(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109715793 A
(43)申请公布日 2019.05.03

(21)申请号 201780041363.1

(22)申请日 2017.05.05

(30)优先权数据
62/332,103 2016.05.05 US

(85)PCT国际申请进入国家阶段日
2018.12.28

(86)PCT国际申请的申请数据
PCT/US2017/031342 2017.05.05

(87)PCT国际申请的公布数据
WO2017/193022 EN 2017.11.09

(71)申请人 科德克希思公司
地址 美国加利福尼亚州

(72)发明人 约瓦娜·纳佐尔 韦丝娜·米切尔 大卫·艾尔加特 卡特里娜·W·莱克萨 尼基·德拉斯 罗伯特·凯文·奥尔 奥斯卡·阿尔维左 拉维·大卫·加西亚 朱迪·维多利亚·安东尼奥·维杜雅 考特尼·黛安娜·莫菲特

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
代理人 李敏春 郑霞

(51)Int.Cl.
C12N 9/84(2006.01)
C12P 21/00(2006.01)

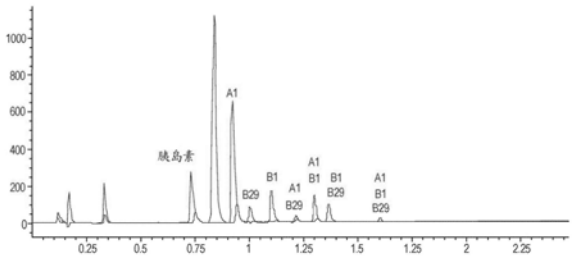
权利要求书7页 说明书83页
序列表（电子公布） 附图9页

(54)发明名称

青霉素-G 酰化酶

(57)摘要

本发明提供了具有改进的特性的工程化青霉素G酰化酶(PGA)、编码该酶的多核苷酸、包括该酶的组合物以及使用该酶的方法。



CN 109715793 A

1.一种工程化青霉素G酰化酶，所述工程化青霉素G酰化酶具有与SEQ ID NO:4、8、14、300、1036、1194、1262和/或1288至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同的多肽序列。

2.根据权利要求1所述的工程化青霉素G酰化酶，其中所述青霉素G酰化酶包括SEQ ID NO:4、8、14、300、1036、1194、1262和/或12884。

3.根据权利要求1所述的工程化青霉素G酰化酶，其中所述青霉素G酰化酶包含与表5.1、6.1、7.1、8.1、8.2、9.1、10.1、12.1、12.2、12.3、12.4、12.5、12.6和/或12.7中列出的至少一种序列至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同的多肽序列。

4.根据权利要求1所述的工程化青霉素G酰化酶，其中所述青霉素G酰化酶包含表5.1、6.1、7.1、8.1、8.2、9.1、10.1、12.1、12.2、12.3、12.4、12.5、12.6和/或12.7中列出的多肽序列。

5.根据权利要求1所述的工程化青霉素G酰化酶，其中所述青霉素G酰化酶包含与SEQ ID NO:8至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同并包含在选自27、27;28;71;74;547、27;71;74;484;547;584;697、71;74、129、253、254、256、348、352、372、373、374、380、380;457、386;390、386;390、386、387;390、451、457、467、470、474、616、623、704、706和708的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列，其中所述位置以SEQ ID NO:8为基准编号。

6.根据权利要求1所述的工程化青霉素G酰化酶，其中所述青霉素G酰化酶包含与SEQ ID NO:14至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列:9,9;103;119;131;233;312;324;432;444;494,9;103;119;131;324;432;494;646,9;103;119;131;233;269;304;444;494;646,9;103;119;131;304;324;432;444;494;646,9;103;119;233;646,9;103;119;494,9;103;233;312;646,9;103;233;432;646,9;103;233;494,9;103;269;304;324;494;646,9;103;304;432;444;646,9;119;131;233,9;119;131;233;304;444;646,9;119;131;233;494;646,9;119;131;233;494;661,9;119;131;312;444;646,9;119;131;432;444;646,9;119;233;269;273;304;312;432;444;646,9;119;233;494;646,9;119;304;444;494;646,9;131;233;269;273;312;432;444;646,9;233;273;304;494;646,9;233;304;494;646,9;233;312;432;646,9;233;494;646,9;312;444;646,9;432;444;494;646,9;494;646,28;374;380,103;119;131;233;273;304;324;432;444;494;646,103;119;131;233;304;312;432;494;646;661,103;119;131;269;312;494;646,103;119;233,103;119;233;273;432,103;119;233;304;646,103;119;233;312;646,103;119;494;646;661,103;119;646,103;131;233;304;324;444;646,103;131;269;273;444;646,103;233;273;312;324;432;444;646,103;233;273;312;432;444;646,103;269;273;444;646,103;273;304;324;444;494,103;312;444;646,103;444;494;646,119;131;444;646;661,119;131;494,119;131;646,119;131;304;432;444;646;661,119;131;444,119;233;304;312;324;432;444;646,119;233;304;646,119;233;312,119;233;646,119;269;273;312;324;494;646,119;269;273;312;432;444;646,119;273;324;444;494;646,119;312;444;646,119;432;444;646,129;254;348;457;704,129;348,129;348;467;470;704;

708,129;348;470;623;704;706,129;348;470;623;704;706;708,129;380;470,129;457;470;474,129;470,129;623,131;233;273;646,131;233;304;444,131;233;432;646,131;273;432;444;494;646,185,233;269;304;312;324;432;444;646,253,253;256,253;256;352;373;374;616,253;256;352;374;380;451,253;256;374;451,253;256;380;451;616,253;352;374;616,253;352;451;616,253;373;451,253;374;451,253;374;451;623,253;451,253;451;457,253;254;352;374;380,253;256,253;256;352;374;451,253;256;352;380;451;616,253;256;352;451,253;256;352;374;380;451,253;256;352;380,253;256;373,253;256;374;616,253;256;380,253;256;380;451;546;616,253;352;373;374;451,253;352;373;374;451;616,253;352;374;616,253;352;623,253;373;374;451;616,253;380,254;255;352,254;256;352;451,254;256;373;374;380;451,254;256;374,254;256;374;451,254;256;451,254;352;380,254;256;380;451;616,254;352;380;451,254;352;451,254;373;374;380;451,254;373;374;451,254;374,254;374;380,254;380,254;451,254;616,256;352;374,256;352;380;451;616,256;374;380;451;616,256;374;616,273;312;444;646,304;312;444;646,312;444;646,312;646,348,348;372;470;623;708,348;704;708,352,352;373;374;451,352;373;380;451,352;373;380;451;616,352;373;451,352;374,352;374;380,352;374;380;451;616,352;374;451,352;374;616,352;380,352;380;451,352;380;451;616,352;380;451;623,352;380;616,352;451,352;451;616,352;616,372;457;470;623,373;374,373;374;451,373;451,373;616,374,374;380,374;380;451,374;380;451;623,374;451,374;451;616,374;616,374;623,380,380;451,380;451;616,380;451;623,380;616,380;616;623,380;623,415,443,444,444;494,444;646,451,451;616,451;623,457,457;470,457;704;708,470;708,492;493,517,560,616,623,723,和748,其中所述位置以SEQ ID NO:14为基准编号。

7.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶包含与SEQ ID NO:300至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列:9;61;444,9;168;185;517;560;748,9;185;415,9;185;415;443;444;517;723;748,9;185;415;443;444;517;560,9;185;415;443;444;517;748,9;185;415;444;517;723;748,9;185;415;444;517;748,9;185;415;444;517;560,9;185;415;444;517;560;723;748,9;185;415;444,9;185;415;444;560,9;185;415;444;723;748,9;185;415;444;517;560;723;748,9;185;415;517;560,9;185;415;517;723,9;185;415;517;748,9;185;415;748,9;185;443;444;517;560,9;185;443;444;723,9;185;443;444;560;723,9;185;443;444;517,9;185;444,9;185;444;517;560,9;185;444;560;723,9;185;444;560;748,9;185;444;517;560,9;185;444;517;560;723,9;185;444;517;560;748,9;185;444;517;723;748,9;185;444;560,9;185;444;723,9;185;444;517;560;723,9;185;517,9;185;517;560;723;748,9;185;748,9;415;443;444;517;560,9;415;443;444;517;748,9;415;443;444;560;723;748,9;415;443;444;517;560,9;415;443;444;517;560;723,9;415;443;444;517;748,9;415;443;560,9;415;443;560;723;748,9;415;444,9;415;444;517;560;748,9;415;444;517,9;415;444;517;560;723;748,9;415;444;560;723;748,9;415;444;723;748,9;415;444;517,9;415;444;517;560;723;748,9;415;444;560;723,9;415;444;

517,9;415;444;517;560;748,9;415;444;517,9;415;444;560,9;415;444;560,9;415;444;517;560;723,9;415;444;560,9;415;444;560;665;723;748,9;415;444;723;748,9;415;517,9;415;517;560;723;748,9;415;517;560;723;748,9;415;517;560;748,9;415;560,9;415;748,9;443;444;517;560;723,9;443;444;560,9;443;444;560;723;748,9;443;444;517,9;443;444;560;723,9;443;444;517,9;443;444;517;560,9;443;444;517;748,9;443;517;748,9;443;723,9;444;517;560,9;444;560;723,9;444;748,9;444;517;,560;723;748,9;444;560,9;444;560;723,9;444;560;748,9;444;560;748,9;444;748,9;444;517;560;723;748,9;444;517,9;444;517;560;723,9;444;517;723,9;444;560;748,9;444;723,9;444,9;444;517;560,9;517,9;517;560;723,9;517;560;748,9;517;560;748,9;517;723,9;517;748,9;560;723;748,9;723,103,103;119,103;119;129;254;256;348;494;646,103;119;129;444;494,103;119;254;348,103;119;254;348;444,103;119;254;444,103;119;256;348;444;494;646,103;119;256;494,103;119;348,103;119;348;457,103;119;348;457;494,103;119;457,103;119;494,103;119;494;646,103;129,103;129;254;444;457;494,103;129;256;348,103;129;348;646,103;254,103;254;256;348;444;494,103;254;646,103;254;348,103;254;348;494,103;256,103;256;444,103;256;457,103;256;494,103;348,103;348;444,103;348;494,103;444,103;494,103;494;646,119;129;254;348;494,119;129;254;457;494,119;129;256;348;457,119;254;348;457,119;256;348 119;256;348;494,119;256;444,119;348,119;348;494,185;415;443;444;517;560,185;415;443;444;517,185;415;444;517;560;748,185;415;444;517;748,185;415;444;560,185;415;560,185;415;560;723,185;415;723;748,185;443;444;560,185;444;446;517;560,185;444;517;560,185;444;517;560;723;748,185;444;517;723,185;444;560,185;517;560;723,185;560,185;560;723,185;560;748,254;457,256;348,256;494,348,348;444,348;444;646,348;457,348;494,415,415;443;444;517;748,415;443;444;560,415;443;517;560;723,415;443;517;723,415;444;723,415;444;517,415;444;517;560;723;748,415;444;517;560;748,415;444;560;723,415;444;517;560;723,415;444;517;723;748 415;517;560;723;748,415;444;560,415;444;560;748,415;444,415;517,415;517;560,415;517;560;748,415;560;723,415;723;748,415;723;748,415;748,443;444;517;560;723;748,443;444;517;748,443;444,443;444:560,443;444;723,443;517,444,444;517;560,444;517;748,444;748,444;517;560;723;748,444;560;723,444;560,444;560;723,444;517,444;517;560,444;517;723;748,444;560;748,444;723,444;517;560,444;560;723,517;560,517;560;748,517;723,517;748,517;748,560,723,和723;748,其中所述位置以SEQ ID NO:300为基准编号。

8.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶包含与SEQ ID NO:1262至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自24,24;27;28;701;729,24;28;56;308;379;701,24;28;56;701,24;28;71;701,24;28;321;701,24;28;457;701,24;31;56;386;701,24;31,24;31;56;697,24;31;56;701,24;31;56;264;701;750,24;31;71;701,24;56;154;270;697,24;56;697;701,24;56;701,24;71;701,24;225;701,24;484;701,24;28,28;31,31;56;701,56;71;701,56;119;146;701,56;154;701,56;322;697;701,56;658;701,56;697;701,56;701,56;701;

711,697,697;701,71;74,71;701,129;511;701,154;754,177,410;697;701,423;701,431,697,和701的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列,其中所述位置以SEQ ID NO:1262为基准编号。

9.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶包含与SEQ ID NO:1288至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自22、31、31;56;264;308;379;484;547;711;750、32、50、57、69;74、71、71;74、71;74;129、71;74;145、71;74;248、71;74;470、71;149、75、141和394的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列,其中所述位置以SEQ ID NO:1288为基准编号。

10.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶包含与SEQ ID NO:1036至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自2、47、176、253、255、384、460、467、536和623的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列,其中所述位置以SEQ ID NO:1036为基准编号。

11.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶包含与SEQ ID NO:1194至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列12;103;119;131;233;384;444;494;S646,12;103;119;131;233;444;494;S646,12;103;119;233;384;444;494;S646,12;103;119;233;444;467;494;S646,12;103;119;233;444;494;536;S646,12;103;119;233;444;494:S646,12;103;131;233;444;467;494;S646,12;103;131;233;444;494;S646,12;103;233;444;494;S646,12;119;233;384;444;494;S646,28;264;384;467;484;536:547,103;119;131;233;384;444;494;S646,103;119;233;444;494;S646,103;233;444;494;S646,264;384;467;484;536;547,384;467;484;536;547,和668,其中所述位置以SEQ ID NO:1194为基准编号。

12.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶包含与SEQ ID NO:1288至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列:20;709,27,27;74;253;254,27;74;253;254;255;348;369;370;381,27;74;253;254;255;348;370;384,27;74;253;254;255;369;370,27;74;253;254;255;370,27;74;253;254;255;370;381;384,27;74;253;254;255;381,27;74;253;254;348,27;74;253;254;384,27;74;253;255,27;74;253;255;348;370;384,27;74;253;255;348;381,27;74;253;255;348;384,27;74;253;348;369;370,27;74;253;348;369;370;381,27;74;253;348;369;370;381;384,27;74;253;348;370,27;74;253;381;384,27;74;253;384,27;74;254;255;348,27;74;254;255;348;369;370;381,27;74;254;255;348;370,27;74;254;255;348;381,27;74;254;255;381,27;74;254;348;381;384,27;74;254;369;370,27;74;255;348,27;74;255;348;369;370,27;74;255;348;369;381;384,27;74;255;370,27;74;348,27;74;369;370,27;74;107;255;348;369;370,27;74;253,27;74;253;254;255,27;74;253;254;255;348;370,27;74;253;254;348;369;370;381,27;74;253;254;348;369;384,27;74;253;254;348;370,27;74;253;254;348;370;381,27;74;253;254;369,27;74;253;254;370;381,27;74;253;255;348;369;370,27;74;253;255;370,27;74;253;348,27;74;253;348;

370;381;384,27;74;253;369;381;384,27;74;254,27;74;254;255;348;369;381,27;74;254;255;348;370;381,27;74;254;255;369,27;74;254;348,27;74;254;348;369;381;384,27;74;254;348;370,27;74;254;348;370;381,27;74;254;384,27;74;255;348;370,27;74;348;384,27;74;253;254;255;348;369;370,27;74;253;254;348;369,27;74;253;254;348;369;370,27;74;253;254;348;370;381;384,27;74;253;254;348;381;384,27;74;253;370,27;74;254;255;348;369;370,27;74;254;255;348;381;384,27;74;370,27;74;253;254;255;348,27;74;253;255;348;370;381,27;74;253;255;384,27;74;253;348;369;370;384,27;74;253;348;381,27;74;254;255;348;369;384,27;74;254;255;348;370;381;384,27;74;254;255;370,27;74;254;348;381,27;74;254;369;384,27;74;255,27;74;255;348;369;381,27;74;348;370,27;74;369;370;381;384,27;253,27;253;254,27;253;254;255,27;253;254;255;260;348;381;384,27;253;254;255;348,27;253;254;255;348;369;370;381;384,27;253;254;255;348;369;384,27;253;254;255;348;370,27;253;254;255;348;370;384,27;253;254;255;348;381;384,27;253;254;348,27;253;254;348;370;381,27;253;254;348;370;384,27;253;254;348;381,27;253;254;348;381;384,27;253;254;348;384,27;253;254;381,27;253;254;381;384,27;253;254;384,27;253;255;348,27;253;255;348;369;370,27;253;255;348;381,27;253;255;348;384,27;253;255;370,27;253;255;370;381;384,27;253;348,27;253;348;370;381;384,27;253;348;370;384,27;253;348;381;384,27;253;369;370,27;253;381;384,27;254;255,27;254;255;348,27;254;255;348;369;370,27;254;255;348;370,27;254;255;348;370;381,27;254;255;348;370;381;384,27;254;255;348;370;384,27;254;255;369;370;381;384,27;254;255;370,27;254;255;381;384,27;254;255;384,27;254;348;369;370;381;384,27;254;348;370,27;254;348;381;384,27;254;348;384,27;254;369;381;384,27;254;449,27;254;470,27;255;348,27;255;348;370,27;255;348;370;381;384,27;255;348;381;384,27;255;370,27;348,27;348;369;370;381;384,27;348;381,27;348;384,69,74;253;254;369;370,74;254;255;348;384,84,128,131,132,133,134,253,253;348;370,254,255,255;348;370,256,317;380,348;467,370,373,377,381,381;672,383,384,388,453,457,467,472,615,616,618,619,620,623,627,701,705,,其中所述位置以SEQ ID NO:1288为基准编号。

13.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶包含与SEQ ID NO:1262至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自24;31;56;701、24;31;56;701、24;31;71;701、24;56;701、和71;74的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列,其中所述位置以SEQ ID NO:1262为基准编号。

14.根据权利要求1-14中任一项所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶能够产生乙酸苯酯单保护的或二保护的胰岛素。

15.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶包含与SEQ ID NO:4-1902的偶数编号的序列中提供的至少一种序列至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同的多肽序列。

16.根据权利要求1所述的工程化青霉素G酰化酶,其中所述青霉素G酰化酶由与SEQ ID

NO:3、7、13、299、1035、1193、1261和/或1287至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同的多核苷酸序列编码。

17.一种工程化多核苷酸序列,所述工程化多核苷酸序列编码根据权利要求1-16中任一项的工程化青霉素G酰化酶。

18.一种工程化多核苷酸序列,所述工程化多核苷酸序列编码一种工程化青霉素G酰化酶,其中所述工程化多核苷酸序列与SEQ ID NO:3-1901的奇数编号的序列中提供的至少一种序列至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同。

19.一种工程化多核苷酸序列,所述工程化多核苷酸序列编码包含与SEQ ID NO:3、7、13、299、1035、1193、1261和/或1287至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同的多肽序列的工程化青霉素G酰化酶。

20.一种载体,所述载体包含根据权利要求17-19中任一项所述的多核苷酸序列。

21.一种宿主细胞,所述宿主细胞包含根据权利要求11所述的载体。

22.根据权利要求12所述的宿主细胞,其中所述宿主细胞为原核细胞或真核细胞。

23.根据权利要求13所述的宿主细胞,其中所述原核宿主细胞为大肠杆菌(E.coli)。

24.一种组合物,所述组合物包含至少一种根据权利要求1-16中任一项中提供的工程化青霉素G酰化酶。

25.一种用于产生根据权利要求1-16中任一项所述的工程化青霉素G酰化酶的方法,所述方法包括在使得产生所述工程化青霉素G酰化酶的条件下,培养根据权利要求22和/或权利要求23所述的宿主细胞。

26.一种用于产生乙酸苯酯单保护的或二保护的胰岛素的方法,所述方法包括:i)提供根据权利要求1-16中任一项所述的工程化青霉素G酰化酶和/或根据权利要求24所述的组合物,和包含A1/B1/B29三-乙酸苯酯保护基团的胰岛素;和ii)在使得所述工程化青霉素G酰化酶从所述胰岛素去除A1、B1和/或B29三-乙酸苯酯保护基团的条件下,将所述工程化青霉素G酰化酶暴露于所述包含A1/B1/B29三-乙酸苯酯保护基团的胰岛素,从而产生乙酸苯酯单保护的或二保护的胰岛素。

27.根据权利要求26所述的方法,其中所述青霉素G酰化酶去除所述胰岛素的A1三-乙酸苯酯保护基团。

28.根据权利要求26所述的方法,其中所述青霉素G酰化酶去除所述胰岛素的B1三-乙酸苯酯保护基团。

29.根据权利要求26所述的方法,其中所述青霉素G酰化酶去除所述胰岛素的B29三-乙酸苯酯保护基团。

30.根据权利要求26所述的方法,其中所述青霉素G酰化酶去除所述胰岛素的A1、B1和B29三-乙酸苯酯保护基团。

31.根据权利要求26-29中任一项所述的方法,其中与由多肽SEQ ID NO:2产生乙酸苯酯单保护的或二保护的胰岛素相比,所述工程化青霉素G酰化酶产生多于90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多的乙酸苯酯单保护的或二保护的胰岛素。

32.一种组合物,所述组合物包含根据权利要求26-31中任一项所述的方法产生的乙酸苯酯单保护的或二保护的胰岛素。

33.一种用于产生乙酸苯酯单保护的或二保护的胰岛素的方法,所述方法包括:i)提供

权利要求1-16中任一项所述的工程化青霉素G酰化酶和/或根据权利要求24所述的组合物，和游离胰岛素；ii)在使得所述工程化青霉素G酰化酶酰化A1、B1和/或B29位置的条件下，将所述工程化青霉素G酰化酶暴露于所述胰岛素，从而产生单保护的或二保护的胰岛素。

34.根据权利要求33所述的方法，其中所述青霉素G酰化酶酰化所述胰岛素的A1位置。

35.根据权利要求33所述的方法，其中所述青霉素G酰化酶酰化所述胰岛素的B1位置。

36.根据权利要求33所述的方法，其中所述青霉素G酰化酶酰化所述胰岛素的B29位置。

37.根据权利要求33所述的方法，其中所述青霉素G酰化酶酰化所述胰岛素的A1、B1和B29位置。

38.根据权利要求32所述的方法，其中与由多肽SEQ ID NO:2产生乙酸苯酯单保护的或二保护的胰岛素相比，所述工程化青霉素G酰化酶产生多于90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多的乙酸苯酯单保护的或二保护的胰岛素。

39.一种组合物，所述组合物包含根据权利要求33-38中任一项所述的方法产生的乙酸苯酯单保护的或二保护的胰岛素。

40.一种用于产生乙酸苯酯单保护的或二保护的胰岛素的方法，所述方法包括：i)提供根据权利要求1-16中任一项所述的工程化青霉素G酰化酶和/或根据权利要求24所述的组合物，和胰岛素；ii)在使得所述工程化青霉素G酰化酶向所述胰岛素添加A1、B1和/或B29三-乙酸苯酯保护基团的条件下，将所述工程化青霉素G酰化酶暴露于所述胰岛素，从而产生乙酸苯酯单保护的或二保护的胰岛素。

41.根据权利要求40所述的方法，其中所述青霉素G酰化酶添加所述胰岛素的A1三-乙酸苯酯保护基团。

42.根据权利要求40所述的方法，其中所述青霉素G酰化酶添加所述胰岛素的B1三-乙酸苯酯保护基团。

43.根据权利要求40所述的方法，其中所述青霉素G酰化酶添加所述胰岛素的B29三-乙酸苯酯保护基团。

44.根据权利要求40所述的方法，其中所述青霉素G酰化酶添加所述胰岛素的A1、B1和B29三-乙酸苯酯保护基团。

45.根据权利要求40-44中任一项所述的方法，其中与由多肽SEQ ID NO:2产生乙酸苯酯单保护的或二保护的胰岛素相比，所述工程化青霉素G酰化酶产生多于90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多的乙酸苯酯单保护的或二保护的胰岛素。

46.一种组合物，所述组合物包含根据权利要求40-45中任一项所述的方法产生的乙酸苯酯单保护的或二保护的胰岛素。

47.根据权利要求26-31、33-38和/或40-45中任一项所述的方法，其中所述青霉素G酰化酶包括SEQ ID NO:837或1219。

48.根据权利要求26-31、33-38和/或40-45中任一项所述的方法，其中所述青霉素G酰化酶包括SEQ ID NO:897或1859。

49.一种组合物，所述组合物包含根据权利要求26-31、33-38和/或40-45中任一项所述的至少一种方法产生的乙酸苯酯单保护的或二保护的胰岛素。

青霉素-G 酰化酶

[0001] 本申请要求于2016年5月5日递交的美国临时专利申请系列号 62/332,103的优先权,出于所有目的将其通过引用以其整体并入本文。

发明领域

[0002] 本发明提供了工程化青霉素G酰化酶(PGA)、编码该酶的多核苷酸、包含该酶的组合物以及使用该工程化PGA酶的方法。

[0003] 对序列表、表格或计算机程序的引用

[0004] 序列表的正式副本作为ASCII格式化文本文件经由EFS-Web以文件名“CX2-154WO1_ST25.txt”与说明书同时提交,创建日期2017年5月5日,且大小为8,962千字节。经由EFS-Web提交的序列表为说明书的一部分并且通过引用以其整体并入本文。

[0005] 发明背景

[0006] 青霉素G酰化酶(PGA)(青霉素酰胺酶,EC 3.5.1.11)催化青霉素G(苄基青霉素)侧链的酰胺键的裂解。该酶在商业上被用于制造6-氨基-青霉烷酸(6-APA)和苯基乙酸(PAA)。6-APA是工业生产半合成的β-内酰胺抗生素诸如阿莫西林、氨苄青霉素和头孢氨苄的关键化合物。天然存在的PGA 酶在商业过程中显示出不稳定性,需要固定在固体基底上以用于商业应用。PGA已经被共价键接到多种支持物上,并且PGA固定系统已经被报道作为用于合成纯光学异构体的有用的工具。然而,附接到固体表面导致受损的酶特性,例如降低的活性和/或选择性和对溶质接触的限制。此外,尽管附接到固体基底允许捕获酶和在另外的加工循环中的再次使用,酶的稳定性使得此类应用可能受到限制。PGA酶促催化青霉素G形成6-APA是配向性(regiospecific)(它不裂解内酰胺酰胺键)和立体特异性的。生产6-APA 可能构成药品生产中酶促催化的最大利用。PGA与苯乙酰基部分相关的酶活性允许大量的伯胺以及醇的苯乙酰基衍生物的立体特异性水解。

发明内容

[0007] 本发明提供了工程化青霉素G酰化酶(PGA)、编码该酶的多核苷酸、包含该酶的组合物以及使用该工程化PGA酶的方法。

[0008] 本发明提供了能够通过向游离胰岛素的A1、B1或B29位置添加保护基团或者从A1/B1/B29三保护的胰岛素去除保护基团来产生乙酸苯酯单保护的或二保护的胰岛素的工程化青霉素G酰化酶。在一些实施方案中,青霉素G酰化酶与SEQ ID NO:4、8、14、300、1036、1194、1262和/或1288 至少约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%、约99%或更多地相同。在一些实施方案中,本发明提供了能够从胰岛素去除A1、B1 或B29三-乙酸苯酯保护基团以产生二-乙酸苯酯保护的胰岛素的工程化青霉素G酰化酶,其中青霉素G酰化酶与SEQ ID NO:4、8、14、300、1036、 1194、1262和/或1288至少85%、86%、87%、88%、89%、90%、91%、 92%、93%、94%、95%、96%、97%、98%、99%或更多地相同。在一些实施方案中,青霉素G酰化酶去除胰岛素的A1三-乙酸苯酯保护基团,而在一些另外的实施方案中,青霉素G酰化酶去除

胰岛素的B1三-乙酸苯酯保护基团，且在又另外的实施方案中，青霉素G酰化酶去除胰岛素的B29 三-乙酸苯酯保护基团。在一些实施方案中，青霉素G酰化酶去除胰岛素的A1、B1和B29三-乙酸苯酯保护基团。在一些另外的实施方案中，与由多肽SEQ ID NO:2产生乙酸苯酯单保护的或二保护的胰岛素相比，工程化青霉素G酰化酶产生多于90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或更多的乙酸苯酯单保护的或二保护的胰岛素。

[0009] 在一些另外的实施方案中，本发明提供了能够从胰岛素去除 A1/B1/B29三-乙酸苯酯保护基团以产生游离胰岛素的工程化青霉素G酰化酶，其中青霉素G酰化酶包括SEQ ID NO:4、8、14、300、1036、1194、 1262和/或1288。在一些另外的实施方案中，青霉素G酰化酶包含如表5.1、 6.1、7.1、8.1、8.2、9.1、10.1、12.1、12.2、12.3、12.4、12.5、12.6和/或 12.7中提供的青霉素G变体中的至少一个突变。在一些实施方案中，本发明提供了包含与表5.1、6.1、7.1、8.1、8.2、9.1、10.1、12.1、12.2、12.3、 12.4、12.5、12.6和/或12.7中列出的至少一种序列至少85％、90％、91％、 92％、93％、94％、95％、96％、97％、98％、99％或更多地相同的序列的工程化青霉素G酰化酶。在一些实施方案中，青霉素G酰化酶包含表5.1、 6.1、7.1、8.1、8.2、9.1、10.1、12.1、12.2、12.3、12.4、12.5、12.6和/或 12.7中提供的变体。

[0010] 本发明还提供了包含与SEQ ID NO:8至少85％、90％、91％、92％、 93％、94％、95％、96％、97％、98％、99％或更多地相同并包含在选自27、 27;28;71;74;547、27;71;74;484;547;584;697、71;74、129、253、254、256、 348、352、372、373、374、380、380;457、386;390、386;390、386、387;390、 451、457、467、470、474、616、623、704、706和708的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶，其中位置以SEQ ID NO:8为基准编号。

[0011] 本发明还提供了包含与SEQ ID NO:14至少85％、90％、91％、92％、 93％、94％、95％、96％、97％、98％、99％或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶:9,9;103;119;131;233;312;324;432;444;494,9;103;119;131;324;432;494;646,9;103;119;131;233;269;304;444;494;646， 9;103;119;131;304;324;432;444;494;646,9;103;119;233;646,9;103;119;494,9;103;233;312;646， 9;103;233;432;646,9;103;233;494,9;103;269;304;324;494;646,9;103;304;432;444;646， 9;119;131;233,9;119;131;233;304;444;646,9;119;131;233;494;646,9;119;131;233;494;661， 9;119;131;312;444;646,9;119;131;432;444;646,9;119;233;269;273;304;312;432;444;646， 9;119;233;494;646,9;119;304;444;494;646,9;131;233;269;273;312;432;444;646， 9;233;273;304;494;646,9;233;304;494;646,9;233;312;432;646,9;233;494;646,9;312;444;646， 9;432;444;494;646,9;494;646,28;374;380,103;119;131;233;273;304;324;432;444;494;646， 103;119;131;233;304;312;432;494;646;661,103;119;131;269;312;494;646,103;119;233， 103;119;233;273;432,103;119;233;304;646,103;119;233;312;646,103;119;494;646;661,103;119;646,103;131;233;304;324;444;646,103;131;269;273;444;646， 103;233;273;312;324;432;444;646,103;233;273;312;432;444;646,103;269;273;444;646， 103;273;304;324;444;494,103;312;444;646,103;444;494;646,119;131;444;646;661,119;131;494， 119;131;646,119;131;304;432;444;646;661,119;131;444,119;233;304;312;324;432;444;646， 119;233;304;646,119;233;312,119;233;646,119;269;273;312;324;494;

646，119;269;273;312;432;444;646,119;273;324;444;494;646,119;312;444;646,119;432;444;646，129;254;348;457;704,129;348,129;348;467;470;704;708,129;348;470;623;704;706，129;348;470;623;704;706;708,129;380;470,129;457;470;474,129;470,129;623,131;233;273;646，131;233;304;444,131;233;432;646,131;273;432;444;494;646,185，233;269;304;312;324;432;444;646,253,253;256,253;256;352;373;374;616,253;256;352;374;380;451,253;256;374;451,253;256;380;451;616,253;352;374;616,253;352;451;616,253;373;451,253;374;451,253;374;451;623,253;451,253;451;457,253;254;352;374;380,253;256,253;256;352;374;451,253;256;352;380;451;616,253;256;352;451，253;256;352;374;380;451,253;256;352;380,253;256;373,253;256;374;616,253;256;380，253;256;380;451;546;616,253;352;373;374;451,253;352;373;374;451;616,253;352;374;616，253;352;623,253;373;374;451;616,253;380,254;255;352,254;256;352;451，254;256;373;374;380;451,254;256;374,254;256;374;451,254;256;451,254;352;380，254;256;380;451;616,254;352;380;451,254;352;451,254;373;374;380;451,254;373;374;451，254;374,254;374;380,254;380,254;451,254;616,256;352;374,256;352;380;451;616，256;374;380;451;616,256;374;616,273;312;444;646,304;312;444;646,312;444;646,312;646,348，348;372;470;623;708,348;704;708,352,352;373;374;451,352;373;380;451,352;373;380;451;616，352;373;451,352;374,352;374;380,352;374;380;451;616,352;374;451,352;374;616,352;380，352;380;451,352;380;451;616,352;380;451;623,352;380;616,352;451,352;451;616,352;616，372;457;470;623,373;374,373;374;451,373;451,373;616,374,374;380,374;380;451，374;380;451;623,374;451,374;451;616,374;616,374;623,380,380;451,380;451;616,380;451;623,380;616,380;616;623,380;623,415,443,444,444;494,444;646,451,451;616,451;623,457，457;470,457;704;708,470;708,492;493,517,560,616,623,723,和748,其中位置以SEQ ID NO:14编号。

[0012] 本发明还提供了包含与SEQ ID NO:300至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶:9;61;444,9;168;185;517;560;748,9;185;415，9;185;415;443;444;517;723;748,9;185;415;443;444;517;560,9;185;415;443;444;517;748，9;185;415;444;517;723;748,9;185;415;444;517;748,9;185;415;444;517;560，9;185;415;444;517;560;723;748,9;185;415;444,9;185;415;444;560,9;185;415;444;723;748，9;185;415;444;517;560;723;748,9;185;415;517;560,9;185;415;517;723,9;185;415;517;748，9;185;415;748,9;185;443;444;517;560,9;185;443;444;723,9;185;443;444;560;723，9;185;443;444;517,9;185;444,9;185;444;517;560,9;185;444;560;723,9;185;444;560;748，9;185;444;517;560,9;185;444;517;560;723,9;185;444;517;560;748,9;185;444;517;723;748，9;185;444;560,9;185;444;723,9;185;444;517;560;723,9;185;517,9;185;517;560;723;748，9;185;748,9;415;443;444;517;560,9;415;443;444;517;748,9;415;443;444;560;723;748，9;415;443;444;517;560,9;415;443;444;517;560;723,9;415;443;444;517;748,9;415;443;560，9;415;443;560;723;748,9;415;444,9;415;444;517;560;748,9;415;444;517，9;

415;444;517;560;723;748,9;415;444;560;723;748,9;415;444;723;748,9;415;444;517, 9;415;444;517;560;723;748,9;415;444;560;723,9;415;444;517,9;415;444;517;560;748, 9;415;444;517,9;415;444;560,9;415;444;560,9;415;444;517;560;723,9;415;444;560, 9;415;444;560;665;723;748,9;415;444;723;748,9;415;517,9;415;517;560;723;748, 9;415;517;560;723;748,9;415;517;560;748,9;415;560,9;415;748,9;443;444;517;560;723, 9;443;444;560,9;443;444;560;723;748,9;443;444;517,9;443;444;560;723,9;443;444;517, 9;443;444;517;560,9;443;444;517;748,9;443;517;748,9;443;723,9;444;517;560,9;444;560;723, 9;444;748,9;444;517;560;723;748,9;444;560,9;444;560;723,9;444;560;748,9;444;560;748, 9;444;748,9;444;517;560;723;748,9;444;517,9;444;517;560;723,9;444;517;723,9;444;560;748, 9;444;723,9;444,9;444;517;560,9;517,9;517;560;723,9;517;560;748,9;517;560;748,9;517;723, 9;517;748,9;560;723;748,9;723,103,103;119,103;119;129;254;256;348;494;646, 103;119;129;444;494,103;119;254;348,103;119;254;348;444,103;119;254;444, 103;119;256;348;444;494;646,103;119;256;494,103;119;348,103;119;348;457, 103;119;348;457;494,103;119;457,103;119;494,103;119;494;646,103;129, 103;129;254;444;457;494,103;129;256;348,103;129;348;646,103;254,103;254;256;348;444;494, 103;254;646,103;254;348,103;254;348;494,103;256,103;256;444,103;256;457,103;256;494,103;348,103;348;444,103;348;494,103;444,103;494,103;494;646,119;129;254;348;494,119;129;254;457;494,119;129;256;348;457,119;254;348;457,119;256;348 119;256;348;494, 119;256;444,119;348,119;348;494,185;415;443;444;517;560,185;415;443;444;517, 185;415;444;517;560;748,185;415;444;517;748,185;415;444;560,185;415;560,185;415;560;723, 185;415;723;748,185;443;444;560,185;444;446;517;560,185;444;517;560, 185;444;517;560;723;748,185;444;517;723,185;444;560,185;517;560;723,185;560,185;560;723, 185;560;748,254;457,256;348,256;494,348,348;444,348;444;646,348;457,348;494,415, 415;443;444;517;748,415;443;444;560,415;443;517;560;723,415;443;517;723,415;444;723, 415;444;517,415;444;517;560;723;748,415;444;517;560;748,415;444;560;723, 415;444;517;560;723,415;444;517;723;748 415;517;560;723;748,415;444;560,415;444;560;748, 415;444,415;517,415;517;560,415;517;560;748,415;560,415;560;723,415;723;748,415;723;748,415;748, 443;444;517;560;723;748,443;444;517;748,443;444,443;444;560,443;444;723,443;517,444,444;517;560,444;517;748,444;748,444;517;560;723;748,444;560;723,444;560,444;560;723,444;517,444;517;560,444;517;723;748,444;560;748,444;723,444;517;560,444;560;723,517;560, 517;560;748,517;723,517;748,517;748,560,723,和723;748,其中位置以SEQ ID NO:300为基准编号。

[0013] 本发明还提供了包含与SEQ ID NO:1262至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶:24,24;27;28;701;729,24;28;56;308;379;701, 24;28;56;701,24;28;71;701,24;28;321;701,24;28;457;701,24;31;56;386;701,24;31,24;31;56;697, 24;31;56;701,24;31;56;264;701;750,24;31;71;

701,24;56;154;270,697,24;56;697;701,24;56;701， 24;71;701,24;225;701,24;484;701,24;28,28;31,31;56;701,56;71;701,56;119;146;701,56;154;701， 56;322;697;701,56;658;701,56;697;701,56;701,56;701;711,697,697;701,71;74,71;701， 129;511;701,154;754,177,410;697;701,423;701,431,697,和701,其中位置以SEQ ID NO:1262为基准编号。

[0014] 本发明还提供了包含与SEQ ID NO:1288至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自22、31、31;56;264;308;379;484;547;711;750、32、50、57、69;74、71、71;74、71;74;129、71;74;145、71;74;248、71;74;470、71;149、75、141和394的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶,其中位置以SEQ ID NO:1288为基准编号。

[0015] 本发明还提供了包含与SEQ ID NO:1036至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自2、47、176、253、255、384、460、467、536和623的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶,其中位置以SEQ ID NO:1036为基准编号。

[0016] 本发明还提供了包含与SEQ ID NO:1194至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶:12;103;119;131;233;384;444;494;S646， 12;103;119;131;233;444;494;S646,12;103;119;233;384;444;494;S646， 12;103;119;233;444;467;494;S646,12;103;119;233;444;494;536;S646， 12;103;119;233;444;494;S646,12;103;131;233;444;467;494;S646,12;103;131;233;444;494;S646， 12;103;233;444;494;S646,12;119;233;384;444;494;S646,28;264;384;467;484;536;547， 103;119;131;233;384;444;494;S646,103;119;233;444;494;S646,103;233;444;494;S646， 264;384;467;484;536;547,384;467;484;536;547,和668,其中位置以SEQ ID NO:1194为基准编号。

[0017] 本发明还提供了包含与SEQ ID NO:1288至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多地相同并包含在选自以下的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶:20;709,27,27;74;253;254,27;74;253;254;255;348;369;370;381,27;74;253;254;255;348;370;384,27;74;253;254;255;369;370， 27;74;253;254;255;370,27;74;253;254;255;370;381;384,27;74;253;254;255;381， 27;74;253;254;348,27;74;253;254;384,27;74;253;255,27;74;253;255;348;370;384， 27;74;253;255;348;381,27;74;253;255;348;384,27;74;253;348;369;370， 27;74;253;348;369;370;381,27;74;253;348;369;370;381;384,27;74;253;348;370， 27;74;253;381;384,27;74;253;384,27;74;254;255;348,27;74;254;255;348;369;370;381， 27;74;254;255;348;370,27;74;254;255;348;381,27;74;254;255;381,27;74;254;348;381;384， 27;74;254;369;370,27;74;255;348,27;74;255;348;369;370,27;74;255;348;369;381;384， 27;74;255;370,27;74;348,27;74;369;370,27;74;107;255;348;369;370,27;74;253， 27;74;253;254;255,27;74;253;254;255;348;370,27;74;253;254;348;369;370;381， 27;74;253;254;348;369;384,27;74;253;254;348;370,27;74;253;254;348;370;381， 27;74;253;254;369,27;74;253;254;370;381,27;74;

253;255;348;369;370,27;74;253;255;370，27;74;253;348,27;74;253;348;370;381;384,27;74;253;369;381;384，27;74;254,27;74;254;255;348;369;381,27;74;254;255;348;370;381,27;74;254;255;369，27;74;254;348,27;74;254;348;369;381;384,27;74;254;348;370,27;74;254;348;370;381，27;74;254;384,27;74;255;348;370,27;74;348;384,27;74;253;254;255;348;369;370，27;74;253;254;348;369,27;74;253;254;348;369;370,27;74;253;254;348;370;381;384，27;74;253;254;348;381;384,27;74;253;370,27;74;254;255;348;369;370，27;74;254;255;348;381;384,27;74;370,27;74;253;254;255;348,27;74;253;255;348;370;381，27;74;253;255;384,27;74;253;348;369;370;384,27;74;253;348;381,27;74;254;255;348;369;384，27;74;254;255;348;370;381;384,27;74;254;255;370,27;74;254;348;381,27;74;254;369;384，27;74;255,27;74;255;348;369;381,27;74;348;370,27;74;369;370;381;384,27;253,27;253;254,27;253;254;255,27;253;254;255;260;348;381;384,27;253;254;255;348，27;253;254;255;348;369;370;381;384,27;253;254;255;348;369;384,27;253;254;255;348;370,27;253;254;255;348;370;384,27;253;254;255;348;381;384,27;253;254;348，27;253;254;348;370;381,27;253;254;348;370;384,27;253;254;348;381,27;253;254;348;381;384，27;253;254;348;384,27;253;254;381,27;253;254;381;384,27;253;254;384,27;253;255;348，27;253;255;348;369;370,27;253;255;348;381,27;253;255;348;384,27;253;255;370，27;253;255;370;381;384,27;253;348,27;253;348;370;381;384,27;253;348;370;384，27;253;348;381;384,27;253;369;370,27;253;381;384,27;254;255,27;254;255;348，27;254;255;348;369;370,27;254;255;348;370,27;254;255;348;370;381，27;254;255;348;370;381;384,27;254;255;348;370;384,27;254;255;369;370;381;384，27;254;255;370,27;254;255;381;384,27;254;255;384,27;254;348;369;370;381;384,27;254;348;370，27;254;348;381;384,27;254;348;384,27;254;369;381;384,27;254;449,27;254;470,27;255;348，27;255;348;370,27;255;348;370;381;384,27;255;348;381;384,27;255;370,27;348，27;348;369;370;381;384,27;348;381,27;348;384,69,74;253;254;369;370,74;254;255;348;384,84，128,131,132,133,134,253,253;348;370,254,255,255;348;370,256,317;380,348;467,370,373，377,381,381;672,383,384,388,453,457,467,472,615,616,618,619,620,623,627,701,705，706,708,和709,其中位置以SEQ ID NO:1288为基准编号。

[0018] 本发明还提供了包含与SEQ ID NO:1262至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同并包含在选自 24;31;56;701、24;31;56;701、24;31;71;701、24;56;701和71;74的一个或更多个氨基酸位置的至少一个取代或取代集的多肽序列的工程化青霉素G酰化酶,其中位置以SEQ ID NO:1262为基准编号。

[0019] 本发明提供了包含与SEQ ID NO:3-1901的奇数编号的序列中提供的至少一种序列至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同的序列的工程化青霉素G酰化酶。

[0020] 本发明还提供了包含至少一种本文提供的青霉素G酰化酶的组合物。在一些实施方案中,组合物包含至少一种被固定的青霉素G酰化酶。

[0021] 本发明还提供了由与选自SEQ ID NO:3、7、13、299、1035、1193、1261和/或1287的

序列具有至少约85％、约86％、约87％、约88％、约 89％、约90％、约91％、约92％、约93％、约94％、约95％、约96％、约 97％、约98％、约99％或更多序列同一性的多核苷酸序列编码的青霉素G 酰化酶。在一些实施方案中，编码青霉素G酰化酶的多核苷酸序列与选自SEQID NO:3、7、13、299、1035、1193、1261和/或1287的序列具有至少85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、 96％、97％、98％、99％或更多序列同一性。在一些实施方案中，多核苷酸序列选自SEQ ID NO:3、7、13、299、1035、1193、1261和/或1287。在一些实施方案中，工程化多核苷酸序列包含与SEQ ID NO:3-1901的奇数编号的序列中提供的至少一种序列至少85％、90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或更多地相同的序列。

[0022] 本发明还提供了包含本文提供的多核苷酸序列(例如，SEQ ID NO:3、 7、13、299、1035、1193、1261和/或1287)的载体。本发明还提供了包含本文提供的载体(例如，包含SEQ ID NO:3、7、13、299、1035、1193、 1261和/或1287的多核苷酸序列的载体)的宿主细胞。在一些实施方案中，载体包含至少一种包含与SEQ ID NO:3-1901的奇数编号的序列中提供的至少一种序列至少85％、90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或更多地相同的序列的工程化多核苷酸序列。

[0023] 本发明还提供了包含本文提供的载体的宿主细胞。在一些实施方案中，宿主细胞为原核细胞或真核细胞。在一些另外的实施方案中，宿主细胞为细菌细胞。在一些另外的实施方案中，宿主细胞为大肠杆菌(E.coli)。

[0024] 本发明还提供了包含至少一种本文提供的工程化青霉素G酰化酶的组合物。另外，本发明提供了用于产生本文提供的工程化青霉素G酰化酶的方法，所述方法包括在使得产生工程化青霉素G酰化酶的条件下培养本文提供的宿主细胞。在一些实施方案中，所述方法还包括回收产生的工程化青霉素G酰化酶。

[0025] 本发明还提供了用于产生乙酸苯酯保护的胰岛素的方法，所述方法包括：i)提供至少一种本文提供的工程化青霉素G酰化酶和/或包含至少一种本文提供的工程化青霉素G酰化酶的组合物，和包含A1/B1/B29三-乙酸苯酯保护基团的胰岛素或游离胰岛素；和ii)在使得工程化青霉素G酰化酶去除A1/B1/B29三-乙酸苯酯保护基团并产生游离胰岛素、二保护的胰岛素或单保护的胰岛素的条件下，将工程化青霉素G酰化酶暴露于包含 A1/B1/B29三-乙酸苯酯保护基团的胰岛素；iii)在使得工程化青霉素G酰化酶添加乙酸苯酯保护基团并产生三保护的胰岛素、二保护的胰岛素或单保护的胰岛素的条件下，将工程化青霉素G酰化酶暴露于胰岛素。在该方法的一些实施方案中，青霉素G酰化酶与SEQ ID NO:4、8、14、300、1036、 1194、1262和/或1288至少约85％、约86％、约87％、约88％、约89％、约90％、约91％、约92％、约93％、约94％、约95％、约96％、约97％、约98％、约99％或更多地相同。在该方法的一些实施方案中，青霉素G酰化酶与SEQ ID NO:4、8、14、300、1036、1194、1262和/或1288至少85％、 86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同。在该方法的一些另外的实施方案中，青霉素G 酰化酶包括SEQ ID NO:4、8、14、300、1036、1194、1262和/或1288。在一些实施方案中，青霉素G酰化酶与SEQ ID NO:4-1902中提供的至少一种偶数编号的序列至少85％、86％、87％、88％、89％、90％、91％、92％、 93％、94％、95％、96％、97％、98％、99％或更多地相同。在一些实施方案中，工程化青霉素G酰化酶比野生型青霉素G酰化酶产生多于90％、 91％、92％、93％、94％、95％、

96％、97％、98％、99％或更多的游离胰岛素。本发明还提供了包含使用本发明的方法产生的游离胰岛素的组合物。在一些实施方案中，本发明提供了包含根据本文提供的任何方法产生的乙酸苯酯单保护的或二保护的胰岛素的组合物。

[0026] 本发明还提供了用于产生乙酸苯酯单保护的或二保护的胰岛素的方法，所述方法包括：i)提供至少一种工程化青霉素G酰化酶和/或包含至少一种本文提供的工程化青霉素G酰化酶的组合物，和游离胰岛素；和ii) 在使得工程化青霉素G酰化酶酰化A1、B1和/或B29位置的条件下，将工程化青霉素G酰化酶暴露于胰岛素，从而产生单保护的或二保护的胰岛素。在一些实施方案中，青霉素G酰化酶酰化胰岛素的A1位置，而在一些其他实施方案中，青霉素G酰化酶酰化胰岛素的B1位置，且在又另外的实施方案中，青霉素G酰化酶酰化胰岛素的B29位置。在一些实施方案中，青霉素G酰化酶酰化胰岛素的A1、B1和B29位置。在一些另外的实施方案中，与由多肽SEQ ID NO:2产生乙酸苯酯单保护的或二保护的胰岛素相比，工程化青霉素G酰化酶产生多于90％、91％、92％、93％、94％、 95％、96％、97％、98％、99％或更多的乙酸苯酯单保护的或二保护的胰岛素。在一些另外的实施方案中，用于酰化A1、B1和/或B29的青霉素G 酰化酶包括与SEQ ID NO:4、8、14、300、1036、1194、1262和/或1288 具有至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、 99％或更多地同一性的序列。在一些实施方案中，青霉素G酰化酶与SEQ ID NO:4-1902中提供的至少一种偶数编号的序列至少85％、86％、87％、 88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多地相同。

[0027] 在该方法的一些另外的实施方案中，青霉素G酰化酶包含与表5.1、 6.1、7.1、8.1、8.2、9.1、10.1、12.1、12.2、12.3、12.4、12.5、12.6和/或 12.7中的至少一种序列具有至少85％、90％、91％、92％、93％、94％、95％、 96％、97％、98％、99％或更多同一性的序列。在该方法的又另外的实施方案中，青霉素G酰化酶包含表5.1、6.1、7.1、8.1、8.2、9.1、10.1、12.1、 12.2、12.3、12.4、12.5、12.6和/或12.7的任一个中列出的序列。

[0028] 本发明还提供了用于产生乙酸苯酯单保护的或二保护的胰岛素的方法，所述方法包括：i)提供权利要求1的工程化青霉素G酰化酶，和胰岛素；ii)在使得工程化青霉素G酰化酶向胰岛素添加A1、B1和/或B29三- 乙酸苯酯保护基团的条件下，将工程化青霉素G酰化酶暴露于胰岛素，从而产生乙酸苯酯单保护的或二保护的胰岛素。在一些实施方案中，青霉素G酰化酶添加胰岛素的A1三-乙酸苯酯保护基团，而在一些另外的实施方案中，青霉素G酰化酶添加胰岛素的B1三-乙酸苯酯保护基团，且在又另外的实施方案中，青霉素G酰化酶添加胰岛素的B29三-乙酸苯酯保护基团。在一些另外的实施方案中，青霉素G酰化酶添加胰岛素的A1、B1和 B29三-乙酸苯酯保护基团。在一些实施方案中，与由多肽SEQ ID NO:2 产生乙酸苯酯单保护的或二保护的胰岛素相比，工程化青霉素G酰化酶产生多于90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或更多的乙酸苯酯单保护的或二保护的胰岛素。在一些另外的实施方案中，青霉素G酰化酶包括与SEQ ID NO:4、8、14、300、1036、1194、1262和/或1288具有至少85％、90％、91％、92％、93％、94％、95％、96％、 97％、98％、99％或更多同一性的序列。在一些实施方案中，青霉素G酰化酶与SEQ ID NO:4-1902中提供的至少一种偶数编号的序列至少85％、 86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、 98％、99％或更多地相同。

[0029] 本发明还提供了包含根据本文提供的任何方法产生的乙酸苯酯单保护的或二保

护的胰岛素的组合物。

[0030] 本发明还提供了其中青霉素G酰化酶包括SEQ ID NO:837、897、1219 或1859的方法。

[0031] 附图简述

[0032] 图1提供了用于定量胰岛素的分析方法的层析谱和酰化产物的洗脱顺序。

[0033] 图2提供了实施例11中描述的实验的结果。

[0034] 发明描述

[0035] 本发明提供了能够将青霉素裂解为苯乙酸和6-氨基青霉烷酸(6-APA) 的工程化青霉素G酰化酶(PGA),6-氨基青霉烷酸是合成各种各样的β-内酰胺抗生素的关键中间体。特别地,本发明提供了能够通过向游离胰岛素的A1、B1或B29位置添加保护基团或从A1/B1/B29三保护的胰岛素去除保护基团或去除A1/B1/B29三-乙酸苯酯保护基团来释放游离胰岛素来产生乙酸苯酯单保护的或二保护的胰岛素的工程化PGA。

[0036] 通常,天然存在的PGA是包括α亚基和β亚基的异二聚体酶。野生型PGA天然地被合成为PGA原前体(pre-pro-PGA)多肽,包含介导易位至周质的N-末端信号肽和连接α亚基的C-末端至β亚基的N-末端的接头区域。蛋白水解加工产生成熟的异二聚体酶。分子间接头区域还可以起促进酶的正确折叠的功能。本文提供的PGA是基于来自Kluyvera citrophila的PGA,其中引入多种修饰以产生如下文详细描述的改进的酶学特性。

[0037] 对于本文提供的描述,单数的使用包括复数(且反之亦然),除非另有明确说明。例如单数形式“一(a)”、“一(an)”和“该(the)”包括复数指代物,除非上下文另有清楚指明。类似地,“包含(comprise、comprises、comprising)”、“包括(include、includes和including)”是可互换的,而不意图限制性的。还应当理解,在多种实施方案的描述中使用术语“包含(comprising)”的情况下,本领域技术人员将理解,在一些特定情况下,可以使用“基本上由...组成”或“由...组成”的语言可选择地描述实施方案。

[0038] 包括附图的上文的一般描述和下文的详细描述两者仅是示例性的和说明性的,而不是限制本公开内容。此外,本文使用的章节标题仅用于组织目的,并且不被解释为限制所描述的主题。

[0039] 定义

[0040] 如本文使用的,以下术语意图具有以下含义。

[0041] 参考本公开内容,本文说明书中使用的技术和科学术语将具有本领域普通技术人员通常理解的含义,除非另有具体定义。因此,以下术语意图具有以下含义。本文提及的全部专利和出版物,包括此类专利和出版物中公开的全部序列被明确地通过引用并入。除非另有指明,本发明的实践包括分子生物学、发酵、微生物学和相关领域中常用的常规技术,这些技术是本领域技术人员已知的。除非本文另有定义,本文使用的所有技术和科学术语具有与本发明所属领域内的普通技术人员通常理解的的含义相同的含义。尽管与本文描述的那些相似或等效的任何方法和材料都可用于实践或测试本发明,但描述了优选的方法和材料。事实上,不意图本发明受限于本文描述的特定方法学、方案和试剂,因为这些可以根据使用它们的环境变化。本文提供的标题不是对本发明的多个方面或实施方案的限制。

[0042] 尽管如此,为了便于理解本发明,许多术语定义如下。数值范围包括定义该范围的数字。因此,本文公开的每个数值范围意图包括落入此类较宽数值范围内的每个较窄的数

值范围，如同此类较窄数值范围在本文被全部清楚地写出。还意图本文公开的每个最大的(或最小的)数值限值包含每个较低(或较高)的数值限值，如同此类较低(或较高)的数值限值在本文中被清楚地写出。

[0043] 如本文使用的，术语“包含(comprising)”及其同源词以其包容性含义被使用(即，等同于术语“包括(including)”及其对应的同源词)。

[0044] 如在本文和在所附权利要求中所使用的，单数“一(a)”、“一(an)”及“该 (the)”包括复数指示物，除非上下文另有清楚地指明。因此，例如提及一种“宿主细胞”包括多于一种(a plurality of)此类宿主细胞。

[0045] 除非另有指明，分别地，核酸以5′至3′方向从左到右书写且氨基酸序列以氨基至羧基方向从左至右书写。

[0046] 本文提供的标题不是对可以通过整体参考本说明书而被具有的本发明的各个方面或实施方案的限制。因此，下文定义的术语通过整体参考本说明书作而被更充分地定义。

[0047] 如本文使用的，术语“蛋白”、“多肽”和“肽”在本文中可互换使用，以表示通过酰胺键共价连接的至少两个氨基酸的聚合物，而不论长度或翻译后修饰(例如糖基化、磷酸化、脂化、肉豆蔻酸化(myristilation)、泛素化等)。此定义中包括D-氨基酸和L-氨基酸以及D-氨基酸和L-氨基酸的混合物。

[0048] 如本文使用的，“多核苷酸”和“核酸”指共价连接在一起的两个或更多个核苷。多核苷酸可以完全包含核糖核苷(即RNA)、完全由2′脱氧核糖核苷酸(即DNA)构成或完全由核糖核苷和2′脱氧核糖核苷的混合物构成。虽然核苷将通常经由标准磷酸二酯键连接在一起，但多核苷酸可以包括一种或更多种非标准连接。多核苷酸可以是单链的或双链的，或者可以包括单链区和双链区两者。此外，尽管多核苷酸将通常包含天然存在的编码核碱基(即腺嘌呤、鸟嘌呤、尿嘧啶、胸腺嘧啶和胞嘧啶)，它可以包含一种或更多种修饰的和/或合成的核碱基(例如肌苷、黄嘌呤、次黄嘌呤等)。优选地，此类修饰的或合成的核碱基将是编码核碱基。

[0049] 如本文使用的，“杂交严格度”指在核酸杂交中的杂交条件，诸如洗涤条件。通常，杂交反应在较低严格度的条件下进行，随后是不同的但较高严格度的洗涤。术语“中度严格杂交”指允许靶DNA结合以下互补的核酸的条件，所述互补的核酸与靶DNA具有约60％同一性，优选地约75％同一性，约85％同一性;与靶多核苷酸具有大于约90％同一性。示例性中度严格条件为等同于在42℃在50％甲酰胺、5×Denhart溶液、5×SSPE、0.2％ SDS中杂交，随后是在42℃在0.2×SSPE、0.2％SDS中洗涤的条件。“高严格度杂交”通常指距如对指定的多核苷酸序列在溶液条件下确定的热解链温度T_m约10℃或更少的条件。在一些实施方案中，高严格度条件指以下的条件:其仅允许在65℃在0.018M NaCl中形成稳定的杂交体的那些核酸序列的杂交(即，如果杂交体在65℃在0.018M NaCl中为不稳定的，它在如本文考虑的高严格度条件下将为不稳定的)。可以例如通过在等同于在42℃在50％甲酰胺、5×Denhart′s溶液、5×SSPE、0.2％SDS的条件中杂交，随后在65℃在0.1×SSPE和0.1％ SDS中洗涤来提供高严格度条件。另一个高严格度条件是在等同于在65℃在包含0.1％(w:v)SDS的5X SSC 中杂交，并在65℃在包含0.1％SDS的0.1×SSC中洗涤。其他高严格度杂交条件以及中度严格度条件是本领域技术人员已知的。

[0050] 如本文使用的，“编码序列”指编码蛋白的氨基酸序列的核酸部分(例如基因)。

[0051] 如本文使用的，“密码子优化”指编码蛋白的多核苷酸的密码子改变为在特定生物体中优先使用的那些密码子，使得编码的蛋白在感兴趣的生物体中有效地表达。在一些实施方案中，可以对编码PGA酶的多核苷酸进行密码子优化，用于从选择用于表达的宿主生物体的优化产生。尽管遗传密码为简并的，因为大多数氨基酸由被称为“同义(synonyms)”或“同义突变 (synonymous)”密码子的几个密码子表示，但熟知的是，特定生物体的密码子使用为非随机的并对于特定的密码子三联体有偏好。就给定的基因、具有共同功能或祖先来源的基因、高表达的蛋白相对于低拷贝数蛋白和生物体基因组的聚集蛋白编码区而言，这种密码子使用偏向可能更高。在一些实施方案中，可以对编码PGA酶的多核苷酸进行密码子优化，用于从选择用于表达的宿主生物体的优化产生。

[0052] 如本文使用的，“优选的、最佳的、高密码子使用偏好密码子”可互换地指在蛋白编码区中的使用频率高于编码相同氨基酸的其他密码子的密码子。优选的密码子可以根据单个基因、共同功能或来源的一组基因、高表达基因中的密码子使用、整个生物体的聚集蛋白编码区中的密码子频率、相关生物体的聚集蛋白编码区中的密码子频率或其组合来确定。其频率随着基因表达的水平而增加的密码子通常是用于表达的最佳密码子。用来确定特定生物体中密码子的频率(例如密码子使用、相对同义突变密码子使用) 和密码子偏好的多种方法是已知的，包括多变量分析，例如使用聚类分析或相关性分析和基因中使用的密码子的有效数目(参见例如GCG CodonPreference,Genetics Computer Group Wisconsin Package;CodonW, John Peden,University of Nottingham;McInerney,Bioinform.,14:372-73 [1998];Stenico等人,Nucl.Acids Res.,222:437-46[1994];和Wright,Gene 87:23-29[1990])。密码子使用表可用于越来越多的生物体(参见例如,Wada 等人,Nucleic Acids Res.,20:2111-2118[1992];Nakamura等人,Nucl.Acids Res.,28:292[2000];Duret,等人,同上;Henaut和Danchin,“Escherichia coli and Salmonella,”Neidhardt,等人.(编著),ASM Press,Washington D.C.[1996],第2047-2066页。用于获得密码子使用的数据源可以依赖于能够编码蛋白的任何可获得的核苷酸序列。这些数据集包括实际已知编码被表达的蛋白的核酸序列(例如完整的蛋白编码序列-CDS)、被表达的序列标签(ESTS)或基因组序列的预测编码区)(参见例如Uberbacher,Meth.Enzymol., 266:259-281[1996];Tiwari等人,Comput.Appl.Biosci.,13:263-270[1997])。

[0053] 如本文使用的，“控制序列”在本文中被定义为包括对本发明的多核苷酸和/或多肽的表达是必需或有利的所有组分。每一个控制序列对于感兴趣的多核苷酸可以是天然的或外源的。此类控制序列包括但不限于前导序列、多聚腺苷酸化序列、前肽序列、启动子、信号肽序列和转录终止子。

[0054] 如本文使用的，“可操作地连接”在本文被定义为如下配置：在所述配置中控制序列被适当放置(即，以功能关系)在相对于感兴趣的多核苷酸的位置处，使得控制序列指导或调控感兴趣的多核苷酸和/或多肽的表达。

[0055] 如本文使用的，“启动子序列”指被宿主细胞识别用于感兴趣的多核苷酸诸如编码序列的表达的核酸序列。控制序列可以包括合适的启动子序列。启动子序列包含介导感兴趣的多核苷酸的表达的转录控制序列。启动子可以是在选择的宿主细胞中显示转录活性的任何核酸序列，包括突变、截短的和杂合启动子，并且可以从编码与宿主细胞同源或异源的细胞外或细胞内多肽的基因来获得。

[0056] 如本文使用的，“天然存在的”或“野生型”指自然界中发现的形式。例如，天然存在或野生型多肽或多核苷酸序列是存在于可以从自然界的来源分离的生物体中并且没有被人工操作有意修饰的序列。

[0057] 如本文使用的，当在本公开内容中提及(例如细胞、核酸或多肽)使用时，“非天然存在的”、“工程化的”和“重组的”指已经以自然界中原本不存在的方式被修饰的材料，或对应于材料的自然或天然形式的材料。在一些实施方案中，材料虽与天然存在的材料相同，但产生自或源自合成的材料和/或通过使用重组技术操作产生。非限制性实例包括，除其他以外，表达在天然(非重组)形式的细胞中未发现的基因或表达原本以不同水平表达的天然基因的重组细胞。

[0058] 如本文使用的，“序列同一性百分比”、“同一性百分比”和“相同百分比”指多核苷酸序列或多肽序列之间的比较，并且通过在比较窗上比较两个最佳比对(aligned)序列来确定，其中与参考序列相比，比较窗中的多核苷酸或多肽序列的部分可以包括添加或缺失(即空位)，用于这两个序列的最佳比对序列。百分比如下计算：通过确定两个序列中出现相同的核酸碱基或氨基酸残基、或者核酸碱基或氨基酸残基与空位对齐的位置的数目，以产生匹配位置的数目，将匹配位置的数目除以比较窗中位置的总数目，并将结果乘以100以得到序列同一性的百分比。确定最佳比对和序列同一性百分比使用BLAST和BLAST 2.0算法进行(参见例如，Altschul等人，J. Mol.Biol.215:403-410[1990]；和Altschul等人，Nucl.Acids Res.3389-3402 [1977])。用于进行BLAST分析的软件为通过美国国家生物技术信息中心 (the+National Center for Biotechnology Information,NCBI)网站公共可得的。

[0059] 简言之，BLAST分析包括首先通过鉴定查询序列中长度W的短字来鉴定高评分序列对(HSP)，当所述短字与数据库序列中相同长度的字比对时匹配或满足一些正值的阀值评分T。T被称为相邻字评分阈值(Altschul 等人，同上)。这些最初的邻近字击中(word hit)用作启动检索的种子以找到更长的包含它们的HSP。然后字击中沿着每个序列在两个方向上延伸直到累积比对评分不能增加的程度。对于核苷酸序列，累积评分使用参数M(对于匹配残基对的奖励评分；永远＞0)和N(对于错配残基的惩罚评分；永远＜0)计算。对于氨基酸序列，评分矩阵用于计算累积评分。当以下情况时，每一个方向上的字击中的延伸被终止：累积比对评分从其最大达到值下降了量X；由于累积一个或更多个负评分残基比对，累积得分达到0或以下；或到达任一序列末端。BLAST算法参数W、T和X确定比对的灵敏度和速度。BLASTN程序(对于核苷酸序列)使用以下作为缺省值：字长(W)为 11、期望值(E)为10、M＝5、N＝-4、以及两个链的比较。对于氨基酸序列， BLASTP程序使用以下作为缺省值：字长(W)为3，期望值(E)为10和 BLOSUM62评分矩阵(参见例如，Henikoff和Henikoff，Proc.Natl.Acad.Sci.USA 89:10915[1989])。

[0060] 许多其他算法是可获得的和本领域已知的，这些算法在提供两个序列的同一性百分比方面与BLAST相似地运作。待比较的序列的最佳比对可以使用本领域已知的任何合适的方法进行(例如通过Smith和Waterman， Adv.Appl.Math.2:482[1981]的局部同源性算法；通过Needleman和Wunsch， J.Mol.Biol.48:443[1970]的同源性比对算法；通过Pearson和Lipman， Proc.Natl.Acad.Sci.USA 85:2444[1988]的搜索相似性的方法；和/或通过这些算法的计算机实现([GCG Wisconsin软件包中的GAP、BESTFIT、FASTA 和TFASTA)，或通过

使用本领域通常已知的方法进行目视检查。另外，序列比对与序列同一性百分比的确定可以使用GCG Wisconsin软件包 (Accelrys、Madison WI)中的BESTFIT或GAP程序，使用提供的缺省参数。

[0061] 如本文使用的，“大体同一性”指在至少20个残基位置的比较窗中、通常在至少30个-50个残基的窗中，与参考序列相比，具有至少80％序列同一性、至少85％同一性和89％至95％序列同一性，更通常至少99％序列同一性的多核苷酸或多肽序列，其中序列同一性的百分比通过在比较窗上比较参考序列和包含总计为参考序列的20％或更少的缺失或添加的序列来计算。在应用于多肽的具体实施方案中，术语“大体同一性”意指当诸如通过程序GAP或BESTFIT使用缺省空位权重进行最佳比对时，两个多肽序列共有至少80％的序列同一性，优选地至少89％的序列同一性、至少95％的序列同一性或更多(例如99％的序列同一性)。在一些优选的实施方案中，不相同的残基位置通过保守氨基酸取代而不同。

[0062] 如本文中使用的，“参考序列”指另一序列被与之比较的特定序列。参考序列可以是更大序列的子集，例如，全长基因或多肽序列的区段。通常，参考序列为至少20个核苷酸或氨基酸残基的长度、至少25个残基的长度、至少50个残基的长度，或核酸或多肽的全长。因为两个多核苷酸或多肽可以各自(1)包括在两个序列之间相似的序列(即，完整序列的一部分)，和 (2)可以还包括在两个序列之间趋异的(divergent)序列，所以两个(或更多个)多核苷酸或多肽之间的序列比较通常通过在比较窗上比较两个多核苷酸的序列以鉴定和比较序列局部区域的相似性来进行。术语“参考序列”不意图限于野生型序列，并且可以包括工程化序列或改变的序列。例如，在一些实施方案中，“参考序列”可以是先前工程化或改变的氨基酸序列。

[0063] 如本文使用的，“比较窗”指至少约20个连续核苷酸位置或氨基酸残基的概念性区段，其中序列可以与至少20个连续核苷酸或氨基酸的参考序列进行比较，并且其中序列在比较窗中的部分与参考序列(其不包含添加或缺失)相比，可以包括20％或更少的添加或缺失(即，空位)以用于两个序列的最佳比对。比较窗可以比20个连续残基更长，并任选地包括30、40、 50、100或更长的窗。

[0064] 如本文使用的，当在给定的氨基酸或多核苷酸序列的编号的情况下使用时，“对应于”、“以.....为基准”和“相对于”指当给定的氨基酸或多核苷酸序列与参考序列相比较时指定的参考序列残基的编号。换言之，给定的聚合物的残基数目或残基位置以参考序列为基准被指定，而不是通过给定的氨基酸或多核苷酸序列内残基的实际数字位置被指定。例如，给定的氨基酸序列，诸如工程化PGA的氨基酸序列可以通过引入空位以与参考序列比对而优化两条序列之间的残基匹配。在这些情况中，尽管存在空位，在给定的氨基酸或多核苷酸序列中的残基的编号以与其比对的参考序列作出为基准。如本文使用的，如下文进一步描述的对残基位置的指代，诸如“Xn”，应被理解为指“对应于.....的残基”，除非另有明确说明。因此，例如，“X94”指多肽序列中位置94处的任何氨基酸。

[0065] 如本文使用的，“改进的酶特性”指与参考PGA相比表现出任何酶特性的改进的PGA。对于本文描述的工程化PGA多肽，比较一般针对野生型 PGA酶进行，尽管在一些实施方案中，参考PGA可以是另一种改进的工程化PGA。需要改进的酶特性包括但不限于酶促活性(其可以用使用指定量的PGA，在指定反应时间，底物转化的百分比表示)、化学选择性、热稳定性、溶剂稳定性、pH活性谱、辅因子需求、对抑制剂的耐性 (refractoriness)(例如产物

抑制)、立体专一性和立体选择性(包括对映异构体选择性)。

[0066] 如本文使用的,“增加的酶促活性”指工程化PGA多肽的改进特性,其可以通过与参考PGA酶相比,比活性(例如,产生的产物/时间/重量蛋白) 的增加或底物向产物的转化百分比(例如使用指定量的PGA,在指定时间段内起始量的底物向产物的转化转化百分比)的增加来表示。确定酶活性的示例性方法被提供于实施例中。可以影响与酶活性相关的任何特性,包括经典酶特性K_m、V_{max}或k_{cat},它们的改变可以导致增加的酶促活性。酶活性的改进可以从对应的野生型PGA酶的酶促活性的约1.5倍至比天然存在的PGA或所述PGA多肽源自其的另一种工程化PGA的酶促活性高多达2 倍、5倍、10倍、20倍、25倍、50倍、75倍、100倍或更多。在具体实施方案中,工程化PGA酶表现出比亲本PGA酶高1.5倍至50倍、1.5倍至100倍的范围内的改进的酶促活性。技术人员应理解,任何酶的活性都是受扩散限制的,使得催化转化速率不能超过底物包括任何需要的辅因子的扩散速率。扩散限值,或k_{cat}/K_m的理论最大值通常为约10^8至10^9($M^{-1}s^{-1}$)。因此,PGA的酶活性的任何改进将具有与被PGA酶作用的底物的扩散速率相关的上限。PGA活性可以通过用于测量青霉素G裂解时乙酸苯酯的释放的任何一种标准测定测量,诸如滴定法(参见例如,Simons和Gibson, Biotechnol.Tech.,13:365-367[1999])。在一些实施方案中,PGA活性可以通过使用6-硝基苯基乙酰氨基苯甲酸(NIPAB)测量,其裂解产物5-氨基-2-硝基-苯甲酸是分光光度法可测量的(λmax=405nm)。酶活性的比较使用定义的酶制品、在设定条件下用定义的测定和一种或更多种定义的底物进行,如本文进一步详细描述的。通常,当比较裂解物时,确定细胞的数目和测定的蛋白的量,并且使用相同的表达系统和相同的宿主细胞来使由宿主细胞产生并存在于裂解物中的酶的量的变化最小化。

[0067] 如本文使用的,“增加的酶促活性”和“增加的活性”指工程化酶的改进特性,其可以通过与本文描述的参考酶相比,比活性(例如,产生的产物/ 时间/重量蛋白)的增加或将底物向产物的转化百分比(例如使用指定量的 PGA,在指定时间段内起始量的底物向产物的转化转化百分比)的增加来表示。可以影响与酶活性相关的任何特性,包括经典酶特性K_m、V_{max}或k_{cat},它们的改变可以导致增加的酶促活性。在一些实施方案中,本文提供的 PGA酶通过从胰岛素的特定残基去除三-乙酸苯酯保护基团来释放胰岛素。酶活性的比较使用定义的酶制品、在设定条件下用定义的测定和一种或更多种定义的底物进行,如本文进一步详细描述的。通常,当比较细胞裂解物中的酶时,确定细胞的数目和测定的蛋白的量,并且使用相同的表达系统和相同的宿主细胞来使由宿主细胞产生并存在于裂解物中的酶的量的变化最小化。

[0068] 如本文使用的,“转化”指将底物酶促转化为对应的产物。

[0069] 如本文使用的,“转化百分比”指在指定条件下在一定时间段内被转化为产物的底物的百分比。因此,例如,PGA多肽的“酶促活性”或“活性”可以表示为底物向产物的“转化百分比”。

[0070] 如本文中使用的,“化学选择性”指相对于一种产物相对于另一种产物在化学或酶促反应中的优先形成。

[0071] 如本文使用的,“热稳定性的(thermostable)”和“热稳定的(thermal stable)”可互换使用以指这样的多肽,其与未处理的酶相比,当暴露于一组温度条件(例如40-80℃)持续一段时间(例如0.5-24小时)时耐受失活,因此在暴露于升高的温度后,保持一定水平的

残余活性(例如多于60％至80％)。

[0072] 如本文使用的,“溶剂稳定的”指与未处理的酶相比,多肽在暴露于不同浓度(例如5％-99％)的溶剂(例如异丙醇、四氢呋喃、2-甲基四氢呋喃、丙酮、甲苯、乙酸丁酯、甲基叔丁基醚等)一段时间(例如0.5-24小时)后,维持相似活性(例如大于例如60％至80％)的能力。

[0073] 如本文使用的,“pH稳定的”指与未处理的酶相比,暴露于高pH或低 pH(例如4.5-6或8至12)一段时间(例如0.5-24小时)后维持相似活性(例如多于60％至80％)的PGA多肽。

[0074] 如本文使用的,“热稳定且溶剂稳定的”指既热稳定也溶剂稳定的PGA 多肽。

[0075] 如本文使用的,“亲水性氨基酸或残基”指具有根据Eisenberg等人的标准化共有疏水性量表(normalized consensus hydrophobicity scale)表现出小于0的疏水性的侧链的氨基酸或残基(Eisenberg等人,J.Mol.Biol., 179:125-142[1984])。遗传编码的亲水性氨基酸包括L-Thr(T)、L-Ser(S)、 L-His(H)、L-Glu(E)、L-Asn(N)、L-GIn(Q)、L-Asp(D)、L-Lys(K)和L-Arg (R)。

[0076] 如本文使用的,“酸性氨基酸或残基”指当氨基酸包含于肽或多肽内时,具有表现出小于约6的pK值的侧链的亲水性氨基酸或残基。由于失去氢离子,酸性氨基酸在生理pH通常具有带负电荷的侧链。遗传编码的酸性氨基酸包括L-Glu(E)和L-Asp(D)。

[0077] 如本文使用的,“碱性氨基酸或残基”指当氨基酸包含于肽或多肽内时,具有表现出大于约6的pK值的侧链的亲水性氨基酸或残基。由于与水合离子的缔合,碱性氨基酸通常在生理pH具有带正电荷的侧链。遗传编码的碱性氨基酸包括L-Arg(R)和L-Lys(K)。

[0078] 如本文使用的,“极性氨基酸或残基”指具有这样的侧链的亲水性氨基酸或残基:该侧链在生理pH不带电荷的侧链,具有至少一个其中两个原子共用的电子对被一个原子更紧密地持有的键。遗传编码的极性氨基酸包括L-Asn(N)、L-Gln(Q)、L-Ser(S)和L-Thr(T)。

[0079] 如本文使用的,“疏水性氨基酸或残基”指具有根据Eisenberg等人的标准化共有疏水性量表表现出大于0的疏水性的侧链的氨基酸或残基 (Eisenberg等人,J.Mol.Biol., 179:125-142[1984])。遗传编码的疏水性氨基酸包括L-Pro(P)、L-Ile(I)、L-Phe(F)、L-Val(V)、L-Leu(L)、L-Trp(W)、 L-Met(M)、L-Ala(A)和L-Tyr(Y)。

[0080] 如本文使用的,“芳族氨基酸或残基”指具有包括至少一个芳环或杂芳环的侧链的亲水性或疏水性氨基酸或残基。遗传编码的芳族氨基酸包括L-Phe(F)、L-Tyr(Y)和L-Trp(W)。尽管由于L-His(H)的杂芳基氮原子的 pKa,它有时被归类为碱性残基或因为它的侧链包括杂芳环而被归类为芳族残基,在本文中组氨酸被归类为亲水性残基或“受限残基(constrained residue)”(参见下文)。

[0081] 如本文使用的,“受限氨基酸或残基”指具有受限几何学的氨基酸或残基。本文中,受限残基包括L-Pro(P)和L-His(H)。组氨酸具有受限几何学,因为它具有相对小的咪唑环。脯氨酸具有受限几何学,因为它也具有五元环。

[0082] 如本文使用的,“非极性氨基酸或残基”指具有在生理pH不带电荷的侧链、并且所述侧链具有其中两个原子共有的电子对通常被两个原子的每一个等同地持有的键(即,侧链不是极性的)的疏水性氨基酸或残基。遗传编码的非极性氨基酸包括L-Gly(G)、L-Leu(L)、L-Val(V)、L-Ile(I)、L-Met (M)和L-Ala(A)。

[0083] 如本文使用的,“脂族氨基酸或残基”指具有脂族烃侧链的疏水性氨基酸或残基。

遗传编码的脂族氨基酸包括L-Ala(A)、L-Val(V)、L-Leu(L) 和L-Ile(I)。

[0084] 应注意,半胱氨酸(或“L-Cys”或“[C]”)之所以与众不同,是因为它可以与其他L-Cys(C)氨基酸或其他含巯基(sulfanyl)或巯基(sulfhydryl)氨基酸形成二硫桥。“半胱氨酸样残基”包括半胱氨酸和包含可用于形成二硫桥的巯基部分的其他氨基酸。L-Cys(C)(和具有含-SH侧链的其他氨基酸)以还原的游离-SH或氧化的二硫桥接形式存在于肽中的能力影响L-Cys(C)是否向肽贡献净的疏水性质或亲水性质。虽然根据Eisenberg的标准化共有量表 (Eisenberg等人,1984,同上),L-Cys(C)表现出0.29的疏水性,但是应理解,为了本公开内容的目的,L-Cys(C)被归入它自己的独特的组。

[0085] 如本文使用的,“小氨基酸或残基“指具有包括总计三个或更少个碳和/ 或杂原子(不包括α-碳和氢)的侧链的氨基酸或残基。根据以上定义,小氨基酸或残基可以进一步分类为脂族、非极性、极性或酸性小氨基酸或残基。遗传编码的小氨基酸包括L-Ala(A)、L-Val(V)、L-Cys(C)、L-Asn(N)、 L-Ser(S)、L-Thr(T)和L-Asp(D)。

[0086] 如本文使用的,“含羟基的氨基酸或残基”指包含羟基(-OH)部分的氨基酸。遗传编码的含羟基的氨基酸包括L-Ser(S)、L-Thr(T)和L-Tyr(Y)。

[0087] 如本文使用的,“氨基酸差异”和“残基差异”指在多肽序列的一个位置处氨基酸残基相对于参考序列中对应位置处的氨基酸残基的差异。本文中氨基酸差异的位置通常被称为“Xn”,其中n指残基差异所基于的参考序列中的对应位置。例如,“与SEQ ID NO:2相比在位置X40处的残基差异”指在对应于SEQ ID NO:2的位置40的多肽位置处的氨基酸残基的差异。因此,如果SEQ ID NO:2的参考多肽在位置40处具有组氨酸,则“与SEQ ID NO:2相比在位置X40处的残基差异”指在对应于SEQ ID NO:2的位置 40的多肽位置处除了组氨酸以外的任何残基的氨基酸取代。在本文的大多数实例中,在一个位置处的特定氨基酸残基差异指示为“XnY”,其中“Xn”指定如以上描述的对应位置,并且“Y”为在工程化多肽中发现的氨基酸的单字母标识符(即,与参考多肽中的不同的残基)。在一些实例中,本公开内容还提供由常规符号“AnB”表示的特定氨基酸差异,其中A为参考序列中的残基的单字母标识符,“n”为在参考序列中的残基位置的数字,并且B 为工程化多肽的序列中残基取代的单字母标识符。在一些情况下,本公开内容的多肽相对于参考序列可以包含一个或更多个氨基酸残基差异,其由相对于参考序列存在残基差异的一列指定位置指示。在其中多于一个氨基酸可以被用于多肽的特定残基位置的一些实施方案中,可以使用的多种氨基酸残基由“/”隔开(例如,X192A/G)。在一些实施方案中,取代集中的取代通过分号(“;”)或斜杠(“/”)隔开(例如,对于相对于SEQ ID NO:8具有以下取代集的变体PGA--Y27T;G71H;D74G;D484N;Q547K;Y584F;M697L)。本公开内容包括包含包括保守和非保守的氨基酸取代的任一个/或两者的一个或更多个氨基酸差异的工程化多肽序列。包括于本公开内容的序列表中的特定重组碳酸酐酶多肽的氨基酸序列包括起始甲硫氨酸(M)残基(即,M 代表残基位置1)。然而,技术人员理解,该起始甲硫氨酸残基可以通过例如宿主细胞中或体外翻译系统中的生物加工机制去除,以产生缺乏起始甲硫氨酸残基但另外保留酶的特性的成熟蛋白。因此,如本文使用的术语“相对于SEQ ID NO:2在位置Xn的氨基酸残基差异”可以指位置“Xn”或已经被加工以缺少起始甲硫氨酸的参考序列的对应位置(例如,位置(X-1)n)。

[0088] 如本文使用的,短语“保守氨基酸取代”指具有相似侧链的残基的可互换性,并且因此通常包括用相同或相似的氨基酸定义类别中的氨基酸取代多肽中的氨基酸。例如但不

限于,在一些实施方案中,具有脂族侧链的氨基酸被另一种脂族氨基酸(例如,丙氨酸、缬氨酸、亮氨酸和异亮氨酸)取代;具有羟基侧链的氨基酸被另一种具有羟基侧链的氨基酸(例如,丝氨酸和苏氨酸)取代;具有芳族侧链的氨基酸被另一种具有芳香族侧链的氨基酸 (例如,苯丙氨酸、酪氨酸、色氨酸和组氨酸)取代;具有碱性侧链的氨基酸被另一种具有碱性侧链的氨基酸(例如,赖氨酸和精氨酸)取代;具有酸性侧链的氨基酸被另一种具有酸性侧链的氨基酸(例如,天冬氨酸或谷氨酸) 取代;和/或疏水性氨基酸或亲水性氨基酸分别被另一种疏水性氨基酸或亲水性氨基酸取代。示例性保守性取代在表1中提供。

[0089]

表 1.示例性保守氨基酸取代

残基	可能的保守取代
A、、L、V、I	其他脂族的(A、L、V、I) 其他非极性的(A、L、V、I、G、M)
G、M	其他非极性的(A、L、V、I、G、M)
D、E	其他酸性的(D、E)
K、R	其他碱性的(K、R)
N、Q、S、T	其他极性的
H、Y、W、F	其他芳族的(H、Y、W、F)
C、P	非极性的

[0090] 如本文使用的,短语“非保守取代”指用具有显著不同的侧链特性的氨基酸取代多肽中的氨基酸。非保守取代可以使用定义的组之间而不是定义的组之内的氨基酸,并且影响:(a)取代区域中的肽骨架的结构(例如,脯氨酸取代甘氨酸),(b)电荷或疏水性,或(c)侧链堆积。例如但不限于,示例性非保守取代可以是用碱性或脂族氨基酸取代酸性氨基酸;用小氨基酸取代芳族氨基酸;和用疏水性氨基酸取代亲水性氨基酸。

[0091] 如本文使用的,“缺失”指通过从参考多肽去除一个或更多个氨基酸的多肽修饰。缺失可以包括去除1个或更多个氨基酸、2个或更多个氨基酸、5个或更多个氨基酸、10个或更多个氨基酸、15个或更多个氨基酸或者20个或更多个氨基酸、多达组成多肽的氨基酸总数的10%、或多组成多肽的达氨基酸总数的20%,同时保留酶促活性和/或保留工程化酶的改进的特性。缺失可以涉及多肽的内部部分和/或末端部分。在多个实施方案中,缺失可以构成连续的区段或可以是不连续的。

[0092] 如本文使用的,“插入”指通过向参考多肽添加一个或更多个氨基酸的多肽修饰。在一些实施方案中,改进的工程化PGA酶包括向天然存在的 PGA多肽中插入一个或更多个氨基酸,以及向工程化PGA多肽中插入一个或更多个氨基酸。插入可以在多肽的内部部分或至羧基或氨基末端。如本文使用的插入包括如本领域已知的融合蛋白。插入可以是氨基酸的连续区段或由天然存在的多肽中的一个或更多个氨基酸隔开。

[0093] 术语“氨基酸取代集(amino acid substitution set)”或“取代集”指与参考序列相比,多肽序列中的一组氨基酸取代。取代集可以具有1、2、3、4、 5、6、7、8、9、10、11、12、13、14、15或更多个氨基酸取代。在一些实施方案中,取代集指存在于实施例中提供的表格中所列的变体PGA的任一个中的氨基酸取代集。

[0094] 如本文使用的,“片段”指具有氨基末端和/或羧基末端缺失但其中剩余的氨基酸序列与序列中的对应位置相同的多肽。片段可以通常具有全长 PGA多肽,例如多肽SEQ ID

NO:2的约80％、约90％、约95％、约98％或约99％。在一些实施方案中，片段是“有生物活性的”(即它表现出与全长序列相同的酶促活性)。

[0095] 如本文使用的，“分离的多肽”指如下多肽:所述多肽与其天然伴随的其他污染物例如蛋白、脂质和多核苷酸大体上分离。该术语包括已从它们天然存在的环境或表达系统(例如，宿主细胞或体外合成)中取出或纯化的多肽。改进的PGA酶可以存在于细胞内、存在于细胞培养基中，或以多种形式制备，诸如裂解物或分离的制品。因此，在一些实施方案中，本公开内容的工程化PGA多肽可以是分离的多肽。

[0096] 如本文使用的，“大体上纯的多肽”指多肽物类为存在的优势物类(即，以摩尔或重量计，它比在该组合物中的任何其他单独的大分子物类更丰富) 的组合物，并且当目标物类构成存在的大分子物类的按摩尔或％重量计至少约50％时，通常为大体上纯化的组合物。通常，以存在于组合物中的所有大分子物质的摩尔或％重量计，大体上纯的工程化PGA多肽组合物占约 60％或更多、约70％或更多、约80％或更多、约90％或更多、约91％或更多、约92％或更多、约93％或更多、约94％或更多、约95％或更多、约 96％或更多、约97％或更多、约98％或更多、或约99％。溶剂物类、小分子(＜500道尔顿)和元素离子物类不被认为是大分子物类。在一些实施方案中，分离的改进的PGA多肽是大体上纯的多肽组合物。

[0097] 如本文中当关于核酸或多肽使用时，术语“异源的”指正常情况下生物体(例如，野生型生物体)不表达和分泌的序列。在一些实施方案中，该术语包括包含两个或更多个子序列的序列，发现所述子序列彼此之间的关系与在自然界中正常存在的关系不同，或所述子序列被重组工程化，使得其表达水平或与细胞中的其他核酸或其他分子的物理关系或结构不是正常存在于自然界中的。例如，异源的核酸通常被重组地产生，具有以在自然界中未发现的方式排列的来自不相关的基因的两个或更多个序列(例如，本发明的核酸开放阅读框(ORF)可操作地连接至被插入到表达盒诸如载体中的启动子序列)。在一些实施方案中，“异源多核苷酸”指通过实验室技术被引入到宿主细胞的任何多核苷酸，并且包括从宿主细胞中取出、进行实验室操作并且然后重新引入到宿主细胞的多核苷酸。

[0098] 如本文使用的，“合适的反应条件”指生物催化反应溶液中的那些条件 (例如酶载量、底物载量、辅因子载量、温度、pH、缓冲液、共溶剂等的范围)，在该条件下，本公开内容的PGA多肽能够通过除去三-乙酸苯酯保护基团来释放游离胰岛素。在本公开内容中提供了并通过实施例说明了示例性的“合适的反应条件”。

[0099] 如本文使用的，诸如“化合物载量”、“酶载量”或“辅因子载量”中的“载量”指在反应开始时反应混合物中组分的浓度或量。

[0100] 如本文使用的，在生物催化剂介导的方法的上下文中，“底物”指由生物催化剂作用的化合物或分子。

[0101] 如本文使用的，在生物催化剂介导的方法的上下文中，“产物”指由生物催化剂的作用产生的化合物或分子。

[0102] 如本文使用的，如本文使用的“平衡”指在化学或酶促反应中产生稳定状态浓度的化学物类的过程(例如两种物类A和B的相互转化)，包括立体异构体的相互转化，如通过该化学或酶促反应的正向速率常数和逆向速率常数确定的。

[0103] 如本文使用的，“酰化酶”和“酰基转移酶”可互换使用以指能够从供体转移酰基基团至受体以形成酯或酰胺的酶。酰化酶介导的逆反应引起酯或酰胺水解。

[0104] 如本文使用的，“青霉素G”和“苄基青霉素”指也被称为(2S,5R,6R)-3,3- 二甲基-7-氧杂-6-(2-苯基乙酰氨基)-4-硫杂-1-氮杂双环[3.2.0]庚烷-2-羧酸 ($C_{16}H_{18}N_2O_4S$)的抗生素。它主要有效针对革兰氏阳性生物体，但是一些革兰氏阴性生物体也对它敏感。

[0105] 如本文使用的，“青霉素G酰化酶”和“PGA”可互换使用以指具有介导青霉素G(苄基青霉素)裂解为苯乙酸(PHA)和6-氨基青霉烷酸(6-APA)的能力的酶。在一些实施方案中，PGA活性可以基于裂解模式底物，例如将 6-硝基-3-(苯基乙酰氨基)苯甲酸裂解为苯乙酸和5-氨基-2-硝基-苯甲酸。PGA还能够进行转移酰基供体的酰基基团至酰基受体的逆反应。如本文使用的，PGA包括天然存在的(野生型)PGA以及包含通过人类操作产生的一种或更多种工程化多肽的非天然存在的PGA酶。野生型PGA基因是由通过54个氨基酸的间隔区连接的α亚基(23.8KDa)和β亚基(62.2KDa)组成的异二聚体。由于存在间隔区，需要自动加工步骤(autoprocessing step)以形成活性蛋白。

[0106] 如本文使用的，“酰基供体”指酰化酶底物中将酰基基团贡献给酰基受体以形成酯或酰胺的部分。

[0107] 如本文使用的，“酰基受体”指酰化酶底物中接受酰基供体的酰基基团以形成酯或酰胺的部分。

[0108] 如本文使用的，“α-链序列”指对应于SEQ ID NO:2的位置27至235 处的残基的氨基酸序列(例如，与这些残基具有至少85％同一性)。如本文使用的，单链多肽可以包含“α-链序列”和另外的序列。

[0109] 如本文使用的，“β-链序列”指对应于SEQ ID NO:2的位置290至846 处的残基的氨基酸序列(例如与这些残基具有至少85％同一性)。如本文使用的，单链多肽可以包含“β-链序列”和另外的序列。

[0110] 如本文使用的，当在工程化PGA酶的上下文中使用时，“源自”标明工程化所基于原始PGA酶，和/或编码该PGA酶的基因。例如，SEQ ID NO：60的工程化PGA酶通过使编码SEQ ID NO:2的K.citrophila PGA α-链和β- 链序列的基因人工进化多代获得。因此，该工程化PGA酶“源自”SEQ ID NO:2的天然存在的或野生型PGA。

[0111] 如本文使用的，“胰岛素”指由正常个体的胰腺的β细胞产生的多肽激素。胰岛素是通过降低血糖水平调控糖类代谢必需的。全身性缺乏胰岛素导致糖尿病。胰岛素包括51个氨基酸并且具有约5800道尔顿的分子量。胰岛素包括2条肽链(指定为“A”和“B”)，包含1个亚基内和2个亚基间二硫键。A链包括21个氨基酸，且B链包括30个氨基酸。两条链形成高度有序的结构，在A链和B链两者中具有数个α-螺旋区。分离的链是无活性的。在溶液中，胰岛素是单体、二聚体或六聚体。它在用于皮下注射的高度浓缩的制品中是六聚体的，但当它在体液中被稀释时变成单体的。该定义意图包括胰岛素原和具有天然存在的胰岛素的部分或全部一级结构构型和至少一种生物学特性的任何纯化的分离的多肽。它还意图包括天然和合成来源的胰岛素，包括糖型(glycoform)以及类似物(例如，具有缺失、插入和/或取代的多肽)。

[0112] 胰岛素包含可以潜在地与乙酸苯酯供体反应并可以被PGA去保护的3 个亲核胺。这些残基包括在B链的位置29(B29)的Lys和两个N-末端游离胺:A链的位置1(A1)的Gly和B链的位置1(B1)的Phe。本文提供了三- 保护的胰岛素(化学连接至人类胰岛素上的A1、B1、B29残基的乙酸苯酯)。PGA先前已经被报道为以对乙酸苯酯酰胺键的排他选择性催化N-乙

酸苯酯保护的肽和胰岛素的水解，使蛋白的剩余肽键不受影响(Brtnik等人，Coll.Czech.Chem.Commun.,46(8),1983-1989[1981];和Wang等人.Biopolym.,25(增刊)，S109-S114[1986])。

[0113] 如本文使用的，“三-乙酸苯酯保护基团”指在B1、B29和A1位置具有三个被苯乙酰基基团保护的伯胺的胰岛素分子。

[0114] 如本文使用的，“二-乙酸苯酯保护基团”指在B1、B29和/或A1位置具有两个被苯乙酰基基团保护的伯胺的胰岛素分子。

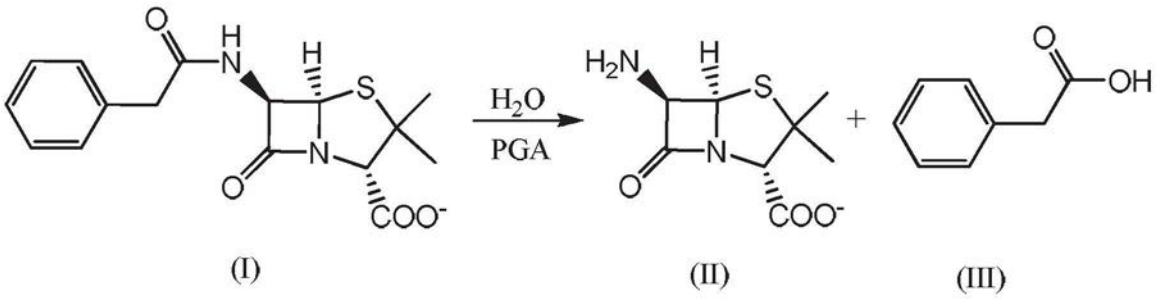
[0115] 如本文使用的，“二-乙酸苯酯保护基团”指在B1、B29或A1位置具有一个被苯乙酰基基团保护的伯胺的胰岛素分子。

[0116] 青霉素G酰化酶

[0117] 青霉素酰化酶首先由Sakaguchi和Murao描述为来自产黄青霉 (Penicillium chrysogenum)Wisc.Q176(Sakaguchi和Murao,J. Agr.Chem.Soc.Jpn.,23:411[1950])。青霉素G酰化酶是作用于青霉素G、头孢菌素G和相关的抗生素的侧链以产生β-内酰胺抗生素中间体6-氨基青霉烷酸和7-氨基脱乙酰氧基头孢菌素酸(cephalosporanic acid)的水解酶，以苯乙酸为常见副产物。这些抗生素中间体是半合成的抗生素，诸如氨苄青霉素、阿莫西林、氯唑西林、头孢氨苄和cefatoxime的潜在构成单元。

[0118] 如上文所述，青霉素G酰化酶(PGA)的特征在于催化具有结构式(I)的共轭碱的青霉素G水解裂解为具有结构式(II)的共轭碱的6-氨基青霉烷酸和结构式(III)的苯乙酸的能力，如方案1中示出的：

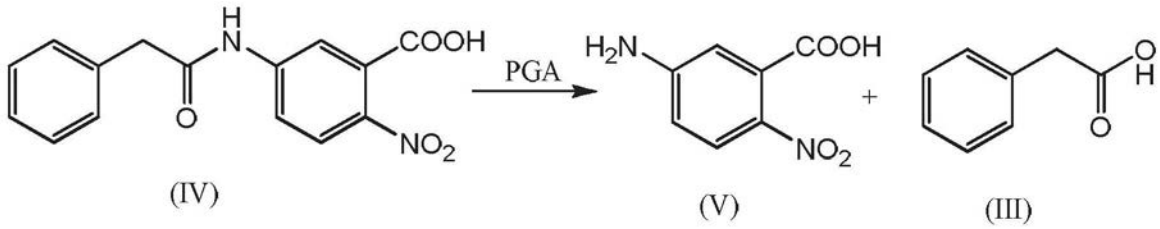
[0119]



方案 1

[0120] 虽然不囿于理论，底物特异性表现为与识别疏水性苯基基团相关，而亲核体充当β-内酰胺和多种其他基团诸如β-氨基酸的受体，该亲核体在一些PGA中是β-链的N-末端的丝氨酸残基。PGA的特征还可以在于裂解类似于青霉素G的模式底物，例如如方案2中示出的，将结构式(IV)的6-硝基-3-(苯乙酰氨基)苯甲酸(NIPAB)裂解为结构式(III)的苯乙酸和结构式(V) 的5-氨基-2-硝基苯甲酸的能力(参见例如，Alkema等人，Anal.Biochem.，275:47-53[1999])：

[0121]



(IV) (V) (III)

方案 2

[0122] 由于5-氨基-2-硝基-苯甲酸是生色的(chromogenic),式(IV)的底物提供了测量PGA活性的方便途径。除了上述反应以外,PGA还可以在DL-叔亮氨酸的动态拆分中使用,用于制备光学纯的叔亮氨酸(参见例如,Liu等人, Prep.Biochem.Biotechnol.,36:235-41[2006])。

[0123] 本公开内容的PGA是基于从生物体Kluyvera citrophila(K.citrophila) 获得的酶。如同来自其他生物体的PGA,K.citrophila的PGA是通过PGA 原前体多肽的蛋白水解加工产生的包括α亚基和β亚基的异二聚体酶。去除信号肽和间隔肽产生成熟的异二聚体(参见例如Barbero等人,Gene 49:69-80[1986])。天然存在的K.citrophila的PGA原前体多肽的氨基酸序列是可公开获得的(参见例如,Genbank登录号P07941,[gi:129551])并且在本文中以SEQ ID NO:2提供。天然存在的K.citrophila PGA的α-链序列对应于SEQ ID NO:2的残基27至235。天然存在的K.citrophila PGA的β- 链序列对应于SEQ ID NO:2的残基290至846。SEQ ID NO:2的残基1至 26对应于信号肽,且SEQ ID NO:2的残基236-289对应于连接前肽,信号肽和连接前肽两者都被去除以产生天然存在的成熟PGA酶,成熟PGA 酶是包含α-链亚基和β-链亚基的异二聚体。

[0124] 在一些实施方案中,本发明提供了具有与SEQ ID NO:4、8、14、300、 1036、1194、1262和/或1288具有至少约85%、86%、87%、88%、89%、 90%、91%、92%、93%、94%、95%、96%、97%、98%或99%或更多序列同一性的氨基酸序列的工程化PGA多肽。

[0125] 本发明提供了适用于商业规模用途的胰岛素特异性脱酰基生物催化剂。使用定向进化来开发能够去除或向胰岛素添加A1/B1/B29-三-乙酸苯酯保护基团的高效酰化酶变体。与野生型PGA相比,本文提供的PGA变体能够接受各种各样的酰基基团,表现出增加的溶剂稳定性和改进的热稳定性。本文提供的变体PGA缺少间隔区。因此,不需要自动加工步骤来产生有活性的酶。

[0126] 本发明还提供了编码工程化PGA多肽的多核苷酸。在一些实施方案中,多核苷酸被可操作地连接至控制基因表达的一个或更多个异源调控序列,以创建能够表达多肽的重组多核苷酸。包含编码工程化PGA多肽的异源多核苷酸的表达构建体可以被引入到适当的宿主细胞以表达对应的 PGA多肽。

[0127] 由于知晓对应于多种氨基酸的密码子,蛋白序列的可得性提供了对能够编码目标的所有多核苷酸的描述。遗传密码的简并性——其中相同的氨基酸由可选择的密码子或同义密码子编码——允许制备极大量的核酸,所有这些核酸都编码本文公开的改进的PGA酶。因此,已经鉴定了特定的氨基酸序列,本领域技术人员可以通过以不改变蛋白的氨基酸序列的方式简单改变一个或更多个密码子的序列来制备任何数目的不同核酸。在这方面,本

公开内容特别设想了通过选择基于可能的密码子选择的组合可以产生的每个和每一个可能的多核苷酸变化，并且所有这些变化应被认为针对本文公开的任何多肽(包括实施例5和6中的表中呈现的氨基酸序列)具体公开。

[0128] 在多种实施方案中，密码子被优选地选择为适应在其中产生蛋白的宿主细胞。例如，在细菌中使用的优选的密码子被用于在细菌中表达基因；在酵母中使用的优选的密码子被用于酵母中的表达；并且在哺乳动物中使用的优选的密码子被用于在哺乳动物细胞中表达。

[0129] 在某些实施方案中，不需要替换所有密码子以优化PGA多肽的密码子使用，因为天然序列将包含优选的密码子并且因为可不需要对所有氨基酸残基使用优选的密码子。因此，编码PGA酶的密码子优化的多核苷酸可以在全长编码区的约40％、50％、60％、70％、80％或大于90％的密码子位置包含优选密码子。

[0130] 在一些实施方案中，多核苷酸包含编码具有与本文描述的任一个参考工程化PGA多肽的α-链和/或β链具有至少约85％、86％、87％、88％、89％、 90％、91％、92％、93％、94％、95％、96％、97％、98％或99％或更多序列同一性的氨基酸序列的PGA多肽的核苷酸序列。相应地，在一些实施方案中，多核苷酸编码与基于SEQ ID NO:4、8、14、300、1036、1194、1262和/或1288的参考α-链和β-链序列至少约85％、86％、87％、88％、 89％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％或更多地相同的氨基酸序列。在一些实施方案中，多核苷酸编码SEQ ID NO:4、 8、14、300、1036、1194、1262和/或1288的α-链和/或β-链氨基酸序列。

[0131] 在一些实施方案中，多核苷酸包含编码与SEQ ID NO:4、8、14、300、 1036、1194、1262和1288具有至少约85％、86％、87％、88％、89％、90％、 91％、92％、93％、94％、95％、96％、97％、98％或99％或更多序列同一性的氨基酸序列的PGA多肽的核苷酸序列。相应地，在一些实施方案中，多核苷酸编码与SEQ ID NO:3、7、13、299、1035、1193、1261和1287 至少约85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、 95％、96％、97％、98％或99％或更多地相同的氨基酸序列。

[0132] 在一些实施方案中，编码改进的PGA多肽的分离的多核苷酸以多种方式被操作以提供改进的多肽活性和/或表达。取决于表达载体，在分离的多核苷酸插入到载体之前对分离的多核苷酸的操作可以是期望的或必要的。利用重组DNA方法修饰多核苷酸和核酸序列的技术是本领域熟知的。

[0133] 例如，诱变和定向进化的方法可以容易地应用至多核苷酸，以产生可以被表达、筛选和测定的变体库。诱变和定向进化方法是本领域熟知的(参见例如美国专利号5,605,793、5,811,238、5,830,721、5,834,252、5,837,458、 5,928,905、6,096,548、6,117,679、6,132,970、6,165,793、6,180,406、6,251,674、 6,265,201、6,277,638、6,287,861、6,287,862、6,291,242、6,297,053、6,303,344、 6,309,883、6,319,713、6,319,714、6,323,030、6,326,204、6,335,160、6,335,198、 6,344,356、6,352,859、6,355,484、6,358,740、6,358,742、6,365,377、6,365,408、 6,368,861、6,372,497、6,337,186、6,376,246、6,379,964、6,387,702、6,391,552、 6,391,640、6,395,547、6,406,855、6,406,910、6,413,745、6,413,774、6,420,175、 6,423,542、6,426,224、6,436,675、6,444,468、6,455,253、6,479,652、6,482,647、6,483,011、6,484,105、6,489,146、6,500,617、6,500,639、6,506,602、6,506,

603、6,518,065、6,519,065、6,521,453、6,528,311、6,537,746、6,573,098、6,576,467、6,579,678、6,586,182、6,602,986、6,605,430、6,613,514、6,653,072、6,686,515、6,703,240、6,716,631、6,825,001、6,902,922、6,917,882、6,946,296、6,961,664、6,995,017、7,024,312、7,058,515、7,105,297、7,148,054、7,220,566、7,288,375、7,384,387、7,421,347、7,430,477、7,462,469、7,534,564、7,620,500、7,620,502、7,629,170、7,702,464、7,747,391、7,747,393、7,751,986、7,776,598、7,783,428、7,795,030、7,853,410、7,868,138、7,783,428、7,873,477、7,873,499、7,904,249、7,957,912、7,981,614、8,014,961、8,029,988、8,048,674、8,058,001、8,076,138、8,108,150、8,170,806、8,224,580、8,377,681、8,383,346、8,457,903、8,504,498、8,589,085、8,762,066、8,768,871、9,593,326、以及所有相关的PCT和所有相关的非美国的对应申请;Ling等人,Anal.Biochem.,254(2):157-78[1997]; Dale等人,Meth.Mol.Biol.,57:369-74[1996];Smith,Ann.Rev.Genet.,19:423-462[1985];Botstein等人,Science,229:1193-1201[1985];Carter, Biochem.J.,237:1-7[1986];Kramer等人,Cell,38:879-887[1984];Wells等人,Gene,34:315-323[1985];Minshull等人,Curr.Op.Chem.Biol.,3:284-290 [1999];Christians等人,Nat.Biotechnol.,17:259-264[1999];Crameri等人, Nature,391:288-291[1998];Crameri,等人,Nat.Biotechnol.,15:436-438 [1997];Zhang等人,Proc.Nat.Acad.Sci.U.S.A.,94:4504-4509[1997];Crameri 等人,Nat.Biotechnol.,14:315-319[1996];Stemmer,Nature,370:389-391 [1994];Stemmer,Proc.Nat.Acad.Sci.USA,91:10747-10751[1994];WO 95/22625;WO 97/0078;WO 97/35966;WO 98/27230;WO 00/42651; WO 01/75767;和WO 2009/152336,其全部通过引用并入本文)。

[0134] 在一些实施方案中,本发明的变体PGA酰化酶还包括不改变酶的编码活性的另外的序列。例如,在一些实施方案中,变体PGA酰化酶被连接至可用于纯化的表位标签或另一个序列。

[0135] 在一些实施方案中,本发明的变体PGA酰化酶多肽由它们在其中被表达的宿主细胞(例如酵母宿主细胞或丝状真菌宿主细胞)分泌,并且被表达为包括信号肽(即连接至多肽的氨基末端并指导所编码的多肽进入细胞分泌途径的氨基酸序列)的前蛋白。

[0136] 在一些实施方案中,信号肽是内源的K.citrophila PGA酰化酶信号肽。在一些其他实施方案中,使用来自其他K.citrophila分泌蛋白的信号肽。在一些实施方案中,根据宿主细胞和其他因素使用其他信号肽。用于丝状真菌宿主细胞有效的信号肽编码区包括但不限于从以下获得的信号肽编码区:米曲霉(Aspergillus oryzae)TAKA淀粉酶、黑曲霉(Aspergillus niger) 中性淀粉酶、黑曲霉葡糖淀粉酶、米黑根毛霉(Rhizomucor miehei)天冬氨酸蛋白酶、特异腐质霉(Humicola insolens)纤维素酶、绵毛状腐质霉(Humicola lanuginosa)脂肪酶和里氏木霉(T.reesei)纤维二糖水解酶II。用于细菌宿主细胞的信号肽编码区包括但不限于从以下的基因获得的信号肽编码区:芽孢杆菌(Bacillus)NCIB 11837麦芽糖淀粉酶、嗜热脂肪芽孢杆菌α-淀粉酶、地衣芽孢杆菌(Bacillus licheniformis)枯草杆菌蛋白酶、地衣芽孢杆菌β-内酰胺酶、嗜热脂肪芽孢杆菌(Bacillus stearothermophilus)中性蛋白酶(nprT、nprS、nprM)、和枯草芽孢杆菌prsA。在一些另外的实施方案中,可以在本发明中使用其他信号肽(参见例如,Simonen and Palva,

Microbiol.Rev.,57：109-137[1993],通过引用并入本文)。另外的可用于酵母宿主细胞的信号肽包括来自酿酒酵母(Saccharomyces cerevisiae)α-因子、酿酒酵母SUC2转化酶的基因的那些(参见例如,Taussig和Carlson,Nucl.Acids Res.,11:1943-54 [1983];SwissProt登录号P00724;和Romanos等人,Yeast 8:423-488[1992])。在一些实施方案中,可以使用这些信号肽和其他信号肽的变体。事实上,本发明不意图限于任何具体的信号肽,因为本领域已知的任何合适的信号肽可以在本发明中使用。

[0137] 在一些实施方案中,本发明提供了编码变体PGA酰化酶多肽的如本文描述的多核苷酸,和/或其生物活性片段。在一些实施方案中,多核苷酸被可操作地连接至控制基因表达的一个或更多个异源调控或控制序列,以创建能够表达多肽的重组多核苷酸。在一些实施方案中,包含编码变体 PGA酰化酶的异源多核苷酸的表达构建体被引入到适当的宿主细胞中以表达变体PGA酰化酶。

[0138] 本领域普通技术人员理解,由于遗传密码的简并性,存在编码本发明的变体PGA多肽的多种核苷酸序列。例如,密码子AGA、AGG、CGA、 CGC、CGG和CGU都编码氨基酸精氨酸。因此,在本发明核酸中的其中精氨酸被密码子指定的每个位置处,该密码子可被更改为以上描述的对应密码子中的任一个而不改变编码的多肽。应理解,RNA序列中的“U”对应于DNA序列中的“T”。本发明设想并提供了可以通过选择基于可能的密码子选择的组合制备的编码本发明的多肽的核酸序列的每种和每一种可能的变异。

[0139] 如上文所指出的,编码PGA的DNA序列还可以被设计为高密码子使用偏好密码子(在蛋白编码区比编码相同氨基酸的其他密码子以更高的频率使用的密码子)。优选的密码子可以根据单个基因、共同功能或来源的一组基因、高表达基因中的密码子使用、整个生物体的聚集蛋白编码区中的密码子频率、相关生物体的聚集蛋白编码区中的密码子频率或其组合来确定。其频率随着基因表达的水平而增加的密码子通常是用于表达的最佳密码子。特别地,DNA序列可以被优化以便在特定宿主生物体中表达。用于确定在特定生物体中的密码子频率(例如密码子使用、相对同义密码子使用) 和密码子偏好的多种方法是本领域已知的,包括对基因中使用的密码子的多变量分析(例如使用聚类分析或对应分析)和有效数目。用于获得密码子使用的数据源可以依赖于能够编码蛋白的任何可获得的核苷酸序列。如本领域熟知的,这些数据集包括实际已知编码被表达的蛋白(例如完整的蛋白编码序列-CDS)、被表达的序列标签(EST)或基因组序列的预测编码区的核酸序列。编码变体PGA的多核苷酸可以使用本领域已知的任何适当的方法制备。一般地,寡核苷酸被单独地合成,然后连接(通过例如酶促连接或化学连接方法或聚合酶介导的方法),以基本上形成任何期望的连续序列。在一些实施方案中,本发明的多核苷酸使用本领域已知的任何合适的方法通过化学合成制备,包括但不限于自动合成方法。例如,在亚磷酰胺方法中,寡核苷酸被合成(例如,在自动的DNA合成仪中)、纯化、退火、连接并克隆入适当的载体。在一些实施方案中,双链DNA片段然后通过合成互补链和将链在适当条件下退火到一起,或通过用适当引物序列使用DNA 聚合酶添加互补链获得。存在多种常规和标准教科书,提供可在本发明中使用的本领域技术人员熟知的方法。

[0140] 如以上讨论的,可以通过对编码天然存在的PGA的多核苷酸进行诱变和/或定向进化方法来获得工程化PGA。诱变可以根据本领域已知的任何技术进行,包括随机诱变和定点诱变。定向进化可以用本领域已知的任何技术进行(包括重排(shuffling)),以筛选改进的

变体。其他可使用的定向进化程序包括但不限于，交错延伸程序(StEP)、体外重组、诱变PCR、盒式诱变、通过重叠延伸剪接(SOEing)、ProSAR™定向进化方法等，以及任何其他合适的方法。

[0141] 对在诱变处理后获得的克隆筛选具有期望的改进的酶特性的工程化 PGA。可以使用监测产物形成速率的标准生物化学技术对来自表达文库的酶活性进行测量。当期望的改进的酶特性是热稳定性时，可以在将酶制品置于指定的温度并测量热处理后剩余的酶活性的量后测量酶活性。然后包含编码PGA的多核苷酸的克隆被分离、测序以鉴定核苷酸序列的变化(如果有)，并且用于在宿主细胞中表达酶。

[0142] 当已知工程化多肽的序列时，根据已知的合成方法可以通过标准固相方法制备编码该酶的多核苷酸。在一些实施方案中，多达约100个碱基的片段可以被单独地合成，然后连接(例如，通过酶促或化学连接方法 (chemical ligation method)或聚合酶介导的方法)以形成任何期望的连续序列。例如，本发明的多核苷酸和寡核苷酸可以通过化学合成制备(例如，使用Beaucage等，Tetra.Lett.，22:1859-69[1981]描述的经典的亚磷酰胺方法或Matthes等人，EMBO J.，3:801-05[1984]描述的方法，因为它通常以自动合成方法实施)。根据亚磷酰胺方法，寡核苷酸被合成(例如，在自动的DNA 合成仪中)、纯化、退火、连接并克隆入适当的载体。另外，基本上任何核酸可以从多种商业来源中的任一个获得(例如，The Midland Certified Reagent Company，Midland，TX，The Great American Gene Company，Ramona，CA，ExpressGen Inc.Chicago，IL，Operon Technologies Inc.， Alameda，CA以及许多其他商业来源)。

[0143] 本发明还提供了包含编码如本文提供的至少一种变体PGA的序列的重组构建体。在一些实施方案中，本发明提供了一种表达载体，所述表达载体包含可操作地连接至异源启动子的变体PGA多核苷酸。在一些实施方案中，本发明的表达载体被用于转化适当的宿主细胞以允许该宿主细胞表达变体PGA蛋白。用于在真菌和其他生物体中重组表达蛋白的方法是本领域熟知的，并且多种表达载体是可获得的或可以使用常规方法构建。在一些实施方案中，本发明的核酸构建体包含本发明的核酸序列插入其中的载体，诸如质粒、黏粒、噬菌体、病毒、细菌人工染色体(BAC)、酵母人工染色体(YAC)等。在一些实施方案中，本发明的多核苷酸被整合到适用于表达变体PGA多肽的多种表达载体的任一种中。合适的载体包括但不限于，染色体、非染色体和合成的DNA序列(例如SV40的衍生物)，以及细菌质粒、噬菌体DNA、杆状病毒(baculovirus)、酵母质粒、源自质粒和噬菌体DNA的组合的载体、病毒DNA诸如牛痘、腺病毒、禽痘病毒、假狂犬病、腺病毒、腺病毒伴随病毒、逆转录病毒以及许多其他载体。本发明可以使用将遗传材料转导到细胞内的任何合适的载体，并且如果需要复制，该载体在相关宿主中是可复制并且可生存的。在一些实施方案中，构建体还包括可操作地连接至蛋白编码序列的调控序列，包括但不限于启动子。大量合适的载体和启动子是本领域的技术人员已知的。事实上，在一些实施方案中，为了在特定宿主中获得高表达水平，在异源启动子的控制下表达本发明的变体PGA经常是有用的。在一些实施方案中，使用本领域已知的任何合适的方法将启动子序列可操作地连接至变体PGA编码序列的5’区。用于表达变体PGA的有用的启动子的实例包括但不限于来自真菌的启动子。在一些实施方案中，可使用真菌菌株中驱动除PGA基因以外的基因表达的启动子序列。作为一个非限制性实例，可以使用来自编码内切葡聚糖酶的基因的真菌启动子。在一些实施方案中，可以使用驱动PGA基因

在除该PGA所来源的真菌菌株以外的真菌菌株中表达的启动子序列。可用于指导本发明的核苷酸构建体在丝状真菌宿主细胞中的转录的其他合适的启动子的实例包括但不限于从以下的基因获得的启动子:米曲霉TAKA淀粉酶、米黑根毛霉天冬氨酸蛋白酶、黑曲霉中性α-淀粉酶、黑曲霉酸稳定性α-淀粉酶、黑曲霉或泡盛曲霉(Aspergillus awamori)葡糖淀粉酶(glaA)、米黑根毛霉脂肪酶、米曲霉碱性蛋白酶、米曲霉磷酸丙糖异构酶、构巢曲霉(Aspergillus nidulans)乙酰胺酶和尖孢镰孢菌(Fusarium oxysporum)胰蛋白酶样蛋白酶(参见例如WO 96/00787,通过引用并入本文),以及NA2-tpi启动子(来自黑曲霉中性α-淀粉酶和米曲霉磷酸丙糖异构酶的基因的启动子的杂合体)、启动子诸如cbh1、cbh2、egl1、egl2、pepA、hfb1、 hfb2、xyn1、amy和glaA(参见例如,Nunberg等人,Mol.Cell Biol.,4:2306 -2315[1984];Boel等人,EMBO J.,3:1581-85[1984];和欧洲专利申请 137280,这些全部通过引用并入本文),以及它们的突变的、截短的和杂合的启动子。

[0144] 在酵母宿主细胞中,有用的启动子包括但不限于,来自以下的基因的那些:酿酒酵母烯醇化酶(eno-1)、酿酒酵母半乳糖激酶(gal1)、酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(ADH2/GAP)、和酿酒酵母3-磷酸甘油酸激酶。可用于酵母宿主细胞的另外的有用的启动子是本领域已知的(参见例如, Romanos等人,Yeast 8:423-488[1992],通过引用并入本文)。另外,与真菌中壳多糖酶产生相关的启动子可用于本发明(参见例如,Blaiseau和Lafay, Gene 120243-248[1992];和Limon等人,Curr.Genet.,28:478-83(1995),两者通过引用并入本文)。

[0145] 对于细菌宿主细胞,用于指导本公开内容的核酸构建体的转录的合适启动子包括但不限于,从以下获得的启动子:大肠杆菌lac操纵子、大肠杆菌trp操纵子、噬菌体λ、天蓝色链霉菌(Streptomyces coelicolor)琼脂糖酶基因(dagA)、枯草芽孢杆菌果聚糖蔗糖酶基因(sacB)、地衣芽孢杆菌α- 淀粉酶基因(amyL)、嗜热脂肪芽孢杆菌麦芽糖淀粉酶基因(amyM)、解淀粉芽孢杆菌(Bacillus amyloliquefaciens)α-淀粉酶基因(amyQ)、地衣芽孢杆菌青霉素酶基因(penP)、枯草芽孢杆菌xylA和xylB基因及原核β-内酰胺酶基因(参见例如,Villa-Kamaroff等人,Proc.Natl.Acad.Sci.USA 75:3727-3731 [1978])、以及tac启动子(参见例如,DeBoer等人,Proc.Natl.Acad.Sci.USA 80: 21-25[1983])。

[0146] 在一些实施方案中,本发明的克隆的变体PGA还具有合适的转录终止子序列,转录终止子序列是由宿主细胞识别以终止转录的序列。终止子序列被可操作地连接至编码多肽的核酸序列的3′末端。在选择的宿主细胞中有功能的任何终止子可用于本发明中。用于丝状真菌宿主细胞的示例性转录终止子包括但不限于,从以下的基因获得的那些:米曲霉TAKA淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、黑曲霉α-葡萄糖苷酶和尖孢镰刀菌胰蛋白酶样蛋白酶(还参见美国专利号7,399,627,其通过引用并入本文)。在一些实施方案中,用于酵母宿主细胞的示例性终止子包括从以下的基因获得的那些:酿酒酵母烯醇化酶、酿酒酵母细胞色素C (CYC1)、和酿酒酵母甘油醛-3-磷酸脱氢酶。用于酵母宿主细胞的其他有用的终止子是本领域已知的(参见例如,Romanos等,Yeast 8:423-88[1992])。

[0147] 在一些实施方案中,合适的前导序列为克隆的变体PGA序列的一部分,前导序列是一种mRNA的对宿主细胞翻译重要的非翻译区。前导序列被可操作地连接至编码多肽的核酸序列的5′末端。在本发明中可以使用在选择的宿主细胞中有功能的任何前导序列。用于丝状真菌宿主细胞的示例性前导序列包括但不限于从以下的基因获得的那些:米曲霉TAKA淀

粉酶和构巢曲霉磷酸丙糖异构酶。用于酵母宿主细胞的合适的前导序列包括,但不限于,从以下的基因获得的那些:酿酒酵母烯醇化酶(ENO-1)、酿酒酵母3-磷酸甘油酸激酶,酿酒酵母α-因子、和酿酒酵母醇脱氢酶/甘油醛-3- 磷酸脱氢酶(ADH2/GAP)。

[0148] 在一些实施方案中,本发明的序列还包括多腺苷酸化序列,多腺苷酸化序列是一种可操作地连接至核酸序列的3′末端的序列,并且其在转录时,被宿主细胞识别为将多腺苷残基添加至转录的mRNA的信号。在本发明中可以使用在选择的宿主细胞中有功能的任何多腺苷酸化序列。用于丝状真菌宿主细胞的示例性多腺苷酸化序列包括,但不限于从以下的基因获得的那些:米曲霉TAKA淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、尖孢镰刀菌胰蛋白酶样蛋白酶和黑曲霉α葡糖苷酶。用于酵母宿主细胞的有用的多腺苷酸化序列是本领域已知的(参见例如Guo和Sherman, Mol.Cell.Biol.,15:5983-5990[1995])。

[0149] 在一些实施方案中,控制序列包括编码连接至多肽的氨基末端的氨基酸序列的信号肽编码区并将编码的多肽引导到细胞的分泌途径中。核酸序列的编码序列的5′末端可以固有地包含信号肽编码区,其符合翻译阅读框地(in translation reading frame)与编码分泌多肽的编码区的区段天然地连接。可选地,编码序列的5′末端可以包含对编码序列而言外源的信号肽编码区。在编码序列不天然包含信号肽编码区的情况下可能需要外源信号肽编码区。

[0150] 可选地,外源信号肽编码区可以简单地替换天然信号肽编码区以增加多肽的分泌。然而,在本发明中可以使用指导所表达的多肽进入选择的宿主细胞的分泌途径的任何信号肽编码区。

[0151] 用于细菌宿主细胞的有效的信号肽编码区包括但不限于从以下的基因获得的信号肽编码区:芽孢杆菌NClB 11837麦芽糖淀粉酶、嗜热脂肪芽孢杆菌α-淀粉酶、地衣芽孢杆菌枯草杆菌蛋白酶、地衣芽孢杆菌β-内酰胺酶、嗜热脂肪芽孢杆菌中性蛋白酶(nprT、nprS、nprM)、和枯草芽孢杆菌prsA。另外的信号肽是本领域已知的(参见例如,Simonen和Palva,Microbiol.Rev.,57:109-137[1993])。

[0152] 用于丝状真菌宿主细胞有效的信号肽编码区包括但不限于从以下的基因获得的信号肽编码区:米曲霉TAKA淀粉酶、黑曲霉中性淀粉酶、黑曲霉葡糖淀粉酶、米黑根毛霉天冬氨酸蛋白酶、特异腐质霉纤维素酶和绵毛状腐质霉(Humicola lanuginosa)脂肪酶。

[0153] 用于酵母宿主细胞的有用的信号肽包括但不限于酿酒酵母α因子和酿酒酵母转化酶的基因。其他有用的信号肽编码区是本领域已知的(参见例如 Romanos等人,[1992],同上)。

[0154] 在一些实施方案中,控制序列包括前肽编码区,其编码定位在多肽的氨基末端处的氨基酸序列。所得多肽被称为酶原(proenzyme)或多肽原(在某些情况下称为酶原(zymogen))。多肽原通常是无活性的,并可以通过前肽从多肽原的催化裂解或自催化裂解被转化为成熟的有活性的PGA多肽。前肽编码区可以从枯草芽孢杆菌碱性蛋白酶(aprE)、枯草芽孢杆菌中性蛋白酶(nprT)、酿酒酵母α-因子、米黑根毛霉天冬氨酸蛋白酶和嗜热毁丝霉 (Myceliophthora thermophila)乳糖酶的基因获得(参见例如WO 95/33836)。

[0155] 当信号肽和前肽区均存在于多肽的氨基末端时,前肽区域紧邻多肽的氨基末端定位并且信号肽区紧邻前肽区的氨基末端定位。

[0156] 在一些实施方案中，还使用调控序列以允许相对于宿主细胞的生长调控多肽的表达。调控系统的实例为引起基因的表达响应于化学或物理刺激 (包括调控性化合物的存在)而开启或关闭的那些。在原核宿主细胞中，合适的调控序列包括但不限于lac、tac和trp操纵子系统。在酵母宿主细胞中，合适的调控系统包括例如ADH2系统或GAL1系统。在丝状真菌中，合适的调控序列包括TAKAα-淀粉酶启动子、黑曲霉葡糖淀粉酶启动子和米曲霉葡糖淀粉酶启动子。

[0157] 调控序列的其他实例是允许基因扩增的那些。在真核系统中，这些包括在甲氨蝶呤存在下扩增的二氢叶酸还原酶基因以及用重金属扩增的金属硫蛋白基因。在这些情况下，编码本发明的PGA多肽的核酸序列将与调控序列可操作地连接。

[0158] 因此，在另外的实施方案中，本发明提供了包含编码工程化PGA多肽或其变体的多核苷酸以及一个或更多个表达调控区诸如启动子和终止子、复制起点等(根据它们将被引入的宿主的类型)的重组表达载体。在一些实施方案中，以上描述的多种核酸和控制序列被连接在一起以产生重组表达载体，其可以包括一个或更多个方便的限制性位点，以允许在此类位点插入或取代编码多肽的核酸序列。可选地，在一些实施方案中，核酸序列通过将该核酸序列或包含该核酸序列的核酸构建体插入到用于表达的合适的载体来表达。在创建表达载体时，编码序列以使编码序列与用于表达的适当的控制序列可操作地连接的方式位于载体中。

[0159] 重组表达载体包括任何合适的载体(例如，质粒或病毒)，其可以方便地进行重组DNA程序并且可以引起多核苷酸序列的表达。载体的选择通常取决于载体与待引入该载体的宿主细胞的相容性。在一些实施方案中，载体是线性质粒或闭合的环状质粒。

[0160] 在一些实施方案中，表达载体为自主复制载体(即，作为染色体外的实体存在的载体，其复制独立于染色体复制，诸如质粒、染色体外元件、微型染色体或人工染色体)。在一些实施方案中，载体包含用于确保自我复制的任何工具(means)。可选地，在一些其他实施方案中，当被引入到宿主细胞时，载体被整合到基因组中并与其整合的染色体一起复制。此外，在另外的实施方案中，可以使用单一载体或质粒或者一起包含待引入到宿主细胞基因组中的总DNA的两种或多种载体或质粒，或者转座子。

[0161] 在一些实施方案中，本发明的表达载体包含一个或更多个可选择标志物，其允许容易选择经转化的细胞。“可选择标志物”为其产物提供杀生物剂或病毒抗性、对抗微生物剂或重金属的抗性、对营养缺陷型的原养型等的基因。可以在本发明中使用用于在丝状真菌宿主细胞中使用的任何合适的可选择标志物，包括但不限于amdS(乙酰胺酶)、argB(鸟氨酸氨甲酰转移酶)、bar(膦丝菌素乙酰转移酶)、hph(潮霉素磷酸转移酶)、niaD(硝酸还原酶)、pyrG(乳清酸核苷-5′-磷酸脱羧酶)、sC(硫酸腺苷酰转移酶(sulfate adenyltransferase))和trpC(邻氨基苯甲酸合成酶)，以及其等同物。在宿主细胞诸如曲霉属真菌(Aspergillus)中可用的另外的标志物包括但不限于构巢曲霉或米曲霉的amdS和pyrG基因，和吸水链霉菌(Streptomyces hygroscopicus)的bar基因。用于酵母宿主细胞的合适的标志物包括但不限于ADE2、HIS3、LEU2、LYS2、MET3、TRP1和URA3。细菌的可选择标志物的实例包括但不限于，来自枯草芽孢杆菌或地衣芽孢杆菌的dal基因，或赋予抗生素抗性诸如氨苄青霉素、卡那霉素、氯霉素和或四环素抗性的标志物。

[0162] 在一些实施方案中，本发明的表达载体包含允许载体整合到宿主细胞的基因组中

或允许载体在细胞中独立于基因组自主复制的元件。在一些涉及整合到宿主细胞基因组中的实施方案中，载体依赖于编码多肽的核酸序列或用于通过同源或非同源重组将载体整合到基因组的载体的任何其他元件。

[0163] 在一些可选的实施方案中，表达载体包含用于指导通过同源重组整合到宿主细胞的基因组中的另外的核酸序列。另外的核酸序列允许载体在染色体的准确位置整合到宿主细胞基因组中。为了增加在准确位置整合的可能性，整合元件优选地包含充足数目的核苷酸，诸如100至10,000碱基对，优选地400至10,000碱基对，且最优选的800至10,000碱基对，所述充足数目的核苷酸与对应的靶序列高度同源以增加同源重组的可能性。整合元件可以是与宿主细胞的基因组中的靶序列同源的任何序列。此外，整合元件可以是非编码的或编码的核酸序列。在另一方面，载体可以通过非同源重组整合到宿主细胞的基因组中。

[0164] 对于自主复制，载体还可以包括允许载体在所研究的宿主细胞中自主复制的复制起点。细菌复制起点的实例是允许在大肠杆菌中复制的P15A ori或质粒pBR322、pUC19、pACYC177(这些质粒具有P15A ori)、或 pACYC184的复制起点、和允许在芽孢杆菌杆菌中复制的pUB110、pE194、 pTA1060或pAMβ1。用于在酵母宿主细胞中使用的复制起点的实例是2微米复制起点(2 micron origin of replication)、ARS1、ARS4、ARS1和CEN3 的组合、以及ARS4和CEN6的组合。复制起点可以是具有使其在宿主细胞中的功能对温度敏感的突变的复制起点的复制起点(参见例如Ehrlich, Proc.Natl.Acad.Sci.USA 75:1433[1978])。

[0165] 在一些实施方案中，多于一个拷贝的本发明的核酸序列被插入到宿主细胞中以增加基因产物的产生。核酸序列的拷贝数的增加可以通过将至少一个另外拷贝的序列整合到宿主细胞基因组中，或通过包括具有核酸序列的可扩增选择标志物基因获得，其中包含选择标志物基因的扩增拷贝并从而包含核酸序列的另外拷贝的细胞可以通过在适当的可选择剂(selectable agent)的存在下培养细胞被选择。

[0166] 用于在本发明中使用的许多表达载体是商购可得的。合适的商购表达载体包括但不限于，p3xFLAGTM™表达载体(Sigma-Aldrich Chemicals)，其包括CMV启动子和用于在哺乳动物宿主细胞中表达的hGH多聚腺苷酸化位点和用于在大肠杆菌中扩增的pBR322复制起点和氨苄青霉素抗性标志物。其他合适的表达载体包括但不限于，pBluescriptII SK(-)和pBK-CMV (Stratagene)，以及源自pBR322(Gibco BRL)、pUC(Gibco BRL)、pREP4、 pCEP4(Invitrogen)或pPoly的质粒(参见例如Lathe等人，Gene 57:193-201 [1987])。

[0167] 因此，在一些实施方案中，包含编码至少一种变体PGA的序列的载体被转化到宿主细胞中以允许载体繁殖和表达变体PGA。在一些实施方案中，变体PGA被翻译后修饰以去除信号肽，并且在一些情况下可以在分泌后被裂解。在一些实施方案中，以上描述的转化的宿主细胞在允许表达变体PGA的条件下在合适的营养培养基中培养。可以在本发明中使用可用于培养宿主细胞的任何合适的培养基，包括但不限于基本培养基或包含适当补充剂的复合培养基。在一些实施方案中，宿主细胞生长于HTP培养基。合适的培养基可以从多种商业供应商获得，或可以根据已公开的配方 (例如，在美国典型培养物保藏中心(the American Type Culture Collection) 的目录中)制备。

[0168] 在另一方面，本发明提供了包含编码本文提供的改进的PGA多肽的多核苷酸的宿主细胞，所述多核苷酸可操作地连接至用于在宿主细胞中表达PGA酶的一个或更多个控制序列。用于在表达由本发明的表达载体编码的PGA多肽中使用的宿主细胞是本领域熟知的，

并且包括但不限于细菌细胞诸如大肠杆菌、巨大芽孢杆菌(Bacillus megaterium)、开菲尔乳杆菌 (Lactobacillus kefir)、链霉菌(Streptomyces)和鼠伤寒沙门氏菌(Salmonella typhimurium)细胞;真菌细胞诸如酵母细胞(例如酿酒酵母或巴斯德毕赤酵母(Pichia pastoris)(ATCC登录号201178));昆虫细胞诸如果蝇S2和灰翅夜蛾属(Spodoptera)Sf9细胞;动物细胞诸如CHO、COS、BHK、293和 Bowes黑素瘤细胞;和植物细胞。用于以上描述的宿主细胞的适当的培养基以及生长条件是本领域熟知的。

[0169] 用于表达PGA的多核苷酸可以通过本领域已知的多种方法引入细胞。技术,除其他以外,包括电穿孔、生物弹射粒子轰击(biolistic particle bombardment)、脂质体介导的转染、氯化钙转染和原生质体融合。用于将多核苷酸引入到细胞的多种方法是本领域技术人员已知的。

[0170] 在一些实施方案中,宿主细胞为真核细胞。合适的真核宿主细胞包括但不限于真菌细胞、藻类细胞、昆虫细胞和植物细胞。合适的真菌宿主细胞包括但不限于子囊菌门(Ascomycota)、担子菌门(Basidiomycota)、半知菌门(Deuteromycota)、接合菌门(Zygomycota)、半知菌(Fungi imperfecti)。在一些实施方案中,真菌宿主细胞是酵母细胞和丝状真菌细胞。本发明的丝状真菌宿主细胞包括亚门真菌亚门(Eumycotina)和卵菌门(Oomycota)的所有丝状形式。丝状真菌的特征在于具有壳多糖、纤维素和其他复杂多糖构成的细胞壁的营养菌丝体。本发明的丝状真菌宿主细胞从形态学上可与酵母不同。

[0171] 在本发明的一些实施方案中,丝状真菌宿主细胞为任何合适的属和种,包括但不限于绵霉属(Achlya)、枝顶孢属(Acremonium)、曲霉属(Aspergillus)、短梗霉属(Aureobasidium)、烟管菌属(Bjerkandera)、拟蜡菌属(Ceriporiopsis)、头孢霉属(Cephalosporium)、金孢子菌属(Chrysosporium)、旋孢腔菌属 (Cochliobolus)、棒囊壳属(Corynascus)、丛赤壳属(Cryphonectria)、隐球菌属(Cryptococcus)、鬼伞属(Coprinus)、革盖菌属(Coriolus)、色二孢属 (Diplodia)、Endothis、镰孢属(Fusarium)、赤霉菌属(Gibberella)、胶枝霉属 (Gliocladium)、腐质霉属(Humicola)、肉座菌属(Hypocrea)、毁丝霉属 (Myceliophthora)、毛霉属(Mucor)、脉孢菌属(Neurospora)、青霉菌属(Penicillium)、柄孢壳菌属(Podospora)、白腐菌属(Phlebia)、瘤胃壶菌属 (Piromyces)、梨形孢属(Pyricularia)、根毛霉属(Rhizomucor)、根霉属 (Rhizopus)、裂褶菌属(Schizophyllum)、柱顶孢霉属(Scytalidium)、孢子丝菌属(Sporotrichum)、踝节菌属(Talaromyces)、热子囊菌属(Thermoascus)、梭孢壳属(Thielavia)、栓菌属(Trametes)、弯颈霉属(Tolypocladium)、木霉属 (Trichoderma)、轮枝孢属(Verticillium)和/或小包脚菇属(Volvariella)、和/或它们的有性型或无性型、和同义型(synonyms)、基原异名或分类学等同物。

[0172] 在本发明的一些实施方案中,宿主细胞为酵母细胞,包括但不限于假丝酵母属(Candida)、汉逊酵母属(Hansenula)、酵母属(Saccharomyces)、裂殖酵母属(Schizosaccharomyces)、毕赤酵母属(Pichia)、克鲁维酵母属 (Kluyveromyces)或亚罗酵母属(Yarrowia)物种的细胞。在本发明的一些实施方案中,酵母细胞是多形汉逊酵母(Hansenula polymorpha)、酿酒酵母、卡氏酵母(Saccharomyces carlsbergensis)、糖化酵母(Saccharomyces diastaticus)、诺地酵母(Saccharomyces norbensis)、克氏酵母(Saccharomyces kluyveri)、粟酒裂殖酵母(Schizosaccharomyces pombe)、巴斯德毕赤酵

母(Pichia pastoris)、芬兰毕赤酵母(Pichiafinlandica)、Pichia trehalophila、Pichia kodamae、膜醭毕赤氏酵母(Pichia membranaefaciens)、Pichia opuntiae、Pichia thermotolerans、Pichia salictaria、Pichia quercuum、Pichia pijperi、树干毕赤酵母(Pichia stipitis)、甲醇毕赤酵母(Pichia methanolica)、安格斯毕赤酵母(Pichia angusta)、乳酸克鲁维酵母(Kluyveromyces lactis)、白色念珠菌 (Candida albicans)或解脂耶氏酵母(Yarrowia lipolytica)。

[0173] 在本发明的一些实施方案中，宿主细胞为藻类细胞，诸如衣藻属(Chlamydomonas)(例如莱茵哈德衣藻(C.reinhardtii))和席藻属 (Phormidium)(P.sp.ATCC29409)。

[0174] 在一些其他实施方案中，宿主细胞为原核细胞。合适的原核细胞包括但不限于革兰氏阳性、革兰氏阴性和革兰氏不定的细菌细胞。可以在本发明中使用的任何合适的细菌生物体包括但不限于土壤农杆菌属 (Agrobacterium)、脂环酸芽孢杆菌属(Alicyclobacillus)、鱼腥藻属(Anabaena)、倒囊藻属(Anacystis)、不动杆菌属(Acinetobacter)、嗜酸菌属(Acidothermus)、节杆菌属(Arthrobacter)、定氮菌(Azobacter)、芽孢杆菌属(Bacillus)、双歧杆菌属(Bifidobacterium)、短杆菌属(Brevibacterium)、丁酸弧菌属 (Butyrivibrio)、布赫纳氏菌属(Buchnera)、Campestris、弯曲杆菌属 (Camplyobacter)、梭菌属(Clostridium)、棒杆菌属(Corynebacterium)、着色菌属(Chromatium)、粪球菌属(Coprococcus)、埃希氏菌属(Escherichia)、肠球菌属(Enterococcus)、肠杆菌属(Enterobacter)、欧文氏菌属(Erwinia)、梭杆菌属(Fusobacterium)、Faecalibacterium、弗朗西斯菌属(Francisella)、黄杆菌属(Flavobacterium)、地芽孢杆菌属(Geobacillus)、嗜血杆菌属 (Haemophilus)、螺杆菌属(Helicobacter)、克雷伯菌属(Klebsiella)、乳杆菌属(Lactobacillus)、乳球菌属(Lactococcus)、泥杆菌属(Ilyobacter)、微球菌属(Micrococcus)、细杆菌属(Microbacterium)、生根瘤菌属(Mesorhizobium)、甲基杆菌属(Methylobacterium)、甲基杆菌属(Methylobacterium)、分枝杆菌属(Mycobacterium)、奈瑟菌属(Neisseria)、泛菌属(Pantoea)、假单胞菌属 (Pseudomonas)、原绿球藻属(Prochlorococcus)、红细菌属(Rhodobacter)、红假单胞菌属(Rhodopseudomonas)、红假单胞菌属(Rhodopseudomonas)、罗氏菌属(Roseburia)、红螺菌属(Rhodospirillum)、红球菌属(Rhodococcus)、栅藻属(Scenedesmus)、链霉菌属(Streptomyces)、链球菌属(Streptococcus)、 Synecoccus、糖单孢菌属(Saccharomonospora)、葡萄球菌属(Staphylococcus)、沙雷氏菌属(Serratia)、沙门氏菌属(Salmonella)、志贺氏菌属(Shigella)、嗜热厌氧杆菌属(Thermoanaerobacterium)、Tropheryma、Tularensis、Temecula、热聚球藻属(Thermosynechococcus)、高温球菌属(Thermococcus)、脲原体属 (Ureaplasma)、黄单胞菌属(Xanthomonas)、木杆菌属(Xylella)、耶尔森氏菌属(Yersinia)和发酵单胞菌属(Zymomonas)。在一些实施方案中，宿主细胞为以下物种：土壤农杆菌属、不动杆菌属、固氮菌属、芽孢杆菌属、双歧杆菌属、布赫纳氏菌属、地芽胞杆菌属(Geobacillus)、弯曲杆菌属、梭菌属、棒杆菌属、埃希氏菌属、肠球菌属、欧文氏菌属、黄杆菌属、乳杆菌属、乳球菌属、泛菌属、假单胞菌属、葡萄球菌属、沙门氏菌属、链球菌属、链霉菌属或发酵单胞菌属。在一些实施方案中，细菌宿主菌株是对人类非致病性的。在一些实施方案中，细菌宿主菌株是工业

菌株。许多细菌工业菌株是已知的，并且在本发明中是合适的。在本发明的一些实施方案中，细菌宿主细胞是土壤杆菌属(Agrobacterium)的物种(例如放射型土壤杆菌(A.radiobacter)、发根土壤杆菌(A.rhizogenes)和悬钩土壤杆菌(A.rubi))。在本发明的一些实施方案中，细菌宿主细胞是节杆菌属的物种(例如金黄节杆菌(A.aurescens)、柠檬色节杆菌(A.citreus)、球形节杆菌(A.globiformis)、裂烃谷氨酸节杆菌(A.hydrocarboglutamicus)、迈索尔节杆菌(A.mysorens)、烟草节杆菌(A.nicotianae)、石蜡节杆菌(A.paraffineus)、A.protophonniae、 A.roseoparqffinus、硫磺节杆菌(A.sulfureus)和产脲节杆菌(A.ureafaciens))。在本发明的一些实施方案中，细菌宿主细胞是芽孢杆菌属的物种(例如苏云金芽孢杆菌(B.thuringensis)、炭疽芽孢杆菌(B.anthracis)、巨大芽胞杆菌(B. megaterium)、枯草芽孢杆菌(B.subtilis)、迟缓芽孢杆菌(B.lentus)、环状芽胞杆菌(B.circulans)、短小芽胞杆菌(B.pumilus)、灿烂芽孢杆菌(B.lautus)、凝结芽孢杆菌(B.coagulans)、短芽胞杆菌(B.brevis)、坚强芽孢杆菌(B.firmus)、B.alkaophius、地衣芽孢杆菌(B.licheniformis)、克劳氏芽孢杆菌(B.clausii)、嗜热脂肪芽孢杆菌(B.stearothermophilus)、耐盐芽孢杆菌(B. halodurans)和解淀粉芽孢杆菌(B.amyloliquefaciens))。在一些实施方案中，宿主细胞是工业芽孢杆菌菌株，包括但不限于枯草芽孢杆菌、短小芽胞杆菌、地衣芽孢杆菌、巨大芽胞杆菌、克劳氏芽孢杆菌、嗜热脂肪芽杆菌或解淀粉芽孢杆菌。在一些实施方案中，芽孢杆菌宿主细胞是枯草芽孢杆菌、地衣芽孢杆菌、巨大芽胞杆菌、嗜热脂肪芽孢杆菌和/或解淀粉芽孢杆菌。在一些实施方案中，细菌宿主细胞是梭菌属的物种(例如丙酮丁醇梭菌(C. acetobutylicum)、破伤风梭菌(C. tetani)E88、C.lituseburense、C. saccharobutylicum、产气荚膜梭菌(C.perfringens)和C.beijerinckii)。在一些实施方案中，细菌宿主细胞是棒杆菌属的物种(例如谷氨酸棒杆菌(C. glutamicum)和嗜乙酰棒杆菌(C.acetoacidophilum))。在一些实施方案中，细菌宿主细胞是埃希氏菌属的物种(例如大肠杆菌(E.coli))。在一些实施方案中，细菌宿主细胞是欧文氏菌属的物种(例如噬夏孢欧文氏菌(E.uredovora)、胡萝卜欧文氏菌(E.carotovora)、菠萝欧文氏菌(E.ananas)、草生欧文氏菌 (E.herbicola)、斑点欧文氏菌(E.punctata)和土欧文氏菌(E.terreus))。在一些实施方案中，细菌宿主细胞是泛菌属的物种(例如柠檬泛菌(P.citrea)和成团泛菌(P.agglomerans))。在一些实施方案中，细菌宿主细胞是假单胞菌属的物种(例如恶臭假单胞菌(P.putida)、铜绿假单胞菌(P.aeruginosa)、P. mevalonii和假单胞菌属的种(P.sp.)D-0110)。在一些实施方案中，细菌宿主细胞是链球菌属的物种(例如S.equisimiles、酿脓链球菌(S.pyogenes)和乳链球菌(S.uberis))。在一些实施方案中，细菌宿主细胞是链霉菌属的物种(例如生二素链霉菌(S.ambofaciens)、不发色链霉菌(S.achromogenes)、阿维链霉菌(S.avermitilis)、天蓝色链霉菌、金黄色链霉菌(S.aureofaciens)、金色链霉菌(S.aureus)、杀真菌素链霉菌(S.fungicidicus)、灰色链霉菌(S.griseus) 和变铅青链霉菌(S.lividans))。在一些实施方案中，细菌宿主细胞是发酵单胞菌属的物种(例如运动发酵单胞菌(Z.mobilis)和Z.lipolytica)。

[0175] 示例性宿主细胞是大肠杆菌W3110。表达载体通过将编码改进的PGA 的多核苷酸可操作地连接到质粒pCK110900构建，所述编码改进的PGA 的多核苷酸可操作地连接至在lacI阻遏子的控制下的lac启动子。表达载体还包含P15a复制起点和氯霉素耐受基因。大肠

杆菌W3110中包含主题多核苷酸的细胞通过对细胞进行氯霉素选择来分离。

[0176] 可以在本发明中使用的许多原核和真核菌株从多个培养物保藏中心可被公众容易地获得,诸如美国典型培养物保藏中心(ATCC)、德国微生物保藏中心(DSM)、真菌菌种保藏中心(CBS)和农业研究机构培养物保藏中心 (Agricultural Research Service Patent Culture Collection,Northern Regional Research Center,NRRL)。

[0177] 在一些实施方案中,宿主细胞被遗传修饰以具有改进蛋白分泌、蛋白稳定性的特征和/或蛋白表达和/或分泌所需的其他特性。遗传修饰可以通过遗传工程技术和/或经典微生物技术(例如化学或UV诱变和随后的选择) 实现。事实上,在一些实施方案中,重组修饰和经典选择技术的组合被用于产生宿主细胞。使用重组技术,核酸分子可以以在宿主细胞中和/或在培养基中导致PGA变体的收率增加的方式被导入、缺失、抑制或修饰。例如,敲除Alp1功能产生蛋白酶缺陷型细胞,而敲除pyr5功能产生具有嘧啶缺陷表型的细胞。在一个遗传工程方法中,同源重组被用于通过体内特异性靶向基因引起靶向性基因修饰,来抑制所编码的蛋白的表达。在可选的方法中,可以使用siRNA、反义和/或核酶技术抑制基因表达。本领域已知多种方法用于降低蛋白在细胞中的表达,包括但不限于缺失编码该蛋白的全部或部分基因和定点诱变(site-specific mutagenesis)以破坏基因产物的表达或活性。(参见例如Chaveroche等人,Nucl.Acids Res.,28:22e9[2000]7; Cho等人,Molec.Plant Microbe Interact.,19:7-15[2006];Maruyama和 Kitamoto,Biotechnol Lett.,30:1811-1817[2008];Takahashi等人, Mol.Gen.Genom.,272:344-352[2004];以及You等人., Arch.Micriobiol.,191:615-622[2009],其全部通过引用并入本文)。也可以使用随机诱变,随后是筛选期望的突变(参见例如Combier等人,FEMS Microbiol.Lett.,220:141-8[2003];和Firon等人,Eukary.Cell 2:247-55(2003),其均通过引用并入)。

[0178] 将载体或DNA构建体导入到宿主细胞可以使用本领域已知的任何合适的方法实现,包括但不限于磷酸钙转染、DEAE-右旋糖酐介导的转染、PEG介导的转化、电穿孔或本领域已知的其他常用技术。

[0179] 在一些实施方案中,本发明的工程化宿主细胞(即“重组宿主细胞”)在被修改为适于激活启动子、筛选转化体或扩增PGA多核苷酸的常规营养培养基中培养。培养条件,诸如温度、pH等,是先前对所选择的用于表达的宿主细胞使用的那些,并且是本领域技术人员熟知的。如所述的,许多标准参考文献和教科书可用于培养和产生许多细胞,包括细菌、植物、动物(特别是哺乳动物)和古细菌来源的细胞。

[0180] 在一些实施方案中,表达本发明的变体PGA多肽的细胞在分批或连续发酵条件下生长。经典的“分批发酵”是封闭的系统,其中培养基的组成在发酵开始时设置并且在发酵期间不进行人工调整。分批系统的变化是也可以在本发明中使用的“补料-分批发酵”。在此变化中,随着发酵进展增量地增加底物。当分解代谢物阻遏可能抑制细胞的新陈代谢时以及当期望在培养基中具有有限量的底物时,可使用补料-分批系统。分批发酵和补料- 分批发酵在本领域中是常见的和熟知的。“连续发酵”是一个开放系统,其中向生物反应器连续添加限定的发酵培养基,并且同时取出等量的条件培养基用于加工。连续发酵通常维持培养物在恒定的高密度,其中细胞主要处于对数期生长。连续发酵系统力求维持稳定状态的生长条件。调节用于连续发酵过程的营养物和生长因子的方法以及用于使产物形成的速率最大化的技术是工业微生物学领域熟知的。

[0181] 在本发明的一些实施方案中，可以使用无细胞转录/翻译系统产生变体 PGA。若干系统是商购可得的并且方法是本领域技术人员熟知的。

[0182] 本发明提供了制备变体PGA多肽或其生物活性片段的方法。在一些实施方案中，所述方法包括：提供用以下多核苷酸转化的宿主细胞，所述多核苷酸编码包含与SEQ ID NO：4、8、14、300、1036、1194、1262和/ 或1288具有至少约70％(或至少约75％、至少约80％、至少约85％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％或至少约 99％)序列同一性的氨基酸序列并包含如本文提供的至少一个突变；在宿主细胞表达编码的变体PGA多肽的条件下在培养基中培养转化的宿主细胞；和任选地回收或分离所表达的变体PGA多肽，和/或回收或分离包含所表达的变体PGA多肽的培养基。在一些实施方案中，方法还提供任选地在表达所编码的PGA多肽后裂解转化的宿主细胞，和任选地从细胞裂解物回收和/或分离所表达的变体PGA多肽。本发明还提供了制备变体PGA多肽的方法，所述方法包括在适于产生变体PGA多肽的条件下培养用变体 PGA多肽转化的宿主细胞，和回收变体PGA多肽。通常，PGA多肽的回收或分离是从宿主细胞培养基、宿主细胞或两者的回收或分离，使用本领域熟知的蛋白回收技术，包括本文描述的那些蛋白回收技术。在一些实施方案中，宿主细胞通过离心收获、用物理或化学手段破坏，且所得粗提取物被保留用于进一步纯化。在蛋白表达中使用的微生物细胞可以通过任何常规方法破坏，包括但不限于冻融循环、超声、机械破坏和/或使用细胞裂解剂，以及本领域技术人员熟知的许多其他合适的方法。

[0183] 可以使用用于蛋白纯化的熟知技术中的任何一种或更多种，将在宿主细胞中表达的工程化PGA酶从细胞和/或培养基中回收，用于蛋白纯化的熟知技术除了其他以外包括，溶菌酶处理、超声、过滤、盐析、超离心和层析。用于裂解和从细菌诸如大肠杆菌高效提取蛋白的合适的溶液是以商标名CelLytic B™(Sigma-Aldrich)商购可得的。因此，在一些实施方案中，回收/分离所得多肽，并任选地通过本领域已知的多种方法的任一种纯化。例如，在一些实施方案中，通过常规程序从营养培养基分离多肽，包括但不限于离心、过滤、提取、喷雾干燥、蒸发、层析(例如离子交换、亲和性、疏水相互作用、聚焦层析和尺寸排阻)或沉淀。在一些实施方案中，按需要使用蛋白重折叠步骤以完成成熟蛋白的构型。另外，在一些实施方案中，在最后纯化步骤中使用高效液层析(HPLC)。例如，在一些实施方案中，本发明可以使用本领域已知的方法(参见例如Parry等人，Biochem.J.，353:117 [2001]；和Hong等人，Appl.Microbiol.Biotechnol.，73:1331[2007]，两者通过引用并入本文)。事实上，本发明可以使用本领域已知的任何合适的纯化方法。

[0184] 用于分离PGA多肽的层析技术包括但不限于反相色层析、高效液相层析、离子交换层析、凝胶电泳和亲和层析。用于纯化特定酶的条件将部分地取决于因素诸如净电荷、疏水性、亲水性、分子量、分子形状等，是本领域技术人员已知的。

[0185] 在一些实施方案中，亲和技术可以用于分离改进的PGA酶。对于亲和层析纯化，可以使用特异性结合PGA多肽的任何抗体。为了产生抗体，可以通过注射PGA免疫多种宿主动物，包括但不限于兔、小鼠、大鼠等。 PGA多肽可以通过侧链官能基团或附接至侧链官能基团的接头的方式附接至合适的运载体(carrier)诸如BSA上。取决于宿主物种，多种佐剂可用来增加免疫反应，包括但不限于弗氏佐剂(完全和不完全)、矿物凝胶例如氢氧化铝、表面活性物质诸如溶血卵磷脂、pluronic多元醇、聚阴离子、肽、油乳剂、钥孔虫戚血兰素、二硝基酚和可能有用的人类佐剂例如BCG(卡介苗)和短棒杆菌(Corynebacterium parvum)。

[0186] 在一些实施方案中,PGA变体以表达酶的细胞的形式、作为粗提取物、或作为分离的或纯化的制品制备和使用。在一些实施方案中,PGA变体制备为冻干物、以粉末形式(例如丙酮粉末)制备或制备为酶溶液。在一些实施方案中,PGA变体呈大体上纯的制品的形式。

[0187] 在一些实施方案中,PGA多肽被附接至任何合适的固体基质上。固体基质包括但不限于固相、表面和/或膜。固体支持物包括但不限于,有机聚合物诸如聚苯乙烯、聚乙烯、聚丙烯、聚氟乙烯、聚氧乙烯(polyethyleneoxy) 和聚丙烯酰胺以及它们的共聚物和接枝物。固体支持物还可以是无机的,诸如玻璃、二氧化硅(silica)、可控孔隙玻璃(CPG)、反相二氧化硅或金属诸如金或铂。基质的形状可以呈珠、球、颗粒(particle)、颗粒剂(granule)、凝胶、膜或表面的形式。表面可以是平面的、大体上平面的或非平面的。固体支持物可以是多孔的或无孔的,并且可以具有溶胀或非溶胀特性。固体支持物可以被配置为孔、凹陷或其他容器(container)、器皿(vessel)、特征或位置的形式。多于一个(a plurality of)支持物可以被配置在阵列的多个位置上,所述多个位置是试剂的自动递送或通过检测方法和/或仪器可寻址的。

[0188] 在一些实施方案中,使用免疫学方法纯化PGA变体。在一种方法中,将使用常规方法针对变体PGA多肽(例如针对包含SEQ ID NO:4、8、14、 300、1036、1194、1262、和/或1288的任一个的多肽、和/或其免疫原性片段)产生的抗体固定在珠上,在其中变体PGA被结合并沉淀的条件下与细胞培养基混合。在一个相关方法中,可以使用免疫层析。

[0189] 在一些实施方案中,变体PGA被表达为包括非酶部分的融合蛋白。在一些实施方案中,变体PGA序列被融合至纯化辅助结构域(purification facilitating domain)。如本文使用的,“纯化辅助结构域”指介导与其融合的多肽的纯化的结构域。合适的纯化结构域包括但不限于,金属螯合肽、允许在固定的金属上纯化的组氨酸-色氨酸模块、结合谷胱甘肽的序列(例如 GST)、血球凝集素(HA)标签(对应于源自流感血球凝集素蛋白的表位;参见例如Wilson等人,Cell 37:767[1984])、麦芽糖结合蛋白序列、在FLAGS 延伸/亲和纯化系统(例如,可从Immunex Corp获得的系统)中使用的FLAG 表位,等。预期用于在本文描述的组合物和方法中使用的一种表达载体提供包含本发明的多肽的融合蛋白的表达,本发明的多肽被融合至多组氨酸区,由肠激酶裂解位点隔开。组氨酸残基有助于在IMIAC(固定金属离子亲和层析;参见例如Porath等人,Prot.J.Exp.Purif.,3:263-281[1992])上纯化,同时肠激酶裂解位点提供用于从融合蛋白分离变体PGA多肽的手段。 pGEX载体(Promega)也可以被用于以与谷胱甘肽S-转移酶融合的融合蛋白表达外源多肽。一般而言,此类融合蛋白是可溶性的并且可以被容易地通过吸附至配体-琼脂糖珠(例如在GST-融合体的情况下是谷胱甘肽-琼脂糖) 从裂解的细胞中纯化,随后在游离配体的存在下洗脱。

[0190] 实验

[0191] 本公开内容的多种特征和实施方案在以下代表性实施例中进行了说明,这些实施例旨在说明而非限制。

[0192] 在以下实验公开内容中,使用以下缩写:ppm(百万分率);M(摩尔/ 升);mM(毫摩/升),uM和μM(微摩/升);nM(纳摩/升);mol(摩尔);gm 和g(克);mg(毫克);ug和μg(微克);L和l(升);ml和mL(毫升);cm(厘米);mm(毫米);um和μm(微米);sec.(秒);min(s)(分钟);h(s)和hr(s)(小时);U(单位);MW(分子量);rpm(转每分);℃(摄氏度);RT(室温); CDS(编码序列);DNA(脱氧核糖核酸);RNA(核糖核酸);TB(Terrific肉汤;12g/L细菌-胰蛋白胨,24g/

L酵母提取物,4mL/L甘油,65mM磷酸钾, pH 7.0,1mM $MgSO_4$);CAM(氯霉素);PMBS(多粘菌素B硫酸盐);IPTG (异丙基硫代半乳糖苷);TFA(三氟乙酸);HPLC(高效液相层析);FIOPC(相对于阳性对照的改进倍数);HTP(高通量);LB(Luria肉汤);Codexis (Codexis,Inc.,Redwood City,CA);Sigma-Aldrich(Sigma-Aldrich,St.Louis, MO);Millipore(Millipore,Corp.,Billerica MA);Difco(Difco Laboratories, BD Diagnostic Systems,Detroit,MI);Daicel(Daicel,West Chester,PA); Genetix(Genetix USA,Inc.,Beaverton,OR);Molecular Devices(Molecular Devices,LLC,Sunnyvale,CA);Applied Biosystems(Applied Biosystems,part of Life Technologies,Corp.,Grand Island,NY),Agilent(Agilent Technologies, Inc.,Santa Clara,CA);Thermo Scientific(Thermo Fisher Scientific,Waltham, MA的一部分);(Infors;Infors-HT,Bottmingen/Basel,Switzerland);Corning (Corning,Inc.,Palo Alto,CA);和Bio-Rad(Bio-Rad Laboratories,Hercules, CA);Microfluidics(Microfluidics Corp.,Newton,MA,United States of America)。

[0193] 实施例1

[0194] 包含重组PGA基因的大肠杆菌表达宿主

[0195] 用于产生本发明的变体的初始PGA酶从Codex®Acylase Panel (Codexis)或共同拥有的美国临时专利申请系列号62/158,118中公开的变体获得。PGA组平板包括与野生型Kluyvera citrophila PGA相比具有改进的特性的工程化PGA多肽的集合。野生型PGA基因是由通过54aa间隔区连接的α亚基(23.8KDa)和β亚基(62.2KDa)组成的异二聚体。由于存在间隔区,需要自动加工步骤(autoprocessing step)以形成活性蛋白。将野生型基因修饰以消除间隔区,从而消除自动加工步骤。Codex®Acylase Panel (Codexis)包含缺乏间隔区的PGA变体(参见例如,美国专利申请公布 2010/0143968 A1)。将PGA编码基因克隆到表达载体pCK110900中(参见,美国专利申请公布第2006/0195947号的图3),可操作地连接至在lacI阻遏子控制下的lac启动子。表达载体还包含P15a复制起点和氯霉素抗性基因。使用本领域已知的标准方法将所得质粒转化到大肠杆菌W3110中。如本领域已知的,通过对细胞进行氯霉素选择来分离转化体(参见例如,美国专利号8,383,346和WO2010/144103)。

[0196] 实施例2

[0197] 制备含有HTP PGA的湿细胞沉淀物

[0198] 将来自单克隆菌落的包含重组PGA编码基因的大肠杆菌细胞接种到 96孔浅孔微量滴定板的孔中的包含1%葡萄糖和30μg/mL氯霉素的180μl LB中。将板用可透O_2的密封物密封,并使培养物在30℃,200rpm和85%湿度生长过夜。然后,将10μl每种细胞培养物转移到包含390mL TB和30μg/mL CAM的96孔深孔板的孔中。将深孔板用可透O_2的密封物密封,并在30℃,250rpm和85%湿度孵育,直到达到OD600 0.6-0.8。然后用1mM 终浓度的IPTG诱导细胞培养物,并在最初使用的相同条件下孵育过夜。然后,使用在4000rpm持续10min的离心使细胞沉淀。弃去上清液并在裂解前将沉淀物在-80℃冷冻。

[0199] 实施例3

[0200] 制备包含HTP PGA的细胞裂解物

[0201] 首先,向如实施例2中描述的产生的每个孔中的细胞团(cell paste)添加包含10mM Tris-HCl缓冲液(pH7.5)、1mg/mL溶菌酶和0.5mg/mL PMBS 的200μl裂解缓冲液。使细

胞伴随在台式振荡器(bench top shaker)上震荡在室温裂解2小时。然后将板在4000rpm和4℃离心15min。澄清的上清液被用于生物催化反应以确定它们的活性水平。

[0202] 实施例4

[0203] 从摇瓶(SF)培养物制备冻干的裂解物

[0204] 将选择的如以上描述生长的HTP培养物铺板于具有1％葡萄糖和30 μg/ml CAM的LB琼脂板上并在37℃生长过夜。将来自每种培养物的单个菌落转移至具有1％葡萄糖和30μg/ml CAM的6ml LB中。将培养物在 30℃,250rpm生长18h,并以约1:50传代培养至含30μg/ml CAM的250 ml TB中,至最终OD_{600}为0.05。使培养物在30℃,250rpm生长约195 分钟,至OD_{600}在0.6-0.8之间,并用1mM IPTG诱导。然后使培养物在 30℃,250rpm生长20h。4000rpm×20min离心培养物。弃去上清液,并在30ml的20mM TRIS-HCl(pH7.5)中重悬沉淀物。使细胞沉淀(4000 rpm×20min)并在-80℃冷冻120分钟。将冷冻的沉淀物重悬于30ml的 20mM TRIS-HCl pH7.5中,并使用Microfluidizer系统((Microfluidics)以 18,000psi裂解。使裂解物沉淀(10,000rpm×60min)并将上清液冷冻并冻干以产生摇瓶(SF)酶。

[0205] 实施例5

[0206] 在B29位置三-保护的胰岛素相比于SEQ ID NO:8在脱酰化方面的改进

[0207] 在筛选美国临时专利申请系列号62/158,118中公开的变体后,选择了 SEQ ID NO:8作为亲本酶用于产生B29脱酰化的产物。使用良好建立的技术(例如,饱和诱变、先前鉴定的有益突变的重组)产生工程化基因的文库。如实施例2中描述,在HTP中产生由每种基因编码的多肽,并如实施例3 中描述产生可溶性裂解物。在30℃在200μL反应物中筛选每种变体3小时,所述反应物包含5g/L A1/B1/B29三-乙酸苯酯胰岛素、200mM Tris 缓冲液pH＝8.3和40μL粗裂解物。将96孔板热密封并在Thermotron®振荡器中以100rpm孵育。用200μl乙腈或二甲基乙酰胺猝灭反应物,并使用台式振荡器混合5分钟。然后将板以4000rpm离心5分钟,将其稀释2倍至水中,并加载到HPLC中用于分析。

[0208] 相对于SEQ ID NO:8的活性(活性FIOP)被计算为由变体形成的产物的转化百分比相对于由SEQ ID NO:8产生的转化百分比并显示在表5.1中。通过将如HPLC分析所观察到的产物峰面积除以底物、产物和杂质/副产物峰的面积总和来计算转化百分比。

[0209]

表5.1 变体相对于SEQ ID NO: 8的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 8)	活性FIOP [1] (相对于SEQ ID NO: 8)
9/10	G71C;D74G	+
11/12	Y27T;G71H;D74G;D484N;Q547K;Y584F;M697L	+
13/14	Y27T;V28A;G71H;D74G;Q547K	+
39/40	L253Y	+
41/42	Y27T	+
29/30	S386G;K390S	+
17/18	L387K;K390S	+
81/82	A474C	+
45/46	S386A;K390S	+
69/70	S386H	+
15/16	S704N	+
27/28	F254N	+
63/64	A470C	+
31/32	A373R	+
23/24	T352K	+
59/60	L253H	+
37/38	A616Y	+
55/56	S372L	+
19/20	A451K	+
77/78	I708L	+
43/44	F256Y	+
35/36	L387R;K390S	+
21/22	D623Q	+
47/48	A467W	+
53/54	D623K	+
73/74	N348H	+
51/52	F254K	+
49/50	S706T	+
33/34	S374T	+
83/84	Q380M	+
75/76	T129W	+
67/68	Q380Y;N457L	+
57/58	A470E	+
61/62	N457A	+
25/26	D623A	+
65/66	N457Q	+
79/80	F256S	+
71/72	A373K	+

[1]增加的活性水平相对于参考多肽SEQ ID NO: 8确定并如下定义："+"活性增加>1倍但小于2.0倍；"++"活性增加>2.0倍但小于3倍；"+++"活性增加>3倍但小于10倍；"++++">10倍。

[0210] 实施例6

[0211] 在B29位置三-保护的胰岛素相比于SEQ ID NO:14在脱酰化方面的改进

[0212] 在实施例5中描述的筛选变体后，选择了SEQ ID NO:14作为亲本酶。使用良好建立的技术(例如，饱和诱变、先前鉴定的有益突变的重组)产生工程化基因的文库。如实施例2中描述，在HTP中产生由每种基因编码的多肽，并如实施例3中描述产生可溶性裂解物。

[0213] 在30℃在200μL反应物中筛选每种变体24小时，所述反应物包含10 g/L A1/B1/B29三-乙酸苯酯胰岛素、200mM Tris缓冲液pH＝8.3和20μL 粗裂解物。将96孔板热密封并在thermotron振荡器中以100rpm孵育。用 200μl乙腈或二甲基乙酰胺猝灭反应物，并使用台式振荡器混合5分钟。然后将板以4000rpm离心5分钟，将其稀释2倍至水中，并加载到HPLC中用于分析。

[0214] 相对于SEQ ID NO:14的活性(活性FIOP)被计算为由变体形成的产物的转化百分比相对于由SEQ ID NO:14产生的转化百分比并显示在表6.1 中。通过将如HPLC分析所观察到的产物峰面积除以底物、产物和杂质/ 副产物峰的面积总和来计算转化百分比。

[0215]

表6.1 变体相对于SEQ ID NO: 14的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 14)	活性FIOP [1] (相对于SEQ ID NO: 14)
419/420	N185D	+
433/434	T560G	+
443/444	A517K	+
415/416	G415H	+
441/442	G444K	+
417/418	G444T	+
421/422	R748D	+
439/440	G444R	+
447/448	T443D	+
423/424	G444S	+
427/428	S492G/G493C	+
431/432	N9D	+
437/438	G444Q	+
429/430	R748S	+
435/436	K723D	+
445/446	R748E	+

[0216]

表6.1 变体相对于SEQ ID NO: 14的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 14)	活性FIOP [1] (相对于SEQ ID NO: 14)
425/426	G444L	+
165/166	N9S;K103E;W119Y;T131D;Q233K;T269R;K304Q;G444N;N494D;S646D	+
177/178	K103E;G444N;N494D;S646D	+
119/120	K103E;W119Y;T131D;Q233K;K273A;K304Q;A324T;Q432E;G444N;N494D;S646D	+
211/212	K103E;W119Y;Q233K;E312A;S646D	+
207/208	N9S;K103E;W119Y;T131D;K304Q;A324T;Q432E;G444N;N494D;S646D	+
149/150	W119Y;Q233K;K304Q;E312A;A324T;Q432E;G444N;S646D	+
175/176	K103E;T131D;Q233K;K304Q;A324T;G444N;S646D	+
85/86	K103E;W119Y;N494D;S646D;K661Q	+
111/112	N9S;W119Y;T131A;Q233K;N494D;S646D	+
125/126	K103E;W119Y;Q233K;K304Q;S646D	+
173/174	N9S;K103E;W119Y;Q233K;S646D	+
215/216	K103E;Q233K;K273A;E312A;Q432E;G444N;S646D	+
133/134	N9S;W119Y;Q233K;N494D;S646D	+
155/156	W119Y;K273A;A324T;G444N;N494D;S646D	+
181/182	N9S;W119Y;T131A;Q233K;K304Q;G444N;S646D	+
121/122	K103E;W119Y;T131A;Q233K;K304Q;E312A;Q432E;N494D;S646D;K661Q	+
163/164	W119Y;Q233K;K304Q;S646D	+
115/116	N9S;W119Y;K304Q;G444N;N494D;S646D	+
109/110	N9S;W119Y;Q233K;T269R;K273A;K304Q;E312A;Q432E;G444N;S646D	+
123/124	W119Y;E312A;G444N;S646D	+
113/114	N9S;W119Y;T131A;E312A;G444N;S646D	+
143/144	K103E;E312A;G444N;S646D	+
103/104	K103E;W119Y;S646D	+
135/136	W119Y;Q233K;S646D	+
129/130	K103E;W119Y;Q233K	+
167/168	W119Y;T131A;G444N;S646D;K661Q	+
157/158	W119Y;Q432E;G444N;S646D	+
189/190	K103E;T269R;K273A;G444N;S646D	+
159/160	N9S;T131A;Q233K;T269R;K273A;E312A;Q432E;G444N;S646D	+
183/184	K103E;K273A;K304Q;A324T;G444N;N494D	+
105/106	E312A;G444N;S646D	+
127/128	N9S;Q233K;N494D;S646D	+
221/222	N9S;Q233K;K304Q;N494D;S646D	+
191/192	T131D;K273A;Q432E;G444N;N494D;S646D	+
219/220	N9S;Q432E;G444N;N494D;S646D	+
171/172	N9S;Q233K;K273A;K304Q;N494D;S646D	+
169/170	K103E;Q233K;K273A;E312A;A324T;Q432E;G444N;S646D	+

[0217]

表6.1 变体相对于SEQ ID NO: 14的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 14)	活性FIOP [1] (相对于SEQ ID NO: 14)
97/98	N9S;K103E;Q233K;N494D	+
201/202	W119Y;T269R;K273A;E312A;Q432E;G444N;S646D	+
187/188	T131A;Q233K;K304Q;G444N	+
101/102	N9S;K103E;W119Y;N494D	+
197/198	K103E;T131D;T269R;K273A;G444N;S646D	+
91/92	N9S;W119Y;T131D;Q233K;N494D;K661Q	+
205/206	K103E;W119Y;T131A;T269R;E312A;N494D;S646D	+
223/224	N9S;K103E;W119Y;T131A;Q233K;E312A;A324T;Q432E;G444N;N494D	+
213/214	N9S;W119Y;T131A;Q432E;G444N;S646D	+
145/146	G444N;N494D	+
95/96	W119Y;T131A;N494D	+
93/94	N9S;K103E;K304Q;Q432E;G444N;S646D	+
185/186	K304Q;E312A;G444N;S646D	+
147/148	W119Y;T131A;S646D	+
139/140	G444N;S646D	+
89/90	K273A;E312A;G444N;S646D	+
199/200	W119Y;Q233K;E312A	+
137/138	T131A;Q233K;K273A;S646D	+
153/154	N9S;E312A;G444N;S646D	+
161/162	N9S;W119Y;T131D;Q233K	+
195/196	E312A;S646D	+
193/194	N9S;K103E;T269R;K304Q;A324T;N494D;S646D	+
107/108	W119Y;T131D;K304Q;Q432E;G444N;S646D;K661Q	+
131/132	W119Y;T131D;G444N	+
209/210	T131A;Q233K;Q432E;S646D	+
99/100	N9S;Q233K;E312A;Q432E;S646D	+
141/142	N9S;K103E;Q233K;E312A;S646D	+
117/118	K103E;W119Y;Q233K;K273A;Q432E	+
179/180	Q233K;T269R;K304Q;E312A;A324T;Q432E;G444N;S646D	+
217/218	N9S;K103E;W119Y;T131A;A324T;Q432E;N494D;S646D	+
203/204	N9S;N494D;S646D	+
87/88	N9S;K103E;Q233K;Q432E;S646D	+
151/152	W119Y;T269R;K273A;E312A;A324T;N494D;S646D	+
237/238	L253Y;F256Y;T352K;A451K;	+++
299/300	L253Y;T352K;A373R;S374T;A451K;A616Y	+++
271/272	L253H;F256Y;T352K;A373R;S374T;A616Y	++
321/322	F256Y;S374T;Q380L;A451K;A616Y	++
307/308	Q380L;A451K;A616Y	++
323/324	L253H;T352K;A451K;A616Y	++
243/244	L253Y;F256Y;T352K;S374T;Q380L;A451K	++
293/294	F254N;A373R;S374T;A451K	++
273/274	T352K;A373R;S374T;A451K	++
337/338	L253Y;T352K;A373R;S374T;A451K	++

[0218]

表6.1 变体相对于SEQ ID NO: 14的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 14)	活性FIOP [1] (相对于SEQ ID NO: 14)
227/228	L253H;A451K;N457S	++
325/326	L253Y;F256Y;T352K;S374T;A451K	++
333/334	L253H;T352K;S374T;A616Y	++
363/364	L253Y;A373R;S374T;A451K;A616Y	++
275/276	T352K;S374T;A451K	++
331/332	L253Y;F254N;T352K;S374T;Q380L	++
343/344	S374T;A451K;A616Y	++
295/296	A451K;A616Y	++
351/352	T352K;S374T;Q380L;A451K;A616Y	++
367/368	T352K;Q380L;A451K;A616Y	++
349/350	L253Y;T352K;S374T;A616Y	++
401/402	L253H;F256Y;T352K;Q380L;A451K;A616Y	++
289/290	T352K;A451K;A616Y	++
381/382	L253Y;F256Y;A373R	++
383/384	L253H;F256Y;S374T;A616Y	++
253/254	S374T;Q380L;A451K	++
335/336	S374T;Q380L;A451K;D623N	++
355/356	L253Y;F256Y;T352K;Q380L	++
405/406	T352K;S374T;A616Y	++
225/226	F254N;T352K;Q380L;A451K	++
231/232	A373R;S374T;A451K	++
379/380	F254N;A616Y	++
393/394	L253H;F256Y;Q380L;A451K;A616Y	++
399/400	F254N;A373R;S374T;Q380L;A451K	++
411/412	L253H;F256Y;T352K;S374T;Q380L;A451K	++
263/264	F254N;F256Y;S374T;A451K	++
341/342	S374T;A616Y	++
283/284	L253H;S374T;A451K	++
375/376	F256Y;S374T;A616Y	++
339/340	F254N;T352K;A451K	++
407/408	L253Y;F256Y;Q380L	++
373/374	F256Y;T352K;S374T	++
409/410	L253H;S374T;A451K;D623Q	++
285/286	L253Y;T352K;D623N	++
279/280	Q380L;A451K;D623A	++
269/270	F254N;A451K	++
257/258	T352K;Q380L;A451K;D623Q	++
403/404	T352K;A373R;A451K	++
359/360	T352K;S374T	++
345/346	Q380L;A616Y	++
305/306	L253H;F256Y	++
245/246	S374T;A451K	++
377/378	L253Y;F256Y	++
385/386	T352K;S374T;Q380L	++
365/366	L253H;A373R;A451K	++
287/288	L253Y;Q380L	++

[0219]

表6.1 变体相对于SEQ ID NO: 14的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 14)	活性FIOP [1] (相对于SEQ ID NO: 14)
291/292	Q380L;A451K	++
353/354	F254N;F256Y;Q380L;A451K;A616Y	++
313/314	F254N;S374T	++
309/310	F254N;S374T;Q380L	++
239/240	T352K;Q380L;A616Y	++
361/362	T352K;Q380L;A451K	++
387/388	L253H;F256Y;S374T;A451K	++
249/250	A373R;S374T	++
371/372	L253H;A451K	++
255/256	L253Y;F256Y;Q380L;A451K;H546Y;A616Y	+
327/328	T352K;A616Y	+
267/268	A616Y	+
347/348	F256Y;T352K;Q380L;A451K;A616Y	+
329/330	S374T;Q380L	+
311/312	S374T;D623Q	+
281/282	F254N;F256Y;A373R;S374T;Q380L;A451K	+
235/236	T352K;A451K	+
319/320	T352K;A373R;Q380L;A451K;A616Y	+
229/230	D623A	+
413/414	Q380L;A616Y;D623Q	+
265/266	A373R;A451K	+
251/252	S374T	+
395/396	F254N;F256Y;T352K;A451K	+
397/398	A451K;D623N	+
247/248	F254N;F256Y;S374T	+
233/234	L253H	+
317/318	F254N;A255E;T352K	+
369/370	A28S;S374T;Q380L	+
301/302	F254N;T352K;Q380L	+
261/262	F254N;Q380L	+
241/242	A451K	+
277/278	T352K	+
259/260	T352K;Q380L	+
389/390	D623N	+
391/392	D623Q	+
303/304	T352K;A373R;Q380L;A451K	+
297/298	Q380L	+
357/358	F254N;F256Y;A451K	+
315/316	A373R;A616Y	+
451/452	N457Q;A470C	+
455/456	T129W;F254K;N348H;N457Q;S704N	+
467/468	N348H;S372L;A470E;D623K;I708L	+
457/458	N457Q;S704N;I708L	+
483/484	T129W;D623K	+
453/454	N457Q	+
485/486	T129W;N348H	+

[0220]

表6.1 变体相对于SEQ ID NO: 14的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 14)	活性FIOP [1] (相对于SEQ ID NO: 14)
461/462	N348H	+
449/450	N348H;S704N;I708L	+
459/460	T129W;Q380M;A470E	+
469/470	T129W;N457Q;A470C;A474C	+
471/472	T129W;N348H;A470E;D623K;S704N;S706T;I708L	+
475/476	T129W;N348H;A467W;A470C;S704N;I708L	+
477/478	S372L;N457Q;A470E;D623K	+
479/480	D623K	+
481/482	A470C;I708L	+
465/466	T129W;N348H;A470E;D623K;S704N;S706T	+
463/464	T129W;A470E	+
473/474	Q380M;D623K	+

[1]增加的活性水平相对于参考多肽SEQ ID NO: 14确定并如下定义："+"=活性增加大于1倍但小于2.0倍；"++"=活性增加大于2.0倍但小于3倍；"+++"=活性增加大于3倍但小于10倍；"++++"=大于10倍。

[0221] 实施例7

[0222] 在B29位置三-保护的胰岛素相比于SEQ ID NO:300在脱酰化方面的改进

[0223] 在实施例6中描述的筛选变体后,选择了SEQ ID NO:300作为亲本酶。使用良好建立的技术(例如,饱和诱变、先前鉴定的有益突变的重组)产生工程化基因的文库。如实施例2中描述,在HTP中产生由每种基因编码的多肽,并如实施例3中描述产生可溶性裂解物。

[0224] 在30℃在200μL反应物中筛选每种变体24小时,所述反应物包含50 g/L A1/B1/B29三-乙酸苯酯胰岛素、200mM Tris缓冲液pH=8.3和2.5μL 粗裂解物。将96孔板热密封并在thermotron振荡器中以100rpm孵育。用 200μl乙腈或二甲基乙酰胺猝灭反应物,并使用台式振荡器混合5分钟。然后将板以4000rpm离心5分钟,将其稀释2倍至水中,并加载到HPLC中用于分析。

[0225] 相对于SEQ ID NO:300的活性(活性FIOP)被计算为由变体形成的产物的转化百分比相对于由SEQ ID NO:300产生的转化百分比并显示在表 7.1中。通过将如HPLC分析所观察到的产物峰面积除以底物、产物和杂质/副产物峰的面积总和来计算转化百分比。

[0226]

表7.1 变体相对于SEQ ID NO: 300的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 300)	活性FIOP [1] (相对于SEQ ID NO: 300)
837/838	N9D;N185D;G415H;G444T;A517K;T560G;K723D;R748E	+++
813/814	N9D;N185D;G415H;T443D;G444R;A517K;K723D;R748D	+++
775/776	N9D;G415H;T443D;T560G;K723D;R748D	+++
839/840	N9D;G415H;G444L;T560G;K723D;R748E	+++
719/720	N9D;G415H;T443D;G444R;T560G;K723D;R748E	+++
797/798	G415H;T443D;A517K;T560G;K723D	+++
695/696	N9D;G415H;T443D;G444T;A517K;T560G;K723D	+++
767/768	N9D;N185D;G415H;G444S;K723D;R748E	+++
597/598	N9D;N185D;G415H;T443D;G444T;A517K;T560G	+++
561/562	N9D;N185D;G415H;G444L;A517K;T560G;K723D;R748S	+++
827/828	N185D;G415H;T443D;G444R;A517K;T560G	+++
771/772	N185D;G415H;G444T;T560G	+++
773/774	N9D;N185D;T443D;G444K;T560G;K723D	+++
765/766	G415H;T443D;G444T;T560G	+++
769/770	G415H;G444Q;T560G;K723D	+++
791/792	N9D;G415H;A517K;T560G;K723D;R748D	+++
783/784	G415H;A517K;T560G;K723D;R748D	+++
673/674	N185D;G415H;T443D;G444T;A517K	++
545/546	N9D;G415H;G444Q;A517K;T560G;K723D;R748S	++
727/728	N9D;N185D;G415H;T443N;G444K;A517K;R748E	++
599/600	N9D;G415H;G444L;A517K;T560G;K723D;R748D	++
633/634	N9D;G415H;T443D;G444S;A517K;T560G	++
651/652	N9D;G415H;T443D;G444T;A517K;R748S	++
683/684	N9D;G415H;G444T;A517K;T560G;K723D	++
817/818	G415H;G444S;A517K;K723D;R748E	++
671/672	N9D;G415H;G444Q;T560G;K723D	++
557/558	N9D;N185D;G415H;G444K;A517K;K723D;R748E	++
627/628	N185D;G415H;G444R;A517K;T560G;R748D	++
699/700	N9D;G415H;T443D;G444K;A517K;T560G	++
679/680	N9D;G415H;A517K;T560G;K723D;R748E	++
569/570	N9D;T443D;G444K;T560G;K723D;R748D	++
589/590	G415H;G444Q;A517K;T560G;K723D;R748S	++
573/574	G415H;G444S;A517K;T560G;K723D	++
799/800	N9D;N185D;G415H;G444K;A517K;R748D	++
525/526	G415H;T443D;G444T;A517K;R748E	++
639/640	G415H;T443D;A517K;K723D	++
807/808	N9D;N185D;G444Q;A517K;T560G;K723D	++
749/750	N9D;N185D;G444T;A517K;T560G;K723D	++
619/620	N9D;T443D;G444K;A517K;T560G;K723D	++
543/544	N185D;G415H;T560G;K723D	++
779/780	N9D;G415H;G444L;K723D;R748S	++
707/708	N9D;N185D;G415H;G444L;A517K;T560G	++
725/726	N9D;N185D;T443D;G444K;A517K;T560G	++
523/524	N9D;G415H;G444T;T560G;S665R;K723D;R748S	++
713/714	G415H;G444S;T560G;R748E	++
587/588	N9D;N185D;T443D;G444S;K723D	++

[0227]

表7.1 变体相对于SEQ ID NO: 300的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 300)	活性FIOP [1] (相对于SEQ ID NO: 300)
539/540	N9D;N185D;G415H;G444Q;T560G	++
551/552	G415H;G444Q;A517K;T560G;R748S	++
553/554	N9D;G415H;G444T;K723D;R748E	++
741/742	N9D;G415H;G444R;A517K;T560G;R748E	++
595/596	N9D;G415H;T443D;G444R;A517K;R748E	++
655/656	N9D;N185D;G415H;A517K;K723D	++
641/642	N9D;N185D;G444Q;A517K;T560G;R748D	++
751/752	G415H;G444S;T560G	++
491/492	N9D;G415H;T443D;T560G	++
823/824	N9D;G444L;A517K;T560G;K723D;R748E	++
717/718	N9D;N185D;G415H;A517K;T560G	++
611/612	N185D;G415H;K723D;R748S	++
593/594	N9D;G415H;G444T;T560G	++
745/746	N9D;N185D;A517K;T560G;K723D;R748E	++
575/576	G415H;T560G;K723D	++
643/644	N9D;T443D;G444R;T560G;K723D	++
609/610	N9D;T443D;G444T;A517K;T560G	++
515/516	G415H;A517K;T560G;R748E	++
489/490	N9D;G415H;G444K;A517K;T560G;R748S	++
583/584	N9D;G415H;G444S;T560G	++
601/602	N185D;G444R;A517K;T560G;K723D;R748S	++
715/716	N185D;G415H;T560G	++
735/736	N185D;G415H;G444R;A517K;R748S	++
513/514	N9D;G415H;A517K;T560G;R748E	++
511/512	N9D;G61D;G444S	++
617/618	G415H;G444K;K723D	++
733/734	N9D;N185D;G444L;T560G;K723D	++
675/676	G415H;K723E;R748E	++
723/724	N9D;G415H;G444R;T560G	++
665/666	N185D;T443D;G444R;T560G	++
607/608	N9D;G444N;A517K;T560G;K723D;R748E	++
653/654	N9D;N185D;G415H;A517K;R748S	++
647/648	N9D;G444Q;A517K;T560G;K723D	++
509/510	N9D;N185D;G444S;A517K;K723D;R748S	++
691/692	N9D;N185D;G415H;R748E	++
787/788	N9D;G415H;G444Q;A517K	++
549/550	G444L;A517K;T560G;K723D;R748S	++
487/488	N185D;A517K;T560G;K723D	++
677/678	T443D;G444T;K723D	++
681/682	T443D;G444K;A517K;R748E	++
649/650	N9D;N185D;G444L;T560G;R748S	++
785/786	N9D;G415H;G444K	++
531/532	N9D;N185D;G415H;G444Q	++
621/622	G415H;K723D;R748S	++
819/820	N185D;G444L;A517K;T560G	++
795/796	G415H;G444T	++

[0228]

表7.1 变体相对于SEQ ID NO: 300的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 300)	活性FIOP [1] (相对于SEQ ID NO: 300)
625/626	N9D;G415H;T560G	++
687/688	G444S;A517K;K723D;R748E	++
585/586	N9D;G444K;T560G;K723D	++
499/500	N9D;T443D;G444T;A517K;R748S	++
541/542	N9D;N185D;G444Q;A517K;T560G	++
753/754	N9D;N185D;G444S;T560G	++
645/646	N9D;N185D;G444L;A517K;T560G	++
581/582	N185D;G444S;A517K;K723D	++
659/660	N9D;G415H;G444L;A517K	++
535/536	T443D;G444S;T560G	++
801/802	N9D;N185D;G444S;K723D	++
579/580	N9D;T443D;G444K;T560G	++
669/670	G444T;T560G;K723D	++
689/690	G415H;G444L;A517K	++
763/764	N9D;N185D;T443D;G444R;A517K	++
777/778	N185D;T560G;K723D	++
757/758	N9D;T443D;K723D	++
637/638	N9D;T443D;G444S;A517K	++
635/636	N9D;G444L;T560G;K723D	++
503/504	N9D;G415H;G444S;A517K	++
815/816	N185D;T560G;R748E	++
825/826	N9D;Q168R;N185D;A517K;T560G;R748S	++
781/782	G444L;T560G;K723D	++
667/668	N9D;G444T;A517K;T560G	++
743/744	N9D;G444L;T560G;R748E	++
711/712	N9D;T443D;A517K;R748E	++
563/564	G444Q;T560G;K723D	++
603/604	N9D;N185D;G415H	++
821/822	G415H;G444L	++
661/662	N9D;G415H;G444R;A517K	++
623/624	N9D;A517K;T560G;K723D	++
731/732	G415H;R748E	++
663/664	N185D;T560G;R748S	++
537/538	N9D;T560G;K723D;R748S	++
793/794	G444Q;T560G	++
567/568	N9D;G444L;T560G;R748S	+
519/520	N9D;G444S;T560G;R748E	+
693/694	N9D;G415H;R748S	+
521/522	N9D;G444Q;A517K;K723D	+
755/756	N185D;G444R;T560G	+
709/710	N9D;G415H;A517K	+
615/616	N9D;A517K;T560G;R748E	+
547/548	G444T;A517K;T560G	+
729/730	N9D;A517K;T560G;R748D	+
803/804	A517K;T560G;R748E	+
555/556	G444K;A517K;T560G	+

[0229]

表7.1 变体相对于SEQ ID NO: 300的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 300)	活性FIOP [1] (相对于SEQ ID NO: 300)
497/498	N9D;G444K;A517K;T560G	+
831/832	N9D;N185D;G444K	+
591/592	G415H;A517K	+
613/614	G444K;A517K;R748E	+
737/738	N185D;T560G	+
739/740	G444S;K723D	+
493/494	G444S;T560G;R748S	+
835/836	N9D;N185D;R748E	+
829/830	N9D;G444K;R748S	+
809/810	G415H	+
701/702	N9D;G444L;T560G	+
685/686	T443D;A517K	+
705/706	N9D;G444S;K723D	+
501/502	T443D;G444S	+
761/762	N9D;N185D;A517K	+
629/630	A517K;T560G	+
805/806	N9D;A517K;K723D	+
833/834	G444K;R748S	+
505/506	N9D;T443D;G444R;A517K	+
507/508	G444R;A517K;T560G	+
747/748	A517K;K723D	+
571/572	N9D;G444Q;A517K	+
517/518	K723D;R748D	+
533/534	N185D;G444K;K446N;A517K;T560G	+
565/566	N9D;G444L;R748E	+
495/496	T443D;G444K;A517K;T560G;K723D;R748S	+
657/658	N9D;K15M;G444L;K723D	+
631/632	N9D;G444T	+
697/698	N9D;A517K;R748S	+
811/812	N9D;N185D	+
529/530	N9D;K723D	+
605/606	A517K;R748E	+
789/790	K723D	+
527/528	G415H;A517K;T560G	+
559/560	T560G	+
721/722	G444K	+
759/760	G444R;A517K	+
577/578	N9D;A517K	+
703/704	A517K;R748S	+
903/904	K103E;W119Y;N348H;N457Q;N494D	+++
897/898	K103E;W119Y;F256R;N348H;G444N;N494D;S646D	++
877/878	K103E;T129W;F254N;G444N;N457Q;N494D	++
943/944	K103E;W119Y;N348H;N457Q	++
895/896	K103E;W119Y;N457Q	++
855/856	W119Y;T129W;F256R;N348H;N457Q	++
945/946	W119Y;T129W;F254N;N457Q;N494D	++

[0230]

表7.1 变体相对于SEQ ID NO: 300的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 300)	活性FIOP [1] (相对于SEQ ID NO: 300)
935/936	K103E;F256R;N457Q	++
881/882	W119Y;F256R;N348H;N494D	++
939/940	N348H;N457Q	++
911/912	K103E;T129W;F256R;N348H	++
923/924	K103E;T129W;N348H;S646D	++
871/872	W119Y;F254N;N348H;N457Q	++
933/934	K103E;W119Y;F254N;N348H;G444N	++
891/892	W119Y;F256R;N348H	++
929/930	F256R;N348H	++
841/842	K103E;W119Y;N348H	+
879/880	K103E;N348H;N494D	+
893/894	K103E;N348H;G444N	+
937/938	K103E;W119Y;T129W;G444N;N494D	+
941/942	K103E;F256R;G444N	+
915/916	K103E;W119Y;F254N;G444N	+
873/874	F254N;N457Q	+
899/900	K103E;N348H	+
901/902	K103E;W119Y;F254N;N348H	+
865/866	K103E;W119Y;N494D;S646D	+
867/868	K103E;F254N;N348H;N494D	+
949/950	N348H;G444N;S646D	+
917/918	W119Y;F256R;G444N	+
845/846	K103E;W119Y;F256R;N494D	+
885/886	K103E;F254N;N348H	+
927/928	W119Y;T129W;F254N;N348H;N494D	+
883/884	K103E;F254N;S646D	+
849/850	W119Y;N348H;N494D	+
861/862	W119Y;N348H	+
913/914	K103E;F256R;N494D	+
925/926	N348H;N494D	+
859/860	K103E;W119Y;N494D	+
951/952	W119Y;F256R	+
857/858	W119Y;N348H;G444N	+
921/922	K103E;T129W	+
875/876	K103E;G444N	+
931/932	K103E;F256R	+
947/948	K103E;N494D;S646D	+
843/844	N348H;G444N	+
853/854	K103E;N494D	+
847/848	K103E;W119Y	+
909/910	F256R;N494D	+
905/906	K103E;W119Y;T129W;F254N;F256R;N348H;N494D;S646D	+
919/920	K103E;F254N	+
869/870	W119Y;S646D	+
887/888	N348H	+
863/864	W119Y;G444N	+

[0231]

表7.1 变体相对于SEQ ID NO: 300的活性

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 300)	活性FIOP[1] (相对于SEQ ID NO: 300)
851/852	W119Y;T129W;S646D	+
907/908	K103E	+
889/890	K103E;F254N;F256R;N348H;G444N;N494D	+

[1]增加的活性水平相对于参考多肽SEQ ID NO: 300确定并如下定义："+"=活性增加大于1倍但小于2.0倍；"++"=活性增加大于2.0倍但小于3倍；"+++"=活性增加大于3倍但小于10倍；"++++"=大于10倍。

[0232] 实施例8

[0233] 在A1位置三-保护的胰岛素相比于SEQ ID NO:1262或1288在脱酰化方面的改进

[0234] 筛选来自Codex®Acylase Panel(Codexis)的变体和美国临时专利申请系列号62/158,118中公开的变体,鉴定了变体SEQ ID NO:1262和1288 作为在位置A1处从A1/B1/B29三-乙酸苯酯胰岛素去除保护基团的最佳酶。选择包含SEQ ID NO:1262的变体作为文库的亲本酶,靶向针对三-保护胰岛素在A1位置脱酰化的改进的活性和化学选择性。使用良好建立的技术 (例如,饱和诱变、先前鉴定的有益突变的重组)产生工程化基因的文库。如实施例2中描述,在HTP中产生由每种基因编码的多肽,并如实施例3 中描述产生可溶性裂解物。在30℃在200μL反应物中筛选每种变体5小时,所述反应物包含5g/L A1/B1/B29三-乙酸苯酯胰岛素、0.1M Tris-HCl 缓冲液pH=8.0、17g/L苯乙酸甲酯和80μL可溶性裂解物。将96孔板热密封并在Thermotron®振荡器中以100rpm孵育。用300μl乙腈猝灭反应物,并使用台式振荡器混合5分钟。然后将板以4000rpm离心5分钟,并加载到HPLC中用于分析。

[0235] 相对于SEQ ID NO:1262或1288的转化百分比(转化百分比FIOP)被计算为由变体形成的产物的转化百分比相对于由SEQ ID NO:1262或1288 产生的转化百分比并显示在表8.1和8.2中。通过将如HPLC分析所观察到的产物峰面积除以底物、产物和杂质/副产物峰的面积总和来计算转化百分比。

[0236] 相对于SEQ ID NO:1262或1288的选择性百分比(选择性百分比FIOP) 被计算为由变体形成的产物的选择性百分比相对于由SEQ ID NO:1262或 1288产生的选择性百分比并显示在表8.1和8.2中。通过将如HPLC分析所观察到的产物峰面积除以产物和杂质/副产物峰的面积总和来计算选择性百分比。

[0237]

表8.1 脱酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1262)	脱酰化转化百分比FIOP [1]	脱酰化选择性百分比FIOP [2]
953/954	W431R	+++	+++
955/956	F701L	++++++	++++++
957/958	F701A	+++++	++++++
959/960	F24A	+++++	++++++
961/962	F71V	+++	+++
963/964	F71K	+++	++++
965/966	F71E	++++++	++++++
967/968	F24Y;V28A	+++++	+++++
969/970	M697L	+++	+++
971/972	I177T	++++	++++
973/974	W154F;L754P	++	++
975/976	F71C;G74D	++	++
977/978	F701V	++++++	++++++
979/980	F701H	++++++	++++++
981/982	F701C	+++++	++++++
983/984	F701I	++++++	++++++
985/986	F701M	++++	+++++
987/988	M697F	++	++
989/990	M697G	++++	++++
991/992	V28A;Y31F	++++++	++++++
993/994	V28A;Y31L	++++++	++++++
995/996	V28A;Y31V	++++++	++++++
997/998	V28A;Y31T	++++	++++
999/1000	V28A;Y31C	+++++	++++++
1001/1002	V28A;Y31N	++++++	++++++
1003/1004	V28A;Y31M	++++++	++++++
1005/1006	V28A;Y31K	++++++	++++++
1007/1008	F24A;Y31W;V56I;F701W	++++++	++++++
1009/1010	F24A;V56I;F701W	++++++	++++++
1011/1012	F24A;Y31C;V56I;S386P;F701Y	++	++++++
1013/1014	F24A;F71C;F701W	++++	++++++
1015/1016	F24A;L225V;F701Y	+++	++++++
1017/1018	F24A;Y31W;V56T;V264A;F701W;S750G	++++++	++++++
1019/1020	F24A;Y31W;F71C;F701W	+++	++++++
1021/1022	F24A;D484N;F701W	++++++	++++++
1023/1024	F24A;V56T;F701W	+++++	++++++
1025/1026	F24A;Y31W	++++++	++++++
1027/1028	F24A;V56I;W154F;I270V;M697L	+++	++++++
1029/1030	F24A;V56I;M697F;F701W	+++	++++++
1031/1032	F24A;Y31W;V56I;F701Y	++++	++++++
1033/1034	F24A;Y31W;V56I;M697L	++++++	++++++

[0238]

表8.1 脱酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1262)	脱酰化转化百分比FIOP [1]	脱酰化选择性百分比FIOP [2]
1035/1036	F24Y;V28A;F71C;F701W	++++++	++++++
1037/1038	F24Y;V28A;V56I;A308T;T379A;F701W	++++++	++++++
1039/1040	F24Y;V28A;D321N;F701H	++++++	++++++
1041/1042	F24Y;V28A;N457Y;F701W	++++++	++++++
1043/1044	F24Y;V28A;V56I;F701W	++++++	++++++
1045/1046	F24Y;Y27C;V28T;F701W;V729F	++++++	++++++
1047/1048	V56T;F71G;F701W	++++++	++++++
1049/1050	M697L;F701Y	++++	++++
1051/1052	V56T;W658R;F701W	++	++
1053/1054	V56T;F71G;F701I	++++	++++++
1055/1056	A410P;M697F;F701W	+++++	++++++
1057/1058	V56L;K322R;M697L;F701V	+++++	++++++
1059/1060	V56L;F701H	++++++	++++++
1061/1062	Y31C;V56L;F701H	++++++	++++++
1063/1064	F71G;F701W	++++++	++++++
1065/1066	T129A;I511F;F701Y	++++	+++++
1067/1068	V56T;F701W;L711Q	++	++
1069/1070	I423T;F701W	++	++
1071/1072	V56I;W154F;F701W	+++	+++
1073/1074	V56L;W119R;K146E;F701Y	++++	++++++
1075/1076	V56T;M697L;F701W	++++++	++++++

[1]每种变体的转化百分比相对于参考多肽SEQ ID NO: 1262确定并如下定义："+"=活性增加大于1倍但小于2.0倍；"++"=活性增加大于2.0倍但小于3倍；"+++"=活性增加大于3倍但小于4倍；"++++"=大于4倍但小于5倍；"+++++"=大于5倍但小于6倍；"++++++"=大于6倍。

[2]每种变体的选择性百分比相对于参考多肽SEQ ID NO: 1262确定并如下定义："+"=选择性增加大于1倍但小于2.0倍；"++"=选择性增加大于2.0倍但小于3倍；"+++"=选择性增加大于3倍但小于4倍；"++++"=选择性增加大于4倍但小于5倍；"+++++"=选择性增加大于5倍但小于6倍；"++++++"=选择性增加大于6倍。

表8.2 脱酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	脱酰化转化百分比FIOP [1]	脱酰化选择性百分比FIOP [2]
1077/1078	G71Y;D74K	+++++	++++++
1079/1080	Y31A	+	+++
1081/1082	G71K;D74L	+	+++
1083/1084	G71K;D74N;T129P	+	+++
1085/1086	T32R	+	++++

[0239]

表8.2 脱酰化结果

SEQ ID NO:	氨基酸差异 (相对于SEQ ID NO: 1288)	脱酰化转化百分比FIOP [1]	脱酰化选择性百分比FIOP [2]
1087/1088	A69L;D74T	+	+++
1089/1090	V75H	+	++
1091/1092	R141A	+	+++
1093/1094	P22C	+	++++
1095/1096	F57W	+	++++
1097/1098	G71E	+	+++++
1099/1100	G71E;D74T	+++	+++++
1101/1102	G71W;D74T	+++	++++
1103/1104	G71W;D74G	+++	++++
1105/1106	G71W;D74S	+++	++++
1107/1108	G71D;D74H	+++	++++
1109/1110	G71D;D74M	+++	++++
1111/1112	G71D;D74L	+++	++++
1113/1114	G71E;D74T;A470V	++	+++
1115/1116	G71W	++	++++
1117/1118	G71D	+++	++++
1119/1120	K394Y	++	++++
1121/1122	G71H;D74M	++	+++
1123/1124	G71I;D74Q;G145S	++	++
1125/1126	G71H;D74H	++	+++
1127/1128	F50W	+	++
1129/1130	G71I;A149V	+	++
1131/1132	G71N;D74Q	++	++
1133/1134	G71N;D74A	+	++
1135/1136	G71H	++	++
1137/1138	Y31L;V56I;V264A;A308T;T379A;D484N;Q547K;L711Q;S750G	+	++
1139/1140	G71N;D74V;Y248C	+	++
1141/1142	G71H;D74G	++	++
1143/1144	G71H;D74P	++	+++
1145/1146	G71N;D74S	++	++
1147/1148	G71F;D74E	++	++
1149/1150	G71Y;D74L	+++	+++
1151/1152	G71H;D74S	++	++
1153/1154	G71H;D74A	++	+++
1155/1156	G71L;D74M	+	+
1157/1158	G71T	++	++

[1]每种变体的转化百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=活性增加大于1倍但小于2.0倍；"++"=活性增加大于2.0倍但小于3倍；"+++"=活性增加大于3倍但小于4倍；"++++"=大于4倍但小于5倍；"+++++"=大于5倍但小于6倍；"++++++"=大于6倍。

[2]每种变体的选择性百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=选择性增加大于1倍但小于2.0倍；"++"=选择性增加大于2.0倍但小于3倍；"+++"=选择性增加大于3倍但小于4倍；"++++"=选择性增加大于4倍但小于5倍；"+++++"=选择性增加大于5倍但小于6倍；

[0240]

表8.2 脱酰化结果

SEQ ID NO:	氨基酸差异 (相对于SEQ ID NO: 1288)	脱酰化转化百分比FIOP [1]	脱酰化选择性百分比FIOP [2]
“++++++”=选择性增加大于6倍。			

[0241] 实施例9

[0242] 在A1位置三-保护的胰岛素相比于SEQ ID NO:1036在脱酰化方面的改进

[0243] 在实施例8中描述的筛选变体后,选择了SEQ ID NO:1036作为亲本酶。使用良好建立的技术(例如,饱和诱变、先前鉴定的有益突变的重组) 产生工程化基因的文库。如实施例2中描述,在HTP中产生由每种基因编码的多肽,并如实施例3中描述产生可溶性裂解物。

[0244] 在30℃在200μL反应物中筛选每种变体5小时,所述反应物包含5g/L A1/B1/B29三-乙酸苯酯胰岛素、200mM Tris缓冲液pH=8.0和10μL可溶性裂解物。将96孔板热密封并在Thermotron®振荡器中以100rpm孵育。用200μl乙腈或二甲基乙酰胺猝灭反应物,并使用台式振荡器混合5分钟。然后将板以4000rpm离心5分钟,将其稀释2倍至水中,并加载到HPLC中用于分析。

[0245] 相对于SEQ ID NO:1036的转化百分比(转化百分比FIOP)被计算为由变体形成的产物的转化百分比相对于由SEQ ID NO:1036产生的转化百分比并显示在表9.1中。通过将如HPLC分析所观察到的产物峰面积除以底物、产物和杂质/副产物峰的面积总和来定量转化百分比。

[0246] 相对于SEQ ID NO:1036的选择性百分比(选择性百分比FIOP)被计算为由变体形成的产物的选择性百分比相对于由SEQ ID NO:1036产生的选择性百分比并显示在表9.1中。通过将如HPLC分析所观察到的产物峰面积除以产物和杂质/副产物峰的面积总和来定量选择性百分比。

[0247]

表9.1 脱酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1036)	脱酰化转化百分比FIOP [1]	脱酰化选择性百分比FIOP [2]
1175/1176	A467L	+	+
1171/1172	T384Y	+	+
1177/1178	A255Y	+	+
1165/1166	D623E	+	+
1169/1170	L253M	+	+
1183/1184	T176C	+	+
1163/1164	A255S	+	+
1181/1182	T384R	+++	+
1159/1160	T384K	+	+
1185/1186	L536T	+	+
1189/1190	L536K	+	+
1193/1194	N2L	+++	+
1179/1180	F460Y	+	+
1167/1168	A255H	+	+
1187/1188	N47R	++	+
1161/1162	L253N	+++	
1173/1174	L253Q	+	
1191/1192	L536R	+	

[1]转化百分比相对于参考多肽SEQ ID NO: 1036确定并如下定义："+"=活性增加大于1倍但小于2.0倍；"++"=活性增加大于2.0倍但小于2.5倍；"+++"=活性增加大于2.5倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1036确定并如下定义："+"=选择性增加大于1倍但小于2.0倍；"++"=选择性增加大于2.0倍但小于2.5倍；"+++"=选择性增加大于2.5倍。

[0248] 实施例10

[0249] 在A1位置三-保护的胰岛素相比于SEQ ID NO:1194在脱酰化方面的改进

[0250] 在实施例9中描述的筛选变体后，选择了SEQ ID NO:1194作为亲本酶。使用良好建立的技术(例如，饱和诱变、先前鉴定的有益突变的重组) 产生工程化基因的文库。如实施例2中描述，在HTP中产生由每种基因编码的多肽，并如实施例3中描述产生可溶性裂解物。

[0251] 在30℃在200μL反应物中筛选每种变体5小时，所述反应物包含15 g/L A1/B1/B29三-乙酸苯酯胰岛素、200mM Tris缓冲液pH＝8.0和10μL 可溶性裂解物。将96孔板热密封并在thermotron中以100rpm孵育。用 200μl乙腈或二甲基乙酰胺猝灭反应物，并使用台式振荡器混合5分钟。然后将板以4000rpm离心5分钟，将其稀释2倍至水中，并加载到HPLC 中用于分析。

[0252] 相对于SEQ ID NO:1194的转化百分比(转化百分比FIOP)被计算为由变体形成的产物的转化百分比相对于由SEQ ID NO:1194产生的转化百分比并显示在表10.1中。通过将如HPLC分析所观察到的产物峰面积除以底物、产物和杂质/副产物峰的面积总和来计算转化百分比。

[0253] 相对于SEQ ID NO:1194的选择性百分比(选择性百分比FIOP)被计算为由变体形成的产物的选择性百分比相对于由SEQ ID NO:1194产生的选择性百分比并显示在表10.1中。通过将如HPLC分析所观察到的产物峰面积除以产物和杂质/副产物峰的面积总和来计算选择性百分比。

[0254]

表10.1 脱酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1194)	脱酰化转化百分比 FIOP [1]	脱酰化选择性百分比 FIOP [2]
1195/1196	V264A;T384K;A467L;D484N;L536T;Q547K	+	+
1197/1198	V264A;T384R;A467L;D484N;L536T;Q547K	+	+
1199/1200	T384K;A467L;D484N;L536T;Q547K	+++	+
1201/1202	V264A;T384R;A467L;D484N;L536T;Q547K	+	+
1203/1204	V264A;T384R;A467L;D484N;L536T;Q547K	+	+
1205/1206	V264A;T384R;A467L;D484N;L536T;Q547K	+	+
1207/1208	T384R;A467L;D484N;L536T;Q547K	+++	
1209/1210	V264A;T384R;A467L;D484N;L536T;Q547K	+	+
1211/1212	A28T;V264A;T384R;A467L;D484N;L536T;Q547K	++	+
1213/1214	Q12A;K103E;W119Y;Q233K;T384R;G444N;N494D;S646D	++	
1215/1216	Q12A;K103E;T131A;Q233K;G444N;N494D;S646D	+	+
1217/1218	Q12A;K103E;W119Y;T131A;Q233K;G444N;N494D;S646D	+	+
1219/1220	Q12A;K103E;T131A;Q233K;G444N;A467L;N494D;S646D	++	+
1221/1222	Q12A;K103E;W119Y;Q233K;G444N;N494D;S646D	+	+
1223/1224	K103E;Q233K;G444N;N494D;S646D	+	+
1225/1226	Q12A;K103E;Q233K;G444N;N494D;S646D	+	
1227/1228	Q12A;K103E;W119Y;Q233K;G444N;N494D;S646D	+	+
1229/1230	K103E;W119Y;Q233K;G444N;N494D;S646D	+	+
1231/1232	Q12A;K103E;W119Y;T131A;Q233K;T384K;G444N;N494D;S646D	++	
1233/1234	Q12A;K103E;W119Y;Q233K;G444N;N494D;S646D	+	+
1235/1236	Q12A;K103E;W119Y;T131A;Q233K;G444N;N494D;S646D	+	+
1237/1238	Q12A;K103E;W119Y;T131A;Q233K;T384K;G444N;N	++	

[0255]

表10.1 脱酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1194)	脱酰化转化百分比 FIOP [1]	脱酰化选择性百分比 FIOP [2]
	494D;S646D		
1239/1240	Q12A;K103E;W119Y;T131A;Q233K;G444N;N494D;S646D	+	+
1241/1242	Q12A;K103E;W119Y;T131A;Q233K;G444N;N494D;S646D	+	+
1243/1244	Q12A;K103E;W119Y;Q233K;G444N;A467L;N494D;S646D	+	+
1245/1246	Q12A;K103E;W119Y;Q233K;G444N;A467L;N494D;S646D	+	+
1247/1248	Q12A;K103E;W119Y;Q233K;G444N;N494D;L536T;S646D	+	+
1249/1250	K103E;W119Y;T131A;Q233K;T384R;G444N;N494D;S646D	++	
1251/1252	Q12A;W119Y;Q233K;T384R;G444N;N494D;S646D	+	+
1253/1254	K668E	+	+

[1]转化百分比相对于参考多肽SEQ ID NO: 1194确定并如下定义："+"=活性增加大于1倍但小于2.0倍；"++"=活性增加大于2.0倍但小于2.5倍；"+++"=活性增加大于2.5倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1194确定并如下定义："+"=选择性增加大于1倍但小于2.0倍；"++"=选择性增加大于2.0倍但小于2.5倍；"+++"=选择性增加大于2.5倍。

[0256] 实施例11

[0257] 在A1、B1和B29位置酰化胰岛素方面的改进

[0258] 筛选Codex®Acylase Panel(Codexis)中和美国临时专利申请系列号 62/158,118中公开的一系列变体，以确定其在A1、B1和B29位置处酰化胰岛素的能力。如实施例2中描述，在HTP中产生由每种基因编码的多肽，并如实施例3中描述产生可溶性裂解物。

[0259] 在30℃在200μL反应物中筛选每种变体20小时，所述反应物包含10 g/L胰岛素、0.1M CHES缓冲液pH＝10、17g/L苯乙酸甲酯和20μL可溶性裂解物。将96孔板热密封并在Thermothron®振荡器中以300rpm孵育。用200μl乙腈猝灭反应物，并使用台式振荡器混合5分钟。然后将板以4000 rpm离心5分钟，并加载到HPLC中用于分析。

[0260] 通过将如HPLC分析所观察到的产物峰面积除以底物、产物和杂质/ 副产物峰的面积总和来计算转化百分比(%conv.)。通过将如HPLC分析所观察到的产物峰面积除以产物和杂质/副产物峰的面积总和来计算选择性百分比(%sel.)。结果提供于图2中。

[0261] 实施例12

[0262] 与SEQ ID NO:1288相比，在A1、B1和B29位置酰化胰岛素方面的改进

[0263] 在实施例1中描述的筛选变体和鉴定在位置B29酰化胰岛素的最佳酶后选择SEQ ID NO:1288作为亲本酶。使用良好建立的技术(例如，饱和诱变、先前鉴定的有益突变的重

组)产生工程化基因的文库。如实施例2中描述,在HTP中产生由每种基因编码的多肽,并如实施例3中描述产生可溶性裂解物。

[0264] 在30℃在200μL反应物中筛选每种变体5小时,所述反应物包含10 g/L胰岛素、0.1M TRIS缓冲液pH=9.25、20%乙腈、17g/L苯乙酸甲酯和 10μL澄清的裂解物。将96孔板热密封并在Thermotron®振荡器中以100 rpm孵育。用200μl乙腈猝灭反应物,并使用台式振荡器混合5分钟。然后将板以4000rpm离心5分钟,并加载到HPLC中用于分析。

[0265] 相对于SEQ ID NO:1288的转化百分比(转化百分比FIOP)被计算为由变体形成的产物的转化百分比相对于由SEQ ID NO:1288产生的转化百分比并显示在表12.1、12,2、12,3、12.4、12.5、12.6和12.7中。通过将如 HPLC分析所观察到的产物峰面积除以底物、产物和杂质/副产物峰的面积总和来计算转化百分比。

[0266] 相对于SEQ ID NO:1288的选择性百分比(选择性百分比FIOP)计算为由变体形成的产物的选择性百分比相对于由SEQ ID NO:1288产生的选择性百分比并显示在表12.1、12,2、12,3、12.4、12.5、12.6和12.7中。通过将如HPLC分析所观察到的产物峰面积除以产物和杂质/副产物峰的面积总和来计算选择性百分比。

[0267]

表12.1 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1酰化转化百分比(FIOP)[1]	A1酰化选择性百分比(FIOP)[2]
1677/1678	Y27T;F254W;A470V	++	+
1679/1680	Y27T;L253V;A255G;N348R	+	+
1685/1686	Y27T;A255G;W370I	+++	

[0268]

表12.1 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1酰化转化百分比(FIOP)[1]	A1酰化选择性百分比(FIOP)[2]
1701/1702	Y27T;D74S;F254W;A255G;N348R;K369C;T384P	+	+
1703/1704	Y27T;D74N;L253V;F254W;N348R;K369C;T384P	++	+
1759/1760	L253M	+	+
1797/1798	T384R	++	+
1799/1800	D623N	+	
1811/1812	L253S	+	+
1817/1818	N457T	+	+
1819/1820	R317S;Q380P	+	+
1821/1822	A373Y	+	+
1859/1860	K128W	+	+
1901/1902	T705S	+	+

[1]转化百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=活性增加大于1倍但小于1.5倍；"++"=活性增加大于1.5倍但小于2.0倍；"+++"=活性增加大于2.0倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=选择性增加大于0.95倍。

表12.2 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B29酰化转化百分比(FIOP)[1]	B29酰化选择性百分比(FIOP)[2]
1371/1372	Y27T;N348R;D381K	+	+
1417/1418	Y27T;D74S;A255G;N348R;K369C;D381K		+
1605/1606	Y27T;D74S;A255G;N348R;D381K;T384P	+	+
1637/1638	Y27T;N348R;K369C;W370I;D381K;T384P		+
1659/1660	Y27T;D74G;F254W;A255G;N348R;K369C;W370I;D381K		+
1689/1690	Y27T;F254W;A255G;N348R;W370I;D381K		+
1707/1708	Y27T;A255G;N348R;W370I;D381K;T384P		+
1711/1712	D381F	+	+
1715/1716	Q134M	+	+
1717/1718	D623W	++	+
1719/1720	L253R	+	+

[0269]

表12.2 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B29酰化转化百分比(FIOP)[1]	B29酰化选择性百分比(FIOP)[2]
1721/1722	N627M	+	+
1723/1724	N627R	+	+
1725/1726	D623N	++	+
1729/1730	K615V	+	+
1731/1732	D381L	+	+
1733/1734	D381R	+	++
1735/1736	A132G	+	+
1737/1738	A467S	+	+
1741/1742	F256Y	+	+
1743/1744	D623V	+	+
1745/1746	K615H		+
1747/1748	D623A	+	+
1749/1750	D381Q	+	+
1753/1754	K615C		+
1755/1756	T384R	+	+
1761/1762	F256H	+	+
1773/1774	T453C		+
1775/1776	D381V		+
1777/1778	D381K	+	+
1779/1780	D381F;Q672K		+
1781/1782	D623Y	+	+
1785/1786	D623R	+++	++
1787/1788	D623F	+	+
1789/1790	D623K	+++	++
1793/1794	D381I		+
1823/1824	A373K	+++	++
1825/1826	S706K	+	++
1829/1830	N348K;A467T	+	+
1831/1832	D709G	+	+
1833/1834	D709A	+	+
1835/1836	F620R	+	+
1837/1838	D709N	+	++
1841/1842	D709H	+	+
1843/1844	E377A		+
1845/1846	F620K	+	+
1847/1848	S706R		+
1849/1850	D709R	+	++++
1853/1854	N20S;D709Q		+
1857/1858	D709S	+	+
1867/1868	V618C	+	+
1871/1872	A69M		+
1875/1876	F254K	+	+
1877/1878	A84V		+
1879/1880	F701H		+
1881/1882	P383K		+
1883/1884	A69L		+++
1885/1886	I708V		+

[0270]

表12.2 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B29酰化转化百分比(FIOP)[1]	B29酰化选择性百分比(FIOP)[2]
1889/1890	A255K	++	++
1893/1894	A255R	++	++
1895/1896	A69V		+
1897/1898	P383R		++++

[1]转化百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=活性增加大于2倍但小于4倍；"++"=活性增加大于4倍但小于6倍；"+++"=活性增加大于6倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=选择性增加大于2.0倍但小于5倍；"++"=选择性增加大于5.0倍但小于10倍；"+++"=选择性增加大于10.0倍但小于15倍；"++++"=选择性增加大于15.0倍。

表12.3 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B1酰化转化百分比(FIOPC)[1]	B1酰化选择性百分比(FIOPC)[2]
1315/1316	Y27T;L253V	+	
1317/1318	Y27T;D74G;L253V;F254W	++	+
1323/1324	Y27T;D74G;L253V;A255G;N348R;W370I;T384P	++	+
1325/1326	Y27T;D74N;L253V;F254W;N348R;W370I;D381K;T384P	++	++
1327/1328	Y27T;D74N;L253V;F254W;A255G;N348R;W370I	++	++
1335/1336	Y27T;D74G;L253V;N348R;K369C;W370I;D381F;T384P	+	+
1337/1338	Y27T;L253V;F254W;N348R	+	
1339/1340	Y27T;D74P;L253V;F254W;A255G;N348R	+	+
1345/1346	Y27T;D74G;L253V;F254W;A255G;W370I	++	++
1347/1348	Y27T;L253V;F254W;A255G	+	
1349/1350	Y27T;D74G;L253V;F254W;T384P	++	+
1351/1352	Y27T;D74N;F254W;N348R;W370I	+	+
1353/1354	Y27T;D74S;L253V;N348R;W370I;D381K;T384P	++	+
1357/1358	Y27T;L253V;A255G;W370I	+	+
1359/1360	Y27T;D74G;F254W;A255G;D381F	+	+
1365/1366	Y27T;L253V;K369C;W370I	+	

[0271]

表12.3 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B1酰化转化百分比(FIOPC)[1]	B1酰化选择性百分比(FIOPC)[2]
1377/1378	Y27T;L253V;F254W;N348R;D381F	+	+
1379/1380	Y27T;L253V;F254W;A255G;N348R;W370I	+	+
1381/1382	Y27T;D74S;L253V;A255G;T384P	+	+
1387/1388	Y27T;D74N;L253V;N348R;W370I;D381W;T384P	++	+
1393/1394	Y27T;D74P;L253V;N348R	+	+
1399/1400	Y27T;D74G;F254W;A255G;N348R;K369C;W370I;D381F	+	+
1401/1402	Y27T;D74P;F254W;A255G;N348R;K369C;W370I	+	+
1403/1404	Y27T;D74N;L253V;F254W;W370I;D381K	++	++
1409/1410	Y27T;D74G;L253V;T384P	+	+
1411/1412	Y27T;D74S;N348R;W370I	+	+
1413/1414	Y27T;D74S;L253V;F254W;A255G;K369C;W370I	+	+
1419/1420	Y27T;D74G;L253V;N348R;K369C;W370I	+	
1425/1426	Y27T;D74N;L253V;F254W;A255G	++	+
1429/1430	Y27T;D74P;L253V;F254W;N348R	+	+
1435/1436	Y27T;D74G;A255G;W370I	+	+
1437/1438	Y27T;D74G;L253V;A255G;N348R;D381W	+	+
1439/1440	Y27T;D74P;L253V;W370I	++	+
1441/1442	Y27T;L253V;N348R;W370I;T384P	+	
1443/1444	Y27T;D74G;L253V;F254W;A255G;W370I;D381K;T384P	++	+
1447/1448	Y27T;D74N;F254W;A255G;N348R;W370I;D381K	+	+
1449/1450	Y27T;D74S;L253V;N348R;K369C;W370I;T384P	+	+
1453/1454	Y27T;D74N;L253V;F254W;N348R;W370I;D381F	++	++
1455/1456	Y27T;D74S;F254W;A255G;N348R;W370I;D381F;T384P	+	+
1457/1458	Y27T;D74G;L253V;F254W;A255G;D381K	++	++
1459/1460	Y27T;D74S;F254W;K369L;W370I	+	+
1463/1464	Y27T;L253V;F254W;D381F;T384P	+	+
1465/1466	Y27T;D74G;L253V;F254W;A255G;K369C;W370I	+	
1467/1468	Y27T;D74G;F254W;A255G;N348R;W370I	++	+
1469/1470	Y27T;D74G;L253V;F254W;A255G;N348R;W370I;T384P	++	++
1471/1472	L253V;N348R;W370I	+	

[0272]

表12.3 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B1酰化转化百分比(FIOPC)[1]	B1酰化选择性百分比(FIOPC)[2]
1475/1476	Y27T;D74N;L253V;F254W;K369C	+	+
1477/1478	Y27T;D74G;F254W;A255G;N348R	+	
1479/1480	Y27T;D74S;L253V;N348R;D381W	+	+
1481/1482	Y27T;D74S;L253V;N348R	+	+
1483/1484	Y27T;L253V;F254W;A255G;N348R	+	
1489/1490	Y27T;D74S;L253V;A255G;N348R;W370I;D381K	++	++
1491/1492	Y27T;F254W;A255G;N348R;K369C;W370I	+	+
1493/1494	Y27T;D74N;L253V;F254W;A255G;N348R;W370I;T384P	++	+++
1495/1496	Y27T;D74S;F254W;A255G;W370I	++	++
1497/1498	Y27T;F254W;A255G;K369C;W370I;D381F;T384P	+	
1501/1502	Y27T;L253V;F254W;D381F	++	+
1503/1504	Y27T;D74N;F254W;N348R	+	
1505/1506	Y27T;F254W;A255G;W370I	+	
1509/1510	Y27T;D74N;A107V;A255G;N348R;K369C;W370I	+	
1511/1512	Y27T;F254W;A255G;N348R;W370I	+	
1515/1516	Y27T;D74N;F254W	+	+
1521/1522	Y27T;D74G;L253V;D381F;T384P	+	+
1523/1524	Y27T;D74P;L253V;A255G	+	+
1525/1526	Y27T;D74S;L253V;F254W;A255G;N348R	+	+
1529/1530	Y27T;D74S;F254W;A255G;N348R;K369C;W370I	+	
1533/1534	Y27T;D74P;W370I	+	
1537/1538	Y27T;L253V;F254W;T384P	+	+
1543/1544	Y27T;D74N;L253V;F254W;A255G;W370I	+++	+++
1545/1546	Y27T;D74G;L253V;N348R;W370I	++	+
1553/1554	Y27T;F254W;A255G;N348R;W370I;T384P	+	
1555/1556	Y27T;D74G;L253V;F254W;N348R	+	+
1563/1564	Y27T;L253V;A255G;W370I;D381F;T384P	+	
1565/1566	Y27T;L253V;A255G;N348R;D381K	+	+
1567/1568	Y27T;L253V;F254W;N348R;W370I;T384P	+	+
1569/1570	Y27T;D74N;F254W;N348R;W370I;D381K	+	+
1571/1572	Y27T;D74N;L253V;A255G;N348R;K369C;W370I	+	+
1573/1574	Y27T;L253V;F254W;N348R;D381W;T384P	+	+
1577/1578	Y27T;L253V;F254W;A255G;G260C	+	+

[0273]

表12.3 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B1酰化转化百分比(FIOPC)[1]	B1酰化选择性百分比(FIOPC)[2]
	;N348R;D381F;T384P		
1579/1580	Y27T;D74P;L253V;F254W;N348R;D381F;T384P	+	+
1583/1584	Y27T;D74G;L253V;N348R;K369C;W370I;D381F	+	+
1585/1586	Y27T;D74N;L253V;N348R	+	+
1587/1588	Y27T;D74G;L253V;A255G;N348R;T384P	+	+
1589/1590	Y27T;D74N;L253V;A255G;W370I	++	++
1591/1592	Y27T;D74N;A255G;N348R;W370I	+	+
1593/1594	Y27T;L253V;N348R;D381F;T384P	+	
1603/1604	Y27T;L253V;D381F;T384P	+	
1607/1608	Y27T;D74P;L253V;F254W;N348R;W370I;D381W;T384P	++	+
1609/1610	Y27T;L253V;F254W;A255G;N348R;W370I;T384P	+	
1627/1628	Y27T;L253V;F254W	+	+
1631/1632	Y27T;L253V;F254W;N348R;W370I;D381F	++	++
1633/1634	Y27T;D74G;L253V;A255G	+	+
1635/1636	Y27T;D74N;L253V;F254W;N348R;W370I	+++	+++
1639/1640	Y27T;D74N;L253V;F254W;N348R;K369C;W370I;D381K	+	+
1643/1644	Y27T;D74G;L253V;F254W;A255G;N348R;K369C;W370I;D381F	+	+
1645/1646	Y27T;D74G;L253V;F254W;N348R;K369C;W370I	+	+
1655/1656	Y27T;D74N;L253V	+	+
1657/1658	Y27T;D74G;L253V;F254W;A255G;N348R	+	+
1661/1662	Y27T;F254W;A255G;T384P	+	
1663/1664	Y27T;F254W;A255G	++	+
1667/1668	Y27T;L253V;F254W;A255G;N348R;D381F;T384P	+	+
1669/1670	Y27T;F254W;A255G;N348R	+	+
1671/1672	Y27T;D74N;F254W;T384P	+	
1673/1674	Y27T;L253V;N348R	+	+
1675/1676	Y27T;D74G;A255G;N348R	+	
1677/1678	Y27T;F254W;A470V	+	
1679/1680	Y27T;L253V;A255G;N348R	+	
1681/1682	Y27T;D74N;L253V;K369C;D381K;T384P	+	+
1683/1684	Y27T;D74G;L253V;A255G;N348R;D381F	++	+
1685/1686	Y27T;A255G;W370I	+	
1687/1688	Y27T;D74S;F254W;A255G;N348R	+	

[0274]

表12.3 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B1酰化转化百分比(FIOPC)[1]	B1酰化选择性百分比(FIOPC)[2]
1693/1694	Y27T;D74S;L253V;F254W;N348R;D381W;T384P	+++	++
1695/1696	Y27T;D74N;F254W;A255G;N348R	+	
1697/1698	Y27T;D74P;L253V;F254W;A255G	++	+
1705/1706	Y27T;L253V;N348R;W370I;D381F;T384P	++	++
1709/1710	Y27T;D74S;A255G;W370I	+	
1815/1816	N388E	+	++
1865/1866	A255P	+	

[1]转化百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=活性增加大于2倍但小于5倍；"++"=活性增加大于5倍但小于10倍；"+++"=活性增加大于10倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=选择性增加大于2.0倍但小于5倍；"++"=选择性增加大于5.0倍但小于7倍；"+++"=选择性增加大于7.0倍。

表12.4 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B29酰化转化百分比(FIOPC)[1]	A1/B29酰化选择性百分比(FIOPC)[2]
1319/1320	Y27T;D74S;F254W;N348R;D381W	+	+
1329/1330	Y27T;A255G;N348R;W370I	+	
1333/1334	Y27T;N348R;T384P	+	+
1363/1364	Y27T;D74G;F254W;A255G;N348R;D381W	+	+
1373/1374	Y27T;D74S;N348R	+	
1499/1500	Y27T;A255G;N348R	+	+
1507/1508	Y27T;L253V;N348R;D381F;T384P	+	+
1541/1542	Y27T;L253V;F254W;N348R;T384P	+	
1551/1552	Y27T;D74P;F254W;A255G;N348R;D381K;T384P	+	+
1595/1596	Y27T;N348R	+	+
1599/1600	Y27T;F254W;A255G;N348R;W370I;D381W;T384P	+	
1605/1606	Y27T;D74S;A255G;N348R;D381K;T384P	+	+
1613/1614	Y27T;D74G;F254W;N348R;D381W;T384P	+	+
1615/1616	Y27T;F254W;A255G;D381K;T384P	+	+

[0275]

表12.4 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B29酰化转化百分比 (FIOPC)[1]	A1/B29酰化选择性百分比 (FIOPC)[2]
1621/1622	Y27T;D74S;F254W;N348R;D381F	+	+
1647/1648	Y27T;A255G;N348R;D381W;T384P	+	+
1651/1652	Y27T;F254W;N348R;D381W;T384P	+	+
1653/1654	Y27T;D74G;N348R	+	+
1661/1662	Y27T;F254W;A255G;T384P	++	+
1663/1664	Y27T;F254W;A255G	++	+
1665/1666	Y27T;D74S;L253V;N348R	+++	++
1667/1668	Y27T;L253V;F254W;A255G;N348R;D381F;T384P	+	+
1669/1670	Y27T;F254W;A255G;N348R	+	+
1671/1672	Y27T;D74N;F254W;T384P	+++	+
1673/1674	Y27T;L253V;N348R	+	
1675/1676	Y27T;D74G;A255G;N348R	++	++
1677/1678	Y27T;F254W;A470V	+	
1679/1680	Y27T;L253V;A255G;N348R	+	+
1683/1684	Y27T;D74G;L253V;A255G;N348R;D381F	++	+
1685/1686	Y27T;A255G;W370I	++	+
1687/1688	Y27T;D74S;F254W;A255G;N348R	+++	++
1691/1692	Y27T;D74N;L253V;F254W	+	
1693/1694	Y27T;D74S;L253V;F254W;N348R;D381W;T384P	++	++
1695/1696	Y27T;D74N;F254W;A255G;N348R	+++	++
1699/1700	Y27T;D74G;L253V;F254W;A255G;N348R	+++	+
1705/1706	Y27T;L253V;N348R;W370I;D381F;T384P	+	+
1707/1708	Y27T;A255G;N348R;W370I;D381K;T384P	+	+
1709/1710	Y27T;D74S;A255G;W370I	++	+
1717/1718	D623W	+	+
1719/1720	L253R	+	+
1747/1748	D623A	+	+
1757/1758	D623N	+	+
1761/1762	F256H	+	+
1763/1764	A616R	+	
1771/1772	D623L	+	+
1781/1782	D623Y	+	+
1787/1788	D623F	+	+
1789/1790	D623K	+	+
1803/1804	D381Q	+	+
1805/1806	T384R	++	++
1813/1814	D623R	++	+++
1823/1824	A373K	+	++
1851/1852	H472R	+	
1855/1856	F620R	+	+
1865/1866	A255P	+	

[0276]

表12.4 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B29酰化转化百分比 (FIOPC)[1]	A1/B29酰化选择性百分比 (FIOPC)[2]
1875/1876	F254K	+	+
1893/1894	A255R	+	+

[1]转化百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=活性增加大于2倍但小于5倍；"++"=活性增加大于5倍但小于10倍；"+++"=活性增加大于10倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=选择性增加大于2.0倍但小于5倍；"++"=选择性增加大于5.0倍但小于7倍；"+++"=选择性增加大于7.0倍。

表12.5 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1酰化转化百分比 (FIOPC)[1]	A1/B1酰化选择性百分比 (FIOPC)[2]
1313/1314	Y27T;L253V;A255G;N348R;T384P	+	+
1315/1316	Y27T;L253V	+	+
1317/1318	Y27T;D74G;L253V;F254W	+	+
1323/1324	Y27T;D74G;L253V;A255G;N348R;W370I;T384P	+	+
1325/1326	Y27T;D74N;L253V;F254W;N348R;W370I;D381K;T384P	+	+
1327/1328	Y27T;D74N;L253V;F254W;A255G;N348R;W370I	+	+
1329/1330	Y27T;A255G;N348R;W370I	+	+
1331/1332	Y27T;D74P;L253V;F254W;A255G;N348R;K369C;W370I	+	+
1335/1336	Y27T;D74G;L253V;N348R;K369C;W370I;D381F;T384P	+	+
1337/1338	Y27T;L253V;F254W;N348R	+	+
1339/1340	Y27T;D74P;L253V;F254W;A255G;N348R	+	+
1341/1342	Y27T	+	
1343/1344	Y27T;D74G;L253V;N348R;K369C;W370I	+	+
1345/1346	Y27T;D74G;L253V;F254W;A255G;W370I	++	+
1347/1348	Y27T;L253V;F254W;A255G	+	+
1349/1350	Y27T;D74G;L253V;F254W;T384P	+	+

[0277]

表12.5 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1酰化转化百分比 (FIOPC)[1]	A1/B1酰化选择性百分比 (FIOPC)[2]
1351/1352	Y27T;D74N;F254W;N348R;W370I	+	+
1353/1354	Y27T;D74S;L253V;N348R;W370I;D381K;T384P	+	+
1357/1358	Y27T;L253V;A255G;W370I	+	+
1359/1360	Y27T;D74G;F254W;A255G;D381F	+	+
1365/1366	Y27T;L253V;K369C;W370I	+	+
1375/1376	Y27T;D74P;F254W;A255G;N348R	+	+
1377/1378	Y27T;L253V;F254W;N348R;D381F	+	+
1381/1382	Y27T;D74S;L253V;A255G;T384P	+	+
1383/1384	Y27T;D74G;A255G;W370I	+	+
1385/1386	Y27T;D74S;N348R	+	
1387/1388	Y27T;D74N;L253V;N348R;W370I;D381W;T384P	+	+
1389/1390	Y27T;D74G;L253V;F254W;N348R	+	+
1393/1394	Y27T;D74P;L253V;N348R	+	+
1395/1396	Y27T;L253V;F254W;A255G;N348R	+	+
1397/1398	Y27T;D74S;L253V;F254W;A255G;N348R	+	+
1399/1400	Y27T;D74G;F254W;A255G;N348R;K369C;W370I;D381F	+	+
1401/1402	Y27T;D74P;F254W;A255G;N348R;K369C;W370I	+	+
1403/1404	Y27T;D74N;L253V;F254W;W370I;D381K	++	+
1405/1406	Y27T;D74G;F254W;K369C;W370I	+	+
1407/1408	Y27T;D74G;L253V;N348R;W370I	+	+
1409/1410	Y27T;D74G;L253V;T384P	+	+
1411/1412	Y27T;D74S;N348R;W370I	+	+
1413/1414	Y27T;D74S;L253V;F254W;A255G;K369C;W370I	+	+
1421/1422	Y27T;D74G;F254W;A255G;N348R	+	+
1423/1424	Y27T;L253V;N348R;W370I;T384P	+	+
1425/1426	Y27T;D74N;L253V;F254W;A255G	+	+
1427/1428	Y27T;L253V;F254W	+	+
1431/1432	Y27T;D74G;K369C;W370I	+	+
1433/1434	Y27T;F254W;A449V	+	+
1437/1438	Y27T;D74G;L253V;A255G;N348R;D381W	+	+
1439/1440	Y27T;D74P;L253V;W370I	+	+
1443/1444	Y27T;D74G;L253V;F254W;A255G;W370I;D381K;T384P	++	+
1447/1448	Y27T;D74N;F254W;A255G;N348R;W370I;D381K	+	+
1449/1450	Y27T;D74S;L253V;N348R;K369C;W370I;T384P	+	+
1451/1452	Y27T;D74N;L253V;N348R	+	+
1453/1454	Y27T;D74N;L253V;F254W;N348R;W	+	+

[0278]

表12.5 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1酰化转化百分比 (FIOPC)[1]	A1/B1酰化选择性百分比 (FIOPC)[2]
	370I;D381F		
1455/1456	Y27T;D74S;F254W;A255G;N348R;W370I;D381F;T384P	+	+
1457/1458	Y27T;D74G;L253V;F254W;A255G;D381K	+	+
1459/1460	Y27T;D74S;F254W;K369L;W370I	+	+
1463/1464	Y27T;L253V;F254W;D381F;T384P	+	+
1465/1466	Y27T;D74G;L253V;F254W;A255G;K369C;W370I	+	+
1467/1468	Y27T;D74G;F254W;A255G;N348R;W370I	+	+
1469/1470	Y27T;D74G;L253V;F254W;A255G;N348R;W370I;T384P	+	+
1471/1472	L253V;N348R;W370I	+	+
1473/1474	Y27T;F254W;N348R;W370I	+	+
1475/1476	Y27T;D74N;L253V;F254W;K369C	+	+
1479/1480	Y27T;D74S;L253V;N348R;D381W	+	+
1485/1486	Y27T;D74S;A255G	+	+
1489/1490	Y27T;D74S;L253V;A255G;N348R;W370I;D381K	+	+
1493/1494	Y27T;D74N;L253V;F254W;A255G;N348R;W370I;T384P	+	+
1495/1496	Y27T;D74S;F254W;A255G;W370I	++	+
1497/1498	Y27T;F254W;A255G;K369C;W370I;D381F;T384P	+	+
1501/1502	Y27T;L253V;F254W;D381F	+	+
1503/1504	Y27T;D74N;F254W;N348R	+	+
1505/1506	Y27T;F254W;A255G;W370I	++	+
1509/1510	Y27T;D74N;A107V;A255G;N348R;K369C;W370I	+	+
1511/1512	Y27T;F254W;A255G;N348R;W370I	+	+
1515/1516	Y27T;D74N;F254W	+	+
1517/1518	Y27T;D74S;F254W;K369C;T384P	+	+
1521/1522	Y27T;D74G;L253V;D381F;T384P	+	+
1523/1524	Y27T;D74P;L253V;A255G	+	+
1527/1528	Y27T;D74G;A255G;N348R;K369C;W370I	+	+
1529/1530	Y27T;D74S;F254W;A255G;N348R;K369C;W370I	+	+
1531/1532	Y27T;F254W;A255G;N348R;K369C;W370I	+	+
1533/1534	Y27T;D74P;W370I	+	+
1535/1536	Y27T;D74P;L253V;F254W;N348R;K369C;W370I	+	+
1537/1538	Y27T;L253V;F254W;T384P	+	+
1541/1542	Y27T;L253V;F254W;N348R;T384P	+	+
1543/1544	Y27T;D74N;L253V;F254W;A255G;W	+	+

[0279]

表12.5 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1酰化转化百分比 (FIOPC)[1]	A1/B1酰化选择性百分比 (FIOPC)[2]
	370I		
1547/1548	Y27T;F254W;A255G;N348R	+	+
1549/1550	Y27T;D74S;F254W;N348R	+	+
1553/1554	Y27T;F254W;A255G;N348R;W370I;T384P	+	+
1557/1558	D74N;L253V;F254W;K369C;W370I	+	+
1561/1562	Y27T;D74S;K369C;W370I;D381K;T384P	+	+
1563/1564	Y27T;L253V;A255G;W370I;D381F;T384P	++	+
1567/1568	Y27T;L253V;F254W;N348R;W370I;T384P	+	+
1569/1570	Y27T;D74N;F254W;N348R;W370I;D381K	+	+
1571/1572	Y27T;D74N;L253V;A255G;N348R;K369C;W370I	+	+
1573/1574	Y27T;L253V;F254W;N348R;D381W;T384P	+	+
1577/1578	Y27T;L253V;F254W;A255G;G260C;N348R;D381F;T384P	+	
1579/1580	Y27T;D74P;L253V;F254W;N348R;D381F;T384P	+	+
1587/1588	Y27T;D74G;L253V;A255G;N348R;T384P	+	+
1589/1590	Y27T;D74N;L253V;A255G;W370I	+	+
1591/1592	Y27T;D74N;A255G;N348R;W370I	+	+
1599/1600	Y27T;F254W;A255G;N348R;W370I;D381W;T384P	+	+
1601/1602	Y27T;L253V;A255G;N348R;K369C;W370I	+	+
1603/1604	Y27T;L253V;D381F;T384P	+	+
1607/1608	Y27T;D74P;L253V;F254W;N348R;W370I;D381W;T384P	+	+
1609/1610	Y27T;L253V;F254W;A255G;N348R;W370I;T384P	+	+
1611/1612	A255G;N348R;W370I	+	+
1613/1614	Y27T;D74G;F254W;N348R;D381W;T384P	+	
1615/1616	Y27T;F254W;A255G;D381K;T384P	+	+
1623/1624	Y27T;D74N;F254W;A255G;K369C	+	+
1625/1626	Y27T;D74P;L253V;F254W;N348R	+	+
1631/1632	Y27T;L253V;F254W;N348R;W370I;D381F	+	+
1633/1634	Y27T;D74G;L253V;A255G	+	+
1635/1636	Y27T;D74N;L253V;F254W;N348R;W370I	+	+
1643/1644	Y27T;D74G;L253V;F254W;A255G;N348R;K369C;W370I;D381F	+	+

[0280]

表12.5 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1酰化转化百分比 (FIOPC)[1]	A1/B1酰化选择性百分比 (FIOPC)[2]
1645/1646	Y27T;D74G;L253V;F254W;N348R;K369C;W370I	+	+
1649/1650	Y27T;L253V;F254W;A255G;N348R;W370I	++	+
1655/1656	Y27T;D74N;L253V	+	+
1661/1662	Y27T;F254W;A255G;T384P	++	++
1663/1664	Y27T;F254W;A255G	++	++
1665/1666	Y27T;D74S;L253V;N348R	++	++
1667/1668	Y27T;L253V;F254W;A255G;N348R;D381F;T384P	+	+
1671/1672	Y27T;D74N;F254W;T384P	+++	++
1673/1674	Y27T;L253V;N348R	+	+
1675/1676	Y27T;D74G;A255G;N348R	+	+
1677/1678	Y27T;F254W;A470V	+	+
1679/1680	Y27T;L253V;A255G;N348R	+	+
1683/1684	Y27T;D74G;L253V;A255G;N348R;D381F	+	+
1685/1686	Y27T;A255G;W370I	++	++
1687/1688	Y27T;D74S;F254W;A255G;N348R	++	+
1691/1692	Y27T;D74N;L253V;F254W	++++	++
1693/1694	Y27T;D74S;L253V;F254W;N348R;D381W;T384P	++	++
1695/1696	Y27T;D74N;F254W;A255G;N348R	++	+
1697/1698	Y27T;D74P;L253V;F254W;A255G	+++	++
1699/1700	Y27T;D74G;L253V;F254W;A255G;N348R	++	++
1701/1702	Y27T;D74S;F254W;A255G;N348R;K369C;T384P	+	
1703/1704	Y27T;D74N;L253V;F254W;N348R;K369C;T384P	+	+
1705/1706	Y27T;L253V;N348R;W370I;D381F;T384P	+	+
1709/1710	Y27T;D74S;A255G;W370I	+++	++
1711/1712	D381F	+	+
1713/1714	A132S	+	+
1715/1716	Q134M	+	+
1717/1718	D623W	+	+
1727/1728	T131L	+	+
1731/1732	D381L	+	+
1735/1736	A132G	+	+
1751/1752	W370V	+	+
1765/1766	D381R	+	+
1769/1770	T384R	+	+
1791/1792	D623Y	+	+
1799/1800	D623N	+	+
1801/1802	D623R	+	+
1807/1808	S619I	+	

[0281]

表12.5 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1酰化转化百分比 (FIOPC)[1]	A1/B1酰化选择性百分比 (FIOPC)[2]
1809/1810	L253V	+	+
1839/1840	T133K	+	
1865/1866	A255P	+	+
1887/1888	I708M	+	
1891/1892	F254T	+	+
1899/1900	T705S	+	

[1]转化百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义：“+”=活性增加大于2倍但小于10倍；“++”=活性增加大于10倍但小于50倍；“+++”=活性增加大于50倍但小于100倍；“++++”=活性增加大于100倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义：“+”=选择性增加大于2倍但小于10倍；“++”=选择性增加大于10倍但小于50倍；“+++”=选择性增加大于50倍但小于100倍；“++++”=选择性增加大于100倍。

表12.6 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B1/B29酰化转化百分比 (FIOPC)[1]	B1/B29酰化选择性百分比 (FIOPC)[2]
1323/1324	Y27T;D74G;L253V;A255G;N348R;W370I;T384P	+	+
1325/1326	Y27T;D74N;L253V;F254W;N348R;W370I;D381K;T384P	+	+
1327/1328	Y27T;D74N;L253V;F254W;A255G;N348R;W370I	+	+
1345/1346	Y27T;D74G;L253V;F254W;A255G;W370I	+	
1349/1350	Y27T;D74G;L253V;F254W;T384P	+	
1353/1354	Y27T;D74S;L253V;N348R;W370I;D381K;T384P	+	+
1387/1388	Y27T;D74N;L253V;N348R;W370I;D381W;T384P	++	+
1389/1390	Y27T;D74G;L253V;F254W;N348R	+	
1391/1392	Y27T;L253V;F254W;A255G;N348R;K369C;W370I;D381W;T384P	+	+
1397/1398	Y27T;D74S;L253V;F254W;A255G;N348R	+	+
1403/1404	Y27T;D74N;L253V;F254W;W370I;D381K	+	+

[0282]

表12.6 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B1/B29酰化转化百分比 (FIOPC)[1]	B1/B29酰化选择性百分比 (FIOPC)[2]
1411/1412	Y27T;D74S;N348R;W370I	+	+
1425/1426	Y27T;D74N;L253V;F254W;A255G	+	
1435/1436	Y27T;D74G;A255G;W370I	+	+
1437/1438	Y27T;D74G;L253V;A255G;N348R;D381W	+	+
1439/1440	Y27T;D74P;L253V;W370I	+	+
1443/1444	Y27T;D74G;L253V;F254W;A255G;W370I;D381K;T384P	++	+
1445/1446	Y27T;L253V;F254W;A255G;N348R;K369C;T384P	+	+
1447/1448	Y27T;D74N;F254W;A255G;N348R;W370I;D381K	+	+
1453/1454	Y27T;D74N;L253V;F254W;N348R;W370I;D381F	+	+
1455/1456	Y27T;D74S;F254W;A255G;N348R;W370I;D381F;T384P	+	+
1457/1458	Y27T;D74G;L253V;F254W;A255G;D381K	+	+
1467/1468	Y27T;D74G;F254W;A255G;N348R;W370I	+	+
1469/1470	Y27T;D74G;L253V;F254W;A255G;N348R;W370I;T384P	+	+
1479/1480	Y27T;D74S;L253V;N348R;D381W	+	+
1485/1486	Y27T;D74S;A255G	+	
1489/1490	Y27T;D74S;L253V;A255G;N348R;W370I;D381K	+	+
1493/1494	Y27T;D74N;L253V;F254W;A255G;N348R;W370I;T384P	+	+
1495/1496	Y27T;D74S;F254W;A255G;W370I	+	+
1503/1504	Y27T;D74N;F254W;N348R	+	
1505/1506	Y27T;F254W;A255G;W370I	+	
1511/1512	Y27T;F254W;A255G;N348R;W370I	+	
1521/1522	Y27T;D74G;L253V;D381F;T384P	+	+
1533/1534	Y27T;D74P;W370I	+	
1543/1544	Y27T;D74N;L253V;F254W;A255G;W370I	+	+
1545/1546	Y27T;D74G;L253V;N348R;W370I	+	+
1567/1568	Y27T;L253V;F254W;N348R;W370I;T384P	+	+
1573/1574	Y27T;L253V;F254W;N348R;D381W;T384P	+	+
1577/1578	Y27T;L253V;F254W;A255G;G260C;N348R;D381F;T384P	+	+
1579/1580	Y27T;D74P;L253V;F254W;N348R;D381F;T384P	+	+
1581/1582	Y27T;D74G;L253V;F254W	+	+

[0283]

表12.6 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异（相对于SEQ ID NO: 1288）	B1/B29酰化转化百分比 (FIOPC)[1]	B1/B29酰化选择性百分比 (FIOPC)[2]
1585/1586	Y27T;D74N;L253V;N348R	+	+
1587/1588	Y27T;D74G;L253V;A255G;N348R;T384P	+	+
1589/1590	Y27T;D74N;L253V;A255G;W370I	+	+
1591/1592	Y27T;D74N;A255G;N348R;W370I	+	+
1597/1598	Y27T;D74P;F254W;A255G;N348R;D381K;T384P	+	+
1599/1600	Y27T;F254W;A255G;N348R;W370I;D381W;T384P	+	+
1601/1602	Y27T;L253V;A255G;N348R;K369C;W370I	+	
1605/1606	Y27T;D74S;A255G;N348R;D381K;T384P	+	+
1607/1608	Y27T;D74P;L253V;F254W;N348R;W370I;D381W;T384P	+	+
1609/1610	Y27T;L253V;F254W;A255G;N348R;W370I;T384P	+	+
1613/1614	Y27T;D74G;F254W;N348R;D381W;T384P	+	+
1617/1618	Y27T;D74N;F254W;N348R;W370I	+	+
1619/1620	Y27T;D74P;L253V;F254W;N348R;K369C	+	+
1625/1626	Y27T;D74P;L253V;F254W;N348R	+	
1629/1630	Y27T;F254W;K369C;D381F;T384P	+	+
1631/1632	Y27T;L253V;F254W;N348R;W370I;D381F	+	+
1635/1636	Y27T;D74N;L253V;F254W;N348R;W370I	+	+
1641/1642	Y27T;F254W;N348R;K369C;W370I;D381N;T384P	+	+
1657/1658	Y27T;D74G;L253V;F254W;A255G;N348R	+	
1767/1768	D623N	+	+
1769/1770	T384R	+	+
1771/1772	D623L	+	+
1835/1836	F620R	++	+++
1837/1838	D709N	+++	+++
1859/1860	K128W	++	++
1861/1862	T705E	++	++
1869/1870	A255E	++	++
1873/1874	F254T	++	++

[1]转化百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=活性增加大于2倍但小于10倍；"++"=活性增加大于10倍但小于50倍；"+++"=活性增加大于50倍但小于100倍；"++++"=活性增加大于100倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=选择性增加

[0284]

表12.6 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	B1/B29酰化转化百分比 (FIOPC)[1]	B1/B29酰化选择性百分比 (FIOPC)[2]
大于2倍但小于10倍；“++”=选择性增加大于10倍但小于50倍；“+++”=选择性增加大于50倍但小于100倍；“++++”=选择性增加大于100倍。			

表12.7 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1/B29酰化转化百分比 (FIOPC)[1]	A1/B1/B29酰化选择性百分比 (FIOPC)[2]
1313/1314	Y27T;L253V;A255G;N348R;T384P	+	+
1315/1316	Y27T;L253V	+	+
1317/1318	Y27T;D74G;L253V;F254W	+	+
1319/1320	Y27T;D74S;F254W;N348R;D381W	+	+
1321/1322	Y27T;D74G;A255G;N348R;K369C;D381F;T384P	+	+
1323/1324	Y27T;D74G;L253V;A255G;N348R;W370I;T384P	+	+
1325/1326	Y27T;D74N;L253V;F254W;N348R;W370I;D381K;T384P	++	+
1327/1328	Y27T;D74N;L253V;F254W;A255G;N348R;W370I	+	+
1329/1330	Y27T;A255G;N348R;W370I	++	+
1333/1334	Y27T;N348R;T384P	+	+
1335/1336	Y27T;D74G;L253V;N348R;K369C;W370I;D381F;T384P	+	+
1339/1340	Y27T;D74P;L253V;F254W;A255G;N348R	+	+
1341/1342	Y27T	+	+
1345/1346	Y27T;D74G;L253V;F254W;A255G;W370I	+	+
1347/1348	Y27T;L253V;F254W;A255G	+	+
1349/1350	Y27T;D74G;L253V;F254W;T384P	+	+
1351/1352	Y27T;D74N;F254W;N348R;W370I	+	+
1353/1354	Y27T;D74S;L253V;N348R;W370I;D381K;T384P	++	++
1355/1356	Y27T;L253V;N348R	+	+
1357/1358	Y27T;L253V;A255G;W370I	+	+
1359/1360	Y27T;D74G;F254W;A255G;D381F	+	+
1361/1362	Y27T;F254W;N348R;T384P	+	
1363/1364	Y27T;D74G;F254W;A255G;N348R;D381W	+	+

[0285]

表12.7 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1/B29 酰化转化百分比 (FIOPC)[1]	A1/B1/B29 酰化选择性百分比 (FIOPC)[2]
1367/1368	D74N;F254W;A255G;N348R;T384P	+	+
1369/1370	Y27T;F254W;A255G	+	+
1373/1374	Y27T;D74S;N348R	+	+
1375/1376	Y27T;D74P;F254W;A255G;N348R	+	+
1377/1378	Y27T;L253V;F254W;N348R;D381F	+	+
1381/1382	Y27T;D74S;L253V;A255G;T384P	+	+
1383/1384	Y27T;D74G;A255G;W370I	++	+
1387/1388	Y27T;D74N;L253V;N348R;W370I;D381W;T384P	++	++
1389/1390	Y27T;D74G;L253V;F254W;N348R	++	+
1393/1394	Y27T;D74P;L253V;N348R	+	+
1395/1396	Y27T;L253V;F254W;A255G;N348R	+	+
1397/1398	Y27T;D74S;L253V;F254W;A255G;N348R	+	+
1399/1400	Y27T;D74G;F254W;A255G;N348R;K369C;W370I;D381F	+	+
1401/1402	Y27T;D74P;F254W;A255G;N348R;K369C;W370I	+	+
1403/1404	Y27T;D74N;L253V;F254W;W370I;D381K	++	++
1405/1406	Y27T;D74G;F254W;K369C;W370I	+	+
1407/1408	Y27T;D74G;L253V;N348R;W370I	++	+
1409/1410	Y27T;D74G;L253V;T384P	+	+
1411/1412	Y27T;D74S;N348R;W370I	++	+
1415/1416	Y27T;D74G;A255G;N348R	+	+
1421/1422	Y27T;D74G;F254W;A255G;N348R	+	+
1423/1424	Y27T;L253V;N348R;W370I;T384P	++	+
1425/1426	Y27T;D74N;L253V;F254W;A255G	+	+
1433/1434	Y27T;F254W;A449V	+	+
1437/1438	Y27T;D74G;L253V;A255G;N348R;D381W	+	+
1439/1440	Y27T;D74P;L253V;W370I	+	+
1443/1444	Y27T;D74G;L253V;F254W;A255G;W370I;D381K;T384P	++	++
1447/1448	Y27T;D74N;F254W;A255G;N348R;W370I;D381K	+	+
1453/1454	Y27T;D74N;L253V;F254W;N348R;W370I;D381F	++	++
1455/1456	Y27T;D74S;F254W;A255G;N348R;W370I;D381F;T384P	++	++
1457/1458	Y27T;D74G;L253V;F254W;A255G;D381K	++	+
1459/1460	Y27T;D74S;F254W;K369L;W370I	+	+
1461/1462	Y27T;D74N;F254W;A255G;N348R;K369C;D381F	+	+
1463/1464	Y27T;L253V;F254W;D381F;T384P	++	+
1467/1468	Y27T;D74G;F254W;A255G;N348R;	++	++

[0286]

表12.7 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1/B29 酰化转化百分比 (FIOPC)[1]	A1/B1/B29 酰化选择性百分比 (FIOPC)[2]
	W370I		
1469/1470	Y27T;D74G;L253V;F254W;A255G;N348R;W370I;T384P	+	+
1471/1472	L253V;N348R;W370I	+	+
1473/1474	Y27T;F254W;N348R;W370I	+	+
1479/1480	Y27T;D74S;L253V;N348R;D381W	++	++
1481/1482	Y27T;D74S;L253V;N348R	+	+
1485/1486	Y27T;D74S;A255G	+	+
1487/1488	Y27T;L253V;A255G;N348R	+	+
1489/1490	Y27T;D74S;L253V;A255G;N348R;W370I;D381K	+	+
1493/1494	Y27T;D74N;L253V;F254W;A255G;N348R;W370I;T384P	+	+
1495/1496	Y27T;D74S;F254W;A255G;W370I	++	+
1497/1498	Y27T;F254W;A255G;K369C;W370I;D381F;T384P	+	+
1499/1500	Y27T;A255G;N348R	+	+
1501/1502	Y27T;L253V;F254W;D381F	+	+
1503/1504	Y27T;D74N;F254W;N348R	+	+
1505/1506	Y27T;F254W;A255G;W370I	++	++
1507/1508	Y27T;L253V;N348R;D381F;T384P	+	+
1511/1512	Y27T;F254W;A255G;N348R;W370I	++	++
1513/1514	Y27T;L253V;F254W;N348R	+	+
1515/1516	Y27T;D74N;F254W	+	+
1519/1520	Y27T;D74N;F254W;N348R;K369C;D381F;T384P	+	+
1521/1522	Y27T;D74G;L253V;D381F;T384P	++	++
1523/1524	Y27T;D74P;L253V;A255G	+	+
1529/1530	Y27T;D74S;F254W;A255G;N348R;K369C;W370I	+	+
1533/1534	Y27T;D74P;W370I	++	+
1537/1538	Y27T;L253V;F254W;T384P	+	+
1539/1540	Y27T;D74N;N348R	+	+
1541/1542	Y27T;L253V;F254W;N348R;T384P	+	+
1543/1544	Y27T;D74N;L253V;F254W;A255G;W370I	+	+
1547/1548	Y27T;F254W;A255G;N348R	+	+
1549/1550	Y27T;D74S;F254W;N348R	+	+
1551/1552	Y27T;D74P;F254W;A255G;N348R;D381K;T384P	+	+
1553/1554	Y27T;F254W;A255G;N348R;W370I;T384P	++	++
1559/1560	Y27T;D74N;N348R;T384P	+	+
1561/1562	Y27T;D74S;K369C;W370I;D381K;T384P	+	+
1563/1564	Y27T;L253V;A255G;W370I;D381F;T384P	++	++

[0287]

表12.7 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1/B29 酰化转化百分比 (FIOPC)[1]	A1/B1/B29 酰化选择性百分比 (FIOPC)[2]
1567/1568	Y27T;L253V;F254W;N348R;W370I;T384P	++	+
1569/1570	Y27T;D74N;F254W;N348R;W370I;D381K	+	+
1573/1574	Y27T;L253V;F254W;N348R;D381W;T384P	++	+
1575/1576	Y27T;L253V;F254W	+	+
1577/1578	Y27T;L253V;F254W;A255G;G260C;N348R;D381F;T384P	+	+
1579/1580	Y27T;D74P;L253V;F254W;N348R;D381F;T384P	++	++
1585/1586	Y27T;D74N;L253V;N348R	+	+
1587/1588	Y27T;D74G;L253V;A255G;N348R;T384P	+	+
1589/1590	Y27T;D74N;L253V;A255G;W370I	+	+
1591/1592	Y27T;D74N;A255G;N348R;W370I	+	+
1595/1596	Y27T;N348R	+	+
1599/1600	Y27T;F254W;A255G;N348R;W370I;D381W;T384P	++	++
1603/1604	Y27T;L253V;D381F;T384P	+	+
1605/1606	Y27T;D74S;A255G;N348R;D381K;T384P	+	+
1607/1608	Y27T;D74P;L253V;F254W;N348R;W370I;D381W;T384P	++	++
1609/1610	Y27T;L253V;F254W;A255G;N348R;W370I;T384P	++	+
1611/1612	A255G;N348R;W370I	+	+
1613/1614	Y27T;D74G;F254W;N348R;D381W;T384P	++	++
1615/1616	Y27T;F254W;A255G;D381K;T384P	++	++
1621/1622	Y27T;D74S;F254W;N348R;D381F	+	+
1625/1626	Y27T;D74P;L253V;F254W;N348R	+	+
1631/1632	Y27T;L253V;F254W;N348R;W370I;D381F	+	+
1633/1634	Y27T;D74G;L253V;A255G	+	+
1635/1636	Y27T;D74N;L253V;F254W;N348R;W370I	+	+
1643/1644	Y27T;D74G;L253V;F254W;A255G;N348R;K369C;W370I;D381F	+	+
1645/1646	Y27T;D74G;L253V;F254W;N348R;K369C;W370I	+	+
1647/1648	Y27T;A255G;N348R;D381W;T384P	+	+
1649/1650	Y27T;L253V;F254W;A255G;N348R;W370I	++	+
1651/1652	Y27T;F254W;N348R;D381W;T384P	+	+
1653/1654	Y27T;D74G;N348R	+	+
1655/1656	Y27T;D74N;L253V	+	+

[0288]

表12.7 酰化结果

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1288)	A1/B1/B29 酰化转化百分比 (FIOPC)[1]	A1/B1/B29 酰化选择性百分比 (FIOPC)[2]
1657/1658	Y27T;D74G;L253V;F254W;A255G;N348R	+	+
1717/1718	D623W	+	+
1743/1744	D623V	+	+
1747/1748	D623A	+	+
1749/1750	D381Q	+	+
1771/1772	D623L	+	+
1795/1796	D623N	+	+
1797/1798	T384R	++	+
1823/1824	A373K	+	+
1827/1828	F620R	+	+
1863/1864	F254T	+	+
1865/1866	A255P	+	+

[1]转化百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=活性增加大于2倍但小于10倍；"++"=活性增加大于10倍但小于50倍；"+++"=活性增加大于50倍。

[2]选择性百分比相对于参考多肽SEQ ID NO: 1288确定并如下定义："+"=选择性增加大于2倍但小于10倍；"++"=选择性增加大于10倍但小于50倍；"+++"=选择性增加大于50倍。

[0289] 实施例13

[0290] 用可选择的酰基供体酰化胰岛素

[0291] 使用(2-(4-羟苯基)乙酸甲酯或2-(4-羟苯基)乙酰胺)作为苯基乙酸甲酯的替代物，评价表13.1中列出的五种变体的酰化。如实施例4中描述产生摇瓶粉末。将反应在96孔深孔板中进行，每个孔包含200μL的包含0.1M CHES(pH 10)、5％乙腈，15g/L胰岛素、26g/L酰基供体(2-(4-羟苯基)乙酸甲酯或2-(4-羟苯基)乙酰胺)和1g/L冻干酶粉末。将HTP板热密封并在Thermotron®振荡器(3mm throw，型号#AJ185，Infors)中在30℃，300rpm 孵育2小时。用200μl乙腈猝灭反应物，并使用台式振荡器混合5分钟。然后将板以4000rpm离心5分钟，并加载到HPLC中用于分析。

[0292] 每种变体的活性计算为转化百分比，其通过将通过如HPLC分析确定的所有产物峰的面积除以底物和各种胰岛素产物的面积之和来定量。

[0293]

表13.1

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于SEQ ID NO: 1262)	2-(4-羟苯基)乙酰胺[1]	2-(4-羟苯基)乙酸甲酯[1]
1287/1288	F71G;G74D	+++	+++
1007/1008	F24A;Y31W;V56I;F701W	++	+
1019/1020	F24A;Y31W;F71C;F701W	++	+
1023/1024	F24A;V56T;F701W	+	++
1031/1032	F24A;Y31W;V56I;F701Y	+	++

[1]转化百分比如下定义："+"=转化大于1%但小于10%；"++"=转化大于10.0％但小于20%；"+++"=转化大于20%。

[0294] 实施例14

[0295] 胰岛素及其酰化产物的分析检测

[0296] 使用表14.1、14.2、14.3、14.4和14.5中的分析方法收集实施例5-12 中描述的数据。本文提供的方法均可以用于分析使用本发明产生的变体。图1中显示的结果对应于这些方法的化合物的洗脱顺序。

[0297]

表 14.1 分析方法

仪器	Thermo LXQ
柱	Waters Xbridge C18 柱： 50 × 3.0 mm, 5 um, 带有 Phenomenex C18 guard 筒： 5 × 3.0 mm, 5 µm
流动相	梯度(A: 在水中的 0.2%甲酸； B: 在 MeCN 中的 0.2%甲酸) 时间(min) %A 0.0 78 1.0 78 4.0 68 5.0 5.0 6.0 78 7.0 78
流速	0.7 mL/min
运行时间	7 min
柱温度	45°C
注入体积	10 µL
MS 检测	LXQ； 在 0-0.5 min 之间转移来自 MS 的流量。 BP 提取离子，用于: 胰岛素产物 = 968.6, 1162.6, 1453.0, 1937.0 单-胰岛素产物 = 1189.5, 1486.5, 1981.5 二-胰岛素产物 = 1216.0, 11519.8 三-胰岛素产物 = 1243.0, 1553.5
MS 条件	MS 极性： 正； 电离： ESI； 模式： Q1 扫描从 300-2000； 源电压：

[0298]

表 14.1 分析方法	
	20；鞘气(Sheath gas)：40；辅助气(Aux gas)：10；加热毛细管温度(Cap temp)：350；离子喷雾电压：5000；Cap V：5；镜筒透镜：55。

表 14.2 分析方法	
仪器	Agilent HPLC 1200 系列
柱	Ascentis Express C18, 4.6 × 100 或 150 mm, 2.7 uM
流动相	梯度 I (A：在水中的 0.05% TFA；B：在 MeCN 中的 0.05% TFA) 时间(min) %A 0.0 95 0.1 70 8、8.5 或 9 50 8.1、8.6 或 9.1 5 8.2、8.7 或 9.2 95 9、9.2 或 9.5 95 梯度 II ((A：在水中的 0.05% TFA；B：在 MeCN 中的 0.05% TFA) 时间(min) %A 0 70 7 或 8 50 7.1 或 8.1 70 9 或 10 70
流速	1.0 mL/min
运行时间	~10 min
产物洗脱顺序	胰岛素；A1-酰化的胰岛素；B 29-酰化的胰岛素；B1-酰化的胰岛素；二-A1、B29-酰化的胰岛素；二-A1、B1-酰化的胰岛素；二-B1、B29-酰化的胰岛素；三-A1、B1、B29-酰化的胰岛素
柱温度	40°C
注入体积	5 µL
检测	UV 218 nm 和 280 nm 检测器：MWD (Agilent 1200 系列)；狭缝=4nm；峰宽 > 0.1min；参考 = 360; BW = 8

表 14.3 分析方法	
仪器	Agilent HPLC 1290 系列
柱	Waters Cortecs UPLC C18 2.1 × 50 mm, 1.6 uM
流动相	梯度(A：在水中的 0.05% TFA；B：在 MeCN 中的 0.05% TFA)用于样品 时间(min) %A 0.0 72 2.5 50 2.51-2.7 10 2.71 72

[0299]

表 14.3 分析方法	
	3 72 梯度(A: 在水中的 0.05% TFA; B: 在 MeCN 中的 0.05% TFA)用于洗涤 时间(min) %A 0.0 72 1.7 0 1.71-2 72
流速	0.9 mL/min
运行时间	3 min
产物洗脱顺序	胰岛素; A1-酰化的胰岛素; B 29-酰化的胰岛素; B1-酰化的胰岛素; 二-A1、B29-酰化的胰岛素; 二-A1、B1-酰化的胰岛素; 二-B1、B29-酰化的胰岛素; 三-A1、B1、B29-酰化的胰岛素
柱温度	40°C
注入体积	0.5 µL
检测	UV 218 nm 和 280 nm 检测器: MWD (Agilent 1290 系列); 狭缝=4 nm; 峰宽 > 0.1 min; 参考 = 360; BW = 8

表 14.4 分析方法	
仪器	Agilent HPLC 1200 系列
柱	Ascentis Express C18, 4.6 × 100 或 150 mm, 2.7 uM
流动相	梯度 I (A: 在水中的 0.05% TFA; B: 在 MeCN 中的 0.05% TFA) 时间(min) %A 0.0 95 0.1 70 5、6 或 8 50 5.5、6.5 或 8.1 5 5.7 或 6.7 5 5.8、6.8 或 8.2 95 6、7 或 9 95
流速	1.0 mL/min
运行时间	~10 min
产物洗脱顺序	胰岛素; A1-酰化的胰岛素; B 29-酰化的胰岛素; B1-酰化的胰岛素; 二-A1、B29-酰化的胰岛素; 二-A1、B1-酰化的胰岛素; 二-B1、B29-酰化的胰岛素; 三-A1、B1、B29-酰化的胰岛素
柱温度	40°C
注入体积	5 µL
检测	UV 218 nm 和 280 nm 检测器: MWD (Agilent 1200 系列); 狭缝=4nm; 峰宽 > 0.1min; 参考 = 360; BW = 8

[0300]

表 14.5 分析方法	
仪器	Agilent HPLC 1290 系列
柱	Waters Cortecs UPLC C18 2.1 × 50 mm, 1.6 uM
流动相	梯度(A: 在水中的 0.05% TFA; B: 在 MeCN 中的 0.05% TFA)用于样品 时间(min) %A 0.0 72 1.5 50 1.51-1.7 10 1.71 72 2 72
流速	0.9 mL/min
运行时间	3 min
产物洗脱顺序	胰岛素; A1-酰化的胰岛素; B 29-酰化的胰岛素; B1-酰化的胰岛素; 二-A1、B29-酰化的胰岛素; 二-A1、B1-酰化的胰岛素; 二-B1、B29-酰化的胰岛素; 三-A1、B1、B29-酰化的胰岛素
柱温度	40°C
注入体积	0.5 µL
检测	UV 218nm 和 280nm 检测器: MWD (Agilent 1290 系列); 狭缝=4nm; 峰宽 > 0.1min; 参考 = 360; BW = 8

[0301] 为了所有目的,本申请中引用的所有出版物、专利、专利申请和其他文件通过引用特此以其整体并入,其程度如同每个单独的出版物、专利、专利申请或其他文件被单独地指出通过引用为了所有目的并入一样。

[0302] 虽然已经说明和描述多个具体的实施方案,但是将理解可以进行各种变化,而不脱离本发明的精神和范围。

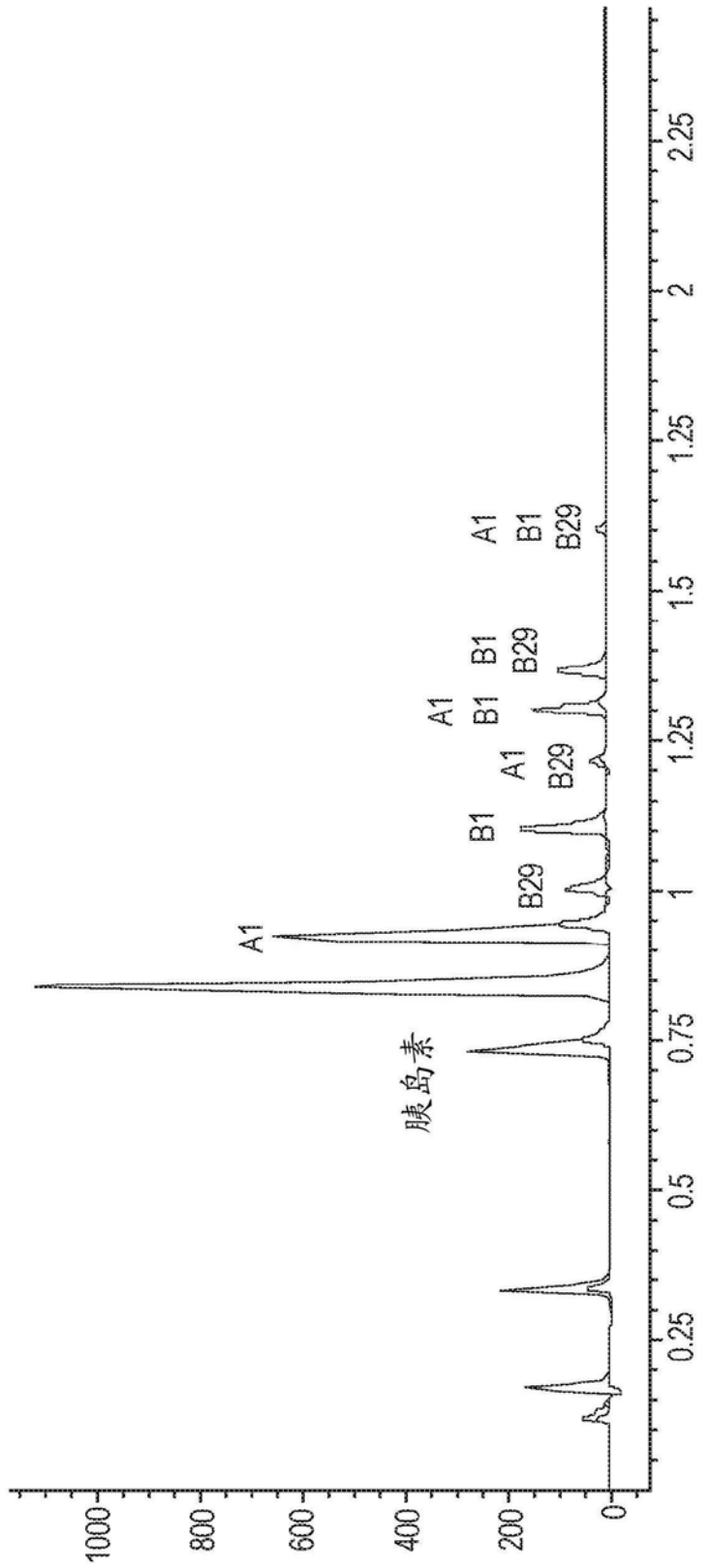

1000
800
600
400
200
0
0.25
0.5
0.75
1
1.25
1.5
1.25
2
2.25
胰岛素
A1
B29
B1
A1
B29
A1
B1
B1
B29
A1
B1
B29

图1

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于 SEQ ID NO: 1262)	A1		B29		B1		A1/B29		A1/B1		B1/B29	
		% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	%conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]
1263/1264	V28A;G74D;Q547K;	+	+++++		+								
953/954	W431R;	+	+++++		+								
955/956	F701L;	+	+++++										
957/958	F701A;	+	+++++										
1265/1266	F701Y;	+	+++++		+								
959/960	F24A;	+	+++++		+								+
1267/1268	F71L;	+	+++++		+				+				
961/962	F71V;	+	+++++				+						
1269/1270	F71C;	+	+++++		+								
963/964	F71K;	+	++++		+				+				
1271/1272	F71R;		+++++										
965/966	F71E;	+	+++++		+								
967/968	F24Y;V28A;	+	+++++		+				+				
969/970	M697L;	+	+++++		+								

图2A

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于 SEQ ID NO: 1262)	A1		B29		B1		A1/B29		A1/B1		B1/B29	
		% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	%conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]
1273/1274	V56T;	+	+++++		+								
1275/1276	V56I;	+	+++++		+								
1277/1278	V56L;	+	+++++		+								
971/972	I177T;	+	+++++		+								
1281/1282	W154F;	+	+++++		+								
973/974	W154F;L754P;	+	+++++		+								
975/976	F71C;G74D;	+	++++	+	+				+				
1285/1286	F71Q;G74D;P164S;	+	++++	+	+				+				
1287/1288	F71G;G74D;	++	+++	+	+	+	+			+	+		
979/980	F701H;	+	+++++										
981/982	F701C;	+	+++++		+								
983/984	F701I;	++	+++++										
985/986	F701M;	+	+++++		+				+				
1289/1290	F701W;	+	+++++		+								

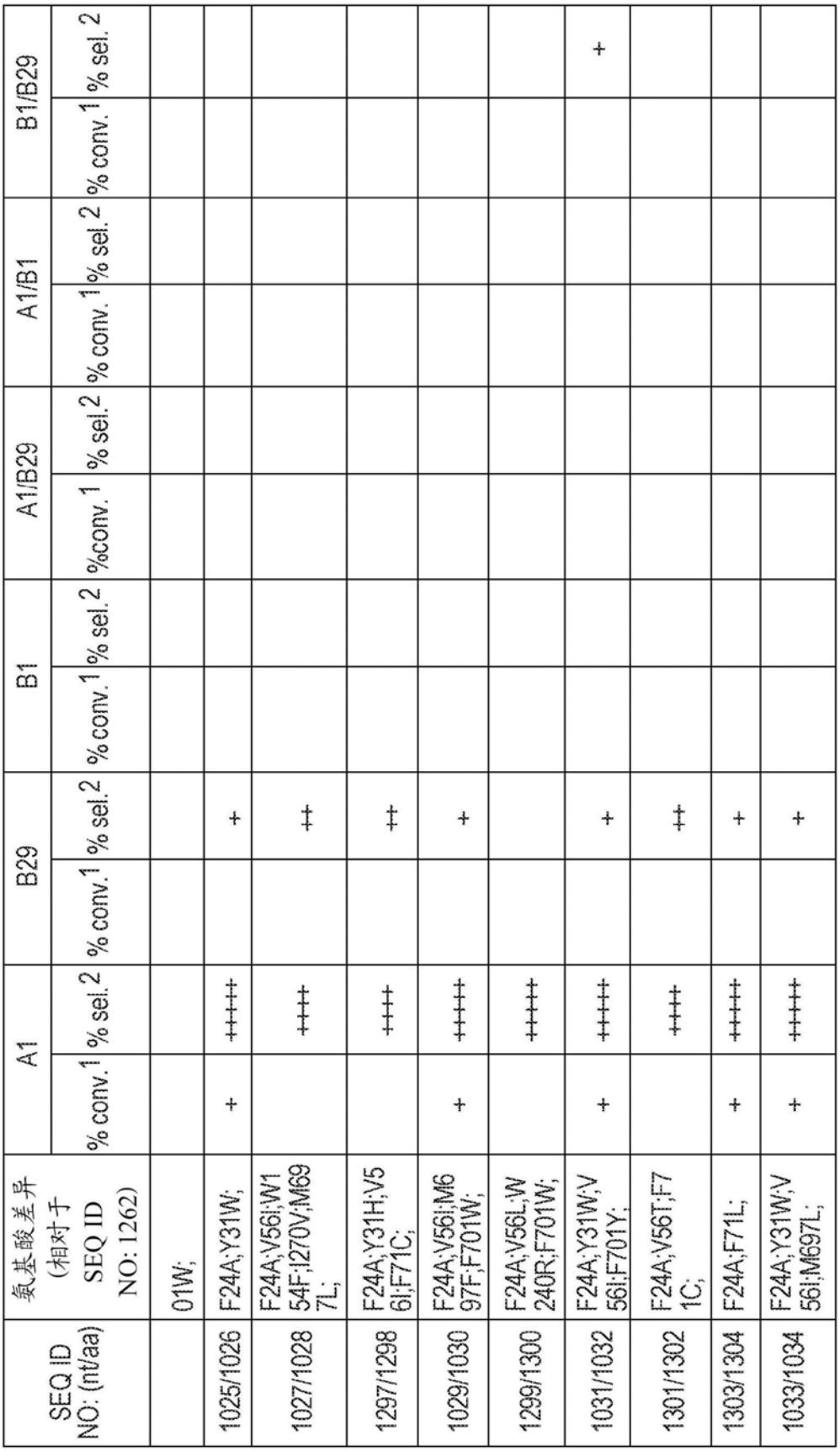
图2B

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于 SEQ ID NO: 1262)	A1		B29		B1		A1/B29		A1/B1		B1/B29	
		% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	%conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]
1291/1292	F701V;	+	+++++						+				
987/988	M697F;	+	+++++		+				+				
989/990	M697G;	+	+++++		+				+				+
991/992	V28A;Y31F;		++++		++								
993/994	V28A;Y31L;	+	+++++		+								
995/996	V28A;Y31V;	+	++++		++								
997/998	V28A;Y31T;	+	+++++		+								
999/1000	V28A;Y31C;	+	+++++		+								
1001/1002	V28A;Y31N;		+++		+++								
1003/1004	V28A;Y31M;	+	+++++										
1005/1006	V28A;Y31K;		++++		++								
1007/1008	F24A;Y31W;V56I;F701W;	+	+++++										
1009/1010	F24A;V56I;F701W;	+	+++++						+				
1011/1012	F24A;Y31C;V56I;S386P;F701												

图2C

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于 SEQ ID NO: 1262)	A1		B29		B1		A1/B29		A1/B1		B1/B29	
		% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	%conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]
	Y;												
1013/1014	F24A;F71C;F7 01W;	+	+++++		+								
1293/1294	F24A;P29S;V5 6L;T352S;N483 S;		++++		++								
1261/1262		+	+++++		+								
1015/1016	F24A;L225V;F 701Y;	+	+++++										+
1017/1018	F24A;Y31W;V 56T;V264A;F7 01W;S750G;		++++		++								
1019/1020	F24A;Y31W;F7 1C;F701W;	+	+++++		+								
1295/1296	F24A;Q340R;V 391G;		++++		++								
1021/1022	F24A;D484N;F 701W;	++	+++++										
1023/1024	F24A;V56T;F7	++	+++++										

图2D



SEQ ID NO: (nt/aa)	氨基酸差异 (相对于 SEQ ID NO: 1262)	A1		B29		B1		A1/B29		A1/B1		B1/B29	
		% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	%conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]
	01W;												
1025/1026	F24A;Y31W;	+	++++		+								
1027/1028	F24A;V56I;W154F;I270V;M697L;		++++		++								
1297/1298	F24A;Y31H;V56I;F71C;		++++		++								
1029/1030	F24A;V56I;M697F;F701W;	+	++++		+								
1299/1300	F24A;V56L;W240R;F701W;		++++										
1031/1032	F24A;Y31W;V56I;F701Y;	+	++++		+								+
1301/1302	F24A;V56T;F71C;		++++		++								
1303/1304	F24A;F71L;	+	++++		+								
1033/1034	F24A;Y31W;V56I;M697L;	+	++++		+								

图2E

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于 SEQ ID NO: 1262)	A1		B29		B1		A1/B29		A1/B1		B1/B29	
		% conv.1	% sel.2	% conv.1	% sel.2	% conv.1	% sel.2	%conv.1	% sel.2	% conv.1	% sel.2	% conv.1	% sel.2
1035/1036	F24Y;V28A;F71C;F701W;	+	+++++		+				+				
1037/1038	F24Y;V28A;V56I;A308T;T379A;F701W;	+	+++++		+				+				
1039/1040	F24Y;V28A;D321N;F701H;	+	+++++		+								
1041/1042	F24Y;V28A;N457Y;F701W;	++	+++++										
1043/1044	F24Y;V28A;V56I;F701W;	+	+++++										
1045/1046	F24Y;Y27C;V28T;F701W;V729F;	+	+++++		+								
1305/1306	V56L;I77T;F701W;	+	+++++										
1047/1048	V56T;F71G;F701W;	++	+++++										
1049/1050	M697L;F701Y;	+	+++++		+								

图2F

SEQ ID NO: (nt/aa)	氨基酸差异 (相对于 SEQ ID NO: 1262)	A1		B29		B1		A1/B29		A1/B1		B1/B29	
		% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	%conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]
1307/1308	W26C;F701W;		++++		+								
1051/1052	V56T;W658R;F701W;	+	+++++		+								
1053/1054	V56T;F71G;F701I;	+	+++++		+								
1055/1056	A410P;M697F;F701W;	+	+++++		+								+
1057/1058	V56L;K322R;M697L;F701V;	+	+++++										
1059/1060	V56L;F701H;	+	+++++		+								
1061/1062	Y31C;V56L;F701H;	+	+++++						+				
1063/1064	F71G;F701W;	++	+++++										
1065/1066	T129A;I511F;F701Y;		++++		+								
1309/1310	D501N;F701W;	+	+++++		+								
1067/1068	V56T;F701W;L711Q;	+	+++++		+								

图2G

SEQ ID NO: (nt/aa)	氨基酸差异（相对于SEQ ID NO: 1262）	A1		B29		B1		A1/B29		A1/B1		B1/B29	
		% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	%conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]	% conv.[1]	% sel.[2]
1069/1070	I423T;F701W;	+	+++++		+								
1071/1072	V56I;W154F;F701W;	+	+++++		+								
1073/1074	V56L;W119R;K146E;F701Y;		++++		+								
1075/1076	V56T;M697L;F701W;	+	+++++										
1311/1312	V56T;F701W;	+	+++++		+								

[1] 转化百分比如下定义：“+”=转化大于1%但小于10%；“++”=转化大于10.0 %但小于20%；“+++”=转化大于20%。

[2] 选择性百分比如下定义：“+”=选择性大于5%但小于20%；“++”=选择性大于20 %但小于40%；“+++”=选择性大于40%但小于60%；“++++”=选择性大于60%但小于80%；“+++++”=选择性大于80%。

图2H