



(12) 发明专利

(10) 授权公告号 CN 111868252 B

(45) 授权公告日 2025. 06. 03

(21) 申请号 201980018867.0

(22) 申请日 2019.03.12

(65) 同一申请的已公布的文献号
申请公布号 CN 111868252 A

(43) 申请公布日 2020.10.30

(30) 优先权数据
62/641590 2018.03.12 US
62/682260 2018.06.08 US
62/695252 2018.07.09 US

(85) PCT国际申请进入国家阶段日
2020.09.11

(86) PCT国际申请的申请数据
PCT/US2019/021876 2019.03.12

(87) PCT国际申请的公布数据
W02019/178116 EN 2019.09.19

(73) 专利权人 科纳根公司
地址 美国马萨诸塞州

(72) 发明人 毛国红 M·巴腾 P·宏特
余晓丹

(74) 专利代理机构 中国专利代理(香港)有限公
司 72001
专利代理师 初明明 黄希贵

(51) Int.Cl.
C12N 9/10 (2006.01)
C12P 19/56 (2006.01)

(56) 对比文件
CN 107105686 A, 2017.08.29
US 2016186225 A1, 2016.06.30
WO 2016073740 A1, 2016.05.12
WO 2018031955 A2, 2018.02.15
审查员 关维

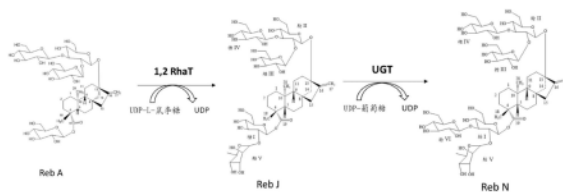
权利要求书2页 说明书32页
序列表27页 附图9页

(54) 发明名称

甜菊糖苷莱苞迪昔J和莱苞迪昔N的生物合成生产

(57) 摘要

本公开涉及通过使用莱苞迪昔A作为底物和涉及各种1,2RhaT-鼠李糖基转移酶的生物合成途径来产生甜菊糖苷莱苞迪昔J和莱苞迪昔N。



1. 一种制备莱苞迪昔N的生物合成方法,所述方法包括:

在具有1,2-鼠李糖基转移酶活性的第一重组多肽的存在下使甜菊糖苷组合物与鼠李糖供体部分反应;其中所述第一重组多肽如SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:9或SEQ ID NO:19的氨基酸序列所示,

其中所述甜菊糖苷组合物包含莱苞迪昔A,所述莱苞迪昔A被所述第一重组多肽转化为莱苞迪昔J;和

在具有葡糖基转移酶活性的第二重组多肽的存在下使所述莱苞迪昔J与葡萄糖供体部分反应;其中所述第二重组多肽如SEQ ID NO:7、SEQ ID NO:11、SEQ ID NO:13或SEQ ID NO:15的氨基酸序列所示。

2. 根据权利要求1所述的方法,所述方法包括在转化的细胞系统中表达所述第一重组多肽。

3. 根据权利要求2所述的方法,其中所述转化的细胞系统选自由以下组成的组:酵母、不产生甜菊糖苷的植物、藻类和细菌。

4. 根据权利要求2所述的方法,其中所述转化的细胞系统是真菌。

5. 根据权利要求3所述的方法,其中所述转化的细胞系统为选自由以下组成的组的细菌或酵母:埃希氏菌属、沙门氏菌属、芽孢杆菌属、不动杆菌属、链霉菌属、棒状杆菌属、甲基弯菌属、甲基单胞菌属、红球菌属、假单胞菌属、红细菌属、集胞藻属、接合酵母属、克鲁维酵母属、假丝酵母属、汉逊酵母属、德巴利酵母属、毛霉菌属、毕赤酵母属、球拟酵母属、曲霉属、关节杆菌属、短杆菌属、微杆菌属、节杆菌属、柠檬酸杆菌属、克雷伯氏菌属、泛菌属和梭菌属。

6. 根据权利要求3所述的方法,其中所述转化的细胞系统为酵母属。

7. 根据权利要求2所述的方法,其中反应步骤在所述转化的细胞系统中进行。

8. 根据权利要求2所述的方法,所述方法包括从所述转化的细胞系统中分离所述第一重组多肽,并且反应步骤在体外进行。

9. 根据权利要求1或2所述的方法,其中所述鼠李糖供体部分为UDP-L-鼠李糖。

10. 根据权利要求1所述的方法,其中所述葡萄糖供体部分为原位生成的。

11. 根据权利要求1所述的方法,其中所述第二重组多肽具有葡糖基转移酶活性和蔗糖合酶活性两者。

12. 根据权利要求1所述的方法,所述方法包括在具有蔗糖合酶活性的第三重组多肽存在下使所述莱苞迪昔J与葡萄糖供体部分反应。

13. 根据权利要求1或2所述的方法,其中所述甜菊糖苷组合物包含莱苞迪昔I。

14. 根据权利要求13所述的方法,其中所述莱苞迪昔I通过在具有葡糖基转移酶活性的第二重组多肽存在下使包含莱苞迪昔A的甜菊糖苷组合物与葡萄糖供体部分反应来制备。

15. 根据权利要求14所述的方法,其中所述葡萄糖供体部分为原位生成的。

16. 根据权利要求14所述的方法,其中所述第二重组多肽具有葡糖基转移酶活性和蔗糖合酶活性两者。

17. 根据权利要求14所述的方法,其中所述第二重组多肽如SEQ ID NO:7、SEQ ID NO:11、SEQ ID NO:13或SEQ ID NO:15的氨基酸序列所示。

18. 根据权利要求14所述的方法,所述方法包括在具有蔗糖合酶活性的第三重组多肽

存在下使所述包含莱苞迪昔A的甜菊糖苷组合物与葡萄糖供体部分反应。

19. 一种制备莱苞迪昔J的生物合成方法,所述方法包括:

在具有1,2-鼠李糖基转移酶活性的重组多肽存在下使包含莱苞迪昔A的甜菊糖苷组合物与鼠李糖供体部分反应;其中所述重组多肽如SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:9或SEQ ID NO:19的氨基酸序列所示。

20. 根据权利要求19所述的方法,所述方法包括在转化的细胞系统中表达所述重组多肽。

21. 根据权利要求20所述的方法,其中所述转化的细胞系统选自由以下组成的组:酵母、不产生甜菊糖苷的植物、藻类、和细菌。

22. 根据权利要求20所述的方法,其中所述转化的细胞系统是真菌。

23. 根据权利要求21所述的方法,其中所述转化的细胞系统为选自由以下组成的组的细菌或酵母:埃希氏菌属、沙门氏菌属、芽孢杆菌属、不动杆菌属、链霉菌属、棒状杆菌属、甲基弯菌属、甲基单胞菌属、红球菌属、假单胞菌属、红细菌属、集胞藻属、接合酵母属、克鲁维酵母属、假丝酵母属、汉逊酵母属、德巴利酵母属、毛霉菌属、毕赤酵母属、球拟酵母属、曲霉属、关节杆菌属、短杆菌属、微杆菌属、节杆菌属、柠檬酸杆菌属、克雷伯氏菌属、泛菌属和梭菌属。

24. 根据权利要求21所述的方法,其中所述转化的细胞系统为酵母属。

25. 根据权利要求20所述的方法,其中反应步骤在所述转化的细胞系统中进行。

26. 根据权利要求20所述的方法,所述方法包括从所述转化的细胞系统中分离所述重组多肽,并且反应步骤在体外进行。

27. 根据权利要求19所述的方法,其中所述鼠李糖供体部分为UDP-L-鼠李糖。

甜菊糖苷莱苞迪苷J和莱苞迪苷N的生物合成生产

[0001] 相关申请的交叉引用

[0002] 本申请要求2018年3月12日提交的美国临时专利申请号62/641,590和2018年6月8日提交的美国临时专利申请号62/682,260的优先权,这两个临时专利申请的内容全文以引用方式并入本文。

技术领域

[0003] 本公开一般涉及甜菊糖苷的生物合成。更具体地,本公开涉及用于制备包含莱苞迪苷J(“Reb J”)和/或莱苞迪苷N(“Reb N”)的组合物的生物催化方法,以及具有在用于产生Reb J和/或Reb N的相关生物合成途径中有用的酶活性的重组多肽。

背景技术

[0004] 甜菊糖苷是在甜叶菊(*Stevia rebaudiana*)植物的叶中发现的一类化合物,其可用作高强度、低热量的甜味剂。这些天然存在的甜菊糖苷共有相同的基本二萜结构(甜菊醇主链),但在甜菊醇主链的C13和C19位置的糖残基(例如葡萄糖、鼠李糖和木糖残基)的数量和类型方面不同。有趣的是,基本甜菊醇结构的糖“装饰”中的这些变化通常显著且不可预测性地影响所得甜菊糖苷的特性。受影响的特性可包括但不限于总体口感特征、任何异味的存在和程度、结晶点、“口感”、溶解度和感觉到的甜度以及其他差异。具有已知结构的甜菊糖苷包括甜菊苷、莱苞迪苷A、莱苞迪苷B、莱苞迪苷C、莱苞迪苷D、莱苞迪苷E、莱苞迪苷F、莱苞迪苷M、莱苞迪苷J、莱苞迪苷N和杜尔可苷A。

[0005] 基于干重计,甜菊苷、莱苞迪苷A、莱苞迪苷C和杜尔可苷A分别占野生型甜叶菊叶中发现的所有甜菊糖苷的总重量的约9.1%、3.8%、0.6%和0.3%。其他甜菊糖苷诸如Reb J和Reb N以显著更低的含量存在。来自甜叶菊植物的提取物是可商购获得的。在此类提取物中,甜菊苷和莱苞迪苷A通常是主要组分,而其他已知的甜菊糖苷则作为微量或痕量组分存在。在任何给定的甜叶菊提取物中的各种甜菊糖苷的实际含量水平可根据例如甜叶菊植物生长的气候和土壤、收获甜叶菊叶的条件和用于提取期望的甜菊糖苷的方法而变化。出于说明,商业制备物中莱苞迪苷A的量可在总甜菊糖苷含量的约20重量%至大于约90重量%之间变化,而莱苞迪苷B、莱苞迪苷C和莱苞迪苷D的量可分别为总甜菊糖苷含量的约1重量%至2重量%、约7重量%至15重量%和约2重量%。在此类提取物中,莱苞迪苷J和莱苞迪苷N各自通常占总甜菊糖苷含量的小于0.5重量%。

[0006] 作为天然甜味剂,不同的甜菊糖苷具有不同程度的甜味、口感和余味。甜菊糖苷的甜度显著高于食用糖(即蔗糖)的甜度。例如,甜菊苷本身的甜度是蔗糖的100-150倍,但具有苦的回味,如在许多味道测试中所注意到的,而莱苞迪苷A和E的甜度是蔗糖的250-450倍,并且回味特征比甜菊苷好得多。然而,这些甜菊糖苷本身仍保持明显的余味。因此,任何甜叶菊提取物的总体味道特征均受提取物中各种甜菊糖苷的相对含量的深刻影响,而所述相对含量又受植物来源、环境因素(诸如土壤含量和气候)和提取方法的影响。特别地,提取条件的变化可导致甜叶菊提取物中甜菊糖苷的组成不一致,使得味道特征在不同批次的

提取产品中变化。甜叶菊提取物的味道特征也可能受提取过程后保留在产品中的植物来源或环境来源的污染物(诸如色素、脂质、蛋白质、酚类和糖类)的影响。这些污染物通常具有对于将甜叶菊提取物用作甜味剂来说不期望的异味。此外,分离在甜叶菊提取物中不丰富的单独或特定组合的甜菊糖苷的方法可能会在成本和资源上令人望而却步。

[0007] 此外,从植物中提取的方法通常采用使用溶剂诸如己烷、氯仿和乙醇的固-液提取技术。溶剂萃取为能量密集型方法,并且可能导致与有毒废物处理有关的问题。因此,需要新的生产方法来降低甜菊糖苷的生产成本以及减轻大规模培养和加工的环境影响。

[0008] 因此,在本领域中需要一种新的制备甜菊糖苷特别是微量甜菊糖苷诸如Reb J和Reb N的方法,其可产生具有更好和更一致的味道特征的产品。更优选地,此类制备方法可利用更丰富的甜菊糖苷诸如Reb A以降低生产成本。

发明内容

[0009] 在一些实施方案中,本公开涵盖从Reb A制备Reb J的方法以及通过Reb J的中间体从Reb A制备Reb N的方法。在一些实施方案中,本公开提供了从Reb I制备Reb N的方法,以及通过Reb I的中间体从Reb A制备Reb N的方法。

[0010] 在一个实施方案中,本公开提供了通过本文所述的各种1,2-鼠李糖基转移酶从Reb A产生甜菊糖苷莱苞迪苷J“Reb J”或13-[(2-O-β-D-吡喃葡萄糖基-3-O-β-D-吡喃葡萄糖基-β-D-吡喃葡萄糖基)氧基]对映-贝壳杉-16-烯-19-羧酸-[(2-O-α-L-吡喃鼠李糖基-β-D-吡喃葡萄糖基)酯]。图1示出了Reb J的化学结构。

[0011] 在另一个实施方案中,本公开提供了通过本文所述的各种UDP-糖基转移酶从Reb J产生甜菊糖苷莱苞迪苷N“Reb N”或13-[(2-O-β-D-吡喃葡萄糖基-3-O-β-D-吡喃葡萄糖基-β-D-吡喃葡萄糖基)氧基]对映-贝壳杉-16-烯-19-羧酸-[(2-O-α-L-吡喃鼠李糖基-3-O-β-D-吡喃葡萄糖基-β-D-吡喃葡萄糖基)酯]。图2示出了Reb N的化学结构。

[0012] 本方法提供了使用某些特定的合成途径合成特定的甜菊糖苷的方法。

[0013] 在产品/商业实用性方面,在美国市场上存在数十种含有甜菊糖苷的产品,并且可用于从食品、饮料和饮食补充剂到止痛药和驱虫剂的任何事物中。含有甜菊糖苷的产品可以是液体、颗粒制剂、凝胶或气溶胶。

[0014] 本文尤其提供了制备莱苞迪苷诸如莱苞迪苷N的生物合成方法,该方法包括在具有1,2-鼠李糖基转移酶活性的第一重组多肽存在下使甜菊糖苷组合物与鼠李糖供体部分反应;其中该第一重组多肽包含与SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19具有至少80%(例如,至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)序列同一性的氨基酸序列。

[0015] 在一些实施方案中,该第一重组多肽包含与SEQ ID NO:3具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的同一性的氨基酸序列。

[0016] 在一些实施方案中,该第一重组多肽包含与SEQ ID NO:9具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的同一性的氨基酸序列。

[0017] 在一些实施方案中,该第一重组多肽包含与SEQ ID NO:19具有至少80%、至少

85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的同一性的氨基酸序列。

[0018] 在一些实施方案中,本文提供的生物合成方法包括在转化的细胞系统中表达该第一重组多肽。在一些实施方案中,转化的细胞系统选自由以下组成的组:酵母、不产生甜菊糖苷的植物、藻类、真菌和细菌。在一些实施方案中,细菌或酵母选自由以下组成的组:埃希氏菌属(*Escherichia*)、沙门氏菌属(*Salmonella*)、芽孢杆菌属(*Bacillus*)、不动杆菌属(*Acinetobacter*)、链霉菌属(*Streptomyces*)、棒状杆菌属(*Corynebacterium*)、甲基弯菌属(*Methylosinus*)、甲基单胞菌属(*Methylomonas*)、红球菌属(*Rhodococcus*)、假单胞菌属(*Pseudomonas*)、红细菌属(*Rhodobacter*)、集胞藻属(*Synechocystis*)、酵母属(*Saccharomyces*)、接合酵母属(*Zygosaccharomyces*)、克鲁维酵母属(*Kluyveromyces*)、假丝酵母属(*Candida*)、汉逊酵母属(*Hansenula*)、德巴利酵母属(*Debaryomyces*)、毛霉菌属(*Mucor*)、毕赤酵母属(*Pichia*)、球拟酵母属(*Torulopsis*)、曲霉属(*Aspergillus*)、关节杆菌属(*Arthrobotlys*)、短杆菌属(*Brevibacteria*)、微杆菌属(*Microbacterium*)、节杆菌属(*Arthrobacter*)、柠檬酸杆菌属(*Citrobacter*)、克雷伯氏菌属(*Klebsiella*)、泛菌属(*Pantoea*)和梭菌属(*Clostridium*)。

[0019] 在一些实施方案中,本文提供的生物合成方法包括在转化的细胞系统中进行的反应步骤。在其他实施方案中,反应步骤可在体外进行。在一些实施方案中,生物合成方法包括从转化的细胞系统中分离第一重组多肽,并且反应步骤可在体外进行。

[0020] 在一些实施方案中,鼠李糖供体为鼠李糖。在一些实施方案中,鼠李糖供体为L-鼠李糖。在一些实施方案中,鼠李糖供体部分为UDP-L-鼠李糖。

[0021] 在一些实施方案中,甜菊糖苷组合物包含莱菔迪苷A,并且反应步骤导致莱菔迪苷J的产生。

[0022] 在一些实施方案中,本文提供的生物合成方法还包括在具有葡糖基转移酶活性的第二重组多肽存在下使莱菔迪苷J与葡萄糖供体部分反应。在一些实施方案中,葡萄糖供体部分为原位生成的。

[0023] 在一些实施方案中,第二重组多肽具有葡糖基转移酶活性和蔗糖合酶活性两者。在一些实施方案中,该第二重组多肽包含与SEQ ID NO:7、SEQ ID NO:11、SEQ ID NO:13或SEQ ID NO:15具有至少80%(例如,至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)序列同一性的氨基酸序列。

[0024] 在一些实施方案中,该第二重组多肽包含与SEQ ID NO:7具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的同一性的氨基酸序列。

[0025] 在一些实施方案中,该第二重组多肽包含与SEQ ID NO:11具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的同一性的氨基酸序列。

[0026] 在一些实施方案中,该第二重组多肽包含与SEQ ID NO:13具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的同一性的氨基酸序列。

[0027] 在一些实施方案中,该第二重组多肽包含与SEQ ID NO:15具有至少80%、至少

85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的同一性的氨基酸序列。

[0028] 在一些实施方案中,本文提供的生物合成方法还包括在具有蔗糖合酶活性的第三重组多肽存在下使莱菔迪昔J与葡萄糖供体部分反应。

[0029] 在一些实施方案中,甜菊糖苷组合物可包括莱菔迪昔I。在一些实施方案中,莱菔迪昔I可通过在具有葡糖基转移酶活性的第二重组多肽存在下使包含莱菔迪昔A的甜菊糖苷组合物与葡萄糖供体部分反应来制备。

[0030] 在一些实施方案中,本文提供的生物合成方法还包括在具有蔗糖合酶活性的第三重组多肽存在下使包含莱菔迪昔A的甜菊糖苷组合物与葡萄糖供体部分反应。

[0031] 本文还尤其提供了制备莱菔迪昔诸如莱菔迪昔J的生物合成方法,该方法包括在具有1,2-鼠李糖基转移酶活性的重组多肽存在下使包含莱菔迪昔A的甜菊糖苷组合物与鼠李糖供体部分反应;其中所述重组多肽包含与SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19具有至少80%(例如,至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)序列同一性的氨基酸序列。

[0032] 在一些实施方案中,具有1,2-鼠李糖基转移酶活性的重组多肽包含与SEQ ID NO:3具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的氨基酸序列。

[0033] 在一些实施方案中,具有1,2-鼠李糖基转移酶活性的重组多肽包含与SEQ ID NO:9具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的氨基酸序列。

[0034] 在一些实施方案中,具有1,2-鼠李糖基转移酶活性的重组多肽包含与SEQ ID NO:19具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的氨基酸序列。

[0035] 本公开的方面还提供了通过本文(包括任何上述实施方案)所述的任何生物合成方法可获得或通过其产生的莱菔迪昔。

[0036] 本公开的方面还提供了编码如本文所述的多肽的核酸。在一些实施方案中,核酸包含编码多肽的序列,该多肽包含与SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的氨基酸序列。在一些实施方案中,核酸包含与SEQ ID NO:4、SEQ ID NO:10或SEQ ID NO:20具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的序列。在一些实施方案中,核酸包含SEQ ID NO:4的序列。在一些实施方案中,核酸包含SEQ ID NO:10的序列。在一些实施方案中,核酸包含SEQ ID NO:20的序列。在一些实施方案中,核酸为质粒或其他载体。

[0037] 本公开的方面还提供了包含本文(包括任何上述实施方案)所述的核酸的细胞。

[0038] 本公开的其他方面提供了包含至少一种多肽的组合物,该多肽包含与SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的氨基酸序列。在一些实施方案中,组合物包含至少一种包含SEQ ID NO:3的序列的多肽。在一些实施方案中,组合物包含至少一种包含SEQ ID NO:9的序列的多肽。在一些实施方案中,组合物包含至少一种包含SEQ ID NO:19的

序列的多肽。在一些实施方案中,组合物是体外反应混合物,例如包含如本文所述的鼠李糖供体部分和如本文所述的甜菊糖苷组合物。

[0039] 本公开的方面提供了包含至少一种多肽的细胞,该多肽包含与SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的氨基酸序列。在一些实施方案中,细胞包含至少一种包含SEQ ID NO:3的序列的多肽。在一些实施方案中,细胞包含至少一种包含SEQ ID NO:9的序列的多肽。在一些实施方案中,细胞包含至少一种包含SEQ ID NO:19的序列的多肽。在一些实施方案中,细胞为酵母细胞、不产生甜菊糖苷的植物细胞、藻类细胞、真菌细胞或细菌细胞。在一些实施方案中,细菌或酵母选自由以下组成的组:埃希氏菌属、沙门氏菌属、芽孢杆菌属、不动杆菌属、链霉菌属、棒状杆菌属、甲基弯菌属、甲基单胞菌属、红球菌属、假单胞菌属、红细菌属、集胞藻属、酵母属、接合酵母属、克鲁维酵母属、假丝酵母属、汉逊酵母属、德巴利酵母属、毛霉菌属、毕赤酵母属、球拟酵母属、曲霉属、关节杆菌属、短杆菌属、微杆菌属、节杆菌属、柠檬酸杆菌属、克雷伯氏菌属、泛菌属和梭菌属。在一些实施方案中,细胞还包含如本文所述具有葡糖基转移酶活性或葡糖基转移酶活性和蔗糖合酶活性的第二多肽,诸如包含与SEQ ID NO:7、SEQ ID NO:11、SEQ ID NO:13或SEQ ID NO:15具有至少80% (例如,至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%) 序列同一性的氨基酸序列的第二多肽。

[0040] 对于该实施方案中的细胞系统,其可选自由以下组成的组:一种或多种细菌、一种或多种酵母、以及它们的组合、或可允许用所选择的基因进行遗传转化并随后生物合成产生期望甜菊糖苷的任何细胞系统。在最优选的微生物系统中,使用大肠杆菌(E. coli)来产生期望的甜菊糖苷化合物。

[0041] 在一些实施方案中,本公开提供了包含SEQ ID NO:3的氨基酸序列并且被鉴定为EUCP1的EU11酶的突变体。在一些实施方案中,本公开提供了包含与SEQ ID NO:3具有至少90% (例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%) 同一性的氨基酸序列的重组多肽。

[0042] 在一些实施方案中,本公开提供了具有对应于EUCP1的序列并且包含SEQ ID NO:4的DNA分子。在一些实施方案中,本公开提供了具有与SEQ ID NO:4具有至少90% (例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%) 同一性的序列的核酸分子。

[0043] 在一些实施方案中,本公开提供了包含SEQ ID NO:1的氨基酸序列的EU11酶的突变体。在一些实施方案中,本公开提供了包含与SEQ ID NO:1具有至少90% (例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%) 同一性的氨基酸序列的重组多肽。

[0044] 在一些实施方案中,本公开提供了具有对应于EU11的序列并且包含SEQ ID NO:2的DNA分子。在一些实施方案中,本公开提供了具有与SEQ ID NO:2具有至少90% (例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%) 同一性的序列的核酸分子。

[0045] 在一些实施方案中,本公开提供了包含SEQ ID NO:9的氨基酸序列的本文称为UGT2E-B的酶。在一些实施方案中,本公开提供了包含与SEQ ID NO:9具有至少90% (例如,

至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的氨基酸序列的重组多肽。

[0046] 在一些实施方案中,本公开提供了具有对应于UGT2E-B的序列并且包含SEQ ID NO:10的DNA分子。在一些实施方案中,本公开提供了具有与SEQ ID NO:10具有至少90%(例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的序列的核酸分子。

[0047] 在一些实施方案中,本公开提供了包含SEQ ID NO:19的氨基酸序列的本文称为NX114的酶。在一些实施方案中,本公开提供了包含与SEQ ID NO:19具有至少90%(例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的氨基酸序列的重组多肽。

[0048] 在一些实施方案中,本公开提供了具有对应于NX114并且包含SEQ ID NO:20的序列的DNA分子。在一些实施方案中,本公开提供了具有与SEQ ID NO:20具有至少90%(例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的序列的核酸分子。

[0049] 在一些实施方案中,本公开提供了包含能够产生本文所述的一种或多种酶的载体的微生物宿主细胞。在某些实施方案中,酶可自由以下组成的组:EUCP1[SEQ ID NO.3]、UGT2E-B[SEQ ID No.9]和NX114[SEQ ID NO.19]。在一些实施方案中,酶可自由以下组成的组:CP1[SEQ ID NO.11]和CP2[SEQ ID NO.13]。在一些实施方案中,该酶可以是本文称为GS[SEQ ID No.15]的融合酶。在一些实施方案中,宿主细胞选自自由以下组成的组:细菌、酵母、丝状真菌、蓝细菌藻类和植物细胞。在一些实施方案中,宿主细胞选自自由以下组成的组:埃希氏菌属、沙门氏菌属、芽孢杆菌属、不动杆菌属、链霉菌属、棒状杆菌属、甲基弯菌属、甲基单胞菌属、红球菌属、假单胞菌属、红细菌属、集胞藻属、酵母属、接合酵母属、克鲁维酵母属、假丝酵母属、汉逊酵母属、德巴利酵母属、毛霉菌属、毕赤酵母属、球拟酵母属、曲霉属、关节杆菌属、短杆菌属、微杆菌属、节杆菌属、柠檬酸杆菌属、克雷伯氏菌属、泛菌属、棒状杆菌属、梭菌属(例如,丙酮丁醇梭菌)。在一些实施方案中,宿主细胞为从植物中分离的细胞,所述植物选自自由以下组成的组:大豆、油菜籽、向日葵、棉花、玉米、烟草、苜蓿、小麦、大麦、燕麦、高粱、稻、西兰花、菜花、卷心菜、欧洲防风草、甜瓜、胡萝卜、芹菜、欧芹、番茄、马铃薯、草莓、花生、葡萄、草籽作物、甜菜、甘蔗、菜豆、豌豆、裸麦、亚麻、硬木树、软木树、牧草、拟南芥、稻(水稻)、大麦、柳枝稷(*Panicum virgatum*)、短柄草属、芸苔属和海甘蓝。

[0050] 在一些实施方案中,本公开提供了产生莱菔迪昔J的方法,该方法包括将底物与重组多肽一起温育,该重组多肽包含与SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19具有至少80%(例如,至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的氨基酸序列。在一些实施方案中,重组多肽为与SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19具有至少80%(例如,至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的1,2鼠李糖基转移酶。在一些实施方案中,底物选自自由以下组成的组:甜茶苷、甜菊苷或莱菔迪昔A以及它们的组合。

[0051] 在一些实施方案中,本公开提供了包含通过上述方法的任何实施方案产生的Reb J的甜味剂。

[0052] 在一些实施方案中,本公开提供了产生莱菔迪昔N的方法,该方法包括将底物与第

一重组多肽一起温育,该第一重组多肽包含与SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19具有至少80% (例如,至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的氨基酸序列。在一些实施方案中,底物选自由以下组成的组:Reb J、Reb I、Reb A、甜菊苷或甜茶苷以及它们的组合。在一些实施方案中,底物包含Reb J。在其他实施方案中,底物包含Reb I。在一些实施方案中,该方法还包括将具有葡萄糖基转移酶活性的第二重组多肽和任选的重组蔗糖合酶与底物和第一重组多肽一起温育。

[0053] 在一些实施方案中,本公开提供了包含通过上述方法的任何实施方案产生的Reb N的甜味剂。

[0054] 在一些实施方案中,本公开提供了一种从莱菔迪苷J合成莱菔迪苷N的方法,该方法包括:制备包含莱菔迪苷J、选自由蔗糖、尿苷二磷酸(UDP)和尿苷二磷酸葡萄糖(UDP-葡萄糖)组成的组的葡萄糖供体部分以及UGT酶的反应混合物,将反应混合物温育足够的时间以产生莱菔迪苷N,其中葡萄糖共价偶联至莱菔迪苷J以产生莱菔迪苷N。在各种实施方案中,UGT酶可以是包含与UGT76G1[SEQ ID No.7]、CP1[SEQ ID No.11]、CP2[SEQ ID No.13]或GS[SEQ ID No.15]具有至少80%序列同一性的氨基酸序列的多肽。在一些实施方案中,该方法还包括向反应混合物中添加蔗糖合酶。在一些实施方案中,蔗糖合酶选自由以下组成的组:拟南芥蔗糖合酶1、拟南芥蔗糖合酶3和绿豆蔗糖合酶。在一些实施方案中,蔗糖合酶为拟南芥蔗糖合酶1(SEQ ID NO:17)。

[0055] 在一些实施方案中,所述Reb N为大于70% (例如,大于80%、大于85%、大于90%、大于95%、大于96%、大于97%、大于98% ,或大于99%)纯。

[0056] 在一些实施方案中,本公开提供了包含能够产生其中氨基酸序列对应于SEQ ID NO:3、SEQ ID NO:9或SEQ ID NO:19的酶的载体的宿主细胞。在一些实施方案中,宿主细胞选自由以下组成的组:细菌、酵母、丝状真菌、蓝细菌藻类和植物细胞。在一些实施方案中,宿主细胞选自由以下组成的组:埃希氏菌属、沙门氏菌属、芽孢杆菌属、不动杆菌属、链霉菌属、棒状杆菌属、甲基弯菌属、甲基单胞菌属、红球菌属、假单胞菌属、红细菌属、集胞藻属、酵母属、接合酵母属、克鲁维酵母属、假丝酵母属、汉逊酵母属、德巴利酵母属、毛霉菌属、毕赤酵母属、球拟酵母属、曲霉属、关节杆菌属、短杆菌属、微杆菌属、节杆菌属、柠檬酸杆菌属、克雷伯氏菌属、泛菌属、棒状杆菌属、梭菌属。在一些实施方案中,宿主细胞为从植物中分离的细胞,所述植物选自由以下组成的组:大豆、油菜籽、向日葵、棉花、玉米、烟草、苜蓿、小麦、大麦、燕麦、高粱、稻、西兰花、菜花、卷心菜、欧洲防风草、甜瓜、胡萝卜、芹菜、欧芹、番茄、马铃薯、草莓、花生、葡萄、草籽作物、甜菜、甘蔗、菜豆、豌豆、裸麦、亚麻、硬木树、软木树、牧草、拟南芥、稻(水稻)、大麦、柳枝稷(*Panicum virgatum*)、短柄草属、芸苔属和海甘蓝。

[0057] 在一些实施方案中,本公开提供了包含以下的饮料产品:至多约125ppm的莱菔迪苷N;和至少一种选自由以下组成的组的非营养性甜味剂:Reb J、Reb W、Reb V、Reb D4、Reb E和Reb M以及它们的组合,其中该至少一种非营养性甜味剂以约30ppm至约600ppm的浓度存在。在一些实施方案中,该至少一种非营养性甜味剂选自由以下组成的组:Reb J、Reb W、Reb V、Reb D4、Reb E和Reb M以及它们的组合,并且其中所述Reb N和所述至少一种非营养性甜味剂以约1:5的重量比存在。

[0058] 在一些实施方案中,本公开提供了一种GS融合酶,其包含具有与UGT76G1至少90%

相同的氨基酸序列的第一结构域和具有与AtSUS1至少90%相同的氨基酸序列的第二结构域。GS融合酶可具有包含SEQ ID NO:15的氨基酸序列。在一些实施方案中,本公开提供了包含与SEQ ID NO:15具有至少90%(例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的氨基酸序列的重组多肽。

[0059] 在一些实施方案中,本公开提供了具有对应于GS的序列并且包含SEQ ID NO:16的DNA分子。在一些实施方案中,本公开提供了具有与SEQ ID NO:16具有至少90%(例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的序列的核酸分子。

[0060] 在一些实施方案中,本公开提供了包含UGT2E-B酶并且具有SEQ ID NO:9的氨基酸序列的UDP-糖基转移酶。在一些实施方案中,本公开提供了包含与SEQ ID NO:9具有至少90%(例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的氨基酸序列的重组多肽。

[0061] 在一些实施方案中,本公开提供了具有对应于UGT2E-B的序列并且包含SEQ ID NO:10的DNA分子。在一些实施方案中,本公开提供了具有与SEQ ID NO:10具有至少90%(例如,至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)同一性的序列的核酸分子。

[0062] 在一些实施方案中,本公开提供了包含能够产生UGT2E-B酶的载体的微生物宿主细胞。在一些实施方案中,宿主细胞选自由以下组成的组:细菌、酵母、丝状真菌、蓝细菌藻类和植物细胞。在一些实施方案中,宿主细胞选自由以下组成的组:埃希氏菌属、沙门氏菌属、芽孢杆菌属、不动杆菌属、链霉菌属、棒状杆菌属、甲基弯菌属、甲基单胞菌属、红球菌属、假单胞菌属、红细菌属、集胞藻属、酵母属、接合酵母属、克鲁维酵母属、假丝酵母属、汉逊酵母属、德巴利酵母属、毛霉菌属、毕赤酵母属、球拟酵母属、曲霉属、关节杆菌属、短杆菌属、微杆菌属、节杆菌属、柠檬酸杆菌属、克雷伯氏菌属、泛菌属、棒状杆菌属、梭菌属。在一些实施方案中,宿主细胞为从植物中分离的细胞,所述植物选自由以下组成的组:大豆、油菜籽、向日葵、棉花、玉米、烟草、苜蓿、小麦、大麦、燕麦、高粱、稻、西兰花、菜花、卷心菜、欧洲防风草、甜瓜、胡萝卜、芹菜、欧芹、番茄、马铃薯、草莓、花生、葡萄、草籽作物、甜菜、甘蔗、菜豆、豌豆、裸麦、亚麻、硬木树、软木树、牧草、拟南芥、稻(水稻)、大麦、柳枝稷(*Panicum virgatum*)、短柄草属、芸苔属和海甘蓝。

[0063] 在一些实施方案中,本公开提供了产生莱菔迪昔N的方法,该方法包括将底物与重组多肽一起温育,该重组多肽包含与SEQ ID NO:11或SEQ ID NO:13或SEQ ID NO:15具有至少80%同一性的氨基酸序列。在一些实施方案中,底物选自由以下组成的组:Reb J、Reb A、甜菊苷或甜茶苷以及它们的组合。在一些实施方案中,该方法还包括将重组蔗糖合酶与底物和重组多肽一起温育。

[0064] 在一些实施方案中,本公开提供了包含能够产生其中氨基酸序列对应于SEQ ID NO:11和13的CP1和CP2酶的载体的宿主细胞。在一些实施方案中,宿主细胞选自由以下组成的组:细菌、酵母、丝状真菌、蓝细菌藻类和植物细胞。在一些实施方案中,宿主细胞选自由以下组成的组:埃希氏菌属、沙门氏菌属、芽孢杆菌属、不动杆菌属、链霉菌属、棒状杆菌属、甲基弯菌属、甲基单胞菌属、红球菌属、假单胞菌属、红细菌属、集胞藻属、酵母属、接合酵母属、克鲁维酵母属、假丝酵母属、汉逊酵母属、德巴利酵母属、毛霉菌属、毕赤酵母属、球拟酵

母属、曲霉属、关节杆菌属、短杆菌属、微杆菌属、节杆菌属、柠檬酸杆菌属、克雷伯氏菌属、泛菌属、棒状杆菌属、梭菌属。在一些实施方案中,宿主细胞为从植物中分离的细胞,所述植物选自自由以下组成的组:大豆、油菜籽、向日葵、棉花、玉米、烟草、苜蓿、小麦、大麦、燕麦、高粱、稻、西兰花、菜花、卷心菜、欧洲防风草、甜瓜、胡萝卜、芹菜、欧芹、番茄、马铃薯、草莓、花生、葡萄、草籽作物、甜菜、甘蔗、菜豆、豌豆、裸麦、亚麻、硬木树、软木树、牧草、拟南芥、稻(水稻)、大麦、柳枝稷(*Panicum virgatum*)、短柄草属、芸苔属和海甘蓝。

[0065] 在一些实施方案中,本公开提供了包含SEQ ID NO:11的氨基酸序列并且被鉴定为CP1的突变酶。

[0066] 在一些实施方案中,本公开提供了包含SEQ ID NO:13的氨基酸序列并且被鉴定为CP2的突变酶。

[0067] 虽然本公开易受各种修改形式和替代形式的影响,但其具体实施方案在附图中以举例的方式示出,并且将在本文中详细描述。然而,应当理解,本文提供的附图和具体实施方式并非旨在将本公开限制为所公开的具体实施方案,相反,目的在于涵盖落入由所附权利要求限定的本公开的实质和范围内的所有修改、等同物和另选的替代形式。

[0068] 结合附图,本发明的其他特征和优点将在以下对本发明的优选实施方案的详细描述中变得显而易见。

附图说明

[0069] 图1示出了莱苞迪昔J(“Reb J”)的化学结构。

[0070] 图2示出了莱苞迪昔N(“Reb N”)的化学结构。

[0071] 图3A示出了用于产生Reb J和Reb N的生物合成途径。更具体地,从莱苞迪昔A(“Reb A”)开始,可使用1,2鼠李糖基转移酶(“1,2 RhaT”)通过催化鼠李糖部分从UDP-鼠李糖供体转移至Reb A受体的19-O-葡萄糖的C-2’来产生Reb J。在随后的反应中,可使用UDP-糖基转移酶(“UGT”)通过催化葡萄糖部分从UDP-葡萄糖供体转移至Reb J受体的19-O-葡萄糖的C-3’而从Reb J产生Reb N。图3B示出了使用UGT从Reb A产生Reb I的生物合成途径。然后可使用1,2 RhaT将Reb I转化为Reb N。

[0072] 图4示出了通过HPLC分析的从Reb A由所选择的1,2 RhaT酶催化的Reb J的体外产生。图A示出了Reb A标准品的保留时间。图B示出了Reb J标准品的保留时间。图C示出了使用Reb A作为底物从由EU11(SEQ ID NO:1)催化的反应系统中获得的产物的保留时间。图D示出了使用Reb A作为底物从由EUCP1(SEQ ID NO:3)催化的反应系统中获得的产物的保留时间。图E示出了使用Reb A作为底物从由HV1(SEQ ID NO:5)催化的反应系统中获得的产物的保留时间。箭头指示Reb J的存在。

[0073] 图5示出了通过HPLC分析的从Reb A由UGT2E-B催化的Reb J的体外产生。图A示出了Reb A标准品的保留时间。图B示出了Reb J标准品的保留时间。图C示出了使用Reb A作为底物从由UGT2E-B(SEQ ID NO:9)催化的反应系统中获得的产物的保留时间。箭头指示Reb J的存在。

[0074] 图6示出了通过HPLC分析的从Reb A由NX114催化的Reb J的体外产生。图A示出了Reb J标准品的保留时间。图B示出了Reb A标准品的保留时间。图C示出了使用Reb A作为底物从由NX114(SEQ ID NO:19)催化的反应系统中获得的产物的保留时间。箭头指示Reb J的

存在。

[0075] 图7示出了通过HPLC分析的从Reb J由UGT76G1催化的Reb N的体外产生。图A示出了包括莱菔迪昔O、Reb N、莱菔迪昔C和杜尔可昔A的各种甜菊糖苷标准品的保留时间。图B示出了使用从EU11-催化的反应中获得的Reb J从由UGT76G1 (SEQ ID NO:7) 催化的反应系统中获得的产物的保留时间。图C示出了使用从EUCP1-催化的反应中获得的Reb J从由UGT76G1 (SEQ ID NO:7) 催化的反应系统中获得的产物的保留时间。箭头指示Reb N的存在。

[0076] 图8示出了通过HPLC分析的从Reb J由所选择的UGT酶催化的Reb N的体外产生。图A示出了包括莱菔迪昔O、Reb N、莱菔迪昔C和杜尔可昔A的各种甜菊糖苷标准品的保留时间。图B示出了Reb J标准品的保留时间。图B示出了从EUCP1-催化的反应中获得的Reb J中间体的保留时间(如箭头所示)。图D示出了使用图B中所示的Reb J中间体从由UGT76G1 (SEQ ID NO:7) 催化的反应系统中获得的Reb N产物的保留时间。图E示出了使用图B中所示的Reb J中间体从由CP1 (SEQ ID NO:11) 催化的反应系统中获得的Reb N产物的保留时间(如箭头所示)。图F示出了使用图B中所示的Reb J中间体从由CP2 (SEQ ID NO:13) 催化的反应系统中获得的Reb N产物的保留时间(如箭头所示)。图G示出了使用图B中所示的Reb J中间体从由融合酶GS (SEQ ID NO:15) 催化的反应系统中获得的Reb N产物的保留时间(如箭头所示)。

[0077] 图9示出了通过HPLC分析的从Reb J在UDP和蔗糖存在下由所选择的UGT酶催化的Reb N的体外产生。图A示出了Reb N标准品的保留时间。图B示出了Reb J标准品的保留时间。图C示出了从UGT76G1-催化的反应中获得的产物的保留时间。没有观察到Reb N的产生。箭头指示Reb J的存在。图D示出了从由UGT76G1和AtSUS1 (SEQ ID NO:17) 两者催化的反应系统中获得的产物的保留时间。箭头指示Reb N的存在。图E示出了从由融合酶GS (SEQ ID NO:15) 催化的反应系统中获得的产物的保留时间。箭头指示Reb N的存在。

具体实施方式

[0078] 对本文所用术语的解释:

[0079] 甜菊糖苷是一类负责南美植物甜叶菊(菊科(Asteraceae))的叶的甜味的化合物,并且可在食品、饲料和饮料中用作甜味剂。

[0080] 定义:

[0081] 细胞系统是提供异位蛋白质表达的任何细胞。其包括细菌、酵母、植物细胞和动物细胞。其包括原核细胞和真核细胞两者。其还包括基于细胞组分(诸如核糖体)的蛋白质的体外表达。

[0082] 编码序列将被赋予其对于本领域普通技术人员来说普通且惯用的含义,并且非限制地用于指编码特定氨基酸序列的DNA序列。

[0083] 使细胞系统生长包括提供使细胞繁殖和分裂的适当培养基。其还包括提供资源,使得细胞或细胞组分可翻译和制备重组蛋白质。

[0084] 蛋白质表达。蛋白质生成可在基因表达后发生。其由DNA已转录为信使RNA (mRNA) 之后的阶段组成。然后将mRNA翻译成多肽链,其最终折叠成蛋白质。DNA通过转染存在于细胞中,转染是一种有意将核酸引入细胞中的方法。该术语常用于真核细胞中的非病毒方法。其还可涉及其他方法和细胞类型,但其他术语是优选的:“转化”更通常用于描述细菌、非动

物真核细胞,包括植物细胞中的非病毒DNA转移。在动物细胞中,转染是优选的术语,因为转化还用于是指这些细胞中的癌性状态(致癌作用)的进展。转导通常用于描述病毒介导的DNA转移。转化、转导和病毒感染包含在本申请的转染的定义下。

[0085] 酵母。根据本公开,如本文所要求保护的酵母是被归类为真菌界的成员的真核、单细胞微生物。酵母是由多细胞祖先进化的单细胞生物体,但是其中可用于本公开的一些物种是具有通过形成被称为伪菌丝(pseudo hyphae)或假菌丝的连接的芽殖细胞串而发展多细胞特性的能力的那些物种。

[0086] UGT名称。本公开中使用的UGT酶的名称与UGT命名委员会(Mackenzie等人,“The UDP glycosyltransferase gene super family:recommended nomenclature updated based on evolutionary divergence,”*P_{HARMACOGENETICS}*,1997,第7卷,第255-269页)采用的命名系统一致,其通过家族编号、表示亚家族的字母和单个基因的编号的组合将UGT基因分类。例如,名称“UGT76G1”是指由属于UGT家族编号76(其为植物来源)、亚家族G和基因编号1的基因编码的UGT酶。

[0087] 结构术语:

[0088] 如本文所用,单数形式“一个”、“一种”和“该”包括复数指代,除非内容另外明确指出。

[0089] 就说明书或权利要求书中使用的术语“包括”、“具有”等来说,此类术语旨在以类似于术语“包含”的方式为包含性的,如同“包含”在权利要求书中用作过渡词时所解释的那样。

[0090] 词语“示例性”在本文中用来表示“用作示例、实例或说明”。本文中被描述为“示例性”的任何实施方案未必被解释为比其他实施方案优选或有利。

[0091] 术语“互补”将被赋予其对于本领域普通技术人员来说普通且惯用的含义,并且非限制性地用于描述能够彼此杂交的核苷酸碱基之间的关系。例如,对于DNA,腺苷与胸腺嘧啶互补,并且胞嘧啶与鸟嘌呤互补。因此,本主题技术还包括与所附序列列表中报告的完整序列互补的分离的核酸片段以及那些基本上相似的核酸序列。

[0092] 术语“核酸”和“核苷酸”将被赋予其相应的对于本领域普通技术人员来说普通且惯用的含义,并且非限制性地用于指单链或双链形式的脱氧核糖核苷酸或核糖核苷酸及其聚合物。除非特别限定,否则该术语涵盖含有天然核苷酸的已知类似物的核酸,所述类似物具有与参考核酸相似的结合特性并且以与天然核苷酸类似的方式被代谢。除非另有说明,否则特定的核酸序列也隐含地涵盖其保守修饰或简并的变体(例如简并密码子取代)和互补序列,以及明确指出的序列。

[0093] 术语“分离的”将被赋予其对于本领域普通技术人员来说普通且惯用的含义,并且当用于分离的核酸或分离的多肽的上下文中时,非限制性地用于指人为地存在于其天然环境之外并且因此不是自然产物的核酸或多肽。分离的核酸或多肽可以纯化的形式存在,或者可存在于非天然环境中,例如存在于转基因宿主细胞中。

[0094] 如本文所用,术语“温育”意指将两种或更多种化学或生物实体(例如化合物和酶)混合并且允许它们在有利于产生甜菊糖苷组合物的条件下相互作用的过程。

[0095] 术语“简并变体”是指具有通过一个或多个简并密码子取代而与参考核酸序列不同的残基序列的核酸序列。简并密码子取代可通过生成其中一个或多个所选择的(或全部)

密码子的第三位置被混合的碱基和/或脱氧肌苷残基取代的序列来实现。核酸序列及其所有简并变体将表达相同的氨基酸或多肽。

[0096] 术语“多肽”、“蛋白质”和“肽”将被赋予其相应的对于本领域普通技术人员来说普通且惯用的含义；这三个术语有时可互换使用，并且无论其大小或功能如何，都非限制地用于指氨基酸或氨基酸类似物的聚合物。尽管“蛋白质”通常用于指相对大的多肽，而“肽”通常用于指小的多肽，但这些术语在本领域中的使用是重叠且变化的。除非另有说明，否则本文所用的术语“多肽”是指肽、多肽和蛋白质。当涉及多核苷酸产物时，术语“蛋白质”、“多肽”和“肽”在本文可互换使用。因此，示例性多肽包括多核苷酸产物、天然存在的蛋白质、同源物、直系同源物、旁系同源物、片段和前述的其他等同物、变体和类似物。

[0097] 术语“多肽片段”和“片段”在参考多肽使用时，将被赋予其对于本领域普通技术人员来说普通且惯用的含义，并且非限制性地用于指与参考多肽本身相比缺失氨基酸残基但是其中剩余的氨基酸序列通常与参考多肽中的相应位置相同的多肽。此类缺失可发生在参考多肽的氨基末端或羧基末端，或两者。

[0098] 术语多肽或蛋白质的“功能片段”是指作为全长多肽或蛋白质的一部分并且具有与全长多肽或蛋白质基本上相同的生物活性，或执行与全长多肽或蛋白质基本上相同的功能（例如，执行相同的酶反应）的肽片段。

[0099] 可互换使用的术语“变体多肽”、“经修饰的氨基酸序列”或“经修饰的多肽”是指与参考多肽相差一个或多个氨基酸（例如相差一个或多个氨基酸取代、缺失和/或添加）的氨基酸序列。在一个方面，变体是保留了参考多肽的一些或全部能力的“功能变体”。

[0100] 术语“功能变体”还包括保守取代的变体。术语“保守取代的变体”是指具有与参考肽相差一个或多个保守氨基酸取代的氨基酸序列并且保持参考肽的一些或全部活性的肽。“保守氨基酸取代”是用功能相似的残基取代氨基酸残基。保守取代的示例包括一个非极性（疏水）残基诸如异亮氨酸、缬氨酸、亮氨酸或甲硫氨酸取代另一个；一个带电或极性（亲水）残基取代另一个，诸如在精氨酸和赖氨酸之间、在谷氨酰胺和天冬酰胺之间、在苏氨酸和丝氨酸之间；一个碱性残基诸如赖氨酸或精氨酸取代另一个；或一个酸性残基诸如天冬氨酸或谷氨酸取代另一个；或一个芳族残基诸如苯丙氨酸、酪氨酸或色氨酸取代另一个。预期这种取代对蛋白质或多肽的表观分子量或等电点几乎没有影响。短语“保守取代的变体”还包括其中残基被化学衍生的残基替代的肽，条件是所得肽保持如本文所述的参考肽的一些或全部活性。

[0101] 与本主题技术的多肽相关的术语“变体”还包括具有与参考多肽的氨基酸序列至少75%、至少76%、至少77%、至少78%、至少79%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%和甚至100%相同的氨基酸序列的功能活性多肽。

[0102] 术语“同源的”在其所有语法形式和拼写变化中是指具有“共同进化起源”的多核苷酸或多肽之间的关系，包括来自超家族的多核苷酸或多肽和来自不同物种的同源多核苷酸或蛋白质（Reeck等人，*C_{ELL}* 50:667, 1987）。此类多核苷酸或多肽具有序列同源性，如其序列相似性所反映的，无论是在同一性百分比方面还是在保守位置处存在特定氨基酸或基序方面。例如，两个同源多肽可具有至少75%、至少76%、至少77%、至少78%、至少79%、至少

80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%和甚至100%相同的氨基酸序列。

[0103] “合适的调节序列”将被赋予其对于本领域普通技术人员来说普通且惯用的含义,并且非限制性地用于指位于编码序列上游(5'非编码序列)、内部、或下游(3'非编码序列)的核苷酸序列,并且其影响相关编码序列的转录、RNA加工或稳定性或翻译。调节序列可包括启动子、翻译前导序列、内含子和多腺苷酸化识别序列。

[0104] “启动子”将被赋予其对于本领域普通技术人员来说普通且惯用的含义,并且非限制性地用于指能够控制编码序列或功能性RNA表达的DNA序列。通常,编码序列位于启动子序列的3'。启动子可全部源自天然基因,或者由源自自然界中发现的不同启动子的不同元件组成,或者甚至包含合成的DNA片段。本领域技术人员应理解,不同的启动子可指导基因在不同组织或细胞类型中、或在不同发育阶段、或响应于不同环境条件的表达。导致基因在大多数细胞类型中在大多数时间表达的启动子通常被称为“组成型启动子”。还认识到,由于在大多数情况下调节序列的精确边界尚未完全确定,因此不同长度的DNA片段可能具有相同的启动子活性。

[0105] 术语“可操作地连接”是指核酸序列在单个核酸片段上的缔合,使得一个的功能受另一个影响。例如,当启动子可影响编码序列的表达时(即,编码序列在启动子的转录控制下),启动子与该编码序列可操作地连接。编码序列可以有义或反义方向与调节序列可操作地连接。

[0106] 如本文所用,术语“表达”将被赋予其对于本领域普通技术人员来说普通且惯用的含义,并且非限制性地用于指源自本主题技术的核酸片段的有义(mRNA)或反义RNA的转录和稳定积累。“过表达”是指在转基因或重组生物体中产生的基因产物超过了正常或非转化生物体中的产生水平。

[0107] “转化”将被赋予其对于本领域技术人员来说普通且惯用的含义,并且非限制性地用于指将多核苷酸转移到靶细胞中。转移的多核苷酸可被结合到靶细胞的基因组或染色体DNA中,从而产生遗传稳定的遗传,或者其可独立于宿主染色体进行复制。含有转化的核酸片段的宿主生物被称为“转基因的”或“转化的”。

[0108] 当在本文中与宿主细胞结合使用时,术语“转化的”、“转基因的”和“重组的”将被赋予其相应的对于本领域普通技术人员来说普通且惯用的含义,并且非限制性地用于指已经向其中引入异源核酸分子的宿主生物体的细胞,诸如植物或微生物细胞。核酸分子可被稳定地整合到宿主细胞的基因组中,或者核酸分子可作为染色体外分子存在。此类染色体外分子可自我复制。转化的细胞、组织或受试者应被理解为不仅涵盖转化过程的终产物,而且涵盖其转基因后代。

[0109] 当在本文中与多核苷酸结合使用时,术语“重组的”、“异源的”和“外源的”将被赋予其对于本领域普通技术人员来说普通且惯用的含义,并且非限制性地用于指源自对特定宿主细胞为外源的来源或者如果源自相同来源则相对于其原始形式被修饰的多核苷酸(例如,DNA序列或基因)。因此,宿主细胞中的异源基因包括对特定宿主细胞为内源的但已经通过例如使用定点诱变或其他重组技术进行修饰的基因。该术语还包括天然存在的DNA序列的非天然存在的多个拷贝。因此,该术语指对细胞为外源或异源的或对细胞为同源的但在

宿主细胞内通常不存在该元件的位置或形式的DNA片段。

[0110] 类似地,当在本文中与多肽或氨基酸序列结合使用时,术语“重组的”、“异源的”和“外源的”是指源自对特定宿主细胞为外源的来源或者如果源自相同来源则相对于其原始形式被修饰的多肽或氨基酸序列。因此,重组DNA片段可在宿主细胞中表达以产生重组多肽。

[0111] 术语“质粒”、“载体”和“盒”将被赋予其相应的对于本领域普通技术人员来说普通且惯用的含义,并且非限制性地用于指通常携带不是细胞中心代谢的一部分的基因并且通常为环状双链DNA分子形式的染色体外元件。此类元件可以是源自任何来源的单链或双链DNA或RNA的线性或环状的自主复制序列、基因组整合序列、噬菌体或核苷酸序列,其中许多核苷酸序列已被连接或重组成独特的构建体,该构建体能够将所选择的基因产物的启动子片段和DNA序列以及适当的3'非翻译序列引入细胞中。“转化盒”是指含有外源基因并且除外源基因外还具有促进特定宿主细胞转化的元件的特定载体。“表达盒”是指含有外源基因并且除外源基因外还具有允许该基因在外源宿主中增强表达的元件的特定载体。

[0112] 在一些实施方案中,本公开涉及使用本文所述的至少一种新型UDP-鼠李糖基转移酶(RhaT)从Reb A产生感兴趣的甜菊糖苷Reb N,所述新型UDP-鼠李糖基转移酶可在鼠李糖基化反应中将鼠李糖部分从UDP-L-鼠李糖转移至甜菊糖苷受体。参考图3,合成途径可涉及莱苞迪昔J作为中间体(图3A)或莱苞迪昔I作为中间体(图3B)。在一些实施方案中,本主题技术提供了具有UDP糖基转移酶活性的重组多肽以用于合成甜菊糖苷。在一些实施方案中,重组多肽可具有1,2-19-0-鼠李糖糖基化活性。在一些实施方案中,重组多肽可具有1,3-19-0-葡萄糖糖基化活性。本主题技术的重组多肽可用于甜菊糖苷化合物的生物合成。在本公开中,UDP-鼠李糖基转移酶(Rha T)是指将鼠李糖糖部分从活化的供体分子(通常为UDP-L-鼠李糖)转移至受体甜菊糖苷分子的酶。1,2 Rha T是指将鼠李糖部分从UDP-L-鼠李糖转移至甜菊糖苷的19-0-葡萄糖或13-0-葡萄糖的C-2'的酶活性。1,2-19-0-鼠李糖糖基化活性是指将鼠李糖部分转移至甜菊糖苷诸如莱苞迪昔A(以产生莱苞迪昔J)或莱苞迪昔I(以产生莱苞迪昔N)的19-0葡萄糖部分的C-2'的酶活性。1,3-19-0-葡萄糖糖基化活性是指将葡萄糖部分转移至甜菊糖苷诸如莱苞迪昔J(以产生莱苞迪昔N)或莱苞迪昔A(以产生莱苞迪昔I)的19-0葡萄糖部分的C-3'的酶活性(图3)。

[0113] 合成生物学

[0114] 本文所用的标准重组DNA和分子克隆技术是本领域众所周知的,并且例如如以下文献中所述:Sambrook,J.、Fritsch,E.F.和Maniatis,T.,*MOLECULAR CLONING: A LABORATORY MANUAL*,第2版;Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y., 1989(以下称为“Maniatis”);和Silhavy,T.J.、Bennan,M.L.和Enquist,L.W.,*EXPERIMENTAL GENETICS WITH BACTERIAL GENES*; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y., 1984;以及Ausubel,F.M.等人,*CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*,由Greene Publishing和Wiley-Interscience于1987年出版;(这些文献中的每一个的全部内容以引用方式并入本文)。

[0115] 除非另外限定,否则本文所用的所有技术和科学术语均具有与本公开所属领域的普通技术人员通常理解的含义相同的含义。尽管在本公开的实践或测试中可使用与本文所述的那些方法和材料相似或等同的任何方法和材料,但下文描述了优选的方法和材料。

[0116] 通过考虑以下非限制性实施例,将更充分地理解本公开。应当理解,这些实施例虽

然指示了本主题技术的优选实施方案,但仅以说明的方式给出。根据以上讨论和这些实施例,本领域技术人员可确定本主题技术的基本特征,并且在不脱离其精神和范围的情况下,可对本主题技术进行各种改变和修改以使其适应各种用途和技术。

[0117] 糖基化通常被认为是控制植物天然产物的生物活性和储存的普遍存在的反应。小分子的糖基化由迄今为止已研究的大多数植物物种中的转移酶超家族催化。这些糖基转移酶(GTs)已被分成60多个家族。其中,GT酶家族(也称为UDP糖基转移酶(UGT)和UDP-鼠李糖基转移酶)将糖部分转移至特定的受体分子。这些是在甜菊糖苷中转移此类糖部分以帮助产生各种莱菔迪苷的分子。这些酶中的每一种都具有其自身的活性特征和优选的它们转移其活化糖部分的结构位置。

[0118] 生产系统

[0119] 原核生物中蛋白质的表达最经常在细菌宿主细胞中用含有指导融合或非融合蛋白质表达的组成型或诱导型启动子的载体进行。融合载体将许多氨基酸添加到其中编码的蛋白质中,通常添加到重组蛋白质的氨基末端。此类融合载体通常用于三个目的:1)增加重组蛋白的表达;2)增加重组蛋白的溶解度;以及3)通过在亲和纯化中作为配体来帮助重组蛋白的纯化。通常,在融合部分和重组蛋白的连接处引入蛋白水解切割位点,以使得重组蛋白能够在融合蛋白纯化后与融合部分分离。此类载体在本公开的范围内。

[0120] 在一个实施方案中,表达载体包括用于在细菌细胞中表达重组多肽的那些遗传元件。用于在细菌细胞中转录和翻译的元件可包括启动子、蛋白质复合物的编码区和转录终止子。

[0121] 本领域普通技术人员将知道可用于制备表达载体的分子生物学技术。如上所述,用于结合到本主题技术的表达载体中的多核苷酸可通过常规技术诸如聚合酶链式反应(PCR)来制备。

[0122] 已经开发了几种分子生物学技术通过互补粘性末端将DNA可操作地连接到载体上。在一个实施方案中,可将互补均聚物束添加到待插入载体DNA中的核酸分子上。然后载体和核酸分子通过互补均聚物末端之间的氢键连接以形成重组DNA分子。

[0123] 在另选的实施方案中,使用含有一个或多个限制性位点的合成接头将本主题技术的多核苷酸可操作地连接至表达载体。在一个实施方案中,多核苷酸通过限制性内切核酸酶消化生成。在一个实施方案中,将核酸分子用噬菌体T4 DNA聚合酶或大肠杆菌DNA聚合酶I处理,所述酶利用其3'-5'-核酸外切活性去除突出的3'-单链末端,并利用其聚合活性填充凹陷的3'-末端,从而生成平末端DNA片段。然后在可催化平末端DNA分子连接的酶诸如噬菌体T4 DNA连接酶的存在下,将平末端片段与大摩尔过量的接头分子一起温育。因此,反应产物是在其末端携带聚合物接头序列的多核苷酸。然后将这些多核苷酸用适当的限制酶切割并连接到表达载体,该表达载体已经用产生与该多核苷酸的末端相容的末端的酶切割。

[0124] 另选地,可使用具有连接非依赖性克隆(LIC)位点的载体。然后可以将所需的PCR扩增的多核苷酸克隆到LIC载体中,而无需进行限制性消化或连接(Aslanidis和de Jong, *N_{UCL}.A_{CID}.R_{ES}* 18 6069-74, (1990), Haun等人, *B_{IOTECHNIQUES}* 13, 515-18 (1992), 其以引用方式并入本文)。

[0125] 在一个实施方案中,为了分离和/或修饰用于插入到所选质粒中的感兴趣的多核苷酸,适合使用PCR。可设计用于序列PCR制备的合适引物,以分离核酸分子的所需编码区、

添加限制性内切核酸酶或LIC位点、将编码区置于所需阅读框中。

[0126] 在一个实施方案中,使用PCR,使用适当的寡核苷酸引物制备用于结合到本主题技术的表达载体中的多核苷酸。扩增编码区,同时引物本身被结合到扩增的序列产物中。在一个实施方案中,扩增引物包含限制性核酸内切酶识别位点,其允许将扩增的序列产物克隆到合适的载体中。

[0127] 可通过常规转化或转染技术将表达载体引入植物或微生物宿主细胞。用本主题技术的表达载体转化合适的细胞是通过本领域已知的方法完成的,并且通常取决于载体和细胞的类型。合适的技术包括磷酸钙或氯化钙共沉淀、DEAE-葡聚糖介导的转染、脂转染、化学穿孔或电穿孔。

[0128] 可通过本领域众所周知的技术鉴定成功转化的细胞,即含有表达载体的那些细胞。例如,可培养用本主题技术的表达载体转染的细胞以产生本文所述的多肽。可通过本领域众所周知的技术检查细胞中表达载体DNA的存在。

[0129] 宿主细胞可含有先前描述的表达载体的单拷贝,或另选地含有表达载体的多拷贝,

[0130] 在一些实施方案中,转化的细胞为动物细胞、昆虫细胞、植物细胞、藻类细胞、真菌细胞或酵母细胞。在一些实施方案中,细胞为选自以下组成的组的植物细胞:低芥酸菜子植物细胞、油菜籽植物细胞、棕榈植物细胞、向日葵植物细胞、棉花植物细胞、玉米植物细胞、花生植物细胞、亚麻植物细胞、芝麻植物细胞、大豆植物细胞和矮牵牛植物细胞。

[0131] 含有指导外源蛋白高水平表达的调节序列的微生物宿主细胞表达系统和表达载体是本领域技术人员众所周知的。这些中的任何一种都可用于构建用于在微生物宿主细胞中表达本主题技术的重组多肽的载体。然后可将这些载体经由转化引入合适的微生物中,以允许本主题技术的重组多肽的高水平表达。

[0132] 用于转化合适的微生物宿主细胞的载体或盒是本领域众所周知的。通常,载体或盒包含指导相关多核苷酸转录和翻译的序列、选择性标记和允许自主复制或染色体整合的序列。合适的载体包括具有转录起始控制的多核苷酸的5'区域和控制转录终止的DNA片段的3'区域。优选两个控制区都源自与转化的宿主细胞同源的基因,尽管应理解此类控制区不需要源自被选作宿主的特定物种的天然基因。

[0133] 用于驱动重组多肽在期望微生物宿主细胞中表达的起始控制区或启动子是众多的,并且是本领域技术人员所熟悉的。实际上,能够驱动这些基因的任何启动子都适用于本主题技术,包括但不限于CYCI、HIS3、GALI、GALIO、ADHI、PGK、PH05、GAPDH、ADCI、TRPI、URA3、LEU2、ENO、TPI(用于在酵母属中表达);AOXI(用于在毕赤酵母属中表达);以及lac、trp、JPL、IPR、T7、tac和trc(用于在大肠杆菌中表达)。

[0134] 终止控制区也可源自微生物宿主天然的各种基因。对于本文所述的微生物宿主,可任选地包括终止位点。

[0135] 在植物细胞中,本主题技术的表达载体可包括与启动子可操作地连接的编码区,所述启动子能够在期望的发育阶段在期望的组织中指导本主题技术的重组多肽的表达。为了方便起见,待表达的多核苷酸可包含源自相同多核苷酸的启动子序列和翻译前导序列。还应该存在编码转录终止信号的3'非编码序列。表达载体还可包含一个或多个内含子以促进多核苷酸表达。

[0136] 对于植物宿主细胞,可在本主题技术的载体序列中使用能够诱导编码区表达的任何启动子和任何终止子的任何组合。启动子和终止子的一些合适的示例包括来自胭脂碱合酶(nos)、章鱼碱合酶(ocs)和花椰菜花叶病毒(CaMV)基因的那些。可使用的一种类型的有效植物启动子是高水平植物启动子。与本主题技术的表达载体可操作连接的此类启动子应能够促进载体的表达。可用于本主题技术的高水平植物启动子包括例如来自大豆的核酮糖-1,5-二磷酸羧化酶的小亚基(ss)的启动子(Berry-Lowe等人,J.MOLECULAR AND APPLIED GENETICS,1:483-498(1982),其整体据此以与本文一致的程度并入本文)和叶绿素alb结合蛋白的启动子。已知这两种启动子在植物细胞中是光诱导的(参见,例如,GENETIC ENGINEERING OF PLANTS,AN AGRICULTURAL PERSPECTIVE,A.Cashmore,Plenum,N.Y.(1983),第29-38页;Coruzzi,G.等人,The Journal of Biological CHEMISTRY,258:1399(1983),以及Dunsmuir,P.等人,JOURNAL OF MOLECULAR AND APPLIED GENETICS,2:285(1983),这些文献中的每个据此以与本文一致的程度以引用方式并入本文)。

[0137] Reb J或Reb I的前体合成

[0138] 如前所述,甜菊糖苷是负责南美植物甜叶菊(菊科)和植物掌叶覆盆子(Rubus chingii)(蔷薇科(Rosaceae))的甜味的化合物。这些化合物是糖基化的二萜。具体地,它们的分子可被视为甜菊醇分子,其中其羟基氢原子被葡萄糖分子替代以形成酯,并且羟基氢与葡萄糖和鼠李糖的组合形成缩醛。

[0139] 制备本公开中的感兴趣的化合物的一种方法是采用化学衍生或经由工程化微生物诸如细菌和/或酵母中的生物合成产生的常见或廉价的前体诸如甜菊醇、甜菊苷、Reb A或甜茶苷,并通过已知或廉价的方法合成目标甜菊糖苷诸如Reb J或Reb I。

[0140] 本公开的方面涉及在能够产生甜菊醇的微生物系统中涉及重组表达酶的方法。通常,此类酶可包括:柯巴基二磷酸合酶(CPS)、贝壳杉烯合酶(KS)和香叶基香叶基二磷酸合酶(GGPPS)。这应该发生在表达内源性类异戊二烯合成途径诸如非甲羟戊酸(MEP)途径或甲羟戊酸途径(MVA)的微生物菌株中。在一些实施方案中,细胞为细菌细胞,包括大肠杆菌或酵母细胞(诸如酵母属细胞、毕赤酵母属细胞或耶氏酵母属(Yarrowia)细胞)。在一些实施方案中,细胞为藻类细胞或植物细胞。

[0141] 此后,从发酵培养物中回收前体以用于化学合成。通常,这是甜菊醇,尽管其可以是贝壳杉烯,或来自细胞培养物的甜菊糖苷。在一些实施方案中,从气相中回收甜菊醇、贝壳杉烯和/或甜菊糖苷,而在其他实施方案中,将有机层或聚合树脂添加到细胞培养物中,然后从有机层或聚合树脂中回收贝壳杉烯、甜菊醇和/或甜菊糖苷。在一些实施方案中,甜菊糖苷选自莱苞迪苷A、莱苞迪苷B、莱苞迪苷C、莱苞迪苷N、莱苞迪苷E、莱苞迪苷F、莱苞迪苷J或杜尔可苷A。在一些实施方案中,产生的萜类化合物为甜菊双糖苷或甜菊苷。还应当理解,在一些实施方案中,至少一个酶促步骤诸如一个或多个糖基化步骤离体进行。

[0142] 本公开的一部分为Reb J甜菊糖苷的产生,其可进一步酶促转化为Reb N。在一些实施方案中,可使用如本文所述的鼠李糖基转移酶(RhaT)(例如EU11、EUCP1、HV1、UGT2E-B或NX114)和鼠李糖供体部分(例如UDP-鼠李糖)将Reb A转化为Reb J。在一些实施方案中,当利用由发明人产生的特异性地鉴定和/或修饰的酶,使用糖部分向甜菊醇主链中的多步骤化学组装,将二萜类甜菊醇转化为甜茶苷和甜菊苷时,发生微生物产生的甜菊醇向期望的甜菊糖苷(例如Reb N)转化的生物合成。除了本文所用的EU11、EUCP1、HV1和UGT76G1酶之

外,还鉴定了也可用来递送这些甜菊醇莱苞迪苷的其他酶。例如,已确定其他UGT酶(分别为CP1和CP2—SEQ ID NO:11和SEQ ID NO:13的氨基酸序列)和UGT76G1-AtSUS1融合酶可将Reb J转化为Reb N。

[0143] 本公开的一部分为从Reb A产生Reb I甜菊糖苷,其可进一步酶促转化为Reb N。在一些实施方案中,通过在具有葡糖基转移酶活性的重组多肽的存在下使Reb A与葡萄糖供体部分反应来进行Reb A向Reb I转化的生物合成。在一些实施方案中,葡萄糖供体部分为原位生成的。在一些实施方案中,将葡萄糖供体部分添加到反应中。在一些实施方案中,具有葡糖基转移酶活性的重组多肽还包含蔗糖合酶活性。例如,在一些实施方案中,鉴定为UGT76G1(SEQ ID NO:7)的酶可将Reb A转化为Reb I。在一些实施方案中,其他UGT酶(例如,分别为CP1和CP2—SEQ ID NO:11和SEQ ID NO:13的氨基酸序列)可将Reb A转化为Reb I。在一些实施方案中,UCT76G1-AtSUS1融合酶可将Reb A转化为Reb I。在一些实施方案中,可使用如本文所述的鼠李糖基转移酶(RhaT)例如EU11、EUCP1 UDP、HV1、UGT2E-B或NX114以及鼠李糖供体部分诸如UDP-鼠李糖将Reb I转化为Reb N。

[0144] 甜菊糖苷的生物合成

[0145] 如本文所述,本发明技术的重组多肽具有UDP-糖基转移酶(UDP-葡糖基转移酶和/或UDP-鼠李糖基转移酶)活性,并且可用于开发用于制备在自然界中不存在或通常在天然来源中具有低丰度的甜菊糖苷分别诸如莱苞迪苷J、莱苞迪苷I和莱苞迪苷N的生物合成方法。本发明技术的重组多肽具有UDP-糖基转移酶活性、可用于开发用于制备甜菊糖苷诸如莱苞迪苷J或莱苞迪苷I的生物合成方法并实现莱苞迪苷N的合成生产。

[0146] 底物或起始甜菊糖苷组合物可以是能够在由一种或多种UDP-鼠李糖基转移酶催化的反应中转化为甜菊糖苷化合物的任何天然或合成化合物。例如,底物可以是天然甜叶菊提取物、甜菊醇、甜菊醇-13-O-葡糖苷、甜菊醇-19-O-葡糖苷、甜菊醇-1,2-双糖苷、甜茶苷、甜菊苷、莱苞迪苷A、莱苞迪苷B或莱苞迪苷I。底物可以是纯化合物或不同化合物的混合物。优选地,底物包括选自以下组成的组的化合物:甜茶苷、甜菊苷、甜菊醇、莱苞迪苷A、莱苞迪苷B、莱苞迪苷J以及它们的组合。

[0147] 本文所述的方法还提供了偶联反应系统,其中本文所述的重组肽可与一种或多种另外的酶组合起作用以改善甜菊糖苷化合物的总体生物合成的效率或改变其结果。例如,该另外的酶可再生鼠李糖基化反应所需的UDP-鼠李糖(参见,例如Pei等人,“Construction of a novel UDP-rhamnose regeneration system by a two-enzyme reaction system and application in glycosylation of flavonoid,”*Biochemical Engineering Journal*,139:33-42(2018),其全部公开内容以引用方式并入本文),并且该另外的酶可通过将由糖基化反应产生的UDP转化回UDP-葡萄糖(例如使用蔗糖作为葡萄糖残基的供体)来再生糖基化反应所需的UDP-葡萄糖,从而提高糖基化反应的效率。

[0148] 在另一个实施方案中,本主题技术的方法还包括将根据本公开的重组和新型UDP-鼠李糖基转移酶(RhaT)与底物和本文所述的一种或多种另外的重组多肽(例如,重组UGT)一起温育。该重组UGT可催化与本主题技术的重组多肽所催化的糖基化反应不同的糖基化反应,从而导致Reb J和Reb N的产生。

[0149] 合适的UDP-糖基转移酶包括本领域已知的能够催化甜菊糖苷化合物生物合成中的一个或多个反应的任何UGT,诸如UGT85C2、UGT74G1、UGT76G1或其功能同源物。

[0150] 在一些实施方案中,在本主题技术的体外方法中,UDP-葡萄糖和/或UDP-L-鼠李糖可以约0.2mM至约5mM、优选约0.5mM至约2mM、更优选约0.7mM至约1.5mM的浓度包括在缓冲液中。在一个实施方案中,当反应中包括重组蔗糖合酶时,蔗糖也以约100mM至约500mM、优选约200mM至约400mM、更优选约250mM至约350mM的浓度包括在缓冲液中。

[0151] 在一些实施方案中,在本主题技术的体外方法中,基于干重计,重组多肽与底物的重量比为约1:100至约1:5、优选约1:50至约1:10、更优选约1:25至约1:15。

[0152] 在一些实施方案中,体外方法的反应温度为约20°C至约40°C,合适地为25°C至约37°C。

[0153] 本领域技术人员将认识到,可将通过本文所述方法产生的甜菊糖苷组合物进一步纯化并与其他甜菊糖苷、风味剂或甜味剂混合以获得期望的风味剂或甜味剂组合物。例如,如本文所述产生的富含Reb J或Reb N的组合物可与含有莱菔迪苷A作为主要甜菊糖苷的天然甜叶菊提取物混合,或与其他合成或天然甜菊糖苷产品混合,以制备期望的甜味剂组合物。另选地,由本文所述方法获得的基本上纯化的甜菊糖苷(例如莱菔迪苷J和莱菔迪苷N)可与其他甜味剂诸如蔗糖、麦芽糖糊精、阿斯巴甜、三氯蔗糖、纽甜、乙酰舒泛钾和糖精组合。如本领域已知的,可调节甜菊糖苷相对于其他甜味剂的量以获得期望的味道。本文所述的甜菊糖苷(包括莱菔迪苷N、莱菔迪苷I、莱菔迪苷J或它们的组合)可包括在食品(诸如饮料、软饮料、冰淇淋、乳制品、糖果、谷物、口香糖、烘焙食品等)、膳食补充剂、医疗营养品以及药物产品中。

[0154] 使用同一性评分分析序列相似性

[0155] 如本文所用,“序列同一性”是指两个最佳比对的多核苷酸或肽序列在组分(例如,核苷酸或氨基酸)的整个比对窗口中不变的程度。测试序列和参考序列的比对片段的“同一性分数”是由两个比对序列共享的相同组分的数目除以参考序列片段中的组分总数,即整个参考序列或参考序列的较小的限定部分。

[0156] 如本文所用,术语“百分比序列同一性”或“百分比同一性”是指当两个序列最佳比对时(在比较窗口范围内具有总计小于20%的参考序列的适当核苷酸插入、缺失或缺口),与测试(“受试者”)多核苷酸分子(或其互补链)相比,参考(“查询”)多核苷酸分子(或其互补链)的线性多核苷酸序列中相同核苷酸的百分比。用于比对比较窗口的序列的最佳比对是本领域技术人员所熟知的,并且可通过诸如以下工具来进行:Smith和Waterman的局部同源性算法、Needleman和Wunsch的同源比对算法、Pearson和Lipman的搜索相似性方法,并且优选地通过这些算法的计算机实施来进行,诸如作为GCG®Wisconsin Package®的一部分获得的GAP、BESTFIT、FASTA和TFASTA(Accelrys Inc., Burlington, MA)。测试序列和参考序列的比对片段的“同一性分数”是由两个比对序列共享的相同组分的数目除以参考序列片段中的组分总数,即整个参考序列或参考序列的较小的限定部分。百分比序列同一性表示为同一性分数乘以100。一个或多个多核苷酸序列的比较可涉及全长多核苷酸序列或其部分,或更长的多核苷酸序列。出于本公开的目的,还可使用BLASTX版本2.0(对于翻译的核苷酸序列)和BLASTN版本2.0(对于多核苷酸序列)来测定“百分比同一性”。

[0157] 序列同一性百分比优选使用Sequence Analysis Software Package™的“最佳拟合”或“空位”程序(版本10;Genetics Computer Group, Inc., Madison, WI)来确定。“空位”利用Needleman和Wunsch的算法(Needleman和Wunsch, JOURNAL OF Molecular Biology 48:

443-453,1970) 来寻找使匹配数最大化并使空位数最小化的两个序列的比对。“最佳拟合”使用Smith和Waterman的局部同源性算法 (Smith和Waterman, *ADVANCES IN APPLIED MATHEMATICS*, 2: 482-489,1981,Smith等人, *NUCLEIC ACIDS RESEARCH* 11:2205-2220,1983) 来进行两个序列之间最佳片段相似性的最佳比对,并插入空位以使匹配数最大化。百分比同一性最优选使用“最佳拟合”程序来确定。

[0158] 用于确定序列同一性的可用方法还公开于局部序列排比检索基本工具 (BLAST) 程序中,所述程序可从National Library of Medicine,National Institute of Health, Bethesda,Md.20894的National Center Biotechnology Information (NCBI) 公开获得;参见BLAST Manual,Altschul等人,NCBI,NLM,NIH;Altschul等人, *J.MOL.BIOL.* 215:403-410 (1990);BLAST程序的2.0版或更高版本允许将空位(缺失和插入)引入比对中;对于肽序列,BLASTX可用于确定序列同一性;并且,对于多核苷酸序列,BLASTN可用于确定序列同一性。

[0159] 如本文所用,术语“显著百分比序列同一性”是指至少约70%序列同一性、至少约80%序列同一性、至少约85%同一性、至少约90%序列同一性、或甚至更大的序列同一性,诸如约98%或约99%序列同一性的百分比序列同一性。因此,本公开的一个实施方案是多核苷酸分子,其与本文所述的多核苷酸序列具有至少约70%的序列同一性,至少约80%的序列同一性,至少约85%的同一性,至少约90%的序列同一性,或甚至更大的序列同一性,诸如约98%或约99%的序列同一性。具有本公开的1,2 RhaT和UGT酶的活性的多核苷酸分子能够引导多种甜菊糖苷的产生,并且与本文提供的多核苷酸序列具有显著百分比序列同一性,并且包括在本公开的范围之内。

[0160] 同一性和相似性

[0161] 同一性是在比对序列(其可仅使用序列信息或结构信息或一些其他信息进行,但通常其仅基于序列信息)后在一对序列之间相同的氨基酸的分数,而相似性是基于使用一些相似性矩阵的比对指定的评分。相似性指数可以为以下BLOSUM62、PAM250或GONNET中的任一者,或者本领域技术人员进行蛋白质序列比对所用的任何矩阵。

[0162] 同一性是两个子序列之间一致性的程度(序列之间没有空位)。25%或更高的同一性意指功能的相似性,然而18%至25%意指结构或功能的相似性。请记住,两个完全不相关或随机序列(其为多于100个残基)可具有高于20%的同一性。相似性是当比较两个序列时,该两个序列之间相似的程度。这取决于它们的同一性。

[0163] 如从上述说明显而易见,本公开的某些方面不受本文说明的实施例的具体细节限制,并且因此考虑到本领域技术人员将进行其他修改和应用或它们的等效形式。因此,权利要求应当旨在覆盖所有不偏离本公开的实质和范围的此类修改和应用。

[0164] 此外,除非另外限定,否则本文所用的所有技术和科学术语均具有与本公开所属领域的普通技术人员通常理解的含义相同的含义。尽管在本公开的实践或测试中可使用与本文所述的那些方法和材料相似或等同的任何方法和材料,但上文描述了优选的方法和材料。

[0165] 尽管出于理解的目的已经以举例说明和实施例的方式对前述发明进行了详细描述,但对于本领域的技术人员将显而易见的是,可实践某些改变和修改。因此,说明书和实施例不应当被理解为对本发明的范围的限制,本发明的范围由所附权利要求书界定。

[0166] 实施例

[0167] 实施例1:1,2 RhaT酶的酶活性筛选

[0168] 使用系统发育、基因簇和蛋白质BLAST分析来鉴定候选1,2 RhaT转移酶基因。根据大肠杆菌(Genscript,NJ)的密码子偏好性来优化和合成候选1,2 RhaT转移酶基因的全长DNA片段。将合成的DNA片段克隆到细菌表达载体pETite N-His SUMO Kan载体(Lucigen)中。

[0169] 将每个表达构建体转化到大肠杆菌BL21(DE3)中,随后将其在含有50 μ g/mL卡那霉素的LB培养基中于37 $^{\circ}$ C生长,直到OD₆₀₀达到0.8-1.0。通过添加1mM的异丙基 β -D-1-硫代吡喃半乳糖苷(IPTG)诱导蛋白表达,并将培养物在16 $^{\circ}$ C下再温育22小时。离心收集细胞(3,000 \times g;10分钟;4 $^{\circ}$ C)。收集细胞团块并立即使用或保存在-80 $^{\circ}$ C。

[0170] 通常将细胞团块重悬于裂解缓冲液(50mM磷酸钾缓冲液,pH7.2,25 μ g/ml溶菌酶,5 μ g/ml DNase I,20mM咪唑,500mM NaCl,10%甘油和0.4%Triton X-100)。通过在4 $^{\circ}$ C下进行超声处理破碎细胞,并通过离心(18,000 \times g;30分钟)澄清细胞碎片。将上清液上样到平衡的(平衡缓冲液:50mM磷酸钾缓冲液,pH7.2,20mM咪唑,500mM NaCl,10%甘油)Ni-NTA(Qiagen)亲和柱中。蛋白质样品上样后,用平衡缓冲液洗涤柱以去除未结合的污染蛋白质。用含有250mM咪唑的平衡缓冲液洗脱His-标记的1,2 RhaT重组多肽。

[0171] 测定经纯化的候选1,2 RhaT重组多肽的UDP-鼠李糖基转移酶活性。通常,重组多肽(20-50 μ g)在200 μ l体外反应系统中进行测试。该反应系统含有50mM的磷酸钾缓冲液(pH7.2)、3mM的MgCl₂、0.25-1mg/ml的莱苞迪昔A和UDP-L-鼠李糖。在30-37 $^{\circ}$ C下进行反应并通过添加1-丁醇终止反应。

[0172] 将产物样品用200 μ L 1-丁醇萃取三次。将合并的级分干燥并溶于100 μ L80%甲醇中,以进行高效液相色谱(HPLC)分析。

[0173] 使用Dionex UPLC ultimate 3000系统(Sunnyvale,CA)进行HPLC分析,该系统包括四元泵、温度受控柱室、自动进样器和UV吸收检测器。使用具有保护柱的Synergi Hydro-RP柱(Phenomenex)表征合并样品中的甜菊糖苷。HPLC分析中使用的检测波长为210nm。

[0174] 活性筛选后,鉴定了几种用于Reb A到Reb J的生物转化的具有UDP-鼠李糖基转移酶活性的1,2 RhaT酶。

[0175] 实施例2:Reb A到Reb J的酶促生物转化

[0176] 莱苞迪昔J的生物合成涉及糖苷配基甜菊醇的葡糖基化和鼠李糖基化。具体地,Reb J可通过Reb A的19-O-葡萄糖的C-2'的鼠李糖基化即经由 α 1,2鼠李糖基化来产生(图3)。

[0177] 图4至图6比较了如通过HPLC分析的Reb A和Reb J标准品的保留时间与从由候选1,2 RhaT酶催化的鼠李糖基化反应中获得的反应产物的保留时间。

[0178] 据报道,EU11、EUCP1和HV1具有葡糖基化活性,并且能够将葡萄糖部分从UDP-葡萄糖转移至甜菊糖苷,从而产生化合物诸如甜菊苷、莱苞迪昔KA、莱苞迪昔E、莱苞迪昔D、莱苞迪昔V和莱苞迪昔D3等。

[0179] 如图4(C)-4(E)所示,Reb A在分别由EU11(SEQ ID NO:1)、本文命名为EUCP1(SEQ ID NO:3)的循环排列的EU11和HV1(SEQ ID NO:5)催化反应系统中被转化为Reb J,表明除了葡糖基化活性之外,这些酶还具有1,2-鼠李糖基转移酶活性。发现EUCP1比EU11具有更高的1,2鼠李糖基转移酶活性,而EU11比HV1显示出高得多的1,2鼠李糖基转移酶活性。

[0180] 图5显示Reb A在由UGT2E-B(SEQ ID NO:9)催化的反应系统中被转化为Reb J。图6显示Reb A在由NX114(SEQ ID NO:19)催化的反应系统中被转化为Reb J。

[0181] 总体而言,这些数据表明,酶EU11、EUCP1、HV1、UGT2E-B和NX114中的每一种均表现出1,2 RhaT活性,并且能够在 α 1,2鼠李糖基化反应中将鼠李糖部分从UDP-L-鼠李糖供体转移至Reb A的19-O-葡萄糖部分的C-2',从而产生Reb J。

[0182] 实施例3:通过LC-MS分析确认RebJ的产生

[0183] 为了确认在由1,2 RhaT候选酶催化的反应中产生的化合物的身份,使用Synergy Hydro-RP柱进行LC-MS分析。流动相A为0.1%甲酸的水溶液,流动相B为0.1%甲酸的乙腈溶液。流速为0.6ml/min。在Q Exactive混合四极-Orbitrap质谱仪(Thermo Fisher Scientific)上用优化的方法以阳离子模式进行样品的质谱分析。

[0184] LC-MS分析确认,从分别由EU11、EUCP1、HV1、UGT2E-B和NX114催化的鼠李糖基化反应中获得的产物具有与Reb J标准品相同的质量[(M+Na⁺) 1135.48m/z]和保留时间(2.61分钟)。

[0185] 实施例4:Reb J到Reb N的酶促生物转化

[0186] 再次参考图3,Reb N可通过在 β 1,3葡糖基化反应中Reb J的19-O-葡萄糖的C-3'的葡糖基化来产生。

[0187] 为了鉴定用于将Reb J转化为Reb N的合适的UGT,使用得自实施例1的Reb J产物来测定酶候选物。

[0188] 在通过加热(以使1,2 RhaT酶变性)终止Reb J产生后,将UGT候选酶(每200 μ l反应混合物5 μ g酶)和1mM UDP-葡萄糖添加到反应混合物中。将葡糖基化反应在37°C下进行3小时。

[0189] 图7比较了如通过HPLC分析的Reb O、Reb N、Reb C和杜尔可昔A("dA")标准品的保留时间与使用分别从实施例1的由EU11-催化(图7(B))和由EUCP1-催化(图7(C))的反应系统中获得的Reb J产物作为底物从由UGT76G1(SEQ ID NO:7)催化的葡糖基化反应中获得的反应产物的保留时间。如图7(B)和7(C)所示,在UGT76G1存在下Reb J被转化为Reb N。

[0190] 基于UGT76G1的结构,发明人能够设计CP1(SEQ ID NO:11)(其为UGT76G1的循环排列)、CP2(SEQ ID NO:13)(其为CP1的突变体并且通过在CP1的C-末端和N-末端之间插入接头生成)和融合酶GS(SEQ ID NO:15)(包括与蔗糖合酶(SUS)结构域AtSUS1偶联的UDP-糖基转移酶(UGT)结构域UGT76G1)。

[0191] 参考图8,图8(A)示出了Reb J标准品的保留时间。图8(B)确认了如由EUCP1催化的Reb A向Reb J的生物转化。使用从EUCP1催化的反应中获得的Reb J产物作为底物,测定UGT76G1、CP1、CP2和GS将Reb J转化为Reb N的葡糖基化活性。与显示Reb N标准品保留时间的图8(C)相比,图8(D)、8(E)、8(F)和8G表明UGT76G1、CP1、CP2和GS酶中的每一种都能够在 β 1,3葡糖基化反应中将葡萄糖部分从UDP-葡萄糖转移至Reb J的19-O-葡萄糖的C-3',从而产生Reb N。

[0192] 为了进一步评估UGT-SUS融合酶在Reb J到Reb N的生物转化中的活性,使用以下酶促系统在UDP和蔗糖存在但无UDP-葡萄糖存在下进行测定:(1)UGT76G1(SEQ ID NO:7)本身,(2)添加UDP-糖基转移酶UGT76G1(SEQ ID NO:7)和蔗糖合酶AtSUS1(SEQ ID NO:17)两者,以及(3)包括与蔗糖合酶(SUS)结构域AtSUS1偶联的UDP-糖基转移酶(UGT)结构域

UGT76G1的融合酶GS (SEQ ID NO:15),其具有UDP-糖基转移酶活性和蔗糖合酶活性两者。

[0193] 如图9 (C) 所示,当反应系统中仅存在UDP和蔗糖时,UGT76G1不能将Reb J转化为Reb N。另一方面,通过使用包括UGT酶 (例如UGT76G1) 和SUS酶 (例如AtSUS1) 两者的催化系统,SUS酶能够从UDP和蔗糖生成UDP葡萄糖,并且UGT酶能够将葡萄糖部分从UDP-葡萄糖转移至Reb J,从而产生Reb N (图9 (D))。类似地,融合酶GS能够从Reb J产生Reb N (图9 (E)),表明该融合酶具有蔗糖合酶活性并且能够从UDP和蔗糖生成UDP-葡萄糖。另外,与其中UGT76G1和AtSUS1作为单独的酶存在的反应系统相比,发现融合酶GS具有更高的酶活性 (产生更多的Reb N)。

[0194] 实施例5:通过LC-MS分析确认RebN的产生

[0195] 为了进一步确认在由UGT酶催化的反应中产生的化合物的身份,使用Synergy Hydro-RP柱进行LC-MS分析。流动相A为0.1%甲酸的水溶液,流动相B为0.1%甲酸的乙腈溶液。流速为0.6ml/min。在Q Exactive混合四极-Orbitrap质谱仪 (Thermo Fisher Scientific) 上用优化的方法以阳离子模式进行样品的质谱分析。

[0196] LC-MS分析确认,使用Reb J (由如实施例1中所述的由EU11/EUCP1/HV1/UGT2E-B/NX114催化的鼠李糖基化反应制得) 作为底物,从分别由UGT76G1、CP1、CP2和GS催化的葡萄糖基化反应中获得的产物具有与Reb N标准品相同的质量 [(M+Na⁺) 1297.53m/z] 和保留时间 (2.43分钟)。

[0197] 感兴趣的序列:

[0198] 序列:

[0199] EU11:氨基酸序列 (SEQ ID NO:1)

[0200] MDSGYSSSYAAAAGMHVVICPWLAFGHLLPCLDLAQRLASRGHRVSVFVSTPRNISRLPPVRPALAPLVA
FVALPLPRVEGLPDGAESTNDVPHDRPDMVELHRRAFDGLAAPFSEFLGTACADWVIVDVFHWA AAAALEHKVPCA
MMLLGSAHMIASIAARRLERAETESPAAGQGRPAAAPTFEVARMKLI RTKGSSGMSLAERFSLTLRSLLVVGRC
VEFEPETVPLLSTLRGKPI TFLGLMPPLHEGRREDGEDATVRWLDQPAKSVVYVALGSEVPLGVEKVHELALGLEL
AGTRFLWALRKPTGVSDADLLPAGFEERTRGRGVVATRWVPMQMSILAHAAVGAFLTHCGWNSTIEGLMFGHPLIMLP
IFGDQGPNARLIEAKNAGLQVARNDGDG SFDREGVAAAIRAVAVEEESSKVFQAKAKKLQEIVADMACHERYIDGFI
QQLRSYKD

[0201] EU11:DNA序列 (SEQ ID NO:2)

[0202] ATG GATTCGGGTTACTCTTCTCCTATGCGGCGGCTGCGGGTATGCACGTTGTTATCTGTCCGTGGCTG
GCTTTTGGTCACCTGCTGCCGTGCTGGATCTGGCACAGCGTCTGGCTTACGCGGCCATCGTGT CAGCTTCGTGTC
TACCCCGCGCAATATTTTCGCTCTGCCGCCGTTTCGTCCGGCACTGGCTCCGCTGGTTGCATTTGTCGCTCTGCCGC
TGCCGCGCGTGAAGGTCTGCCGGATGGTGC GAAAAGTACCAACGACGTGCCGCATGATCGCCCGGACATGGTTGAA
CTGCACCGTCGTGCATTCGATGGTCTGGCAGCACCGTTTTCCGAATTTCTGGGTACGGCGTGCGCCGATTGGGTGAT
CGTTGACGTCTTTTCACTACTGGGCGGCGGCGGCGCTGGAACATAAAGTCCGTGTGCAATGATGCTGCTGGGCT
CAGCTCACATGATTGCGTCGATCGCAGACCGTCGCCTGGAACGTGCAGAAACCGAAAGTCCGGCTGCGGCCGCCAG
GGTCGCCCCGCGAGCTGCGCCGACCTTCGAAGTGCCCGCATGAACTGATTCGTACGAAAGGCAGCTCTGGTATGAG
CCTGGCAGAACGCTTTAGTCTGACCCTGTCCGTAGTTCCTGGTGGTTGGTCGAGTTGCGTTGAATTTGAACCGG
AAACCGTCCCGCTGCTGTCCACGCTGCGTGGTAAACCGATCACCTTTCTGGGTCTGATGCCGCCGCTGCATGAAGGC
CGTCGCGAAGATGGTGAAGACGCAACGGTTCGTTGGCTGGATGCACAGCCGGCTAAAAGCGTCGTGTATGTCCGCT

GGGCTCTGAAGTGCCGCTGGGTGTGGAAAAAGTTCACGAACTGGCACTGGGCCTGGAAGTGGCTGGCACCCGCTTCC
TGTGGGCACTGCGTAAACCGACGGGTGTGAGCGATGCGGACCTGCTGCCGCGCGTTTTGAAGAACGTACCCGCGGC
CGTGGTGTTCGCAACCGCTTGGGTCCCGCAAATGAGCATTCTGGCGCATGCCGCAGTGGGCGCCTTTCTGACCCA
CTGTGGTTGGAACAGCACGATCGAAGGCCTGATGTTTGGTCACCCGCTGATTATGCTGCCGATCTTCGGCGATCAGG
GTCCGAACGCACGTCTGATTGAAGCGAAAAATGCCGGCCTGCAAGTTGCGCGCAACGATGGCGACGGTTCTTTCGAC
CGTGAGGGTGTGGCTGCGGCCATTCGCGCAGTGGCTGTTGAAGAAGAATCATCGAAAGTTTTTCAGGCGAAAGCCAA
AAAAGTGAAGAAATCGTCGCGGATATGGCCTGCCACGAACGCTACATTGATGGTTTTATTAGCAACTGCGCTCCT
ACAAAGACTAA

[0203] EUCP1:氨基酸序列 (SEQ ID NO:3)

[0204] MGSSGMSLAERFSLTLRSSSLVVGSRSCVEFEPETVPLLSTLRGKPIITFLGLMPPLHEGRREDGEDATVR
WLDAQPAKSVVYVALGSEVPLGVEKVHELALGLELAGTRFLWALRKPTGVSDADLLPAGFEERTRGRGVVATRWVPPQ
MSILAHAAVGAFLTHCGWNSTIEGLMFGHPLIMLP IFGDQGP NARL IEAKNAGLQVARNDGDGSFDREGVAAAIRAV
AVEEESKVFQAKAKKLQEI VADMACHERYIDGFIQQLRSYKDDSGYSSSYAAAAGMHVVICPWLAFGHLLPCLDLA
QRLASRGHRVSFVSTPRNISRLPPVRPALAPLVAFFVALPLPRVEGLPDGAESTNDVPHDRPDMVELHRRAFDGLAAP
FSEFLGTACADWVIVDFVHHWAAAAALEHKVPCAMMLLGS AHMIAS IADRR LERAETESPAAGQGRPAAAPTFEVA
RMKLIRTK

[0205] EUCP1:DNA序列 (SEQ ID NO:4)

[0206] ATGGGTAGCTCGGGCATGTCCCTGGCGGAACGCTTTTCGCTGACGCTGAGTCGCTCATCCCTGGTTGTT
GGTCGCAGTTGTGTTGAATTTGAACCGGAAACCGTTCGCTGCTGTCTACGCTGCGCGGCAAACCGATTACCTTCT
GGTCTGATGCCGCCGCTGCATGAAGGCCGTCGCGAAGATGGTGAAGACGCCACGGTTCGTTGGCTGGATGCTCAGC
CGGCGAAATCGGTGGTTTATGTGCGACTGGGCAGCGAAGTGCCGCTGGGTGTCGAAAAAGTGCACGAACTGGCCCTG
GGCCTGGAAGTGGCAGGCACCCGCTTTCTGTGGGCACTGCGTAAACCGACGGCGTTAGCGATGCTGACCTGCTGCC
GGCGGGTTTTCGAAGAACGCACCCGCGGCCGTTGGTGTCTGTTGGCCACCCGTTGGGTGCCGAAATGTCCATTCTGGCTC
ATGCGGCCGTTGGCGCATTCTGACCCACTGCGGTTGGAACAGCACGATCGAAGGCCTGATGTTTGGTCATCCGCTG
ATTATGCTGCCGATCTTCGGCGATCAGGGTCCGAACGCACGCCTGATCGAAGCCAAAAATGCAGGCCTGCAAGTTGC
GCGTAACGATGGCGACGGTAGCTTTGACCGCGAAGGTGTGCGCAGCTGCGATTCTGTGCTGTGGCGGTTGAAGAAGAAA
GCAGCAAAGTCTTCCAGGCCAAAGCGAAAAAACTGCAAGAAATCGTGGCTGATATGGCGTGTGATGAACGCTATATT
GACGGCTTTATCCAGCAACTGCGTTCTTACAAAAGATGACAGTGGCTATAGTTCCTCATAACCCGAGCTGCGGGTAT
GCATGTTGTCAATTTGCCCGTGGCTGGCGTTTGGTCACCTGCTGCCGTGTCTGGATCTGGCACAGCGCCTGGCATCTC
GCGGTCACCGTGTTCGTTGTCAGCACCCCGCGCAATATCAGTCGCTGCCGCCGTTTCGTCGGCGCTGGCGCCG
CTGTTTGCCTTCGTTGCACTGCCGCTGCCGCTGTGGAAGGTCTGCCGGATGGTGCCGAATCGACCAACGACGTTCC
GCATGATCGTCCGGACATGGTTCGAACTGCATCGTCGCGCCTTTGATGGCCTGGCCGACCGTTTTAGCGAATTTCTGG
GTACGGCCTGCGCAGATTGGGTCAATGTGGACGTTTTTACCACCTGGGCGGCGGCGGCGGCTGGAACATAAAGTG
CCGTGTGCGATGATGCTGCTGGGTTCCGCCACATGATTGCTTCAATCGCGGATCGTCGCTGGAACGTGCCGAAAC
CGAAAGTCCGGCGGCGGCGAGGCCAGGGTCGTCGGCGGCGGCACCGACCTTTGAAGTGGCACGTATGAAACTGATTC
GCACGAAATAA

[0207] HV1:氨基酸序列 (SEQ ID NO:5)

[0208] MDGNSSSSPLHVVICPWLAALGHLLPCLDIAERLASRGHRVSFVSTPRNIARLPPLRPVAVPLVDFVALP
LPHVDGLPEGAESTNDVPYDKFELHRKAFFDGLAAPFSEFLRAACAEGAGSRPDWLVDTFHHWAAAAAVENKVPCVM

LLLGAATVIAGFARGVSEHAAAAVGKERPAAEAPSFETERRKLMTTQNASGMTVAERYFLTLMRSDLVAIRSCAEWE
 PESVAALTTLAGKPVVPLGLLPPSPEGGRGVSKEDAAVRWLDAPAKSVVYVALGSEVPLRAEQVHELALGLELSGA
 RFLWALRKPTDAPDAAVLPPGFEERTRGRGLVVTGWVPQIGVLAHGAVAAFLTHCGWNSTIEGLLFGHPLIMLPISS
 DQGPNARLMEGRKVMQVPRDESDGSFRREDVAATVRAVAVEEDGRRVFTANAKKMQEIVADGACHERCIDGFIQQL
 RSYKA

[0209] HV1:DNA序列 (SEQ ID NO:6)

[0210] ATGGATGGTAACTCCTCCTCCTCGCCGCTGCATGTGGTCATTTGTCCGTGGCTGGCTCTGGGTACCTG
 CTGCCGTGTCTGGATATTGCTGAACGTCTGGCGTCACGCGCCATCGTGTGAGTTTTGTGTCCACCCCGCGCAACAT
 TGCCCGTCTGCCGCCGCTGCGTCCGGCTGTTGCACCGCTGGTTGATTCGTGCGACTGCCGCTGCCGATGTTGACG
 GTCTGCCGGAGGGTGCAGGAATCGACCAATGATGTGCCGTATGACAAATTTGAACTGCACCGTAAGGCGTTCGATGGT
 CTGGCGGCCCGTTTTAGCGAATTTCTGCGTGCAGCTTGCAGCAGAAGGTGCAGGTTCTCGCCCGGATTGGCTGATTGT
 GGACACCTTTCATCACTGGGCGGCGGCGGCGGCGGTGGAAAACAAAGTGCCGTGTGTTATGCTGCTGCTGGGTGCAG
 CAACGGTGATCGCTGGTTTTCGCGCGTGGTGTAGCGAACATGCGGCGGCGGCGGTGGGTAAAGAACGTCCGGCTGCG
 GAAGCCCCGAGTTTTGAAACCGAACGTCGCAAGCTGATGACCACGCAGAATGCCTCCGGCATGACCGTGGCAGAACG
 CTATTTCTGACGCTGATGCGTAGCGATCTGGTTGCCATCCGCTCTGCGCAGAATGGGAACCGAAAGCGTGGCAG
 CACTGACCACGCTGGCAGGTAACCGGTGGTTCGCTGGGTCTGCTGCCGCCGAGTCCGGAAGGCGGTCTGGCGTT
 TCCAAAGAAGATGCTGCGGTCCGTTGGCTGGACGCACAGCCGGCAAAGTCAGTCGTGTACGTCGACTGGGTTCCGA
 AGTGCCGCTGCGTGCGGAACAAGTTCACGAACTGGCACTGGCCTGGAAGTACGCGGTGCTCGCTTTCTGTGGGCGC
 TGCGTAAACCGACCGATGCACCGGACGCCGAGTGTGCCGCCGGTTTTCAAGAACGTACCCGCGGCCGTGGTCTG
 GTTGTACGGGTTGGGTGCCGAGATTGGCGTCTGGCTCATGGTGCAGTGGCTGCGTTTTCTGACCCACTGTGGCTG
 GAACTCTACGATCGAAGGCCTGCTGTTCCGGTTCATCCGCTGATTATGCTGCCGATCAGCTCTGATCAGGGTCCGAATG
 CGCGCCTGATGGAAGGCCGTAAAGTCGGTATGCAAGTGCCGCGTATGAATCAGACGGCTCGTTTTCGTCGCGAAGAT
 GTTGCCGCAACCGTCCGCGCCGTGGCAGTTGAAGAAGACGGTCGTGCGCTTTCACGGCTAACCGGAAAAAGATGCA
 AGAAATTGTGGCCGATGGCGCATGCCACGAACGTTGTATTGACGGTTTTATCCAGCAACTGCGCAGTTACAAGGCGT
 AA

[0211] UGT76G1:氨基酸序列 (SEQ ID NO:7)

[0212] MENKTETTVRRRRRIILFPVPFQGHINPILQLANVLYSKGFSITIFHTNFNPKPNTSNYPHFTFRFILDN
 DPQDERISNLPHTGLAGMRIPIINEHGADELRRLELLMLASEEDEEVSCLITDALWYFAQSVADSLNLRRLVLMT
 SSLFNFHAHVSLPQFDELGYLDPDDKTRLEEQASGFPMLKVKDIKSAYSNWQILKEILGKMIKQTKASSGVIWNSFK
 ELEESELETVIREIPAPSFLIPLPKHLTASSSSLLDHDRTVFQWLDQQPPSSVLYVSGSTSEVDEKDFLEIARGLV
 DSKQSFLWVVRPGFVKGSTWVEPLPDGFLGERGRIVKWPQQEVLHGAIGAFWTHSGWNSTLESVCEGVPMIFSDF
 GLDQPLNARYMSDVLKGVVYLENGWERGEIANAIRRVMVDEEGEYIRQNARVLKQKADVSLMKGGSSYESLESLSVSY
 ISSL

[0213] UGT76G1:DNA序列 (SEQ ID NO:8)

[0214] ATGGAGAATAAGACAGAAACAACCGTAAGACGGAGGCGGAGGATTATCTTGTCCCTGTACCATTTACG
 GGCCATATTAATCCGATCCTCCAATTAGCAAACGTCCTTACTCCAAGGATTTTCAATAACAATCTTCCATACTAA
 CTTTAACAAGCCTAAAACGAGTAATTATCCTCACTTACATTCAGGTTTATTCTAGACAACGACCCTCAGGATGAGC
 GTATCTCAAATTTACCTACGCATGGCCCTTGGCAGGTATGCGAATACCAATAATCAATGAGCATGGAGCCGATGAA
 CTCCGTCGCGAGTTAGAGCTTCTCATGCTCGCAAGTGAGGAAGACGAGGAAGTTTCGTGCCTAATAACTGATGCGCT

TTGGTACTTCGCCCAATCAGTCGCAGACTCACTGAATCTACGCCGTTTGGTCCTTATGACAAGTTCATTATTCAACT
 TTCACGCACATGTATCACTGCCGCAATTTGACGAGTTGGGTACCTGGACCCGGATGACAAAACGCGATTGGAGGAA
 CAAGCGTCGGGCTTCCCATGCTGAAAAGTCAAAGATATTAAGAGCGCTTATAGTAATTGGCAAATTCTGAAAGAAAT
 TCTCGGAAAAATGATAAAGCAAACCAAAGCGTCTCTGGAGTAATCTGGAACCTTCAAGGAGTTAGAGGAATCTG
 AACTTGAAACGGTCATCAGAGAAATCCCCGCTCCCTCGTTCTTAATTCCACTACCCAAGCACCTTACTGCAAGTAGC
 AGTTCCCTCCTAGATCATGACCGAACCGTGTTCAGTGGCTGGATCAGCAACCCCGTCGTCAGTTCTATATGTAAG
 CTTTGGGAGTACTTCGGAAGTGGATGAAAAGGACTTCTTAGAGATTGCGCGAGGGCTCGTGGATAGCAAACAGAGCT
 TCCTGTGGGTAGTGAGACCGGGATTCTGTTAAGGGCTCGACGTGGGTCGAGCCGTTGCCAGATGGTTTTCTAGGGGAG
 AGAGGGAGAATCGTGAAATGGGTCCACAGCAAGAGGTTTTGGCTCACGGAGCTATAGGGCCTTTTGGACCCACTC
 TGGTTGGAATTCTACTCTTGAAAAGTGTCTGTGAAGGCGTCCAATGATATTTCTGATTTTGGGCTTGACCAGCCTC
 TAAACGCTCGCTATATGTCTGATGTGTGAAGGTTGGCGTGTACCTGGAGAATGGTTGGGAAAGGGGGAAATTGCC
 AACGCCATACGCCGGTAATGGTGGACGAGGAAGGTGAGTACATACGTCAGAACGCTCGGGTTTTAAAACAAAAGC
 GGACGTCAGCCTTATGAAGGGAGGTAGCTCCTATGAATCCCTAGAATCCTTGGTAAGCTATATATCTTCGTTATAA

[0215] UGT2E-B:氨基酸序列 (SEQ ID NO:9)

[0216] MATSDSIVDDRKQLHVATFPWLAFGHILPYLQLSKLIAEKGHKVSFLSTTRNIQRLSSHISPLINVVQL
 TLPRVQELPEDAEATTDVHPEDIPYLKKASDGLQPEVTRFLEQHSPDWI IYDYTHYWLPSIAASLGISRHFVSTTP
 WAIAYMGPSADAMINGS DGRRTTVEDLTPPKWFFPFTKVCWRKHLRLVYPKAPGISDGYRMGMVLKGS DCLLSKC
 YHEFGTQWLPLETLHQVPVVPVGLLPPEIPGDEKDETWVSIKKWLDGKQKGSVVYVALGSEALVSQTEVVELALGL
 ELSGLPFVWAYRKP KPAKSDSVELPDGFVERTRDRGLVWTSWAPQLRILSHESVCGFLTHCGSGSIVEGLMFGHPL
 IMLPIFGDQPLNARLLEDKQVGIEIPRNEEDGCLTKESVARSLRSVVVEKEGEIYKANARELSKIYNDTKVEKEYVS
 QFVDYLEKNARAVAI DHES

[0217] UGT2E-B:DNA序列 (SEQ ID NO:10)

[0218] ATGGCTACCAGTGACTCCATAGTTGACGACCGTAAGCAGCTTCATGTTGCGACGTTCCCATGGCTTGCT
 TTCGGTCACATCCTCCCTTACCTTACGCTTTCGAAATTGATAGCTGAAAAGGGTCACAAAGTCTCGTTTCTTTCTAC
 CACCAGAAACATTCAACGTCTCTTCTCATATCTCGCCACTCATAAATGTTGTTCAACTCACACTTCCACGTGTCC
 AAGAGCTGCCGGAGGATGCAGAGGCGACCACTGACGTCCACCCTGAAGATATTCATATCTCAAGAAGGCTTCTGAT
 GGTCTTCAACCGGAGGTCACCCGTTTCTAGAACAACTCTCCGGACTGGATTATTTATGATTATACTCACTACTG
 GTTGCCATCCATCGCGGCTAGCCTCGGTATCTCACGAGCCACTTCTCCGTCACCACTCCATGGGCCATTGCTTATA
 TGGGACCCTCAGCTGACCCATGATAAATGGTTCAGATGGTCGAACCACGGTTGAGGATCTCACGACACCGCCCAAG
 TGGTTTTCCCTTTCCGACCAAAGTATGCTGGCGGAAGCATGATCTGCCCGACTGGTGCCTTACAAAGCTCCGGGGAT
 ATCTGATGGATAACCGTATGGGGATGGTTCTTAAGGGATCTGATTGTTTGGCTTTCCAAATGTTACCATGAGTTTGGAA
 CTCAATGGCTACCTCTTTTGGAGACACTACACCAAGTACCGGTGGTTCCGGTGGGATTACTGCCACCGGAAATACCC
 GGAGACGAGAAAGATGAAACATGGGTGTCAATCAAGAAAATGGCTCGATGGTAAACAAAAGGCAGTGTGGTGTACGT
 TGCATTAGGAAGCGAGGCTTTGGTGAGCCAAACCGAGGTTGTTGAGTTAGCATTGGGTCTCGAGCTTTCTGGGTTGC
 CATTTGTTTGGGCTTATAGAAAACAAAAGTCCCGGAAGTCAGACTCGGTGGAGTTGCCAGACGGGTTCTGGGAA
 CGAACTCGTGACCGTGGGTTGGTCTGGACGAGTTGGGCACCTCAGTTACGAATACTGAGCCATGAGTCGGTTTGTGG
 TTTCTTGACTCATTGTGGTTCTGGATCAATTGTGGAAGGGCTAATGTTTGGTCACCCTCTAATCATGCTACCGATTT
 TTGGGGACCAACCTCTGAATGCTCGATTACTGGAGGACAAACAGGTGGGAATCGAGATACCAAGAAATGAGGAAGAT
 GGTTGCTTGACCAAGGAGTCGGTTGCTAGATCACTGAGGTCCGTTGTTGTGAAAAAGAAGGGGAGATCTACAAGGC

GAACGCGAGGGAGCTGAGTAAATCTATAACGACACTAAGGTTGAAAAAGAATATGTAAGCCAATTCGTAGACTATT
TGGAAAAGAATGCGCGTGC GGTTGCCATCGATCATGAGAGTTAA

[0219] CP1:氨基酸序列 (SEQ ID NO:11)

[0220] MNWQILKEILGKMIKQTKASSGVIWNSFKELEESELETVIREIPAPSFLIPLPKHLTASSSSLLDHDRT
VFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSKQSFLWVVRPGFVKGSTWVEPLPDGFLGERGRIVKWPV
QQEVLAHGAIGAFWTHSGWNSTLESVCEGVPMIFSDFGLDQPLNARYMSDVLKGVYLENGWERGEIANAIRVMVD
EEGEYIRQNRVLKQKADVSLMKGGSSYESLESLSVSYISSLENKTETT VRRRRRIILFPVPFQGHINPILQLANVLY
SKGFSITIFHTNFKPKTSNYPHFTFRFILDNDPQDERISNLPTHGPLAGMRIPIINEHGADELRRLELLMLASEE
DEEVSLITDALWYFAQSVADSLNLRRLVMTSSLFNFHAHVSLPQFDELGYLDPDDKTRLEEQASGFPMLKVKDIK
SAYS

[0221] CP1:DNA序列 (SEQ ID NO:12)

[0222] ATGAACTGGCAAATCCTGAAAGAAATCCTGGGTAAAATGATCAAACAAACAAAGCGTCGTCGGGCGTT
ATCTGAACTCCTTCAAAGAACTGGAAGAATCAGAACTGGAACCGTTATTCGCGAAATCCCGGCTCCGTCGTTCTCT
GATTCGCTGCCGAAACATCTGACCGCGAGCAGCAGCAGCCTGCTGGATCACGACCGTACGGTCTTTCAGTGGCTGG
ATCAGCAACCGCCGTCATCGGTGCTGTATGTTTCATTCGGTAGCACCTCTGAAGTCGATGAAAAAGACTTTCTGGAA
ATCGCTCGCGGCTGGTGGATAGTAAACAGTCCTTCTGTGGGTGGTTCGTCGGGTTTTGTGAAAGGCAGCACGTG
GGTTGAACCGCTGCCGATGGCTTCTGGGTGAACGCGCCGCTATTGTCAAATGGGTGCCGCAGCAAGAAGTGTGG
CACATGGTGTATCGGCGGTTTTGGACCCACTCTGGTTGGAACAGTACGCTGGAATCCGTTTGCGAAGGTGTCCCG
ATGATTTTCAGCGATTTTGGCCTGGACCAGCCGCTGAATGCCCGCTATATGTCTGATGTTCTGAAAGTCGGTGTGTA
CCTGGAACCGTTGGGAACGTGGCGAAATTGCGAATGCCATCCGTCGCGTTATGGTCGATGAAGAAGGCGAATACA
TTCGCCAGAACGCTCGTGTCTGAAACAAAAAGCGGACGTGAGCCTGATGAAAGGCGGTAGCTCTTATGAATCACTG
GAATCGCTGGTTAGCTACATCAGTTCCTGGAATAAAAACCGAAACCAGGTGCGTCGCCGTCGCCGTTATATCCT
GTTCCCGTTCCGTTTCAGGGTCATATTAACCCGATCCTGCAACTGGCGAATGTTCTGTATTCAAAGGCTTTTCGA
TCACCATCTTCCATACGAACTTCAACAAAACCGAAAACAGTAACTACCCGCACTTTACGTTCCGCTTTATTCTGGAT
AACGACCCGAGGATGAACGTATCTCCAATCTGCCGACCCACGGCCCGCTGGCCGATGCGCATTCCGATTATCAA
TGAACACGGTGAGATGAACTGCGCCGTAAGTGGAACTGCTGATGCTGGCCAGTGAAGAAGATGAAGAAGTGTCTCT
GTCTGATCACCGACGCACTGTGGTATTTCCGCCAGAGCGTTGCAGATTCTCTGAACCTGCGCCGCTCTGGTCTGATG
ACGTCATCGCTGTTCAATTTTCATGCGCACGTTTCTCTGCCGCAATTTGATGAACTGGGCTACCTGGACCCGGATGA
CAAACCCGCTCTGGAAGAACAAGCCAGTGGTTTTCCGATGCTGAAAGTCAAAGACATTAATCCGCCTATTCTGTAA

[0223] CP2:氨基酸序列 (SEQ ID NO:13)

[0224] MNWQILKEILGKMIKQTKASSGVIWNSFKELEESELETVIREIPAPSFLIPLPKHLTASSSSLLDHDRT
VFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLVDSKQSFLWVVRPGFVKGSTWVEPLPDGFLGERGRIVKWPV
QQEVLAHGAIGAFWTHSGWNSTLESVCEGVPMIFSDFGLDQPLNARYMSDVLKGVYLENGWERGEIANAIRVMVD
EEGEYIRQNRVLKQKADVSLMKGGSSYESLESLSVSYISSLYKDDSGYSSSYAAAAGMENKTETT VRRRRRIILFPV
PFQGHINPILQLANVLYSKGFSITIFHTNFKPKTSNYPHFTFRFILDNDPQDERISNLPTHGPLAGMRIPIINEHG
ADELRRLELLMLASEE DEEVSLITDALWYFAQSVADSLNLRRLVMTSSLFNFHAHVSLPQFDELGYLDPDDKTR
LEEQASGFPMLKVKDIKSAYS

[0225] CP2:DNA序列 (SEQ ID NO:14)

[0226] ATGAACTGGCAAATCCTGAAAGAAATCCTGGGTAAAATGATCAAACAAACAAAGCGTCGTCGGGCGTT

ATCTGGAACCTCCTTCAAAGAACTGGAAGAATCAGAACTGGAAACCGTTATTCGCGAAATCCCGGCTCCGTCGTTCT
 GATTCCGCTGCCGAAACATCTGACCGCGAGCAGCAGCAGCCTGCTGGATCACGACCGTACGGTCTTTCAGTGGCTGG
 ATCAGCAACCGCCGTCATCGGTGCTGTATGTTTCATTCGGTAGCACCTCTGAAGTCGATGAAAAAGACTTTCTGGAA
 ATCGCTCGCGGCTGGTGGATAGTAAACAGTCCCTCCTGTGGGTGGTTCGTCCGGGTTTTGTGAAAGGCAGCACGTG
 GGTTGAACCGCTGCCGATGGCTTCTGGGTGAACCGCGCCGTTATTGTCAAATGGGTGCCGAGCAAGAAGTGTGTG
 CACATGGTGTATCGGCGGTTTTGGACCCACTCTGGTTGGAACAGTACGCTGGAATCCGTTTGCGAAGGTGTCCCG
 ATGATTTTCAGCGATTTTGGCCTGGACCAGCCGCTGAATGCCCGCTATATGTCTGATGTTCTGAAAGTCGGTGTGTA
 CCTGGAAAACGGTTGGGAACGTGGCGAAAATTGCGAATGCCATCCGTCGCGTTATGGTCGATGAAGAAGGCGAATACA
 TTCGCCAGAACGCTCGTGTCTGAAAACAAAAAGCGGACGTGAGCCTGATGAAAGGCGGTAGCTCTTATGAATCACTG
 GAATCGCTGGTTAGCTACATCAGTTCCTGTACAAAAGATGACAGCGGTTATAGCAGCAGCTATGCGGCGGCGGCGGG
 TATGGAAAATAAAACCGAAACCACGGTGCCTCGCCGTCGCCGTTATCCTGTTCCCGTTCCGTTTCAGGGTCATA
 TTAACCCGATCCTGCAACTGGCGAATGTTCTGTATTCAAAGGCTTTTCGATCACCATCTTCCATACGAACTTCAAC
 AAACCGAAAACAGTAACCTACCCGACTTTACGTTCCGCTTATTCTGGATAACGACCCGAGGATGAACGTATCTC
 CAATCTGCCGACCCACGGCCCGCTGGCCGGTATGCGCATTCCGATTATCAATGAACACGGTGCAGATGAACTGCGCC
 GTGAACTGGAACCTGCTGATGCTGGCCAGTGAAGAAGATGAAGAAGTGCCTGTCTGATCACCGACGCACTGTGGTAT
 TTCGCCAGAGCGTTGCAGATTCTCTGAACCTGCGCCGCTGCTGCTGATGACGTCATCGCTGTTCAATTTTCATGC
 GCACGTTTCTCTGCCGAATTTGATGAACTGGGCTACCTGGACCCGGATGACAAAACCCGCTCTGGAAGAACAAGCCA
 GTGGTTTTCCGATGCTGAAAGTCAAAGACATTAATCCGCCTATTCGTAA

[0227] GS:氨基酸序列 (SEQ ID NO:15)

[0228] MENKTETTVRRRRRIILFPVPFQGHINPILQLANVLYSKGFSITIFHTNFKPKTSNYPHFTFRFILDN
 DPQDERISNLPTHGLAGMRIPIINEHGADLRRELELLMLASEEDEEVSLITDALWYFAQSVADSLNLRRLVLMT
 SSLFNFAHVSPLPQFDELGYLDPDDKTRLEEQASGFPMKVKDIKSAYSNWQILKEILGMIKQTKASSGVIWNSFK
 ELEESELETVIREIPAPSFLIPLPKHLTASSSSLLDHDRTVFQWLDQQPPSSVLYVSFGSTSEVDEKDFLEIARGLV
 DSKQSFLWVVRPGFVKGSTWVEPLPDGFLGERGRIVKWPQQEVLHGAIGAFWTHSGWNSTLESVCEGVPMIFSDF
 GLDQPLNARYMSDVLKGVVYLENGWERGEIANAIRRMVDEEGEYIRQNARVLKQKADVSLMKGSSYESLESLSVSY
 ISSLGSGANAERMITRVHSQRERLNETLVSERNEVLALLSRVEAKGKGIQQNQIIAEFEALPEQTRKKLEGGPFFD
 LLKSTQEAIIVLPPWVALAVRPRPGVWEYLRVNLHALVVEELQPAEFLHFKEELVDGVKNGNFTLELDFEPNASIPR
 PTLHKYIGNVDVFLNRHLSAKLFHDKESLLPLLKFLRLHSHQGNLMLSEKIQLNNTLQHTLRKAEYLAELKSETL
 YEEFEAKFEEIGLERGWGDAERVLDMIRLLLDLLEAPDPTLETFLGRVPMVFNVILSPHGYFAQDNVLGYPDTG
 GQVVYILDQVRALEIEMLRQIKQQGLNIKPRILILTRLLPDAVGTTCGERLERVYDSEYCDILRVPFRTEKGIVRKW
 ISRFEVWPYLETYTEDAAVELSKELNGKPDIIIGNYSDGNLVAELLAHKLGVQCTIAHALEKTKYPDSDIYWKLD
 DKYHFSCQFTADIFAMNHTDFIITSTFQEIAGSKETVGQYESHTAFTLPLGLYRVVHGIDVFDPKFNIVSPGADMSIY
 FPYTEEKRLTKFHSEIEELLYSDVENKEHLCVLKDKKKPILFTMARLDRVKNLSGLVEWYGKNTRLRELANLVVVG
 GDRRKESKDNEEKAEMKKMYDLIEEYKLNQFRWISSQMDRVRNGLYRYICDTKGAFVQPALYEAFGLTVVEAMTC
 GLPTFATCKGGPAEIVHKGSGFHIDPYHGDQAADTLADFFTKCKEDPSHWDEISKGGLQRIEEKYTWQIYSQRLLT
 LTGVYGFWKHVSNDRLREARRYLEMFYALKYRPLAQAVPLAQDD

[0229] GS:DNA序列 (SEQ ID NO:16)

[0230] ATGGAGAATAAGACAGAAACAACCGTAAGACGGAGCGGAGGATTATCTTGTCCCTGTACCATTTACG
 GGCCATATTAATCCGATCCTCCAATTAGCAAACGTCCTCTACTCCAAGGGATTTTCAATAACAATCTTCCATACTAA

CTTTAACAAGCCTAAAACGAGTAATTATCCTCACTTTACATTTCAGGTTTCATTCTAGACAACGACCCTCAGGATGAGC
GTATCTCAAATTTACCTACGCATGGCCCTTGGCAGGTATGCCAATACCAATAATCAATGAGCATGGAGCCGATGAA
CTCCGTCGCGAGTTAGAGCTTCTCATGCTCGCAAGTGAGGAAGACGAGGAAGTTTCGTGCCTAATAACTGATGCGCT
TTGGTACTTCGCCCAATCAGTCGCAGACTCACTGAATCTACGCCGTTTGGTCCTTATGACAAGTTCATTATTCAACT
TTCACGCACATGTATCACTGCCGCAATTTGACGAGTTGGGTACCTGGACCCGGATGACAAAACGCGATTGGAGGAA
CAAGCGTCGGGCTTCCCATGCTGAAAAGTCAAAGATATTAAGAGCGCTTATAGTAATTGGCAAATTCTGAAAGAAAT
TCTCGAAAAATGATAAAGCAAACCAAAGCGTCTCTGGAGTAATCTGGAACCTTCAAGGAGTTAGAGGAATCTG
AACTTGAAACGGTCATCAGAGAAATCCCCGCTCCCTCGTTCTTAATCCACTACCCAAGCACCTTACTGCAAGTAGC
AGTTCCCTCCTAGATCATGACCGAACCGTGTTCAGTGGCTGGATCAGCAACCCCGTCGTCAGTTCTATATGTAAG
CTTTGGGAGTACTTCGGAAGTGGATGAAAAGGACTTCTTAGAGATTGCGCGAGGGCTCGTGGATAGCAAACAGAGCT
TCCTGTGGGTAGTGAGACCGGGATTTCGTTAAGGGCTCGACGTGGTTCGAGCCGTTGCCAGATGGTTTTCTAGGGGAG
AGAGGGAGAATCGTGAAATGGGTCCACAGCAAGAGGTTTTGGCTCACGGAGCTATAGGGGCCTTTTGACCCACTC
TGTTTGAATTCTACTCTTGAAAGTGTCTGTGAAGGCGTTCCAATGATATTTCTGATTTTGGGCTTGACCAGCCTC
TAAACGCTCGCTATATGTCTGATGTGTGAAGGTTGGCGGTACCTGGAGAATGGTTGGGAAAGGGGGAAATTGCC
AACGCCATACGCCGGTAATGGTGGACGAGGAAGGTGAGTACATACGTCAGAACGCTCGGGTTTTAAAACAAAAGC
GGACGTCAGCCTTATGAAGGGAGGTAGCTCCTATGAATCCCTAGAATCCTTGGTAAGCTATATATCTTCGTTAGGTT
CTGGTGCAAACGCTGAACGTATGATAACGCGCTCCACAGCCAACGTGAGCGTTTGAACGAAACGCTTGTCTGAG
AGAAACGAAGTCCTTGCCTTGCTTCCAGGGTTGAAGCCAAAGGTAAAGGTATTTACAACAAAACCAGATCATTGC
TGAATTCGAAGCTTTGCCTGAACAAAACCCGGAAGAAACTTGAAGGTGGTCCTTCTTTGACCTTCTCAAATCCACTC
AGGAAGCAATTGTGTTGCCACCATGGGTTGCTCTAGCTGTGAGGCCAAGGCCTGGTGTGGGAATACTTACGAGTC
AATCTCCATGCTCTTGTGCTTGAAGAACTCCAACCTGCTGAGTTTCTTCATTTCAAGGAAGAACTCGTTGATGGAGT
TAAGAATGGTAATTTCACTCTTGAGCTTGATTTGAGCCATTCAATGCGTCTATCCCTCGTCCAACACTCCACAAAT
ACATTGGAATGGTGTGACTTCCCTAACCGTCATTTATCGGCTAAGCTCTTCCATGACAAGGAGAGTTTGTCTCCA
TTGCTTAAGTTCCTTCGCTTTCACAGCCACCAGGGCAAGAACCTGATGTTGAGCGAGAAGATTCAGAACCTCAACAC
TCTGCAACACACCTTGAGGAAAGCAGAAGAGTATCTAGCAGAGCTTAAAGTCCGAAACACTGTATGAAGAGTTTGAGG
CCAAGTTTGAGGAGATTGGTCTTGAGAGGGGATGGGGAGACAATGCAGAGCGTGTCTTGACATGATACGTCTTCTT
TTGGACCTTCTTGAGGCGCCTGATCCTTGCACTCTTGAGACTTTTCTTGAAGAGTACCAATGGTGTTCACGTTGT
GATCCTCTCTCCACATGGTTACTTTGCTCAGGACAATGTTCTTGGTTACCCTGACACTGGTGGACAGGTTGTTTACA
TTCTTGATCAAGTTCGTGCTCTGGAGATAGAGATGCTTCAACGTATTAAGCAACAAGGACTCAACATTAACCAAGG
ATTCTCATTCTAACTCGACTTCTACCTGATGCGGTAGGAACTACATGCGGTGAACGCTCTCGAGAGAGTTTATGATTC
TGAGTACTGTGATATTCTTCGTGTGCCCTCAGAACAGAGAAGGGTATTGTTTCGCAAATGGATCTCAAGGTTCTGAAG
TCTGGCCATATCTAGAGACTTACACCGAGGATGCTGCGGTTGAGCTATCGAAAGAATTGAATGGCAAGCCTGACCTT
ATCATTGGTAACTACAGTGATGGAATCTTGTGCTTCTTTATTGGCTCACAACTGGTGTCACTCAGTGTACCAT
TGCTCATGCTCTTGAGAAAACAAAGTACCCGGATTCTGATATCTACTGGAAGAAGCTTGACGACAAGTACCATTTCT
CATGCCAGTTCCTGCGGATATTTTCGCAATGAACCACACTGATTTTCATCATCACTAGTACTTTCCAAGAAATTGCT
GGAAGCAAAGAACTGTTGGGCAGTATGAAAACCACACAGCCTTTACTCTCCCGGATTGTATCGAGTTGTTACCGG
GATTGATGTGTTTGATCCCAAGTTCAACATTGTCTCTCCTGGTGTGATATGAGCATCTACTTCCCTTACACAGAGG
AGAAGCGTAGATTGACTAAGTTCCTCTGAGATCGAGGAGCTCCTCTACAGCGATGTTGAGAACAAAGAGCACTTA
TGTGTGCTCAAGGACAAGAAGAAGCCGATTCTTTCACAATGGCTAGGCTTGATCGTGTCAAGAAGTGTGAGGCT

TGTTGAGTGGTACGGGAAGAACACCCGCTTGC GTGAGCTAGCTAACTTGGTTGTTGTTGGAGGAGACAGGAGGAAAG
 AGTCAAAGGACAATGAAGAGAAAAGCAGAGATGAAGAAAATGTATGATCTCATTGAGGAATACAAGCTAAACGGTCAG
 TTCAGGTGGATCTCCTCTCAGATGGACCGGGTAAGGAACGGTGAGCTGTACCGGTACATCTGTGACACCAAGGGTGC
 TTTTGTCCAACCTGCATTATATGAAGCCTTTGGGTTAACTGTTGTGGAGGCTATGACTTGTGGTTTACCGACTTTCG
 CCACTTGCAAAGGTGGTCCAGCTGAGATCATTGTGCACGGTAAATCGGGTTTCCACATTGACCCTTACCATGGTGAT
 CAGGCTGCTGATACTCTTGCTGATTTCTTCACCAAGTGTAAGGAGGATCCATCTCACTGGGATGAGATCTCAAAGG
 AGGGCTTCAGAGGATTGAGGAGAAAATACACTTGGCAAATCTATTCACAGAGGCTCTTGACATTGACTGGTGTGTATG
 GATTCTGGAAGCATGTCTCGAACCTTGACCGTCTTGAGGCTCGCCGTTACCTTGAAATGTTCTATGCATTGAAGTAT
 CGCCCATTGGCTCAGGCTGTTCTCTTGACAAGATGATTGA

[0231] AtSUS1:氨基酸序列 (SEQ ID NO:17)

[0232] MANAERMITRVHSQRERLNETLVSERNEVLALLSRVEAKGKGI LQQNQI IAEFEALPEQTRKKLEGGPF
 FDLLKSTQEAI VLPWVALAVRPRPGVWEYLRVNLHALVVEELQPAEFLHFKEELVDGVKNGNFTLELDFEPFNASI
 PRPTLHKYIIGNVDFLNRHLSAKLFHDKESLLPLLKFLRLHSHQGNLMLSEKIQLNLTQLHTLRKAEYLAELKSE
 TLYEEFEAKFEEI GLERGWG DNAERVLDMIRLLLDLEAPDPCTLETFLGRVPMVFNVVILSPHGYFAQDNVLGYPD
 TGGQVVYILDQVRALEIEMLRIKQQGLNIKPRIL IL TRLLPDAVGTTCGERLERYVDSEYCDILRVPFRTEKGIVR
 KWISRFVWPYLETYTEDAAVELSKELNGKPDLIIGNYSDGNL VASLLAHKLGVTQCTIAHALEKTKYPDSDIYWKK
 LDDKYHFSCQFTADIFAMNHTDFIITSTFQEIAGSKETVGGYESHAF TLPGLYRVVHGIDVFDPKFNIVSPGADMS
 IYFPYTEEKRRLTKFHSEIEELLYSDVENKEHL CVLKDKKKPILFTMARLDRVKNLSGLVEWYGKNTRLRELANLVV
 VGGDRRKESKDNEEKAEMKMYDLIEEYKLNQFRWISSQMDRVN GELYRYICDTKGAFVQPALYEAFLTVVEAM
 TCGLPTFATCKGGPAEIVHKGSGFHIDPYHGDQAADTLADFFTKCKEDPSHWDEISKGGLQRIEEKYTWQIYSQRL
 LTLTG VYGFWKHVS NLDRL EARRYLEM FYALKYRPLAQAVPLAQDD

[0233] AtSUS1:DNA序列 (SEQ ID NO:18)

[0234] ATGGCAAACGCTGAACGTATGATTACCCGTGTCCACTCCCAACGCGAACGCCTGAACGAAACCCTGGTG
 TCGGAACGCAACGAAGTTCTGGCACTGCTGAGCCGTGTGGAAGCTAAGGGCAAAGGTATTCTGCAGCAAAACCAGAT
 TATCGCGAATTTGAAGCCCTGCCGGAACAAACCCGCAAAAAGCTGGAAGGCGGTCCGTTTTTCGATCTGCTGAAAT
 CTACGCAGGAAGCGATCGTTCTGCCCGGTGGGTCGCACTGGCAGTGCCTCCGCGTCCGGCGTTTTGGGAATATCTG
 CGTGTCAACCTGCATGCACTGGTGGTTGAAGAACTGCAGCCGGCTGAATTTCTGCACTTCAAGGAAGAACTGGTTGA
 CGGCGTCAAAAACGGTAATTTTACCCTGGAACCTGGATTTTGAACCGTTCAATGCCAGTATCCCGCGTCCGACGCTGC
 ATAAATATATTGGCAACGGTGTGGACTTTCTGAATCGCCATCTGAGCGCAAAGCTGTTCCACGATAAAGAATCTCTG
 CTGCCGCTGCTGAAATTCCTGCGTCTGCATAGTCACCAGGGCAAGAACCTGATGCTGTCCGAAAAAATTCAGAACCT
 GAATACCCTGCAACACACGCTGCGCAAGGCGGAAGAATACCTGGCCGAACGAAAAGTGAACCCCTGTACGAAGAAT
 TCGAAGCAAAGTTCGAAGAAATTGGCCTGGAACGTGGCTGGGGTGACAATGCTGAACGTGTTCTGGATATGATCCGT
 CTGCTGCTGGACCTGCTGGAAGCACCGGACCCGTGCACCCTGGAACGTTTCTGGGTCCGCTGCCGATGGTTTTCAA
 CGTCGTGATTCTGTCCCCGCATGGCTATTTTGCACAGGACAATGTGCTGGGTTACCCGGATACCCGGCGGTGAGGTTG
 TCTATATTCTGGATCAAGTTCGTGCGCTGGAATTTGAAATGCTGCAGCGCATCAAGCAGCAAGGCCTGAACATCAAA
 CCGCGTATTCTGATCCTGACCCGTCTGCTGCCGGATGCAGTTGGTACCACGTGCGGTGAACGTCTGGAACGCGTCTA
 TGACAGCGAATACTGTGATATTCTGCGTGTCCCGTTTCGCACCGAAAAGGGTATTGTGCGTAAATGGATCAGTCGCT
 TCGAAGTTTGGCCGTATCTGGAACCTACACGGAAGATGCGGCCGTGGAACGTCCAAGGAACCTGAATGGCAAACCG
 GACCTGATTATCGGCAACTATAGCGATGGTAATCTGGTCGCATCTCTGCTGGCTCATAAACTGGGTGTGACCCAGTG

CACGATTGCACACGCTCTGGAAAAGACCAAAATATCCGGATTTCAGACATCTACTGGAAAAAGCTGGATGACAAATATC
ATTTTTCGTGTCAAGTTACCGCGGACATTTTTGCCATGAACCACACGGATTTTATTATCACCAGTACGTTCCAGGAA
ATCGCGGGCTCCAAAGAAAACCGTGGGTCAATACGAATCACATACCGCCTTCACGCTGCCGGGCTGTATCGTGTGGT
TCACGGTATCGATGTTTTGACCCGAAATTCAATATTGTCAGTCCGGGCGCGGATATGTCCATCTATTTTCCGTACA
CCGAAGAAAAGCGTCGCTGACGAAATTCATTTCAGAAATTGAAGAACTGCTGTACTCGGACGTGGAAAACAAGGAA
CACCTGTGTGTTCTGAAAGATAAAAAAGAAACCGATCCTGTTTACCATGGCCCGTCTGGATCGCGTGAAGAATCTGTC
AGGCCTGGTTGAATGGTATGGTAAAAACACGCGTCTGCGGAACTGGCAAATCTGGTCTGGTTGGCGGTGACCGTC
GCAAGGAATCGAAAGATAACGAAGAAAAGGCTGAAATGAAGAAAATGTACGATCTGATCGAAGAATACAAGCTGAAC
GGCCAGTTTTCGTTGGATCAGCTCTCAAATGGACCGTGTGCGCAATGGCGAACTGTATCGCTACATTTGCGATACCA
GGGTGCGTTTTGTTACCGGCACTGTACGAAGCTTTCGGCCTGACCGTCGTGGAAGCCATGACGTGCGGTCTGCCGA
CCTTTGCGACGTGTAAGGCGGTCCGGCCGAAATATCGTGCATGGCAAATCTGGTTTCCATATCGATCCGTATCAC
GGTGTATCAGGCAGCTGACACCCTGGCGGATTTCTTTACGAAGTGTAAAGAAGACCCGTCACACTGGGATGAAATTC
GAAGGGCGGTCTGCAACGTATCGAAGAAAAATATACCTGGCAGATTTACAGCCAACGCCTGCTGACCCTGACGGGCG
TCTACGGTTTTTGGAAACATGTGTCTAATCTGGATCGCCTGGAAGCCCGTCGCTATCTGGAATGTTTTACGCACTG
AAGTATCGCCCGTGGCACAAGCCGTTCCGCTGGCACAGGACGACTAA

[0235] NX114:氨基酸序列 (SEQ ID NO:19)

[0236] MENGSSPLHVVFIPWLAFLGHLPLFLDLAERLAARGHRVSFVSTPRNLARLRPVRPALRGLVDLVALPLP
RVHGLPDGAEATSDVPFEKFEHRKAFDGLAAPFSAFLDAACAGDKRPDWVIPDFMHWVAAAAQKRGVPCAVLIPC
SADVMALYGQPTETSTEQPEAIARSMAAEAPSFEAERNTTEYGTAGASGVSIMTRFSLTLKWSKLVALRSCPELEPG
VFTTLTRVYSKPVVFPGLLPPRRDGAHGVRKNGEDDGAIRWLDEQPAKSVVYVALGSEAPVSADLLRELAHGLELA
GTRFLWALRRPAGVNDGDSILPNGFLERTGERGLVTTGWVQVSI LAHAAVCAFLTHCGWGSVVEGLQFGHPLIMLP
IIGDQGPNARFLEGRKVGAVPRNHADGSFDRSGVAGAVRAVAVEEEGKAFAANARKLQEI VADRERDERCTDGF IH
HLTSWNELEA

[0237] NX114:DNA序列 (SEQ ID NO:20)

[0238] ATGGAAAATGGTAGCAGTCCGCTGCATGTTGTTATTTTTCCGTGGCTGGCATTGGTTCATCTGCTGCCG
TTTCTGGATCTGGCAGAACGTCTGGCAGCACGTGGTCATCGTGTAGCTTTGTTAGCACACCGCGTAATCTGGCAGC
TCTGCGTCCGGTTCGTCGGCACTGCGTGGTCTGGTTGATCTGGTTGCACTGCCGCTGCCTCGTGTTCATGGTCTGC
CGGATGGTGCCGAAGCAACCAGTGATGTTCCGTTTAAAAATTTGAACTGCACCGCAAAGCATTGATGGCCTGGCT
GCACCGTTTAGCGCATTCTGGATGCAGCATGTGCCGGTGATAAACGTCCGGATTGGGTTATTCCGGATTTTATGCA
TTATTGGGTTGCAGCAGCAGCACAGAAACGTGGTGTTCGTTGTCAGTTCTGATTCCGTGTAGCGCAGATGTTATGG
CACTGTATGGTCAGCCGACCGAAAACCAGCACCGAACAGCCGGAAGCAATTGCACGTAGCATGGCAGCAGAAGCACCG
AGCTTTGAAGCAGAACGTAATACCGAAGAATATGGTACAGCCGGTGCAAGCGGTGTAGCATTATGACCCGTTTTAG
TCTGACCCTGAAATGGTCAAACTGGTTGCCCTGCGTAGCTGTCCGAACTGGAACCGGGTGTTTTTACCACACTGA
CCCGTGTATATAGCAAACCGGTTGTGCCGTTTGGTCTGCTGCCTCCGCGTCGTGATGGTGCACATGGTGTTCGTA
AATGGTGAAGATGATGGTGCCATTATTCGTTGGCTGGATGAACAGCCTGCAAAAAGCGTTGTTTATGTTGCACTGGG
TAGCGAAGCACCGGTTTCAGCCGATCTGCTGCGTGAACGGCACATGGTCTGGAATTAGCAGGCACCCGTTTTCTGT
GGGCTCTGCGTCGTCTGCGGTGTTAATGATGGTGATAGCATTCTGCCGAATGGTTTTCTGGAACGTACCGGTGAA
CGCGGTCTGGTTACCACCGGTTGGGTTCCGAGGTTAGTATTCTGGCCATGCAGCAGTTGTGCATTTCTGACCCA
TTGTGGTTGGGGTAGCGTTGTTGAAGGTTTACAGTTTGGCCATCCGCTGATTATGCTGCCGATTATTGGTGTACAG

GTCCGAATGCACGCTTTCTGGAAGGTCGTAAAGTTGGTGTTCAGTTCCGCGTAACCATGCAGATGGTAGCTTTGAT
CGTAGCGGTGTTGCCGGTGCCGTTTCGTGCAGTTGCAGTTGAAGAAGAAGGTAAAGCCTTGCAGCAAATGCCCGTAA
ACTGCAAGAAATTGTTGCAGATCGTGAACGTGATGAACGTTGTACCGATGGTTTTATTCATCATCTGACCAGCTGGA
ATGAACTGGAAGCATAA

	225	230	235	240			
	Pro Ile Thr Phe Leu Gly Leu Met Pro Pro Leu His Glu Gly Arg Arg						
		245	250	255			
	Glu Asp Gly Glu Asp Ala Thr Val Arg Trp Leu Asp Ala Gln Pro Ala						
	260	265	270				
	Lys Ser Val Val Tyr Val Ala Leu Gly Ser Glu Val Pro Leu Gly Val						
	275	280	285				
	Glu Lys Val His Glu Leu Ala Leu Gly Leu Glu Leu Ala Gly Thr Arg						
	290	295	300				
	Phe Leu Trp Ala Leu Arg Lys Pro Thr Gly Val Ser Asp Ala Asp Leu						
	305	310	315	320			
	Leu Pro Ala Gly Phe Glu Glu Arg Thr Arg Gly Arg Gly Val Val Ala						
	325	330	335				
	Thr Arg Trp Val Pro Gln Met Ser Ile Leu Ala His Ala Ala Val Gly						
	340	345	350				
	Ala Phe Leu Thr His Cys Gly Trp Asn Ser Thr Ile Glu Gly Leu Met						
	355	360	365				
	Phe Gly His Pro Leu Ile Met Leu Pro Ile Phe Gly Asp Gln Gly Pro						
	370	375	380				
	Asn Ala Arg Leu Ile Glu Ala Lys Asn Ala Gly Leu Gln Val Ala Arg						
	385	390	395	400			
	Asn Asp Gly Asp Gly Ser Phe Asp Arg Glu Gly Val Ala Ala Ala Ile						
[0002]	405	410	415				
	Arg Ala Val Ala Val Glu Glu Glu Ser Ser Lys Val Phe Gln Ala Lys						
	420	425	430				
	Ala Lys Lys Leu Gln Glu Ile Val Ala Asp Met Ala Cys His Glu Arg						
	435	440	445				
	Tyr Ile Asp Gly Phe Ile Gln Gln Leu Arg Ser Tyr Lys Asp						
	450	455	460				
	<210>	2					
	<211>	1389					
	<212>	DNA					
	<213>	水稻					
	<400>	2					
	atggattcgg	gttactcttc	ctcctatgcg	gcbgctgcgg	gtatgcacgt	tgttatctgt	60
	ccgtggctgg	cttttggtca	cctgctgccg	tgcttgatc	tgccacagcg	tctggcttca	120
	cgcgccatc	gtgtcagctt	cggtctacc	ccgcgcaata	tttcgctct	gccgccggtt	180
	cgccggcac	tggtccgct	ggttcattt	gtcgtctgc	cgctgccgcg	cggtgaaggt	240
	ctgccgatg	gtgcgaaag	taccaacgac	gtgccgatg	atgcccgga	catggttgaa	300
	ctgaccctc	gtgattcga	tggtctggca	gcaccgtttt	ccgaatttct	gggtacggcg	360
	tgcccgatt	gggtgatcgt	tgactcttt	catcactggg	cggcggcggc	ggcgctggaa	420
	cataaagttc	cggtgcaat	gatgctgctg	ggctcagctc	acatgattgc	gtcgatcgca	480
	gaccgtgcc	tggaactgc	agaaaccgaa	agtccggctg	cggccggcca	gggtcgcccc	540
	gcagctgcg	cgaccttca	agtggccccg	atgaaactga	ttcgtacgaa	aggcagctct	600

	ggatgagcc tggcagaacg ctttagtctg accctgtccc gtagttccct ggtggttgg	660	
	cgcagttgcg ttgaatttga accggaacc gtcccgtgc tgtccacgct gcgttgtaaa	720	
	ccgatcacct tctgggtct gatgcccgct ctgcatgaag gccgtcgcga agatggtgaa	780	
	gacgcaacgg tgcgttgct ggatgcacag ccggctaaaa gcgtcgtgta tgcgccctg	840	
	ggctctgaag tggcctggg tgggaaaaa gttcacgaac tggcactggg cctggaactg	900	
	gctggcacc gcttctgtg ggactgctg aaaccgacgg gtgtgagcga tgcggacctg	960	
	ctgccggccg gtttgaaga acgtaccgc ggccgtggg ttgtcgaac gcgttggtc	1020	
	ccgaaatga gcattctggc gcatgccgca gtgggcct tctgacca ctgtggtg	1080	
	aacgacgca tcgaaggct gatgttggc caccgctga ttatgctgcc gatctcggc	1140	
	gatcagggtc cgaacgcacg tctgattgaa gcgaaaaatg ccggcctgca agttgcgcgc	1200	
	aacgatggcg acggtcttt cgaccgtgag ggtgtggctg cggcattcg cgcagtggt	1260	
	gttgaagaag aatcatcga agttttcag gcgaaagcca aaaaactgca agaaatcgtc	1320	
	gcggatatgg cctgccacga acgctacatt gatggttca ttcagcaact gcgctcctac	1380	
	aaagactaa		1389
	<210> 3		
	<211> 462		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成多肽		
	<400> 3		
[0003]	Met Gly Ser Ser Gly Met Ser Leu Ala Glu Arg Phe Ser Leu Thr Leu		
	1 5 10 15		
	Ser Arg Ser Ser Leu Val Val Gly Arg Ser Cys Val Glu Phe Glu Pro		
	20 25 30		
	Glu Thr Val Pro Leu Leu Ser Thr Leu Arg Gly Lys Pro Ile Thr Phe		
	35 40 45		
	Leu Gly Leu Met Pro Pro Leu His Glu Gly Arg Arg Glu Asp Gly Glu		
	50 55 60		
	Asp Ala Thr Val Arg Trp Leu Asp Ala Gln Pro Ala Lys Ser Val Val		
	65 70 75 80		
	Tyr Val Ala Leu Gly Ser Glu Val Pro Leu Gly Val Glu Lys Val His		
	85 90 95		
	Glu Leu Ala Leu Gly Leu Glu Leu Ala Gly Thr Arg Phe Leu Trp Ala		
	100 105 110		
	Leu Arg Lys Pro Thr Gly Val Ser Asp Ala Asp Leu Leu Pro Ala Gly		
	115 120 125		
	Phe Glu Glu Arg Thr Arg Gly Arg Gly Val Val Ala Thr Arg Trp Val		
	130 135 140		
	Pro Gln Met Ser Ile Leu Ala His Ala Ala Val Gly Ala Phe Leu Thr		
	145 150 155 160		
	His Cys Gly Trp Asn Ser Thr Ile Glu Gly Leu Met Phe Gly His Pro		
	165 170 175		
	Leu Ile Met Leu Pro Ile Phe Gly Asp Gln Gly Pro Asn Ala Arg Leu		

	180	185	190	
	Ile Glu Ala Lys Asn Ala Gly Leu Gln Val Ala Arg Asn Asp Gly Asp			
	195	200	205	
	Gly Ser Phe Asp Arg Glu Gly Val Ala Ala Ala Ile Arg Ala Val Ala			
	210	215	220	
	Val Glu Glu Glu Ser Ser Lys Val Phe Gln Ala Lys Ala Lys Lys Leu			
	225	230	235	240
	Gln Glu Ile Val Ala Asp Met Ala Cys His Glu Arg Tyr Ile Asp Gly			
	245	250	255	
	Phe Ile Gln Gln Leu Arg Ser Tyr Lys Asp Asp Ser Gly Tyr Ser Ser			
	260	265	270	
	Ser Tyr Ala Ala Ala Ala Gly Met His Val Val Ile Cys Pro Trp Leu			
	275	280	285	
	Ala Phe Gly His Leu Leu Pro Cys Leu Asp Leu Ala Gln Arg Leu Ala			
	290	295	300	
	Ser Arg Gly His Arg Val Ser Phe Val Ser Thr Pro Arg Asn Ile Ser			
	305	310	315	320
	Arg Leu Pro Pro Val Arg Pro Ala Leu Ala Pro Leu Val Ala Phe Val			
	325	330	335	
	Ala Leu Pro Leu Pro Arg Val Glu Gly Leu Pro Asp Gly Ala Glu Ser			
	340	345	350	
[0004]	Thr Asn Asp Val Pro His Asp Arg Pro Asp Met Val Glu Leu His Arg			
	355	360	365	
	Arg Ala Phe Asp Gly Leu Ala Ala Pro Phe Ser Glu Phe Leu Gly Thr			
	370	375	380	
	Ala Cys Ala Asp Trp Val Ile Val Asp Val Phe His His Trp Ala Ala			
	385	390	395	400
	Ala Ala Ala Leu Glu His Lys Val Pro Cys Ala Met Met Leu Leu Gly			
	405	410	415	
	Ser Ala His Met Ile Ala Ser Ile Ala Asp Arg Arg Leu Glu Arg Ala			
	420	425	430	
	Glu Thr Glu Ser Pro Ala Ala Ala Gly Gln Gly Arg Pro Ala Ala Ala			
	435	440	445	
	Pro Thr Phe Glu Val Ala Arg Met Lys Leu Ile Arg Thr Lys			
	450	455	460	
	<210> 4			
	<211> 1389			
	<212> DNA			
	<213> 人工序列			
	<220>			
	<223> 合成多核苷酸			
	<400> 4			
	atgggtagct cgggcatgtc cctggcggaa cgcttttcgc tgacgctgag tcgctcatcc		60	
	ctggtgttg gtcgcagttg tgttgaatt gaaccgaaa ccgttccgct gctgtctacg		120	

	ctgcgcggca aaccgattac cttcctgggt ctgatgccgc cgctgcatga aggccgtcgc	180
	gaagatgggt aagacgccac ggtgcgttgg ctggatgctc agccggcgaa atcggtggtt	240
	tatgtcgcac tgggcagcga agtgccgctg ggtgtcgaaa aagtgcacga actggccctg	300
	ggcctggaac tggcaggcac ccgctttctg tgggcactgc gtaaaccgac gggcgtagc	360
	gatgctgacc tgctgccggc gggtttcgaa gaacgcaccc gcggccgtgg tgcctggcc	420
	accggtggg tgccgaaat gtccattctg gtcctatgcg ccgttgccgc atttctgacc	480
	cactgcggtt ggaacagcac gatcgaaggc ctgatgttg gtcctcgcct gattatgctg	540
	ccgatctcg gcatcaggg tccgaacgca cgctgatcg aagccaaaaa tgcaggcctg	600
	caagttgcgc gtaacgatgg cgacggtagc tttgaccgcg aaggtgtcgc agctgcgatt	660
	cgtgctgtgg cggttgaaga agaaagcagc aaagtcttc aggccaaagc gaaaaaactg	720
	caagaaatcg tggctgatat ggcgtgtcat gaacgctata ttgacggctt tatccagcaa	780
	ctgcttctt acaaagatga cagtggtat agttcctcat acgccgagc tgcgggtatg	840
	catgttgca tttcccgtg gctggcgtt ggacacctgc tggcgtgtc ggatctggca	900
	cagccctgg catctcgcgg tcaccgtgtt tcgttcgca gcaccccgcg caatatcagt	960
	cgtctgccgc cggctcgtc ggcctggcg ccgctggtg cgttcgttc actgccgctg	1020
	ccgctgtgg aaggtctgcc ggatggtgcc gaatcgacca acgacgttc gcatgatcgt	1080
	ccggacatgg tcgaactgca tcgtcgcgc tttgatggcc tggccgacc gtttagcga	1140
	tttctggta cggcctgcgc agattgggtc attgtggacg ttttcacca ctggcggcg	1200
	gcggcggcg tggaaacata agtgccgtg gcatgatgc tgcctggctc cgccacatg	1260
	attgctcaa tcgctgatc tcgctggaa cgtgccgaaa ccgaaagtcc ggcggcggca	1320
	ggccagggtc gtccggcggc ggcaccgacc tttgaagtgg cacgtatgaa actgattcgc	1380
	acgaaataa	1389
[0005]	<210> 5	
	<211> 459	
	<212> PRT	
	<213> 大麦	
	<400> 5	
	Met Asp Gly Asn Ser Ser Ser Ser Pro Leu His Val Val Ile Cys Pro	
	1 5 10 15	
	Trp Leu Ala Leu Gly His Leu Leu Pro Cys Leu Asp Ile Ala Glu Arg	
	20 25 30	
	Leu Ala Ser Arg Gly His Arg Val Ser Phe Val Ser Thr Pro Arg Asn	
	35 40 45	
	Ile Ala Arg Leu Pro Pro Leu Arg Pro Ala Val Ala Pro Leu Val Asp	
	50 55 60	
	Phe Val Ala Leu Pro Leu Pro His Val Asp Gly Leu Pro Glu Gly Ala	
	65 70 75 80	
	Glu Ser Thr Asn Asp Val Pro Tyr Asp Lys Phe Glu Leu His Arg Lys	
	85 90 95	
	Ala Phe Asp Gly Leu Ala Ala Pro Phe Ser Glu Phe Leu Arg Ala Ala	
	100 105 110	
	Cys Ala Glu Gly Ala Gly Ser Arg Pro Asp Trp Leu Ile Val Asp Thr	
	115 120 125	
	Phe His His Trp Ala Ala Ala Ala Val Glu Asn Lys Val Pro Cys	

	130	135	140	
	Val Met Leu Leu Leu Gly Ala Ala Thr Val Ile Ala Gly Phe Ala Arg			
	145	150	155	160
	Gly Val Ser Glu His Ala Ala Ala Val Gly Lys Glu Arg Pro Ala			
	165	170		175
	Ala Glu Ala Pro Ser Phe Glu Thr Glu Arg Arg Lys Leu Met Thr Thr			
	180	185		190
	Gln Asn Ala Ser Gly Met Thr Val Ala Glu Arg Tyr Phe Leu Thr Leu			
	195	200		205
	Met Arg Ser Asp Leu Val Ala Ile Arg Ser Cys Ala Glu Trp Glu Pro			
	210	215		220
	Glu Ser Val Ala Ala Leu Thr Thr Leu Ala Gly Lys Pro Val Val Pro			
	225	230		240
	Leu Gly Leu Leu Pro Pro Ser Pro Glu Gly Gly Arg Gly Val Ser Lys			
	245	250		255
	Glu Asp Ala Ala Val Arg Trp Leu Asp Ala Gln Pro Ala Lys Ser Val			
	260	265		270
	Val Tyr Val Ala Leu Gly Ser Glu Val Pro Leu Arg Ala Glu Gln Val			
	275	280		285
	His Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly Ala Arg Phe Leu Trp			
	290	295		300
[0006]	Ala Leu Arg Lys Pro Thr Asp Ala Pro Asp Ala Ala Val Leu Pro Pro			
	305	310		320
	Gly Phe Glu Glu Arg Thr Arg Gly Arg Gly Leu Val Val Thr Gly Trp			
	325	330		335
	Val Pro Gln Ile Gly Val Leu Ala His Gly Ala Val Ala Ala Phe Leu			
	340	345		350
	Thr His Cys Gly Trp Asn Ser Thr Ile Glu Gly Leu Leu Phe Gly His			
	355	360		365
	Pro Leu Ile Met Leu Pro Ile Ser Ser Asp Gln Gly Pro Asn Ala Arg			
	370	375		380
	Leu Met Glu Gly Arg Lys Val Gly Met Gln Val Pro Arg Asp Glu Ser			
	385	390		400
	Asp Gly Ser Phe Arg Arg Glu Asp Val Ala Ala Thr Val Arg Ala Val			
	405	410		415
	Ala Val Glu Glu Asp Gly Arg Arg Val Phe Thr Ala Asn Ala Lys Lys			
	420	425		430
	Met Gln Glu Ile Val Ala Asp Gly Ala Cys His Glu Arg Cys Ile Asp			
	435	440		445
	Gly Phe Ile Gln Gln Leu Arg Ser Tyr Lys Ala			
	450	455		
	<210> 6			
	<211> 1380			
	<212> DNA			

	Glu Glu Asp Glu Glu Val Ser Cys Leu Ile Thr Asp Ala Leu Trp Tyr			
	115	120	125	
	Phe Ala Gln Ser Val Ala Asp Ser Leu Asn Leu Arg Arg Leu Val Leu			
	130	135	140	
	Met Thr Ser Ser Leu Phe Asn Phe His Ala His Val Ser Leu Pro Gln			
	145	150	155	160
	Phe Asp Glu Leu Gly Tyr Leu Asp Pro Asp Asp Lys Thr Arg Leu Glu			
	165	170	175	
	Glu Gln Ala Ser Gly Phe Pro Met Leu Lys Val Lys Asp Ile Lys Ser			
	180	185	190	
	Ala Tyr Ser Asn Trp Gln Ile Leu Lys Glu Ile Leu Gly Lys Met Ile			
	195	200	205	
	Lys Gln Thr Lys Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu			
	210	215	220	
	Leu Glu Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro			
	225	230	235	240
	Ser Phe Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser			
	245	250	255	
	Leu Leu Asp His Asp Arg Thr Val Phe Gln Trp Leu Asp Gln Gln Pro			
	260	265	270	
	Pro Ser Ser Val Leu Tyr Val Ser Phe Gly Ser Thr Ser Glu Val Asp			
	275	280	285	
[0008]	Glu Lys Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys Gln			
	290	295	300	
	Ser Phe Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr Trp			
	305	310	315	320
	Val Glu Pro Leu Pro Asp Gly Phe Leu Gly Glu Arg Gly Arg Ile Val			
	325	330	335	
	Lys Trp Val Pro Gln Gln Glu Val Leu Ala His Gly Ala Ile Gly Ala			
	340	345	350	
	Phe Trp Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys Glu			
	355	360	365	
	Gly Val Pro Met Ile Phe Ser Asp Phe Gly Leu Asp Gln Pro Leu Asn			
	370	375	380	
	Ala Arg Tyr Met Ser Asp Val Leu Lys Val Gly Val Tyr Leu Glu Asn			
	385	390	395	400
	Gly Trp Glu Arg Gly Glu Ile Ala Asn Ala Ile Arg Arg Val Met Val			
	405	410	415	
	Asp Glu Glu Gly Glu Tyr Ile Arg Gln Asn Ala Arg Val Leu Lys Gln			
	420	425	430	
	Lys Ala Asp Val Ser Leu Met Lys Gly Gly Ser Ser Tyr Glu Ser Leu			
	435	440	445	
	Glu Ser Leu Val Ser Tyr Ile Ser Ser Leu			
	450	455		

	65	70	75	80
	Ala Glu Ala Thr Thr Asp Val His Pro Glu Asp Ile Pro Tyr Leu Lys			
		85	90	95
	Lys Ala Ser Asp Gly Leu Gln Pro Glu Val Thr Arg Phe Leu Glu Gln			
	100	105	110	
	His Ser Pro Asp Trp Ile Ile Tyr Asp Tyr Thr His Tyr Trp Leu Pro			
	115	120	125	
	Ser Ile Ala Ala Ser Leu Gly Ile Ser Arg Ala His Phe Ser Val Thr			
	130	135	140	
	Thr Pro Trp Ala Ile Ala Tyr Met Gly Pro Ser Ala Asp Ala Met Ile			
	145	150	155	160
	Asn Gly Ser Asp Gly Arg Thr Thr Val Glu Asp Leu Thr Thr Pro Pro			
	165	170	175	
	Lys Trp Phe Pro Phe Pro Thr Lys Val Cys Trp Arg Lys His Asp Leu			
	180	185	190	
	Ala Arg Leu Val Pro Tyr Lys Ala Pro Gly Ile Ser Asp Gly Tyr Arg			
	195	200	205	
	Met Gly Met Val Leu Lys Gly Ser Asp Cys Leu Leu Ser Lys Cys Tyr			
	210	215	220	
	His Glu Phe Gly Thr Gln Trp Leu Pro Leu Leu Glu Thr Leu His Gln			
	225	230	235	240
[0010]	Val Pro Val Val Pro Val Gly Leu Leu Pro Pro Glu Ile Pro Gly Asp			
	245	250	255	
	Glu Lys Asp Glu Thr Trp Val Ser Ile Lys Lys Trp Leu Asp Gly Lys			
	260	265	270	
	Gln Lys Gly Ser Val Val Tyr Val Ala Leu Gly Ser Glu Ala Leu Val			
	275	280	285	
	Ser Gln Thr Glu Val Val Glu Leu Ala Leu Gly Leu Glu Leu Ser Gly			
	290	295	300	
	Leu Pro Phe Val Trp Ala Tyr Arg Lys Pro Lys Gly Pro Ala Lys Ser			
	305	310	315	320
	Asp Ser Val Glu Leu Pro Asp Gly Phe Val Glu Arg Thr Arg Asp Arg			
	325	330	335	
	Gly Leu Val Trp Thr Ser Trp Ala Pro Gln Leu Arg Ile Leu Ser His			
	340	345	350	
	Glu Ser Val Cys Gly Phe Leu Thr His Cys Gly Ser Gly Ser Ile Val			
	355	360	365	
	Glu Gly Leu Met Phe Gly His Pro Leu Ile Met Leu Pro Ile Phe Gly			
	370	375	380	
	Asp Gln Pro Leu Asn Ala Arg Leu Leu Glu Asp Lys Gln Val Gly Ile			
	385	390	395	400
	Glu Ile Pro Arg Asn Glu Glu Asp Gly Cys Leu Thr Lys Glu Ser Val			
	405	410	415	
	Ala Arg Ser Leu Arg Ser Val Val Val Glu Lys Glu Gly Glu Ile Tyr			

	420	425	430
	Lys Ala Asn Ala Arg Glu Leu Ser Lys Ile Tyr Asn Asp Thr Lys Val		
	435	440	445
	Glu Lys Glu Tyr Val Ser Gln Phe Val Asp Tyr Leu Glu Lys Asn Ala		
	450	455	460
	Arg Ala Val Ala Ile Asp His Glu Ser		
	465	470	
	<210> 10		
	<211> 1422		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成多核苷酸		
	<400> 10		
	atggctacca gtgactccat agttgacgac cgtaagcagc ttcattgttc gacgttccca	60	
	tggttgctt tcggtcacat cctcccttac cttcagcttt cgaattgat agctgaaaag	120	
	ggtcacaaaag tctcgtttct tctaccacc agaaacattc aacgtctctc ttctcatatc	180	
	tcgacctca taaatgttgt tcaactcaca ctccacgtg tccaagagct gccggaggat	240	
	gcagaggcga ccaactgacgt ccaccctgaa gatattccat atctcaagaa ggcttctgat	300	
	ggtcttcaac cggaggctac ccggttcta gaacaacact ctccggactg gattattat	360	
	gattatactc actactggtt gccatccatc gcggctagcc tcggtatctc acgagcccac	420	
	ttctccgtca ccaactccatg ggccattgct tatatgggac cctcagctga cgccatgata	480	
[0011]	aatggttcag atggtcgaac cacggtgag gatctcacga caccgcccga gtggtttccc	540	
	ttccgacca aagtatgctg gcggaagcat gatctgtccc gactggtgcc ttacaaagct	600	
	ccggggatat ctgatggata ccgatgggg atggttctta agggatctga ttgtttgctt	660	
	tccaaatgtt accatgagtt tggaactcaa tggctacctc ttttgagac actacaccaa	720	
	gtaccggtgg ttccggtggg attactgcca ccggaatac ccggagacga gaaagatgaa	780	
	acatgggtgt caatcaagaa atggctcgat ggtaaacaaa aaggcagtggt ggtgtacgtt	840	
	gcattaggaa gcgaggcttt ggtgagccaa accgaggttg ttgagttagc attgggtctc	900	
	gagctttctg ggttgccatt tgtttgggct tatagaaaac caaaagggtcc cgcaagtca	960	
	gactcggtag agttgccaga cgggttcgtg gaacgaactc gtgaccgtgg gttggtctgg	1020	
	acgagttggg cacctcagtt acgaatactg agccatgagt cggtttggg tttcttgact	1080	
	cattgtggtt ctgatcaat tgtggaaggg ctaatgttg gtcaccctct aatcatgcta	1140	
	ccgatttttg gggaccaacc tctgaatgct cgattactgg aggacaaaca gttgggaatc	1200	
	gagataccaa gaaatgagga agatggttc ttgaccaagg agtcggttc tagatcactg	1260	
	aggtccgttg ttgtgaaaaa agaaggggag atctacaagg cgaacgcgag ggagctgagt	1320	
	aaaatctata acgacactaa ggtgaaaaa gaatatgtaa gccaatcgt agactatttg	1380	
	gaaaagaatg cgcgtcggg tgccatcgat catgagagtt aa	1422	
	<210> 11		
	<211> 458		
	<212> PRT		
	<213> 人工序列		
	<220>		
	<223> 合成多肽		

	<400> 11			
	Met Asn Trp Gln Ile Leu Lys Glu Ile Leu Gly Lys Met Ile Lys Gln			
	1	5	10	15
	Thr Lys Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu Leu Glu			
	20	25		30
	Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro Ser Phe			
	35	40	45	
	Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser Leu Leu			
	50	55	60	
	Asp His Asp Arg Thr Val Phe Gln Trp Leu Asp Gln Gln Pro Pro Ser			
	65	70	75	80
	Ser Val Leu Tyr Val Ser Phe Gly Ser Thr Ser Glu Val Asp Glu Lys			
	85	90		95
	Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys Gln Ser Phe			
	100	105		110
	Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr Trp Val Glu			
	115	120	125	
	Pro Leu Pro Asp Gly Phe Leu Gly Glu Arg Gly Arg Ile Val Lys Trp			
	130	135	140	
	Val Pro Gln Gln Glu Val Leu Ala His Gly Ala Ile Gly Ala Phe Trp			
	145	150	155	160
[0012]	Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys Glu Gly Val			
	165	170		175
	Pro Met Ile Phe Ser Asp Phe Gly Leu Asp Gln Pro Leu Asn Ala Arg			
	180	185		190
	Tyr Met Ser Asp Val Leu Lys Val Gly Val Tyr Leu Glu Asn Gly Trp			
	195	200	205	
	Glu Arg Gly Glu Ile Ala Asn Ala Ile Arg Arg Val Met Val Asp Glu			
	210	215	220	
	Glu Gly Glu Tyr Ile Arg Gln Asn Ala Arg Val Leu Lys Gln Lys Ala			
	225	230	235	240
	Asp Val Ser Leu Met Lys Gly Gly Ser Ser Tyr Glu Ser Leu Glu Ser			
	245	250		255
	Leu Val Ser Tyr Ile Ser Ser Leu Glu Asn Lys Thr Glu Thr Thr Val			
	260	265		270
	Arg Arg Arg Arg Arg Ile Ile Leu Phe Pro Val Pro Phe Gln Gly His			
	275	280	285	
	Ile Asn Pro Ile Leu Gln Leu Ala Asn Val Leu Tyr Ser Lys Gly Phe			
	290	295	300	
	Ser Ile Thr Ile Phe His Thr Asn Phe Asn Lys Pro Lys Thr Ser Asn			
	305	310	315	320
	Tyr Pro His Phe Thr Phe Arg Phe Ile Leu Asp Asn Asp Pro Gln Asp			
	325	330		335
	Glu Arg Ile Ser Asn Leu Pro Thr His Gly Pro Leu Ala Gly Met Arg			

	340	345	350
Ile Pro Ile Ile Asn Glu His Gly Ala Asp Glu Leu Arg Arg Glu Leu			
	355	360	365
Glu Leu Leu Met Leu Ala Ser Glu Glu Asp Glu Glu Val Ser Cys Leu			
	370	375	380
Ile Thr Asp Ala Leu Trp Tyr Phe Ala Gln Ser Val Ala Asp Ser Leu			
385	390	395	400
Asn Leu Arg Arg Leu Val Leu Met Thr Ser Ser Leu Phe Asn Phe His			
	405	410	415
Ala His Val Ser Leu Pro Gln Phe Asp Glu Leu Gly Tyr Leu Asp Pro			
	420	425	430
Asp Asp Lys Thr Arg Leu Glu Glu Gln Ala Ser Gly Phe Pro Met Leu			
	435	440	445
Lys Val Lys Asp Ile Lys Ser Ala Tyr Ser			
	450	455	

- <210> 12
- <211> 1377
- <212> DNA
- <213> 人工序列
- <220>
- <223> 合成多核苷酸
- <400> 12

[0013]

atgaactggc aaatcctgaa agaaatcctg ggtaaaatga tcaaacaaac caaagcgtcg	60
tcgggcggtta tctggaactc ctcaaagaa ctggaagaat cagaactgga aaccgttatt	120
cgcgaaatcc cggtccgctc gttcctgatt ccgctgccga aacatctgac cgcgagcagc	180
agcagcctgc tggatcacga ccgtacggtc tttcagtggc tggatcagca accgcctca	240
tcggtgctgt atgtttcatt cggtagcacc tctgaagtcg atgaaaaaga ctttctggaa	300
atcgctcgcg gcctgggtgga tagtaaacag tccttcctgt gggtggttcg tccgggtttt	360
gtgaaaggca gcacgtgggt tgaaccgctg ccggatggct tctgggtga acgcgccgt	420
attgtcaaat gggtgccgca gcaagaagtg ctggcacatg gtgctatcgg cgcgttttgg	480
accactctg gttggaacag tacgctggaa tccgtttgcg aagggtccc gatgatttc	540
agcgattttg gcctggacca gccgctgaat gcccctata tgtctgatgt tctgaaagtc	600
ggtgtgtacc tggaaaacgg ttgggaacgt ggcgaaattg cgaatgcat ccgtcgcgtt	660
atggtcgatg aagaaggcga atacattcgc cagaacgctc gtgtcctgaa acaaaaagcg	720
gacgtgagcc tgatgaaagg cggtagctct tatgaatcac tggaatcgtc ggtagctac	780
atcagttccc tggaaaataa aaccgaaacc acggtcgtc gccctgcgcg tattatcctg	840
ttcccgttc cgtttcaggg tcatattaac ccgacccctg aactggcgaa tgttctgtat	900
tcaaaaggct tttgatcac catcttccat acgaacttca acaaaccgaa aaccagtaac	960
taccgcact ttacgttccg ctttattctg gataacgacc cgcaggatga acgtatctcc	1020
aatctgccga cccacggccc gctggccggt atgcgcattc cgattatcaa tgaacacggt	1080
gcagatgaac tgcgccgtga actggaactg ctgatctgag ccagtgaaga agatgaagaa	1140
gtgtcctgtc tgatcaccga cgcactgtgg tatttcgccc agagcgttgc agattctctg	1200
aacctgcgcc gtctggtcct gatgacgtca tcgctgttca atttcatgc gcacgtttct	1260
ctgcccaat ttgatgaact gggctacctg gaccggatg acaaaaccg tctggaagaa	1320

caagccagtg gtttccgat gctgaaagtc aaagacatta aatccgccta ttcgtaa 1377
 <210> 13
 <211> 475
 <212> PRT
 <213> 人工序列
 <220>
 <223> 合成多肽
 <400> 13
 Met Asn Trp Gln Ile Leu Lys Glu Ile Leu Gly Lys Met Ile Lys Gln
 1 5 10 15
 Thr Lys Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu Leu Glu
 20 25 30
 Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro Ser Phe
 35 40 45
 Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser Leu Leu
 50 55 60
 Asp His Asp Arg Thr Val Phe Gln Trp Leu Asp Gln Gln Pro Pro Ser
 65 70 75 80
 Ser Val Leu Tyr Val Ser Phe Gly Ser Thr Ser Glu Val Asp Glu Lys
 85 90 95
 Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys Gln Ser Phe
 100 105 110
 [0014] Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr Trp Val Glu
 115 120 125
 Pro Leu Pro Asp Gly Phe Leu Gly Glu Arg Gly Arg Ile Val Lys Trp
 130 135 140
 Val Pro Gln Gln Glu Val Leu Ala His Gly Ala Ile Gly Ala Phe Trp
 145 150 155 160
 Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys Glu Gly Val
 165 170 175
 Pro Met Ile Phe Ser Asp Phe Gly Leu Asp Gln Pro Leu Asn Ala Arg
 180 185 190
 Tyr Met Ser Asp Val Leu Lys Val Gly Val Tyr Leu Glu Asn Gly Trp
 195 200 205
 Glu Arg Gly Glu Ile Ala Asn Ala Ile Arg Arg Val Met Val Asp Glu
 210 215 220
 Glu Gly Glu Tyr Ile Arg Gln Asn Ala Arg Val Leu Lys Gln Lys Ala
 225 230 235 240
 Asp Val Ser Leu Met Lys Gly Gly Ser Ser Tyr Glu Ser Leu Glu Ser
 245 250 255
 Leu Val Ser Tyr Ile Ser Ser Leu Tyr Lys Asp Asp Ser Gly Tyr Ser
 260 265 270
 Ser Ser Tyr Ala Ala Ala Ala Gly Met Glu Asn Lys Thr Glu Thr Thr
 275 280 285

Val Arg Arg Arg Arg Arg Ile Ile Leu Phe Pro Val Pro Phe Gln Gly			
290	295	300	
His Ile Asn Pro Ile Leu Gln Leu Ala Asn Val Leu Tyr Ser Lys Gly			
305	310	315	320
Phe Ser Ile Thr Ile Phe His Thr Asn Phe Asn Lys Pro Lys Thr Ser			
325	330		335
Asn Tyr Pro His Phe Thr Phe Arg Phe Ile Leu Asp Asn Asp Pro Gln			
340	345		350
Asp Glu Arg Ile Ser Asn Leu Pro Thr His Gly Pro Leu Ala Gly Met			
355	360	365	
Arg Ile Pro Ile Ile Asn Glu His Gly Ala Asp Glu Leu Arg Arg Glu			
370	375	380	
Leu Glu Leu Leu Met Leu Ala Ser Glu Glu Asp Glu Glu Val Ser Cys			
385	390	395	400
Leu Ile Thr Asp Ala Leu Trp Tyr Phe Ala Gln Ser Val Ala Asp Ser			
405	410		415
Leu Asn Leu Arg Arg Leu Val Leu Met Thr Ser Ser Leu Phe Asn Phe			
420	425		430
His Ala His Val Ser Leu Pro Gln Phe Asp Glu Leu Gly Tyr Leu Asp			
435	440	445	
Pro Asp Asp Lys Thr Arg Leu Glu Glu Gln Ala Ser Gly Phe Pro Met			
450	455	460	
[0015] Leu Lys Val Lys Asp Ile Lys Ser Ala Tyr Ser			
465	470	475	
<210> 14			
<211> 1428			
<212> DNA			
<213> 人工序列			
<220>			
<223> 合成多核苷酸			
<400> 14			
atgaactggc aaatcctgaa agaaatcctg ggtaaaatga tcaaacaaac caaagcgtcg			60
tgggctgta tctggaactc ctcaaagaa ctggaagaat cagaactgga aaccgttatt			120
cgcgaaatcc cggctccgtc gttcctgatt ccgctgccga aacatctgac cgcgagcagc			180
agcagcctgc tggatcacga ccgtacggtc tttcagtggc tggatcagca accgccgtca			240
tcggtgctgt atgtttcatt cggtagcacc tctgaagtcg atgaaaaaga ctttctggaa			300
atcgtctcgc gcctggtgga tagtaaacag tccttctgtt gggtggttcg tccgggtttt			360
gtgaaaggca gcacgtgggt tgaaccgctg ccggatggct tcttgggtga acgcgccct			420
attgtcaaat gggcgccgca gcaagaagtg ctggcacatg gtgctatcgg cgcgttttgg			480
accactctg gttggaacag tacgctggaa tccgtttcgc aagggtgcc gatgatcttc			540
agcgattttg gcctggacca gccgctgaat gcccgctata tgtctgatgt tctgaaagtc			600
ggtgtgtacc tggaaaacgg ttgggaactg ggcgaaattg cgaatccat ccgtcgcgtt			660
atggtcgtat aagaaggcga atacattcgc cagaacgctc gtgtcctgaa acaaaaagcg			720
gacgtgagcc tgatgaaagg cggtagctct tatgaatcac tggaatcgt ggttagctac			780

	Lys Gln Thr Lys Ala Ser Ser Gly Val Ile Trp Asn Ser Phe Lys Glu		
	210	215	220
	Leu Glu Glu Ser Glu Leu Glu Thr Val Ile Arg Glu Ile Pro Ala Pro		
	225	230	235
	Ser Phe Leu Ile Pro Leu Pro Lys His Leu Thr Ala Ser Ser Ser Ser		240
	245	250	255
	Leu Leu Asp His Asp Arg Thr Val Phe Gln Trp Leu Asp Gln Gln Pro		
	260	265	270
	Pro Ser Ser Val Leu Tyr Val Ser Phe Gly Ser Thr Ser Glu Val Asp		
	275	280	285
	Glu Lys Asp Phe Leu Glu Ile Ala Arg Gly Leu Val Asp Ser Lys Gln		
	290	295	300
	Ser Phe Leu Trp Val Val Arg Pro Gly Phe Val Lys Gly Ser Thr Trp		
	305	310	315
	Val Glu Pro Leu Pro Asp Gly Phe Leu Gly Glu Arg Gly Arg Ile Val		320
	325	330	335
	Lys Trp Val Pro Gln Gln Glu Val Leu Ala His Gly Ala Ile Gly Ala		
	340	345	350
	Phe Trp Thr His Ser Gly Trp Asn Ser Thr Leu Glu Ser Val Cys Glu		
	355	360	365
	Gly Val Pro Met Ile Phe Ser Asp Phe Gly Leu Asp Gln Pro Leu Asn		
[0017]	370	375	380
	Ala Arg Tyr Met Ser Asp Val Leu Lys Val Gly Val Tyr Leu Glu Asn		
	385	390	395
	Gly Trp Glu Arg Gly Glu Ile Ala Asn Ala Ile Arg Arg Val Met Val		400
	405	410	415
	Asp Glu Glu Gly Glu Tyr Ile Arg Gln Asn Ala Arg Val Leu Lys Gln		
	420	425	430
	Lys Ala Asp Val Ser Leu Met Lys Gly Gly Ser Ser Tyr Glu Ser Leu		
	435	440	445
	Glu Ser Leu Val Ser Tyr Ile Ser Ser Leu Gly Ser Gly Ala Asn Ala		
	450	455	460
	Glu Arg Met Ile Thr Arg Val His Ser Gln Arg Glu Arg Leu Asn Glu		
	465	470	475
	Thr Leu Val Ser Glu Arg Asn Glu Val Leu Ala Leu Leu Ser Arg Val		480
	485	490	495
	Glu Ala Lys Gly Lys Gly Ile Leu Gln Gln Asn Gln Ile Ile Ala Glu		
	500	505	510
	Phe Glu Ala Leu Pro Glu Gln Thr Arg Lys Lys Leu Glu Gly Gly Pro		
	515	520	525
	Phe Phe Asp Leu Leu Lys Ser Thr Gln Glu Ala Ile Val Leu Pro Pro		
	530	535	540
	Trp Val Ala Leu Ala Val Arg Pro Arg Pro Gly Val Trp Glu Tyr Leu		
	545	550	555
			560

	Arg Val Asn Leu His Ala Leu Val Val Glu Glu Leu Gln Pro Ala Glu		
	565	570	575
	Phe Leu His Phe Lys Glu Glu Leu Val Asp Gly Val Lys Asn Gly Asn		
	580	585	590
	Phe Thr Leu Glu Leu Asp Phe Glu Pro Phe Asn Ala Ser Ile Pro Arg		
	595	600	605
	Pro Thr Leu His Lys Tyr Ile Gly Asn Gly Val Asp Phe Leu Asn Arg		
	610	615	620
	His Leu Ser Ala Lys Leu Phe His Asp Lys Glu Ser Leu Leu Pro Leu		
	625	630	635
	Leu Lys Phe Leu Arg Leu His Ser His Gln Gly Lys Asn Leu Met Leu		
	645	650	655
	Ser Glu Lys Ile Gln Asn Leu Asn Thr Leu Gln His Thr Leu Arg Lys		
	660	665	670
	Ala Glu Glu Tyr Leu Ala Glu Leu Lys Ser Glu Thr Leu Tyr Glu Glu		
	675	680	685
	Phe Glu Ala Lys Phe Glu Glu Ile Gly Leu Glu Arg Gly Trp Gly Asp		
	690	695	700
	Asn Ala Glu Arg Val Leu Asp Met Ile Arg Leu Leu Leu Asp Leu Leu		
	705	710	715
	Glu Ala Pro Asp Pro Cys Thr Leu Glu Thr Phe Leu Gly Arg Val Pro		
	725	730	735
[0018]	Met Val Phe Asn Val Val Ile Leu Ser Pro His Gly Tyr Phe Ala Gln		
	740	745	750
	Asp Asn Val Leu Gly Tyr Pro Asp Thr Gly Gly Gln Val Val Tyr Ile		
	755	760	765
	Leu Asp Gln Val Arg Ala Leu Glu Ile Glu Met Leu Gln Arg Ile Lys		
	770	775	780
	Gln Gln Gly Leu Asn Ile Lys Pro Arg Ile Leu Ile Leu Thr Arg Leu		
	785	790	795
	Leu Pro Asp Ala Val Gly Thr Thr Cys Gly Glu Arg Leu Glu Arg Val		
	805	810	815
	Tyr Asp Ser Glu Tyr Cys Asp Ile Leu Arg Val Pro Phe Arg Thr Glu		
	820	825	830
	Lys Gly Ile Val Arg Lys Trp Ile Ser Arg Phe Glu Val Trp Pro Tyr		
	835	840	845
	Leu Glu Thr Tyr Thr Glu Asp Ala Ala Val Glu Leu Ser Lys Glu Leu		
	850	855	860
	Asn Gly Lys Pro Asp Leu Ile Ile Gly Asn Tyr Ser Asp Gly Asn Leu		
	865	870	875
	Val Ala Ser Leu Leu Ala His Lys Leu Gly Val Thr Gln Cys Thr Ile		
	885	890	895
	Ala His Ala Leu Glu Lys Thr Lys Tyr Pro Asp Ser Asp Ile Tyr Trp		
	900	905	910

Lys Lys Leu Asp Asp Lys Tyr His Phe Ser Cys Gln Phe Thr Ala Asp	915	920	925
Ile Phe Ala Met Asn His Thr Asp Phe Ile Ile Thr Ser Thr Phe Gln	930	935	940
Glu Ile Ala Gly Ser Lys Glu Thr Val Gly Gln Tyr Glu Ser His Thr	945	950	955
Ala Phe Thr Leu Pro Gly Leu Tyr Arg Val Val His Gly Ile Asp Val	965	970	975
Phe Asp Pro Lys Phe Asn Ile Val Ser Pro Gly Ala Asp Met Ser Ile	980	985	990
Tyr Phe Pro Tyr Thr Glu Glu Lys Arg Arg Leu Thr Lys Phe His Ser	995	1000	1005
Glu Ile Glu Glu Leu Leu Tyr Ser Asp Val Glu Asn Lys Glu His	1010	1015	1020
Leu Cys Val Leu Lys Asp Lys Lys Lys Pro Ile Leu Phe Thr Met	1025	1030	1035
Ala Arg Leu Asp Arg Val Lys Asn Leu Ser Gly Leu Val Glu Trp	1040	1045	1050
Tyr Gly Lys Asn Thr Arg Leu Arg Glu Leu Ala Asn Leu Val Val	1055	1060	1065
Val Gly Gly Asp Arg Arg Lys Glu Ser Lys Asp Asn Glu Glu Lys	1070	1075	1080
Ala Glu Met Lys Lys Met Tyr Asp Leu Ile Glu Glu Tyr Lys Leu	1085	1090	1095
Asn Gly Gln Phe Arg Trp Ile Ser Ser Gln Met Asp Arg Val Arg	1100	1105	1110
Asn Gly Glu Leu Tyr Arg Tyr Ile Cys Asp Thr Lys Gly Ala Phe	1115	1120	1125
Val Gln Pro Ala Leu Tyr Glu Ala Phe Gly Leu Thr Val Val Glu	1130	1135	1140
Ala Met Thr Cys Gly Leu Pro Thr Phe Ala Thr Cys Lys Gly Gly	1145	1150	1155
Pro Ala Glu Ile Ile Val His Gly Lys Ser Gly Phe His Ile Asp	1160	1165	1170
Pro Tyr His Gly Asp Gln Ala Ala Asp Thr Leu Ala Asp Phe Phe	1175	1180	1185
Thr Lys Cys Lys Glu Asp Pro Ser His Trp Asp Glu Ile Ser Lys	1190	1195	1200
Gly Gly Leu Gln Arg Ile Glu Glu Lys Tyr Thr Trp Gln Ile Tyr	1205	1210	1215
Ser Gln Arg Leu Leu Thr Leu Thr Gly Val Tyr Gly Phe Trp Lys	1220	1225	1230
His Val Ser Asn Leu Asp Arg Leu Glu Ala Arg Arg Tyr Leu Glu	1235	1240	1245

[0019]

	Met Phe Tyr Ala Leu Lys Tyr Arg Pro Leu Ala Gln Ala Val Pro	
	1250	1255 1260
	Leu Ala Gln Asp Asp	
	1265	
	<210> 16	
	<211> 3807	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成多核苷酸	
	<400> 16	
	atggagaata agacagaaac aaccgtaaga cggaggcggga ggattatctt gttccctgta	60
	ccatttcagg gccatattaa tccgatcctc caattagcaa acgtcctcta ctccaaggga	120
	ttttcaataa caatcttcca tactaacttt aacaagccta aaacgagtaa ttatcctcac	180
	tttaccattca ggttcattct agacaacgac cctcaggatg agcgtatctc aaatttacct	240
	acgcatggcc cttggcagg tatgcgaata ccaataatca atgagcatgg agccgatgaa	300
	ctccgtcgcg agttagagct tctcatgctc gcaagtgagg aagacgagga agtttcgtgc	360
	ctaataactg atgcgctttg gtacttcgcc caatcagtcg cagactcact gaatctacgc	420
	cgtttggtcc ttatgacaag ttattatc aactttcacg cacatgtatc actgccgcaa	480
	tttgacgagt tgggttacct ggaccggat gacaaaacgc gattggagga acaagcgtcg	540
	ggcttcccca tgctgaaagt caaatatatt aagagcgtt atagtaattg gcaaattctg	600
[0020]	aaagaaattc tcggaanaat gataaagcaa accaaagcgt cctctggagt aatctggaac	660
	tccttcaagg agttagagga atctgaactt gaaacggtca tcagagaaat ccccgtccc	720
	tcgttcttaa ttccactacc caagcacctt actgcaagta gcagttcct ctagatcat	780
	gaccgaaccg tgtttcagtg gctggatcag caacccccgt cgtcagttct atatgtaagc	840
	ttgggagta cttcggaaagt ggatgaaaag gacttcttag agattgcgcg agggctcgtg	900
	gatagcaaac agagcttctt gtggtagtg agaccgggat tcgttaaggg ctgcacgtgg	960
	gtcagaccgt tgccagatgg tttctaggg gagagagggga gaatcgtgaa atgggttcca	1020
	cagcaagagg ttttgctca cggagctata ggggcctttt ggaccactc tggttggaat	1080
	tctactcttg aaagtgtctg tgaaggcgtt ccaatgatat tttctgattt tgggcttgac	1140
	cagcctctaa acgctcgtca tatgtctgat gtgtgaagg ttggcgtgta cctggagaat	1200
	ggttgggaaa ggggggaaaat tgccaacgcc atacccggg taatggtgga cgaggaaggt	1260
	gagtacatac gtcagaacgc tcgggttta aaacaaaaag cggacgtcag cttatgaag	1320
	ggaggtagct cctatgaatc ctagaatcc ttgtaagct atatatcttc gttaggttct	1380
	ggtgcaaacg ctgaacgtat gataacgcgc gtccacagcc aacgtgagcg ttgaaacgaa	1440
	acgctgtttt ctgagagaaa cgaagtctt gccttgcttt ccagggttga agccaaaggt	1500
	aaaggtattt tacaacaaa ccagatcatt gctgaattcg aagctttgcc tgaacaaacc	1560
	cggaagaaac ttgaaggtgg tcctttcttt gaccttctca aatccactca ggaagcaatt	1620
	gtgttccac catgggttc tctagctgtg aggccaaggc ctggtgttg ggaatactta	1680
	cgagtcaatc tccatgctct tctcgttgaa gaactccaac ctgctgagtt tcttcatttc	1740
	aaggaagaac tcgttgatgg agttaagaat ggtaattca ctcttgagct tgatttcgag	1800
	ccattcaatg cgtctatccc tcgtccaaca ctccacaaat acattgaaa tgggttgac	1860
	ttccttaacc gtcattatc ggctaagctc ttccatgaca aggagagttt gcttcattg	1920
	cttaagtcc ttcgtctca cagccaccag ggcaagaacc tgatgttgag cgagaagatt	1980

	Glu Gly Gly Pro Phe Phe Asp Leu Leu Lys Ser Thr Gln Glu Ala Ile		
	65	70	75 80
	Val Leu Pro Pro Trp Val Ala Leu Ala Val Arg Pro Arg Pro Gly Val		
		85	90 95
	Trp Glu Tyr Leu Arg Val Asn Leu His Ala Leu Val Val Glu Glu Leu		
		100	105 110
	Gln Pro Ala Glu Phe Leu His Phe Lys Glu Glu Leu Val Asp Gly Val		
		115	120 125
	Lys Asn Gly Asn Phe Thr Leu Glu Leu Asp Phe Glu Pro Phe Asn Ala		
		130	135 140
	Ser Ile Pro Arg Pro Thr Leu His Lys Tyr Ile Gly Asn Gly Val Asp		
	145	150	155 160
	Phe Leu Asn Arg His Leu Ser Ala Lys Leu Phe His Asp Lys Glu Ser		
		165	170 175
	Leu Leu Pro Leu Leu Lys Phe Leu Arg Leu His Ser His Gln Gly Lys		
		180	185 190
	Asn Leu Met Leu Ser Glu Lys Ile Gln Asn Leu Asn Thr Leu Gln His		
		195	200 205
	Thr Leu Arg Lys Ala Glu Glu Tyr Leu Ala Glu Leu Lys Ser Glu Thr		
		210	215 220
	Leu Tyr Glu Glu Phe Glu Ala Lys Phe Glu Glu Ile Gly Leu Glu Arg		
[0022]	225	230	235 240
	Gly Trp Gly Asp Asn Ala Glu Arg Val Leu Asp Met Ile Arg Leu Leu		
		245	250 255
	Leu Asp Leu Leu Glu Ala Pro Asp Pro Cys Thr Leu Glu Thr Phe Leu		
		260	265 270
	Gly Arg Val Pro Met Val Phe Asn Val Val Ile Leu Ser Pro His Gly		
		275	280 285
	Tyr Phe Ala Gln Asp Asn Val Leu Gly Tyr Pro Asp Thr Gly Gly Gln		
		290	295 300
	Val Val Tyr Ile Leu Asp Gln Val Arg Ala Leu Glu Ile Glu Met Leu		
	305	310	315 320
	Gln Arg Ile Lys Gln Gln Gly Leu Asn Ile Lys Pro Arg Ile Leu Ile		
		325	330 335
	Leu Thr Arg Leu Leu Pro Asp Ala Val Gly Thr Thr Cys Gly Glu Arg		
		340	345 350
	Leu Glu Arg Val Tyr Asp Ser Glu Tyr Cys Asp Ile Leu Arg Val Pro		
		355	360 365
	Phe Arg Thr Glu Lys Gly Ile Val Arg Lys Trp Ile Ser Arg Phe Glu		
		370	375 380
	Val Trp Pro Tyr Leu Glu Thr Tyr Thr Glu Asp Ala Ala Val Glu Leu		
	385	390	395 400
	Ser Lys Glu Leu Asn Gly Lys Pro Asp Leu Ile Ile Gly Asn Tyr Ser		
		405	410 415

	Asp Gly Asn Leu Val Ala Ser Leu Leu Ala His Lys Leu Gly Val Thr		
	420	425	430
	Gln Cys Thr Ile Ala His Ala Leu Glu Lys Thr Lys Tyr Pro Asp Ser		
	435	440	445
	Asp Ile Tyr Trp Lys Lys Leu Asp Asp Lys Tyr His Phe Ser Cys Gln		
	450	455	460
	Phe Thr Ala Asp Ile Phe Ala Met Asn His Thr Asp Phe Ile Ile Thr		
	465	470	475
	Ser Thr Phe Gln Glu Ile Ala Gly Ser Lys Glu Thr Val Gly Gln Tyr		
	485	490	495
	Glu Ser His Thr Ala Phe Thr Leu Pro Gly Leu Tyr Arg Val Val His		
	500	505	510
	Gly Ile Asp Val Phe Asp Pro Lys Phe Asn Ile Val Ser Pro Gly Ala		
	515	520	525
	Asp Met Ser Ile Tyr Phe Pro Tyr Thr Glu Glu Lys Arg Arg Leu Thr		
	530	535	540
	Lys Phe His Ser Glu Ile Glu Glu Leu Leu Tyr Ser Asp Val Glu Asn		
	545	550	555
	Lys Glu His Leu Cys Val Leu Lys Asp Lys Lys Lys Pro Ile Leu Phe		
	565	570	575
	Thr Met Ala Arg Leu Asp Arg Val Lys Asn Leu Ser Gly Leu Val Glu		
	580	585	590
[0023]	Trp Tyr Gly Lys Asn Thr Arg Leu Arg Glu Leu Ala Asn Leu Val Val		
	595	600	605
	Val Gly Gly Asp Arg Arg Lys Glu Ser Lys Asp Asn Glu Glu Lys Ala		
	610	615	620
	Glu Met Lys Lys Met Tyr Asp Leu Ile Glu Glu Tyr Lys Leu Asn Gly		
	625	630	635
	Gln Phe Arg Trp Ile Ser Ser Gln Met Asp Arg Val Arg Asn Gly Glu		
	645	650	655
	Leu Tyr Arg Tyr Ile Cys Asp Thr Lys Gly Ala Phe Val Gln Pro Ala		
	660	665	670
	Leu Tyr Glu Ala Phe Gly Leu Thr Val Val Glu Ala Met Thr Cys Gly		
	675	680	685
	Leu Pro Thr Phe Ala Thr Cys Lys Gly Gly Pro Ala Glu Ile Ile Val		
	690	695	700
	His Gly Lys Ser Gly Phe His Ile Asp Pro Tyr His Gly Asp Gln Ala		
	705	710	715
	Ala Asp Thr Leu Ala Asp Phe Phe Thr Lys Cys Lys Glu Asp Pro Ser		
	725	730	735
	His Trp Asp Glu Ile Ser Lys Gly Gly Leu Gln Arg Ile Glu Glu Lys		
	740	745	750
	Tyr Thr Trp Gln Ile Tyr Ser Gln Arg Leu Leu Thr Leu Thr Gly Val		
	755	760	765

	Tyr Gly Phe Trp Lys His Val Ser Asn Leu Asp Arg Leu Glu Ala Arg			
	770	775		
	Arg Tyr Leu Glu Met Phe Tyr Ala Leu Lys Tyr Arg Pro Leu Ala Gln	780		
	785	790	795	800
	Ala Val Pro Leu Ala Gln Asp Asp			
	805			
	<210> 18			
	<211> 2427			
	<212> DNA			
	<213> 拟南芥			
	<400> 18			
	atggcaaacg ctgaacgtat gattaccctg gtccactccc aacgcgaacg cctgaacgaa	60		
	accctggtgt cggaacgcaa cgaagtctgt gcaactgctga gccgtgtgga agctaagggc	120		
	aaaggtattc tgcagcaaaa ccagattatc gcggaatttg aagccctgcc ggaacaaacc	180		
	cgcaaaaagc tggaggcgcg tccgttttc gatctgctga aatctacgca ggaagcgatc	240		
	gttctgccgc cgtgggtcgc actggcagtg cgtccgcgtc cggcgcttg ggaatatctg	300		
	cgtgtcaacc tgcattgact ggtggtgaa gaactgcagc cggctgaatt tctgacttc	360		
	aaggaagaac tggttgacgg cgtcaaaaac ggtaatttta cctggaact ggatttgaa	420		
	ccgttcaatg ccagtatccc gcgtccgacg ctgcataaat atattggcaa cggtgtggac	480		
	tttctgaatc gccatctgag cgcaaaagct ttccacgata aagaatctct gctgccgctg	540		
	ctgaaattcc tgcgtctgca tagtcaccag ggcaagaacc tgatgctgtc cgaaaaaatt	600		
[0024]	cagaacctga ataccctgca acacacgctg cgcaaggcgg aagaatacct ggccgaactg	660		
	aaaagtgaac ccctgtacga agaattcga gcaaaagttc aagaaattgg cctggaactg	720		
	ggctgggggt acaatgctga acgtgtctg gatgatgacc gtctgctgtg ggacctgctg	780		
	gaagcaccgg acccgtgcac cctggaaacg tttctgggtc gcgtgccgat ggtttcaac	840		
	gtcgtgattc tgcctccgca tggctatctt gcacaggaca atgtgctggg ttaccggat	900		
	accggcggtc aggttgtcta tattctggat caagttcgtg cgctggaaat tgaatgctg	960		
	cagcgcatac agcagcaagg cctgaacatc aaaccgcgta ttctgatcct gaccgtctg	1020		
	ctgccggatg cagttggtac cacgtcgggt gaacgtctgg aacgcgtcta tgacagcgaa	1080		
	tactgtgata ttctgcgtgt cccgtttcgc accgaaaagg gtattgtgctg taaatggatc	1140		
	agtcgcttcg aagtttgcc gtatctggaa acctacacgg aagatgcggc cgtggaactg	1200		
	tccaaggaac tgaatggcaa accggacctg attatcggca actatagcga tggtaatctg	1260		
	gtcgcattct tgcggctca taaactgggt gtgaccagtg gcacgattgc acacgtctg	1320		
	gaaaagacca aatatccgga ttcagacatc tactggaaaa agctggatga caaatatcat	1380		
	ttttcgtgac agttaccgc ggacattttt gccatgaacc acacggattt tattatcacc	1440		
	agtacgttcc aggaaatcgc gggctccaaa gaaaccgtgg gtcaatacga atcacatacc	1500		
	gccttcacgc tgccgggct gtatcgtgtg gttcacggta tcgatgtttt tgaccgaaa	1560		
	ttcaatattg tcagtccggg cgcggatag tccatctatt ttccgtacac cgaagaaaag	1620		
	cgtcgcctga cgaaattcca ttcagaaatt gaagaactgc tgtactcggc cgtggaaaac	1680		
	aaggaacacc tgtgtgttct gaaagataaa aagaaaccga tctgtttac catggccctg	1740		
	ctggatcgcg tgaagaatct gtcaggcctg gttgaatggt atggtaaaaa cacgcgtctg	1800		
	cgcaactgg caaatctggt cgtggtggc ggtgaccgtc gcaaggaatc gaaagataac	1860		
	gaagaaaagg ctgaaatgaa gaaatgtac gatctgatcg aagaatacaa gctgaacggc	1920		
	cagtttctgt ggatcagctc tcaaatggac cgtgtcgcga atggcgaact gtatcgctac	1980		

	245	250	255
	Glu Asp Asp Gly Ala Ile Ile Arg Trp Leu Asp Glu Gln Pro Ala Lys		
	260	265	270
	Ser Val Val Tyr Val Ala Leu Gly Ser Glu Ala Pro Val Ser Ala Asp		
	275	280	285
	Leu Leu Arg Glu Leu Ala His Gly Leu Glu Leu Ala Gly Thr Arg Phe		
	290	295	300
	Leu Trp Ala Leu Arg Arg Pro Ala Gly Val Asn Asp Gly Asp Ser Ile		
	305	310	315
	Leu Pro Asn Gly Phe Leu Glu Arg Thr Gly Glu Arg Gly Leu Val Thr		
	325	330	335
	Thr Gly Trp Val Pro Gln Val Ser Ile Leu Ala His Ala Ala Val Cys		
	340	345	350
	Ala Phe Leu Thr His Cys Gly Trp Gly Ser Val Val Glu Gly Leu Gln		
	355	360	365
	Phe Gly His Pro Leu Ile Met Leu Pro Ile Ile Gly Asp Gln Gly Pro		
	370	375	380
	Asn Ala Arg Phe Leu Glu Gly Arg Lys Val Gly Val Ala Val Pro Arg		
	385	390	395
	Asn His Ala Asp Gly Ser Phe Asp Arg Ser Gly Val Ala Gly Ala Val		
	405	410	415
[0026]	Arg Ala Val Ala Val Glu Glu Glu Gly Lys Ala Phe Ala Ala Asn Ala		
	420	425	430
	Arg Lys Leu Gln Glu Ile Val Ala Asp Arg Glu Arg Asp Glu Arg Cys		
	435	440	445
	Thr Asp Gly Phe Ile His His Leu Thr Ser Trp Asn Glu Leu Glu Ala		
	450	455	460
	<210> 20		
	<211> 1395		
	<212> DNA		
	<213> 短花药野生稻		
	<400> 20		
	atggaaaatg gtagcagtc gctgcatggt gttatttttc cgtggctggc atttggcat 60		
	ctgctgccgt ttctggatct ggcagaacgt ctggcagcac gtggtcatcg ttttagcttt 120		
	gtagcacac cgcgtaatct ggcacgtctg cgtccggttc gtccggcact gctggtctg 180		
	gttgatctgg ttgactgcc gctgcctcgt gttcatggtc tgccggatgg tgccgaagca 240		
	accagtgatg ttccgttga aaaattgaa ctgcaccgca aagcatttga tggcctggct 300		
	gcaccgttta ggcatttct ggatgcagca tgtgccggtg ataaacttcc ggattgggtt 360		
	attccggatt ttatgcatta ttgggttga gcagcagcac agaaacttgg tttccgtgt 420		
	gcagttctga ttccgttag cgagatggt atggcactgt atggtcagcc gaccgaaacc 480		
	agcaccgaac agccggaagc aattgcacgt agcatggcag cagaagcacc gagctttgaa 540		
	gcagaacgta ataccgaaga atatggtaca gccggtgcaa gcggtgtag cattatgacc 600		
	cgttttagtc tgaccctgaa atggtcaaaa ctggttcccc tgcgtagctg tccggaactg 660		
	gaaccgggtg ttttaccac actgaccctg gtttatagca aaccggttgt gccgtttggt 720		

	ctgctgcctc cgcgtcgtga tggcgacat ggtgttcgta aaaatggtga agatgatggt	780
	gccattattc gttggctgga tgaacagcct gcaaaaagcg ttgtttatgt tgcactgggt	840
	agcgaagcac cggtttcagc cgatctgctg cgtgaactgg cacatggtct ggaattagca	900
	ggcaccggtt ttctgtgggc tctgcgtcgt cctgccggtg ttaatgatgg tgatagcatt	960
	ctgccgaatg gttttctgga acgtaccggt gaacgcggtc tggttaccac cggttgggtt	1020
[0027]	ccgcaggtta gtattctggc ccatgcagca gtttgcgat ttctgacca ttgtggttg	1080
	ggtagcgttg ttgaaggttt acagtttggc catccgctga ttatgctgcc gattattggt	1140
	gatcagggtc cgaatgcagc ctttctgga ggtcgtaaag ttggtgttgc agttccgct	1200
	aacatgcag atggtagctt tgatcgtagc ggtgttgccg gtgccgttcg tgcagttgca	1260
	gttgaagaag aaggtaaagc ctttgcagca aatgcccgta aactgcaaga aattgttga	1320
	gatcgtgaac tgatgaacg ttgtaccgat ggtttattc atcatctgac cagctggaat	1380
	gaactggaag cataa	1395

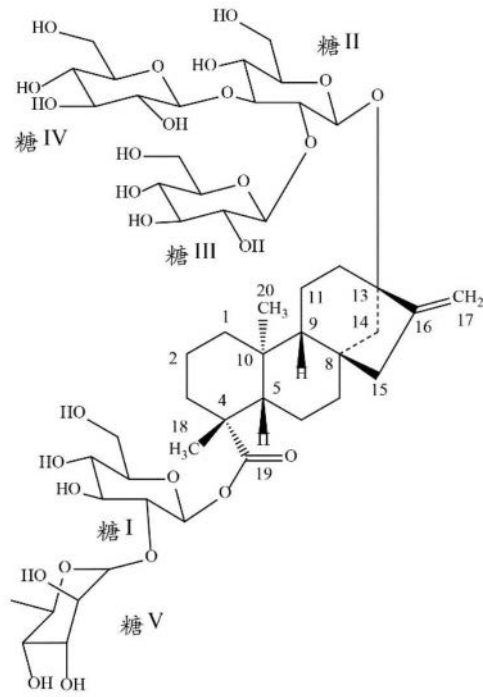


图1

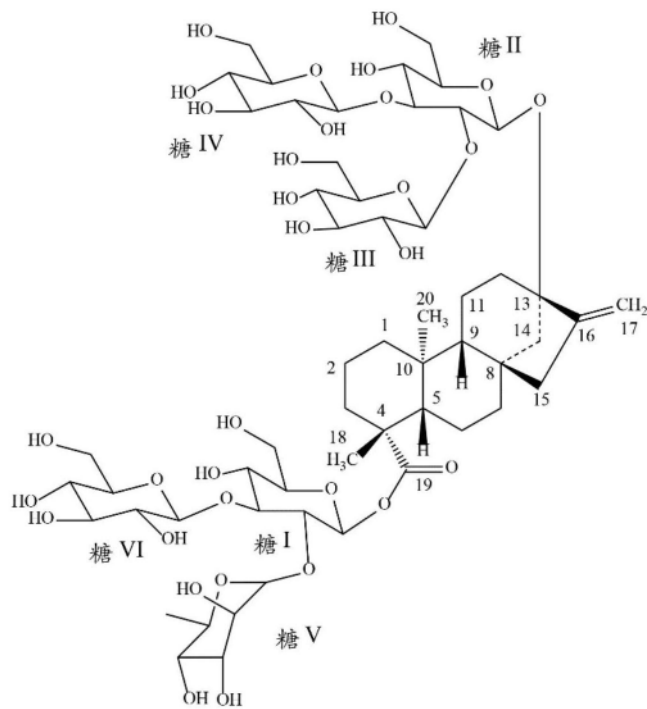


图2

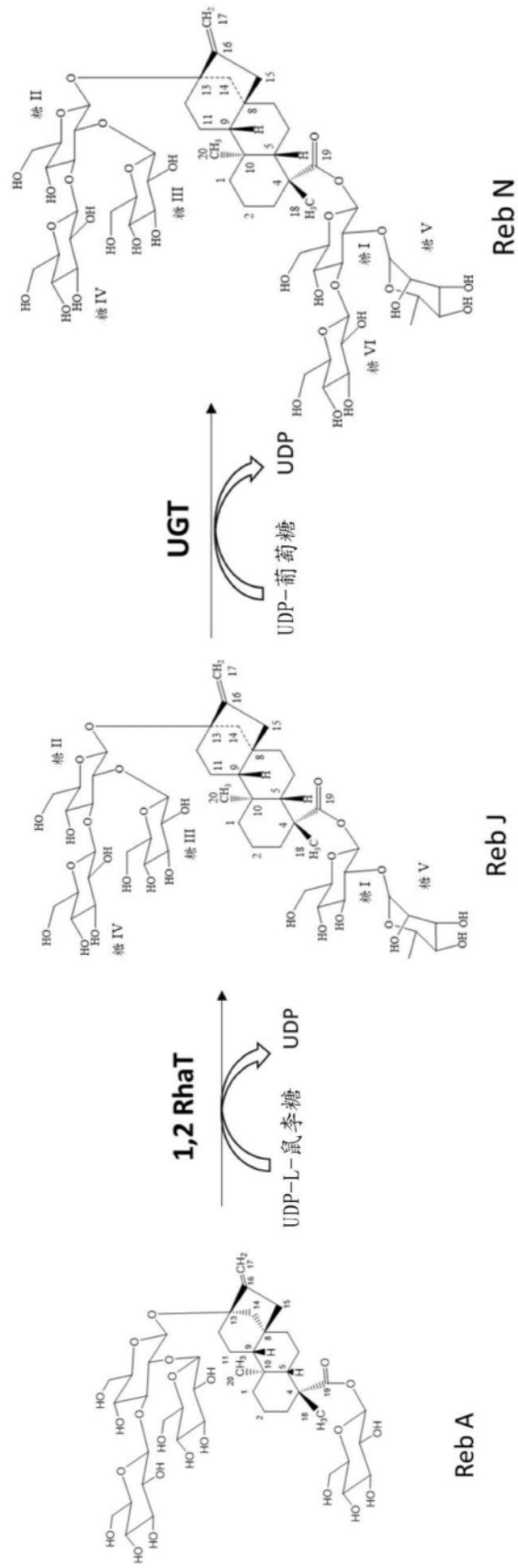


图3A

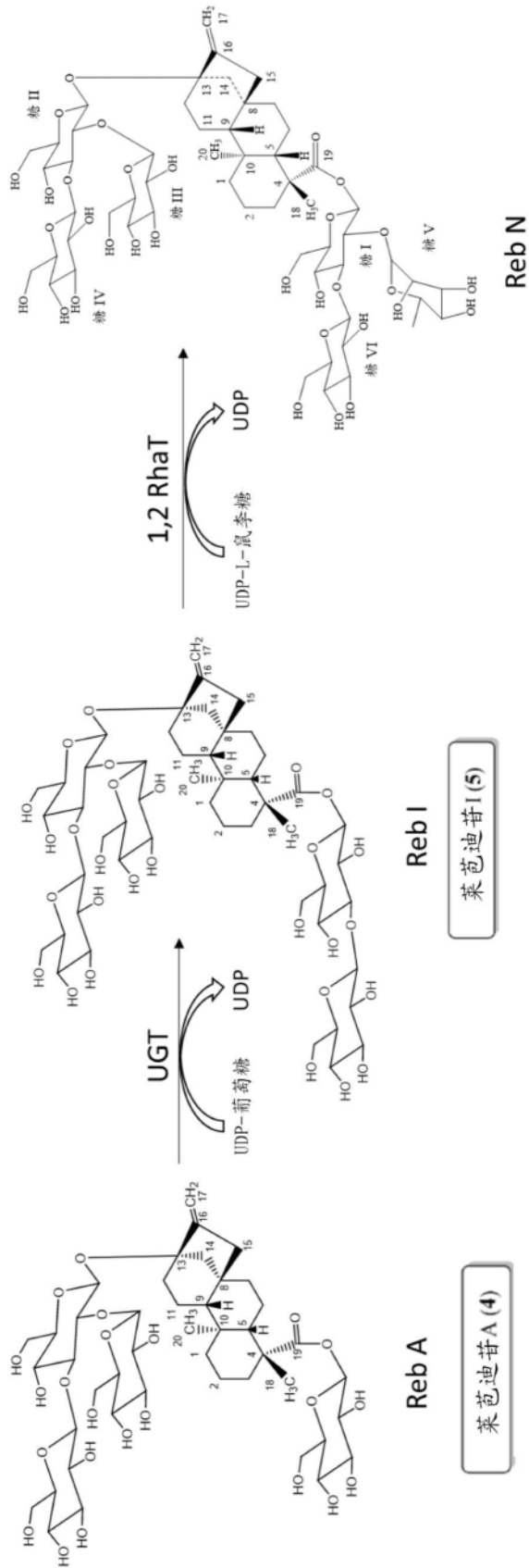


图3B

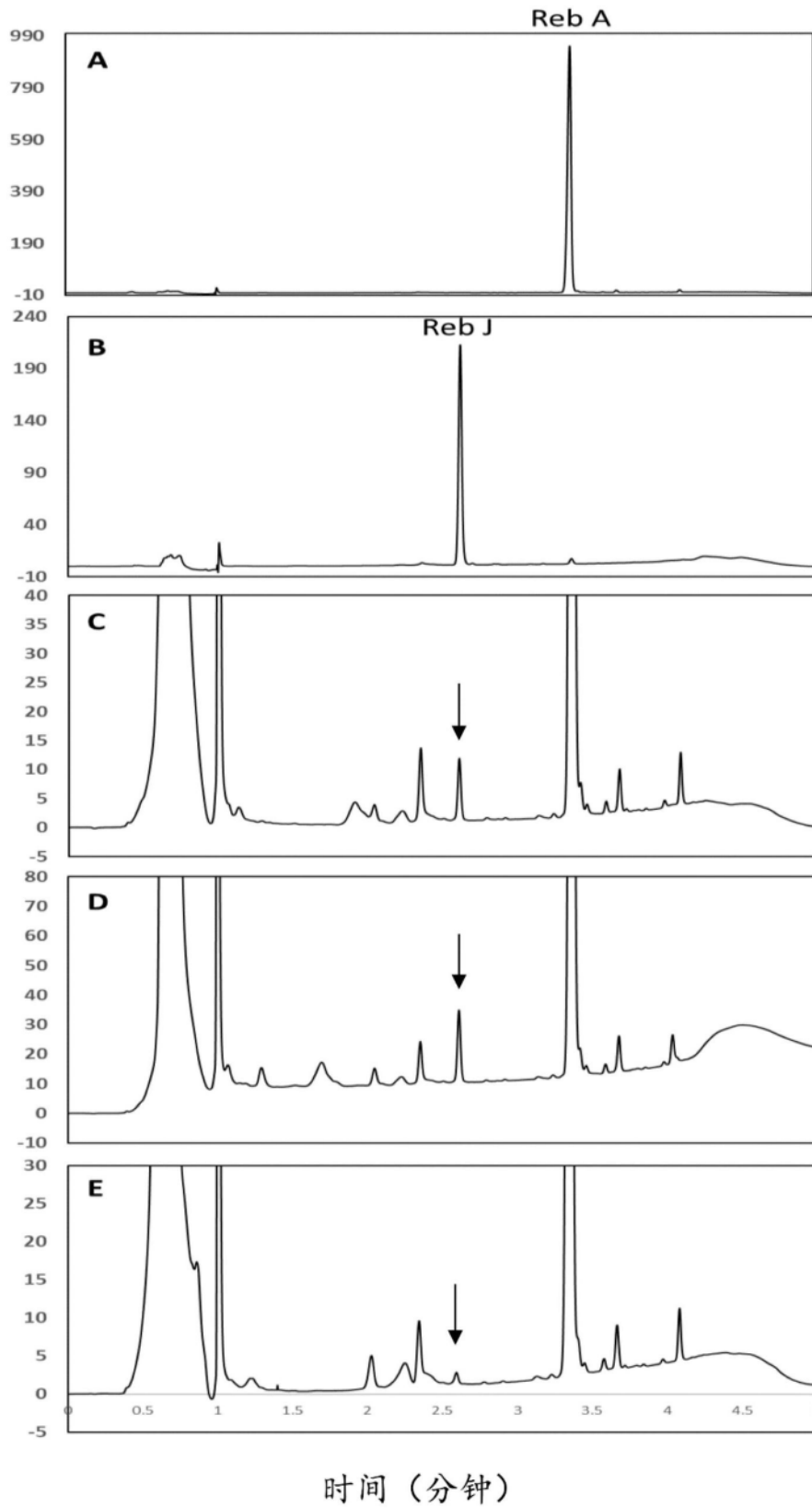


图4

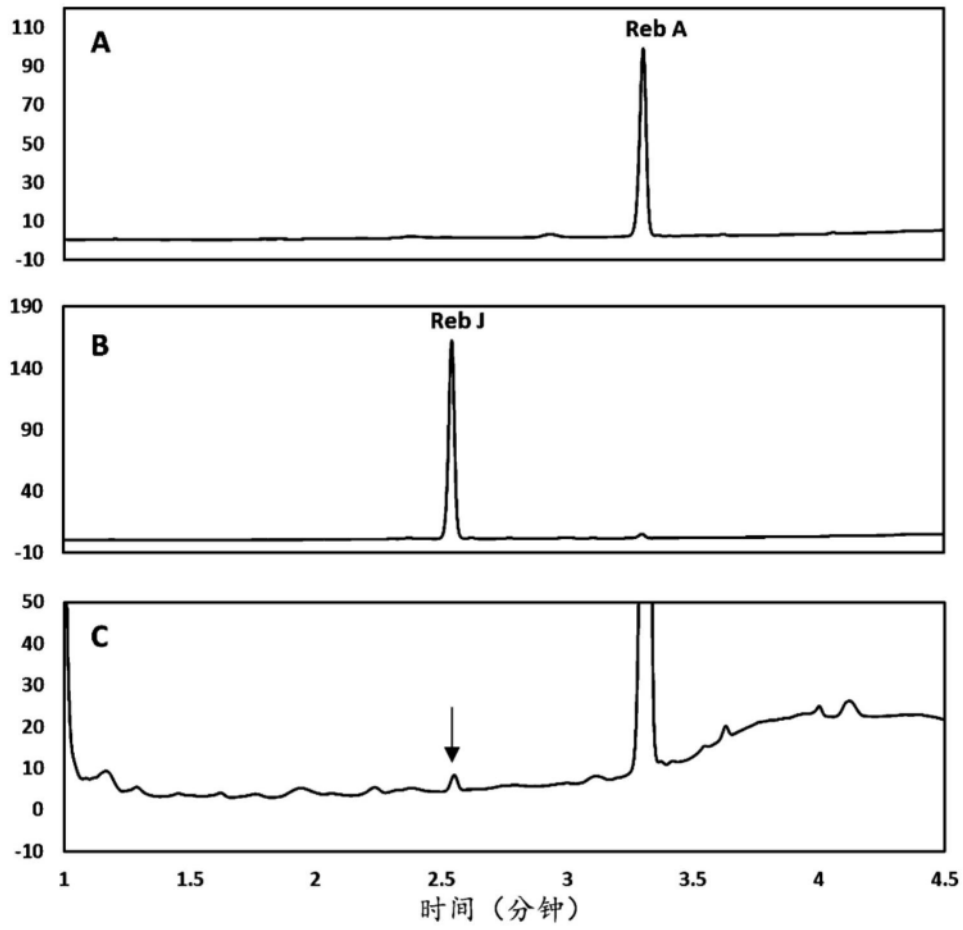


图5

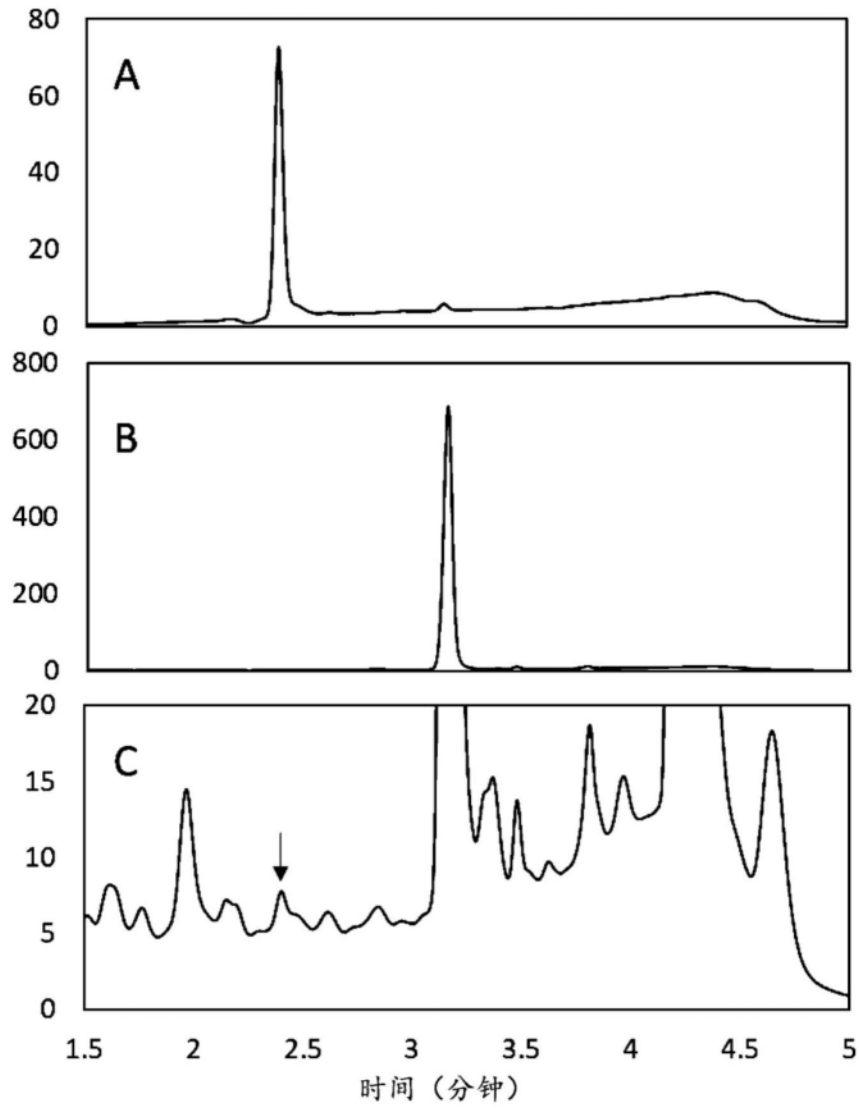


图6

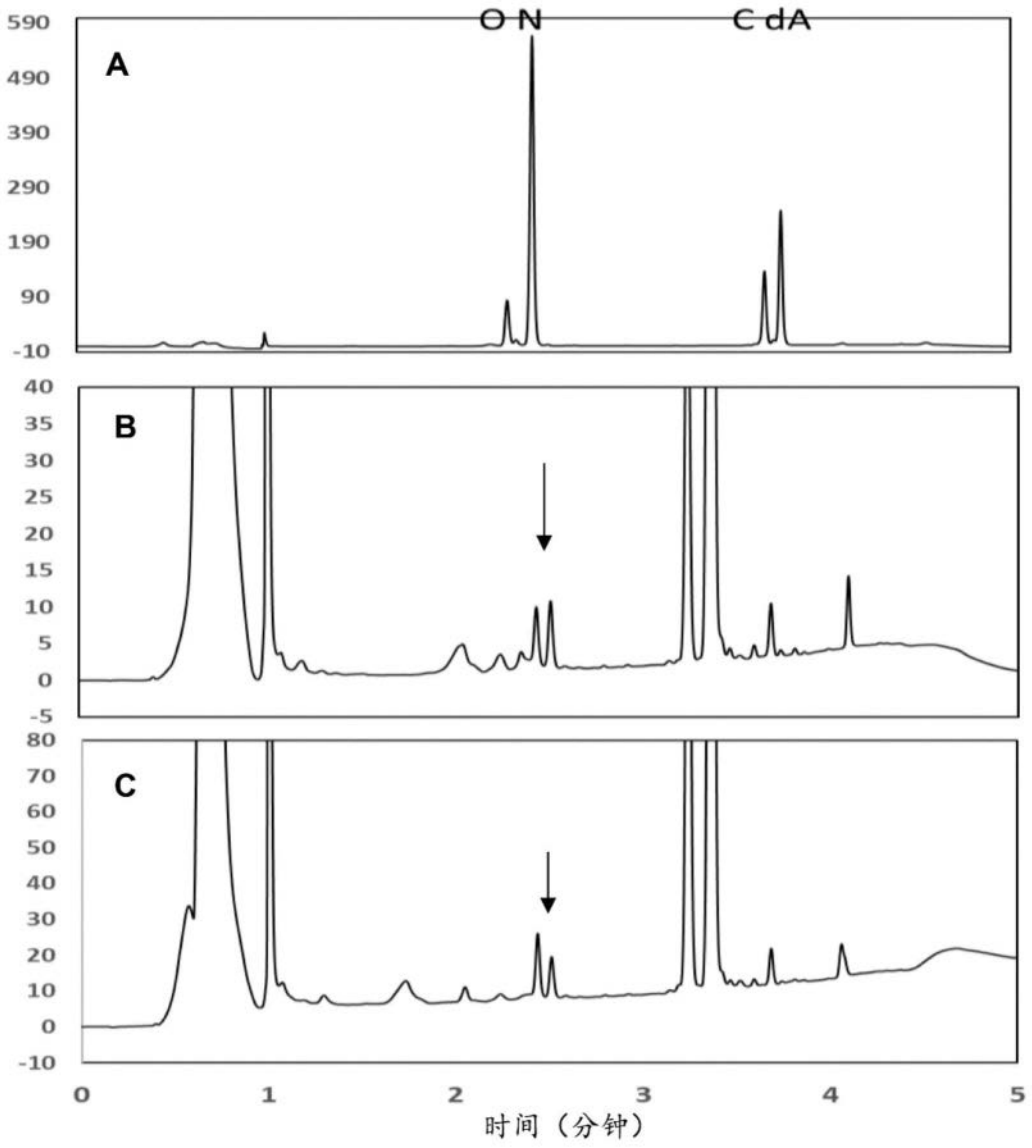


图7

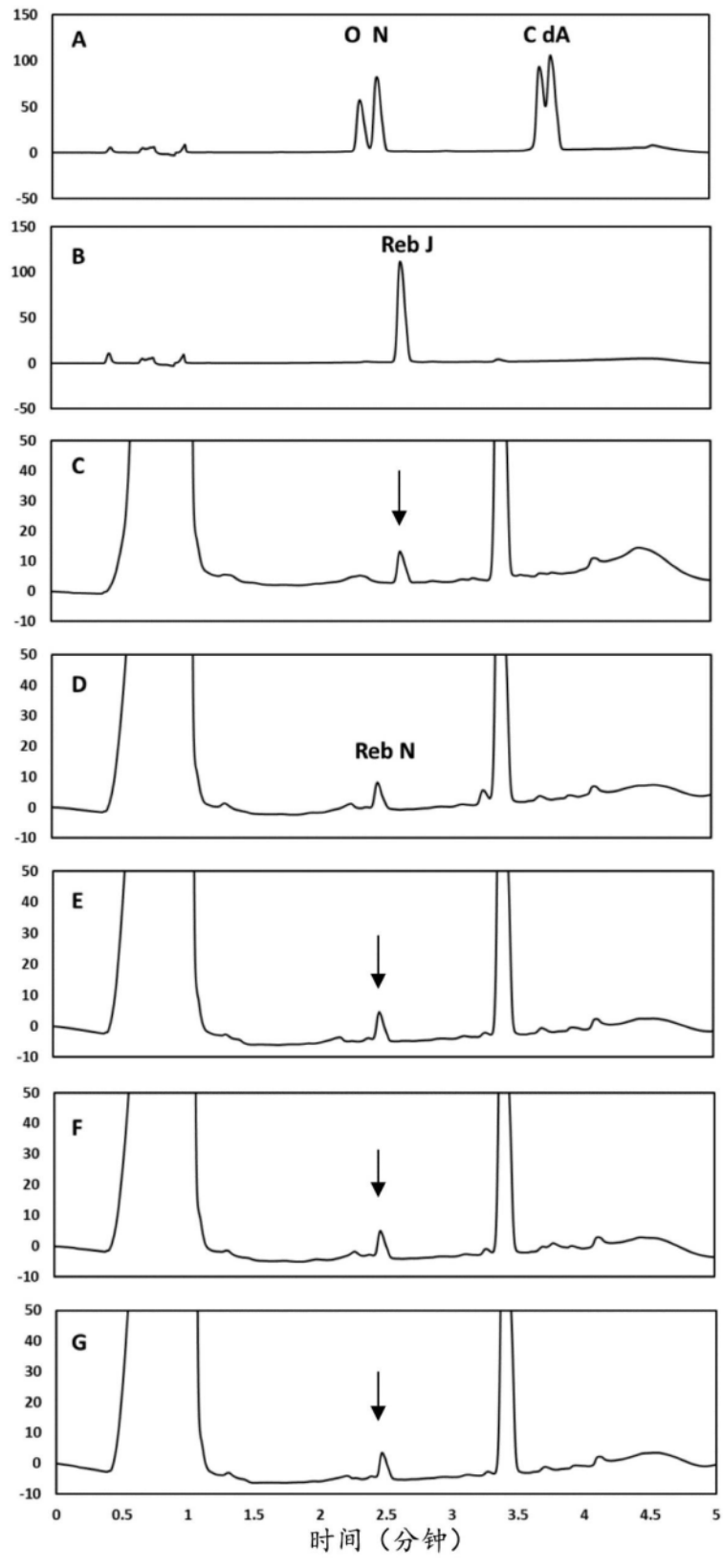


图8

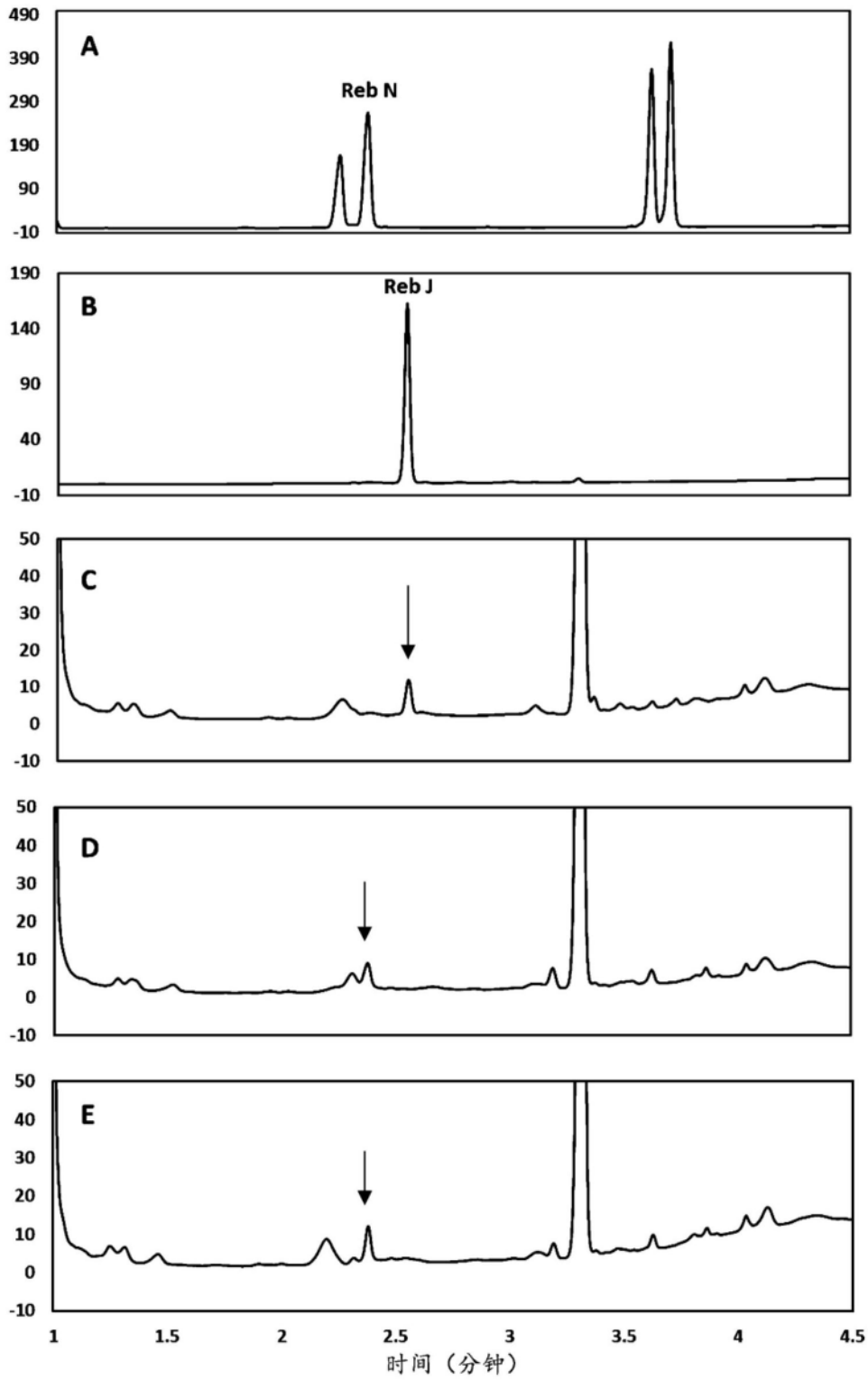


图9