



(12) 发明专利



(10) 授权公告号 CN 113195509 B

(45) 授权公告日 2025. 02. 21

(21) 申请号 201980083180.5	(51) Int.Cl.
(22) 申请日 2019.12.18	C07H 3/06 (2006.01)
(65) 同一申请的已公布的文献号	C07H 5/06 (2006.01)
申请公布号 CN 113195509 A	C12N 9/10 (2006.01)
(43) 申请公布日 2021.07.30	C12P 19/18 (2006.01)
(30) 优先权数据	(56) 对比文件
18213728.1 2018.12.18 EP	W0 2016040531 A1,2016.03.17
(85) PCT国际申请进入国家阶段日	Sun,H.Y., et al..Accession No:WP_
2021.06.16	042442472.1,glycosyltransferase family 10
(86) PCT国际申请的申请数据	[Azospirillum sp. B510].Genbank
PCT/EP2019/085841 2019.12.18	database.2015,ORIGIN和FEATURES.
(87) PCT国际申请的公布数据	Sun,H.Y.,et al..Accession No:WP_
W02020/127417 EN 2020.06.25	038499425.1,glycosyltransferase family 10
(73) 专利权人 因比奥斯公司	[Basilea psittacipulmonis].Genbank
地址 比利时根特	database.2014,ORIGIN和FEATURES.
(72) 发明人 J·博普雷兹 N·兰诺	Sun,H.Y.,et al..Accession No:WP_
K·梵德瓦勒 A·瓦考特兰	085088362.1,glycosyltransferase family 10
(74) 专利代理机构 中国贸促会专利商标事务所	[Azospirillum oryzae].Genbank
有限公司 11038	database.2017,ORIGIN和FEATURES.
专利代理师 陈晓娜	审查员 高赞
	权利要求书5页 说明书35页
	序列表32页 附图14页

(54) 发明名称

生产3-岩藻糖基乳糖和转化乳糖的 α -1,3-岩藻糖基转移酶

(57) 摘要

本发明涉及生产3-岩藻糖基乳糖(3-FL)的方法以及新型岩藻糖基转移酶,更具体地新型乳糖结合 α -1,3-岩藻糖基转移酶多肽,以及它们的应用。此外,本发明提供了使用新型乳糖结合 α -1,3-岩藻糖基转移酶生产3-岩藻糖基乳糖(3-FL)的方法。

1. 一种生产 α -1,3-岩藻糖基乳糖的方法,所述方法包括以下步骤:
 - a) 提供具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽,其中所述多肽选自自由以下项目组成的组:SEQ ID NO: 6、2或4,
 - b) 在其中所述多肽催化岩藻糖残基从供体底物转移至受体底物的条件下,使步骤a)的具有 α -1,3-岩藻糖基转移酶活性的多肽与包含GDP-岩藻糖作为供体底物和乳糖作为受体底物的混合物接触,从而产生 α -1,3-岩藻糖基乳糖,
 - c) 任选地分离所述 α -1,3-岩藻糖基乳糖。
2. 根据权利要求1所述的方法,其中所述多肽在无细胞系统中提供。
3. 根据权利要求1所述的方法,其中所述多肽由包含编码所述多肽的多核苷酸的细胞产生。
4. 根据权利要求1或3中任一项所述的方法,其中所述GDP-岩藻糖和/或乳糖由产生所述GDP-岩藻糖和/或乳糖的细胞提供。
5. 根据权利要求1所述的方法,所述方法包括以下步骤:
 - i) 提供经基因修饰以产生 α -1,3-岩藻糖基乳糖的细胞,所述细胞包含至少一种编码用于 α -1,3-岩藻糖基乳糖合成的酶的核酸序列,所述细胞包含具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的所述多肽的表达,
 - ii) 在允许产生 α -1,3-岩藻糖基乳糖的条件下在培养基中培养细胞。
6. 根据权利要求5所述的方法,所述方法进一步包括从培养中分离 α -1,3-岩藻糖基乳糖的步骤。
7. 根据权利要求3所述的方法,所述方法包括以下步骤:
 - a) 提供表达具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的所述多肽的宿主细胞;
 - b) 在允许产生 α -1,3-岩藻糖基乳糖并允许表达具有 α -1,3-岩藻糖基转移酶活性的所述多肽的合适的营养条件下使所述宿主细胞生长;
 - c) 在步骤b)的同时或之后提供供体底物GDP-岩藻糖和受体底物乳糖,以便 α -1,3-岩藻糖基转移酶多肽催化岩藻糖残基从GDP-岩藻糖向乳糖的转移,从而产生 α -1,3-岩藻糖基乳糖;
 - d) 任选地从宿主细胞或其生长的培养基中分离所述 α -1,3-岩藻糖基乳糖。
8. 根据权利要求7所述的方法,其中宿主细胞经转化或转染以表达具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的外源多肽。
9. 根据权利要求5或7中任一项所述的方法,其特征在于通过在宿主细胞中同时表达的酶或通过宿主细胞的代谢来提供GDP-岩藻糖和/或乳糖。
10. 根据权利要求1所述的方法,其进一步包括 α -1,3-岩藻糖基乳糖的纯化。
11. 根据权利要求1所述的方法,所述方法进一步包括以下步骤中的至少一个:
 - i) 向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖;
 - ii) 以连续方式向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少

50克、至少75克、至少100克、至少120克或至少150克乳糖；

iii) 向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖,并且使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍,或不大于2倍,或小于2倍;

iv) 以连续方式向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖,并且使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍,或不大于2倍,或小于2倍;

v) 在1天、2天、3天、4天或5天的过程中,通过进料溶液以连续方式向培养基中添加乳糖进料;

vi) 在1天、2天、3天、4天或5天的过程中,通过进料溶液以连续方式向培养基中添加乳糖进料,并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L;

vii) 在1天、2天、3天、4天或5天的过程中,通过进料溶液以连续方式向培养基中添加乳糖进料,并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L;并且其中将所述溶液的pH设置在3至7;

viii) 在1天、2天、3天、4天或5天的过程中,通过进料溶液以连续方式向培养基中添加乳糖进料,并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L;并且其中将所述溶液的pH设置在3至7,并且其中将所述进料溶液的温度保持在20℃至80℃;

所述方法导致在所述培养基的最终体积中 α -1,3-岩藻糖基乳糖浓度为至少50 g/L、至少75g/L、至少90 g/L、至少100 g/L、至少125 g/L、至少150 g/L、至少150 g/L、至少175 g/L或至少200 g/L。

12. 根据权利要求3或5中任一项所述的方法,其中所述细胞选自由微生物、除能够发育成完整植物的植物细胞外的植物细胞、或除能够发育成完整动物的动物细胞外的动物细胞组成的组,其中所述微生物是细菌或真菌,所述植物细胞来源于水稻、棉花、油菜籽、大豆、玉米或谷类植物,所述动物细胞来源于昆虫、鱼类、鸟类或非人类哺乳动物。

13. 根据权利要求12所述的方法,其中所述细胞是大肠杆菌(*Escherichia coli*)细胞或酵母细胞。

14. 经基因修饰以产生 α -1,3-岩藻糖基乳糖的宿主细胞,其中宿主细胞包含至少一种编码参与 α -1,3-岩藻糖基糖合成的酶的核酸序列;所述细胞包含表达具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽,其中所述多肽选自由以下项目组成的组:SEQ ID NO: 6、2或4,

所述细胞还包含:(1) 产生适于通过所述 α -1,3-岩藻糖基转移酶转移至乳糖的GDP-岩藻糖的生物合成途径,和/或(2) 产生乳糖的生物合成途径或乳糖从培养基内化到细胞中的机制,在细胞中乳糖被岩藻糖基化以产生 α -1,3-岩藻糖基乳糖,

并且其中所述细胞选自由微生物、除能够发育成完整植物的植物细胞外的植物细胞、

或除能够发育成完整动物的动物细胞外的动物细胞组成的组,其中所述微生物是细菌或真菌,所述植物细胞来源于水稻、棉花、油菜籽、大豆、玉米或谷类植物,所述动物细胞来源于昆虫、鱼类、鸟类或非人类哺乳动物。

15. 根据权利要求14所述的宿主细胞,其中所述细胞是大肠杆菌细胞或酵母细胞。

16. 根据权利要求14或15任一项所述的宿主细胞,所述宿主细胞包含:i) 包含编码具有乳糖结合 α -1,3-岩藻糖基转移酶活性的所述多肽的多核苷酸的序列,其中所述序列对于所述宿主细胞是外源的序列并且其中所述序列整合在所述宿主细胞的基因组中,或ii) 含有包含编码所述多肽的多核苷酸的载体,其中所述多核苷酸可操作地连接至由用所述载体转染的宿主细胞识别的控制序列。

17. 根据权利要求14所述的宿主细胞,其特征在于所述宿主细胞是细菌细胞。

18. 根据权利要求17所述的宿主细胞,其中所述宿主细胞是大肠杆菌菌株的细胞。

19. 根据权利要求18所述的宿主细胞,其中所述大肠杆菌菌株是K12菌株。

20. 根据权利要求19所述的宿主细胞,其中所述大肠杆菌K12菌株是大肠杆菌MG1655。

21. 根据权利要求14所述的宿主细胞,其特征在于编码具有乳糖结合 α -1,3-岩藻糖基转移酶活性的多肽的多核苷酸适合于相应宿主细胞的密码子使用。

22. 一种生产 α -1,3-岩藻糖基乳糖的方法,其包括以下步骤:

a) 提供根据权利要求14至21中任一项所述的宿主细胞,

b) 在允许产生 α -1,3-岩藻糖基转移酶的条件下在培养基中培养所述宿主细胞。

23. 根据权利要求22所述的方法,所述方法进一步包括从培养中分离所述 α -1,3-岩藻糖基转移酶的步骤。

24. 根据权利要求14至21中任一项所述的宿主细胞用于生产 α -1,3-岩藻糖基乳糖的用途。

25. 权利要求1至13中任一项所述的多肽用于生产 α -1,3-岩藻糖基乳糖的用途。

26. 异源表达乳糖结合 α -1,3-岩藻糖基转移酶多肽的微生物,其中所述多肽选自自由以下项目组成的组:SEQ ID NO: 6、2或4,

所述微生物还包含:(1) 产生适于通过所述 α -1,3-岩藻糖基转移酶转移至乳糖的GDP-岩藻糖的生物合成途径,和/或(2) 产生乳糖的生物合成途径或乳糖从培养基内化到微生物的细胞中的机制,在细胞中乳糖被岩藻糖基化以产生 α -1,3-岩藻糖基乳糖。

27. 根据权利要求26所述的微生物用于生产 α -1,3-岩藻糖基乳糖的用途。

28. 根据权利要求1或22中任一项所述的方法,其还包括从宿主细胞或其生长的培养基中分离所述 α -1,3-岩藻糖基乳糖的步骤。

29. 根据权利要求28所述的方法,其中所述分离包括以下步骤中的至少一个:澄清、超滤、纳米过滤、反渗透、微滤、活性木炭或碳处理、切向流高效过滤、亲和色谱、离子交换色谱、疏水相互作用色谱和/或凝胶过滤、配体交换色谱。

30. 根据权利要求29所述的方法,其中所述超滤是切向流超滤。

31. 根据权利要求28所述的方法,其进一步包括 α -1,3-岩藻糖基乳糖的纯化。

32. 根据权利要求31所述的方法,其中所述 α -1,3-岩藻糖基乳糖的所述纯化包括以下步骤中的至少一个:使用活性木炭或碳、使用木炭、纳米过滤、超滤或离子交换、使用醇、使用含水醇混合物、结晶、蒸发、沉淀或干燥。

33. 根据权利要求32所述的方法, 其中所述干燥是喷雾干燥或冻干。
34. 根据权利要求1或22中任一项所述的方法, 其中所述多肽在真菌、细菌、昆虫、动物和植物表达系统中产生, 所述表达系统不包含: 1) 能够发育成完整昆虫的昆虫细胞, 2) 能够发育成完整动物的动物细胞, 以及3) 能够发育成完整植物的植物细胞。
35. 根据权利要求34所述的方法, 其中所述宿主细胞是细菌细胞。
36. 根据权利要求35所述的方法, 其中所述宿主细胞是大肠杆菌菌株的细胞。
37. 根据权利要求36所述的方法, 其中所述大肠杆菌菌株是K12菌株。
38. 根据权利要求37所述的方法, 其中所述大肠杆菌K12菌株是大肠杆菌MG1655。
39. 根据权利要求34所述的方法, 其中所述宿主细胞是酵母细胞。
40. 根据权利要求1或22中任一项所述的方法, 其中培养基中的乳糖浓度在50至150 g/L的范围内。
41. 根据权利要求1或22中任一项所述的方法, 其中 α -1,3-岩藻糖基乳糖的最终浓度在70 g/L至200 g/L的范围内。
42. 根据权利要求1或22中任一项所述的方法, 其中所述生产导致在发酵结束时乳糖浓度与 α -1,3-岩藻糖基乳糖的浓度比率小于1:5。
43. 根据权利要求1或22中任一项所述的方法, 其中所述生产导致在发酵结束时 α -1,3-岩藻糖基乳糖纯度为80%或更高。
44. 根据权利要求22所述的方法, 所述方法还包括以下步骤中的至少一个:
- i) 向培养基中添加乳糖进料, 所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖;
 - ii) 以连续方式向培养基中添加乳糖进料, 所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖;
 - iii) 向培养基中添加乳糖进料, 所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖, 并且使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍, 或不大于2倍, 或小于2倍;
 - iv) 以连续方式向培养基中添加乳糖进料, 所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖, 并且使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍, 或不大于2倍, 或小于2倍;
 - v) 在1天、2天、3天、4天或5天的过程中, 通过进料溶液以连续的方式向培养基中添加乳糖进料;
 - vi) 在1天、2天、3天、4天或5天的过程中, 通过进料溶液以连续的方式向培养基中添加乳糖进料, 并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L;
 - vii) 在1天、2天、3天、4天或5天的过程中, 通过进料溶液以连续方式向培养基中添加乳糖进料, 并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L; 并且其中将所述溶液的pH设置在3至7;
 - viii) 在1天、2天、3天、4天或5天的过程中, 通过进料溶液以连续方式向培养基中添加

乳糖进料,并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L;并且其中将所述溶液的pH设置在3至7,并且其中将所述进料溶液的温度保持在20℃至80℃;

所述方法导致在所述培养基的最终体积中 α -1,3-岩藻糖基乳糖浓度是至少50 g/L、至少75 g/L、至少90 g/L、至少100 g/L、至少125 g/L、至少150 g/L、至少175 g/L或至少200 g/L,并且任选地在所述培养物的最终体积中乳糖浓度与 α -1,3-岩藻糖基乳糖浓度比率低于1:5、1:10、1:20或1:40。

45. 根据权利要求22所述的方法,所述方法还包括以下步骤中的至少一个:

i) 向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖;

ii) 以连续方式向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖;

iii) 向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖,并且使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍,或不大于2倍,或小于2倍;

iv) 以连续方式向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少50克、至少75克、至少100克、至少120克或至少150克乳糖,并且使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍,或不大于2倍,或小于2倍;

v) 在1天、2天、3天、4天或5天的过程中,通过进料溶液以连续的方式向培养基中添加乳糖进料;

vi) 在1天、2天、3天、4天或5天的过程中,通过进料溶液以连续的方式向培养基中添加乳糖进料,并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L;

vii) 在1天、2天、3天、4天或5天的过程中,通过进料溶液以连续方式向培养基中添加乳糖进料,并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L;并且其中将所述溶液的pH设置在3至7;

viii) 在1天、2天、3天、4天或5天的过程中,通过进料溶液以连续方式向培养基中添加乳糖进料,并且其中所述乳糖进料溶液的浓度是50 g/L、75 g/L、100 g/L、125 g/L、150 g/L、175 g/L、200 g/L、225 g/L、250 g/L、275 g/L、300 g/L、325 g/L、350 g/L、375 g/L、400 g/L、450 g/L、500 g/L、550 g/L或600 g/L;并且其中将所述溶液的pH设置在3至7,并且其中将所述进料溶液的温度保持在20℃至80℃;

所述方法导致在所述培养基的最终体积中 α -1,3-岩藻糖基乳糖浓度是至少50 g/L、至少75 g/L、至少90 g/L、至少100 g/L、至少125 g/L、至少150 g/L、至少175 g/L或至少200 g/L,并且任选地在所述培养物的最终体积中 α -1,3-岩藻糖基乳糖纯度是80%或更高。

生产3-岩藻糖基乳糖和转化乳糖的 α -1,3-岩藻糖基转移酶

技术领域

[0001] 本发明涉及生产3-岩藻糖基乳糖 (3-FL) 的方法以及新鉴定的岩藻糖基转移酶,更具体地新鉴定的乳糖结合 α -1,3-岩藻糖基转移酶多肽,以及它们的应用。此外,本发明提供了使用新鉴定的乳糖结合 α -1,3-岩藻糖基转移酶生产3-岩藻糖基乳糖 (3-FL) 的方法。

技术背景

[0002] 现在已经有80多种属于人乳寡糖 (HMO) 家族的化合物得到结构表征。这些HMO代表一类作为益生元起作用的复杂寡糖。此外,HMO与上皮表位的结构同源性说明了针对对细菌病原体的保护作用。在婴儿胃肠道内,HMO选择性地滋养选定细菌菌株的生长,并因此,在母乳喂养的婴儿中引发独特肠道微生物群的发展。

[0003] 这些人乳寡糖中的一些需要存在特定的岩藻糖基化的结构,这些结构很可能表现出特定的生物活性。生产这些岩藻糖基化的寡糖需要岩藻糖基转移酶的作用。这些岩藻糖基转移酶属于糖基转移酶家族,广泛表达于脊椎动物、无脊椎动物、植物、真菌、酵母和细菌中。它们催化岩藻糖残基从供体 (通常是鸟苷二磷酸岩藻糖 (GDP-岩藻糖)) 转移至受体 (acceptor),包括寡糖、糖 (蛋白) 和糖 (脂)。因此,岩藻糖基化的受体底物参与了多种生物和病理过程。

[0004] 已经鉴定出几种岩藻糖基转移酶,例如在细菌幽门螺杆菌 (*Helicobacter pylori*)、大肠杆菌 (*Escherichia coli*)、肠道沙门氏菌 (*Salmonella enterica*)、哺乳动物、秀丽隐杆线虫 (*Caenorhabditis elegans*) 和曼氏血吸虫 (*Schistosoma mansoni*) 以及植物中鉴定到岩藻糖基转移酶。

[0005] 岩藻糖基转移酶基于岩藻糖添加的位点分类为例如 α -1,2-、 α -1,3-、 α -1,4和O-岩藻糖基转移酶。

[0006] 本领域已经描述了几种 α -1,3-岩藻糖基转移酶。WO 1998/055630描述了幽门螺杆菌的细菌 α -1,3-岩藻糖基转移酶基因,其可用于生产寡糖,例如Lewis X、Lewis Y和唾液酸Lewis X。WO 2016/040531描述了几种用于生产岩藻糖基化的寡糖的 α -1,3-岩藻糖基转移酶。在这里, α -1,3-岩藻糖基转移酶被描述为与诺德拟杆菌 (*Bacteroides nordii*) CafC酶具有25%至100%的序列同一性。然而,在该文件的表1中,作者清楚地表明,他们测试的酶 (其中许多与CafC具有>25%的序列同一性) 中超过一半 (7/12) 不能使用乳糖作为受体底物产生3-岩藻糖基乳糖。这说明明显地并非所有假定的岩藻糖基转移酶都具有乳糖结合岩藻糖基转移酶活性。WO2012/049083描述了一些新的 α -1,3-岩藻糖基转移酶及其在生产岩藻糖基化的产物中的用途。Huang等人2017对各种外源性 α -1,3-岩藻糖基转移酶候选物以及一系列大肠杆菌宿主菌株进行了比较,并证明使用大肠杆菌BL21 (DE3) 为宿主菌株的幽门螺杆菌futA产生最高滴度的3-岩藻糖基乳糖 (人乳寡糖之一)。

[0007] 通常,已知 α -1,3-岩藻糖基转移酶 (也称为3-岩藻糖基转移酶) 对乳糖的亲合力低。生产HMO 3-岩藻糖基乳糖需要3-岩藻糖基转移酶。低亲合力对3-岩藻糖基乳糖的生产率有负面影响。为了提高转化率和生产率,需要具有足够乳糖亲和力、优选地更高乳糖亲和

力的转移酶。

[0008] 因此,本发明的一个目的是提供工具和方法,通过这些工具和方法,可以以高效的、有时间和成本效益的方式生产或合成3-岩藻糖基乳糖,并且与现有技术方法相比产生类似或更高数量的所需产物。

发明内容

[0009] 令人惊讶的是,现在已经发现,本发明的新鉴定的乳糖结合 α -1,3-岩藻糖基转移酶提供与目前已知的乳糖结合 α -1,3-岩藻糖基转移酶相比具有相似或更高的乳糖结合和/或转移酶性质的转移酶。

[0010] 因此,本发明提供使用新鉴定的乳糖结合 α -1,3-岩藻糖基转移酶生产3-岩藻糖基乳糖(3FL)的方法。3FL可通过在 α -1,3-岩藻糖基转移酶存在下使乳糖反应而获得, α -1,3-岩藻糖基转移酶能够催化从乳糖和GDP-岩藻糖形成3-岩藻糖基乳糖寡糖。备选地,也可以从产生根据本发明的 α -1,3-岩藻糖基转移酶的微生物中获得。

[0011] 定义

[0012] 本说明书中用于描述本发明及其各种实施方式的词语不仅应理解为其常规定义的含义,而且应通过在本说明书中的特殊定义包括超出常规定义含义范围的结构、材料或行为。因此,如果在本说明书的上下文中可以将一个元素理解为包括一个以上的含义,那么其在权利要求中的使用必须理解为对说明书和该词本身支持的所有可能含义的通用。

[0013] 本文公开的本发明的各种实施方案和实施方案的方面不仅应按照本说明书中具体描述的顺序和上下文来理解,而且还应包括其任何顺序及任何组合。当上下文需要时,所有以单数使用的词语应视为包含复数,反之亦然。除非另有定义,否则本文中使用的所有技术和科学术语通常具有与本发明所属领域的普通技术人员通常理解的相同含义。一般而言,本文所使用的术语以及本文所描述的细胞培养、分子遗传学、有机化学和核酸化学及杂交中的实验室程序是本领域公知且常用的。核酸和肽的合成使用标准技术。通常,酶促反应和纯化步骤根据制造商的说明书进行。

[0014] 在附图和说明书中,已经公开了本发明的实施方案,并且尽管使用了特定术语,但是这些术语仅在描述性意义上使用,而不是为了限制的目的,本发明的范围在后附权利要求书中阐述。必须理解,所示出的实施方案仅为示例的目的而阐述,并且不应将其视为限制本发明。对于本领域的技术人员显而易见的是,可以与本公开内容的文字和精神一致以及在本公开内容的范围内进行修改、其他实施方案、改进、细节和使用,根据专利法解释本公开内容仅受权利要求书的限制,包括等同原则。在随后的权利要求书中,提供用于指定权利要求步骤的参考字符只是为了便于描述,并不意在暗示执行步骤的任何特定顺序。

[0015] 根据本发明,术语“多核苷酸”通常指任何多核糖核苷酸或多脱氧核糖核苷酸,其可以是未修饰的RNA或DNA或者修饰的RNA或DNA。“多核苷酸”包括但不限于单链和双链DNA,作为单链和双链区域或单链、双链和三链区域的混合物的DNA,单链和双链RNA,以及作为单链和双链区域的混合物的RNA,包含DNA和RNA(其可以是单链,或更典型的双链或三链区域,或单链和双链区域的混合物)的杂交分子。此外,本文所使用的“多核苷酸”指包含RNA或DNA或者RNA和DNA两者的三链区域。这些区域中的链可以来自同一个分子或来自不同的分子。这些区域可以包括所有的一个或多个分子,但更典型地只涉及一些分子的区域。三螺旋区

域的分子之一通常是寡核苷酸。如本文所用,术语“多核苷酸”还包括如上所述的含有一个或多个修饰碱基的DNA或RNA。因此,具有出于稳定性或其他原因而修饰的主链的DNA或RNA是根据本发明的“多核苷酸”。此外,包含不寻常碱基(例如肌苷)或修饰的碱基(例如三酰化碱基)的DNA或RNA应理解为涵盖在术语“多核苷酸”中。应当理解,已经对DNA和RNA进行了多种修饰,其用于本领域技术人员已知的许多有用目的。本文中使用的术语“多核苷酸”包括这种经化学、酶或代谢修饰的多核苷酸形式,以及病毒和细胞(包括例如简单和复杂细胞)所特有的DNA和RNA的化学形式。术语“多核苷酸”也包括通常称为寡核苷酸的短多核苷酸。

[0016] “多肽”是指包含通过肽键或修饰的肽键彼此连接的两个或更多个氨基酸的任何肽或蛋白质。“多肽”既指短链(通常称为肽、寡肽和寡聚物),也指长链(通常称为蛋白质)。多肽可以含有20种基因编码氨基酸以外的氨基酸。“多肽”包括通过自然过程(例如加工和其他翻译后修饰)修饰的多肽,也包括通过化学修饰技术修饰的多肽。这样的修饰在基础教科书和更详细的专著中以及多卷研究文献中充分描述,并且其对于本领域技术人员是公知的。相同类型的修饰可以以相同的程度或不同的程度存在于给定多肽中的数个位点上。此外,给定的多肽可以包含许多类型的修饰。修饰可以在多肽的任意位置发生,包括肽主链、氨基酸侧链和氨基末端或羧基末端。修饰包括,例如,乙酰化、酰化、ADP-核糖基化、酰胺化、黄素的共价连接、血红素部分的共价连接、核苷酸或核苷酸衍生物的共价连接、脂质或脂质衍生物的共价连接、磷脂酰肌醇的共价连接、交联、环化、二硫键形成、去甲基化作用、共价交联的形成、焦谷氨酸的形成、甲酰化作用、 γ -羧基化、糖基化、GPI锚定形成、羟基化、碘化、甲基化、豆蔻酰化作用、氧化、蛋白水解加工、磷酸化、异戊二烯化、外消旋化、脂质连接、硫化、谷氨酸残基的 γ -羧基化和ADP-核糖基化、硒化(selenoylation)、转移RNA介导的向蛋白添加氨基酸(诸如精氨酸化)和泛蛋白化。多肽可以是分支的或有或无分支的环状。环状、分支的和分支环状的多肽可以由翻译后天然过程形成,并且也可以通过完全合成法制得。

[0017] “分离的”意指“通过人工”由其天然状态改变,即,如果其天然存在,则其已经改变或从其原始环境中移出,或者二者。例如,天然存在于活的生物体中的多核苷酸或多肽不是“分离的”,但是与其天然状态的共存物质分开的相同的多核苷酸或多肽是“分离的”,如该术语在本文中所用的。类似地,如本文所使用的术语“合成的”序列是指合成产生而不是从天然来源直接分离的任意序列。如本文所使用的术语“合成的”是指任何合成生成的序列,并且不是直接从天然来源分离出来的。

[0018] “重组”意指通过将来自一个物种的基因移植或剪接到不同物种的宿主生物体的细胞中而制备的基因工程化的DNA。这种DNA成为宿主基因组成的一部分并被复制。在本发明的上下文中使用的“突变”细胞或微生物是指经基因工程化或具有改变的基因组成的细胞或微生物。

[0019] 在本发明的上下文中,术语“经基因修饰用于生产3-岩藻糖基乳糖的细胞”是指微生物的细胞,其经基因操作以包含以下中的至少一种:i) 编码合成所述3-岩藻糖基乳糖所必需的 α 1,3岩藻糖基转移酶的重组基因,ii) 产生适于通过所述岩藻糖基转移酶转移至乳糖的GDP岩藻糖的生物合成途径,和/或iii) 产生乳糖的生物合成途径或乳糖从培养基内化到细胞中的机制,在细胞中乳糖被岩藻糖基化以产生3-岩藻糖基乳糖。

[0020] 术语“编码用于3-岩藻糖基乳糖合成的酶的核酸序列”涉及编码用于3-岩藻糖基

乳糖合成途径中必需的酶的核酸序列,例如能够将GDP-岩藻糖供体底物的岩藻糖部分转移到乳糖的半乳糖部分的3羟基上从而产生3-岩藻糖基乳糖的酶。

[0021] 在本公开内容的上下文中,术语“内源性”是指任何多核苷酸、多肽或蛋白质序列,其是细胞的天然部分并且存在于其在细胞染色体中的天然位置。术语“外源性”是指任何多核苷酸、多肽或蛋白质序列,其来源于所研究的细胞外部,并且不是细胞的天然部分,或不存在于细胞染色体或质粒中的其天然位置。

[0022] 术语“异源的”当用于提及多核苷酸、基因、核酸、多肽或酶时,是指来自或衍生自宿主物种以外的来源的多核苷酸、基因、核酸、多肽或酶。相反,本文使用“同源”多核苷酸、基因、核酸、多肽或酶来表示衍生自宿主生物体物种的多核苷酸、基因、核酸、多肽或酶。当涉及用于维持或操纵基因序列的基因调控序列或辅助核酸序列时(例如,启动子、5'非翻译区、3'非翻译区、poly A附加序列、内含子序列、剪接位点、核糖体结合位点、内部核糖体进入序列、基因组同源区、重组位点等),“异源”是指调控序列或辅助序列与在构建体、基因组、染色体或附加体中与调控或辅助核酸序列并列的基因不天然相关联。因此,可操作地连接至在其天然状态下(即在非基因工程生物体的基因组中)不是可操作地连接至的基因的启动子在本文中被称为“异源启动子”,即使该启动子可衍生自与其所连接的基因相同的物种(或在某些情况下,同一个生物体)。

[0023] 如本文所使用的术语“编码多肽的多核苷酸”包括包含编码本发明多肽的序列的多核苷酸,所述多肽尤其是具有所附序列表中SEQ ID NO 2、4、6、8、10、12、14、16、20、22、28、30或32所示氨基酸序列的 α -1,3-岩藻糖基转移酶。为清楚起见,编码SEQ ID NO 18、24和26的多肽的多核苷酸也是该定义所涵盖的多核苷酸,但是SEQ ID NO 18的多核苷酸是用作参考的现有技术 α -1,3-岩藻糖基转移酶,并且SEQ ID NO 24和26的多核苷酸是对乳糖作为受体不起作用的 α -1,3-岩藻糖基转移酶。该术语还包括多核苷酸,所述多核苷酸包括编码所述多肽的单个连续区域或不连续区域(例如,被整合的噬菌体或插入序列或编辑所间隔)以及也可包含编码和/或非编码序列的附加区域。

[0024] 在本文中使用的术语“变体”是分别与参照多核苷酸或多肽不同但是保留基本特性的多核苷酸或多肽。多核苷酸的典型变体在核苷酸序列上与另一种参照多核苷酸不同。变体的核苷酸序列中的变化可以改变或不改变由参照多核苷酸编码的多肽的氨基酸序列。核苷酸变化可以导致由参照序列编码的多肽中的氨基酸取代、添加、缺失、融合和截短,如下文所讨论那样。多肽的典型变体在氨基酸序列上与另一种参照多肽不同。通常,限制差异以使参照多肽和变体的序列在整体上紧密相似,并且在许多区域中是相同的。变体和参照多肽可以在氨基酸序列中有以任意组合的一个或多个取代、添加、缺失的不同。取代或插入的氨基酸残基可以是或可以不是由遗传密码编码的氨基酸残基。多核苷酸或多肽的变体可以是天然存在的,诸如等位基因变体,或者其可以是不已知天然存在的变体。多核苷酸和多肽的非天然存在的变体可以通过诱变技术、通过直接合成和通过本领域技术人员已知的其他重组方法制得。在一些实施方案中,本发明设想通过修饰本发明中使用的膜蛋白的结构来制备功能性变体。可以通过氨基酸取代、缺失、添加或其组合来产生变体。例如,有理由期望用异亮氨酸或缬氨酸单独替换亮氨酸、用谷氨酸单独替换天冬氨酸、用丝氨酸单独替换苏氨酸、或用结构上相关的氨基酸对氨基酸的相似替换(例如,保守突变)不会对所得分子的生物活性产生重大影响。保守替换是发生在与其侧链相关的氨基酸家族内的替

换。通过评估变体多肽以类似于野生型多肽的方式在细胞中产生应答的能力,在本发明的情况下,比没有变体的细胞提供更好的产率、生产率和/或生长速度,可以容易地确定本公开内容的多肽的氨基酸序列中的改变是否导致功能同源物。

[0025] 本文所使用的术语“功能同源物”描述具有序列相似性并且还共享诸如生化活性的至少一个功能特征的那些分子。功能同源物通常对于相同的特征产生相似的,但不一定相同的程度。功能上同源的蛋白质具有相同的特征,其中一个同源物产生的定量测量值为另一个的至少10%;更典型,至少为20%,在约30%和约40%之间;例如,在约50%和约60%之间;在约70%和约80%之间;或者在约90%和约95%之间;在约98%和约100%之间,或者超过原始分子所产生的定量测量值的100%。因此,当分子具有酶活性时,功能同源物将具有与原始酶相比的上述酶活性百分比。如果分子是DNA结合分子(例如,多肽),则同源物将具有上述结合亲和力百分比,通过结合分子的重量与原始分子相比进行测量。

[0026] 功能同源物和参考多肽可能是天然存在的多肽,并且序列相似性可能是由于趋同或发散的进化事件造成的。功能同源物有时被称为直系同源物,其中“直系同源物”是指在另一物种中与参考基因或蛋白质功能等同的同源基因或蛋白质。

[0027] 功能同源物可以通过核苷酸和多肽序列比对分析来鉴定。例如,对核苷酸或多肽序列的数据库执行查询可以鉴定生物量调节多肽的同源物。序列分析可以涉及使用生物量调节多肽的氨基酸序列作为参照序列的非冗余数据库的BLAST、反向BLAST或PSI-BLAST分析。在某些情况下,氨基酸序列是从核苷酸序列推导出来的。通常,数据库中那些序列同一性大于40%的多肽是进一步评估作为生物量调节多肽的适宜性的候选物。氨基酸序列相似性允许保守的氨基酸替换,例如一个疏水性残基替换另一个疏水性残基或者一个极性残基替换另一个极性残基。如果需要,可以对这类候选物进行人工检查,以缩小待进一步评估的候选物的数量。人工检查可以通过选择那些似乎具有在生产率调控多肽中存在的结构域(例如,保守的功能结构域)的候选物来执行。

[0028] “片段”,就多核苷酸而言,是指克隆或多核苷酸分子的任何部分,特别是保留可用的功能特征的多核苷酸的部分。有用的片段包括寡核苷酸和多核苷酸,它们可用于杂交或扩增技术或者复制、转录或翻译的调控。“多核苷酸片段”是指多核苷酸的任何子序列,通常具有本文提供的任何序列的至少约9个连续核苷酸,例如至少约30个核苷酸或至少约50个核苷酸。示例性片段可另外或备选地包括包含编码多肽的保守家族结构域的区域、基本上由其组成或由其组成的片段。示例性片段可另外或备选地包括包含多肽的保守结构域的片段。

[0029] 片段可另外或备选地包括多肽和蛋白质分子的子序列,或多肽的子序列。在某些情况下,片段或结构域是多肽的子序列,其以与完整多肽基本相同的方式或类似程度执行完整多肽的至少一种生物功能。例如,多肽片段可包含可识别的结构基序或功能结构域,例如DNA结合位点或结构域,其与DNA启动子区、激活结构域或用于蛋白质-蛋白质相互作用的结合域结合,并且可启动转录。片段的大小可以从少至3个氨基酸残基到完整多肽的全长不等,例如长度至少约20个氨基酸残基,例如长度至少约30个氨基酸残基。优选地,片段是具有衍生片段的多肽的至少一种性质或活性的功能片段,例如,片段可以包括多肽的功能结构域或保守结构域。例如,可以通过Pfam或保守结构域数据库(CDD)指定来表征结构域。

[0030] 本发明中使用的术语“alpha-1,3-岩藻糖基转移酶”、“alpha 1,3-岩藻糖基转移

酶”、“3-岩藻糖基转移酶”、“ α -1,3-岩藻糖基转移酶”、“ α 1,3岩藻糖基转移酶”、“3岩藻糖基转移酶”、“3-FT”或“3FT”可交换使用,并指催化岩藻糖从供体底物GDP-L-岩藻糖转移至受体分子乳糖在 α -1,3-键中的糖基转移酶。编码“ α -1,3-岩藻糖基转移酶”或上述任一术语的多核苷酸是指编码这种糖基转移酶的多核苷酸,其催化岩藻糖从供体底物GDP-L-岩藻糖转移至受体分子乳糖在 α -1,3-键中。

[0031] 本发明中使用的术语“3-岩藻糖基乳糖”、“ α 1,3-岩藻糖基乳糖”、“ α 1,3岩藻糖基乳糖”、“ α -1,3-岩藻糖基乳糖”、“ α 1,3岩藻糖基乳糖”、“Gal β -4 (Fuc α 1-3) Glc”、“3FL”或“3-FL”可互换使用,并指通过 α -1,3-岩藻糖基转移酶催化岩藻糖残基从GDP-L-岩藻糖转移至乳糖在 α -1,3-键中而获得的产物。

[0032] “寡糖”作为本文中使用的术语并且如在现有技术中通常理解的,是指含有少量(通常为3到10个)简单糖(即单糖)的糖聚合物。

[0033] 术语“纯化的”是指实质上或基本上不含干扰生物分子的活性的组分材料。对于细胞、糖类、核酸和多肽,术语“纯化的”是指实质上或基本上不含通常在材料以其天然状态存在时伴随材料的组分材料。典型地,本发明的纯化的糖类、寡糖、蛋白质或核酸是通过银染凝胶上的条带强度或其他测定纯度的方法测量的至少约50%、55%、60%、65%、70%、75%、80%或85%纯度,通常至少约为90%、91%、92%、93%、94%、95%、96%、97%、98%或99%纯度。纯度或均一性可通过本领域公知的许多方法来指示,例如蛋白质或核酸样品的聚丙烯酰胺凝胶电泳,随后在染色时可视化。出于某些目的,需要高分辨率,并使用HPLC或类似的纯化方法。对于寡糖,例如3-岩藻糖基乳糖,可使用诸如但不限于薄层色谱、气相色谱、NMR、HPLC、毛细管电泳或质谱的方法来测定纯度。

[0034] 术语“相同”或百分比“同一性”或%“同一性”在两个或更多个核酸或多肽序列的情形中,是指两个或更多个序列或子序列,当用序列比较算法或目测法测量就最大对应性进行比较和比对时,其是相同的或具有特定百分比的相同的氨基酸残基或核苷酸。对于序列比较,一个序列作为参照序列,将测试序列与之进行比较。当使用序列比较算法时,将测试序列和参照序列输入计算机,必要时指定子序列坐标,并指定序列算法程序参数。然后,序列比较算法根据指定的程序参数,计算测试序列相对于参照序列的序列同一性百分比。可使用BLAST和PSI-BLAST确定同一性百分比(Altschul等人,1990,J Mol Biol 215:3, 403-410;Altschul等人,1997,Nucleic Acids Res 25:17,3389-402)。为了本发明的目的,使用MatGAT2.01确定同一性百分比(Campanella等人,2003,BMC Bioinformatics 4:29)。MatGAT使用Myers和Miller全局比对算法进行配对比对。使用蛋白质的如下默认参数:(1)缺口罚分存在:12和延伸:2;(2)采用的矩阵为BLOSUM50。

[0035] 术语“控制序列”是指由宿主细胞转录和翻译系统识别的序列,允许将多核苷酸序列转录和翻译成多肽。因此,这种DNA序列对于在特定宿主细胞或生物体中表达可操作连接的编码序列是必需的。这种控制序列可以是但不限于启动子序列、核糖体结合序列、Shine Dalgarno序列、Kozak序列、转录终止子序列。例如,适于原核生物的控制序列包括启动子、任选的操纵子序列和核糖体结合位点。已知真核细胞利用启动子、多聚腺苷酸化信号和增强子。如果前序列或可分泌前导物的DNA作为参与多肽分泌的前蛋白表达,则将前序列或可分泌前导物的DNA可操作地连接到多肽的DNA;如果启动子或增强子影响序列的转录,则将启动子或增强子可操作地连接到编码序列;或者如果核糖体结合位点影响序列的转录,则

将核糖体结合位点可操作地连接到编码序列;或者如果核糖体结合位点的定位便于翻译,则将核糖体结合位点可操作地连接到编码序列。所述控制序列还可利用外部化学品(例如但不限于IPTG、阿拉伯糖、乳糖、别乳糖、鼠李糖或岩藻糖)经由可诱导启动子或经由诱导或抑制所述多核苷酸转录或翻译为多肽的遗传回路而得到另外控制。

[0036] 本发明中使用的术语“发酵结束”是指收获发酵物以进行产品纯化的时机。

[0037] 一般来说,“可操作连接”是指被连接的DNA序列是连续的,并且,在可分泌前导物的情况下,是连续的并且处于阅读框中。但是,增强子不必是连续的。

[0038] 发明详述

[0039] 根据第一实施方案,本发明提供一种生产 α -1,3-岩藻糖基乳糖的方法。该方法包括以下步骤:

[0040] a) 提供一种具有 α -1,3-岩藻糖基转移酶活性且具有使用乳糖作为受体底物的能力的多肽,其中所述多肽包含

[0041] i) 编码保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)的氨基酸序列;

[0042] ii) 编码保守的[K/D][L/K/M]XXX[F/Y]结构域(SEQ ID NO 34)的氨基酸序列,以及

[0043] iii) 其中如果ii)的结构域等于DM[A/S]VSF(SEQ ID NO 36),则另外地在N-末端区域存在保守的基序[N/H]XDPAXLD(SEQ ID NO 35);

[0044] 其中X可以是任何不同的氨基酸;并且

[0045] 其中,所述多肽的C-末端具有从GDP-岩藻糖结合结构域的第一个氨基酸开始的小于或等于100个氨基酸;

[0046] b) 在其中所述多肽催化岩藻糖残基从供体底物转移至受体底物的条件下,将步骤a)中具有 α -1,3-岩藻糖基转移酶活性的所述多肽与包含GDP岩藻糖作为供体底物和乳糖作为受体底物的混合物接触,由此产生 α -1,3-岩藻糖基乳糖。

[0047] 相对于目前已知的 α -1,3-岩藻糖基转移酶,包含上述结构域中的两个(或SEQ ID NO 33至36的全部,视情况而定)的这些新鉴定的多肽提供了一种具有使用乳糖作为受体底物的能力的替代型 α -1,3-岩藻糖基转移酶。包含上述结构域中的两个(或SEQ ID NO 33至36的全部,视情况而定)的多肽提供了与目前已知的 α -1,3-岩藻糖基转移酶相比具有相似或更高乳糖结合和/或相似或更高转移酶性质的转移酶。

[0048] 在本发明的第一优选实施方案中,用于本发明的多肽包含具有SEQ ID NO33到34或36的结构域中的两个(或SEQ ID NO 33至36的全部,视情况而定)并且其中所述SEQ ID NO 33是具有氨基酸序列YXTEK(SEQ ID NO:37)的保守结构域,其中X可以是任何不同的氨基酸。

[0049] 在本发明的第二优选实施方案中,用于本发明的多肽包含具有SEQ ID NO33至34或36的结构域中的两个(或SEQ ID NO 33至36的全部,视情况而定)并且其中所述SEQ ID NO 34是具有氨基酸序列[K/D]LX[I/L/M]G[F/Y](SEQ ID NO:38)、[K/D][L/K]xL[S/G][F/Y](SEQ ID NO:39)或[K/D]LXLG[F/Y](SEQ ID NO:40)的保守的结构域,其中X可以是任何不同的氨基酸。

[0050] 使用一些新鉴定具有使用乳糖作为受体底物的能力和具有 α -1,3-岩藻糖基转移

酶活性并且具有新鉴定的结构域的多肽的另一个优点在于在反应或发酵结束时,以比使用具有SEQ ID NO 18的参照现有技术多肽获得的纯度更高的纯度产生3-岩藻糖基乳糖,这是由于新鉴定的3-岩藻糖基转移酶具有更好的转化能力,可以将乳糖用于3FL生产。更具体地说,乳糖浓度与3-岩藻糖基乳糖浓度比率小于1:5、优选地小于1:10、更优选地小于1:20、优选地小于1:40。在另一优选实施方案中,3-岩藻糖基乳糖纯度在发酵结束时为80%或更高。

[0051] 根据本发明,生产 α -1,3-岩藻糖基乳糖的方法可在无细胞系统或含有细胞的系统中进行。使底物GDP岩藻糖和乳糖与 α -1,3-岩藻糖基转移酶多肽反应足够的时间并在足够的条件下反应,以形成酶产物。这些条件将取决于底物和酶的数量和纯度,以及系统是无细胞系统还是基于细胞的系统而变化。本领域技术人员将容易地调整这些变量。

[0052] 在无细胞系统中,根据本发明的多肽、受体底物、供体底物和视情况而定的其他反应混合物成分(包括其他糖基转移酶和辅助酶)通过混合在水性反应介质中组合以执行酶促反应。酶可以在溶液中自由利用,或者可以结合或固定化在支持物(例如聚合物)上,并且底物可以添加到支持物上。例如,支持物可以包装在柱中。

[0053] 本文所述的用于合成3-岩藻糖乳糖的含细胞系统或基于细胞的系统可包括经基因修饰的宿主细胞。根据本发明的一个方面,具有 α -1,3-岩藻糖基转移酶活性的多肽由产生该多肽的细胞(例如,如本文所述的宿主细胞)产生。根据本发明的另一方面,GDP-岩藻糖和/或乳糖由生产所述GDP-岩藻糖和/或乳糖的细胞提供。细胞可以是宿主细胞,其也产生 α -1,3-岩藻糖基转移酶。备选地,细胞可以是产生 α -1,3-岩藻糖基转移酶的宿主细胞之外的另一细胞,在这种情况下,技术人员将讨论细胞偶联。这种生产GDP-岩藻糖的细胞可以表达将例如要添加到宿主细胞的岩藻糖转化为GDP-岩藻糖的酶。这种酶可以是,例如,双功能岩藻糖激酶/岩藻糖-1-磷酸鸟苷酰转移酶,如脆弱拟杆菌(*Bacteroides fragilis*)的Fkp,或一个单独的岩藻糖激酶与一个单独的岩藻糖-1-磷酸鸟苷酰转移酶的组合,如已知的,它们来自于几个物种,包括智人(*Homo sapiens*)、猪(*Sus scrofa*)和褐家鼠(*Rattus norvegicus*)。

[0054] 在另一实施方案中,本发明涉及一种生产 α -1,3-岩藻糖基乳糖的方法,包括以下步骤:

[0055] i) 提供一种经基因修饰用于生产 α -1,3-岩藻糖基乳糖的细胞,所述细胞包含至少一个编码用于 α -1,3-岩藻糖基乳糖合成的酶的核酸序列,所述细胞包含具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽的表达,其中所述多肽包含:

[0056] a) 编码保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)的氨基酸序列;

[0057] b) 编码保守的[K/D][L/K/M]XXX[F/Y]结构域(SEQ ID NO 34)的氨基酸序列,以及

[0058] c) 其中如果b)的结构域等于DM[A/S]VSF(SEQ ID NO 36),则在N-末端区域存在保守的基序[N/H]XDPAXLD(SEQ ID NO 35);

[0059] 其中X可以是任何不同的氨基酸;并且

[0060] 其中所述多肽的C-末端具有从GDP-岩藻糖结合结构域的第一个氨基酸开始的小于或等于100个氨基酸,以及

[0061] ii) 在允许生产 α -1,3-岩藻糖基乳糖的条件下,在培养基中培养细胞,从而产生3-FL。

[0062] 在另一个实施方案中,本发明涉及一种生产 α -1,3-岩藻糖基乳糖的方法,该方法包括以下步骤:

[0063] a) 提供表达本文所定义的具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的所述多肽的宿主细胞;

[0064] b) 在允许产生 α -1,3-岩藻糖基乳糖并允许表达具有 α -1,3-岩藻糖基转移酶活性的多肽的合适营养条件下使所述宿主细胞生长;

[0065] c) 在步骤b)的同时或之后提供供体底物GDP-岩藻糖和受体底物乳糖,以使 α -1,3-岩藻糖基转移酶多肽催化岩藻糖残基从GDP-岩藻糖转移至乳糖,从而生产 α -1,3-岩藻糖基乳糖。

[0066] 然后任选地从宿主细胞和/或其生长的培养基分离产生的3FL。

[0067] 根据又一个实施方案,本文所述方法中所述3-岩藻糖基乳糖的生产是通过细胞对编码 α -1,3-岩藻糖基转移酶的多核苷酸的异源或同源(过)表达来进行的。

[0068] 在本文所述的本发明方法中,宿主细胞可以经转化或转染以表达本文所述的并具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的外源多肽。因此,本发明涉及一种使用宿主细胞生产 α -1,3-岩藻糖基乳糖的方法,包括以下步骤:

[0069] a) 使经转化或转染以表达具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的外源多肽的宿主细胞生长,其中所述多肽如本文所述;和

[0070] b) 在步骤a)的同时或之后,提供供体底物GDP-岩藻糖和受体底物乳糖,其中 α -1,3-岩藻糖基转移酶多肽催化岩藻糖残基从供体底物转移至受体底物,从而产生 α -1,3-岩藻糖基乳糖。

[0071] 优选地,如本文所使用的具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的外源多肽产生3FL,其在发酵结束时乳糖浓度与3FL浓度的比率小于1:5。

[0072] 乳糖浓度与3FL浓度的比率可以小于1:5、1:6、1:7、1:8、1:9、1:10、1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:21、1:22、1:23、1:24、1:25、1:26、1:27、1:28、1:29、1:30、1:31、1:32、1:33、1:34、1:35、1:36、1:37、1:38、1:39、1:40、1:45、1:50、1:55、1:60、1:65、1:70、1:75、1:80、1:85、1:90、1:95、1:100、1:110、1:120、1:130、1:140、1:150、1:160、1:170、1:180、1:190、1:200、1:300、1:400、1:500、1:600、1:700、1:800、1:900、1:1000。

[0073] 在优选实施方案中,在生产过程中获得乳糖浓度与3FL浓度的比率小于1:5、1:6、1:7、1:8、1:9、1:10、1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:21、1:22、1:23、1:24、1:25、1:26、1:27、1:28、1:29、1:30、1:31、1:32、1:33、1:34、1:35、1:36、1:37、1:38、1:39、1:40、1:45、1:50、1:55、1:60、1:65、1:70、1:75、1:80、1:85、1:90、1:95、1:100、1:110、1:120、1:130、1:140、1:150、1:160、1:170、1:180、1:190、1:200、1:300、1:400、1:500、1:600、1:700、1:800、1:900、1:1000,导致最终乳糖浓度低于25g/L、20g/L、15g/L、10g/L、9g/L、8g/L、7g/L、6g/L、5g/L、4g/L、3g/L、2g/L、1g/L、0.5g/L、0.25g/L、0.1g/L或0g/L。

[0074] 在另一个实施方案中,在生产过程中获得乳糖浓度与3FL浓度的比率小于1:5、1:6、1:7、1:8、1:9、1:10、1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:21、1:22、1:23、1:24、1:25、1:26、1:27、1:28、1:29、1:30、1:31、1:32、1:33、1:34、1:35、1:36、1:37、1:38、1:39、1:40、1:45、1:50、1:55、1:60、1:65、1:70、1:75、1:80、1:85、1:90、1:95、1:

100、1:110、1:120、1:130、1:140、1:150、1:160、1:170、1:180、1:190、1:200、1:300、1:400、1:500、1:600、1:700、1:800、1:900、1:1000,其中乳糖浓度在底物限制条件下进料,其中底物限制定义为决定底物的转化率的在生物反应器中的浓度。

[0075] 在另一个实施方案中,在生产过程中获得乳糖浓度和3FL浓度的比率小于1:5、1:6、1:7、1:8、1:9、1:10、1:11、1:12、1:13、1:14、1:15、1:16、1:17、1:18、1:19、1:20、1:21、1:22、1:23、1:24、1:25、1:26、1:27、1:28、1:29、1:30、1:31、1:32、1:33、1:34、1:35、1:36、1:37、1:38、1:39、1:40、1:45、1:50、1:55、1:60、1:65、1:70、1:75、1:80、1:85、1:90、1:95、1:100、1:110、1:120、1:130、1:140、1:150、1:160、1:170、1:180、1:190、1:200、1:300、1:400、1:500、1:600、1:700、1:800、1:900、1:1000,其中乳糖在限速条件下在细胞中形成。

[0076] 在另一个实施方案中,基于肉汤中(乳糖和3FL)总和的3-岩藻糖基乳糖在肉汤中的纯度高于约80%,例如80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%、99.6%、99.7%、99.8%、99.9%。如本文所用,3-岩藻糖基乳糖纯度定义为3FL浓度与3FL浓度和乳糖浓度之和的比率($[3FL]/([3FL]+[乳糖])$)。

[0077] 根据本发明,可在发酵培养基或水性培养基中将GDP-岩藻糖和/或乳糖进料给宿主细胞。备选地,GDP-岩藻糖和/或乳糖可由在宿主细胞中同时表达的酶或由宿主细胞的代谢来提供。因此,宿主细胞也将在GDP-岩藻糖和/或乳糖后产生 α -1,3-岩藻糖基转移酶。在另一实施方案中,GDP-岩藻糖和/或乳糖可由作为产生 α -1,3-岩藻糖基转移酶的宿主细胞之外的另一细胞的细胞产生,在这种情况下,本领域技术人员将谈论细胞偶联。这种产生GDP-岩藻糖的细胞可以表达将例如要添加到宿主细胞的岩藻糖转化为GDP-岩藻糖的酶。这种酶例如可以是双功能岩藻糖激酶/岩藻糖-1-磷酸鸟苷酰转移酶,如脆弱拟杆菌的Fkp,或一种单独的岩藻糖激酶与一种单独的岩藻糖-1-磷酸鸟苷酰转移酶的组合,如已知它们来自几种物种,包括智人、猪和褐家鼠。

[0078] 根据又一实施方案,所述 α -1,3-岩藻糖基乳糖的产生通过如本文所述的宿主细胞进行,所述宿主细胞包含编码 α -1,3-岩藻糖基转移酶的多核苷酸的异源或同源(过度)表达。

[0079] 在另一方面中,本发明提供如本文所述的用于生产 α -1,3-岩藻糖基乳糖的方法,其中该方法进一步包括从宿主细胞或其生长的培养基分离 α -1,3-岩藻糖基乳糖的步骤。

[0080] 如本文所用,术语“分离”意指从反应混合物和/或从产生 α -1,3-岩藻糖基转移酶的细胞中收获、收集或回收根据本发明的由 α -1,3-岩藻糖基转移酶产生的 α -1,3-岩藻糖基乳糖。

[0081] 如果通过细胞或发酵制备 α -1,3-岩藻糖基乳糖,则3-FL可通过常规方式从制备混合物的水性培养基中分离。如果 α -1,3-岩藻糖基乳糖仍存在于产生 α -1,3-岩藻糖基乳糖的细胞中,则可使用从细胞中释放或提取 α -1,3-岩藻糖基乳糖的常规方法,例如使用高pH、热激、超声波、法式滤压、均化、酶促水解、化学水解、溶剂水解、洗涤剂、水解……来破坏细胞。培养基、反应混合物和/或细胞提取物一起并单独称为含3-FL的混合物,然后可进一步用于分离3-FL。这优选地涉及澄清含3-FL的混合物以去除悬浮的颗粒和污染物,特别是细胞、细胞组分,通过培养经基因修饰的细胞和/或进行酶促反应产生的不溶性代谢物和碎片。在此步骤中,可采用常规方式澄清含有3-FL的混合物。优选地,通过离心、絮凝、倾析和/或过滤

来澄清含有3-FL的混合物。从含有3-FL的混合物中分离3-FL的第二步骤优选地涉及从含有3-FL的混合物(优选地在其已被澄清之后)中去除基本上所有蛋白质以及肽、氨基酸、RNA和DNA以及可能干扰随后分离步骤的任何内毒素和糖脂。在该步骤中,蛋白质和相关杂质可按常规方式从含有3-FL的混合物中去除。优选地,通过超滤、纳米过滤、反渗透、微滤、活性炭或碳处理、切向流高效过滤、切向流超滤、亲和色谱、离子交换色谱(例如但不限于阳离子交换、阴离子交换、混合床离子交换)、疏水相互作用色谱和/或凝胶过滤(即尺寸排阻色谱),特别是通过色谱,更特别是通过离子交换色谱或疏水相互作用色谱或配体交换色谱从含有3-FL的混合物中去除蛋白质、盐、副产物、颜色和其他相关杂质。除了尺寸排阻色谱以外,蛋白质和相关杂质被色谱介质或选定的膜保留,而3-FL保留在含有3-FL的混合物中。

[0082] 通过蒸发、冻干、结晶、沉淀和/或干燥、喷雾干燥的进一步纯化步骤或无需进一步纯化步骤从反应混合物和/或培养基和/或细胞进一步分离3-FL。

[0083] 在更进一步方面,本发明还提供 α -1,3-岩藻糖基乳糖的进一步纯化。所述 α -1,3-岩藻糖基乳糖的进一步纯化可例如通过使用(活性)木炭或碳、纳米过滤、超滤或离子交换来完成,以去除任何剩余的DNA、蛋白质、LPS、内毒素或其他杂质。也可以使用乙醇等醇和含水醇混合物。另一种纯化步骤是通过结晶、蒸发或沉淀产物来完成的。另一种纯化步骤是干燥、喷雾干燥或冷冻干燥 α -1,3-岩藻糖基乳糖。

[0084] 分离的且优选地也是纯化的3-FL可用作婴儿配方食品中的补充剂并且用于治疗新生婴儿中的各种疾病。

[0085] 本发明的另一方面提供了一种方法,其中多肽和也是优选的3-FL是在和/或由本文所述的真菌、酵母、细菌、昆虫、动物和植物表达系统或细胞中产生的。表达系统或细胞选自包括细菌、酵母或真菌的列表,或指植物或动物细胞。后者细菌优选地属于变形菌(Proteobacteria)门或厚壁菌(Firmicutes)门或蓝细菌(Cyanobacteria)门或者异常球菌-栖热菌(Deinococcus-Thermus)门。属于变形菌门的后者细菌优选地属于肠杆菌科(Enterobacteriaceae),优选地属于大肠杆菌种。后者细菌优选地涉及属于大肠杆菌种的任何菌株,例如但不限于大肠杆菌B、大肠杆菌C、大肠杆菌W、大肠杆菌K12、大肠杆菌Nissle。更具体地,后者术语涉及培养的大肠杆菌菌株——称为大肠杆菌K12菌株——其非常适合于实验室环境,并且与野生型菌株不同,其已丧失了肠中繁殖的能力。大肠杆菌K12菌株的公知实例是K12野生型、W3110、MG1655、M182、MC1000、MC1060、MC1061、MC4100、JM101、NZN111和AA200。因此,本发明特别涉及如上所述突变和/或转化的大肠杆菌宿主细胞或菌株,其中所述大肠杆菌菌株是K12菌株。更具体地,大肠杆菌K12菌株是大肠杆菌MG1655。属于厚壁菌门的后者细菌优选地属于芽孢杆菌纲(Bacilli),优选地属于乳杆菌目(Lactobacilliales),其成员诸如乳酸乳杆菌(*Lactobacillus lactis*)、肠系膜明串珠菌(*Leuconostoc mesenteroides*),或芽孢杆菌目(Bacillales),其成员诸如来自芽孢杆菌属,诸如枯草芽孢杆菌(*Bacillus subtilis*)或解淀粉芽孢杆菌(*B. amyloliquefaciens*)。属于放线菌门的后者细菌优选地属于棒杆菌科(Corynebacteriaceae),其成员为谷氨酸棒杆菌(*Corynebacterium glutamicum*)或非发酵棒杆菌(*C. afermentans*),或属于链霉菌科(Streptomycetaceae),其成员为灰链霉菌(*Streptomyces griseus*)或弗雷迪链霉菌(*S. fradiae*)。后者酵母优选地属于子囊菌门(Ascomycota)或担子菌门(Basidiomycota)或半知菌门(Deuteromycota)或接合菌门(Zygomycete)。后者酵母优选地属于酵母属

(*Saccharomyces*)、毕赤酵母属(*Pichia*)、*Komagataella*、汉逊酵母属(*Hansunella*)、克鲁维酵母属(*Kluyveromyces*)、耶氏酵母属(*Yarrowia*)或斯塔摩酵母属(*Starmerella*)。后者真菌优选地属于根霉属(*Rhizopus*)、盘基网柄菌属(*Dictyostelium*)、青霉属(*Penicillium*)、毛霉属(*Mucor*)或曲霉属(*Aspergillus*)。

[0086] 根据本发明的另一方面,编码具有 α -1,3-岩藻糖基转移酶活性的多肽的多核苷酸适于相应细胞或表达系统的密码子使用。

[0087] 在另一优选实施方案中,本发明的方法使用用于包含本发明的 α -1,3-岩藻糖基转移酶的宿主细胞或微生物生长的培养基,其中培养基中的乳糖浓度在50至150g/L的范围内。培养基中的所述乳糖浓度可为50g/L、55g/L、60g/L、65g/L、70g/L、75g/L、80g/L、85g/L、90g/L、95g/L、100g/L、105g/L、110g/L、115g/L、120g/L、125g/L、130g/L、135g/L、140g/L、145g/L或150g/L。

[0088] 在另一优选实施方案中,本发明方法产生3-岩藻糖基乳糖的最终浓度范围为70g/L至200g/L。这样的3-FL浓度为70g/L、75g/L、80g/L、85g/L、90g/L、95g/L、100g/L、105g/L、110g/L、115g/L、120g/L、125g/L、130g/L、135g/L、140g/L、145g/L、145g/L、150g/L、155g/L、160g/L、165g/L、170g/L、175g/L、180g/L、185g/L、190g/L、195g/L或200g/L。培养基中较高的乳糖浓度可提供在生产方法中获得的甚至更高的3-FL最终浓度。

[0089] 在另一优选实施方案中,本发明的方法产生3FL的最终浓度范围为70g/L至200g/L,如上所述,并且其中所述肉汤中的3FL纯度为80%或更高。根据本发明的3FL纯度为至少约80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%、99.6%、99.7%、99.8%、99.9%。

[0090] 在如本文所述的本发明方法中,具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽包含:

[0091] a) 编码保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)的氨基酸序列;

[0092] b) 编码保守的[K/D][L/K/M]XXX[F/Y]结构域(SEQ ID NO 34)的氨基酸序列,和

[0093] c) 其中如果b)的结构域等于DM[A/S]VSF(SEQ ID NO 36),则另外地在N-末端区域存在保守的基序[N/H]XDPAXLD(SEQ ID NO 35);

[0094] 其中X可以是任何不同的氨基酸;并且

[0095] 其中所述多肽的C-末端具有从所述GDP-岩藻糖结合结构域的第一个氨基酸开始的少于或等于100个氨基酸。

[0096] 在本发明的范围内,所述多肽被证明具有乳糖结合 α -1,3-岩藻糖基转移酶活性,并且与目前已知的 α -1,3-岩藻糖基转移酶相比优选地具有更好的乳糖转化效率。

[0097] 在本发明的优选实施方案中,所述具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽包含选自以下项目组成的组的氨基酸序列:

[0098] i) 所附序列表中SEQ ID NO 6、2、4、8、10、12、14、16、20、22、28、30或32的任何一个;

[0099] ii) 与SEQ ID NO 2、20或22的全长氨基酸序列具有87%或更高序列同一性的氨基酸序列;

[0100] iii) 与SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个的全长氨基酸序列具有80%或更高序列同一性的氨基酸序列；

[0101] iv) SEQ ID NO 2、20或22中所示的氨基酸序列的片段，其中所述片段包含其至少45个连续氨基酸；

[0102] v) SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个所示的氨基酸序列的片段，其中所述片段包含其至少10个连续氨基酸并且具有乳糖结合 α -1,3-岩藻糖基转移酶活性。

[0103] 任选地，所述多肽通过N-末端和/或C-末端氨基酸延伸片段进一步修饰。

[0104] 本文所用多肽的氨基酸序列可以是选自所附序列表中SEQ ID NO 6、2、4、8、10、12、14、16、20、22、28、30或32的序列。氨基酸序列也可以是与SEQ ID NO 2、20或22中任何一个的全长氨基酸序列具有大于约87%序列同一性、87%、88%、89%、90%、91%、92%、93%、94%、95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%、99.6%、99.7%、99.8%、99.9%序列同一性的氨基酸序列。氨基酸序列也可以是与SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个的全长氨基酸序列具有大于约80%序列同一性、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%、99.6%、99.7%、99.8%、99.9%序列同一性的氨基酸序列。

[0105] 此外，在本发明的范围内，氨基酸序列可以是SEQ ID NO 2、20或22中任何一个所示的氨基酸序列的片段，其中所述片段包含其至少45个连续氨基酸；替代地，氨基酸序列可以是SEQ ID NO 6、8、10、12、14、16、28、30或32中任何一个所示的氨基酸序列的片段，其中所述片段包含其至少10个连续氨基酸并且具有乳糖结合 α -1,3-岩藻糖基转移酶活性。

[0106] 进一步包括在本发明范围内的是如本文所述的 α -1,3-岩藻糖基转移酶多肽，其任选地通过N-末端和/或C-末端氨基酸延伸片段进一步修饰。这种氨基酸延伸片段应理解为在多肽的N-末端和/或C-末端添加多肽序列。例如，多肽序列可与 α -1,3-岩藻糖基转移酶多肽融合以实现另外的酶活性。这种氨基酸延伸片段可以是特定的标签和/或HQ-标签；多至20个氨基酸的延伸，例如1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个氨基酸；这种延伸也可以是25、30、35、40、45、50、55、60、65、70或更多个氨基酸的长度。可选的N-末端和/或C-末端氨基酸延伸片段还可以是用于纯化的标签、用于增加多肽溶解性的标签、用于代谢物形成的标签或氨基酸延伸片段、用于蛋白质代谢组学的标签、用于底物结合的标签、在基因融合中具有相同或不同功能的另一种多肽，例如但不限于编码GDP-岩藻糖合酶、半乳糖基转移酶、岩藻糖基转移酶、双功能岩藻糖激酶/岩藻糖-1-磷酸鸟苷酰转移酶或岩藻糖-1-磷酸鸟苷酰转移酶的多肽，其中所述另一多肽任选地经由肽接头融合至 α -1,3-岩藻糖基转移酶多肽。例如，如本文所述的 α -1,3-岩藻糖基转移酶多肽任选地包括一个或多个外源亲和力标签，例如纯化或底物结合标签，例如6His标签序列、GST标签、HQ标签、HA标签序列、多个6His标签序列、多个GST标签、多个HA标签序列、SNAP标签、SUMOstar标签。其他实例包括蛋白水解裂解位点、保留位点、裂解位点、多组氨酸标签、生物素、抗生物素蛋白、BiTag序列、S标签、肠激酶位点、凝血酶位点、抗体或抗体结构域、抗体片段、抗原、受体(receptor)、受体结构域、受体片段、配体、染料、接受体、猝灭剂或其组合。

[0107] 此外， α -1,3-岩藻糖基转移酶多肽可包括代表功能等效多肽的蛋白质或多肽。这

种等效的 α -1,3-岩藻糖基转移酶多肽可包含由本文所述的 α -1,3-岩藻糖基转移酶多核苷酸编码的氨基酸序列内的氨基酸残基的缺失、添加或替换,但其导致沉默改变,从而产生功能等效的 α -1,3-岩藻糖基转移酶。氨基酸取代可基于所涉及残基的极性、电荷、溶解性、疏水性、亲水性和/或两亲性性质的相似性来进行。例如,非极性(疏水性)氨基酸包括丙氨酸、亮氨酸、异亮氨酸、缬氨酸、脯氨酸、苯丙氨酸、色氨酸和蛋氨酸;平面中性氨基酸包括甘氨酸、丝氨酸、苏氨酸、半胱氨酸、酪氨酸、天冬酰胺和谷氨酰胺;带正电荷(碱性)氨基酸包括精氨酸、赖氨酸和组氨酸;以及带负电荷(酸性)氨基酸包括天冬氨酸和谷氨酸。在本发明的上下文中,如本文所使用的“功能等效物”是指通过包括但不限于酶活性的许多标准中的任一判断的,能够表现出与本发明的乳糖结合 α -1,3-岩藻糖基转移酶多肽基本相似的体内活性的多肽。

[0108] 包括在本发明范围内的是 α -1,3-岩藻糖基转移酶蛋白质、多肽和衍生物(包括片段),其在翻译期间或之后被差异修饰。此外,可将非经典氨基酸或化学氨基酸类似物作为取代或添加而引入 α -1,3-岩藻糖基转移酶多肽序列中。

[0109] α -1,3-岩藻糖基转移酶多肽可通过使用本领域公知技术通过经由重组DNA技术产生的多核苷酸的表达而产生。本领域技术人员熟知的方法可用于构建包含 α -1,3-岩藻糖基转移酶编码序列和适当的转录和/或翻译控制信号的表达载体。这些方法包括,例如,体外重组DNA技术、合成技术和体内基因重组。例如,见Sambrook等人(2001) *Molecular Cloning: a laboratory manual*,第3版,Cold Spring Harbor Laboratory Press, CSH, New York或*Current Protocols in Molecular Biology*, John Wiley and Sons, N.Y. (1989年和每年更新)中描述的技术。备选地, α -1,3-岩藻糖基转移酶多肽可通过直接合成、通过提取在自然界或者在无细胞和/或体外系统内产生该多肽的细胞来产生。

[0110] 新鉴定的具有结合乳糖能力的 α -1,3-岩藻糖基转移酶用于生产3-岩藻糖基乳糖并优选地生产纯度为80%或更高的这种3FL的适宜性是非常令人惊讶的,并且因此,它们的用途代表了一种容易地、高效地和成本有效地生产3-岩藻糖基乳糖的优秀工具。

[0111] 编码 α -1,3-岩藻糖基转移酶多肽的多核苷酸可使用本领域公知的技术经由重组DNA技术产生。本领域技术人员熟知的方法可用于构建包含 α -1,3-岩藻糖基转移酶编码序列和适当的转录和/或翻译控制信号的表达载体。这些方法包括,例如,体外重组DNA技术、合成技术和体内基因重组。例如,见Sambrook等人(2001) *Molecular Cloning: a laboratory manual*,第3版,Cold Spring Harbor Laboratory Press, CSH, New York或*Current Protocols in Molecular Biology*, John Wiley and Sons, N.Y. (1989年和每年更新)中描述的技术。

[0112] 根据本发明的另一方面,提供一种载体,其包含如本文所述的编码具有 α -1,3-岩藻糖基转移酶活性的多肽的多核苷酸,其中多核苷酸可操作地连接至被用载体转化的宿主细胞所识别的控制序列。在特别优选的实施方案中,载体是表达载体,并且根据本发明的另一方面,载体可以以质粒、粘粒、噬菌体、脂质体或病毒的形式存在。

[0113] 因此,根据本发明的多核苷酸可例如包含在将稳定转化/转染到宿主细胞中的载体中。在载体中,本发明的多核苷酸受启动子控制。启动子可以是例如诱导型启动子,使得基因/多核苷酸的表达可以被特异性地靶向,并且,如果需要,基因可以以此方式过度表达。启动子也可以是组成型启动子。

[0114] 多种表达系统可用于生产本发明的多肽。这些载体尤其包括染色体、附加体和病毒衍生的载体,例如,衍生自细菌质粒、噬菌体、转座子、酵母附加体、插入元件、酵母染色体元件、病毒的载体,以及衍生自其组合的载体,例如衍生自质粒和细菌噬菌体遗传元件的那些,如粘粒和噬菌粒。这些载体可包含选择标志物,例如但不限于抗生素标志物、营养缺陷型标志物、毒素-抗毒素标志物、RNA正义/反义标志物。表达系统构建体可能包含调控和产生表达的控制区域。一般而言,适合在宿主中维持、繁殖或表达多核苷酸和/或表达多肽的任何系统或载体可用于在这方面的表达。适当的DNA序列可以通过各种已知的和常规的技术中的任何一种插入到表达系统中,例如,Sambrook等人中阐述的那些,参见上文。

[0115] 对于重组生产,宿主细胞可以经基因工程化以掺入表达系统或其部分或本发明的多核苷酸。将多核苷酸引入宿主细胞可以通过许多标准实验室手册中描述的方法来实现,例如Davis等人,Basic Methods in Molecular Biology, (1986)和Sambrook等人,1989,上文。

[0116] 根据本发明的另一方面,提供了一种宿主细胞,其包含如上所述的载体。

[0117] 根据另一方面,本发明提供了一种用于生产 α -1,3-岩藻糖基乳糖的经基因修饰宿主细胞,其中宿主细胞包含至少一种编码用于3-岩藻糖基乳糖合成的酶的核酸序列并且其中所述细胞包含具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽的表达。所述多肽如本文所述。

[0118] 如本文所使用的,术语“宿主细胞”现在定义为已被转化或转染或者能够通过外源性多核苷酸序列转化或转染的细胞,因此包含至少一种不天然存在于所述宿主细胞中的序列。

[0119] 可以利用多种宿主表达载体系统来表达本发明的 α -1,3-岩藻糖基转移酶多核苷酸。这种宿主表达系统表示可产生和随后纯化目标编码序列的媒介物,但也表示在用适当的核苷酸编码序列进行转化或转染时,在原位显示本发明的 α -1,3-岩藻糖基转移酶基因产物的细胞。

[0120] 根据本发明的另一方面,提供了一种用于生产3-岩藻糖乳糖的宿主细胞,其中宿主细胞包含由多核苷酸组成的序列,所述多核苷酸编码具有乳糖结合 α -1,3-岩藻糖基转移酶活性的多肽,如本文所述,其中序列是对于宿主细胞外源的序列并且其中序列整合在宿主细胞的基因组中。多核苷酸可操作地与宿主细胞所识别的控制序列相连接。

[0121] 根据本发明的备选方面,提供了一种用于生产3-岩藻糖基乳糖的宿主细胞,其中宿主细胞包含载体,所述载体包含本文描述的多核苷酸,其中多核苷酸可操作地连接到由用所述载体转化的宿主细胞识别的控制序列。

[0122] 在另一方面,本发明还提供了一种 α -1,3-岩藻糖基乳糖的生产方法,包括以下步骤:a)提供本文所述的细胞,和b)在允许产生 α -1,3-岩藻糖基转移酶的条件下在培养基中培养细胞。

[0123] 优选地,所述 α -1,3-岩藻糖基转移酶从本文所述的培养中分离。优选地,也可以按照本文所述进行纯化。

[0124] 在另一个方面,本发明提供了本发明所述的细胞用于生产3-岩藻糖基乳糖的用途。

[0125] 根据本发明的另一方面,提供了一种微生物,其表达本文所述的并且优选地由本

文所述的多核苷酸编码的 α -1,3-岩藻糖基转移酶。

[0126] 本文所使用的术语微生物或生物体或细胞或宿主细胞是指从包括细菌、酵母或真菌的列表中选择微生物,或指植物或动物细胞。后者细菌优选地属于变形菌(Proteobacteria)门或厚壁菌(Firmicutes)门或蓝细菌(Cyanobacteria)门或者异常球菌-栖热菌(Deinococcus-Thermus)门。属于变形菌门的后者细菌优选地属于肠杆菌科(Enterobacteriaceae),优选地属于大肠杆菌种。后者细菌优选地涉及属于大肠杆菌种的任何菌株,例如但不限于大肠杆菌B、大肠杆菌C、大肠杆菌W、大肠杆菌K12、大肠杆菌Nissle。更具体地,后者术语涉及培养的大肠杆菌菌株——称为大肠杆菌K12菌株——其非常适合于实验室环境,并且与野生型菌株不同,其已丧失了肠中繁殖的能力。大肠杆菌K12菌株的公知实例是K12野生型、W3110、MG1655、M182、MC1000、MC1060、MC1061、MC4100、JM101、NZN111和AA200。因此,本发明特别涉及如上所述突变和/或转化的大肠杆菌宿主细胞或菌株,其中所述大肠杆菌菌株是K12菌株。更优选地,大肠杆菌K12菌株是大肠杆菌MG1655。属于厚壁菌门的后者细菌优选地属于芽孢杆菌纲(Bacilli),优选地属于乳杆菌目(Lactobacilliales),其成员诸如乳酸乳杆菌(*Lactobacillus lactis*)、肠系膜明串珠菌(*Leuconostoc mesenteroides*),或芽孢杆菌目(Bacillales),其成员诸如来自芽孢杆菌属,诸如枯草芽孢杆菌(*Bacillus subtilis*)或解淀粉芽孢杆菌(*B.amyloliquefaciens*)。后者细菌属于放线菌门,优选地属于棒杆菌科(Corynebacteriaceae),其成员为谷氨酸棒杆菌(*Corynebacterium glutamicum*)或非发酵棒杆菌(*C.afermentans*),或属于链霉菌科(Streptomycetaceae),其成员为灰链霉菌(*Streptomyces griseus*)或弗雷迪链霉菌(*S.fradiae*)。后者酵母优选地属于子囊菌门(Ascomycota)或担子菌门(Basidiomycota)或半知菌门(Deuteromycota)或接合菌门(Zygomycete)。后者酵母优选地属于酵母属(*Saccharomyces*)、毕赤酵母属(*Pichia*)、Komagataella、汉逊酵母属(*Hansunella*)、克鲁维酵母属(*Kluyveromyces*)、耶氏酵母属(*Yarrowia*)或斯塔摩酵母属(*Starmerella*)。后者真菌优选地属于根霉属(*Rhizopus*)、盘基网柄菌属(*Dictyostelium*)、青霉属(*Penicillium*)、毛霉属(*Mucor*)或曲霉属(*Aspergillus*)。

[0127] 根据本发明的另一方面,编码具有乳糖结合 α -1,3-岩藻糖基转移酶活性的多肽的多核苷酸适宜于相应宿主细胞的密码子使用。

[0128] 本发明的另一方面提供了本文所述的多肽用于生产 α -1,3-岩藻糖乳糖的用途。本发明的另一方面提供了如本文所述的多核苷酸或本文所述的载体用于生产 α -1,3-岩藻糖乳糖的用途。

[0129] 根据另一实施方案,提供迄今未知的乳糖结合 α -1,3-岩藻糖基转移酶。本发明提供一种具有乳糖结合 α -1,3-岩藻糖基转移酶活性的分离和/或合成多肽,其中所述多肽包括:

[0130] -编码保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)和保守的[K/D][L/K/M]XXX[F/Y]结构域(SEQ ID NO 34)的氨基酸序列,其中如果该结构域等于DM[A/S]VSF(SEQ ID NO 36),则另外地在N-末端区域存在保守的基序[N/H]XDPAXLD(SEQ ID NO 35),其中X可以是任何不同的氨基酸,并且所述氨基酸序列的C-末端具有从GDP-岩藻糖结合结构域的第一个氨基酸开始的小于或等于100个氨基酸。

[0131] 优选地,所述多肽选自以下项目组成的组:

[0132] i) 所附序列列表中的SEQ ID NO 2、4、6、8、10、12、14、16、20、22、28、30或32;

[0133] ii) 与SEQ ID NO 2、20或22的全长具有87%或更高序列同一性的氨基酸序列;

[0134] iii) 包含与SEQ ID NO 6、8、10、12、14、16、28、30或32中任何一个的全长氨基酸序列具有至少80%序列同一性的氨基酸序列;

[0135] iv) SEQ ID NO 2、20或22中任何一个所示的序列的片段,其中所述片段包括其至少45个连续氨基酸;

[0136] v) SEQ ID NO 6、8、10、12、14、16、28、30或32中任何一个所示的氨基酸序列的片段,其中所述片段包括其至少10个连续氨基酸。

[0137] 任选地,所述多肽通过N-末端和/或C-末端氨基酸延伸片段进一步修饰。

[0138] 在本发明的范围内,分离和/或合成的多肽具有乳糖结合 α -1,3-岩藻糖基转移酶活性。这样的多肽包含编码保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)和保守[K/D][L/K/M]XXX[F/Y]结构域(SEQ ID NO 34)的氨基酸序列,其中如果该结构域等于DM[A/S]VSF(SEQ ID NO 36),则另外地在N-末端区域存在保守的基序[N/H]XDPAXLD(SEQ ID NO 35),其中X可以是任何不同的氨基酸,且所述氨基酸序列的C-末端具有从以上定义的保守的GDP-岩藻糖结合结构域的第一个氨基酸开始的小于或等于100个氨基酸,例如100、99、98、97、96、95、94、93、92、91、90、89、88、87、86、85、84、83、82、81、80、79、78、77、76、75、74、73、72、71、70、69、68、67、66、65、64、63、62、61、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40个氨基酸。

[0139] 还包括在本发明的范围内是如本文所述的 α 1,3-岩藻糖基转移酶多肽,其任选地通过N-末端和/或C-末端氨基酸延伸片段进一步修饰。

[0140] 令人惊讶地发现,新鉴定的乳糖结合 α -1,3-岩藻糖基转移酶可用于进行非天然发生的反应。此外,已经发现上述鉴定的 α -1,3-岩藻糖基转移酶能够使用乳糖作为底物,具有与目前已知的 α -1,3-岩藻糖基转移酶相似或更高的乳糖结合特性,并且能够产生3-岩藻糖基乳糖。

[0141] 直到今天,如表1所示,还没有描述新鉴定的本发明岩藻糖基转移酶具有乳糖结合 α -1,3-岩藻糖基转移酶活性。

[0142] 表1:

[0143]

SEQ ID NO	DNA 标识	NCBI_名称	生物体
SEQ ID 1 - 2	SMF69967.1	糖基转移酶家族 10 (岩藻糖基转移酶) C-term [稻固氮螺菌 (<i>Azospirillum oryzae</i>)]	稻固氮螺菌 (<i>Azospirillum oryzae</i>) A2P
SEQ ID 3 - 4	WP_042442472.1	推定的糖基转移酶[生脂固氮螺菌 4B]	生脂固氮螺菌 (<i>Azospirillum lipoferum</i>)
SEQ ID 5 - 6	AIL32582.1	假定蛋白 IX83_03990 [<i>Basilea psittaculmonis</i> DSM 24701]	<i>Basilea psittaculmonis</i>
SEQ ID 7 - 8	ADG66884.1	推测的 LPS 生物合成相关糖基转移酶[<i>Planctopirus limnophila</i> DSM 3776]	<i>Planctopirus limnophila</i> (菌株 ATCC 43296 / DSM 3776 / IFAM 1008 / 290) (<i>Planctomyces limnophilus</i>)
SEQ ID 9 - 10	KHJ37904.1	糖基转移酶家族 10 (岩藻糖基转移酶)[<i>Pedobacter glucosidilyticus</i>]	<i>Pedobacter glucosidilyticus</i>
SEQ ID 11 - 12	WP_081748371.1	假定蛋白[卡托氏吡啉单胞菌 (<i>Porphyromonas catoniae</i>)]	卡托氏吡啉单胞菌 (<i>Porphyromonas catoniae</i>) (WGS, 在 genbank 中: JDF01000001 至 JDF010000025)
SEQ ID 13 - 14	WP_052080772.1	假定蛋白[吡啉单胞菌属 (<i>Porphyromonas</i>) 物种 COT-239 OH1446]	吡啉单胞菌属 (<i>Porphyromonas</i>) 物种 COT-239 OH1446 (contig_18; NZ_JRA01000018.1)
SEQ ID 15 - 16	EHG19535.1	假定蛋白 HMPREF9334_01850 [损伤新月形单胞菌 ATCC 43532]	损伤新月形单胞菌 (<i>Selenomonas infelix</i>) ATCC 43532
SEQ ID 19 - 20	WP_109445332	α -1,3-岩藻糖基转移酶[固氮螺菌属 (<i>Azospirillum</i>) 物种[TSH64]	固氮螺菌属 (<i>Azospirillum</i>) 物种 TSH64
SEQ ID 21 - 22	QCG87584.1	α -1,3-岩藻糖基转移酶[固氮螺菌属 (<i>Azospirillum</i>) 物种 TSH100]	固氮螺菌属 (<i>Azospirillum</i>) 物种 TSH100
SEQ ID 27 - 28	SEQ36152.1	糖基转移酶家族 10 (岩藻糖基转移酶) C-term [丁酸弧菌属物种 TB]	丁酸弧菌属 (<i>Butyrivibrio</i>) 物种 TB
SEQ ID 29 - 30	EKX99948.1	假定蛋白 HMPREF9134_01857 [卡托氏吡啉单胞菌 F0037]	卡托氏吡啉单胞菌 (<i>Porphyromonas catoniae</i>) F0037
SEQ ID 31 - 32	SHI32494.1	糖基转移酶家族 10 (岩藻糖基转移酶) C-term [溶纤维丁酸弧菌]	溶纤维丁酸弧菌 (<i>Butyrivibrio fibrisolvens</i>)

[0144] 如表2所示,还发现新鉴定的具有结合乳糖的能力、被用于生产3-岩藻糖基乳糖的 α -1,3-岩藻糖基转移酶,所有这些都具有相同的特殊特征,即具有包含保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)的氨基酸序列,其中x可以是任何不同的氨基酸并且其中所述氨基酸序列的C-末端具有从GDP-岩藻糖结合结构域的第一个氨基酸开始的小于或等于100个氨基酸。这与已知的乳糖结合3-岩藻糖基转移酶相反,已知的乳糖结合3-岩藻糖基转移酶例如在W02012/049083所述具有从上述定义的GDP-岩藻糖结合结构域的第一个氨基酸开始的长于100个氨基酸的C-末端。

[0145] 表2:

[0146]

SEQ ID NO	DNA 标识	生物体	从 GDP-岩藻糖结合结构域的 C-末端长度
SEQ ID 1 - 2	SMF69967.1	稻固氮螺菌 A2P	89
SEQ ID 3 - 4	WP_042442472.1	生脂固氮螺菌 B510	89
SEQ ID 5 - 6	AIL32582.1	<i>Basilea psittacipulmonis</i>	97
SEQ ID 7 - 8	ADG66884.1	<i>Planctopirus limnophila</i>	96
SEQ ID 9 - 10	KHJ37904.1	<i>Pedobacter glucosidilyticus</i>	94
SEQ ID 11 - 12	WP_081748371.1	卡托氏吡啶单胞菌	93
SEQ ID 13 - 14	WP_052080772.1	吡啶单胞菌属物种 COT-239 OH1446	94
SEQ ID 15 - 16	EHG19535.1	损伤新月形单胞菌	96
SEQ ID 19 - 20	WP_109445332	固氮螺菌属物种 TSH64	89
SEQ ID 21 - 22	QCG87584.1	固氮螺菌属物种 TSH100	89
SEQ ID 27 - 28	SEQ36152.1	丁酸弧菌属物种 TB	93
SEQ ID 29 - 30	EKX99948.1	卡托氏吡啶单胞菌 F0037	93
SEQ ID 31 - 32	SHI32494.1	溶纤维丁酸弧菌 DSM 3071	93

[0147] 此外,还发现SEQ ID NO 6、SEQ ID NO 10、SEQ ID NO 12、SEQ ID NO 14和SEQ ID NO 16的多肽序列共享结构域PENXXXXXXXXTEK (SEQ ID NO 37),其中X可以是任何不同的氨基酸,如图1所示,其中将该结构域放在方框中。所有比对均采用MAFFT v7.307,可视化采用Jalview 2.10。

[0148] 此外,还发现,除了保守的GDP-岩藻糖结合域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R] (SEQ ID NO 33) 之外,新鉴定的多肽都共享共同的保守的基序[K/D][L/K/M]XXX[F/Y] (SEQ ID NO 34),其中X可以是任何不同的氨基酸,以及对乳糖结合重要的保守的氨基酸区[FW]W,如图11所示,其中将该结构域放在方框中。

[0149] 此外,我们注意到,当这个共同特征是DM[A/S]VSF (SEQ ID NO 36)时,另外地在蛋白质的N-末端结构域中需要保守的[N/H]XDPAXLD (SEQ ID NO 35) 基序,其中X可以是任何不同的氨基酸,以使酶对作为受体底物的乳糖具有 α -1,3-岩藻糖基转移酶活性,如图12的比对中所示,其中将该结构域放在方框中。如实施例14中所例示,具有SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 20和SEQ ID NO 22的多肽包含两种共有基序并且似乎对作为受体底物的乳糖具有 α -1,3-岩藻糖基转移酶活性,而具有SEQ ID NO 24和SEQ ID NO 26的多肽不包含N-末端[NH]XDPAXLD基序(其中X可以是任何不同的氨基酸 (SEQ ID NO 35))并且确实显示出这种活性。

[0150] 此外,还发现SEQ ID NO 6、SEQ ID NO 12、SEQ ID NO 14、SEQ ID NO 16、SEQ ID NO 28、SEQ ID NO 30和SEQ ID NO 32的多肽序列共享结构域K[IV]F[FL]XGEN (SEQ ID NO 41) 和RFPLW (SEQ ID NO 42),其中x可以是任何不同的氨基酸,如图13的比对所示,其中将该结构域放在方框中。

[0151] 在第二实施方案中,本发明还涉及一种分离和/或合成的多核苷酸,其编码如上所

述具有乳糖结合 α -1,3-岩藻糖基转移酶活性的多肽。

[0152] 在本发明的范围内,多核苷酸可以是编码SEQ ID NO 2、6、8、10、12、14、16、20、22、28、30、32所示氨基酸序列中任何一个的多核苷酸的等位变体。

[0153] 因此,本发明还涉及一种分离和/或合成的多核苷酸,其编码具有 α -1,3-岩藻糖基转移酶活性的多肽并且其包含选自以下项目组成的组的序列:a)所附序列表中的SEQ ID NO 1、5、7、9、11、13、15、19、21、27、29、31;b)与SEQ ID NO 1、5、7、9、11、13、15、19、21、27、29、31互补的核酸序列;c)与SEQ ID NO 1、5、7、9、11、13、15、19、21、27、29、31具有80%或更高序列同一性的核酸序列。

[0154] 因此,本发明还涉及通过根据本发明的方法获得的3-岩藻糖基乳糖,以及如上所述的多核苷酸、载体、宿主细胞、微生物或多肽用于生产3-岩藻糖基乳糖的用途。 α -1,3-岩藻糖基乳糖可用作食品添加剂,益生元,共生体(symbiotic),用于婴儿食品、成人食品或饲料的补充,或用作治疗或医药活性化合物。通过这些新方法,可以容易而有效地提供 α -1,3-岩藻糖基乳糖,而不需要复杂、耗时和耗费成本的合成过程。

[0155] 除非另有定义,否则本文中使用的所有技术和科学术语通常具有与本发明所属领域的普通技术人员通常理解的相同含义。一般而言,本文中使用的术语以及上下文描述的细胞培养、分子遗传学、有机化学和核酸化学中的实验室程序是本领域中公知和常用的。核酸和肽的合成使用标准技术。通常,酶促反应和纯化步骤根据制造商的说明书进行。

[0156] 其他优点来自具体实施方案、实施例和附图。

[0157] 不言而喻,在不脱离本发明的范围的情况下,上述特征和下面仍要说明的特征不仅可以在各自指定的组合中使用,还可以在其他组合中使用或单独使用。

[0158] 本发明涉及以下具体实施方案:

[0159] 1.一种生产 α -1,3-岩藻糖基乳糖的方法,该方法包括以下步骤:

[0160] a)提供具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽,其中所述多肽包含

[0161] i)编码保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)的氨基酸序列;

[0162] ii)编码保守的[K/D][L/K/M]XXX[F/Y]结构域(SEQ ID NO 34)的氨基酸序列,和

[0163] iii)其中如果ii)的结构域等于DM[A/S]VSF(SEQ ID NO 36),则另外地在N-末端区域存在保守的基序[N/H]XDPAXLD(SEQ ID NO 35);

[0164] 其中X可以是任何不同的氨基酸;和

[0165] 其中所述多肽的C-末端具有从GDP-岩藻糖结合结构域的第一个氨基酸开始的小于或等于100个氨基酸;

[0166] b)在其中所述多肽催化岩藻糖残基从供体底物转移至受体底物的条件下,使步骤a)的具有 α -1,3-岩藻糖基转移酶活性的所述多肽与包含GDP-岩藻糖作为供体底物和乳糖作为受体底物的混合物接触,

[0167] 从而产生 α -1,3-岩藻糖基乳糖,

[0168] c)任选地分离所述 α -1,3-岩藻糖基乳糖。

[0169] 2.根据实施方案1的方法,其中所述多肽在无细胞系统中提供。

[0170] 3.根据实施方案1的方法,其中所述多肽由包含编码所述多肽的多核苷酸的细胞

产生。

[0171] 4. 根据实施方案1或3中任一项的方法,其中所述GDP-岩藻糖和/或乳糖由产生所述GDP-岩藻糖和/或乳糖的细胞提供。

[0172] 5. 根据实施方案1、3或4中的任一项的方法,该方法包括以下步骤:

[0173] i) 提供经基因修饰以产生 α -1,3-岩藻糖基乳糖的细胞,所述细胞包含至少一种编码用于 α -1,3-岩藻糖基乳糖合成的酶的核酸序列,

[0174] 所述细胞包含具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的所述多肽的表达,

[0175] ii) 在允许产生 α -1,3-岩藻糖基乳糖的条件下在培养基中培养细胞,

[0176] iii) 优选地从培养中分离 α -1,3-岩藻糖基乳糖。

[0177] 6. 根据实施方案3的方法,该方法包括以下步骤:

[0178] a) 提供表达具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的所述多肽的宿主细胞;

[0179] b) 在允许产生 α -1,3-岩藻糖基乳糖并允许表达具有 α -1,3-岩藻糖基转移酶活性的所述多肽的合适的营养条件下使所述宿主细胞生长;

[0180] c) 在步骤b)的同时或之后提供供体底物GDP-岩藻糖和受体底物乳糖,以便 α -1,3-岩藻糖基转移酶多肽催化岩藻糖残基从GDP-岩藻糖向乳糖的转移,从而产生 α -1,3-岩藻糖基乳糖;

[0181] d) 任选地从宿主细胞或其生长的培养基中分离所述 α -1,3-岩藻糖基乳糖。

[0182] 7. 根据实施方案5或6中任一项的方法,其中宿主细胞经转化或转染以表达具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的外源多肽。

[0183] 8. 根据实施方案3至7中任一项的方法,其特征在于通过在宿主细胞中同时表达的酶或通过宿主细胞的代谢来提供GDP-岩藻糖和/或乳糖。

[0184] 9. 根据实施方案1至8中任一项的方法,其进一步包括 α -1,3-岩藻糖基乳糖的纯化。

[0185] 10. 根据前述实施方案中任一项的方法,其中所述多肽选自由以下项目组成的组:

[0186] i) 所附序列表中SEQ ID NO 6、2、4、8、10、12、14、16、20、22、28、30或32的任何一个;

[0187] ii) 与SEQ ID NO 2、20或22的全长氨基酸序列具有87%或更高序列同一性的氨基酸序列;

[0188] iii) 与SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个的全长氨基酸序列具有80%或更高序列同一性的氨基酸序列;

[0189] iv) SEQ ID NO 2、20或22所示的氨基酸序列的片段,其中所述片段包含其至少45个连续氨基酸;

[0190] v) SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个所示的氨基酸序列的片段,其中所述片段包含其至少10个连续氨基酸并具有乳糖结合 α -1,3-岩藻糖基转移酶活性;

[0191] 任选地,所述多肽通过N-末端和/或C-末端氨基酸延伸片段进一步修饰。

[0192] 11. 根据前述实施方案中任一项的生产3-岩藻糖基乳糖的方法,该方法进一步包

括以下步骤中的至少一个：

[0193] i) 向培养基中添加乳糖进料,所述乳糖进料每初始反应器体积包含至少50克、更优选地至少75克、更优选地至少100克、更优选地至少120克、更优选地至少150克乳糖,优选地以连续方式添加,并且优选地使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍,优选地不大于2倍,更优选地小于2倍;

[0194] ii) 在1天、2天、3天、4天、5天的过程中,通过进料溶液以连续方式向培养基中添加乳糖进料;

[0195] iii) 在1天、2天、3天、4天、5天的过程中,通过进料溶液以连续方式向培养基中添加乳糖进料,并且其中所述乳糖进料溶液的浓度是50g/L,优选地75g/L,更优选地100g/L,更优选地125g/L,更优选地150g/L,更优选地175g/L,更优选地200g/L,更优选地225g/L,更优选地250g/L,更优选地275g/L,更优选地300g/L,更优选地325g/L,更优选地350g/L,更优选地375g/L,更优选地400g/L,更优选地为450g/L,更优选地500g/L,甚至更优选地550g/L,最优选地600g/L;并且其中优选地将所述溶液的pH设置在3至7并且其中优选地将所述进料溶液的温度保持在20℃至80℃;

[0196] iv) 所述方法导致在所述培养基的最终体积中3-岩藻糖基乳糖浓度为至少50g/L,优选地至少75g/L,更优选地至少90g/L,更优选地至少100g/L,更优选地至少125g/L,更优选地至少150g/L,更优选地至少150g/L,更优选地至少175g/L,更优选地至少200g/L。

[0197] 12. 经基因修饰以产生 α -1,3-岩藻糖基乳糖的宿主细胞,其中宿主细胞包含至少一种编码参与 α -1,3-岩藻糖基糖合成的酶的核酸序列;所述细胞包含表达具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽,其中所述多肽包含:

[0198] i) 编码保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)的氨基酸序列;

[0199] ii) 编码保守的[K/D][L/K/M]XXX[F/Y]结构域(SEQ ID NO 34)的氨基酸序列,和

[0200] iii) 其中如果ii)的结构域等于DM[A/S]VSF(SEQ ID NO 36),则另外地在N-末端区域存在保守的基序[N/H]XDPAXLD(SEQ ID NO 35);

[0201] 其中X可以是任何不同的氨基酸;和

[0202] 其中所述多肽的C-末端具有从GDP-岩藻糖结合结构域的第一个氨基酸开始的小于或等于100个氨基酸。

[0203] 13. 根据实施方案12的细胞,所述宿主细胞包含:i) 包含编码具有乳糖结合 α -1,3-岩藻糖基转移酶活性的所述多肽的多核苷酸的序列,其中所述序列对于所述宿主细胞是外源的序列,并且其中所述序列整合在所述宿主细胞的基因组中,或ii) 含有包含编码所述多肽的多核苷酸的载体,其中所述多核苷酸可操作地连接至由用载体转染的宿主细胞识别的控制序列。

[0204] 14. 根据实施方案12或13中任一项的细胞,其中所述多肽包含选自以下项目组成的组的氨基酸序列:

[0205] i) 所附序列表中SEQ ID NO 6、2、4、8、10、12、14、16、20、22、28、30或32的任何一个;

[0206] ii) 与SEQ ID NO 2、20或22的全长氨基酸序列具有87%或更高序列同一性的氨基酸序列;

[0207] iii) 与SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个人的全长氨基酸序列具有80%或更高序列同一性的氨基酸序列;

[0208] iv) SEQ ID NO 2、20或22所示的氨基酸序列的片段,其中所述片段包含其至少45个连续氨基酸;

[0209] v) SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个人所示的氨基酸序列的片段,其中所述片段包含其至少10个连续氨基酸并具有乳糖结合 α -1,3-岩藻糖基转移酶活性;

[0210] 任选地,所述多肽通过N-末端和/或C-末端氨基酸延伸片段进一步修饰。

[0211] 15. 根据实施方案3至11中任一项的方法或根据实施方案12、13或14中任一项的细胞,其中所述细胞选自由微生物、植物或动物细胞组成的组,优选地所述微生物是细菌、真菌或酵母,优选地所述植物是水稻、棉花、油菜籽、大豆、玉米或谷类植物,优选地所述动物是昆虫、鱼类、鸟类或非人类哺乳动物;优选地细胞是大肠杆菌细胞。

[0212] 16. 根据实施方案12至15中任一项的宿主细胞,其特征在于宿主细胞是细菌、优选地是大肠杆菌菌株、更优选地是K12菌株的大肠杆菌菌株的细胞,甚至更优选地大肠杆菌K12菌株是大肠杆菌MG1655。

[0213] 17. 根据实施方案12至15中任一项的宿主细胞,其特征在于宿主细胞是酵母细胞。

[0214] 18. 根据实施方案12至17中任一项的宿主细胞,其特征在于编码具有乳糖结合 α -1,2-岩藻糖基转移酶活性的多肽的多核苷酸适合于相应宿主细胞的密码子使用。

[0215] 19. 一种生产 α -1,3-岩藻糖基乳糖的方法,包括以下步骤:

[0216] a) 提供根据实施方案12至18中任一项的细胞,

[0217] b) 在允许产生 α -1,3-岩藻糖基转移酶的条件下在培养基中培养细胞,

[0218] c) 优选地,从培养中分离所述 α -1,3-岩藻糖基转移酶。

[0219] 20. 根据实施方案12至18中任一项的宿主细胞用于生产 α -1,3-岩藻糖基乳糖的用途。

[0220] 21. 如实施方案1或11中任一项的方法中所述的多肽用于生产 α -1,3-岩藻糖基乳糖的用途。

[0221] 22. 异源表达乳糖结合 α -1,3-岩藻糖基转移酶多肽的微生物,其中所述多肽包含:

[0222] i) 编码保守的GDP-岩藻糖结合结构域[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R](SEQ ID NO 33)的氨基酸序列;

[0223] ii) 编码保守的[K/D][L/K/M]XXX[F/Y]结构域(SEQ ID NO 34)的氨基酸序列,和

[0224] iii) 其中如果ii)的结构域等于DM[A/S]VSF(SEQ ID NO 36),则另外地在N-末端区域存在保守的基序[N/H]XDPAXLD(SEQ ID NO 35);

[0225] 其中X可以是任何不同的氨基酸;和

[0226] 其中所述多肽的C-末端具有从GDP-岩藻糖结合结构域的第一个氨基酸开始的小于或等于100个氨基酸。

[0227] 23. 根据实施方案22的微生物,其中所述多肽包含选自由以下项目组成的组的氨基酸序列:

[0228] i) 所附序列表中SEQ ID NO 6、2、4、8、10、12、14、16、20、22、28、30或32的任何一个;

[0229] ii) 与SEQ ID NO 2、20或22的全长氨基酸序列具有87%或更高序列同一性的氨基酸序列;

[0230] iii) 与SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个的全长氨基酸序列具有80%或更高序列同一性的氨基酸序列;

[0231] iv) SEQ ID NO 2、20或22所示的氨基酸序列的片段,其中所述片段包含其至少45个连续氨基酸;

[0232] v) SEQ ID NO 6、8、10、12、14、16、28、30或32的任何一个所示的氨基酸序列的片段,其中所述片段包含其至少10个连续氨基酸并具有乳糖结合 α -1,3-岩藻糖基转移酶活性;

[0233] 任选地,所述多肽通过N-末端和/或C-末端氨基酸延伸片段进一步修饰。

[0234] 24. 根据实施方案22或23的微生物用于生产 α -1,3-岩藻糖基乳糖的用途。

[0235] 25. 根据实施方案1至11、15或19中任一项的方法,还包括从宿主细胞或其生长的培养基中分离所述 α -1,3-岩藻糖基乳糖的步骤。

[0236] 26. 根据实施方案1至11、15、19或25中任一项的方法,其中所述分离包括以下步骤中的至少一个:澄清、超滤、纳米过滤、反渗透、微滤、活性木炭或碳处理、切向流高效过滤、切向流超滤、亲和色谱、离子交换色谱、疏水相互作用色谱和/或凝胶过滤、配体交换色谱。

[0237] 27. 实施方案1至11、15、19、25或26中任一项的方法,进一步包括 α -1,3-岩藻糖基乳糖的纯化。

[0238] 28. 实施方案27的方法,其中所述 α -1,3-岩藻糖基乳糖的所述纯化包括以下步骤中的至少一个:使用活性木炭或碳,使用木炭,纳米过滤,超滤或离子交换,使用醇,使用含水醇混合物,结晶,蒸发,沉淀,干燥,喷雾干燥或冻干。

[0239] 29. 实施方案1至11、15、19、25至28中任一项的方法,其中多肽在真菌、酵母、细菌、昆虫、动物和植物表达系统中产生。

[0240] 30. 实施方案29的方法,其中宿主细胞是细菌、优选地是大肠杆菌菌株、更优选地是K12菌株的大肠杆菌菌株的细胞,甚至更优选地大肠杆菌K12菌株是大肠杆菌MG1655。

[0241] 31. 实施方案29的方法,其中宿主细胞是酵母细胞。

[0242] 32. 实施方案1至11、15、19、25至31中任一项的方法,其中培养基中的乳糖浓度在50至150g/L的范围内。

[0243] 33. 实施方案1至11、15、19、25至32中任一项的方法,其中3-岩藻糖基乳糖的最终浓度在70g/L至200g/L的范围内。

[0244] 34. 实施方案1至11、15、19、25至33中任一项的方法,其中所述生产导致在发酵结束时乳糖浓度与3-岩藻糖基乳糖的浓度比率小于1:5。

[0245] 35. 实施方案1至11、15、19、25至34中任一项的方法,其中所述产生导致在发酵结束时3-岩藻糖基乳糖纯度为80%或更高。

[0246] 36. 一种生产 α -1,3-岩藻糖基乳糖的方法,该方法包括以下步骤:

[0247] a) 提供具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽,

[0248] b) 在其中所述多肽催化岩藻糖残基从供体底物转移至受体底物的条件下,使步骤a)的具有 α -1,3-岩藻糖基转移酶活性的所述多肽与包含GDP-岩藻糖作为供体底物和乳糖

作为受体底物的混合物接触,

[0249] 从而产生 α -1,3-岩藻糖基乳糖,

[0250] c) 其中所述催化导致在发酵结束时乳糖浓度与3-岩藻糖基乳糖的浓度比率小于1:5,

[0251] d) 任选地分离所述 α -1,3-岩藻糖基乳糖。

[0252] 37. 一种生产 α -1,3-岩藻糖基乳糖的方法, 该方法包括以下步骤:

[0253] a) 提供具有 α -1,3-岩藻糖基转移酶活性和具有使用乳糖作为受体底物的能力的多肽,

[0254] b) 在其中所述多肽催化岩藻糖残基从供体底物转移至受体底物的条件下, 使步骤a) 的具有 α -1,3-岩藻糖基转移酶活性的所述多肽与包含GDP-岩藻糖作为供体底物和乳糖作为受体底物的混合物接触,

[0255] 从而产生 α -1,3-岩藻糖基乳糖,

[0256] c) 其中所述催化导致在发酵结束时3-岩藻糖基乳糖纯度为80%或更高,

[0257] d) 任选地分离所述 α -1,3-岩藻糖基乳糖。

[0258] 38. 生产3-岩藻糖基乳糖的方法, 包括以下步骤中的至少一个:

[0259] i) 向培养基中添加乳糖进料, 所述乳糖进料每初始反应器体积包含至少50克、更优选地至少75克、更优选地至少100克、更优选地至少120克, 更优选地至少150克乳糖, 优选地以连续的方式添加, 并且优选地使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍, 优选地不大于2倍, 更优选地小于2倍;

[0260] ii) 在1天、2天、3天、4天、5天的过程中, 通过进料溶液以连续的方式向培养基中添加乳糖进料;

[0261] iii) 在1天、2天、3天、4天、5天的过程中, 通过进料溶液以连续的方式向培养基中添加乳糖进料, 并且其中所述乳糖进料溶液的浓度是50g/L, 优选地75g/L, 更优选地100g/L, 更优选地125g/L, 更优选地150g/L, 更优选地175g/L, 更优选地200g/L, 更优选地225g/L, 更优选地250g/L, 更优选地275g/L, 更优选地300g/L, 更优选地325g/L, 更优选地350g/L, 更优选地375g/L, 更优选地400g/L, 更优选地450g/L, 更优选地500g/L, 甚至更优选地550g/L, 最优选地600g/L; 并且其中优选地将所述溶液的pH设置在3至7并且其中优选地将所述进料溶液的温度保持在20℃至80℃;

[0262] 所述方法导致在所述培养基的最终体积中3-岩藻糖基乳糖浓度是至少50g/L, 优选地至少75g/L, 更优选地至少90g/L, 更优选地至少100g/L, 更优选地至少125g/L, 更优选地至少150g/L, 更优选地至少175g/L, 更优选地至少200g/L, 并且优选地在所述培养物的最终体积中乳糖浓度与3FL浓度比率低于1:5, 更优选地1:10, 甚至更优选地1:20, 最优选地1:40。

[0263] 39. 生产3-岩藻糖基乳糖的方法, 其包括以下步骤中的至少一个:

[0264] i) 向培养基中添加乳糖进料, 所述乳糖进料每初始反应器体积包含至少50克、更优选地至少75克、更优选地至少100克、更优选地至少120克, 更优选地至少150克乳糖, 优选地以连续的方式添加, 并且优选地使得培养基的最终体积是添加所述乳糖进料之前的培养基体积的不大于3倍, 优选地不大于2倍, 更优选地小于2倍;

[0265] ii) 在1天、2天、3天、4天、5天的过程中, 通过进料溶液以连续的方式向培养基中添

加乳糖进料;

[0266] iii) 在1天、2天、3天、4天、5天的过程中,通过进料溶液以连续的方式向培养基中添加乳糖进料,并且其中所述乳糖进料溶液的浓度是50g/L,优选地75g/L,更优选地100g/L,更优选地125g/L,更优选地150g/L,更优选地175g/L,更优选地200g/L,更优选地225g/L,更优选地250g/L,更优选地275g/L,更优选地300g/L,更优选地325g/L,更优选地350g/L,更优选地375g/L,更优选地400g/L,更优选地450g/L,更优选地500g/L,甚至更优选地550g/L,最优选地600g/L;并且其中优选地将所述溶液的pH设置在3至7并且其中优选地将所述进料溶液的温度保持在20℃至80℃;

[0267] 所述方法导致在所述培养基的最终体积中3-岩藻糖基乳糖浓度是至少50g/L,优选地至少75g/L,更优选地至少90g/L,更优选地至少100g/L,更优选地至少125g/L,更优选地至少150g/L,更优选地至少175g/L,更优选地至少200g/L,并且优选地在所述培养物的最终体积中3FL纯度是80%或更高。

[0268] 以下附图和实施例将用作本发明的进一步说明和澄清,并且不旨在限制本发明。

附图说明

[0269] 图1显示了SEQ ID NO 6、SEQ ID 10、SEQ ID 12、SEQ ID 14和SEQ ID NO 16的多肽序列的比对。

[0270] 图2显示了生长实验中3-岩藻糖基乳糖的标准化的产量。

[0271] 图3显示以在培养基中从低到高乳糖量的生长实验中3-岩藻糖基乳糖的标准化的产量。

[0272] 图4显示在培养基中低乳糖量的生长实验中3-岩藻糖基乳糖的标准化的产量。

[0273] 图5显示所鉴定的乳糖结合 α -1,3-岩藻糖基转移酶之一的转化为3-FL的乳糖百分比。

[0274] 图6显示由不同启动子驱动的不同已鉴定的乳糖结合 α -1,3-岩藻糖基转移酶的转化为3-FL的乳糖百分比。

[0275] 图7显示了进一步实验中3-岩藻糖基乳糖的标准化的产量。

[0276] 图8显示由不同启动子驱动的已鉴定的乳糖结合 α -1,3-岩藻糖基转移酶的子集的3-岩藻糖基乳糖的标准化的产量。

[0277] 图9显示了从2个不同启动子表达幽门螺杆菌fucT(SEQ ID 18)的菌株的3-岩藻糖基乳糖的标准化的产量。

[0278] 图10显示了表达具有DM[AS]VSF共有基序的多肽的菌株的3-岩藻糖基乳糖的标准化的产量

[0279] 图11显示了SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 6、SEQ ID NO 8、SEQ ID NO 10、SEQ ID NO 12、SEQ ID NO 14、SEQ ID NO 16、SEQ ID NO 20、SEQ ID NO 22、SEQ ID NO 28、SEQ ID NO 30和SEQ ID NO 32的多肽序列的比对。用方框标记共有基序[Y/W/L/H/F/M]X[T/S/C][E/Q/D/A][K/R]、[K/D][L/K/M]XXX[F/Y]和[FW]W,其中X可以是任何不同的氨基酸。

[0280] 图12显示了SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 20、SEQ ID NO 22、SEQ ID NO 24和SEQ ID NO 26的多肽序列的比对。用方框标记共有基序DM[A/S]VSF和[N/H]XDPAXLD,

其中X可以是任何不同的氨基酸(和不相关基序)。

[0281] 图13显示了SEQ ID NO 6、SEQ ID 12、SEQ ID 14、SEQ ID NO 16、SEQ ID NO 28、SEQ ID NO 30和SEQ ID NO 32的多肽序列的比对。用方框标记共有基序K[I/V]F[F/L]XGEN (SEQ ID NO 41)和RFPLW(SEQ ID NO 42),其中X可以是任何不同的氨基酸。

实施例

[0282] 实施例1:材料和方法大肠杆菌

[0283] 培养基

[0284] Luria肉汤 (LB) 培养基由1%胰蛋白胨 (Difco, Erembodegem, 比利时)、0.5%酵母提取物 (Difco) 和0.5%氯化钠 (VWR, Leuven, 比利时) 组成。用于摇瓶实验的培养基包含2.00g/L NH_4Cl 、5.00g/L $(\text{NH}_4)_2\text{SO}_4$ 、2.993g/L KH_2PO_4 、7.315g/L K_2HPO_4 、8.372g/L MOPS、0.5g/L NaCl 、0.5g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、14.26g/L蔗糖或实施例中指定的其他碳源、1ml/L维生素溶液、100 μL /L钼酸盐溶液和1ml/L硒溶液。用1M KOH将培养基的pH值设置为7。维生素溶液由3.6g/L $\text{FeCl}_2 \cdot 4\text{H}_2\text{O}$ 、5g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 、1.3g/L $\text{MnCl}_2 \cdot 2\text{H}_2\text{O}$ 、0.38g/L $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$ 、0.5g/L $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ 、0.94g/L ZnCl_2 、0.0311g/L H_3BO_3 、0.4g/L $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$ 和1.01g/L盐酸硫胺组成。钼酸盐溶液含有0.967g/L $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$ 。硒溶液含42g/L SeO_2 。

[0285] 用于发酵的基本培养基含有6.75g/L NH_4Cl 、1.25g/L $(\text{NH}_4)_2\text{SO}_4$ 、2.93g/L KH_2PO_4 和7.31g/L KH_2PO_4 、0.5g/L NaCl 、0.5g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 、14.26g/L蔗糖、1mL/L维生素溶液、100 μL /L钼酸盐溶液和1mL/L硒溶液,组成与上述相同。

[0286] 通过高压灭菌 (121 $^\circ\text{C}$, 21') 对复合培养基进行灭菌,并通过过滤 (0.22 μm Sartorius) 对基本培养基进行灭菌。必要时,通过添加抗生素 (例如氯霉素 (20mg/L)、羧苄青霉素 (100mg/L)、壮观霉素 (40mg/L) 和/或卡那霉素 (50mg/L)) 使培养基具有选择性。

[0287] 质粒

[0288] pKD46 (Red辅助质粒,氨苄青霉素抗性)、pKD3 (包含FRT侧接的氯霉素抗性 (cat) 基因)、pKD4 (包含FRT侧接的卡那霉素抗性 (kan) 基因) 和pCP20 (表达FLP重组酶活性) 质粒从R.Cunin教授 (比利时布鲁塞尔大学, 2007年) 获得。

[0289] 利用Golden Gate组装法在pMB1 ori载体中构建 α -1,3-岩藻糖基转移酶表达质粒。使用启动子apFAB305 (“PROM0012”)、apFAB146 (“PROM0032”) (两者均如Mutalik等人 (Nat. Methods 2013, No. 10, 354-360) 所述) 以及p14 (“PROM0016”与“UTR0019”组合) (如De Mey等人 (BMC Biotechnology 2007) 所述) 和UTRs Gene10-LeuAB-BCD2 (“UTR0002”) (如Mutalik等人 (Nat. Methods 2013, No. 10, 354-360) 所述) 表达基因。

[0290] 质粒维持在从Invitrogen购买的宿主大肠杆菌DH5 α (F^- , phi80dlacZdeltaM15, delta(lacZYA-argF)U169, deoR, recA1, endA1, hsdR17 (rk^- , mk^+), phoA, supE44, lambda $^-$, thi-1, gyrA96, relA1) 中。

[0291] 菌株和突变

[0292] 大肠杆菌K12-MG1655 [lambda^- , F^- , rph-1] 于2007年3月从Coli Genetic Stock Center (US), CGSC菌株#: 7740获得。使用Datsenko和Wanner (PNAS 97 (2000), 6640-6645) 发表的技术进行基因破坏和基因导入。这项技术是基于通过lambda Red重组酶进行的同源重组后的抗生素选择。翻转酶重组酶的后续催化确保在最终生产菌株中去除抗生素选择盒。

[0293] 将携带Red辅助质粒pKD46的转化子在含有氨苄青霉素(100mg/L)和L-阿拉伯糖(10mM)的10ml LB培养基中于30℃培养至OD_{600nm}为0.6。通过第一次用50ml冰冷水洗涤,第二次用1ml冰冷水洗涤,使细胞成为电转感受态的。然后,将细胞重悬在50μl的冰冷水中。使用Gene Pulser™(BioRad) (600Ω、25μFD和250伏)用50μl细胞和10-100ng线性双链DNA产物进行电穿孔。

[0294] 电穿孔后,将细胞加入1mL LB培养基中于37℃孵育1h,并最后涂布于含有25mg/L氯霉素或50mg/L卡那霉素的LB琼脂上,以选择抗生素抗性转化子。所选择的突变体通过PCR用修饰区上游和下游的引物进行验证,并于42℃在LB琼脂中生长以失去辅助质粒。对突变体测试氨苄青霉素敏感性。

[0295] 使用pKD3、pKD4及其衍生物作为模板,通过PCR获得线性ds-DNA扩增子。所用的引物有一部分序列与模板互补,另一部分与染色体DNA上必须发生重组的那一侧互补。对于基因组敲除,同源区域被设计为目标基因的起始密码子和终止密码子的上游50-nt和下游50-nt。对于基因组敲入,必须考虑转录起点(+1)。PCR产物经PCR纯化,DpnI消化,从琼脂糖凝胶再纯化,并悬浮于洗脱缓冲液(5mM Tris, pH8.0)中。

[0296] 将选择的突变体(氯霉素或卡那霉素抗性)用pCP20质粒转化,pCP20质粒为氨苄青霉素和氯霉素抗性质粒,其显示温度敏感复制和热诱导的FLP合成。于30℃选择氨苄青霉素抗性转化子,然后于42℃在LB中纯化少数菌落,并且然后测试所有抗生素抗性和FLP辅助质粒的丢失。用对照引物(Fw/Rv-gene-out)检查基因敲除和敲入。

[0297] 通过敲除基因lacZ、lacY lacA、glgC、agp、pfkA、pfkB、pgi、arcA、iclR、wcaJ、pgi、lon和thyA,制造衍生自大肠杆菌K12-MG1655的突变菌株。此外,大肠杆菌lacY基因、来自运动发酵单胞菌(*Zymomonas mobilis*)的果糖激酶基因(frk)和来自青春双歧杆菌(*Bifidobacterium adolescentis*)的蔗糖磷酸化酶(SP)被敲入基因组并组成性表达。组成型启动子来源于De Mey等人(BMC Biotechnology, 2007)描述的启动子文库。W02016075243和W02012007481中也描述了这些基因修饰。

[0298] 所有构建的带有假定的α-1,3-岩藻糖基转移酶基因的质粒都在这株衍生自大肠杆菌K12 MG1655的突变菌株中进行了评估。所有菌株都储存在-80℃的冷冻小瓶中(过夜LB培养物与70%甘油按1:1比率混合)。表3中提供了所有成功的乳糖结合α-1,3-岩藻糖基转移酶(SEQ ID NO 1至16、19至22和27至32)以及现有技术α-1,3-岩藻糖基转移酶(SEQ ID NO 17-18)和两种非功能性α-1,3-岩藻糖基转移酶(SEQ ID NO 23至26)。

[0299] 表3:

SEQ ID	生物体	国家来源
SEQ ID 1 - 2	稻固氮螺菌 A2P	日本
SEQ ID 3 - 4	生脂固氮螺菌 B510	日本
[0300] SEQ ID 5 - 6	<i>Basilea psittacipulmonis</i>	瑞士
SEQ ID 7 - 8	<i>Planctopirus limnophila</i> (菌株 ATCC 43296 / DSM 3776 / IFAM 1008 / 290) (<i>Planctomyces limnophilus</i>)	德国
SEQ ID 9 - 10	<i>Pedobacter glucosidilyticus</i>	德国

[0301]	SEQ ID 11 - 12	卡托氏卟啉单胞菌 (WGS, 在 genbank 中: JDF01000001 至 JDF010000025)	美国
	SEQ ID 13 - 14	卟啉单胞菌属物种 COT-239 OH1446 (contig_18; NZ_JRAO01000018.1)	英国
	SEQ ID 15 - 16	损伤新月形单胞菌 ATCC 43532	未知
	SEQ ID 17 - 18	幽门螺杆菌	澳大利亚
	SEQ ID 19 - 20	固氮螺菌属物种 TSH64	日本
	SEQ ID 21 - 22	固氮螺菌属物种 TSH100	日本
	SEQ ID 23 - 24	Azospirillum brasilense	未知
	SEQ ID 25 - 26	固氮螺菌属物种 B510	日本
	SEQ ID 27 - 28	丁酸弧菌属物种 TB	未知
	SEQ ID 29 - 30	卡托氏卟啉单胞菌 F0037	未知
	SEQ ID 31 - 32	溶纤维丁酸弧菌 DSM 3071	未知

[0302] 异源和同源表达

[0303] 所有需要表达的潜在 α -1,3-岩藻糖基转移酶基因,无论是用于质粒还是用于基因组插入,都是在Twist Biosciences (旧金山,美国)合成的。通过根据表达宿主的密码子使用优化密码子使用,可以进一步促进表达。利用供应商的工具对基因进行了优化。

[0304] 培养条件

[0305] 96孔微量滴定板实验的预培养从冷冻小瓶开始,在150 μ L LB中并于37 $^{\circ}$ C在800rpm的轨道摇床上孵育过夜。将该培养物用作96孔方形微量滴定板的接种物,用400 μ L MMsf培养基稀释400倍。然后将这些最终的96孔培养板于37 $^{\circ}$ C在800rpm的轨道摇床上孵育72h或更短或更长时间。在培养实验结束时,从每个孔中采集样品,以测量肉汤上清液中的糖浓度(细胞外糖浓度,在将细胞离心之后),或在将细胞离心之前将培养液于90 $^{\circ}$ C煮沸15min(=整个肉汤测量值,细胞内和细胞外糖浓度的平均值)。

[0306] 生物反应器的预培养从特定菌株的整个1mL冷冻小瓶开始,接种在1L或2.5L摇瓶中的250mL或500mL MMsf培养基中,并于37 $^{\circ}$ C在200rpm的轨道摇床上孵育24h。然后接种5L生物反应器(在2L分批培养基中接种250mL);该过程由MFCS控制软件(Sartorius Stedim Biotech,Melsungen,德国)控制。培养条件设置为37 $^{\circ}$ C,最大搅拌;压力气体流速取决于菌株和生物反应器。使用0.5M H₂SO₄和20%NH₄OH将pH控制在6.8。冷却废气。发酵过程中发泡时加入10%的有机硅消泡剂溶液。

[0307] 光密度

[0308] 通过测量600nm处的光密度(Implen Nanophotometer NP80,Westburg,比利时或Spark 10M微孔板阅读器,Tecan,瑞士)经常监测培养物的细胞密度。

[0309] 液相色谱法

[0310] 自制3-岩藻糖基乳糖标准品。其他标准品,如但不限于乳糖、蔗糖、葡萄糖、果糖购自Sigma。

[0311] 经由HPLC-RI (Waters,美国) 方法分析碳水化合物,其中RI (折射率) 检测含有样品时流动相折射率的变化。使用X桥柱 (Waters X桥HPLC柱,美国) 和含有75ml乙腈、25ml超纯水和0.15ml三乙胺的流动相在恒流中分离糖。柱尺寸为4.6x 150mm,粒径为3.5 μ m。柱温设为35 $^{\circ}$ C,泵流速为1mL/min。

[0312] 实施例2:对掺入大肠杆菌中的不同乳糖结合 α -1,3-岩藻糖基转移酶的评估

[0313] 建立了一个实验来评估一些编码潜在 α -1,3-岩藻糖基转移酶的基因,这些酶能够从GDP-岩藻糖和乳糖产生3-岩藻糖基乳糖(3-FL)。根据实施例1中提供的培养条件进行生长试验。

[0314] 图2显示使用生产培养基中20g/L乳糖,在使用两种不同的启动子(PROM0012和PROM0016)的成功表达各种乳糖结合 α -1,3-岩藻糖基转移酶的菌株的生长实验中获得的3-岩藻糖基乳糖的标准化的产量。每个数据点对应于一个孔的数据。水平虚线表示对所有数据点进行标准化的设定点。

[0315] 实验鉴定了与具有先前证实的乳糖结合 α -1,3-岩藻糖基转移酶活性的含有SEQ ID NO 18的菌株相比,以下具有乳糖结合3-岩藻糖基转移酶活性的多肽:SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 6、SEQ ID NO 8、SEQ ID NO 10、SEQ ID NO 12和SEQ ID NO 14具有相似的或更好的乳糖结合 α -1,3-岩藻糖基转移酶活性。SEQ ID NO 4的多肽与SEQ ID NO 2具有90,8%的全序列同一性,由此表明与SEQ ID NO 2具有87%或更高序列同一性的序列也具有乳糖结合 α -1,3-岩藻糖基转移酶活性。

[0316] 实施例3:对掺入大肠杆菌中的乳糖结合 α -1,3-岩藻糖基转移酶在基本培养基中以低到高乳糖浓度产生3-FL的能力的评估

[0317] 对编码SEQ ID NO 6的基因(与PROM0016组合)在含有不同浓度乳糖的基本培养基中产生3-FL的能力进行评估。根据实施例1中提供的培养条件进行生长试验。如上所述,具有SEQ ID NO 6和SEQ ID NO 18(由PROM0016驱动)的菌株在96孔板的多个孔中生长。SEQ ID NO 18先前已证实对乳糖的 α -1,3-岩藻糖基转移酶活性。

[0318] 图3显示了3-岩藻糖基乳糖的标准化的产量,其中6种不同浓度的乳糖作为3-FL的前体(90g/L及其1:2的稀释系列,直到2.8g/L,如图所示)。每个数据点对应于一个孔的数据。水平虚线表示对所有数据点进行标准化的设定点。

[0319] 实验鉴定了与表达SEQ ID NO 18的菌株相比,SEQ ID NO 6的多肽在所有乳糖浓度下具有更好的乳糖结合 α -1,3-岩藻糖基转移酶活性,SEQ ID NO 18是具有先前证实的乳糖结合 α -1,3-岩藻糖基转移酶活性的多肽。

[0320] 实施例4:对掺入大肠杆菌中的各种乳糖结合 α -1,3-岩藻糖基转移酶在基本培养基中以低浓度乳糖产生3-FL的能力的评估

[0321] 在低浓度乳糖的生长实验中,对上述几种鉴定的具有编码SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 6、SEQ ID NO 12和SEQ ID NO 14的基因的菌株评估从GDP-岩藻糖和乳糖产生3-岩藻糖基乳糖的能力。根据实施例1中提供的培养条件进行生长试验。

[0322] 图4显示了表达各种 α -1,3-岩藻糖基转移酶(使用两种不同的启动子PROM0012和PROM0016)并在含有少量乳糖(2.8g/L乳糖)的培养基中生长的菌株的3-岩藻糖基乳糖的标准化的产量。每个数据点对应于一个孔的数据。水平虚线表示对所有数据点进行标准化的设定点。

[0323] 实验鉴定了与含有具有先前证实的乳糖结合 α -1,3-岩藻糖基转移酶活性的SEQ ID NO 18的菌株相比,当提供有低浓度的乳糖时,下述具有SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 6、SEQ ID NO 12和SEQ ID NO 14的多肽具有相似的或更好的乳糖结合 α -1,3-岩藻糖基转移酶活性。

[0324] 实施例5:对掺入大肠杆菌中的SEQ ID NO 6的多肽对两种低浓度乳糖的酶活性的评估

[0325] 在提供2.8g/L或5.62g/L的乳糖和30g/L的蔗糖的生长试验中,对编码SEQ ID NO 6的基因(并与PROM0016组合)评估在产生GDP-岩藻糖的菌株中将乳糖转化为3-岩藻糖基乳糖的能力。根据实施例1中提供的培养条件进行生长试验。

[0326] 图5显示了转化为3-FL的乳糖%,通过将测量的3-FL量除以基于乳糖输入浓度理论上可获得的量来计算。理论上,如果所有乳糖都被转化,则得到100%的值。每个数据点对应于一个孔的数据。

[0327] 将表达如SEQ ID NO 6所示的多肽的菌株与表达如SEQ ID NO 18所示的多肽的菌株(由PROM0016驱动)进行比较,所述SEQ ID NO 18先前被证实具有对乳糖的 α -1,3-岩藻糖基转移酶活性。在两种乳糖浓度下,对于给定量的碳源(30g/L蔗糖),与表达SEQ ID NO 18所示多肽的菌株相比,表达SEQ ID NO 6所示多肽的菌株能够将多得多的乳糖转化为3-FL。

[0328] 实施例6:在限制浓度的乳糖和蔗糖下,对掺入大肠杆菌中的各种乳糖结合 α -1,3-岩藻糖基转移酶的酶活性的评估

[0329] 在低浓度乳糖(2g/L)和蔗糖(7.5g/L)的生长实验中,对编码上述鉴定的多肽SEQ ID NO 2、SEQ ID NO 6、SEQ ID NO 12和SEQ ID NO 14的基因评估在产生GDP-岩藻糖的菌株中将乳糖转化为3-岩藻糖基乳糖的能力。根据实施例1中提供的培养条件进行生长试验。

[0330] 图6显示了转化为3-FL的乳糖%,通过将测量的3-FL量除以基于乳糖输入浓度理论上可获得的量来计算。每个数据点对应于一个孔的数据。

[0331] 对于给定量的碳源(7.5g/L蔗糖),表达具有SEQ ID NO 2、SEQ ID NO 6、SEQ ID NO 12或SEQ ID NO 14的多肽的菌株能够比表达具有SEQ ID NO 18的多肽的菌株将更多乳糖转化为3-FL。

[0332] 实施例7:在分批发酵中对表达各种乳糖结合 α -1,3-岩藻糖基转移酶的大肠杆菌菌株的评估

[0333] 进行生物反应器规模的分批发酵以评估如实施例1所述的衍生自突变体大肠杆菌K12 MG1655菌株背景的菌株,其表达具有SEQ ID NO 2、SEQ ID NO 6、SEQ ID NO 12和SEQ ID NO 14的各种 α -1,3-岩藻糖基转移酶。如实施例1所述进行生物反应器运行。在这些实例中,蔗糖被用作碳源。在分批培养基中以90g/L添加乳糖作为3-FL形成的前体。

[0334] 图7显示了在分批发酵中使用成功表达各种乳糖结合 α -1,3-岩藻糖基转移酶的菌株与在生产培养基中乳糖作为前体获得的3-岩藻糖基乳糖的标准化的产量。每个数据点对应于一次发酵运行的数据。水平虚线表示对所有数据点进行标准化的设定值。

[0335] 实验表明,与表达具有SEQ ID NO 18的多肽的菌株相比,表达具有SEQ ID NO 2、SEQ ID NO 6、SEQ ID NO 12或SEQ ID NO 14的乳糖结合 α -1,3-岩藻糖基转移酶基因的突变体大肠杆菌菌株产生更多量的3-FL。

[0336] 实施例8:对掺入大肠杆菌中的不同乳糖结合 α -1,3-岩藻糖基转移酶的评估

[0337] 用表达具有SEQ ID NO 2、SEQ ID NO 6、SEQ ID NO 12、SEQ ID NO 14和SEQ ID NO 16的酶的菌株进行进一步的实验,并评估在产生GDP-岩藻糖的菌株中这些酶是否能够从乳糖产生3-岩藻糖基乳糖。根据实施例1中提供的培养条件进行生长试验。

[0338] 图8显示了使用成功表达各种乳糖结合 α -1,3-岩藻糖基转移酶(使用三种不同的启动子PROM0012、PROM0016和PROM0026)的菌株在生产培养基中使用20g/L乳糖的3-岩藻糖基乳糖的标准化的产量。每个数据点对应于一个孔的数据。水平虚线表示对所有数据点进行标准化的设定点。

[0339] 实验证实了来自实施例2的表达具有SEQ ID NO 12、SEQ ID NO 6、SEQ ID NO 12和SEQ ID NO 14的多肽的菌株的结果,并且鉴定了与具有先前证实的乳糖结合 α -1,3-岩藻糖基转移酶活性的含有SEQ ID 18的菌株相比,具有SEQ ID NO 16的多肽也具有更好的乳糖结合 α -1,3-岩藻糖基转移酶活性。

[0340] 实施例9:材料和方法酿酒酵母

[0341] 培养基

[0342] 菌株生长在合成的成分确定的酵母培养基上,所述培养基具有完全补充混合物(SD CSM)或CSM drop-out(SD CSM-Ura),含有6.7g/L不含氨基酸的酵母氮基(YNB w/o AA, Difco)、20g/L琼脂(Difco)(固体培养物)、22g/L葡萄糖一水合物或20g/L乳糖和0.79g/L CSM或0.77g/L CSM Ura(MP Biomedicals)。

[0343] 菌株

[0344] 由Bachmann等人(Yeast (1998) 14:115-32)创建的酿酒酵母(*Saccharomyces cerevisiae*) BY4742从Euroscarf培养物中心获得。所有突变体菌株均采用Gietz的方法(Yeast 11:355-360,1995)通过同源重组或质粒转化产生。马克思乳酸克鲁维酵母(*Kluyveromyces marxianus lactis*)可在LMG培养物中心获得(Ghent,比利时)。

[0345] 质粒

[0346] 酵母表达质粒p2a_2 μ _sia_GFA1(Chan 2013(质粒70(2013)2-17))用于在酿酒酵母中表达外源基因。该质粒包含氨苄青霉素抗性基因和细菌复制起点,以允许在大肠杆菌中进行选择和维持。质粒中还含有质粒中还含有2 μ 酵母ori和Ura3选择标志物用于在酵母中的选择和维持。下一步,该质粒可以经修饰成p2a_2 μ _f1以含有乳糖透性酶(例如来自乳酸克鲁维酵母的LAC12)、GDP甘露糖4,6-脱水酶(例如来自大肠杆菌的Gmd)和GDP-L-岩藻糖合酶(例如来自大肠杆菌的fc1)。

[0347] 酵母表达质粒p2a_2 μ _f1_3ft基于p2a_2 μ _ft,但以也表达SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 6、SEQ ID NO 8、SEQ ID NO 10、SEQ ID NO 12、SEQ ID NO 14、SEQ ID NO 16或SEQ ID NO 18的方式进行修饰。优选地但不是必要的,岩藻糖基转移酶蛋白质在N-末端融合到SUM0star标签(例如,从pYSUM0star,Life Sensors,Malvern,PA获得)以增强岩藻糖基转移酶的溶解性。

[0348] 质粒维持在从Invitrogen购买的宿主大肠杆菌DH5 α (F⁻,phi80dlacZdeltaM15,delta(lacZYA-argF)U169,deoR,recA1,endA1,hsdR17(rk⁻,mk⁺),phoA,supE44,lambda⁻,thi-1,gyrA96,relA1)。

[0349] 基因表达启动子

[0350] 使用合成的组成型启动子来表达基因,如Blazeck(Biotechnology and

Bioengineering, Vol.109, No.11, 2012) 所述。

[0351] 异源和同源表达

[0352] 需要表达的基因,无论是来自质粒还是来自基因组,都是由以下公司之一合成的: DNA2.0、Gen9或IDT。

[0353] 通过针对表达宿主的密码子使用优化密码子使用,可以进一步促进表达。利用供应商的工具对基因进行了优化。

[0354] 培养条件

[0355] 一般来说,酵母菌株最初生长在SD-CSM平板上以获得单一菌落。这些平板于30℃生长2-3天。

[0356] 从单个菌落开始,预培养物于30℃,以200rpm摇动在5mL中过夜生长。随后的125mL摇瓶实验在25mL培养基中接种2%的这种预培养物。这些摇瓶于30℃使用200rpm的轨道振荡进行孵育。不需要使用诱导剂,因为所有基因都是组成性表达的。

[0357] 实施例10:使用各种乳糖结合 α -1,3-岩藻糖基转移酶在酿酒酵母中生产3-岩藻糖基乳糖

[0358] 另一实例提供以酿酒酵母形式的真核生物体用于本发明的用途。使用如实施例9所述的菌株、质粒和方法,创建表达SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 6、SEQ ID NO 8、SEQ ID NO 10、SEQ ID NO 12、SEQ ID NO 14、SEQ ID NO 16或SEQ ID NO 18的菌株。

[0359] 除此之外,为了生产3-岩藻糖基乳糖,还进行了进一步的修饰。这些修饰包括添加乳糖透性酶、GDP-甘露糖4,6-脱水酶和GDP-L-岩藻糖合酶。优选的乳糖透性酶是来自乳酸克鲁维酵母的KILAC12基因(WO 2016/075243)。优选的GDP-甘露糖4,6-脱水酶和GDP-L-岩藻糖合酶分别是来自大肠杆菌的gmd和fc1。

[0360] 这些菌株能够以葡萄糖或甘油为碳源生长,将碳源转化为GDP-L-岩藻糖,吸收乳糖,并使用GDP-L-岩藻糖和乳糖作为酶的底物生产3-岩藻糖基乳糖,所述酶由SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 6、SEQ ID NO 8、SEQ ID NO 10、SEQ ID NO 12、SEQ ID NO 14、SEQ ID NO 16或SEQ ID NO 18表示,以SEQ ID NO 18为参照。

[0361] 所述菌株的预培养在5mL合成的成分确定的培养基SD-CSM(含有22g/L葡萄糖)中进行,并于30℃生长,如实施例9所述。将这些预培养物接种在25mL培养基中,置于摇瓶中,以10g/L蔗糖作为唯一碳源,并于30℃生长。按照实施例1所述,采集常规样品并测量3-岩藻糖基乳糖的产生。

[0362] 实施例11:3-岩藻糖基乳糖的酶促生产

[0363] 另一实例提供了本发明的具有SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 6、SEQ ID NO 8、SEQ ID NO 10、SEQ ID NO 12、SEQ ID NO 14或SEQ ID NO 16的酶的用途。这些酶在无细胞表达系统(例如但不限于PURExpress系统(NEB)),或在宿主生物体(例如但不限于大肠杆菌或酿酒酵母)中产生,之后可分离并任选地进一步纯化上述酶。

[0364] 将上述酶提取物或纯化的酶中的每一个与GDP-岩藻糖、乳糖和缓冲组分(例如Tris-HCl或HEPES)一起添加到反应混合物中。然后将所述反应混合物于特定温度(例如37℃)孵育一定量的时间(例如24小时),在此期间,乳糖将由使用GDP-岩藻糖的酶转化为3-岩藻糖基乳糖。然后通过本领域已知的方法从反应混合物中分离3-岩藻糖基乳糖。如果优选,可进一步纯化3-FL。在反应结束时或者分离和/或纯化后,如实施例1所述测量3-岩藻糖基

乳糖的生产。

[0365] 实施例12:不同乳糖浓度的3-岩藻糖基乳糖生产

[0366] 如实施例1和实施例7所述的发酵过程,其中培养基中的乳糖浓度在50至150g/L的范围内。在该过程中,所述乳糖转化为3-岩藻糖基乳糖,直到剩余很少量的乳糖。乳糖与3-岩藻糖基乳糖的最终比率可在该过程中通过稍早停止该过程(较高的乳糖与3-岩藻糖基乳糖比率)或稍晚停止该过程(较低的乳糖与3-岩藻糖基乳糖比率)来操纵。可通过向生物反应器进料高浓度的乳糖溶液(含或不含另一碳源)来增加容器中的乳糖浓度。所述乳糖进料含有100至700g/L的乳糖浓度并保持在一定温度,使得乳糖在低于或等于6的pH保持可溶,以避免在该过程中形成乳果糖,这是乳制品工业中使用的标准方法。在这种生产过程中达到的3-岩藻糖基乳糖的最终浓度的范围在70g/L(在如上所述的过程中使用较低乳糖浓度时)至200g/L或更高(在如上所述的过程中使用较高乳糖浓度时)。

[0367] 实施例13:对从各种启动子表达的幽门螺杆菌 α -1,3-岩藻糖基转移酶fucT (SEQ ID 18)的评估

[0368] 将编码幽门螺杆菌 α -1,3-岩藻糖基转移酶fucT的基因(SEQ ID NO 18)克隆在表达载体中在启动子PROM0012或PROM0016的控制下,并将所得质粒转化到大肠杆菌突变菌株中,如实施例1所述。然后在生长试验中对这些菌株评估其产生3-FL的能力。这两个菌株在96孔板的多个孔中生长。

[0369] 图9显示了菌株产生的3-岩藻糖基乳糖的标准化的产量。每个数据点对应于一个孔的数据。水平虚线表示对所有数据点进行标准化的设定点。

[0370] 实验表明,使用启动子PROM0012表达幽门螺杆菌FucT的菌株中的3-FL产量下降到从启动子PROM0016表达岩藻糖基转移酶类似菌株中观察到的水平的 $\pm 30\%$ 。

[0371] 通过外推实施例2、4和8中提供的数据,我们可以得出结论,当岩藻糖基转移酶从同一启动子(PROM0012或PROM0016)表达时,与具有 α 1,3-岩藻糖基转移酶fucT(SEQ ID NO 18)的对照菌株相比,含有SEQ ID NO 2-16中任一个的所有菌株显示出显著更高的产量,例外的是,具有SEQ ID NO 10的菌株显示与对照菌株相似的产量。

[0372] 实施例14:对表达具有DM[AS]VSF共有基序的多肽的菌株评估3-岩藻糖基乳糖的生产

[0373] 在如实施例1所述的生长实验中,对含有用于SEQ ID NO 4、SEQ ID NO 20、SEQ ID NO 22、SEQ ID NO 24和SEQ ID NO 26任一的表达构建体的突变大肠杆菌菌株评估其3-FL产生。如图12所示,所有多肽序列均包含共有结构域DM[AS]VSF(SEQ ID NO 36),但只有SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 20和SEQ ID NO 22另外地在蛋白质的N-末端区域包含共有基序[NH]XDPAxLD(SEQ ID NO 35)。将含有幽门螺杆菌 α -1,3-岩藻糖基转移酶fucT(SEQ ID NO 18)的菌株作为阳性对照。所有菌株在96孔板的多个孔中生长,并在含30g/L蔗糖和20g/L乳糖的标准培养基中进行测试。

[0374] 图10显示了菌株产生的3-岩藻糖基乳糖的标准化的产量。每个数据点对应于一个孔的数据。水平虚线表示对所有数据点进行标准化的设定点。

[0375] 实验表明,只有含有具有共有基序[NH]xDPAXLD和DM[AS]VSF两者的多肽(即SEQ ID NO 2、SEQ ID NO 4、SEQ ID NO 20或SEQ ID NO 22)的菌株才能产生3-FL,而含有具有DM[AS]VSF但缺乏[NH]xDPAXLD的多肽(即SEQ ID NO 24和SEQ ID NO 26)的菌株不产生任

何3-FL。基于此数据,我们可以得出结论,在具有DM[AS]VSF(SEQ ID NO 36)结构域的多肽的N-末端区域存在[NH]_xDPA_xLD(SEQ ID NO 35)共有基序对于酶具有乳糖结合 α -1,3-岩藻糖基转移酶活性至关重要。

[0376] 此外,SEQ ID NO 22的多肽与SEQ ID NO 2具有92%的全局序列同一性,由此表明与SEQ ID NO 2具有87%或更高序列同一性的序列也具有乳糖结合 α -1,3-岩藻糖基转移酶活性。

[0377] 实施例15:对表达SEQ ID NO 28、SEQ ID NO 30或SEQ ID NO 32的多肽的菌株的评估

[0378] 在如实施例1所述的生长实验中,可对含有用于SEQ ID NO 28、SEQ ID NO 30或SEQ ID NO 32的表达构建体的突变大肠杆菌菌株评估其3-FL生产。在生长试验结束时,可以在培养肉汤中观察到3-岩藻糖基乳糖的生产。

[0379] 实施例16:对分批进料发酵结束时3FL纯度的评估

[0380] 进行生物反应器规模的分批进料发酵以评估衍生自突变体大肠杆菌K12MG1655菌株背景的菌株,所述菌株如实施例1所述,其表达具有SEQ ID NO 2、SEQ ID NO 6和SEQ ID NO 18的各种 α -1,3-岩藻糖基转移酶。如实施例1所述进行生物反应器运行。在这些示例中,蔗糖被用作碳源。在分批培养基中以90g/L添加乳糖作为用于3-FL形成的前体,并在分批进料期间添加浓缩蔗糖溶液。对于每个菌株,进行三次独立的发酵。

[0381] 在发酵结束时,对肉汤分析存在乳糖和3-FL的情况,并使用公式 $3FL(g/L) / (3FL(g/L) + 乳糖(g/L))$ 来计算3-FL纯度。对于含SEQ ID NO 18的菌株,获得85%的平均纯度,而对于含有SEQ ID NO 2或6的菌株,获得分别超过98%和99%的平均纯度。

[0382] 实验表明,在生物反应器规模的分批进料发酵中,表达具有SEQ ID NO 2或SEQ ID NO 6的乳糖结合 α -1,3-岩藻糖基转移酶基因的突变大肠杆菌菌株比含有SEQ ID NO 18的类似菌株产生具有更高3-FL纯度的肉汤。

[0001]	序列表	
[0002]	<110>	因比奥斯公司
[0003]	<120>	生产3-岩藻糖基乳糖和转化乳糖的 α -1,3-岩藻糖基转移酶
[0004]	<130>	005-PCT
[0005]	<160>	42
[0006]	<170>	PatentIn version 3.5
[0007]	<210>	1
[0008]	<211>	978
[0009]	<212>	DNA
[0010]	<213>	稻固氮螺菌 (<i>Azospirillum oryzae</i>)
[0011]	<400>	1
[0012]	atgattgatc agcgtaccag cgactttcta agcgagttcc tggcgtcaag ccaccgtgat	60
[0013]	ccggcgcgtt tagatagctt cctgcttcac gggccaggcc gcggcgctcg cgcggcgaaa	120
[0014]	ccgcgtctga aaattgcgtt ttttgacttt tggccggaat ttgaccggc tgcgaacttt	180
[0015]	tttgtggaca ttctgagcgc gcgctttgat gtaagcgttg tggataacga tagcgactta	240
[0016]	gcgattgtga gcgtgtttgg caccgcccat cgcgaagcgc gtaccgcgcg tagcatgttt	300
[0017]	tttaccggcg aaaatgtgcg tccgccgta gatggcgtgg atatgagcgt gagctttgat	360
[0018]	cgcattgatg acccacgcca ctatcgctg ccgttatatg tgatgcacgc gtgggatcat	420
[0019]	aggcgtgaag gggcgacgcc gcacttttgc cagagcgtgc tgccgccggt gccgccgacg	480
[0020]	cgtgaagaag cggcgaaacg taagttttgc gcgtttttat ataaaaatcc aaactgcgcg	540
[0021]	aggcgcaacg actttttcca gatgttatgc gcgaggcgc acgtggaaag cgtgggctgg	600
[0022]	ctgctgaata ataccggcag cgtggtgaaa atgggctggc tgccgaaaat tcgtgtgttc	660
[0023]	agccgctacc gctttgcgtt tgcgtttgaa aatgcgagcc acccaggcta cctgacggag	720
[0024]	aaaatttttag atgcgtttca agcgggcgcg gtgccgttat attggggcga cccaggagtg	780
[0025]	ctgcgcgacg ttgcggcccg cagctttatt gatgtgagcc gctacgcgag cgatgaagaa	840
[0026]	gcgtgcgacg cgattctggc cgcggatgat gattacgata cctaccgccg ctaccgcagc	900
[0027]	acgccgccgt ttctgggcgc ggaagatttt tactttgatg cgtttcggct ggccgagtgg	960
[0028]	attgagagcc gtctataa	978
[0029]	<210>	2
[0030]	<211>	325
[0031]	<212>	PRT
[0032]	<213>	稻固氮螺菌
[0033]	<400>	2
[0034]	Met Ile Asp Gln Arg Thr Ser Asp Phe Leu Ser Glu Phe Leu Ala Ser	
[0035]	1 5 10 15	
[0036]	Ser His Arg Asp Pro Ala Arg Leu Asp Ser Phe Leu Leu His Gly Pro	
[0037]	20 25 30	
[0038]	Gly Arg Gly Ala Arg Ala Ala Lys Pro Arg Leu Lys Ile Ala Phe Phe	
[0039]	35 40 45	
[0040]	Asp Phe Trp Pro Glu Phe Asp Pro Ala Ala Asn Phe Phe Val Asp Ile	
[0041]	50 55 60	

[0042]	Leu Ser Ala Arg Phe Asp Val Ser Val Val Asp Asn Asp Ser Asp Leu	
[0043]	65	70 75 80
[0044]	Ala Ile Val Ser Val Phe Gly Thr Arg His Arg Glu Ala Arg Thr Ala	
[0045]	85	90 95
[0046]	Arg Ser Met Phe Phe Thr Gly Glu Asn Val Arg Pro Pro Leu Asp Gly	
[0047]	100	105 110
[0048]	Val Asp Met Ser Val Ser Phe Asp Arg Ile Asp Asp Pro Arg His Tyr	
[0049]	115	120 125
[0050]	Arg Leu Pro Leu Tyr Val Met His Ala Trp Asp His Arg Arg Glu Gly	
[0051]	130	135 140
[0052]	Ala Thr Pro His Phe Cys Gln Ser Val Leu Pro Pro Val Pro Pro Thr	
[0053]	145	150 155 160
[0054]	Arg Glu Glu Ala Ala Lys Arg Lys Phe Cys Ala Phe Leu Tyr Lys Asn	
[0055]	165	170 175
[0056]	Pro Asn Cys Ala Arg Arg Asn Asp Phe Phe Gln Met Leu Cys Ala Arg	
[0057]	180	185 190
[0058]	Arg His Val Glu Ser Val Gly Trp Leu Leu Asn Asn Thr Gly Ser Val	
[0059]	195	200 205
[0060]	Val Lys Met Gly Trp Leu Pro Lys Ile Arg Val Phe Ser Arg Tyr Arg	
[0061]	210	215 220
[0062]	Phe Ala Phe Ala Phe Glu Asn Ala Ser His Pro Gly Tyr Leu Thr Glu	
[0063]	225	230 235 240
[0064]	Lys Ile Leu Asp Ala Phe Gln Ala Gly Ala Val Pro Leu Tyr Trp Gly	
[0065]	245	250 255
[0066]	Asp Pro Gly Val Leu Arg Asp Val Ala Ala Gly Ser Phe Ile Asp Val	
[0067]	260	265 270
[0068]	Ser Arg Tyr Ala Ser Asp Glu Glu Ala Cys Asp Ala Ile Leu Ala Ala	
[0069]	275	280 285
[0070]	Asp Asp Asp Tyr Asp Thr Tyr Arg Arg Tyr Arg Ser Thr Pro Pro Phe	
[0071]	290	295 300
[0072]	Leu Gly Ala Glu Asp Phe Tyr Phe Asp Ala Phe Arg Leu Ala Glu Trp	
[0073]	305	310 315 320
[0074]	Ile Glu Ser Arg Leu	
[0075]	325	
[0076]	<210> 3	
[0077]	<211> 978	
[0078]	<212> DNA	
[0079]	<213> 生脂固氮螺菌 (Azospirillum lipoferum)	
[0080]	<400> 3	
[0081]	atgattgatc agcgtacgag cgattttctg agcgagtttc tagcgtcacc aaaccgtgat	60
[0082]	ccagcggtgt tagatcggtt tctgcttcat gggccggagc gcgggggccc tgcggccaga	120
[0083]	ccgcgcctga aaattgcgtt ttttgacttt tggccggaat ttgaccgag cgcaattttt	180

[0084]	ttttagaaaa ttctgagcag cgcctttgat gtgagcgtgg ttgataatga tagcgattta	240
[0085]	gcgattctga gcgtgtttgg cgaacgccac cgcgaagcgc gtaccgcgcg cgcgctgttt	300
[0086]	tttaccggcg aaaatgtgcg cccgccgtta gatggcgtgg atatgagcgt gagctttgat	360
[0087]	cgcattgatc atccacgtca ttatcgctg cgtttatacg tgatgcatgc gtgggatcac	420
[0088]	cgtcgcgaag gggcgacccc gcatttttgc catccggtgc tgccgccggt gccgccgacg	480
[0089]	cgtgaagaag cggcgaaacg taagttttgc gcgtttttat ataaaaatcc tcaactgcgcg	540
[0090]	cgccgaacg atttttttca gatgctgtgc gcgcggcgcc atgtggaaag cgtgggctgg	600
[0091]	ctgctgaata ataccggcag cgtggtgaaa atgggctggc tgccgaaaat tcgcgtgttt	660
[0092]	gcgcgctatc gctttgcgtt tgcgtttgaa aacgcggcgc atccaggcta tctgaccgag	720
[0093]	aaaattcttg atgcgtttca ggccgggacg gtaccgttat actggggcga cagcggcgtg	780
[0094]	ctgcgcgacg ttgcggccgg cagctttatt gatgtgagcc gctatgcgag cgatgaagaa	840
[0095]	gcgattgaag cgattctggc gattgatgat gactatgata gctatcgccg gtaccgcggc	900
[0096]	acggcgccat ttttaggcac cgaggacttt tactttgacg cgtaccggct ggccgagtgg	960
[0097]	attgagagcc gcctgtag	978
[0098]	<210> 4	
[0099]	<211> 325	
[0100]	<212> PRT	
[0101]	<213> 生脂固氮螺菌	
[0102]	<400> 4	
[0103]	Met Ile Asp Gln Arg Thr Ser Asp Phe Leu Ser Glu Phe Leu Ala Ser	
[0104]	1 5 10 15	
[0105]	Pro Asn Arg Asp Pro Ala Val Leu Asp Arg Phe Leu Leu His Gly Pro	
[0106]	20 25 30	
[0107]	Glu Arg Gly Gly Arg Ala Ala Arg Pro Arg Leu Lys Ile Ala Phe Phe	
[0108]	35 40 45	
[0109]	Asp Phe Trp Pro Glu Phe Asp Pro Ser Ala Asn Phe Phe Val Glu Ile	
[0110]	50 55 60	
[0111]	Leu Ser Ser Arg Phe Asp Val Ser Val Val Asp Asn Asp Ser Asp Leu	
[0112]	65 70 75 80	
[0113]	Ala Ile Leu Ser Val Phe Gly Glu Arg His Arg Glu Ala Arg Thr Ala	
[0114]	85 90 95	
[0115]	Arg Ala Leu Phe Phe Thr Gly Glu Asn Val Arg Pro Pro Leu Asp Gly	
[0116]	100 105 110	
[0117]	Val Asp Met Ser Val Ser Phe Asp Arg Ile Asp His Pro Arg His Tyr	
[0118]	115 120 125	
[0119]	Arg Leu Pro Leu Tyr Val Met His Ala Trp Asp His Arg Arg Glu Gly	
[0120]	130 135 140	
[0121]	Ala Thr Pro His Phe Cys His Pro Val Leu Pro Pro Val Pro Pro Thr	
[0122]	145 150 155 160	
[0123]	Arg Glu Glu Ala Ala Lys Arg Lys Phe Cys Ala Phe Leu Tyr Lys Asn	
[0124]	165 170 175	
[0125]	Pro His Cys Ala Arg Arg Asn Asp Phe Phe Gln Met Leu Cys Ala Arg	

[0126]	180	185	190
[0127]	Arg His Val Glu Ser Val Gly Trp Leu Leu Asn Asn Thr Gly Ser Val		
[0128]	195	200	205
[0129]	Val Lys Met Gly Trp Leu Pro Lys Ile Arg Val Phe Ala Arg Tyr Arg		
[0130]	210	215	220
[0131]	Phe Ala Phe Ala Phe Glu Asn Ala Ala His Pro Gly Tyr Leu Thr Glu		
[0132]	225	230	235
[0133]	Lys Ile Leu Asp Ala Phe Gln Ala Gly Thr Val Pro Leu Tyr Trp Gly		
[0134]	245	250	255
[0135]	Asp Ser Gly Val Leu Arg Asp Val Ala Ala Gly Ser Phe Ile Asp Val		
[0136]	260	265	270
[0137]	Ser Arg Tyr Ala Ser Asp Glu Glu Ala Ile Glu Ala Ile Leu Ala Ile		
[0138]	275	280	285
[0139]	Asp Asp Asp Tyr Asp Ser Tyr Arg Arg Tyr Arg Gly Thr Ala Pro Phe		
[0140]	290	295	300
[0141]	Leu Gly Thr Glu Asp Phe Tyr Phe Asp Ala Tyr Arg Leu Ala Glu Trp		
[0142]	305	310	315
[0143]	Ile Glu Ser Arg Leu		
[0144]	325		
[0145]	<210> 5		
[0146]	<211> 990		
[0147]	<212> DNA		
[0148]	<213> Basilea psittacipulmonis		
[0149]	<400> 5		
[0150]	atgtcagtgc tgaaaaaact ggtaagaacg ctgaaaaaga aaaaagatat tcccagcgaa	60	
[0151]	aatcaggaag atattaagcc gcaggaatth ggacatatca aacattatca tttctggcct	120	
[0152]	ctgtcgaacg aaacgttttt taaccaatth gcgcaggaaa aaaaccttga tctgagccag	180	
[0153]	acggccctaa ttagctgctt tggggaacta tcagcgattc ctaaaatccc tgaacggtat	240	
[0154]	aaagtgtttt ttacggggga aaatatctat catccggatc gcatctcata ttcgatccg	300	
[0155]	gagctctatc gcatggtgga tctgtattta ggctttgaat accggacgga accgaagtat	360	
[0156]	ctacgctttc cgctgtgggt ttggtattta tgcggcctga cgaaaaaacc gcatttttca	420	
[0157]	catgaatcaa ttgcggaatt tattcggaat atgaaccaac ctgagtttcg cctgcagtca	480	
[0158]	tcaaggaatc gcttttgcag ccatattagc agccatgata cgaacggcat tcggaaacgg	540	
[0159]	atgattgatt taattttacc gattgcgtca gtggattgcg cggcaagtt tatgaacaac	600	
[0160]	acggatgagt taaaagcgaa gtttaatgat gacaagatcg actatctaaa acagtatcgg	660	
[0161]	tttaattctct gtcctgaaaa ctcagaatca gtcgggtata tcacggaaaa gatttttgaa	720	
[0162]	tcaattatgg ccgggtgcat tccaatttat tggggaggag tcaagcagct ttttgtcgaa	780	
[0163]	ccggatattt taaatccgga agcatttatt tactacgaaa aagggaaga agagcaatta	840	
[0164]	gcgaacacagg ttgaagaact ttgatataca cctaaacggg atgaagagtt tgcagcaatt	900	
[0165]	gccccgttta aagaggacgc agccgaagtg atttatacgt ggattgaaga actggaaaaa	960	
[0166]	cggctacggg catttgaacc aaaagcctaa	990	
[0167]	<210> 6		

[0168]	<211>	329	
[0169]	<212>	PRT	
[0170]	<213>	Basilea psittacipulmonis	
[0171]	<400>	6	
[0172]	Met Ser Val Leu Lys Lys Leu Val Arg Thr Leu Lys Lys Lys Lys Asp		
[0173]	1	5	10 15
[0174]	Ile Pro Ser Glu Asn Gln Glu Asp Ile Lys Pro Gln Glu Phe Gly His		
[0175]		20	25 30
[0176]	Ile Lys His Tyr His Phe Trp Pro Leu Ser Asn Glu Thr Phe Phe Asn		
[0177]		35	40 45
[0178]	Gln Phe Ala Gln Glu Lys Asn Leu Asp Leu Ser Gln Thr Ala Leu Ile		
[0179]		50	55 60
[0180]	Ser Cys Phe Gly Glu Leu Ser Ala Ile Pro Lys Ile Pro Glu Arg Tyr		
[0181]	65	70	75 80
[0182]	Lys Val Phe Phe Thr Gly Glu Asn Ile Tyr His Pro Asp Arg Ile Ser		
[0183]		85	90 95
[0184]	Tyr Ser Asp Pro Glu Leu Tyr Arg Met Val Asp Leu Tyr Leu Gly Phe		
[0185]		100	105 110
[0186]	Glu Tyr Arg Thr Glu Pro Lys Tyr Leu Arg Phe Pro Leu Trp Val Trp		
[0187]		115	120 125
[0188]	Tyr Leu Cys Gly Leu Thr Lys Lys Pro His Phe Ser His Glu Ser Ile		
[0189]		130	135 140
[0190]	Ala Glu Phe Ile Arg Lys Met Asn Gln Pro Glu Phe Arg Leu Gln Ser		
[0191]	145	150	155 160
[0192]	Ser Arg Asn Arg Phe Cys Ser His Ile Ser Ser His Asp Thr Asn Gly		
[0193]		165	170 175
[0194]	Ile Arg Lys Arg Met Ile Asp Leu Ile Leu Pro Ile Ala Ser Val Asp		
[0195]		180	185 190
[0196]	Cys Ala Gly Lys Phe Met Asn Asn Thr Asp Glu Leu Lys Ala Lys Phe		
[0197]		195	200 205
[0198]	Asn Asp Asp Lys Ile Asp Tyr Leu Lys Gln Tyr Arg Phe Asn Leu Cys		
[0199]		210	215 220
[0200]	Pro Glu Asn Ser Glu Ser Val Gly Tyr Ile Thr Glu Lys Ile Phe Glu		
[0201]	225	230	235 240
[0202]	Ser Ile Met Ala Gly Cys Ile Pro Ile Tyr Trp Gly Gly Val Lys Gln		
[0203]		245	250 255
[0204]	Leu Phe Val Glu Pro Asp Ile Leu Asn Pro Glu Ala Phe Ile Tyr Tyr		
[0205]		260	265 270
[0206]	Glu Lys Gly Lys Glu Glu Gln Leu Ala Lys Gln Val Glu Glu Leu Trp		
[0207]		275	280 285
[0208]	Ile Ser Pro Lys Arg Tyr Glu Glu Phe Ala Ala Ile Ala Pro Phe Lys		
[0209]		290	295 300

[0210]	Glu Asp Ala Ala Glu Val Ile Tyr Thr Trp Ile Glu Glu Leu Glu Lys	
[0211]	305	310 315 320
[0212]	Arg Leu Arg Ala Phe Glu Pro Lys Ala	
[0213]	325	
[0214]	<210> 7	
[0215]	<211> 930	
[0216]	<212> DNA	
[0217]	<213> Planctopirus limnophila	
[0218]	<400> 7	
[0219]	atgcaaaaaa ctttttctct acgtccagtc ccagtggatg caattcgatg gaactttgca	60
[0220]	aacttttggg cgggtttcga tgcattagca ttcgagcggc acctgttagg tgtctctggg	120
[0221]	aaacataagt tccgattag tgaacaaaac cctcagattg tgttcgaatc ggtctttggg	180
[0222]	actccaggga aggggcgcga gcatggcca aaggcacgac aggtgtggtg tacgggagaa	240
[0223]	aacgtcgcac caccactgga tcagtttgat aaatgtttat cgttccatcg ggatattaag	300
[0224]	gatccacggc atttgcgttg gccatactat ctactgcaact tagcaagctt accaatgtct	360
[0225]	ttcaatgatt tggatgaagt ccagagctca gtgtcgactt gggcagaacg tccagggttc	420
[0226]	tgtgcattta ttgcatttaa cgagggatgc cagacgcgga accggtttgt ggaaaagcta	480
[0227]	agtcgatacc gtcgagtgga ttgtccaggt cgggtcttaa ataatatgac gagtgagaca	540
[0228]	ctgggtcagc gagggaactt gcatgggaaa attaaactcc tgaagcaata taaatacgca	600
[0229]	gtgtgcttcg agaatactag cacgcgagga tcagaagggt atgtgacgga aaagctggtt	660
[0230]	gacgcgatgc tggcaggtcg catacatta tactggggcg accaccgggt tggggaagat	720
[0231]	tttaacgaga actcgttcat taacttagga gtatacggga acgatgtgaa tgcaatggtc	780
[0232]	cagcatgtga ttgaactgga ttctgacgaa cggttgcaa ataacctgtt tcaagagcca	840
[0233]	tggttgccag aaattaagtc gtcagagcac ttctctttcg aaacgagcaa ggatgcaatt	900
[0234]	ctgaagttag tggcaaacgt aaataaatga	930
[0235]	<210> 8	
[0236]	<211> 309	
[0237]	<212> PRT	
[0238]	<213> Planctopirus limnophila	
[0239]	<400> 8	
[0240]	Met Gln Lys Thr Phe Ser Leu Arg Pro Val Pro Val Asp Ala Ile Arg	
[0241]	1	5 10 15
[0242]	Trp Asn Phe Ala Asn Phe Trp Ser Gly Phe Asp Ala Leu Ala Phe Glu	
[0243]	20	25 30
[0244]	Arg His Leu Leu Gly Val Ser Gly Lys His Lys Phe Arg Ile Ser Glu	
[0245]	35	40 45
[0246]	Gln Asn Pro Gln Ile Val Phe Glu Ser Val Phe Gly Thr Pro Gly Lys	
[0247]	50	55 60
[0248]	Gly Arg Glu Arg Trp Pro Lys Ala Arg Gln Val Trp Tyr Thr Gly Glu	
[0249]	65	70 75 80
[0250]	Asn Val Ala Pro Pro Leu Asp Gln Phe Asp Lys Cys Leu Ser Phe His	
[0251]	85	90 95

[0252]	Arg Asp Ile Lys Asp Pro Arg His Leu Arg Trp Pro Tyr Tyr Leu Leu	
[0253]	100	105 110
[0254]	His Leu Ala Ser Leu Pro Met Ser Phe Asn Asp Leu Val Lys Cys Gln	
[0255]	115	120 125
[0256]	Ser Ser Val Ser Thr Trp Ala Glu Arg Pro Gly Phe Cys Ala Phe Ile	
[0257]	130	135 140
[0258]	Ala Phe Asn Glu Gly Cys Gln Thr Arg Asn Arg Phe Val Glu Lys Leu	
[0259]	145	150 155 160
[0260]	Ser Arg Tyr Arg Arg Val Asp Cys Pro Gly Arg Val Leu Asn Asn Met	
[0261]	165	170 175
[0262]	Thr Ser Glu Thr Leu Gly Gln Arg Gly Asn Leu His Gly Lys Ile Asn	
[0263]	180	185 190
[0264]	Phe Leu Lys Gln Tyr Lys Tyr Ala Val Cys Phe Glu Asn Thr Ser Thr	
[0265]	195	200 205
[0266]	Arg Gly Ser Glu Gly Tyr Val Thr Glu Lys Leu Val Asp Ala Met Leu	
[0267]	210	215 220
[0268]	Ala Gly Cys Ile Pro Leu Tyr Trp Gly Asp His Arg Val Gly Glu Asp	
[0269]	225	230 235 240
[0270]	Phe Asn Glu Asn Ser Phe Ile Asn Leu Gly Val Tyr Gly Asn Asp Val	
[0271]	245	250 255
[0272]	Asn Ala Met Val Gln His Val Ile Glu Leu Asp Ser Asp Glu Arg Leu	
[0273]	260	265 270
[0274]	Gln Asn Asn Leu Phe Gln Glu Pro Trp Leu Pro Glu Ile Lys Ser Ser	
[0275]	275	280 285
[0276]	Glu His Phe Ser Phe Glu Thr Ser Lys Asp Ala Ile Leu Lys Leu Val	
[0277]	290	295 300
[0278]	Ala Asn Val Asn Lys	
[0279]	305	
[0280]	<210> 9	
[0281]	<211> 951	
[0282]	<212> DNA	
[0283]	<213> Pedobacter glucosidilyticus	
[0284]	<400> 9	
[0285]	atgaagtatt ttctgctgct ggctaaacgc cacaagaaat acctgaaaga gaagaaaatc	60
[0286]	ttccgcaact cgactatctc tttttacaac ttctgggaaa tcgaggatta caacaacttt	120
[0287]	tggctgcaga aattcatcgt agaccgtaac ctgaacccaa aaaacaaatc catcaacttc	180
[0288]	ttttctgtgt tcggcccgcg ctatgtcctt aaaaagcaaa aagcagcgat caatatatttc	240
[0289]	ttctcgggcg aaaccatgag ccgtttcaaa aaataccacg actattgtct gcctgaagtt	300
[0290]	gatctggcgc tgggtttcga cgatctgcaa cagagaagt acttccgtct gccgctgtgg	360
[0291]	atcctggact ttttcgaacc gactgttgac cttgaaaaag ctaaagaaaa actgaaacag	420
[0292]	ctgaactact acaaaaacaa taaaccgatc gtgcgtgaaa agttctgctc tctgatcgcc	480
[0293]	cgtcacgacg aaaacggcat ccgtaaaaag attgtgaaca cgctgaaccc aatcgaaacg	540

[0294]	gttgactgtg caggcaaact gttcaacaat actgctcgct tacagaccga attcgcgaaac	600
[0295]	aacaaagtaa aattttctgga gaactacaag tttaacatct gcccggaata caccaaccag	660
[0296]	gaatcctaca ccaccgaaaa acttttcgaa agcttcgctg caggctgtat cccgatctac	720
[0297]	tggggttctg ctcagaaacc ggaaccgaac atcttcaaac cgtctagcat catctttttc	780
[0298]	gatgagttca aaaacaccct gtctgaggat gttgaacgtc tacataaaga tccgaaactg	840
[0299]	tacctggact ttattagcca gaaccggttt caggacacag ctgctgaata catcatccag	900
[0300]	actatctcca acctggaact caaactgaaa gagatcatta accaggcata a	951
[0301]	<210>	10
[0302]	<211>	316
[0303]	<212>	PRT
[0304]	<213>	Pedobacter glucosidilyticus
[0305]	<400>	10
[0306]	Met Lys Tyr Phe Leu Leu Leu Ala Lys Arg His Lys Lys Tyr Leu Lys	
[0307]	1 5 10 15	
[0308]	Glu Lys Lys Ile Phe Arg Asn Ser Thr Ile Ser Phe Tyr Asn Phe Trp	
[0309]	20 25 30	
[0310]	Glu Ile Glu Asp Tyr Asn Asn Phe Trp Leu Gln Lys Phe Ile Val Asp	
[0311]	35 40 45	
[0312]	Arg Asn Leu Asn Pro Lys Asn Lys Ser Ile Asn Phe Phe Ser Val Phe	
[0313]	50 55 60	
[0314]	Gly Pro Arg Tyr Val Leu Lys Lys Gln Lys Ala Ala Ile Asn Ile Phe	
[0315]	65 70 75 80	
[0316]	Phe Ser Gly Glu Thr Met Ser Arg Phe Lys Lys Tyr His Asp Tyr Cys	
[0317]	85 90 95	
[0318]	Leu Pro Glu Val Asp Leu Ala Leu Gly Phe Asp Asp Leu Gln His Glu	
[0319]	100 105 110	
[0320]	Lys Tyr Phe Arg Leu Pro Leu Trp Ile Leu Asp Phe Phe Glu Pro Thr	
[0321]	115 120 125	
[0322]	Val Asp Leu Glu Lys Ala Lys Glu Lys Leu Lys Gln Leu Asn Tyr Tyr	
[0323]	130 135 140	
[0324]	Lys Asn Asn Lys Pro Ile Val Arg Glu Lys Phe Cys Ser Leu Ile Ala	
[0325]	145 150 155 160	
[0326]	Arg His Asp Glu Asn Gly Ile Arg Lys Lys Ile Val Asn Thr Leu Asn	
[0327]	165 170 175	
[0328]	Pro Ile Glu Thr Val Asp Cys Ala Gly Lys Leu Phe Asn Asn Thr Ala	
[0329]	180 185 190	
[0330]	Arg Leu Gln Thr Glu Phe Ala Asn Asn Lys Val Lys Phe Leu Glu Asn	
[0331]	195 200 205	
[0332]	Tyr Lys Phe Asn Ile Cys Pro Glu Asn Thr Asn Gln Glu Ser Tyr Thr	
[0333]	210 215 220	
[0334]	Thr Glu Lys Leu Phe Glu Ser Phe Ala Ala Gly Cys Ile Pro Ile Tyr	
[0335]	225 230 235 240	

[0336]	Trp Gly Ser Ala Gln Lys Pro Glu Pro Asn Ile Phe Lys Pro Ser Ser	
[0337]	245	250 255
[0338]	Ile Ile Phe Phe Asp Glu Phe Lys Asn Thr Leu Ser Glu Asp Val Glu	
[0339]	260	265 270
[0340]	Arg Leu His Lys Asp Pro Lys Leu Tyr Leu Asp Phe Ile Ser Gln Asn	
[0341]	275	280 285
[0342]	Pro Phe Gln Asp Thr Ala Ala Glu Tyr Ile Ile Gln Thr Ile Ser Asn	
[0343]	290	295 300
[0344]	Leu Glu Leu Lys Leu Lys Glu Ile Ile Asn Gln Ala	
[0345]	305	310 315
[0346]	<210> 11	
[0347]	<211> 1029	
[0348]	<212> DNA	
[0349]	<213> 卡托氏卟啉单胞菌 (Porphyromonas catoniae)	
[0350]	<400> 11	
[0351]	atgccgatat atgatataaa agccatgaat accccgtcga agcaaccatt acgagaacga	60
[0352]	ctgcatatga tgcgtcgacg taatcgtgtg cggaagcgtt cggttatagc cctgattaag	120
[0353]	tcgcathtag atagctcacg ttatcaggat tataactggg gggattctca tgcgtctacc	180
[0354]	ttctggttac cacggtttat agatttgcatt ctggaaccga agaagaaaat taattttattt	240
[0355]	tcgtgctttc aaaatccgtt aatgctgatt cgttattata aagggtgtta aatttttttta	300
[0356]	tcagggtgaaa acctgaccaa taacgaacat ttccggtttc acccgcgtat gctggatcac	360
[0357]	cgaataaatg aagttgatct agcgttaggt ttccaatttc gtaaagacct gaaatattac	420
[0358]	cgttttccgt tatggattta ccagaatgaa tttattagcc cgtctgccag cttagaggac	480
[0359]	atatgtgttc tggtaggcca gataaacgac ccgtcgacct gtcgtagcgc caagcgttct	540
[0360]	cgttttattg gtcagatttc gagccatgat aagggtggca tgcggggacg gctgatagat	600
[0361]	ctgttatctc cgattgggca aattgattgc gccggtaaat ttctgcataa taccgacgaa	660
[0362]	ctgttagaag tgtatggtga cgataaatc aaatatttag ccaactatcg ttttaactta	720
[0363]	tgcccagaaa attcattagg tgagggtat attaccgaaa aagtgtttga tagcatacgt	780
[0364]	gccggttgta ttccgattta ctggggcgcc tatttagaac cagggtattt aaaccgaaa	840
[0365]	gcgattttac gttttgaaga gggcaaagag caagaatttt ataaccgggt taaagaatta	900
[0366]	tgggaaaacg aggaagcgta tgaacagttc attctggaac caccgtttgt cgaaggggcc	960
[0367]	gccgaacgta tttgggaaat tttgcagggt ttacgtgaac gtttagcgcc attagttgaa	1020
[0368]	gaagggtaa	1029
[0369]	<210> 12	
[0370]	<211> 342	
[0371]	<212> PRT	
[0372]	<213> 卡托氏卟啉单胞菌	
[0373]	<400> 12	
[0374]	Met Pro Ile Tyr Asp Ile Lys Ala Met Asn Thr Pro Ser Lys Gln Pro	
[0375]	1	5 10 15
[0376]	Leu Arg Glu Arg Leu His Met Met Arg Arg Arg Asn Arg Val Arg Lys	
[0377]	20	25 30

[0378]	Arg Ser Val Ile Ala Leu Ile Lys Ser His Leu Asp Ser Ser Arg Tyr
[0379]	35 40 45
[0380]	Gln Asp Tyr Asn Trp Trp Asp Ser His Ala Ser Thr Phe Trp Leu Pro
[0381]	50 55 60
[0382]	Arg Phe Ile Asp Leu His Leu Glu Pro Lys Lys Lys Ile Asn Leu Phe
[0383]	65 70 75 80
[0384]	Ser Cys Phe Gln Asn Pro Leu Met Leu Ile Arg Tyr Tyr Lys Gly Val
[0385]	85 90 95
[0386]	Lys Ile Phe Leu Ser Gly Glu Asn Leu Thr Asn Asn Glu His Phe Gly
[0387]	100 105 110
[0388]	Phe His Pro Arg Met Leu Asp His Arg Ile Asn Glu Val Asp Leu Ala
[0389]	115 120 125
[0390]	Leu Gly Phe Glu Phe Arg Lys Asp Pro Lys Tyr Tyr Arg Phe Pro Leu
[0391]	130 135 140
[0392]	Trp Ile Tyr Gln Asn Glu Phe Ile Ser Pro Ser Ala Ser Leu Glu Asp
[0393]	145 150 155 160
[0394]	Ile Cys Val Leu Val Gly Gln Ile Asn Asp Pro Ser Thr Arg Arg Ser
[0395]	165 170 175
[0396]	Ala Lys Arg Ser Arg Phe Ile Gly Gln Ile Ser Ser His Asp Lys Gly
[0397]	180 185 190
[0398]	Gly Met Arg Gly Arg Leu Ile Asp Leu Leu Ser Pro Ile Gly Gln Ile
[0399]	195 200 205
[0400]	Asp Cys Ala Gly Lys Phe Arg His Asn Thr Asp Glu Leu Leu Glu Val
[0401]	210 215 220
[0402]	Tyr Gly Asp Asp Lys Phe Lys Tyr Leu Ala Asn Tyr Arg Phe Asn Leu
[0403]	225 230 235 240
[0404]	Cys Pro Glu Asn Ser Leu Gly Glu Gly Tyr Ile Thr Glu Lys Val Phe
[0405]	245 250 255
[0406]	Asp Ser Ile Arg Ala Gly Cys Ile Pro Ile Tyr Trp Gly Ala Tyr Leu
[0407]	260 265 270
[0408]	Glu Pro Gly Ile Leu Asn Pro Lys Ala Ile Leu Arg Phe Glu Glu Gly
[0409]	275 280 285
[0410]	Lys Glu Gln Glu Phe Tyr Asn Arg Val Lys Glu Leu Trp Glu Asn Glu
[0411]	290 295 300
[0412]	Glu Ala Tyr Glu Gln Phe Ile Leu Glu Pro Pro Phe Val Glu Gly Ala
[0413]	305 310 315 320
[0414]	Ala Glu Arg Ile Trp Glu Ile Leu Gln Gly Leu Arg Glu Arg Leu Ala
[0415]	325 330 335
[0416]	Pro Leu Val Glu Glu Gly
[0417]	340
[0418]	<210> 13
[0419]	<211> 972

[0420]	<212>	DNA	
[0421]	<213>	卟啉单胞菌属物种(Porphyrromonas sp.)	
[0422]	<400>	13	
[0423]	atgaacgcgg	tcgagcgagt	acggaatata cttaattatt gtattaacga agtccaaatg 60
[0424]	taccggcagt	gtcccaattc	aaaatactat aattttctggc cctgtgatta taataataat 120
[0425]	tggtttaacc	atttcgtaga	acaccgaggc ttagctaaag aacggcaccg gcttaacttt 180
[0426]	ttctcggctc	ttggtaaccc	tctactgccg cggattatac cggggaagaa agtgttcttc 240
[0427]	actggggaga	atcttgcaga	taactcaata cactcaatag ggcgagcttt caaaaagacc 300
[0428]	tttccggtat	atgatctggt	acttgggttc gactatgaag tagaggatag ccgggtgaat 360
[0429]	tatatgcggt	ttccattatg	gatagccttc ctgatagatc cgaccgccga ttatcagaaa 420
[0430]	ataaaggaaa	cgattgaacg	gattaacgac ccgtcaacgc ggcttaacgc gagccgggat 480
[0431]	cgtttcgect	gccttggtgc	cagccacgat aaaactggta tacggcagaa attatatgat 540
[0432]	gtccttatgc	cgatagcgtc	agtaacttgc ccaggacggt tccagaataa tacgaacgag 600
[0433]	cttcacgatt	tatatgcaaa	cgacaagcgg gaatatttaa aactgtttaa atttaacgta 660
[0434]	tgtccagaaa	attcatcgac	tccgggttat ataactgaaa agttattcga ttcgttcgca 720
[0435]	tcaggttgta	ttcccatata	cttcggtggg ggaactgagg aaatagagcc cgatattgta 780
[0436]	aaccaaggag	cgttcatacg	gtactgggat gatgggcgaa tggactggat ggacacgta 840
[0437]	cgggaacttt	gggaatcgcc	gtcagcatat cgggccgttg ccgagatacc accgttcaaa 900
[0438]	gaacaagcag	cagatgtaat	ttatgcctat atggaaaacc ttcacgacaa acttgcggca 960
[0439]	atagtgcggt	ga	972
[0440]	<210>	14	
[0441]	<211>	323	
[0442]	<212>	PRT	
[0443]	<213>	卟啉单胞菌属物种	
[0444]	<400>	14	
[0445]	Met Asn Ala Val	Glu Arg Val Arg Asn Ile Leu Asn Tyr Cys Ile Asn	
[0446]	1	5	10 15
[0447]	Glu Val Gln Met Tyr Arg Gln Cys Pro Asn Ser Lys Tyr Tyr Asn Phe		
[0448]	20	25	30
[0449]	Trp Pro Cys Asp Tyr Asn Asn Asn Trp Phe Asn His Phe Val Glu His		
[0450]	35	40	45
[0451]	Arg Gly Leu Ala Lys Glu Arg His Arg Leu Asn Phe Phe Ser Val Phe		
[0452]	50	55	60
[0453]	Gly Asn Pro Leu Leu Pro Arg Ile Ile Pro Gly Lys Lys Val Phe Phe		
[0454]	65	70	75 80
[0455]	Thr Gly Glu Asn Leu Ala Asp Asn Ser Ile His Ser Ile Gly Arg Ala		
[0456]	85	90	95
[0457]	Phe Lys Lys Thr Phe Pro Val Tyr Asp Leu Val Leu Gly Phe Asp Tyr		
[0458]	100	105	110
[0459]	Glu Val Glu Asp Ser Arg Val Asn Tyr Met Arg Phe Pro Leu Trp Ile		
[0460]	115	120	125
[0461]	Ala Phe Leu Ile Asp Pro Thr Ala Asp Tyr Gln Lys Ile Lys Glu Thr		

[0462]	130	135	140
[0463]	Ile Glu Arg Ile Asn Asp Pro Ser Thr Arg Leu Asn Ala Ser Arg Asp		
[0464]	145	150	155 160
[0465]	Arg Phe Ala Cys Leu Val Ala Ser His Asp Lys Thr Gly Ile Arg Gln		
[0466]	165	170	175
[0467]	Lys Leu Tyr Asp Val Leu Met Pro Ile Ala Ser Val Thr Cys Pro Gly		
[0468]	180	185	190
[0469]	Arg Phe Gln Asn Asn Thr Asn Glu Leu His Asp Leu Tyr Ala Asn Asp		
[0470]	195	200	205
[0471]	Lys Arg Glu Tyr Leu Lys Leu Phe Lys Phe Asn Val Cys Pro Glu Asn		
[0472]	210	215	220
[0473]	Ser Ser Thr Pro Gly Tyr Ile Thr Glu Lys Leu Phe Asp Ser Phe Ala		
[0474]	225	230	235 240
[0475]	Ser Gly Cys Ile Pro Ile Tyr Phe Gly Gly Gly Thr Glu Glu Ile Glu		
[0476]	245	250	255
[0477]	Pro Asp Ile Val Asn Gln Gly Ala Phe Ile Arg Tyr Trp Asp Asp Gly		
[0478]	260	265	270
[0479]	Arg Met Asp Trp Met Asp Thr Val Arg Glu Leu Trp Glu Ser Pro Ser		
[0480]	275	280	285
[0481]	Ala Tyr Arg Ala Val Ala Glu Ile Pro Pro Phe Lys Glu Gln Ala Ala		
[0482]	290	295	300
[0483]	Asp Val Ile Tyr Ala Tyr Met Glu Asn Leu His Asp Lys Leu Ala Ala		
[0484]	305	310	315 320
[0485]	Ile Val Arg		
[0486]	<210> 15		
[0487]	<211> 951		
[0488]	<212> DNA		
[0489]	<213> 损伤新月形单胞菌 (Selenomonas infelix)		
[0490]	<400> 15		
[0491]	atgttaatgc gtgcactcag aaaaatgaag cgatggggac gtgtggcggt tgattacacg	60	
[0492]	aatactacga aggatggagc tgtttgtctat cataattggt ggccgtgtaa ttatgaggaa	120	
[0493]	gagtggtttc atcgttttgt tgtacaaaat attggaacag aacgttgcta tcatctcttt	180	
[0494]	tctgtatttg gtccacgtat tgcgttgacg ctgccaacac cgaataaagt ttttttctgt	240	
[0495]	ggtgaaaatg tgcataacgc agagtggccc tataaaagct atcaagatca tgcacttgga	300	
[0496]	gatgtcaagc tggctctcgg atatgatgat atacaggatg aacgatatat tcgatttcct	360	
[0497]	ctgtgggtgc tctatatgtt cgatcctgtt gttgaccgat atgccatccg tgagcgaatt	420	
[0498]	gaagaaatca atcatgcaga gaatacaaga aaatatgaat gtgtattgat ttccagacac	480	
[0499]	gataagtgga atatgcgtgg tccaatttat gatgcattga aagatcattt ggctatttcc	540	
[0500]	tgtgctggga aatggaagca aaacactgat gaactgtgga cggttttacaa tgatgataaa	600	
[0501]	ccacgctatc taaaagagtt taagttaaat atctgcccgg agaattttga cagcccgat	660	
[0502]	tatgttacag agaagctgtt tgaagccttt cggagtggaa caattcctat ttatgcaggc	720	
[0503]	ggaggcgcac atccggagcc ggaaattgtg aatcgaagcg cactactcct ttgggagcga	780	

[0504] ggacaaagtg atcatagtg cttggtacag gaagttatac ggctcgcacg cgatgagata 840
 [0505] tactatgata aatttgtaca tcaggttcgt ttgcttccgt atacggaaga gtttgtttat 900
 [0506] gaacagtttt catcgtgaa agagcggttg ttgcagataa gacgagggtg a 951
 [0507] <210> 16
 [0508] <211> 316
 [0509] <212> PRT
 [0510] <213> 损伤新月形单胞菌
 [0511] <400> 16
 [0512] Met Leu Met Arg Ala Leu Arg Lys Met Lys Arg Trp Gly Arg Val Ala
 [0513] 1 5 10 15
 [0514] Phe Asp Tyr Thr Asn Thr Thr Lys Asp Gly Ala Val Cys Tyr His Asn
 [0515] 20 25 30
 [0516] Trp Trp Pro Cys Asn Tyr Glu Glu Glu Trp Phe His Arg Phe Val Val
 [0517] 35 40 45
 [0518] Gln Asn Ile Gly Thr Glu Arg Cys Tyr His Phe Phe Ser Val Phe Gly
 [0519] 50 55 60
 [0520] Pro Arg Ile Ala Leu Thr Leu Pro Thr Pro Asn Lys Val Phe Phe Cys
 [0521] 65 70 75 80
 [0522] Gly Glu Asn Val His Asn Ala Glu Trp Pro Tyr Lys Ser Tyr Gln Asp
 [0523] 85 90 95
 [0524] His Ala Leu Gly Asp Val Lys Leu Ala Leu Gly Tyr Asp Asp Ile Gln
 [0525] 100 105 110
 [0526] Asp Glu Arg Tyr Ile Arg Phe Pro Leu Trp Leu Leu Tyr Met Phe Asp
 [0527] 115 120 125
 [0528] Pro Val Val Asp Arg Tyr Ala Ile Arg Glu Arg Ile Glu Glu Ile Asn
 [0529] 130 135 140
 [0530] His Ala Glu Asn Thr Arg Lys Tyr Glu Cys Val Leu Ile Ser Arg His
 [0531] 145 150 155 160
 [0532] Asp Lys Trp Asn Met Arg Gly Pro Ile Tyr Asp Ala Leu Lys Asp His
 [0533] 165 170 175
 [0534] Leu Ala Ile Ser Cys Ala Gly Lys Trp Lys Gln Asn Thr Asp Glu Leu
 [0535] 180 185 190
 [0536] Trp Thr Val Tyr Asn Asp Asp Lys Pro Arg Tyr Leu Lys Glu Phe Lys
 [0537] 195 200 205
 [0538] Phe Asn Ile Cys Pro Glu Asn Phe Asp Thr Pro Tyr Tyr Val Thr Glu
 [0539] 210 215 220
 [0540] Lys Leu Phe Glu Ala Phe Arg Ser Gly Thr Ile Pro Ile Tyr Ala Gly
 [0541] 225 230 235 240
 [0542] Gly Gly Asp His Pro Glu Pro Glu Ile Val Asn Arg Ser Ala Leu Leu
 [0543] 245 250 255
 [0544] Leu Trp Glu Arg Gly Gln Ser Asp His Ser Ala Leu Val Gln Glu Val
 [0545] 260 265 270

[0546]	Ile Arg Leu Ala Arg Asp Glu Ile Tyr Tyr Asp Lys Phe Val His Gln	
[0547]	275	280 285
[0548]	Val Arg Leu Leu Pro Tyr Thr Glu Glu Phe Val Tyr Glu Gln Phe Ser	
[0549]	290	295 300
[0550]	Ser Leu Lys Glu Arg Leu Leu Gln Ile Arg Arg Gly	
[0551]	305	310 315
[0552]	<210>	17
[0553]	<211>	1437
[0554]	<212>	DNA
[0555]	<213>	幽门螺杆菌 (Helicobacter pylori)
[0556]	<400>	17
[0557]	atgtttcagc cgctgctgga tgcataatgtt gaaagcgcaa gcattgaaaa aatggcaagc	60
[0558]	aaaagtccgc ctccgctgaa aattgcagtt gcaaattggt ggggtgatga agaaatcaaa	120
[0559]	gagtttaaaa acagcgtcct gtactttatt ctgagccagc gttataccat taccctgcat	180
[0560]	cagaatccga atgaattttc cgatctgggtt tttggtaatc cgctgggtag cgcacgtaaa	240
[0561]	attctgagct atcagaatgc aaaacgcgtg ttttataccg gtgaaaatga aagcccgaac	300
[0562]	ttcaacctgt ttgattatgc cattgggttc gatgagctgg attttaatga tcgttatctg	360
[0563]	cgtatgccgc tgtattatga tcgtctgcat cataaagcag aaagcgtaa tgataccacc	420
[0564]	gcaccgtata aactgaaaga taatagcctg tacgcactga aaaaaccgag ccattgcttt	480
[0565]	aaagaaaaac atccgaatct gtgtgccgtg gttaatgatg aaagcgatcc tctgaaacgt	540
[0566]	ggttttgcaa gctttgttgc aagcaatccg aacgcaccga ttcgtaatgc attctatgat	600
[0567]	gcactgaata gcattgaacc ggttaccggt ggtggtagcg ttcgtaatac cctgggttat	660
[0568]	aatgtgaaaa acaaaaacga attcctgagc cagtataaat tcaatctgtg ctttgaaaac	720
[0569]	accaggggtt atggttatgt gaccgaaaaa atcatcgatg cctatttcag ccataccatt	780
[0570]	ccgatttatt ggggtagccc gagcgttgca aaagatttca atccgaaaag ctttgtgaac	840
[0571]	gtgcacgact tcaaaaactt tgatgaagcc atcgattata tcaaatacct gcacacccat	900
[0572]	aaaaacgcct atctggatat gctgtatgaa aatccgctga atacactgga tggtaaagcc	960
[0573]	tatttttacc agaacctgag cttcaaaaaa atcctggcct ttttcaaaac catcctggaa	1020
[0574]	aacgatacca tctatcacga taaccctgtt atcttttgcc gtgatctgaa tgaaccgctg	1080
[0575]	gttaccattg atgatctgcg tgtaattat gatgacctgc gcgtgaacta cgacgatctg	1140
[0576]	cgcatacaatt atgatgatct gcgggtaaac tatgatgatc tgcgtatcaa ctacgacgac	1200
[0577]	ctgcgtgtga actacgatga cctgcgggtt aattatgatg atctgcggat taattatgat	1260
[0578]	gatctgcgtg tgaactatga cgatctgcgt gtgaattacg agcgtctgct gagcaaagca	1320
[0579]	accctctgc tggaactgag ccagaatacc accagtaaaa tctatcgtaa agcgtaccag	1380
[0580]	aaaagcctgc ctctgctgcg tgcaattcgt cgttgggtta aaaaactggg tctgtaa	1437
[0581]	<210>	18
[0582]	<211>	478
[0583]	<212>	PRT
[0584]	<213>	幽门螺杆菌
[0585]	<400>	18
[0586]	Met Phe Gln Pro Leu Leu Asp Ala Tyr Val Glu Ser Ala Ser Ile Glu	
[0587]	1	5 10 15

[0588]	Lys Met Ala Ser Lys Ser Pro Pro Pro Leu Lys Ile Ala Val Ala Asn
[0589]	20 25 30
[0590]	Trp Trp Gly Asp Glu Glu Ile Lys Glu Phe Lys Asn Ser Val Leu Tyr
[0591]	35 40 45
[0592]	Phe Ile Leu Ser Gln Arg Tyr Thr Ile Thr Leu His Gln Asn Pro Asn
[0593]	50 55 60
[0594]	Glu Phe Ser Asp Leu Val Phe Gly Asn Pro Leu Gly Ser Ala Arg Lys
[0595]	65 70 75 80
[0596]	Ile Leu Ser Tyr Gln Asn Ala Lys Arg Val Phe Tyr Thr Gly Glu Asn
[0597]	85 90 95
[0598]	Glu Ser Pro Asn Phe Asn Leu Phe Asp Tyr Ala Ile Gly Phe Asp Glu
[0599]	100 105 110
[0600]	Leu Asp Phe Asn Asp Arg Tyr Leu Arg Met Pro Leu Tyr Tyr Asp Arg
[0601]	115 120 125
[0602]	Leu His His Lys Ala Glu Ser Val Asn Asp Thr Thr Ala Pro Tyr Lys
[0603]	130 135 140
[0604]	Leu Lys Asp Asn Ser Leu Tyr Ala Leu Lys Lys Pro Ser His Cys Phe
[0605]	145 150 155 160
[0606]	Lys Glu Lys His Pro Asn Leu Cys Ala Val Val Asn Asp Glu Ser Asp
[0607]	165 170 175
[0608]	Pro Leu Lys Arg Gly Phe Ala Ser Phe Val Ala Ser Asn Pro Asn Ala
[0609]	180 185 190
[0610]	Pro Ile Arg Asn Ala Phe Tyr Asp Ala Leu Asn Ser Ile Glu Pro Val
[0611]	195 200 205
[0612]	Thr Gly Gly Gly Ser Val Arg Asn Thr Leu Gly Tyr Asn Val Lys Asn
[0613]	210 215 220
[0614]	Lys Asn Glu Phe Leu Ser Gln Tyr Lys Phe Asn Leu Cys Phe Glu Asn
[0615]	225 230 235 240
[0616]	Thr Gln Gly Tyr Gly Tyr Val Thr Glu Lys Ile Ile Asp Ala Tyr Phe
[0617]	245 250 255
[0618]	Ser His Thr Ile Pro Ile Tyr Trp Gly Ser Pro Ser Val Ala Lys Asp
[0619]	260 265 270
[0620]	Phe Asn Pro Lys Ser Phe Val Asn Val His Asp Phe Lys Asn Phe Asp
[0621]	275 280 285
[0622]	Glu Ala Ile Asp Tyr Ile Lys Tyr Leu His Thr His Lys Asn Ala Tyr
[0623]	290 295 300
[0624]	Leu Asp Met Leu Tyr Glu Asn Pro Leu Asn Thr Leu Asp Gly Lys Ala
[0625]	305 310 315 320
[0626]	Tyr Phe Tyr Gln Asn Leu Ser Phe Lys Lys Ile Leu Ala Phe Phe Lys
[0627]	325 330 335
[0628]	Thr Ile Leu Glu Asn Asp Thr Ile Tyr His Asp Asn Pro Phe Ile Phe
[0629]	340 345 350

[0630]	Cys Arg Asp Leu Asn Glu Pro Leu Val Thr Ile Asp Asp Leu Arg Val	
[0631]	355	360 365
[0632]	Asn Tyr Asp Asp Leu Arg Val Asn Tyr Asp Asp Leu Arg Ile Asn Tyr	
[0633]	370	375 380
[0634]	Asp Asp Leu Arg Val Asn Tyr Asp Asp Leu Arg Ile Asn Tyr Asp Asp	
[0635]	385	390 395 400
[0636]	Leu Arg Val Asn Tyr Asp Asp Leu Arg Val Asn Tyr Asp Asp Leu Arg	
[0637]	405	410 415
[0638]	Ile Asn Tyr Asp Asp Leu Arg Val Asn Tyr Asp Asp Leu Arg Val Asn	
[0639]	420	425 430
[0640]	Tyr Glu Arg Leu Leu Ser Lys Ala Thr Pro Leu Leu Glu Leu Ser Gln	
[0641]	435	440 445
[0642]	Asn Thr Thr Ser Lys Ile Tyr Arg Lys Ala Tyr Gln Lys Ser Leu Pro	
[0643]	450	455 460
[0644]	Leu Leu Arg Ala Ile Arg Arg Trp Val Lys Lys Leu Gly Leu	
[0645]	465	470 475
[0646]	<210>	19
[0647]	<211>	978
[0648]	<212>	DNA
[0649]	<213>	固氮螺菌属物种(Azospirillum sp.)
[0650]	<400>	19
[0651]	atgattgatc agcggaccgg cgtgtttctg agcgaatttc ttgatacccg caacagggac	60
[0652]	ccggcagtagc tggatcgctt tctactgcag ggaccggatg gcggaaggcg gggagcgaaa	120
[0653]	ccgaacctga aagtggcggtt ttttgacttt tggccagaat ttgacccag cgcaaacttt	180
[0654]	tttgtggaga ttctgagcgc gcgctttcag gtgagcgtgg tggaaaacga tagcgatctt	240
[0655]	gcgattgtga gcgtgtttgg caccggacca cgggaaatac ggactgcgcg gagcatgttt	300
[0656]	tttaccggag aaaatgtgcg gccgccgctt gatggcattg atatgagcgt gagctttgat	360
[0657]	cgcattgatg acccacggca ttttcgcctg ccgctatatg tggatgcagc gtatgatcat	420
[0658]	ctgcgcgagg gagcagcacc gtatttttgc cagccagtgc tgccgccagt gccgccgact	480
[0659]	cgggaagatg cggcagaacg gaagttttgc gcgtttcttt ataaaaacc aaactgcgcg	540
[0660]	cgccgcaacg atttttttca tatgcttggc gcgcggcgcc atgttgatag cgtgggctgg	600
[0661]	ctgctgaaca acaccggcag cgtggtgaaa atgggatgac taccgaaaat tcgggtgttt	660
[0662]	agccgtatc gctttgcgtt tgcgtttgaa aacgctagcc atccaggcta tctgaccgaa	720
[0663]	aaaattctgg atgcttttca ggccgggagca gtgccgcttt attggggcga cccaggcgtg	780
[0664]	ctgcgcgacg tggcagcggg cagctttatt gatgtgagca ggtatagcag cgatgaagaa	840
[0665]	gcgattgaag cgattctggc gattgatgat gactatggcg cgtatcgccg ctatcgacg	900
[0666]	actccgccct ttcttggcac tgaagacttt cattttgacg cgtatcgact ggcggagtgg	960
[0667]	attgagagcc gactataa	978
[0668]	<210>	20
[0669]	<211>	325
[0670]	<212>	PRT
[0671]	<213>	固氮螺菌属物种

[0672]	<400> 20															
[0673]	Met	Ile	Asp	Gln	Arg	Thr	Gly	Val	Phe	Leu	Ser	Glu	Phe	Leu	Asp	Thr
[0674]	1				5					10					15	
[0675]	Arg	Asn	Arg	Asp	Pro	Ala	Val	Leu	Asp	Arg	Phe	Leu	Leu	Gln	Gly	Pro
[0676]				20					25					30		
[0677]	Asp	Gly	Gly	Arg	Arg	Gly	Ala	Lys	Pro	Asn	Leu	Lys	Val	Ala	Phe	Phe
[0678]			35					40					45			
[0679]	Asp	Phe	Trp	Pro	Glu	Phe	Asp	Pro	Ser	Ala	Asn	Phe	Phe	Val	Glu	Ile
[0680]		50					55					60				
[0681]	Leu	Ser	Ala	Arg	Phe	Gln	Val	Ser	Val	Val	Glu	Asn	Asp	Ser	Asp	Leu
[0682]	65					70					75					80
[0683]	Ala	Ile	Val	Ser	Val	Phe	Gly	Thr	Gly	Pro	Arg	Glu	Ile	Arg	Thr	Ala
[0684]					85					90					95	
[0685]	Arg	Ser	Met	Phe	Phe	Thr	Gly	Glu	Asn	Val	Arg	Pro	Pro	Leu	Asp	Gly
[0686]			100						105					110		
[0687]	Ile	Asp	Met	Ser	Val	Ser	Phe	Asp	Arg	Ile	Asp	Asp	Pro	Arg	His	Phe
[0688]			115					120					125			
[0689]	Arg	Leu	Pro	Leu	Tyr	Val	Val	His	Ala	Tyr	Asp	His	Leu	Arg	Glu	Gly
[0690]		130					135					140				
[0691]	Ala	Ala	Pro	Tyr	Phe	Cys	Gln	Pro	Val	Leu	Pro	Pro	Val	Pro	Pro	Thr
[0692]	145					150					155					160
[0693]	Arg	Glu	Asp	Ala	Ala	Glu	Arg	Lys	Phe	Cys	Ala	Phe	Leu	Tyr	Lys	Asn
[0694]				165						170					175	
[0695]	Pro	Asn	Cys	Ala	Arg	Arg	Asn	Asp	Phe	Phe	His	Met	Leu	Gly	Ala	Arg
[0696]			180						185					190		
[0697]	Arg	His	Val	Asp	Ser	Val	Gly	Trp	Leu	Leu	Asn	Asn	Thr	Gly	Ser	Val
[0698]			195					200					205			
[0699]	Val	Lys	Met	Gly	Trp	Leu	Pro	Lys	Ile	Arg	Val	Phe	Ser	Arg	Tyr	Arg
[0700]		210					215					220				
[0701]	Phe	Ala	Phe	Ala	Phe	Glu	Asn	Ala	Ser	His	Pro	Gly	Tyr	Leu	Thr	Glu
[0702]	225					230					235					240
[0703]	Lys	Ile	Leu	Asp	Ala	Phe	Gln	Ala	Gly	Ala	Val	Pro	Leu	Tyr	Trp	Gly
[0704]				245						250					255	
[0705]	Asp	Pro	Gly	Val	Leu	Arg	Asp	Val	Ala	Ala	Gly	Ser	Phe	Ile	Asp	Val
[0706]			260						265					270		
[0707]	Ser	Arg	Tyr	Ser	Ser	Asp	Glu	Glu	Ala	Ile	Glu	Ala	Ile	Leu	Ala	Ile
[0708]			275					280					285			
[0709]	Asp	Asp	Asp	Tyr	Gly	Ala	Tyr	Arg	Arg	Tyr	Arg	Ser	Thr	Pro	Pro	Phe
[0710]		290					295					300				
[0711]	Leu	Gly	Thr	Glu	Asp	Phe	His	Phe	Asp	Ala	Tyr	Arg	Leu	Ala	Glu	Trp
[0712]	305					310					315					320
[0713]	Ile	Glu	Ser	Arg	Leu											

[0714]	325
[0715]	<210> 21
[0716]	<211> 978
[0717]	<212> DNA
[0718]	<213> 固氮螺菌属物种
[0719]	<400> 21
[0720]	atgattgaca ggcggacaag cgatcttctg gcggagttcc tagcaagcgc taacaaagat 60
[0721]	ccggcagtag ttgatcgatt cctactacat ggaccggacc ggggaggccg cagcgcgaaa 120
[0722]	ccgcggctga aaattgcgtt ttttgacttt tggccggagt ttgaccggc agcaaatttt 180
[0723]	tttttggaat ttctgagcgc gcgctttgat ctgagcgtgg tggataatga tagcgatcta 240
[0724]	gcgattgtga gcgtgtttgg aattcgccat cgggaagctc ggactgcgcg aagcctgttt 300
[0725]	tttaccggcg aaaatgtgcg gccccgctt gatggcgtgg atatgagcgt gagctttgat 360
[0726]	cgcatgtatg acccacggca ttatcggtcg ccgctttatg tgatgcatgc gtgggatcat 420
[0727]	cggcgcgagg gagcaactcg gcatttttgc catagcgtgc tgccgccggt gccccgact 480
[0728]	cgggaagaag cagataggcg gaagttttgc gcgtttcttt ataaaaatcc aaactgcgag 540
[0729]	cggcgcaacg actttttccg gatgctttgc gcgcgccgcc atgtggaaag cgtgggatgg 600
[0730]	ctgctgaaca acaccggcag cgtggtgaaa atgggctggc tgccgaaaat tcgggtgttt 660
[0731]	agccgctatc gctttgcgtt tgcgtttgaa aatgcgagcc atccaggcta tctgaccgaa 720
[0732]	aaaattcttg atgcgtttca ggccggcgct gtgccgcttt attgggggga cccaggcgtg 780
[0733]	ctgcgggacg tagcggcggg cagctttatt gacgtgagcc gctatagcag cgatgaagaa 840
[0734]	gcgattgatg cgattctggc aattgatgac gactatgata cctatcgccg ccatcgacgc 900
[0735]	actgctccat ttcttggcac tgaagacttt tattttgacg cgtttcgact ggcggagtgg 960
[0736]	attgagagcc gactataa 978
[0737]	<210> 22
[0738]	<211> 325
[0739]	<212> PRT
[0740]	<213> 固氮螺菌属物种
[0741]	<400> 22
[0742]	Met Ile Asp Arg Arg Thr Ser Asp Phe Leu Ala Glu Phe Leu Ala Ser
[0743]	1 5 10 15
[0744]	Ala Asn Lys Asp Pro Ala Val Leu Asp Arg Phe Leu Leu His Gly Pro
[0745]	20 25 30
[0746]	Asp Arg Gly Gly Arg Ser Ala Lys Pro Arg Leu Lys Ile Ala Phe Phe
[0747]	35 40 45
[0748]	Asp Phe Trp Pro Glu Phe Asp Pro Ala Ala Asn Phe Phe Val Glu Ile
[0749]	50 55 60
[0750]	Leu Ser Ala Arg Phe Asp Leu Ser Val Val Asp Asn Asp Ser Asp Leu
[0751]	65 70 75 80
[0752]	Ala Ile Val Ser Val Phe Gly Ile Arg His Arg Glu Ala Arg Thr Ala
[0753]	85 90 95
[0754]	Arg Ser Leu Phe Phe Thr Gly Glu Asn Val Arg Pro Pro Leu Asp Gly
[0755]	100 105 110

[0756]	Val Asp Met Ser Val Ser Phe Asp Arg Ile Asp Asp Pro Arg His Tyr	
[0757]	115	120 125
[0758]	Arg Leu Pro Leu Tyr Val Met His Ala Trp Asp His Arg Arg Glu Gly	
[0759]	130	135 140
[0760]	Ala Thr Arg His Phe Cys His Ser Val Leu Pro Pro Val Pro Pro Thr	
[0761]	145	150 155 160
[0762]	Arg Glu Glu Ala Asp Arg Arg Lys Phe Cys Ala Phe Leu Tyr Lys Asn	
[0763]	165	170 175
[0764]	Pro Asn Cys Glu Arg Arg Asn Asp Phe Phe Arg Met Leu Cys Ala Arg	
[0765]	180	185 190
[0766]	Arg His Val Glu Ser Val Gly Trp Leu Leu Asn Asn Thr Gly Ser Val	
[0767]	195	200 205
[0768]	Val Lys Met Gly Trp Leu Pro Lys Ile Arg Val Phe Ser Arg Tyr Arg	
[0769]	210	215 220
[0770]	Phe Ala Phe Ala Phe Glu Asn Ala Ser His Pro Gly Tyr Leu Thr Glu	
[0771]	225	230 235 240
[0772]	Lys Ile Leu Asp Ala Phe Gln Ala Gly Ala Val Pro Leu Tyr Trp Gly	
[0773]	245	250 255
[0774]	Asp Pro Gly Val Leu Arg Asp Val Ala Ala Gly Ser Phe Ile Asp Val	
[0775]	260	265 270
[0776]	Ser Arg Tyr Ser Ser Asp Glu Glu Ala Ile Asp Ala Ile Leu Ala Ile	
[0777]	275	280 285
[0778]	Asp Asp Asp Tyr Asp Thr Tyr Arg Arg His Arg Ser Thr Ala Pro Phe	
[0779]	290	295 300
[0780]	Leu Gly Thr Glu Asp Phe Tyr Phe Asp Ala Phe Arg Leu Ala Glu Trp	
[0781]	305	310 315 320
[0782]	Ile Glu Ser Arg Leu	
[0783]	325	
[0784]	<210> 23	
[0785]	<211> 978	
[0786]	<212> DNA	
[0787]	<213> Azospirillum brasilense	
[0788]	<400> 23	
[0789]	atgcttgacc agcggacaag cgcattccta gaagagtttc tggcaaaacc aggcggcgat	60
[0790]	ccggaacggc ttgatcgctt tctgctgcat ggcccatatc gcggacggcg cggcggcagg	120
[0791]	ccacggctga aactggcggt ttatgatttt tggccggaat ttgatactgg caggaaacttt	180
[0792]	tttattgaaa ttctgagcag ccgcttcgac ctgagcgtgg tggaagatga tagcgacctt	240
[0793]	gcgattgtga gcgtgtttgg cggacggcat cgcgcagcac gcagccggcg caccctgttt	300
[0794]	tttaccggag aaaatgtgcg cccccgctg gatggctttg atatggcagt gagctttgat	360
[0795]	cgcgtgggcg atccgcgcca ttatcgctg ccgctttatg tgatgcatgc gtatgaacat	420
[0796]	atgcgggaag gagcagtgcc gcatTTTTgc agcccagtgc tgccgccggt gccgccaagc	480
[0797]	cgggcagcgt ttgcagaacg caacttttgc gcgtttcttt ataaaaaccc gaacggagaa	540

[0798]	cgccgcaacc gcttttttcc ggcacttgat gcacggcggc gcgtggacag cgtgggctgg	600
[0799]	catcttaaca acaccggcag cgtggtgaaa atgggctggc tggcaaaaat tcgcgtgttt	660
[0800]	gagcgctatc gctttgcgtt tgcgtttgaa aacgcgagcc atccaggcta tctgactgag	720
[0801]	aaaattcttg atgtgtttca ggcgggagca gtgccgcttt attgggggga cccagacgtg	780
[0802]	gaacgggaag tggcagcagg cagctttatt gatgtgagcc gctttgcgac tgatgaagaa	840
[0803]	gcagcagaac atattctggc actggatgga gactatgatg cgtattgcgc gtatcgcgcg	900
[0804]	gtggcaccat ttctgggaac tgaagaattt cattttgatg cgtatcgctt tgcggattgg	960
[0805]	attgaaagcc ggctgtag	978
[0806]	<210>	24
[0807]	<211>	325
[0808]	<212>	PRT
[0809]	<213>	Azospirillum brasilense
[0810]	<400>	24
[0811]	Met Leu Asp Gln Arg Thr Ser Ala Phe Leu Glu Glu Phe Leu Ala Lys	
[0812]	1 5 10 15	
[0813]	Pro Gly Gly Asp Pro Glu Arg Leu Asp Arg Phe Leu Leu His Gly Pro	
[0814]	20 25 30	
[0815]	Tyr Arg Gly Arg Arg Gly Gly Arg Pro Arg Leu Lys Leu Ala Phe Tyr	
[0816]	35 40 45	
[0817]	Asp Phe Trp Pro Glu Phe Asp Thr Gly Arg Asn Phe Phe Ile Glu Ile	
[0818]	50 55 60	
[0819]	Leu Ser Ser Arg Phe Asp Leu Ser Val Val Glu Asp Asp Ser Asp Leu	
[0820]	65 70 75 80	
[0821]	Ala Ile Val Ser Val Phe Gly Gly Arg His Arg Ala Ala Arg Ser Arg	
[0822]	85 90 95	
[0823]	Arg Thr Leu Phe Phe Thr Gly Glu Asn Val Arg Pro Pro Leu Asp Gly	
[0824]	100 105 110	
[0825]	Phe Asp Met Ala Val Ser Phe Asp Arg Val Gly Asp Pro Arg His Tyr	
[0826]	115 120 125	
[0827]	Arg Leu Pro Leu Tyr Val Met His Ala Tyr Glu His Met Arg Glu Gly	
[0828]	130 135 140	
[0829]	Ala Val Pro His Phe Cys Ser Pro Val Leu Pro Pro Val Pro Pro Ser	
[0830]	145 150 155 160	
[0831]	Arg Ala Ala Phe Ala Glu Arg Asn Phe Cys Ala Phe Leu Tyr Lys Asn	
[0832]	165 170 175	
[0833]	Pro Asn Gly Glu Arg Arg Asn Arg Phe Phe Pro Ala Leu Asp Ala Arg	
[0834]	180 185 190	
[0835]	Arg Arg Val Asp Ser Val Gly Trp His Leu Asn Asn Thr Gly Ser Val	
[0836]	195 200 205	
[0837]	Val Lys Met Gly Trp Leu Ala Lys Ile Arg Val Phe Glu Arg Tyr Arg	
[0838]	210 215 220	
[0839]	Phe Ala Phe Ala Phe Glu Asn Ala Ser His Pro Gly Tyr Leu Thr Glu	

[0840]	225	230	235	240
[0841]	Lys Ile Leu Asp Val Phe Gln Ala Gly Ala Val Pro Leu Tyr Trp Gly			
[0842]		245	250	255
[0843]	Asp Pro Asp Val Glu Arg Glu Val Ala Ala Gly Ser Phe Ile Asp Val			
[0844]		260	265	270
[0845]	Ser Arg Phe Ala Thr Asp Glu Glu Ala Ala Glu His Ile Leu Ala Leu			
[0846]		275	280	285
[0847]	Asp Gly Asp Tyr Asp Ala Tyr Cys Ala Tyr Arg Ala Val Ala Pro Phe			
[0848]		290	295	300
[0849]	Leu Gly Thr Glu Glu Phe His Phe Asp Ala Tyr Arg Leu Ala Asp Trp			
[0850]	305	310	315	320
[0851]	Ile Glu Ser Arg Leu			
[0852]		325		
[0853]	<210> 25			
[0854]	<211> 912			
[0855]	<212> DNA			
[0856]	<213> 固氮螺菌属物种			
[0857]	<400> 25			
[0858]	atgttagatc ggtttctgct tcatgggccg gagcgcgggg gccgtgcggc cagaccgcgc	60		
[0859]	ctgaaaattg cgttttttga ctttttgccg gaatttgacc cgagcgcgaa tttttttgta	120		
[0860]	gaaattctga gcagccgctt tgatgtgagc gtggttgata atgatagcga tttagcgatt	180		
[0861]	ctgagcgtgt ttggcgaacg ccaccgcgaa gcgcgtaccg cgcgcgcgct gttttttacc	240		
[0862]	ggcgaaaatg tgcgcccgcc gttagatggc gtggatatga gcgtgagctt tgatcgcat	300		
[0863]	gatcatccac gtcattatcg cctgccgtta tacgtgatgc atgcgtggga tcaccgtcgc	360		
[0864]	gaaggggcga ccccgcatit ttgccatccg gtgctgccgc cggtgccgcc gacgcgtgaa	420		
[0865]	gaagcggcga aacgtaagtt ttgcgcgttt ttatataaaa atcctcactg cgcgcgccgc	480		
[0866]	aacgattttt ttcagatgct gtgcgcgcgg cgccatgtgg aaagcgtggg ctggctgctg	540		
[0867]	aataataaccg gcagcgtggt gaaaatgggc tggctgccga aaattcgctg gtttgcgcg	600		
[0868]	tatcgctttg cgttttgcgt tgaaaacgcg gcgcatccag gctatctgac cgagaaaatt	660		
[0869]	ctggatgcgt ttcaggccgg gacggtaccg ttatactggg gcgacagcgg cgtgctgcgc	720		
[0870]	gacgttgccg ccggcagctt tattgatgtg agccgctatg cgagcgatga agaagcgatt	780		
[0871]	gaagcgattc tggcgattga tgatgactat gatagctatc gccggtaccg cggcacggcg	840		
[0872]	ccatttttag gcaccgagga cttttacttt gacgcgtacc ggctggccga gtggattgag	900		
[0873]	agccgcctgt ag			912
[0874]	<210> 26			
[0875]	<211> 303			
[0876]	<212> PRT			
[0877]	<213> 固氮螺菌属物种			
[0878]	<400> 26			
[0879]	Met Leu Asp Arg Phe Leu Leu His Gly Pro Glu Arg Gly Gly Arg Ala			
[0880]	1	5	10	15
[0881]	Ala Arg Pro Arg Leu Lys Ile Ala Phe Phe Asp Phe Trp Pro Glu Phe			

[0882]	20	25	30
[0883]	Asp Pro Ser Ala Asn Phe Phe Val Glu Ile Leu Ser Ser Arg Phe Asp		
[0884]	35	40	45
[0885]	Val Ser Val Val Asp Asn Asp Ser Asp Leu Ala Ile Leu Ser Val Phe		
[0886]	50	55	60
[0887]	Gly Glu Arg His Arg Glu Ala Arg Thr Ala Arg Ala Leu Phe Phe Thr		
[0888]	65	70	75
[0889]	Gly Glu Asn Val Arg Pro Pro Leu Asp Gly Val Asp Met Ser Val Ser		
[0890]	85	90	95
[0891]	Phe Asp Arg Ile Asp His Pro Arg His Tyr Arg Leu Pro Leu Tyr Val		
[0892]	100	105	110
[0893]	Met His Ala Trp Asp His Arg Arg Glu Gly Ala Thr Pro His Phe Cys		
[0894]	115	120	125
[0895]	His Pro Val Leu Pro Pro Val Pro Pro Thr Arg Glu Glu Ala Ala Lys		
[0896]	130	135	140
[0897]	Arg Lys Phe Cys Ala Phe Leu Tyr Lys Asn Pro His Cys Ala Arg Arg		
[0898]	145	150	155
[0899]	Asn Asp Phe Phe Gln Met Leu Cys Ala Arg Arg His Val Glu Ser Val		
[0900]	165	170	175
[0901]	Gly Trp Leu Leu Asn Asn Thr Gly Ser Val Val Lys Met Gly Trp Leu		
[0902]	180	185	190
[0903]	Pro Lys Ile Arg Val Phe Ala Arg Tyr Arg Phe Ala Phe Ala Phe Glu		
[0904]	195	200	205
[0905]	Asn Ala Ala His Pro Gly Tyr Leu Thr Glu Lys Ile Leu Asp Ala Phe		
[0906]	210	215	220
[0907]	Gln Ala Gly Thr Val Pro Leu Tyr Trp Gly Asp Ser Gly Val Leu Arg		
[0908]	225	230	235
[0909]	Asp Val Ala Ala Gly Ser Phe Ile Asp Val Ser Arg Tyr Ala Ser Asp		
[0910]	245	250	255
[0911]	Glu Glu Ala Ile Glu Ala Ile Leu Ala Ile Asp Asp Asp Tyr Asp Ser		
[0912]	260	265	270
[0913]	Tyr Arg Arg Tyr Arg Gly Thr Ala Pro Phe Leu Gly Thr Glu Asp Phe		
[0914]	275	280	285
[0915]	Tyr Phe Asp Ala Tyr Arg Leu Ala Glu Trp Ile Glu Ser Arg Leu		
[0916]	290	295	300
[0917]	<210> 27		
[0918]	<211> 1032		
[0919]	<212> DNA		
[0920]	<213> 丁酸弧菌属物种(Butyrivibrio sp.)		
[0921]	<400> 27		
[0922]	atgaatataa ttcactttta tgcaagatat ttaagagaat cacataactg gaacagagaa	60	
[0923]	cgtgaagtta ctcgtaacgg tgttatgact tttgctaatt ggtggagaga agaccgcac	120	

[0924]	aagaactggt ttgcaagatt tattgacgct ggaagcaaag accctgaacg caggattagg	180
[0925]	ttctatagta tttttggacc atatatgaaa ttaaaagaag attttgatgg agctaagata	240
[0926]	ttcttttccg gagagaatct tgagcagccg gtctatcaca gaatattaaa gacagatcct	300
[0927]	atagaagata gaatatgggc tgacagaagg aagctgtatg gtaattatgg agcaggagat	360
[0928]	gtggatcttg ctataggatt tggcaatagg gaagaggatt cacttatggg atttgaagga	420
[0929]	agcaggaaga ctaaatatat ccgctttcct ttatggctta catatgtttt tgatcctgat	480
[0930]	tgtactcatg atgatattaa gagaaccatt gatgagataa atgcagttcg ttccacaggc	540
[0931]	aggaaggata ctctgcttct tgcacgcat gatttctggg ggacaaggtc agatatctta	600
[0932]	aagagtctag aagggtgatg cgatgttagt attgccggta aatggcgcaa caacaccaaa	660
[0933]	gaactctggg aagattataa caatgacaag aataaatac tgtcagaatt taaatttaat	720
[0934]	atatgccctg aaaatgttga tgcaccggga tatgtgacag agaagatatt tgatgctttt	780
[0935]	aaatgcgag ctattcctat atacagggc tgtcttgga agcctgagcc ggatgtgata	840
[0936]	aatacagatg cagttctttt atgggacttc gatggagata attcagatac tatatccttg	900
[0937]	attaaaaaac taaattcgga taatgtatac tatgataact ttgtatctca gcccaaattc	960
[0938]	aaaccggatg cggcagagta tgtggttga tgtatggatg agctgaggcg aagctttgat	1020
[0939]	cagtcacatc ga	1032
[0940]	<210>	28
[0941]	<211>	343
[0942]	<212>	PRT
[0943]	<213>	丁酸弧菌属物种
[0944]	<400>	28
[0945]	Met Asn Ile Ile His Phe Tyr Ala Arg Tyr Leu Arg Glu Ser His Asn	
[0946]	1 5 10 15	
[0947]	Trp Asn Arg Glu Arg Glu Val Thr Arg Asn Gly Val Met Thr Phe Ala	
[0948]	20 25 30	
[0949]	Asn Trp Trp Arg Glu Asp Pro His Lys Asn Trp Phe Ala Arg Phe Ile	
[0950]	35 40 45	
[0951]	Asp Ala Gly Ser Lys Asp Pro Glu Arg Arg Ile Arg Phe Tyr Ser Ile	
[0952]	50 55 60	
[0953]	Phe Gly Pro Tyr Ser Lys Leu Lys Glu Asp Phe Asp Gly Ala Lys Ile	
[0954]	65 70 75 80	
[0955]	Phe Phe Ser Gly Glu Asn Leu Glu Gln Pro Val Tyr His Arg Ile Leu	
[0956]	85 90 95	
[0957]	Lys Thr Asp Pro Ile Glu Asp Arg Ile Trp Ala Asp Arg Arg Lys Leu	
[0958]	100 105 110	
[0959]	Tyr Gly Asn Tyr Gly Ala Gly Asp Val Asp Leu Ala Ile Gly Phe Gly	
[0960]	115 120 125	
[0961]	Asn Arg Glu Glu Asp Ser Leu Met Gly Phe Glu Gly Ser Arg Lys Thr	
[0962]	130 135 140	
[0963]	Lys Tyr Ile Arg Phe Pro Leu Trp Leu Thr Tyr Val Phe Asp Pro Asp	
[0964]	145 150 155 160	
[0965]	Cys Thr His Asp Asp Ile Lys Arg Thr Ile Asp Glu Ile Asn Ala Val	

[0966]	165	170	175
[0967]	Arg Ser Thr Gly Arg Lys Asp Thr Leu Leu Leu Ala Ser His Asp Phe		
[0968]	180	185	190
[0969]	Trp Gly Thr Arg Ser Asp Ile Leu Lys Ser Leu Glu Gly Val Cys Asp		
[0970]	195	200	205
[0971]	Val Ser Ile Ala Gly Lys Trp Arg Asn Asn Thr Lys Glu Leu Trp Glu		
[0972]	210	215	220
[0973]	Asp Tyr Asn Asn Asp Lys Asn Lys Tyr Leu Ser Glu Phe Lys Phe Asn		
[0974]	225	230	235
[0975]	Ile Cys Pro Glu Asn Val Asp Ala Pro Gly Tyr Val Thr Glu Lys Ile		
[0976]	245	250	255
[0977]	Phe Asp Ala Phe Lys Cys Gly Ala Ile Pro Ile Tyr Gln Gly Cys Leu		
[0978]	260	265	270
[0979]	Gly Lys Pro Glu Pro Asp Val Ile Asn Thr Asp Ala Val Leu Leu Trp		
[0980]	275	280	285
[0981]	Asp Phe Asp Gly Asp Asn Ser Asp Thr Ile Ser Leu Ile Lys Lys Leu		
[0982]	290	295	300
[0983]	Asn Ser Asp Asn Val Tyr Tyr Asp Asn Phe Val Ser Gln Pro Lys Phe		
[0984]	305	310	315
[0985]	Lys Pro Asp Ala Ala Glu Tyr Val Val Ala Cys Met Asp Glu Leu Arg		
[0986]	325	330	335
[0987]	Arg Ser Phe Asp Gln Leu Ile		
[0988]	340		
[0989]	<210> 29		
[0990]	<211> 1059		
[0991]	<212> DNA		
[0992]	<213> 卡托氏卟啉单胞菌		
[0993]	<400> 29		
[0994]	atgtatgccc catacaaaag ccctatcttt gtgcccatac acgacactaa ggcaatgaat	60	
[0995]	ccccccacca aacaaccact tagagagagg ctccacatga tgcgtaggcg caatcgtatc	120	
[0996]	cgaaaacgct ctgtgatagc tctcatcaaa tctcaccttg acagctcagc ctaccaggac	180	
[0997]	tacaactgggt gggacagtca cgcctcgacc ttttggtgc caggttcat tgacctacac	240	
[0998]	ctcgagccca agaaggaggat caatctcttc tcttgettcc aaaatcccct aatgctcatc	300	
[0999]	cgctactaca aaggggtgaa gatcttccta tcaggtgaga accttgccaa taacgagcac	360	
[1000]	tttggettcc atccccgcac gctcgatcat aggatcaacg aggtggactt agccctaggc	420	
[1001]	tttgagttcc gcaaggatcc caagtactat cgcttcccc tttggatcta tcagaatgag	480	
[1002]	ttcatcagcc ccagtgctag cctagaggat atacgtgcgc tccttgagca gatcaacgat	540	
[1003]	ccctccaccc gtcgtagcac gggacgcagt cgcttcatcg ggcagatctc cagccacgac	600	
[1004]	aaaggcgga tgcgaggacg gctcattgac ctctgaatc ccacggaca aatcgactgc	660	
[1005]	gcagggaagt tccgtcaca caccgatgag ctctcgagg tctacgggga tgacaagttt	720	
[1006]	aagtacctag ctaactaccg cttcaacctc tgcccagaga attcactagg cgaaggctac	780	
[1007]	atcaccgaga aggtcttcga cagcatacgc gcaggtgtga tccccatcta ttgggggtgc	840	

[1008] tacctagagc ctggcatcct taacccaag gctatcctac gcttcgagga agggaaggaa 900
 [1009] caagagttct ataaccgagt gaaggagctg tgggagaacg aagcggccta cgagcagttt 960
 [1010] atcctcgagc ccccttcgt agaggagca gcagagcgca tctgggaaat cctccagggg 1020
 [1011] cttcgtgagc gccttgcccc tcttgaggag gaagataa 1059
 [1012] <210> 30
 [1013] <211> 352
 [1014] <212> PRT
 [1015] <213> 卡托氏卟啉单胞菌
 [1016] <400> 30
 [1017] Met Leu Ala Pro Tyr Lys Ser Pro Ile Phe Val Pro Ile Tyr Asp Thr
 [1018] 1 5 10 15
 [1019] Lys Ala Met Asn Pro Pro Thr Lys Gln Pro Leu Arg Glu Arg Leu His
 [1020] 20 25 30
 [1021] Met Met Arg Arg Arg Asn Arg Ile Arg Lys Arg Ser Val Ile Ala Leu
 [1022] 35 40 45
 [1023] Ile Lys Ser His Leu Asp Ser Ser Arg Tyr Gln Asp Tyr Asn Trp Trp
 [1024] 50 55 60
 [1025] Asp Ser His Ala Ser Thr Phe Trp Leu Pro Arg Phe Ile Asp Leu His
 [1026] 65 70 75 80
 [1027] Leu Glu Pro Lys Lys Arg Ile Asn Leu Phe Ser Cys Phe Gln Asn Pro
 [1028] 85 90 95
 [1029] Leu Met Leu Ile Arg Tyr Tyr Lys Gly Val Lys Ile Phe Leu Ser Gly
 [1030] 100 105 110
 [1031] Glu Asn Leu Ala Asn Asn Glu His Phe Gly Phe His Pro Arg Met Leu
 [1032] 115 120 125
 [1033] Asp His Arg Ile Asn Glu Val Asp Leu Ala Leu Gly Phe Glu Phe Arg
 [1034] 130 135 140
 [1035] Lys Asp Pro Lys Tyr Tyr Arg Phe Pro Leu Trp Ile Tyr Gln Asn Glu
 [1036] 145 150 155 160
 [1037] Phe Ile Ser Pro Ser Ala Ser Leu Glu Asp Ile Arg Ala Leu Leu Glu
 [1038] 165 170 175
 [1039] Gln Ile Asn Asp Pro Ser Thr Arg Arg Ser Thr Gly Arg Ser Arg Phe
 [1040] 180 185 190
 [1041] Ile Gly Gln Ile Ser Ser His Asp Lys Gly Gly Met Arg Gly Arg Leu
 [1042] 195 200 205
 [1043] Ile Asp Leu Leu Asn Pro Ile Gly Gln Ile Asp Cys Ala Gly Lys Phe
 [1044] 210 215 220
 [1045] Arg His Asn Thr Asp Glu Leu Leu Glu Val Tyr Gly Asp Asp Lys Phe
 [1046] 225 230 235 240
 [1047] Lys Tyr Leu Ala Asn Tyr Arg Phe Asn Leu Cys Pro Glu Asn Ser Leu
 [1048] 245 250 255
 [1049] Gly Glu Gly Tyr Ile Thr Glu Lys Val Phe Asp Ser Ile Arg Ala Gly

[1050]	260	265	270
[1051]	Cys Ile Pro Ile Tyr Trp Gly Ala Tyr Leu Glu Pro Gly Ile Leu Asn		
[1052]	275	280	285
[1053]	Pro Lys Ala Ile Leu Arg Phe Glu Glu Gly Lys Glu Gln Glu Phe Tyr		
[1054]	290	295	300
[1055]	Asn Arg Val Lys Glu Leu Trp Glu Asn Glu Ala Ala Tyr Glu Gln Phe		
[1056]	305	310	315
[1057]	Ile Leu Glu Pro Pro Phe Val Glu Gly Ala Ala Glu Arg Ile Trp Glu		
[1058]	325	330	335
[1059]	Ile Leu Gln Gly Leu Arg Glu Arg Leu Ala Pro Leu Val Glu Glu Gly		
[1060]	340	345	350
[1061]	<210> 31		
[1062]	<211> 1032		
[1063]	<212> DNA		
[1064]	<213> 溶纤维丁酸弧菌(<i>Butyrivibrio fibrisolvens</i>)		
[1065]	<400> 31		
[1066]	atgaacataa ttcactttta tgcaagatat ttaagagaat cacataattg gaacagagaa	60	
[1067]	cgtgaagtta ctcgtaatgg tgttatgact tttgctaact ggtggagaga agatccgcat	120	
[1068]	aagaactggg ttgcaagatt tattgatgct gggaataaag accctgagcg caggatcaga	180	
[1069]	ttctatagta tttttggacc ttatagtaaa ttgaaggaag attttgatgg agccaagata	240	
[1070]	ttcttttcg gagaaaatct tgaacagccg gttttacaca gaatactaaa gacagatcct	300	
[1071]	atagaagaca ggatatgggc tgacagaaga aagctgtatg gtaattatgg agctggagaa	360	
[1072]	gtggatcttg ctataggttt tggtaataga gaagaggatt cacttctggg atttgaaggg	420	
[1073]	agcaggaaga caaaatatat ccgctttcct ttatggctta catatgtctt tgatcctgac	480	
[1074]	tgtactcatg atgatattaa gagaaccata gatgagataa atgcagttcg ttctacaggc	540	
[1075]	aggaaggata ccctccttct tgcacgcgat gatttctggg ggacaaggtc agatatctta	600	
[1076]	aagagtcttg aaggtgtatg tgatattagt attgccggtg aatggcgcaa taacaccaa	660	
[1077]	gagctctggg aagattatga caatgacaag aataaatatc tgtcagaatt taaatttaac	720	
[1078]	atatgccctg aaaatgttga tgcaccggga tatgtaacag agaagatatt tgatgctttt	780	
[1079]	aaatgccggag ctattcctat atatcagggc tgcctaggta agcctgagcc gaatgtgata	840	
[1080]	aatacagatg cagtacttct atgggacttc gatggagata attcagatac tatagccttg	900	
[1081]	attaaaaaac taaattcgga taatgtatac tatgataact ttgtatctca gcctaaattc	960	
[1082]	aaaccggatg cggcagagta tgtagttgca tgtatggatg agctaaggcg tagctttgac	1020	
[1083]	aggctgatct ga	1032	
[1084]	<210> 32		
[1085]	<211> 343		
[1086]	<212> PRT		
[1087]	<213> 溶纤维丁酸弧菌		
[1088]	<400> 32		
[1089]	Met Asn Ile Ile His Phe Tyr Ala Arg Tyr Leu Arg Glu Ser His Asn		
[1090]	1	5	10
[1091]	Trp Asn Arg Glu Arg Glu Val Thr Arg Asn Gly Val Met Thr Phe Ala		

[1092]	20	25	30
[1093]	Asn Trp Trp Arg Glu Asp Pro His Lys Asn Trp Phe Ala Arg Phe Ile		
[1094]	35	40	45
[1095]	Asp Ala Gly Asn Lys Asp Pro Glu Arg Arg Ile Arg Phe Tyr Ser Ile		
[1096]	50	55	60
[1097]	Phe Gly Pro Tyr Ser Lys Leu Lys Glu Asp Phe Asp Gly Ala Lys Ile		
[1098]	65	70	75
[1099]	Phe Phe Ser Gly Glu Asn Leu Glu Gln Pro Val Leu His Arg Ile Leu		
[1100]	85	90	95
[1101]	Lys Thr Asp Pro Ile Glu Asp Arg Ile Trp Ala Asp Arg Arg Lys Leu		
[1102]	100	105	110
[1103]	Tyr Gly Asn Tyr Gly Ala Gly Glu Val Asp Leu Ala Ile Gly Phe Gly		
[1104]	115	120	125
[1105]	Asn Arg Glu Glu Asp Ser Leu Leu Gly Phe Glu Gly Ser Arg Lys Thr		
[1106]	130	135	140
[1107]	Lys Tyr Ile Arg Phe Pro Leu Trp Leu Thr Tyr Val Phe Asp Pro Asp		
[1108]	145	150	155
[1109]	Cys Thr His Asp Asp Ile Lys Arg Thr Ile Asp Glu Ile Asn Ala Val		
[1110]	165	170	175
[1111]	Arg Ser Thr Gly Arg Lys Asp Thr Leu Leu Leu Ala Ser His Asp Phe		
[1112]	180	185	190
[1113]	Trp Gly Thr Arg Ser Asp Ile Leu Lys Ser Leu Glu Gly Val Cys Asp		
[1114]	195	200	205
[1115]	Ile Ser Ile Ala Gly Lys Trp Arg Asn Asn Thr Lys Glu Leu Trp Glu		
[1116]	210	215	220
[1117]	Asp Tyr Asp Asn Asp Lys Asn Lys Tyr Leu Ser Glu Phe Lys Phe Asn		
[1118]	225	230	235
[1119]	Ile Cys Pro Glu Asn Val Asp Ala Pro Gly Tyr Val Thr Glu Lys Ile		
[1120]	245	250	255
[1121]	Phe Asp Ala Phe Lys Cys Gly Ala Ile Pro Ile Tyr Gln Gly Cys Leu		
[1122]	260	265	270
[1123]	Gly Lys Pro Glu Pro Asn Val Ile Asn Thr Asp Ala Val Leu Leu Trp		
[1124]	275	280	285
[1125]	Asp Phe Asp Gly Asp Asn Ser Asp Thr Ile Ala Leu Ile Lys Lys Leu		
[1126]	290	295	300
[1127]	Asn Ser Asp Asn Val Tyr Tyr Asp Asn Phe Val Ser Gln Pro Lys Phe		
[1128]	305	310	315
[1129]	Lys Pro Asp Ala Ala Glu Tyr Val Val Ala Cys Met Asp Glu Leu Arg		
[1130]	325	330	335
[1131]	Arg Ser Phe Asp Arg Leu Ile		
[1132]	340		
[1133]	<210> 33		

[1134]	<211>	5
[1135]	<212>	PRT
[1136]	<213>	人工序列
[1137]	<220>	
[1138]	<223>	保守的GDP-岩藻糖结合结构域
[1139]	<220>	
[1140]	<221>	变体
[1141]	<222>	(1) .. (1)
[1142]	<223>	Xaa可以是Tyr, Trp, Leu, His, Phe或Met
[1143]	<220>	
[1144]	<221>	不确定
[1145]	<222>	(2) .. (2)
[1146]	<223>	Xaa可以是任意天然存在的氨基酸
[1147]	<220>	
[1148]	<221>	变体
[1149]	<222>	(3) .. (3)
[1150]	<223>	Xaa可以是Thr, Ser或Cys
[1151]	<220>	
[1152]	<221>	变体
[1153]	<222>	(4) .. (4)
[1154]	<223>	Xaa可以是Glu, Gln, Asp或Asn
[1155]	<220>	
[1156]	<221>	变体
[1157]	<222>	(5) .. (5)
[1158]	<223>	Xaa可以是Lys或Arg
[1159]	<400>	33
[1160]	Xaa Xaa Xaa Xaa Xaa	
[1161]	1	5
[1162]	<210>	34
[1163]	<211>	6
[1164]	<212>	PRT
[1165]	<213>	人工序列
[1166]	<220>	
[1167]	<223>	保守结构域
[1168]	<220>	
[1169]	<221>	变体
[1170]	<222>	(1) .. (1)
[1171]	<223>	Xaa可以是Lys或Asp
[1172]	<220>	
[1173]	<221>	变体
[1174]	<222>	(2) .. (2)
[1175]	<223>	Xaa可以是Leu, Lys或Met

[1176]	<220>
[1177]	<221> 不确定
[1178]	<222> (3) .. (5)
[1179]	<223> Xaa可以是任意天然存在的氨基酸
[1180]	<220>
[1181]	<221> 变体
[1182]	<222> (6) .. (6)
[1183]	<223> Xaa可以是Phe或Tyr
[1184]	<400> 34
[1185]	Xaa Xaa Xaa Xaa Xaa Xaa
[1186]	1 5
[1187]	<210> 35
[1188]	<211> 8
[1189]	<212> PRT
[1190]	<213> 人工序列
[1191]	<220>
[1192]	<223> 保守的基序
[1193]	<220>
[1194]	<221> 变体
[1195]	<222> (1) .. (1)
[1196]	<223> Xaa可以是Asn或His
[1197]	<220>
[1198]	<221> 不确定
[1199]	<222> (2) .. (2)
[1200]	<223> Xaa可以是任意天然存在的氨基酸
[1201]	<220>
[1202]	<221> 不确定
[1203]	<222> (6) .. (6)
[1204]	<223> Xaa可以是任意天然存在的氨基酸
[1205]	<400> 35
[1206]	Xaa Xaa Asp Pro Ala Xaa Leu Asp
[1207]	1 5
[1208]	<210> 36
[1209]	<211> 6
[1210]	<212> PRT
[1211]	<213> 人工序列
[1212]	<220>
[1213]	<223> 保守的基序
[1214]	<220>
[1215]	<221> 变体
[1216]	<222> (3) .. (3)
[1217]	<223> Xaa可以是Ala或Ser

[1218]	<400>	36
[1219]	Asp Met Xaa Val Ser Phe	
[1220]	1	5
[1221]	<210>	37
[1222]	<211>	5
[1223]	<212>	PRT
[1224]	<213>	人工序列
[1225]	<220>	
[1226]	<223>	保守的GDP-岩藻糖结合结构域
[1227]	<220>	
[1228]	<221>	不确定
[1229]	<222>	(2) .. (2)
[1230]	<223>	Xaa可以是任意天然存在的氨基酸
[1231]	<400>	37
[1232]	Tyr Xaa Thr Glu Lys	
[1233]	1	5
[1234]	<210>	38
[1235]	<211>	6
[1236]	<212>	PRT
[1237]	<213>	人工序列
[1238]	<220>	
[1239]	<223>	保守结构域
[1240]	<220>	
[1241]	<221>	变体
[1242]	<222>	(1) .. (1)
[1243]	<223>	Xaa可以是Lys或Asp
[1244]	<220>	
[1245]	<221>	不确定
[1246]	<222>	(3) .. (3)
[1247]	<223>	Xaa可以是任意天然存在的氨基酸
[1248]	<220>	
[1249]	<221>	变体
[1250]	<222>	(4) .. (4)
[1251]	<223>	Xaa可以是Ile, Leu或Met
[1252]	<220>	
[1253]	<221>	变体
[1254]	<222>	(6) .. (6)
[1255]	<223>	Xaa可以是Phe或Tyr
[1256]	<400>	38
[1257]	Xaa Leu Xaa Xaa Gly Xaa	
[1258]	1	5
[1259]	<210>	39

[1260]	<211>	6
[1261]	<212>	PRT
[1262]	<213>	人工序列
[1263]	<220>	
[1264]	<223>	保守结构域
[1265]	<220>	
[1266]	<221>	变体
[1267]	<222>	(1) .. (1)
[1268]	<223>	Xaa可以是Lys或Asp
[1269]	<220>	
[1270]	<221>	变体
[1271]	<222>	(2) .. (2)
[1272]	<223>	Xaa可以是Leu或Lys
[1273]	<220>	
[1274]	<221>	不确定
[1275]	<222>	(3) .. (3)
[1276]	<223>	Xaa可以是任意天然存在的氨基酸
[1277]	<220>	
[1278]	<221>	变体
[1279]	<222>	(5) .. (5)
[1280]	<223>	Xaa可以是Ser或Gly
[1281]	<220>	
[1282]	<221>	变体
[1283]	<222>	(6) .. (6)
[1284]	<223>	Xaa可以是Phe或Tyr
[1285]	<400>	39
[1286]	Xaa Xaa Xaa Leu Xaa Xaa	
[1287]	1	5
[1288]	<210>	40
[1289]	<211>	6
[1290]	<212>	PRT
[1291]	<213>	人工序列
[1292]	<220>	
[1293]	<223>	保守结构域
[1294]	<220>	
[1295]	<221>	变体
[1296]	<222>	(1) .. (1)
[1297]	<223>	Xaa可以是Lys或Asp
[1298]	<220>	
[1299]	<221>	不确定
[1300]	<222>	(3) .. (3)
[1301]	<223>	Xaa可以是任意天然存在的氨基酸

[1302]	<220>
[1303]	<221> 不确定
[1304]	<222> (6) .. (6)
[1305]	<223> Xaa可以是Phe或Tyr
[1306]	<400> 40
[1307]	Xaa Leu Xaa Leu Gly Xaa
[1308]	1 5
[1309]	<210> 41
[1310]	<211> 8
[1311]	<212> PRT
[1312]	<213> 人工序列
[1313]	<220>
[1314]	<223> 保守结构域
[1315]	<220>
[1316]	<221> 变体
[1317]	<222> (2) .. (2)
[1318]	<223> Xaa可以是Ile或Val
[1319]	<220>
[1320]	<221> 变体
[1321]	<222> (4) .. (4)
[1322]	<223> Xaa可以是Phe或Leu
[1323]	<220>
[1324]	<221> USNURE
[1325]	<222> (5) .. (5)
[1326]	<223> Xaa可以是任意天然存在的氨基酸
[1327]	<400> 41
[1328]	Lys Xaa Phe Xaa Xaa Gly Glu Asn
[1329]	1 5
[1330]	<210> 42
[1331]	<211> 5
[1332]	<212> PRT
[1333]	<213> 人工序列
[1334]	<220>
[1335]	<223> 保守结构域
[1336]	<400> 42
[1337]	Arg Phe Pro Leu Trp
[1338]	1 5

		10	20	30	40	
SEQ_ID_06/1-329	1	-----MSVLKKLVRTLKKKKDIPSEN-----	QEDIKPQEF	30		
SEQ_ID_10/1-316	1	-----MKYFLLAKRHKKYLKE-----	KKIFRNSTI	26		
SEQ_ID_12/1-342	1	MPIDIKAMNTPS--KQPLRERLHMMRRNRVRKRSVIALIKSHLDSSRY	48			
SEQ_ID_14/1-323	1	-----MNAVERVRNILNYCINEVQMYRQCPNS-----	27			
SEQ_ID_16/1-316	1	-----MLMRALRKMGRVAFDY-----	TNTTKDGAV	28		
		60	70	80	90	
SEQ_ID_06/1-329	31	GHIKHYHFWPLS--NETFFNQFAQEKNL--DLS-----	QTALISCFGEL	70		
SEQ_ID_10/1-316	27	---SFYNFWEIEDYNNFWLQKFI-----	VDRNLNPKNKSINFFSVFGPR	67		
SEQ_ID_12/1-342	49	---QDYNWWDH--ASTFWLPRFI-----	DLHLEPK-KKINLFSCFQNP	86		
SEQ_ID_14/1-323	28	---KYYNFWPCD--YNNNWFNHFVEHRGLAKERH-----	RLNFFSVFGN-	66		
SEQ_ID_16/1-316	29	---CYHNWWPCN--YEEWFHFRFV--VQNIGTERC-----	YHFFSVFGPR	66		
		110	120	130	140	
SEQ_ID_06/1-329	71	SAIPKIPERYKVFFFTGENI-----	YHPDRISYSDPELYRMVDLYLG	111		
SEQ_ID_10/1-316	68	YVLKKQKAAINIFFSGETMSRFK---KYH---DYCLPE---	VDLALG	105		
SEQ_ID_12/1-342	87	LMLIRYYKGVKIFLSGENLTNNEH--FGFHPRMLDHRINE---	VDLALG	130		
SEQ_ID_14/1-323	67	PLLPRIIPGKKVFFFTGENLADN-----	SIH--SIGRAFKKTFPVYDLVLG	109		
SEQ_ID_16/1-316	67	IALT-LPTPNKVFFCGENVHNAEWPKSYQ---	DHALGD---	VKLALG	107	
		160	170	180	190	
SEQ_ID_06/1-329	112	FEYRTEP---KYLRFPPLWWV--YLCGLTKKPHFSHESI	AEFIRKMNQPE	155		
SEQ_ID_10/1-316	106	FDDLQHE---KYFRLPLWIL--DFF-----	EPTVDLEKAKEKLKQLNYYK	145		
SEQ_ID_12/1-342	131	FEFRKDP---KYRFPPLWIYQNEFI-----	SPSASLEDICVLVGQINDPS	172		
SEQ_ID_14/1-323	110	FDYEVEDSRVNYMRFPWLIA--FLI---	DPTADYQKIKETIERINDPS	152		
SEQ_ID_16/1-316	108	YDDIQDE---RYIRFPPLWLL--YMF---	DPVVDRYAIRERIEEINHAE	147		
		210	220	230	240	
SEQ_ID_06/1-329	156	FRLQSSRNRFCSHISSHDTNGIRKRMIDLILPIASVDCAGKFMNNTDELK	205			
SEQ_ID_10/1-316	146	NNKPIVREKFCSLIARHDENGIRKKIVNTLNPIETVDCAGKLFNNTARLQ	195			
SEQ_ID_12/1-342	173	TRRSAKRSRFIGQISSHDKGGMGRGLIDLLSPIGQIDCAGKFRHNTDELL	222			
SEQ_ID_14/1-323	153	TRLNASRDRFACLVASHDKTGIRQKLYDVLMPIASVTCPRGFQNNTNELH	202			
SEQ_ID_16/1-316	148	N---TRKYECVLSRHDKWNMRGPIDALKDHLAISGAGKWKQNTDELW	193			
		260	270	280	290	
SEQ_ID_06/1-329	206	AKFNDDKIDYLKQYRFNLCPENSESVGYITEKIFESIMAGCIPYWGCVK	255			
SEQ_ID_10/1-316	196	TEFANNKVKFLNYKFNICPENTNQESYTTTEKIFESFAAGCIPYWGSAQ	245			
SEQ_ID_12/1-342	223	EYVGDDKFKYLANYRFNLCPENSLGEGYITEKIFDSIRAGCIPYWGAY-	271			
SEQ_ID_14/1-323	203	DLYANDKREYLKLFKFNVCPENSSTPGYITEKIFDSFASGCIPYFGGGT	252			
SEQ_ID_16/1-316	194	TVYNDDKPRYLKEFKFNICPENFDTPYYVTEKIFEAFRSGTIPYAGGGD	243			
		310	320	330	340	
SEQ_ID_06/1-329	256	QLFVEPDILNPEAFI--YYEKGKEE--QLAKQVEELWISPKRYEEFAAIAP	302			
SEQ_ID_10/1-316	246	K--PEPNIFKPSSII--FFDEFK-N--TLSEDVERLHKDPKLYLDFISQNP	289			
SEQ_ID_12/1-342	272	---LEPGILNPKAIL--RFEEGKEQ--EFYNRVKELWENEEAYEQFILLEPP	315			
SEQ_ID_14/1-323	253	EE--IEPDIVNQGAFIYWDGGRMD--WM-DTVRELWESPSAYRAVAEIPP	298			
SEQ_ID_16/1-316	244	H--PEPEIVNRSALL--LWERGQSDHSALVQEVIRLARDEIYYDKFVHQVR	290			
		360	370			
SEQ_ID_06/1-329	303	FKEDAAEVITYWIEELEKRLRAFEPKA	329			
SEQ_ID_10/1-316	290	FQDTAAEYIIQTISNLELKLKEIINQA	316			
SEQ_ID_12/1-342	316	FVEGAAERIWEILQGLRERLAPLVEEG	342			
SEQ_ID_14/1-323	299	FKEQAADVITYAYMENLHDKLAAIVR--	323			
SEQ_ID_16/1-316	291	LLPYTEEFVYEQFSSLKERLLQIRRG-	316			

图1

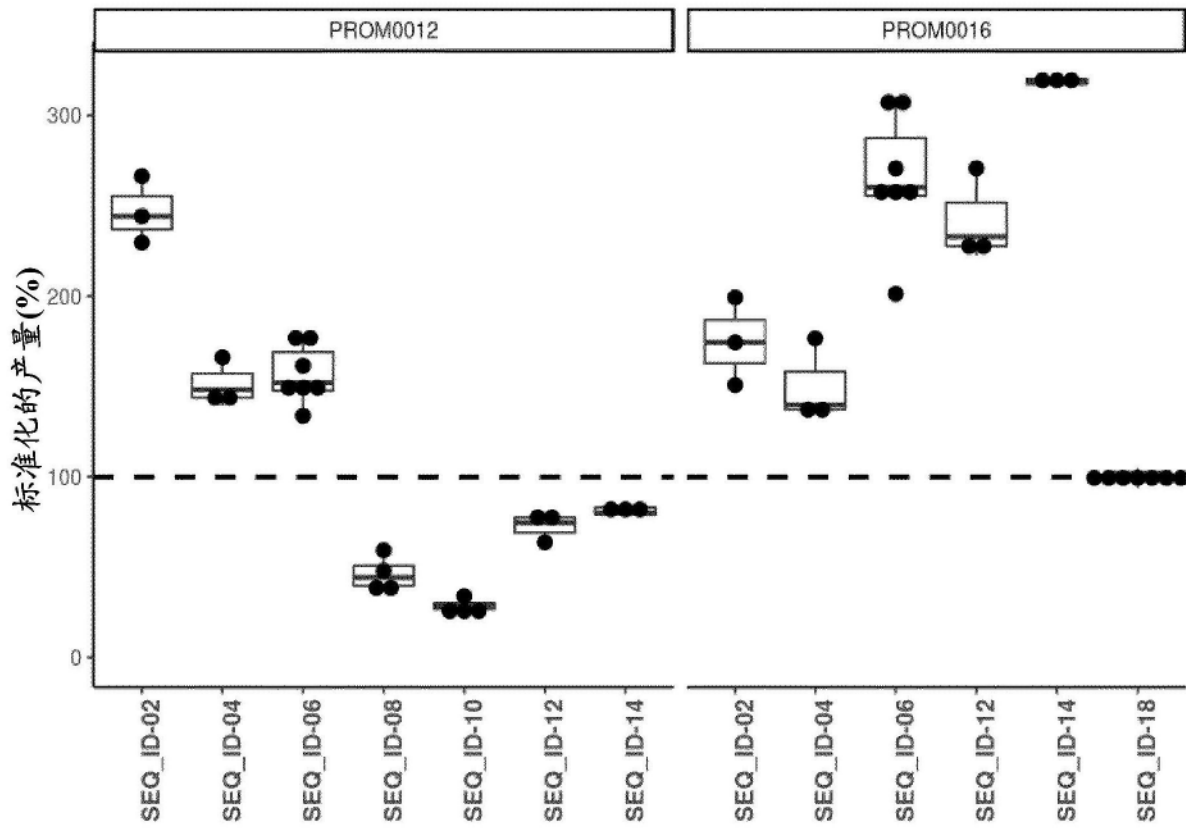


图2

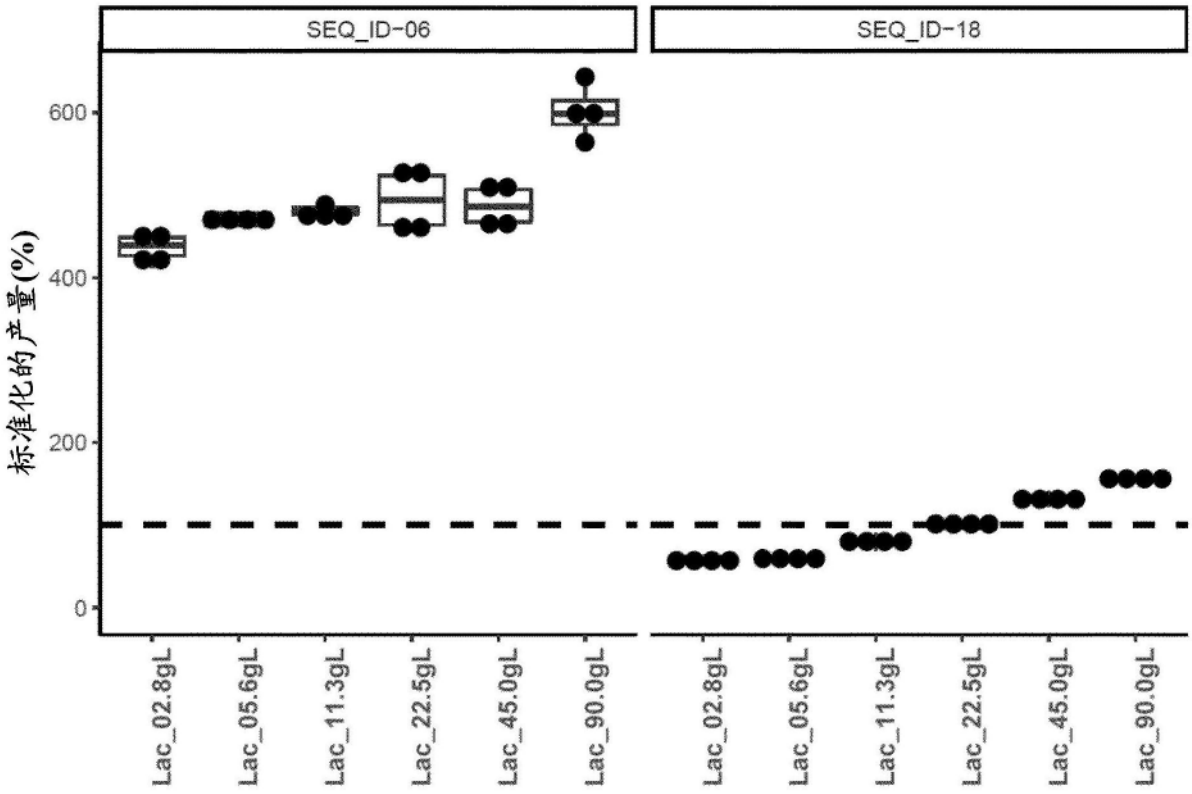


图3

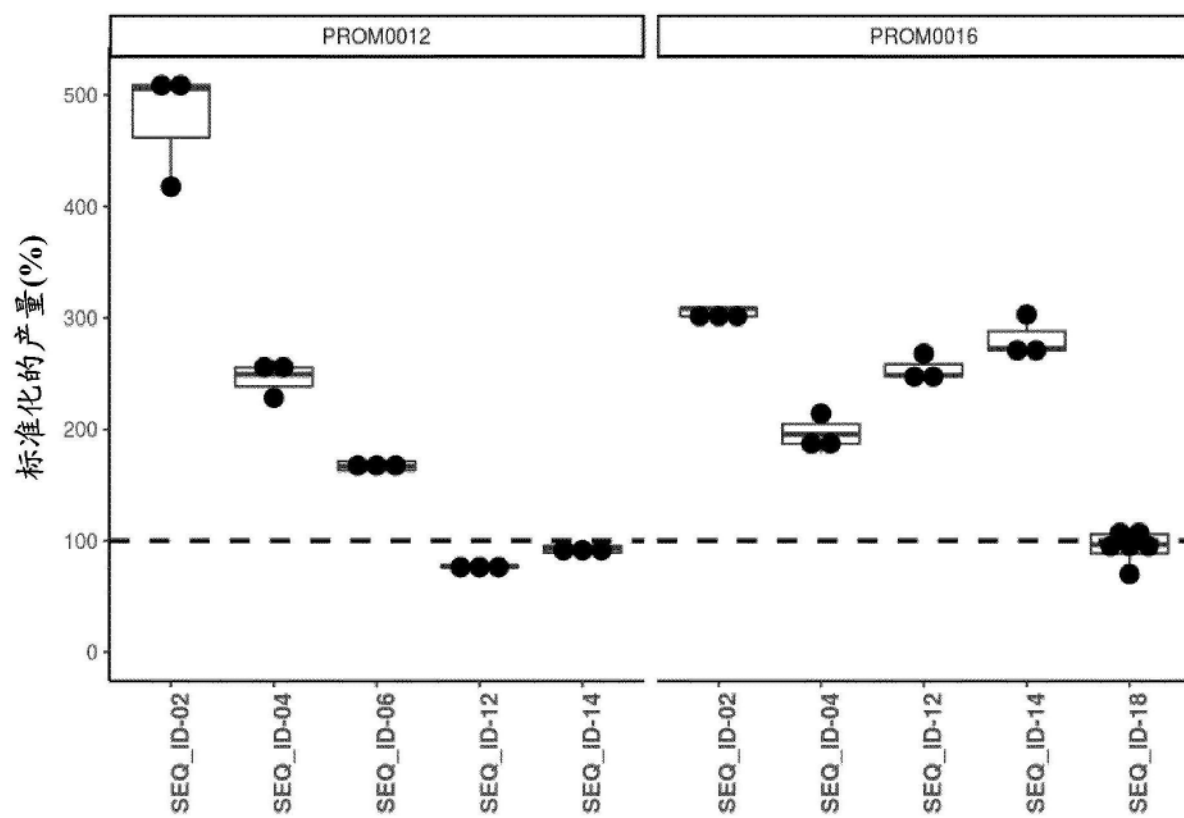


图4

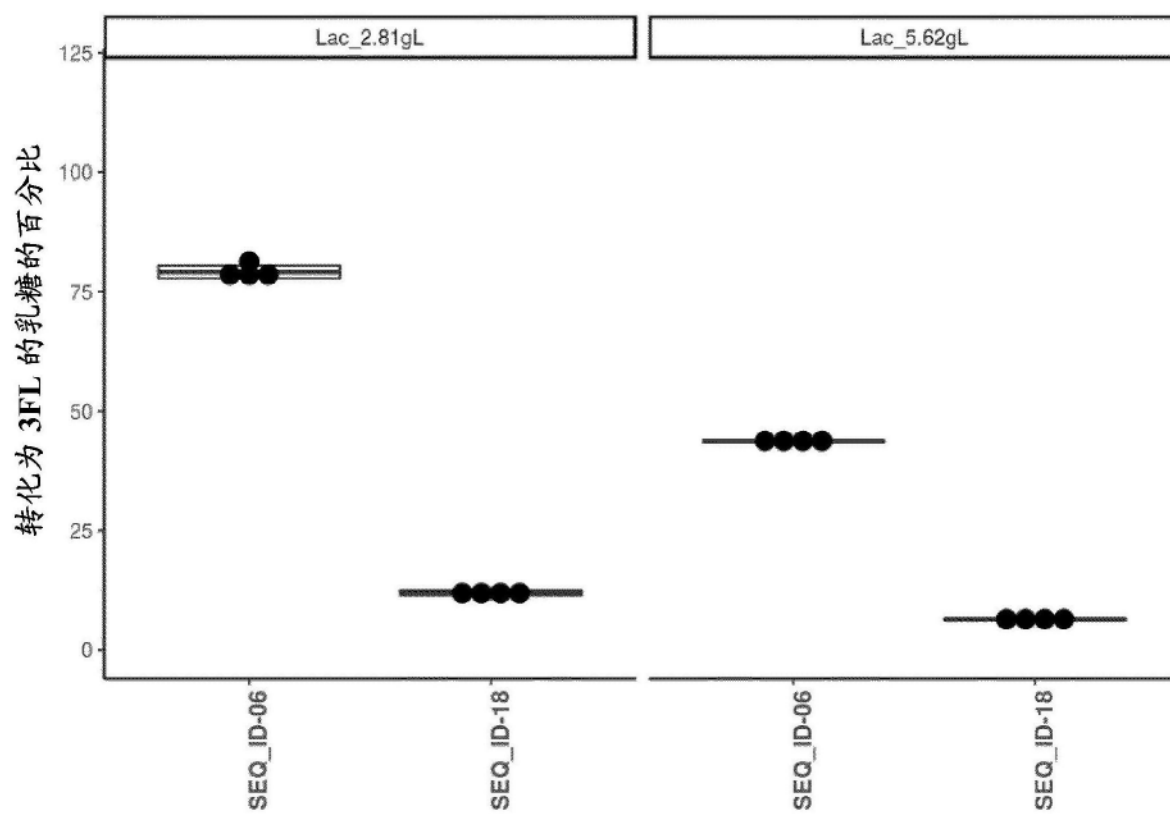


图5

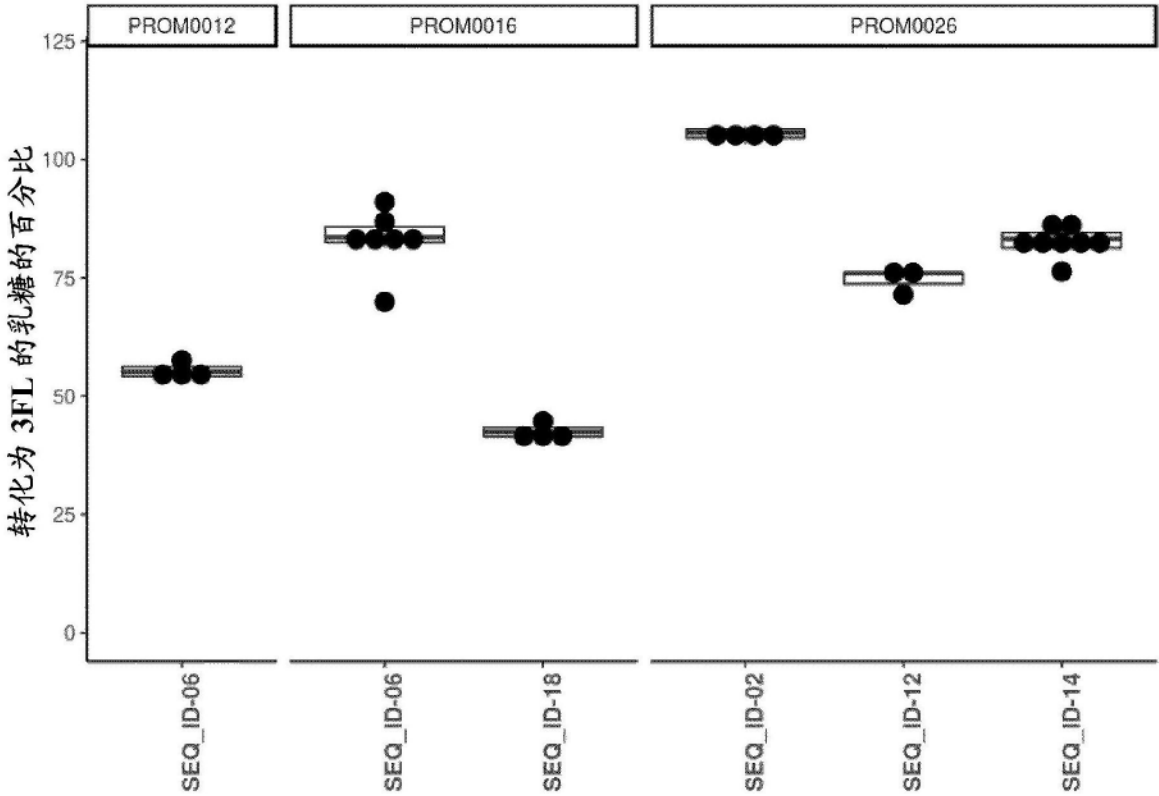


图6

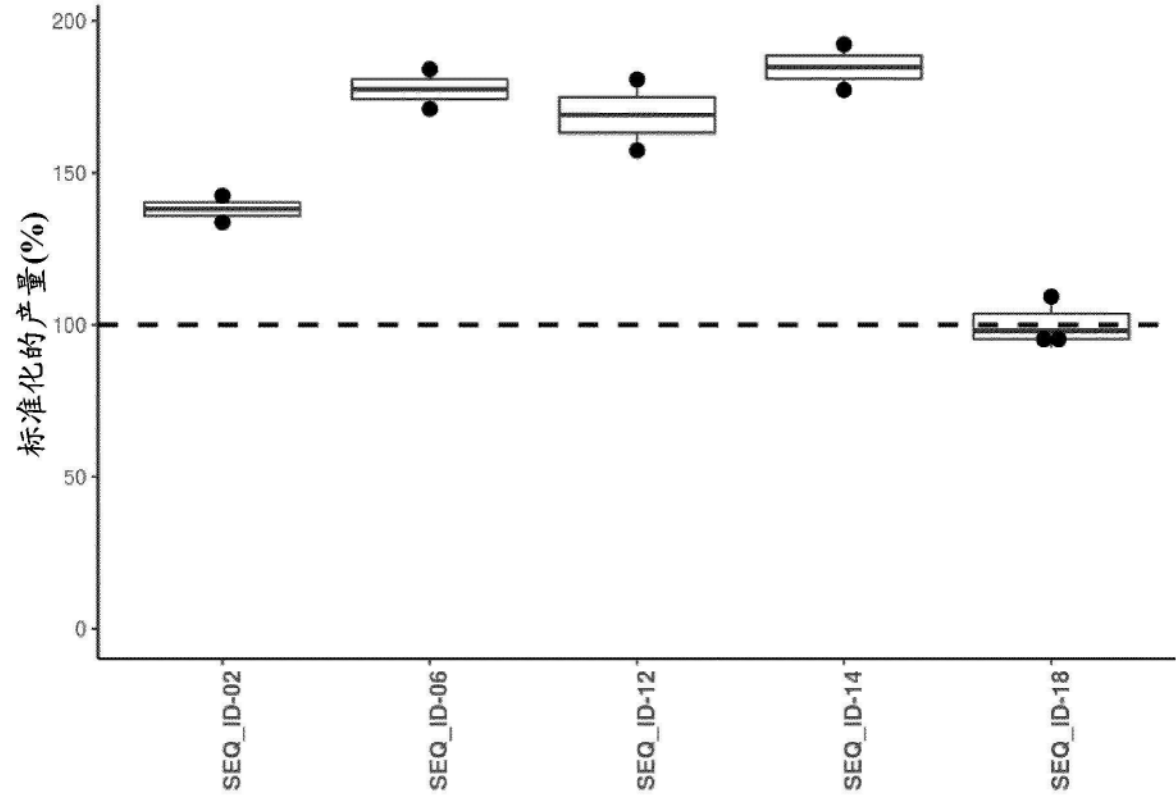


图7

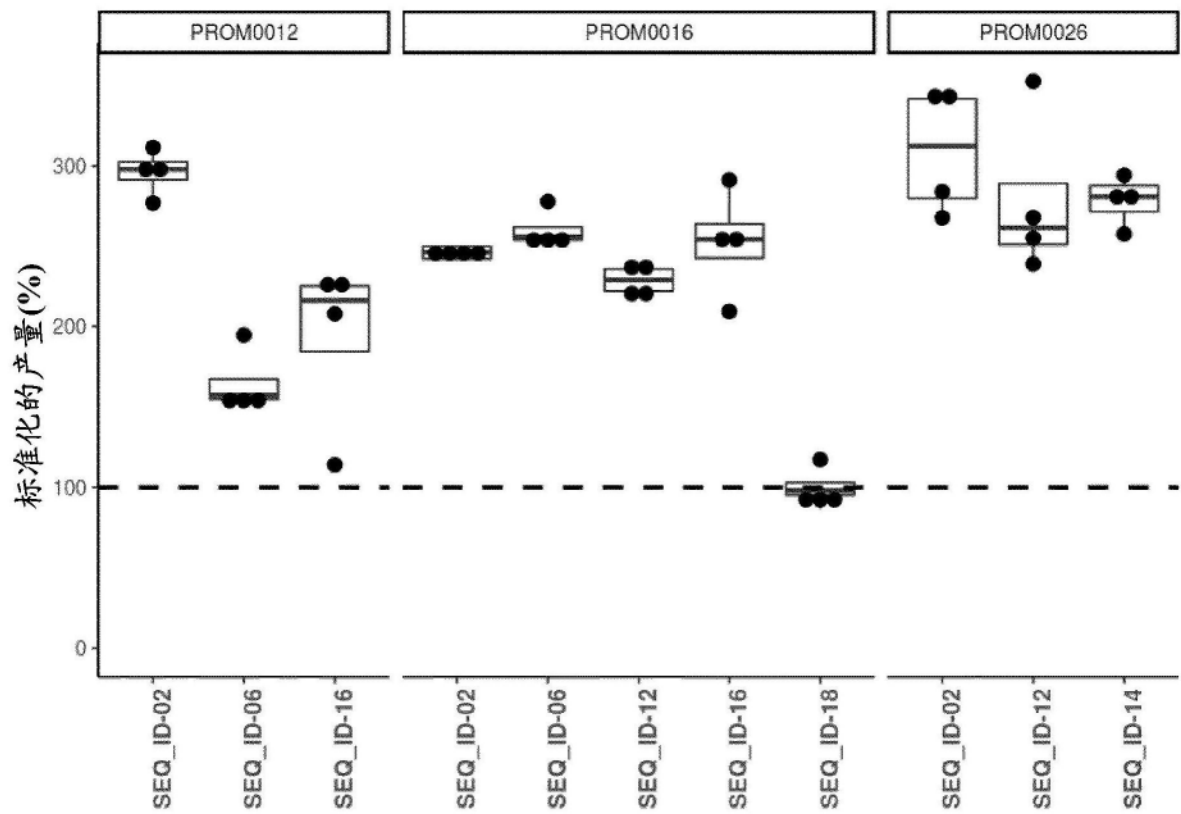


图8

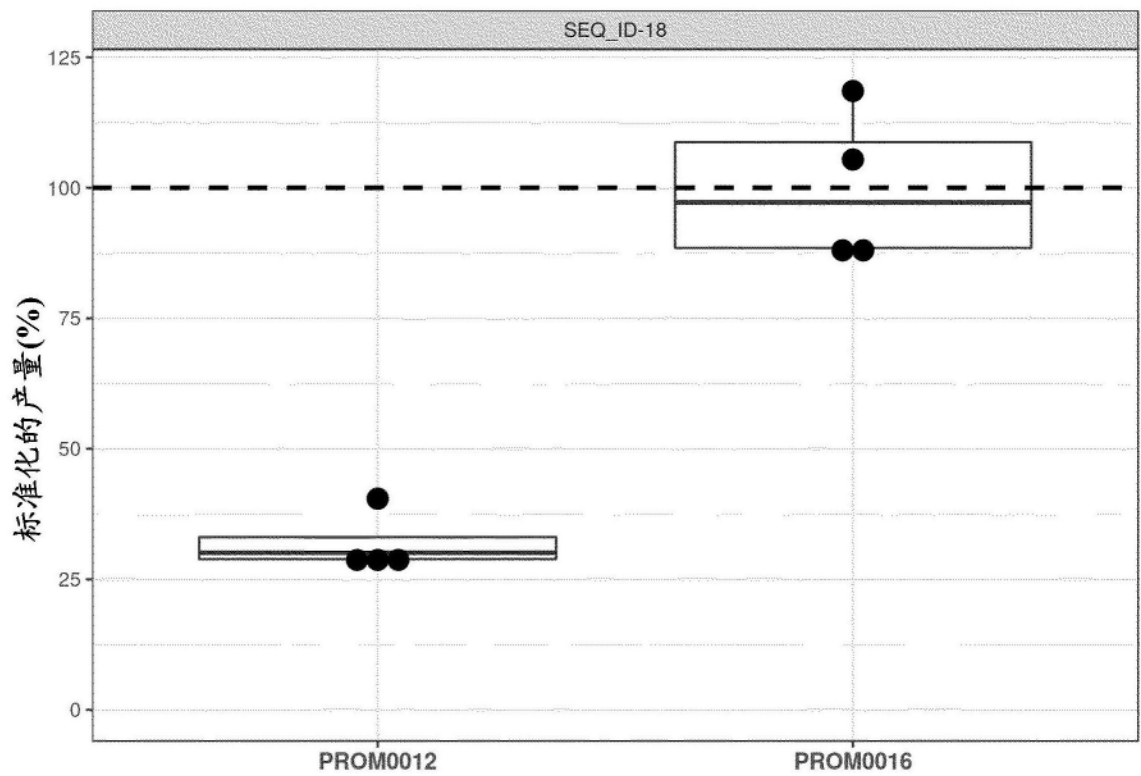


图9

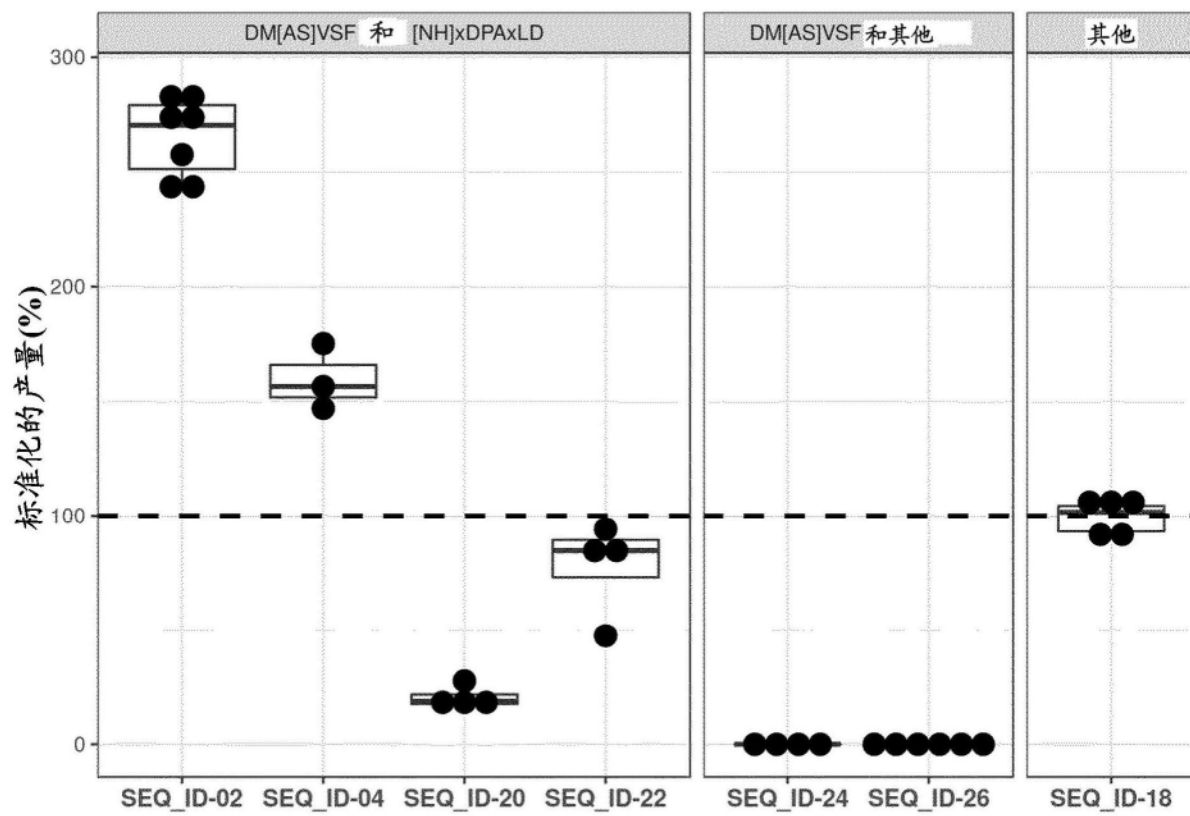


图10

SEQ_ID_02/1-325	1	-----MID--QRTSDFLSEFLASSHRDPAFLDSDFLHGPGRGAAA--KPRLK-----IAFF	48
SEQ_ID_04/1-325	1	-----MID--QRTSDFLSEFLASPNRDPVLDLFLHGPGRGAAA--RPRLK-----IAFF	48
SEQ_ID_20/1-325	1	-----MID--QRTGVFLSEFLDTRNRPVLDLFLHGPGRGAAA--KPRLK-----IAFF	48
SEQ_ID_22/1-325	1	-----MID--RRTSDFLAEFLASANKDPVLDLFLHGPGRGAAA--KPRLK-----IAFF	48
SEQ_ID_08/1-309	1	-----MKTFSLRPVVDA-----IR-----WNFA	20
SEQ_ID_10/1-316	1	-----MKYFLLLAKRHKKYLKE-----KKIFRNSTI-----SFY	29
SEQ_ID_06/1-329	1	-----MSVLKKLVRTLKKKKDIPSEN-----QEDIKQEFEGHIKHY	36
SEQ_ID_12/1-342	1	-----MPIYDIKAMNTPS--KQPLRERLHMRRNRVRKRSVI-----ALIKSHLDSSRY--QDY	51
SEQ_ID_14/1-323	1	-----MNAVERVRNIIYNICINEVQMYRQCPS-----KYY	30
SEQ_ID_16/1-316	1	-----MLMRALRKMKGWRVAFDY-----TNTTKDGAV--CYH	31
SEQ_ID_28/1-343	1	-----MNI--IIHFYARYLRESHNWNR-----E-----REVTNRNGVM--TFA	32
SEQ_ID_30/1-352	1	MLAPYKSPIFVPIYDTKAMNPPT--KQPLRERLHMRRNRIRKRSVI-----ALIKSHLDSSRY--QDY	61
SEQ_ID_32/1-343	1	-----MNI--IIHFYARYLRESHNWNR-----E-----REVTNRNGVM--TFA	32
SEQ_ID_02/1-325	49	FWPFDPAANFFVDIL--SAR-FDVSVDNDSDIAIVSVFGT--RHREARTARSMFFTGENVVRPPL--	110
SEQ_ID_04/1-325	49	FWPFDPSANFFVEIL--SSR-FDVSVDNDSDIAILSVFGE--RHREARTARALFFTGENVRPPL--	110
SEQ_ID_20/1-325	49	FWPFDPSANFFVEIL--SAR-FQVSVDNDSDIAIVSVFGT--GPREIARTARSMFFTGENVVRPPL--	110
SEQ_ID_22/1-325	49	FWPFDPAANFFVEIL--SAR-FDLSVDNDSDIAIVSVFGI--RHREARTARSLFFTGENVRPPL--	110
SEQ_ID_08/1-309	21	NFWSQFDALA-FERHLLGVSGK-HKFRISEQNPQIVFESVFGTTPGKGRERWPKARQVWYTGENVAPPL--	86
SEQ_ID_10/1-316	30	NFWEIEDYNNFWLQKFI-----VDRNLNPKNKSINFFSVFGPRY-VLKKQKAAINIFFSGETMSR--	88
SEQ_ID_06/1-329	37	NFWPIS--NETFFNQFAQEKNL--DLS-----QTALISCFGELS-AIPKIPERYKVFFTGENTI--	89
SEQ_ID_12/1-342	52	NWWDH--ASTFWLPRFI-----DLHLEPK-KKINLFCFQNPPL-MLIRYKGVKIFLSGENLNTN--	107
SEQ_ID_14/1-323	31	NFWPD--YNNNWFNHFVEHRLAKERH-----RLNFFSVFGN-P-LLPRIIPGKKVFFTGENTLAD--	87
SEQ_ID_16/1-316	32	NWPCN--YEEEFHFRFV-----VQNGTE-RCYHFFSVFGPRI-ALTLPN-KVFFCGEN--	83
SEQ_ID_28/1-343	33	NWWRD--PHKNWFARFI-----DAGSKDPE-RRIRFYSIFGPYS-KLKEDFDGAKIFFSGENLEQPVYHRLKLT	98
SEQ_ID_30/1-352	62	NWWDH--ASTFWLPRFI-----DLHLEPK-KRINLFCFQNPPL-MLIRYKGVKIFLSGENLAN--	117
SEQ_ID_32/1-343	33	NWWRD--PHKNWFARFI-----DAGNKDPE-RRIRFYSIFGPYS-KLKEDFDGAKIFFSGENLEQPVYHRLKLT	98

图11

SEQ_ID	02/1-325	111	-----DG-----	160	170	180	190	200	210	220	143
SEQ_ID	04/1-325	111	-----DG-----				VDMSVSFIR-IDDP-----				143
SEQ_ID	20/1-325	111	-----DG-----				VDMSVSFIR-IDHP-----				143
SEQ_ID	22/1-325	111	-----DG-----				VDMSVSFIR-IDDP-----				143
SEQ_ID	08/1-309	87	-----DQ-----				FDKCLSFIRDIKDP-----				113
SEQ_ID	10/1-316	89	-----FK--KYH----				VDLALGFID-LQHE-----				122
SEQ_ID	06/1-329	90	-----YHPDRISYSDPELYRM-----				VDLYLGFY-RTEP-----				128
SEQ_ID	12/1-342	108	-----NEHFGFHPRLDHRINE-----				VDLALGFIF-RKDP-----				148
SEQ_ID	14/1-323	88	-----N--SIH--SIGRAFKKTFPV-----				VDLVLGFY-EVEDS-----				129
SEQ_ID	16/1-316	84	-----VHNAEWP--YK--SYQ----				VKLALGYID-IQDE-----				124
SEQ_ID	28/1-343	99	-----DPEDRIWADRRK--LYG--NYGAGD-----				VDLALGFON-REEDSLMGFEGSRKTKYIRFPLWLT-----				154
SEQ_ID	30/1-352	118	-----NEHFGFHPRLDHRINE-----				VDLALGFIF-RKDP-----				158
SEQ_ID	32/1-343	99	-----DPEDRIWADRRK--LYG--NYGAGE-----				VDLALGFON-REEDSLMGFEGSRKTKYIRFPLWLT-----				154
SEQ_ID	02/1-325	144	GATPHFCQSVLPVPPTREE-----	230	240	250	260	270	280	290	201
SEQ_ID	04/1-325	144	GATPHFCHPVLPPVPTREE-----				-----AAKRKFCAFLYKNPNPCARRNDFFQMLCARRHVESVGWL				201
SEQ_ID	20/1-325	144	GAAPYFCQPVLPVPPTRED-----				-----AAKRKFCAFLYKNPNPCARRNDFFQMLCARRHVESVGWL				201
SEQ_ID	22/1-325	144	GATRHFCHSVLPVPPTREE-----				-----AAERKFCAFLYKNPNPCARRNDFFHMLGARRHVDVGWL				201
SEQ_ID	08/1-309	114	-----LASLPMSFNDLVK-----				-----ADRRKFCAFLYKNPNPCARRNDFFRMLCARRHVESVGWL				201
SEQ_ID	10/1-316	123	-----DFF-----				-----CQSSVSTWAERPFGCAFIANECCQTRNRFVEKLSRYRRVDCPGRV				172
SEQ_ID	06/1-329	129	-----YLKG--LTKKPHFSHESIAEFIRKMNQPEFRLQSSRNRFCSHISSHDTNGIRKRMIDILIPIASVDCAGKF				-----ADRRKFCAFLYKNPNPCARRNDFFHMLGARRHVDVGWL				187
SEQ_ID	12/1-342	149	-----NEFI-----				-----SPSASLEDICVIVGQINDPSTRSAKRSRFIGQISSHDKGGMGRGLDLSPIGQIDCAGKF				197
SEQ_ID	14/1-323	130	-----FLI-----				-----DPTADYQIKIKETIERINDPSTRNASRDRFACLVASHDKTGIRQKLYDVIMPIASVTCGRF				214
SEQ_ID	16/1-316	125	-----YMF-----				-----DPVVDRYAIRERIEEINHAEN-----				194
SEQ_ID	28/1-343	155	-----YVF-----				-----TRKYECLVLSRHDKNWNRGPIYDALKDHLAISCAKW				185
SEQ_ID	30/1-352	159	-----NEFI-----				-----TGRKDTLLASHDFWGTGRSDILKSLEGVCDVSIAGKW				215
SEQ_ID	32/1-343	155	-----YVF-----				-----TGRKDTLLASHDFWGTGRSDILKSLEGVCDVSIAGKW				224

图11(续)

SEQ_ID	ID	02/1-325	202	LNNTGS--VVKMGWL-PKIRVFSRYRFAFENAS---	HPQ	YLTEK	LDAFQAGAVPLYWGDPGL---	RDVAA	266
SEQ_ID	ID	04/1-325	202	LNNTGS--VVKMGWL-PKIRVFSRYRFAFENAA---	HPQ	YLTEK	LDAFQAGTVPLYWGDSGL---	RDVAA	266
SEQ_ID	ID	20/1-325	202	LNNTGS--VVKMGWL-PKIRVFSRYRFAFENAS---	HPQ	YLTEK	LDAFQAGAVPLYWGDPGL---	RDVAA	266
SEQ_ID	ID	22/1-325	202	LNNTGS--VVKMGWL-PKIRVFSRYRFAFENAS---	HPQ	YLTEK	LDAFQAGAVPLYWGDPGL---	RDVAA	266
SEQ_ID	ID	08/1-309	173	LNNTSETLGQRGNLHGKINFLKQKYAVCFENTSTRGSE---	QES	YTTEK	VDAMLACIPLYWGDRVG---	EDFNE	243
SEQ_ID	ID	10/1-316	188	FNNTAR--LQTEFAN-NKVKFLENYKFNICPENTN---	QES	YTTEK	FESFAAGCIPIYWGSAQK---	PEPNI	254
SEQ_ID	ID	06/1-329	198	MNNTDE--LKAKFND-DKIDYLKQYRNLCPENSE---	SVQ	YITEK	FESIMAGCIPIYWGKVQLFVEPDILNP	266	
SEQ_ID	ID	12/1-342	215	RHNTDE--LLEVYGD-DKFKYLANRYRNLCPENSL---	GEQ	YITEK	FDSIRAGCIPIYWGAY---	LEPGILNP	279
SEQ_ID	ID	14/1-323	195	QNNTNE--LHDLNAN-DKREYKLKFKNVCPENSS---	TPQ	YITEK	FDSFASGCIPIYFGGGTEE-	IEPDI	262
SEQ_ID	ID	16/1-316	186	KQNTDE--LWTVYND-DKPRYLKEFKNICPENFD---	TPQ	YITEK	FEAFRSCTIPIYAGGGDH---	PEPEIVNR	252
SEQ_ID	ID	28/1-343	216	RNNTKE--LWEDYNN-DKNKYLSEFKNICPENVD---	APQ	YITEK	FDFAKCGAIPIYQGCLGK---	PEPDVINT	282
SEQ_ID	ID	30/1-352	225	RHNTDE--LLEVYGD-DKFKYLANRYRNLCPENSL---	GEQ	YITEK	FDSIRAGCIPIYWGAY---	LEPGILNP	289
SEQ_ID	ID	32/1-343	216	RNNTKE--LWEDYND-DKNKYLSEFKNICPENVD---	APQ	YITEK	FDFAKCGAIPIYQGCLGK---	PEPNVINT	282

SEQ_ID	ID	02/1-325	267	GSFIDVSRY-ASDEEA-CDAILAADDDYDTRYRST---	PPFL	---	GAEDFYFDA---	FRLAEWIESRL	325
SEQ_ID	ID	04/1-325	267	GSFIDVSRY-ASDEEA-IEAILAIDDDYDSYRRYRG---	APFL	---	GTEDFYFDA---	YRLAEWIESRL	325
SEQ_ID	ID	20/1-325	267	GSFIDVSRY-SSDEEA-IEAILAIDDDYGYRRYRST---	PPFL	---	GTEDFHFDA---	YRLAEWIESRL	325
SEQ_ID	ID	22/1-325	267	GSFIDVSRY-SSDEEA-IDAILAIDDDYDTRYRHRST---	APFL	---	GTEDFYFDA---	FRLAEWIESRL	325
SEQ_ID	ID	08/1-309	244	NSFINLGVY-GNDVNAMVQHVIELDSD---	ERLQNNLFQEPWLP	PEIKSSEHFSFETSKDAILKLVANVNK---	309		
SEQ_ID	ID	10/1-316	255	SSII-FFD--EFK-NTLSEDEVRLHKDPKLYLDIFSQ---	NPFQD	---	TAAEYIIQ	TISNLEKLKEIINQA-	316
SEQ_ID	ID	06/1-329	267	EAFI-YYE--KGKEEQIAKQVEELWISPKRYEEFAAI---	APFKE	---	DAAEVIYTWIEELEKRLRAFEPA-	329	
SEQ_ID	ID	12/1-342	280	KAIL-RFE--EGKEQEFYNRVKELWENEEAYEQFILE---	PPFVE	---	GAAERIWEILOGLRERLAPLVEEG-	342	
SEQ_ID	ID	14/1-323	263	GAFIRYWD--DGRMDWM-DTVRELWESPAYRAVAEI---	PPFKE	---	QAADVIYAYMENLHDKLAAIVR---	323	
SEQ_ID	ID	16/1-316	253	SALL-LWERGSDHSALVQEVIRLARDEIYYDKFVHQ---	VRLLP	---	YTEEFVYEQFSSLKERLLQIRRG---	316	
SEQ_ID	ID	28/1-343	283	DAVL-LWDF-DGDNSDTISLKKLNSDNVYDNFVSQ---	PKFKP	---	DAAEYVACMDLRRSFDQLI---	343	
SEQ_ID	ID	30/1-352	290	KAIL-RFE--EGKEQEFYNRVKELWENEEAYEQFILE---	PPFVE	---	GAAERIWEILOGLRERLAPLVEEG-	352	
SEQ_ID	ID	32/1-343	283	DAVL-LWDF-DGDNSDTIALIKKLNSDNVYDNFVSQ---	PKFKP	---	DAAEYVACMDLRRSFDRLI---	343	

图11(续)

		10	20	30	40	
SEQ_ID_02/1-325	1	MIDQRTSDFLSEFLAS	SHRDPARLDS	FLLHGPGRGARA	AKPRLKIAFFDF	50
SEQ_ID_04/1-325	1	MIDQRTSDFLSEFLAS	PNRDPAVLDF	FLLHGPERGGRA	ARPRLKIAFFDF	50
SEQ_ID_20/1-325	1	MIDQRTGVFLSEFLDT	RNRDPAVLDF	FLLQGPDGGRG	AKPNLKVAFFDF	50
SEQ_ID_22/1-325	1	MIDRRTSDFLAFLAS	ANKDPAVLDF	FLLHGPDGRGGR	SAKPRLKIAFFDF	50
SEQ_ID_24/1-325	1	MLDQRTSAFLEEF	LAKPGGDPERLDF	FLLHGPYRGRRG	GRPRLKLAFYDF	50
SEQ_ID_26/1-303	1	-----	-----MLDF	FLLHGPERGGRA	ARPRLKIAFFDF	28
		60	70	80	90	
SEQ_ID_02/1-325	51	WPEFDPAANFFVDI	LSARFDVSVVDN	SDLAIVSVFGTR	HREARTARSMF	100
SEQ_ID_04/1-325	51	WPEFDPSANFFVEI	LSSRFDVSVVDN	SDLAIVSVFGER	HREARTARALF	100
SEQ_ID_20/1-325	51	WPEFDPSANFFVEI	LSARFQVSVVEN	SDLAIVSVFGTG	PREIR TARSMF	100
SEQ_ID_22/1-325	51	WPEFDPAANFFVEI	LSARFDLSVVDN	SDLAIVSVFGIR	HREARTARSLF	100
SEQ_ID_24/1-325	51	WPEFDTGRNFFIEI	LSSRFDLSVVED	SDLAIVSVFGGR	HRAARSRTLF	100
SEQ_ID_26/1-303	29	WPEFDPSANFFVEI	LSSRFDVSVVDN	SDLAIVSVFGER	HREARTARALF	78
		110	120	130	140	
SEQ_ID_02/1-325	101	FTGENVRPPLDG	V/DMSVSFDR	IDDP RHYRLP	LYVMHAWDHR	REGATPHFC 150
SEQ_ID_04/1-325	101	FTGENVRPPLDG	V/DMSVSFDR	IDHPRHYRLP	LYVMHAWDHR	REGATPHFC 150
SEQ_ID_20/1-325	101	FTGENVRPPLDG	V/DMSVSFDR	IDDP RHFRLP	LYVVHAYDHL	REGAAPYFC 150
SEQ_ID_22/1-325	101	FTGENVRPPLDG	V/DMSVSFDR	IDDP RHYRLP	LYVMHAWDHR	REGATRHFC 150
SEQ_ID_24/1-325	101	FTGENVRPPLDG	V/DMAVSFDR	VGDP RHYRLP	LYVMHAYEHM	REGAVPHFC 150
SEQ_ID_26/1-303	79	FTGENVRPPLDG	V/DMSVSFDR	IDHPRHYRLP	LYVMHAWDHR	REGATPHFC 128
		160	170	180	190	
SEQ_ID_02/1-325	151	QSVLPPVPPTREEA	AKRKFCAFLYKN	PNCARRNDF	FQMLCARRHVES	VGW 200
SEQ_ID_04/1-325	151	HPVLPPVPPTREEA	AKRKFCAFLYKN	PHCARRNDF	FQMLCARRHVES	VGW 200
SEQ_ID_20/1-325	151	QSVLPPVPPTRED	AERKFCAFLYKN	PNCARRNDF	FHMLGARRHVS	VGW 200
SEQ_ID_22/1-325	151	HSVLPPVPPTREEA	DRRKFCAFLYKN	PNCERRNDF	FRMLCARRHVES	VGW 200
SEQ_ID_24/1-325	151	SPVLPPVPPSRAA	FAERNFCAFLYK	NPNGERRN	FFPALDARRRV	DSVGW 200
SEQ_ID_26/1-303	129	HPVLPPVPPTREEA	AKRKFCAFLYKN	PHCARRNDF	FQMLCARRHVES	VGW 178
		210	220	230	240	
SEQ_ID_02/1-325	201	LLNNTGSVVKM	GWLPKIRVFS	RYRFAFAFEN	ASHPGYLTEKI	LDAFQAGA 250
SEQ_ID_04/1-325	201	LLNNTGSVVKM	GWLPKIRVF	ARYRFAFAFEN	AHHPGYLTEKI	LDAFQAGT 250
SEQ_ID_20/1-325	201	LLNNTGSVVKM	GWLPKIRVFS	RYRFAFAFEN	ASHPGYLTEKI	LDAFQAGA 250
SEQ_ID_22/1-325	201	LLNNTGSVVKM	GWLPKIRVFS	RYRFAFAFEN	ASHPGYLTEKI	LDAFQAGA 250
SEQ_ID_24/1-325	201	HLNNTGSVVKM	GWLAKIRVF	ERYRFAFAFEN	ASHPGYLTEKI	LDVFQAGA 250
SEQ_ID_26/1-303	179	LLNNTGSVVKM	GWLPKIRVF	ARYRFAFAFEN	AHHPGYLTEKI	LDAFQAGT 228
		260	270	280	290	
SEQ_ID_02/1-325	251	VPLYWGD	PGVLRDVAAGS	FIDVSR	YASDEEACDAI	LAADDYD TYRRYRS 300
SEQ_ID_04/1-325	251	VPLYWGD	SGVLRDVAAGS	FIDVSR	YASDEEAIEAI	LAIDDDYD SYRRYRG 300
SEQ_ID_20/1-325	251	VPLYWGD	PGVLRDVAAGS	FIDVSR	YSSDEEAIEAI	LAIDDDYD GAYRRYRS 300
SEQ_ID_22/1-325	251	VPLYWGD	PGVLRDVAAGS	FIDVSR	YSSDEEAIDAI	LAIDDDYD TYRRHRS 300
SEQ_ID_24/1-325	251	VPLYWGD	PDVEREVAAGS	FIDVSR	FATDEEAAEHIL	ALDGDYD DAYCAYRA 300
SEQ_ID_26/1-303	229	VPLYWGD	SGVLRDVAAGS	FIDVSR	YASDEEAIEAI	LAIDDDYD SYRRYRG 278
		310	320			
SEQ_ID_02/1-325	301	TPPFLGAEDFY	FD AFR LAEWI	ESRL		325
SEQ_ID_04/1-325	301	TAPFLGTEDFY	FD AFR LAEWI	ESRL		325
SEQ_ID_20/1-325	301	TPPFLGTEDFH	FD AFR LAEWI	ESRL		325
SEQ_ID_22/1-325	301	TAPFLGTEDFY	FD AFR LAEWI	ESRL		325
SEQ_ID_24/1-325	301	VAPFLGTEEFH	FD AFR LAEWI	ESRL		325
SEQ_ID_26/1-303	279	TAPFLGTEDFY	FD AFR LAEWI	ESRL		303

图12

		10	20	30	40	50	
SEQ_ID_06/1-329	1	-----	-----	-----	-----	-----	25
SEQ_ID_12/1-342	1	-----	-----	-----	-----	-----	43
SEQ_ID_14/1-323	1	-----	-----	-----	-----	-----	27
SEQ_ID_16/1-316	1	-----	-----	-----	-----	-----	20
SEQ_ID_28/1-343	1	-----	-----	-----	-----	-----	21
SEQ_ID_30/1-352	1	MLAPYKSP	IFVPIYDTKAMNPPT	-----	-----	-----	53
SEQ_ID_32/1-343	1	-----	-----	-----	-----	-----	21
		60	70	80	90	100	
SEQ_ID_06/1-329	26	KPQEF	GHIKHYHFWPLS	-----	-----	-----	72
SEQ_ID_12/1-342	44	DSSRY	-----	-----	-----	-----	88
SEQ_ID_14/1-323	28	-----	-----	-----	-----	-----	68
SEQ_ID_16/1-316	21	NTTKD	GAVCYHNWPCNYEEWFHFRV	-----	-----	-----	68
SEQ_ID_28/1-343	22	EVTRNG	VMTFANWWREDPHKNWFARF	-----	-----	-----	70
SEQ_ID_30/1-352	54	DSSRY	-----	-----	-----	-----	98
SEQ_ID_32/1-343	22	EVTRNG	VMTFANWWREDPHKNWFARF	-----	-----	-----	70
		120	130	140	150	160	
SEQ_ID_06/1-329	73	IPKIPER	KVFFFTGEN	-----	-----	-----	106
SEQ_ID_12/1-342	89	LIRYYKG	KIFLSGEN	-----	-----	-----	125
SEQ_ID_14/1-323	69	LPRIIPG	KVFFFTGEN	-----	-----	-----	104
SEQ_ID_16/1-316	69	LTLP	TPNKVFFFCGEN	-----	-----	-----	102
SEQ_ID_28/1-343	71	LKEDFDG	AKIFFSGEN	EQPVYHRI	LKTDP	IEDRIWADRRKLYGNYGAGD	121
SEQ_ID_30/1-352	99	LIRYYKG	KIFLSGEN	-----	-----	-----	135
SEQ_ID_32/1-343	71	LKEDFDG	AKIFFSGEN	EQPVLHRI	LKTDP	IEDRIWADRRKLYGNYGAGE	121
		170	180	190	200	210	
SEQ_ID_06/1-329	107	DLYLG	FHEYRTEP	-----	-----	-----	147
SEQ_ID_12/1-342	126	DLALG	FEFRKDP	-----	-----	-----	164
SEQ_ID_14/1-323	105	DLVLG	FDYEVEDS	-----	-----	-----	144
SEQ_ID_16/1-316	103	KLALG	YDDIQDE	-----	-----	-----	139
SEQ_ID_28/1-343	122	DLAIG	FGNREEDSLMGFE	GSRKTKY	-----	-----	169
SEQ_ID_30/1-352	136	DLALG	FEFRKDP	-----	-----	-----	174
SEQ_ID_32/1-343	122	DLAIG	FGNREEDSLMGFE	GSRKTKY	-----	-----	169
		230	240	250	260	270	
SEQ_ID_06/1-329	148	IRKMNQ	PEFRLQSSRNRFCSH	ISSHD	TNGIRKRMIDLI	LPIASVDCAGKFMNNTD	202
SEQ_ID_12/1-342	165	VGQIN	DPSTRRS	AKRSRFIGQ	ISSHD	KGGMRGRLIDLLSPIGQIDCAGKFRHNTD	219
SEQ_ID_14/1-323	145	IERIN	DPSTRLNASDRFAC	LVDVLMPIASVTCPGRFQNTN	199		
SEQ_ID_16/1-316	140	IEEIN	-----	-----	-----	-----	190
SEQ_ID_28/1-343	170	IDEIN	-----	-----	-----	-----	220
SEQ_ID_30/1-352	175	LEQIN	DPSTRRSTGRSRFIGQ	ISSHD	KGGMRGRLIDLLNPIGQIDCAGKFRHNTD	229	
SEQ_ID_32/1-343	170	IDEIN	-----	-----	-----	-----	220
		280	290	300	310	320	
SEQ_ID_06/1-329	203	ELKAK	FNDDKIDYLKQYRNLCPENSESVGYITEKIFESIMAGCIP	IYWGGV	KQL	257	
SEQ_ID_12/1-342	220	ELLE	VGDDKFKYLANRYRNLCPENSELGEYITEKVFD	SIRAGCIP	IYWGAY	271	
SEQ_ID_14/1-323	200	ELHDL	YANDKREYLLKLFKNVCPENSSTPGYITEKLFD	SFASGCIP	IYFGGGT	254	
SEQ_ID_16/1-316	191	ELWT	VYNDKPRYLKEFKNICPENFDTPYYVTEKLFEAF	RSGTIP	IYAGGGD	244	
SEQ_ID_28/1-343	221	ELWED	YNDKKNKYLSEFKNICPENVDAPGYVTEKIFDAFKCGAIP	IYQGCLGK	274		
SEQ_ID_30/1-352	230	ELLE	VGDDKFKYLANRYRNLCPENSELGEYITEKVFD	SIRAGCIP	IYWGAY	281	
SEQ_ID_32/1-343	221	ELWED	YNDKKNKYLSEFKNICPENVDAPGYVTEKIFDAFKCGAIP	IYQGCLGK	274		
		340	350	360	370	380	
SEQ_ID_06/1-329	258	FVEPD	ILNPEAFI	-----	-----	-----	309
SEQ_ID_12/1-342	272	LEPGI	LNPKAIL	-----	-----	-----	322
SEQ_ID_14/1-323	255	IEPD	IVNQGA	FIYWD	-----	-----	305
SEQ_ID_16/1-316	245	PEPE	IVNRSALL	-----	-----	-----	297
SEQ_ID_28/1-343	275	PEPD	VINTDAVL	-----	-----	-----	326
SEQ_ID_30/1-352	282	LEPGI	LNPKAIL	-----	-----	-----	332
SEQ_ID_32/1-343	275	PEPN	VINTDAVL	-----	-----	-----	326
		390	400				
SEQ_ID_06/1-329	310	VIYT	WIEELEKRLRAFEPKA				329
SEQ_ID_12/1-342	323	RIWE	ILQGLRERLAPLVEEG				342
SEQ_ID_14/1-323	306	VIY	AYMENLHDKLAAIVR				323
SEQ_ID_16/1-316	298	FVYE	QFSSLKERLLQIRRG				316
SEQ_ID_28/1-343	327	YVV	ACMDELRRSFDQLI				343
SEQ_ID_30/1-352	333	RIWE	ILQGLRERLAPLVEEG				352
SEQ_ID_32/1-343	327	YVV	ACMDELRRSFDRLI				343

图13