

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200910004743.5

[51] Int. Cl.

C12Q 1/25 (2006.01)

C12N 9/00 (2006.01)

C12P 21/02 (2006.01)

[43] 公开日 2009 年 9 月 16 日

[11] 公开号 CN 101532046A

[22] 申请日 2004.10.13

[21] 申请号 200910004743.5

分案原申请号 200480030137.6

[30] 优先权

[32] 2003.10.14 [33] US [31] 60/511, 532

[71] 申请人 斯克利普斯研究院

地址 美国加利福尼亚州

[72] 发明人 L·阿方达 P·G·舒尔茨

张致文

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 韦 东

权利要求书 2 页 说明书 48 页 附图 6 页

[54] 发明名称

将氧化还原活性氨基酸位点特异地掺入蛋白质

[57] 摘要

本发明涉及将氧化还原活性氨基酸位点特异地掺入蛋白质。具体而言，本发明提供了产生蛋白生物合成机制各组分的组合物和方法，所述蛋白生物合成机制包含正交 tRNA、正交氨酰 tRNA 合成酶和 tRNA/合成酶正交对，该生物合成机制可将氧化还原活性氨基酸掺入到蛋白质中。还提供了鉴定这些正交对的方法以及利用这些正交对产生含氧化还原活性氨基酸的蛋白质的方法。

1. 一种鉴定利用氧化还原活性氨基酸特异性氨酰化 O-tRNA 的正交氨酰 tRNA 合成酶的方法，所述方法包括：

a) 选择第一物种的细胞群，此群细胞各包含：

i) 多种氨酰 tRNA 合成酶(RS)中的一个成员；

ii) 来自一个或多个物种的正交 tRNA(O-tRNA)；

iii) 编码选择标记并含有至少一个被所述 O-tRNA 识别的选择者密码子的多核苷酸；

iv) 所述氧化还原活性氨基酸；和

v) 内源性翻译系统；

其中，抑制效率增强的所述细胞群中的细胞含有多种 RS 中的成员，所述成员是用所述氧化还原活性氨基酸特异性氨酰化 O-tRNA 的活性 RS；和

b)选择用氧化还原活性氨基酸氨酰化 O-tRNA 的活性 RS，从而鉴定特异性氨酰化 O-tRNA 的正交氨酰 tRNA 合成酶，其中所述活性 RS 在所述第一物种的细胞中的效率为包含所述内源性翻译系统、SEQ ID NO: 2 所示的 O-tRNA、3,4-二羟基-L-苯丙氨酸 (DHP) 和含有 SEQ ID NO:1 所示氨基酸序列的氨酰 tRNA 合成酶的对比翻译系统效率的至少 50%；和

所述活性 RS 用选自以下的氧化还原活性氨基酸优先氨酰化所述 O-tRNA：3,4-二羟基-L-苯丙氨酸 (DHP)、3,4,5-三羟基-L-苯丙氨酸、3-硝基-酪氨酸、4-硝基-苯丙氨酸和 3-硫醇-酪氨酸。

2. 如权利要求 1 所述的方法，其特征在于，所述选择包括正选择，所述选择标记包括正选择标记。

3. 如权利要求 1 所述的方法，其特征在于，所述多种 RS 包括突变型 RS、衍生自第一物种之外的一个或多个物种的 RS，或突变型 RS 和衍生自第一物种之外某物种的 RS 两者。

4. 一种用权利要求 1 所述方法鉴定的正交氨酰 tRNA 合成酶。

5. 一种在细胞内产生特定位置含有氧化还原活性氨基酸的蛋白质的方法，所述方法包括：

a) 在合适的培养基中培养细胞，所述细胞含有包含至少一个选择者密码子并编码

蛋白质的核酸；和

b) 提供氧化还原活性氨基酸；

其中所述细胞还包含：

内源性翻译系统；

在细胞中具有功能并识别该选择者密码子的正交 tRNA(O-tRNA)；和

用氧化还原活性氨基酸优先氨酰化 O-tRNA 的正交氨酰 tRNA 合成酶(O-RS)，

其中所述 O-RS 在所述细胞中的效率为包含所述内源性翻译系统、SEQ ID NO: 2 所示的 O-tRNA、3,4-二羟基-L-苯丙氨酸 (DHP) 和含有 SEQ ID NO:1 所示氨基酸序列的氨酰 tRNA 合成酶的对比翻译系统效率的至少 50%；和

所述 O-RS 用选自以下的氧化还原活性氨基酸优先氨酰化所述 O-tRNA：3,4-二羟基-L-苯丙氨酸 (DHP)、3,4,5-三羟基-L-苯丙氨酸、3-硝基-酪氨酸、4-硝基-苯丙氨酸和 3-硫醇-酪氨酸；和

c) 在含有至少一个选择者密码子的核酸翻译时将氧化还原活性氨基酸掺入蛋白质的特定位置，从而产生在特定位置含有氧化还原活性氨基酸的蛋白质。

6. 如权利要求 5 所述的方法，其特征在于，所述 O-RS 具有含有 SEQ ID NO: 1 的氨基酸序列，或在 SEQ ID NO: 1 所示氨基酸序列中取代、删除或添加一个或多个氨基酸并编码具有 SEQ ID NO: 1 氨酰-tRNA 合成酶活性的蛋白质的保守变体。

7. 如权利要求 5 所述的方法，其特征在于，所述细胞是非真核细胞。

8. 如权利要求 7 所述的方法，其特征在于，所述非真核细胞是大肠杆菌细胞。

9. 一种含有蛋白质的组合物，其特征在于，所述蛋白质含有两种或多种氧化还原活性氨基酸，所述氧化还原活性氨基酸选自 3,4-二羟基-L-苯丙氨酸 (DHP)、3,4,5-三羟基-L-苯丙氨酸、3-硝基-酪氨酸、4-硝基-苯丙氨酸和 3-硫醇-酪氨酸。

10. 如权利要求 9 所述的组合物，其特征在于，所述蛋白质含有的氨基酸序列与野生型治疗蛋白、诊断蛋白、工业用酶或其部分至少 75%相同。

11. 如权利要求 9 所述的组合物，其特征在于，所述组合物含有药学上可接受的载体。

12. 如权利要求 9 所述的组合物，其特征在于，所述两种或多种氧化还原活性氨基酸各不相同。

## 将氧化还原活性氨基酸位点特异地掺入蛋白质

本发明专利申请是国际申请号 PCT/US2004/034089、国际申请日为 2004 年 10 月 13 日，进入中国国家阶段的申请号为 200480030137.6，名称为“将氧化还原活性氨基酸位点特异地掺入蛋白质”的发明专利申请的分案申请。

### 相关申请的交叉参考

本申请要求 2003 年 10 月 14 日提交的美国临时专利申请 USSN 60/511,532 的优先权益，本申请已将其公开的内容完整纳入作为参考。

### 有关由联邦政府资助的研究和开发项目所形成的发明的权利声明

本发明是在政府资助下完成的，资助来自于国立卫生研究院的基金 No.GM 66494 和能源部的基金 DE-FG03-OOER45812。因此政府对本发明享有一定的权利。

### 发明领域

本发明属于翻译生物化学领域。本发明涉及产生和使用正交 tRNA、正交氨酰 tRNA 合成酶及正交 tRNA/正交氨酰 tRNA 合成酶对的组合物和方法，这种对中的蛋白质掺入有氧化还原活性氨基酸。本发明还涉及利用这种对及相关组合物在细胞内产生蛋白质的方法。

### 发明背景

在 20 种常见的遗传编码的氨基酸中只有半胱氨酸已被氧化还原，其结果是它可参与各种各样酶催化的氧化还原反应(Surdhar 和 Armstrong(1987) J. Phys. Chem., 91:6532-6537; Licht 等, (1996) Science 271:477-481)。因此，大多数生物氧化还原过程需要诸如黄素、烟酰胺和金属离子之类的辅因子。少数情况下，酪氨酸和色氨酸侧链翻译后修饰所产生的醌类被用作氧化还原辅因子(Stubbe 和 Van der Donk (1998) Chem. Rev., 98:705-762)。例如，牛血浆铜胺氧化酶可利用 3,4,6-三羟基-L-苯丙氨酸

(TOPA)将伯胺和分子氧分别转化成醛和过氧化氢(Janes 等, (1990) Science 248:981-987)。这些氨基酸衍生的氧化还原催化剂是由自由基介导的反应和酶反应产生的(Rodgers 和 Dean(2000) Int. J. Biochem. Cell Biol., 32:945-955)。毫无疑问, 遗传编码其它氧化还原活性氨基酸而并非通过复杂的翻译后机制产生它们的能力将显著增强对蛋白质研究和工程构建电子传递过程的能力。本发明满足了这些要求和其它要求, 且通过以下描述这将是显而易见的。

### 发明概述

本发明提供了制造正交组分的组合物和方法, 所述正交组分用于例如在体内响应选择者密码子(selector codon), 例如, 终止密码子、无义密码子、四碱基或多碱基密码子等, 将氧化还原活性氨基酸掺入到正在延伸的多肽链中。例如, 本发明提供了正交 tRNA(O-tRNA)、正交氨酰 tRNA 合成酶(O-RS)和它们组成的对。这些对可用来将氧化还原活性氨基酸掺入到正在延伸的多肽链中。

一些实施方案中, 本发明的组合物可包括正交氨酰 tRNA 合成酶(ORS), 该 O-RS 能优先用氧化还原活性氨基酸氨酰化 O-tRNA。在某些实施方案中, 所述 O-RS 包括含有 SEQ ID NO:1 所示氨基酸序列或其保守变体。在本发明的某些实施方案中, 所述 O-RS 能优先氨酰化 O-tRNA, 其效率为含有 SEQ ID NO:1 所示氨基酸序列多肽的效率的至少 50%。

含有 O-RS 的组合物还可任选含有正交 tRNA(O-tRNA), 其中 O-tRNA 能识别选择者密码子。通常, 与含有 SEQ ID NO: 2 所示多核苷酸序列的 O-tRNA 或由该多核苷酸序列编码的 O-tRNA 相比, 在存在相关合成酶时, 本发明的 O-tRNA 响应选择者密码子的抑制效率至少约为, 例如, 45%、50%、60%、75%、80%或 90%或者更高。一实施方案中, O-RS 和 O-tRNA 的抑制效率加在一起比缺乏 O-RS 时 O-tRNA 的抑制效率高例如 5 倍、10 倍、15 倍、20 倍、25 倍或更高。一方面, O-RS 和 O-tRNA 的抑制效率加在一起为衍生自詹氏甲烷球菌(*Methanococcus jannaschii*)的正交酪氨酰 tRNA 合成酶对的抑制效率的至少 45%。

含有 O-tRNA 的组合物可任选包括细胞(例如, 非真核细胞, 如大肠杆菌(*E.coli*)细胞等, 或真核细胞), 和/或翻译系统。

本发明也提供了包含翻译系统的细胞(例如, 非真核细胞或真核细胞), 所述翻译系统包含正交 tRNA(O-tRNA); 正交氨酰 tRNA 合成酶(O-RS); 和氧化还原活性氨基酸。O-RS 通常可优先氨酰化 O-tRNA, 其效率为含有 SEQ ID NO:1 所示氨基酸序列

多肽的至少 50%。O-tRNA 能识别第一选择者密码子，而 O-RS 能优先用氧化还原活性氨基酸氨酰化 O-tRNA。一实施方案中，所述 O-tRNA 包含 SEQ ID NO:2 所示多核苷酸序列或其互补性多核苷酸序列，或由它们编码。一实施方案中，所述 O-RS 包含 SEQ ID NO:1 中任一所示氨基酸序列或其保守变体。

本发明的细胞还可任选包含其它不同的 O-tRNA/O-RS 对和第二非天然氨基酸，例如，该 O-tRNA 能识别第二选择者密码子而 O-RS 能优先用第二非天然氨基酸氨酰化 O-tRNA。任选地，本发明的细胞包括含有编码感兴趣多肽的多核苷酸的核酸，所述多核苷酸包含 O-tRNA 可识别的选择者密码子。

一些实施方案中，本发明的细胞包括大肠杆菌细胞，所述细胞包含正交 tRNA(O-tRNA)、正交氨酰 tRNA 合成酶(O-RS)、氧化还原活性氨基酸以及含有编码感兴趣多肽的多核苷酸的核酸，所述多核苷酸包括 O-tRNA 可识别的选择者密码子。在本发明的某些实施方案中，所述 O-RS 优先氨酰化 O-tRNA，其效率为含有 SEQ ID NO:1 所示氨基酸序列的多肽的至少 50%。

在本发明的某些实施方案中，本发明的 O-tRNA 包含 SEQ ID NO:2 所示多核苷酸序列或其互补性多核苷酸序列，或由它们编码。在本发明的某些实施方案中，O-RS 包含 SEQ ID NO:1 所示的氨基酸序列或其保守变体。一实施方案中，所述 O-RS 或其一部分由编码 SEQ ID NO:1 所示氨基酸的多核苷酸序列或其互补多核苷酸序列编码。

本发明的 O-tRNA 和/或 O-RS 可衍生自多种生物体中的任何一种(例如，真核和/或非真核生物)。

多核苷酸也是本发明的一个特征。本发明的多核苷酸包括人造(例如，人造或非天然产生的)多核苷酸，其含有编码 SEQ ID NO:1 所示氨基酸的核苷酸序列和/或与上述多核苷酸序列互补或编码上述多核苷酸序列的核苷酸序列。本发明的多核苷酸还包括能在高度严格条件下以基本全长的核酸与上述多核苷酸杂交的核酸。本发明的多核苷酸还包括，例如与天然产生的 tRNA 至少有 75%、至少 80%、至少 90%、至少 95%、至少 98%或更高相同性的多核苷酸(但本发明的多核苷酸不是天然产生的 tRNA)。例如，与上述任何一种和/或含有上述任何一种的保守变体的多核苷酸至少有 80%、至少 90%、至少 95%、至少 98%或更高相同性的人造多核苷酸也包括在本发明的多核苷酸之内。

含有本发明多核苷酸的载体也是本发明的一个特征。例如，本发明的载体可包括质粒、粘粒、噬菌体、病毒、表达载体等。含有本发明载体的细胞也是本发明的一

个特征。

产生 O-tRNA/O-RS 对诸组分的方法也是本发明的特征。用这些方法产生的组分也是本发明的特征。例如，产生至少一种能与细胞正交的 tRNA(O-tRNA)的方法包括产生突变型 tRNA 文库；使突变型 tRNA 文库每个成员的反密码子环突变以识别选择者密码子，从而提供潜在 O-tRNA 文库，并对第一物种的第一细胞群进行负选择，其中所述细胞含有潜在 O-tRNA 文库的成员。负选择可消除潜在 O-tRNA 文库中的含有可被该细胞内源性氨酰 tRNA 合成酶(RS)氨酰化的成员。这提供了对第一物种的细胞为正交性的 tRNA 库，从而提供至少一种 O-tRNA。还提供了用本发明方法产生的 O-tRNA。

在某些实施方案中，所述方法还包括对第一物种的第二细胞群进行正选择，其中所述细胞含有对第一物种的细胞为正交性的 tRNA 库的成员、相关氨酰 tRNA 合成酶和正选择标记。使用正选择可选出或筛选到含有可被相关氨酰 tRNA 合成酶氨酰化并在正选择标记存在时显示出所需反应的 tRNA 库成员的细胞，从而提供 O-tRNA。在某些实施方案中，第二细胞群包括未被负选择去除的细胞。

还提供了鉴定与 O-tRNA 一起使用的氧化还原活性氨基酸的正交氨酰 tRNA 合成酶的方法。例如，所述方法包括选择第一物种的细胞群，该群中的细胞各包含：1)多种氨酰 tRNA 合成酶(RS)中的一个成员，(例如，所述多种 RS 可包括突变型 RS、衍生自第一物种之外的物种的 RS，或突变型 RS 和衍生自第一物种之外的物种的 RS 两者)；2)正交 tRNA(O-tRNA)(例如，来自一个或多个物种)；和 3)编码正选择标记并含有至少一个选择者密码子的多核苷酸。

选出或筛选到的细胞(例如，宿主细胞)与缺乏所述多种 RS 中的一个成员或含有减少数目的该成员的细胞相比，显示出增强的抑制效率。这些选出/筛选到的细胞含有能氨酰化 O-tRNA 的活性 RS。用这种方法鉴定正交氨酰 tRNA 合成酶也是本发明的一个特征。

在细胞(例如，非真核细胞，如大肠杆菌细胞等，或真核细胞)内特定位置制造含氧化还原活性氨基酸的蛋白质的方法也是本发明的一个特征。例如，一种方法包括使细胞在合适的培养基中生长，其中所述细胞含有包含至少一个选择者密码子并编码蛋白质的核酸，提供氧化还原活性氨基酸，和在含有至少一个选择者密码子的核酸翻译过程中将氧化还原活性氨基酸掺入该蛋白质的特定位置，从而产生该蛋白质。所述细胞还包含：在细胞中具有功能和识别选择者密码子的正交 tRNA(O-tRNA)；和能优先用氧化还原活性氨基酸氨酰化 O-tRNA 的正交氨酰 tRNA 合成酶(O-RS)。用这

种方法产生的蛋白质也是本发明的一个特征。

本发明还提供含有蛋白质的组合物，其中所述蛋白质包含氧化还原活性氨基酸(例如，3,4-二羟基-L-苯丙氨酸(phenylalanine)(DHP)、3,4,5-三羟基-L-苯丙氨酸、3-硝基-酪氨酸、4-硝基-苯丙氨酸、3-硫醇-酪氨酸等)。在某些实施方案中，所述蛋白质含有的氨基酸序列与治疗蛋白、诊断蛋白、工业用酶或其部分至少 75%相同。任选地，所述组合物包含药学上可接受的载体。

### 定义

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

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

正交：本文所用的术语“正交”指一种分子(如正交 tRNA(O-tRNA)和/或正交氨酰 tRNA 合成酶(O-RS))具有细胞内源性组分的功能，但其效力与细胞或翻译系统内的相应内源性分子相比降低，或者不具有细胞内源性组分的功能。就 tRNA 和氨酰-tRNA 合成酶而言，正交是指正交的 tRNA 具有内源性 tRNA 合成酶的功能但其效率与内源性 tRNA 和内源性 tRNA 合成酶相比降低，如低于 20%、10%、5%或 1%。正交分子缺乏细胞内内源性互补分子的正常功能。例如，与内源性的 tRNA 被内源性 RS 氨酰化相比，细胞内的正交 tRNA 被细胞的内源性 RS 氨酰化时其效率降低，甚至为零。在另外一个实施例中，与内源性的 RS 氨酰化内源性的 tRNA 相比，正交的 RS 在感兴趣细胞内氨酰化任何内源性的 tRNA 时其效率降低，或者甚至为零。第二正交分子可以被引入到细胞内和第一正交分子一起发挥功能。例如，正交 tRNA/RS 对包括引入的互补组分，在细胞内一起行使功能时其效率与对照，如相应的内源性 tRNA/RS 对或活性正交对(如酪氨酰正交 tRNA/RS 对)相比可以达到 45%、50%、60%、70%、75%、80%、90%、95%、99%或更高。

正交酪氨酰-tRNA：本文所用的术语正交酪氨酰-tRNA(酪氨酰-O-tRNA)是与感兴趣翻译系统正交的 tRNA，其中 tRNA：(1)与天然产生的酪氨酰 tRNA 相同或基本类似，(2)来源于发生天然突变或人工诱变的天然酪氨酰 tRNA，(3)由考虑到(1)或(2)的野生型或突变型酪氨酰 tRNA 序列的方法所产生，(4)与野生型或突变型酪氨酰 tRNA



同源, (5)与表 1 中所列称为酪氨酰 tRNA 合成酶底物的典型 tRNA 同源, 或(6)表 1 中称为酪氨酰 tRNA 合成酶底物的典型 tRNA 的保守变体。酪氨酰 tRNA 可以携带一个氨基酸, 或者处于非负载状态。还需要说明“酪氨酰-O-tRNA”可以被相关的合成酶加载(氨酰化)赖氨酸之外的氨基酸, 如高谷氨酰胺。应理解本发明的酪氨酰-O-tRNA 确实可用于将任何必需的氨基酸, 无论是天然的或是人工合成的, 掺入到正在延伸的多肽内, 如翻译期间, 以响应选择者密码子。

**正交酪氨酰氨基酸合成酶:** 本文所用的正交酪氨酰氨基酸合成酶(酪氨酰-O-RS)是一种可以在感兴趣翻译系统内优先用氨基酸氨酰化酪氨酰-O-tRNA 的合成酶。被酪氨酰-O-RS 加载到酪氨酰-O-tRNA 上的氨基酸可以是任何氨基酸, 无论是天然的还是人工合成的, 不限于本文所述的。该合成酶还任选可以与天然的酪氨酰氨基酸合成酶相同或同源, 或者与表 1 中称为酪氨酰-O-RS 的合成酶相同或同源。例如, 酪氨酰-O-RS 可以是表 1 所列的酪氨酰-O-RS 的保守变体, 和/或与表 1 所列的酪氨酰-O-RS 的序列至少有 50%、60%、70%、80%、90%、95%、98%、99%或更高的相同性。

**同源:** 术语“同源”指能联合发挥功能的组分, 如正交 tRNA 和正交氨酰-tRNA 合成酶。此种组分也称为互补组分。

**优先氨酰化:** 术语“优先氨酰化”是指, 与 O-RS 氨酰化天然产生的 tRNA 或氨酰化用于产生 O-tRNA 的原料的效率相比, O-RS 用氧化还原活性氨基酸氨酰化 O-tRNA 的效率为, 例如 70%、75%、85%、90%、95%、99%或更高。

**选择者密码子:** 术语“选择者密码子”指在翻译过程中能被 O-tRNA 识别但通常不被内源性 tRNA 识别的密码子。O-tRNA 反密码子环能识别 mRNA 上的选择者密码子并能在多肽的该位点掺入此氨基酸, 例如非天然氨基酸, 如氧化还原活性氨基酸。选择者密码子包括无义密码子, 如终止密码子, 例如琥珀密码子、赭石密码子和乳白密码子; 四碱基或多碱基密码子; 稀有密码子; 天然或非天然碱基对产生的密码子等。

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

**抑制活性:** 如本文采用的, 术语“抑制活性”通常指 tRNA(如抑制型 tRNA)翻译阅读通过某密码子(如为琥珀密码子或四碱基或多碱基密码子的选择者密码子)的能力, 如果阅读不通过则翻译终止或产生错译(如移码突变)。抑制型 tRNA 的抑制活性可以表示为与第二抑制型 tRNA 或与对照系统如缺乏 O-RS 的对照系统相比所观察到

的翻译通读活性的百分数。

本发明提供了定量测定抑制活性的各种方法。一种特定 O-tRNA 和 ORS 对在对感兴趣的选择者密码子(如琥珀密码子)反应时的抑制百分率,指一种位于编码可表达检测标记的核酸中的某给定表达检测标记(如 LacZ),包括选择者密码子,在感兴趣翻译系统内的活性与阳性对照构建物活性相比的百分数,其中的感兴趣的翻译系统含有 O-RS 和 O-tRNA,而阳性对照不含 O-tRNA、O-RS 和选择者密码子。因此,如果某缺乏选择者密码子的活性阳性对照标记构建物在一个给定翻译系统内的活性是 X,那么含有选择者密码子的被检测构建物的抑制百分率为被检测构建物在与阳性对照标记基本相同的表达环境条件下所显示的 X 的百分比,其中该被检测标记构建物所表达的翻译系统也含有 O-tRNA 和 O-RS。一般来说,表达被检测标记的翻译系统还包含能被 O-RS 和 O-tRNA 识别的氨基酸。任选地,抑制百分率的测定可以通过将该被检测标记与“背景”或“阴性”对照标记构建物比较而精细确定,后者含有与被检测标记相同的选择者密码子,但是该系统中不包含 O-tRNA、O-RS 和/或能被 O-tRNA 和/或 O-RS 识别的相关氨基酸。这种阴性对照可用于对该标记在感兴趣翻译系统所产生的背景信号进行处理使抑制百分率测定标准化。

抑制效率可通过本领域熟知的多种方法测定。例如,可利用  $\beta$  半乳糖苷酶报告分子试验,如将衍生的 lacZ 质粒(该构建物在 lacZ 核酸序列中含有选择者密码子)和含有本发明的 O-tRNA 的质粒一起导入到合适生物(如可利用此正交组分的生物)的细胞内。也可导入相关合成酶(或者以多肽或编码相关合成酶的多核苷酸形式)。将这些细胞在培养基内培养到所需的密度,如 OD<sub>600</sub> 达到约 0.5 时,采用 BetaFluor™  $\beta$ -半乳糖苷酶检测试剂盒(Novagen)进行  $\beta$  半乳糖苷酶试验。抑制百分率可以计算为样品相对于可比较对照,如衍生的 lacZ 构建物所显示的活性的百分比,其中所述构建物在所需位置上含有相应的有义密码子而不是选择者密码子。

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

**非天然氨基酸:** 本文所用术语“非天然氨基酸”指 20 个常见天然氨基酸或硒代半胱氨酸或吡咯赖氨酸之外的任何氨基酸、修饰的氨基酸和/或氨基酸类似物,如氧化还原活性氨基酸。

衍生于：本文所用的术语“衍生于”指从特定分子或生物、或者根据该特定分子或生物的信息分离的组分，或者利用该特定分子或生物、或者根据该特定分子或生物的信息产生的组分。

阳性选择或筛选标记：本文所用的术语“阳性选择标记”指当一种标记存在时，可被表达、活化等，会导致含有该标记相应性状的细胞得以鉴定，如可以将含有该阳性选择标记的细胞与无此性状的细胞相区别。

阴性选择或筛选标记：本文所用的术语“阴性选择标记”指当一种标记存在时，如被表达、活化等，可以鉴定不含有所选特性或性状的细胞(如，与具有该特性或性状的细胞相比较)。

报告分子：本文所用的术语“报告分子”是指可用于鉴定和/或筛选感兴趣系统内靶组分的组分。例如，报告分子可以包括蛋白质，如能赋予抗生素抗性或敏感性的酶(如  $\beta$ -内酰胺酶、氯霉素乙酰转移酶(CAT)等)、荧光选择标记(如绿色荧光蛋白(如 GFP)、YFP、EGFP、RFP 等)、冷光标记(如萤火虫荧光素酶蛋白)、亲和力选择标记，或者正或负选择标记基因如 lacZ、 $\beta$ -gal/lacZ( $\beta$ -半乳糖苷酶)、Adh(乙醇脱氢酶)、his3、ura3、leu2、lys2 等。

真核生物：本文所用的术语“真核生物”指在系统进化树内属于真核生物域的生物，如动物(如哺乳动物、昆虫、爬行动物、鸟类等)、纤毛虫、植物(如单子叶植物、双子叶植物、藻类等)、真菌、酵母、鞭毛虫、微孢子虫、原生生物等。

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

保守变体：本文所用术语“保守变体”就翻译组分而言指一种翻译组分，如保守变体 O-tRNA 或保守变体 O-RS，其功能特性与基本组分 O-tRNA 或 O-RS 相似，但

是在序列上与参照 O-tRNA 或 O-RS 相比有些不同。例如, O-RS 可用非天然氨基酸, 例如 3,4-二羟基-L-苯丙氨酸(DHP)等的氧化还原活性氨基酸, 氨酰化互补的 O-tRNA 或保守变体 O-tRNA, 但是 O-tRNA 和保守变体 O-tRNA 的序列不相同。保守变体的序列可含有一个、两个、三个、四个、五个或更多突变, 只要保守变体与相应的 O-tRNA 或 O-RS 互补。

**选择或筛选试剂:** 本文所用的术语“选择或筛选试剂”指当一种试剂存在时可以从一群组分中选出或筛选出某种特定组分。例如, 选择或筛选试剂包括但不限于营养素、抗生素、光的波长、抗体、表达的多核苷酸等。筛选试剂可以使用不同的浓度、强度等。

**响应:** 本文所用的术语“响应”指本发明的 tRNA 识别选择者密码子并介导将结合于 tRNA 的氧化还原活性氨基酸掺入正在衍生的多肽链的过程。

**编码:** 本文所用的术语“编码”指利用多聚大分子或序列符号串上的信息指导产生第二分子或序列符号串的过程, 所述第二分子或序列符号串与第一分子或序列符号串不同。如本文所描述, 该术语的使用范围很宽, 可用于许多方面。一方面, 术语“编码”描述 DNA 的半保守性复制过程, 其中双链 DNA 分子的一条链作为 DNA 依赖的 DNA 聚合酶的模板编码出新合成的一条互补姊妹链。

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

### 附图简述

图 1 提供了从 3,4-二羟基-L-苯丙氨酸(DHP; 结构 1)的氧化产物到 DHP-半醌基 2, 其易被氧化成 DHP-醌 3 的示意图。

图 2A 和 2B 提供了抹香鲸肌红蛋白响应 Mb 基因第 4 位的琥珀密码子的 DHP 依赖性表达。图 2A 提供了银染胶和 western 印迹。图 2B 提供了 DHPMb 的 ESI-QqTOF 质谱分析。

图 3A 和 3B 提供了循环伏安图。图 3A 提供了 wtMb 和 DHPMb 中的血红素组的

循环伏安图。图 3B 提供了不同溶液中的 DHP 的循环伏安图, 所述溶液含有: 100 $\mu$ M DHP、wtMb 或 DHPMb。所有伏安图在 pH 7.4 的 0.1 M 磷酸缓冲液中以氩气下记录; 扫描速度: 1 V s<sup>-1</sup> 对 SCE。

图 4 提供了用电化学方法氧化蛋白质中 DHP 的示意图。

图 5 提供了詹氏甲烷球菌酪氨酰 tRNA 合成酶(MjTyrRS)的核苷酸序列和氨基酸序列。

图 6 提供了 3,4-二羟基-L-苯丙氨酸(DHP)-tRNA 合成酶(DHPRS)的核苷酸序列和氨基酸序列, 基于詹氏甲烷球菌酪氨酰 tRNA 合成酶(MjTyrRS)具有以下氨基酸改变: Tyr32 $\rightarrow$ Leu、Ala67 $\rightarrow$ Ser、His70 $\rightarrow$ Asn 和 Ala167 $\rightarrow$ Gln。改变的氨基酸和相应的三联密码子(相对野生型序列)被框出。

### 发明详述

为了能在体内将其他合成的氨基酸如氧化还原活性氨基酸掺入到遗传密码内, 需要一个能在翻译系统中发挥正常功能但是对组织翻译系统来说是“正交的”的新型氨酰-tRNA 合成酶和 tRNA 的正交对, 也就是说该正交对对内源翻译系统而言可独立地行使合成酶和 tRNA 的功能。该正交对所需要的特性包括 tRNA 只能解码或识别一个特异性的新密码, 如选择者密码子, 该密码子不能被任何内源性 tRNA 所解码, 而氨酰-tRNA 合成酶只能用特异性的氧化还原活性氨基酸优先氨酰化(或加载)其相关 tRNA。O-tRNA 通常也不能被内源性合成酶氨酰化。例如, 在大肠杆菌内, 一个正交对包含一个不能与任何内源性 tRNA(大肠杆菌内有 40 种)发生交叉反应的氨酰 tRNA 合成酶和一个不能被任何内源性合成酶(大肠杆菌内有 21 种)氨酰化的正交 tRNA。

本发明提供了鉴定和产生其它正交 tRNA-氨酰 tRNA 合成酶对(例如可用来掺入氧化还原活性氨基酸的 O-tRNA/O-RS 对)的组合物和方法。本发明的 O-tRNA 能够介导将氧化还原活性氨基酸掺入到含有例如在体内可被 O-tRNA 识别的选择者密码子的多核苷酸所编码的蛋白质中。O-tRNA 的反密码子环可识别 mRNA 上的选择者密码子并能在多肽的该位点掺入氨基酸, 如氧化还原活性氨基酸。本发明的正交氨酰 tRNA 合成酶能只用特异性的氧化还原活性氨基酸优先氨酰化(或加载)其相关 O-tRNA。

例如, 氧化还原活性氨基酸 3,4-二羟基-L-苯丙氨酸(DHP)有两个电子可被氧化成为醌而可响应选择者密码子(例如 TAG 密码子)被选择性地且有效地掺入到大肠埃希

杆菌(大肠杆菌)生物体的蛋白质中。参见图 1。蛋白质内的 DHP 可用电化学方法氧化。若能将氧化还原活性氨基酸位点特异性掺入蛋白质中将有助于研究蛋白质内的电子传递,并能工程构建具有新特性的氧化还原蛋白质。见图 4。例如,氧化还原活性蛋白质的表达有助于研究并能够改变蛋白质内电子传递途径、改变酶的催化功能、将蛋白质与小分子和生物分子交联等。

#### 正交 tRNA/正交氨酰-tRNA 合成酶及它们组成的对

适于产生含有一个或多个非天然氨基酸,例如氧化还原活性氨基酸的蛋白质的翻译系统描述于题为“METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGANOL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS”(产生正交 tRNA-氨酰-tRNA 合成酶对的方法和组合物)的国际专利申请 WO2002/086075 和题为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”(非天然氨基酸的体内掺入)的 WO2002/085923。另外,还可以参见 2004 年 4 月 16 日提交的国际专利 PCT/US2004/011786。这些专利申请都已完整纳入本文作为参考。这种翻译系统一般都包括含有正交 tRNA(O-tRNA)、正交氨酰 tRNA 合成酶(O-RS)以及氧化还原活性氨基酸的细胞(例如非真核细胞或真核细胞),其中 O-RS 可用氧化还原活性氨基酸氨酰化 O-tRNA。本发明的正交对包括 O-tRNA,如抑制型 tRNA、移码 tRNA 等,以及 O-RS。本发明还提供了其他各组分。

O-tRNA 可识别选择者密码子,与包含或编码于 SEQ ID NO: 2 所示多核苷酸序列的 O-tRNA 相比,在相关合成酶存在时,本发明的 O-tRNA 为了响应选择者密码子而产生的抑制效率至少可达 45%、50%、60%、75%、80%、90%或更高。O-RS 可用氧化还原活性氨基酸氨酰化 O-tRNA。细胞利用该组分对将氧化还原活性氨基酸掺入到正在延伸的多肽链内,例如通过包含编码感兴趣多肽的多核苷酸的核酸,该多核苷酸含有能被 O-tRNA 识别的选择者密码子。在本发明的某些实施方案中,细胞如大肠杆菌细胞包含正交 tRNA(O-tRNA)、正交氨酰-tRNA 合成酶(O-RS)、氧化还原活性氨基酸,和含有编码感兴趣多肽的多核苷酸的核酸,该多核苷酸含有能被 O-tRNA 识别的选择者密码子。翻译系统也可以是体外系统。

在一个实施方案中,O-RS 和 O-tRNA 联合在一起的抑制效率至少是缺少 O-RS 时 O-tRNA 的抑制效率的 5 倍、10 倍、15 倍、20 倍、25 倍或更高。在一方面,O-RS 和 O-tRNA 联合在一起的抑制效率是衍生自詹氏甲烷球菌的正交酪氨酰 tRNA 合成酶对抑制效率的至少约 35%、40%、45%、50%、60%、75%、80%、90%或更高。

本发明任选可在细胞内包括多个 O-tRNA/O-RS 对,从而可以掺入一个以上的非

天然氨基酸，如氧化还原活性氨基酸和其他的非天然氨基酸。例如，细胞还可以包含其他不同的 O-tRNA/O-RS 对和第二种非天然氨基酸，此另一种 O-tRNA 可识别第二选择者密码子，此另一种 O-RS 可用第二种非天然氨基酸优选氨酰化 O-tRNA。例如，含有一种 O-tRNA/O-RS 对(其中 O-tRNA 可识别琥珀选择者密码子)的细胞还可以含有第二种正交对，如亮氨酰、赖氨酰、谷氨酰对等(其中第二 O-tRNA 可识别不同的选择者密码子，如乳白密码子、四碱基密码子等)。

O-tRNA 和/或 O-RS 可以是天然产生的，或可通过突变天然的 tRNA 和/或 RS 而产生，例如这可从各种生物体产生 tRNA 文库和/或 RS 文库。例如，一种产生正交 tRNA/氨酰-tRNA 合成酶对的策略涉及输入宿主细胞之外生物体或多种生物体来源的异源性(相对于该宿主细胞而言)tRNA/合成酶对到宿主细胞内。该候选异源合成酶的特性包括不能加载宿主细胞的任何 tRNA，该候选异源 tRNA 的特性包括不能被宿主细胞的任何合成酶氨酰化。另外，该异源 tRNA 对于宿主细胞的合成酶而言是正交的。

产生正交对的第二种策略是产生突变文库，然后从中筛选出 O-tRNA 或 O-RS。这些策略也可联合使用。

### 正交 tRNA(O-tRNA)

正交 tRNA(O-tRNA)可以在体内和体外介导将氧化还原活性氨基酸掺入蛋白质内，所述蛋白质是由含选择者密码子的多核苷酸编码的，这种选择者密码子可被 O-tRNA 识别。在某些实施方案中，本发明的 O-tRNA 与包含或编码于 SEQ ID NO: 2 所列多核苷酸序列的 O-tRNA 相比，在相关合成酶存在的条件下，为了响应选择者密码子而产生的抑制效率至少可达 45%、50%、60%、75%、80%、90%或更高。

抑制效率可通过本领域熟知的多种试验检测。例如，可以利用  $\beta$  半乳糖苷酶报告分子试验，如将衍生的 lacZ 质粒(该构建物含有选择者密码子和 lacZ 核酸序列)和含有本发明 O-tRNA 的质粒一起导入到含合适生物(如可利用正交组分的生物)的细胞内。也要导入相关合成酶(或者以多肽或编码该相关合成酶的多核苷酸形式)。将细胞在培养基内培养到所需密度，如 OD<sub>600</sub> 约 0.5 时，采用 BetaFluor™  $\beta$ -半乳糖苷酶检测试剂盒(Novagen)进行  $\beta$  半乳糖苷酶试验。抑制百分率可以计算为样品相对于可比较对照(如衍生的 lacZ 构建物所显示的活性值)的活性百分数，所述构建物在所需位置上含有有义密码子而不是选择者密码子。

本发明的 O-tRNA 的例子是 SEQ ID NO: 2。可参见本文的表 1 和实施例 2 中所显

示的典型 O-tRNA 和 O-RS 分子的序列。还可参见本文的“核酸和多肽序列及其变体”部分。该 tRNA 分子中的胸腺嘧啶(T)被尿嘧啶(U)替换。也可以有其他的碱基修饰。本发明还包括 O-tRNA 的保守变体。例如, O-tRNA 的保守变体包括的那些与 SEQ ID NO: 2 的 O-tRNA 功能类似并保留 tRNA 的 L 型结构, 但不含有相同序列(且不同于野生型 tRNA 分子)的变体。也可参见本文的“核酸和多肽序列及其变体”部分。

含 O-tRNA 的组合物还可以包含正交氨酰-tRNA 合成酶(O-RS), 其中的 O-RS 优先用氧化还原活性氨基酸氨酰化 O-tRNA。在某些实施方案中, 含有 O-tRNA 的组合物还可以包含翻译系统(如在体外或体内)。含有编码感兴趣多肽的多核苷酸的核酸, 或者这些物质的一种或多种组合也可以存在于所述细胞内, 其中所述多核苷酸含有能被 O-tRNA 识别的选择者密码子。参见本文的“正交氨酰-tRNA 合成酶”部分。

产生正交 tRNA(O-tRNA)的方法也是本发明的特征。用该方法产生的 O-tRNA 也是本发明的特征。在本发明的某些实施方案中, O-tRNA 可通过构建突变体文库来产生。突变体 tRNA 文库可用本领域已知的各种诱变技术构建。例如, 突变体 tRNA 可通过位点特异性突变、随机位点诱变、同源重组、DNA 改组技术或其他重复诱变(recursive mutagenesis)方法、嵌合体构建、或者它们的组合来产生。

也可将其他的突变引入到特定位置上, 如 tRNA 环或区域中的非保守区、保守区、随机区或者混合区, 如反密码子环、接受臂、D 臂或环、可变环、T $\Psi$ C 臂或环、tRNA 分子的其他区域或其组合。一般来说, tRNA 分子中的突变包括突变 tRNA 突变体文库内各成员的反密码子环以使其能够识别选择者密码子。该方法还可包括将一个额外的序列(CCA)加到 O-tRNA 的末端。一般来说, O-tRNA 对感兴趣生物的正交性相比起始材料(如 tRNA 序列群)有所改进, 但保留了对所需 RS 的亲合力。

本发明的方法任选包括分析 tRNAs 和/或 tRNA 合成酶序列的同源性以确定对特定生物具有正交性的潜在候选 O-tRNA、O-RS 和/或 O-tRNA/O-RS 对。可利用本领域已知的和本文所描述的计算机程序都进行这种分析, 如 BLAST 和 pileup 程序。在一个实施例中, 为了筛选潜在的正交翻译组分用于原核生物大肠杆菌, 可选择不显示与原核生物具有罕见同源性的合成酶和/或 tRNA。

一般来说, 可以通过负选择第一物种的细胞群来获得 O-tRNA, 该细胞可能含有 O-tRNA 群的成员。负选择可去除含有可被细胞内源性氨酰-tRNA 合成酶(RS)氨酰化的 O-tRNA 文库的成员细胞。这样就可以得到对第一物种细胞具有正交性的 tRNA 库。

在某些实施方案中, 进行负选择时需要将选择者密码子引入到编码负选择标记的多核苷酸中, 所述负选择标记例如有能赋予抗生素抗性的酶如  $\beta$  内酰胺酶, 能产生



可检测产物的酶如  $\beta$ 半乳糖苷酶、氯霉素乙酰转移酶(CAT)，毒性产物如芽孢杆菌 RNA 酶引入到非必须位置上(还能产生功能性的芽孢杆菌 RNA 酶)等选择。筛选还可任选通过在存在选择试剂(比如抗生素，如氨苄青霉素)的培养基中培养细胞群来完成。在一个实施方案中，选择试剂的浓度是变化的。

例如，为了测定抑制型 tRNA 的活性，根据选择者密码子如无义密码子或移码突变在体内的抑制使用一个选择系统，将选择者密码子引入到编码负性选择标记的多核苷酸(如  $\beta$ 半乳糖苷酶基因(*bla*))内。例如，构建在某个位置上含有选择者密码子的多核苷酸变体如 *bla* 变体。用这些多核苷酸转化细胞，如细菌。如果存在无法被大肠杆菌内源性合成酶不能有效加载的正交 tRNA，这种抗生素抗性如氨苄青霉素抗性应当或多或少使质粒不含转化细菌。如果该 tRNA 不是正交的，或者能加载 tRNA 的异源合成酶在此系统内共表达，应可观察到更高水平的抗生素抗性，如氨苄青霉素抗性。挑出那些无法在抗生素浓度约等于未用质粒转化的细胞的 LB 琼脂平板上生长的细胞，如细菌。

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

在一个实施方案中，对能与所需生物正交的 tRNA 库进行正选择时，将选择者密码子置于正选择标记中，例如抗药性基因(如  $\beta$ 内酰胺酶基因)编码的正选择标记中。在含有编码或包含能与细胞正交的 tRNA 库某成员的多核苷酸、编码正选择标记的多核苷酸和编码同源性 RS 的多核苷酸的细胞内进行正选择。在某些实施方案中，第二群细胞含有负选择不能去除的细胞。由于该多核苷酸在细胞内表达，使细胞可在含选择试剂(如氨苄青霉素)的培养基内生长。然后选出能被共表达的相关合成酶氨酰化以及响应其选择者密码子而掺入氨基酸的 tRNA。一般来说，这些细胞与含有非功能性 tRNA 或不能被感兴趣合成酶有效识别的 tRNA 的细胞相比，抑制效应显示升高。含有非功能性的 tRNA 或不能被感兴趣合成酶有效识别的 tRNA 的细胞对该抗生素敏感。因此那些(i)不是宿主(如大肠杆菌)内源性合成酶的底物；(ii)能被感兴趣合成酶氨酰化；以及(iii)在翻译过程中具有功能的 tRNA 可在两种筛选中保留下来。

在上面所描述的方法中，所述选择(如正选择、负选择或者正负选择)的严格程度还包括改变筛选的严格条件。例如，由于芽孢杆菌 RNA 酶是毒性很高的蛋白质，负选择条件的严格程度可通过将不同数感兴趣选择者密码子引入到该酶基因内和/或通

过使用可诱导启动子来加以调控。在另一个实施例中，选择试剂的浓度是可以改变的(如氨苄青霉素的浓度)。在本发明的一个方面，筛选条件的严格程度是随时变化的，因为在前几轮筛选中感兴趣活性较低。因此，在前几轮中筛选标准的严格程度较低，而在后几轮筛选中需要更严格的筛选标准。在某些实施方案中，负选择、正选择或者正负选择可重复多次。也可以使用多个不同的负选择标记、正选择标记或正负选择标记。在某些实施方案中，正选择标记和负选择标记可以相同。

其他类型的选择/筛选方法也可用于本发明中以产生正交性翻译组分，如 O-tRNA、O-RS 和能利用氧化还原活性氨基酸的 O-tRNA/O-RS 对。例如，负选择标记、正选择标记或正负选择标记可以是在适当反应物存在时能发荧光或催化发光反应的标记。在另一个实施方案中，标记的产物可通过荧光激活的细胞分选法(FACS)或发光检测技术检测。另外，标记还可以是亲和选择标记。参见 Francisco, J. A.等, (1993) *Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragement on the external surface*. Proc Natl Acad Sci USA. 90: 10444-8。

产生重组正交 tRNA 的其他方法见于题为“Methods and compositions for the production of orthogonal tRNA-aminoacyl tRNA synthetase pairs”(用于产生正交 tRNA-氨酰-tRNA 合成酶对的方法和组合物)的国际专利申请 WO 2002/086075 和题为“EXPANDING THE EUKARYOTIC GENETIC CODE”(扩充真核遗传密码)的 USSN 60/479,931 及 60/496,548。还见于 Forster 等, (2003) *Programming peptidomimetic synthetases by translating genetic codes designed de novo* PNAS 100(11):6353-6357; 和 Feng 等, (2003), *Expanding tRNA recognition of a tRNA synthetase by a single amino acid change*, PNAS 100(10):5676-5681。

#### 正交氨酰-tRNA 合成酶(O-RS)

本发明的 O-RS 在体外或体内能用氧化还原活性氨基酸优选氨酰化 O-tRNA。可通过含 O-RS 的多肽或编码 O-RS 或 O-RS 片段的多核苷酸将本发明的 O-RS 提供给翻译系统。例如，O-RS 包含 SEQ ID NO: 1 所列的氨基酸序列或其保守变体。在另一个实施例中，O-RS 或 O-RS 片段是由多核苷酸或其互补序列编码的，该多核苷酸编码含有 SEQ ID NO: 1 的氨基酸序列。参见本文表 1 和实施例 2 中所列的典型 O-RS 分子的序列。也可参见本文的“核酸和多肽序列及其变体”部分。

鉴定可与 O-tRNA 一起应用的正交氨酰-tRNA 合成酶(O-RS)的方法也是本发明的特征。例如，该方法包括选择，如正选择，第一物种的细胞群，该群中的个别细胞

含有：1) 一组氨酰-tRNA 合成酶(RS)的一个成员(如这组 RS 可包括突变体 RS、第一物种之外物种来源的 RS，或包括这二者)；2) 正交 tRNA(O-tRNA)(如一个或多个物种来源的)；以及 3) 编码选择标记(如正选择标记)并含有至少一个选择者密码子的多核苷酸。筛选出那些与缺乏该组 RS 成员或 RS 成员数目少的细胞相比抑制效率升高的细胞。抑制效率可用本领域已知的和本文所描述的方法测定。抑制效率升高的细胞所包含的活性 RS 可氨酰化 O-tRNA。将来自第一物种的第一组 tRNA 被活性 RS 氨酰化的水平(体外或体内)，与来自第二物种的第二组 tRNA 被活性 RS 氨酰化的水平相比较。氨酰化水平可通过检测底物(标记的氨基酸或非天然氨基酸，例如，诸如 DHP 等氧化还原活性氨基酸)来确定。通常选出氨酰化第二组 tRNA 的效率比氨酰化第一组 tRNA 更高的活性 RS，从而获得能与 O-tRNA 一起应用或更有效(优化)的正交氨酰-tRNA 合成酶。利用这种方法鉴定出的 O-RS 也是本发明的特征。

许多试验可用于确定氨酰化效率。这些方法可在体外或体内进行。例如，体外氨酰化试验的描述可见 Hoben 和 Soll(1985) *Methods Enzymol.* 113:55-59 中。氨酰化效率还可利用一个与正交翻译组分相连的报告分子，和检测能表达含有一个选择者密码子的多核苷酸编码的蛋白质的细胞在该报告分子的水平来确定。见题为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS(非天然氨基酸的体内掺入)”的 WO 2002/085923 和 2004 年 4 月 16 日提交的国际专利申请 PCT/US2004/011786。

O-RS 可以被进一步修饰以改变该合成酶的底物特异性，以使该 O-RS 只能将所需要的非天然氨基酸(如 DHP 等的氧化还原活性氨基酸)而不是其他任何常见的 20 种氨基酸加载到 O-tRNA 上。产生对非天然氨基酸具有底物特异性的正交氨酰 tRNA 合成酶的方法包括突变该合成酶，例如在该合成酶的活性位点、编辑机制位点上突变合成酶，在不同的位点上组合合成酶的不同结构域等，然后通过筛选过程进行选择。可用的一种策略是依据联用正选择然后负选择。在正选择过程中，抑制引入到正选择标记非必需位置上的选择者密码子，可使细胞在正选择压力下存活下来。在同时存在天然氨基酸和非天然氨基酸时，存活细胞编码的活性合成酶可利用天然或非天然氨基酸氨酰化正交的抑制型 tRNA。在负选择过程中，抑制引入到负选择标记非必需位置的选择者密码子，可去除具有天然氨基酸特异性的合成酶。经过负选择和正选择存活的细胞所编码的合成酶只能用非天然氨基酸来氨酰化(加载)正交的抑制型 tRNA。然后对这些合成酶进行进一步的诱变，如 DNA 改组或其他重复诱变方法。

突变体 O-RS 文库可利用本领域已知的各种诱变技术来构建。例如，可通过位点特异性突变、随机位点突变、同源重组、DNA 改组或其他重复诱变方法、嵌合构建、

或者它们的组合来制备。例如，可从两个或多个其它较小的趋异性较低的“亚文库”来构建突变体 RS 文库。本发明还包括 RS 的嵌合文库。应注意可任选构建各种生物(例如微生物如真细菌或古细菌)的 tRNA 合成酶文库，如具有天然多样性的文库(见 Short 等人的美国专利 No. 6,238, 884; Schallenberger 等人的美国专利 No. 5,756, 316; Petersen 等人的美国专利 No. 5,783, 431; Thompson 等人的美国专利 No. 5,824, 485; Short 等人的美国专利 No. 5, 958, 672)并筛选其中的正交对。

合成酶一旦通过正选择和负选择过程，可对这些合成酶进一步诱变。例如，分离编码 O-RS 的核酸；从该核酸产生一组编码突变的 O-RS(如通过随机诱变、位点特异性诱变、重组或它们的组合)的多核苷酸；这些步骤的每一步或者这些步骤的组合都可以重复进行直到获得能够优先用非天然氨基酸(如氧化还原活性氨基酸)氨酰化 O-tRNA 的突变体 O-RS。在本发明的一个方面，这些步骤可进行多次，如至少两次。

其他严格程度的选择/筛选条件也可用于本发明的方法中以产生 O-tRNA、O-RS 或 O-tRNA/O-RS 对。选择或筛选条件的严格程度在产生 O-RS 的一个步骤或两个步骤中可以不同。这包括改变所用选择/筛选试剂的浓度等。还可以进行额外轮次的正选择和/或负选择。选择或筛选还可以包括氨基酸渗透性的一种或多种改变，翻译效率的改变、翻译精确度的改变等。一般来说，所述一种或多种改变的基础是能利用正交的 tRNA-tRNA 合成酶对产生蛋白质的生物体的一种或多种基因中的突变。

产生 O-RS 以及改变该合成酶的底物特异性的其他通用细节可参见题为 *Methods and compositions for the production of orthogonal tRNA-aminoacyl-tRNA synthetase pairs*” (产生正交 tRNA-氨酰-tRNA 合成酶对的方法和组合物)的国际专利申请 WO 2002/086075 和 2004 年 4 月 16 日提交的国际专利申请 PCT/US2004/011786。

### 来源和宿主生物

本发明的翻译组分可衍生自非真核生物。例如，正交的 O-tRNA 可衍生自非真核生物(或生物组合)，例如古细菌，如詹氏甲烷球菌、热自养甲烷球菌，嗜盐菌如沃氏盐杆菌和盐杆菌 NRC-1，闪烁古生球菌，激烈火球菌，超嗜热古菌，*Aeuryopyrum pernix*，海沼甲烷球菌，坎氏甲烷嗜热菌，马氏微球菌(Mm)，超耐高温热棒菌，火球菌(*Pyrococcus abyssi*)，硫磺矿硫化叶菌(Ss)，硫化叶菌(*Sulfolobus tokodaii*)，噬酸热原体，火山热原体等，或者真细菌，如大肠杆菌、嗜热栖热菌(*Thermus thermophilus*)、嗜热脂肪芽孢杆菌(*Bacillus stearothermophilus*)等，正交的 O-RS 可衍生自非真核生物(或生物混合物)，例如古细菌，如詹氏甲烷球菌、热自养甲烷球菌，嗜盐菌属如沃氏

盐杆菌和盐杆菌 NRC-1, 闪烁古生球菌, 激烈火球菌, 超嗜热古菌, *Aeuiropyrum pernix*, 海沼甲烷球菌, 坎氏甲烷嗜热菌, 马氏微球菌, 超耐高温热棒菌, 火球菌(*Pyrococcus abyssi*), 硫磺矿硫化叶菌, 硫化叶菌(*Sulfolobus tokodaii*), 噬酸热原体, 火山热原体等, 或者真细菌, 如大肠杆菌、嗜热栖热菌、嗜热脂肪芽孢杆菌等。在一个实施方案中, 真核生物如植物、藻类、原生生物、真菌、酵母、动物(如哺乳动物、昆虫、节肢动物等)等也可作为 O-tRNA 和 O-RS 的来源。

O-tRNA/O-RS 对的各个组分可以衍生自同一生物, 也可以衍生自不同生物。在一个实施方案中, O-tRNA/O-RS 对衍生自同一生物。或者 O-tRNA/O-RS 对的 O-tRNA 和 O-RS 也可以衍生自不同生物。

可在体内或体外, 或所用的细胞(如非真核细胞或真核细胞)中选择或筛选 O-tRNA、O-RS 或 O-tRNA/O-RS 对以产生出含氧化还原活性氨基酸的多肽。非真核细胞可以是各种来源的, 如真细菌, 如大肠杆菌、嗜热栖热菌、嗜热脂肪芽孢杆菌等, 或者古细菌, 如詹氏甲烷球菌、热自养甲烷球菌, 嗜盐菌属如沃氏盐杆菌和盐杆菌 NRC-1, 闪烁古生球菌, 激烈火球菌, 超嗜热古菌, *Aeuiropyrum pernix*, 海沼甲烷球菌, 坎氏甲烷嗜热菌, 马氏微球菌, 超耐高温热棒菌, 火球菌(*Pyrococcus abyssi*), 硫磺矿硫化叶菌, 硫化叶菌(*Sulfolobus tokodaii*), 噬酸热原体, 火山热原体等。真核细胞也可以是各种来源的, 如植物(如复杂植物, 如单子叶植物或双子叶植物)、藻类、原生生物、真菌、酵母(如酿酒酵母)、动物(如哺乳动物、昆虫、节肢动物等)等。含本发明的翻译组分的细胞组合物也是本发明的特征。

筛选一个物种中的 O-tRNA 和/或 O-RS 用于另外一个物种的描述见 2004 年 4 月 16 日提交的国际专利申请 PCT/US2004/011786。

### 选择者密码子

本发明的选择者密码子扩大了蛋白质生物合成机的遗传密码框架。例如, 选择者密码子包括独特的三碱基密码子, 无义密码子如终止密码子, 例如琥珀密码子(UAG)、或乳白密码子, 非天然密码子, 至少一个四碱基密码子, 稀有密码子等。可将一些选择者密码子导入到所需基因内, 如一个、两个、三个或三个以上选择者密码子。通过采用不同的选择者密码子, 可利用多个正交 tRNA/合成酶对同时将多个氧化还原活性氨基酸(例如非天然氨基酸)同时掺入到特定位点上。

在一个实施方案中, 本发明的方法包括利用终止密码子作为选择者密码子在体内将氧化还原活性氨基酸引入到细胞内。例如, 产生可识别该终止密码子的 O-tRNA,

然后被 O-RS 用氧化还原活性氨基酸氨酰化。这种 O-tRNA 不能被天然产生的宿主氨酰-tRNA 合成酶所识别。可利用常规的定点诱变技术将终止密码子引入到编码感兴趣多肽的多核苷酸的感兴趣位点上。参见 Sayers, J. R.等, (1988), "5', 3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis." *Nucleic Acids Res*, 791-802。当 O-RS、O-tRNA 和编码感兴趣多肽的核酸在体内组合时, 氧化还原活性氨基酸因响应终止密码子而被掺入, 产生在该特定位置上含氧化还原活性氨基酸的多肽。本发明的一个实施方案中, 用作选择者密码子的终止密码子不是琥珀密码子 UAG 和/或乳白密码子 UGA。一实施例中, 将 UAG 和 UGA 都用作选择者密码子的遗传密码可编码 22 个氨基酸, 同时保留赭石无义密码子 UAA, UAA 是最丰富的终止信号。

体内掺入氧化还原活性氨基酸对宿主细胞不会产生显著影响。例如, 在非真核细胞如大肠杆菌内, 由于 UAG 密码子的抑制效率依赖于 O-tRNA(如琥珀抑制型 tRNA)和释放因子 1(RF1)(RF1 可结合 UAG 密码子, 而引起正在延伸的肽链从核糖体上释放下来)之间的竞争, 因此可通过提高 O-tRNA(如抑制型 tRNA)的表达水平, 或利用 RF1 缺陷株来调节抑制效率。在真核细胞内, 由于 UAG 密码子的抑制效率依赖于 O-tRNA(如琥珀抑制型 tRNA)和真核释放因子(如 eRF)(eRF 可结合终止密码子, 引起正在延伸的肽链从核糖体上释放下来)之间的竞争, 因此可通过提高 O-tRNA(如抑制型 tRNA)的表达水平来调节抑制的效率。另外, 也可存在其他化合物, 例如还原剂二硫苏糖醇(DTT)。

氧化还原活性氨基酸也可以由稀有密码子编码。例如, 当体外蛋白合成反应中的精氨酸浓度下降时, 稀有的精氨酸密码子 AGG 被证实可通过丙氨酸酰化的合成 tRNA 有效地将丙氨酸掺入到多肽链内。见 Ma 等, *Biochemistry*, 32:7939(1993)。在这种情况下, 合成的 tRNA 可与天然的 tRNA<sub>Arg</sub> 竞争, 后者作为稀有成分存在于大肠杆菌内。另外, 某些生物不能利用所有的三联密码子。藤黄微球菌(*Micrococcus luteus*)内的独立密码子 AGA 可用于在体外将氨基酸插入到转录/翻译提取物内。见 Kowal 和 Oliver, *Nucl. Acid.Res.*, 25:4685(1997)。体内利用这些稀有密码子可以产生出本发明的各组分。

选择者密码子还包括扩充的密码子, 例如四碱基或多碱基密码子, 如四碱基、五碱基、六碱基或多碱基密码子。四碱基密码子的例子包括 AGGA、CUAG、UAGA、CCCU 等。五碱基密码子的例子包括 AGGAC、CCCCU、CCCUC、CUAGA、CUACU、UAGGC 等。本发明的方法包括利用基于移码抑制的扩充密码子。四碱基或多碱基密

码子可用于将一个或多个非天然氨基酸如氧化还原活性氨基酸掺入到同一个蛋白质内。在其他的实施方案中,反密码子环可以解码至少一个四碱基密码子、至少一个五碱基密码子、至少一个六碱基密码子等。由于有 256 个可能的四碱基密码子,因此,利用四碱基或多碱基密码子可在同一个细胞内编码多个非天然氨基酸。见 Anderson 等, (2002) *Exploring the Limits of Codon and Anticodon Size*, Chemistry and Biology, 9:237-244; 以及 Magliery, (2001) *Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in Escherichia coli*, J. Mol. Biol. 307:755-769。

例如,用体外生物合成方法利用四碱基密码子将非天然氨基酸掺入到蛋白质内。见 Ma 等, (1993) Biochemistry, 32:7939; 和 Hohsaka 等, (1999) J. Am. Chem. Soc., 121:34。利用 CGGG 和 AGGU 与两个用化学方法酰化的移码抑制型 tRNA, 将赖氨酸的 2-萘基丙氨酸和 NBD 衍生物同时掺入到链亲和素内。见 Hohsaka 等, (1999) J. Am. Chem. Soc., 121:12194。在一个体内研究中, Moore 等人检查了 tRNA<sup>Leu</sup> 衍生物和 NCUA 反密码子能否抑制 UAGN 密码子(N 可以是 U、A、G 或 C), 结果发现四联密码子 UAGA 可被含 UCUA 反密码子的 tRNA<sup>Leu</sup> 解码, 其效率为 13~26%, 在 0 或-1 框架处几乎不解码。见 Moore 等, (2000) J. Mol. Biol. 298:195。在一个实施方案中, 基于稀有密码子或无义密码子的扩充密码子可用于本发明中, 这些密码子可降低在其他不需要位点发生错义通读和移码抑制。

在一个给定的系统中, 选择者密码子还包括一个天然的三碱基密码子, 该内源系统不能利用(或很少利用)这种天然的三碱基密码子。例如, 包括缺乏能识别天然三碱基密码子的 tRNA 的系统 and/或三碱基密码子是稀有密码子的系统。

选择者密码子任选包括非天然碱基对。这些非天然碱基对进一步扩充了现有的遗传字母表。一种额外的碱基对可将三联密码子的数目从 64 个增加到 125 个。三碱基对的特性包括稳定的选择性碱基配对、可有效地被聚合酶高度精确地掺入到 DNA 内、以及在新非天然碱基对合成后继续有效地引物延伸。适合本发明方法和组合物的非天然碱基对的描述见 Hirao 等, (2002) *An unnatural base pair for incorporating amino acid analogues into protein*, Nature Biotechnology, 20:177-182 和 Wu, Y.等, (2002) J. Am. Chem. Soc. 124:14626-14630。其他相关文献将在下文列出。

非天然核苷在用于体内时可以穿透生物膜, 被磷酸化后形成相应的三磷酸盐形式。另外, 所增加的遗传信息是稳定的, 不含被细胞的酶破坏。Benner 及其同事在先前曾尝试利用与经典的 Watson-Crick 碱基对不同的氢键模式, 最值得注意的是

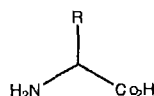
iso-C:iso-G 对。见 Switzer 等, (1989) *J. Am. Chem. Soc.*, 111:8322 和 Piccirilli 等, (1990) *Nature*, 343:33; Kool, (2000) *Curr. Opin. Chem. Biol.*, 4:602。一般来说, 这些碱基会与天然碱基发生某种程度的错配, 并且不能用酶学方法复制。Kool 及其同事的研究证明碱基之间的疏水性包裹相互作用可代替氢键, 驱使碱基对的形成。见 Kool, (2000) *Curr. Opin. Chem. Biol.*, 4:602, Guckian 和 Kool, (1998) *Angew. Chem. Int. Ed. Engl.*, 36, 2825。为尝试开发出能满足上述所有要求的非天然碱基对, Schultz、Romesberg 及其同事系统地合成和研究了一系列的非天然疏水碱基。发现 PICS:PICS 自身对比天然碱基对更稳定, 可被大肠杆菌 DNA 聚合酶 I 的 Klenow 片段(KF)有效掺入到 DNA 内。见 McMinn 等, (1999) *J. Am. Chem. Soc.*, 121:11586 和 Ogawa 等, (2000) *J. Am. Chem. Soc.*, 122:3274。3MN:3MN 自身配对可由 KF 合成, 其活性和选择性足以使其行使生物学功能。见 Ogawa 等, (2000) *J. Am. Chem. Soc.* 122:8803。但是, 这两个碱基的作用是进一步复制的链末端。最近已开发了一种可用于复制 PICS 自身对的突变型 DNA 聚合酶。另外, 7AI 自身对也可被复制。见 Tae 等, (2001) *J. Am. Chem. Soc.*, 123:7439。一种新型的金属碱基对 Dipic:Py 也被开发出来, 该碱基对在结合 Cu(II)时可形成稳定的碱基对。见 Meggers 等, (2000) *J. Am. Chem. Soc.*, 122:1071。由于扩充的密码子和非天然密码子本来就是与天然密码子正交的, 因此本发明的方法可利用这一特性来产生这些密码子的正交 tRNA。

一种翻译旁路系统也可用于将氧化还原活性氨基酸掺入到所需多肽内。在翻译旁路系统中, 可将一个大序列插入到基因内不能翻译成蛋白质。该序列含有一种可作为提示信号的结构, 诱导核糖体跳过该序列恢复掺入序列的下游翻译。

### 非天然氨基酸

如本文所述, 非天然氨基酸指除硒氨酸和/或吡咯赖氨酸和下列 20 种遗传密码编码的  $\alpha$  氨基酸之外的所有氨基酸、修饰氨基酸或氨基酸类似物:  $\alpha$  氨基酸是丙氨酸、精氨酸、天冬酰胺、天冬氨酸、半胱氨酸、谷氨酰胺、谷氨酸、甘氨酸、组氨酸、异亮氨酸、亮氨酸、赖氨酸、蛋氨酸、苯丙氨酸、脯氨酸、丝氨酸、苏氨酸、色氨酸、酪氨酸、缬氨酸。 $\alpha$  氨基酸的通用结构可用结构式 I 表示:

I



非天然氨基酸一般都具有结构式 I 的结构, 其中 R 基团代表 20 种天然氨基酸中所用基团之外的任何取代基。见 L. Stryer 编的《生物化学》(*Biochemistry*)(第三版,



1988, Freeman and Company, New York)以了解 20 种天然氨基酸的结构。注意, 本发明的非天然氨基酸可以是上述 20 种  $\alpha$  氨基酸之外的天然化合物。

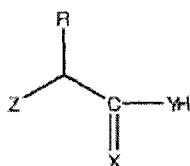
一般来说, 因为本发明的非天然氨基酸在侧链上与天然氨基酸不同, 因此非天然氨基酸可以与其他氨基酸(如天然或天然的)形成酰胺键, 形成的方式与天然蛋白中的酰胺键一样。但是, 非天然氨基酸含有与天然氨基酸可区别的侧链基团。

在将非天然氨基酸掺入到蛋白质中时特别令人感兴趣的是能否将氧化还原活性氨基酸(例如含有可将分子中的电子和/或质子传递入或传递出结构部分的非天然氨基酸)掺入到蛋白质中。例如, 在一类氧化还原活性氨基酸中, 式 I 中的 R 包括但不限于: 酮基-、叠氨基-、羟基-、卤代-(例如碘代-)、硝基-、硫醇基-、硒基-、磺酰基-、杂环、醛(aldehyde)、硫代酸等, 或它们的任何组合。本发明的氧化还原活性氨基酸的例子包括但不限于: 3,4-二羟基-L-苯丙氨酸(DHP)、3,4,6-三羟基-L-苯丙氨酸、3,4,5-三羟基-L-苯丙氨酸、3-硝基-酪氨酸、4-硝基-苯丙氨酸、3-硫醇-酪氨酸等。也可参见图 1。

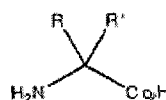
在其他的非天然氨基酸中, 式 I 中的 R 基团可以是烷基-、芳基-、酰基-、肼、氰基-、卤代-、酰肼、烯基、炔基、醚、硼酸基、磷、膦酰基、膦、烯酮、亚胺、酯、羟胺、胺等, 或上述基团的组合。我们感兴趣的其他非天然氨基酸包括但不限于含光活化交链接头的氨基酸、自旋-标记氨基酸、荧光标记氨基酸、金属结合氨基酸、含金属的氨基酸、放射性氨基酸、含有新官能团的氨基酸、能与其他分子共价或非共价结合的氨基酸、光陷阱(photocaged)和/或光促异构(photoisomerizable)氨基酸、含生物素或生物素类似物的氨基酸、含酮基氨基酸、糖基化氨基酸、侧链连有糖部分的氨基酸、含聚乙二醇或聚醚的氨基酸、重原子取代的氨基酸、化学方法可裂解或光解的氨基酸、与天然氨基酸相比侧链延长的氨基酸(如超过约 5 个、约 10 个碳原子的聚醚或长链烃)、含碳连接糖的氨基酸、含氨基硫代酸的氨基酸、以及含一个或多个毒性结构部分的氨基酸。

非天然氨基酸除了可含新型侧链以外, 还任选可含有经修饰的骨架结构, 如结构式 II 和 III 所示的结构:

II



III



式中, Z 一般为 OH、NH<sub>2</sub>、SH、NH-R'或 S-R'; X 和 Y 可以相同或不同, 一般包括 S 或 O, R 和 R'可相同也可不同, 一般选自与上述结构式 I 所示的非天然氨基酸的 R 基团相同的组成物以及氢。例如, 本发明的非天然氨基酸还可以在结构式 II 和 III 所示的氨基或羧基上含有取代基。这种类型的非天然氨基酸包括但不限于带 20 个常见氨基酸相应的侧链或非天然侧链的  $\alpha$ -羟酸、 $\alpha$ -硫代酸、 $\alpha$ -氨基硫代羧酸。另外,  $\alpha$  碳原子上的取代基还包括 L,D 或  $\alpha$ - $\alpha$ -双取代的氨基酸, 如 D-谷氨酸、D-丙氨酸、D-甲基-O-酪氨酸、氨基丁酸等。其他的结构形式包括环形氨基酸, 如脯氨酸类似物以及 3、4、6、7、8 和 9 元环脯氨酸类似物,  $\beta$  和  $\gamma$  氨基酸, 如取代的  $\beta$  丙氨酸和  $\gamma$ -氨基丁酸。

例如, 许多非天然氨基酸衍生自天然氨基酸, 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、正位取代的酪氨酸和间位取代的酪氨酸, 这些酪氨酸取代基团包括: 乙酰基、苯甲酰基、氨基、肼、羟氨基、巯基、羧基、异丙基、甲基、C<sub>6</sub>-C<sub>20</sub> 直链或支链烷基、饱和或不饱和烷基、O-甲基、聚醚基等。另外, 也可考虑多个取代的芳环。本发明的谷氨酰胺类似物包括但不限于:  $\alpha$ -羟基衍生物、 $\gamma$ -取代衍生物、环状衍生物和酰胺基取代的谷氨酰胺衍生物。苯丙氨酸类似物的例子包括但不限于: 对位取代的苯丙氨酸、正位取代的苯丙氨酸和间位取代的苯丙氨酸, 这些取代基包括: 羟基、甲氧基、甲基、丙稀基、醛基、硝基、硫醇基或酮基等。非天然氨基酸的具体例子包括但不限于: 3,4-二羟基-L-苯丙氨酸(DHP)、3,4,6-三羟基-L-苯丙氨酸、3,4,5-三羟基-L-苯丙氨酸、4-硝基-苯丙氨酸、*p*-乙酰基-L-苯丙氨酸、*p*-炔丙氧基苯丙氨酸、O-甲基-L-酪氨酸、L-3-(2-萘基)丙氨酸、3-甲基-苯丙氨酸、O-4-丙稀基-L-酪氨酸、4-丙基-L-酪氨酸、3-硝基-酪氨酸、3-硫醇-酪氨酸、三-O-乙酰基-GlcNAc  $\beta$ -丝氨酸、L-多巴、含氟苯丙氨酸、异丙基-L-苯丙氨酸、*p*-叠氨基-L-苯丙氨酸、*p*-酰基-L-苯丙氨酸、*p*-苯甲酰基-L-苯丙氨酸、L-磷丝氨酸、磷酰基丝氨酸、磷酰基酪氨酸、*p*-碘代-苯丙氨酸、*p*-溴代苯丙氨酸、*p*-氨基-L-苯丙氨酸以及异丙基-L-苯丙氨酸等。各种非天然氨基酸的结构显示在本文的图 1 和题为“非天然氨基酸的体内掺入”的 WO 2002/085923 的图 16、17、18、19、26 和 29 中。

### 非天然氨基酸的化学合成

上述许多非天然氨基酸都可以从 Sigma(USA)或 Aldrich(Milwaukee, WI, USA) 等公司购买到。那些不能从市场上购买到的非天然氨基酸可任选按照各种发表的或本领域技术人员已知的方法合成。见 Fessendon 和 Fessendon 的《有机化学》(Organic Chemistry) (1982, 第二版, Willard Grant Press, Boston Mass.); March 的《高级有机

化学》(Advanced Organic Chemistry)(第三版, 1985, Wiley and Sons, New York); 以及 Carey 和 Sundberg 的《高级有机化学》(Advanced Organic Chemistry)(第三版, A 和 B 部分, 1990, Plenum Press, New York)。其他描述非天然氨基酸合成的文献包括: 题为“非天然氨基酸的体内掺入”的国际专利申请 WO 2002/085923; Matsoukas 等, (1995) J. Med. Chem., 38, 4660-4669; King, F. E. 和 Kidd, D. A. A. (1949) *A New Synthesis of Glutamine and of  $\gamma$ -Dipeptides of Glutamic Acid from Phthylated Intermediates*. J. Chem. Soc., 3315-3319; Friedman, O. M. 和 Chatterji, R. (1959) *Synthesis of Derivatives of Glutamine as Model Substrates for Anti-tumor Agents*. J. Am. Chem. Soc. 81, 3750-3752; Craig, J. C. 等 (1988) *Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine)*. J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmont, M. 和 Frappier, F. (1991) *Glutamine analogues as Potential Antimalarials*, Eur. J. Med. Chem. 26, 201-5; Koskinen, A. M. P. 和 Rapoport, H. (1989) *Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues*. J. Org. Chem. 54, 1859-1866; Christie, B. D. 和 Rapoport, H. (1985) *Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization*. J. Org. Chem. 1989:1859-1866; Barton 等, (1987) *Synthesis of Novel  $\alpha$ -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- $\alpha$ -Amino-Adipic Acids, L- $\alpha$ -aminopimelic Acid and Appropriate Unsaturated Derivatives*. Tetrahedron Lett. 43:4297-4308; 以及 Subasinghe 等, (1992) *Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site*. J. Med. Chem. 35:4602-7。还可参见 2003 年 12 月 22 日提交的题为“Protein Arrays(蛋白质阵列)”的国际专利申请 PCT/US03/41346。

### 非天然氨基酸的细胞摄取

一般认为在设计和选择非天然氨基酸, 如用于掺入到蛋白质中的非天然氨基酸时要考虑的一个问题是细胞对非天然氨基酸的摄取。例如, 带高密度电荷  $\alpha$  氨基酸的化合物不可能透过细胞膜。天然氨基酸通过蛋白质转运系统的收集摄入到细胞内, 但是这些转运系统一般都有不同程度的氨基酸特异性。可进行一种快速筛选来评价哪种非天然氨基酸可被细胞摄取。见下列文献中的毒性试验: 2003 年 12 月 22 日提交的题为“Protein Arrays(蛋白质阵列)”的国际专利申请 PCT/US03/41346; 以及 Liu

和 Schultz (1999) *Progress toward the evolution of an organism with an expanded genetic code*. PNAS 96:4780-4785。虽然用各种方法不难分析摄取情况，但是，设计能利用细胞摄取通路的非天然氨基酸的另外一种方法是在体内提供生物合成通路所产生的氨基酸。

### 非天然氨基酸的生物合成

细胞内已经存在许多生物合成通路可产生氨基酸和其他化合物。虽然某种具体非天然氨基酸的生物合成方法在自然界中如在细胞内可能不存在，但本发明提供了这样的方法。例如，可以通过加入新的酶或者修饰现有的宿主细胞通路来产生非天然氨基酸的生物合成通路。加入的新酶可以是天然产生的酶，也可以是人工改造的酶。例如，*p*-氨基苯丙氨酸(如上述 WO 2002/085923 实施例中提到的)的生物合成依赖于加入其他生物的已知酶的组合物。可以通过用含这些酶的基因的质粒转化细胞而将这些酶的基因导入到细胞内。这些基因在细胞内表达时即提供了酶催化通路来合成所需要的化合物。可任选加入的这类酶的例子见下面实施例的描述。加入的酶的序列可在 Genbank 中找到。也可以同样方式将人工修饰的酶加入到细胞内。以这种方式操纵细胞内的合成机制和物质来产生非天然氨基酸。

许多方法确实都可用于在体外或体内产生生物合成通路所用的新型酶、改造现有的通路来产生非天然氨基酸。许多修饰酶的方法和生物合成通路的其他组分都可用于本发明来产生非天然氨基酸(或者修饰酶使其具有新的底物特异性或其他感兴趣的活性)。例如，DNA 改组可用于产生新的酶和/或这种酶的通路以在体外或体内产生非天然氨基酸(或产生新的合成酶)。见 Stemmer(1994), *Rapid evolution of a protein in vitro by DNA shuffling*, Nature 370(4):389-391; 以及 Stemmer, (1994), *DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution*, Proc. Natl. Acad. Sci. USA., 91:10747-10751。一种改组相关基因(如同源的基因)家族的方法可快速地修改酶使其具有所需特性。这种“家族基因改组”方法的一个例子见 Cramer 等(1998) *DNA shuffling of a family of genes from diverse species accelerates directed evolution* Nature, 391(6664):288-291 中的描述。新型酶(不论生物合成通路的组分或合成酶)也可以通过 DNA 重组方法产生，这种方法被称为“产生杂交酶的渐进式截断技术”(ITCHY)，如 Ostermeier 等, (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" Nature Biotech 17:1205 中所描述的。这种方法也可用于产生酶或其他通路变体的文库，这些酶或通路变体可作为一种或多种体外

或体内重组方法的底物。见 Ostermeier 等(1999) "Combinatorial Protein Engineering by Incremental Truncation," Proc.Natl.Acad.Sci.USA, 96:3562-67, 和 Ostermeier 等(1999), "Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts," Biological and Medicinal Chemistry, 7:2139-44。另外一种方法采用指数整体诱变来产生酶或其他通路变体的文库, 选出文库中能够催化生物合成反应产生非天然氨基酸(或新的合成酶)的酶和通路变体。在这种方法中, 感兴趣序列中的小簇残基被用于随机地平行鉴定位置发生改变从而导致产生功能性蛋白的每一个氨基酸。能用于本发明来产生可生产非天然氨基酸的新酶(或新的合成酶)的这种方法的例子见于 Delegrave 和 Youvan(1993) Biotechnology Research 11:1548-1552 的描述。在另一种方法中, 利用掺入的(doped)或退化的寡核苷酸进行随机或半随机诱变来改造酶和/或通路组分, 例如采用 Arkin 和 Youvan(1992) "Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis" Biotechnology 10:297-300 或 Reidhaar-Olson 等, (1991) "Random mutagenesis of protein sequences using oligonucleotide cassettes" Methods Enzymol. 208:564-86 所描述的通用诱变方法。通常称为“非随机”诱变的另一种利用多核苷酸重新组装和位点饱和诱变方法可用来产生酶和/或通路组分, 然后筛选能够自行一种或多种合成酶或生物合成通路的能(如在体内产生非天然氨基酸)的那些酶或组分。见 Short 的题为“基因疫苗和酶的非随机产生”的专利申请 WO 00/46344。

这种突变方法的一种替代方法包括重组生物整个基因组并选择产生的具有特定通路功能的子代(通常称为“全基因组改组”)。这种方法也可用于本发明, 如通过基因组重组和选择能够产生出非天然氨基酸(或其中间体)的生物(如大肠杆菌或其他细胞)。例如, 下列文献中所描述的方法可用来设计通路以修饰细胞内现有的和/或新的通路, 在体内产生非天然氨基酸: Patnaik 等(2002) "Genome shuffling of lactobacillus for improved acid tolerance" Nature Biotechnology, 20(7):707-712; 以及 Zhang 等(2002) "Genome shuffling leads to rapid phenotypic improvement in bacteria" Nature, 02-7, 415(6872):644-646。

现在已有一些其他技术可用来改造生物和代谢通路来产生所需化合物, 这些方法也可用于产生非天然氨基酸。描述这种通路改造方法的文献包括: Nakamura 和 White(2003) "Metabolic engineering for the microbial production of 1,3 propanediol" Curr. Opin. Biotechnol. 14(5):454-9; Berry 等(2002) "Application of Metabolic Engineering to improve both the production and use of Biotech Indigo" J. Industrial Microbiology and

Biotechnology 28:127-133; Banta 等(2002) "Optimizing an artificial metabolic pathway: Engineering the cofactor specificity of *Corynebacterium* 2,5-diketo-D-gluconic acid reductase for use in vitamin C biosynthesis" Biochemistry, 41(20), 6226-36; Selivonova 等, (2001) "Rapid Evolution of Novel Traits in Microorganisms" Applied and Environmental Microbiology, 67:3645 等等。

一般来说, 无论用什么方法, 用本发明的经改造的生物合成通路所产生出的非天然氨基酸的浓度足以有效的生物合成蛋白质, 例如可达到细胞内的天然含量, 但该浓度对细胞内的其他氨基酸的浓度不会产生显著影响, 或者耗竭掉细胞内的资源。以这种方法在体内产生的浓度一般约为 10 mM 到 0.05 mM。一旦细胞经过改造产生了可用于特异性通路和非天然氨基酸制备的酶后, 任选进行体内选择以进一步优化用于核糖体蛋白合成和细胞生长的非天然氨基酸的产量。

#### 用于掺入 3,4-二羟基-L-苯丙氨酸(DHP)的正交组分

本发明提供了在体内产生能将氧化还原活性氨基酸, 例如 3,4-二羟基-L-苯丙氨酸(DHP)掺入到正在合成的多肽链中以响应选择者密码子, 如终止密码子、无义密码子、四碱基或多碱基密码子等的正交组分的产生方法和组合物。例如, 本发明提供了正交-tRNA(O-tRNA)、正交氨酰-tRNA 合成酶(O-RS)及两者组成的对。这些正交对可用于将 DHP 掺入到正在延伸的多肽链中。

本发明的组合物包括正交氨酰-tRNA 合成酶(O-RS), 其中的 O-RS 优先用 DHP 氨酰化 O-tRNA。在某些实施方案中, O-RS 含有 SEQ ID NO:1 所示的氨基酸序列或其保守变体。在本发明的某些实施方案中, O-RS 优先用氧化还原活性氨基酸氨酰化 O-tRNA, 其效率至少是含氨基酸序列 SEQ ID NO:1 多肽的 50%。

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

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

包含翻译系统的细胞(如非真核细胞或真核细胞)也是本发明的内容, 其中的翻译系统包括正交-tRNA(O-tRNA); 正交氨酰-tRNA 合成酶(O-RS)和氧化还原活性氨基酸, 例如 3,4-二羟基-L-苯丙氨酸(DHP)。一般来说, O-RS 优先氨酰化 O-tRNA, 其效率至少是含氨基酸序列 SEQ ID NO.:1 多肽的 50%。O-tRNA 可识别第一选择者密码子, O-RS 优先用 3,4-二羟基-L-苯丙氨酸(DHP)氨酰化 O-tRNA。在一个实施方案中, O-tRNA 含有 SEQ ID NO.:2 所示的序列或其互补多核苷酸序列, 或者被该序列编码。在一个实施方案中, O-RS 含有任何一个 SEQ ID NO.:1 或其保守变体所示的氨基酸序列。

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

在某些实施方案中, 本发明的细胞包括大肠杆菌细胞, 细胞内含有正交-tRNA(O-tRNA)、正交氨酰-tRNA 合成酶(O-RS)、氧化还原活性氨基酸和含有感兴趣多肽编码多核苷酸的核酸, 其中的多核苷酸含有能被 O-tRNA 识别的选择者密码子。

在本发明的某些实施方案中, 本发明的 O-tRNA 含有本文列表中和实施例中所示的多核苷酸序列或其互补的多核苷酸序列, 或者被这些序列编码。在本发明的某些实施方案中, O-RS 含有序列表中所示的氨基酸序列或其保守变体。在一个实施方案中, O-RS 或其片段被编码本文列表中和实施例中所示的氨基酸序列的多核苷酸序列或其互补的多核苷酸序列编码。

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

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

含本发明的多核苷酸的载体也是本发明的特征。例如, 本发明的载体可以是质粒、

粘粒、噬菌体、病毒、表达载体等。含本发明的载体的细胞也是本发明的特征。

O-tRNA/O-RS 对的组分的产生方法也是本发明的特征。利用这些方法产生出的组分也是本发明的特征。例如，至少产生一个与细胞正交的 tRNA(O-tRNA)的方法包括产生突变的 tRNA 文库；突变 tRNA 突变文库内每一个成员的反密码子环使其能识别选择者密码子，从而形成一个 O-tRNA 文库，负选择第一物种的第一细胞群，其中的细胞含有 O-tRNA 文库的成员。负选择可去除含 O-tRNA 文库的成员的细胞，其中的 O-tRNA 可被细胞内源性的氨酰-tRNA 合成酶(RS)氨酰化。这样就形成一个与第一物种的细胞正交的 tRNA 库，从而至少可以获得一个 O-tRNA。利用本发明的方法产生出的 O-tRNA 也包括在本发明之内。

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

本发明还提供了能用氧化还原活性氨基酸加载 O-tRNA 的正交-氨酰-tRNA 合成酶的鉴定方法。例如，该方法包括筛选第一物种的细胞群，其中的细胞都含有 1) 一组氨酰-tRNA 合成酶(RS)的成员(如一组 RS 可包括突变的 RS、第一物种之外的物种来源的 RS、或者二者都包括)；2) 正交-tRNA(O-tRNA)(如衍生自一个或多个物种)；以及 3) 编码正选择标记并且至少包含一个选择者密码子的多核苷酸。

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

在细胞(例如非真核细胞如大肠杆菌细胞等，或者真核细胞)内产生特定位置上为 3,4-二羟基-L-苯丙氨酸(DHP)的蛋白质的方法也是本发明的特征。例如，一种方法是在适当的培养基内培养细胞，其中的细胞含有核酸，该核酸至少含有一个选择者密码子，编码一个蛋白，加入 DHP，在至少含有一个选择者密码子的核酸的翻译过程中 DHP 被掺入到蛋白质的特定位置上，生成蛋白质。细胞还含有：在细胞内有功能并且能识别选择者密码子的正交-tRNA(O-tRNA)；以及优先用 DHP 氨酰化 O-tRNA 的正交氨酰-tRNA 合成酶(O-RS)。利用这种方法产生的蛋白质也是本发明的特征。

本发明还提供了包含蛋白质的组合物，其中的蛋白质含有 DHP。在某些实施方案中，蛋白质的氨基酸序列与已知蛋白如治疗蛋白、诊断蛋白、工业用酶的氨基酸序



列或其部分至少 75% 相同性。另外，组合物还可以包含药学上可接受的载体。

### 核酸和多肽序列及其变体

如上文和下文所描述，本发明提供了编码例如 O-tRNA 和 O-RS 的核酸的多核苷酸序列和多肽的氨基酸序列，例如 O-RS，以及包含所述序列的组合物、系统和方法。所述序列的例子，如 O-tRNA 和 O-RS 的氨基酸和核苷酸序列也在本文中列出(见表 1，例如 SEQ ID NO: 1-3)。但是，本领域的技术人员应该了解本发明并不仅仅局限于这些序列，如实施例和列表中的序列。本领域的技术人员应当了解本发明还包括具有本文所描述的功能如编码合适 O-tRNA 或 O-RS 的许多相关和不相关序列。

本发明提供了多肽(O-RS)和多核苷酸，如 O-tRNA、编码 O-RS 或其片段的多核苷酸，用于分离氨酰-tRNA 合成酶克隆的寡核苷酸等。本发明的多核苷酸包括那些编码本发明的感兴趣蛋白或多肽的多核苷酸，其中的蛋白或多肽含有一个或多个选择者密码子。另外，本发明的多核苷酸还包括含有 SEQ ID NO: 2 中所列核苷酸序列的多核苷酸；与多核苷酸序列互补的多核苷酸或者编码多核苷酸序列的多核苷酸。本发明的多核苷酸还可以编码包含 SEQ ID NO: 1 的氨基酸序列。本发明的多核苷酸还包括编码本发明的多肽的多核苷酸。同样，在高严格条件下，可与上述多核苷酸在全长核酸序列上杂交的人工核酸也包括在本发明的多核苷酸范围内。在一个实施方案中，组合物包含本发明的多肽和各种赋形剂(如缓冲液、水、药学上可接受的赋形剂等)。本发明还提供了可与本发明的多肽发生特异性免疫反应的抗体或抗血清。人工多核苷酸是由人产生的，不是天然发生的。

本发明的多核苷酸还包括人工的多核苷酸，这种多核苷酸与天然 tRNA(但是天然 tRNA 除外)至少有 75%、80%、90%、95%、98% 或更高的相同性。多核苷酸还包括与天然 tRNA 的多核苷酸至少有 75%、80%、90%、95%、98% 或更高相同性的人工多核苷酸。

在某些实施方案中，载体(如质粒、粘粒、噬菌体、病毒等)含有本发明的多核苷酸。在一个实施方案中，载体是表达载体。在另一个实施方案中，表达载体包含与本发明的一个或多个多核苷酸功能相连的启动子。在另一个实施方案中，细胞包含携带本发明的多核苷酸的载体。

本领域的技术人员还应该理解的是本发明还包括公开序列的多种变体。例如，可形成功能相同序列的公开序列的保守变体也包含在本发明的范围内。核酸多核苷酸序列的变体也被认为包括在本发明的范围之内，其中的变体可与至少一个公开序列杂交。本文公开的序列的独特序列，例如通过标准序列比较技术确定的独特序列，

也包括在本发明的范围之内。

保守变体

由于遗传密码的简并性，“静默替代”(即核酸序列中不会导致其编码多肽改变的替代)是每个编码氨基酸的核酸所具有的特性。同样，氨基酸序列中的一个或几个氨基酸发生“保守氨基酸替代”是指被具有高度相似性能的不同氨基酸的替代，也很容易鉴定为与公开的构建物高度相似。每一个公开序列的这种保守变体也是本发明的特征。

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

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

| 非极性和/或脂肪族侧链                             | 极性不带电荷的侧链                                 | 芳香族侧链              | 带正电荷的侧链           | 带负电荷的侧链     |
|-----------------------------------------|-------------------------------------------|--------------------|-------------------|-------------|
| 甘氨酸<br>丙氨酸<br>缬氨酸<br>亮氨酸<br>异亮氨酸<br>脯氨酸 | 丝氨酸<br>苏氨酸<br>半胱氨酸<br>蛋氨酸<br>天冬酰胺<br>谷氨酰胺 | 苯丙氨酸<br>酪氨酸<br>色氨酸 | 赖氨酸<br>精氨酸<br>组氨酸 | 天冬氨酸<br>谷氨酸 |

核酸杂交

对比杂交可用于鉴定本发明的核酸，如 SEQ ID NO: 2，包括本发明的核酸的保守变体，这种对比杂交方法是一种区分本发明核酸的优选方法。另外，能在高、超高、

超超高严格条件下与 SEQ ID NO: 2 所代表的那些核酸杂交的靶核酸也是本发明的特征。这类核酸的典型例子包括那些与一个给定核酸序列相比发生一个或几个静默核酸替代或保守核酸替代的核酸。

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

当核酸相遇时发生杂交,一般是在溶液中。核酸之所以可以发生杂交是由各种已知的物理-化学作用力造成的,如氢键、溶剂排斥、碱基堆积等。较全面的有关核酸杂交的指导手册是 Tijssen(1993)的《生物化学和分子生物学实验技术—与核酸探针杂交》(Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes)中第 I 部分第 2 章,“杂交原理和核酸探针分析策略概述(Overview of principles of hybridization and the strategy of nucleic acid probe assays)”,(Elsevier, New York),以及上述 Ausubel 的文献。Hames 和 Higgins(1995)编辑的 Gene Probes 1(基因探针 1),牛津大学出版社 IRL 出版社,英国牛津(Hames 和 Higgins 1)以及 Hames 和 Higgins(1995)编辑的 Gene Probes 2(基因探针 2),牛津大学出版社 IRL 出版社,英国牛津(Hames 和 Higgins 2)详细描述了 DNA 和 RNA,其中包括寡核苷酸的合成、标记、检测和定量方法。

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

“严格的杂交洗涤条件”用于核酸杂交试验如 DNA 印迹和 RNA 印迹时是指序列依赖性的,但是在不同的环境参数下该条件是不同的。较全面的有关核酸杂交的指导手册是上述的 Tijssen(1993)和 Hames 和 Higgins, 1 和 2。任何被检测核酸所用的杂交和洗涤条件的严格程度可以根据经验很容易地确定。例如,在确定杂交和洗涤条件的严格程度时,可以逐渐提高杂交和洗涤条件的严格程度(如,在杂交或洗涤过程中通过升高温度、降低盐浓度、增加洗涤剂的浓度和/或增加有机溶剂如甲醛的

浓度),直到满足了所选择的一组标准。例如,在高严格杂交和洗涤条件下,逐渐增加杂交和洗涤条件直到探针与互补靶核酸完全匹配,所产生的信噪比至少是探针与不匹配靶核酸杂交所产生的信噪比的5倍。

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

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

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

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

### 独特的亚序列

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

同样,本发明还包括含独特亚序列的多肽,这些独特亚序列位于选自本文所描述的多肽的O-RS序列的多肽内。在这里,与任何已知的多肽序列相应的多肽相比,独特亚序列都是独特的。

本发明还提供了能在严格条件下与独特编码寡核苷酸杂交的靶核酸,该寡核苷酸编码选自O-RS序列的多肽内的独特亚序列,其中的独特亚序列与任何对照多肽(如本发明的合成酶来源的亲本序列,如通过突变获得)相应的多肽相比都是独特的。独

特序列可通过上述方法确定。

### 序列比较、相同性和同源性

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

术语“基本相同的”用于两个或多个核酸或多肽(如编码 O-tRNA 或 O-RS 的 DNA, O-RS 的氨基酸序列)的比较时是指两个或多个序列或亚序列排列对比并比较其最大相似性时，这两个或多个序列或亚序列至少有约 60%、80%、90-95%、98%、99%或更多的核苷酸或氨基酸残基是相同的，这种比较可利用序列比较算法或通过肉眼观察来完成。无论实际衍生自何种家系，这种“基本相同的”序列一般被认为是“同源的”。“基本上相同”指在序列的一个区域内的比较，这个区域长度宜至少约为 50 个残基，更优选至少约 100 个残基，最优选二序列至少约含 150 个残基基本上相同，或者在两个序列的全长上进行比较。

当蛋白和/或蛋白序列衍生自共同的亲代蛋白或蛋白序列时，不论是天然的还是人造的，这些蛋白和/或蛋白序列是“同源的”。同样，当核酸和/或核酸序列衍生自共同的亲代核酸或核酸序列时，不论是天然的还是人造的，这些核酸和/或核酸序列是“同源的”。例如，任何天然核酸都可用现有的诱变方法修饰使其包含一个或多个选择者密码子。当该突变的核酸表达时，其编码的多肽含有一个或多个非天然氨基酸，如氧化还原活性氨基酸。当然，该突变过程可能改变一个或多个标准密码子，从而使产生的突变蛋白中发生一个或多个标准氨基酸的改变。同源性一般是根据两个或多个核酸或蛋白质(或其序列)的序列相同性推算出来的。用于确定序列之间同源性的精确百分相同性根据所要比较的蛋白质和核酸而不同，但是一般来说至少要有 25%的序列相同性才能建立同源性。更高的序列类似性，如 30%、40%、50%、60%、70%、80%、90%、95%、99%或更高，也可用于建立同源性。确定序列相同性百分比的方法(如 BLASTP 和 BLASTN，使用默认参数)见本文的描述，这些方法都是现有的。

为了进行序列比较并确定其同源性，一般需要把一个序列看作参考序列，被检测序列与之比较。在使用序列比较算法时，将被测序列与参考序列输入到计算机内，如果需要，设定亚序列匹配值和序列算法程序的参数。序列比较算法就可以根据设定的程序参数计算出被测序列相对于参考序列的序列相同性百分比。

进行序列比较的理想序列排列对比方法可用下列算法完成：局部同源性算法，Smith & Waterman, Adv. Appl. Math. 2:482(1981)；同源性排列对比算法，Needleman & Wunsch, J. Mol. Biol. 48:443(1970)；相似性检索方法，Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444(1988)，以及这些算法的计算机程序(Wisconsin Genetics Software Package 内的 GAP、BESTFIT、FASTA 和 TFASTA, Genetics Computer Group, 575 Science Dr., Madison, WI)，或者通过肉眼检查(见下面 Ausubel 等的文献)。

适于确定序列相同性和序列相似性百分比的算法的一个例子是 BLAST 算法，该算法的描述见 Altschul 等, J. Mol. Biol. 215:403-410(1990)。执行 BLAST 分析的软件公众可通过国家生物技术信息中心的网站([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/))下载。这种算法包括先通过鉴定查询序列中长度为 W 的短字符来确定高评分序列对(HSP)，所述短字符在与序列数据库中相同长度的字符排列对比时可匹配或满足某些正阈值评分 T。T 称为相邻字符评分阈值(Altschul 等，同上)。这些初始的相邻字符采样数(hits)作为种子启动检索来寻找含有这些字符的 HSP。然后字符采样数在每一序列的两个方向上扩展，直到累积排列对比评分升高。用参数 M(匹配残基对的奖赏分数；总是大于 0)和 N(错配残基的惩罚分数；总是小于 0)来计算核苷酸序列的累积评分。对于氨基酸来说，可利用计分矩阵来计算累积评分。字符采样数在每个方向上的延伸停止于下列时刻：累积排列对比评分从其曾达到的最大数值开始回落时；由于累积了一个或多个负评分残基排列对比使累积评分达到 0 或 0 以下时；或者达到序列的末端时。BLAST 算法的参数 W、T 和 X 决定了算法的灵敏度和速度。BLASTN 程序(用于核苷酸序列)的默认参数为：字长(W)=11，期望值(E)=10，终止值=100，M=5，N=-4，两条链进行比较。对于氨基酸序列来说，BLASTP 程序的默认参数为：字长(W)=3，期望值(E)=10，使用 BLOSUM62 计分矩阵(见 Henikoff 和 Henikoff(1989) Proc.Natl.Acad.Sci.USA 89: 10915)。

除了能够计算序列相同性百分比以外，BLAST 算法还可以进行两个序列之间相似性的统计分析(见 Karlin 和 Altschul, Proc.Nat'l.Acad.Sci.USA 90:5873-5787(1993))。BLAST 算法所提供的一个相似性数据是最小总概率(P(N))，这一数据说明了两个核苷酸或氨基酸序列之间相匹配的可能性。例如，如果将被测核酸与参考核酸比较后得到的最小总概率约在 0.1 以下、优选约 0.01 以下、最优选的是约 0.001 以下，则这个核酸被认为与参考序列是相似的。

#### 诱变和其他分子生物学技术

本发明的和用于本发明的多核苷酸和多肽可以用分子生物学技术进行操纵。描述

分子生物学技术的综合性教科书包括: Berger 和 Kimmel 的《分子克隆技术指南》(Guide to Molecular Cloning Techniques)第 152 卷“酶学方法”, Academic Press, Inc., San Diego, CA(Berger); Sambrook 等人的《分子克隆实验手册》(Molecular Cloning—A Laboratory Manual)(第三版)1—3 卷, 冷泉港实验室, 纽约冷泉港, 2001("Sambrook")和《现代分子生物学操作指南》(Current Protocols in Molecular biology), F. M. Ausubel 等人编辑,《现代操作指南》(Current Protocols), Greene Publishing Associates, Inc. 和 John Wiley & Sons, Inc., 共同投资出版, (2003 年增补)("Ausubel")。这些教科书描述了诱变、载体的使用、启动子以及许多其他相关主题, 如与基因产生有关的主题, 包括可用于制备含有氧化还原活性氨基酸(例如 DHP)的蛋白质、正交 tRNA、正交合成酶和它们组成的正交对的选择者密码子。

各种突变技术都可用于本发明中进行 tRNA 分子的突变、tRNA 文库的产生、合成酶文库的产生、编码氧化还原活性氨基酸的选择者密码子向感兴趣蛋白或多肽内的掺入。这些技术包括但不限于位点定向诱变、随机位点诱变、同源重组、DNA 重组或其他的重复诱变方法、嵌合体构建、利用含尿嘧啶的模板进行诱变、寡核苷酸定向突变、硫代磷酸修饰的 DNA 突变、利用带缺口的双链 DNA 进行的突变等。其他合适的方法包括点错配修复、利用修复缺陷宿主系进行的突变、限制性筛选和限制性纯化、缺失突变、通过总基因合成的诱变、双链断裂修复等。利用嵌合体构建进行的诱变也包括在本发明中。在一个实施方案中, 可以利用已知的有关天然分子或者改变的或突变的天然分子的信息指导诱变, 如序列、序列比较、物理特性、晶体结构等。

宿主细胞可用本发明的多核苷酸或包含本发明多核苷酸的构建物进行遗传改造(如转化、转导或转染), 如本发明的载体, 这些载体可以是克隆载体, 也可以是表达载体。例如, 正交 tRNA、正交 tRNA 合成酶及其衍生的蛋白质的编码区可以被可控地连接到能在感兴趣宿主细胞内发挥功能的基因表达调节元件上。典型的载体都包含转录和翻译终止子、转录和翻译起始序列和用于调节特定靶核酸表达的启动子。载体还可以包含基因表达盒, 其中至少含有一个独立的终止子序列, 允许表达盒在真核或原核宿主、或者二者内(如穿梭载体)复制的序列以及用于原核和真核系统内的选择标记。载体可以在原核或真核宿主内表达和/或整合, 最好是既可在真核宿主内又可在原核宿主内表达和/或整合。见 Gilman 和 Smith, Gene 8:81(1979); Roberts 等, Nature, 328:731(1987); Schneider, B.等, Protein Expr. Purif. 6435:10(1995); Ausubel, Sambrook, Berger(这三篇文献同上)。载体可以是质粒、细菌、病毒、裸多核苷酸或

共轭多核苷酸的形式。可用标准方法将载体导入到细胞和/或微生物内,如电穿孔(From 等, *Proc. Natl. Acad. Sci. USA* 82, 5824(1985));用病毒载体感染;用含核酸的小粒子高压渗透,其中核酸位于小珠或离子的基质内或表面上(Klein 等, *Nature* 327,70-73(1987))。

ATCC 提供了可用于克隆的细菌和噬菌体的目录,如 ATCC 出版的《细菌和噬菌体的 ATCC 目录》(*The ATCC Catalogue of Bacteria and Bacteriophage*)(1996), Gherna 等(编)。用于测序、克隆和分子生物学其他方面的其他基础方法及其理论阐述见于 Sambrook(同上)、Ausubel(同上)和 Watson 等(1992) *科学美国人丛书第二版重组 DNA(Recombinant DNA Second Edition Scientific American Books)*, NY。另外,几乎所有核酸(以及几乎所有的标记核酸,无论是标准的还是非标准的)都可以从各个公司里购买或订购,如 Midland Certified Reagent Company(Midland, TX [mcr.com](http://mcr.com))、The Great American Gene Company(Ramona, CA, 网址是 [www.genco.com](http://www.genco.com))、ExpressGen Inc.(Chicago, IL, 网址是 [www.expressgen.com](http://www.expressgen.com))、Operon Technologies Inc.(Alameda, CA) 等。

经遗传改造的宿主细胞可用经过适当改变以便于进行下列操作的常规营养培养基进行培养:筛选、活化启动子或筛选转化子。这些细胞还可以培养成转基因生物。有关细胞分离和培养(如用于亚序列核酸分离)的其他有用文献包括 Freshney(1994)《动物细胞培养—基本操作技术》(*Culture of Animal Cells, a Manual of Basic Technique*), 第三版, Wiley-Liss, New York 及其引用的参考文献;Payne 等(1992)《在液体系统中进行植物细胞培养和组织培养》(*Plant Cell and Tissue Culture in Liquid Systems*), John Wiley & Sons, Inc., New York, NY;Gamborg 和 Phillips(编)(1995)《植物细胞、组织和器官培养》(*Plant Cell, Tissue and Organ Culture*), Fundamental Methods Springer Lab Manual, Springer-Verlag(Berlin Heidelberg New York)以及 Atlas 和 Parks(编)《微生物学培养基手册》(*The Handbook of Microbiological Media*)(1993), CRC Press, Boca Raton, FL。

### 感兴趣的蛋白和多肽

氧化还原活性氨基酸的一个优点是它们可被用来改造蛋白质中的电子传递过程。其它优点包括但不限于,氧化还原活性蛋白质的表达有助于研究和能够改变蛋白质电子传递途径、改变酶的催化功能、使蛋白质与小分子和生物分子交联等。含有至少一种氧化还原活性氨基酸的感兴趣蛋白或多肽是本发明的一个特征。本发明还包括含有至少用本发明的组合物和方法制造的氧化还原活性氨基酸的多肽或蛋白质。



赋形剂(如药学上可接受的赋形剂)也可以和蛋白质一起存在。另外,本发明的蛋白还可以进行翻译后修饰。

在细胞内产生特定位置上为氧化还原活性氨基酸的蛋白质的方法也是本发明的特征。例如,一种方法是在适宜的培养基内培养细胞,其中的细胞含有至少包含一个选择者密码子并能编码一个蛋白质的核酸;提供氧化还原活性氨基酸;其中的细胞还包含:在细胞内有功能并能识别选择者密码子的正交-tRNA(O-tRNA)和优先用氧化还原活性氨基酸氨酰化 O-tRNA 的正交氨酰-tRNA 合成酶(O-RS)。在某些实施方案中,与包含或编码于 SEQ ID NO: 2 所列多核苷酸序列的 O-tRNA 相比,在相关合成酶存在的条件下,这种 O-tRNA 为了响应选择者密码子而产生的抑制效率至少可达 45%、50%、60%、75%、80%、90%或更高。利用这种方法产生的蛋白也是本法发明的特征。

本发明还提供了含有蛋白的组合物,其中的蛋白包含氧化还原活性氨基酸。在某些实施方案中,蛋白所包含的氨基酸序列与治疗蛋白、诊断蛋白、工业用酶或其片段的序列至少有 75%的相同性。

本发明的组合物和利用本发明的方法产生出的组合物也可以存在于细胞内。因此本发明的 O-tRNA/O-RS 对其单个组分可被用于宿主系统的翻译机器中,从而将氧化还原活性氨基酸掺入到蛋白质内。国际专利申请 PCT/US2004/011786(申请日为 2004 年 4 月 16 日,名称为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS(扩充真核遗传密码”)和 WO 2002/085923(题为“非天然氨基酸的体内掺入”)描述了这一过程,本文已纳入作为参考。例如,当 O-tRNA/O-RS 对被引入到宿主如大肠杆菌细胞内后,O-tRNA/O-RS 对在体内响应选择者密码子后可将氧化还原活性氨基酸如 DHP,例如一种合成的氨基酸,掺入到蛋白质内,如酪氨酸或苯丙氨酸衍生物,这是作为外源性物质加入到培养基内的。另外,本发明的组合物既可以用于体内翻译系统,也可以用于体外翻译系统。

利用本发明的细胞可以大规模地合成含非天然氨基酸的蛋白质。一方面,组合物可以包含至少 10 $\mu$ g、50 $\mu$ g、75 $\mu$ g、100 $\mu$ g、200 $\mu$ g、250 $\mu$ g、500 $\mu$ g、1mg、10mg 或更多含氧化还原活性氨基酸的蛋白质,或者含有蛋白质的量能够达到体内蛋白合成方法的要求(本文提供了重组蛋白产生和纯化方法的详细描述)。另一方面,组合物中所含的蛋白质浓度可以是每升细胞裂解物、缓冲液、药学上可接受的缓冲液或其他液体悬液中至少含 10 $\mu$ g、50 $\mu$ g、75 $\mu$ g、100 $\mu$ g、200 $\mu$ g、250 $\mu$ g、500 $\mu$ g、1mg、10mg 或更高(如以 1nL 到 100L 的体积)。在细胞内大量产生(例如大于其他方法,如体外翻

译,在一般情况下所能达到的规模)至少含至少一个氧化还原活性氨基酸的蛋白质也是本发明的特征。

掺入氧化还原活性氨基酸可导致蛋白质结构和/或功能的变化,如大小、酸性、亲核性、氢键、疏水性、蛋白酶靶位点的易接近性、对基序的靶向性(如用于蛋白芯片时)等。含有氧化还原活性氨基酸的蛋白质的催化活性或物理特性可能会得以改善,或者获得全新的催化活性或物理特性。例如,下列特性可通过在蛋白质内掺入氧化还原活性氨基酸加以改变:毒性、生物分布、结构特性、分光光度特性、化学和/或光化学特性、催化能力、半衰期(如血清半衰期)、与其他分子相互作用的能力(共价的或非共价的)。包含至少携带一个氧化还原活性氨基酸的蛋白质的组合物可用作新型的治疗蛋白、诊断蛋白、催化酶、工业用酶、结合蛋白(如抗体)以及可用于蛋白质结构和功能的研究。见 Dougherty, (2000) *Unnatural Amino Acids as Probes of Protein Structure and Function*, Current Opinion in Chemical Biology, 4:645-652。

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

一般来说,包含氧化还原活性氨基酸(及任何相应的编码核酸,如包含一个或多个选择者密码子的核酸)都可用本发明的组合物和方法产生。无需将已知的成千上万个蛋白质都一一鉴定,其中的任何一个都可以被修饰成含一个或多个非天然氨基酸的形式,如通过任何已知的突变方法将一个或多个适宜的选择者密码子掺入到相关的翻译系统内。已知蛋白序列的常用数据库包括 GenBank EMBL、DDBJ 和 NCBI。其他数据库通过搜索互联网也可以很容易地找到。

一般来说,本发明的蛋白质与已知的蛋白(如治疗蛋白、诊断蛋白、工业用酶,或其片段等)至少有 60%、70%、75%、80%、90%、95%、99%或更高的相同性,其

中包含一个或多个非天然氨基酸。可被修饰使其包含一个或多个氧化还原活性氨基酸的治疗蛋白、诊断蛋白以及其他蛋白可在下列文献中找到,但是又不仅仅局限于这些文献:2004年4月16日提交的题为“Expanding the Eukaryotic Genetic Code(扩充真核遗传密码)”的国际专利申请 PCT/US2004/011786;以及 WO 2002/085923,题为“非天然氨基酸的体内掺入”。可被修饰使其包含一个或多个氧化还原活性氨基酸的治疗蛋白、诊断蛋白以及其他蛋白包括但不限于  $\alpha$ -1 抗胰蛋白酶、血管他丁、抗溶血因子、抗体(有关抗体的详细描述见下文)、载脂蛋白、脱辅基蛋白、心房利钠因子、心房利钠多肽、心房肽、C-X-C 趋化因子(如 T39765、NAP-2、ENA-78、Gro-a、Gro-b、Gro-c、IP-10、GCP-2、NAP-4、SDF-1、PF4、MIG)、降钙素、CC 趋化因子(如单核细胞趋化蛋白-1、单核细胞趋化蛋白-2、单核细胞趋化蛋白-3、单核细胞炎症蛋白-1  $\alpha$ 、单核细胞炎症蛋白-1  $\beta$ 、RANTES、I309、R83915、R91733、HCC1、T58847、D31065、T64262)、CD40 配基、C-kit 配基、胶原蛋白、集落刺激因子(CSF)、补体因子 5a、补体抑制因子、补体受体 1、细胞因子(如上皮细胞中性粒细胞活化肽-78、GRO $\alpha$ /MGSA、GRO $\beta$ 、GRO $\gamma$ 、MIP-1 $\alpha$ 、MIP-1 $\delta$ 、MCP-1)表皮生长因子(EGF)、红细胞生成素(EPO)、剥脱毒素 A 和 B、因子 IX、因子 VII、因子 VIII、因子 X、成纤维细胞生长因子(FGF)、纤维蛋白原、纤连蛋白、G-CSF、GM-CSF、葡萄糖脑苷脂酶、绒毛膜促性腺激素、生长因子、Hedgehog 蛋白(如 Sonic、Indian、Desert)、血红素、肝细胞生长因子(HGF)、水蛭素、人血清白蛋白、胰岛素、胰岛素样生长因子(IGF)、干扰素(如 IFN- $\alpha$ 、IFN- $\beta$ 、IFN- $\gamma$ )、白细胞介素(如 IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12 等)、角质化细胞生长因子(KGF)、乳铁蛋白、白血病抑制因子、荧光素酶、Neurturin、中性粒细胞抑制因子(NIF)、抑瘤素 M、骨生成蛋白、甲状旁腺激素、PD-ECSF、PDGF、肽类激素(如人生长激素、Pleiotropin、蛋白 A、蛋白 G、热原性外毒素 A、B 和 C、松弛素、肾素、SCF、可溶性补体受体 1、可溶性 I-CAM1、可溶性白细胞介素受体(IL-1、2、3、4、5、6、7、9、10、11、12、13、14、15)、可溶性 TNF 受体、生长调节素、生长抑素、生长激素、链激酶、超抗原即葡萄球菌肠毒素(SEA、SEB、SEC1、SEC2、SEC3、SED、SEE)、超氧化物歧化酶(SOD)、中毒性休克综合征毒素 TSST-1)、胸腺素  $\alpha$ 1、组织纤溶酶原激活因子、肿瘤坏死因子  $\beta$ (TNF $\beta$ )、肿瘤坏死因子受体(TNFR)、肿瘤坏死因子- $\alpha$ (TNF $\alpha$ )、血管内皮生长因子(VEGEF)、尿激酶等。

可以利用本发明的组合物和方法在体内产生出的含本文所描述的氧化还原活性氨基酸的一类蛋白质是转录调节因子或其片段。转录调节因子的例子包括能调节细

胞生长、分化、调控等的基因和转录调节蛋白。转录调节因子存在于原核细胞、病毒和真核生物如真菌、植物、酵母、昆虫和动物如哺乳动物，提供了一个大范围的治疗靶位。应当理解的是表达和转录活化因子调节转录的机制很多，如与受体结合、刺激信号转导级联、调节转录因子的表达、与启动子和增强子结合、与启动子和增强子结合的蛋白结合、解链 DNA、剪接 mRNA 前体、多聚腺苷酸化 RNA 和降解 RNA。

本发明的一类蛋白质(如携带一个或多个氧化还原活性氨基酸的蛋白质)是表达活化因子，包括细胞因子、炎症分子、生长因子及其受体、以及癌基因产物，如白细胞介素(如 IL-1、IL-2、IL-8 等)、干扰素、FGF、IGF-I、IGF-II、FGF、PDGF、TNF、TGF- $\alpha$ 、TGF- $\beta$ 、EGF、KGF、SCF/c-Kit、CD40L/CD40、VLA-4/VCAM-1、ICAM-1/LFA-1 和 hyalurin/CD44；信号转导分子及其相应的癌基因产物，如 Mos、Ras、Raf 和 Met；以及转录活化因子和抑制因子，如 p53、Tat、Fos、Myc、Jun、Myb、Rel，以及类固醇激素受体，如雌激素、孕酮、睾丸酮、醛固酮、LDL 受体的配基和皮质酮。

本发明还提供了至少携带一个氧化还原活性氨基酸的酶(如工业用酶)或其片段。酶的例子包括但不限于酰胺酶、氨基酸消旋酶、酰基转移酶、脱氢酶、二氧合酶、二芳基丙烷过氧化物酶、表位酶、环氧化物水解酶、酯酶、异构酶、激酶、葡萄糖异构酶、糖苷酶、葡萄糖酰基转移酶、卤代过氧化物酶(haloperoxidase)、单加氧酶(如 p450)、脂肪酶、木素过氧化物酶、胍水合酶、胍水解酶、蛋白酶、磷酸酯酶、枯草杆菌蛋白酶、转氨酶和核酶。

这些蛋白中有许多是可以从公司购买到的(见 Sigma BioSciences 的 2002 年产品目录和价目表)，其相应的蛋白序列和基因及其变体通常都是已知的(见 Genbank)。它们都可以按照本发明的方法通过掺入一个或多个氧化还原活性氨基酸或其他非天然氨基酸来加以修饰，例如改变蛋白质的一种或多种治疗、诊断或酶催化特性。与治疗相关的特性包括血清半衰期、保存半衰期、稳定性、免疫原性、治疗活性、可检测性(如通过非天然氨基酸如氧化还原活性氨基酸上掺入报告基团(如标记物或标记结合位点))、LD<sub>50</sub> 或其他负效应的降低、通过胃肠道进入体内的能力(如口服利用度)等。诊断特性包括保存半衰期、稳定性、诊断活性、可检测性等。相关的酶催化特性包括保存半衰期、稳定性、酶活性、生产能力等。

其他的多种蛋白也可加以修饰使其包含一个或多个本发明的氧化还原活性氨基酸或其他非天然氨基酸。例如，本发明包含的蛋白质可以用氧化还原活性氨基酸替代一个或多个天然氨基酸的疫苗蛋白，这些蛋白来自于感染性真菌，如曲霉属、念珠菌属的真菌；细菌，特别是大肠杆菌，可以作为致病菌的模型，和医学上有重

要影响的细菌如葡萄球菌属(如金黄色葡萄糖球菌)或链球菌属(如肺炎链球菌)的细菌;原虫如孢子虫(如刚地弓形虫)、根足虫(如内阿米巴原虫)和鞭毛虫(锥形虫、利什曼原虫、毛滴虫、鞭毛虫等);病毒如(+) RNA 病毒(例子包括痘病毒类如痘苗病毒;细小核糖核酸病毒如脊髓灰质炎病毒;披膜病毒如风疹病毒;黄病毒如 HCV;以及冠状病毒), (-) RNA 病毒(如弹状病毒如 VSV;副黏液病毒如 RSV;正粘病毒如流感病毒;布亚病毒;以及沙粒病毒), dsDNA 病毒(如呼肠孤病毒), RNA 到 DNA 的病毒,即逆转录病毒如 HIV 和 HTLV,以及某些 DNA 到 RNA 的病毒如乙肝病毒。

与农业有关的蛋白质如昆虫抗性蛋白(如 Cry 蛋白)、淀粉和脂质生成酶、植物和昆虫毒素、毒素抗性蛋白、霉菌毒素解毒蛋白、植物生长酶(如核酮糖 1,5-二磷酸羧化酶/氧合酶, "RUBISCO")、脂肪加氧酶(LOX)、以及磷酸烯醇丙酮酸(PEP)羧化酶也是氧化还原活性氨基酸修饰的适宜靶标。

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

感兴趣蛋白或多肽的编码基因可用本领域技术人员熟知的方法以及本文“突变和其他分子生物学技术”部分描述的方法加以突变使其携带用于掺入氧化还原活性氨基酸的一个或多个选择者密码子。例如,感兴趣蛋白的编码核酸可被突变使其携带一个或多个选择者密码子,因而可以掺入一个或多个氧化还原活性氨基酸。本发明包括任何蛋白的这种突变版本,比如至少含有一个氧化还原活性氨基酸。同样,本发明还包括相应的核酸,即含有一个或多个编码一个或多个氧化还原活性氨基酸的选择者密码子的任何核酸。

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

#### 根据免疫活性定义多肽

由于本发明的多肽提供了多种新的多肽序列(如在本文的翻译系统中合成的蛋白中含有氧化还原活性氨基酸或者在新型合成酶、标准氨基酸的新型序列中含有氧化还原活性氨基酸),因此多肽可能会提供能被免疫分析方法识别的新型结构特征。能

特异性结合本发明多肽的抗血清的产生以及能被这些抗血清结合的多肽也是本发明的特征。本文所用的术语“抗体”包括但不限于免疫球蛋白基因或其片段编码的多肽，这些多肽可特异性识别并结合分析物(抗原)。例子包括多克隆抗体、单克隆抗体、嵌合抗体和单链抗体等。免疫球蛋白的片段，其中包括 Fab 片段和表达文库如噬菌体展示文库产生的片段，也包括在本文所用的术语“抗体”之中。见 Paul 编辑的《基础免疫学》(Fundamental Immunology)，第四版，1999，Raven Press，New York，中关于抗体结构和方法学的描述。

为了产生用于免疫分析中的抗血清，可按照本文所描述的方法产生和纯化一种或多种免疫原性多肽。例如，用重组细胞产生重组蛋白。纯系小鼠(用于此分析中，因为纯系小鼠具有遗传相同性，因此得到的结果重复性更高)用含标准佐剂如福氏完全佐剂的免疫原性蛋白按照标准的免疫方法免疫(见 Harlow 和 Lane(1988)，《抗体—实验室操作手册》(Antibodies, A Laboratory Manual)(冷泉港出版社，纽约)中关于抗体产生、用于确定特异性免疫反应的免疫分析方法和条件的描述)。关于蛋白、抗体、抗血清等的其他详细描述见 USSN 60/479,931、60/463,869 和 60/496,548，题为“扩充真核遗传密码”(Expanding the Eukaryotic Genetic Code)；WO 2002/085923，题为“非天然氨基酸的体内掺入”；专利申请 USSN 60/441, 450，名称为“糖蛋白合成”，申请日为 2003 年 1 月 16 日；以及专利申请“蛋白质芯片”，代理人备案号为 P1001US00，申请日为 2002 年 12 月 22 日。

#### O-tRNA 和 O-RS 以及 O-tRNA/O-RS 对的应用

本发明的组合物以及用本发明的方法产生的组合物存在于细胞内。因而本发明的 O-tRNA/O-RS 对或其各个组分可用于宿主系统的翻译机器中，结果导致氧化还原活性氨基酸被掺入到蛋白质内。相应的 Schultz 等人的专利申请“非天然氨基酸的体内掺入”(WO2002/085923)描述了这一过程，本文已纳入作为参考。例如，当 O-tRNA/O-RS 对被导入到宿主如大肠杆菌内时，O-tRNA/O-RS 对就可在体内将氧化还原活性氨基酸掺入到蛋白质如肌红蛋白或治疗蛋白内以响应选择者密码子如琥珀无义密码子，其中的氧化还原活性氨基酸可外源性地加入到培养基中。另外，本发明的组合物既可用于体外翻译系统中，也可用于体内系统中。含氧化还原活性氨基酸的蛋白质可用作治疗蛋白，并可用来在蛋白质内改变酶的催化功能和/或电子传递途径、将蛋白质与小分子和/或生物分子交联，并且有助于研究蛋白质的结构、与其他蛋白质的相互作用、蛋白质内的电子传递过程等。

#### 试剂盒

试剂盒也是本发明的特征。例如，本发明提供了用于在细胞内产生至少含一个氧化还原活性氨基酸的蛋白质的试剂盒，其中的试剂盒包括一个盛有 O-tRNA 和/或 O-tRNA 编码多核苷酸序列、和/或 O-RS 编码多核苷酸序列、和/或 O-RS 的容器。在一个实施方案中，试剂盒还包括氧化还原活性氨基酸。在另一个实施方案中，试剂盒还包括关于产生蛋白质的说明书。

## 实施例

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

### 实施例 1:将氧化还原活性氨基酸位点特异性掺入蛋白质中

最近的报道显示，可用正交 tRNA-氨酰 tRNA 合成酶对将许多非天然氨基酸被选择性地掺入到大肠杆菌和酵母的蛋白质中(Wang 等，(2001) Science 292:498-500; Zhang 等，(2003) Biochemistry 42:6735-6746; Chin 等，(2003) Science 301:964-967)。这些正交对不与宿主细胞翻译机制的内源组分发生交叉反应，但能识别所需的非天然氨基酸并响应琥珀无义密码子 TAG 将其掺入蛋白质中(Wang 等，(2000) J. Am. Chem. Soc., 122:5010-5011; Wang 和 Schultz(2001) Chem. Biol., 8:883-890)。为在大肠杆菌内遗传编码 3,4-二羟基-L-苯丙氨酸(DHP; 见图 1 的化合物 1)，正交詹氏甲烷球菌 tRNA-合成酶(MjTyrRS; 提供在图 5 和表 1 中，其氨基酸序列提供于 SEQ ID NO:4，核苷酸序列提供于 SEQ ID NO:5)的特异性被改变，以使该合成酶用 DHP 而非任何 20 种常见氨基酸来氨酰化突变型酪氨酸 tRNA 琥珀抑制基因(*mutRNA<sup>Tyr</sup><sub>CUA</sub>*)。这些突变型合成酶选自两种突变型 MjTyrRS 文库(Wang 等，(2001) Science 292:498-500; Zhang 等，(2002) Angew. Chem. Int. Ed., 41:2840-2842)。在基于嗜热脂肪芽孢杆菌的同源 TyrRS 晶体结构的分析的第一文库中(Brick 等，(1989) J. Mol. Biol., 208:83-98)，位于酪氨酸芳环对位 6.5Å 内的 MjTyrRS 活性位点中的 5 个残基(Tyr32、Glu107、Asp158、Ile159 和 Leu162)被随机突变(在质粒 pBK-lib 上编码)。在第二文库中，使位于酪氨酸芳环间位 6.5Å 内的 6 个残基(Tyr32、Ala67、His70、Gln155、Asp158、Ala167)随机突变(在质粒 pBK-lib-m 上编码)。

为改变 TyrRS 的特异性，以使能特异性掺入 DHP 而非其它天然氨基酸，采用由若干轮正选择和负选择构成的遗传选择。在正选择中，根据对氯霉素乙酰转移酶(CAT)

基因(pRep(2)/YC)中非必需位置(Asp112)的琥珀密码子的抑制,选出两个突变型 TyrRS 文库。将用突变型 TyrRS 文库、*mutRNA<sub>CUA</sub><sup>Tyr</sup>* 基因和琥珀突变型 CAT 基因转化的细胞在厌氧条件下培育在含有 1 mM DHP 和 70 $\mu$ g/ml 氯霉素的基础培养基中以避免氧化 DHP。存活的细胞含有用 DHP 或内源氨基酸氨酰化 *mutRNA<sub>CUA</sub><sup>Tyr</sup>* 的突变型 TyrRS。然后根据对引入在毒性芽孢杆菌 RNA 酶基因(pLWJ17B3)中非必需位置(Gln2、Asp44、Gly55)的三个琥珀密码子的抑制,用负选择除去含有天然氨基酸的突变型 TyrRS。将含有得自先前正选择的突变型 TyrRS、*mutRNA<sub>CUA</sub><sup>Tyr</sup>* 和琥珀突变型芽孢杆菌 RNA 酶基因的细胞培养在不含 DHP 的 Luria-Bertani(LB)培养基中。在这些条件下,编码对内源氨基酸具有特异性的突变型 TyrRS 的细胞将产生全长的芽孢杆菌 RNA 酶并死亡。只有那些含有对 DHP 具有特异性的突变型 TyrRS 的细胞可存活。交替进行三轮正选择和两轮负选择后,得到能在高浓度氯霉素(90 mg/L)下存活并依赖于 DHP 存在、含有所选突变型 TyrRS 基因(DHPRS)、*mutRNA<sub>CUA</sub><sup>Tyr</sup>* 和 Asp112TAG CAT 基因的克隆。然而,在缺乏 DHP 时,同样的细胞仅在 20 mg/L 氯霉素中存活。这一结果说明,所选 DHPRS 酶对 DHP 的特异性高于天然氨基酸。测序显示所选 DHPRS 中存在以下突变:Tyr32 $\rightarrow$ Leu、Ala67 $\rightarrow$ Ser、His70 $\rightarrow$ Asn、Ala167 $\rightarrow$ Gln。所述 DHPRS 合成酶示于图 6 和表 1。同时,其氨基酸序列提供于 SEQ ID NO:1,核苷酸序列提供于 SEQ ID NO:3。

为测定用所选 pDHPRS 克隆掺入 DHP 的保真性和效率,我们响应琥珀密码子在 C 末端用六组氨酸标记的突变型抹香鲸肌红蛋白(Mb; 见 Chin 等, (2002) Proc.Natl.Acad.Sci.U.S.A., 99:11020-11024)暴露于表面的第 4 个残基处掺入 DHP。含有 DHP 的全长肌红蛋白(DHPMb)用 GMML(含有亮氨酸的甘油基础培养基)作为生长培养基在还原条件(100 $\mu$ M 二硫苏糖醇(DTT))下表达以防止 DHP 在掺入蛋白质之前被氧化。突变蛋白质的产率约为 1 毫克/升(相同条件下野生型 Mb(wtMb)的产率不可测)。非全长 Mb 在不存在 DHP 时表达;在不存在 DTT 时大多数细胞由于氧化的醌(见图 1 的化合物 3)的毒性而死亡。全长 DHPMb 用钴基 IMAC(固定的金属亲和层析)树脂纯化。将存在和不存在 DHP 时表达的突变蛋白质的纯化产物加样到 SDS-PAGE 凝胶上,银染并对凝胶进行 western 印迹。采用抗 His6-tag 抗体,非全长 Mb 在不存在 DHPRS 或 *mutRNA<sub>CUA</sub><sup>Tyr</sup>* 时表达(示于图 2A)。用带有四极-四极飞行时间(QqTOF)的电喷雾电离(ESI)质谱仪来测量蛋白质的分子量。图 2B 显示了质量为 18,448.5 道尔顿的 DHPMb 的 ESI-QqTOF 质谱。含有 Mb 的 DHP 该值在 70ppm 之内,计算得到的质量为 18447.2 道尔顿(相邻峰的质量为 18,432.3 道尔顿,根据对照试验用测量技术得到,由于失去氧或失去氧和质子所致)。



用循环伏安法来确定当将裸金电极浸入含有 DHPMb 的溶液时是否可观察到氧化的氢醌的氧化还原波。图 3A 显示在厌氧条件下含有 wtMb 和 DHPMb 的溶液不可逆的伏安应答(Bard 和 Faulkner, 见《电化学方法》(*Electrochemical Methods*); John W. Wiley & Sons, Inc.:New York, 1980; 第 213-248、429-487 和 675-698 页)。在  $E = -320$  mV 时观察到 wtMbFe(III)产生的还原峰电位, 而该突变蛋白的还原峰电位迁移到更低的负电位  $E = -400$  mV。这种迁移是由于存在 DHP, DHP 在比缺乏 DHP 时低得多的电位上促进了 Fe(III)的还原。观察到的不可逆的伏安数是由于电子传递速率慢, 这可能是由于电子接近电极有限所造成的。图 3B 显示了含有  $100\mu\text{M}$  DEP、wtMb 和 DHPMb 的溶液的伏安反应。DHP 氧化产生的电流只在存在突变的 Mb 时或在不含 DHP 的溶液中可见, 分别为  $E = 580$  mV 和  $E = 385$  mV。这些结果清楚显示, Mb 中存在 DHP 对 Fe(III)-血红素基团的氧化还原电位有显著影响, 反之亦然。

这里的描述证明, 氧化还原活性氨基酸(例如 DHP)可被有效和选择性地掺入到大肠杆菌之类生物体的蛋白质中。这些氨基酸可在蛋白质中通过电化学方法氧化。这种将氧化还原活性氨基酸位点特异性掺入蛋白质的能力有助于研究蛋白质中的电子传递, 并能基因改造产生具有新特性的氧化还原蛋白。将 DHP 等氧化还原活性氨基酸位点特异性掺入到模型蛋白如 Mb 和其它蛋白质的各个位点可用来研究这种蛋白质和其它蛋白质中的电子传递途径(Mayo 等, (196) *Science* 233:948-952; Gray 和 Malmstrom (1989) *Biochemistry* 28:7499-7505)。

#### 实施例 4: 用来掺入氧化还原活性氨基酸的示例性 O-RS 和 O-tRNA

一种示例性 O-tRNA 包含 SEQ ID NO.:2(见表 1)。实施例 O-RS 包含 SEQ ID NO:1(见表 1)和图 6 提供的氨基酸序列。编码 O-RS 或其一部分的多核苷酸的例子包括编码含有 SEQ ID NO:1 氨基酸序列的多核苷酸。例如, 图 6 和 SEQ ID NO:3 提供的多核苷酸可编码示例性 O-RS。

应当理解的是本文所描述的实施例和实施方案只是为了说明的目的, 本领域的技术人员可以对这些实施例和实施方案进行各种修改或改变, 这些经修改的实施例或实施方案也包括在本申请的精神和范围之内以及本发明附属权利要求的范围之内。

虽然为了便于理解本发明, 在上文已经对本发明做了某种程度的详细描述, 但是应该清楚的是, 只要不偏离本发明的范围, 本领域的技术人员通过阅读本说明书就可以对本发明的形式和细节作出各种修改。例如, 上面所描述的各种技术方法和装置都可以不同组合的方式加以使用。出于所有目的, 本申请所引用的所有文献、专

利、专利申请和/或其他文献都已完整纳入本文作为参考，这就好比各个文献、专利、专利申请和/或其他文献都被单独纳入本文作为参考。

表 1:序列

| SEQ ID NO: | 描述                                                                                                                  | 序列                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------------|---------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1          | DHPRS(合成酶)的氨基酸序列,<br>根据詹氏甲烷球菌酪氨酸 tRNA-合成酶<br>(MjTyrRS)具有以下氨基酸改变:<br>Tyr32→Leu, Ala67→Ser, His70→Asn,<br>Ala167→Gln  | MDEFEMIKRNTSEIISEELREVLKKDEKSA<br>LIGFEPGSKIHLGHYLQIKKMIDLQNAFDDI<br>IILLSDLNAYLNQKGELDEIRKIGDYNKKVF<br>EAMGLKAKYVYGSEFQLDKDYTLNVYRLALK<br>TTLKRARRSMELIAREDENPKVAEVIYPIMQ<br>VNDIHYLGVDVQVGGMEQRKIHMLARELLPK<br>KVVCIHNFVLTGLDGEKMSSSKGNFIAVDD<br>SPEEIRAKIKKAYCPAGVVEGNPIMEIAKYF<br>LEYPLTIKRPEKFGGDLTVNSYEELSLFKN<br>KELHPMDLKNVAEELIKILEPIRKRL                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 2          | <i>mutRNA<sup>Tyr</sup><sub>CUA</sub></i>                                                                           | CCGCCGGUAGUUCACAGGGCAGAACGCCGG<br>ACUCUAAAUCGCCAUGGCCUGGUUCAAUUC<br>CGCCCCGCCGACCA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 3          | DHPRS(合成酶)的核苷酸序列,<br>根据詹氏甲烷球菌酪氨酸 tRNA-合成酶<br>(MjTyrRS)编码的氨基酸改变如下:<br>Tyr32→Leu, Ala67→Ser, His70→Asn,<br>Ala167→Gln | ATGGACGAATTTGAAATGATAAAGAGAAACACAT<br>CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT<br>TTTAAAAAAGATGAAAAATCTGCTCTCATAGGT<br>TTTGAACCAAGTGGTAAAAATACATTTAGGGCATT<br>ATCTCCAAATAAAAAAGATGATTGATTACAAAA<br>TGCTGGATTGATATATATATGTTGAGCGAT<br>TTAAACGCCATTAAACCAGAAAGGAGAGTTGG<br>ATGAGATTAGAAAAATAGGAGATTAAACAAAA<br>AGTTTGTGAAGCAATGGGGTTAAAGGCAAAATAT<br>GTTTATGGAAGTGAATTCAGCTTGATAAGGATT<br>ATACACTGAATGTCATAGATTGGCTTTAAAAAC<br>TACCTTAAAAAGAGCAAGAGGAGTATGGAACCT<br>ATAGCAAGAGAGGATGAAAAATCCAAGGTTGCTG<br>AAGTTATCTATCCAATAATGCAGGTTAATGATAT<br>TCATTATTAGGCGTTGATGTTTCAGGTTGGAGGG<br>ATGGAGCAGAGAAAAATACACATGTTAGCAAGGG<br>AGCTTTTACCAAAAAAGGTTGTTTGTATTCACAA<br>CCCTGTCTTAACGGGTTTGGATGGAGAAGGAAAG<br>ATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTG<br>ATGACTCTCCAGAAGAGATTAGGGCTAAGATAAA<br>GAAAGCATACTGCCAGCTGGAGTTGTTGAAGGA<br>AATCCAATAATGGAGATAGCTAAATACTTCCTTG<br>AATATCCTTTAACCATAAAAAAGGCCAGAAAAATT<br>TGGTGGAGATTTGACAGTTAATAGCTATGAGGAG<br>TTAGAGAGTTTATTTAAAAATAAGGAATTCATC<br>CAATGGATTAAAAAAATGCTGTAGCTGAAGAATC<br>TATAAAGATTTTAGAGCCAATTAGAAAGAGATTA |
| 4          | 詹氏甲烷球菌酪氨酸 tRNA-合成酶<br>(MjTyrRS)的氨基酸序列                                                                               | MDEFEMIKRNTSEIISEELREVLKKDEKSA<br>YIGFEPGSKIHLGHYLQIKKMIDLQNAFDDI<br>IILLADLHAYLNQKGELDEIRKIGDYNKKVF<br>EAMGLKAKYVYGSEFQLDKDYTLNVYRLALK<br>TTLKRARRSMELIAREDENPKVAEVIYPIMQ<br>VNDIHYLGVDVAVGGMEQRKIHMLARELLPK<br>KVVCIHNFVLTGLDGEKMSSSKGNFIAVDD<br>SPEEIRAKIKKAYCPAGVVEGNPIMEIAKYF<br>LEYPLTIKRPEKFGGDLTVNSYEELSLFKN<br>KELHPMDLKNVAEELIKILEPIRKRL                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

|   |                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5 | 詹氏甲烷球菌酪氨酸 tRNA-合成酶<br>(MjTyrRS)的核苷酸序列 | ATGGACGAATTTGAAATGATAAAGAGAAACACAT<br>CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT<br>TTTAAAAAAGATGAAAAATCTGCTTACATAGGT<br>TTTGAACCAAGTGTTAAAAATACATTTAGGGCATT<br>ATCTCCAAATAAAAAAGATGATTGATTTACAAAA<br>TGCTGGATTGTATATAATTATATTGTTGGCTGAT<br>TTACACGCCATTTTAAACCAGAAAGGAGAGTTGG<br>ATGAGATTAGAAAAATAGGAGATTATAACAAAAA<br>AGTTTGTGAAGCAATGGGGTTAAAGGCCAAATAT<br>GTTTATGGAAGTGAATTCCAGCTTGATAAGGATT<br>ATACACTGAATGTCATAGATTGGCTTTAAAAAC<br>TACCTTAAAAAGAGCAAGAAGGAGTATGGAAGTT<br>ATAGCAAGAGAGGATGAAATCCAAAGGTTGCTG<br>AAGTTATCTATCCAATAATGCAGGTTAATGATAT<br>TCATTATTTAGGCGTTGATGTTGCAGTTGGAGGG<br>ATGGAGCAGAGAAAAATACACATGTTAGCAAGGG<br>AGCTTTTACCAAAAAAGGTTGTTTGTATTACAA<br>CCCTGTCCTTAACGGGTTTGGATGGAGAAGGAAAG<br>ATGAGTTCCTTCAAAGGGGAATTTTATAGCTGTTG<br>ATGACTCTCCAGAAAGAGATTAGGGCTAAGATAAA<br>GAAAGCATACTGCCAGCTGGAGTTGTTGAAGGA<br>AATCCAATAATGGAGATAGCTAAATACTTCCTTG<br>AATATCCTTTAACCATAAAAAGGCCAGAAAAATT<br>TGGTGGAGATTTGACAGTTAATAGCTATGAGGAG<br>TTAGAGAGTTTATTTAAAAATAAGGAATTGCATC<br>CAATGGATTTAAAAAATGCTGTAGCTGAAGAAGT<br>TATAAAGATTTTAGAGCCAAATTAGAAGAGATTAA |
|---|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

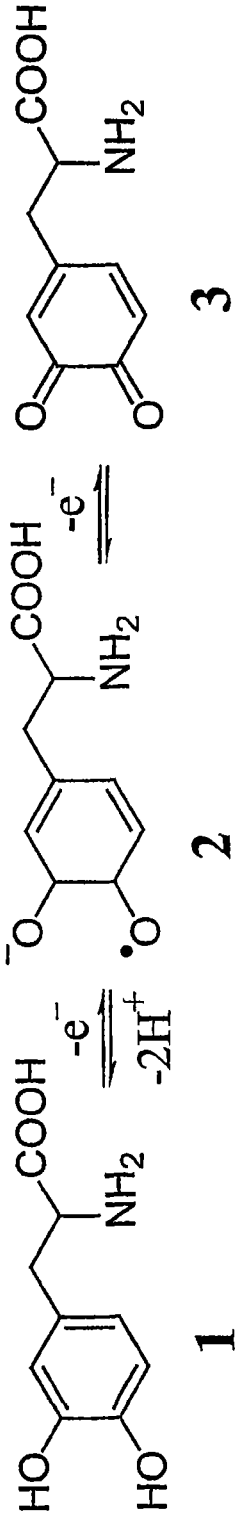


图 1

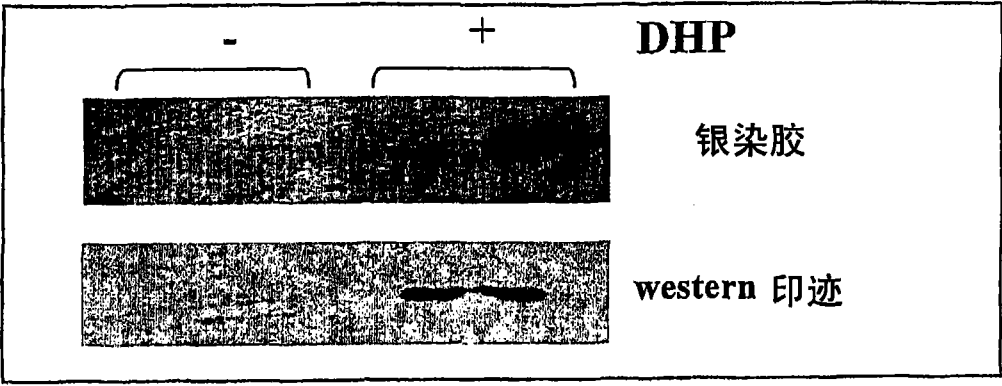


图 2A

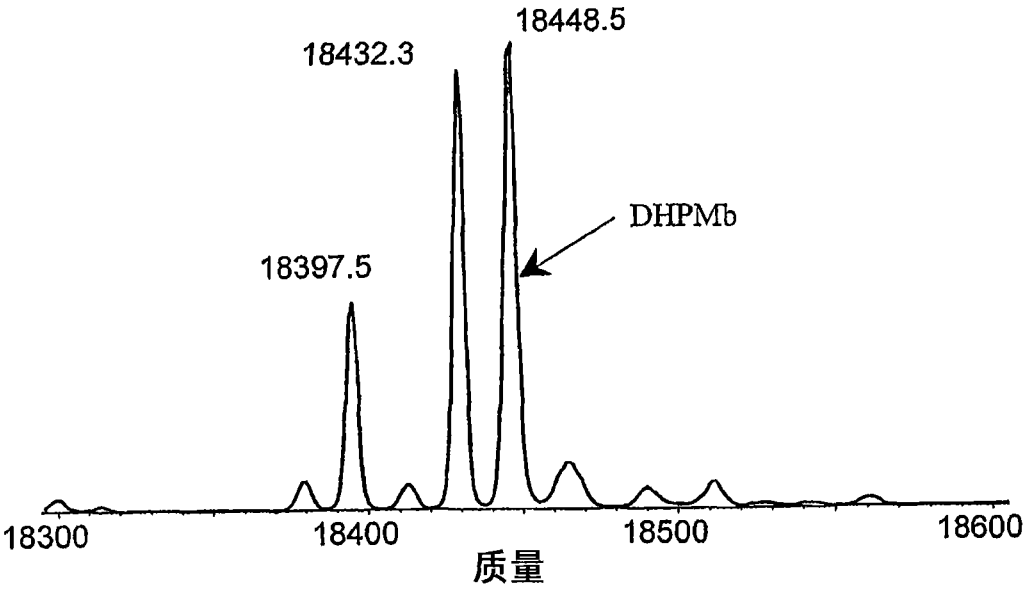


图 2B

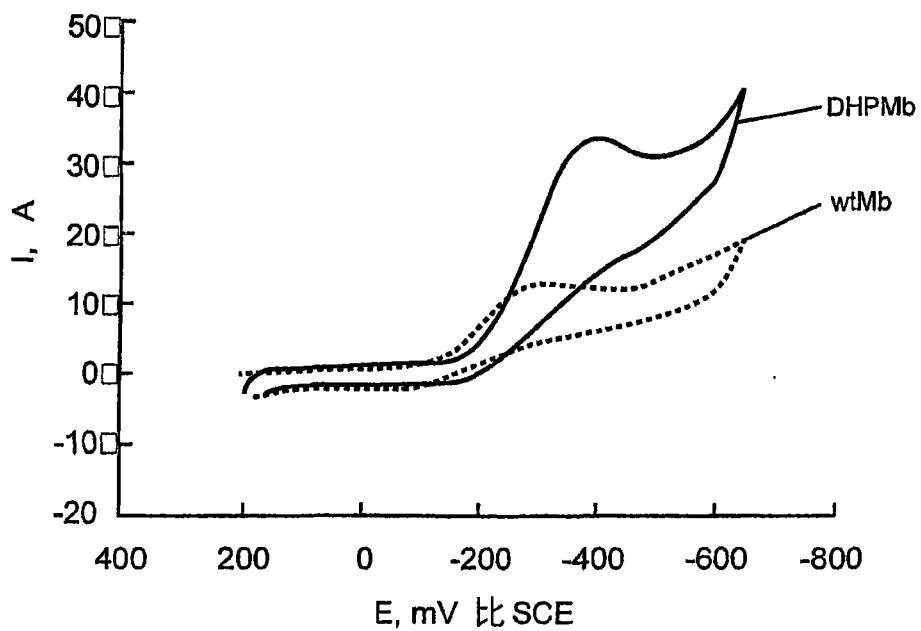


图 3A

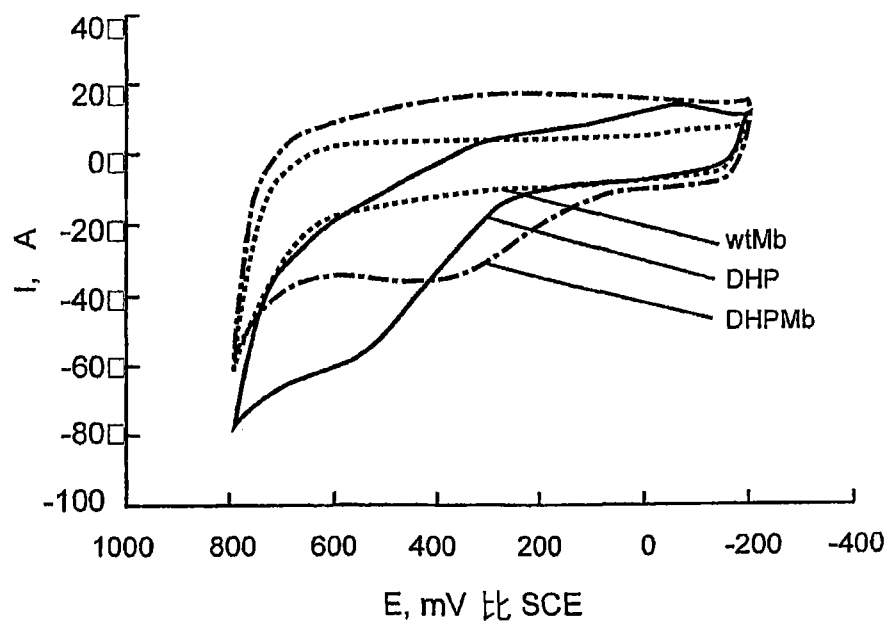


图 3B

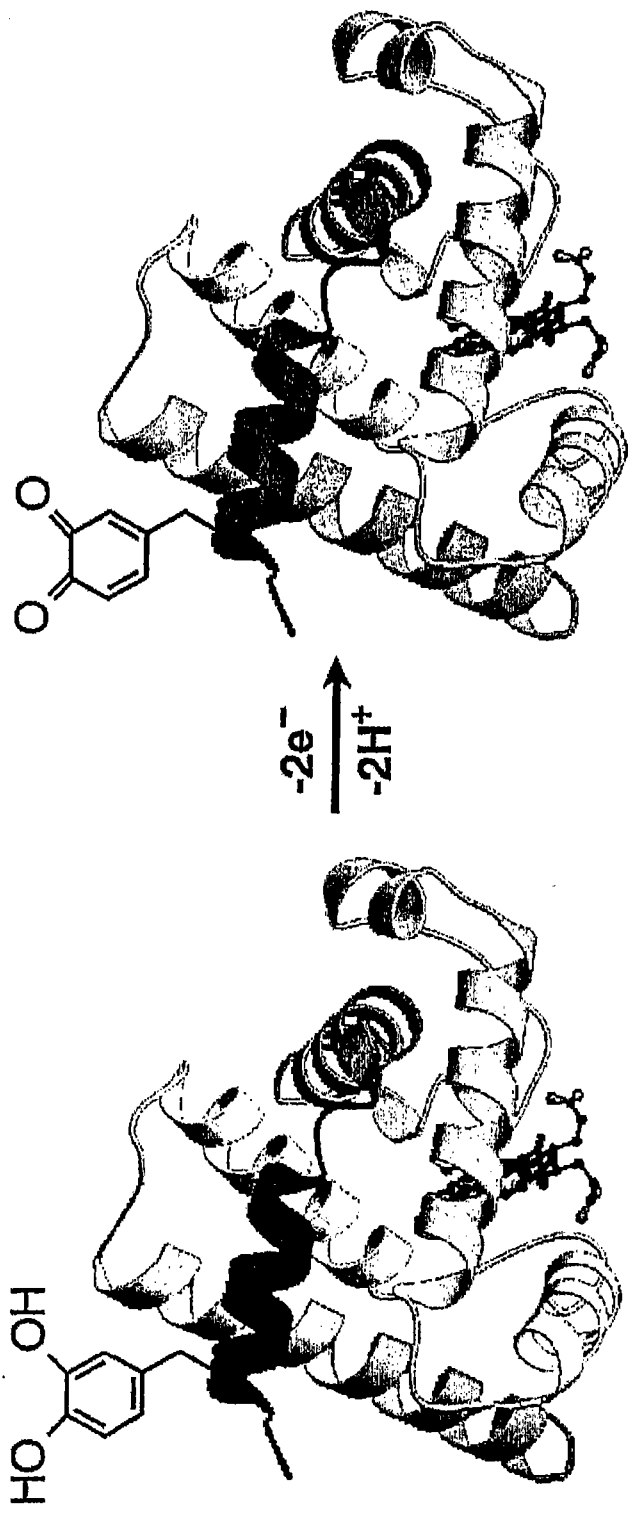


图4

1/1  
 ATG GAC GAA TTT GAA ATG ATA AAG AGA AAC ACA TCT GAA ATT ATC AGC GAG GAA GAG TTA AGA GAG GTT TTA AAA AAA GAT GAA AAA TCT GCT TAC ATA GGT TTT GAA CCA AGT GGT AAA  
 Met asp glu phe glu met ile lys arg asn thr ser glu ile ile ser glu glu glu leu arg glu val leu lys lys asp glu lys ser ala tyr ile gly phe glu pro ser gly lys  
 121/41  
 ATA CAT TTA GGG CAT TAT CTC CAA ATA AAA AAG ATG ATT GAT TTA CAA AAT GCT GGA TTT GAT ATA ATT ATA TTG TTG GCT GAT TTA CAC GCC TAT TTA AAC CAG AAA GGA GAG TTG GAT  
 ile his leu glu gly his tyr leu gln ile lys lys met ile asp leu gln asn ala gly phe asp ile ile leu leu ala asp leu his ala tyr leu asn gln lys gly glu leu asp  
 241/81  
 GAG ATT AGA AAA ATA GGA GAT TAT AAC AAA GAT TTT GAA GCA ATG GGG TTA AAG GCA AAA TAT GTT TAT GGA AGT GAA TTC CAG CTT GAT AAG GAT TAT ACA CTG AAT GTC TAT AGA  
 glu ile arg lys ile gly asp tyr asn lys lys val phe glu ala met gly leu lys ala lys tyr val tyr gly ser glu phe gln leu asp lys asp tyr thr leu asn val tyr arg  
 361/121  
 TTG GCT TTA AAA ACT ACC TTA AAA AGA GCA AGA AGG AGT ATG GAA CTT ATA GCA AGA GAG GAT GAA AAT CCA AAG GTT GCT GAA GTT ATC TAT CCA ATA ATG CAG GTT AAT GAT ATT CAT  
 leu ala leu lys thr thr leu lys arg ala arg arg ser met glu leu ile ala arg glu asp glu asn pro lys val ala glu val ile tyr pro ile met gln val asn asp ile his  
 481/161  
 TAT TTA GGC GTT GAT GTT GCA GTT GGA GGG ATG GAG CAG AGA AAA ATA CAC ATG TTA GCA AGG GAG CTT TTA CCA AAA AAG GTT GTT TGT ATT CAC AAC CCT GTC TTA ACG GGT TTG GAT  
 tyr leu glu val asp val ala val gly gly met glu gln arg lys ile his met leu ala arg glu leu leu pro lys lys val val cys ile his asn pro val leu thr gly leu asp  
 601/201  
 GGA GAA GGA AAG ATG AGT TCT TCA AAA GGG AAT TTT ATA GCT GTT GAT GAC TCT CCA GAA GAG ATT AGG GCT AAG ATA AAG AAA GCA TAC TGC CCA GCT GGA GTT GTT GAA GGA AAT CCA  
 gly glu gly lys met ser ser lys gly asn phe ile ala val asp asp ser pro glu glu ile arg ala lys ile lys lys ala tyr cys pro ala gly val val glu gly asn pro  
 721/241  
 ATA ATG GAG ATA GCT AAA TAC TTC CTT GAA TAT CCT TTA ACC ATA AAA AGG CCA GAA AAA TTT GGT GGA GAT TTG ACA GTT AAT AGC TAT GAG GAG TTA GAG AGT TTA TTT AAA AAT AAG  
 ile met glu ile ala lys tyr phe leu glu tyr pro leu thr ile lys arg pro glu lys phe gly gly asp leu thr val asn ser tyr glu glu leu glu ser leu phe lys asn lys  
 841/281  
 GAA TTG CAT CCA ATG GAT TTA AAA AAT GCT GTA GCT GAA GAA CTT ATA AAG ATT TTA GAG CCA ATT AGA AAG AGA TTA (SEQ ID NO:5)  
 glu leu his pro met asp leu lys asn ala val ala glu leu ile lys ile leu glu pro ile arg lys arg leu (SEQ ID NO:4)



5



1/1 31/11 61/21 91/31  
 ATG GAC GAA TTT GAA ATG ATA AAG AGA AAC ACA TCT GCT CTC ATA GGT TTT GAA CCA AGT GGT AAA  
 Met asp glu phe glu met ile lys arg asn thr ser glu ile ile ser glu glu glu leu arg glu val leu lys lys asp glu lys ser ala leu  
 121/41 151/51 181/61 211/71  
 ATA CAT TTA GGG CAT TAT CTC CAA ATA AAA AAG ATG ATT GAT TTA CAA AAT GCT GGA TTT GAT ATA ATT ATA TTG TTG AGC GAT TTA AAC  
 ile his leu gly his tyr leu gln ile lys lys met ile asp leu gln asn ala gly phe asp ile ile ile leu leu ser asp leu  
 241/81 271/91 301/101 331/111  
 GAG ATT AGA AAA ATA GGA GAT TAT AAC AAA GTT TTT GAA GCA ATG GGG TTA AAG CCA AAA TAT GTT TAT GGA AGT GAA TTC CAG CTT GAT AAG  
 glu ile arg lys ile gly asp tyr asn lys lys val phe glu ala met gly leu lys ala lys tyr val tyr gly ser glu phe glu leu asp  
 361/121 391/131 421/141 451/151  
 TTG GCT TTA AAA ACT ACC TTA AAA AGA GCA AGA AGG AGT ATG GAA CTT ATA GCA AGA GAG GAT GAA AAT CCA AAG GTT GCT GAA GTT ATC TAT CCA ATA ATG CAG GTT AAT GAT AAT CAT  
 leu ala leu lys thr leu lys arg ala arg arg ser met glu leu ile ala arg glu asp glu asn pro lys val ala glu val ile tyr pro ile met gln val asn asp ile his  
 481/161 511/171 541/181 571/191  
 TAT TTA GGC GTT GAT GTT CAG GTT GGA CGG ATG GAG CAG AGA AAA ATA CAC ATG TTA GCA AGG GAG CTT TTA CCA AAA AAG GTT GTT TGT ATT CAC AAC CCT GTC TTA ACG GGT TTG GAT  
 tyr leu gly val asp val gln val gly gly met glu arg lys ile his met leu ala arg glu leu leu pro lys lys val val cys ile his asn pro val leu thr gly leu asp  
 601/201 631/211 661/221 691/231  
 GGA GAA GGA AAG ATG AGT TCT TCA AAA GGG AAT TTT ATA GCT GTT GAT GAC TCT CCA GAA GAG ATT AGG GCT AAG ATA AAG AAA GCA TAC TGC CCA GCT GGA GTT GTT GAA GGA AAT CCA  
 gly glu gly lys met ser ser lys gly asn phe ile ala val asp asp ser pro glu glu ile arg ala lys ile lys lys ala tyr cys pro ala gly val val glu gly asn pro  
 721/241 751/251 781/261 811/271  
 ATA ATG GAG ATA GCT AAA TAC CTT GAA TAT CCT TTA ACC ATA AAA AGG CCA GAA AAA TTT GGT GGA GAT TTG ACA GTT AAT AGC TAT GAG GAG TTA GAG AGT TTA TTT AAA AAT AAG  
 ile met glu ile ala lys tyr phe leu glu tyr pro leu thr ile lys arg pro glu lys phe gly phe gly gly asp leu thr val asn ser tyr glu leu phe lys asn lys  
 841/281 871/291 901/301  
 GAA TTG CAT CCA ATG GAT TTA AAA AAT GCT GTA GCT GAA GAA CTT ATA AAG ATT TTA GAG CCA ATT AGA AAG AGA TTA (SEQ ID NO:3)  
 glu leu his pro met asp leu lys asn ala val ala glu glu leu ile lys ile leu glu pro ile arg lys arg leu (SEQ ID NO:1)



6