



(12) 发明专利申请



(10) 申请公布号 CN 114729334 A

(43) 申请公布日 2022. 07. 08

(21) 申请号 202080070644.1

(22) 申请日 2020.08.07

(30) 优先权数据

62/884,454 2019.08.08 US

62/884,465 2019.08.08 US

62/936,860 2019.11.18 US

(85) PCT国际申请进入国家阶段日

2022.04.07

(86) PCT国际申请的申请数据

PCT/US2020/045506 2020.08.07

(87) PCT国际申请的公布数据

W02021/026506 EN 2021.02.11

(71) 申请人 布里克生物公司

地址 美国马萨诸塞州

(72) 发明人 詹姆斯·塞巴斯蒂安·依塔利亚

权利要求书3页 说明书56页 附图32页

(54) 发明名称

用于非天然氨基酸的位点特异性掺入的氨酰tRNA合成酶和细胞系

(57) 摘要

本发明总的来说涉及工程化改造的tRNA、工程化改造的氨酰tRNA合成酶、非天然氨基酸和包含它们的细胞以及它们在将非天然氨基酸掺入到蛋白质中的用途。

1. 一种原核亮氨酰tRNA合成酶突变蛋白,其能够将非天然氨基酸装载到tRNA上以掺入到蛋白质中,所述tRNA合成酶突变蛋白包含SEQ ID NO:1的氨基酸序列和(i)对应于His537位置处的至少一个替换(例如用疏水氨基酸的替换),(ii)选自E20V、E20M、L41V、L41A、Y499H、Y499A、Y527I、Y527V、Y527G及其任何组合的至少一个氨基酸替换,(iii)选自E20K和L41S及其任何组合的至少一个氨基酸替换和选自M40I、T252A、Y499I和Y527A及其任何组合的至少一个氨基酸替换,或(iv) (i)、(ii)和(iii)中的两者或更多者的组合。

2. 权利要求1所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、M40I、L41S、T252A、Y499I、Y527A或H537G或其组合。

3. 权利要求2所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、M40I、L41S、T252A、Y499I、Y527A和H537G。

4. 权利要求1-3中的任一项所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含在第20位处用Glu或Lys之外的氨基酸的替换。

5. 权利要求4所述的tRNA合成酶突变蛋白,其中所述替换使用疏水氨基酸(例如Leu、Val或Met)。

6. 权利要求5所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20M、M40I、L41S、T252A、Y499I、Y527A和H537G。

7. 权利要求5所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20V、M40I、L41S、T252A、Y499I、Y527A和H537G。

8. 权利要求1-3中的任一项所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含在第41位处用Leu或Ser之外的氨基酸的替换。

9. 权利要求8所述的tRNA合成酶突变蛋白,其中所述替换使用Leu之外的疏水氨基酸(例如Gly、Ala、Val或Met)。

10. 权利要求9所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、M40I、L41V、T252A、Y499I、Y527A和H537G。

11. 权利要求9所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、M40I、L41A、T252A、Y499I、Y527A和H537G。

12. 权利要求1-3中的任一项所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含在第499位处用Tyr、Ile或Ser之外的氨基酸的替换。

13. 权利要求12所述的tRNA合成酶突变蛋白,其中所述替换使用小疏水氨基酸(例如Gly、Ala或Val)。

14. 权利要求13所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、M40I、L41S、T252A、Y499A、Y527A和H537G。

15. 权利要求12所述的tRNA合成酶突变蛋白,其中所述替换使用带正电荷氨基酸(例如Lys、Arg或His)。

16. 权利要求15所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、M40I、L41S、T252A、Y499H、Y527A和H537G。

17. 权利要求1-3中的任一项所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含在第527位处用Ala或Leu之外的疏水氨基酸(例如Gly、Ile、Met或Val)的替换。

18. 权利要求17所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、

M40I、L41S、T252A、Y499I、Y527I和H537G。

19. 权利要求17所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、M40I、L41S、T252A、Y499I、Y527V和H537G。

20. 权利要求17所述的tRNA合成酶突变蛋白,其中所述tRNA合成酶突变蛋白包含E20K、M40I、L41S、T252A、Y499I、Y527G和H537G。

21. 一种核酸,其编码权利要求1-20中的任一项所述的tRNA合成酶突变蛋白。

22. 一种转移载体,其包含权利要求21所述的核酸。

23. 权利要求22所述的转移载体,其中所述载体能够将所述核酸引入到细胞中,随后所述核酸可以整合到所述细胞的基因组中。

24. 一种工程化改造的细胞,其包含权利要求1-20中的任一项所述的tRNA合成酶突变蛋白。

25. 一种工程化改造的细胞,其包含权利要求21所述的核酸。

26. 一种工程化改造的细胞,其包含权利要求22或23所述的转移载体。

27. 权利要求25所述的细胞,其中核酸被稳定整合到所述细胞的基因组中。

28. 权利要求27所述的细胞,其中所述核酸能够在所述细胞中表达以产生相应的tRNA合成酶突变蛋白。

29. 权利要求24-28中的任一项所述的细胞,所述细胞还包含抑制型亮氨酸tRNA,其能够将非天然氨基酸掺入到在所述细胞中进行表达的蛋白质中。

30. 权利要求29所述的细胞,其中所述抑制型亮氨酸tRNA选自SEQ ID NOs:16-42中的任一者。

31. 权利要求29或30所述的细胞,其中编码所述抑制型亮氨酸tRNA的核酸被稳定整合在所述细胞的基因组中。

32. 权利要求31所述的细胞,其中所述核酸能够在所述细胞中表达以产生相应的抑制型tRNA。

33. 权利要求24-32中的任一项所述的细胞,其中所述非天然氨基酸是亮氨酸类似物。

34. 权利要求33所述的细胞,其中所述亮氨酸类似物选自包含炔烃、叠氮化物、环丙烯、烯炔、酮、醛、二氮丙啶或四嗪官能团的直链烷基卤和直链脂族链。

35. 权利要求24-34中的任一项所述的细胞,其中所述蛋白质从包含过早终止密码子的核酸序列表达。

36. 权利要求35所述的细胞,其中所述tRNA合成酶突变蛋白能够将非天然氨基酸装载到抑制型亮氨酸tRNA上,将其在对应于所述过早终止密码子的位置处掺入到所述蛋白质中。

37. 权利要求35所述的细胞,其中所述抑制型tRNA包含反密码子序列,其与所述过早终止密码子杂交,并允许将所述非天然氨基酸在对应于所述过早终止密码子的位置处掺入到所述蛋白质中。

38. 权利要求24-37中的任一项所述的细胞,其中在所述细胞中表达的蛋白质是抗体(或其片段)、双特异性抗体、纳米抗体、亲和体、病毒蛋白、趋化因子、抗原、凝血因子、激素、生长因子、酶或任何其他多肽或蛋白质。

39. 权利要求24-41中的任一项所述的细胞,其中所述细胞是原核细胞(例如细菌细

胞)。

40. 权利要求24-41中的任一项所述的细胞,其中所述细胞是真核细胞(例如哺乳动物细胞)。

41. 一种表达含有非天然氨基酸的蛋白质的方法,所述方法包括将权利要求24-43中的任一项所述的细胞在允许将所述非天然氨基酸掺入到正在所述细胞中表达的蛋白质中的条件下培养或生长。

42. 权利要求44所述的方法,其中所述蛋白质被表达至少5天。

43. 权利要求44所述的方法,其中所述蛋白质被表达至少10天。

44. 权利要求45所述的方法,其中所述蛋白质被表达至少15、20、25、30或35天。

用于非天然氨基酸的位点特异性掺入的氨酰tRNA合成酶和细胞系

[0001] 与相关申请的交叉引用

[0002] 本申请要求2019年8月8日提交的美国临时专利申请号62/884,454、2019年8月8日提交的62/884,465和2019年11月18日提交的62/936,860的利益和优先权,每个所述临时申请为所有目的整体通过参考并入本文。

技术领域

[0003] 本发明总的来说涉及工程化改造的tRNA、工程化改造的氨酰tRNA合成酶、非天然氨基酸和包含它们的细胞以及它们在将非天然氨基酸掺入到蛋白质中的用途。

背景技术

[0004] 在自然界中,蛋白质通过被称为转录和翻译的过程在细胞中产生。在转录期间,包含共同编码目的蛋白的一系列密码子的基因被转录成信使RNA(mRNA)。在翻译过程中,核糖体附着在mRNA上并沿其移动,将特定的氨基酸在对应于密码子的位置处掺入到从所述mRNA合成(翻译)的多肽链中,以产生蛋白质。在翻译过程中,与转运RNA(tRNA)偶联的特定的天然存在的氨基酸进入核糖体。含有反密码子序列的tRNA与它们在mRNA中相应的密码子序列杂交,并在合成蛋白质时将它们携带的氨基酸在适合的位置处转移到新生蛋白质链中。

[0005] 在过去几十年中,已做出重大努力尝试以商业规模的量生产位点特异性修饰的蛋白质例如哺乳动物蛋白质的均质制剂,用于包括例如治疗学和诊断学在内的各种不同应用中。此外,已做出努力尝试在真核细胞(例如哺乳动物细胞)中生产这些修饰的哺乳动物蛋白质,因为所述蛋白质可能更容易以正确折叠和具有完全活性的形式生产,和/或更容易以类似于在哺乳动物细胞中天然生产的本源蛋白质的方式进行翻译后修饰。

[0006] 生产含有位点特异性修饰的蛋白质的一种方法涉及将一个或多个非天然氨基酸(UAA)位点特异性掺入到目的蛋白质中。在体内将UAA位点特异性掺入到蛋白质中的能力已成为增强蛋白质功能或引入在自然界中未发现的新的化学官能团的有力工具。这种技术所需的核心元件包括:工程化改造的tRNA,为所述tRNA装载UAA的工程化改造的氨酰tRNA合成酶(aaRS),以及指导所述UAA在蛋白质合成时掺入到所述蛋白质中的独特密码子,例如终止密码子。

[0007] 这种方法的核心是使用工程化改造的tRNA/aaRS对,其中所述aaRS将所述目的UAA装载到所述tRNA,而不与通常存在于所述表达宿主细胞中的tRNA和氨基酸发生交叉反应。这通过使用源自于与所述表达宿主细胞不同的生命域的生物体的工程化改造的tRNA/aaRS对来实现,以便最大化所述工程化改造的tRNA/aaRS对(例如工程化改造的细菌tRNA/aaRS对)与天然存在于所述表达宿主细胞(例如哺乳动物细胞)中的tRNA/aaRS对之间的正交性。通过所述aaRS装载有所述UAA的工程化改造的tRNA,结合或杂交到存在于编码所述待表达的蛋白质的mRNA中的独特密码子例如过早终止密码子(UAG、UGA、UAA)。参见例如图1,其示出了使用来自于所述表达宿主细胞的内源tRNA和内源aaRS以及被引入到所述宿主细胞中

的工程化改造的正交tRNA和正交aaRS来合成蛋白质,以便于在蛋白质合成时通过核糖体将UAA掺入到所述蛋白质中。到目前为止,已为某些天然存在的氨基酸产生了各种不同的正交tRNA/aaRS对(参见例如美国专利出版物US2017/0349891和Zheng等,(2018) BIOCHEM. 57: 441-445)。所述方法有助于表达含有位点特异性修饰例如生物偶联手柄和光活化交联剂的蛋白质,其可用作治疗剂(例如抗体药物偶联物(ADC)、双特异性抗体(例如双特异性单克隆抗体)、纳米抗体、趋化因子、疫苗、凝血因子、激素和酶)。

[0008] 尽管瞬时转染技术已被用于将工程化改造的tRNA/aaRS对引入到表达宿主细胞中,但不能长时间并以高滴度可重复地表达所述蛋白质,使得这种技术不适合于商业化规模的基于蛋白质的产品的可靠生产。尽管迄今为止已做出努力,但对解决了所需遗传组分的瞬时递送限制的基于哺乳动物细胞的表达平台仍存在着需求,以产生优化的表达系统,用于以高滴度长时间地表达目的蛋白质。

发明内容

[0009] 本公开总的来说涉及在蛋白质合成时使用正交tRNA/氨酰tRNA合成酶对将UAA掺入到目的蛋白质中的领域。本公开涉及用于将非天然氨基酸掺入到蛋白质中的tRNA、氨酰tRNA合成酶和/或非天然氨基酸的优化,并涉及通过基因组或分子生物学工程来构建和优化表达平台(细胞系),用于带有非天然氨基酸的蛋白质的商业规模生产。

[0010] 一方面,本发明提供了一种原核亮氨酰tRNA合成酶突变蛋白,其能够将非天然氨基酸装载到tRNA上以掺入到蛋白质中。所述tRNA合成酶突变蛋白包含SEQ ID NO:1的氨基酸序列和(i)对应于His537位置处的至少一个替换(例如用疏水氨基酸的替换),(ii)选自E20V、E20M、L41V、L41A、Y499H、Y499A、Y527I、Y527V、Y527G及其任何组合的至少一个氨基酸替换,(iii)选自E20K和L41S及其任何组合的至少一个氨基酸替换和选自M40I、T252A、Y499I和Y527A及其任何组合的至少一个氨基酸替换,或(iv) (i)、(ii)和(iii)中的两者或更多者的组合,例如(i)和(ii)、(i)和(iii)、(ii)和(iii)以及(i)、(ii)和(iii)的组合。

[0011] 在某些实施方式中,所述tRNA合成酶突变蛋白可以包含E20K、M40I、L41S、T252A、Y499I、Y527A或H537G或其任何组合(例如所述tRNA合成酶突变蛋白可以包含E20K、M40I、L41S、T252A、Y499I、Y527A和H537G)。

[0012] 在某些实施方式中,所述亮氨酰tRNA合成酶突变蛋白包含在第20位处用Glu或Lys之外的氨基酸的替换,例如用疏水氨基酸(例如Leu、Val或Met)的替换。例如,所述tRNA合成酶突变蛋白可以包含:E20M、M40I、L41S、T252A、Y499I、Y527A和H537G;或E20V、M40I、L41S、T252A、Y499I、Y527A和H537G。

[0013] 在某些实施方式中,所述亮氨酰tRNA合成酶突变蛋白包含在第41位处用Leu或Ser之外的氨基酸的替换,例如用Leu之外的疏水氨基酸(例如Gly、Ala、Val或Met)的替换。例如,所述tRNA合成酶突变蛋白可以包含:E20K、M40I、L41V、T252A、Y499I、Y527A和H537G;或E20K、M40I、L41A、T252A、Y499I、Y527A和H537G。在某些实施方式中,所述tRNA合成酶突变蛋白包含L41V。

[0014] 在某些实施方式中,所述亮氨酰tRNA合成酶突变蛋白包含在第499位处用Tyr、Ile或Ser之外的氨基酸的替换,例如用小疏水氨基酸(例如Gly、Ala或Val)的替换。例如,所述tRNA合成酶突变蛋白可以包含E20K、M40I、L41S、T252A、Y499A、Y527A和H537G。在某些实施

方式中,所述tRNA合成酶突变蛋白包含在第499位处用带正电荷氨基酸(例如Lys、Arg或His)的替换。例如,所述tRNA合成酶突变蛋白可以包含E20K、M40I、L41S、T252A、Y499H、Y527A和H537G。

[0015] 在某些实施方式中,所述tRNA合成酶突变蛋白包含在第527位处用Ala或Leu之外的疏水氨基酸(例如Ile或Val)的替换。例如,所述tRNA合成酶突变蛋白可以包含:E20K、M40I、L41S、T252A、Y499I、Y527I和H537G;E20K、M40I、L41S、T252A、Y499I、Y527V和H537G;或E20K、M40I、L41S、T252A、Y499I、Y527G和H537G。

[0016] 在某些实施方式中,所述亮氨酰tRNA合成酶突变蛋白包含SEQ ID NOs:2-13中的任一者的氨基酸序列。

[0017] 另一方面,本发明提供了一种核酸,其编码上述tRNA合成酶突变蛋白中的任一者。

[0018] 另一方面,本发明提供了一种转移载体,其包含上述核酸中的任一者。在某些实施方式中,所述转移载体能够将所述核酸引入到细胞中。在某些实施方式中,所述转移载体(或来自于所述转移载体的核酸)可以稳定地整合到所述细胞的基因组中。在某些实施方式中,所述转移载体(或来自于所述转移载体的核酸)可以稳定地维持在所述细胞中而不整合到所述细胞的基因组中。

[0019] 另一方面,本发明提供了一种工程化改造的细胞,其包含上述tRNA合成酶突变蛋白中的任一者。

[0020] 另一方面,本发明提供了一种工程化改造的细胞,其包含上述核酸中的任一者,例如其中所述核酸被稳定地整合在所述细胞的基因组中,和/或所述核酸能够在所述细胞中表达,以产生相应的tRNA合成酶突变蛋白。

[0021] 另一方面,本发明提供了一种工程化改造的细胞,其包含上述转移载体中的任一者。在某些实施方式中,所述转移载体(或来自于所述转移载体的核酸)被稳定整合在所述细胞的基因组中。在某些实施方式中,所述转移载体(或来自于所述转移载体的核酸)不被整合在所述细胞的基因组中,但被稳定地维持在所述细胞中。

[0022] 在上述工程化改造的细胞中的任一者的某些实施方式中,所述细胞还包含抑制型亮氨酰tRNA,其能够将非天然氨基酸掺入到在所述细胞中经历表达的蛋白质中。例如,所述抑制型亮氨酰tRNA可以选自SEQ ID NOs:16-42中的任一者。在某些实施方式中,编码所述抑制型亮氨酰tRNA的核酸被稳定整合在所述细胞的基因组中,并且例如所述核酸能够在所述细胞中表达以产生相应的抑制型tRNA。

[0023] 在上述工程化改造的细胞中的任一者的某些实施方式中,所述非天然氨基酸是亮氨酸类似物,例如选自包含炔烃、叠氮化物、环丙烯、烯炔、酮、醛、二氮丙啶、四嗪或任何其他官能团的直链烷基卤和直链脂族链的亮氨酸类似物。

[0024] 在上述工程化改造的细胞中的任一者的某些实施方式中,所述蛋白质从包含过早终止密码子的核酸序列表达,例如,所述tRNA合成酶突变蛋白能够将非天然氨基酸装载到抑制型亮氨酰tRNA上,将其在对应于所述过早终止密码子的位置处掺入到所述蛋白质中。在某些实施方式中,所述抑制型tRNA包含反密码子序列,其与所述过早终止密码子杂交,并允许将所述非天然氨基酸在对应于所述过早终止密码子的位置处掺入到所述蛋白质中。

[0025] 在上述工程化改造的细胞中的任一者的某些实施方式中,将要在所述细胞中表达的蛋白质是抗体(或其片段)、双特异性抗体、纳米抗体、亲和体、病毒蛋白、趋化因子、抗原、

凝血因子、激素、生长因子、酶或任何其他多肽或蛋白质。

[0026] 在上述工程化改造的细胞中的任一者的某些实施方式中,所述细胞是原核细胞(例如细菌细胞)或真核细胞(例如哺乳动物细胞)。

[0027] 另一方面,本发明提供了一种表达含有非天然氨基酸的蛋白质的方法。所述方法包括将上述工程化改造的细胞中的任一者在允许将所述非天然氨基酸掺入到正在所述细胞中表达的蛋白质中的条件下培养或生长。在某些实施方式中,所述蛋白质被表达至少5、10、15、20、25、30或35天。在某些实施方式中,所述蛋白质在所述靶蛋白的初始表达后被表达(例如连续地)至少5、10、15、20、25、30或35天。

[0028] 本发明的这些和其他方面和特点被描述在下面的详细描述和权利要求书中。

附图说明

[0029] 参考下述附图,可以更完整地理解本发明。

[0030] 图1描绘了使用非天然氨基酸(UAA)的遗传密码扩展的示意图。

[0031] 图2A-2C描绘了作为突变体亮氨酰tRNA合成酶的示例性底物的一部分UAA。

[0032] 图3描绘了作为突变体色氨酰tRNA合成酶的示例性底物的一部分UAA。

[0033] 图4示出了亮氨酰tRNA合成酶突变蛋白活性测定的结果,使用荧光活性作为UAA掺入的报告物。图4A中示出了亮氨酰tRNA合成酶突变,图4B中示出了代表终止密码子抑制和UAA掺入的量化的荧光,并且图4C中示出了代表性荧光图像。

[0034] 图5示出了所描绘的亮氨酰tRNA合成酶突变蛋白的多特异性(即单个合成酶掺入不同非天然氨基酸的能力)。图5A描绘了在实施例1中描述的多特异性测定法中使用的UAA,图5B示出了表明多特异性的荧光图像。

[0035] 图6展示了用于产生稳定细胞系的示例性工作流程。图6A是描绘了示例性稳定细胞系产生过程的步骤的流程图,图6B是展示了从稳定转染、使用双荧光报告基因表征到通过克隆分离进行整合的过程的示意图。

[0036] 图7示出了在转染和抗生素选择后稳定的亮氨酰抑制细胞系的荧光激活细胞分选(FACS)合并物分析和克隆分离。群体迁移到45度轴中表示条件GFP信号和UAA掺入的变化。图7A-7C描绘了对照的结果,图7D-7E描绘了通过基于脂质体(LF)的转染获得的稳定细胞系的结果,图7F-7H描绘了通过基于核转染(NF)的转染获得的稳定细胞系的结果。图7I中鉴定了克隆群体。嘌呤霉素(Puro)浓度被注明。

[0037] 图8A和图8B示出了在细胞分选后繁殖2-4周之间,使用条件mCherry-GFP*报告基因通过荧光显微术对克隆分离物进行的重新表征。如第二行图像中所示的GFP荧光表明UAA的成功掺入。

[0038] 图9A-9L描绘的FACS直方图示出了在细胞分选后繁殖2-4周之间使用mCherry-GFP*报告基因对克隆分离物进行的重新表征。图9A-9J是各种不同克隆群体的FACS图。图9K是使用转移载体的瞬时转染对照,图9L展示了来自于所述FACS的样品门选。

[0039] 图10是指定细胞系中的克隆抑制效率和蛋白质表达的比较。图10A描绘了在克隆细胞系中使用条件双报告基因时的平均mCherry和GFP荧光,图10B描绘了在克隆细胞系中使用条件双报告基因时的平均mCherry和GFP荧光的比率,并且图10C描绘了在克隆细胞系中使用条件双报告基因时的mCherry阳性细胞和GFP阳性细胞的百分率的比率。

[0040] 图11是SDS-PAGE考马斯凝胶,展示了示例性亮氨酸稳定细胞系的生产率,其中使用GFP蛋白产量作为抑制活性的读数。

[0041] 图12是在相同的选择条件下使用“野生型”亮氨酸tRNA琥珀抑制基因或h1亮氨酸tRNA琥珀抑制基因产生的稳定细胞系合并物的FACS比较,其中使用条件双报告基因作为读数。图12A-12B是荧光对照,图12C-12D分别示出了使用“野生型”亮氨酸tRNA的结果和使用h1亮氨酸tRNA的结果。图12E描绘了在每个靶门P5和P6中鉴定的所选克隆的数目。

[0042] 图13是在转染和抗生素选择后稳定的色氨酸抑制细胞系的FACS合并物分析和克隆分离。图13A-13B代表荧光对照,图13C-13D示出了用于色氨酸合成酶克隆分离物的嘌呤霉素选择的两种不同条件。

[0043] 图14A描绘了UAA C5AzMe、LCA和AzW。图14B描绘了C5Az的合成路线。图14C描绘了5-AzW的合成路线。图14D-F描绘了LCA的合成路线。

[0044] 详细描述

[0045] 本公开总的来说涉及使用正交tRNA/氨酰tRNA合成酶对将非天然氨基酸掺入到目的蛋白质中的领域。本公开涉及用于将非天然氨基酸掺入到蛋白质中的工程化改造的正交tRNA、工程化改造的氨酰tRNA合成酶和/或非天然氨基酸的优化,并且涉及通过基因组或分子生物学工程进行的表达平台(细胞系)的优化,用于具有非天然氨基酸的蛋白质的商业规模生产。

[0046] 当在本文中使用时,术语“正交”是指被目的表达系统(例如内源细胞内翻译系统)以降低的效率使用的分子(例如正交tRNA或正交氨酰tRNA合成酶)。例如,在目的翻译系统中,正交tRNA被所述目的翻译系统的任何内源氨酰tRNA合成酶以与内源tRNA被内源氨酰tRNA合成酶的氨酰化相比降低的或甚至为零的效率氨酰化。在另一个实例中,在目的翻译系统中,正交氨酰tRNA合成酶以与内源tRNA被内源氨酰tRNA合成酶的氨酰化相比降低的或甚至为零的效率氨酰化任何内源tRNA。

[0047] 下面将更详细地讨论本发明的各个不同特点和方面。

[0048] I. 氨酰tRNA合成酶

[0049] 本发明涉及工程化改造的氨酰tRNA合成酶(或aaRS),其能够将非天然氨基酸装载到tRNA上以掺入到蛋白质中。当在本文中使用时,术语“氨酰tRNA合成酶”是指将或能够将氨基酸(例如非天然氨基酸)装载到tRNA上以掺入到蛋白质中的任何酶或其功能性片段。当在本文中使用时,术语氨酰tRNA合成酶的“功能性片段”是指全长氨酰tRNA合成酶的片段,其保留了相应全长tRNA合成酶(例如天然存在的tRNA合成酶)的例如至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或100%的酶活性。氨酰tRNA合成酶的酶活性可以通过本领域中已知的任何方法来测定。例如,体外氨酰化测定法描述在Hoben等,(1985) *METHODS ENZYMOLOGY* 113:55-59和美国专利申请公开No.2003/0228593中,并且基于细胞的氨酰化测定法描述在美国专利申请公开Nos.2003/0082575和2005/0009049中。在某些实施方式中,所述功能性片段包含在全长tRNA合成酶(例如天然存在的氨酰tRNA合成酶)中存在的至少100、150、200、250、300、350、400、450、500、550、600、650、700、750或800个连续氨基酸。

[0050] 术语氨酰tRNA合成酶包括相对于野生型氨酰tRNA合成酶序列具有一个或多个突变(例如氨基酸替换、缺失或插入)的变体(即突变蛋白)。在某些实施方式中,氨酰tRNA合成

酶突变蛋白可能包含单个突变(例如本文中设想的突变)或2、3、4、5、6、7、8、9、10、11、12、13、14、15个或超过15个突变(例如本文中设想的突变)的组合,由所述突变构成或基本上由所述突变构成。设想了氨酰tRNA合成酶突变蛋白可能包含1-15、1-10、1-7、1-6、1-5、1-4、1-3、1-2、2-15、2-10、2-7、2-6、2-5、2-4、2-3、3-15、3-10、3-7、3-6、3-5或4-10、4-7、4-6、4-5、5-10、5-7、5-6、6-10、6-7、7-10、7-8或8-10个突变(例如本文中设想的突变),由所述突变构成或基本上由所述突变构成。

[0051] 氨酰tRNA合成酶突变蛋白相对于野生型序列或本文描述的序列可能包含保守替换。当在本文中使用时,术语“保守替换”是指使用结构上相似的氨基酸进行的替换。例如,保守替换可以包括下述组内的替换:Ser和Cys;Leu,Ile和Val;Glu和Asp;Lys和Arg;Phe、Tyr和Trp;以及Gln,Asn,Glu,Asp和His。保守替换也可以由BLAST(基本局部比对搜索工具(Basic Local Alignment Search Tool))算法、BLOSUM替换矩阵(例如BLOSUM 62矩阵)或PAM替换:p矩阵(例如PAM 250矩阵)定义。

[0052] 在某些实施方式中,所述氨酰tRNA合成酶突变蛋白相对于相应的(或模板)野生型氨酰tRNA合成酶的底物特异性被改变,使得只有所需的非天然氨基酸而不是常见的20种氨基酸中的任一者被装载到底物tRNA上。

[0053] 氨酰tRNA合成酶可以源自于细菌来源,例如大肠埃希氏杆菌(*Escherichia coli*)、嗜热栖热菌(*Thermus thermophilus*)或嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)。氨酰tRNA合成酶也可以源自于古菌来源,例如源自于甲烷八叠球菌(*Methanosarcinaceae*)或Desulfitobacterium家族,M.barkeri(Mb)、M.alvus(Ma)、M.mazei(Mm)或D.hafnisen(Dh)家族中的任一者,Methanobacterium thermoautotrophicum、Haloflex volcanii、Halobacterium菌种NRC-1或Archaeoglobus fulgidus。在其他实施方式中,也可以使用真核来源,例如植物、藻类、原生生物、真菌、酵母或动物(例如哺乳动物、昆虫、节肢动物等)。当在本文中使用时,术语“衍生物”或“源自于”是指从指定分子或生物体分离或使用来自于指定分子或生物体的信息制造的组分。当在本文中使用时,术语“类似物”是指源自于参比组分(例如野生型tRNA、野生型tRNA合成酶或天然氨基酸)或与参比组分类似(在结构和/或功能方面)的组分(例如tRNA、tRNA合成酶或非天然氨基酸)。在某些实施方式中,衍生物或类似物具有参比或原始组分(例如野生型组分)的给定活性的至少40%、50%、60%、70%、80%、90%、100%或更高的活性。

[0054] 设想了所述氨酰tRNA合成酶可以在体外或体内将底物tRNA氨酰化,并且可以作为多肽或蛋白质或作为编码所述氨酰tRNA合成酶的多核苷酸提供给翻译系统(例如体外翻译系统或细胞)。

[0055] 在某些实施方式中,所述氨酰tRNA合成酶源自于大肠杆菌亮氨酰tRNA合成酶,并且例如所述氨酰tRNA合成酶相对于天然存在的亮氨酸偏好性地用亮氨酸类似物将大肠杆菌亮氨酰tRNA(或其变体)氨酰化。

[0056] 例如,所述氨酰tRNA合成酶可以包含SEQ ID NO:1或与SEQ ID NO:1具有至少85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。在某些实施方式中,所述氨酰tRNA合成酶包含SEQ ID NO:1或其功能性片段或变体,并具有1、2、3、4、5个或更多个下述突变:(i)对应于SEQ ID NO:1的第2位的位置处谷氨酰胺残基的替换,例如用谷氨酸替换(Q2E);(ii)对应于SEQ ID NO:1的第20位的位置处谷氨酸残基的替换,例如用赖氨酸

(E20K)、甲硫氨酸(E20M)或缬氨酸(E20V)替换;(iii)对应于SEQ ID NO:1的第40位的位置处甲硫氨酸残基的替换,例如用异亮氨酸(M40I)或缬氨酸(M40V)替换;(iv)对应于SEQ ID NO:1的第41位的位置处亮氨酸残基的替换,例如用丝氨酸(L41S)、缬氨酸(L41V)或丙氨酸(L41A)替换;(v)对应于SEQ ID NO:1的第252位的位置处苏氨酸残基的替换,例如用丙氨酸(T252A)或精氨酸(T252R)替换;(vi)对应于SEQ ID NO:1的第499位的位置处酪氨酸残基的替换,例如用异亮氨酸(Y499I)、丝氨酸(Y499S)、丙氨酸(Y499A)或组氨酸(Y499H)替换;(vii)对应于SEQ ID NO:1的第527位的位置处酪氨酸残基的替换,例如用丙氨酸(Y527A)、亮氨酸(Y527L)、异亮氨酸(Y527I)、缬氨酸(Y527V)或甘氨酸(Y527G)替换;或(viii)对应于SEQ ID NO:1的第537位的位置处组氨酸残基的替换,例如用甘氨酸(H537G)替换,或上述突变的任何组合。

[0057] 在某些实施方式中,所述氨酰tRNA合成酶包含:(i)对应于SEQ ID NO:1的His537位置处的至少一个替换(例如用疏水氨基酸的替换),(ii)选自E20V、E20M、L41V、L41A、Y499H、Y499A、Y527I、Y527V、Y527G及其任何组合的至少一个氨基酸替换,(iii)选自E20K和L41S及其任何组合的至少一个氨基酸替换和选自M40I、T252A、Y499I和Y527A及其任何组合的至少一个氨基酸替换,或(iv)(i)、(ii)和(iii)中的两者或更多者的组合,例如(i)和(ii)、(i)和(iii)、(ii)和(iii)以及(i)、(ii)和(iii)的组合。

[0058] 在某些实施方式中,所述氨酰tRNA合成酶包含对应于SEQ ID NO:1的第20位的位置处谷氨酸残基的替换,例如用Glu或Lys之外的氨基酸替换,例如用疏水氨基酸(例如Leu、Val或Met)替换。在某些实施方式中,所述氨酰tRNA合成酶包含对应于SEQ ID NO:1的第41位的位置处亮氨酸残基的替换,例如用Leu或Ser之外的氨基酸替换,例如用Leu之外的疏水氨基酸(例如Gly、Ala、Val或Met)替换。在某些实施方式中,所述氨酰tRNA合成酶包含对应于SEQ ID NO:1的第499位的位置处酪氨酸残基的替换,例如用小疏水氨基酸(例如Gly、Ala或Val)替换或用带正电荷氨基酸(例如Lys、Arg或His)替换。在某些实施方式中,所述氨酰tRNA合成酶包含对应于SEQ ID NO:1的第527位的位置处酪氨酸残基的替换,例如用Ala或Leu之外的疏水氨基酸(例如Gly、Ile、Met或Val)替换。在某些实施方式中,所述tRNA合成酶突变蛋白包含L41V。

[0059] 在某些实施方式中,所述氨酰tRNA合成酶包含选自下述的突变的组合:(i)Q2E、E20K、M40I、L41S、T252A、Y499I、Y527A和H537G;(ii)Q2E、E20K、M40V、L41S、T252R、Y499S、Y527L和H537G;(iii)Q2E、M40I、T252A、Y499I、Y527A和H537G;(iv)Q2E、E20M、M40I、L41S、T252A、Y499I、Y527A和H537G;(v)Q2E、E20V、M40I、L41S、T252A、Y499I、Y527A和H537G;(vi)Q2E、E20K、M40I、L41V、T252A、Y499I、Y527A和H537G;(vii)Q2E、E20K、M40I、L41A、T252A、Y499I、Y527A和H537G;(viii)Q2E、E20K、M40I、L41S、T252A、Y499A、Y527A和H537G;(ix)Q2E、E20K、M40I、L41S、T252A、Y499H、Y527A和H537G;(x)Q2E、E20K、M40I、L41S、T252A、Y499I、Y527I和H537G;(xi)Q2E、E20K、M40I、L41S、T252A、Y499I、Y527V和H537G;(xii)Q2E、E20K、M40I、L41S、T252A、Y499I、Y527G和H537G;(xiii)E20K、M40I、L41S、T252A、Y499I、Y527A和H537G;(xiv)E20M、M40I、L41S、T252A、Y499I、Y527A和H537G;(xv)E20V、M40I、L41S、T252A、Y499I、Y527A和H537G;(xvi)E20K、M40I、L41V、T252A、Y499I、Y527A和H537G;(vii)E20K、M40I、L41A、T252A、Y499I、Y527A和H537G;(xviii)E20K、M40I、L41S、T252A、Y499A、Y527A和H537G;(xix)E20K、M40I、L41S、T252A、Y499H、Y527A和H537G;(xx)E20K、M40I、L41S、T252A、

Y499I、Y527I和H537G；(xxi) E20K、M40I、L41S、T252A、Y499I、Y527V和H537G；以及 (xxii) E20K、M40I、L41S、T252A、Y499I、Y527G和H537G。

[0060] 在某些实施方式中，所述氨酰tRNA合成酶包含SEQ ID NOs:2-13中的任一者的氨基酸序列或与SEQ ID NOs:2-13中的任一者具有至少85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

[0061] 在某些实施方式中，所述tRNA合成酶突变蛋白包含SEQ ID NO:14的氨基酸序列，其中 X_2 是Q或E， X_{20} 是E、K、V或M， X_{40} 是M、I或V， X_{41} 是L、S、V或A， X_{252} 是T、A或R， X_{499} 是Y、A、I、H或S， X_{527} 是Y、A、I、L或V，并且 X_{537} 是H或G，并且所述tRNA合成酶突变蛋白相对于SEQ ID NO:1包含至少一个突变(例如2、3、4、5、6、7、8、9个或更多个突变)。在某些实施方式中，所述tRNA合成酶突变蛋白包含SEQ ID NO:15的氨基酸序列，其中 X_{20} 是K、V或M， X_{41} 是S、V或A， X_{499} 是A、I或H，并且 X_{527} 是A、I或V，并且所述tRNA合成酶突变蛋白相对于SEQ ID NO:1包含至少一个突变。

[0062] 在某些实施方式中，所述氨酰tRNA合成酶源自于大肠杆菌色氨酰tRNA合成酶，并且例如所述氨酰tRNA合成酶相对于天然存在的色氨酸偏好性地用色氨酸类似物将大肠杆菌色氨酰tRNA(或其变体)氨酰化。

[0063] 例如，所述氨酰tRNA合成酶可以包含SEQ ID NO:43或与SEQ ID NO:43具有至少85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。在某些实施方式中，所述氨酰tRNA合成酶包含SEQ ID NO:43或其功能性片段或变体，但具有下述突变中的一者或多者：(i) 对应于SEQ ID NO:43的第8位的位置处丝氨酸残基的替换，例如用丙氨酸(S8A)替换；(ii) 对应于SEQ ID NO:43的第144位的位置处缬氨酸残基的替换，例如用丝氨酸(V144S)、甘氨酸(V144G)或丙氨酸(V144A)替换；(iii) 对应于SEQ ID NO:43的第146位的位置处缬氨酸残基的替换，例如与丙氨酸(V146A)、异亮氨酸(V146I)或半胱氨酸(V146C)替换。在某些实施方式中，所述氨酰tRNA合成酶包含选自下述的突变的组合：(i) S8A、V144S和V146A，(ii) S8A、V144G和V146I，(iii) S8A、V144A和V146A，以及(iv) S8A、V144G和V146C。

[0064] 在某些实施方式中，所述氨酰tRNA合成酶包含SEQ ID NOs:44-47中的任一者的氨基酸序列或与SEQ ID NOs:44-47中的任一者具有至少85%、90%、95%、96%、97%、98%或99%序列同一性的氨基酸序列。

[0065] 序列同一性可以以本领域技术人员的技术范围之内的各种不同方式确定，例如使用可公开获得的计算机软件例如BLAST、BLAST-2、ALIGN或Megalign (DNASTAR) 软件。BLAST(基本局部比对搜索工具(Basic Local Alignment Search Tool))分析使用了被程序blastp、blastn、blastx、tblastn和tblastx所利用的算法(Karlin等，(1990) $P_{ROC} \cdot N_{ATL} \cdot A_{CAD} \cdot S_{CI} \cdot USA$ 87:2264-2268; Altschul, (1993) $J. M_{OL} \cdot E_{VOL}$.36:290-300; Altschul等，(1997) $N_{UCLEIC} \cdot A_{CIDS} \cdot R_{ES}$.25:3389-3402, 其通过参考并入本文)，其为序列相似性搜索量身定制。对于搜索序列数据库中的基本问题的讨论，参见Altschul等，(1994) $N_{ATURE} \cdot G_{ENETICS}$ 6:119-129, 其通过参考完全并入本文。本领域技术人员可以确定用于度量比对的适合的参数，包括在待比较的序列的全长上实现最大比对所需的任何算法。用于直方图、描述、期望值(即报告针对数据库序列的匹配的统计显著性阈值)、截止值、矩阵和过滤器的搜索参数采用默认设置。由blastp、blastx、tblastn和tblastx使用的默认评分矩阵是BLOSUM62矩阵(Henikoff等，(1992) $P_{ROC} \cdot N_{ATL} \cdot A_{CAD} \cdot S_{CI} \cdot USA$ 89:10915-10919, 其通过参考完全并入本文)。

四个blastn参数可以如下调整:Q=10(空位创建罚分);R=10(空位扩展罚分);wink=1(在沿着查询序列的每个第wink位置处产生字命中);和gapw=16(设置在其中产生带空位比对的窗口宽度)。等效的blastp参数设置可以是Q=9;R=2;wink=1;和gapw=32。搜索也可以使用NCBI(美国国家生物技术信息中心(National Center for Biotechnology Information))BLAST高级选项参数来进行(例如:-G,创建空位的成本[整数]:默认值=对于核苷酸来说5/对于蛋白质来说11;-E,扩展空位的成本[整数]:默认值=对于核苷酸来说2/对于蛋白质来说1;-q,核苷酸错配的惩罚[整数]:默认值=-3;-r,核苷酸匹配的奖励[整数]:默认值=1;-e,期望值[实数]:默认值=10;-W,字长[整数]:默认值=对于核苷酸来说11/对于megablast来说28/对于蛋白质来说3;-y,以比特为单位的blast扩展的衰减(X):默认值=对于blastn来说20/对于其他来说7;-X,带空位比对的X衰减值(以比特为单位):默认值=对于所有程序来说15,不适用于blastn;和-Z,带空位比对的最终X衰减值(以比特为单位):对于blastn来说50,对于其他来说25)。也可以使用用于成对蛋白质比对的ClustalW(默认参数可以包括例如Blosom62矩阵和空位创建罚分=10和空位扩展罚分=0.1)。可以在GCG软件包10.0版中获得的序列之间的Bestfit比较使用DNA参数GAP=50(空位创建罚分)和LEN=3(空位扩展罚分)。Bestfit蛋白质比较中的等效设置是GAP=8和LEN=2。

[0066] 用于生产蛋白质例如氨酰tRNA合成酶的方法在本领域中是已知的。例如,编码目的蛋白质的DNA分子可以被化学合成或通过重组DNA方法合成。得到的编码所述目的蛋白质的DNA分子可以被连接到其他适合的核苷酸序列,包括例如表达控制序列,以产生编码所需蛋白质的常规基因表达构建物(即表达载体)。确定的基因构建物的生产在本领域的常规技术范围之内。

[0067] 可以将编码所需蛋白质(例如氨酰tRNA合成酶)的核酸并入(连接)到表达载体中,所述表达载体可以通过常规的转染或转化技术引入到宿主细胞中。示例性的宿主细胞是大肠杆菌细胞、中华仓鼠卵巢(CHO)细胞、人胚胎肾293(HEK 293)细胞、HeLa细胞、幼仓鼠肾(BHK)细胞、猴肾细胞(COS)、人肝细胞癌细胞(例如Hep G2)和骨髓瘤细胞。转化的宿主细胞可以在允许所述宿主细胞表达所需蛋白质的条件下培养。

[0068] 具体的表达和纯化方法随着所使用的表达系统而变。例如,如果基因将要在大肠杆菌中表达,则首先通过将所述工程化改造的基因放置在适合的细菌启动子例如Trp或Tac和原核信号序列的下游,将它克隆到表达载体中。所述表达的蛋白质可能被分泌。所述表达的蛋白质可能积累在折射体或包涵体中,它们可以在通过法式压滤法或超声处理破碎细胞后收获。然后将所述折射体溶解,并且可以通过本领域中已知的方法将所述蛋白质重新折叠和/或切割。

[0069] 如果所述工程化改造的基因将要在真核宿主细胞例如CHO细胞中表达,则首先将它插入到含有适合的真核启动子、分泌信号、poly A序列和终止密码子的表达载体中。任选地,所述载体或基因构建物可以含有增强子和内含子。可以使用常规技术将所述基因构建物引入到真核宿主细胞中。

[0070] 目的蛋白质(例如氨酰tRNA合成酶)可以通过将用编码这种蛋白质的表达载体转染的宿主细胞在允许所述蛋白质表达的条件下生长(培养)来产生。在表达后,可以使用本领域中已知的技术例如亲和标签例如谷胱甘肽-S-转移酶(GST)或组氨酸标签来收获并纯化或分离所述蛋白质。

[0071] 用于产生氨酰tRNA合成酶和用于改变所述合成酶的底物特异性的其他方法,可以在下述文献中找到:美国专利申请公开Nos.2003/0108885和2005/0009049,Hamano-Takaku等,(2000) JOURNAL OF BIOLOGICAL CHEMISTRY 275 (51):40324-40328,Kiga等,(2002) PROC.NATL.ACAD.SCI.USA 99 (15):9715-9723,和Francklyn等,(2002) RNA,8:1363-1372。

[0072] 本发明还涵盖了编码本文公开的氨酰tRNA合成酶的核酸。例如,编码本文公开的亮氨酰tRNA合成酶突变蛋白的核苷酸序列描绘在SEQ ID NOs:55-66中。因此,本发明提供了一种核酸,其包含SEQ ID NOs:55-66中的任一项的核苷酸序列,或与SEQ ID NOs:55-66中的任一者具有至少85%、90%、95%、96%、97%、98%或99%序列同一性的核苷酸序列。本发明还提供了一种核酸,其包含编码由SEQ ID NOs:55-66中的任一者编码的氨基酸序列的核苷酸序列,或与由SEQ ID NOs:55-66中的任一者编码的氨基酸序列的编码核苷酸序列具有至少85%、90%、95%、96%、97%、98%或99%序列同一性的核苷酸序列。

[0073] II. tRNA

[0074] 本发明涉及介导非天然氨基酸在蛋白质中的掺入的转运RNA (tRNA)。

[0075] 在蛋白质合成期间,tRNA分子将氨基酸递送到核糖体,以掺入到生长的蛋白质(多肽)链中。tRNA的长度通常为约70至100个核苷酸。有活性的tRNA含有3'CCA序列,其可以在合成期间被转录到所述tRNA中,或者可以在晚些时候在转录后加工期间添加。在氨酰化期间,附连到给定tRNA分子的氨基酸被共价附连到3'-端核糖的2'或3'羟基,以形成氨酰tRNA(aa-tRNA)。应该理解,氨基酸可以自发地从2'-羟基迁移到3'-羟基,反之亦然,但它在核糖体处从3'-OH位置掺入到生长的蛋白质链中。折叠的aa-tRNA分子的另一个末端处的环含有一个被称为反密码子的三个碱基的序列。当这个反密码子序列与核糖体结合的mRNA中互补的三碱基密码子序列杂交或碱基配对时,所述aa-tRNA与所述核糖体结合,并且它的氨基酸被掺入到正被所述核糖体合成的多肽链中。由于与特定密码子碱基配对的所有tRNA均被单一特定氨基酸氨酰化,因此遗传密码的翻译受到tRNA影响。mRNA中的61个非终止密码子中的每一者都指导其同源aa-tRNA的结合,并向正被所述核糖体合成的生长的多肽链添加单一特定氨基酸。术语“同源”是指一起发挥作用的组分,例如tRNA和氨酰tRNA合成酶。

[0076] 抑制型tRNA是在给定翻译系统中改变mRNA的阅读的修饰的tRNA。例如,抑制型tRNA可以通读诸如终止密码子、四碱基密码子或稀有密码子的密码子。使用抑制这个词是基于下述事实,即在某些情况下,所述修饰的tRNA“抑制”mRNA中的密码子的典型表型效应。抑制型tRNA通常在反密码子中含有突变(修饰),改变了密码子特异性,或在改变tRNA的氨酰化身份的某些位置处含有突变(修饰)。术语“抑制活性”是指tRNA例如抑制型tRNA通读在目的系统中不会被内源翻译机制通读的密码子(例如过早终止密码子)的能力。

[0077] 在某些实施方式中,tRNA(例如抑制型tRNA)含有修饰的反密码子区,使得所述修饰的反密码子与相应的天然存在的反密码子杂交不同的密码子。

[0078] 在某些实施方式中,tRNA包含与选自UAG(即“琥珀”终止密码子)、UGA(即“蛋白石”终止密码子)和UAA(即“赭石”终止密码子)的密码子杂交的反密码子。

[0079] 在某些实施方式中,tRNA包含与非标准密码子例如4或5核苷酸密码子杂交的反密码子。4碱基密码子的实例包括AGGA、CUAG、UAGA和CCCU。5碱基密码子的实例包括AGGAC、CCCCU、CCCUC、CUAGA、CUACU和UAGGC。包含与非标准密码子例如4或5核苷酸密码子杂交的反密码子的tRNA和使用此类tRNA将非天然氨基酸掺入到蛋白质中的方法,描述在例如下述文

献中:Moore等,(2000) J.MOL.BIOL.298:195;Hohsaka等,(1999) J.AM.CHEM.SOC.121:12194;Anderson等,(2002) C_{CHEMISTRY AND BIOLOGY} 9:237-244;Magliery (2001) J.MOL.BIOL.307:755-769;和PCT公开No.WO2005/007870。

[0080] 当在本文中使用时,术语“tRNA”包括相对于参比(例如野生型)tRNA序列具有一个或多个突变(例如核苷酸替换、缺失或插入)的变体。在某些实施方式中,tRNA可以包含单一突变(例如本文中设想的突变)或2、3、4、5、6、7、8、9、10、11、12、13、14、15个或超过15个突变(例如本文中设想的突变)的组合,由所述突变构成或基本上由所述突变构成。设想了tRNA可以包含1-15、1-10、1-7、1-6、1-5、1-4、1-3、1-2、2-15、2-10、2-7、2-6、2-5、2-4、2-3、3-15、3-10、3-7、3-6、3-5或3-4个突变(例如本文中设想的突变),由所述突变构成或基本上由所述突变构成。

[0081] 在某些实施方式中,变体抑制型tRNA相对于对应的野生型抑制型tRNA(在这种情况下,野生型抑制型tRNA是指对应于野生型tRNA分子的抑制型tRNA,但对反密码子区进行了提供抑制活性的任何修饰),将非天然氨基酸(例如本文中设想的非天然氨基酸)掺入到哺乳动物蛋白中的活性提高。相对于野生型抑制型tRNA,所述变体抑制型tRNA的活性可以被提高例如约2.5至约200倍、约2.5至约150倍、约2.5至约100倍、约2.5至约80倍、约2.5至约60倍、约2.5至约40倍、约2.5至约20倍、约2.5至约10倍、约2.5至约5倍、约5至约200倍、约5至约150倍、约5至约100倍、约5至约80倍、约5至约60倍、约5至约40倍、约5至约20倍、约5至约10倍、约10至约200倍、约10至约150倍、约10至约100倍、约10至约80倍、约10至约60倍、约10至约40倍、约10至约20倍、约20至约200倍、约20至约150倍、约20至约100倍、约20至约80倍、约20至约60倍、约20至约40倍、约40至约200倍、约40至约150倍、约40至约100倍、约40至约80倍、约40至约60倍、约60至约200倍、约60至约150倍、约60至约100倍、约60至约80倍、约80至约200倍、约80至约150倍、约80至约100倍、约100至约200倍、约100至约150倍或约150至约200倍。

[0082] 设想了所述tRNA可以在体外或体内发挥作用,并且可以作为成熟tRNA(例如氨酰化的tRNA)或作为编码所述tRNA的多核苷酸提供给翻译系统(例如体外翻译系统或细胞)。

[0083] tRNA可以源自于细菌来源,例如大肠埃希氏杆菌(*Escherichia coli*)、嗜热栖热菌(*Thermus thermophilus*)或嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)。tRNA也可以源自于古菌来源,例如源自于甲烷八叠球菌(*Methanosarcinaceae*)或Desulfitobacterium家族,M.barkeri (Mb)、M.alvus (Ma)、M.mazei (Mm)或D.hafnisense (Dh)家族中的任一者,Methanobacterium thermoautotrophicum、Haloferax volcanii、Halobacterium菌种NRC-1或Archaeoglobus fulgidus。在其他实施方式中,也可以使用真核来源,例如植物、藻类、原生生物、真菌、酵母或动物(例如哺乳动物、昆虫、节肢动物等)。

[0084] 在某些实施方式中,所述tRNA源自于大肠杆菌亮氨酰tRNA,并且例如通过源自于大肠杆菌亮氨酰tRNA合成酶的氨酰tRNA合成酶、例如本文中设想的氨酰tRNA合成酶,相对于天然存在的亮氨酸优选地装载亮氨酸类似物。

[0085] 例如,所述tRNA可以包含SEQ ID NOs:16-42中的任一者的核苷酸序列或与SEQ ID NOs:16-42中的任一者具有至少85%、90%、95%、96%、97%、98%或99%序列同一性的核苷酸序列,由所述序列构成或基本上由所述序列构成。

[0086] 在某些实施方式中,所述tRNA源自于大肠杆菌色氨酰tRNA,并且例如通过源自于

大肠杆菌色氨酸tRNA合成酶的氨酰tRNA合成酶、例如本文中设想的氨酰tRNA合成酶，相对于天然存在的色氨酸偏好性地装载色氨酸类似物。

[0087] 例如，所述tRNA可以包含SEQ ID NOs:49-53中的任一者的核苷酸序列或与SEQ ID NOs:49-53中的任一者具有至少85%、90%、95%、96%、97%、98%或99%序列同一性的核苷酸序列，由所述序列构成或基本上由所述序列构成。

[0088] 应该理解，在整个本说明书中，在tRNA包含包括一个或多个胸腺嘧啶(T)的核苷酸序列、基本上由所述序列构成或由所述序列构成的每种情况下，也设想了包含包括尿嘧啶(U)代替一个或多个所述胸腺嘧啶(T)或尿嘧啶(U)代替所有所述胸腺嘧啶(T)的同一核苷酸序列、基本上由所述序列构成或由所述序列构成的tRNA。同样地，在tRNA包含包括一个或多个尿嘧啶(U)的核苷酸序列、基本上由所述序列构成或由所述序列构成的每种情况下，也设想了包含包括胸腺嘧啶(T)代替一个或多个所述尿嘧啶(U)或胸腺嘧啶(T)代替所有所述尿嘧啶(U)的核苷酸序列、基本上由所述序列构成或由所述序列构成的tRNA。

[0089] 用于生产重组tRNA的方法描述在下述文献中：美国专利申请公开Nos.2003/0108885和2005/0009049,Forster等，(2003)PROC.NATL.ACAD.SCI.USA 100(11):6353-6357,和Feng等，(2003),PROC.NATL.ACAD.SCI.USA 100(10):5676-5681。

[0090] 可以通过包括酶法或化学法在内的任何方法将tRNA用所需的非天然氨基酸(UAA)氨酰化(即装载)。

[0091] 能够装载tRNA的酶分子包括氨酰tRNA合成酶，例如本文公开的氨酰tRNA合成酶。能够装载tRNA的其他酶分子包括核酶，例如在下述文献中描述的核酶：Illangakekare等，(1995)SCIENCE 267:643-647;Lohse等，(1996)NATURE 381:442-444;Murakami等，(2003)CHEMISTRY AND BIOLOGY 10:1077-1084;美国专利申请公开No.2003/0228593。

[0092] 化学氨酰化方法包括在下述文献中描述的方法：Hecht(1992)ACC.CHEM.RES.25:545;Heckler等，(1988)BIOTCHEM.1988,27:7254;Hecht等，(1978)J.BIOL.CHEM.253:4517;Cornish等，(1995)ANGEW.CHEM.INT.ED.ENG.34:621;Robertson等，(1991)J.AM.CHEM.SOC.113:2722;Noren等，(1989)SCIENCE 244:182;Bain等，(1989)J.AM.CHEM.SOC.111:8013;Bain等，(1992)NATURE 356:537;Gallivan等，(1997)CHEM.BIOL.4:740;Turcatti等，(1996)J.BIOL.CHEM.271:19991;Nowak等，(1995)SCIENCE 268:439;Saks等，(1996)J.BIOL.CHEM.271:23169;和Hohsaka等，(1999)J.AM.CHEM.SOC.121:34。

[0093] III. 非天然氨基酸(UAA)

[0094] 本发明涉及非天然氨基酸(UAA)和它们在蛋白质中的掺入。

[0095] 当在本文中使用时，非天然氨基酸是指下述20种遗传编码的 α -氨基酸之外的任何氨基酸、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。关于所述20种天然氨基酸的结构，参见例如L.Stryer的《生物化学》(Biochemistry)，第3版，1988, Freeman and Company, New York。术语非天然氨基酸还包括通过天然氨基酸的修饰(例如翻译后修饰)出现，但它们本身不被翻译复合体天然掺入到生长的多肽链中的氨基酸。

[0096] 由于非天然氨基酸与天然氨基酸通常仅在侧链的结构上有差异，因此非天然氨基酸可以例如与其他氨基酸以与在天然存在的蛋白质中形成酰胺键相同的方式形成酰胺键。

然而,非天然氨基酸具有使它们有别于天然氨基酸的侧链基团。例如,所述侧链可以包含烷基、芳基、酰基、酮、叠氮基、羟基、肼、氰基、卤代、酰肼、烯基、烷基、醚、硫醇、硒代、磺酰基、硼酸酯、硼酸基、磷酸基、膦酰基、膦、杂环、烯酮、亚胺、醛、酯、硫代酸、羟胺、胺等或其任何组合。其他非天然存在的氨基酸包括但不限于包含可光活化交联剂的氨基酸、自旋标记的氨基酸、荧光氨基酸、金属结合性氨基酸、含金属氨基酸、放射活性氨基酸、具有新官能团的氨基酸、与其他分子共价或非共价相互作用的氨基酸、光笼化和/或光异构化氨基酸、包含生物素或生物素类似物的氨基酸,糖基化氨基酸例如糖取代的丝氨酸、其他碳水化合物修饰的氨基酸、含酮氨基酸、包含聚乙二醇或聚醚的氨基酸、重原子取代的氨基酸、可化学切割和/或可光解的氨基酸、与天然氨基酸相比具有延长的侧链的氨基酸(包括但不限于聚醚或长链烃,包括但不限于大于约5或大于约10个碳)、碳连接的含糖氨基酸、氧化还原活性氨基酸、含氨基硫代酸的氨基酸和包含一个或多个毒性部分的氨基酸。

[0097] 除了含有新侧链的非天然氨基酸之外,非天然氨基酸还任选地包含修饰的骨架结构。

[0098] 许多非天然氨基酸是基于天然氨基酸例如酪氨酸、谷氨酰胺、苯丙氨酸等。酪氨酸类似物包括对位取代的酪氨酸、邻位取代的酪氨酸和间位取代的酪氨酸,其中所述取代的酪氨酸包含酮基(包括但不限于乙酰基)、苯甲酰基、氨基、肼、羟胺、硫醇基、羧基、异丙基、甲基、 C_6-C_{20} 直链或支链烃、饱和或不饱和烃、0-甲基、聚醚基、硝基等。此外,也设想了多取代的芳环。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、环状衍生物和酰胺取代的谷氨酰胺衍生物。示例性的苯丙氨酸类似物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,其中所述取代基包含羟基、甲氧基、甲基、烯丙基、醛、叠氮基、碘、溴、酮基(包括但不限于乙酰基)等。非天然氨基酸的具体实例包括但不限于对-乙酰基-L-苯丙氨酸、对-炔丙基-苯丙氨酸、0-甲基-L-酪氨酸、L-3-(2-萘基)丙氨酸、3-甲基-苯丙氨酸、0-4-烯丙基-L-酪氨酸、4-丙基-L-酪氨酸、三-0-乙酰基-GlcNAc β -丝氨酸、L-多巴、氟代苯丙氨酸、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、膦酰基丝氨酸、膦酰基酪氨酸、对-碘代-苯丙氨酸、对-溴代苯丙氨酸、对-氨基-L-苯丙氨酸、异丙基-L-苯丙氨酸和对-炔丙氧基-苯丙氨酸等。

[0099] 各种不同的非天然氨基酸的结构实例提供在下述文献中:美国专利申请公开Nos.2003/0082575和2003/0108885,PCT公布No.WO 2002/085923,和Kiick等,(2002) PROC.NATL.ACAD.SCI.USA 99:19-24。

[0100] 多肽中的非天然氨基酸可用于将另一个分子附连到所述多肽,包括但不限于标记物、染料、聚合物、水溶性聚合物、聚乙二醇衍生物、可光活化交联剂、放射性核素、细胞毒性化合物、药物、亲和标记、光亲和标记、反应性化合物、树脂、第二个蛋白质或多肽或多肽类似物、抗体或抗体片段、金属螯合剂、辅因子、脂肪酸、碳水化合物、多核苷酸、DNA、RNA、反义多核苷酸、糖类、水溶性树枝状聚合物、环糊精、抑制性核糖核酸、生物材料、纳米粒子、自旋标记物、荧光团、含金属部分、放射活性部分、新型官能团、与其他分子共价或非共价相互作用的基团、光笼化部分、可激发光化辐射的部分、光异构化部分、生物素、生物素衍生物、生物素类似物、掺有重原子的部分、可化学切割的基团、可光解基团、延长的侧链、碳连接的糖、氧化还原活性剂、氨基硫代酸、毒性部分、同位素标记部分、生物物理探针、磷光基团、化

学发光基团、电子致密基团、磁性基团、嵌入基团、生色团、能量转移剂、生物活性剂、可检测标记物、小分子、量子点、纳米递质、免疫调节分子或上述物质的任何组合。

[0101] 任何适合的非天然氨基酸均可以与本文描述的方法一起使用,以掺入到目的蛋白质中。

[0102] 所述非天然氨基酸可以是亮氨酸类似物。本发明提供了一种图2A中描绘的亮氨酸类似物或包含所述亮氨酸类似物的组合物。例如,图2A中的式A描绘了一种氨基酸类似物,其含有的侧链包括长为n个单元(0-20个单元)的含碳链。O、S、CH₂或NH存在于X位置中,后面可以跟有长为n个单元(0-20个单元)的另一个含碳链。官能团Y附连到第二含碳链的末端碳(例如图2A中所描绘的官能团1-12,其中R表示与所述第二含碳侧链的末端碳原子的连键)。在一个实例中,这些官能团可用于将任何适合的配体生物偶联到适合于位点特异性UAA掺入的任何目的蛋白质。图2A中的式B描绘了一种氨基酸类似物,其含有的侧链被表示为分别附连到所述第二含碳链或第一含碳链的Z-Y₂或Z-Y₃。Z表示包含(CH₂)_n单元的碳链,其中n是0-20的任何整数。Y₂或Y₃独立地可以是与Y₁相同或不同的基团。本发明还提供了图2B中描绘的亮氨酸类似物(LCA、LKET或ACA)或包含图2B中描绘的亮氨酸类似物的组合物。其他示例性亮氨酸类似物包括选自包含官能团例如炔烃、叠氮化物、环丙烯、烯炔、酮、醛、二氮丙啶或四嗪官能团的直链烷基卤和直链脂族链的亮氨酸类似物,以及图2C中示出的结构1-6。然而,可以设想,在所述亮氨酸类似物被掺入到蛋白质或多肽链中时,附连到图2A-2C中示出的任何氨基酸的第一个碳的氨基和羧基两者将构成肽键的部分。

[0103] 在某些实施方式中,所述非天然氨基酸是色氨酸类似物(在本文中也称为衍生物)。示例性的色氨酸类似物包括5-叠氮基色氨酸、5-炔丙氧基色氨酸、5-氨基色氨酸、5-甲氧基色氨酸、5-O-烯丙基色氨酸或5-溴色氨酸。其他示例性色氨酸类似物描绘在图3中。然而,可以设想,在所述色氨酸类似物被掺入到蛋白质或多肽链中时,附连到图3中的色氨酸类似物的第一个碳的氨基和羧基两者将构成肽键的部分。

[0104] 此外,在图14A中阐述的被称为C5AzMe(亮氨酸类似物)、LCA(亮氨酸类似物)和AzW(色氨酸类似物)的UAA可用于本发明的实践。

[0105] C5AzMe(图14B中所示的化合物5)可以以与图14B中概括的合成相似的方式制备。化合物5可以通过例如化合物4的去保护来提供。化合物4的去保护包括除去保护基团(例如Boc)。用于去保护的条件可以包括但不限于HCl,在DCM中。化合物4可以例如通过化合物3在暴露于适合的亲核试剂(例如N₃⁻)时的亲核取代来产生。用于亲核取代的示例性条件包括但不限于NaN₃,在DMF中和80℃下。化合物3可以例如通过化合物1向化合物2的亲核加成来制备。用于亲核加成的示例性条件包括但不限于K₂CO₃,在0℃至RT下。此外,此外,如果需要,化合物5的酯可以通过暴露于温和的水性碱性条件来去除,以产生所述UAA的羧酸形式。

[0106] AzW(图14C中所示的化合物15)可以以与图14C中概括的合成相似的方式制备。化合物15可以例如在碱性条件下从其盐酸盐14制备。示例性的碱性条件包括但不限于KOtBu,在THF中。盐酸盐14可以例如通过化合物13的皂化和随后的去保护来制备。用于皂化和保护基团(例如Boc)去保护的条件对于本领域普通技术人员来说是已知的。例如,皂化可以使用MeOH中的1M NaOH来实现。在某些实施方式中,用于去保护的条件包括但不限于HCl(aq.)。化合物13可以例如通过化合物12与适合的叠氮化合物源的金属介导的偶联来合成。化合物13可以例如使用NaN₃·Cu(OAc)₂,在MeOH中从化合物12制备。化合物12可以例如通过化合物

11的金属催化的硼化,从化合物11制备。用于金属催化的硼化的示例性条件包括但不限于 B_2pin_2 、 $PdCl_2 \cdot dppf$ 和 $KOAc$,在1,4-二噁烷中。化合物11可以例如使用适合的保护基团(例如Boc),通过化合物10的保护来制备。化合物10的保护可以使用 Boc_2O 、 Et_3N 和DMAP,在 CH_2Cl_2 中实现。化合物10可以例如在适合于将肟还原的条件下,例如使用Zn在AcOH中,从化合物9合成。化合物9可以例如通过吡啶8向化合物7的亲核加成来合成。化合物8向化合物7的亲核加成可以在 Na_2CO_3 存在下,在 CH_2Cl_2 中进行。化合物7可以例如通过将化合物6暴露于甲醇中的盐酸羟胺来制备。

[0107] LCA(图14F中所示的化合物21)可以以与图14F中概括的合成相似的方式制备。化合物21可以例如通过将化合物20暴露于适合的酸例如但不限于二噁烷中的4M HCl,从化合物20制备。化合物20可以通过亚胺19的水解来产生。亚胺19的水解可以例如使用1M HCl(aq.),在THF中来实现。化合物19可以例如通过化合物18在暴露于适合的亲核试剂(例如 N_3^-)时的亲核取代来产生。用于亲核取代的示例性条件包括但不限于 NaN_3 ,在DMF中。化合物18可以通过化合物16的烯醇化物向化合物17的亲核加成来制备。用于实现从化合物16和17合成化合物18的适合的条件包括但不限于四丁基硫酸氢铵(TBAHS)和10%NaOH,在DCM中。用于合成LCA的其他方法示出在图14D和14E中。

[0108] 许多非天然氨基酸是可商购的,例如从Sigma-Aldrich(St.Louis,Mo.,USA)、Novabiochem(Darmstadt,Germany)或Peptech(Burlington,Mass.,USA)。不可商购的非天然氨基酸可以使用本领域普通技术人员已知的标准方法来合成。关于有机合成技术,参见例如《有机化学》(Organic Chemistry),Fessenden和Fessenden,(1982,第2版,Willard Grant Press,Boston Mass.);《高等有机化学》(Advanced Organic Chemistry),March(第3版,1985,Wiley and Sons,New York);和《高等有机化学》(Advanced Organic Chemistry),Carey和Sundberg(第3版,部分A和B,1990,Plenum Press,New York)。描述非天然氨基酸的合成的其他示例性出版物出现在下述文献中:PCT公开No.WO2002/085923;美国专利申请公开No.2004/0198637;Matsoukas等,(1995)J.M_{ED}.C_{HEM}.38:4660-4669;King等,(1949)J.C_{HEM}.SOC.3315-3319;Friedman等,(1959)J.A_M.C_{HEM}.SOC.81:3750-3752;Craig等,(1988)J.O_{RG}.C_{HEM}.53:1167-1170;Azoulay等,(1991)E_{UR}.J.M_{ED}.C_{HEM}.26:201-5;Koskinen等,(1989)J.O_{RG}.C_{HEM}.54:1859-1866;Christie等,(1985)J.O_{RG}.C_{HEM}.50:1239-1246;Barton等,(1987)T_{ETRAHEDRON} 43:4297-4308;和Subasinghe等,(1992)J.M_{ED}.C_{HEM}.35:4602-7。

[0109] IV. 载体

[0110] tRNA、氨酰tRNA合成酶或任何其他目的分子,可以通过将编码所述分子的基因并入到适合的表达载体中,在目的细胞中表达。当在本文中使用时,“表达载体”是指一种包含重组多核苷酸的载体,其包含可操作连接到待表达的多核苷酸序列的表达控制序列。表达载体包含足够用于表达的顺式作用元件;其他用于表达的元件可以由宿主细胞或在体外表达系统中提供。

[0111] tRNA、氨酰tRNA合成酶或任何其他目的分子,可以通过将编码所述分子的基因并入到适合的转移载体中,引入到目的细胞中。术语“转移载体”是指一种包含重组多核苷酸的载体,其可用于将所述多核苷酸递送到细胞内部。应该理解,载体可以既是表达载体又是转移载体。

[0112] 载体(例如表达载体或转移载体)包括本领域中已知的所有载体,例如并入目的重

组多核苷酸的粘粒、质粒(例如裸露的或包含在脂质体中的)、反转录转座子(例如背驮、睡美人)和病毒(例如慢病毒、反转录病毒、腺病毒和腺相关病毒)。

[0113] 典型的载体含有可用于调控特定靶核酸的表达的转录和翻译终止子、转录和翻译起始序列和启动子。所述载体任选地包含含有至少一个独立的终止子序列的基因表达盒,允许所述表达盒在真核生物或原核生物或两者(包括但不限于穿梭载体)中复制的序列,和用于原核和真核系统两者的选择标记。

[0114] 在某些实施方式中,所述载体包含可操作连接到编码所述抑制型tRNA和/或tRNA合成酶的核苷酸序列的调控序列或启动子。术语“可操作连接”是指多核苷酸元件以功能性关系连接。当一个核苷酸序列与另一个核苷酸序列以功能性关系放置时,它是“可操作连接的”。例如,如果一个启动子或增强子影响一个基因的转录,则它被可操作连接到所述基因。可操作连接的核苷酸序列通常是毗邻的。然而,由于增强子通常在与启动子相隔几千碱基的情况下起作用并且内含子序列可能具有可变的长度,因此某些多核苷酸元件可能是可操作连接的但并不直接相邻,并且甚至可能从不同等位基因或染色体上反式起作用。

[0115] 可以使用的示例性启动子包括但不限于反转录病毒LTR、SV40启动子、人巨细胞病毒(CMV)启动子、U6启动子、EF1 α 启动子、CAG启动子、H1启动子、UbiC启动子、PGK启动子、7SK启动子、pol II启动子、pol III启动子或任何其他启动子(例如细胞启动子例如真核细胞启动子,包括但不限于组蛋白、pol III和 β -肌动蛋白启动子)。可以使用的其他病毒启动子包括但不限于腺病毒启动子、TK启动子和B19细小病毒启动子。对于本领域技术人员来说,适合的启动子的选择从本文中包含的教导将是显而易见的。在某些实施方式中,载体包含可操作连接到CMV或EF1 α 启动子的编码氨酰tRNA合成酶的核苷酸序列和/或可操作连接到U6或H1启动子的编码抑制型tRNA的核苷酸序列。

[0116] 在某些实施方式中,所述载体是病毒载体。术语“病毒”在本文中用于指称一类没有蛋白质合成或能量产生机制的专性细胞内寄生物。示例性的病毒载体包括反转录病毒载体(例如慢病毒载体)、腺病毒载体、腺相关病毒载体、疱疹病毒载体、epstein-barr病毒(EBV)载体、多瘤病毒载体(例如猴空泡病毒40(SV40)载体)、痘病毒载体和假型病毒载体。

[0117] 所述病毒可以是RNA病毒(具有由RNA构成的基因组)或DNA病毒(具有由DNA构成的基因组)。在某些实施方式中,所述病毒载体是DNA病毒载体。示例性的DNA病毒包括细小病毒(例如腺相关病毒)、腺病毒、非洲猪瘟病毒、疱疹病毒(例如单纯性疱疹病毒1和2(HSV-1和HSV-2)、epstein-barr病毒(EBV)、巨细胞病毒(CMV))、乳头瘤病毒(例如HPV)、多瘤病毒(例如猴空泡病毒40(SV40))和痘病毒(例如痘苗病毒、牛痘病毒、天花病毒、鸡痘病毒、羊痘病毒、粘液瘤病毒)。在某些实施方式中,所述病毒载体是RNA表达载体。示例性的RNA病毒包括布尼亚病毒(例如汉坦病毒)、冠状病毒、黄病毒(例如黄热病病毒、西尼罗病毒、登革热病毒)、肝炎病毒(例如甲型肝炎病毒、丙型肝炎病毒、戊型肝炎病毒)、流感病毒(例如甲型流感病毒、乙型流感病毒、丙型流感病毒)、麻疹病毒、腮腺炎病毒、诺如病毒(例如诺沃克病毒)、脊髓灰质炎病毒、呼吸道合胞病毒(RSV)、反转录病毒(例如人免疫缺陷病毒-1(HIV-1))和环曲病毒。

[0118] 腺相关病毒(AAV)载体

[0119] 在某些实施方式中,载体是腺相关病毒(AAV)载体。AAV是一种小的、无包膜的二十面体病毒,属于细小病毒科和依赖细小病毒属。AAV具有约4.7kb的单链线性DNA基因组。AAV

能够感染几种组织类型的分裂和静止细胞,不同的AAV血清型表现出不同的组织嗜性。

[0120] AAV包括大量血清学可区分的类型,包括血清型AAV-1至AAV-12,以及来自于非人类灵长动物的超过100种血清型(参见例如Srivastava (2008) *J. CELL BIOCHEM.*, 105 (1):17-24, 和Gao等, (2004) *J. VIROLOGY*, 78 (12), 6381-6388)。在本发明中使用的AAV载体的血清型可以由本领域技术人员根据递送效率、组织嗜性和免疫原性来选择。例如,AAV-1、AAV-2、AAV-4、AAV-5、AAV-8和AAV-9可用于递送到中枢神经系统;AAV-1、AAV-8和AAV-9可用于递送到心脏;AAV-2可用于递送到肾脏;AAV-7、AAV-8和AAV-9可用于递送到肝脏;AAV-4、AAV-5、AAV-6、AAV-9可用于递送到肺,AAV-8可用于递送到胰腺,AAV-2、AAV-5和AAV-8可用于递送到感光细胞;AAV-1、AAV-2、AAV-4、AAV-5和AAV-8可用于递送到视网膜色素上皮;AAV-1、AAV-6、AAV-7、AAV-8和AAV-9可用于递送到骨骼肌。在某些实施方式中,所述AAV衣壳蛋白包含在美国专利No.7,198,951中公开的序列,例如但不限于AAV-9(美国专利No.7,198,951的SEQ ID NOs:1-3)、AAV-2(美国专利No.7,198,951的SEQ ID NO:4)、AAV-1(美国专利No.7,198,951的SEQ ID NO:5)、AAV-3(美国专利No.7,198,951的SEQ ID NO:6)和AAV-8(美国专利No.7,198,951的SEQ ID NO:7)。本发明中还考虑到了从恒河猴鉴定到的AAV血清型例如rh.8、rh.10、rh.39、rh.43和rh.74。除了天然的AAV血清型之外,已开发了修饰的AAV衣壳,用于改善递送效率、组织嗜性和免疫原性。示例性的天然和修饰的AAV衣壳公开在美国专利Nos.7,906,111、9,493,788和7,198,951以及PCT公开No.WO2017189964A2中。

[0121] 野生型AAV基因组含有两个145个核苷酸的反向末端重复序列(ITR),其含有指导AAV复制、基因组衣壳化和整合的信号序列。除了ITR之外,三个AAV启动子p5、p19和p40驱动编码rep和cap基因的两个开放阅读框的表达。与单个AAV内含子的差异剪接偶联的两个rep启动子导致从rep基因产生四种rep蛋白(Rep 78、Rep 68、Rep52和Rep 40)。Rep蛋白负责基因组复制。Cap基因从p40启动子表达,并编码三种衣壳蛋白(VP1、VP2和VP3),它们是cap基因的剪接变体。这些蛋白质形成AAV粒子的衣壳。

[0122] 由于用于复制、衣壳化和整合的顺式作用信号包含在所述ITR内,因此4.3kb的内部基因组的一部分或全部可以被外来DNA例如目的外源基因的表达盒代替。因此,在某些实施方式中,所述AAV载体包含的基因组包含两侧带有5' ITR和3' ITR的外源基因的表达盒。所述ITR可以源自于与衣壳相同的血清型或其衍生物。或者,所述ITR可以是与衣壳不同的血清型,从而产生假型AAV。在某些实施方式中,所述ITR源自于AAV-2。在某些实施方式中,所述ITR源自于AAV-5。所述ITR中的至少一者可以被修饰以突变或缺失末端解离位点,从而允许产生自身互补的AAV载体。

[0123] 为了产生AAV载体,所述rep和cap蛋白可以反式提供在例如质粒上。允许AAV复制的宿主细胞系必须表达rep和cap基因、两侧带有ITR的表达盒和由辅助病毒提供的辅助功能例如腺病毒基因E1a、E1b55K、E2a、E4orf6和VA(Weitzman等,《腺相关病毒生物学》(Adeno-associated virus biology),腺相关病毒:方法和方案(Adeno-associated virus:Methods and Protocols),pp.1-23,2011)。用于产生和纯化AAV载体的方法已被详细描述(参见例如Mueller等,(2012) *CURRENT PROTOCOLS IN MICROBIOLOGY*, 14D.1.1-14D.1.21,新的重组腺相关病毒载体的生产和发现(Production and Discovery of Novel Recombinant Adeno-Associated Viral Vectors))。大量细胞类型适用于生产AAV载体,包括HEK293细胞、COS细胞、HeLa细胞、BHK细胞、Vero细胞以及昆虫细胞(参见例如美国专利Nos.6,156,

303,5,387,484,5,741,683,5,691,176,5,688,676和8,163,543,美国专利公开No.20020081721,和PCT公开Nos.W000/47757、W000/24916和W096/17947)。AAV载体通常通过一个含有两侧带有ITR的表达盒的质粒和一个或多个提供其他AAV和辅助病毒基因的另外的质粒,在这些细胞类型中产生。

[0124] 在本发明中可以使用任何血清型的AAV。同样,设想了可以使用任何腺病毒类型,并且本领域技术人员能够鉴定适合于生产所需重组AAV载体(rAAV)的AAV和腺病毒类型。AAV粒子可以例如通过亲和层析、iodixonal梯度或CsCl梯度来纯化。

[0125] AAV载体可以具有尺寸为4.7kb或大于或小于4.7kb的单链基因组,包括大至5.2kb或小至3.0kb的超尺寸基因组。因此,在将从AAV载体表达的目的外源基因小的情况下,AAV基因组可以包含填塞序列。此外,载体基因组可以是基本上自身互补的,从而允许在细胞中快速表达。在某些实施方式中,所述自身互补的AAV载体的基因组从5'至3'包含:5' ITR;第一核酸序列,其包含可操作连接到目的基因的编码序列的启动子和/或增强子;修饰的ITR,其不具有有功能的末端解离位点;第二核酸序列,其与所述第一核酸序列互补或基本上互补;和3' ITR。所有类型的含有基因组的AAV载体均适合用于本发明的方法。

[0126] AAV载体的非限制性实例包括pAAV-MCS(Agilent Technologies)、pAAVK-EF1 α -MCS(System Bio目录号AAV502A-1)、pAAVK-EF1 α -MCS1-CMV-MCS2(System Bio目录号AAV503A-1)、pAAV-ZsGreen1(Clontech目录号6231)、pAAV-MCS2(Addgene质粒编号46954)、AAV-Stuffer(Addgene质粒编号106248)、pAAVscCBPIGpluc(Addgene质粒编号35645)、AAVS1_Puro_PGK1_3xFLAG_Twin_Strep(Addgene质粒编号68375)、pAAV-RAM-d2TTA::TRE-MCS-WPRE-pA(Addgene质粒编号63931)、pAAV-UbC(Addgene质粒编号62806)、pAAVS1-P-MCS(Addgene质粒编号80488)、pAAV-Gateway(Addgene质粒编号32671)、pAAV-Puro_siKD(Addgene质粒编号86695)、pAAVS1-Nst-MCS(Addgene质粒编号80487)、pAAVS1-Nst-CAG-DEST(Addgene质粒编号80489)、pAAVS1-P-CAG-DEST(Addgene质粒编号80490)、pAAVf-EnhCB-lacZnlS(Addgene质粒编号35642)和pAAVS1-shRNA(Addgene质粒编号82697)。这些载体可以被修饰以适合于治疗用途。例如,目的外源基因可以被插入到多克隆位点中,并且选择标记(例如puro或编码荧光蛋白的基因)可以被缺失或用另一个(相同或不同的)目的外源基因代替。AAV载体的其他实例公开在下述文献中:美国专利Nos.5,871,982、6,270,996、7,238,526、6,943,019、6,953,690、9,150,882和8,298,818,美国专利公开No.2009/0087413,和PCT公开Nos.W02017075335A1、W02017075338A2和W02017201258A1。

[0127] 慢病毒载体

[0128] 在某些实施方式中,所述病毒载体可以是反转录病毒载体。反转录病毒载体的实例包括莫洛尼鼠白血病毒载体、脾坏死病毒载体和源自于反转录病毒例如劳斯肉瘤病毒、哈维肉瘤病毒、禽白血病毒、人免疫缺陷病毒、骨髓增殖性肉瘤病毒和乳腺肿瘤病毒的载体。反转录病毒载体可用作介导反转录病毒介导的真核细胞中的基因转移的试剂。

[0129] 在某些实施方式中,所述反转录病毒载体是慢病毒载体。示例性的慢病毒载体包括源自于人免疫缺陷病毒1(HIV-1)、人免疫缺陷病毒2(HIV-2)、猴免疫缺陷病毒(SIV)、猫免疫缺陷病毒(FIV)、牛免疫缺陷病毒(BIV)、杰姆布拉纳病病毒(JDV)、马传染性贫血病毒(EIAV)和山羊关节炎脑炎病毒(CAEV)的载体。

[0130] 反转录病毒载体通常被构建使得大部分编码病毒结构基因的序列被缺失并被

目的基因代替。通常,使用本领域中已知的遗传工程技术将结构基因(即gag、pol和env)从反转录病毒骨架移除。因此,最小的反转录病毒载体从5'至3'包含:5'长末端重复序列(LTR),包装信号,任选的外源启动子和/或增强子,目的外源基因,和3'LTR。如果不提供外源启动子,则基因表达由5'LTR驱动,其是一种弱启动子,并需要Tat的存在来激活表达。用于制造慢病毒的结构基因可以提供在单独的载体中,从而使产生的病毒粒子复制缺陷。具体对于慢病毒来说,包装系统可以包含编码Gag、Pol、Rev和Tat基因的单个包装载体,以及编码包膜蛋白Env(通常为VSV-G,因为它具有广泛的感染性)的第三个独立的载体。为了提高包装系统的安全性,可以将包装载体分离,从一个载体表达Rev,并从另一个载体表达Gag和Pol。通过使用包含嵌合5'LTR的反转录病毒载体,其中所述5'LTR的U3区被异源调控元件代替,也可以从所述包装系统中消除Tat。

[0131] 基因可以以几种通用方式并入到前病毒骨架中。最直接的结构是其中将反转录病毒的结构基因用在LTR内的病毒调控序列的控制之下转录的单个基因代替的结构。还构建了可以将超过一个基因引入到靶细胞中的反转录病毒载体。通常,在此类载体中,一个基因在病毒LTR的调控控制之下,而第二个基因摆脱被剪接的信息表达,或在自身内部启动子的调控之下。

[0132] 因此,所述新的基因两侧带有5'和3'LTR,它们分别用于促进病毒粒子RNA的转录和多聚腺苷酸化。术语“长末端重复序列”或“LTR”是指位于反转录病毒DNA的末端处的碱基对的结构域,它们在其天然序列的背景中是正向重复序列,并含有U3、R和U5区。LTR通常提供对反转录病毒基因的表达(例如基因转录本的促进、起始和多聚腺苷酸化)和病毒复制来说基础的功能。LTR含有大量调控信号,包括转录控制元件、多聚腺苷酸化信号和病毒基因组的复制和整合所需的序列。U3区含有增强子和启动子元件。U5区是引物结合位点与R区之间的区域,并含有多聚腺苷酸化序列。R(重复序列)区在U3与U5区之间。在某些实施方式中,所述R区包含反式激活响应(TAR)遗传元件,其与反式激活(tat)遗传元件相互作用,以增强病毒复制。在其中5'LTR的U3区被异源启动子代替的实施方式中,不需要这个元件。

[0133] 在某些实施方式中,所述反转录病毒载体包含修饰的5'LTR和/或3'LTR。对3'LTR进行修饰通常是为了通过赋予病毒以复制缺陷来提高慢病毒或反转录病毒系统的安全性。在特定实施方式中,所述反转录病毒载体是自失活(SIN)载体。当在本文中使用,时,SIN反转录病毒载体是指一种复制缺陷型反转录病毒载体,其中3'LTR U3区已被修饰(例如通过缺失或替换)以防止第一轮病毒复制之外的病毒转录。这是因为所述3'LTR U3区在病毒复制期间被用作5'LTR U3区的模板,因此在没有U3增强子-启动子的情况下不能制造病毒转录本。在另一个实施方式中,3'LTR被修饰,使得U5区被例如理想的多聚腺苷酸化序列代替。应该指出,对LTR的修饰例如对3'LTR、5'LTR或3'和5'LTR两者的修饰,也包括在本发明中。

[0134] 在某些实施方式中,将5'LTR的U3区用在病毒粒子生产期间驱动病毒基因组转录的异源启动子代替。可以使用的启动子的实例包括例如猿猴病毒40(SV40)(例如早期或晚期)、巨细胞病毒(CMV)(例如立即早期)、莫洛尼鼠白血病病毒(MoMLV)、劳斯肉瘤病毒(RSV)和单纯性疱疹病毒(HSV)(胸苷激酶)启动子。典型的启动子能够以Tat不依赖性方式驱动高水平的转录。这种代替降低了重组以产生可复制病毒的可能性,因为在所述病毒生产系统中没有完整的U3序列。

[0135] 与5'LTR相邻的是基因组的反转录(tRNA引物结合位点)和病毒DNA在粒子中的高

效包装(Psi位点)所必需的序列。当在本文中使用时,术语“包装信号”或“包装序列”是指位于反转录病毒的基因组内,在病毒粒子形成期间反转录病毒RNA链的衣壳化所需的序列(参见例如Clever等,1995 J.V_{IROLOGY}, 69(4):2101-09)。包装信号可以是病毒基因组的衣壳化所需的最小包装信号(也被称为psi[Ψ]序列)。

[0136] 在某些实施方式中,所述反转录病毒载体(例如慢病毒载体)还包含FLAP。当在本文中使用时,术语“FLAP”是指一种其序列包括反转录病毒例如HIV-1或HIV-2的中心多聚嘌呤道和中心终止序列(cPPT和CTS)的核酸。适合的FLAP元件描述在美国专利No.6,682,907和Zennou等,(2000) C_{ELL}, 101:173中。在反转录期间,正链DNA在cPPT处的中心起始和在CTS处的中心终止导致形成三链DNA结构,即中心DNA瓣。尽管不希望受到任何理论限制,所述DNA瓣可以作为慢病毒基因组核输入的顺式活性决定子和/或可以提高病毒的滴度。在特定实施方式中,反转录病毒载体骨架包含在所述载体中目的异源基因的上游或下游的一个或多个FLAP元件。例如,在特定实施方式中,转移质粒包括FLAP元件。在一个实施方式中,本发明的载体包含从HIV-1分离的FLAP元件。

[0137] 在某些实施方式中,所述反转录病毒载体(例如慢病毒载体)还包含输出元件。在一个实施方式中,反转录病毒载体包含一个或多个输出元件。术语“输出元件”是指调控RNA转录本从细胞核向细胞质的运输的顺式作用转录后调控元件。RNA输出元件的实例包括但不限于人免疫缺陷病毒(HIV) RRE(参见例如Cullen等,(1991) J.V_{IROL}. 65:1053;和Cullen等,(1991) C_{ELL} 58:423)和乙型肝炎病毒转录后调控元件(HPRE)。通常,RNA输出元件位于基因的3'UTR内,并且可以作为一个或多个拷贝插入。

[0138] 在某些实施方式中,所述反转录病毒载体(例如慢病毒载体)还包含转录后调控元件。各种不同的转录后调控元件可以提高异源核酸的表达,例如土拨鼠肝炎病毒转录后调控元件(WPRE;参见Zufferey等,(1999) J.V_{IROL}., 73:2886);乙型肝炎病毒中存在的转录后调控元件(HPRE)(Huang等,M_{OL}.C_{ELL}.B_{IOL}., 5:3864)等(Liu等,(1995), G_{ENES} D_{EV}., 9:1766)。转录后调控元件通常位于异源核酸序列的3'末端。这种配置导致合成5'部分包含所述异源核酸编码序列并且3'部分包含所述转录后调控元件序列的mRNA转录本。在某些实施方式中,本发明的载体缺少或不包含转录后调控元件例如WPRE或HPRE,因为在某些情况下这些元件提高细胞转化的风险和/或不实质性或显著增加mRNA转录本的量或提高mRNA稳定性。因此,在某些实施方式中,本发明的载体缺少或不包含WPRE或HPRE作为附加的安全措施。

[0139] 指导异源核酸转录本的高效终止和多聚腺苷酸化的元件提高异源基因表达。转录终止信号通常存在于多聚腺苷酸化信号的下游。因此,在某些实施方式中,所述反转录病毒载体(例如慢病毒载体)还包含多聚腺苷酸化信号。当在本文中使用时,术语“多聚腺苷酸化信号”或“多聚腺苷酸化序列”是指通过RNA聚合酶H指导新生RNA转录本的终止和多聚腺苷酸化两者的DNA序列。重组转录本的高效多聚腺苷酸化的合乎需要的,因为缺少多聚腺苷酸化信号的转录本不稳定并迅速降解。可用于本发明的自体的多聚腺苷酸化信号的说明性实例包括理想多聚腺苷酸化序列(例如AATAAA、ATTAAA、AGTAAA)、牛生长激素多聚腺苷酸化序列(BGHpA)、兔β-珠蛋白多聚腺苷酸化序列(rβgpA)或本领域中已知的另一种适合的异源或内源多聚腺苷酸化序列。

[0140] 在某些实施方式中,反转录病毒载体还包含绝缘子元件。绝缘子元件可能有助于保护反转录病毒表达的序列例如治疗性基因免于整合位点效应,所述效应可以由基因组

DNA中存在的顺式作用元件介导,并导致被转移序列的表达失调(即位置效应;参见例如 Burgess-Beusse等,(2002) $P_{ROC} \cdot N_{ATL} \cdot A_{CAD} \cdot S_{CI}$, USA, 99:16433;和Zhan等,2001, $H_{UM} \cdot G_{ENET}$, 109:471)。在某些实施方式中,所述反转录病毒载体在一个或两个LTR中或所述载体整合到细胞基因组中的区域内的别处包含绝缘子元件。适合用于本发明的绝缘子包括但不限于鸡 β -珠蛋白绝缘子(参见Chung等,(1993) C_{ELL} 74:505;Chung等,(1997) $P_{ROC} \cdot N_{ATL} \cdot A_{CAD} \cdot S_{CI}$, USA 94:575;和Bell等,1999 C_{ELL} 98:387)。绝缘子元件的实例包括但不限于从 β -珠蛋白基因座分离的绝缘子,例如鸡HS4。

[0141] 慢病毒载体的非限制性实例包括pLVX-EF1 α -AcGFP1-C1(Clontech目录号631984)、pLVX-EF1 α -IRES-mCherry(Clontech目录号631987)、pLVX-Puro(Clontech目录号632159)、pLVX-IRES-Puro(Clontech目录号632186)、pLenti6/V5-DESTTM(Thermo Fisher)、pLenti6.2/V5-DESTTM(Thermo Fisher)、pLK0.1(Addgene的质粒编号10878)、pLK0.3G(Addgene的质粒编号14748)、pSico(Addgene的质粒编号11578)、pLJM1-EGFP(Addgene的质粒编号19319)、FUGW(Addgene的质粒编号14883)、pLVTHM(Addgene的质粒编号12247)、pLVUT-tTR-KRAB(Addgene的质粒编号11651)、pLL3.7(Addgene的质粒编号11795)、pLB(Addgene的质粒编号11619)、pWPXL(Addgene的质粒编号12257)、pWPI(Addgene的质粒编号12254)、EF.CMV.RFP(Addgene的质粒编号17619)、pLenti CMV Puro DEST(Addgene的质粒编号17452)、pLenti-puro(Addgene的质粒编号39481)、pULTRA(Addgene的质粒编号24129)、pLX301(Addgene的质粒编号25895)、pHIV-EGFP(Addgene的质粒编号21373)、pLV-mCherry(Addgene的质粒编号36084)、pLionII(Addgene的质粒编号1730)、pInducer10-mir-RUP-PheS(Addgene的质粒编号44011)。这些载体可以被修饰以适合于治疗性用途。例如,可以将选择标记(例如puro、EGFP或mCherry)缺失或用第二个目的外源基因代替。慢病毒载体的其他实例公开在美国专利Nos.7,629,153、7,198,950、8,329,462、6,863,884、6,682,907、7,745,179、7,250,299、5,994,136、6,287,814、6,013,516、6,797,512、6,544,771、5,834,256、6,958,226、6,207,455、6,531,123和6,352,694以及PCT公开No.WO2017/091786中。

[0142] 腺病毒载体

[0143] 在某些实施方式中,所述病毒载体可以是腺病毒载体。腺病毒是中等尺寸(90-100nm)、无包膜的(裸露的)二十面体病毒,由核衣壳和双链线性DNA基因组构成。术语“腺病毒”是指腺病毒属的任何病毒,包括但不限于人类、牛、羊、马、犬、猪、鼠和猿猴腺病毒亚属。通常,腺病毒载体通过在腺病毒的腺病毒基因组中引入一个或多个突变(例如缺失、插入或替换)来产生,以便适应于非本源核酸序列在所述腺病毒中的插入,例如用于基因转移。

[0144] 人类病毒可用作所述腺病毒载体的腺病毒基因组的来源。例如,腺病毒可以是亚群A(例如血清型12、18和31)、亚群B(例如血清型3、7、11、14、16、21、34、35和50)、亚群C(例如血清型1、2、5和6)、亚群D(例如血清型8、9、10、13、15、17、19、20、22-30、32、33、36-39和42-48)、亚群E(例如血清型4)、亚群F(例如血清型40和41)、未分类的亚群(例如血清型49和51)或任何其他腺病毒血清群或血清型。腺病毒血清型1至51可以从美国典型培养物保藏中心(American Type Culture Collection)(ATCC,Manassas,Virginia)获得。非C群腺病毒载体、生产非C群腺病毒载体的方法和使用非C群腺病毒载体的方法公开在例如美国专利Nos.5,801,030、5,837,511和5,849,561以及PCT公开Nos.WO1997/012986和WO1998/053087

中。

[0145] 非人类腺病毒(例如猿、猴、禽、犬、绵羊或牛腺病毒)可用于产生所述腺病毒载体(即作为所述腺病毒载体的腺病毒基因组的来源)。例如,所述腺病毒载体可以基于猴腺病毒,包括新世界和旧世界猴两者(参见例如“病毒分类:国际病毒分类委员会的VHIth报告”(Virus Taxonomy:VHIth Report of the International Committee on Taxonomy of Viruses) (2005))。感染灵长动物的腺病毒的系统发育分析公开在例如Roy等,(2009) *PLoS Pathog.* 5(7):e1000503中。大猩猩腺病毒可用作腺病毒载体的腺病毒基因组的来源。大猩猩腺病毒和腺病毒载体描述在例如PCT公开Nos. W02013/052799、W02013/052811和W02013/052832中。腺病毒载体也可以包含亚型的组合,从而成为“嵌合”腺病毒载体。

[0146] 所述腺病毒载体可以是复制型、条件复制型或复制缺陷型的。复制型腺病毒载体可以在典型的宿主细胞、即通常能够被腺病毒感染的细胞中复制。条件复制型腺病毒载体是已被工程化改造以在预定条件下复制的腺病毒载体。例如,可以将复制必需的基因功能例如由腺病毒早期区域编码的基因功能可操作连接到诱导型、阻遏型或组织特异性转录控制序列,例如启动子。条件复制型腺病毒载体被进一步描述在美国专利No. 5,998,205中。复制缺陷型腺病毒载体是作为例如一个或多个复制必需的基因功能或区域的缺陷的结果而需要补充腺病毒基因组的复制所需的一个或多个基因功能或区域的腺病毒载体,因此所述腺病毒载体不在典型的宿主细胞、特别是将要被所述腺病毒载体感染的人类的细胞中复制。

[0147] 优选地,所述腺病毒载体是复制缺陷的,使得所述复制缺陷型腺病毒载体需要补充所述腺病毒基因组的至少一个复制必需的基因功能或一个或多个区域才能繁殖(例如形成腺病毒载体粒子)。所述腺病毒载体可以仅在腺病毒基因组的早期区域(即E1-E4区)、仅在腺病毒基因组的晚期区域(即L1-L5区域)、在腺病毒基因组的早期和晚期区域两者中缺陷一个或多个复制必需的基因功能,或缺陷所有的腺病毒基因(即高容量腺病毒载体(HC-Ad))。参见例如Morsy等,(1998) *P_{ROC}.N_{ATL}.A_{CAD}.S_{CI}.USA* 95:965-976;Chen等,(1997) *P_{ROC}.N_{ATL}.A_{CAD}.S_{CI}.USA* 94:1645-1650;和Kochanek等,(1999) *H_{UM}.G_{ENE}.T_{HER}*.10(15):2451-9。复制缺陷型腺病毒载体的实例公开在美国专利Nos. 5,837,511、5,851,806、5,994,106、6,127,175、6,482,616和7,195,896以及PCT公开Nos. W01994/028152、W01995/002697、W01995/016772、W01995/034671、W01996/022378、W01997/012986、W01997/021826和W02003/022311中。

[0148] 本发明的复制缺陷型腺病毒载体可以在互补细胞系中生产,所述互补细胞系以适合的水平提供在所述复制缺陷型腺病毒载体中不存在但为病毒繁殖所需的基因功能,以便产生高滴度的病毒载体储用物。这种互补细胞系是已知的,并包括但不限于293细胞(描述在例如Graham等,(1977) *J.G_{EN}.V_{IROL}*.36:59-72中)、PER.C6细胞(描述在例如PCT公开No. W01997/000326和美国专利Nos. 5,994,128和6,033,908中)和293-ORF6细胞(描述在例如PCT公开No. W01995/034671和Brough等,(1997) *J.V_{IROL}*.71:9206-9213中)。用于生产本发明的复制缺陷型腺病毒载体的其他适合的互补细胞系包括为了繁殖编码其表达在宿主细胞中抑制病毒生长的转入基因的腺病毒载体而产生的互补细胞(参见例如美国专利公开No. 2008/0233650)。其他适合的互补细胞描述在例如美国专利Nos. 6,677,156和6,682,929以及PCT公开No. W02003/020879中。含有腺病毒载体的组合物的配方进一步描述在例如美

国专利Nos.6,225,289和6,514,943以及PCT公开No.WO2000/034444中。

[0149] 其他示例性腺病毒载体和/或制备或繁殖腺病毒载体的方法描述在美国专利Nos.5,559,099、5,837,511、5,846,782、5,851,806、5,994,106、5,994,128、5,965,541、5,981,225、6,040,174、6,020,191、6,083,716、6,113,913、6,303,362、7,067,310和9,073,980中。

[0150] 可商购的腺病毒载体系统包括可以从Thermo Fisher Scientific获得的ViraPower™腺病毒表达系统、可以从Agilent Technologies获得的AdEasy™腺病毒载体系统和可以从Takara Bio USA, Inc获得的Adeno-X™表达系统3。

[0151] V. 宿主细胞和细胞系

[0152] 本发明还涵盖了包含本文公开的tRNA、氨酰tRNA合成酶、非天然氨基酸、核酸和/或载体的宿主细胞或细胞系(例如原核或真核宿主细胞或细胞系)。编码所述工程化改造的tRNA和氨酰tRNA合成酶的核酸,可以作为表达宿主细胞中的自主复制载体(例如质粒或病毒粒子)或通过表达宿主细胞例如哺乳动物宿主细胞的基因组中的稳定整合的元件或一系列稳定整合的元件,在所述表达宿主细胞中表达。

[0153] 宿主细胞被遗传工程改造(包括但不限于转化、转导或转染),例如使用本文公开的核酸或载体。例如,在某些实施方式中,一个或多个载体包含正交tRNA、正交氨酰tRNA合成酶和任选的将要通过纳入一个或多个UAA进行修饰的蛋白质的编码区,所述编码区被可操作连接到在所需的宿主细胞或细胞系中有功能的基因表达控制元件。例如,所述编码tRNA合成酶和tRNA和任选的选择标记(例如抗生素抗性基因,例如嘌呤霉素抗性表达盒)的基因可以被整合在转移载体(例如质粒,其在转染前可以被线性化)中,其中例如所述编码tRNA合成酶的基因可以在聚合酶II启动子(例如CMV、EF1α、UbiC或PGK,例如CMV或EF1α)的控制之下,并且所述编码tRNA的基因可以在聚合酶III启动子(例如U6、7SK或H1,例如U6)的控制之下。将所述载体通过包括电穿孔或用病毒载体感染在内的标准方法转染到细胞和/或微生物中,并且可以通过选择标记的表达(例如通过抗生素抗性)来选择克隆。

[0154] 示例性的原核宿主细胞或细胞系包括源自于细菌的细胞,例如大肠埃希氏杆菌(*Escherichia coli*)、嗜热栖热菌(*Thermus thermophilus*)、嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)、荧光假单胞菌(*Pseudomonas fluorescens*)、铜绿假单胞菌(*Pseudomonas aeruginosa*)和恶臭假单胞菌(*Pseudomonas putida*)。示例性的真核宿主细胞或细胞系包括源自于植物(例如复杂植物美丽如单子叶或双子叶植物)、藻类、原生生物、真菌、酵母(包括酿酒酵母(*Saccharomyces cerevisiae*))或动物(包括哺乳动物、昆虫、节肢动物等)的细胞。其他的示例性宿主细胞或细胞系包括HEK293、HEK293T、Expi293、CHO、CHOK1、Sf9、Sf21、HeLa、U2OS、A549、HT1080、CAD、P19、NIH 3T3、L929、N2a、MCF-7、Y79、S0-RB50、HepG2、DUKX-X11、J558L、BHK、COS、Vero、NS0或ESC。应该理解,宿主细胞或细胞系可以包括单个集落、分离的群体(单克隆)或细胞的非均质混合物。

[0155] 设想的细胞或细胞系包含例如正交tRNA/氨酰tRNA合成酶对的一个或多个拷贝,其任选地稳定维持在所述细胞的基因组或由所述细胞维持的另一个DNA片段中。例如,所述细胞或细胞系可以含有一个或多个拷贝的(i)被所述细胞稳定维持的色氨酰tRNA/氨酰tRNA合成酶对(野生型或工程化改造的),和/或(ii)被所述细胞稳定维持的亮氨酰tRNA/氨酰tRNA合成酶对(野生型或工程化改造的)。

[0156] 例如,在某些实施方式中,所述细胞系是稳定细胞系,并且所述细胞系包含的基因组已在其中稳定地整合有(i)编码氨酰tRNA合成酶(例如能够将非天然氨基酸装载到tRNA上的原核色氨酰tRNA合成酶突变蛋白或能够将非天然氨基酸装载到tRNA上的原核亮氨酰tRNA合成酶突变蛋白,例如本文公开的tRNA合成酶突变蛋白)的核酸序列;和/或(ii)编码抑制型tRNA(例如能够装载非天然氨基酸的原核抑制型色氨酰tRNA或能够装载非天然氨基酸的原核抑制型亮氨酰tRNA,例如本文公开的抑制型tRNA)的核酸序列。

[0157] 在某些实施方式中,所述细胞系能够表达所述靶蛋白至少5、10、15、20、25、30或35天(例如当所述细胞被维持在连续培养中时)。在某些实施方式中,所述细胞系能够表达所述靶蛋白5至30天、5至20天、5至10天、10至30天、10至20天或20至30天(例如当所述细胞被维持在连续培养中时)。

[0158] 在某些实施方式中,所述细胞系能够以在相应的细胞系中从缺少过早终止密码子的基因表达的模板蛋白的表达水平的至少30%、40%、50%、60%、70%、80%或90%或约100%的表达水平表达所述靶蛋白,例如,所述细胞系能够以所述表达水平表达所述靶蛋白至少5、10、15、20、25、30或35天(例如当所述细胞被维持在连续培养中时)。

[0159] 将编码tRNA和/或氨酰tRNA合成酶的核酸引入到目的细胞的基因组中或将所述核酸稳定地维持在被所述细胞复制的在所述基因组之外的DNA中的方法,在本领域中是公知的。

[0160] 所述编码tRNA和/或氨酰tRNA合成酶的核酸可以在表达载体、转移载体或DNA表达盒例如本文公开的表达载体、转移载体或DNA表达盒中提供给所述细胞。所述编码tRNA和/或氨酰tRNA合成酶的表达载体、转移载体或DNA表达盒可以含有所述tRNA和/或氨酰tRNA合成酶的一个或多个拷贝,其任选地在具有诱导或组成活性的启动子的控制之下。所述表达载体、转移载体或DNA表达盒可以例如含有其他标准组分(增强子、终止子等)。设想了所述编码tRNA的核酸和编码氨酰tRNA合成酶的核酸可以在相同或不同载体上,可以以相同或不同的比率存在,并且可以被同时或顺序地引入到所述细胞中或稳定整合在细胞基因组中。

[0161] 编码所述tRNA和/或氨酰tRNA合成酶的DNA表达盒的一个或多个拷贝,可以通过转座子系统(例如PiggyBac)、病毒载体(例如慢病毒载体或其他反转录病毒载体)、基于CRISPR/Cas9的重组、电穿孔或自然重组、BxB1重组酶系统或使用复制/维持型DNA片段(例如源自于Epstein-Barr病毒的复制/维持型DNA片段),整合到宿主细胞基因组中或稳定维持在所述细胞中。

[0162] 为了选择稳定维持编码所述tRNA和/或氨酰tRNA合成酶的核酸和/或将UAA高效掺入到目的蛋白质中的细胞系,可以使用选择标记。示例性的选择标记包括吉欧霉素、嘌呤霉素、新霉素、二氢叶酸还原酶(DHFR)、谷氨酰胺合成酶(GS)、mCherry-EGFP融合体或其他荧光蛋白。在某些实施方式中,编码选择标记蛋白的基因(或在选择标记存在下编码可检测的功能例如存活所需的蛋白质的基因)可以包含过早终止密码子,使得只有在所述细胞系能够在所述过早终止密码子的位点处掺入UAA的情况下所述蛋白质才被表达。

[0163] 在某些实施方式中,为了开发包括两个或更多个tRNA/氨酰tRNA合成酶对的宿主细胞或细胞系,人们可以使用多个相同或不同的UAA指导密码子,以便鉴定通过迭代的多轮基因组整合和选择并入所述两个或更多个tRNA/氨酰tRNA合成酶对的多个拷贝的宿主细胞或细胞系。首先可以通过含有一个或多个终止密码子的选择标记来分离含有增强的UAA掺

入效率、低背景且毒性降低的宿主细胞或细胞系。随后,可以对所述宿主细胞或细胞系执行选择计划,以鉴定含有tRNA/氨酰tRNA合成酶对的所需拷贝并表达含有一个或多个终止密码子的目的基因(基因组整合或未整合的)的宿主细胞或细胞系。蛋白质表达可以使用本领域中已知的任何方法来测定,包括例如使用结合目的蛋白质的抗体或C-端标签的Western印迹。

[0164] 将所述宿主细胞或细胞系在常规营养培养基中培养,所述培养基经过改良以适合于诸如筛选步骤、激活启动子或选择转化体等的行动。这些细胞可以被任选地培养成转基因生物体。用于例如细胞分离和培养(例如用于后续核酸分离)的其他有用参考文献包括Freshney(1994),《动物细胞培养基本技术手册》(Culture of Animal Cells,a Manual of Basic Technique),第3版,Wiley-Liss,New York;Payne等,(1992),《液体系统中的植物细胞和组织培养》(Plant Cell and Tissue Culture in Liquid Systems),John Wiley& Sons,Inc.New York,N.Y.;Gamborg和Phillips主编,(1995),《植物细胞、组织和器官培养,基础方法Springer实验室手册》(Plant Cell,Tissue and Organ Culture;Fundamental Methods Springer Lab Manual),Springer-Verlag(Berlin Heidelberg N.Y.);以及Atlas和Parks主编,《微生物培养基手册》(The Handbook of Microbiological Media),(1993)CRC Press,Boca Raton,Fla。

[0165] 能够产生掺入UAA的抗体的示例性细胞系的产生描述在Roy等,(2020)MABS 12(1),e1684749中。产生用于将UAA掺入到蛋白质中的稳定细胞系的其他示例性方法描述在本文的实施例2和图6中。

[0166] 在某些实施方式中,用于将UAA掺入到蛋白质中的稳定细胞系的产生方法包括一个或多个下述步骤:(i)将细胞用编码抑制型tRNA和氨酰tRNA合成酶的一个或多个质粒转染,其中所述一个或多个质粒包含选择标记(例如抗生素抗性基因);(ii)使用所述选择标记选择含有所述一个或多个质粒的细胞;(iii)将所述细胞用在UAA掺入到蛋白质中之后给出可检测信号的报告基因构建物(例如荧光报告基因构建物)瞬时转染;(iv)使用所述报告基因构建物选择能够掺入UAA的细胞;和(v)进一步繁殖所述细胞。在某些实施方式中,所述方法还包括(vi)将细胞用所述报告基因构建物再次瞬时转染,并使用所述报告基因构建物选择具有维持的掺入UAA的能力的细胞。

[0167] VI. 包含非天然氨基酸(UAA)的蛋白质及其制备方法

[0168] 本发明还涵盖了包含非天然氨基酸(UAA)的蛋白质及其制备方法。

[0169] 非天然氨基酸的掺入可以出于各种不同的目的,包括定制蛋白质结构和/或功能的变化,改变尺寸、酸度、亲核性、氢键形成、疏水性、蛋白酶靶位点的可及性,靶向部分(例如用于蛋白质阵列),添加生物活性分子,附连聚合物,附连放射性核素,调节血清半衰期,调节组织穿透(例如肿瘤),调节主动运输,调节组织、细胞或器官特异性或分布,调节免疫原性,调节蛋白酶抗体等。包含非天然氨基酸的蛋白质可以具有增强的或甚至全新的催化或生物物理性质。例如,通过将非天然氨基酸包含在蛋白质中,任选地修改下述性质:毒性,生物分布,结构性质,光谱性质,化学和/或光化学性质,催化能力,半衰期(包括但不限于血清半衰期),与其他分子反应、包括但不限于共价或非共价反应的能力等。包括包含至少一个非天然氨基酸的蛋白质的组合物可用于包括但不限于新的治疗剂、诊断剂、酶和结合蛋白(例如治疗性抗体)。

[0170] 蛋白质可能具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个UAA。所述UAA可以相同或不同。例如,在所述蛋白质中可以存在1、2、3、4、5、6、7、8、9或10个或更多个不同位点,其包含1、2、3、4、5、6、7、8、9或10个或更多个不同UAA。蛋白质可以具有所述蛋白质中存在的至少一个但少于所有的特定氨基酸被所述UAA替换。对于具有超过一个UAA的给定蛋白质来说,所述UAA可以相同或不同(例如,所述蛋白质可以包含两种或更多种不同类型的UAA,或者可以包括两个相同的UAA)。对于具有超过两个UAA的给定蛋白质来说,所述UAA可以相同、不同或者是同一类型的多个非天然氨基酸与至少一个不同UAA的组合。

[0171] 在某些实施方式中,所述蛋白质是抗体(或其片段)、双特异性抗体、纳米抗体、亲和体、病毒蛋白、趋化因子、抗原、凝血因子、激素、生长因子、酶或任何其他多肽或蛋白质。

[0172] 当在本文中使用时,除非另有指明,否则术语“抗体”被理解为意味着完整抗体(例如完整的单克隆抗体)或其片段,例如抗体的Fc片段(例如单克隆抗体的Fc片段)或抗体的抗原结合片段(例如单克隆抗体的抗原结合片段),包括已被修饰、工程化改造或化学偶联的完整抗体、抗原结合片段或Fc片段。抗原结合片段的实例包括Fab、Fab'、(Fab')₂、Fv、单链抗体(例如scFv)、微抗体和双体抗体。已被修饰或工程化改造的抗体的实例包括嵌合抗体、人源化抗体和多特异性抗体(例如双特异性抗体)。化学偶联的抗体的实例是偶联到毒素部分的抗体。

[0173] 可以被修饰以包含一个或多个非天然氨基酸的治疗、诊断和其他蛋白质的其他实例描述在美国专利申请公开Nos.2003/0082575和2005/0009049中。

[0174] 本文公开的tRNA、氨酰tRNA合成酶和/或非天然氨基酸可用于使用任何适合的翻译系统将非天然氨基酸掺入到目的蛋白质中。

[0175] 术语“翻译系统”是指包含将氨基酸掺入到生长的多肽链(蛋白质)中所必需的组分的系统。翻译系统的组分可以包括例如核糖体、tRNA、合成酶、mRNA等。翻译系统可以是细胞或无细胞的,并且可以是原核或真核的。例如,翻译系统可以包括或源自于非真核细胞例如细菌(例如大肠杆菌)、真核细胞例如酵母细胞、哺乳动物细胞、植物细胞、藻类细胞、真菌细胞或昆虫细胞。

[0176] 翻译系统包括宿主细胞或细胞系,例如本文中设想的宿主细胞或细胞系。为了在宿主细胞中表达具有非天然氨基酸的多肽,人们可以将编码所述多肽的多核苷酸克隆到表达载体中,所述表达载体含有例如指导转录的启动子、转录/翻译终止子,以及如果用于编码蛋白质的核酸的话用于翻译起始的核糖体结合位点。

[0177] 翻译系统还包括全细胞制备物,例如可以在其中将所需核酸序列转录成mRNA并表达所述mRNA的通透化的细胞或细胞培养物。无细胞翻译系统是可商购的,并且许多不同类型和系统是公知的。无细胞系统的实例包括但不限于原核裂解物例如大肠埃希氏杆菌裂解物,和真核裂解物例如小麦胚芽提取物、昆虫细胞裂解物、兔网织红细胞裂解物、兔卵母细胞裂解物和人细胞裂解物。也可以使用重构的翻译系统。重构的翻译系统可以包含纯化的翻译因子的混合物以及裂解物的组合或增补有纯化的翻译因子例如起始因子-1(IF-1)、IF-2、IF-3(α 或 β)、延伸因子T(EF-Tu)或终止因子的裂解物。无细胞系统也可以是偶联的转录/翻译系统,其中将DNA引入到所述系统,转录成mRNA并翻译所述mRNA。

[0178] 本发明提供了表达含有非天然氨基酸的蛋白质的方法和在蛋白质中的指定位置

处具有一个或多个非天然氨基酸的蛋白质的生产方法。所述方法包括在允许将所述非天然氨基酸掺入到将要在细胞中表达的蛋白质中的条件下温育翻译系统(例如培养或生长宿主细胞或细胞系,例如本文公开的宿主细胞或细胞系)。所述翻译系统可以在适合于将一个或多个非天然氨基酸掺入到所述蛋白质中的条件下与所述一个或多个非天然氨基酸(例如亮氨酸色氨酰腺类似物)接触(例如可以将细胞培养基与其接触)。

[0179] 在某些实施方式中,所述蛋白质从包含过早终止密码子的核酸序列表达。所述翻译系统(例如宿主细胞或细胞系)可以例如含有亮氨酸tRNA合成酶突变蛋白(例如本文公开的亮氨酸tRNA合成酶突变蛋白),其能够将非天然氨基酸(例如亮氨酸类似物)装载到抑制型亮氨酸tRNA(例如本文公开的抑制型亮氨酸tRNA)上,将其在对应于过早终止密码子的位置处掺入到所述蛋白质中。在某些实施方式中,所述亮氨酸抑制型tRNA包含反密码子序列,其与所述过早终止密码子杂交,并允许所述非天然氨基酸在对应于所述过早终止密码子的位置处掺入到所述蛋白质中。

[0180] 在某些实施方式中,所述蛋白质从包含过早终止密码子的核酸序列表达。所述翻译系统(例如宿主细胞或细胞系)可以例如含有色氨酸tRNA合成酶突变蛋白(例如本文公开的色氨酸tRNA合成酶突变蛋白),其能够将非天然氨基酸(例如色氨酸类似物)装载到抑制型色氨酸tRNA(例如本文公开的抑制型色氨酸tRNA)上,将其在对应于过早终止密码子的位置处掺入到所述蛋白质中。在某些实施方式中,所述色氨酸抑制型tRNA包含反密码子序列,其与所述过早终止密码子杂交,并允许所述非天然氨基酸在对应于所述过早终止密码子的位置处掺入到所述蛋白质中。

[0181] 在整个说明书中,在组合物被描述成具有、包括或包含特定组分的情况下,或在过程和方法被描述成具有、包括或包括特定步骤的情况下,设想了另外存在基本上由所叙述的组分组成或由所述组分组成的本发明的组合物,并且存在基本上由所叙述的过程步骤组成或由所述过程步骤组成的根据本发明的过程和方法。

[0182] 在本申请中,在元件或组分被称为包含在所叙述的元件或组分的列表中和/或选自所述列表的情况下,应该理解所述元件或组分可以是所叙述的元件或组分中的任一者,或者所述元件或组分可以选自所述叙述的元件或组分中的两者或更多者。

[0183] 此外应该理解,本文中描述的组合物或方法的要素和/或特点可以以各种不同的方式组合,而不背离在本文中不论是明示还是暗示的本发明的精神和范围。例如,在提及特定化合物时,除非从上下文另有理解,否则该化合物可用于本发明的组合物的各个不同实施方式和/或本发明的方法中。换句话说,在本申请中,实施方式以能够编写和绘制清晰简洁的申请的方式来描述和描绘,但意在并且应该理解,所述实施方式可以以各种不同方式组合或分离,而不脱离本教导和发明。例如,应该理解,本文中描述和描绘的所有特点均可适用于本文描述和描绘的本发明的所有方面。

[0184] 应该理解,除非从上下文和用法另有理解,否则表述“……中的至少一者”包括在所述表述之后的每一个所叙述的对象以及所述对象中的两者或更多者的组合。除非从上下文另有理解,否则与三个或更多个叙述的对象相关联的表述“和/或”应该被理解为具有相同的含义。

[0185] 术语“包括”、“具有”、“含有”,包括其语法同义语的使用,通常应该被理解为是开放性和非限制性的,例如不排除另外的未叙述的要素或步骤,除非另有明确陈述或从上下

文另有理解。

[0186] 当在定量值之前使用术语“约”时，本发明还包括所述特定定量值本身，除非另有明确陈述。当在本文中使用时，除非另有指明或推断，否则术语“约”是指与标称值相差±10%。

[0187] 应该理解，步骤的顺序或执行某些行动的顺序是无关紧要的，只要本发明仍可操作即可。此外，两个或更多个步骤或行动可以同时进行。

[0188] 本文中使用的任何和所有实例或示例性语言如“例如”或“包括”，仅打算更好地说明本发明，并且不对本发明的范围构成限制，除非提出权利要求。本说明书中的所有语言均不应被解释为表明任何未提出权利要求的要素对于本发明的实践来说是必不可少的。

[0189] 序列表

[0190]

SEQ ID NO	描述	序列
1	大肠埃希氏杆菌 (<i>E. coli</i>) LeuRS	MQEQYRPEEIESKVQLHWDEKRTFEVTEDESKEKYYCLSM PYPGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAF GLPAEGAAVKNNAPAPWTYDNIAYMKNQLKMLGFGYDWS RELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWC PNDQTVLANEQVIDGCCWRC DTKVERKEIPQWFIKITAYADELLND LDKLDHWPDTVKTMQRNWIGRSEGVEITFNVNDYDNTLT VYTTPRPTFMGCTYLAVAAGHPLAQKAAENNP ELAAAFIDECRN TKVAEAEAMATMEKKGVDTGFKAVHPLTGEEIPVWAANFV LMEYGTGAVMAVPGHDQRDYEFA SKYGLNIKPVILAADGSEP DLSQQALTEKGVLFNSGEFNGLDHEAAFN AIADKLTAMGVGERKVN YRLRDWGVSRQRYWGAPIPMVTLEDGTV MPTPDDQLPVILPEDVVM DGITSPIKADPEWAKTTVNGMPALRETD TFDTFMESSWYYARYTCPQYKEGMLDSE AANYWLPVDIYIGGIE HAIMHLLYFRFFHKL MRDAGMVNSDEPAKQLLCQGMVLAD AFYYVGENGERNWVSPVDAIVERDEKGRIV KAKDAAGHEL VYTGMSKMSKSKNNGIDPQVMVERYGADTV RLEFMMFASPAD MTLEWQESGVEGANRFLKRVWKL VYEHTAKGDVAALNVD ALTENQKALRRDVHKTIAKVTD DIGRRQTFNTAIAAIMELMN KLAKAPTDGEQDRALMQEALLAVVRMLNP FTHICFTLWQE LKGEGDIDNAPWPVVADEKAMVEDSTLVV VQVNGKVRAKIT VPVDATEEQVRERAGQEHLVAKYLDGVT VRKVIYVPGKLLN LVVGGPV (SEQ ID NO: 1)
2	LeuRS.v1	MEEQYRPEEIESKVQLHWDKKRTFEVTEDESKEKYYCL SISPYPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQ PIGWDAFGLPAEGAAVKNNAPAPWTYDNIAYMKNQL KMLGFGYDWSRELATCTPEYYRWEQKFFTELYKKGL VYKKTSAVNWC PNDQTVLANEQVIDGCCWRC DTKVERKEIPQWFIKITAYADELLNDL

[0191]

		DKLDHWPDTVKTMQRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNP ELAAFIDE CRNT KVAEAEMATMEKKGVD TGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRD YEFASKYGLNKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVN YRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVM DGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSWIYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEHA IMGLLYFRFFHKL MRDAGMVNSDEPAKQLLCQGMVLADAF YYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHEL VYT GMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADMT LEWQESGVEGANRFLKRVWKL VYEHTAKGDVAALNVDALT ENQKALRRDVHKTI AKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELKG EGDIDNAPWPV ADEKAMVEDSTLVVVQVNGKVRAKITVPV DATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLV GGPV (SEQ ID NO: 2)
3	LeuRS.v2	MEEQYRPEEIESKVQLHWDKKRTFEVTEDESKEKYYCLSVSP YPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFG LPAEGA AVKNNTAPAPWTYDNIA YMKNQLKMLGFGYDWSR ELATCTPEYYRWEQKFFTELYKKGLVYKKTS AVNWCPNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMQRNWIGRSEGVEITFNVDYDNTLTVYT TRPDRFMGCTYLAVAAGHPLAQKAAENNP ELAAFIDE CRNT KVAEAEMATMEKKGVD TGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRD YEFASKYGLNKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVN YRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVM DGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSWSYARYTCPQYKEGMLDSEAANYWLPVDILIGGIEH AIMGLLYFRFFHKL MRDAGMVNSDEPAKQLLCQGMVLADA FYYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHEL VY TGMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADM TLEWQESGVEGANRFLKRVWKL VYEHTAKGDVAALNVDAL TENQKALRRDVHKTI AKVTDDIGRRQTFNTAIAAIMELMNKLA AKAPTDGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELK GEGDIDNAPWPV ADEKAMVEDSTLVVVQVNGKVRAKITVP VDATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLV VGGPV (SEQ ID NO: 3)
4	LeuRS.v3	MEEQYRPEEIESKVQLHWDEKRTFEVTEDESKEKYYCLSI LP YPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFG LPAEGA AVKNNTAPAPWTYDNIA YMKNQLKMLGFGYDWSR ELATCTPEYYRWEQKFFTELYKKGLVYKKTS AVNWCPNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMQRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNP ELAAFIDE CRNT KVAEAEMATMEKKGVD TGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRD YEFASKYGLNKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVN YRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVM DGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSWIYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEHA IMGLLYFRFFHKL MRDAGMVNSDEPAKQLLCQGMVLADAF YYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHEL VYT GMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADMT LEWQESGVEGANRFLKRVWKL VYEHTAKGDVAALNVDALT ENQKALRRDVHKTI AKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELKG EGDIDNAPWPV ADEKAMVEDSTLVVVQVNGKVRAKITVPV

[0192]

		DATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLVV GGPV (SEQ ID NO: 4)
5	V9	MEEQYRPEEIESKVQLHWDMMKRTFEVTEDESKEKYYCLSISP YPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFG LPAEGAAVKNNNTAPAPWTYDNIAYMKNQLKMLGFGYDWSR ELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMQRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPALAFAFIDECRNT KVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAASKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNIAADKLTAMGVGE RKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVMGITSPIKADPEWAKTTVNGMPALRETDFTDT FMESWIYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEHA IMGLLYFRFFHKLMDAGMVNSDEPAKQLLCQGMVLADAF YYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVT GMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADMT LEWQESGVEGANRFLKRVWKLVEHTAKGDVAALNVDALT ENQKALRRDVHKTIKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTHICFTLWQELKG EGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRAKITVPV DATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLVV GGPV (SEQ ID NO: 5)
6	V10	MEEQYRPEEIESKVQLHWDVKRTFEVTEDESKEKYYCLSISP YPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFG LPAEGAAVKNNNTAPAPWTYDNIAYMKNQLKMLGFGYDWSR ELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMQRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPALAFAFIDECRNT KVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAASKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNIAADKLTAMGVGE RKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVMGITSPIKADPEWAKTTVNGMPALRETDFTDT FMESWIYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEHA IMGLLYFRFFHKLMDAGMVNSDEPAKQLLCQGMVLADAF YYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVT GMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADMT LEWQESGVEGANRFLKRVWKLVEHTAKGDVAALNVDALT ENQKALRRDVHKTIKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTHICFTLWQELKG EGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRAKITVPV DATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLVV GGPV (SEQ ID NO: 6)
7	V26	MEEQYRPEEIESKVQLHWDKKRTFEVTEDESKEKYYCLSIVP YPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFG LPAEGAAVKNNNTAPAPWTYDNIAYMKNQLKMLGFGYDWSR ELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMQRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPALAFAFIDECRNT KVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAASKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNIAADKLTAMGVGE RKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL

[0193]

		PVILPEDVVMGDGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSWIYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEHA IMGLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMVLADAF YYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVT GMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADMT LEWQESGVEGANRFLKRVWKLVEHTAKGDVAALNVDALT ENQKALRRDVHKTIKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELKG EGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRAKITVPV DATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLVV GGPV (SEQ ID NO: 7)
8	V29	MEEQYRPEEIESKVQLHWDKRTFEVTEDESKEKYYCLSIAP YPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFG LPAEGAAVKNNAPAPWTYDNIAYMKNQLKMLGFGYDWSR ELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPPELAFFIDECRNT KVAEAEMATMEKKGVDGTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVMGDGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSWIYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEHA IMGLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMVLADAF YYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVT GMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADMT LEWQESGVEGANRFLKRVWKLVEHTAKGDVAALNVDALT ENQKALRRDVHKTIKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELKG EGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRAKITVPV DATEEQVRERAGQEHLVAKYLDGVTVRK VIYVPGKLLNLVVGGPV (SEQ ID NO: 8)
9	V48	MEEQYRPEEIESKVQLHWDKRTFEVTEDESKEKYYCLSISPY PSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFGL PAEGAAVKNNAPAPWTYDNIAYMKNQLKMLGFGYDWSRE LATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPPELAFFIDECRNT KVAEAEMATMEKKGVDGTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVMGDGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSWAYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEH AIMGLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMVLADA FYYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVT TGMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADM TLEWQESGVEGANRFRKRVWKLVEHTAKGDVAALNVDALT ENQKALRRDVHKTIKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELKG EGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRAKITVPV DATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLVV GGPV (SEQ ID NO: 9)
10	V50	MEEQYRPEEIESKVQLHWDKRTFEVTEDESKEKYYCLSISPY PSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFGL

[0194]

		PAEGAAVKNNNTAPAPWPTYDNIAYMKNQLKMLGFGYDWSRE LATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPPELAFAFIDECRNT KVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAASKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVMGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSEWHYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEH AIMGLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMVLADA FYYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVE TGMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADM TLEWQESGVEGANRFLKRVWKLVEHTAKGDVAALNVDAL TENQKALRRDVHKTIAKVTDIGRRQTFNTAIAAIMELMNKL AKAPTDGEQDRALMQEALLAVVRMLNPFTHICFTLWQELK GEGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRKITVP VDATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLV VGGPV (SEQ ID NO: 10)
11	V54	MEEQYRPEEIESKVQLHWDKRTFEVTEDESKEKYYCLSISPY PSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFGL PAEGAAVKNNNTAPAPWPTYDNIAYMKNQLKMLGFGYDWSRE LATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPPELAFAFIDECRNT KVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAASKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVMGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSEWHYARYTCPQYKEGMLDSEAANYWLPVDIAIGGIEHAI MGLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMVLADAFY YVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVEYTG MSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADM TLEWQESGVEGANRFLKRVWKLVEHTAKGDVAALNVDALTE NQKALRRDVHKTIAKVTDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTHICFTLWQELKG EGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRKITVPV VDATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLV VGGPV (SEQ ID NO: 11)
12	V55	MEEQYRPEEIESKVQLHWDKRTFEVTEDESKEKYYCLSISP YPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFG LPAEGAAVKNNNTAPAPWPTYDNIAYMKNQLKMLGFGYDWSR ELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPPELAFAFIDECRNT KVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAASKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVMGITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSEWHYARYTCPQYKEGMLDSEAANYWLPVDIVIGGIEHA IMGLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMVLADAF YVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVEYTG GMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADM TLEWQESGVEGANRFLKRVWKLVEHTAKGDVAALNVDALT

[0195]

		ENQKALRRDVHKTIKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELKG EGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRKITVPV DATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLVV GGPV (SEQ ID NO: 12)
13	V56	MEEQYRPEEIESKVQLHWDKRTFEVTEDESKEKYYCLSISP YPSGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDAFG LPAEGAAVKNNNTAPAPWTYDNIAYMKNQLKMLGFGYDWSR ELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNPNDQT VLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLNDL DKLDHWPDTVKTMQRNWIGRSEGVEITFNVDYDNTLTVYT TRPDAFMGCTYLAVAAGHPLAQKAAENNPAAAFIDECRNT KVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWAANFVLM EYGTGAVMAVPGHDQRDYEFAKYGLNIKPVILAADGSEPD LSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLTAMGVGE RKVNRYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDDQL PVILPEDVVMGDITSPIKADPEWAKTTVNGMPALRETDTFDT FMESSWIYARYTCPQYKEGMLDSEAANYWLPVDIGIGGIEHA IMGLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMVLADAF YYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAGHELVT GMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFASPADMT LEWQESGVEGANRFLKRVWKLVEHTAKGDVAALNVDALT ENQKALRRDVHKTIKVTDDIGRRQTFNTAIAAIMELMNKLA KAPTDGEQDRALMQEALLAVVRMLNPFTPHICFTLWQELKG EGDIDNAPWPVADEKAMVEDSTLVVVQVNGKVRKITVPV DATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGKLLNLVV GGPV (SEQ ID NO: 13)
14		MX ₂ EQYRPEEIESKVQLHWDX ₂₀ KRTFEVTEDESKEKYYCLSX ₄₀ X ₄₁ PYP SGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGW DAFGLPAEGAAVKNNNTAPAPWTYDNIAYMKNQLKMLGFGY DWSRELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWC NPNDQTVLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADE LLNDLKDLDHWPDTVKTMQRNWIGRSEGVEITFNVDYDNT LTVYTTRPD _{X₂₅₂} FMGCTYLAVAAGHPLAQKAAENNPAAAFI DECRNTKVAEAEMATMEKKGVDTGFKAVHPLTGEEIPVWA ANFVLMEYGTGAVMAVPGHDQRDYEFAKYGLNIKPVILAA DGSEPDLSQQALTEKGVLFNSGEFNGLDHEAAFNAIADKLT AMGVGERKVNRYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPT PDDQLPVILPEDVVMGDITSPIKADPEWAKTTVNGMPALRET DTFDTFMESSWX ₄₉₉ YARYTCPQYKEGMLDSEAANYWLPVDI X ₅₂₇ IGGIEHAIX ₅₃₇ LLYFRFFHKLMRDAGMVNSDEPAKQLLC QGMVLADAFYYVGENGERNWVSPVDAIVERDEKGRIVKAK DAAGHELVTGMSKMSKSKNNGIDPQVMVERYGADTVRLF MMFASPADMTLEWQESGVEGANRFLKRVWKLVEHTAKG DVAALNVDALTENQKALRRDVHKTIKVTDDIGRRQTFNTA IAAIMELMNKLAKAPTDGEQDRALMQEALLAVVRMLNPFT PHICFTLWQELKGEGDIDNAPWPVADEKAMVEDSTLVVVQVN GKVRKITVPVDATEEQVRERAGQEHLVAKYLDGVTVRKVI YVPGKLLNLVVGGPV 其中 X ₂ 是 Q 或 E, X ₂₀ 是 E、K、V 或 M, X ₄₀ 是 M、I 或 V, X ₄₁ 是 L、S、V 或 A, X ₂₅₂ 是 T、A 或 R, X ₄₉₉ 是 Y、A、I、H 或 S, X ₅₂₇ 是 Y、A、I、L 或 V, 并且 X ₅₃₇ 是 H 或 G (SEQ ID NO: 14)
15		MEEQYRPEEIESKVQLHWDX ₂₀ KRTFEVTEDESKEKYYCLSIX ₄₁ PYP SGRLHMGHVRNYTIGDVIARYQRMLGKNVLQPIGWDA FGLPAEGAAVKNNNTAPAPWTYDNIAYMKNQLKMLGFGYDW SRELATCTPEYYRWEQKFFTELYKKGLVYKKTSAVNWCNPND QTVLANEQVIDGCCWRCDTKVERKEIPQWFIKITAYADELLN

[0196]

		DLDKLDHWPDTVKTMRNWRSEGVETFNVDYDNTLTV YTRPDAFMGCTYLAVAAGHPLAQKAAENNPelaarAFIDECR NTKVAEAEMATMEKKGVDGTGFAVHPLTGEEIPVWAANFV LMEYGTGAVMAVPGHDQRDYEFASKYGLNIKPVILAADGSE PDLSQQALTEKGVLFNSGEFNGLDHEAAFNIAIDKLTAMGV GERKVNYRLRDWGVSRQRYWGAPIPMVTLEDGTVMPTPDD QLPVILPEDVVMVGITSPIKADPEWAKTTVNGMPALRETDTF DTFMESSWX₄₉₉YARYTCPQYKEGMLDSEAANYWLPVDIX₅₂₇I GGIEHAIMGLLYFRFFHKLMRDAGMVNSDEPAKQLLCQGMV LADAFYYVGENGERNWVSPVDAIVERDEKGRIVKAKDAAG HELVTGMSKMSKSKNNGIDPQVMVERYGADTVRLFMMFA SPADMTLEWQESGVEGANRFLKRVWKLVEHTAKGDVAAL NVDALTENQKALRRDVHKTIAKVTDIGRRQTFNTAIAAIME LMNKLAKAPTDGEQDRALMQEALLAVVRMLNPTPHICFTL WQELKGECDIDNAPWPVADKAMVEDSTLVVVQVNGKVRA KITVPVDATEEQVRERAGQEHLVAKYLDGVTVRKVIYVPGK LLNLVVGGPV 其中 X₂₀ 是 K、V 或 M, X₄₁ 是 S、V 或 A, X₄₉₉ 是 A、I 或 H, 并且 X₅₂₇ 是 A、I 或 V (SEQ ID NO: 15)
16	Leu- tRNA.wtCU A	GCCCGGATGGTGAATCGGTAGACACAAGGGATTctaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG tA (SEQ ID NO: 16)
17	Leu- tRNA.wtUC A	GCCCGGATGGTGAATCGGTAGACACAAGGGATTtcaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG tA (SEQ ID NO: 17)
18	Leu- tRNA.wtUU A	GCCCGGATGGTGAATCGGTAGACACAAGGGATTttaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG tA (SEQ ID NO: 18)
19	Leu- tRNA.h1CU A	GCCCGGATGGTGAATCGGTAGACACAAGGGACTctaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG cA (SEQ ID NO: 19)
20	Leu- tRNA.h1UC A	GCCCGGATGGTGAATCGGTAGACACAAGGGACTtcaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG cA (SEQ ID NO: 20)
21	Leu- tRNA.h1UU A	GCCCGGATGGTGAATCGGTAGACACAAGGGACTttaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG cA (SEQ ID NO: 21)
22	Leu- tRNA.wtCU A_cca	GCCCGGATGGTGAATCGGTAGACACAAGGGATTctaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG tACCA (SEQ ID NO: 22)
23	Leu- tRNA.wtUC A_cca	GCCCGGATGGTGAATCGGTAGACACAAGGGATTtcaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG tACCA (SEQ ID NO: 23)
24	Leu- tRNA.wtUU A_cca	GCCCGGATGGTGAATCGGTAGACACAAGGGATTttaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG tACCA (SEQ ID NO: 24)
25	Leu- tRNA.h1CU A_cca	GCCCGGATGGTGAATCGGTAGACACAAGGGACTctaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG cACCA

[0197]

		(SEQ ID NO: 25)
26	Leu-tRNA.h1UC A_cca	GCCCGGATGGTGGGAATCGGTAGACACAAGGGACTtcaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG cACCA (SEQ ID NO: 26)
27	Leu-tRNA.h1UU A_cca	GCCCGGATGGTGGGAATCGGTAGACACAAGGGACTttaAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGGG cACCA (SEQ ID NO: 27)
28	Leu-tRNA.003.3 0	GCCCGGATGGTGGGAATCGGTAGACACAAGGGACTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCCGG GCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 27)
29	Leu-tRNA.003.3 1	GGGCGTGTGGTGGGAATCGGTAGACACAAGGGATTnnnAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCGCGCC CACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 28)
30	Leu-tRNA.003.3 2	GGGCGCGTGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCGCGC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 29)
31	Leu-tRNA.003.3 3	GGGCATGTGGTGGGAATCGGTAGACACAAGGGATTnnnAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCATGCC CACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 30)
32	Leu-tRNA.003.3 4	GGGCACGTGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCGTGC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 31)
33	Leu-tRNA.003.3 5	GGGGTGTGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCGCCC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 32)
34	Leu-tRNA.003.3 6	GGGGGCGTGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCGTCC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 33)
35	Leu-tRNA.003.3 7	GGGGATGTGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCGTCC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 34)
36	Leu-tRNA.003.3 8	GGGGACGTGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCGTCC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 35)
37	Leu-tRNA.003.3 9	GCCCGTATGGTGGGAATCGGTAGACACAAGGGATTnnnAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTGCGG GCACCA 其中 nnn 是 CTA、TCA 或 TTA

[0198]

		(SEQ ID NO: 36)
38	Leu-tRNA.003.40	GGGATAGTGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCTATC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 37)
39	Leu-tRNA.003.41	GGGCATGTGGTGGGAATCGGTAGACACAAGGGATTnnnAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCGTGCC CACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 38)
40	Leu-tRNA.003.42	GGGCAGATGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTCTGC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 39)
41	Leu-tRNA.003.43	GGGCGTATGGTGGGAATCGGTAGACACAAGGGATTnnnAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTGCGCC CACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 40)
	Leu-tRNA.003.44	GGGCAAGTGGTGGGAATCGGTAGACACAAGGGATTnnnAAT CCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCCTTGC CCACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 41)
42	Leu-tRNA.003.45	GCACACATGGTGGGAATCGGTAGACACAAGGGATTnnnAATC CCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCCCGCTGTGTG CACCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 42)
43	大肠埃希氏杆菌 (<i>E. coli</i>) TrpRS.wt	MTKPIVFSGAQPSGELTIGNYMGALRQWVNMQDDYHCIYCI VDQHAITVRQDAQKLRKATLDTLALYLACGIDPEKSTIFVQS HVPEHAQLGWALNCYTYFGELSRMTQFKDKSARYAENINAG LFDYPVLMAADILLYQTNLVPVGEDQKQHLELSRDIAQRFNA LYGEIFKVPEPFIPKSGARVMSLLEPTKKMSKSDDNRNNVIGL LEDPKSVVKKIKRAVTDSEPPVVRYDVQNKAGVSNLLDILS AVTGQSIPELEKQFEGKMYGHLKGEVADAVSGMLTELQERY HRFRNDEAFLQQVMKDGAEKASAHASRTLKAVYEAIQFVAK P (SEQ ID NO: 43)
44	TrpRS.h14	MTKPIVFAGAQPSGELTIGNYMGALRQWVNMQDDYHCIYCI VDQHAITVRQDAQKLRKATLDTLALYLACGIDPEKSTIFVQS HVPEHAQLGWALNCYTYFGELSRMTQFKDKSARYAENINAG LFDYPVLMAADILLYQTNLGPGCEDQKQHLELSRDIAQRFNA LYGEIFKVPEPFIPKSGARVMSLLEPTKKMSKSDDNRNNVIGL LEDPKSVVKKIKRAVTDSEPPVVRYDVQNKAGVSNLLDILS AVTGQSIPELEKQFEGKMYGHLKGEVADAVSGMLTELQERY HRFRNDEAFLQQVMKDGAEKASAHASRTLKAVYEAIQFVAK P (SEQ ID NO: 44)
45	TrpRS.h13	MTKPIVFAGAQPSGELTIGNYMGALRQWVNMQDDYHCIYCI VDQHAITVRQDAQKLRKATLDTLALYLACGIDPEKSTIFVQS HVPEHAQLGWALNCYTYFGELSRMTQFKDKSARYAENINAG LFDYPVLMAADILLYQTNLAPAGEDQKQHLELSRDIAQRFNA LYGEIFKVPEPFIPKSGARVMSLLEPTKKMSKSDDNRNNVIGL LEDPKSVVKKIKRAVTDSEPPVVRYDVQNKAGVSNLLDILS AVTGQSIPELEKQFEGKMYGHLKGEVADAVSGMLTELQERY HRFRNDEAFLQQVMKDGAEKASAHASRTLKAVYEAIQFVAK

[0199]

		P (SEQ ID NO: 45)
46	TrpRS.h9	MTKPIVFAGAQPSGELTIGNYMGALRQWVNMQDDYHCIYCI VDQHAITVRQDAQKLRKATLDTLALYLACGIDPEKSTIFVQS HVPEHAQLGWALNCYTYFGELSRMTQFKDKSARYAENINAG LFDYPVLMAADILLYQTNLSPAGEDQKQHLELSRDIAQRFNA LYGEIFKVPEPFIPKSGARVMSLLEPTKKMSKSDDNRNNVIGL LEDPKSVVKKIKRAVTDSEPPVVRVDVQNKAGVSNLLDILS AVTGQSIPELEKQFEGKMYGHLKGEVADAVSGMLTELQERY HRFRNDEAFLQQVMKDGAEKASAHASRTLKAVYEAIGFVAK P (SEQ ID NO: 46)
47	TrpRS.h10	MTKPIVFAGAQPSGELTIGNYMGALRQWVNMQDDYHCIYCI VDQHAITVRQDAQKLRKATLDTLALYLACGIDPEKSTIFVQS HVPEHAQLGWALNCYTYFGELSRMTQFKDKSARYAENINAG LFDYPVLMAADILLYQTNLGPAGEDQKQHLELSRDIAQRFNA LYGEIFKVPEPFIPKSGARVMSLLEPTKKMSKSDDNRNNVIGL LEDPKSVVKKIKRAVTDSEPPVVRVDVQNKAGVSNLLDILS AVTGQSIPELEKQFEGKMYGHLKGEVADAVSGMLTELQERY HRFRNDEAFLQQVMKDGAEKASAHASRTLKAVYEAIGFVAK P (SEQ ID NO: 47)
48	大肠埃希氏 杆菌 (<i>E. coli</i>) Trp- tRNA	AGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTccaAAACC GGGTGTTGGGAGTTCGAGTCTCTCCGCCCCTGCCA (SEQ ID NO: 48)
49	Trp- tRNA _{CUA}	AGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTctaAAACC GGGTGTTGGGAGTTCGAGTCTCTCCGCCCCTGCCA (SEQ ID NO: 49)
50	Trp- tRNA _{UCA}	AGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTtcaAAACC GGGTGTTGGGAGTTCGAGTCTCTCCGCCCCTGCCA (SEQ ID NO: 50)
51	Trp- tRNA _{UUA}	AGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTttaAAACC GGGTGTTGGGAGTTCGAGTCTCTCCGCCCCTGCCA (SEQ ID NO: 51)
52	Trp- tRNA.001.4	TGGGGTATCGCCAAGCGGTAAGGCACCGGATTcnnnTTCCG GCATTCCGAGGTTTGAATCCTCGTACCCAGCCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 52)
53	Trp- tRNA.001.5	AGGGGCATAGCTCAAGCGGTAAAGCACCGGACTnnnAAAC CGGCAGTCCGAAGTTCGAATCCCCCAGCCAGCCA 其中 nnn 是 CTA、TCA 或 TTA (SEQ ID NO: 53)
54	大肠埃希氏 杆菌 (<i>E. coli</i>) LeuRS	ATGCAGGAACAGTATCGCCCGGAAGAAATTGAAAGCAA GTGCAGCTGCATTGGGATGAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCAT GCTGCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATAACATTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCCTGCCGCGGAAGGCGCGCGGTGAAAAAC AACACCGCGCCGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACAGCTGAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATAACCAAAGTGGAACGCAAAG AAATTCCGCAGTGGTTTATTAAAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG

[0200]

		ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCGCAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACCACCGCCCGGATACCTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAAGTGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGGCGGAAGCGG AAATGGCGACCATGGAAAAAAAAGGCGTGGATACCGGCT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCC ATTAAGCGGATCCGGAATGGGCGAAAAACACCGTGAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT TTATGGAAGCAGCTGGTATTATGCGCGCTATACCTGCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTTATATTGGCGGCATTGAAC ATGCGATTATGCATCTGCTGTATTTTCGCTTTTTTCATAAA CTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCGG CGAAACAGCTGCTGTGCCAGGGCATGGTGTGGCGGATGC GTTTTATTATGTGGGCGAAAAACGGCGAACGCAACTGGGTG AGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGGCC GCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTGGT GTATACCGCATGAGCAAAATGAGCAAAAGCAAAAACAA CGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCGCG GATACCGTGCGCCTGTTTATGATGTTTGCGAGCCCGGCGG ATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGCG CGAACCGCTTTCTGAAACGCGTGTGGAAGTGGTGTATGA ACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGGAT GCGCTGACCGAAAAACCAGAAAGCGCTGCGCCGCGATGTGC ATAAAACCATTTGCGAAAGTGACCGATGATATTGGCCGCCG CCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAGTGA ATGAACAACTGGCGAAAGCGCCGACCGATGGCGAACAG GATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGTGC GCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCCTG TGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGCG CCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAAGAT AGCACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCAGCG CGAAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAGGT GCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAATAT CTGGATGGCGTGACCGTGCGCAAAGTGATTTATGTGCCGG GCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 54)
55	LeuRS.v1	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCAT TAGCCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATAACATTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCTGCCGGCGGAAGGCGCGGCGGTGAAAAAC AACACCGCGCCGGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG

[0201]

		GCCTGGTGTATAAAAAAACCAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAG AAATTCCGCAGTGGTTTATTAAAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCGCAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACCACCCGCCCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCCGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAACGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGCGGAAGCGG AAATGGCGACCATGGAAAAAAAAGGCGTGATACCGGT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCGCGCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCC ATTAAAGCGGATCCGGAATGGGCGAAAAACACCGTGAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTGCGATTGGCGGCATTGAA CATCGGATTATGGGCTGCTGTATTTTCGCTTTTTCATAA ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG GCGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATG CGTTTTATTATGTGGGCGAAAAACGGCGAACGCAACTGGGT GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC AACGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCG CGGATACCGTGCGCCTGTTTATGATGTTTTCGAGCCCGGCG GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC GCGAACCGCTTTCTGAAACGCGTGTTGGAACCTGATG AACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGG ATGCGCTGACCGAAAAACCAGAAAGCGCTGCGCCGCGATGT GCATAAAACCATTTGCGAAAGTGACCGATGATATTGGCCGC CGCCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAC TGATGAACAACTGGCGAAAGCGCCGACCGATGGCGAAC AGGATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGT GCGCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCC TGTGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGC GCCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAGA TAGCACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGC GCGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAG GTGCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAAT ATCTGGATGGCGTGACCGTGCGCAAAGTGATTATGTGCC GGGCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 55)
56	LeuRS.v2	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCGT GAGCCCGTATCCGAGCGGCCCGCTGCATATGGGCCATGTG CGCAACTATACCATTTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA

[0202]

		TGCCTTTGGCCTGCCGGCGGAAGGCGCGGCGGTGAAAAAC AACACCGCGCCGGCGCGTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAG AAATTCCGCAGTGGTTTATTAATAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCCGAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGATACCAACCCGCCCGGATCGCTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAACGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGCGGAAGCGG AAATGGCGACCATGGAAAAAAAAGGCGTGATACCGGCT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAATATGGCCTGAACATTAAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCC ATTAAGCGGATCCGGAATGGGCGAAAAACACCGTGAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT TTATGAAAGCAGCTGGAGCTATGCGCGCTATACCTGCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTCTGATTGGCGGCATTGAAC ATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAAA CTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCGG CGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATGC GTTTTATTATGTGGGCGAAAAACGGCGAACGCAACTGGGT AGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGGCC GCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTGGT GTATACCGGCATGAGCAAAATGAGCAAAAGCAAAACAA CGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCGCG GATACCGTGCGCCTGTTTATGATGTTTGCGAGCCCGGCGG ATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGCG CGAACCGCTTTCTGAAACGCGTGTTGAAACTGGTGTATGA ACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGGAT GCGCTGACCGAAAAACCAGAAAGCGCTGCGCCGCGATGTGC ATAAAACCATTTGCGAAAGTGACCGATGATATTGGCCGCCG CCAGACCTTTAACACCGCGATTGCGGCGATTATGGAACGTG ATGAACAACTGGCGAAAGCGCCGACCGATGGCGAACAG GATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGTG GCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCCTG TGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGCG CCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAGAT AGCACCTGGTGGTGGTGAGGTGAACGGCAAAGTGCGCG CGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAGGT GCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAATAT CTGGATGGCGTGACCGTGCGCAAAGTGATTTATGTGCCGG GCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 56)
57	LeuRS.v3	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAA

[0203]

GTGCAGCTGCATTGGGATGAAAAACGCACCTTTGAAGTGA
 CCGAAGATGAAAGCAAAGAAAAATATTATTGCCTGAGCAT
 TCTGCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG
 CGCAACTATAACCATTGGCGATGTGATTGCGCGCTATCAGC
 GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA
 TGCCTTTGGCCTGCCGGCGGAAGGCGCGGCGGTGAAAAAC
 AACACCGCGCCGGCGCCGTGGACCTATGATAACATTGCGT
 ATATGAAAAACCAGCTGAAAATGCTGGGCTTTGGCTATGA
 TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT
 CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG
 GCCTGGTGTATAAAAAAACAGCGCGGTGAACCTGGTGCCC
 GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT
 GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAG
 AAATTCCGCAGTGGTTTATTAAAATTACCGCGTATGCGGAT
 GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG
 ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCGCGAG
 CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT
 AACACCCTGACCGTGTATACACCCGCCCGGATGCGTTTAT
 GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG
 GCGCAGAAAGCGGCGGAAAAACAACCCGGAAGTGGCGGCG
 TTTATTGATGAATGCCGCAACACCAAAGTGGCGGAAGCGG
 AAATGGCGACCATGGAAAAAAAAGGCGTGATACCGGCT
 TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCGGT
 GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC
 GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG
 AATTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT
 TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG
 GCGCTGACCGAAAAAGGCGTGCTGTTTAAACAGCGGCGAAT
 TTAACGGCCTGGATCATGAAGCGGCGTTTAAACGCGATTGC
 GGATAAACTGACCGCATGGGCGTGCGGCGAACGCAAAAGT
 GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC
 TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG
 GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT
 TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCG
 ATTAAAGCGGATCCGGAATGGGCGAAAACACCGTGAAAC
 GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT
 TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCCC
 GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA
 CTATTGGCTGCCGCTGGATATTGCGATTGGCGGCATTGAA
 CATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAA
 ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG
 GCGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATG
 CGTTTTATTATGTGGGCGAAAACGGCGAACGCAACTGGGT
 GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG
 CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG
 GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC
 AACGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCG
 CGGATACCGTGCGCCTGTTTATGATGTTTGCAGGCCCGCG
 GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC
 GCGAACCGCTTTCTGAAACGCGTGTTGAAACTGGTGTATG
 AACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGG
 ATGCGCTGACCGAAAACAGAAAGCGCTGCGCCGCGATGT
 GCATAAAACCATTCGAAAGTGACCGATGATATTGGCCGC
 CGCCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAC
 TGATGAACAACTGGCGAAAGCGCCGACCGATGGCGAAC
 AGGATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGT
 GCGCATGCTGAACCCGTTTACCCCGCATATTGCTTTACCC
 TGTGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGC
 GCCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAGA
 TAGCACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGC
 GCGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAG

[0204]

		GTGCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAAT ATCTGGATGGCGTGACCGTGCGCAAAGTGATTATGTGCC GGGCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 57)
58	V9	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAA GTGCAGCTGCATTGGGATATGAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCAT TAGCCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATAACATTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCCTGCCGGCGGAAGGCGCGGCGGTGAAAAAC AACACCGCGCCGGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACAGCTGAAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAG AAATTCCGCAGTGGTTTATTAATAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCCGCAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACCACCCGCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAACCTGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGGCGGAAGCGG AAATGGCGACCATGGAAAAAAAAGGCGTGATACCGGCT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACCAGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAACGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCCG ATTAAGCGGATCCGGAATGGGCGAAAACACCGTGAAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTGCGATTGGCGGCATTGAA CATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAA ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG GCGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATG CGTTTTATTATGTGGGCGAAAACGGCGAACGCAACTGGGT GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC AACGGCATTGATCCGCAAGGTGATGGTGGAACGCTATGGCG CGGATACCGTGCGCCTGTTTATGATGTTTTCGAGCCCGGCG GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC GCGAACCGCTTTCTGAAACGCGTGTTGAAACTGGTGTATG AACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGG ATGCGCTGACCGAAAACAGAAAGCGCTGCGCCGCGATGT GCATAAAACCATTTGCGAAAGTGACCGATGATATTGGCCGC CGCCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAC TGATGAACAACTGGCGAAAGCGCCGACCGATGGCGAAC AGGATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGT

[0205]

		GCGCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCC TGTGGCAGGAACGTGAAAGGCGAAGGCGATATTGATAACGC GCCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAAGA TAGCACCTTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGC GCGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAG GTGCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAAT ATCTGGATGGCGTGACCGTGCGCAAAGTGATTATGTGCC GGGCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 58)
59	V10	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAAA GTGCAGCTGCATTGGGATGTGAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCCTGAGCAT TAGCCCGTATCCGAGCGGCCCGCTGCATATGGGCCATGTG CGCAACTATAACATTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TCGTTTTGGCCTGCCGGCGGAAGGCGCGCGGTGAAAAAC AACACCGCGCCGGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAAG AAATTCCGCAGTGGTTTATTAATAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCGCAG CGAAGGCGTGGAATTAACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACCACCCGCCCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAACCTGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGGCGGAAGCGG AAATGGCGACCATGGAAAAAAAGGCGTGGATACCGGT TTAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCCGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAAATATGGCCTGAACATTAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAACAGCGGCGAAT TTAACGGCTGGATCATGAAGCGGCGTTTAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGTGGATGGCATTACCAGCCCG ATTAAGCGGATCCGGAATGGGCGAAAACACCGTGAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTGCGATTGGCGGCATTGAA CATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAA ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG GCGAAACAGCTGCTGTGCCAGGCGATGGTGGTGGCGGATG CGTTTTATTATGTGGGCGAAAACGGCGAACGCAACTGGGT GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC AACGGCATTGATCCGCAGGTGATGGTGAACGCTATGGCG CGGATACCGTGCGCCTGTTTATGATGTTTTCGAGCCCGGCG GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC GCGAACCCTTTCTGAAACGCGTGTTGAAACTGGTGTATG AACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGG

[0206]

		ATGCGCTGACCGAAAAACCAGAAAGCGCTGCGCCGCGATGT GCATAAAACCAATTGCGAAAGTGACCGATGATATTGGCCGC CGCCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAC TGATGAACAACTGGCGAAAGCGCCGACCGATGGCGAAC AGGATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGT GCGCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCC TGTGGCAGGAAGTGAAGGCGAAGGCGATATTGATAACGC GCCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGGAAGA TAGCACCTTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGC GCGAAAATTACCGTGCCGGTGGATGCGACCGAAGAAGCAG GTGCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAAT ATCTGGATGGCGTGACCGTGCGCAAAGTGATTATGTGCC GGGCAAAGTCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 59)
60	V26	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATTGCCTGAGCAT TGTGCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATACCAATTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TCGGTTTGGCCTGCCGGCGGAAGGCGCGGCGGTGAAAAAC AACACCGCGCCGGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACCAAGCTGAAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGTGCGGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAG AAATTCGCGAGTGGTTTATTAATAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCGCGAG CGAAGGCGTGGAAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACCAACCGCCCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAAGTGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGGCGGAAGCGG AAATGGCGACCATGGAAAAAAAGGCGTGATACCGGT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAAATATGGCCTGAACATTAAACGGGTGAT TCTGGCGGCGGATGGCAGCGAACCGGATCTAGCGGACAG GCGCTGACCGAAAAAGGCGTGCTGTTTAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGTGATGGATGGCATTACCAGCCCC ATTAAGCGGATCCGGAATGGGCGAAAACACCGTGAAC GGCATGCCGGCGCTGCGCGAAACCGGATACCTTTGATACCT TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTGCGATTGGCGGCATTGAA CATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAA ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG GCGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATG CGTTTTATTATGTGGGCGAAAACGGCGAACGCAACTGGGT GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC

[0207]

		AACGGCATTGATCCGCAGGTGATGGTGGAAACGCTATGGCG CGGATACCGTGCGCCTGTTTATGATGTTTGCGAGCCCGCG GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC GCGAACCGCTTTCTGAAACGCGTGTTGGAACTGGTGTATG AACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGG ATGCGCTGACCGAAAACCAGAAAGCGCTGCGCCGCGATGT GCATAAAACCATTGCGAAAGTGACCGATGATATTGGCCGC CGCCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAC TGATGAACAACTGGCGAAAGCGCCGACCGATGGCGAAC AGGATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGT GCGCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCC TGTGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGC GCCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAAG TAGCACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGC GCGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAG GTGCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAAT ATCTGGATGGCGTGACCGTGCGCAAAGTGATTTATGTGCC GGGCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 60)
61	V29	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCAT TGCGCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATACCATTTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCCTGCCGGCGGAAGGCGCGCGGTGAAAAAC AACACCGCGCCGGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAG AAATTCCGCAGTGGTTTATTAATAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCGCAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACCAACCCGCCCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAACCTGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGCGGGAAGCGG AAATGGCGACCATGGAAAAAAAAGGCGTGATACCGCT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGTGATGGATGGCATTACCAGCCC ATTAAAGCGGATCCGGAATGGGCGAAAACCACCGTGAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTGCGATTGGCGGCATTGAA CATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAA ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG

[0208]

		GCGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATG CGTTTTATTATGTGGCGGAAAACGGCGAACGCAACTGGGT GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC AACGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCG CGGATACCGTGCGCCTGTTTATGATGTTTGCAGAGCCCGCG GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC GCGAACCGCTTTCTGAAACGCGTGTTGAAACTGGTGTATG AACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGG ATGCGCTGACCGAAAACAGAAAGCGCTGCGCCGCGATGT GCATAAAACCATTGCGAAAGTGACCGATGATATTGGCCGC CGCCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAC TGATGAACAAACTGGCGAAAGCGCCGACCGATGGCGAAC AGGATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGT GCGCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCC TGTGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGC GCCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAGA TAGCACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGC GCGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAG GTGCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAAT ATCTGGATGGCGTGACCGTGCGCAAAGTGATTTATGTGCC GGGCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 61)
62	V48	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCAT TAGCCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATACCATTTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCCTGCCGGCGGAAGGCGCGCGGTGAAAAAC AACACCGCGCCGCGCGCTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAAG AAATTCGCGAGTGGTTTATTAATAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCCGAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACCACCCGCCCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAACAACCCGGAACCTGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGCGGGAAGCGG AAATGGCGACCATGGAAAAAAAAGGCGTGATACCGGCT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGTGGATGGCATTACCAGCCCCG ATTAAAGCGGATCCGGAATGGGCGAAAACACCGTGAAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT

[0209]

		TTATGGAAAGCAGCTGGGCGTATGCGCGCTATACCTGCCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTGCGATTGGCGGCATTGAA CATGCGATTATGGGCTGCTGTATTTTCGCTTTTTTCATAA ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG GCGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATG CGTTTTATTATGTGGGCGAAAACGGCGAACGCAACTGGGT GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC AACGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCG CGGATACCGTGCGCCTGTTTATGATGTTTGCAGAGCCCGCG GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC GCGAACCGCTTTAAACGCGTGTGGAAGTGGTGTATGAAC ATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGGATGC GCTGACCGAAAACAGAAAGCGCTGCGCCGCGATGTGCAT AAAACCATTTGCGAAAGTGACCGATGATATTGGCCGCCGCC AGACCTTTAACACCGCGATTGCGGCGATTATGGAAGTGT GAACAACTGGCGAAAGCGCCGACCGATGGCGAACAGGA TCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGTGCGC ATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCTGTG GCAGGAACTGAAAGGCGAAGGCGATATTGATAACGCGCC GTGGCCGGTGGCGGATGAAAAAGCGATGGTGAAGATAG CACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGCGCG AAAATTACCGTGCCGGTGGATGCGACCGAAGAACAGGTGC GCGAACGCGCGGGCCAGGAACATCTGGTGCGGAAATATCT GGATGGCGTGACCGTGCGCAAAGTGATTTATGTGCCGGGC AAACTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 62)
63	V50	ATGGAAGAAGCAGTATCGCCCGGAAGAAATTGAAAGCAAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCAT TAGCCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATACCATTTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCCTGCCGGCGGAAGGCGCGGCGGTGAAAAAC AACACCGCGCCGGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGATATAAAAAAACAGCGCGGTGAACATGGTGCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGAACGCAAAAG AAATTCCGCAGTGGTTTATTAATAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAACCATGCAGCGCAACTGGATTGGCCGCAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACACCCGCCCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAACCTGGCGCG TTATTGATGAATGCCGCAACACCAAAGTGCGGGAAGCGG AAATGGCGACCATGGAAAAAAAAGGCGTGATACCGCT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTTGCGAGCAATATGGCCTGAACATTAAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGCGCAACGCAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC

[0210]

		TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCC ATTAAAGCGGATCCGGAATGGGCGAAAACACCGTGAAC GGCATGCCGGCGCTGCGCGAAAACCGATACCTTTGATACCT TTATGGAAAGCAGCTGGCATTATGCGCGCTATACCTGCCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTGCGATTGGCGGCATTGAA CATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAA ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG GCGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATG CGTTTTATTATGTGGGCGAAAACGGCGAACGCAACTGGGT GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC AACGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCG CGGATACCGTGCGCCTGTTTATGATGTTTGCAGCCCGGCG GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC GCGAACCGCTTTCTGAAACGCGTGTTGAAACTGGTGTATG AACAATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGG ATGCGCTGACCGAAAACAGAAAGCGCTGCGCCGCGATGT GCATAAAACCATTGCGAAAGTGACCGATGATATTGGCCGC CGCCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAC TGATGAACAAACTGGCGAAAGCGCCGACCGATGGCGAAC AGGATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGT GCGCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCC TGTGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGC GCCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAGA TAGCACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGC GCGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAG GTGCGCGAACGCGCGGGCCAGGAACATCTGGTGCGGAAAT ATCTGGATGGCGTGACCGTGCGCAAAGTGATTTATGTGCC GGGCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 63)
64	V54	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCAT TAGCCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATACCATTTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCCTGCCGGCGGAAGGCGCGCGGTGAAAAAC AACACCGCGCGCGCGCGCTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAG AAATTCCGCAGTGGTTTATTAAAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCCGCAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCTGACCGTGTATACCAACCGCCCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAACCTGGCGGCG TTATTGATGAATGCCGCAACACCAAAGTGCGGGAAGCGG AAATGGCGACCATGGAAAAAAAAGGCGTGATACCGGCT TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCCGGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG AATTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT

[0211]

		TCTGGCGGCGGATGGCAGCGAACC GGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCCG ATTAAGCGGATCCGGAATGGGCGAAAAACCACCGTGAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCCCG GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTATTATTGGCGGCATTGAAC ATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAAA CTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCGG CGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATGC GTTTTATTATGTGGGCGAAAACGGCGAACGCAACTGGGTG AGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGGCC GCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTGGT GTATACCGGCATGAGCAAAATGAGCAAAAGCAAAACAA CGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCGCG GATACCGTGCGCCTGTTTATGATGTTTGCGAGCCCGCGCG ATATGACCCCTGGAATGGCAGGAAAGCGGCGTGGAAGGCG CGAACCGCTTTCTGAAACGCGTGTTGGAACCTGGTGTATGA ACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGGAT GCGCTGACCGAAAACCAGAAAGCGCTGCGCCGCGATGTGC ATAAAACCATTTGCGAAAGTGACCGATGATATTGGCCGCCG CCAGACCTTTAACACCGCGATTGCGGCGATTATGGAACCTG ATGAACAACTGGCGAAAGCGCCGACCGATGGCGAACAG GATCGCGCGTGATGCAGGAAGCGCTGCTGGCGGTGGTGCG GCATGCTGAACCCGTTTACCCCGCATATTGCTTTACCCTG TGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGCG CCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGAAGAT AGCACCCCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGCG CGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAGGT GCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAATAT CTGGATGGCGTGACCGTGCGCAAAGTGATTTATGTGCCGG GCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 64)
65	V55	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATTCCTGAGCAT TAGCCCGTATCCGAGCGGCGCCTGCATATGGGCCATGTG CGCAACTATACCATTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCCTGCCGGCGGAAGGCGGCGGTGAAAAAC AACACCGCGCCGGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACAGCGCGGTGAACCTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAG AAATTCCGCAGTGGTTTATTAAAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCGCAG CGAAGGCGTGGAATTACCTTTAACGTGAACGATTATGAT AACACCCTGACCGTGTATACCACCCGCCCGGATGCGTTTAT GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG GCGCAGAAAGCGGCGGAAAAACAACCCGGAACCTGGCGGCG TTTATTGATGAATGCCGCAACACCAAAGTGCGCGAAGCGG

[0212]

		AAATGGCGACCATGGAAAAAAGGCGTGGATACCGGT TAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCCGT GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC GCGGTGATGGCGGTGCCGGCCATGATCAGCGCGATTATG AATTTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT TCTGGCGGCGGATGGCAGCGAACC GGATCTGAGCCAGCAG GCGCTGACCGAAAAAGGCGTGCTGTTTAACAGCGGCGAAT TTAACGGCCTGGATCATGAAGCGGCGTTTAACGCGATTGC GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCCG ATTAAAGCGGATCCGGAATGGGCGAAAAACCGTGAAC GGCATGCCGGCGCTGCGCGAAACCGATACCTTTGATACCT TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCCC GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA CTATTGGCTGCCGGTGGATATTGTGATTGGCGGCATTGAAC ATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAAA CTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCGG CGAAACAGCTGCTGTGCCAGGGCATGGTGCTGCCGGATGC GTTTTATTATGTGGGCGAAAAACGGCGAACGCAACTGGGTG AGCCCCGTGGATGCGATTGTGGAACGCGATGAAAAAGGCC GCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTGGT GTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAACAA CGGCATTGATCCGCAGGTGATGGTGGAACGCTATGGCGCG GATACCGTGCGCCTGTTTATGATGTTTGCGAGCCCGGCGG ATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGCG CGAACCGCTTTCTGAAACGCGTGTTGGAACCTGGTGTATGA ACATACCGCGAAAGGCGATGTGGCGGCGTGACCGTGATG GCGCTGACCGAAAAACAGAAAGCGCTGCGCCGCGATGTGC ATAAAACCATTTGCGAAAGTGACCGATGATATTGGCCGCCG CCAGACCTTTAACACCGCGATTGCGGCGATTATGGAACCTG ATGAACAAACTGGCGAAAGCGCCGACCGATGGCGAACAG GATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGTGC GCATGCTGAACCCGTTTACCCCGCATATTGCTTTACCCTG TGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGCG CCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGGAAAGAT AGCACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGCG CGAAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAGGT GCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAATAT CTGGATGGCGTGACCGTGCGCAAAGTGATTTATGTGCCGG GCAAACCTGCTGAACCTGGTGGTGGGCGGCCCGGTG (SEQ ID NO: 65)
66	V56	ATGGAAGAACAGTATCGCCCGGAAGAAATTGAAAGCAAA GTGCAGCTGCATTGGGATAAAAAACGCACCTTTGAAGTGA CCGAAGATGAAAGCAAAGAAAAATATTATGCTGAGCAT TAGCCCGTATCCGAGCGGCCGCTGCATATGGGCCATGTG CGCAACTATACCATTTGGCGATGTGATTGCGCGCTATCAGC GCATGCTGGGCAAAAACGTGCTGCAGCCGATTGGCTGGGA TGCGTTTGGCCTGCCGGCGGAAGGCGCGCGGTGAAAAAC AACACCGCGCCGCGCCGTGGACCTATGATAACATTGCGT ATATGAAAAACCAGCTGAAAAATGCTGGGCTTTGGCTATGA TTGGAGCCGCGAACTGGCGACCTGCACCCCGGAATATTAT CGCTGGGAACAGAAATTTTTTACCGAACTGTATAAAAAAG GCCTGGTGTATAAAAAAACCAGCGCGGTGAACTGGTGCCC GAACGATCAGACCGTGCTGGCGAACGAACAGGTGATTGAT GGCTGCTGCTGGCGCTGCGATACCAAAGTGGAACGCAAAG AAATTCCGCAGTGGTTTATTAATAATTACCGCGTATGCGGAT GAACTGCTGAACGATCTGGATAAACTGGATCATTGGCCGG ATACCGTGAAAACCATGCAGCGCAACTGGATTGGCCCGAG

[0213]

CGAAGGCGTGGA AATTACCTTTAACGTGAACGATTATGAT
AACACCCGTGACCGTGTATACACCCGCCCGGATGCGTTTAT
GGGCTGCACCTATCTGGCGGTGGCGGCGGGCCATCCGCTG
GCGCAGAAAGCGGCGGAAAACAACCCGGAAGTGGCGGCG
TTTATTGATGAATGCCGCAACACCAAAGTGGCGGAAGCGG
AAATGGCGACCATGGAAAAAAAAGGCGTGGATACCGGCT
TTAAAGCGGTGCATCCGCTGACCGGCGAAGAAATTCCGGT
GTGGGCGGCGAACTTTGTGCTGATGGAATATGGCACCGGC
GCGGTGATGGCGGTGCCGGGCCATGATCAGCGCGATTATG
AATTTGCGAGCAAATATGGCCTGAACATTAAACCGGTGAT
TCTGGCGGCGGATGGCAGCGAACCGGATCTGAGCCAGCAG
GCGCTGACCGAAAAAGGCGTGCTGTTTAACAGCGGCGAAT
TTAACGGCCTGGATCATGAAGCGGCGTTTAACGCGATTGC
GGATAAACTGACCGCGATGGGCGTGGGCGAACGCAAAGT
GAACTATCGCCTGCGCGATTGGGGCGTGAGCCGCCAGCGC
TATTGGGGCGCGCCGATTCCGATGGTGACCCTGGAAGATG
GCACCGTGATGCCGACCCCGGATGATCAGCTGCCGGTGAT
TCTGCCGGAAGATGTGGTGATGGATGGCATTACCAGCCCCG
ATTAAGCGGATCCGGAATGGGCGAAAACACCGTGAAC
GGCATGCCGGCGCTGCGCGAAAACCGATACCTTTGATACCT
TTATGGAAAGCAGCTGGATTTATGCGCGCTATACCTGCC
GCAGTATAAAGAAGGCATGCTGGATAGCGAAGCGGCGAA
CTATTGGCTGCCGGTGGATATTGGCATTGGCGGCATTGAA
CATGCGATTATGGGCCTGCTGTATTTTCGCTTTTTTCATAA
ACTGATGCGCGATGCGGGCATGGTGAACAGCGATGAACCG
GCGAAACAGCTGCTGTGCCAGGGCATGGTGCTGGCGGATG
CGTTTTATTATGTGGGCGAAAACGGCGAACGCAACTGGGT
GAGCCCGGTGGATGCGATTGTGGAACGCGATGAAAAAGG
CCGCATTGTGAAAGCGAAAGATGCGGCGGGCCATGAACTG
GTGTATACCGGCATGAGCAAAATGAGCAAAAGCAAAAAC
AACGGCATTGATCCGCAGGTGATGGTGGAACGCTATTGGCG
CGGATACCGTGCGCCTGTTTATGATGTTTTCGAGCCCGGCG
GATATGACCCTGGAATGGCAGGAAAGCGGCGTGGAAGGC
GCGAACCGCTTTCTGAAACGCGTGTGGAACTGGTGTATG
AACATACCGCGAAAGGCGATGTGGCGGCGCTGAACGTGG
ATGCGCTGACCGAAAACAGAAAGCGCTGCGCCGCGATGT
GCATAAAACCATTTGCGAAAGTGACCGATGATATTGGCCGC
CGCCAGACCTTTAACACCGCGATTGCGGCGATTATGGAAC
TGATGAACAACTGGCGAAAGCGCCGACCGATGGCGAAC
AGGATCGCGCGCTGATGCAGGAAGCGCTGCTGGCGGTGGT
GCGCATGCTGAACCCGTTTACCCCGCATATTTGCTTTACCC
TGTGGCAGGAACTGAAAGGCGAAGGCGATATTGATAACGC
GCCGTGGCCGGTGGCGGATGAAAAAGCGATGGTGAAGA
TAGCACCTGGTGGTGGTGCAGGTGAACGGCAAAGTGCGC
GCGAAAATTACCGTGCCGGTGGATGCGACCGAAGAACAG
GTGCGCGAACGCGCGGGCCAGGAACATCTGGTGGCGAAAT
ATCTGGATGGCGTGACCGTGCGCAAAGTGATTATGTGCC
GGGCAAACTGCTGAACCTGGTGGTGGGCGGCCCGGTG
(SEQ ID NO: 66)

实施例

[0214] 下面的实施例仅仅是说明性的,并且不打算以任何方式限制本发明的范围或内容。

[0215] 实施例1-改进的亮氨酰tRNA合成酶突变蛋白的构建和选择

[0216] 本实施例描述了亮氨酰tRNA合成酶突变蛋白的构建。

[0217] 将野生型大肠埃希氏杆菌亮氨酰tRNA合成酶 (SEQ ID NO:1) 克隆到质粒中CMV启动子的控制之下。所述质粒还含有编码抑制型tRNA (带有CUA反密码子的大肠埃希氏杆菌亮氨酰tRNA h1, SEQ ID NO:19) 的4x U6-LeutRNA_{CUA} DNA表达盒。所述编码亮氨酰trRNA合成

酶和亮氨酸抑制型tRNA的质粒用作文库模板构建物,并被称为pBBK-LeuRS.wt-LtR-TAG。

[0218] 亮氨酸合成酶突变蛋白V2 (SEQ ID NO:3) 和V3 (SEQ ID NO:4) 通过野生型活性位点残基的标准诱变来产生(参见Zheng等,(2018)同上)。亮氨酸合成酶突变蛋白V1 (SEQ ID NO:2) 通过合并V2和V3突变蛋白的不同活性位点突变来设计。

[0219] 然后将编码亮氨酸tRNA合成酶突变蛋白V1 (SEQ ID NO:2;在本文中被称为LeuRS.v1)的质粒作为模板,用于产生编码另外的亮氨酸tRNA合成酶变体的质粒的文库。所述文库包括编码将Q2、E20、M40、L41、T252、Y499、Y527和H537中的每一者用20种天然氨基酸中的每一者单独替换的亮氨酸tRNA合成酶变体的质粒。

[0220] 利用编码在Y39处具有琥珀密码子并在C-端处融合到His标签的GFP蛋白 (GFP39-TAG) 的报告质粒的sGFP-39TAG报告基因荧光测定法,被用于评估哺乳动物细胞中的亮氨酸合成酶突变蛋白活性。将HEK293T细胞在增补有10%FBS和0.5x青霉素-链霉素的Dulbecco改良的Eagle培养基 (DMEM) 中,在加湿培养箱中,在37℃和8%CO₂存在下培养。在转染前24小时,在12孔板中每个孔接种0.7x 10⁶个HEK293T细胞。将聚乙烯亚胺和DNA以4μL PEI (1mg/mL) 比1μg DNA的比例在DMEM中混合。对于每个转染混合物来说,将500ng GFP39-TAG报告质粒与pBBK-LeuRS.v#-LtR-TAG混合。非天然氨基酸 (UAA) 以0.5mM LCA、1mM LKET或1mM ACA的浓度添加到培养基或排除在培养基之外。在48小时时使用Olympus显微镜,通过488带通滤光片组获得荧光图像。为了获得GFP39-TAG表达数据,以12,000x g离心收获细胞。将细胞用PBS清洗,并用增补有1x Halt蛋白酶抑制剂和0.01%Pierce通用核酸酶的CellLytic M裂解缓冲液裂解。在20分钟温育后,以15,000x g离心10分钟将裂解物澄清,并转移到透明底96孔测定板中。使用Fluoroskan Ascent II (Ex.485nm;Em.535nm) 测量荧光。

[0221] 如上所述测定LeuRS.v1的变体 (SEQ ID NOs:5-14) 的LCA掺入。LeuRS.v1的变体中的点突变示出在图4A中,并且GFP-39TAG报告基因测定中的相应荧光活性示出在图4B中。正如在图4B中看到的,从高通量筛选凭经验选择的亮氨酸tRNA合成酶与亲本LeuRS.v1相比在0.5mM LCA存在下具有不同的活性。测试了超过50种变体;在GFP-39TAG报告基因测定中选择的LeuRS变体 (LeuRS变体v1、v2、v9、v10、v26、v29、v48、v50、v54、v55和v63,其氨基酸序列和突变如下面的表1中所示) 的代表性荧光图像示出在图4C中。发现当与亲本序列LeuRS.v1相比时,变体v9、v10、v26、v29、v48、v50、v54、v55和v63具有增强的LCA掺入活性。

[0222] 对图4B-4C中描绘的LeuRS变体进行多特异性分析,以测试它们是否接受除了LCA之外的UAA底物。如上所述在0.5mM LCA、1mM LKET或1mM ACA存在下进行GFP-39TAG表达分析(图5A)。如上所述测量活性,结果描绘在图5B中。

[0223] 本实施例中描述的与野生型相比具有增强的活性的亮氨酸合成酶突变蛋白概述在表1中。

[0224] 表1

[0225]

LeuRS	Q2	E20	M40	L41	T252	Y499	Y527	H537
WT (SEQ ID NO: 1)	Q	E	M	L	T	Y	Y	H
V1 (SEQ ID NO: 2)	E	K	I	S	A	I	A	G
V2 (SEQ ID NO: 3)	E	K	V	S	R	S	L	G
V3 (SEQ ID NO: 4)	E	E	I	L	A	I	A	G
V9 (SEQ ID NO: 5)	E	M	I	S	A	I	A	G
V10 (SEQ ID NO: 6)	E	V	I	S	A	I	A	G

[0226]

V26 (SEQ ID NO: 7)	E	K	I	V	A	I	A	G
V29 (SEQ ID NO: 8)	E	K	I	A	A	I	A	G
V48 (SEQ ID NO: 9)	E	K	I	S	A	A	A	G
V50 (SEQ ID NO: 10)	E	K	I	S	A	H	A	G
V54 (SEQ ID NO: 11)	E	K	I	S	A	I	I	G
V55 (SEQ ID NO: 12)	E	K	I	S	A	I	V	G
V63 (SEQ ID NO: 13)	E	K	I	S	A	I	G	G

[0227] 实施例2-表达亮氨酰tRNA和亮氨酰tRNA合成酶突变蛋白的稳定细胞系的构建

[0228] 本实施例描述了表达亮氨酰抑制型tRNA和亮氨酰tRNA合成酶的细胞系、例如稳定细胞系的构建(示意描绘在图6A-6B中)。

[0229] CHO-dhFr贴壁细胞从ATCC获得。选择CHO-dhFr细胞作为亲代细胞系是由于它们固有的代谢dhFr选择策略对于目的靶基因的未来研究(即在平台细胞系产生后)的灵活性。将CHO-dhFr细胞按照ATCC的方案维持在增补有10%胎牛血清、0.1mM次黄嘌呤、0.016mM胸苷和0.002mM甲氨蝶呤的Gibco™ IMDM中。对于稳定细胞系的产生来说,培养处于第15次传代下的CHO-dhFr细胞,用于使用脂质体2000(Thermo Fisher Scientific)或Nucleofector 4D-X装置和相关的试剂盒(Lonza)的pCLD质粒转染。

[0230] 对于脂质体2000转染来说,在转染前24小时,将2mL的每mL 2.5×10^5 个CHO-dhFr细胞在6孔板的每个孔中铺板。24小时后,使用脂质体2000,按照Thermo Fisher Scientific的标准方案将细胞用3μg质粒转染,所述质粒含有(i) 4x U6启动子,亮氨酰抑制型tRNA h1重复序列盒,和(ii) EF1α启动子,LeuRS.v1-IRES-嘌呤霉素表达盒(质粒2,表2)。

[0231] 对于Lonza Nucleofector 4D-X转染来说,使用Lonza试剂盒SE,按照制造商的说明书将 1×10^6 个CHO-dhFr细胞用2-4μg质粒转染,所述质粒含有(i) 4x U6启动子,亮氨酰抑制型tRNA重复序列盒,和(ii) EF1α启动子,LeuRS.v1-IRES-嘌呤霉素表达盒(质粒2,表2)。将转染后的细胞在6孔板中铺板。在基于脂质体2000的转染或核转染后,将培养板在37℃和5%CO₂下温育恢复24-48小时,然后施加1.5μg-6μg嘌呤霉素进行选择(图6A的步骤1-2)。

[0232] 质粒2中的亮氨酰抑制型tRNA描绘在SEQ ID NO:19中,质粒2中的LeuRS.v1亮氨酰tRNA合成酶描绘在SEQ ID NO:2中。

[0233] 在嘌呤霉素选择后2周,使用脂质体2000,如制造商的方案中所述,将所述细胞用包含GFP和mCherry荧光报告基因两者的双报告基因构建物转染。所述双报告基因构建物组成性地产生mCherry(红色)并通过包含终止密码子的连接物连接到GFP,使得如果tRNA/aaRS对有活性,则所述报告物条件性产生GFP(绿色)(图6A的步骤3)。将转染的细胞在0.25mM LCA存在下培养15-48小时,然后在BD Melody FACS分选仪(BD)上进行无菌单细胞分选。UAA掺入活性通过荧光激活细胞分选(FACS)和荧光显微镜(图7-8)来确定。图7A-7C展示了荧光报告物阳性对照的预期表型,证明mCherry-GFP野生型导致细胞沿着45度轴群集。图7D-7E表示如上所述使用基于脂质体的转染获得并用嘌呤霉素选择的稳定细胞群体,而图7F-7H表示如上所述使用基于核转染的转染获得并用嘌呤霉素选择的稳定细胞群体。如图7A-7H中所示,指示了条件GFP信号变化的群体向45度轴的迁移,证实了核转染具有最大的迁移以及因此条件双报告基因的最大激活。总之,沿着45度轴的每个点均为一个稳定细胞系,可以被选择用于分离、繁殖和重新表征(图7)。在FACS合并物分析期间,选择mCherry和GFP双阳性克隆,并通过作为图7中描绘的最宽的选择的“全”门或沿着45度轴的45门将单克隆分选到96孔板中(图7I)。将这些克隆培养回收,然后通过条件双报告基因进行重新表征(图6A的步骤4)。

[0234] 然后将分离的克隆单独地制备用于转染,并通过与上述相同的mCherry-GFP条件报告基因进行重新表征,其用作UAA掺入的代表性读数(图6A的步骤5)。使用LCA如上所述(除非另有描述)将克隆群体用荧光报告基因瞬时转染,并通过荧光显微镜(图8)、FACS(图9)筛选。来自于FACS结果的抑制效率和蛋白质产量的定量示出在图10中。在图8中,首先通过荧光显微术将克隆分离物与表达mCherry-GFPwt(简称为MGwt)的亲本细胞或用mCherry-GFP*(简称为MG*,*是指TAG突变体)和被称为“瞬时对照”或“pCLD瞬时”的最初用于产生所述稳定细胞系的pCLD抑制型质粒(质粒2,表2)共转染的亲本细胞进行比较。图8A和图8B描绘了对从基于脂质体或核转染的细胞系产生分离到的克隆进行的上述标准的基于脂质体的表征测定(在48小时分析)。克隆1.L1w.6、2.2N4S.3和2.2N6S.3的代表性图像证实了这些克隆与亲本细胞系相比具有总体更高的蛋白质表达和更高的显示出MG*报告基因通读的细胞的百分率。

[0235] 通过FACS定量所述转染的细胞的荧光强度。在克隆群体中表达的MG*报告基因的直方图分析(图9)有助于了解所述群体中可以掺入UAA的细胞的总体百分率,并允许比较蛋白质表达生产率和抑制效率。稳定克隆(图9A-9J)相对于含有报告基因和抑制型质粒的CHO-dhFr的瞬时对照(图9K;表2的质粒2)中MG*表达的直方图分析显示,如图9I中所示门选的稳定克隆在0.25mM LCA存在下具有总体更高的细胞转染效率以及因此更高的报告基因表达。带有通过核转染递送的Leu-tR-RS基因的克隆显示出更高的细胞转染效率,正如在图9中所证实的。

[0236] 图10中描绘的直方图使用BD Melody软件来确定。图10A中示出的平均mCherry或GFP荧光的定量描绘了与在图9中看到的相同的趋势。为了增加对相对抑制效率的了解,比较了细胞系之间平均GFP荧光与平均mCherry荧光的比率。图10B中描绘的稳定细胞系表现出合理的抑制效率,并证实4x Leu-tRNA/1x LeuRS.v1的比率不需另外的抑制型质粒DNA即可产生可以掺入LCA的稳定细胞系的非凡能力。此外,图10C描绘了与瞬时转染相比表达GFP的细胞的百分率与表达mCherry的细胞的百分率之间的更高比率,证实了稳定群体中更高

的UAA掺入频率。对tRNA:aaRS比率和整合位点的进一步实验和分析可能改进这些特征。

[0237] 使用GFP*报告基因进行了稳定克隆的生产率分析,与亲本细胞系或克隆1.L1w.6进行了比较(图11)。使用核转染,在0.25mM LCA存在下将2ml细胞培养物用3μg编码在Y39处具有琥珀密码子并在C-端处融合有His标签的GFP蛋白(GFP39-TAG)的报告质粒转染,并在转染后允许其在如上所述的培养箱条件下表达72小时。收获细胞,将其在150μL增补有1x Halt蛋白酶抑制剂和0.01%Pierce通用核酸酶的CellLytic M裂解缓冲液中裂解,并按照制造商的方案亲和捕获在50μL Ni-NTA珠子上。将Ni-NTA珠子用PBS加20mM咪唑清洗4次,然后用50μL PBS加300mM咪唑洗脱。将每个样品在4X-SDS样品缓冲液中变性,并通过考马斯凝胶进行分析。将来自于每个样品的14μL的相同上样体积在4-12%Bis-Tris凝胶上,在1X MES运行缓冲液中分离(如图11中所示)。第1道含有在亲本细胞系CHO-DhFr中转染的GFPwt,第2道含有在亲本细胞系CHO-DhFr中转染的GFP-TAG和pCLD抑制型质粒(质粒2,表2),第3道含有在克隆1.L1.6中转染的GFPwt,第4道含有在克隆1.L1.6中转染的GFP-TAG并且没有其他质粒(图11)。观察到GFP的正确尺寸并用箭头指示。显示出在亲本和克隆细胞系中表达的蛋白质水平相当,并且与野生型相比含有UAA的蛋白质的相对比例明显更高(例如第3相对于第4道与第2相对于第1道相比)。

[0238] 通过上述方法在30天的时间段中进一步测定了克隆细胞系的UAA掺入能力的稳定性。将克隆融化并如上所述按照方案进行培养。在培养期结束时,将细胞融化并为每个时间点如上所述重复荧光分析,然后如上所述进行总体生产率分析(图6A的步骤6)。对于维持UAA掺入能力的细胞系来说,产生其中将目的基因稳定整合在平台细胞系的基因组中的生产细胞系(图6A的步骤7)。

[0239] 用于构建本实施例中描述的表达亮氨酸抑制型tRNA和亮氨酸tRNA合成酶的稳定细胞系的构建物的概述描绘在表2中。

[0240] 表2

[0241]	# 质粒	# tRNA	# aaRS	抗性基因
[0242]	1 pCLD-4xLeutRwt-LeuRS.v1-Puro	4	1	嘌呤霉素
	2 pCLD-4xLeutR.h1-LeuRS.v1-Puro	4	1	嘌呤霉素

[0243] 实施例3-使用WT亮氨酸tRNA琥珀抑制基因与H1亮氨酸tRNA琥珀抑制基因产生的稳定细胞系合并物的比较

[0244] 利用如上实施例2中描述的核转染,使用pCLD-4xLeutRwt-LeuRS.v1-Puro(表2的质粒1)或带有野生型或工程化改造以含有CUA反密码子的突变蛋白tRNA h1的pCLD-4xLeutR.h1-LeuRS.v1-Puro(表2的质粒2)产生平行的细胞系合并物,以便比较野生型和h1亮氨酸tRNA分别对稳定克隆产生效率的影响(过程如图6A和图6B的步骤1-3中所示)。质粒1中的亮氨酸抑制型tRNA LeutRwt描绘在SEQ ID NO:16中,质粒2中的亮氨酸抑制型tRNA LeutR.h1描绘在SEQ ID NO:19中,并且质粒1和2中的LeuRS.v1亮氨酸tRNA合成酶描绘在SEQ ID NO:2中。两种合并物经历相同的选择条件,并通过FACS分析使用如上文实施例2中所述的MG*报告基因进行分析。使用表达h1 tRNA的pCLD质粒的瞬时转染被用作对照以鉴定靶门P6。当细胞用h1 tRNA而不是“野生型”抑制基因转染时,在45度门中鉴定到多大约2倍

的克隆 (P6) 并鉴定到多2倍的阳性克隆 (P5) (图12)。

[0245] 实施例4-表达色氨酰tRNA和色氨酰tRNA合成酶突变蛋白的稳定细胞系的构建

[0246] 通过如上在实施例2中所述的核转染,使用pCLD-4xTrptR-TGA-TrpRS.h14-Puro产生细胞系合并物。这个版本的pCLD质粒含有 (i) 4x U6启动子, Trp-tRNA-UCA重复序列盒, 和 (ii) EF1 α 启动的TrpRS.h14-IRES-嘌呤霉素表达盒。质粒1中的色氨酰抑制型tRNA Trp-tRNA-UCA描绘在SEQ ID NO:50中, 并且质粒1中的TrpRS.h14色氨酰tRNA合成酶描绘在SEQ ID NO:44中。

[0247] 使用了在1.5 μ g/mL或4 μ g/mL嘌呤霉素中7-10天的缩短的选择, 并使用了具有TGA终止密码子而不是TAG终止密码子的mCherry-GFP*报告基因。在哺乳动物细胞中, 对于色氨酰对来说, TGA终止密码子与TAG终止密码子相比表现出更高的效率。稳定的色氨酸细胞系的合并物经历相同的选择条件, 并通过FACS分析使用如上文实施例2中所述修饰的MG* (TGA) 报告基因, 在1mM 5-羟基色氨酸 (HTP) 这种用于色氨酰对的UAA存在下进行分析 (图13)。在靶45度门内鉴定到稳定的克隆, 并将其分选到克隆分离物中。这些稳定克隆的鉴定和分选证实了4:1的tRNA:aaRS比例对于稳定的色氨酰细胞系的产生来说是可行的。

[0248] 进行了克隆分离物的进一步表征, 例如使用在上文实施例2中描述的亮氨酰克隆分离物所进行的, 以确定所述稳定的色氨酰细胞系的蛋白质产量和稳定性特征。

[0249] 通过参考并入

[0250] 本文中提及的每个专利和科学文献的全部公开内容为所有目的通过参考并入本文。

[0251] 等同性

[0252] 在不背离本发明的精神或本质特征的情况下, 本发明可以以其他特定形式体现。因此, 上述实施方式在所有方面都应被认为是说明性的, 而不是对本文中描述的本发明的限制。因此, 本发明的范围由随附的权利要求书而不是由上述描述指定, 并且落于所述权利要求书的等同性含义和范围之内所有变化都旨在包含在其中。

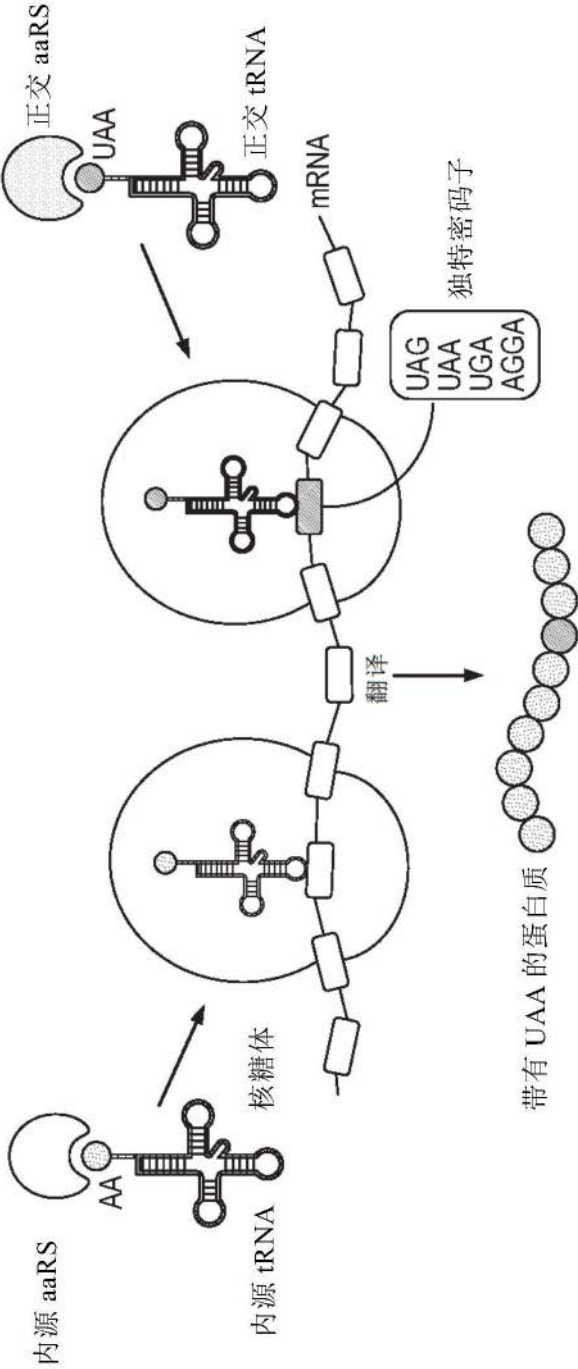
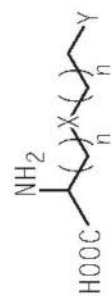
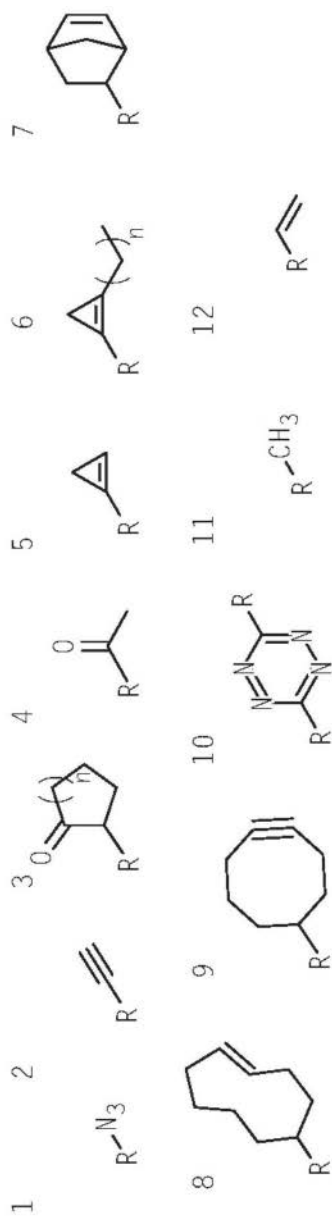


图1

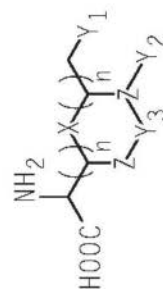
式 A



X = O, S, CH₂, NH
n = 0-20 的任何整数
Y = 官能团 1-12 (上面概述的)



式 B



X = O, S, CH₂, NH
n = 0-20 的任何整数
Z = 任何长度的碳链(CH₂)n, 其中 n = 0-20 的任何整数
Y₁ = 官能团 1-12 (上面概述的)
Y₂ = 相同或不同的官能团 1-12 (上面概述的)
Y₃ = 相同或不同的官能团 1-12 (上面概述的)

图2A

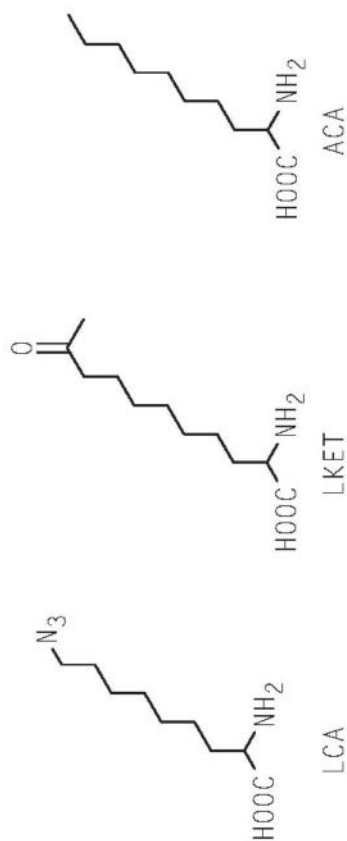


图2B

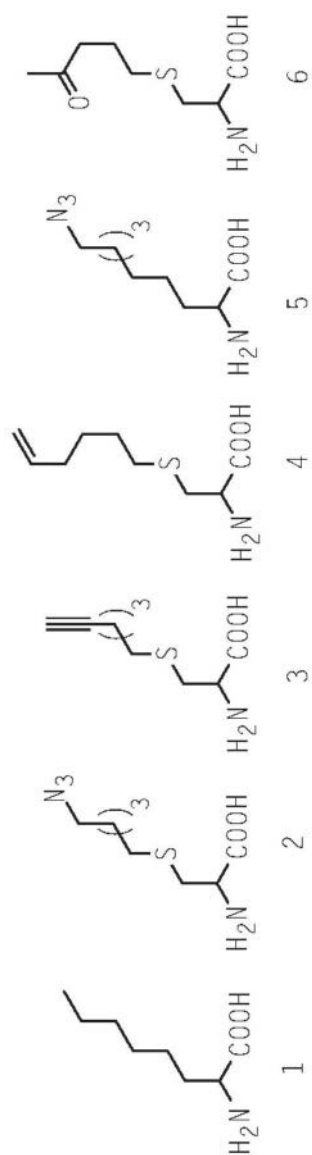


图2C

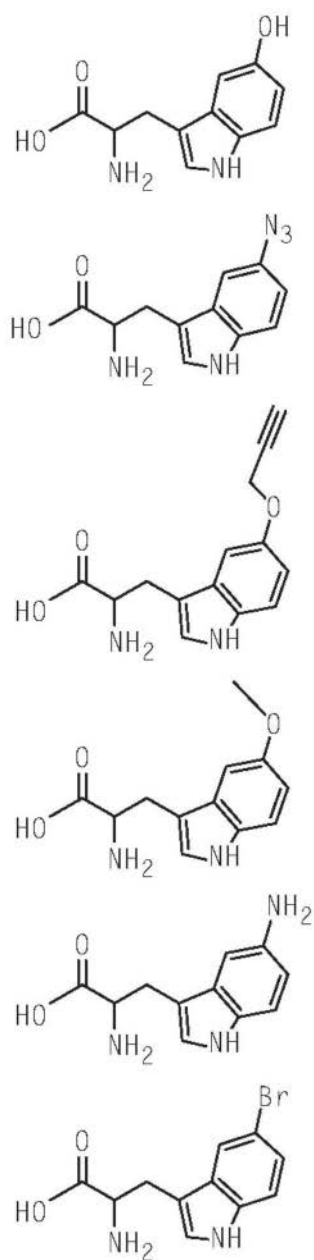


图3

LeuRS.v#	突变
1	LeuRS.v1
9	E20M
10	E20V
11	E20S
12	E20D
13	E20E
15	M40M
16	M40V
18	M40P
20	M40A
23	L41L
24	L41I
26	L41V
29	L41A
31	L41D
32	L41C
33	L41G
36	T252T
37	T252N
39	T252D
40	T252E
44	Y499V
45	Y499S
46	Y499P
48	Y499A
49	Y499Y
50	Y499H
54	Y527I
55	Y527V
63	Y527G

图4A

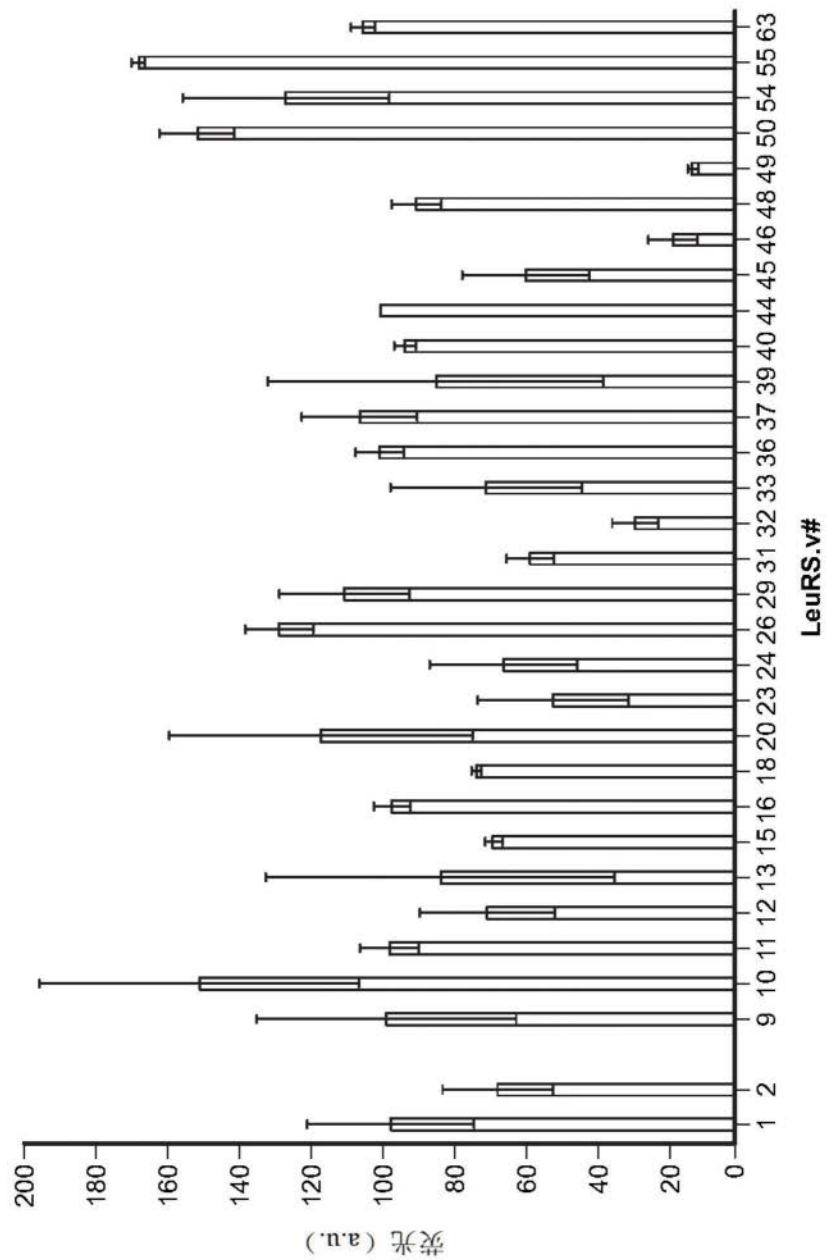


图4B

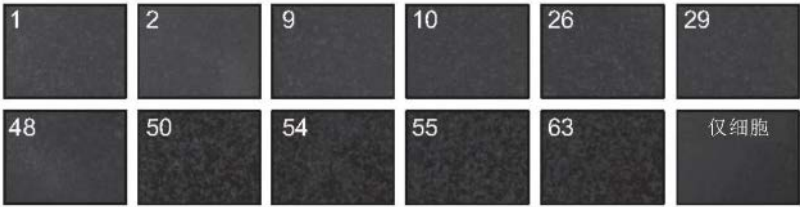


图4C

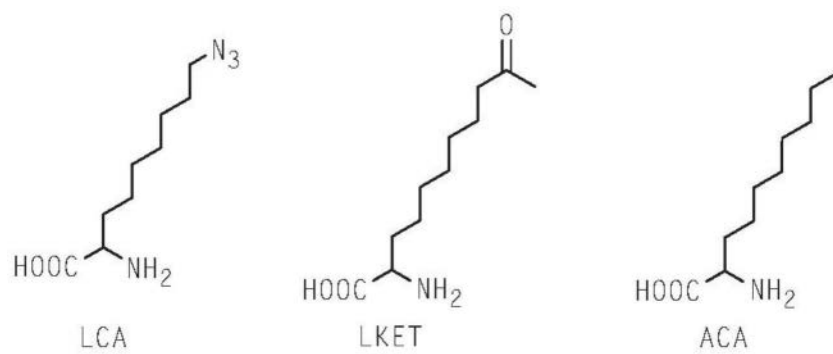


图5A

LeuRS.v#	突变
1	LeuRS.v1
9	E20M
10	E20V
26	L41V
48	Y499A
50	Y499H
54	Y527I

图5B

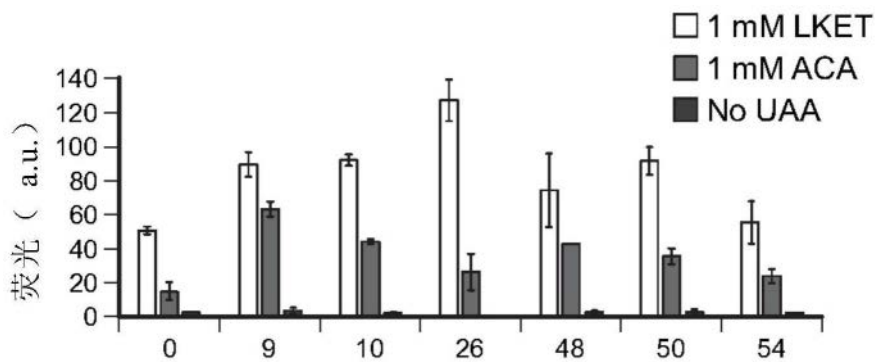


图5C

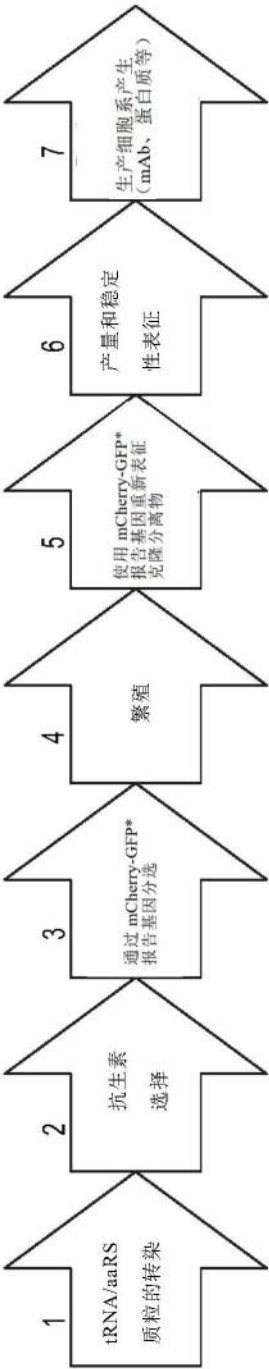


图6A

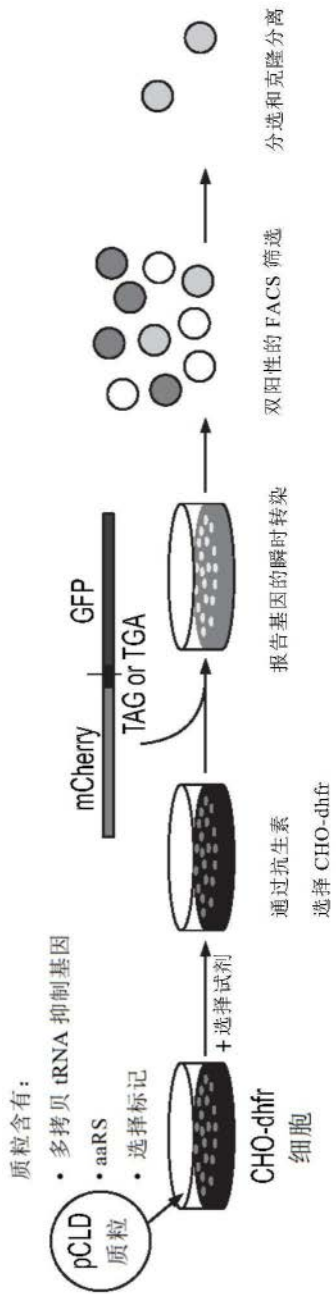
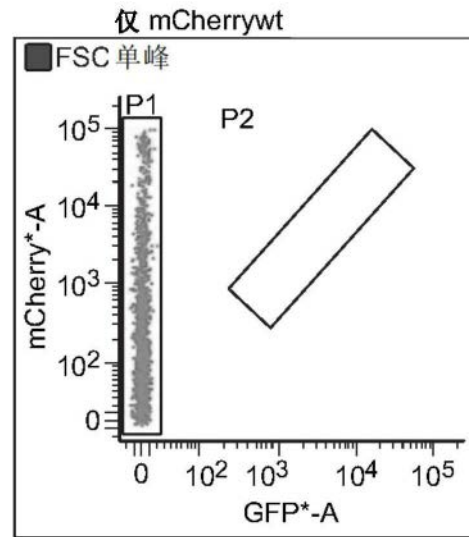
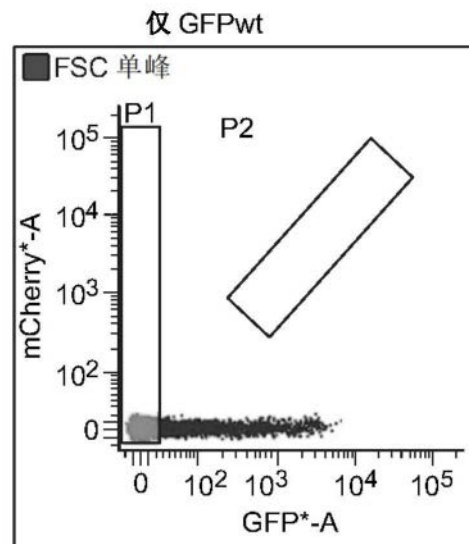


图6B



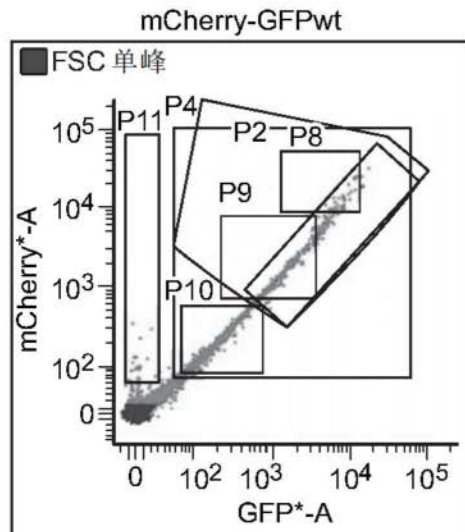
P1: mCherry
P2: 对分选阳性

图7A



P1: mCherry
P2: 对分选阳性

图7B



P11: mCherry 可行性对照

P4: GFP-mCherry 阳性

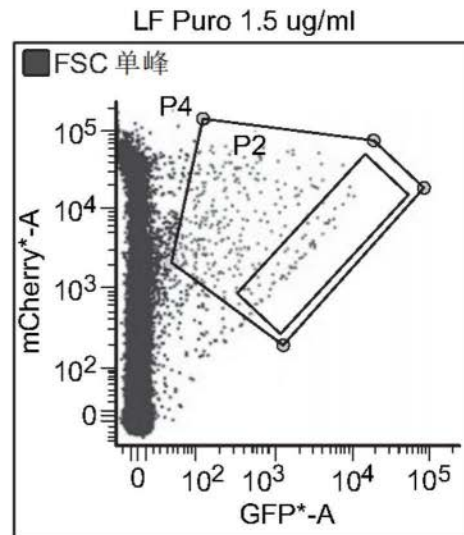
P2: 45 度对角线门

P8: 门 1 水平线高 GFP

P9: 门 2 水平线中 GFP

P10: 门 3 水平线低 GFP

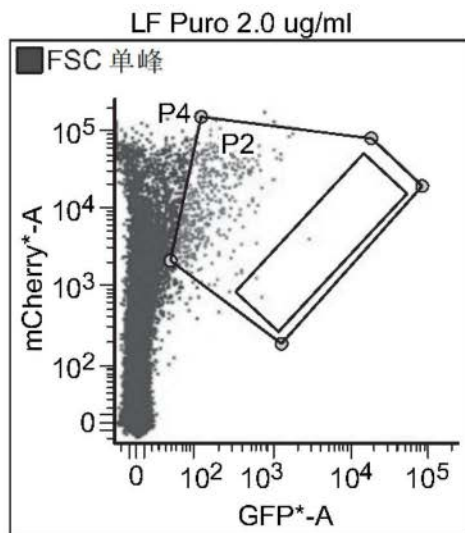
图 7C



P4: GFP 强分选

P2: 45 度对角线门

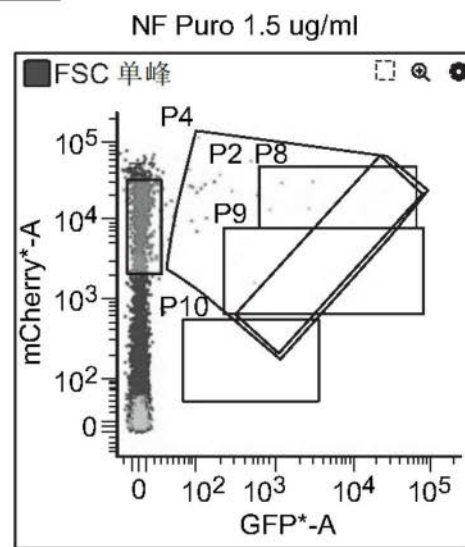
图 7D



P4: GFP 强分选

P2: 45 度对角线门

图 7E



P11: mCherry 可行性对照

P4: GFP-mCherry 阳性

P2: 45 度对角线门

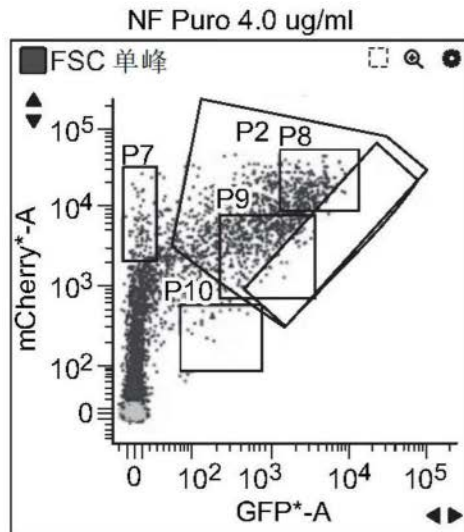
P7: mCherry 可行性对照

P8: 门 1 水平线高 GFP

P9: 门 2 水平线中 GFP

P10: 门 3 水平线低 GFP

图 7F



P2: 45 度对角线门

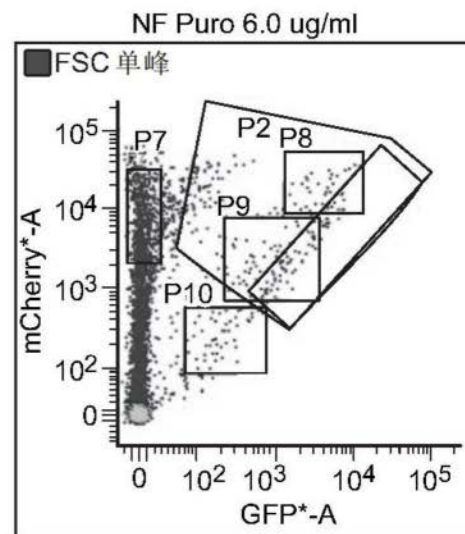
P7: mCherry 可行性对照

P8: 门 1 水平线高 GFP

P9: 门 2 水平线中 GFP

P10: 门 3 水平线低 GFP

图 7G



P2: 45 度对角线门

P7: mCherry 可行性对照

P8: 门 1 水平线高 GFP

P9: 门 2 水平线中 GFP

P10: 门 3 水平线低

图 7H

克隆	条件	试剂	[Puro, ug/mL]	分选用门
1.L1w.3	D	LF	1.5	全部
1.L1w.6	D	LF	1.5	全部
1.L1w.9	D	LF	1.5	全部
1.L2w.12	E	LF	2.0	全部
2-1.N1S.5	F	NF	1.5	45 度门
2-2.N4S.2	G	NF	4.0	45 度门
2-2.N4S.3	G	NF	4.0	45 度门
2-2.N4S.4	G	NF	4.0	45 度门
2-2.N4S.5	G	NF	4.0	45 度门
2-2.N6S3	H	NF	6.0	45 度门

图7I

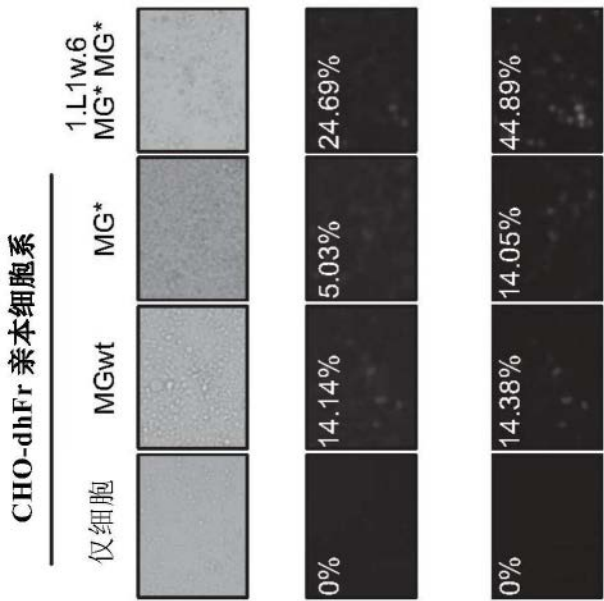


图8A

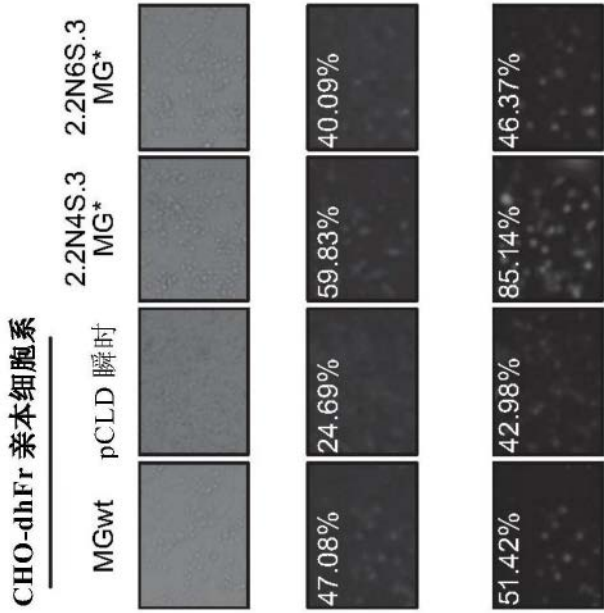


图8B

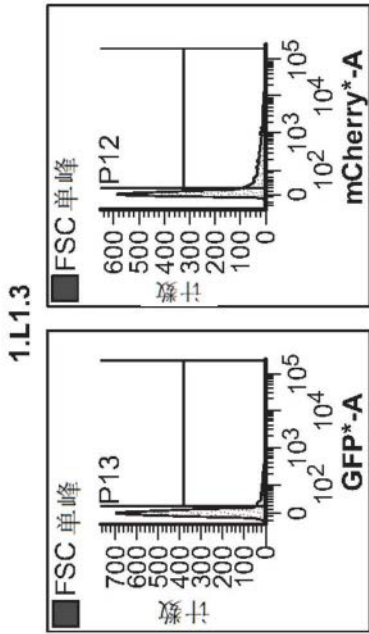


图9A

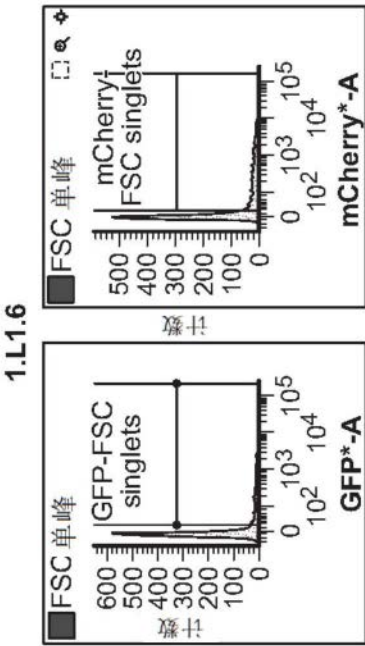


图9B

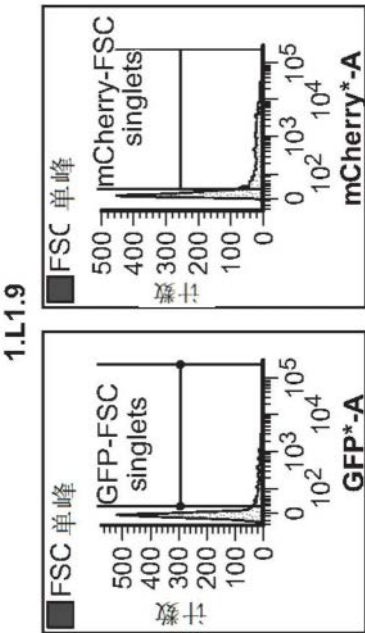


图9C

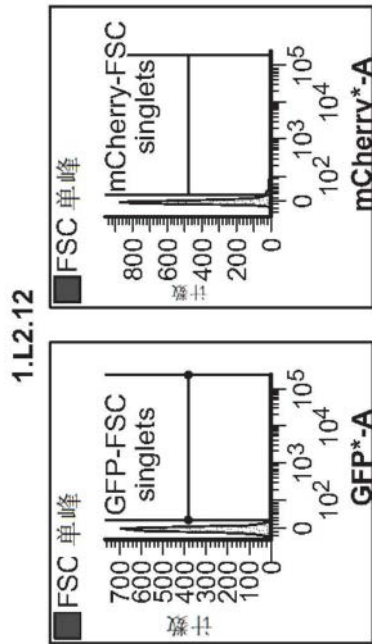


图9D

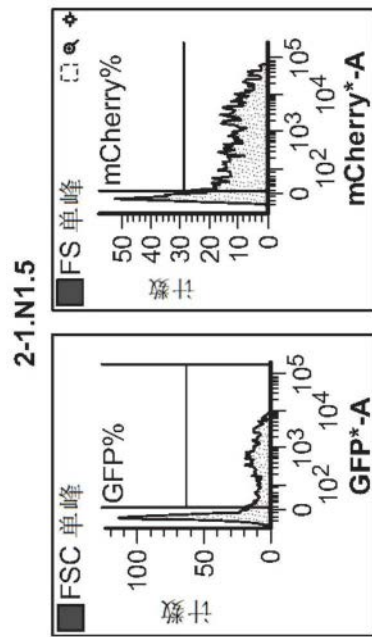


图9E

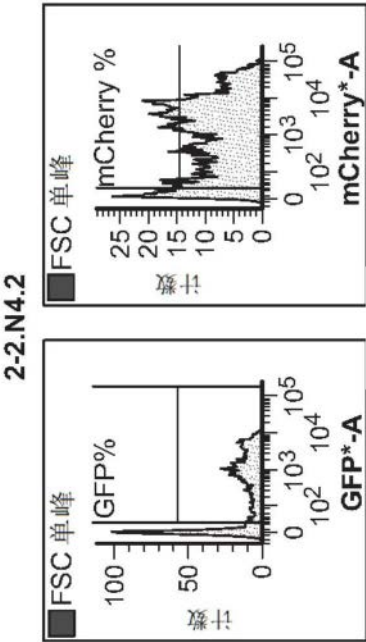


图9F

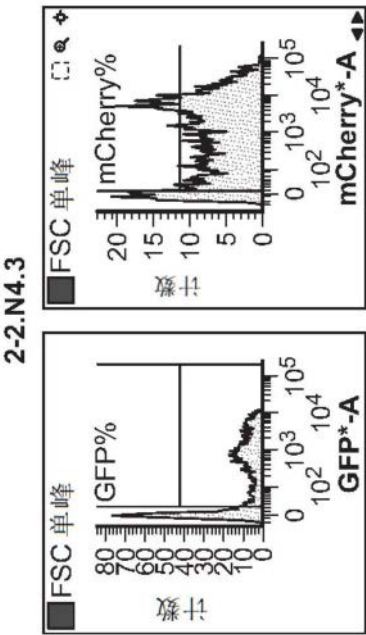


图9G

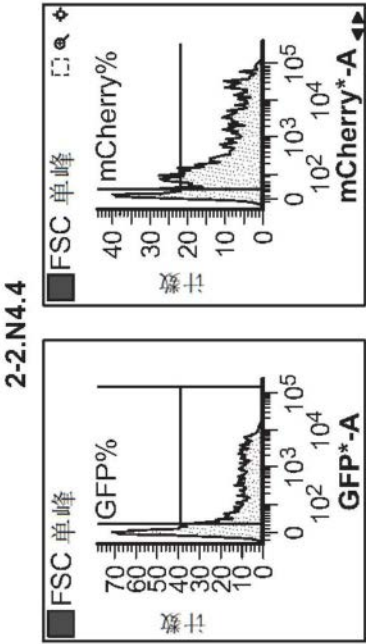


图9H

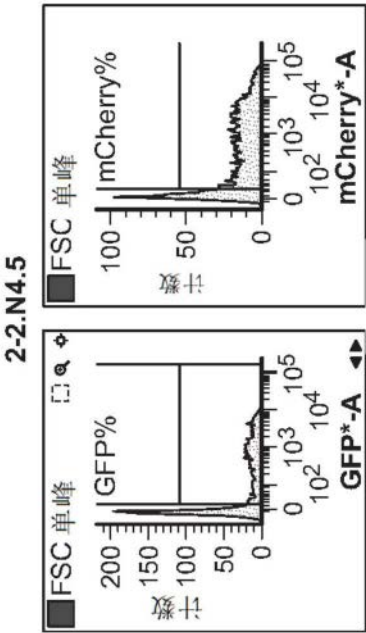


图9I

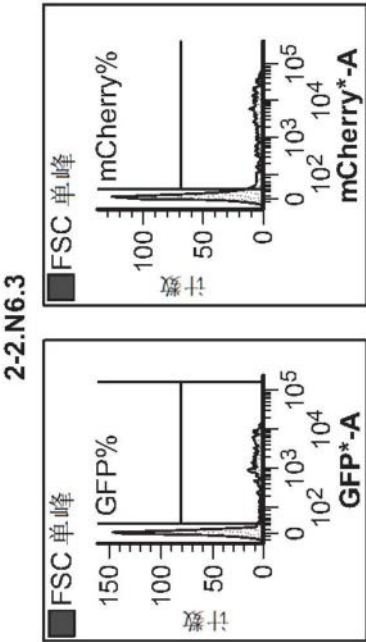


图9J

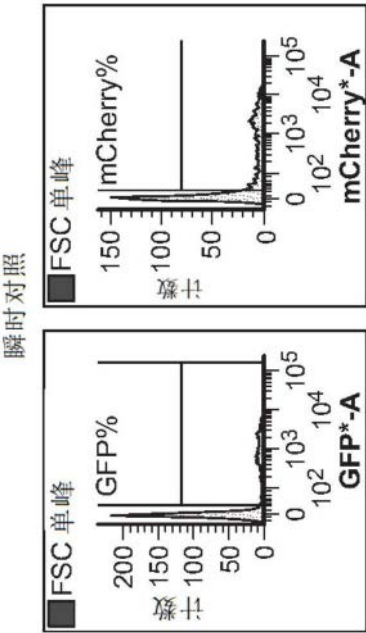


图9K

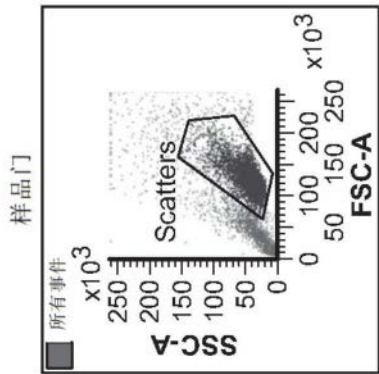


图9L

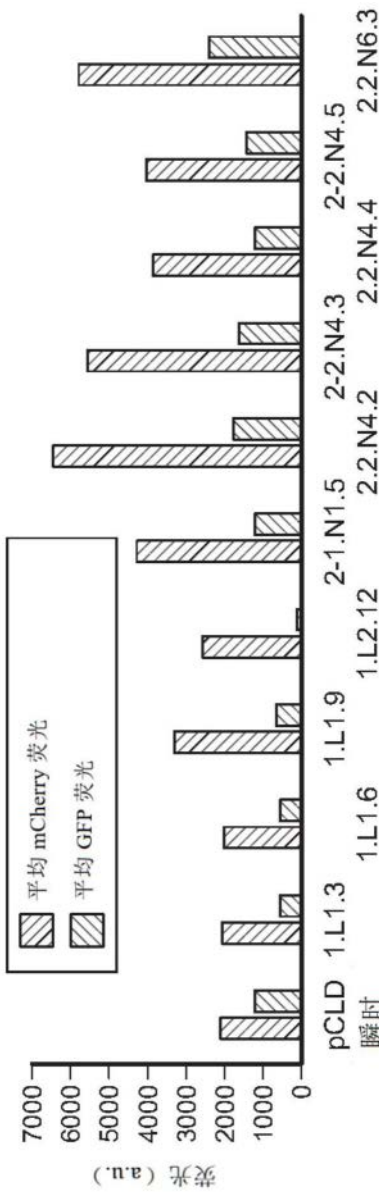


图10A

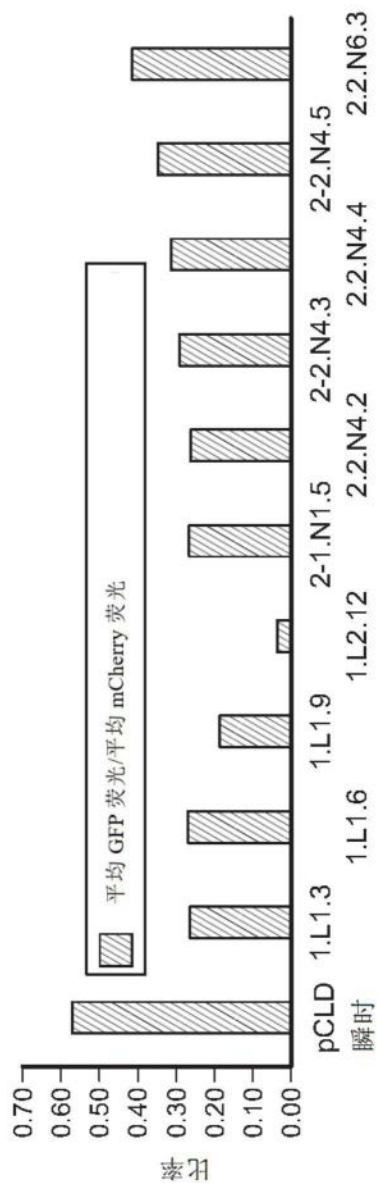


图10B

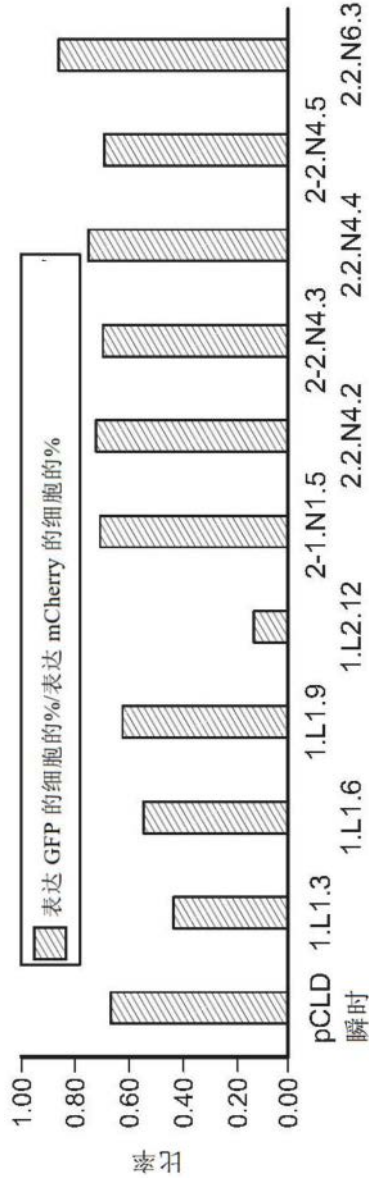


图10C

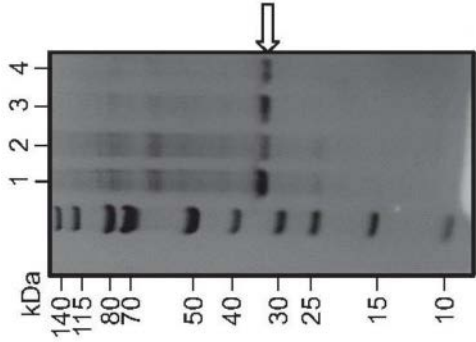


图11

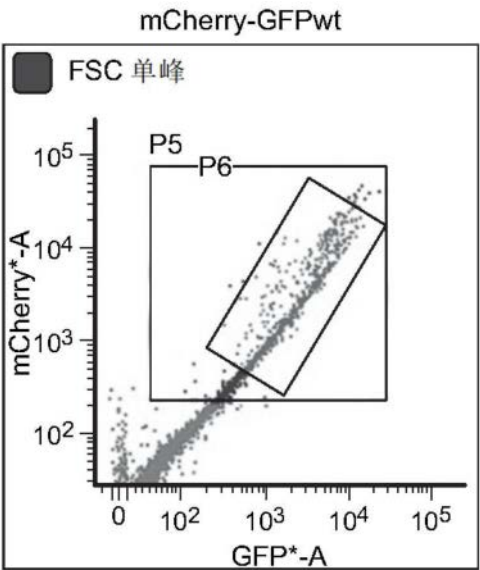


图12A

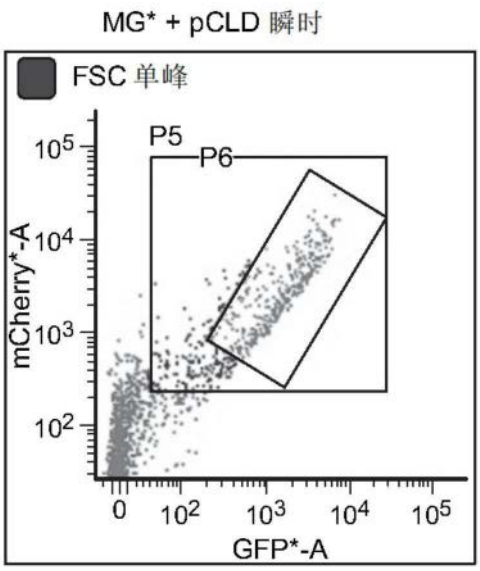


图12B

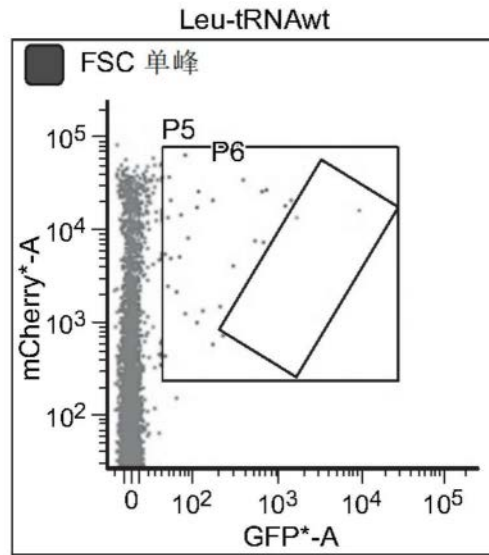


图12C

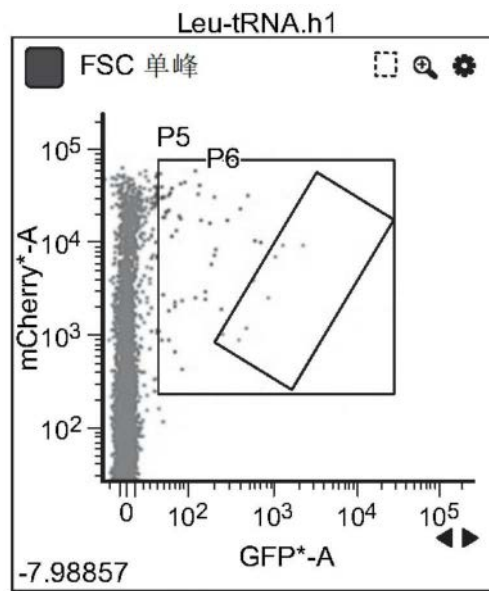


图12D

4xU6 Leu-tRNA	#P5 门的克隆数目	#P6 门的克隆数目
WT	31	2
h1	55	8

图12E

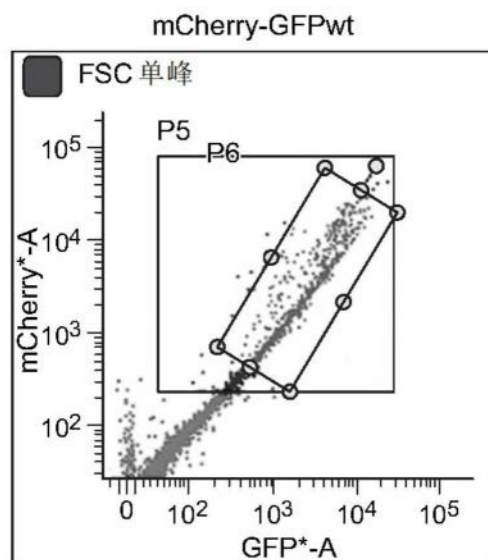


图13A

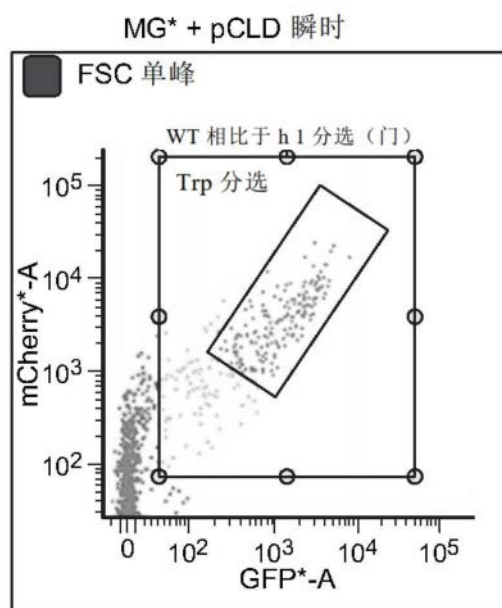


图13B

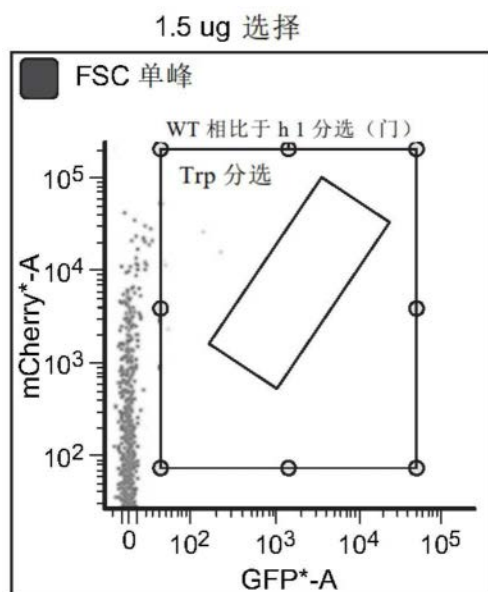


图13C

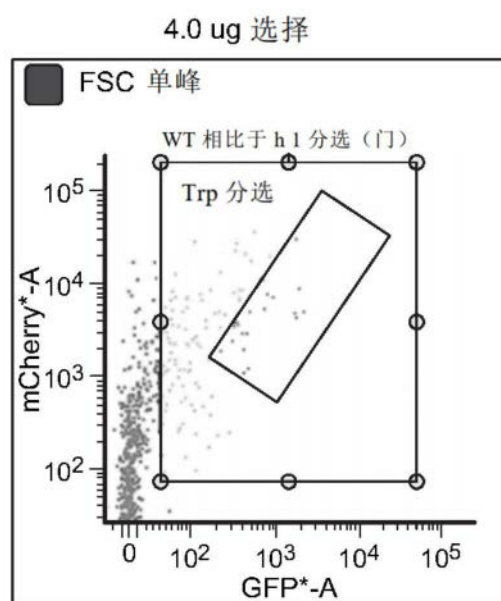


图13D

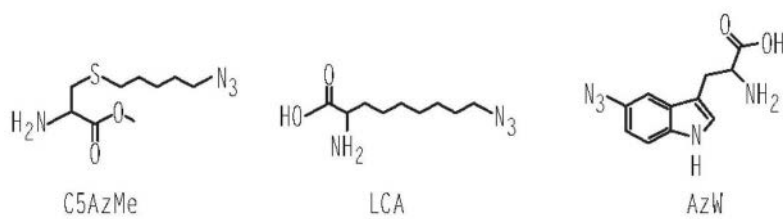


图14A

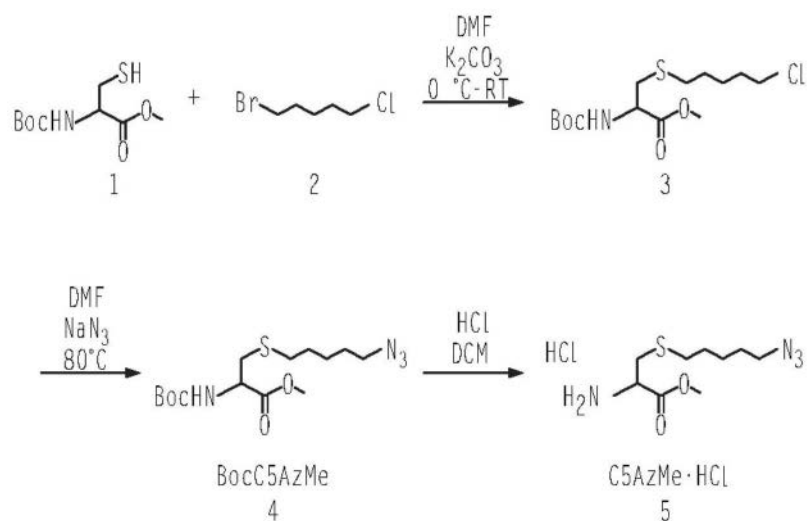


图14B

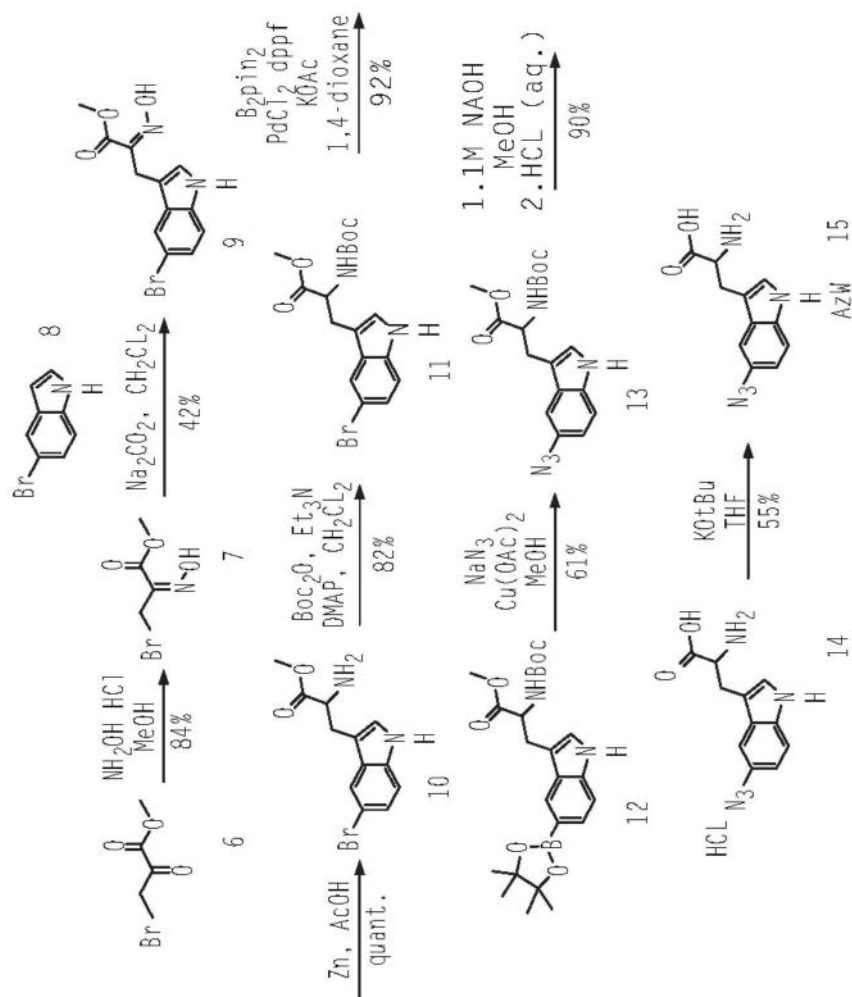


图14C

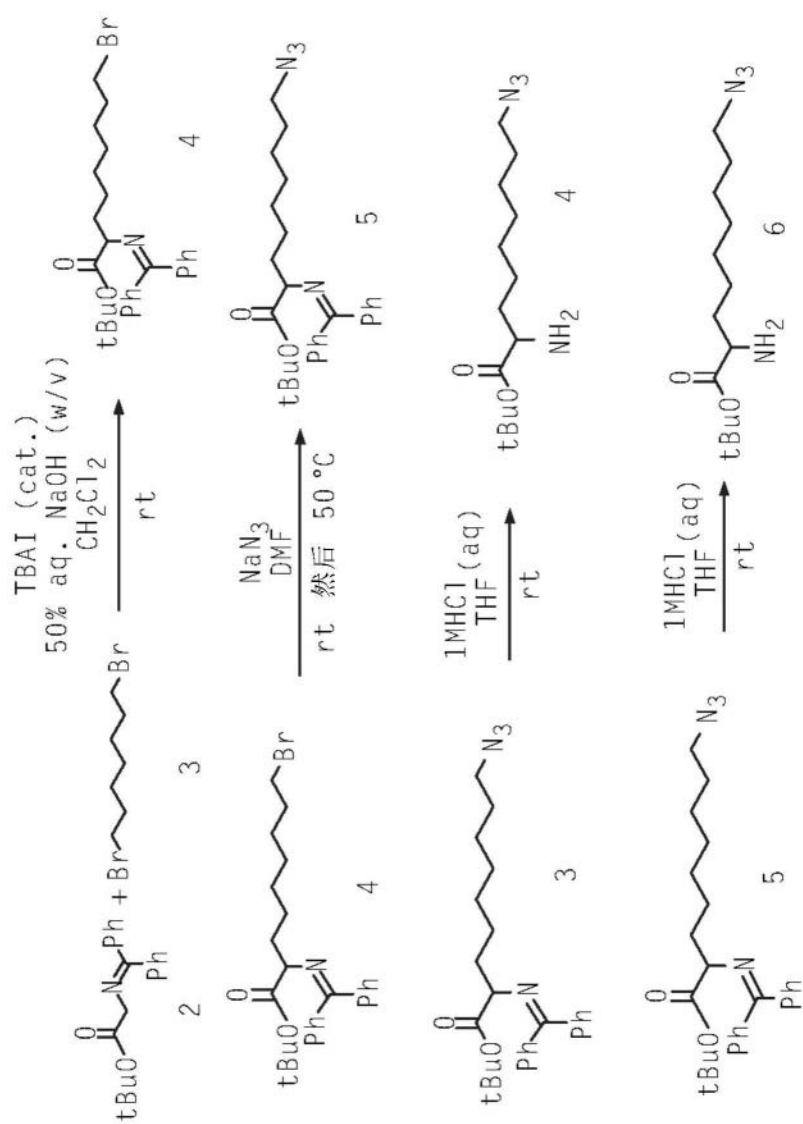


图14D

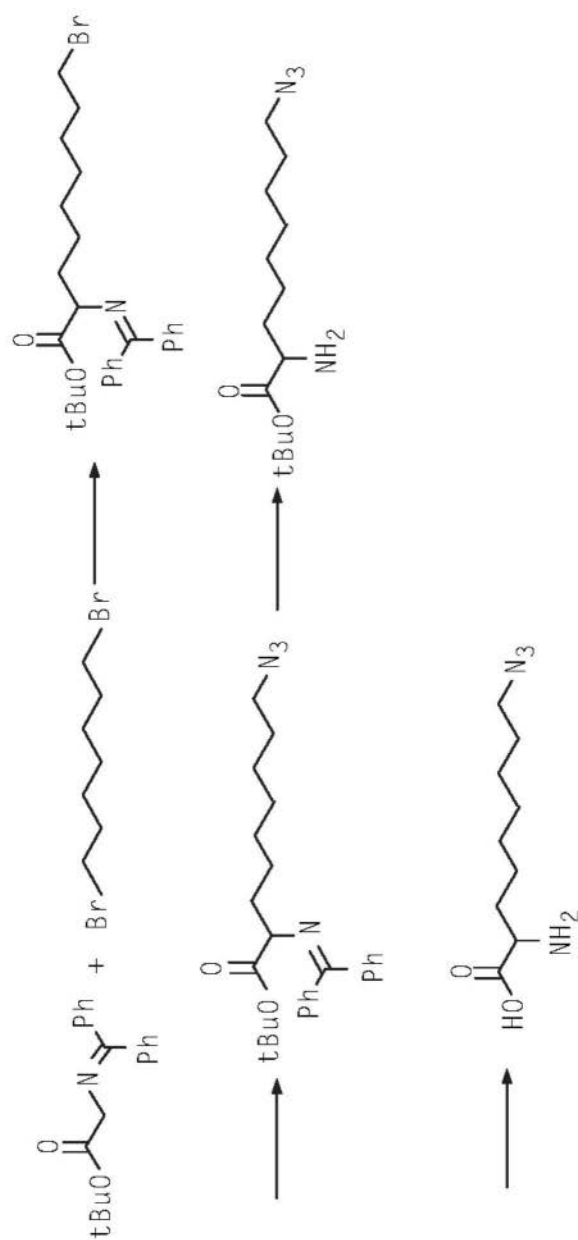


图14E

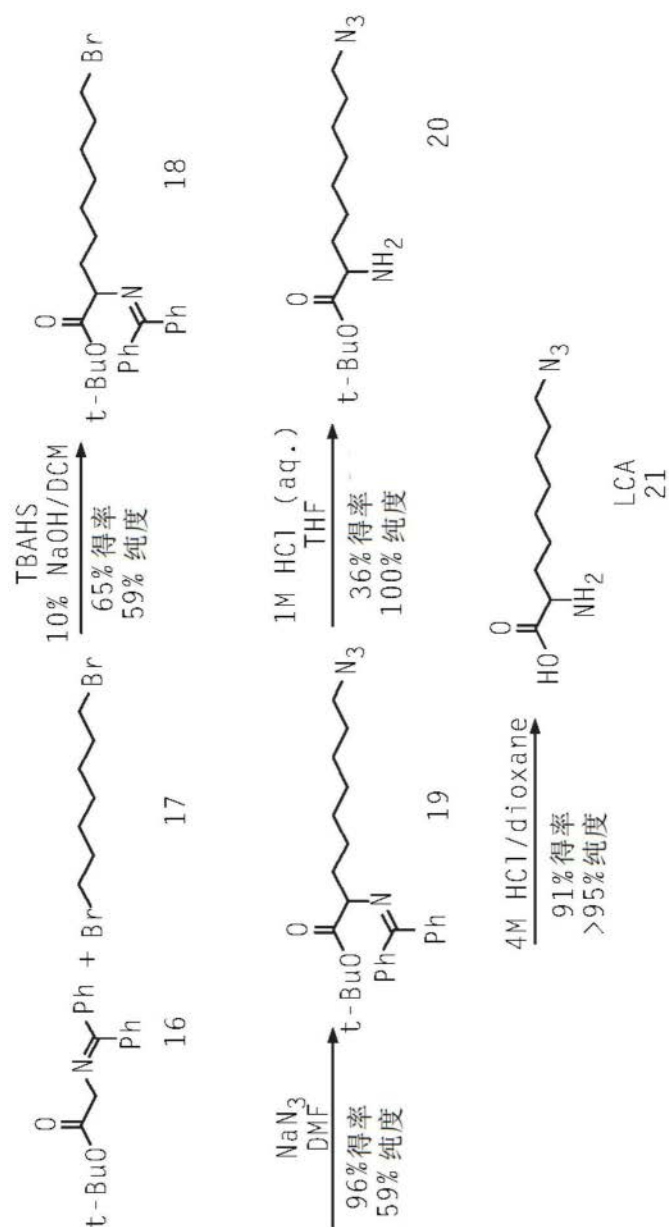


图14F