

一种非天然氨基酸及其应用、包含其的重组蛋白以及重组蛋白偶联物

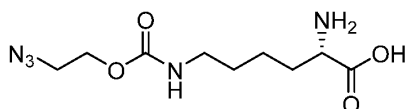
技术领域

本发明涉及生物制药领域，具体涉及一种非天然氨基酸、包含所述非天然氨基酸的重组蛋白以及所述重组蛋白形成的偶联物。

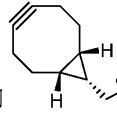
背景技术

通过向蛋白质中引入含有特殊基团的非天然氨基酸，可以实现多种科学研究与产品开发应用，例如，向蛋白质中引入光敏非天然氨基酸，或者对非天然氨基酸进行特殊标记，便于研究蛋白质之间的相互作用；又例如，利用非天然氨基酸进行酶的定向改造，以提高酶活、酶的稳定性，或者便于进行酶的高效固定化；再例如，利用在常规宿主中不可实现非天然氨基酸插入的特点，可制备安全的活菌或活病毒疫苗。使用密码子扩展技术向蛋白质中引入非天然氨基酸的一项重要应用在于对蛋白质进行定点修饰，改变蛋白质的功能、稳定性、半衰期等特征，并可用于创新型生物药的开发。目前已有一系列成果，如 PEG 定点偶联重组人生长激素制备长效重组蛋白、小分子毒素定点偶联单抗开发抗体偶联药物等。由此可见，非天然氨基酸具有非常重要的作用和非常广泛的应用领域。

现有技术（例如 CN 102838671B、CN 106146663A、*J. Am. Chem. Soc.* **2009**, 131, 8720 等）公开了一种非天然氨基酸 Lys-azido，结构式如下所示：



20

Lys-azido 末端的叠氮结构(-N₃)能与含炔烃结构（如 BCN，即 ) 修饰的载体药物（如 PEG 等）进行化学连接得到偶联物（例如，中国专利 CN 103153927B），具有很高的特异选择性。但是，该种偶联方法和化学修饰方法需引入成本较高的炔烃类结构，且使用当量较大时才可获得可接受的药物抗体偶联比率，由此增加了相应的生产成本，工艺过程也较为复杂，工艺条件苛刻。

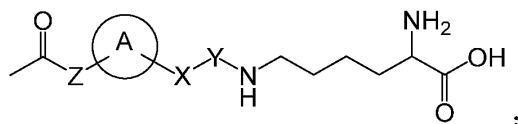
因此，需要开发一种结构新颖、制备简便、成本低廉的非天然氨基酸，以扩大氨基酸的种类以及应用。

发明内容

为克服现有技术中存在的上述不足，本发明的一个目的是提供一种非天然氨基酸。本发明的另一个目的是提供所述非天然氨基酸的应用。

本发明的还一个目的是提供一种重组蛋白以及一种重组蛋白偶联物。

本发明提供的非天然氨基酸为具有如式(I)所示结构的化合物或其对映异构体，



I

- 5 其中，X 和 Z 各自独立地表示取代或未取代的 C0~C20 直链或支链亚烷基，其中的一个或多个 -CH₂- 可任选地替换为 -O-、-S-、-NH-、-C(O)-、-S(O)- 中的一个或多个，Y 表示 -C(O)-、-S(O)- 或 -CH₂-，以及 A 表示取代或未取代的 C6~C20 芳基；

- 当所述 X、Z 和 A 各自独立地表示取代的基团时，取代基可以选自羟基、巯基、卤素、硝基、氰基、烷基、烯基、炔基、烷氧基、酰基、酰胺基、羧基、酯基、氨基、磺酰基、
10 亚磺酰基、环烷基、杂环基、芳基、杂芳基中的一种或多种。

如实施例 11 所示，本发明的发明人发现，除了成本昂贵、工艺复杂之外，Lys-azido 末端的叠氮结构 (-N₃) 在插入到蛋白中时容易被还原为氨基结构 (-NH₂)，从而失去偶联的活性，因此该还原反应降低了工艺制备过程中的收率。



- 15 本发明提供的非天然氨基酸在其末端引入了羰基作为活性反应基团，不仅结构新颖、制备简便，而且偶联条件温和，生产成本较低，插入蛋白序列时不易发生结构变化而导致反应活性丧失；本发明提供的非天然氨基酸中还包含与端基羰基相连的芳基，芳基的引入可进一步增强所得偶联物的稳定性，偶联物即使在较低的 pH 条件下也不容易发生分解。此外，本发明提供的非天然氨基酸中还包含有一定链长的亚烷基，化合物的柔性较好，更
20 容易形成各种偶联物。

- 本发明提供的非天然氨基酸中，“C0~Cn”包括 C0~C1、C0~C2、.....C0~Cn，当表示 C0 时，意为该基团不存在，其两端的 C 原子直接连接成键。举例而言，所述“C0~C6”基团是指该部分中具有 0~6 个碳原子，即基团不存在、包含 1 个碳原子、2 个碳原子、3 个碳原子、4 个碳原子、5 个碳原子或 6 个碳原子；所述“C6~C10”基团是指该部分中具有
25 6~10 个碳原子，即包含 6 个碳原子、7 个碳原子、8 个碳原子、9 个碳原子或 10 个碳原子。

本发明提供的非天然氨基酸中，“芳基”是指含有一个或者两个环的碳环芳香系统，其中所述环可以以稠合的方式连接在一起。“芳基”包括单环或双环的芳基，比如苯基、萘基、

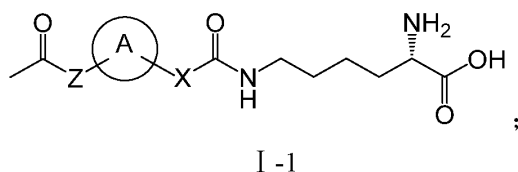
四氢萘基的芳香基团。优选芳基为 C6~C10 芳基,更优选芳基为苯基和萘基,最优选为苯基。

在根据本发明的一些优选的实施方式中,所述取代基可以选自羟基、巯基、卤素、硝基、氰基、C1~C6 烷基、C1~C6 烷氧基、酰基、酰胺基、羧基、酯基、氨基、磺酰基、亚磺酰基、C3~C8 环烷基、C3~C8 杂环基、C6~C20 芳基、C4~C10 杂芳基中的一种或多种。

在根据本发明的一些优选的实施方式中,所述 X 和 Z 可以各自独立地表示 C0~C10 直链或支链亚烷基,其中的一个或多个-CH₂-可任选地替换为-O-、-S-、-NH-中的一个或多个;在根据本发明的一些更优选的实施方式中,所述 X 和 Z 可以各自独立地表示 C0~C6 直链亚烷基,其中的一个或多个-CH₂-可任选地替换为-O-、-S-、-NH-中的一个或多个;在根据本发明的一些更优选的实施方式中,所述 X 和 Z 不同时为 C0 的亚烷基,即 X 和 Z 基团不能同时不存在。

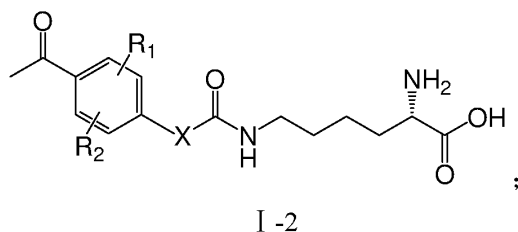
在根据本发明的一些优选的实施方式中,所述 A 可以表示取代或未取代的 C6~C10 芳基,更优选地,所述 A 可以表示取代或未取代的苯基或萘基。

在根据本发明的一些优选的实施方式中,所述非天然氨基酸可以为具有如式(I-1)所示结构的化合物,



其中,所述 X、Z 和 A 各自独立地如上述技术方案任一项所定义。

在根据本发明的一些优选的实施方式中,所述非天然氨基酸可以为具有如式(I-2)所示结构的化合物,

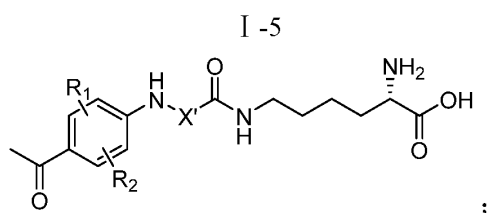
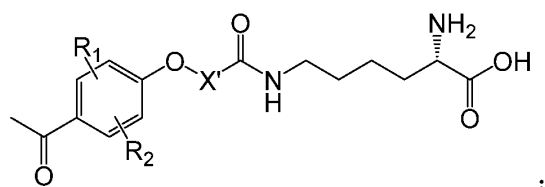
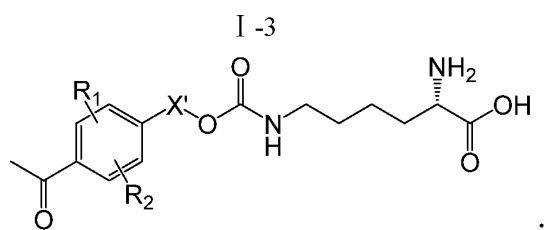
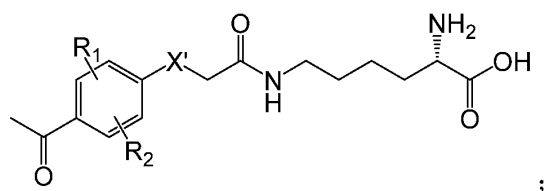


其中,所述 X 如上述技术方案任一项所定义;

R₁ 和 R₂ 各自独立地表示氢、羟基、巯基、卤素、硝基、氰基、C1~C6 烷基、C1~C6 烷氧基、酰基、酰胺基、羧基、酯基、氨基、磺酰基、亚磺酰基、C3~C8 环烷基、C3~C8 杂环基、C6~C20 芳基或 C4~C10 杂芳基。

在根据本发明的一些较优选的实施方式中,所述非天然氨基酸为具有如式(I-3)、式

(I-4)、式(I-5)或式(I-6)所示结构的化合物，



其中，X'表示 C0~C6 的直链亚烷基，更优选表示 C0~C4 的直链亚烷基，其中的一个或多个-CH₂-可任选地替换为-O-和/或-NH-；

所述 R₁ 和 R₂ 各自独立地如上述技术方案任一项所定义。

本发明提供的非天然氨基酸包括光学纯的对映异构体和外消旋体。

在根据本发明的一些更优选的实施方式中，本发明提供的非天然氨基酸为具有以下结构之一的化合物：

(1) NBOK	
----------	--

(2) NPAK	
(3) NBPk	
(4) NPOK	
(5) NBGK	
(6) NPOK-2	
(7) NBGK-2	

本发明还提供了上述技术方案任一项所述的非天然氨基酸在制备重组蛋白或重组蛋白偶联物中的应用。

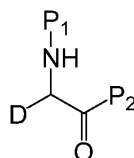
本发明所述的应用中，重组蛋白可以是如上述技术方案任一项所述的非天然氨基酸以任意位点、任意数量插入至本领域常见的任意种类的蛋白之中而得到重组蛋白。重组蛋白偶联物可以是所得的任意重组蛋白与本领域常见的偶联部分偶联后所得的偶联物，其中，偶联部分可以包括但不限于聚合物（例如，任意分子量的聚乙二醇）、蛋白、多肽或小分子药物中的一种或多种。

在根据本发明的一些优选的实施方式中，重组蛋白可以为重组人生长激素，重组蛋白

偶联物可以为重组人生长激素-聚乙二醇偶联物。

本发明还提供了一种重组蛋白，其中，所述重组蛋白的氨基酸序列中的至少一个位点上为上述技术方案任一项所述的非天然氨基酸。

进一步地，所述重组蛋白可以具有如式(II)所示的结构，



II

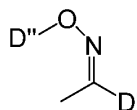
式(II)中，D表示如上述技术方案任一项所述的非天然氨基酸除去氨基酸部分的残基，P₁、P₂分别表示所述非天然氨基酸的氨基和羧基在所述氨基酸序列中的连接部分。

本发明提供的重组蛋白可以使用本领域常见的制备方法制备得到，包括使用基因密码子扩展技术实现含非天然氨基酸的重组蛋白的克隆与表达。

本发明提供的重组蛋白可以是如上述技术方案任一项所述的非天然氨基酸以任意位点、任意数量插入至本领域常见的任意种类的蛋白之中而得到重组蛋白，例如重组人生长激素。

本发明还提供了一种重组蛋白偶联物，其中，通过上述技术方案任一项所述的重组蛋白中的非天然氨基酸的端基羰基与含有“NH₂-O-”端基的偶联部分形成脒键而形成。

进一步地，所述重组蛋白偶联物可以具有如式(III)所示的结构，



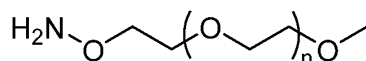
III

式(III)中，D'表示如上述技术方案任一项所述的重组蛋白除去非天然氨基酸的端基羰基部分的残基，D''表示偶联部分除去“NH₂-O-”端基的残基。

本发明提供的重组蛋白偶联物中，偶联部分可以包括聚合物（例如，任意分子量的聚乙二醇）、蛋白、多肽或小分子药物中的一种或多种。不同种类的偶联部分可以单独与重组蛋白偶联，也可以不同种类的偶联部分先形成连接后再与重组蛋白偶联。

在根据本发明的一些优选的实施方式中，所述重组蛋白可以为重组人生长激素，所述含有“NH₂-O-”端基的偶联部分可以为含有“NH₂-O-”端基的聚乙二醇。

在根据本发明的一些更优选的实施方式中，含有“NH₂-O-”端基的聚乙二醇具有如下结构式：



其中, 含有“NH₂-O-”端基的聚乙二醇的分子量可以为 10~100 KD, 包括但不限于约 10 KD、20 KD、30 KD、40 KD、50 KD、60 KD、70 KD、80 KD、90 KD、100 KD 等分子量值或任意组合的分子量区间。优选地, 含有“NH₂-O-”端基的聚乙二醇的分子量可以为 20~50 KD。

5 在根据本发明的一些更优选的实施方式中, 所述重组人生长激素的氨基酸序列如 SEQ ID NO: 1 所示; 进一步优选地, 对应于所述氨基酸序列 SEQ ID NO: 1 中的第 107 位点上如上述技术方案任一项所述的非天然氨基酸。

当重组蛋白为重组人生长激素时, 本发明还提供了上述技术方案任一项所述的重组蛋白偶联物在制备用于治疗内源性生长激素分泌不足所引起的生长发育障碍、特纳综合征引
10 起的生长发育障碍或成人生长激素缺乏症的药物中的用途。

本发明提供的技术方案具有以下优点:

(1) 本发明提供的非天然氨基酸在其结构中引入了端基羰基以及与之相连的芳基, 相对于目前的端基叠氮基非天然氨基酸 (例如, Lys-azido), 制备更简便, 安全性更好, 插入蛋白时不易失活, 与偶联部分的结合率更高, 所得偶联物的稳定性也更高。

15 (2) 本发明提供的非天然氨基酸作为一种氨基酸衍生物, 本身可具备氨基酸的性质, 由此扩大了氨基酸的潜在种类, 使其可作为氨基酸衍生物应用至众多领域之中, 尤其是在制备重组蛋白或重组蛋白偶联物中。

(3) 本发明提供的非天然氨基酸可以在原核表达系统和真核表达系统中被成功识别并插入至蛋白中, 从而产生特定位点含有非天然氨基酸的蛋白, 由此可形成具有不同理化性质、生化活性的重组蛋白, 扩大了蛋白的种类和潜在的应用范围。此外, 本发明的非天然
20 氨基酸在蛋白中具有较高的表达效率, 实用性更好。

(4) 本发明提供的重组蛋白由于含有本发明的非天然氨基酸, 其所含的末端羰基活性基团可方便地形成蛋白偶联物 (或缀合物), 例如聚乙二醇偶联物、聚乙二醇-活性药物偶联物等。这些蛋白偶联物由于经过了性能改造, 因而可具有多种改进的生物活性 (例如抗
25 肿瘤活性)。

(5) 本发明还提供了一种新型的蛋白偶联物平台, 通过蛋白质所含的新型非天然氨基酸以及所连接的偶联部分, 可实现多种蛋白、多种偶联部分的结合。

附图说明

30 图 1 为实施例 8 中重组人生长激素表达质粒 pET21-rhGH107 的图谱。

图 2 为实施例 8 所得的添加不同种类非天然氨基酸获得的发酵产物 SDS-PAGE 电泳图, 其中各泳道分别表示如下: 泳道 1: 蛋白分子量 Marker; 泳道 2: 野生型重组人生长激素; 泳道 3: 投料 NBOK 的重组人生长激素表达产物; 泳道 4: 投料 NPAK 的重组人生长激素表达产物; 泳道 5: 投料 NBPK 的重组人生长激素表达产物; 泳道 6: 投料 NBGK

的重组人生长激素表达产物；泳道 7：投料 NPOK 的重组人生长激素表达产物；泳道 8：投料 NPOK-2 的重组人生长激素表达产物；泳道 9：投料 NBGK-2 的重组人生长激素表达产物；泳道 10：不添加任何非天然氨基酸的重组人生长激素表达产物。

图 3 为实施例 8 中的含非天然氨基酸的重组人生长激素与 PEG 偶联产物的 SDS-PAGE 电泳图，其中各泳道分别表示如下：泳道 1：分子量 Marker；泳道 2：野生型重组人生长激素；泳道 3：含 NBOK 的重组人生长激素与 PEG 偶联产物；泳道 4：含 NPAK 的重组人生长激素与 PEG 偶联产物；泳道 5：含 NBPK 的重组人生长激素与 PEG 偶联产物；泳道 6：含 NBGK 的重组人生长激素与 PEG 偶联产物；泳道 7：含 NPOK 的重组人生长激素与 PEG 偶联产物；泳道 8：含 NPOK-2 的重组人生长激素与 PEG 偶联产物。

图 4 为实施例 8 中经纯化的 PEG 偶联重组人生长激素的 SDS-PAGE 电泳图，其中各泳道分别表示如下：泳道 1：分子量 Marker；泳道 2：重组人生长激素标准品；泳道 3：含 NPOK 的重组人生长激素与 PEG 偶联产物；泳道 4：含 NBOK 的重组人生长激素与 PEG 偶联产物。

图 5A-图 5G 为实施例 8 中的细胞活性曲线图，其中，图 5A 为中检院 rhGH 标准品的细胞活性曲线图；图 5B 为赛增®的细胞活性曲线图；图 5C 为金赛增®的细胞活性曲线图；图 5D 为 rhGH (NPOK) 的细胞活性曲线图；图 5E 为 PEG-rhGH (NPOK) 的细胞活性曲线图；图 5F 为 rhGH (NBOK) 的细胞活性曲线图；图 5G 为 PEG-rhGH (NBOK) 的细胞活性曲线图。

图 6 为实施例 9 中 CHO 细胞瞬时表达插入非天然氨基酸的荧光显微图像。

图 7 为实施例 10 中表达质粒 pCDNA3.1-Trastuzumab-UAG142 的图谱。

图 8A-图 8C 分别为实施例 10 中含 NBPK 曲妥珠单抗、含 NBOK 曲妥珠单抗以及含 NPOK-2 曲妥珠单抗与毒素偶联后的 HIC-HPLC 谱图。

图 9 为实施例 10 中含 NBOK 的曲妥珠单抗和 DM1 修饰的含 NBOK 曲妥珠单抗分别对 BT-474 细胞的抑制效果图。

图 10A、图 10B 分别为实施例 11 中第 140 位突变为 Lys-azido 的 rhGH 和第 140 位突变为 NBOK 的 rhGH 的质谱图。

图 11A、图 11B 分别为实施例 11 中第 140 位突变为 Lys-azido 的 rhGH 和第 140 位突变为 NBOK 的 rhGH 分别与 30KD PEG 偶联过程的 SDS-PAGE 电泳图。其中图 11A 各泳道分别表示如下：泳道 1：分子量 Marker；泳道 2：野生型重组人生长激素；泳道 3：rhGH-Lys-azido-140；泳道 4：rhGH-Lys-azido-140：30K BCN-PEG 为 1:15 (摩尔比，下同) 偶联反应 72 h 的产物；泳道 5：rhGH-Lys-azido-140：30K BCN-PEG 为 1:25 偶联反应 72 h 的产物；图 11B 各泳道分别表示如下：泳道 1：分子量 Marker；泳道 2：rhGH-NBOK-140；泳道 3：rhGH-NBOK-140：30K BCN-PEG 为 1:15 偶联反应 6 h 的产物；泳道 4：rhGH-NBOK-140：30K BCN-PEG 为 1:15 偶联反应 9 h 的产物；泳道 5：rhGH-NBOK-140：

30K BCN-PEG 为 1:15 偶联反应 12 h 的产物；泳道 6：rhGH-NBOK-140：30K BCN-PEG 为 1:15 偶联反应 24 h 的产物；泳道 7：rhGH-NBOK-140：30K BCN-PEG 为 1:15 偶联反应 48 h 的产物。

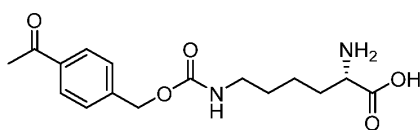
5 具体实施方式

以下结合具体实施例对本发明的技术方案做进一步详细说明。

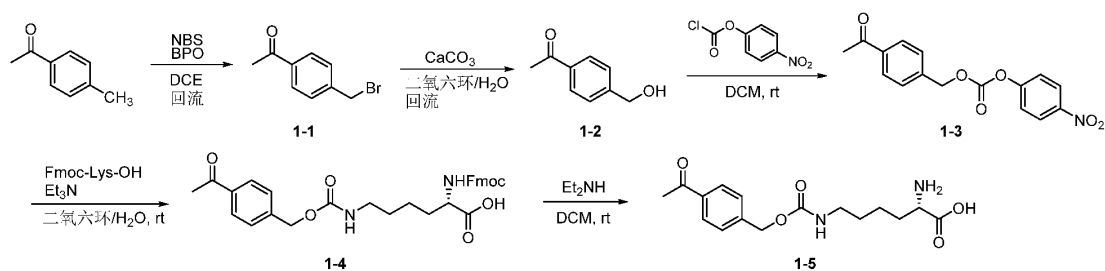
本发明的实施例中所使用的试剂或原料如无特别说明均为商购产品。

实施例 1 非天然氨基酸 NBOK 的制备

10 NBOK 的结构式如下所示：



反应过程如下图所示：



15 制备过程包括以下步骤：

a) 在反应烧瓶中，加入对甲基苯乙酮(4.0 mL, 30.0 mmol)，加入溶剂 DCE (50.0 mL)，加入 NBS (6.41 g, 36.0 mmol)和 BPO (0.05 g, 0.3 mmol)，混合物在 80℃下回流 24 小时后，容器放入冰水中冷却，析出固体，过滤除去固体，用饱和 Na_2CO_3 洗涤 3 次，用 DCM 萃取 3 次，合并有机相，加入无水硫酸钠干燥，过滤，减压浓缩后得到粗产物 **1-1** (5.46 g, 收率 85%)，无需纯化直接用于下一步。

b) 在反应烧瓶中，加入产物 **1-1** (2.73 g, 12.80 mmol)，加入溶剂二氧六环(40 mL)和水 (40 mL)，再加入碳酸钙(7.68 g, 76.8 mmol)，混合物在 105℃回流 24 小时，冷却至室温，过滤除去固体，用 DCM 萃取 3 次，合并有机相，减压浓缩，经柱层析(洗脱剂：PE:EA=3:1)纯化后得到产物 **1-2** (1.80 g, 收率 94%)。

25 c) 在两口反应烧瓶中，加入对硝基苯基氯甲酸酯(2.90 g, 14.4 mmol)，加入溶剂 DCM (10.0 mL)，降温至 0℃，加入产物 **1-2** (1.80 g, 12.0 mmol)和吡啶(1.2 mL, 14.4 mmol)，室温搅拌 18 小时后，反应液中加入饱和碳酸钠溶液(10 mL)，用 DCM (50 mL)萃取 3 次，合并有机相，用水洗涤 2 次，无水硫酸钠干燥，过滤，减压浓缩，经柱层析(洗脱剂：PE:EA=5:1)

纯化得到产物 **1-3** (3.14 g, 收率 83%)。

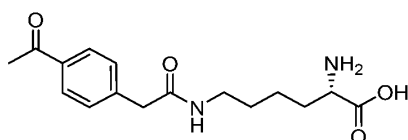
d) 在反应烧瓶中, 加入产物 **1-3** (1.26 g, 4.0 mmol) 和 Fmoc-Lys-OH 盐酸盐 (1.40 g, 3.33 mmol), 加入溶剂二氧六环 (15 mL) 和水 (5 mL), 再加入三乙胺 (1.2 mL, 8.3 mmol), 混合物在室温下反应 24 小时后, 加入适量 1M HCl 溶液, 用 DCM 萃取, 减压浓缩, 得到粗产物 **1-4**, 直接用于下一步。

e) 在反应烧瓶中, 将产物 **1-4** (1.10 g, 0.19 mmol) 溶于 DCM (10 mL) 中, 加入二乙胺 (5.0 mL), 在室温下反应 6 小时, 产物析出, 过滤后用 DCM 打浆 3 次, 即得目标产物 NBOK (**1-5**, 817 mg, 两步收率 63%)。

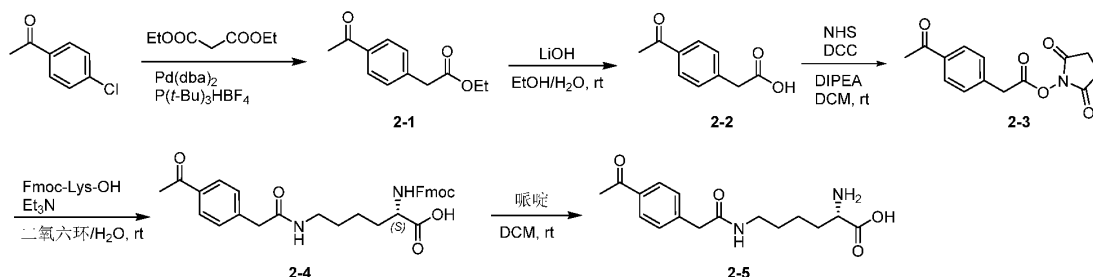
¹H-NMR (400 MHz, 重水) δ 8.04 (d, $J = 8.4$ Hz, 2H), 7.55 (d, $J = 8.0$ Hz, 2H), 5.21 (s, 2H), 3.74 (t, $J = 6.0$ Hz, 1H), 3.17 (t, $J = 6.4$ Hz, 2H), 2.70 (s, 3H), 1.95 – 1.83 (m, 2H), 1.62 – 1.52 (m, 2H), 1.47 – 1.35 (m, 2H)。

实施例 2 非天然氨基酸 NPAK 的制备

NPAK 的结构式如下所示:



反应过程如下图所示:



制备过程包括以下步骤:

a) 在反应烧瓶中, 加入对氯苯乙酮 (1.00 g, 6.47 mmol), 氮气氛围下, 加入丙二酸二乙酯 (6.84 g, 47.70 mmol)、KHCO₃ (0.97 g, 9.70 mmol) 和 K₂CO₃ (1.34 g, 9.70 mmol), 然后加入 Pd(dba)₃ (0.019 g, 0.030 mmol) 和 P(*t*-Bu)₃HBF₄ (0.021 g, 0.071 mmol), 加入完毕后置换氮气保护, 升温至 160°C 反应 40 h。TLC 检测反应完全后, 反应液中加入水 (30 mL), 用 EA 萃取 3 次, 合并有机相, 用水洗涤 2 次, 无水硫酸钠干燥, 过滤, 在 0-5°C 减压浓缩, 得到无色透明液体, 柱层析 (洗脱剂: PE:EA=10:1) 纯化得到产物 **2-1** (0.80 g, 收率 60%)。

b) 在反应烧瓶中, 加入 LiOH (0.30 g, 11.64 mmol) 和水 (5.0 mL), 加入乙醇 (10 mL), 加入产物 **2-1** (0.80 g, 3.88 mmol), 室温搅拌 2 h 后, TLC 检测反应完全, 反应液中加入 2 M

HCl 溶液调节 pH = 1~2, 用 EA 萃取 3 次, 合并有机相, 用水洗涤 2 次, 无水硫酸钠干燥, 过滤, 减压浓缩, 得到产物 **2-2** (0.5 g, 收率 72%)。

c) 在反应烧瓶中, 加入产物 **2-2** (0.20 g, 1.12 mmol), 依次加入 N-羟基琥珀酰亚胺 (NHS, 0.19 g, 1.68 mmol)、DIPEA (0.07 g, 0.56 mmol)、DCM (2.0 mL)。降温至 0~5℃, 加入 DCC (0.23 g, 1.12 mmol) 和 DCM (2.0 mL) 的溶液, 保温反应 2 h。升至室温搅拌过夜。TLC 检测反应完全后, 过滤, 用 DCM 洗涤, 母液减压浓缩, 柱层析(洗脱剂: PE:EA=5:1) 纯化得到产物 **2-3** (0.19 g, 收率 62%)。

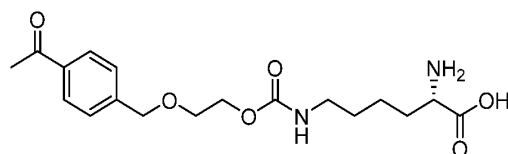
d) 在反应烧瓶中, 加入产物 **2-3** (0.10 g, 0.36 mmol), 依次加入三乙胺(0.04 g, 0.36 mmol)、Fmoc-Lys-OH 盐酸盐 (0.13 g, 0.36 mmol)、二氧六环(2.0 mL)和水(2.0 mL), 室温下搅拌反应 18 h。TLC 检测反应完全后, 减压浓缩, 用 EA 萃取 3 次, 无水硫酸钠干燥, 减压浓缩, 柱层析(洗脱剂: DCM:MeOH=15:1)纯化得到油状液体产物 **2-4** (0.03 g, 收率 61%)。

e) 在反应烧瓶中, 加入产物 **2-4** (0.08 g, 0.15 mmol)、DCM (1.0 mL) 和哌啶(0.04 g, 0.47 mmol), 在室温下搅拌 3 h。TLC 检测反应完全后, 减压浓缩, 用石油醚(5 mL)打浆 1 h, 过滤, 滤饼再用石油醚(5 mL)打浆 1 h, 过滤, 滤饼用乙醇重复打浆 4 次除去残余的哌啶, 最终得到类白色固体 **2-5** (0.02 g, 收率 43%)。

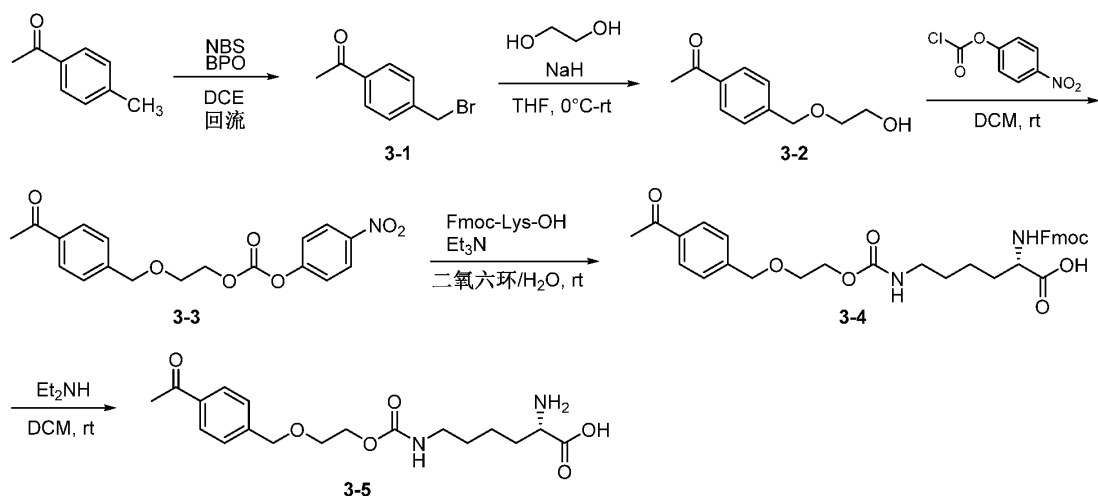
¹H-NMR (400 MHz, 重水) δ 7.85 (d, *J* = 8.2 Hz, 2H), 7.33 (d, *J* = 8.2 Hz, 2H), 3.94 (t, *J* = 6.3 Hz, 1H), 3.56 (s, 2H), 3.12 (t, *J* = 6.8 Hz, 2H), 2.54 (s, 3H), 1.80 – 1.70 (m, 2H), 1.54 – 1.45 (m, 2H), 1.40 – 1.224 (m, 2H)。

20 实施例 3 非天然氨基酸 NBPK 的制备

NBPK 的结构式如下所示:



反应过程如下图所示:



制备过程包括以下步骤:

a) 在反应烧瓶中, 加入对甲基苯乙酮(4.0 mL, 30.0 mmol), 加入溶剂 DCE (50.0 mL), 加入 NBS (6.41 g, 36.0 mmol)和 BPO (0.05 g, 0.3 mmol), 混合物在 80℃下回流 24 小时。TLC 检测反应完全后, 容器放入冰水中冷却, 析出固体, 过滤除去固体, 用饱和 Na_2CO_3 洗涤 3 次, 用 DCM 萃取 3 次, 合并有机相, 加入无水硫酸钠干燥, 过滤, 减压浓缩后得到粗产物 **3-1** (5.46 g, 收率 85%), 无需纯化直接用于下一步。

b) 在反应烧瓶中, 加入 NaH (0.58 g, 14.64 mmol, 60%), 加入干燥的溶剂 THF (20 mL), 冰浴冷却下, 缓慢加入乙二醇(6.7 mL, 122.0 mmol), 在室温条件下搅拌 1 小时。然后加入产物 **3-1** (2.60 g, 12.2 mmol), 70℃加热回流 48 小时至反应完全。冰浴冷却下, 缓慢滴入饱和 NH_4Cl 淬灭 NaH, 用水洗涤, EtOAc 萃取 3 次, 合并有机相, 加入无水硫酸钠干燥, 过滤, 减压浓缩后, 经柱层析(洗脱剂: PE:EA=2:1)纯化得到产物 **3-2** (1.39 g, 收率 59%)。

c) 在反应烧瓶中, 加入产物 **3-2** (1.39 g, 7.2 mmol), 加入溶剂 DCM (10 mL), 冰浴冷却下, 加入对硝基苯基氯甲酸酯(1.74 g, 8.64 mmol)和吡啶(0.7 mL, 8.64 mmol), 室温搅拌 18 小时。TLC 检测反应完全后, 加入水洗涤, EtOAc 萃取 3 次, 合并有机相, 加入无水硫酸钠干燥, 过滤, 减压浓缩后, 经柱层析(洗脱剂: PE:EA=3:1)纯化得到产物 **3-3** (2.27 g, 收率 88%)。

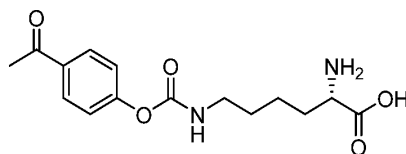
d) 在反应烧瓶中, 加入产物 **3-3** (2.27 g, 6.32 mmol), 加入溶剂二氧六环(16 mL)和水(4 mL), 加入 Fmoc-Lys-OH 盐酸盐(2.13 g, 5.27 mmol), 再加入三乙胺(1.85 mL, 13.2 mmol), 在室温下搅拌 18 小时至反应完全。加入适量 1 M HCl 调 pH 至 2 左右, 用乙酸乙酯萃取, 合并有机相, 加入无水硫酸钠干燥, 过滤, 减压浓缩, 得到粗产物 **3-4**, 直接用于下一步。

e) 在反应烧瓶中, 将上一步的产物 **3-4** 溶于 DCM (10 mL), 加入二乙胺(5 mL), 在室温下反应 6 小时。TLC 检测反应完全后, 减压浓缩, 经柱层析(洗脱剂: DCM:MeOH:H₂O=40:10:1)纯化得到产物白色固体(**3-5**, 0.95 g, 两步收率 49%)。

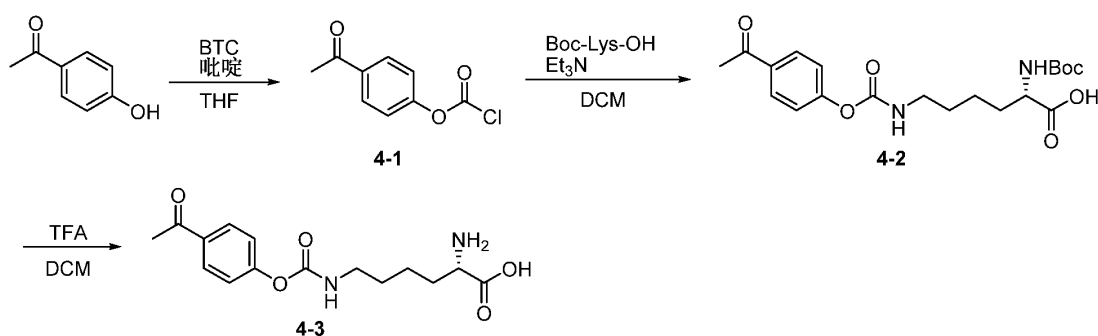
$^1\text{H-NMR}$ (400 MHz, 重水) δ 8.04 (d, $J = 8.0$ Hz, 2H), 7.56 (d, $J = 8.0$ Hz, 2H), 4.72 (s, 2H), 4.25 (s, 2H), 3.81 (s, 2H), 3.74 (t, $J = 6.0$ Hz, 1H), 3.16 – 3.08 (m, 2H), 2.70 (s, 3H), 1.97 – 1.79 (m, 2H), 1.60 – 1.48 (m, 2H), 1.48 – 1.35 (m, 2H).

5 实施例 4 非天然氨基酸 NPOK 的制备

NPOK 的结构式如下所示:



反应过程如下图所示:



10 制备过程包括以下步骤:

a) 在反应烧瓶中, 加入三光气(BTC, 2.18 g, 7.35 mmol), 加入溶剂 THF (10.0 mL), 在冰浴冷却下, 加入对羟基苯乙酮(2.0 g, 14.7 mmol)和吡啶(1.5 mL, 17.64 mmol), 混合物在室温下反应 24 小时。TLC 检测反应完全后, 加入适量水, 用 EtOAc 萃取 3 次, 合并有机相。加入无水硫酸钠干燥, 过滤, 减压浓缩后得到粗产物 **4-1** (1.20 g), 直接用于下一步。

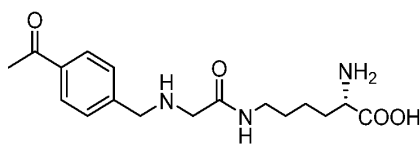
15 b) 在反应烧瓶中, 加入 Boc-赖氨酸(1.1 g, 5.0 mmol), 加入溶剂 DCM (10.0 mL), 加入产物 **4-1** (1.20 g)和三乙胺(2 mL, 15 mmol)。在室温下搅拌 24 小时, TLC 检测反应完全后, 加入适量 1 M HCl 调 pH 至弱酸性, 用 DCM 萃取 3 次, 合并有机相。加入无水硫酸钠干燥, 过滤, 减压浓缩后, 经柱层析(洗脱剂: DCM:MeOH=5:1)得到产物 **4-2** (1.70 g, 收率 84%)。

20 c) 在反应烧瓶中, 加入产物 **4-2** (1.70 g, 4.2 mmol), 加入溶剂 DCM (5 mL), 加入三氟乙酸(5 mL), 混合物在室温下反应 1 小时。TLC 检测反应完全后, 直接减压浓缩, 经柱层析(洗脱剂: DCM:MeOH:H₂O=40:10:1)后得到产物 **4-3** (1.19 g, 收率 92%)。

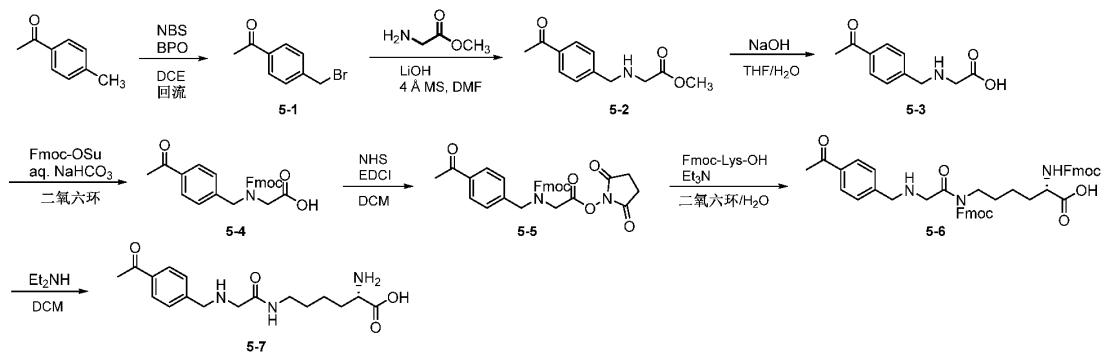
$^1\text{H-NMR}$ (400 MHz, 重水) δ 8.09 (d, $J = 8.6$ Hz, 2H), 7.31 (d, $J = 8.6$ Hz, 2H), 3.78 (t, $J = 6.0$ Hz, 1H), 3.27 (t, $J = 6.8$ Hz, 2H), 2.70 (s, 3H), 1.98 – 1.87 (m, 2H), 1.72 – 1.60 (m, 2H), 1.55 – 1.43 (m, 2H).

实施例 5 非天然氨基酸 NBGK 的制备

NBGK 的结构式如下所示：



反应过程如下图所示：



制备过程包括以下步骤：

a) 在反应烧瓶中，加入对甲基苯乙酮(8.0 mL, 60.0 mmol)，加入溶剂 DCE (80.0 mL)，加入 NBS (12.82 g, 72.0 mmol)和 BPO (145 mg, 0.6 mmol)，混合物在 90℃ 下回流 24 小时。TLC 检测反应完全后，容器放入冰水中冷却，析出固体，过滤除去固体，用饱和 Na_2CO_3 洗涤 3 次，用 DCM 萃取 3 次，合并有机相。加入无水硫酸钠干燥，过滤，减压浓缩后得到粗产物 **5-1** (11.12 g, 收率 87%)，无需纯化直接用于下一步。

b) 在反应烧瓶中，将 4 Å MS (14 g)和 LiOH (1.45 g, 34.54 mmol)用 DMF (70 mL)溶解，在室温下搅拌 20 min 后，加入甘氨酸甲酯盐酸盐(2.0 g, 15.7 mmol)，再搅拌 45 min 后，加入产物 **5-1** (4.0 g, 18.8 mmol)，室温搅拌 18 小时。TLC 监测反应完全后，过滤除去固体，用 EA 洗滤饼，滤液用水洗涤两次。无水硫酸钠干燥，减压浓缩后得粗产物 **5-2**，直接用于下一步。

c) 在反应烧瓶中，将上一步产物 **5-2** 溶于二氧六环(20 mL)，缓慢滴加 1 M NaOH，反应 2 小时后，TLC 监测到水解反应完成得到产物 **5-3**。加入 20 mL 饱和 NaHCO_3 后，缓慢加入溶于二氧六环(10 mL)的 Fmoc-OSu，室温搅拌过夜，TLC 监测反应完全后，用 1 M HCl 调至弱酸性，用 EA 萃取，无水硫酸钠干燥，减压浓缩，经柱层析(洗脱剂：DCM:MeOH=10:1)得到产物 **5-4** (3.60 g, 收率 89%)。

d) 在反应烧瓶中，加入产物 **5-3** (3.60 g, 8.0 mmol)、NBS(1.10 g, 9.6 mmol)、EDCI (1.85 g, 9.6 mmol)，加入溶剂 DCM (50 mL)，混合物在室温下反应 18 小时。TLC 监测反应完全后，用水洗涤 3 次，无水硫酸钠干燥，减压浓缩后得到产物 **5-5** (3.20 g, 收率 75%)。

e) 在反应烧瓶中，加入产物 **5-5** (3.20 g, 6.0 mmol)，加入溶剂二氧六环(40 mL)和水(10

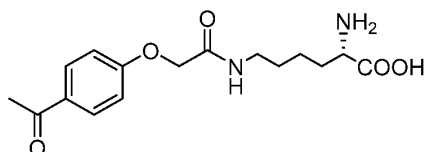
mL), 加入 Fmoc-Lys-OH 盐酸盐 (3.0 g, 7.2 mmol), 再加入三乙胺(2.0 mL, 15.0 mmol), 在室温下搅拌 18 小时至反应完全。加入适量 1 M HCl 调 pH 至 2 左右, 用 EA 萃取, 合并有机相, 加入无水硫酸钠干燥, 过滤, 减压浓缩, 经柱层析(洗脱剂: DCM:MeOH:AcOH=20:1:0.5)得到产物 **5-5** (3.50 g, 收率 75%)。

- 5 f) 在反应烧瓶中, 将产物 **5-5** 溶于 DCM (20 mL), 加入二乙胺(20 mL), 在室温下反应 6 小时。TLC 监测反应完全后, 减压浓缩, 经柱层析(洗脱剂: DCM:MeOH:H₂O=30:10:1)得到终产物白色粉末 **5-7** (0.55 g, 收率 37%)。

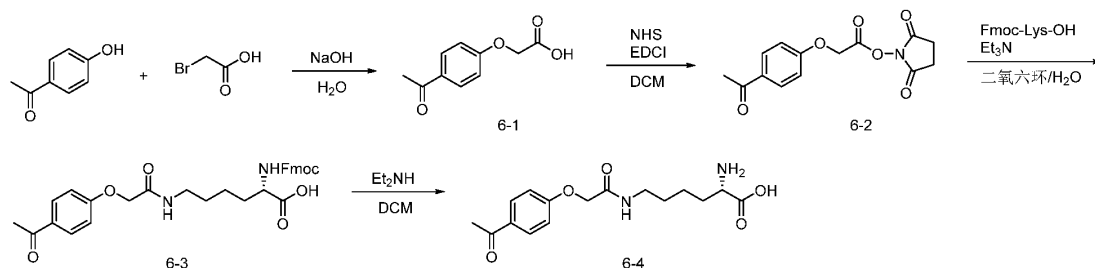
¹H-NMR (400 MHz, 重水) δ 7.98 (d, J = 8.2 Hz, 2H), 7.50 (d, J = 8.2 Hz, 2H), 3.84 (s, 2H), 3.71 (s, 1H), 3.31 (s, 2H), 3.17 (t, J = 6.9 Hz, 2H), 2.67 (s, 3H), 1.97 – 1.73 (m, 2H), 1.58 – 1.45 (m, 2H), 1.44– 1.27 (m, 2H).

实施例 6 非天然氨基酸 NPOK-2 的制备

NPOK-2 的结构式如下所示:



- 15 反应过程如下图所示:



制备过程包括以下步骤:

- a) 在反应烧瓶中, 加入对乙酰基苯酚(2.05 g, 15.0 mmol)和溴乙酸(2.50 g, 18.0 mmol), 再加入 NaOH(1.20 g, 30 mmol)的水溶液(6 mL)。混合物在 100℃下回流 24 小时至反应完全。反应冷却至室温, 加 1 M 盐酸调至酸性, 析出固体, 过滤得到白色粗产物 **6-1** (3.32 g, 收率 113%), 无需纯化直接用于下一步。

- b) 在反应烧瓶中, 将上一步粗产物 **6-1** (3.32 g, 17.0 mmol)溶于 DCM (50 mL)中, 加入 NHS (2.35 g, 20.4 mmol)和 EDCI (3.90 g, 20.4 mmol)。混合物在常温下搅拌 18 小时至反应完全。用 DCM 萃取, 合并有机相, 加入无水硫酸钠干燥, 过滤, 减压浓缩。再经柱层析(洗脱剂: DCM: MeOH: AcOH=20:1:0.5)纯化, 得到产物 **6-2** (1.67 g, 收率 38%)。

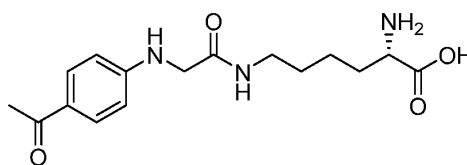
- c) 在反应烧瓶中, 加入产物 **6-2** (1.67 g, 5.7 mmol), 加入溶剂二氧六环(20 mL)和水(50 mL), 加入 Fmoc-Lys-OH 盐酸盐 (1.9 g, 4.8 mmol), 再加入三乙胺(1.7 mL, 12.0 mmol), 在

室温下搅拌 18 小时至反应完全, 加入适量 1 M HCl 调 pH 至 2 左右, 用 EA 萃取, 合并有机相, 加入无水硫酸钠干燥, 过滤, 减压浓缩。将所得粗产物 **6-3** 直接溶于 DCM(10 mL), 再加入二乙胺(5 mL)。混合物在常温下搅拌 18 小时至反应完全。减压浓缩, 经柱层析(洗脱剂: DCM:MeOH:H₂O=40:10:1)纯化得到终产物 **6-4** (709 mg, 两步收率 39%)。

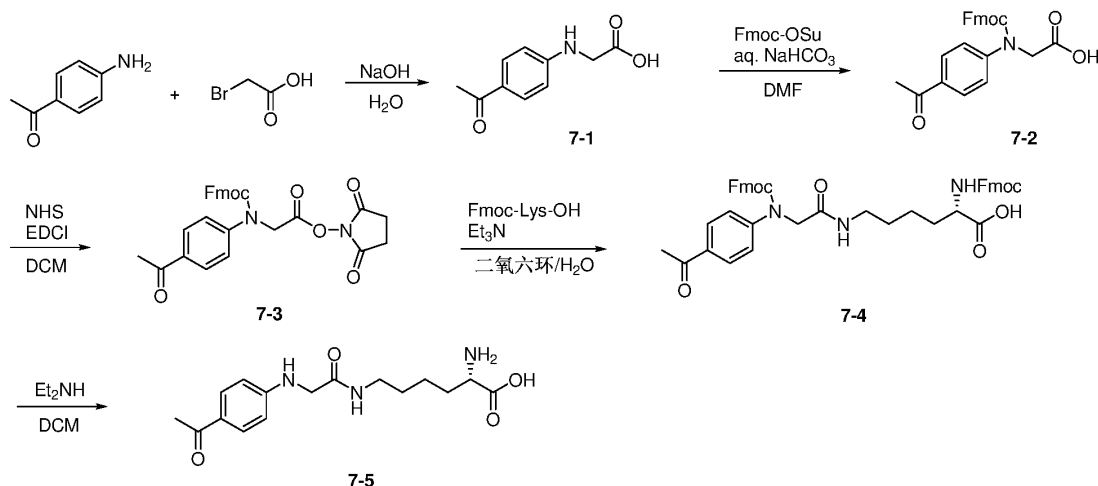
- 5 ¹H-NMR (400 MHz, 重水) δ 7.91 (d, $J = 8.8$ Hz, 2H), 6.98 (d, $J = 8.8$ Hz, 2H), 4.60 (s, 2H), 3.59 (t, $J = 6.4$ Hz, 1H), 3.19 (t, $J = 6.8$ Hz, 2H), 2.52 (s, 3H), 1.87 – 1.65 (m, 2H), 1.56 – 1.40 (m, 2H), 1.35 – 1.15 (m, 2H)。

实施例 7 非天然氨基酸 NGBK-2 的制备

10 NGBK-2 的结构式如下所示:



反应过程如下图所示:



制备过程包括以下步骤:

- 15 a) 在反应烧瓶中, 加入溴乙酸(2.10 g, 15.0 mmol)和 NaOH(0.80 g, 20 mmol)的水溶液(10 mL), 搅拌 10 分钟。然后加入对乙酰基苯胺(1.40 g, 10.0 mmol), 混合物在 100℃下回流 18 小时至反应完全。反应冷却至室温, 过滤, 用水洗涤, 得到白色粗产物 **7-1** (1.30 g, 收率 67%), 无需纯化直接用于下一步。
- 20 b) 在反应烧瓶中, 加入产物 **7-1** (1.30 g, 6.7 mmol)和 NaHCO₃(1.70 g, 20.1 mmol)的水溶液(20 mL)。再加入 Fmoc-OSu(2.80 g, 8.1 mmol)和 DMF(20 mL)。混合物在 60℃下搅拌 18 小时至反应完全。冷却至室温, 用 EA 萃取。保留水相用 1M 盐酸调 pH 至 2 左右, 再用 EA 萃取, 得有机相。加入无水硫酸钠干燥, 过滤, 减压浓缩。得产物 **7-2**, 无需纯化直

接用于下一步。

c) 在反应烧瓶中,将上一步产物 **7-2** (约 6.7 mmol), NHS(0.90 g, 8.0 mmol)和 EDCI(1.50 g, 8.0 mmol)溶解于 DMF(50 mL)。反应混合物在室温下搅拌 24 小时至反应完全。加水,用 DCM 萃取,得有机相。加入无水硫酸钠干燥,过滤,减压浓缩。得到产物 **7-3**, 无需纯化
5 直接用于下一步。

d) 在反应烧瓶中,加入上一步产物 **7-3** (约 6.7 mmol), Fmoc-Lys-OH 盐酸盐(2.30 g, 5.6 mmol), 以及三乙胺(2.0 mL, 14.0 mmol)。反应混合物在室温下搅拌 3 小时至反应完全。然后用 1M 盐酸调 pH 至 2 左右,用 EA 萃取。加入无水硫酸钠干燥,过滤,减压浓缩。得到产物 **7-4**, 无需纯化直接用于下一步。

10 e) 在反应烧瓶中,加入上一步产物 **7-4**。加入溶剂 DCM(20 mL)和二乙胺(10 mL)。反应混合物在室温下搅拌 12 小时至反应完全。先减压浓缩,再加入乙腈(50 mL)重新溶解,再减压浓缩,反复 3 次操作除去多余二乙胺。加入 DCM 打浆两次,得到最终产物 **7-5** (1.65 g, 总收率 51%)。

¹H-NMR (400 MHz, 重水) δ 7.71 (d, *J* = 8.8 Hz, 2H), 6.51 (d, *J* = 8.8 Hz, 2H), 3.80 (s, 2H), 3.53 (t, *J* = 6.8 Hz, 1H), 3.09 (t, *J* = 6.8 Hz, 2H), 2.39 (s, 3H), 1.73 – 1.60 (m, 2H), 1.42 – 1.33 (m, 2H), 1.25 – 1.14 (m, 2H)。
15

实施例 8

使用实施例 1-7 制备的 NPOK、NPAK、NBOK、NBPK、NBGK、NPOK-2、NBGK-2
20 在原核表达系统中制备重组人生长激素,并制备 PEG 定点偶联物。

(1) 辅助质粒的获得

辅助质粒 pSupAR-MbPylRS 购自质粒保藏组织 Addgene (货号#91705),该质粒可在大肠杆菌中表达特异识别吡咯赖氨酸衍生非天然氨基酸的 tRNA 和 tRNA 合成酶,经添加有 37.5 mg/L 氯霉素的 LB 培养基进行摇瓶培养后提取得到辅助质粒。

25 (2) 阅读框内部含终止密码子的重组人生长激素表达质粒构建

从美国国家生物信息中心数据库中获得智人生长激素(氨基酸序列如 SEQ ID NO: 1 所示)的编码基因的 mRNA 序列,在其翻译的蛋白质 C 端添加 6 个组氨酸组成的纯化标签,同时将 SEQ ID NO: 1 第 107 位氨基酸的密码子改成琥珀密码子(TAG),再通过全基因合成的方式合成该完整的 DNA 序列得到重组人生长激素的基因序列(SEQ ID NO: 2)。

30 智人生长激素的氨基酸序列(SEQ ID NO: 1)如下:

FPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIP
TPSNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLEEG
IQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRI
VQCRSVEGSCGF

35 重组人生长激素的基因序列(SEQ ID NO: 2)如下:

ATGTTTCCTACTATACCACTATCTCGTCTATTCGATAACGCTATGCTTCGGGCCCCAT
CGTCTTCATCAGCTGGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTATATCCCCAA
AGGAACAGAAGTATTCATTCCTGCAGAACCCCCAGACCTCCCTCTGTTTCTCAGAGT
CTATTCCGACACCCTCCAACAGGGAGGAAACACAACAGAAATCCAACCTAGAGCTGC
5 TCCGCATCTCCCTGCTGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTCCTCAGGAG
TGTCTTCGCCAACAGCCTGGTGTACGGCGCCTCTTAGAGCAACGTCTATGACCTCCTA
AAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGAAGATGGCAGCCCC
CCGGACTGGGCAGATCTTCAAGCAGACCTACAGCAAGTTCGACACAACTCACACA
ACGATGACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTTCAGGAAGGACATGG
10 ACAAGGTCGAGACATTCTGCGCATCGTGCAGTGCCGCTCTGTGGAGGGCAGCTGTG
GCTTCCTCGAGCACCACCACCACCACCTAA

通过一步亚克隆，将重组人生长激素的基因序列（SEQ ID NO: 2）克隆至 pET21a（Novagen，货号#69740-3）的 *Nde* I 和 *Xho* I 两个酶切位点之间，得到表达质粒 pET21-rhGH107，经测序验证与预期序列一致。pET21-rhGH107 可以用于表达第 107 位氨基酸密码子被替换为琥珀密码子的重组人生长激素，且在该蛋白的 C 端包含 6 个组氨酸纯化标签。重组人生长激素表达质粒 pET21-rhGH107 的图谱如图 1 所示。

（3）目的表达菌株的获得

将上述辅助质粒 pSupAR-MbPylRS 和表达质粒 pET21-rhGH107 共转化至大肠杆菌 OrigamiB (DE3) (Novagen，货号#70911-3) 感受态细胞，并使用含 100 mg/L 氨苄青霉素和 37.5 mg/L 氯霉素的 LB 培养基筛选获得双抗性菌株，即为重组人生长激素表达菌株。

（4）插入非天然氨基酸的重组蛋白质表达

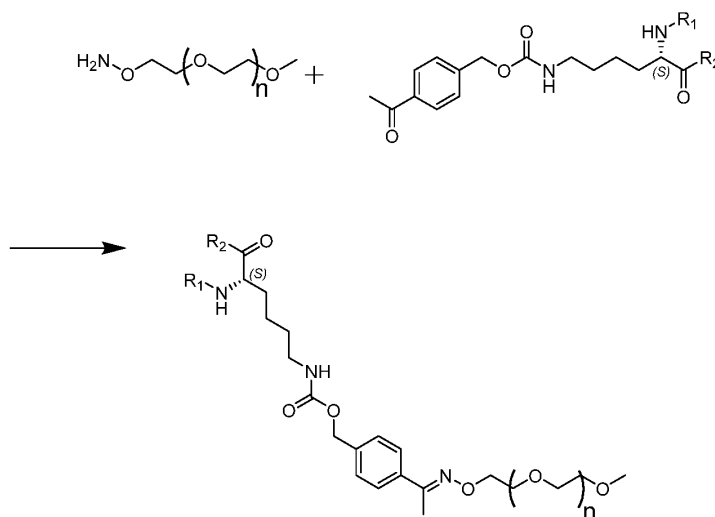
将筛选后的重组人生长激素表达菌株接种至 8 份 2×YT 培养基（酵母提取物 16 g/L，胰蛋白胨 10 g/L，NaCl 5 g/L，含 100 mg/L 氨苄青霉素和 37.5 mg/L 氯霉素）中，37℃培养至菌液 OD600 为 2.0 ± 0.2，向 8 份菌液中分别添加 IPTG（终浓度为 1 mM）和阿拉伯糖（终浓度为 0.2%），向第 1 份菌液中添加 NBOK（终浓度为 1 mM）、向第 2 份菌液中添加 NPAK（终浓度为 1 mM）、向第 3 份菌液中添加 NBPK（终浓度为 1 mM）、向第 4 份菌液中添加 NBGK（终浓度为 1 mM）、向第 5 份菌液中添加 NPOK（终浓度为 1 mM）、向第 6 份菌液中添加 NPOK-2（终浓度为 1 mM），向第 7 份菌液中添加 NBGK-2（终浓度为 1 mM），第 8 份菌液不添加非天然氨基酸，作为阴性对照。37℃培养诱导表达 5-6 小时后取各培养液 1 ml，10000 rpm 离心 1 min，用 PBS 重悬至 OD600 为 10，分别取各菌悬液进行 SDS-PAGE 电泳，各个菌株的 SDS-PAGE 电泳图见图 2。图 2 结果表明表达菌株在分别添加了 7 种非天然氨基酸的情况下可以表达目的蛋白。

（5）插入非天然氨基酸的重组蛋白质的纯化

上述诱导表达后的各菌液 10000 rpm 离心 5 min，取沉淀，加入 1/20 细胞生长体积的 NTA Buffer（20 mM Tris-HCl pH7.9，0.5 M NaCl，10%甘油）和终浓度为 1 mM 的 PMSF，超声波匀浆破碎细胞，10000 rpm 离心 20 min，取上清，使用 Ni-NTA 层析柱吸附带有 His

标签的目的蛋白，再用洗脱液 NTA Buffer (20 mM Tris-HCl pH7.9, 0.5 M NaCl, 10%甘油, 250 mM 咪唑) 洗脱，得到纯度约 90%的目的蛋白(即，插入非天然氨基酸的重组蛋白质)样品。将各样品置于 10 K 硝酸纤维素透析袋，置于 pH7.0 的 PBS 缓冲液中透析(每 1 ml 蛋白溶液对应 200 ml 透析液)过夜，并换液 1 次，收集透析后的蛋白溶液，SDS-PAGE 检测蛋白浓度。

(6) 插入非天然氨基酸的重组蛋白质与 PEG 偶联反应



如上述合成路线所示(其中 R_1 到 R_2 的方向为氨基酸序列的 N 末端到 C 末端方向)。

以 30 KD 氨氧基 PEG 肟化反应偶联插入非天然氨基酸的重组蛋白质为例，偶联反应操作如下：在偶联反应前，用 pH7.0 PBS 缓冲液调节上述获得的目的蛋白至 0.5 mg/ml，按照 1: 15 (摩尔比，重组蛋白：氨氧基 PEG) 加入 30 KD 氨氧基 PEG 固体(来源北京键凯科技有限公司)，充分摇晃溶解，得到澄清透明溶液，之后将反应液密封，在恒温摇床(25℃, 100 rpm)中摇晃反应。48 h 后使用 SDS-PAGE 分析偶联情况，见图 3。图 3 结果显示 6 种目标蛋白均实现了 30 KD PEG 偶联，进一步表明了上述 7 种非天然氨基酸插入了目标蛋白中。插入非天然氨基酸的重组蛋白质与 PEG 偶联产物简称为“PEG-rhGH”。

(7) PEG-rhGH 的纯化

偶联后的 PEG-rhGH 经 Butyl HP 疏水层析柱分离纯化(上样缓冲液：50 mM $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, 0.8 M $(\text{NH}_4)_2\text{SO}_4$, pH8.5; 洗脱缓冲液：20 mM Tris-HCl, pH8.5, 40 倍柱体积梯度洗脱)，获得电泳纯级(>95%)纯品 PEG-rhGH。以含有 NPOK 和含有 NBOK 的 PEG-rhGH 为例，电泳图见图 4。

其他的含有本发明的非天然氨基酸的 PEG-rhGH 在纯化后也达到了电泳纯级。

(8) PEG-rhGH 的活性分析

具体过程如下：HEK293-GHR 细胞(细胞株来源见中国专利：2020112649827)用含

0.1 mg/mL 潮霉素 B(生工生物工程(上海)股份有限公司, 货号 A600230-0001)、0.75 mg/mL 的 G418 (生工生物工程(上海)股份有限公司, 货号 A600958-0005)、10%胎牛血清 (Gibco, 货号 10099141) 的 DMEM (Gibco, 货号 11995040) 培养基, 于 37℃、5 %二氧化碳条件下培养至足够量, 在黑色 96 孔细胞培养板 (Corning, 货号 3904) 中每孔接入 9×10^4 个 HEK293-GHR 细胞, 90 μ L/孔, 37℃、5 %二氧化碳条件下饥饿培养 16 h。将 rhGH 标准品 (购自中国食品药品检定研究院, 简称“中检院”)、含 NPOK 的 rhGH (即, rhGH (NPOK))、PEG 修饰 NPOK 的 PEG-rhGH (即, PEG-rhGH (NPOK))、含 NBOK 的 PEG-rhGH (即, rhGH (NBOK))、PEG 修饰 NBOK 的 PEG-rhGH (即, PEG-rhGH (NBOK))、市售的 rhGH 药品赛增[®] (长春金赛药业有限责任公司)、市售的 PEG-rhGH 药品金赛增[®] (长春金赛药业有限责任公司) 样品进行梯度稀释, 共 9 个浓度, 于 37℃、5 %二氧化碳条件下培养 4 h。每孔加入 50 μ L ONE-Glo[™] (Promega, 货号 E6120) 检测试剂, 酶标仪读取化学发光值。根据样品浓度和酶标仪的读值拟合曲线, 计算 EC50。结果见图 5A-图 5G 和表 1。结果表明偶联前各位点细胞活性与中检院 rhGH 标准品相当。偶联后细胞活性降低, 约为标准品的 25%~50%, 其生物学活性变化规律与上市的 PEG-rhGH 一致。

其他的含有本发明的非天然氨基酸的 PEG-rhGH 在偶联之后细胞活性也降低为标准品的 25%~50%, 生物学活性变化规律也与上市的 PEG-rhGH 一致。

表 1

样品名称	EC50 (ng/mL)	相对活性(%)
中检院 rhGH 标准品	12.0	100%
rhGH (NPOK)	14.3	84%
PEG-rhGH (NPOK)	44.3	27%
rhGH (NBOK)	13.0	92%
PEG-rhGH (NBOK)	25.3	47%
赛增 [®]	12.6	95%
金赛增 [®]	68.6	18%

20 实施例 9

使用实施例 1-7 制备的非天然氨基酸 NPOK、NPAK、NBOK、NBPK、NBGK、NPOK-2 和 NBGK-2 在真核表达系统中制备重组 GFP 蛋白。

(1) 辅助质粒的获得

辅助质粒 pCMV-MbPylRS 购自质粒保藏组织 addgene (货号#91706), 该质粒编码哺乳动物细胞中特异性识别吡咯赖氨酸衍生非天然氨基酸的氨酰 tRNA 合成酶及对应的 tRNA (识别琥珀密码子 UAG)。

5 (2) 基因阅读框内部含有琥珀密码子的绿色荧光蛋白表达载体的构建

将表达含有纯化标签的野生型绿色荧光蛋白 (其基因编码序列如 SEQ ID NO: 3 所示) 的表达载体进行点突变, 获取阅读框第 40 位氨基酸密码子被突变为琥珀密码子的表达质粒 pEGFP40, 其完整序列见 SEQ ID NO: 4。

含有纯化标签的野生型绿色荧光蛋白的基因序列 (SEQ ID NO: 3) 如下:

10 ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCTGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTAGGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG
TGCCCTGGCCACCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCT
ACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG
15 TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACA
ACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCC
GCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACC
20 CCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCC
GCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT
GACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGAAGTACGAAGATTA
CAAGGATGACGACGATAAGTAA

表达质粒 pEGFP40 的基因序列 (SEQ ID NO: 4) 如下:

25 GTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATT
GATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATAT
GGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGA
CCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACT
TTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
30 AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCG
CTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACG
TATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
TAGCGGTTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTT
TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCAT
35 GACGCAAATGGGCGGTAGGCGTGACGGTGGGAGGTCTATATAAGCAGAGCTCGTTT
AGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAG
ACACCGGGACCGATCCAGCCTCCGGACTCTAGAGGATCCCCGCCACCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC

GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAGGG
CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCAC
CCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCA
5 CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGG
CAACATCCTGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGTGAAGTTCAAGATCCGCCACAACATCG
AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC
10 GGCCCCGTGCTGCTGCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAA
GACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGAAGTACGAAGATTACAAGGATGACGA
CGATAAGTAAGAATTCCGGAAGGGTTCGATCCCTACCGGTTAGTAATGAGTTTAAACG
GGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGA
15 CGGCAATAAAAAGACAGAATAAAACGCACGGGTGTTGGGTCGTTTGTTCATAAACGC
GGGGTTCGGTCCCAGGGCTGGCACTCTGTGATACCCACCGAGACCCCATTTGGGGC
CAATACGCCCCGCGTTTCTTCCTTTTCCCCACCCCAAGTTCGGGTGAAGGCC
CAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCAGATCTGCGCAGC
TGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT
20 GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCT
TTCGCTTCTTCCCTTCCCTTTCTCGCCACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAA
TCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAA
CTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC
CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAAC
25 ACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTTCGGCCTA
TTGGTTAAAAAATGAGCTGATTTAAACAAAAATTTAACGCGAATTAATTCTGTGGAATG
TGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAA
GCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAG
GCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAAC
30 TCCGCCCATCCCGCCCCCTAACTCCGCCAGTTCCGCCATTCTCCGCCCATGGCTGA
CTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAA
GTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTG
TATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAAC
AAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTTCGGCTATG
35 ACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGC
AGGGGCGCCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGC
AGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCTTGCGCAGCT
GTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCG
GGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTG

ATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTGACCACCAAGC
GAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGA
TGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAGTGTTCGCCAGGCTCA
AGGCGCGCATGCCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGC
5 CGAATATCATGGTGGAATAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGG
TGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTT
GGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCG
AGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTC
GCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTCGATTCCACCG
10 CCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGAT
CCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGCA
GCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTT
TTCATGCTATTCTAGTTGTGGTTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTAT
ACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGA
15 AATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAG
CCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCG
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
GAGAGGCGGTTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGC
TCGGTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTAT
20 CCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAA
GGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC
TGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGAC
CCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGCGCTTTCT
25 CAATGCTCACGCTGTAGGTATCTCAGTTCCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCT
GTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCTTATCCGGTAAGTATCGTCT
TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA
CTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACC
30 TTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTAAAGGG
ATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATAAAAATG
AAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCT
35 TAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTCGTTTCATCCATAGTTGCCTGA
CTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG
CAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCC
AGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTC
TATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC

GTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATT
CAGCTCCGGTTCCTCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAA
AGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCAGTGTTA
TCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATG
5 CTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGA
CCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATC
TTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAA
10 AAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATAT
TATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAAGTGCCACCTGACGTC
GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGA
TGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAG
15 TGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAG
AATCTGCTTAGG

(3) 插入非天然氨基酸的重组 GFP 蛋白表达

从美国模式培养物集存库 (ATCC) 购买中国仓鼠卵巢细胞 CHO-K1 (货号 #CCL-61-ATC), 并使用含 10%胎牛血清的 RPMI1640 培养基进行贴壁培养。提取上述步
20 骤的辅助质粒 pCMV-MbPylRS 与绿色荧光蛋白表达质粒 pEGFP40, 使用 lipo2000 转染试
剂 (invitrogen 公司, 货号#12566014), 根据说明书进行瞬时转染, (将细胞以 50000 个细
胞/孔的接种量接种到 24 孔板中, 共 8 个孔, 在接种培养 24 h 后进行转染, 每孔 500 ng 质
粒) 并在转染的同时向第 1 孔中添加 NPOK (终浓度为 1 mM)、向第 2 孔中添加 NPAK (终
浓度为 1 mM)、向第 3 孔中添加 NBOK (终浓度为 1 mM)、向第 4 孔中添加 NBPK (终浓
25 度为 1 mM)、向第 5 孔中添加 NBGK (终浓度为 1 mM), 向第 6 孔中添加 NPOK-2 (终浓
度为 1 mM), 向第 7 孔中添加 NBGK-2 (终浓度为 1 mM), 第 8 孔不添加非天然氨基酸,
作为阴性对照。在二氧化碳培养箱中静置培养 48 h 后在荧光显微镜下观察细胞, 结果显示
第 1 孔~第 7 孔存在明显的绿色荧光 (见图 6), 证明非天然氨基酸能够插入到绿色荧光蛋
白中, 得到完整的绿色荧光蛋白, 且不影响荧光蛋白的功能。

30

实施例 10

使用实施例 1-7 制备的 NPOK、NPAK、NBOK、NBPk、NBGK、NPOK-2、NBGK-2
在真核表达系统中表达含非天然氨基酸的抗 HER2 单抗, 并制备与毒素的偶联物。

(1) 辅助质粒的获得

35 辅助质粒 pCMV-MbPylRS 购自质粒保藏组织 addgene (货号#91706), 该质粒编码哺
乳动物细胞中特异性识别吡咯赖氨酸衍生非天然氨基酸的氨酰 tRNA 合成酶及对应的

tRNA（识别琥珀密码子 UAG）。

(2) 基因阅读框内部含有琥珀密码子的抗 **HER2** 抗体（曲妥珠单抗）表达载体的构建

通过全基因合成的方式合成编码曲妥珠单抗的重链和轻链 DNA（对应的氨基酸序列分别为 SEQ ID NO: 5 和 SEQ ID NO: 6），并亚克隆至真核表达载体 pCDNA3.1+，再将得到的表达载体进行点突变，获取重链阅读框第 142 位氨基酸密码子被突变为琥珀密码子的表达质粒 pCDNA3.1-Trastuzumab-UAG142，其图谱见图 7，其完整序列见 SEQ ID NO: 7。

曲妥珠单抗的重链氨基酸序列（SEQ ID NO: 5）如下：

```
EVQLVESGGGLVQPGGSLRLSCAASGFNIDKYIHWRQAPGKGLEWVARIYPTNG  
YTRYADSVKGRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQG  
10 TLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTWSNSGALTSGVHTFP  
AVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPA  
PELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTK  
PREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVY  
TLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSK  
15 LTVDKSRWQQGNVFCFSVMHEALHNHYTQKSLSLSPGK
```

曲妥珠单抗的轻链氨基酸序列（SEQ ID NO: 6）如下：

```
DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYS  
GVPSTRFSGSRSGTDFLTITSLQPEDFATYYCQGHYTPPTFGQGTKVEIKRTVAAPSVFIFPP  
SDEQLKSGTASVVCCLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKSTYLSSTL  
20 TLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
```

表达质粒 pCDNA3.1-Trastuzumab-UAG142 的基因序列（SEQ ID NO: 7）如下：

```
GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTC  
TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAG  
TAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATG  
25 AAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATA  
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTC  
ATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGA  
CCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC  
CAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTT  
30 GGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGT  
AAATGGCCCGCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGC  
AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATC  
AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATGAC  
GTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCAAAATGTCGTAAC  
35 AACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATA  
AGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATA  
CGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAACTTAAGCTTATGTACCGG  
ATGCAGCTGCTGAGCTGTATCGCCCTGTCTCTGGCCCTCGTGACCAACAGCGAAGTG
```

CAGCTGGTGGAAAGCGGCGGAGGACTGGTGCAGCCTGGCGGATCTCTGAGACTGAG
CTGTGCCGCCAGCGGCTTCAACATCAAGGACACCTACATCCACTGGGTGCGCCAGGC
CCCTGGCAAGGGACTGGAATGGGTGGCCAGAATCTACCCACCAACGGCTACACCAG
ATACGCCGACAGCGTGAAGGGCCGGTTCACCATCAGCGCCGACACCAGCAAGAACA
5 CCGCTACCTGCAGATGAACAGCCTGCGGGCCGAGGACACCGCCGTGTACTACTGTA
GTAGATGGGGAGGCGACGGCTTCTACGCCATGGACTATTGGGGCCAGGGCACCCCTCG
TGACAGTGTCTAGTGCGTAGACCAAGGGGCCCTCGGTCTTCCCCCTGGCACCCCTCCT
CCAAGAGCACCTCTGGGGGCACAGCGGCCCTGGGCTGCCTGGTCAAGGACTACTTC
CCCGAACCGGTGACGGTGTCGTGGAACCTCAGGCGCCCTGACCAGCGGCGTGCACAC
10 CTTCCCGGCTGTCCTACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTG
CCCTCCAGCAGCTTGGGCACCCAGACCTACATCTGCAACGTGAATCACAAGCCCAGC
AACACCAAGGTGGACAAGAAAGTTGAGCCCAAATCTTGTGACAAAACCTCACACATG
CCCACCGTGCCAGCACCTGAACCTCGGGGGGACCGTCAGTCTTCTCTTCCCCC
AAAACCCAAGGACACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGT
15 GGACGTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGCGTGG
AGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAACAGCACGTACCGT
GTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTGAATGGCAAGGAGTACAA
GTGCAAGGTCTCCAACAAAGCCCTCCAGCCCCATCGAGAAAACCATCTCCAAGC
CAAAGGGCAGCCCCGAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGATGAGC
20 TGACCAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCAGCGACA
TCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACCTACAAGACCACGCCT
CCCGTGCTGGACTCCGACGGCTCCTTCTTCTCTACAGCAAGCTCACCGTGGACAAG
AGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCAC
AACCCTACACGCAGAAGAGCCTCTCCCTGTCTCCCGGCAAGTGATAAGGCCGGCCC
25 CTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGT
GCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCG
GAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGGGGTCTTCCCCTCTCGCCAAA
GGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGA
AGACAAACAACGTCTGTAGCGACCCTTTCAGGCAGCGGAACCCCCACCTGGCGA
30 CAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACA
ACCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCCTC
AAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCATTGTATGGGATC
TGATCTGGGGCCTCGGTACACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAACGT
CTAGGCCCCCGAACCACGGGGACGTGGTTTTCTTTGAAAAACACGATGATAATATG
35 GCCACAACCATGTACCGCATGCAACTCCTGTCTTGCAATTGCACTAAGTCTTGCACTTG
TCACAAACAGTGATATCCAGATGACCCAGAGCCCCAGCAGCCTGTCTGCCAGCGTGG
GCGACAGAGTGACCATCACCTGTAGAGCCAGCCAGGACGTGAACACCGCCGTGGCC
TGGTATCAGCAGAAGCCTGGCAAGGCCCCCAAGCTGCTGATCTACAGCGCCAGCTTC
CTGTACAGCGGCGTGCCAGCAGATTCAGCGGCAGCAGATCCGGCACCGACTTCACC

CTGACCATCAGCTCCCTGCAGCCCGAGGACTTCGCCACCTACTACTGCCAGCAGCAC
TACACCACCCCCCCCCACATTTGGCCAGGGCACCAGGTGGAAATCAAGCGTACGGTG
GCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACG
CCTCTGTTGTGTGCCTGCTGAATAACTTCTACCCCAGAGAAGCCAAAGTGCAGTGGA
5 AGGTGGACAACGCCCTGCAGAGCGGAAACAGCCAGGAAAGCGTGACAGAGCAGGA
TTCCAAGGATTCCACATACAGCCTGAGCAGCACACTGACACTGTCCAAGGCCGACTA
CGAGAAGCACAAGGTGTACGCCTGCGAAGTGACACACCAGGGACTGTCTCCCCCTG
TGACAAAGAGCTTCAACAGAGGAGAATGCTGATGATCTAGAGGGCCCGTTTAAACCC
GCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCC
10 CGTGCCTTCCTTGACCTGGAAGGTGCCACTCCCCTGTCTTCTCCTAATAAAATGAG
GAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTG
GGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCAC
GCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACC
15 GCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGC
CACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGA
TTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTA
GTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTT
TAATAGTGGACTCTTGTTCCAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTT
20 TTGATTTATAAGGGATTTTGCCGATTTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAA
CAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTC
CCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAAC
CAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCT
CAATTAGTCAGCAACCATAGTCCCGCCCCTAACCTCCGCCCATCCCGCCCCTAACCTCCG
25 CCCAGTTCCGCCCATTTCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGC
CGAGGGCCGCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTTGGAGG
CCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAA
GAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCT
CCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGC
30 TGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGTTCTTTTTGTCA
AGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGT
GGCTGGCCACGACGGGCGTTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGG
GAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTATCTCACC
TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCT
35 TGATCCGGCTACCTGCCCATTTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACG
TACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGG
GCTCGCGCCAGCCGAAGTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGG
ATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCG
CTTTTCTGGATTTCAGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATA

GCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTC
CTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGCAGCGCATCGCCTTCTATCGCCTTCT
TGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCC
AACCTGCCATCACGAGATTTGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCG
5 GAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGG
AGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAAT
AGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTC
CAAACATCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGG
CGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC
10 AACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAA
CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCTGTGCC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCT
CTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGGCTGCGGCGAGCGGT
ATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGG
15 AAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCG
TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCT
CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC
CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT
20 TCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACATATCGTCTTGAGTCCAACCCGTAAGACACGACT
TATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCG
GTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTAT
TTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTG
25 ATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTTGTTTGCAAGCAGCAGATTAC
GCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGC
TCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGAT
CTTCACCTAGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATATATGA
GTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATC
30 TGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACG
GGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC
GGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAAGTG
GTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTA
AGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGG
35 TGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTCCCAACGATCAAGGCG
AGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCTCCGATC
GTTGTGCAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATA
ATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACC
AAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATAC

GGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTC
 TTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCC
 ACTCGTGACCCAACTGATCTTCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAG
 CAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATG
 5 TTGAATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTC
 ATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA
 CATTTCCCGAAAAGTGCCACCTGACGTC

(3) 插入非天然氨基酸

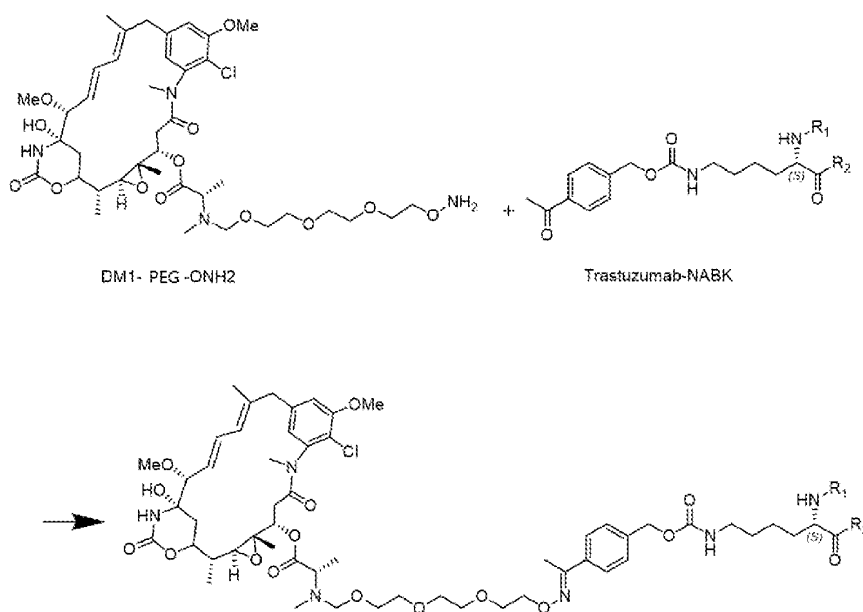
使用悬浮驯化的 HEK293 细胞, 以 $0.3 \times 10^5/\text{mL}$ 密度接种至 Wayne293™ 培养基(Quacell
 10 Biotechnology, 货号 A21501), 使用 1 L 的摇瓶, 装液量 240 mL, 在 120 rpm、5%CO₂、
 80%湿度的条件下震荡培养, 当细胞密度达到约 $1 \times 10^6/\text{mL}$ 时进行转染, 提取步骤(1)和
 (2)描述的辅助质粒 pCMV-MbPylRS 与表达质粒 pCDNA3.1-Trastuzumab-UAG142, 并去
 除内毒素备用, 每个摇瓶转染时使用表达质粒和辅助质粒各 120 μg, 将质粒加入离心管中,
 用 1×PBS 缓冲液稀释至 7.2 mL, 再在另一个离心管中加入 PEI 转染试剂(聚乙烯亚胺,
 15 Polyethyleneimine) 720 μg, 再加入 1×PBS 缓冲液稀释至 7.2 mL, 混匀后将两个离心管各
 静置 5 分钟, 再将两个离心管中的液体轻混匀后静置 10 分钟, 缓慢滴加共 14.4 mL 至 1 L
 摇瓶的 240 mL 的细胞悬液中, 并在滴加过程中不断轻轻摇匀, 120 rpm、5%CO₂ 和 80%湿
 度的条件下震荡培养 4 h 后添加终浓度为 1 mM 的非天然氨基酸, 再继续 120 rpm、5%CO₂、
 80%湿度的条件下震荡培养 5 天, 收获细胞培养上清用于纯化抗体。

(4) 纯化

采用 HiTrap Protein A, 1 ml 预装柱纯化细胞培养上清, 洗脱缓冲液为 100 mmol/L 甘
 氨酸, 200 mmol/L 醋酸盐, pH3.5, 得到纯化后的插入非天然氨基酸的曲妥珠单抗。

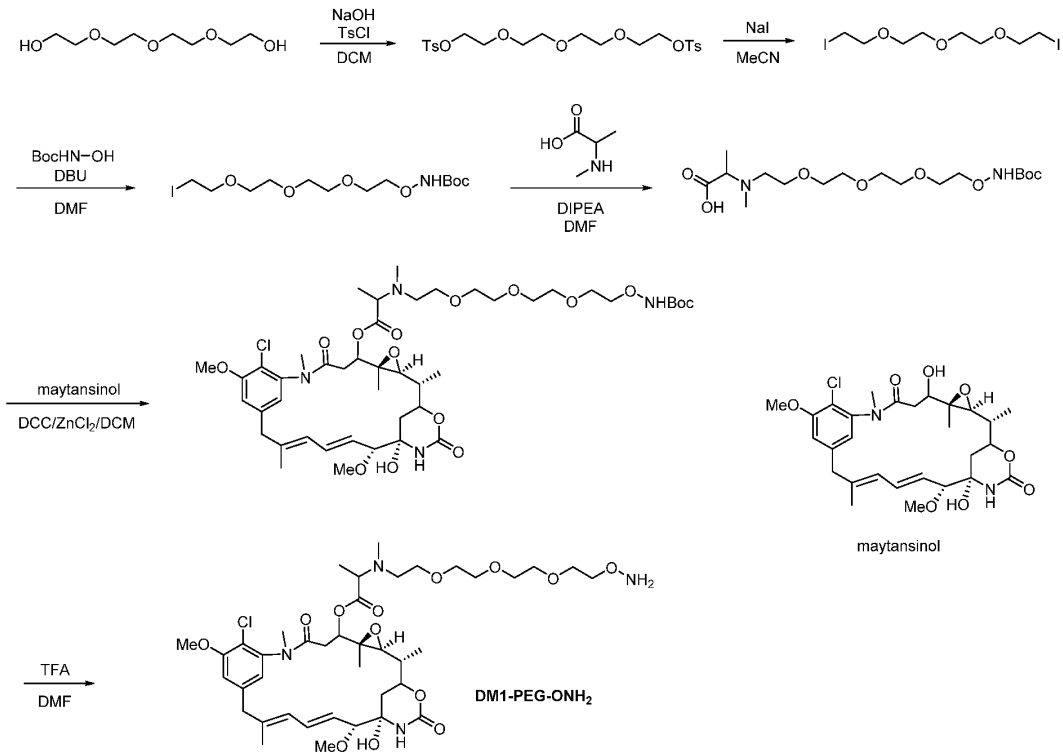
(5) 偶联

合成路线如下所示(以 NBOK 为例):



通过脲化反应将含有氨氧末端基团的毒素（DM1-PEG-ONH₂）与纯化得到的插入非天然氨基酸的曲妥珠单抗（其中 R₁ 到 R₂ 的方向为氨基酸序列的 N 末端到 C 末端方向）进行定点偶联，得到单抗-毒素偶联物。

DM1-PEG-ONH₂ 制备过程如下：



偶联过程的具体操作如下：将纯化得到的插入非天然氨基酸的曲妥珠单抗与 DM1-PEG-ONH₂ 以 1:15 的摩尔比进行混合溶清，然后采用 10 M 醋酸调节 pH 至 4.0，在摇床上摇晃震荡（25℃，200 rpm），48 h 取样，采用疏水层析原理的 HPLC（HIC-HPLC）检测曲妥珠单抗的反应情况。结果显示 93.0% 的含 NBPK 曲妥珠单抗被毒素修饰，93.0% 的含 NBOK 曲妥珠单抗被毒素修饰，87.4% 的含 NPOK-2 曲妥珠单抗被毒素修饰，如图 8A-8C 所示。曲妥珠单抗偶联比例=100%- 剩余未偶联原料的比例。

偶联后的单抗-毒素偶联物通过 50 kDa 超滤离心管进行换液，去除未反应的毒素原料，置换缓冲液为 20 mM 组氨酸缓冲液（pH6.5）。

HIC-HPLC 分析条件如下：

流动相 A（2 M 硫酸铵，75 mM K₂HPO₄，pH 7.2±0.2）；

流动相 B（75 mM K₂HPO₄，25% 异丙醇，pH 7.2±0.2）。

色谱柱	TSK gel Butyl-NPR column, 4.6 × 100 mm, 2.5 μm		
检测波长	280 nm		
柱箱温度	40℃		
进样器温度	5℃		
流速	0.5 ml/min		
进样量	10 μl		
梯度洗脱	时间 (min)	A(%)	B(%)
	0	80	20
	30	40	60
	35	25	75
	40	25	75
	45	80	20
	60	80	20

(6) 活性分析

采用 BT-474 细胞株 (ATCC, 货号 HTB-20), 检测样品对细胞增殖抑制作用。

具体过程如下: BT-474 细胞用含 10%胎牛血清 (Gibco, 货号 10099141) 的 ATCC
5 Hybri-Care Medium (ATCC, 货号 46-X) 培养基, 于 37℃、5 %二氧化碳条件下培养至足
够量, 在 96 孔细胞培养板 (Corning, 货号 3599) 中每孔接种 1.5×10^4 个 BT-474 细胞 (包
括溶媒对照孔), 空白对照孔中加入不含细胞的培养基, 80 μL/孔。将含 NBOK 的曲妥珠
单抗和 DM1 修饰的含 NBOK 曲妥珠单抗从 17 μg/mL 梯度稀释至 0.13 μg/mL, 共 8 个浓度,
每一稀释度 2 个复孔。将稀释好的样品转移至已接种 BT-474 细胞的培养板内, 每孔 10 μL,
10 在溶媒对照孔和空白对照孔中加入 10 μL 培养基, 于 37℃、5 %二氧化碳条件下培养。第
5 天每孔加入 10 μL CCK-8 (碧云天, 货号 C0043) 检测试剂, 于 37℃、5 %二氧化碳条件
下显色 6 h, 酶标仪于 450 nm 波长下读取光密度值。用下列公式来计算检测样品的抑制率
(Growth inhibition): 抑制率(%) = (OD 化合物 - OD 空白对照) / (OD 溶媒对照 - OD 空
白对照) × 100%。在 Excel 中计算不同浓度化合物的抑制率, 然后用 GraphPad Prism7 软件
15 作抑制曲线图和计算 IC50。结果见图 9。该结果显示 DM1 修饰的含 NBOK 曲妥珠单抗相
比含 NBOK 曲妥珠单抗, IC50 低约 3 倍, 肿瘤杀伤效果得到显著提升。

(7) DM1 修饰的含非天然氨基酸的曲妥珠单抗的稳定性考察

偶联后的单抗-毒素偶联物通过 50 kDa 超滤离心管 (Millipore, 货号 UFC905024#) 进
行换液, 去除未反应的毒素原料。置换缓冲液为 20 mM 组氨酸缓冲液 (pH6.0)。蛋白溶液
20 分成三份, 分别用 1M 柠檬酸溶液或 1M Tris 溶液调节 pH4.0、pH6.0、pH8.0, 置于 25℃
水浴中, 分别于 24 h、48 h、72 h、96 h 和 120 h 取样, HIC-HPLC 检测 (检测条件同上),
结果见表 2。

表 2 DM1 修饰的含 NBOK 的曲妥珠单抗在 25℃、不同 pH 下的稳定性

取样时间	未偶联的含 NOPK 的曲妥珠单抗含量/%		
	pH 4.0	pH 6.0	pH 8.0
0 h	7.0%		
24 h	7.3%	7.50%	7.61%
48 h	9.2%	7.50%	7.50%
72 h	11.5%	7.62%	7.63%
96 h	13.8%	7.68%	7.71%
120 h	15.2%	7.50%	7.61%

表 2 的结果表明，本发明提供的 DM1 修饰的含非天然氨基酸的曲妥珠单抗中，所形成的脲键在低 pH 下仍相对较稳定。

5

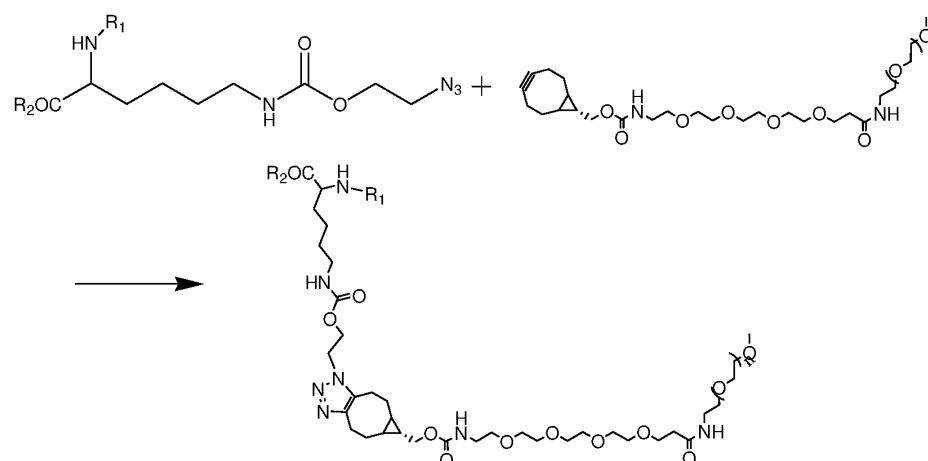
实施例 11 Lys-azido 的活性基团还原现象

(1) 高分辨质谱分析产物分子量

参考文献（宋礼华等，生物学杂志，16（1）：6-8,1999）的分泌型生长激素表达方法，根据文献中提供的信号肽序列，在表达载体上 rhGH 的 N 端添加 OmpA 信号肽编码序列，结合实施例 8 中的辅助质粒 pSupAR-MbPylRS，构建 K140 密码子突变为琥珀密码子（TAG）的 rhGH 的 OrigamiB（DE3）表达菌株，通过在发酵过程中添加 Lys-azido，表达第 140 位突变为 Lys-azido 的天然 rhGH，编号为 rhGH-Lys-azido-140；通过在发酵过程中添加 NBOK，表达第 140 位突变为 NBOK 的天然 rhGH，编号为 rhGH-NBOK-140，并参考上述参考文献相应纯化手段提纯。将 rhGH-Lys-azido-140 和 rhGH-NBOK-140 通过液质联用（高分辨质谱仪：XevoG2-XS Q-Tof，Waters 公司；超高效液相色谱：UPLC(Acquity UPLC I-Class)，Waters 公司）分析完整分子量。

如图 10A 所示，rhGH-Lys-azido-140 样品中出现了比理论分子量（22265 Da）小 26 Da 的组分，该组分推断为由 Lys-azido 末端的叠氮结构(-N₃)被还原为（-NH₂）的产物，说明目前常用的非天然氨基酸 Lys-azido 在插入蛋白时叠氮基不稳定，容易被还原形成还原产物，因此会失去偶联活性。如图 10B 所示，rhGH-NBOK-140 样品未出现比例较高的杂质，表明 NBOK 相比 Lys-azido，制备获得的重组蛋白产物更加稳定。

(2) 偶联效率比较



如上述反应路线所示，通过 Click 反应将 30 KD BCN-PEG（参照中国专利 CN 112279906A 自制）与 rhGH-Lys-azido-140 定点偶联（其中 R₁ 到 R₂ 的方向为氨基酸序列的 N 末端到 C 末端方向）。用 pH7.0 PBS 缓冲液调节上述获得的目的蛋白至 0.5 mg/ml，分别按照 1: 15 和 1: 25（摩尔比，重组蛋白：30 KD BCN-PEG），将 BCN-PEG 固体投入 rhGH-Lys-azido 溶液中，充分摇晃溶解，得到澄清透明的溶液，之后将反应液密封，在恒温摇床（25℃，70 rpm）中摇晃反应。每隔一段时间，取样用 SDS-PAGE 检测反应结果，72 h 后反应停止，转化率约 50%-70%。反应结果见图 11A。

参考实施例 8，通过脲化反应将 30 KD 氨氧基 PEG 与 rhGH-NBOK-140 定点偶联。每隔一段时间，取样用 SDS-PAGE 检测反应结果，48 h 后反应停止，转化率约 90%。反应结果见图 11B。

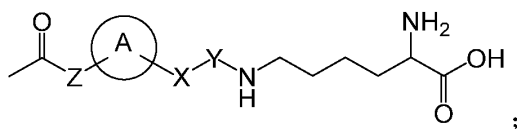
结合高分辨质谱分子量检测结果可以判断 Lys-azido 的还原导致产物无法与 BCN-PEG 偶联，由此降低了偶联率。而包含本发明的非天然氨基酸的重组蛋白与 PEG 进行偶联时，转化率明显高于包含 Lys-azido 的重组蛋白，而且反应时间也明显缩短，因此显著提高了反应效率。

除非特别限定，本发明所用术语均为本领域技术人员通常理解的含义。

本发明所描述的实施方式仅出于示例性目的，并非用以限制本发明的保护范围，本领域技术人员可在本发明的范围内作出各种其他替换、改变和改进，因而，本发明不限于上述实施方式，而仅由权利要求限定。

权利要求书

1. 一种非天然氨基酸，其特征在于，为具有如式(I)所示结构的化合物或其对映异构体，



I

其中，X和Z各自独立地表示取代或未取代的C₀~C₂₀直链或支链亚烷基，其中的一个或多个-CH₂-可任选地替换为-O-、-S-、-NH-、-C(O)-、-S(O)-中的一个或多个，Y表示-C(O)-、-S(O)-或-CH₂-，以及A表示取代或未取代的C₆~C₂₀芳基；

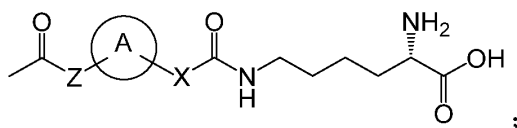
10 当所述X、Z和A各自独立地表示取代的基团时，取代基选自羟基、巯基、卤素、硝基、氰基、烷基、烯基、炔基、烷氧基、酰基、酰胺基、羧基、酯基、氨基、磺酰基、亚磺酰基、环烷基、杂环基、芳基、杂芳基中的一种或多种。

2. 根据权利要求1所述的非天然氨基酸，其特征在于，所述取代基选自羟基、巯基、卤素、硝基、氰基、C₁~C₆烷基、C₁~C₆烷氧基、酰基、酰胺基、羧基、酯基、氨基、磺酰基、亚磺酰基、C₃~C₈环烷基、C₃~C₈杂环基、C₆~C₂₀芳基、C₄~C₁₀杂芳基中的一种或多种；

优选地，所述X和Z各自独立地表示C₀~C₁₀直链或支链亚烷基，优选表示C₀~C₆直链亚烷基，其中的一个或多个-CH₂-可任选地替换为-O-、-S-、-NH-中的一个或多个；
20 更优选地，所述X和Z不同时为C₀的亚烷基；和/或

所述A表示取代或未取代的C₆~C₁₀芳基，更优选地，所述A表示取代或未取代的苯基或萘基。

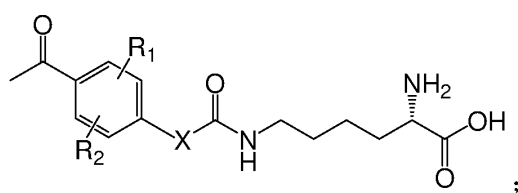
3. 根据权利要求1或2所述的非天然氨基酸，其特征在于，所述非天然氨基酸为具有如式(I-1)所示结构的化合物，



I-1

其中，所述X、Z和A各自独立地如权利要求1或2所定义；

优选地，所述非天然氨基酸为具有如式(I-2)所示结构的化合物，

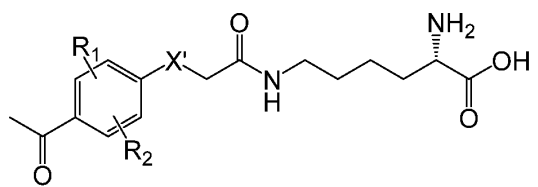


I -2

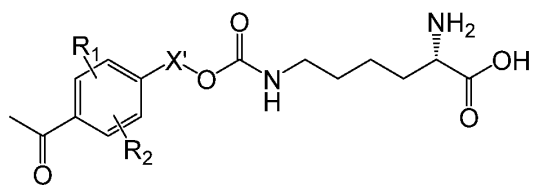
其中, 所述 X 如权利要求 1 或 2 所定义;

R_1 和 R_2 各自独立地表示氢、羟基、巯基、卤素、硝基、氰基、C1~C6 烷基、C1~C6 烷氧基、酰基、酰胺基、羧基、酯基、氨基、磺酰基、亚磺酰基、C3~C8 环烷基、C3~C8 杂环基、C6~C20 芳基或 C4~C10 杂芳基。

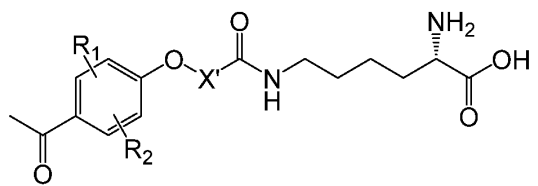
4. 根据权利要求 3 所述的非天然氨基酸, 其特征在于, 所述非天然氨基酸为具有如式(I -3)、式(I -4)、式(I -5)或式(I -6)所示结构的化合物,



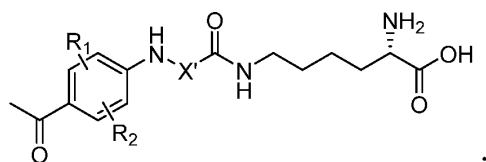
I -3



I -4



I -5

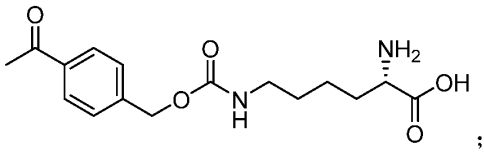
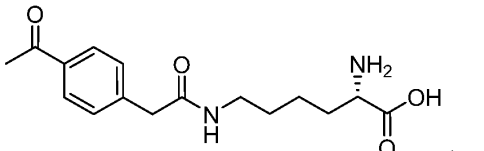
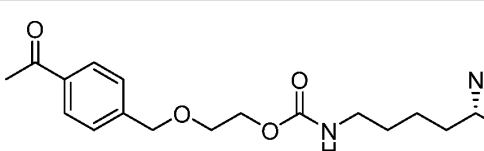
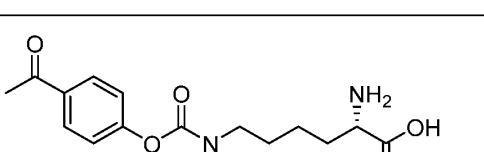
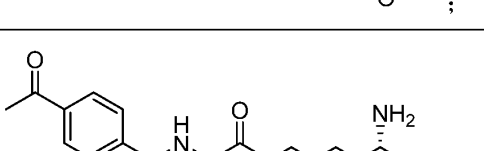
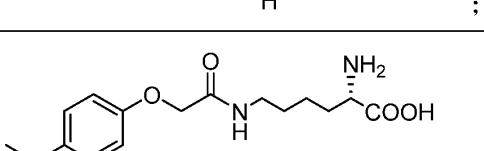


I -6

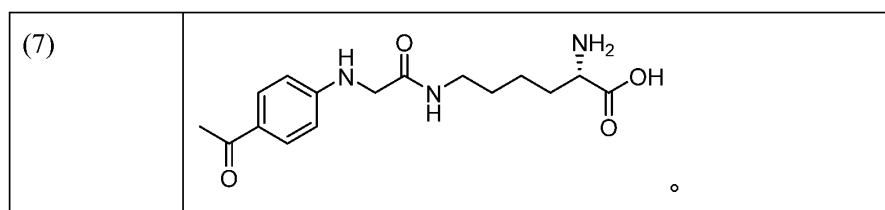
其中, X'表示 C0~C6 的直链亚烷基, 优选表示 C0~C4 的直链亚烷基, 其中的一个或多个-CH₂-可任选地替换为-O-和/或-NH-;

所述 R₁ 和 R₂ 各自独立地如权利要求 3 所定义。

- 5 5. 根据权利要求 1-4 任一项所述的非天然氨基酸, 其特征在于, 所述非天然氨基酸为具有以下结构之一的化合物:

(1)	
(2)	
(3)	
(4)	
(5)	
(6)	

或



6. 权利要求 1-5 任一项所述的非天然氨基酸在制备重组蛋白或重组蛋白偶联物中的应用；优选地，所述重组蛋白为重组人生长激素；或者，优选地，所述重组蛋白偶联物为重组人生长激素-聚乙二醇偶联物。

5

7. 一种重组蛋白，其特征在于，所述重组蛋白的氨基酸序列中的至少一个位点上为权利要求 1-5 任一项所述的非天然氨基酸。

8. 一种重组蛋白偶联物，其特征在于，通过权利要求 7 所述的重组蛋白中的非天然氨基酸的端基羰基与含有“NH₂-O-”端基的偶联部分形成脲键而形成。

9. 根据权利要求 8 所述的重组蛋白偶联物，其特征在于，所述重组蛋白为重组人生长激素，所述含有“NH₂-O-”端基的偶联部分为含有“NH₂-O-”端基的聚乙二醇，优选分子量为 10~100 KD 的聚乙二醇；更优选地，所述重组人生长激素的氨基酸序列如 SEQ ID NO: 1 所示；进一步优选地，对应于所述氨基酸序列 SEQ ID NO: 1 中的第 107 位点上包含权利要求 1-5 任一项所述的非天然氨基酸。

10. 权利要求 9 所述的重组蛋白偶联物在制备用于治疗内源性生长激素分泌不足所引起的生长发育障碍、特纳综合征引起的生长发育障碍或成人生长激素缺乏症的药物中的用途。

20