发明名称
正交赖氨酸 - tRNA 和氨酰 - tRNA 合成酶对的组合物及其应用

摘要
本发明提供了制备蛋白生物合成机组分的组合物和方法，所述蛋白生物合成机组分包含正交赖氨酸 tRNA、正交赖氨酸 - 氨酰 tRNA 合成酶和赖氨酸 - tRNA/合成酶正交对，该生物合成机组分将高谷氨酰胺插入到蛋白质中以响应四碱基密码子。本发明还提供了鉴定这些正交对的方法以及利用这些正交对制备含高谷氨酰胺的蛋白质的方法。
1. 一种翻译系统，该翻译系统包括：
正交氨酰 tRNA(赖氨酰 O-tRNA)或其修饰变体；
正交氨酰 tRNA 合成酶(O-RS)，其优先将一个或多个氨基酸开辟于正交氨酰 tRNA 或其修饰变体，或
正交氨酰 tRNA(赖氨酰 O-tRNA)或其修饰变体以及优先将一个或多个氨基酸开辟于赖氨酰 O-tRNA 或其修饰变体的正交氨酰 tRNA 合成酶(O-RS)。

2. 如权利要求 1 所述的翻译系统，其特征在于，所述翻译系统包含细胞。

3. 如权利要求 2 所述的翻译系统，其特征在于，所述细胞是大肠杆菌细胞。

4. 如权利要求 1 所述的翻译系统，其特征在于，所述氨基酸是非天然氨基酸。

5. 如权利要求 4 所述的翻译系统，其特征在于，所述非天然氨基酸是高谷氨酰胺。

6. 如权利要求 1 所述的翻译系统，其特征在于，所述赖氨酰 O-tRNA 或其修饰变体 O-RS 或者二者都来源于嗜热古菌 Pyrococcus horikoshii(PhKRS)。

7. 如权利要求 6 所述的翻译系统，其特征在于，所述 O-RS 是 PhKRS、E444G、Ph Δ AD 或是 Ph Δ AD 的 I41 和/或 S268 突变体。

8. 如权利要求 6 所述的翻译系统，其特征在于，当所述 O-RS 在大肠杆菌内表达时，它所具有的毒性等同于低的 Ph Δ AD 的 I41 和/或 S268 突变体。

9. 如权利要求 1 所述的翻译系统，其特征在于，所述赖氨酰 O-tRNA 或其修饰变体包含四碱基密码子或琥珀密码子的识别序列。

10. 如权利要求 1 所述的翻译系统，其特征在于，所述赖氨酰 O-tRNA 或其修饰变体包含 AGGA 的识别序列。

11. 如权利要求 1 所述的翻译系统，其特征在于，所述赖氨酰 O-tRNA 或其修饰变体包含一个含有 CU(X)nXXXAA 序列的反密码子环。

12. 如权利要求 11 所述的翻译系统，其特征在于，所述 CU(X)nXXXAA 序列包括 CUCUAAA 或 CUUCCUAA。

13. 如权利要求 1 所述的翻译系统，其特征在于，所述赖氨酰 O-tRNA 或其修饰变体包括 SEQ ID NO:24 或 SEQ ID NO:26。

14. 如权利要求 1 所述的翻译系统，其特征在于，所述 O-RS 和赖氨酰 O-tRNA 或其修饰变体抑制终止或移码选择者密码子的效率至少相当于 E444G、Ph Δ AD 或
Ph AD 的 I41 和/或 S268 突变体与 SEQ ID NO:24 或 SEQ ID NO:26 的 O-tRNA 联合时的 50%。

15. 如权利要求 1 所述的翻译系统，该系统包括其它 O-RS 和其它 O-tRNA，其中所述其它 O-RS 和其它 O-tRNA 所抑制的移码选择者密码子不同于赖氨酸 O-tRNA 或其修饰变体和优先加载赖氨酸 O-tRNA 或其修饰变体的 O-RS 所抑制的移码选择者密码子。

16. 如权利要求 1 所述的翻译系统，其特征在于，所述翻译系统对靶多肽编码靶核酸中的四碱基选择者密码子和终止选择者密码子均抑制。

17. 如权利要求 16 所述的翻译系统，其特征在于，所述四碱基选择者密码子包括序列 AGGA，且终止选择者密码子包括序列 TAG 或 UAG。

18. 如权利要求 1 所述的翻译系统，该系统包括含四碱基选择者密码子的靶核酸。

19. 如权利要求 18 所述的翻译系统，该系统包括靶核酸编码的蛋白质。

20. 如权利要求 19 所述的翻译系统，其特征在于，所述蛋白质含有高谷氨酰胺。

21. 如权利要求 1 所述的翻译系统，该系统包括含有四碱基选择者密码子和终止选择者密码子的靶核酸。

22. 如权利要求 21 所述的翻译系统，该系统包括由靶核酸编码的蛋白质，其中所述蛋白质含有至少两个不同的非天然氨基酸。

23. 一种翻译系统，该系统包括：
能识别四碱基选择者密码子的第一正交 tRNA(O-tRNA);
能用第一非天然氨基酸优先加载 O-tRNA 的第一正交氨酰 tRNA 合成酶(O-RS);
能识别终止选择者密码子的第二 O-tRNA；以及
能用第二非天然氨基酸优先加载第二 O-tRNA 的第二 O-RS。

24. 如权利要求 23 所述的翻译系统，其特征在于，所述四碱基选择者密码子是 AGGA。

25. 如权利要求 23 所述的翻译系统，其特征在于，所述终止密码子是 UAG。

26. 如权利要求 23 所述的翻译系统，其特征在于，所述翻译系统包含细胞。

27. 如权利要求 23 所述的翻译系统，其特征在于，所述第一或第二 O-tRNA 是正交赖氨酰 tRNA(赖氨酰 O-tRNA)或其修饰变体。

28. 如权利要求 23 所述的翻译系统，其特征在于，所述第一 O-tRNA 是正交赖氨酰 tRNA(赖氨酰 O-tRNA)或其修饰变体，且所述第二 O-tRNA 是正交酪氨酰 tRNA(酪氨酰 O-tRNA)或其修饰变体。
29. 如权利要求 23 所述的翻译系统，该系统还包括含有至少一个四碱基选择者密码子和一个终止选择者密码子的核酸。

30. 如权利要求 29 所述的翻译系统，其特征在于，所述四碱基选择者密码子是 AGGA，且所述终止选择者密码子是 TAG 或 UAG。

31. 如权利要求 29 所述的翻译系统，其特征在于，所述核酸是表达的 RNA。

32. 如权利要求 29 所述的翻译系统，其特征在于，所述翻译系统包含由核酸编码的蛋白质，该蛋白质含有至少两个不同的非天然氨基酸。

33. 如权利要求 32 所述的翻译系统，其特征在于，所述蛋白质含有高谷氨酰胺。

34. 如权利要求 32 所述的翻译系统，其特征在于，所述蛋白质与肌红蛋白同源。

35. 如权利要求 23 所述的翻译系统，该系统包含高谷氨酰胺。

36. 一种组合物，该组合物包含：PhKRS、E444G、PhΔAD、PhΔAD 的 I41 和/或 S268 突变体、或其保守变体。

37. 一种编码 PhKRS、E444G、PhΔAD、PhΔAD 的 I41 和/或 S268 突变体、或其保守变体的核酸。

38. 一种包含或编码相应于 SEQ ID NO:24 或 SEQ ID NO:26 的 tRNA 或其保守变异的核酸。

39. 一种包含正交氨酰-tRNA 合成酶(O-RS)的组合物，其特征在于，所述 O-RS 优先用高谷氨酰胺使 O-tRNA 氨酰化。

40. 如权利要求 39 所述的组合物，其特征在于，所述 O-RS 包括 PhΔAD 的 I41 和/或 S268 突变体或其保守变异。

41. 如权利要求 39 所述的组合物，其特征在于，所述 O-RS 优先氨酰化 O-tRNA，其效率至少是 PhΔAD 的 I41 和/或 S268 突变体效率的 50%。

42. 如权利要求 39 所述的组合物，其特征在于，所述 O-RS 来源于超嗜热古菌。

43. 如权利要求 39 所述的组合物，该组合物包含 O-tRNA，其中所述 O-tRNA 识别四碱基选择者密码子。

44. 如权利要求 43 所述的组合物，其特征在于，所述四碱基选择者密码子包括 AGGA 序列。

45. 如权利要求 39 所述的组合物，该组合物包含细胞，其中所述 O-RS 被细胞内的一种或多种核酸编码。

46. 如权利要求 45 所述的组合物，其特征在于，所述细胞是大肠杆菌细胞。

47. 如权利要求 39 所述的组合物，该组合物包含翻译系统。
48. 如权利要求 39 所述的组合物，该组合物包含细胞，其中所述 O-RS 被细胞内的一种或多种核酸编码，该细胞还包含：
正交-tRNA(O-tRNA); 和
高谷氨酰胺;
其中，所述 O-tRNA 识别第一选择者密码子，且所述 O-RS 优先用第一高谷氨酰胺使 O-tRNA 氨酰化。

49. 如权利要求 48 所述的组合物，其特征在于，所述细胞包含编码目的多肽的靶核酸，其中所述靶核酸包含能被 O-tRNA 识别的选择者密码子。

50. 如权利要求 48 所述的组合物，其特征在于，所述 O-tRNA 包含 SEQ ID NO:24 或 SEQ ID NO:26 中所列出的多核苷酸序列或其互补多核苷酸序列，或由上述序列编码，其中所述 O-RS 包含相应于 E444G、Ph Δ AD、Ph Δ AD 的 I41 和/或 S268 突变体，或其保守变异的氨基酸序列。

51. 如权利要求 48 所述的组合物，其特征在于，所述 O-RS 和 O-tRNA 抑制终止选择者密码子或移码选择者密码子的效率至少相当于 E444G、Ph Δ AD 或 Ph Δ AD 的 I41 和/或 S268 突变体联合序列为 SEQ ID NO:24 或 SEQ ID NO:26 的 O-tRNA 时的 50%。

52. 如权利要求 48 所述的组合物，其特征在于，所述细胞是大肠杆菌细胞。

53. 如权利要求 48 所述的组合物，其特征在于，所述细胞还包括其它不同的 O-tRNA/O-RS 对和其它不同的非天然氨基酸，其中所述 O-tRNA 识别第二选择者密码子，且所述 O-RS 优先用第二非天然氨基酸氨酰化 O-tRNA。

54. 如权利要求 53 所述的组合物，其特征在于，所述细胞包含含有第一和第二选择者密码子的靶核酸。

55. 如权利要求 54 所述的组合物，其特征在于，所述细胞含有由靶核酸编码的蛋白质，该蛋白质含有至少两个不同的非天然氨基酸。

56. 一种含蛋白质的组合物，其特征在于，所述蛋白质含有高谷氨酰胺。

57. 如权利要求 56 所述的组合物，其特征在于，所述蛋白质含有与野生型治疗蛋白、诊断蛋白、工业用酶的氨基酸序列或其部分至少 75% 相同的氨基酸序列。

58. 如权利要求 56 所述的组合物，其特征在于，所述组合物含有药学上可接受的载体。

59. 一种筛选能将高谷氨酰胺加栽到正交 tRNA(O-tRNA) 上的活性正交-氨酰-tRNA 合酶(O-RS)的方法，该方法包括：
以一群细胞为筛选对象，其中所述细胞都包含：

1) O-tRNA，其中所述 O-tRNA 与细胞群中含 O-tRNA 的成员交交；

2) 一组 O-RS，其包含的一个或多个活性 O-RS 成员可在细胞群中的一个或多个细胞内将高谷氨酰胺加载到 O-tRNA 上：

3) 编码可选择标记的多核苷酸，其中所述多核苷酸含有至少一个能被 O-tRNA 识别的选择子密码子；以及

4) 高谷氨酰胺；

其中，所述细胞群内含活性 O-RS 的靶细胞是通过其对可选择标记的抑制效率相对于缺少 RS 群但含有 O-tRNA 的对照细胞的抑制效率的增强而被识别的；以及

筛选靶细胞，从而筛选出活性 O-RS。

60. 如权利要求 59 所述的方法，其特征在于，所述细胞还经进一步筛选以去除那些含有非靶 O-RS 的细胞，这些非靶 O-RS 用高谷氨酰胺之外的其它氨基酸加载 O-tRNA。

61. 如权利要求 59 所述的方法，其特征在于，所述筛选包括正筛选，且所述可选择标记包括正选择标记。

62. 如权利要求 59 所述的方法，其特征在于，所述一组 RS 包含突变的 RS、第一物种之外的一个或多个物种来源的 RS、或者同时含有突变的 RS 和第一物种之外的物种来源的 RS。

63. 一种用权利要求 59 所述方法识别出的正交氨酰-tRNA 合成酶。

64. 一种在细胞内制备特定位置上的高谷氨酰胺的蛋白质的方法，该方法包括：

在合适的培养基内培养细胞，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，以及

提供高谷氨酰胺；

其中，所述细胞还包含：

能识别选择者密码子的正交-tRNA(O-tRNA)；以及

优先用高谷氨酰胺氨酰化 O-tRNA 的正交氨酰-tRNA 合成酶(O-RS)；以及

根据选择者密码子而将高谷氨酰胺插入到特定位置上，从而制备出所述蛋白质。

65. 如权利要求 64 所述的方法，其特征在于，所述 O-RS 含有相应于 E444G、Ph Δ AD、Ph Δ AD 的 I41 和/或 S268 突变体或其保守变异的氨基酸序列。
说明书

正交赖氨酸-tRNA 和氨基-tRNA 合成酶对的组合物及其应用

相关申请的交叉参考
本申请要求 2003 年 7 月 7 日提交的临时专利申请 US 60/485,451；2003 年 12 月 10 日提交的临时专利申请 US 60/528,815 和 2004 年 1 月 15 日提交的临时专利申请 US 60/537,149 的优先权，并获益于上述临时专利申请，本申请已将其公开的内容完整纳入作为参考。

有关由联邦政府资助的研究和开发项目所形成的发明的权利声明
本发明是在政府资助下完成的，资助来自于能源部的基金 DE-FG0300ER45812。因此政府对本发明也享有一定的权利。

发明领域
本发明属于翻译生物化学领域。本发明涉及制备正交 tRNA、正交氨酰 tRNA 合成酶及 tRNA/氨酰 tRNA 合成酶对的方法和组合物，其中的蛋白质插入有非天然氨基酸，如高谷氨酰胺 (homoglutamine)，以响应选择者密码子 (selector coden) 如四碱基和终止选择者密码子。其中包括在一个蛋白链内插入多个不同的非天然氨基酸以响应终止选择者密码子和四碱基选择者密码子。本发明还涉及利用这种 tRNA/氨酰 tRNA 合成酶对及其相关的组合物在细胞内制备蛋白质的方法。

发明背景


发明概述

本发明提供了利用正交 tRNA 和正交氨酰 tRNA 合成酶制备蛋白产物的新型翻译
系统。本发明涉及装备的翻译系统、使用该翻译系统的方法以及该系统制备出的翻译产物。该技术可用于本文所描述的许多方面，例如而不限于治疗性产品、诊断试剂和工业用酶的制备。

一方面，本发明提供了一种翻译系统，该系统可利用正交赖氨酸 tRNA(赖氨酸 O-tRNA)或 O-tRNA 的修饰变体和/或优先用一个或多个氨基酸加载正交氨酰 tRNA 的正交氨酰 tRNA 合成酶(O-RS)或 O-RS 的修饰变体。在某些实施方式中，翻译系统位于细胞外，如大肠杆菌内。结合到 O-tRNA 上的氨基酸还可以是非天然氨基酸，如高谷氨酰胺。

在翻译系统的某些实施方式中，赖氨酰 O-tRNA、O-tRNA 的变体、O-RS，或者二者都来源于超嗜热古菌(Pyrococcus horikoshii)(PhKRS)。现已发现许多 O-RS 都可用于该翻译系统，其中包括 PhKRS、E444G、PhΔΔ 和 PhΔΔ 的 141 和/或 S268 突
变体。在某些实施方式中，当超嗜热古菌 O-RS 在大肠杆菌细胞内表达时所具有的毒性等于或低于 PhΔAD 的 I41 和/或 S268 突变体。

赖氨酰 O-tRNA 或突变的 O-tRNA 还可以包含四碱基密码子或琥珀密码子的识别序列。例如，赖氨酰 O-tRNA 或其变体可包含 AGGA 的识别序列。在翻译系统的一些相关方面，赖氨酰 O-tRNA 或其变体可利用一个序列为 CU(X),XXXAA 的反密码子环。在这一方面的某些实施方式中，CU(X),XXXAA 序列可以是 CUCUAAA 或 CUUCCUAA。实施例包括包含序列 SEQ ID NO:24 或 SEQ ID NO:26 的赖氨酰 O-tRNA 或其变体。

在本发明的某些实施方式中，O-RS、赖氨酰 O-tRNA 或其变体抑制终止选择者密码子或移码选择者密码子的效率至少要达到 E444G、PhΔAD 或 PhΔAD 的 I41 和/或 S268 突变体联合具有序列 SEQ ID NO:24 或 SEQ ID NO:26 的 O-tRNA 的 50%。

在另外的实施方式中，翻译系统可利用其他的 O-RS 和其他的 O-tRNA，这些其他的组分可抑制移码选择者密码子，这些移码选择者密码子与赖氨酰 O-tRNA 或 O-tRNA 变体和优先加载赖氨酰 O-tRNA 或 O-tRNA 变体的 O-RS 所抑制的移码选择者密码子不同。在其他的实施方式中，翻译系统既抑制靶多肽编码靶核酸上的四碱基选择者密码子，也抑制终止选择者密码子。四碱基选择者密码子还可以使用序列 AGGA，终止选择者密码子还可以使用序列 TAG 或 UAG。在某些实施方式中，翻译系统包括含四碱基选择者密码子的靶核酸。翻译系统还可以包含靶核酸编码的蛋白质。例如，这个蛋白质可含有高谷氨酰胺残基。

翻译系统还可以包含编码四碱基选择者密码子和终止选择者密码子的靶核酸。在某些方面，翻译系统包含靶核酸编码的蛋白质，其中的蛋白质包含至少两个不同的非天然氨基酸。

例如，本发明提供了一种翻译系统，该系统利用可识别四碱基选择者密码子的第一正交 tRNA (O-tRNA)、将第一非天然氨基酸优先加载到 O-tRNA 上的第一正交氨基酸 tRNA 合成酶(O-RS)、可识别终止选择者密码子的第二 O-tRNA 以及将第二非天然氨基酸优先加载到第二 O-tRNA 上的第二 O-RS。在一个实施方式中，四碱基选择者密码子是 AGGA，终止密码子是 UAG。翻译系统还可以位于细胞内。

在翻译系统的某些实施方式中，第一或第二 O-tRNA 是正交赖氨酰 tRNA (赖氨酰 O-tRNA) 或其修饰变体。另外，第一 O-tRNA 可以是正交赖氨酰 tRNA (赖氨酰 O-tRNA) 或其适宜变体，第二 O-tRNA 是正交酪氨酰 tRNA (酪氨酰 O-tRNA) 或其适宜变体。翻译系统还可以包含至少编码一个四碱基选择者密码子和一个终止选择者密
码子的核酸。在这些实施方式中，四碱基选择者密码子可以是 AGGA，终止选择者密码子可以是 TAG 或 UAG，被翻译的核酸一般是表达出的 RNA。在某些实施方式中，核酸编码的蛋白质内至少包含两个不同的非天然氨基酸，如高谷氨酰胺和第二非天然氨基酸，如亲电子的氨基酸。翻译系统可以包含各种非天然氨基酸，其中包括高谷氨酰胺。在一个实施例中，翻译系统内的蛋白质与肌红蛋白同源，但是其中包含一个非天然氨基酸如高谷氨酰胺。

本发明提供了含有蛋白质的组合物，其中的蛋白质包含但不仅限于 PhKRS、E444G、PhΔAD、PhΔAD 的 I41 和/或 S268 突变体的氨基酸序列，或者这些蛋白质的保守变体。本发明同样提供了编码 PhKRS、E444G、PhΔAD、PhΔAD 的 I41 和/或 S268 突变体的核酸，或者这些核酸的保守变体。本发明提供了包含 SEQ ID NO:24 或 SEQ ID NO:26 相应的 tRNA 的部分或完整序列的核酸，或者这些核酸的任意保守变体。

在其他方面，本发明提供了包含但不仅限于正交氨酰-tRNA 合成酶(O-RS)的组合物，其中的 O-RS 优先用高谷氨酰胺氨酰化 O-tRNA。这种 O-RS 含有 PhΔAD 的 I41 和/或 S268 突变体相应的突变，或者这个蛋白的任何保守变体。在某些实施方式中，O-RS 优先氨酰化 O-tRNA，其效率至少为 PhΔAD 的 I41 和/或 S268 突变体的 50%。在某些实施方式中，O-RS 来源于超嗜热古菌。在 O-RS 和 O-tRNA 都包含于组合物内的某些实施方式中，O-tRNA 可识别四碱基选择者密码子，如 AGGA。

在上述组合物包含细胞或存在于细胞内的某些实施方式中，O-RS 可被细胞内的一种或多种核酸编码，这些细胞可以是大肠杆菌细胞。组合物可包含翻译系统。在包含细胞并且其中的 O-RS 是被细胞内的一种或多种核酸编码的组合物中，细胞还包含能识别第一选择者密码子和高谷氨酰胺的正交 tRNA (O-tRNA)，例如，其中的 O-RS 优先用高谷氨酰胺氨酰化 O-tRNA。在某些实施方式中，细胞包含编码目的多肽的靶核酸，其中的靶核酸编码一个能被 O-tRNA 识别的选择者密码子。

O-tRNA 还可以被多核苷酸的部分或完整序列编码，如上述的 SEQ ID NO:24 或 SEQ ID NO:26，或其互补的多核苷酸序列，O-RS 包含相应于 PhKRS、E444G、PhΔAD、PhΔAD 的 I41 和/或 S268 突变体的氨基酸序列，或者该序列的任何保守变体。需要说明的是本文所描述的任一核酸序列都可以 RNA 或 DNA 的形式出现；因此，除非在上下文中特别说明，序列如 SEQ ID NO:24 或 25 可以是 RNA 的形式，也可以是 DNA 的形式，无论是否清晰地指出。在某些实施方式中，O-RS 和 O-tRNA 抑制终止选择者密码子或移码选择者密码子的效率至少要达到 E444G、PhΔAD 或 PhΔAD 的
I41 和/或 S268 突变体联合具有序列 SEQ ID NO:24 或 SEQ ID NO:26 的 O-tRNA 的 50%。在组合物包含细胞的实施方式中，细胞可以是大肠杆菌细胞。这个组合物的细胞还可以包含其他不同的 O-tRNA/O-RS 对和其他不同的非天然氨基酸，其中 O-tRNA 可识别第二选择者密码子，O-RS 优先用第二非天然氨基酸氨酰化 O-tRNA。细胞还可以包含编码第一和第二选择者密码子的靶核酸。另外，这种组合物的细胞还可以包含靶核酸编码的蛋白质，其中的蛋白质至少插入有两个不同的非天然氨基酸。

本发明还提供了至少包含一个高谷氨酰胺的蛋白质。在某些实施方式中，蛋白质所包含的氨基酸序列与野生型治疗性蛋白、诊断蛋白、工业用酶或这些蛋白的任何部分的序列至少有 75% 相同。另外，蛋白质上还可以连接药学上可接受的载体。

在其他方面，本发明提供了筛选有活性的正交-氨酰-tRNA 合成酶(O-RS) 的方法，其中的 O-RS 可将高谷氨酰胺加栽到正交 tRNA(O-tRNA) 上。这些方法的第一是提供一群细胞用于筛选，这些细胞都包含 1) O-tRNA，其中 O-tRNA 对于细胞群内含有 O-tRNA 的细胞来说是正交的；2) 包括一种或多种有活性的 O-RS 成员的 O-RS 群，这些成员可在该细胞群的一种或多种细胞内将高谷氨酰胺加栽到 O-tRNA 上；3) 编码选择标记的多核苷酸，其中多核苷酸至少编码一个可被 O-tRNA 识别的选择者密码子；以及 4) 高谷氨酰胺。在这个筛选过程中，与无 RS 但是有 O-tRNA 的对照细胞相比，细胞群内含有活性 O-RS 的靶细胞因其对选择标记的抑制效应增强而被筛选出来。该方法的第二步是筛选含有活性 O-RS 的靶细胞。本发明还提供了通过任意筛选方法鉴定正交氨酰-tRNA 合成酶的方法。

在这些方法的某些实施方式中，细胞还可以被进一步筛选以去除那些含有非靶 O-RS 的细胞，其中的非靶 O-RS 可以将高谷氨酰胺之外的氨基酸加栽到 O-tRNA 上。在某些实施方式中，筛选是正筛选，选择标记是正选择标记。

在这些方法的不同实施方式中，RS 群可以是不同来源的，其中包括而不限于突变的 RS、第一物种之外的一个或多个物种来源的 RS，或者二者都有。

本发明还提供了在细胞内制备在特定位置上是高谷氨酰胺的蛋白质的方法。该方法包括在适宜培养基中培养细胞的步骤，其中的细胞含有至少编码一个选择者密码和一个蛋白质的核酸，例如，其中的细胞还可以包含能识别选择者密码子的正交-tRNA (O- tRNA) 和优选用高谷氨酰胺氨酰化 O-tRNA 的正交氨酰-tRNA 合成酶 (O-RS)，提供高谷氨酰胺；将高谷氨酰胺插入到特定位置以响应选择者密码子，从而制备出蛋白质。在该方法的一个实施方式中，O-RS 的氨基酸序列利用了 E444G、PhΔAD、PhΔAD 的 I41 和/或 S268 突变体、或者其保守变体的完整或部分序列。
定义

在详细描述本发明前，需要说明的是本发明并不局限于某些特定的生物系统，本发明可以有各种变化。还应当说明的是本文所用的方法只是为了描述特定的实施方式，并不意味着对本发明的限制。如本发明的说明书和附属权利要求所使用的，除非特别指明，单数形式“一个”、“一种”和“这种”也包括复数。因此，“一个细胞”也包括两个或多个细胞；“细菌”也包括细菌的混合物等。

除了在此处和说明书的其余部分所定义的，本文所用的所有技术术语和科学术语都与本发明所属领域的技术人员所通常理解的意义相同。

正文氨基酸-tRNA：本文所用的术语正交氨基酸-tRNA(赖氨酸-O-tRNA)是与目的翻译系统正交的tRNA，其中tRNA：(1)与天然的赖氨酸tRNA相同或基本相似，(2)来源于发生自然突变或人工诱变的天然赖氨酸tRNA，(3)由考虑到(1)或(2)的野生型或突变赖氨酸tRNA序列的任意方法所制备，(4)与野生型或突变的赖氨酸tRNA同源，(5)与表1中所列的被定义为赖氨酰tRNA合成酶底物的典型tRNA同源，或(6)表1中所列的被定义为赖氨酰tRNA合成酶底物的典型tRNA的保守变体。赖氨酰tRNA可以携带一个氨基酸，或者处于非负载状态。还需要说明的是“赖氨酰-O-tRNA”还可以被同源的合成酶加载(氨酰化)上赖氨酸之外的氨基酸，如高谷氨酰胺。本发明的赖氨酰-O-tRNA确实可用于将任何必需的氨基酸，无论是天然的或是人工合成的，插入到正在合成的多肽内，如翻译期间，以响应选择者密码子。

正文氨基酸氨基酸合成酶：本文所用的正文氨基酸氨基酸合成酶(赖氨酰-O-RS)是一种可以在目的翻译系统中优先将氨基酸加载到赖氨酰-O-tRNA上的合成酶。被赖氨酰-O-RS加载到赖氨酰-O-tRNA上的氨基酸可以是任何氨基酸，无论是天然的还是人工合成的，仅限于本文所描述的。合成酶还可以与天然的赖氨酰氨基酸合成酶相同或同源，或者与表1所列的赖氨酰-O-RS相同或同源。例如，赖氨酰-O-RS可以是表1所列的赖氨酰-O-RS的保守变体，和或与表1所列的赖氨酰-O-RS的序列至少有50%、60%、70%、80%、90%、95%、98%、99%或更高的相同性。

正文：本文所用的术语“正交”是指一个分子(如正交tRNA(O-tRNA)和/或正交氨酰tRNA合成酶(O-RS))的功能与细胞的内源性组分一样，但是其活性与细胞或翻译系统内的其相应的内源性分子相比下降，或者不具有细胞内源性组分的功能。在用于tRNA和氨酰tRNA合成酶时，正交的是指正交的tRNA与内源性tRNA合成酶配合时其效率与内源性tRNA和内源性tRNA合成酶配合时相比下降，如下降到20%、
10%、5%或1%以下。正交分子缺乏细胞内内源性互补分子的正常功能。例如，与内源性的tRNA被内源性RS氨酰化相比，细胞内的正交tRNA被细胞的内源性RS氨酰化时其效率降低，甚至根本检测不到。在另外两个实施例中，与内源性的RS氨酰化内源性的tRNA相比，正交的RS在目的细胞内氨酰化任何内源性的tRNA时其效率降低，或者根本检测不到。第二正交分子可以被引入到细胞内和第一正交分子一起发挥功能。例如，正交tRNA/RS对包括引入的互补组分，在细胞内一起行使功能时其效率与对照，如相应的内源性tRNA/RS对或活性正交(tRNA/RS对)相比可以达到45%、50%、60%、70%、75%、80%、90%、95%、99%或更高。

同源：术语“同源”是指能联合发挥功能的组分，如正交tRNA和优先氨酰化正交tRNA的正交氨酰-tRNA合成酶。该组分也被称为“互补的”。

优先氨酰化：术语“优先氨酰化”是指O-RS在将一个特定氨基酸加载到特定tRNA上时的效率比加载其他tRNA上时要高。例如，O-RS加载同源O-tRNA时的效率比氨酰化非同源tRNA(如作为底物用于制备同源O-tRNA的tRNA，如通过突变)时的效率要高(如70%、75%、85%、90%、95%、99%或更高)。

选择者密码子：术语“选择者密码子”是指在翻译过程中能被O-tRNA识别但通常不被内源性tRNA识别的密码子。典型的例子包括终止密码子、含四个或更多碱基接的密码子等。O-tRNA反密码子环可识别选择者密码子，如表达的RNA如mRNA内的选择者密码子，通过翻译系统的翻译可将其相应的氨基酸插入到多肽内。例如，在本文的一个实施方式中，O-tRNA可识别选择者密码子如四碱基密码子，通过翻译过程将一个非天然氨基酸如高谷氨酰胺插入到多肽内。选择者密码子包括无义密码子如终止密码子，例如琥珀密码子、脯氨酸密码子和乳白密码子；四碱基或多碱基密码子；稀有密码子；天然或非天然碱基对来源的密码子等。

抑制型tRNA：抑制型tRNA是一种可在给定的翻译系统内改变信使RNA(mRNA)的阅读的tRNA，例如通过将氨基酸插入到多肽链内以响应选择者密码子的机制。例如，抑制型tRNA可阅读终止密码子、四碱基密码子、稀有密码子等。

抑制活性：如本文所述，术语“抑制活性”通常是指tRNA(如抑制型tRNA)在翻译过程中阅读过密码子(如为琥珀密码子或四碱基或多碱基密码子的选择者密码子)的抑制作用，如果阅读不过则翻译终止或产生错误译(如移码突变)。抑制型tRNA的抑制活性可以表达为与第二抑制型tRNA或对照系统如缺乏O-RS的对照系统相比所观察到的翻译通读活性的百分数。

本发明提供了定量测定抑制活性的各种方法。一种特定O-tRNA和ORS对目的
选择者密码子的抑制百分率是指一种位于编码表达检测标记的核酸上的给定表达检测标记（如 LacZ），包括选择者密码子，在目的翻译系统内的活性相对于阳性对照构建体活性的百分数，其中的目的翻译系统包含 O-RS 和 O-tRNA，而阳性对照不包含 O-tRNA、O-RS 和选择者密码子。因此，如果缺乏选择者密码子的有活性的阳性对照标记构建体在一个给定翻译系统内的活性是 X，那么含有选择者密码子的被检测构建体的抑制百分率为被检测构建体在与阳性对照标记基本相同的表达环境下所展示的 X 的百分比，其中被检测标记构建体所表达的翻译系统还包含 O-tRNA 和 O-RS。一般说来，表达被检测标记的翻译系统还包含能被 O-RS 和 O-tRNA 识别的氨基酸。另外，百分抑制率还可以通过比较被检测标记与“背景”或“阴性”对照标记构建体来加以精炼，后者含有与被检测标记相同的选择者密码子，但是系统中不包含 O-tRNA、O-RS 和/或能被 O-tRNA 和/或 O-RS 识别的相关氨基酸。这种阴性对照由于考虑到了标记在目的翻译系统中所产生的背景信号，因此可以使百分抑制率更加准确。

抑制效率可通过本领域熟知的多种方法检测。例如，可以利用 β 半乳糖苷酶报告基因分析方法，如将衍生的 lacZ 质粒（其中构建体含有选择者密码子和 lacZ 核酸序列）和含有本发明的 O-tRNA 的质粒一起导入到适宜有机体（如可使用正交组分的有机体）的细胞内。同源的合成酶也要导入（或者以多肽或编码同源合成酶的多核苷酸的形成）。细胞在培养基内生长到预期的密度时，如 OD600 达到约 0.5 时，利用 BetaFluor™ β-半乳糖苷酶检测试剂盒（Novagen）进行 β 半乳糖苷酶活性分析。百分抑制率可以计算为样品相对于可比较对照的活性百分数，如衍生的 lacZ 构建体所展示的活性，其中构建体在预期位置上含有有义密码子而不是选择者密码子。

翻译系统：术语“翻译系统”是指可以将氨基酸插入到正在合成的多肽链（蛋白）上的组分。翻译系统的组分包括核糖体、tRNA、合成酶、mRNA 等。本发明的 O-tRNA 和/或 O-RS 也可以加入到体外或体内翻译系统中，或者作为翻译系统的一部分，如非真核细胞如细菌（如大肠杆菌）或真核细胞如酵母细胞、哺乳动物细胞、植物细胞、藻类细胞、真菌细胞、昆虫细胞等内的翻译系统。

非天然氨基酸：如本文所述，术语“非天然氨基酸”是指 20 个天然氨基酸或稀有天然氨基酸硒氨酸或吡咯赖氨酸之外的任何氨基酸、修饰氨基酸和/或氨基酸类似物，如高谷氨酰胺。

来源于：本文所用的术语“来源于”是指从特定分子或有机体、或者特定分子或有机体的信息中分离的组分，或者利用特定分子或有机体、或者特定分子或有机
体的信息制备的组分。

阳性选择标记：本文所用的术语“阳性选择标记”是指当一个标记存在时，如表达、活化等，会导致具有该标记相应性状的细胞被鉴定出来，如可以将具有阳性选择标记的细胞与无此形状的细胞区别开来。

阴性选择标记：本文所用的术语“阴性选择标记”是指一个标记存在时，如表达、活化等，可以鉴定出不具有被筛选特性或形状的细胞(如，与具有该特性或形状的细胞比较时)。

报告子：本文所用的术语“报告子”是指可用于鉴定和/或筛选目的系统内靶组分的组分。例如，报告子可以是蛋白，如能赋予抗生素耐药性或敏感性的酶(如β-内酰胺酶、氯霉素乙酰转移酶(CAT)等)、荧光选择标记(如绿色荧光蛋白(如 GFP)、YFP、EGFP、RFP 等)、冷光标记(如萤火虫荧光素酶蛋白)、亲和性选择标记，或者正或负选择标记基金如 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”是指在系统进化树内属于真核生物域的有机体，如动物(如哺乳动物、昆虫、爬行动物、鸟类等)、纤毛虫、植物(如单子叶植物、双子叶植物、藻类等)、真菌、酵母、鞭毛虫、微孢子虫、原生生物等。

非真核生物：本文所用的术语“非真核生物”是指非真核有机体。例如，非真核有机体可能属于原核细菌系统发育区(如大肠杆菌、嗜热栖杆菌、嗜热脂肪芽孢杆菌等)或古细菌(如詹氏甲烷球菌(Methanococcus jannaschii) (Mj)、马氏微球菌(Methanosarcina mazei)(Mm)、热自养甲烷球菌(Methanobacterium thermoautotrophicum) (Mtl)、海沼甲烷球菌(Methanococcus maripaludis)、坎氏甲烷嗜热菌(Methanopyrus kandleri)、盐杆菌属如沃氏盐杆菌(Haloferax volcanii)和嗜盐菌 NRC-1(Halobacterium species NRC-1)、闪烁古生球菌(Archaeoglobus fulgidus)(Af)、激烈火球菌(Pyrococcus furiosus)(Pf)、超嗜热古菌(Ph)、超耐高温热棒菌(Pyrobaculum aerophilum)、火球菌(Pyrococcus abyssi)、硫磺矿硫化叶菌(Sulfolobus solfataricus) (Ss)、硫化叶菌(Sulfolobus tokodaiii)、Aeuropyrum pertinax (Ap)、噬酸热原体(Thermoplasma acidophilum)、火山热原体(Thermoplasma volcanium)等)系统发育区。

保守变体：如本文所描述，术语“保守变体”用于指翻译组分时是指一个翻译组分，如保守突变的 O-tRNA 或保守突变的 O-RS，其功能特性与基本组分 O-tRNA 或 O-RS 相似，但是在序列与参照 O-tRNA 或 O-RS 相比有些不同。例如，O-RS 可用非天然氨基酸如高谷氨酰胺酰化互补的 O-tRNA 或保守突变的 O-tRNA，但是
O-tRNA 和保守突变的 O-tRNA 在序列上是有差异的。保守变体在序列上可能含有一个、两个、三个、四个、五个或更多个突变，只要保守变体与相应的 O-tRNA 或 O-RS 能够互补。

筛选试剂：本文所用的术语“筛选试剂”是指当一个试剂存在时可以从一群组分中筛选出某种特定组分。例如，筛选试剂包括而不限于营养素、抗生素、光的波长、抗体、表达的多核酸等。筛选试剂可以使用不同的浓度、密度等。

响应：本文所用的术语“响应”是指本发明的 tRNA 识别选择者密码子并将相应的氨基酸，如非天然氨基酸如高谷氨酰胺，插入到正在合成的多肽链中的过程，其中的氨基酸是由 tRNA 携带的。

编码：如本文所描述，术语“编码”是指利用多聚大分子或序列符号串上的信息指导合成第二分子或序列符号串的过程，其中的第二分子或序列符号串与第一分子或序列符号串是不同的。如本文所描述，该术语的使用范围很宽，可用于许多方面。一方面，术语“编码”可用于描述 DNA 的半保留复制过程，其中双链 DNA 分子的一条链作为模板由 DNA 依赖的 DNA 聚合酶编码出新合成的一条互补姊妹链。

另一方面，术语“编码”是指利用一个分子上的信息指导合成化学性质与第一分子不同的第二分子的过程。例如，DNA 分子编码一个 RNA 分子（如通过 DNA 依赖的 RNA 聚合酶参与的转录过程）。另外，RNA 分子也可以编码多肽，比如在翻译过程中。当该术语用于描述翻译过程时，该术语也可以扩展到编码氨基酸的三联密码子。在某些方面，RNA 分子也可编码 DNA 分子，比如通过 RNA 依赖的 DNA 聚合酶参与的反转录过程。在另一个方面，DNA 分子可编码多肽，应当理解的是，在这种情况下，术语编码是指转录和翻译两个过程。

附图简述

图 1 描述的是古细菌 tRNA^{1-38} 序列的排列。Pa，火球菌（Pyrococcus abyssi）；Pf，激烈火球菌；Ph，超嗜热古菌；Pya，超耐高温热棒菌；Ta，噬酸热原体；Tv，火山热原体（Thermoplasma volcanum）；Af，闪烁古生球菌；Hh，嗜碱菌 NRC-1；Mj，詹氏甲烷球菌；Mt，热自养甲烷球菌；Mm，马氏微球菌；St，硫化叶菌（Sulfolobus tokodaii）；Ss，硫磺矿硫化叶菌；Ap，Aeropyrum pernix 来源的基因组序列用 GCG 程序 pileup 排列，并用 prettybox 程序展示。

图 2 A 和 B 都是直方图，描述的是在体外的种间氨酰化。（A）大肠杆菌完整 tRNA 或（B）嗜盐菌完整 tRNA 被 EcKRS（□）、PhKRS（■）或无合成酶（▲）氨酰化。试验在 20μl
的反应液内进行，其中含有 50 mM Tris-Cl, pH 7.5, 30 mM KCl, 20 mM MgCl₂, 3 mM谷胱甘肽, 0.1 mg/mL BSA, 10 mM ATP, 1μM [³H]赖氨酸(Amersham), 750 nM 合成酶和 0、2、10 或 40μM 全 tRNA，反应在 37°C 进行 20 分钟。

图 3 用图表的形式描述了有序序列来源的琥珀抑制型 tRNA。古细菌 tRNA Gly 序列家族的有序序列在于三叶草构型。反密码子环从有序序列变成 CUCUAAA，形成 AK CUA。

图 4 是一个直方图，描述的是正交合成酶-tRNA 对的体内活性。GeneHogs 细胞(Invitrogen)中图中所示的质粒转染，或者不用质粒转染，设四复孔，测定其中的 β-半乳糖苷酶的活性。质粒 pKQ 表达 PhKRS 的 E44G 突变体。质粒 pKQ 作为无合成酶的样品。质粒 pACGFP 表达 AK CUA，质粒 pACGFP 作为不含 tRNA 的样品。LacZ 报告基因来自于质粒 pLASC-lacZ。细胞在 2YT 培养基内生长，加入适量的抗生素，当细胞生长到 OD₆₀₀ 等于 0.5 时用 Miller 的方法进行分析(Miller, 1972),《分子遗传学实验》(Experiments in molecular genetics)，冷泉港实验室，冷泉港，纽约)。

图 5.琥珀抑制型 tRNA 和四碱基抑制型 tRNA 的构建。琥珀抑制型 tRNA 构建自多种 tRNA Gly 序列的多序列排列。正交 AGGA 抑制型 tRNA 通过筛选受体文库而获得。

图 6A 和 6B.图 6A 显示的是 PhKRS 活性位点的结构。残基 E41 和 Y268 可与赖氨酸底物特异性接触。这些残基可同时随机用于构建活性位点库。图 6B 显示的是各种氨基酸的结构。

图 7A 和 7B 提供的是化学荧光磷光光像(chemiluminescence phosphoimage)，显示的是 AGGA 抑制子对肌红蛋白表达的影响。图 7A，在 AKUCU tRNA 和 PhΔAD，一种 hGln 特异性变体，或 JYRS 存在的情况下，G24 位为 AGGA 密码子的肌红蛋白基因的表达。琥珀抑制子在 S4 位上对肌红蛋白表达的抑制作用，其企图与用 PhΔAD 或 JYRS 一样。b, hGln 通过 AGGA 抑制子在 24 位上插入，AzPhe 通过琥珀抑制子在 75 位上插入，二者插入到同一条多肽链上。

图 8.含高谷氨酰胺或赖氨酸的肽酶水解片段的基质辅助的电荷电离飞行时间(MALDI-TOF)分析。

图 9. 在 24 位含高谷氨酰胺、在 75 位含 O-甲基-酪氨酸的全长肌红蛋白的电喷雾质谱分析。

详细描述
为了在体内将其他的合成氨基酸如高谷氨酰胺插入到基因密码内，需要设计一个能在翻译系统中发挥正常功能但是对于组织内的翻译系统来说是“正交的”的新型氨酰-tRNA合成酶/tRNA 正交对；也就是说该正交对在翻译系统内可独立地行使合成酶和 tRNA 的功能。我们希望正交对所具有的特性包括 tRNA 只能解码或识别一个特异性的新密码，如筛选密码，该密码不被任何内源性的 tRNA 所解码，氨酰-tRNA 合成酶只用特异性的非天然氨基酸如高谷氨酰胺优先氨酰化（或加载）其同源的 tRNA。O-tRNA 也不能被内源性的合成酶氨酰化。例如，在大肠杆菌内，一个正交对包含一个基本不能氨酰化任何内源性 tRNA 的氨酰 tRNA 合成酶和一个基本不能被任何内源性合成酶氨酰化的正交 tRNA，在大肠杆菌内，内源性的 tRNA 有 40 种，内源性的氨酰 tRNA 合成酶有 21 种。

在本文中我们描述了以古细菌 tRNA^{Lys} 序列制备新的正交合成酶/tRNA 对的过程，该正交对可有效而特异地将氨基酸高谷氨酰胺（hGln）插入到肌红蛋白内以响应四碱基选择者密码子 AGGA。hGln 所产生的移码校正抑制不会显著影响蛋白的产量或细胞的生长速率，与第二 O-tRNA-ORS 对所产生的 TAG 抑制相正交。这一研究的结果表明可用的三联密码子的数量和翻译机器本身都不再成为进一步扩展密码子的显著障碍。

为了使四联密码子能在体内编码非天然氨基酸，人们必须制备出能特异识别这个密码子的正交 tRNA (O-tRNA) 以及相应的合成酶，这个合成酶只能利用非天然的目的氨基酸来氨酰化这个 O-tRNA。由于以前所制备的正交氨基酸烷球菌琥珀抑制型 tRNA 的反密码子环是同型合成酶 JYRS 的关键识别元件，因此使用这个环来解码四碱基密码子是比较困难的。虽然有可能使 JYRS 反密码子环的结合特异性下降，但是如果只用反密码子环序列来构建与琥珀抑制子和四碱基抑制子不同的互补正交对可能是比较困难的。因此，本发明利用不同来源的正交对来构建一个系统，该系统可将两个或多个不同的非天然氨基酸同时插入到多肽中以响应两个或多个不同的选择者密码子。

本发明提供了鉴定和制备另外的正交 tRNA-氨酰-tRNA 合成酶对，如 O-tRNA/O-ORS 对的组合物和方法，其中的正交对可用于插入非天然氨基酸，如高谷氨酰胺。本发明的典型 O-tRNA 能够在体内使高谷氨酰胺插入到多核苷酸编码的蛋白质内，其中的多核苷酸含有能被 O-tRNA 识别的选择者密码子。O-tRNA 的反密码子环可识别 mRNA 上的选择者密码子，并将其对应的氨基酸如高谷氨酰胺插入到多肽的这个位点上。本发明的正交氨酰-tRNA 合成酶只用特异性的非天然氨基酸优先
氨酰化(或加载)其O-tRNA。

正交tRNA/正交氨酰-tRNA合成酶及其对


一般来说，当正交对识别选择者密码子并加载氨基酸以响应选择者密码子时，正交对被称为“抑制”了选择者密码子。这就是说，不被翻译系统(如细胞)的内源性机器识别的选择者密码子通常不被翻译，导致多肽合成的中断，否则多肽就可以从核酸上翻译出来。本发明的O-tRNA可识别选择者密码子，与包含或编码于上文所列的多核苷酸序列的O-tRNA相比，在同源合成酶存在的条件下，本发明的O-tRNA为了响应选择者密码子而产生的抑制效率至少可达45%、50%、60%、75%、80%、90%或更高。O-RS可用非天然的目的氨基酸如高谷氨酰胺氨酰化O-tRNA。细胞利用O-tRNA/O-RS对将非天然氨基酸插入到正在合成的多肽链内，例如通过一个包含编码目的多肽的多核苷酸的核酸，其中的多核苷酸含有能被O-tRNA识别的选择者密码子。在某些我们期望的方面，细胞可以包含其他的O-tRNA/O-RS对，其中的O-tRNA可被其他的O-RS加载上不同的非天然氨基酸。例如，一个O-tRNA可识别四碱基密码子，而另一个可以识别终止密码子。另外，多个不同的终止密码子或多个不同的四碱基密码子也可以特异性地识别不同的选择者密码子。

在本发明的某些实施方式中，细胞如大肠杆菌细胞包含正交tRNA(O-tRNA)、正交氨酰-tRNA合成酶(O-RS)、高谷氨酰胺和包含编码目的多肽的多核苷酸的核酸，其中的多核苷酸含有能被O-tRNA识别的选择者密码子。翻译系统也可以是无细胞系统，如各种的市售“体外”转录/翻译系统再加上本文所描述的O-tRNA/ORS对和
非天然氨基酸。

在一个实施方式中，O-RS 和 O-tRNA 联合在一起所产生的抑制效率至少是缺少 O-RS 的 O-tRNA 的抑制效率的 5 倍、10 倍、15 倍、20 倍、25 倍或更高。在一方面，O-RS 和 O-tRNA 联合在一起所产生的抑制效率至少是如上文所述序列的正交合成酶对抑制效率的约 35%、40%、45%、50%、60%、75%、80%、90%或更高。

如上所述，本发明还可以包括位于细胞或其他翻译系统内的多个 O-tRNA/O-RS 对，这样就可以插入不止一个非天然氨基酸，如高谷氨酰胺和其他的非天然氨基酸。例如，细胞还可以包含其他的 O-tRNA/O-RS 对和第二种非天然氨基酸，其中这个其他的 O-tRNA 可识别第二选择者密码子，这个其他的 O-RS 优先用第二种非天然氨基酸氮酰化 O-tRNA。例如，包含一种 O-tRNA/O-RS 对(其中 O-tRNA 可识别琥珀选择者密码子)的细胞还可以包含第二种正交对，如亮氨酸、赖氨酸、谷氨酸 tRNA 等(其中第二种 O-tRNA 可识别不同的选择者密码子，如乳白密码子、四碱基密码子等)。不同的正交对最好来自不同，这样有利于对不同选择者密码子的识别。

O-tRNA 和或 O-RS 可以是天然的，也可以是通过突变天然的 tRNA 和或 RS 而获得的，例如通过制备各种有机体来源的 tRNA 文库和或 RS 文库，或者通过各种已有的突变技术来获得。例如，一种制备正交 tRNA/氨酰-tRNA 合成酶对的策略是将宿主之外的有机体或多种有机体来源的异源(相对于宿主细胞来说)tRNA/合成酶对导入到宿主细胞内。候选异源合成酶的特性包括能够加载宿主细胞的任何 tRNA，优选异源 tRNA 的特性包括不能被宿主细胞的任何合成酶氨酰化。另外，异源 tRNA 对于宿主细胞的合成酶来说是正交的。

制备正交对的第二种策略是制备突变文库，然后从中筛选出 O-tRNA 或 O-RS。这些策略可联合使用。

正交 tRNA(O-tRNA)

我们期望本发明的正交 tRNA(O-tRNA)可以在体内和体外介导非天然氨基酸如高谷氨酰胺向蛋白质内的插入，其中的蛋白质是由含选择者密码子的多核苷酸编码的，这种选择者密码子可被 O-tRNA 识别。在某些实施方式中，本发明的 O-tRNA 与包含或编码于上文所列的多核苷酸序列的 O-tRNA 相比，在同源合成酶存在的条件下，为了响应选择者密码子而产生的抑制效率至少可达 45%、50%、60%、75%、80%、90%或更高。

抑制效率可通过本领域熟知的多种方法检测。例如，可以利用 β 半乳糖苷酶报告基因分析方法，如将衍生的 lacZ 质粒(其中构建体含有选择者密码子和 lacZ 核酸序
列)和含有本发明的 O-tRNA 的质粒一起导入到适宜有机体(如可使用正交组分的有机体)的细胞内。同源的合成酶也要导入(或者以多肽或编码同源合成酶的多核苷酸的形式)。细胞在培养基内生长到预期的密度时,如 OD_{600} 达到约 0.5 时,利用 BetaFluor™ β-半乳糖苷酶检测试剂盒 (Novagen) 进行 β 半乳糖苷酶活性分析。百分抑制率可以计算为样品相对于可比较对照的活性百分数,如衍生的 lacZ 构建体所展示的活性,其中构建体在预期位置上含有有义密码子而不是选择者密码子。

本发明的典型 O-tRNA 为本文所列的序列。也可参见本文的表、实施例和图中所显示的典型 O-tRNA 和 O-RS 分子的序列。还可参见本文的 “核酸和多肽序列及其变体” 部分。在 RNA 分子,如 O-RS mRNA 或 O-tRNA 分子中,其尿嘧啶是由一个给定序列(或者其编码 DNA)或其互补链的胸腺嘧啶替换而来的。序列中还可以有其他的碱基修饰。

本发明还包括本文特定的 O-tRNA 相应的保守突变的 O-tRNA。例如, O-tRNA 的保守变体包括那些与特定 O-tRNA 功能类似的分子,如本文所列的序列以及那些依赖本身互补而保持 tRNA L 型结构但是其序列与本文的列表、图或实施例中所列序列(最好不是野生型的 tRNA 分子)不一样的序列。参见本文的 “核酸和多肽序列及其变体” 部分。

含 O-tRNA 的组合物还可以包含正交氨基酰-tRNA 合成酶(O-RS),其中的 O-RS 优先用非天然氨基酸如高谷氨酰胺酰化 O-tRNA。在某些实施方式中,含有 O-tRNA 的组合物还可以包含翻译系统(如在体外或体内)。含有编码目的多肽的多核苷酸的核酸,或者这些物质的一种或多种的混合物也可以存在于细胞内,其中的多核苷酸含有能被 O-tRNA 识别的选择者密码子。参见本文的 “正交氨基酰-tRNA 合成酶” 部分。

制备正交 tRNA(O-tRNA)的方法也是本发明的特征。利用该方法制备的 O-tRNA 也是本发明的特征。在本发明的某些实施方式中, O-tRNA 可通过构建突变体文库来制备。突变的 tRNA 文库可用本领域已知的各种技术构建。例如, 突变的 tRNA 可通过位于特异性突变技术、随机点突变技术、同源重组技术、DNA 重组技术或其他回归突变方法、嵌合构建、或者混合使用上述技术来制备。

其他的突变也可以被引入到特定位置上, 如 tRNA 环或区的非保守区、保守区、随机区或者混合区, 如反密码子环、接受臂、D 择或环、可变环、TPC 或 tRNA 分子上的其他区域。一般来说, tRNA 分子上的突变包括突变 tRNA 突变体文库内每个成员的反密码子环使利用其能够识别选择者密码子。该方法还能够将一个额外的序列 (CCA) 加到 O-tRNA 的末端。一般来说, O-tRNA 对目的有机体的正交性要高于其起
始材料，如 tRNA 序列群，但是却保留了对目的 RS 的亲和性。

本发明的方法还包括分析 tRNAs 和或 tRNA 合成酶序列的相似性(和或理论上源性)以确定对特定有机体具有正交性的潜在候选 O-tRNA、O-RS 和或 O-tRNA/O-RS 对。本领域熟知的以及本文所描述的计算机程序都可用于这种分析，如 BLAST 和 pileup 程序。在一个实施例中，为了筛选潜在的正交翻译组分用于原核有机体大肠杆菌，需要选择序列与原核有机体不一致的合成酶和/或 tRNA。

一般来说，可以通过负筛选第一物种的细胞群来获得 O-tRNA，其中的细胞含有 O-tRNA 群的成员。负筛选可去除含有 O-tRNA 文库的成员的细胞，其中的 O-tRNA 成员可被细胞内源性的氨酰-tRNA 合成酶 (RS) 氨酰化。这样就可以得到对第一物种的细胞具有正交性的 tRNA 库。

在某些实施方式中，进行负筛选时需要将选择者密码子引入到编码负选择标记的多核苷酸中，如能赋予抗生素耐药性的酶如 β 内酰胺酶，编码可检测产物的酶如 β 半乳糖苷酶、氯霉素乙酰转移酶 (CAT)，毒性产物如自杀基因引入到非必须位置上(还能产生功能性的自杀产物)等。筛选还可以通过使细胞群在含筛选试剂(比如抗生素，如氨苄青霉素)的培养基中生长来完成。在一个实施方式中，筛选试剂的浓度是变化的。

例如，为了测定抑阻型 tRNA 的活性，根据选择者密码子如无义密码子或移码突变在体内的抑阻活性使用一种筛选系统，其中的选择者密码子被引入到编码负性选择标记如 β 半乳糖苷酶基因 (bla) 的多核苷酸中。例如，构建在某个位置(如 A184)上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。如果存在无法被大肠杆菌内源性合成酶有效加载的正交 tRNA，细菌所具有的抗生素耐药性如氨苄青霉素耐药性应当约等于或低于未被质粒转化的细菌的耐药性。如果 tRNA 不是正交的，或者能加载 tRNA 的异源合成酶在系统内共表达，那么应该可以观察到更高水平的抗生素耐药性，如氨苄青霉素耐药性。挑出那些无法在抗生素浓度约等于未用质粒转化的细胞的 LB 琼脂平板上生长的细胞，如细菌。

如果存在毒性产物(如 RNA 酶或自杀酶)，当一群候选 tRNA 成员被宿主如大肠杆菌的内源性合成酶氨酰化时(即对于宿主如大肠杆菌的合成酶来说不是正交的)，选择者密码子被抑制，所产生的毒性多核苷酸产物可导致细胞的死亡。而含有正交 tRNA 或非功能性 tRNA 的细胞存活下来。

在一个实施方式中，目的有机体正交的 tRNA 库需要进行正筛选，其中选择者密码子位于正选择标记，例如耐药基因如 β 内酰胺酶基因编码的正选择标记内。正筛
选在含有编码或包含细胞正交的tRNA的多重核苷酸编码正选择标记的多重核苷酸编码同源性RS的多重核苷酸的细胞内进行。在某些实施方式中，第二群细胞含有无法通过负筛选去除的细胞。多重核苷酸在细胞内表达，细胞可含筛选试剂如氨苄青霉素的培养基内生长。然后根据tRNA被共表达的同源合成酶氨酰化的能力以及响应其选择者密码子而插入氨基酸的能力来筛选tRNA。一般来说，这些细胞与含有非功能性的tRNA或不能被目的合成酶有效识别的tRNA的细胞相比其抑制效应应该是升高的。含有非功能性的tRNA或不能被目的合成酶有效识别的tRNA的细胞对抗生素敏感。因而那些(i)不是宿主如大肠杆菌内源性合成酶的底物；(ii)能被目的合成酶氨酰化；以及(iii)在翻译过程中有功能的tRNA可在两轮筛选中保留下来。

相应的，根据其被筛选的环境不同，同一个标记既可以是正选择标记，也可以是负选择标记。就是说，如果为了筛选它则是正选择标记，如果为了耐受它则是负选择标记。

在上面所描述的方法中，筛选，如正筛选、负筛选或者正负筛选的严格程度还包括改变筛选的严格条件。例如，由于自杀酶是毒性很高的蛋白质，负筛选条件的严格程度可通过将不同数目的选择者密码子引入到自杀基因内和/或通过使用可诱导启动子来加以调控。在另一个实施例中，筛选试剂的浓度是可以改变的(如氨苄青霉素的浓度)。在本发明的一个方面，筛选条件的严格程度是随时变化的，因为前几轮筛选中目的活性较高。因此，在前几轮筛选中筛选程度度是较低的，而在后几轮筛选中需要更严格的筛选条件。在某些实施方式中，负筛选、正筛选或者正负筛选可重复多次。也可以使用多个不同的负选择标记、正选择标记或正负选择标记。在某种实施方式中，正选择标记和负选择标记可以是相同的。

其他类型的筛选方法也可用于本发明中以制备出正交的翻译组分，如O-tRNA、O-RS和能加载非天然氨基酸如高谷氨酰胺以响应选择者密码子的O-tRNA/O-RS对。例如，负选择标记、正选择标记或正负选择标记可以是在适当反应物存在的情况下能发荧光或催化发光反应的标记。在另一个实施方式中，标记的产物可通过荧光激活的细胞分选法(FACS)或发光检测技术检测。另外，标记还可以是亲和选择标记。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. Proc Natl Acad Sci USA, 90:10444-8。

制备重组正交tRNA的其他方法见于题为“Methods and compositions for the production of orthogonal tRNA-aminoacyl-tRNA synthetase paris”(用于制备正交tRNA-

正交氨酰-tRNA 合成酶(O-RS)

本发明的 O-RS 在体外或体内优先用非天然氨基酸如高谷氮酰胺氨酰化 O-tRNA。本发明的 O-RS 可通过加入含 O-RS 的多肽或编码 O-RS 或 O-RS 片段的多核苷酸而提供给翻译系统。例如，O-RS 的例子包含本文列表和实施例中所列的氨基酸序列或其保守变体。在另一个实施例中，O-RS 或 O-RS 片段是由多核苷酸或其互补序列编码的，该多核苷酸编码本文列表或实施例中所列的氨基酸序列。参见本文列表和实施例中所列的典型 O-RS 分子的序列。也可参见本文的“核酸和多肽序列及其变体”部分。

鉴定可与 O-tRNA 作用的正交氨酰-tRNA 合成酶(O-RS)，如 O-RS 的方法也是本发明的特征。例如，该方法包括筛选，如正筛选，第一物种的细胞群，其中的个别细胞含有：1) 一组氨酰-tRNA 合成酶(RS)的一个成员(如这组 RS 内可能包括突变的 RS、第一物种以外的物种来源的 RS，或者二者都包括)；2) 正交 tRNA(O-tRNA)(如一个或多个物种来源的)；以及 3) 编码选择标记(如正选择标记)并且含有至少一个选择者密码子的多核苷酸。筛选出那些与缺乏 RS 成员或 RS 成员数目少的细胞相比抑制效率升高的细胞。抑制效率可用本领域熟知的和本文所描述的方法测定。抑制效率升高的细胞包含有活性的 RS，这些 RS 可氨酰化 O-tRNA。来自于第一物种的第一组 tRNA 被有活性的 RS 氨酰化的水平(体外或体内)与来自于第二物种的第二组 tRNA 被有活性的 RS 氨酰化的水平相比较。氨酰化的水平可通过检测底物(标记的氨基酸或非天然氨基酸，如标记的高谷氨酰胺)来确定。一般来说被筛选出的有活性的 RS 氨酰化第二组 tRNA 的效率比第一组更高，因此可获得能更有效作用于 O-tRNA 的正交氨酰-tRNA 合成酶。利用这种方法鉴定出的 O-RS 也是本发明的特征。


鉴定出的 O-RS 还可以被进一步修饰以改变合成酶的底物特异性，这样 O-RS 就只能将我们所期望的非天然氨基酸如高谷氨酸胺而不是其他任何的 20 种常见氨基酸加载到 O-tRNA 上。制备对非天然氨基酸具有底物特异性的正交氨酰 tRNA 合成酶的方法包括突变合成酶，例如在合成酶的活性位点、编辑机制位点上突变合成酶，在不同的位点上将合成酶的不同区域连接起来等等，然后通过筛选过程加以筛选。还有一种策略是根据先使用正筛选，随后通过负筛选来完成。在正筛选过程中，被引入到正选择标记非必需位置上的选择者密码子所产生的抑制效应可使细胞在正筛选压力下存活下来。如果同时存在天然氨基酸和非天然氨基酸，存活下来的细胞就可以编码有活性的合成酶，这些合成酶利用天然的或非天然的氨基酸可氨酰化正交的抑制型 tRNA。在负筛选过程中，被引入到负选择标记非必需位置的选择者密码子所产生的抑制效应可以去除具有天然氨基酸特异性的合成酶。经过负筛选和正筛选存活下来的细胞所编码的合成酶只能用非天然氨基酸来氨酰化(加载)正交的抑制型 tRNA。这些合成酶可被进一步突变，如通过 DNA 重组技术或其他回归突变方法突变。

突变的 O-RS 文库可利用本领域熟知的各种技术来构建。例如，可通过位点特异性突变技术、随机点突变技术、同源重组技术、DNA 重组技术或其他回归突变方法、嵌合构建、或者混合使用上述技术来制备。例如，可利用两个或多个较小的、低分散的“亚文库”来构建突变的 RS 文库。本发明还包括 RS 的嵌合文库。应当注意的是也可以构建各种有机体(例如微生物如真细菌或古细菌)的 tRNA 合成酶文库，如具有天然多样性的文库(见 Short 等人的美国专利 No. 6,238, 884; Schallenger 等人的美国专利 No. 5,756, 316; Petersen 等人的美国专利 No. 5,783, 431; Thompson 等人的美国专利 No. 5,824, 485; Short 等人的美国专利 No. 5, 958, 672)来筛选正交对。

合成酶一旦通过正筛选和负筛选过程，这些合成酶就应该进一步被诱变。例如，分离编码 O-RS 的核酸：利用该核酸制备出一组编码突变的 O-RS(如通过随机突变、位点特异性突变、重组或联合技术)的多核苷酸；这些步骤的每一步或者所有这些步骤都可以重复进行直到获得能够优先用非天然氨基酸如高谷氨酸氨酰化 O-tRNA 的突变 O-RS。在本发明的一个方面，步骤可进行多次，如至少两次。
其他严格程度的筛选条件也可用于本发明的方法中以制备 O-tRNA、O-RS 或 O-tRNA/O-RS 对。筛选条件的严格程度在制备 O-RS 的一个步骤或两个步骤中可以是不同的。这包括改变筛选试剂的使用浓度等。还可以进行额外轮次的正筛选和/或负筛选。筛选还可能包括氨基酸渗透性的一次或多次改变，翻译效率的改变、翻译精确度的改变等。一般来说，要根据有机体的一个或多个基因上的突变来确定进行一次或多次改变，其中正交的 tRNA-tRNA 合成酶对被用于制备蛋白质。

制备 O-RS 以及改变合成酶的底物特异性的其他通用细节可参见名称为 Methods and compositions for the production of orthogonal tRNA-aminoacyl-tRNA synthetase paris” (用于制备正交 tRNA-氨酰-tRNA 合成酶对的方法和组合物)的国际专利申请 WO 2002/086075 和申请日为 2004 年 4 月 16 日的国际专利申请 PCT/US2004/011786。

来源和宿主有机体

本发明的翻译组分可来源于非真核有机体。例如，正交的 O-tRNA 可来源于非真核有机体(或有机体混合物)，例如古细菌，如詹氏甲烷球菌、热自养甲烷球菌，嗜盐菌如沃氏盐杆菌和盐杆菌 NRC-1，闪烁古生球菌，激烈火球菌，超嗜热古菌，Aeuropyrum pernix，海沼甲烷球菌，坎氏甲烷嗜热菌，马氏微球菌(Mm)，超耐高温热棒菌，火球菌(Pyroccocus abyssi)，硫磺矿硫化叶菌(Ss)，硫化叶菌(Sulfolobus tokodaii)，噬酸热原体，火山热原体等，或者真细菌，如大肠杆菌、嗜热肠热菌(Thermus thermophilus)、嗜热脂肪芽孢杆菌(Bacillus stearothermophilus)等，正交的 O-RS 也可以来源于非真核有机体(或有机体混合物)，例如古细菌，如詹氏甲烷球菌、热自养甲烷球菌，嗜盐菌属如沃氏盐杆菌和盐杆菌 NRC-1，闪烁古生球菌，激烈火球菌，超嗜热古菌，Aeuropyrum pernix，海沼甲烷球菌，坎氏甲烷嗜热菌，马氏微球菌，超耐高温热棒菌，火球菌(Pyroccocus abyssi)，硫磺矿硫化叶菌，硫化叶菌(Sulfolobus tokodaii)，噬酸热原体，火山热原体等，或者真细菌，如大肠杆菌、嗜热肠热菌、嗜热脂肪芽孢杆菌等。在一个实施方式中，真核有机体如植物、藻类、原生生物、真菌、酵母、动物(如哺乳动物、昆虫、节肢动物等)等也可作为 O-tRNA 和 O-RS 的来源。

O-tRNA/O-RS 对的各种组分可以来源于同一有机体，也可以来源于不同有机体。在一个实施方式中，O-tRNA/O-RS 对来源于同一有机体。另外 O-tRNA/O-RS 对的 O-tRNA 和 O-RS 也可以来源于不同有机体。在一个优选实施方式中，古细菌超嗜热古菌的赖氨酰合成酶/tRNA 对被用作大肠杆菌翻译系统正交对。如本文所描述，
这个正交对被修饰使其能够识别四碱基选择者密码子，也可以被修饰成能够用非天然氨基酸如高谷氨酰胺加载 O-tRNA 的形式。这个正交对（或其修饰形式）也可以与以前描述的正交对联合使用，如那些荚氏甲烷球菌来源的，经修饰后能识别终止选择者的密码子的正交对。这样就可以通过在目的翻译系统中加入包含两个或多个能被 O-tRNA/O-RS 对识别的选择者密码子的蛋白质编码核酸，从而使制备出的蛋白质含有两个不同的非天然氨基酸。

可在体内或体外，或者利用细胞，如非真核细胞或真核细胞，来筛选 O-tRNA、O-RS 或 O-tRNA/O-RS 对以制备出含高谷氨酰胺或其他非天然的目的氨基酸的多肽。非真核细胞可以是各种来源的，如真细菌，如大肠杆菌、嗜热栖热菌、嗜热脂肪芽孢杆菌等，或者古细菌，如詹氏甲烷球菌、热自养甲烷球菌，嗜盐菌属如沃氏盐杆菌和盐杆菌 NRC-1，闪烁古球菌，激烈火球菌，超嗜热古菌，Aeuropyrum pernix，海沼甲烷球菌，坎氏甲烷嗜热菌，马氏微球菌，超耐高温热棒菌，火球菌（Pyrococcus abyssii），硫磺矿硫化叶菌，硫化叶菌（Sulfolobus tokodaii），嗜酸热原体，火山热原体等。真核细胞也可以是各种来源的，如植物（如复杂植物，如单子叶植物或双子叶植物）、藻类、原生生物、真菌、酵母（如酿酒酵母）、动物（如哺乳动物、昆虫、节肢动物等）等。含本发明的翻译组分的细胞组合物也是本发明的特征。

筛选一个物种的 O-tRNA 和或 O-RS 用于另外一个物种的描述见申请日为 2004 年 4 月 16 日的国际专利申请 PCT/US2004/011786。

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的遗传密码框架。例如，选择者密码子包括独特的三碱基密码子，无义密码子如终止密码子，例如琥珀密码子（UAG）、或乳白密码子，一种非天然密码子，至少一个四碱基密码子（如 AGGA），一个稀有密码子等。不同数目的选择者密码子都可以被插入到目的基因内，如一个、两个、三个或三个以上等。通过使用不同的选择者密码子，可利用多个正交 tRNA/合成酶对同时将多个不同的非天然氨基酸同时引入到特异性位点上。

在一个实施方式中，本发明的方法包括利用终止密码子作为选择者密码子在体内将高谷氨酰胺引入到细胞内。例如，制备出可识别四碱基选择者密码子的 O-tRNA，然后被 O-RS 氨酰化加载上高谷氨酰胺。这个 O-tRNA 不被翻译系统内源性的氨酰-tRNA 合成酶所识别。可利用常规的位点特异性突变技术将选择者密码子引入到编码目的多肽的靶多核苷酸的特定位点上。参见 Sayers, J. R. 等，(1988)，"5′，3′
Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis." Nucleic Acids Res., 791-802. 当 O-RS, O-tRNA 和编码目的多肽的核酸在体内结合时，高谷氨酸酵胺因响应选择者密码子而被插入，制备出一个在特定位置上含高谷氨酸酵胺的多肽。

非天然氨基酸如高谷氨酸酵胺在体内的插入对宿主细胞可以不产生显著影响。例如，在非真核细胞如大肠杆菌内，由于终止选择者密码子 UAG 的抑制效率依赖于 O-tRNA 如琥珀抑制型 tRNA 和释放因子 1(RF1)(RF1 可结合 UAG 密码子，启动正在合成的肽从核糖体上释放下来)之间的竞争，因此可通过增加 O-tRNA 如抑制型 tRNA 的表达水平或者利用 RF1 缺陷株来调节抑制的效率。在真核细胞内，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)(eRF 可结合终止密码子，启动正在合成的肽从核糖体上释放下来)之间的竞争，因此可通过增加 O-tRNA 如抑制型 tRNA 的表达水平来调节抑制的效率。另外，也可以通过添加其他的化合物来调节释放因子的活性，例如还原剂二硫苏糖醇(DTT)。

非天然氨基酸，包括高谷氨酸酵胺，也可以被稀有密码子编码。例如，当体外蛋白合成反应体系内的精氨酸浓度下降时，稀有的精氨酸密码子 AGG 被证实可通过一个丙氨酸酰化的合成 tRNA 有效地将丙氨酸插入到多肽链内。见 Ma 等, Biochemistry, 32:7939 (1993)。在这种情况下，合成的 tRNA 可与天然的 tRNA_{Arg} 竞争，后者作用于稀有成分存在于大肠杆菌内。另外，某些有机体无法利用所有的三联密码子。藤黄微球菌(Micrococcus luteus)内的独立密码子 AGA 可被用于氨基酸在体外转录/翻译提取物内的插入。见 Kowal 和 Oliver, Nucl. Acid Res., 25:4685 (1997)。在体内利用这些稀有密码子可以制备出本发明的组分。


在一个给定的系统中，选择者密码子还可以是一个天然的三碱基密码子，其中内源性的翻译系统不能利用(或很少利用)这个天然的三碱基密码子。例如，缺乏能识别天然三碱基密码子的 tRNA 的系统和/或三碱基密码子是稀有密码子的系统。


一种翻译旁路系统也可用于将高谷氨酸酰胺或其他非天然氨基酸插入到目的多肽内。在翻译旁路系统中，一个不能被翻译成蛋白的大片段序列被插入到基因内。这个序列含有一个可作为提示符的结构，诱导核糖体跳过这个序列，在这个插入片段的下游恢复翻译。

**非天然氨基酸**

如本文所述，非天然氨基酸是指除硒氨基酸和/或吡咯赖氨酸和下列 20 个遗传密码编码的 α 氨基酸之外的所有氨基酸，修饰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、蛋氨酸、苯丙氨酸、脯氨酸、丝氨酸、苏氨酸、色氨酸、酪氨酸、缬氨酸。α 氨基酸的通用结构可用结构式 I 表示：

![结构式 I](image)

非天然氨基酸一般都具有结构式 I 的结构，其中R 基团代表 20 个天然氨基酸使用的基团之外的任何取代基。见 L. Stryer 编的《生物化学》(Biochemistry) (第三版，1988，Freeman and Company，New York) 以了解 20 个天然氨基酸的结构。需要注意
的是，本发明的非天然氨基酸也可以是上述 20 个 
α 氨基酸之外的天然化合物(或者人工制备的合成化合物)

一般来说，本发明的非天然氨基酸在侧链上与天然氨基酸是不同的，因此非天然
氨基酸可以与其他氨基酸形成酰胺键，形成的方式与天然蛋白中的酰胺键一样。
但是，非天然氨基酸含有将其与天然氨基酸区别开来的侧链基团。

在将非天然氨基酸插入蛋白质的过程中令我们非常感兴趣的是它有能力插入高
谷氨酰胺。对于其他的非天然氨基酸来说，结构式 I 中的 R 基团可以是烷基、芳基、
酰基、酮基、叠氮基、羟基、烯基、卤素-、酰肼、烯基、炔基、醚、疏醇、硒
基、磺酰基、硼酸基、磷酸基、磷、杂环、烯酮、亚胺、醛、酯、硫代酸、羟胺、
胺等，或上述基团的混合物。我们感兴趣的其他非天然氨基酸包括但不限于含光活
化交联物的氨基酸、自旋-标记的氨基酸、荧光标记的氨基酸、金属结合的氨基酸、
含金属的氨基酸、放射性氨基酸、含有新官能团的氨基酸、与其他分子共价或非共
价结合的氨基酸、光陷阱(photocaged)和/或可感光异构的(photoisomerizable)氨基酸、
含生物素或生物素类似物的氨基酸、含酮类的氨基酸、糖基化的氨基酸、含聚乙二
醇或聚醚的氨基酸、重原子取代氨基酸、可化学裂解或光裂解的氨基酸、与天然氨
基酸相比侧链延长的氨基酸(如超过约 5 个、约 10 个碳原子的聚醚或长链烃)、含碳
连接糖的氨基酸、具有氧化还原活性的氨基酸、含氨基硫代酸的氨基酸、以及含一
个或多个毒性基序的氨基酸。在某些实施方式中，非天然氨基酸含有光活化交联物。
在一个实施方式中，非天然氨基酸含有连接于氨基酸侧链上的糖基和/或其他糖类修
饰。

非天然氨基酸除了可以包含新型侧链以外，还可以包含经修饰的骨架结构，如
结构式 II 和 III 所示的结构:

```
II

III
```

式中，Z 一般为 OH、NH₂、SH、NH-R'或 S-R'; X 和 Y 可以相同或不同，一般
包括 S 或 O，R 和 R'可以相同，也可以不同，一般选自与上述结构式 I 所示的非天然
氨基酸的 R 基团相同的一组取代基团以及氢。例如，本发明的非天然氨基酸还可以在结构式 II 和 III 所示的氨基或羧基上含有取代基团。这一类型的非天然氨基酸包括而不限于带 20 个常见氨基酸相应的侧链或非天然侧链的 α-羟酸、α-硫代酸、α-氨基硫代羧酸盐。另外，α 碳原子上的取代基还包括 L,D 或 α-α-双取代基氨基酸，如 D-谷氨酸、D-丙氨酸、D-甲基-O-酯氨酸、氨基酸等。其他的结构形式包括环形氨基酸，如脯氨酸类似物以及 3,4,6,7,8 和 9 环脯氨酸类似物，β 和 γ 氨基酸，如取代的 β 丙氨酸和 γ-苯丙氨酸。本发明其他结构的非天然氨基酸包括同型-β-型结构，比如其中位于 α 碳原子处有一个亚甲基或氨基夹心，例如同型-β-酯氨酸、α-苯基-酯氨酸的同分异构体。如图所示：

许多非天然氨基酸来源于天然氨基酸，如酪氨酸、谷氨酰胺、苯丙氨酸等。例如，酯氨基酸类似物包括取代酯氨基酸、正取代酯氨基酸和偏取代酯氨基酸，其中取代酯氨基酸携带乙酰基、苯甲酰基、氨基、甲基、羧基和酰胺基等。另外，它也可以包含多取代芳环。本发明的谷氨酰胺类似物包括而不限于 α-羟基衍生物、γ-取代衍生物、环状衍生物和酰胺取代基谷氨酰胺衍生物。典型的苯丙氨酸类似物的例子包括而不限于 α-羟基苯丙氨酸、正取代苯丙氨酸和偏取代苯丙氨酸，其中的取代基包括羟基、甲酰基、甲基、丙酰基、酯基或酮基等。非天然氨基酸的特殊例子包括而不限于高谷氨酰胺、3,4-羟基-L-苯丙氨酸、p-乙酰基-L-苯丙氨酸、p-炔丙基苯丙氨酸、O-甲基-L-酯氨酸、L-3-(2-萘基)丙氨酸、3-甲基-苯丙氨酸、O-4-丙酰基-L-酯氨酸、4-丙基-L-酯氨酸、三-O-乙酰基-GlcNAc-丝氨酸、L-多巴、含氮苯丙氨酸、异丙基-L-苯丙氨酸、p-叠氮基-L-苯丙氨酸、p-酰基-L-苯丙氨酸、p-苯甲酰基-L-苯丙氨酸、L-硫醇氨酸、磷酰基丝氨酸、磷酰基酯氨基酸、p-碘-苯丙氨酸、p-溴苯丙氨酸、p-甲基-L-苯丙氨酸以及异丙基-L-苯丙氨酸等。各种非天然氨基酸的结构显示在题为“非天然氨基酸的体内插入”的 WO 2002/085923 的图 16, 17, 18, 19, 26 和 29 中。

非天然氨基酸的化学合成

上面所提到的许多非天然氨基酸都可以从 Sigma(USA) 或 Aldrich(Milwaukee, WI, USA) 等公司购买到。那些未商业化的非天然氨基酸可以利用文献中或本领域技术人员熟知的方法合成。见 Fessendon 和 Fessendon 的《有机化学》(Organic Chemistry)
非天然氨基酸的细胞摄取

一般来说，在设计和筛选非天然氨基酸，如用于插入到蛋白质中，的时候需要考虑的一个问题就是细胞对非天然氨基酸的摄取。例如，带高密度电荷的 α 氨基酸是不可能透过细胞膜的。天然氨基酸通过以蛋白质为基础的转运系统的收集进入到细胞内，但是这些转运系统一般都有不同程度的氨基酸特异性。有一种快速筛选方

### 非天然氨基酸的生物合成

在细胞内已经存在许多生物合成通路用于合成氨基酸和其他化合物。但是某种特定非天然氨基酸的生物合成方法在自然界中可能是不存在的，如在细胞内，本发明提供了这样的方法。例如，可以通过加入新的酶或者修饰现有的宿主细胞通路来设计出非天然氨基酸的生物合成通路。额外的新酶可以是天然的酶，也可以是经人工改造的酶。例如，p-氨基苯丙氨酸（如上述WO 2002/085923 的实施例中存在的氨基酸）的生物合成依赖于添加其他有机体来源的已知酶的混合物。可以通过用含这些酶的基因的质粒转化细胞而将这些基因导入到细胞内。这些基因在细胞内表达时，可以提供一个酶催化通路来合成所需要的化合物。可用于本发明的这类酶的典型例子见下面实施例的描述。这些额外的酶的序列可在 Genbank 中找到。人工修饰的酶也可以同样的方式加入到细胞内。在这种方式中，可以通过选择细胞内合成机器和细胞的来源来制备非天然氨基酸。


现在还有一些其他的技术可被用来改造用于制备目的化合物的有机体和代谢通

一般来说，无论用什么方法，用本发明的人工生物合成通路制备出的非天然氨基酸的浓度都足以满足蛋白质的生物合成，如细胞内的天然含量，但是其浓度还没有对细胞内的其他氨基酸的浓度产生显著影响，或者耗竭掉细胞内的资源。用这种方法在体内制备出的浓度一般约为 10 mM 到 0.05 mM。一旦细胞经过改造可以产生用于特异性通路和非天然氨基酸制备的酶以后，就需要进行体内筛选以进一步优化用于核糖体蛋白合成和细胞生长的非天然氨基酸的制备。

用于插入高谷氨酰胺的正交组分

本发明提供了用于在体内将高谷氨酰胺插入到至少合成的多肽链中以响应选择者密码子、编码密码子、义码密码子、四碱基或多碱基密码子等的正交组分的制备方法和组合物。例如，本发明提供了正交-tRNA(O-tRNA)、正交氨酰-tRNA 合成酶(O-RS)及其正交对。这些正交对可用于将高谷氨酰胺插入到正在合成的多肽链中。

本发明的组合物包括正交氨酰-tRNA 合成酶(O-RS)，其中的 O-RS 优先用高谷氨酰胺酰化 O-tRNA。在某些实施方式中，O-RS 含有携带 SEQ ID NO:1 或其保守变体的氨基酸序列。在本发明的某些实施方式中，O-RS 优先氨酰化 O-tRNA，其效率至少是含氨基酸序列 SEQ ID NO:1 的多肽的 50%。

含有 O-RS 的组合物还可以含有正交 tRNA(O-tRNA)，其中的 O-tRNA 可识别选择者密码子。一般来说，与包含或编码于本文序列表和实施例所列的多核苷酸序列的 O-tRNA 相比，在同源合成酶存在的条件下，本发明的 O-tRNA 为了响应选择者密码子而产生的抑制效率至少可达 45%、50%、60%、75%、80%、90%或更高。在一个实施方式中，O-RS 联合 O-tRNA 所产生的抑制效率至少比缺少 O-RS 的 O-tRNA 所产生的抑制效率高 5 倍、10 倍、15 倍、20 倍、25 倍或更高。一方面，O-RS 联合 O-tRNA 所产生的抑制效率至少是乙烷烷球菌来源的正交氨酰-tRNA 合成酶
对所产生的抑制效率的 45%。

包含 O-tRNA 的组合物还可以包含细胞(例如非真核细胞如大肠杆菌等，或真核细胞)和/或翻译系统。

包含翻译系统的细胞(如非真核细胞或真核细胞)也是本发明的内容。其中的翻译系统包括正交-tRNA(O-tRNA)、正交氨酰-tRNA合成酶(O-RS)和高谷氨酰胺。一般来说，O-RS 优先氨酰化 O-tRNA，其效率至少是含氨基酸序列 SEQ ID NO:1 多肽的 50%。O-tRNA 可识别第一选择者密码子，O-RS 优选用高谷氨酰胺氨基酰化 O-tRNA。在一个实施方式中，O-tRNA 含有 SEQ ID NO:2 所示的序列或其互补多核苷酸序列，或者被该序列编码。在一个实施方式中，O-RS 含有任何一个 SEQ ID NO:1 或其保守变体所示的氨基酸序列。

本发明的细胞还可以包含其他的 O-tRNA/O-RS 对和第二种非天然氨基酸，例如，其中的这个 O-tRNA 可识别第二选择者密码子，这个 O-RS 优选用第二种非天然氨基酸氨基酰化 O-tRNA。另外，本发明的细胞还可以包含含有目的多肽编码多核苷酸的核酸，其中的多核苷酸含有能被 O-tRNA 识别的选择者密码子。

在某些实施方式中，本发明的细胞包括大肠杆菌细胞，细胞内含有正交-tRNA(O-tRNA)、正交氨酰-tRNA合成酶(O-RS)、高谷氨酰胺和含有目的多肽编码多核苷酸的核酸，其中的多核苷酸含有能被 O-tRNA 识别的选择者密码子。

在本发明的某些实施方式中，本发明的 O-tRNA 含有本文列表中和实施例中所述的多核苷酸序列或其互补的多核苷酸序列，或者被这些序列编码。在本发明的某些实施方式中，O-RS 含有序列中所示的氨基酸序列或其保守变体。在一个实施方式中，O-RS 或其片段被编码本文列表中和实施例中所示的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 可来源于各种有机体(如真核和/或非真核有机体)。

多核苷酸也是本发明的特征。本发明的多核苷酸包括人工的(如人造的和非天然产生的)多核苷酸，其中含有编码本文列表所示多肽的核苷酸序列，和/或与该多核苷酸序列互补的多核苷酸序列。本发明的多核苷酸还包括在较严格的条件下能与上述多核苷酸在核酸全长上杂交的核酸。本发明的多核苷酸还包括与天然 tRNA 及其相应编码核酸的序列具有至少 75%、80%、90%、95%、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 及其相应的编码核酸)，其中的 tRNA 可识别选择者密码子，如四碱基密码子。与上述任何多核苷酸序列至少有 80%、90%、95%、98%或更高相同性的人工多核苷酸序列，和/或含上述序列保守变体的多
核苷酸也包括在本发明的多核苷酸中。

含本发明的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。含本发明的载体的细胞也是本发明的特征。

O-tRNA/O-RS 对的组分的制备方法也是本发明的特征。利用这些方法制备出的组分也是本发明的特征。例如，至少制备一个与细胞正交的 tRNA(O-tRNA)的方法包括制备突变的 tRNA 文库；突变 tRNA 突变文库内每一个成员的反密码子环使其能识别选择者密码子，从而形成一个 O-tRNA 文库，负筛选第一物种的第一细胞群，其中的细胞含有 O-tRNA 文库的成员。负筛选可去除含 O-tRNA 文库的成员的细胞，其中的 O-tRNA 可被细胞内源性的氨酰-tRNA 合成酶(RS)氨酰化。这样就形成一个与第一物种的细胞正交的 tRNA 库，从而至少可以获得一个 O-tRNA。利用本发明的方法制备出的 O-tRNA 也包括在本发明之内。

在某些实施方式中，本发明的方法还包括正筛选第一物种的第二细胞群，其中的细胞含有与第一物种的细胞正交的 tRNA 库的成员，同源的氨酰-tRNA 合成酶和正选择标记。通过正筛选筛选出的那些细胞含有能被同源氨酰-tRNA 合成酶氨酰化的 tRNA 库的成员，在正选择标记存在的情况下可出现预期的反应，从而可以获得 O-tRNA。在某些实施方式中，第二细胞群含有没有被负筛选去除的细胞。

本发明还提供了能将高谷氨酰胺加到 O-tRNA 上的正交-氨酰-tRNA 合成酶的鉴定方法。例如，该方法包括筛选第一物种的细胞群，其中的细胞都含有 1) 一组氨酰-tRNA 合成酶(RS)的成员(如一组 RS 可包括突变的 RS、第一物种之外的物种来源的 RS、或者二者都包括)；2) 正交-tRNA(O-tRNA)(如来源于一个或多个物种)；以及 3) 编码正选择标记并且至少包含一个选择者密码子的多核苷酸。

筛选细胞(如宿主细胞)以挑出与不含 RS 群的成员或含量较少的细胞相比抑制效率升高的那些细胞。这些筛选出的细胞含有能氨酰化 O-tRNA 的有活性的 RS。利用这种方法鉴定出的正交氨酰-tRNA 合成酶也是本发明的特征。

在细胞(例如非真核细胞如大肠杆菌细胞等，或者真核细胞)内制备特定位点上为高谷氨酰胺的蛋白质的方法也是本发明的特征。例如，一种方法是在适当的培养基内培养细胞，其中的细胞含有核酸，该核酸至少含有一个选择者密码子，编码一个蛋白，加入高谷氨酰胺，在至少含有一个选择者密码子的核酸的翻译过程中高谷氨酰胺被插入到蛋白质的特定位点上，生成蛋白质。细胞还含有：在细胞内有功能并且能识别选择者密码子的正交-tRNA(O-tRNA)；以及优先用高谷氨酰胺氨酰化 O-tRNA 的正交氨酰-tRNA 合成酶(O-RS)。利用这种方法制备的蛋白质也是本发明的
特征。

本发明还提供了包含蛋白质的组合物，其中的蛋白质含有高谷氨酰胺。在某些
实施方式中，蛋白质的氨基酸序列与已知蛋白如治疗性蛋白、诊断蛋白、工业用酶
的氨基酸序列或其部分至少75%相似性。另外，组合物还可以包含药学上可接受的
载体。

核酸和多肽序列及其变体

如上文和下文所描述，本发明提供了核酸的多核苷酸序列，如O-tRNA和O-RS，
多肽的氨基酸序列，如O-RS，以及包含所述序列的组合物、系统和方法。所述序列
的例子如O-tRNA和O-RS也在本文中列出(见本文的序列表和实施例)。但是，本领
域的技术人员应该了解本发明并不仅仅局限于这些序列，如实施例和列表中的序列。
本领域的技术人员应当了解本发明还包括具有本文所描述的功能如编码合适
O-tRNA或O-RS的许多相关和不相关序列。

本发明提供了多肽(O-RS)和多核苷酸，如O-tRNA、编码O-RS或其片段的多核
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那些
编码本发明的目的蛋白或多肽的多核苷酸，其中的蛋白或多肽含有一个或多个选择
者密码子。另外，本发明的多核苷酸还包括含有序列表中所列核苷酸序列的多核苷
酸；与多核苷酸序列互补的多核苷酸或编码多核苷酸序列的多核苷酸。本发明的
多核苷酸还包括编码本发明的多肽的多核苷酸。同样，在高严格条件下，可与上
述多核苷酸在全长核酸序列上杂交的人工核酸也包括在本发明的多核苷酸范围内。
在一个实施方式中，组合物包含本发明的多肽和各种赋形剂(如缓冲液、水、药学上
可接受的赋形剂等)。本发明还提供了可与本发明的多肽发生特异性免疫反应的抗体
或抗血清。人工多核苷酸是由人制备的，不是天然发生的。

本发明的多核苷酸还包括人工的多核苷酸，这种多核苷酸与天然tRNA、本文列
表或实施例中的任何tRNA或其编码核酸的多核苷酸(但是天然tRNA除外)至少有
75%、80%、90%、95%、98%或更高的相似性。多核苷酸还包括与天然tRNA的多
核苷酸至少有75%、80%、90%、95%、98%或更高相似性的人工多核苷酸。

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，载体是表达载体。在另一个实施方式中，表达载体包含与
本发明的一个或多个多核苷酸功能相连的启动子。在另一个实施方式中，细胞包含
携带本发明的多核苷酸的载体。
本领域的技术人员还应该理解的是本发明还包括公开序列的多种变体。例如，可形成功能相似序列的公开序列的保守变体也包含在本发明的范围内。核酸多核苷酸序列的变体也被认为包括在本发明的范围之内，其中的变体可与至少一个公开序列杂交，并且能识别选择者密码子。本文公开的序列的独特序列，例如通过标准序列比较技术确定的独特序列，也包括在本发明的范围之内。

保守变体

由于遗传密码的简并性，“静默替代”（即核酸序列上的替代不会导致其编码多肽的改变）是每个编码氨基酸的核酸的隐含特性。同样，氨基酸序列上的一个或几个氨基酸发生“保守氨基酸替代”是指用特性高度相似的不同氨基酸来替代，也很容易被鉴定为与公开的构建体高度相似。每一个公开序列的这种保守变体也是本发明的特征。

特定核酸序列的“保守变体”是指那些编码一样或基本一样的氨基酸序列的核酸，或者是指虽然不编码氨基酸序列，但序列基本相同的核酸。本领域的技术人员应该认识到导致编码序列上单个氨基酸或一小部分氨基酸（一般少于5%，更常见的是少于4%、2%或1%）发生改变、添加或删除的单个替代、缺失或插入是“保守修饰的变体”，其中的改变可导致氨基酸的删除、氨基酸的添加或以化学相似的氨基酸替代原来的氨基酸。因此，本发明所列出的多肽序列的“保守变体”包括用具有相同保守取代基的氨基酸替代一小部分氨基酸，一般低于多肽氨基酸总数的5%，更常见的是低于2%或1%。最后，不改变核酸分子编码活性的序列添加，如添加非功能序列，被称为基础氨基酸的保守变体。

可提供功能相似氨基酸的保守替换表是本领域熟知的，其中一个氨基酸残基被另一个具有相似化学特性（芳香族侧链或带正电荷的侧链）的氨基酸替代，因此不会在实质上改变多肽分子的功能。下面列出了几组具有相似化学特性的天然氨基酸，其中每一组内的互相替换都是“保守替代”。

<table>
<thead>
<tr>
<th>非极性/或脂肪族侧链</th>
<th>极性不带电荷的侧链</th>
<th>芳香族侧链</th>
<th>带正电荷的侧链</th>
<th>带负电荷的侧链</th>
</tr>
</thead>
<tbody>
<tr>
<td>甘氨酸</td>
<td>丝氨酸</td>
<td>苯丙氨酸</td>
<td>赖氨酸</td>
<td>天冬氨酸</td>
</tr>
<tr>
<td>丙氨酸</td>
<td>苏氨酸</td>
<td>酪氨酸</td>
<td>精氨酸</td>
<td>谷氨酸</td>
</tr>
<tr>
<td>缬氨酸</td>
<td>半胱氨酸</td>
<td>色氨酸</td>
<td>组氨酸</td>
<td></td>
</tr>
<tr>
<td>亮氨酸</td>
<td>蛋氨酸</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>异亮氨酸</td>
<td>天冬酰胺</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>脯氨酸</td>
<td>谷氨酰胺</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
核酸杂交

对比杂交可用于鉴定本发明的核酸，如本文序列表中所列的那些核酸，包括本发明的核酸的保守变体，这种对比杂交方法是一种将本发明的核酸与不相关核酸区别开来的方法。另外，能与序列表中所列的那些核酸在高、超高、超超高严格条件下杂交的靶核酸也是本发明的特征。这类核酸的典型例子包括那些与一个给定核酸序列相比发生一个或几个静默核酸替代或保守核酸替代的核酸。

如果一个被检测核酸被说成可与一个核酸探针特异性杂交，那么这个被检测核酸与探针的匹配程度至少相当于与完全匹配的互补靶核酸匹配程度的1/2，即信噪比至少是探针与靶核酸杂交的1/2，杂交的条件为完全匹配的探针与完全匹配的互补靶核酸结合所产生的信噪比至少等于与任何不匹配的靶核酸杂交所产信生的信噪比的约5倍到10倍。


在 DNA 印迹或 RNA 印迹中，含有超过 100 个互补残基的互补核酸杂交到滤膜上的一个典型杂交条件是 50% 甲醇，1 mg 肝素，42°C 杂交过夜。严格洗涤条件的例子是 65°C 用 0.2×SSC 洗涤 15 分钟(见上述文献 Sambrook 有关 SSC 缓冲液的描述)。一般情况下先用低严格条件再用高严格条件洗涤以去除背景探针信号。低严格洗涤条件的一个例子是 40°C 用 2×SSC 洗涤 15 分钟。如果所检测到的信噪比比不相关探针在特定杂交试验中所产生的信噪比高 5 倍(或更高)，就说明这是一个特异性杂交。

“严格的杂交洗涤条件”用于核酸杂交试验如 DNA 印迹和 RNA 印迹时是指序列依赖性的，但是在不同的环境参数下该条件是不同的。较全面的有关核酸杂交的指导手册是上述的 Tijssen (1993)和 Hames 和 Higgins，1 和 2。任何被检测核酸所用
的杂交和洗涤条件的严格程度可以根据经验很容易地确定。例如，在确定杂交和洗涤条件的严格程度时，可以逐渐提高杂交和洗涤条件的严格程度（如，在杂交或洗涤过程中通过升高温度、降低盐浓度、增加洗涤剂的浓度和/或增加有机溶剂如甲醇的浓度），直到满足了所选择的一组标准。例如，在高压杂交和洗涤条件下，逐渐增加杂交和洗涤条件直到探针与互补靶核酸完全匹配，所产生的信噪比至少是探针与不匹配靶核酸杂交所产生的信噪比的 5 倍。

对于一个特定探针来说，十分严格的条件等同于其热熔点（T_m）。T_m 是指 50%的被检测核酸与完全匹配的探针杂交时的温度。一般来说，在一定的离子强度和 pH 条件下，本发明所选择的“高严格”杂交和洗涤条件比一个特定序列的 T_m 约低 5℃。

“超高严格程度”的杂交和洗涤条件是指杂交和洗涤条件的严格程度逐渐升高，直到探针与完全匹配的互补靶核酸结合所产生的信噪比至少达到探针与不匹配核酸杂交所产生的信噪比的 10 倍。靶核酸在这种条件下与探针杂交，如果所产生的信噪比至少是完全匹配的互补靶核酸所产生的信噪比的 1/2，那么就可以说靶核酸可以在超高严格条件下与探针结合。

同样可以通过逐渐提高预杂交试验的杂交和/或洗涤条件的严格程度来确定更高严格程度的杂交条件。例如，逐渐升高杂交和洗涤条件的严格程度，直到探针与完全匹配的互补靶核酸结合所产生的信噪比至少达到探针与不匹配核酸杂交所产生的信噪比的 10 倍、20 倍、50 倍、100 倍、500 倍或更高。靶核酸在这种条件下与探针杂交，如果所产生的信噪比至少是完全匹配的互补靶核酸所产生的信噪比的 1/2，那么就可以说靶核酸可以在超超严格条件下与探针结合。

如果核酸编码的多肽是基本相同的，那么即使核酸不能彼此杂交，这两个核酸也是基本相同的。这种情况发生在核酸的拷贝是充分利用遗传密码所允许的最大简并性来制备的时候。

**独特的亚序列**

一方面，本发明提供了含独特亚序列的核酸，这些亚序列位于选自本文所描述的 O-tRNA 和 O-RS 序列的核酸内。与任何已知的 tRNA 或 RS 核酸序列相应的核酸相比，独特亚序列都是独特的。可利用 BLAST 进行核酸的排列，使用默认参数。任何独特亚序列都可作为探针用于鉴定本发明的核酸。

同样，本发明还包括含独特亚序列的多肽，这些独特亚序列位于选自本文所描述的 O-RS 序列的多肽内。在这里，与任何已知的 RS 序列相应的多肽相比，独特亚序列都是独特的。
本发明还提供了能在严格条件下与独特编码寡核苷酸杂交的靶核酸，该寡核苷酸编码选自 O-RS 序列的多肽内的独特亚序列，其中的独特亚序列与任何对照多肽(如本发明的合成酶来源的亲本序列，如通过突变获得)相应的多肽相比都是独特的。独特序列可通过上述方法确定。

序列比较、相同性和同源性

术语“相同的”或“百分相同性”用于两个或多个核酸或多肽序列的比较时是指两个或多个序列或亚序列在一起排列并比较其最大相似性时，这两个或多个序列或亚序列是相同的，或者有特定百分比的氨基酸残基或核苷酸是相同的，从而利用下面所描述的一种序列比较算法(或本领域技术人员所熟知的其他算法)或通过肉眼观察来完成。

术语“基本相同的”用于两个或多个核酸或多肽(如编码 O-trNA 或 O-RS 的 DNA，O-RS 的氨基酸序列)的比较时是指两个或多个序列或亚序列在一起排列并比较其最大相似性时，这两个或多个序列或亚序列至少有约 60%、80%、90-95%、98%、99%或更多的核苷酸或氨基酸残基是相同的，这种比较可利用序列比较算法或通过肉眼观察来完成。无论实际来源于何种家系，这种“基本相同的”序列一般被认为是“同源的”。“实质相同性”是指在序列的一个区域内的比较，这个区域至少约含 50 个残基，更优选的是至少约含 100 个残基，最优选的是至少约含 150 个残基，或者在两个序列的全长上进行比较。

当蛋白和/或蛋白序列来源于同一个亲本蛋白或蛋白序列时，不论是自然获得的还是经过改造的，这些蛋白和/或蛋白序列都是“同源的”。同样，当核酸和/或核酸序列来源于同一个亲本核酸或核酸序列时，不论是自然获得的还是经过改造的，这些核酸和/或核酸序列都是“同源的”。例如，任何天然核酸都可以用现有的突变方法加以修饰使其包含一个或多个选择者密码子。当这个突变的核酸被表达时，就编码出一条含一个或多个非天然氨基酸，如谷氨酰胺的多肽。当然，突变过程也可以改变一个或多个标准密码子，从而使其编码的突变蛋白上产生一个或多个标准氨基酸的改变。同源性一般是根据两个或多个核酸或蛋白(或其序列)的序列相同性推算出来的。可用于推算同源性的序列之间的精确百分相同性根据所要比较的蛋白和核酸的不同而不同，但是一般来说至少要有 25%的序列相同性才能建立同源性。更高的序列类似性，如 30%、40%、50%、60%、70%、80%、90%、95%、99%或更高，也可用于建立同源性。确定序列相同性百分比的方法(如 BLASTP 和 BLASTN，使用默认参数)见本文的描述，这些方法都是现有的。
为了进行序列比较并确定其同源性，一般需要把一个序列看作参考序列，被检测序列与之比较。在使用序列比较算法时，将被测序列与参考序列输入到计算机内，如果需要，设定亚序列匹配值和序列算法程序的参数。序列比较算法就可以根据设定的程序参数计算出被测序列相对于参考序列的百分序列相同性。


得到的最小总概率约在 0.1 以下、优选约 0.01 以下、最优选的是约 0.001 以下，则这个核酸被认为与参考序列是相似的。

### 突变和其他分子生物学技术


各种突变技术都可用于本发明中进行 tRNA 分子的突变、tRNA 文库的制备、合成酶文库的制备、编码高谷氨酰胺的选择者密码子向目的蛋白或多肽内的插入。这些技术包括但不限于定点定向突变、随机点突变、同源重组、DNA 重组或其他的回归突变方法、嵌合构建体、利用含尿嘧啶的模板进行的突变、寡核苷酸定向突变、硫代磷酸修饰的 DNA 突变、利用带缺口的双链 DNA 进行的突变等。其他合适的方法包括点突变修复、利用修复缺陷宿主系进行的突变、限制性筛选和限制性纯化、缺失突变、通过总基因合成进行的突变、双链断裂修复等。利用嵌合构建体进行的突变也包括在本发明中。在一个实施方式中，可以利用已知的有关天然分子或者改变的或突变的天然分子的信息指导突变，如序列、序列比较、物理特性、晶体结构等。

宿主细胞可用本发明的多核苷酸或包含本发明多核苷酸的构建体进行遗传改造（如转化、转导或转染），如本发明的载体，这些载体可以是克隆载体，也可以是表达载体。例如，正交 tRNA、正交 tRNA 合成酶及其衍生的蛋白质的编码区可以被可操控地连接到能在给定宿主细胞内发挥功能的基因表达调节元件上。典型的载体都包含转录和翻译终止子、转录和翻译起始序列和用于调节特定靶核酸表达的启动子。载体还可以包含基因表达盒，其中至少含有一个独立的终止子序列，允许表达盒在真核或原核宿主、或者二者内(如穿梭载体)复制的序列以及用于原核和真核系统内的


目的蛋白和多肽
至少含有一个高谷氨酰胺的目的蛋白或多肽是本发明的特征，含有两个或多个
不同的非天然氨基酸的多肽也是本发明的特征。赋形剂（如药学上可接受的赋形剂）也可以和蛋白质一起存在。另外，本发明的蛋白还可以进行翻译后修饰。

在细胞内制备特定位置上为高谷氨酰胺或其它非天然氨基酸的蛋白质的方法也是本发明的特征。例如，一种方法是在适宜的培养基内培养细胞，其中的细胞含有至少包含一个选择者密码子并能编码一个蛋白质的核酸；提供高谷氨酰胺或其他非天然氨基酸；其中的细胞还包含：在细胞内有功能并能识别选择者密码子的正交-tRNA(O-tRNA)和优先用高谷氨酰胺或其他非天然氨基酸氨酰化 O-tRNA 的正交氨酰-tRNA 合成酶(O-RS)。在某些实施方式中，与包含或编码于本系列列表和实施例所列的多核苷酸序列的 O-tRNA 相比，在同源合成酶存在的条件下，这种 O-tRNA 为了响应选择者密码子而产生的抑制效率至少可达 45%、50%、60%、75%、80%、90% 或更高。利用这种方法制备的蛋白也是本发明的特征。

本发明还提供了含有蛋白的组合物，其中的蛋白包含高谷氨酰胺。在某些实施方式中，蛋白所包含的氨基酸序列与靶蛋白如治疗性蛋白、诊断蛋白、工业用蛋白或其片段的序列至少有 75% 的相同性，如那些因为插入了一个或多个非天然氨基酸如高谷氨酰胺与靶蛋白不同的蛋白质。


利用本发明的细胞可以大规模地合成含非天然氨基酸的蛋白质。一方面，组合物可以包含至少 10μg、50μg、75μg、100μg、200μg、250μg、500μg、1mg、10mg 或更多含高谷氨酰胺或多种非天然氨基酸的蛋白质，或者含有蛋白质的量能够达到体内蛋白合成方法的要求(本文提供了重组蛋白制备和纯化方法的详细描述)。另一方面，组合物中所含的蛋白质浓度可以是每升细胞裂解物、缓冲液、药学上可接受的缓冲液或其他液体溶液中至少含 10μg、50μg、75μg、100μg、200μg、250μg、500μg、
1mg、10mg 或更高(如以1mL到100L的体积)。在细胞内大量制备(例如大于其他方法，如体外翻译，在一般情况下所能达到的规模)至少含一个高谷氨酰胺的蛋白质也是本发明的特征。

插入高谷氨酰胺或其他非天然氨基酸可导致蛋白质结构和/或功能的变化，如大小、酸性、亲核性、氢键、疏水性、蛋白酶靶点的易接近性、对基序的靶向性(如用于蛋白芯片时)等。含有高谷氨酰胺的蛋白质的催化活性或物理特性可能会得以改善，或者获得全新的催化活性或物理特性。例如，下列特性可通过在蛋白质内插入高谷氨酰胺或其他非天然氨基酸加以改变：毒性、生物分布、结构特性、分光光度特性、化学和/或光化学特性、催化能力、半衰期(如血清半衰期)、与其他分子相互作用的能力(共价的或非共价的)。包含至少携带一个高谷氨酰胺的蛋白质的组合物可用作新型的治疗蛋白、诊断蛋白、催化酶、工业用酶、结合蛋白(如抗体)以及可用于蛋白质结构和功能的研究。见 Dougherty，(2000) Unnatural Amino Acids as Probes of Protein Structure and Function, Current Opinion in Chemical Biology, 4:645-652。另外，一个或多个非天然氨基酸可以被插入到多肽内以形成分子标签，如用于将多肽固定到固体支持物上。见 Wang 和 Schultz 的专利申请“蛋白质芯片”，申请日为2003年12月22日，代理人备案号为54-000810PC，参见其中的“利用含非天然氨基酸的多肽制备芯片的方法”的延伸讨论。

在本发明的一个方面，组合物包含至少一个蛋白质，其中的蛋白质至少含有1、2、3、4、5、6、7、8、9、10 或更多个非天然氨基酸，如高谷氨酰胺或其他非天然氨基酸。这些非天然氨基酸可以是相同的，也可以是不同的，如在蛋白质的1、2、3、4、5、6、7、8、9、10 或更多个不同的位点上含有1、2、3、4、5、6、7、8、9、10 或更多个不同的非天然氨基酸。另一方面，组合物包含一个蛋白质，这个蛋白质至少有一个但不是全部的特定氨基酸被高谷氨酰胺替代。对于一个给定的含多个非天然氨基酸的蛋白质来说，其中的非天然氨基酸可以是相同的，也可以是不同的(如蛋白质包含两个或多个不同类型的非天然氨基酸，或者包含两个相同的非天然氨基酸)。对于一个给定的含两个以上非天然氨基酸的蛋白质来说，其中的非天然氨基酸可以是相同的，也可以是不同的，或者其中有多个非天然氨基酸是相同的，但是至少有一个不同非天然氨基酸。

一般来说，包含非天然氨基酸如高谷氨酰胺的任何蛋白质(或其片段)，或者编码多个不同非天然氨基酸的任何蛋白(及其相应的编码核酸，如包含一个或多个选择者密码子的核酸)都可用本发明的组合物和方法制备。无需将已知的成千上万个蛋白质
都一一鉴定，其中的任何一个都可以被修饰成含一个或多个非天然氨基酸的形式，如通过任何已知的突变方法将一个或多个适宜的选择者密码子插入到相关的翻译系统内。已知蛋白序列的常用数据库包括 GenBank EMBL、DDBJ 和 NCBI。其他数据库通过搜索互联网也可以很容易地找到。

一般来说，本发明的蛋白质与己知的蛋白，如治疗性蛋白、诊断蛋白、工业用蛋白，或其片段等，至少有 60%、70%、75%、80%、90%、95%、99% 或更高的相同性，其中包含一个或多个非天然氨基酸。可被修饰使其包含一个或多个高谷氨酰胺的治疗性蛋白，诊断蛋白以及其他蛋白可在下列文献中找到，但是又不仅仅局限于这些文献：2004 年 4 月 16 日提交的题为“Expanding the Eukaryotic Genetic Code”(扩充真核遗传密码)的国际专利申请 PCT/US2004/011786；以及WO 2002/085923，题为“非天然氨基酸的体内插入”。可被修饰使其包含一个或多个高谷氨酰胺的治疗性蛋白，诊断蛋白以及其他蛋白包括但不限于α-1 抗胰蛋白酶、血管性因子、抗凝血因子、抗体 (有关抗体的详细描述见下文)、载脂蛋白、脱辅基蛋白、心房利钠因子、心房利钠多肽、心房肽、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、单核细胞炎症蛋白-1 α、单核细胞炎症蛋白-1 β、RANTES、I309、R83915、R91733、HCC1、T58847、D31065、T64262)、CD40 配基、C-kit 配基、胶原蛋白、集落刺激因子 (CSF)、补体因子 5a、补体抑制因子、补体受体 1、细胞因子 (如上皮细胞中性粒细胞活化肽-78、GROα/MGSA、GROβ、GROγ、MIP-1α、MIP-1β、MCP-1 表皮生长因子 (EGF)、红细胞生成素 (EPO)、剥脱因子 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)、纤维蛋白原、纤溶酶、G-CSF、GM-CSF、葡萄糖脑苷脂酶、绒毛膜促性腺激素、生长因子、Hedgehog 蛋白 (如 Sonic、Indian、Desert)、血红素、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)、白细胞介素 (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)、角质化细胞生长因子 (KGF)、乳铁蛋白、白血病抑制因子、荧光素酶、Neurturin、中性粒细胞抑制因子 (NIF)、抑制素 M、骨形成蛋白、甲状旁腺激素、PD-ECSF、PDGF、肽类激素 (如人生长激素、Pleiotropin、蛋白质 A、蛋白质 G、发热外毒素 A、B 和 C、松弛素、肾素、SCF、可溶性补体受体 1、可溶性 I-CAM1、可溶性白细胞介素受体 (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、生长抑素、生长激素、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)、毒性性休克综合症毒素 TSST-1)、胸腺激素 α1、组织纤溶酶原激活因子、肿瘤杀死因子 β(TNFβ)、肿瘤杀死因子受体(TNFR)、肿瘤杀死因子-α(TNFα)、血管内皮生长因子(VEGF)、尿激酶等。

可以利用本发明的组合物和方法在体内制备出的含本文所描述的高谷氨酰胺的一类蛋白质是转录调节因子或其片段。转录调节因子的例子包括能调节细胞生长、分化、调控等的基因和转录调节蛋白。转录调节因子存在于原核细胞、病毒和真核生物如真菌、植物、酵母、昆虫和动物如哺乳动物，提供了一个大范围的治疗靶位。应当理解的是表达和转录活化因子调节转录的机制很多，如与受体结合、刺激信号转导级联、调节转录因子的表达、与启动子和增强子结合、与启动子和增强子结合的蛋白结合、解链 DNA、剪接 mRNA 前体、多聚腺苷酸化 RNA 和降解 RNA。

本发明的一类蛋白质(如携带一个或多个高谷氨酰胺的蛋白质)是表达活化因子，包括细胞因子、炎症分子、生长因子及其受体、以及癌基因产物，如白细胞介素(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/Kit、CD40L/CD40、VLA-4/VCAM-1、ICAM-1/LFA-1 和 hyaluron/CD44；信号转导分子及其相应的癌基因产物，如 Mos、Ras、Raf 和 Met；以及转录活化因子和抑制因子，如 p53、Tat、Fos、Myc、Jun、Myb、Rel，以及类固醇激素受体，如雌激素、孕酮、睾丸酮、醛固酮、LDL 受体的配基和皮质酮。

本发明还提供了至少携带一个同型高谷酰胺的酶(如工业用酶)或其片段。酶的例子包括但不限于酰胺酶、氨基酸消旋酶、酰基转移酶、脱氢酶、二氢酶、二芳基丙烷过氧化物酶、表位酶、环氧化物水解酶、酯酶、异构酶、激酶、葡萄糖异构酶、糖苷酶、葡萄糖酰基转移酶、卤素过氧化物酶(haloperoxidase)、单加氧酶(如 p450)、脂肪酶、木素过氧化物酶、脂水合酶、脂水解酶、蛋白酶、磷酸酯酶、枯草杆菌蛋白酶、转氨酶和核酶。

这些蛋白中有许多是可以在公司购买到的(见 Sigma BioSciences 的 2003 年产品目录和价目表)，其相应的蛋白序列和基因及其变体通常都是已知的(见 Genbank)。它们都可以按照本发明的方法通过插入一个或多个高谷氨酰胺或其他非天然氨基酸来加以修饰，例如改变蛋白质的一种或多种治疗、诊断或酶催化特性。与治疗相关的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(如通过在非天然氨基酸如高谷氨酰胺上插入报告基团(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。诊断特性包括
保存半衰期、稳定性、诊断活性、可检测性等。相关的酶催化特性包括保存半衰期、
稳定性、酶活性、生产能力等。

其他的多种蛋白也可加以修饰使其包含一个或多个本发明的高谷氨酰胺或其他
非天然氨基酸。例如，本发明包含的蛋白质可以是用高谷氨酰胺替代一个或多个天
然氨基酸的疫苗蛋白，这些蛋白来自于感染性真菌，如曲霉属、念珠菌属的真菌；
细菌，特别是大肠杆菌，可以作为致病菌的模型，和医学上有重要影响的细菌如葡萄
球菌属(如金黄色葡萄球菌)或链球菌属(如肺炎链球菌)的细菌；原虫如孢子虫(如
刚地弓形虫)、根足虫(如内阿米巴原虫)和鞭毛虫(锥形虫、利什曼原虫、毛滴虫、鞭
毛虫等)；病毒如(+R)NA病毒(例如包括痘病毒类如痘苗病毒；细小核糖核酸病毒如
脊髓灰质炎病毒；披膜病毒如风疹病毒；黄病毒如HCV；以及冠状病毒)，(-)RNA
病毒(如弹状病毒如 VS；副黏液病毒如RSV；正粘病毒如流感病毒；布亚病毒；以
及沙粒病毒)，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，以及某些DNA到RNA的病毒如乙肝病毒。

与农业有关的蛋白质如昆虫抗性蛋白(如Cry蛋白)、淀粉和脂质生成酶、植物和
昆虫毒素、毒素抗性蛋白、霉菌毒素解毒蛋白、植物生长酶(如糖醇糖1,5-二磷酸
羧化酶/合酶，"RUBISCO")、脂肪加氧酶(LOX)、以及磷酸烯醇丙酮酸(PEP)羧化酶也
是高谷氨酰胺或其他非天然氨基酸修饰的适宜靶位。

在某些实施方式中，本发明的方法和/或组合物内的目的蛋白或多肽(或其片段)
是由核酸编码的。一般来说，这种核酸至少含有1,2,3,4,5,6,7,8,9,10
或更多个选择者密码子。

目的蛋白或多肽的编码基因可用本领域技术人员熟知的方法以及本文"突变和
其他分子生物学技术"部分描述的方法加以突变使其携带用于插入高谷氨酰胺的一
个或多个选择者密码子。例如，目的蛋白的编码核酸可被突变使其携带一个或多个
选择者密码子，因而可以插入一个或多个高谷氨酰胺。本发明包括任何蛋白的这种
突变，外如至少含有一个高谷氨酰胺。同样，本发明还包括相应的核酸，即含
有一个或多个编码一个或多个高谷氨酰胺的选择者密码子的任何核酸。

为了使蛋白包含高谷氨酰胺，需要使用适于通过正交tRNA/RS对在体内插入高
谷氨酰胺的宿主细胞和有机体。宿主细胞用一种或多种载体进行遗传改造(如转化、
转导或转染)，这些载体表达正交tRNA、正交tRNA合成酶和编码衍生蛋白的载体。
这些组分可以位于同一个载体上，或者位于不同的载体上，或者两个组分位于一个
载体上而第三个组分位于第二个载体上。载体可以是质粒、细菌、病毒、裸多核苷酸。
酸或共轭多核苷酸。

**根据免疫活性定义多肽**

由于本发明的多肽提供了多种新的多肽序列（如在本文的翻译系统中合成的蛋白中含有高谷氨酰胺或者在新型合成酶、标准氨基酸的新型序列存在的情况下合成的蛋白中含有高谷氨酰胺），因此多肽可能会提供能被免疫分析方法识别的新型结构特征。能特异性结合本发明多肽的抗血清的制备以及能被这些抗血清结合的多肽也是本发明的特征。本文所用的术语“抗体”包括而不限于免疫球蛋白基因或其片段编码的多肽，这些多肽可特异性识别并结合分析物（抗原）。例子包括多克隆抗体、单克隆抗体、嵌合抗体和单链抗体等。免疫球蛋白的片段，其中包括 Fab 片段和表达文库如噬菌体展示文库产生的片段，也包括在本文所用的术语“抗体”之中。见 Paul 编辑的《基础免疫性》(Fundamental Immunology)，第四版，1999，Raven Press, New York，中关于抗体结构和方法学的描述。


**O-tRNA 和 O-RS 以及 O-tRNA/O-RS 对的应用**

本发明的组合物以及用本发明的方法制备的组合物存在于细胞内。因而本发明的 O-tRNA/O-RS 对或其各个组分可用于宿主系统的翻译机器中，结果导致高谷氨酰胺被插入到蛋白质内。Schultz 等人的专利申请“非天然氨基酸的体内插入”(WO2002/085923) 描述了这一过程，本文已纳入作为参考。例如，当 O-tRNA/O-RS 对被导入到宿主如大肠杆菌内时，O-tRNA/O-RS 对就可在体内将高谷氨酰胺插入到蛋白质如肌红蛋白或治疗性蛋白内以响应选择者密码子如琥珀无义密码子，其中的高谷氨酰胺可外源性地加入到培养基中。另外，本发明的组合物既可用于体外翻译
系统中，也可用于体内系统中。含高谷氨酰胺的蛋白质可用作治疗性蛋白，也可用于研究蛋白质的结构、与其他蛋白质的相互作用、蛋白质内的电子转移过程等。

试剂盒

试剂盒也是本发明的特征。例如，本发明提供了用于在细胞内制备至少含一个高谷氨酰胺的蛋白质的试剂盒，其中试剂盒包括一个盛有 O-tRNA 和/或 O-tRNA 编码多核苷酸序列、和/或 O-RS 编码多核苷酸序列、和/或 O-RS 的容器。在一个实施方式中，试剂盒还包括高谷氨酰胺。在另一个实施方式中，试剂盒还包括于制备蛋白质的说明书。本发明的任何组合物、系统或装置都可以用适当的包装材料（如容器等）包装起来制成试剂盒的形式。

实施例

下面所提供的实施例是为了说明本发明，并不意味着对本发明权利要求的限制。本领域的技术人员应该知道在不偏离本发明权利要求范围的情况下可以对各种非关键性参数做出修改。

实施例 1: 衍生于 ARCHAEAL tRNA135 的正交合成酶/tRNA 对的制备

正交合成酶-tRNA 对来源于超嗜热古菌的赖氨酸 tRNA 合成酶。根据古细菌 tRNALys 序列的多序列排列的共有序列得到的琥珀抑制型 tRNA 可以在 β 半乳糖苷酶分析试验中产生 32% 的琥珀抑制效应。因此，这个正交对是一个高效的系统，可用于在大肠杆菌细胞内将非天然氨基酸插入到蛋白质内。

将有机体的遗传密码扩充到 20 个常见氨基酸之外的其他氨基酸最少需要两个新基因：可选择性活化非天然氨基酸但是不能将氨基酸转移到内源 tRNA 上的氨酰-

最近诞生了一种新的方法，该方法利用热自养甲烷球菌亮氨酸-tRNA 合成酶和盐杆菌 NRC-1 来源的同源正交琥珀、乳白和四碱基抑制型 tRNA 构建正交对(Anderson 和 Schultz, (2003) Biochemistry, 42 (32):9598-608)。这些根据多序列排列的古细菌亮氨酸 tRNA 的相同序列而设计出的抑制型 tRNA 可提供有效的正交移码和乳白抑制型 tRNA。正常的正交抑制型 tRNA 含有 CU(X)XXXAA 反密码子环(其中的(X)XXX 是密码子的反向互补序列)，其茎区无错配碱基。我们现在描述的是利用 “有序列抑制型” 策略进行古细菌 tRNA^{lys} 序列和古细菌超嗜热古菌赖氨酸-tRNA 合成酶来源的正交合成酶-tRNA 对的理论设计。


为了确定 PhKRS 是否是大肠杆菌正交的，需要证明合成酶不能有效地加载大肠杆菌 tRNA。PhKRS 的基因是从基因组 DNA 中通过 PCR 扩增出来的，插到质粒 pBAD-Myc/HisA (Invitrogen) 内过量表达。所得到的 PhKRS 蛋白加以纯化使之成为单一物质。用盐杆菌 NRC-1 或大肠杆菌来源的 tRNA 进行氨酰化分析试验。盐杆菌 NRC-1 来源的 tRNA 序列和超嗜热古菌来源的 tRNA 序列是高度同源的(图 1)。因此，我们预测嗜盐菌 tRNA 可以很容易地被 PhKRS 加载。确实，在 20 分钟内，PhKRS 加载的嗜盐菌 tRNA 的量比加载的大肠杆菌 tRNA 的量高出 14 倍以上(图 2; 10 μM [tRNA])。虽然 PhKRS 也能少量加载大肠杆菌 tRNA，但是，在相同浓度的大肠

为了进一步了解 PhKRS 的正交性，我们试图将 PhKRS 基因插入到组成型表达载体 pKQ 内。这个质粒含有组成型氨酰胺启动子控制下的核糖体结合位点、多克隆位点和质粒 pBAD-Myc/HisA (Invitrogen) 来源的 rrnB 终止子。这个质粒还含有一个 ColE1 复制起点和一个用于质粒筛选的卡那霉素耐药选择标记。不幸的是，野生型 PhKRS 基因在以组成型的方式表达时是有毒性的。但是，我们偶然发现 E444G 突变体(质粒 pKQ-PhE444G)在大肠杆菌内表达时其毒性是降低的。E444G 突变可导致对系统的毒性明显下降的机制还不清楚，一种可能是突变阻止了大肠杆菌 tRNA 的低水平种间错误加载，这种情况在体外可以观察到。


个质粒含有氨苄青霉素耐药选择标记和处于 lpp 启动子控制之下的编码 β-半乳糖苷酶的 lacZ 基因。在 lacZ 基因的允许位 25 位残基上有一个琥珀型密码子，这个密码子可导致翻译的提前终止。在无琥珀抑制型 tRNA 存在的情况下，含 pLASC-lacZ(TAG) 的细胞所观察到的 β 半乳糖苷酶活性只有质粒 pLASC-lacZ(Lys)的细胞所具有活性的 0.17%(质粒 pLASC-lacZ(Lys)在 25 位上含有一个 AAA(赖氨酸)有义密码子)，比不含质粒的细胞所具有的活性也只高 2 倍。如果 AKCUB 不能被内源性的大肠杆菌合成酶加载，当 tRNA 与 pLASC-lacZ (TAG)共表达时不应该产生琥珀型抑制。含 pAC-AKCU 和 pLASC-lacZ(TAG)的细胞只有 1.7%的抑制效率(相对于 lacZ(Lys)表达时)。如果 PhKRS 能够加载正交 tRNA，那么当质粒 pKQ-PhE444G 被导入到系统内时应该可以观察到更高水平的抑制。确实，所观察到的抑制水平达 32%。与此相对应的是，上面所描述的大肠杆菌正交的詹氏甲烷球菌来源的酪氨酸合成酶-tRNA 对在同源合成酶存在的情况下展示出 18.5%的抑制效率，如果没有同源合成酶则抑制效率只有 0.2%。在同源合成酶存在和不存在的情况下，热自养甲烷球菌 (M.thermoautotrophicum)来源的亮氨酸琥珀型正交对的抑制效率分别为 33.2%和 1.5% (Anderson 和 Schultz, (2003) Biochemistry, 42 (32):9598-608)。

在这个实施例中，我们鉴定出了 I 型赖氨酰-tRNA 合成酶，并且通过占细菌 tRNA^{Lys} 序列的多序列排列设计出了一种琥珀抑制型 tRNA，二者一起组成正交合成酶-tRNA对，可用于非天然氨基酸在大肠杆菌内的位点特异性插入。PhKRS/AKCU 所具有的高效率(32%的抑制效率)证明了共有序列策略可有效地用于正交抑制型 tRNA 的构建。

实施例 2：利用非天然氨基酸进行移码校正抑制可消除 PHKRS 的毒性

超嗜热古菌的 I 型赖氨酰 tRNA 合成酶来源的正交 tRNA-合成酶对已被用于大肠杆菌中。该系统的 tRNA 部分可以很好地行使正交琥珀酸抑制子的功能。合成酶表达质粒 pKQ-PhE444G 能够加载这种 tRNA，但是依然可以观察到毒性效应。当其单独表达时，含 pKQ-PhE444G 的细胞的生长密度只有不含质粒的细胞的 56%。报告质粒 pAC-AKCU 和 pAC-AKCU 也有中等程度的毒性，其生长密度分别是不含质粒的细胞的 72%和 52%。当 pKQ-PhE444G 与质粒 pAC-AKCU 共同转化时，细胞的密度下降到 17%。另外，当与 β 半乳糖苷酶报告质粒 pAC-AKCU(质粒 pACKO-A184TAG 的衍生物)共表达时，细胞的密度只有 5%。这清楚地表明 tRNA 和合成酶在这个系统内都有毒性效应。另外，用两个质粒共转化细胞时可能具有协同效应，使细胞的存活
能力大大降低。为了解决这一问题，我们需要找到 PhKRS 的低毒性突变体。

我们猜测在 PhKRS 上进行点突变或其它类型突变可能能够降低合成酶的毒性同时又保留其加载活性。因此，我们将 pKQ-PhE444G 转化到化学感受态 XL1-red 细胞(Stratagene)内，然后将细胞接种到含 25μg/ml 卡那霉素的 LB 琼脂培养板上。这个细胞株含有几个能导致转化质粒发生高频突变的基因组突变。从培养板上挑出约 100 个克隆，然后在含卡那霉素的 25ml 液体 LB 培养基内扩增。预计 pKQ-PhE444G 的非毒性突变株的生长速度要快于野生型的细胞株，通过细胞的系列培养可导致这些突变株的累积。经过 2 个系列培养，每一步稀释 10000 倍，吸出少量细胞导入到含质粒 pAC-AK CUA 的 Genehog 细胞中，然后接种到含 25μg/ml 卡那霉素和 25μg/ml 氮霉素以及各种浓度的氨苄青霉素的 LB 琼脂培养板上。90%以上的转化细胞无明显的毒性，可以在含 1000μg/ml 氨苄青霉素的 LB 琼脂培养板上生长。甚至在含 1500μg/ml 氨苄青霉素的培养板上也看到了较小的克隆，说明产生了有效的琥珀型抑制。分离出一个突变的合成酶，命名为 pKQ-PhKep，通过限制性酶切图谱和 PhKRS 开放阅读框架测序的方法确定其特征。突变的基因含有一个 778bp 的插入片段，位于残基 S357 之后，但是其他的序列都与质粒 pKQ-PhE444G 相同。通过 BLAST 检索发现这个插入片段与质粒 p1658/97 上被称为 “insAcp1” 的序列同源，但是在这个文献中再也没有提到这个序列，因此也无法知道这个序列的来源。当 PhKRS 从起始密码子开始翻译时，这个基因的预测产物在 S357 的下游被裁掉 6 个氨基酸。为了确定 PhKRS 的这种截断是否就是毒性下降的原因，合成了一个序列 5'-CAGTGGAAATTCAGTAAAGTGGCAGCATCAC-3' 的引物 CA510R，用它在质粒 pKQ 上构建截断突变体。利用引物 CA279 和 CA510R 通过 PCR 扩增质粒 pKQ-PhKep，扩增的产物亚克隆到质粒 pKQ 的 NcoI 和 EcoRI 位点之间。所得到的质粒 pKQ-PhΔAD(也被称为 pKQ-Ph510)与质粒 pAC-AK CUA 共转化，结果显示得到的转化子与 pKQ-PhKep 转化细胞具有相同的 IC50，没有明显的毒性。

在残基 357 后的截断好像可以删除 PhKRS 的反密码子结合区，我们想要了解这种删除对合成酶的 tRNA 识别特性会产生何种影响。因此，我们在体外过量表达合成酶并作氮酰化分析。用引物 CA279 和 CA511 (5'-CATTGGGAATTGAGATTTGGCAGCATCAC-3') 通过 PCR 从 pKQ-PhKep 上扩增这个基因，然后亚克隆到 pBAD-Myc/HisA 的 NcoI 和 EcoRI 位点之间，使其与 C 末端 Myc/His 标记处于同一个阅读框架内。利用 Ni-NTA 层析技术纯化表达的蛋白质。
实施例 3：用四碱基密码子和非天然氨基酸扩充遗传密码


为了在体内用四联密码子编码非天然氨基酸，我们必需制备出一种能特异性识别这种密码子的 tRNA 和一种只能用目的非天然氨基酸特异性地氨酰化这种 tRNA 的相应合成酶。由于以前所制备的正交性甲烷球菌琥珀抑制型 tRNA 的反密码子环是同源合成酶 JYRS 的关键识别元件，因此要改造这个 tRNA 使其解码四碱基密码子是很困难的。虽然也有可能降低 JYRS 的反密码子结合特异性，但要构建出用一个反密码子序列就可以区别琥珀型抑制子和四碱基抑制子的互相正交的对可能是很困难的。因此，最可能的情况是一个系统利用两个不同来源的正交对将两个不同的非天然氨基酸同时插入到多肽链，这样就需要设计出另一个新的正交 tRNA-合成酶对。


如实施例 1 所详细描述的，为了确定这个截断突变体在大肠杆菌内是否有功能、是否是正交的，需要证明这个合成酶不能加载大肠杆菌的 tRNA，但是要保留加载同源古细菌 tRNA 的活性。克隆 PhΔAD 和 EcKRS 的基因并使其过量表达，纯化所得到的蛋白使其成为均一的物质。用古细菌盐杆菌 NRC-1 或大肠杆菌来源的全 tRNA
进行氨酰化分析试验。在 20 分钟内，Ph\DeltaAD 加载的嗜盐菌全 tRNA 的量比加载的大肠杆菌全 tRNA 的量要多很多(图 2：10 \mu M \text{[tRNA]} )。虽然 Ph\DeltaAD 也能少量加载大肠杆菌 tRNA，但是，在相同浓度的大肠杆菌全 tRNA 存在的情况下，其加载速率比大肠杆菌合成酶低 28 倍，只是背景加载速率的 7 倍。另外，Ph\DeltaAD 不能弥补 PAL\DeltaS\DeltaUTR(pMAKlysU) 株在 43°C 下的生长(Chen 等，1994) Properties of the lysyl-tRNA synthetase gene and product from the extreme thermophile Thermus thermophilus. J. Bacteriol. 176:2699-705)，这是一个大肠杆菌赖氨酰 tRNA 合成酶发生 lysS lysU 双缺陷的突变株。但是，EcKRS(克隆自大肠杆菌 HB101 株的 lysU 位点)可使其正常生长。因此，Ph\DeltaAD 不能替代 PhKRS，在体内可能很难与内源性的大肠杆菌合成酶竞争加载大肠杆菌的 tRNA。

为了证明 Ph\DeltaAD 在大肠杆菌内是有功能的，我们利用正交琥珀抑制型 tRNA 进行 β 半乳糖苷酶抑制试验来分析 Ph\DeltaAD。这种方法还可以使我们利用以前设计的詹氏甲烷球菌正交对筛选试验方案。抑制型 tRNA_{CUA}(AK\text{CUA}) 是利用最近描述的共有序列抑制策略(Anderson 和 Schultz, P. G. (2003) Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair for Four-base, Amber, and Opal Suppression. Biochemistry 42:9598-608) 设计出来的。从现有的基因组序列中挑出所有的古细菌 tRNA^{lys} 序列进行多序列排列，确定其共有序列。将反密码子环序列改变成 CUCUAAA 从而形成一个琥珀抑制型 tRNA(图 1)。合成 AK\text{CUA} 的基因，并将其插入质粒 pACKO-A184TAG(Anderson 和 Schultz, P. G. (2003) Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair for Four-base, Amber, and Opal Suppression. Biochemistry 42:9598-608) 以确定其琥珀型抑制效率。这个质粒含有 β 半乳糖苷酶的基因(bla)，该基因的允许位 A184 上是一个 TAG 密码子。当没有 tRNA 时，检测不到全长 β 半乳糖苷酶的表达，对 5μg/ml 的氨苄青霉素敏感。当 AK\text{CUA} 表达时，对氨苄青霉素的耐药性没有升高，说明 tRNA 没有被大肠杆菌的合成酶加载。当与 Ph\DeltaAD 共表达时，正交 tRNA 被加载，产生有效的琥珀型抑制，对氨苄青霉素的耐药性升高到 1000\mu g/ml。这些结果表明 Ph\DeltaAD/AK\text{CUA} 正交对是有效的正交琥珀型抑制系统。

以前的研究结果表明四碱基密码子可被大肠杆菌的 tRNA^{agg}_{CU} 有效抑制。在这种情况下，四碱基密码子的抑制可与稀有密码子 AGG 竞争，从而使抑制型 tRNA 的毒性消失。AGGA 抑制型 tRNA 来源于 AK\text{CUA}，将其反密码子环改变成 CUUCCUAA，将 AGGA 抑制型 tRNA 插入到质粒 pACKO-A184AGGA 中。与 pACKO-A184TAG 一样，这个质粒在 A184 位点上也含有 AGGA 密码子，该密码子可导致翻译流产，
对氨苄青霉素的耐药性只有 5μg/ml。不幸的是，这个 tRNA 不再与大肠杆菌的合成酶正交。含有这种 tRNA 的细胞无论是否存在 PhΔAD 都可以在 200μg/ml 的氨苄青霉素中存活。

为了鉴定正交变体，我们构建了一个 tRNA 文库，其中位于 1-4 位和 69-72 位的 tRNA 受体臂的最后四个碱基对都是随机分布的 (Anderson 和 Schultz, P. G. (2003) Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair for Four-base, Amber, and Opal Suppression. Biochemistry 42:9598-608)。这些 tRNA 与 PhΔAD 共表达，细胞经两轮 200μg/ml 的氨苄青霉素筛选后得到一组有活性的 AGGA 抑制型 tRNA。为了鉴定正交的变体，分离 tRNA 质粒，挑出 384 个在不存在 PhΔAD 时对氨苄青霉素敏感的克隆。这些克隆中，活性最高的正交克隆在有 PhΔAD 时对氨苄青霉素的耐药性达到 700μg/ml，而在没有 PhΔAD 时对氨苄青霉素的耐药性只有 5μg/ml。这个 tRNA，AK_{UCCU}，包含多个受体臂上的碱基替代，其中一个偶然发现的 A37C 变化重要性是未知的(图 5)。

为了插入能响应 AGGA 的高谷氨酰基，下一步要做的是改变 PhΔAD 的特异性。选择高谷氨酰基作为起始靶位来确定突变的合成酶的保真度，因为它与赖氨酸的大小一样，既有氢键提供特性，也有氢键接受特性。经过对 PhKRS 晶体结构的分析，发现只有两个残基 E41 和 Y268 可以特异性地识别赖氨酸的 ε-氨基基团。在构建 PhΔAD 的小活性位点文库时通过饱和突变使这两个残基随机分布。为了筛选出 hGln 特异性的合成酶，我们使用了一个 GFP 报告质粒 pREP2-AK_{CUA}(Santoro 等, (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. Nat. Biotechnol. 20:1044-8)。这个质粒编码 AK_{CUA} 的基因，在 M1 和 Q107 位置含有两个 TAG 密码子的 T7 RNA 聚合酶基因和位于 T7 启动子之下的 GFPuv。当 pREP2-AK_{CUA} 和 PhΔAD 共同转化时，T7RNAP 上的琥珀型密码子被抑制，导致 GFPuv 的表达，产生绿色荧光。在没有合成酶时，细胞是白色的。因此，通过观察是否能发荧光就可以鉴定出含活性合成酶的细胞。细胞用 pREP2-AK_{CUA} 共转化，然后在含 hGln 的培养基上培养文库，分离出每一个绿色克隆，然后用含和不含非天然氨基酸的培养基培养，挑出那些需要 hGln 才能发荧光的克隆。在挑出的 15 个克隆中，有 5 个能在含 hGln 的培养基上产生较强的荧光。在这些克隆中，除了一个之外其他都含有保守的 Y268S 替代；这些合成酶中特异性最高的合成酶 hGlnRS 具有 I41 和 S268 突变，对其做进一步鉴定。与 PhΔAD 相比，5 个克隆相应的 5 个突变体具有如下的序列变化:

克隆 Y268(密码子) E41(密码子)
含有 Gly24→AGGA 突变的肌红蛋白用正交 hGlnRS/AKTCTTT 对表达以确定所观察到的 hGln 依赖性表型是否是因为非天然氨基酸的特异性插入。在 hGlnRS 不存在的情况下，突变的肌红蛋白基因表达时，检测不到蛋白质的产生。而与 hGlnRS 共表达时，可以分离出 1.8 mg/ml 肌红蛋白。与此相比，含野生型肌红蛋白基因的质粒 pBAD-JYAMB 表达时所产生的肌红蛋白的量为 3.8 mg/ml。对纯化蛋白的胰酶水解片段进行基质辅助的激光解吸/电离飞行时间(MALDI-TOF)分析，结果发现多肽的分子量为 1676.85 Da，与预测的 1676.88 Da 一致。没有检测到在 24 位上含赖氨酸或谷氨酰胺的多肽。另外，我们观察到在为响应 AGGA 而将 hGln 插入期间，几乎没有毒性。在无阿拉伯糖诱导的肌红蛋白表达所使用的培养条件下，GeneHogs 细胞在指数生长期间的倍增时间是 hGln 的细胞的两倍。无论是否存在 hGln 都可以观察到生长速率的下降，这说明所产生的轻微毒性是合成酶和 tRNA 表达的结果，而不是 AGGA 抑制的结果。因此，AGGA 抑制子和稀有 AGG 密码子之间的交叉反应并不会限制移码抑制的实际使用。

图 7A 说明了通过 AGGA 抑制而插入 hGln 残基时肌红蛋白的表达情况。这个图是一个蛋白印迹杂交图，所用探针为抗 His C 末端抗体。只有在 AK514 tRNA、hGlnRS (PhKRSΔ 的一种 hGln 特异性变体) 和 hGln(2-5 道) 这三个组分都存在的情况下在 G24 位上含 AGGA 密码子的肌红蛋白基因(Myco24AGGA) 才能表达出蛋白。只有在 JYRS 及其同源琥珀型 tRNA 存在的情况下在 75 位上为琥珀型抑制的肌红蛋白(Myco75TAG)才有可能表达，这说明赖氨酸和酰胺酸对(6-9 道) 是互相正交的。

够联合使用而没有交叉反应。


图 7B 描述的是利用两个正交 tRNA 系统同时插入两个不同的非天然氨基酸的肌红蛋白模板系统。这个图是一个蛋白印迹图，所用探针为抗 His C 端抗体。在赖氨酸和酪氨酸正交对都存在的情况下，24 位上含 AGGA 密码子和 75 位上含 TAG 密码子的肌红蛋白基因(Myo24AGGA/75TAG)可表达出蛋白质。利用 hGlnRS 和 OmeTyrRS，24 位上的 AGGA 抑制密码子和 75 位上的琥珀抑制型密码子分别将 hGln 和 O-甲基-L-酪氨酸(OmeTyr)插入到一条多肽链上。如图所示，只有在两种非天然氨基酸、赖氨酸正交系统和酪氨酸正交系统都存在的条件下才能制备出蛋白来。

图 7A 和 7B 的蛋白印迹结果被图 8 和图 9 的分析结果进一步证实。图 8 显示的是如图 7A 所示的野生型肌红蛋白(MyoWT; 据推测在 24 位上为赖氨酸)和突变的肌红蛋白(24AGGA; 据推测可能含有 hGln)的肽酶水解片段的基质辅助的激光解吸/电离飞行时间分析结果。通过对纯化蛋白的肽酶水解片段的分析发现多肽的分子量为 1,676.85 Da，与预测的分子量 1676.88 Da 一致。没有观察到在 24 位上为赖氨酸(用 PhKRS3 表达的蛋白质的肽酶裂解片段的分子量为 950.46 Da，计算出的分子量为 950.53Da，全长肽的分子量为 1,661.91 Da)或谷氨酰胺(计算出的分子量为 1661.87 Da)的多肽。

图 9 显示的是全长双突变蛋白(如图 7B 所示)的电喷雾质谱(MS)分析结果。MS 结果表明含两个非天然氨基酸的肌红蛋白的分子量为 18,546.40 Da(SD 0.11)，与预测的分子量 18,546.60(SD 0.81)一致，含 G24K 和 A75Y 替代的肌红蛋白的理论分子量为 18,518.3 Da。

总之，我们的试验结果证明了利用移码抑制可在体内进行非天然氨基酸的位点特异性插入。另外，我们的实验结果还表明这个超嗜热古菌(Photorikosshii)来源的 tRNA 合成酶和詹氏甲烷球菌酪氨酸-tRNA 合成酶来源的正交对是互相正交的，我们能够在一条多肽链上同时插入两个非天然氨基酸。筛选出能将互相正交的活性氨基酸甚
至是荧光氨基酸插入的合成酶变体应该也是可能的。利用这种材料就有可能将荧光对插入到多肽中用于在体外进行 FRET 研究。同样，通过插入氨基酸之外的基序如 α 
羟酸来制备非天然多聚物的核糖体产物也是可能的。其他的密码子如 CCCU 或 
CUAG 也可以被使用，这些密码子在大肠杆菌内都是被抑制的。另外，利用有限的 
密码子的简并性来从新合成大肠杆菌的基因组，从而可以使可用的密码子达到 43 个。 

材料和方法

合成酶基因的克隆和表达：制备从美国典型培养物收集中心获得的超嗜热古菌 
(ATCC：#700860) 的基因组 DNA。通过 PCR 扩增 PhKRS，然后插入到质粒 
pBAD/Myc-HisA(Invitrogen) 的 NcoI 和 EcoRI 位点之间，用于过量表达 PhKRS。为了 
使其进行组成型表达，将这个合成酶基因克隆到 pGLN 和 pKQ 内(Anderson 和 
Schultz，2003) Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair 
for Four-base，Amber，and Opal Suppression。Biochemistry 42:9598-608)，处于谷氨酰 
胺启动子的控制之下。这些质粒来源于 pBAD/Myc-HisA，分别具有氨苄青霉素和卡 
那霉素抗性。同样，从大肠杆菌 HB101 株的 lysU 位点中克隆出 EcKRS。按照 Qiagen 
QIA 表达试剂盒所描述的方法在 SYT 培养基中过量表达蛋白，然后用 100 
mM Tris-HCl，pH 7.5；100 mM NaCl 和 10% 甘油透析。

体外氨酰化分析：大肠杆菌全 tRNA 购自 Roche，嗜盐菌 tRNA 用 RNA/DNA 提 
取试剂盒(Qiagen)从盐杆菌 NRC-1(ATCC：#700922) 培养物中提取。反应总体系为 
20μl，其中含有 50 mM Tris-HCl，pH 7.5；30 mM KCl；20 mM MgCl2；3 mM 谷胱甘 
肽；0.1 mg/mL BSA；10 mM ATP；1 μM [3H] 赖氨酸(Amersham); 750 nM 合成酶和 0、 
2、10 或 40 μM 全 tRNA，37℃ 反应 20 分钟。

补偿分析：PALΔSÅUTR(pMAKllysU) 细胞用 pGLN 匿骨生转化以表达 PhΔAD、 
EcKRS、或不表达合成酶，在 30℃、含 50μg/ml 氨苄青霉素的 LB 琼脂培养板上培养 
细胞。在不含抗生素的 LB 琼脂培养板或液体培养基内进行合成酶缺失生长缺陷突变 
株 PALASÅUTR 的补偿分析，试验温度为 43℃。GeneHogs 的生长状况以为阳性对 
照以消除合成酶表达所产生的毒性效应的可能性。对于液体培养基来说，饱和培养 
物用新鲜培养基稀释 1000 倍，在 600nm 处监测其 8 小时的生长情况。

文库构建：通过合成寡核苷酸(Genosys)的重叠延伸构建 tRNA 的基因，亚克隆到 
pACKO-A184TAG 或 pACKO-A184AGGA 的 EcoRI 和 PstI 位点之间。这些 
pACYC184 来源的报告质粒含有 p15A 复制起点、氯霉素耐药基因和一个强的组成型 
启动子来控制 tRNA 基因的表达。PhΔAD 来源的文库用质粒 pKQ 通过 EIPCR 构建

抑制效率的确定：通过在添加 25μg/ml 卡那霉素和氯霉素以及从 5 到 1000μg/ml 之间不同浓度的氨苄青霉素的 LB 琼脂培养基内接种来确定抑制效率。抑制效率定义为细胞能存活并形成克隆的一系列培养板中的最高浓度，其次高或次低的浓度应该在该值的 20% 以内。

含 hGln 的肌红蛋白的表达和特性检测：为了进行单位点插入试验，构建含 p15A 复制起点和编码氯霉素乙酰转移酶 AK_{UCU} 和复制位点肌红蛋白基因的质粒 pMyo-AK_{UCU}。肌红蛋白的基因被置于阿拉伯糖启动子的控制之下，在其 24 位是 AGGA 密码子。为了进行双位点插入，通过在质粒 pMyo-AK_{UCU} 的肌红蛋白基因的 A74 位上引入 TAG 密码子来构建另一个质粒。将正交 tRNA^{35} - J17 置于第二 lpp 启动子的控制之下。合成酶 hGlnRS 和 AzPheRS 用质粒 pQK 表达，受谷氨酰胺独立启动子的控制。含有合成酶酶和肌红蛋白表达质粒的 GeneHogs 细胞在 37°C 培养，使之生长到 OD_{800} = 0.7，所用的培养基为 GMML 培养基，其中添加了除赖氨酸之外的其他 19 种氨基酸，每种 0.4mg/ml。1 mg/ml hGln (Sigma)，用 0.02% 的阿拉伯糖诱导，然后使其生长到饱和状态。超声裂解细胞，用 Qiagen QIA 表达试剂盒纯化肌红蛋白。其肽酶水解片段用 TSRI Proteomics Facility 进行 MALDI-MS 分析。

实施例 4：示例性 赖氨酸 O-RS 和 赖氨酸 O-tRNA
示例性 O-tRNA 见于实施例和/或表 1 中。示例性 O-RS 也见于实施例和/或表 1 中。因此，编码 O-RS 或其片段的多核苷酸也包括实施例和/或表 1 中所列的那些多核苷酸。


应当理解的是本发明所描述的实施例和实施方式只是为了说明的目的，本领域的技术人员可以对这些实施例和实施方式进行各种修改或改变，这些经修改的实施例或实施方式也包括在本申请的精神和范围之内以及本发明附属权利要求的范围之内。
虽然为了便于理解本发明，在上文已经对本发明做了某种程度的详细描述，但是应该清楚的是，只要不偏离本发明的范围，本领域的技术人员通过阅读本说明书就可以对本发明的形式和细节作出各种修改。例如，上面所描述的各种技术方法和装置都可以不同组合的方式加以使用。出于所有目的，本申请所引用的所有文献、专利、专利申请和/或其他文献都已完整纳入本文作为参考，这就好比各个文献、专利、专利申请和/或其他文献都被单独纳入本文作为参考。
<table>
<thead>
<tr>
<th>SEQ ID:</th>
<th>标号</th>
<th>序列</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ ID: 1</td>
<td>Pa15</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 2</td>
<td>Pf32</td>
<td>GGGCGGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 3</td>
<td>Ph36</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 4</td>
<td>Pa34</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 5</td>
<td>Ph9</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 6</td>
<td>Pf4</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 7</td>
<td>Pya26</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 8</td>
<td>Ta1</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 9</td>
<td>Tv40</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 10</td>
<td>Af18</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 11</td>
<td>Hh1</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 12</td>
<td>Ta5</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 13</td>
<td>Tv24</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 14</td>
<td>Mj21</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 15</td>
<td>Mt15</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 16</td>
<td>Mm4</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 17</td>
<td>St1</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 18</td>
<td>St43</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 19</td>
<td>Pya13</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 20</td>
<td>Af15</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 21</td>
<td>Ss43</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 22</td>
<td>Ap47</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 23</td>
<td>isus</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 24</td>
<td>AKC</td>
<td>GGGCCGAGUAGCUAGCGUAGCAGGCGGACUCUUAUACGCCAGUAGGGGUUCAAAUUCCCCGGGCGCCGCA</td>
</tr>
<tr>
<td>SEQ ID: 25</td>
<td>AK</td>
<td>NNNNCCGUAGCUCAGCCUUGGUAGAGCGCGGCCUUCCUAAACCCGQ</td>
</tr>
<tr>
<td>SEQ ID: 26</td>
<td>AKUCC</td>
<td>UGGUCGCUAGCUCAGCCUUGGUAGAGCGCGGCCUUCCUAAACCCGQ</td>
</tr>
<tr>
<td>SEQ ID: 27</td>
<td>PhAA</td>
<td>ATGGTTCATT GGGCCGATTA TATTGctag ataataata gagagagggg</td>
</tr>
<tr>
<td>SEQ ID: 28</td>
<td>PhADAD</td>
<td>MVHWDYIAD KIRERGKEKY KYVUESGTPSY GVHYQFRFLPTATYTVGH</td>
</tr>
<tr>
<td>SEQ ID: 29</td>
<td>PhE444G</td>
<td>atyttgctatt ggccgcatta tattgctgctg ataataata gagagagggg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tcctagtgg gaatttacg tggaggttag aatacgccaa aagttgtctag tttggggtggct gctcgaaggg cactatccaa tgtgggtgagt aatagtggaga gttcagagtcc tggtttttag cagtttaattgctgagt gataagagagt gctggtgtgg ggtggtgtgg ggtggtgtgg ggtggtgtgg ggtggtgtgg</td>
</tr>
<tr>
<td>SEQ ID: 30</td>
<td>PhE444G.ep 也称为 PhKPS</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------</td>
<td></td>
</tr>
<tr>
<td>MVHWDYTAAD KIIRGERGEKE KYVUESGTP SGYVHGNFR BLPAYIVGH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALDRGKYEVBR MIHMPWYDRY FRKVFRNPQW EWDLYGMPB ETGPDZGWCH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESYASHFMKR FEEVEKLEG horrified YKRGYSEEY RLFARKRDK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEILNKVRIR AQQPPLERPW WMAMYCPHE RHAREBIEN WGWKVKYKCP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BGHSGVRWDR SGVPHLKNWV DWPMRHSWPQ FVBDVPARSH LVAGSSYDGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIEIKVEYHR EAPLSLMNYEF VPXGIQKYM SGSKGVVLV SDLYEVLEPG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVRFYIYHRH FKXKELDGGL LIGILNALDPF EVKREIPFGV EGGKGDDEEL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRTYELMSPK FPKBPVQAF QFPFLVLAVQF HLFREEDIN LVIKSHIHPF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLRSKVDVEVR KIAKLNAM RVKYAPEDVK FSIILEKPEV ESFGVDREAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEVARWLENHI BEFSEBFTHN ILFEVAKRKG ISSRFEWSTL YRFIPEGK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRLSAFLAPL DRSFVYKLRLE LEG*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEQ ID: 31</th>
<th>pACKO-A184TAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaaacgcccga tgcagctcga tccaggcggcc aagaatagta ataaggccg</td>
<td></td>
</tr>
<tr>
<td>gataaactt gttgcttatt tctttactag tcctttatatata gggccgtaat</td>
<td></td>
</tr>
<tr>
<td>tccagctgaa ttcggctgt ttattgagta acagagatag aacgcttgatg</td>
<td></td>
</tr>
<tr>
<td>ccctaaatgt ctcttacagt gcattgagga tataacagct gttggtatct</td>
<td></td>
</tr>
<tr>
<td>ctttgatttt cttccctcct tctctctctctct tctctctctctct ccgaggtggt</td>
<td></td>
</tr>
<tr>
<td>gcttgacagc aaagggcgag cggcagcag cggcagcag cggcagcag cggcagcag</td>
<td></td>
</tr>
<tr>
<td>ggggttactat gttgccagct aaggggagt gttggtgct gttggtgct gttggtgct</td>
<td></td>
</tr>
<tr>
<td>cggggagaga aaggggagc ggtggtgctt tctctctctctct cggcagcag</td>
<td></td>
</tr>
<tr>
<td>tatattcccg ttcctccgct tccggcctgc aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ggcgaggcgg aaggggagc ggcgaggcgg ggcgaggcgg ggcgaggcgg ggcgaggcgg</td>
<td></td>
</tr>
<tr>
<td>ggggttactat gttgccagct aaggggagt gttggtgct gttggtgct gttggtgct</td>
<td></td>
</tr>
<tr>
<td>cggggagaga aaggggagc ggtggtgctt tctctctctctct cggcagcag</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>aattggttgcc gcgctgcgcgc gcgctgcgcgc gcgctgcgcgc gcgctgcgcgc gcgctgcgcgc</td>
<td></td>
</tr>
<tr>
<td>cggcagcag cggcagcag cggcagcag cggcagcag cggcagcag cggcagcag</td>
<td></td>
</tr>
<tr>
<td>gcagagagag gcagagagag gcagagagag gcagagagag gcagagagag gcagagagag</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>ccattgccct ggcgaggcgg aaggggagc ggcgaggcgg aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>gagagagaga aaggggaga aaggggaga aaggggaga aaggggaga aaggggaga</td>
<td></td>
</tr>
<tr>
<td>atgcacctcaat</td>
<td>aatattgaa</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>tgcgctctctt</td>
<td>atctccctct</td>
</tr>
<tr>
<td>accgcaacgc</td>
<td>gctgagaaaa</td>
</tr>
<tr>
<td>cggaaatcgc</td>
<td>aagagagccg</td>
</tr>
<tr>
<td>gctgagaggg</td>
<td>ctgctggctc</td>
</tr>
<tr>
<td>gagactgtcctt</td>
<td>tgtgctggctt</td>
</tr>
<tr>
<td>ggtggagcctt</td>
<td>gggaggtgcctt</td>
</tr>
<tr>
<td>gagactgtcctt</td>
<td>tgtgctggctt</td>
</tr>
<tr>
<td>ggtggagcctt</td>
<td>gggaggtgcctt</td>
</tr>
<tr>
<td>SEQ ID: 32</td>
<td>PACKO-A184AGGA</td>
</tr>
<tr>
<td>-----------</td>
<td>--</td>
</tr>
<tr>
<td>aacggtgtaa caagggtgaa cactatccca tatcaccacgc tcaccgtctt tcattgcccact acg</td>
<td></td>
</tr>
<tr>
<td>gaaactccgga tggactctca tcagggccggc aagaatgtga aataaagcgcg</td>
<td></td>
</tr>
<tr>
<td>gataaactct tggcttatatt ttctttaaaggg ttctttaaaaa ggcccgtataa</td>
<td></td>
</tr>
<tr>
<td>tccagctgaa cggctcgggttat ataggtcactc tgaagcaacgc actgaaatgc</td>
<td></td>
</tr>
<tr>
<td>cttcagatgt tcttctacag caccttgga aactataacgc ttaaggtggtc tggcagtatc</td>
<td></td>
</tr>
<tr>
<td>caggtatatt ctttctcatat tttgcttcgt tagtctcggta aatatcgcgtat</td>
<td></td>
</tr>
<tr>
<td>aacctaaaat atacgcccgg agttgatcct attttcatat attgctggattg acgttgaaagt</td>
<td></td>
</tr>
<tr>
<td>gaaaccccttt aogtggccgt caaggtcctta bttctgcccc aaacctgg caggtgagt</td>
<td></td>
</tr>
<tr>
<td>ggctggtccc gcaagtttcc gagatcgtta cattcttctac gtcgaaagtac</td>
<td></td>
</tr>
<tr>
<td>gctatcact ctgatctcact gagcagctg cgcgttcgtta ctgcttcctctac gtcgaaagtac</td>
<td></td>
</tr>
<tr>
<td>gctggtttcg aatcgtcgg gctggctacc cgggtctggctg actcggccag aacctggcgt</td>
<td></td>
</tr>
<tr>
<td>gccggtgtca cgtggctggc aatcgtcgg gctggctacc cgggtctggctg actcggccag</td>
<td></td>
</tr>
<tr>
<td>gccggtgtca cgtggctggc aatcgtcgg gctggctacc cgggtctggctg actcggccag</td>
<td></td>
</tr>
<tr>
<td>gctggtttcg aatcgtcgg gctggctacc cgggtctggctg actcggccag aacctggcgt</td>
<td></td>
</tr>
<tr>
<td>gccggtgtca cgtggctggc aatcgtcgg gctggctacc cgggtctggctg actcggccag</td>
<td></td>
</tr>
<tr>
<td>gctggtttcg aatcgtcgg gctggctacc cgggtctggctg actcggccag aacctggcgt</td>
<td></td>
</tr>
<tr>
<td>gccggtgtca cgtggctggc aatcgtcgg gctggctacc cgggtctggctg actcggccag</td>
<td></td>
</tr>
<tr>
<td>gctggtttcg aatcgtcgg gctggctacc cgggtctggctg actcggccag aacctggcgt</td>
<td></td>
</tr>
</tbody>
</table>
atCTCATGAC CAAATAATCCCT TAAACGgcatg cacccatattct tgcggcgccg
gtgctcacaag gcctcaacact actactggcgc TGCTCTCTTAAT GGCAGGAGGC
GCAATAAGGG GAGGGCTTGGG CGAAAGGAGG ATCTGCTGCC AGGCAGATTA
GTTGGCTAAG GGCAGAGGTT TCCCGACTAC GAGCTGTTAA AAGCAGGGCC
AGTGGCAAGC TTAAAAaaa atccttagcct tgcctagag tCTCTGATTA
TAATACTCTT CTATTGGGCT CTAAAATACTC TATAGTGGT TCTACGTAGC
CATAAAAATA ATCCATTGGAG TGGACATTT GATTTTTATT TAAACATCTC
AATCCCTCAT AGCTTAAAAC TGCCAAATTT CACCAAAATT TTCTCGGCTC
TTTATATGTT AAAAGAGATT CAAGGAGATT TTATAGTTAC TCTACATTAG
GAATCTCCTTT AACAATCCCA GCTGGGCAAT ATCTCTCTTT TATCTACAGC
CTAACCTCTT CTGAGGACAT ATCAACAGCT ATAAAAATCTC TTATAGGAGA
AAACAACCTT TTTTGTAAA AGCTCTCTTG CTAACAGGCG TATTTTTTTC
TGCTCCATCC TCCCAACTGC ACCATAAACC CTTAAAAATT GAACTACAT
AAGCGCCATT AGTGGAGATA TCTACCTAC AGCCTTTGGG TTTCCTACCT
CTTGGGCTTAT AGTGTCCTAATA CTTCTCTTCTG CTCTTTTTGA AGTAGTTTTT
AACCGGCTTC TTATGACATTT CAGTGTTAATA TCTCTATTTAA GCTTGAATT
agcgttacaag gatattacaca aagtttattta tgttgagaat attttttttaa
tgggaggccac ccattttttttag atcctttctgct caaagAAGCG CCGCAAGGNN
TGGTTTCTAA TTCCACAGTT AGACGGCCAA AGACAGCAGCA TGAACGctct
cattttcttat tctgaggttaa aacaggttccc accgtgttccc ggtagctcct
tccggtcgccgc gctcttgccag ctagctgctc cgtctgcttct
ttttatagct acactgttagc acagttggtcc gcaggccgctca acagtccccc
gcggccacatt tccctgcaac tccaagttgg acctgaagcc cctttgccca
atatgttcct gatcttctac gcagagttgc tcgcttcattc ctcggtgaaa
cctcattctg ttaattattag aagccttccag ttttttatttt ctttctttgag
gcgaatattaa tcgatcatttc gctcaagttat acctgctacc gcagagttgg
gacatattac ccggatttaa cagcataagct catgctagct ctcggtgaaa
ccctgggatt acctccatattg cagctgagct ctttctttttag aagattaagcc
tgtctgtttat aatttcattaa gcattttctcg gactatggagc cctattacaga
gccagctgct aggcagcagc tgcagcagca atttttttttac gcagagttgg
cgctgctttatt tccagttaata gcctcctgct aagtttattta tgttgagaat
aaaagctacctg ttaaagaggtg aacacgcttgg acgcgtaggg cacccaggg
ccaaatatattt tccattattta cgggttagttt gaaattttg accaatggtt
cgttaaagcg cctctttcagg gatgatattt tgtgagctg cgcagacgg
agtctcttttcatt gtcggttacgt aatggtggcct cagcttatttt gccggtcctc
tggcagaggg tggacgttcc gcgtgctcag gcagaatttt tcgatagcttt
cagagagatt ctttcagcttt cgggggtcag cttttttttag aagtttattta
SEQ ID: 34
pKQ

| ATGGATCCGA | GCTGAGATGC | TGCAGCTGGT | ACCATATGGG | AATTCTAAGC |
| TGGGCCCCGA | ACAAAAATCT | ATTCGAGAGG | AGGAGTCTGA | TACGCTTGTC |
| GACCATCATC | ATCATATCGA | TGGAAGTTAA | AGCGTCTCCA | GCTGGTGGCT |
| TGGGGGACAT | GAGAAGAGAT | TTTAGAGCTT | ATACAGATTTA | AATCATACG |
| CAGAGGGTGT | CTGATAAACG | GAAATTTACG | TGCGCGAGAT | AGCAGGTTGG |
| TCAGCTCAGA | CCCATCGCGC | AACAGAAGG | TGAAGACCAG | TACCGCGGT |
| GTCTGCTCGA | GTCTGCTCGA | TGGACAGAGA | GAGAAGTCGC | AGGGATCAGA |
| TAAATACAA | GCTCAGTGGG | AAAGACTTGG | CCTTCTCTTT | TATCATCTGT |
| TGGGATGGA | AAGATATTGTG | TTTCATATA | CTGCTGATG | GCGCCAGC |
| TGGAGGTCA | AGGATTTTTC | TCAGCTGATC | GTTCAGGCA | CCGCAGC |
| AGGAGTCTA | CCCCAGAGG | TCAGGTCTG | GGTTCGAG | AGTGGTGCT |
| TCTATGCTT | TGATCTTCTC | GACGAGGTG | GCCGAGAAG | AAACGAGA |
| GCAAGACCT | GCTTACGGTG | GAAAGAAT | CCAAGAAG | AGAGGAGA |
| GCTATGTCG | AACGTTTCA | ATACGATTC | CTGCGTCG | AATATCATC |
| TCACGAGGC | GAGCTGATC | TATCGAGG | TGGTCTATA | AAGAAGATG |
| GTGGTGCTC | GGGGACCGC | GTAATATTC | CTTACTCAG | GCATGAC |
| TGATGTTGTA | GACCAGAGG | AAGAGACT | AATACGATG | ATGAGATG |
| TGCTCATGT | TGACGCTGA | GCTTAAAT | ATGAGATG | CTGACGAG |
| CAGGTTGG | AGCTTGG | CCAAGATTC | ATGAGATG | ATGAGATG |
| TAGTTTGTG | GTGAGGCTA | GCGGACTAC | CTGCGTCG | AATATCATC |
| GTGGGATG | GGGGACCGC | GTAATATTC | CTTACTCAG | GCATGAC |
| CTGAGAGG | AGCTTGG | CCAAGATTC | ATGAGATG | ATGAGATG |
| TGATGTTGTA | GACCAGAGG | AAGAGACT | AATACGATG | ATGAGATG |
图 3
图 5