arg Tyr Gln Gln Trp Ala Met
530 535 540

Ser Ala Glu Gln His Tyr Leu Pro Thr Ser Ala Pro Ala Gln Ala Ala
545 550 555 560

Gln Ala Val Ala Thr Leu
565

<210> 12
<211> 1453
<212> DNA
<213> E. coli

<400> 12
atggcgattg caattggcct cgattttggc agtgattctg tgcgagcttt ggcggtggac 60
tgcgccagcg gtgaagagat cgccaccagc gtagagtggc atccccgttg gcaaaaaggg 120
caattttgtg atgccccgaa taaccagttc cgtcatcatc cgcgtgacta cattgagtca 180
atggaagcgg cactgaaaac cgtgcttgca gagcttagcg tcgaacagcg cgcagctgtg 240
gtcgggattg gcgttgacag taccggctcg acgccgcac cgattgatgc cgacggtaac 300
gtgctggcgc tgcgcccgga gtttgccgaa aaccggaacg cgatgttcgt attgtggaaa 360
gaccacactg cggttgaaag aagcgaagag attaccggtt tgtgccacgc gccgggcaat 420
gttgactact cccgctatat tggcgggtatt tattccagcg aatggttctg ggcaaaaatc 480
ctgcatgtga ctgcgcagga cagcgcctg gcgcaatctg ccgcatcgtg gattgagctg 540
tgcgactggg tgccagctct gctttccggc accaccgcc cgcaggatat tcgtcgcgga 600
cgttgcagcg ccgggcataa atctctgtgg cagcaaagct ggggcggcct gccgccagcc 660
agtttctttg atgagctgga cccgatcctc aatcgccatt tgccttcccc gctgttcact 720
gacacctgga ctgccgatat tccggtgggc accttatgcc cggaatgggc gcagcgtctc 780
ggcctgcctg aaagcgtggc gatttccggc ggcgcgtttg actgccatat gggcgagtt 840
ggcgagggcg cacagcctaa cgcactggta aaagtattcg gtacttcac ctgcgacatt 900
ctgattgccg acaaacagag cgttggcgag cgggcagtta aaggtatttg cggtcaggtt 960
gatggcagcg tggcgcctgg atttatcggc ctggaagcag gccaatcggc gtttggtgat 1020
atctacgcct gggtcggctg cgtactcagc tggccgctgg aacagcttgc cgccagcat 1080
ccggaactga aagcgcaaat caacgccagc cagaaacaac tgcttccggc gctgaccgaa 1140
gcatgggcca aaaatccgtc tctggatcac ctgccgggtg tgctcgactg gtttaacggc 1200
cgtcgctcgc caaacgctaa ccaacgcctg aaaggggtga ttaccgatct taacctcgct 1260
accgacgctc cgtgctgtt cggcgggttg attgctgcca ccgcctttgg cgcacgcgca 1320
atcatggagt gctttaccga tcaggggatc gccgtcaata acgtgatggc gctgggcggc 1380

[0016]

```

atcgcgcgga aaaaccaagt cattatgcag gcctgctgcg acgtgctgaa tcgcccgtg 1440
caaattgttg cct 1453

<210> 13
<211> 231
<212> PRT
<213> E. coli
<400> 13

Met Leu Glu Asp Leu Lys Arg Gln Val Leu Glu Ala Asn Leu Ala Leu
1 5 10 15

Pro Lys His Asn Leu Val Thr Leu Thr Trp Gly Asn Val Ser Ala Val
20 25 30

Asp Arg Glu Arg Gly Val Phe Val Ile Lys Pro Ser Gly Val Asp Tyr
35 40 45

Ser Ile Met Thr Ala Asp Asp Met Val Val Val Ser Ile Glu Thr Gly
50 55 60

Glu Val Val Glu Gly Ala Lys Lys Pro Ser Ser Asp Thr Pro Thr His
65 70 75 80

[0017] Arg Leu Leu Tyr Gln Ala Phe Pro Ser Ile Gly Gly Ile Val His Thr
85 90 95

His Ser Arg His Ala Thr Ile Trp Ala Gln Ala Gly Gln Ser Ile Pro
100 105 110

Ala Thr Gly Thr Thr His Ala Asp Tyr Phe Tyr Gly Thr Ile Pro Cys
115 120 125

Thr Arg Lys Met Thr Asp Ala Glu Ile Asn Gly Glu Tyr Glu Trp Glu
130 135 140

Thr Gly Asn Val Ile Val Glu Thr Phe Glu Lys Gln Gly Ile Asp Ala
145 150 155 160

Ala Gln Met Pro Gly Val Leu Val His Ser His Gly Pro Phe Ala Trp
165 170 175

Gly Lys Asn Ala Glu Asp Ala Val His Asn Ala Ile Val Leu Glu Glu
180 185 190

Val Ala Tyr Met Gly Ile Phe Cys Arg Gln Leu Ala Pro Gln Leu Pro
195 200 205

Asp Met Gln Gln Thr Leu Leu Asn Lys His Tyr Leu Arg Lys His Gly
210 215 220

```

Ala Lys Ala Tyr Tyr Gly Gln
225 230

<210> 14
<211> 696
<212> DNA
<213> E coli

<400> 14
atgttagaag atctcaaacg ccaggtatta gaggcccaacc tggcgctgcc aaaacataac 60
ctgggtcacgc tcacatgggg caacgtcagc gccgttgatc gcgagcgcg cgctctttgtg 120
atcaaacctt ccggcgctcga ttacagcatc atgaccgctg acgatattggt cgtgggttagc 180
atcgaaaccg gtgaagtggg tgaagggtgcg aaaaagccct cctccgatac gccaaactcac 240
cgactgctct atcaggcatt cccgtccatt ggcggcattg tgcacacaca ctcgcgccac 300
gccactatct gggcgagcggc gggccagtcg attccagcaa ccggcaccac ccacgccgac 360
tattttctacg gcaccattcc ctgcacccgc aaaatgaccg acgcagaaat caacgggtgaa 420
tatgagtggg aaaccggtaa cgtcacgta gaaaccttcg aaaaacaggg tatcgatgca 480
gcgcgaatgc ccggcgctcct ggtccattct cacggcccat ttgcatgggg caaaaatgcc 540
gaagatgcgg tgcataacgc catcgtgctg gaagaggctg cttatatggg gatattctgc 600
cgtcagttag cgccgcagtt accggatatg cagcaaagcg tgctgaataa acactatctg 660
cgtaagcatg gcgcgaaggc atattacggg cagtaa 696

[0018]

<210> 15
<211> 438
<212> PRT
<213> Bacteroides thetaiotaomicron

<400> 15
Met Ala Thr Lys Glu Phe Phe Pro Gly Ile Glu Lys Ile Lys Phe Glu
1 5 10 15
Gly Lys Asp Ser Lys Asn Pro Met Ala Phe Arg Tyr Tyr Asp Ala Glu
20 25 30
Lys Val Ile Asn Gly Lys Lys Met Lys Asp Trp Leu Arg Phe Ala Met
35 40 45
Ala Trp Trp His Thr Leu Cys Ala Glu Gly Gly Asp Gln Phe Gly Gly
50 55 60
Gly Thr Lys Gln Phe Pro Trp Asn Gly Asn Ala Asp Ala Ile Gln Ala
65 70 75 80
Ala Lys Asp Lys Met Asp Ala Gly Phe Glu Phe Met Gln Lys Met Gly
85 90 95
Ile Glu Tyr Tyr Cys Phe His Asp Val Asp Leu Val Ser Glu Gly Ala

	100	105	110
	Ser Val Glu Glu Tyr Glu Ala Asn Leu Lys Glu Ile Val Ala Tyr Ala 115 120 125		
	Lys Gln Lys Gln Ala Glu Thr Gly Ile Lys Leu Leu Trp Gly Thr Ala 130 135 140		
	Asn Val Phe Gly His Ala Arg Tyr Met Asn Gly Ala Ala Thr Asn Pro 145 150 155 160		
	Asp Phe Asp Val Val Ala Arg Ala Ala Val Gln Ile Lys Asn Ala Ile 165 170 175		
	Asp Ala Thr Ile Glu Leu Gly Gly Glu Asn Tyr Val Phe Trp Gly Gly 180 185 190		
	Arg Glu Gly Tyr Met Ser Leu Leu Asn Thr Asp Gln Lys Arg Glu Lys 195 200 205		
	Glu His Leu Ala Gln Met Leu Thr Ile Ala Arg Asp Tyr Ala Arg Ala 210 215 220		
	Arg Gly Phe Lys Gly Thr Phe Leu Ile Glu Pro Lys Pro Met Glu Pro 225 230 235 240		
[0019]	Thr Lys His Gln Tyr Asp Val Asp Thr Glu Thr Val Ile Gly Phe Leu 245 250 255		
	Lys Ala His Gly Leu Asp Lys Asp Phe Lys Val Asn Ile Glu Val Asn 260 265 270		
	His Ala Thr Leu Ala Gly His Thr Phe Glu His Glu Leu Ala Val Ala 275 280 285		
	Val Asp Asn Gly Met Leu Gly Ser Ile Asp Ala Asn Arg Gly Asp Tyr 290 295 300		
	Gln Asn Gly Trp Asp Thr Asp Gln Phe Pro Ile Asp Asn Tyr Glu Leu 305 310 315 320		
	Thr Gln Ala Met Met Gln Ile Ile Arg Asn Gly Gly Leu Gly Thr Gly 325 330 335		
	Gly Thr Asn Phe Asp Ala Lys Thr Arg Arg Asn Ser Thr Asp Leu Glu 340 345 350		
	Asp Ile Phe Ile Ala His Ile Ala Gly Met Asp Ala Met Ala Arg Ala 355 360 365		
	Leu Glu Ser Ala Ala Ala Leu Leu Asp Glu Ser Pro Tyr Lys Lys Met		

	370	375	380	
	Leu Ala Asp Arg Tyr Ala Ser Phe Asp Gly Gly Lys Gly Lys Glu Phe 385 390 395 400			
	Glu Asp Gly Lys Leu Thr Leu Glu Asp Val Val Ala Tyr Ala Lys Thr 405 410 415			
	Lys Gly Glu Pro Lys Gln Thr Ser Gly Lys Gln Glu Leu Tyr Glu Ala 420 425 430			
	Ile Leu Asn Met Tyr Cys 435			
	<210> 16			
	<211> 1317			
	<212> DNA			
	<213> Bacteroides thetaiotaomicron			
	<400> 16			
	atggcaacaa aagaattttt tccggaatt gaaaagatta aatttgaagg taaagatagt 60			
	aagaacccga tggcattccg ttattacgat gcagagaagg tgattaatgg taaaaagatg 120			
	aaggattggc tgagattcgc tatggcatgg tggcacacat tgtgcgctga aggtggtgat 180			
	cagttcggtg gcggaacaaa gcaattccca tggaatggta atgcagatgc tatacaggca 240			
[0020]	gcaaaagata agatggatgc aggatttgaa ttcattgcaga agatgggtat cgaatactat 300			
	tgcttccatg acgtagactt ggtttcggaa ggtgccagtg tagaagaata cgaagctaac 360			
	ctgaaagaaa tcgtagctta tgcaaaacag aaacaggcag aaaccggtat caaactactg 420			
	tggggtactg ctaatgtatt cggtcacgcc cgctatatga acggtgcagc taccaatcct 480			
	gacttcgatg tagtagctcg tgctgctgtt cagatcaaaa atgcgattga tgcaacgatt 540			
	gaacttggcg gagagaatta tgtgttttgg ggtggtcgtg aaggctatat gtctcttctg 600			
	aacacagatc agaaacgtga aaaagaacac cttgcacaga tgttgacgat tgctcgtgac 660			
	tatgcccgtg cccgtgggtt caaagggtact ttcctgatcg aaccgaaacc gatggaaccg 720			
	actaaacatc aatatgacgt agatacggaa actgtaatcg gcttcctgaa agctcatggt 780			
	ctggataaagg atttcaaagt aaatatcgag gtgaatcacg caactttggc aggtcacact 840			
	ttcagcatg aattggctgt agctgtagac aatggtatgt tgggctcaat tgacgccaat 900			
	cgtggtgact atcagaatgg ctgggataca gaccaattcc cgatcgacaa ttatgaactg 960			
	actcaggcta tgatgcagat tatccgtaat ggtggtctcg gtaccggtgg tacgaacttt 1020			
	gatgctaaaa cccgtcgtaa ttctactgat ctggaagata tctttattgc tcacatcgca 1080			
	ggtatggacg ctatggcccg tgactcgaag agtcagcgg ctctgctcga cgaatctccc 1140			
	tataagaaga tgctggctga ccgttatgct tcatttgatg ggggcaaagg taaagaattt 1200			
	gaagacggca agctgactct ggaggatgtg gttgcttatg caaaaacaaa aggcgaaccg 1260			
	aaacagacta gcggcaagca agaactttat gagggcaattc tgaatatgta ttgctaa 1317			

<210> 17
 <211> 258
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 17

Met Ala Ala Gly Val Pro Lys Ile Asp Ala Leu Glu Ser Leu Gly Asn
 1 5 10 15

Pro Leu Glu Asp Ala Lys Arg Ala Ala Tyr Arg Ala Val Asp Glu
 20 25 30

Asn Leu Lys Phe Asp Asp His Lys Ile Ile Gly Ile Gly Ser Gly Ser
 35 40 45

Thr Val Val Tyr Val Ala Glu Arg Ile Gly Gln Tyr Leu His Asp Pro
 50 55 60

Lys Phe Tyr Glu Val Ala Ser Lys Phe Ile Cys Ile Pro Thr Gly Phe
 65 70 75 80

Gln Ser Arg Asn Leu Ile Leu Asp Asn Lys Leu Gln Leu Gly Ser Ile
 85 90 95

[0021] Glu Gln Tyr Pro Arg Ile Asp Ile Ala Phe Asp Gly Ala Asp Glu Val
 100 105 110

Asp Glu Asn Leu Gln Leu Ile Lys Gly Gly Gly Ala Cys Leu Phe Gln
 115 120 125

Glu Lys Leu Val Ser Thr Ser Ala Lys Thr Phe Ile Val Val Ala Asp
 130 135 140

Ser Arg Lys Lys Ser Pro Lys His Leu Gly Lys Asn Trp Arg Gln Gly
 145 150 155 160

Val Pro Ile Glu Ile Val Pro Ser Ser Tyr Val Arg Val Lys Asn Asp
 165 170 175

Leu Leu Glu Gln Leu His Ala Glu Lys Val Asp Ile Arg Gln Gly Gly
 180 185 190

Ser Ala Lys Ala Gly Pro Val Val Thr Asp Asn Asn Asn Phe Ile Ile
 195 200 205

Asp Ala Asp Phe Gly Glu Ile Ser Asp Pro Arg Lys Leu His Arg Glu
 210 215 220

Ile Lys Leu Leu Val Gly Val Val Glu Thr Gly Leu Phe Ile Asp Asn
 225 230 235 240

Ala Ser Lys Ala Tyr Phe Gly Asn Ser Asp Gly Ser Val Glu Val Thr
 245 250 255

Glu Lys

<210> 18
 <211> 2467
 <212> DNA
 <213> *Saccharomyces cerevisiae*

[0022] <400> 18
 ggatccaaga ccattattcc atcagaatgg aaaaaagttt aaaagatcac ggagattttg 60
 ttctttctgag cttctgctgt ccttgaaaac aaattattcc gctggccgcc ccaaacaaaa 120
 acaaccccgga ttttaataaca ttgtcacagt attagaaatt ttctttttac aaattaccat 180
 ttccagctta ctacttccta taatcctcaa tcttcagcaa gcgacgcagg gaatagccgc 240
 tgagggtgcat aactgtcact tttcaattcg gccaatgcaa tctcaggcgg acgaataagg 300
 gggccctctc gagaaaaaca aaaggaggat gagattagta ctttaatggt gtgttcagta 360
 attcagagac agacaagaga ggtttccaac acaatgtctt tagactcata ctatcttggg 420
 tttgatcttt cgaccaaca actgaaatgt ctcgccatta accaggacct aaaaattgtc 480
 cattcagaaa cagtggaatt tgaaaaggat cttccgcatt atcacacaaa gaagggtgtc 540
 tatatacacg gcgacactat cgaatgtccc gtagccatgt ggtagggggc tctagatctg 600
 gttctctcga aatatcgca ggctaaattt ccattgaaca aagttatggc cgtctcaggg 660
 tcctgccagc agcacgggtc tgtctactgg tcctcccaag ccgaatctct gttagagcaa 720
 ttgaataaga aaccggaaaa agatttattg cactacgtga gctctgtagc atttgcaagg 780
 caaacgccc ccaattggca agaccacagt actgcaaagc aatgtcaaga gtttgaagag 840
 tgcatagggt ggccctgaaaa aatggctcaa ttaacagggt ccagagccca ttttagattt 900
 actggtcctc aaattctgaa aattgcacaa ttagaaccag aagcttacga aaaaacaaag 960
 accatttctt tagtgtctaa ttttttgact tctatcttag tgggccatct tgttgaatta 1020
 gaggaggcag atgcctgtgg tatgaacctt tatgatatac gtgaaagaaa attcatgtat 1080
 gagctactac atctaattga tagttcttct aaggataaaa ctatcagaca aaaattaatg 1140
 agagcaccca tgaaaaattt gatagcgggt accatctgta aatattttat tgagaagtac 1200
 ggtttcaata caaactgcaa ggtctctccc atgactgggg ataatttagc cactatatgt 1260
 tctttacccc tgcggaagaa tgacgttctc gtttccctag gaacaagtac tacagttctt 1320
 ctgggtcaccg ataagtatca cccctctccg aactatcatc ttttcattca tccaactctg 1380
 ccaaaccatt atatgggtat gatttggtat tgtaatggtt ctttggcaag ggagaggata 1440
 agagacgagt taaacaaaga acgggaaaat aattatgaga agactaacga ttggactctt 1500
 tttaatcaag ctgtgctaga tgactcagaa agtagtgaag atgaattagg tgtatatattt 1560
 cctctggggg agatcgttcc tagcgtaaaa gccataaaca aaagggttat cttcaatcca 1620

```

aaaacgggta tgattgaaag agaggtggcc aagttcaaag acaagaggca cgatgccaaa 1680
aatattgtag aatcacaggc ttttaagttgc agggtaagaa tatctcccct gctttcggat 1740
tcaaacgcaa gtcacaaca gagactgaac gaagatacaa tcgtgaagtt tgattacgat 1800
gaatctccgc tgcgggacta cctaaataaa aggccagaaa ggactttttt tgtagggtggg 1860
gcttctaaaa acgatgctat tgtgaagaag tttgctcaag tcattggtgc tacaaagggg 1920
aatttttaggc tagaaacacc aaactcatgt gcccttggtg gttgttataa ggccatgtgg 1980
tcattgttat atgactctaa taaaattgca gttccttttg ataaatttct gaatgacaat 2040
tttccatggc atgtaatgga aagcatatcc gatgtggata atgaaaattg gatcgctata 2100
attccaagat tgtcccctta agcgaactgg aaaagactct catctaaaat atgtttgaat 2160
aatttatcat gccctgacaa gtacacacaa acacagacac ataataatac tacatatata 2220
tatatcaccg ttattatgcg tgcacatgac aatgcccttg tatgtttcgt atactgtagc 2280
aagtagtcat cattttgttc cccgttcgga aaatgacaaa aagtaaaatc aataaatgaa 2340
gagtaaaaaa caatttatga aaggggtgagc gaccagcaac gagagagaca aatcaaatta . 2400
gcgctttcca gtgagaatat aagagagcat tgaaagagct aggttattgt taaatcatct 2460
cgagctc 2467

```

[0023]

```

<210> 19
<211> 238
<212> PRT
<213> Saccharomyces cerevisiae

```

```

<400> 19

```

```

Met Val Lys Pro Ile Ile Ala Pro Ser Ile Leu Ala Ser Asp Phe Ala
1           5           10           15

```

```

Asn Leu Gly Cys Glu Cys His Lys Val Ile Asn Ala Gly Ala Asp Trp
          20           25           30

```

```

Leu His Ile Asp Val Met Asp Gly His Phe Val Pro Asn Ile Thr Leu
          35           40           45

```

```

Gly Gln Pro Ile Val Thr Ser Leu Arg Arg Ser Val Pro Arg Pro Gly
50           55           60

```

```

Asp Ala Ser Asn Thr Glu Lys Lys Pro Thr Ala Phe Phe Asp Cys His
65           70           75           80

```

```

Met Met Val Glu Asn Pro Glu Lys Trp Val Asp Asp Phe Ala Lys Cys
          85           90           95

```

```

Gly Ala Asp Gln Phe Thr Phe His Tyr Glu Ala Thr Gln Asp Pro Leu
100           105           110

```

```

His Leu Val Lys Leu Ile Lys Ser Lys Gly Ile Lys Ala Ala Cys Ala

```

	115	120	125	
	Ile Lys Pro Gly Thr Ser Val Asp Val Leu Phe Glu Leu Ala Pro His 130 135 140			
	Leu Asp Met Ala Leu Val Met Thr Val Glu Pro Gly Phe Gly Gly Gln 145 150 155 160			
	Lys Phe Met Glu Asp Met Met Pro Lys Val Glu Thr Leu Arg Ala Lys 165 170 175			
	Phe Pro His Leu Asn Ile Gln Val Asp Gly Gly Leu Gly Lys Glu Thr 180 185 190			
	Ile Pro Lys Ala Ala Lys Ala Gly Ala Asn Val Ile Val Ala Gly Thr 195 200 205			
	Ser Val Phe Thr Ala Ala Asp Pro His Asp Val Ile Ser Phe Met Lys 210 215 220			
	Glu Glu Val Ser Lys Glu Leu Arg Ser Arg Asp Leu Leu Asp 225 230 235			
[0024]	<210> 20			
	<211> 1328			
	<212> DNA			
	<213> Saccharomyces cerevisiae			
	<400> 20			
	gttaggcact tacgtatctt gtatagtagg aatggctcgg tttatgtata ttaggagatc	60		
	aaaacgagaa aaaaatacca tatcgtatag tatagagagt ataaatataa gaaatgccgc	120		
	atatgtacaa ctaatctagc aaatctctag aacgcaattc cttcgagact tcttctttca	180		
	tgaaggagat aacatcgtgc gggtcagctg cagtga aaac actggtacca gcgacaataa	240		
	cgttggcacc ggctttggcg gctttcggga tggctctcctt gcccaaacca ccacgcactt	300		
	ggatattcaa atgggggaac ttggctctca aagtttccac ttttggcatc atgtctttcca	360		
	tgaatttttg gcctccaaac ccagggtcca cagtcataac aagagccata tccaaatgag	420		
	gagctagttc aaataaaacg tcaacagaag taccagggtt gatggcgcat gcagctttga	480		
	tgcccttaga cttaatcaac ttaactaaat gcaaagggtc ttgtgtggcc tcgtagtgga	540		
	acgtaaattg gtcagcacca catttagcaa aatcgtcgac ccatttttca ggattttcaa	600		
	ccatcatgtg acaatcgaag aacgcagtgg gcttcttttc tgtgttgcta gcatcgccag	660		
	ggcgtggcac agaacgacgt agggaggtaa caattggttg gccagagta atgtttggaa	720		
	caaaatggcc gtccatgaca tcgatatgta accaatctgc gccggcgttg atgaccttat	780		
	gacattcgca acccaagttg gcgaagtcag aagcaaggat actgggagct ataattgggtt	840		
	tgaccatttt ttcttgtgtg ttacctcgc tcttggaatt agcaaaggc cttcttgcac	900		
	gaaattgtat cgagtttgct ttatttttct ttttacgggc ggattctttc tattctggct	960		

```

ttcctataac agagatcatg aaagaagttc cagcttacgg atcaagaaag tacctataca 1020
tatacaaaaa tctgattact ttcccagctc gacttggata gctgttcttg ttttctcttg 1080
gcgacacatt ttttgtttct gaagccacgt cctgctttat aagaggacat ttaaagttgc 1140
aggacttgaa tgcaattacc ggaagaagca accaaccggc atggttcagc atacaataca 1200
catttgatta gaaaagcaga gaataaatag acatgatacc tctcttttta tcctctgcag 1260
cgtattattg tttattccac gcaggcatcg gtcgttggct gttgttatgt ctcagataag 1320
cgcgtttg 1328

```

```

<210> 21
<211> 680
<212> PRT
<213> Saccharomyces cerevisiae

```

```
<400> 21
```

```
Met Thr Gln Phe Thr Asp Ile Asp Lys Leu Ala Val Ser Thr Ile Arg
1 5 10 15
```

```
Ile Leu Ala Val Asp Thr Val Ser Lys Ala Asn Ser Gly His Pro Gly
20 25 30
```

```
Ala Pro Leu Gly Met Ala Pro Ala Ala His Val Leu Trp Ser Gln Met
35 40 45
```

[0025]

```
Arg Met Asn Pro Thr Asn Pro Asp Trp Ile Asn Arg Asp Arg Phe Val
50 55 60
```

```
Leu Ser Asn Gly His Ala Val Ala Leu Leu Tyr Ser Met Leu His Leu
65 70 75 80
```

```
Thr Gly Tyr Asp Leu Ser Ile Glu Asp Leu Lys Gln Phe Arg Gln Leu
85 90 95
```

```
Gly Ser Arg Thr Pro Gly His Pro Glu Phe Glu Leu Pro Gly Val Glu
100 105 110
```

```
Val Thr Thr Gly Pro Leu Gly Gln Gly Ile Ser Asn Ala Val Gly Met
115 120 125
```

```
Ala Met Ala Gln Ala Asn Leu Ala Ala Thr Tyr Asn Lys Pro Gly Phe
130 135 140
```

```
Thr Leu Ser Asp Asn Tyr Thr Tyr Val Phe Leu Gly Asp Gly Cys Leu
145 150 155 160
```

```
Gln Glu Gly Ile Ser Ser Glu Ala Ser Ser Leu Ala Gly His Leu Lys
165 170 175
```

```
Leu Gly Asn Leu Ile Ala Ile Tyr Asp Asp Asn Lys Ile Thr Ile Asp

```

	180	185	190
	Gly Ala Thr Ser Ile Ser Phe Asp Glu Asp Val Ala Lys Arg Tyr Glu 195 200 205		
	Ala Tyr Gly Trp Glu Val Leu Tyr Val Glu Asn Gly Asn Glu Asp Leu 210 215 220		
	Ala Gly Ile Ala Lys Ala Ile Ala Gln Ala Lys Leu Ser Lys Asp Lys 225 230 235 240		
	Pro Thr Leu Ile Lys Met Thr Thr Thr Ile Gly Tyr Gly Ser Leu His 245 250 255		
	Ala Gly Ser His Ser Val His Gly Ala Pro Leu Lys Ala Asp Asp Val 260 265 270		
	Lys Gln Leu Lys Ser Lys Phe Gly Phe Asn Pro Asp Lys Ser Phe Val 275 280 285		
	Val Pro Gln Glu Val Tyr Asp His Tyr Gln Lys Thr Ile Leu Lys Pro 290 295 300		
	Gly Val Glu Ala Asn Asn Lys Trp Asn Lys Leu Phe Ser Glu Tyr Gln 305 310 315 320		
[0026]	Lys Lys Phe Pro Glu Leu Gly Ala Glu Leu Ala Arg Arg Leu Ser Gly 325 330 335		
	Gln Leu Pro Ala Asn Trp Glu Ser Lys Leu Pro Thr Tyr Thr Ala Lys 340 345 350		
	Asp Ser Ala Val Ala Thr Arg Lys Leu Ser Glu Thr Val Leu Glu Asp 355 360 365		
	Val Tyr Asn Gln Leu Pro Glu Leu Ile Gly Gly Ser Ala Asp Leu Thr 370 375 380		
	Pro Ser Asn Leu Thr Arg Trp Lys Glu Ala Leu Asp Phe Gln Pro Pro 385 390 395 400		
	Ser Ser Gly Ser Gly Asn Tyr Ser Gly Arg Tyr Ile Arg Tyr Gly Ile 405 410 415		
	Arg Glu His Ala Met Gly Ala Ile Met Asn Gly Ile Ser Ala Phe Gly 420 425 430		
	Ala Asn Tyr Lys Pro Tyr Gly Gly Thr Phe Leu Asn Phe Val Ser Tyr 435 440 445		
	Ala Ala Gly Ala Val Arg Leu Ser Ala Leu Ser Gly His Pro Val Ile		

	450	455	460
	Trp Val Ala Thr His Asp Ser Ile Gly Val Gly Glu Asp Gly Pro Thr		
	465	470	475 480
	His Gln Pro Ile Glu Thr Leu Ala His Phe Arg Ser Leu Pro Asn Ile		
		485	490 495
	Gln Val Trp Arg Pro Ala Asp Gly Asn Glu Val Ser Ala Ala Tyr Lys		
		500	505 510
	Asn Ser Leu Glu Ser Lys His Thr Pro Ser Ile Ile Ala Leu Ser Arg		
		515	520 525
	Gln Asn Leu Pro Gln Leu Glu Gly Ser Ser Ile Glu Ser Ala Ser Lys		
		530	535 540
	Gly Gly Tyr Val Leu Gln Asp Val Ala Asn Pro Asp Ile Ile Leu Val		
		545	550 555 560
	Ala Thr Gly Ser Glu Val Ser Leu Ser Val Glu Ala Ala Lys Thr Leu		
		565	570 575
	Ala Ala Lys Asn Ile Lys Ala Arg Val Val Ser Leu Pro Asp Phe Phe		
		580	585 590
[0027]	Thr Phe Asp Lys Gln Pro Leu Glu Tyr Arg Leu Ser Val Leu Pro Asp		
		595	600 605
	Asn Val Pro Ile Met Ser Val Glu Val Leu Ala Thr Thr Cys Trp Gly		
		610	615 620
	Lys Tyr Ala His Gln Ser Phe Gly Ile Asp Arg Phe Gly Ala Ser Gly		
		625	630 635 640
	Lys Ala Pro Glu Val Phe Lys Phe Phe Gly Phe Thr Pro Glu Gly Val		
		645	650 655
	Ala Glu Arg Ala Gln Lys Thr Ile Ala Phe Tyr Lys Gly Asp Lys Leu		
		660	665 670
	Ile Ser Pro Leu Lys Lys Ala Phe		
		675	680
	<210> 22		
	<211> 2046		
	<212> DNA		
	<213> Saccharomyces cerevisiae		
	<400> 22		
	atggcacagt tctccgacat tgataaactt gcggtttcca ctttaagatt actttccgtt		60
	gaccaggtgg aaagcgaca atctggccac ccaggtgcac cactaggatt ggcaccagtt		120

	gcccattgtaa ttttcaagca actgcgctgt aaccctaaca atgaacattg gatcaataga	180
	gacaggtttg ttctgtcgaa cggtcactca tgcgtcttct tgtactcaat gctccatcta	240
	ttaggatacg attactctat cgaggacttg agacaattta gacaagtaaa ctcaaggaca	300
	ccgggtcatc cagaattcca ctacagcggga gtggaaatca cttccgggcc gctaggccag	360
	ggtatctcaa atgctgttgg tatggcaata gcgcaggcca actttgccgc cacttataac	420
	gaggatggct ttcccatttc cgactcatat acgtttgcta ttgtagggga tggttgctta	480
	caagaggggtg tttcttcgga gacctcttcc ttagcgggac atctgcaatt gggtaacttg	540
	attacgtttt atgacagtaa tagcatttcc attgacggta aaacctcgta ctctgttcgac	600
	gaagatgttt tgaagcgata cgaggcatat ggttgggaag tcatggaagt cgataaagga	660
	gacgacgata tggaatccat ttctagcgct ttggaaaagg caaaactatc gaaggacaag	720
	ccaaccataa tcaaggtaac tactacaatt ggatttgggt ccctacaaca ggggtactgct	780
	ggtgttcatg ggtccgcttt gaaggcagat gatgttaaac agttgaagaa gaggtggggg	840
	tttgacccaa ataaatcatt tgtagtacct caagaggtgt acgattatta taagaagact	900
	gttgtggaac ccggtcaaaa acttaatgag gaatgggata ggatgtttga agaatacaaa	960
	accaaatttc ccgagaaggg taaagaattg caaagaagat tgaatggtga gttaccggaa	1020
	ggttgggaaa agcatttacc gaagtttact ccggacgacg atgctctggc aacaagaaag	1080
[0028]	acatcccagc aggtgctgac gaacatggct caagttttgc ctgaattgat cgggtggttct	1140
	gccgatttga caccttcgaa tctgacaagg tgggaaggcg cggtagattt ccaacctccc	1200
	attaccaaac taggtaacta tgcaggaagg tacattagat acggtgtgag ggaacacgga	1260
	atgggtgcca ttatgaacgg tatctctgcc tttggtgcaa actacaagcc ttacggtggt	1320
	acctttttga acttcgtctc ttatgctgca ggagccgtta ggttagccgc cttgtctggt	1380
	aatccagtca tttgggttgc aacacatgac tctatcgggc ttggtgagga tgggtccaacg	1440
	caccaacctt ttgaaactct ggctcacttg agggctattc caaacatgca tgtatggaga	1500
	cctgctgatg gtaacgaaac ttctgctgca tattattctg ctatcaaatac tgggtcgaaca	1560
	ccatctgttg tggctttatc acgacagaat ctctctcaat tggagcattc ctcttttgaa	1620
	aaagccttga aggggtggcta tgtgatccat gacgtggaga atcctgatat tatcctggtg	1680
	tcaacaggat cagaagtctc catttctata gatgcagcca aaaaattgta cgatactaaa	1740
	aaaatcaaag caagagttgt ttccctgcca gacttttata cttttgacag gcaaagtga	1800
	gaatacagat tctctgttct accagacggt gttccgatca tgtcctttga agtattggct	1860
	acttcaagct ggggtaagta tgctcatcaa tcgttcggac tcgacgaatt tgggtcgttca	1920
	ggcaaggggc ctgaaattta caaattgttc gatttcacag cggacggtgt tgcgtcaagg	1980
	gctgaaaaga caatcaatta ctacaaagga aagcagttgc tttctcctat ggggaagagct	2040
	ttctaa	2046

<210> 23
 <211> 335
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 23

Met Ser Glu Pro Ala Gln Lys Lys Gln Lys Val Ala Asn Asn Ser Leu
 1 5 10 15

Glu Gln Leu Lys Ala Ser Gly Thr Val Val Val Ala Asp Thr Gly Asp
 20 25 30

Phe Gly Ser Ile Ala Lys Phe Gln Pro Gln Asp Ser Thr Thr Asn Pro
 35 40 45

Ser Leu Ile Leu Ala Ala Ala Lys Gln Pro Thr Tyr Ala Lys Leu Ile
 50 55 60

Asp Val Ala Val Glu Tyr Gly Lys Lys His Gly Lys Thr Thr Glu Glu
 65 70 75 80

Gln Val Glu Asn Ala Val Asp Arg Leu Leu Val Glu Phe Gly Lys Glu
 85 90 95

Ile Leu Lys Ile Val Pro Gly Arg Val Ser Thr Glu Val Asp Ala Arg
 100 105 110

[0029]

Leu Ser Phe Asp Thr Gln Ala Thr Ile Glu Lys Ala Arg His Ile Ile
 115 120 125

Lys Leu Phe Glu Gln Glu Gly Val Ser Lys Glu Arg Val Leu Ile Lys
 130 135 140

Ile Ala Ser Thr Trp Glu Gly Ile Gln Ala Ala Lys Glu Leu Glu Glu
 145 150 155 160

Lys Asp Gly Ile His Cys Asn Leu Thr Leu Leu Phe Ser Phe Val Gln
 165 170 175

Ala Val Ala Cys Ala Glu Ala Gln Val Thr Leu Ile Ser Pro Phe Val
 180 185 190

Gly Arg Ile Leu Asp Trp Tyr Lys Ser Ser Thr Gly Lys Asp Tyr Lys
 195 200 205

Gly Glu Ala Asp Pro Gly Val Ile Ser Val Lys Lys Ile Tyr Asn Tyr
 210 215 220

Tyr Lys Lys Tyr Gly Tyr Lys Thr Ile Val Met Gly Ala Ser Phe Arg
 225 230 235 240

Ser Thr Asp Glu Ile Lys Asn Leu Ala Gly Val Asp Tyr Leu Thr Ile

	245	250	255	
	Ser Pro Ala Leu Leu Asp Lys Leu Met Asn Ser Thr Glu Pro Phe Pro			
	260	265	270	
	Arg Val Leu Asp Pro Val Ser Ala Lys Lys Glu Ala Gly Asp Lys Ile			
	275	280	285	
	Ser Tyr Ile Ser Asp Glu Ser Lys Phe Arg Phe Asp Leu Asn Glu Asp			
	290	295	300	
	Ala Met Ala Thr Glu Lys Leu Ser Glu Gly Ile Arg Lys Phe Ser Ala			
	305	310	315	320
	Asp Ile Val Thr Leu Phe Asp Leu Ile Glu Lys Lys Val Thr Ala			
	325	330	335	
	<210> 24			
	<211> 2046			
	<212> DNA			
	<213> Saccharomyces cerevisiae			
	<400> 24			
	atggcacagt tctccgacat tgataaactt gcggtttcca ctttaagatt actttccggt			60
	gaccaggttg aaagcgcaca atctggccac ccaggtgcac cactaggatt ggcaccagtt			120
[0030]	gcccattgtaa ttttcaagca actgcgctgt aaccctaaca atgaacattg gatcaataga			180
	gacaggtttg ttctgtcgaa cggtcactca tgcgctcttc tgtactcaat gctccatcta			240
	ttaggatacg attactctat cgaggacttg agacaattta gacaagtaaa ctcaaggaca			300
	ccgggtcatc cagaattcca ctccagcggga gtggaaatca cttccggtcc gctaggccag			360
	ggtatctcaa atgctgtttg tatggcaata ggcagggcca actttgccgc cacttataac			420
	gaggatggct ttcccatttc cgactcatat acgtttgcta ttgtagggga tggttgctta			480
	caagaggggtg tttcttcgga gacctcttcc ttagcgggac atctgcaatt gggttaacttg			540
	attacgtttt atgacagtaa tagcatttcc attgacggta aaacctcgta ctcgttcgac			600
	gaagatgttt tgaagcgata cgaggcatat ggttggaag tcatggaagt cgataaagga			660
	gacgacgata tggaatccat ttctagcgct ttggaaaagg caaaactatc gaaggacaag			720
	ccaaccataa tcaaggtaac tactacaatt ggatttggtt ccctacaaca gggtagtgct			780
	ggtgttcatg ggtccgcttt gaaggcagat gatgttaaag agttgaagaa gaggtggggg			840
	tttgacccaa ataaatcatt ttagtagacct caagaggtgt acgattatta taagaagact			900
	gttggtggaac ccggtcaaaa acttaatgag gaatgggata ggatgtttga agaatacaaa			960
	accaaatttc ccgagaaggg taaagaattg caaagaagat tgaatggtga gttaccggaa			1020
	ggttgggaaa agcattttacc gaagtttact ccggacgacg atgctctggc aacaagaaag			1080
	acatcccagc aggtgctgac gaacatgggt caagttttgc ctgaattgat cggtggttct			1140
	gccgatttga caccttcgaa tctgacaagg tgggaaggcg cggtagattt ccaacctccc			1200

```

attacccaac taggtaacta tgcaggaagg tacattagat acggtgtgag ggaacacgga 1260
atgggtgcc a ttatgaacgg tatctctgcc tttggtgcaa actacaagcc ttacgggtggt 1320
acctttttga acttcgtctc ttatgctgca ggagccgta ggttagccgc cttgtctggt 1380
aatccagtca tttgggttgc aacacatgac tctatcgggc ttggtgagga tggccaacg 1440
caccaacctt ttgaaactct ggctcacttg agggctattc caaacatgca tgtatggaga 1500
cctgctgatg gtaacgaaac ttctgctgcg tattattctg ctatcaaatt tggcgaaca 1560
ccatctgttg tggctttatc acgacagaat ctctctcaat tggagcattc ctcttttgaa 1620
aaagccttga aggggtggcta tgtgatccat gacgtggaga atcctgatat tatcctgggtg 1680
tcaacaggat cagaagtctc catttctata gatgcagcca aaaaattgta cgatactaaa 1740
aaaatcaaag caagagttgt ttccctgcc aacttttata cttttgacag gcaaagtgaa 1800
gaatacagat tctctgttct accagacggg gttccgatca tgtcctttga agtattggct 1860
acttcaagct ggggtaagta tgctcatcaa tcgttcggac tcgacgaatt tggcgttca 1920
ggcaaggggc ctgaaattta caaattgttc gatttcacag cggacggtgt tgcgtcaagg 1980
gctgaaaaga caatcaatta ctacaaagga aagcagttgc tttctctat gggaagagct 2040
ttctaa 2046

```

[0031]

```

<210> 25
<211> 600
<212> PRT
<213> Saccharomyces cerevisiae
<400> 25

```

```

Met Leu Cys Ser Val Ile Gln Arg Gln Thr Arg Glu Val Ser Asn Thr
1           5           10          15

```

```

Met Ser Leu Asp Ser Tyr Tyr Leu Gly Phe Asp Leu Ser Thr Gln Gln
          20          25          30

```

```

Leu Lys Cys Leu Ala Ile Asn Gln Asp Leu Lys Ile Val His Ser Glu
          35          40          45

```

```

Thr Val Glu Phe Glu Lys Asp Leu Pro His Tyr His Thr Lys Lys Gly
          50          55          60

```

```

Val Tyr Ile His Gly Asp Thr Ile Glu Cys Pro Val Ala Met Trp Leu
65          70          75          80

```

```

Glu Ala Leu Asp Leu Val Leu Ser Lys Tyr Arg Glu Ala Lys Phe Pro
          85          90          95

```

```

Leu Asn Lys Val Met Ala Val Ser Gly Ser Cys Gln Gln His Gly Ser
          100          105          110

```

```

Val Tyr Trp Ser Ser Gln Ala Glu Ser Leu Leu Glu Gln Leu Asn Lys

```

	115	120	125
	Lys Pro Glu Lys Asp Leu	Leu His Tyr Val Ser	Ser Val Ala Phe Ala
	130	135	140
	Arg Gln Thr Ala Pro	Asn Trp Gln Asp His	Ser Thr Ala Lys Gln Cys
	145	150	155 160
	Gln Glu Phe Glu	Glu Cys Ile Gly Gly	Pro Glu Lys Met Ala Gln Leu
		165	170 175
	Thr Gly Ser Arg	Ala His Phe Arg	Phe Thr Gly Pro Gln Ile Leu Lys
		180	185 190
	Ile Ala Gln Leu Glu Pro Glu	Ala Tyr Glu Lys Thr	Lys Thr Ile ser
	195	200	205
	Leu Val Ser Asn Phe Leu	Thr Ser Ile Leu Val	Gly His Leu Val Glu
	210	215	220
	Leu Glu Glu Ala Asp	Ala Cys Gly Met Asn	Leu Tyr Asp Ile Arg Glu
	225	230	235 240
	Arg Lys Phe Ser	Asp Glu Leu Leu His	Leu Ile Asp Ser Ser Ser Lys
		245	250 255
[0032]	Asp Lys Thr	Ile Arg Gln Lys Leu	Met Arg Ala Pro Met Lys Asn Leu
		260	265 270
	Ile Ala Gly Thr	Ile Cys Lys Tyr	Phe Ile Glu Lys Tyr Gly Phe Asn
		275	280 285
	Thr Asn Cys Lys Val	Ser Pro Met Thr Gly	Asp Asn Leu Ala Thr Ile
		295	300
	Cys Ser Leu Pro Leu	Arg Lys Asn Asp Val	Leu Val Ser Leu Gly Thr
	305	310	315 320
	Ser Thr Thr Val	Leu Leu Val Thr Asp	Lys Tyr His Pro Ser Pro Asn
		325	330 335
	Tyr His Leu Phe	Ile His Pro Thr	Leu Pro Asn His Tyr Met Gly Met
		340	345 350
	Ile Cys Tyr	Cys Asn Gly Ser	Leu Ala Arg Glu Arg Ile Arg Asp Glu
		355	360 365
	Leu Asn Lys Glu Arg	Glu Asn Asn Tyr Glu	Lys Thr Asn Asp Trp Thr
		370	375 380
	Leu Phe Asn Gln	Ala Val Leu Asp Asp	Ser Glu Ser ser Glu Asn Glu

385	390	395	400
Leu Gly Val Tyr Phe 405	Pro Leu Gly Glu Ile Val 410	Pro Ser Val Lys Ala 415	
Ile Asn Lys Arg Val 420	Ile Phe Asn Pro Lys Thr Gly Met 425	Ile Glu Arg 430	
Glu Val Ala Lys Phe Lys Asp Lys Arg His Asp Ala Lys Asn Ile Val 435	440	445	
Glu Ser Gln Ala Leu Ser Cys Arg Val Arg Ile Ser Pro Leu Leu Ser 450	455	460	
Asp Ser Asn Ala Ser Ser Gln Gln Arg Leu Asn Glu Asp Thr Ile Val 465	470	475	480
Lys Phe Asp Tyr Asp Glu Ser Pro Leu Arg Asp Tyr Leu Asn Lys Arg 485	490	495	
Pro Glu Arg Thr Phe Phe Val Gly Gly Ala Ser Lys Asn Asp Ala Ile 500	505	510	
Val Lys Lys Phe Ala Gln Val Ile Gly Ala Thr Lys Gly Asn Phe Arg 515	520	525	
[0033] Leu Glu Thr Pro Asn Ser Cys Ala Leu Gly Gly Cys Tyr Lys Ala Met 530	535	540	
Trp Ser Leu Leu Tyr Asp Ser Asn Lys Ile Ala Val Pro Phe Asp Lys 545	550	555	560
Phe Leu Asn Asp Asn Phe Pro Trp His Val Met Glu Ser Ile Ser Asp 565	570	575	
Val Asp Asn Glu Asn Trp Asp Arg Tyr Asn Ser Lys Ile Val Pro Leu 580	585	590	
Ser Glu Leu Glu Lys Thr Leu Ile 595	600		
<210> 26			
<211> 2467			
<212> DNA			
<213> Saccharomyces cerevisiae			
<400> 26			
ggatccaaga ccattattcc atcagaatgg aaaaaagttt aaaagatcac ggagattttg			60
ttcttctgag cttctgctgt ccttgaaaac aaattattcc gctggccgcc ccaaacaaaa			120
acaaccccga ttttaataaca ttgtcacagt attagaaatt ttctttttac aaattaccat			180
ttccagctta ctacttccta taatcctcaa tcttcagcaa gcgaagcagg gaatagccgc			240

[0034]	tgagggtgcat aactgtcact tttcaattcg gccaatgcaa tctcaggcgg acgaataagg	300
	gggccctctc gagaaaaaca aaaggaggat gagattagta ctttaatgtt gtgttcagta	360
	attcagagac agacaagaga ggtttccaac acaatgtctt tagactcata ctatcttggg	420
	tttgatcttt cgaccaaca actgaaatgt ctcgccatta accaggacct aaaaattgtc	480
	cattcagaaa cagtggaatt tgaaaaggat cttccgcatt atcacacaaa gaagggtgtc	540
	tatatacacg gcgacactat cgaatgtccc gtagccatgt ggtaggggc tctagatctg	600
	gttctctcga aatatcgga ggctaaattt ccattgaaca aagttatggc cgtctcaggg	660
	tcctgccagc agcacgggtc tgtctactgg tcctcccaag ccgaatctct gttagagcaa	720
	ttgaataaga aaccggaaaa agatttattg cactacgtga gctctgtagc atttgcaagg	780
	caaaccgccc ccaattggca agaccacagt actgcaaagc aatgtcaaga gtttgaagag	840
	tgcataggtg ggcctgaaaa aatggctcaa ttaacagggt ccagagccca ttttagattt	900
	actggtcctc aaattctgaa aattgcacaa ttagaaccag aagcttacga aaaaacaaag	960
	accatttctt tagtgtctaa ttttttgact tctatcttag tgggccatct tgttgaatta	1020
	gaggaggcag atgcctgtgg tatgaacctt tatgatatac gtgaaagaaa attcatgtat	1080
	gagctactac atctaattga tagttcttct aaggataaaa ctatcagaca aaaattaatg	1140
	agagcaccca tgaaaaattt gatagcgggt accatctgta aatattttat tgagaagtac	1200
	ggtttcaata caaactgcaa ggtctctccc atgactgggg ataatttagc cactatatgt	1260
	tctttacccc tgcggaagaa tgacgttctc gtttccctag gaacaagtac tacagttctt	1320
	ctgggtcaccg ataagtatca cccctctccg aactatcatc ttttcattca tccaactctg	1380
	ccaaaccatt atatgggtat gatttggtat tgtaatgggt ctttggaag ggagaggata	1440
	agagacgagt taaacaaaga acgggaaaat aattatgaga agactaacga ttggactctt	1500
	tttaatcaag ctgtgctaga tgactcagaa agtagtgaaa atgaattagg tgtatatattt	1560
	cctctggggg agatcgttcc tagcgtaaaa gccataaaca aaagggttat cttcaatcca	1620
	aaaacgggta tgattgaaag agagggtggc aagttcaaag acaagaggca cgatgccaaa	1680
	aatattgtag aatcacaggc ttttaagttgc agggtaagaa tatctcccct gctttcggat	1740
	tcaaacgcaa gctcacaaca gagactgaac gaagatacaa tcgtgaagtt tgattacgat	1800
	gaatctccgc tgcgggacta cctaaataaa aggccagaaa ggactttttt tgtagggtgg	1860
	gcttctaaaa acgatgctat tgtgaagaag tttgctcaag tcattgggtg taaaaaggt	1920
	aatttttaggc tagaaacacc aaactcatgt gcccttggtg gttgttataa ggccatgtgg	1980
	tcattgttat atgactctaa taaaattgca gtcccttttg ataaatttct gaatgacaat	2040
	tttccatggc atgtaatgga aagcatatcc gatgtggata atgaaaattg gatcgctata	2100
	attccaagat tgtcccctta agcgaactgg aaaagactct catctaaaat atgtttgaat	2160
	aatttatcat gccctgacaa gtacacacaa acacagacac ataatatata tacatatata	2220
	tatatcaccg ttattatgcg tgcacatgac aatgcccttg tatgtttcgt atactgtagc	2280

aagtagtcat ctttttgttc cccgttcgga aaatgacaaa aagtaaaatc aataaatgaa 2340
 gagtaaaaaa caatttatga aagggtgagc gaccagcaac gagagagaca aatcaaatta 2400
 gcgctttcca gtgagaatat aagagagcat tgaaagagct aggttattgt taaatcatct 2460
 cgagctc 2467

<210> 27
 <211> 494
 <212> PRT
 <213> Piromyces的种
 <400> 27

Met Lys Thr Val Ala Gly Ile Asp Leu Gly Thr Gln Ser Met Lys Val
 1 5 10 15
 Val Ile Tyr Asp Tyr Glu Lys Lys Glu Ile Ile Glu Ser Ala Ser Cys
 20 25 30
 Pro Met Glu Leu Ile Ser Glu Ser Asp Gly Thr Arg Glu Gln Thr Thr
 35 40 45
 Glu Trp Phe Asp Lys Gly Leu Glu Val Cys Phe Gly Lys Leu Ser Ala
 50 55 60
 [0035] Asp Asn Lys Lys Thr Ile Glu Ala Ile Gly Ile Ser Gly Gln Leu His
 65 70 75 80
 Gly Phe Val Pro Leu Asp Ala Asn Gly Lys Ala Leu Tyr Asn Ile Lys
 85 90 95
 Leu Trp Cys Asp Thr Ala Thr Val Glu Glu Cys Lys Ile Ile Thr Asp
 100 105 110
 Ala Ala Gly Gly Asp Lys Ala Val Ile Asp Ala Leu Gly Asn Leu Met
 115 120 125
 Leu Thr Gly Phe Thr Ala Pro Lys Ile Leu Trp Leu Lys Arg Asn Lys
 130 135 140
 Pro Glu Ala Phe Ala Asn Leu Lys Tyr Ile Met Leu Pro His Asp Tyr
 145 150 155 160
 Leu Asn Trp Lys Leu Thr Gly Asp Tyr Val Met Glu Tyr Gly Asp Ala
 165 170 175
 Ser Gly Thr Ala Leu Phe Asp Ser Lys Asn Arg Cys Trp Ser Lys Lys
 180 185 190
 Ile Cys Asp Ile Ile Asp Pro Lys Leu Leu Asp Leu Leu Pro Lys Leu
 195 200 205

Ile Glu Pro Ser Ala Pro Ala Gly Lys Val Asn Asp Glu Ala Ala Lys
 210 215 220
 Ala Tyr Gly Ile Pro Ala Gly Ile Pro Val Ser Ala Gly Gly Gly Asp
 225 230 235 240
 Asn Met Met Gly Ala Val Gly Thr Gly Thr Val Ala Asp Gly Phe Leu
 245 250 255
 Thr Met Ser Met Gly Thr Ser Gly Thr Leu Tyr Gly Tyr Ser Asp Lys
 260 265 270
 Pro Ile Ser Asp Pro Ala Asn Gly Leu Ser Gly Phe Cys Ser Ser Thr
 275 280 285
 Gly Gly Trp Leu Pro Leu Leu Cys Thr Met Asn Cys Thr Val Ala Thr
 290 295 300
 Glu Phe Val Arg Asn Leu Phe Gln Met Asp Ile Lys Glu Leu Asn Val
 305 310 315 320
 Glu Ala Ala Lys Ser Pro Cys Gly Ser Glu Gly Val Leu Val Ile Pro
 325 330 335
 [0036] Phe Phe Asn Gly Glu Arg Thr Pro Asn Leu Pro Asn Gly Arg Ala Ser
 340 345 350
 Ile Thr Gly Leu Thr Ser Ala Asn Thr Ser Arg Ala Asn Ile Ala Arg
 355 360 365
 Ala Ser Phe Glu Ser Ala Val Phe Ala Met Arg Gly Gly Leu Asp Ala
 370 375 380
 Phe Arg Lys Leu Gly Phe Gln Pro Lys Glu Ile Arg Leu Ile Gly Gly
 385 390 395 400
 Gly Ser Lys Ser Asp Leu Trp Arg Gln Ile Ala Ala Asp Ile Met Asn
 405 410 415
 Leu Pro Ile Arg Val Pro Leu Leu Glu Glu Ala Ala Ala Leu Gly Gly
 420 425 430
 Ala Val Gln Ala Leu Trp Cys Leu Lys Asn Gln Ser Gly Lys Cys Asp
 435 440 445
 Ile Val Glu Leu Cys Lys Glu His Ile Lys Ile Asp Glu Ser Lys Asn
 450 455 460
 Ala Asn Pro Ile Ala Glu Asn Val Ala Val Tyr Asp Lys Ala Tyr Asp
 465 470 475 480

Glu Tyr Cys Lys Val Val Asn Thr Leu Ser Pro Leu Tyr Ala
485 490

<210> 28

<211> 2041

<212> DNA

<213> Piromyces的种

<400> 28

[0037] attatataaa ataactttaa ataaaacaat ttttatttgt ttatttaatt attcaaaaaa 60
aattaaagta aaagaaaaat aatacagtag aacaatagta ataatatcaa aatgaagact 120
gttgctggta ttgatcttgg aactcaaagt atgaaagtcg ttatttacga ctatgaaaag 180
aaagaaatta ttgaaagtgc tagctgtcca atggaattga tttccgaaag tgacggtacc 240
cgtgaacaaa ccaactgaatg gtttgacaag ggtcttgaag tttgttttgg taagcttagt 300
gctgataaca aaaagactat tgaagctatt ggtatttctg gtcaattaca cggttttgtt 360
cctcttgatg ctaacggtaa ggctttatac aacatcaaac tttgggtgta tactgctacc 420
gttgaagaat gtaagattat cactgatgct gccggtgggtg acaaggctgt tattgatgcc 480
cttggttaacc ttatgctcac cggtttcacc gtcctcaaaga tcctctggct caagcgcaac 540
aagccagaag ctttcgctaa cttaaagtac attatgcttc cacacgatta cttaaactgg 600
aagcttactg gtgattacgt tatggaatac ggtgatgcct ctggtaccgc tctcttcgat 660
tctaagaacc gttgctggtc taagaagatt tgcgatatca ttgacccaaa acttttagat 720
ttacttccaa agttaattga accaagcgct ccagctggta aggttaattga tgaagccgct 780
aaggcttacg gtattccagc cggtattcca gtttccgctg gtggtgggtga taacatgatg 840
gggtgctgtg gtactggtag tgttgcctgat gggttccctta ccatgtctat ggggtacttct 900
gggtactctt acggttacag tgacaagcca attagtgacc cagctaattgg ttaagtgggt 960
ttctgttctt ctactggtag atggcttcca ttactttgta ctatgaactg tactgttgcc 1020
actgaattcg ttcgtaacct ctccaaatg gatattaagg aacttaattgt tgaagctgcc 1080
aagtctccat gtggtagtga aggtgtttta gttattccat tcttcaatgg tgaaagaact 1140
ccaaacttac caaacggtcg tgctagtatt actggtctta cttctgctaa caccagccgt 1200
gctaacattg ctgctgctag tttcgaatcc gccgttttcg ctatgcgtgg tggtttagat 1260
gctttccgta agttagggtt ccaaccaaag gaaattcgtc ttattgggtg tggttctaag 1320
tctgatctct ggagacaaat tgccgctgat atcatgaacc ttccaatcag agttccactt 1380
ttagaagaag ctgctgctct tgggtgggtg gttcaagctt tatgggtgtct taagaaccaa 1440
tctggttaagt gtgatattgt tgaactttgc aaagaacaca ttaagattga tgaatctaag 1500
aatgctaacc caattgccga aaatgttgct gtttacgaca aggcttacga tgaatactgc 1560
aagggtgtaa atactctttc tccattatat gcttaaattg ccaatgtaaa aaaaaatata 1620
atgccatata attgccttgt caatacactg ttcattgttca tataatcata ggacattgaa 1680
tttacaagggt ttatacaatt aatatctatt atcatattat tatacagcat ttcattttct 1740

aagattagac gaaacaattc ttggttcctt gcaatataca aaatttacat gaatttttag 1800
aatagtctcg tatttatgcc caataatcag gaaaattacc taatgctgga ttcttggttaa 1860
taaaaacaaa ataaataaat taaataaaca aataaaaatt ataagtaaat ataaatatat 1920
aagtaatatata aaaaaaaagt aaataaataa ataaataaat aaaaattttt tgcaaatata 1980
taaataaata aataaaatat aaaaataatt tagcaataa attaaaaaaa aaaaaaaaaa 2040
a 2041

<210> 29
<211> 327
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 29

Met Ser Ser Leu Val Thr Leu Asn Asn Gly Leu Lys Met Pro Leu Val
1 5 10 15

Gly Leu Gly Cys Trp Lys Ile Asp Lys Lys Val Cys Ala Asn Gln Ile
20 25 30

Tyr Glu Ala Ile Lys Leu Gly Tyr Arg Leu Phe Asp Gly Ala Cys Asp
35 40 45

[0038] Tyr Gly Asn Glu Lys Glu Val Gly Glu Gly Ile Arg Lys Ala Ile Ser
50 55 60

Glu Gly Leu Val Ser Arg Lys Asp Ile Phe Val Val Ser Lys Leu Trp
65 70 75 80

Asn Asn Phe His His Pro Asp His Val Lys Leu Ala Leu Lys Lys Thr
85 90 95

Leu Ser Asp Met Gly Leu Asp Tyr Leu Asp Leu Tyr Tyr Ile His Phe
100 105 110

Pro Ile Ala Phe Lys Tyr Val Pro Phe Glu Glu Lys Tyr Pro Pro Gly
115 120 125

Phe Tyr Thr Gly Ala Asp Asp Glu Lys Lys Gly His Ile Thr Glu Ala
130 135 140

His Val Pro Ile Ile Asp Thr Tyr Arg Ala Leu Glu Glu Cys Val Asp
145 150 155 160

Glu Gly Leu Ile Lys Ser Ile Gly Val Ser Asn Phe Gln Gly Ser Leu
165 170 175

Ile Gln Asp Leu Leu Arg Gly Cys Arg Ile Lys Pro Val Ala Leu Gln
180 185 190

Ile Glu His His Pro Tyr Leu Thr Gln Glu His Leu Val Glu Phe Cys
 195 200 205

Lys Leu His Asp Ile Gln Val Val Ala Tyr Ser Ser Phe Gly Pro Gln
 210 215 220

Ser Phe Ile Glu Met Asp Leu Gln Leu Ala Lys Thr Thr Pro Thr Leu
 225 230 235 240

Phe Glu Asn Asp Val Ile Lys Lys Val Ser Gln Asn His Pro Gly Ser
 245 250 255

Thr Thr Ser Gln Val Leu Leu Arg Trp Ala Thr Gln Arg Gly Ile Ala
 260 265 270

Val Ile Pro Lys Ser Ser Lys Lys Glu Arg Leu Leu Gly Asn Leu Glu
 275 280 285

Ile Glu Lys Lys Phe Thr Leu Thr Glu Gln Glu Leu Lys Asp Ile Ser
 290 295 300

Ala Leu Asn Ala Asn Ile Arg Phe Asn Asp Pro Trp Thr Trp Leu Asp
 305 310 315 320

Gly Lys Phe Pro Thr Phe Ala
 325

[0039]

<210> 30
 <211> 984
 <212> DNA
 <213> Saccharomyces cerevisiae

<400> 30
 atgtcttcac tgggttactct taataacggt ctgaaaatgc ccctagtcgg cttaggggtgc 60
 tggaaaattg acaaaaaagt ctgtgcgaat caaatattatg aagctatcaa attaggctac 120
 cgtttattcg atggtgcttg cgactacggc aacgaaaagg aagttggtga aggtatcagg 180
 aaagccatct ccgaagggtct tgtttctaga aaggatatat ttgttgtttc aaagttatgg 240
 aacaattttc accatcctga tcatgtaaaa ttagctttta agaagacctt aagcgatatg 300
 ggacttgatt atttagacct gtattatatt cacttcccaa tcgccttcaa atatgttcca 360
 tttgaagaga aataccctcc aggattctat acgggcgag atgacgagaa gaaagggtcac 420
 atcaccgaag cacatgtacc aatcatagat acgtaccggg ctctggaaga atgtgttgat 480
 gaaggcttga ttaagtctat tgggtgttcc aactttcagg gaagcttgat tcaagattta 540
 ttacgtggtt gtagaatcaa gcccgtggct ttgcaaattg aacaccatcc ttatttgact 600
 caagaacacc tagttgagtt ttgtaaatta cacgatatcc aagtagttgc ttactcctcc 660
 ttcggtcctc aatcattcat tgagatggac ttacagttgg caaaaaccac gccaaactctg 720
 ttcgagaatg atgtaatcaa gaagggtctca caaaaccatc caggcagtag cacttcccaa 780

	gtattgctta gatgggcaac tcagagagggc attgccgtca ttccaaaatc ttccaagaag	840
	gaaagggttac ttggcaacct agaaatcgaa aaaaagttca ctttaacgga gcaagaattg	900
	aaggatattt ctgcactaaa tgccaacatc agatttaatg atccatggac ctggttggat	960
	ggtaaattcc ccacttttgc ctga	984
	<210> 31	
	<211> 31	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 31	
	gactagtcga gtttatcatt atcaatactg c	31
	<210> 32	
	<211> 49	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 32	
	ctcataatca ggtactgata acattttggt tgtttatgtg tgtttattc	49
[0040]	<210> 33	
	<211> 49	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 33	
	gaataaacac acataaaca acaaaatggt atcagtacct gattatgag	49
	<210> 34	
	<211> 48	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 34	
	aatcataaat cataagaaat tcgcttactt taagaatgcc ttagtcat	48
	<210> 35	
	<211> 48	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 35	
	atgactaagg cattcttaaa gtaagcgaat ttcttatgat ttatgatt	48

	<210> 36	
	<211> 36	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 36	
	cactagtctc gagtgtggaa gaacgattac aacagg	36
	<210> 37	
	<211> 31	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 37	
	cgagctcgtg ggtgtattgg attataggaa g	31
	<210> 38	
	<211> 48	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
[0041]	<400> 38	
	ttgggctggt tcaactaaat tcatttttag gctggatatct tgattcta	48
	<210> 39	
	<211> 48	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 39	
	tagaatcaag ataccagcct aaaaatgaat ttagttgaaa cagcccaa	48
	<210> 40	
	<211> 48	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 40	
	aatcataaat cataagaaat tcgctctaatt atttgattgc ttgcccag	48
	<210> 41	
	<211> 48	
	<212> DNA	
	<213> 人工	
	<220>	

	<223> 引物	
	<400> 41 ctgggcaagc aatcaaata tagagcgaat ttcttatgat ttatgatt	48
	<210> 42 <211> 31 <212> DNA <213> 人工	
	<220> <223> 引物	
	<400> 42 tgagctcgtg tggaagaacg attacaacag g	31
	<210> 43 <211> 28 <212> DNA <213> 人工	
	<220> <223> 引物	
	<400> 43 acgcgtcgcac tcgtaggaac aatttcgg	28
	<210> 44 <211> 50 <212> DNA <213> 人工	
[0042]	<220> <223> 引物	
	<400> 44 cttcttggtt taatgcttct agcatttttt gattaaaatt aaaaaaactt	50
	<210> 45 <211> 50 <212> DNA <213> 人工	
	<220> <223> 引物	
	<400> 45 aagttttttt aattttaatc aaaaaatgct agaagcatta aaacaagaag	50
	<210> 46 <211> 46 <212> DNA <213> 人工	
	<220> <223> 引物	
	<400> 46 ggtatatatt taagagcgat ttgtttactt gcgaactgca tgatcc	46
	<210> 47 <211> 46	

	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 47	
	ggatcatgca gttcgcaagt aaacaaatcg ctcttaaata tataacc	46
	<210> 48	
	<211> 33	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 引物	
	<400> 48	
	cgcagtcgac cttttaaaca gttgatgaga acc	33
	<210> 49	
	<211> 676	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 启动子	
	<400> 49	
	tcgagtttat cattatcaat actgccattt caaagaatac gtaaataatt aatagtagtg	60
[0043]	attttcctaa ctttatcttag tcaaaaaatt agccttttaa ttctgctgta acccgtagat	120
	gccccaaaata gggggcggggt tacacagaat atataacatc gtaggtgtct gggatgaacag	180
	tttattcctg gcatccacta aatataatgg agcccgcttt ttaagctggc atccagaaaa	240
	aaaaagaatc ccagcaccaa aatattgttt tcttcaccaa ccatcagttc ataggtccat	300
	tctcttagcg caactacaga gaacaggggc acaaacaggc aaaaaacggg cacaacctca	360
	atggagtgat gcaacctgcc tggagtaaatt gatgacacaa ggcaattgac ccacgcatgt	420
	atctatctca ttttcttaca ctttctatta ctttctgctc tctctgattt ggaaaaagct	480
	gaaaaaaaag gttgaaacca gttccctgaa attattcccc tacttgacta ataagtatat	540
	aaagacggta ggtattgatt gtaattctgt aaatctatct cttaaacttc ttaaattcta	600
	cttttatagt tagtcttttt tttagtttta aaacaccaag aacttagttt cgaataaaca	660
	cacataaaca aacaaa	676
	<210> 50	
	<211> 326	
	<212> DNA	
	<213> 人工	
	<220>	
	<223> 终止子	
	<400> 50	
	gcgaatttct tatgatttat gatttttatt attaaataag ttataaaaaa aataagtgt	60
	tacaaatttt aaagtgactc ttaggtttta aaacgaaaat tcttattctt gagtaactct	120

```

ttcctgtagg tcaggttgct ttctcaggta tagcatgagg tcgctcttat tgaccacacc 180
tctaccggca tgccgagcaa atgctgcaa atcgctcccc atttcacca attgtagata 240
tgctaaactcc agcaatgagt tgatgaatct cgggtgtgtat tttatgtcct cagaggacaa 300
cacctgttgt aatcgttctt ccacac 326

```

<210> 51
 <211> 374
 <212> DNA
 <213> 人工

<220>
 <223> 启动子

```

<400> 51
gtgggtgtat tggattatag gaagccacgc gctcaacctg gaattacagg aagctggtaa 60
ttttttgggt ttgcaatcat caccatctgc acgttggtat aatgtcccggt gtctatatat 120
atccattgac ggtattctat ttttttgcta ttgaaatgag cgttttttgt tactacaatt 180
ggtttttacag acggaatttt ccctatttgt ttcgctcccat ttttcctttt ctcatgttgc 240
tcatatctta aaaaggtcct ttcttcataa tcaatgcttt cttttactta atattttact 300
tgcattcagt gaattttaat acatattcct ctagtcttgc aaaatcgatt tagaatcaag 360
ataccagcct aaaa 374

```

[0044]

<210> 52
 <211> 390
 <212> DNA
 <213> 人工

<220>
 <223> 启动子

```

<400> 52
ctcgtaggaa caatttcggg cccctgcgtg ttcttctgag gttcatcttt tacatttgct 60
tctgctggat aattttcaga ggcaacaagg aaaaattaga tggcaaaaag tcgtctttca 120
aggaaaaatc cccaccatct ttcgagatcc cctgtaactt attggcaact gaaagaatga 180
aaaggaggaa aatacaaaat atactagaac tgaaaaaaa aaagtataaa tagagacgat 240
atatgccaat acttcacaat gttcgaatct attcttcatt tgcagctatt gtaaaataat 300
aaaacatcaa gaacaaacaa gctcaacttg tcttttctaa gaacaaagaa taaacacaaa 360
aacaanaagt ttttttaatt ttaatcaaaa 390

```

<210> 53
 <211> 302
 <212> DNA
 <213> 人工

<220>
 <223> 终止子

```

<400> 53
acaaatcgct cttaaataa tacctaaaga acattaaagc tatattataa gcaaagatac 60
gtaaattttg cttatattat tatacacata tcatatttct atatttttaa gatttggtta 120
tataatgtac gtaatgcaa ggaaataaat ttatacatt attgaacagc gtccaagtaa 180
[0045] ctacattatg tgcactaata gtttagcgct gtgaagactt tattgtgtcg cgaaaagtaa 240
aaatttttaa aattagagca ccttgaactt gcgaaaagg ttctcatcaa ctgtttaaaa 300
gg 302

```

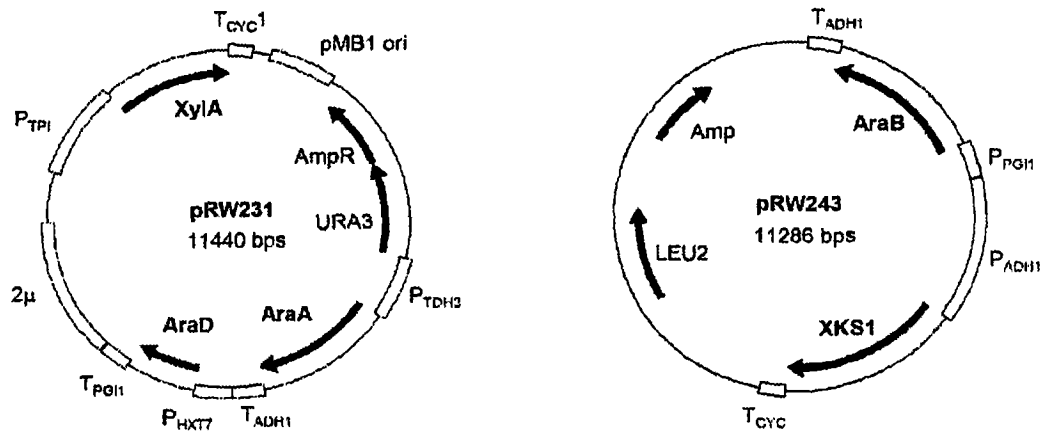


图1

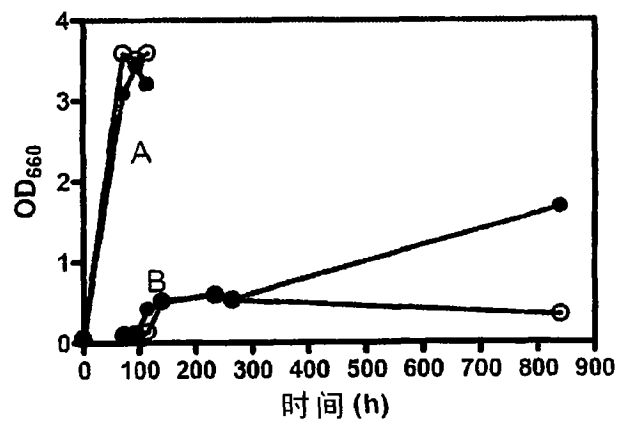


图2

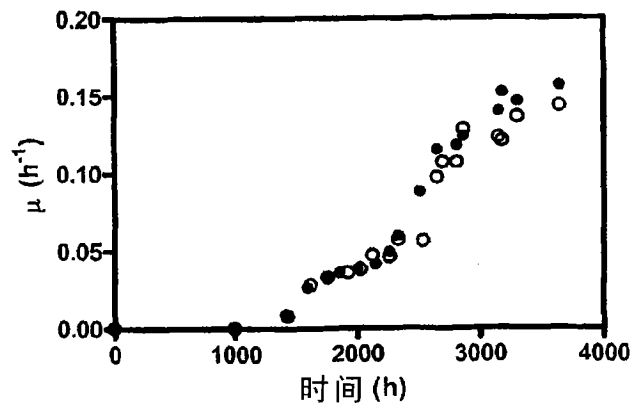


图3

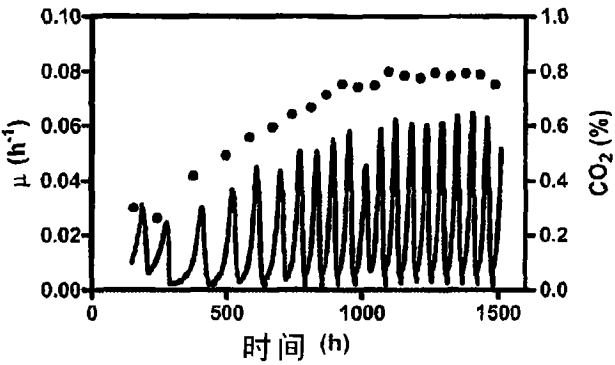


图4

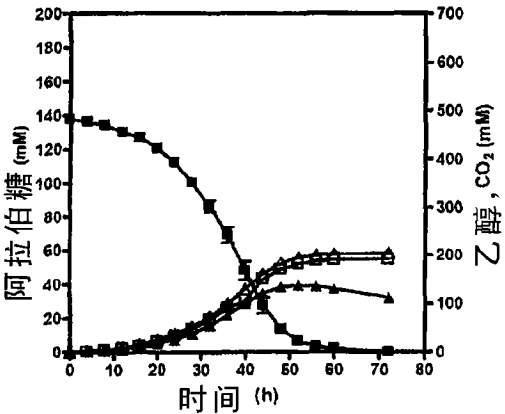


图5A

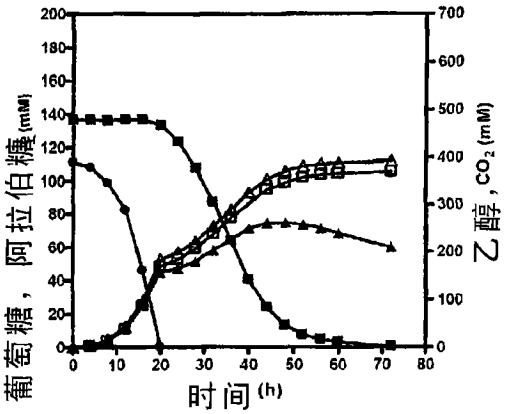


图5B

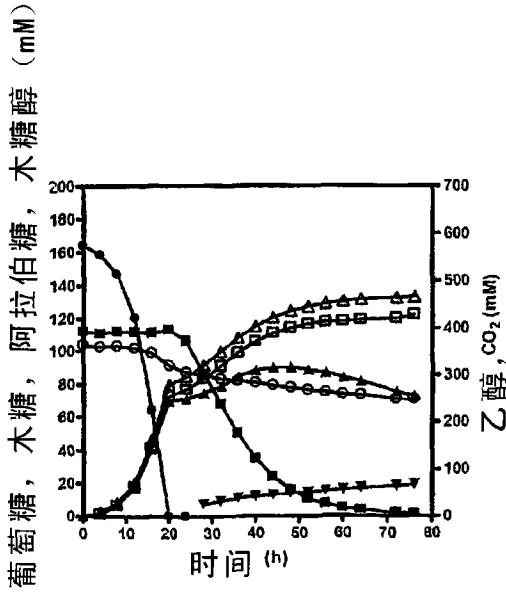


图5C

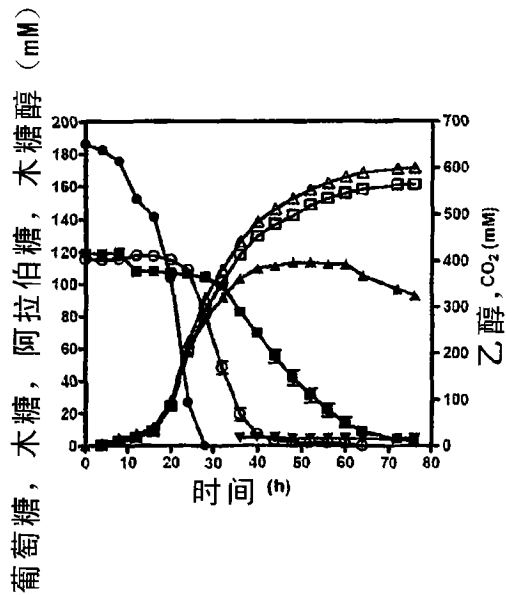


图5D

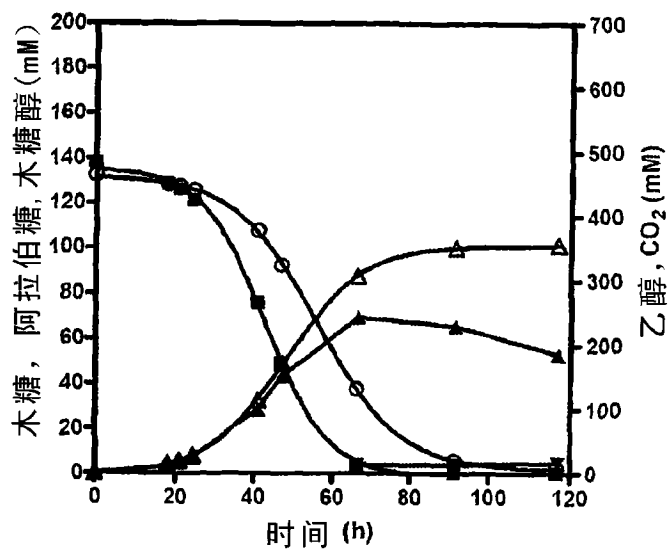


图6

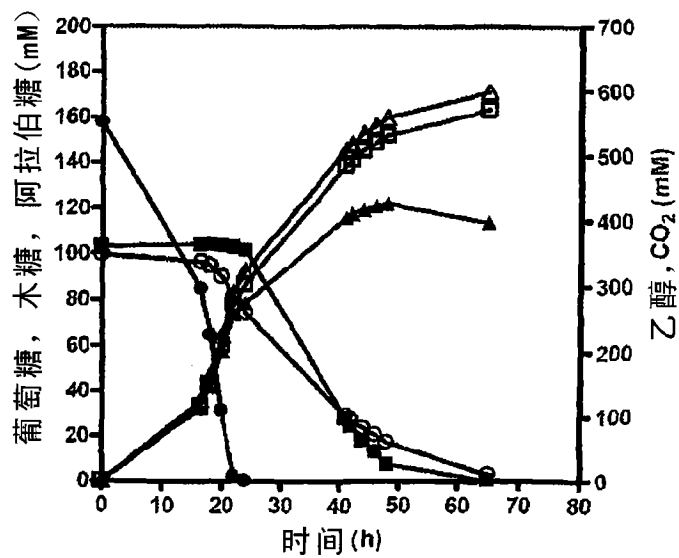


图7