



(12)发明专利申请

(10)申请公布号 CN 110117580 A

(43)申请公布日 2019.08.13

(21)申请号 201810116409.8

(22)申请日 2018.02.05

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

(72)发明人 王江云 安晓景 王天元 韩明杰  
黄爱萍 陈超 江欢欢

(74)专利代理机构 中科专利商标代理有限责任  
公司 11021

代理人 陈晓娜

(51) Int. Cl.

C12N 9/00(2006.01)

C12N 15/52(2006.01)

C12N 9/16(2006.01)

C12P 21/02(2006.01)

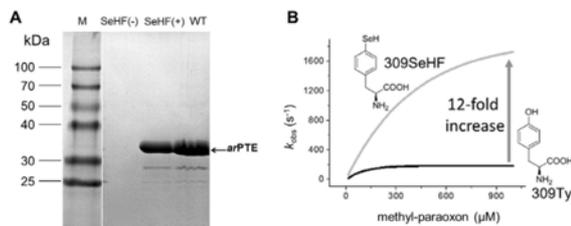
权利要求书2页 说明书13页  
序列表8页 附图5页

(54)发明名称

硒代酪氨酸翻译系统及其应用

(57)摘要

本发明涉及一种氨酰基-tRNA合成酶突变体,其作为一种正交氨酰基-tRNA合成酶,其含有的氨基酸序列选自SEQ ID NO:4所示的氨基酸序列或其具有相同酶活性的保守性变体或同源物。本发明还提供一种硒代酪氨酸翻译系统,其包含:(i)硒代酪氨酸;(ii)所述正交氨酰基-tRNA合成酶;(iii)正交tRNA,其中所述正交氨酰基-tRNA合成酶用硒代酪氨酸优先氨酰化所述正交tRNA;和(iv)编码目标蛋白质的核酸,其中所述核酸含有所述正交tRNA特异性识别的至少一个选择密码子。本发明还提供利用上述硒代酪氨酸翻译系统设计改造蛋白酶的方法,以及该方法编码产生的蛋白酶。



1. 一种正交氨酰基-tRNA合成酶,其含有的氨基酸序列选自由SEQ ID NO:4所示氨基酸序列、SEQ ID NO:4所示的氨基酸序列的保守性变体和与SEQ ID NO:4所示的氨基酸序列具有90%以上的序列同源性的同源物组成的组,其中所述保守性变体或同源物具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性。

2. 一种硒代酪氨酸翻译系统,所述系统包含:

(i) 硒代酪氨酸;

(ii) 权利要求1所述的正交氨酰基-tRNA合成酶;

(iii) 正交tRNA,其包含SEQ ID NO:1所示的多核苷酸序列;其中所述正交氨酰基-tRNA合成酶用所述硒代酪氨酸优先氨酰化所述正交tRNA;和

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

3. 根据权利要求2所述的翻译系统,其特征在于,所述正交tRNA是琥珀抑制型tRNA,所述选择密码子是琥珀密码子,并且所述翻译系统还包含编码所述正交氨酰基-tRNA合成酶的核苷酸序列。

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

5. 根据权利要求4所述的宿主细胞,其中所述宿主细胞是真细菌细胞,优选大肠杆菌细胞。

6. 一种产生在至少一个所选位置定点特异性插入硒代酪氨酸的突变蛋白质的方法,所述方法包括下述步骤:

(a) 提供权利要求2所述的硒代酪氨酸翻译系统,所述系统包含:

(i) 硒代酪氨酸;

(ii) 权利要求1所述的正交氨酰基-tRNA合成酶;

(iii) 正交tRNA,其包含SEQ ID NO:1所示的多核苷酸序列;其中所述正交氨酰基-tRNA合成酶用所述硒代酪氨酸优先氨酰化所述正交tRNA;和

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

(b) 将所述正交tRNA序列、编码所述正交氨酰基-tRNA合成酶的核苷酸序列和编码所述目标蛋白质的核酸序列克隆并转化到适当的宿主细胞中,在培养基中加入硒代酪氨酸,在所述目标蛋白质的翻译期间,硒代酪氨酸氨酰化的正交tRNA识别编码所述目标蛋白质的mRNA上的选择密码子以及硒代酪氨酸,从而将硒代酪氨酸定点特异性插入所述选择密码子对应的氨基酸位置,从而产生在所选位置含硒代酪氨酸的所述目标蛋白质。

7. 根据权利要求6所述的方法,其中所述正交tRNA是琥珀抑制型tRNA,并且所述选择密码子是琥珀密码子。

8. 生产含有硒代酪氨酸的目标蛋白酶突变体的方法,其利用权利要求6所述的方法,其中所用的编码目标蛋白酶突变体的核酸序列在选定的位置包含所述正交tRNA特异性识别的选择密码子,在目标蛋白酶的翻译期间,硒代酪氨酸定点插入到所述选择密码子对应的氨基酸位置,从而产生在选定位置含有硒代酪氨酸的目标蛋白酶突变体。

9. 一种用权利要求8所述的方法设计改造蛋白酶的方法,所述方法包括利用权利要求8

的方法将目标蛋白酶活性位点上的酪氨酸取代为硒代酪氨酸,由此得到含有硒代酪氨酸的目标蛋白酶突变体,从而显著提高酶活性。

10.由权利要求9所述的方法获得的含有硒代酪氨酸的目标蛋白酶突变体,包括但不限于,放射形土壤杆菌 (*Agrobacterium radiobacter*) 磷酸三酯酶,其氨基酸序列如SEQ ID NO:8所示。

## 硒代酪氨酸翻译系统及其应用

### 技术领域

[0001] 本发明属于生物化学领域。具体地,本发明提供一种氨酰基-tRNA合成酶突变体,其为一种正交氨酰基-tRNA合成酶,其含有的氨基酸序列选自由SEQ ID NO:4所示的氨基酸序列和SEQ ID NO:4所示的氨基酸序列的保守性变体组成的组,其中所述保守性变体具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性。本发明还涉及一种硒代酪氨酸(简称为SeHF)翻译系统及其应用。

[0002] 具体而言,本发明涉及利用正交tRNA、正交氨酰基-tRNA合成酶的配对将硒代酪氨酸定点特异性插入目标蛋白质的硒代酪氨酸翻译系统,和利用所述翻译系统在目标蛋白质中定点特异性插入硒代酪氨酸的方法。本发明还涉及一种用这套翻译系统和这种方法设计改造目标蛋白酶的方法,以及该方法产生的含有硒代酪氨酸的突变蛋白酶,例如,插入硒代酪氨酸的放射形土壤杆菌磷酸三酯酶(简称为arPTE)突变体,以及插入硒代酪氨酸的放射形土壤杆菌磷酸三酯酶突变体的应用。

### 背景技术

[0003] 酪氨酸在许多酶催化剂中普遍存在,它的作用包括:1、底物直接亲核攻击(如DNA拓扑异构酶等);2、去质子化过程中的广义碱(如糖苷水解酶等);3、离去基团的广义酸质子化(如酪氨酸苯酚裂解酶等);4、活性部位的氢键供体/受体或电排斥(如酮类固醇异构酶等);5、阳离子- $\pi$ 相互作用(如萜类环化酶等);6、参与氧化还原和电子转移反应(如核糖核苷酸还原酶、Ftm0x1等);7、直接金属配位(如过氧化氢酶、半乳糖氧化酶等)。为了探讨酪氨酸残基在酶催化作用中的机理并更好地设计改造酶,我们将酶活性位点上的酪氨酸突变为带有金属或者邻位取代的非天然氨基酸(简称为UAA)。但是,由于空间位阻,在酶活性位点引入外来原子可能会显著损害其酶活性,因此,我们需要采取一种最小化降低酶活性损伤的策略,即对酶活性位点的氨基酸采取单原子置换。

[0004] 硒与氧的主族相同,硒的范德华半径为1.9埃,而氧的半径为1.52埃。我们选择合成并遗传插入2-氨基-3-(4-氢硒苯基)丙酸(简称硒代酪氨酸或SeHF),它与酪氨酸相比有一个侧链上单个氧-硒原子取代,因此可以提供一种有价值的机制工具来研究酪氨酸在酶催化剂中的作用。同时,与苯酚( $pK_a=10$ )相比,苯基烯醇侧链的 $pK_a$ ( $pK_a=5.9$ )显著降低,苯基硒酸盐阴离子的亲核性增强,这也为更好地设计催化酶提供了良好机会。此外,硒可用于通过多波长反常散射定相法来测定蛋白质结构。受含有硒代半胱氨酸的天然或设计酶启发,硒代酪氨酸也可用于设计具有过氧化物酶和脱卤素酶活性的人造酶。

[0005] 磷酸三酯酶(简称为PTE)是具有多种底物的双金属蛋白,底物之一的对氧磷是一种最常用的杀虫剂。使用金属活化的水作为亲核试剂,PTE可以通过简单的一步 $S_N2$ 置换反应裂解甲基对氧磷(简称为DNP)的P-O键,生成磷酸二甲酯(DMP)和对硝基苯酚(PNP),而不涉及任何共价的酶-底物中间体,使其成为理想的酶设计模型。此外,更好地了解和改进PTE催化活性对保护环境也具有一定的重要性。

[0006] 细菌PTEs的结构和机制已被广泛报道,其转换率由若干组分确定:米氏配合物

(ES) 结合系数 ( $k_1$ ) 和解离系数 ( $k_2$ ), P-O键断裂系数 ( $k_3$ ) 和产物释放系数 ( $k_5$ )。据报道, P-O键断裂 $k_3$ 步骤不是限速步骤, 而产物释放 $k_5$ 则是催化循环中最慢和限制速率的步骤。PTEs的底物结合袋, 同时也是产物释放袋, 由许多疏水性残基组成, 包括Tyr309、Phe132和Trp131, 对氧磷底物的芳环被以上这些氨基酸的侧链包围, PTEs的转化率被认为是已经接近它们的自然进化极限。众多的PTEs突变体通过实验室定向进化或合理设计分离出来, 以提高对各种底物的催化效率。然而, 野生型和突变型PTEs的最适pH值通常大于8.5, 这可能限制了其在环境修复、农药解毒和临床应用中的应用。

[0007] 我们推测, 带负电的SeHF可以促进产物释放, 并且SeHF侧链的 $pK_a$ 值 (pH5.9) 可以显著改善PTE在中性pH环境下的催化活性。因此本研究旨在通过遗传密码扩展, 用非天然氨基酸——SeHF替代酪氨酸, 从而显著提高PTE的酶活性。同时, 本研究现已开发了在原核和真核生物中将各种非天然氨基酸体内位点特异性地定点插入蛋白质的通用方法。这些方法依赖于正交蛋白质翻译组分, 所述组分识别合适的选择密码子 (selector codon), 从而能在体内多肽翻译期间将所需的非天然氨基酸插入限定位置。这些方法利用识别选择密码子的正交tRNA (O-tRNA), 而相应的特异性正交氨酰基-tRNA合成酶 (O-RS) 用非天然氨基酸加载该O-tRNA。这些组分不与宿主生物体内的任何内源性tRNA、氨酰基-tRNA合成酶 (RS)、氨基酸或密码子交叉反应 (即, 它必须是正交的)。利用这种正交tRNA-RS配对可能遗传编码大量结构各异的非天然氨基酸。

[0008] 本领域普遍知道利用适合于制备含一个或多个非天然氨基酸的蛋白质的正交翻译系统, 例如产生正交翻译系统的通用方法。例如, 参见国际公布号WO 2002/086075, 其发明名称为“METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS”; WO 2002/085923, 其发明名称为“IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS”; WO 2004/094593, 其发明名称为“EXPANDING THE EUKARYOTIC GENETIC CODE”。定点特异性插入非天然氨基酸的正交翻译系统及它们的产生和使用方法的其他讨论还可参见Wang和Schultz, Chem. Commun. (Camb) 1:1-11 (2002); Wang和Schultz, Angewandte Chemie Int. Ed. 44 (1):34-66 (2005); Xie和Schultz, Methods 36 (3):227-238 (2005); Xie和Schultz, Curr. Opinion in Chemical Biology 9 (6):548-554 (2005); Wang等, Annu. Rev. Biophys. Biomol. Struct. 35:225-249 (2006)。

## 发明内容

### [0009] I、技术方案

[0010] 本发明人经过筛选, 首次获得一种氨酰基-tRNA合成酶突变体, 其为一种正交氨酰基-tRNA合成酶, 其含有的氨基酸序列选自由SEQ ID NO:4所示氨基酸、SEQ ID NO:4所示的氨基酸序列的保守性变体和与SEQ ID NO:4所示的氨基酸序列具有一定的序列同源性的同源物组成的组, 其中所述保守性变体和同源物具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性, 其中所述序列同源性可以为80%以上, 优选85%以上, 更优选90%以上, 甚至更优选95%以上, 甚至更优选99%以上。这种氨酰基-tRNA合成酶突变体能够用硒代酪氨酸 (简称为SeHF) 优先氨酰化与之配对的正交tRNA, 从而在翻译的氨基酸序列中插入SeHF。这是本发明人首次发现的, 相应地, 在本发明中将其命名为正交硒代酪氨酸氨酰基-tRNA合成酶 (简称为SeHF<sub>RS</sub>)。

[0011] 在一个优选的实施方案中,所述正交硒代酪氨酸氨酰基-tRNA合成酶的氨基酸序列如SEQ ID NO:4所示。

[0012] 在上述发现的基础上,本发明提供一种利用正交tRNA、正交氨酰基-tRNA合成酶的配对将硒代酪氨酸定点特异性插入目标蛋白质的硒代酪氨酸翻译系统,和利用所述翻译系统在目标蛋白质中定点特异性插入硒代酪氨酸的方法。本发明还涉及用这种翻译系统和这种方法产生的含有硒代酪氨酸的突变蛋白质及其应用。

[0013] 因此,本发明的目的在于提供一种正交硒代酪氨酸氨酰基-tRNA合成酶,提供利用正交tRNA、正交氨酰基-tRNA合成酶的配对将硒代酪氨酸定点特异性插入蛋白质的硒代酪氨酸翻译系统,并且提供利用该翻译系统在目标蛋白质中定点特异性插入硒代酪氨酸的方法。

[0014] 本发明还提供利用本发明的硒代酪氨酸翻译系统设计改造目标蛋白酶的方法,以及利用该方法产生的含有至少一个硒代酪氨酸的突变蛋白酶。在本发明的优选方面中,本发明人利用这种方法将硒代酪氨酸定点特异性插入目的蛋白酶中以增强酶活性,所述目的蛋白酶包括,但不限于,磷酸三酯酶,例如,放射形土壤杆菌磷酸三酯酶。然而,本领域技术人员应该理解,本发明的方法也可以用于在磷酸三酯酶之外的多种蛋白酶中,并不局限于该蛋白酶。

[0015] 本领域技术人员应该理解,在本发明中,除了SEQ ID NO:4所示的氨基酸序列之外,术语“本发明的正交氨酰基-tRNA合成酶”或“正交硒代酪氨酸氨酰基-tRNA合成酶”还包括SEQ ID NO:4所示氨基酸序列的保守性变体,只要所述保守性变体具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性即可;并且还包括将SEQ ID NO:4所示的氨基酸序列经过一个或多个氨基酸的取代、缺失或添加,且具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性的由SEQ ID NO:4所示的氨基酸序列衍生的氨基酸序列;并且还包括与SEQ ID NO:4所示的氨基酸序列具有一定的序列同源性的同源物,其中所述序列同源性可以为80%以上,优选85%以上,更优选90%以上,甚至更优选95%以上,甚至更优选99%以上,并且所述同源物具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性。

[0016] 并且,编码本发明的正交硒代酪氨酸氨酰基-tRNA合成酶(SeHFRS)的核苷酸序列也包括在本发明的范围内。优选地,所述编码核苷酸序列为SEQ ID NO:3所示。

[0017] 具体来说,本发明提供在体内(例如在宿主细胞内)识别选择密码子(Selector codon),如琥珀终止密码子(TAG),从而将非天然氨基酸——硒代酪氨酸定点特异性插入到多肽链中的硒代酪氨酸翻译系统。所述硒代酪氨酸翻译系统包含不与宿主细胞翻译机制相互作用的正交-tRNA(O-tRNA)和正交氨酰基-tRNA合成酶(O-RS)配对。即,宿主细胞内源性氨酰基-tRNA合成酶不会识别O-tRNA。类似地,本发明提供的O-RS不以显著水平或者某些情况下不以可检测水平地识别内源性tRNA。利用所述翻译系统能够产生在翻译过程中定点特异性插入硒代酪氨酸的大量蛋白质。

[0018] 在本发明的优选方面中,本发明提供硒代酪氨酸翻译系统。所述翻译系统包含:

[0019] (a) 硒代酪氨酸,

[0020] (b) 本发明的正交氨酰-tRNA合成酶(O-RS),和

[0021] (c) 正交tRNA(O-tRNA),其包含SEQ ID NO:1所示的多核苷酸序列,其中所述正交氨酰-tRNA合成酶用硒代酪氨酸优先氨酰化所述O-tRNA;

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

[0023] 优选地,本发明的硒代酪氨酸翻译系统还包含编码正交氨酰基-tRNA合成酶的核苷酸序列。

[0024] 所述系统中所用的“本发明的正交氨酰基-tRNA合成酶(O-RS)”即为本发明人首次发现的氨酰基tRNA合成酶突变体,其含有的氨基酸序列选自由SEQ ID NO:4所示氨基酸序列和SEQ ID NO:4所示的氨基酸序列的保守性变体组成的组,其中所述保守性变体具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性。换言之,在本发明中,术语“本发明的正交氨酰基-tRNA合成酶”可与“本发明的正交硒代酪氨酸氨酰基-tRNA合成酶”互换使用。

[0025] 该翻译系统中的各种组分可以衍生自各种物种来源,例如,该翻译系统中的各组分衍生自詹氏甲烷球菌(*Methanococcus jannaschii*)。例如,正交tRNA(O-tRNA)为古菌来源的反密码子突变为与琥珀密码互补的酪氨酸tRNA。在一些实施方式中,O-tRNA是琥珀抑制型tRNA。在一些实施方式中,O-tRNA包含SEQ ID NO:1所示的多核苷酸序列,优选地,O-tRNA的序列如SEQ ID NO:1所示。在一个实施方式中,用于该系统的正交氨酰基-tRNA合成酶可以包含SEQ ID NO:4所示的氨基酸序列及该序列的保守变体。在优选的实施方式中,用于该系统的正交氨酰基-tRNA合成酶的氨基酸序列为SEQ ID NO:4所示。

[0026] 在一些方面中,本发明的硒代酪氨酸翻译系统还包含编码目标蛋白质的核酸,其中所述核酸具有由正交tRNA(O-tRNA)特异性识别的至少一个选择密码子。在优选方面中,所述正交tRNA是琥珀抑制型tRNA,并且所述选择密码子是琥珀密码子。

[0027] 在一些方面中,本发明提供包含编码本发明的正交氨酰基-tRNA合成酶的核苷酸序列和相对应的正交tRNA序列的宿主细胞。所用的宿主细胞不作具体限定,只要正交氨酰基-tRNA合成酶和正交tRNA在它们的宿主细胞环境中保留它们的正交性即可。例如,所述宿主细胞可以是真细菌细胞,优选大肠杆菌。

[0028] 本发明还提供产生在至少一个所选位置定点特异性插入硒代酪氨酸的突变蛋白质的方法。所述方法利用上述硒代酪氨酸翻译系统。所述方法通常包括下述步骤:

[0029] (a) 提供含有以下组分的硒代酪氨酸翻译系统的步骤:

[0030] (i) 硒代酪氨酸;

[0031] (ii) 本发明的正交氨酰基-tRNA合成酶(O-RS);

[0032] (iii) 正交tRNA(O-tRNA),其包含SEQ ID NO:1所示的多核苷酸序列,其中所述O-RS用硒代酪氨酸优先氨酰化所述O-tRNA;和

[0033] (iv) 编码目标蛋白质的核酸,其中所述核酸含有O-tRNA特异性识别的至少一个选择密码子(任选地为琥珀密码子);

[0034] (b) 将所述正交tRNA序列、编码所述正交氨酰基-tRNA合成酶的核苷酸序列和编码所述目标蛋白质的核酸序列克隆并转化到适当的宿主细胞中,在培养基中加入硒代酪氨酸,在所述目标蛋白质的翻译过程中,硒代酪氨酸氨酰化的正交tRNA识别编码所述目标蛋白质的mRNA上的选择密码子以及硒代酪氨酸,从而将硒代酪氨酸定点特异性插入所述选择密码子对应的氨基酸位置,从而产生在所选位置含有硒代酪氨酸的突变蛋白质。

[0035] 本领域技术人员应该理解,适当的重组载体的构建和宿主细胞的筛选可以通过常规分子克隆技术和筛选技术实现。

[0036] 本领域技术人员应该理解,在步骤(b)中,将所述正交tRNA序列、编码所述正交氨酰基-tRNA合成酶的核苷酸序列和编码所述目标蛋白质的核酸序列克隆并转化到适当的宿主细胞中可以通过多种方式进行,例如,将所述正交tRNA序列、编码所述正交氨酰基-tRNA合成酶的核苷酸序列和编码所述目标蛋白质的核酸序列分别可操作性地连接到适当的载体中,再以任意次序或三者共同转化到适当的宿主细胞中;或者,也可以将所述正交tRNA序列和编码所述正交氨酰基-tRNA合成酶的核苷酸序列可操作性地连接到一个适当的载体中(两种序列之间有或无适当的接头连接),将编码所述目标蛋白质的核酸序列可操作性地连接到另一种不同的适当的载体中,然后将构建好的两种重组载体共同转化到适当的宿主细胞中;或者,也可以将所述正交tRNA序列和编码所述目标蛋白质的核酸序列可操作性地连接到一个适当的载体中(两种序列之间有或无适当的接头连接),将编码所述正交氨酰基-tRNA合成酶的核苷酸序列可操作性地连接到另一种不同的适当的载体中,然后将构建好的两种重组载体共同转化到适当的宿主细胞中。或者,也可以将正交tRNA序列、编码所述正交氨酰基-tRNA合成酶的核苷酸序列和编码目标蛋白质的核酸序列以任意适当的顺序可操作性地连接在一起,然后克隆到一个载体上,最后转化到适当的宿主细胞中。上述克隆方案都是可行的,本领域技术人员可以根据实验的需要容易地进行适当的选择。

[0037] 另外,本领域技术人员还应该理解,为了避免宿主细胞对外源重组载体的“踢除”效应,往往选择用带有不同抗生素标记的载体来构建需要共同转化到同一宿主细胞中的核酸序列片段。对于适当的载体的选择、重组载体的构建、宿主细胞的转化或转染等等,都是本领域的常规技术,例如,可以参见美国冷泉港实验室出版的分子克隆手册。

[0038] 在所述方法的一些实施方式中,提供翻译系统的步骤包括通过定点诱变使野生型氨酰基-tRNA合成酶的氨基酸结合口袋发生突变,筛选出用所述非天然氨基酸(即硒代酪氨酸)优先氨酰化所述O-tRNA的氨酰基-tRNA合成酶突变体(即,本发明所用的正交氨酰基-tRNA合成酶)。所述筛选步骤包括定点诱变后从得到的氨酰基-tRNA合成酶分子库进行所述O-RS的正选择和负选择(参见下述实施例2)。在一些实施方式中,提供翻译系统的步骤还包括提供O-tRNA的序列,O-tRNA为古菌来源的反密码子突变为与琥珀密码互补的酪氨酸tRNA,例如,所述O-tRNA是琥珀抑制型tRNA,或者O-tRNA包含SEQ ID NO:1所示的多核苷酸序列。在这些方法中,提供翻译系统的步骤还包括提供含有所述翻译系统所用的琥珀选择密码子的编码目标蛋白质的核酸。

[0039] 在一个实施方案中,还可在宿主细胞内实施产生含有硒代酪氨酸的突变蛋白质的方法。在这些情况中,提供的宿主细胞包含本发明的硒代酪氨酸翻译系统(即,包含编码本发明的O-RS的核苷酸序列、O-tRNA序列和含有至少一个选择密码子的编码目标蛋白质的核酸),而在适宜的培养条件下(例如,在培养基中添加硒代酪氨酸等)培养该宿主细胞可导致在所述目标蛋白质中定点特异性插入硒代酪氨酸。在一些实施方式中,提供步骤包括提供真细菌宿主细胞(例如,大肠杆菌等)。

[0040] 本发明还提供生产含有硒代酪氨酸的蛋白酶突变体的方法,其利用上述产生在至少一个所选位置定点特异性插入硒代酪氨酸的突变蛋白质的方法进行,其中所用的编码蛋白酶突变体的核酸序列在选定的位置包含所述正交tRNA特异性识别的选择密码子,在蛋白酶的翻译期间,硒代酪氨酸定点插入到所述选择密码子对应的氨基酸位置,从而产生在选定位置含有硒代酪氨酸的蛋白酶突变体。

[0041] 优选地,本发明还提供生产含有硒代酪氨酸的放射形土壤杆菌磷酸三酯酶(arPTE)突变体的方法,所述方法利用上述硒代酪氨酸翻译系统进行,所述方法通常包括下述步骤:

[0042] (a) 提供含有以下组分的硒代酪氨酸翻译系统的步骤:

[0043] (i) 硒代酪氨酸;

[0044] (ii) 正交氨酰基-tRNA合成酶(O-RS);

[0045] (iii) 正交tRNA(O-tRNA),其包含SEQ ID NO:1所示的多核苷酸序列,其中所述O-RS用所述硒代酪氨酸优先氨酰化所述O-tRNA;和

[0046] (iv) 编码所述放射形土壤杆菌磷酸三酯酶(arPTE)的核酸,例如,但不限于,SEQ ID NO:7(注:为了在大肠杆菌中表达纯化有活性的arPTE蛋白,将arPTE第1-24个氨基酸删除,第364位的V被突变为S,故在SEQ ID NO:7中不包含编码氨基酸1-24的核酸序列),其中所述核酸含有所述O-tRNA特异性识别的至少一个选择密码子(任选地为琥珀密码子);

[0047] (b) 将所述正交tRNA序列、编码所述正交氨酰基-tRNA合成酶的核苷酸序列和编码放射形土壤杆菌磷酸三酯酶的核酸序列克隆并转化到适当的宿主细胞中,在培养基中加入硒代酪氨酸,在所述放射形土壤杆菌磷酸三酯酶(arPTE)的翻译过程中,硒代酪氨酸氨酰化的正交tRNA识别编码磷酸三酯酶的mRNA上的选择密码子以及硒代酪氨酸,从而将硒代酪氨酸定点插入所述磷酸三酯酶的特定位置(即,所述选择密码子对应的氨基酸位置)。

[0048] 本发明还提供一种新颖的设计改造蛋白酶的方法,所述方法包括利用本发明的硒代酪氨酸翻译系统将所述目标蛋白酶活性位点上的酪氨酸取代为硒代酪氨酸,由此得到含有硒代酪氨酸的目标蛋白酶突变体,以增强酶活性。

[0049] 优选地,本发明人利用所述硒代酪氨酸翻译系统设计改造放射形土壤杆菌磷酸三酯酶(arPTE),并且产生了含有硒代酪氨酸的arPTE突变体。具体而言,在野生型arPTE的309位(注:由于本发明所表达的arPTE去除了野生型的第1-24位氨基酸,因此野生型arPTE的309位氨基酸对应本发明的arPTE的第285位氨基酸)引入硒代酪氨酸,所述arPTE突变体的氨基酸序列为SEQ ID NO:8,在SEQ ID NO:8中,第285位氨基酸为硒代酪氨酸(在图5所示的SEQ ID NO:8中以“Y\*”表示),所述arPTE突变体较野生型arPTE相比,酶活性显著提高。

[0050] 综上所述,本发明提供下述技术方案:

[0051] 1. 一种正交氨酰基-tRNA合成酶,其含有的氨基酸序列选自由SEQ ID NO:4所示氨基酸序列、SEQ ID NO:4所示的氨基酸序列的保守性变体和与SEQ ID NO:4所示的氨基酸序列具有90%以上的序列同源性的同源物组成的组,其中所述保守性变体或同源物具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性。

[0052] 2. 根据第1项所述的正交氨酰基-tRNA合成酶,其含有的氨基酸序列是与SEQ ID NO:4所示的氨基酸序列具有95%以上的序列同源性的同源物,其中所述同源物具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性。

[0053] 3. 根据第2项所述的正交氨酰基-tRNA合成酶,其含有的氨基酸序列是与SEQ ID NO:4所示的氨基酸序列具有99%以上的序列同源性的同源物,其中所述同源物具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性。

[0054] 4. 一种硒代酪氨酸翻译系统,所述系统包含:

[0055] (i) 硒代酪氨酸;

[0056] (ii) 第1-3项中任一项所述的正交氨酰基-tRNA合成酶;

[0057] (iii) 正交tRNA,其包含SEQ ID NO:1所示的多核苷酸序列;其中所述正交氨酰基-tRNA合成酶用所述硒代酪氨酸优先氨酰化所述正交tRNA;和

[0058] (iv) 编码目标蛋白质的核酸,其中所述核酸含有所述正交tRNA特异性识别的至少一个选择密码子。

[0059] 5. 根据第4项所述的翻译系统,其特征在于,所述正交tRNA是琥珀抑制型tRNA,所述选择密码子是琥珀密码子,并且所述翻译系统还包含编码所述正交氨酰基-tRNA合成酶的核苷酸序列。

[0060] 6. 一种宿主细胞,其包含编码第1-3项中任一项所述的正交氨酰基-tRNA合成酶的核苷酸序列和相对应的正交tRNA序列。

[0061] 7. 根据第4项所述的宿主细胞,其中所述宿主细胞是真细菌细胞,优选大肠杆菌细胞。

[0062] 8. 一种产生在至少一个所选位置定点特异性插入硒代酪氨酸的突变蛋白质的方法,所述方法包括下述步骤:

[0063] (a) 提供第4项所述的硒代酪氨酸翻译系统,所述系统包含:

[0064] (i) 硒代酪氨酸;

[0065] (ii) 第1-3项中任一项所述的正交氨酰基-tRNA合成酶;

[0066] (iii) 正交tRNA,其包含SEQ ID NO:1所示的多核苷酸序列;其中所述正交氨酰基-tRNA合成酶用所述硒代酪氨酸优先氨酰化所述正交tRNA;和

[0067] (iv) 编码所述目标蛋白质的核酸,其中所述核酸在所选的位置包含所述正交tRNA特异性识别的至少一个选择密码子;和

[0068] (b) 将所述正交tRNA序列、编码所述正交氨酰基-tRNA合成酶的核苷酸序列和编码所述目标蛋白质的核酸序列克隆并转化到适当的宿主细胞中,在培养基中加入硒代酪氨酸,在所述目标蛋白质的翻译期间,硒代酪氨酸氨酰化的正交tRNA识别编码所述目标蛋白质的mRNA上的选择密码子以及硒代酪氨酸,从而将硒代酪氨酸定点特异性插入所述选择密码子对应的氨基酸位置,从而产生在所选位置含硒代酪氨酸的所述目标蛋白质。

[0069] 9. 根据第8项所述的方法,其中所述正交tRNA是琥珀抑制型tRNA,并且所述选择密码子是琥珀密码子。

[0070] 10. 生产含有硒代酪氨酸的目标蛋白酶突变体的方法,其利用第8项所述的方法,其中所用的编码目标蛋白酶突变体的核酸序列在选定的位置包含所述正交tRNA特异性识别的选择密码子,在目标蛋白酶的翻译期间,硒代酪氨酸定点插入到所述选择密码子对应的氨基酸位置,从而产生在选定位置含有硒代酪氨酸的目标蛋白酶突变体。

[0071] 11. 一种用第10项所述的方法设计改造蛋白酶的方法,所述方法包括利用第10项的方法将目标蛋白酶活性位点上的酪氨酸取代为硒代酪氨酸,由此得到含有硒代酪氨酸的目标蛋白酶突变体,从而显著提高酶活性。

[0072] 12. 由第11项所述的方法获得的含有硒代酪氨酸的目标蛋白酶突变体,包括但不限于,放射形土壤杆菌 (*Agrobacterium radiobacter*) 磷酸三酯酶,其氨基酸序列如SEQ ID NO:8所示。

[0073] II、有益效果

[0074] 本研究通过筛选得到了一种正交氨酰基-tRNA合成酶,并且,在此基础上研发了硒代酪氨酸翻译系统。通过该系统可以在目标蛋白酶——例如,但不限于,arPTE中定点特异性插入硒代酪氨酸(SeHF),产生309位(对应SEQ ID NO:8中的第285位氨基酸)含有SeHF的arPTE突变体,较野生型arPTE相比,其酶活性显著提高, $K_{cat}$ 增加12倍,pH7.0时 $k_{cat}/K_m$ 增加3.2倍。通过MD模拟,本研究揭示了arPTE突变体酶活性增强机制,即打开产物释放通道的氢键转换机制。

[0075] 由于SeHF和酪氨酸只有一个单独原子不同,因此SeHF在酶活性位置替代酪氨酸仅产生了最小的结构变化和空间位阻。重要的是:1、受益于SeHF的高亲核性,含有酪氨酸亲核试剂的酶可能通过SeHF的掺入而表现出较高的活性。2、作为广义的酸/碱,酪氨酸的 $pK_a$ 必须通过周围的残基来调节,以促进在接近中性pH条件下的质子提取/添加,而拥有低 $pK_a$ (5.9)的SeHF可以在中性pH条件下作为广义酸/碱,不需要来自周围残基的帮助,这种独特的性质可以显著简化合理的计算酶设计。3、由于SeHF在中性pH下携带一个负电荷,所以它可以表现出较强的阳离子- $\pi$ 相互作用,用于稳定酶的转换状态。

[0076] 本研究结果表明,SeHF的掺入可以用作提高酶催化效率的有力工具,并作为研究酶活性位点中酪氨酸功能的探针。另外,如硒配体的引入可以显著调节金属酶的氧化还原电位和催化性能,那么SeHF在金属酶的设计中也是非常有用的。

[0077] 上述方法同样可以用来设计和探索一些活性位点存在酪氨酸残基的蛋白酶,比如:酪氨酸苯酚裂解酶、类固醇异构酶、萜类环化酶、核糖核苷酸还原酶、FtmOx1和半乳糖氧化酶等。

## 附图说明

[0078] 从下面结合附图的详细描述中,本发明的上述特征和优点将更明显,其中:

[0079] 图1是硒代酪氨酸(SeHF)的化学合成路线图;

[0080] 图2:(A)是含有硒代酪氨酸的放射形土壤杆菌磷酸三酯酶突变体(SeHF-arPTE)的SDS-PAGE电泳图;(B)是野生型和309SeHF突变体PTE酶活曲线对比图;

[0081] 图3是胰蛋白酶消化的PTE309SeHF质谱图;

[0082] 图4:(A)是野生型(WT)和309SeHF突变体PTE在室温下水解甲基对氧磷的pH-速率曲线( $\log k_{cat}$ 对pH);(B)是野生型和309SeHF突变体PTE在室温下水解甲基对氧磷的pH-速率曲线( $\log k_{cat}/K_m$ 对pH);(C)是在室温及中性条件(pH7.0、pH7.5)下,野生型和309SeHF突变体PTE水解甲基对氧磷的反应速率图;

[0083] 图5是正交tRNA、野生型酪氨酰tRNA合成酶、本发明的正交氨酰基-tRNA合成酶和放射形土壤杆菌磷酸三酯酶突变体的序列。

[0084] 序列说明

[0085]

序列编号	说明
SEQ ID NO:1	正交 tRNA 序列
SEQ ID NO:2	野生型酪氨酰 tRNA 合成酶 ( <i>MjYRS</i> ) 的氨基酸序列, 来源于詹氏甲烷球菌 ( <i>Methanococcus jannaschii</i> )
SEQ ID NO:3	本发明的正交氨酰基-tRNA 合成酶 (SeHF <sub>RS</sub> ) 的编码核苷酸序列
SEQ ID NO:4	本发明的正交氨酰基-tRNA 合成酶 (SeHF <sub>RS</sub> ) 的氨基酸序列
SEQ ID NO:5	用于本发明的野生型放射形土壤杆菌磷酸三酯酶 ( <i>arPTE</i> ) 的编码核苷酸序列, 其中不包含全长野生型第 1-24 位氨基酸的编码序列
SEQ ID NO:6	本发明得到的野生型放射形土壤杆菌磷酸三酯酶 ( <i>arPTE</i> ) 的氨基酸序列 (对应全长野生型第 25-365 位氨基酸), 其中第 1 位的 M 为引入的起始氨基酸, 最后的 6 个 H 为用于纯化的组氨酸 (His) 标签, 6 个 H 之前的 LE 为蛋白和 His 标签之间的连接点
SEQ ID NO:7	用于本发明的含有 309 位 SeHF 的放射形土壤杆菌磷酸三酯酶突变体 (PTE309SeHF) 的编码核苷酸序列 (其中不

[0086]

	包含全长野生型第 1-24 位氨基酸的编码序列)
SEQ ID NO:8	本发明得到的含有 309 位 SeHF 的放射形土壤杆菌磷酸三酯酶突变体 (PTE309SeHF) 的氨基酸序列, 其中第 285 位的 Y* (对应全长野生型 <i>arPTE</i> 的第 309 位) 表示引入的硒代酪氨酸

### 具体实施方式

[0087] 以下通过实施例来进一步阐明本发明。但是应该理解, 所述实施例只是举例说明的目的, 并不意欲限制本发明的范围和精神。

[0088] 本领域技术人员应该理解, 除非特别说明, 下述实施例中所用的化学试剂均为可通过商业途径购得的分析纯级别的试剂。

[0089] 实施例1: 硒代酪氨酸 (SeHF) 的化学合成 (图1)

[0090] 步骤一:合成2-乙酰氨基-2-(4-硝基苄基)丙二酸二乙酯(化合物2)

[0091] 将1-(溴甲基)-4-硝基苯(1.0g,1当量)、2-乙酰氨基丙二酸二乙酯(1.2g,1.2当量)和tBuOK(0.63g,1.2当量)的混合物放置于含有20mL EtOH的圆底烧瓶中,通过TLC监测回流过夜。反应完成后,将混合物冷却至室温,过滤并用EtOH洗涤。使用乙酸乙酯/己烷(1:1混合)作为洗脱液,将残余物加到硅胶的快速色谱上,得到白色固体的目标化合物(1.23g,75.5%收率)。

[0092]  $^1\text{H NMR}$  (500MHz,  $\text{CDCl}_3$ )  $\delta$  8.12 (d, 2H), 7.18 (d, 2H), 6.56 (s, 1H), 4.28 (q, 4H), 3.77 (s, 2H), 2.04 (s, 3H), 1.3 (t, 6H).  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ )  $\delta$  169.40, 167.05, 147.27, 143.15, 130.72, 123.46, 66.86, 63.02, 37.59, 23.02, 14.01。

[0093] 步骤二:合成2-乙酰氨基-2-(4-氨基苄基)丙二酸二乙酯(化合物3)

[0094] 在TLC监测下,将15mL化合物2(1.0g,1当量)和Pd/C(0.6g,0.1当量)的MeOH混合物在氢气中(4atm)搅拌,反应完成后,过滤混合物,减压浓缩,并使用乙酸乙酯/己烷(1:1混合)作为洗脱液进行硅胶快速色谱分离,得到白色固体的目标化合物(1.0g,100%产量)。

[0095]  $^1\text{H NMR}$  (500MHz,  $\text{CDCl}_3$ )  $\delta$  6.78 (d, 2H), 6.57 (d, 2H), 4.25 (q, 4H), 3.52 (s, 2H), 2.01 (s, 3H), 1.28 (t, 6H).  $^{13}\text{C NMR}$  ( $\text{CDCl}_3$ )  $\delta$  168.94, 167.67, 145.50, 130.67, 124.78, 115.01, 67.39, 62.50, 37.07, 23.04, 14.02。

[0096] 步骤三:合成2-乙酰氨基-2-(4-硒代氰基苄基)丙二酸二乙酯(化合物5)

[0097] 将化合物3(1.0g)和6N HCl(5mL)放置于圆底烧瓶,冷却至0 $^{\circ}\text{C}$ ,滴加NaNO<sub>2</sub>溶液(0.25g,1.25当量,在5mL H<sub>2</sub>O中)(注意:该步骤可能产生有毒的氮-氧化合物并存在爆炸危险,必须采取必要的保护程序)。5分钟后,用Na<sub>2</sub>CO<sub>3</sub>将pH值调节至6。将混合物保持在0 $^{\circ}\text{C}$ 下,并滴加KSeCN(0.5g,1.1当量,在5mL H<sub>2</sub>O中)。然后将混合物温热至室温并继续反应1小时,再将50mL水倒入反应混合物中。水层用二氯甲烷萃取3次,有机层经无水Na<sub>2</sub>SO<sub>4</sub>干燥,过滤,浓缩并用硅胶色谱分离,使用乙酸乙酯/己烷(1:2混合)作为洗脱液,得到目标化合物(0.55g,40%收率)。 $^1\text{H NMR}$  (500MHz,  $\text{CDCl}_3$ )  $\delta$  7.54 (d, 2H), 7.05 (d, 2H), 6.55 (s, 1H), 4.33-4.22 (m, 4H), 3.69-3.66 (m, 2H), 2.04 (s, 3H), 1.3 (m, 6H)。

[0098] 步骤四:合成2,2'-((亚烷基二(4,1-亚苯基))双(亚甲基))二(2-乙酰氨基丙二酸酯)(化合物6)

[0099] 将化合物5(0.5g,1当量)溶解于三颈圆底烧瓶中氮气保护的5mL EtOH中,通过注射器注入KOH溶液(0.34g,5当量),随后在室温下搅拌1小时。然后加入10mL水,过滤沉淀的固体产物,洗涤并干燥,得到目标化合物(0.42g,90%收率)(注意:该步骤产生有毒的KCN化合物,必须采取必需的保护和后处理程序)。

[0100] 步骤五:合成3,3'-((亚烷基二(4,1-亚苯基))二(2-氨基丙酸)(化合物7,即硒代酪氨酸)

[0101] 向圆底烧瓶中加入化合物6(0.42g)。加入浓HCl(10mL)并将混合物搅拌、回流10小时,通过TLC监测。反应完成后,将溶剂减压浓缩一半。混合物在4 $^{\circ}\text{C}$ 冰箱中静置过夜。过滤沉淀出的产物并干燥,得到目标化合物(0.19g,70%收率)。 $^1\text{H NMR}$  (500MHz,  $\text{D}_2\text{O}$ )  $\delta$  7.55 (d, 2H), 7.26 (d, 2H), 4.10 (t, 1H), 3.09 (m, 2H)。质谱: $m/z$  489, 487, 485,  $[\text{M}+\text{H}]^+$ 。

[0102] 以上合成反应所需化学试剂如无特别说明,均购自百灵威科技有限公司、西格玛奥德里奇贸易有限公司或北京化工厂,均为分析纯以上级别。使用Bruker AMX-600仪器记

录<sup>1</sup>H NMR光谱,使用TMS作为内标(TMS,0.00),以ppm为单位报告化学位移。多重性报道如下:s=单峰,d=双峰,t=三重峰,q=四重峰,m=多重峰。在75.4MHz记录<sup>13</sup>C NMR光谱,并且使用氘代溶剂作为内部标准(CDC1<sub>3</sub>,77.0),以ppm为单位报告化学位移。

[0103] 实施例2:进化SeHF特异性氨酰基-tRNA合成酶

[0104] 为了在基因中位点特异性插入SeHF,需要在所用的E.coli宿主细胞中引入氨酰基-tRNA合成酶/tRNA正交对,这个正交对来源于詹氏甲烷球菌(Methanococcus jannaschii)琥珀抑制酪氨酰tRNA(MjtRNA<sup>Tyr</sup>CUA)/酪氨酰tRNA合成酶(MjYRS,野生型,其氨基酸序列为SEQ ID NO:2)对。MjYRS突变库构建在卡纳霉素抗性pBK质粒(购自美国scripps研究所Peter G.Schultz实验室)中,位于该质粒上E.coli谷氨酰胺合成酶的启动子和终止子之间。所使用的合成酶突变库为pBk-lib-iw1库,该突变库的构建方法为:在MjYRS基因上挑选6个位点(Tyr32,Leu65,Phe108,Gln109,Asp158,和Leu162)引入NNK突变(N=A+T+C+G;K=T+G),另外6个位点(Ile63,Ala67,His70,Tyr114,Ile159,Val164)或随机突变为Gly或保持不变(参见Xie,J.;Liu,W.S.;Schultz,P.G.Angew.Chem.,Int.Ed.2007,46,9239-9242;Wang,JY.;Zhang W.;Song WJ;et al.J.Am.Chem.Soc.2010,132,14812-14818)。

[0105] 通过正负筛选来进化特异性识别SeHF的氨酰基-tRNA合成酶。正筛选质粒包含MjtRNA<sup>Tyr</sup>CUA,TAG突变的氯霉素乙酰转移酶基因,启动表达绿色荧光蛋白的琥珀突变的T7RNA聚合酶,四环素抗性基因。负筛选质粒包含MjtRNA<sup>Tyr</sup>CUA,在阿拉伯糖操纵子下的琥珀突变芽孢杆菌RNA酶基因,以及氨苄青霉素抗性基因。进行3轮正负筛选:包含有正筛选质粒的E.coli DH10B细胞作为正筛选寄主细胞。细胞电转pbk-lib-jw1库,SOC培养基(2% (W/V) 胰蛋白胨,0.5% (W/V) 酵母粉,0.05% (W/V) NaCl,2.5mM KCl,10mM MgCl<sub>2</sub>,20mM葡萄糖)在37℃培养1小时。之后换用极限培养基(GMML极限培养基的配方:M9盐/甘油:764g Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O或者30g Na<sub>2</sub>HPO<sub>4</sub>,15g KH<sub>2</sub>PO<sub>4</sub>,2.5g NaCl,5g NH<sub>4</sub>Cl,50ml甘油,高压灭菌,pH7.0;1M MgSO<sub>4</sub>:高压灭菌;50mM CaCl<sub>2</sub>:高压灭菌;25mM FeCl<sub>2</sub>:过滤灭菌;0.3M亮氨酸:溶解于0.3M NaOH中,过滤灭菌;1L液体GMML培养基:200ml M9盐/甘油,2ml MgSO<sub>4</sub>,2ml CaCl<sub>2</sub>,2ml FeCl<sub>2</sub>,1ml亮氨酸)洗两次,铺板固体极限培养基(在液体GMML培养基中加入500ml 3%琼脂粉,1mM SeHF,50mg/L卡那霉素,60mg/L氯霉素,15mg/L四环素),37℃培养60小时。收取细胞,提取质粒DNA,电泳分离,胶回收。然后,将经过正筛选的pBK-lib-jw1转化到包含负筛选质粒的DH10B感受态细胞中。SOC培养基中恢复1小时。之后涂板包含0.2%阿拉伯糖(购自sigma公司)的LB固体培养基(每升培养基含10g胰蛋白胨,5g酵母粉,10g NaCl)。37℃培养8-12小时。共重复3轮。

[0106] 最后一轮正筛选挑384个克隆,分别点板在含有1mM SeHF、氯霉素60,80,120,160μg/mL的GMML固体培养基上,及不包含SeHF、但包含氯霉素0,20,40,60μg/mL的GMML固体培养基。挑选在在1mM SeHF160μg/mL氯霉素的培养基上生长,而在0mM SeHF,浓度大于20μg/mL氯霉素培养基中不生长的克隆进行进一步验证。最终挑出1个克隆,插入硒代酪氨酸效率最高,测序表明,克隆所包含的氨酰基-tRNA合成酶突变体(SeHFRS)的氨基酸序列为SEQ ID NO:4所示,其中的突变位点为:Tyr32Leu,Leu65Ser,Asp87Asn,Gly97Arg,Tyr114Ser,Asp158Gly和Leu162Thr。

[0107] 本领域技术人员应该理解,在本发明中,除了SEQ ID NO:4所示的氨基酸序列之外,术语“正交氨酰基-tRNA合成酶”或“正交硒代酪氨酸氨酰基-tRNA合成酶”还包括SEQ ID

NO:4所示氨基酸序列的保守性变体,只要所述保守性变体具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性即可;并且还包含将SEQ ID NO:4所示的氨基酸序列经过一个或多个氨基酸的取代、缺失或添加,且具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性的由SEQ ID NO:4所示的氨基酸序列衍生的氨基酸序列;并且还包含与SEQ ID NO:4所示的氨基酸序列具有一定的序列同源性的同源物,其中所述序列同源性可以为80%以上,优选85%以上,更优选90%以上,甚至更优选95%以上,甚至更优选99%以上,并且所述同源物具有与SEQ ID NO:4所示的氨基酸序列相同的酶活性。

[0108] 实施例3:表达SeHF-arPTE及质谱鉴定

[0109] 为了测定SeHF掺入蛋白质的效率和保真度,我们在C-末端含有His6标签的arPTE中用琥珀终止密码子取代Tyr309。在SeHFRS、MjtRNA<sup>Tyr</sup><sub>CUA</sub>和0.5mM SeHF共同存在的条件下,通过大肠杆菌表达arPTE,同时以不添加SeHF作为阴性对照。

[0110] 具体步骤为:将正交tRNA (SEQ ID NO:1) 和筛选出来的编码SeHFRS的核苷酸序列 (SEQ ID NO:3) 构建到pEVOL载体 (购自美国scripps研究所Peter G.Schultz实验室) 上;编码arPTE的核苷酸序列 (SEQ ID NO:5) 构建到pET-22b载体 (购自novagen公司) 上,利用TransStartFastPfu (购自全式金公司) DNA聚合酶及引物设计,经PCR在arPTE-pET22b中引入Y309TAG突变。两种质粒共转化到BL21 (DE3) 细胞 (购自全式金公司) 中。挑取单克隆在37℃培养到OD<sub>600</sub>约等于0.6时,将温度降至30℃,并向TB培养基 (1.2%胰蛋白胨,2.4%酵母提取物,72mM K<sub>2</sub>HPO<sub>4</sub>,17mM KH<sub>2</sub>PO<sub>4</sub>,0.4%甘油) 中加入0.2mM IPTG、0.2%阿拉伯糖、0.2mM CoCl<sub>2</sub>及0.5mM SeHF培养细胞,对照不加入SeHF。16小时之后,收菌,先后用Ni-NTA (购自南京金斯瑞公司) 和Superdex 20010/300凝胶过滤柱 (购自GE Healthcare公司) 纯化蛋白,最后进行SDS-PAGE电泳分析和LC/MS分析。

[0111] SDS-PAGE显示,只有在存在SeHF的培养基中才能纯化出全长的arPTE309SeHF突变体 (以下简称为PTE309SeHF) (图2A),这说明筛选出来的SeHFRS可以特异性的识别SeHF。

[0112] LC/MS图谱显示,胰蛋白酶消化的PTE309SeHF显示全长蛋白质的覆盖率为93.5%,其中包含309SeHF的肽能够被很好地鉴定 (图3),表明我们成功获得了含有SeHF309的全长arPTE蛋白。

[0113] 测序表明,野生型arPTE的氨基酸序列如SEQ ID NO:6所示 (对应全长野生型的第25-365位氨基酸序列),arPTE309SeHF突变体的氨基酸序列如SEQ ID NO:8所示,其中第285位 (即,对应全长野生型的第309位) 氨基酸被硒代酪氨酸替换 (图5)。

[0114] 实施例4:PTE酶活性测定

[0115] 为了确定SeHF是否可用于增强酶活性,我们测定了野生型 (简称为WT) 和309SeHF突变体PTE的催化活性。由于SeHF和酪氨酸结构相似,所以插入的309SeHF应该不会显著影响PTE结构。

[0116] 室温下,当底物甲基对氧磷 (购自Sigma,CAS#:950-35-6) 浓度从12.5μM逐渐增加到1mM时,我们通过监测产物4-硝基苯酚在347nm ( $\epsilon_{347} = 5176\text{M}^{-1}\text{cm}^{-1}$ ) 处的UV-Vis光谱来测定PTE活性。然后将数据拟合到米氏方程中来计算动力学常数 $k_{\text{cat}}$ 和 $K_{\text{m}}$ ,分别绘制WT和309SeHF突变体PTE的 $\log k_{\text{cat}}$ 对pH曲线 (pH-速率曲线,图4A)。当pH为4.7时,WT和309SeHF突变体PTE具有相同的 $\text{p}K_{\text{e}}/\text{p}K_{\text{es}}$ 值。然而,当pH为6.0时,仅有PTE309SeHF突变体显示出第二个 $\text{p}K_{\text{es}}$ 。根据曲线 (图4B),WT PTE的pH-速率曲线在pH为5.0时到达平台期,而309SeHF突变体的

pH-速率曲线在pH为6.5时到达平台期,反映了SeHF侧链的去质子化。

[0117] 为了进一步验证SeHF去质子化和提高酶活性之间的关系,我们比较了WT和309SeHF突变体PTE在接近中性pH (pH7.0-pH7.5) 条件下的反应速度。如图4C所示,309SeHF突变体展现出明显加速的酶反应速率。当pH为7.0时,309SeHF突变体具有比WT PTE高12倍的 $k_{cat}$ 值,这表明309SeHF突变体与WT PTE相比具有更快的产物释放速率,同时证明了309SeHF参与产物释放这一关键限速步骤。 $K_m$ 增加表示 $k_1$ 降低,或者形成米氏复合物(ES)的结合效率较低,进一步证明了309SeHF参与米氏复合物的形成。

[0118] 总体而言,309SeHF突变导致PTE在pH为7时的 $k_{cat}/K_m$ 增加3.2倍。虽然SeHF的负电荷和增加的范德华半径导致底物结合亲和力降低,但是SeHF能够显著地加快产物释放步骤,最终使得PTE在中性pH条件下酶活性大大提高。

[0119] 应该理解,尽管参考其示例性的实施方案,已经对本发明进行具体地显示和描述,但是本领域的普通技术人员应该理解,在不背离由权利要求书所定义的本发明的精神和范围的条件下,可以在其中进行各种形式和细节的变化,可以进行各种实施方案的任意组合。

IB187163序列表

序列表

- <110> 中国科学院生物物理研究所
- <120> 硒代酪氨酸翻译系统及其应用
- <130> IB187163
- <160> 8
- <170> PatentIn version 3.1
- <210> 1
- <211> 77
- <212> RNA
- <213> 人工序列(Artificial Sequence)

<400> 1  
 ccggcgguag uucagcaggg cagaacggcg gacucuaau ccgcauggcg cugguacaaa 60  
 uccggccccg cggacca 77

- <210> 2
- <211> 306
- <212> PRT
- <213> Methanococcus jannaschii

[0001]

<400> 2  
 Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser  
 1 5 10 15  
 Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Tyr  
 20 25 30  
 Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln  
 35 40 45  
 Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile  
 50 55 60  
 Leu Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp  
 65 70 75 80  
 Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met  
 85 90 95  
 Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu Phe Gln Leu Asp Lys  
 100 105 110  
 Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys  
 115 120 125  
 Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro  
 130 135 140

Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Asp Ile His  
 145 150 155 160

Tyr Leu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile  
 165 170 175

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His  
 180 185 190

Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser  
 195 200 205

Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala  
 210 215 220

Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro  
 225 230 235 240

Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys  
 245 250 255

Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu  
 260 265 270

Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Asp Leu Lys  
 275 280 285

[0002] Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys  
 290 295 300

Arg Leu  
 305

<210> 3

<211> 921

<212> DNA

<213> 人工序列(Artificial Sequence)

<400> 3

atggacgaat ttgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60

agagaggttt taaaaaaga tgaaaaatct gctctgatag gttttgaacc aagtggtaaa 120

atacatttag ggcattatct ccaaataaaa aagatgattg atttcaaaa tgctggattt 180

gatataatta taagtttggc tgatttacet gcctatttaa accagaaagg agagttggat 240

gagattagaa aaataggaaa ttatacaaaa aaagtttttg aagcaatgag gttaaaggca 300

aaatattgttt atggaagtga attccagctt gataaggata gcacactgaa tgtctataga 360

ttggctttaa aaactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420

gatgaaaatc caaaggttgc tgaagttatc tatccaataa tgcagggttaa tgggattcat 480

tatactggcg ttgatgttgc agttggaggg atggagcaga gaaaaataca catgttagca 540

agggagcttt taccaaaaaa ggttgtttgt attcacaacc ctgtcttaac gggtttggat 600

ggagaaggaa agatgagttc ttcaaaaggg aatattatag ctgittgatga ctctccagaa 660  
gagattaggg ctaagataaa gaaagcatac tgcccagctg gagttgttga aggaaatcca 720  
ataatggaga tagctaaata cttccttgaa tatectttaa ccataaaaag gccagaaaaa 780  
tttggtggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag 840  
gaattgcatc caatggattt aaaaaatgct gtagctgaag aactataaaa gattttagag 900  
ccaattagaa agagattata a 921

<210> 4

<211> 306

<212> PRT

<213> 人工序列(Artificial Sequence)

<400> 4

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser  
1 5 10 15

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Leu  
20 25 30

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln  
35 40 45

[0003] Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile  
50 55 60

Ser Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp  
65 70 75 80

Glu Ile Arg Lys Ile Gly Asn Tyr Asn Lys Lys Val Phe Glu Ala Met  
85 90 95

Arg Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu Phe Gln Leu Asp Lys  
100 105 110

Asp Ser Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys  
115 120 125

Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro  
130 135 140

Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His  
145 150 155 160

Tyr Thr Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile  
165 170 175

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His  
180 185 190

Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser  
195 200 205

Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala  
 210 215 220

Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro  
 225 230 235 240

Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys  
 245 250 255

Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu  
 260 265 270

Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Asp Leu Lys  
 275 280 285

Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys  
 290 295 300

Arg Leu  
 305

- <210> 5
- <211> 1010
- <212> DNA
- <213> Agrobacterium radiobacter

[0004]

<400> 5  
 ataccgttcg cgggtccgatt cgggtgagcg aagcaggctt taccctgacc catgaacata 60  
 tttgtggctc tagcgcaggc tttctgcgtg catggccgga atttttcggg agccgcaaag 120  
 cactggcaga aaaagcagtt cgcggcctgc gtcatgcgcg aagcgcaggt gttcagacca 180  
 ttgtggatgt gagtaccitt gatattggtc gtgatgttcg tctgctggca gaagttagtc 240  
 gcgcagcaga ttttcatatt gttgcagcaa cgggcctgtg gtttgatccg ccgctgtcaa 300  
 tgcgtatgcg tagtgtggaa gaactgacce agttctttct gcgcgaaatt cagcatggta 360  
 ttgaagatac cggatttcgc gcaggattta ttaaagtgc aaccaccggt aaagcaacce 420  
 cgtttcagga actgggtgctg aaagcagcag ctagggcatac tctggcaacc ggtgtgccgg 480  
 tgaccacca taccagtgca tcacagcgcg atggegaaaca gcagcagca atttttgaat 540  
 cagaaggtct gageccgagt cgtgtgtgta ttggtcatag cgatgatacc gatgatctgt 600  
 cttatctgac cggectggca gcacgcgct atctgggtggg cctggatcgt atgcegtata 660  
 gcgcaattgg tctggaaggt aatgcaagt cactggcact gtttgggacg cgctcttggc 720  
 agaccgcgc actgctgatt aaagcactga ttgatcgcgg ctacaaagat cgcattctgg 780  
 ttagtcatga ttgctgttt ggcttttagta gctatgttac caatattatg gatgttatgg 840  
 atgcattaa tccggatggt atggcatttg tcccgctcgc tgttattccg tttctgcgtg 900  
 aaaaaggcgt tccgccgaa accctggcag gcgttacctg tgcaaatccg gcacgcttcc 960  
 tgtctccgac cgttctgca agtctcgagc accaccacca ccaccactga 1010

<210> 6

<211> 349  
 <212> PRT  
 <213> 人工序列(Artificial Sequence)

<400> 6  
 Met Ser Met Ala Arg Pro Ile Gly Thr Gly Asp Leu Ile Asn Thr Val  
 1 5 10 15  
 Arg Gly Pro Ile Pro Val Ser Glu Ala Gly Phe Thr Leu Thr His Glu  
 20 25 30  
 His Ile Cys Gly Ser Ser Ala Gly Phe Leu Arg Ala Trp Pro Glu Phe  
 35 40 45  
 Phe Gly Ser Arg Lys Ala Leu Ala Glu Lys Ala Val Arg Gly Leu Arg  
 50 55 60  
 His Ala Arg Ser Ala Gly Val Gln Thr Ile Val Asp Val Ser Thr Phe  
 65 70 75 80  
 Asp Ile Gly Arg Asp Val Arg Leu Leu Ala Glu Val Ser Arg Ala Ala  
 85 90 95  
 Asp Val His Ile Val Ala Ala Thr Gly Leu Trp Phe Asp Pro Pro Leu  
 100 105 110  
 Ser Met Arg Met Arg Ser Val Glu Glu Leu Thr Gln Phe Phe Leu Arg  
 115 120 125  
 Glu Ile Gln His Gly Ile Glu Asp Thr Gly Ile Arg Ala Gly Ile Ile  
 130 135 140  
 Lys Val Ala Thr Thr Gly Lys Ala Thr Pro Phe Gln Glu Leu Val Leu  
 145 150 155 160  
 Lys Ala Ala Ala Arg Ala Ser Leu Ala Thr Gly Val Pro Val Thr Thr  
 165 170 175  
 His Thr Ser Ala Ser Gln Arg Asp Gly Glu Gln Gln Ala Ala Ile Phe  
 180 185 190  
 Glu Ser Glu Gly Leu Ser Pro Ser Arg Val Cys Ile Gly His Ser Asp  
 195 200 205  
 Asp Thr Asp Asp Leu Ser Tyr Leu Thr Gly Leu Ala Ala Arg Gly Tyr  
 210 215 220  
 Leu Val Gly Leu Asp Arg Met Pro Tyr Ser Ala Ile Gly Leu Glu Gly  
 225 230 235 240  
 Asn Ala Ser Ala Leu Ala Leu Phe Gly Thr Arg Ser Trp Gln Thr Arg  
 245 250 255  
 Ala Leu Leu Ile Lys Ala Leu Ile Asp Arg Gly Tyr Lys Asp Arg Ile

[0005]

	260		265		270
Leu Val Ser	His Asp Trp	Leu Phe Gly	Phe Ser Ser	Tyr Val Thr	Asn
	275		280		285
Ile Met Asp	Val Met Asp	Arg Ile Asn	Pro Asp Gly	Met Ala Phe	Val
	290		295		300
Pro Leu Arg	Val Ile Pro	Phe Leu Arg	Glu Lys Gly	Val Pro Pro	Glu
	305		310		320
Thr Leu Ala	Gly Val Thr	Val Ala Asn	Pro Ala Arg	Phe Leu Ser	Pro
		325		330	335
Thr Val Arg	Ala Ser Leu	Glu His His	His His His	His His	
	340		345		

<210> 7  
 <211> 1050  
 <212> DNA  
 <213> 人工序列(Artificial Sequence)

[0006]

<400> 7  
 atgtcaatgg cacgcccgat tggaacaggc gatctgatta ataccgttcg cggtcagatt 60  
 ccggtgagcg aagcaggctt taccctgacc catgaacata tttgtggctc tagcgcagcg 120  
 tttctgcgtg catggccgga atttttcggt agccgcaaag cactggcaga aaaagcagtt 180  
 cgcggcctgc gtcatgcgcg aagcgcaggt gttcagacca ttgtggatgt gagtaccttt 240  
 gatattggtc gtgatgttcg tctgctggca gaagttagtc gcgcagcaga tgttcatatt 300  
 gttgcagcaa ccggcctgtg gtttgatccg ccgctgtcaa tgcgtatgcg tagtgtggaa 360  
 gaactgacce agttctttct gcgcgaaatt cagcatggta ttgaagatac cggatttcgc 420  
 gcaggtatta ttaaagttgc aaccaccggt aaagcaacc cgtttcagga actggtgctg 480  
 aaagcagcag ctagggcacc tctggcaacc ggtgtgcccg tgaccacca taccagtga 540  
 tcacagcgeg atggcgaaca gcaggcagca attttgaat cagaaggtct gagcccaggt 600  
 cgtgtgtgta ttggtcatag cgatgatacc gatgatctgt cttatctgac cggcctggca 660  
 gcacgcggct atctggiggg cctggatcgt atgccgtata gcgcaattgg tctggaaggt 720  
 aatgcaagtg cactggcact gtttgggacg cgctcttggc agacccegc actgctgatt 780  
 aaagcactga ttgatecggg ctacaagat cgcattctgg ttagtcatga ttgctgttt 840  
 ggcttttagta gctaggttac caatattatg gatgttatgg atcgattaa tccgatggt 900  
 atggcatttg tgccgctgcg tgttattccg tttctgcgtg aaaaaggcgt tccgccggaa 960  
 accctggcag gcgttaccgt tgcaaatccg gcacgcttcc tgtctccgac cgttctgca 1020  
 agtctcgagc accaccacca ccaccactga 1050

<210> 8  
 <211> 349  
 <212> PRT

<213> 人工序列(Artificial Sequence)

<220>

<221> VARIANT

<223> 第285位的酪氨酸被硒代酪氨酸置换

<400> 8

Met Ser Met Ala Arg Pro Ile Gly Thr Gly Asp Leu Ile Asn Thr Val  
1 5 10 15

Arg Gly Pro Ile Pro Val Ser Glu Ala Gly Phe Thr Leu Thr His Glu  
20 25 30

His Ile Cys Gly Ser Ser Ala Gly Phe Leu Arg Ala Trp Pro Glu Phe  
35 40 45

Phe Gly Ser Arg Lys Ala Leu Ala Glu Lys Ala Val Arg Gly Leu Arg  
50 55 60

His Ala Arg Ser Ala Gly Val Gln Thr Ile Val Asp Val Ser Thr Phe  
65 70 75 80

Asp Ile Gly Arg Asp Val Arg Leu Leu Ala Glu Val Ser Arg Ala Ala  
85 90 95

Asp Val His Ile Val Ala Ala Thr Gly Leu Trp Phe Asp Pro Pro Leu  
100 105 110

[0007]

Ser Met Arg Met Arg Ser Val Glu Glu Leu Thr Gln Phe Phe Leu Arg  
115 120 125

Glu Ile Gln His Gly Ile Glu Asp Thr Gly Ile Arg Ala Gly Ile Ile  
130 135 140

Lys Val Ala Thr Thr Gly Lys Ala Thr Pro Phe Gln Glu Leu Val Leu  
145 150 155 160

Lys Ala Ala Ala Arg Ala Ser Leu Ala Thr Gly Val Pro Val Thr Thr  
165 170 175

His Thr Ser Ala Ser Gln Arg Asp Gly Glu Gln Gln Ala Ala Ile Phe  
180 185 190

Glu Ser Glu Gly Leu Ser Pro Ser Arg Val Cys Ile Gly His Ser Asp  
195 200 205

Asp Thr Asp Asp Leu Ser Tyr Leu Thr Gly Leu Ala Ala Arg Gly Tyr  
210 215 220

Leu Val Gly Leu Asp Arg Met Pro Tyr Ser Ala Ile Gly Leu Glu Gly  
225 230 235 240

Asn Ala Ser Ala Leu Ala Leu Phe Gly Thr Arg Ser Trp Gln Thr Arg  
245 250 255

Ala Leu Leu Ile Lys Ala Leu Ile Asp Arg Gly Tyr Lys Asp Arg Ile  
260 265 270

Leu Val Ser His Asp Trp Leu Phe Gly Phe Ser Ser Tyr Val Thr Asn  
           275                                  280                                  285

Ile Met Asp Val Met Asp Arg Ile Asn Pro Asp Gly Met Ala Phe Val  
           290                                  295                                  300

[0008] Pro Leu Arg Val Ile Pro Phe Leu Arg Glu Lys Gly Val Pro Pro Glu  
           305                                  310                                  315                                  320

Thr Leu Ala Gly Val Thr Val Ala Asn Pro Ala Arg Phe Leu Ser Pro  
                                   325                                  330                                  335

Thr Val Arg Ala Ser Leu Glu His His His His His His  
                                   340                                  345

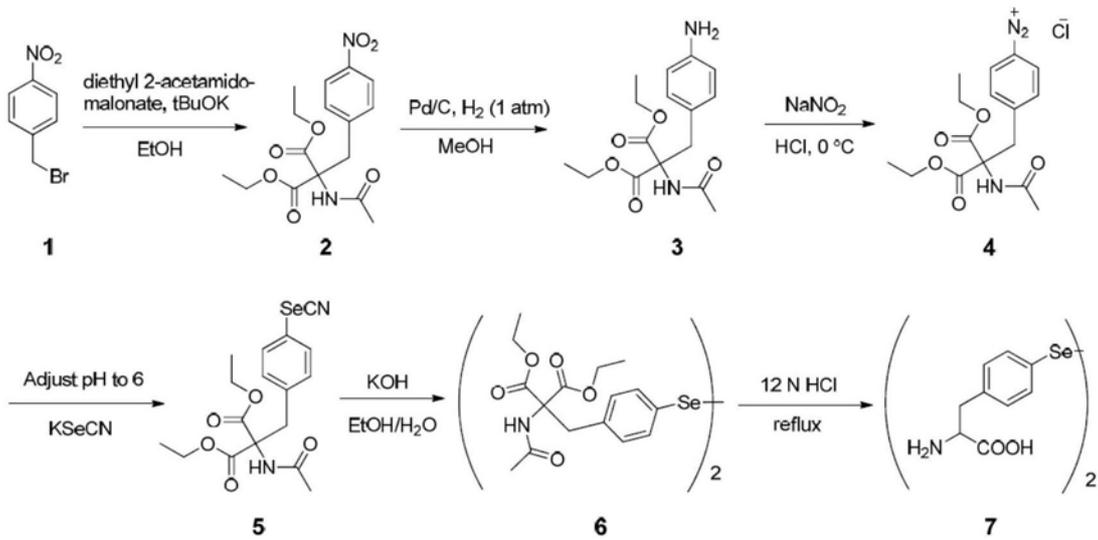


图1

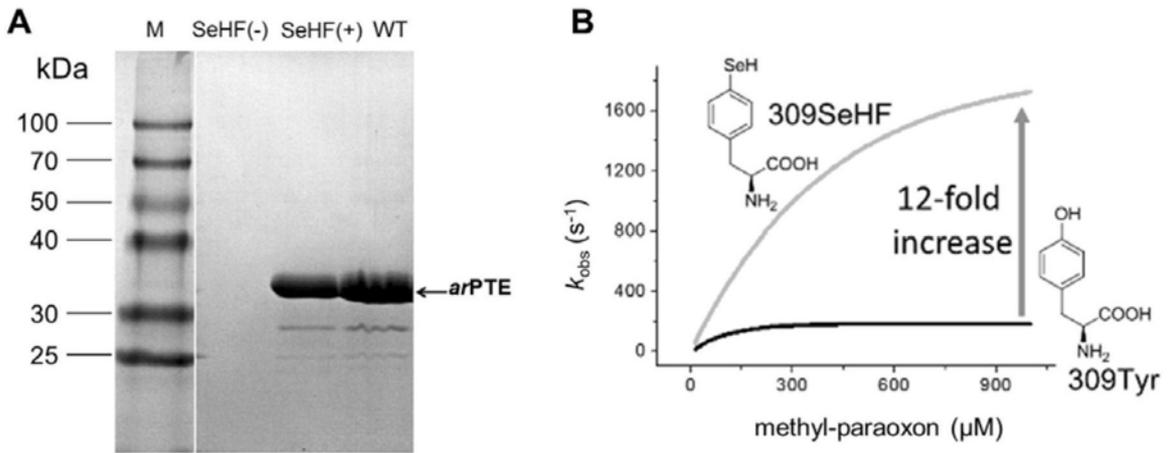


图2

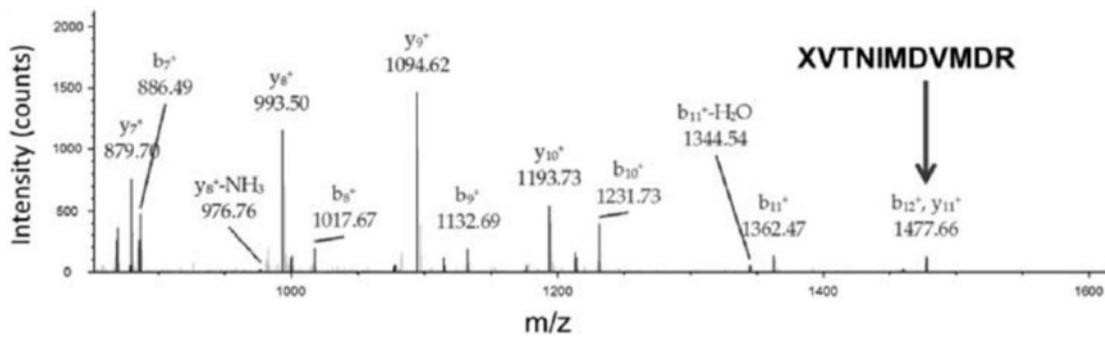


图3

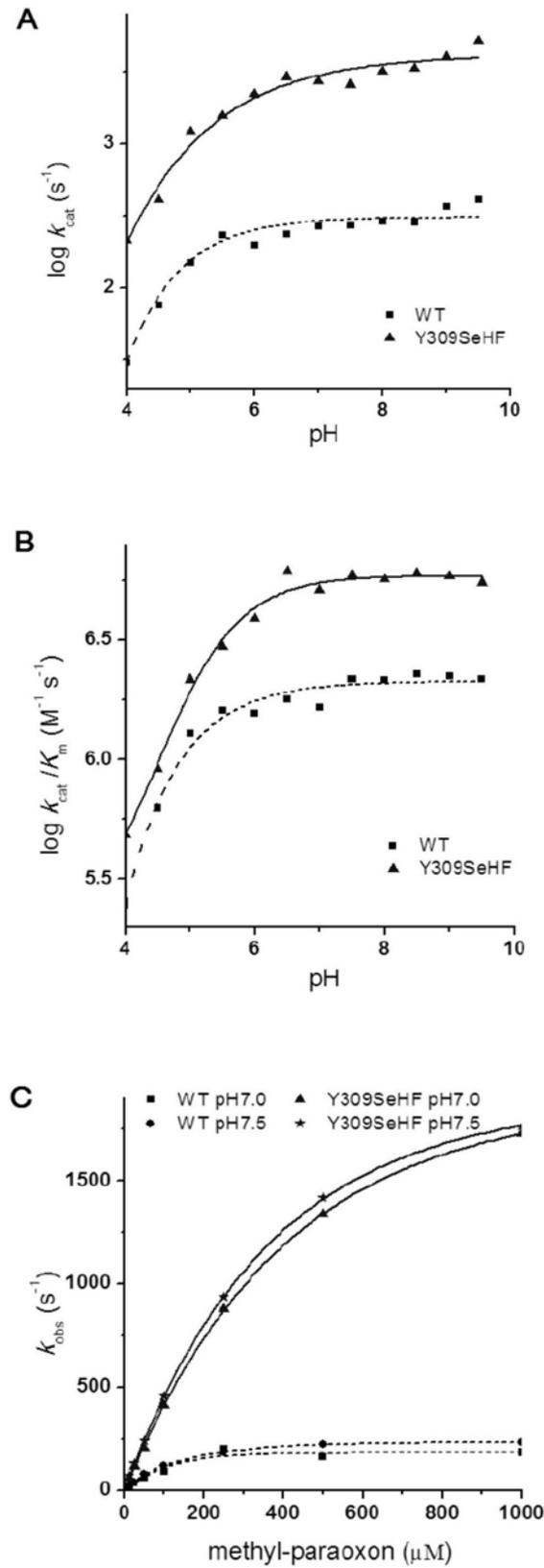


图4

名称	核苷酸/氨基酸序列
正交 tRNA	SEQ ID NO: 1 ccggcgguaguucagcagggcagaacggcgacucuaaauccgcauggcgcuuguuc aaauccggccccggacca
野生型酪氨酰 tRNA 合成酶 ( <i>MjYRS</i> ), 来源于詹氏甲烷 球 菌 ( <i>Methanococcus jannaschii</i> )	氨基酸序列 (SEQ ID NO: 2): <u>MDEFEMIKRNTSEIISEEELREVLKKDEKSAYIGFEPGKIHL</u> <u>GHYLQIKKMIDLQONAGFDIIIILLADLHAYLNQKGELDEIRKI</u> <u>GDYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALK</u> <u>TTLKRARRSMELIAREDENPKVAEVIYPIMQVNDIHYLGVD</u> <u>VAVGGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKM</u> <u>SSSKGNFIAVDDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYF</u> <u>LEYPLTIKRPEKFGGDLTVNSYEELESFLFNKELHPMDLKNA</u> <u>VAEELIKILEPIRKRL</u>
本发明的正交氨 酰基-tRNA 合成 酶 (SeHFRS)	核苷酸序列 (SEQ ID NO: 3): <u>ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAA</u> <u>TTATCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAG</u> <u>ATGAAAAATCTGCTCTGATAGGTTTTGAACCAAGTGGTAA</u> <u>AATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTG</u> <u>ATTTACAAAATGCTGGATTTGATATAATTATAAGTTTGGCT</u> <u>GATTTACATGCCTATTTAAACCAGAAAGGAGAGTTGGATG</u> <u>AGATTAGAAAAATAGGAAATTATAACAAAAAAGTTTTTGA</u> <u>AGCAATGAGGTTAAAGGCCAAAATATGTTTATGGAAGTGAA</u> <u>TCCAGCTTGATAAGGATAGCACACTGAATGTCTATAGATT</u> <u>GGCTTTAAAACTACCTTAAAAAGAGCAAGAAGGAGTAT</u> <u>GGAACCTTATAGCAAGAGAGGATGAAAATCCAAAGGTTGC</u> <u>TGAAGTTATCTATCCAATAATGCAGGTTAATGGGATTCATT</u> <u>ATACTGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGA</u> <u>GAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA</u> <u>AGGTTGTTTGTATTCACAACCCTGTCTTAACGGGTTTGGGA</u> <u>TGGAGAAGGAAAGATGAGTTCTTCAAAGGGAATTTTAT</u> <u>AGCTGTTGATGACTCTCCAGAAGAGATTAGGGCTAAGATA</u> <u>AAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAAT</u> <u>CCAATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTT</u> <u>AACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTTGAC</u> <u>AGTTAATAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATA</u> <u>AGGAATTGCATCCAATGGATTAAAAAATGCTGTAGCTGA</u> <u>AGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGATTA</u> <u>TAA</u>
野生型放射形土	氨基酸序列 (SEQ ID NO:4): <u>MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPGKIHL</u> <u>GHYLQIKKMIDLQONAGFDIIISLADLHAYLNQKGELDEIRKI</u> <u>GNYNKKVFEAMRLKAKYVYGSEFQLDKDSTLNVYRLALK</u> <u>TTLKRARRSMELIAREDENPKVAEVIYPIMQVNGIHYTGVD</u> <u>VAVGGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKM</u> <u>SSSKGNFIAVDDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYF</u> <u>LEYPLTIKRPEKFGGDLTVNSYEELESFLFNKELHPMDLKNA</u> <u>VAEELIKILEPIRKRL</u>
野生型放射形土	核苷酸序列 (SEQ ID NO: 5, 不包含第 1-24 位氨基酸的编码序

<p>壤杆菌磷酸三酯酶 (<i>ar</i>PTE)</p>	<p>列, 仅为第 25-365 位氨基酸的编码序列):  <u>ATGTCAATGGCACGCCCGATTGGAACAGGCGATCTGATTA</u>  <u>ATACCGTTCGCGGTCCGATTCCGGTGAGCGAAGCAGGCTT</u>  <u>TACCCTGACCCATGAACATATTTGTGGCTCTAGCGCAGGC</u>  <u>TTTCTGCGTGCATGGCCGGAATTTTTCGGTAGCCGCAAAG</u>  <u>CACTGGCAGAAAAAGCAGTTCGCGGCCTGCGTCATGCGC</u>  <u>GAAGCGCAGGTGTTTCAGACCATTGTGGATGTGAGTACCTT</u>  <u>TGATATTGGTCGTGATGTTTCGTCTGCTGGCAGAAGTTAGT</u>  <u>CGCGCAGCAGATGTTTCATATTGTTGCAGCAACCGGCCTGT</u>  <u>GGTTTGATCCGCCGCTGTCAATGCGTATGCGTAGTGTGGA</u>  <u>AGAACTGACCCAGTTCCTTTCTGCGCGAAATTCAGCATGGT</u>  <u>ATTGAAGATACCGGTATTCGCGCAGGTATTATTAAGTTGC</u>  <u>AACCACCGGTAAAGCAACCCCGTTTCAGGAACTGGTGCT</u>  <u>GAAAGCAGCAGCTAGGGCATCTCTGGCAACCGGTGTGCC</u>  <u>GGTGACCACCCATACCAGTGCATCACAGCGCGATGGCGA</u>  <u>ACAGCAGGCAGCAATTTTTGAATCAGAAGGTCTGAGCCC</u>  <u>GAGTCGTGTGTGATTGGTCATAGCGATGATACCGATGATC</u>  <u>TGTCTTATCTGACCGCCTGGCAGCACGCGGCTATCTGGT</u>  <u>GGGCCTGGATCGTATGCCGTATAGCGCAATTGGTCTGGAA</u>  <u>GGTAATGCAAGTGCCTGGCACTGTTTGGGACGCGCTCTT</u>  <u>GGCAGACCCGCGCACTGCTGATTAAGCACTGATTGATCG</u>  <u>CGGCTACAAAGATCGCATTCTGGTTAGTCATGATTGGCTG</u>  <u>TTTGGCTTTAGTAGCTATGTTACCAATATTATGGATGTTATG</u>  <u>GATCGCATTAATCCGGATGGTATGGCATTGTGCCGCTGCG</u>  <u>TGTTATTCCGTTTCTGCGTGAAAAAGGCGTTCGCCGGAA</u>  <u>ACCCTGGCAGGCGTTACCGTTGCAAATCCGGCACGCTTTC</u>  <u>TGTCTCCGACCGTTCGTGCAAGTCTCGAGCACCACCACC</u>  <u>ACCACCACTGA</u></p> <p>氨基酸序列 (SEQ ID NO:6, 对应全长野生型的第 25-365 位氨基酸序列, 第 1 位的 M 为引入的起始氨基酸, 最后的 6 个 H 为用于纯化的组氨酸 (His) 标签, 6 个 H 之前的 LE 为蛋白和 His 标签之间的连接点):  <u>MSMARPIGTGDLINTVRGPIPVSEAGFTLTHEHICGSSAGFLR</u>  <u>AWPEFFGSRKALAEKAVRGLRHARSAGVQTIQVSTFDIGR</u>  <u>DVRLLAEVSRAADVHIVAATGLWFDPPLSMRMRSVEELTQF</u>  <u>FLREIQHGIEDTGIRAGIHKVATTGKATPFQELVLKAAARASL</u>  <u>ATGVPVTTHTSASQRDGEQQAIFESEGLSPSRVCIGHSDDT</u>  <u>DDLSTGLAARGYLVLDRMPYSAIGLEGNASALALFGTR</u>  <u>SWQTRALLIKALIDRGYKDRILVSHDWLFGFSSYVTNIMDV</u>  <u>MDRINPDGMAFVPLRVIPFLREKGVPPETLAGVTVANPARFL</u>  <u>SPTVRASLEHHHHHH</u></p>
<p>含有 309 位 SeHF 的放射形土壤杆菌磷酸三酯酶突变体 (PTE309SeHF)</p>	<p>核苷酸序列 (SEQ ID NO: 7, 不包含第 1-24 位氨基酸的编码序列, 仅为第 25-365 位氨基酸的编码序列):  <u>ATGTCAATGGCACGCCCGATTGGAACAGGCGATCTGATTA</u>  <u>ATACCGTTCGCGGTCCGATTCCGGTGAGCGAAGCAGGCTT</u>  <u>TACCCTGACCCATGAACATATTTGTGGCTCTAGCGCAGGC</u>  <u>TTTCTGCGTGCATGGCCGGAATTTTTCGGTAGCCGCAAAG</u>  <u>CACTGGCAGAAAAAGCAGTTCGCGGCCTGCGTCATGCGC</u>  <u>GAAGCGCAGGTGTTTCAGACCATTGTGGATGTGAGTACCTT</u></p>

<p> <u>TGATATTGGTCGTGATGTTTCGTCTGCTGGCAGAAGTTAGT</u>  <u>CGCGCAGCAGATGTTTCATATTGTTGCAGCAACCGGCCTGT</u>  <u>GGTTTGATCCCGCGCTGTCAATGCGTATGCGTAGTGTGGA</u>  <u>AGAACTGACCCAGTTCCTTCTGCGCGAAATTCAGCATGGT</u>  <u>ATTGAAGATACCGGTATTCGCGCAGGTATTATTAAGTTGC</u>  <u>AACCACCGGTAAAGCAACCCCGTTTCAGGAACTGGTGCT</u>  <u>GAAAGCAGCAGCTAGGGCATCTCTGGCAACCGGTGTGCC</u>  <u>GGTGACCACCCATAACCAGTGCATCACAGCGCGATGGCGA</u>  <u>ACAGCAGGCAGCAATTTTTGAATCAGAAGGTCTGAGCCC</u>  <u>GAGTCGTGTGTATTGGTCATAGCGATGATACCGATGATC</u>  <u>TGTCTTATCTGACCGGCCTGGCAGCACGCGGCTATCTGGT</u>  <u>GGGCCTGGATCGTATGCCGTATAGCGCAATTGGTCTGGAA</u>  <u>GGTAATGCAAGTGCCTGGCACTGTTTGGGACGCGCTCTT</u>  <u>GGCAGACCCGCGCACTGCTGATTAAGCACTGATTGATCG</u>  <u>CGGCTACAAAGATCGCATTCTGGTTAGTCATGATTGGCTG</u>  <u>TTTGGCTTTAGTAGCTAGGTTACCAATATTATGGATGTTATG</u>  <u>GATCGCATTAATCCGGATGGTATGGCATTGTGCCGCTGCG</u>  <u>TGTTATTCCGTTTCTGCGTGAAAAAGGCGTTCGCCGGAA</u>  <u>ACCCTGGCAGGCGTTACCGTTGCAAATCCGGCACGCTTTC</u>  <u>TGTCTCCGACCGTTCGTGCAAGTCTCGAGCACCACCACC</u>  <u>ACCACCACTGA</u> </p> <p>           氨基酸序列 (SEQ ID NO: 8), 其中第 285 位的 Y* (对应全长野生型 <i>arPTE</i> 的第 309 位) 表示引入的硒代酪氨酸:         </p> <p>           MSMARPIGTGDLINTVRGPIPVSEAGFTLTHEHICGSSAGFLR            AWPEFFGSRKALAEKAVRGLRHARSAGVQTIVDVSTFDIGR            DVRLLAEVSRAADVHIVAATGLWFDPPLSMRMRSVEELTQF            FLREIQHGIEDTGIRAGIHKVATTGKATPFQELVLKAAARASL            ATGVPVTTHTSASQRDGEQQAIFESEGLSPSRVCIGHSDDT            DDLSYLTGLAARGYLVGLDRMPYSAIGLEGNASALALFGTR            SWQTRALLIKALIDRGYKDRILVSHDWLFGFSSY*VTNIMD            VMDRINPDGMAFVPLRVIPFLREKGVPPETLAGVTVANPAR            FLSPTVRASLEHHHHHH         </p>
---

图5