



(12)发明专利

(10)授权公告号 CN 105820252 B

(45)授权公告日 2020.07.21

(21)申请号 201610177621.6

谢丽尔·L·奎恩

(22)申请日 2011.05.03

(74)专利代理机构 北京英赛嘉华知识产权代理

有限责任公司 11204

(65)同一申请的已公布的文献号

申请公布号 CN 105820252 A

代理人 王达佐 洪欣

(43)申请公布日 2016.08.03

(51)Int.Cl.

(30)优先权数据

61/330,828 2010.05.03 US

61/330,829 2010.05.03 US

C07K 19/00(2006.01)

C12N 9/00(2006.01)

C12N 15/11(2006.01)

C07K 16/40(2006.01)

C12P 21/02(2006.01)

A61K 38/53(2006.01)

(62)分案原申请数据

201180031304.9 2011.05.03

A61K 39/395(2006.01)

(73)专利权人 ATYR 医药公司

地址 美国加利福尼亚州

专利权人 盘古生物制药有限公司

A61P 35/00(2006.01)

A61P 25/00(2006.01)

A61P 3/10(2006.01)

A61P 29/00(2006.01)

(72)发明人 莱斯利·安·格林尼

凯尔·P·基昂格 洪非

阿兰·P·瓦瑟洛特 罗咏诗

杰弗里·D·沃特金斯

约翰·D·门德莱恩

审查员 张宁

权利要求书1页 说明书139页

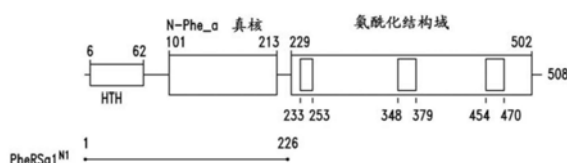
序列表48页 附图5页

(54)发明名称

与苯丙氨酰- $\alpha$ -tRNA合成酶的蛋白片段相关的治疗、诊断和抗体组合物的创新发现

(57)摘要

本文提供了包含氨酰-tRNA合成酶的新鉴定的蛋白片段的组合物、编码所述蛋白片段的多核苷酸及其互补物、相关剂、及其在诊断、药物发现、研究和治疗应用中的使用方法。



1. 治疗组合物,包含分离的氨酰-tRNA合成酶(AARS)多肽,所述多肽由SEQ ID NO:45、12、19和43中的任一个组成,或者通过N-端的一个氨基酸的缺失而区别于SEQ ID NO:45、12、19和43中的任一个,其中所述多肽具有细胞外信号传导活性并且具有至少5mg/mL的溶解度,并且其中所述组合物以蛋白基计具有至少95%的纯度,并小于10EU内毒素/mg蛋白。

2. 如权利要求1所述的治疗组合物,其中所述AARS多肽由SEQ ID NO:45、12、19和43中任一个所示的氨基酸序列组成。

3. 如权利要求1或2所述的治疗组合物,其中所述AARS多肽与异源多肽融合,其中所述异源多肽选自纯化标签、表位标签、靶向序列、信号肽、膜易位序列和PK调节物。

4. 分离的氨酰-tRNA合成酶(AARS)多肽,所述多肽由SEQ ID NO:45、12、19和43中任一个组成,或者通过N-端的一个氨基酸的缺失而区别于SEQ ID NO:45、12、19和43中的任一个,其中所述AARS多肽具有细胞外信号传导活性。

5. 如权利要求4所述的分离的AARS多肽,其中所述AARS多肽由SEQ ID NO:45、12、19和43中任一个所示的氨基酸序列组成。

6. 如权利要求4或5所述的分离的AARS多肽,其中所述AARS多肽与异源多肽融合,其中所述异源多肽选自纯化标签、表位标签、靶向序列、信号肽、膜易位序列和PK调节物。

7. 组合物,包含权利要求4-6任一项所述的分离的氨酰-tRNA合成酶(AARS)多肽,其中所述多肽具有至少5mg/mL的溶解度,并且其中所述组合物以蛋白基计具有至少95%的纯度,并小于10EU内毒素/mg蛋白。

8. 如权利要求7所述的组合物,其中至少一个部分或固体基底与所述AARS多肽共价或非共价连接,其中所述部分是水溶性聚合物或可检测的标记物。

9. 细胞组合物,包含权利要求4-6任一项所述的氨酰-tRNA合成酶(AARS)多肽;和工程化的细胞群,其中至少一个细胞包含编码所述AARS多肽的多核苷酸,其中所述细胞能够在无血清培养基中生长。

10. 细胞生长装置,包含权利要求4-6任一项所述的氨酰-tRNA合成酶(AARS)多肽;和工程化的细胞群,其中至少一个细胞包含编码所述AARS多肽的多核苷酸;至少10升的无血清生长培养基;和无菌容器。

11. 鉴定特异性结合权利要求4-6任一项所述的氨酰-tRNA合成酶(AARS)多肽的化合物的方法,所述方法包括:

- a) 使所述AARS多肽与至少一种测试化合物在适合条件下混合,和
- b) 检测所述AARS多肽与所述测试化合物的结合,从而鉴定特异性结合所述AARS多肽的化合物。

12. 分离的氨酰-tRNA合成酶(AARS)多核苷酸,其由以下组成:

编码权利要求4-6任一项所述的AARS多肽的核苷酸序列,以及任选的一个或多个控制元件或调节序列或非编码序列。

13. 药物组合物,包含权利要求4-6任一项所述的AARS多肽、或权利要求12所述的分离的AARS多核苷酸,和药学上可接受的载体。

## 与苯丙氨酰- $\alpha$ -tRNA合成酶的蛋白片段相关的治疗、诊断和抗体组合物的创新发现

[0001] 相关申请的交叉引用

[0002] 本申请根据35U.S.C. §119(e), 要求于2010年5月3日提交的美国临61/330,828时专利申请第号和2010年5月3日提交的美国临时专利申请第61/330,829号的权益, 每一个临时申请的全部内容通过引用并入本文。

[0003] 关于序列表的声明

[0004] 以文本格式代替纸印本提供本申请相关的序列表, 并且在此通过引用并入本说明书。含有序列表的文本文件的名称为120161\_465PC\_SEQUENCE\_LISTING.txt。该文本文件是大约124KB, 创建于2011年5月2日, 并通过EFS-Web以电子方式提交。

### 技术领域

[0005] 本发明大体涉及包含氨酰-tRNA合成酶和其他蛋白的新鉴定的蛋白片段的组合物, 编码它们的多核苷酸及其互补物, 相关剂, 及其在诊断、药物发现、研究和治疗应用中的使用方法。

### 背景技术

[0006] 在过去的四十年, 氨酰-tRNA合成酶(AARS)被认为是催化tRNA分子氨酰化所必需的管家蛋白, 而tRNA分子氨酰化是蛋白翻译过程中遗传信息解码的一部分。AARS在这方面已经得到广泛研究, 并且克隆了许多它们的全长序列以进行序列分析并提供生化实验的丰富来源。然而, 一些AARS片段和其他蛋白具有意想不到的与氨酰化无关的活性, 包括调节超出蛋白翻译的途径的细胞外信号传导活性。一般而言, 在全长或亲本蛋白序列的情况下未发现这些意想不到的活性; 反而在从其亲本序列去除或切除AARS蛋白片段之后, 或者通过表达并充分纯化AARS片段序列, 然后测试新的非合成酶相关的活性, 发现了这些活性。

[0007] 虽然AARS的全长序列已经知道了一段时间, 但是还没有进行系统的实验分析来阐释此类AARS蛋白片段或来自有关或相关蛋白的蛋白片段, 或者评价全长AARS蛋白在氨基酸合成之外的新生物活性方面的潜在作用。在本说明书的多个部分, 此类AARS蛋白片段、AARS结构域或AARS选择性剪接变体在本文被称为“切除蛋白(resectin)”。在其最广的范围内, 术语“切除蛋白”指已经从其天然的全长或亲本蛋白序列切离或限制(通过蛋白水解、选择性剪接、诱变或重组基因工程)的一部分蛋白, 否则其常常掩盖其新的生物活性。同样, 还没有进行系统的实验分析以考察此类切除蛋白在治疗各种医学病症中作为生物治疗剂、诊断剂或药物靶的用途, 或者它们与人类疾病的潜在关联。作为具有对哺乳动物生命至关重要的已知功能的必需的管家基因, AARS既没有被考虑作为哺乳动物的药物靶, 也没有通过标准的基因组测序、生物信息学或类似工作分析它们以鉴定具有非合成酶活性的切除蛋白。同样, 标准的生化研究工作已经远离鉴定AARS切除蛋白的生物性质及其潜在的治疗和诊断相关性的方向, 这主要归因于之前所理解的其相应的全长亲本AARS的作用。

[0008] 附图简述

[0009] 图1显示与N端AARS多肽的相对位置和大小重叠的苯丙氨酰- $\alpha$ -tRNA合成酶的域结构示意图。图1A代表通过质谱分析鉴定的片段,图1B代表通过转录组的深度测序鉴定的片段,以及图1C代表通过生物信息学分析鉴定的片段。

[0010] 图2显示与C端AARS多肽的相对位置和大小重叠的苯丙氨酰- $\alpha$ -tRNA合成酶的域结构示意图。图2A代表通过转录组的深度测序鉴定的片段,以及图2B代表通过生物信息学分析鉴定的片段。

[0011] 发明概述

[0012] 本发明的实施方案一般涉及氨酰-tRNA合成酶(AARS)的蛋白片段的发现,所述蛋白片段具有非常规的生物活性,例如细胞外信号传导活性,和/或治疗和诊断相关的其他特性。AARS是所有生物中存在的蛋白合成器的通用且必要元件,但是人类AARS及其相关蛋白具有天然存在的切除变体,具有强大的细胞信号传导活性,促进人类的正常功能。这些蛋白片段的活性不同于AARS公知的蛋白合成活性,并且本发明包括发现和开发这些切除蛋白作为新治疗剂、新发现研究试剂和定向生物制剂和诊断剂的新抗原/靶,可能用于治疗或诊断许多人类疾病,例如炎性、血液学、神经退行性、自身免疫性、血细胞生成、心血管和代谢疾病或疾患。

[0013] 因此,本发明的AARS蛋白片段可以被称为“切除蛋白”或者“appendacrine”。如上所述,术语“切除蛋白”源自从其全长亲本AARS序列环境切割或切除给定的AARS蛋白片段的过程,而其全长亲本AARS序列通常掩盖其非常规活性。在某些情况下,本发明的AARS蛋白片段和多核苷酸通过该切除过程的发生被鉴定,而不论是天然存在的(例如,蛋白水解的、剪接变体)、人工诱导的还是预测的。术语“appendacrine”衍生自“附加”(来自拉丁语-appendere)和“分离”或“分辨”(来自希腊语-crines),并且也反映了AARS蛋白片段的一个或多个附加结构域与其相应的全长或亲本AARS序列分离。

[0014] 尽管之前已经显示几种AARS片段具有非合成酶活性,但是有关生物治疗、发现或诊断功用的此类片段的表达、分离、纯化和表征是有限的,并且本领域技术人员不容易将此类活性与整个AARS家族的每个成员或替代片段相联系。在此,利用了系统的方法发现并证实了用于生物治疗发现和诊断功用的20个线粒体AARS和20个胞质AARS(和相关蛋白)的AARS蛋白片段。例如,使用主要鉴定蛋白水解片段的质谱(MS)从生物样品鉴定了本发明的AARS蛋白片段和编码它们的多核苷酸中的某些,并且通过主要鉴定剪接变体的深度测序技术鉴定了其他。使用计算机模拟预测氨基酸序列,例如通过计算机比较来自人类和低等生物的合成酶以及关键区别(例如蛋白酶位点),鉴定了其他AARS蛋白片段;该方法利用了全长AARS的序列分析,基于分辨具有非常规生物活性的蛋白水解片段和功能结构域的具体标准。

[0015] AARS的新的切除蛋白是意想不到的,并且它们的差异表达也是意想不到的。特定的切除通常见于不同的处理(例如,在有或没有血清的培养基中生长的细胞)、不同的生长阶段(例如,成体脑相对于胎儿脑)和不同的组织类型(例如,胰腺相对于肝脏)。尽管在相同的细胞位置以相对比例的量需要所有氨酰tRNA合成酶的常规功能,但所有氨酰tRNA合成酶的表达模式是不同的。人们不会预期一种氨酰tRNA合成酶活性的水平在其他氨酰tRNA合成酶活性不增加的同时增加。质谱和深度测序数据指示氨酰tRNA合成酶切除蛋白确实具有不同的水平并且出现在不同的部位和不同的阶段。

[0016] 此外,AARS蛋白片段可以被表达和纯化至足够高的纯度以分辨其生物性质。之前,片段常常不具有足够的纯度、折叠和稳定性来实现非合成酶活性的适当的生物表征。例如,配合足够纯的、稳定的、可溶的且折叠的切除蛋白使用基于细胞的测定,以揭示其重要的生物治疗、发现或诊断活性。

[0017] 具体说,本发明涉及具有生物治疗、发现或诊断功用的苯丙氨酰- $\alpha$ -tRNA合成酶的蛋白片段,相关剂和组合物及其使用方法。如本文所描述的,本发明的组合物用于许多诊断、药物发现和治疗应用。优选地,AARS蛋白和片段是纯化的并且保存在适合条件下,达到所述生物治疗、发现或诊断用途所需的程度。

[0018] 某些实施方案包括组合物,包含至少约100、90、80、70、60、50或40个氨基酸的分离的氨酰-tRNA合成酶(AARS)蛋白片段,所述蛋白片段包含表1-3或表4-6或表7-9任一个中所列的氨基酸序列,并且具有至少约5mg/mL的溶解度,并且其中所述组合物以蛋白基计具有至少约95%的纯度和小于约10EU内毒素/mg蛋白。一方面,所述组合物是治疗组合物。在具体实施方案中,组合物基本没有血清。在一些实施方案中,AARS蛋白片段包含非常规活性。在一些实施方案中,非常规生物活性选自:细胞外信号传导的调节、细胞增殖的调节、细胞分化的调节、基因转录的调节、细胞因子生成或活性的调节、细胞因子受体活性的调节、和炎症的调节。在一些实施方案中,对于基于细胞的非常规生物活性,AARS蛋白片段具有小于约1nM、约5nM、约10nM、约50nM、约100nM或约200nM的EC<sub>50</sub>。

[0019] 在某些实施方案中,AARS蛋白片段与异源多肽融合。在一些实施方案中,AARS融合蛋白基本上保持所述AARS蛋白片段的非常规活性。在一些实施方案中,AARS融合蛋白抑制所述AARS蛋白片段的非常规活性。在某些实施方案中,异源多肽与AARS蛋白片段的N端连接。在某些实施方案中,异源多肽与AARS蛋白片段的C端连接。一方面,在这些实施方案的任何一个中,异源多肽选自由以下组成的组:纯化标签、表位标签、靶向序列、信号肽、膜转位序列和PK调节剂。

[0020] 在某些实施方案中,组合物包含浓度最少约10mg/mL的AARS蛋白片段。在某些实施方案中,组合物包含至少90%单分散的AARS蛋白片段。在某些实施方案中,组合物包含小于约3%的高分子量聚集蛋白。在某些实施方案中,组合物在4℃下在PBS中以至少10mg/mL的浓度保存1周时表现出小于3%的聚集。在某些实施方案中,组合物在室温下在PBS中以至少10mg/mL的浓度保存1周时表现出小于3%的聚集。

[0021] 本文描述了用于测量切除蛋白的此类特征的各种测定,并可以用来限定本发明的各方面。在某些方面中,这些特征对于本文描述的AARS蛋白片段的生物治疗功用是优选的。

[0022] 某些实施方案包括组合物,包含至少35个氨基酸的分离的氨酰-tRNA合成酶(AARS)蛋白片段,该蛋白片段与表1-3或表4-6或表7-9任一个中所列的氨基酸序列相差约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个氨基酸的取代、缺失和/或添加,其中改变的蛋白片段基本上保持未改变蛋白的非常规活性或者具有与非常规活性相关的显性负表型,其中所述蛋白片段具有至少约5mg/mL的溶解度,并且其中所述组合物以蛋白基计具有至少约95%的纯度和小于约10 EU内毒素/mg蛋白。在具体实施方案中,组合物基本没有血清。

[0023] 其他实施方案包括组合物,包含与表1-3或表4-6或表7-9中所列的分离的氨酰-tRNA合成酶(AARS)蛋白片段特异性结合的分离的抗体,其中所述抗体对所述AARS蛋白片段

的亲和力比所述抗体对相应的全长AARS多肽的亲和力强约10X。本发明的令人惊讶的一个方面包括具有抗体或其他定向生物制剂可及的“新”表面的某些切除蛋白，而全长AARS“隐藏”或用其他序列或相邻结构域覆盖这些表面。切除过程还能产生更大的水可及性，用于揭示之前未鉴定的生物性质。一些实施方案包括组合物，包含与表1-3或表4-6或表7-9所列的分离的氨酰-tRNA合成酶(AARS)蛋白片段特异性结合的分离的抗体，其中所述抗体具有至少约10nM的对所述AARS蛋白片段的亲和力，和至少约100nM的对相应全长AARS多肽的亲和力。在某些实施方案中，抗体结合位于表1-3或表4-6或表7-9任一个中所列的AARS多肽独特剪接点内的表位，或者结合该剪接位点的C端氨基酸序列。在某些实施方案中，抗体拮抗AARS蛋白片段的非常规活性。此类拮抗剂可以任选地结合相应的亲本或全长AARS。

[0024] 其他方面涉及生物测定系统，包含至少90个氨基酸的基本上纯的氨酰-tRNA合成酶(AARS)蛋白片段和结合所述AARS蛋白片段的结合配体，所述蛋白片段包含表1-3或表4-6或表7-9中所列的氨基酸序列。一方面，结合配体选自自由以下组成的组：细胞表面受体蛋白、核酸、脂质膜、细胞调节蛋白、酶和转录因子。任选地，此类受体可以是细胞的部分，所述细胞优选是与所揭示的切除蛋白生物学相关的细胞。

[0025] 某些实施方案包括细胞组合物，包含：至少90个氨基酸的分离的氨酰-tRNA合成酶(AARS)蛋白片段，所述蛋白片段包含表1-3或表4-6或表7-9中所列的氨基酸序列；和工程化的细胞群，其中至少一个细胞包含编码所述AARS蛋白片段的多核苷酸。一方面，所述细胞能够在无血清培养基中生长。

[0026] 还包括检测系统，包含：至少50或100个氨基酸的基本上纯的氨酰-tRNA合成酶(AARS)蛋白片段，所述蛋白片段包含表1-3或表4-6或表7-9中所列的氨基酸序列；包含与所述蛋白片段结合的细胞表面受体或其细胞外部分的细胞；和调节所述AARS蛋白片段和细胞外受体的结合或相互作用的小于约2000道尔顿的分子或第二多肽。

[0027] 特定实施方案包括诊断系统，包含：至少90个氨基酸的基本上纯的氨酰-tRNA合成酶(AARS)蛋白片段，所述蛋白片段包含表1-3或表4-6或表7-9中所列的氨基酸序列；和包含与所述AARS蛋白片段结合的细胞表面受体或其细胞外部分的细胞，其中所述系统或细胞包含指示分子，所述指示分子允许检测所述细胞表面受体或其细胞外部分的水平或活性的变化。

[0028] 某些实施方案包括细胞生长装置，包含：至少35个氨基酸的分离的氨酰-tRNA合成酶(AARS)蛋白片段，所述蛋白片段包含表1-3或表4-6或表7-9中所列的氨基酸序列；工程化的细胞群，其中至少一个细胞包含编码所述AARS蛋白片段的多核苷酸；至少约10升的无血清生长培养基；和无菌容器。在具体实施方案中，英语本文描述的方法或组合物的任何一种的细胞能够在无血清培养基中生长，所述培养基任选含有抗生素和诱导物。

[0029] 一些实施方案涉及反义或RNA干扰(RNAi)剂，包含靶向对抗表1-3或表4-6或表7-9中所列的AARS剪接变体的独特剪接点的序列。

[0030] 还包括治疗组合物，包含至少35个氨基酸的分离的氨酰-tRNA合成酶(AARS)蛋白片段，所述蛋白片段包含表1-3或表4-6或表7-9中所列的氨基酸序列，其中所述蛋白片段特异性结合结合配体并且具有至少约5mg/mL的溶解度，并且其中所述组合物以蛋白基计具有至少约95%的纯度。在一些方面，所述组合物可以具有小于10EU内毒素/mg蛋白。

[0031] 还包括组合物，包含至少35个氨基酸的分离的氨酰-tRNA合成酶(AARS)蛋白片段，

该蛋白片段与表1-3或表4-6或表7-9中所列的氨基酸序列至少80%、85%、90%、95%、98%或100%同一,其中所述蛋白片段具有至少约5mg/mL的溶解度,并且其中所述组合物以蛋白基计具有至少约95%的纯度和小于约10EU内毒素/mg蛋白。在这些实施方案的任何一个中,组合物可以包含就其表观分子量而言是至少约50%、约60%、约70%、约80%、约90%或约95%单分散的AARS蛋白片段。在这些实施方案的任何一个的另一方面,组合物包含小于约10% (以蛋白基计) 高分子量聚集蛋白,或小于约5%高分子量聚集蛋白,或小于约4%高分子量聚集蛋白,或小于约3%高分子量聚集蛋白,或小于2%高分子量聚集蛋白,或小于约1%高分子量聚集蛋白。

[0032] 在这些实施方案的任何一个的另一方面,组合物在4°C下在PBS中以至少10mg/mL的浓度保存1周时表现出小于10%的聚集,或者在4°C下在PBS中以至少10mg/mL的浓度保存1周时表现出小于5%的聚集,或者在4°C下在PBS中以至少10mg/mL的浓度保存1周时表现出小于3%的聚集,或者在4°C下在PBS中以至少10mg/mL的浓度保存1周时表现出小于2%的聚集,或者在4°C下在PBS中以至少10mg/mL的浓度保存1周时表现出小于1%的聚集。

[0033] 某些实施方案包括组合物,包含至少35个氨基酸的基本上纯的氨酰-tRNA合成酶(AARS)蛋白片段和至少一个与之共价或非共价连接的部分,所述蛋白片段包含表1-3或表4-6或表7-9中所列的氨基酸序列。在某些实施方案中,所述部分时可检测的标记物。在某些实施方案中,所述部分是水溶性聚合物。在某些实施方案中,所述部分是PEG。在这些实施方案的任何一个的一个方面,所述部分与所述蛋白片段的N端连接。在这些实施方案的任何一个的一个方面,所述部分与所述蛋白片段的C端连接。

[0034] 具体实施方案包括组合物,包含与至少35个氨基酸的分离的氨酰-tRNA合成酶(AARS)蛋白片段或其生物活性片段或变体连接的固体基底,所述蛋白片段包含表1-3或表4-6或表7-9中所列的氨基酸序列,其中所述蛋白片段具有至少约5mg/mL的溶解度,并且所述组合物以蛋白基计具有至少约95%的纯度。

[0035] 还包括组合物,包含与表1-3或表4-6或表7-9中所列的分离的氨酰-tRNA合成酶(AARS)蛋白片段特异性结合的结合剂,其中所述结合剂具有至少约1nM的对所述蛋白片段的亲和力。一方面,所述结合剂结合位于表1-3或表4-6或表7-9任一个中所列的AARS多肽独特剪接点内的表位,或者结合该剪接位点的C端氨基酸序列。在某些实施方案中,所述结合剂拮抗AARS多肽的非常规活性。

[0036] 某些实施方案包括分离的氨酰-tRNA合成酶(AARS)多肽,包含本文所述的AARS蛋白片段的氨基酸序列、由本文所述的AARS多核苷酸编码的氨基酸序列、或其变体或片段。某些AARS多肽包含与表1-3或表4-6或表7-9或表E2中公开的AARS参照序列至少80%、85%、90%、95%、98%或100%同一的氨基酸序列。某些AARS多肽基本上由与表1-3或表4-6或表7-9或表E2中公开的AARS参照序列至少80%、85%、90%、95%、98%或100%同一的氨基酸序列组成。在某些实施方案中,所述多肽包含非常规生物活性。在具体实施方案中,非常规生物活性选自:细胞信号传导(例如细胞外信号传导)的调节、细胞增殖的调节、细胞迁移的调节、细胞分化的调节、凋亡或细胞死亡的调节、血管生成的调节、细胞结合的调节、细胞代谢的调节、细胞摄取的调节、基因转录的调节、或分泌的调节、细胞因子生成或活性的调节、细胞因子受体活性的调节、和炎症的调节。

[0037] 其他方面包括抗体和其他结合剂,其表现出本文描述的分离的AARS多肽、所述

AARS多肽的结合配体或两者复合物的结合特异性。在某些实施方案中,抗体或结合剂对AARS多肽的亲合力比其对相应的全长AARS多肽的亲合力强约10X。在具体实施方案中,结合剂选自肽、肽模拟物、adnectin、适体和小分子。在某些实施方案中,抗体或结合剂拮抗AARS多肽的非常规活性。在其他实施方案中,抗体或结合剂激动AARS多肽的非常规活性。

[0038] 某些实施方案包括分离的氨酰-tRNA合成酶(AARS)多核苷酸,包含本文所述的AARS多核苷酸的核苷酸序列、编码本文描述的AARS蛋白片段的核苷酸序列、或其变体、片段或互补物。某些AARS多核苷酸包含与表1-3或表4-6或表7-9或表E2中公开的AARS参照多核苷酸至少80%、85%、90%、95%、98%或100%同一的核苷酸序列或其互补物。在某些实施方案中,核苷酸序列是为细菌表达而密码子优化的。一方面,核苷酸序列与表E2中公开的多核苷酸序列至少80%同一。

[0039] 具体的AARS多核苷酸由与表1-3或表4-6或表7-9或表E2中公开的AARS参照多核苷酸至少80%、85%、90%、95%、98%或100%同一的核苷酸序列或其互补物组成。其他AARS多核苷酸包含与表1-3或表4-6或表7-9或表E2中公开的AARS参照多核苷酸特异性杂交的核苷酸序列,或者基本由与表1-3或表4-6或表7-9或表E2中公开的AARS参照多核苷酸特异性杂交的核苷酸序列组成。在某些实施方案中,多核苷酸选自引物、探针和反义寡核苷酸。在具体实施方案中,引物、探针或反义寡核苷酸靶向AARS多核苷酸内的特异性或独特的剪接点、和/或该剪接位点的3'序列。

[0040] 某些实施方案包括测定样品中AARS蛋白片段的的存在或水平的方法,所述方法包括使所述样品与特异性结合本文描述的AARS蛋白片段的一种或多种结合剂接触,检测所述结合剂的存在或不存在,并从而测定所述AARS蛋白片段的的存在或水平。其他实施方案包括测定样品中AARS蛋白片段的的存在或水平的方法,所述方法包括用能够特异性鉴定本文描述的蛋白片段的检测器分析所述样品,并从而测定所述AARS蛋白片段的的存在或水平。在具体实施方案中,所述检测器是质谱仪(MS)、流式细胞仪、蛋白成像设备、酶联免疫吸附测定(ELISA)或蛋白微阵列。某些实施方案包括将AARS蛋白片段的的存在或水平与对照样品或预定值相比较。某些实施方案包括表征样品的状态以使之与所述对照相区分。在具体实施方案中,样品和对照包含细胞或组织,并且所述方法包括区分不同物种的细胞或组织、不同组织或器官的细胞、不同细胞发育状态的细胞、不同细胞分化状态的细胞、不同生理状态的细胞、或健康细胞与患病细胞。例如,选择的切除蛋白可以在诸如应激或刺激条件下更丰富。

[0041] 某些实施方案包括用于鉴定特异性结合本文描述的氨酰-tRNA合成酶(AARS)多肽或其细胞结合配体的一种或多种的化合物的发现方法和相关组合物,所述方法包括a)使所述AARS多肽或其细胞结合配体或两者与至少一种测试化合物在适合条件下混合,并b)检测所述AARS多肽或其细胞结合配体或两者与所述测试化合物的结合,从而鉴定特异性结合所述AARS多肽或其细胞结合配体或两者的化合物。在某些实施方案中,测试化合物是多肽或肽、抗体或其抗原结合片段、肽模拟物或小分子。在某些实施方案中,测试化合物激动所述AARS多肽或其细胞结合配体的非常规生物活性。在其他实施方案中,测试化合物拮抗所述AARS多肽或其细胞结合配体的非常规生物活性。某些实施方案包括通过上述方法鉴定的化合物,例如激动剂(例如,小分子、肽)。

[0042] 某些实施方案包括测定样品中AARS剪接变体的多核苷酸序列的的存在或水平的方法,所述方法包括使所述样品与特异性杂交本文所述的AARS多核苷酸的一种或多种寡核苷



酸接触,检测所述样品中所述寡核苷酸的存在或不存在,并从而测定所述AARS剪接变体的多核苷酸序列的存在或水平。其他实施方案包括测定样品中AARS剪接变体的多核苷酸序列的存在或水平的方法,所述方法包括使所述样品与特异性扩增本文所述的AARS多核苷酸的至少两种寡核苷酸接触,进行扩增反应,检测扩增产物的存在或不存在,并从而测定所述AARS剪接变体的多核苷酸序列的存在或水平。在具体实施方案中,所述寡核苷酸特异性杂交或特异性扩增所述AARS剪接变体独特的剪接点。某些实施方案包括将所述AARS蛋白片段或剪接变体的存在或水平与对照样品或预定值相比较。某些实施方案包括表征所述样品的状态以使之与所述对照相区分。在具体实施方案中,所述样品和对照包含细胞或组织,并且所述方法包括区分不同物种的细胞或组织、不同组织或器官的细胞、不同细胞发育状态的细胞、不同细胞分化状态的细胞、或健康细胞与患病细胞。

[0043] 一些实施方案包括药物组合物,包含本文所述的AARS多核苷酸、本文所述的AARS多肽、本文所述的结合剂或通过上述方法鉴定的或本文描述的化合物,和药学上可接受的赋形剂或载体。

[0044] 某些实施方案包括调节细胞的细胞活性的方法,所述方法包括使所述细胞与本文描述的AARS多核苷酸、本文描述的AARS多肽、本文所述的结合剂、上述方法或本文描述的化合物或本文所述的药物组合物接触。在具体实施方案中,所述细胞活性选自细胞增殖、细胞迁移、细胞分化、凋亡或细胞死亡、细胞信号传导、血管生成、细胞结合、细胞摄取、细胞分泌、代谢、细胞因子生成或活性、细胞因子受体活性、基因转录和炎症。一方面,所述细胞选自自由以下组成的组:前脂肪细胞、骨髓、嗜中性粒细胞、血细胞、肝细胞、星形细胞、充质干细胞和骨骼肌细胞。

[0045] 在某些实施方案中,所述细胞在受试者中。某些实施方案包括治疗受试者,其中所述受试者具有与肿瘤疾病相关的病症、免疫系统疾病或病症、感染性疾病、代谢疾病、炎症疾患、神经元/神经系统疾病、肌肉/心血管疾病、与异常血细胞生成相关的疾病、与异常血管生成相关的疾病或与异常细胞存活相关的疾病。

[0046] 还包括制备药物化合物的方法,包括:a)在包含表1-3或表4-6或表7-9中所列的氨基酸序列的至少35个氨基酸的AARS蛋白片段存在下进行一种或多种候选化合物的体外筛选,以鉴定特异性结合所述AARS蛋白片段的化合物;b)使用步骤a)中鉴定的化合物进行基于细胞的或生化或受体测定,以鉴定调节所述AARS蛋白片段的一种或多种非常规活性的化合物;c)任选地评估在步骤b)中鉴定的化合物的结构-活性关系(SAR),以使其结构与所述非常规活性的调节相关联,并任选地衍生化所述化合物以改变其调节所述非常规活性的能力;和d)产生对于人类使用而言足量的在步骤b)中鉴定的化合物或在步骤c)中衍生的化合物,从而制备所述药物化合物。

[0047] 其他实施方案包括制备药物化合物的方法,包括:a)在特异性结合表1-3或表4-6或表7-9的AARS蛋白片段的细胞表面受体或其细胞外部分存在下进行一种或多种候选化合物的体外筛选,以鉴定特异性结合所述细胞表面受体或其细胞外部分的化合物;b)使用步骤a)中鉴定的化合物进行基于细胞的或生化或受体测定,以鉴定调节所述AARS蛋白片段的一种或多种非常规活性的化合物;c)任选地评估在步骤b)中鉴定的化合物的结构-活性关系(SAR),以使其结构与所述非常规活性的调节相关联,并任选地衍生化所述化合物以改变其调节所述非常规活性的能力;和d)产生对于人类使用而言足量的在步骤b)中鉴定的化合

物或在步骤c)中衍生的化合物,从而制备所述药物化合物。

[0048] 一些实施方案包括细胞组合物,包含工程化的细胞群,其中至少一个细胞包含编码异源全长半胱氨酰-tRNA合成酶(AARS)蛋白的多核苷酸,其中所述细胞能够在无血清培养基中生长。一方面,所述全长氨酰-tRNA合成酶(AARS)蛋白包含异源纯化标签或表位标签以利于AARS蛋白片段的纯化。另一方面,所述全长氨酰-tRNA合成酶(AARS)蛋白包含异源蛋白水解位点以使AARS蛋白片段能够在裂解时生成。

[0049] 一些实施方案包括在细胞内原位生成表1-3或表4-6或表7-9或表E2所列的AARS多肽的方法,所述方法包括:i)在所述细胞内表达异源全长半胱氨酰-tRNA合成酶(AARS)蛋白,其中所述细胞包含能够裂解所述异源全长氨酰-tRNA合成酶(AARS)蛋白以生成所述AARS多肽的蛋白酶。

[0050] 一些实施方案包括生成表1-3或表4-6或表7-9或表E2所列的AARS多肽的方法,所述方法包括使分离的全长氨酰-tRNA合成酶(AARS)蛋白与能够裂解所述异源全长氨酰-tRNA合成酶(AARS)蛋白并产生AARS多肽的蛋白酶接触。

[0051] 一些实施方案包括工程化的全长氨酰-tRNA合成酶(AARS)蛋白,包含能够实现表1-3或表4-6或表7-9或表E2任一个中所列的AARS蛋白片段的蛋白水解生成的异源蛋白水解位点。

[0052] 一些实施方案包括组合物,包含分离的全长氨酰-tRNA合成酶蛋白,其中所述组合物以蛋白基计具有至少约95%的纯度,小于约10EU内毒素/mg蛋白,并且基本上没有血清。一方面,所述全长氨酰-tRNA合成酶蛋白以至少10mg/mL的浓度存在,并且是至少90%单分散的。

[0053] 另一实施方案包括治疗由tRNA合成酶的表达、活性或时空位置的异常调节介导的疾病或疾患的方法,所述方法通过施用表1-3或表4-6或表7-9或表E2任一个中所列的AARS蛋白片段或编码所述ARRS蛋白片段的核酸。在该实施方案的一个方面,所述疾病选自癌症、神经病、糖尿病和炎性疾患。

[0054] 发明详述

[0055] 目录

	I. 概述	14
	II. 定义	15
	III. 纯化的 AARS 蛋白片段和变体	27
	IV. AARS 多核苷酸	58
	V. 抗体	70
	VI. 抗体替代物和其他结合剂	75
	VII. 生物测定和分析测定	80
	VIII. 表达和纯化系统	82
[0056]	IX. 诊断方法和组合物	95
	X. 反义剂和 RNAi 剂	110
	A. 反义剂	111
	B. RNA 干扰剂	119
	XI. 药物发现	127
	XII. 使用方法	135
	XIII. 药物制剂、施用和试剂盒	139
	XIV. 实施例	148

[0057] I. 概述

[0058] 本发明至少部分涉及新的AARS多肽的发现,及其制备和使用方法,这代表天然野生型蛋白转化成与天然存在的全长苯丙氨酰- $\alpha$ 亚单位-tRNA合成酶基因相比表现出明显不同特性的新形式。此类AARS多肽的鉴定是基于广泛序列和在不同组织中表达的苯丙氨酰- $\alpha$ 亚单位-tRNA合成酶的质谱分析,随后系统生成并测试每个潜在AARS多肽以鉴定代表稳定且可溶的蛋白结构域的蛋白序列,其表现出新的生物活性和有利的治疗药物特性。

[0059] 基于该分析,已经鉴定了源自苯丙氨酰- $\alpha$ 亚单位-tRNA合成酶的AARS多肽的至少两个新家族。

[0060] 一方面,此类苯丙氨酰- $\alpha$ 亚单位tRNA合成酶衍生的AARS多肽包括包含苯丙氨酰- $\alpha$ 亚单位tRNA合成酶的大约氨基酸1-226的多肽序列。

[0061] 第二方面,此类苯丙氨酰- $\alpha$ 亚单位tRNA合成酶衍生的AARS多肽包括包含苯丙氨酰- $\alpha$ 亚单位tRNA合成酶的大约氨基酸1-152的多肽序列。

[0062] 这些新的AARS多肽家族代表了新的之前未知的蛋白产物,尤其表现出i) 新的生物活性, ii) 有利的蛋白稳定性和聚集特性, 和iii) 在原核表达系统中以高水平表达和生成的能力,这些是未在完整的野生型蛋白中发现的显著不同的特性。

[0063] II. 定义

[0064] 除非另外指明,本文所用的所有技术和科学术语都具有与本发明所属领域普通技术人员所通常理解的含义相同的含义。尽管在实施或检验本发明时可以使用与本文所述的

相似或等同的任何方法和材料,但是本文描述了优选的方法和材料。出于本发明的目的,下文定义了以下术语。

[0065] 本文使用冠词“a(一个)”和“an(一个)”指一个或多于一个(即,指至少一个)的该冠词的语法对象。举例来说,“一个元件”表示一个元件或多于一个的元件。

[0066] “约”表示与参照的量、水平、值、数、频率、百分比、尺度、尺寸、数额、重量或长度相比,改变多达30、25、20、15、10、9、8、7、6、5、4、3、2或1%的量、水平、值、数、频率、百分比、尺度、尺寸、数额、重量或长度。

[0067] “激动剂”是指增强或模拟活性的分子。例如,AARS或另一蛋白的非常规生物活性。激动剂可以包括蛋白、核酸、碳水化合物、小分子或任何其他化合物或组合物,这些激动剂通过与AARS或其结合配体直接相互作用或通过AARS参与的生物途径中的成分起作用来调节AARS的活性。包括部分激动剂和完全激动剂。

[0068] 如本文所用的,术语“氨基酸”意图表示天然存在的和非天然存在的氨基酸以及氨基酸类似物和模拟物。天然存在的氨基酸包括蛋白生物合成中使用的20种(L)-氨基酸以及其他氨基酸,例如4-羟脯氨酸、羟赖氨酸、锁链素、异锁链素、高半胱氨酸、瓜氨酸和鸟氨酸。非天然存在的氨基酸包括例如(D)-氨基酸、正亮氨酸、正缬氨酸、p-氟苯丙氨酸、乙基硫酸等,其是本领域技术人员已知的。氨基酸类似物包括天然和非天然存在的氨基酸的修饰形式。这种修饰可以包括例如取代或替代氨基酸上的化学基团和部分,或者通过氨基酸的衍生化。氨基酸模拟物包括例如表现出功能上类似性质的有机结构,所述性质例如参照氨基酸的电荷和电荷空间特性。例如,模拟精氨酸(Arg或R)的有机结构具有位于类似分子空间并且具有与天然存在的Arg氨基酸的侧链的ε-氨基相同程度的移动性的正电荷部分。模拟物还包括约束结构以维持氨基酸或氨基酸官能团的最佳空间和电荷相互作用。本领域技术人员知道或者可以确定什么结构构成功能上等效的氨基酸类似物和氨基酸模拟物。

[0069] 在某些方面,非天然氨基酸的使用可用于修饰(例如增加)AARS蛋白片段的选定的非常规活性,或者改变蛋白的体内或体外半衰期。可以使用非天然氨基酸来促进AARS蛋白的(选择性)化学修饰(例如,聚乙二醇化)。例如,某些非天然氨基酸允许聚合物例如PEG与给定蛋白连接,从而改善其药代动力学性质。

[0070] 氨基酸类似物和模拟物的具体实例可以参见例如Roberts和Vellaccio, *The Peptides: Analysis, Synthesis, Biology*, 编辑Gross和Meinhofer, 第5卷, p.341, Academic Press, Inc., New York, N.Y. (1983), 其全卷通过引用并入本文。其他实例包括全烷基化氨基酸,特别是全甲基化氨基酸。参见例如Combinatorial Chemistry, 编辑Wilson和Czarnik, Ch.11, p.235, John Wiley&Sons Inc., New York, N.Y. (1997), 其全书通过引用并入本文。而其他实例包括其酰胺部分(和因此,得到的肽的酰胺骨架)已经被例如糖环、类固醇、苯并二氮杂环庚烷或碳环所替代的氨基酸。参见例如, *Burger's Medicinal Chemistry and Drug Discovery*, 编辑Manfred E. Wolff, Ch.15, pp.619-620, John Wiley&Sons Inc., New York, N.Y. (1995), 其全书通过引用并入本文。用于合成肽、多肽、肽模拟物和蛋白的方法是本领域公知的(参见例如,美国专利号5,420,109; M. Bodanzky, *Principles of Peptide Synthesis* (第1版&第2修订版), Springer-Verlag, New York, N.Y. (1984&1993), 见第7章; Stewart和Young, *Solid Phase Peptide Synthesis*, (第2版), Pierce Chemical Co., Rockford, Ill. (1984), 其各自通过引用并入本文)。因此,本发明的AARS多肽可以由天

然存在和非天然存在的氨基酸以及氨基酸类似物和模拟物构成。

[0071] 术语“拮抗剂”指减小或减弱一种活性的分子。例如，AARS或另一蛋白的非常规活性生物活性。拮抗剂可以包括诸如抗体的蛋白、核酸、碳水化合物、小分子或任何其他化合物或组合物，这些拮抗剂通过与AARS或其结合配体直接相互作用或通过AARS参与的生物途径中的组分起作用来调节AARS或其结合配体的活性。包括部分和完全拮抗剂。

[0072] 术语“氨酰-tRNA合成酶”（AARS）一般指以其天然或野生型形式能够催化特定氨基酸或其前体酯化成所有其相容的相关tRNA之一以形成氨酰-tRNA的酶。在这种“常规”活性中，氨酰-tRNA合成酶催化一个两步骤反应：首先，它们通过形成氨酰-腺苷酸而活化其各自的氨基酸，其中氨基酸的羧基通过替代焦磷酸而与ATP的 $\alpha$ -磷酸连接，然后，当结合正确的tRNA时，氨酰-腺苷酸的氨酰基被转移至tRNA的2'或3'末端OH。

[0073] I类氨酰-tRNA合成酶通常具有两个高度保守的序列基序。这些酶在腺苷核苷酸的2'-OH处氨基酰化，并且通常是单体或二聚体。II类氨酰-tRNA合成酶通常具有三个高度保守的序列基序。这些酶在同一腺苷的3'-OH处氨基酰化，并且通常是二聚体或四聚体。II类酶的活性位点主要由侧翼为 $\alpha$ -螺旋的七链反平行 $\beta$ -片层构成。尽管苯丙氨酸-tRNA合成酶是II类，但是它在2'-OH处氨基酰化。

[0074] AARS多肽包括酪氨酰-tRNA合成酶(TyrRS)、色氨酰-tRNA合成酶(TrpRS)、谷氨酰胺氨酰-tRNA合成酶(GlnRS)、甘氨酰-tRNA合成酶(GlyRS)、组氨酰-tRNA合成酶(HisRS)、丝氨酰-tRNA合成酶(SerRS)、苯丙氨酰-tRNA合成酶(PheRS)、丙氨酰-tRNA合成酶(AlaRS)、天冬酰胺氨酰-tRNA合成酶(AsnRS)、天冬氨酰-tRNA合成酶(AspRS)、半胱氨酰-tRNA合成酶(CysRS)、谷氨酰-tRNA合成酶(GluRS)、脯氨酰-tRNA合成酶(ProRS)、精氨酰-tRNA合成酶(ArgRS)、异亮氨酰-tRNA合成酶(IleRS)、亮氨酰-tRNA合成酶(LeuRS)、赖氨酰-tRNA合成酶(LysRS)、苏氨酰-tRNA合成酶(ThrRS)、甲硫氨酰-tRNA合成酶(MetRS)或缬氨酰-tRNA合成酶(ValRS)的线粒体和胞质形式。这些AARS多肽的野生型或亲本序列是本领域已知的。

[0075] “编码序列”表示负责编码基因的多肽产物的任何核酸序列。与之相反，术语“非编码序列”是指并非负责编码基因的多肽产物的任何核酸序列。

[0076] 整个说明书中，除非上下文另外要求，词语“包含(comprise)”、“包含(comprises)”和“包含(comprising)”应理解为意指包括陈述的步骤或元素或者步骤或元素的组，但不排除任何其他步骤或元素或者步骤或元素的组。

[0077] “由...组成”表示包括且限于跟随在短语“由...组成”后的任何内容。因此，短语“由...组成”表明所列出的元素是必需的或强制性的，并且可以不存在其他元素。“基本上由...组成”表示包括列在该短语后所列的任何元素，并限于不妨碍或不促进在所列元素的公开内容中指明的活性或作用的其他元素。因此，短语“基本上由...组成”表明所列元素是必需的或强制性的，但其他元素是任选的，可以存在或可以不存在，这取决于它们是否实质影响所列元素的活性或作用。

[0078] “不含内毒素”或“基本不含内毒素”的表述一般涉及含有至多痕量(例如，对受试者没有临床上不良的生理效应的量)内毒素并且优选不可检测的量的内毒素的组合物、溶剂和/或容器。内毒素是与某些细菌、通常是革兰氏阴性细菌有关的毒素，但是内毒素也可存在于革兰氏阳性细菌中，例如产单核细胞李斯特菌。最普遍的内毒素是存在于各种革兰氏阴性细菌外膜的脂多糖(LPS)或脂寡糖(LOS)，并且其代表了这些细菌引起疾病的能力的

主要致病特征。人类中的小量内毒素可以产生发热、血压降低和炎症和凝血的激活,以及其他不良的生理效应。

[0079] 因此,在AARS多肽的药物生产中,经常希望从药物产品和/或药物容器去除大部分或所有痕量的内毒素,因为即使小量也可能引起人类的不良反应。除热源的烘箱可用于该目的,因为超过300°C的温度通常是分解大多数内毒素所需要的。例如,基于主要的包装材料,例如注射器或管形瓶,250°C的玻璃温度与30分钟的维持时间的组合经常足以实现内毒素水平减少3个数量级。考虑去除内毒素的其他方法,包括例如色谱和过滤方法,如本文所述和本领域已知的。还包括在真核细胞例如哺乳动物细胞中产生AARS多肽并从中分离它们的方法,以减少(如果不是消除)本发明组合物中存在的内毒素的风险。在无血清细胞中产生AARS多肽并从中分离它们的方法是优选的。这种包含AARS多肽的组合物代表了新制剂,表现出污染了血清或内毒素的AARS多肽组合物所没有的新颖的生物和治疗特性,血清或内毒素具有结合AARS多肽并改变AARS多肽的新颖生物性质的潜力。

[0080] 可以使用本领域的常规技术检测内毒素。例如,利用了鲎的血液的鲎变形细胞溶解物测定是用于测定内毒素存在的非常灵敏的测定,并且基于该测定检测内毒素的试剂、试剂盒和仪器可商业途径获自例如Lonza Group。在该测试中,非常低水平的LPS可以引起可检测的鲎溶解物的凝结,这归因于放大该反应的强大的酶促级联。还可以通过酶联免疫吸附测定(ELISA)来定量内毒素。要成为基本上无内毒素的,内毒素水平可以小于约0.001、0.005、0.01、0.02、0.03、0.04、0.05、0.06、0.08、0.09、0.1、0.5、1.0、1.5、2、2.5、3、4、5、6、7、8、9或10EU/mg蛋白。通常,1ng脂多糖(LPS)对应于约1-10EU。

[0081] 在某些实施方案中,组合物中任何给定剂(例如,AARS蛋白片段)的“纯度”可以是具体限定的。例如,某些组合物可以包含至少80、85、90、91、92、93、94、95、96、97、98、99或100%纯的剂,包括其间所有小数,例如但不限于通过高效液相色谱(HPLC)所测量的,高效液相色谱(HPLC)是生物化学和分析化学中用于分离、鉴定和定量化合物的常用公知形式的柱色谱。

[0082] 如本文所使用的,“功能”和“功能的”等术语是指生物学功能、酶血功能或治疗功能。

[0083] “基因”表示可以占据染色体上的特定基因座的遗传单元,并且由转录调节序列和/或翻译调节序列、和/或编码区、和/或非翻译序列(即,内含子、5'和3'非翻译序列)组成。

[0084] “同源性”是指同一的或组成型保守取代的氨基酸的百分数。同源性可以利用诸如GAP(Deveraux等人,1984,Nucleic acids Research12,387-395)的序列比较程序来确定,其通过引用并入本文。通过这种方式,可以通过在比对中插入空位来比较与本文中所引用的序列相似或明显不同长度的序列,这种空位可以通过例如利用GAP的比较算法来确定。

[0085] 术语“宿主细胞”包括个体细胞或细胞培养物,其可以是或已经是本发明的任何重组载体或分离的多核苷酸的接受者。宿主细胞包括单个宿主细胞的子代,且由于天然的、偶然的或有意的突变和/或改变,该子代可以不必与原始的亲代细胞完全一致(在形态学上或在总DNA互补物中)。宿主细胞包括用本发明的重组载体或多核苷酸在体内或体外转染或感染的细胞。包含本发明的重组载体的宿主细胞是重组宿主细胞。

[0086] “分离的”表示基本上或实质上不含在其天然状态下通常伴随它的组分的物质。例

如,本文使用的“分离的多核苷酸”包括已经从天然存在状态中位于其侧翼的序列中纯化的多核苷酸,例如,从通常邻近DNA片段的序列中移出所述片段。可选地,本文所用的“分离的肽”或“分离的多肽”等包括从其天然的细胞环境中以及从与细胞的其他组分的结合中体外分离和/或纯化的肽或多肽分子,即,它明显不与体内物质结合。

[0087] 如本文所使用的,术语“mRNA”或有时称为“mRNA转录物”包括但不限于:前mRNA转录物、转录加工中间体、准备用于翻译的成熟mRNA以及一个或多个基因的转录物,或来源于mRNA转录物的核酸。转录物加工可以包括剪接、编辑和降解。如本文所使用的,来源于mRNA转录物的核酸是指所述mRNA转录物或其子序列被最终用作模板所合成的核酸。从mRNA逆转录的cDNA,从该cDNA转录的RNA,从该cDNA扩增的DNA,从扩增的DNA转录的RNA等都是来源于该mRNA的转录物,并且多这种来源的产物的检测能指示样品中原始转录物的存在和/或丰度。因此,mRNA来源的样品包括但不限于,一个或多个基因的mRNA转录物、从该mRNA逆转录的cDNA、从该cDNA转录的cRNA、从基因扩增的DNA、从扩增的DNA转录的RNA等。

[0088] 本文使用的“非常规的”活性通常是指i) 本发明的AARS多肽所具有的新活性,而在任何显著程度上,完整的全长亲本蛋白不具有该活性,或ii) 完整的天然全长亲本蛋白所具有的活性,其中AARS多肽与完整的天然全长亲本蛋白相比表现出显著更高(即,至少大20%)的比活性,或者表现出新环境中的活性;例如,与完整的天然全长亲本蛋白具有的其他活性相独立的活性。在AARS多肽的情况下,非常规活性的非限制性实例包括细胞外信号传导、RNA结合、氨基酸结合、细胞增殖的调节、细胞迁移的调节、细胞分化的调节(例如血细胞生成、神经发生、肌发生、骨发生和脂肪生成)、基因转录的调节、细胞凋亡或其它形式的细胞死亡的调节、细胞信号传导的调节、细胞摄取或分泌的调节、血管生成的调节、细胞结合的调节、细胞代谢的调节、细胞因子产生或活性的调节、细胞因子受体活性的调节、炎症的调节等。

[0089] 术语“半数有效浓度”或“EC<sub>50</sub>”指本文所述的AARS蛋白片段、抗体或其他剂在一段指定的暴露时间之后诱导介于基线和最大值中间的响应的浓度;因此,分级剂量响应曲线的EC<sub>50</sub>代表观察到其最大效应的50%时的化合物浓度。在某些实施方案中,本文提供的剂的EC<sub>50</sub>是与如上所述的“非常规”活性相关的。EC<sub>50</sub>还代表在体内获得最大效应的50%所需的血浆浓度。类似地,“EC<sub>90</sub>”指观察到其最大效应的90%时的剂或化合物的浓度。“EC<sub>90</sub>”可以从“EC<sub>50</sub>”和H<sub>11</sub>斜率计算,或者可以使用本领域的常规知识从数据直接确定。在一些实施方案中,AARS蛋白片段、抗体或其他剂的EC<sub>50</sub>小于约0.01、0.05、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50、60、70、80、90或100nM。优选地,生物治疗组合物将具有约1nM或更小的EC<sub>50</sub>值。

[0090] 术语“调节”包括与对照相比,通常以统计学上显著或生理上显著的量“增加”或“刺激”,以及“减少”或“降低”。因此,根据使用条件,“调节剂”可以是激动剂、拮抗剂或其任何混合物。“增加的”或“提高的”量通常是“统计学上显著的”量,并且可以包括没有组合物(缺乏试剂或化合物)或有对照组合物时所产生的量的1.1、1.2、2、3、4、5、6、7、8、9、10、15、20、30或更大倍数(例如,500、1000倍)(包括其间大于1的所有整数和小数,例如1.5、1.6、1.7、1.8等)的增加。“减少的”或降低的量通常是“统计学上显著的”量,并且可以包括没有组合物(缺乏试剂或化合物)或有对照组合物时所产生的量的1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、

35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%的降低,包括其间所有整数。作为一个非限制性实例,比较常规和非常规活性时的对照可能包括与其相应的全长AARS相比感兴趣的AARS蛋白片段,或者具有与其相应的全长AARS相当的常规活性的片段AARS。“统计学上显著的”量的其他实例在下文描述。

[0091] “获自”表示诸如例如,多核苷酸提取物或多肽提取物的样品分离自或源自受试者的特定来源。例如,提取物可以获自从受试者直接分离的组织或生物流体。“源自”或“获自”还可以指多肽或多核苷酸序列的来源。例如,本发明的AARS序列可以“源自”AARS蛋白水解片段或AARS剪接变体或其部分的序列信息,不管它们是天然存在的还是人工产生的,并因此可以包括该序列,基本上由该序列组成,或者由该序列组成。

[0092] 术语“多肽”和“蛋白”在本文可互换使用,指氨基酸残基的聚合物及其变体和合成的和天然存在的类似物。因此,这些术语适用于其中一个或多个氨基酸残基是合成的非天然存在的氨基酸(诸如相应的天然存在的氨基酸的化学类似物)的氨基酸聚合物,以及适用于天然存在的氨基酸聚合物及其天然存在的化学衍生物。此类衍生物包括例如翻译后修饰和降解产物,包括AARS参照片段的焦谷氨酰、异天冬氨酰、蛋白水解的、磷酸化的、糖基化的、氧化的、异构化的和脱氨基化的变体。

[0093] 本文所用的表述“序列同一性”或者例如包含“与...50%同一的序列”指序列在比较窗内在逐个核苷酸基础或逐个氨基酸基础上同一的程度。因此,“序列同一性百分比”可以通过以下来计算:在比较窗中比较两个最优比对的序列,确定两条序列中出现的相同的核酸碱基(例如A、T、C、G、I)或相同的氨基酸残基(例如,Ala、Pro、Ser、Thr、Gly、Val、Leu、Ile、Phe、Tyr、Trp、Lys、Arg、His、Asp、Glu、Asn、Gln、Cys和Met)的位置数量以得到匹配位置数,将该匹配位置数除以比较窗中的位置总数(即,窗大小),并将结果乘以100以得到序列同一性百分比。

[0094] 用于描述两个或更多个多核苷酸或多肽间的序列关系的术语包括“参照序列”、“比较窗”、“序列同一性”、“序列同一性百分比”和“基本同一性”。“参照序列”的长度为至少12个,但通常是15至18个,经常是至少25个单体单元(包括核苷酸和氨基酸残基)。因为两个多核苷酸各自可以包含(1)在两个多核苷酸间相似的序列(即,完整多核苷酸序列的仅一部分)和(2)在两个多核苷酸间不同的序列,两个(或更多个)多核苷酸间的序列比较通常通过比较“比较窗”中的两个多核苷酸的序列来进行以鉴定和比较序列相似性的局部区域。“比较窗”指至少6个连续位置、通常为约50至约100个、更通常约100至150个的连续位置的概念节段,其中在一序列与具有相同连续位置数的参照序列最优比对后,将该序列与参照序列比较。比较窗可以包含与参照序列(其不包含添加或缺失)相比约20%或更少的添加或缺失(即空位)用于两个序列的最优比对。用于比对比较窗的序列的最优比对可以通过算法的计算机运行(Wisconsin遗传学软件包发行版7.0(Genetics Software Package Release7.0)中的GAP、BESTFIT、FASTA和TFASTA,Genetics Computer Group,575 Science Drive Madison,WI,USA)或通过目视检查以及选择的各种方法的任一种产生的最佳比对(即,产生整个比较窗的最高同源性百分比)来实施。例如Altschul等人,1997,Nucl.Acids Res.25:3389公开的BLAST程序家族也可以作为参考。可以在Ausubel等人,“Current Protocols in Molecular Biology,”John Wiley&Sons Inc,1994-1998,第15章的19.3单元中找到序列分析的详细讨论。



[0095] 两序列间的序列相似性或序列同一性(术语在本文可互换使用)的计算如下实施。为了确定两个氨基酸序列或两个核酸序列的同一性百分比,出于最优比较的目的,将序列比对(例如,可以在第一和第二氨基酸或核酸序列中的一个或两者中引入空位以用于最优比对,出于比较目的,可以不管非同源性序列)。在某些实施方案中,出于比较目的比对的参照序列的长度是参照序列的长度的至少30%,优选至少40%,更优选至少50%、60%和甚至更优选至少70%、80%、90%、100%。然后比较在对应氨基酸位置或核苷酸位置的氨基酸残基或核苷酸。当在第一序列的一个位置被与在第二序列的对应位置相同的氨基酸残基或核苷酸占据时,则所述分子在该位置是同一的。

[0096] 两序列间的同一性百分比是被所述序列共享的相同位置数的函数,将空位数和每一空位的长度考虑在内,其需要被引入用于两序列的最优比对。

[0097] 两序列间的序列比较和同一性百分比的确定可以利用数学算法来实现。在一个优选的实施方案中,两氨基酸序列间的同一性百分比利用已并入GCG软件包(可以在<http://www.gcg.com>得到)的GAP程序的Needleman和Wunsch算法(1970, J. Mol. Biol. 48:444-453)来确定,利用Blossum62矩阵或PAM250矩阵,空位权重为16、14、12、10、8、6或4以及长度权重为1、2、3、4、5或6。在另一优选的实施方案中,两核苷酸序列间的同一性百分比利用GCG软件包(可以在<http://www.gcg.com>得到)的GAP程序来确定,利用NWSgapdna.CMP矩阵和40、50、60、70或80的空位权重以及1、2、3、4、5或6的长度权重。特别优选组的参数(除非特别指明,应该利用的一组)是Blossum62得分矩阵,空位罚分为12,空位延伸罚分为4,以及移码空位罚分为5。

[0098] 两氨基酸或核苷酸序列之间的同一性百分比可以利用E. Meyers和W. Miller(1989, Cabios, 4:11-17)的算法来确定,该算法并入ALIGN程序(2.0版)中,利用PAM120权重残基表,空位长度罚分为12,空位罚分为4。

[0099] 本文描述的核酸和蛋白序列可以用作“查询序列”以进行对公开数据库的检索,例如,以鉴定其他家族成员或相关序列。可以利用Altschul, 等人,(1990, J. Mol. Biol, 215:403-10)的NBLAST和XBLAST程序(2.0版)进行这类检索。可以利用NBLAST程序,得分=100,字长=12来进行BLAST核苷酸检索以获得与本发明的核酸分子同源的核苷酸序列。可以利用XBLAST程序,得分=50,字长=3来进行BLAST蛋白检索以获得与本发明的蛋白分子同源的氨基酸序列。为了获得用于比较目的的带空位的比对,可以利用Altschul等人,(1997, Nucleic acids Res, 25:3389-3402)中描述的带空位的BLAST。当利用BLAST和带空位的BLAST程序时,可以利用各自程序(例如,XBLAST和NBLAST)中的缺省参数。

[0100] 术语“溶解度”指本文提供的剂在液体溶剂中溶解并形成均质溶液的性质。溶解度通常表示为浓度,借助每单位体积溶剂的溶质质量(g溶质/kg溶剂、g/dL(100mL)、mg/ml等)、摩尔浓度、重量克分子浓度、摩尔分数或其他类似的浓度描述。在指定条件下,单位量的溶剂可以溶解的溶质的最大平衡量是该溶质在该溶剂中的溶解度,所述指定条件包括温度、压力、pH和溶剂性质。在某些实施方案中,在生理pH下测量溶解度。在某些实施方案中,在水或生理缓冲液例如PBS中测量溶解度。在某些实施方案中,在诸如血液或血清的生物液体(溶剂)中测量溶解度。在某些实施方案中,温度可以是约室温(例如,约20、21、22、23、24、25°C)或约体温(37°C)。在某些实施方案中,诸如AARS蛋白片段的剂在室温或37°C下具有至少约0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、

16、17、18、19、20、25或30mg/ml的溶解度。

[0101] 本文所使用的“剪接点”包括成熟mRNA转录物或所编码的多肽中第一个外显子的3'端与第二个外显子的5'端连接的区域。该区域的大小可以变化,并在一个外显子的3'端与另一个外显子的5'端连接处的确切残基的任一端可以包括2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100或更多个(包括其间所有整数)核苷酸或氨基酸残基。“外显子”指已经通过顺式剪接移除前体RNA的一部分(内含子)或已经通过反式剪接将两个或更多个前体RNA分子连接在一起后,在成熟形式的RNA分子中呈现的核酸序列。成熟RNA分子可以是信使RNA或诸如rRNA或tRNA的非编码RNA的功能形式。取决于上下文,外显子可以指DNA或其RNA转录物中的序列。“内含子”指基因中不翻译成蛋白的非编码核酸区域。非编码的内含子节段被转录到前体mRNA(前mRNA)和某些其他RNA(例如长的非编码RNA)中,并随后在加工为成熟RNA的过程中通过剪接去除。

[0102] “剪接变体”指通过选择性剪接产生的成熟mRNA或其编码的蛋白,所述选择性剪接是指在RNA剪接期间,RNA(初级基因转录物或前mRNA)的外显子以多种途径重新连接的过程。所得的不同mRNA可以翻译成不同的蛋白异构体,从而允许单一基因编码多种蛋白。

[0103] 本文所用的“受试者”包括呈现能够用本发明的AARS多核苷酸或多肽治疗或诊断的症状或处于呈现该症状风险中的任何动物。还包括为诊断或其他目的需要描绘本发明AARS多肽和/或多核苷酸水平的受试者。合适的受试者(患者)包括实验室动物(诸如小鼠、大鼠、兔或豚鼠)、农场动物、和家养动物或宠物(诸如猫或狗)。包括非人灵长类动物和优选人类患者。

[0104] 如本文所用的,“治疗(treatment)”或“治疗(treating)”包括对能够被本文所述的AARS多核苷酸或多肽的非常规活性影响的疾病或病症的症状或病理产生的任何期望的效应,并且可以包括被治疗的疾病或病症的一个或多个可测量的标志物的甚至极小的变化或改善。还包括与非AARS疗法相关的治疗,其中本文所述的AARS序列提供治疗的临床标志物。“治疗(treatment)”或“治疗(treating)”并不必然表示疾病或病症或其相关症状的彻底根除或治愈。接受该治疗的受试者是有相应需要的任何受试者。临床改善的示例性标志物对本领域技术人员是明显的。

[0105] 除非文中有明确相反指示,本发明的实施可利用本技术领域内的常规分子生物学方法和重组DNA技术,为了说明的目的,其中很多方法和技术在下文进行了描述。这些技术在文献中有充分的讲解。参见,例如Sambrook,等人,Molecular Cloning:A Laboratory Manual(第3版,2000);DNA Cloning:A Practical Approach,第I&II卷(D.Glover编辑);Oligonucleotide Synthesis(N.Gait编辑,1984);Oligonucleotide Synthesis:Methods and Applications(P.Herdewijn编辑,2004);Nucleic acid Hybridization(B.Hames&S.Higgins编辑,1985);Nucleic acid Hybridization:Modern Applications(Buzdin and Lukyanov编辑,2009);Transcription and Translation(B.Hames&S.Higgins编辑,1984);Animal Cell Culture(R.Freshney,ed.,1986);Freshney,R.I.(2005)Culture of Animal Cells,a Manual of Basic Technique,第5版.Hoboken NJ,John Wiley&Sons;B.Perbal,A Practical Guide to Molecular Cloning(第3版2010);Farrell,R.,RNA Methodologies:A Laboratory Guide for Isolation and Characterization(第3版2005),Methods of

Enzymology:DNA Structure Part A:Synthesis and Physical Analysis of DNA Methods in Enzymology,Academic Press;Using Antibodies:A Laboratory Manual: Portable Protocol NO.I,Edward Harlow,David Lane,Ed Harlow(1999,Cold Spring Harbor Laboratory Press,ISBN 0-87969-544-7);Antibodies:A Laboratory Manual,Ed Harlow (编辑),David Lane (编辑) (1988,Cold Spring Harbor Laboratory Press,ISBN0-87969-3,4-2),1855.Handbook of Drug Screening,Ramakrishna Seethala,Prabhavathi B.Fernandes编辑(2001,New York,N.Y.,Marcel Dekker,ISBN0-8247-0562-9);和Lab Ref:A Handbook of Recipes,Reagents,and Other Reference Tools for Use at the Bench,编辑Jane Roskams和Linda Rodgers,(2002,Cold Spring Harbor Laboratory, ISBN0-87969-630-3)。

[0106] 本文引用的所有出版物、专利和专利申请通过引用整体并入本文。

[0107] III. 用于治疗剂和其他应用的纯化的AARS蛋白片段和变体

[0108] 令人惊讶地,不同于仅知道其氨酰化活性的其全长亲本序列,已经发现AARS片段具有对于生物治疗、发现和诊断应用重要的生物活性。因此,本发明的实施方案包括氨酰-tRNA合成酶(AARS)的全长蛋白、成熟蛋白同种型和蛋白片段及其生物活性变体和片段。在某些实施方案中,蛋白和片段可以通过内源性蛋白水解、体外蛋白水解、剪接变化或计算机模拟预测以及其他机制产生。本文描述的AARS蛋白片段及其变体可以具有至少一种“非常规”生物活性。本发明的AARS蛋白片段在本文还称为“AARS多肽”或“AARS参照多肽”。在某些实施方案中,本文提供的AARS多肽包括以下或主要由以下组成:以下表1-3或表4-6或表7-9列出的AARS多肽“参照序列”的全部或一部分,其代表各种苯丙氨酰- $\alpha$ 亚单位tRNA合成酶片段的氨基酸序列。小鼠和人AARS蛋白序列是高度相关的,通常在完整序列、特定结构域或特定蛋白片段内相差不超过几个氨基酸。

[0109] N端AARS多肽:(表1、2&3)

[0110]

表 1A 通过质谱鉴定的 AARS 多肽			
名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
PheRSa1 <sup>N1</sup>	蛋白 / 人类 / 1-226	MADGQVAELLLRRLEASDGGGLDSAELAAELGM EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFSKAMSNKWIRVDKSAADGPRVFRVVDSD MEDEVQRRLQLVRRGGQAEKLGKERSSELKRK	SEQ.ID. NO.12

[0111]

表 1A 通过质谱鉴定的 AARS 多肽			
名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
		LLAEVTLKTYWVSKGSASFSTSISKQETELSPEMI SSGSWRDRPFKPYNFLAHGVLPSGHLHP	
PheRSa1 <sup>NI</sup>	DNA / 人类 /	ATGGCGGATGGTCAGGTGGCGGAACTGCTGC TCCGGCGGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCGTGTGTTTCGAAG CATTCCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCCG GTGGACAAGAGTGCGGCTGACGGGCCCGGG TGTTCCGAGTGGTGGACAGCATGGAGGATGA GGTGCAGCGGCGGCTCCAGCTGGTCCGGGGG GGACAGGCTGAGAAGCTGGGGGAGAAGGAG AGGAGCGAGCTGAGGAAGAGGAAGCTGTTG GCTGAAGTGA CTCTGAAGACCTACTGGGTGA GCAAAGGCAGTGCCTTTAGTACCAGCATCTCC AAGCAAGAGACAGAGCTGAGCCCAGAGATGA TCTCCAGTGGCTCTTGCGGGACCGGCCCTTC AAGCCCTACA ACTTCTTGGCCACCGGTGTCTC CCCCGACAGCGGCCACCTTCACCCG	SEQ.ID. NO.13

[0112]

表 1B PheRSa1 <sup>NI</sup> 检测的质谱肽和推断的连接肽		
类型/ 物种	序列	SEQ.ID. NO.
蛋白 / 小鼠	RLEVADGGGLDSAELATQLGVEHQAVVGAVK	SEQ.ID. NO.14
蛋白 / 小鼠	SLQALGEVIEAELR	SEQ.ID. NO.15
蛋白 / 小鼠	STKCWELTTEGEEIAREGSHEARVFRSIPLEGLVQSELMHLPSGK VGFSKAMSNKWIRVDKSAADGPRVFRVDSIEDEVQKR	SEQ.ID. NO.16
蛋白 /	VVDSIEDEVQKR	SEQ.ID.

[0113]

表 1B PheRSa1 <sup>N1</sup> 检测的质谱肽和推断的连接肽		
类型/ 物种	序列	SEQ.ID. NO.
小鼠		NO.17

[0114]

表 1C PheRSa1 <sup>N1</sup> 基于检测的质谱肽的连锁序列		
类型/ 物种	序列	SEQ.ID. NO.
蛋白 / 小鼠	<u>RLEVADGGLDSAELATQLGVEHQAVVGAVKSLQALGEVIEAEL</u> <u>RSTKCWELTTEGEEIAREGSHEARVFRSIPLEGLVQSELMHLPSG</u> <u>KVGFSKAMSNKWIRVDKSAADGPRVFRVVDSDIEDEVQKR</u>	SEQ.ID. NO.18

[0115]

表 2 通过深度测序鉴定的 AARS 多肽和替代转录物			
名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
PheRSa1 <sup>N3</sup>	蛋白 / 人类 / 1-128 + 13 aa	MADGQVAELLLRRLEASDGGGLDSAELAAELGM EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFSKAMSNKWIRVDKSAADGPRVFRVDSE DLLGEQRQCL	SEQ.ID. NO.19
PheRSa1 <sup>N3</sup>	DNA / 人类	ATGGCGGATGGTCAGGTGGCGGAACTGCTGC TCCGGCGGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCGTGTGTTTCGAAG CATTCCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCGG GTGGACAAGAGTGCGGCTGACGGGCCCCGGG TGTCCGAGTGGACTCTGAAGACCTACTGGGT GAGCAAAGGCAGTGCCTTTAG	SEQ.ID. NO.20
PheRSa1 <sup>N4</sup>	蛋白 /	MADGQVAELLLRRLEASDGGGLDSAELAAELGM	SEQ.ID.

表 2  
通过深度测序鉴定的 AARS 多肽和替代转录物

名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
	人类 / 1-168 + 200-508	EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFSKAMSNKWIRVDKSAADGPRVFRVVDS MEDEVQRRLQLVRGGQAEKLGKERSELKRK LLAEVGSWRDRPFKPYNFLAHGVLPDSGHLHP LLKVR SQFRQIFLEMGFTEMPD NFISSFWNFD ALFQPQHPARDQHDTFFLRDPAEALQLPMDY VQRVKRTHSQGGYGSQGYKYNWKLDEARKNL LRHTTSASARALYRLAQKKPFTPVKYFSIDRVF RNETLDATHLAEFHQIEGVVADHGLTLGHLMG VLREFFTKLGITQLRFKPAYNPYTEPSMEVFSYH QGLKKWVEVGNSGVFRPEMLLPMGLPENVSVI AWGLSLERPTMIKYGINNIRELVGHKVNLMV YDSPLCRLDAEPRPPPTQEA	NO.21
PheRSa1 <sup>N4</sup>	DNA / 人类	ATGGCGGATGGTCAGGTGGCGGA ACTGCTGC TCCGCGCGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCGTGTGTTTCGAAG CATTCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCCG GTGGACAAGAGTGCGGCTGACGGGCCCGGG TGTCCGAGTGGTGGACAGCATGGAGGATGA GGTGCAGCGGCGGCTCCAGCTGGTCCGGGGG GGACAGGCTGAGAAGCTGGGGGAGAAGGAG AGGAGCGAGCTGAGGAAGAGGAAGCTGTTG GCTGAAGTTGGCTCTTGGCGGGACCGGCCCT TCAAGCCCTACA ACTTCTTGGCCCACGGTGTC CTCCCCGACAGCGGCCACCTTCACCCGCTGCT CAAGGTCCGCTCCCAGTTCGACAGATCTTCC TGGAGATGGGGTTCACCGAGATGCCGACTGAT AACTTCATTGAGAGCTCCTTCTGGA ACTTGA CGCCCTCTTCCAGCCCCAGCAGCACCCAGCC CGTGACCAGCACGACACCTTCTTCTTCGAGA TCCAGCGGAGGCCCTGCAGCTCCCAATGGAC	SEQ.ID. NO.22

[0116]

[0117]

表 2 通过深度测序鉴定的 AARS 多肽和替代转录物			
名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
		TATGTCCAGCGGGTCAAGCGGACCCACTCTCA GGGCGGCTACGGCTCACAGGGGTACAAGTAT AACTGGAAGCTGGACGAGGCCCGGAAAAAC CTACTGCGAACCCACACCACATCAGCCAGCG CCCGTGCCTCTACCGCCTTGCCAGAAAGAA GCCCTTCACTCCGGTCAAGTACTTCTCCATCG ACCGCGTATTCCGGAATGAGACCCTGGACGCC ACGCACCTGGCTGAGTTCCACCAGATCGAGG GCGTGGTGGCGGATCATGGTCTCACCTTGGGC CACCTCATGGGCGTTCTGCGGGAGTTCTTAC CAAGCTGGGTATCACGCAACTCCGCTTCAAGC CAGCCTACAACCCATACACAGAGCCAGCATG GAGGTGTTTACGCTACCACCAAGGCCTGAAGA AGTGGGTGGAGGTCGGAAACTCGGGGGTCTT CCGTCCAGAGATGCTGCTGCCCATGGGGCTTC CCGAGAACGTGTCGGTCATTGCCTGGGGCCTC TCCCTGGAGCGCCCAACGATGATCAAATATGG CATCAACAATATCCGGGAGCTGGTGGGCCACA AGGTGAACCTGCAGATGGTGTATGACAGTCCC CTGTGCCGCTGGATGCCGAGCCGAGGCCCC CTCCCACACAGGAGGCTGCGTGA	
PheRSa1 <sup>N5</sup>	蛋白 / 人 类 /1-198 + S199R + 243-508	MADGQVAELLLRRLEASDGLDSAELAEELGM EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFSKAMSNKWIRVDKSAADGPRVFRVVD MEDEVQRRLQLVRGGQAEKLGEKERSELRKRK LLAEVTLKTYWVSKGSFSTTSISKQETELSPEMI SRFTEPTDNFISSFWNFDALFQPQQHPARDQ HDTFFLRDPAEALQLPMDYVQRVKRTHSQGGY GSQGYKYNWKLDEARKNLLRTHHTTSASARALY RLAQKKPFTPVKYFSIDRVFRNETLDATHLAEF HQIEGVVADHGLTLGHLMGVLEFFTKLGITQL RFKPAYNPYTEPSMEVFSYHQGLKKWVEVGNS GVFRPEMLLPMLPENVSVIAWGLSLERPTMIK YGINNIRELVGHKVNLMVYDSPLCRLDAEPRP PPTQEAA	SEQ.ID. NO.23
PheRSa1 <sup>N5</sup>	DNA / 人类	ATGGCGGATGGTCAGGTGGCGGAACTGCTGC TCCGCGCGGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC	SEQ.ID. NO.24

表 2  
通过深度测序鉴定的 AARS 多肽和替代转录物

名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
		ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCGTGTGTTTCGAAG CATTCCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCCGG GTGGACAAGAGTGCGGCTGACGGGCCCCGGG TGTTCCGAGTGGTGGACAGCATGGAGGATGA GGTGCAGCGGCGGCTCCAGCTGGTCCGGGGG GGACAGGCTGAGAAGCTGGGGGAGAAGGAG AGGAGCGAGCTGAGGAAGAGGAAGCTGTTG GCTGAAGTGACTCTGAAGACCTACTGGGTGA GCAAAGGCAGTGCCTTTAGTACCAGCATCTCC AAGCAAGAGACAGAGCTGAGCCCAGAGATGA TCTCCAGGTTACCGAGATGCCGACTGATAAC TTCATTGAGAGCTCCTTCTGGAACTTTGACGC CCTCTCCAGCCCCAGCAGCACCCAGCCCGT GACCAGCACGACACCTTCTTCCTTCGAGATCC AGCGGAGGCCCTGCAGCTCCCAATGGACTAT GTCCAGCGGGTCAAGCGGACCCACTCTCAGG GCGGCTACGGCTCACAGGGGTACAAGTATAAC TGGAAGCTGGACGAGGCCCGGAAAAACCTAC TGCGAACCCACACCACATCAGCCAGCGCCCG TGCGCTCTACCGCCTTGCCCAGAAGAAGCCCT TCACTCCGGTCAAGTACTTCTCCATCGACCGC GTATTCCGGAATGAGACCCTGGACGCCACGC ACCTGGCTGAGTTCCACCAGATCGAGGGCGT GGTGGCGGATCATGGTCTCACCTTGGGCCACC TCATGGGCGTTCTGCGGGAGTTCTTCACCAAG CTGGGTATCACGCAACTCCGCTTCAAGCCAGC CTACAACCCATACACAGAGCCCAGCATGGAG GTGTTCAAGTACCACCAAGGCCTGAAGAAGT GGGTGGAGGTCGGAAACTCGGGGGTCTCCG TCCAGAGATGCTGCTGCCCATGGGGCTTCCCG AGAACGTGTCGGTCATTGCCTGGGGCCTCTCC CTGGAGCGCCCAACGATGATCAAATATGGCAT CAACAATATCCGGGAGCTGGTGGGCCACAAG	

[0118]



表 2 通过深度测序鉴定的 AARS 多肽和替代转录物			
名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
		GTGAACCTGCAGATGGTGTATGACAGTCCCCT GTGCCGCCTGGATGCCGAGCCGAGGCCCCCT CCCACACAGGAGGCTGCGTGA	
PheRSa1 <sup>N6</sup>	蛋白 / 人类 / 1-168 + 31 aa	MADGQVAELLRRLEASDGGGLDSAELAAELGM EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFSKAMSNKWIRVDKSAADGPRVFRVVD MEDEVQRRLQLVRGGQAEKLGEKERSELRKRK LLAEVSSGGPAAPNGLCPAGQADPLSGRLRLTG VQV	SEQ.ID. NO.25
PheRSa1 <sup>N6</sup>	DNA / 人类	ATGGCGGATGGTCAGGTGGCGGAACTGCTGC TCCGGCGGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCCTGTGTTTCGAAG CATTCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCGG GTGGACAAGAGTGCGGCTGACGGGCCCCGGG TGTTCCGAGTGGTGGACAGCATGGAGGATGA GGTGCAGCGGCGGCTCCAGCTGGTCCGGGGG GGACAGGCTGAGAAGCTGGGGGAGAAGGAG AGGAGCGAGCTGAGGAAGAGGAAGCTGTTG GCTGAAGTATCCAGCGGAGGCCCTGCAGCTC CCAATGGACTATGTCCAGCGGGTCAAGCGGA CCCACTCTCAGGGCGGCTACGGCTCACAGGG GTACAAGTATAA	SEQ.ID. NO.26
PheRSa1 <sup>N7</sup>	蛋白 / 人 类 /1-399 + 426-508	MADGQVAELLRRLEASDGGGLDSAELAAELGM EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFSKAMSNKWIRVDKSAADGPRVFRVVD MEDEVQRRLQLVRGGQAEKLGEKERSELRKRK LLAEVTLKTYWVSKGSFSTISISKQETELSP SSGSRDRPFKPYNFLAHGVLPDSGHLHPLLK VRSQFRQIFLEMGFTEMPDNISSFWNFDALF QPQQHPARDQHDTFFLRDPAEALQLPMDYVQR	SEQ.ID. NO.27

[0119]

表 2 通过深度测序鉴定的 AARS 多肽和替代转录物			
名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
		VKRTHSQGGYGSQGYKYNWKLDEARKNLLRT HTTSASARALYRLAQQKPFTPVKYFSIDRVFRN ETLDATHLAEFHQIEGVVADHGLTLGHLMGVL REFFTKLGLKKWVEVGNSGVFRPEMLLPMGLP ENVSVIAWGLSLERPTMIKYGINNIRELVGHKV NLQMVYDSPLCRLDAEPRPPPTQEA	
PheRSa1 <sup>N7</sup>	DNA / 人类	ATGGCGGATGGTCAGGTGGCGGAAGTCTGCTGC TCCGCGGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCCTGTGTTTCGAAG CATTCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCGG GTGGACAAGAGTGCGGCTGACGGGCCCCGGG TGTTCCGAGTGGTGGACAGCATGGAGGATGA GGTGCAGCGGCGGCTCCAGCTGGTCCGGGGG GGACAGGCTGAGAAGCTGGGGGAGAAGGAG AGGAGCGAGCTGAGGAAGAGGAAGCTGTTG GCTGAAGTGAAGTCTGAAGACCTACTGGGTGA GCAAAGGCAGTGCCTTTAGTACCAGCATCTCC AAGCAAGAGACAGAGCTGAGCCCAGAGATGA TCTCCAGTGGCTCTTGGCGGGACCGGCCCTTC AAGCCCTACAACCTTCTTGGCCCACGGTGTCTC CCCCGACAGCGGCCACCTTACCCGCTGCTC AAGGTCCGCTCCCAGTTCCGACAGATCTTCTC GGAGATGGGGTTCACCGAGATGCCGACTGAT AACTTCATTGAGAGCTCCTTCTGGAACCTTGA CGCCCTCTTCCAGCCCAGCAGCACCCAGCC CGTGACCAGCACGACACCTTCTTCTTCGAGA TCCAGCGGAGGCCCTGCAGCTCCCAATGGAC TATGTCCAGCGGGTCAAGCGGACCCACTCTCA GGGCGGCTACGGCTCACAGGGGTACAAGTAT AACTGGAAGCTGGACGAGGCCCGGAAAAAC CTACTGCGAACCCACACCACATCAGCCAGCG	SEQ.ID. NO.28

[0120]

表 2  
通过深度测序鉴定的 AARS 多肽和替代转录物

名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
		CCCGTGCGCTCTACCGCCTTGCCCAGAAGAA GCCCTTCACTCCGGTCAAGTACTTCTCCATCG ACCGCGTATTCCGGAATGAGACCCTGGACGCC ACGCACCTGGCTGAGTTCCACCAGATCGAGG GCGTGGTGGCGGATCATGGTCTCACCTGGGC CACCTCATGGGCGTTCTGCGGGAGTTCTTAC CAAGCTGGGCCTGAAGAAGTGGGTGGAGGTC GGAAACTCGGGGTCTTCCGTCCAGAGATGC TGCTGCCCATGGGGCTTCCCGAGAACGTGTCTG GTCATTGCCTGGGGCCTCTCCCTGGAGCGCCC AACGATGATCAAATATGGCATCAACAATATCC GGGAGCTGGTGGGCCACAAGGTGAACCTGCA GATGGTGTATGACAGTCCCCTGTGCCGCTGG ATGCCGAGCCGAGGCCCCCTCCCACACAGGA GGCTGCGTGA	
PheRSal <sup>N8</sup>	蛋白 / 人 类 /1-242 + 31 aa	MADGQVAELLLRRLEASDGGGLDSAELAAELGM EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFASKAMSNKWIRVDKSAADGPRVFRVVD MEDEVQRRLQLVRGGQAEKLGKERSSELKRK LLAEVTLKTYWVSKGSFSTISISKQETELSP SSGSWRDRPFKPYNFLAHGVLPDSGHLHPLLK VRSQFRQIFLEMSSGGPAAPNGLCPAGQADPL SGRLRLTGTVQV	SEQ.ID. NO.29
PheRSal <sup>N8</sup>	DNA / 人类	ATGGCGGATGGTCAGGTGGCGGAAGTCTGCTGC TCCGCGGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCGTGTGTTTCAAG CATTCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCCGG GTGGACAAGAGTGC GGCTGACGGGCCCCGGG TGTTCAGAGTGGTGGACAGCATGGAGGATGA GGTGCAGCGGCGGCTCCAGCTGGTCCGGGGG GGACAGGCTGAGAAGCTGGGGGAGAAGGAG	SEQ.ID. NO.30

[0121]

[0122]

表 2 通过深度测序鉴定的 AARS 多肽和替代转录物			
名称	类型/ 物种/ 残基	氨基酸和核酸序列	SEQ.ID. NO.
		AGGAGCGAGCTGAGGAAGAGGAAGCTGTTG GCTGAAGTGACTCTGAAGACCTACTGGGTGA GCAAAGGCAGTGCCTTTAGTACCAGCATCTCC AAGCAAGAGACAGAGCTGAGCCCAGAGATGA TCTCCAGTGGCTCTTGGCGGGACCGGCCCTTC AAGCCCTACA ACTTCTTGGCCCACGGTGTCTC CCCCGACAGCGGCCACCTTCACCCGCTGCTC AAGGTCCGCTCCCAGTTCGACAGATCTTCTC GGAGATGGGATCCAGCGGAGGCCCTGCAGCT CCCAATGGACTATGTCCAGCGGGTCAAGCGG ACCCACTCTCAGGGCGGCTACGGCTCACAGG GGTACAAGTATAA	

[0123]

表 2B AARS 多肽独特剪接点			
名称	类型/ 物种	独特剪接点附近的氨基酸和核酸序列	SEQ.ID. NO.
FA-AS01	DNA / 人类 /	TGACGGGCCCCGGGTGTTCCGAGTG GACTCTGA AGACCTACTGGGTGAGC	SEQ.ID. NO.31
	蛋白 / 人类 /	DGPRVFRVDSEDLLGE	SEQ.ID. NO.32
FA-AS02	DNA / 人类 /	GGAAGAGGAAGCTGTTGGCTGAAGT TGGCTCTT GGCGGGACCGGCCCTTC	SEQ.ID. NO.33
	蛋白 / 人类 /	KRKLLAEVGSWRDRPF	SEQ.ID. NO.34
FA-AS03	DNA / 人类 /	AGCTGAGCCCAGAGATGATCTCCAG GTTACCCG AGATGCCGACTGATAAC	SEQ.ID. NO.35
	蛋白 / 人类 /	LSPEMISRFTEMPDND	SEQ.ID. NO.36
FA-AS04	DNA / 人类 /	GGAAGAGGAAGCTGTTGGCTGAAGT ATCCAGC GGAGGCCCTGCAGCTCCC	SEQ.ID. NO.37
	蛋白 / 人类 /	KRKLLAEVSSGGPAAP	SEQ.ID. NO.38
FA-AS05	DNA / 人类 /	CTGCGGGAGTTCTTCACCAAGCTGG GCCTGAAG AAGTGGGTGGAGGTCGG	SEQ.ID. NO.39
	蛋白 / 人类 /	LREFFTKLGLKKWVEV	SEQ.ID. NO.40
FA-AS06	DNA /	TCCGACAGATCTTCTGGAGATGGG ATCCAGCG	SEQ.ID.

[0124]

表 2B AARS 多肽独特剪接点			
名称	类型/物种	独特剪接点附近的氨基酸和核酸序列	SEQ.ID. NO.
	人类 /	GAGGCCCTGCAGCTCCC	NO.41
	蛋白 / 人类 /	RQIFLEMGSSGGPAAP	SEQ.ID. NO.42

[0125]

表 3 通过生物信息学鉴定的 AARS 多肽和核酸			
名称	类型/物种 / 残基	氨基酸和核酸序列	SEQ.ID. NO.
PheRSa1 <sup>N2</sup>	蛋白 / 人类 / 1-219	MADGQVAELLRRLEASDGGGLDSAELAEELGM EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFSKAMSNKWIRVDKSAADGPRVFRVVDSD MEDEVQRRLQLVRGGQAEKLGKERSELRKRK LLAEVTLKTYWVSKGSFSTSISKQETELSPEMI SSGSWRDRPFKPYNFLAHGVLP	SEQ.ID. NO.43
PheRSa1 <sup>N2</sup>	DNA/ 人类	ATGGCGGATGGTCAGGTGGCGGAAGTCTGCTGC TCCGGCGGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCGTGTGTTTCGAAG CATTCCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCCGG GTGGACAAGAGTGCGGCTGACGGGCCCCGGG TGTTCCGAGTGGTGGACAGCATGGAGGATGA GGTGCAGCGGCGGCTCCAGCTGGTCCGGGGG GGACAGGCTGAGAAGCTGGGGGAGAAGGAG AGGAGCGAGCTGAGGAAGAGGAAGCTGTTG GCTGAAGTGAAGTCTGAAGACCTACTGGGTGA GCAAAGGCAGTGCCTTTAGTACCAGCATCTCC AAGCAAGAGACAGAGCTGAGCCCAGAGATGA TCTCCAGTGGCTCTTGGCGGGACCGGCCCTTC AAGCCCTACAACCTTCTTGGCCCACGGTGTCTC CCCC	SEQ.ID. NO.44
PheRSa1 <sup>N9</sup>	蛋白 / 人类 /	MADGQVAELLRRLEASDGGGLDSAELAEELGM EHQAVVGAVKSLQALGEVIEAELRSTKHWELTA	SEQ.ID. NO.45

表 3 通过生物信息学鉴定的 AARS 多肽和核酸			
名称	类型/物种 /残基	氨基酸和核酸序列	SEQ.ID. NO.
	1-152	EGEEIAREGSHEARVFRSIPPEGLAQSELMRLPS GKVGFSKAMSNKWIRVDKSAADGPRVFRVDS MEDEVQRRLQLVRGGQAEKLG	
[0126] PheRSa1 <sup>N9</sup>	DNA/ 人 类	ATGGCGGATGGTCAGGTGGCGGAACTGCTGC TCCGCGGCTGGAGGCGTCTGATGGCGGCCT GGACAGCGCCGAGTTGGCGGCTGAGCTGGGC ATGGAGCACCAGGCGGTGGTGGGCGCCGTGA AGAGCCTTCAGGCGCTGGGCGAGGTCATCGA GGCTGAACTTCGGTCCACCAAGCACTGGGAG CTTACTGCGGAGGGCGAGGAGATTGCCCGGG AGGGCAGCCATGAGGCCCGTGTGTTTCGAAG CATTCCCCAGAGGGCCTGGCCCAGAGCGAG CTTATGCGACTGCCAGTGGCAAAGTGGGCTT CAGCAAGGCCATGTCCAACAAGTGGATTCGG GTGGACAAGAGTGCGGCTGACGGGCCCCGGG TGTTCCGAGTGGTGGACAGCATGGAGGATGA GGTGCAGCGGCGGCTCCAGCTGGTCCGGGGG GGACAGGCTGAGAAGCTGGGG	SEQ.ID. NO.46

[0127] C-端AARS多肽: (表4, 5&6)

表 4A 通过质谱鉴定的 AARS 多肽			
名称	类型/物种 /残基	氨基酸和核酸序列	SEQ.ID. NO.
[0128]			

表 4B 检测的质谱肽和推断的连接肽		
类型/物种	序列	SEQ.ID. NO.
[0129]		

表 4C 基于检测的质谱肽的连锁序列		
类型/物种	序列	SEQ.ID. NO.
[0130]		

表 5  
通过深度测序鉴定的 AARS 多肽和替代转录物

名称	类型/物种 /残基	氨基酸和核酸序列	SEQ.ID. NO.
PheRSa1 <sup>C2</sup>	蛋白 / 人类 / 53 aa + 169-508	MRPVCFEAFPQRAWPRASLDCDPVAKWASARP CPTSGFGWTRVRLTGPGCSEWTLKTYWVSKGS AFSTSISKQETELSPMISSGSWRDRPFKPYNFL AHGVLPDSGHLHPLLK VRSQFRQIFLEMGFTEM PTDNFISSFWNFDALFQPQQHPARDQHDTFFL RDPAEALQLPMDYVQRVKRTHSQGGYGSQGY KYNWKLDEARKNLLRTHHTTSASARALYRLAQQ KPFTPVKYFSIDRVFRNETLDATHLAEFHQIEGV VADHGLTLGHLMGVLRFFTKLGITQLRFKPAY NPYTEPSMEVFSYHQGLKKWVEVGNVSGVFRPE MLLPMGLPENVSVIAWGLSLERPTMIKYGINNI RELVGHKVNLMVYDSPLCRLDAEPRPPPTQE AA	SEQ.ID. NO.75
PheRSa1 <sup>C2</sup>	DNA / 人 类	ATGAGGCCCGTGTGTTTCGAAGCATTCCCCCA GAGGGCCTGGCCCAGAGCGAGCTTATGCGAC TGCCCAGTGGCAAAGTGGGCTTCAGCAAGGC CATGTCCAACAAGTGGATTCGGGTGGACAAG AGTGC GGCTGACGGGCCCCGGGTGTTCCGAG TGGACTCTGAAGACCTACTGGGTGAGCAAAG GCAGTGCCTTTAGTACCAGCATCTCCAAGCAA GAGACAGAGCTGAGCCCAGAGATGATCTCCA GTGGCTCTTGGCGGGACCGGCCCTTCAAGCC CTACAACCTTCTGGCCCACGGTGTCTCCCCG ACAGCGGCCACCTTCACCCGCTGCTCAAGGT CCGCTCCCAGTTCGACAGATCTTCTGGAGA TGGGGTTCACCGAGATGCCGACTGATAACTTC ATTGAGAGCTCCTTCTGGAACCTTGACGCCCT CTCCAGCCCCAGCAGCACCCAGCCCGTGAC CAGCACGACACCTTCTTCTTCGAGATCCAGC GGAGGCCCTGCAGCTCCAATGGACTATGTCC AGCGGGTCAAGCGGACCCACTCTCAGGGCGG CTACGGCTCACAGGGGTACAAGTATAACTGGA AGCTGGACGAGGCCCGAAAAACCTACTGCG AACCACACCACATCAGCCAGCGCCCGTGCG CTCTACCGCCTTGCCCAGAAGAAGCCCTTAC TCCGGTCAAGTACTTCTCCATCGACCGCGTAT TCCGGAATGAGACCCTGGACGCCACGCACCT GGCTGAGTTCCACCAGATCGAGGGCGTGGTG GCGGATCATGGTCTCACCTTGGGCCACCTCAT GGGCGTCTGCGGGAGTTCTTACCAAGCTG	SEQ.ID. NO.76

[0131]

表 5 通过深度测序鉴定的 AARS 多肽和替代转录物			
名称	类型/物种 /残基	氨基酸和核酸序列	SEQ.ID. NO.
		GGTATCACGCAACTCCGCTTCAAGCCAGCCTA CAACCCATACACAGAGCCCAGCATGGAGGTG TTCAGCTACCACCAAGGCCTGAAGAAGTGGG TGGAGGTTCGAAACTCGGGGGTCTTCCGTCC AGAGATGCTGCTGCCCATGGGGCTTCCCGAG AACGTGTCGGTCATTGCCTGGGGCCTCTCCCT GGAGCGCCCAACGATGATCAAATATGGCATCA ACAATATCCGGGAGCTGGTGGGCCACAAGGT GAACCTGCAGATGGTGTATGACAGTCCCCTGT GCCGCCTGGATGCCGAGCCGAGGCCCCCTCC CACACAGGAGGCTGCGTGA	
PheRSa1 <sup>C3</sup>	蛋白 / 人类 / 290-508	MDYVQRVKRTHSQGGYGSQGYKYNWKLDEA RKNLLRHTTSASARALYRLAQKKPFTPVKYFS IDRVFRNETLDATHLAEFHQIEGVVADHGLTLG HLMGVLREFFTKLGITQLRFKPAYNPYTEPSME VFSYHQGLKKWVEVGNSGVFRPEMLLPMGLPE NVSVIAWGLSLERPTMIKYGINNIRELVGHKVN LQMVYDSPLCRLDAEPRPPPTQEAA	SEQ.ID. NO.77
PheRSa1 <sup>C3</sup>	DNA / 人类	ATGGACTATGTCCAGCGGGTCAAGCGGACCC ACTCTCAGGGCGGCTACGGCTCACAGGGGTA CAAGTATAACTGGAAGCTGGACGAGGCCCGG AAAAACCTACTGCGAACCCACACCACATCAG CCAGCGCCCGTGCCTCTACCGCCTTGCCAG AAGAAGCCCTTCACTCCGGTCAAGTACTTCTC CATCGACCGCGTATTCCGGAATGAGACCCTGG ACGCCACGCACCTGGCTGAGTTCCACCAGAT CGAGGGCGTGGTGGCGGATCATGGTCTCACCT TGGGCCACCTCATGGGCGTTCTGCGGGAGTTC TTCACCAAGCTGGGTATCACGCAACTCCGCTT CAAGCCAGCCTACAACCCATACACAGAGCCC AGCATGGAGGTGTTCACTACCACCAAGGCC TGAAGAAGTGGGTGGAGGTTCGAAACTCGGG GGTCTTCCGTCCAGAGATGCTGCTGCCCATGG GGCTTCCCAGAACGTGTCGGTCATTGCCTGG GGCCTCTCCCTGGAGCGCCCAACGATGATCAA ATATGGCATCAACAATATCCGGGAGCTGGTGG GCCACAAGGTGAACCTGCAGATGGTGTATGA CAGTCCCCTGTGCCGCCTGGATGCCGAGCCG AGGCCCCCTCCCACACAGGAGGCTGCGTGA	SEQ.ID. NO.78

[0132]



[0133]

表 5B AARS 多肽独特剪接点			
名称	类型/物种	独特剪接点附近的氨基酸和核酸序列	SEQ.ID. NO.
FA-AS01	DNA / 人类 /	TGACGGGCCCCGGGTGTTCCGAGTG GACTCTG AAGACCTACTGGGTGAGC	SEQ.ID. NO.79
	蛋白 / 人类 /	TGPGCSEWTLKTYWVS	SEQ.ID. NO.80
FA-AS04	DNA / 人类 /	GGAAGAGGAAGCTGTTGGCTGAAGT ATCCAGC GGAGGCCCTGCAGCTCCC	SEQ.ID. NO.81
	蛋白 / 人类 /	N/A	

[0134]

表 6 通过生物信息学鉴定的 AARS 多肽和核酸			
名称	类型/物种/残基	氨基酸和核酸序列	SEQ.ID. NO.
PheRSa1 <sup>C1</sup>	蛋白 / 人类 / 459-508	LSLERPTMIKYGINNIRELVGHKVNLMVYDSP LCRLDAEPRPPTQAAA	SEQ.ID. NO.82
PheRSa1 <sup>C1</sup>	DNA/ 人类	CTCTCCCTGGAGCGCCCAACGATGATCAAATA TGGCATCAACAATATCCGGGAGCTGGTGGGCC ACAAGGTGAACCTGCAGATGGTGTATGACAG TCCCCTGTGCCGCCTGGATGCCGAGCCGAGG CCCCTCCACACAGGAGGCTGCGTGA	SEQ.ID. NO.83

[0135]

内部AARS多肽: (表7, 8&9)

[0136]

表 7A 通过质谱鉴定的 AARS 多肽			
名称	类型/物种/残基	氨基酸和核酸序列	SEQ.ID. NO.

[0137]

表 7B 检测的质谱肽和推断的连接肽		
类型/物种	序列	SEQ.ID. NO.

[0138]

表 7C 基于检测的质谱肽的连锁序列		
类型/物种	序列	SEQ.ID. NO.

表 8 通过深度测序鉴定的 AARS 多肽和替代转录物			
[0139]	名称	类型/ 物种/ 残基	氨基酸和核酸序列 SEQ.ID. NO.

表 8B AARS 多肽独特剪接点			
[0140]	名称	类型/ 物种	独特剪接点附近的氨基酸和核酸序列 SEQ.ID. NO.

表 9 通过生物信息学鉴定的 AARS 多肽和核酸			
[0141]	名称	类型/ 物种/ 残基	氨基酸和核酸序列 SEQ.ID. NO.

[0142] “蛋白片段”或蛋白片段的氨基酸序列，例如蛋白水解片段或剪接变体片段，可以根据许多技术来表征、鉴定或衍生。例如，可以通过诸如深度测序的技术来鉴定剪接变体（参见例如Xing等人，RNA.14:1470-1479,2008;和Zhang等人，Genome Research.17:503-509,2007）。作为进一步实例，例如通过将全长或其他AARS多肽与选择的蛋白酶一起孵育，可以在体外鉴定蛋白片段，例如蛋白水解片段，或者可以内源地（例如在体内）鉴定它们。在某些实施方案中，例如通过在选择的微生物或真核细胞中重组表达全长或其他AARS多肽，可以产生或鉴定蛋白片段，例如内源性蛋白水解片段，并且从其中分离并表征内源产生的蛋白片段，所述微生物或真核细胞已经被修饰为含有一种或多种选择的蛋白酶或者天然含有能够作用于选择的AARS多肽的一种或多种蛋白酶。

[0143] 在某些实施方案中，诸如内源（例如，天然存在的）蛋白水解片段的蛋白片段可以从以下级分产生或鉴定：例如各种细胞级分（例如，胞质、膜、核）和/或各种细胞类型的生长培养基，所述细胞类型包括例如，免疫细胞，诸如单核细胞、树突状细胞、巨噬细胞（例如，RAW264.7巨噬细胞）、嗜中性粒细胞、嗜酸性粒细胞、嗜碱性细胞，和淋巴细胞，例如B细胞和T细胞（例如，CD4+辅助细胞和CD8+杀伤细胞），包括原代T细胞和T细胞系，诸如Jurkat T细胞系，以及自然杀伤（NK）细胞等。

[0144] 在某些实施方案中，不论如何产生的诸如内源蛋白水解片段的蛋白片段可以通过诸如质谱的技术或等效技术来鉴定。一旦体外或内源鉴定的蛋白片段已经产生或被鉴定，它就可以被作图或测序，并且例如，被克隆进表达载体用于重组产生，或合成产生。

[0145] 大量的蛋白酶可以用于产生、鉴定、衍生或表征AARS蛋白片段例如蛋白水解片段的序列。一般而言，通常根据三个主要标准将蛋白酶分类：(i) 所催化的反应，(ii) 催化位点的化学性质，和(iii) 由结构揭示的进化关系。通过催化机理分类的蛋白酶或蛋白水解酶的一般实例包括天冬氨酸蛋白酶、丝氨酸蛋白酶、半胱氨酸蛋白酶和金属蛋白酶。

[0146] 大多数天冬氨酸蛋白酶属于胃蛋白酶家族。该家族包括消化酶，例如胃蛋白酶和凝乳酶，以及溶酶体组织蛋白酶D和诸如肾素的加工酶，和某些真菌蛋白酶（例如，青霉蛋白酶（penicillopepsin）、根霉菌蛋白酶（rhizopuspepsin）、壳室囊菌蛋白酶

(endothiapepsin)。天冬氨酸蛋白酶的第二个家族包括病毒蛋白水解酶,例如来自AIDS病毒(HIV)的蛋白酶,还被称为逆转录病毒蛋白酶(retropepsin)。

[0147] 丝氨酸蛋白酶包括两个不同的家族。第一,胰凝乳蛋白酶家族,包括哺乳动物酶,例如胰凝乳蛋白酶、胰蛋白酶、弹性蛋白酶和激肽释放酶,第二,枯草杆菌蛋白酶家族,包括诸如枯草杆菌蛋白酶的细菌酶。这两个家族之间的通用3D结构不同,但它们具有相同的活性位点几何构型,并且通过相同的机理进行催化。丝氨酸蛋白酶表现出不同的底物特异性、差异性,这主要与各个酶亚位点(底物残基作用位点)中的氨基酸取代有关。某些丝氨酸蛋白酶具有扩展的与底物的作用位点,而其他丝氨酸蛋白酶具有限于P1底物残基的特异性。

[0148] 半胱氨酸蛋白酶家族包括诸如木瓜蛋白酶、猕猴桃蛋白酶和菠萝蛋白酶的植物蛋白酶,数种哺乳动物溶酶体组织蛋白酶,胞质钙蛋白酶(钙激活的)以及数种寄生虫蛋白酶(例如,锥虫(*Trypanosoma*)、血吸虫(*Schistosoma*))。木瓜蛋白酶是原型,并且是该家族中研究最多的成员。对白介素-1- $\beta$ 转化酶的X射线结构的最新解释已经揭示半胱氨酸蛋白水解酶的一种新的折叠类型。

[0149] 金属蛋白酶是较老的蛋白酶种类中的一种,在细菌、真菌和高等生物中发现。它们在序列和3D结构上差异显著,但大多数酶都含有具有催化活性的锌原子。在某些情况下,锌可以被诸如钴或镍的另一金属替换而不丧失蛋白水解活性。细菌嗜热菌蛋白酶已被良好表征并且其结晶结构显示锌由两个组氨酸和一个谷氨酸结合。许多金属蛋白酶含有序列基序HEXXH,其为锌提供两个组氨酸配体。第三配体是谷氨酸(嗜热菌蛋白酶、脑啡肽酶、丙氨酰氨肽酶)或组氨酸(虾红素、粘质沙雷氏菌酶(*serralysin*))。

[0150] 示例性的蛋白酶包括例如,无色肽酶、氨肽酶、安克洛酶、血管紧张素转化酶、菠萝蛋白酶、钙蛋白酶、钙蛋白酶I、钙蛋白酶II、羧肽酶A、羧肽酶B、羧肽酶G、羧肽酶P、羧肽酶W、羧肽酶Y、胰天蛋白酶1、胰天蛋白酶2、胰天蛋白酶3、胰天蛋白酶4、胰天蛋白酶5、胰天蛋白酶6、胰天蛋白酶7、胰天蛋白酶8、半胰天冬酶9、胰天蛋白酶10、胰天蛋白酶11、胰天蛋白酶12、胰天蛋白酶13、组织蛋白酶B、组织蛋白酶C、组织蛋白酶D、组织蛋白酶E、组织蛋白酶G、组织蛋白酶H、组织蛋白酶L、木瓜凝乳蛋白酶、胃促胰酶、胰凝乳蛋白酶、梭菌蛋白酶、胶原酶、补体C1r、补体C1s、补体因子D、补体因子I、黄瓜素、二肽基肽酶IV、弹性蛋白酶(白细胞)、弹性蛋白酶(胰腺)、内蛋白酶Arg-C、内蛋白酶Asp-N、内蛋白酶Glu-C、内蛋白酶Lys-C、肠激酶、因子Xa、无花果蛋白酶、弗林蛋白酶、粒酶A、粒酶B、HIV蛋白酶、IGase、组织激肽释放酶、亮氨酸氨基肽酶(总的)、亮氨酸氨基肽酶(细胞溶质的)、亮氨酸氨基肽酶(微粒体的)、基质金属蛋白酶、蛋氨酸氨基肽酶、中性蛋白酶、木瓜蛋白酶、胃蛋白酶、纤维蛋白溶酶、氨酰基脯氨酸二肽酶、链酶蛋白酶E、前列腺特异性抗原、来自灰色链霉菌(*Streptomyces griseus*)的碱性蛋白酶、来自曲霉属(*Aspergillus*)的蛋白酶、来自佐氏曲霉(*Aspergillus saitoi*)的蛋白酶、来自酱油曲霉(*Aspergillus sojae*)的蛋白酶、蛋白酶(地衣芽孢杆菌(*B.licheniformis*)) (碱性的或碱性蛋白酶)、来自多黏芽孢杆菌(*Bacillus polymyxa*)的蛋白酶、来自芽孢杆菌(*Bacillus sp.*)的蛋白酶、来自根霉菌(*Rhizopus sp.*)的蛋白酶、蛋白酶S、蛋白酶体、来自米曲霉(*Aspergillus oryzae*)的蛋白酶、蛋白酶3、蛋白酶A、蛋白酶K、蛋白C、焦谷氨酸氨基肽酶、凝乳酶、链激酶、枯草杆菌蛋白酶、嗜热菌蛋白酶、凝血酶、组织纤维蛋白溶酶原激活剂、胰蛋白酶、类胰蛋白酶和尿激酶。

[0151] 某些实施方案涉及分离的AARS多肽,包含所述片段的药物组合物,和使用它们的

方法,所述分离的AARS多肽包含来源于内源天然存在的AARS多肽片段的氨基酸序列,或基本上由所述氨基酸序列组成或由所述氨基酸序列组成。这些和相关的实施方案可以在体内、离体和/或体外产生或鉴定。在某些优选的体外实施方案中,通过将AARS多肽例如全长AARS多肽与一种或多种分离的人蛋白酶孵育而产生或鉴定AARS蛋白水解片段,这些蛋白酶的大部分是人类内源性或天然的,例如本文描述和本领域已知的弹性蛋白酶和其他蛋白酶。某些实施方案涉及分离的AARS多肽,包含所述片段的药物组合物,和使用它们的方法,所述分离的AARS多肽包含来源于内源天然存在的AARS剪接变体的氨基酸序列,或基本上由所述氨基酸序列组成或由所述氨基酸序列组成。基本上,AARS蛋白片段可以分离自已经在体内或体外暴露于蛋白酶的样品。

[0152] 在某些实施方案中,可以通过诸如质谱的技术或等效技术来鉴定AARS蛋白片段。仅作为示例而非限制,在某些实施方案中,来自许多生理状态(例如,缺氧、饮食、年龄、疾病)的各种细胞类型、组织或体液的蛋白质组或其级分可以通过1D SDS-PAGE进行分离并按固定间隔将凝胶泳道切成条带;然后该条带可以任选地用诸如胰蛋白酶的合适的蛋白酶消化以释放肽,该肽随后可以通过1D反相LC-MS/MS来分析。所得的蛋白质组数据可被整合成所谓的肽图(peptograph),其中在左图中,相对于垂直维度中的SDS-PAGE迁移(从上到下是分子量从高到低),以水平维度绘出指定蛋白的序列范围(从左到右是N端到C端)。然后可以对具体肽片段进行测序或定位。在某些实施方案中,AARS参照片段可以通过其例如与相应的全长AARS的分子量相比的独特的分子量来表征。

[0153] 如上所述,本发明的实施方案包括表1-3或表4-6或表7-9所列的AARS多肽。还包括AARS参考多肽的“变体”。表述多肽“变体”指通过至少一个氨基酸残基的添加、缺失和/或取代而区别于参照AARS多肽的多肽,其通常保持(例如,模拟)或调解(例如拮抗)参考AARS多肽的一种或多种非常规活性。

[0154] 而且,人苯丙氨酰-tRNA合成酶包括数百种高度相关的多态形式,这些在本领域已知是至少部分功能上可互换的。因此,选择天然存在的苯丙氨酰-tRNA合成酶变体(包括例如表A所列单核苷酸多态形式)以产生基于人类以及其他物种的苯丙氨酰-tRNA合成酶的同源物、直系同源物和天然存在的同种型的任何一种的序列含有一个或多个氨基酸变化的AARS多肽是常规的事情。

**表 A**  
**人类苯丙氨酰- $\alpha$ tRNA 合成酶 SNP**

Gene Bank 登录号	核苷酸变化	Gene Bank 登录号	核苷酸变化
rs117997664	G/T	rs78379286	-/A
rs117532528	A/C	rs78214285	C/G
rs117397150	C/T	rs78009727	C/G
rs117345957	A/G	rs77667575	C/G
rs117111613	A/G	rs77471408	G/T
rs116999707	C/T	rs77464836	A/G
rs116992783	C/T	rs77258442	A/G
rs116324221	A/T	rs77136463	A/G
rs116144034	G/T	rs77092376	C/T
rs115925430	C/T	rs76848673	A/G
rs115255943	C/T	rs76720345	A/T
rs115117231	A/C	rs76491506	A/C
rs114790582	A/G	rs75098395	G/T
rs114717980	C/T	rs74181681	A/G
rs114495724	A/G	rs73925236	A/G
rs114312478	A/G	rs73004683	A/T
rs114095713	A/G	rs71168639	-/AA
rs113987774	C/T	rs71168638	-/A
rs113870503	C/T	rs71168637	-/TTT
rs113802496	C/G	rs71168636	-/TT
rs113779830	A/G	rs66630592	A/G
rs113375740	-/T	rs62109865	A/G
rs113324767	A/T	rs61737507	G/T
rs113303874	C/T	rs61242190	C/G
rs112863124	A/T	rs60848235	A/C

[0155]

**表 A**  
**人类苯丙氨酰- $\alpha$ tRNA 合成酶 SNP**

Gene Bank 登录号	核苷酸变化	Gene Bank 登录号	核苷酸变化
rs112443302	A/G	rs60167882	A/G
rs112332431	A/G	rs59533321	A/G
rs112235079	-/A	rs57449367	A/T
rs111851259	C/T	rs55923989	A/G
rs111757774	A/G	rs45512394	A/C
rs111730933	A/C	rs36074715	-/G
rs111325496	C/G	rs35705523	-/T
rs111317860	A/C	rs35636827	-/G
rs79300376	C/T	rs35503400	-/C
rs79130552	A/C	rs35498150	-/A
rs78937739	A/G	rs35087277	C/T
rs78601726	A/G	rs34795408	C/T
rs34586259	-/AAA	rs5016037	A/G
rs34118249	A/G	rs4542773	A/G
rs34026093	-/A	rs3111317	A/G
rs33925420	A/G	rs3111316	C/T
rs33919090	-/A	rs2974750	A/C
rs16978813	C/T	rs2974749	A/G
rs12151216	C/T	rs2967893	A/C
rs11673346	C/T	rs2967890	C/T
rs11538254	C/T	rs2967889	G/T
rs11538253	A/G	rs2293683	A/G
rs10419627	A/G	rs2009222	C/T
rs8108826	A/G	rs2009218	A/G
rs8107173	A/G	rs1804567	C/T
rs7508226	A/G	rs1045913	C/T
rs7508225	A/G	rs7252705	A/G

[0156]

[0157] 在某些实施方案中,多肽变体通过一个或多个取代而区别于参照多肽,所述取代可以是本文所述的并且是本领域内已知的保守或非保守取代。在某些实施方案中,多肽变体包含保守取代,在这方面,本领域熟知一些氨基酸可以改变为具有广泛相似特性的其他氨基酸,而不会改变多肽的活性性质。

[0158] 在某些实施方案中,变体多肽包括与本文描述的AARS参考多肽的相应序列具有至少约50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或更大序列同一性或相似性的氨基酸序列,并且基本上保持所述参考多肽的非常规活性。还包括通过1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、110、120、130、140、150或更多个氨基酸的添加、缺失或取代而区别于参照AARS序列,但保持参照AARS多肽性质的序列。在某些实施方案中,氨基酸添加或缺

失发生在AARS参照多肽的C端和/或N端。在某些实施方案中,氨基酸添加包括邻近这些AARS参照多肽的C端和/或N端的1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50或更多个野生型残基(即,来自相应的全长AARS多肽)。

[0159] 在某些实施方案中,变体多肽与相应的AARS参照序列相差至少1%但少于20%、15%、10%或5%的残基。(如果这种比较需要比对,则序列应该进行最大相似性的比对。从缺失或插入或错配“圈”出的序列被认为是差异)。适宜地,差异是在非必需残基处的差异或变化或保守取代。在某些实施方案中,变体AARS多肽的分子量与AARS参照多肽相差约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%或更多。

[0160] 还包括AARS参照多肽的生物活性“片段”,即AARS蛋白片段的生物活性片段。代表性的生物活性片段通常参与相互作用,例如分子内的或分子间的相互作用。分子间相互作用可以是特异性结合相互作用或酶促相互作用。分子间相互作用可以在AARS多肽与细胞结合配体之间发生,所述细胞结合配体例如参与AARS多肽的非常规活性的细胞受体或其他宿主分子。在一些实施方案中,AARS蛋白、其变体和生物活性片段以至少约0.01、0.05、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40或50nM的亲和力结合一种或多种细胞结合配体。AARS蛋白片段对选择的细胞结合配体、特别是参与非常规活性的结合配体的结合亲和力通常比对应于全长AARS多肽的AARS蛋白片段强至少约1.5x、2x、2.5x、3x、3.5x、4x、4.5x、5x、6x、7x、8x、9x、10x、15x、20x、25x、30x、40x、50x、60x、70x、80x、90x、100x、200x、300x、400x、500x、600x、700x、800x、900x、1000x或更多(包括其间所有整数)。AARS蛋白片段对参与AARS的至少一种常规活性的结合配体的结合亲和力通常比对应于全长AARS多肽的AARS蛋白片段弱至少约1.5x、2x、2.5x、3x、3.5x、4x、4.5x、5x、6x、7x、8x、9x、10x、15x、20x、25x、30x、40x、50x、60x、70x、80x、90x、100x、200x、300x、400x、500x、600x、700x、800x、900x、1000x或更多。

[0161] 通常,生物活性片段包含具有AARS参照多肽的至少一种活性的结构域或基序,并且可以包括各个活性结构域的一个或多个(在一些情况下,是全部)并且包括具有非常规活性的片段。在一些情况下,AARS多肽的生物活性片段具有对于特定的截短的片段独特的生物活性,使得全长AARS多肽可能没有该活性。在一些情况下,生物活性可以通过将生物活性AARS多肽片段与其他的全长AARS多肽序列分离,或者通过改变全长AARS野生型多肽序列的某些残基以暴露生物活性结构域而被揭示。

[0162] AARS参照多肽的生物活性片段可以是这样的多肽片段,该多肽片段例如是本文描述的AARS参照多肽任何一个中所示的氨基酸序列的10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、110、120、130、140、150、160、170、180、190、200、220、240、260、280、300、320、340、360、380、400、450、500、550、600、650、700、750个或更多个连续或非连续的氨基酸,包括其间所有整数(例如101、102、103)和范围(例如,50-100、50-150、50-200),但通常不包括全长AARS。在某些实施方案中,生物活性片段包括非常规活性相关的序列、结构域或基序。在某些实施方案中任何AARS参照多肽的C端或N端区域可以被截短约1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、250、300、350、

400、450、500、550、600、650或700或更多氨基酸,或截短约10-50、20-50、50-100、100-150、150-200、200-250、250-300、300-350、350-400、400-450、450-500、500-550、550-600、600-650、650-700或更多氨基酸,包括其间所有整数和范围(例如,101、102、103、104、105),只要截短的AARS多肽保持参照多肽的非常规活性。通常,生物活性片段具有衍生它的生物活性(即,非常规活性)AARS参照多肽的活性的不小于约1%、5%、约10%、约25%或约50%。测量这种非常规活性的示例性方法描述于实施例。

[0163] 如上所述AARS多肽可以以不同方式改变,包括氨基酸取代、缺失、截短和插入。用于这种操纵的方法通常是本领域已知的。例如,AARS参照多肽的氨基酸序列变体可以通过DNA中的突变制备。用于诱变和核苷酸序列改变的方法是本领域已知的。参见例如Kunkel(1985,Proc.Natl.Acad.Sci.USA.82:488-492),Kunkel等人(1987,Methods in Enzymol,154:367-382),美国专利号4,873,192,Watson,J.D.等人(“Molecular Biology of the Gene”,第四版,Benjamin/Cummings,Menlo Park,Calif.,1987)以及其中引用的参考文献。关于不影响感兴趣蛋白的生物活性的合适氨基酸取代的指导可以见于Dayhoff等人,(1978)Atlas of Protein Sequence and Structure(Natl.Biomed.Res.Found.,Washington,D.C.)的模型中。

[0164] 类似地,如果使用AARS多肽产生免疫原性问题,则解决和/或减少免疫原性在本领域技术范围内,例如,通过使用自动化计算机识别程序来鉴定潜在的T细胞表位,以及鉴定较小免疫原性形式的定向进化方法。

[0165] 本领域已知用于筛选通过点突变或截短构成的组合文库基因产物的方法以及用于筛选cDNA文库中具有选定特性的基因产物的方法。这些方法适用于通过AARS多肽的组合诱变产生的基因文库的快速筛选。增强文库中的功能突变体频率的技术,递归集合诱变(Recursive ensemble mutagenesis)(REM)可以与筛选测定组合用于鉴定AARS多肽变体(Arkin和Yourvan(1992)Proc.Natl.Acad.Sci.USA89:7811-7815;Delgrave等人,(1993)Protein Engineering,6:327-331)。保守取代,诸如将一个氨基酸与具有相似特性的另一个氨基酸交换,可以是理想的,如下文更详细地讨论的。

[0166] 与参照AARS氨基酸序列相比较,生物活性截短的和/或变体AARS多肽可以在沿它们序列不同位置包含保守氨基酸取代。此外,AARS蛋白的天然存在的变体已经被测序,并且在本领域已知是至少部分功能上可互换的。因此,基于人类以及AARS蛋白的其他物种的已知AARS蛋白同源物、直系同源物和天然存在的同种型中天然存在的序列变化来选择将保守或非保守突变引入AARS多肽的氨基酸位置是常规的事情。

[0167] “保守氨基酸取代”是其中氨基酸残基被具有相似侧链的氨基酸残基替代的取代。具有类似侧链的氨基酸残基家族已在本领域中被定义,其通常可以如下进行亚分类:

[0168] 酸性的:残基由于丢失了H离子在生理pH下带负电荷,以及该残基被水性溶液吸引而使它在包含它的肽在生理pH的水性介质中寻找该肽构象的表面位置。具有酸性侧链的氨基酸包括谷氨酸和天冬氨酸。

[0169] 碱性的:残基由于结合H离子在生理pH下或其一个或两个pH单位内带正电荷(例如,组氨酸),以及该残基被水性溶液吸引以使它在包含它的肽在生理pH的水性介质中时寻找该肽构象的表面位置。具有碱性侧链的氨基酸包括精氨酸、赖氨酸和组氨酸。

[0170] 带电荷的:残基在生理pH下带电荷,并且因此包括具有酸性或碱性侧链的氨基酸



(即,谷氨酸、天冬氨酸、精氨酸、赖氨酸和组氨酸)。

[0171] 疏水的:残基在生理pH下不带电荷,并且该残基在水性溶液中被排斥以使它在包含它的肽在水性介质中时寻求该肽构象的内部位置。具有疏水侧链的氨基酸包括酪氨酸、缬氨酸、异亮氨酸、亮氨酸、甲硫氨酸、苯丙氨酸和色氨酸。

[0172] 中性/极性的:残基在生理pH下不带电荷,但该残基不足以被水性溶液排斥以使它会在包含它的肽在水性介质中时寻求该肽构象的内部位置。具有中性/极性侧链的氨基酸包括天冬酰胺、谷氨酰胺、半胱氨酸、组氨酸、丝氨酸和苏氨酸。

[0173] 该说明还将某些氨基酸表征为“小的”,因为它们的侧链不足够大到提供疏水性,即便缺乏极性基团。除脯氨酸外,“小”氨基酸是具有4个或更少的碳的那些,其时至少一个极性基团在侧链上,另三个碳或更少不在侧链上。具有小侧链的氨基酸包括甘氨酸、丝氨酸、丙氨酸和苏氨酸。基因编码的仲氨基酸脯氨酸是特例,由于它已知的对肽链的二级构象的影响。脯氨酸的结构与所有其他天然存在的氨基酸的不同在于它的侧链结合到 $\alpha$ -氨基基团的氮,以及 $\alpha$ -碳。然而,几个氨基酸相似矩阵(例如,PAM120矩阵和PAM250矩阵,如例如Dayhoff等人,1978,A model of evolutionary change in proteins),Atlas of Protein sequence and structure,第5卷,pp.345-358,National Biomedical Research Foundation,Washington DC;以及Gonnet等人,(Science,256:14430-1445,1992)将脯氨酸与甘氨酸、丝氨酸、丙氨酸和苏氨酸包括在同一组中。因此,出于本发明的目的,脯氨酸被分类为“小”氨基酸。

[0174] 用于极性或非极性分类所需的吸引或排斥程度是随机的,因此,本发明特别包括的氨基酸已经被分类为一种或另一种。没有具体命名的大多数氨基酸可以基于已知的行为进行分类。

[0175] 氨基酸残基可以进一步亚分类为环状的或非环状的,芳香族的或非芳香族的,与残基侧链取代基因有关的自明的分类,以及分类为小的或大的。如果残基包括羧基碳在内总共含4个碳原子或更少,被视为小的,但是存在额外的极性取代基;如果不存在额外的极性取代基,则总共含三个碳原子或更少。当然,小残基总是非芳香族的。取决于它们的结构特性,氨基酸残基可以分为两类或更多类。对于天然存在的蛋白氨基酸,根据该方案的亚分类呈现在表B中。

[0176] 表B:氨基酸亚类

亚类	氨基酸
酸性的	天冬氨酸、谷氨酸
碱性的	非环状的：精氨酸、赖氨酸；环状的：组氨酸
带电的	天冬氨酸、谷氨酸、精氨酸、赖氨酸、组氨酸
小的	甘氨酸、丝氨酸、丙氨酸、苏氨酸、脯氨酸
极性的/中性的	天冬酰胺、组氨酸、谷氨酰胺、半胱氨酸、丝氨酸、苏氨酸
[0177]	
极性的/大的	天冬酰胺、谷氨酰胺
疏水性的	酪氨酸、缬氨酸、异亮氨酸、亮氨酸、蛋氨酸、苯丙氨酸、色氨酸
芳香族的	色氨酸、酪氨酸、苯丙氨酸
影响侧链取向的	甘氨酸和脯氨酸
残基	

[0178] 保守性氨基酸取代还包括基于侧链的分组。例如，具有脂肪族侧链的氨基酸组是甘氨酸、丙氨酸、缬氨酸、亮氨酸和异亮氨酸；具有脂肪族-羟基侧链的氨基酸组是丝氨酸和苏氨酸；具有含酰胺侧链的氨基酸组是天冬酰胺和谷氨酰胺；具有芳香侧链的氨基酸组是苯丙氨酸、酪氨酸和色氨酸；具有碱性侧链的氨基酸组是赖氨酸、精氨酸和组氨酸；以及具有含硫侧链的氨基酸组是半胱氨酸和甲硫氨酸。例如，预期异亮氨酸或缬氨酸取代亮氨酸、谷氨酰胺取代天冬酰胺、丝氨酸取代苏氨酸、或者用结构相关的氨基酸取代氨基酸的类似取代不会对产生的变体多肽的特性产生大的影响是合理的。如本文所描述的，氨基酸变化是否产生功能的截短和/或变体AARS多肽可以通过检测它的活性容易地确定。保守取代显示在标题为“示例性取代”的表C中。落入本发明范围的氨基酸取代一般通过选择在它们以下效应中不会显著不同的取代来实现：维持(a)取代区域内的肽骨架的结构，(b)分子在靶部位的电荷或疏水性，或(c)侧链的体积。引入取代后，根据生物活性筛选变体。

[0179] 表C：示例性氨基酸取代

原始残基	示例性取代	优选取代
Ala	Val, Leu, Ile	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln, His, Lys, Arg	Gln
Asp	Glu	Glu
Cys	Ser	Ser
Gln	Asn, His, Lys,	Asn
Glu	Asp, Lys	Asp
Gly	Pro	Pro
His	Asn, Gln, Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe, Norleu	Leu
Leu	Norleu, Ile, Val, Met, Ala, Phe	Ile
Lys	Arg, Gln, Asn	Arg

原始残基	示例性取代	优选取代
Met	Leu, Ile, Phe	Leu
Phe	Leu, Val, Ile, Ala	Leu
Pro	Gly	Gly
Ser	Thr	Thr
Thr	Ser	Ser
Trp	Tyr	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	Ile, Leu, Met, Phe, Ala, Norleu	Leu

[0182] 可选择地,用于产生保守性取代的类似氨基酸可以基于侧链的鉴定分为三类。第一组包括谷氨酸、天冬氨酸、精氨酸、赖氨酸、组氨酸,其都具有带电荷的侧链;第二组包括甘氨酸、丝氨酸、苏氨酸、半胱氨酸、酪氨酸、谷氨酰胺、天冬酰胺;以及第三组包括亮氨酸、异亮氨酸、缬氨酸、丙氨酸、脯氨酸、苯丙氨酸、色氨酸、甲硫氨酸,如Zubay, G., Biochemistry, third edition, Wm.C. Brown Publishers (1993) 所描述的。

[0183] 因此,在截短的和/或变体AARS多肽中的预测的非必需氨基酸残基通常被来自同一侧链家族的另一氨基酸残基取代。可选择地,突变可以沿着AARS编码序列的全部或部分随机引入,诸如通过饱和诱变,以及可以筛选产生的突变体的母体多肽的活性以鉴定保持该活性的突变体。编码序列诱变后,可以重组表达编码的肽,以及确定肽的活性。“非必需”

氨基酸残基是能够从实施方案多肽的野生型序列改变的残基,而没有消除或基本上改变其一种或多种活性。适宜地,改变基本上不会消除这些活性的一种,例如,活性至少为野生型的20%、40%、60%、70%或80%、100%、500%、1000%或更多。“必需的”氨基酸残基是当从参照截短的多肽的野生型序列改变后,导致母体分子的活性消除的残基,致使存在的野生型活性不到20%。例如,这类必需氨基酸残基包括在不同种类间的AARS多肽中保守的那些,包括在来自不同来源的AARS多肽的血小板生成刺激结合部位或基序中保守的那些序列。

[0184] 一般而言,多肽和融合多肽(以及它们的编码核苷酸)是分离的。“分离的”多肽或多核苷酸是从其原始环境中移出的那些。例如,如果与天然系统中的共存在物质的一些或全部分开,则天然存在的蛋白是分离的。优选地,这种多肽是至少约90%纯的,更优选地,至少约95%纯的,以及最优选地,至少约99%纯的。例如,如果多核苷酸被克隆到不是天然环境的一部分的载体中,它被认为是分离的。

[0185] 某些实施方案还包括AARS多肽的二聚体。二聚体可以包括,例如两个相同的AARS多肽间的同二聚体、两个不同AARS多肽的异二聚体(例如,全长的YRS多肽和截短的YRS多肽;截短的YRS多肽和截短的WRS多肽),和/或AARS多肽以及异源多肽间的异二聚体。如本文所述的,某些异二聚体,诸如AARS多肽和异源多肽间的那些,可以是双官能的。

[0186] 还包括AARS多肽的单体,包括分离的AARS单体,所述分离的AARS单体无论是否由于一个或多个取代、截短、缺失、添加、化学修饰或这些改变的组合,都基本上不与第二种AARS多肽二聚化。在某些实施方案中,单体AARS多肽具有生物活性,包括二聚体或多聚体AARS多肽复合物不具有的非常规活性。

[0187] 如本文所述,本发明的某些实施方案还考虑修饰的AARS多肽的用途,包括改善AARS多肽所需特性的修饰。本发明AARS多肽的示例性修饰包括,但不限于,一个或多个组成氨基酸的化学和/或酶促衍生作用,包括侧链修饰、主链修饰和N端与C端修饰,这些修饰包括乙酰化、羟基化、甲基化、酰胺化;和碳水化合物或脂质部分、辅因子等的连接。示例性修饰还包括AARS多肽的聚乙二醇化(参见,例如,Veronese和Harris,Advanced Drug Delivery Reviews54:453-456,2002;和Pasut等人,Expert Opinion.Ther.Patents14(6)859-894 2004,其通过引用并入本文)。

[0188] PEG是在水中和在许多有机溶剂中具有溶解性、缺少毒性并缺少免疫原性的公知聚合物。它也是澄清、无色、无味和化学上稳定的。处于这些原因和其他原因,PEG已经被选作连接的优选聚合物,但是它仅仅被用来示例说明而非限制。可以使用其他水溶性聚合物获得类似产物,包括但不限于聚乙烯醇,其他聚(环氧烷)例如聚(丙二醇)等,聚(乙氧基化多元醇)例如聚(乙氧基化甘油)等,羧甲基纤维素,葡聚糖,聚乙烯醇,聚乙烯吡咯烷酮,聚-1,3-二氧戊环,聚-1,3,6-三氧杂环己烷,乙烯/马来酸酐,和聚氨基酸。本领域技术人员能够基于所需剂量、循环时间、对蛋白水解的抗性和其他考虑因素来选择所需聚合物。

[0189] 具体地,许多PEG衍生物既是可获得的,也适合用于制备PEG-轭合物。例如,以商标SUNBRIGHT®系列销售的NOF Corp.的PEG试剂提供了许多PEG衍生物,包括甲氧基聚乙二醇和活化的PEG衍生物,例如甲氧基-PEG胺,马来酰亚胺,N-羟基琥珀酰亚胺酯,和羧酸,通过各种方法与AARS多肽的N端、C端或任何内部氨基酸连接。Nektar Therapeutics的先进的聚乙二醇化技术也提供了多种PEG-偶联技术而可能改善基于AARS多肽的治疗剂的安全性和效力。

[0190] 对专利、公布的专利申请和相关出版物的检索也会为阅读本公开的本领域技术人员提供显著可能的PEG-偶联技术和PEG-衍生物。例如,美国专利号6,436,386;5,932,462;5,900,461;5,824,784;和4,904,584;其内容通过引用并入本文,描述了此类技术和衍生物以及它们的生产方法。

[0191] 在某些方面,可以利用化学选择性连接技术来修饰本发明的AARS多肽,诸如通过以位点特异性和受控的方式连接聚合物。这种技术通常依赖于通过化学的或重组方式将化学选择性锚定物并入蛋白骨架,以及用携带互补连接物的聚合物进行随后的修饰。因此,可以控制组装过程和产生的蛋白-聚合物缀合物的共价结构,这使得诸如功效和药物代谢动力学特性的药物性质的合理优化成为可能(参见,例如,Kochendoerfer, *Current Opinion in Chemical Biology*9:555-560,2005)。

[0192] 在其他实施方案中,还包括AARS多肽与其他蛋白的融合蛋白,并且这些融合蛋白可以增加AARS多肽的生物活性、分泌、靶向、生物寿命、穿过细胞膜或血脑屏障的能力、或药代动力学性质。改善药代动力学性质的融合蛋白的实例(“PK调节剂”)包括但不限于,与人白蛋白(Osborn等人:*Eur. J. Pharmacol.* 456(1-3):149-158, (2002))、抗体Fc结构域、聚Glu或聚Asp序列、和运铁蛋白的融合体。此外,与由氨基酸Pro、Ala和Ser(‘PAS化’)构成的构象上混乱的多肽序列或羟乙基淀粉的融合体(以商标 HESYLATION®销售)提供了增加AARS多肽的流体动力学体积的简单方式。这种额外的延伸采用了庞大的随机结构,这显著增加了所得到的融合蛋白的大小。通过这种方式,通常快速的小AARS多肽经由肾过滤的清除被延迟几个数量级。此外,还已经显示IgG融合蛋白的使用使一些融合蛋白能够穿过血脑屏障(Fu等人, (2010) *Brain Res.* 1352:208-13)。

[0193] 改善跨细胞膜穿过的融合蛋白的实例包括与膜转位序列的融合体。在此上下文中,术语“膜转位序列”指天然存在的和合成的氨基酸序列,其能够跨细胞膜进行膜转位。代表性的膜转位序列包括基于源自Tat蛋白和同源转录蛋白Antennapedia的天然存在的膜转位序列的那些,以及完全或部分基于聚精氨酸和赖氨酸残基的合成膜转位序列。代表性的膜转位序列包括例如以下专利中公开的那些:US5,652,122;US5,670,617;US5,674,980;US5,747,641;US5,804,604;US6,316,003;US7,585,834;US7,312,244;US7,279,502;US7,229,961;US7,169,814;US7,453,011;US7,235,695;US6,982,351;US6,605,115;US7,306,784;US7,306,783;US6,589,503;US6,348,185;US6,881,825;US7,431,915;W00074701A2;W02007111993A2;W02007106554A2;W002069930A1;W003049772A2;W003106491A2;和W02008063113A1。

[0194] 要理解,柔性分子连接体(或间隔区)任选地可以插入其间,并且共价连接AARS多肽和本文公开的融合蛋白的任何一个。

[0195] 此外,在某些实施方案中,AARS多肽可以包括合成的或天然存在的分泌信号序列,源自其他公知的分泌蛋白。在一些实施方案中,此类蛋白可以通过蛋白水解裂解来加工,以原位形成AARS多肽。此类融合蛋白包括例如AARS多肽与以下的融合体:泛素,以提供新的N端氨基酸,或者使用分泌信号以介导AARS多肽高水平分泌进入细胞外介质,或者N端或C端表位标签以改善纯化或检测。

[0196] 本文所述的AARS多肽可以通过本领域人员已知的任何合适的方法制备,例如通过重组技术。除了重组产生方法,还可以通过使用固相技术的定向肽合成来产生本发明的多

肽及其片段 (Merrifield, J. Am. Chem. Soc. 85:2149-2154 (1963))。可以用手动技术或通过自动化来进行蛋白合成。例如,可以用Applied Biosystems的431A肽合成仪 (Perkin Elmer) 实现自动化合成。可选择地,不同的片度可以分别通过化学方式合成并利用化学方法进行组合以制备所需的分子。

#### [0197] IV. AARS多核苷酸

[0198] 本发明实施方案包括编码氨酰-tRNA合成酶 (AARS) 的一个或多个新鉴定的蛋白片段的多核苷酸以及其互补物、变体和片段。在某些实施方案中,AARS多核苷酸编码表1-3或表4-6或表7-9中所列的AARS多肽参照序列的全部或一部分,其代表苯丙氨酰-tRNA合成酶的剪接变体、蛋白水解片段或其他类型的片段。某些实施方案包括编码包含那些剪接变体的一个或多个剪接点的序列的多肽或蛋白的多核苷酸,以及其互补物、变体和片段。在某些实施方案中,通常由于选择的AARS剪接变体(其以新的或特别的方式组合外显子)的单一性质,AARS多核苷酸参照序列包含独特的或特别的剪接点。某些实施方案排出了相应的全长AARS多核苷酸。

[0199] 如本文所述,本发明的AARS多核苷酸还包括引物、探针、反义寡核苷酸和RNA干扰剂,所述引物、探针、反义寡核苷酸和RNA干扰剂包含这些参照多核苷酸的全部或部分,它们与这些参照多核苷酸的全部或部分互补或者与这些参照多核苷酸特异性杂交。

[0200] 本文使用的术语“多核苷酸”或“核酸”指mRNA、RNA、cRNA、cDNA或DNA。该术语通常指至少10个碱基长度的核苷酸多聚形式,所述核苷酸是核糖核苷酸或脱氧核苷酸或任一类型的核苷酸的修饰形式。该术语包括单链和双链形式的DNA。如本文所用的术语“DNA”和“多核苷酸”和“核酸”是指从特定物种的总基因组DNA分离出来的DNA分子。因此,编码多肽的DNA节段是指这样的DNA节段,其含有一种或多种编码序列,而基本上从获得所述DNA节段的物种的总基因组DNA分离出来或纯化出来。还包括不编码AARS多肽的非编码多核苷酸(例如,引物、探针、寡核苷酸)。术语“DNA节段”和“多核苷酸”包括DNA节段和这些节段的较小片段,并且还包括重组载体,包括例如质粒、粘粒、噬菌粒、噬菌体、病毒等。

[0201] 额外的编码或非编码序列可以但不一定存在于本发明的多核苷酸内,并且多核苷酸可以但不一定被连接到其他分子和/或支持材料。因此,本发明的多核苷酸,不管其编码序列的长度,可以与其他DNA序列诸如启动子、聚腺苷化信号、额外的限制性酶切位点、多克隆位点、其他编码节段等组合,以使它们总长度可以显著不同。

[0202] 因此考虑可以利用几乎任意长度的多核苷酸片段,其总长度优选受制备的简便和在预期重组DNA方案中的用途限制。包括长度约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、41、43、44、45、46、47、48、49、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、220、240、260、270、280、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000、2100、2200、2300、2400、2500、2600、2700、2800、2900、3000或更多(包括其间所有整数)个碱基的多核苷酸,包括AARS参照多核苷酸(例如,碱基数目X-Y,其中X是约1-3000或更多,并且Y是约10-3000或更多)或其互补物的任何部分或片段(例如,长度大于约6、7、8、9或10个核苷酸)。

[0203] 本发明的实施方案还包括AARS参照多核苷酸序列的“变体”。多核苷酸“变体”与参照多核苷酸相比含有一个或多个取代、添加、缺失和/或插入。一般而言,如通过本文其他地

方描述的序列比对程序,使用默认参数测定的,AARS参照多核苷酸序列的变体可以与特定的核苷酸序列具有至少约30%、40%、50%、55%、60%、65%、70%、一般至少约75%、80%、85%,希望约90%至95%或更多,以及更适合约98%或更多的序列同一性。在某些实施方案中,变体与参照序列相差约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100(包括其间所有整数)或更多碱基。在某些实施方案中,例如当多核苷酸变体编码具有非常规活性的AARS多肽时,编码的AARS多肽的希望活性相对于未修饰多肽而言基本上没有减小。对编码多肽活性的影响一般可以如本文所述来评估。

[0204] 某些实施方案包括在下文所述的严格条件下与参照AARS多核苷酸序列或它们的互补物杂交的多核苷酸。如本文使用的,术语“在低严格、中等严格、高严格或极高严格条件下的杂交”描述了杂交和洗涤的条件。进行杂交反应的指导可以见于Ausubel等人,(1998,同上),Sections 6.3.1-6.3.6。在该参考文献中描述了水性和非水性方法,可以使用任一种。

[0205] 本文参考的低严格条件包括以及涵盖用至少约1%v/v到至少约15%v/v的甲酰胺,和至少约1M到至少约2M盐在42°C杂交;以及用至少约1M到至少约2M盐在42°C洗涤。低严格条件还可以包括用1%牛血清白蛋白(BSA)、1mM EDTA、0.5M NaHPO<sub>4</sub>(pH7.2)、7%SDS在65°C杂交,以及用(i) 2×SSC、0.1%SDS;或(ii) 0.5%BSA、1mM EDTA、40mM NaHPO<sub>4</sub>(pH4.7.2)、5%SDS在室温洗涤。低严格条件的一个实施方案包括在约45°C下在6×氯化钠/柠檬酸钠(SSC)中杂交,随后至少在50°C(对于低严格条件,洗涤的温度可以提高到55°C)下,在0.2×SSC、0.1%SDS中洗涤两次。

[0206] 中等严格条件包括和涵盖用至少约16%v/v到至少约30%v/v的甲酰胺和至少约0.5M到至少约0.9M的盐在42°C杂交,以及用至少约0.1M到至少约0.2M盐在55°C洗涤。中等严格条件还可以包括用1%牛血清白蛋白(BSA)、1mM EDTA、0.5M NaHPO<sub>4</sub>(pH7.2)、7%SDS在65°C杂交,以及用(i) 2×SSC、0.1%SDS;或(ii) 0.5%BSA、1mM EDTA、40mM NaHPO<sub>4</sub>(pH4.7.2)、5%SDS在60-65°C洗涤。中等严格条件的一个实施方案包括在约45°C下在6×SSC中杂交,随后在60°C下,在0.2×SSC、0.1%SDS中洗涤一次或多次。高严格条件包括和涵盖用至少约31%v/v到至少约50%v/v的甲酰胺和至少约0.01M到至少约0.15M的盐在42°C杂交,以及用约0.01M到约0.02M的盐在55°C洗涤。

[0207] 高严格条件还可以包括用1%BSA、1mM EDTA、0.5M NaHPO<sub>4</sub>(pH7.2)、7%SDS在65°C杂交,以及用(i) 0.2×SSC、0.1%SDS;或(ii) 0.5%BSA、1mM EDTA、40mM NaHPO<sub>4</sub>(pH7.2)、41%SDS在超过65°C的温度下洗涤。高严格条件的一个实施方案包括在约45°C下在6×SSC中杂交,随后在65°C下,在0.2×SSC、0.1%SDS中洗涤一次或多次。极高严格条件的一个实施方案包括用0.5M磷酸钠、7%SDS在65°C杂交,随后在在65°C下,在0.2×SSC、0.1%SDS中洗涤一次或多次。

[0208] 本领域熟知其他严格条件,并且本领域技术人员会认识到能够操纵多种因素来优化杂交的特异性。最后洗涤的严格性优化可以确保高度的杂交。对于具体的实例,参见Ausubel等人,同上,在第2.10.1至2.10.16页以及Sambrook等人(1989,同上)的第1.101至1.104节。

[0209] 尽管严格洗涤通常在约42°C至68°C的温度下进行,但是本领域技术人员应当理

解,其他温度可以适用于严格条件。最大杂交速率通常发生在低于 $T_m$ 约20°C至25°C下以形成DNA-DNA杂交体。本领域公知, $T_m$ 是解链温度,或者是两条互补的多核苷酸序列分离的温度。用于估算 $T_m$ 的方法是本领域熟知的(参见Ausubel等人,同上在第2.10.8页)。

[0210] 通常,完美匹配的DNA双链的 $T_m$ 可以根据以下公式近似预测: $T_m = 81.5 + 16.6 (\log_{10}M) + 0.41 (\%G+C) - 0.63 (\% \text{甲酰胺}) - (600/\text{长度})$ ,其中: $M$ 是 $\text{Na}^+$ 的浓度,范围优选为0.01摩尔至0.4摩尔;%G+C是鸟嘌呤和胞嘧啶碱基之和占碱基总数的百分比,在30%至75%G+C的范围内;%甲酰胺是甲酰胺浓度的体积百分比;长度是DNA双螺旋中碱基对的数目。随机错配的碱基对的数目每增加1%,双链DNA的 $T_m$ 降低约1°C。对于高严格条件,一般在 $T_m - 15^\circ\text{C}$ 下进行洗涤,或对于中等严格条件,在 $T_m - 30^\circ\text{C}$ 下进行洗涤。

[0211] 在杂交方案的一个实例中,将包含固定的DNA的膜(例如,硝酸纤维素膜或尼龙膜)在含有标记的探针的杂交缓冲液(50%去离子甲酰胺、5×SSC、5×Denhardt's溶液(0.1%聚蔗糖、0.1%聚乙烯吡咯烷酮和0.1%牛血清白蛋白)、0.1%SDS和200mg/mL变性的鲑鱼精子DNA)中于42°C下过夜杂交。然后将膜进行两次相继的中等严格洗涤(即,用2×SSC、0.1%SDS,在45°C下进行15分钟;随后是用2×SSC、0.1%SDS,在50°C下进行15分钟),随后是两次相继的更高严格洗涤(即,用0.2×SSC、0.1%SDS,在55°C下进行12分钟;随后是用0.2×SSC和0.1%SDS溶液,在65-68°C下进行12分钟)。

[0212] 如上所述,某些实施方案涉及编码AARS多肽的AARS多核苷酸。除了其他应用,这些实施方案还可用于重组产生所需的AARS多肽或其变体,或在选定的细胞或受试者中表达AARS多肽。本领域普通技术人员将理解,由于遗传密码的简并性,有很多核苷酸序列编码本文所述的多肽。这些多核苷酸中的某些可以与任何天然基因的核苷酸序列具有低的同源性。虽然如此,本发明特别考虑到由于密码子使用中的差异而不同的多核苷酸,例如,为人和/或灵长类动物密码子选择而优化的多核苷酸。

[0213] 因此,多个多核苷酸可以编码本发明的AARS多肽。而且,多核苷酸序列可以为各种原因而被操作。实例包括但不限于加入偏好密码子以增强多核苷酸在各种生物中的表达(大体参见Nakamura等人,Nuc. Acid. Res. (2000) 28(1):292)。此外,可以加入沉默突变以引入或消除限制性位点,降低CpG二核苷酸基序的密度(参见例如Kameda等人,Biochem. Biophys. Res. Commun. (2006) 349(4):1269-1277),或减少单链序列形成茎环结构的能力:(参见例如Zuker M.,Nucl. Acid Res. (2003);31(13):3406-3415)。此外,可以通过在起始密码子处加入Kozak共有序列[即,(a/g)cc(a/g)ccATGg]来进一步优化哺乳动物表达。用于该目的的Kozak共有序列是本领域已知的(Mantyh等人PNAS92:2662-2666(1995);Mantyh等人Prot. Exp. & Purif. 6,124(1995))。

[0214] 本发明的多核苷酸,不管其编码序列自身的长度,可以与其他DNA序列组合,所述其他DNA序列例如启动子、聚腺苷化信号、其他限制性酶切位点、多克隆位点、其他编码片段等,使得它们的总长度可以显著不同。因此考虑可以利用几乎任意长度的多核苷酸片段;总长度优选地受预期重组DNA方案中制备和使用的便利性的限制。

[0215] 可以利用本领域内已知的和可获得的多种完善的技术中的任何一种来制备、操控和/或表达多核苷酸及其融合物。例如,编码本发明的多肽的多核苷酸序列或其融合蛋白或功能等同物可以用于重组DNA分子中以指导AARS多肽在适当的宿主细胞中表达。由于遗传密码子固有的简并性,可以产生编码基本上相同或在功能上等同的氨基酸序列的其他DNA



序列,并且这些序列可以用于克隆和表达给定的多肽。

[0216] 如本领域技术人员所理解地,在一些情况下,产生具有非天然存在的密码子的编码多肽的核苷酸序列是有利的。例如,可以选择特定原核或真核宿主偏好的密码子以增加蛋白表达率或以产生具有所需特性的重组RNA转录物,所述所需特性例如半衰期长于从天然存在序列产生的转录物的半衰期。此类多核苷酸常被称为“密码子优化的”。本文描述的多核苷酸的任何一个可以密码子优化形式使用。在某些实施方案中,多核苷酸可以被密码子优化而用于特定细菌,例如大肠杆菌或酵母例如酿酒酵母(参见例如Burgess-Brown等人,Protein Expr Purif.59:94-102,2008;Ermolaeva MD(2001)Curr.Iss.Mol.Biol.3(4)91-7;Welch等人,PLoS ONE4(9):e7007 doi:10.1371/journal.pone.0007002)。

[0217] 此外,可以使用本领域内公知的方法改造本发明的多核苷酸序列,从而为了多种原因改变多肽编码序列的多核苷酸,包括但不限于修饰基因产物的克隆、加工、表达和/或活性的改变。

[0218] 按照本发明的另一方面,可以例如使用基因治疗技术将编码本发明多肽的多核苷酸体内递送至受试者。基因治疗通常是指将异源核酸转移至患有该治疗所针对的疾患或病症的哺乳动物,尤其是人的某些细胞、靶细胞。以下述方式将核酸引入选择的靶细胞,使得异源DNA表达并从而产生所编码的治疗产物。

[0219] 如本文所教导的能用于基因治疗的各种病毒载体包括腺病毒、疱疹病毒、牛痘病毒、腺相关病毒(AAV),或优选地诸如逆转录病毒的RNA病毒。优选地,所述逆转录病毒载体是鼠或鸟逆转录病毒的衍生物或是慢病毒载体。优选的逆转录病毒载体是慢病毒载体。可以插入单一外源基因的逆转录病毒载体的实例包括但不限于:莫洛尼鼠白血病毒(MoMuLV)、哈维鼠肉瘤病毒(HaMuSV)、鼠乳腺肿瘤病毒(MuMTV)、SIV、BIV、HIV和劳氏肉瘤病毒(RSV)。多种其他的逆转录病毒载体可以并入多个基因。所有的这些载体能转移或并入选择性标志物基因,以便鉴定和产生转染的细胞。例如,通过将感兴趣的来源于锌指的DNA结合多肽序列以及编码对于特异性靶细胞上的受体的配体的另一基因插入病毒载体,可以使所述载体成为靶特异性的。例如,通过插入编码蛋白(二聚体)的多核苷酸,可以使逆转录病毒载体成为靶特异性的。通过使用抗体来靶向于逆转录病毒载体可以实现示例性的靶向。本领域技术人员应当知道或利用有限的实验就能容易确定具体的多核苷酸序列,所述具体的多核苷酸序列能被插入逆转录病毒基因组中,从而允许含有锌指-核苷酸结合蛋白多核苷酸的逆转录病毒载体的靶特异性递送。

[0220] 由于重组逆转录病毒是缺陷型的,所以它们需要辅助以产生感染性载体颗粒。例如,可以通过使用辅助细胞系来提供该辅助,所述辅助细胞系含有编码在LTR中的调节序列控制下的所有逆转录病毒结构基因的质粒。这些质粒缺少能使包装机制识别用于封装的RNA转录物的核苷酸序列。缺失包装信号的辅助细胞系包括但不限于,例如PSI.2、PA317和PA12。由于未包装基因组,所以这些细胞系产生空的病毒体。如果将逆转录病毒载体导入包装信号完整的这类细胞中,而用其他目的基因替代结构基因,则所述载体可以被包装并可以产生载体病毒体。通过该方法产生的载体病毒体随后可以用来感染诸如NIH3T3细胞的组织细胞系,以产生大量嵌合逆转录病毒体。

[0221] 也能使用用于基因治疗的“非病毒”递送技术,包括,例如,DNA-配体复合物、腺病毒-配体-DNA复合物、DNA的直接注射、CaPO<sub>4</sub>沉淀、基因枪技术、电穿孔、脂质体、脂转染等。

这些方法中的任何一种对于本领域技术人员而言是可广泛获得的,并适于用在本发明中。其他合适的方法对于本领域技术人员而言是可获得的,并且应当理解,可以使用可获得的任何转染方法来实现本发明。通过将分离的DNA分子封装于脂质体颗粒中并将该脂质体颗粒与靶细胞的细胞膜接触,能实现脂转染。脂质体是自我组装的胶体颗粒,其中由诸如磷脂酰丝氨酸或磷脂酰胆碱的两亲性分子组成的脂质双分子层封装了周围介质的一部分,使得所述脂质双分子层包围了亲水内部。可以这样构建单层或多层脂质体,使得内部含有所需的化学品、药物或本发明中的分离的DNA分子。

[0222] 另一方面,编码本发明多肽的多核苷酸可用于通过细胞疗法表达和递送AARS多肽。因此在另一方面,本发明包括用于治疗疾病或疾患的细胞疗法,包括施用表达或能够表达AARS多肽的宿主细胞。

[0223] 细胞疗法包括施用已经在体外被选择、繁殖和药理学上处理或改变的(即,基因修饰的)的细胞(Bordignon,C.等人,Cell Therapy:Achievements and Perspectives (1999),Haematologica,84,pp.1110-1149)。例如,此类宿主细胞包括已经被基因修饰为表达AARS多肽的原代细胞,包括巨噬细胞和干细胞。细胞疗法的目的是替代、修复或增强受损组织或器官的生物功能。

[0224] 已经考察了移植细胞用于治疗许多内分泌疾患的用途,所述内分泌疾患例如贫血和侏儒症、血液系统疾患、肾和肝衰竭、脑垂体和CNS缺陷和糖尿病(Uludag等人,Technology of Mammalian Cell Encapsulation(2000),Advanced Drug Delivery Reviews,42,pp.29-64)。移植细胞可以通过释放生物活性化合物例如本发明的AARS多肽而发挥功能,以替代在作用系统中缺乏或不足量产生的内源性AARS多肽。

[0225] 本发明的实施方案还包括无论是用于检测、扩增、反义治疗还是其他目的的寡核苷酸。为了这些和相关的目的,术语“寡核苷酸(oligonucleotide)”或“寡核苷酸(oligo)”或“寡聚体”意图包括单数形式的“寡核苷酸”以及复数形式的“寡核苷酸”,并且是指能在本发明的扩增方法中以及随后的检测方法中用作试剂的两个或更多个核苷酸、核苷、核碱基或相关化合物的任何聚合物。寡核苷酸可以是DNA和/或RNA和/或它们的类似物。

[0226] 术语寡核苷酸并不必指出该试剂的任何特定功能,而是仅一般地用于覆盖本文所述的所有这类试剂。寡核苷酸可以具有各种不同的功能,例如如果寡核苷酸能够与互补链杂交并在核酸聚合酶的存在下能被进一步延伸,则它可以起引物的作用,如果寡核苷酸包含能被RNA聚合酶识别的序列并允许转录,则它起启动子的作用,如果适当放置和/或修饰,它可以起防止杂交或妨碍引物延伸的作用。寡核苷酸可以起探针或反义剂的作用。寡核苷酸实际上可以是任何长度的,该长度仅受其具体功能所限,例如在扩增反应中、在检测扩增反应的扩增产物中或在反义或RNA干扰应用中的作用。本文所述的任何寡核苷酸可以用作引物、探针、反义寡聚体或RNA干扰剂。

[0227] 本文所用的术语“引物”是指在由诸如缓冲液和温度所限定的合适的条件下,在四种不同的核苷三磷酸和诸如DNA或RNA聚合酶或逆转录酶的聚合剂的存在下,能够作为模板指导的DNA合成的起始位点的单链寡核苷酸。在任何给定的情况下,引物的长度取决于例如引物的预期用途,并且范围通常为约15至30个核苷酸,但是可以使用更短和更长的引物。更短引物分子通常需要更低的温度以便与模板形成足够稳定的杂交复合物。引物不需要反映模板的确切序列,但必须足够互补以便与该模板杂交。引物位点是模板上与引物杂交的区

域。引物对是一组引物,包括与待扩增的序列的5'端杂交的5'上游引物和与待扩增的序列的3'端的互补物杂交的3'下游引物。

[0228] 本文所用的术语“探针”包括表面固定的或可溶的但可以被特定靶识别的固定分子。参见,例如美国专利第6,582,908号,具有10、12和更多碱基的探针的所有可能组合的阵列的实例。本文所用的探针和引物通常包含已知序列的至少10-15个连续核苷酸。为了增强特异性,还可以使用更长的探针和引物,例如包含AARS参照序列或其互补物的至少20、25、30、40、50、60、70、80、90、100或至少150个核苷酸的探针和引物。探针和引物可以显著长于这些实例,并且应当理解可以使用由本领域知识以及本说明书(包括表格、附图和序列列表)支持的任何长度。

[0229] 制备和使用探针和引物的方法描述于参考文献中,例如Sambrook, J.等人(1989) *Molecular Cloning: A Laboratory Manual*, 第2增补版, 第1-3卷, Cold Spring Harbor Press, Plainview N.Y.; Ausubel, F.M.等人(1987) *Current Protocols in Molecular Biology*, Greene Publ. Assoc. & Wiley-Intersciences, New York N.Y.; Innis, M.等人(1990) *PCR Protocols. A Guide to Methods and Applications*, Academic Press, San Diego Calif. PCR引物对可以来源于已知序列,例如通过使用用于该目的的计算程序,例如Primer(版本0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge Mass.)。

[0230] 可以使用领域内已知的软件选择用作引物或探针的寡核苷酸。例如, OLIGO4.06软件用于选择每条高达100个核苷酸的PCR引物对,和用于分析寡核苷酸和来自高达32kb的输入多核苷酸序列的高达5,000个核苷酸的更大的多核苷酸。为了扩展能力,将其他特征并入了类似的引物选择程序。例如, PrimOU引物选择程序(可从Genome Center at University of Texas South West Medical Center, Dallas Tex. 公开获得)能够从兆碱基选择特定引物,并因此用于设计全基因组范围的引物。

[0231] Primer3引物选择程序(可从Whitehead Institute/MIT Center for Genome Research, Cambridge Mass. 公开获得)允许使用者输入“引物错配库”,其中避免作为引物结合位点的序列由使用者指定。通过上述任何选择方法鉴定的寡核苷酸和多核苷酸片段可用于杂交技术中,例如,用作PCR或测序引物、微阵列元件或特异性探针以鉴定核酸样品中完全或部分互补的多核苷酸。寡核苷酸选择的方法不限于本文所述的那些方法。Primer3特别用于选择用于微阵列的寡核苷酸(后两个引物选择程序的源代码还可获自其各自来源,并且被修改以满足用户的特定需求)。PrimeGen程序(可从UK Human Genome Mapping Project Resource Centre, Cambridge UK 公开获得)设计了基于多个序列比对的引物,从而允许选择与比对核酸序列的最保守或最不保守区域杂交的引物。因此,该程序用于鉴定独特和保守的寡核苷酸和多核苷酸片段。通过上述选择方法的任何一个鉴定的寡核苷酸和多核苷酸片段用于杂交技术,例如,作为PCR或测序引物、微阵列元件或特异性探针,以鉴定核酸样品中的完全或部分互补的多核苷酸。寡核苷酸选择的方法不限于本文描述的那些。

[0232] 在某些实施方案中,可以通过逐步固相合成,利用上述参考文献中详述的方法以及下文关于具有混合的或不带电的和阳离子的主链键合的寡核苷酸的合成的方法制备寡核苷酸。在某些情况下,可取的是,将其他化学部分添加到寡核苷酸上,例如,以增强药代动力学或促进化合物的俘获或检测。这些部分通常可以根据标准合成方法共价连接到寡聚体

的末端。例如，聚乙二醇部分或其他亲水性聚合物（例如具有10-100个单体亚单元的一种）的添加，可以用于提高溶解度。一种或多种带电基团，例如带负电荷的基团，例如有机酸，可以增强细胞摄取。

[0233] 许多可检测分子可以用于使寡核苷酸或蛋白可被检测，例如可以直接（例如通过光发射）或间接（例如通过与荧光标记的抗体结合）检测的放射性同位素、荧光色素、染料、酶、纳米粒子、化学发光标志物、生物素或本领域内已知的其他单体。

[0234] 放射性同位素提供了可用于本发明某些方面的可检测分子的实例。几种放射性同位素可用作标记核苷酸或蛋白的可检测分子，包括例如<sup>32</sup>P、<sup>33</sup>P、<sup>35</sup>S、<sup>3</sup>H和<sup>125</sup>I。这些放射性同位素具有不同的半衰期、衰变类型和能量水平，其可被调整以满足特定方案的需要。例如，<sup>3</sup>H是低能量的发射体，导致低的背景水平，然而，该低能量也导致放射自显影的长时段。放射性标记的核糖核苷酸、脱氧核糖核苷酸和氨基酸是可商业途径获得的。可以获得在第一或 $\alpha$ 磷酸基团或者在第三或 $\gamma$ 磷酸基团放射性标记的核苷酸。例如， $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ 和 $[\gamma\text{-}^{32}\text{P}]\text{dATP}$ 是可商业途径获得的。此外，放射性标记的核苷酸的不同特异性活性也是可商业途径获得的，并且可以根据不同方案而被调整。

[0235] 可用于检测寡核苷酸的可检测分子的其他实例包括荧光团。几种荧光团可用于标记核苷酸，包括例如荧光素、四甲基若丹明、德克萨斯红和许多其他荧光团（例如Haugland, Handbook of Fluorescent Probes-第9版, 2002, Molec. Probes, Inc., Eugene OR; Haugland, The Handbook: A Guide to Fluorescent Probes and Labeling Technologies-第10版, 2005, Invitrogen, Carlsbad, CA）。

[0236] 作为一个实例，可以在化学合成期间荧光标记寡核苷酸，因为在核苷酸合成期间胺或硫醇的加入允许添加荧光团。荧光标记的核苷酸是可商业途径获得的。例如，与涵盖谱图的十种不同荧光团匹配的尿苷和脱氧尿苷三磷酸是可用的。可以与核苷酸直接结合的荧光染料也可用作可检测分子。例如，FAM、JOE、TAMRA和ROX是已经与核苷酸连接的胺反应性荧光染料，并且用于自动化DNA测序。这些荧光标记的核苷酸是可商业途径获得的，例如ROX-ddATP、ROX-ddCTP、ROX-ddGTP和ROX-ddUTP。

[0237] 非放射性和非荧光的可检测分子也是可用的。如上所述，生物素可以与核苷酸直接连接，并且通过与亲和素或链霉亲和素的特异性和高亲和力结合而被检测，所述亲和素或链霉亲和素已经化学偶联至催化比色反应的酶（例如，磷酸酶、荧光素酶或过氧化物酶）。地高辛配基标记的核苷酸也可类似用于核酸的非同位素检测。生物素化和地高辛配基标记的核苷酸是可商业途径获得的。

[0238] 极小颗粒，称为纳米粒子，也可用于标记寡核苷酸探针。这些粒子大小范围1-1000nm，并且包括多种化学结构，例如金和银粒子和量子。当用倾斜入射的白光照射时，范围40-120nm的银或金纳米粒子将以高密度散射单色光。散射光的波长取决于颗粒的大小。紧密靠近的4至5个不同颗粒会相互散射单色光，其重叠时将产生特定的独特颜色。这些颗粒由诸如Genicon Sciences (Carlsbad, CA) 等公司生产。衍生的银或金颗粒可以与多种分子连接，所述分子包括蛋白、抗体、小分子、受体配体和核酸。例如，颗粒表面可以被化学衍生化以允许连接核苷酸。

[0239] 可用于检测可检测分子的纳米粒子的其他类型包括量子。量子是直径为1-5nm的荧光晶体，其可由大范围波长的光激发。当由具有适当波长的光激发时，这些晶体发射光，

例如单色光,波长取决于其化学组成和大小。诸如CdSe、ZnSe、InP或InAs的量子具有独特的光学性质:这些和类似量子可获自许多商业来源(例如,NN-Labs,Fayetteville,AR;Ocean Nanotech,Fayetteville,AR;Nanoco Technologies,Manchester,UK;Sigma-Aldrich, St.Louis,MO)。

[0240] 根据量子晶体的尺寸类别数目,可以产生数十个类别的颗粒。晶体的尺寸类别通过以下产生:1)严格控制晶体形成参数以产生每个所需的尺寸类别的颗粒,或2)通过在宽松控制的晶体形成参数下产生数批晶体,随后根据所需尺寸和/或发射波长来分类。其中量子包埋于半导体发光/检测装置的固有硅外延层内的参考文献的两个实例是Chapple Sokol,等人的美国专利第5,293,050和5,354,707号。

[0241] 在某些实施方案中,寡核苷酸引物或探针可以用一种或多种发光或以其他方式可检测的染料标记。染料发射的光可以是可见光或不可见光,例如紫外光或红外光。在示例性实施方案中,染料可以是荧光共振能量转移(FRET)染料;咕吨染料,例如荧光素和若丹明;在 $\alpha$ 或 $\beta$ 位置具有氨基的染料(例如萘胺染料、1-二甲基氨基萘基-5-磺酸酯、1-苯胺基-8-萘磺酸酯和2-对甲苯胺基-6-萘磺酸酯);具有3-苯基-7-异氰酸香豆素的染料;吡啶,例如9-异硫氰酸吡啶和吡啶橙;苝,苯并恶二唑和苝;具有3-( $\epsilon$ -羧基戊基)-3'-乙基-5,5'-二甲基氧杂羧花青素(CYA)的染料;6-羧基荧光素(FAM);5&6-羧基若丹明-110(R110);6-羧基若丹明-6G(R6G);N,N,N',N'-四甲基-6-羧基若丹明(TAMRA);6-羧基-X-若丹明(ROX);6-羧基-4',5'-二氯-2',7'-二甲氧基荧光素(JOE);ALEXA FLUOR™;Cy2;德克萨斯红和若丹明红;6-羧基-2',4,7,7'-四氯荧光素(TET);6-羧基-2',4,4',5',7,7'-六氯荧光素(HEX);5-羧基-2',4',5',7'-四氯荧光素(ZOE);NAN;NED;Cy3;Cy3.5;Cy5;Cy5.5;Cy7;和Cy7.5;IR800CW、ICG、Alexa Fluor 350;Alexa Fluor 488;Alexa Fluor 532;Alexa Fluor 546;Alexa Fluor 568;Alexa Fluor 594;Alexa Fluor 647;Alexa Fluor 680或Alexa Fluor 750。

[0242] 本发明的AARS多核苷酸和寡核苷酸可用于本文描述的治疗、诊断、研究或药物发现组合物和方法的任何一个。

[0243] V. 抗体

[0244] 根据另一方面,本发明还提供了对于AARS多肽、或其天然细胞结合配体(即,细胞受体、脂质、糖类、蛋白或核酸结合配体)或其复合物表现出结合特异性的抗体以及使用所述抗体的方法。如本文描述和本领域已知的,术语抗体包括抗体的各种变型,例如FAB、人抗体、修饰的人抗体、单链、非人抗体和免疫球蛋白折叠的其他衍生物,其构成针对抗原的免疫系统配体的基础。抗体可以用于本文提供的治疗、诊断、药物发现或蛋白表达/纯化方法以及组合物。

[0245] 本发明的某些抗体不同于某些之前制备的抗体,因为它们可以区分表1-3或表4-6或表7-9的AARS蛋白片段与其相应的全长AARS,通常通过以比相应全长AARS更大的亲和力结合AARS蛋白片段。一般而言,这种抗体可以结合通过剪接变化、蛋白水解或产生本发明AARS蛋白片段的其他细胞加工(例如,翻译后加工,包括但不限于磷酸化和改变蛋白结构的其他修饰)而产生或揭示的独特序列或结构。在一些方面,抗体可以结合独特剪接点附近的序列(例如,结合选自表2B、5B或8B所列的剪接点序列的至少5个连续氨基酸的一个或多个区域,或者结合表2B、5B或8B所列的该剪接位点的C端任何氨基酸序列)。例如,这种抗体可以对以下具有结合特异性:在AARS蛋白片段中暴露而不在全长AARS中暴露的一个或多个非

溶剂暴露面,或者在全长AARS不存在或者以其他方式不可及的序列。抗体还可以结合由于AARS蛋白片段与全长AARS之间的折叠差异导致的独特三维结构。这种结合差异可以是局部的(例如,在特定结构域或区域)或总体的。作为一个实例,AARS蛋白片段的折叠可以产生在相应或亲本AARS中不存在的独特的连续或不连续表位。实例还包括特异性结合通过剪接变化、蛋白水解或其他细胞加工而产生的N端或C端的抗体;此类末端与全长AARS相比可以是独特的,或者可以不暴露于全长形式的抗体结合,因为它们的末端被完全或部分包埋在较大的AARS亲本分子的总体结构中。

[0246] 在某些实施方案中,本文提供的抗体不形成聚集物,具有所需的溶解度,和/或具有适合用于人类的免疫原性特征,如本文描述和本领域已知的。还包括适合生产工作,例如纯化本文描述的AARS蛋白片段的抗体。优选地,活性抗体可以被浓缩为至少约10mg/ml,并且被任选配制用于生物治疗用途。

[0247] 在某些实施方案中,抗体有效调节本发明的AARS多肽所介导的一种或多种非常规活性。例如,在一些实施方案中,抗体是与AARS多肽和/或其结合配体结合、抑制它们相互作用的能力和/或拮抗AARS多肽的非常规活性的抗体。例如,在某些实施方案中,抗体结合AARS多肽的细胞结合配体并模拟AARS多肽活性,例如通过增加或激动由AARS多肽介导的非常规活性。因此,例如通过部分或完全拮抗或激动本发明AARS多肽的活性,抗体可以用来治疗、治疗或预防所述AARS多肽介导的疾病、疾患或其他病症。

[0248] 如果抗体或其抗原结合片段与多肽以可检测的水平(例如,在ELISA测定中)发生反应,并且在相似条件下不与无关多肽以统计学显著方式发生可检测的反应,则称所述抗体或其抗原结合片段与本发明多肽“特异性结合”、“免疫结合”和/或是“免疫反应性的”。在某些情况下,结合剂不与AARS多肽的全长形式显著地相互作用。

[0249] 在此背景下使用的免疫结合通常是指免疫球蛋白分子和该免疫球蛋白特异性针对的抗原之间发生的非共价型相互作用。可以用相互作用的解离常数( $K_d$ )来表示结合(诸如免疫结合相互作用)的强度或亲和力,其中 $K_d$ 越小表示亲和力越高。可以用本领域内公知的方法对所选多肽的免疫结合特性进行定量。参见例如Davies等人(1990) Annual Rev. Biochem. 59:439-473。在某些示例性实施方案中,抗体具有至少约0.01、0.05、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40或50nM的对AARS蛋白片段的亲和力。在某些实施方案中,抗体对AARS蛋白片段的亲和力强于其对相应的全长AARS多肽的亲和力,通常强约1.5x、2x、2.5x、3x、3.5x、4x、4.5x、5x、6x、7x、8x、9x、10x、15x、20x、25x、30x、40x、50x、60x、70x、80x、90x、100x、200x、300x、400x、500x、600x、700x、800x、900x、1000x或更多(包括其间所有整数)。在某些实施方案中,抗体对相应全长AARS蛋白的亲和力是至少约0.05、0.1、0.25、0.5、0.75、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20 $\mu$ M。在某些实施方案中,抗体弱或基本上不可检测地与全长AARS蛋白结合。

[0250] 抗体的“抗原结合位点”或“结合部分”是指免疫球蛋白分子参与抗原结合的部分。抗原结合位点由重(“H”)和轻(“L”)链N端可变区(“V”)的氨基酸残基形成。重链和轻链的V区中的三个高度不同的区段被称为“高变区”,其插在称为“框架区”或“FR”的较保守的侧翼区段之间。因此,术语“FR”是指天然见于免疫球蛋白中高变区之间并邻近高变区的氨基酸序列。在抗体分子中,轻链的三个高变区和重链的三个高变区在三维空间彼此相对排列,形

成抗原结合表面。该抗原结合表面与结合的抗原的三维表面互补,并且每条重和轻链的三个高变区被称作“互补决定区”或“CDR”。

[0251] 可以通过本领域普通技术人员已知的多种技术中的任何一种制备抗体。参见,例如Harlow和Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988。对感兴趣的多肽特异的单克隆抗体可以使用例如Kohler和Milstein, *Eur. J. Immunol.* 6:511-519, 1976的技术以及对其改进的技术来制备。还包括利用转基因动物例如小鼠来表达人抗体的方法。参见例如,Neuberger等人, *Nature Biotechnology* 14:826, 1996; Lonberg等人, *Handbook of Experimental Pharmacology* 113:49-101, 1994; 和Lonberg等人, *Internal Review of Immunology* 13:65-93, 1995。特定实例包括 **REGERNEREX®** 的 **VELOCIMMUNE®** 平台(参见例如美国专利号6,596,541)。还可以通过使用噬菌体展示或酵母展示文库来产生或鉴定抗体(参见例如美国专利号7,244,592; Chao等人, *Nature Protocols* 1:755-768, 2006)。可用文库的非限制性实例包括克隆的或合成的文库,例如人组合抗体文库(HuCAL),其中人抗体谱的结构多样性由七个重链和七个轻链可变区基因代表。这些基因的组合在主库中产生49个框架。通过在框架上叠加高变基因盒(CDR=互补决定区),可以再生成大量的人类抗体谱。还包括设计为具有编码轻链可变区、重链CDR-3的人类供体来源的片段、编码重链CDR-1多样性的合成DNA、和编码重链CDR-2多样性的合成DNA的人类文库。适合使用的其他文库对于本领域技术人员而言是明显的。本发明多肽可用于纯化过程,例如亲和色谱步骤。

[0252] 通过优先蛋白水解切割IgM可以产生“Fv”片段,而蛋白水解切割IgG或IgA免疫球蛋白分子很少产生“Fv”片段。然而,更通常使用本领域内已知的重组技术衍生出Fv片段。Fv片段包括非共价V<sub>H</sub>:V<sub>L</sub>异二聚体,该异二聚体包括保持了天然抗体分子的大部分抗原识别和结合能力的抗原结合位点。参见例如Inbar等人(1972) *Proc. Nat. Acad. Sci. USA* 69:2659-2662; Hochman等人(1976) *Biochem* 15:2706-2710; 和Ehrlich等人(1980) *Biochem* 19:4091-4096。

[0253] 单链Fv(“sFv”)多肽是从融合基因表达的共价连接的V<sub>H</sub>:V<sub>L</sub>异二聚体,所述融合基因包括通过肽编码连接物连接的V<sub>H</sub>-和V<sub>L</sub>-编码基因。Huston等人(1988) *PNAS USA* 85(16):5879-5883。已有多种方法描述了分辨化学结构,用于将来自抗体V区的天然聚集、但化学分离的轻多肽链和重多肽链转换成sFv分子,所述sFv分子会折叠成基本上相似于抗原结合位点的结构的三维结构。参见,例如,Huston等人的美国专利第5,091,513号和第5,132,405号; 和Ladner等人的美国专利第4,946,778号。

[0254] 每个上述分子包括分别插在重链和轻链FR组之间的重链和轻链CDR组,所述FR组为CDR提供支持并且限定了CDR相对于彼此的空间关系。如本文所用的术语“CDR组”指重链或轻链V区的三个高变区。从重链或轻链的N端开始,将这些区分别表示为“CDR1”、“CDR2”和“CDR3”。因此,抗原结合位点包括6个CDR,包含来自每条重链和轻链V区的CDR组。在本文中,包含单一CDR(例如,CDR1、CDR2或CDR3)的多肽被称为“分子识别单元”。多个抗原抗体复合物的晶体分析已经证实,CDR的氨基酸残基与结合的抗原形成广泛接触,其中最广泛的抗原接触是与重链CDR3。因此,所述分子识别单元主要负责抗原结合位点的特异性。

[0255] 如本文所用的术语“FR组”指为重链或轻链V区的CDR组的CDR提供框架的四个侧翼氨基酸序列。一些FR残基可以接触结合的抗原;然而,FR主要负责将V区折叠为抗原结合位

点,尤其是直接邻近CDR的FR残基。在FR中,某些氨基酸残基和某些结构特征是高度保守的。在这点上,所有的V区序列包含约90个氨基酸残基的内部二硫环。当V区折叠成结合位点时,CDR表现为突出的环状基序,该环状基序形成抗原结合表面。普遍公认的是,FR的保守结构区影响CDR环成某些“常规”结构的折叠形状,与具体的CDR氨基酸序列无关。而且,已知某些FR残基参与非共价的结构域间接触,该结构域间接触使抗体重链和轻链的相互作用稳定。

[0256] 某些实施方案包括单结构域抗体(sdAb或“纳米抗体”),其指由单个单体可变抗体结构域组成的抗体片段(参见例如美国专利号5,840,526;5,874,541;6,005,079,6,765,087,5,800,988;5,874,541;和6,015,695)。这样的sdAB通常具有约12-15kDa的分子量。在某些方面,sdAB和其他抗体分子可以源自或分离自免疫的骆驼和美洲驼的独特的重链抗体,常称为骆驼抗体。参见例如Conrath等人,JBC.276:7346-7350,2001。

[0257] 很多包含源自非人免疫球蛋白的抗原结合位点的“人源化”抗体分子已有描述,包括如下嵌合抗体:具有与人恒定结构域融合的啮齿动物V区及其相关CDR(Winter等人(1991)Nature 349:293-299;Lobuglio等人(1989)Proc.Nat.Acad.Sci.USA 86:4220-4224;Shaw等人(1987)J Immunol.138:4534-4538;和Brown等人(1987)Cancer Res.47:3577-3583),在与合适的人抗体恒定结构域融合之前移植入人支持FR中的啮齿动物CDR(Riechmann等人(1988)Nature 332:323-327;Verhoeyen等人(1988)Science 239:1534-1536;和Jones等人(1986)Nature 321:522-525),以及由重组相嵌的啮齿动物FR支持的啮齿动物CDR(欧洲专利公开第519,596号,于1992年12月23日公布)。这些“人源化”分子被设计为对于啮齿动物抗人抗体分子的不希望的免疫应答降至最低,该免疫应答限制了那些部分在人接受体中的治疗应用的持续时间和功效。参见例如美国专利号5,530,101;5,585,089;5,693,762;6,180,370;和7,022,500。

[0258] 本发明抗体可用于本文描述的治疗、诊断、药物发现、蛋白纯化和分析方法以及组合物的任何一种。

[0259] VI. 抗体替代物和其他结合剂

[0260] 根据另一方面,本发明还提供了抗体替代物或其他结合剂以及使用它们的组合物和方法,所述抗体替代物或其他结合剂例如可溶性受体、adnectin、肽、肽模拟物、小分子、适体等,其表现出对本文公开的AARS多肽或其细胞结合配体或者其部分、变体或衍生物的结合特异性。结合剂可用于本文描述的治疗、诊断、药物发现或蛋白表达/纯化和分析方法以及组合物的任何一种。生物基础的结合剂例如adnectin、可溶性受体、avimer和trinectin是特别有用的。

[0261] 在某些实施方案中,这些结合剂能有效地调节本发明的AARS多肽所介导的一种或多种非常规活性。例如,在一些实施方案中,所述结合剂是与本发明的AARS多肽和/或其结合配体结合、抑制它们相互作用的能力、和/或拮抗所述AARS多肽非常规活性的结合剂。例如,在某些实施方案中,结合剂结合AARS多肽的细胞结合配体并模拟AARS多肽活性,例如通过增加或激动由AARS多肽介导的非常规活性。因此,例如通过部分或完全拮抗或激动本发明AARS多肽的活性,这些结合剂可以用来治疗、治疗或预防所述AARS多肽介导的疾病、疾患或其他病症。

[0262] 如果结合剂与多肽或其细胞结合配体以可检测的水平(例如,在ELISA测定中)发生反应,并且在相似条件下不与无关多肽以统计学显著方式发生可检测的反应,则称所述



结合剂与本发明AARS多肽“特异性结合”。在某些情况下,结合剂不与AARS多肽的全长形式显著地相互作用。在某些示例性实施方案中,结合剂具有至少约0.01、0.05、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40或50nM的对AARS蛋白片段或其细胞结合配体的亲和力。在某些实施方案中,结合剂对AARS蛋白片段的亲和力强于其对相应的全长AARS多肽的亲和力,通常强约1.5x、2x、2.5x、3x、3.5x、4x、4.5x、5x、6x、7x、8x、9x、10x、15x、20x、25x、30x、40x、50x、60x、70x、80x、90x、100x、200x、300x、400x、500x、600x、700x、800x、900x、1000x或更多(包括其间所有整数)。在某些实施方案中,结合剂对相应全长AARS蛋白的亲和力是至少约0.05、0.1、0.25、0.5、0.75、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20 $\mu$ M。

[0263] 如上文所述,结合剂包括“肽”。术语肽通常是指氨基酸残基的聚合物及其变体和合成类似物。在某些实施方案中,术语“肽”是指相对短的多肽,包括由约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45或50个(包括其间所有整数和范围(例如,5-10、8-12、10-15))氨基酸组成并且与AARS多肽、其细胞结合配体或二者相互作用的肽。如本文所述,肽可以由天然存在的氨基酸或非天然存在的氨基酸组成。

[0264] 还提供了仅由天然存在的氨基酸组成的肽、肽模拟物或肽类似物。肽类似物普遍用于药物行业,作为具有模板肽类似的性质的非肽药物。这些类型的非肽化合物被称为“肽的模拟物”或“肽模拟物”(Luthman,等人,A Textbook of Drug Design and Development, 14:386-406,第2版,Harwood Academic Publishers(1996);Joachim Grante, Angew.Chem.Int.Ed.Engl., 33:1699-1720(1994);Fauchere,J.,Adv Drug Res., 15:29(1986);Veber和Freidinger TINS,p.392(1985);和Evans,等人,J.Med.Chem.30:229(1987))。肽模拟物是模拟肽的生物活性但是在化学性质上不再是肽的分子。肽模拟物化合物是本领域已知的并且描述于例如美国专利第6,245,886号。

[0265] 本发明还包括类肽。肽的类肽衍生物代表另一种形式的修饰肽,其保持对生物活性重要的结构决定子,但是消除了肽键,从而赋予对蛋白水解的抗性(Simon,等人,PNAS USA.89:9367-9371,1992)。类肽是N取代的甘氨酸的寡聚体。已经描述了许多N-烷基,各自对应于天然氨基酸的侧链。本发明的肽模拟物包括其中至少一个氨基酸、几个氨基酸或所有氨基酸残基被相应的N取代甘氨酸所替代的化合物。类肽文库描述于例如美国专利第5,811,387号。

[0266] 结合剂可以包括一种或多种小分子。“小分子”指合成来源或生物来源(生物分子)的有机化合物,但通常不是聚合物。有机化合物指其分子包含碳的一大类化合物,通常排除仅含有碳酸盐、碳的单纯氧化物或氰化物的那些化合物。“生物分子”通常是指由活的生物产生的有机分子,包括大的聚合分子(生物聚合物),例如肽、多糖和核酸以及小分子,例如初级代谢产物、次级代谢产物、脂质、磷脂、糖脂、甾醇、甘油酯、维生素和激素。“聚合物”通常指由重复结构单元组成的大分子或高分子,所述重复结构单元通常由共价化学键连接。

[0267] 在某些实施方案中,小分子具有小于1000-2000道尔顿的分子量,通常为约300至700道尔顿,并包括约50、100、150、200、250、300、350、400、450、500、550、500、650、600、750、700、850、800、950或1000道尔顿。小分子文库在本文其他地方描述。

[0268] 结合剂还包括适体(参见例如,Ellington等人,Nature.346,818-22,1990;和

Tuerk等人, *Science*. 249, 505-10, 1990)。适体的实例包括核酸适体(例如, DNA适体、RNA适体)和肽适体。核酸适体通常指已经通过例如SELEX(通过指数富集的配体的系统性进化)的体外选择或等效方法的重复轮次被工程化改造,从而与各种分子靶结合的核酸种类,所述分子靶例如小分子、蛋白、核酸并且甚至是细胞、组织和生物。参见例如美国专利号6,376,190;和6,387,620。因此,包括与本文所述的AARS多肽和/或它们的细胞结合配体结合的核酸适体。

[0269] 肽适体通常包括两端连接到蛋白支架(即,通常将肽适体的结合亲和力增强到与抗体的结合亲和力相当的水平(例如,在纳摩尔范围内)的双结构约束)的可变的肽环。在某些实施方案中,可变环长度可以由约10-20个(包括其间所有整数)氨基酸组成,并且该支架可以包括具有良好溶解度和容纳性质的任何蛋白。某些示例性的实施方案可以使用细菌蛋白硫氧还蛋白A作为支架蛋白,待插入还原活性位点的可变环(野生型蛋白中的-Cys-Gly-Pro-Cys-环)具有能够形成二硫键的两个半胱氨酸侧链。例如,美国申请号2003/0108532中描述了用于鉴定肽适体的方法。因此,包括与本文所述的AARS多肽和/或它们的细胞结合配体结合的肽适体。可以利用本领域内已知的不同系统,包括酵母双杂交系统,进行肽适体选择。

[0270] 还包括与本发明的AARS蛋白片段特异性结合的ADNECTINS™、AVIMERS™、anaphone和anticalin。ADNECTINS™指一类衍生自人类纤连蛋白的靶向生物制剂,纤连蛋白是与其他蛋白天然结合的丰富的细胞外蛋白。参见例如美国申请号2007/0082365;2008/0139791;和2008/0220049。ADNECTINS™通常由天然纤连蛋白骨架以及人类纤连蛋白特定部分的多个靶向结构域组成。靶向结构域可以被工程化以使ADNECTIN™能够特异性识别感兴趣的治疗靶,例如本发明的AARS蛋白片段。

[0271] AVIMERS™指使用体外的外显子改组和噬菌体展示工程化的多聚体结合蛋白或肽。多个结合结构域连接,导致与单表位免疫球蛋白结构域相比更大的亲和力和特异性。参见例如Silverman等人, *Nature Biotechnology*. 23:1556-1561, 2005;美国专利号7,166,697;和美国申请号2004/0175756、2005/0048512、2005/0053973、2005/0089932和2005/0221384。

[0272] 还包括设计的锚蛋白重复蛋白(DARPin),其包括能够在药物发现和药物开发中提供超越抗体的靶结合优点的一类非免疫球蛋白。除了其他用途,DARPin完美适合毒素或其他治疗性负载的体内成像或递送,这是因为它们有利的分子性质,包括小尺寸和高稳定性。许多靶特性DARPin在细菌中的低成本生产和快速产生使得DARPin方法可用于药物发现。此外,DARPin可以容易地以多特异性形式产生,提供了靶向效应DARPin至特定器官或用几个DARPin组成的一个分子靶向多个受体的潜力。参见例如Stumpff等人, *Curr Opin Drug Discov Devel*. 10:153-159, 2007;美国申请号2009/0082274;和PCT/EP2001/10454。

[0273] 某些实施方案包括“单体”,其通常利用人类纤连蛋白的第10个纤连蛋白III型结构域(FNfn10)作为支架以展示靶向结合的多个表面环。FNfn10是小的(94个残基)蛋白,具有类似于免疫球蛋白折叠的 $\beta$ -夹心结构。它在没有二硫键或金属离子时是高度稳定的,并且它可以在细菌中以正确折叠形式高水平表达。FNfn10支架与实际上任何展示技术相容。参见例如Batori等人, *Protein Eng*. 15:1015-20, 2002;和Wojcik等人, *Nat Struct Mol Biol*., 2010;和美国专利号6,673,901。

[0274] Anticalin指一类抗体模拟物,通常从人类脂笼蛋白合成,具有由结构上刚性的框架支持的高变环区的结合蛋白家族。参见例如美国申请号2006/0058510。Anticalin通常具有约20kDa的大小。Anticalin可以由八个反平行 $\beta$ -链和一个连接的 $\alpha$ -螺旋形成的筒状结构(稳定的 $\beta$ -筒状支架)表征,所述八个反平行 $\beta$ -链由四个肽环配对连接。在某些方面中,在高变环区域形成实现特异性结合的构象偏差。参见例如Skerra, FEBS J. 275:2677-83, 2008, 通过引用并入本文。

[0275] VII. 有关药物释放测定以及产品规格、诊断和试剂的生物测定和分析测定

[0276] 还包括涉及作为治疗和诊断试剂的AARS蛋白片段和相关剂的生物测定。实例包括测量纯度、生物活性、亲和力、溶解度、pH、内毒素水平等的生物测定和分析测定,其许多在本文描述。还包括建立了剂量响应曲线和/或提供两批剂之间比较的一个或多个基础的测定。批比较可以基于化学表征、生物表征和临床表征的任何一个或多个。对于蛋白剂,还包括评价选定剂的效价、稳定性、药代动力学和免疫原性的方法。除了其他用途以外,这些和其他方法可用于生物或化学剂的大量释放测试,所述生物或化学剂包括AARS蛋白片段、抗体、结合剂、多核苷酸,例如反义剂和载体,以及本文描述的其他剂。

[0277] 某些实施方案包括使用生物亲和测定。此类测定可用于评估例如AARS蛋白片段与细胞结合配体之间或者AARS蛋白片段与抗体之间的结合亲和力。还可以测量AARS蛋白片段和替代结合剂之间或者AARS细胞结合配体与候选物或前导测试化合物之间的结合亲和力,所述替代结合剂例如候选物或前导测试化合物(例如,AARS的小分子调节剂)。某些示例性结合亲和测定可以利用本文所述和本领域已知的ELISA测定。某些测定利用高效受体结合色谱(参见例如Roswall等人, Biologicals. 24:25-39, 1996)。其他示例性结合亲和测定利用基于表面等离子体共振 (SPR) 的技术。实例包括BIACore技术,其中某些集成了SPR技术与微流体系统以实时监测范围从pM至mM的浓度的分子相互作用。还包括KINEXA™测定,其提供了结合特异性、结合亲和力和结合动力学/速率常数的准确测量。

[0278] 某些实施方案涉及用于评价或最佳化蛋白剂的免疫原性的免疫测定。实例包括离体细胞测定和体外免疫酶测定,以提供对治疗蛋白的免疫原性潜力的有用信息。例如,可以使用离体细胞应答测定来再现抗原呈递细胞 (APC) 和T细胞之间的细胞合作,从而测量与感兴趣的蛋白接触之后的T细胞激活。某些体外酶测定可以利用涵盖大部分相关人群的重组HLA-DR分子集合,并且可以包括使用HLA-DR分子测试肽(源自治疗蛋白的片段化)结合的自动化免疫酶测定。还包括降低选定蛋白的免疫原性的方法,例如通过使用这些和相关方法来鉴定并然后去除或改变蛋白剂的一个或多个T细胞表位。

[0279] 还包括用于测量诸如特定生物活性的参数的生物释放测定(例如,基于细胞的测定),所述生物活性包括非常规生物活性和细胞毒性。某些特定生物测定包括,例如,利用与读数器功能上偶联的选定的AARS蛋白片段的细胞结合配体(例如细胞表面受体)的基于细胞的测定,所述读数器例如本文描述的非常规生物活性的荧光或发光指示器。例如,具体实施方案包括,包含与AARS蛋白片段结合的细胞表面受体或其细胞外部分的细胞,其中所述细胞包含检测器或读数器。还包括表征诸如AARS多肽或抗体的剂的药代动力学的体内生物测定,通常利用工程化小鼠或其他哺乳动物(参见例如Lee等人, The Journal of Pharmacology. 281:1431-1439, 1997)。基于细胞毒性的生物测定的实例包括但不限于释放测定(例如,测量凋亡的铬或铯释放测定;参见例如von Zons等人, Clin Diagn Lab

Immunol. 4:202-207, 1997), 其可以评估AARS蛋白片段的细胞毒性, 无论是用于建立剂量响应曲线、批测试或是与各种管理机构例如美国食品药品监督管理局 (FDA) 审批相关的其他性质。

[0280] 例如, 可以使用此类测定来开发选定的AARS蛋白片段或其他剂的剂量响应曲线, 和/或比较蛋白或其他剂的不同批次的剂量响应曲线。剂量响应曲线是使刺激强度与受体响应关联的X-Y曲线; 响应可以是生理或生化响应, 例如体外细胞或体内细胞或组织中的非常规生物活性, 体内测量的治疗有效量 (例如通过EC<sub>50</sub>所测量的), 或者无论是在体外还是在体内测量的死亡 (例如, 细胞死亡、生物死亡)。死亡通常被指示为LD<sub>50</sub>, 一种对于模型群体的50%致死的统计学派生剂量, 但是它也可以通过LC<sub>01</sub> (1%动物测试群体的致死剂量)、LC<sub>100</sub> (100%动物测试群体的致死剂量) 或LC<sub>L0</sub> (引起致死的最低剂量) 来指示。几乎任何期望的效应或终点可以通过该方式来表征。

[0281] 响应曲线的测量剂量通常绘制在X轴, 并且响应绘制在Y轴。更通常地, 剂量的对数绘制在X轴, 最常产生S形曲线, 最陡峭部分在中间。没有可观察的效应水平 (NOEL) 指没有观察到可测量的效应的最低实验剂量, 并且阈值剂量指曲线上指示高于零的响应的第一点。作为一般原则, 越强的药物产生越陡的剂量响应曲线。对于许多药物而言, 实现预期效应的剂量略大于阈值剂量, 通常是因为更低的剂量相对无效, 而更高的剂量导致不期望的副作用。对于体内产生的剂量响应曲线而言, 如果需要, 可以通过诸如 $\mu\text{g}/\text{kg}$ 、 $\text{mg}/\text{kg}$ 或 $\text{g}/\text{kg}$ 体重的值来表征曲线。

[0282] 对于批比较而言, 计算不同批次的不同剂量响应曲线之间 (例如, AARS蛋白片段、抗体或其他剂的不同批次之间) 的变异系数 (CV) 可能是有用的, 部分原因是CV允许采用不同单位或不同方式的数据集之间的比较。例如, 在某些示例性实施方案中, 对于4、5、6、7或8点剂量曲线而言, 2或3或更多不同批次AARS蛋白片段或其他剂之间具有小于约15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%或1%的CV。在某些实施方案中, 在基于细胞的测定中测量剂量响应曲线, 并且其读数涉及AARS蛋白片段的选定的非常规活性的增加或减少。在某些实施方案中, 在细胞释放测定或动物模型 (例如小鼠模型) 中测量剂量响应曲线, 其读数涉及细胞死亡或动物死亡。其他变异对于本领域技术人员而言是明显的。

#### [0283] VIII. 表达和纯化系统

[0284] 本发明实施方案包括用于表达和纯化本发明的AARS蛋白片段或基于其他多肽的剂的方法和相关组合物。可以利用以下所述的标准方案方便地制备此类重组AARS多肽: 例如 Sambrook, 等人, (1989, 同上), 特别是第16和17节; Ausubel等人, (1994, 同上), 特别是第10和16章; 以及 Coligan等人, Current Protocols in Protein Science (John Wiley & Sons, Inc. 1995-1997), 特别是第1、5和6章。作为一个一般性实例, 可以通过包括下述步骤的一个或多个的方法来制备AARS多肽: (a) 制备包含多核苷酸序列的构建体, 该多核苷酸序列编码AARS多肽并被可操作地连接到调节元件; (b) 将构建体导入宿主细胞; (c) 培养宿主细胞以表达AARS多肽; 以及 (d) 从宿主细胞分离AARS多肽。

[0285] AARS多核苷酸在本文其他地方描述。为了表达所需的多肽, 可以将编码多肽的核苷酸序列或功能等同物插入合适的表达载体中, 即包含插入的编码序列的转录和翻译所必需的元件的载体。可以用本领域技术人员公知的方法构建含有编码目的多肽的序列和合适的转录和翻译控制元件的表达载体。这些方法包括体外重组DNA技术、合成技术、以及体内

遗传重组。这些技术描述于Sambrook等人, *Molecular Cloning, A Laboratory Manual* (1989), 和Ausubel等人, *Current Protocols in Molecular Biology* (1989)。

[0286] 多种表达载体/宿主系统是已知的,并且可以用来包含和表达多核苷酸序列。这些表达载体/宿主系统包括但不限于微生物,诸如用重组噬菌体、质粒、或粘粒DNA表达载体转化的细菌;用酵母表达载体转化的酵母;用病毒表达载体(例如,杆状病毒)感染的昆虫细胞系统;用病毒表达载体(例如,花椰菜花叶病毒CaMV;烟草花叶病毒TMV)或用细菌表达载体(例如,Ti或pBR322质粒)转化的植物细胞系统;或动物细胞系统,包括哺乳动物细胞和更特别是人类细胞系统。

[0287] 表达载体中存在的“控制元件”或“调节序列”是与宿主细胞蛋白相互作用来进行转录和翻译的载体的那些非翻译区—增强子、启动子、5'和3'非翻译区。这些元件在它们的强度和特异性上可以不同。根据所用的载体系统和宿主,可以使用任何数量的合适的转录和翻译元件,包括组成型启动子和诱导型启动子。例如,当在细菌系统克隆时,可以使用诱导型启动子,例如PBLUESCRIPT噬菌粒(Stratagene, LaJolla, Calif.)或SPORT1质粒(GibcoBRL, Gaithersburg, Md.)的杂合lacZ启动子等。在哺乳动物细胞系统中,通常优选来自哺乳动物基因或来自哺乳动物病毒的启动子。如果生成包含编码多肽的序列的多个拷贝的细胞系是必需的,则基于SV40或EBV的载体可以有利地与适当的选择标志物一起使用。

[0288] 在细菌系统中,根据表达的多肽的希望的用途,可以选择多种表达载体。例如,当大量需要时,可以使用指导容易纯化的融合蛋白的高水平表达的载体。这类载体包括但不限于,诸如BLUESCRIPT(Stratagene)的多功能大肠杆菌克隆和表达载体,其中将编码目的多肽的序列与载体连接,与 $\beta$ -半乳糖苷酶的氨基末端Met和随后的7个残基在同一读码框中,以便产生杂交蛋白;pIN载体(Van Heeke&Schuster, *J. Biol. Chem.* 264:55035509 (1989))等。pGEX载体(Promega, Madison, Wis.)也可以用来将外来多肽表达为具有谷胱甘肽巯基转移酶(GST)的融合蛋白。通常,这类融合蛋白是可溶的,并且可以通过谷胱甘肽琼脂糖珠的吸附并随后在游离谷胱甘肽的存在下洗脱来容易地从裂解的细胞中纯化。这些系统中产生的蛋白可以被设计为包括肝素、凝血酶或因子XA蛋白酶切割位点,以便任意地从GST部分释放克隆的目的多肽。

[0289] 某些实施方案可以采用基于大肠杆菌的表达系统(参见例如,Structural Genomics Consortium等人, *Nature Methods*. 5:135-146, 2008)。这些和相关实施方案可以部分或完全依赖于连接独立性克隆(LIC)以产生适合的表达载体。在具体实施方案中,蛋白表达可以由T7 RNA聚合酶(例如,pET载体系列)控制。这些和相关实施方案可以利用表达宿主菌株BL21(DE3),一种支持T7介导的表达并且为改善的靶蛋白稳定性而缺乏lon和ompT蛋白酶的BL21的 $\lambda$ DE3溶源性细菌。还包括携带编码大肠杆菌中很少使用的tRNA的质粒的表达宿主菌株,例如ROSETTA<sup>TM</sup>(DE3)和Rosetta2(DE3)菌株。还可以使用以商标BENZONASE<sup>®</sup>核酸酶和BUGBUSTER<sup>®</sup>蛋白提取试剂销售的试剂改良细胞溶解和样品操作。对于细胞培养,自诱导培养基可以提高许多表达系统的效率,包括高通量表达系统。该类型的培养基(例如,OVERNIGHT EXPRESS<sup>TM</sup>自诱导系统)通过代谢转变而逐渐引发蛋白表达,而不添加人工诱导剂,例如IPTG。具体实施方案采用六组氨酸标签(例如以商标HIS•TAG<sup>®</sup>融合体销售的那些),随后是固定化金属亲和色谱(IMAC)纯化或相关技术。然

而,在某些方面中,临床级蛋白可分离自大肠杆菌包涵体,使用或没有使用亲和标签(参见例如,Shimp等人,Protein Expr Purif.50:58-67,2006)。作为其他实例,某些实施方案可以采用冷激诱导的大肠杆菌高产率生产系统,因为蛋白在大肠杆菌中低温下的过表达提高了其溶解度和稳定性(参见例如,Qing等人,Nature Biotechnology.22:877-882,2004)。

[0290] 还包括高密度细菌发酵系统。例如,高细胞密度培养真养产碱杆菌(*Ralstonia eutropha*)允许以高于150g/L的细胞密度生产蛋白,并且以超过10g/L的滴度表达重组蛋白。

[0291] 在酿酒酵母(*Saccharomyces cerevisiae*)中,可以使用多种含有诸如 $\alpha$ 因子、醇氧化酶和PGH的组成型或诱导型启动子的载体。关于综述,参见Ausubel等人(同上)和Grant等人,Methods Enzymol.153:516-544(1987)。还包括毕赤酵母(*Pichia pastoris*)表达系统(参见例如Li等人,Nature Biotechnology.24,210-215,2006;和Hamilton等人,Science,301:1244,2003)。某些实施方案包括被工程化为选择性糖基化蛋白的酵母系统,包括具有人源化N-糖基化途径的酵母以及其他酵母(参见例如Hamilton等人,Science.313:1441-1443,2006;Wildt等人,Nature Reviews Microbiol.3:119-28,2005;和Gerngross等人,Nature-Biotechnology.22:1409-1414,2004;美国专利号7,629,163;7,326,681;和7,029,872)。仅作为实例,可以在Fernbach烧瓶或15L、50L、100L和200L发酵罐以及其他中培养重组酵母培养物。

[0292] 如果使用植物表达载体,编码多肽的序列的表达可以由多种启动子中的任何一种所驱动。例如,诸如CaMV的35S和19S启动子的病毒启动子可以单独使用或与来自TMV的 $\Omega$ 前导序列组合使用(Takamatsu,EMBO J.6:307-311(1987))。可选择地,可以使用诸如RUBISCO的小亚基或热休克启动子的植物启动子(Coruzzi等人,EMBO J.3:1671-1680(1984); Broglie等人,Science 224:838-843(1984);和Winter等人,Results Probl.Cell Differ.17:85-105(1991))。可以通过直接的DNA转化或病原体介导的转染将这些构建体引入植物细胞。这类技术描述于很多可普遍获得的综述中(参见例如,Hobbs in McGraw Hill,Yearbook of Science and Technology,pp.191-196(1992))。

[0293] 还可以用昆虫系统来表达目的多肽。例如,在一种这类系统中,将苜蓿银纹夜蛾(*Autographa californica*)核型多角体病毒(AcNPV)用作载体以在草地贪夜蛾(*Spodoptera frugiperda*)细胞或粉纹夜蛾(*Trichoplusia*)细胞中表达外来基因。可以将编码多肽的序列克隆入病毒的非必需区,例如多角体蛋白基因,并置于多角体蛋白启动子的控制下。编码多肽的序列的成功插入会使多角体蛋白基因失活,并产生缺少衣壳蛋白的重组病毒。随后,可以将所述重组病毒用于感染例如草地贪夜蛾细胞或粉纹夜蛾细胞,目的多肽可以在其中表达(Engelhard等人,Proc.Natl.Acad.Sci.U.S.A.91:3224-3227(1994))。还包括杆状病毒表达系统,包括利用了SF9、SF21和Tni细胞的那些(参见例如,Murphy和Piwnica-Worms,Curr Protoc Protein Sci.第5章:第5.4单元,2001)。昆虫系统可以提供类似于哺乳动物系统的翻译后修饰。

[0294] 在哺乳动物宿主细胞中,通常可以使用多种基于病毒的表达系统。例如,如果将腺病毒用作表达载体,则可以将编码目的多肽的序列连入由晚期启动子和三联前导序列组成的腺病毒转录/翻译复合物中。病毒基因组的非必需E1或E3区中的插入可以用来获得能在感染的宿主细胞中表达多肽的活病毒(Logan&Shenk,Proc.Natl.Acad.Sci.U.S.A.81:

3655-3659(1984))。此外,可以使用诸如鲁斯氏肉瘤病毒(RSV)增强子的转录增强子来增强哺乳动物宿主细胞中的表达。

[0295] 有用的哺乳动物宿主细胞系的实例包括由SV40转化的猴肾CV1细胞系(COS-7, ATCC CRL 1651);人胚胎肾细胞系(为在悬浮培养基中生长而亚克隆的293或293细胞, Graham等人, *J. Gen. Virol.* 36:59(1977));幼仓鼠肾细胞(BHK, ATCC CCL 10);小鼠睾丸支持细胞(TM4, Mather, *Biol. Reprod.* 23:243-251(1980));猴肾细胞(CV1 ATCC CCL 70);非洲绿猴肾细胞(VERO-76, ATCC CRL-1587);人类宫颈癌细胞(HELA, ATCC CCL 2);犬肾细胞(MDCK, ATCC CCL 34);水牛鼠肝细胞(BRL 3A, ATCC CRL 1442);人肺细胞(W138, ATCC CCL 75);人肝细胞(Hep G2, HB 8065);小鼠乳腺肿瘤(MMT 060562, ATCC CCL51);TR1细胞(Mather等人, *Annals N.Y. Acad. Sci.* 383:44-68(1982));MRC5细胞;FS4细胞;和人肝癌细胞系(Hep G2)。其他有用的哺乳动物宿主细胞系包括中国仓鼠卵巢(CHO)细胞,包括DHFR-CHO细胞(Urlaub等人, *PNAS USA* 77:4216(1980));和骨髓瘤细胞系,例如NS0和Sp2/0。有关适合抗体生产的某些哺乳动物宿主细胞系的综述,参见例如Yazaki和Wu, *Methods in Molecular Biology*, 第248卷(B.K.C Lo编辑, Humana Press, Totowa, N.J., 2003), pp.255-268。某些优选的哺乳动物细胞表达系统包括基于CHO和HEK293-细胞的表达系统。哺乳动物表达系统可以利用例如在T烧瓶、滚瓶或细胞工厂中附着的细胞系,或者例如在1L和5L旋转器、5L、14L、40L、100L和200L搅拌罐生物反应器、或20/50L和100/200L WAVE生物反应器中的悬浮培养物,以及本领域已知的其他表达系统。

[0296] 还包括蛋白的无细胞表达。这些和相关的实施方案通常利用纯化的RNA聚合酶、核糖体、tRNA和核糖核苷酸;这些试剂可以通过从细胞或从基于细胞的表达系统提取来生产。

[0297] 也可以使用特定的起始信号来实现编码目的多肽的序列的更高效的翻译。这些信号包括ATG起始密码子和邻近的序列。如果将编码多肽的序列、其起始密码子和上游序列插入合适的表达载体中,则不需要其他的转录或翻译控制信号。然而,如果仅将编码序列或其一部分插入,则应该提供包括ATG起始密码子在内的外源翻译控制信号。此外,起始密码子应该在正确的读码框中以确保全部插入物的翻译。外源翻译元件和起始密码子可以是不同来源的,可以是天然的或合成的。可以通过包含适于所用的特定细胞系统的增强子来增强表达效率,诸如文献中描述的那些增强子(Scharf. 等人, *Results Probl. Cell Differ.* 20:125-162(1994))。

[0298] 此外,可以基于宿主细胞株调节插入的序列的表达或以所需的方式加工所表达的蛋白的能力来选择宿主细胞株。多肽的这些修饰包括但不限于翻译后修饰,诸如乙酰化、羧基化、糖基化、磷酸化、脂质化和酰基化。切割蛋白的“前原(prepro)”形式的翻译后加工也可以用来促进正确的插入、折叠和/或功能。除了细菌细胞,可以选择对于这类翻译后活性具有或者甚至缺乏特异性的细胞器和特征机制的诸如酵母、CHO、HeLa、MDCK、HEK293和W138的不同宿主细胞,以确保外来蛋白的正确修饰和加工。

[0299] 为了长期、高产量地产生重组蛋白,通常优选稳定的表达。例如,可以用表达载体转化稳定表达目的多核苷酸的细胞系,所述表达载体可以在同一载体或分别的载体上含有病毒复制起点和/或内源表达元件以及选择标记基因。引入载体后,在将细胞转至选择培养基之前,可以允许细胞在富集培养基中生长约1-2天。选择标记的目的是赋予对于选择的抗性,并且选择标记的存在允许成功表达引入的序列的细胞的生长和恢复。可以用适合细胞

类型的组织培养技术增殖稳定转化的细胞的抗性克隆。还可以采用瞬时生成,例如通过瞬时转染或感染。适合瞬时生成的示例性哺乳动物表达系统包括基于HEK293和CHO的系统。

[0300] 可以用任何数量的选择系统来恢复转化或转导的细胞系。这些选择系统包括,但不限于,可以分别可用于tk-或aprt-细胞的单纯疱疹病毒胸苷激酶(Wigler等人,Cell111:223-232(1977))和腺嘌呤磷酸核糖转移酶(Lowy等人,Cell122:817-823(1990))基因。另外,抗代谢物、抗生素或除草剂抗性可以作为选择的基础;例如,提供甲氨蝶呤抗性的dhfr(Wigler等人,Proc.Natl.Acad.Sci.U.S.A.77:3567-70(1980));提供对氨基糖苷类、新霉素和G-418抗性的npt(Colbere-Garapin等人,J.Mol.Biol.150:1-14(1981));以及分别提供对氯磺隆和草丁膦(phosphinotricin)乙酰转移酶的抗性的als或pat(Murry,同上)。其他选择基因已有描述,例如,允许细胞利用吡啶来代替色氨酸的trpB,或允许细胞利用组胺醇来代替组氨酸的hisD(Hartman&Mulligan,Proc.Natl.Acad.Sci.U.S.A.85:8047-51(1988))。可视标记已获得普遍应用,这类标记如绿色荧光蛋白(GFP)和其他荧光蛋白(例如RFP、YFP)、花青素、 $\beta$ -葡糖苷酸酶及其底物GUS、以及荧光素酶及其底物荧光素,这类标记不仅广泛用于鉴定转化株,而且用于对归因于特异性载体系统的瞬时或稳定蛋白表达进行定量(参见例如,Rhodes等人,Methods Mol.Biol.55:121-131(1995))。

[0301] 本发明的实施方案还包括高通量蛋白生产系统或微量生产系统。某些方面可以利用例如用于在金属螯合物改性的载玻片表面或MagneHis Ni-颗粒上蛋白表达和纯化的六组氨酸融合标签(参见例如,Kwon等人,BMC Biotechnol.9:72,2009;和Lin等人,Methods Mol Biol.498:129-41,2009))。还包括高通量的无细胞蛋白表达系统(参见例如,Sitaraman等人,Methods Mol Biol.498:229-44,2009)。例如,这些和相关的实施方案可以用于产生AARS蛋白片段的微阵列,其然后可用于筛选文库以鉴定与AARS蛋白片段相互作用的剂。

[0302] 用于检测和测量多核苷酸编码的产物的表达的多种方法是本领域已知的,所述方法利用本领域已知的对所述产物特异的多克隆或单克隆抗体。实例包括酶联免疫吸附测定(ELISA)、蛋白质免疫印迹、放射免疫测定(RIA)和荧光活化的细胞分选(FACS)。在Hampton等人,Serological Methods,a Laboratory Manual(1990)和Maddox等人,J.Exp.Med.158:1211-1216(1983)中以及其他地方描述了这些和其他测定。

[0303] 广泛多种的标记和缀合技术是本领域技术人员已知的并且可以用于各种核酸和氨基酸测定中。制备用于检测多核苷酸相关序列的标记的杂交或PCR探针的方法包括寡核苷酸标记、切口平移、末端标记或利用标记的核苷酸的PCR扩增。可选择地,可以将序列或其任何部分克隆入载体中,用于产生mRNA探针。这种载体在本领域是已知的,并且可以商购,并且可以通过添加诸如T7、T3或SP6的合适的RNA聚合酶和标记的核苷酸来体外合成RNA探针。可以使用多种可商业购得的试剂盒进行这些程序。可以使用的合适的报告分子或标记物包括放射性核素、酶、荧光剂、化学发光剂或生色剂以及底物、辅因子、抑制剂、磁性颗粒等。

[0304] 可以在适合蛋白从细胞培养物中表达和回收的条件下,培养用目的多核苷酸序列转化的宿主细胞。某些具体实施方案利用无血清的细胞表达系统。实例包括可以在无血清培养基上生长的HEK293细胞和CHO细胞(参见例如Rosser等人,蛋白Expr.Purif.40:237-43,2005;和美国专利号6,210,922)。



[0305] 根据所用的序列和/或载体,重组细胞产生的蛋白可以是分泌的或包含在细胞内。如本领域技术人员所理解的,可以将包含本发明多核苷酸的表达载体设计为包含指导编码的多肽通过原核生物或真核生物细胞膜分泌的信号序列。可以用其他的重组构造将编码目的多肽的序列连接至编码促进可溶蛋白纯化和/或检测的多肽结构域的核苷酸序列。此类结构域的实例包括可剪切和不可剪切的亲和纯化以及表位标签,例如亲和素、FLAG标签、多聚组氨酸标签(例如6xHis)、cMyc标签、V5-标签、谷胱甘肽S-转移酶(GST)标签和其他。

[0306] 可以根据本领域已知的许多技术来纯化和表征由重组细胞产生的蛋白。用于进行蛋白纯化和分析蛋白纯度的示例性系统包括快速蛋白液相色谱(FPLC)(例如,AKTA和Bio-Rad FPLC系统)、高效液相色谱(HPLC)(例如,Beckman and Waters HPLC)。用于纯化的示例性化学包括离子交换色谱(例如Q,S)、尺寸排阻色谱、盐梯度、亲和纯化(例如Ni、Co、FLAG、麦芽糖、谷胱甘肽、蛋白A/G)、凝胶过滤、反相陶瓷 HYPERD® 离子交换色谱和疏水性相互作用柱(HIC)和本领域已知的其他。还包括分析方法,例如SDS-PAGE(例如,考马斯,银染色)、免疫印迹、Bradford和ELISA,其可以在生产或纯化过程的任何步骤中利用,通常测量蛋白组合物的纯度。

[0307] 还包括浓缩AARS蛋白片段和包含浓缩的可溶性蛋白的组合物的方法。在不同的方面,此类AARS多肽浓缩溶液可以包含浓度约5mg/mL;或约8mg/mL;或约10mg/mL;约15mg/mL;或约20mg/mL的蛋白。

[0308] 一方面,此类组合物可以基本上是单分散的,意味着当例如通过尺寸排阻色谱、动态光散射或分析型超速离心评估时,AARS多肽组合物主要(即,至少约90%或更多)以一种表观分子量形式存在。

[0309] 另一方面,此类组合物具有至少约90%的纯度(以蛋白基计),或者在一些方面至少约95%纯度,或者在某些实施方案中,至少98%纯度。可以通过本领域已知的任何常规分析方法测定纯度。

[0310] 另一方面,与存在的蛋白总量相比,此类组合物具有小于约10%的高分子量聚集物含量,或者在一些实施方案中,此类组合物具有小于约5%的高分子量聚集物含量,或者在一些方面,此类组合物具有小于约3%的高分子量聚集物含量,或者在一些实施方案中,小于约1%的高分子量聚集物含量。高分子量聚集物含量可以通过多种分析型技术测定,包括例如尺寸排阻色谱、动态光散射或分析型超速离心。

[0311] 在某些实施方案中,如本文所述,AARS多肽组合物具有小于约10EU/mg AARS多肽、或小于约5EU/mg AARS多肽、小于约3EU/mg AARS多肽、或小于约1EU/mg AARS多肽的内毒素含量。

[0312] 本文涵盖的浓缩方法的实例包括冻干,其通常在溶液含有很少不同于感兴趣的蛋白的可溶性组分时采用。冻干通常在HPLC运行之后进行,并且可以从混合物去除大部分或所有挥发性组分。还包括超速离心技术,其通常采用一种或多种选择性可透过膜来浓缩蛋白溶液。所述膜允许水和小分子穿过并保留蛋白;可以通过机械泵、气压或离心以及其他技术来迫使溶液穿过膜。

[0313] 在某些实施方案中,如根据本领域常规技术所测量的,试剂、AARS蛋白片段或相关剂(例如抗体)具有至少约90的纯度。在某些实施方案中,例如诊断组合物或某些治疗组合物,本发明的AARS组合物具有至少约95%的纯度。在具体实施方案中,例如治疗或药物组合

物,本发明的AARS组合物具有至少约97%或98%或99%的纯度。在其他实施方案中,例如当用作参照或研究试剂时,AARS蛋白片段可以具有较低的纯度,并且可以具有至少约50%、60%、70%或80%的纯度。可以总体上或者相对于选定的组分(例如其他蛋白)来测量纯度,例如以蛋白基计的纯度。

[0314] 还可以根据其生物特性来表征纯化的AARS蛋白片段。如本文所述,实例包括对选定配体(例如,AARS蛋白片段的细胞结合配体,例如细胞表面受体或其细胞外结构域)的结合亲和力或结合动力学,以及一种或多种常规或非规律生物活性的存在或水平。可以根据本领域已知的多种技术来测量结合亲和力和结合动力学,所述技术例如利用表面等离子体共振 (SPR) 的 **BIACORE®** 和相关技术,表面等离子体共振 (SPR) 是一种光学现象,能够实现未标记的相互作用物的实时检测。就亲和力和动力学两方面而言,基于SPR的生物传感器可用于测定活性浓度、筛选和表征。可以根据基于细胞的测定来测量一种或多种常规或非规律生物活性的存在或水平,包括利用与读数器或指示器功能上偶联的选定的AARS蛋白片段的细胞结合配体(例如细胞表面受体)的那些,所述指示器例如本文所述的非规律生物活性的荧光或发光指示器。

[0315] 在某些实施方案中,如上所述,AARS多肽组合物基本上无内毒素,包括例如约95%无内毒素,优选约99%无内毒素,和更优选约99.99%无内毒素。如本文描述的,可以根据本领域常规技术来检测内毒素的存在。在具体实施方案中,可以在基本上无血清的培养基中从真核细胞例如哺乳动物细胞或人细胞制备AARS组合物。

[0316] 在某些实施方案中,AARS多肽组合物包含小于约10%wt/wt的高分子量聚集物,或小于约5%wt/wt高分子量聚集物,或小于约2%wt/wt高分子量聚集物,或小于约或小于约1%wt/wt高分子量聚集物。

[0317] 还包括基于蛋白的分析型测定和方法,其可用于评估例如蛋白纯度、大小、溶解度和聚集程度以及其他特性。可以多种方式评估蛋白纯度。例如,可以基于主要结构、高级结构、大小、电荷、疏水性和糖基化来评估纯度。用于评估主要结构的方法的实例包括N-端和C-端测序和肽作图(参见例如,Allen等人,Biologicals.24:255-275,1996)。用于评估高级结构的实例包括圆二色谱(参见例如,Kelly等人,Biochim Biophys Acta.1751:119-139,2005)、荧光光谱(参见例如,Meagher等人,J.Biol.Chem.273:23283-89,1998)、FT-IR、酰胺氢-氘交换动力学、差示扫描量热法、NMR光谱、与构象敏感性抗体的免疫反应性。还可以将高级结构评估为多种参数(例如pH、温度或加入的盐)的函数。用于评估蛋白特性(例如大小)的方法的实例包括分析型超速离心和尺寸排阻HPLC (SEC-HPLC),并且用于测量电荷的示例性方法包括离子交换色谱和等电聚焦。例如,可以通过反相HPLC和疏水性相互作用色谱HPLC来评估疏水性。糖基化可以影响药代动力学(例如,清除)、构象或稳定性、受体结合和蛋白功能,并且可以通过例如质谱和核磁共振 (NMR) 光谱来评估。

[0318] 如上所述,某些实施方案包括使用SEC-HPLC来评估蛋白特性,例如纯度、尺寸(例如,尺寸均一性)或聚集程度,和/或纯化蛋白,以及其他用途。SEC,还包括凝胶过滤色谱 (GFC) 和凝胶渗透色谱 (GPC),指其中溶液中的分子根据其尺寸或者更具体说根据其流体动力学体积、扩散系数和/或表面性质而在多孔材料中被分离的色谱方法。该方法一般用于分离生物分子,并且测定聚合物的分子量和分子量分布。通常,将生物样品或蛋白样品(例如,根据本文提供和本领域已知的蛋白表达方法产生的蛋白提取物)加载入具有确定的固定相

(多孔材料)的选择的尺寸排阻柱,所述固定相优选是不与样品中的蛋白相互作用的相。在某些方面中,固定相由在玻璃或不锈钢柱中包装成密集三维基质的惰性颗粒构成。流动相可以是纯水、水缓冲液、有机溶剂或其混合物。固定相颗粒通常具有仅允许低于某一尺寸的分子进入的小孔和/或通道。因此,大颗粒被排除在这些孔和通道之外,并且它们与固定相的有限的相互作用引导它们在实验开始时作为“完全排除的”峰洗脱。可以进入孔中的较小分子从流动的流动相去除,并且它们在固定相孔中的停留时间不分取决于它们进入孔的程度。它们从流动相流的去流导致它们需要更长的时间从柱中洗脱,并且导致颗粒之间根据其尺寸差异而被分离。给定的尺寸排阻柱具有可以被分离的一定范围的分子量。总体而言,大于上限的分子将不被固定相捕获,小于下限的分子将完全进入固定相并作为单条带洗脱,并且该范围内的分子将以不同的速率洗脱,速率由它们的性质例如流体动力学体积决定。关于这些方法在药物蛋白实践中的实例,参见 Bruner 等人, *Journal of Pharmaceutical and Biomedical Analysis*.15:1929-1935,1997。

[0319] 例如, Anicetti 等人 (*Trends in Biotechnology*.7:342-349,1989) 也讨论了临床应用的蛋白纯度。用于分析蛋白纯度的最新技术包括但不限于 LabChip GXII, 一种用于快速分析蛋白和核酸的自动化平台,其提供了蛋白的滴度、大小分级和纯度分析的高通量分析。在某些非限制性实施方案中,可以通过利用色谱材料的组合以至少两个正交步骤以及其他方法来获得临床级蛋白,例如蛋白片段和抗体(参见例如, *Therapeutic proteins: Methods and Protocols*. 第308卷,编辑 Smales 和 James, Humana Press Inc., 2005)。通常,如根据本领域已知和本文描述的技术所测量的,蛋白剂(例如, AARS 蛋白片段、抗体、结合剂)和其他剂(例如,反义、RNAi、小分子)基本上无内毒素。

[0320] 还包括蛋白溶解度测定。例如,可以利用此类测定来测定重组生产的最佳生长和纯化条件,以优化缓冲液的选择,并优化 AARS 蛋白片段或其变体的选择。可以根据许多参数评价溶解度或聚集,包括温度、pH、盐和其他添加剂的存在或不存在。溶解度筛选测定的实例包括但不限于,使用浑浊度或其他量度作为终点来测量蛋白溶解度的基于微板的方法,用于分析纯化的重组蛋白的溶解度的高通量测定(参见例如, Stenvall 等人, *Biochim Biophys Acta*.1752:6-10,2005),利用遗传标记蛋白的结构互补来监测和测量蛋白的体内折叠和溶解度的测定(参见例如, Wigley 等人, *Nature Biotechnology*.19:131-136,2001),以及使用扫描电化学显微镜 (SECM) 在大肠杆菌中电化学筛选重组蛋白溶解度(参见例如, Nagamine 等人, *Biotechnology and Bioengineering*.96:1008-1013,2006),以及其他。可以根据本领域常规技术鉴定或选择具有增加的溶解度(或减少的聚集)的 AARS 蛋白片段,包括对于蛋白溶解度的简单的体内测定(参见例如, Maxwell 等人, *Protein Sci*.8:1908-11,1999)。

[0321] 还可以通过动态光散射技术来测量蛋白溶解度和聚集。聚集是涵盖几种类型的相互作用或特性的一般术语,包括可溶的/不可溶的、共价的/非共价的、可逆的/不可逆的、和天然的/变性的相互作用和特性。对于蛋白治疗剂,聚集物的存在通常被认为是不希望的,因为考虑聚集物可能引起免疫原性反应(例如,小聚集物)或者可能引起有关施用的不良事件(例如,微粒)。动态光散射指可用于测定悬浮液或聚合物(例如溶液中的蛋白)中小颗粒的尺寸分布特征的技术。该技术还称为光相关光谱法 (PCS) 或拟弹性光散射 (QELS),利用散射光来测量蛋白颗粒的扩散速率。由于溶液中分子和颗粒的布朗运动,可以观察到散射强

度的波动。可以规范加工该运动数据以推导样品的尺寸分布,其中尺寸由蛋白颗粒的斯托克斯半径或流体动力学半径给出。流体动力学半径取决于质量和形状(构象)。动态散射可以监测极少量聚集蛋白(<0.01%重量)的存在,即使是在含有大范围质量的样品中。它还可用于比较不同制剂的稳定性,包括例如依赖实时监测升高温度下变化的应用。因此,某些实施方案包括使用动态光散射来分析含有本发明的AARS蛋白片段、抗体或其他剂的样品中聚集物的溶解度和/或存在。

#### [0322] IX. 诊断方法和组合物

[0323] 本文所述的AARS剂诸如AARS蛋白片段、AARS多核苷酸和抗体以及其他结合剂可以用于诊断测定和诊断组合物。包括生化、组织和基于细胞的方法和组合物以及其他。

[0324] 这些和相关的实施方案包括检测一种或多种新鉴定的AARS蛋白片段(还称为AARS多肽)的AARS多核苷酸序列或相应的AARS多肽序列或其一部分。例如,某些方面包括检测一种或多种新鉴定的AARS剪接变体和/或这些剪接变体的一个或多个剪接点的AARS多核苷酸序列或相应的多肽序列或其部分。在某些实施方案中,所述剪接点的至少一个的多核苷酸或相应的多肽序列对于特定AARS剪接变体是独特的。

[0325] 还包括直接检测AARS蛋白片段,包括剪接变体、蛋白水解片段和其他。在某些实施方案中,一种或多种新鉴定的AARS蛋白片段的存在或水平与一种或多种细胞类型或细胞状态有联系或相关。因此,AARS多肽或多核苷酸的存在或水平可以用于区分不同的细胞类型或不同的细胞状态。如本文描述和本领域已知的,可以根据基于多核苷酸和/或多肽的诊断技术来检测AARS蛋白片段及其相关多核苷酸的存在或水平。

[0326] 某些方面可以采用AARS蛋白片段、抗体或AARS多核苷酸作为伴随诊断方法的部分,通常来评估一个受试者或者群体受试者是否顺利地响应特定医疗。例如,基于受试者是否具有针对给定疾病或病症选定的一个或多个生物标志,可以将给定的AARS治疗剂(例如,蛋白片段、反义、RNAi、抗体、结合剂)鉴定为适合一个受试者或某个群体的受试者。生物标志的实例包括血清/组织标志以及可以通过医学成像技术鉴定的标志。在某些实施方案中,天然存在的AARS蛋白片段(或其相应的多核苷酸)本身可以提供可用于测量药物结果或评估药物使用在一个特定受试者或特定群体的受试者中的价值的血清和/或组织生物标志。在某些方面中,AARS多肽或多核苷酸参照序列的鉴定可以包括鉴定无论是在选定的受试者、选定的组织还是其他中该序列的差异表达,如本文所述和本领域已知的。

[0327] 本文提供的某些方法依赖于AARS多肽或多核苷酸的差异表达,来表征细胞、组织或受试者的病症或状态并且将其与另一细胞、组织或受试者区分。非限制性的实例包括,检测生物样品中AARS多肽或多核苷酸的存在或水平以区分不同类型的细胞或组织,不同组织或器官的细胞,诸如新生的和成体的细胞发育状态,细胞分化状态,诸如健康、患病和受治疗的状况,细胞内和细胞外级分,以及原代细胞培养物和其他细胞培养物,例如永生化细胞培养物。

[0328] 差异表达包括AARS多核苷酸或多肽参照序列的一种或多种基因表达水平与相同序列在合适的对照中的表达水平相比较的统计学显著差异。统计学显著差异可以指所测量的表达水平的增加或降低,所述测量是通过RNA水平、蛋白水平、蛋白功能或诸如本文所述的那些基因表达的任何其他相关测量进行的。还包括本发明AARS多核苷酸或多肽和通常是相同或相应类型的全长或野生型胞质或线粒体AARS序列之间的比较。可以通过本领域和本

文描述的许多技术来检测差异表达,包括基于多核苷酸和多肽的技术,例如实时PCR、差减杂交、多核苷酸和多肽阵列和其他。

[0329] 如果结果不可能是偶然发生的,则结果通常被称为统计学上显著的。测试或结果的显著性水平通常涉及概率统计假定检验概念。在简单情况下,统计学上的显著性可以被定义为当零假设事实上为真时做出否决该零假设的判定的概率(被称为I型错误或“假阳性确定”的判定)。该判定通常用p值进行:如果p值小于显著性水平,则否决零假设。p值越小,结果越显著。贝叶斯因子也可用于确定统计学上的显著性(参见例如Goodman S., Ann Intern Med 130:1005-13,1999)。

[0330] 在更复杂的但实际上重要的情况下,测试或结果的显著性水平可以反映在这样的分析中,其中当零假设事实上为真的时做出否决该零假设的判定的概率不超过规定的概率。该类分析允许那些应用,其中决定否决的概率可以远小于零假设中所包含的某些假定集合的显著性水平。

[0331] 在某些示例性的实施方案中,统计学上显著的差异表达可以包括这样的情况,其中相对于合适的对照,可疑的生物样品中给定的AARS序列的表达水平提供表达上至少约1.2X、1.3X、1.4X、1.5X、1.6X、1.7X、1.8X、1.9X、2.0X、2.2X、2.4X、2.6X、2.8X、3.0X、4.0X、5.0X、6.0X、7.0X、8.0X、9.0X、10.0X、15.0X、20.0X、50.0X、100.0X或更大的差异(即,可以更高或更低表达的差异表达),包括期间所有整数和小数(例如,1.24X、1.25X、2.1X、2.5X、60.0X、75.0X等)。在某些实施方案中,统计学上显著的差异表达可以包括这样的情况,其中相对于合适的对照,可疑的生物样品中给定的AARS序列的表达水平提供表达上至少约4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000百分比(%)或更大的差异(即,可以更高或更低的差异表达),包括其间所有整数和小数。

[0332] 作为额外的实例,还可以通过进行本文所述及领域内已知的Z检验确定差异表达,即计算绝对Z评分(参见实施例1)。Z检验通常用于鉴定样品平均值和群体平均值之间的显著性差异。例如,相对于95%置信区间(即5%的显著性水平)的标准正常表(例如,对照组),绝对值大于1.96的Z评分指示非随机性。对于99%的置信区间,如果绝对Z大于2.58,意味着 $p < .01$ ,并且差异更加显著—可以更高置信地否决零假设。在这些和相关实施方案中,1.96、2、2.58、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或更高的绝对Z评分,包括其间所有小数(例如,10.1、10.6、11.2等),可以为统计学上的显著性提供有力标准。在某些实施方案中,大于6的绝对Z评分提供特别高的统计学上的显著性。

[0333] 基本上相似地通常是指生物样品与参考对照之间的表达水平缺乏统计学上的显著性差异。基本上相似的表达水平的实例可以包括这样的情况,其中相对于参照样品,可疑的生物样品中给定的SSCIGS的表达水平提供表达上少于约.05X、0.1X、0.2X、0.3X、0.4X、0.5X、0.6X、0.7X、0.8X、0.9X、1.0X、1.1X、1.2X、1.3X或1.4X的差异(即,可以更高或更低表达的差异表达),包括其间所有小数(例如,.15X、0.25X、0.35X等)。在某些实施方案中,差异表达可以包括这样的情况,其中相对于参照样品,可疑的生物样品中给定的AARS序列的表达水平提供表达上少于约0.25、0.5、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50百分比(%)的差异(即,可以更高或更低的差异表达),包括其间所有小数。

[0334] 在某些实施方案中,例如当使用Affymethx Microarray微阵列测量AARS多核苷酸

或多肽参照序列的表达水平时,还可以通过平均表达值确定差异表达,所述平均表达值由 Affymethx Microarray Suite5 软件 (Affymetrix, Santa Clara, CA) 或其他相似软件所概括,通常折合成 1000 的平均表达值。

[0335] 本发明的实施方案包括检测 AARS 多核苷酸或多肽参照序列或其部分的存在或水平以区分不同生物或物种的细胞或组织或其他生物样品的方法,其中该序列的存在或水平与选择的生物或物种相关。通常的实例包括在人与细菌、真菌、植物和其他非人动物的任意组合之间进行区分的方法。动物包括在人与脊椎动物和无脊椎动物的任意组合之间进行区分的方法,包括诸如鱼、两栖动物、爬行类、鸟和非人哺乳动物的脊椎动物以及诸如昆虫、软体动物、甲壳类和珊瑚虫等无脊椎动物。非人哺乳动物包括在人与非人哺乳动物的任意组合之间进行区分的方法,所述非人哺乳动物来自非洲猿目、象鼩目、管齿目、蹄兔目、长鼻目、海牛目、带甲目、披毛目、树鼩目、皮翼目、灵长目、啮齿目、兔形目、猬形目、鼯形目、翼手目、鳞甲目、鲸目、食肉目、奇蹄目或偶蹄目。灵长目包括猴子、猿、大猩猩和黑猩猩以及领域内已知的其他种类。因此,本文所述的 AARS 多核苷酸或多肽参照序列或变体的存在或水平可以用于鉴定诸如细胞、组织或器官的给定的生物样品的来源,所述鉴定通过区分这些生物的任意组合,或通过区分人与这些生物的任意一种或多种 (例如生物组) 来进行。在某些实施方案中,还可以通过将 AARS 序列或其部分的存在或水平与预定值比较来确定给定的生物样品的来源。

[0336] 本发明的实施方案包括检测 AARS 多核苷酸或多肽参照序列或其部分的存在或水平,以区分来自不同组织或器官的细胞或其他生物样品的方法。非限制性的实例包括区分来自以下的任意组合的细胞或其他生物样品的方法:皮肤 (例如,真皮、表皮、皮下组织层)、毛囊、神经系统 (例如,脑、脊髓、外周神经)、听觉系统或平衡器官 (如,内耳,中耳,外耳)、呼吸系统 (例如,鼻、气管、肺)、胃食管组织、胃肠道系统 (例如,口腔、食道、胃、小肠、大肠、直肠)、血管系统 (例如,心脏、血管和动脉)、肝脏、胆囊、淋巴/免疫系统 (例如,淋巴结、淋巴滤泡、脾脏、胸腺、骨髓)、泌尿生殖系统 (如,肾、输尿管、膀胱、尿道、子宫颈、输卵管、卵巢、子宫、外阴、前列腺、尿道球腺体、附睾、前列腺、精囊、睾丸)、肌肉骨骼系统 (例如,骨骼肌、平滑肌、骨、软骨、肌腱、韧带)、脂肪组织、乳房以及内分泌系统 (例如,下丘脑、垂体、甲状腺、胰腺、肾上腺)。因此,根据本文所述的 AARS 多核苷酸或多肽序列的联合,这些方法可以用于鉴定或表征细胞或其他生物样品所来源的组织或器官。

[0337] 本发明的实施方案包括检测 AARS 多核苷酸或多肽参照序列或其部分的存在或水平,以区分或表征细胞的发育或分化状态的方法。还包括区分生殖细胞、干细胞和体细胞的方法。发育状态的实例包括新生的和成体的。细胞分化状态的实例包括全能细胞、多能细胞、多能祖干细胞和成熟的、完全分化的细胞之间的所有精细且可确认的阶段。

[0338] 全能细胞具有全部潜能,通常在有性和无性繁殖时出现,并包括孢子与合子,尽管在某些情况下细胞可以去分化并恢复全能性。多能细胞包括具有分化为三种胚层中任何一种的潜能的干细胞,所述胚层包括内胚层 (内胃壁、胃肠道、肺)、中胚层 (肌肉、骨骼、血液、泌尿生殖系统) 和外胚层 (表皮组织和神经系统)。多潜能祖细胞通常可以分化为有限数目的组织类型。多潜能细胞的实例包括,但不限于,来自骨髓的产生诸如红细胞、白细胞和血小板的免疫细胞的造血干细胞 (成体干细胞),来自骨髓的产生基质细胞、脂肪细胞和各种类型的骨细胞的间质干细胞 (成体干细胞),产生各种类型的皮肤细胞的上皮干细胞 (祖细

胞)和产生分化的肌肉组织的肌卫星细胞(祖细胞)。因此,与对照或预定水平相比,特定AARS多核苷酸或多肽序列(例如,AARS剪接变体的剪接点、AARS蛋白水解片段)的存在或水平可用于区分或表征上述细胞分化状态。

[0339] 本发明的实施方案包括检测AARS多核苷酸或多肽参照序列的存在或水平以表征或诊断细胞、组织、器官或受试者的状况的方法,其中该状况可以表征为健康的、患病的、有患病风险的或受治疗的。为了该诊断目的,术语“诊断的(diagnostic)”或“诊断的(diagnosed)”包括鉴定病理状态的存在或性质,表征发展为这种状态的风险和/或检测该病理状态响应于治疗的改变(或未改变)。诊断方法在其灵敏度和特异性上有差异。在某些实施方案,诊断测定的“灵敏度”是指测试为阳性的患病细胞、组织或个体的百分比(“真阳性”的百分比)。未被该测定检测到的患病细胞、组织或个体通常被称为“假阴性”。未患病且在测定中测试为阴性的细胞、组织或个体可以称为“真阴性”。在某些实施方案中,诊断测定的“特异性”可以定义为 $1 - (\text{假阳性比例})$ ,其中“假阳性”比例被定义为没有疾病且测试为阳性的那些样品或个体的比例。虽然特定诊断方法可能不能提供状态的最终诊断,但是如果该方法能提供辅助诊断的阳性指示,它就足够了。

[0340] 在某些情况下,可以通过将与病理状态相关的一种或多种选择的AARS多核苷酸或多肽参照序列或其部分的存在或水平与合适的对照相比(无论水平是增加的还是降低)来诊断病理状态的存在或发展为病理状态的风险。“合适的对照”或“适当的对照”包括组织或生物的细胞或其他生物样品中确定的值、水平、特征、特性或性质,例如,表现出诸如不存在所述状态的正常性状的对照或正常细胞、组织或生物。在某些实施方案中,“合适的对照”或“适当的对照”是预先确定的值、水平、特征、特性或性质。其他合适的对照对本领域技术人员是显而易见的。疾病和病症的实例在本文他处描述。

[0341] 本发明的实施方案包括基于AARS多核苷酸或核酸的检测技术,该技术由于检测的灵敏度而提供某些优势。因此,某些实施方案涉及作为诊断方法或测定的一部分的AARS多核苷酸的使用或检测。AARS多核苷酸的存在和/或水平可以通过本领域已知的任何方法测量,包括诸如Northern印迹的杂交测定、定量或定性聚合酶链式反应(PCR)、定量或定性逆转录酶PCR(RT-PCR)、微阵列、点印迹或槽印迹、或诸如荧光原位杂交(FISH)的原位杂交以及其他。这些方法中的某些在下文更详细描述。

[0342] 可以通过本领域内已知的方法从血液、生物流体、组织、器官、细胞系或其他相关样品中收集和/或产生诸如DNA和RNA的AARS多核苷酸,例如描述于以下的那些方法: Kingston. (2002 Current Protocols in Molecular Biology, Greene Publ. Assoc. Inc. & John Wiley & Sons, Inc., NY, NY (参见例如,如描述于Nelson等人 Proc Natl Acad Sci U S A, 99:11890-11895, 2002)等。而且,许多商业途径可获得的用于构建RNA的试剂盒用于制备用于本发明的RNA。可以从获自正常健康受试者的器官/组织/细胞来构建RNA;然而,本发明还涵盖从患病受试者构建RNA。某些实施方案考虑使用来自任何类型的受试者或动物的任何类型的器官。对于测试样品,RNA可以获自具有或没有可见疾病的个体(例如,任何动物,包括哺乳动物)和组织样品、生物流体(例如全血)等。

[0343] 在某些实施方案中,cDNA序列的扩增或构建可能有助于提高检测能力。本公开以及本领域提供了进行此类任务的细节的必需水平。在一个示例性实施方案中,全血被用作RNA来源,因此,RNA稳定试剂是任选使用的,例如PAX管,如例如描述于Thach等人,

J. Immunol. Methods. Dec 283 (1-2) : 269-279, 2003和Chai等人, J. Clin. Lab Anal. 19 (5) : 182-188, 2005 (两篇参考文献均通过引用并入)。可以使用本领域已知的技术产生互补DNA (cDNA) 文库, 例如描述于以下文献中的那些技术: Ausubel等人 (2001 Current Protocols in Molecular Biology, Greene Publ. Assoc. Inc. & John Wiley & Sons, Inc., NY, NY); Sambrook等人 (1989 Molecular Cloning, Second Ed., Cold Spring Harbor Laboratory, Plainview, NY); Maniatis等人 (1982 Molecular Cloning, Cold Spring Harbor Laboratory, Plainview, NY) 和其他。而且, 用于构建cDNA文库的许多可商业途径获得的试剂盒可用于制备本发明的cDNA文库。可以从获自正常的健康受试者的器官/组织/细胞构建文库。

[0344] 某些实施方案可以利用用于检测AARS多核苷酸序列的杂交方法。用于进行多核苷酸杂交测定的方法在领域内发展良好。杂交测定程序和条件根据应用而变化, 并且根据已知的通用结合方法进行选择, 包括描述于以下的那些方法: Maniatis等人 Molecular Cloning: A Laboratory Manual (第2版, Cold Spring Harbor, N.Y., 1989); Berger和Kimmel Methods in Enzymology, 第152卷, Guide to Molecular Cloning Techniques (Academic Press, Inc., San Diego, Calif., 1987); Young和Davis, PNAS. 80:1194 (1983)。进行重复和受控的杂交反应的方法和装置已经描述于美国专利第5,871,928号、第5,874,219号、第6,045,996号和第6,386,749号、第6,391,623号, 通过引用将每篇并入本文。

[0345] 某些实施方案可以利用用于检测AARS多核苷酸序列的核酸扩增方法。术语“扩增”或“核酸扩增”是指包含预期的特异性靶核酸序列的至少一部分的靶核酸的多个拷贝的产生。多个拷贝可以被称为扩增子或扩增产物。在某些实施方案中, 扩增的靶所包含的序列小于完整的靶基因序列 (内含子和外显子) 或表达的靶基因序列 (外显子和位于侧翼的非翻译序列的剪接的转录物)。例如, 可以利用与靶多核苷酸的内部位置杂交并从该位置启动聚合的扩增引物通过扩增该靶多核苷酸的一部分来产生特异的扩增子。优选地, 扩增出的部分包含可以利用多种公知方法的任一种进行检测的可检测的靶序列。

[0346] 本文所使用的“选择性扩增”或“特异性扩增”指扩增本发明的靶核酸序列, 其中靶序列的可检测的扩增基本上限于从待测试的目的核酸样品扩增靶序列而不是来自某些其他样品来源的靶核酸序列扩增, 所述其他样品来源例如在扩增反应期间使用的试剂中或进行扩增反应的环境中存在的污染。

[0347] 术语“扩增条件”指允许按照本发明进行核酸扩增的条件。在某些实施方案中, 扩增条件可以不如本文所述的“严格杂交条件”严格。在扩增条件下, 本发明的扩增反应中使用的寡核苷酸与它们的预期靶杂交, 但在严格杂交条件下可能杂交也可能不杂交。另一方面, 本发明的检测探针通常在严格杂交条件下杂交。本领域技术人员根据所利用的特定扩增方法可易于确定按照本发明实施核酸扩增的可接受的条件。

[0348] 许多公知的核酸扩增方法需要热循环以交替地使双链核酸变性并使其与引物杂交; 然而, 其他公知的核酸扩增方法是等温的。通常被称为PCR的聚合酶链式反应 (美国专利第4,683,195号、第4,683,202号、第4,800,159号、第4,965,188号) 使用变性, 引物对与互补链的退火和引物延伸的多个循环以指数方式增加靶序列的拷贝数。在被称作RT-PCR的变形中, 使用逆转录酶 (RT) 从mRNA产生互补DNA (cDNA), 然后通过PCR扩增cDNA以产生DNA的多个拷贝。



[0349] 如上所述,术语“PCR”指选择性地扩增靶核酸种类的多个扩增循环。包括定量PCR (qPCR)、实时(PCR)、逆转录PCR (RT-PCR) 和定量逆转录PCR (qRT-PCR) 在领域内有良好的描述。术语“pPCR”指定量聚合酶链式反应,术语“qRT-PCR”是指定量逆转录聚合酶链式反应。qPCR和qRT-PCR可以用于扩增靶向的cDNA分子并同时对其进行定量。它使得可以对cDNA库中的特定序列,例如选择的AARS基因或转录物,同时进行检测和定量。

[0350] 术语“实时PCR”可以使用DNA结合染料,与PCR中所有双链(ds)DNA结合,引起染料发出荧光。因此,PCR期间DNA产物的增加导致荧光强度的增加,并在每个循环进行测量,从而允许对DNA浓度进行定量。然而,诸如SYBRGreen的dsDNA染料将与所有dsDNA PCR产物结合。在实时PCR热循环仪中检测并测量荧光,并且与产物的指数增加相对应的荧光的几何级数增加用来确定每个反应中的阈值循环(“CT”)。

[0351] 术语“Ct评分”指阈值循环数,这是PCR扩增已超过阈值水平时的循环。如果样品中特定基因的mRNA的量较高,则其将比低表达的基因更早跨越阈值,因为有更多的起始RNA用于扩增。因此,Ct评分低表示样品中的基因表达高,Ct评分高表示基因表达低。

[0352] 某些实施方案可以采用通常被称为LCR的连接酶链式反应(Weiss,R.1991, Science254:1292),其使用与靶核酸的邻近区域杂交的两套互补的DNA寡核苷酸。DNA寡核苷酸可以在热变性、杂交和连接的重复循环中通过DNA连接酶共价连接,从而产生可检测的双链连接的寡核苷酸产物。

[0353] 另一方法是链置换扩增(Walker,G.等人,1992,Proc.Natl.Acad.Sci.USA89:392-396;美国专利号5,270,184和5,455,166),通常称为SDA,其利用以下的循环:使引物序列与靶序列的相对链退火配对,在dNTPs存在下的引物延伸以产生双链体半硫代磷酸酯化的引物延伸产物,半修饰的限制内切核酸酶识别位点的内切核酸酶介导的切割,和聚合酶介导的从切口的3'端的引物延伸以置换现有链并产生用于下一轮引物退火、切割和链置换的链,导致产物的几何扩增。嗜热的SDA (tSDA) 以基本相同的方法在更高的温度下使用嗜热的内切核酸酶和聚合酶(欧洲专利号0 684 315)。

[0354] 其他扩增方法包括例如:基于核酸序列的扩增(美国专利号5,130,238),通常称为NASBA;利用通常称为QB复制酶的RNA复制酶扩增探针分子本身的方法(Lizardi,P.等人,1988,BioTechnol.6:1197-1202);基于转录的扩增方法(Kwoh,D.等人,1989, Proc.Natl.Acad.Sci.USA86:1173-1177);自维持序列复制(Guatelli,J.等人,1990, Proc.Natl.Acad.Sci.USA87:1874-1878);和转录介导的扩增(美国专利号5,480,784和5,399,491),常称为TMA。关于已知扩增方法的进一步讨论,参见Persing,David H.,1993,“In Vitro Nucleic Acid Amplification Techniques”,Diagnostic Medical Microbiology: Principles and Applications(Persing等人编辑),pp.51-87(American Society for Microbiology,Washington,DC)。

[0355] 本发明的示例性的基于转录的扩增系统包括TMA,其利用RNA聚合酶来产生靶区域的多个RNA转录物(美国专利号5,480,784和5,399,491)。TMA利用了逆转录酶和RNA聚合酶存在下与靶核酸杂交以形成双链启动子的“启动子-引物”,RNA聚合酶从所述双链启动子产生RNA转录物。这些转录物可以变成在能够与RNA转录物杂交的第二引物存在下其他轮次TMA的模板。不同于PCR、LCR或需要热变性的其他方法,TMA是等温方法,其利用RNA酶H活性来消化RNA:DNA杂交体的RNA链,从而使DNA链可用于与引物或启动子-引物杂交。一般而言,

利用了为扩增提供的逆转录酶相关的RNA酶H活性。

[0356] 在一个示例性的TMA方法中,一个扩增引物是包含启动子序列的寡核苷酸启动子-引物,当双链时变成功能性的,位于靶结合序列的5',靶结合序列能够在待扩增序列的3'位置与靶RNA的结合位点杂交。当针对T7RNA聚合酶识别时,启动子-引物可以被称为“T7引物”。在某些情况下,启动子-引物此类启动子-引物的亚群的3'端可以被修饰以阻断或减少引物延伸。从未修饰的启动子-引物,逆转录酶产生靶RNA的cDNA拷贝,而RNA酶H活性降解靶RNA。然后第二扩增引物结合至cDNA。该引物可以被称为“非T7引物”以使之区别于“T7引物”。从该第二扩增引物,逆转录酶产生另一条DNA链,得到在一端具有功能性启动子的双链DNA。当双链时,启动子序列能够结合RNA聚合酶以开始启动子-引物杂交的靶序列的转录。RNA聚合酶利用该启动子序列产生多个RNA转录物(即,扩增子),一般约100至1,000个拷贝。每个新合成的扩增子可以与第二扩增引物退火。然后逆转录酶可以产生DNA拷贝,而RNA酶H活性降解该RNA:DNA双链体的RNA。然后,启动子-引物可以结合新合成的DNA,允许逆转录酶产生双链DNA,由此RNA聚合酶产生多个扩增子。因此,使用两个扩增引物可以实现十亿倍的等温扩增。

[0357] 在某些实施方案中,其他技术可用于评估来自特定cDNA文库转录本的RNA转录本,包括微阵列分析(Han,M.,等人,Nat Biotechnol,19:631-635,2001;Bao,P.,等人,Anal Chem,74:1792-1797,2002;Schena等人,Proc.Natl.Acad.Sci.USA 93:10614-19,1996;和Heller等人,Proc.Natl.Acad.Sci.USA 94:2150-55,1997)和SAGE(基因表达系列分析)。与MPSS相似,SAGE是数字化的并可以产生大量的信号序列(参见例如Velculescu,V.E.,等人,Trends Genet,16:423-425.,2000;Tuteja R.和Tuteja N.Bioessays.2004 Aug;26(8):916-22),但是数量级小于可从诸如MPSS的技术获得的信号序列。

[0358] 在某些实施方案中,术语“微阵列”包括具有与基底结合的多个核酸的“核酸微阵列”,与所述多个结合的核酸中的每个的杂交可被分别检测。基底可以是实心的或多孔的,平面的或非平面的,整体的或分散的。核酸微阵列包括在Schena(编辑),DNA Microarrays: A Practical Approach(Practical Approach Series),Oxford University Press(1999);Nature Genet.21(1)(增刊):1-60(1999);Schena(编辑),Microarray Biochip: Tools and Technology,Eaton Publishing Company/BioTechniques Books Division(2000)中所述的所有装置。核酸微阵列可以包含与基底结合的多个核酸,其中所述多个核酸位于多个珠子上而不是整体的平面基底上,例如如在Brenner等人,Proc.Natl.Acad.Sci.USA97(4):1665-1670(2000)中所描述的。核酸微阵列的实例可以见于美国专利第6,391,623号、第6,383,754号、第6,383,749号、第6,380,377号、第6,379,897号、第6,376,191号、第6,372,431号、第6,351,712号、第6,344,316号、第6,316,193号、第6,312,906号、第6,309,828号、第6,309,824号、第6,306,643号、第6,300,063号、第6,287,850号、第6,284,497号、第6,284,465号、第6,280,954号、第6,262,216号、第6,251,601号、第6,245,518号、第6,263,287号、第6,251,601号、第6,238,866号、第6,228,575号、第6,214,587号、第6,203,989号、第6,171,797号、第6,103,474号、第6,083,726号、第6,054,274号、第6,040,138号、第6,083,726号、第6,004,755号、第6,001,309号、第5,958,342号、第5,952,180号、第5,936,731号、第5,843,655号、第5,814,454号、第5,837,196号、第5,436,327号、第5,412,087号和第5,405,783号,通过引用将这些专利的公开内容并入本文。

[0359] 其他实例包括可从Affymetrix (Santa Clara, Calif.) 以商标名GENECHIP™商购的核酸阵列。制备和使用阵列的其他示例性的方法提供于例如美国专利第7,028,629号、第7,011,949号、第7,011,945号、第6,936,419号、第6,927,032号、第6,924,103号、第6,921,642号和第6,818,394号中。

[0360] 涉及阵列和微阵列的本发明还考虑连接到固体基底的聚合物的多种用途。这些用途包括基因表达监测、谱型分析、文库筛选、基因分型和诊断。基因表达监测和谱型分析方法以及用于基因表达监测和谱型分析的方法描述于美国专利第5,800,992号、第6,013,449号、第6,020,135号、第6,033,860号、第6,040,138号、第6,177,248号和第6,309,822号中。基因分型及其用途描述于美国申请系列第10/442,021号、第10/013,598号(美国申请第2003/0036069号)以及美国专利第5,925,525号、第6,268,141号、第5,856,092号、第6,267,152号、第6,300,063号、第6,525,185号、第6,632,611号、第5,858,659号、第6,284,460号、第6,361,947号、第6,368,799号、第6,673,579和第6,333,179号中。可以与本文所公开的方法联合使用的核酸扩增、标记和分析的其他方法示例于美国专利第5,871,928号、第5,902,723号、第6,045,996号、第5,541,061号和第6,197,506号。

[0361] 如本文所述,某些实施方案可以利用诸如引物或探针的寡核苷酸用于扩增或检测,这对本领域技术人员是显而易见的。可以通过本领域普通技术人员已知的技术产生确定序列和化学结构的寡核苷酸,例如通过化学或生化合成,以及通过从重组核酸分子例如细菌或病毒载体的体外或体内表达。在某些实施方案中,寡核苷酸不仅仅由野生型染色体DNA或其体内转录产物组成。

[0362] 可以任何方式修饰寡核苷酸或引物,只要给定修饰与给定寡核苷酸的预期功能相容。本领域普通技术人员能够容易确定给定修饰是否适合本发明任何给定寡核苷酸或是其所需的。本文其他地方更详细描述了相关的AARS寡核苷酸。

[0363] 如本文所述,尽管寡核苷酸的设计和序列取决于其功能,但是通常应考虑数个变量。最相关的变量有:长度、解链温度( $T_m$ )、特异性,与系统内其他寡核苷酸的互补性、G/C含量、多嘧啶(A、G)或多嘌呤(T、C)区段以及3'端序列。控制这些和其他变量是寡核苷酸设计标准和公知的方面,容易获得各种计算机程序来筛选大量的潜在寡核苷酸以获得最佳的寡核苷酸。

[0364] 因此,某些实施方案包括检测样品中靶AARS多核苷酸的方法,所述多核苷酸包含本文所述的参照AARS多核苷酸的序列,所述方法包括a)将样品与探针杂交,所述探针包含与样品中靶多核苷酸互补的序列,并且在所述探针和所述靶多核苷酸或其片段之间形成杂交复合物的条件下,所述探针与所述靶多核苷酸特异性杂交,和b)检测是否存在所述杂交复合物,并且任选地,如果存在,检测其含量。还包括检测样品中靶AARS多核苷酸的方法,所述多核苷酸包含本文所述的参照AARS多核苷酸序列,所述方法包括a)扩增所述靶多核苷酸或其片段,和b)检测是否存在所述扩增的靶多核苷酸或其片段,并且任选地,如果存在,检测其含量。具体实施方案涉及检测AARS剪接变体,例如通过检测剪接变体的独特剪接点,无论是通过杂交、扩增还是其他检测方法。

[0365] 本发明的实施方案包括多种基于AARS多肽的检测技术,包括基于抗体的检测技术。这些实施方案包括用AARS多肽来产生抗体或其他结合剂的用途,然后所述抗体或其他结合剂可以用于诊断方法和组合物中,从而检测通常来自受试者的细胞或其他生物样品中

所选的AARS多肽或对其进行定量。

[0366] 某些实施方案可以利用标准方法和检测器,例如蛋白质印迹和免疫沉淀反应,酶联免疫吸附测定(ELISA)、流式细胞术和利用成像设备的免疫荧光测定(IFA)。这些公知的方法通常利用本文所述的一种或多种单克隆或多克隆抗体,所述抗体与本发明的选择的AARS多肽或该AARS多肽的独特区域特异性结合,并通常不与诸如全长AARS多肽的其他AARS多肽显著结合。在某些实施方案中,AARS多肽的独特区域可以代表由AARS的新鉴定的蛋白片段具有的独特三维结构。

[0367] 某些实施方案可以利用“阵列”,例如“微阵列”。在某些实施方案中,“微阵列”还可以指具有与基底结合的多肽集合或多个多肽的“肽微阵列”或“蛋白微阵列”,与所述多个结合的多肽的结合可以被单独地检测。可选地,肽微阵列可以具有多种结合剂,包括,但不限于,可以特异性检测本文所述的AARS多肽的结合的单克隆抗体、多克隆抗体、噬菌体展示结合剂,酵母双杂交结合剂和适体。阵列可以基于这些AARS多肽的自身抗体检测,例如,如在Robinson等人,Nature Medicine 8(3):295-301(2002)中所描述。肽阵列的实例可见于W002/31463、W002/25288、W001/94946、W001/88162、W001/68671、W001/57259、W000/61806、W000/54046、W000/47774、W099/40434、W099/39210和W097/42507以及美国专利第6,268,210号、第5,766,960号和第5,143,854号,通过引用将每篇并入本文。

[0368] 某些实施方案可以利用MS或其他基于分子量的方法用于诊断性地检测AARS多肽序列。质谱法(MS)通常是指用于确定样品或分子的元素组成的分析技术。MS还可用于确定诸如肽和其他化合物的分子的化学结构。

[0369] 通常,MS原理包括将化合物电离以产生带电的分子或分子片段,然后测量它们的质荷比。在示例性的MS方案中:将样品加载到MS装置上,并经过汽化,通过多种方法中的一种将样品的组分电离(例如,通过用电子束轰击它们),这导致形成带正电的粒子,然后通过磁场加速阳离子,当所述离子穿过电磁场时根据它们的运动轨迹对粒子的质荷比( $m/z$ )进行计算,并且根据 $m/z$ 对之前的步骤中检测的离子进行分类。

[0370] 示例性的MS装置具有三个模块:离子源,其将气相的样品分子转化成离子(或在电喷雾电离的情况下,将溶液中存在的离子移动到气相中);质量分析器,其使用电磁场通过离子的质量将其分类;和检测器,其测量指示剂量的值,并因而提供数据用于计算存在的每种离子的丰度。

[0371] MS技术同时具有定性和定量用途,包括鉴定未知化合物,确定分子中元素的同位素组成,和通过观察化合物的断裂确定化合物的结构。其他用途包括对样品中化合物的量进行定量或研究气相离子化学的原理(真空中离子和中性粒子化学)。包括气相色谱-质谱(GC/MS或GC-MS)、液相色谱质谱(LC/MS或LC-MS)和离子迁移光谱/质谱(IMS/MS或IMMS)。因此,根据本文所提供的任何方法可以将MS技术用于测量生物样品中本发明的AARS多肽的存在或水平,并将这些水平与对照样品或预定值进行比较。

[0372] 某些实施方案可以采用细胞分选或细胞可视化或成像设备/技术来检测或定量AARS多核苷酸或多肽的存在或水平。实例包括流式细胞术或FACS、免疫荧光分析(IFA)和原位杂交技术,例如荧光原位杂交(FISH)。

[0373] 某些实施方案可以采用常规的为诊断目的的生物学方法、软件和系统。本发明的计算机软件产品通常包括具有计算机可执行指令的计算机可读介质,所述计算机可执行指

令用于进行本发明方法的逻辑步骤。适合的计算机可读介质包括软盘、CD-ROM/DVD/DVD-ROM、硬盘驱动、闪存、ROM/RAM、磁带等。计算机可执行指令可以适合的计算机语言或几种语言的组合来编写。基本的计算机生物学方法描述于例如Setubal和Meidanis等人, *Introduction to Computational Biology Methods* (PWS Publishing Company, Boston, 1997); Salzberg, Searles, Kasif, (编辑.), *Computational Methods in Molecular Biology*, (Elsevier, Amsterdam, 1998); Rashidi和Buehler, *Bioinformatics Basics: Application in Biological Science and Medicine* (CRC Press, London, 2000), 以及 Ouelette和Bzevanis *Bioinformatics: A Practical Guide for Analysis of Gene and proteins* (Wiley&Sons, Inc., 第2版, 2001)。参见美国专利号6,420,108。

[0374] 某些实施方案可以采用多种用途的各种计算机程序产品和软件, 例如探针设计、数据管理、分析和仪器操作。参见, 美国专利号5,593,839、5,795,716、5,733,729、5,974,164、6,066,454、6,090,555、6,185,561、6,188,783、6,223,127、6,229,911和6,308,170。

[0375] 全基因组采样测定(WGSA)描述于例如Kennedy等人, *Nat. Biotech.* 21, 1233-1237 (2003), Matsuzaki等人, *Gen. Res.* 14:414-425, (2004) 和Matsuzaki, 等人, *Nature Methods* 1:109-111 (2004)。用于作图测定的算法描述于例如Liu等人, *Bioinformatics.* 19: 2397-2403 (2003) 和Di等人 *Bioinformatics.* 21:1958 (2005)。与WGSA相关的其他方法和用于WGSA的测定以及WGSA的应用公开于例如美国专利申请号60/676,058 (2005年4月29日提交)、60/616,273 (2004年10月5日提交)、10/912,445、11/044,831、10/442,021、10/650,332和10/463,991。使用作图测定的全基因组关联研究描述于例如Hu等人, *Cancer Res.*; 65 (7):2542-6 (2005), Mitra等人, *Cancer Res.*, 64 (21):8116-25 (2004), Butcher等人, *Hum Mol Genet.*, 14 (10):1315-25 (2005), 和Klein等人, *Science.* 308 (5720):385-9 (2005)。

[0376] 此外, 某些实施方案可以包括用于在网络例如因特网上提供遗传信息的方法, 如例如以下中显示的: 美国申请号10/197,621、10/063,559 (美国公布号2002/0183936)、10/065,856、10/065,868、10/328,818、10/328,872、10/423,403和60/482,389。

#### [0377] X. 反义剂和RNAi剂

[0378] 本发明实施方案还包括靶向AARS多核苷酸序列的反义寡核苷酸和RNAi剂, 以及使用它们减少选定AARS转录物和/或蛋白片段表达的方法。某些实施方案涉及靶向一个或多个剪接点(通常是独特的), 产生剪接变体, 即本发明的AARS蛋白片段。还包括靶向某些剪接形式的反义或RNAi抑制的方法, 以鼓励或阻碍选定蛋白片段的剪接。在某些优选实施方案中, 产生AARS蛋白片段的剪接点在特定组织过表达, 并且对于该剪接变体是独特的。在这些和相关实施方案中, 这样的剪接变体不是靶向细胞类型中胞质AARS活性的唯一来源。例如, 待靶向的某些剪接变体可以代表给定细胞或组织中AARS RNA剪接变体总拷贝数的约10%至50%, 并且优选代表给定细胞或组织中AARS RNA剪接变体总拷贝数的约1-10%。还可以靶向占给定细胞或组织中AARS RNA剪接变体总拷贝数的约<1%的剪接变体。

[0379] 在某些实施方案中, 反义剂或RNAi剂不靶向全长蛋白, 因为此类全长蛋白负责蛋白合成的关键步骤, 并且从而避免经常源于野生型AARS敲除的致死性。因此, 本文描述的某些方法可用于避免在慢性和急性治疗中不希望效应, 例如毒性, 并且选择性调节AARS蛋白片段的非常规活性。然而, 某些实施方案可以在类属上靶向AARS序列, 包括全长AARS序列, 例如杀伤或基本上扰乱靶细胞或组织的细胞生理学。

[0380] 在某些实施方案中,待靶向的AARS剪接变体具有非常规生物活性。在某些实施方案中,AARS剪接变体具有减少的或不可检测的常规AARS活性,并且反义或RNAi相关方法更具体地调节其非常规活性。在某些实施方案中,反义或RNAi相关剂可以与靶向或局部递送方法联合以减少对非靶向细胞或组织的不希望的系统效应。除了本文描述的其他细胞或组织以外,可以这种方式被靶向的示例性细胞或组织包括癌细胞以及经由局部应用使它们本身被定位靶向的细胞至组织,例如肿瘤或上皮细胞。

[0381] 反义剂

[0382] 术语“反义寡聚体”或“反义化合物”或“反义寡核苷酸”可互换使用,并且指各自携带碱基配对部分的环状亚单元序列,其通过亚单元间的键合连接,所述亚单元间的键合允许碱基配对部分通过Watson-Crick碱基配对与核酸(通常是RNA)中靶序列杂交,从而在靶序列中形成核酸:寡聚体异源双链体,并通常以此防止该RNA的翻译。还包括使用它来调节选择的AARS转录物和/或其相应的多肽的表达的方法,所述选择的AARS转录物例如剪接变体或蛋白水解片段。

[0383] 反义寡核苷酸可以包含约8至40个亚单元,通常约8-25个亚单元,并优选约12至25个亚单元。在某些实施方案中,寡核苷酸可以具有与靶序列的精确序列互补性或接近的互补性,如下文所定义。在某些实施方案中,靶和反义靶向序列之间互补性程度足以形成稳定的双链体。反义寡聚体与靶RNA序列之间互补的区域可以短至8-11个碱基,但优选12-15个碱基或更多,例如12-20个碱基,或12-25个碱基,包括这些范围之间的所有整数。约14-15个碱基的反义寡聚体通常是足够长的,从而使之在靶向选择的AARS转录物时具有独特的互补序列。在某些实施方案中,互补碱基的最小长度可以是实现如上讨论的必需的结合 $T_m$ 所需要的。

[0384] 在某些实施方案中,长达40个碱基的反义寡聚体可以是合适的,其中至少有例如10-12个碱基的最少碱基数与靶序列互补。然而,通常在小于约30的寡聚体长度时,细胞容易摄取或主动摄取是最佳的。对于下文进一步描述的某些寡聚体,结合稳定性和摄取的最佳平衡通常发生在18-25个碱基的长度。包括由约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个碱基组成的反义寡聚体(例如PNA、LNA、2'-OMe、MOE),其中至少约6、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39或40个连续或非连续碱基与它们的AARS靶序列或其变体互补。

[0385] 在某些实施方案中,反义寡聚体可以与AARS核酸靶序列100%互补,或其可以包含错配,例如,以适应变体,只要在寡聚体和AARS核酸靶序列之间形成的异源双链体足以稳定地耐受细胞核酸酶的作用和体内可能发生的其它降解模式。术语“靶序列”指寡核苷酸所针对的靶RNA的一部分,即通过互补序列的Watson-Crick碱基配对与寡核苷酸杂交的序列。在某些实施方案中,靶序列可以是AARS mRNA的连续区域(例如,AARS mRNA的独特剪接点),或可以由该mRNA的非连续区域组成。

[0386] 下文讨论了不易受核酸酶切割的寡聚体主链。如果存在错配,相对于杂交体双链体的中间区域,其对末端区域产生的去稳定作用较小。根据双链体稳定性的公知原理,允许错配的数目将取决于寡聚体的长度,双链体中G:C碱基对的百分比,和双链体中错配的位置。尽管这样的反义寡聚体不一定与AARS核酸靶序列100%互补,但是其对于与靶序列稳定

且特异性地结合,从而调节核酸靶标的生物活性,例如AARS蛋白的表达是有效的。

[0387] 寡聚体和靶序列之间形成的双链体的稳定性是结合 $T_m$ 和双链体对细胞酶促切割的敏感性的函数。反义寡核苷酸与互补序列RNA的 $T_m$ 可以通过常规方法测量,例如描述于Hames等人,Nucleic acid Hybridization, IRL Press, 1985, pp.107-108或Miyada C.G.和Wallace R.B., 1987, Oligonucleotide hybridization techniques, Methods Enzymol. 第154卷pp.94-107中的那些方法。在某些实施方案中,反义寡聚体与互补序列RNA之间的结合 $T_m$ 可以高于体温并优选高于 $50^{\circ}\text{C}$ 。 $60-80^{\circ}\text{C}$ 范围内或更高的 $T_m$ 是优选的。根据公知原理,可以通过增加双链体中C:G配对碱基的比例和/或通过增加异源双链体的(碱基对)长度来增加寡聚体化合物对于互补碱基的RNA杂交体的 $T_m$ 。同时,为了优化细胞摄取的目的,限制反义寡聚体大小可能是有利的。为此原因,在25个碱基或更小长度显示高 $T_m$ ( $50^{\circ}\text{C}$ 或更大)的化合物一般比需要多于25个碱基以获得高 $T_m$ 值的那些化合物更优选。

[0388] 可以设计反义寡聚体用于阻断或抑制mRNA的翻译或用于抑制天然前mRNA剪接加工或诱导靶mRNA的降解,并可以称该反义寡聚体“针对”或“靶向”其杂交的靶序列。在某些实施方案中,靶序列可以包含AARS mRNA转录物的任何编码或非编码序列,并因而可以在外显子中或在内含子中。在某些实施方案中,靶序列在AARS(例如,全长AARS)中是相对独特或异常的,并被选择用于降低选择的AARS蛋白水解片段或剪接变体的表达。在某些实施方案中,靶位点包含加工前mRNA的3'或5'剪接位点或分支点。剪接位点的靶序列可以包含这样的mRNA序列,该mRNA序列在其5'端具有加工前mRNA中剪接受体点下游或剪接供体点上游的1至约25至约50个碱基对。在某些实施方案中,靶序列可以包括选择性剪接AARS mRNA的剪接点,例如在全长AARS中不存在或者对于该转录物独特或异常的剪接点,因为它在其他AARS剪接变体中不存在或很少存在。当寡聚体以本文所述的方式靶向靶核酸时,该寡聚体更常被称为“靶向”生物学相关的靶,例如参照AARS多核苷酸。

[0389] 寡核苷酸通常与诸如靶DNA或RNA的靶序列互补。术语“互补的”和“互补性”是指通过碱基配对原则而相关的多核苷酸(即,核苷酸序列)。例如,序列“A-G-T”与序列“T-C-A”互补。互补性可以是“部分地”,其中只有部分核酸碱基根据碱基配对原则匹配。或者,核酸之间可以是“完全”或“全部”的互补性(100%)。核酸链间的互补性程度对核酸链间的杂交的效率和强度具有重要的影响。尽管通常期望完美的互补性,但是某些实施方案可以包括与靶序列的一个或多个,但优选20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3、2或1个错配。包括寡聚体中任意位置处的变异。在某些实施方案中,相对于内部的变异,序列中接近寡聚体末端的变异通常是更优选的,并且如果存在,通常在5'和/或3'末端的约10、9、8、7、6、5、4、3、2、或1个核苷酸内。

[0390] 术语“靶向序列”或某些实施方案中的“反义靶向序列”是指寡核苷酸中与DNA或RNA靶分子中的靶序列互补(还表示基本上互补)的序列。反义化合物的全部序列或仅一部分可以与靶序列互补。例如,在具有20-30个碱基的寡核苷酸中,约6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28或29个可以是与靶区域互补的靶向序列。通常,靶向序列由连续碱基形成,但是可选地,靶向序列可以由非连续序列形成,当将所述非连续序列例如从寡核苷酸的相反末端放置在一起时构成跨越靶序列的序列。

[0391] 当以反向平行构型发生杂交时,靶序列和靶向序列被描述为彼此“互补”。靶向序列可以具有与靶序列“近似的”或“大体的”互补性,并仍为了本发明的目的起作用,即其仍

可以是功能“互补的”。在某些实施方案中,在10个核苷酸中,寡核苷酸可以与靶序列具有至多一个错配,并且优选地在20个核苷酸中具有至多一个错配。或者,寡核苷酸可以与本文描述的AARS参照多核苷酸序列或其互补物具有至少约80%、85%、90%序列同源性和优选至少95%序列同源性。

[0392] 如果寡聚体在生理条件下( $T_m$ 基本上高于 $45^\circ\text{C}$ ,优选至少 $50^\circ\text{C}$ 并且通常为 $60^\circ\text{C}$ - $80^\circ\text{C}$ 或更高)与靶(例如,AARS参照多核苷酸或其互补物)杂交,则寡核苷酸与靶多核苷酸序列“特异性杂交”。这类杂交优选符合严格杂交条件。在给定的离子强度和pH下, $T_m$ 是50%的靶序列与互补多核苷酸杂交时的温度。同样,在反义寡聚体与靶序列具有“近似的”或“大体的”互补性,以及具有精确的互补性的条件下,可以发生这类杂交。

[0393] 术语特异性结合或特异性杂交通常是指在选择的杂交条件下寡核苷酸探针或多核苷酸序列不仅与样品中其预期的靶基因序列结合,并且不与样品中其他靶序列显著结合,从而在靶池中区分其预期的靶和所有其他靶。如本文所述,与其预期靶序列特异性杂交的探针还可以在选择的杂交条件下检测浓度差异。

[0394] “核酸酶抗性”低聚分子(寡聚体)指其主链以非杂交或杂交的形式基本上抗核酸酶切割的寡聚体;所述核酸酶是体内常见的细胞外和细胞内核酸酶;即,寡聚体在寡聚体所暴露的体内正常核酸酶条件下显示很少或没有核酸酶切割。

[0395] “异源双链体”指一个寡核苷酸和靶多核苷酸(例如靶DNA或RNA)的互补部分之间的双链体。“核酸酶抗性异源双链体”指通过寡聚体与其互补靶结合而形成的异源双链体,使得异源双链体基本上抗细胞内和细胞外核酸酶的体内降解,所述核酸酶例如RNA酶H,能够切割双链RNA/RNA或RNA/DNA复合物。

[0396] 寡核苷酸的“亚单元”是指一个核苷酸(或核苷酸类似物)单元。该术语可以指有或不具有连接的亚单元间的键的核苷酸,但是当涉及“带电的亚单元”时,电荷通常位于亚单元间的键(例如,磷酸酯或硫代磷酸酯键或阳离子键)当中。

[0397] 寡核苷酸的环状亚单元可以基于核糖或另一戊糖,或在某些实施方案中,基于替代的或修饰的基团。修饰的寡核苷酸主链的实例包括,但不限于,硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨烷基磷酸三酯、甲基和其他烷基磷酸酯(包括3'亚烷基磷酸酯和手性磷酸酯)、亚磷酸酯、氨基磷酸酯(包括3'氨基氨基磷酸酯和氨烷基氨基磷酸酯)、硫代氨基磷酸酯、硫代烷基磷酸酯、硫代烷基磷酸三酯和具有正常3'-5'键的硼烷磷酸酯,它们的2'-5'连接的类似物和具有反极性的那些,其中邻近的核苷单元对以3'-5'至5'-3'连接或以2'-5'至5'-2'连接。还考虑肽核酸(PNA)、锁核酸(LNA)、2'-O-甲基寡核苷酸(2'-Ome)、2'-甲氧基乙氧基寡核苷酸(MOE)、吗啉代、以及领域内已知的其他寡核苷酸。

[0398] 嘌呤或嘧啶碱基配对部分通常是腺嘌呤、胞嘧啶、鸟嘌呤、尿嘧啶、胸腺嘧啶或次黄嘌呤。还包括以下碱基,例如吡啶-4-酮、吡啶-2-酮、苯基、假尿嘧啶、2,4,6-三甲115氧基苯(2,4,6-trimethylthoxy benzene)、3-甲基尿嘧啶、二氢尿苷、蔡基、氨基苯基、5'烷基胞苷(例如,5-甲基胞苷)、5-烷基尿苷(例如,核糖胸苷)、5-卤代尿苷(例如,5-溴代尿苷)或6-氮杂嘧啶或6-烷基嘧啶(例如,6-甲基尿苷)、丙炔、癸苷、2-硫脲核苷、4-硫脲核苷、怀丁苷、怀丁氧苷(wybutoxosine)、4-乙酰胞苷、5-(羧基羟甲基)尿苷、5'-羧基甲基氨基甲基-2-硫脲核苷、5-羧基甲基氨基甲基尿苷、 $\beta$ -D-半乳糖癸苷、1-甲基腺苷、1-甲基肌酐、2,2-二甲基鸟苷、3-甲基胞苷、2-甲基腺苷、2-甲基鸟苷、N6-甲基腺苷、7-甲基鸟苷、5-甲氧基氨基甲



基-2-硫脲核苷、5-甲基氨基甲基尿苷、5-甲基羰基甲基尿苷(5-methylcarbonylnethyluridine)、5-甲氧基尿苷、5-甲基-2-硫脲核苷、2-甲硫基-N6-异戊烯基腺苷、 $\beta$ -D-甘露糖苷、尿苷-5-羟丙酸、2-巯基胞苷、苏氨酸衍生物和其他(Burgin et al., 1996, *Biochemistry*, 35, 14090; Uhlman & Peyman, 同上)。在这方面,“修饰的碱基”是指上文所列举的除了腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)、胸腺嘧啶(T)和尿嘧啶(U)之外的核苷酸碱基;这类碱基可以在反义分子中的任何位置使用。本领域技术人员应当理解,根据寡聚体的用途,T和U是可互换的。例如,与更像RNA的诸如2'-O-甲基反义寡核苷酸的其他反义化学相比,T碱基可以显示为U。

[0399] 如上文所述,本文提供的某些寡核苷酸包含肽核酸(PNA)。肽核酸(PNA)是其中骨架与脱氧核糖骨架基本上同形态的DNA类似物,由嘧啶或嘌呤碱基与之连接的N-(2-氨基乙基)甘氨酸单元组成。含有天然嘧啶和嘌呤碱基的PNA与互补寡核苷酸杂交,遵从Watson-Crick碱基配对原则,并且在碱基对识别方面模拟DNA(Egholm, Buchardt等人1993)。PNA的骨架由肽键而非磷酸二酯键形成,使得它们非常适合反义应用(参见以下结构)。骨架是不带电荷的,得到表现出比正常热稳定性更大的热稳定性的PNA/DNA或PNA/RNA双链体。PNA不被核酸酶或蛋白酶识别。

[0400] 可以使用本领域已知的任何技术合成产生PNA。PNA是其中聚酰胺骨架替代DNA的传统磷酸核糖环的DNA类似物。尽管与天然结构存在基本结构变化,但是PNA能够以螺旋形式序列特异性结合DNA或RNA。PNA的特性包括与互补DNA或RNA的高结合亲和力、由单碱基错配引起的去稳定效应、对核酸酶和蛋白酶的抗性、独立于盐浓度的与DNA或RNA的杂交、和与同型嘌呤DNA形成三链体。Panagene<sup>TM</sup>已经开发其专有的Bts PNA单体(Bts; 苯并噻唑-2-磺酰基)和专有的寡聚方法。使用Bts PNA单体的PNA寡聚由脱保护、偶联和加帽的反复循环组成。Panagene对该技术的专利包括US 6969766、US 7211668、US 7022851、US 7125994、US 7145006和US 7179896。教导PNA化合物的制备的代表性的美国专利包括但不限于美国专利号5,539,082; 5,714,331; 和5,719,262,其各自在此通过引用并入。PNA化合物的进一步教导可以参见Nielsen等人, *Science*, 1991, 254, 1497。

[0401] 还包括“锁核酸”亚单元(LNA)。LNA的结构在领域内是已知的,例如,Wengel, 等人, *Chemical Communications* (1998) 455; *Tetrahedron* (1998) 54, 3607, 和 *Accounts of Chem. Research* (1999) 32, 301; Obika, 等人, *Tetrahedron Letters* (1997) 38, 8735; (1998) 39, 5401, 和 *Bioorganic Medicinal Chemistry* (2008) 16, 9230。

[0402] 寡核苷酸可以掺入一个或多个LNA; 在一些情况下,化合物可以完全由LNA构成。用于合成个体LNA核苷亚单元以及将它们掺入寡核苷酸的方法是本领域已知的: 美国专利7,572,582; 7,569,575; 7,084,125; 7,060,809; 7,053,207; 7,034,133; 6,794,499; 和6,670,461。典型的亚单元间连接体包括磷酸二酯和硫代磷酸酯部分; 或者,可以使用不含磷的连接体。一个优选的实施方式是含有LNA的化合物,其中每个LNA亚单元被DNA亚单元(即,脱氧核糖核苷酸)分开。其他优选的化合物由交替的LNA和DNA亚单元构成,其中亚单元间连接体是硫代磷酸酯。

[0403] 某些寡核苷酸可以包含通过不带电的或基本上不带电的键连接的具有碱基配对部分的基于吗啉代的亚单元。术语“吗啉代寡聚体”或“PMO”(氨基磷酸酯或磷酸二胺吗啉代寡聚体)是指由吗啉代亚单元结构组成的寡核苷酸类似物,其中(i) 该结构通过含有磷的键

连接在一起,所述含有磷的键为1至3个原子长,优选2个原子长,并且优选是不带电的或是阳离子,将一个亚单元的吗啉代氮与邻近亚单元的5'环外碳连接,并且(ii)每个吗啉代环具有通过碱基特异氢键而与多核苷酸中的碱基有效结合的嘌呤或嘧啶或等效的碱基配对部分。

[0404] 可以对该键合做出改变,只要它们不干扰结合或活性。例如,与磷连接的氧可以被硫取代(硫代二氨基磷酸酯)。可以用氨基或低级烷基取代的氨基取代5'氧。与磷连接的侧链氮可以是未取代的、用(任选取代的)低级烷基单取代或二取代的。嘌呤或嘧啶碱基配对部分通常是腺嘌呤、胞嘧啶、鸟嘌呤、尿嘧啶、胸腺嘧啶或肌苷。吗啉代寡聚体的合成、结构和结合特性详述于美国专利号5,698,685、5,217,866、5,142,047、5,034,506、5,166,315、5,521,063和5,506,337以及PCT申请号PCT/US07/11435(阳离子键合)和US08/012804(改进的合成),其全部通过引用并入本文。

[0405] 如下文进一步描述的,还可以通过不基于磷的亚单位间键合来连接吗啉代亚单位,其中至少一个键合被上述侧链阳离子基团修饰。可以使用其未修饰状态下无电荷但是也可携带侧链胺取代基的其他寡核苷酸类似键合。例如,吗啉环上的5'氮原子可以在硫酰胺键合或尿素键合(其中磷分别被碳或硫替代)中使用,并且以与上述结构(b3)中的5'-氮原子类似的方式修饰。

[0406] 某些实施方案包括基本上不带电荷的吗啉代寡聚体,例如基本上不带电荷的二氨基磷酸酯连接的吗啉代寡聚体。寡核苷酸类似物中基本上不带电荷的含磷骨架是这样的骨架,其中大部分亚单元键合,例如其键合的50-100%、通常至少60%至100%或75%或80%在生理pH下不带电荷,并且含有单个磷原子。具有含磷骨架键合的吗啉代寡核苷酸的实例包括氨基磷酸酯和二氨基磷酸酯连接的吗啉代寡核苷酸。某些实施方案可以含有带正电荷的基团,优选其骨架键合的约10%-50%。

[0407] 基于吗啉代的亚单元的性质包括,例如,通过稳定的不带电荷或带正电荷的骨架键合以寡聚体形式连接的能力,支持核苷酸碱基(例如,腺嘌呤、胞嘧啶、鸟嘌呤、胸苷、尿嘧啶和次黄嘌呤)而使得形成的聚合物可以与 $T_m$ 值高于约45°C的相对短的寡核苷酸(例如10-15个碱基)中的互补碱基靶核酸(包括靶RNA)杂交的能力,寡核苷酸主动或被动转运入哺乳动物细胞的能力,和反义寡核苷酸:RNA异源双链体分别抗RNA酶和RNA酶H降解的能力。

[0408] 在某些实施方案中,基本上不带电荷的寡核苷酸可以被修饰为包括带电荷的键合,例如约1个带电荷的键合/2-5个不带电荷的键合,例如约4-5个带电荷的键合/10个不带电荷的键合。在某些实施方案中,当约25%的骨架键合是阳离子时,可以看到反义活性的最佳提高。在某些实施方案中,使用小量(例如10-20%)阳离子键合,或者当阳离子键合数目在范围50-80%内、例如约60%时,可以看到增强。在某些实施方案中,可以通过将大量电荷分布在反义寡核苷酸的“中心区域”骨架键合附近来进一步增强阳离子骨架电荷,例如,在具有8个阳离子骨架键合的20mer寡核苷酸中,使至少70%的这些带电键合位于10个最中心键合。

[0409] 靶向AARS多核苷酸参照序列的一个或多个部分的寡核苷酸或其互补物可用于本文描述和本领域技术人员已知的任何治疗、诊断或药物筛选方法中。

[0410] B. RNA干扰剂

[0411] 某些实施方案涉及RNA干扰(RNAi)剂,所述RNA干扰剂靶向氨酰-tRNA合成酶

(AARS) 参照多核苷酸的一种或多种mRNA转录物,包括其片段和剪接变体。还包括使用该试剂调节选择的AARS转录物的水平的方法,所述AARS转录物例如AARS剪接变体或内源性蛋白水解片段。

[0412] 术语“双链的”是指包含其中链的至少一部分足以互补成氢键并形成双链体结构的两个单独的核酸链。术语“双链体”或“双链体结构”是指双链分子的区域,其中两个单独的链基本上互补并因此彼此杂交。“dsRNA”是指具有双链体结构的核糖核酸分子,所述双链体结构包含两条互补的和反向平行的核酸链(即有义链和反义链)。不是所有的dsRNA的核苷酸必须展现Watson-Crick碱基对,两条RNA链可以是基本上互补的。RNA链可以具有相同或不同数目的核苷酸。

[0413] 在某些实施方案中,dsRNA是或包括其至少部分互补于靶RNA的区域。在某些实施方案中,dsRNA完全互补于靶RNA。dsRNA与靶之间不一定存在完美互补,但是对应性必须足以使dsRNA或其裂解产物指导序列特异性沉默,例如通过对靶RNA的RNAi裂解。与靶链的互补性或同源性程度通常在反义链中是最关键的。虽然特别在反义链中,完美的互补性通常是希望的,但一些实施方案可以包括与靶RNA相比一个或多个但优选6、5、4、3、2或更少的错配。错配在末端区域是最耐受的,并且如果存在,则优选在末端区域,例如在5'端和/或3'端的6、5、4或3个核苷酸内。有义链仅需要与反义链基本互补以维持分子的整体双链特征。

[0414] 如本文使用的,“修饰的dsRNA”指包含至少一个改变的dsRNA分子,所述改变使它比识别相同靶RNA的相同dsRNA分子对更能抵抗核酸酶(例如蛋白激酶)。修饰的dsRNA可以包括单链核苷酸突出端和/或至少一个取代的核苷酸。

[0415] 如本文使用的,“核苷酸突出端”指当RNA链的3'端延伸超过另一互补链的5'端或者反过来时,从双链体结构突出的未配对核苷酸。“平的”或“平端”表示在dsRNA末端没有未配对的核苷酸,即没有核苷酸突出端。“平端”dsRNA是在其整个长度为双链的dsRNA,即,在分子的任一端没有核苷酸突出端。

[0416] 如本文使用的,术语“末端碱基对”指在双链分子的双链体区域的一端的最后一个核苷酸碱基对。例如,如果dsRNA或其他分子是平端(即,没有核苷酸突出端),则分子两端的最后核苷酸碱基对是末端碱基对。如果dsRNA或其他分子在双链体结构的一端或两端具有核苷酸突出端,则紧邻核苷酸突出端的最后一个核苷酸碱基对是该分子该末端的末端碱基对。

[0417] 在某些实施方案中,本文提供的方法可以利用双链核糖核酸(dsRNA)分子作为调节剂,用于减少AARS转录物例如选定的片段或剪接变体的表达。dsRNA一般包括两条单链。dsRNA的一条链包含与靶基因或靶区域(the“有义”链)的一部分基本同一的核苷酸序列,另一条链(“互补”或“反义”链)包含与靶基因的一部分基本互补的序列。这两条链足够互补以杂交形成双链体结构。在某些实施方案中,互补的RNA链可以小于30个核苷酸,长度小于25个核苷酸,或甚至长度为19至24个核苷酸。在某些方面,互补的核苷酸序列的长度可以为20-23个核苷酸或长度为22个核苷酸。

[0418] 在某些实施方案中,RNA链的至少一条包含长度为1至4个核苷酸的核苷酸突出端。在其他实施方案中,dsRNA还可以包含至少一个化学修饰的核苷酸。在某些方面,包含1至4个核苷酸的单链突出端的dsRNA可以包括其中与末端核苷酸对直接相邻的单链突出端的未配对核苷酸含有嘌呤碱基的分子。在其他方面,dsRNA的两端上的最后一个互补核苷酸对事

G-C对,或者后四个末端核苷酸对的至少两个是G-C对。

[0419] 本发明的某些实施方案可以包含微小RNA。微小RNA代表生物中天然产生的一大类小RNA,其中某些调节靶基因的表达。通过切酶从约70个核苷酸的单链发夹前体转录物形成微小RNA。(V.Ambros等人Current Biology13:807,2003)。某些微小RNA可以被转录为发夹RNA前体,然后通过切酶加工成其成熟形式。

[0420] 某些实施方案还可以利用短干扰RNA(siRNA)。在某些实施方案中,双链寡核苷酸的第一链比第二链多出两个核苷残基。在其他实施方案中,第一链和第二链具有相同数目的核苷;然而,第一链和第二链可以抵消,使得第一链和第二链上的两个末端核苷不与互补链上的残基配对。在某些情况下,两个不配对的核苷是胸苷残基。

[0421] 还包括短发夹RNA(shRNA)和微小RNA(miRNA)。shRNA的双链结构由单条自互补的RNA链形成,RNA双链体的形成可以开始于细胞的内部或外部。微小RNA(miRNA)是20-22个核苷酸的小的非编码RNA,通常是从称为前miRNA的~70核苷酸回折RNA前体结构切离的。

[0422] 在调节剂包含siRNA的情况下,该剂应该包括与靶区域足够同源的区域,并且在核苷酸方面具有足够长度,使得siRNA剂或其片段可以介导靶RNA的下调。将理解,在修饰的RNA或核苷酸替代物的情况下,术语“核糖核苷酸”或“核苷酸”还可以指在一个或多个位置的修饰核苷酸或替代物替代部分。因此,siRNA剂是或包括与本文所述的靶RNA至少部分互补的区域。

[0423] 此外,siRNA调节剂可以被修饰或者包括核苷替代物。siRNA剂的单链区域可以被修饰或者包括核苷替代物,例如发夹结构的一个或多个未配对区域,例如连接两个互补区的区域可以具有修饰或核苷替代物。稳定siRNA剂(例如对抗内切核酸酶)的一个或多个3'端或5'端或者有利于反义siRNA剂进入RISC的修饰也是有用的。修饰可以包括C3(或C6、C7、C12)氨基连接体、硫醇连接体、羧基连接体、非核苷酸间隔区(C3、C6、C9、C12、无碱基的、三甘醇、六甘醇)、作为亚磷酰胺引入并且具有另外DMT保护的羟基的特别生物素或荧光素试剂,允许RNA合成期间的多重偶联。

[0424] 例如,siRNA剂可以包括足够长以引发干扰素反应的分子(其可以由切酶裂解(Bernstein等人2001.Nature,409:363-366)并进入RISC(RNAi诱导的沉默复合物)),以及足够短以使它们不能引发干扰素反应的分子(该分子也可由切酶裂解和/或进入RISC),例如具有允许进入RISC的大小的分子,例如类似于切酶裂解产物的分子。siRNA调节剂或其裂解产物可以下调靶基因,例如通过诱导与靶RNA、优选AARS靶(例如选定的剪接变体)相关的RNAi。

[0425] siRNA剂的每条链的长度可以等于或小于35、30、25、24、23、22、21、20、19、18、17、16或15个核苷酸。该链优选是长度至少19个核苷酸。例如,每条链可以是长度为21至25个核苷酸。优选的siRNA剂具有17、18、19、20、21、22、23、24或25核苷酸对的双链体区和一个或多个2-3个核苷酸的突出端,优选一个或两个3'突出端。

[0426] 除了与靶RNA的同源性和下调靶基因的能力,siRNA剂可以具有以下性质的一种或多种:尽管有甚至大量的或全部核苷的修饰,它可以具有可以适当的三维框架呈现碱基(或修饰碱基)的反义链,以能够与同源靶RNA形成正确的碱基配对并形成双链体结构,这足以使靶下调,例如通过裂解靶RNA;尽管有甚至大量的或全部核苷的修饰,它依然具有“RNA样”性质,即,它可以即使非排他地或甚至部分地具有基于脱氧核苷酸内容的RNA分子的整体结

构、化学和物理性质。例如，siRNA剂可以含有例如有义和/或反义链，其中所有核苷酸糖含有例如替代2' 羟基的2' 氟。依然可以预期该含有脱氧核苷酸的剂表现出RNA样性质。虽然不希望受理论束缚，但电负性的氟与核糖的C2' 位置连接时更喜欢轴向定位。氟的这种空间偏好反过来迫使糖采取C<sub>3'</sub>-内折叠。这是与RNA分子中观察到的相同的折叠模式，并且产生RNA特性的A家族类型螺旋。而且，因为氟是良好的氢键受体，它可以参与已知稳定RNA结构的与水分子相同的氢键相互作用。一般优选的是，2' 糖位置的修饰部分能够进入H键合，这是核糖核苷酸的OH部分比脱氧核糖核苷酸的H部分更特性的。

[0427] 如本文所用的“单链RNAi剂”是由单分子制成的RNAi剂。RNAi剂可以包含通过链内配对形成的双链体区，例如其可以是或可以包含发夹或盘-柄结构。单链RNAi调节剂优选地对于靶分子是反义的。单链RNAi剂的长度应该足以使它进入RISC并参与靶mRNA的RISC介导的裂解。单链RNAi剂的长度为至少14个核苷酸，并且更优选至少15、20、25、29、35、40或50个核苷酸。其长度优选小于200、100或60个核苷酸。

[0428] 发夹RNAi调节剂可以具有等于或至少12、18、19、29、21、22、23、24或25个核苷酸对的双链体区。双链体区的长度可以优选等于或小于200、100或50。双链体区的长度的某些范围为15-30、17-23、19-23和19-21个核苷酸。发夹可以具有单链突出端或末端非配对区，优选在3' 端，并优选在发夹的反义侧。在某些实施方案中，突出端的长度是2-3个核苷酸。

[0429] 根据本文提供的方法使用的某些调节剂可以包括RNAi寡核苷酸，例如嵌合寡核苷酸或“嵌合体”，其含有两个或多个化学上不同的区域，各自由至少一个单体单元（在寡核苷酸化合物的情况下即是核苷酸）构成。这些寡核苷酸通常含有至少一个区，该区中的寡核苷酸被修饰以赋予寡核苷酸增加的对核酸酶降解的抗性、增加的细胞摄取和/或增加的对靶核酸的结合亲和力。因此，当使用嵌合寡核苷酸时，与硫代磷酸酯寡聚脱氧核苷酸相比，常常可以使用较短的寡核苷酸获得相当的结果。如上所述，嵌合寡核苷酸可以形成为两个或多个寡核苷酸、修饰的寡核苷酸、寡核苷酸和/或寡核苷酸模拟物的复合结构。此类寡核苷酸还在本领域称为杂合体或gapmer。教导此类杂合结构的制备的代表性美国专利包括但不限于美国专利号5,013,830;5,149,797;5,220,007;5,256,775;5,366,878;5,403,711;5,491,133;5,565,350;5,623,065;5,652,355;5,652,356;5,700,922;和5,955,589，其每一个通过引用并入本文。在某些实施方案中，嵌合寡核苷酸是RNA-DNA、DNA-RNA、RNA-DNA-RNA、DNA-RNA-DNA或RNA-DNA-RNA-DNA，其中寡核苷酸长度为5至60个核苷酸。

[0430] 在本发明的一个方面，RNAi剂涉及包含至少一个与改变的或非天然核碱基连接的配体的寡核苷酸。许多化合物可用作改变的碱基。改变的碱基的结构的重要性程度在于改变的碱基不应该明显防止寡核苷酸与其靶例如mRNA的结合。在某些实施方案中，改变的碱基是二氟甲基基、硝基吡咯基、硝基咪唑基、硝基吡啶基、萘基、蒽基、吡啶基、喹啉基、茈基或本文描述的非天然核碱基的任何一种的二价自由基。在某些实施方案中，非天然核碱基是二氟甲基基、硝基吡咯基或硝基咪唑基。在某些实施方案中，非天然核碱基是二氟甲基基。许多配体是本领域已知的并且适合本发明。例如，配体可以是类固醇、胆汁酸、脂质、叶酸、吡哆醛、B12、核黄素、生物素、芳香族化合物、多环化合物、冠醚、嵌入剂、切割分子、蛋白结合剂或碳水化合物。在某些实施方案中，配体是类固醇或芳香族化合物。在某些情况下，配体是胆甾烯基。

[0431] 在其他实施方案中，RNAi剂是为改善细胞靶向和摄取的目的而与配体连接的寡核

苷酸。例如, RNAi 剂可以与抗体或其抗原结合片段连接。作为额外实例, RNAi 剂可以与特定的配体结合分子连接, 例如与特定细胞表面受体特异性结合的多肽或多肽片段。

[0432] 在其他实施方案中, 调节剂包含本文所述的非天然核碱基。在某些实施方案中, 核苷中天然存在的核糖部分被己糖替代。在某些方面中, 己糖是阿洛糖、阿卓糖、葡萄糖、甘露糖、古洛糖、艾杜糖、半乳糖、塔罗糖或其衍生物。在一个优选实施方案中, 己糖是D-己糖。在某些情况下, 核苷中天然存在核糖部分被多环杂烷基环或环己烯基替代。在某些情况下, 多环杂烷基是在环中含有一个氧原子的双环。在某些情况下, 多环杂烷基是双环[2.2.1]庚烷、双环[3.2.1]辛烷或双环[3.3.1]壬烷。如本文所述, 修饰的RNAi剂的实例还包括含有修饰的骨架或非天然核苷内键合的寡核苷酸。

[0433] 本发明还包括采用合酶的寡核苷酸。催化高特异性内切核糖核酸酶的合成RNA分子及其衍生物被称为核酶(参见例如Haseloff等人的美国专利号5,543,508,和Goodchild等人的美国专利号5,545,729)。裂解反应由RNA分子自身催化。在天然存在的RNA分子中, 自催化裂解位点位于RNA二级结构的高度保守区内(Buzayan等人, Proc.Natl.Acad.Sci.U.S.A., 1986, 83, 8859; Forster等人, Cell, 1987, 50, 9)。天然存在的自催化RNA分子已经被修饰为产生可以高度特异性靶向特定细胞或致病RNA的核酶。因此, 核酶的一般用途与反义寡核苷酸相同(即, 调节特定基因的表达), 并且与寡核苷酸一样是具有明显的单链部分的核酸。

[0434] 在某些情况下, 用于本文提供的方法的RNAi剂或反义寡核苷酸可以被非配体基团修饰。已经将许多非配体分子与寡核苷酸轭合, 以增强寡核苷酸的活性、细胞分布、细胞靶向或细胞摄取, 并且进行此类轭合的程序是科学文献中可获得的。此类非配体部分包括脂质部分, 例如胆固醇(Letsinger等人, Proc.Natl.Acad.Sci.USA, 1989, 86:6553)、富含精氨酸的肽、胆酸(Manoharan等人, Bioorg.Med.Chem.Lett., 1994, 4:1053)、硫醚例如己基-5-三苯甲基硫醇(Manoharan等人, Ann.N.Y.Acad.Sci., 1992, 660:306; Manoharan等人, Bioorg.Med.Chem.Lett., 1993, 3:2765)、硫代胆固醇(Oberhauser等人, Nucl.Acids Res., 1992, 20:533)、脂肪族链例如十二烷二醇或十一烷基残基(Saison-Behmoaras等人, EMBO J., 1991, 10:111; Kabanov等人, FEBS Lett., 1990, 259:327; Svinarchuk等人, Biochimie, 1993, 75:49)、磷脂例如双十六烷基-rac-甘油或1,2-二-0-十六烷基-rac-丙三氧基-3-H-膦酸三乙基铵(Manoharan等人, Tetrahedron Lett., 1995, 36:3651; Shea等人, Nucl.Acids Res., 1990, 18:3777)、多胺或聚乙二醇链(Manoharan等人, Nucleosides&Nucleotides, 1995, 14:969)或金刚烷乙酸(Manoharan等人, Tetrahedron Lett., 1995, 36:3651)、棕榈酰部分(Mishra等人, Biochim.Biophys.Acta, 1995, 1264:229)、或十八胺或己基氨基-羰基-羟胆固醇部分(Crooke等人, J.Pharmacol.Exp.Ther., 1996, 277:923)。教导此类寡核苷酸轭合物的制备的代表性美国专利已经在上文列出。典型的轭合方案包括在序列的一个或多个位置携带氨基连接体的寡核苷酸的合成。然后使用适当的偶联或激活试剂, 使氨基与轭合的分子反应。可以用依然与固体支持体结合的寡核苷酸或者在溶液相中裂解寡核苷酸之后进行轭合反应。HPLC对寡核苷酸轭合物的纯化通常提供了纯的轭合物。

[0435] RNAi剂的其他实例可以参见美国申请公布号2007/0275465、2007/0054279、2006/0287260、2006/0035254、2006/0008822, 其在此通过引用并入。还包括能够表达本文所述AARS-靶向序列的载体递送系统。包括表达siRNA或其他双链体形成RNA干扰分子的载体。

[0436] 载体或核酸构建体系统可以包括单个载体或质粒、两个或多个载体或质粒,合在一起含有待引入宿主细胞基因组或转座子的总DNA。载体的选择通常取决于载体与待引入载体的宿主细胞的相容性。在本情况下,载体或核酸构建体优选是在哺乳动物细胞例如肌肉细胞中具有可操作功能的载体或核酸构建体。载体还可以包含选择性标记,例如抗生素或药物抗性基因,或者报告基因(即,绿色荧光蛋白、荧光素酶),其可用于选择或鉴定适合的转化子或转染子。示例性的递送系统可以包括病毒载体系统(即,病毒介导的转导),包括但不限于逆转录病毒(例如慢病毒)载体、腺病毒载体、腺相关病毒载体和疱疹病毒载体以及本领域已知的其他病毒载体。

[0437] XI. 药物发现

[0438] 某些实施方案涉及AARS多肽、抗体或多核苷酸在药物发现中的用途,通常涉及鉴定调节参照AARS多肽(例如,AARS蛋白片段)的一种或多种非常规活性的试剂。

[0439] 例如,某些实施方案包括鉴定AARS参照多肽的一种或多种“细胞结合配体”的方法,所述细胞结合配体例如与AARS多肽直接或物理上相互作用的细胞蛋白、脂质、核酸或其他宿主分子。具体实例包括例如细胞表面受体(例如GPCR)、蛋白-蛋白相互作用结构域及其细胞外或细胞内结构域。

[0440] 还包括鉴定参与AARS多肽(包括与细胞结合配体直接或间接相互作用的分子)的一种或多种非常规活性,并且调节其在非常规活性中作用或者受结合配体调节的宿主分子。此类宿主分子包括非常规途径的上游和下游组分,通常与该途径中约1、2、3、4、5或更多可鉴定的步骤相关,与细胞结合配体/AARS蛋白相互作用相关。

[0441] 某些方面包括鉴定激动或拮抗AARS参照多肽或其活性变体的非常规活性的化合物(例如,多肽)或其他剂的方法,例如通过与AARS多肽和/或其细胞结合配体的一个或多个相互作用。还包括鉴定调节AARS剪接变体的表达(例如剪接)或者调节蛋白酶活性的剂的方法,所述蛋白酶否则在蛋白水平上调节内源性AARS蛋白片段(切除蛋白)的生成。

[0442] 因此,某些实施方案包括鉴定AARS参照多肽的结合配体的方法,包括a) 在合适的条件下使AARS多肽与生物样品组合,和b) 检测AARS多肽与结合配体的特异性结合,从而鉴定与AARS参照多肽特异性结合的结合配体。还包括筛选与AARS参照多肽或AARS多肽的结合配体特异性结合的化合物的方法,所述方法包括a) 在合适的条件下使所述多肽或结合配体与至少一种测试化合物组合,和b) 检测所述多肽或结合配体与所述测试化合物的结合,从而鉴定与所述多肽或其结合配体特异性结合的化合物。在某些实施方案中,所述化合物是多肽或肽。在某些实施方案中,所述化合物是小分子或其他(例如,非生物的)化合物。在某些实施方案中,所述化合物是肽模拟物。

[0443] 可以利用适合检测蛋白-蛋白相互作用的任何方法来鉴定与AARS参照多肽相互作用,与其一种或多种细胞结合配体相互作用,或与以上二者相互作用的细胞蛋白。可以利用的常规方法的实例包括对细胞裂解物或从细胞裂解物获得的蛋白进行免疫共沉淀反应,交联和通过梯度或色谱柱共纯化,主要用于鉴定裂解物中与AARS多肽相互作用的蛋白。

[0444] 在这些和相关实施方案中,使用本领域技术人员公知的技术,例如通过Edman降解技术,可以确定与AARS多肽或其结合配体相互作用的蛋白的至少一部分氨基酸序列。参见例如Creighton proteins: Structures and Molecular Principles, W.H. Freeman & Co., N.Y., pp. 34-49, 1983。获得的氨基酸序列可以用作产生寡核苷酸混合物的引导,所述寡核

核苷酸混合物可以用于筛选编码这类蛋白的基因序列。例如,可以通过本文所述的和本领域内已知的标准杂交或PCR技术完成筛选。用于产生寡核苷酸混合物和进行筛选的技术是公知的。参见例如Ausubel等人Current Protocols in Molecular Biology Green Publishing Associates and Wiley Interscience,N.Y.,1989;和Innis等人编辑PCR Protocols:A Guide to Methods and Applications Academic Press,Inc.,New York,1990。

[0445] 此外,可以利用同时鉴定编码结合配体或其他多肽的基因的方法。这些方法包括,例如,以与 $\lambda$ -gt11文库抗体探测的公知技术相似的方式,利用标记的AARS蛋白或另一多肽、肽或融合蛋白,例如与标志物(例如,酶、荧光剂、发光蛋白或染料)融合的变体AARS多肽或AARS结构域或Ig-Fc结构域,探测表达文库。

[0446] 仅为示例说明而非限制地详细描述了检测体内蛋白相互作用的一种方法,双杂交系统。该系统的实例描述于(Chien等人,PNAS USA 88:9578-9582,1991)中并且可从Clontech(Palo Alto,Calif.)商购。

[0447] 简言之,利用这种系统,可以构建编码两个杂交蛋白的质粒:一个质粒由与AARS参考核苷酸序列(或者在某些实施方案中,其结合配体)或其变体融合的编码转录激活蛋白的DNA结合域的核苷酸组成,另一个质粒由与已经作为cDNA文库的部分重组入质粒的、编码未知蛋白的cDNA(或cDNA集合)融合的编码转录激活蛋白激活结构域的核苷酸组成。DNA结合结构域融合质粒和激活子cDNA文库可以被转化进入含有报告基因(例如,HBS或lacZ)的酿酒酵母菌株,所述报告基因的调节区含有转录激活子的结合位点。任何一个单独的杂交蛋白不能激活报告基因的转录:DNA结合结构域杂交体不能是因为它不提供激活功能,而激活结构域杂交体不能是因为它不能定位于激活子的结合位点。两个杂交蛋白的相互作用重新构成功能性激活蛋白并导致报告基因的表达,这通过针对报告基因产物的测定来检测。

[0448] 双杂交系统或其他这类方法可用于筛选与“诱饵”基因产物相互作用的蛋白激活结构域文库。作为实例而非限制,AARS参照多肽或变体可用作诱饵基因产物。AARS结合配体也可用作“诱饵”基因产物。总基因组或cDNA序列被融合到编码激活结构域的DNA上。将该文库和编码诱饵AARS基因产物与DNA结合结构域融合的杂交体的质粒共转化进酵母报告菌株中,从所得的转化体筛选表达报告基因的那些菌株。

[0449] 要从中检测与诱饵AARS基因产物相互作用的蛋白的细胞系的cDNA文库可以使用本领域常用方法进行。例如,可以将cDNA片段插入载体,使得它们与GAL4的转录激活结构域翻译融合。该文库可以与诱饵基因-GAL4融合质粒一起共转化进入酵母菌株,该酵母菌株含有通过含有GAL4激活序列的启动子驱动的lacZ基因。与诱饵基因产物相互作用的、与GAL4转录激活结构域融合的cDNA编码蛋白将重新构建活性GAL4蛋白并从而驱动HIS3基因的表达。表达HIS3的菌落可以通过它们在含有基于半固体琼脂、缺乏组氨酸的培养基的陪替式培养皿上的生长来检测。然后,可以使用本领域常用技术从这些菌株纯化cDNA并用于产生和分离诱饵AARS基因相互作用蛋白。

[0450] 还包括三杂交系统,其允许在酵母中检测RNA-蛋白相互作用。参见,例如Hook等人,RNA.11:227-233,2005。因此,这些和相关方法可用于鉴定AARS多肽的细胞结合配体,并且鉴定与AARS多肽、细胞结合配体或二者相互作用的其他蛋白或核酸。

[0451] 某些实施方案涉及作用组筛选方法的使用。具体实例包括基于蛋白结构域的筛选



(参见例如Boxem等人, *Cell*.134:534-545,2008;和Yu等人, *Science*.322:10-110,2008)。

[0452] 如上所述,一旦分离,可以鉴定结合配体,然后相应地将该结合配体与标准技术联合使用以鉴定与其相互作用的蛋白或其他化合物。因此,某些实施方案涉及筛选与AARS参照多肽的结合配体特异性结合的化合物的方法,所述方法包括a) 在合适的条件下使所述结合配体与至少一种测试化合物组合,和b) 检测所述结合配体与所述测试化合物的结合,从而鉴定与结合配体特异性结合的化合物。在某些实施方案中,所述测试化合物是多肽。在某些实施方案中,所述测试化合物是化合物,例如小分子化合物或肽模拟物。

[0453] 某些实施方案包括筛选调节AARS参照多肽的活性的化合物的方法,所述方法包括a) 在容许所述多肽具有活性的条件下使所述多肽与至少一种测试化合物组合,b) 评估存在测试化合物时所述多肽的活性,和c) 将存在测试化合物时所述多肽的活性与不存在测试化合物时所述多肽的活性进行比较,其中存在所述测试化合物时多肽活性的改变表示化合物调节所述多肽的活性。某些实施方案包括筛选调节AARS参照多肽的结合配体的活性的化合物的方法,所述方法包括a) 在容许所述结合配体具有活性的条件下使所述多肽与至少一种测试化合物组合,b) 评估存在测试化合物时所述结合配体的活性,和c) 将存在测试化合物时所述结合配体的活性与不存在测试化合物时所述结合配体的活性进行比较,其中存在所述测试化合物时结合配体活性的改变指示化合物调节所述结合配体的活性。通常,这些和相关实施方案包括评估与AARS多肽或其结合配体相关的选择的非常规活性。包括体外和体内条件,例如细胞培养条件。

[0454] 某些实施方案包括筛选有效地作为AARS参照多肽或其活性片段或变体的完全或部分激动剂的化合物的方法,所述方法包括a) 将包含所述多肽的样品暴露于化合物,和b) 通常通过测量AARS多肽的非常规活性的增加来检测所述样品中的激动剂活性。某些方法包括a) 将包含AARS多肽的结合配体的样品暴露于化合物,和b) 通常通过测量AARS多肽的选择的非常规活性的增加来检测所述样品中的激动剂活性。某些实施方案包括这样的组合物,该组合物包含通过所述方法鉴定的激动剂化合物以及药学上可接受的载体或赋形剂。

[0455] 还包括筛选有效地作为AARS参照多肽的完全或部分拮抗剂的化合物的方法,所述方法包括a) 将包含所述多肽的样品暴露于化合物,和b) 通常通过测量AARS多肽的非常规活性的降低来检测所述样品中的拮抗剂活性。某些方法包括a) 将包含AARS多肽的结合配体的样品暴露于化合物,和b) 通常通过测量AARS多肽的选择的非常规活性的降低来检测所述样品中的拮抗剂活性。某些实施方案包括这样的组合物,该组合物包含通过所述方法鉴定的拮抗剂化合物以及药学上可接受的载体或赋形剂。

[0456] 在某些实施方案中,可以设计体外系统以鉴定可以与AARS参照序列或其结合配体相互作用或调节AARS参照序列或其结合配体的化合物。通过这类系统所鉴定出的某些化合物可用于,例如,调节途径的活性和加工所述途径的组分本身。它们还可在筛选中用于鉴定破坏所述途径的组分之间的相互作用,或可以直接破坏这类相互作用的化合物。一个示例性的方法涉及在允许AARS多肽和测试化合物相互作用和结合的条件和足够的时间下制备它们的反应混合物,从而形成可以从反应混合物移出和/或检测的复合物。

[0457] 可以以多种方式进行体外筛选测定。例如,可以将AARS多肽、细胞结合配体或测试化合物锚定在固相上。在这些和相关实施方案中,在反应结束时可以在固相上捕获和检测所得的复合物。在这类方法的一个实例中,将AARS多肽和/或其结合配体锚定在固相表面

上,而未锚定的测试化合物可被直接地或间接地标记,使得可以检测固相表面上的组分对所述测试化合物的捕获。在其他实例中,将测试化合物锚定在固相表面上,而未锚定的AARS多肽和/或其结合配体是标记的或以某种方法可检测。在某些实施方案中,微量滴定板通常可用作固相。锚定的组分(或测试化合物)可以通过非共价或共价连接被固定。通过用蛋白溶液简单包被固相表面并干燥可以实现非共价连接。可选择地,可以使用对待固定的蛋白特异的固定的抗体,优选单克隆抗体,将所述蛋白锚定在固相表面上。所述表面可以预先制备并储存。

[0458] 为了进行示例性的测定,通常将未固定的组分添加到含有锚定的组分的包被表面。反应完成之后,在所形成的任何特异性复合物仍将固定在固相表面上的条件下去除(例如,通过冲洗)未反应的组分。可以以多种方式实现锚定在固相表面上的复合物的检测。例如,如果之前未固定的组分是预先标记的,则对固定在表面上的标记的检测表明形成了复合物。如果之前未固定的组分是未预先标记的,则可以使用间接标记来检测锚定在表面上的复合物,例如,使用对之前未固定的组分特异的标记抗体(随后可以用标记的抗Ig抗体直接标记或间接标记所述抗体)。

[0459] 可选择地,例如,可以利用表面等离子共振技术(SPR)和共振角的改变作为指标来确定是测试化合物的结合的存在或不存在,其中根据常规方法将AARS多肽或细胞结合配体固定在可商购的传感器芯片(例如,由Biacore<sup>TM</sup>制造)的表面上,将测试化合物与之接触,并且用特定波长的光从特定角度照射所述传感器芯片。还可以通过检测与测试化合物相应的峰的出现来测定测试化合物的结合,所述检测通过这样的方法,其中将AARS多肽或细胞结合配体固定在适用于质谱仪的蛋白芯片的表面上,将测试化合物与之接触,并且将诸如MALDI-MS、ESI-MS、FAB-MS等的电离方法与质谱仪(例如,双聚焦质谱仪、四级杆质谱仪、飞行时间质谱仪、傅里叶变换质谱仪、离子回旋质谱仪等)组合。

[0460] 在某些实施方案中,基于细胞的测定,基于膜囊的测定,或基于膜级分的测定可用于鉴定调节选择的AARS多肽在非常规途径中的相互作用的化合物。为此目的,可以使用表达AARS多肽和/或结合配体的细胞系、或表达包含这类蛋白的结构域或片段(或其组合)的融合蛋白的细胞系、或已经基因工程改造以表达这类蛋白或融合蛋白的细胞系(例如,COS细胞、CHO细胞、HEK293细胞、HeLa细胞等)。通过相对于对照或预定量监测非常规活性的变化(例如,统计学上的显著性变化)可以鉴定影响该活性的测试化合物。

[0461] 对于涉及反义剂和RNAi剂的实施方案,例如,还包括筛选有效地改变AARS参照多核苷酸的表达的化合物的方法,所述方法包括a)将包含AARS参照多核苷酸的样品暴露于诸如潜在的反义寡核苷酸的化合物,和b)检测AARS多核苷酸表达的改变。在某些非限制性的实例中,根据本领域常规技术,这些和相关实施方案可以用于基于细胞的测定或用于无细胞翻译的测定中。还包括通过这类方法所鉴定的反义剂和RNAi剂。

[0462] AARS蛋白片段的抗体还可用于筛选测定,例如鉴定与AARS特异性结合的剂,证实与AARS蛋白片段结合的剂的特异性或亲和力,或者鉴定该剂与AARS蛋白片段之间相互作用的位点。包括其中抗体用作该剂的竞争性抑制剂的测定。例如,以已知亲和力特异性结合AARS蛋白片段的抗体可用作选定剂的竞争性抑制剂,并且用于计算该剂对AARS蛋白片段的亲和力。而且,特异性结合AARS蛋白片段的已知表位或位点的一种或多种抗体可用作竞争性抑制剂以证实该剂是否结合该相同位点。其他变化形式对于本领域技术人员是明显的。

[0463] 还包括适用于高通量筛选 (HTS) 的任何上述方法, 或本领域已知的其他筛选方法。HTS通常使用自动化以运行针对候选化合物文库的测定的筛选, 例如, 测量本文所述的非常规活性的增加或降低的测定。

[0464] 本文提供的任何筛选方法可以利用由组合化学产生的小分子文库。化学和/或生物混合物例如真菌、细菌或海藻提取物的文库是本领域已知的, 并且可以用本发明的任何测定来筛选。用于合成分子文库的方法的实例可以参见: (Carell等人, 1994a; Carell等人, 1994b; Cho等人, 1993; DeWitt等人, 1993; Gallop等人, 1994; Zuckermann等人, 1994)。

[0465] 化合物文库可以提供于溶液中 (Houghten等人, 1992) 或珠上 (Lam等人, 1991) 或芯片上 (Fodor等人, 1993)、细菌、孢子 (Ladner等人, 美国专利号5, 223, 409, 1993)、质粒 (Cull等人, 1992) 或噬菌体上 (Cwirlla等人, 1990; Devlin等人, 1990; Felici等人, 1991; Ladner等人, 美国专利号5, 223, 409, 1993; Scott和Smith, 1990)。本发明的实施方案包括使用不同文库鉴定一种或多种AARS蛋白片段、其细胞结合配体和/或其相关非常规活性的小分子调节剂。用于本发明目的的文库包括但不限于 (1) 化学文库, (2) 天然产物文库, 和 (3) 由随机肽、寡核苷酸和/或有机分子组成的组合文库。

[0466] 化学文库由通过天然产物筛选鉴定为“击中 (hits)”或“前导 (leads)”的一种或多种已知化合物的结构类似物组成。天然产物文库衍生自微生物、动物、植物或海洋生物的集合, 其用于产生用于以下筛选的混合物: (1) 来自土壤、植物或海洋微生物的肉汤培养基的发酵和提取或 (2) 植物或海洋生物的提取。天然产物文库包括多酮化合物 (polyketides)、非核糖体肽及其变体 (非天然存在的)。参见例如Cane等人, *Science* 282:63-68, 1998。组合文库可以由作为混合物的大量肽、寡核苷酸或有机化合物构成。它们相对容易地通过常规自动化合成方法、PCR、克隆或专有的合成方法制备。

[0467] 更具体地, 组合化学文库是通过化学合成或生物合成产生的多种化合物的集合, 通过将许多化学“构件”例如试剂组合在一起。例如, 线性组合化学文库例如多肽文库是通过将一组化学构件 (氨基酸) 以给定化合物长度 (即, 多肽化合物中的氨基酸数目) 的每种可能方式组合而形成的。数百万化合物可以通过化学构件的这种组合混合来合成。

[0468] 关于组合化学和从其产生的文库的综述, 参见例如Huc, I. and Nguyen, R. (2001) *Comb.Chem.High Throughput Screen*4:53-74; Lepre, C A. (2001) *Drug Discov.Today*6:133-140; Peng, S.X. (2000) *Biomed.Chromatogr.*14:430-441; Bohm, H.J. 和Stahl, M. (2000) *Curr.Opin.Chem.Biol.*4:283-286; Barnes, C和Balasubramanian, S. (2000) *Curr.Opin.Chem.Biol.*4:346-350; Lepre, Enjalbal, C, 等人, (2000) *Mass Spectrom Rev.*19:139-161; Hall, D.G., (2000) *Nat.Biotechnol.*18:262-262; Lazo, J.S., 和Wipf, P. (2000) *J.Pharmacol.Exp.Ther.*293:705-709; Houghten, R.A., (2000) *Ann.Rev.Pharmacol.Toxicol.*40:273-282; Kobayashi, S. (2000) *Curr.Opin.Chem.Biol.* (2000) 4:338-345; Kopylov, A.M. 和Spiridonova, V.A. (2000) *Mol.Biol. (Mosk)* 34:1097-1113; Weber, L. (2000) *Curr.Opin.Chem.Biol.*4:295-302; Dolle, R.E. (2000) *J.Comb.Chem.*2:383-433; Floyd, C D., 等人, (1999) *Prog.Med.Chem.*36:91-168; Kundu, B., 等人, (1999) *Prog.Drug Res.*53:89-156; Cabilly, S. (1999) *Mol.Biotechnol.*12:143-148; Lowe, G. (1999) *Nat.Prod.Rep.*16:641-651; Dolle, R.E. 和Nelson, K.H. (1999) *J.Comb.Chem.*1:235-282; Czarnick, A.W. 和Keene, J.D. (1998) *Curr.Biol.*8:R705-R707;

Dolle, R.E. (1998) *Mol. Divers.* 4:233-256; Myers, P.L., (1997) *Curr. Opin. Biotechnol.* 8:701-707; 以及 Pluckthun, A. 和 Cortese, R. (1997) *Biol. Chem.* 378:443。

[0469] 用于制备组合文库的装置是商业途径可获得的(参见例如, 357 MPS, 390 MPS, Advanced Chem Tech, Louisville Ky., Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.)。此外, 许多组合文库本身可商业途径获得(参见例如, ComGenex, Princeton, N.J., Asinex, Moscow, Ru, Tripos, Inc., St. Louis, Mo., ChemStar, Ltd., Moscow, RU, 3D Pharmaceuticals, Exton, Pa., Martek Biosciences, Columbia, Md., 等)。

[0470] 使用方法

[0471] 本发明的实施方案包括治疗性的治疗方法。因此, 本文所述的AARS剂, 包括AARS多肽、AARS多核苷酸、基于AARS多核苷酸的载体、表达AARS的宿主细胞、反义寡核苷酸、RNAi剂以及诸如肽、抗体和抗原结合片段、肽模拟物和其他小分子的结合剂, 可用于治疗与参照AARS的非常规活性相关的多种非限制性疾病或病症。这类非常规活性的实例包括细胞外信号传导的调节、细胞增殖的调节、细胞迁移的调节、细胞分化的调节(例如血细胞生成、神经发生、肌发生、骨发生和脂肪生成)、细胞凋亡或其它形式的细胞死亡的调节、血管生成的调节、细胞结合的调节、细胞代谢的调节、细胞因子产生或活性的调节、细胞因子受体活性的调节、细胞摄取或分泌的调节、免疫调节、炎症的调节、代谢过程例如血糖控制的调节、等。

[0472] 包括基于多核苷酸的疗法, 例如反义疗法和RNAi干扰疗法, 其通常涉及降低靶分子的表达, 所述靶分子例如贡献其非常规活性的AARS多肽的内源性片段或AARS多肽的细胞结合配体。反义或RNAi疗法通常拮抗所述非常规活性, 例如通过降低AARS参照多肽的表达。还包括激动或拮抗AARS参照多肽的非常规活性的基于多肽或肽、抗体或抗原结合片段、肽模拟物或其他小分子的疗法, 例如通过与AARS多肽、其细胞结合配体或二者直接相互作用。

[0473] 这些和相关的实施方案包括使用本发明的AARS剂或组合物治疗细胞、组织或受试者的方法。本发明可以调节的细胞或组织优选为哺乳动物细胞或组织, 或更优选为人细胞或组织。这些细胞或组织可以是健康状态或患病状态。

[0474] 在某些实施方案中, 例如, 提供了用于调节治疗相关的细胞活性的方法, 所述治疗相关的细胞活性包括但不限于细胞代谢、细胞分化、细胞增殖、细胞摄取、细胞分泌、细胞死亡、细胞动员、细胞迁移、基因转录、mRNA翻译、细胞阻抗、免疫应答、炎症应答等, 所述方法包括将细胞与本文所述的AARS剂或AARS组合物接触。在某些实施方案中, 细胞在受试者中。因此, AARS组合物可以用来治疗基本上任何可以获益于一种或多种这些活性的调节的细胞或组织或受试者。

[0475] AARS剂和组合物还可以用于多种治疗环境中的任何一种, 例如包括与治疗或预防以下疾病相关的那些治疗环境: 肿瘤疾病、免疫系统疾病或病症(例如, 自身免疫性疾病和炎症)、感染性疾病、代谢疾病、神经元/神经病学疾病、肌肉/心血管疾病、异常血细胞生成相关的疾病、异常肌发生相关的疾病、异常神经发生相关的疾病、异常脂肪发生相关的疾病、异常骨发生相关的疾病、异常血管生成相关的疾病、异常细胞存活相关的疾病、异常脂质摄取相关的疾病、衰老相关的疾病(例如, 听力丧失、外周或自主神经病变、老年性痴呆、视网膜病)和其他疾病。

[0476] 例如, 在某些示例性实施方案中, 本发明的AARS组合物可以用于调节血管生成, 例

如,通过内皮细胞增殖和/或信号传导的调节。可以使用合适的细胞系(例如,人微脉管内皮肺细胞(HMVEC-L)和人脐带静脉内皮细胞(HUVEC))并使用合适的测定(例如,内皮细胞迁移测定、内皮细胞增殖测定、成管测定、基质胶栓测定等)来监测内皮细胞的增殖和/或信号传导,所述测定中的多数在本领域内是已知的并且是可以获得的。

[0477] 因此,在相关的实施方案中,本发明的组合物可以用来治疗基本上任何可以获益于血管生成调节的细胞或组织或受试者。例如,在一些实施方案中,可以将经历或易感于血管生成的细胞或组织或受试者(例如,血管生成病症)与本发明合适的组合物接触,从而抑制血管生成病症。在其他的实施方案中,可以将经历或易感于不足的血管生成的细胞或组织(例如,血管生成抑制病症)与本发明合适的组合物接触,从而干扰血管生成抑制活性和/或促进血管生成。

[0478] 还包括调节血细胞生成和相关病症的方法。本发明的AARS多肽可以调节的血细胞生成过程的实例包括,但不限于,骨髓细胞(例如,红系细胞,肥大细胞、单核细胞/巨噬细胞,髓样树突状细胞,诸如嗜碱性粒细胞、嗜中性粒细胞和嗜酸性粒细胞的粒细胞,巨核细胞,血小板)和淋巴细胞(例如,自然杀伤细胞、淋巴样树突状细胞、B细胞和T细胞)的形成。某些特定的血细胞生成过程包括红细胞生成、粒细胞生成、淋巴细胞生成、巨核细胞生成、血小板生成和其他。还包括调节造血细胞的转运或动员的方法,所述造血细胞包括造血干细胞、祖细胞、红细胞、粒细胞、淋巴细胞、巨核细胞和血小板。

[0479] 可以在体内、体外、离体或以上的任意组合进行调节血细胞生成的方法。这些方法可以对含有造血干细胞,造血祖细胞,或能够沿造血谱系分化的其他干细胞或祖细胞(例如,脂肪组织来源的干细胞)的任何生物样品、细胞培养物或组织进行。对于体外和离体方法,可以根据本文所述且本领域内已知的技术和特征分离和/或鉴定无论是造血来源还是其他来源的干细胞和祖细胞。

[0480] 本发明的组合物还可以用作免疫调节剂,用于通过调节直接或间接介导自身免疫和/或炎症疾病、病症和疾患的细胞来治疗抗炎症或促炎症的适应症。可以使用许多本领域内已知和可获得的技术的任何一种来监测本发明组合物作为免疫调节剂或炎症调节剂的功效,所述技术包括例如迁移测定(例如,使用白细胞或淋巴细胞)或细胞活力测定(例如,使用B细胞、T细胞、单核细胞或NK细胞)。

[0481] “炎症”通常是指组织对诸如病原体、受损伤细胞(例如,创伤)和刺激物的有害刺激的生物应答。术语“炎症应答”是指获取和调节炎症的具体机理,包括例如免疫细胞的激活或迁移,细胞因子的产生,血管扩张(包括激肽释放、纤溶和凝血)以及本文所述的本领域内已知的其他机理。

[0482] 慢性炎症的临床体征取决于疾病的持续时间、炎性病变、原因和受影响的解剖学区域(参见例如Kumar等人,Robbins Basic Pathology-第8版,2009 Elsevier,London; Miller,LM,Pathology Lecture Notes,Atlantic Veterinary College,Charlottetown, PEI,Canada)。慢性炎症与多种病理状态或疾病相关,包括例如过敏症、阿尔茨海默氏病、贫血、主动脉瓣狭窄、诸如类风湿性关节炎和骨关节炎的关节炎、癌症、充血性心脏衰竭、纤维肌痛、纤维化、心脏病发作、肾功能衰竭、红斑狼疮、胰腺炎、中风、手术并发症、炎性肺部疾病、炎性肠病、动脉粥样硬化、神经系统疾病、糖尿病、代谢紊乱、肥胖症和银屑病以及本文所述和本领域内已知的其他疾病。因此,AARS组合物可用于治疗或处理慢性炎症,调节任何

一种或多种的个体慢性炎症反应,或治疗任何一种或多种与慢性炎症相关的疾病或病症。

[0483] 评估炎症和其他病症的体征和症状的标准,包括出于做出鉴别诊断以及监测治疗的目的,例如通过根据已接受的临床标准确定改善来确定在治疗期间是否已经施用了治疗有效剂量,对本领域技术人员是明显的,并由以下的教导示例,例如Berkow等人编辑,The Merck Manual,第16版,Merck and Co.,Rahway,N.J.,1992;Goodman等人编辑,Goodman and Gilman's The Pharmacological Basis of Therapeutics,第10版,Pergamon Press, Inc.,Elmsford,N.Y.,(2001);Avery's Drug Treatment:Principles and Practice of Clinical Pharmacology and Therapeutics,第3版,ADIS Press,Ltd.,Williams and Wilkins,Baltimore,MD.(1987);Ebadi,Pharmacology, Little,Brown and Co.,Boston,(1985);Osolci al.编辑,Remington's Pharmaceutical Sciences,第18版,Mack Publishing Co.,Easton,PA(1990);Katzung,Basic and Clinical Pharmacology,Appleton and Lange,Norwalk,CT(1992)。

[0484] 在其他实施方案中,本发明的AARS组合物可以用于调节细胞增殖和/或存活,并且因此,用于治疗或预防特征为细胞增殖和/或存活异常的疾病、疾患或病症。例如,在某些实施方案中,AARS组合物可以用于调节凋亡和/或治疗异常凋亡相关的疾病或病症。可以使用许多本领域内已知的和可获得的技术的任何一种监测凋亡,包括例如测量DNA片段化、膜不对称性改变、凋亡性胱天蛋白酶激活和/或细胞色素C和AIF释放的测定。

[0485] 可以通过常规方法和测定并基于本领域医生或其他技术人员已知的标准容易地监测这些和其他疗法(例如体外疗法)的进展。

[0486] 药物制剂、施用和试剂盒

[0487] 本发明的实施方案包括AARS多核苷酸、AARS多肽、表达AARS多肽的宿主细胞、结合剂、调节剂或本文描述的其他化合物,其在药学上可接受的或生理可接受的溶液中配制,用于单独或与一种或多种其他治疗形式联合施用给细胞或动物。还应当理解,如果需要,本发明的组合物也可以与其他剂联合施用,所述其他剂例如其他蛋白或多肽或各种药物活性剂。对所述组合物中还包括的其他组分几乎没有限制,但其他试剂对需要达到的调节或其他效应没有不良影响。

[0488] 在本发明的药物组合物中,药学上可接受的赋形剂和载体溶液的配制是本领域技术人员所公知的,在许多治疗方案中使用本文所描述的特定组合物的合适的给药和治疗方案的开发也是本领域技术人员所公知的,包括例如口服、胃肠外、静脉内、鼻内、皮下和肌肉内施用和制剂。

[0489] 在某些实施方案中,本发明的药物组合物或治疗组合物不刺激免疫应答。在其他实施方案中,本发明的药物组合物或治疗组合物通常包含一种或多种AARS多肽或多核苷酸,例如通过用作疫苗或相关组合物中的佐剂或者与单独佐剂或刺激免疫应答的剂一起存在于组合物中,来刺激免疫应答。

[0490] 在某些实施方案中,诸如AARS多肽、AARS多核苷酸和抗体的AARS剂具有特定施用模式(例如静脉内施用)所需的溶解度。所需溶解度的实例包括至少约1mg/ml、至少约10mg/ml、至少约25mg/ml和至少约50mg/ml。

[0491] 在某些应用中,可以将本文所公开的药物组合物通过口服给药递送至受试者。因此,可以将这些组合物与惰性稀释剂或与可吸收可食用的载体一起配制,或可以将它们封

装于硬壳或软壳的明胶胶囊中,或可以将它们压成片剂,或可以将它们直接混入饮食食物中。

[0492] 在某些情况下,需要将本文所公开的药物组合物胃肠外、皮下、静脉内、肌内、动脉内、鞘内、实质内、心室内、尿道内、胸内、头盖内、滑膜内或甚至腹膜内递送,如例如在美国专利第5,543,158号、美国专利第5,641,515号和美国专利第5,399,363号(其每一篇通过引用整体并入本文)中所描述的。适合胃肠外施用的装置包括针(包括显微针)注射器、无针注射器和输注技术。

[0493] 作为游离碱或药学上可接受的盐的活性化合物溶液可以在水中制备,适宜地与表面活性剂诸如羟基丙基纤维素混合。分散液还可以在甘油、液体聚乙二醇或其混合物中以及在油中配制。在正常储存和使用的条件下,这些制剂包含防止微生物生长的防腐剂。

[0494] 适于注射使用的药物形式包括无菌水性溶液或分散液和用于无菌可注射溶液或分散液的即时制备的无菌粉末(美国专利第5,466,468号,明确地通过引用整体并入本文)。在所有情况下,所述形式应该是无菌的,并且流动性应达到易于注射的程度。它应该在制备和保存的条件下是稳定的,并且应该对抗诸如细菌和真菌的微生物的污染作用。所述载体可以是含有例如水、乙醇、多元醇(例如,丙三醇、丙二醇和液态聚乙二醇等)、其合适的混合物和/或植物油的溶剂或分散介质。通过使用诸如卵磷脂的包衣,对于分散液保持所需的粒度和通过使用表面活性剂,可以保持合适的流动性。各种抗细菌剂和抗真菌剂可以有利于微生物作用的预防,例如,对羟基苯甲酸酯、氯丁醇、苯酚、山梨酸、硫柳汞等。在许多情况下,优选包括等渗剂,例如,糖或氯化钠。通过在组合物中使用诸如单硬脂酸铝和明胶的延迟吸收剂,可以实现注射组合物的延长吸收。

[0495] 对于水性溶液的胃肠外给药,例如,如果需要,应当使溶液适当缓冲,并且首先用足够的盐水或葡萄糖使液体稀释剂成为等渗的。这些特定的水性溶液尤其适于静脉内施用、肌内施用、皮下施用和腹膜内施用。在该方面,根据本公开,可以利用的无菌水性介质是本领域技术人员已知的。例如,可将1个剂量溶于1ml等渗NaCl溶液中,并且或者加至1000ml的皮下输液或者在计划的输注部位进行注射(参见,例如,Remington's Pharmaceutical Sciences,第5版,pp.1035-1038和1570-1580)。根据待治疗受试者的病症,剂量必然会发生某些变化。在任何情况下,负责给药的人将确定对于个体受试者的合适剂量。而且,对于人类给药,制剂应该满足FDA生物制剂标准处所要求的无菌、热原性以及一般的安全和纯度标准。

[0496] 可以通过以下制备无菌的可注射溶液:将合适的溶剂中的所需量的活性化合物与上面列举的各种其他成分合并,如果需要,随后进行过滤灭菌。通常,通过将各种灭菌的活性成分合并入无菌媒介中来制备分散液,所述无菌媒介含有基础分散介质和来自那些上文列举的所需的其他成分。对于用于制备无菌可注射溶液的无菌粉末而言,优选的制备方法是真空干燥和冷冻干燥技术,这些技术可以从其事先无菌过滤的溶液产生活性成分外加任何其他所需成分的粉末。

[0497] 可以将本文所公开的组合物配制为中性形式或盐形式。药学上可接受的盐包括酸加成盐(与蛋白的自由氨基形成的),并且所述酸加成盐可以与诸如盐酸或磷酸的无机酸形成或与诸如醋酸、草酸、酒石酸、扁桃酸等的有机酸形成。与自由羧基形成的盐也可以来源于诸如氢氧化钠、氢氧化钾、氢氧化铵、氢氧化钙或氢氧化铁的无机碱和诸如异丙胺、三甲

胺、组氨酸、普鲁卡因等的有机碱。配制后,按照与剂型相容的方式和治疗有效量施用溶液。可以以多种剂型容易地施用制剂,例如可注射溶液、药物释放胶囊等。

[0498] 如本文所用的“载体”包括任何和所有的溶剂、分散介质、媒介、包衣、稀释剂、抗菌剂和抗真菌剂、等渗剂和吸收延迟剂、缓冲剂、载体溶液、悬浮液、胶体等。这样的介质和试剂用于药物活性物质的用途在本领域是众所周知的。除非任何常规介质或试剂与活性成分不相容,否则考虑在治疗组合物中使用它。补充的活性成分也能合并入组合物中。

[0499] 短语“药学上可接受的”是指这样的分子实体和组合物,当将其施用给人时,不会产生过敏性或相似的不良反应。包含作为活性成分的蛋白的水性组合物的制备是本领域公知的。通常,将这类组合物制备成为液体溶液或悬浮液的注射液;还可以制备成适于在注射前溶解或悬浮于液体中的固体形式。也可以将制剂乳化。

[0500] 在某些实施方案中,可以通过鼻内喷雾、吸入和/或其他气溶胶递送媒介递送药物组合物。通过鼻内气溶胶喷雾将基因、多核苷酸和肽组合物定向递送至肺的方法已有描述,例如,美国专利第5,756,353号和美国专利第5,804,212号(通过引用将每一篇的全部特别并入本文)。同样,使用鼻内微粒树脂(Takenaga等人,1998)和溶血磷脂酰甘油化合物(美国专利第5,725,871号,通过引用将其全部特别并入本文)的药物递送在制药行业内也是公知的。同样,聚四氟乙烯支持基质的形式的跨粘膜的药物递送描述于美国专利第5,780,045号(通过引用将其全部特别并入本文)。

[0501] 药物组合物可以被配制为立即和/或持续释放。持续释放组合物包括延迟的、修饰的、脉冲的、控制的、靶向的和程序化的释放。因此,组合物可以被配置为悬浮液或固体、半固体或摇溶液体以作为植入剂施用,提供AARS多核苷酸、AARS多肽、结合剂、调节剂和其他活性剂的持续释放。此类制剂的实例包括但不限于药物包衣支架以及包含以下的半固体和悬浮液:药物负载的(DL-乳酸-羟基乙酸)(PLGA)、聚(DL-丙交酯-乙交酯)(PLG)或聚(丙交酯)(PLA)层状囊泡或微粒,水凝胶(Hoffman AS:Ann.N.Y.Acad.Sci.944:62-73(2001)),Flamel Technologies Inc.开发的以商标MEDUSA®销售的聚氨基酸纳米粒子系统,Atrix,Inc.开发的以商标ATRIGEL®销售的非水性凝胶系统,和Durect Corporation开发的以商标SABER®销售的蔗糖醋酸异丁酸酯持续释放制剂,和由SkyePharma开发并以商标DEPOFOAM®销售的基于脂质的系统。

[0502] 能够在延长的时段递送所需剂量的药物组合物的持续释放装置是本领域已知的。例如,美国专利号5,034,229;5,557,318;5,110,596;5,728,396;5,985,305;6,113,938;6,156,331;6,375,978;和6,395,292;教导了能够在延长的时段(即,范围从超过一周至一年或更长的时段)以希望的速率递送活性剂制剂例如溶液或悬液的渗透驱动装置。其他示例性的持续释放装置包括可获自Medtronic的调节型泵,其提供有益剂制剂的恒速流动、可调节的流动或可程序化的流动,包括以商标SYNCHROMED INFUSION SYSTEM®销售的泵,以商标CODMAN®分水泵销售的Johnson and Johnson系统和INSET®技术泵。装置的其他实例描述于美国专利号6,283,949;5,976,109;5,836,935;和5,511,355。

[0503] 在某些实施方案中,可以通过使用脂质体、纳米胶囊、微颗粒、微球、脂质颗粒、囊泡等进行递送,用于将本发明组合物引入合适的宿主细胞中。特别是,可以配制本发明组合



物用于封装在脂质颗粒、脂质体、囊泡、纳米球、纳米颗粒等中递送。可以用已知的和常规的技术进行这类递送媒介的配制和使用。

[0504] 在某些实施方案中,本文提供的剂可以连接至药学上可接受的固体基底,包括生物相容和可生物降解的基底,例如聚合物和基质。此类固体基底的实例包括但不限于聚酯、水凝胶(例如,聚(2-羟基乙基-甲基丙烯酸酯)或聚(乙烯醇))、聚交酯(美国专利号3,773,919)、L-谷氨酸和 $\gamma$ -乙基-L-谷氨酸的共聚物、不可降解的乙烯-乙酸乙烯酯、可降解的乳酸-羟基乙酸共聚物,例如聚(乳酸-羟基乙酸)(PLGA)和LUPRON DEPOT™(由乳酸-羟基乙酸共聚物和乙酸亮丙瑞林构成的可注射微球)、聚-D-(-)-3-羟基丁酸、胶原、金属、羟基磷灰石、生物玻璃、铝酸盐、生物陶瓷材料和纯化蛋白。

[0505] 在一个具体实施方案中,固体基底包括以商标ATRIGEL™(QLT, Inc., Vancouver, B.C.)销售的可生物降解的聚合物。ATRIGEL®药物递送系统由溶解于生物相容性载体的可生物降解聚合物组成。药物可以在生产时混入该液体递送系统,或者根据产品可以后来在使用时由医师添加。当液体产品通过小号针头注入皮下空间或通过导管放入可及组织部位时,组织液体中的水导致聚合物沉淀并将药物捕获在固体植入剂中。然后,随着聚合物基质随时间的生物降解,包封在植入剂内的药物以控制的方式释放。

[0506] 用于本发明的药物组合物还可以局部、皮(内)或经皮施用给皮肤或粘膜。用于该目的的典型制剂包括凝胶、水凝胶、洗剂、溶液、乳膏、软膏、撒粉剂、敷剂、泡沫剂、膜剂、皮肤贴片、薄片、植入剂、海绵剂、纤维剂、绷带和微乳液。还可以使用脂质体。典型的载体包括醇、水、矿物油、液体石蜡、白矿脂、甘油、聚乙二醇和丙二醇。可以加入穿透增强剂,参见例如Finnin和Morgan: J. Pharm. Sci. 88 (10): 955-958, (1999)。局部施用的其他方式包括通过电穿孔、离子电渗、声透、超声促渗和显微针或无针注射,例如使用以商标POWDERJECT™和BIOJECT™销售的系统。

[0507] 配制方法是本领域公知的并且公开于例如Remington: The Science and Practice of Pharmacy, Mack Publishing Company, Easton, Pa., 第20版, ISBN: 0683306472 (2000)。本文提供的组合物和剂可以根据本发明的方法以任何治疗有效的给药方案施用。选择剂量和频率来产生没有有害效应的该剂的有效水平。本发明化合物的有效量将取决于施用途径、待治疗的温血动物类型和考虑的特定温血动物的身体特性。这些因素及其与测定该量的关系是医疗领域技术人员公知的。可以调整该量和施用方法以实现最佳效力,但是将取决于如下因素:诸如体重、饮食、共存药物治疗和医疗领域技术人员意识到的其他因素。

[0508] 在特定实施方案中,当口服或静脉内施用,施用的组合物或剂的量一般范围从约0.1至约100mg/kg/天并且通常从约0.1至10mg/kg的剂量。在特定实施方案中,剂量是5mg/kg或7.5mg/kg。在不同实施方案中,剂量是约50-2500mg/天、100-2500mg/天、300-1800mg/天或500-1800mg/天。在一个实施方案中,剂量是约100至600mg/天。在另一实施方案中,剂量是约300至1200mg/天。在特定实施方案中,组合物或剂以100mg/天、240mg/天、300mg/天、600mg/天、1000mg/天、1200mg/天或1800mg/天的剂量施用,每天一剂或多剂(即,此时合并的剂实现所需的日剂量)。在相关实施方案中,剂量是100mg bid、150mg bid、240mg bid、300mg bid、500mg bid或600mg bid。在不同实施方案中,组合物或剂以单次或重复给药施用。初始剂量和随后剂量可以相同或不同。

[0509] 在某些实施方案中,以0.1至10mg/kg或0.5至5mg/kg的单剂量施用组合物或剂。在其他实施方案中,以0.1至50mg/kg/天、0.5至20mg/kg/天或5至20mg/kg/天的剂量施用组合物或剂。

[0510] 在某些实施方案中,例如通过在约例如10分钟至90分钟的时段内输注来口服或静脉内施用组合物或剂。在其他相关实施方案中,通过在一段时间内例如以约0.1至约10mg/kg/h的剂量连续输注来施用组合物或剂。虽然时段可以变化,但在某些实施方案中,时段可以在约10分钟至约24小时或者在约10分钟至约三天。

[0511] 在特定实施方案中,有效量或治疗有效量是足以在受试者的血浆中实现约0.1 $\mu$ g/ml至约20 $\mu$ g/ml或者约0.3 $\mu$ g/ml至约20 $\mu$ g/ml的 $C_{max}$ 的组合物或剂的总浓度的量。在某些实施方案中,口服剂量是足以实现约0.1 $\mu$ g/ml至约5 $\mu$ g/ml或约0.3 $\mu$ g/ml至约3 $\mu$ g/ml的血浆浓度( $C_{max}$ )的量。在某些实施方案中,静脉内剂量是足以实现约1 $\mu$ g/ml至约10 $\mu$ g/ml或约2 $\mu$ g/ml至约6 $\mu$ g/ml的血浆浓度( $C_{max}$ )的量。在相关实施方案中,受试者血浆中的剂的总浓度具有小于约20 $\mu$ g/ml的平均谷浓度和/或小于约20 $\mu$ g/ml的稳态浓度。在进一步实施方案中,受试者血浆中剂的总浓度具有小于约10 $\mu$ g/ml的平均谷浓度和/或小于约10 $\mu$ g/ml的稳态浓度。

[0512] 而在另一实施方案中,受试者血浆中剂的总浓度具有约1ng/ml至约10 $\mu$ g/ml的平均谷浓度和/或约1ng/ml至约10 $\mu$ g/ml的稳态浓度。在一个实施方案中,受试者血浆中剂的总浓度具有约0.3 $\mu$ g/ml至约3 $\mu$ g/ml的平均谷浓度和/或约0.3 $\mu$ g/ml至约3 $\mu$ g/ml的稳态浓度。

[0513] 在特定实施方案中,组合物或剂以足以在哺乳动物中实现具有约1ng/ml至约10 $\mu$ g/ml的平均谷浓度和/或约1ng/ml至约10 $\mu$ g/ml的稳态浓度的血浆浓度的量施用。在相关实施方案中,哺乳动物血浆中剂的总浓度具有约0.3 $\mu$ g/ml至约3 $\mu$ g/ml的平均谷浓度和/或约0.3 $\mu$ g/ml至约3 $\mu$ g/ml的稳态浓度。

[0514] 在本发明的特定实施方案中,组合物或剂的有效量,或者组合物或剂的血浆浓度被实现或维持例如至少15分钟、至少30分钟、至少45分钟、至少60分钟、至少90分钟、至少2小时、至少3小时、至少4小时、至少8小时、至少12小时、至少24小时、至少48小时、至少3天、至少4天、至少5天、至少6天、至少1周、至少2周、至少1个月、至少2个月、至少4个月、至少6个月、至少1年、至少2年或超过2年。

[0515] 在某些基于多肽的实施方案中,施用的多肽的量通常在约0.1 $\mu$ g/kg至约0.1mg/kg至约50mg/kg患者体重的范围内。根据疾病的类型和严重程度,例如无论是以一次或多次单独施用或者通过连续输注,施用给患者的初始候选物剂量可以是约0.1 $\mu$ g/kg至约0.1mg/kg至约50mg/kg体重(例如,约0.1–15mg/kg/剂)的多肽。例如,给药方案可以包括施用约4mg/kg的初始给药剂量,随后约2mg/kg的每周维持剂量的多肽,或者负载剂量的大约一半。然而,可以使用其他剂量方案。典型的日剂量范围可以是约0.1 $\mu$ g/kg至约1 $\mu$ g/kg至100mg/kg或更多,取决于上述因素。关于在数日或更长时间的重复施用,根据病症,治疗持续至出现预期的疾病症状的抑制。

[0516] 在特定实施方案中,有效剂量实现本文所述的组合物或剂的血浆水平或平均谷浓度。这些可以使用常规实验容易地确定。

[0517] 在其他方面,本发明提供了包含一个或多个容器的试剂盒,所述容器填充了如本文所述的本发明的一种或多种多肽、多核苷酸、抗体、多单元复合物、其组合物等。所述试剂盒包括如何使用这些组合物的书面说明书(例如,调节细胞信号传导、血管生成、癌症、炎性

条件、诊断等)。

[0518] 本文的试剂盒还可以包含一种或多种其他的治疗剂或待治疗的适应症或期望的诊断应用所适合或需要的其他组分。如果需要,其他的治疗剂可以容纳在第二容器中。其他治疗剂的实例包括但不限于抗肿瘤剂、抗炎剂、抗细菌剂、抗病毒剂、血管生成剂等。

[0519] 本文的试剂盒还可以包含一种或多种注射器或促进预期递送方式所必需或所需要的其他组分(例如,支架、可植入的贮库等)。

[0520] 通过引用将本说明书所引用的所有出版物、专利申请和授权专利并入本文,如同明确和单独指出通过引用并入每一个单独的出版物、专利申请或授权专利。

[0521] 尽管出于清楚理解的目的,以示例说明和实例的方式详细地描述了上述发明,但是对于本领域技术人员明显的是,根据本发明的教导,在不脱离所附权利要求的精神和范围的情况下可以进行某些改变和修改。提供以下实施例仅为说明而非限制。本领域技术人员容易想到,可以改变和修改许多非关键参数而产生基本相似的结果。

#### XIV. 实施例

[0522] 一般方法 除非在以下实施例中另外指明,使用以下关于基因优化、小和大规模蛋白表达、蛋白纯化、转录分析和筛选的一般方法来制备和表征以下实施例中描述的AARS多肽。

[0523] 基因合成和克隆入表达载体

[0524] 使用下列方法将编码AARS多肽的表位标记形式的多核苷酸序列密码子优化并克隆入细菌表达载体。

[0525] 在方法(1)中,通过DNA 2.0(Menlo Park,CA)合成编码每种AARS多肽的大肠杆菌密码子优化的DNA(Welch等人,PLoS ONE 4(9):e7007 doi:10.1371/journal.pone.0007002),并且合成每种AARS多肽的含有N端或C端组合表位标签的两种形式,两者都包含六组氨酸标签和V5表位标签。

[0526] 利用以5'至3'方向编码核糖体结合位点(rbs(以下加下划线))、NdeI限制位点、六组氨酸标签和V5表位标签的5'延伸来合成编码N端标记的AARS多肽的DNA(AGGAGGTAAAACA TATGCATCATCATCATCATCACGGTAAGCCTATCCCTAACCCCTTTGCTCGGTCTCGATTCTACG)(SEQ.ID.No.1),其与预测的AARS多肽开放阅读框以框内融合。在AARS多肽包含预测的天然起始蛋氨酸(ATG)残基或者预测的AARS多肽的第一氨基酸残基是Met的情况下,缺失它。在预测的AARS多肽开放阅读框末端,添加两个终止密码子和XhoI位点(TAATGACTCGAG)(SEQ.ID.No.2)。

[0527] 利用编码rbs(以下加下划线)和NdeI限制位点的5'延伸来合成编码C端标记的AARS多肽的DNA,其再现了AARS多肽的预测的天然起始密码子或者与预测的AARS多肽开放阅读框在框内插入ATG(AGGAGATAAAACATATG)(SEQ.ID.No.3)。在不同的实施方案中,核糖体结合位点可以包含序列“AGGAGGTAAAACAT”(SEQ.ID.No.4)、“AGGAGATAAAACAT”(SEQ.ID.No.5)或GAAGGAGATATACAT(SEQ.ID.No.6)。在预测的AARS多肽开放阅读框的3'端,合成了以5'至3'顺序编码V5表位标签、六组氨酸标签、两个终止密码子和XhoI位点的3'延伸(GGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTCGATTCTACGCACCACCATCATCACCATTAATGACTCGAG)(SEQ.ID.No.7),其与预测的AARS多肽开放阅读框在框内融合。如果AARS多肽包括预测的天

然终止密码子,则缺失它。

[0528] 将编码AARS多肽的合成的DNA序列亚克隆入pJExpress411载体(DNA 2.0)。在测序以证实正确产物合成之后,将表达载体转化进入细菌以蛋白表达,如下更全面描述的。

[0529] 在方法(2)中,通过GENEWIZ(South Plainfield,NJ)合成编码每个AARS多肽的大肠杆菌密码子优化的DNA(Ermolaeva MD(2001)Curr.Iss.Mol.Biol.3(4)91-7)。利用包含用于随后克隆的独特限制位点的短的5'和3'延伸合成了每个编码AARS多肽的多核苷酸序列。

[0530] 具体地,将BamHI限制位点插入在预测的开放阅读框的5'端。在AARS多肽包含预测的天然起始氨基酸残基(ATG)或者预测的AARS多肽的第一氨基酸残基是Met的情况下,缺失它。此外,在预测的开放阅读框的3'末端插入XhoI限制位点。在AARS多肽包含预测的天然终止密码子的情况下,缺失它。

[0531] 在限制性消化之后,将得到的DNA序列亚克隆入含有N端(pET24b\_N-6XHis/V5)或C端(pET24b\_C-V5/6XHis)组合表位标签的修饰的pET-24b载体(EMD,Gibbstown,NJ),两者都包含六组氨酸和V5表位标签(通过GENEWIZ,(South Plainfield,NJ)进行载体修饰)。

[0532] 在限制性消化和克隆之后,将编码N标记的AARS多肽的DNA克隆入N-标记的载体(pET24b\_N-6XHis/V5),其包含编码六组氨酸和V5表位标签的5'DNA序列(CATATGCATCATCATCATCATCACGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGGGATCC)(SEQ.ID.No.8),与嵌入NdeI限制位点的起始密码子(ATG)在框内。该5'延伸通过短的2氨基酸连接体(GS)与预测的AARS多肽开放阅读框融合。

[0533] 在预测的开放阅读框的3'端,编码N-标记AARS多肽的DNA包含编码2氨基酸延伸(LE)和随后两个终止密码子的DNA序列(CTCGAGTAATGA)(SEQ.ID.No.9)。

[0534] 在限制性消化和克隆之后,克隆入C-标记载体(pET24b\_C-V5/6XHis)的编码C-标记AARS多肽的DNA包含通过短的2氨基酸连接体(GS)与预测的AARS多肽开放阅读框融合的编码嵌入NdeI限制位点的起始密码子(ATG)的5'序列(CATATGGGATCC)(SEQ.ID.No.10)。

[0535] 在预测的开放阅读框的3'端,编码C-标记的AARS多肽的DNA包含编码短的连接体2氨基酸连接体(LE)和随后V5表位标签和随后六组氨酸和两个终止密码子的3'DNA序列CTCGAGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCACCACCACCACCACCACCTAATGA(SEQ.ID.No.11)。

[0536] 在方法(3)(制备AARS多肽用于动物研究)中,使用LR克隆酶重组反应将人类全长苯丙氨酰-tRNA合成酶(PheRSa)(Invitrogen Ultimate ORF collection Clone ID IOH5778)通过重组亚克隆到pET301(Invitrogen)中。使用AccuPrime Pfx SuperMix(Invitrogen,cat no 12344-040)将所得载体用作模板以PCR扩增编码PheRSa的DNA,从而增加5'NdeI位点和3'XhoI位点,使用下列寡核苷酸:

[0537] 5'-GGCCAACATATGGCGGATGGTCAGGTGG-3'(SEQ.ID.No.93)

[0538] 5'-GATTCTCGAGCGCAGCCTCCTGTGTGG-3'(SEQ.ID.No.94)

[0539] 所得PCR产物使用Qiagen PCR Purification Kit(Qiagen,cat no 28104)纯化。使用上述限制性位点,将纯化的PCR产物亚克隆到含有框内C-端V5表位标签和八组氨酸标签版本的pET21a(Novagen)。编码克隆到C-标记的载体(pET21a\_C-V5/8XHis)的C-标记的AARS的DNA包含编码嵌入载体序列中所含的NdeI限制性位点的起始密码子(ATG)的5'序列,

其融合预计的AARS多肽开放阅读框。在预计的开放阅读框的3'端,编码C-标记的PheRSa的DNA包含编码2个氨基酸延伸(LE)的DNA序列、然后是V5表位标签和终止密码子:

[0540] 5'-CTCGAGGGTAAGCCTATCCCTAACCTCTCTCGGTCTCGATTCTACGCATCATCACCACCACCA  
CCACCACTGA-3' (SEQ.ID.No.95)

[0541] DNA测序用于证实预计的多核苷酸序列(Retrogen, San Diego)。

[0542] ssDNA制备:为了制备pET21\_C-V5/8XHis-PheRSa ssDNA,将dsDNA载体转换为CJ236细菌细胞(NEB, cat no E4141S)中,并且接种到含有氨苄西林(100 $\mu$ g/mL)和氯霉素(30 $\mu$ g/mL)的LB-琼脂板上。将板在37 $^{\circ}$ C下孵育过夜。将菌落用于接种含有氨苄西林和氯霉素的LB培养基,并且在225rpm和37 $^{\circ}$ C下孵育过夜。将含有氨苄西林和氯霉素的20mL的LB接种200 $\mu$ L的整夜培养物,并且在225rpm和37 $^{\circ}$ C下生长2小时。使培养物感染5e9 pfu的M13K07辅助噬菌体(NEB, cat no N0315S)。在1小时后,将卡那霉素加入培养物中至最终浓度为50 $\mu$ g/mL,并且在225rpm和37 $^{\circ}$ C下孵育过夜。通过在1900x g下两次离心将细菌从培养物中分离并丢弃。通过和最终浓度4%PEG-8000和500mM乙酸钠在4 $^{\circ}$ C下孵育过夜使ssDNA析出。使ssDNA在12000x g下离心,并且重新混悬于1.4mL LB培养基。通过随后以14,500x g离心除去细胞碎片。使用Qiagen QIAprep M13试剂盒(Qiagen, cat no 27704)从上清中纯化ssDNA。

[0543] Kunkel突变:进行Kunkel突变以使用下列寡核苷酸来制备PheRSa physiocrine PheRSa1<sup>N9</sup>:

[0544] (5'-GTTAGGGATAGGCTTACCCTCGAGCCCCAGCTTCTCAGCCTGTCCCCC-3') SEQ.ID.No.96

[0545] 将100ng的寡核苷酸和5U PNK激酶(Roche, cat no 10633542001)在1X PNK激酶缓冲液和0.5mM ATP的存在下孵育。使反应在37 $^{\circ}$ C下孵育1小时。

[0546] 在75 $^{\circ}$ C的加热块上,使100ng的ssDNA载体和退火缓冲液(20mM Tris, pH7.4, 2mM MgCl<sub>2</sub>, 50mM NaCl, 最终浓度)中的6.9ng的激酶寡核苷酸孵育5分钟。使反应冷却至室温,同时保留在加热块中。

[0547] 为了延伸质粒,将1U的T4 DNA聚合酶(Roche, cat no 11004786001)和1U T4 DNA连接酶(Roche, cat no 10481220001)加入反应液中。另外,加入合成缓冲液至最终浓度0.45mM dNTPs, 0.91mM ATP, 9.1mM Tris, pH7.4, 4.5mM MgCl<sub>2</sub>和1.8mM DTT)。将该反应在冰上孵育5分钟,然后在37 $^{\circ}$ C下孵育90分钟。使5 $\mu$ L的延伸反应液转换到200 $\mu$ L DH5a细胞中。将转换物接种到氨苄西林板上,并且在37 $^{\circ}$ C下孵育过夜。

[0548] 将单独菌落用于接种含有氨苄西林的LB培养基。使培养物在37 $^{\circ}$ C下生长过夜。使用Qiagen Spin Miniprep试剂盒(Qiagen, cat no 27106)制备DNA质粒,并且验证序列(Retrogen, San Diego)。在测序证实合成正确产物后,将表达载体转换到细菌中以如下面更详细所述用于蛋白表达。

[0549] AARS多肽表达、纯化和生物物理表征

[0550] 根据所需的蛋白量,6xHis-标记的AARS多肽以中等通量形式和/或以较大规模烧瓶培养在细菌中表达。使用如下所述并且为具体实验指定的亲和色谱和离子交换色谱纯化AARS多肽。

[0551] 细菌培养:在PCR板中42 $^{\circ}$ C下将100ng包含编码每种AARS多肽的密码子优化的DNA的表达载体(如上所述)转化进入BL21(DE3)(EMD chemicals, 目录号69450)感受态大肠杆

菌细菌,持续30秒钟。还评价了C41 (DE3) (Lucigen, 目录号60442)、HMS174 (DE3) (EMD chemicals, 目录号69453) 和Origami2 (DE3) (EMD chemicals, 目录号71345) 菌株。将板在冰上放置2分钟,添加100 $\mu$ L SOC培养基,随后在37 $^{\circ}$ C孵育1小时。向24孔深孔板(Qiagen, 目录号19583) 的每个孔加入5mL补充了卡那霉素(100 $\mu$ g/mL) 的自动诱导培养基(EMD chemicals, 目录号71491)。向个体孔添加转化反应物,将深孔板用粘附膜(VWR, 目录号60941-078) 密封并在37 $^{\circ}$ C摇动器中以250rpm孵育过夜。当使用低温(25 $^{\circ}$ C) 条件时,改为进行孵育48小时。

[0552] 为了较大规模表达,将补充了卡那霉素(100 $\mu$ g/mL) 的200mL自动诱导培养基添加入带有通气帽的500-mL锥形烧瓶(Corning, 目录号431401)。向个体烧瓶添加转化反应物,并在37 $^{\circ}$ C摇动器中以250rpm孵育30小时。

[0553] 蛋白分离:在培养物达到稳定期(通常OD<sub>600</sub>是3-6) 时,将深孔板在3600x g离心10分钟。小心吸出培养基并使深孔板在-80 $^{\circ}$ C或-20 $^{\circ}$ C冷冻10分钟。然后使深孔板在室温解冻并向每个孔添加1mL裂解缓冲液(补充了200 $\mu$ L lysonase (EMD chemicals, 目录号71370) 和蛋白酶抑制剂“complete mini EDTA-free” (Roche, 目录号11 836 170 001) 的100mL Bugbuster)。通过反复吹吸重悬沉淀直至没有可见的块,并转移至eppendorf管,随后在室温下在摇动器上孵育10-20分钟。在4 $^{\circ}$ C下以16,000g离心10分钟之后,将裂解物加载到Ni-NTA Superflow 96 BioRobot试剂盒(Qiagen, 目录号969261) 中包含的TurboFilter96板并以500g离心5-10分钟。

[0554] 为了较大规模表达,将稳定期培养物转移入500-mL瓶子并以6,000g离心10分钟。弃掉培养基并将沉淀在进一步处理之前保存在-80 $^{\circ}$ C或-20 $^{\circ}$ C。然后允许沉淀在室温下解冻,并向每个瓶子添加20mL裂解缓冲液。通过反复吹吸悬浮沉淀直至没有可见的块,随后在室温下在摇动器上孵育20分钟。在4 $^{\circ}$ C下以10,000g离心30分钟之后,将裂解物转移至洁净的管或瓶中。如果在转移期间留存痕量碎片,则将样品再次离心或通过0.45 $\mu$ m醋酸纤维素膜(Corning, 目录号430314) 以进一步澄清。

[0555] 亲和纯化:将QIAFilter96板加载Ni-NTA Superflow 96 BioRobot试剂盒中包含的200 $\mu$ L Ni-NTA Superflow浆体,并且通过添加600 $\mu$ L结合缓冲液(20mM磷酸钠、500mM氯化钠和10mM咪唑,pH7.5) 来平衡树脂。施加-15英寸汞柱的真空,直至所有缓冲液穿过树脂。然后将来自之前步骤的澄清的细胞裂解物加载至QIAFilter<sup>®</sup>96板并允许结合5分钟。施加-3英寸汞柱的真空大约5分钟,直至所有样品穿过树脂。然后用1mL结合缓冲液洗涤树脂,随后用含有0.1%Triton X-100的1mL结合缓冲液洗涤两次。然后,用不含Triton X-100的1mL结合缓冲液洗涤树脂10次。结合的6xHis-标记的AARS多肽用450 $\mu$ L洗脱缓冲液(20mM磷酸钠、500mM氯化钠和500mM咪唑,pH7.5) 洗脱并在4 $^{\circ}$ C贮存。

[0556] 为了较大规模表达,用1mL Ni-NTA Superflow浆体(Qiagen, 目录号30450) 加载空的一次性柱“Poly-Prep” (Bio-Rad, 目录号731-1550), 并通过添加5mL结合缓冲液来平衡0.5mL树脂。然后将来自之前步骤的澄清的细胞裂解物加载到柱上,并允许通过重力穿过。树脂首先用含有0.1%Triton X-100的50mL结合缓冲液洗涤,然后用不含Triton X-100的50mL结合缓冲液洗涤。结合的6xHis-标记的AARS多肽用2mL洗脱缓冲液洗脱并在4 $^{\circ}$ C贮存。

[0557] 脱盐和精制步骤:对于分子量>10kDa的AARS多肽,将AcroPrep96滤板(Pall, 目录号5034) 的Omega10K膜用20 $\mu$ L 1X PBS冲洗,并将该板放在真空歧管上(>10英寸汞柱) 直至

所有液体穿过。将来自之前步骤 (Ni-NTA) 的洗脱液分配入每个孔并施加真空直至所有液体穿过。重复这些步骤,直至处理完总的洗脱液体积 (450 $\mu$ L)。

[0558] 通过向每个孔添加180 $\mu$ L 1X PBS pH7.4来回收AARS多肽,小心吹吸10次,然后转移至洁净的深孔板。重复该步骤以产生每孔360 $\mu$ L的总体积,并将深孔板贮存在4 $^{\circ}$ C。对于分子量<10kDa的AARS多肽,将来自Ni-NTA的洗脱液加载至具有Ultracel-3膜 (Millipore, 目录号UFC900308)的Amicon Ultra-15离心过滤装置,随后添加10mL 1X PBS并以3,600g离心10-30分钟,直至体积小于360 $\mu$ L。回收样品并添加1X PBS至360 $\mu$ L的终体积。

[0559] 为了去除内毒素,将具有Mustang Q膜的AcroPrep Advance滤板 (Pall, 目录号8171)用300 $\mu$ L 1X PBS冲洗并以1,000g离心5分钟以去除缓冲液。向滤板添加脱盐的AARS多肽 (360 $\mu$ L/孔)并在摇动器上孵育5-10分钟。然后将板以1,000g离心5-10分钟,收集含有AARS多肽的流经级分并在4 $^{\circ}$ C贮存。

[0560] 为了较大规模表达,根据AARS多肽的分子量,将来自Ni-NTA的洗脱液加载至具有Ultracel-3或Ultracel-10膜 (Millipore, 目录号UFC900308或UFC901008)的Amicon Ultra-15离心过滤装置,然后以3,600g离心10-30分钟,直至体积减少至250 $\mu$ L。将样品在10mL 1X PBS, pH7.4中混合并以3,600g再次离心10-30分钟,直至体积约250 $\mu$ L。再重复一次该步骤,回收上清液并添加1X PBS至1.5mL的终体积。

[0561] 为了去除内毒素,用1mL 1X PBS冲洗Sartobind Q5强阴离子交换膜 (Sartorius, 目录号Q5F),使用塑料注射器使AARS多肽缓慢穿过该膜。将含有AARS多肽的流经级分收集在96深孔板中,将该板密封并贮存在4 $^{\circ}$ C。

[0562] 如下所述,使用亲和色谱和一系列再折叠步骤纯化在细菌中表达并存在于包涵体中的6xHis-标记的AARS多肽。

[0563] 细菌培养:在PCR板中42 $^{\circ}$ C下将100ng编码每种AARS多肽的质粒转化进入BL21 (DE3) (EMD chemicals, 目录号69450)或C41 (DE3) (Lucigen, 目录号60442)感受态大肠杆菌细菌,持续30秒钟。将板放在冰上2分钟并添加100 $\mu$ L SOC培养基,随后在37 $^{\circ}$ C孵育1小时。将补充了卡那霉素 (100 $\mu$ g/mL)的5mL自动诱导培养基 (EMD chemicals, 目录号71491)添加至24孔深孔板 (Qiagen, 目录号19583)的每个孔。向个体孔添加转化反应物,将深孔板用粘附膜 (VWR, 目录号60941-078)密封并在37 $^{\circ}$ C摇动器中以250rpm孵育过夜。

[0564] 为了较大规模表达,将补充了卡那霉素 (100 $\mu$ g/mL)的200mL自动诱导培养基加入带有通气帽的500-mL锥形烧瓶 (Corning, 目录号431401)。向个体烧瓶添加转化反应物,并在37 $^{\circ}$ C摇动器中以250rpm孵育30小时。

[0565] 分离:在培养物达到稳定期 (通常OD<sub>600</sub>是3-6)时,将深孔板在3600x g离心10分钟。小心吸出培养基并使深孔板在-80 $^{\circ}$ C或-20 $^{\circ}$ C冷冻10分钟。然后使深孔板在室温解冻并向每个孔添加1mL裂解缓冲液 (补充了200 $\mu$ L lysonase (EMD chemicals, 目录号71370)和蛋白酶抑制剂“complete mini EDTA-free” (Roche, 目录号11 836 170 001)的100mL Bugbuster)。通过反复吹吸重悬沉淀直至没有可见的块,并转移至eppendorf管,随后在室温下在摇动器上孵育10-20分钟。在4 $^{\circ}$ C下以16,000g离心10分钟之后,弃掉可溶性裂解物,并将包涵体完全重悬于变性结合缓冲液 (20mM磷酸钠、500mM氯化钠、6M盐酸胍、10mM咪唑, pH7.5)。将样品以16,000g离心10分钟,并将上清液加载至Ni-NTA Superflow96 BioRobot试剂盒 (Qiagen, 目录号969261)中包含的TurboFilter96板,随后以500g离心5-10分钟。将

滤液收集在洁净的96孔深孔板(Greiner,目录号780286)中。

[0566] 为了较大规模表达,将稳定期培养物转移入500-mL瓶子并以6,000g离心10分钟。弃掉培养基并将沉淀在进一步处理之前保存在-80℃或-20℃。然后允许沉淀在室温下解冻,并向每个瓶子添加20mL裂解缓冲液。通过反复吹吸悬浮沉淀直至没有可见的块,随后在室温下在摇动器上孵育20分钟。在4℃下以10,000g离心30分钟之后,将可溶性裂解物弃掉,并将不溶性包涵体完全重悬于变性结合缓冲液中。

[0567] 亲和纯化:将QIAFilter 96板加载Ni-NTA Superflow 96 BioRobot试剂盒中包含的200μL Ni-NTA Superflow浆体,并且通过添加600μL变性结合缓冲液(见上文)来平衡树脂。施加-15英寸汞柱的真空,直至所有缓冲液穿过树脂。然后将来自之前步骤的澄清的变性样品加载至QIAFilter® 96板并允许结合5分钟。施加大约3英寸汞柱的真空大约5分钟,直至所有样品穿过树脂。然后用1mL变性结合缓冲液洗涤树脂,随后用含有0.1%Triton X-100的1mL变性结合缓冲液洗涤五次。然后,用不含Triton X-100的1mL变性结合缓冲液洗涤树脂15次。然后,用450μL变性洗脱缓冲液(20mM磷酸钠、500mM氯化钠、6M盐酸胍和500mM咪唑,pH7.5)洗脱结合的6xHis-标记的AARS多肽并在4℃贮存。

[0568] 为了较大规模表达,用1mL Ni-NTA Superflow浆体(Qiagen,目录号30450)加载空的一次性柱“Poly-Prep”(Bio-Rad,目录号731-1550),并通过添加5mL变性结合缓冲液(见上文)来平衡0.5mL树脂。然后将来自之前步骤的变性包涵体加载到柱上,并允许通过重力穿过。树脂首先用含有0.1%Triton X-100的50mL变性结合缓冲液洗涤,然后用不含Triton X-100的50mL变性结合缓冲液洗涤。结合的6xHis-标记的AARS多肽用2mL变性洗脱缓冲液洗脱并在4℃贮存。

[0569] 再折叠:

[0570] 对于>10kDa的AARS多肽,将AcroPrep96滤板(Pall,目录号5034)的Omega10K膜用20μL 1X PBS冲洗,并将板放入真空歧管(>10英寸汞柱),直至所有液体穿过。将来自之前步骤(Ni-NTA)的洗脱液分配入每个孔并施加真空直至所有液体穿过。重复这些步骤,直至处理完总的洗脱液体积(450μL)。通过向每个孔添加200μL再折叠缓冲液(含有50mM Tris、250mM氯化钠、10mM氯化钾、2mM氯化镁、2mM氯化钙、400mM蔗糖、500mM精氨酸、1mM DTT和0.01%聚山梨酯80,pH7.4)来回收AARS多肽,小心吹吸10次,然后转移至洁净的深孔板。重复该步骤以产生每孔400μL的总体积,并将深孔板放在摇动器上4℃过夜。对于<10kDa的AARS多肽,将来自Ni-NTA的洗脱液加载至具有Ultracel-3膜(Millipore,目录号UFC900308)的Amicon Ultra-15离心过滤装置,随后添加10mL再折叠缓冲液并以3,600g离心10-30分钟,直至体积小于400μL。回收样品并添加额外的再折叠缓冲液至400μL的终体积。将样品转移至96孔深孔板,用膜密封并放在摇动器上4℃过夜。

[0571] 为了较大规模培养,将来自Ni-NTA的洗脱液加载至具有Ultracel-3或Ultracel-10膜(Millipore,目录号UFC900308或UFC901008,取决于AARS多肽的分子量)的Amicon Ultra-15离心过滤装置,然后以3,600g离心10-30分钟,直至体积减少至约500μL。对于pI>7的AARS多肽,将样品在以下缓冲液中稀释20倍:50mM乙酸钠、10mM氯化钠、0.4mM氯化钾、1mM EDTA、400mM蔗糖、500mM精氨酸、1mM DTT和0.01%聚山梨酯80,pH6.0。对于pI<7的AARS多肽,将样品在以下缓冲液中稀释20倍:50mM Tris、250mM氯化钠、10mM氯化钾、2mM氯化镁、2mM氯化钙、400mM蔗糖、500mM精氨酸、1mM DTT和0.01%聚山梨酯80,pH8.0。将样品在摇动



器上4℃下孵育过夜。

[0572] 脱盐和精制步骤:在过夜孵育之后,将96孔深孔板以3,600g离心以去除任何潜在的聚集物。然后使上清液经受1X PBS (Invitrogen, 目录号10010)的缓冲液交换。对于>10kDa的AARS多肽,AcroPrep96滤板的Omega 10K膜用20μL 1X PBS冲洗,并将板放入真空歧管(>10英寸汞柱),直至所有液体穿过。将再折叠缓冲液中的样品分配入每个孔并施加真空,直至所有液体穿过。重复这些步骤,直至处理完总样品体积(400μL)。通过向每个孔添加180μL 1X PBS pH 7.4来回收AARS多肽,小心吹吸10次,然后转移至洁净的深孔板。重复该步骤以产生每孔360μL的总体积,并将深孔板贮存在4℃。对于<10kDa的AARS多肽,将再折叠样品加载至具有Ultracel-3膜(Millipore, 目录号UFC900308)的Amicon Ultra-15离心过滤装置,随后添加10mL 1X PBS并以3,600g离心10-30分钟,直至体积小于360μL。回收样品并添加1X PBS至360μL的终体积。

[0573] 为了去除内毒素,将具有Mustang Q膜的AcroPrep Advance滤板(Pall, 目录号8171)用300μL 1X PBS冲洗并以1,000g离心5分钟以去除缓冲液。向滤板添加AARS多肽(360μL/孔)并在摇动器上孵育5-10分钟。然后将板以1,000g离心5-10分钟,收集含有AARS多肽的流经级分并在4℃贮存。

[0574] 为了较大规模培养,在过夜孵育之后,将再折叠样品以10,000g离心10分钟以去除任何不溶性聚集物。将上清液加载至Amicon Ultra-15离心过滤装置,并以3,600g离心,直至体积减少至250μL。将样品在10mL 1X PBS中混合并以3,600g再次离心10-30分钟,直至体积约250μL。注意,调节1X PBS的pH以符合再折叠缓冲液的pH,pH6.0或pH8.0。再重复一次该步骤,回收上清液并添加1X PBS至1.5mL的终体积。

[0575] 为了去除内毒素,用1mL 1X PBS冲洗Sartobind Q5强阴离子交换膜(Sartorius, 目录号Q5F),使用塑料注射器使AARS多肽缓慢穿过该膜。将含有AARS多肽的流经级分收集在96深孔板中,将该板密封并贮存在4℃。

[0576] 用于体内试验的蛋白制备

[0577] 细菌培养:将包含DNA编码各AARS多肽(如基因合成和克隆方法(3)中所述)的100ng的表达载体在42℃下转换到BL21(DE3)-RIPL(Agilent Technologies cat.#230280)成分大肠杆菌中30秒。将500μL的LB培养基加入细胞,并且以250rpm在37℃摇动器中孵育1小时,并且使150μl的转换反应物接种到氨比西林LB琼脂板上,并且在37℃下孵育过夜。

[0578] 将单独菌落挑出以在30ml的LB-Amp中开始种子培养,并且在250rpm下在37℃摇动器中孵育过夜。然后将种子培养物用于在6L锥形烧瓶中接种2.5L LB-Amp。在培养物达到稳定期(通常OD<sub>600</sub>是0.6-0.8)后,将烧瓶冰化30分钟,然后使用1M IPTG诱导至最终浓度200μM。然后将单独培养物在250rpm下在30℃摇动器中孵育过夜。

[0579] 蛋白分离:然后将培养物转移至500ml Nalgene瓶(Cat#3141-0500)中,并且在4℃下以8,000x g离心10分钟。小心吸出培养基并使小球在-20℃冷冻。

[0580] 然后使小球解冻,并重悬于具有50μlβ-ME和一种蛋白酶抑制剂表(Roche#11873580001)的50ml Ni-NTA pH8.0缓冲液中。加入300mg的溶菌酶(Sigma#L6878),并且使混合物在4℃下旋转30分钟。然后使重悬的小球以25,50和75%各自超声1分钟(开始10秒,关闭5秒)。然后使样品以35,000x g在4℃下旋转45分钟。

[0581] 亲和纯化:然后将上清加入2ml缓冲液平衡的Ni-NTA琼脂(Qiagen#30230),并且在

4℃下旋转1小时。然后使镍结合的蛋白混合物灌到通过Bio-RAD (Cat#737-4151)的缓冲液平衡的eco柱,然后使用具有0.5% Triton-X114 (Sigma#X114)的1L Ni-NTA缓冲液pH8.0 (50mM Tris, pH8, 300mM NaCl, 25mM咪唑)洗涤,然后使用100ml的不含内毒素的Ni-NTA缓冲液pH8.0洗涤。然后使用10ml的不含内毒素的洗脱缓冲液pH8.0 (50mM Tris, pH8, 300mM NaCl, 300mM咪唑)将纯化的蛋白从Ni-NTA琼脂上洗脱,并且在slide-a-lyzers (Pierce)中针对1X PBS pH7.4 (Invitrogen#10010)透析过夜,在后日相隔1小时改变两种缓冲液。

[0582] 浓缩和内毒素去除:使来自Ni-NTA的透析的洗脱物上载到具有Ultracel-10膜 (Millipore, cat.#UFC901008)的Amicon Ultra-15 Centrifugal Filter Unit上,然后以3,600x g离心10-30分钟,直到达到期望的浓度(通常1.7mg/ml)。

[0583] 为了除去内毒素,用1mL 1X PBS冲洗Sartobind Q15强阴离子交换膜 (Sartorius, cat.#Q15X),使用塑料注射器使AARS多肽缓慢穿过该膜。将含有AARS多肽的流经级分收集并等分。然后将AARS多肽在液氮中速冻,并在-80℃下贮存。

[0584] 生物物理表征:所有纯化的AARS多肽通过SDS-PAGE分析,基于A<sub>280</sub>测定其浓度并计算吸光系数(ProtParam on ExPASy server)。通过QCL-1000终点生色LAL测定 (Lonza, 目录号50-648U)根据生产商说明书测量内毒素水平。

[0585] 动态光散射:在实验之前使Wyatt Technology DynaPro99仪器和温度控制器 (20℃)升温15分钟,随后将Dynamics软件连接至仪器。将多次采集的采集时间设为10秒钟,并且激光功率设为100%。在添加蛋白样品 (15μL PBS中浓度约1mg/ml)之前,用去离子水和甲醇充分洗涤石英比色皿。在将比色皿以磨砂侧向左插入支持器之前,通过轻敲它而去除气泡。如果强度过高(屏幕上显示警告信息),将样品进一步用PBS稀释,直至强度降低至正常范围。收集的数据包括流体动力学半径、多分散性、预测的平均分子量、强度百分比和质量百分比。

[0586] 尺寸排阻色谱:在加载至General Electric AKTA FPLC上的100μL样品环中之前,将蛋白样品在PBS中稀释至约5-10mg/ml的浓度。使用Superdex 200 10/300 GL尺寸排阻柱 (General Electric, 目录号17-5175-01)来分离。首先用1.5柱体积(CV)的1X PBS缓冲液平衡柱,随后样品注射。将柱在1CV的1x PBS缓冲液(等梯度流动)中运行,在280nm监测吸光度。集成峰面积并用Unicorn软件计算百分比。基于与凝胶过滤校准试剂盒 (General Electric, 目录号28-4038-41和28-4038-42)的比较,使用洗脱体积来估计分子量。

[0587] 在高浓度贮存后的蛋白回收:将使用Amicon Ultra-15过滤装置 (Millipore, 目录号UFC901024或UFC900324,取决于分子量)浓缩至≥10mg/ml的10μL AARS多肽转移至洁净微量离心管。将样品在室温贮存1周,随后以16,000g离心10分钟以沉淀任何沉淀物。通过Bradford蛋白测定来测定上清液浓度并与在为期一周暴露于室温之前测量的浓度相比较。将回收表示为起始浓度的百分比。

[0588] 通过LC-MS表征AARS多肽:将纯化的AARS多肽 (1mg/ml)稀释1:10进入0.1%甲酸,并用Dionex自动进样器将0.6μg蛋白加载至C4毛细管柱上。通过切割150mm熔融二氧化硅管 (0.36mm OD×0.1mm ID, Polymicro Technologies, 目录号2000023)来制备毛细管柱。使用Suter Instrument激光纤维牵拉器在一端牵拉毛细管并用熔融二氧化硅切割器切割以产生5μm尖端。使用压弹,将毛细管用C4树脂 (5μm, 300Å, Michrom, 目录号PM5/64300/00)包裹至75mm的长度。在与Dionex Ultimate3000 HPLC系统偶联的ThermoFisher LTQ离子阱质谱

仪上进行LC-MS分析。使用35-分钟梯度的5-70%乙腈的0.1%甲酸溶液以0.9 $\mu$ L/min的流速从柱洗脱分析物。在全MS扫描模式(300-2,000m/z)以2.5kV的喷雾电压进行LTQ。

[0589] 数据收集和分析:通过在LTQ XL质谱仪上运行的XCalibur产生的原始文件中贮存原始质谱数据。色谱上主要峰的MS光谱用ThermoFisher去卷积算法ProMass进一步分析以获得AARS多肽分子量。

[0590] AARS多肽的功能性分析

[0591] 转录谱

[0592] 背景和治疗相关性:除了传统的靶鉴定技术,最近已经出现了作为帮助阐明AARS多肽作用机制的重要方法的基因组工具,并且可以在药物发现过程的早期提供对治疗相关性的直接了解。为了利于理解潜在的治疗功用,将原代人类细胞类型与AARS多肽一起培养,并且在与AARS多肽孵育之后的两个单独时间点评估转录分析。

[0593] 被选择用于转录分析的细胞类型是基于有关细胞的多能性能力以及鉴定AARS多肽直接治疗价值的潜力。例如,间充质干细胞(MSC)在暴露于特定刺激时可以分化成骨发生、脂肪生成、软骨形成、心肌或神经细胞系,使它们有潜力用于理解AARS多肽与宽范围细胞类型和疾病的潜在相关性。

[0594] 除了支持造血细胞,骨髓基质细胞还可以被诱导分化成不同结缔组织细胞系的细胞,例如骨、软骨和脂肪。人类间充质干细胞(hMSC)保持多潜能性以及体外广泛增殖的潜力开辟了恢复受损或患病组织的细胞疗法的新道路。最近报道还指示,hMSC的细胞命运能够跨越胚层界限。除了分化成中胚层的多细胞系,这些细胞还能分化成内胚层起源的神经元和内胚层起源的肝细胞样细胞。在分化过程中,这些细胞可以改变某些细胞系特定转录物的表达模式。

[0595] 因此,特定AARS多肽以时间依赖性方式调节hMSC中基因特定模式的能力证明这些蛋白在广泛的分化途径以及由于这些过程或相应细胞类型的功能障碍或恶化所导致的疾病和疾患中可能起着重要作用。而且,具有调节MSC中基因转录的能力的AARS多肽具有实现体外或体内调节血细胞生成、神经发生、肌发生、骨发生和脂肪生成以及在许多疾患和疾病中重要的治疗功用,所述疾患和疾病包括例如炎症反应、自身免疫性、癌症、神经变性、肌肉萎缩、骨质疏松和脂肪萎缩。

[0596] 人类骨骼肌细胞(HSkMC)能够经历分化以表现出肌动蛋白和肌球蛋白肌丝,并且已经用于研究遗传肌肉疾病,例如恶性高热<sup>1</sup>。HSkMC还具有用作心脏移植的潜力,修补心脏损伤。最近,培养的人类骨骼肌细胞已经用于微重力实验以研究低重力环境对人类骨骼肌的影响。

[0597] 因此,特定AARS多肽以时间依赖性方式调节HSkMC中基因特定模式的能力证明这些蛋白在肌发生过程以及由于这些过程以及肌细胞发育或代谢的功能障碍或恶化所导致的疾病和疾患中可能起着重要作用。因此,具有调节肌细胞中基因转录的能力的AARS多肽在许多疾病中具有治疗功用,例如治疗代谢疾病、恶病质、各种肌肉损耗病症以及骨骼肌疾病。

[0598] 方法:使用高通量微流体实时定量PCR(RT-qPCR)方法(Fluidigm Corporation)(参见Petriv等人,(2010)PNAS(10.1073/pnas.1009320107)评估AARS多肽在人类骨髓基质细胞(HMSC)和人类骨骼肌细胞(HSkMC)中调节基因表达的能力。在这里报道的实验中,

人类HskMC(目录号150-05f)和HMSC(目录号492-05f)购自Cell Applications。HMSC细胞以第二代冷冻保存,并且可以培养并繁殖至10次群体倍增。这里使用第6代HMSC。人类骨骼肌细胞(HskMC)以第二代冷冻保存,并且可以培养并繁殖至少15次群体倍增。这里报道的实验中使用了从正常人供体收集后的第6代HskMC。

[0599] 在这两种情况下,细胞以50000细胞/mL接种在100 $\mu$ L体积的生长培养基中并暴露于浓度为250nM或如下所示的AARS多肽,持续24小时和72小时。对照包括具有标准混合物的分化培养基,标准混合物促进(1)脂肪生成、(2)骨发生、(3)软骨形成和(4)骨骼肌管形成。另外的对照包括仅含有生长培养基的未处理细胞。为每个分化对照运行两个孔。对照:所有培养基使用DMEM作为基础培养基来制成。依据标准文献,并且分化培养基购自Cell Applications。根据供应商,分化培养基含有以下添加剂:骨骼肌分化混合物:FBS、胰岛素、谷氨酰胺、FGF、EGF;脂肪生成混合物:胰岛素、地塞米松和IBMX;骨发生混合物:FBS、地塞米松、抗坏血酸2磷酸、 $\beta$ -磷酸甘油;软骨形成混合物:胰岛素、抗坏血酸2磷酸和TGF- $\beta$ 1。

[0600] 利用ABI (Applied Biosystems, 品目号AM1728) **TAQMAN**<sup>®</sup> Gene Expression Cells-to-CT<sup>™</sup>试剂盒的标准方案来裂解细胞并收集基因组材料。使用ABI Pre-Amp Mix (Applied Biosystems, Item#4391128) 来起始前扩增。基因特异性引物使用Primer3程序产生并购自IDT technologies。使用Fluidigm分析阵列(品目号BMK-M-96.96)进行实际定量PCR,使用标准Fluidigm加样试剂和吸取装置,下表E1列出了被分析的基因。

表 E1  
转录分析中评估的基因列表

汇编的独特列表	refseq_nt	全称	同义
ABCA1	NM_005502	ATP 结合盒, 子族 A(ABC1), 成员 1	ABC-1 ABC1 CERP FLJ14958 HDLDT1  MGC164864 MGC165011 TGD
ACTB	NM_001101	肌动蛋白, $\beta$	PS1TP5BP1
ACTG1	NM_001614	肌动蛋白, $\gamma 1$	ACT ACTG DFNA20 DFNA26
ACVR2B	NM_001106	活化素 A 受体, IIB 型	ACTRIIB ActR-IIB MGC116908
APOA1	NM_000039	载脂蛋白 A-I	MGC117399
ARNT	NM_178427	芳基羟受体核转运蛋白	HIF-1 $\beta$  HIF1B HIF1B TANGO  bHLHe2
BAD	NM_032989	细胞死亡的 BCL2 相关激动剂	BBC2 BCL2L8
BCL2	NM_000657	B 细胞 CLL/淋巴瘤 2	Bcl-2
BMP2	NM_001200	骨形态发生蛋白 2	BMP2A
BMP4	NM_130851	骨形态发生蛋白 4	BMP2B BMP2B1 MCOPS6 OFC 11 ZYME
C3AR1	NM_004054	补体成分 3a 受体 1	AZ3B C3AR HNFAG09
CASP3	NM_032991	胱天蛋白酶 3, 凋亡相关的半胱氨酸蛋白酶	CPP32 CPP32B SCA-1
CAV1	NM_001753	小窝蛋白 1, 小凹蛋白, 22kDa	BSCL3 CGL3 MSTP085 VIP21
CDH5	NM_001795	钙粘着蛋白 5, 2 型(血管内皮)	7B4 CD144 FLJ17376
CFLAR	NM_003879	CASP8 和 FADD 样凋亡调节剂	CASH CASP8AP1 CLARP Casper r  FLAME FLAME-1 FLAME1 FLI P I-FLICE  MRIT  c-FLIP c-FLIPL c-FLIPR c-FLIPS
COMP	NM_000095	软骨寡聚体基质蛋白	EDM1 EPD1 MED MGC131819  MGC149768  PSACH THBS5

[0601]

表 E1  
转录分析中评估的基因列表

汇编的独特列表	refseq_nt	全称	同义
CSF1	NM_172212	集落刺激因子 1(巨噬细胞)	MCSF MGC31930
CTGF	NM_001901	结缔组织生长因子	CCN2 HCS24 IGFBP8 MGC102839 NOV2
CTNNB1	NM_001904	连环蛋白(钙粘着蛋白相关蛋白), $\beta$ 1, 88kDa	CTNNB DKFZp686D02253 FLJ25606 FLJ37923
DAAM1	NM_014992	形态发生的蓬乱相关激活剂 1	FLJ41657 KIAA0666
ELN	NM_001081755	弹性蛋白	FLJ38671 FLJ43523 SVAS WBS WS
ENO1	NM_001428	烯醇化酶 1, $\alpha$ )	ENO1L1 MPB1 NNE PPH
FABP3	NM_004102	脂肪酸结合蛋白 3, 肌肉和心脏(乳腺衍生的生长抑制剂)	FABP11 H-FABP MDGI O-FABP
FAK	NM_001199649	粘着斑激酶	fak1
FGF4	NM_002007	成纤维细胞生长因子 4	HBGF-4 HST HST-1 HSTF1 K-FGF KFGF
FIGF	NM_004469	c-fos 诱导的生长因子(血管内皮生长因子 D)	VEGF-D VEGFD
FLT1	NM_002019	fms 相关的酪氨酸激酶 1(血管内皮生长因子/血管通透因子受体)	FLT VEGFR1
FOXA1	NM_004496	叉头盒 A1	HNF3A MGC33105 TCF3A
GAPDH	NM_002046	甘油醛-3-磷酸脱氢酶	G3PD GAPD MGC88685
GFAP	NM_002055	胶质纤维酸性蛋白	FLJ45472
SLC2A4	NM_001042	溶质载体家族 2(促葡萄糖转运蛋白), 成员 4	GLUT4
HAND1	NM_004821	心脏和神经脊衍生物表达的 1	Hxt Thing1 bHLHa27 eHand
HIF1A	NM_181054	缺氧诱导因子 1, $\alpha$ 亚基(基本的螺旋-环-螺旋转录因子)	HIF-1 $\alpha$  HIF1 HIF1-A MOP1 PASD8 bHLHe78
HK2	NM_000189	己糖激酶 2	DKFZp686M1669 HKII HXK2
HMGB1	NM_002128	高迁移组框 1	DKFZp686A04236 HMG1 HMG3 SBP-1
HNF4A	NM_178850	肝细胞核因子 4, $\alpha$	FLJ39654 HNF4 HNF4a7 HNF4a8 HNF4a9 HNF4 $\alpha$  MODY MODY1 NR2A1 NR2A21 TCF TCF14
HPRT1	NM_000194	次黄嘌呤磷酸核糖转移酶 1	HGPRT HPRT
HSPB1	NM_001540	热激 27kDa 蛋白 1	CMT2F DKFZp586P1322 HMN2B HS.76067 HSP27 HSP28 Hsp25 SRP27
ICAM1	NM_000201	细胞间黏附分子 1	BB2 CD54 P3.58
IFNG	NM_000619	干扰素, $\gamma$	IFG IFI

[0602]

表 E1  
转录分析中评估的基因列表

汇编的独特列表	refseq_nt	全称	同义
IGF1	NM_001111285	胰岛素样生长因子 1(生长调节素 C)	IGF-I IGF1A IGFI
IGF2	NM_001127598	胰岛素样生长因子 2(生长调节素 A)	C11orf43 FLJ22066 FLJ44734 INSIGF pp9974
IGFBP3	NM_001013398	胰岛素样生长因子结合蛋白 3	BP-53 IBP3
IGFBP5	NM_000599	胰岛素样生长因子结合蛋白 5	IBP5
IKKB	NM_001556	B 细胞中 $\kappa$ 轻链多肽基因增强子的抑制剂, 激酶 $\beta$	FLJ33771 FLJ36218 FLJ38368 FLJ40509 IKK- $\beta$  IKK2 IKKB MGC131801 NFKBIKB
IL10	NM_000572	白介素 10	CSIF IL-10 IL10A MGC126450 MGC126451 TGIF
IL1B	NM_000576	白介素 1, $\beta$	IL-1 IL1-B IL1F2
IL3	NM_000588	白介素 3(集落刺激因子, 多样的)	IL-3 MCGF MGC79398 MGC79399 MULTI-CSF
IL4	NM_172348	白介素 4	BCGF-1 BCGF1 BSF1 IL-4 MGC79402
IL5	NM_000879	白介素 5(集落刺激因子, 嗜伊红性粒细胞)	EDF IL-5 TRF
IL6R	NM_181359	白介素 6 受体	CD126 IL-6R-1 IL-6R- $\alpha$  IL6RA MGC104991
IL8	NM_000584	白介素 8	CXCL8 GCP-1 GCP1 LECT LUCT LYNAP MDNCF MONAP NAF NAP-1 NAP1
ITGA5	NM_002205	整联蛋白, $\alpha$ 5(纤连蛋白受体, $\alpha$ 多肽)	CD49e FNRA VLA5A
KDR	NM_002253	激酶插入域受体(III 型受体酪氨酸激酶)	CD309 FLK1 VEGFR VEGFR2
LEP	NM_000230	瘦素	FLJ94114 OB OBS
LPL	NM_000237	脂蛋白脂肪酶	HDLCQ11 LIPD
MAPK11	NM_002751	丝裂原活化蛋白激酶 11	P38B P38B2 PRKM11 SAPK2 SAPK2B p38-2 p38B
MMP1	NM_002421	基质金属肽酶 1(间质胶原酶)	CLG CLGN
MMP3	NM_002422	基质金属肽酶 3(基质溶素 1, 明胶酶原)	CHDS6 MGC126102 MGC126103 MGC126104 MMP-3 SL-1 STMY STMY1 STR1
MYH1	NM_005963	肌球蛋白, 重链 1, 骨骼肌, 成人	MGC133384 MYHSA1 MYHa MyHC-2X/D MyHC-2x

[0603]

表 E1  
转录分析中评估的基因列表

汇编的独特列表	refseq_nt	全称	同义
MYH11	NM_022844	肌球蛋白, 重链 11, 平滑肌	AAT4 DKFZp686D10126 DKFZp686D19237  FAA4 FLJ35232 MGC126726 MGC32963 SMHC SMMHC
MYH7	NM_000257	肌球蛋白, 重链 7, 心肌, $\beta$	CMD1S CMH1 DKFZp451F047 MGC138376  MGC138378 MPD1 MYHCB SPMD SPMM
MYOD1	NM_002478	肌源性分化 1	MYF3 MYOD PUM bHLHc1
NFATC1	NM_172390	活化的 T 细胞的核因子, 细胞质的, 钙调磷酸酶依赖性的 1	MGC138448 NF-ATC NFAT2 NFATc
NFATC2	NM_173091	活化的 T 细胞的核因子, 细胞质的, 钙调磷酸酶依赖性的 2	NFAT1 NFATP
NFKB1	NM_003998	B 细胞中 $\kappa$ 轻链多肽基因增强子的核因子 1	DKFZp686C01211 EBP-1 KBF1 MGC54151  NF- $\kappa$ -B NF- $\kappa$ B NFKB-p105 NFKB-p50 p105 p50
NOS2	NM_000625	一氧化氮合酶 2, 诱导型	HEP-NOS INOS NOS NOS2A
NOTCH1	NM_017617	notch 1	TAN1 hN1
NR3C1	NM_001024094	核受体 子族 3, 组 C, 成员 1(糖皮质激素受体)	GCCR GCR GR GRL
NRP2	NM_201279	神经毡蛋白 2	MGC126574 NP2 NPN2 PRO2714 VEGF165R2
PAX7	NM_013945	配对盒 7	FLJ37460 HUP1 PAX7B RMS2
PDGFB	NM_033016	血小板衍生的生长因子 $\beta$ 多肽 (猿肉瘤病毒(v-sis)癌基因同源物)	FLJ12858 PDGF2 SIS SSV c-sis
PDK4	NM_002612	丙酮酸脱氢酶激酶, 同工酶 4	FLJ40832
PLA2G1B	NM_000928	磷脂酶 A2, 组 IB(胰腺)	MGC119834 MGC119835 PLA2  PLA2A  PPLA2
PLIN1	NM_002666	脂滴相关蛋白	围脂滴蛋白
PPARG	NM_138712	过氧化物酶体增殖剂活化受体 $\gamma$	CIMT1 GLM1 NR1C3 PPARG1 PPARG2  PPAR $\gamma$
QARS	NM_005051	谷氨酰胺酰基-tRNA 合成酶	GLNRS PRO2195
RHOA	NM_001664	Ras 同源基因家族, 成员 A	ARH12 ARHA RHO12 RHOH12
RUNX1	NM_001754	Runt 相关的转录因子 1	AML1 AML1-EVI-1 AMLCR1 CBFA2 EVI-1 PEBP2aB

[0604]



表 E1  
转录分析中评估的基因列表

汇编的独特列表	refseq_nt	全称	同义
RXRA	NM_002957	维甲酸 X 受体, $\alpha$	FLJ00280 FLJ00318 FLJ16020 FLJ16733  MGC102720 NR2B1
SERPINE1	NM_001165413	丝氨酸蛋白酶抑制剂肽酶抑制剂, 分支 E(连接蛋白, 纤溶酶原激活抑制剂 1 型), 成员 1	PAI PAI-1 PAI1 PLANH1
SMAD2	NM_005901	SMAD 家族成员 2	JV18 JV18-1 MADH2 MADR2 MGC22139  MGC34440 hMAD-2 hSMAD2
SMAD4	NM_005359	SMAD 家族成员 4	DPC4 JIP MADH4
TERT	NM_198255	端粒酶逆转录酶	EST2 TCS1 TP2 TRT hEST2
TGFB1	NM_000660	转化生长因子, $\beta$ 1	CED DPD1 LAP TGFB TGF $\beta$
TGFB3	NM_003239	转化生长因子, $\beta$ 3	ARVD FLJ16571 TGF- $\beta$ 3
THBS4	NM_003248	血小板反应蛋白 4	TSP4
TNF	NM_000594	肿瘤坏死因子	DIF TNF- $\alpha$  TNFA TNFSF2
TUBB	NM_178014	微管蛋白, $\beta$	M40 MGC117247 MGC16435 OK/SW-cl.56 TUBB1  TUBB5
TUBB1	NM_030773	微管蛋白, $\beta$ 1	微管蛋白同种型 $\beta$ (1)
TUBG1	NM_001070	微管蛋白, $\gamma$ 1	GCP-1 TUBG TUBGCP1
VCAM1	NM_080682	血管细胞黏附分子 1	CD106 DKFZp779G2333 INCAM-100 MGC99561
VEGFA	NM_003376	血管内皮生长因子 A	MGC70609 MVCD1 VEGF VPF
VIM	NM_003380	波形蛋白	FLJ36605
WISP1	NM_080838	WNT1 诱导型信号传导通路蛋白 1	CCN4 WISP1c WISP1i WISP1tc
WNT1	NM_005430	无翅型 MMTV 整合位点家族, 成员 1	INT1

[0605]

[0606] 生物信息学分析: 将以.csv格式从Fluidigm的Biomark机器检索的数据转化成表格形式, 包含样品、mRNA和复制信息以及原始荧光数据。失败的PCR反应被标记为缺失。在标准化为mRNA物类的总表达之后, 组合多个实验。基于所有测试的生物复制物的至少2个的检测要求, 过滤所有测量的mRNA表达。我们评估了整个数据集中的技术、生物和设定偏差平均值。

[0607] 为了数据分析, 首先将所有感兴趣基因的Ct值标准化为相应样品的管家基因的平均Ct值, 以获得 $\Delta$ Ct值 ( $\Delta$ Ct=Ct基因-Ct平均管家基因)。然后将每个样品的基因标准化为未处理对照中的相同基因, 以获得 $\Delta\Delta$ Ct值 ( $\Delta\Delta$ Ct= $\Delta$ Ct对照样品- $\Delta$ Ct测试样品)。

[0608] 为了获得倍变值, 对上调的基因(即,  $\Delta\Delta$ Ct大于0)进行以下计算: 倍变= $2^{\Delta\Delta$ Ct}。对于下调基因(即,  $\Delta\Delta$ Ct小于0): 倍变= $-(2^{-|\Delta\Delta$ Ct|})。

[0609] 细胞增殖测定(以下数据表中的测定A1-A11)

[0610] 背景和/治疗相关性: 调节不同细胞类型的细胞增殖和凋亡速率的能力是许多治疗化合物的基本性质, 并且与许多疾病和疾患的治疗和预防直接相关。

[0611] 因此,具有调节细胞增殖和或凋亡速率的能力的AARS多肽在许多疾病中具有显著的治疗功用,包括作为干细胞的生长因子和分化因子,以及在治疗方案中增强或抑制感兴趣的特定细胞类型(包括例如造血细胞、免疫调节细胞、癌症)在体内或体外的增殖,以及用于治疗 and 预防与衰老相关的疾病,包括例如神经变性、外周神经病变和肌肉及软组织张力的损失。

[0612] 方法:使用下列方法的一种或多种评估AARS多肽对细胞增殖的影响,如在以下方法中更具体阐述的。

[0613] Hoechst 33432.使用Hoechst 33432进行评估增殖的标准细胞计数,Hoechst 33432是细胞透性的核复染剂,当结合dsDNA时发射蓝色荧光。它可作为溶液获得(Invitrogen目录号H-3570),在培养基或PBS中以1 $\mu$ g/mL的终浓度使用。使细胞在AARS多肽存在下在96孔板中生长48小时的标准生长时间,或者根据细胞类型而生长更长时间,如以下实施例中所描述的。

[0614] ATP-lite.细胞ATP水平与细胞健康相关,并且可以使用许多商业途径可获得的试剂盒容易地测定。ATP-lite(Perkin-Elmer,目录号6016947 Boston,MA 02481),它是裂解溶液和ATP检测试剂的均质混合物,在使用前预先混合并且与培养细胞以1:1v:v比率使用。将板孵育5分钟以促进裂解并使用发光板读数器测量板。使细胞在AARS多肽存在下在96孔板中生长48小时的标准生长时间,或者根据细胞类型而生长更长时间,如以下实施例中所描述的。

[0615] **ALAMARBLUE®**(刃天青)是基于细胞氧化还原状态的细胞活力指示剂。活性成分刃天青是无毒的细胞透性化合物,当以其氧化形式存在时颜色为蓝色并且几乎无荧光。然而,当进入正常活力细胞时,刃天青快速还原成试卤灵,试卤灵产生红色荧光信号。活力细胞连续转化刃天青为试卤灵,从而产生活力和细胞毒性的定量量度。毒性缺乏允许细胞长期暴露于刃天青而没有负面影响;如通过流式细胞术分析测定的,发现在刃天青存在下生长的细胞产生与对照细胞类似数量的活细胞。

[0616] 通过向细胞添加刃天青/**ALAMARBLUE®**溶液来进行测量,将它们孵育1-4小时,并读取荧光或吸光度。荧光或吸光度的量与活细胞数量成比例,并且对应于细胞代谢活性。受损和非活力细胞具有较低的先天代谢活性并因此产生比健康细胞成比例降低的信号。在与**ALAMARBLUE®**孵育之后,可以在荧光和吸光度仪器上容易地测量样品。对于荧光读数:使用530nm激发和590nm发射滤波器设置。

[0617] 使细胞在AARS多肽存在下在96孔板中生长48小时的标准生长时间,或者根据细胞类型而生长更长时间,如以下实施例中所描述的。

[0618] HepG2C3a人类肝细胞中的乙酰化LDL摄取(以下数据表中的测定B1)

[0619] 背景和治疗相关性:LDL是血液中胆固醇的主要载体,负责血浆中超过60%的胆固醇。在人类中,肝脏LDL受体负责清除循环中大约70%的血浆LDL。内化的LDL在溶酶体中降解为游离胆固醇和氨基酸。肝脏是人类有关LDL分解代谢和LDL受体活性的最重要器官。未内化并保持在循环中的LDL可以被内皮细胞转运进入血管壁,导致动脉粥样硬化斑块的形成。循环LDL也可被巨噬细胞摄取,并且这也促进斑块形成。认为增加LDL摄取进入肝脏组织对于人类健康是有益的,并且发现可以正向调节该过程的安全有效治疗剂可以提供心血管

和代谢疾病的新疗法。为了考察AARS多肽的独特性质是否可以调节乙酰化LDL的摄取,在HepG2C3a细胞中采用了测量乙酰化LDL摄取的标准测定。

[0620] 因此,具有调节LDL摄取的能力的AARS多肽在许多疾病中具有重要的治疗功用,包括例如治疗高胆固醇血症、高血脂、1型和2型糖尿病、代谢综合征和血管疾病,包括动脉硬化。

[0621] 方法:HEPG2C3a细胞(ATCC#CRL-10741)维持在添加了10%FBS(HyClone目录号SH30910.03)、50u/mL青霉素/50μg/mL链霉素的(Invitrogen)Eagles最低必需(EMEM)培养基中,在75mL烧瓶中15mL培养基。将细胞在37°C,5%CO<sub>2</sub>加湿环境中生长,并且在BSL2认证的组织培养罩中使用,使用了无菌技术和适当的个人防护设备,包括护目镜、手套和实验服。当在清澈见底的胶原涂布板上生长时,HEPG2C3a表达LDL受体并且有能力进行乙酰化LDL摄取。将100μL体积细胞铺板于胶原涂布板(Invitrogen目录号A11428),以50,000细胞/mL的细胞密度在完全培养基(上述)中过夜。细胞用PBS(Invitrogen目录号10010)洗涤一次,向每个孔添加80μL的无血清EMEM。向每个孔中以250nM每孔的终浓度添加AARS多肽,以无菌PBS中一致的体积。将独特的AARS多肽放入每个孔。使细胞血清饥饿并暴露于AARS多肽16小时。在16小时孵育之后,收集上清液并使用RND Systems的标准ELISA试剂盒(目录号DY643)测量可溶性ICAM,向每个孔添加补充了5μg/mL ac-LDL(Alexa Fluor 488标记的,目录号L23380,Invitrogen)的无血清培养基。在37°C5%CO<sub>2</sub>下孵育2小时之后,用无菌PBS洗涤细胞两次,之后向每孔添加100μL PBS来定量。使用在Victor X5荧光板读数器(Perkin Elmer)上底部读数来分析板的总荧光强度,激发波长中心在大约485nm,发射波长中心在大约535nm。用Hoechst染料对细胞染色,读取荧光强度(405nm激发/450nm发射)以确认整个板的总细胞数是一致的。

[0622] 人类嗜中性粒细胞氧化猝发和弹性蛋白酶生成的调节(以下数据表中的测定C1-C3)

[0623] 嗜中性粒细胞氧化猝发

[0624] 背景和治疗相关性:多形核嗜中性粒细胞和单核细胞的噬菌作用构成了宿主防御微生物(包括细菌和真菌)感染的必要方面。吞噬过程可以被分为几个主要阶段:趋化(噬菌细胞迁移至炎症部位),颗粒粘附至噬菌细胞的细胞表面,通过依赖氧(氧化猝发)和不依赖氧的机制的摄取(吞噬)和细胞内杀伤。在慢性肉芽肿病(CGD)的天生缺陷中观察到减少或缺失的猝发活性。CGD是通常在生命前两年自己显现的一组异质性的遗传性疾患。该疾病的特征是由细菌和真菌生物体引起的反复且威胁生命的感染。这些感染通常由肺炎、淋巴腺炎或涉及淋巴结、肺和肝的脓肿组成。NADPH氧化酶是负责产生过氧化物阴离子的酶系统,过氧化物阴离子被快速转化为过氧化氢和羟基自由基。NADPH氧化酶系统的组成肽的异常导致CGD的功能障碍特征。来自CGD患者的嗜中性粒细胞不能在刺激之后产生明显的氧化猝发。描述了不同形式的CGD(经典的X连接CGD和常染色体隐性模式)。粒细胞的氧化猝发在移植、HIV感染后期和老年人中受损,使得这些群体更易受继发性感染和炎症疾病的恶化。各种免疫调节剂(例如,细胞因子(GM-CSF、G-CSF、TNF)或药物)也似乎对氧化猝发具有影响。具有以治疗方式上调或下调氧化猝发的能力的蛋白存在对于各种不同疾病状态有用的潜力。

[0625] 方法:作为氧化猝发过程的激动剂,蛋白激酶C配体佛波醇12-豆蔻酸酯13-乙酸酯

(PMA) 可用于该测定。肝素化的全血与无菌葡聚糖(0.6%终浓度)混合1小时并允许分层。下层含有嗜中性粒细胞、单核细胞和红细胞。使用氯化铵裂解步骤去除所有RBC,在裂解步骤之后剩下具有大约3%单核细胞污染的97%纯的嗜中性粒细胞群体。刺激时,粒细胞和单核细胞产生反应性氧代谢物(过氧化物阴离子、过氧化氢和次氯酸),其破坏吞噬体内的细菌。可以通过Amplex Red的添加和氧化来监测氧化猝发期间反应性氧化剂的形成。然后分析产生了反应性氧自由基的细胞的百分比,并使用荧光板读数器分析其平均荧光强度。该反应的典型时程是10分钟,在2分钟时看到明显的猝发,在20分钟时看到信号减弱。该测定可以在缺乏PMA下以激动模式运行,或者在共同施用浓度低于该化合物EC50的AARS多肽和PMA下以拮抗模式运行。

[0626] 人类嗜中性粒细胞弹性蛋白酶生成的调节

[0627] 背景和治疗相关性:嗜中性粒细胞弹性蛋白酶是丝氨酸蛋白酶,已经暗示其在许多人类疾病发展中具有特定作用,包括肺和心血管系统的炎性疾病。尽管其主要的生理作用在于天生的宿主防御,但是它也可以参与组织重构,并且具有目前公认为对局部炎症信号重要的促分泌素作用。数十年来,将嗜中性粒细胞弹性蛋白酶活性牵涉于肺气肿的发展,然而,只有近期将致病功能归因于该丝氨酸蛋白酶,这种情况下发生过量的细胞外基质沉积。遗传操纵的动物模型的使用开始揭示其作用可能影响肺纤维化修复的可能途径。新出现的证据表明,对纤维发生调节剂生成和胶原合成具有更直接影响的细胞途径的参与,似乎支持嗜中性粒细胞弹性蛋白酶促进肺基质累积的作用。人类嗜中性粒细胞弹性蛋白酶还存在于动脉粥样硬化斑块中,它在那里促进基质降解和与动脉瘤形成和斑块破裂合并症相关的血管壁的弱化。这些作用中还涉及其他细胞外蛋白酶,但是该酶宽的底物范围和效价以及嗜中性粒细胞脱粒相关活性使该破坏性蛋白酶被选作动脉粥样硬化疾病的治疗靶。

[0628] 方法:该测定使用ENZCHEK®弹性蛋白酶测定试剂盒(Invitrogen目录号E-12056)。使用6%葡聚糖溶液从新鲜人类血液制备嗜中性粒细胞,在将细胞铺板于RPMI培养基(培养基应该是无添加的,没有血清,没有抗生素)之前裂解红细胞。通过向含有冻干底物的三个管形瓶之一直接添加1.0mL去离子水(dH<sub>2</sub>O)并混合溶解,来制备DQ弹性蛋白底物的1.0mg/mL贮液。通过在54mL dH<sub>2</sub>O中稀释6mL 10X反应缓冲液来制备1X反应缓冲液。通过在1X反应缓冲液中十倍稀释DQ弹性蛋白贮液来制备DQ弹性蛋白的100μg/mL工作溶液。通过在dH<sub>2</sub>O中制备100U/mL贮液来制备猪胰弹性蛋白酶贮液。为测定弹性蛋白酶活性,将50μL 1X反应缓冲液吸入含有30μL体积500,000嗜中性粒细胞/mL的每个测定孔中。每孔添加8μL的每种AARS多肽,在37°C下孵育样品20分钟。向每孔添加50μL 100μg/mL DQ弹性蛋白工作溶液并混合。在室温下孵育样品,避光30分钟。可以在多个时间点测量配有标准荧光素滤器(ex 485/Em 535)荧光的荧光微量板读数器中的荧光强度。

[0629] 与Toll样受体的结合以及NFκB的激活(以下数据表中的测定D1-D4)

[0630] 背景和治疗相关性:巨噬细胞是天生免疫系统的主要参与者,并且表达大量不同类别的模式识别受体(PRR),包括Toll样受体(TLR)家族,其是免疫应答的有力调节剂和控制剂。

[0631] 微生物病原体和内源性配体对TLR的刺激起始诱导促炎细胞因子和指导下游适应性免疫反应的效应细胞因子分泌的信号传导级联。内源性配体以及微生物成分被TLR识别并且可以激活TLR,提高了这些受体可能成为开发针对多种疾病的新疗法的关键靶标的可

能性。

[0632] 因此,调节TLR受体活性的AARS多肽在许多疾病和疾患中具有治疗功用,包括例如炎症性疾病和疾患、自身免疫疾病、组织移植/器官排斥,癌症预防或治疗,调节血细胞生成和感染。

[0633] RAW-BLUE细胞中TLR激活的测量

[0634] 以商标RAW-BLUE™细胞 (Invivogen,产品代码:raw-sp) 销售的小鼠巨噬细胞表达除了TLR5之外的所有TLR,并且包括可被NF- $\kappa$ B和AP-1转录因子诱导的分泌胚胎碱性磷酸酶(SEAP)基因。在TLR刺激时,RAW-BLUE™细胞激活NF- $\kappa$ B和/或AP-1,导致使用SEAP检测培养基可测量的SEAP分泌。

[0635] 方法:RAW-BLUE™细胞用PBS洗涤两次,胰蛋白酶化,并重悬于新鲜生长培养基(生长培养基:DMEM,4.5g/1葡萄糖,10%热失活胎牛血清(56℃下30分钟),100mg/mL ZEOCIN™,2mM L-谷氨酰胺)。以100 $\mu$ L总体积在96孔板中以50,000细胞/孔的浓度铺板,以下文描述的实验中显示的浓度向每孔添加AARS多肽、对照或AARS多肽(+LPS)。在37℃、5%CO<sub>2</sub>培养箱中孵育细胞18小时。在实验第2天,根据说明书制备SEAP检测培养基(QUANTI-BLUE™) (Invivogen产品代码:rep-qb1),向透明平底96孔板的每孔添加120 $\mu$ L,添加细胞上清液(20 $\mu$ L)。在37℃孵育样品约30分钟至2小时。使用分光光度计测定SEAP水平并在650nm读取吸光度。

[0636] 为了检测特异性阻止TLR激活的AARS多肽,可以调整该测定以鉴定潜在的TLR拮抗剂。在这种情况下,在添加50ng/mL LPS之前1小时,以约250nM每孔(或者以下实施例中另外指定的)终浓度向细胞添加AARS多肽。如上所述孵育细胞并检测SEAP。使用没有LPS或AARS多肽单独添加的PBS对照孔来确定测量时TLR刺激的基础水平。对照孔用PBS和已知的TLR激动剂和拮抗剂预处理。使用减去背景的[PBS+LPS信号]与[AARS多肽+LPS信号]的比率来确定百分比拮抗。

[0637] 在Hek293细胞中筛选人类TLR

[0638] 人类HEK293细胞是遗传修饰的并且以商标HEK-Blue™TLR细胞 (Invivogen) 销售。该细胞类型的TLR2和TLR4形式选择性表达所有TLR2或TLR4,并且包括在与五个NF- $\kappa$ B和AP-1转录因子结合位点融合的IFN- $\beta$ 最小启动子控制下的分泌胚胎碱性磷酸酶(SEAP)报告基因。使用特异性TLR2或4激动剂(分别地),Hek-Blue™TLR2和Hek-Blue™TLR4细胞激活NF- $\kappa$ B和/或AP-1,导致使用SEAP检测试剂可测量的SEAP分泌。Hek-Blue™TLR2细胞用LPS共受体蛋白CD14共转染以增强TLR2响应性并改善信号质量。亲本细胞表达内源性水平的TLR1、3、5、6和NOD1。

[0639] 方法:Hek-Blue™-TLR2或Hek-Blue™-TLR4细胞用PBS洗涤两次,胰蛋白酶化,并重悬于新鲜生长培养基(生长培养基:DMEM,4.5g/1葡萄糖,10%热失活胎牛血清(56℃下30分钟),100mg/mL ZEOCIN™,2mM L-谷氨酰胺)。以100 $\mu$ L总体积在96孔板中以50,000细胞/孔的浓度铺板,以下文描述的实验中显示的浓度向每孔添加AARS多肽、对照或AARS多肽(+LPS)。在37℃、5%CO<sub>2</sub>培养箱中孵育细胞18小时。在实验第2天,根据说明书制备SEAP检测培养基(QUANTI-BLUE™) (Invivogen产品代码:rep-qb1),向透明平底96孔板的每孔添加120 $\mu$ L,添加细胞上清液(20 $\mu$ L)。在37℃孵育样品约30分钟至2小时。使用分光光度计测定SEAP水平并在650nm读取吸光度。对照孔用PBS和已知的TLR激动剂例如UltraPure LPS (TLR-4) 或

PAM3CSK4 (TLR-2) 预处理。使用减去背景的 [PBS+LPS信号] 与 [AARS多肽+LPS信号] 的比率来确定百分比激动。

[0640] 细胞因子释放 (以下数据表中的测定E1-E16)

[0641] 背景和治疗相关性: 细胞因子是一类多样的小细胞信号传导蛋白分子, 其广泛用于细胞间通信, 并且在正常体内稳态中起重要作用, 包括免疫调节和调控。因此, 调节细胞因子释放或生物活性的AARS多肽在许多疾病和疾患中具有治疗功用, 包括例如炎症性疾病和疾患、自身免疫疾病、组织移植/器官排斥, 癌症预防或治疗, 调节血细胞生成和感染。

[0642] 细胞因子从培养细胞的释放

[0643] 方法: 将测试细胞以约1百万细胞/孔的密度在1mL生长培养基中接种于24孔板。细胞用任一种AARS多肽 (下文实施例中所示的浓度) 或等体积PBS处理, 并在37°、5%CO<sub>2</sub> 孵育过夜。细胞处理之后, 在4°C水平离心机中以2,000x g离心样品5分钟。小心取出培养基以不扰动细胞沉淀并转移至新管。立即测定样品或在液氮中快速冷冻以后续分析。使用商业途径可获得的试剂盒 (R&D Systems, Inc, MN, USA) 或者经由合同研究组织 (MD Biosciences (St. Paul, MN) 测定细胞因子释放 (包括细胞因子MIF、IL-8、IL-10、丝氨酸蛋白酶抑制剂E1、GM-CSF、GRO、IL-1 $\alpha$ 、IL-1 $\beta$ 、IL-1ra、IL-6、MCP-1、MIP-1、RANTES和TNF- $\alpha$ )。

[0644] 细胞因子从人类全血的释放

[0645] 方法: 从正常人供体获得人类全血并收集在含有肝素的标准收集管中。血液在收集当日使用, 以确保充分的细胞健康。小心混合血液并以100 $\mu$ L体积铺板入96孔聚碳酸酯V底板。添加AARS多肽并使用设为50 $\mu$ L的多通道移液器缓慢混入血液2X。滤头用于所有实验, 并且穿戴全套PPE。所有实验在适合人类血液实验的专用生物安全罩中进行。血液在37°C、5%CO<sub>2</sub> 孵育过夜。细胞处理之后, 在水平离心机中以2,000x g离心样品5分钟。收集上清液进行细胞因子ELISA。ELISA如前所述进行。

[0646] 细胞因子从PBMC的释放

[0647] 方法: 为了分离外周血单核细胞, 在室温下使新鲜分离的人类全血在50mL锥形管中以1:1的比率在Sigma HISTOPAQUE®-1077上轻轻分层。使分层的样品在室温下在临床水平离心机中以400x g连续离心30分钟。然后通过移液器取出在血浆和密度梯度之间的界面处的白色细胞层。通过稀释并以250x g离心10分钟, 将这些外周血单核细胞用RPMI-1640 (Invitrogen#22400-105) 洗涤两次。将洗涤的PBMC重悬于RPMI-1640+10%FBS并以1x10<sup>6</sup>细胞/mL铺板。

[0648] 细胞因子从人类滑膜细胞的释放

[0649] 背景和治疗相关性: 许多研究已经证明IL-6和IL-8在几种疾病中过量生成, 因此可能在炎症疾病病理中发挥基础作用。IL-6激活内皮细胞生成, 导致IL-8和单核细胞趋化蛋白的释放、粘附分子的表达和白细胞募集到炎症部位。这些细胞因子在与炎症疾病相关的细胞类型中表达, 所述细胞类型包括系统性青少年关节炎、系统性红斑狼疮、克罗恩病和类风湿性关节炎发病中牵涉的细胞。细胞因子生成的最重要的系统作用之一是诱导急相反应。急相蛋白主要由肝脏生成, 并且包括通过激活补体、诱导促炎细胞因子和刺激嗜中性粒细胞趋化而促进免疫应答的蛋白。或者, 急相反应可以是有帮助的, 并且急相蛋白例如蛋白酶拮抗剂、调理素和促凝血剂通过解决炎症来帮助限制组织破坏。具体地, IL-6可以刺激滑膜细胞增殖和破骨细胞活化, 导致滑膜血管翳形成和修复。IL-6与IL-1作用以增加基质金

属蛋白酶的生成,这可以促进关节和软骨破坏。然而,IL-6还可以具有保护关节的作用,如以下发现所表明的:当注射入具有抗原诱导的关节炎的小鼠关节中时,该细胞因子诱导金属蛋白酶的抑制剂的表达并刺激蛋白聚糖合成。人类成纤维细胞样滑膜细胞类风湿性关节炎(HFLS-RA)分离自从患有类风湿性关节炎(RA)的患者获得的滑膜组织。它们在第二代被冻存,并且可以培养并繁殖至少5次群体倍增。HFLS因其通过产生促进软骨降解的细胞因子和金属蛋白酶而破坏关节的作用而久负盛名。

[0650] 因此,具有调节成纤维细胞样滑膜细胞-类风湿性关节炎(HFLS-RA)的生长、分化或细胞因子释放特征的能力的AARS多肽在许多疾病中具有治疗功用,所述疾病包括系统性青少年关节炎、系统性红斑狼疮、克罗恩病和类风湿性关节炎。

[0651] 方法:HFLS-RA,成年细胞(Cell Applications目录号408RA-05a)在使用之前维持在滑膜细胞生长培养基(Cell Applications目录号415-50)(125mL烧瓶中15mL培养基)中传1代。细胞维持在37°C,5%CO<sub>2</sub>加湿环境中,并且在BSL2认证的组织培养罩中使用,使用了无菌技术和适当的个人防护设备,包括护目镜、手套和实验服。将80μL体积细胞以50,000细胞/mL的细胞密度在生长培养基中铺板过夜。在过夜粘附之后,向每个孔以250nM每孔(或者以下实施例中另外指定的)终浓度添加无菌PBS中的AARS多肽。对照孔含有未处理的细胞,并且与等体积PBS孵育。使细胞暴露于基础培养基(Cell Applications目录号310-470)中的蛋白或PBS,持续24小时。取出上清液,根据生产商说明书(RND Systems,目录号DY206和DY-208,DY-210 Duo-set试剂盒)运行IL-8、IL-6和TNFα ELISA测定。如前所述,通过在取出上清液之后向板添加含有刃天青的新鲜培养基并在37°C孵育3小时,使用刃天青评估增殖。在荧光板读数器上读板,活力/增殖表示为AARS多肽处理孔的刃天青相关荧光除以仅PBS处理孔的刃天青相关荧光的函数。

[0652] 人类星形细胞增殖和炎性细胞因子生成

[0653] 背景和治疗相关性:人类星形细胞(HA)源自人类大脑皮层。它们在第二代冻存,并且可以培养并增殖10次群体倍增。HA是中枢神经系统中最丰富的细胞,并且它们执行许多功能,例如为神经元提供机械支持和营养,并从神经元去除废物。除了对最佳神经元功能起到关键的支持作用,它们还提供形成血脑屏障的内皮细胞的生化支持。最近研究显示,星形细胞能够通过教导干细胞采取神经元命运并控制单突触功能来调节神经发生,积极参与脑内信息的传递和储存。对星形细胞在神经系统功能中的重要性的认识正在增加,HA可以用作探索星形细胞功能多样性的有效体外模型。已经显示星形细胞响应IL6和TNFα而增殖。此外,这些细胞能够生成它们自己的IL6和TNFα。因此,调节HA增殖和细胞因子生成的AARS多肽在许多神经疾病中具有治疗功用,所述神经疾病包括神经炎症、神经变性、脑肿瘤发生和脑缺血和修复。

[0654] 方法:根据生产商说明书,将来自Cell Applications(目录号882K-05f)的人类星形细胞(HA)保持在Cell Applications HA细胞生长培养基(目录号821-500)。细胞维持在37°C,5%CO<sub>2</sub>加湿环境中,并且在BSL2认证的组织培养罩中使用,使用了无菌技术和适当的个人防护设备,包括护目镜、手套和实验服。将80μL体积细胞以50,000细胞/mL的细胞密度在完全培养基(上述)铺板于胶原涂布板上过夜。细胞用PBS洗涤一次,向每个孔添加80μL的无血清生长培养基。向每个孔中以250nM每孔(或以下实施例中另外描述)的终浓度添加AARS多肽,以无菌PBS中一致的体积。使细胞暴露于AARS多肽48小时,取出使用的培养基进

行细胞因子评估(如前所述)。使细胞暴露于基础培养基(Cell Applications目录号310-470)中的蛋白或PBS,持续48小时。取出上清液,根据生产商说明书(RND Systems,目录号DY206和DY-208,DY-210 Duo-set试剂盒)运行IL-8和IL-6 ELISA测定。如前所述,通过在取出上清液之后向板添加含有刃天青的新鲜培养基并在37℃孵育3小时,使用刃天青评估增殖。在荧光板读数器上读板,活力/增殖表示为AARS多肽处理孔的刃天青相关荧光除以仅PBS处理孔的刃天青相关荧光的函数。

[0655] 人类肺微血管内皮细胞(HLMVEC)增殖和炎性细胞因子生成

[0656] 背景和治疗相关性:肺血管具有极大的生理/病理意义。目前认为它是由代谢上活性的功能上响应性的细胞构成的组织,以调节全身动脉血组成的方式与循环底物和形成的元素相互作用,影响靶器官功能,并促进血栓形成、止血和免疫反应以及肿瘤代谢。人类肺微血管内皮细胞(HLMVEC)表现出在急性肺损伤过程中提供白细胞定向迁移进入肺的关键提示的趋化细胞因子和细胞粘附分子的提高的表达。这种原发性细胞类型可以是在体外研究肺部微血管的病理学和生物学各个方面的有用工具。认为微血管响应于炎性刺激的结构和功能改变是器官损伤中的关键因素,并且在适当条件下可以提供修复刺激。这些血管改变的重要原因是涉及白细胞浸润的炎症反应的诱导。集中于粒细胞与内皮的粘附的许多研究已经揭示,白细胞募集和迁移包括精心策划的粘附级联。该粘附级联在粒细胞附着内皮时开始,并且开始沿低速液体流动方向展开。随着粒细胞展开,它变得活化,随后牢固附着于内皮,并跨内皮迁移进入血管外空间。这些粘附事件部分地受粒细胞表面上的CAM与内皮上存在的同源糖蛋白之间存在的分子相互作用介导。许多研究揭示,内皮细胞粘附分子E-选择蛋白可以与S<sub>Lex</sub>型聚糖呈递粒细胞配体相互作用,以介导粘附级联的附着和展开步骤。级联的下游步骤包括内皮表达的细胞间粘附分子与粒细胞表达的CD18整联蛋白的相互作用。

[0657] 因此,调节人肺微血管内皮细胞的增殖和/或细胞因子生成的AARS多肽在许多血管和肺疾病中具有治疗功用,所述疾病包括炎性和阻塞性肺病,包括例如肺高血压、慢性阻塞性肺病、特发性肺纤维化和哮喘。

[0658] 方法:HLMVEC(Cell Applications,目录号540-05)保持在Cell Applications微血管内皮细胞生长培养基(目录号111-500)中。为了适当生长,在铺板细胞之前,使用含有胶原的附着因子溶液(Cell Applications,目录号123-100)涂布板和烧瓶。细胞维持在37℃,5%CO<sub>2</sub>加湿环境中,并且在BSL2认证的组织培养罩中使用,使用了无菌技术和适当的个人防护设备,包括护目镜、手套和实验服。将80μL体积细胞以50,000细胞/mL的细胞密度在完全培养基(上述)铺板于胶原涂布板上过夜。细胞用PBS洗涤一次,向每个孔添加80μL的无血清生长培养基。向每个孔中以250nM每孔(或以下实施例中另外描述)的终浓度添加AARS多肽,以无菌PBS中一致的体积。使细胞暴露于AARS多肽48小时,取出使用的培养基进行ELISA,以评估细胞粘附分子和细胞因子(如前所述)。使用RND Systems的标准ELISA试剂盒(目录号分别为DY643和DY720)测量包括可溶性VCAM和/或ICAM的细胞粘附分子。如前所述,通过在取出上清液之后向板添加含有刃天青的新鲜培养基并在37℃孵育3小时,使用刃天青评估增殖。在荧光板读数器上读板,活力/增殖表示为AARS多肽处理孔的刃天青相关荧光除以仅PBS处理孔的刃天青相关荧光的函数。

[0659] 细胞粘附(以下数据表中的测定F1-F7)



[0660] 背景和治疗相关性:细胞粘附分子(CAM)是位于细胞表面上的蛋白,参与称为细胞粘附的过程中与其他细胞或与细胞外基质(ECM)的结合。这些蛋白通常是跨膜受体,并且由以下三个结构域构成:与细胞骨架相互作用的细胞内结构域,跨膜结构域,和与相同类型的其他CAM(同类分子结合)或其他CAM或细胞外基质(异类分子结合)相互作用的细胞外结构域。大多数CAM属于四个蛋白家族:Ig(免疫球蛋白)超家族(IgSF CAM)、整联蛋白、钙粘着蛋白和选择蛋白。免疫球蛋白超家族(IgSF)细胞粘附分子是不依赖钙的跨膜糖蛋白,包括:神经细胞粘附分子(NCAM)、细胞间细胞粘附分子(ICAM)、血管细胞粘附分子(VCAM)、血小板内皮细胞粘附分子(PECAM-1)、内皮细胞选择性粘附分子(ESAM)、连接粘附分子(JAM)、结合素和其他细胞粘附分子。

[0661] 细胞粘附分子是在许多炎性肝病中对于白细胞与肝窦内皮的粘附以及迁移和细胞毒性重要的细胞表面糖蛋白。ICAM-1在炎症中发挥重要作用,并且ICAM-1在内皮细胞上增加的表达反映在内皮细胞的活化。ICAM-1是特别重要的,因为它介导牢固的内皮粘附并促进白细胞迁移。研究显示,在炎性肝病中肝窦内皮细胞和肝细胞上都存在ICAM-1的上调,所述炎性肝病例如乙型肝炎病毒感染、自身免疫肝病、酒精性肝炎和肝同种异体移植排斥。

[0662] 因此,调节细胞粘附分子生成和与内皮细胞的细胞粘附的AARS多肽在许多炎性疾病中具有治疗功用,所述炎性疾病包括例如心血管疾病、粥样硬化、自身免疫和肺高血压。

[0663] 方法:人类脐静脉细胞(ATCC,目录号CRL-2873)(HUVEC)以约 $1.2 \times 10^5$ 细胞/孔的浓度接种于12孔板并根据生产商说明书培养,所述板涂布了在所示的ATCC培养基和补充物中的人类纤连蛋白附着溶液。细胞用所示浓度的AARS多肽或单独的PBS刺激,并在生长培养基中孵育过夜。人类急性单核细胞白血病(THP-1(TIB-202))细胞重悬于含有钙黄绿素AM( $6 \mu\text{L}/\text{mL}$ ;Invitrogen目录号C1430)的0.1%BSA/RPMI无血清培养基中,并且孵育30分钟。收集标记的细胞并重悬于含有10%FBS的RPMI培养基,调整密度至 $2 \times 10^6$ 细胞/mL。

[0664]  $100 \mu\text{L}$  ( $2 \times 10^5$ ) 标记的THP-1细胞在1mL生长培养基中放入HUVEC单层的每个孔中,并且孵育15分钟。用PBS洗涤孔两次以去除未结合的细胞,然后通过荧光板读数器读取细胞,使用488nm的激发波长和530nm的发射波长。

[0665] 细胞分化(以下数据表中的测定G1-G4)

[0666] 原代人类前脂肪细胞中的脂肪细胞分化和增殖

[0667] 背景和治疗相关性:肥胖和脂肪萎缩都普遍与包括糖尿病和心血管疾病的病理有关。目前公认脂肪组织是分泌许多因子的内分泌器官,失调的分泌影响脂肪生成以及全身葡萄糖/胰岛素稳态。导致肥胖的过量脂肪组织已经成为严重的公众健康威胁。脂肪组织发育可能受遗传背景、激素平衡、饮食和体力活动的影响。当脂肪细胞大小由于较高的三酰基甘油累积而增加时,脂肪组织质量可以增加。此外,还可以发生由于前体细胞分化成脂肪细胞而导致的脂肪细胞数目增加,即使在成年中,如在严重人类肥胖中和喂食高糖或高脂肪饮食的啮齿类动物中观察到的。具体地,认为脂肪细胞源自经历定型和分化过程(脂肪生成)的间充质细胞。前脂肪细胞系在用合成的糖皮质激素组成的脂肪生成剂处理时可以经历脂肪分化,地塞米松(DEX)、异丁基甲基黄嘌呤(IBMX)和胰岛素在这些研究中具有价值。已经非常明确,过氧化物酶体增殖剂激活受体 $\gamma$ (PPAR $\gamma$ )和转录因子的CCAAT增强剂结合蛋白(C/EBP)家族在脂肪分化中发挥重要作用。脂肪分化早期,C/EBP $\beta$ 和C/EBP $\delta$ 分别被DEX

和IBMX诱导,其然后一起诱导PPAR  $\gamma$  和C/EBP $\alpha$ 以激活脂肪功能所需的各种脂肪标志。已经报道其他转录因子也正向或负向调节脂肪生成,并且各种生长因子和激素可以通过调节脂肪生成转录因子的表达来影响脂肪细胞分化。事实上,除了通过贮存三乙酰甘油并在需要时释放脂肪酸而成为哺乳动物能量贮存的主要部位以外,脂肪组织分泌参与多种生理过程的许多分子,所述生理过程包括免疫应答、血管功能和能量稳态。诸如TNF- $\alpha$ 和IL-6的细胞因子从脂肪细胞分泌。这些因子的一些还可以通过自分泌/旁分泌作用来影响脂肪组织的生长和发育。

[0668] 因此,具有调节正常人类前脂肪组织的分化和/或增殖的能力的AARS多肽在许多疾病中具有治疗功用,包括例如治疗和预防代谢疾病、心血管疾病、肥胖和脂肪萎缩,以及糖尿病的长期并发症。

[0669] 方法:HPAd(人类前脂肪细胞)(Cell Application目录号803sD)根据供应商说明书来保持。为了培养,将细胞快速解冻并立即转移入15mL的脂肪细胞生长培养基(Cell Application目录号811M-250)并铺板入标准无菌组织培养处理烧瓶。每隔一天用新鲜的脂肪细胞生长培养基更换培养基,直至细胞>60%汇合。将细胞在37 $^{\circ}$ C,5%CO $_2$ 加湿环境中生长,并且在BSL2认证的组织培养罩中使用,使用了无菌技术和适当的个人防护设备,包括护目镜、手套和实验服。细胞以约50,000细胞/mL的浓度铺板于透明底黑壁96孔组织培养处理测定板以分化。向每个孔中以250nM每孔(或以下实施例中另外指定)的终浓度添加AARS多肽。所有细胞在生长培养基中维持2天,除了阳性对照用脂肪生成分化培养基(Cell Applications目录号811D-250)刺激。使细胞暴露于AARS多肽48小时。使用RND Systems的标准ELISA试剂盒(目录号分别为DY643和DY720)测量包括可溶性VCAM和/或ICAM的细胞粘附分子。如前所述,通过在取出上清液之后向板添加含有刃天青的新鲜培养基并在37 $^{\circ}$ C孵育3小时,使用刃天青评估增殖。在荧光板读数器上读板,活力/增殖表示为AARS多肽处理孔的刃天青相关荧光除以仅PBS处理孔的刃天青相关荧光的函数。添加新鲜培养基,分化维持初始培养基更换后16天,每隔一天更换新鲜培养基以维持细胞健康。在第15天,将细胞铺板于无血清培养基。在第16天,使用Nile Red(Invitrogen,3 $\mu$ M终浓度)染色来评估向成熟脂肪细胞的分化,并使用适当波长的荧光板读数器定量。为了进行该测定,细胞用10%低聚甲醛固定,在PBS中洗涤,并且在含有0.5%BSA和0.1%Triton X-100的PBS中通透。如前所述,用终浓度1 $\mu$ g/mL的Hoechst染料33432在荧光读数器上的强度测量来评估细胞增殖。脂肪生成表示为Nile Red信号的强度。Hoechst染料信号用来评估细胞数目。

[0670] 人类骨骼肌细胞分化和增殖

[0671] 背景和治疗相关性:骨骼肌的发育是多步骤过程,包括多潜能中胚层细胞定向生成肌细胞,成肌细胞从细胞周期退出并分化成肌细胞,以及最后骨骼肌纤维的生长和成熟。骨骼肌分化包括成肌细胞排列、延伸和融合成多核肌管,同时诱导调节和结构肌肉特异性基因。在分子水平上,生肌定型和肌肉特异性基因表达包括骨骼肌特异性螺旋-环-螺旋(bHLH)MyoD家族蛋白,其包括MyoD、肌细胞生成素、myf-5和MRF4以及肌细胞增强剂结合因子2(MEF2)。MyoD家族蛋白的DNA结合活性被Id减弱,Id与增殖细胞中的E2a基因产物形成复合物并且当它们被诱导分化时被下调。分化成肌管的决定受几种因子的负向影响。已知用牛血清、基础成纤维细胞生长因子2或转化生长因子 $\beta$ 1处理成肌细胞抑制成肌细胞的分化。肌发生还受癌基因的负向调节,所述癌基因例如c-myc、c-jun、c-fos、H-ras和E1a。关于在

血清退出时成肌细胞中触发的信号传导的信息很少,所述信号传导导致MyoD家族基因表达的诱导和肌肉分化。生肌分化似乎取决于成肌细胞血浆膜上存在的整联蛋白的激活,表明“从外到内”生化途径的运作,其中整联蛋白是上游分子物类。胰岛素样生长因子(IGF)-I和-II及其受体的相互作用也是骨骼肌分化的正向调节剂。

[0672] 因此,具有调节肌肉发育的能力的AARS多肽在许多疾病中具有治疗功用,包括例如治疗代谢疾病、恶病质、各种肌肉损耗病症以及肌肉骨骼疾病,其中肌肉萎缩在发病和症候学中起关键作用。人类骨骼肌细胞(HSkMC)可以经历分化以表现出肌动蛋白和肌球蛋白肌丝。HSkMC已经用于遗传肌肉疾病的研究,例如恶性高热。HSkMC还具有用作心脏移植物的潜力,修补心脏损伤,因此,具有调节肌肉发育的能力的AARS多肽也具有作为肌发生的体外和体内调节剂的功用。

[0673] 方法:为了评价AARS多肽在该过程中的潜在作用,采用了骨骼肌细胞分化的标准测定。对于该测定,从健康人供体的缘骨骼肌分离人类成年骨骼肌细胞(HSkMC,Cell Application目录号150-05f)。将细胞保持在HSkMC生长培养基(Cell Applications,目录号151-500)。这些细胞可以培养并繁殖至少15次群体倍增。为了分化,将细胞维持在生长培养基传1代,然后以50,000细胞/mL培养基铺板于用100 $\mu$ L/孔胶原处理的96孔透明底黑壁TC处理板。允许细胞粘附过夜。向每个孔以250nM每孔(或者以下实施例中另外指定的)终浓度添加PBS中的AARS多肽或单独的PBS。此时,对照孔接收相同体积的分化培养基(Cell Applications目录号151D-250)。细胞与蛋白或分化培养基孵育48小时,在48小时时,从所有孔收集细胞培养上清液,以150 $\mu$ L的体积向整个板添加分化培养基,除了对照孔仅维持在生长培养基中。如前所述,用上清液评估包括IL6和IL8的细胞因子生成。如前所述,通过在取出上清液之后向板添加含有刃天青的新鲜培养基并在37 $^{\circ}$ C孵育3小时,使用刃天青评估增殖。在显微镜下监测细胞并且每2天更换培养基为新鲜分化培养基。在第10天,取出培养基,用10%低聚甲醛固定细胞30分钟。用PBS中的0.1%Triton X-100通透细胞18分钟,细胞用TR标记的鬼笔环肽和Hoechst 33432(如前描述的)染色以分别确定肌动蛋白和细胞核。使用核强度测定每个孔中的细胞增殖,并且使用鬼笔环肽强度测定总肌动蛋白含量。细胞还用 $\alpha$ 肌动蛋白骨骼肌抗体(GenTex目录号GTX101362)染色。对所有孔的数字照片使用荧光显微镜以及目视检查并评分。

[0674] 人类骨髓间充质肝细胞分化和增殖

[0675] 背景和治疗相关性:间充质肝细胞(MSC)是可以分化成多种细胞类型的多能干细胞,所述细胞类型包括成骨细胞、软骨细胞、肌细胞、脂肪细胞、 $\beta$ -胰岛细胞和潜在的神经细胞。许多不同事件促进MSC定型为其他细胞系,包括转录因子、辅因子和来自众多途径的信号传导中间体的复杂网络的协调。MSC具有重要的治疗意义,因为它们代表具有治疗许多急性和变性疾病的潜力的细胞群体。

[0676] 而且,具有调节MSC分化成不同发育途径的能力的AARS多肽具有显著的治疗功用,能够在体外或体内调节血细胞生成、神经发生、肌发生、骨发生和脂肪生成以及许多疾患和疾病,包括例如炎症反应、自身免疫、癌症、神经变性、肌肉萎缩、骨质疏松和脂肪萎缩。人类MSC是免疫豁免的,并且代表同种异体移植的有利细胞类型,减少移植排斥和移植并发症风险。最近,在使用自体间充质肝细胞再生人类组织,包括软骨和半月板、腱和骨折中,已经取得了重要进展。许多研究还考察了MSC用于基因疗法,包括移植血管内皮生长因子转染的

MSC以改善大鼠MI之后的心脏功能, MSC作为干扰素- $\beta$ 的载体递送入小鼠肿瘤, 以及使用表达BMP的基因疗法以促进骨形成。因此, 由于MSC作为直接和改性治疗剂的重要意义以及AARS多肽用作调节MSC体内分化的治疗剂的潜力, 测试了作为MSC增殖和分化的潜在诱导剂的AARS多肽。

[0677] 方法:hMSC(人类骨髓基质细胞)(Cell Application目录号492-05f)根据供应商说明书来维持。为了培养, 将细胞快速解冻并立即转移入15mL骨髓基质细胞生长培养基(Cell Application目录号419-500)并铺板入标准无菌组织培养处理烧瓶。每隔一天用新鲜的骨髓基质细胞生长培养基更换培养基, 直至细胞 $>60\%$ 汇合。将细胞在 $37^{\circ}\text{C}$ ,  $5\%\text{CO}_2$ 加湿环境中生长, 并且在BSL2认证的组织培养罩中使用, 使用了无菌技术和适当的个人防护设备, 包括护目镜、手套和实验服。细胞以约50,000细胞/mL的浓度铺板于透明底黑壁96孔组织培养处理测定板以分化。向每个测定孔中以250nM每孔(或以下实施例中另外指定)的终浓度添加tRNA合成酶衍生的蛋白。所有细胞在生长培养基中维持2天, 除了阳性对照用生骨和生软骨的分化培养基(StemPro, Invitrogen, 目录号分别是A10072-01和A10071-01)刺激。使细胞暴露于AARS多肽48小时。使用来自RND Systems的标准ELISA试剂盒(目录号DY643)测量可溶性VCAM。如前所述, 通过在取出上清液之后向板添加含有刃天青的新鲜培养基并在 $37^{\circ}\text{C}$ 孵育3小时, 使用刃天青评估增殖。在荧光板读数器上读板, 活力/增殖表示为AARS多肽处理孔的刃天青相关荧光除以仅PBS处理孔的刃天青相关荧光的函数。在评估细胞活力之后, 通过两次培养基更换而去除刃天青, 并向所有孔添加0.5X分化培养基。通过所有孔的目视检查来监测分化, 持续培养基更换后10天, 每隔一天更换新鲜培养基以维持细胞健康。在第一次分化更换后第10天, 使用ELF-97染剂(Invitrogen目录号E6601)的碱性磷酸酶染色来评估分化。(Yang等人, Nature Protocols (6) 187-213 (2011) doi:10.1038/nprot.2010.189)。

[0678] 人类肺动脉平滑肌细胞(hPASMC)增殖和分化。

[0679] 背景和治疗相关性: 正常人类成年肺血管中的肺动脉平滑肌细胞(PASMC)大部分是静止、非迁移的, 并且主要致力于在肺内发挥其收缩功能。然而, PASMC不是终末分化的, 并且具有响应于变化的局部环境提示而调节其表型并退出其静止状态的能力。该分化状态可以响应于组织需要的变化而出现在发育、组织损伤和血管重构中。肺高血压(PH)与许多潜在病症相关, 包括由于增加的血管张力和PASMC收缩性以及血管重构而导致的外周肺血管阻力的增加。血管重构包括PASMC生长、基质材料合成、以及小肺动脉(PA)壁中细胞-细胞和细胞-基质相互作用的改变, 这导致血管壁平滑肌组分厚度增加以及正常非肌化的末端PA的异常肌化。该过程促使管腔直径减小和外周阻力增加。尽管PASMC在疾病起因中的确切作用还存在争议, 但是发生的变化在该疾病的临床后果中起关键作用。研究细胞分化的关键步骤是鉴定一组细胞特异性或细胞选择性基因, 其促成细胞的差异化功能。已经鉴定了许多平滑肌细胞(SMC)基因, 其用作血管SMC的分化或成熟的相对状态的有用标志, 例如SM $\alpha$ -肌动蛋白、SM MHC、h1-calponin、SM22- $\alpha$ 、结蛋白、变黏着斑蛋白、平滑肌蛋白和其他。最广泛使用的标志是SM $\alpha$ -肌动蛋白, 部分原因是该蛋白的大量极高亲和力和高度选择性抗体的商业可获得性。PASMC的变化是源于其固有特性还是源于控制PASMC生长的分子事件的失调, 依然是未决的问题。然而, 测定调控提示和管理潜在的失调为管理许多血管和肺疾病提供了重要的治疗启示, 所述疾病包括肺高血压、血管疾病。

[0680] 因此,具有调节衍生自成人的正常人类PASMC的分化和/或增值的能力的AARS多肽在许多血管和肺疾病中具有治疗功用,所述疾病包括炎性和阻塞性肺病,包括例如肺高血压、慢性阻塞性肺病、特发性肺纤维化和哮喘。

[0681] 方法:在使用之前,将HPASMC (Cell Applications目录号352-05a) 维持在HPASMC培养基 (Cell Applications目录号352-05a) (125mL烧瓶中15mL培养基) 中传一代。细胞维持在37°C, 5%CO<sub>2</sub>加湿环境中,并且在BSL2认证的组织培养罩中使用,使用了无菌技术和适当的个人防护设备,包括护目镜、手套和实验服。将80μL体积细胞以50,000细胞/mL的细胞密度在生长培养基中铺板在涂布的胶原上过夜。向每个孔以250nM(或者以下实施例中另外指定的) 终浓度添加无菌PBS中的AARS多肽。对照孔仅含有等体积的PBS。使阳性对照样品与供应商提供的HPASMC分化培养基 (Cell Applications目录号311D-250) 一起孵育。使细胞暴露于基础培养基 (Cell Applications目录号310-470) 中的AARS多肽或PBS,持续48小时,随后将整个板的培养基更换为分化培养基。如前所述,收集上清液并用来评估包括IL6和IL8的细胞因子生成。如前所述,通过在取出上清液之后向板添加含有刃天青的新鲜培养基并在37°C孵育3小时,使用刃天青评估增殖。监测细胞10天,每隔一天更换培养基。如上所述固定并用0.1% Triton X-100通透之后评估分化,通过使用抗SMA-α抗体 (GeneTex目录号GTX101362) 和Alexa 405耦连的二抗定量对平滑肌肌动蛋白-α染色的染色。在10%甲醛中固定细胞30分钟之后,使用Hoechst染色来评估增殖。使用底部读数的荧光板读数器读取Hoechst染料,使用405nm的激发波长 (Ex) 和450nm的发射波长 (Em)。通过使用Alexa-488标记的鬼笔环肽染色剂 (Invitrogen目录号A12379) 评估总的肌动蛋白染色。

[0682] AARS多肽与细胞结合的分析 (以下数据表中的测定H1-H10)

[0683] 背景和治疗相关性:AARS多肽与特定细胞类型的结合证明关注的细胞类型表达关注AARS多肽的特异性受体。根据关注的细胞类型,细胞结合意味着AARS多肽在体内调节该细胞或类似类型的细胞的活性或行为的潜在作用。这种调节作用的具体实例包括例如结合和调节B细胞和T细胞 (免疫调节/趋化/自身免疫/炎症); HepG2细胞 (控制代谢、胆固醇摄取或代谢); THP-1、jurkat、Raji细胞 (免疫调节/趋化/自身免疫/炎症)、血小板 (血栓形成)、3T3L1脂肪细胞 (脂肪生成/代谢) 和C2C12小鼠成肌细胞 (肌生成、骨发生)。

[0684] 与血细胞结合

[0685] 方法:从健康供体收集血液于EDTA管。将2mL全血放入5mL Falcon FACS管。添加2mL染色缓冲液 (PBS+2%FBS), 涡旋3-5秒钟,以300x g离心5分钟。吸取上清液,反复洗涤,沉淀重悬于2mL染色缓冲液。

[0686] 将100μl洗涤的血液转移入洁净的5mL FACS样品管。以下文所列的特定实验中指示的浓度向管中添加His6-或V5-His6-标记的AARS多肽并在冰上孵育45分钟。孵育之后,向管中添加不同细胞类型表面标志的抗体 (BD Pharmigen目录号560910、555398、555415、340953、560361) 和FITC标记的抗V5标签抗体 (V5-FITC, Invitrogen目录号R96325) 或FITC标记的抗His6抗体 (AbCam目录号ab1206), 并在冰上黑暗孵育30分钟。在孵育之后,向管中添加2mL BD FACS裂解溶液 (目录号349202)。涡旋样品并在冰上放置15分钟。用1x2mL PBS洗涤样品并重悬于2mL 2%甲醛的PBS溶液,之后FACS分析。当单独抗体没有显著信号时,结合超过25%细胞群体的AARS多肽被认为是击中。

[0687] 血小板结合测定:将50μl洗涤的血液转移入洁净的5mL FACS样品管。以下文所列

的特定实验中指示的浓度向管中添加His6-或V5-His6-标记的AARS多肽并将管在冰上放置45分钟。向每个管添加20 $\mu$ L CD61泛血小板抗体 (BD Pharmigen, 目录号555754) 和0.5 $\mu$ L抗V5-FITC标记抗体 (Invitrogen, R96325) 或FITC标记的抗His6抗体 (AbCam目录号ab1206)。将管放在冰上并避光30分钟。使样品在1%甲醛的PBS溶液中达到2mL的总体积,并在24小时内通过流式细胞术分析。当单独抗体没有显著信号时,结合超过25%细胞群体的AARS多肽被认为是击中。

[0688] 与培养细胞的结合:将100 $\mu$ L完全RPMI培养基中大约 $1 \times 10^6$ 细胞放入5mL FACS管。以下文所列的特定实验中指示的浓度向管中添加His6-或V5-His6-标记的AARS多肽并将管在冰上放置45分钟。细胞样品用1mL染色缓冲液 (PBS+2%FBS) 洗涤两次,然后添加含有200 $\mu$ g/mL人IgG的染色缓冲液中的0.5 $\mu$ L抗V5-FITC抗体 (Invitrogen R96325) 或FITC标记的抗His6抗体 (AbCam目录号ab1206),并将样品在冰上孵育,避光30分钟。将样品用1mL染色缓冲液洗涤两次,然后在1%甲醛的PBS溶液中达到1mL的总体积,并在24小时内通过流式细胞术分析。当单独抗体没有显著信号时,结合超过25%细胞群体的AARS多肽被认为是击中。

[0689] 动物研究:调节血细胞生成和循环细胞因子

[0690] 背景和治疗相关性:血细胞生成(或者说造血作用或生血作用)是血液细胞成分的形成。所有血液细胞成分源自造血干细胞(HSC),造血干细胞存在于骨髓质(骨髓)并且具有产生所有不同类型成熟血细胞的独特能力。HSC是自我更新的:当它们增殖时,其子细胞中的至少一些保持为HSC,使得干细胞池不会耗尽。然而,HSC的其他子细胞(骨髓和淋巴祖细胞)可以各自定向于可选分化途径的任一个,使得产生血细胞的一种或多种特定类型,但是它们本身不能自我更新。因此,血液成分响应于AARS多肽暴露的变化表明,AARS多肽能够调节血细胞生成,并且调节造血干细胞的发育。

[0691] 所有血细胞可以分成三个细胞系:类红细胞、淋巴细胞和髓细胞。

[0692] 类红细胞是携氧红血球。网织红细胞和红细胞是功能性的并且被释放入血液。因此,网织红细胞计数估计了红细胞生成速率,红血球计数变化表明AARS多肽调节红细胞生成。

[0693] 淋巴细胞是适应性免疫系统的基石。它们源自共同的淋巴祖细胞。淋巴细胞系主要由T细胞和B细胞(白血球类型)组成。因此,响应于AARS多肽暴露的白血球计数或组成的变化表明AARS多肽调节淋巴细胞生成。

[0694] 髓细胞包括粒细胞、巨核细胞和巨噬细胞,并且源自共同的髓祖细胞,牵涉许多作用,包括先天免疫、适应性免疫和血液凝固。因此,响应于AARS多肽暴露的髓细胞计数或组成的变化表明AARS多肽调节髓细胞生成。通过测量响应于AARS多肽暴露的粒细胞数目变化,可以利用相同的原理来确定AARS多肽是否调节粒细胞生成。可以通过血液中巨核细胞或血细胞组成或数目的变化来推断AARS多肽调节巨核细胞生成的作用。

[0695] 野生型小鼠或各种炎症动物模型系统中的细胞因子释放提供了AARS多肽调节炎症反应的潜在能力的最初评估。例如,可以使用饮食诱导的肥胖(DIO)的小鼠模型来容易地评估AARS多肽调节急性慢性炎症过程的作用。DIO模型重点在于使啮齿类动物置于高脂肪饮食数月,导致肥胖增加、胰岛素抵抗和免疫系统功能障碍。该免疫系统失调的具体后果导致DIO动物中促炎细胞因子生成增加,导致慢性系统性炎症病症。越来越多的证据表明,低度炎症促进肥胖和在称为代谢综合征的人类病症中同样发现的糖尿病表型的发展和维持。

这样,AARS多肽调节免疫系统并恢复体内平衡向着解决这种慢性炎症状态发展的能力将特别有益于许多疾病和病患,包括但不限于治疗和预防代谢疾病、糖尿病、心血管疾病、粥样硬化、肥胖以及各种自身免疫疾病和疾患的症状和副作用,所述疾病和疾患包括例如多发性硬化、血管疾患和变应性疾患。

[0696] 方法:雄性野生型对照(C57BL/6)或饮食诱导的肥胖小鼠(C57BL/6NHsd)购自Harlan(Indianapolis,IN)并单独圈养。DIO小鼠被喂食高脂肪饮食(目录号TD.06414-60% kcal来自脂肪),对照小鼠被喂食正常饮食(目录号2018S-18%kcal来自脂肪)。将DIO小鼠置于高脂肪饮食,在6周龄时开始,持续共10周。允许DIO和对照小鼠随意进食和饮水。在16周龄时,基于体重分选小鼠并随机分成5只动物一组。在第2天,称重小鼠并尾静脉取血(100  $\mu$ L)以进行治疗前全血计数(CBC)分析。在第1天,称重小鼠并经由尾静脉以10mg/kg静脉内注射介质(PBS)或个体AARS多肽。注射后4小时,面静脉取血(150-200 $\mu$ L)以进行随后的细胞因子分析。在第2、3和4天,小鼠如第1天一样静脉内施用。在第5天,称重小鼠,终止,并通过心脏穿刺收集血液以进行全血计数(CBC分析)(血浆-EDTA)和细胞因子检查(血清)。

[0697] CBC和细胞因子分析:从注射前(第2天)和最后注射之后24小时(第5天)抽取的血液分析全血计数。评估白血球总计数和总体红血球形态的CBC值。通过嗜中性粒细胞、淋巴细胞、单核细胞、嗜伊红性粒细胞和嗜碱粒细胞的总百分比和各自百分比来进一步表征白血球。红血球分解包括血红蛋白(dL)、血细胞比容(%)、平均血细胞比容(fL)、平均血细胞血红蛋白、平均血细胞血红蛋白浓度(%)和总血小板计数( $10^3/\mu$ L)的测量。CBC分析通过Antech Diagnostics(Fishers,IN)进行。如果physiocrine处理的动物中细胞或分析物水平和媒介物处理的动物(计算用于对照组)的区别大于一个标准偏差,则认为CBC测量中的改变为击中,只要该偏差大于从对照组计算的平均值15%即可。

[0698] 在注射后4小时(第1天)和最后注射后24小时(第5天)检查循环细胞因子水平。分离血清,快速冷冻,并送至Rules Based Medicine(Austin,TX)以进行多分析物分析。使用涵盖59个独特生物标志的RodentMap板分析血清样品,所述生物标志包括Apo A-1、CD40、CD40-L、CRP、ET-1、嗜酸细胞活化趋化因子、EGF、因子VII、血纤蛋白原、FGF-9、FGF-basic、GST- $\alpha$ 、GCP-2、GM-CSF、KC/GRO $\alpha$ 、触珠蛋白、IgA、IFN  $\gamma$ 、IP-10、IL-1 $\alpha$ 、IL-1 $\beta$ 、IL-10、IL-11、IL-12p70、Il-17A、IL-18、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、LIF、淋巴细胞趋化因子、M-CSF-1、MIP-1 $\alpha$ 、MIP-1 $\beta$ 、MIP-1  $\gamma$ 、MIP-2、MIP-3 $\beta$ 、MDC、MMP-9、MCP-1、MCP-3、MCP-5、MPO、肌红蛋白、SAP、SGOT、SCF、RANTES、TPO、组织因子、TIMP-1、TNF- $\alpha$ 、VCAM-1、VEGF-A和vWF。如果与介质对照相比,细胞因子增加至少2倍或减少至少50%,则将细胞因子水平变化计数为击中。

[0699] 实施例1

[0700] 使用蛋白拓扑和迁移分析平台鉴定从AARS的蛋白水解片段和选择性剪接产物

[0701] 为了从细胞系、条件培养基和组织鉴定AARS片段,以如下方式制备样品:

[0702] 小鼠巨噬细胞(RAW 264.7)、细胞溶质和条件培养基:使用无血清DMEM培养基以的 $15 \times 10^6$ 细胞/烧瓶的密度处理细胞。48小时后,收集条件培养基和细胞沉淀并加工。通过SDS-PAGE从分泌和细胞溶质蛋白组级分分离200 $\mu$ g蛋白,制备凝胶切片以通过质谱分析。

[0703] 小鼠胰腺组织:从三只小鼠切下胰腺,dounce匀浆,并在含有蛋白酶抑制剂的PBS中超声。通过离心分离细胞溶质蛋白组,并通过SDS-PAGE分离200 $\mu$ g蛋白,制备凝胶切片以

通过质谱分析。

[0704] 小鼠肝组织:切下三只小鼠的肝,dounce匀浆,并在含有蛋白酶抑制剂的PBS中超声。通过离心分离细胞溶质蛋白组,并通过SDS-PAGE分离200 $\mu$ g蛋白,并制备凝胶切片以通过质谱分析。

[0705] 通过配有终极3000 $\mu$ LC系统(Dionex)的LTQ XL离子阱质谱仪(ThermoFisher)分析凝胶内消化物。首先使用Dionex自动进样器,将样品用5%乙腈的0.1%甲酸溶液上样至PepTrap(michrom),持续10分钟。然后,用含有10cm C18树脂(michrom)的100 $\mu$ m(内径)熔融二氧化硅毛细管柱分析样品。使用在0.1%甲酸中5-33.5%乙腈的线性梯度在110分钟内以0.45 $\mu$ l/min的流速将肽从柱洗脱进入质谱仪。

[0706] 以数据依赖性筛选模式操作LTQ,使得一次全MS扫描之后进行7个最丰富离子的7次MS/MS扫描。进行动态排除,重复计数等于1,重复持续时间等于20秒,排除列表大小是300,并且排除持续时间是60秒。

[0707] 在LC-MS/MS分析之后,使用小鼠IPI数据库的连锁目标/诱饵变体,用BioWorks3.3.1(SEQUEST)检索原始数据。过滤SEQUEST数据并用DTASelect分选。表1、4和7显示了以这种方式鉴定的序列。

[0708] 实施例2

[0709] 使用深度测序鉴定剪接变体

[0710] 使用富集了氨酰基tRNA合成酶转录物的cDNA文库的高通量测序,鉴定氨酰基tRNA合成酶的剪接变体。从诸如成人脑和胎儿脑的组织的总RNA提取物制备cDNA模板,并使用特异性针对所有注释人类氨酰基tRNA合成酶及其相关蛋白的所有注释外显子的引物序列富集氨酰基tRNA合成酶转录物。

[0711] 人类总RNA获自Clontech。对于细胞系和小鼠组织样品,使用RNA提取II试剂盒(MN)提取总RNA。通过DNA酶I消化总RNA样品中的基因组DNA。为了获得成熟信使RNA(mRNA),通过结合多聚A+RNA和通过5'-磷酸依赖性外切核酸酶消化无5'-帽的RNA,两次富集RNA样品。使用与氨酰基tRNA合成酶基因的外显子序列退火的引物,从成熟RNA合成互补DNA(cDNA)。使用氨酰基tRNA合成酶-外显子特异性cDNA和氨酰基tRNA合成酶-外显子引物的不同组合,通过多重PCR扩增富集了氨酰基tRNA合成酶基因的转录组。在向修复片段的3'端添加A-突出端之前,在两端酶促修复双链氨酰基tRNA合成酶富集的转录组PCR产物。然后向氨酰基tRNA合成酶富集的转录组PCR产物添加测序衔接子和索引序列以产生供Illumina的多重测序试剂盒深度测序的cDNA文库。简言之,将具有3'-A突出端的氨酰基tRNA合成酶富集的转录组PCR产物连接至试剂盒中提供的InPE衔接寡核苷酸。将索引序列添加至具有InPE衔接子的PCR产物。为了获得供深度测序的足够DNA片段,通过PCR进一步扩增具有索引序列的PCR产物。汇集具有不同索引的氨酰基tRNA合成酶富集的cDNA文库并使用Illumina DNA测序仪测序,以获得50碱基对末端读数。测序读数与人类或小鼠基因组作图以鉴定选择性剪接事件。使用“Splicemap”软件(可在<http://www-stat.stanford.edu/~kinfai/SpliceMap/>公共下载获得)来鉴定剪接点。

[0712] 进行这些cDNA的深度测序以产生长度约50核苷酸的约1百万测序读数。针对注释外显子接合处查询特异性针对氨酰基tRNA合成酶的外显子的序列,并将新的外显子接合处鉴定为选择性剪接事件。



[0713] 表2、5和8中各列标记的“5' 外显子”和“3' 外显子”指示,存在时,这些外显子在cDNA序列中融合在一起。表2、5和8显示根据选择性剪接事件鉴定的序列,含有此类剪接事件的转录物,和由这些转录物表达的多肽。通过深度测序鉴定的选择性剪接变体在表2、5和8中鉴定为在成人或胎儿脑中标记为“测序读数”的列中存在大于零的数值的那些变体。

[0714] 实施例3

[0715] 使用生物信息学鉴定AARS多肽

[0716] 使用生物信息学鉴定AARS蛋白片段(切除蛋白或appendacrine肽)。使用诸如FASTA(可获自网站[http://fasta.bioch.virginia.edu/fasta\\_www2/fasta\\_www.cgi](http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi))或NCBI的BLASTP程序(可获自网站[http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PROGRAM=blastp&BLAST\\_PROGRAMS=blastp&PAGE\\_TYPE=BlastSearch&SHOW\\_DEFAULTS=on&LINK\\_LOC=blasthom](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthom)),将全长人类氨酰基tRNA合成酶的氨基酸序列与其来自细菌大肠杆菌的直系同源物的全长氨基酸序列比对。当比对的细菌序列中存在缺口,或者两个物种之间具有低同源性的区域时,将来自人蛋白的切除蛋白序列鉴定为序列覆盖区。表3、6和9中的肽和相应的DNA序列包括以这种方式鉴定的实例。

[0717] 实施例4

[0718] 通过质谱鉴定的AARS多肽的差异表达

[0719] 如实施例1所述使用PROTOMAP技术来比较不同组织/细胞类型中苯丙氨酰- $\alpha$ 亚单位tRNA合成酶的差异表达(有关序列和比较,参考表1、4和7):在小鼠肝组织和小鼠胰腺组织之间比较表达。比较RAW264.7的细胞溶质与在血清饥饿48小时之后收获的RAW264.7细胞的条件培养基之间的氨酰-tRNA合成酶切除蛋白表达。

[0720] 实施例5

[0721] 通过深度测序鉴定AARS多肽的差异表达

[0722] 为了测试剪接变体的差异表达,对从不同组织制备的cDNA进行深度测序。

[0723] 氨酰基tRNA合成酶的特异性选择性剪接变体事件的表达是意想不到的,并且指示生物重要性。不同转录组样品的深度测序中所见的读数相对数目的变化指示,氨酰基tRNA合成酶的选择性剪接事件受差异化调节并且不仅仅是由于样品处理的假象。

[0724] 实施例6

[0725] 抗体筛选

[0726] 为利于发现表现出与特定氨酰基tRNA合成酶片段优先结合(例如,与亲本全长酶相比 $\geq 10$ 倍更高的亲和力)的抗体,使用亲和富集技术(淘选),通过AbD Serotec (MorphoSys<sup>TM</sup>的一个部门, Martinsried/Planegg, Germany) 筛选人类抗体噬菌体展示文库。在使用氨酰基tRNA合成酶片段多轮筛选之后富集的抗体随后通过ELISA表征与这些片段的反应性和与亲本全长酶的反应性。还进一步表征了证明与氨酰基tRNA合成酶片段优先结合(例如, $\geq 10$ 倍更高的亲和力)的克隆。

[0727] 如果在该过程结束时没有实现必要的特异性,减法策略,例如与全长酶的预吸附步骤和/或反筛选,被用来清除交叉反应的抗体并驱使选择过程朝向氨酰基tRNA合成酶片段上的独特表位。

[0728] 实施例7

[0729] 使用系统PCR鉴定剪接变体

[0730] PCR反应的cDNA模板从组织或细胞(例如,人脑、IMR-32和HEK293T)的总RNA提取物逆转录。使用氨酰基tRNA合成酶特异性引物进行PCR反应,使被设计为与基因的5'半部分中的5'未翻译区或外显子退火的正向引物(FP1)与被设计为与基因的3'半部分或3'UTR中的外显子退火的反向引物(FP1)配对。通过琼脂糖凝胶电泳分析扩增的DNA产物以鉴定不同大小的PCR产物,然后片段扩增形成常规转录物。这些不同的PCR产物从凝胶切离并纯化,并连接入标准克隆载体以进行DNA序列分析。将选择性剪接变体鉴定为来自常规转录物的不同序列。通过该系统PCR方法鉴定的剪接变体示于表2、5和8。

[0731] 实施例8

[0732] 选择的AARS多核苷酸的密码子优化

[0733] 基于以下标准的一个或多个来选择代表性AARS多肽(总结于表E2)以进行进一步的生化、生物物理和功能表征:i) AARS多肽蛋白水解片段的鉴定,ii) AARS多肽剪接变体的鉴定,iii) 通过生物信息学分析鉴定AARS多肽,iv) 特定AARS多肽的差异表达证据,v) AARS蛋白的结构域结构,vi) AARS多肽的大小,和vii) 类似重复序列的最小化。

**表 E2**  
**密码子优化和细菌表达而选择的 AARS 多肽的总结**

AARS 多肽名称	表位标记 AARS 多肽的 SEQ. ID No.	AARS 多核苷酸的 SEQ. ID No.	AARS 蛋白的残基	表位标签的位置	使用的克隆/合成方法
PheRSa1 <sup>N1</sup>	SEQ.ID. NO. 47	SEQ.ID. NO. 66	1-226	N-端	1
PheRSa1 <sup>N1</sup>	SEQ.ID. NO. 48	SEQ.ID. NO. 66	1-226	C-端	1
PheRSa1 <sup>N2</sup>	SEQ.ID. NO. 49	SEQ.ID. NO. 67	1-219	N-端	1
PheRSa1 <sup>N2</sup>	SEQ.ID. NO. 50	SEQ.ID. NO. 67	1-219	C-端	1
PheRSa1 <sup>N3</sup>	SEQ.ID. NO. 51	SEQ.ID. NO. 68	1-128 + 13 aa	N-端	1
PheRSa1 <sup>N3</sup>	SEQ.ID. NO. 52	SEQ.ID. NO. 68	1-128 + 13 aa	C-端	1
PheRSa1 <sup>N4</sup>	SEQ.ID. NO. 53	SEQ.ID. NO. 69	1-168 + 200-508	N-端	1
PheRSa1 <sup>N4</sup>	SEQ.ID. NO. 54	SEQ.ID. NO. 69	1-168 + 200-508	C-端	1
PheRSa1 <sup>N5</sup>	SEQ.ID. NO. 55	SEQ.ID. NO. 70	1-198 + S199R + 243-508	N-端	1
PheRSa1 <sup>N5</sup>	SEQ.ID. NO. 56	SEQ.ID. NO. 70	1-198 + S199R + 243-508	C-端	1
PheRSa1 <sup>N6</sup>	SEQ.ID. NO. 57	SEQ.ID. NO. 71	1-168 + 31 aa	N-端	1
PheRSa1 <sup>N6</sup>	SEQ.ID. NO. 58	SEQ.ID. NO. 71	1-168 + 31 aa	C-端	1
PheRSa1 <sup>N7</sup>	SEQ.ID. NO. 59	SEQ.ID. NO. 72	1-399 + 426-508	N-端	1
PheRSa1 <sup>N7</sup>	SEQ.ID. NO. 60	SEQ.ID. NO. 72	1-399 + 426-508	C-端	1
PheRSa1 <sup>N8</sup>	SEQ.ID. NO. 61	SEQ.ID. NO. 73	1-242 + 31 aa	N-端	1
PheRSa1 <sup>N8</sup>	SEQ.ID. NO. 62	SEQ.ID. NO. 73	1-242 + 31 aa	C-端	1
PheRSa1 <sup>N9</sup>	SEQ.ID. NO. 63	SEQ.ID. NO. 74	1-152	N-端	1
PheRSa1 <sup>N9</sup>	SEQ.ID. NO. 64&65	SEQ.ID. NO. 74	1-152	C-端	1
PheRSa1 <sup>C1</sup>	SEQ.ID. NO. 84	SEQ.ID. NO. 90	459-508	N-端	1
PheRSa1 <sup>C1</sup>	SEQ.ID. NO. 85	SEQ.ID. NO. 90	459-508	C-端	1
PheRSa1 <sup>C2</sup>	SEQ.ID. NO. 86	SEQ.ID. NO. 91	53 aa + 169-508	N-端	1
PheRSa1 <sup>C2</sup>	SEQ.ID. NO. 87	SEQ.ID. NO. 91	53 aa + 169-508	C-端	1
PheRSa1 <sup>C3</sup>	SEQ.ID. NO. 88	SEQ.ID. NO. 92	290-508	N-端	1
PheRSa1 <sup>C3</sup>	SEQ.ID. NO. 89	SEQ.ID. NO. 92	290-508	C-端	1

[0734]

[0735] 使用表E2所列的基因合成方法,如遗传材料和方法章节所述,合成并克隆了编码表E2所列的选定AARS多肽以及适当N或C端表位标签的多核苷酸。

[0736] 实施例9

[0737] 小规模细菌表达和纯化

[0738] 如遗传材料和方法章节所述,在大肠杆菌中表达了表E2所列的AARS多肽。以下表E3中总结了可溶性的和包涵体定位的AARS多肽的相对表达。

表 E3  
AARS 多肽细菌表达特征的总结

AARS 多肽	表位标签的位置	从可溶性级分回收的蛋白的量	从包涵体回收的蛋白的量
PheRSa1 <sup>N1</sup>	N-端	+++	ND
PheRSa1 <sup>N1</sup>	C-端	++	ND
PheRSa1 <sup>N2</sup>	N-端	+++	ND
PheRSa1 <sup>N2</sup>	C-端	++++	ND
PheRSa1 <sup>N3</sup>	N-端	+	ND
PheRSa1 <sup>N3</sup>	C-端	++++	ND
PheRSa1 <sup>N4</sup>	N-端	+	ND
PheRSa1 <sup>N4</sup>	C-端	+	ND
PheRSa1 <sup>N5</sup>	N-端	+	ND
PheRSa1 <sup>N5</sup>	C-端	+	ND
PheRSa1 <sup>N6</sup>	N-端	+	ND
PheRSa1 <sup>N6</sup>	C-端	+	ND
PheRSa1 <sup>N7</sup>	N-端	+	ND
PheRSa1 <sup>N7</sup>	C-端	+	ND
PheRSa1 <sup>N8</sup>	N-端	+	ND
PheRSa1 <sup>N8</sup>	C-端	+	ND
PheRSa1 <sup>N9</sup>	N-端	+	ND
PheRSa1 <sup>N9</sup>	C-端	+	ND
PheRSa1 <sup>C1</sup>	N-端	+	+
PheRSa1 <sup>C1</sup>	C-端	+	+
PheRSa1 <sup>C2</sup>	N-端	+	+
PheRSa1 <sup>C2</sup>	C-端	+	+
PheRSa1 <sup>C3</sup>	N-端	+	+
PheRSa1 <sup>C3</sup>	C-端	+	+
“+” 代表 0-1 mg/L AARS 多肽表达 “++” 代表 1-5 mg/L AARS 多肽表达; “+++” 代表 5-10 mg/L AARS 多肽表达; “++++” 代表 10-15 mg/L AARS 多肽表达; “+++++” 代表 ≥15 mg/L AARS 多肽表达;			
ND: 未测定			

[0739]

[0740] 惊人惊讶地,蛋白表达数据证明,当在大肠杆菌中表达时表现出高水平的可溶性蛋白表达的至少两个蛋白结构域的存在。具体地,数据证明,AARS多肽PheRSa1<sup>N1</sup>(氨基酸1-226)和PheRSa1<sup>N2</sup>(氨基酸1-129)确定了在大肠杆菌中高度表达的第一新的蛋白结构域的界限。另外,数据证明,AARS多肽PheRSa1<sup>N3</sup>(氨基酸1-128+13aa)确定了在大肠杆菌中高度表达的第二新的蛋白结构域的界限。

[0741] 实施例10

[0742] AARS多肽的大规模生产

[0743] 以较大量制备代表性AARS多肽,以实现进一步的功能和生物物理表征。如遗传材料和方法章节所述,表E4所列AARS多肽在大规模培养的大肠杆菌中表达。每个表达的可溶性蛋白的产率和指定的生物物理特性在下面总结于表E4。

表 E4 代表性 AARS 多肽产率和生物物理表征的总结								
AARS 多肽	表位标签 的位置	产率 [mg/L] (1)	纯度 [%]	内毒素 [EU/mg]	分子量	工作贮液 浓度 [mg/ml]	稳定性[回 收百分 比] <sup>(2)</sup>	聚集[DLS]
PheRSa1 <sup>N2</sup>	C-端	5.7	90	1.1	C: 26,471 D: 26,343 (3)	33.4	96	+
注解 (1): 通过在最后纯化步骤之后测量蛋白回收而测定的产率 (2): 测定为在 25°C 下 1 周后非聚集材料的回收百分比 (3): 可能代表无 N 端蛋氨酸的 MW C: 计算的 D: 测定的 图例: “+” 代表小于 1% 高分子蛋白聚集物 “++” 代表小于 2% 高分子蛋白聚集物 “+++” 代表小于 5% 高分子蛋白聚集物 “++++” 代表小于 10% 高分子蛋白聚集物 “+++++” 代表大于 10% 高分子蛋白聚集物 ND: 未测定								

[0744] [0745] 这些研究的结果确定,来自AARS蛋白PheRSa1<sup>N2</sup>家族的代表性AARS蛋白表现出合理的初始蛋白表达产率和溶解度特性。

[0746] 实施例11

[0747] 代表性AARS多肽的转录分析

[0748] 为了测试AARS多肽调节基因表达的能力,将选择的AARS多肽与间充质肝细胞或人类骨骼肌细胞孵育,时间和浓度如表E5所示。

测试样品描述		细胞类型和暴露时间				
AARS 多肽	表位标签的位置	浓度 nM	MSC 24 小时	MSC 72 小时	HSkMC 24 小时	HSkMC 72 小时
PheRSa1 <sup>N1</sup>	N-端	250	3	2	4	8
PheRSa1 <sup>N1</sup>	C-端	250	2	3	6	13
PheRSa1 <sup>N2</sup>	N-端	250	1	6	5	8
PheRSa1 <sup>N2</sup>	C-端	250	7	13	9	18
PheRSa1 <sup>N3</sup>	C-端	250	0	4	5	8
PheRSa1 <sup>N9</sup>	C-端	250	5	4	3	11
对照						
筛选的所有 AARS 多肽的平均值			3	5	6	7
骨发生混合物			17	20	11	16
软骨形成混合物			17	19	14	19
脂肪生成混合物			19	15	16	18
SKMC 阳性对照			11	8	5	4
骨发生混合物			0	0	1	1

[0749]

[0750] 在表E5中,如一般方法章节中描述的,每列中的数字代表与对照基因相比被正或负调节至少4倍的基因数目。数据显示,当添加至间充质肝细胞(MSC)和/或人类骨骼肌细胞(HSkMC)时,测试的AARS多肽的特定形式具有调节转录并因此潜在调节发育命运或分化状态的惊人能力。表中用加粗数字显示的单元格代表其中AARS多肽对表中所示细胞系和时间的基因转录调节表现出显著影响的实例。

[0751] 结论,PheRSa1<sup>N1</sup>,PheRSa1<sup>N2</sup>和PheRSa1<sup>N3</sup>显示是间充质肝细胞和/或人类骨骼肌细胞基因表达的主要调节剂。。

[0752] 实施例12

[0753] AARS多肽的功能分析

[0754] 为了测试AARS多肽调节许多表型过程的能力,将选择的AARS多肽与一般方法章节和表E5和E6中提供的细胞类型和条件孵育。

[0755]

<b>表E6</b>	
<b>测定的关键和指示击中的标准</b>	
<b>增殖测定</b>	
来源和细胞类型	测定号
人类巨核细胞白血病细胞/Mo7e	A1
人类急性早幼粒细胞白血病细胞/HL60	A2
人类成淋巴细胞(癌细胞系)/RPMI8226	A3
人类间充质干细胞/hMSC	A4
人类星形细胞	A5
人类骨髓穿刺细胞/骨髓细胞	A6
人类骨髓穿刺细胞/骨髓细胞(长期培养)	A7
人类滑膜细胞/HFLS-SynRA	A8
人类前脂肪细胞/hPAD	A9
人类肺动脉平滑肌细胞/hPASMC	A10
人类骨骼肌细胞/hSKMC	A11
<p>通过用测定孔中的数值除以测定板的平均PBS值来进行增殖测定的数据分析。如果沿正向偏离PBS值超过3SD,则认为AARS多肽是增殖性的。如果沿负向偏离PBS值超过3SD,则认为tRNA合成酶衍生的AARS多肽是细胞毒性的。细胞毒性化合物用作阴性对照,并且其平均值总是偏离PBS平均值超过3SD。</p>	
<b>细胞分化和表型测定</b>	
测定描述	测定号
人类肝细胞(HepG2C3a细胞)乙酰化LDL摄取	B1
<p>通过用测定孔中的数值除以测定板的平均PBS值来进行ac-LDL摄取测定的数据分析。如果沿正向或负向偏离PBS值超过2SD,则认为AARS多肽是ac-LDL摄取的调节剂。使用荧光显微镜进行目视检查以确认板读数结果。</p>	
<b>人类嗜中性粒细胞测定</b>	
测定描述	测定号
嗜中性粒细胞弹性蛋白酶	C1

[0756]

<b>表E6</b>	
<b>测定的关键和指示击中的标准</b>	
<b>增殖测定</b>	
嗜中性粒细胞氧化猝发(激动剂)	C2
嗜中性粒细胞氧化猝发(拮抗剂)	C3
<p>通过用测定孔中的数值除以测定板的平均 PBS 值来进行嗜中性粒细胞测定的数据分析。如果沿正向或负向偏离 PBS 值超过 2SD, 则认为 AARS 多肽是嗜中性粒细胞弹性蛋白酶生成或氧化猝发生物学的调节剂。</p>	
<b>样受体(TLR)的调节</b>	
测定描述	测定号
RAW BLUE 细胞中的 TLR 激活	D1
RAW BLUE 细胞中的 TLR 拮抗	D2
hTLR2 的激活	D3
hTLR4 的激活	D4
<p>通过用测定孔中的数值除以测定板的平均PBS值来进行TLR调节测定的数据分析。如果沿正向或负向偏离PBS值超过3SD, 则认为AARS多肽是TLR特异性生物学的调节剂。阳性对照, 包括LPS和检测试剂, 总是显著不同的并且偏离PBS平均值&gt; 3 SD。</p>	
<b>细胞因子释放</b>	
测定描述	测定号
人类滑膜细胞细胞因子生成(IL6 释放)	E1
人类肺动脉平滑肌细胞(hPASMC)细胞因子生成(IL6 释放)	E2
人类骨骼肌细胞(hSKMC)细胞因子生成(IL6 释放)	E3
人类星形细胞细胞因子生成(IL6 释放)	E4
全血 IL6 释放	E5
人类肺动脉平滑肌细胞(hPASMC)细胞因子生成(IL8 释放)72 h 孵育	E6
<b>IL8生成</b>	
测定描述	测定号



<b>表E6</b>	
<b>测定的关键和指示击中的标准</b>	
<b>增殖测定</b>	
人类滑膜细胞细胞因子生成(IL8 释放)	E7
人类肺动脉平滑肌细胞(hPASMC)细胞因子生成(IL8 释放)	E8
人类骨骼肌细胞(hSKMC)细胞因子生成(IL8 释放)	E9
人类星形细胞细胞因子生成(IL8 释放)	E10
人类肝细胞(HepG2C3a 细胞)IL8 释放	E11
人急性早幼粒细胞白血病细胞/HL60(IL8 释放)	E12
人类成淋巴细胞(癌细胞系)/RPMI8226(IL8 释放)	E13
<b>TNF <math>\alpha</math>生成</b>	
人类滑膜细胞细胞因子生成(TNF $\alpha$ 释放)	E14
全血 TNF $\alpha$ 释放	E15
<b>IL10释放</b>	
人急性早幼粒细胞白血病细胞/HL60 IL10 释放	E16
人类原发性血液单核细胞(IL10 释放)	E17
<p>通过用测定孔中的数值除以测定板的平均PBS值来进行细胞因子释放测定的数据分析。如果测量值沿正向或负向偏离PBS值超过2SD, 则认为AARS多肽是细胞因子生成或细胞因子相关生物学的调节剂。在每个板上运行蛋白标准(针对每个测定试剂盒)以确保良好的测定质量。只有具有 &gt; 0.9的R<sup>2</sup>值的蛋白标准曲线的测定被选择进行数据分析。</p>	
<b>细胞粘附和趋化</b>	
测定描述	测定号
单核细胞 THP 1/人类脐静脉内皮细胞(HUVEC)细胞粘附	F1
人类肝细胞(HepG2C3a 细胞)(ICAM 释放)	F2
人类肺微血管内皮细胞(HLMVEC)细胞粘附调节(ICAM 释放)	F3
人类脐静脉内皮细胞(HUVEC)细胞粘附调节(VCAM释放)	F4
人类间充质干细胞(hMSC)细胞粘附调节(VCAM释放)	F5
人类骨骼肌细胞(hSKMC)细胞粘附调节(VCAM释放)	F6
人类肺动脉平滑肌细胞(hPASMC)细胞粘附调节(VCAM释放)	F7

[0757]

表E6

## 测定的关键和指示击中的标准

增殖测定	
<p>通过用测定孔中的数值除以测定板的平均PBS值来进行细胞粘附调节测定的数据分析。如果获得值沿正向或负向偏离PBS值超过2SD, 则认为AARS多肽是细胞粘附调节剂或细胞粘附相关生物学的调节剂。在ELISA测定的情况下, 在每个板上运行蛋白标准(针对每个测定试剂盒)以确保良好的测定质量。只有具有 <math>&gt; 0.9</math>的<math>R^2</math>值的蛋白标准曲线的测定被选择进行数据分析。</p>	
细胞分化	
测定描述	测定号
人类前脂肪细胞(hPAD)细胞分化	G1
人类骨骼肌(hSKMC)细胞分化	G2
人类间充质细胞(hMSC)细胞分化	G3
人类肺动脉平滑肌细胞(hPASMC)分化	G4
<p>通过用测定孔中的数值除以测定板的平均PBS值来进行细胞分化测定的数据分析。如在方法章节所述的, 基于特定抗体的荧光强度来评分分化测定。如果在给定处理的孔中, 对于特定标志物分化的强度值沿正向或负向偏离PBS值超过2SD, 则认为AARS多肽是细胞分化的调节剂。对于hSKMC分析, 拍摄所有孔的数字照片, 由三人以盲的方式对照片评分, 使用4分评分系统, 其中4分指示浓密的骨骼肌肌动蛋白染色和明显的肌管形成, 1分指示缺少任何分化或抑制分化。使用来自视觉评分的平均值, 只有平均值<math>&gt; 3</math>的孔被认为击中。该测定中分化对照处理孔通常评分<math>&gt; 2</math>, 而PBS处理孔评分<math>&lt; 2</math>。</p>	
细胞结合	
测定描述	测定号
PBMC	H1
原发性T细胞	H2
原发性B细胞	H3
原发性单核细胞	H4
HepG2	H5
3T3L1	H6
C2C12	H7
THP1	H8
Jurkat	H9
Raji	H10
<p>当抗体单独不具有明显信号时, 如果蛋白具有大于25%的细胞聚集的特定结合, 则认为</p>	

[0758]

[0759]

表E6			
测定的关键和指示击中的标准			
增殖测定			
AARS 多肽与特定细胞类型结合。			

[0760]

表 E7			
AARS 多肽的功能分析研究结果			
AARS 多肽	表位标签的位置	浓度 [nM]	测定击中
PheRSa1 <sup>N1</sup>	N-端	250	A7(增殖)
PheRSa1 <sup>N1</sup>	C-端	250	
PheRSa1 <sup>N2</sup>	N-端	250	E6(细胞因子释放) G2(分化)
PheRSa1 <sup>N2</sup>	C-端	250	E7(细胞因子释放) G1(分化) G2(分化)
PheRSa1 <sup>N3</sup>	C-端	250	C2(嗜中性粒细胞活化) G2(分化)
PheRSa1 <sup>N9</sup>	C-端	250	E4(细胞因子释放)

[0761] 结论, PheRSa1<sup>N1</sup>, PheRSa1<sup>N2</sup>, PheRSa1<sup>N3</sup>和PheRSa1<sup>N9</sup>显示是增殖、分化、细胞因子释放和嗜中性粒细胞活化的主要调节剂。

[0762] 当根据转录分析数据看时, 表型筛选数据证明, AARS多肽PheRSa1<sup>N1</sup>(氨基酸1-226)和PheRSa1<sup>N2</sup>(氨基酸1-219)确定了在许多表型筛选测定中高度活性的第一新的蛋白结构域的界限。注意在多种情况下, AARS多肽的N-端标记和C-端标记的版本特别在转录分析数据试验和表型筛选试验中的相对活性具有显著差别。该数据和这些AARS多肽的假设一致, 当AARS多肽是完整tRNA合成酶的一部分或者翻译性地融合另外蛋白时, 抑制新的生物活性, 但当AARS多肽具有游离氨基端或C端时, 展现该生物活性。

[0763] 因此, 结论是包含苯丙氨酰- $\alpha$ 亚单位tRNA合成酶的氨基酸1-226的AARS多肽确定了第一新的高度活性AARS多肽结构域的适当界限(即, 在约+/-5个氨基酸内), 即i) 高度功能活性, ii) 可以在大肠杆菌中容易地制备和生产, 和iii) 表现出有利的蛋白稳定性和聚集特性。

[0764] 本领域技术人员将理解, 包含少至苯丙氨酰- $\alpha$ 亚单位-tRNA合成酶的约前219个氨基酸到长至苯丙氨酰- $\alpha$ 亚单位-tRNA合成酶的约前226个氨基酸的任何AARS多肽代表本文描述的特定AARS多肽的功能等同物。

[0765] 表型筛选数据还证明, AARS多肽PheRSa1<sup>N3</sup>(氨基酸1-128+13aa)和PheRSa1<sup>N9</sup>(氨基酸1-152)确定了在许多表型筛选测定中高度活性的第二新的蛋白结构域的界限。

[0766] 因此, 结论是包含苯丙氨酰- $\alpha$ 亚单位tRNA合成酶的氨基酸1-152的AARS多肽确定了第二新的高度活性AARS多肽结构域的适当界限(即, 在约+/-5个氨基酸内), 即i) 高度功

能活性, ii) 可以在大肠杆菌中容易地制备和生产, 和 iii) 表现出有利的蛋白稳定性和聚集特性。

[0767] 本领域技术人员将理解, 包含少至苯丙氨酰- $\alpha$ 亚单位-tRNA合成酶的约氨基酸1-128到长至苯丙氨酰- $\alpha$ 亚单位-tRNA合成酶的氨基酸1-152的任何AARS多肽代表本文描述的特定AARS多肽的功能等同物。

[0768] 实施例13

[0769] 动物研究

[0770] 为了测试AARS多肽调节血细胞生成和炎性应答的能力, 根据通用方法所述, 将代表性AARS多肽PheRSa1<sup>N9</sup>施用至野生型和DIO小鼠。涉及细胞血液计数和细胞因子释放的结果分别示于表E8和E9。

表 E8 细胞结合研究										
AARS 多肽	细胞类型									
	PBMC	初级 T 细胞	初级 B 细胞	初级单核细胞	HepG2	3T3L1	C2C12	THP1	Jurkat	Raji
PheRSa1 <sup>N9</sup>	ND	0	0	0	ND	ND	ND	ND	ND	ND

表 E9 PheRSa1 <sup>N9</sup> 对于野生型和食物引导的肥胖(DIO)小鼠中血细胞生成的影响		
细胞类型	野生型小鼠	DIO 小鼠
白血细胞	0	0

表 E9 PheRSa1 <sup>N9</sup> 对于野生型和食物引导的肥胖(DIO)小鼠中血细胞生成的影响		
细胞类型	野生型小鼠	DIO 小鼠
嗜中性粒细胞	0	0
淋巴细胞	0	0
单核细胞	0	0
HGB	0	0
HCT	0	0
MCV	0	0
MCH	0	0
MCHC	0	0
血小板	0	0

表 E10 PheRSa1 <sup>N9</sup> 对于野生型和食物诱导的肥胖小鼠中细胞血液计数和细胞因子释放的影响			
正常饮食的 C57BL/6 小鼠			
[0774] AARS 多肽	CBC	急性细胞因子	慢性细胞因子
PheRSa1 <sup>N9</sup>	ND	12	2
高脂肪饮食的 C57BL/6 小鼠			
PheRSa1 <sup>N9</sup>	ND	8	2

[0775] 体内动物研究的结果显示,在野生型小鼠或和食物诱导的肥胖小鼠中,以10mg/kg 每日施用PheRSa1<sup>N9</sup>4天在全部动物中对于血细胞形成具有很小影响或没有影响。PheRSa1<sup>N9</sup>影响急性细胞因子释放,从而暗示在免疫功能和炎性应答中发挥作用。而且,动物研究证明,在该剂量下PheRSa1<sup>N9</sup>是安全和良好耐受的,并且在处理的动物中不产生过度不良反应。

## 序列表

- <110> aTyr Pharma Inc.  
Pangu Biopharma Limited  
Greene, Leslie Ann  
Chiang, Kyle P.  
Hong, Fei  
Vasserot, Alain P.  
Lo, Wing-Sze  
Watkins, Jeffrey D.  
Mendlein, John D.  
Quinn, Cheryl L.
- <120> 与苯丙氨酰-a-tRNA合成酶的蛋白片段相关的治疗、诊断和抗体组合物的创新发现
- <130> 120161.465PC
- <140> PCT  
<141> 2011-05-03
- <150> US 61/330,828  
<151> 2010-05-03
- <150> US 61/330,829  
<151> 2010-05-03
- [0001] <160> 96
- <170> FastSEQ for Windows Version 4.0
- <210> 1  
<211> 77  
<212> DNA  
<213> 人工序列
- <220>  
<223> 寡核苷酸
- <400> 1  
aggaggtaaa acatatgcat catcatcadc atcacggtaa gcctatccct aaccctttgc 60  
tcggtctcga ttctacg 77
- <210> 2  
<211> 12  
<212> DNA  
<213> 人工序列
- <220>  
<223> 寡核苷酸
- <400> 2  
taatgactcg ag 12
- <210> 3  
<211> 17

	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 3	
	aggagataaa acatatg	17
	<210> 4	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 4	
	aggaggtaaa acat	14
	<210> 5	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
[0002]	<400> 5	
	aggagataaa acat	14
	<210> 6	
	<211> 15	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 6	
	gaaggagata tacat	15
	<210> 7	
	<211> 72	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 7	
	ggtaagccta tcctaacc tctcctcggc ctcgatteta cgcaccacca tcataccat	60
	taatgactcg ag	72
	<210> 8	
	<211> 72	
	<212> DNA	

	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 8	
	catatgcatc atcatcatca tcacggtaag cctatcccta accctctcct cggctctcgat	60
	tctacgggat cc	72
	<210> 9	
	<211> 12	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 9	
	ctcgagtaat ga	12
	<210> 10	
	<211> 12	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
[0003]	<400> 10	
	catatgggat cc	12
	<210> 11	
	<211> 72	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 11	
	ctcgagggta agcctatccc taaccctctc ctcggtctcg attctacgea ccaccaccac	60
	caccactaat ga	72
	<210> 12	
	<211> 226	
	<212> PRT	
	<213> 智人(Homo sapiens)	
	<400> 12	
	Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala	
	1 5 10 15	
	Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met	
	20 25 30	
	Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly	
	35 40 45	
	Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr	
	50 55 60	



Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
 65 70 75 80  
 Phe Arg Ser Ile Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
 85 90 95  
 Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
 100 105 110  
 Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
 115 120 125  
 Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
 130 135 140  
 Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
 145 150 155 160  
 Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr Tyr Trp Val Ser  
 165 170 175  
 Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu  
 180 185 190  
 Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys  
 195 200 205  
 Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His Leu  
 210 215 220  
 His Pro  
 225

<210> 13  
 <211> 678  
 <212> DNA  
 <213> 智人

[0004]

<400> 13  
 atggcggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60  
 ctggacagcg ccgagttggc ggctgagctg ggcatggagc accaggcggg ggtgggcgcc 120  
 gtgaagagcc ttcaggcgct gggcgaggct atcgaggctg aacttcggtc caccaagcac 180  
 tgggagctta ctgcggaggg cgaggagatt gcccgggagg gcagccatga ggcccgtgtg 240  
 tttcgaagca ttccccaga gggcctggcc cagagcgagc ttatgcgact gcccagtggc 300  
 aaagtgggct tcagcaaggc catgtccaac aagtggattc gggtaggcaa gagtgccggt 360  
 gacgggcccc ggggtgttccg agtgggtggac agcatggagg atgaggtgca gcgggcggtc 420  
 cagctggtcc gggggggaca ggctgagaag ctgggggaga aggagaggag cgagctgagg 480  
 aagaggaagc tgttggtga agtgactctg aagacctact gggtaggcaa aggcagtgcc 540  
 tttagtacca gcatctcaa gcaagagaca gagctgagcc cagagatgat ctccagtggc 600  
 tcttgccggg accggcctt caagccctac aacttcttgg cccacggtgt cctccccgac 660  
 agcggccacc ttcaccgc 678

<210> 14  
 <211> 30  
 <212> PRT  
 <213> 小鼠(Mus musculus)

<400> 14  
 Arg Leu Glu Val Ala Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Thr  
 1 5 10 15  
 Gln Leu Gly Val Glu His Gln Ala Val Val Gly Ala Val Lys  
 20 25 30

<210> 15  
 <211> 14

<212> PRT  
<213> 小鼠

<400> 15  
Ser Leu Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg  
1 5 10

<210> 16  
<211> 83  
<212> PRT  
<213> 小鼠

<400> 16  
Ser Thr Lys Cys Trp Glu Leu Thr Thr Glu Gly Glu Glu Ile Ala Arg  
1 5 10 15  
Glu Gly Ser His Glu Ala Arg Val Phe Arg Ser Ile Pro Leu Glu Gly  
20 25 30  
Leu Val Gln Ser Glu Leu Met His Leu Pro Ser Gly Lys Val Gly Phe  
35 40 45  
Ser Lys Ala Met Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala  
50 55 60  
Asp Gly Pro Arg Val Phe Arg Val Val Asp Ser Ile Glu Asp Glu Val  
65 70 75 80  
Gln Lys Arg

[0005] <210> 17  
<211> 12  
<212> PRT  
<213> 小鼠

<400> 17  
Val Val Asp Ser Ile Glu Asp Glu Val Gln Lys Arg  
1 5 10

<210> 18  
<211> 127  
<212> PRT  
<213> 小鼠

<400> 18  
Arg Leu Glu Val Ala Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Thr  
1 5 10 15  
Gln Leu Gly Val Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu  
20 25 30  
Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys Cys  
35 40 45  
Trp Glu Leu Thr Thr Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His  
50 55 60  
Glu Ala Arg Val Phe Arg Ser Ile Pro Leu Glu Gly Leu Val Gln Ser  
65 70 75 80  
Glu Leu Met His Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met  
85 90 95  
Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg

Val Phe Arg Val Val Asp Ser Ile Glu Asp Glu Val Gln Lys Arg  
 100 105 110  
 115 120 125

<210> 19  
 <211> 141  
 <212> PRT  
 <213> 智人

<400> 19  
 Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45  
 Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
 50 55 60  
 Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
 65 70 75 80  
 Phe Arg Ser Ile Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
 85 90 95  
 Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
 100 105 110  
 Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
 115 120 125  
 Asp Ser Glu Asp Leu Leu Gly Glu Gln Arg Gln Cys Leu  
 130 135 140

[0006]

<210> 20  
 <211> 426  
 <212> DNA  
 <213> 智人

<400> 20  
 atggcggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60  
 ctggacagcg ccgagttggc ggctgagctg ggcatggagc accaggcggg ggtgggcgcc 120  
 gtgaagagcc ttcaggcgct gggcgaggct atcgaggctg aacttcggtc caccaagcac 180  
 tgggagctta ctgctggagg cgaggagatt gcccgaggag gcagccatga ggcccgtgtg 240  
 tttcgaagca ttccccaga gggcctggcc cagagcgagc ttatgcgact gccagtggtg 300  
 aaagtgggct tcagcaaggc catgtccaac aagtggattc gggaggacaa gaggcgggct 360  
 gacgggcccc ggggtgtccg agtggactct gaagacctac tgggtgagca aaggcagtg 420  
 ctttag 426

<210> 21  
 <211> 477  
 <212> PRT  
 <213> 智人

<400> 21  
 Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly

		35				40					45					
	Glu	Val	Ile	Glu	Ala	Glu	Leu	Arg	Ser	Thr	Lys	His	Trp	Glu	Leu	Thr
		50					55				60					
	Ala	Glu	Gly	Glu	Glu	Ile	Ala	Arg	Glu	Gly	Ser	His	Glu	Ala	Arg	Val
	65					70					75					80
	Phe	Arg	Ser	Ile	Pro	Glu	Gly	Leu	Ala	Gln	Ser	Glu	Leu	Met	Arg	
					85					90					95	
	Leu	Pro	Ser	Gly	Lys	Val	Gly	Phe	Ser	Lys	Ala	Met	Ser	Asn	Lys	Trp
				100					105					110		
	Ile	Arg	Val	Asp	Lys	Ser	Ala	Ala	Asp	Gly	Pro	Arg	Val	Phe	Arg	Val
				115					120					125		
	Val	Asp	Ser	Met	Glu	Asp	Glu	Val	Gln	Arg	Arg	Leu	Gln	Leu	Val	Arg
		130					135					140				
	Gly	Gly	Gln	Ala	Glu	Lys	Leu	Gly	Glu	Lys	Glu	Arg	Ser	Glu	Leu	Arg
	145					150					155					160
	Lys	Arg	Lys	Leu	Leu	Ala	Glu	Val	Gly	Ser	Trp	Arg	Asp	Arg	Pro	Phe
					165					170					175	
	Lys	Pro	Tyr	Asn	Phe	Leu	Ala	His	Gly	Val	Leu	Pro	Asp	Ser	Gly	His
				180					185					190		
	Leu	His	Pro	Leu	Leu	Lys	Val	Arg	Ser	Gln	Phe	Arg	Gln	Ile	Phe	Leu
			195					200					205			
	Glu	Met	Gly	Phe	Thr	Glu	Met	Pro	Thr	Asp	Asn	Phe	Ile	Glu	Ser	Ser
		210					215					220				
	Phe	Trp	Asn	Phe	Asp	Ala	Leu	Phe	Gln	Pro	Gln	Gln	His	Pro	Ala	Arg
	225					230					235					240
	Asp	Gln	His	Asp	Thr	Phe	Phe	Leu	Arg	Asp	Pro	Ala	Glu	Ala	Leu	Gln
					245					250					255	
[0007]	Leu	Pro	Met	Asp	Tyr	Val	Gln	Arg	Val	Lys	Arg	Thr	His	Ser	Gln	Gly
				260					265					270		
	Gly	Tyr	Gly	Ser	Gln	Gly	Tyr	Lys	Tyr	Asn	Trp	Lys	Leu	Asp	Glu	Ala
			275					280					285			
	Arg	Lys	Asn	Leu	Leu	Arg	Thr	His	Thr	Thr	Ser	Ala	Ser	Ala	Arg	Ala
		290					295					300				
	Leu	Tyr	Arg	Leu	Ala	Gln	Lys	Lys	Pro	Phe	Thr	Pro	Val	Lys	Tyr	Phe
	305					310					315					320
	Ser	Ile	Asp	Arg	Val	Phe	Arg	Asn	Glu	Thr	Leu	Asp	Ala	Thr	His	Leu
					325					330					335	
	Ala	Glu	Phe	His	Gln	Ile	Glu	Gly	Val	Val	Ala	Asp	His	Gly	Leu	Thr
				340					345					350		
	Leu	Gly	His	Leu	Met	Gly	Val	Leu	Arg	Glu	Phe	Phe	Thr	Lys	Leu	Gly
			355					360					365			
	Ile	Thr	Gln	Leu	Arg	Phe	Lys	Pro	Ala	Tyr	Asn	Pro	Tyr	Thr	Glu	Pro
		370					375					380				
	Ser	Met	Glu	Val	Phe	Ser	Tyr	His	Gln	Gly	Leu	Lys	Lys	Trp	Val	Glu
	385					390					395					400
	Val	Gly	Asn	Ser	Gly	Val	Phe	Arg	Pro	Glu	Met	Leu	Leu	Pro	Met	Gly
				405						410					415	
	Leu	Pro	Glu	Asn	Val	Ser	Val	Ile	Ala	Trp	Gly	Leu	Ser	Leu	Glu	Arg
				420					425					430		
	Pro	Thr	Met	Ile	Lys	Tyr	Gly	Ile	Asn	Asn	Ile	Arg	Glu	Leu	Val	Gly
			435					440					445			
	His	Lys	Val	Asn	Leu	Gln	Met	Val	Tyr	Asp	Ser	Pro	Leu	Cys	Arg	Leu
		450					455					460				
	Asp	Ala	Glu	Pro	Arg	Pro	Pro	Pro	Thr	Gln	Glu	Ala	Ala			
	465					470					475					

<210> 22  
 <211> 1434  
 <212> DNA  
 <213> 智人

<400> 22  
 atggcggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60  
 ctggacagcg ccgagttggc ggctgagctg ggcatggagc accaggcggg ggtgggcggc 120  
 gtgaagagcc ttcaggcget gggcgaggtc atcgaggctg aacttcggtc caccaagcac 180  
 tgggagctta ctgctggagg cgaggagatt gcccgggagg gcagccatga ggcccgtgtg 240  
 tttcgaagca ttccccaga gggcctggcc cagagcgagc ttatgcgact gcccagtggc 300  
 aaagtgggct tcagcaaggc catgtccaac aagtggattc ggtgggacaa gagtgcggct 360  
 gacgggcccc ggggtgttcc agtgggtggc agcatggagg atgagggtga gcggcggctc 420  
 cagctggtcc gggggggaca ggctgagaag ctgggggaga aggagaggag cgagctgagg 480  
 aagaggaagc tgttggctga agttggctct tggcgggacc ggcccttcaa gccctacaac 540  
 ttcttggccc acggtgtcct ccccgacagc ggccacctc acccgctgct caaggtccgc 600  
 tcccagttcc gacagatctt cctggagatg gggttcaccg agatgccgac tgataacttc 660  
 attgagagct ccttctggaa ctttgacgcc ctcttcacg cccagcagca cccagcccgt 720  
 gaccagcaag acaccttctt ccttcgagat ccagcggagg ccttcgagct cccaatggac 780  
 tatgtccagc ggggtcaagc gaccactct cagggcggct acggctcaca ggggtacaag 840  
 tataactgga agctggacga ggcccggaaa aacctactgc gaaccacac cacatcagcc 900  
 agcggcccgt cgctctaccg ccttgcccag aagaagcct tcaactcggg caagtacttc 960  
 tccatgacc gcgtattccg gaatgagacc ctggacgcca cgcacctggt tgagttccac 1020  
 cagctgagg gcgtgtggc ggatcatggt ctgacctgg gccacctcat gggcgttctg 1080  
 cgggagtctt tcaccaagct gggatcacg caactccgct tcaagccagc ctacaacca 1140  
 tacacagagc ccagcatgga ggtgttcagc taccaccaag gctgaagaa gtgggtggag 1200  
 gtcgaaact cgggggtctt cgtccagag atgctgctgc ccatggggct tcccgagaac 1260  
 gtgtcggtea ttgectgggg cctctccctg gagegcccac cgatgatcaa atatggcatc 1320  
 [0008] aacaatatec gggagctggg gggccacaag gtgaacctgc agatggtgta tgacagtccc 1380  
 ctgtgcccgc tggatgccga gccgaggccc cctccacac aggaggctgc gtga 1434

<210> 23  
 <211> 465  
 <212> PRT  
 <213> 智人

<400> 23  
 Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45  
 Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
 50 55 60  
 Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
 65 70 75 80  
 Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
 85 90 95  
 Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
 100 105 110  
 Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
 115 120 125  
 Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
 130 135 140  
 Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
 145 150 155 160

Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr Tyr Trp Val Ser  
 165 170 175  
 Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu  
 180 185 190  
 Ser Pro Glu Met Ile Ser Arg Phe Thr Glu Met Pro Thr Asp Asn Phe  
 195 200 205  
 Ile Glu Ser Ser Phe Trp Asn Phe Asp Ala Leu Phe Gln Pro Gln Gln  
 210 215 220  
 His Pro Ala Arg Asp Gln His Asp Thr Phe Phe Leu Arg Asp Pro Ala  
 225 230 235 240  
 Glu Ala Leu Gln Leu Pro Met Asp Tyr Val Gln Arg Val Lys Arg Thr  
 245 250 255  
 His Ser Gln Gly Gly Tyr Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys  
 260 265 270  
 Leu Asp Glu Ala Arg Lys Asn Leu Leu Arg Thr His Thr Thr Ser Ala  
 275 280 285  
 Ser Ala Arg Ala Leu Tyr Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro  
 290 295 300  
 Val Lys Tyr Phe Ser Ile Asp Arg Val Phe Arg Asn Glu Thr Leu Asp  
 305 310 315 320  
 Ala Thr His Leu Ala Glu Phe His Gln Ile Glu Gly Val Val Ala Asp  
 325 330 335  
 His Gly Leu Thr Leu Gly His Leu Met Gly Val Leu Arg Glu Phe Phe  
 340 345 350  
 Thr Lys Leu Gly Ile Thr Gln Leu Arg Phe Lys Pro Ala Tyr Asn Pro  
 355 360 365  
 Tyr Thr Glu Pro Ser Met Glu Val Phe Ser Tyr His Gln Gly Leu Lys  
 370 375 380  
 [0009] Lys Trp Val Glu Val Gly Asn Ser Gly Val Phe Arg Pro Glu Met Leu  
 385 390 395 400  
 Leu Pro Met Gly Leu Pro Glu Asn Val Ser Val Ile Ala Trp Gly Leu  
 405 410 415  
 Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn Ile Arg  
 420 425 430  
 Glu Leu Val Gly His Lys Val Asn Leu Gln Met Val Tyr Asp Ser Pro  
 435 440 445  
 Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro Thr Gln Glu Ala  
 450 455 460  
 Ala  
 465

<210> 24  
 <211> 1398  
 <212> DNA  
 <213> 智人

<400> 24  
 atggcggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60  
 ctggacagcg ccgagttggc ggctgagctg ggcattggagc accaggcggt ggtgggcgcc 120  
 gtgaagagcc ttcaggcgct gggcgaggtc atcgaggctg aacttcggtc caccaagcac 180  
 tgggagctta ctgctggagg cgaggagatt gcccgggagg gcagccatga ggcccgtgtg 240  
 tttcgaagca ttccccaga gggcctggcc cagagcgagc ttatgcgact gccagtggtg 300  
 aaagtgggct tcagcaaggc catgtccaac aagtggattc ggggtggacaa gagtgcggct 360  
 gacgggcccc ggggtgttccg agtgggtggac agcatggagg atgaggtgca gcggcggctc 420  
 cagctggtcc gggggggaca ggctgagaag ctgggggaga aggagaggag cgagctgagg 480  
 aagaggaagc tgttgctga agtgactctg aagacctact gggtagcaaa aggcagtgcc 540

```

tttagtacca gcatctccaa gcaagagaca gagctgagcc cagagatgat ctccaggttc 600
accgagatgc cgactgataa cttcattgag agctccttct ggaactttga cgccctcttc 660
cagccccagc agcaccagc cegtgaccag cagcacacct tcttccttcg agatccagcg 720
gaggccctgc agtcccaat ggactatgtc cagcgggtca agcggacca ctctcagggc 780
ggctacggct cacaggggta caagtataac tggagctgg acgaggcccc gaaaaacct 840
ctgccaacc acaccacatc agccagcgcc cgtgcgctct accgccttgc ccagaagaag 900
cccttcactc cggtaagta cttctccatc gaccgcgtat tccggaatga gaccctggac 960
gccacgcacc tggctgagtt ccaccagatc gagggcgtgg tggcggatca tggctctacc 1020
ttgggccacc tcatgggctg tctgcgggag ttcttcacca agctgggtat cagcgaactc 1080
cgcttcaagc cagcctacaa cccatacaca gagcccagca tggaggtgtt cagctaccac 1140
caaggcctga agaagtgggt ggaggtcgga aactcggggg tcttccttcc agagatgctg 1200
ctgcccattg ggcttcccga gaacgtgtcg gtcattgcct ggggcctctc cctggagcgc 1260
ccaacgatga tcaaatatgg catcaacaat atccgggagc tgggggcca caaggtgaac 1320
ctgcagatgg tgtatgacag tcccctgtgc cgcctggatg ccgagccgag gccccctcc 1380
acacaggagg ctgcgtga 1398

```

<210> 25

<211> 199

<212> PRT

<213> 智人

<400> 25

[0010] Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
1 5 10 15  
Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
20 25 30  
Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
35 40 45  
Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
50 55 60  
Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
65 70 75 80  
Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
85 90 95  
Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
100 105 110  
Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
115 120 125  
Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
130 135 140  
Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
145 150 155 160  
Lys Arg Lys Leu Leu Ala Glu Val Ser Ser Gly Gly Pro Ala Ala Pro  
165 170 175  
Asn Gly Leu Cys Pro Ala Gly Gln Ala Asp Pro Leu Ser Gly Arg Leu  
180 185 190  
Arg Leu Thr Gly Val Gln Val  
195

<210> 26

<211> 600

<212> DNA

<213> 智人

<400> 26

atggcggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60

```

ctggacagcg ccgagttggc ggctgagctg ggcattggagc accaggcggg ggtggggcgcc 120
gtgaagagcc ttcaggcgct gggcgaggtc atcgaggctg aacttcggtc caccaagcac 180
tgggagctta ctgcggaggg cgaggagatt gcccgggagg gcagccatga gggccgtgtg 240
tttcaagca tccccaga gggcctggcc cagagcgagc ttatgcgact gcccagtggc 300
aaagtgggct tcagcaaggc catgtccaac aagtggattc gggaggacaa gaggcggtg 360
gacgggcccc ggggtgttccg agtgggtggac agcatggagg atgagggtgca gggcggtgct 420
cagctggtcc ggggggggaca ggctgagaag ctgggggaga aggagaggag cgagctgagg 480
aagaggaagc tgttggtgta agtatccagc ggaggccctg cagctcccaa tgactatgt 540
ccagcgggtc aagcggacc actctcaggg cggctacggc tcacagggtt acaagtataa 600

```

<210> 27  
<211> 482  
<212> PRT  
<213> 智人

<400> 27

[0011]

```

Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala
1      5      10     15
Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met
20     25     30
Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly
35     40     45
Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr
50     55     60
Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val
65     70     75     80
Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg
85     90     95
Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp
100    105    110
Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val
115    120    125
Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg
130    135    140
Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg
145    150    155    160
Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr Tyr Trp Val Ser
165    170    175
Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu
180    185    190
Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys
195    200    205
Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His Leu
210    215    220
His Pro Leu Leu Lys Val Arg Ser Gln Phe Arg Gln Ile Phe Leu Glu
225    230    235    240
Met Gly Phe Thr Glu Met Pro Thr Asp Asn Phe Ile Glu Ser Ser Phe
245    250    255
Trp Asn Phe Asp Ala Leu Phe Gln Pro Gln Gln His Pro Ala Arg Asp
260    265    270
Gln His Asp Thr Phe Phe Leu Arg Asp Pro Ala Glu Ala Leu Gln Leu
275    280    285
Pro Met Asp Tyr Val Gln Arg Val Lys Arg Thr His Ser Gln Gly Gly
290    295    300
Tyr Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys Leu Asp Glu Ala Arg
305    310    315    320

```



Lys Asn Leu Leu Arg Thr His Thr Thr Ser Ala Ser Ala Arg Ala Leu  
325 330 335  
Tyr Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro Val Lys Tyr Phe Ser  
340 345 350  
Ile Asp Arg Val Phe Arg Asn Glu Thr Leu Asp Ala Thr His Leu Ala  
355 360 365  
Glu Phe His Gln Ile Glu Gly Val Val Ala Asp His Gly Leu Thr Leu  
370 375 380  
Gly His Leu Met Gly Val Leu Arg Glu Phe Phe Thr Lys Leu Gly Leu  
385 390 395 400  
Lys Lys Trp Val Glu Val Gly Asn Ser Gly Val Phe Arg Pro Glu Met  
405 410 415  
Leu Leu Pro Met Gly Leu Pro Glu Asn Val Ser Val Ile Ala Trp Gly  
420 425 430  
Leu Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn Ile  
435 440 445  
Arg Glu Leu Val Gly His Lys Val Asn Leu Gln Met Val Tyr Asp Ser  
450 455 460  
Pro Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro Pro Thr Gln Glu  
465 470 475 480  
Ala Ala

<210> 28  
<211> 1449  
<212> DNA  
<213> 智人

[0012]

<400> 28  
atggcgggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60  
ctggacagcgc ccgagttggc ggctgagctg ggcattggagc accaggcggg ggtggggcgcc 120  
gtgaagagacc ttcaggcgct gggcgaggct atcgaggctg aacttcggtc caccaagcac 180  
tgggagctta ctgctggagg cgaggagatt gcccgggagg gcagccatga ggcccgtgtg 240  
tttcaagca ttccccaga gggcctggcc cagagcgagc ttatgcgact gcccagtggc 300  
aaagtgggct tcagcaaggc catgtccaac aagtggattc ggttgacaa gagtgcggct 360  
gacgggcccc ggggtgttccg agtgggtggac agcatggagg atgagggtgca gggcgggctc 420  
cagctggtcc gggggggaca ggctgagaag ctgggggaga aggagaggag cgagctgagg 480  
aagaggaagc tgttggtga agtgactctg aagacctact ggttgagcaa aggcagtgcc 540  
tttagtacca gcatctccaa gcaagagaca gagctgagcc cagagatgat ctccagtggc 600  
tcttgccggg accggccctt caagccctac aacttcttgg cccacgggtg cctccccgac 660  
agcggccacc ttcaccgct gctcaaggtc cgtcccagt tccgacagat cttcctggag 720  
atggggttca cggagatgcc gactgataac ttcattgaga gtccttctg gaactttgac 780  
gccctcttcc agccccagca gcacccagcc cgtgaccagc acgacacctt cttccttcca 840  
gatccagcgg aggcctgca gtcaccaatg gactatgtcc agcgggtcaa gcgaccacc 900  
tctcagggcg gctacggctc acaggggtac aagtataact ggaagctgga cgaggcccgg 960  
aaaaacctac tgcaaccca caccacatca gccagcggcc gtgcgctcta ccgccttggc 1020  
cagaagaagc ccttcaactc ggtcaagtac ttctccatcg accgcgtatt ccggaatgag 1080  
accctggaag ccacgcacct ggctgagttc caccagatcg agggcgtggg ggcggatcat 1140  
ggtctcaact tgggccacct catgggcggt ctgcccggag tcttcccaa gctgggctctg 1200  
aagaagtggg tggaggtcgg aaactcgggg gtcttccgct cagagatgct gctgcccatt 1260  
gggcttcccg agaactgtgc ggtcattgcc tggggcctct cctggagcgc cccaacgatg 1320  
atcaaatatg gcatcaaaa tatccgggag ctggtgggccc acaaggtgaa cctgcagatg 1380  
gtgtatgaca gtcccctgtg ccgcctggat gccgagccga ggccccctcc cacacaggag 1440  
gctgcgtga 1449

<210> 29

<211> 273  
 <212> PRT  
 <213> 智人

<400> 29

Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45  
 Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
 50 55 60  
 Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
 65 70 75 80  
 Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
 85 90 95  
 Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
 100 105 110  
 Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
 115 120 125  
 Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
 130 135 140  
 Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
 145 150 155 160  
 Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr Tyr Trp Val Ser  
 165 170 175  
 Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu  
 180 185 190  
 Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys  
 195 200 205  
 Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His Leu  
 210 215 220  
 His Pro Leu Leu Lys Val Arg Ser Gln Phe Arg Gln Ile Phe Leu Glu  
 225 230 235 240  
 Met Gly Ser Ser Gly Gly Pro Ala Ala Pro Asn Gly Leu Cys Pro Ala  
 245 250 255  
 Gly Gln Ala Asp Pro Leu Ser Gly Arg Leu Arg Leu Thr Gly Val Gln  
 260 265 270  
 Val

[0013]

<210> 30  
 <211> 822  
 <212> DNA  
 <213> 智人

<400> 30

atggcggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60  
 ctggacagcg ccgagttggc ggctgagctg ggcattggagc accaggcggg ggtggggcgc 120  
 gtgaagagcc ttcaggcgct gggcgaggct atcgaggctg aacttcggtc caccaagcac 180  
 tgggagctta ctgctggagg cgaggagatt gcccgggagg gcagccatga ggcccgtgtg 240  
 tttcgaagca ttccccaga gggcctggcc cagagcgagc ttatgcgact gccagtggtg 300  
 aaagtgggct tcagcaagc catgtccaac aagtggattc ggggtggaca gagtgcggct 360  
 gacgggcccc ggggtgttcc agtgggtggac agcatggagg atgaggtgca gcggcggctc 420  
 cagctggtcc gggggggaca ggctgagaag ctgggggaga aggagaggag cgagctgagg 480

aagaggaagc tgttgctga agtgactctg aagacctact gggtgagcaa aggcagtgcc 540  
 tttagtacca gcatctccaa gcaagagaca gagctgagcc cagagatgat ctccagtggc 600  
 tcttgccggg accggccctt caagccctac aacttcttgg cccacgggtg cctccccgac 660  
 agcggccacc ttcaccgct gctcaaggtc cgctcccagt tccgacagat cttcctggag 720  
 atgggatcca gcgaggccc tgcagctccc aatggactat gtccagcggg tcaagcggac 780  
 ccactctcag ggcggctacg gctcacaggg gtacaagtat aa 822

<210> 31  
 <211> 50  
 <212> DNA  
 <213> 智人

<400> 31  
 tgacgggccc cgggtgttcc gagtggactc tgaagaccta ctgggtgagc 50

<210> 32  
 <211> 16  
 <212> PRT  
 <213> 智人

<400> 32  
 Asp Gly Pro Arg Val Phe Arg Val Asp Ser Glu Asp Leu Leu Gly Glu  
 1 5 10 15

<210> 33  
 <211> 50  
 <212> DNA  
 <213> 智人

[0014]

<400> 33  
 ggaagaggaa gctgttgct gaagttgct cttggcggga ccggcccttc 50

<210> 34  
 <211> 16  
 <212> PRT  
 <213> 智人

<400> 34  
 Lys Arg Lys Leu Leu Ala Glu Val Gly Ser Trp Arg Asp Arg Pro Phe  
 1 5 10 15

<210> 35  
 <211> 50  
 <212> DNA  
 <213> 智人

<400> 35  
 agctgagccc agagatgac tccaggttca ccgagatgcc gactgataac 50

<210> 36  
 <211> 16  
 <212> PRT  
 <213> 智人

<400> 36



<210> 43  
 <211> 219  
 <212> PRT  
 <213> 智人

<400> 43

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

[0016]

<210> 44  
 <211> 657  
 <212> DNA  
 <213> 智人

<400> 44

```

atggcggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60
ctggacacgc ccgagttggc ggctgagctg ggcatggagc accaggcggg ggtgggcgcc 120
gtgaagagcc ttcagggcgt gggcgaggct atcgaggctg aacttcggtc caccaagcac 180
tgggagctta ctgcggaggg cgaggagatt gcccgaggag gcagccatga ggcccgtgtg 240
tttcaagca ttccccaga gggcctggcc cagagcgagc ttatgcgact gcccagtggc 300
aaagtgggct tcagcaaggc catgtccaac aagtggattc ggtggacaa gagtgcggct 360
gacgggcccc ggggtgttcc agtggtggac agcatggagg atgaggtgca gcggcggctc 420
cagctggtcc gggggggaca ggctgagaag ctgggggaga aggagaggag cgagctgagg 480
aagaggaagc tgttggctga agtgactctg aagacctact ggtgagcaa aggcagtgcc 540
tttagtacca gcatctcaa gcaagagaca gagctgagcc cagagatgat ctccagtggc 600
tcttggcggg accggcctt caagccctac aacttcttgg cccacgggtg cctcccc 657
    
```

<210> 45  
 <211> 152

<212> PRT  
<213> 智人

<400> 45

```

Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala
 1          5          10          15
Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met
          20          25          30
Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly
          35          40          45
Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr
          50          55          60
Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val
65          70          75          80
Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg
          85          90          95
Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp
          100         105         110
Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val
          115         120         125
Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg
          130         135         140
Gly Gly Gln Ala Glu Lys Leu Gly
145          150

```

<210> 46  
<211> 456  
[0017] <212> DNA  
<213> 智人

<400> 46

```

atggcgggatg gtcaggtggc ggaactgctg ctccggcggc tggaggcgtc tgatggcggc 60
ctggacagcgc ccgagttggc ggctgagctg ggcattggagc accaggcggg ggtgggcggc 120
gtgaagagcc ttcaggcgtt gggcgaggct atcgaggctg aacttcggtc caccaagcac 180
tgaggagctta ctgaggagg cgaggagatt gcccgaggag gcagccatga ggcccgtgtg 240
tttcgaagca ttccccaga gggcctggcc cagagcgagc ttatgcgact gccagtggtg 300
aaagtgggct tcagcaaggc catgtccaac aagtggattc ggggtggaca gagtgcggct 360
gacgggcccc ggggtgttccg agtgggtggac agcatggagg atgaggtgca gcggcggctc 420
cagctggtcc gggggggaca ggctgagaag ctggggg          456

```

<210> 47  
<211> 246  
<212> PRT  
<213> 人工序列

<220>

<223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 47

```

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
 1          5          10          15
Gly Leu Asp Ser Thr Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg
          20          25          30
Arg Leu Glu Ala Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala
          35          40          45
Glu Leu Gly Met Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu

```

50	Gln	Ala	Leu	Gly	Glu	Val	Ile	Glu	Ala	Glu	Leu	Arg	Ser	Thr	Lys	His
65	Trp	Glu	Leu	Thr	Ala	Glu	Gly	Glu	Glu	Ile	Ala	Arg	Glu	Gly	Ser	His
				85	Phe	Arg	Ser	Ile	Pro	Pro	Glu	Gly	Leu	Ala	Gln	Ser
				100					105					110		
				115	Glu	Leu	Met	Arg	Leu	Pro	Ser	Gly	Lys	Val	Gly	Phe
				130	Ser	Asn	Lys	Trp	Ile	Arg	Val	Asp	Lys	Ser	Ala	Ala
				145	Val	Phe	Arg	Val	Val	Asp	Ser	Met	Glu	Asp	Glu	Val
				160	Gln	Leu	Val	Arg	Gly	Gly	Gln	Ala	Glu	Lys	Leu	Gly
				175					180							
				190	Ser	Glu	Leu	Arg	Lys	Arg	Lys	Leu	Leu	Ala	Glu	Val
				205	Tyr	Trp	Val	Ser	Lys	Gly	Ser	Ala	Phe	Ser	Thr	Ser
				220					225							
				240	Glu	Thr	Glu	Leu	Ser	Pro	Glu	Met	Ile	Ser	Ser	Gly
				245	Arg	Pro	Phe	Lys	Pro	Tyr	Asn	Phe	Leu	Ala	His	Gly
					Ser	Gly	His	Leu	His	Pro						

[0018] <210> 48  
 <211> 246  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 48	Met	Ala	Asp	Gly	Gln	Val	Ala	Glu	Leu	Leu	Leu	Arg	Arg	Leu	Glu	Ala
1				5					10					15		
	Ser	Asp	Gly	Gly	Leu	Asp	Ser	Ala	Glu	Leu	Ala	Ala	Glu	Leu	Gly	Met
				20					25					30		
	Glu	His	Gln	Ala	Val	Val	Gly	Ala	Val	Lys	Ser	Leu	Gln	Ala	Leu	Gly
				35					40					45		
	Glu	Val	Ile	Glu	Ala	Glu	Leu	Arg	Ser	Thr	Lys	His	Trp	Glu	Leu	Thr
				50					55					60		
	Ala	Glu	Gly	Glu	Glu	Ile	Ala	Arg	Glu	Gly	Ser	His	Glu	Ala	Arg	Val
65				70					75					80		
	Phe	Arg	Ser	Ile	Pro	Pro	Glu	Gly	Leu	Ala	Gln	Ser	Glu	Leu	Met	Arg
				85					90					95		
	Leu	Pro	Ser	Gly	Lys	Val	Gly	Phe	Ser	Lys	Ala	Met	Ser	Asn	Lys	Trp
				100					105					110		
	Ile	Arg	Val	Asp	Lys	Ser	Ala	Ala	Asp	Gly	Pro	Arg	Val	Phe	Arg	Val
				115					120					125		
	Val	Asp	Ser	Met	Glu	Asp	Glu	Val	Gln	Arg	Arg	Leu	Gln	Leu	Val	Arg
				130					135					140		
	Gly	Gly	Gln	Ala	Glu	Lys	Leu	Gly	Glu	Lys	Glu	Arg	Ser	Glu	Leu	Arg
145				150					155					160		
	Lys	Arg	Lys	Leu	Leu	Ala	Glu	Val	Thr	Leu	Lys	Thr	Tyr	Trp	Val	Ser

```

          165              170              175
Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu
          180              185              190
Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys
          195              200              205
Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His Leu
          210              215              220
His Pro Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr
225              230              235
His His His His His His
          245

```

<210> 49  
 <211> 239  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 49

```

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
 1              5              10              15
Gly Leu Asp Ser Thr Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg
          20              25              30
Arg Leu Glu Ala Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala
          35              40              45
[0019] Glu Leu Gly Met Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu
          50              55              60
Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His
65              70              75              80
Trp Glu Leu Thr Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His
          85              90              95
Glu Ala Arg Val Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser
          100              105              110
Glu Leu Met Arg Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met
          115              120              125
Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg
          130              135              140
Val Phe Arg Val Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu
145              150              155              160
Gln Leu Val Arg Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg
          165              170              175
Ser Glu Leu Arg Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr
          180              185              190
Tyr Trp Val Ser Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln
          195              200              205
Glu Thr Glu Leu Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp
          210              215              220
Arg Pro Phe Lys Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro
225              230              235

```

<210> 50  
 <211> 239  
 <212> PRT



<213> 人工序列

<220>

<223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 50

Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45  
 Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
 50 55 60  
 Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
 65 70 75 80  
 Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
 85 90 95  
 Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
 100 105 110  
 Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
 115 120 125  
 Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
 130 135 140  
 Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
 145 150 155 160  
 Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr Tyr Trp Val Ser  
 165 170 175  
 [0020] Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu  
 180 185 190  
 Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys  
 195 200 205  
 Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Gly Lys Pro Ile Pro  
 210 215 220  
 Asn Pro Leu Leu Gly Leu Asp Ser Thr His His His His His His  
 225 230 235

<210> 51

<211> 161

<212> PRT

<213> 人工序列

<220>

<223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 51

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg  
 20 25 30  
 Arg Leu Glu Ala Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala  
 35 40 45  
 Glu Leu Gly Met Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu  
 50 55 60  
 Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His  
 65 70 75 80

Trp Glu Leu Thr Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His  
85 90 95  
Glu Ala Arg Val Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser  
100 105 110  
Glu Leu Met Arg Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met  
115 120 125  
Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg  
130 135 140  
Val Phe Arg Val Asp Ser Glu Asp Leu Leu Gly Glu Gln Arg Gln Cys  
145 150 155 160  
Leu

<210> 52  
<211> 161  
<212> PRT  
<213> 人工序列

<220>  
<223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 52  
Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
1 5 10 15  
Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
20 25 30  
Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
35 40 45  
Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
50 55 60  
Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
65 70 75 80  
Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
85 90 95  
Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
100 105 110  
Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
115 120 125  
Asp Ser Glu Asp Leu Leu Gly Glu Gln Arg Gln Cys Leu Gly Lys Pro  
130 135 140  
Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr His His His His His  
145 150 155 160  
His

[0021]

<210> 53  
<211> 497  
<212> PRT  
<213> 人工序列

<220>  
<223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 53  
Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu

	1			5					10				15		
	Gly	Leu	Asp	Ser	Thr	Ala	Asp	Gly	Gln	Val	Ala	Glu	Leu	Leu	Arg
				20					25				30		
	Arg	Leu	Glu	Ala	Ser	Asp	Gly	Gly	Leu	Asp	Ser	Ala	Glu	Leu	Ala
			35				40					45			
	Glu	Leu	Gly	Met	Glu	His	Gln	Ala	Val	Val	Gly	Ala	Val	Lys	Ser
			50				55				60				
	Gln	Ala	Leu	Gly	Glu	Val	Ile	Glu	Ala	Glu	Leu	Arg	Ser	Thr	Lys
	65				70					75				80	
	Trp	Glu	Leu	Thr	Ala	Glu	Gly	Glu	Glu	Ile	Ala	Arg	Glu	Gly	Ser
				85					90					95	
	Glu	Ala	Arg	Val	Phe	Arg	Ser	Ile	Pro	Pro	Glu	Gly	Leu	Ala	Gln
			100						105					110	
	Glu	Leu	Met	Arg	Leu	Pro	Ser	Gly	Lys	Val	Gly	Phe	Ser	Lys	Ala
			115					120					125		
	Ser	Asn	Lys	Trp	Ile	Arg	Val	Asp	Lys	Ser	Ala	Ala	Asp	Gly	Pro
			130				135					140			
	Val	Phe	Arg	Val	Val	Asp	Ser	Met	Glu	Asp	Glu	Val	Gln	Arg	Arg
	145					150				155					
	Gln	Leu	Val	Arg	Gly	Gln	Ala	Glu	Lys	Leu	Gly	Glu	Lys	Glu	Arg
				165					170					175	
	Ser	Glu	Leu	Arg	Lys	Arg	Lys	Leu	Leu	Ala	Glu	Val	Gly	Ser	Trp
			180					185						190	
	Asp	Arg	Pro	Phe	Lys	Pro	Tyr	Asn	Phe	Leu	Ala	His	Gly	Val	Leu
			195					200					205		
	Asp	Ser	Gly	His	Leu	His	Pro	Leu	Leu	Lys	Val	Arg	Ser	Gln	Phe
			210				215					220			
	Gln	Ile	Phe	Leu	Glu	Met	Gly	Phe	Thr	Glu	Met	Pro	Thr	Asp	Asn
	225				230						235				
	Ile	Glu	Ser	Ser	Phe	Trp	Asn	Phe	Asp	Ala	Leu	Phe	Gln	Pro	Gln
				245					250					255	
	His	Pro	Ala	Arg	Asp	Gln	His	Asp	Thr	Phe	Phe	Leu	Arg	Asp	Pro
			260						265					270	
	Glu	Ala	Leu	Gln	Leu	Pro	Met	Asp	Tyr	Val	Gln	Arg	Val	Lys	Arg
			275					280					285		
	His	Ser	Gln	Gly	Gly	Tyr	Gly	Ser	Gln	Gly	Tyr	Lys	Tyr	Asn	Trp
			290				295					300			
	Leu	Asp	Glu	Ala	Arg	Lys	Asn	Leu	Leu	Arg	Thr	His	Thr	Thr	Ser
	305					310					315				
	Ser	Ala	Arg	Ala	Leu	Tyr	Arg	Leu	Ala	Gln	Lys	Lys	Pro	Phe	Thr
				325						330				335	
	Val	Lys	Tyr	Phe	Ser	Ile	Asp	Arg	Val	Phe	Arg	Asn	Glu	Thr	Leu
			340						345					350	
	Ala	Thr	His	Leu	Ala	Glu	Phe	His	Gln	Ile	Glu	Gly	Val	Val	Ala
			355					360					365		
	His	Gly	Leu	Thr	Leu	Gly	His	Leu	Met	Gly	Val	Leu	Arg	Glu	Phe
			370				375					380			
	Thr	Lys	Leu	Gly	Ile	Thr	Gln	Leu	Arg	Phe	Lys	Pro	Ala	Tyr	Asn
	385					390					395				
	Tyr	Thr	Glu	Pro	Ser	Met	Glu	Val	Phe	Ser	Tyr	His	Gln	Gly	Leu
				405						410				415	
	Lys	Trp	Val	Glu	Val	Gly	Asn	Ser	Gly	Val	Phe	Arg	Pro	Glu	Met
			420						425					430	
	Leu	Pro	Met	Gly	Leu	Pro	Glu	Asn	Val	Ser	Val	Ile	Ala	Trp	Gly
			435					440					445		
	Ser	Leu	Glu	Arg	Pro	Thr	Met	Ile	Lys	Tyr	Gly	Ile	Asn	Asn	Ile
			450				455						460		

[0022]

Glu Leu Val Gly His Lys Val Asn Leu Gln Met Val Tyr Asp Ser Pro  
 465 470 475 480  
 Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro Pro Thr Gln Glu Ala  
 485 490 495  
 Ala

<210> 54  
 <211> 497  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 54  
 Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45  
 Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
 50 55 60  
 Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
 65 70 75 80  
 Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
 85 90 95  
 Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
 100 105 110  
 Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
 115 120 125  
 Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
 130 135 140  
 Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
 145 150 155 160  
 Lys Arg Lys Leu Leu Ala Glu Val Gly Ser Trp Arg Asp Arg Pro Phe  
 165 170 175  
 Lys Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His  
 180 185 190  
 Leu His Pro Leu Leu Lys Val Arg Ser Gln Phe Arg Gln Ile Phe Leu  
 195 200 205  
 Glu Met Gly Phe Thr Glu Met Pro Thr Asp Asn Phe Ile Glu Ser Ser  
 210 215 220  
 Phe Trp Asn Phe Asp Ala Leu Phe Gln Pro Gln Gln His Pro Ala Arg  
 225 230 235 240  
 Asp Gln His Asp Thr Phe Phe Leu Arg Asp Pro Ala Glu Ala Leu Gln  
 245 250 255  
 Leu Pro Met Asp Tyr Val Gln Arg Val Lys Arg Thr His Ser Gln Gly  
 260 265 270  
 Gly Tyr Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys Leu Asp Glu Ala  
 275 280 285  
 Arg Lys Asn Leu Leu Arg Thr His Thr Thr Ser Ala Ser Ala Arg Ala  
 290 295 300  
 Leu Tyr Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro Val Lys Tyr Phe  
 305 310 315 320

[0023]

Ser Ile Asp Arg Val Phe Arg Asn Glu Thr Leu Asp Ala Thr His Leu  
 325 330 335  
 Ala Glu Phe His Gln Ile Glu Gly Val Val Ala Asp His Gly Leu Thr  
 340 345 350  
 Leu Gly His Leu Met Gly Val Leu Arg Glu Phe Phe Thr Lys Leu Gly  
 355 360 365  
 Ile Thr Gln Leu Arg Phe Lys Pro Ala Tyr Asn Pro Tyr Thr Glu Pro  
 370 375 380  
 Ser Met Glu Val Phe Ser Tyr His Gln Gly Leu Lys Lys Trp Val Glu  
 385 390 395 400  
 Val Gly Asn Ser Gly Val Phe Arg Pro Glu Met Leu Leu Pro Met Gly  
 405 410 415  
 Leu Pro Glu Asn Val Ser Val Ile Ala Trp Gly Leu Ser Leu Glu Arg  
 420 425 430  
 Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn Ile Arg Glu Leu Val Gly  
 435 440 445  
 His Lys Val Asn Leu Gln Met Val Tyr Asp Ser Pro Leu Cys Arg Leu  
 450 455 460  
 Asp Ala Glu Pro Arg Pro Pro Pro Thr Gln Glu Ala Ala Gly Lys Pro  
 465 470 475 480  
 Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr His His His His  
 485 490 495  
 His

[0024]

&lt;210&gt; 55

&lt;211&gt; 485

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

&lt;400&gt; 55

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Ala Asp Gly Gln Val Ala Glu Leu Leu Arg  
 20 25 30  
 Arg Leu Glu Ala Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala  
 35 40 45  
 Glu Leu Gly Met Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu  
 50 55 60  
 Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His  
 65 70 75 80  
 Trp Glu Leu Thr Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His  
 85 90 95  
 Glu Ala Arg Val Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser  
 100 105 110  
 Glu Leu Met Arg Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met  
 115 120 125  
 Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg  
 130 135 140  
 Val Phe Arg Val Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu  
 145 150 155 160  
 Gln Leu Val Arg Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg  
 165 170 175

Ser Glu Leu Arg Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr  
 180 185 190  
 Tyr Trp Val Ser Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln  
 195 200 205  
 Glu Thr Glu Leu Ser Pro Glu Met Ile Ser Arg Phe Thr Glu Met Pro  
 210 215 220  
 Thr Asp Asn Phe Ile Glu Ser Ser Phe Trp Asn Phe Asp Ala Leu Phe  
 225 230 235 240  
 Gln Pro Gln Gln His Pro Ala Arg Asp Gln His Asp Thr Phe Phe Leu  
 245 250 255  
 Arg Asp Pro Ala Glu Ala Leu Gln Leu Pro Met Asp Tyr Val Gln Arg  
 260 265 270  
 Val Lys Arg Thr His Ser Gln Gly Gly Tyr Gly Ser Gln Gly Tyr Lys  
 275 280 285  
 Tyr Asn Trp Lys Leu Asp Glu Ala Arg Lys Asn Leu Leu Arg Thr His  
 290 295 300  
 Thr Thr Ser Ala Ser Ala Arg Ala Leu Tyr Arg Leu Ala Gln Lys Lys  
 305 310 315 320  
 Pro Phe Thr Pro Val Lys Tyr Phe Ser Ile Asp Arg Val Phe Arg Asn  
 325 330 335  
 Glu Thr Leu Asp Ala Thr His Leu Ala Glu Phe His Gln Ile Glu Gly  
 340 345 350  
 Val Val Ala Asp His Gly Leu Thr Leu Gly His Leu Met Gly Val Leu  
 355 360 365  
 Arg Glu Phe Phe Thr Lys Leu Gly Ile Thr Gln Leu Arg Phe Lys Pro  
 370 375 380  
 Ala Tyr Asn Pro Tyr Thr Glu Pro Ser Met Glu Val Phe Ser Tyr His  
 385 390 395 400  
 [0025] Gln Gly Leu Lys Lys Trp Val Glu Val Gly Asn Ser Gly Val Phe Arg  
 405 410 415  
 Pro Glu Met Leu Leu Pro Met Gly Leu Pro Glu Asn Val Ser Val Ile  
 420 425 430  
 Ala Trp Gly Leu Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr Gly Ile  
 435 440 445  
 Asn Asn Ile Arg Glu Leu Val Gly His Lys Val Asn Leu Gln Met Val  
 450 455 460  
 Tyr Asp Ser Pro Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro Pro  
 465 470 475 480  
 Thr Gln Glu Ala Ala  
 485

<210> 56

<211> 485

<212> PRT

<213> 人工序列

<220>

<223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 56

Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45

[0026]

Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
50 55 60  
Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
65 70 75 80  
Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
85 90 95  
Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
100 105 110  
Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
115 120 125  
Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
130 135 140  
Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
145 150 155 160  
Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr Tyr Trp Val Ser  
165 170 175  
Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu  
180 185 190  
Ser Pro Glu Met Ile Ser Arg Phe Thr Glu Met Pro Thr Asp Asn Phe  
195 200 205  
Ile Glu Ser Ser Phe Trp Asn Phe Asp Ala Leu Phe Gln Pro Gln Gln  
210 215 220  
His Pro Ala Arg Asp Gln His Asp Thr Phe Phe Leu Arg Asp Pro Ala  
225 230 235 240  
Glu Ala Leu Gln Leu Pro Met Asp Tyr Val Gln Arg Val Lys Arg Thr  
245 250 255  
His Ser Gln Gly Gly Tyr Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys  
260 265 270  
Leu Asp Glu Ala Arg Lys Asn Leu Leu Arg Thr His Thr Thr Ser Ala  
275 280 285  
Ser Ala Arg Ala Leu Tyr Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro  
290 295 300  
Val Lys Tyr Phe Ser Ile Asp Arg Val Phe Arg Asn Glu Thr Leu Asp  
305 310 315 320  
Ala Thr His Leu Ala Glu Phe His Gln Ile Glu Gly Val Val Ala Asp  
325 330 335  
His Gly Leu Thr Leu Gly His Leu Met Gly Val Leu Arg Glu Phe Phe  
340 345 350  
Thr Lys Leu Gly Ile Thr Gln Leu Arg Phe Lys Pro Ala Tyr Asn Pro  
355 360 365  
Tyr Thr Glu Pro Ser Met Glu Val Phe Ser Tyr His Gln Gly Leu Lys  
370 375 380  
Lys Trp Val Glu Val Gly Asn Ser Gly Val Phe Arg Pro Glu Met Leu  
385 390 395 400  
Leu Pro Met Gly Leu Pro Glu Asn Val Ser Val Ile Ala Trp Gly Leu  
405 410 415  
Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn Ile Arg  
420 425 430  
Glu Leu Val Gly His Lys Val Asn Leu Gln Met Val Tyr Asp Ser Pro  
435 440 445  
Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro Pro Thr Gln Glu Ala  
450 455 460  
Ala Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr His  
465 470 475 480  
His His His His His  
485

<210> 57  
 <211> 219  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 57  
 Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg  
 20 25 30  
 Arg Leu Glu Ala Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala  
 35 40 45  
 Glu Leu Gly Met Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu  
 50 55 60  
 Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His  
 65 70 75 80  
 Trp Glu Leu Thr Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His  
 85 90 95  
 Glu Ala Arg Val Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser  
 100 105 110  
 Glu Leu Met Arg Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met  
 115 120 125  
 Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg  
 130 135 140  
 Val Phe Arg Val Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu  
 145 150 155 160  
 Gln Leu Val Arg Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg  
 165 170 175  
 Ser Glu Leu Arg Lys Arg Lys Leu Leu Ala Glu Val Ser Ser Gly Gly  
 180 185 190  
 Pro Ala Ala Pro Asn Gly Leu Cys Pro Ala Gly Gln Ala Asp Pro Leu  
 195 200 205  
 Ser Gly Arg Leu Arg Leu Thr Gly Val Gln Val  
 210 215

[0027]

<210> 58  
 <211> 219  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 58  
 Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45  
 Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
 50 55 60



Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
65 70 75 80  
Phe Arg Ser Ile Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
85 90 95  
Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
100 105 110  
Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
115 120 125  
Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
130 135 140  
Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
145 150 155 160  
Lys Arg Lys Leu Leu Ala Glu Val Ser Ser Gly Gly Pro Ala Ala Pro  
165 170 175  
Asn Gly Leu Cys Pro Ala Gly Gln Ala Asp Pro Leu Ser Gly Arg Leu  
180 185 190  
Arg Leu Thr Gly Val Gln Val Gly Lys Pro Ile Pro Asn Pro Leu Leu  
195 200 205  
Gly Leu Asp Ser Thr His His His His His His  
210 215

<210> 59  
<211> 502  
<212> PRT  
<213> 人工序列

[0028] <220>  
<223> 具有n-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 59  
Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
1 5 10 15  
Gly Leu Asp Ser Thr Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg  
20 25 30  
Arg Leu Glu Ala Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala  
35 40 45  
Glu Leu Gly Met Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu  
50 55 60  
Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His  
65 70 75 80  
Trp Glu Leu Thr Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His  
85 90 95  
Glu Ala Arg Val Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser  
100 105 110  
Glu Leu Met Arg Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met  
115 120 125  
Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg  
130 135 140  
Val Phe Arg Val Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu  
145 150 155 160  
Gln Leu Val Arg Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg  
165 170 175  
Ser Glu Leu Arg Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr  
180 185 190  
Tyr Trp Val Ser Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln  
195 200 205

Glu Thr Glu Leu Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp  
 210 215 220  
 Arg Pro Phe Lys Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp  
 225 230 235 240  
 Ser Gly His Leu His Pro Leu Leu Lys Val Arg Ser Gln Phe Arg Gln  
 245 250 255  
 Ile Phe Leu Glu Met Gly Phe Thr Glu Met Pro Thr Asp Asn Phe Ile  
 260 265 270  
 Glu Ser Ser Phe Trp Asn Phe Asp Ala Leu Phe Gln Pro Gln Gln His  
 275 280 285  
 Pro Ala Arg Asp Gln His Asp Thr Phe Phe Leu Arg Asp Pro Ala Glu  
 290 295 300  
 Ala Leu Gln Leu Pro Met Asp Tyr Val Gln Arg Val Lys Arg Thr His  
 305 310 315 320  
 Ser Gln Gly Gly Tyr Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys Leu  
 325 330 335  
 Asp Glu Ala Arg Lys Asn Leu Leu Arg Thr His Thr Thr Ser Ala Ser  
 340 345 350  
 Ala Arg Ala Leu Tyr Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro Val  
 355 360 365  
 Lys Tyr Phe Ser Ile Asp Arg Val Phe Arg Asn Glu Thr Leu Asp Ala  
 370 375 380  
 Thr His Leu Ala Glu Phe His Gln Ile Glu Gly Val Val Ala Asp His  
 385 390 395 400  
 Gly Leu Thr Leu Gly His Leu Met Gly Val Leu Arg Glu Phe Phe Thr  
 405 410 415  
 Lys Leu Gly Leu Lys Lys Trp Val Glu Val Gly Asn Ser Gly Val Phe  
 420 425 430  
 [0029] Arg Pro Glu Met Leu Leu Pro Met Gly Leu Pro Glu Asn Val Ser Val  
 435 440 445  
 Ile Ala Trp Gly Leu Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr Gly  
 450 455 460  
 Ile Asn Asn Ile Arg Glu Leu Val Gly His Lys Val Asn Leu Gln Met  
 465 470 475 480  
 Val Tyr Asp Ser Pro Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro  
 485 490 495  
 Pro Thr Gln Glu Ala Ala  
 500

<210> 60

<211> 502

<212> PRT

<213> 人工序列

<220>

<223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 60

Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45  
 Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
 50 55 60

Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
65 70 75 80  
Phe Arg Ser Ile Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
85 90 95  
Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
100 105 110  
Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
115 120 125  
Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
130 135 140  
Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
145 150 155 160  
Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr Tyr Trp Val Ser  
165 170 175  
Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu  
180 185 190  
Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys  
195 200 205  
Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His Leu  
210 215 220  
His Pro Leu Leu Lys Val Arg Ser Gln Phe Arg Gln Ile Phe Leu Glu  
225 230 235 240  
Met Gly Phe Thr Glu Met Pro Thr Asp Asn Phe Ile Glu Ser Ser Phe  
245 250 255  
Trp Asn Phe Asp Ala Leu Phe Gln Pro Gln Gln His Pro Ala Arg Asp  
260 265 270  
Gln His Asp Thr Phe Phe Leu Arg Asp Pro Ala Glu Ala Leu Gln Leu  
275 280 285  
Pro Met Asp Tyr Val Gln Arg Val Lys Arg Thr His Ser Gln Gly Gly  
290 295 300  
Tyr Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys Leu Asp Glu Ala Arg  
305 310 315 320  
Lys Asn Leu Leu Arg Thr His Thr Thr Ser Ala Ser Ala Arg Ala Leu  
325 330 335  
Tyr Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro Val Lys Tyr Phe Ser  
340 345 350  
Ile Asp Arg Val Phe Arg Asn Glu Thr Leu Asp Ala Thr His Leu Ala  
355 360 365  
Glu Phe His Gln Ile Glu Gly Val Val Ala Asp His Gly Leu Thr Leu  
370 375 380  
Gly His Leu Met Gly Val Leu Arg Glu Phe Phe Thr Lys Leu Gly Leu  
385 390 395 400  
Lys Lys Trp Val Glu Val Gly Asn Ser Gly Val Phe Arg Pro Glu Met  
405 410 415  
Leu Leu Pro Met Gly Leu Pro Glu Asn Val Ser Val Ile Ala Trp Gly  
420 425 430  
Leu Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn Ile  
435 440 445  
Arg Glu Leu Val Gly His Lys Val Asn Leu Gln Met Val Tyr Asp Ser  
450 455 460  
Pro Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro Pro Thr Gln Glu  
465 470 475 480  
Ala Ala Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr  
485 490 495  
His His His His His His  
500

[0030]

<210> 61  
 <211> 293  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 61  
 Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg  
 20 25 30  
 Arg Leu Glu Ala Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala  
 35 40 45  
 Glu Leu Gly Met Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu  
 50 55 60  
 Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His  
 65 70 75 80  
 Trp Glu Leu Thr Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His  
 85 90 95  
 Glu Ala Arg Val Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser  
 100 105 110  
 Glu Leu Met Arg Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met  
 115 120 125  
 Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg  
 130 135 140  
 Val Phe Arg Val Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu  
 145 150 155 160  
 Gln Leu Val Arg Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg  
 165 170 175  
 Ser Glu Leu Arg Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr  
 180 185 190  
 Tyr Trp Val Ser Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln  
 195 200 205  
 Glu Thr Glu Leu Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp  
 210 215 220  
 Arg Pro Phe Lys Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp  
 225 230 235 240  
 Ser Gly His Leu His Pro Leu Leu Lys Val Arg Ser Gln Phe Arg Gln  
 245 250 255  
 Ile Phe Leu Glu Met Gly Ser Ser Gly Gly Pro Ala Ala Pro Asn Gly  
 260 265 270  
 Leu Cys Pro Ala Gly Gln Ala Asp Pro Leu Ser Gly Arg Leu Arg Leu  
 275 280 285  
 Thr Gly Val Gln Val  
 290

[0031]

<210> 62  
 <211> 293  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 62  
 Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala  
 1 5 10 15  
 Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met  
 20 25 30  
 Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly  
 35 40 45  
 Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr  
 50 55 60  
 Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val  
 65 70 75 80  
 Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg  
 85 90 95  
 Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp  
 100 105 110  
 Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val  
 115 120 125  
 Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg  
 130 135 140  
 Gly Gly Gln Ala Glu Lys Leu Gly Glu Lys Glu Arg Ser Glu Leu Arg  
 145 150 155 160  
 Lys Arg Lys Leu Leu Ala Glu Val Thr Leu Lys Thr Tyr Trp Val Ser  
 165 170 175  
 Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu  
 180 185 190  
 Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys  
 195 200 205  
 [0032] Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His Leu  
 210 215 220  
 His Pro Leu Leu Lys Val Arg Ser Gln Phe Arg Gln Ile Phe Leu Glu  
 225 230 235 240  
 Met Gly Ser Ser Gly Gly Pro Ala Ala Pro Asn Gly Leu Cys Pro Ala  
 245 250 255  
 Gly Gln Ala Asp Pro Leu Ser Gly Arg Leu Arg Leu Thr Gly Val Gln  
 260 265 270  
 Val Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr His  
 275 280 285  
 His His His His His  
 290

<210> 63

<211> 172

<212> PRT

<213> 人工序列

<220>

<223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 63

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg  
 20 25 30  
 Arg Leu Glu Ala Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala

```

          35          40          45
Glu Leu Gly Met Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu
  50          55          60
Gln Ala Leu Gly Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His
65          70          75          80
Trp Glu Leu Thr Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His
          85          90          95
Glu Ala Arg Val Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser
  100          105          110
Glu Leu Met Arg Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met
  115          120          125
Ser Asn Lys Trp Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg
  130          135          140
Val Phe Arg Val Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu
145          150          155          160
Gln Leu Val Arg Gly Gly Gln Ala Glu Lys Leu Gly
          165          170

```

<210> 64  
 <211> 172  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

[0033]

```

<400> 64
Met Ala Asp Gly Gln Val Ala Glu Leu Leu Leu Arg Arg Leu Glu Ala
  1          5          10          15
Ser Asp Gly Gly Leu Asp Ser Ala Glu Leu Ala Ala Glu Leu Gly Met
  20          25          30
Glu His Gln Ala Val Val Gly Ala Val Lys Ser Leu Gln Ala Leu Gly
  35          40          45
Glu Val Ile Glu Ala Glu Leu Arg Ser Thr Lys His Trp Glu Leu Thr
  50          55          60
Ala Glu Gly Glu Glu Ile Ala Arg Glu Gly Ser His Glu Ala Arg Val
65          70          75          80
Phe Arg Ser Ile Pro Pro Glu Gly Leu Ala Gln Ser Glu Leu Met Arg
          85          90          95
Leu Pro Ser Gly Lys Val Gly Phe Ser Lys Ala Met Ser Asn Lys Trp
  100          105          110
Ile Arg Val Asp Lys Ser Ala Ala Asp Gly Pro Arg Val Phe Arg Val
  115          120          125
Val Asp Ser Met Glu Asp Glu Val Gln Arg Arg Leu Gln Leu Val Arg
  130          135          140
Gly Gly Gln Ala Glu Lys Leu Gly Gly Lys Pro Ile Pro Asn Pro Leu
145          150          155          160
Leu Gly Leu Asp Ser Thr His His His His His His
          165          170

```

<210> 65  
 <211> 176  
 <212> PRT  
 <213> 人工序列

<220>

<223> 具有C-端8xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 65

Met	Ala	Asp	Gly	Gln	Val	Ala	Glu	Leu	Leu	Leu	Arg	Arg	Leu	Glu	Ala
1				5				10						15	
Ser	Asp	Gly	Gly	Leu	Asp	Ser	Ala	Glu	Leu	Ala	Ala	Glu	Leu	Gly	Met
		20						25					30		
Glu	His	Gln	Ala	Val	Val	Gly	Ala	Val	Lys	Ser	Leu	Gln	Ala	Leu	Gly
		35					40					45			
Glu	Val	Ile	Glu	Ala	Glu	Leu	Arg	Ser	Thr	Lys	His	Trp	Glu	Leu	Thr
		50					55				60				
Ala	Glu	Gly	Glu	Glu	Ile	Ala	Arg	Glu	Gly	Ser	His	Glu	Ala	Arg	Val
65					70					75				80	
Phe	Arg	Ser	Ile	Pro	Pro	Glu	Gly	Leu	Ala	Gln	Ser	Glu	Leu	Met	Arg
				85					90					95	
Leu	Pro	Ser	Gly	Lys	Val	Gly	Phe	Ser	Lys	Ala	Met	Ser	Asn	Lys	Trp
			100					105					110		
Ile	Arg	Val	Asp	Lys	Ser	Ala	Ala	Asp	Gly	Pro	Arg	Val	Phe	Arg	Val
		115					120					125			
Val	Asp	Ser	Met	Glu	Asp	Glu	Val	Gln	Arg	Arg	Leu	Gln	Leu	Val	Arg
		130					135				140				
Gly	Gly	Gln	Ala	Glu	Lys	Leu	Gly	Leu	Glu	Gly	Lys	Pro	Ile	Pro	Asn
145					150					155					160
Pro	Leu	Leu	Gly	Leu	Asp	Ser	Thr	His	His	His	His	His	His	His	His
				165					170						175

[0034]

<210> 66

<211> 687

<212> DNA

<213> 人工序列

<220>

<223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

<400> 66

gctgatggtc	aagtcgctga	actgctgctg	agacgtctgg	aggcgagcga	cggtggtttg	60
gacagcgcag	aactggcagc	agagctgggc	atggagcacc	aggcggttgt	gggcgctgtc	120
aaatccctgc	aggcactggg	tgaagttatt	gaggcggagc	tgcgttcgac	caagcactgg	180
gagctgacgg	cggagggtga	agaaatcgct	cgcgaaagta	gccatgaagc	gcgtgtcttt	240
cgcagcatcc	cgccggaggg	tctggcccaa	agcgagctga	tgcgtttgcc	gtctggtaaa	300
gttggcttta	gcaaggccat	gagcaataaa	tggatccgtg	ttgacaagag	cgccgcccag	360
ggcccgcgtg	tgttccgctg	tgtggatagc	atggaagatg	aggtccagcg	ccgcctgcag	420
ctggctcgtg	gcggtcaagc	ggaaaagctg	ggtgagaaag	aacgttccga	gctgcgtaag	480
cgtaaattgc	tggcggaggt	gactctgaaa	acctactggg	tgtctaaggg	cagcgcattt	540
agcacgtcta	ttagcaaaaca	agaaaccgag	ctgagcccag	agatgattag	ctccggtagc	600
tggcgcgatac	gtccgttcaa	gccgtataac	ttcctggcgc	acggcgttct	gcctgatagc	660
ggtcacctgc	atccgtaatg	actcgag				687

<210> 67

<211> 666

<212> DNA

<213> 人工序列

<220>

<223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

<400> 67  
gctgatggtc aagtcgctga actgctgctg agacgtctgg aggcctctga cgggtggtctg 60  
gatagcgcgg agttggccgc ggagctgggc atggaacatc aggcagttgt gggcgcagtt 120  
aagagcttgc aggcgctggg cgaagtattt gaggcagaac tgcgttcac gaaacattgg 180  
gagctgaccg ccgagggcga agaaattgct cgtgagggtt cccacgaagc gcgtgtgttc 240  
cgttctatcc cgccggaggg tctggcgcaa agcgaactga tgcgcctgcc gagcggtaag 300  
gtgggcttta gcaaagcaat gagcaacaag tggatccgtg tcgataagag cgcagcggat 360  
ggtccgcgcg tgtttcgcgt ggttgactcg atggaggacg aggtccaacg tcgtctgcag 420  
ctgggtccgtg gcggtcaagc ggagaagtgg ggtgagaaaag aacgtagcga actgcgcaaa 480  
cgtaagctgc tggcggaagt caccctgaaa acgtactggg tcagcaaagg cagcgccttt 540  
agcaccagca tcagcaaaaa ggagactgag ctgagcccgg agatgatcag ctcggtttcc 600  
tggcgcgacc gccctttcaa gccatataat ttcttggtc acggtgtttc gccgtaatga 660  
ctcgag 666

<210> 68  
<211> 432  
<212> DNA  
<213> 人工序列

<220>  
<223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

[0035] <400> 68  
gctgatggtc aagtcgctga actgctgctg agacgtctgg aggcagcga cgggtggtctg 60  
gacagcgcgg agctggcagc ggaactgggt atggagcatc aagcggtcgt tggcgcggtt 120  
aaatccctgc aggccctggg tgaagtgatt gaagcagaac tgcgcagcac caaacactgg 180  
gagctgaccg ccgagggcga agaaatcgc cgtgagggca gccacgaggc gcgtgtcttt 240  
cgcagcatcc cgccggaggg tctggcgcag agcgagctga tgcgcctgcc ttctggtaag 300  
gtgggcttca gcaaggcaat gagcaaaaa tggattcgtg tcgacaagtc cgcagctgat 360  
ggtccgcgctg tgttccgtgt tgattctgag gatttgttgg gcgaacagcg tcaatgctg 420  
taatgactcg ag 432

<210> 69  
<211> 1440  
<212> DNA  
<213> 人工序列

<220>  
<223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

<400> 69  
gctgatggtc aagtcgctga actgctgctg agacgtctgg aggcagcga tgggtggtctg 60  
gactctgcgg agctggccgc agagctgggt atggagcacc aggcggtcgt tggcgcggtt 120  
aaaagcctgc aagctctggg cgaggtgatc gaggccgagc tgcgttcgac gaaacattgg 180  
gaattgacgg ctgagggcga agaaatcgc cgtgaaggta gccacgaggc acgtgttttc 240  
cgttccatcc cgccagaagg tctggcgcag agcgaactga tgcgtctgcc tagcggcaaa 300  
gtgggcttca gcaaggccat gagcaacaag tggattcgcg tcgataaaaag cgcagccgat 360  
ggtccgcgcg ttttccgtgt ggtggacagc atggaggacg aagtgcaacg ccgtctgcag 420  
ctgggtccgcg gcggtcaagc ggagaaactg ggtgagaaaag aacgctctga actgcgtaag 480  
cgtaagctgc tggcggagggt cggcagctgg cgcgaccgcc cgttcaaacc gtataatfff 540  
ctggcgcacg gtgttctgcc agacagcggc cactgcacc cgtgtttgaa ggtgcttcc 600  
cagtttcgtc agatctttct ggagatgggt ttaccgaga tgcgaccga caacttcatt 660  
gaaagctcgt tctggaattt tgatgcgctg ttcaaccgc agcagcacc ggctcgtgac 720  
cagcatgata cgttcttctt gcgtgatccg gcggaagctc tgcagttgcc gatggactac 780  
gtccagcgcg ttaaagcac ccatagccaa ggtggttacg gttctcaagg ttataaatat 840  
aactggaagc tggatgagc gcgtaagaat ctgttgcga cgcacacgac ttccgcatcc 900



```

gcgcgtgcgc tgtaccgtct ggctcagaag aaaccgttta ccccggttaa atatttcagc 960
atcgaccgcg tctttcgtaa tgaaacctct gatgcaactc atctggcaga gttccacca 1020
attgaggcgc ttgtggcggg tcatggcctg accctgggcc atctgatggg cgttctgcgc 1080
gagttcttca cgaagtggg tattaccaa ctgcgtttca aaccggccta caaccctat 1140
accgaaccgt cgatggaggt ctttagctac caccaaggtt tgaagaagtg ggtcgaagtc 1200
ggcaatagcg gtgtgtttcg tcctgagatg ttgctgccga tgggtttgcc ggaaaacgtt 1260
agcgtgattg catggggcct gagcctgga cgtccgacca tgattaagta cggtatcaat 1320
aacatccgtg agctggtggg tcacaaagtt aacctgcaga tgggttacga cagcccgtg 1380
tgccgtcttg acgccgaacc acgcccgcg cctaccaag aggcggcgta atgactcgag 1440

```

<210> 70  
 <211> 1404  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

[0036]

```

<400> 70
gctgatggtc aagtcgctga actgctgctg agacgtcttg aggcctccga cggcggctctg 60
gacagcgcgg agttggcggc ggaactgggt atggaacacc aagcagtggg cgggtgccgtt 120
aagagcctgc aagcgtggg tgaggtgatt gaagcagagc tgcgtagcac gaaacattgg 180
gagctgaccg ccgaggggta agagatcgcc cgcaaggca gccacgaggc tcgtgttttc 240
cgtagcattc cgccggaggg tctggcgcag agcgaactga tgcgtctgcc gagcggtaaa 300
gttggtttta gcaaagcgt gtctaacaag tggatccgcg tggataaaag cgcggccgac 360
ggtcgcgctg tctttcgcgt tctcagacgc atggaagatg aagtccaacg tcgtctgcag 420
ttgggtgcgt gcggtcaggc ggagaagctg ggcgagaaag agcgttcgga gttgcgtaaa 480
cgtaaactgc tggcggaagt taccctgaaa acttattggg tcagcaaggg tagcgcattc 540
tctaccagca tctccaaaca ggaactgag ctgagcccgg aaatgatcag ccgtttcacg 600
gagatgccga ccgataactt tattgaatcg tcttctgga atttcgatgc gctgtttcag 660
ccgcaacaac atccggcacg cgatcaacat gacaccttct ttctgcgcga cccggcggaa 720
gcaactgcaac tgccaatgga ctacgttcag cgtgtgaagc gcacccattc gcagggcggc 780
tacggtagcc agggttataa gtataattgg aagctggatg aggcacgtaa gaatctgctg 840
cgtaccata ccacctcggc gtctgcgcgt gcgctgtacc gcctggcaca aaagaaaccg 900
tttacgccgg ttaagtactt cagcattgat cgtgtgttcc gtaatgaaac gctggacgcc 960
accacactgg ccgagttcca ccagatcgag ggcgtcgtcg ctgaccacgg cttgacgctg 1020
ggtcacctga tgggtgtgct gcgtgagttc ttacgaagt tgggtatcac tcagctgcgc 1080
tttaaaccgg catataatcc gtacaccgag ccgagcatgg aggtctttag ctaccaccaa 1140
ggtttgaaga aatgggtgga ggtgggcaac agcggcgttt tccgccctga gatgttctg 1200
ccgatgggcc tgccggaaaa cgttagcgtg attgcttggg gcctgtccct ggaacgcccg 1260
accatgatta agtacggtat caataacatt cgtgaactgg ttggtcacia agtgaacctg 1320
caaatggttt atgatagccc gctgtgccgc ctggacgcgg agccacgtcc tccaccgacg 1380
caggaagcag cttaatgact cgag 1404

```

<210> 71  
 <211> 606  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

```

<400> 71
gctgatggtc aagtcgctga actgctgctg agacgtcttg aggcgagcga cggcggctctg 60
gacagcgcgt agctggcggc agaactgggc atggaacacc aagcagtcgt tggcgcgcgtc 120
aaaagcttgc aggcgtggg cgaggttatt gaggcggaac tgcgcagcac caagcactgg 180

```

gagctgacgg cggagggcga agaaatcgcc cgcgaaggca gccatgaagc acgtgtgttc 240  
 cgctcgattc cgccggaagg tctggcgagc agcgagctga tgcgtttgcc gtctggtaaa 300  
 gtgggtttca gcaaggctat gtccaacaaa tggatccgtg ttgataaatc cgcagcagac 360  
 ggtccacgtg tctttcgcgt cgttgatagc atggaagatg aggttcaacg tcgtctgcaa 420  
 ctggttcgtg gtggtcaagc ggagaagctg ggtgagaaag agcgctccga gttgcgtaag 480  
 cgtaagtgc tggcggaggt gtctagcggc ggtccggcgg ctccgaatgg tctgtgcccg 540  
 gccggtcagg cggacctctt gagcggtcgc ctgcgtctga ccggcgtgca ggtgtaatga 600  
 ctcgag 606

<210> 72

<211> 1455

<212> DNA

<213> 人工序列

<220>

<223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

<400> 72

gctgatggtc aagtcgctga actgctgctg agacgtcttg aggcgagcga tgggtggtctg 60  
 gactcggcag aactggctgc cgagctgggt atggagcacc aggcggctcgt tggcgcggtc 120  
 aaaagcctgc aagctttggg tgaagttatt gaggcggaac tgcgcagcac gaaacactgg 180  
 gagctgaccg cagaggcga agaaattgca cgtgaaggta gccacgaggc ccgtgtcttt 240  
 cgttccattc cgccagaggg tttggcccag agcgaactga tgcgtctgcc gagcggcaag 300  
 gtgggtttct cgaaggccat gtccaacaag tggattcgtg tggacaaatc cgccgcagat 360  
 ggcccgcgtg tgtttcgcgt cgttgacagc atggaggac aggtgcaacg tcgcctgcag 420  
 ctggtccgtg gggccaagc ggagaaactg ggtgagaaag aacgtagcga gctgcgtaag 480  
 cgtaagctgc tggctgaagt taccctgaaa acctactggg tgtctaaagg tagcgcgttt 540  
 tctacgagca tctctaaaca ggagactgaa ctgagcccgg aaatgattag cagcggcagc 600  
 [0037] tggcgcgacc gccattcaa gccgtacaac tttctggctc acggcgtgct gccggactcc 660  
 ggccacctgc atccgctgct gaaagttcgt agccagttcc gtcaaactct cctggagatg 720  
 ggcttcaccg aatgccgac cgataatttc atcgagagca gcttctggaa cttcgacgca 780  
 ctgtttcagc cgcagcagca tccggcgcgt gaccagcacg ataccttctt tctgcgtgat 840  
 ccggctgaag cgtgcaact gccgatggat tacgttcagc gcgtcaagcg cacgcacagc 900  
 caaggtggct acggtagcca gggctataag tacaattgga aactggatga ggcgcgcaag 960  
 aatctgttgc gcaccatac taccagcgcg agcgcgcgtg cgctgtaccg tctggctcaa 1020  
 aagaaaccgt ttacgccagt taaatatttc agcattgacc gcgtgtttcg taatgaaacc 1080  
 ttggatgcaa cgcacctggc ggagtttcat caaatcgagg gtgttgtggc agaccacggt 1140  
 ctgacctggt gtcactgat ggggtgtgctg cgcgaattct ttacgaagtt gggctctgaag 1200  
 aaatgggttg aggttggtaa tagcgggtgtt ttccgcctg agatgttgtt gccgatgggt 1260  
 ctgccgaaa acgtgtctgt catcgcgtgg ggctgagcc tggagcgtcc gaccatgatc 1320  
 aagtatggta ttaacaatat ccgtgagctg gtgggccata aagtcaactt gcaaatgggt 1380  
 tatgacagcc cgtgtgccc cctggatgcg gagccgcgtc ctccgccgac ccaggaagca 1440  
 gcctaagac tcgag 1455

<210> 73

<211> 828

<212> DNA

<213> 人工序列

<220>

<223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

<400> 73

gctgatggtc aagtcgctga actgctgctg agacgtcttg aggccagcga cggcggcttg 60  
 gacagcgcgg aattggccgc ggaactgggt atggaacacc aagctgttgt tgggtgcagtg 120  
 aagtcgctgc aggcactggg tgaagtcatt gaagctgaac tgcgtagcac caagcactgg 180  
 gagctgactg cggaggcga agaaatcgca cgtgaaggca gccatgagcc gcgtgttttc 240

```

cgcagcattc cgccggaggg tctggctcag agcgagctga tgcgcttgcc gtccggtaag 300
gttggtttta gcaaagccat gagcaacaaa tggatccgcg tggacaaaag cgcggctgat 360
ggtccacgtg tgtttcgtgt cgtggacagc atggaggatg aggtccaacg tcgcctgcag 420
ctggttcgtg gcggccaagc ggaaaagctg ggtgagaaag agcgtagcga gctgcgtaaa 480
cgtaaactgc tggcggaggt gacgctgaaa acgtattggg tgtccaaggg ttctgcgttc 540
agcaccagca ttagcaagca ggaaaccgag ctgtccccgg agatgatcag cagcggtagc 600
tggcgcgacc gtccgttcaa accgtacaat tttctggcgc acggcgttct gccggatagc 660
ggccacctgc atccgctggt gaaggtccgc tcgcaattcc gtcagatctt tctggagatg 720
ggttctagcg gtggtccggc agcaccgaac ggtttgtgcc cggcgggtca ggcggatcct 780
ctgtccggcc gtctgcgctt gaccggcgtc caagtgtaat gactcgag 828

```

<210> 74  
<211> 465  
<212> DNA  
<213> 人工序列

<220>  
<223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

```

<400> 74
gctgatggtc aagtcgctga actgctgctg cgccgtcttg aggcacgga tggcggctctg 60
gacagcgtag agctggcggc ggaactgggt atggagcacc aagccgtggt gggtagcgtt 120
aaatccctgc aagccctggg tgaggtcatt gaggctgagt tgcgtagcac gaaacattgg 180
gagctgaccg ccgaagtgta agaaatcgcg cgtgaaggta gccacgagc acgcgttttc 240
cgtagcattc cgccagaggg tctggcgcaa agcgagctga tgcgtctgcc gagcggcaag 300
gtcggtttct ccaaagcgat gagcaacaag tggatccgcg tggataaatc tgcagcggac 360
ggtccgcgcg tctttcgtgt cgttgacagc atggaagatg aggtgcagcg tcgtctgcag 420
ctggttcgcg gcggccaggc tgaaaagttg ggctaataac tcgag 465

```

[0038]

<210> 75  
<211> 393  
<212> PRT  
<213> 智人

```

<400> 75
Met Arg Pro Val Cys Phe Glu Ala Phe Pro Gln Arg Ala Trp Pro Arg
1          5          10          15
Ala Ser Leu Cys Asp Cys Pro Val Ala Lys Trp Ala Ser Ala Arg Pro
20          25          30
Cys Pro Thr Ser Gly Phe Gly Trp Arg Val Arg Leu Thr Gly Pro
35          40          45
Gly Cys Ser Glu Trp Thr Leu Lys Thr Tyr Trp Val Ser Lys Gly Ser
50          55          60
Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu Leu Ser Pro Glu
65          70          75          80
Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys Pro Tyr Asn
85          90          95
Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His Leu His Pro Leu
100         105         110
Leu Lys Val Arg Ser Gln Phe Arg Gln Ile Phe Leu Glu Met Gly Phe
115         120         125
Thr Glu Met Pro Thr Asp Asn Phe Ile Glu Ser Ser Phe Trp Asn Phe
130         135         140
Asp Ala Leu Phe Gln Pro Gln Gln His Pro Ala Arg Asp Gln His Asp
145         150         155         160
Thr Phe Phe Leu Arg Asp Pro Ala Glu Ala Leu Gln Leu Pro Met Asp
165         170         175

```

Tyr Val Gln Arg Val Lys Arg Thr His Ser Gln Gly Gly Tyr Gly Ser  
 180 185 190  
 Gln Gly Tyr Lys Tyr Asn Trp Lys Leu Asp Glu Ala Arg Lys Asn Leu  
 195 200 205  
 Leu Arg Thr His Thr Thr Ser Ala Ser Ala Arg Ala Leu Tyr Arg Leu  
 210 215 220  
 Ala Gln Lys Lys Pro Phe Thr Pro Val Lys Tyr Phe Ser Ile Asp Arg  
 225 230 235 240  
 Val Phe Arg Asn Glu Thr Leu Asp Ala Thr His Leu Ala Glu Phe His  
 245 250 255  
 Gln Ile Glu Gly Val Val Ala Asp His Gly Leu Thr Leu Gly His Leu  
 260 265 270  
 Met Gly Val Leu Arg Glu Phe Phe Thr Lys Leu Gly Ile Thr Gln Leu  
 275 280 285  
 Arg Phe Lys Pro Ala Tyr Asn Pro Tyr Thr Glu Pro Ser Met Glu Val  
 290 295 300  
 Phe Ser Tyr His Gln Gly Leu Lys Lys Trp Val Glu Val Gly Asn Ser  
 305 310 315 320  
 Gly Val Phe Arg Pro Glu Met Leu Leu Pro Met Gly Leu Pro Glu Asn  
 325 330 335  
 Val Ser Val Ile Ala Trp Gly Leu Ser Leu Glu Arg Pro Thr Met Ile  
 340 345 350  
 Lys Tyr Gly Ile Asn Asn Ile Arg Glu Leu Val Gly His Lys Val Asn  
 355 360 365  
 Leu Gln Met Val Tyr Asp Ser Pro Leu Cys Arg Leu Asp Ala Glu Pro  
 370 375 380  
 Arg Pro Pro Pro Thr Gln Glu Ala Ala  
 385 390

[0039]

<210> 76  
 <211> 1182  
 <212> DNA  
 <213> 智人

<400> 76  
 atgaggcccg tgtgtttcga agcattcccc cagagggcct ggcccagagc gagcttatgc 60  
 gactgcccag tggcaaagtg ggcttcagca aggccatgtc caacaagtgg attcgggtgg 120  
 acaagagtgc ggctgacggg ccccggtgt tccgagtgga ctctgaagac ctactgggtg 180  
 agcaaaggca gtgcctttag taccagcatc tccaagcaag agacagagct gagcccagag 240  
 atgatctcca gtggctcttg ggggaccgg cccttcaagc cctacaactt cttggcccac 300  
 ggtgtcctcc cggacagcgg ccaccttca cggctgctca aggtccgctc ccagttccga 360  
 cagatcttcc tggagatggg gttcaccgag atgccgactg ataacttcat tgagagctcc 420  
 ttctggaact ttgacgcctt cttccagccc cagcagcacc cagcccgtga ccagcacgac 480  
 accttcttcc ttcgagatcc agcggaggcc ctgcagctcc caatggacta tgtccagcgg 540  
 gtcaagcggg cccactctca gggcggctac ggctcacagg ggtacaagta taactggaag 600  
 ctggacgagg cccggaaaaa cctactgcca acccacacca catcagccag cgcccgtgcg 660  
 ctctaccgcc ttgcccagaa gaagccctt actccggcca agtacttctc catcgaccgc 720  
 gtattccgga atgagaccct ggacgccac cacctggctg agttccacca gatcgagggc 780  
 gtgggtggcgg atcatggtct cacctgggc cacctcatgg gcgttctgcg ggagttcttc 840  
 accaagctgg gtatcacgca actccgctt aagccagcct acaaccata cacagagccc 900  
 agcatggagg tgttcagcta ccaccaaggc ctgaagaagt ggggtggagg cggaaactcg 960  
 ggggtcttcc gtccagagat gctgctgcc atggggcttc ccgagaacgt gtcggtcatt 1020  
 gcctggggcc tctccctgga gcgcccacg atgatcaaat atggcatcaa caatatccgg 1080  
 gagctggtgg gccacaagg gaaactgcag atgggtgatg acagtcccct gtgccgctg 1140  
 gatgccgagc cgaggccccc tcccacacag gaggtgcgt ga 1182

<210> 77  
 <211> 219  
 <212> PRT  
 <213> 智人

<400> 77

Met Asp Tyr Val Gln Arg Val Lys Arg Thr His Ser Gln Gly Gly Tyr  
 1 5 10 15  
 Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys Leu Asp Glu Ala Arg Lys  
 20 25 30  
 Asn Leu Leu Arg Thr His Thr Thr Ser Ala Ser Ala Arg Ala Leu Tyr  
 35 40 45  
 Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro Val Lys Tyr Phe Ser Ile  
 50 55 60  
 Asp Arg Val Phe Arg Asn Glu Thr Leu Asp Ala Thr His Leu Ala Glu  
 65 70 75 80  
 Phe His Gln Ile Glu Gly Val Val Ala Asp His Gly Leu Thr Leu Gly  
 85 90 95  
 His Leu Met Gly Val Leu Arg Glu Phe Thr Lys Leu Gly Ile Thr  
 100 105 110  
 Gln Leu Arg Phe Lys Pro Ala Tyr Asn Pro Tyr Thr Glu Pro Ser Met  
 115 120 125  
 Glu Val Phe Ser Tyr His Gln Gly Leu Lys Lys Trp Val Glu Val Gly  
 130 135 140  
 Asn Ser Gly Val Phe Arg Pro Glu Met Leu Leu Pro Met Gly Leu Pro  
 145 150 155 160  
 Glu Asn Val Ser Val Ile Ala Trp Gly Leu Ser Leu Glu Arg Pro Thr  
 165 170 175  
 [0040] Met Ile Lys Tyr Gly Ile Asn Asn Ile Arg Glu Leu Val Gly His Lys  
 180 185 190  
 Val Asn Leu Gln Met Val Tyr Asp Ser Pro Leu Cys Arg Leu Asp Ala  
 195 200 205  
 Glu Pro Arg Pro Pro Pro Thr Gln Glu Ala Ala  
 210 215

<210> 78  
 <211> 660  
 <212> DNA  
 <213> 智人

<400> 78

atggactatg tccagcgggt caagcggacc cactctcagg gcggctacgg ctcacagggg 60  
 tacaagtata actggaagct ggacgaggcc cggaaaaacc tactgcgaac ccacaccaca 120  
 tcagccagcg cccgtgcgct ctaccgcctt gcccagaaga agcccttcac tccgggtcaag 180  
 tactttctcca tcgaccgcgt attccggaat gagaccctgg acgccacgca cctggctgag 240  
 ttccaccaga tcgagggcgt ggtggcggat catggtctca ccttgggcca cctcatgggc 300  
 gttctgcggg agttcttcac caagctgggt atcacgcaac tccgcttcaa gccagcctac 360  
 aaccataca cagagcccag catggagggtg ttcagctacc accaaggcct gaagaagtgg 420  
 gtggaggteg gaaactcggg ggtcttccgt ccagagatgc tgctgccat ggggcttccc 480  
 gagaacgtgt cggtcattgc ctggggcctc tccttgagc gcccaacgat gatcaaatat 540  
 ggcataca atatecggga gctgggtggc cacaaggtga acctgcagat ggtgtatgac 600  
 agtcccctgt gccgcctgga tgccgagccg aggccccctc ccacacagga ggctgcgtga 660

<210> 79  
 <211> 50

<212> DNA  
<213> 智人

<400> 79  
tgacgggccc cgggtgttcc gagtggactc tgaagaccta ctgggtgagc 50

<210> 80  
<211> 16  
<212> PRT  
<213> 智人

<400> 80  
Thr Gly Pro Gly Cys Ser Glu Trp Thr Leu Lys Thr Tyr Trp Val Ser  
1 5 10 15

<210> 81  
<211> 50  
<212> DNA  
<213> 智人

<400> 81  
ggaagaggaa gctgttgct gaagtatcca gcggaggccc tgcagctccc 50

<210> 82  
<211> 50  
<212> PRT  
<213> 智人

[0041]

<400> 82  
Leu Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn Ile  
1 5 10 15  
Arg Glu Leu Val Gly His Lys Val Asn Leu Gln Met Val Tyr Asp Ser  
20 25 30  
Pro Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro Pro Thr Gln Glu  
35 40 45  
Ala Ala  
50

<210> 83  
<211> 153  
<212> DNA  
<213> 智人

<400> 83  
ctctccctgg agcgcceaac gatgatcaaa tatggcatca acaatatccg ggagctggtg 60  
ggccacaagg tgaacctgca gatggtgtat gacagteccc tgtgccgcct ggatgccgag 120  
ccgaggcccc ctcccacaca ggaggctgcg tga 153

<210> 84  
<211> 71  
<212> PRT  
<213> 人工序列

<220>  
<223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 84  
 Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Leu Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr  
 20 25 30  
 Gly Ile Asn Asn Ile Arg Glu Leu Val Gly His Lys Val Asn Leu Gln  
 35 40 45  
 Met Val Tyr Asp Ser Pro Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro  
 50 55 60  
 Pro Pro Thr Gln Glu Ala Ala  
 65 70

<210> 85  
 <211> 71  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

[0042] <400> 85  
 Met Leu Ser Leu Glu Arg Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn  
 1 5 10 15  
 Ile Arg Glu Leu Val Gly His Lys Val Asn Leu Gln Met Val Tyr Asp  
 20 25 30  
 Ser Pro Leu Cys Arg Leu Asp Ala Glu Pro Arg Pro Pro Thr Gln  
 35 40 45  
 Glu Ala Ala Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser  
 50 55 60  
 Thr His His His His His His  
 65 70

<210> 86  
 <211> 413  
 <212> PRT  
 <213> 人工序列

<220>  
 <223> 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 86  
 Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15  
 Gly Leu Asp Ser Thr Arg Pro Val Cys Phe Glu Ala Phe Pro Gln Arg  
 20 25 30  
 Ala Trp Pro Arg Ala Ser Leu Cys Asp Cys Pro Val Ala Lys Trp Ala  
 35 40 45  
 Ser Ala Arg Pro Cys Pro Thr Ser Gly Phe Gly Trp Thr Arg Val Arg  
 50 55 60  
 Leu Thr Gly Pro Gly Cys Ser Glu Trp Thr Leu Lys Thr Tyr Trp Val  
 65 70 75 80  
 Ser Lys Gly Ser Ala Phe Ser Thr Ser Ile Ser Lys Gln Glu Thr Glu  
 85 90 95  
 Leu Ser Pro Glu Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe

100 105 110  
 Lys Pro Tyr Asn Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His  
 115 120 125  
 Leu His Pro Leu Leu Lys Val Arg Ser Gln Phe Arg Gln Ile Phe Leu  
 130 135 140  
 Glu Met Gly Phe Thr Glu Met Pro Thr Asp Asn Phe Ile Glu Ser Ser  
 145 150 155 160  
 Phe Trp Asn Phe Asp Ala Leu Phe Gln Pro Gln Gln His Pro Ala Arg  
 165 170 175  
 Asp Gln His Asp Thr Phe Phe Leu Arg Asp Pro Ala Glu Ala Leu Gln  
 180 185 190  
 Leu Pro Met Asp Tyr Val Gln Arg Val Lys Arg Thr His Ser Gln Gly  
 195 200 205  
 Gly Tyr Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys Leu Asp Glu Ala  
 210 215 220  
 Arg Lys Asn Leu Leu Arg Thr His Thr Thr Ser Ala Ser Ala Arg Ala  
 225 230 235 240  
 Leu Tyr Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro Val Lys Tyr Phe  
 245 250 255  
 Ser Ile Asp Arg Val Phe Arg Asn Glu Thr Leu Asp Ala Thr His Leu  
 260 265 270  
 Ala Glu Phe His Gln Ile Glu Gly Val Val Ala Asp His Gly Leu Thr  
 275 280 285  
 Leu Gly His Leu Met Gly Val Leu Arg Glu Phe Phe Thr Lys Leu Gly  
 290 295 300  
 Ile Thr Gln Leu Arg Phe Lys Pro Ala Tyr Asn Pro Tyr Thr Glu Pro  
 305 310 315 320  
 Ser Met Glu Val Phe Ser Tyr His Gln Gly Leu Lys Lys Trp Val Glu  
 325 330 335  
 Val Gly Asn Ser Gly Val Phe Arg Pro Glu Met Leu Leu Pro Met Gly  
 340 345 350  
 Leu Pro Glu Asn Val Ser Val Ile Ala Trp Gly Leu Ser Leu Glu Arg  
 355 360 365  
 Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn Ile Arg Glu Leu Val Gly  
 370 375 380  
 His Lys Val Asn Leu Gln Met Val Tyr Asp Ser Pro Leu Cys Arg Leu  
 385 390 395 400  
 Asp Ala Glu Pro Arg Pro Pro Pro Thr Gln Glu Ala Ala  
 405 410

[0043]

&lt;210&gt; 87

&lt;211&gt; 413

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

&lt;400&gt; 87

Met Arg Pro Val Cys Phe Glu Ala Phe Pro Gln Arg Ala Trp Pro Arg  
 1 5 10 15  
 Ala Ser Leu Cys Asp Cys Pro Val Ala Lys Trp Ala Ser Ala Arg Pro  
 20 25 30  
 Cys Pro Thr Ser Gly Phe Gly Trp Thr Arg Val Arg Leu Thr Gly Pro  
 35 40 45  
 Gly Cys Ser Glu Trp Thr Leu Lys Thr Tyr Trp Val Ser Lys Gly Ser



50 Ala Phe Ser Thr Ser Ile 55 Ser Lys Gln Glu Thr 60 Glu Leu Ser Pro Glu  
 65 Met Ile Ser Ser Gly Ser Trp Arg Asp Arg Pro Phe Lys Pro Tyr Asn 80  
 85 Phe Leu Ala His Gly Val Leu Pro Asp Ser Gly His Leu His Pro Leu 95  
 100 Leu Lys Val Arg Ser Gln Phe Arg Gln Ile Phe Leu Glu Met Gly Phe 110  
 115 Thr Glu Met Pro Thr Asp Asn Phe Ile Glu Ser Ser Phe Trp Asn Phe 125  
 130 Asp Ala Leu Phe Gln Pro Gln Gln His Pro Ala Arg Asp Gln His Asp 140  
 145 Thr Phe Phe Leu Arg Asp Pro Ala Glu Ala Leu Gln Leu Pro Met Asp 155  
 165 Tyr Val Gln Arg Val Lys Arg Thr His Ser Gln Gly Gly Tyr Gly Ser 170  
 180 Gln Gly Tyr Lys Tyr Asn Trp Lys Leu Asp Glu Ala Arg Lys Asn Leu 185  
 195 Leu Arg Thr His Thr Thr Ser Ala Ser Ala Arg Ala Leu Tyr Arg Leu 200  
 210 Ala Gln Lys Lys Pro Phe Thr Pro Val Lys Tyr Phe Ser Ile Asp Arg 205  
 225 Val Phe Arg Asn Glu Thr Leu Asp Ala Thr His Leu Ala Glu Phe His 215  
 230 Gln Ile Glu Gly Val Val Ala Asp His Gly Leu Thr Leu Gly His Leu 235  
 245 Met Gly Val Leu Arg Glu Phe Phe Thr Lys Leu Gly Ile Thr Gln Leu 250  
 255 Arg Phe Lys Pro Ala Tyr Asn Pro Tyr Thr Glu Pro Ser Met Glu Val 265  
 260 Phe Ser Tyr His Gln Gly Leu Lys Lys Trp Val Glu Val Gly Asn Ser 270  
 275 Gly Val Phe Arg Pro Glu Met Leu Leu Pro Met Gly Leu Pro Glu Asn 285  
 280 Val Ser Val Ile Ala Trp Gly Leu Ser Leu Glu Arg Pro Thr Met Ile 290  
 295 Lys Tyr Gly Ile Asn Asn Ile Arg Glu Leu Val Gly His Lys Val Asn 300  
 305 Leu Gln Met Val Tyr Asp Ser Pro Leu Cys Arg Leu Asp Ala Glu Pro 315  
 310 Arg Pro Pro Pro Thr Gln Glu Ala Ala Gly Lys Pro Ile Pro Asn Pro 320  
 325 Leu Leu Gly Leu Asp Ser Thr His His His His His His 330  
 335 340 345 350 355 360 365 370 375 380 385 390 395 400  
 405 410

[0044]

&lt;210&gt; 88

&lt;211&gt; 239

&lt;212&gt; PRT

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 具有N-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

&lt;400&gt; 88

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu

```

      1           5           10           15
Gly Leu Asp Ser Thr Asp Tyr Val Gln Arg Val Lys Arg Thr His Ser
      20           25           30
Gln Gly Gly Tyr Gly Ser Gln Gly Tyr Lys Tyr Asn Trp Lys Leu Asp
      35           40           45
Glu Ala Arg Lys Asn Leu Leu Arg Thr His Thr Thr Ser Ala Ser Ala
      50           55           60
Arg Ala Leu Tyr Arg Leu Ala Gln Lys Lys Pro Phe Thr Pro Val Lys
      65           70           75           80
Tyr Phe Ser Ile Asp Arg Val Phe Arg Asn Glu Thr Leu Asp Ala Thr
      85           90           95
His Leu Ala Glu Phe His Gln Ile Glu Gly Val Val Ala Asp His Gly
      100          105          110
Leu Thr Leu Gly His Leu Met Gly Val Leu Arg Glu Phe Phe Thr Lys
      115          120          125
Leu Gly Ile Thr Gln Leu Arg Phe Lys Pro Ala Tyr Asn Pro Tyr Thr
      130          135          140
Glu Pro Ser Met Glu Val Phe Ser Tyr His Gln Gly Leu Lys Lys Trp
      145          150          155          160
Val Glu Val Gly Asn Ser Gly Val Phe Arg Pro Glu Met Leu Leu Pro
      165          170          175
Met Gly Leu Pro Glu Asn Val Ser Val Ile Ala Trp Gly Leu Ser Leu
      180          185          190
Glu Arg Pro Thr Met Ile Lys Tyr Gly Ile Asn Asn Ile Arg Glu Leu
      195          200          205
Val Gly His Lys Val Asn Leu Gln Met Val Tyr Asp Ser Pro Leu Cys
      210          215          220
Arg Leu Asp Ala Glu Pro Arg Pro Pro Pro Thr Gln Glu Ala Ala
[0045] 225          230          235

```

<210> 89

<211> 239

<212> PRT

<213> 人工序列

<220>

<223> 具有C-端6xHis亲和标签的苯丙氨酰tRNA合成酶多肽的Alpha亚单位

<400> 89

```

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

```

130	135	140
Asn Ser Gly Val Phe Arg Pro Glu Met Leu Leu Pro Met Gly Leu Pro		
145	150	155
Glu Asn Val Ser Val Ile Ala Trp Gly Leu Ser Leu Glu Arg Pro Thr		160
	165	170
Met Ile Lys Tyr Gly Ile Asn Asn Ile Arg Glu Leu Val Gly His Lys		175
	180	185
Val Asn Leu Gln Met Val Tyr Asp Ser Pro Leu Cys Arg Leu Asp Ala		190
	195	200
Glu Pro Arg Pro Pro Pro Thr Gln Glu Ala Ala Gly Lys Pro Ile Pro		205
	210	215
Asn Pro Leu Leu Gly Leu Asp Ser Thr His His His His His His		220
225	230	235

<210> 90  
 <211> 162  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

<400> 90  
 ctgagcctgg aacgcccgac catgatcaaa tacggcatta acaatatccg tgagctgggtg 60  
 ggtcaciaaag ttaacctgca gatggtctac gacagcccgt tgtgccgect ggatgctgaa 120  
 ccgcgtccac cgccgacca agaggcagcg taatgactcg ag 162

[0046] <210> 91  
 <211> 1188  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位

<400> 91  
 cgcccagtct gttcgaagc attcccacaa cgcgcctggc cgcgcgctag cttgtgcgat 60  
 tgcccggttg cgaagtgggc cagcgcgcgc ccatgcccga cctctggttt tggttgact 120  
 cgtgttcgtc tgacgggtcc gggttgtagc gactggacc tgaaaaccta cctgggtgac 180  
 aagggtagcg cgtttagcac cagcatcagc aaacaggaaa ccgagttgct cccggagatg 240  
 attagctctg gtagctggcg tgatcgtcca ttttaagccgt acaattttct ggcgcacggt 300  
 gtcttgccgg acagcgggtca cttgcaccct ctgctgaagg tgcgctccca attccgcca 360  
 atctttctgg agatgggttt taccgagatg ccgacggaca atttcatcga atcctcgttc 420  
 tggaatttgc atgactggtt ccagccgcaa caacacccgg cacgtgacca gcacgatacc 480  
 ttctttctgc gcgactctgc ggaagcgtg caattgcccga tggactatgt gcagcgtgctc 540  
 aaacgcacgc acagccaggg tggctatggc agccagggt acaagtacaa ctggaaactg 600  
 gatgaagctc gcaagaatct gctgcgtacc catacagcga gcgcaagcgc gcgtgcgctg 660  
 taccgtctgg cgcagaagaa accgttact ccggtgaagt atttctcgat tgaccgtgtg 720  
 ttccgtaacg aaaccctgga cgccacgcat ctggcagagt tcatcaaat tgagggtgtg 780  
 gttgccgacc acggcctgac cctgggcat ctgatggcg tcttgcgcga attctttact 840  
 aaattgggta tcacgcaact gcgtttcaaa ccggcgtaca acccgtagac cgagccgagc 900  
 atggagggtct tttcttatca tcagggttg agaagtggg tggaaagtcg caacagcggc 960  
 gttttccgcc cagaaatgct gctgccgatg ggtctgcccga gaagtgttag cgtcatcga 1020  
 tggggtctga gcctggagcg tccgacgatg atcaatacgc gattaacaa tattcgtgag 1080  
 ctggttggtc acaaagttaa cctgcagatg gtttatgact ccccgctgtg tegtctggat 1140  
 gccgagccgc gtcgcctcc gaccaagaa gcagcttaat gactcgag 1188

	<210> 92	
	<211> 666	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 密码子优化的苯丙氨酰tRNA合成酶多核苷酸的alpha亚单位	
	<400> 92	
	gactatgtac aacgcgtaa ggcacccac agccaaggcg gctacggtag ccagggctac 60	
	aaatacaact ggaaactgga tgaggcccg c aagaacctgc tgcgtactca taccacgtcc 120	
	gcgagcgcac gtgctctgta tcgcctggcg caaaagaagc cttttacgcc ggtgaaatac 180	
	ttctctattg acccggtttt ccgcaatgaa accctggacg ctacgcacct ggcggagttt 240	
	caccagatcg aagggtgtgtt ggccgatcac ggtctgacct tgggcatct gatgggtgtc 300	
	ttgcgtgagt tctttactaa gctgggcatt acgcaactgc gttttaaac ggcgtataat 360	
	ccgtacaccg agccgagcat ggaggttttc agctatcacc agggcctgaa gaagtgggtt 420	
	gaggtcggta acagcgggtg gttccgtccg gaaatgctgc tgccgatggg ttgtccgga 480	
	aatgtcagcg tgatcgcatt gggcctgtcc ctggaacgtc cgaccatgat caaatatggt 540	
	atcaataaca ttcgtgagct ggttggatcat aaagttaacc tgcaaatggt ctacgatagc 600	
	ccgttgtgcc gcctggacgc ggagcccgct ccaccaccga cccaggaagc cgcataatga 660	
	ctcgag	666
	<210> 93	
	<211> 28	
	<212> DNA	
	<213> 人工序列	
[0047]	<220>	
	<223> 寡核苷酸	
	<400> 93	
	ggccaacata tggcggatgg tcaggtgg	28
	<210> 94	
	<211> 27	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 94	
	gattctcgag cgcagcctcc tgtgtgg	27
	<210> 95	
	<211> 75	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 寡核苷酸	
	<400> 95	
	ctcgagggtg agcctatccc taaccctctc ctcggtctcg attctacgca tcataccacc 60	
	caccaccacc actga	75

- <210> 96
- <211> 48
- <212> DNA
- <213> 人工序列

[0048]

- <220>
- <223> 寡核苷酸

<400> 96  
gttaggata ggcttacct cgagcccag cttctcagcc tgtcccc

48

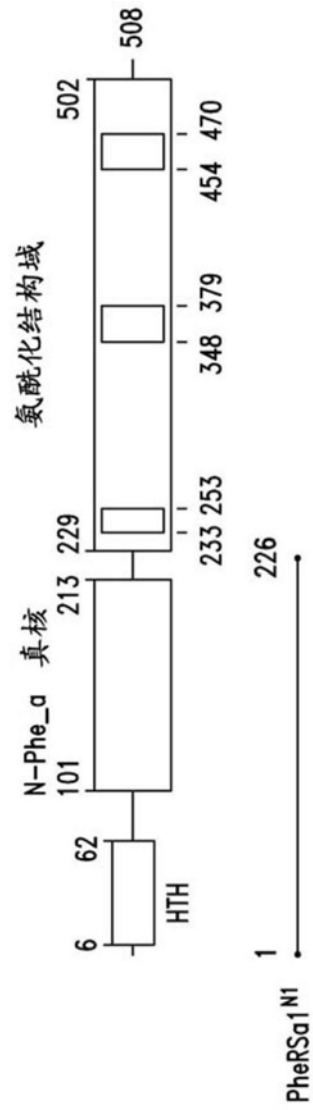


图1A

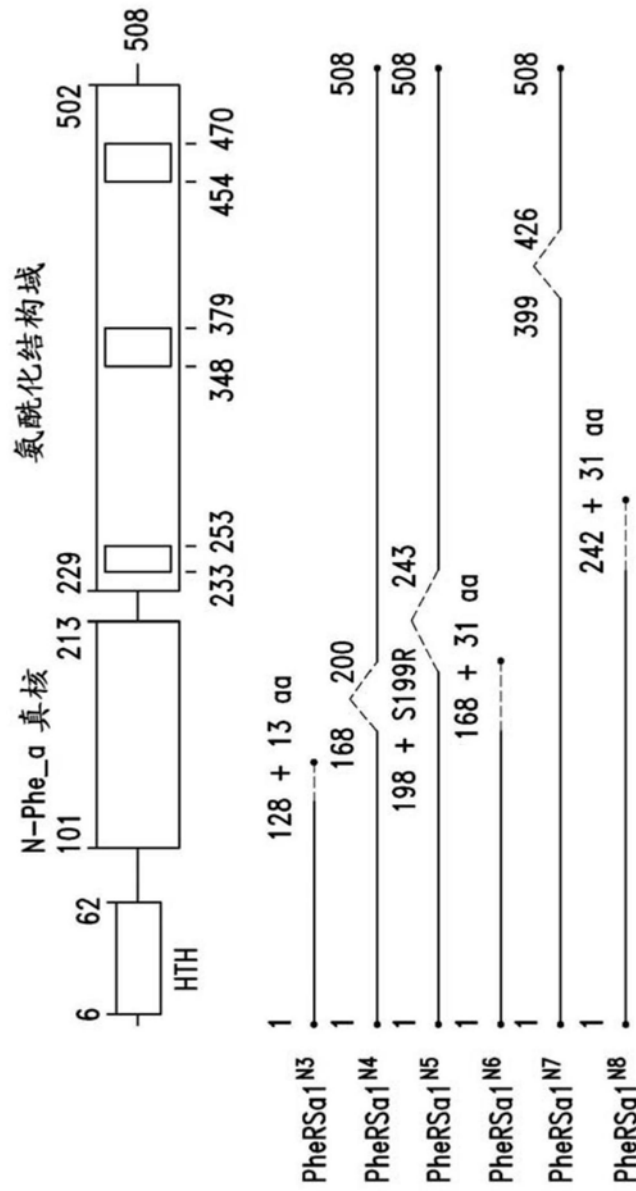


图1B

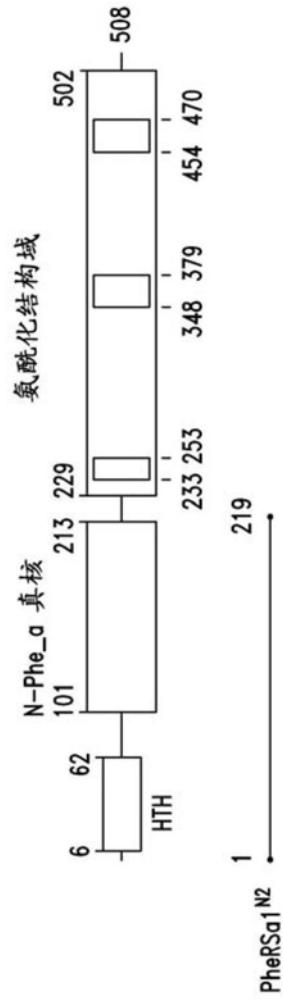


图1C



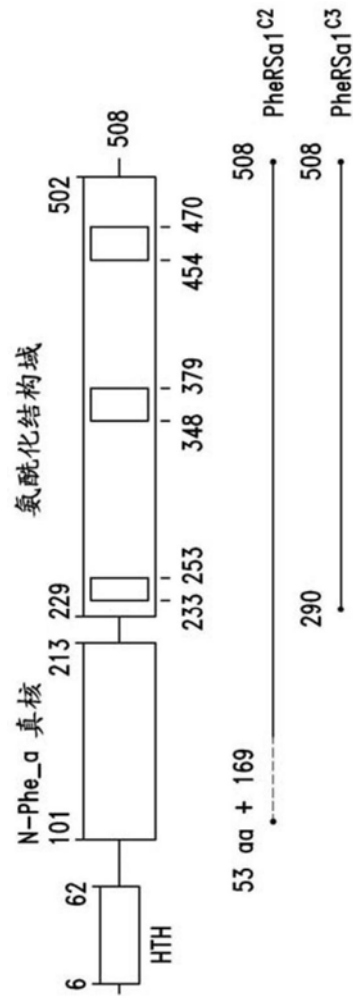


图2A

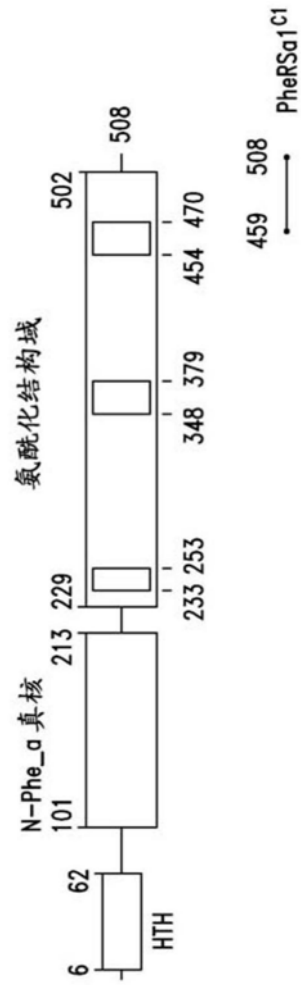


图2B