

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12N 15/74 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200780018984.4

[43] 公开日 2009 年 9 月 16 日

[11] 公开号 CN 101535482A

[22] 申请日 2007.5.21

[21] 申请号 200780018984.4

[30] 优先权

[32] 2006.5.23 [33] US [31] 60/808,099

[32] 2006.6.14 [33] US [31] 60/813,856

[86] 国际申请 PCT/US2007/012361 2007.5.21

[87] 国际公布 WO2007/139870 英 2007.12.6

[85] 进入国家阶段日期 2008.11.24

[71] 申请人 斯克利普斯研究院

地址 美国加利福尼亚州

[72] 发明人 王江云 谢建明 P·G·舒尔茨

[74] 专利代理机构 上海专利商标事务所有限公司
代理人 韦东

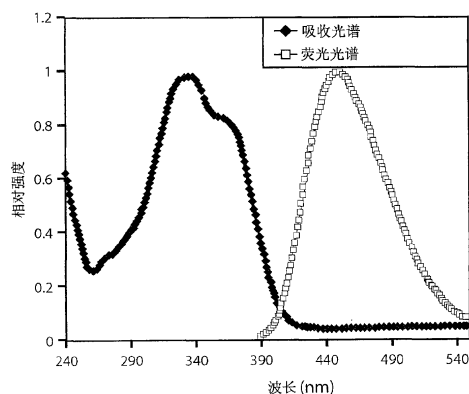
权利要求书 4 页 说明书 69 页 附图 6 页

[54] 发明名称

遗传编码的荧光香豆素氨基酸

[57] 摘要

本发明涉及能将香豆素非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸掺入在真细菌宿主细胞,例如大肠杆菌中产生的蛋白质的 tRNA 和氨酰-tRNA 正交配对。本发明提供,例如但不限于新型正交合成酶,鉴定和制备所述新型合成酶的方法,产生含有非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸的蛋白质的方法以及相关的翻译系统。



1. 一种翻译系统，其包含：

(a) 第一非天然氨基酸，即 L-(7-羟基香豆素-4-基)乙基甘氨酸非天然氨基酸；

(b) 第一正交氨酰-tRNA 合成酶(O-RS)；和

(c) 第一正交 tRNA(O-tRNA)；

其中所述第一 O-RS 用所述第一非天然氨基酸优先氨酰化所述第一 O-tRNA。

2. 如权利要求 1 所述的翻译系统，其特征在于，所述第一 O-RS 衍生自詹氏甲烷球菌 (*Methanococcus jannaschii*) 氨酰-tRNA 合成酶。

3. 如权利要求 1 所述的翻译系统，其特征在于，所述第一 O-RS 衍生自野生型詹氏甲烷球菌酪氨酰-tRNA 合成酶。

4. 如权利要求 1 所述的翻译系统，其特征在于，所述第一 O-RS 包含 SEQ ID NO: 4 所示氨基酸序列或其保守变体。

5. 如权利要求 1 所述的翻译系统，其特征在于，所述第一 O-tRNA 是琥珀抑制基因 tRNA。

6. 如权利要求 1 所述的翻译系统，其特征在于，所述第一 O-tRNA 包含 SEQ ID NO: 1 所示多核苷酸序列或由其编码。

7. 如权利要求 1 所述的翻译系统，还包括编码感兴趣蛋白质的核酸，所述核酸含有至少一个选择者密码子，其中所述选择者密码子为所述第一 O-tRNA 识别。

8. 如权利要求 7 所述的翻译系统，还包括第二 O-RS 和第二 O-tRNA，其中所述第二 O-RS 用不同于所述第一非天然氨基酸的第二非天然氨基酸优先氨酰化所述第二 O-tRNA，而其中所述第二 O-tRNA 识别不同于所述第一 O-tRNA 识别的选择者密码子的选择者密码子。

9. 如权利要求 1 所述的翻译系统，其特征在于，所述系统包括含有所述第一非天然氨基酸、所述第一 O-RS 和所述第一 O-tRNA 的宿主细胞。

10. 如权利要求 9 所述的翻译系统，其特征在于，所述宿主细胞是真

细菌细胞。

11. 如权利要求 10 所述的翻译系统，其特征在于，所述真细菌细胞是大肠杆菌细胞。

12. 如权利要求 9 所述的翻译系统，其特征在于，所述宿主细胞含有编码所述第一 O-RS 的多核苷酸。

13. 如权利要求 12 所述的翻译系统，其特征在于，所述多核苷酸包含 SEQ ID NO: 5 所示核苷酸序列。

14. 如权利要求 9 所述的翻译系统，其特征在于，所述宿主细胞包含编码所述第一 O-tRNA 的多核苷酸。

15. 一种在翻译系统内产生在选择位置上含有非天然氨基酸的蛋白质的方法，所述方法包括：

(a) 提供一种翻译系统，其包含：

(i) 第一非天然氨基酸，即 L-(7-羟基香豆素-4-基)乙基甘氨酸非天然氨基酸；

(ii) 第一正交氨酰-tRNA 合成酶(O-RS)；

(iii) 第一正交 tRNA (O-tRNA)，其中所述第一 O-RS 用所述非天然氨基酸优先氨酰化所述第一 O-tRNA；和

(iv) 编码所述蛋白质的核酸，其中所述核酸包含所述第一 O-tRNA 识别的至少一个选择者密码子；和

(b) 在所述蛋白质的翻译过程中对所述选择者密码子起反应而将所述非天然氨基酸掺入所述蛋白质的所述选择位置，从而产生在选择位置含有所述非天然氨基酸的所述蛋白质。

16. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供编码所述 O-RS 的多核苷酸。

17. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供衍生自詹氏甲烷球菌氨酰-tRNA 合成酶的 O-RS。

18. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供衍生自野生型詹氏甲烷球菌酪氨酰-tRNA 合成酶的 O-RS。

19. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括

提供含有 SEQ ID NO: 4 所示氨基酸序列的 O-RS 和其保守变体。

20. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括通过定点诱变突变野生型氨酰-tRNA 合成酶的氨基酸结合袋，并选择能用所述非天然氨基酸优先氨酰化所述 O-tRNA 的所得 O-RS。

21. 如权利要求 20 所述的方法，其特征在于，所述选择步骤包括在定点诱变之后从含有多个所得氨酰-tRNA 合成酶分子的集合中对所述 O-RS 进行正选择和负选择。

22. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供编码所述 O-tRNA 的多核苷酸。

23. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供 O-tRNA，该 O-tRNA 是琥珀抑制型 tRNA。

24. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供含有 SEQ ID NO: 1 所示多核苷酸序列或由其编码的 O-tRNA。

25. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供含有琥珀选择者密码子的核酸。

26. 如权利要求 15 所述的方法，其特征还在于，所述蛋白质含有不同于所述第一非天然氨基酸的第二非天然氨基酸，且其中所述翻译系统还包括第二 O-RS 和第二 O-tRNA，其中所述第二 O-RS 用不同于所述第一非天然氨基酸的第二非天然氨基酸优先氨酰化所述第二 O-tRNA，且其中的第二 O-tRNA 识别核酸中不同于所述第一 O-tRNA 识别的选择者密码子的选择者密码子。

27. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供宿主细胞，其中所述宿主细胞含有所述第一非天然氨基酸、所述第一 O-RS、所述第一 O-tRNA 和所述核酸，且其中所述掺入步骤包括培养所述宿主细胞。

28. 如权利要求 27 所述的方法，其特征在于，所述提供宿主细胞包括提供真细菌宿主细胞。

29. 如权利要求 28 所述的方法，其特征在于，所述提供真细菌宿主细胞包括提供大肠杆菌宿主细胞。

30. 如权利要求 27 所述的方法，其特征在于，所述提供宿主细胞包括提供含有编码所述 O-RS 的多核苷酸的宿主细胞。

31. 如权利要求 30 所述的方法，其特征在于，所述提供含有编码所述 O-RS 的多核苷酸的宿主细胞的步骤包括提供含有含 SEQ ID NO: 5 所示核苷酸序列的多核苷酸的宿主细胞。

32. 如权利要求 15 所述的方法，其特征在于，所述提供翻译系统包括提供细胞提取物。

33. 一种包含多肽的组合物，所述多肽含有 SEQ ID NO: 4 所示氨基酸序列，或其保守变体，其中所述保守变体多肽用非天然氨基酸氨酰化关联的正交 tRNA (O-tRNA)，效率至少为观察到含有所述 O-tRNA、所述非天然氨基酸和含有 SEQ ID NO: 4 所示氨基酸序列的氨酰-tRNA 合成酶的翻译系统效率的 50%。

34. 一种编码权利要求 33 所述多肽的多核苷酸。

35. 如权利要求 34 所述的多核苷酸，其特征在于，所述多核苷酸包含 SEQ ID NO: 5 所示核苷酸序列。

36. 如权利要求 33 所述的组合物，其特征在于，所述组合物是含有所述多肽的细胞。

37. 一种含有多核苷酸的组合物，所述多核苷酸包含 SEQ ID NO: 5 所示核苷酸序列。

38. 一种含权利要求 37 所述多核苷酸的载体。

39. 一种含权利要求 37 所述多核苷酸的表达载体。

40. 一种包含载体的细胞，所述载体含权利要求 37 所述的多核苷酸。

遗传编码的荧光香豆素氨基酸

相关申请的交叉引用

本申请要求以下申请的优先权：2006年5月23日提交的美国临时申请序列号 60/808,099；和 2006年6月14日提交的美国临时申请序列号 60/813,856；两篇申请的内容通过引用全文纳入本文。

对联邦资助研发下所作发明的权利的声明

本发明在能源部合同号 ER46051 的政府支持下做出。政府对本发明享有一定权利。

发明领域

本发明属于翻译生物化学领域。本发明涉及制备和利用正交(orthogonal) tRNA、正交氨酰-tRNA 合成酶和它们的配对将非天然氨基酸掺入蛋白质的组合物和方法。本发明还涉及用这种配对在细胞中产生蛋白质的方法以及用这种方法产生的蛋白质。

发明背景

因为高度灵敏且操作安全，荧光已成为生物技术中最重要的检测信号之一。诸如荧光共振能量转移(FRET)或荧光极化等方法能实时分析生物分子结合情况、运动或构象改变。用荧光探针选择性修饰蛋白质的能力极其有助于蛋白质结构和功能的体外和体内研究(Hermanson (1996) 刊于 *Bioconjugate Techniques*(生物偶联技术)，学术出版社(Academic Press): 圣迭戈；和 Tsien (1998) *Annu. Rev. Biochem.*, 67:509-544)。

目前体内研究蛋白质的荧光方法常依赖于含大荧光蛋白，例如绿色荧光蛋白(GFP)的融合构建物。虽然已证实该探针可用于蛋白质表达、定位和生物分子相互作用，但其尺寸大会导致明显的结构干扰。GFP 融合还局限于靶蛋白的

C-或 N-末端, 并且对其环境相对不灵敏(Tsien (1998) *Annu. Rev. Biochem.*, 67:509-44)。GFP 还需要许多转录物以获得合适的信号, 其折叠和荧光团成熟需要滞后时间。

还可采用化学方法用各种小的、合成的荧光团选择性修饰蛋白质从而尽可能降低结构干扰(Hermanson, 刊于 *Bioconjugate Techniques*(生物偶联技术), 学术出版社: 圣迭戈 (1996); Martin 等, *Nat. Biotech.*, 23:1308-1314 (2005); Keppler 等, *Nat. Biotech.*, 21:86-89 (2003); Lin 等, *J. Am. Chem. Sci.*, 128:4542-3 (2006))。然而, 这些技术通常局限于分离蛋白质上独特的反应性表面可接近残基(例如, 用马来酰亚胺衍生物修饰半胱氨酸; 参见 Hermanson, 刊于 *Bioconjugate Techniques*(生物偶联技术), (1996), 学术出版社: 圣迭戈), 表现出区域选择性不佳, 具有细胞毒性或需要引入染料结合蛋白基序。

已在实验上证明可利用化学错酰化(misacylated)tRNA 的生物合成标记方法(Mendel 等, *Annu. Rev. Biophys. Biomol. Struct.* (1995) 24:435-462; Hohsaka 等, *FEBS Lett.* (2004) 560:173-177), 但蛋白质的产率有限而且通常只能在体外进行。

直接在活生物体内将遗传编码的荧光氨基酸掺入蛋白质的确定位点可以克服荧光标记蛋白质的许多局限性(Wang 和 Schultz, *Angew. Chem. Int. Ed.* (2005) 44:34-66; Wang 等, *Annu. Rev. Biophys. Biomol. Struct.*, (2006) 35:225-249)。位点特异性掺入荧光氨基酸能在宿主蛋白质中尽可能少地引入干扰, 还能以更高的精确度进行荧光共振能量转移(FRET)检测(Truong 和 Ikura, *Curr. Opin. Struct. Bio.* 2001, 11:573-578)。此外, 利用荧光氨基酸能检测各氨基酸位置的局部环境, 通过改变蛋白质中荧光氨基酸的位置来查明介导与其它细胞组分相互作用的残基。这对于研究蛋白质折叠也非常有用(Lakowicz, *Principles of Fluorescence Spectroscopy*(荧光光谱学原理), 第二版; Kluwer Academic/Plenum Publishers: 纽约, 1999), 特别是在单分子系统中 (Lipman 等, *Science* (2003) 301:1233-1235), 因为一个蛋白质分子通常含有多个色氨酸残基而用荧光探针化学特异性标记蛋白质也成问题。

为研究蛋白质结构和功能, 本领域需要能将荧光非天然氨基酸掺入蛋白质的新方案。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者密码子(selector codon)从而能在体内多肽翻译期间将所需的非天然氨基酸插入限定位置。这些方法利用识别选择者密码子的正交 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、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。

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，其名为“METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS(产生正交 tRNA-氨酰-tRNA 合成酶配对的方法和组合物)”；WO 2002/085923，其名为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS(非天然氨基酸的体内掺入)”；WO 2004/094593，其名为“EXPANDING THE EUKARYOTIC GENETIC CODE(扩展真核遗传密码)”；2004年7月7日提交的 WO 2005/019415；2004年7月7日提交的 WO 2005/007870；2004年7月7日提交的 WO 2005/007624；和 2005年10月27日提交的 WO 2006/110182，其名为“ORTHOGONAL TRANSLATION COMPONENTS FOR THE IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS(体内掺入非天然氨基酸的正交翻译组分)”。这些申请各自通过引用全文纳入本文。掺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它讨论还可参见 Wang 和 Schultz，“Expanding the Genetic Code(扩展遗传密码)”，*Chem. Commun. (Camb.)* 1:1-11 (2002)；Wang 和 Schultz “Expanding the Genetic Code(扩展遗传密码)”，*Angewandte Chemie Int. Ed.*，44(1):34-66 (2005)；Xie 和 Schultz，“An Expanding Genetic Code(扩展遗传密码)”，*Methods* 36(3):227-238 (2005)；Xie 和 Schultz，“Adding Amino Acids to the Genetic Repertoire(将氨基酸加入遗传库)” *Curr. Opinion in Chemical Biology* 9(6):548-554 (2005)；Wang 等，“Expanding the Genetic Code(扩展遗传密

码)”, *Annu. Rev. Biophys. Biomol. Struct.*, 35:225-249 (2006); Xie 和 Schultz, “A chemical toolkit for proteins - an expanded genetic code(蛋白质的化学工具箱-扩展的遗传密码)”, *Nat. Rev. Mol. Cell Biol.*, 7(10):775-782 (2006), 它们的内容各自通过引用全文纳入本文。

本领域需要开发能将荧光非天然氨基酸掺入蛋白质的正交翻译组分, 其中所述荧光非天然氨基酸掺入在指定位置。纵览下文后可明白本文描述的本发明满足了这些和其它需求。

发明概述

本发明提供在体内(例如在宿主细胞内)对选择者密码子, 如琥珀终止密码子起反应而将香豆素非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸掺入延伸中的多肽链的组合物和方法。这些组合物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-tRNA 合成酶(O-RS)配对。即, 内源性宿主细胞氨酰-tRNA 合成酶不会用氨基酸(天然或非天然)加载 O-tRNA(或加载水平不明显)。类似地, 本发明提供的 O-RS 不以显著或可检测的水平用氨基酸(天然或非天然)加载内源性 tRNA。这些新组合物能够产生含有翻译掺入的 L-(7-羟基香豆素-4-基)乙基甘氨酸的大量蛋白质。这些荧光标记的蛋白质具有各种应用。

在一些方面, 本发明提供翻译系统。这些系统包含第一正交氨酰-tRNA 合成酶(O-RS)、第一正交 tRNA (O-tRNA)和第一非天然氨基酸, 即 L-(7-羟基香豆素-4-基)乙基甘氨酸, 其中所述第一 O-RS 用所述第一非天然氨基酸优先氨酰化所述第一 O-tRNA。

该翻译系统可使用衍生自各种来源的组分。在一个实施方式中, 所述第一 O-RS 衍生自詹氏甲烷球菌(*Methanococcus jannaschii*)氨酰-tRNA 合成酶, 例如野生型詹氏甲烷球菌酪氨酰-tRNA 合成酶。用于该系统的 O-RS 可包含 SEQ ID NO: 4 所示氨基酸序列及该序列的保守变体。在一些实施方式中, 所述 O-tRNA 是琥珀抑制型 tRNA。在一些实施方式中, 所述 O-tRNA 包含 SEQ ID NO:1 或由其编码。

在一些方面, 翻译系统还包含编码感兴趣蛋白质的核酸, 其中所述核

酸具有 O-tRNA 识别的至少一个选择者密码子。

在一些方面，翻译系统包括利用第二非天然氨基酸的第二正交配对(即第二 O-RS 和第二 O-tRNA)，现在该系统能在多肽的不同所选位置掺入至少两个不同的非天然氨基酸。在这种双重系统中，第二 O-RS 用不同于第一非天然氨基酸的第二非天然氨基酸优先氨酰化第二 O-tRNA，而第二 O-tRNA 识别不同于第一 O-tRNA 所识别的选择者密码子的选择者密码子。

在一些实施方式中，翻译系统位于宿主细胞中(包括该宿主细胞)。所用的宿主细胞不作具体限定，只要 O-RS 和 O-tRNA 在它们的宿主细胞环境中能保留其正交性。所述宿主细胞可以是真细菌细胞，如大肠杆菌。所述宿主细胞可含有一种或多种编码包括 O-RS 或 O-tRNA 在内的翻译系统组分的多核苷酸。在一些实施方式中，编码 O-RS 的多核苷酸包含 SEQ ID NO:5 所示核苷酸序列。

本发明还提供产生在所选位置含有一个或多个非天然氨基酸的蛋白质的方法。这些方法利用上述翻译系统。这些方法通常始于提供含有以下组分的翻译系统的步骤：(i) 第一非天然氨基酸，即 L-(7-羟基香豆素-4-基)乙基甘氨酸；(ii) 第一正交氨酰-tRNA 合成酶(O-RS)；(iii) 第一正交 tRNA (O-tRNA)，其中所述 O-RS 用所述非天然氨基酸优先氨酰化所述 O-tRNA；和(iv) 编码蛋白质的核酸，其中所述核酸含有 O-tRNA 识别的至少一个选择者密码子。然后在所述蛋白质的翻译过程中该方法对选择者密码子起反应而将所述非天然氨基酸掺入所述蛋白质的所选位置，从而产生在所选位置含有所述非天然氨基酸的蛋白质。

可利用各种试剂和步骤广泛实施该方法。在一些实施方式中，提供编码 O-RS 的多核苷酸。在一些实施方式中，提供衍生自詹氏甲烷球菌氨酰-tRNA 合成酶的 O-RS，例如可以提供野生型詹氏甲烷球菌酪氨酰-tRNA 合成酶。在一些实施方式中，该提供步骤包括提供含有 SEQ ID NO: 4 所示氨基酸序列及其保守变体的 O-RS。

在这些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-tRNA 合成酶的氨基酸结合袋(binding pocket)发生突变，选择用所述非天然氨基酸优先氨酰化所述 O-tRNA 的所得 O-RS。所述选择步

骤可包括定点诱变后从得到的氨酰-tRNA 合成酶分子库正选择和负选择所述 O-RS。在一些实施方式中,提供步骤提供编码 O-tRNA 的多核苷酸,例如, O-tRNA 是琥珀抑制型 tRNA,或者 O-tRNA 包含 SEQ ID NO:1 所示多核苷酸或由其编码。在这些方法中,提供步骤还包括提供含有翻译系统所用的琥珀选择者密码子的核酸。

还可改进这些方法以在蛋白质中掺入一个以上非天然氨基酸。在那些方法中,联用第二正交翻译系统与第一翻译系统,其中第二系统具有不同的氨基酸和选择者密码子特异性。例如,提供步骤可包括提供第二 O-RS 和第二 O-tRNA,其中第二 O-RS 用不同于第一非天然氨基酸的第二非天然氨基酸优先氨酰化第二 O-tRNA,且第二 O-tRNA 识别核酸中不同于第一 O-tRNA 所识别的选择者密码子的选择者密码子。

还可在宿主细胞环境中实施产生含非天然氨基酸的蛋白质的方法。在这些情况中,提供的宿主细胞含有非天然氨基酸、O-RS、O-tRNA 和编码蛋白质的含至少一个选择者密码子的核酸,而培养该宿主细胞可导致非天然氨基酸的掺入。在一些实施方式中,提供步骤包括提供真细菌宿主细胞(例如,大肠杆菌)。在一些实施方式中,提供步骤包括提供含有编码 O-RS 的多核苷酸的宿主细胞。例如,编码 O-RS 的多核苷酸可包含 SEQ ID NO:5 所示核苷酸序列。在一些实施方式中,通过提供细胞提取物实现提供翻译系统的步骤。

本发明还提供包含核酸和蛋白质的各种组合物。除了组合物含有所述核酸或蛋白质外,对该组合物的性质不作具体限制。本发明组合物可含有任何数量、任何性质的其它组分。

例如,本发明提供含有 O-RS 多肽的组合物,其中所述多肽含有 SEQ ID NO: 4 所示氨基酸序列或其保守变体,其中所述保守变体多肽用非天然氨基酸氨酰化关联(cognate)正交 tRNA (O-tRNA)的效率至少为含有该 O-tRNA、该非天然氨基酸和含有 SEQ ID NO: 4 所示氨基酸序列的氨酰-tRNA 合成酶的翻译系统所观察到的效率的 50%。本发明还提供编码上述任何多肽的多核苷酸。在一些实施方式中,这些多核苷酸可含有 SEQ ID NO:5 所示核苷酸序列。在一些实施方式中,多肽在细胞中。

本发明还提供含有 SEQ ID NO:5 所示核苷酸序列的多核苷酸组合物。在一些实施方式中，本发明提供含有所述多核苷酸的载体，如表达载体。在一些实施方式中，本发明提供含有上述载体的细胞。

定义

在详细描述本发明前，应该知道本发明不局限于具体的生物系统，本发明当然可以有各种变化。还应知道本文所用的术语只是为了描述具体的实施方式，而非限制性的。除非另有明确指出，本说明书和随附的权利要求书中使用的单数形式“一个”、“一种”和“该”也包括复数对象。因此，例如述及“一个细胞”包括两个或多个细胞的组合；“一个多核苷酸”实际上包括该多核苷酸的多份拷贝。

除了此处和说明书以下其余部分所定义的，本文所用的所有技术和科学术语都与本发明所属领域普通技术人员常规理解的意义相同。

正交：本文所用的术语“正交”指与某细胞或翻译系统的内源性相应分子相比，某分子(例如，正交 tRNA(O-tRNA)和/或正交氨酰 tRNA 合成酶(O-RS))与该细胞内源性组分起作用的效率降低，或不能与该细胞的内源性组分起作用。就 tRNA 和氨酰-tRNA 合成酶而言，正交指与和内源性 tRNA 合成酶起作用的内源性 tRNA 相比，正交 tRNA 不能与内源性 tRNA 合成酶起作用或起作用的效率降低；或者与和内源性 tRNA 起作用的内源性 tRNA 合成酶相比，正交氨酰 tRNA 合成酶与内源性 tRNA 不起作用或起作用的效率降低，所述效率降低例如，小于 20%的效率，小于 10%的效率，小于 5%的效率，或小于 1%的效率。正交分子缺乏细胞中功能正常的内源性互补分子。例如，与内源性 RS 氨酰化内源性 tRNA 相比，细胞的任何内源性 RS 氨酰化细胞中正交 tRNA 的效率降低或者甚至为 0。在另一例子中，与内源性 RS 氨酰化内源性 tRNA 相比，正交 RS 氨酰化感兴趣细胞中任何内源性 tRNA 的效率降低或者甚至为 0。可将能与第一正交分子起作用的第二正交分子引入细胞。例如，正交 tRNA/RS 配对包括引入的互补组份，与对照，例如相应的 tRNA/RS 内源性配对或活性正交配对(如酪氨酰正交 tRNA/RS 配对)相比，它们在细胞中共同起作用的效率是，例如 45%效率、50%效率、60%效率、70%效率、75%效率、80%

效率、90%效率、95%效率或 99%效率，或更高。

正交酪氨酰-tRNA：本文所用的正交酪氨酰-tRNA(酪氨酰-O-tRNA)是与感兴趣翻译系统正交的 tRNA，其中所述 tRNA 是：(1) 与天然酪氨酰-tRNA 相同或基本类似；(2) 通过天然或人工诱变衍生自天然酪氨酰 tRNA；(3) 由考虑了(1)或(2)的野生型或突变型酪氨酰 tRNA 序列的任何方法产生；(4) 与野生型或突变型酪氨酰 tRNA 同源；(5) 与图 9 中称为酪氨酰 tRNA 合成酶底物的任何示例性 tRNA 同源；或(6) 图 9 中称为酪氨酰 tRNA 合成酶底物的任何示例性 tRNA 的保守变体。酪氨酰 tRNA 可加载有氨基酸，或处于非负载状态。还应知道“酪氨酰-O-tRNA”任选被关联(cognate)合成酶分别加载(氨酰化)除酪氨酸以外的氨基酸，例如非天然氨基酸。应该知道，实际上优选利用本发明酪氨酰-O-tRNA 在翻译期间对选择者密码子起反应而基本上可将任何氨基酸(无论天然或非天然的)掺入延伸的多肽内。

正交酪氨酰氨基酸合成酶：本文所用的正交酪氨酰氨基酸合成酶(酪氨酰-O-RS)是在感兴趣翻译系统内用氨基酸优先氨酰化酪氨酰-O-tRNA 的酶。酪氨酰-O-RS 加载到酪氨酰-O-tRNA 上的氨基酸可以是任何氨基酸，无论是天然的、非天然的还是人工的，并且不限于本文所述的。该合成酶任选与天然酪氨酰氨基酸合成酶相同或同源，或者与图 9 中称为 O-RS 的合成酶相同或同源。例如，所述 O-RS 可以是图 9 的酪氨酰-O-RS 的保守变体，和/或与图 9 中 O-RS 的序列至少有 50%、60%、70%、80%、90%、95%、98%、99%或以上相同。

关联(cognate)：术语“关联”指共同起作用或彼此具有一定特异性的组分，例如正交 tRNA 和正交氨酰基-tRNA 合成酶。这些组分也可互称为“互补的”。

优先氨酰化：对于本文所用正交翻译系统而言，当某表达系统中 O-RS 使 O-tRNA 带上氨基酸的效率高于其使任何内源性 tRNA 带上氨基酸的效率时，该 O-RS “优先氨酰化”关联 O-tRNA。即，当翻译系统中存在摩尔比大致相等的 O-tRNA 与任何给定的内源性 tRNA 时，O-RS 加载 O-tRNA 的频率高于它加载内源性 tRNA 的频率。当翻译系统中存在等摩尔浓度的 O-tRNA 与内源性 tRNA 时，由 O-RS 加载的 O-tRNA 与由 O-RS 加载的内源性 tRNA 的相对比例优选较高，最好导致 O-RS 仅加载或几乎仅加载 O-tRNA。当存在等摩尔浓度

的 O-tRNA 与 O-RS 时, O-RS 加载的 O-tRNA 与内源性 tRNA 的相对比例大于 1:1, 优选至少约 2:1, 更优选 5:1, 更优选 10:1, 更优选 20:1, 更优选 50:1, 更优选 75:1, 更优选 95:1、98:1、99:1、100:1、500:1、1,000:1、5,000:1 或更高。

当(a) 与内源性 tRNA 相比, O-RS 优先氨酰化 O-tRNA, 和(b) 与 O-RS 用任何天然氨基酸氨酰化 O-tRNA 相比, 氨酰化对非天然氨基酸特异时, 称 O-RS “优先用非天然氨基酸氨酰化 O-tRNA”。即, 当包含 O-RS 和 O-tRNA 的翻译系统中存在等摩尔量的非天然和天然氨基酸时, O-RS 用非天然氨基酸加载 O-tRNA 的频率高于用天然氨基酸加载的频率。加载有非天然氨基酸的 O-tRNA 与加载有天然氨基酸的 O-tRNA 的相对比例优选较高。O-RS 最好使 O-tRNA 仅加载, 或者几乎仅加载有非天然氨基酸。当翻译系统中存在等摩尔浓度的天然和非天然氨基酸时, 使 O-tRNA 带上非天然氨基酸与使 O-tRNA 带上天然氨基酸的相对比例大于 1:1, 优选至少约 2:1, 更优选 5:1, 更优选 10:1, 更优选 20:1, 更优选 50:1, 更优选 75:1, 更优选 95:1、98:1、99:1、100:1、500:1、1,000:1、5,000:1 或更高。

选择者密码子: 术语“选择者密码子”指翻译过程中为 O-tRNA 所识别而不为内源性 tRNA 所识别的密码子。O-tRNA 反密码子环识别 mRNA 上的选择者密码子并在多肽的此位置掺入其氨基酸, 例如非天然氨基酸。选择者密码子可包括, 例如无义密码子, 如终止密码子(如, 琥珀、赭石和乳白密码子); 四碱基或四碱基以上密码子; 罕用密码子; 衍生自天然或非天然碱基对的密码子, 等等。

抑制型 tRNA: 抑制型 tRNA 是在多肽翻译期间通常对终止密码子起反应而掺入氨基酸(即, “连读”)来改变给定的翻译系统中信使 RNA(mRNA)阅读的 tRNA。在一些方面, 本发明的选择者密码子是抑制型密码子, 例如, 终止密码子(如, 琥珀、赭石和乳白密码子)、四碱基密码子、罕用密码子等。

抑制活性: 本文所用的术语“抑制活性”总体上指 tRNA(例如, 抑制型 tRNA)能翻译连读在其它情况中可导致翻译终止或错译(例如, 移码)的密码子(例如, 作为选择者密码子的琥珀密码子或四个或以上个碱基的密码子)的能力。抑制型 tRNA 的抑制活性可表示为与第二抑制型 tRNA, 或与对照系统, 例如缺乏

O-RS 的对照系统相比, 观察到的翻译连读活性的百分比。

本发明提供定量测定抑制活性的各种方法。特定 O-tRNA 和 O-RS 对感兴趣选择者密码子(例如, 琥珀密码子)的抑制百分比指, 在感兴趣的翻译系统中, 在编码表达测试标记的核酸中含有选择者密码子的该给定表达测试标记(例如, LacZ)的活性与阳性对照构建物相比的百分比, 所述感兴趣的翻译系统包含 O-RS 和 O-tRNA, 所述阳性对照没有 O-tRNA、O-RS 和选择者密码子。因此, 例如, 如果在给定翻译系统中观察到不含选择者密码子的活性阳性对照标记构建物的活性为 X(其单位与所述标记试验相关), 则含有选择者密码子的测试构建物的抑制百分比是在与阳性对照标记的表达基本相同的环境条件下(除了该测试标记构建物在也含有 O-tRNA 和 O-RS 的翻译系统中表达外), 该测试标记构建物显示的 X 百分比。表达该测试标记的翻译系统一般也包含 O-RS 和 O-tRNA 识别的氨基酸。可任选通过将测试标记与“背景”或“阴性”对照标记构建物比较来校正抑制百分比的测量值, 所述“背景”或“阴性”对照标记构建物含有与测试标记相同的选择者密码子, 但其所在的系统不含 O-tRNA、O-RS 和/或 O-tRNA 和/或 O-RS 识别的相关氨基酸。该阴性对照可用于标准化抑制百分比测定值以补偿感兴趣的翻译系统中标记的背景信号的影响。

可通过本领域已知的许多试验测定抑制效率。例如, 可采用 β -半乳糖苷酶报道试验, 如可将衍生的 lacZ 质粒(该构建物在 lacZ 核酸序列中含有选择者密码子)连同含有本发明的 O-tRNA 的质粒引入合适生物(例如, 可利用正交组分的生物)的细胞。还可引入关联合成酶(可以是多肽或表达时能编码该关联合成酶的多核苷酸)。细胞在培养基中生长至所需密度, 例如至 OD₆₀₀ 约为 0.5, 进行 β -半乳糖苷酶试验, 例如用诺瓦金公司(Novagen)的 BetaFluor™ β -半乳糖苷酶试验试剂盒。可将抑制百分比计算为样品相对于可比较对照, 例如, 衍生的 lacZ 构建物的观察值的活性百分比, 该构建物在所需位置具有相应的有义密码子而非选择者密码子。

翻译系统: 术语“翻译系统”指将氨基酸掺入延伸中的多肽链(蛋白质)的各组分。翻译系统的组分可包括, 例如核糖体、tRNA、合成酶、mRNA 等。本发明的 O-tRNA 和/或 O-RS 可加入体外或体内翻译系统或是其一部分, 例如存在于非真核细胞, 如细菌(如大肠杆菌)中, 或存在于真核细胞中, 如酵母菌、

哺乳动物细胞、植物细胞、藻类细胞、真菌细胞、昆虫细胞等。

非天然氨基酸：本文所用的术语“非天然氨基酸”指不在 20 种常见天然氨基酸或硒代半胱氨酸或吡咯赖氨酸(pyrrolysine)之列的任何氨基酸，修饰的氨基酸和/或氨基酸类似物。例如，本发明可使用非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸 (还写作 7-羟基香豆素-乙基甘氨酸；参见图 1，结构 1)。

衍生自：本文所用的术语“衍生自”指分离自特定分子或生物，或是用特定分子或生物的信息制得的某组份。例如，衍生自第二多肽的多肽含有与该第二多肽的氨基酸序列相同或基本上类似的氨基酸序列。以多肽为例，可通过，例如天然产生的诱变、人工定向诱变或人工随机诱变获得衍生的种类。用于衍生多肽的诱变可以是有意定向或有意随机的，或混用两种方法。诱变多肽以产生衍生自第一(多肽)的不同多肽可以是随机的(例如，聚合酶失真所致)，可通过合适的筛选方法，例如本文所述的方法鉴定衍生的多肽。诱变多肽通常需要对编码该多肽的多核苷酸进行操作。

正选择或筛选标记：本文所用的术语“正选择或筛选标记”指当存在该标记时(例如被表达或激活等)可不具有特征的细胞中鉴定出具有特征的细胞，例如具有正选择标记的细胞。

负选择或筛选标记：本文所用的术语“负选择或筛选标记”指当存在该标记时(例如被表达或激活等)可鉴定不含有所选特性或特征的细胞(例如，与确实具有该特性或特征的细胞相比)。

报道分子：本文所用的术语“报道分子”是指可用于鉴定和/或选择感兴趣系统的靶组分的组分。例如，报道分子可以包括蛋白质，如能赋予抗生素抗性或敏感性的酶(如 β -内酰胺酶、氯霉素乙酰转移酶(CAT)等)、荧光筛选标记(如绿色荧光蛋白(如 GFP)、YFP、EGFP、RFP 等)、冷光标记(如萤火虫萤光素酶蛋白)、亲和力筛选标记，或者可选择性的正或负标记基因如 lacZ、 β -gal/lacZ(β -半乳糖苷酶)、ADH(乙醇脱氢酶)、his3、ura3、leu2、lys2 等。

真核生物：本文所用的术语“真核生物”指属于真核生物界(Kingdom Eucarya)的生物。真核生物通常因其以下特征而区别于原核生物：典型的多细胞组织(但不都是多细胞，例如酵母)，存在膜限制的核与其它膜限制的细胞器，

线形遗传物质(即, 线形染色体), 不存在操纵子, 存在内含子、信使(RNA)加帽和 poly-A mRNA, 以及其它生物化学特征, 如不同的核糖体结构。真核生物包括, 例如动物(如哺乳动物、昆虫、爬行动物、鸟等), 纤毛虫, 植物(如单子叶植物、双子叶植物、藻类等)、真菌、酵母菌、鞭毛虫类、微孢子虫、原生动物等。

原核生物: 本文所用的术语“原核生物”指属于原核生物界(也称为原核生物(Procarya))的生物。原核生物通常因其以下特征而区别于真核生物: 单细胞组织, 通过出芽或分裂的无性生殖, 缺乏膜限制的核与其它膜限制的细胞器, 环状染色体, 存在操纵子, 不存在内含子、信使(RNA)加帽和 poly-A mRNA, 以及其它生物化学特征, 如不同的核糖体结构。原核生物包括真细菌和古细菌 (Archaea) (有时称为“古细菌(Archaeobacteria)”)亚界。在原核生物界有时将蓝细菌(蓝绿藻)与支原体分为不同类别。

细菌: 本文所用的术语“细菌”和“真细菌”指不同于古细菌的原核生物。类似地, 古细菌指不同于真细菌的原核生物。可根据许多形态和生物化学标准区分真细菌和古细菌。例如, 可采用核糖体 RNA 序列、RNA 聚合酶结构的差异, 内含子的存在与否, 抗生素敏感性, 细胞壁肽聚糖和其它细胞壁组分的存在与否, 膜脂质的分枝与不分枝结构, 存在/不存在组蛋白和组蛋白样蛋白而将某生物指定为真细菌或古细菌。

真细菌的例子包括大肠杆菌(*Escherichia coli*)、嗜热栖热菌(*Thermus thermophilus*)、枯草芽孢杆菌(*Bacillus subtilis*)和嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)等)。古细菌的例子包括詹氏甲烷球菌(*Methanococcus jannaschii*)(Mj)、梅氏甲烷八叠球菌(*Methanosarcina mazei*)(Mm)、嗜热碱甲烷杆菌 (*Methanobacterium thermoautotrophicum*)(Mt)、海沼甲烷球菌 (*Methanococcus maripaludis*)、甲烷嗜热菌(*Methanopyrus kandleri*)、盐细菌 (*Halobacterium*) (例如沃氏富盐菌(*Haloferax volcanii*)和盐细菌种 NRC-1)、闪烁古生球菌(*Archaeoglobus fulgidus*)(Af)、激烈火球菌(*Pyrococcus furiosus*)(Pf)、极端嗜热古菌(*Pyrococcus horikoshii*)(Ph)、好氧火球菌(*Pyrobaculum aerophilum*)、深海火球菌(*Pyrococcus abyssi*)、硫磺矿硫化叶菌(*Sulfolobus solfataricus*)(Ss)、超嗜热泉古菌(*Sulfolobus tokodaii*)、嗜热泉生古细菌(*Aeuiropyrum pernix*)(Ap)、

嗜酸热原体 (*Thermoplasma acidophilum*) 和火山热原体 (*Thermoplasma volcanium*)。

保守性变体：就翻译组分而言，本文所用的术语“保守性变体”指某翻译组份，例如保守性变体 O-tRNA 或保守性变体 O-RS，其所执行的功能与与其相似的基本组分（例如 O-tRNA 或 O-RS）相类似，但与参比 O-tRNA 或 O-RS 相比序列中有变异的。例如，O-RS 或该 O-RS 的保守性变体可用非天然氨基酸，例如 L-(7-羟基香豆素-4-基)乙基甘氨酸非天然氨基酸氨酰化关联 O-tRNA。在该例子中，所述 O-RS 及保守性变体 O-RS 的氨基酸序列不相同。所述保守性变体在序列中可具有例如一处变异、两处变异、三处变异、四处变异或五处或更多处变异，只要该保守性变体仍与关联的相应 O-tRNA 或 O-RS 互补(例如，与之起作用)。

在一些实施方式中，保守性变体 O-RS 与衍生其的 O-RS 相比，含有一个或多个保守性氨基酸取代。在一些实施方式中，保守性变体 O-RS 与衍生其的 O-RS 相比，含有一个或多个保守性氨基酸取代，此外还保留了 O-RS 的生物学活性；例如，保守性变体 O-RS 保留了衍生其的亲本 O-RS 分子至少 10%，或者至少 20%，至少 30%，或至少 40%的生物学活性。在一些优选实施方式中，所述保守性变体 O-RS 保留了衍生其的亲本 O-RS 分子至少 50%的生物学活性。保守性变体 O-RS 的保守性氨基酸取代可出现在该 O-RS 的任何结构域内，包括氨基酸结合袋。

选择或筛选试剂：本文所用的术语“选择或筛选试剂”指当存在时可从某群体中选择/筛选某些组分的试剂。例如，选择或筛选试剂可以是但不限于，如营养物质、抗生素、某波长的光、抗体、表达的多核苷酸等。选择试剂可因例如浓度、强度等而有所不同。

起反应：本文所用的术语“起反应”指本发明的 O-tRNA 识别选择者密码子并介导与该 tRNA 偶联的非天然氨基酸掺入延伸中的多肽链的过程。

编码：本文所用的术语“编码”指用多聚大分子或序列链中的信息来指导不同于该第一分子或序列链的第二种分子或序列链产生的任何过程。本文所用的该术语应用广泛，可用于各领域。在一些方面，术语“编码”描述了半保留 DNA 复制的过程，其中双链 DNA 分子的一条链用作模板，借助依赖 DNA 的

DNA 合成酶来编码新合成的互补姊妹链。

在另一方面，术语“编码”指用一种分子的信息来指导产生化学性质不同于该第一分子的第二种分子的任何过程。例如，DNA 分子可编码 RNA 分子(例如，包含依赖 DNA 的 RNA 聚合酶的转录过程)。RNA 分子也可编码多肽，例如翻译过程。当术语“编码”用于描述翻译过程时，其含义也延伸至编码氨基酸的三联密码子。在一些方面，RNA 分子可编码 DNA 分子，例如通过包含依赖 RNA 的 DNA 合成酶的逆转录过程。在另一方面，DNA 分子可编码多肽，应该知道该情况用“编码”同时包括转录和翻译过程。

附图简述

图 1 是化学合成 L-(7-羟基香豆素-4-基)乙基甘氨酸 (结构 1)的示意图。

图 2 是 L-(7-羟基香豆素-4-基)乙基甘氨酸的吸收和发射光谱。以 pH 7.4, 100 mM 磷酸钠缓冲液记录两种光谱。在此 pH 下，L-(7-羟基香豆素-4-基)乙基甘氨酸存在酚和酚盐两种形式(比例约为 2:1)。360 nM 酚盐形式的消光系数为 17000，量子产额(quantum yield)为 0.63。

图 3 提供了抹香鲸肌红蛋白的结构(pdb 编码 105M)。标出了残基 Ser4 和 His37 的 R-基团。

图 4 提供野生型和突变型形式全肌红蛋白(holomyoglobin)的脲诱导未折叠的圆二色性分析结果。分析的突变形式是 Ser4 → 7-羟基香豆素-乙基甘氨酸或 His37 → 7-羟基香豆素-乙基甘氨酸突变型全肌红蛋白。通过 7-羟基香豆素-乙基甘氨酸的 CD 和荧光强度监测未折叠情况。两种实验均在 100 mM 磷酸钠缓冲液，pH 7.4 中进行，脲的各种浓度如图所示。与 0 M 脲相比，有 5 M 脲存在下，肌红蛋白突变体显示荧光强度增加 30%。根据完全未折叠的状态，通过标准化 5 M 脲的荧光强度或摩尔椭圆率计算“折叠分数”。

图 5 提供 TAG4 突变体肌红蛋白表达(黑箭头所示)的 SDS-PAGE 分析的照片。中图显示在有和没有 1 mM L-(7-羟基香豆素-4-基)乙基甘氨酸存在下，累积蛋白质的考马斯-染色的 SDS-PAGE 分离情况。右图显示 SDS-PAGE 分离后野生型和 TAG4 突变体肌红蛋白的荧光图像。左图显示染色的分子量标记。

图 6 提供 TAG4 突变型肌红蛋白的 ESI-MS 图谱。插页显示去卷积图

(deconvoluted spectrum)。预计分子量：18512 Da；实测值：18511 Da。

图 7 提供 60 μ M 香豆素基氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸在 100 mM 磷酸钠缓冲液中的吸收光谱。

图 8 提供突变型 *MjTyrRS* (CouRS-D8)活性位点的结构模型。采用野生型 *MjTyrRS* 结构 (pdb 编码 1J1U)产生该模型。用 L-(7-羟基香豆素-4-基)乙基甘氨酸替代酪氨酸底物，突变如下所示：Tyr32Glu、Leu65His、Ala67Gly、His70Gly、Phe108Tyr、Gln109His、Asp158Gly 和 Leu162Gly。采用美国加利福尼亚州圣迭戈市阿克塞勒斯软件公司的 InsightII[®]软件 (Accelrys Software, Inc.[®], San Diego, CA, USA)，通过能量最小化优化结构。

图 9 提供各种核苷酸和氨基酸序列。

发明详述

本发明提供能在大肠杆菌中对选择者密码子（例如琥珀终止密码子 TAG）起反应而将荧光氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸(参见图 1，结构 1)体内选择性引入蛋白质的正交 tRNA/氨酰-tRNA 合成酶配对。本发明提供用非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸特异性加载相关的正交 tRNA(O-tRNA)的新正交氨酰-tRNA 合成酶(O-RS)多肽。

7-羟基香豆素分子最初因其荧光量子产率高、Stoke 位移较大(参见图 2)、尺寸小、对 pH(参见图 7)和溶剂极性灵敏而得到研究(Zinsli, J. Photochem (1974) 3:55-69)。

在某些方面，为证明(但不是限制)本发明，本文内容证明可将非天然氨基酸部分掺入模型蛋白质(参见实施例 4 和 5)。掺入非天然氨基酸不必局限于任何特定的蛋白质。从本发明可以明白，将非天然荧光氨基酸掺入感兴趣的特定蛋白质对于各种目的是有利的。

本文还描述了开发能在真细菌中起作用而对选择者密码子起反应，从而位点特异性地掺入非天然荧光氨基酸(例如，图 1 所示的非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸，结构 1)的新型正交 tRNA/氨酰-tRNA 合成酶配对的方法。简言之，我们鉴定了在大肠杆菌宿主细胞中用非天然氨基酸选择性加载抑制型 RNA 的詹氏甲烷球菌酪氨酰-tRNA 合成酶的新型突

变体。

可利用这些开发的 tRNA-合成酶配对将非天然荧光氨基酸位点特异性地掺入蛋白质。可通过工程改造编码感兴趣蛋白质的多核苷酸序列使之含有能发出掺入非天然氨基酸信号的选择者密码子，从而能将非天然氨基酸编程掺入任何所需位置。

本文所述发明提供在真细菌，例如大肠杆菌中将 L-(7-羟基香豆素-4-基)乙基甘氨酸非天然氨基酸遗传编码并掺入蛋白质的正交配对，其中所述正交组分不与宿主细胞翻译机制的内源性大肠杆菌组分交叉反应，但能识别所需的非天然氨基酸并对选择者密码子(例如，琥珀无义密码子，TAG)起反应而将其掺入蛋白质。本发明提供的正交组分包括衍生自詹氏甲烷球菌酪氨酰 tRNA-合成酶的正交氨酰-tRNA 合成酶，和突变型酪氨酰 tRNA_{CUA} 琥珀抑制物，二者在真细菌宿主细胞中用作正交配对。

本发明提供鉴定和产生其它正交 tRNA-氨酰-tRNA 合成酶配对，例如 O-tRNA/O-RS 配对的组合物和方法，所述配对可用于将 L-(7-羟基香豆素-4-基)乙基甘氨酸非天然氨基酸掺入蛋白质。本发明的 O-tRNA/O-RS 配对能介导 L-(7-羟基香豆素-4-基)乙基甘氨酸掺入多核苷酸编码的蛋白质，其中所述多核苷酸包含所述 O-tRNA 识别的选择者密码子。所述 O-tRNA 的反密码子环识别 mRNA 上的选择者密码子，进而将非天然氨基酸掺入多肽中的该位点。本发明的正交氨酰-tRNA 合成酶通常只用一种特定的非天然氨基酸优先氨酰化(或加载)其 O-tRNA。

正交 tRNA/氨酰-tRNA 合成酶技术

理解与正交 tRNA 和正交氨酰-tRNA 合成酶配对有关的活性有助于理解本发明的新型组合物和方法。为在遗传密码中加入额外的非天然氨基酸，需要含有氨酰-tRNA 合成酶和适当 tRNA 的新正交配对，它们可在宿主翻译机制中有效起作用，但对于所述翻译系统是“正交”的，这意味着它可不依赖于翻译系统的内源性合成酶和 tRNA 而起作用。正交配对的所需特征包括能解码或识别不由任何内源性 tRNA 解码的仅一种特定密码子(例如选择者密码子)的 tRNA，和只能用一种特定非天然氨基酸优先氨酰化(或“加载”)其关联 tRNA 的氨酰-tRNA 合成酶。内源性合成酶通常不氨酰化 O-tRNA(或氨

酰化, 即加载不佳)。例如, 在大肠杆菌宿主系统中, 正交配对包括不与任何内源性 tRNA(例如, 大肠杆菌中有 40 种)交叉反应的氨酰-tRNA 合成酶和不被任何内源性合成酶(例如, 大肠杆菌中有 21 种)氨酰化的正交 tRNA。

本领域已知适于制备含有一个或多个非天然氨基酸的蛋白质的正交翻译系统的通用原则, 例如制备正交翻译系统的通用方法。例如, 可参见国际公开号 WO 2002/086075, 名为“METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS”(产生正交 tRNA-氨酰-tRNA 合成酶配对的方法和组合物); WO 2002/085923, 名为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”(非天然氨基酸的体内掺入); WO 2004/094593, 名为“EXPANDING THE EUKARYOTIC GENETIC CODE”(扩展真核遗传密码); 2004 年 7 月 7 日提交的 WO 2005/019415; 2004 年 7 月 7 日提交的 WO 2005/007870; 2004 年 7 月 7 日提交的 WO 2005/007624; 2005 年 10 月 27 日提交的 WO 2006/110182, 名为“ORTHOGONAL TRANSLATION COMPONENTS FOR THE IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”(用于体内掺入非天然氨基酸的正交翻译组分)和 2007 年 3 月 7 日提交的美国实用申请序列号 11/715,672, 名为“SYSTEMS FOR THE EXPRESSION OF ORTHOGONAL TRANSLATION COMPONENTS IN EUBACTERIAL HOST CELLS(在真细菌宿主细胞中表达正交翻译组分的系统)”。这些申请各自通过引用全文纳入本文。掺入非天然氨基酸的正交翻译系统和它们的产生及使用方法的讨论还可参见 Wang 和 Schultz, “Expanding the Genetic Code (扩展遗传密码)”, *Chem. Commun. (Camb.)* 1:1-11(2002); Wang 和 Schultz, “Expanding the Genetic Code (扩展遗传密码)”, *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005); Xie 和 Schultz, “An Expanding Genetic Code (扩展的遗传密码)”, *Methods* 36(3):227-238 (2005); Xie 和 Schultz, “Adding Amino Acids to the Genetic Repertoire (将氨基酸加入遗传库)”, *Curr. Opinion in Chemical Biology* 9(6):548-554 (2005); Wang 等, “Expanding the Genetic Code (扩展遗传密码)”, *Annu. Rev. Biophys. Biomol. Struct.*, 35:225-249 (2006); Xie 和 Schultz,

“A chemical toolkit for proteins - an expanded genetic code(蛋白质的化学工具箱-扩展的遗传密码)”，*Nat. Rev. Mol. Cell Biol.*, 7(10):775-782 (2006); 这些文献的内容通过引用全文纳入本文。

正交翻译系统

正交翻译系统一般包括含有正交 tRNA(O-tRNA)、正交氨酰 tRNA 合成酶(O-RS)和非天然氨基酸的细胞(可以是原核细胞，例如大肠杆菌)，其中所述 O-RS 用所述非天然氨基酸氨酰化所述 O-tRNA。本发明的正交配对可以包括 O-tRNA，例如抑制型 tRNA、移码 tRNA 等和关联 O-RS。本发明的正交系统通常包含处于宿主细胞环境中或无宿主细胞的 O-tRNA/O-RS 配对。除多组分系统外，本发明还提供新型单组分，例如，新型正交氨酰-tRNA 合成酶多肽(例如，SEQ ID NO:4)和编码那些多肽的多核苷酸(例如，SEQ ID NO:5)。

当正交配对识别选择者密码子并对该选择者密码子起反应而加载氨基酸时，通常称该正交配对“抑制”该选择者密码子。即，不被翻译系统的(例如，细胞的)内源性机制识别的选择者密码子通常不被加载，从而阻断了多肽产生，否则可自核酸翻译该多肽。在正交配对系统中，O-RS 用特定的非天然氨基酸氨酰化 O-tRNA。加载的 O-tRNA 识别选择者密码子并抑制选择者密码子所致的翻译阻断。

在一些方面，与包含本文序列表所示多核苷酸序列或由其编码的 O-tRNA 的抑制效率相比，本发明的 O-tRNA 识别选择者密码子并在关联合成酶存在下对选择者密码子起反应而具有至少约，例如，45%、50%、60%、75%、80%或 90%或更高的抑制效率。

在一些实施方式中，联用 O-RS 和 O-tRNA 的抑制效率是缺乏 O-RS 的 O-tRNA 的抑制效率的约，例如 5 倍、10 倍、15 倍、20 倍或 25 倍或更高。在一些方面，联用 O-RS 和 O-tRNA 的抑制效率是本文序列表所示正交合成酶配对的抑制效率的至少约，例如 35%、40%、45%、50%、60%、75%、80%或 90%或更高。

宿主细胞利用 O-tRNA/O-RS 配对将非天然氨基酸掺入延伸中的多肽链，例如经由包含编码感兴趣多肽的多核苷酸的核酸，其中所述多核苷酸包含所述

O-tRNA 识别的选择者密码子。在某些优选方面，细胞可包含一种或多种其它 O-tRNA/O-RS 配对，其中所述其它 O-RS 用不同的非天然氨基酸加载所述其它 O-tRNA。例如，O-tRNA 之一可识别四碱基密码子，其它 O-tRNA 可识别终止密码子。或者，多个不同的终止密码子或多个不同的四碱基密码子可用于同一编码核酸。

应注意，在一些实施方式中，细胞或其它翻译系统中可存在多个 O-tRNA/O-RS 配对，从而能将多个非天然氨基酸掺入多肽。例如，细胞还可包含额外的不同 O-tRNA/O-RS 配对和第二非天然氨基酸，其中该额外的 O-tRNA 识别第二选择者密码子，该额外的 O-RS 用该第二非天然氨基酸优先氨酰化该额外的 O-tRNA。例如，包含 O-tRNA/O-RS 配对的细胞(其中所述 O-tRNA 识别，例如琥珀选择者密码子)还可包含第二正交配对，其中该第二 O-tRNA 识别不同的选择者密码子，例如乳白密码子、四碱基密码子等。不同的正交配对优选衍生自不同来源，从而有助于识别不同的选择者密码子。

在某些实施方式中，系统包含例如大肠杆菌细胞等细胞，这些细胞含有正交 tRNA(O-tRNA)、正交氨酰 tRNA 合成酶(O-RS)、非天然氨基酸和包含编码感兴趣多肽的多核苷酸的核酸，其中所述多核苷酸包含所述 O-tRNA 识别的选择者密码子。翻译系统还可以是无细胞系统，例如与本文所述 O-tRNA/O-RS 配对和非天然氨基酸组合的各种市售可得“体外”转录/翻译系统。

O-tRNA 和/或 O-RS 可以是天然产生的，或者可以是，例如天然 tRNA 和/或 RS 经突变衍生，例如，通过产生各种生物的 tRNA 文库和/或 RS 文库和/或采用各种可用的突变方案。例如，产生正交 tRNA/氨酰-tRNA 合成酶配对的一种方案包括将，例如得自除宿主细胞外的来源或多个来源的异源(对宿主细胞而言)tRNA/合成酶配对输入宿主细胞。候选异源合成酶的特性包括，例如不加载任何宿主细胞 tRNA，候选异源 tRNA 的特性包括，例如不被任何宿主细胞合成酶所氨酰化。此外，异源 tRNA 对于所有宿主细胞合成酶是正交的。产生正交配对的第二种方案包括产生用于筛选和/或选择 O-tRNA 或 O-RS 的突变体文库。也可联用这些方案。

正交 tRNA(O-tRNA)

本发明的正交 tRNA(O-tRNA)优选在例如体内或体外介导将非天然氨基酸掺入蛋白质内，所述蛋白质由含选择者密码子的多核苷酸编码，而这种选择者密码子可被该 O-tRNA 识别。在某些实施方式中，与包含本文序列表中 O-tRNA 序列所示多核苷酸序列或其编码的 O-tRNA 相比，本发明的 O-tRNA 在关联合成酶存在下，对选择者密码子起反应而具有至少 45%、50%、60%、75%、80%、90%或更高的抑制效率。

可通过本领域已知的多种试验测定抑制效率。例如，可采用 β 半乳糖苷酶报道分子试验，例如，将衍生的 lacZ 质粒(该构建物在 lacZ 核酸序列中含有选择者密码子)和含有本发明 O-tRNA 的质粒一起导入合适生物(如可利用正交组分的生物)的细胞内。还可导入关联合成酶(多肽或表达时可编码关联合成酶的多核苷酸)。将细胞在培养基内培养到所需密度，如 OD₆₀₀ 约 0.5 时，采用例如诺瓦金公司(Novagen)的 BetaFluor™ β -半乳糖苷酶检测试剂盒进行 β 半乳糖苷酶试验。抑制百分率可以计算为样品相对于可比较对照(如衍生的 lacZ 构建物的观察值，所述构建物在所需位置上含有相应的有义密码子而不是选择者密码子)的活性百分数。

本发明 O-tRNA 的例子见本文序列表，例如参见图 9 和 SEQ ID NO:1。本文内容还提供了设计其它等价 O-tRNA 的指导。在 RNA 分子(例如 O-RS mRNA 或 O-tRNA 分子)中，与给定序列(或对编码 DNA 而言反之亦然)或其互补序列相比，胸腺嘧啶(T)被尿嘧啶(U)取代。还可存在碱基的其它修饰以大量产生功能等价分子。

本发明还包括对应于本文具体 O-tRNA 的 O-tRNA 保守性变体。例如，O-tRNA 保守性变体包括功能与具体 O-tRNA(例如本文序列表中的)相似的，和因合适的自身互补性而保留了 tRNA 的 L-形结构但不具有与例如序列表或图 9 所示那些(序列)相同的序列并且最好也不是野生型 tRNA 分子的那些分子。

含 O-tRNA 的组合物还可包含正交氨酰-tRNA 合成酶(O-RS)，其中所述 O-RS 用非天然氨基酸优先氨酰化 O-tRNA。在某些实施方式中，含有 O-tRNA 的组合物还可包含(例如体外或体内)翻译系统。细胞中也可存在含有编码感兴趣多肽的多核苷酸的核酸或这些物质中一个或多个的组合，其

中所述多核苷酸含有 O-tRNA 能识别的选择者密码子。

产生正交 tRNA(O-tRNA)的方法也是本发明的特征之一。通过该方法产生的正交 tRNA(O-tRNA)也是本发明的特征之一。在本发明的某些实施方式中,可通过构建突变体文库来产生 O-tRNA。可采用本领域已知的各种诱变技术构建突变体 tRNA 文库。例如,可通过位点特异性突变、随机位点突变、同源重组、DNA 改组或其它递归诱变(recursive mutagenesis)方法、嵌合体构建、或者这些技术的任何组合产生突变体 tRNA,例如 SEQ ID NO: 1 所示 O-tRNA。

也可将其它突变引入特定位置,例如 tRNA 的所需环或区域(如反密码子环、接纳茎、D 臂或环、可变环、TPC 臂或环、tRNA 分子的其它区域或其组合)中一个或多个非保守位置、保守位置、一个或多个随机位置或者二者的组合位置。tRNA 中的突变通常包括使突变型 tRNA 文库内各成员的反密码子环突变以使其能识别选择者密码子。该方法还可包括给 O-tRNA 加上额外序列。与起始材料(例如多种 tRNA 序列)相比, O-tRNA 通常提高了对所需生物的正交性,同时保留其对所需 RS 的亲和力。

这些方法任选包括分析 tRNA 和/或氨酰-tRNA 合成酶序列的相似性(和/或所推断的同源性)以确定显示与特定生物正交的潜在候选 O-tRNA、O-RS 和/或其配对。可利用本领域已知和本文所述的计算机程序进行这种分析,例如可用 BLAST 和 pileup 程序。在一个实施例中,为筛选用于大肠杆菌的可能的正交翻译组分,可选择与真细菌生物不显示接近的序列相似性的合成酶和/或 tRNA。

通常可通过,例如负选择第一物种的细胞群以获得 O-tRNA,其中所述细胞含有多种潜在 O-tRNA 的某成员。负选择可去除含有被细胞内源性氨酰-tRNA 合成酶(RS)氨酰化的潜在 O-tRNA 文库某成员的细胞。这样就可以得到第一物种细胞的正交 tRNA 库。

某些实施方式在负选择中将选择者密码子引入编码负选择标记(例如能赋予抗生素抗性的酶如 β 内酰胺酶;能得到可检测产物的酶如 β 半乳糖苷酶、氯霉素乙酰转移酶(CAT),所述产物例如是毒性产物,如芽孢杆菌 RNA 酶)的多核苷酸的非必须位置(例如,仍能产生功能性芽孢杆菌 RNA 酶)

等。还任选在选择性试剂(比如抗生素,如氨苄青霉素)存在下培养细胞群来进行筛选/选择。在一个实施方式中,选择性试剂的浓度不同。

例如,为检测抑制型 tRNA 的活性,可利用基于选择者密码子的体内抑制的选择系统,例如将无义(如终止)或移码突变引入编码负选择标记的多核苷酸(如编码 β -内酰胺酶的基因(*bla*))中。例如,构建在某位置(如, A184)含有选择者密码子的多核苷酸变体,如 *bla* 变体。用这些多核苷酸转化细胞,如细菌。以不能被内源性大肠杆菌合成酶有效加载的正交 tRNA 为例,抗生素抗性(例如氨苄青霉素抗性)应约为或小于未用质粒转化的细菌的抗生素抗性。如果 tRNA 不是正交的,或者能加载该 tRNA 的异源合成酶在此系统内共同表达,应可观察到更高水平的抗生素(如氨苄青霉素)抗性。选出那些在抗生素浓度与未用质粒转化的细胞大致相等的 LB 琼脂平板上不能生长的细胞,如细菌。

以毒性产物(如核糖核酸酶或芽孢杆菌 RNA 酶)为例,当多种潜在的 tRNA 中的成员被内源性宿主(如大肠杆菌)合成酶(即与宿主如大肠杆菌合成酶不是正交的)氨酰化时,选择者密码子被抑制,所产生的毒性多核苷酸产物导致细胞死亡。而含有正交 tRNA 或非功能性 tRNA 的细胞可存活。

然后,在一个实施方式中,对与所需生物正交的 tRNA 库进行正选择,其中将选择者密码子置于正选择标记中(例如抗药性基因(如 β 内酰胺酶基因)编码的标记)。对以下细胞进行正选择:含有编码与该细胞正交的 tRNA 库某成员的多核苷酸或包含该成员的多核苷酸、含有编码正选择标记的多核苷酸和含有编码关联 RS 的多核苷酸。在某些实施方式中,第二群细胞含有不能通过负选择除去的细胞。该多核苷酸在胞内表达,细胞在有选择试剂(如氨苄青霉素)存在下生长。然后选择能被共同表达的关联合成酶氨酰化以及对此选择者密码子起反应而插入氨基酸的 tRNA。与一种或多种含有非功能性 tRNA 或不能被感兴趣合成酶有效识别的 tRNA 的细胞相比,这些细胞通常显示抑制效率升高。含有非功能性 tRNA 或不能被感兴趣合成酶有效识别的 tRNA 的细胞对该抗生素敏感。因此在两次选择中能保留下的 tRNA 是:(i) 不是内源性宿主(如大肠杆菌)合成酶的底物;(ii) 能被感兴趣的合成酶氨酰化;以及(iii) 能在翻译过程中起作用。

因此,取决于筛选所处的环境,同一标记可以是正或负标记。即,如果为了筛选该标记,则它是正标记,但如果为了抵御该标记则它是负标记。

上述方法中,筛选,如正选择、负选择或者正负选择的严格性任选包括改变选择的严格性。例如,由于芽孢杆菌 RNA 酶是毒性极高的蛋白质,可通过将不同数目的选择者密码子引入到芽孢杆菌 RNA 酶基因内和/或使用可诱导启动子来控制负选择的严格性。在另一实施例中,选择或筛选试剂的浓度(如氨苄青霉素的浓度)可以不同。在本发明的一些方面,因为在前几轮中所需活性较低,严格性可以不同。因此,前几轮适用较低的严格性筛选标准,而后几轮选择适用更严谨的标准。在某些实施方式中,负选择、正选择或者正负选择可重复多次。也可使用多个不同的负选择标记、正选择标记或正负选择标记。在某些实施方式中,正选择标记和负选择标记可以相同。

其他类型的选择/筛选方法也可用于本发明以制备正交翻译组分,如 O-tRNA、O-RS 和能对选择者密码子起反应而加载非天然氨基酸的 O-tRNA/O-RS 配对。例如,负选择标记、正选择标记或正负选择标记可包括发荧光的或在合适反应物存在下能催化发光反应的标记。在另一实施方式中,可通过荧光激活细胞分选(FACS)或发光检测标记的产物。标记任选可包括亲和筛选标记。也参见 Francisco, J. A. 等, (1993), “*Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface* (在外表面表达功能性抗体片段的大肠杆菌的制备与荧光激活细胞分选)”, Proc Natl Acad Sci USA, 90: 10444-8。

制备重组正交 tRNA 的其它方法可见,例如名为“METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS (制备正交 tRNA-氨酰-tRNA 合成酶配对的方法和组合物)”的国际公布 WO 2002/086075; 名为“EXPANDING THE EUKARYOTIC GENETIC CODE (扩展真核生物遗传密码)”的 WO 2004/094593; 与 2004 年 7 月 7 日提交的 WO 2005/019415。也参见 Forster 等, (2003), “*Programming peptidomimetic synthetases by translating genetic codes designed de novo* (通过翻译从头设计的遗传密码来

编程肽模拟合成酶)”，PNAS，100 (11): 6353-6357；和 Feng 等，(2003)，
“*Expanding tRNA recognition of a tRNA synthetase by a single amino acid change* (通过单氨基酸改变扩大 tRNA 合成酶的 tRNA 识别)”，PNAS，100 (10): 5676-5681。

正交氨酰-tRNA 合成酶(O-RS)

本发明的 O-RS 在体外或体内能用非天然氨基酸优先氨酰化 O-tRNA。可通过含 O-RS 的多肽和/或编码 O-RS 或其部分的多核苷酸将本发明的 O-RS 提供给翻译系统，如细胞。例如，示例性 O-RS 包含 SEQ ID NO:4 所示氨基酸序列或其保守变体。在另一实施例中，O-RS 或其部分是由编码含本文序列表和实施例所示的氨基酸序列的多核苷酸序列或其互补多核苷酸序列编码。可参见，例如 SEQ ID NO:5 所示多核苷酸。

鉴定可与 O-tRNA 一起应用的正交氨酰-tRNA 合成酶(O-RS)，例如 O-RS 的方法也是本发明的特征之一。例如，一种方法包括选择(如正选择)第一物种的细胞群，其中所述细胞各自含有：1) 多种氨酰-tRNA 合成酶(RS)的成员之一(例如，所述多种 RS 可包括突变型 RS、衍生自第一物种之外的物种的 RS，或突变型 RS 和衍生自第一物种之外的物种的 RS 二者)；2) 正交 tRNA(O-tRNA)(例如，一个或多个物种的)；以及 3) 编码选择(例如正选择)标记并含有至少一个选择者密码子的多核苷酸。与缺乏所述多种 RS 的成员或成员数目少的细胞相比，选择或筛选显示抑制效率提高的那些细胞。可用本领域已知和本文所述的方法测定抑制效率。抑制效率提高的细胞包含能氨酰化 O-tRNA 的活性 RS。将第一物种的第一组 tRNA 的活性 RS 氨酰化的水平(体外或体内)与第二物种的第二组 tRNA 被活的 RS 氨酰化的水平作比较。通过可检测物质(例如，标记的非天然氨基酸)测定氨酰化水平。通常选择与第一组 tRNA 相比更有效地氨酰化第二组 tRNA 的活性 RS，从而获得能与 O-tRNA 联用的有效(优化)正交氨酰-tRNA 合成酶。采用这种方法鉴定的 O-RS 也是本发明的特征之一。

可用许多试验来测定氨酰化。这些试验可在体外或体内进行。例如，体外氨酰化试验可见例如 Hoben 和 Soll，(1985)，*Methods Enzymol.*，113:

55-59。也可联用报道分子和正交翻译组分并检测细胞中的报道分子来测定氨酰化，所述细胞表达含有至少一个选择者密码子并编码某蛋白质的多核苷酸。也参见名为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS(非天然氨基酸的体内掺入)”的 WO 2002/085923 和名为“EXPANDING THE EUKARYOTIC GENETIC CODE(扩展真核生物遗传密码)”的 WO 2004/094593。

还可进一步改造鉴定的 O-RS 以改变该合成酶的底物特异性，从而使得 O-tRNA 只加载上所需的非天然氨基酸，而不是任何常见的 20 种氨基酸。产生对非天然氨基酸具有底物特异性的正交氨酰 tRNA-合成酶的方法包括，例如通过组合合成酶的不同结构域而使合成酶在其活性位点、在其编辑机制位点、在不同位点发生突变等，并应用选择方法。也可以采用先正选择再负选择的方案。在正选择中，抑制引入阳性标记的一个或多个非重要位置的选择者密码子使得细胞在正选择压力下存活。在同时有天然和非天然氨基酸存在时，如此存活的细胞编码使正交抑制型 tRNA 带有天然或非天然氨基酸的活性合成酶。在负选择中，抑制引入阴性标记的一个或多个非重要位置的选择者密码子使合成酶失去天然氨基酸特异性。经正、负选择而存活的细胞编码只用非天然氨基酸氨酰化(加载)正交抑制型 tRNA 的合成酶。然后通过例如 DNA 改组或其它递归诱变方法进一步诱变这些合成酶。

可采用本领域已知的各种诱变技术产生突变型 O-RS 文库。例如，可用位点特异性突变、随机点突变、同源重组、DNA 改组或其它递归诱变方法，嵌合构建或它们的组合来制备突变型 RS。例如，可从两种或更多种其它如较小、多样性较低的“亚文库”来制备突变型 RS 文库。本发明也包括 RS 的嵌合文库。应注意，可以构建各种生物(例如微生物，如真细菌或古细菌)的 tRNA 合成酶文库，例如具有天然多样性的文库(参见，例如授予 Short 等的美国专利号 6,238,884；授予 Schallenberger 等的美国专利号 5,756,316；授予 Petersen 等的美国专利号 5,783,431；授予 Thompson 等的美国专利号 5,824,485；授予 Short 等的美国专利号 5,958,672)，并任选构建和筛选正交配对。

一旦合成酶经历正和负选择/筛选方法,就可进一步诱变这些合成酶。例如可分离编码 O-RS 的核酸;从该核酸产生编码突变 O-RS 的一组多核苷酸(例如通过随机诱变、位点特异性诱变、重组或它们的任何组合);可重复各步骤或各步骤的组合直至获得能用非天然氨基酸优先氨酰化 O-tRNA 的突变 O-RS。在本发明的一些方面,这些步骤可进行多次,例如至少两次。

本发明方法也可采用其它水平的选择/筛选严格性来产生 O-tRNA、O-RS 或其配对。可改变 O-RS 产生方法中一个或两个步骤的选择或筛选严格性。这包括,例如,改变选择/筛选试剂的用量等。也可额外进行几轮正和/或负选择。选择或筛选也可包括以下一种或多种改变:氨基酸渗透性、翻译效率、翻译保真度(translational fidelity)等。一种或多种改变通常依据用正交 tRNA-tRNA 合成酶配对来产生蛋白质的生物中一种或多种基因突变。

产生 O-RS 并改变合成酶底物特异性的其它常规细节见名为“METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS(产生正交 tRNA-氨酰基 tRNA 合成酶配对的方法和组合物)”的 WO 2002/086075 和名为“EXPANDING THE EUKARYOTIC GENETIC CODE(扩展真核生物遗传密码)”的 WO 2004/094593。也可参见 Wang 和 Schultz, “Expanding the Genetic Code(扩展遗传密码)”, *Angewandte Chemie Int. Ed.*, 44(1): 34-66 (2005), 其内容以引用方式全文纳入。

来源和宿主生物

本发明正交翻译组分(O-tRNA 和 O-RS)可得自任何生物(或生物的组合),可用于任何其它物种的宿主翻译系统,只要所述 O-tRNA/O-RS 组分与宿主系统能以正交方式起作用。某正交配对中的 O-tRNA 和 O-RS 无需得自同一生物。在一些方面,正交组分得自用于真细菌宿主系统的古菌(即,古细菌)基因。

例如,正交 O-tRNA 可衍生自古菌生物,例如古细菌,如詹氏甲烷球菌、嗜热碱甲烷杆菌;盐细菌,如沃氏富盐菌和盐细菌种 *NRC-1*;闪烁古生球

菌、激烈火球菌、极端嗜热古菌、嗜热泉生古细菌、海沼甲烷球菌、甲烷嗜高热菌(*Methanopyrus kandleri*)、梅氏甲烷八叠球菌(*Mm*)、耐超高温热棒菌(*Pyrobaculum aerophilum*)、深海火球菌、硫磺矿硫化叶菌(*Ss*)、超嗜热古菌(*Sulfolobus tokodaii*)、嗜酸热原体、火山热原体等；或真细菌，如大肠杆菌、嗜热栖热菌、枯草芽胞杆菌(*Bacillus subtilis*)、嗜热脂肪芽孢杆菌等；而正交 O-RS 可得自以下生物(或生物的组合)，例如古细菌，如詹氏甲烷球菌、嗜热碱甲烷杆菌；盐细菌，如沃氏富盐菌和盐细菌种 *NRC-1*；闪烁古生球菌、激烈火球菌、极端嗜热古菌、嗜热泉生古细菌、海沼甲烷球菌、甲烷嗜高热菌、梅氏甲烷八叠球菌、耐超高温热棒菌、深海火球菌、硫磺矿硫化叶菌、超嗜热古菌、嗜酸热原体、火山热原体等；或真细菌，如大肠杆菌、嗜热栖热菌、枯草芽胞杆菌、嗜热脂肪芽孢杆菌等。在一个实施方式中，真核生物来源，例如植物、藻类、原生动物、真菌、酵母菌、动物(例如哺乳动物、昆虫、节肢动物等)等也可用作 O-tRNA 和 O-RS 的来源。

O-tRNA/O-RS 配对的各组分可得自同一生物或不同生物。在一个实施方式中，O-tRNA/O-RS 配对可来自同一生物。或者，O-tRNA/O-RS 配对的 O-tRNA 和 O-RS 可来自不同生物。

O-tRNA、O-RS 或 O-tRNA/O-RS 配对可在体内或体外选择或筛选和/或可用于细胞，例如真细菌细胞，从而产生含有非天然氨基酸的多肽。所用的真细菌细胞例如但不限于大肠杆菌、嗜热栖热菌、枯草芽胞杆菌、嗜热脂肪芽孢杆菌等。含本发明翻译组分的真细菌细胞组合物也是本发明特征之一。

为在一种生物中筛选 O-tRNA 和/或 O-RS 以用于另一种生物，也可参见 2004 年 4 月 16 日提交的名为“EXPANDING THE EUKARYOTIC GENETIC CODE(扩展真核生物遗传密码)”的国际申请公布号 WO 2004/094593。

虽然正交翻译系统(例如，含有 O-RS、O-tRNA 和非天然氨基酸)可利用培养的宿主细胞产生含非天然氨基酸的蛋白质，但并非要求本发明的正交翻译系统需要完整的、有活力的宿主细胞。例如，在细胞提取物存在下，正交翻译系统可利用无细胞系统。实际上，使用无细胞的体外转录/翻译系

统产生蛋白质是公认的技术。利用本文所述的正交翻译系统组分，使这些体外系统适用于产生含非天然氨基酸的蛋白质也属于本发明的范围。

选择者密码子

本发明的选择者密码子扩大了蛋白质生物合成机理的遗传密码子框架。例如，选择者密码子包括，如独特的三碱基密码子、无义密码子(如终止密码子(如琥珀密码子(UAG)或乳白密码子(UGA)))、非天然密码子、至少四碱基的密码子、罕用密码子等。可将许多选择者密码子引入所需基因，例如一个以上、两个以上、三个以上等。利用不同的选择者密码子，可利用多对正交 tRNA/合成酶配对，从而能利用这些不同的选择者密码子同时位点特异性地掺入多个非天然氨基酸，例如包含至少一个非天然氨基酸。

在一个实施方式中，这些方法包括利用作为终止密码子的选择者密码子在细胞中体内掺入非天然氨基酸。例如，制备能识别终止密码子的 O-tRNA，并由 O-RS 用非天然氨基酸氨酰化该 O-tRNA。宿主的天然氨酰-tRNA 合成酶不识别该 O-tRNA。可采用常规的定点诱变将终止密码子引入编码感兴趣多肽的多核苷酸中感兴趣的位点。可参见，例如 Sayers, J.R. 等, (1988), “5', 3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis(硫代磷酸酯寡核苷酸定向诱变中的 5', 3' 核酸外切酶)”, *Nucleic Acids Res*, 791-802。当例如在体内联用 O-RS、O-tRNA 和编码感兴趣多肽的核酸时，可对终止密码子起反应而掺入非天然氨基酸，从而获得在特定位点含有非天然氨基酸的多肽。在本发明的一个实施方式中，用作选择者密码子的终止密码子是琥珀密码子 UAG 和/或乳白密码子 UGA。在一实施例中，UAG 和 UGA 都用作选择者密码子的遗传密码可编码 22 个氨基酸，同时保留赭石无义密码子，UAA，其是最丰富的终止信号。

体内掺入非天然氨基酸不会显著干扰宿主细胞。例如，在非真核细胞，如大肠杆菌中，由于 UAG 密码子的抑制效率依赖于 O-tRNA(如琥珀抑制型 tRNA)与释放因子 1(RF1)(其可结合 UAG 密码子，进而引起延伸中的肽从核糖体释放)之间的竞争，因此可通过，例如提高 O-tRNA(如抑制型 tRNA)的表达水平，或利用 RF1 缺陷型菌株调节抑制效率。在真核细胞中，由于

UAG 密码子的抑制效率依赖于 O-tRNA(如琥珀抑制型 tRNA)和真核释放因子(如 eRF)(其可结合终止密码子, 进而引起延伸中的肽从核糖体释放)之间的竞争, 因此可通过, 例如提高 O-tRNA(如抑制型 tRNA)的表达水平调节抑制效率。此外, 也可存在其它化合物, 例如还原剂, 如二硫苏糖醇(DTT)。

非天然氨基酸也可由罕用密码子编码。例如, 当体外蛋白合成反应中的精氨酸浓度下降时, 证实罕用的精氨酸密码子 AGG 可利用丙氨酸酰化的合成 tRNA 而有效插入 ala。可参见 Ma 等, (1993), *Biochemistry*, 32:7939。在此情况中, 合成 tRNA 可与大肠杆菌中作为次要成分存在的天然 tRNA^{Arg} 竞争。此外, 一些生物不能利用所有三联密码子。已可在体外转录/翻译提取物中利用藤黄微球菌(*Micrococcus luteus*)的未指定密码子 AGA 插入氨基酸。可参见, 例如 Kowal 和 Oliver, (1997), *Nucl. Acid.Res.*, 25:4685。可产生本发明的各组分以体内利用这些罕用密码子。

选择者密码子也可包括延伸密码子, 例如四个或四个以上碱基的密码子, 如四个、五个、六个或更多碱基的密码子。四碱基密码子的例子包括, 例如 AGGA、CUAG、UAGA、CCCU 等。五碱基密码子的例子包括, 例如 AGGAC、CCCCU、CCCUC、CUAGA、CUACU、UAGGC 等。本发明方法可包括使用基于移码抑制的延伸密码子。四个或四个以上碱基的密码子可将例如一个或多个非天然氨基酸插入同一蛋白质。在其它实施方式中, 反密码子环可解码例如至少一种四碱基密码子、至少一种五碱基密码子或至少一种六碱基密码子或更多。由于可能的四碱基密码子有 256 种, 所以同一细胞中可用四个或四个以上碱基的密码子编码多种非天然氨基酸。也可参见, Anderson 等, (2002), “Exploring the Limits of Codon and Anticodon Size(密码子和反密码子大小极限的研究)”, *Chemistry and Biology*, 9: 237-244 和 Magliery, (2001), “Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in *Escherichia coli*(扩增遗传密码: 在大肠杆菌中选择有效的四碱基密码子抑制剂并用文库方法鉴定“变化的”四碱基密码子)”, *J. Mol. Biol.*, 307: 755-769。

例如, 已采用体外生物合成方法利用四碱基密码子将非天然氨基酸掺

入蛋白质。参见, 例如 Ma 等, (1993), *Biochemistry*, 32: 7939 和 Hohsaka 等, (1999), *J. Am. Chem. Soc.*, 121: 34。联用 CGGG 和 AGGU 及两种化学酰化的移码抑制型 tRNA 在体外将 2-萘基丙氨酸和赖氨酸的 NBD 衍生物同时掺入链霉素亲和素。可参见, 例如 Hohsaka 等, (1999), *J. Am. Chem. Soc.*, 121: 12194。在体内研究中, Moore 等检测了含 NCUA 反密码子的 tRNA^{Leu} 衍生物抑制 UAGN 密码子(N 可以是 U、A、G 或 C)的能力, 发现含 UCUA 反密码子的 tRNA^{Leu} 能解码四联 UAGA, 其效率为 13 到 26%, 而在 0 或 -1 读框中几乎不解码。可参见 Moore 等, (2000), *J. Mol. Biol.*, 298: 195。在一个实施方式中, 本发明可利用基于罕用密码子或无义密码子的延伸密码子, 它们能降低在其它不期望位点的错义连读(missense readthrough)和移码抑制。四碱基密码子已在各种正交系统中用作选择者密码子。可参见, 例如, WO 2005/019415; WO 2005/007870 和 WO 2005/07624。也可参见 Wang 和 Schultz, “Expanding the Genetic Code(扩展遗传密码)”, *Angewandte Chemie Int. Ed.*, 44(1): 34-66 (2005), 其内容通过引用的方式全文纳入。虽然以下例子利用琥珀选择者密码子, 但通过改进本文的实例使之包含四碱基 O-tRNA 和经修饰包含与之前描述的各非天然氨基酸 O-RS 的突变相似的突变的合成酶, 也可使用四碱基或更多碱基的密码子。

对于给定系统, 选择者密码子还可包括天然三碱基密码子中内源性系统不用(或很少用)的那个天然碱基密码子。例如, 这包括缺乏能识别该天然三碱基密码子的 tRNA 的系统, 和/或该三碱基密码子是罕用密码子的系统。

选择者密码子任选包含非天然碱基对。这些非天然碱基对进一步扩大了现有的遗传字符集(genetic alphabet)。一种额外的碱基对可将三联密码子的数目从 64 增加至 125。第三碱基对的特性包括: 稳定和选择性的碱基配对, 利用聚合酶高保真地有效酶促掺入 DNA, 新生非天然碱基对合成后引物继续有效延伸。适用于这些方法和组合物的非天然碱基对的描述包括: 例如 Hirao 等, (2002), “An unnatural base pair for incorporating amino acid analogues into protein(将氨基酸类似物掺入蛋白质的非天然碱基对)”, *Nature Biotechnology*, 20: 177-182。也可参见 Wu, Y.等, (2002), *J. Am. Chem. Soc.*, 124: 14626-14630。下文列出了其它的相关出版物。

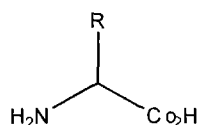
对于体内使用, 非天然核苷是膜可渗透的, 并可磷酸化形成相应的磷酸三酯。此外, 增加的遗传信息稳定且不为细胞酶所破坏。Benner 与他人先前的工作利用了不同于经典 Watson-Crick 配对的氢键模式, 其中最值得注意的例子是异-C: 异-G 配对(iso-C: iso-G pair)。参见, 例如 Switzer 等, (1989), *J. Am. Chem. Soc.*, 111: 8322; Piccirilli 等, (1990), *Nature*, 343: 33; Kool, (2000), *Curr. Opin. Chem. Biol.*, 4: 602。这些碱基通常与天然碱基有一定程度的错配, 因而不能酶促复制。Kool 和同事证明, 碱基间的疏水包装相互作用(hydrophobic packing interaction)可替代氢键以驱动碱基对形成。参见 Kool, (2000), *Curr. Opin. Chem. Biol.*, 4: 602; Guckian 和 Kool, (1998), *Angew. Chem. int. Ed. Engl.*, 36: 2825。在致力于开发符合以上所有要求的非天然碱基对的过程中, Schultz、Romesberg 和同事已系统地合成并研究了一系列非天然疏水碱基。发现 PICS: PICS 自身配对比天然碱基对更稳定, 并可用大肠杆菌 DNA 聚合酶 I 的 Klenow 片段(KF)有效地掺入 DNA。可参见, 例如 McMinn 等, (1999), *J. Am. Chem. Soc.*, 121: 11586; 和 Ogawa 等, (2000), *J. Am. Chem. Soc.*, 122: 3274。就生物学功能而言, KF 合成 3MN: 3MN 自身配对的效率和选择性足够。可参见, 例如 Ogawa 等, (2000), *J. Am. Chem. Soc.*, 122: 8803。然而, 这两种碱基均作为进一步复制的链终止子而起作用。近来开发了可用于复制 PICS 自身配对的突变型 DNA 聚合酶。此外, 可复制 7AI 自身配对。可参见, 例如 Tae 等, (2001), *J. Am. Chem. Soc.*, 123: 7439。还已开发了与 Cu(II)结合后可形成稳定配对的金属碱基(metallobase)对 Dipis: Py。可参见, Meggers 等, (2000), *J. Am. Chem. Soc.*, 122: 10714。因为延伸密码子和非天然密码子在本质上与天然密码子正交, 所以本发明方法可利用该特性产生它们的正交 tRNA。

还可利用翻译旁路系统将非天然氨基酸掺入所需多肽。在翻译旁路系统中, 可将一个大序列插入基因, 但该序列不翻译成蛋白质。该序列含有可作为提示的结构, 从而能诱导核糖体跳过该序列然后恢复该插入序列的下游翻译。

非天然氨基酸

本文所用的非天然氨基酸指除硒代半胱氨酸和/或吡咯赖氨酸和以下 20 种遗传编码的 α -氨基酸以外的任何氨基酸、修饰的氨基酸或氨基酸类似物：丙氨酸、精氨酸、天冬酰胺、天冬氨酸、半胱氨酸、谷氨酰胺、谷氨酸、甘氨酸、组氨酸、异亮氨酸、亮氨酸、赖氨酸、甲硫氨酸、苯丙氨酸、脯氨酸、丝氨酸、苏氨酸、色氨酸、酪氨酸、缬氨酸。 α -氨基酸的通用结构如式 I 所示：

I



非天然氨基酸一般具有式 I 所示的任何结构，其中 R 基团是除 20 种天然氨基酸中所用以外的任何取代基。20 种天然氨基酸的结构可参见，例如《生物化学》(Biochemistry), L.Stryer 编，第三版，1988，福里曼公司(Freeman and Company)，纽约。注意，本发明的非天然氨基酸可以是除以上 20 种 α -氨基酸以外的天然化合物。

由于本发明的非天然氨基酸与天然氨基酸通常区别在侧链，所以非天然氨基酸与其它氨基酸(例如天然或非天然的)形成酰胺键的方式与天然蛋白质中酰胺键的形成方式相同。然而，非天然氨基酸的侧链不同于天然氨基酸的侧链。

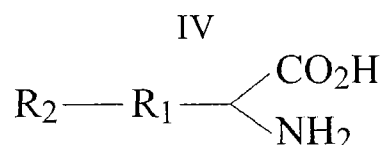
本文特别感兴趣的是非天然香豆素氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸 (也写作 7-羟基香豆素-乙基甘氨酸；参见图 1，结构 1)。除了 7-羟基香豆素-乙基甘氨酸非天然氨基酸外，可将其它非天然氨基酸同时掺入感兴趣的多肽，例如联用合适的第二 O-RS/O-tRNA 配对与本发明提供的正交配对。已知许多这样的其它非天然氨基酸和合适的正交配对。参见本文内容和本文引用的参考文献。例如，参见 Wang 和 Schultz，“Expanding the Genetic Code(扩展遗传密码)”，*Angewandte Chemie Int. Ed.*，44(1):34-66 (2005)；Xie 和 Schultz，“An Expanding Genetic Code(扩展遗传密码)”，*Methods* 36(3):227-238 (2005)；Xie 和 Schultz，“Adding Amino Acids to the Genetic Repertoire(将氨基酸加入遗传库)”，*Curr. Opinion in Chemical Biology* 9(6):548-554 (2005)；和 Wang 等，“Expanding the Genetic Code(扩

展遗传密码)”，*Annu. Rev. Biophys. Biomol. Struct.*，35:225-249 (2006)；这些文献各自的内容以引用的方式全文纳入本文作为参考。

虽然本文所述实施例中主要感兴趣的是 L-(7-羟基香豆素-4-基)乙基甘氨酸 (图 1, 结构 1 所示), 但并非要将本发明严格限制于该结构。实际上, 不难制备各种易于获得且结构相关的类似物, 这些类似物保留主要的荧光特性, 而且本发明的氨酰-tRNA 合成酶(例如, SEQ ID NO: 4 所示 O-RS)还特异性识别这些类似物。这些相关的荧光类似物属于本发明的范围。

在其它非天然氨基酸中, 例如式 I 中的 R 可任选含有烷基-、芳基-、酰基-、胍、氰基-、卤代-、酰肼、烯基、醚、硼酸基(borate)、硼酸酯基(boronate)、磷、膦酰基、膦、烯酮、亚胺、酯、羟胺、胺等, 或上述基团的任何组合。感兴趣的其它非天然氨基酸包括但不限于: 含光可活化交联剂的氨基酸、自旋-标记的氨基酸、荧光氨基酸、金属结合氨基酸、含金属的氨基酸、放射性氨基酸、含有新官能团的氨基酸、能与其它分子共价或非共价相互作用的氨基酸、光锁定(photocaged)和/或光致异构(photoisomerizable)的氨基酸、含生物素或生物素类似物的氨基酸、含酮基的氨基酸、糖基化的氨基酸、与氨基酸侧链相连的糖部分、含聚乙二醇或聚醚的氨基酸、重原子取代的氨基酸、化学方法可切割或光可切割的氨基酸、与天然氨基酸相比侧链延长的氨基酸(如聚醚或长链烃, 如超过约 5 个、约 10 个碳原子等)、含碳连接糖的氨基酸、含氨基硫代酸的氨基酸以及含一个或多个毒性部分的氨基酸。

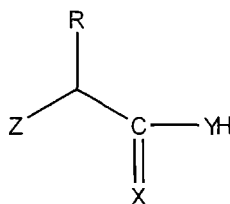
本发明另一方面提供具有以下式 IV 所示通用结构的非天然氨基酸:



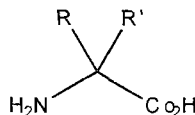
具有此结构的非天然氨基酸一般可以是任何结构, 其中 R_1 是 20 种天然氨基酸之一(例如, 酪氨酸或苯丙氨酸)所用的取代基, R_2 是取代基。因此, 此类非天然氨基酸可视作天然氨基酸衍生物。

除了含有图 1, 结构 1 所示香豆素侧链的非天然氨基酸外, 非天然氨基酸还可任选含有经修饰的骨架结构, 例如式 II 和 III 所示的结构:

II



III



其中，Z 一般包含 OH、NH₂、SH、NH-R'或 S-R'；X 和 Y 可以相同或不同，一般包含 S 或 O，R 和 R'可任选相同或不同，通常选自与上述式 I 所示非天然氨基酸的 R 基团相同的组成以及氢。例如，本发明的非天然氨基酸还任选在式 II 和 III 所示氨基或羧基中包含取代基。该类型的非天然氨基酸包括但不限于：例如含有 20 种常见氨基酸的相应侧链或非天然侧链的 α-羟酸、α-硫代酸、α-氨基硫代羧酸酯等。另外，α 碳上的取代基还任选包括 L、D 或 α-α-双取代氨基酸，如 D-谷氨酸、D-丙氨酸、D-甲基-O-酪氨酸、氨基丁酸等。其它替代结构包括环形氨基酸，如脯氨酸类似物以及 3、4、6、7、8 和 9 元环脯氨酸类似物，β 和 γ 氨基酸，如取代的 β 丙氨酸和 γ-氨基丁酸。

在一些方面，本发明使用 L-构型的非天然氨基酸。然而，本发明并非仅限于使用 L-构型的非天然氨基酸。还考虑了这些非天然氨基酸的 D-对映体可用于本发明。

本发明所用的非天然氨基酸不严格局限于香豆素非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸。本领域技术人员知道不难获得天然氨基酸的各种非天然类似物。例如，但不限于不难制备衍生自酪氨酸的非天然(氨基酸)。酪氨酸类似物包括，例如对位取代的酪氨酸、邻位取代的酪氨酸和间位取代的酪氨酸，其中该取代的酪氨酸含有炔基、乙酰基、苯甲酰基、氨基、肼、羟胺、巯基、羧基、异丙基、甲基、C₆-C₂₀ 直链或支链烃、饱和或不饱和的烃、O-甲基、聚醚基团、硝基等。此外，也包括多取代的芳环。本发明的谷氨酰胺类似物包括但不限于：α-羟基衍生物、γ-取代的衍生物、环形衍生物和酰胺取代的谷氨酰胺衍生物。苯丙氨酸类似物的例子包括但

不限于：对位取代的苯丙氨酸、邻位取代的苯丙氨酸和间位取代的苯丙氨酸，其中所述取代基含有炔基、羟基、甲氧基、甲基、烯丙基、醛基、硝基、巯基或酮基等。非天然氨基酸的具体例子包括但不限于：对-乙基硫代羰基-L-苯丙氨酸、对-(3-氧代丁酰基)-L-苯丙氨酸、1,5-丹酰-丙氨酸、7-氨基-香豆素氨基酸、7-羟基-香豆素氨基酸、硝基苄基-丝氨酸、O-(2-硝基苄基)-L-酪氨酸、对-羧甲基-L-苯丙氨酸、对-氰基-L-苯丙氨酸、间-氰基-L-苯丙氨酸、联苯基丙氨酸、3-氨基-L-酪氨酸、联吡啶基丙氨酸、对-(2-氨基-1-羟乙基)-L-苯丙氨酸、对-异丙基硫代羰基-L-苯丙氨酸、3-硝基-L-酪氨酸和对-硝基-L-苯丙氨酸。还包括对-炔丙氧基苯丙氨酸、3,4-二羟基-L-苯丙氨酸(DHP)、3,4,6-三羟基-L-苯丙氨酸、3,4,5-三羟基-L-苯丙氨酸、4-硝基-苯丙氨酸、对-乙酰基-L-苯丙氨酸、O-甲基-L-酪氨酸、L-3-(2-萘基)丙氨酸、3-甲基-苯丙氨酸、O-4-烯丙基-L-酪氨酸、4-丙基-L-酪氨酸、3-硝基-酪氨酸、3-巯基-酪氨酸、三-O-乙酰基-GlcNAc β -丝氨酸、L-多巴(Dopa)、氟化苯丙氨酸、异丙基-L-苯丙氨酸、对-叠氮基-L-苯丙氨酸、对-酰基-L-苯丙氨酸、对-苯甲酰基-L-苯丙氨酸、L-磷酸丝氨酸、磷酰丝氨酸、磷酰酪氨酸、对-碘代-苯丙氨酸、对-溴代苯丙氨酸、对-氨基-L-苯丙氨酸以及异丙基-L-苯丙氨酸等。本文引用的参考文献披露了各种非天然氨基酸的结构。还可参见国际公布 WO 2004/094593，名为“EXPANDING THE EUKARYOTIC GENETIC CODE(扩展真核细胞遗传密码)”；和 2005 年 10 月 27 日提交的 WO 2006/110182，名为“ORTHOGONAL TRANSLATION COMPONENTS FOR THE *IN VIVO* INCORPORATION OF UNNATURAL AMINO ACIDS(体内掺入非天然氨基酸的正交翻译组分)”。

非天然氨基酸的化学合成

上述许多非天然氨基酸可以从，例如美国的西格玛公司 (Sigma)或美国威斯康星州密尔沃基市的阿尔德里奇公司 (Aldrich)等购得。那些不能商品化购得的非天然氨基酸可任选如各种出版物所述或采用本领域技术人员已知的标准方法合成。有机合成技术可参见，例如 Fessenden 和 Fessenden, Organic Chemistry(有机化学), (1982, 第二版, 威拉德格兰特出版社(Willard

Grant Press), 波士顿, 马萨诸塞州); March, *Advanced Organic Chemistry*(高级有机化学)(第三版, 1985, 威立父子公司(Wiley and Sons), 纽约); 以及 Carey 和 Sundberg 的 *Advanced Organic Chemistry*(高级有机化学)(第三版, A 和 B 部分, 1990, 普莱纳姆出版社(Plenum Press), 纽约)。其它描述非天然氨基酸合成的出版物包括: 名为“非天然氨基酸的体内掺入”的国际专利申请 WO 2002/085923; Matsoukas 等, (1995) *J. Med. Chem.*, 38, 4660-4669; King 和 Kidd, (1949) “A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates”(从邻苯二甲酰化中间体合成谷氨酰胺和 γ -二肽的新方法). *J. Chem. Soc.*, 3315-3319; Friedman 和 Chatterji, (1959) “Synthesis of Derivatives of Glutamine as Model Substrates for Anti-tumor Agents(合成谷氨酰胺衍生物作为抗肿瘤药物的模拟底物)”. *J. Am. Chem. Soc.* 81, 3750-3752; Craig 等, (1988) “Absolute Configuration of the Enantiomers of 7-Chloro-4[[4-(diethylamino)-1-methylbutyl]amino]quinoline(Chloroquine)(7-氯-4[[4-(二乙基氨基)-1-甲基丁基]氨基]喹啉(氯喹)对映体的绝对构型)”. *J. Org. Chem.* 53, 1167-1170; Azoulay 等, (1991) “Glutamine analogues as Potential Antimalarials(谷氨酰胺类似物作为潜在的抗疟药)”, *Eur. J. Med. Chem.* 26: 201-205; Koskinen 和 Rapoport (1989) “Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues(合成 4-取代的脯氨酸作为构象约束的氨基酸类似物)”. *J. Org. Chem.* 54: 1859-1866; Christie 和 Rapoport, (1985) “Synthesis of Optically Pure Pípecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization(从 L-天冬酰胺合成光学纯的哌啶酸酯, 应用于通过氨基酸脱羰基和亚氨鎓离子环化总合成(+)-阿扑长春胺)”. *J. Org. Chem.* 1989:1859-1866; Barton 等, (1987) “Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L-and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives(采用自由基化学方法合成新的 α -氨基酸和衍生物: 合成 L-和 D- α 氨基酸、L- α -氨基庚二酸以及合适的饱和衍生物)”.

Tetrahedron Lett. 43:4297-4308; 以及 Subasinghe 等, (1992) “Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site(使君子氨酸类似物: β -杂环 2-氨基丙酸衍生物的合成及其在使君子氨酸敏化部位的活性)”. *J. Med. Chem.* 35:4602-7. 还可参见 2003 年 12 月 22 日提交的名为 “Protein Arrays(蛋白质阵列)” 的国际专利申请 WO2004/058946。

非天然氨基酸的细胞摄取

细胞摄取非天然氨基酸通常是设计和选择非天然氨基酸, 例如用于掺入蛋白质中应考虑的问题之一。例如, α -氨基酸的高电荷密度提示这些化合物不可能透过细胞。天然氨基酸通过蛋白质转运系统的收集(作用)摄取入细胞, 这些系统往往显示程度不同的氨基酸特异性。可进行快速筛选来评估细胞将摄取哪种非天然氨基酸(如果有的话)。可参见, 例如 2003 年 12 月 22 日提交的名为 “Protein Arrays(蛋白质阵列)” 的国际公布 WO 2004/058946 中的毒性试验; Liu 和 Schultz, (1999), “Progress toward the evolution of an organism with an expanded genetic code(含扩增遗传密码的生物演化的进展)”, *PNAS*, 96: 4780-4785。虽然不难采用各种试验分析摄取情况, 但设计适合于细胞摄取途径的非天然氨基酸的另一种方法是提供生物合成途径, 从而能在体内产生氨基酸。

非天然氨基酸的生物合成

细胞中已有许多生物合成途径来产生氨基酸和其它化合物。虽然自然界中, 例如细胞中, 可能不存在特定非天然氨基酸的生物合成方法, 但本发明提供了这种方法。例如, 通过加入新的酶或修饰现有的宿主细胞途径可以在宿主细胞中任选产生非天然氨基酸的生物合成途径。其它新的酶任选是天然产生的酶或人工开发的酶。例如, 生物合成对氨基苯丙氨酸(如 WO 2002/085923 中的实施例所示)依赖加入其它生物的已知酶的混合物。可通过用含这些酶的基因的质粒转染细胞而将这些基因引入细胞。当这些基因在细胞中表达时, 它们提供了合成所需化合物的酶途径。可任选加入的

酶类型的例子见以下实施例。其它酶序列见例如 Genbank。也可以相同方式将人工开发的酶任选加入细胞。可以此方式操控细胞机理和资源用以产生非天然氨基酸。

实际上,可采用各种方法产生新的酶从而在体内或体外用于生物合成途径,改进现有途径,产生非天然氨基酸。开发酶和其它生物合成途径组分的许多现有方法适用于本发明以产生非天然氨基酸(或者,实际上,用于开发合成酶使之具有新的底物特异性或其它感兴趣的活性)。例如,可任选采用 DNA 改组来开发新的酶和/或这些酶的途径,从而能在体外或体内产生非天然氨基酸(或产生新的合成酶)。可参见,例如 Stemmer, (1994),

“Rapid evolution of a protein in vitro by DNA shuffling(通过 DNA 改组在体外快速进化蛋白质)”, *Nature*, 370(4): 389-391; Stemmer, (1994), “DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution(通过随机断裂和装配的 DNA 改组: 分子进化的体外重组)”, *Proc. Natl. Acad. Sci. USA.*, 91: 10747-10751。相关方法改组关联的(例如同源的)基因家族,从而能快速开具有所需特性的酶。这种“家族基因改组”方法的例子见 Cramer 等, (1998), “DNA shuffling of a family of genes from diverse species accelerates directed evolution(不同种类基因家族的 DNA 改组加速了定向进化)”, *Nature*, 391(6664): 288-291。也可采用称为“产生杂交酶的递增截短”(ITCHY)的 DNA 重组方法来产生新的酶(无论是生物合成途径组分或合成酶),例如 Ostermeier 等, (1999), “A combinatorial approach to hybrid enzymes independent of DNA homology(不依赖 DNA 同源性的杂交酶的组合方法)”, *Nature Biotech*, 17: 1205 中所述。该方法也可用于产生用作一种或多种体外或体内重组方法底物的酶或其它途径变体的文库。也可参见, Ostermeier 等, (1999), “Combinatorial Protein Engineering by Incremental Truncation(采用递增截短的组合蛋白质工程)”, *Proc. Natl. Acad. Sci. USA*, 96: 3562-67; Ostermeier 等, (1999), “Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts(作为工程改造新生物催化剂的方法的递增截短)”, *Biological and Medicinal Chemistry*, 7: 2139-44。另一种方法采用指数集合诱变

(exponential ensemble mutagenesis)来产生酶或其它途径变体的文库,其能例如催化与产生非天然氨基酸(或新合成酶)有关的生物合成反应。在该方法中,平行地随机选取(randomized)感兴趣序列中的小基团残基,以在各不同位置鉴定可产生功能性蛋白质的氨基酸。适用于本发明以产生新酶进而产生非天然氨基酸(或新合成酶)的此类方法的例子见 Delegrave 和 Youvan, (1993), *Biotechnology Research*, 11: 1548-1552。在另一方法中,利用掺杂或简并寡核苷酸的随机或半随机诱变可用于工程改造酶和/或途径成分,例如采用如以下文献所述的通用诱变方法: Arkin 和 Youvan, (1992), “Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis” (优化核苷酸混合物来编码用于半随机诱变的特定氨基酸亚组), *Biotechnology* 10:297-300; 或 Reidhaar-Olson 等, (1991), “Random mutagenesis of protein sequences using oligonucleotide cassettes” (用寡核苷酸盒随机诱变蛋白质序列), *Methods Enzymol.*, 208: 564-86。利用多核苷酸重装配和位点饱和诱变的另一种方法(常称为“非随机”诱变)可用于产生酶和/或途径组分,然后筛选它们行使一种或多种合成酶或生物合成途径功能(例如在体内产生非天然氨基酸)的能力。可参见,例如 Short, “NON-STOCHASTIC GENERATION OF GENETIC VACCINES AND ENZYMES(遗传疫苗和酶的非随机产生)”, WO 00/46344。

这种突变方法的替代方法包括重组生物整个基因组并选择得到的后代的特定途径功能(常称为“全基因组改组”)。该方法适用于本发明,例如通过基因组重组并选择能够产生非天然氨基酸(或其中间体)的生物(大肠杆菌或其它细胞)。例如,以下出版物所指导的方法可应用于途径设计,从而能开发细胞中现有和/或新的途径进而在体内产生非天然氨基酸: Patnaik 等, (2002), “Genome shuffling of lactobacillus for improved acid tolerance(乳酸杆菌基因组改组来提高酸耐受性)”, *Nature Biotechnology*, 20(7): 707-712; 和 Zhang 等, (2002), “Genome shuffling leads to rapid phenotypic improvement in bacteria(基因组改组导致细菌表型快速改善)”, *Nature*, 2月7日, 415: 644-646。

其它可用于生物和代谢途径工程改造(例如为产生所需化合物)的技术

也适用于产生非天然氨基酸。指导有用的路径工程改造方法的出版物的例子包括:Nakamura 和 White,(2003), “Metabolic engineering for the microbial production of 1,3 propanediol(微生物生产 1,3 丙二醇的代谢工程)”, *Curr. Opin. Biotechnol.*, 14(5): 454-9; Berry 等,(2002), “Application of Metabolic Engineering to improve both the production and use of Biotech Indigo(应用代谢工程来提高 Biotech Indigo 的生产和用途)”, *J. Industrial Microbiology and Biotechnology*, 28: 127-133; Banta 等,(2002), “Optimizing an artificial metabolic pathway: Engineering the cofactor specificity of Corynebacterium 2,5-diketo-D-gluconic acid reductase for use in vitamin C biosynthesis(优化人工代谢途径: 工程改造用于维生素 C 生物合成的棒状杆菌 2,5-二酮基-D-葡萄糖酸还原酶的辅助因子特异性)”, *Biochemistry*, 41(20): 6226-36; Selivonova 等,(2001), “Rapid Evolution of Novel Traits in Microorganisms(微生物中新特性的快速进化)”, *Applied and Environmental Microbiology*, 67: 3645, 和许多其它出版物。

无论采用什么方法,用本发明工程改造的生物合成途径产生的非天然氨基酸的浓度足以有效地生物合成蛋白质,例如天然细胞含量,但不应达到显著影响其它细胞氨基酸的浓度或耗尽细胞资源的程度。以此方式在体内所产生的浓度一般为约 10 mM 到约 0.05 mM。一旦细胞被工程改造从而产生了特定途径所需的酶和非天然氨基酸,为了核糖体蛋白质合成和细胞生长,可任选采用体内选择以进一步优化非天然氨基酸的产生。

掺入非天然氨基酸的正交组分

本发明提供制备正交组分从而能在体内对选择者密码子,如琥珀终止密码子、无义密码子、四碱基或多碱基密码子等起反应而将非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸(参见图 1, 结构 1)掺入延伸的多肽链中的方法和组合物。例如,本发明提供了正交-tRNA(O-tRNA)、正交氨酰-tRNA 合成酶(O-RS)及其配对。这些配对可用于将非天然氨基酸掺入延伸的多肽链中。

本发明的组合物包含正交氨酰-tRNA 合成酶(O-RS),所述 O-RS 用 7-

羟基香豆素-乙基甘氨酸优先氨酰化 O-tRNA。在某些实施方式中，O-RS 包含含有 SEQ ID NO:4 的氨基酸序列及其保守变体。在本发明的某些实施方式中，O-RS 用特定的非天然氨基酸优先氨酰化 O-tRNA 胜过任何内源性 tRNA，其中所述 O-RS 对 O-tRNA 有偏好(bias)，其中加载有非天然氨基酸的 O-tRNA 与加载有相同非天然氨基酸的内源性 tRNA 的比例大于 1:1，更优选 O-RS 排他性地或者几乎排他性地加载 O-tRNA。

含 O-RS 的组合物还可任选含有正交 tRNA(O-tRNA)，该 O-tRNA 识别选择者密码子。与包含本文序列表(例如，SEQ ID NO:1)和实施列所列的多核苷酸序列或由其编码的 O-tRNA 的抑制效率相比，在关联合成酶存在下，本发明 O-tRNA 对选择者密码子起反应而通常具有至少约，例如 45%、50%、60%、75%、80%、90%或更高的抑制效率。在一个实施方式中，O-RS 结合 O-tRNA 的抑制效率至少是没有 O-RS 存在下 O-tRNA 的抑制效率高，例如 5 倍、10 倍、15 倍、20 倍、25 倍或更高。在一些方面，O-RS 与 O-tRNA 一起的抑制效率至少是衍生自詹氏甲烷球菌的正交酪氨酰-tRNA 合成酶的抑制效率的 45%。

包含 O-tRNA 的组合物还可任选包含细胞(例如真细菌细胞，如大肠杆菌细胞等，或真核细胞，如酵母细胞)和/或翻译系统。

本发明还提供包含翻译系统的细胞(如真细菌细胞或酵母细胞)，所述翻译系统包含正交-tRNA(O-tRNA)；正交氨酰-tRNA 合成酶(O-RS)和 7-羟基香豆素-乙基甘氨酸非天然氨基酸。O-RS 通常用非天然氨基酸优先氨酰化 O-tRNA 胜过任何内源性 tRNA，该 O-RS 对 O-tRNA 有偏好，加载有非天然氨基酸的 O-tRNA 与加载有相同非天然氨基酸的内源性 tRNA 的比例大于 1:1，更优选该 O-RS 排他性地或者几乎排他性地加载 O-tRNA。O-tRNA 识别第一选择者密码子，O-RS 能用非天然氨基酸优先氨酰化 O-tRNA。在一个实施方式中，O-tRNA 含有 SEQ ID NO:1 所示多核苷酸序列或其互补多核苷酸序列，或者由其编码。在一个实施方式中，该 O-RS 含有 SEQ ID NO:4 所示氨基酸序列和其保守变体。

本发明的细胞还可任选包含其它不同的 O-tRNA/O-RS 配对和第二非天然氨基酸，例如，此 O-tRNA 识别第二选择者密码子，此 O-RS 用第二非

天然氨基酸优先氨酰化相应的 O-tRNA, 其中所述第二氨基酸不同于第一非天然氨基酸。本发明的细胞可任选包含含有编码感兴趣多肽的多核苷酸的核酸, 所述多核苷酸含有所述 O-tRNA 能识别的选择者密码子。

在某些实施方式中, 本发明的细胞是真细菌细胞(如大肠杆菌), 所述细胞含有正交-tRNA(O-tRNA)、正交氨酰-tRNA 合成酶(O-RS)、非天然氨基酸和含有编码感兴趣多肽的多核苷酸的核酸, 其中所述多核苷酸含有该 O-tRNA 能识别的选择者密码子。在本发明的某些实施方式中, 所述 O-RS 用所述非天然氨基酸优先氨酰化所述 O-tRNA, 其效率高于该 O-RS 氨酰化任何内源性 tRNA 的效率。

在本发明的某些实施方式中, 本发明的 O-tRNA 含有本文序列表(例如, SEQ ID NO:1)或实施例中所示的多核苷酸序列或其互补多核苷酸序列, 或由这些序列编码。在本发明的某些实施方式中, O-RS 含有序列表中所示氨基酸序列或其保守变体。在一个实施方式中, O-RS 或其部分由编码本文序列表和实施例所示氨基酸序列的多核苷酸序列或其互补的多核苷酸序列编码。

本发明的 O-tRNA 和/或 O-RS 可衍生自各种生物(如真核和/或非真核生物)。

多核苷酸也是本发明的特征。本发明的多核苷酸(例如, SEQ ID NO:5)包括人工的(如人造的和非天然产生的)多核苷酸, 所述多核苷酸序列含有编码本文序列表所示多肽的核苷酸序列, 和/或与该多核苷酸序列互补。本发明的多核苷酸还可包括在高度严谨条件下能与上述多核苷酸在基本全长的核酸上杂交的核酸。本发明的多核苷酸还包括与天然 tRNA 或相应编码核酸的序列具有, 例如至少 75%、至少 80%、至少 90%、至少 95%、至少 98% 或更高相同性的多核苷酸(但是本发明的多核苷酸不是天然 tRNA 或相应的编码核酸), 所述 tRNA 识别选择者密码子, 如四碱基密码子。与上述任何多核苷酸序列和/或含上述任何序列的保守变体的多核苷酸有, 例如至少 80%、至少 90%、至少 95%、至少 98% 或更高相同性的人工多核苷酸也包括在本发明的多核苷酸内。

含本发明多核苷酸的载体也是本发明的特征。例如, 本发明的载体可

以是质粒、粘粒、噬菌体、病毒、表达载体等。含本发明载体的细胞也是本发明的特征。

O-tRNA/O-RS 配对组分的制备方法也是本发明的特征。用这些方法制备的组分也是本发明的特征。例如，产生至少一种与细胞正交的 tRNA(O-tRNA)的方法包括制备突变体 tRNA 文库；使突变体 tRNA 文库中各成员的反密码子环突变使得其能识别选择者密码子，藉此提供潜在的 O-tRNA 文库，并对第一物种的第一细胞群进行负选择，所述细胞含有潜在 O-tRNA 文库的成员。负选择可去除含潜在 O-tRNA 文库成员的细胞，该潜在 O-tRNA 可被细胞的内源性氨酰-tRNA 合成酶(RS)氨酰化。由此提供与第一物种的细胞正交的 tRNA 库，从而至少提供一种 O-tRNA。还提供用本发明方法产生的 O-tRNA。

在某些实施方式中，这些方法还包括正选择第一物种的第二细胞群，其中所述细胞含有与第一物种的细胞正交的 tRNA 库的成员、关联氨酰-tRNA 合成酶和阳性选择标记。采用正选择可选择或筛选出含有能被关联氨酰-tRNA 合成酶氨酰化而且在阳性选择标记存在下可显示所需反应的 tRNA 库成员的那些细胞，藉此提供 O-tRNA。在某些实施方式中，第二细胞群含有未被负选择去除的细胞。

还提供能用非天然氨基酸加载 O-tRNA 的正交-氨酰-tRNA 合成酶的鉴定方法。例如，该方法包括对第一物种的细胞群进行筛选，其中所述细胞各自含有：1) 多种氨酰-tRNA 合成酶(RS)的成员（例如，所述多种 RS 可包括突变型 RS、衍生自第一物种以外的物种的 RS、或者突变型 RS 和衍生自第一物种以外的物种的 RS 二者）；2) 正交-tRNA(O-tRNA) (如衍生自一个或多个物种)；和 3) 编码阳性选择标记并且包含至少一个选择者密码子的多核苷酸。

选择或筛选与不含所述多种 RS 的成员或所含成员数目较少的细胞相比细胞(如宿主细胞)中显示抑制效率提高的那些细胞。这些选择/筛选的细胞含有能氨酰化 O-tRNA 的活性 RS。用这些方法鉴定到的正交氨酰-tRNA 合成酶也是本发明的特征。

在细胞(例如真细菌细胞，如大肠杆菌细胞等，或酵母细胞)内产生在所

选位置含有非天然氨基酸的蛋白质的方法也是本发明的特征。例如，一种方法包括在适当的培养基内培养细胞，其中所述细胞含有核酸，而该核酸含有至少一个选择者密码子并编码一种蛋白，从而提供非天然氨基酸并在含所述至少一个选择者密码子的核酸的翻译过程中将该非天然氨基酸掺入该蛋白质的特定位置，从而产生该蛋白质。细胞还含有：在细胞内起作用并识别选择者密码子的正交-tRNA(O-tRNA)；以及用非天然氨基酸优先氨酰化 O-tRNA 的正交氨酰-tRNA 合成酶(O-RS)。用这种方法产生的蛋白质也是本发明的特征。

本发明还提供含有蛋白质的组合物，其中所述蛋白质包含 7-羟基香豆素-乙基甘氨酸。在某些实施方式中，蛋白质包含的氨基酸序列与已知蛋白，如治疗蛋白、诊断蛋白和工业用酶或其一部分的氨基酸序列有至少 75% 相同性。组合物任选包含药学上可接受的载体。

核酸和多肽序列及变体

如本文所述，本发明提供编码例如 O-tRNA 和 O-RS 的多核苷酸序列，多肽氨基酸序列，例如 O-RS，以及例如包含所述多核苷酸或多肽序列的组合物、系统和方法。本文披露了所述序列的例子，如 O-tRNA 和 O-RS 的氨基酸和核苷酸序列(见图 9，例如 SEQ ID NO: 4 和 5)。然而，本领域技术人员应知道本发明并不限于本文，例如实施例和序列表中披露的那些序列。本领域技术人员应知道本发明还提供具有本文所述功能的许多相关序列，例如编码本文所披露 O-RS 的保守变体的多核苷酸和多肽。

实施例 3-5 描述了能用 7-羟基香豆素-乙基甘氨酸非天然氨基酸氨酰化 O-tRNA 的正交合成酶种类(O-RS)的构建和分析。这些实施例描述能掺入非天然氨基酸 7-羟基香豆素-乙基甘氨酸的 O-RS 种类的构建和分析。

本发明提供多肽(O-RS)和多核苷酸，如 O-tRNA、编码 O-RS 或其诸部分的多核苷酸，用于分离氨酰-tRNA 合成酶克隆的寡核苷酸等。本发明的多核苷酸包括编码本发明的感兴趣蛋白或多肽并含有一个或多个选择者密码子的那些多核苷酸。另外，本发明多核苷酸包括，例如含有 SEQ ID NO:5 所示核苷酸序列的多核苷酸；与其多核苷酸序列互补或编码其多核苷酸序

列的多核苷酸。本发明的多核苷酸还包括编码含有 SEQ ID NO: 4 的 O-RS 氨基酸序列的任何多核苷酸。类似地, 在高严谨条件下, 可与上述多核苷酸在基本全长的核酸上杂交(并且不是天然多核苷酸序列)的人工核酸也是本发明多核苷酸。在一个实施方式中, 组合物包含本发明的多肽和赋形剂(如缓冲液、水、药学上可接受的赋形剂等)。本发明还提供可与本发明多肽发生特异性免疫反应的抗体或抗血清。人工多核苷酸是人造而不是天然产生的多核苷酸。

本发明的多核苷酸还包括人工多核苷酸, 即与天然 tRNA 具有如至少 75%、至少 80%、至少 90%、至少 95%、至少 98%或更高相同性(但不是天然 tRNA)。多核苷酸还包括与天然 tRNA 具有如至少 75%、至少 80%、至少 90%、至少 95%、至少 98%或更高相同性(但不是 100%相同)的人工多核苷酸。

在某些实施方式中, 载体(如质粒、粘粒、噬菌体、病毒等)含有本发明的多核苷酸。在一个实施方式中, 所述载体是表达载体。在另一个实施方式中, 所述表达载体包含与本发明的一种或多种多核苷酸操作性相连的启动子。在另一个实施方式中, 细胞包含含有本发明多核苷酸的载体。

本领域技术人员还知道本发明包括所披露序列的多种变体。例如, 产生功能相同序列的所披露序列的保守变体也属于本发明。核酸多核苷酸序列的变体也属于本发明, 所述变体与至少一种所披露序列杂交。本文披露序列的独特亚序列, 例如可通过标准序列比较技术测定的独特亚序列也属于本发明。

保守性变异

由于遗传密码的简并性, “沉默取代”(即, 核酸序列中的取代不导致所编码多肽改变)是编码氨基酸序列的每条核酸序列所隐含的特征。类似地, 也不难鉴定氨基酸序列中有一个或少数氨基酸被具有高度相似特性的不同氨基酸取代的“保守性氨基酸取代”, 因为它与披露的构建物高度相似。披露的各序列的这种保守性变异是本发明的特征之一。

具体核酸序列的“保守性变异”指编码相同或基本相同的氨基酸序列

的那些核酸，或者，如果核酸不编码氨基酸序列，则指基本上相同的序列。技术人员知道改变、加入或删除所编码序列中一个氨基酸或少部分氨基酸(一般低于 5%、更常见低于 4%、2%或 1%)的各种取代、缺失或插入是“保守性修饰变异”，这些改变导致氨基酸的缺失、插入或被化学性质相似的氨基酸所取代。因此，本发明所列举多肽序列的“保守性变异”包括用同一保守性取代组的氨基酸取代该多肽序列的少部分(通常低于 5%，更常见是低于 2%或 1%)的氨基酸。最后，加入不改变某核酸分子所编码活性的序列(例如加入无功能序列)是基础核酸的保守性变异。

本领域熟知提供功能类似的氨基酸的保守性取代表，其中一个氨基酸残基被另一个具有相似化学特性(例如芳族侧链或带正电荷的侧链)的氨基酸残基取代，因此基本上不改变该多肽分子的功能特性。以下列举多组化学特性相似的天然氨基酸，其中，各组内的取代即“保守性取代”。

保守性氨基酸取代

非极性和/或脂族侧链	极性,不带电荷的侧链	芳族侧链	带正电荷的侧链	带负电荷的侧链
甘氨酸 丙氨酸 缬氨酸 亮氨酸 异亮氨酸 脯氨酸	丝氨酸 苏氨酸 半胱氨酸 甲硫氨酸 精氨酸 谷氨酰胺	苯丙氨酸 酪氨酸 色氨酸	赖氨酸 精氨酸 组氨酸	天冬氨酸 谷氨酸

核酸杂交

可采用比较杂交来鉴定本发明的核酸，包括本发明核酸的保守性变体，该比较杂交方法是区别本发明核酸的优选方法。此外，在高度、超高度和超超高度严谨性条件下能与 SEQ ID NO:5 所示核酸杂交的靶核酸是本发明的特征。这种核酸的例子包括与某给定的核酸序列相比，含有一个或少许沉默或保守性核酸取代的核酸。

当测试核酸与探针的杂交程度是与完美匹配的互补靶标杂交程度的至少 50%，即信噪比至少是该探针与靶标在完美匹配的探针与完美匹配的互

补靶标结合的条件下杂交的信噪比的一半高时，可称该测试核酸与探针核酸特异性杂交；此条件下，完美匹配的探针与完美匹配的互补靶核酸结合的信噪比至少是与任何不匹配靶核酸杂交所观察到的信噪比的约 5-10 倍。

当核酸结合(一般在溶液中)时，称其“杂交”。核酸因各种已充分表征的理化力，例如氢键、溶剂排斥、碱基堆积等而杂交。核酸杂交的广泛指南见 Tijssen, (1993), *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*(生物化学与分子生物学的实验室技术——核酸探针杂交), 第一部分第二章, “Overview of principles of hybridization and the strategy of nucleic acid probe assays(杂交原理概述与核酸探针试验方法)”, (艾尔赛弗公司(Elsevier), 纽约); 以及 *Current Protocols in Molecular Biology*(最新分子生物学方法), Ausubel 等编, “Current Protocols”(最新方法), 格林出版合伙公司(Greene Publishing Associates, Inc.) 和约翰威立父子公司(John Wiley & Sons)的合资企业, (2006 年增补); Hames 和 Higgins, (1995), *Gene Probes 1*(基因探针 1)和 *Gene Probes 2*(基因探针 2), IRL 出版社, 牛津大学出版社, 牛津, 英格兰, 它们提供了合成、标记、检测和定量测定 DNA 及 RNA, 包括寡核苷酸的细节。

在 Southern 或 northern 印迹滤膜上具有 100 个以上互补残基的互补核酸杂交的严谨性杂交条件的例子是: 含 1 mg 肝素的 50%福尔马林, 42℃ 杂交过夜。严谨性洗涤条件的例子是 65℃, 用 $0.2 \times \text{SSC}$ 洗涤 15 分钟 (SSC 缓冲液的描述可参见 Sambrook 等; Sambrook 等, *Molecular Cloning - A Laboratory Manual*(分子克隆-实验室手册), (第三版), 第 1-3 卷, 冷泉港实验室, 冷泉港, 纽约, 2001)。通常先进行低严谨性洗涤除去背景探针信号, 再进行高严谨性洗涤。示例性低严谨性洗涤是 40℃, 用 $2 \times \text{SSC}$ 洗涤 15 分钟。具体杂交试验中信噪比是无关探针观察到的信噪比的 5 倍(或更高)通常表示检测到特异性杂交。

核酸杂交实验(例如, Southern 和 northern 杂交)的“严谨性杂交洗涤条件”取决于序列, 并随不同的环境参数而不同。核酸杂交的广泛指南见 Tijssen, (1993), *生物化学与分子生物学的实验室技术——核酸探针杂交*, 第一部分第二章, “杂交原理概述与核酸探针试验方法”, (艾尔赛弗公司,

纽约); Hames 和 Higgins, (1995), 基因探针 1, IRL 出版社, 牛津大学出版社, 牛津, 英格兰, 和 Hames 和 Higgins, (1995), 基因探针 2, IRL 出版社, 牛津大学出版社, 牛津, 英格兰。不难凭经验确定适合任何测试核酸的严谨性杂交和洗涤条件。例如, 在确定严谨性杂交和洗涤条件时, 可逐渐提升杂交和洗涤条件(例如, 通过升高杂交或洗涤的温度、降低杂交或洗涤中的盐浓度、增加杂交或洗涤中的洗涤剂浓度和/或增加杂交或洗涤中的有机溶剂如福尔马林的浓度)直至符合选定的标准。例如, 在高度严谨性杂交和洗涤条件中, 逐渐提升杂交和洗涤条件直至探针与完美匹配的互补靶标结合的信噪比至少是该探针与不匹配靶标杂交所观察到的信噪比的 5 倍。

选择的“非常严谨”的条件等于具体探针的热解链温度(T_m)。 T_m 是 50% 的测试序列与完美匹配的探针杂交时的温度(在规定的离子强度和 pH 下)。为本发明的目的, “高度严谨”的杂交和洗涤条件一般选择为在规定的离子强度和 pH 下比具体序列的 T_m 约低 5℃。

“超高严谨”的杂交和洗涤条件指, 增加杂交和洗涤条件的严谨性直至探针与完美匹配的互补靶核酸结合的信噪比至少是任何不匹配靶核酸杂交时观察到的信噪比的 10 倍。靶核酸在这种条件下与探针杂交的信噪比至少是完美匹配的互补靶核酸的 1/2 时, 可称其在超高严谨性条件下与探针结合。

类似地, 可通过逐渐提升相关杂交试验的杂交和/或洗涤条件来确定甚至更高的严谨性水平。例如, 提升杂交和洗涤的条件严谨性直至探针与完美匹配的互补靶核酸结合的信噪比至少是任何不匹配靶核酸杂交信噪比的 10、20、50、100 或 500 倍或更高。靶核酸在这种条件下与探针杂交的信噪比至少是完美匹配的互补靶核酸的 1/2 时, 可称其在超超高度严谨性条件下与探针结合。

如果在严谨性条件下彼此不杂交的核酸所编码的多肽基本相同, 则这些核酸仍基本相同。例如, 当利用遗传密码所允许的最大密码简并性产生核酸的拷贝时会发生这种情况。

独特亚序列

在一些方面,本发明提供含有选自本文披露的 O-tRNA 和 O-RS 序列的核酸中独特亚序列的核酸。与对应于任何已知 O-tRNA 或 O-RS 核酸序列的核酸相比,所述独特亚序列是独特的。可采用,例如设置默认参数的 BLAST 进行核酸比对。任何独特亚序列可用作,例如探针来鉴定本发明核酸或相关核酸。

类似地,本发明包括含有选自本文披露的 O-RS 序列的多肽中独特亚序列的多肽。与对应于任何已知多肽序列的多肽相比,本文的独特亚序列是独特的。

本发明还提供能在严谨性条件下与独特的编码寡核苷酸杂交的靶核酸,所述寡核苷酸编码选自 O-RS 序列的多肽中的独特亚序列,其中与对应于任何对照多肽(例如,通过突变获得本发明合成酶的亲代序列)的多肽相比,所述独特亚序列是独特的。可如上所述测定独特序列。

序列比较、相同性和同源性

对于两个或多个核酸或多肽序列,术语“相同的”或“相同性百分比”指当用下文所述序列比较算法(或技术人员可用的其它算法)或通过目测观察来比较和比对最大相应性时,这两个或多个序列或亚序列相同,或者有特定百分比的氨基酸残基或核苷酸相同。

对于两个核酸或多肽(例如,编码 O-tRNA 或 O-RS 的 DNA,或 O-RS 的氨基酸序列),术语“基本相同”指当利用序列比较算法或通过目测观察来比较和比对最大相应性时,两个或多个序列或亚序列至少有约 60%、约 80%、约 90-95%、约 98%、约 99%或更多的核苷酸或氨基酸残基相同。这种“基本相同”的序列通常认为是“同源的”,而不论实际祖先。优选在至少长约 50 个残基的序列区域,更优选在至少约 100 个残基的区域存在“基本相同性”,待比较两条序列最好在至少约 150 个残基或者在全长上基本相同。

当蛋白质和/或蛋白质序列(天然或人工地)衍生自同一祖先蛋白或蛋白序列时,它们是“同源的”。类似地,当核酸和/或核酸序列(天然或人工地)衍生自同一祖先核酸或核酸序列时,它们是“同源的”。例如,可通过任

何可用的诱变方法来修饰任何天然核酸使之含有一个或多个选择者密码子。当该诱变的核酸表达时，它编码的多肽含有一个或多个非天然氨基酸。当然，该突变方法还可改变一个或多个标准密码子，从而也改变了所得突变型蛋白质中的一个或多个标准氨基酸。通常可从两种或多种核酸或蛋白质(或其序列)间的序列相似性推断同源性。可用于确认同源性的序列间精确的相似性百分比因所研究的核酸与蛋白质而有所不同，但常规利用低至25%的序列相似性来确认同源性。还可利用更高水平的序列相似性，例如30%、40%、50%、60%、70%、80%、90%、95%或99%或更高来确认同源性。本文描述了测定序列相似性百分比的方法(例如，采用参数默认的BLASTP和BLASTN)，这些方法众所周知。

对于序列比较和同源性测定，通常将一条序列用作参比序列与测试序列相比较。当采用序列比较算法时，将测试序列与参比序列输入计算机，如果需要可指定亚序列坐标，并指定序列算法程序参数。然后根据所指定的程序参数，序列比较算法可计算测试序列相比于参比序列的序列相同性百分比。

可用以下方法进行比较序列的最佳比对，例如Smith和Waterman的局部同源性算法，*Adv. Appl. Math.*, 2: 482, (1981); Needleman和Wunsch的同源性比对算法，*J. Mol. Biol.*, 48: 443, (1970); Pearson和Lipman的相似性检索方法，*Proc. Nat'l. Acad. Sci. USA*, 85: 2444, (1988); 计算机执行这些算法(威斯康星遗传软件包中的GAP, BESTFIT, FASTA和TFASTA, 遗传学计算机组(Genetics Computer Group), 575 Science Dr., 麦迪逊, 威斯康星州); 或者目测观察(一般可参见*Current Protocols in Molecular Biology*(最新分子生物学方法), Ausubel等编, “*Current Protocols*(最新方法)”, 格林出版合伙公司和约翰威立父子公司的合资企业, 2006年增补)。

适用于测定序列相同性和序列相似性百分比的算法的一个例子是Altschul等, *J. Mol. Biol.*, 215: 403-410, (1990)所述的BLAST算法。进行BLAST分析的软件由国家生物技术信息中心对公众开放(参见NCBI网址)。该算法包括: 首先通过在查询序列中鉴定长为W的短字符串来鉴定高评分序列配对(HSP), 这些字符串与数据库序列中相同长度的字符串比对时符合或满足

某些正值阈值评分 T 。 T 称为邻近字串评分阈值(Altschul 等, (1990), *J. Mol. Biol.*, 215: 403-410)。这些原始邻近字串选中(hit)作为启动检索的种子来找寻含有它们的较长 HSP。然后使这些字串选中沿着各条序列双向延伸, 以致累积比对评分增加。对于核苷酸序列, 用参数 M (一对匹配残基的奖励评分; 恒大于 0)和 N (错配残基的罚分; 恒小于 0)计算累积评分。对于氨基酸序列, 则用评分矩阵来计算累积评分。当出现以下情况时终止各方向的字串选中延伸: 累积比对评分从其达到的最高值下降 X ; 因一个或多个负评分残基比对的累积导致累积评分降至 0 或 0 以下; 或者到达各序列的末端。BLAST 算法参数 W 、 T 和 X 决定比对的灵敏度和速度。BLAST 程序(用于核苷酸序列)默认 11 为字长(W)、10 为期望值(E)、100 为截断值、 $M=5$ 、 $N=-4$, 并进行双链比较。对于氨基酸序列, BLAST 程序默认 3 为字长(W)、10 为期望值(E)并采用 BLOSUM62 评分矩阵(参见 Henikoff 和 Henikoff, (1989), *Proc. Natl. Acad. Sci. USA*, 89: 10915)。

除了计算序列相同性百分比外, BLAST 算法也可对两条序列间的相似性进行统计学分析(参见, 例如 Karlin 和 Altschul, *Proc. Nat'l. Acad. Sci. USA*, 90: 5873-5787, (1993))。BLAST 算法提供的相似性量度之一是最小总概率($P(N)$), 其指示两条核苷酸或氨基酸序列之间发生随机匹配的概率。例如, 如果在测试核酸与参比核酸的比较中最小总概率小于约 0.1, 更优选小于约 0.01, 最优选小于约 0.001, 则可认为该核酸与参比核酸相似。

诱变与其它分子生物学技术

可采用分子生物学技术操作本发明的和用于本发明的多核苷酸和多肽。描述分子生物学的通用教材包括 Berger 和 Kimmel, *Guide to Molecular Cloning Techniques*(分子克隆技术指南), *Methods in Enzymology*(酶学方法), 第 152 卷, 学术出版社公司(Academic Press, Inc.), 圣迭戈, 加利福尼亚州; Sambrook 等, *Molecular Cloning - A Laboratory Manual*(分子克隆-实验室手册), (第三版), 第 1-3 卷, 冷泉港实验室, 冷泉港, 纽约, 2001 和 *Current Protocols in Molecular Biology*(最新分子生物学方法), Ausubel 等编; “*Current Protocols in Molecular Biology*(分子生物学最新方法)”, 格林出

版合伙公司和约翰威立父子公司的合资企业, (2006 年增补)。这些教材描述了诱变、载体的应用、启动子和涉及, 例如产生含有选择者密码子的基因从而产生含非天然氨基酸的蛋白质, 正交 tRNA、正交合成酶及其配对的许多其它相关课题。

本发明可采用各种类型的诱变, 例如以使 tRNA 分子突变、产生 tRNA 文库、产生合成酶文库、插入编码感兴趣蛋白质或多肽中的非天然氨基酸的选择者密码子。它们包括但不限于: 定点诱变、随机点诱变、同源重组、DNA 改组或其它递归诱变方法、嵌合构建、利用含尿嘧啶的模板诱变、寡核苷酸指导的诱变、硫代磷酸酯修饰的 DNA 诱变、利用缺口双螺旋 DNA 的诱变等, 或者是它们的任何组合。其它合适的方法包括点错配修复、利用修复缺陷宿主株的诱变、限制性选择和限制性纯化、缺失诱变、总基因合成诱变、双链断裂修复等。本发明也包括, 例如涉及嵌合构建物的诱变。在一个实施方式中, 可用天然分子或经改变或突变的天然分子的已知信息, 例如序列, 序列比较、物理特性、晶体结构等来指导诱变。

可利用本发明多核苷酸或含有本发明多核苷酸的构建物, 例如本发明的载体(例如可以是克隆载体或表达载体)来遗传改造(例如转化、转导或转染)宿主细胞。例如, 可将正交 tRNA、正交 tRNA 合成酶和待衍生蛋白质的编码区操作性连接于可在所需宿主细胞中起作用的基因表达控制元件。典型的载体含有转录和翻译终止子、转录和翻译起始序列和用于调节具体靶核酸表达的启动子。这些载体可任选地包含遗传表达盒, 所述表达盒含有至少一个独立的终止子序列, 允许该表达盒在真核细胞或原核细胞或二者中复制的序列(例如, 穿梭载体), 和适用于原核与真核系统的选择标记。载体适用于在原核细胞、真核细胞或优选二者中复制和/或整合。参见 Gilman 和 Smith, *Gene*, 8: 81, (1979); Roberts 等, *Nature*, 328: 731, (1987); Schneider 等, *Protein Expr. Purif.*, 6435: 10, (1995); Berger 和 Kimmel, 分子克隆技术指南, 酶学方法, 第 152 卷, 学术出版社公司, 圣迭戈, 加利福尼亚州; Sambrook 等, 分子克隆-实验室手册, (第三版), 第 1-3 卷, 冷泉港实验室, 冷泉港, 纽约, 2001; 和最新分子生物学方法, Ausubel 等编; 最新方法, 格林出版合伙公司和约翰威立父子公司的合资企业, (2006

年增补)。例如,载体可以是质粒、细菌、病毒、裸多核苷酸或缀合的多核苷酸形式。可通过标准方法将载体导入细胞和/或微生物,所述方法包括电穿孔(From 等, *Proc. Natl. Acad. Sci. USA*, 82, 5824, (1985)), 病毒载体感染, 利用在小珠或颗粒的基质内或其表面上含有核酸的小颗粒的高速弹丸穿透(Klein 等, *Nature*, 327: 70-73, (1987))等。

开发了能在大肠杆菌中对琥珀终止密码子(UAG)起反应而将非天然氨基酸位点特异性掺入蛋白质的高效且通用的单质粒系统。在该新系统中, 詹氏甲烷球菌抑制型 tRNA^{tyr}(CUA)和酪氨酰-tRNA 合成酶配对由与大多数大肠杆菌表达载体相容的单个质粒编码。构建在 proK 启动子和终止子控制下的单顺反子 tRNA 操纵子, 用于优化二级结构和 tRNA 加工。引入突变形式的合成酶 glnS 启动子显著提高了抑制效率和保真度。利用多拷贝的 tRNA 基因以及该合成酶上的特定突变(D286R)也可提高抑制效率 (Kobayashi 等, (2003), “Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion(用于遗传密码扩展的酪氨酰-tRNA 合成酶的正交 tRNA 特异性的结构基础)”, *Nat. Struct. Biol.*, 10(6): 425-432)。几种不同非天然氨基酸的掺入高效且精确也证明该优化系统的通用性, 所述非天然氨基酸在研究蛋白质功能和结构中的独特用途已得到证实。

ATCC 提供了可用于克隆的细菌和细菌噬菌体目录, 例如 ATCC 出版的 *The ATCC Catalogue of Bacteria and Bacteriophage*(ATCC 细菌和细菌噬菌体目录), (1996), Gherna 等编。测序、克隆的其它基本方法和分子生物学的其它方面及基础理论思考也可参见 Sambrook 等, 分子克隆-实验室手册, (第三版), 第 1-3 卷, 冷泉港实验室, 冷泉港, 纽约, 2001; 最新分子生物学方法, Ausubel 等编; 最新方法, 格林出版合伙公司和约翰威立父子公司的合资企业, (2006 年增补); 和 Watson 等, (1992), *Recombinant DNA*(重组 DNA), 第二版, 科学美国书籍公司(Scientific American Books), 纽约。此外, 可以向各种商业来源定制或标准订购基本上任何核酸(和实际上任何标记的核酸, 无论标准或非标准的), 例如得克萨斯州米德兰市的米德兰检定试剂公司(Midland Certified Reagent Company)、加利福尼亚州雷蒙纳市的大美国基因公司(The Great American Gene Company, Ramona, CA)、伊利

诺斯州芝加哥市的表达基因公司(ExpressGen Inc.)、加利福尼亚州阿拉米达市的操纵子技术公司(Operon Technologies Inc.)和许多其它公司。

可利用为例如筛选步骤、激活启动子或选择转化子的活性而作适当改进的常规营养培养基培养工程改造的宿主细胞。这些细胞任选培养成转基因生物。其它可用的参考文献,例如细胞分离和培养(例如,用于随后核酸分离)的文献包括: Freshney, (1994), *Culture of Animal Cells, a Manual of Basic Technique*(动物细胞培养,基本技术手册),第三版,威立利斯公司(Wiley-Liss),纽约及其所引用的参考文献; Payne 等, (1992), *Plant Cell and Tissue Culture in Liquid Systems*(液体系统中的植物细胞与组织培养),约翰威立父子公司,纽约,纽约州; Gamborg 和 Phillips 编, (1995), *Plant Cell, Tissue and Organ Culture*(植物细胞、组织和器官培养); *Fundamental Methods*(基本方法), Springer Lab Manual(斯普林格实验室手册), S-V 公司(Springer-Verlag) (伯林海德尔堡 纽约); Atlas 和 Parks 编, *The Handbook of Microbiological Media*(微生物培养基手册), (1993), CRC 出版社(CRC Press), 伯克莱屯, 佛罗里达州。

感兴趣的蛋白质与多肽

在细胞中产生在特定位置含非天然氨基酸的蛋白质的方法也是本发明特征之一。例如,一种方法包括用适当的培养基培养细胞,其中所述细胞包含含有至少一个选择者密码子并编码蛋白质的核酸;和提供非天然氨基酸;其中所述细胞还包含:在细胞内具有功能并识别选择者密码子的正交-tRNA(O-tRNA);和用非天然氨基酸优先氨酰化 O-tRNA 的正交氨酰-tRNA 合成酶(O-RS)。该方法所产生的蛋白质也是本发明特征之一。

在某些实施方式中,与表达系统中任何内源性 tRNA 相比, O-RS 偏向于氨酰化关联的 O-tRNA。当 O-tRNA 和 O-RS 以等摩尔浓度存在时, O-RS 所加载的 O-tRNA 与内源性 tRNA 的相对比例大于 1:1, 优选至少约 2:1, 更优选 5:1, 更优选 10:1, 更优选 20:1, 更优选 50:1, 更优选 75:1, 更优选 95:1, 98:1, 99:1, 100:1, 500:1, 1,000:1, 5,000:1 或更高。

本发明还提供包含蛋白质的组合物,所述蛋白质含有非天然氨基酸。在某

些实施方式中,所述蛋白质含有的氨基酸序列与治疗性蛋白、诊断性蛋白、工业用酶或其部分的序列至少有 75%相同。

本发明的组合物和本发明方法制备的组合物任选处于细胞中。然后将本发明的 O-tRNA/O-RS 配对或各组分用于宿主系统的翻译机制中,从而将非天然氨基酸掺入蛋白质。2004 年 4 月 16 日提交的,名为“EXPANDING THE EUKARYOTIC GENETIC CODE(扩展真核遗传密码)”的国际公布号 WO 2004/094593 和名为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS(非天然氨基酸的体内掺入)”的国际公布号 WO 2002/085923 描述了该方法,这两篇文献以引用的方式纳入本文。例如,当 O-tRNA/O-RS 配对引入宿主,例如大肠杆菌细胞时,该配对对选择者密码子起反应而将非天然氨基酸,例如 7-羟基香豆素-乙基甘氨酸在体内掺入蛋白质。加入系统的非天然氨基酸可以是合成氨基酸,例如可以外源性加入生长培养基的苯丙氨酸或酪氨酸衍生物。本发明的组合物任选为体外翻译系统,或体内系统。

本发明的细胞能大量合成包含非天然氨基酸的蛋白质。在一些方面,组合物任选包含,例如至少 10 微克、至少 50 微克、至少 75 微克、至少 100 微克、至少 200 微克、至少 250 微克、至少 500 微克、至少 1 毫克、至少 10 毫克或更多含非天然氨基酸的蛋白质,或者体内蛋白质产生方法所能达到的含量(本文提供重组蛋白质制备和纯化的细节)。在另一方面,组合物中包含在例如细胞裂解液、缓冲液、药学缓冲液或其它液体混悬液中(例如,体积为约 1nL-约 100L)的蛋白质浓度任选是,例如每升至少 10 微克蛋白质、每升至少 50 微克蛋白质、每升至少 75 微克蛋白质、每升至少 100 微克蛋白质、每升至少 200 微克蛋白质、每升至少 250 微克蛋白质、每升至少 500 微克蛋白质、每升至少 1 毫克蛋白质或每升至少 10 毫克蛋白质或更高。在细胞中制备大量(例如,大于采用其它方法如体外翻译通常可能得到的量)包含至少一个非天然氨基酸的蛋白质是本发明特征之一。

可掺入非天然氨基酸来,例如改变蛋白质结构和/或功能,如改变大小、酸性、亲核性、氢键、疏水性、蛋白酶靶位的易接近性、对某部分的靶向(如用于蛋白阵列)、掺入标记物或反应基团等。含有非天然氨基酸的蛋白质的

催化或物理特性可得到改善,甚至获得全新的催化或物理特性。例如,可通过将非天然氨基酸掺入蛋白质来任选改进以下特性:毒性、生物分布、结构特性、光谱特性、化学和/或光化学特性、催化能力、半衰期(例如血清半衰期)、与其它分子的反应能力(如共价或非共价的)等。包含含有至少一个非天然氨基酸的蛋白质的组合物可用作,例如新型治疗剂、诊断剂、催化酶、工业用酶、结合蛋白(如抗体)以及用于例如研究蛋白质结构和功能。参见例如 Dougherty, (2000), “*Unnatural Amino Acids as Probes of Protein Structure and Function*(作为蛋白质结构和功能探针的非天然氨基酸”), *Current Opinion in Chemical Biology*, 4: 645-652。

在本发明的一些方面,组合物包含至少一种含有至少一个,例如至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个或至少十个或更多非天然氨基酸的蛋白质。所述非天然氨基酸可以相同或不同,例如蛋白质中可以在 1、2、3、4、5、6、7、8、9 或 10 个或更多不同的位点上含有 1、2、3、4、5、6、7、8、9 或 10 个或更多不同的非天然氨基酸。在另一方面,组合物包含蛋白质,该蛋白质中存在的至少一个(但不是全部)具体氨基酸是非天然氨基酸。对于给定的含多个非天然氨基酸的蛋白质,所述非天然氨基酸可以相同或不同(如蛋白质可包含两个或多个不同类型的非天然氨基酸,或者可包含两个相同的非天然氨基酸)。对于含两个以上非天然氨基酸的给定蛋白质,所述非天然氨基酸可以相同,不同,或是多个同一类型的非天然氨基酸与至少一个不同的非天然氨基酸的组合。

利用本文组合物和方法可制备基本上任何含非天然氨基酸的蛋白质(或其一部分)(与任何相应的编码核酸,例如含有一个或更多选择者密码子的核酸)。未试图鉴定成百上千的已知蛋白质,可修饰它们中的任一种使之含有一个或多个非天然氨基酸,例如通过改进任何可用的突变方法从而使相关翻译系统中包含一个或多个合适的选择者密码子。已知蛋白质的共有序列库包括 GenBank、EMBL、DDBJ 和 NCBI。通过检索互联网不难鉴定其它库。

这些蛋白质通常与任何可用的蛋白质(例如,治疗性蛋白、诊断性蛋白、工业用酶或其一部分等)有,例如至少 60%、至少 70%、至少 75%、至少 80%、至少 90%、至少 95%或至少 99%或更高的相同性,同时这些蛋白质含有一个

或多个非天然氨基酸。可经修饰包含一个或多个非天然氨基酸的治疗性、诊断性和其它蛋白质的例子参见但不限于 2004 年 4 月 16 日提交的名为“Expanding the Eukaryotic Genetic Code(扩展真核生物遗传密码)”的国际公布号 WO 2004/094593 和名为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS.(非天然氨基酸的体内掺入)”的 WO 2002/085923。可经修饰包含一个或多个非天然氨基酸的治疗性、诊断性和其它蛋白质的例子包括但不限于：例如 α -1 抗胰蛋白酶、血管生成抑制素、抗溶血因子(Antihemolytic factor)、抗体(抗体的其它细节见下文)、载脂蛋白、脱辅蛋白、心房利钠因子、心房利钠多肽、心房肽(Atrial peptides)、C-X-C 趋化因子(如 T39765、NAP-2、ENA-78、Gro-a、Gro-b、Gro-c、IP-10、GCP-2、NAP-4、SDF-1、PF4、MIG)、降钙素、CC 趋化因子(如单核细胞趋化蛋白-1、单核细胞趋化蛋白-2、单核细胞趋化蛋白-3、单核细胞炎性蛋白-1 α 、单核细胞炎性蛋白-1 β 、RANTES、I309、R83915、R91733、HCC1、T58847、D31065、T64262)、CD40 配体、C-kit 配体、胶原、集落刺激因子(CSF)、补体因子 5a、补体抑制剂、补体受体 1、细胞因子(如，上皮嗜中性活化肽-78(epithelial Neutrophil Activating Peptide-78)、GRO α /MGSA、GRO β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1)、表皮生长因子(EGF)、促红细胞生成素(“EPO”)、剥落性毒素 A 和 B、因子 IX、因子 VII、因子 VIII、因子 X、成纤维细胞生长因子(FGF)、血纤蛋白原、纤连蛋白、G-CSF、GM-CSF、葡萄糖脑苷脂酶、促性腺激素、生长因子、Hedgehog 蛋白(如 Sonic, Indian, Desert)、血红蛋白、肝细胞生长因子(HGF)、水蛭素、人血清白蛋白、胰岛素、胰岛素样生长因子(IGF)、干扰素(如，IFN- α 、IFN- β 、IFN- γ)、白介素(如，IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12 等)、角质形成细胞生长因子(KGF)、乳铁蛋白、白血病抑制因子、萤光素酶、神经营养因子(Neurturin)、嗜中性白细胞抑制因子(NIF)、抑瘤蛋白 M、成骨蛋白、甲状腺激素、PD-ECSF、PDGF、肽激素(如，人生长激素)、多效营养因子、A 蛋白、G 蛋白、热源性外毒素 A、B 和 C、松弛素、肾素、SCF、可溶性补体受体 1、可溶性 I-CAM 1、可溶性白介素受体(IL-1、2、3、4、5、6、7、9、10、11、12、13、14、15)、可溶性 TNF 受体、生

长调节素、促生长素抑制剂、促生长素、链激酶、超抗原，即葡萄球菌肠毒素(SEA、SEB、SEC1、SEC2、SEC3、SED、SEE)、超氧化物歧化酶(SOD)、中毒性休克综合征毒素(TSST-1)、胸腺素 α 1、组织纤溶酶原激活物、肿瘤坏死因子 β (TNF β)、肿瘤坏死因子受体(TNFR)、肿瘤坏死因子- α (TNF α)、血管内皮生长因子(VEGEF)、尿激酶和许多其它蛋白。

可采用本文所述能在体内掺入非天然氨基酸的组合物和方法制备的一类蛋白质包括转录调节剂或其一部分。示例性转录调节剂包括调节细胞生长、分化、调控等的基因和转录调节蛋白。转录调节剂存在于原核生物、病毒和真核生物(包括真菌、植物、酵母菌)、昆虫和动物(包括哺乳动物)中，从而提供了广泛的治疗靶点。应该理解表达和转录激活物通过许多机理调控转录，例如通过与受体结合、刺激信号转导级联反应、调控转录因子表达、与启动子和增强子结合、与结合启动子和增强子的蛋白质结合、DNA 解链、前-mRNA 剪接、RNA 聚腺苷酸化和 RNA 降解。

本发明的一类蛋白质(例如，含有一个或多个非天然氨基酸的蛋白质)包括生物学活性蛋白，例如细胞因子，炎性分子，生长因子，它们的受体和癌基因产物，如白介素(如 IL-1、IL-2、IL-8 等)、干扰素、FGF、IGF-I、IGF-II、FGF、PDGF、TNF、TGF- α 、TGF- β 、EGF、KGF、SCF/c-Kit、CD40L/CD40、VLA-4/VCAM-1、ICAM-1/LFA-1 和透明素(hyalurin)/CD44；信号转导分子和相应的癌基因产物，例如 Mos、Ras、Raf 和 Met；转录激活物和阻遏物，例如 p53、Tat、Fos、Myc、Jun、Myb、Rel 和类固醇激素受体，例如雌激素、孕酮、睾酮、醛甾酮的那些受体，LDL 受体配体和皮质酮。

本发明也提供含有至少一个非天然氨基酸的酶(例如工业用酶)或其一部分。酶的例子包括但不限于：酰胺酶、氨基酸消旋酶、酰基转移酶、脱卤酶、双加氧酶、二芳基丙烷过氧化物酶(diarylpropane peroxidases)、差向异构酶、环氧化物水解酶、酯酶、异构酶、激酶、葡糖异构酶、糖苷酶、糖基转移酶、卤素过氧化物酶(haloperoxidases)、单加氧酶(例如 p450)、脂酶、木质素过氧化物酶、胍水合酶、胍水解酶、蛋白酶、磷酸酶、枯草杆菌蛋白酶、转氨酶和核酶。

许多这些蛋白质可商品化购得(参见，例如西格玛生物科学公司(Sigma

BioSciences)), 相应的蛋白质序列和基因及其许多变体通常是熟知的(参见, 例如 Genbank)。可根据本发明通过插入一个或多个非天然氨基酸来修饰任何这些蛋白, 从而(例如)改变这些蛋白质一种或多种感兴趣的、治疗、诊断或酶活性。治疗相关特性包括血清半衰期、保存半衰期、稳定性、免疫原性、治疗活性、可检测性(例如, 通过在非天然氨基酸中加入报道基团(如标记物或标记结合位点))、LD₅₀ 或其他负效应的降低、通过胃肠道进入体内的能力(如口服利用度)等。诊断特性的例子包括保存半衰期、稳定性、诊断活性、可检测性等。相关的酶特性的例子包括保存半衰期、稳定性、酶活性、生产能力等。

也可采用本发明方法和组合物修饰各种其它蛋白质使之包含一个或多个非天然氨基酸。例如, 本发明可包括用非天然氨基酸取代一种或多种疫苗蛋白质中的一个或多个天然氨基酸, 例如以下来源的蛋白质: 感染性真菌, 如曲霉(*Aspergillus*)、假丝酵母(*Candida*)种; 细菌, 特别是用作病原性细菌模型的大肠杆菌, 以及医学上重要的细菌, 如葡萄球菌属(*Staphylococci*)(如, 金黄色葡萄球菌)或链球菌属(*Streptococci*)(如, 肺炎链球菌); 原生动物, 如孢子纲(如, 疟原虫(*Plasmodia*))、根足虫(*rhizopods*)(如, 内变形虫(*Entamoeba*))和鞭毛虫类(锥虫(*Trypanosoma*)、利什曼原虫(*Leishmania*)、毛滴虫(*Trichomonas*)、贾第虫(*Giardia*)等); 病毒, 如(+)RNA 病毒(例子包括痘病毒, 如痘苗病毒(*vaccinia*); 小 RNA 病毒, 如脊髓灰质炎病毒(*polio*); 披膜病毒, 如风疹病毒(*rubella*); 黄病毒(*Flaviviruses*), 如 HCV; 和冠状病毒)、(-)RNA 病毒(如棒状病毒(*Rhabdovirus*), 如 VSV; 副粘病毒, 如 RSV; 正粘病毒, 如流感病毒; 布尼亚病毒和嵌沙样病毒)、dsDNA 病毒(如呼肠孤病毒)、RNA 到 DNA 的病毒, 即逆转录病毒, 如 HIV 和 HTLV 和某些 DNA 到 RNA 的病毒, 如乙肝病毒。

农业相关的蛋白也是非天然氨基酸修饰的合适靶位, 例如抗虫蛋白(如 Cry 蛋白)、淀粉和脂质产生酶、植物和昆虫毒素、毒素耐受性蛋白、真菌毒素解毒蛋白、植物生长酶(如核酮糖 1,5-二磷酸羧化酶/加氧酶, “RUBISCO”)、脂氧合酶(LOX)和磷酸烯醇式丙酮酸(PEP)羧化酶。

在某些实施方式中, 本发明方法和/或组合物中的感兴趣多肽或蛋白质(或

其一部分)由核酸编码。所述核酸通常含有至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、十个或更多个选择者密码子。

可采用本领域技术人员熟知和本文在“诱变与其它分子生物学技术”中所述的方法诱变编码感兴趣蛋白质或多肽的基因，从而使之含有(例如)一个或多个选择者密码子来掺入非天然氨基酸。例如，可诱变感兴趣蛋白质的核酸，使之含有一个或多个选择者密码子以插入一个或多个非天然氨基酸。本发明包括任何蛋白质的这种变体(例如突变体)形式，例如含有至少一个非天然氨基酸。类似地，本发明也包括相应的核酸，即具有编码一个或多个非天然氨基酸的一个或多个选择者密码子的任何核酸。

为制备含有非天然氨基酸的蛋白质，可利用适合通过正交 tRNA/RS 对在体内掺入非天然氨基酸的宿主细胞和生物。可用表达正交 tRNA、正交 tRNA 合成酶的一个或多个载体和编码待衍生蛋白质的载体来遗传改造(例如转化、转导或转染)宿主细胞。各组分可以位于同一载体或各自位于不同载体，或者可以两个组分位于一个载体而第三组分位于第二载体。载体可以是例如质粒、细菌、病毒、裸多核苷酸或偶联多核苷酸的形式。

通过免疫反应性测定多肽

因为本发明多肽提供了各种新的多肽序列(例如，以在本文翻译系统中合成的蛋白质为例，多肽含有非天然氨基酸；或者，以新的合成酶为例，则是标准氨基酸的新序列)，这些多肽也提供了可在例如免疫测定中识别的新结构特征。产生能与本发明多肽特异性结合的抗血清以及能与这种血清结合的多肽是本发明的特征之一。本文所用的术语“抗体”包括但不限于：基本上由一个或多个免疫球蛋白基因编码的多肽，或其能特异性结合并识别分析物(抗原)的片段。例子包括多克隆、单克隆、嵌合型和单链抗体等。本文所用的术语“抗体”也包括免疫球蛋白片段，包括 Fab 片段和由表达文库(包括噬菌体展示文库)产生的片段。抗体的结构与术语可参见，例如 Paul, *Fundamental Immunology*(基础免疫学)，第四版，1999，雷文出版社(Raven Press)，纽约。

为产生用于免疫测定的抗血清，可如本文所述产生并纯化一种或多种免疫

原性多肽。例如，可在重组细胞中产生重组蛋白。采用标准小鼠免疫方案(关于用于测定特异性免疫反应性的抗体产生、免疫测定形式和条件的标准说明可参见，例如 Harlow 和 Lane, (1988), *Antibodies, A Laboratory Manual*(抗体，实验室手册)，冷泉港出版社，纽约)，用免疫原性蛋白和标准佐剂(如弗氏佐剂)免疫小鼠的近交品系(此类小鼠因其实际遗传相同性使得实验结果重现性较高而用于测定)。蛋白质、抗体、抗血清等的其它细节可见名为“EXPANDING THE EUKARYOTIC GENETIC CODE(扩展真核生物遗传密码)”的国际公布号 WO 2004/094593；名为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS(非天然氨基酸的体内掺入)”的 WO 2002/085923；名为“GLYCOPROTEIN SYNTHESIS(糖蛋白合成)”的 WO 2004/035605 和名为“PROTEIN ARRAYS(蛋白质阵列)”的 WO 2004/058946。

O-tRNA 和 O-RS 以及 O-tRNA/O-RS 配对的应用

本发明的组合物和用本发明方法制备的组合物任选存在于细胞内。然后可将本发明的 O-tRNA/O-RS 配对或各组分用于宿主系统的翻译机制中，从而将非天然氨基酸掺入蛋白质。Schultz 等人的名为“非天然氨基酸的体内掺入”的国际公布号 WO 2002/085923 描述了该方法，该篇文献以引用的方式纳入本文。例如，当 O-tRNA/O-RS 配对引入宿主，例如大肠杆菌细胞或酵母时，该配对对选择者密码子，例如琥珀无义密码子起反应而将可外源性加入生长培养基的非天然氨基酸在体内掺入蛋白质，例如肌红蛋白测试蛋白或治疗蛋白中。本发明的组合物可以任选处于体外翻译系统，或细胞体内系统中。含非天然氨基酸的蛋白质的应用范围广泛。例如掺入蛋白质的非天然部分可作为广泛修饰的靶标，如与其它蛋白质、与小分子如标记物或染料和/或生物分子交联。通过这些修饰，掺入非天然氨基酸可产生改进的治疗蛋白，可用于改变或提高酶的催化功能。在一些方面，蛋白质中掺入非天然氨基酸和随后进行修饰有助于研究蛋白质的结构、与其它蛋白质的相互作用，等等。

试剂盒

试剂盒也是本发明特征之一。例如，提供用于在细胞中产生含至少一个非天然氨基酸的蛋白质的试剂盒，所述试剂盒中装有至少一个容器，而该容器含有编码 O-tRNA 的多核苷酸序列、和/或 O-tRNA、和/或编码 O-RS 的多核苷酸序列、和/或 O-RS。在一个实施方式中，所述试剂盒还包含非天然氨基酸 7-羟基香豆素-乙基甘氨酸。在另一实施方式中，所述试剂盒还包含制备蛋白质和/或宿主细胞的使用说明材料。

实施例

提供以下实施例只是为了说明，并非要限制本发明。技术人员应该知道可以改变各种非关键性参数而不会脱离本发明的范围。应该知道，本文所述的实施例和实施方式只是为说明目的，本领域技术人员借鉴这些实施例和实施方式可以作出各种改进或改变而仍属于本申请的构思和权限以及随附权利要求书的范围内。

实施例 1

通用方法

所有化学试剂均获自商业来源，未进一步纯化即使用。用德国莱茵斯滕/卡尔斯鲁厄的布鲁克生物自旋公司 (Bruker BioSpin GmbH, Rheinstetten/Karlsruhe, Germany) 的 Bruker AMX-400 多核分光光度计记录 ^1H NMR 图谱，报道相比于四甲基硅烷的化学位移。在加利福尼亚州拉霍亚市的斯克里普斯质谱中心 (Scripps Center for Mass Spectrometry, La Jolla, CA) 获得蛋白质质谱。

质粒和细胞系

质粒 pBK-lib5 编码詹氏甲烷球菌酪氨酰 tRNA 合成酶 (*Mj*TyrRS) 突变体文库，其中随机选取 (randomized) 残基 Tyr32、Leu65、Phe108、Gln109、Asp158 和 Leu162；此外，使 His70 突变成 Gly，Ala67 突变成 Ala 或 Gly。质粒 pREP(2)/YC 编码 *Mj*tRNA_{CUA}^{Tyr}，在残基 112 处为 TAG 密码子的氯霉素乙酰转移酶 (CAT) 基因，在 T7 启动子控制下的 GFP 基因，在残基 1 和

107 处为 TAG 密码子的突变型 T7 RNA 聚合酶, 和 Tet^r 标记。质粒 pLWJ17B3 编码在 *lpp* 启动子和 *rrnC* 终止子控制下的 *MjtRNA*_{CUA}^{Tyr}, 在 *ara* 启动子控制下的芽孢杆菌 RNA 酶基因 (含有残基 2、44 和 65 处的三个琥珀密码子), 和 Amp^r 标记。质粒 pBAD/JYAMB-4TAG-Myo 编码含阿拉伯糖启动子和 *rrnB* 终止子的突变型抹香鲸肌红蛋白基因, 含 *lpp* 启动子和 *rrnC* 终止子的 *MjtRNA*^{Tyr}_{CUA}, 和四环素抗性标记。利用大肠杆菌菌株 GeneHog®-Fis, F⁻*mcrA* Δ(*mrr-hsdRMS-mcrBC*) 80*lacZ*ΔM15 Δ*lacX74* *recA1* *endA1* *araD139* Δ(*ara-leu*)7697 *galU* *galK* *λrpsL*(Str^R) *nupG*, *fis*::Tn7 (DHFR)研究所表达的含非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸的蛋白质。

实施例 2

合成非天然氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸

非天然香豆素氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸 (也写作 7-羟基香豆素-乙基甘氨酸; 参见图 1, 结构 1)的合成是首先将 N-α-Cbz-L-谷氨酸α-苄酯转变成侧链β-酮酸酯(keto ester), 然后在甲磺酸中使其与间苯二酚反应 (von Pechmann 反应; Brun 等, *Angew. Chem., Int. Ed.*, (2004) 43:3432-3436)从而获得氨基酸 L-(7-羟基香豆素-4-基)乙基甘氨酸。下文详述的这些步骤。

步骤 1-合成乙基丙二酸镁

向搅拌中的 1.5 M 氯化镁和 3 M 乙基丙二酸钾的水溶液中加入 5 体积的异丙醇。过滤后, 真空干燥乙基丙二酸镁。

步骤 2-合成(2S)-2-苄氧基羰基氨基-5-氧代-庚二酸 1-苄酯 7-乙酯 (图 1, 结构 3)

室温下将 Z-Glu-Obzl (N-α-Cbz-L-谷氨酸α-苄酯) (5g, 13.46 mmol)溶解于 50 ml 无水 THF 中。缓慢加入 N,N'-羰基二咪唑 (1.1 当量), 然后在室温下搅拌混合物 2 小时。加入乙基丙二酸镁 (0.55 当量), 室温下搅拌混合物过夜。用乙醚萃取产物, 用 10% NaHCO₃、水和盐水洗涤。通过硅胶快速层析(乙酸乙酯: 己烷 50:50)纯化含有 3 的残留物, 旋转蒸发仪浓缩获得产率为 40%的白色固体。¹H-NMR (400 MHz, CDCl₃): δ 1.26 (t, *J* = 7.2, 3H),

1.9-2.01 (m, 1H), 2.1-2.28 (m, 1H), 2.5-2.7 (m, 2H), 3.37 (s, 2H), 4.18 (q, $J = 7.2$, 2H), 4.41 (s, 1H), 5.11 (s, 2H), 5.18 (s, 2H), 5.44 (d, $J = 8$, 1H), 7.25-7.41 (m, 10H)。C₂₄H₂NO₇ (M+ H⁺)的 LC-MS (ESI)计算值为 442.18, 实测值为 442.0。

步骤 3-合成 L-(7-羟基香豆素-4-基)乙基甘氨酸 (图 1, 结构 1)

将上一步骤的(2S)-2 苄氧基羰基氨基-5-氧代-庚二酸 1-苄酯 7-乙酯 (2.35 g, 5 mmol)缓慢加入甲磺酸(20 ml)配制的间苯二酚 (3 g, 25 mmol)中, 室温下搅拌 2 小时。然后向混合物中加入 5 体积乙醚, 冷却至-30℃。用冷的乙醚洗涤沉淀物, 将其溶解于水中, 过滤并冻干。反相 HPLC 纯化 L-(7-羟基香豆素-4-基)乙基甘氨酸获得产率为 50%的棕色固体 (YMC AA12S052503WT 柱, 12 ml/分钟流速, 水配制的 10%-90% CH₃CN, 0.1% TFA (w/v), 总共 12 分钟)。¹H-NMR (DMSO): δ 2.06 (s, 2H), 2.70-2.90 (m, 2H), 3.82-4.0 (m, 1H), 6.06 (s, 1H), 6.75 (dd, $J = 8.8, 2.4$, 1H), 6.82 (d, $J = 2.4$, 1H), 7.62 (d, $J = 8.8$, 1H), 8.17 (s, 3H)。C₁₃H₁₄NO₅ (M+ H⁺)的 LC-MS (ESI)计算值为 264.08 Da, 实测值为 264.0 Da。

实施例 3

遗传选择对 7-羟基香豆素-乙基甘氨酸具有特异性的突变型合成酶

为在大肠杆菌中将 7-羟基香豆素-乙基甘氨酸选择性掺入蛋白质中的指定位点, 开发了能对 TAG 选择者密码子起反应并对 7-羟基香豆素-乙基甘氨酸具有特异性的突变型詹氏甲烷球菌酪氨酰琥珀抑制型 tRNA (MjtRNA^{Tyr}_{CUA}) / 酪氨酰-tRNA 合成酶 (MjTyrRS)配对。

文库构建

为容纳非天然氨基酸 7-羟基香豆素-乙基甘氨酸较大的香豆素侧链, 制备 MjTyrRS 文库 pBK-lib5, 其中野生型詹氏甲烷球菌酪氨酰-tRNA 合成酶 (图 9; SEQ ID NO: 2 和 3)His70 突变成 Gly, Ala67 固定为 Ala 或 Gly 以增加活性位点大小。随机选取结合的酪氨酸附近的 6 个残基: Tyr-32、Leu-65、Phe-108、Gln-109、Asp-158 和 Leu-162(Kobayashi 等, *Nat. Struct. Biol.*, (2003) 10:425-432; Zhang 等, *Protein Sci.*, (2005) 14:1340-1349)。pBK-lib5 文库

由 2×10^9 个 TyrRS 独立克隆构成, 其采用标准 PCR 方法构建。

正/负选择

随后, 对该文库进行几轮正和负选择。在正选择中, 当用 pBK-lib 和 $MjtRNA^{Tyr}_{CUA}$ 共转染的细胞在有 1 mM 非天然氨基酸(UAA)和氯霉素存在下生长时, 细胞存活依赖于对引入氯霉素乙酰转移酶(CAT)基因容许位点的琥珀密码子的抑制(Xie 和 Schultz, *Methods* (2005) 36:227-238)。然后将正选择的克隆转化入含有 $MjtRNA^{Tyr}_{CUA}$ 和编码毒性芽孢杆菌 RNA 酶蛋白并在容许位点引入 3 个琥珀突变的基因的细胞中。使这些细胞在没有 UAA 存在下生长以除去利用内源性氨基酸的任何克隆(负选择)。下文详述了这些正和负选择。

正选择-含有 pREP(2)/YC 质粒的大肠杆菌 DH10B 用作正选择的宿主菌株。用 pBK-lib5 文库转化细胞, 在 SOC 中回收 1 小时, 先用含亮氨酸的甘油基本培养基(glycerol minimal media with leucine) (GMML)洗涤两次, 再涂布到补加了浓度分别是 50 $\mu\text{g/mL}$ 、60 $\mu\text{g/mL}$ 、15 $\mu\text{g/mL}$ 和 1 mM 的卡那霉素、氯霉素、四环素和 7-羟基香豆素-乙基甘氨酸的 GMML-琼脂平板上。37°C 温育平板 60 小时, 刮下存活的细胞, 提取质粒 DNA 并通过凝胶电泳纯化。

负选择-然后将 pBK-lib5 DNA 转化入含有负选择质粒 pLWJ17B3 的电感受态细胞, 在 SOC 中回收 1 小时, 然后涂布到含 0.2%阿拉伯糖、50 $\mu\text{g/mL}$ 氨苄西林和 50 $\mu\text{g/mL}$ 卡那霉素的 LB-琼脂平板上。然后在 37°C 温育平板 8-12 小时, 如上所述提取存活克隆的 pBK-lib5 DNA。

随后对该文库进行一轮正选择, 接着进行负选择和最后一轮正选择(用 70 $\mu\text{g/mL}$ 的氯霉素)。在此阶段, 选择 96 个单独的克隆, 将它们悬浮在 96-孔板的 50 μL GMML 中, 一式两份点样在两组 GMML 平板上。一组 GMML-琼脂平板补加有四环素 (15 $\mu\text{g/mL}$)、卡那霉素 (50 $\mu\text{g/mL}$)和浓度为 60、80、100 和 120 $\mu\text{g/mL}$ 氯霉素以及 1 mM 7-羟基香豆素-乙基甘氨酸。另一组平板相同, 但不含 7-羟基香豆素-乙基甘氨酸, 所用的氯霉素浓度是 0、20、40 和 60 $\mu\text{g/mL}$ 。37°C 温育 60 小时后, 发现有 1 mM 7-羟基香豆素-乙基甘氨酸存在时, 一个克隆在氯霉素为 100 $\mu\text{g/mL}$ 下可存活, 而没有 7-羟基香豆素-

乙基甘氨酸存在时，只在氯霉素为 20 µg/mL 下可存活。因此，3 轮正选择和 2 轮负选择后，该文库产生在有 1 mM 7-羟基香豆素-乙基甘氨酸存在时在氯霉素为 100 µg/mL 下可存活，而没有 7-羟基香豆素-乙基甘氨酸存在时，只在氯霉素为 20 µg/mL 下可存活的一个克隆。

该克隆(称为 CouRS-D8)经测序发现含有以下 8 个突变：Tyr32Glu、Leu65His、Ala67Gly、His70Gly、Phe108Tyr、Gln109His、Asp158Gly 和 Leu162Gly (参见图 9 及 SEQ ID NO: 4 和 5)。图 9 中 SEQ ID NO: 4 所示的突变型合成酶多肽序列以粗体字表示突变位置。下表进一步总结了突变型合成酶蛋白质。

	氨基酸位置 (密码子)								
<i>Mj</i> 酪氨酰-tRNA 合成酶	32	65	67	70	108	109	158	162	SEQ ID NO:
野生型	Tyr (TAC)	Leu (TTG)	Ala (GCT)	His (CAC)	Phe (TTC)	Gln (CAG)	Asp (GAT)	Leu (TTA)	2 (3)
CouRS-D8 突变体	Glu (GAG)	His (CAT)	Gly (GGT)	Gly (GGC)	Tyr (TAT)	His (CAT)	Gly (GGG)	Gly (GGT)	4 (5)

4 个残基突变成甘氨酸，最可能产生足够的空间以容纳 7-羟基香豆素-乙基甘氨酸残基。在野生型酶中通过氢键与结合的酪氨酸结合的 Tyr32 和 Asp158 也发生突变，这与针对酪氨酸的活性丧失是一致的 (Kobayashi 等, *Nat. Struct. Biol.*, (2003) 10:425-432; Zhang 等, *Protein Sci.*, (2005) 14:1340-1349)。

图 8 显示突变型 *Mj*TyrRS (CouRS-D8)活性位点的预测结构模型。利用野生型 *Mj*TyrRS 结构 (pdb 密码 1J1U)产生模型。酪氨酸底物为 L-(7-羟基香豆素-4-基)乙基甘氨酸所替代，显示有以下突变：Tyr32Glu、Leu65His、Ala67Gly、His70Gly、Phe108Tyr、Gln109His、Asp158Gly 和 Leu162Gly。然后利用美国加利福尼亚州圣迭戈市阿克塞勒斯软件公司 (Accelrys Software, Inc.[®], San Diego, CA, USA)的 InsightII[®]软件，通过能量最小化优化结构。

实施例 4

含有 7-羟基香豆素-乙基甘氨酸的突变型肌红蛋白模型蛋白 的表达和表征

为测定 7-羟基香豆素-乙基甘氨酸掺入蛋白质的效率和保真度,取代含 C-末端 His₆ 标签的抹香鲸肌红蛋白(参见图 3)中位置四(Ser-4)或位置 37(His-37)的琥珀选择者密码子。在选择的合成酶 CouRS-D8 和 *MjtRNA*^{Tyr}_{CUA} 存在下,在基本培养基或补加了 1 mM 7-羟基香豆素-乙基甘氨酸的培养基中培养大肠杆菌以表达蛋白质。在没有 7-羟基香豆素-乙基甘氨酸存在下表达蛋白质作为负对照。该蛋白质的产生和分析如下所详述。

产生突变型蛋白质

将质粒 pBAD/JYAMB-4TAG-Myo 或 pBAD/JYAMB-37TAG-Myo 与表达 CouRS-D8 的 pBK 载体共同独立转染入 GeneHog[®]-Fis 大肠杆菌细胞。在补加了卡那霉素(50 µg/mL)和四环素 (15 µg/mL)的 2YT 培养基 (5 mL)中扩增细胞。利用起始培养液 (1 mL)接种补加了合适抗生素和 1 mM 7-羟基香豆素-乙基甘氨酸的 100 mL 液体 GMML 或 2YT。然后,在 37°C 将细胞培养至 OD₆₀₀ 为 0.5,通过加入 0.2%阿拉伯糖诱导蛋白质表达。另一 4-12 小时后,离心收集细胞。在天然条件下,通过 Ni-NTA 亲和层析纯化 TAG4 和 TAG37 肌红蛋白突变体。然后使样品脱盐。各培养基中 TAG4 和 TAG37 突变型肌红蛋白的产量是 2 mg/L。作为比较,在相似条件下野生型肌红蛋白的产量是 5 mg/L。

蛋白质质谱分析

对纯化的 TAG4 突变型肌红蛋白进行 ESI 质谱分析。突变型肌红蛋白的 ESI-质谱分析获得实测平均分子量 M_{avg} 为 18511 Da,与计算分子量 M_{avg} 为 18512 Da 相一致(参见图 6)。因为当利用 5%的卷积阈值时在 ESI-质谱中未发现对应于天然氨基酸的物质的掺入,因此从蛋白质谱的信噪比获得的 7-羟基香豆素-乙基甘氨酸掺入纯度最低为 95%。如此设置检测阈值,使得只能检测丰度超过主峰的 5%的物质。因为只观察到主峰,因此突变型蛋白的纯度 > 95%。

通过 SDS-PAGE 拆分分析蛋白质

通过 SDS-PAGE 分析纯化的蛋白质显示全长蛋白质只在有 7-羟基香豆素-乙基甘氨酸 (参见图 5)存在下表达。此外,用 SDS-PAGE 拆分后,掺入遗传编码的 7-羟基香豆素-乙基甘氨酸的 TAG4 突变型肌红蛋白显示荧光条带,而缺乏遗传编码的 7-羟基香豆素-乙基甘氨酸的野生型肌红蛋白未显示荧光。荧光成像利用新泽西州皮斯卡塔韦市分子动力/GE®保健生命科学公司的 STORM® 荧光扫描仪 (Molecular Dynamics/GE® Healthcare Life Sciences, Piscataway, NJ)获得。

实施例 5

利用遗传编码的 7-羟基香豆素-乙基甘氨酸作为蛋白质构型的探针

为证明 7-羟基香豆素-乙基甘氨酸的潜在应用之一,将该氨基酸用作全肌红蛋白(holomyoglobin)的脲依赖性变性的探针。因为香豆素荧光对溶剂极性敏感,其荧光强度应与蛋白质 7-羟基香豆素-乙基甘氨酸附近的局部未折叠相关。肌红蛋白结构由通过短环和转角相连的 8 个螺旋(A 到 H)构成 (Delli Castelli 等, *Protein Sci.*, (2002) 11:2273-2276)。将螺旋 A 中的 Ser4 和螺旋 C 中的 His37 (参见图 3; 二残基主要接触溶剂,不会明显与其它附近的残基相互作用)各自独立突变成选择者密码子 TAG,从而在有特异性正交 O-tRNA/O-RS 配对存在下体内遗传编码掺入 7-羟基香豆素-乙基甘氨酸。

如图 4 所示,通过圆二色性监测的野生型、Ser4 → 7-羟基香豆素-乙基甘氨酸和 His37 → 7-羟基香豆素-乙基甘氨酸全肌红蛋白的脲诱导未折叠曲线实质上相同,提示将 7-羟基香豆素-乙基甘氨酸引入螺旋 A 或螺旋 C 不会明显干扰蛋白质稳定性。脲浓度为 2 M 时,全肌红蛋白中位置 4 的 7-羟基香豆素-乙基甘氨酸的荧光强度增加 30% (在 2 M-5 M 脲时维持该水平),提示蛋白质的该区域是无序的 (Zhang 等, *Protein Sci.*, (2005) 14:1340-1349)。相反,脲为 2 M 时,在残基 37 处含有 7-羟基香豆素-乙基甘氨酸的突变型肌红蛋白显示其荧光强度改变不大,但在脲为 3 M 时经历相似的荧光增加。与该结果一致的是,以前的 NMR 研究显示当脲浓度高于 2.2 M 时,螺旋 A 和 B 主要是无序的,如该区域中短和中等范围的 NOE 的消失所示,而螺旋 C、D 和 F 随后在脲浓度高于 3.0 M 时未折叠 (Delli Castelli 等, *Protein Sci.*,

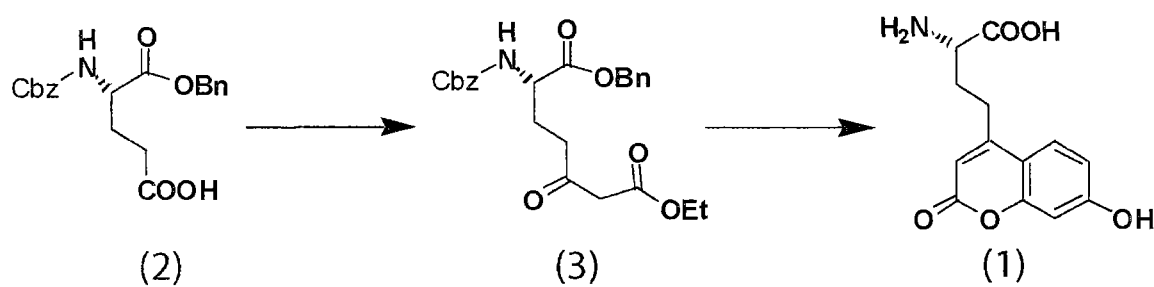
(2002) 11:2273-2276)。因此，与报道整个结构上平均总体改变的圆二色性相反，氨基酸 7-羟基香豆素-乙基甘氨酸似乎是蛋白质构象改变的位点特异性探针。

还测试了 7-羟基香豆素-乙基甘氨酸对 pH 的敏感性，如图 7 所示。在 100 mM 磷酸钠缓冲液中检测 60 μ M 香豆素基氨基酸 7-羟基香豆素-乙基甘氨酸的吸收光谱。360 nm 的吸光度随 pH 升高而增加。从这些检测值计算 pKa 为 7.8。

7-羟基香豆素-乙基甘氨酸对 pH (参见图 7)和溶剂极性的敏感性 (Zinsli, J. *Photochem* (1974) 3:55-69)使其成为许多生物学研究的有用体外和体内探针 (Brun 等, *Angew. Chem., Int. Ed.*, (2004) 43:3432-3436)。例如，可利用氨基酸 7-羟基香豆素-乙基甘氨酸监测生物分子相互作用或蛋白质中的构象改变，或膜结合蛋白质的拓扑学特性。此外，因为 7-羟基香豆素-乙基甘氨酸的 pKa 为 7.8, 该 pKa 可通过取代香豆素环而系统地改变 (Sun 等, *Bioorg. Med. Chem. Lett.*, (1998) 8:3107-3110)，其应该是细胞器 pH 和 pH 依赖性细胞过程的有用探针。此外，激发状态的 7-羟基香豆素是强烈的光酸(photo-acid) (Cohen 和 Huppert, *Phys. Chem. A* (2001) 105:7157-7164)，可促进蛋白质中质子转运过程的研究。

* * *

虽然出于清晰和理解的目的描述了上述发明的一些细节，但应该知道本领域技术人员通过阅读本文内容可对形式和细节作出各种改变而不脱离本发明的范围。如同每一份出版物、专利、专利申请和/或其它文件单独表明出于所有目的以引用的方式纳入本文一样，本申请引用的所有出版物、专利、专利申请和/或其它文件出于所有目的以引用的方式全文纳入本文。



- (1) L-(7-羟基香豆素-4-基)乙基甘氨酸
- (2) N-α-Cbz-L-谷氨酸α-苄酯
- (3) (2S)-2苄氧基羰基氨基-5-氧代-庚二酸1-苄酯7-乙酯

图 1

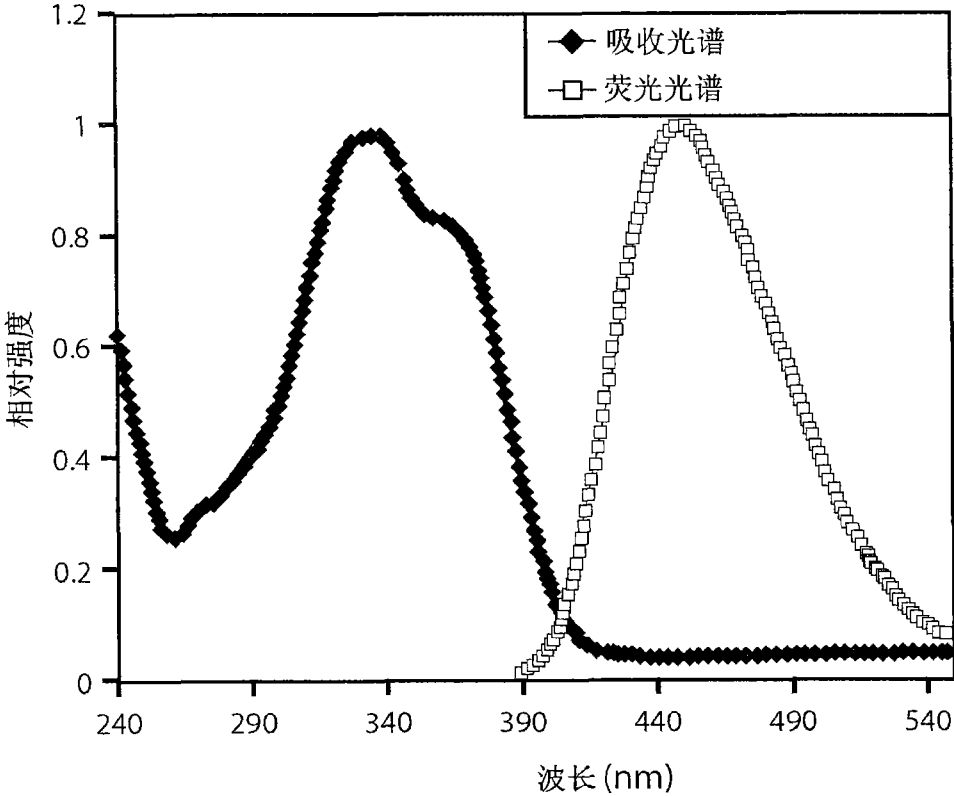


图 2

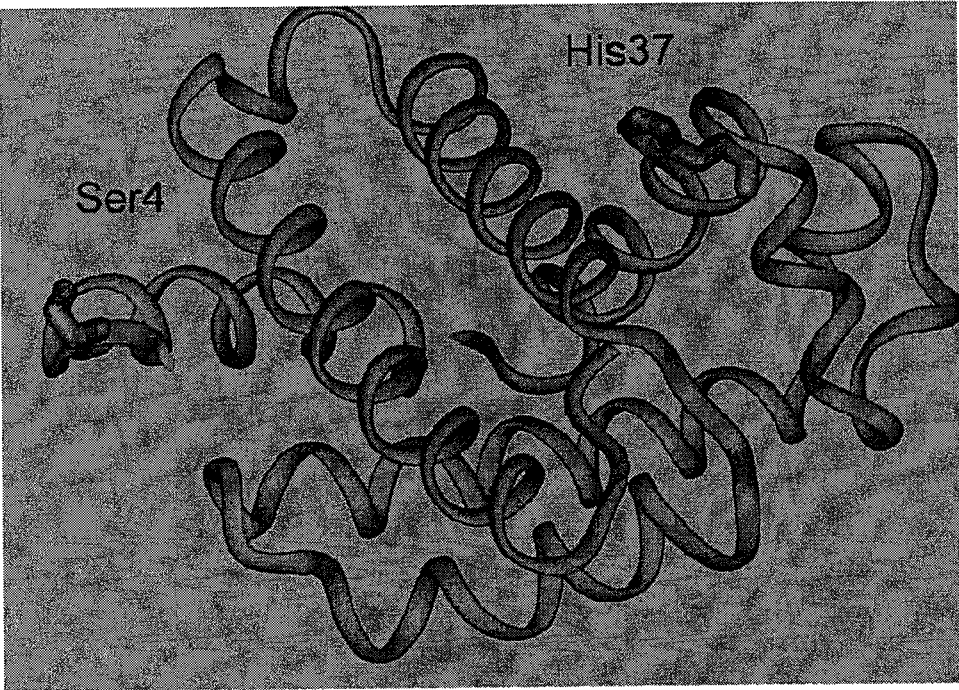


图 3

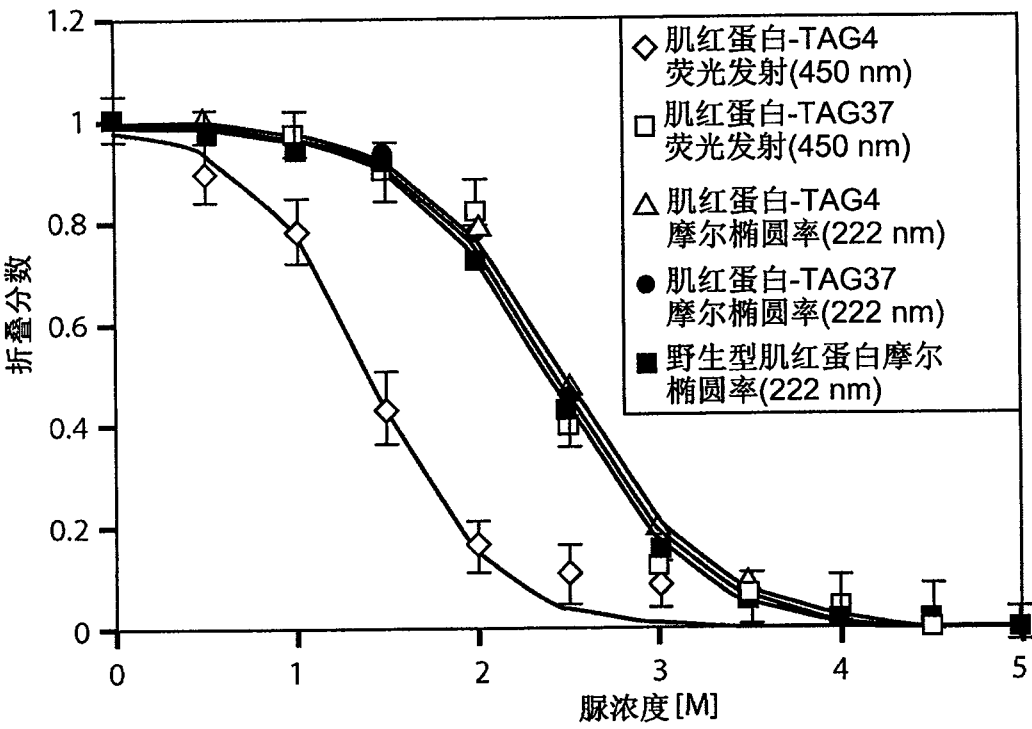


图 4

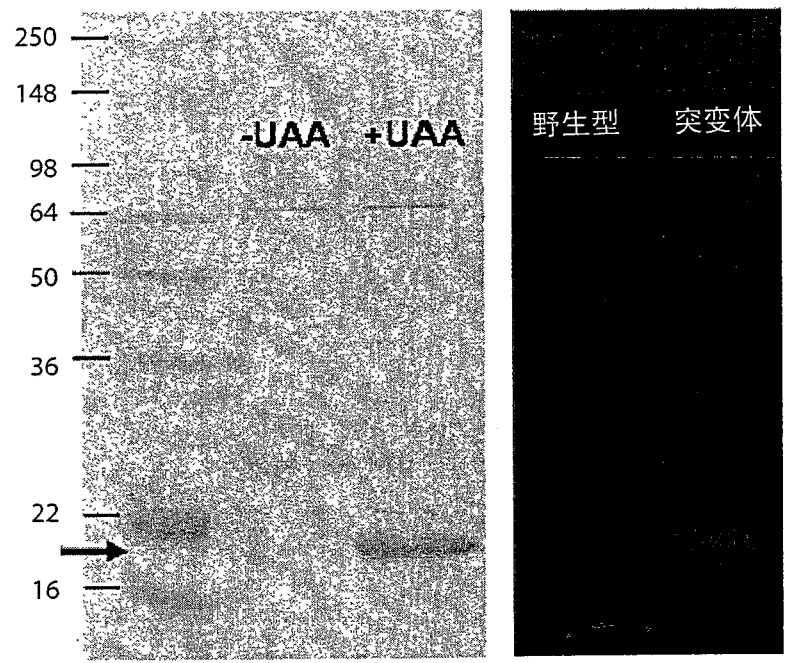


图 5

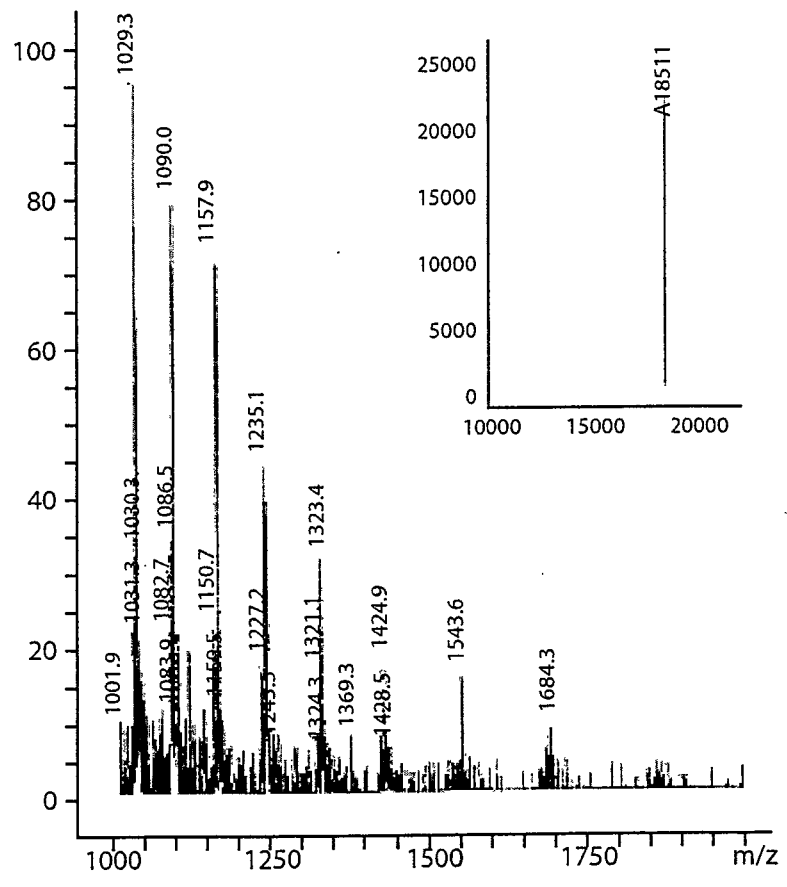


图 6

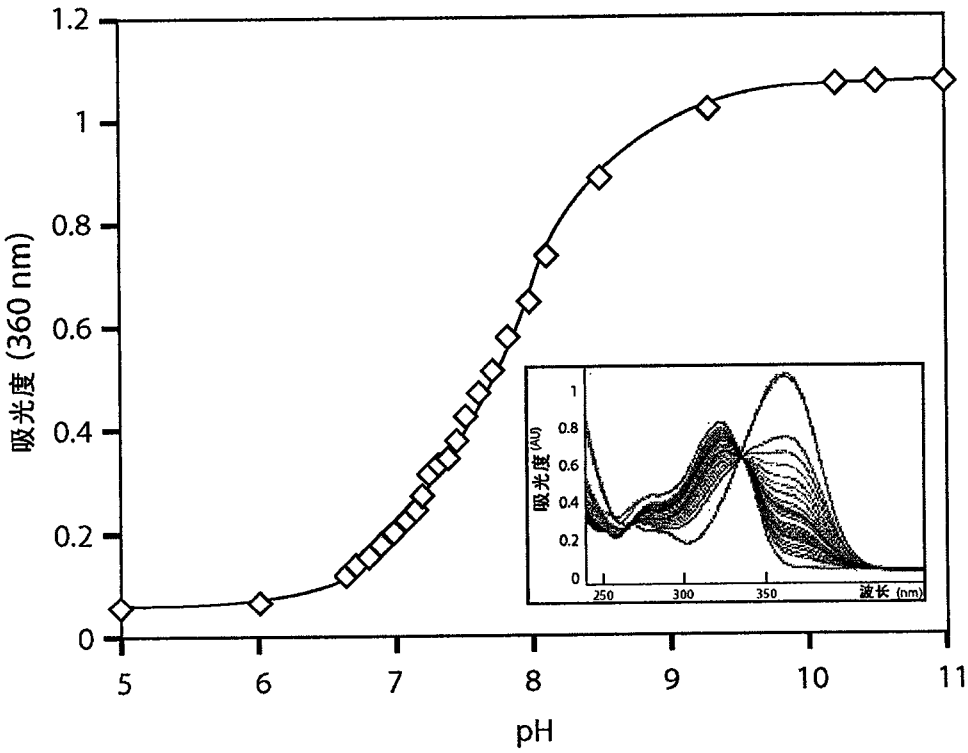


图 7

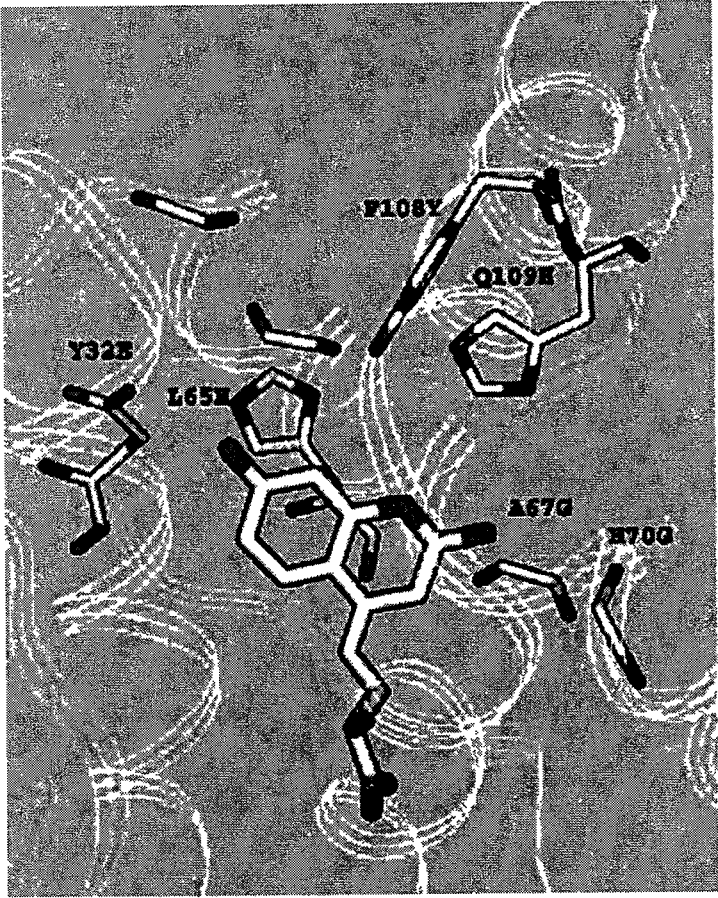


图 8

核苷酸和氨基酸序列

SEQ ID NO:	描述	序列
1	詹氏甲烷球菌抑制型酪氨酰-tRNA ^{CUA} aka MjtRNA-Tyr(CUA) 或 mutRNA ^{Tyr CUA}	CCGGCGGUAGUUCAGCAGGGCAGAACGGCGGACUCUAAAUCCG CAUGGCGCUGGUUCAAUCCGGCCCGCCGACCA
2	野生型詹氏甲烷球菌酪氨酰-tRNA合成酶 (MjTyrRS)氨基酸序列	MDEFEMIKRNTSEIIISEEELREVLKKDEKSAYIGFEPGSKIHL GHYLQIKKMIDLQNAGFDIIILLADLHAYLNQKGELDEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTLKR ARRSMEIAREDENPKVAEVIYPIMQVNDIHYLGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELESLEFKNKELHPMDLKNAVAEELIKILEP IRKRL
3	野生型詹氏甲烷球菌酪氨酰-tRNA合成酶 (MjTyrRS)核苷酸序列	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGC GAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAAATCTGCTTAC ATAGGTTTTGAACCAAGTGGTAAAATACATTTAGGGCATTATCTCCAA ATAAAAAGATGATTGATTTACAAAATGCTGGATTTGATATAATTATA TTGTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGAT GAGATTAGAAAAATAGGAGATTATAACAAAAAGTTTTGAAGCAATG GGGTTAAAGGCAAAATATGTTTATGGAAGTGAATCCAGCTTGATAAG GATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAA AGAGCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGATATTCAT TATTTAGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTAC AACCTGTCTTAACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCA AAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCA ATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAA AGGCCAGAAAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTAAAAA AATGCTGTAGCTGAAGAAGCTTATAAAGATTTTAGAGCCAATTAGAAAG AGATTA
4	L-(7-羟基香豆素-4-基) 乙基甘氨酸氨酰-tRNA 合成酶氨基酸序列 (衍生自野生型詹氏甲烷 球菌酪氨酰tRNA-合成酶)	MDEFEMIKRNTSEIIISEEELREVLKKDEKSAEIGFEPGSKIHL GHYLQIKKMIDLQNAGFDIIHLLGDLGAYLNQKGELDEIRKI GDYNKKVFEAMGLKAKYVYGSEYHLDKDYTLNVYRLALKTTL KRARRSMELIAREDENPKVAEVIYPIMQVNGIHYGGVDVAVG GMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNF IAVDDDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTI KRPEKFGGDLTVNSYEELESLEFKNKELHPMDLKNAVAEELIKI LEPIRKRL

图 9

SEQ ID NO:	描述	序列
5	L-(7-羟基香豆素-4-基) 乙基甘氨酸氨酰-tRNA 合成酶核苷酸序列	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGC GAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAAATCTGCTGAG ATAGGTTTTGAACCAAGTGGTAAAAATACATTTAGGGCATTATCTCCAA ATAAAAAAGATGATTGATTTACAAAATGCTGGATTGATATAATTATA CATTTGGGTGATTTAGGCGCCTATTTAAACCAGAAAGGAGAGTTGGAT GAGATTAGAAAAATAGGAGATTATAACAAAAAAGTTTTTGAAGCAATG GGGTTAAAGGC AAAATATGTTTATGGAAGTGAATATCATCTTGATAAG GATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAA AGAGCAAGAAGGAGTATGGAACCTTAGCAAGAGAGGATGAAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGGGATTCAT TATGGTGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGGAGCTTTTACCAAAAAAGTTGTTTGTATTCAC AACCCGTCTTAACGGGTTTGGATGGAGAAGGAAAGATGAGTTCCTTCA AAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCCAGCTGGAGTTGTTGAAGGAAATCCA ATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAA AGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTAAAA AATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAG AGATTATAA

图 9 (续)