



(12) 发明专利申请

(10) 申请公布号 CN 103667202 A

(43) 申请公布日 2014. 03. 26

(21) 申请号 201210343156. 0

C12P 21/02 (2006. 01)

(22) 申请日 2012. 09. 14

(71) 申请人 中国科学院生物物理研究所
地址 100101 北京市朝阳区大屯路 15 号

(72) 发明人 王江云 潘延超 江欢欢 高峰

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021

代理人 王旭

(51) Int. Cl.

C12N 9/10 (2006. 01)

C12N 1/21 (2006. 01)

C12N 5/10 (2006. 01)

C12P 21/00 (2006. 01)

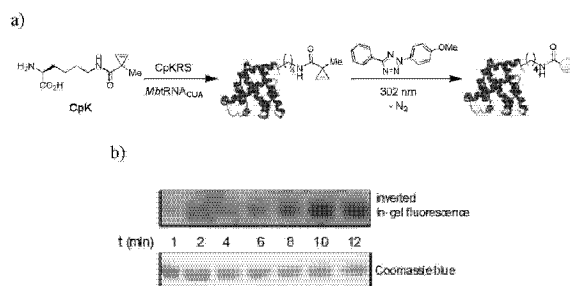
权利要求书2页 说明书10页
序列表8页 附图12页

(54) 发明名称

N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统及其应用

(57) 摘要

本发明涉及氨酰基-tRNA 合成酶突变体,其含有的氨基酸序列选自 SEQ ID NO :4 所示的氨基酸和它们的保守性变体构成的组。本发明提供利用正交 tRNA、正交氨酰基-tRNA 合成酶的配对将 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸(简称为 CpK) 定点特异插入目标蛋白质的 CpK 翻译系统,和利用所述翻译系统在目标蛋白质中定点特异插入 CpK 的方法。所述 CpK 翻译系统包含:(i) N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸;(ii) 本发明的正交氨酰基-tRNA 合成酶;(iii) 正交 tRNA,其中所述正交氨酰基-tRNA 合成酶用所述 CpK 优先氨酰化所述正交 tRNA;和(iv) 编码目标蛋白质的核酸,其中所述核酸含有所述正交 tRNA 特异性识别的至少一个选择密码子。本发明还涉及定点特异插入 CpK 的突变蛋白质通过与四唑类化合物进行光点击反应从而进行蛋白标记方面的应用。



1. 一种正交氨酰基-tRNA 合成酶,其含有的氨基酸序列选自 SEQ ID NO:4 所示的氨基酸和它们的保守性变体构成的组。

2. 一种 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统,所述系统包含:

(i) N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸;

(ii) 权利要求 1 所述的正交氨酰基-tRNA 合成酶;

(iii) 正交 tRNA,其包含 SEQ ID NO:1 所示的多核苷酸序列,其中所述正交氨酰基-tRNA 合成酶用所述 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸优先氨酰化所述正交 tRNA;和

(iv) 编码目标蛋白质的核酸,其中所述核酸含有所述正交 tRNA 特异性识别的至少一个选择密码子。

3. 如权利要求 2 所述的翻译系统,其特征在于,所述正交 tRNA 是含有琥珀反密码子的 tRNA,并且所述选择密码子是对应的琥珀密码子,并且其还包含编码正交氨酰基-tRNA 合成酶的核苷酸序列。

4. 一种宿主细胞,其包含正交 tRNA 序列和编码权利要求 1 所述的正交氨酰基-tRNA 合成酶的核苷酸序列。

5. 如权利要求 4 所述的宿主细胞,其中所述宿主细胞是真细菌细胞,或哺乳动物细胞,优选大肠杆菌细胞和人胚肾 (HEK) 293 细胞。

6. 一种产生在至少一个特定位置定点特异插入 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸的突变蛋白质的方法,所述方法包括下述步骤:

(a) 提供权利要求 2 所述的 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统,该系统包含:

(i) N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸;

(ii) 权利要求 1 所述的正交氨酰基-tRNA 合成酶;

(iii) 正交 tRNA,其包含 SEQ ID NO:1 所示的多核苷酸序列,其中所述正交氨酰基-tRNA 合成酶用所述 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸优先氨酰化所述正交 tRNA;和

(iv) 编码所述目标蛋白质的核酸,其中所述核酸在所述特定的位置包含所述正交 tRNA 特异性识别的至少一个选择密码子;和

(b) 将所述正交 tRNA 序列和编码所述正交氨酰基-tRNA 合成酶的核苷酸序列以及编码所述目标蛋白质的核酸序列克隆并转化到适当的宿主细胞中,在培养基中加入 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸,在所述蛋白质的翻译期间, N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸氨酰化的正交 tRNA 识别 mRNA 上的所述选择密码子以及 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸,从而介导 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸定点插入所述目标蛋白质的特定位置,从而产生在所选位置含 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸的所述目标蛋白质。

7. 如权利要求 6 所述的方法,其中所述正交 tRNA 是含有琥珀反密码子的 tRNA,所述选择密码子是对应的琥珀密码子。

8. 生产含有 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸的肌红蛋白突变体的方法,其利用权利要求 6 所述的方法,其中所用的编码肌红蛋白突变体的核酸序列为 SEQ ID NO:5,在野

生型肌红蛋白的 4 位引入 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸,所述肌红蛋白突变体的氨基酸序列为 SEQ ID NO :6。

9. 生产含有 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸的增强型绿色荧光蛋白突变体的方法,其利用权利要求 6 所述的方法,其中所用的编码增强型绿色荧光蛋白突变体的核酸序列为 SEQ ID NO :7,在野生型增强绿色荧光蛋白的 37 位引入 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸,所述增强型绿色荧光蛋白突变体的氨基酸序列为 SEQ ID NO :8。

10. 由权利要求 6-8 中任一项所述的方法获得的含有 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸的突变蛋白质的应用,所述突变蛋白质中的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸在光激发下与四唑官能团化合物反应,生成具有荧光的吡唑啉环,从而实现在体内或体外有效的标记蛋白。

N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统及其应用

技术领域

[0001] 本发明属于生物化学领域。具体地,本发明提供一种氨酰基-tRNA合成酶突变体,其含有的氨基酸序列选自SEQ ID NO:4所示的氨基酸和它们的保守性变体构成的组。本发明还涉及一种N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸(简称为CpK)翻译系统。更具体地,本发明涉及利用正交tRNA、正交氨酰基-tRNA合成酶的配对将CpK定点特异插入目标蛋白质的CpK翻译系统,和利用所述翻译系统在目标蛋白质中定点特异插入CpK的方法。本发明还涉及用这种翻译系统和这种方法产生的含有CpK的突变蛋白质,例如,插入CpK的肌红蛋白和增强型绿色荧光蛋白突变体,以及插入CpK的突变蛋白质的应用。

背景技术

[0002] 蛋白荧光标记技术目前已经被广泛应用于蛋白质功能的可视化研究中。荧光蛋白常被用来研究蛋白质在生物体内的表达和定位,但由于它本身体积比较大,往往会影响目标蛋白的生理功能。小分子荧光探针以其体积小、膜透性好、背景噪音低以及制备方便的优点成为蛋白质研究的一个有力工具。但是,把各种不同功能的荧光分子引入到蛋白质中的方法的短缺是限制其应用的瓶颈。因此,发展一种高效的蛋白标记技术对生命科学的研究有着重要的意义。

[0003] 为此,生物正交化学提供了令人兴奋的研究生物体中生物分子动力学及功能的新策略。与以配体为基础的方法相比,生物正交化学则需要能与目标生物分子特异性共价结合的探针分子。因此这种方法具有以下独特的优势:(1)能应用到各种各样的生物体分子中,包括蛋白、核酸、碳水化合物和脂质体。(2)用途广泛,探针分子的选择取决于研究人员的想象力。(3)具有高扩展性,适合活细胞中单个目标生物分子功能注释以及一类生物分子的全基因功能分析。由于这些特征,生物正交化学被成功的运用到可视化蛋白表达,跟踪蛋白定位,测定蛋白活性,鉴定蛋白相互作用和生物体系中蛋白转换等功能研究。

[0004] 与其他生物正交反应相比,点击化学由于其具有高选择性、水兼容性和高产率的特点使其在细胞生物学研究中具有巨大的应用前景。光点击化学是由纽约大学的Lin Q.教授提出来的概念,其的主要优点是不需要Cu(I)催化,而是通过光诱导引发反应,并且对细胞没有毒性。这类生物正交反应提供了一种用于研究生物时空分辨和可控引发的化学工具。由于四唑类化合物在紫外光照射下能够释放氮气而原位生成腈亚胺偶极子,该偶极子与烯烃发生瞬间环合生成吡唑啉环加成产物,因此这类反应能够实现时空可控引发。而且吡唑啉环加成产物是有荧光的,这样我们就能用荧光分析来证明特定的环加成产物的生成。近期,已有报道证明降冰片烯能与大环四唑类化合物进行高效地环加成反应,但是由于降冰片烯体积相对较大,有可能会影响蛋白的自身构象,因此我们希望通过扩展基因密码的方法在蛋白的特异性位点掺入含有环丙烯官能团的氨基酸,利用环丙烯官能团与四唑类化合物在一定波长紫外光照射下发生环加成反应,从而进行高效地、可控引发地位点特异性蛋白标记。

[0005] 为了能在蛋白的特异性位点掺入含有环丙烯官能团的氨基酸,本领域需要能将环

丙烯赖氨酸掺入到蛋白质中的新方案。现已开发了在原核和真核生物中将各种非天然氨基酸体内位点特异性地掺入蛋白质的通用方法。这些方法依赖于正交蛋白质翻译组分,所述组分识别合适的选择密码子(selector codon)从而能在体内多肽翻译期间将所需的非天然氨基酸掺入限定位置。这些方法利用识别选择密码子的正交 tRNA(O-tRNA),而相应的特异性正交氨酰 tRNA 合成酶(O-RS)用非天然氨基酸加载该 O-tRNA。这些组分不与宿主生物体内的任何内源性 tRNA、氨酰 tRNA 合成酶(RS)、氨基酸或密码子交叉反应(即,它必须是正交的)。利用这种正交 tRNA-RS 配对可能遗传编码大量结构各异的非天然氨基酸。

[0006] 本领域普遍知道利用适合于制备含一个或多个非天然氨基酸的蛋白质的正交翻译系统,例如产生正交翻译系统的通用方法。例如,参见国际公布号 WO 2002/086075,其发明名称为“METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS”;WO 2002/085923,其发明名称为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”;WO 2004/094593,其发明名称为“EXPANDING THE EUKARYOTIC GENETIC CODE”。定点特异插入非天然氨基酸的正交翻译系统及它们的产生和使用方法的其他讨论还可参见 Wang 和 Schultz, Chem. Commun. (Camb) 1 :1-11(2002); Wang 和 Schultz, Angewandte Chemie Int. Ed. 44(1) :34-66(2005); Xie 和 Schultz, Methods 36(3) :227-238(2005); Xie 和 Schultz, Curr. Opinion in Chemical Biology 9(6) :548-554(2005); Wang 等, Annu. Rev. Biophys. Biomol. Struct. 35 :225-249(2006)。

发明内容

[0007] 1、技术问题

[0008] 本发明提供一种氨酰基-tRNA 合成酶突变体,其含有的氨基酸序列选自由 SEQ ID NO :4 所示的氨基酸和它们的保守性变体构成的组。这种氨酰基-tRNA 合成酶突变体能够用 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸(简称为 CpK) 优先氨酰化与之配对的正交 tRNA,从而在翻译的氨基酸序列中插入 CpK。这是本发明人首次发现的,相应地,在本发明中将其命名为正交 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸氨酰基-tRNA 合成酶(CpKRS)。

[0009] 本发明涉及利用正交 tRNA、正交氨酰基-tRNA 合成酶的配对将 CpK 定点特异插入目标蛋白质的 CpK 翻译系统,和利用所述翻译系统在目标蛋白质中定点特异插入 CpK 的方法。本发明还涉及用这种翻译系统和这种方法产生的含有 CpK 的突变蛋白质及其应用。

[0010] 因此,本发明的目的在于提供利用正交 tRNA、正交氨酰基-tRNA 合成酶的配对将 CpK 定点特异插入蛋白质的 CpK 翻译系统,并且提供利用该翻译系统在目标蛋白质中定点特异插入 CpK 的方法。

[0011] 本发明还提供利用本发明的 CpK 翻译系统产生的含有至少一个 CpK 的突变蛋白质。在本发明的优选方面中,本发明人利用这种方法将 CpK 定点特异插入目的蛋白中,所述目的蛋白包括,但不限于,肌红蛋白(Myoglobin)和增强型绿色荧光蛋白(EGFP)等。通过本发明的方法得到的包含 CpK 的突变蛋白通过与四唑类化合物发生光点击反应生成具有荧光的化合物进而实现蛋白标记。然而,本领域技术人员应该理解,本发明的方法也可以用于在肌红蛋白和增强型绿色荧光蛋白之外的多种蛋白中定点特异插入 CpK,并不局限于这两种蛋白。

[0012] 2、技术方案

[0013] 本发明人经过筛选,首次获得一种正交氨酰基-tRNA 合成酶,其含有的氨基酸序列选自 SEQ ID NO:4 所示的氨基酸和它们的保守性变体构成的组。并且,本发明人利用所述正交氨酰基-tRNA 合成酶,研发了 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统。

[0014] 具体来说,本发明提供在体内(例如在宿主细胞内)识别选择密码子如琥珀终止密码子(TAG)从而将非天然氨基酸 CpK 定点特异插入到多肽链中的 CpK 翻译系统。所述 CpK 翻译系统包含不与宿主细胞翻译机制相互作用的正交-tRNA(O-tRNA)和正交氨酰基-tRNA 合成酶(O-RS)配对。即,宿主细胞内源性氨酰基-tRNA 合成酶不会识别 O-tRNA。类似地,本发明提供的 O-RS 不以显著水平或者某些情况下不以可检测水平地识别内源性 tRNA。利用所述翻译系统能够生产在翻译过程中定点特异插入 CpK 的多种突变体蛋白质。

[0015] 在一些方面中,本发明提供 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统。所述翻译系统包含:(a) 非天然氨基酸,即 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸(简称为 CpK), (b) 本发明的正交氨酰基-tRNA 合成酶(O-RS),和 (c) 正交 tRNA(O-tRNA),其包含 SEQ ID NO:1 所示的多核苷酸序列,其中所述正交氨酰基-tRNA 合成酶用所述非天然氨基酸(即 CpK),优先氨酰化所述 O-tRNA。

[0016] 优选地,本发明的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统还包含编码目标蛋白质的核酸,其中所述核酸含有由正交 tRNA(O-tRNA)特异性识别的至少一个选择密码子,优选地为琥珀密码子。更优选地,本发明的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统还包含编码正交氨酰基-tRNA 合成酶的核苷酸序列。

[0017] 所述系统中所用的正交氨酰基-tRNA 合成酶即为本发明人首次发现的氨酰基 tRNA 合成酶突变体,其含有的氨基酸序列选自 SEQ ID NO:4 所示的氨基酸和它们的保守性变体构成的组。

[0018] 在本发明的优选方面中,本发明提供一种 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统,所述系统包含:

[0019] (i) N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸;

[0020] (ii) 本发明的正交氨酰基-tRNA 合成酶;

[0021] (iii) 正交 tRNA,其包含 SEQ ID NO:1 所示的多核苷酸序列;其中所述正交氨酰基-tRNA 合成酶用所述 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸优先氨酰化所述正交 tRNA;和

[0022] (iv) 编码目标蛋白质的核酸,其中所述核酸含有所述正交 tRNA 特异性识别的至少一个选择密码子。

[0023] 优选地,所述 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统还包含编码本发明的正交氨酰基-tRNA 合成酶的核苷酸序列。

[0024] 该翻译系统中的各种组分可以衍生自各种物种来源,例如,该翻译系统中的各组分衍生自巴氏甲烷八叠球菌(*Methanosarcina barkeri*)。例如,正交 tRNA(O-tRNA)为古菌来源的反密码子突变为与琥珀密码互补的吡咯赖氨酸 tRNA。在一些实施方式中,O-tRNA 是含有琥珀反密码子的 tRNA。在一些实施方式中,O-tRNA 包含 SEQ ID NO:1 所示的多核苷酸序列,优选地,O-tRNA 的序列如 SEQ ID NO:1 所示。在一个实施方式中,用于该系统的正交氨酰基-tRNA 合成酶可以包含 SEQ ID NO:4 所示的氨基酸序列及该序列的保守变体。在优选的实施方式中,用于该系统的正交氨酰基-tRNA 合成酶的氨基酸序列为 SEQ ID NO:

4 所示。

[0025] 在一些方面中,本发明的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统还包含编码目标蛋白质的核酸,其中所述核酸具有由正交 tRNA (O-tRNA) 特异性识别的至少一个选择密码子。在优选方面中,所述正交 tRNA 是含有琥珀反密码子的 tRNA,并且所述选择密码子是对应的琥珀密码子。

[0026] 在一些方面中,本发明提供包含正交 tRNA 序列和编码本发明的正交氨酰基-tRNA 合成酶的核苷酸序列的宿主细胞。所用的宿主细胞不作具体限定,只要正交氨酰基-tRNA 合成酶 (O-RS) 和正交 tRNA (O-tRNA) 在它们的宿主细胞环境中保留它们的正交性即可。例如,所述宿主细胞可以是真细菌细胞,也可以是哺乳动物细胞,优选大肠杆菌和人胚肾 (HEK) 293 细胞。可以将包含正交 tRNA 序列的重组载体和包含编码本发明的正交氨酰 tRNA 合成酶的核苷酸序列的重组载体共转化到宿主细胞中,从而获得包含正交 tRNA 序列和编码本发明的正交氨酰 tRNA 合成酶的核苷酸序列的宿主细胞。所述包含正交 tRNA 序列和编码本发明的正交氨酰 tRNA 合成酶的核苷酸序列的宿主细胞构成本发明的另一个方面。

[0027] 本发明还提供产生在至少一个特定位置定点特异插入 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸的突变蛋白质的方法。所述方法利用上述 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统。所述方法通常包括下述步骤:(a) 提供含有以下组分的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统的步骤:(i) 非天然氨基酸,即 CpK;(ii) 本发明的正交氨酰基-tRNA 合成酶 (O-RS);(iii) 正交 tRNA (O-tRNA),其包含 SEQ ID NO:1 所示的多核苷酸序列,其中所述 O-RS 用所述非天然氨基酸(即 CpK) 优先氨酰化所述 O-tRNA;和(iv) 编码目标蛋白质的核酸,其中所述核酸含有 O-tRNA 特异性识别的至少一个选择密码子(任选地为琥珀密码子);(b) 将所述正交 tRNA 序列和编码所述正交氨酰基-tRNA 合成酶的核苷酸序列以及编码所述目标蛋白质的核酸序列克隆并转化到适当的宿主细胞中,在培养基中加入 CpK,在所述蛋白质的翻译过程中,CpK 氨酰化的 O-tRNA 识别 mRNA 上的所述选择密码子以及 CpK,从而介导 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸定点插入所述目标蛋白质的特定位置,从而产生在所选位置含有 CpK 的突变蛋白质。本领域技术人员应该理解,适当的重组载体的构建和宿主细胞的筛选可以通过常规分子克隆技术和筛选技术实现。

[0028] 本领域技术人员应该理解,在步骤 (b) 中,将所述正交 tRNA 序列和编码所述正交氨酰基-tRNA 合成酶的核苷酸序列以及编码所述目标蛋白质的核酸序列克隆并转化到适当的宿主细胞中可以通过多种方式进行,例如,将所述正交 tRNA 序列、编码所述正交氨酰基-tRNA 合成酶的核苷酸序列以及编码所述目标蛋白质的核酸序列分别可操作性地连接到适当的载体中,再以任意次序或三者共同转化到适当的宿主细胞中;或者,也可以将所述正交 tRNA 序列和编码所述正交氨酰基-tRNA 合成酶的核苷酸序列可操作性地连接到一个适当的载体中(两种序列之间有或无适当的接头连接),将编码所述目标蛋白质的核酸序列可操作性地连接到另一种不同的适当的载体中,然后将构建好的两种重组载体共同转化到适当的宿主细胞中;或者,也可以将所述正交 tRNA 序列和编码所述目标蛋白质的核酸序列可操作性地连接到一个适当的载体中(两种序列之间有或无适当的接头连接),将编码所述正交氨酰基-tRNA 合成酶的核苷酸序列可操作性地连接到另一种不同的适当的载体中,然后将构建好的两种重组载体共同转化到适当的宿主细胞中。上述克隆方案都是可行的,本领域技术人员可以根据实验的需要容易地进行适当的选择。

[0029] 另外,本领域技术人员还应该理解,为了避免宿主细胞对外源重组载体的“踢除”效应,往往选择用带有不同抗生素标记的载体来构建需要共同转化到同一宿主细胞中的核酸序列片段。对于适当的载体的选择、重组载体的构建、宿主细胞的转化或转染等等,都是本领域的常规技术,例如,可以参见美国冷泉港实验室出版的分子克隆手册。

[0030] 在所述方法的一些实施方式中,提供翻译系统的步骤包括通过定点诱变使野生型氨酰基-tRNA 合成酶的氨基酸结合口袋发生突变,选择用所述非天然氨基酸(即 CpK) 优先氨酰化所述 O-tRNA 的氨酰基-tRNA 合成酶突变体(即,本发明所用的正交氨酰基-tRNA 合成酶)。所述选择步骤包括定点诱变后从得到的氨酰基-tRNA 合成酶分子库进行所述 O-RS 的正选择和负选择(参见下述实施例 2)。在一些实施方式中,提供翻译系统的步骤还包括提供 O-tRNA 的序列,O-tRNA 为古菌来源的反密码子突变为与琥珀密码子互补的赖氨酸 tRNA,例如,所述 O-tRNA 是含有琥珀反密码子的 tRNA,或者 O-tRNA 包含 SEQ ID NO:1 所示的多核苷酸序列。在这些方法中,提供翻译系统的步骤还包括提供含有所述翻译系统所用的琥珀选择密码子的编码目标蛋白质的核酸。

[0031] 还可在宿主细胞内实施产生含有 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸的突变蛋白质的方法。在这些情况中,提供的宿主细胞包含本发明的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统(即,包含编码本发明的 O-RS 的核苷酸序列、O-tRNA 序列和含有至少一个选择密码子的编码目标蛋白质的核酸),而在适宜的培养条件下(例如,在培养基中添加 CpK 等)培养该宿主细胞可导致在所述目标蛋白质中定点特异插入 CpK。在一些实施方式中,提供步骤包括提供真细菌宿主细胞和哺乳动物细胞(例如,大肠杆菌 和人胚肾 (HEK) 293 细胞)。

[0032] 本发明还提供蛋白标记的遗传方法,所述方法利用上述 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统进行。这些方法通常始于提供含有以下组分的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统的步骤:(i) N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸,简称为 CpK;(ii) 本发明的正交氨酰 tRNA 合成酶,其含有的氨基酸序列选自 SEQ ID NO:4 所示的氨基酸和它们的保守性变体构成的组;(iii) 正交 tRNA,其包含 SEQ ID NO:1 所示的多核苷酸序列,其中所述正交氨酰 tRNA 合成酶用所述 CpK 优先氨酰化所述正交 tRNA;和(iv) 编码所述目标蛋白质的核酸,其中所述核酸含有所述正交 tRNA 特异性识别的至少一个选择密码子(任选地为琥珀密码子);然后在所述蛋白质的翻译过程中,CpK 氨酰化的正交 tRNA 对所述选择密码子起反应而将培养基中的所述 CpK 掺入所述蛋白的所选位置,之后通过与含四唑官能团的化合物在一定波长紫外光照射下发生环加成反应,生成具有荧光的化合物以实现位点特异性蛋白标记。

[0033] 因此,本发明还提供利用本发明的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统获得的包含至少一个 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸的突变蛋白质的应用,所述突变蛋白质中的 N^ε-(1-甲基环丙-2-烯酰胺)-赖氨酸在光引发下通过与含有四唑官能团的化合物特异反应,产生荧光,用于在体内或体外有效地标记蛋白。所述活体细胞选自真细菌细胞,或哺乳动物细胞,优选大肠杆菌细胞和人胚肾 (HEK) 293 细胞。

[0034] 3、有益效果

[0035] 通过生物正交化学的方法选择性的修饰蛋白,可以实现蛋白位点特异性插入生物正交反应基团。应用琥珀密码子在细胞中编码含有环丙烯官能团的氨基酸(CpK),实现在特

定位点插入该非天然氨基酸的蛋白质的高效表达,进而在光引发下通过与含有四唑官能团的化合物特异反应,生成具有荧光的化合物,从而在体内或者体外有效地标记蛋白。

附图说明

[0036] 从下面结合附图的详细描述中,本发明的上述特征和优点将更明显,其中:

[0037] 图 1 是 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸 (CpK) 的化学合成,图 1a 是 CpK 的化学合成路线图,图 1b 是化合物 1a 的氢谱图,图 1c 是化合物 1b 的氢谱图,图 1d 是化合物 1b 的碳谱图,图 1e 是化合物 1c 的氢谱图,图 1f 是化合物 1c 的碳谱图,图 1g 是化合物 1d 的氢谱图,图 1h 是化合物 1d 的碳谱图,图 1i 是化合物 CpK 的氢谱图,图 1j 是化合物 CpK 的碳谱图;

[0038] 图 2 是四唑类化合物 4 的化学合成,图 2a 是四唑类化合物 4 的化学合成路线图,图 2b 是化合物 4c 的氢谱图,图 2c 是化合物 4c 的碳谱图,图 2d 是化合物 4 的氢谱图,图 2e 是化合物 4 的碳谱图;

[0039] 图 3 是正交 tRNA,氨酰基-tRNA 合成酶,肌红蛋白及增强型绿色荧光蛋白突变体序列;

[0040] 图 4a 是 CpK-肌红蛋白的 SDS-PAGE 电泳图,图 4b 是 CpK-肌红蛋白的一级质谱图,图 4c 是 CpK-肌红蛋白的二级质谱图;

[0041] 图 5 是 CpK-肌红蛋白与四唑类化合物 3 的体外光点击反应图,图 5a 表示 CpK-肌红蛋白与四唑类化合物 3 的光点击反应示意图,图 5b 表示 CpK-肌红蛋白与四唑类化合物 3 点击反应后的 SDS-PAGE 胶图;

[0042] 图 6 是 CpK-增强型绿色荧光蛋白与四唑类化合物 4(在图 6 中以“4”表示四唑类化合物 4) 的体内光点击反应图。

具体实施方式

[0043] 以下通过实施例来进一步阐明本发明。但是应该理解,所述实施例只是举例说明的目的,并不意欲限制本发明的范围和精神。

[0044] 本领域技术人员应该理解,除非特别说明,下述实施例中所用的化学试剂均为可通过商业途径购得的分析纯级别的试剂。

[0045] 实施例 1:化学合成

[0046] 1、 N^{ϵ} -(1-甲基环丙-2-烯酰胺)-赖氨酸 (CpK) 的化学合成(图 1)

[0047] CpK 的化学合成路线参见图 1。具体合成反应步骤如下:

[0048] 向装有磁力搅拌和温度计的 1000mL 圆底烧瓶中加入叠氮化钠 (6.5g, 100mmol), 2M 氢氧化钠溶液 (200mL, 40mmol), 四丁基溴化铵 (TBAB) (80mg, 0.25mmol) 和 100mL 己烷 (100mL), 于 0℃ 冰浴上剧烈搅拌。待完全溶解后,用注射器缓慢滴加三氟甲基磺酸酐 (8.2mL, 50mmol), 反应 10 分钟。另取 2-甲基乙酰乙酸乙酯 (3.53mL, 25mmol) 溶于 100mL 乙腈中,然后通过漏斗将 2-甲基乙酰乙酸乙酯溶液加入反应烧瓶中,接着再加入 10mL 乙腈,可以发现初始无色的反应液迅速变黄。继续在 0℃ 冰浴上搅拌 30 分钟后,用冰浴后的水 (100mL) 和乙醚 (100mL) 稀释反应液,并转移到分液漏斗中,收集有机相,用冰浴后的乙醚萃取,再加入无水 $MgSO_4$ 干燥,过滤,浓缩,得到亮黄色的油状化合物 1a (3.0g, 产率 84%)。

[0049] 取一 250mL 圆底烧瓶,加入四三苯基醋酸二铯 (29mg, 0.02mmol), 二氯甲烷 (80mL) 和三甲基乙炔基硅 (5.7mL, 40mmol), 同时取化合物 1a (2.84g, 20mmol) 溶于二氯甲烷 (15mL) 中。然后,通过注射泵于室温下将 1a 溶液泵入圆底烧瓶中,速率 0.6mL/hr。待 1a 溶液全部泵入烧瓶中后,减压蒸馏,并用 10% 的醚 / 己烷作为洗脱剂,通过硅胶层析柱进行分离,得到浅黄色的油状化合物 1b (1.9g, 产率 46%)。

[0050] 在 0°C 冰浴条件下,将化合物 1b (0.594g, 3.0mmol) 溶于 15mL 甲醇溶液中,并加入 20mL 2M KOH 溶液。反应混合液于室温下搅拌过夜,然后减压,移除甲醇,用 6M 盐酸调节 pH 至 1-2,再用二氯甲烷萃取三次,收集有机相,用无水 $MgSO_4$ 干燥,过滤,浓缩,得到白色固体化合物 1c (0.217g, 产率 74%)。

[0051] 取 N-羟基丁二酰亚胺 (0.242g, 2.1mmol) 于 0°C 冰浴条件下溶于 24mL 乙酸乙酯 / 二恶烷 (1 : 1) 溶液中,再依次加入化合物 1c (0.196g, 2mmol) 和二环己基碳二亚胺 (0.433g, 2.1mmol)。混合液在室温下搅拌 5h,然后过滤,取滤液进行减压蒸馏,再用 1 : 1 的己烷 / 乙酸乙酯作为洗脱剂,通过硅胶层析柱进行分离,得到化合物 1d (0.337g, 产率 84%)。

[0052] 往烧瓶中加入 Fmoc-赖氨酸 -OH · HCl (0.689g, 1.7mmol) 和 N, N-二异丙基乙胺 (0.490g, 3.8) 的二甲基甲酰胺 (8mL) 溶液,另取化合物 1d (331mg, 1.7mmol) 溶于 3mL 二甲基甲酰胺中,室温下将 1d 溶液滴加到先前所述混合液中,反应 5h,然后进行减压蒸馏,再用 50mL 乙酸乙酯溶解,并用柠檬酸 (25mL × 2), 水 (25mL) 和盐水 (25mL) 的混合液进行连续洗脱,收集有机相,用无水 $MgSO_4$ 干燥,过滤,浓缩,得到 0.500g 粗提物。接着将粗提物溶于 10mL 二氯甲烷中,加入 5mL 二乙胺反应 5h,得到的混合物进行减压蒸馏,再用乙醚 (25mL × 2) 洗脱,干燥,得到终产物 CpK (0.180g, 产率 88%)。

[0053] 合成的 N^e-(1-甲基环丙-2-烯酰胺)-赖氨酸 (CpK) 的化学式为 $C_{11}H_8N_2O_3$, 分子量为 226.27, 结构式参见图 1。

[0054] 2、四唑类化合物 3 的化学合成

[0055] 本发明人合成的四唑类化合物 3 的化学式为 $C_{14}H_{12}N_4O$, 分子量为 252.27, 结构式为:



[0057] 具体合成反应步骤如下:

[0058] 将苯甲醛 (1.06g, 10.0mmol) 溶于 100mL 无水乙醇中,加入等当量的苯磺酰肼 (1.72g, 10.0mmol), 室温下搅拌过夜,待有大量白色固体析出后,过滤,收集固体,干燥后得 2.92g 产物,收率 92%,不需进一步提纯而直接用于下一步反应。将对甲氧基苯胺 (615mg, 5.0mmol) 溶于无水乙醇:水 (1 : 1, 10mL) 中,冰浴冷却后加入 1.0mL 浓盐酸,搅拌 10min 后,缓慢滴加 5mL $NaNO_2$ (363mg, 5.25mmol) 的水溶液。冰浴下继续搅拌 1h 后,将其滴加到上述合成的 Schiff 碱 (1.59g, 5.0mmol) 的 20mL 吡啶溶液中。TLC 跟踪反应进程,待反应结束后,加入等体积的水,有大量固体析出,过滤,用乙醚和乙酸乙酯的混合溶剂 (1 : 1) 洗涤,得到浅粉色固体,即为目标产物,收率为 52%。

[0059] 3、四唑类化合物 4 的化学合成 (图 2)

[0060] 本发明人合成的四唑类化合物 4 的化学式为 $C_{27}H_{30}N_6O_8$, 分子量为 556.56, 结构式参见图 2。

[0061] (E)-甲基 4-((2-苯磺酰肼叉)甲基))苯甲酸和 6-甲酯基-萘-2-氯在嘧啶溶液中生成四唑 4a, 然后用氢氧化锂水解, 反应溶液为乙醇/三氟乙酸/水的混合液, 从而得到 4b。取 4b (88.5mg, 0.246mmol) 和三乙胺 (83.5 μ L, 0.6mmol) 溶解于 0.3mL 2-(2-苄氧基)乙氧基乙醇, 2.7mL 甲苯和 3.0mL 二氧杂环乙烷中, 并加入 3 \AA 的分子筛, 在氩气保护下加入叠氮磷酸二苯酯 (130 μ L, 0.6mmol), 100 $^{\circ}$ C 搅拌回流。待反应完全后, 过滤除去分子筛, 减压蒸馏, 并用硅胶柱分离纯化得到 4c。取 4c (75mg, 0.1mmol) 溶于 2mL 乙醇/三氟乙酸 (1:1) 混合液中, 在氩气保护下加入 20 μ L 2M 的盐酸, 10% 钨和 7.5mg 碳, 然后通入氢气, 室温反应 12 小时, 过滤, 取滤液进行真空干燥, 最后通过重结晶作用获得目标产物 (47mg, 产率 83%)。

[0062] 以上合成反应所需化学试剂如无特别说明, 均购自北京百灵威公司和北京化工厂, 均为分析纯以上级别。

[0063] 实施例 2: 进化 CpK 特异性氨酰基-tRNA 合成酶

[0064] 为了在基因中位点特异性插入 CpK, 需要在所用的 *E. coli* 宿主细胞中引入氨酰基-tRNA 合成酶/tRNA 正交对, 这个正交对来源于巴氏甲烷八叠球菌 (*Methanosarcina barkeri*) 琥珀抑制赖氨酰 tRNA (Mb tRNA_{CUA}^{Py1})/赖氨酰 tRNA 合成酶 (Mb Py1RS, 野生型, 其氨基酸序列为 SEQ ID NO:2) 对。

[0065] Mb Py1RS 突变库构建在卡纳霉素抗性 pBk 质粒 (购自美国 scripps 研究所 Peter G. Schultz 实验室) 中, 位于该质粒上 *E. coli* 谷氨酰胺合成酶的启动子和终止子之间。所使用的合成酶突变库为 pBk-CpK 库, 该突变库的构建方法为: 在 Mb Py1RS 基因上挑选 5 个位点 (L266, L270, Y271, L274 和 C313), 通过 PCR 的方法进行随机突变。然后进行正负筛选来进化特异性识别 CpK 的氨酰基-tRNA 合成酶。正筛选质粒包含 Mb tRNA_{CUA}^{Py1}, TAG 突变的氯霉素乙酰转移酶基因, 启动表达绿色荧光蛋白的琥珀突变的 T7RNA 聚合酶, 四环素抗性基因。负筛选质粒包含 Mb tRNA_{CUA}^{Py1}, 在阿拉伯糖操纵子下的琥珀突变芽孢杆菌 RNA 酶基因, 以及氨苄青霉素抗性基因。进行 3 轮正筛选和 2 轮负筛选: 包含有正筛选质粒的 *E. coli* DH10B 细胞作为正筛选寄主细胞。细胞电转 pBk-CpK 库, SOC 培养基 (2% (W/V) 胰蛋白胨, 0.5% (W/V) 酵母粉, 0.05% (W/V) NaCl, 2.5mM KCl, 10mM MgCl₂, 20mM 葡萄糖) 在 37 $^{\circ}$ C 培养 1 小时。之后换用极限培养基 (GMM 极限培养基的配方: M9 盐/甘油: 764g Na₂HPO₄·7H₂O 或者 30g Na₂HPO₄, 15g KH₂PO₄, 2.5g NaCl, 5g NH₄Cl, 50ml 甘油, 高压灭菌, pH 7.0; 1mM MgSO₄: 高压灭菌; 50mM CaCl₂: 高压灭菌; 25mM FeCl₂: 过滤灭菌; 0.3M 亮氨酸: 溶解于 0.3M NaOH 中, 过滤灭菌; 1L 液体 GMM 培养基: 200ml M9 盐/甘油, 2ml MgSO₄, 2ml CaCl₂, 2ml FeCl₂, 1ml 亮氨酸) 洗两次, 铺板固体极限培养基 (在液体 GMM 培养基中加入 500ml 3% 琼脂粉水溶液, 1mM CpK, 50mg/L 卡那霉素, 60mg/L 氯霉素, 15mg/L 四环素), 37 $^{\circ}$ C 培养 60 小时。收取细胞, 提取质粒 DNA, 电泳分离, 胶回收。然后, 将经过正筛选的 pBk-CpK 转化到包含负筛选质粒的 DH10B 感受态细胞中。SOC 培养基中恢复 1 小时。之后涂板包含 0.2% 阿拉伯糖 (购自 sigma 公司) 的 LB 固体培养基 (每升培养基含 10g 胰蛋白胨, 5g 酵母粉, 10g NaCl)。37 $^{\circ}$ C 培养 8-12 小时。

[0066] 最后一轮正筛选挑 384 个克隆, 分别点板在含有 1mM CpK、氯霉素 60, 80, 100,

120mg/L 的 GMML 固体培养基上,及不包含 CpK、但包含氯霉素 0, 20, 40, 60mg/L 的 GMML 固体培养基。挑选在在 1mM CpK 120mg/L 氯霉素的培养基上生长,而在 0mM CpK 40mg/L 氯霉素培养基中不生长的克隆进行进一步验证。最终挑出 1 个克隆,插入 CpK 效率最高,测序表明,克隆所包含的氨酰基-tRNA 合成酶突变体 (CpKRS) 的氨基酸序列为 SEQ ID NO :4 所示,其中突变位点为 L266M, L270I, Y271L, L274A 和 C313I。

[0067] 实施例 3 :表达 CpK-肌红蛋白并进行体外光点击反应

[0068] 将正交 tRNA (SEQ ID NO :1) 和编码肌红蛋白的核苷酸序列 (4TAG) (SEQ ID NO :5) 构建到 pBAD 载体 (购自美国 scripps 研究所 Peter G. Schultz 实验室) 上,筛选出来的编码 CpKRS 的核苷酸序列 (SEQ ID NO :3) 构建到 pBK 载体 (购自美国 scripps 研究所 Peter G. Schultz 实验室) 上,然后共转化到 DH10B 细胞 (购自全式金公司) 中。挑取单个克隆在 37°C 培养到 OD₆₀₀ 约等于 0.5 时,向 LB 培养基中加入 1mM CpK 及 0.2% 阿拉伯糖 (购自 sigma 公司) 培养细胞,对照不加入 CpK。6-7 小时之后,收菌, Ni-NTA 纯化蛋白,并用 SDS-PAGE 电泳分析 (图 4a)。

[0069] 我们发现,只有在存在 CpK 的培养基中才能纯化出全长的肌红蛋白,这说明筛选出来的 CpKRS 可以特异性的识别 CpK。在 LB 培养基中 CpK-肌红蛋白的产率为 2mg/L,而野生型肌红蛋白的产率为 33mg/L。为了检测 CpK 仅仅插入到肌红蛋白的 4 位琥珀突变位点,我们对 CpK-肌红蛋白进行了 ESI-TOF 质谱检测,检测结果分子量为 18476Da (图 4b),与计算的分子量 18476Da 吻合。为了进一步验证 CpK 插入的特异性,我们做了二级质谱验证, b 离子和 y 离子分布显示 CpK 特异插入到肌红蛋白的第四位而不是其他位置 (图 4c)。

[0070] 为了验证 CpK 能否和四唑类化合物进行光点击反应,我们取纯化的 CpK-肌红蛋白 (0.5mg/mL) 与四唑类化合物 3 (100 μM) 在 PBS 缓冲液中孵育,然后用 302nm 手提紫外灯分别照射不同时间 (1-12 分钟),将反应后的混合物进行 SDS-PAGE 电泳。待电泳结束后,取出凝胶,在紫外灯下观察是否有荧光。观察结果表明,随着照射时间的延长,荧光强度逐渐增加,在 10min 时达到最高值 (图 5b)。

[0071] 实施例 4 :表达 CpK-增强型绿色荧光蛋白并进行体内光点击反应

[0072] 既然细菌和哺乳动物细胞中均存在氨酰 tRNA 合成酶 /tRNA 正交对,我们接下来验证了正交 tRNA/正交氨酰 tRNA 合成酶 (Mb tRNA_{CUA}^{Py1}/MbAcrKRS) 是否可以将 CpK 遗传掺入到哺乳动物细胞蛋白中。我们把 Mb tRNA_{CUA}^{Py1} (SEQ ID NO :1) 和 CpKRS 编码序列 (SEQ ID NO :3) 克隆到 pCMV-NBK-1 载体 (购自美国 scripps 研究所 Peter G. Schultz 实验室) 上,获得重组构建体 pCMV-tRNA-CpKRS,由 CMV 启动子控制 CpKRS 的表达, U6 控制 Mb tRNA_{CUA}^{Py1} 的转录。同时按照常规分子克隆方法构建 pSwan-EGFP 质粒 (pSwan 质粒购自美国 scripps 研究所 Peter G. Schultz 实验室),并在 EGFP 的 37 位引入 TAG 突变 (SEQ ID NO :7)。在 35-mm 玻底皿 (购自杭州生友公司) 上用 10% FBS DMEM 高糖培养基 (购自 Invitrogen 公司) 培养人胚肾 (HEK) 293 细胞 (购自中国协和医科大学基础医学院),待其生长至 50-60% 时,用 Lipofectamine 2000 试剂 (购自 Invitrogen 公司) 将质粒 pCMV-tRNA-CpKRS 和 pSwan-EGFP37TAG 共转染到 293 细胞中,在加入 4mM CpK 的情况下培养细胞 36 小时后,再加入 40 μM 四唑化合物 4,培养 1.5 小时,然后用 365nm 手提紫外灯照射 2 分钟,在荧光显微镜下观察,并以未加 CpK 或者四唑化合物 4 的细胞作为对照。结果如图 6 所示:在 EGFP 通道 (488nm 激发, 499-578nm 发射) 下,只有加入 CpK 的细胞可以观察到荧光信号,而没有

CpK 的情况下无荧光信号,表明 CpK 成功地通过 Mb tRNA_{CUA}^{Py1}/Mb AcrKRS 正交对掺入到哺乳动物细胞的遗传密码子编码的蛋白序列中;在吡唑啉通道(pyrazoline channel)(405nm 激发,410-498nm 发射)下,只有同时加入 CpK 和四唑化合物 4 的细胞才能发出氰色荧光,证明 CpK 在哺乳动物细胞内也能进行高效地光点击反应,从而可以特异性地标记蛋白。

[0073] 应该理解,尽管参考其示例性的实施方案,已经对本发明进行具体地显示和描述,但是本领域的普通技术人员应该理解,在不背离由后附的权利要求所定义的本发明的精神和范围的条件下,可以在其中进行各种形式和细节的变化,可以进行各种实施方案的任意组合。

[0001]

IB126226序列表.txt

序列表

<110> 中国科学院生物物理所

<120> N ϵ -(1-甲基环丙-2-烯酰胺)-赖氨酸翻译系统及其应用

<130> IB126226

<160> 8

<170> PatentIn version 3.1

<210> 1

<211> 72

<212> DNA

<213> 正交tRNA

<400> 1

gggaacctga icatgtagat cgaatggact ctaaaccgt tcagccgggt tagattcccg 60

gggtttccgc ca 72

<210> 2

<211> 419

<212> PRT

<213> 野生型赖氨酰tRNA合成酶 (Mb PyIRS), 来源于巴氏甲烷八叠球菌

<400> 2

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

[0002]

Val Val Ser Ala Pro Lys Val Lys Lys Ala Met Pro Lys Ser Val Ser
 100 105 110

Arg Ala Pro Lys Pro Leu Glu Asn Ser Val Ser Ala Lys Ala Ser Thr
 115 120 125

Asn Thr Ser Arg Ser Val Pro Ser Pro Ala Lys Ser Thr Pro Asn Ser
 130 135 140

Ser Val Pro Ala Ser Ala Pro Ala Pro Ser Leu Thr Arg Ser Gln Leu
 145 150 155 160

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

Met Ala Lys Pro Phe Arg Glu Leu Glu Pro Glu Leu Val Thr Arg Arg
 180 185 190

Lys Asn Asp Phe Gln Arg Leu Tyr Thr Asn Asp Arg Glu Asp Tyr Leu
 195 200 205

Gly Lys Leu Glu Arg Asp Ile Thr Lys Phe Phe Val Asp Arg Gly Phe
 210 215 220

Leu Glu Ile Lys Ser Pro Ile Leu Ile Pro Ala Glu Tyr Val Glu Arg
 225 230 235 240

Met Gly Ile Asn Asn Asp Thr Glu Leu Ser Lys Gln Ile Phe Arg Val
 245 250 255

Asp Lys Asn Leu Cys Leu Arg Pro Met Leu Ala Pro Thr Leu Tyr Asn
 260 265 270

Tyr Leu Arg Lys Leu Asp Arg Ile Leu Pro Gly Pro Ile Lys Ile Phe
 275 280 285

Glu Val Gly Pro Cys Tyr Arg Lys Glu Ser Asp Gly Lys Glu His Leu
 290 295 300

Glu Glu Phe Thr Met Val Asn Phe Cys Gln Met Gly Ser Gly Cys Thr
 305 310 315 320

Arg Glu Asn Leu Glu Ala Leu Ile Lys Glu Phe Leu Asp Tyr Leu Glu
 325 330 335

Ile Asp Phe Glu Ile Val Gly Asp Ser Cys Met Val Tyr Gly Asp Thr
 340 345 350

Leu Asp Ile Met His Gly Asp Leu Glu Leu Ser Ser Ala Val Val Gly
 355 360 365

Pro Val Ser Leu Asp Arg Glu Trp Gly Ile Asp Lys Pro Trp Ile Gly
 370 375 380

Ala Gly Phe Gly Leu Glu Arg Leu Leu Lys Val Met His Gly Phe Lys
 385 390 395 400

[0003]

Asn Ile Lys Arg Ala Ser Arg Ser Glu Ser Tyr Tyr Asn Gly Ile Ser
 405 410 415

Thr Asn Leu

<210> 3

<211> 1260

<212> DNA

<213> 正交氨基-tRNA合成酶 (CpKRS) 的基因序列

<400> 3

```

atggataaaa aaccgctgga tgtgctgatt agcgcgaccg gectgtggat gagccgtacc 60
ggcaccctgc ataaaatcaa acatcatgaa gtgagccgca gcaaaatcta tattgaaatg 120
gcgctcggcg atcatctggt ggtgaacaac agccgtagct gccgtaccgc gcgtgcgctt 180
cgtcatcata aataccgcaa aacctgcaaa cgttgccgtg tgagcggatga agatatacaac 240
aaatttctga cccgtagcac cgaaagcaaa aacagcgtga aagtgcgtgt ggtgagcgcg 300
ccgaaagtga aaaaagcgat gccgaaaagc gtgagccgtg cgcgaaacc gcctgaaaaat 360
agcgtgagcg cgaaagcgag caccaacacc agccgtagcg ttccgagccc ggcgaaaagc 420
accccgaaca gcagcgttcc ggctctctgcg ccggcaccga gcctgaccgc cagccagctg 480
gatcgtgtgg aagcgtctct gtcctccgaa gataaaatta gcctgaacat ggcgaaaccg 540
ttcgtgaaac tggaaccgga actggtgacc cgtctgtaaaa acgattttca gcgccgtgat 600
accaacgate gtgaagatta tctgggcaaa ctggaacgtg atatacacia attttttctg 660
gatcgcggct ttctggaat taaaagcccg attctgatte cggcggaaata tgtggaacgt 720
atgggcatta acaacgacac cgaactgagc aaacaaattt tccgcgtgga taaaaacctg 780
tgccctgcgtc cgatgatggc cccgaccatt tttaactatg ctctgaaact ggatcgtatt 840
ctgccgggtc cgatcaaaat ttttgaagtg ggcccgtgct atcgcaaaga aagcgtatgc 900
aaagaacacc tggaagaatt caccatggtt aactttatc aaatgggcag cgctgcacc 960
cgtgaaaaac tggaagcgt gatecaaaaga ttcttgatt atctggaat cgaactcgaa 1020
attgtgggag atagctgcat ggigtatggc gataccctgg atattatgca tggcgtatct 1080
gaactgagca gcgcggtggt gggctccggtt agcctggatc gtgaatgggg cattgataaa 1140
ccgiggattg gcgcgggttt tggcctggaa cgtctgctga aagtgatgca tggettcaaa 1200
aacattaaac gtgcgagccg tagcgaaagc tactataacg gcattagcac gaacctgtaa 1260

```

<210> 4

<211> 419

<212> PRT

<213> 正交氨基-tRNA合成酶 (CpKRS)

<400> 4

Met Asp Lys Lys Pro Leu Asp Val Leu Ile Ser Ala Thr Gly Leu Trp
 1 5 10 15

[0004]

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

Arg Ser Lys Ile Tyr Ile Glu Met Ala Cys Gly Asp His Leu Val Val
35 40 45

Asn Asn Ser Arg Ser Cys Arg Thr Ala Arg Ala Phe Arg His His Lys
50 55 60

Tyr Arg Lys Thr Cys Lys Arg Cys Arg Val Ser Gly Glu Asp Ile Asn
65 70 75 80

Asn Phe Leu Thr Arg Ser Thr Glu Ser Lys Asn Ser Val Lys Val Arg
85 90 95

Val Val Ser Ala Pro Lys Val Lys Lys Ala Met Pro Lys Ser Val Ser
100 105 110

Arg Ala Pro Lys Pro Leu Glu Asn Ser Val Ser Ala Lys Ala Ser Thr
115 120 125

Asn Thr Ser Arg Ser Val Pro Ser Pro Ala Lys Ser Thr Pro Asn Ser
130 135 140

Ser Val Pro Ala Ser Ala Pro Ala Pro Ser Leu Thr Arg Ser Gln Leu
145 150 155 160

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

Met Ala Lys Pro Phe Arg Glu Leu Glu Pro Glu Leu Val Thr Arg Arg
180 185 190

Lys Asn Asp Phe Gln Arg Leu Tyr Thr Asn Asp Arg Glu Asp Tyr Leu
195 200 205

Gly Lys Leu Glu Arg Asp Ile Thr Lys Phe Phe Val Asp Arg Gly Phe
210 215 220

Leu Glu Ile Lys Ser Pro Ile Leu Ile Pro Ala Glu Tyr Val Glu Arg
225 230 235 240

Met Gly Ile Asn Asn Asp Thr Glu Leu Ser Lys Gln Ile Phe Arg Val
245 250 255

Asp Lys Asn Leu Cys Leu Arg Pro Met Met Ala Pro Thr Ile Phe Asn
260 265 270

Tyr Ala Arg Lys Leu Asp Arg Ile Leu Pro Gly Pro Ile Lys Ile Phe
275 280 285

Glu Val Gly Pro Cys Tyr Arg Lys Glu Ser Asp Gly Lys Glu His Leu
290 295 300

Glu Glu Phe Thr Met Val Asn Phe Ile Gln Met Gly Ser Gly Cys Thr
305 310 315 320

[0005]

Arg Glu Asn Leu Glu Ala Leu Ile Lys Glu Phe Leu Asp Tyr Leu Glu
 325 330 335

Ile Asp Phe Glu Ile Val Gly Asp Ser Cys Met Val Tyr Gly Asp Thr
 340 345 350

Leu Asp Ile Met His Gly Asp Leu Glu Leu Ser Ser Ala Val Val Gly
 355 360 365

Pro Val Ser Leu Asp Arg Glu Trp Gly Ile Asp Lys Pro Trp Ile Gly
 370 375 380

Ala Gly Phe Gly Leu Glu Arg Leu Leu Lys Val Met His Gly Phe Lys
 385 390 395 400

Asn Ile Lys Arg Ala Ser Arg Ser Glu Ser Tyr Tyr Asn Gly Ile Ser
 405 410 415

Thr Asn Leu

<210> 5

<211> 492

<212> DNA

<213> 含有CpK的肌红蛋白突变体 (Mb-4CpK) 的编码核苷酸序列

<400> 5
 atggttctgt aggaaggatga atggcagctg gttctgcatg ttgggctaa agttgaagct 60
 gacgtcctgt gtcattggca ggacatcttg attcagctgt tcaaatctca tccggaaaact 120
 ctgaaaaaat tcgactgttt caaacatctg aaaactgaag ctgaaatgaa agcttctgaa 180
 gatctgaaaa aacatgggtgt taacctgtta actgccctag gtctatcct taagaaaaaa 240
 gggcatcatg aagctgagct caaacctgtt gcacaatcgc atctactaa acataagatc 300
 ccgatcaaat acctggaatt catctctgaa gcatcatcc atgttctgca ttctagacat 360
 ccaggtgact tcggtctga cgctcagggt gctatgaaca aagctctega gctgttccgt 420
 aaagatateg ctgctaagta caaagaactg ggttaccagg gtggtcggg acatcatcac 480
 catcaaccatt ga 492

<210> 6

<211> 163

<212> PRT

<213> 含有CpK的肌红蛋白突变体 (Mb-4CpK) 的氨基酸序列，其中*表示在第4位引入的CpK

<400> 6

Met Val Leu * Glu Gly Glu Trp Gln Leu Val Leu His Val Trp Ala
 1 5 10 15

[0006]

Lys Val Glu Ala Asp Val Ala Gly His Gly Gln Asp Ile Leu Ile Arg
 20 25 30
 Leu Phe Lys Ser His Pro Glu Thr Leu Glu Lys Phe Asp Arg Phe Lys
 35 40 45
 His Leu Lys Thr Glu Ala Glu Met Lys Ala Ser Glu Asp Leu Lys Lys
 50 55 60
 His Gly Val Thr Val Leu Thr Ala Leu Gly Ala Ile Leu Lys Lys Lys
 65 70 75 80
 Gly His His Glu Ala Glu Leu Lys Pro Leu Ala Gln Ser His Ala Thr
 85 90 95
 Lys His Lys Ile Pro Ile Lys Tyr Leu Glu Phe Ile Ser Glu Ala Ile
 100 105 110
 Ile His Val Leu His Ser Arg His Pro Gly Asp Phe Gly Ala Asp Ala
 115 120 125
 Gln Gly Ala Met Asn Lys Ala Leu Glu Leu Phe Arg Lys Asp Ile Ala
 130 135 140
 Ala Lys Tyr Lys Glu Leu Gly Tyr Gln Gly Gly Ser Gly His His His
 145 150 155 160
 His His His

<210> 7

<211> 861

<212> DNA

<213> 含有CpK的增强型绿色荧光蛋白突变体(EGFP-37CpK)的核苷酸序列

<400> 7

atgagtattc aacatttcg igtgcccett attccctttt ttgcggcatt ttgccttcc 60
 gtttttgcgc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgta 120
 ggagtgggtt acatcgaact ggatcacaac agcggtaaga tccttgagag ttttgccecc 180
 gaagaacgtt ttccaatgat gageactttt aaagtctctgc tatgtgcegc ggtattatcc 240
 cgtattgacg ccgggcaaga gcaactcggc cgcgcatac actattctca gaatgacttg 300
 gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta 360
 tgcagtgetg ccataacct gagtgataac actgcggcca acttaettct gacaacgate 420
 ggaggaccga aggagcfaac cgcctttttg cacaacatgg gggatcatgt aactcgcctt 480
 gatcgttggg aaccggagct gaatgaagcc ataccaaaac acgagcgtga caccacgatg 540
 cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct 600
 tcccggcaac aattaataga ctggatggag cgggataaag ttgcaggacc acttctgcgc 660
 tcggcccttc cggctggctg gtttattgct gataaatctg gagecggatga gcgtgggtct 720

[0007]

cgcggtataca ttgcagcaact ggggccagat ggtaagccct cccgtaicgt agttatctac 780
 acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgetga gataggtgcc 840
 tcactgatta agcattggta a 861

<210> 8

<211> 359

<212> PRT

<213> 含有CpK的增强型绿色荧光蛋白突变体 (EGFP-37CpK) 的氨基酸序列, 其中*表示在第37位引入的CpK

<400> 8

Met Ser Thr Asn Leu Ser Val Ile Lys Asn Pro Arg Val Gln Ser Asp
 1 5 10 15

Gln Arg Arg Leu Val Arg Arg Pro Asp Val Lys Pro Asn Ile Pro Leu
 20 25 30

Ile Val Ile Leu * Ser Thr Leu Glu Asp Pro Arg Val Pro Val Ala
 35 40 45

Thr Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile
 50 55 60

Leu Gly Arg Ala Gly Arg Arg Arg Lys Arg Pro Gln Val Gln Arg Val
 65 70 75 80

Arg Arg Gly Arg Gly Arg Cys His Leu Arg Gln Ala Asp Pro Glu Val
 85 90 95

His Leu His His Arg Gln Ala Ala Arg Ala Leu Ala His Pro Arg Asp
 100 105 110

His Pro Asp Leu Arg Arg Ala Val Leu Gln Pro Leu Pro Arg Pro His
 115 120 125

Glu Ala Ala Arg Leu Leu Gln Val Arg His Ala Arg Arg Leu Arg Pro
 130 135 140

Gly Ala His His Leu Leu Gln Gly Arg Arg Gln Leu Gln Asp Pro Arg
 145 150 155 160

Arg Gly Glu Val Arg Gly Arg His Pro Gly Glu Pro His Arg Ala Glu
 165 170 175

Gly His Arg Leu Gln Gly Gly Arg Gln His Pro Gly Ala Gln Ala Gly
 180 185 190

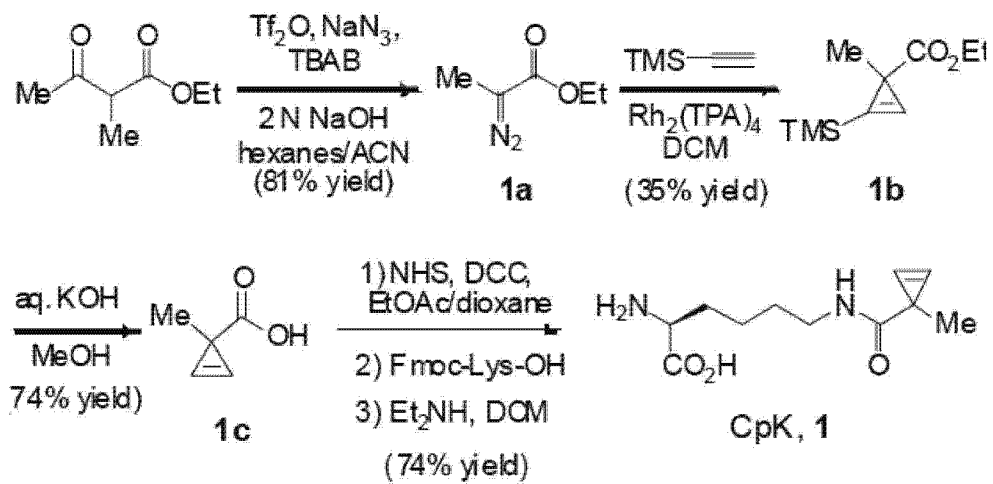
Val Gln Leu Gln Gln Pro Gln Arg Leu Tyr His Gly Arg Gln Ala Glu
 195 200 205

Glu Arg His Gln Gly Glu Leu Gln Asp Pro Pro Gln His Arg Gly Arg
 210 215 220

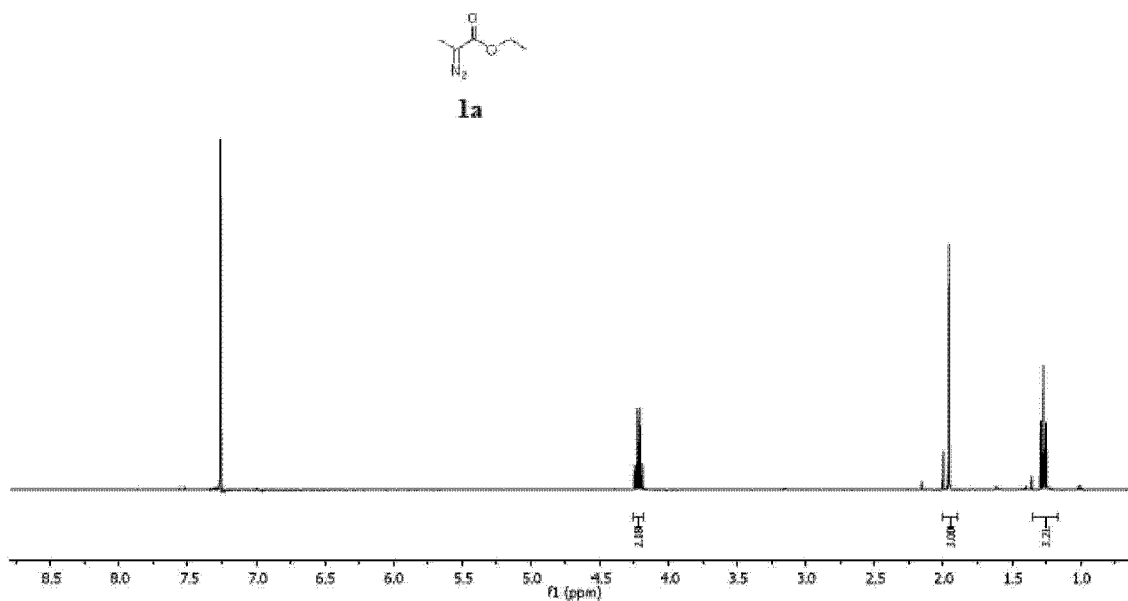
[0008]

Gln Arg Ala Ala Arg Arg Pro Leu Pro Ala Glu His Pro His Arg Arg
 225 230 235 240
 Arg Pro Arg Ala Ala Ala Arg Gln Pro Leu Pro Glu His Pro Val Arg
 245 250 255
 Pro Glu Gln Arg Pro Gln Arg Glu Ala Arg Ser His Gly Pro Ala Gly
 260 265 270
 Ile Pro Thr Glu Arg Arg Ser Leu Pro Leu Pro Thr Cys Leu Val Ser
 275 280 285
 Lys Ile Ile Gly Leu Leu Val Gly Arg Thr Gly Pro Phe Val Ser Arg
 290 295 300
 Val Ser Val Met Thr Val Lys Thr Ser Asp Thr Cys Ser Ser Arg Arg
 305 310 315 320
 Arg Ser Gln Leu Val Cys Lys Arg Met Pro Gly Ala Asp Lys Pro Val
 325 330 335
 Arg Ala Arg Gln Arg Val Leu Ala Gly Val Gly Ala Gly Leu Thr Met
 340 345 350
 Arg His Gln Ser Arg Leu Tyr
 355

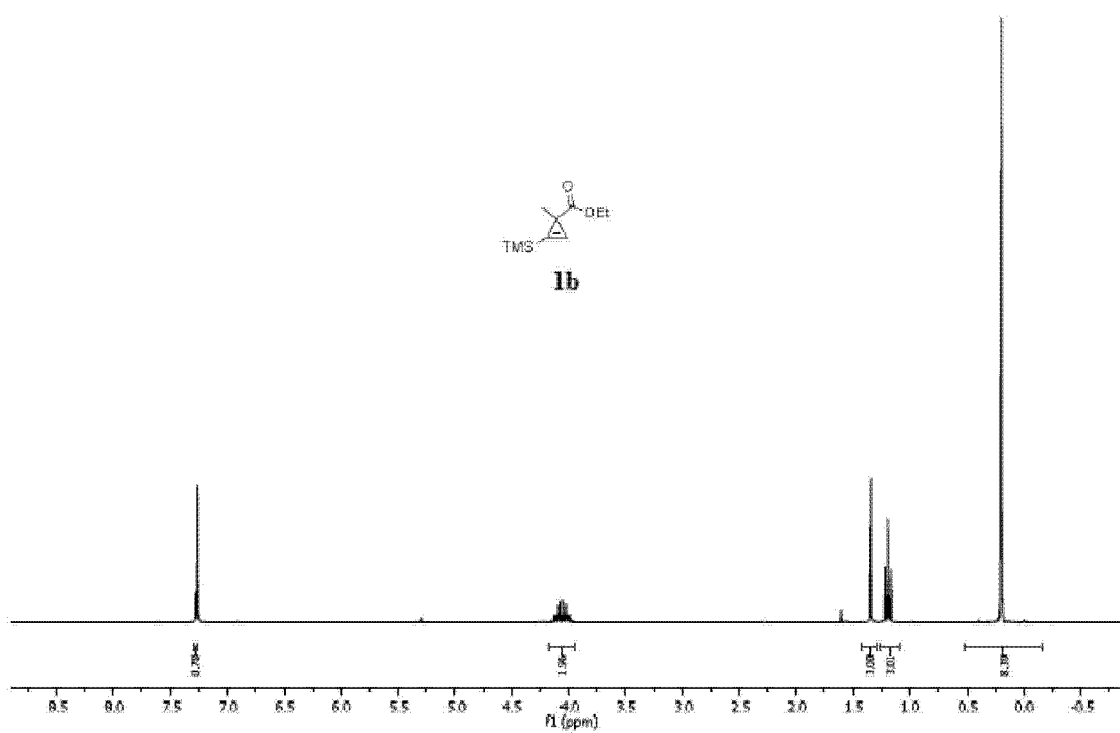
a)



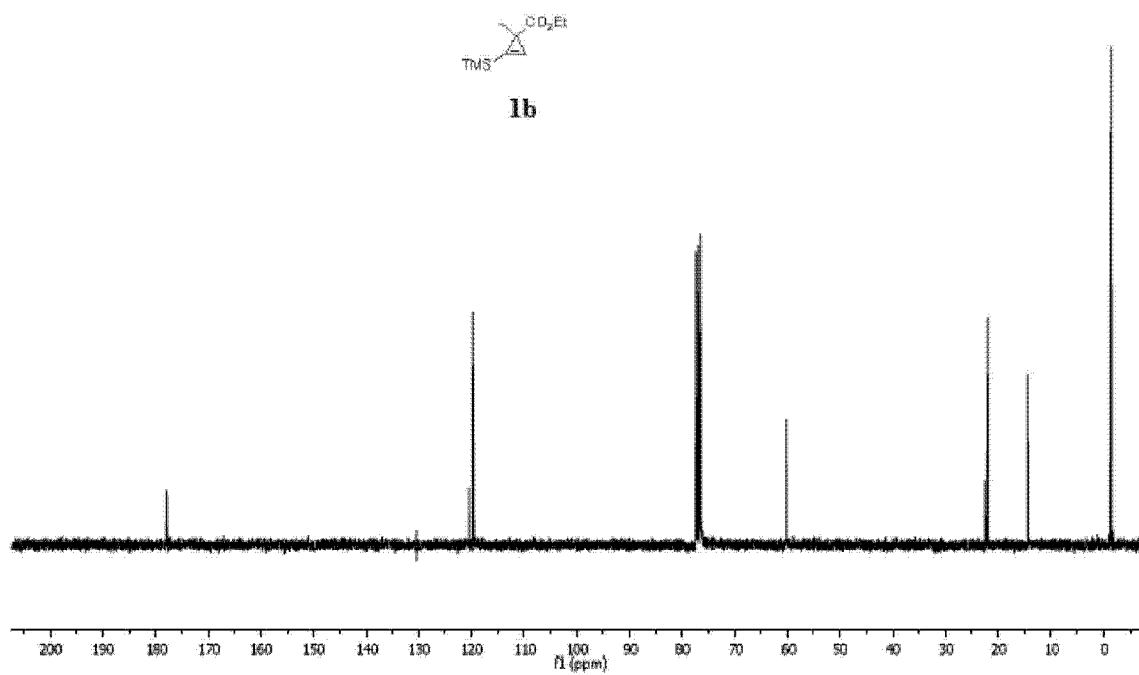
b)



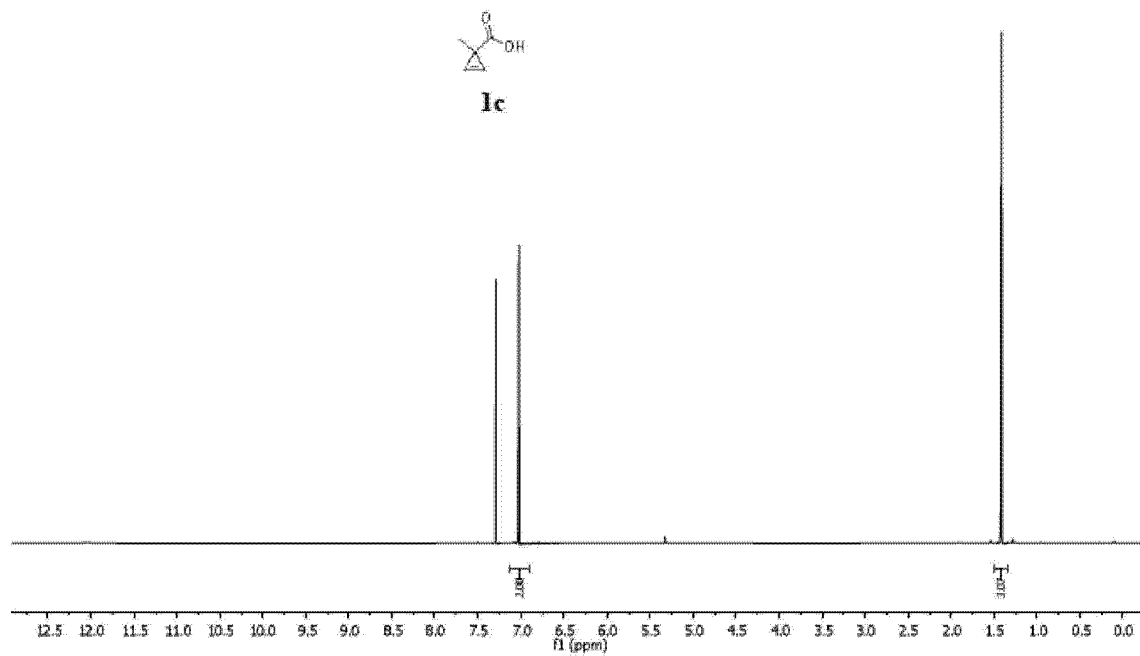
c)



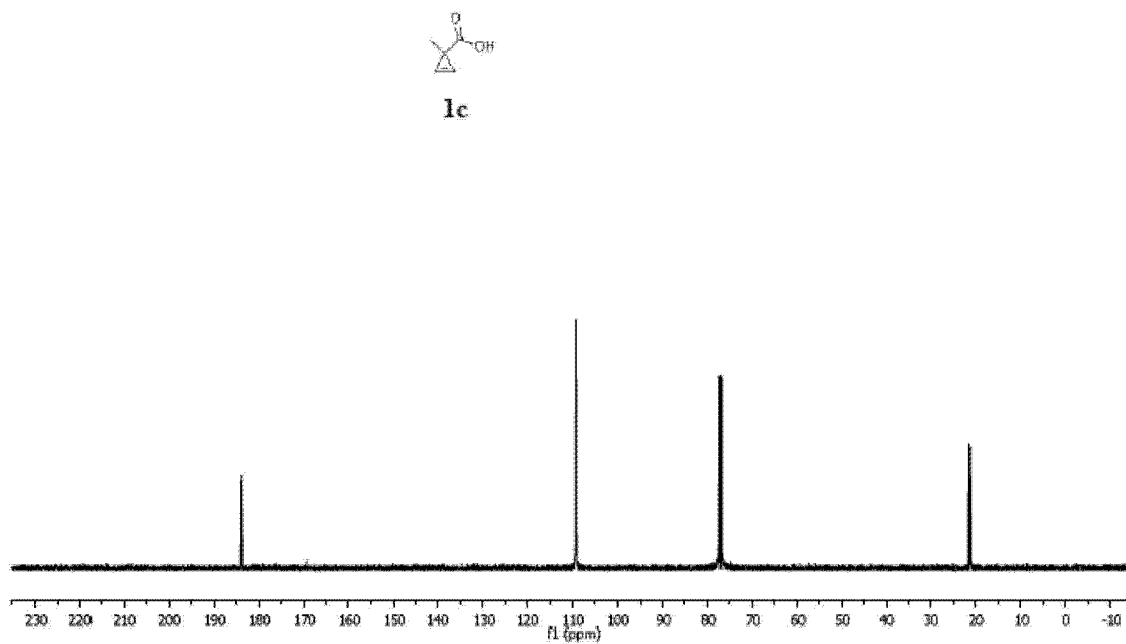
d)



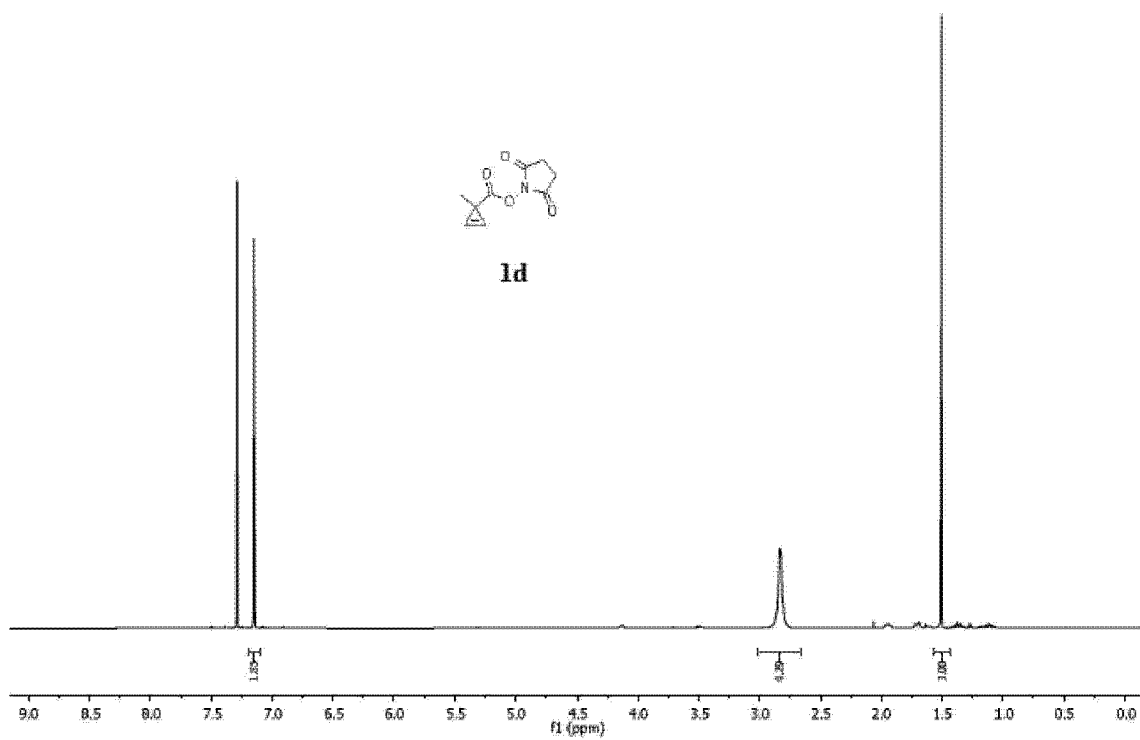
e)



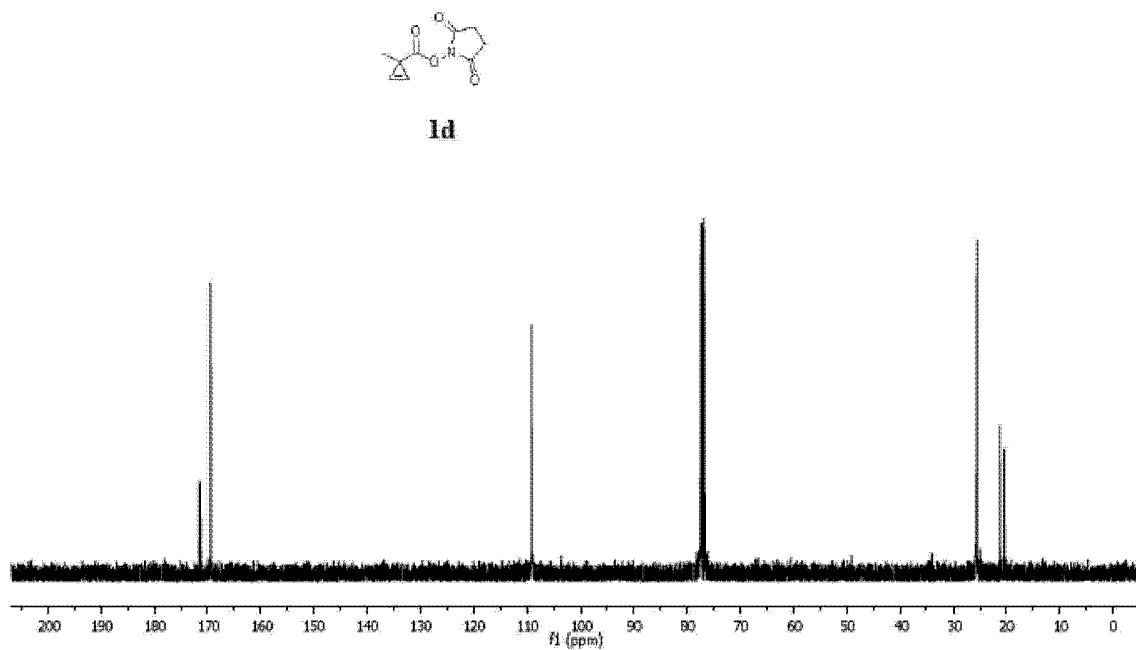
f)



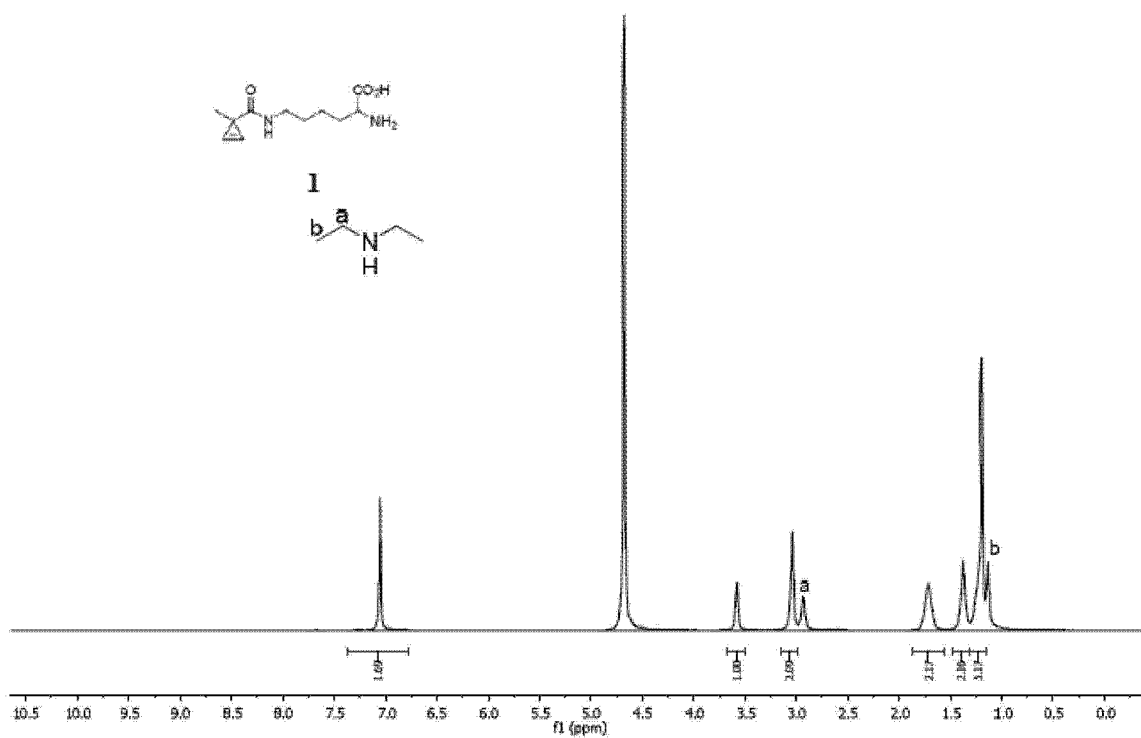
g)



h)



i)



j)

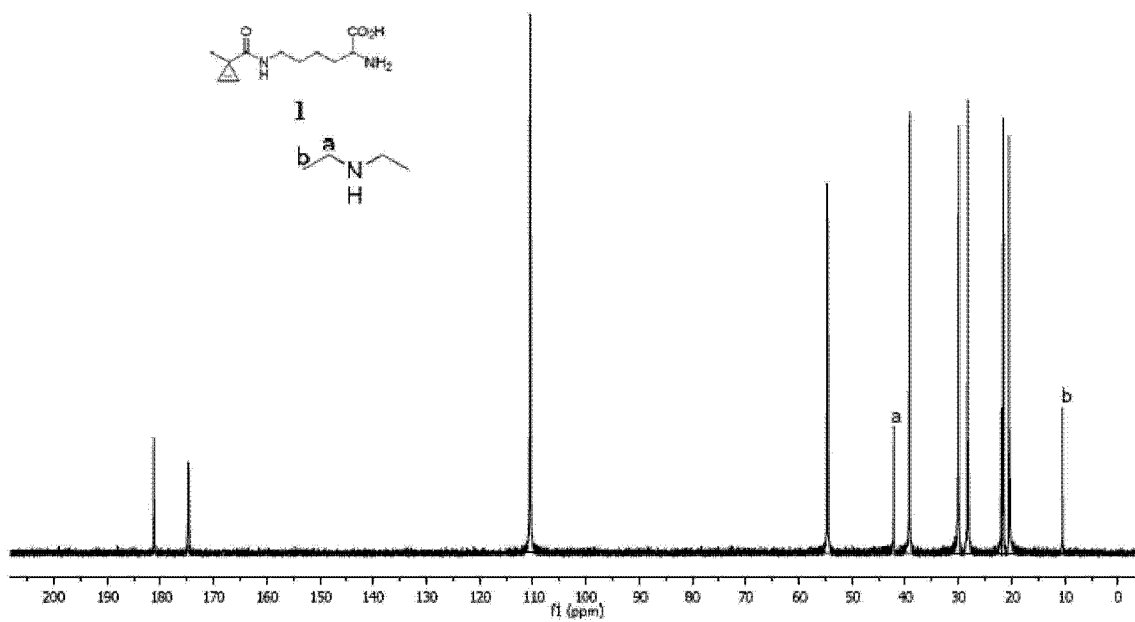
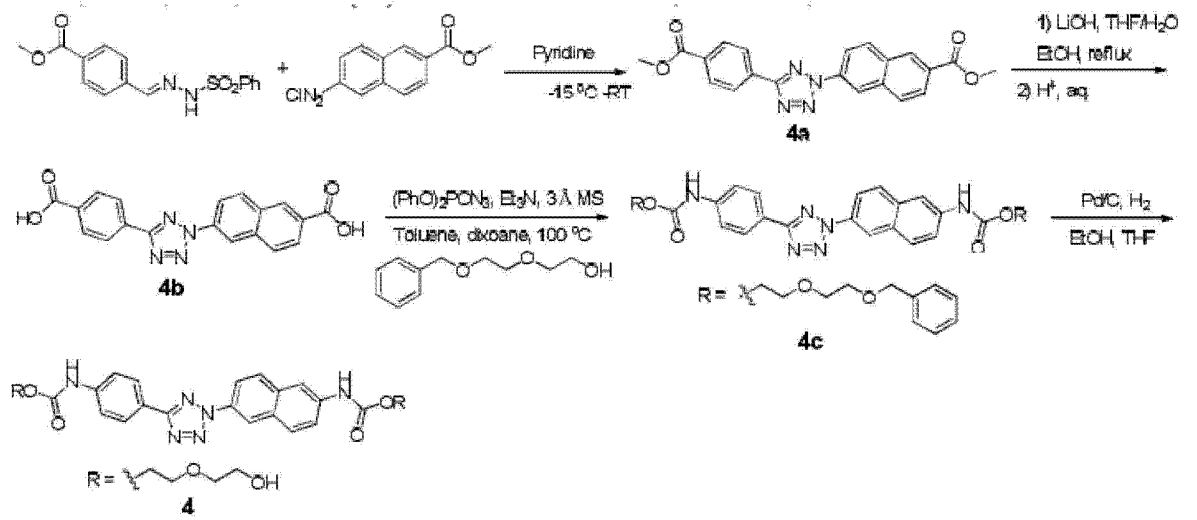
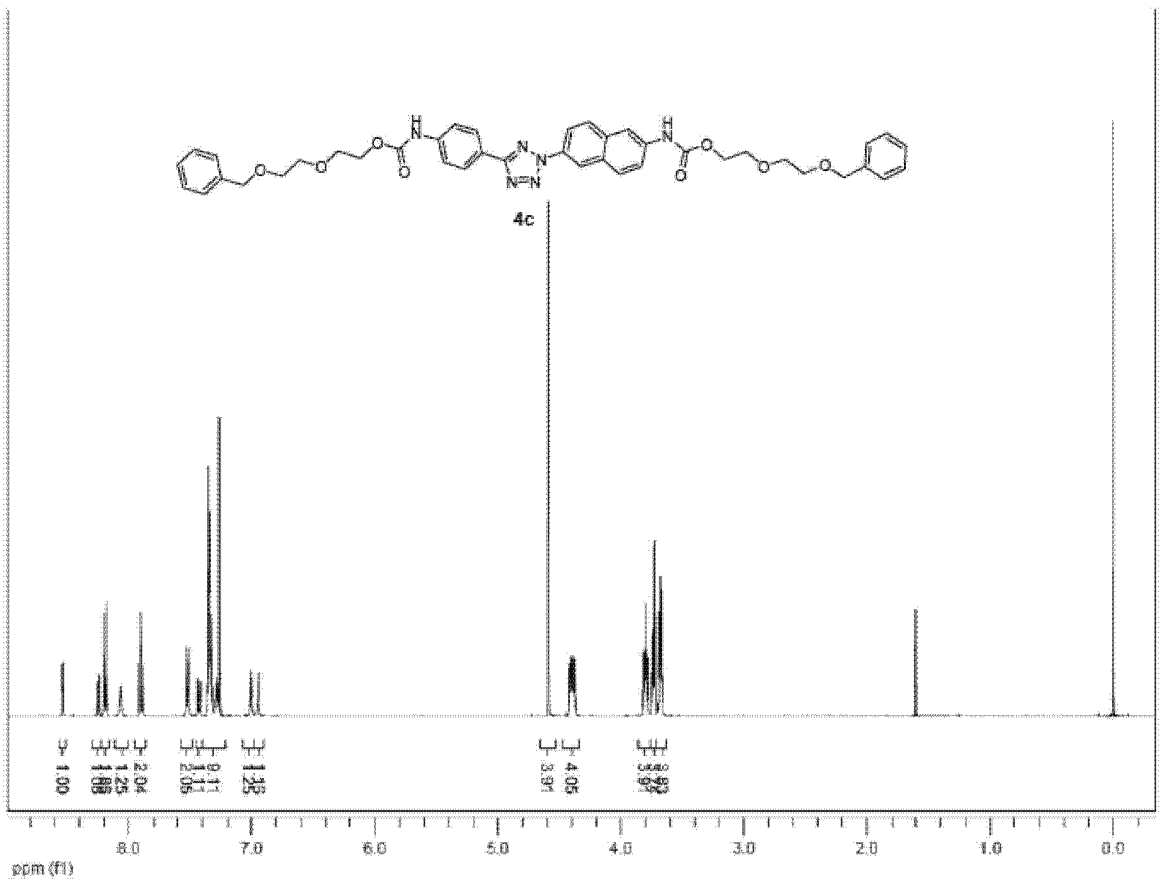


图 1

a)



b)



e)

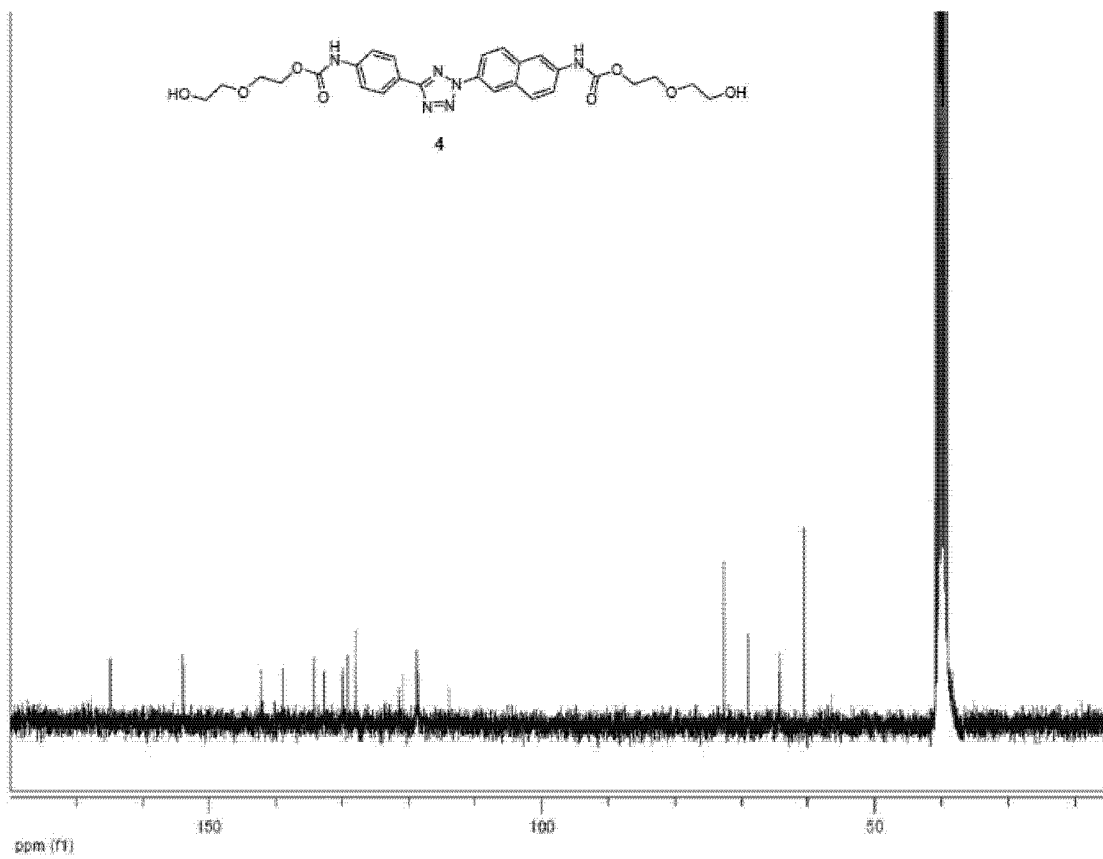
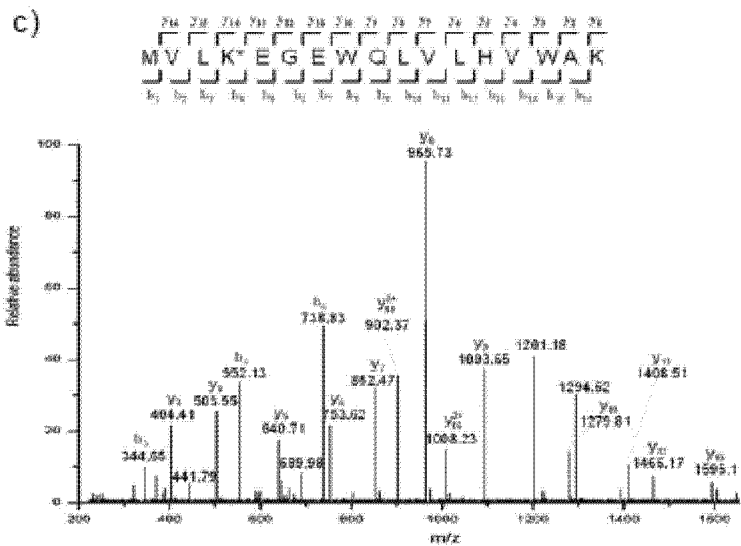
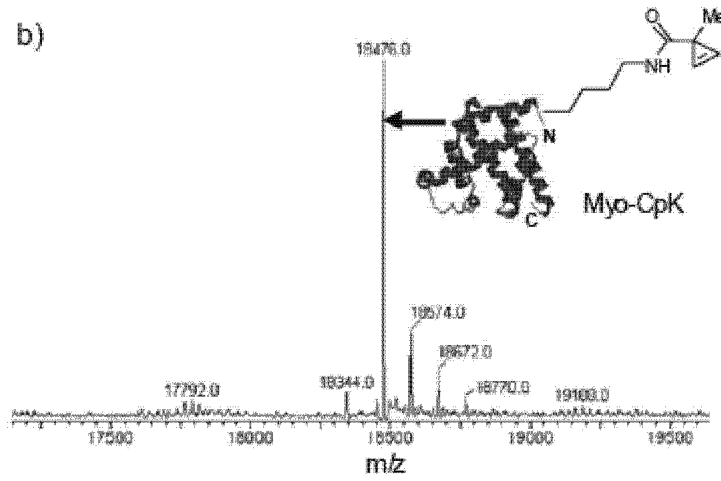
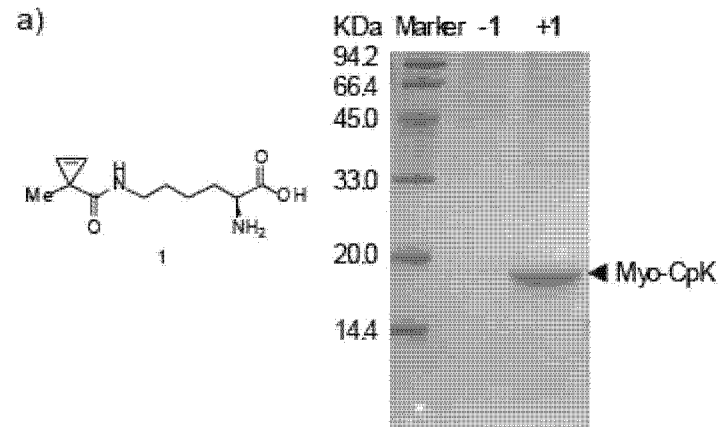


图 2

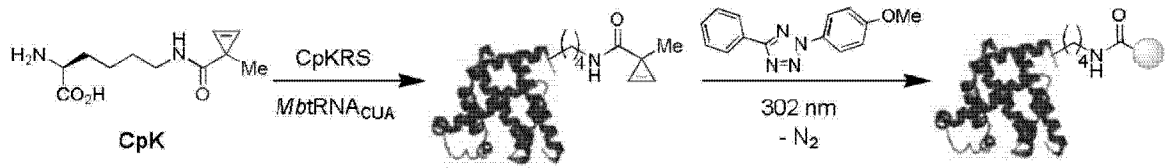
名称	核苷酸/氨基酸序列
正交 tRNA	<p><u>SEQ ID NO: 1</u> gggaacctgatcatgtagatcgaatggactctaataccgttcagccgggtagattcccggggt ttcgcca</p>
野生型赖氨酰 tRNA 合成酶 (Mb PylRS), 来源于巴氏甲烷 八叠球菌	<p><u>氨基酸序列 (SEQ ID NO: 2):</u> MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMA CGDHLVVNNSRSCRTARAFRHHKYRKTCRRCRVSDDEDINNF LTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVS AKASTNTRSVPSPAKSTPNSSVPASAPAPSLTRSQLDRVEAL LSPEDKISLNMAKPFRELEPELVTRRKNDFFQRLYTNDREDYL GKLERDITKFFVDRGFLEIKSPILIPA EYVERMGINNDTELSK QIFRVDKNLCLRPMLAPTLYNYLRKLDRLPGPIKIFEVGPCY RKESDGKEHLEEFMVNFQMGSGCTRENLEALIKEFLDYL EIDFEIVGDSCMVYGD TLDIMHGDLELSSAVVGPVSLDREW GIDKPWIGAGFGLERLLKVMHGFKNIKRASRSSESYNGIST NL</p>
正交氨酰基 -tRNA 合成酶 (CpKRS)	<p><u>核苷酸序列 (SEQ ID NO:3):</u> atggataaaaaaccgctggatgtgctgattagcgcgaccggcctgtggatgagccgtaccgg caccctgcataaaatcaaacatcatgaagtgagccgcagcaaaatctataftgaaatggcgtgc ggcgatcatctggtggtgaacaacagccgtagctgccgtaccgcgctgcttctgcatcata aataccgcaaacctgcaaacgttgccgtgtgagcggatgaagatatcaacaacttctgacccg tagcaccgaaagcaaaaacagcgtgaaagtgcgtgtggtgagcgcgccgaaagtgaaaaa gcgatgccgaaaagcgtgagccgtgcgccgaaaccgtggaaaatagcgtgagcgcgaaa gcgagcaccacaccagccgtagcgtccgagccggcgaaaagcaccggaaacagcagc gttccggcgtctgcgccggcaccgagcctgaccgcagccagctggatcgtgtggaagcgc tgcgtctccggaagataaaaattagcctgaacatggcgaaaccgttctgtaactggaaccgga actggtgaccctcgtaaaaacgatttctagcgcctgtatccaacgatcgtgaagattatctgg gcaactggaacgtgatacaccaaattttgtggatcgccgcttctggaattaaaaagcccg attctgattccggcggaatgtggaacgtatgggcattacaacgacaccgaaactgagcaaa caaatttccgctggataaaaacctgtgcctgcgtccgatgatggcccggaccattttaaactat gctcgtaaactggatcgtattctgccgggtccgatcaaaattttgaaagtggcccctgctatcg caaagaaagcgtggcaagaacacctggaagaattcaccatggtaacttattcaaatggcg agcggctgaccctgaaaacctggaagcgtgatcaagaattcctggattatctggaate gactcgaattgtggcgatagctgcatggtgtatggcgataccctggatattatgcatggcga tctggaactgagcagcgggtgtgggtccggttagcctggatcgtgaatggggcattgataa accgtgattggcgcgggtttggcctggaacgtctgtgaaagtgatgcatggctcaaaaac attaaacgtgcgagccgtagcgaagctactataacggcattagcacgaaacctgtaa</p> <p><u>氨基酸序列 (SEQ ID NO:4):</u> MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMA CGDHLVVNNSRSCRTARAFRHHKYRKTCRRCRVSGEDINNF LTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVS AKASTNTRSVPSPAKSTPNSSVPASAPAPSLTRSQLDRVEAL LSPEDKISLNMAKPFRELEPELVTRRKNDFFQRLYTNDREDYL GKLERDITKFFVDRGFLEIKSPILIPA EYVERMGINNDTELSK QIFRVDKNLCLRPMMAPTIFNYARKLDRLPGPIKIFEVGPCY RKESDGKEHLEEFMVNFIQMGSGCTRENLEALIKEFLDYL EIDFEIVGDSCMVYGD TLDIMHGDLELSSAVVGPVSLDREW GIDKPWIGAGFGLERLLKVMHGFKNIKRASRSSESYNGIST NL</p>

<p>含有 CpK 的肌红蛋白突变体 (Mb-4CpK) 的序列</p>	<p><u>核苷酸序列 (SEQ ID NO:5):</u> Atggttctgtaggaaggtgaatggcagctggttctgcatgtttgggctaaagttgaagctgacg tcgctggtcatggtcaggacatcttgatcgactgttcaaatctcatccggaaactctggaaaaat tcgatcgtttcaaacatctgaaaactgaagctgaaatgaaagcttctgaagatctgaaaaacat ggtgtaccgtgtaactgccctaggtgctatcctaagaaaaagggcacatgaagctgagct caaaccgctgcacaatgcatgctactaaacataagatcccgatcaaatacctggaattcatct ctgaagcgatcatccatgttctgcattctagacatccaggtgactcggctgctgacgctcagggt gctatgaacaaagctctcgagctgtccgtaaagatcgcgtgctaagtacaaagaactgggta ccagggtggctcgggacatcatcaccatcaccattga</p> <p><u>氨基酸序列 (SEQ ID NO:6), 其中*表示引入的 CpK</u> MVL*EGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPET LEKFDRFKHLKTEAEMKASEDLKKHGVTVLTALGAILKKK GHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGD FGADAQGAMNKALELFRKDIAAKYKELGYQGGSQHIIHHH H</p>
<p>含有 CpK 的增强型绿色荧光蛋白突变体 (EGFP-37CpK) 的序列</p>	<p><u>核苷酸序列 (SEQ ID NO:7):</u> atgagtattcaacatttccgtgtgcccttattccctttttgcggcatttgccttctgttttgctca cccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgtaggagtggttacat cgaactggatctcaacagcggtaagatccttgagagtttgcggcgaagaacgtttccaatga tgagcacttttaaagttctgctatgtggcgcgggtattatcccgtattgacgcccggcaagagcaa ctcggctgccgcatacactattctcagaatgacttggtgagtactaccagtcacagaaaagc atcttacggatggcatgacagtaagagaattatgagtgctgccataacatgagtgataaacat gcccgaacttactctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaac atgggggatcatgtaactgccttgatcgttgggaaccggagctgaatgaagccataccaac gacgagcgtgacaccacgatgcctgtagcaatggcaacaacggtgcgcaactattaactggc gaactacttacttagcttcccggcaacaattaatagactggatggagggcgataaagttgcag gaccacttctgcgctcggcccttccggctggctggttattgctgataaatctggagccggtgag cgtgggtctcgcggtatcattgcagcactggggccagatggaagccctcccgtatcgtagtta tctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgcgtgagataggt gcctcactgattaagcattggtaa</p> <p><u>氨基酸序列 (SEQ ID NO:8), 其中*表示引入的 CpK</u> mstnlsviknprvqsdqrrlvrrpdvknplivil*stledprvpvatmvskeelftgvvp ilgragrrrkrpqvrgrgrchlrqadpevhllhrqaaralahprdhpdrravlqplprp heaarllqvrharrlrpgahllqgrrqlqdprrgevrgrhpgephraeghrlqggrrqhpqag agvqlqqpqrlyhgrqaerhqgelqdpqhrgrqraarrplpaehphrrrpraarqplpe hpvrpeqrpqrearshgp agipterrslplptclvskiiigllvgrtgpfvsrvsvmtvktsdcssrrrsqvlckrmpgadkp vrrarqrvlagvgagltmrhqsrlly</p>

图 3



a)



b)

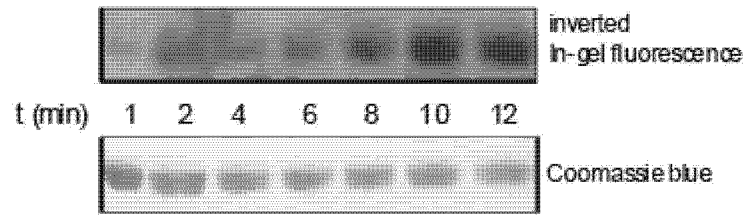


图 5

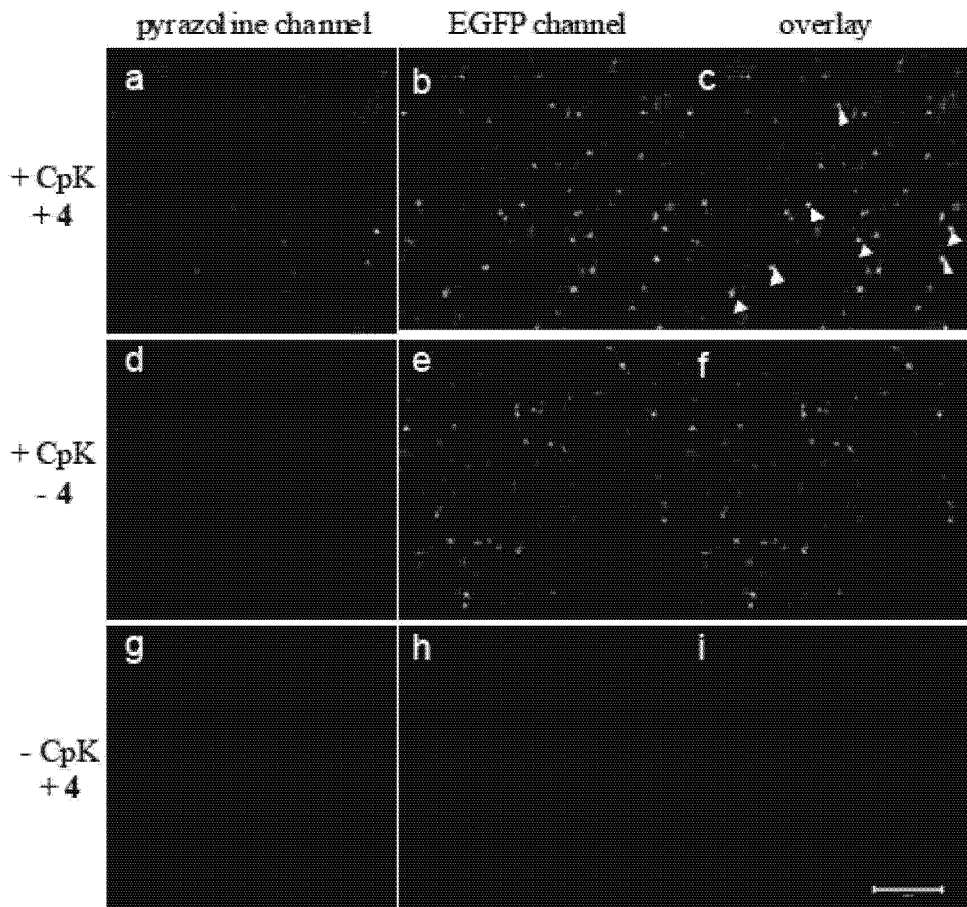


图 6