



(12) 发明专利

(10) 授权公告号 CN 1759176 B

(45) 授权公告日 2012. 10. 03

(21) 申请号 200480006490. 0

A61K 38/44 (2006. 01)

(22) 申请日 2004. 02. 10

A61K 31/7088 (2006. 01)

G01N 33/68 (2006. 01)

(30) 优先权数据

60/447, 747 2003. 02. 11 US

(56) 对比文件

WO 0160991 A2, 2001. 08. 23, 摘要、序列表第 4, 18 页、权利要求 1-28.

(85) PCT 申请进入国家阶段日

2005. 09. 09

Jens Fey. Characterization of Posttranslational Formylglycine Formation by Luminal Components of the Endoplasmic Reticulum. The journal of biological chemistry 276 50. 2001, 276 (50), 47021-47028.

(86) PCT 申请的申请数据

PCT/US2004/003632 2004. 02. 10

(87) PCT 申请的公布数据

W02004/072275 EN 2004. 08. 26

Andreas Schirmer. Computational analysis of bacterial sulfatases and their modifying enzymes. Chemistry and biology 5. 1998, (5), R181-R186.

(73) 专利权人 夏尔人类遗传性治疗公司

地址 美国马萨诸塞州

审查员 黄磊

(72) 发明人 K·冯菲古拉 B·施密特

T·蒂尔克斯 M·W·希尔特莱恩

A·巴拉比奥 M·P·考斯玛

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 程泳

(51) Int. Cl.

C12N 9/02 (2006. 01)

C12N 15/52 (2006. 01)

C12Q 1/68 (2006. 01)

A61K 38/36 (2006. 01)

权利要求书 4 页 说明书 66 页

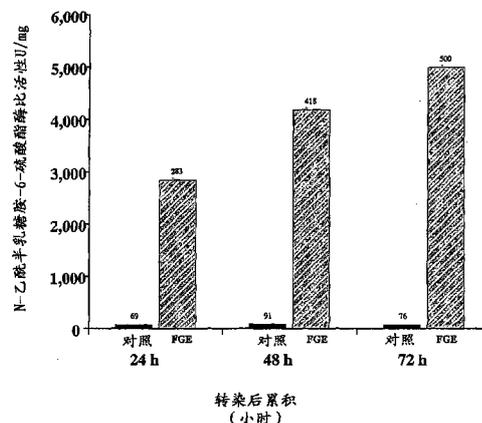
序列表 175 页 附图 7 页

(54) 发明名称

使用甲酰-甘氨酸生成酶 (FGE) 对多种硫酸酯酶缺乏症和其它病症进行诊断和治疗

(57) 摘要

本发明涉及诊断和治疗多种硫酸酯酶缺乏症 (MSD) 及其它硫酸酯酶缺乏症的方法和组合物。更特异地, 本发明涉及被分离的调节硫酸酯酶翻译后修饰的分子。这类修饰为硫酸酯酶的正确功能所必需。



1. 鉴定可用于调节 C α -甲酰甘氨酸生成活性的物质的方法,包括:
 - (a) 将具有 C α -甲酰甘氨酸生成活性的分子与候选物质接触,
 - (b) 测量所述分子的 C α -甲酰甘氨酸生成活性,和
 - (c) 将测得的所述分子的 C α -甲酰甘氨酸生成活性与对照比较以确定候选物质是否调节所述分子的 C α -甲酰甘氨酸生成活性,其中所述分子是 SEQ ID NO:1 所示的核苷酸序列的核酸分子或其表达产物。
2. 与分子选择性结合的试剂在制备用于在受试者中诊断多种硫酸酯酶缺乏症的药物中的用途,所述分子是:
 - (i) SEQ ID NO:1 所示核苷酸序列的核酸分子,或
 - (ii) (i) 的核酸分子的表达产物。
3. 选择性结合 SEQ ID NO:1 所示的核苷酸序列的核酸分子、或其表达产物的试剂在制备用于诊断以所述核酸分子或其表达产物的异常表达为特征的病症的药物中的用途,其中所述病症是多种硫酸酯酶缺乏症。
4. 包含包装的试剂盒,所述包装包含:
 - (i) 选择性结合选自下列的核酸分子的试剂:
 - (a) 核酸分子,其选自 SEQ ID NO:1 和 3 的序列,或
 - (b) 核酸分子,其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子;和
 - (ii) 第二试剂,其选择性结合选自下列的多肽:艾杜糖醛酸 2-硫酸酯酶,硫酸胺酶, N-乙酰半乳糖胺 6-硫酸酯酶, N-乙酰葡萄糖胺 6-硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 HSulf-6, 或其肽,以及用于与所述第二试剂和所述多肽或其肽结合的测量值进行比较的对照。
5. 调节 C α -甲酰甘氨酸生成活性的试剂在制备用于治疗受试者中多种硫酸酯酶缺乏症的药物中的用途,其中调节 C α -甲酰甘氨酸生成活性的试剂是选自下列的核酸分子:
 - (a) 核酸分子,其选自 SEQ ID NO:1 和 3 的序列,或
 - (b) 核酸分子,其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子。
6. 调节 C α -甲酰甘氨酸生成活性的试剂在制备用于治疗受试者中多种硫酸酯酶缺乏症的药物中的用途,其中调节 C α -甲酰甘氨酸生成活性的试剂是:

由选自下列的核酸分子编码的多肽:

 - (a) 核酸分子,其是选自 SEQ ID NO:1 和 3 的序列,或
 - (b) 核酸分子,其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子;或是SEQ ID NO:2 的序列的多肽。
7. 调节 C α -甲酰甘氨酸生成活性的试剂在制备用于治疗受试者中多种硫酸酯酶缺乏症的药物中的用途,其中调节 C α -甲酰甘氨酸生成活性的试剂是由表达选自下列的 FGE 核酸分子的细胞生产的:
 - (a) 核酸分子,其选自 SEQ ID NO:1 和 3 的序列,或
 - (b) 核酸分子,其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子。
8. 权利要求 5、6 或 7 的用途,其中所述药物进一步包括第二试剂,其中所述第二试剂

是编码选自下列的硫酸酯酶的核酸分子：艾杜糖醛酸 2- 硫酸酯酶，硫酸胺酶，N- 乙酰半乳糖胺 6- 硫酸酯酶，N- 乙酰葡萄糖胺 6- 硫酸酯酶，芳基硫酸酯酶 A，芳基硫酸酯酶 B，芳基硫酸酯酶 C，芳基硫酸酯酶 D，芳基硫酸酯酶 E，芳基硫酸酯酶 F，芳基硫酸酯酶 G，HSulf-1，HSulf-2，HSulf-3，HSulf-4，HSulf-5，或 HSulf-6，和所述核酸分子的表达产物。

9. 权利要求 7 的用途，其中表达 FGE 核酸分子的细胞表达外源 FGE 核酸分子。

10. 权利要求 7 的用途，其中表达 FGE 核酸分子的细胞表达内源 FGE 核酸分子。

11. 分离的 FGE 核酸分子或其表达产物在制备药物中的应用，其中所述药物用于在受试者中增加 C α -甲酰甘氨酸生成活性，所述核酸分子选自：

(a) 核酸分子，其选自 SEQ ID NO:1 和 3 的序列，或

(b) 核酸分子，其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子。

12. 有效量的调节 C α -甲酰甘氨酸生成活性的试剂在制备用于在受试者中治疗多种硫酸酯酶缺乏症的药物中的用途，其中所述有效量足以使受试者中的 C α -甲酰甘氨酸生成活性增加，其中调节 C α -甲酰甘氨酸生成活性的试剂是选自下列的有义核酸分子：

(a) 核酸分子，其选自 SEQ ID NO:1 和 3 的序列，或

(b) 核酸分子，其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子。

13. 有效量的调节 C α -甲酰甘氨酸生成活性的试剂在制备用于在受试者中治疗多种硫酸酯酶缺乏症的药物中的用途，其中所述有效量足以使受试者中的 C α -甲酰甘氨酸生成活性增加，其中调节 C α -甲酰甘氨酸生成活性的试剂是：

由选自下列的核酸分子编码的分离的多肽：

(a) 核酸分子，其选自 SEQ ID NO:1 和 3 的序列，或

(b) 核酸分子，其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子；或者

SEQ ID NO. 2 的序列的多肽。

14. 选自下列的核酸分子或其表达产物在制备药物中的用途，所述药物用于在细胞中增加 C α -甲酰甘氨酸生成活性：

(a) 核酸分子，其选自 SEQ ID NO:1 和 3 的序列，或

(b) 核酸分子，其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子。

15. 药物组合物，包含：

治疗多种硫酸酯酶缺乏症的药理学有效量的试剂，所述试剂包含选自下列的分离的核酸分子或其表达产物：

(a) 核酸分子，其选自 SEQ ID NO:1 和 3 的序列，或

(b) 核酸分子，其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子，和药学上可接受的载体。

16. 鉴定可用于治疗多种硫酸酯酶缺乏症的候选物质的方法，包括：

测定一套核酸分子在细胞或组织中的表达，条件为在候选物质缺乏时允许该套核酸分子的最初量表达，其中该套核酸分子包含至少一种选自下面的核酸分子：

(a) 核酸分子，其选自 SEQ ID NO:1 和 3 的序列，或

(b) 由于遗传密码的简并性，密码子序列与 (a) 或 (b) 的核酸分子不同的核酸分子，将细胞或组织与候选物质体外接触，并

检测该套核酸分子表达的测试量,其中在候选物质存在的条件下表达的测试量相对于表达最初量的增加表明候选物质可用于治疗多种硫酸酯酶缺乏症。

17. 固相核酸分子阵列,其由固定于固体基质的一套核酸分子组成,每一核酸分子编码选自下面的多肽:SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 和 78, 艾杜糖醛酸 2- 硫酸酯酶, 硫酸胺酶, N- 乙酰半乳糖胺 6- 硫酸酯酶, N- 乙酰葡萄糖胺 6- 硫酸酯酶, 芳基硫酸酯酶 A, 芳基硫酸酯酶 B, 芳基硫酸酯酶 C, 芳基硫酸酯酶 D, 芳基硫酸酯酶 E, 芳基硫酸酯酶 F, 芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 和 HSulf-6, 其中该套核酸分子编码下列多肽的至少一种:SEQ ID NO :2。

18. 权利要求 17 的固相核酸分子阵列,进一步包含至少一种对照核酸分子。

19. 权利要求 17 的固相核酸分子阵列,其中该套核酸分子包含至少两种核酸分子,每一核酸分子编码选自下列的多肽:SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 和 78, 艾杜糖醛酸 2- 硫酸酯酶, 硫酸胺酶, N- 乙酰半乳糖胺 6- 硫酸酯酶, N- 乙酰葡萄糖胺 6- 硫酸酯酶, 芳基硫酸酯酶 A, 芳基硫酸酯酶 B, 芳基硫酸酯酶 C, 芳基硫酸酯酶 D, 芳基硫酸酯酶 E, 芳基硫酸酯酶 F, 芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 和 HSulf-6。

20. 选自下列的分离的核酸分子或其表达产物在制备药物中的用途,所述药物用于在细胞中增加硫酸酯酶活性:

(a) 核酸分子,其选自 SEQ ID NO :1 和 3 的序列,或

(b) 由于遗传密码的简并性,密码子序列与 (a) 或 (b) 的核酸分子不同的核酸分子。

21. 权利要求 20 的用途,其中细胞表达内源硫酸酯酶。

22. 权利要求 20 的用途,其中细胞表达外源硫酸酯酶。

23. 权利要求 20 的用途,其中内源硫酸酯酶是被激活的。

24. 根据权利要求 21-23 中任一项的用途,其中硫酸酯酶选自艾杜糖醛酸 2- 硫酸酯酶, 硫酸胺酶, N- 乙酰半乳糖胺 6- 硫酸酯酶, N- 乙酰葡萄糖胺 6- 硫酸酯酶, 芳基硫酸酯酶 A, 芳基硫酸酯酶 B, 芳基硫酸酯酶 C, 芳基硫酸酯酶 D, 芳基硫酸酯酶 E, 芳基硫酸酯酶 F, 芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 和 HSulf-6。

25. 权利要求 20 的用途,其中细胞是哺乳动物细胞。

26. 药物组合物,包含:

治疗硫酸酯酶缺乏症的由细胞产生的硫酸酯酶,和

药学上可接受的载体,

其中所述细胞已与试剂接触过,所述试剂包含选自下列的分离的核酸分子或其表达产物:

(a) 核酸分子,其选自 SEQ ID NO :1 和 3 的序列,或

(b) 核酸分子,其由于遗传密码的简并性而在密码子序列中不同于 (a) 的核酸分子。

27. 人 FGE 基因的分离的变体等位基因,其编码变体 FGE 多肽,包含:

在 SEQ ID NO :2 中含有至少一种变异的氨基酸序列,其中所述至少一种变异包括:Met1Arg ;Met1Val ;Ser155Pro ;Cys218Tyr ;Ala279Val ;Arg327Stop ;Cys336Arg ;Arg345Cys ;Arg349Trp ;Arg349Gln ;Ser359Stop ;或其组合。

28. 分离的变体人 FGE 多肽,包含:

在 SEQ ID NO:2 中包含至少一种变异的氨基酸序列,其中所述至少一种变异包含:Met1Arg;Met1Val;Ser155Pro;Cys218Tyr;Ala279Val;Arg327Stop;Cys336Arg;Arg345Cys;Arg349Trp;Arg349Gln;Ser359Stop;或其组合。

29. 抗体,其以权利要求 28 的变体人 FGE 多肽作为免疫原。
30. 权利要求 29 的抗体,其是多克隆抗体。
31. 权利要求 29 的抗体,其是单克隆抗体。
32. 权利要求 29 的抗体,其是嵌合抗体。
33. 权利要求 29 的抗体,其所述抗体被可检测标记所标记。
34. 权利要求 33 的抗体,其中所述可探测性的标记包含放射性元素,荧光化学物或酶。

使用甲酰 - 甘氨酸生成酶 (FGE) 对多种硫酸酯酶缺乏症和 其它病症进行诊断和治疗

发明领域

[0001] 此发明涉及诊断和治疗多种硫酸酯酶缺乏症 (MSD) 以及其它硫酸酯酶缺乏症的方法和组合物。更特异地, 本发明涉及被分离的调节硫酸酯酶翻译后修饰的分子。这类修饰为硫酸酯酶的正确功能所必需。

[0002] 发明背景

[0003] 硫酸酯酶是高度保守的基因家族的成员, 共享广泛的序列同源性 (Franco, B. 等, *Cell*, 1995, 81 :15-25 ;Parenti, G. 等, *Curr. Opin. Gen. Dev.*, 1997, 7 :386-391), 高度的结构类似性 (Bond, C. S. 等, *Structure*, 1997, 5 :277-289 ;Lukatela, G. 等, *Biochemistry*, 1998, 37 :3654-64), 和独特的为硫酸酯断裂所必需的翻译后修饰 (Schmidt, B. 等, *Cell*, 1995, 82 :271-278 ;Selmer, T. 等, *Eur. J. Biochem.*, 1996, 238 :341-345)。翻译后修饰包括对保守的半胱氨酸 (在真核细胞中) 或丝氨酸 (在某些原核细胞中) 残基在 C_β 处的氧化, 得到 L-C_α-甲酰甘氨酸 (也叫做 FGly, 2-氨基-3-氧代丙酸), 其中醛基代替了侧链的硫代甲基基团。醛基是硫酸酯酶催化位点的必需部分, 很可能作为醛水合物起作用。成对的羟基之一在硫酸酯断裂时接受硫酸基团, 导致共价硫酸化的酶中间物的形成。其它羟基是随后的硫酸基团消去和醛基再生所需要的。此修饰在初生硫酸酯酶多肽输入期间或间隔很短时间之后发生于内质网, 并被围绕着将被修饰的半胱氨酸 (或丝氨酸) 残基的短线性序列所调控。此高度保守序列是六肽 L/V-C(S)-X-P-S-R (SEQ ID NO :32), 出现在所有真核细胞硫酸酯酶的 N 末端区域, 并最频繁地在残基 X 上携带羟基或硫羟基 (Dierks, T. 等, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, 94 :11963-11968)。

[0004] 迄今为止, 已在人体中鉴别出 13 个硫酸酯酶基因。它们编码具有不同底物专一性和亚细胞定位如溶酶体、高尔基体和内质网定位的酶。这些基因中的四种 ARSC, ARSD, ARSE, 和 ARSF (分别编码芳基硫酸酯酶 C, D, E 和 F), 位于相同的染色体区域 (Xp22.3) 内。它们共享显著的序列类似性和几乎相同的基因组组织方式, 说明它们来源于在进化中近来才发生的复制事件 (Franco B 等, *Cell*, 1995, 81 :15-25 ;Meroni G 等, *Hum Mol Genet*, 1996, 5 :423-31)。

[0005] 由单独硫酸酯酶活性的缺乏所引起的至少八种人单基因疾病的鉴定, 强调了硫酸酯酶在人体代谢中的重要性。这些病症中的大部分是溶酶体存贮病, 其中, 表型结果源于被储存物质的类型和组织分布。在它们中, 有五种不同类型的因硫酸酯酶对粘多糖分解代谢作用的缺乏而引起的粘多糖病 (MPS 类型 II, IIIA, IIID, IVA 和 VI) (Neufeld 和 Muenzer, 2001, *The mucopolysaccharidoses*, In *The Metabolic and Molecular Bases of Inherited Disease*, C. R. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, B. Childs, K. W. Kinzler 和 B. Vogelstein 编, New York :Mc Graw-Hill, pp. 3421-3452), 和以脑硫脂在中枢和外周神经系统中储存并引起严重和渐进性神经退化为特征的异常染色性脑白质营养不良 (MLD)。两种另外的人类疾病由非溶酶体硫酸酯酶缺乏引起。这些包括 X-连锁的鱼鳞病, 因类固醇硫酸酯酶 (STS/ARSC) 缺乏而引起的皮肤病 ;和点状软骨发育不全, 由芳基硫酸酯酶 E (ARSE)

缺乏引起的影响骨和软骨的疾病。硫酸酯酶也牵涉到药物诱导的人畸形综合症,例如 Warfarin 胚胎病,由怀孕期间在宫内暴露于杀鼠灵而抑制 ARSE 活性所引起。

[0006] 在引起人兴趣的人类单基因疾病中,多种硫酸酯酶缺乏症 (MSD) 是所有硫酸酯酶活性同时有缺陷。因此,此严重的多系统性疾病的表型结合了在单独硫酸酯酶缺乏中所观察到的特征。来自多种硫酸酯酶缺乏症患者的细胞即使以编码人硫酸酯酶的 cDNAs 转染之后也缺乏硫酸酯酶活性,暗示所有硫酸酯酶活性所需要的普遍机制的存在 (Rommerskirch 和 von Figura, Proc. Natl. Acad. Sci., USA, 1992, 89 :2561-2565)。硫酸酯酶的翻译后修饰被发现在多种硫酸酯酶缺乏症患者中有缺陷,暗示此疾病由基因的突变所引起,该基因牵涉到半胱氨酸至甲酰甘氨酸的转化机制 (Schmidt, B. 等, Cell, 1995, 82 :271-278)。尽管存在强烈的生物学和医学兴趣,以鉴别此基因为目标的努力却被多种硫酸酯酶缺乏症患者的稀有性和随之带来的缺乏合适的家族病例而无法完成遗传图谱绘制所妨碍。

[0007] 发明概述

[0008] 本发明提供诊断和治疗多种硫酸酯酶缺乏症 (MIM 272200) 及治疗其它硫酸酯酶缺乏症的方法和组合物。更特异地,已鉴别出编码甲酰甘氨酸生成酶 (FGE) 的基因,此酶负责发生于硫酸酯酶独特的翻译后修饰 (形成 L-C_α-甲酰甘氨酸;也叫做 FGly 和 / 或 2-氨基-3-氧代丙酸),该翻译后修饰为硫酸酯酶功能所必需。已发现,出乎意料的是 FGE 基因中的突变导致受试者中的多种硫酸酯酶缺乏症 (MSD) 的发展。也发现,出乎意料的是 FGE 增强了硫酸酯酶的活性,包括但不限于,艾杜糖醛酸 2-硫酸酯酶,硫酸胺酶, N-乙酰半乳糖胺 6-硫酸酯酶, N-乙酰葡萄糖胺 6-硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 HSulf-6。考虑到这些发现,本发明的分子可用于诊断和治疗多种硫酸酯酶缺乏症以及其它硫酸酯酶缺乏症。

[0009] 在多种硫酸酯酶缺乏症的诊断中应用本发明的分子的方法被提供。

[0010] 另外,为调节硫酸酯酶上 FGly 形成的目的而在体内或体外应用这些分子的方法,治疗与这种修饰相关的病症的方法,和对制备治疗多种硫酸酯酶缺乏症及其它硫酸酯酶缺乏症的药学制剂有用的组合物,也被提供。

[0011] 本发明因此在几个方面包括调节硫酸酯酶上 FGly 形成的多肽,分离的编码那些多肽的核酸,它们的功能修饰物和变体,它们的有用的片段,以及与其相关的治疗、诊断和研究的方法、组合物与工具。

[0012] 根据本发明的一个方面,选自下列的分离的核酸分子被提供:(a) 与 SEQ ID NO: 1 所示的核苷酸序列所组成的分子在严格条件下杂交的核酸分子,该核酸分子编码具有 C_α-甲酰甘氨酸生成活性的甲酰甘氨酸生成酶 (FGE);(b) 与 (a) 的核酸分子在密码子序列上因遗传密码简并性而有所区别的核酸分子;和 (c) (a) 或 (b) 的互补链。在一定的实施方案中,分离的核酸分子包括 SEQ ID NO: 1 所示的核苷酸序列。在某些实施方案中,分离的核酸分子由 SEQ ID NO: 3 所示的核苷酸序列或其片段组成。

[0013] 本发明的另一方面提供分离的选自下列的核酸分子:(a) SEQ ID NO: 1 所示的核苷酸序列的独特 (unique) 片段,和 (b) (a) 的互补链,条件是 (a) 中的独特片段包括了不与某些序列相同的连续核苷酸序列,所述某些序列选自 (1) 与 SEQ ID NO. 4 和 / 或 SEQ ID NO. 4 的核苷酸 20-1141 相同的序列,和 (2) (1) 中核酸分子的互补链。在任何前述的实施方案中,

互补链指全长互补链。

[0014] 在一个实施方案中,连续核苷酸序列选自(1)至少两个不与序列组完全相同的连续核苷酸,(2)至少三个不与序列组完全相同的连续核苷酸,(3)至少四个不与序列组完全相同的连续核苷酸,(4)至少五个不与序列组完全相同的连续核苷酸,(5)至少六个不与序列组完全相同的连续核苷酸,和(6)至少七个不与序列组完全相同的连续核苷酸。

[0015] 在另一实施方案中,片段大小选自至少:8个核苷酸,10个核苷酸,12个核苷酸,14个核苷酸,16个核苷酸,18个核苷酸,20个核苷酸,22个核苷酸,24个核苷酸,26个核苷酸,28个核苷酸,30个核苷酸,40个核苷酸,50个核苷酸,75个核苷酸,100个核苷酸,200个核苷酸,1000个核苷酸和其间的每一整数长度。

[0016] 根据又一方面,本发明提供包含上述核酸分子的表达载体和以该表达载体转化或转染的宿主细胞。

[0017] 根据又一方面,本发明提供表达内源 FGE 基因的活化形式的细胞。在一个实施方案中,内源 FGE 基因的活化通过同源重组发生。

[0018] 根据本发明又一方面,分离的多肽被提供。分离的多肽被前述的本发明核酸分子所编码。在某些实施方案中,分离的多肽被 SEQ ID NO:1 的核酸编码,产生出具有 SEQ ID NO:2 序列和 C_α-甲酰甘氨酸生成活性的多肽。在其它的实施方案中,分离的多肽也许是前述的片段或变体,其长度足以代表人基因组中独特序列并且是具有 C_α-甲酰甘氨酸生成活性的多肽,条件是片段包括不与任何被具有 SEQ ID NO.4 的核酸序列所编码的序列相同的连续氨基酸的序列。在另一实施方案中,上述多肽分子的免疫源性片段被提供。免疫源性片段也许具有也许不具有 C_α-甲酰甘氨酸生成活性。

[0019] 根据本发明又一方面,分离的结合多肽被提供,其选择性地与前述的本发明核酸分子所编码的多肽结合。优选分离的结合多肽选择性地结合包含 SEQ ID NO:2 序列的多肽、其片段或属于分离的具有 C_α-甲酰甘氨酸生成活性的多肽家族而在本文其它部分所述的多肽。在优选的实施方案中,分离的结合多肽包括抗体和抗体片段(如 Fab, F(ab)₂, Fd 和包括了与 FGE 多肽选择性结合的 CDR3 区域的抗体片段)。在一定的实施方案中,抗体是人抗体。在某些实施方案中,抗体是单克隆抗体。在某一实施方案中,抗体是多克隆抗血清。在进一步的实施方案中,抗体是人源化的。在更进一步的实施方案中,抗体是嵌合的。

[0020] 根据本发明又一方面,具有 C_α-甲酰甘氨酸生成活性的分离的多肽的家族被提供。每一所述多肽从氨基端到羧基端包含:(a)氨基端亚结构域 1;亚结构域 2;包含 35 到 45 个氨基酸的羧基端亚结构域 3,而其中亚结构域 3 与选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的多肽的亚结构域 3 具有至少约 75% 的同源性和大致相同的长度。在重要的实施方案中,亚结构域 2 含有 120 到 140 个氨基酸。在进一步的重要实施方案中,亚结构域 2 中至少 5% 的氨基酸是色氨酸。在某些实施方案中,亚结构域 2 与选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的多肽的结构域 2 之间具有至少 50% 的同源性。在一定的实施方案中,每一多肽的亚结构域 3 与选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的多肽的亚结构域 3 具有介于约 80% 和约 100% 之间的同源性。

[0021] 根据本发明进一步的方面,用于测定某一受试者中的 FGE 表达水平的方法被提供。方法包括测量来自某一受试者的测试样本中的 FGE 表达以确定 FGE 在受试者中的表达

水平。在一定的实施方案中,测得的测试样本中的 FGE 表达与对照样本(含有已知 FGE 表达水平)中的 FGE 表达进行比较。表达被定义成 FGE mRNA 表达, FGE 多肽表达,或如本文其它部分所定义的 FGE 的 C_α-甲酰甘氨酸生成活性。多种方法能用于测量表达。本发明优选的实施方案包括用于测量 mRNA 表达的 PCR 和 RNA 印迹,作为测量 FGE 多肽表达的试剂的 FGE 单克隆抗体或 FGE 多克隆抗血清,以及测量 FGE C_α-甲酰甘氨酸生成活性的方法。

[0022] 在一定的实施方案中,测试样本例如活检样本和生物液体如血液被用作测试样本。FGE 在某一受试者的测试样本中的表达与对照样本中的 FGE 表达比较。

[0023] 根据本发明另一方面,用于鉴别在调节分子的 C_α-甲酰甘氨酸生成活性中有用的物质的方法被提供。方法包括 (a) 将具有 C_α-甲酰甘氨酸生成活性的分子与候选物质接触, (b) 测量该分子的 C_α-甲酰甘氨酸生成活性,和 (c) 将测得的该分子 C_α-甲酰甘氨酸生成活性与对照进行比较以确定候选物质是否调节分子的 C_α-甲酰甘氨酸生成活性,其中该分子是具有选自 SEQ ID NO :1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的核苷酸序列的核酸分子,或其表达产物(例如,具有选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的序列的肽)。在一定的实施方案中,对照是在缺乏候选物质的条件下测得的该分子的 C_α-甲酰甘氨酸生成活性。

[0024] 根据本发明又一方面,在受试者中诊断多种硫酸酯酶缺乏症的方法被提供。方法包括将生物样本与试剂接触,所述生物样本来自被怀疑具有多种硫酸酯酶缺乏症的受试者,所述试剂特异结合选自下面的分子:(i) 具有 SEQ ID NO :1, 3, 或 4 核苷酸序列的 FGE 核酸分子, (ii) 核酸分子 (i) 的表达产物,或 (iii) (ii) 的表达产物的片段;以及测量结合的试剂数量并由此确定所述核酸分子或其表达产物的表达是否异常,异常表达表明受试者患有多种硫酸酯酶缺乏症。

[0025] 根据本发明又一方面,用于对特征为核酸分子或其表达产物的异常表达的病症进行诊断的方法被提供。方法包括将来自受试者的生物样本与试剂接触,其中所述试剂特异结合所述核酸分子、其表达产物或其表达产物的片段;并测量结合的试剂数量并由此确定所述核酸分子或其表达产物的表达是否异常,异常表达表明具有所述病症,其中所述核酸分子具有 SEQ ID NO :1 核苷酸序列而所述病症是多种硫酸酯酶缺乏症。

[0026] 根据本发明又一方面,用于在受试者中测量特征为核酸分子或其表达产物的异常表达的多种硫酸酯酶缺乏症的方法被提供。方法包括对来自患者的样本监测选自下面的参数:(i) 具有 SEQ ID NO :1, 3, 4 的核苷酸序列的核酸分子或具有来源于 FEG 基因组位点的序列的核酸分子, (ii) 所述核酸分子编码的多肽, (iii) 来源于所述多肽的肽,和 (iv) 选择性结合所述多肽或肽的抗体,将它们作为对受试者中的多种硫酸酯酶缺乏症的测定。在某些实施方案中,样本是如前述实施方案任一项所描述的生物液体或组织。在一定的实施方案中,监测步骤包括将样本与选自下面的可检测的试剂接触:(a) 在严格条件下与 (i) 中的核酸分子选择性杂交的分离的核酸分子, (b) 选择性结合 (ii) 中的多肽或 (iii) 中的肽的抗体,和 (c) 与 (iv) 中的抗体结合的多肽或肽。抗体、多肽、肽或核酸能用放射性标记或酶进行标记。在进一步的实施方案中,方法进一步包含检验样品中的肽。

[0027] 根据本发明又一方面,试剂盒被提供。试剂盒包含包装,包装包含选择性结合前述 FGE 的分离的核酸或其表达产物的任一种的试剂,和用于与测量值比较的对照,此测量值为所述试剂与前述 FGE 的分离的核酸或其表达产物的任一种结合的测量值。在某些实施方案

中,对照是用于与测量值比较的预定值。在一定的实施方案中,对照包含前述 FGE 的分离的核酸的任一种的表达产物的表位。在一种实施方案中,试剂盒进一步包含选择性结合选自下面的多肽的第二试剂:艾杜糖醛酸 2- 硫酸酯酶,硫酸胺酶 (sulfamidase), N- 乙酰半乳糖胺 6- 硫酸酯酶,N- 乙酰葡萄糖胺 6- 硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 HSulf-6, 或其肽段,和用于与前述第二试剂结合到所述多肽或其肽段的测量值进行比较的对照。

[0028] 根据本发明进一步的方面,治疗多种硫酸酯酶缺乏症的方法被提供。方法包括对需要这种治疗的受试者施用调节 C_α- 甲酰甘氨酸生成活性的试剂,用量为治疗受试者的多种硫酸酯酶缺乏症的有效量。在某些实施方案中,方法进一步包含某一试剂的共施用,所述试剂选自编码艾杜糖醛酸-2- 硫酸酯酶,硫酸胺酶,N- 乙酰半乳糖胺-6- 硫酸酯酶,N- 乙酰葡萄糖胺-6- 硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 或 HSulf-6 的核酸分子,该核酸分子的表达产物,和该核酸分子的表达产物的片段。在一定的实施方案中,调节 C_α- 甲酰甘氨酸生成活性的试剂是本发明的分离的核酸分子(例如如权利要求 1-8 所要求的核酸分子或具有选自 SEQ ID NO :1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的核酸)。在重要的实施方案中,调节 C_α- 甲酰甘氨酸生成活性的试剂是本发明的肽(例如,如权利要求 11-15, 19, 20 所要求的肽,或具有选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的序列的肽)。调节 C_α- 甲酰甘氨酸生成活性的试剂可通过表达内源和 / 或外源 FGE 核酸分子的细胞产生。在重要的实施方案中,内源 FGE 核酸分子可被活化。

[0029] 根据本发明的一个方面,用于在受试者中增加 C_α- 甲酰甘氨酸生成活性的方法被提供。方法包括施用本发明的分离的 FGE 核酸分子(例如,权利要求 1-8 的核酸分子,或具有选自 SEQ ID NO :1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的核酸),和 / 或其表达产物到受试者中,用量为在受试者中增加 C_α- 甲酰甘氨酸生成活性的有效量。

[0030] 根据本发明的一个方面,用于治疗患多种硫酸酯酶缺乏症的受试者的方法被提供。方法包括对需要这种治疗的受试者施用调节 C_α- 甲酰甘氨酸生成活性的试剂,用量为在受试者中增加 C_α- 甲酰甘氨酸生成活性的有效量。在某些实施方案中,调节 C_α- 甲酰甘氨酸生成活性的试剂是本发明的有义核酸(例如,权利要求 1-8 的核酸分子,或具有选自 SEQ ID NO :1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的核酸)。在一定的实施方案中,调节 C_α- 甲酰甘氨酸生成活性的试剂是本发明的分离的多肽(例如,权利要求 11-15, 19, 20 的多肽或具有选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的序列的肽)。

[0031] 根据本发明又一方面,用于在细胞中增加 C_α- 甲酰甘氨酸生成活性的方法被提供。方法包括将细胞与本发明的分离的核酸分子接触(例如,权利要求 1-8 的核酸分子,或具有选自 SEQ ID NO :1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的核酸),或其表达产物,用量为在细胞中增加 C_α- 甲酰甘氨酸生成活性的有效量。在重要的实施方案中,方法包括活化内源 FGE 基因以在细胞中增加 C_α- 甲酰甘氨酸

生成活性。

[0032] 根据本发明进一步的方面,药物组合物被提供。组合物包含治疗多种硫酸酯酶缺乏症的药学有效量的某种试剂,该试剂含有本发明的分离的核酸分子(例如,权利要求1-8中任一项的分离的核酸分子,具有选自SEQ ID NO:1,3,4,45,47,49,51,53,55,57,59,61,63,65,67,69,71,73,75,77,和80-87序列的FGE核酸分子),或其表达产物,以及药学上可接受的载体。

[0033] 根据本发明的一个方面,用于鉴别在治疗多种硫酸酯酶缺乏症中有用的候选物质的方法被提供。方法包括测定一套核酸分子在细胞或组织中的表达,条件是缺乏候选物质时,允许该套核酸分子的最初量的表达,其中该套核酸分子包含至少一种选自下面的核酸分子:(a)严格条件下与由SEQ ID NO:1所示核苷酸序列组成的分子杂交并编码具有C_α-甲酰甘氨酸生成活性(FGE)的多肽的核酸分子,(b)遗传密码简并性引起的在密码子序列中与(a)或(b)中核酸分子不同的核酸分子,(c)具有选自SEQ ID NO:1,3,4,45,47,49,51,53,55,57,59,61,63,65,67,69,71,73,75,77,和80-87的序列的核酸分子,和(d)(a)或(b)或(c)的互补链,将细胞或组织与候选物质接触并探测该套核酸分子表达的测试量,其中在候选物质存在的条件下表达的测试量相对于表达最初量的增加表明候选物质在治疗多种硫酸酯酶缺乏症中是有用的。

[0034] 根据本发明进一步的方面,制备在治疗多种硫酸酯酶缺乏症和/或其它硫酸酯酶缺乏症中有用的药剂的方法被提供。

[0035] 根据本发明又一方面,固相核酸分子阵列被提供。阵列基本上由固定于固相基质的一套核酸分子、其表达产物或其片段(或是核酸的或是多肽分子的)组成,每一核酸分子编码选自下面的多肽:SEQ ID NO.2,5,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,和78,艾杜糖醛酸-2-硫酸酯酶,硫酸胺酶,N-乙酰半乳糖胺-6-硫酸酯酶,N-乙酰葡萄糖胺-6-硫酸酯酶,芳基硫酸酯酶A,芳基硫酸酯酶B,芳基硫酸酯酶C,芳基硫酸酯酶D,芳基硫酸酯酶E,芳基硫酸酯酶F,芳基硫酸酯酶G,HSulf-1,HSulf-2,HSulf-3,HSulf-4,HSulf-5,和HSulf-6。在某些实施方案中,固相阵列进一步包含至少一种对照核酸分子。在一定的实施方案中,该套核酸分子包含至少一种、至少两种、至少三种、至少四种或甚至至少五种核酸分子,每一种选自SEQ ID NO.2,5,46,48,50,52,54,56,58,60,62,64,66,68,70,72,74,76,和78,艾杜糖醛酸-2-硫酸酯酶,硫酸胺酶,N-乙酰半乳糖胺-6-硫酸酯酶,N-乙酰葡萄糖胺-6-硫酸酯酶,芳基硫酸酯酶A,芳基硫酸酯酶B,芳基硫酸酯酶C,芳基硫酸酯酶D,芳基硫酸酯酶E,芳基硫酸酯酶F,芳基硫酸酯酶G,HSulf-1,HSulf-2,HSulf-3,HSulf-4,HSulf-5,和HSulf-6。

[0036] 根据本发明进一步的方面,用于在受试者中治疗硫酸酯酶缺乏症的方法被提供。方法包括对需要这种治疗的受试者施用已根据本发明生产出的硫酸酯酶,用量为在受试者中治疗硫酸酯酶缺乏症的有效量,而此硫酸酯酶缺乏症不是多种硫酸酯酶缺乏症。在重要的实施方案中,通过已与调节C_α-甲酰甘氨酸生成活性的试剂接触的细胞产生硫酸酯酶。在一定的实施方案中,硫酸酯酶缺乏症包括但不限于粘多糖病II(MPS II;Hunter综合症),粘多糖病IIIA(MPS IIIA;Sanfilippo综合症A),粘多糖病VIII(MPS VIII),粘多糖病IVA(MPS IVA;Morquio综合症A),粘多糖病VI(MPS VI;Maroteaux-Lamy综合症),异常染性脑白质营养不良(MLD),X-连锁的隐性点状软骨发育不全1,或X-连锁的鱼鳞病(类固醇硫

酸酯酶缺乏症)。在一定的实施方案中,调节 C_α-甲酰甘氨酸生成活性的试剂可以是本发明的核酸分子或肽。在一种实施方案中,硫酸酯酶和调节 C_α-甲酰甘氨酸生成活性的试剂在同一细胞中共表达。硫酸酯酶和 / 或调节 C_α-甲酰甘氨酸生成活性的试剂可以是内源性或外源性的来源。如果是内源性来源,它可被活化(例如,通过在本领域中已知的合适的位置插入强启动子和 / 或其它元件)。如果是外源性的,其表达可被表达载体上的元件所驱动,或它可靶向于细胞基因组中合适的位置以允许其被增强的表达(例如,强启动子下游)。

[0037] 根据本发明又一方面,药物组合物被提供。组合物包含治疗硫酸酯酶缺乏症的药理学有效量的某种试剂,该试剂含有本发明的分离的核酸分子或其表达产物,及药学上可接受的载体。

[0038] 根据本发明更进一步的方面,用于在细胞中增加硫酸酯酶活性的方法被提供。方法包括将表达硫酸酯酶的细胞与本发明的分离的核酸分子(例如,权利要求 1-8 中任一项的分离的核酸分子,具有选自 SEQ ID NO :1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的 FGE 核酸分子)或其表达产物(例如,权利要求 11-15, 19, 20 的多肽,或具有选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的序列的肽)接触,用量为在细胞中增加硫酸酯酶活性的有效量。细胞可表达内源性和 / 或外源性的硫酸酯酶。在重要的实施方案中,内源性的硫酸酯酶被活化。在一定的实施方案中,硫酸酯酶是艾杜糖醛酸-2-硫酸酯酶,硫酸胺酶,N-乙酰半乳糖胺-6-硫酸酯酶,N-乙酰葡萄糖胺-6-硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 / 或 HSulf-6。在一定的实施方案中,细胞是哺乳动物细胞。

[0039] 根据本发明又一方面,药物组合物被提供。组合物包含治疗硫酸酯酶缺乏症的药理学有效量的被细胞产生的硫酸酯酶,和药学上可接受的载体,其中所述细胞已与包含本发明的分离的核酸分子(例如,权利要求 1-8 中的,或具有选自 SEQ ID NO :1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的核酸分子),或其表达产物(例如,选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的肽)的试剂接触。

[0040] 根据本发明又一方面,分离的人 FGE 基因的变体等位基因(其编码变体 FGE 多肽)被提供。分离的变体等位基因包含某一氨基酸序列,该氨基酸序列在 SEQ ID NO :2 中至少包含一个变异,其中这至少一个变异包含:Met1Arg ;Met1Val ;Leu20Phe ;Ser155Pro ;Ala177Pro ;Cys218Tyr ;Arg224Trp ;Asn259Ile ;Pro266Leu ;Ala279Val ;Arg327Stop ;Cys336Arg ;Arg345Cys ;Ala348Pro ;Arg349Gln ;Arg349Trp ;Arg349Trp ;Ser359Stop ;或其组合。

[0041] 根据本发明又一方面,分离的人类变体 FGE 多肽被提供。分离的人类变体 FGE 多肽包含某一氨基酸序列,该氨基酸序列在 SEQ ID NO :2 中至少含有一个变异,其中这至少一个变异包含:Met1Arg ;Met1Val ;Leu20Phe ;Ser155Pro ;Ala177Pro ;Cys218Tyr ;Arg224Trp ;Asn259Ile ;Pro266Leu ;Ala279Val ;Arg327Stop ;Cys336Arg ;Arg345Cys ;Ala348Pro ;Arg349Gln ;Arg349Trp ;Arg349Trp ;Ser359Stop ;或其组合。

[0042] 以前述任何人类变体 FGE 多肽为免疫原的抗体也被提供。这类抗体包括多克隆抗体、单克隆、嵌合的抗体,也可被进行可探测性的标记。可探测性的标记也许包含放射性元

素, 荧光化学物或酶。

[0043] 根据本发明又一方面, 产生硫酸酯酶的细胞被提供, 其中被细胞产生的活性硫酸酯酶对总硫酸酯酶的比率得到增加。细胞包含: (i) 表达增强的硫酸酯酶, 和 (ii) 表达增强的甲酰甘氨酸生成酶, 其中被细胞产生的活性硫酸酯酶对总硫酸酯酶的比率 (即, 硫酸酯酶的比活性) 相对于由缺乏甲酰甘氨酸生成酶的细胞产生的活性硫酸酯酶对总硫酸酯酶的比率至少增加 5%。在一定的实施方案中, 被细胞产生的活性硫酸酯酶对总硫酸酯酶的比率相对于由缺乏甲酰甘氨酸生成酶的细胞产生的活性硫酸酯酶对总硫酸酯酶的比率, 增加了至少 10%, 15%, 20%, 50%, 100%, 200%, 500%, 1000%。

[0044] 根据本发明进一步的方面, 在受试者中治疗硫酸酯酶缺乏症的改进方法被提供。方法包括以在受试者中治疗硫酸酯酶缺乏症的有效量对需要这种治疗的受试者施用硫酸酯酶, 其中硫酸酯酶与有效增加硫酸酯酶比活性的量的甲酰甘氨酸生成酶接触。在重要的实施方案中, 硫酸酯酶选自艾杜糖醛酸-2-硫酸酯酶, 硫酸胺酶, N-乙酰半乳糖胺-6-硫酸酯酶, N-乙酰葡萄糖胺-6-硫酸酯酶, 芳基硫酸酯酶 A, 芳基硫酸酯酶 B, 芳基硫酸酯酶 C, 芳基硫酸酯酶 D, 芳基硫酸酯酶 E, 芳基硫酸酯酶 F, 芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 HSulf-6。在一定的实施方案中, 甲酰甘氨酸生成酶被权利要求 1-8 的核酸分子或具有选自 SEQ ID NO: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的核酸所编码。在某些实施方案中, 甲酰甘氨酸生成酶是权利要求 11-15, 19, 20 的肽, 或是具有选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的序列的肽。

[0045] 本发明的这些和其它目标将被进一步结合本发明详细描述进行详述。

[0046] 序列简述

[0047] SEQ ID NO: 1 是人 FGE cDNA 的核苷酸序列。

[0048] SEQ ID NO: 2 是人 FGE cDNA (SEQ ID NO: 1) 的翻译产物的预期氨基酸序列。

[0049] SEQ ID NO: 3 是编码 SEQ ID NO: 2 的多肽的人 FGE cDNA 的核苷酸序列 (即, SEQ ID NO: 1 的核苷酸 20-1141)。

[0050] SEQ ID NO: 4 是 GenBank Acc. No. AK075459 的核苷酸序列。

[0051] SEQ ID NO: 5 是 SEQ ID NO: 4 翻译产物的预期氨基酸序列, 一未命名的具有 GenBank Acc. No. BAC11634 的蛋白产物。

[0052] SEQ ID NO: 6 是人艾杜糖醛酸-2-硫酸酯酶 cDNA (GenBank Acc. No. M58342) 的核苷酸序列。

[0053] SEQ ID NO: 7 是人艾杜糖醛酸-2-硫酸酯酶 cDNA (SEQ ID NO: 6) 的翻译产物的预期氨基酸序列。

[0054] SEQ ID NO: 8 是人硫酸胺酶 cDNA (GenBank Acc. No. U30894) 的核苷酸序列。

[0055] SEQ ID NO: 9 是人硫酸胺酶 cDNA (SEQ ID NO: 8) 的翻译产物的预期氨基酸序列。

[0056] SEQ ID NO: 10 是人 N-乙酰半乳糖胺-6-硫酸酯酶 cDNA (GenBank Acc. No. U06088) 的核苷酸序列。

[0057] SEQ ID NO: 11 是人 N-乙酰半乳糖胺-6-硫酸酯酶 cDNA (SEQ ID NO: 10) 的翻译产物的预期氨基酸序列。

[0058] SEQ ID NO: 12 是人 N-乙酰葡萄糖胺-6-硫酸酯酶 cDNA (GenBank Acc. No. Z12173)

的核苷酸序列。

[0059] SEQ ID NO :13 是人 N-乙酰葡萄糖胺-6-硫酸酯酶 cDNA (SEQ IDNO :12) 的翻译产物的预期氨基酸序列。

[0060] SEQ ID NO :14 是人芳基硫酸酯酶 A cDNA (GenBank Acc.No. X52151) 的核苷酸序列。

[0061] SEQ ID NO :15 是人芳基硫酸酯酶 A cDNA (SEQ ID NO :14) 的翻译产物的预期氨基酸序列。

[0062] SEQ ID NO :16 是人芳基硫酸酯酶 B cDNA (GenBank Acc.No. J05225) 的核苷酸序列。

[0063] SEQ ID NO :17 是人芳基硫酸酯酶 B cDNA (SEQ ID NO :16) 的翻译产物的预期氨基酸序列。

[0064] SEQ ID NO :18 是人芳基硫酸酯酶 C cDNA (GenBank Acc.No. J04964) 的核苷酸序列。

[0065] SEQ ID NO :19 是人芳基硫酸酯酶 C cDNA (SEQ ID NO :18) 的翻译产物的预期氨基酸序列。

[0066] SEQ ID NO :20 是人芳基硫酸酯酶 D cDNA (GenBank Acc.No. X83572) 的核苷酸序列。

[0067] SEQ ID NO :21 是人芳基硫酸酯酶 D cDNA (SEQ ID NO :20) 的翻译产物的预期氨基酸序列。

[0068] SEQ ID NO :22 是人芳基硫酸酯酶 E cDNA (GenBank Acc.No. X83573) 的核苷酸序列。

[0069] SEQ ID NO :23 是人芳基硫酸酯酶 E cDNA (SEQ ID NO :22) 的翻译产物的预期氨基酸序列。

[0070] SEQ ID NO :24 是人芳基硫酸酯酶 F cDNA (GenBank Acc.No. X97868) 的核苷酸序列。

[0071] SEQ ID NO :25 是人芳基硫酸酯酶 F cDNA (SEQ ID NO :24) 的翻译产物的预期氨基酸序列。

[0072] SEQ ID NO :26 是人芳基硫酸酯酶 G cDNA (GenBank Acc.No. BC012375) 的核苷酸序列。

[0073] SEQ ID NO :27 是人芳基硫酸酯酶 G (SEQ ID NO :26) 的翻译产物的预期氨基酸序列。

[0074] SEQ ID NO :28 是 HSulf-1cDNA (GenBank Acc.No. AY101175) 的核苷酸序列。

[0075] SEQ ID NO :29 是 HSulf-1cDNA (SEQ ID NO :28) 的翻译产物的预期氨基酸序列。

[0076] SEQ ID NO :30 是 HSulf-2cDNA (GenBank Acc.No. AY101176) 的核苷酸序列。

[0077] SEQ ID NO :31 是 HSulf-2cDNA (SEQ ID NO :30) 的翻译产物的预期氨基酸序列。

[0078] SEQ ID NO :32 是出现于硫酸酯酶的高度保守的六肽 L/V-FGly-X-P-S-R。

[0079] SEQ ID NO :33 是合成的 FGly 形成底物 ;其一级序列是来源于人芳基硫酸酯酶 A。

[0080] SEQ ID NO :34 是混杂寡肽 PVSLPTRSCAALLTGR。

[0081] SEQ ID NO :35 是 Ser69 寡肽 PVSLSTPSRAALLTGR。

- [0082] SEQ ID NO :36 是人 FGE- 特异性引物 1199nc。
- [0083] SEQ ID NO :37 是人 FGE- 特异性正向引物 1c。
- [0084] SEQ ID NO :38 是人 FGE- 特异性反向引物 1182c。
- [0085] SEQ ID NO :39 是包含 EcoRI 位点的人 5' -FGE- 特异性引物。
- [0086] SEQ ID NO :40 是 HA- 特异性引物。
- [0087] SEQ ID NO :41 是 c-myc- 特异性引物。
- [0088] SEQ ID NO :42 是 RGS-His₆- 特异性引物。
- [0089] SEQ ID NO :43 是来自人 FGE 制备物的胰蛋白酶解寡肽 SQNTPDSSASNLGFR。
- [0090] SEQ ID NO :44 是来自人 FGE 制备物的胰蛋白酶解寡肽 MVPIPAGVFTMGTDPPQIK。
- [0091] SEQ ID NO :45 是人 FGE2 平行进化同源物 (paralog) (GenBank GI :24308053) 的核苷酸序列。
- [0092] SEQ ID NO :46 是人 FGE2 平行进化同源物 (SEQ ID NO :45) 的翻译产物的预期氨基酸序列。
- [0093] SEQ ID NO :47 是小鼠 FGE 平行进化同源物 (GenBank GI :26344956) 的核苷酸序列。
- [0094] SEQ ID NO :48 是小鼠 FGE 平行进化同源物 (SEQ ID NO :47) 的翻译产物的预期氨基酸序列。
- [0095] SEQ ID NO :49 是小鼠 FGE 定向进化同源物 (ortholog) (GenBankGI :22122361) 的核苷酸序列。
- [0096] SEQ ID NO :50 是小鼠 FGE 定向进化同源物 (SEQ ID NO :49) 的翻译产物的预期氨基酸序列。
- [0097] SEQ ID NO :51 是果蝇 FGE 定向进化同源物 (GenBank GI :20130397) 的核苷酸序列。
- [0098] SEQ ID NO :52 是果蝇 FGE 定向进化同源物 (SEQ ID NO :51) 的翻译产物的预期氨基酸序列。
- [0099] SEQ ID NO :53 是蚊子 FGE 定向进化同源物 (GenBank GI :21289310) 的核苷酸序列。
- [0100] SEQ ID NO :54 是蚊子 FGE 定向进化同源物 (SEQ ID NO :53) 的翻译产物的预期氨基酸序列。
- [0101] SEQ ID NO :55 是密切相关的 *S. coelicolor* FGE 定向进化同源物 (GenBank GI :21225812) 的核苷酸序列。
- [0102] SEQ ID NO :56 是 *S. coelicolor* FGE 定向进化同源物 (SEQ IDNO :55) 的翻译产物的预期氨基酸序列。
- [0103] SEQ ID NO :57 是密切相关的 *C. efficiens* FGE 定向进化同源物 (GenBank GI :25028125) 的核苷酸序列。
- [0104] SEQ ID NO :58 是 *C. efficiens* FGE 定向进化同源物 (SEQ ID NO :57) 的翻译产物的预期氨基酸序列。
- [0105] SEQ ID NO :59 是 *N. aromaticivorans* FGE 定向进化同源物 (GenBank GI :23108562) 的核苷酸序列。

[0106] SEQ ID NO :60 是 *N. aromaticivorans* FGE 定向进化同源物 (SEQID NO :59) 的翻译产物的预期氨基酸序列。

[0107] SEQ ID NO :61 是 *M. loti* FGE 定向进化同源物 (GenBank GI :13474559) 的核苷酸序列。

[0108] SEQ ID NO :62 是 *M. loti* FGE 定向进化同源物 (SEQ ID NO :61) 的翻译产物的预期氨基酸序列。

[0109] SEQ ID NO :63 是 *B. fungorum* FGE 定向进化同源物 (GenBank GI :22988809) 的核苷酸序列。

[0110] SEQ ID NO :64 是 *B. fungorum* FGE 定向进化同源物 (SEQ ID NO :63) 的翻译产物的预期氨基酸序列。

[0111] SEQ ID NO :65 是 *S. meliloti* FGE 定向进化同源物 (GenBank GI :16264068) 的核苷酸序列。

[0112] SEQ ID NO :66 是 *S. meliloti* FGE 定向进化同源物 (SEQ ID NO :65) 的翻译产物的预期氨基酸序列。

[0113] SEQ ID NO :67 是微颤菌属物种 FGE 定向进化同源物 (GenBank GI :14518334) 的核苷酸序列。

[0114] SEQ ID NO :68 是微颤菌属物种 FGE 定向进化同源物 (SEQ ID NO :67) 的翻译产物的预期氨基酸序列。

[0115] SEQ ID NO :69 是 *P. putida* KT2440FGE 定向进化同源物 (GenBankGI :26990068) 的核苷酸序列。

[0116] SEQ ID NO :70 是 *P. putida* KT2440FGE 定向进化同源物 (SEQ IDNO :69) 的翻译产物的预期氨基酸序列。

[0117] SEQ ID NO :71 是 *R. metallidurans* FGE 定向进化同源物 (GenBankGI :22975289) 的核苷酸序列。

[0118] SEQ ID NO :72 是 *R. metallidurans* FGE 定向进化同源物 (SEQ IDNO :71) 的翻译产物的预期氨基酸序列。

[0119] SEQ ID NO :73 是 *P. marinus* FGE 定向进化同源物 (GenBank GI :23132010) 的核苷酸序列。

[0120] SEQ ID NO :74 是 *P. marinus* FGE 定向进化同源物 (SEQ ID NO :73) 的翻译产物的预期氨基酸序列。

[0121] SEQ ID NO :75 是 *C. crescentus* CB15 FGE 定向进化同源物 (GenBank GI :16125425) 的核苷酸序列。

[0122] SEQ ID NO :76 是 *C. crescentus* CB15 FGE 定向进化同源物 (SEQID NO :75) 的翻译产物的预期氨基酸序列。

[0123] SEQ ID NO :77 是 *M. tuberculosis* Ht37Rv FGE 定向进化同源物 (GenBank GI :15607852) 的核苷酸序列。

[0124] SEQ ID NO :78 是 *M. tuberculosis* Ht37Rv FGE 定向进化同源物 (SEQ ID NO :77) 的翻译产物的预期氨基酸序列。

[0125] SEQ ID NO :79 是出现在 FGE 定向进化同源物和平行进化同源物的亚结构域 3 的

高度保守六肽。

[0126] SEQ ID NO :80 是具有 GenBank Acc.No. :CA379852 的 FGE 定向进化同源物 EST 片段的核苷酸序列。

[0127] SEQ ID NO :81 是具有 GenBank Acc.No. :AI721440 的 FGE 定向进化同源物 EST 片段的核苷酸序列。

[0128] SEQ ID NO :82 是具有 GenBank Acc.No. :BJ505402 的 FGE 定向进化同源物 EST 片段的核苷酸序列。

[0129] SEQ ID NO :83 是具有 GenBank Acc.No. :BJ054666 的 FGE 定向进化同源物 EST 片段的核苷酸序列。

[0130] SEQ ID NO :84 是具有 GenBank Acc.No. :AL892419 的 FGE 定向进化同源物 EST 片段的核苷酸序列。

[0131] SEQ ID NO :85 是具有 GenBank Acc.No. :CA064079 的 FGE 定向进化同源物 EST 片段的核苷酸序列。

[0132] SEQ ID NO :86 是具有 GenBank Acc.No. :BF189614 的 FGE 定向进化同源物 EST 片段的核苷酸序列。

[0133] SEQ ID NO :87 是具有 GenBank Acc.No. :AV609121 的 FGE 定向进化同源物 EST 片段的核苷酸序列。

[0134] SEQ ID NO :88 是 HSulf-3cDNA 核苷酸序列。

[0135] SEQ ID NO :89 是 HSulf-3cDNA (SEQ ID NO :88) 的翻译产物的预期氨基酸序列。

[0136] SEQ ID NO :90 是 HSulf-4cDNA 核苷酸序列。

[0137] SEQ ID NO :91 是 HSulf-4cDNA (SEQ ID NO :90) 的翻译产物的预期氨基酸序列。

[0138] SEQ ID NO :92 是 HSulf-5cDNA 核苷酸序列。

[0139] SEQ ID NO :93 是 HSulf-5cDNA (SEQ ID NO :92) 的翻译产物的预期氨基酸序列。

[0140] SEQ ID NO :94 是 HSulf-6cDNA 核苷酸序列。

[0141] SEQ ID NO :95 是 HSulf-6cDNA (SEQ ID NO :94) 的翻译产物的预期氨基酸序列。

[0142] 附图简述

[0143] 图 1 :在缺乏 (A) 或存在 (B) 来自牛睾丸微粒体的可溶性抽提物的条件下培育后的 P23 的 MALDI-TOF 质谱图示。

[0144] 图 2 :来源于人 FGE 和 PFAM-DUF323 种子的 21 种蛋白的排列的系统发生树。

[0145] 图 3 :人和鼠 FGE 基因位点的组织。外显子显示为以盒子和亮盒子 (鼠的位点) 表示。内含子线上的数字指出了以 kb 表示的内含子大小。

[0146] 图 4 :显示 FGE 表达质粒 pXMG. 1. 3 结构图的图表。

[0147] 图 5 :柱状图描述 FGE 表达质粒瞬时转染的 36F 细胞中的 N-乙酰半乳糖胺-6-硫酸酯酶活性。

[0148] 图 6 :柱状图描述 FGE 表达质粒瞬时转染的 36F 细胞中的 N-乙酰半乳糖胺-6-硫酸酯酶比活性。

[0149] 图 7 :柱状图描述 FGE 表达质粒瞬时转染的 36F 细胞中的 N-乙酰半乳糖胺-6-硫酸酯酶生产。

[0150] 图 8 :描述以 FGE 表达质粒瞬时转染的 30C6 细胞中的艾杜糖醛酸-2-硫酸酯酶活

性。

[0151] 图 9 :描述了囊括本发明的特征的试剂盒。

[0152] 发明详述

[0153] 本发明包括对编码甲酰甘氨酸生成酶 (FGE) 的基因的发现,此酶负责发生在硫酸酯酶上的对硫酸酯酶功能必需的独特翻译后修饰:形成 L-C_α-甲酰甘氨酸 (a. k. a. FGly 和 / 或 2-氨基-3-氧代丙酸)。已发现,出人意料地, FGE 基因中的突变引起受试者中的多种硫酸酯酶缺乏症 (MSD) 的发展。也已发现,出人意料地, FGE 增强硫酸酯酶的活性,包括但不限于,艾杜糖醛酸-2-硫酸酯酶,硫酸胺酶, N-乙酰半乳糖胺-6-硫酸酯酶, N-乙酰葡萄糖胺-6-硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 HSulf-6, 及在申请号为 20030073118, 20030147875, 20030148920, 20030162279 和 20030166283 的 U. S. 临时申请中所述的硫酸酯酶 (其内容被清楚地整合在本文中)。考虑到这些发现,本发明的分子可用于诊断和 / 或治疗多种硫酸酯酶缺乏症,以及其它硫酸酯酶缺乏症的治疗。

[0154] 在诊断多种硫酸酯酶缺乏症中应用本发明的分子的方法被提供。

[0155] 此外,为了对硫酸酯酶上 FGly 的形成进行调节的目的而在体内或体外应用这些分子的方法,治疗与这种修饰相关的病症的方法,及在制备用于治疗多种硫酸酯酶缺乏症及其它硫酸酯酶缺乏症的治疗性制剂中有用的组合物也被提供。

[0156] 本发明因此在几个方面中包括调节硫酸酯酶上的 FGly 的形成的多肽,编码这些多肽的分离的核酸,前述的功能性修饰物和变体,前述的有用片段,以及治疗、诊断和研究的方法、组合物和与其相关的工具。

[0157] “C_α-甲酰甘氨酸生成活性”指分子在底物上形成 FGly 或增强 FGly 形成的能力。底物也许是如本文其它部分所述的硫酸酯酶,或合成的寡肽 (参见,例如 SEQ ID NO :33, 和实施例)。底物优选包含 SEQ IDNO :32 的保守六肽 [L/V-C(S)-X-P-S-R]。分析 FGly 形成的方法如本领域中 (参见,例如 Dierks, T. 等, Proc. Natl. Acad. Sci. U. S. A., 1997, 94 : 11963-11968), 和本文其它部分 (参见,例如实施例) 所描述。在本文中所用的“分子”包含“核酸”和“多肽”。FGE 分子能在体内和体外形成 FGly, 或增强 / 增加 FGly 的形成。

[0158] 在本文中所用的“增强 (或“增加”)”C_α-甲酰甘氨酸生成活性,典型地指增加 FGE 和 / 或它编码的多肽表达。增加的表达指增加 (即,到可探测的程度) 本发明任意核酸 (如本文其它部分所描述的 FGE 核酸) 的复制、转录和 / 或翻译,既然上调任何这些过程将导致基因 (核酸) 所编码多肽的浓度 / 数量的增加。增强 (或增加) C_α-甲酰甘氨酸生成活性也指防止或抑制 FGE 的降解 (例如通过增加的泛素化作用)、下调等,所述降解和下调导致例如相对于对照被增加的或稳定的 FGE 分子 t_{1/2} (半衰期)。下调或减少的表达指基因和 / 或它编码的多肽减少的表达。通过使用本领域已知的任何适合的方法例如核酸杂交或抗体探测方法,分别探测基因 (例如 FGE) mRNA 水平或基因所编码的多肽的蛋白表达水平相对于对照的增加或是减少,可直接测定基因表达的上调或下调。FGE 基因表达的上调或下调也能通过探测 C_α-甲酰甘氨酸生成活性的变化而间接测定。

[0159] 在本文中所用的“表达”指核酸和 / 或多肽表达,以及多肽分子的活性 (例如分子的 C_α-甲酰甘氨酸生成活性)。

[0160] 本发明的一个方面包括对编码 FGE 的 cDNA 的克隆。根据本发明的 FGE 是包含 SEQ ID NO :1 的核酸分子的分离的核酸分子,并编码具有 C_α-甲酰甘氨酸生成活性的多肽。人 FGE cDNA 的序列以 SEQ ID NO :1 而出现,此 cDNA 编码的蛋白产物的预期氨基酸序列以 SEQ ID NO :2 出现。

[0161] 在本文中所用的受试者是哺乳动物或非人类的哺乳动物。在所有实施方案中,人 FGE 和人受试者都是优选。

[0162] 本发明因此在一个方面包括分离的 FGE 多肽,编码此多肽的 cDNA,前述的功能性修饰物和变体,前述的有用片段,以及与其相关的诊断和治疗。

[0163] 在本文中所用的关于核酸的术语“分离的”表示:(i) 在体外通过例如聚合酶链式反应 (PCR) 所扩增的;(ii) 通过克隆所重组性地产生的;(iii) 如通过切断和凝胶分离所纯化的;或 (iv) 通过例如化学合成所合成。分离的核酸是容易通过本领域众所周知的重组 DNA 技术操作的类型。因此,包含在一定载体中的核苷酸序列被认为是分离的,其中所述载体的 5' 和 3' 限制性位点已知或者其聚合酶链式反应 (PCR) 引物序列已被公开,但在其天然宿主中以其天然状态存在的核酸不是分离的。分离的核酸可被充分纯化,但不必这样。例如,在克隆或表达载体内的分离的核酸不纯,这在于它在其寄生的细胞中也许只包含很少百分比的物质。然而,这种核酸是分离的,如本文中所用的术语一样,因为它容易通过本领域中一般技术人员所知道的标准技术被操作。

[0164] 在本文中所用的关于多肽的术语“分离的”表示从其天然环境中被以充分纯的形式分开以至于它可被操作于或用于本发明的任一目的。因此,“分离的”表示纯度足以 (i) 用于生产和 / 或分离抗体, (ii) 用作分析中的试剂, (iii) 用于测序, (iv) 用作治疗等。

[0165] 根据本发明,编码具有 C_α-甲酰甘氨酸生成活性的 FGE 多肽的分离的核酸分子包括:(a) 核酸分子,其在严格条件下与 SEQ ID NO :1 的核酸所组成的分子杂交并编码具有 C_α-甲酰甘氨酸生成活性的 FGE 多肽, (b) (a) 的删除、添加和置换物,其编码各自的具有 C_α-甲酰甘氨酸生成活性的 FGE 多肽, (c) 因为遗传密码简并性而与 (a) 或 (b) 的核酸分子在密码子序列中不同的核酸分子,和 (d) (a), (b) 或 (c) 的互补链。在本文中所用的“互补链”包括“(a), (b) 或 (c) 的全长互补链或 100% 的互补链”。

[0166] 也具有 C_α-甲酰甘氨酸生成活性的本发明的 FGE 核酸的同源物和等位基因也被本发明所包含。本文所述的同源物,包括本文其它部分所鉴别的分子(参见例如 SEQ ID NOs : 4, 5, 45-78, 和 80-87) 即定向进化同源物和平行进化同源物。进一步的,同源物能依照本发明的指导以及传统技术所鉴别。既然本文所述的 FGE 同源物全都具有 C_α-甲酰甘氨酸生成活性,它们能与入 FGE 分子在本发明的所有方面中互换性使用。

[0167] 因此,本发明的一个方面是那些编码 FGE 多肽并在严格条件下与 SEQ ID NO :1 的编码区所组成的核酸分子杂交的核酸序列。在重要的实施方案中,如在此所用的术语“严格条件”指本领域所熟悉的参数。对核酸,被称为严格的杂交条件典型的是在低离子强度和恰好低于 DNA 杂交复合物熔点 (T_m) (典型地,低于杂交物 T_m 约 3°C) 的温度下。更高的严格性限定了探针序列和靶之间更专一的相关性。核酸杂交中所用的严格条件在本领域中众所周知,可在汇编这类方法的参考文献中找到,例如 Molecular Cloning :A Laboratory Manual, J.Sambrook 等 编, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, 或 Current Protocols in Molecular Biology,

F. M. Ausubel 等编, John Wiley & Sons, Inc., New York. “严格条件”的实例是在 6xSSC 中于 65°C 下杂交。另一严格条件的实例是在杂交缓冲液中于 65°C 下杂交, 杂交缓冲液由 3.5xSSC, 0.02% Ficoll, 0.02% 聚乙烯吡咯烷酮, 0.02% 牛血清白蛋白, 2.5mM NaH_2PO_4 [pH7], 0.5% SDS, 2mM EDTA 组成 (SSC 是 0.15M 氯化钠 / 0.15M 柠檬酸钠, pH7; SDS 是十二烷基硫酸钠; 而 EDTA 是乙二胺四乙酸)。杂交之后, 转上 DNA 的膜在 2xSSC 中于室温下清洗, 然后在最高 68°C 的温度下于 0.1xSSC/0.1xSDS 中清洗。在进一步的实施例中, 杂交水溶液的使用的替代是杂交甲酰胺溶液的使用。应用例如 50% 甲酰胺溶液和 42°C, 严格的杂交条件能因此被实现。有其它能被应用的条件、试剂等, 并将导致类似的严格程度。技术人员将熟悉这类条件, 因此它们不在此处给出。然而, 需要被理解的是, 技术人员将能以允许清晰鉴别本发明 FGE 核酸的同源物和等位基因的方式操控条件。技术人员也将熟悉对于细胞和之后将被常规分离的这类分子的表达库进行筛选, 然后再分离相关的核酸分子和序列的方法。

[0168] 一般地, 同源物和等位基因典型地将分别与 SEQ ID NO:1 和 SEQ ID NO:2 具有至少 40% 核苷酸同一性和 / 或至少 50% 氨基酸同一性, 在某些情况下将具有至少 50% 核苷酸同一性和 / 或至少 65% 氨基酸同一性, 而在其它情况下将具有至少 60% 核苷酸同一性和 / 或至少 75% 氨基酸同一性。在进一步的情形中, 同源物和等位基因典型地将分别与 SEQ ID NO:1 和 SEQ ID NO:2 具有至少 90%, 95% 或甚至 99% 的核苷酸同一性和 / 或至少 95%, 98% 或甚至 99% 的氨基酸同一性。同一性可应用多种由 NCBI (Bethesda, Maryland) 开发的公开可得软件工具计算得到。示例的工具包括 Altschul SF 等的启发式算法 (J Mol Biol, 1990, 215:403-410), 也称为 BLAST。Pairwise 和 ClustalW 比对 (BLOSUM30 矩阵设置) 以及 Kyte-Doolittle 水疗分析可应用公开 (EMBL, Heidelberg, Germany) 和商业 (例如来自 Oxford Molecular Group/Genetics Computer Group, Madison, WI 的 MacVector 序列分析软件) 的类型而获得。前述核酸的 Watson-Crick 互补链也被本发明所包括。

[0169] 在对 FGE 相关基因如 FGE 的同源物和等位基因的筛选中, Southern 印迹可应用前述条件以放射性探针来完成。对 DNA 最终转移上去的膜进行清洗之后, 此膜可放置到 X-射线胶片或磷成像平板 (phosphorimager plate) 以探测放射性信号。

[0170] 在此给定关于全长人 FGE cDNA 克隆的指导, 则对应于人 FGE 基因的其它哺乳动物序列如小鼠 cDNA 克隆可应用标准菌落杂交技术从 cDNA 库中分离。

[0171] 本发明也包括含有天然物质中出现的那些密码子的替代物的简并核酸。例如, 丝氨酸残基被密码子 TCA, AGT, TCC, TCG, TCT 和 AGC 编码。因此, 本领域一般技术人员将很明白, 任何丝氨酸编码核苷酸三联体可被用于在体内或体外指导蛋白质合成装置, 以将丝氨酸残基整合进入正在延伸的 FGE 多肽。类似地, 编码其它氨基酸残基的核苷酸序列三联体包括但不限于: CCA, CCC, CCG 和 CCT (脯氨酸密码子); CGA, CGC, CGG, CGT, AGA 和 AGG (精氨酸密码子); ACA, ACC, ACG 和 ACT (苏氨酸密码子); AAC 和 AAT (天冬酰胺密码子); 及 ATA, ATC 和 ATT (异亮氨酸密码子)。其它氨基酸残基可类似地被若干核苷酸序列编码。因此, 本发明包括与生物学上分离的核酸在密码子序列中因遗传密码简并性而有所不同的简并核酸。

[0172] 本发明也提供分离的 SEQ ID NO:1 或 SEQ ID NO:3 或其互补链的独特片段。独特片段是这样的类型, 它是更大的核酸的“标志”。例如, 独特片段长度足以确保其精确序列

无法在人基因组中位于上述 FGE 核酸（及人的等位基因）之外的分子中找到。那些本领域中的一般技术人员可不必应用常规之外的程序来测定片段是否在人基因组中是独特的。然而，独特片段排除完全由选自 SEQ ID NO :4 和 / 或其它如本申请的申请日以前发表过的序列的核苷酸序列所组成的片段。

[0173] 完全由前述 GenBank 保藏库描述的序列所组成的片段不包括任何对于本发明序列而言独特的核苷酸。因此，根据本发明的独特片段必须包含除那些 GenBank 保藏库中的确切序列或其片段之外的核苷酸序列。区别可以是相对于 GenBank 序列的添加、删除或置换，或可以是完全不同于 GenBank 序列的序列。

[0174] 独特片段能在 Southern 和 Northern 印迹分析中用作探针以鉴别这类核酸，或能用于如那些采用 PCR 的扩增分析。如本领域技术人员已知的，大探针如 200, 250, 300 或更多的核苷酸优选用于一定的应用如 Southern 和 Northern 印迹，而更小的片段将优选用于例如 PCR。独特片段也能被用于产生融合蛋白，以产生抗体或测定如实施例证明的多肽片段的结合或产生免疫分析组分。同样地，独特片段可被用于产生例如在抗体制备、免疫分析或治疗应用中有效的 FGE 多肽非融合片段。独特片段进一步地可被用作反义分子以抑制 FGE 核酸和多肽各自的表达。

[0175] 如将被本领域技术人员所认识到的一样，独特片段的大小将依赖于其遗传密码中的保守性。因此，SEQ ID NO :1 或 SEQ ID NO :3 的部分区域和互补链将需要更长的节段来实现独特，而其它的只需要短节段，典型的是长 12 和 32 个核苷酸之间（例如 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 和 32 碱基）或更长，长到公开序列的全长。如上述所提及的，此公开内容倾向于包含每一序列的每一片段，起点在第一核苷酸、第二核苷酸等等，直到离末端剩下 8 个核苷酸，而终点在从第 8、9、10 核苷酸等等开始直到恰好最后一个核苷酸的任何位置（前提为序列是如上述的独特片段）。事实上 SEQ ID NO :1 区域的以核苷酸 1 开始而在核苷酸 1180 终止或 SEQ ID NO :3 区域的以核苷酸 1 开始而在核苷酸 1122 终止的任何节段或其互补链，长度为 20 或更多核苷酸都将是独特的。本领域技术人员对选择这类序列的方法（一般是基于独特片段选择性区分目的序列与人类基因组中其它序列的能力）很精通，虽然可以进行体外的证实性的杂交和序列分析。

[0176] 如上述所提及的，本发明包含选择性结合编码 FGE 多肽的核酸分子以减少 FGE 活性的反义寡核苷酸。

[0177] 在本文中所用的术语“反义寡核苷酸”或“反义”描述一定的寡核苷酸，该寡核苷酸是在生理条件下与包含特定基因的 DNA 或该基因的 mRNA 转录产物杂交，而因此抑制该基因的转录和 / 或该 mRNA 的翻译的寡核糖核苷酸，寡脱氧核糖核苷酸，修饰后的寡核糖核苷酸，或修饰后的寡脱氧核糖核苷酸。反义分子被设计成与靶基因或转录物杂交以干扰靶基因的转录或翻译。本领域技术人员将认识到反义寡核苷酸的确切长度和它与其目标互补的程度将依赖于所选的特异靶，包括靶的序列和构成那一序列的特定碱基。优选反义寡核苷酸被构建和安排成在生理条件下选择性地与靶结合，即相对于靶细胞中的任何其它序列在生理条件下更充分地与其靶序列杂交。基于 SEQ ID NO :1 或基于等位基因或同源基因组和 / 或 cDNA 序列，本领域中的技术人员能容易地挑选和合成出大量合适的反义分子中的任何类型以根据本发明而应用。为了具有充分的选择性和充分有能力进行抑制，这类反义寡核苷酸应包含与靶互补的至少 10 个和更多连续碱基，优选至少 15 个，虽然在某些情形中长度短至

7 个碱基的修饰后的寡核苷酸被成功用作反义寡核苷酸 (Wagner 等, Nat. Med, 1995, 1(11) : 1116-1118 ; Nat. Biotech. , 1996, 14 : 840-844)。最优选的是, 反义寡核苷酸包含 20-30 个碱基的互补序列。虽然对基因或 mRNA 转录物任何区段反义的寡核苷酸可被选出, 在优选实施方案中, 反义寡核苷酸对应于 N 端或 5' 上游位点例如翻译起点、转录起点或启动子位点。此外, 3' - 非翻译区可被反义寡核苷酸所靶向。对 mRNA 拼接位点的靶向也在本领域中应用, 但如果替代性的 mRNA 拼接发生, 则可能不是那么被优选。此外, 反义物优选靶向于不希望出现 mRNA 二级结构 (参见, 例如 Sainio 等, Cell Mol. Neurobiol. 14(5) : 439-457, 1994) 和不希望出现蛋白结合的位点。最后, 虽然 SEQ ID No : 1 公开了 cDNA 序列, 本领域中的一般技术人员可容易地得到对应于此序列的基因组 DNA。因此, 本发明也提供与对应于 SEQ ID NO : 1 的基因组 DNA 互补的反义寡核苷酸。类似地, 等位基因或同源 FGE cDNAs 和基因组 DNAs 的反义物也能应用, 而不需要过度的实验。

[0178] 在一组实施方案中, 本发明的反义寡核苷酸可由“天然”脱氧核糖核苷酸, 核糖核苷酸, 或它们的任意组合所组成。也就是, 一种天然核苷酸的 5' 末端和另一天然核苷酸的 3' 末端可通过核苷间的磷酸二酯键共价连接, 如在天然系统中的一样。这些寡核苷酸可通过领域内认可的方法制备, 所述方法可人工或通过自动合成仪而实施。它们也可通过载体被重组性地产生。

[0179] 然而, 在优选的实施方案中, 本发明的反义寡核苷酸也可包括“修饰后的”寡核苷酸。也就是, 寡核苷酸可通过许多不会阻止它们杂交至其目标但增强它们的稳定性或靶向性或者增强它们的治疗效力的方式被修饰。

[0180] 在本文中所用的术语“修饰后的寡核苷酸”描述这样的寡核苷酸, 其中 (1) 其核苷酸中的至少两个通过合成性的核苷间连接而共价连接 (即, 一个核苷酸 5' 末端和另一核苷酸 3' 末端之间除磷酸二酯键之外的连接) 和 / 或 (2) 通常不与核酸连接的化学基团已被共价附着到寡核苷酸。优选的合成性核苷间连接是硫代磷酸酯, 烷基磷酸酯, 二硫代磷酸酯, 磷酸酯, 烷基硫代硫酸酯 (alkylphosphonothioates), 氨基磷酸酯, 氨基甲酸酯, 碳酸酯, 磷酸三酯, acetamidates, 羧甲基酯和肽。

[0181] 术语“修饰后的寡核苷酸”也包括具有被共价修饰的碱基和 / 或糖的寡核苷酸。例如, 修饰后的寡核苷酸包括具有已在除 3' 位置羟基和 5' 位置磷酸基之外的位置共价附着到低分子量有机基团的糖骨架的寡核苷酸。因此修饰后的寡核苷酸可包括 2'-O- 烷基化核糖基团。此外, 修饰后的寡核苷酸可包括糖, 例如代替核糖的阿拉伯糖。

[0182] 本发明因此涉及含有修饰后的反义分子与药学上可接受的载体的药物制剂, 所述反义分子与编码 FGE 多肽的核酸互补并在生理条件下杂交。反义寡核苷酸可作为药物组合物的一部分被施用。这类药物组合物可包括与任何本领域中已知的生理性和 / 或药学上可接受的标准载体组合的反义寡核苷酸。组合物应是消毒过的, 并以适合对患者施用的单位重量或体积而含有药学有效量的反义寡核苷酸。术语“药学上可接受的”表示不干扰活性成分的生物活性的效力的非毒性物质。术语“生理上可接受的”指与生物系统如细胞、细胞培养物、组织或器官相容的非毒性物质。载体的特征将依赖于施用途径。生理上和药学上可接受的载体包括本领域众所周知的稀释剂、填充物、盐、缓冲剂、稳定物、增溶剂和其它物质。

[0183] 本发明也包括在细胞中增加 C_α- 甲酰甘氨酸生成活性的方法。在重要的实施方案

中,这通过应用载体(“表达载体”和/或“靶向载体”)而完成。

[0184] 在本文中所述的“载体”可以是多种某类核酸中的任何类型,所需序列可通过限制和连接而插入所述核酸以在不同遗传环境间运输或在宿主细胞中表达。载体典型地由 DNA 组成,虽然 RNA 载体也可用。载体包括但不限于质粒、噬菌粒和病毒基因组。克隆载体是能在宿主细胞中复制,并进一步以一个或更多核酸内切酶限制性位点为特征的类型,载体能以可测方式在所述核酸内切酶限制性位点处被切断,所需 DNA 序列可连接进入所述位点处而使得新重组载体保持其在宿主细胞中复制的能力。在质粒的情形中,所需序列的复制可因为质粒在宿主菌内增加拷贝数目而发生很多次,或在宿主通过有丝分裂而复制之前在每一宿主中恰好单独一次。在噬菌体的情形中,复制可在裂解期期间主动发生或在溶原期被动发生。“表达载体”是这样的类型,所需 DNA 序列(例如 SEQ ID NO :3 的 FGE cDNA)通过限制和连接插入其中而使得所需 DNA 序列被可操作性地连接于调节序列并可表达为 RNA 转录物。载体可进一步包含一个或更多适合用于鉴别细胞已被或未被载体转化或转染的标记序列。标记包括,例如增加或减少对抗生素或其它化合物的抗性或敏感性的蛋白的编码基因,其活性通过本领域中已知的标准分析可测的酶(例如 β -半乳糖苷酶或碱性磷酸酶)的编码基因,及可视性地影响转化或转染细胞、宿主、菌落或菌斑的表型(如绿色荧光蛋白)的基因。

[0185] “靶向载体”是典型地含有一定靶向性结构/序列的类型,所述靶向性结构/序列用于例如在内源性基因中(例如,在外显子和/或内含子序列内),内源性基因启动子序列中,或内源性基因启动子序列上游插入调节序列。在另一实施例中,靶向载体可包含目的基因(例如 SEQ ID NO :1 的 cDNA 所编码的)和对基因靶向于基因组中优选位置所需的其它序列(例如,转录活跃位置如无关基因的内源启动子下游)。靶向性构建体和载体的构建在 U. S. Patents 5,641,670 和 6,270,989 中详细描述,其被清楚地整合在本文中作为参考。

[0186] 事实上,任何能被异源性 DNA 或 RNA 转化并能在培养基中生长或保存的细胞(原核或真核),可被用在本发明的实践中。实施例包括细菌细胞如大肠杆菌,昆虫细胞,和哺乳动物如人、小鼠、仓鼠、猪、山羊、灵长类等的细胞。它们可以是原代或次级细胞株(其在培养中显示有限数目的平均群体倍增而不是永久的)和永久的细胞系(其在培养中显示明显的无限寿命)。原代和次级细胞包括,例如,成纤维细胞,角质化细胞,上皮细胞(例如乳腺上皮细胞,肠上皮细胞),内皮细胞,神经胶质细胞,神经细胞,血液的组成成分(例如淋巴细胞,骨髓细胞),肌肉细胞和这些体细胞类型的前体包括胚胎干细胞。在细胞将被在基因治疗中应用时,原代细胞优选从施予操作后的细胞的个体中获得。然而,原代细胞能从相同物种的供体(而不是受体)获得。可与本发明的 DNA 构建物和方法一起应用的永久的人细胞系实例包括但不限于 HT-1080 细胞(ATCC CCL 121),HeLa 细胞和 HeLa 细胞衍生物(ATCC CCL 2,2.1 和 2.2),MCF-7 乳腺癌细胞(ATCC BTH 22),K-562 白血病细胞(ATCC CCL 243),KB 癌细胞(ATCC CCL 17),2780AD 卵巢癌细胞(Van der Blick, A. M. 等, Cancer Res, 48 : 5927-5932(1988),Raji 细胞(ATCC CCL 86),WiDr 结肠腺癌细胞(ATCC CCL 218),SW620 结肠腺癌细胞(ATCC CCL 227),Jurkat 细胞(ATCC TIB 152),Namalwa 细胞(ATCC CRL1432),HL-60 细胞(ATCC CCL 240),Daudi 细胞(ATCC CCL 213),RPMI 8226 细胞(ATCC CCL 155),U-937 细胞(ATCC CRL 1593),Bowes 黑色素瘤细胞(ATCC CRL 9607),WI-38VA13 亚系 2R4 细胞(ATCC CLL 75.1),和 MOLT-4 细胞(ATCC CRL 1582),CHO 细胞,和 COS 细胞,以及通过

人细胞和另一物种细胞融合而产生的异杂交瘤细胞。次级人成纤维细胞株,如 WI-38(ATCC CCL 75) 和 MRC-5(ATCC CCL 171) 也可被应用。对可在实践本发明的方法中应用的细胞类型的进一步讨论,在 U. S. Patents 5,641,670 和 6,270,989 中被描述。无细胞的转录系统也可代替细胞。

[0187] 本发明的细胞被保存在本领域已知的条件下,这将引起 FGE 蛋白或其功能性片段的表达。应用所述方法,表达的蛋白可从细胞裂解物或细胞上清中纯化。根据此方法制得的蛋白质能被制成药学上有用的配剂,并通过本领域中已知的传统药学途径而递送至人或非人的动物(例如口服,静脉内,肌肉内,鼻内,气管内或皮下)。如本文其它部分所描述的,重组细胞可以是永生化的、原代的或次级的细胞,优选人的细胞。来自其它物种的细胞的应用可在一定的情形中是合意的,所述情形为非人细胞有利于在所产生的非人 FGE 在药学上有用的情况下的蛋白产生的目的。

[0188] 在本文中所用的编码序列和调控序列,当它们以将编码序列的表达或转录置于调控序列的影响或控制之下的方式而共价连接时,被表述成“可操作性地”连接。如果期望编码序列被翻译成功能性蛋白质,而假如 5' 调控序列中的启动子的诱导导致编码序列的转录,并且两个 DNA 序列间的连接 (1) 不会导致移码突变的引入,(2) 不会干扰启动子区域指导编码序列转录的能力,或 (3) 不会干扰相应的 RNA 转录物被翻译成为蛋白的能力,则两个 DNA 序列被表述成“可操作性地”连接。因此,如果启动子区域能影响那一 DNA 序列的转录从而所得的转录物可被翻译成所期望的蛋白或多肽,则启动子区域将可操作性地连接编码序列。

[0189] 基因表达所需的调控序列的精确本质可在物种或细胞类型间变动,但将一般性地包括,如所必需的,于转录和翻译各自启动有关的 5' 非转录和 5' 非翻译序列,如 TATA 盒,加帽序列,CAAT 序列和类似物。尤其是,这类 5' 非转录调控序列将包括包含对可操控性连接的基因进行转录控制的启动子序列的启动子区域。调控序列也可包括所期望的增强子序列或上游激动子序列。本发明的载体可随意包括 5' 前导或信号序列。对合适载体的选择和设计是在本领域一般技术人员的能力和指导之内。

[0190] 包含所有表达必需的元件的表达载体可商业性地获得,并为本领域技术人员所知。参见,例如 Sambrook 等, *Molecular Cloning: A Laboratory Manual, Second Edition*, Cold Spring Harbor Laboratory Press, 1989。细胞通过编码 FGE 多肽或其片段或变体的异源性 DNA (RNA) 被引入而基因工程化。那一 DNA (RNA) 被置于转录元件的可操作性控制之下,以允许异源 DNA 在宿主细胞中的表达。

[0191] 用于在哺乳动物细胞中表达 mRNA 的优选系统例如包含选择标记如给予 G418 抗性的基因(其促进对稳定性转染的细胞系的选择)和人细胞巨化病毒(CMV) 增强子-启动子序列的 pRc/CMV(可从 Invitrogen, Carlsbad, CA 得到)。此外,适合在灵长类和犬类细胞系中进行表达的类型有 pCEP4 载体(Invitrogen, Carlsbad, CA),其包含促进质粒作为多拷贝的染色体外元件而保存的 EB 病毒(EBV) 复制起点。另一表达载体是包含多肽延伸因子 1 α 的启动子的 pEF-BOS 质粒,其有效刺激了体外的转录。质粒被 Mishizuma 和 Nagata(*Nuc. Acids Res.* 18 :5322, 1990) 描述,而其在转染实验中的应用已得到公开,例如, Demoulin(*Mol. Cell. Biol.* 16 :4710-4716, 1996)。又一优选的表达载体是 Stratford-Perricaudet 所述的腺病毒,其 E1 和 E3 蛋白有缺陷(*J. Clin. Invest.* 90 :

626-630, 1992)。腺病毒作为 Adeno. P1A 重组体的应用被 Warnier 等所公开,用于在小鼠皮内注射以得到抗 P1A 的免疫化 (Int. J. Cancer, 67 :303-310, 1996)。

[0192] 本发明也包括所谓的表达试剂盒,其允许技术人员制备所期望的表达载体或载体。这类表达试剂盒至少包括每一前面讨论的编码序列的单独部分。其它成分可按需要添加,只要前面提到的所需的序列被包括在内。

[0193] 也应认识到,本发明包含上述的包含表达 FGE cDNA 序列的载体转染宿主细胞和细胞系的应用,这些细胞是原核(例如大肠杆菌)或真核性(例如 CHO 细胞, COS 细胞, 酵母表达系统和在昆虫细胞中表达的杆状病毒)的。尤其有用的是哺乳动物细胞如人,小鼠,仓鼠,猪,山羊,灵长动物等的细胞。它们可以是多种组织类型,包括如本文其它部分所描述的原代细胞和永生化的细胞系。具体实例包括 HT-1080 细胞, CHO 细胞, 树状细胞, U293 细胞, 外周血白细胞, 骨髓干细胞, 胚胎干细胞, 和昆虫细胞。本发明也允许细胞和动物中的 FGE 基因“敲除”的构建,为研究 FGE 活性的某些方面提供材料。

[0194] 本发明也提供被前述 FGE 核酸编码的分离的多肽(包括完整蛋白和部分蛋白),也包括 SEQ ID NO :2 的多肽和其独特片段。这类多肽可用于,例如,单独或作为融合蛋白的部分去产生作为免疫分析成分的抗体。多肽可以是生物样本包括组织或细胞匀浆中分离的,也可是在多种原核或真核表达系统中重组性表达的,所述重组性表达通过构建适合于表达系统的表达载体,将表达载体引入表达系统,及分离重组性表达的蛋白而完成。短多肽,包括抗原性肽(如被 MHC 分子呈递在细胞表面以进行免疫识别的)也能应用已良好建立的肽合成方法化学合成。

[0195] 一般而言, FGE 多肽的独特片段具有如上述讨论的与核酸相关的独特片段的特征和特性。如本领域技术人员将认识到的,独特片段的大小将依赖于一定的因素,例如片段是否组成了保守蛋白结构域的一部分。因此, SEQ ID NO :2 的部分区域将需要更长的片段以保证独特,而其它的只需要短片段,典型的是在 5 和 12 个氨基酸之间(例如 5, 6, 7, 8, 9, 10, 11 和 12 氨基酸长或者更多,包括每一整数直到全长, 287 个氨基酸长)。

[0196] 多肽的独特片段优选保持明显的多肽功能能力的那些片段。可被保持在多肽的独特片段中的功能能力包括与抗体的相互作用、与其它多肽或其片段的相互作用,与其它分子的相互作用等。一种重要的活性是作为鉴别多肽的标签而起作用的能力。本领域技术人员很精通于选择独特氨基酸序列的方法,典型的是基于独特片段从非家族成员中选择性识别出目的序列的能力。片段的序列与那些数据库中已知序列进行比较即是所需的。

[0197] 本发明包括上述 FGE 多肽的变体。在本文中所用的 FGE 多肽的“变体”是包含一种或更多对 FGE 多肽一级氨基酸序列的修饰的多肽。创造 FGE 多肽变体的修饰,典型的是被施加给编码 FGE 多肽的核酸,可包括删除、点突变、截断、氨基酸置换和氨基酸或非氨基酸部分的添加,以:1) 减少或消除 FGE 多肽的活性;2) 增强 FGE 多肽的性质,如表达系统中的蛋白稳定性或蛋白-配体结合的稳定性;3) 对 FGE 多肽提供新活性或性质,如抗原性表位的添加或可探测部分的添加;或4) 提供与 FGE 多肽受体或其它分子的相等或更好的结合。替代性地,修饰可直接施加给多肽,如通过切断,连接分子的添加,可探测部分如生物素的添加,脂肪酸的添加和类似的方式。修饰也包括包含全部或部分 FGE 氨基酸序列的融合蛋白。本领域技术人员将会熟悉预测蛋白序列变化对蛋白构象的影响的方法,并能因此根据已知方法而“设计”FGE 多肽的变体。这种方法的一个实例被 Dahiyat 和 Mayo 在 Science 278:

82-87, 1997 中描述, 其中蛋白质可被重新设计。此方法能被应用于已知的蛋白, 以仅改变多肽序列的一部分。通过应用 Dahiyat 和 Mayo 的计算方法, 具体的 FGE 多肽的变体可得到提议和被测试以测定变体是否保持了所期望的构象。

[0198] 变体可包括被专一修饰的 FGE 多肽, 所述修饰是为了改变多肽的与其生理活性无关的特征。例如, 半胱氨酸残基能被置换或删除以防止不必要的二硫键连接。类似地, 一定的氨基酸可被改变以通过消除表达系统中由蛋白酶作用的蛋白水解而增强 FGE 多肽的表达 (例如存在 KEX2 蛋白酶活性的酵母表达系统中的双碱性氨基酸残基)。

[0199] 编码 FGE 多肽的核酸的突变优选保存编码序列的氨基酸阅读框, 并优选不在核酸中创造出很可能杂交以形成二级结构的区域, 二级结构例如发卡或环结构能对变体多肽的表达有害。

[0200] 突变可通过选择氨基酸置换或多肽编码核酸中选定位置的随机突变而发生。变体多肽随后得到表达, 并进行对一种或更多活性的检测, 以测定哪个突变为变体多肽提供了期望的性质。进一步的对多肽的氨基酸序列而言沉默的突变可提供给变体 (或非变体的 FGE 多肽), 但其提供在特定宿主中优选的翻译密码子, 或改变 mRNA 的结构以, 例如, 增强稳定性和 / 或表达。优选的在例如大肠杆菌、哺乳动物细胞等中的核酸翻译密码子, 对本领域中的一般技术人员而言众所周知。其它突变也可提供给 FGE 基因或 cDNA 克隆的非编码序列以增强多肽的表达。

[0201] 技术人员将意识到保守氨基酸置换可发生在 FGE 多肽中以提供功能等同的前述多肽的变体, 即, 变体保持 FGE 多肽的功能能力。在本文中所用的“保守氨基酸置换”指不显著改变三级结构和 / 或多肽活性的氨基酸置换。变体可根据本领域一般技术人员已知的改变多肽序列的方法而制备, 并包括那些在汇编这类方法的参考文献中找到的类型, 例如 Molecular Cloning :A Laboratory Manual, J. Sambrook 等编, Second Edition, Cold Spring Harbor Laboratory Press, ColdSpring Harbor, New York, 1989, 或 Current Protocols in Molecular Biology, F. M. Ausubel 等编, John Wiley & Sons, Inc., New York。示例性的 FGE 多肽的功能等同性变体包括对 SEQ ID NO :2 的保守氨基酸置换。保守氨基酸置换包括发生在下列组群内部氨基酸中间的置换 : (a) M, I, L, V ; (b) F, Y, W ; (c) K, R, H ; (d) A, G ; (e) S, T ; (f) Q, N ; 和 (g) E, D。

[0202] 因此 FGE 多肽的功能等同的变体, 即, 保持天然 FGE 多肽功能的 FGE 多肽的变体被本发明所涉及。FGE 多肽氨基酸序列中产生功能等同性变体的保守氨基酸置换, 典型地是通过编码 FGE 多肽的核酸 (SEQ IDNOs :1, 3) 的改变所提供。这类置换可通过多种本领域一般技术人员已知的方法而得到。例如, 氨基酸置换可通过 PCR 导向突变, 根据 Kunkel 的方法 (Kunkel, Proc. Nat. Acad. Sci. U. S. A. 82 :488-492, 1985) 的定点突变, 或编码 FGE 多肽的基因的化学合成而得到。FGE 多肽的功能等同片段的活性能通过 :将编码改变后的 FGE 多肽的基因克隆进细菌或哺乳动物表达载体, 将载体引入合适的宿主细胞, 表达改变后的 FGE 多肽, 和测试在本文中公开的 FGE 多肽的功能能力 (例如 C_α- 甲酰甘氨酸生成活性等), 而得到测试。

[0203] 在本文中所述的发明具有许多应用, 其中的一部分在本文其它部分被描述。首先, 本发明允许对 FGE 多肽的分离。技术从业人员众所周知的多种方法能被用于得到分离的 FGE 分子。多肽可从天然产生多肽的细胞中通过层析方式或免疫识别而纯化。替代性地, 表

达载体可被引入细胞以引起多肽的产生。在另一方法中,mRNA 转录物可被微注射或另外地被引入细胞以引起被编码的多肽的产生。FGE mRNA 在无细胞抽提物如网织红细胞降解系统中的翻译也可被用于产生 FGE 多肽。本领域技术人员也可容易地遵循已知的分离 FGE 多肽的方法。这些包括但不限于免疫层析,HPLC,大小排阻层析,离子交换层析和免疫亲和层析。

[0204] 在一定的实施方案中,本发明也提供来源于 FGE 多肽的“显性失活”多肽。显性失活多肽是蛋白质的失活变体,其通过与细胞机制相互作用而从活性蛋白与细胞机制的相互作用中取代了活性蛋白,或与活性蛋白竞争,而减小了活性蛋白的效应。例如,与配体结合但不会转导响应于配体结合的信号显性失活受体能减小配体表达的生物学效应。同样地,与靶蛋白正常相互作用但不磷酸化靶蛋白的显性失活催化惰性激酶,能减少靶蛋白响应于细胞信号的磷酸化。类似地,结合到基因调控区域的启动子位置但不增加基因转录的显性失活转录因子能通过占据启动子结合位置而不增加转录来减小正常转录因子的效应。

[0205] 显性失活多肽在细胞中表达的最终结果是活性蛋白质的功能的减少。本领域一般技术人员能评估显性失活蛋白变体的潜力,并用标准突变技术创造一种或更多显性失活变体蛋白。参见,例如 U. S. Patent No. 5, 580, 723 和 Sambrook 等, *Molecular Cloning :A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, 1989*。技术人员随后能测试经诱变的蛋白群体的选定活性的减少和 / 或对这种活性的保持力。其它类似的用于创造和测试蛋白的显性失活变体的方法对本领域一般技术人员将是很显然的。

[0206] FGE cDNA 的分离也使技术人员可能诊断以 FGE 异常表达为特征性疾病。这些方法包括测定 FGE 基因和 / 或源于其的 FGE 多肽的表达。在前者的情形下,这类测定能通过任何标准的核酸测定分析而进行,包括聚合酶链式反应,或如下面作为例子的采用标记性杂交探针的分析。在后者的情形下,这类测定能通过任何标准的(应用例如结合到分泌出的 FGE 蛋白的抗体的)免疫分析而进行。优选的可根据本发明诊断的疾病是多种硫酸酯酶缺乏症。

[0207] 本发明也包括分离的肽结合试剂,所述试剂,例如,可以是具有选择性结合到 FGE 多肽的能力的抗体或抗体片段(“结合性多肽”)。抗体包括根据传统方法制备的多克隆和单克隆抗体。在一定的实施方案中,本发明排除了与 SEQ ID NO :4 的核酸所编码的多肽结合的结合性试剂(例如抗体)。

[0208] 值得注意的是,如本领域中众所周知的,只有抗体分子的一小部分,即抗原互补位,涉及抗体与其表位的结合(参见,一般地,Clark, W. R. (1986) *The Experimental Foundations of Modern Immunology* Wiley & Sons, Inc., New York ;Roitt, I. (1991) *Essential Immunology*, 7th Ed., Blackwell Scientific Publications, Oxford)。例如,pFc' 和 Fc 区域是补体级联的效应子但不涉及抗原结合。pFc' 区域已被酶切掉的抗体或产生时不含 pFc' 区域的抗体,它们被称为 F(ab')₂ 片段,保持了完整抗体的两个抗原结合位点。类似地,Fc 区域已被酶切掉的抗体或产生时不含 Fc 区域的抗体,它们被称为 Fab 片段,保持了完整抗体分子的一个抗原结合位点。进一步地,Fab 片段由共价结合的抗体轻链和被表示为 Fd 的抗体重链的一部分所组成。Fd 片段是抗体专一性的主要决定因素(单一 Fd 片段可与多达 10 条的不同轻链相连而不改变抗体专一性),并且 Fd 片段在分离后保持表位-结合能力。

[0209] 如本领域中众所周知的,在抗体的抗原结合部分内部有直接与抗原表位相互作用

的互补性决定区域 (CDRs), 也有保持抗体结合部位三级结构的构架区域 (FRs) (参见, 一般地, Clark, 1986 ;Roitt, 1991)。在 IgG 免疫球蛋白重链 Fd 片段和轻链中, 有通过三个互补决定区域 (CDR1 到 CDR3) 分别分开的四个构架区域 (FR1 到 FR4)。CDRs 尤其是 CDR3 区域, 更尤其是重链 CDR3, 很大程度上对抗体专一性负责。

[0210] 现在在本领域中已很好地认识到, 哺乳动物抗体的非 CDR 区域可被同种或异种特异性抗体的类似区域所替代, 而保持原始抗体的表位专一性。这在制备和应用“人源化”抗体中表现得最清晰, 其中非 - 人 CDRs 共价连接到人 FR 和 / 或 Fc/pFc' 区域以产生功能抗体。参见, 例如 U. S. patents 4, 816, 567, 5, 225, 539, 5, 585, 089, 5, 693, 762 和 5, 859, 205。因此, 例如, PCT 国际公开号 WO 92/04381 教授了人源化的鼠类 RSV 抗体的生产和应用, 其中至少鼠 FR 区域的一部分被人源性的 FR 区域所取代。这类抗体, 包括具有抗原结合能力的完整抗体的片段, 经常被表示成“嵌合的”抗体。

[0211] 因此, 正如对本领域一般技术人员而言是很明白的, 本发明也提供 $F(ab')_2$, Fab, Fv 和 Fd 片段; 其中 Fc 和 / 或 FR 和 / 或 CDR1 和 / 或 CDR2 和 / 或轻链 CDR3 区域已被同源的人或非 - 人序列所取代的嵌合的抗体; 其中 FR 和 / 或 CDR1 和 / 或 CDR2 和 / 或轻链 CDR3 区域已经被同源的人或非序列所取代的嵌合 $F(ab')_2$ 片段抗体; 其中 FR 和 / 或 CDR1 和 / 或 CDR2 和 / 或轻链 CDR3 区域已被同源的人或非 - 人序列所取代的嵌合的 Fab 片段抗体; 以及其中 FR 和 / 或 CDR1 和 / 或 CDR2 区域已被同源的人或非 - 人序列所取代的嵌合的 Fd 片段抗体。本发明也包括所谓的单链抗体。

[0212] 因此, 本发明包括与 FGE 多肽专一结合的多种大小和类型的多肽, 及 FGE 多肽与其结合伴侣的复合物。这些多肽也可由抗体技术之外的其它来源得来。例如, 这类多肽结合试剂能由简并肽库而提供, 所述简并肽库可很容易以溶液、以固定的形式作为细菌鞭毛肽展示库或噬菌体展示库而制备。也可制备含有一个或更多氨基酸的肽组合文库。文库能进一步由肽和非肽的合成部分合成。

[0213] 噬菌体展示能在鉴别根据本发明而有用的结合肽中特别有效。简而言之, 可应用传统程序制备显示 4 到约 80 个氨基酸残基插入物的噬菌体库 (应用例如 m13, fd, 或 λ 噬菌体)。插入物可代表, 例如, 完全的降解物或有偏差的阵列。然后可挑选噬菌体承载性插入物, 其结合到 FGE 多肽或 FGE 和结合伴侣的复合物。此过程可通过复选结合到 FGE 多肽或复合物的噬菌体的几个循环而重复。重复的循环导致承载特定序列的噬菌体的富集。DNA 序列分析可以进行, 以鉴别所表达多肽的序列。结合 FGE 多肽或复合物的序列的最小线性部分可被测定。应用含有插入物的偏差性的库, 可重复这些程序, 所述插入物包含最小线性部分的一部分或全部外加一个或更多附加的其上游或下游的简并残基。酵母双杂交筛选方法也可用于鉴别结合到 FGE 多肽的多肽。因此, 本发明的 FGE 多肽或其片段或 FGE 与结合伴侣的复合物能被用于筛选肽库, 包括噬菌体展示库, 以鉴别和选择本发明 FGE 多肽的肽结合伴侣。这类分子能用于, 如所述的, 筛选分析, 纯化方案, 对 FGE 功能的直接干扰, 及其它对本领域一般技术人员而言很显然的目的。

[0214] FGE 多肽或其片段也能用于分离它们的天然结合伴侣。对结合伴侣的分离可根据众所周知的方法完成。例如, 分离的 FGE 多肽能被附着于某一基质, 然后被怀疑含有 FGE 结合伴侣的溶液可应用到基质上。如果 FGE 多肽的结合伴侣在溶液中存在, 那么它将结合到基质固着的 FGE 多肽。之后结合伴侣可被分离出。充当 FGE 结合伴侣的其它蛋白质可通过

类似方法分离出,而不需要过度的实验。优选的结合伴侣是硫酸酯酶。

[0215] 本发明也提供测量 FGE 在受试者中表达水平的方法。这可通过首先得到来自受试者的测试样本而完成。测试样本可以是组织或生物液体。组织包括脑,心脏,血清,乳房,结肠,膀胱,子宫,前列腺,胃,睾丸,卵巢,胰,垂体腺,肾上腺,甲状腺,唾液腺,乳腺,肾,肝,肠,脾,胸腺,血管,骨髓,气管,和肺。在一定的实施方案中,测试样本源自心脏和血管组织,生物液体包括血液,唾液和尿。创伤性和非创伤性技术都能用于获得这类样本,并在本领域中被很好地记载。在分子水平上,应用本发明在此所述的产物和可在汇编这类方法的参考文献中找到的本领域众所周知的方案,PCR 和 Northern 印迹都能用于测定 FGE mRNA 的水平。在蛋白质水平上,应用多克隆或单克隆抗 FGE 血清结合标准免疫分析,FGE 表达可被测定。优选的方法将会把测得的测试样本的 FGE 表达水平与对照进行比较。对照可包括已知量的核酸探针、FGE 表位(如 FGE 表达产物),或来自具有对照或“正常”水平 FGE 表达的受试者的类似测试样本。

[0216] FGE 多肽优选重组性产生,虽然这类蛋白可从生物抽提物中分离出。重组性产生的 FGE 多肽包括含有 FGE 蛋白与另一多肽的融合物的嵌合的蛋白质,所述多肽例如能提供或增强蛋白-蛋白结合、序列专一性核酸结合(如 GAL4),增强 FGE 多肽在分析条件下的稳定性,或提供可探测性的部分例如绿色荧光蛋白。融合到 FGE 多肽或片段的多肽也可提供易于探测融合蛋白的手段,例如通过免疫识别或通过荧光标记。

[0217] 本发明也在产生非人的转基因动物中 useful。在本文中所用的“非人的转基因动物”包括具有整合在生殖系细胞和/或体细胞中的一种或更多的外生性核酸分子的非人的动物。因此转基因动物包括通过同源重组而具有纯合或杂合基因断裂的“敲除的”动物,具有游离或染色体整合的表达载体的动物等。敲除动物能通过应用本领域众所周知的胚胎干细胞同源重组而制得。重组可应用例如本领域一般技术人员已知的 cre/lox 系统或其它重组酶系统而促进。在一定的实施方案中,重组酶系统本身是条件性地表达,例如在一定组织或细胞类型中,在一定胚胎或胚胎后发育阶段,因增加或减弱表达的化合物或类似物质的加入而被诱导。一般地,在这类系统中应用的条件表达载体应用多种可赋予所期望的基因表达模式(例如时间或空间性的)的启动子。条件启动子也能可操纵性地连接到 FGE 核酸分子以受调节或条件性的方式增加 FGE 的表达。FGE 活性或表达的反式-作用负调节子也能可操纵性地连接到上述的条件启动子。这类反式-作用调节子包括反义 FGE 核酸分子,编码显性失活 FGE 分子的核酸分子,对 FGE 核酸专一的核酶分子,和类似物。非人的转基因动物在用于对诊断或治疗方法的生化或生理效应进行测试的实验中有用,所述诊断或治疗方法针对以增加或减少 FGE 表达为特征的病症。其它应用对本领域一般技术人员而言是很显然的。

[0218] 本发明也涉及基因治疗。完成离体基因治疗的程序在 U. S. Patent 5,399,346 中概述,也显示于该专利审查中所提交的文件,其所有内容都是公开可得的文件。一般地,它包括在体外将基因的功能拷贝引入到包含基因缺陷性拷贝的受试者的细胞中,将基因工程化的细胞回置到受试者中。基因的功能拷贝是处于允许基因在基因工程化细胞中表达的调控元件的可操作性控制下。许多转染和转导技术以及适当的表达载体为本领域一般技术人员所熟知,其中的一部分在 PCT 申请 W095/00654 中被描述。应用例如腺病毒,逆转录病毒,疱疹病毒,和靶向的脂质体之类载体的体内基因治疗也根据本发明而涉及。

[0219] 本发明进一步提供鉴别试剂或引导化合物以得到在 FGE 或 FGE 片段依赖性细胞功能的水平上有活性的试剂的有效方法。这类功能尤其包括与其它多肽或片段的相互作用。一般地,筛选方法包括对干扰 FGE 活性(如 C_α-甲酰甘氨酸生成活性)的化合物的分析,虽然增强 FGE C_α-甲酰甘氨酸生成活性的化合物也能应用该筛选方法而被分析。这类方法能被适应于自动化的高产出的化合物筛选。靶指标包括被 FGE 调节的细胞过程如 C_α-甲酰甘氨酸生成活性。

[0220] 多种对备选(药理上的)试剂的分析被提供,包括标记的体外蛋白质-配体结合分析,电泳迁移改变分析,免疫分析,基于细胞的分析例如双杂交或三杂交筛选,表达分析等。转染的核酸能编码例如,组合肽库或 cDNA 库。便于这类分析的试剂例如 GAL4 融合蛋白在本领域中众所周知。可作为范例的基于细胞的分析包括,以编码融合到 GAL4DNA 结合结构域的 FGE 多肽的核酸和编码可操作性地连接到基因表达调节区(如一个或更多 GAL4 结合位点)的报告基因的核酸而转染细胞。报告基因转录的活化发生在 FGE 和报告融合多肽结合以例如激活报告基因的转录时。之后,调节 FGE 多肽介导的细胞功能的试剂通过报告基因表达的变化而被探测。测定报告基因表达变化的方法在本领域中众所周知。

[0221] 方法中所用的 FGE 片段当不是由转染的核酸所产生时,作为分离的多肽加入到分析混合物中。FGE 多肽优选重组产生,虽然此多肽可从生物抽提物中分离。重组产生的 FGE 多肽包括包含 FGE 蛋白与另一多肽的融合体的嵌合蛋白质,另一多肽例如能提供或增强蛋白质-蛋白质结合、序列专一性核酸结合(如 GAL4)、增强 FGE 多肽在分析条件下的稳定性或提供可探测部分例如绿色荧光蛋白或 Flag 表位。

[0222] 分析混合物包含能与 FGE 相互作用的天然的细胞内 FGE 结合靶。当天然 FGE 结合靶可用时,常常优选应用 FGE 结合靶的局部(例如肽-参见例如 SEQ ID NO:33 的肽-或核酸片段)或类似物(即,为分析目的而模拟天然结合靶的 FGE 结合性质的试剂),只要局部或类似物提供分析中可测量的对 FGE 片段的结合亲和性和强烈结合倾向。

[0223] 分析混合物也包含候选物质。典型地,多个分析混合物以不同试剂浓度平行进行,以得到对多种浓度的不同反应。典型地,这些浓度中的某一个用作负对照,即,试剂的零浓度或在分析的探测极限之下的试剂浓度。候选物质包括许多化学种类,虽然它们典型的是有机化合物。优选地,候选物质是小分子有机化合物,即那些具有大于 50 然而小于约 2500 的分子量的类型,优选小于约 1000,而更优选小于约 500。候选物质包含与多肽和/或核酸的结构性相互作用所必需的功能化学基团,并典型地包括至少一个氨基、羰基、羟基或羧基基团,优选至少两个功能化学基团,而更优选至少三个功能化学基团。候选物质可包含被一种或更多上述确定的功能基团所取代的环碳或杂环结构和/或芳香或聚芳香结构。候选物质也可是生物分子例如肽,糖,脂肪酸,固醇,类异戊二烯,嘌呤,嘧啶,上述物质的衍生物或结构类似物,或者其组合,以及类似物质。试剂是核酸的情形中,试剂典型地是 DNA 或 RNA 分子,虽然在此所定义的修饰后的核酸也被涉及。

[0224] 候选物质从多种来源包括合成或天然化合物的库得到。例如,对于多种有机化合物和生物分子的随机和定向合成,有许多方式可用,包括随机寡核苷酸的表达、合成性的有机组合库、随机肽段的噬菌体展示库和类似方式。替代性地,细菌、真菌、植物和动物抽提物形式的天然化合物的库唾手可得或易于产生。此外,天然和合成性产生的库和化合物能通过传统化学、物理和生化方式被修饰。进一步地,已知的(药理学上)试剂可进行定向或随

机化学修饰例如酰化,烷基化,酯化,amidification,等,以产生试剂的结构类似物。

[0225] 多种其它试剂也能被包括在混合物中。这些包括试剂例如盐,缓冲剂,中性蛋白质(例如白蛋白),去污剂,等,其可被用于促进最佳的蛋白质-蛋白质和/或蛋白质-核酸结合。这样的试剂也可减少反应成分的非专一性的或背景性的相互作用。其它改善分析效率的试剂如蛋白酶,抑制剂,核酸酶抑制剂,抗微生物剂和类似物也可被应用。

[0226] 前述分析材料的混合物在由此确定的条件下培育,但对于候选物质存在的情形中,FGE 多肽特异性地与细胞结合靶、其部分或其类似物结合。组分加入的顺序,培育温度,培育时间和分析的其它参数可容易地确定。这类实验只包括分析参数的最优化,而不涉及分析的基本组成。培育温度典型地是在 4°C 和 40°C 之间。培育时间优选最小化以促进快速、高产出的筛选,而典型的是在 0.1 和 10 小时之间。

[0227] 培育之后,FGE 多肽与一种或更多结合靶的专一性结合的存在与否通过使用者可用的任何方便的方法而探测。对于非细胞结合类型分析,分离步骤常被用于从未结合的组分中分离结合的组分。分离步骤可以多种方式完成。方便的是,组分中的至少一种在固相基质上被固定,未结合组分可容易地从其中分离。固相基质可由多种材料制成多种形式,例如微量滴定平板,微珠,计量棒(dipstick),树脂颗粒,等。优选基质被选择成最大的信噪比,主要是最小化背景结合,以及使分离程序 and 成本简化。

[0228] 分离可被多种因素影响,例如从库中除去珠子或计量棒,清空或稀释库如微量滴定平板孔,以清洗溶液或溶剂清洗珠子,颗粒,层析柱或过滤器。分离步骤优选包括多次漂洗或清洗。例如,当固相基质是微量滴定平板时,孔可典型地被包括培育混合物的那些不参与特异性结合的组分的清洗溶液清洗数次,所述组分如盐,缓冲剂,去污剂,非专一性蛋白质等。在固相基质是磁珠的情形中,珠子可用清洗溶液清洗一次或更多,并用磁体而分离出。

[0229] 探测可以任何方便的方式被影响,以进行基于细胞的分析例如双杂交或三杂交筛选。与靶分子相互作用的 FGE 多肽的报告基因转录分析所得到的转录物典型地编码直接或间接可测的产物,例如 β -半乳糖苷酶活性,荧光素酶活性,和类似物。对于非细胞结合分析,组分的一种通常包含或被偶连到可探测的标记。多种标记都能用,例如那些提供直接探测的类型(例如放射性,发光,视觉或电子密度等),或提供间接探测的类型(例如,表位标签如 FLAG 表位、酶标签如辣根过氧化物酶等)。标记可被固着于 FGE 结合伴侣,或整合进结合伴侣的结构中。

[0230] 多种方法可被用于探测标记,这取决于标记的天性和其它的分析组分。例如,当固着到固相基质时或从固相基质分离之后,标记可被探测。标记可通过视觉或电子密度,放射性辐射,非放射性能量转移等直接探测,或以抗体缀合物、链霉抗生物素-生物素缀合物等间接探测。探测标记的方法在本领域中众所周知。

[0231] 本发明提供 FGE 专一结合试剂,鉴别和生产这类试剂的方法,及它们在诊断、治疗和药物开发中的应用。例如,FGE 专一性药理试剂在多种诊断和治疗性应用中有用,尤其是在疾病或疾病预后与改变的 FGE 结合特征相关的情形例如多种硫酸酯酶缺乏症中。新的 FGE 专一结合试剂包括 FGE 专一抗体,细胞表面受体,被例如双杂交筛选的分析法所鉴别的其它天然细胞内和细胞外结合试剂,及被对化学库的筛选所鉴别的非天然细胞内和细胞外结合试剂,和类似物。

[0232] 一般而言, FGE 结合到特异分子的专一性通过结合平衡常数被测定。能选择性结合 FGE 多肽的靶优选具有至少约 10^7M^{-1} 的结合常数, 更优选至少约 10^8M^{-1} , 而最优选至少约 10^9M^{-1} 。多种基于细胞的和非细胞的分析可被用于证明 FGE 专一结合。基于细胞的分析包括单、双和三杂交筛选, 其中 FGE 介导的转录得到抑制或增加等的分析。非细胞分析包括 FGE 蛋白结合分析、免疫分析等。其它对筛选结合 FGE 多肽的试剂有用的分析包括荧光共振能量转移 (FRET) 和电泳迁移改变分析 (EMSA)。

[0233] 根据本方面的另一方面, 鉴别在调节本发明的分子的 C_α -甲酰甘氨酸生成活性中有用的试剂的方法被提供。方法包括 (a) 将具有 C_α -甲酰甘氨酸生成活性的分子与候选物质相接触, (b) 测量该分子的 C_α -甲酰甘氨酸生成活性, 和 (c) 将测得的分子的 C_α -甲酰甘氨酸生成活性与对照比较以确定是否候选物质调节分子的 C_α -甲酰甘氨酸生成活性, 其中分子是本发明的 FGE 核酸分子或其表达产物。“接触”指具有 C_α -甲酰甘氨酸生成活性的分子与候选物质的直接和间接的接触。“间接”接触表示候选物质通过第三者试剂 (例如信使分子, 受体等) 对分子的 C_α -甲酰甘氨酸生成活性施加了其影响。在一定实施方案中, 对照是在缺乏候选物质时测得的分子 C_α -甲酰甘氨酸生成活性。分析方法和候选物质如前述有关 FGE 的实施方案中所述。

[0234] 根据本发明的另一方面, 诊断以核酸分子、其表达产物或其表达产物片段的异常表达为特征的疾病的方法被提供。方法包括将从受试者中分离的生物样本与专一结合到核酸分子、其表达产物或其表达产物片段的试剂接触, 并测定作为疾病判定依据的试剂和核酸分子或表达产物之间的相互作用, 其中核酸分子是根据本发明的 FGE 分子。疾病是多种硫酸酯酶缺乏症。导致 FGE 分子异常表达的 FGE 基因中的突变引起下面的 SEQ ID NO :2 上的氨基酸改变: Met1Arg; Met1Val; Leu20Phe; Ser155Pro; Ala177Pro; Cys218Tyr; Arg224Trp; Asn259Ile; Pro266Leu; Ala 279Val; Arg327Stop; Cys336Arg; Arg345Cys; Ala348Pro; Arg349Gln; Arg349Trp; Arg349Trp; Ser359Stop; 或其组合。

[0235] 在分子是核酸分子的情形中, 这类测定能通过任何标准的核酸测定分析进行, 包括聚合酶链式反应或在本文中作为例子的以标记性杂交探针进行的分析。在分子是核酸分子表达产物或核酸分子表达产物片段的情形中, 这类测定能通过应用例如结合到任何多肽表达产物的抗体的任何标准的免疫分析而进行。

[0236] “异常表达”指 FGE 分子 (核酸和 / 或多肽) 相对于对照 (即, 相同分子在健康或“正常”受试者中的表达) 的减少的表达 (表达不足) 或增加的表达 (过表达)。在本文中所述的“健康受试者”, 指根据标准的医学标准没有出现多种硫酸酯酶缺乏症或具有发展多种硫酸酯酶缺乏症的风险的受试者。健康受试者也没有另外显现出病症。换句话说, 如果被医学专业人员检查, 这类受试者将被表征为健康和不带多种硫酸酯酶缺乏症的症状。这些包括异常染色性脑白质营养不良和粘多糖病的特征, 例如在几种组织中的酸性粘多糖的量增加, 轻微‘脂肪软骨营养不良’, 快速神经性衰退, 尿中的脑硫脂和粘多糖的过量存在, 增加的脑脊液蛋白和外周神经中的髓磷脂的异染性退化。

[0237] 本发明也提供新的试剂盒, 其将被用于测量本发明的核酸和本发明的表达产物的水平。

[0238] 在某一实施方案中, 试剂盒包含包装, 该包装包含下列物质: 选择性结合任何前述 FGE 的分离的核酸或其表达产物的试剂, 及用于与前述试剂和任何前述 FGE 的分离的核酸

或其表达产物结合的测得值进行比较的对照。在某些实施方案中,对照是用于与测得值比较的预先测定值。在一定实施方案中,对照包含任何前述 FGE 的分离的核酸的表达产物的表位。在某一实施方案中,试剂盒进一步包含选择性结合选自下列的多肽的第二试剂:艾杜糖醛酸-2-硫酸酯酶,硫酸胺酶,N-乙酰半乳糖胺-6-硫酸酯酶,N-乙酰葡萄糖胺-6-硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 HSulf-6,或其肽段,和用于与前述第二试剂和所述多肽或其肽段的结合的测得值进行比较的对照。

[0239] 在核酸探测的情形中,用于扩增本发明的核酸分子的引物对可被包括进去。优选的试剂盒将包括对照,例如已知量的核酸探针、表位(例如艾杜糖醛酸-2-硫酸酯酶,硫酸胺酶,N-乙酰半乳糖胺-6-硫酸酯酶,N-乙酰葡萄糖胺-6-硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 HSulf-6,表达产物)或抗表位的抗体,以及指南或其它复印材料。在一定的实施方案中,复印材料能基于分析结果而表征发展硫酸酯酶缺乏状态的风险。试剂可按预先测定的量被包在容器中和/或被涂在小孔上,而试剂盒可包括标准物质,例如标记好的免疫试剂(如标记好的抗-IgG 抗体)和类似物。一种试剂盒是包好的用 FGE 蛋白涂层的聚苯乙烯微量滴定平板和含有标记好的抗人 IgG 抗体的容器。平板的小孔与例如生物液体接触,清洗,然后与抗-IgG 抗体接触。之后探测标记。体现本发明特征的试剂盒,通常以数字 11 指定,在图 25 中显示。试剂盒 11 包含下列主要元件:包装 15,本发明的试剂 17,对照试剂 19 和指南 21。包装 15 是用于盛含有本发明的试剂 17 的一个小瓶(或多个小瓶),含有对照试剂 19 的一个小瓶(或多个小瓶)和指南 21 的类似盒子的结构。本领域技术人员可易于修饰包装 15 以适合个体的需要。

[0240] 本发明也包含在受试者中治疗多种硫酸酯酶缺乏症的方法。方法包括对有这类治疗需要的受试者施用调节 C_α-甲酰甘氨酸生成活性的试剂,用量为在受试者中有效增加 C_α-甲酰甘氨酸生成活性的量。在某些实施方案中,方法进一步包含一定试剂的共施用,所述试剂选自编码下列蛋白的核酸分子:艾杜糖醛酸-2-硫酸酯酶,硫酸胺酶,N-乙酰半乳糖胺-6-硫酸酯酶,N-乙酰葡萄糖胺-6-硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 HSulf-6;核酸分子的表达产物,和/或核酸分子的表达产物的片段。

[0241] 在本文中所用的“调节核酸或多肽的表达的试剂”在本领域中已知,指有义和反义核酸,显性失活核酸,多肽的抗体和类似物。调节分子表达(和如本文中所述的,调节其活性)的任何试剂根据本发明有用。在一定实施方案中,调节 C_α-甲酰甘氨酸生成活性的试剂是本发明的分离的核酸分子(例如 SEQ ID NO. 3 的核酸)。在重要的实施方案中,调节 C_α-甲酰甘氨酸生成活性的试剂是本发明的肽(例如 SEQ ID NO. 2 的肽)。在某些实施方案中,调节 C_α-甲酰甘氨酸生成活性的试剂是本发明的有义核酸。

[0242] 根据本发明的一个方面,在受试者中增加 C_α-甲酰甘氨酸生成活性的方法被提供。方法包括对受试者施用本发明的分离的核酸分子和/或其表达产物,用量为在受试者中增加 C_α-甲酰甘氨酸生成活性的有效量。

[0243] 根据本发明的另一方面,增加细胞中的 C_α-甲酰甘氨酸生成活性的方法被提供。方法包括将细胞与本发明的分离的核酸分子(例如 SEQID NO. 1 的核酸),或其表达产物(例如 SEQ ID NO. 2 的肽)接触,用量为在细胞中有效增加 C_α-甲酰甘氨酸生成活性的量。在重要的实施方案中,方法包括激活内源性的 FGE 基因以增加细胞中的 C_α-甲酰甘氨酸生成活性。

[0244] 在任何的前述实施方案中,核酸可被操作性地偶联到指导真核细胞例如 HT-1080 细胞内的核酸分子表达的基因表达序列。“基因表达序列”是任何调节性的核苷酸序列,例如启动子序列或启动子-增强子组合,其促进它所可操作性地连接的核酸的有效转录和翻译。基因表达序列可以是,例如,哺乳动物或病毒的启动子如组成型或可诱导型的启动子。组成型的哺乳动物启动子包括但不限于下列基因的启动子:次黄嘌呤磷酸核糖基转移酶 (HPTR),腺苷脱氨酶,丙酮酸激酶,α-肌动蛋白启动子和其它组成型启动子。可作为范例的在真核细胞中组成性地发挥功能的病毒启动子包括,例如,来自猿猴病毒,乳头状瘤病毒,腺病毒,人免疫缺陷病毒 (HIV),劳斯肉瘤病毒,细胞巨化病毒,莫洛尼氏白血病毒的长末端重复序列 (LTR) 和其它逆转录病毒的启动子,和单纯疱疹病毒的胸苷激酶启动子。其它组成型启动子为本领域一般技术人员所知。作为本发明的基因表达序列而有用的启动子也包括可诱导的启动子。可诱导的启动子在诱导剂存在时被活化。例如,金属硫蛋白启动子在一定的金属离子存在时被激活而增加转录和翻译。其它可诱导的启动子为本领域一般技术人员所知。

[0245] 一般而言,基因表达序列将包括(如果必要)分别涉及转录和翻译起始的 5' 非转录和 5' 非翻译序列,例如 TATA 盒,加帽序列,CAAT 序列和类似物。尤其是,这类 5' 非转录序列将包括启动子区域,所述启动子区域包括用于对可操作性地连接的核酸进行转录控制的启动子序列。基因表达序列可选性地包括所期望的增强子序列或上游激活子序列。

[0246] 优选任何本发明的 FGE 核酸分子连接到允许核酸分子在特定细胞系的细胞如神经细胞中表达的基因表达序列。允许核酸分子在细胞如神经细胞中表达的序列,是在这类细胞类型中选择性地有活性从而引起核酸分子在这些细胞中表达的类型。例如,突触蛋白 -1 启动子能被用于在神经细胞中表达任何前述的本发明的核酸分子;而例如 vonWillebrand 因子基因启动子能被用于在血管内皮细胞中表达核酸分子。本领域中一般技术人员将能易于鉴别能在本发明的任何优选细胞中表达核酸分子的替代性的启动子。

[0247] 当它们以将核酸编码序列(例如,在 FGE 的情形中,SEQ ID NO. 3)的转录和/或翻译置于基因表达序列的影响或控制下这样的方式而被共价连接时,核酸序列和基因表达序列被表述为“可操作性地连接”。如果期望核酸序列被翻译成为功能性蛋白,并且如果在 5' 基因表达序列中的启动子的诱导导致核酸序列的转录,且如果在两个 DNA 序列之间的连接的本性不会(1) 导致移码突变的引入,(2) 干扰启动子区域指导核酸序列转录的能力,和/或(3) 干扰对应的 RNA 转录产物翻译成为蛋白质的能力时,两个 DNA 序列被表述成可操作性地连接。因此,如果基因表达序列能影响核酸序列的转录以至于得到的转录产物可被翻译成为所期望的蛋白质或多肽,则基因表达序列将被认为可操作性地连接到核酸序列。

[0248] 本发明的分子能单独的或与载体(也见对于载体的更早的讨论)一起被运送到本发明优选细胞类型。按其最广泛的含义(且与本文中其它部分对表达和靶向载体的描述一致),“载体”是能促进:(1) 分子到靶细胞的运送,和/或(2) 分子被靶细胞的吸取,

的任何载体。优选运送载体以有所减少的降解运输分子到靶细胞中,所述有所减少的降解相对于载体缺乏时所产生的降解的程度而言。可选地,“靶向配体”能被附着于载体以选择性地将载体运送到其表面表达靶向配体的相关受体的细胞。以这种方式,载体(含有核酸或蛋白质)能被选择性地运送到神经细胞。靶向的方法包括缀合,例如那些在 Priest 的 U. S. Patent 5, 391, 723 中所述的。另一众所周知的靶向载体的实例是脂质体。脂质体从 Gibco BRL 可商业性地获得。生产靶向脂质体的多种方法被发表。

[0249] 一般而言,在本发明中有用的载体包括但不限于质粒,噬菌粒,病毒,源于病毒或细菌来源的其它载体,所述其它载体已通过本发明核酸序列和能被附着到本发明的核酸序列的另外的核酸片段(例如增强子、启动子)的插入或整合而被操作。病毒载体是载体的优选类型,包括但不限于来自下列病毒的核酸序列:腺病毒;腺伴随病毒;逆转录病毒,例如鼠莫洛尼白血病毒;鼠哈维肉瘤病毒;鼠乳腺肿瘤病毒;劳斯肉瘤病毒;SV40-型病毒;多瘤病毒;EB病毒;乳头状瘤病毒;疱疹病毒;牛痘病毒;脊髓灰质炎病毒;和 RNA 病毒例如逆转录病毒。也能容易地采用没有指出但在本领域中已知的其它载体。

[0250] 对某些应用特别优选的病毒是腺伴随病毒,一种双链 DNA 病毒。腺伴随病毒能感染广泛的细胞类型和物种,且能被工程化为复制缺陷型。它进一步地具有优点例如热和脂溶剂稳定性、在多种株系细胞包括造血细胞中的高转导频率,和缺乏超感染抑制因此而允许多重系列的转导。经报道,腺伴随病毒能以位点特异的方式整合进入细胞 DNA,从而将插入性基因突变的可能性和插入基因的表达的可变性最小化。此外,野生型腺伴随病毒感染已在组织培养物中于缺乏选择压力的条件下存在超过 100 代,意味着腺伴随病毒基因组整合是相对稳定的事件。腺伴随病毒也能以染色体外方式发挥功能。

[0251] 一般而言,其它优选的病毒载体基于非细胞病变性真核病毒,在这些病毒中非必需基因已被目的基因所替代。非细胞病变性病毒包括逆转录病毒,其生命周期包括基因组病毒 RNA 到 DNA 的逆转录及随后的前病毒整合到宿主细胞 DNA 中。腺病毒和逆转录病毒已被许可用于人基因治疗的试验。一般而言,逆转录病毒是复制缺陷型(即,能指导所需蛋白的合成,但不能制造出感染性的颗粒)。这类基因改变的逆转录病毒表达载体具有对基因在体内的高效转导的一般性用途。产生复制缺陷型逆转录病毒的标准流程(包括步骤:外源性遗传物质整合进质粒,质粒对包装细胞系的转染,重组性逆转录病毒通过包装细胞系的产生,从组织培养基中收集病毒颗粒和病毒颗粒对靶细胞的感染)在 Kriegler, M., “Gene Transfer and Expression, A Laboratory Manual,” W. H. Freeman C. O., New York (1990) 和 Murry, E. J. Ed. “Methods in Molecular Biology,” 卷 7, Humana Press Inc., Clifton, New Jersey (1991) 中被提供。

[0252] 另一优选的逆转录病毒载体是源自鼠莫洛尼白血病毒的载体,如在 Nabel, E. G. 等, Science, 1990, 249:1285-1288 中所述的。这些载体据报道对基因运送到动脉壁的所有三层(包括中间层)都有效。其它优选载体在 Flugelman 等, Circulation, 1992, 85:1110-1117 中公开。另外的对运送本发明的分子有用的载体在 U. S. Patent No. 5, 674, 722 中被 Mulligan 等描述。

[0253] 除前述载体外,其它传送方法可被用于传送本发明的分子到细胞例如神经细胞,肝,成纤维细胞,和/或血管内皮细胞,并促进在那里的吸收。

[0254] 本发明的优选的这类运送方法是胶态分散体系统。胶态分散体系统包括基于脂

质的系统,包括水包油乳液,微囊,混合微囊,和脂质体。本发明优选的胶态系统是脂质体。脂质体是人造膜容器,其作为体内或体外的运送载体而有用。已有显示,大小范围在 0.2-4.0 μm 的单层容器 (LUV) 能包裹进大的大分子。RNA, DNA 和完整病毒粒子能被封装在含水内部,并以生物活性形式运送到细胞中 (Fraley 等, Trends Biochem. Sci., 1981, 6: 77)。为了脂质体是有效的基因转运营载体,一种或更多的下列特征应具备:(1) 目的基因高效被封装,而保持生活活性;(2) 相对于非靶细胞的与靶细胞的优选和牢固的结合;(3) 囊泡的水内容物到靶细胞细胞质的高效运送;和(4) 遗传信息的准确和有效表达。

[0255] 脂质体可通过脂质体到特异配体例如单克隆抗体,糖,糖脂,或蛋白质的偶联,而靶向于特定组织,例如心肌或血管细胞壁。可对靶向脂质体于血管壁有用的配体包括但不限于,仙台病毒的病毒衣壳蛋白。此外,载体可偶联到核靶向性肽上,其将定向核酸到宿主细胞核。

[0256] 脂质体可从 Gibco BRL 商业性地获得,例如由阳离子脂质如 N-[1-(2,3 双油氧基)-丙基]-N, N, N-三甲基氯化铵 (DOTMA) 和二甲基双十八烷基溴化铵 (DDAB) 形成的 LIPOFECTIN™ 和 LIPOFECTACE™。生产脂质体的方法在本领域中众所周知,也在许多出版物中被描述。脂质体也已被 Gregoriadis, G. 在 Trends in Biotechnology, 卷 3, 235-241 页 (1985) 中所回顾。用于大分子包括核酸的细胞内运送的新的脂质体也在 PCT 国际申请 PCT/US96/07572 (公开号 W0 96/40060, 名称为“*Intracellular Delivery of Macromolecules*”) 中被描述。

[0257] 在一个特定的实施方案中,优选的载体是适合植入哺乳动物受体的生物相容性的微颗粒或植入物。作为范例的根据此方法而有用的生物侵蚀性植入物在 PCT 国际申请 PCT/US/03307 (公开号 W0 95/24929, 名称为“*Polymeric Gene Delivery System*”, 其要求 1994 年 3 月 15 日提交的美国专利申请系列号 213,668 的优先权) 中被描述。PCT/US/0307 描述了用于包含(受适当的启动子调控的)外源性基因的生物相容性的聚合性基质,优选生物降解性的聚合性基质。聚合性基质被用于实现外源性基因在患者中的持续性释放。根据本发明,在此所述的核酸在 PCT/US/03307 中所公开的生物相容性优选生物降解性的聚合性基质中被封装或分散。聚合性基质优选微颗粒形式例如微球体(其中核酸被分散到整个固体聚合性基质中)或微囊(其中核酸被储存在聚合性壳的核心中)。包含本发明的核酸的聚合性基质的其它形式包括膜,涂层,胶,植入物和支架。聚合性基质装置的大小和组成被选作引起基质装置被植入的组织中的良好的释放动力学。进一步设计的聚合性基质的大小根据将被应用的运送方法而选择,所述运送方法典型地是组织注射或悬浮液通过雾化施用到鼻部和/或肺部区域。聚合性基质组合物能被选成同时具有良好的降解速率,及由生物粘附性物质形成,以在设计物被施用到血管表面时进一步增加转运的有效性。基质组合物也能被选成不降解,但更适合通过持续相当一段时间的扩散而释放。

[0258] 非生物降解性和生物降解性的聚合性基质都能被用作将本发明的核酸运送到受试者。优选生物降解性基质。这类聚合物可以是自然或合成性的聚合物。优选合成性聚合物。根据释放所需要的时间长度而选择聚合物,一般是几个小时到一年或更长时间的级别。典型地,范围从几个小时到三至十二个月的时间长度的释放最合意。聚合物可选性地是能吸收高到其重量的约 90% 的水的水凝胶的形式,而进一步地,可选性地与多价离子或其它聚合物交联。

[0259] 一般而言,本发明的核酸应用生物侵蚀性植入物以扩散方式而被运送,或更优选地,通过聚合性基质的降解。作为范例的能被用于形成生物降解性运送系统的合成性聚合物包括:聚酰胺,聚碳酸酯,聚亚烷基,聚亚烷基二醇,聚环氧烷,聚亚烷基对苯二酸酯,聚乙烯醇,聚乙烯醚,聚乙烯酯,聚-乙烯卤化物,聚乙烯吡咯烷酮,聚乙交酯,聚硅氧烷,聚氨基甲酸乙酯和其共聚物,烷基纤维素,羟基烷基纤维素,纤维素醚,纤维素酯,硝基纤维素,丙烯酸酯和甲基丙烯酸酯的聚合物,甲基纤维素,乙基纤维素,羟丙基纤维素,羟丙基甲基纤维素,羟丁基甲基纤维素,醋酸纤维素,丙酸纤维素,纤维素醋酸丁酸酯,纤维素醋酸邻苯二甲酸酯,羧乙基纤维素,三醋酸纤维素,纤维素硫酸钠盐,聚(甲基异丁烯酸酯),聚(乙基异丁烯酸酯),聚(丁基异丁烯酸酯),聚(异丁基异丁烯酸酯),聚(己基异丁烯酸酯),聚(异癸基异丁烯酸酯),聚(十二烷基异丁烯酸酯),聚(苯基异丁烯酸酯),聚(甲基丙烯酸酯),聚(异丙基丙烯酸酯),聚(异丁基丙烯酸酯),聚(十八烷基丙烯酸酯),聚乙烯,聚丙烯,聚(乙二醇),聚(环氧乙烷),聚(对苯二甲酸乙二酯),聚(乙烯醇),聚乙烯醋酸,聚氯乙烯,聚苯乙烯和聚乙烯吡咯烷酮。

[0260] 非生物降解性聚合物的实例包括亚乙基乙烯醋酸,聚(甲)丙烯酸,聚酰胺,其共聚物和混合物。

[0261] 生物降解性聚合物的实例包括合成性聚合物例如乳酸和乙二酸的聚合物,聚酸酐,聚(原)酸酯,聚氨基甲酸乙酯,聚(丁酸),聚(戊酸),聚(丙交酯-共己内酯),天然聚合物例如藻酸和其它多糖包括葡聚糖和纤维素,胶原,其化学衍生物(化学基团例如,烷基,亚烷基,羟化,氧化的取代和增加,和其它由本领域技术人员常规进行的修饰),白蛋白和其它亲水蛋白质,玉米醇溶蛋白和其它醇溶谷蛋白及疏水蛋白质,其共聚物和混合物。一般而言,这些物质或者通过酶水解或者在体内暴露于水时通过表面或整体侵蚀而降解。

[0262] 特别受关注的生物粘附性聚合物包括 H. S. Sawhney, C. P. Pathak 和 J. A. Hubell 在 *Macromolecules*, 1993, 26, 581-587 中所述的生物侵蚀性水凝胶,其教导在此被整合,聚透明质酸,酪蛋白,明胶,谷胶酪蛋白,聚酸酐,聚丙烯酸,藻酸,脱乙酰壳多糖,聚(甲基异丁烯酸酯),聚(乙基异丁烯酸酯),聚(丁基异丁烯酸酯),聚(异丁基异丁烯酸酯),聚(己基异丁烯酸酯),聚(异癸基异丁烯酸酯),聚(十二烷基异丁烯酸酯),聚(苯基异丁烯酸酯),聚(甲基丙烯酸酯),聚(异丙基丙烯酸酯),聚(异丁基丙烯酸酯),和聚(十八烷基丙烯酸酯)。因此,本发明提供用作药物的本发明的上述分子的组合物,制备此药物的方法和在体内用于持续释放药物的方法。

[0263] 压缩试剂也能与本发明的载体结合应用。在本文中所用的“压缩试剂”指试剂例如组蛋白,它中和核酸上的负电荷从而允许核酸压缩成细颗粒。核酸的压缩促进了核酸被靶细胞的吸收。压缩试剂能单独使用,即以被细胞更有效地吸收的形式运送分离的本发明的核酸或,更优选地,与一种或更多上述载体组合。

[0264] 其它作为范例的能被用于促进靶细胞对本发明的核酸的吸收的组合物包括磷酸钙及细胞内运输、微注射组合物和电穿孔的其它化学调节物。

[0265] 本发明包含在细胞中增加硫酸酯酶活性的方法。这类方法包括以有效地在细胞中增加硫酸酯酶活性的量将分离的本发明的核酸分子(例如如权利要求 1-8 中任一项所要求的分离的核酸分子,具有选自 SEQ ID NO:1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的 FGE 核酸分子)或其表达产物(例如如权利要求

11-15, 19, 20 中所要求的多肽或具有选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的序列的肽) 与表达硫酸酯酶的细胞接触。在本文中所用的“增加”硫酸酯酶活性, 指增加的对硫酸酯酶的特异性底物的亲和性和 / 或对其的转化, 典型地导致硫酸酯酶分子上的 FGly 的形成的增加。在某一实施方案中, 细胞以比野生型细胞更高的水平表达硫酸酯酶。“在细胞中增加硫酸酯酶活性”也指增加细胞所分泌的硫酸酯酶的活性。细胞可表达内源性和 / 或外源性的硫酸酯酶。与 FGE 分子的所述接触也指激活细胞的内源性 FGE 基因。在重要的实施方案中, 内源性硫酸酯酶被活化。在一定实施方案中, 硫酸酯酶是艾杜糖醛酸-2-硫酸酯酶, 硫酸胺酶, N-乙酰半乳糖胺-6-硫酸酯酶, N-乙酰葡萄糖胺-6-硫酸酯酶, 芳基硫酸酯酶 A, 芳基硫酸酯酶 B, 芳基硫酸酯酶 C, 芳基硫酸酯酶 D, 芳基硫酸酯酶 E, 芳基硫酸酯酶 F, 芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 / 或 HSulf-6。在一定的实施方案中, 细胞是哺乳动物细胞。

[0266] 根据本方面的另一方面, 药物组合物被提供。组合物以药学上有效治疗硫酸酯酶缺乏症的量包含细胞产生的硫酸酯酶, 也包含药学上可接受的载体, 其中所述细胞已与包含分离的本发明的核酸分子(例如如权利要求 1-8 所要求的分离的核酸分子或具有选自 SEQ ID NO :1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 和 80-87 的序列的核酸分子)或其表达产物(例如选自 SEQ ID NO. 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 和 78 的肽)的试剂接触。在重要的实施方案中, 硫酸酯酶以高于正常 / 对照细胞的水平表达。

[0267] 本发明也包含产生硫酸酯酶的细胞, 其中细胞所产生的活性硫酸酯酶对总硫酸酯酶的比率增加。细胞包含:(i) 相对于对照具有增加的活性的硫酸酯酶, 和 (ii) 相对于对照具有增加的活性的甲酰甘氨酸生成酶, 其中相对于甲酰甘氨酸生成酶缺乏的细胞所产生的活性硫酸酯酶对总硫酸酯酶的比率, 细胞所产生的活性硫酸酯酶对总硫酸酯酶的比率增加至少 5%。在本领域中已知, 硫酸酯酶的过表达能减少内源性硫酸酯酶的活性 (Anson 等, *Biochem. J.*, 1993, 294 :657-662)。此外, 仅有重组硫酸酯酶的一部分有活性。出乎意料地, 我们发现在具有硫酸酯酶的增加的表达 / 活性的细胞中, FGE 增加的表达 / 活性导致更有活性的硫酸酯酶的产生。既然 FGly 在硫酸酯酶分子上的存在与硫酸酯酶活性相关, 则“有活性的硫酸酯酶”就能通过应用 MALDI-TOF 质谱(如本文中其它部分所描述的)对 FGly 在硫酸酯酶细胞产物上的存在的测定而定量。然后对总硫酸酯酶的比率能很容易测定。

[0268] 本发明也提供用于硫酸酯酶缺乏症的诊断和治疗的方法。这类疾病包括但不限于, 多种硫酸酯酶缺乏症, 粘多糖病 II (MPS II ;Hunter 综合症), 粘多糖病 IIIA (MPS IIIA ; Sanfilippo 综合症 A), 粘多糖病 VIII (MPS VIII), 粘多糖病 IVA (MPS IVA ;Morquio 综合症 A), 粘多糖病 VI (MPS VI ;Maroteaux-Lamy 综合症), 异常染性脑白质营养不良 (MLD), X-连锁的隐性点状软骨发育不全 1, 和 X-连锁的鱼鳞病(类固醇硫酸酯酶缺乏症)。

[0269] 本发明的方法在对任何前述病症的急性或预防性治疗中都有用。在此所用的急性治疗指对具有特定病症的受试者的治疗。预防性治疗指对可能具有此病症但现在并没有或并未经历此病症的症状的受试者的治疗。

[0270] 就其最广泛的意义而言, 术语“治疗”表示急性和预防性治疗两者。如果需要治疗的受试者正经历病症(或具有或正具有特定病症), 那么治疗病症指改善、减弱或消除病症或来自病症的一种或更多的症状。在某些优选的实施方案中, 治疗此病症指改善、减弱或消

除与病症相关的特异的症状或特异亚型的症状。如果需要治疗的受试者是可能具有此病症的受试者,那么对受试者治疗指减小受试者产生此病症的风险。

[0271] 本发明的治疗剂的施用模式和剂量将随着被治疗的病症的特定阶段、被治疗的受试者的年龄和生理状态、同时进行治疗(如果有)的本性、施用的特定途径和在医学实践者的知识和专业技术范围内的类似因素而变化。

[0272] 如本文中所述的,本发明的试剂以治疗任何前述硫酸酯酶缺乏症的有效量而被施用。一般而言,有效量是能引起受试者的所期望的组织中的有益变化的任何量。有效量优选在特定病症中足以引起良好表型变化例如症状或病症的整体性减轻、缓和或消除的量。

[0273] 一般而言,有效量是药制剂单独或与进一步的药剂一起产生出所期望的反应那样的量。这可包括只短暂减慢病症的进程,虽然更优选它包括长期性地阻止病症进程,或延缓病症的发作,或防止病症发生。这可通过常规方法监测。一般地,活性化合物的剂量将从约 0.01mg/kg 每天到 1000mg/kg 每天。期望范围在 50 μ g-500mg/kg 的剂量将是适合,优选口服和每天施用一次或几次。

[0274] 当然,这类量将取决于被治疗的特定病症,病症的严重程度,患者个体的参数包括年龄、生理状况、身材尺寸和体重,治疗的持续时间,同时进行治疗(如果有)的本性,施用的特异途径和在医学实践者的知识和专业技术范围内的类似因素。低剂量将源自施用的一定形式,例如静脉内施用。在应用初始剂量时受试者中的反应不充分的情形中,更高的剂量(或通过不同的更局部化的运送途径的有效更高剂量)可应用到患者耐受力所允许的程度。每天多重剂量被预期能实现化合物的合适的系统水平。一般地,优选应用最大剂量,也就是,根据合理的医学判断的最高安全剂量。然而将被本领域中一般技术人员所理解的是,患者可出于医学原因、心理原因或事实上的其它任何原因而坚持较低剂量或可耐受性剂量。

[0275] 本发明的试剂可选性地与药学上可接受的载体组合以形成药制剂。在本文中所述的术语“药学上可接受的载体”表示适合对人施用的一种或更多相容性的固体或液体填充物、稀释剂或封装物质。术语“载体”表示天然或合成性的有机或无机成分,活性成分与其组合以促进应用。药学组合物的成分也能被与本发明的分子共混合,及彼此混合,其方式是确保没有实质性破坏所期望的药学效力的相互作用存在。在某些方面,药制剂以有效治疗疾病的量包含本发明的试剂。

[0276] 药制剂可包含合适的缓冲剂,包括:盐形式的醋酸、盐形式的柠檬酸、盐形式的硼酸或盐形式的磷酸。药学组合物可选性地,也可包含合适的防腐剂例如苯扎氯;氯代丁醇;对羟基苯甲酸酯或乙基汞硫代水杨酸钠(thimerosal)。

[0277] 多种施药途径可用。当然,选出的特定模式将取决于选定的特定药物,被治疗的病症的严重程度和治疗效力所需的剂量。本发明的方法,一般而言,可应用医学上可接受的任何施用模式,即产生活性化合物的有效水平而不引起临床上无法接受的不利效应的有意义的任何模式而实行。施用的这类模式包括口服,直肠,局部,鼻,皮内,经皮,或非肠道途径。术语“非肠道的”包括皮下,静脉内, intraoental, 肌肉内,或输注。静脉内或肌肉内途径不是特别适合长期性治疗和预防。作为实例,用于对具有偏头痛的受试者的急性治疗的药学组合物可配制成多种不同的方式和多种施用模式,包括片剂,胶囊,粉末,栓剂,注射液和鼻喷雾剂。

[0278] 药制剂可方便地以单位剂型存在,也可通过任何制药领域中众所周知的方法制备。所有方法包括将活性试剂与载体相连的步骤,所述载体组成了一种或更多附属成分。一般而言,组合物通过统一地和密切地将活性化合物与液体载体、很好地分隔的固体载体或同时二者相连而制备,然后如果需要,将产物定形。

[0279] 适合口服施用的组合物可作为离散性单位存在,例如胶囊,片剂,锭剂,每一个都包含预定量的活性化合物。其它组合物包括水性或非水性的悬浮液例如糖浆,酏剂或乳剂。

[0280] 适合非肠道施用的组合物方便地包含消毒过的本发明试剂的水制剂,其优选与接受者的血液等渗。此水制剂可根据已知方法应用合适的分散或润湿剂和悬浮试剂进行配制。无菌的可注射性制剂也可是非毒性的注射用可接受的稀释剂或溶剂中的无菌的注射溶液或悬浮液,例如在 1,3-丁二醇中的溶液。可接受的可用载体和溶剂有水、Ringer's 溶液和等渗氯化钠溶液。此外,无菌的固定油传统上被用作溶剂或悬浮介质。为了这一目的,任何温和的固定油可被应用,包括合成性的甘油单酯或双酯。此外,脂肪酸例如油酸可在可注射的制剂中应用。适合口服,皮下,静脉内,肌肉内等施用的配方能在 Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, PA 中找到。

[0281] 根据本发明的一个发明,用于在细胞中增加 C_α-甲酰甘氨酸生成活性的方法被提供。方法包括将分离的本发明的核酸分子(例如 SEQ ID NO. 1 的核酸)或其表达产物(例如 SEQ ID NO. 2 的肽)以在细胞中增加 C_α-甲酰甘氨酸生成活性的有效量与细胞接触。在重要的实施方案中,方法包括激活内源性 FGE 基因以在细胞中增加 C_α-甲酰甘氨酸生成活性。在某些实施方案中,接触在允许本发明的分子进入细胞的条件下进行。

[0282] 根据本发明的术语“允许分子进入”细胞具有下面基于分子本性的含义。对分离的核酸,它用于描述核酸通过细胞膜而进入细胞核,在其基础上“核酸转基因”能利用细胞机制产生核酸所编码的功能多肽。“核酸转基因”用于描述所有的有或没有相连接载体的本发明的核酸。对于多肽,它用于描述多肽通过细胞膜进入细胞质,及如果需要,对细胞的细胞质机制的利用以功能性地修饰多肽(例如修饰成活性形式)。

[0283] 多种技术可被用于将本发明的核酸引入细胞,其依赖于核酸是在体外还是体内而被引入宿主。这类技术包括核酸-CaPO₄ 沉淀的转染,与 DEAE 相连的核酸的转染,包括目的核酸的逆转录病毒的转染,脂质体介导的转染和类似方法。对于一定的应用,优选将核酸靶向于特定细胞。在这类例子中,用于将本发明的核酸运送进细胞的载体(例如逆转录病毒或其它病毒;脂质体)可以具有附着于其上的靶向分子。例如,分子例如对靶细胞上的表面膜蛋白专一的抗体或靶细胞上的受体的配体能被固定到核酸运送载体上或整合在其中。例如,脂质体被用于运送本发明的核酸的情形中,结合到与内吞作用相关的表面膜蛋白的蛋白质可被整合进用于靶向和/或促进吸收的脂质体的配剂。这类蛋白质包括对特定细胞类型有亲和力的衣壳蛋白或其片段,针对循环中经历内化的蛋白质的抗体,靶向细胞内定位和增强细胞内半衰期的蛋白质,和类似物。聚合性运送系统也已被成功用于运送核酸进入细胞,如本领域技术人员所已知的。这类系统甚至允许核酸的口服运送。

[0284] 其它运送系统能包括定时释放性、延迟释放性或持续释放性运送系统。这类系统能避免本发明的试剂的重复施用,增加对受试者和医生的方便性。多种类型的释放运送系统可用,且为本领域一般技术人员所已知。它们包括以聚合物为基础的系统例如聚(丙交酯-乙交酯),共聚草酸,聚己内酯,聚酯酰胺,聚原酸酯,聚羟基丁酸和聚酸酐。包含药物的

前述聚合物的微囊在例如 U. S. Patent 5, 075, 109 中被描述。运送系统也包括非聚合物系统：脂类，包括固醇例如胆固醇，胆固醇酯和脂肪酸或中性脂肪例如单-二-和三-甘油酯；水凝胶释放系统；sylastic 系统；以肽为基础的系统；蜡涂层；应用传统粘合剂和赋形剂的压制成的药片；部分融合的植入物；和类似物。具体的实例包括但不限于：(a) 本发明的试剂以某一形式被包含在基质内的侵蚀性系统例如 U. S. Patent Nos. 4, 452, 775, 4, 675, 189 和 5, 736, 152 中所描述的那些类型，和 (b) 活性成分以受控的速率从聚合物中渗出的扩散系统例如 U. S. Patent Nos. 3, 854, 480, 5, 133, 974 和 5, 407, 686 中所述的。此外，以泵为基础的硬件运送系统能被应用，其中部分适于植入。

[0285] 长期持续释放性植入物的应用可以是合乎期望。在本文中所用的长期释放，表示植入物被构建和计划成至少 30 天运送治疗水平的活性成分，优选 60 天。长期持续释放性植入物是本领域中一般技术人员众所周知的，包括部分上述的释放系统。具体的实例包括但不限于，在 U. S. Patent No. 4, 748, 024 和加拿大 No. 1330939 中所述的长期持续释放性植入物。

[0286] 本发明也包括本发明的 FGE 分子之外的试剂的施用，和在某些实施方案中的共施用，所述试剂当以有效剂量施用时可与本发明的分子合作性地、加成性地或协同性地起作用，以：(i) 调节 C_α-甲酰甘氨酸生成活性，和 (ii) 治疗涉及本发明分子的 C_α-甲酰甘氨酸生成活性的任何病症（例如硫酸酯酶缺乏症包括多种硫酸酯酶缺乏症）。本发明的分子之外的试剂包括艾杜糖醛酸-2-硫酸酯酶，硫酸胺酶，N-乙酰半乳糖胺-6-硫酸酯酶，N-乙酰葡萄糖胺-6-硫酸酯酶，芳基硫酸酯酶 A，芳基硫酸酯酶 B，芳基硫酸酯酶 C，芳基硫酸酯酶 D，芳基硫酸酯酶 E，芳基硫酸酯酶 F，芳基硫酸酯酶 G，HSulf-1，HSulf-2，HSulf-3，HSulf-4，HSulf-5，或 HSulf-6，（核酸和多肽，和 / 或其片段），和 / 或其组合。

[0287] 在本文中所用的“共施用”指同时施用两种或更多的本发明的化合物（例如已知在对例如硫酸酯酶缺乏症的治疗中有益的 FGE 核酸和 / 或多肽和试剂，——例如治疗 MPSII 中的艾杜糖醛酸-2-硫酸酯酶），其作为单一组合物中的混合物，或以足够的时间间隔顺序施用以至于化合物可发挥出相加性或甚至协同性的效力。

[0288] 本发明也包含固相核酸分子阵列。阵列基本上由固定于固体基质的一套核酸分子、其表达产物或其（或者核酸或者多肽分子的）片段组成，每一核酸分子选自 FGE，艾杜糖醛酸-2-硫酸酯酶，硫酸胺酶，N-乙酰半乳糖胺-6-硫酸酯酶，N-乙酰葡萄糖胺-6-硫酸酯酶，芳基硫酸酯酶 A，芳基硫酸酯酶 B，芳基硫酸酯酶 C，芳基硫酸酯酶 D，芳基硫酸酯酶 E，芳基硫酸酯酶 F，芳基硫酸酯酶 G，HSulf-1，HSulf-2，HSulf-3，HSulf-4，HSulf-5，和 HSulf-6。在某些实施方案中，固相阵列进一步包含至少一种对照核酸分子。在一定实施方案中，这套核酸分子包含至少一种，至少两种，至少三种，至少四种或甚至至少五种核酸分子，每一种选自 FGE，艾杜糖醛酸-2-硫酸酯酶，硫酸胺酶，N-乙酰半乳糖胺-6-硫酸酯酶，N-乙酰葡萄糖胺-6-硫酸酯酶，芳基硫酸酯酶 A，芳基硫酸酯酶 B，芳基硫酸酯酶 C，芳基硫酸酯酶 D，芳基硫酸酯酶 E，芳基硫酸酯酶 F，芳基硫酸酯酶 G，HSulf-1，HSulf-2，HSulf-3，HSulf-4，HSulf-5，和 HSulf-6。在优选的实施方案中，这套核酸分子包含最多 100 种不同核酸分子。在重要的实施方案中，这套核酸分子包含最多 10 种不同核酸分子。

[0289] 根据此发明，标准的微阵列技术的杂交技术被用于评估核酸表达的模式和鉴别核酸表达。微阵列技术，其也称为其它名字，包括：DNA 芯片技术，基因芯片技术和固相核酸

阵列技术,为本领域一般技术人员已知,并基于但不限于,获得在固定的基质上的已鉴别出的核酸探针的阵列(例如在本文中其它部分所述如 FGE,艾杜糖醛酸-2-硫酸酯酶,硫酸胺酶, N-乙酰半乳糖胺-6-硫酸酯酶, N-乙酰葡萄糖胺-6-硫酸酯酶,芳基硫酸酯酶 A,芳基硫酸酯酶 B,芳基硫酸酯酶 C,芳基硫酸酯酶 D,芳基硫酸酯酶 E,芳基硫酸酯酶 F,芳基硫酸酯酶 G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, 和 / 或 HSulf-6 的分子),以报告分子(例如放射性的,化学发光的,或荧光性的标记例如荧光素, Cye3-dUTP 或 Cye5-dUTP) 标记靶分子,将靶核酸与探针杂交,和评估靶-探针的杂交。一般而言,具有与靶序列完全匹配的核酸序列的探针,将比不完全匹配的探针检测到更强的报告分子信号。在核酸微阵列技术中应用的多种成分和技术在 TheChipping Forecast, Nature Genetics, 卷 21, Jan 1999 中呈现,其全部内容在本文中被整合作为参考。

[0290] 根据本发明,微阵列基质可包括但不限于玻璃,硅土,铝硅酸盐,硼硅酸盐,金属氧化物例如氧化铝和氧化镍,多种粘土,硝基纤维素或尼龙。在所有实施方案中,优选玻璃基质。根据本发明,探针选自核酸,其包括但不限于:DNA,基因组 DNA, cDNA 和寡核苷酸;且可以是天然的或合成的。寡核苷酸探针优选 20 到 25 聚体寡核苷酸,而 DNA/cDNA 探针优选长度为 500 到 5000 个碱基,虽然其它长度也可被应用。合适的探针长度可被本领域一般技术人员通过本领域内已知的程序而测定。在某一实施方案中,优选的探针是以 SEQ ID NOs: 1, 3, 4, 6, 8, 10, 和 / 或 12 给出的核酸分子中的两种或更多的组合。探针可应用本领域一般技术人员已知的标准方法例如凝胶过滤或沉淀纯化,以除去污染物。

[0291] 在某一实施方案中,微阵列基质可用化合物涂层以增强探针在基质上的合成。这类化合物包括但不限于,寡聚乙二醇。在另一实施方案中,基质上的偶联试剂或基团能被用于共价连接第一核苷酸或寡核苷酸到基质。这些试剂或基团可包括但不限于:氨基,羟基,溴基和羧基。这些活性基团优选通过烃基例如亚烷基或亚苯基二价基团附着于基质,一个化合价位置被链连接占据而余下的一个附着到活性基团。这些烃基基团可包含多至约 10 个碳原子,优选多至约 6 个碳原子。亚烷基通常优选在主链中包含 2 到 4 个碳原子。本方法的这些和另外的细节在例如 U. S. Patent 4, 458, 066 中被公开,其全部内容被整合作为参考。

[0292] 在某一实施方案中,应用方法例如光导向的化学合成、光化学去保护或核苷酸前体到基质的运送及随后的探针产生,探针以预先确定的栅格模式直接在基质上被合成。

[0293] 在另一实施方案中,基质可用化合物涂层以增强探针与基质的结合。这类化合物包括但不限于:聚赖氨酸,氨基硅烷,氨基-反应性硅烷 (Chipping Forecast, 1999) 或铬 (Gwynne 和 Page, 2000)。

[0294] 在此实施方案中,预合成的探针以精确的、预先确定体积的和栅格的模式被应用于基质,利用了计算机控制的机器人将探针以接触打印方式或非接触方式例如喷墨或压电运送而应用于基质。以包括但不限于 UV 照射的方法,探针可共价连接到基质。在另一实施方案中,探针热连接到基质。

[0295] 靶标是选自包括但不限于下列的核酸:DNA,基因组 DNA, cDNA, RNA, mRNA, 且可以是天然的或合成性的。在所有实施方案中,优选来自被怀疑为发展或具有硫酸酯酶缺乏症的受试者的核酸分子。在本发明的一定的实施方案中,一种或更多对照核酸分子被附着于基质。优选地,对照核酸分子允许了对包括但不限于下列的因素的测定:核酸质量和结合特

征 ;试剂质量和有效性 ;杂交成功性 ;和分析阈值和成功性。对照核酸可包括但不限于,基因如管家基因的表达产物或其片段。

[0296] 为选择一套硫酸酯酶缺乏疾病的标记物,优选分析由例如基因表达的微阵列分析产生的表达数据,以确定在不同患者类别(每一患者类别是一种不同的硫酸酯酶缺乏疾病)中的哪个基因是显著差异性地表达。基因表达的显著性可应用 Permax 计算机软件被测定,虽然任何能区分表达的显著差异的标准统计包也可被使用。Permax 实现了对数据的大规模阵列的排列双样本 t- 检验。对于观察中的高维度载体,Permax 软件计算了对每一属性的 t- 统计,并应用全部属性的最大值与最小值的排列分布评估了显著性。主要应用是确定两组之间差异最大的属性(基因)(例如对照的健康受试者和具有特定硫酸酯酶缺乏症的受试者),应用 t- 统计的值测量“最大差异”和它们的显著性水平。

[0297] 硫酸酯酶缺乏疾病相关性核酸分子的表达也能应用蛋白测量方法测定,以测定 SEQ ID NOs :2 的表达,例如通过测定 SEQ ID NOs :1 和 / 或 3 编码的多肽的表达。特异和定量测量蛋白质的优选方法包括但不限于:基于质谱的方法例如表面增强的激光解吸电离(SELDI ;例如 Ciphergen 蛋白质芯片系统),非基于质谱的方法,和基于免疫组织化学的方法例如二维凝胶电泳。

[0298] 通过本领域一般技术人员已知的程序,SELDI 方法可被用于蒸发显微量的蛋白,和创造个体蛋白质的“指纹”,从而允许对在单一样本中的多种蛋白质的含量的同时测量。优选地,基于 SELDI 的分析可被用于鉴别多种硫酸酯酶缺乏症的特征,以及这类病症的阶段。这类分析优选包括,但不限于下列实例。RNA 微阵列所发现的基因产物可通过被专一(抗体介导的)捕获到 SELDI 蛋白质盘(例如选择性 SELDI)而被选择性地测量。蛋白筛选(例如以 2 维凝胶)所发现的基因产物可通过最优化的“总蛋白 SELDI”解析以可视化那些来自 SEQ ID NOs :1,6,8,10,12,14,16,18,20,22,24,26,和 / 或 28 中的特定的目的标记。来自 SEQ ID NOs :1,6,8,10,12,14,16,18,20,22,24,26,和 / 或 28 中的多种标记物的 SELDI 测量的特定硫酸酯酶缺乏症的预测模型可被用于 SELDI 策略。

[0299] 任何前述微阵列方法在测定硫酸酯酶缺乏疾病相关核酸的蛋白中的应用,能以本领域一般技术人员已知的常规方法完成,而通过蛋白质测量方法测定的表达可与标记物的预测水平相关,用作对硫酸酯酶缺乏疾病患者治疗策略进行选择的预测性方法。

[0300] 本发明也包含硫酸酯酶产生性细胞,其中细胞产生的活性硫酸酯酶对总硫酸酯酶的比率(即比活性)被增加。细胞包含:(i) 具有增强的表达的硫酸酯酶,和(ii) 具有增强的表达的甲酰甘氨酸生成酶,其中细胞产生的活性硫酸酯酶对总硫酸酯酶的比率,相对于甲酰甘氨酸生成酶缺乏时细胞所产生的活性硫酸酯酶对总硫酸酯酶的比率,增加至少 5%。

[0301] 在本文中所用的“具有增强的表达的硫酸酯酶”典型地指硫酸酯酶和 / 或其编码的多肽相对于对照的增加的表达。增加的表达指增加(即到可探测的程度)任何硫酸酯酶核酸(如本文中其它部分所描述的本发明的硫酸酯酶核酸)的复制、转录和 / 或翻译,因为任何这些过程的上调会导致基因(核酸)所编码的多肽的浓度 / 数量的增加。这能应用大量本领域中已知也在本文中其它部分描述的方法而完成,例如以硫酸酯酶 cDNA 和 / 或包含硫酸酯酶位点的基因组 DNA 对细胞的转染,通过应用同源重组将例如强启动子元件置于内源性硫酸酯酶基因的基因组位点上游而激活内源性硫酸酯酶基因(参见,例如 U. S. Patents Nos. 5, 733, 761、6, 270, 989 和 6, 565, 844 中详细描述基因激活技术,所有这

些被整合在本文中作为参考)等。典型的对照将是载体质粒所转染的同样的细胞。增强(或增加)硫酸酯酶活性也指防止或抑制硫酸酯酶的降解(例如通过增强的泛素化),下调等,其导致例如相对于对照的增加的或稳定的硫酸酯酶分子 $t_{1/2}$ (半衰期)。下调或降低的表达指基因和/或其编码的多肽的降低的表达。基因表达的上调或下调能通过一定方式直接测定,所述方式为应用任何本领域已知的适合的方式例如核酸杂交或抗体探测方法分别探测基因(例如,硫酸酯酶)的 mRNA 水平或基因所编码多肽的蛋白质表达水平的增加或降低,并与对照比较。硫酸酯酶基因表达的上调或下调也能通过探测硫酸酯酶活性的变化而间接测定。

[0302] 类似地,在本文中所用的“具有增强的表达的甲酰甘氨酸生成酶”,典型地指本发明的 FGE 核酸和/或其编码的多肽相对于对照的增加的表达。增加的表达指增加(即到可探测的程度)任何本发明的 FGE 核酸(如本文中其它部分所描述的)的复制、转录和/或翻译,因为任何这些过程的上调会导致基因(核酸)所编码的多肽的浓度/数量的增加。这能应用上述的(对硫酸酯酶)和在本文中其它部分描述的方法而完成,

[0303] 在一定实施方案中,相对于甲酰甘氨酸生成酶缺乏的细胞所产生的活性硫酸酯酶对总硫酸酯酶的比率,细胞所产生的活性硫酸酯酶对总硫酸酯酶的比率增加至少 10%, 15%, 20%, 50%, 100%, 200%, 500%, 1000%。

[0304] 本发明进一步包含在受试者中治疗硫酸酯酶缺乏症的改进的方法。方法包括对需要这类治疗的受试者以有效剂量施用硫酸酯酶以治疗受试者中的硫酸酯酶缺乏症,其中硫酸酯酶与一定量的甲酰甘氨酸生成酶接触,所述量是增加硫酸酯酶比活性的有效量。如本文中其它部分所描述的,“比活性”指所产生的活性硫酸酯酶对总硫酸酯酶的比率。在此所用的“接触”指如本文中其它部分所描述的 FGE 翻译后修饰硫酸酯酶。将对一般技术人员显然的是,如果编码 FGE 和硫酸酯酶的核酸在细胞中共表达,或甚至如果分离的 FGE 多肽与分离的硫酸酯酶多肽在体内或体外接触,则 FGE 能接触硫酸酯酶并对它进行修饰。即使分离的 FGE 多肽能与分离的硫酸酯酶多肽共施用于受试者,以治疗受试者中的硫酸酯酶缺乏症,在 FGE 和硫酸酯酶之间的接触也优选在硫酸酯酶施用于受试者之前在体外发生。既然更低量的硫酸酯酶需要被施用,和/或以更低频率施用(因为硫酸酯酶具有更高的比活性),此改进的治疗方法对受试者有益。

[0305] 参考下面的实施例,本发明将被更充分地理解。然而,这些实施例只倾向于举例说明本发明的实施方案,而不应被解释为限制本发明的范围。

[0306] 实施例

[0307] 实施例 1:

[0308] 多种硫酸酯酶缺乏症由编码人 C_α-甲酰甘氨酸生成酶(FGE)的基因中的突变所引起。

[0309] 实验程序

[0310] 材料和方法

[0311] FGE 的体外分析

[0312] 为监测 FGE 的活性, N-乙酰化和 C-酰胺化的 23 聚肽 P23(MTDFYVPVSLCTPSRAALLTGRS)(SEQ ID NO:33)被用作底物。位点 11 的半胱氨酸残基向 FG1y 转化通过 MALDI-TOF 质谱监测。30%乙氰和 0.1%三氯乙酸(TFA)中的 6 μM 的

P23 原液被制备。在标准条件下,6pmol 的 P23 与高至 10 μ l 的酶一起在最终体积为 30 μ l 的 50mM Tris/HCl (pH 9.0, 含有 67mM NaCl, 15 μ M CaCl₂, 2mM DTT, 和 0.33mg/ml 牛血清白蛋白) 中于 37°C 培育。为停止酶反应, 加入 1.5 μ l 10% TFA。然后 P23 被固着于 ZipTip C18 (Millipore), 以 0.1% TFA 清洗, 并在 3 μ l 的 50% 乙腈和 0.1% TFA 中洗脱。0.5 μ l 的洗脱液与 0.5 μ l 的基质溶液 (50% 乙腈和 0.1% TFA 中的 5mg/ml α -氰基-4-羟基-肉桂酸 (BrukerDaltonics, Billerica, MA)) 在不锈钢的靶上混合。MALDI-TOF 质谱应用反射模式和刚好高于解吸 / 电离阈值的激光能量以 Reflex III (Bruker Daltonics) 完成。所有谱线是来自靶标上几个点的 200-300 次射击 (shot) 的平均值。质量轴应用分子质量范围在 1000 到 3000Da 的肽作为外部标准而校准。P23 的单同位素 MH⁺ 是 2526.28, 而包含 FGly 产物的是 2508.29。活性 (pmol 产物 / 小时) 根据产物的峰值高度除以 P23 和产物的峰值高度总和而计算。

[0313] 来自牛睾丸的 FGE 的纯化

[0314] 牛睾丸得自当地屠宰场, 冰上保存最多 20 小时。实质 (Parenchyme) 从结缔组织中释放, 并在韦林氏搅切器中和通过三轮次的马达陶制 (motor pottering) 而匀浆。通过所得到的匀浆的细胞分级分离, 粗微粒体 (RM) 的制备按 (Meyer 等, J. Biol. Chem., 2000, 275:14550-14557) 所述并进行下面的修正而完成。三次差异离心步骤, 每次 4°C 下 20 分钟, 在 500g (JA10 转子), 3000g (JA10) 和 10000g (JA20) 下完成。RM 膜从最后一次的上清液中沉淀 (125000g, Ti45 转子, 45 分钟, 4°C), 通过马达陶制匀浆并在蔗糖垫层上分层 (50mM Hepes, pH 7.6, 50mM KAc, 6mM MgAc₂, 1mM EDTA, 1.3M 蔗糖, 5mM β -巯基乙醇)。在 Ti45 转子中以 45000rpm 4°C 下离心 210 分钟后, RM 从沉淀中回收。通常 100000-150000 等价的 RM, 如 Walter 和 Blobel (Methods Enzymol., 1983, 96:84-93) 所定义, 从 1kg 睾丸组织中得到。reticuloplasm, 即 RM 的腔内容物通过低浓度脱氧 Big Chap 下的分级抽提而得到, 如 Fey 等, J. Biol. Chem., 2001, 276:47021-47028 所述。对 FGE 的纯化, 95ml 的 reticuloplasm 4°C 下对 20mM Tris/HCl, pH 8.0, 2.5mM DTT 透析 20 小时, 而通过 125000g 下离心 1 小时澄清。32ml 澄清 reticuloplasm 等份物室温下被装在 MonoQ HR10/10 柱 (Amersham Biosciences, Piscataway, NJ) 上, 清洗和用线性梯度的 80ml Tris 缓冲液中的 0 到 0.75M NaCl 以 2ml/min 洗脱。三轮洗脱的包含 FGE 活性的级分 (50-165mM NaCl 洗脱) 被收集 (42ml), 与 2ml 的伴刀豆球蛋白 A-Sepharose (Amersham Biosciences, 已用含有 0.5M KCl, 1mM MgCl₂, 1mM MnCl₂, 1mM CaCl₂, 和 2.5mM DTT 的 50mM Hepes 缓冲液, pH 7.4 清洗过) 混合。4°C 下培育 16 小时后, 伴刀豆球蛋白 A-Sepharose 在柱中被收集, 并用 6ml 相同的 Hepes 缓冲液清洗。固着的物质通过柱在室温下与 6ml 0.5M α -甲基甘露糖苷在 50mM Hepes, pH 7.4, 2.5mM 中培育 1 小时而洗脱。洗脱用 4ml 的相同洗脱剂重复。来自伴刀豆球蛋白 A-Sepharose 的合并的洗脱液 (10ml) 以 0.5M Tris/HCl (pH 9.0) 调至 pH 8.0, 并与用 10mg 混杂肽 (PVSLPTRSCAALLTGR) (SEQ ID NO:34) 衍生化的 2ml 的 Affigel 10 (Bio-Rad Laboratories, Hercules, CA) 混合, 以缓冲液 A (50mM Hepes, pH 8.0, 含有 0.15M 乙酸钾, 0.125M 蔗糖, 1mM MgCl₂, 和 2.5mM DTT) 清洗。4°C 下培育 3 小时后, 亲和基质在柱中被收集。收集经过的液体 (flow through) 和用 4ml 缓冲液 A 清洗的组分, 合并, 并与已用 10mg 的 Ser69 肽 (PVSLSTPSRAALLTGR) (SEQ ID NO:35) 取代的 2ml Affigel 10 混合, 用缓冲液 A 清洗。4°C 下培育过夜, 亲和基质在柱中被收集, 用 6ml 的缓冲液 B (含有 2M NaCl 和 20 种组成

蛋白质的氨基酸（每一浓度为 50mg/ml）的混合物的缓冲液 A）清洗 3 次。通过将 Affigel 与含有 25mM Ser69 肽的 6ml 缓冲液 B 培育两次，每次 90min，而将固着的物质从亲和基质洗脱。洗脱液的等分组分以 1mg/ml 牛血清白蛋白取代，对缓冲液 A 透析，并分析活性。活性的保留部分（11.8ml）在 Vivaspin500 浓缩器（Vivascience AG, Hannover, Germany）中浓缩，并在 95°C 下溶于 Laemmli SDS 样品缓冲液。起始物质和层析步骤后得到的制剂的多肽组成以 SDS-PAGE（15% 丙烯酰胺，0.16% 双丙烯酰胺）监测，和以 SYPRO Ruby（Bio-Rad Laboratories）染色。

[0315] 通过质谱对 FGE 的鉴定

[0316] 对于肽质量指纹分析，纯化的多肽用胰蛋白酶在凝胶内被消化（Shevchenko 等，Anal. Chem., 1996, 68 :850-855），在 C18 ZipTip 上脱盐，并应用二氢苯甲酸作为基质和两种来自胰蛋白酶的自降解肽（m/z 842.51 和 2211.10）作为内标准而以 MALDI-TOF 质谱进行分析。对串联质谱分析，选出的肽通过 MALDI-TOF post-source decay 质谱而被分析。它们对应的带双电荷的离子被分离出，并被 offline nano-ESI 离子阱质谱（EsquireLC, Bruker Daltonics）所片段化。质谱数据通过 Mascot 检索算法用于 NCBI 蛋白质数据库中和 NCBI EST 核苷酸数据库中的蛋白质鉴定。

[0317] 生物信息学

[0318] 信号肽和切断的位置根据在 EMBOSS（Rice 等，Trends in Genetics, 2000, 16 : 276-277）中实行的 von Heijne（von Heijne, Nucleic Acids Res., 1986, 14 :4683-90）的方法被描述。N-糖基化位点应用 Brunak 的算法（Gupta 和 Brunak, Pac. Symp. Biocomput., 2002, 310-22）被预测。功能结构域通过搜索 PFAM-Hidden-Markov-Models（version 7.8）（Sonnhammer 等，Nucleic Acids Res., 1998, 26 :320-322）而被探测到。为搜索 FGE 同源物，National Center for Biotechnology Information 的数据库（Wheeler 等，Nucleic Acids Res., 2002, 20 :13-16）用 BLAST（Altschul 等，Nucleic Acids Res., 1997, 25 :3389-3402）查询。序列类似度应用来自 EMBOSS 的标准工具而计算。基因组位点组织和同线性应用 NCBI 的人和小鼠基因组资源及也来自 NCBI, Bethesda, MD 的人类 - 小鼠同源性图谱而确定。

[0319] 人 FGE cDNA 的克隆

[0320] 从人成纤维细胞应用 RNEASY™ Mini 试剂盒（Qiagen, Inc., Valencia, CA）制备的总 RNA，应用 OMNISCRIPRT™ 试剂盒（Qiagen, Inc., Valencia, CA）以及寡（dT）引物或者 FGE 专一的引物 1199nc（CCAATGTAGGTCAGACACG）（SEQ ID NO :36）而被反转录。cDNA 的第一链通过应用正向引物 1c（ACATGGCCCGCGGGAC）（SEQ ID NO :37）和作为反向引物的 1199nc 或 1182nc（CGACTGCTCCTTGGACTGG）（SEQ ID NO :38）的 PCR 而扩增。PCR 产物被直接克隆进 pCR4-TOPO™ 载体（Invitrogen Corporation, Carlsbad, CA）。通过对克隆 PCR 产物（其从不同的个体和独立的 RT-PCR 反应得到）的多种测序，FGE cDNA 的编码序列被确定（SEQ ID NOs :1 和 3）。

[0321] 突变探测，基因组测序，定点诱变和 RNA 印迹分析

[0322] 此研究中采用的标准流程本质上如 Lübke 等（Nat. Gen., 2001, 28 :73-76）和 Hansske 等（J. Clin. Invest., 2002, 109 :725-733）所述。Northern 点与覆盖整个编码区 cDNA 探针杂交，并与 β -肌动蛋白 cDNA 探针杂交而作为 RNA 负载的对照。

[0323] 细胞系和细胞培养

[0324] 来自多种硫酸酯酶缺乏症患者1-6的成纤维细胞分别从 E. Christenson (Rigshospitalet Copenhagen), M. Beck (**Universitätskinderklinik Mainz**), A. Kohlschütter (**Universitätskrankenhaus Eppendorf, Hamburg**), E. Zammarchi (Meyer Hospital, University of Florence), K. Harzer (Institut **für Hirnforschung, Universität Tübingen**), 和 A. Fensom (Guy's Hospital, London) 得到。人皮肤成纤维细胞, HT-1080, BHK21 和 CHO 细胞被以 37°C 和 5% CO₂ 下保存在含有 10% 胎牛血清的 Dulbecco 的修饰后的 Eagle 培养基中。

[0325] 转染, 间接免疫荧光, Western 印迹分析和 FGE 活性的探测

[0326] 通过 add-on PCR (其应用了 Pfu 聚合酶 (Stratagene, La Jolla, CA) 和下列引物: GGAATTCGGGACAACATGGCTGCG (EcoRI) (SEQ ID NO:39), CCAAGCTTATGCGTAGTCAGGCACATCATACGGATAGTCCATGGTGGGCAGGC (HA) (SEQ ID NO:40), CCAAGCTTACAGGTCTTCTTCAGAAATCAGCTTTTGTTCGTCCATGGTGGGCAGGC (c-Myc) (SEQ ID NO:41), CCAAGCTTAGTGATGGTATGGTATGGC GATCCTCTGTCCATGGTGGGCAGGC (RGS-His₆) (SEQ ID NO:42)), FGE cDNA 被装上 5' EcoRI-位点以及 3' HA-, c-Myc 或者 RGS-His₆- 标记序列, 及其后的终止密码子和 HindIII 位点。得到的 PCR 产物作为 EcoRI/HindIII 片段被克隆进 pMPSVEH (Artelt 等, Gene, 1988, 68: 213-219)。应用 EFFECTENE™ (Qiagen) 作为转染试剂, 得到的质粒被瞬时转染进生长在盖玻片上的 HT-1080, BHK21 和 CHO 细胞。转染 48 小时后, 细胞通过如前面所述的间接免疫荧光 (Lübke 等, Nat. Gen., 2001, 28:73-76; Hansske 等, J. Clin. Invest., 2002, 109:725-733), 应用抗 HA 的单克隆 IgG1 抗体 (Berkeley 抗体 Company, Richmond, CA), 抗 c-Myc 的单克隆 IgG1 抗体 (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) 或抗 RGS-His 的单克隆 IgG1 抗体 (Qiagen) 作为第一抗体而被分析。内质网标记蛋白即蛋白二硫键异构酶 (PDI) 被用不同亚型的单克隆抗体 (IgG2A, Stressgen Biotech., Victoria BC, Canada) 所探测。第一抗体被用分别偶联到 CY2 或 CY3 的同型-专一性的山羊第二抗体 (Molecular Probes, Inc., Eugene, OR) 所探测。免疫荧光图像在 Leica TCS Sp2 AOBIS 激光扫描显微镜上得到。对 Western 印迹分析, 应用相同的单克隆抗体, HRP-缀合的抗-小鼠 IgG 被用作第二抗体。对 FGE 活性的测定, 胰蛋白酶化的细胞用含有 2.5mM DTT、蛋白酶抑制剂和 1% Triton X-100 的 10mM Tris (pH 8.0) 所溶解的磷酸缓冲盐溶液清洗, 其含有蛋白酶抑制剂的混合物 (208 μM 4-(2-氨基乙基) 盐酸苯磺酰基氟, 0.16 μM 抑酶肽, 4.2 μM 亮抑酶肽, 7.2 μM 苯丁抑制素, 3 μM 抑胃酶肽 A, 2.8 μM ME-64), 并通过在 125,000g 离心 1 小时而澄清。上清液在 MonoQ PC 1.6/5 柱上应用上述条件进行层析。在 50-200mM NaCl 中洗脱的组分被收集, 冷冻干燥, 并重构成原始体积的十分之一, 随后用肽 P23 测定 FGE 活性。

[0327] 逆转录病毒转导

[0328] 目的 cDNAs 被克隆进基于莫洛尼鼠白血病病毒的载体 pLPCX 和 pLNCX2 (BD Biosciences Clontech, Palo Alto, CA)。嗜亲性的 FNX-Eco 细胞 (ATCC, Manassas, VA) 的转染, 及兼嗜性 RETROPACK™ PT67 细胞 (BD Biosciences Clontech) 和人成纤维细胞的转导按 Lübke 等, Nat. Gen., 2001, 28:73-76; Thiel 等, Biochem. J., 2002, 376, 195-201 所述完成。对部分实验, 在测定硫酸酯酶活性前, 用嘌呤霉素选择 pLPCX 转导的 PT67 细胞

[0329] 硫酸酯酶分析

[0330] ASA, STS 和 GalNAc6S 的活性按 Rommerskirch 和 von Figura, Proc. Natl. Acad. Sci., USA, 1992, 89 :2561-2565 ;Glössl 和 Kresse, Clin. Chim. Acta, 1978, 88 :111-119 中所述被测定。

[0331] 结果

[0332] 对 FGE 活性的基于肽的快速分析

[0333] 我们已发展出应用体外合成的 [³⁵S]ASA 片段作为底物而用于在微粒体抽提物中测定 FGE 活性的分析法。片段被加入分析混合物而作为核糖体相连的初生链复合物。对产物的定量包括胰蛋白酶消化、肽通过 RP-HPLC 的分离, 及通过化学衍生为脎、RP-HPLC 分离和液闪计数的组合而对含有 [³⁵S]- 标记的 FGly 的胰蛋白酶降解肽的鉴别和定量 (Fey 等, J. Biol. Chem., 2001, 276 :47021-47028)。为监测纯化过程中的酶活性, 此麻烦的程序需要被修正。对应于 ASA 残基 65-80 和含有 FGly 形成所需的序列基序的合成性的 16 聚肽在体外分析中抑制了 FGE 活性。这暗示了肽例如 ASA65-80 可作为 FGE 的底物而起作用。我们合成了 23 聚肽 P23 (SEQ ID NO :33), 其对应于 ASA 残基 60-80 并附加了 N- 乙酰化的甲硫氨酸和 C- 酰胺化的丝氨酸残基以分别保护 N- 和 C- 末端。含有半胱氨酸和 FGly 的 P23 形式能通过基质辅助的激光解析 / 电离飞行时间 (MALDI-TOF) 质谱鉴别和定量。FGly 残基在 P23 位点 11 的存在通过 MALDI-TOF post source decay 质谱证实 (参见 Peng 等, J. MassSpec., 2003, 38 :80-86)。P23 和来自牛胰或牛睾丸微粒体的抽提物的培育将高至 95% 的肽转化成为含有 FGly 的衍生物 (图 1)。在标准条件下, 反应与酶的数量和培育的时间成正比例, 只要少于 50% 的底物被消耗和培育时间不超过 24 小时。P23 的 k_m 是 13nM。减少的和氧化的谷胱甘肽、Ca²⁺ 和 pH 的效应与在应用核糖体相连初生链复合物作为底物的分析中所见到的那些相当 (Fey 等, J. Biol. Chem., 2001, 276 :47021-47028)。

[0334] FGE 的纯化

[0335] 对 FGE 的纯化, 牛睾丸微粒体的可溶性的级分 (reticuloplasm) 作为起始物质。FGE 的比活性比那些来自牛胰微粒体的 reticuloplasm 中的类型高 10-20 倍 (Fey 等, J. Biol. Chem., 2001, 276 :47021-47028)。FGE 的纯化通过四个层析步骤的组合而完成。前两个步骤是在 MonoQ 阴离子交换柱和伴刀豆球蛋白 A-Sepharose 柱上的层析。在 pH 8, FGE 活性物被固着到 MonoQ, 并在 50-165mM NaCl 被洗脱, 而具有 60-90% 的回收率。当此级分与伴刀豆球蛋白 A-Sepharose 混合时, FGE 被固着。起始活性的 30-40% 能用 0.5M α - 甲基甘露糖苷洗脱。后面两个纯化步骤是以 16 聚肽衍生的亲和基质上的层析。第一亲和基质是 ASA65-80 肽的变体所取代的 Affigel 10, 其中对 FGly 形成很关键的残基 Cys69, Pro71 和 Arg73 被搅乱 (混杂肽 PVSLPTRSCAALLTGR-SEQ ID NO :34)。当以 10mM 浓度加入体外的分析中时, 此肽不会抑制 FGE 活性, 且当固定到 Affigel 10 时不会保持 FGE 活性。在混杂肽亲和基质上的层析除去了肽结合性蛋白, 包括内质网的伴侣。第二亲和基质是 ASA65-80 肽的变体所取代的 Affigel 10, 其中 Cys69 被丝氨酸取代 (Ser69 肽 PVSLSTPSRAALLTGR-SEQ IDNO :35)。Ser69 肽亲和基质有效结合 FGE。FGE 活性物能用 2M KSCN 或者 25mM Ser69 肽洗脱, 而具有 20-40% 的回收率。在活性测定之前, KSCN 或 Ser69 肽必须通过透析除去。Cys69 被丝氨酸的取代对活性 FGE 的洗脱至关重要。被野生型 ASA65-80 肽所取代的 Affigel 10 有效结合 FGE。然而, 几乎没有活性能在用离液盐 (KSCN, MgCl₂), 肽 (ASA65-80 或 Ser69 肽), 或者低或高 pH 值缓冲液洗脱的洗脱液中回收。在图 2 中显示了起始物质和在经历了典型

纯化过程的四个层析步骤之后得到的活性组分的多肽模式。在最终级分中,5%的起始 FGE 活性和 0.0006%的起始蛋白被回收(8333-倍纯化)。

[0336] 纯化的 39.5 和 41.5kDa 多肽被单个基因编码

[0337] 在纯化的 FGE 制剂中的 39.5 和 41.5kDa 多肽被进行肽质量指纹分析。通过 MALDI-TOF 质谱得到的两个多肽的胰蛋白酶肽的质谱很大程度地重叠,暗示着两个蛋白质源自相同基因。在两个多肽的胰蛋白酶肽中,两个富含肽($MH^+1580.73$, SQNTPDSSASNLGFR(SEQ ID NO:43) 和 $MH^+2049.91$, MVPIPAGVFTMGTDPPQIK-SEQ ID NO:44 外加两处甲硫氨酸氧化)被发现,其与具有 GenBank Acc. No. AK075459(SEQ ID NO:4) 的 cDNA 所编码的蛋白质吻合。这两个肽的氨基酸序列被 MALDI-TOFpost source decay 谱和应用 offline nano-electrospray ionisation (ESI) 离子阱质谱的 MS/MS 分析所确认。人 cDNA 的牛定向进化同源物的 EST 序列,其覆盖 FGE 的 C 端部分并与两个肽的序列都吻合,提供了牛 FGE 的另外的序列信息。

[0338] FGE 的进化保守和结构域结构

[0339] 人 FGE 基因被 SEQ ID NOs :1 和 / 或 3 的 GDNA 编码,并位于染色体 3p26。它跨度约为 105kb,而编码序列分散在 9 个外显子中。人 FGE 基因的三个定向进化同源物在小鼠中(87%一致),黄果蝇(48%一致),和疟蚊(47%一致)被发现。定向进化同源物的 EST 序列在更多的 8 个物种中被发现,包括奶牛,猪,非洲爪蟾, *Silurana tropicalis*, 斑马鱼, 鲑鱼和其它鱼物种(对于细节,参见实施例 2)。外显子-内含子结构在人与小鼠基因间保守,而在染色体 6E2 上的小鼠基因位于与人染色体 3p26 同线的区域内。*S. cerevisiae* 和 *C. elegans* 的基因组缺乏 FGE 同源物。在原核生物中,12 种人 FGE 的同源物被发现。人 FGE 的 cDNA 被预测编码一个 374 残基的蛋白(图 3 和 SEQ ID NO:2)。此蛋白包含一个 33 残基的可切断的信号序列,其表明 FGE 到内质网中的迁移,此蛋白也包含了在 Asn141 处的单一 N-糖基化位点。FGE 与伴刀豆球蛋白 A 的结合暗示着此 N-糖基化位点被应用。FGE 的残基 87-367 被列在 PFAM 蛋白质基序数据库中,作为未知功能的结构域(PFAM :DUF323)。人 FGE 和其在数据库中已鉴定出的真核生物的定向进化同源物的序列比较分析说明此结构域由三个不同的亚结构域组成。

[0340] 在四种已知的真核生物 FGE 定向进化同源物中, N 端亚结构域(人 FGE 中的残基 91-154) 具有 46%的序列一致性和 79%的相似性。在人 FGE 中,此结构域在 Asn 141 携带 N-糖基化位点,这在其它定向进化同源物中保守。FGE 的中间部分(人 FGE 中的残基 179-308) 为富含色氨酸的亚结构域(每 129 残基中 12 个色氨酸)。真核生物定向进化同源物在此亚结构域内的一致性是 57%,相似性是 82%。C 端亚结构域(人 FGE 中的残基 327-366) 是 FGE 家族内最高度保守的序列。人 C 端亚结构域与真核生物定向进化同源物(3 个全长序列和 8 个 EST) 的序列一致性是 85%,相似性是 97%。在亚结构域 3 的 40 个残基内,四个半胱氨酸残基完全保守。三个半胱氨酸也在原核生物 FGE 定向进化同源物中保守。FGE 家族的 12 个原核生物成员(细节参见实施例 2) 与真核生物 FGE 共享亚结构域结构。三个亚结构域之间的边界在原核生物 FGE 家族中更加明显,因为可变长度的非保守序列将亚结构域彼此分开。人和小鼠基因组编码两个密切相关的 FGE 同源物(SEQ ID NOs :43 和 44, GenBank Acc. No. NM_015411, 在人中;和 SEQ ID NOs :45 和 46, GenBank Acc. No. AK076022, 在小鼠中)。两个平行进化同源物为 86%一致。它们的基因位于同线染色体区域(在人中

7q11, 在小鼠中 5G1)。两个平行进化同源物都与 FGE 定向进化同源物分享亚结构域结构, 与人 FGE 有 35% 一致性和 47% 的相似性。在两个同源物中都 100% 一致的第三亚结构域中, 含有半胱氨酸的亚结构域 3 的十一体序列缺失。

[0341] 表达, 亚细胞定位和分子形式

[0342] 2.1kb 的单一转录子通过来自成纤维细胞的总 RNA 和来自心脏、脑、胎盘、肺、肝脏、骨骼肌、肾和胰的聚 A⁺RNA 的 RNA 印迹分析可被探测。相对于 β -肌动蛋白 RNA, 含量有一个数量级的变化, 而在胰和肾中最高, 脑中最低。多种真核生物细胞系, 其稳定地或瞬时表达人 FGE 的 cDNA 或 C 端被 HA-, Myc- 或 His₆-tag 所延长的 FGE 衍生物的 cDNA, 被用于对 FGE 活性和 FGE 亚细胞定位进行分析。标记或非标记的 FGE 的瞬时表达增加了 1.6-3.9 倍的 FGE 活性。FGE 在 PT67 细胞中的稳定表达增加了约 100 倍的 FGE 活性。在 BHK 21, CHO 和 HT1080 细胞中通过间接免疫荧光对带标记的 FGE 形式的探测显示多种带标记的 FGE 形式与蛋白质二硫键异构酶 (一种内质网的腔蛋白) 共定位。来自 BHK 21 细胞 (被编码带标记的 FGE 形式的 cDNA 所瞬时转染) 的抽提物的 Western 印迹分析显示单一的免疫活性条带, 其表观大小在 42 到 44kDa 之间。

[0343] FGE 基因携带多种硫酸酯酶缺乏症中的突变

[0344] 多种硫酸酯酶缺乏症由硫酸酯酶中无法产生 FGly 残基的缺陷所引起 (Schmidt, B. 等, Cell, 1995, 82 :271-278)。FGE 基因因此是多种硫酸酯酶缺乏症的候选基因。我们对 7 位多种硫酸酯酶缺乏症患者的 FGE 编码性 cDNA 进行了扩增和测序, 发现十个不同的突变, 其被基因组 DNA 测序所确证 (表 1)。

[0345] 表 1 :MSD 患者中的突变

[0346]

| 突变 | 对蛋白质的作用 | 说明 | 患者 |
|-------------|--------------|---------------------------|------|
| 1076C > A | S359X | C-端 16 残基截短 | 1* |
| IVS3+5-8del | 删去残基 149-173 | 按读码框删去外显子 3 | 1, 2 |
| 979C > T | R327X | 失去亚结构域 3 | 2 |
| 1045C > T | R349W | 置换亚结构域 3 中的保守残基 | 3, 7 |
| 1046G > A | R349Q | 置换亚结构域 3 中的保守残基 | 4 |
| 1006T > C | C336R | 置换亚结构域 3 中的保守残基 | 4 |
| 836C > T | A279V | 置换亚结构域 2 中的保守残基 | 5 |
| 243delC | 移码和截短 | 丢失所有三个亚结构域 | 5 |
| 661delG | 移码和截短 | 丢失 FGE 的 C 末端第三, 包括亚结构域 3 | 6** |

| | | | |
|-------------|--------------|-------------|---|
| IVS6-1G > A | 删去残基 281-318 | 按读码框删去外显子 7 | 5 |
|-------------|--------------|-------------|---|

[0347] * 患者 1 是在 Schmidt, B. 等, Cell, 1995, 82 :271-278 和 Rommerskirch 和 vonFigura, Proc. Natl. Acad. Sci., USA, 1992, 89 :2561-2565 中的多种硫酸酯酶缺乏症患者 Mo.。

[0348] ** 患者 6 是被 Burk 等, J. Pediatr., 1984, 104 :574-578 报道的多种硫酸酯酶缺乏症患者。

[0349] 其它患者代表未发表的病例。

[0350] 第一位患者是 1076C > A 置换 (将丝氨酸 359 的密码子转换成了终止密码子) (S359X) 和导致 149-173 的 25 个残基 (被外显子 3 所编码和将蛋白质的第一与第二结构域分开) 的删除的突变的杂合子。基因组测序显示第三个内含子的 +5-8 核苷酸的删除 (IVS3+5-8del), 其从而破坏了内含子 3 的剪切供体位点。第二位患者是引起外显子 3 丢失 (IVS3+5-8del) 的突变和 979C > T 置换 (将精氨酸 327 的密码子转换成了终止密码子) (R327X) 的杂合子。由 979C > T 等位基因编码的截短的 FGE 缺乏了亚结构域 3 的绝大部分。第三位患者是 1045C > T 置换 (亚结构域 3 中的保守精氨酸 349 被色氨酸取代) (R349W) 的纯合子。第四位患者是 FGE 结构域中取代保守残基的两个错义突变的杂合子: 1046 > T 置换 (精氨酸 349 被谷氨酰胺取代) (R349Q) 和 1006T > C 置换 (半胱氨酸 336 被精氨酸取代) (C336R)。第五位患者是 836C > T 置换 (保守丙氨酸 279 被缬氨酸取代) (A279V) 的杂合子。第二突变是一种单一核苷酸删除 (243delC), 其在脯氨酸 81 之后改变序列并导致残基 139 之后的翻译停止。第六位患者是单一核苷酸删除 (661delG) (其在残基 220 之后改变氨基酸序列并在残基 266 之后引入终止密码子) 的杂合子。第二突变是内含子 6 的剪切受体位点突变 (IVS6-1G > A), 其导致编码残基 281-318 的外显子 7 的按读码框删除。在第七位患者中, 与在第三位患者中相同的 1045C > T 置换被发现。此外, 我们探测来自对照和多种硫酸酯酶缺乏症患者的 18FGE 等位基因的编码区中的两种多态性。22% 携带 188G > A 置换 (丝氨酸 63 被天冬酰胺取代) (S63N), 而 28% 有沉默的 1116C > T 置换。

[0351] 多种硫酸酯酶缺乏症成纤维细胞以野生型和突变体 FGE cDNA 的转导

[0352] 为了确认 FGE 缺陷是多种硫酸酯酶缺乏症中合成的硫酸酯酶无活性的原因, 我们在多种硫酸酯酶缺乏症成纤维细胞中应用逆转录病毒基因转移来表达 FGE cDNA。作为对照, 我们转导了没有 cDNA 插入物的逆转录病毒载体。为监测代谢缺陷的补偿 (complementation), ASA、类固醇硫酸酯酶 (STS) 和 N-乙酰半乳糖胺 6-硫酸酯酶 (GalNAc6S) 的活性在选择之前或之后在被转导的成纤维细胞中测量。野生型 FGE 的转导部分恢复了两种多种硫酸酯酶缺乏症 - 细胞系中的三种硫酸酯酶的催化活性 (表 2), 而在第三种多种硫酸酯酶缺乏症细胞系中部分恢复了 STS 的活性。应当指出的是, 对于 ASA 和 GalNAc6S, 恢复只是在对成纤维细胞选择之后的部分恢复, 达到正常活性的 20 到 50%。对于 STS, 活性被发现在选择后恢复到对照成纤维细胞的水平。选择增加了 ASA 和 STS 的活性 50 到 80%, 其与早些时候的观察即 15 到 50% 的成纤维细胞被转导 (Lübke 等, Nat. Gen., 2001, 28 :73-76) 相一致。仅以逆转录病毒载体转导的多种硫酸酯酶缺乏症成纤维细胞中的硫酸酯酶活性 (表 2) 与非转导性多种硫酸酯酶缺乏症成纤维细胞 (未显示) 中的那些结果相当。携带 IVS3+5-8del 突变的 FGE cDNA 的转导无法恢复硫酸酯酶活性 (表 2)。

[0353] 表 2:多种硫酸酯酶缺乏症成纤维细胞通过野生型或突变体 FGE cDNA 转导的补偿
[0354]

| 成纤维细胞 | FGE-插入 | 硫酸酯酶 | | |
|--------------|-------------------------------|------------|---------|-----------------------|
| | | ASA | STS | GalNAc6S ¹ |
| 多种硫酸酯酶缺乏症 3° | - | 1.9 ± 0.2 | < 3 | 56.7 ± 32 |
| | FGE ⁺ | 7.9 | 13.5 | n. d. |
| | FGE ⁺⁺ | 12.2 ± 0.2 | 75.2 | 283 ± 42 |
| | FGE-IVS3+5-8del ⁺ | 1.8 | < 3 | n. d. |
| | FGE-IVS3+5-8del ⁺⁺ | 2.1 | < 3 | 98.5 |
| 多种硫酸酯酶缺乏症 4° | - | 1.1 ± 0.3 | < 3 | n. d. |
| | FGE ⁺ | 4.7 | 17.0 | n. d. |
| 对照成纤维细胞 | | 58 ± 11 | 66 ± 31 | 828 ± 426 |

[0355] ¹ 数值给出了 ASA (mU/mg 细胞蛋白质), STS (μ U/mg 细胞蛋白质), GalNAc6S (μ U/mg 细胞蛋白质) 和 β-己糖胺酶 (U/mg 细胞蛋白质) 之间的比率。对于对照成纤维细胞, 6-11 个细胞系的均值和方差被给出。指明了平行转导的两种培养物的范围的情形针对多种硫酸酯酶缺乏症成纤维细胞给出。

[0356] ° 多种硫酸酯酶缺乏症成纤维细胞的数目指表 1 中的患者的数目

[0357] + 选择前活性测定

[0358] ++ 选择后活性测定

[0359] n. d. 没有测定

[0360] 讨论

[0361] FGE 是内质网的高度保守的糖蛋白

[0362] 来自牛睾丸的 FGE 的纯化得到了源自相同基因的 39.5 和 41.5kDa 的两种多肽。三种 FGE 不同标记形式在三种不同真核生物细胞系中作为单一形式的表达暗示, 在从牛睾丸纯化的 FGE 制剂中观察到的两种形式中的一种可能由纯化过程中的有限蛋白水解所产生。ASA65-80 肽中 Cys69 被丝氨酸的置换, 对于 FGE 通过亲和层析的纯化至关重要。FGE 具有可切断的信号序列, 其介导穿过内质网膜的移位。成熟蛋白的更多部分 (340 个残基中的 275 个) 定义了独有的结构域, 其很可能由三个亚结构域组成 (参见实施例 2), 对于这三个亚结构域, 都没有在具有已知功能的蛋白质中找到同源物。对线性 FGly 修饰基元在新合成的硫酸酯酶多肽中的识别 (Dierks 等, EMBO J., 1999, 18 :2084-2091) 可能是 FGE 亚结构域的功能。催化结构域能以数种方式催化 FGly 形成。已有建议提出, FGE 从半胱氨酸的巯基吸取电子, 并将它们转移到受体。得到的硫醛将自发地水解成 FGly 和 H₂S (Schmidt, B. 等, Cell, 1995, 82 :271-278)。替代性地, FGE 能扮演混合功能的加氧酶 (单加氧酶), 在电子供体如 FADH₂ 的协助下将 O₂ 的一个原子引入到半胱氨酸, 而另一个引入 H₂O。得到的半胱氨酸硫醛水合衍生物将自发地反应成 FGly 和 H₂S。部分纯化的 FGE 制剂的初步实验显示 FGly 形成对分子氧的重要依赖性。这将暗示 FGE 扮演混合功能的加氧酶。亚结构域 3 特别高的保守性和三个完全保守的半胱氨酸残基在其中的存在使得此亚结构域成为催化位点很可能的

候选者。将会有趣的是研究,介导 FGly 基元的识别和电子受体或电子供体的结合的结构元件是否与 FGE 的结构域结构相关。

[0363] 重组 FGE 定位于内质网,这与所建议的其行为位置一致。FGly 残基在新合成的硫酸酯酶中产生,这发生在它们移位到内质网的过程中或短时间之后 (Dierks 等, Proc. Natl. Acad. Sci. U. S. A., 1997, 94 :11963-11968 ;Dierks 等, FEBS Lett., 1998, 423 : 61-65)。FGE 本身不含有 KDEL 类型的 ER- 保留信号。因此它在内质网中的保留可被与其它 ER 蛋白的相互作用所介导。移位 /N- 糖基化机制的成分是这类相互作用的伴侣的有吸引力的候选者。

[0364] 导致多种硫酸酯酶缺乏症的 FGE 中的突变

[0365] 我们已显示,编码 FGE 的基因中的突变导致多种硫酸酯酶缺乏症。FGE 也可与其它成分相互作用,而编码后者的基因中的缺陷能同等有效地导致多种硫酸酯酶缺乏症。在七位多种硫酸酯酶缺乏症患者中,我们确实发现了 FGE 基因中的十种不同突变。通过在亚结构域 3 (三个突变) 或亚结构域 2 (一个突变) 中取代高度保守的残基或 C 端不同长度的截短 (四个突变) 或大的按读码框删除 (两个突变),所有突变对 FGE 蛋白有严重的影响。对于两种多种硫酸酯酶缺乏症 - 细胞系和一种多种硫酸酯酶缺乏症的突变,已有显示,野生型而不是突变体的 FGE cDNA 的转导部分恢复硫酸酯酶的活性。这清晰地鉴定出 FGE 基因作为突变位点以及突变的致病本质。多种硫酸酯酶缺乏症是临床和生化方面都异型。在出生时期存在而发展出脑积水的罕见的新生形式,起初象幼儿异染性脑白质营养不良而后来发展出鱼鳞癣和粘多糖症类似特征的普遍形式,以及较少发生的温和形式 (在此形式中,粘多糖症临床特征流行),已被区分开。生化上其特征是硫酸酯酶的残余活性能被测到,对于大多数病例而言,其在培养的皮肤成纤维细胞中低于对照的 10% (Burch 等, Clin. Genet., 1986, 30 :409-15 ;Basner 等, Pediatr. Res., 1979, 13 :1316-1318)。然而,在某些多种硫酸酯酶缺乏症细胞系中,选出的硫酸酯酶的活性能达到正常范围 (Yutaka 等, Clin. Genet., 1981, 20 :296-303)。此外,残余活性已被报道经历了基于细胞培养条件和未知因素的变化。生化上,多种硫酸酯酶缺乏症被分类为两组。在组 I 中,硫酸酯酶的残余活性低于 15%,包括 ASB 的那些部分。在组 II 中,硫酸酯酶的残余活性更高一些,特别是 ASB 的那些部分可达到高至对照的 50-100% 的值。所有在此报道的患者除了患者 5 都属于组 I,患者 5 属于生化表型的组 II (ASB 活性在对照的范围内)。基于临床标准,患者 1 和 6 是新生型病例,而患者 2-4 和 7 是普遍型,患者 5 是多种硫酸酯酶缺乏症的类粘多糖症形式。

[0366] 表型异质性暗示多种硫酸酯酶缺乏症患者中的不同突变与不同的 FGE 残余活性相关。稳定表达 FGE IVS3+5-8del 的 PT67 细胞上的初步数据表明外显子 3 的按读码框删除完全终止了 FGE 活性。应用最近发展出的高敏感性质谱方法对多种硫酸酯酶缺乏症中的突变、变体 FGE 的生化性质的特征鉴定,及对硫酸酯酶中的 FGly 的残余含量的特征鉴定 (Peng 等, J. Mass Spec., 2003, 38 :80-86) 将为多种硫酸酯酶缺乏症中基因型 - 表型相关性提供更好的理解。

[0367] 实施例 2:

[0368] 人 FGE 基因定义了修饰硫酸酯酶的从原核生物到真核生物都保守的新基因家族

[0369] 生物信息学

[0370] 信号肽和切断位点以 von Heijne 所述的方法 (Nucleic Acids Res., 1986, 14 :

4683),该方法在EMBOSS中的实行(Rice等,Trends inGenetics,2000,16 :276-277),和Nielsen等的方法(蛋白质Engineering,1997,10 :1-6)来描述。N-糖基化位点应用Brunak的算法(Gupta和Brunak,Pac. Symp. Biocomput.,2002,310-22)预测。

[0371] 功能结构域通过搜索PFAM-Hidden-Markov-Models(version 7.8)而被探测(Sonnhammer等,Nucleic Acids Res.,1998,26 :320-322)。来自PFAM DUF323种子的序列从TrEMBL(Bairoch, A. 和Apweiler, R., Nucleic Acids Res.,2000,28 :45-48)得到。多重排列和系统发生树的构建用Clustal W(Thompson, J. 等,NucleicAcids Res.,1994,22 :4673-4680)完成。对于系统发生树的估计,缺口位置被排斥,而多重置换被修正。完成树引导程序(Treebootstrapping)以得到显著的结果。树应用Njplot(Perriere,G. 和Gouy, M., Biochimie,1996,78 :364-369)而可见。比对应用来自EMBOSS的pret-typlot命令而作出曲线。

[0372] 为搜索FGE同源物,National Center for BiotechnologyInformation(NCBI)的数据库NR,NT和EST(Wheeler等,NucleicAcids Res.,2002,20 :13-16)被用BLAST(Altschul等,NucleicAcids Res.,1997,25 :3389-3402)查询。对于蛋白质序列,对当前版本NR数据库应用迭代收敛性Psi-Blast完成搜索,使用期望截止值 10^{-40} 和默认参数。收敛在5次迭代后实现。对于核苷酸序列,搜索以Psi-Tblastn完成:应用NR和人FGE的蛋白质序列作为输入值,用迭代收敛性Psi-Blast建立针对hFGE的分值矩阵。此矩阵被用作对blastall的输入值,以查询核苷酸数据库NT和EST。对于这两个步骤,都用期望截止值 10^{-20} 。

[0373] 蛋白质二级结构预测应用Psipred(Jones,D.,J Mol Biol.,1999,292 :1950-202; McGuffin,L. 等,Bioinformatics,2000,16 :404-405)完成。

[0374] 亚结构域的相似性分数应用EMBOSS的cons算法以默认参数从比对中计算。次比对(meta alignments)通过比对FGE家族亚类的共有序列而产生。基因组位点组织和同线性应用在NCBI(Bethesda,MD)的NCBI人和小鼠基因组资源和Softberry's(Mount Kisco, NY)Human-Mouse-Rat Synteny而测定。细菌基因组序列从NCBI-FTP-服务器下载。NCBI微生物基因组注解被用于获得对细菌FGE基因的基因组位点的总览。

[0375] 结果和讨论

[0376] 人FGE与相关蛋白的基本特征和基元

[0377] 人FGE基因(SEQ ID NOs :1,3)编码预测具有374个残基的FGE蛋白(SEQ ID NO :2)。在残基22-33之间的切断信号(15.29的Heijne值)和在1.7和3.3之间的残基17-29的亲水值(Kyte,J. 和Doolittle,R.,J Mol Biol.,1982,157 :105-132)表明,33个N端残基在ER移位后被切下来。然而应用Nielsen等人的算法(Protein Engineering,1997,10 :1-6),预测信号序列的切断发生在残基34之后。蛋白质在Asn141处具有单一的潜在N-糖基化位点。

[0378] 以FGE蛋白序列对蛋白质基元数据库PFAM(Sonnhammer等人,Nucleic Acids Res.,1998,26 :320-322)的搜索显示人FGE的残基87-367能被分类成具有 7.9×10^{-114} 的高显著期望值的蛋白结构域DUF323(“未知功能的结构域”,PF03781)。确定DUF323的PFAM一种子由25个蛋白质序列组成,其受试者是源自测序数据的假设性蛋白质。为分析人FGE和DUF323间的关系,进行具有DUF323种子序列的FGE的多重比对。基于此,构建系统发生

树并进行自展分析。假设性序列中的四个 (TrEMBL-IDs Q9CK12, Q9I761, 094632 和 Q9Y405) 与种子的其它成员有如此强的差异以至于它们阻止成功的自展分析而不得不在这套序列中除去。图 2 显示经自展的树,其显示人 FGE 和余下的 21 种 DUF323 种子蛋白质之间的关系。此树可用于将这些种子成员亚分成两个类别:与人 FGE 密切相关的同源物,和余下的相关性较小的基因。

[0379] 最顶上的 7 个蛋白质具有在 0.41 和 0.73 之间的对人 FGE 的系统发生距离。它们只含有单一结构域, DUF323。在本组内的同源性延伸到整个氨基酸序列,其较大部分由 DUF323 结构域组成。DUF323 结构域在此组同源物内非常保守,而种子的其它 15 个蛋白与人 FGE 的相关性较小(系统发生距离在 1.14 和 1.93 之间)。它们的 DUF323 结构域与高度保守的第一组的 DUF323- 结构域(参见“FGE 亚结构域和 FGE 基因中的突变”这一节)有明显的差异。这 15 个蛋白质中的大多数是假设的,它们中的 6 个已被进一步研究。它们中的一个,来自 *C. trachomatis* 的丝氨酸 / 苏氨酸激酶 (TrEMBL:084147) 包含除 DUF 323 之外的其它结构域:ATP- 结合结构域和激酶结构域。来自 *R. sphaeroides* (TrEMBL:Q9ALV8) 和假单孢菌属物种 (TrEMBL:052577) 的序列编码蛋白质 NirV,它是与含铜的亚硝酸盐还原酶 nirK 共转录的基因 (Jain, R. 和 Shapleigh, J., *Microbiology*, 2001, 147:2505-2515)。CarC (TrEMBL:Q9XB56) 是涉及来自 *E. carotovora* 的 β -内酰胺抗生素的合成的加氧酶 (McGowan, S. 等人, *Mol Microbiol.*, 1996, 22:415-426; Khaleeli N, T. C., 和 Busby RW, *Biochemistry*, 2000, 39:8666-8673)。XylR (TrEMBL:031397) 和 BH0900 (TrEMBL:Q9KEF2) 是在芽孢杆菌科 (bacillaceae) 和梭菌科 (clostridiaceae) 中涉及戊糖利用的调节的增强子结合蛋白质 (Rodionov, D. 等人, *FEMS Microbiol Lett.*, 2001, 205:305-314)。FGE 和 DUF323 的比较导致建立区分 FGE 家族与具有不同功能的、疏远的、含有 DUF323 的同源物的同源性阈值。后者包括丝氨酸 / 苏氨酸激酶和 XylR (转录增强子) 以及 FGE (产生 FGly 的酶) 和 CarC (加氧酶)。如本文中其它部分所讨论的, FGE 也可作为加氧酶而发挥其半胱氨酸修饰功能,这暗示 DUF323 种子的 FGE 和非 FGE 成员可共享加氧酶功能。

[0380] FGE 的同源物

[0381] 人 FGE 的密切相关的同源物在 DUF323 种子中的存在引导我们在 NCBI 的 NR 数据库中搜索人 FGE 的同源物 (Wheeler 等人, *Nucleic Acids Res.*, 2002, 20:13-16)。搜索的阈值以这样的方式选择,即获得在 DUF323 种子中存在的全部 6 个同源物和其它密切相关的同源物而没有发现其它的种子成员。此搜索导致鉴定出真核生物中的 3 种 FGE 定向进化同源物、原核生物中的 12 种定向进化同源物和人与小鼠中的 2 种平行进化同源物 (表 3)。

[0382] 表 3:真核生物和原核生物中的 FGE 基因家族

[0383]

| SEQ ID NOs : NA, AA [GI] | 物种 | 长度 [AA] | 亚类 |
|--------------------------------|----------------------------|---------|----|
| 1/3, 2 | 智人 (<i>Homo sapiens</i>) | 374 | E1 |

| | | | |
|----------------------|--|------|----|
| 49, 50 [22122361] | 小家鼠 (<i>Mus musculus</i>) | 372f | E1 |
| 51, 52 [20130397] | 黄果蝇 (<i>Drosophila melanogaster</i>) | 336 | E1 |
| 53, 54 [21289310] | 疟蚊 (<i>Anopheles gambiae</i>) | 290 | E1 |
| 47, 48 [26344956] | 小家鼠 | 308 | E2 |
| 45, 46 [24308053] | 智人 | 301 | E2 |
| 55, 56 [21225812] | 天蓝色链霉菌 (<i>Streptomyces coelicolor</i>)A3(2) | 314 | P1 |
| 57, 58 [25028125] | <i>Corynebacterium efficiens</i> YS-314 | 334 | P1 |
| 59, 60 [23108562] | <i>Novosphingobium aromaticivorans</i> | 338 | P2 |
| 61, 62 [13474559] | 根瘤菌 (<i>Mesorhizobium loti</i>) | 372 | P2 |
| 63, 64 [22988809] | <i>Burkholderia fungorum</i> | 416 | P2 |
| 65, 66 [16264068] | 苜蓿中华根瘤菌 (<i>Sinorhizobium meliloti</i>) | 303 | P2 |
| 67, 68 [14518334] | 微颤菌属物种 (<i>Microscilla</i> sp.) | 354 | P2 |
| 69, 70 [26990068] | 恶臭假单孢菌 (<i>Pseudomonas putida</i>) KT2440 | 291 | P2 |
| 71, 72 [22975289] | <i>Ralstonia metallidurans</i> | 259 | P2 |
| 73, 74 [23132010] | 海洋原绿球藻 (<i>Prochlorococcus marinus</i>) | 291 | P2 |

| | | | |
|----------------------|--|-----|----|
| 75, 76 [16125425] | 新月柄杆菌 (<i>Caulobacter crescentus</i>) CB15 | 338 | P2 |
| 77, 78 [15607852] | 结核分枝杆菌 (<i>Mycobacterium tuberculosis</i>)Ht37Rv | 299 | P2 |

[0384] GI-GenBank 蛋白质标识符

[0385] NA- 核酸 AA- 氨基酸

[0386] E1 真核生物定向进化同源物 E2- 真核生物平行进化同源物

[0387] P1 密切相关的原核生物定向进化同源物 P2 其它原核

[0388] 生物定向进化同源物

[0389] f-GenBank 中误预测的蛋白质序列

[0390] 注意小鼠序列 GI 22122361 在 GenBank 中是被预测成编码一种 284 个氨基酸的蛋白质, 虽然 cDNA 序列 NM 145937 编码一种 372 个残基的蛋白质。此误预测是由于忽略了鼠 FGE 基因第一外显子。所有在 NR 数据库中发现的序列都来自更高等的真核生物或原核生物。FGE- 同源物没有从古细菌或植物中被探测到。以甚至更低的阈值在被完全测序的秀丽线虫 (*C. elegans*) 和啤酒糖酵母 (*S. cerevisiae*) 基因组及相关的 ORF 数据库中所进行的搜索没有显示出任何同源物。在 NT 和 EST 核苷酸数据库的真核生物序列中的搜索导致鉴定出 8 种另外的 FGE 定向进化同源的 EST, 它们具有在蛋白质水平上显示高度保守性的 3'-端 cDNA 序列片段, 并且没有在 NR 数据库中列出。这些序列不包含 mRNA 的全部编码部分, 且全都来自更高等的真核生物 (表 4)。

[0391] 表 4 :真核生物中的 FGE 定向进化同源物 EST 片段

[0392]

| SEQ ID NOs : NA [GB] | 物种 |
|----------------------------|-----------------------------------|
| 80 [CA379852] | 虹鳟 (<i>Oncorhynchus mykiss</i>) |
| 81 [AI721440] | 斑马鱼 (<i>Danio rerio</i>) |
| 82 [BJ505402] | 青鳉 (<i>Oryzias latipes</i>) |
| 83 [BJ054666] | 非洲爪蟾 (<i>Xenopus laevis</i>) |
| 84 [AL892419] | <i>Silurana tropicalis</i> |

| | |
|------------------|------------------------------|
| 85 [CA064079] | 加拿大鲑鱼 (<i>Salmo salar</i>) |
| 86 [BF189614] | 野猪 (<i>Sus scrofa</i>) |
| 87 [AV609121] | 黄牛 (<i>Bos taurus</i>) |

[0393] GB-GenBank 登录号 ;NA- 核酸

[0394] 来自NR数据库的编码序列的多重比对和系统发生树的构建(应用ClustalW)允许了对同源物四个亚组的定义:真核生物定向进化同源物(人、小鼠、蚊子和果蝇 FGE),真核生物平行进化同源物(人和小鼠 FGE 平行进化同源物),与 FGE 密切相关的原核生物定向进化同源物(链霉菌 (*Streptomyces*) 和棒状菌 (*Corynebacterium*)),和其它原核生物定向进化同源物(柄细菌 (*Caulobacter*),假单胞菌 (*Pseudomonas*),分支杆菌 (*Mycobacterium*), *Prochlorococcus*, *Mesorhizobium*, 根瘤菌 (*Sinorhizobium*), *Novosphingobium*, *Ralstonia*, 伯克霍尔德氏菌 (*Burkholderia*), 和微颤菌 (*Microscilla*))。真核生物定向进化同源物显示对人 FGE 的 87% (小鼠), 48% (果蝇) 和 47% (疟蚊) 的总体一致性。虽然 FGE 定向进化同源物在原核生物和更高等的真核生物中被发现,但是它们在系统发生上位于啤酒糖酵母和黄果蝇之间的更低等真核生物的全测序基因组中不存在。此外, FGE 同源物在大肠杆菌和 the pufferfish 的全测序基因组中也不存在。

[0395] 如本文其它部分所讨论的,在人和小鼠中发现的 FGE 平行进化同源物可具有较小的产生 FGly 的活性,并对在多种硫酸酯酶缺乏症患者中发现的残余硫酸酯酶活性作出贡献。

[0396] FGE 的亚结构域

[0397] FGE 基因家族的成员具有三个高度保守的部分/结构域(如本文其它部分所述)。除了分开前者的两个非保守序列外,它们在 N- 和 C- 端还具有非保守的延伸。三个保守部分被认为代表了 DUF323 结构域的亚结构域,因为它们被不同长度的非保守部分所间隔开。间隔开亚结构域 1 和 2 的部分的长度在 22 和 29 个残基之间变化,而间隔开亚结构域 2 和 3 的部分的长度在 7 到 38 个氨基酸之间变化。N- 和 C- 端非保守部分显示甚至更强的长度变化(N- 端:0-90 个氨基酸, C- 端:0-28 个氨基酸)。来自 *Ralstonia metallidurans* 的 FGE 基因的序列很有可能不完整,因为它缺乏第一亚结构域。

[0398] 为验证对 DUF323 的亚结构域的定义有道理,我们应用 Psipred 完成了对人 FGE 蛋白的二级结构预测。疏水性的 ER- 信号(残基 1-33)被预测为含有螺旋结构,确认了 von-Heijne 算法的信号预测。N- 端非保守区(氨基酸 34-89)和亚结构域 2 和 3 之间的分隔区(氨基酸 308-327)包含卷绕部分。分隔亚结构域 1 和 2 的区域包含卷绕。氨基酸 65/66 处的 α -螺旋具有低的预测置信度,并且很可能是预测假象。亚结构域边界位于卷绕内,而不打断 α -螺旋或 β -链。第一亚结构域由几个 β -链和一个 α -螺旋组成,第二亚结构域包含两个 β -链和四个 α -螺旋。第三亚结构域在亚结构域的起点和终点具有侧面与折叠片相接的 α -螺旋。简而言之,二级结构与所提出的亚结构域结构一致,因为亚结构

域边界位于卷绕内且亚结构域包含结构性的元件 α -螺旋和 β -链。

[0399] 应当指出的是,没有亚结构域以数据库所列出的序列中的单独的模块存在。在FGE家族的四个亚组的每一种内,亚结构域高度保守,而第三亚结构域显示最高的同源性(表5)。此亚结构域在亚组之间也显示最强的同源性。

[0400] 表5:FGE家族亚结构域的同源性(%相似性)

[0401]

| 亚家族 成员 | | 亚结构域 | | |
|--------|----|------|----|-----|
| | | 1 | 2 | 3 |
| E1 | 4 | 79 | 82 | 100 |
| E2 | 2 | 90 | 94 | 100 |
| P1 | 2 | 70 | 79 | 95 |
| P2 | 10 | 59 | 79 | 80 |

[0402] E1-真核生物定向进化同源物;E2-真核生物平行进化同源物

[0403] P1密切相关的原核生物定向进化同源物;P2其它原核生物定向进化同源物

[0404] FGE家族的第一亚结构域在亚组间显示出最弱的同源性。在真核生物定向进化同源物中,它携带N-糖基化位点:人中的残基Asn 141,小鼠中的Asn 139和果蝇中的Asn 120。在疟蚊中,在与黄果蝇Asn120同源的残基130处没有发现天冬酰胺。然而,两个核苷酸的改变将在疟蚊中创造出一个N-糖基化位点Asn 130。因此,包含残基130的序列需要再次测序。第二亚结构域富含色氨酸,人FGE的129个残基中有12个Trp。这些色氨酸中的十个在FGE家族中保守。

[0405] 亚结构域3的高保守性:真核生物定向进化同源物之间的亚结构域3是100%类似和90%一致。第三亚结构域对蛋白质功能的重要性被这样的观察结果所强调,即此亚结构域是多种硫酸酯酶缺乏症患者中的致病性突变的热点。实施例1中所述的在6位多种硫酸酯酶缺乏症患者中鉴别的9个突变中的7个是位于编码亚结构域3的40个残基的序列中。残基包含4个半胱氨酸,其中3个在原核生物和真核生物定向进化同源物中保守。两种真核生物平行进化同源物显示出与FGE家族其它成员的最小同源性,例如它们缺乏亚结构域3的3个保守半胱氨酸中的2个。在定向进化同源物和平行进化同源物的亚结构域3序列之间保守的特征是起始的RVXXGG(A)S基元(SEQ ID NO:79),一种含有3个精氨酸的七肽(亚结构域共有序列的残基19-25),和末端的GFR基元。对15个不是FGE密切同源物的种子序列的DUF323结构域进行比较,显示出显著的序列差异:15个种子序列具有相对不保守的第一和第二亚结构域,虽然整个亚结构域结构也可见到。当与原核生物FGE家族成员(约60%的相似性)比较时,在FGE家族中非常保守的亚结构域3更短,并具有对真核生物亚结构域3的显著更弱的同源性(约20%的相似性)。因此它们缺乏所有的亚结构域3的保守半胱氨酸残基。仅有的保守特征是起始的RVXXGG(A)S基元(SEQ ID NO:79)和末端的GFR基元。

[0406] 人和鼠FGE基因的基因组组织

[0407] 人FGE基因位于染色体3p26。它含有105kb和9个用于翻译序列的外显子。鼠FGE基因长度为80Kb,位于染色体6E2。鼠FGE基因的9个外显子具有几乎与人外显子同样

的大小(图3)。人和小鼠基因之间的主要差异是外显子9中的3'-UTR的更低的保守性和外显子9的长度,在鼠基因中外显子9的长度多了461bp。小鼠染色体6的区段6E2与人染色体区段3p26是高度同线性的。在朝着端粒的方向上,人和鼠FGE座位的侧面都与编码LMCD1、KIAA0212、ITPR1、AXCAM和IL5RA的基因相接。在着丝粒方向上,两个FGE座位在侧面都与CAV3和OXTR的座位相接。

[0408] 原核生物FGE基因的基因组组织

[0409] 在原核生物中,硫酸酯酶根据在它们的活性中心中转换成FGly的残基而被分类为或者半胱氨酸型或者丝氨酸型的硫酸酯酶(Miech, C. 等人, J Biol Chem., 1998, 273 : 4835-4837 ;Dierks, T. 等人, J Biol Chem., 1998, 273 :25560-25564)。在肺炎克雷伯氏菌(*Klebsiellapneumoniae*)、大肠杆菌(*E. coli*)和鼠疫耶尔森氏菌(*Yersiniapestis*)中,丝氨酸型硫酸酯酶是具有AtsB的操纵子的一部分,其编码含有铁-硫簇的基元的胞质蛋白质,并对从丝氨酸残基产生FGly是至关重要的(Marquardt, C. 等人, J Biol Chem., 2003, 278 :2212-2218 ;Szameit, C. 等人, J Biol Chem., 1999, 274 :15375-15381)。

[0410] 因此有意思的是,检测原核生物FGE基因是否位于作为FGE底物的半胱氨酸型硫酸酯酶的邻近位置。在表3中所示的原核生物FGE基因中,7种具有已被完全测序的基因组,从而允许了对FGE座位的相邻相关效应的分析。实际上,在7个基因组的4个(*C. efficiens* :PID 25028125, 恶臭假单胞菌 :PID 26990068, 新月柄杆菌 :PID 16125425和结核分枝杆菌 :PID 15607852)中,发现半胱氨酸型硫酸酯酶直接与FGE相邻,这与FGE和硫酸酯酶的共转录相一致。在它们中的两种(*C. efficiens*和恶臭假单胞菌)中,FGE和硫酸酯酶甚至具有重叠的ORF,这强烈地指明了它们的共表达。此外,FGE和硫酸酯酶基因在四种原核生物中的基因组相邻相关效应为细菌FGE是功能性的定向进化同源物的假设提供了另外的证据。

[0411] 余下的三种生物含有半胱氨酸型硫酸酯酶(天蓝色链霉菌 :PID24413927, 根瘤菌 :PID 13476324, 苜蓿中华根瘤菌 :PIDs 16262963, 16263377, 15964702),然而,在这些生物中邻近FGE的基因既不含有规范的硫酸酯酶特征(Dierks, T. 等人, J Biol Chem., 1998, 273 :25560-25564)也不含有将会指明它们功能的结构域。因此在这些生物中,FGE和半胱氨酸型硫酸酯酶的表达很可能被反式调节。

[0412] 结论

[0413] 对其缺乏可引起常染色体隐性传递的溶酶体储存性疾病(多种硫酸酯酶缺乏症)的人FGE进行鉴定使得能够定义一个新的基因家族,所述基因家族包含来自原核生物和真核生物的FGE定向进化同源物以及小鼠和人中的FGE平行进化同源物。在全测序的大肠杆菌,啤酒糖酵母,秀丽线虫和*Fugu rubripes*的基因组中没有发现FGE。此外,在原核生物和更高等的真核生物之间存在系统发生上的间隔,在系统发生上位于原核生物和黄果蝇之间的任何物种中没有FGE。然而,这些更低等真核生物中的一部分,例如秀丽线虫,具有半胱氨酸型硫酸酯酶基因。这指明了作用于半胱氨酸型硫酸酯酶的第二产生FGly的系统的存在。此假设被这样的观察结果支持,即没有FGE的大肠杆菌能在半胱氨酸型硫酸酯酶中产生FGly(Dierks, T. 等人, J Biol Chem., 1998, 273 :25560-25564)。

[0414] 实施例3:

[0415] FGE表达引起硫酸酯酶过表达的细胞系中的硫酸酯酶活性的显著

[0416] 增加

[0417] 我们想检测 FGE 对表达 / 过表达硫酸酯酶的细胞的影响。为此目标, 表达艾杜糖醛酸-2-硫酸酯酶 (I2S) 或 N-乙酰半乳糖胺 6-硫酸酯酶 (GALNS) 这些人硫酸酯酶的 HT-1080 细胞用 FGE 表达构建体 pXMG. 1. 3 (表 7 和图 4) 或用对照质粒 pXMG. 1. 2 (反义方向的 FGE, 不能产生功能性 FGE, 表 7) 进行两次重复的转染。在进行一次 24 小时电穿孔后的培养基改变之后 24、48 和 72 小时采集培养基样本。培养基样本通过活性分析来测试各自的硫酸酯酶活性, 并通过艾杜糖醛酸-2-硫酸酯酶或 N-乙酰半乳糖胺 6-硫酸酯酶专一的 ELISA 来评估总硫酸酯酶蛋白水平。

[0418] 表 6. 表达硫酸酯酶的转染的细胞系, 该硫酸酯酶用作转染的底物

[0419]

| 细胞株 | 质粒 | 表达的硫酸酯酶 |
|------|-----------|-----------------|
| 36F | pXFM4A. 1 | N-乙酰半乳糖胺 6-硫酸酯酶 |
| 30C6 | pXI2S6 | 艾杜糖醛酸-2-硫酸酯酶 |

[0420] 表 7. 用于转染表达艾杜糖醛酸 2-硫酸酯酶和 N-乙酰半乳糖胺 6-硫酸酯酶的 HT-1080 细胞的 FGE 和对照质粒

[0421]

| 质粒 | 主要 DNA 序列元件的构造* |
|--------------------------|--|
| pXMG. 1. 3 (FGE 表达) | > 1. 6kb CMV 增强子 / 启动子 > 1. 1kb FGE cDNA > hGH 3' 非翻译序列 < amp < DHFR 盒 < Cneo 盒 (新霉素磷酸转移酶) |
| pXMG. 1. 2 (对照, FGE 反方向) | > 1. 6kb CMV 增强子 / 启动子 < 1. 1kb FGE cDNA < hGH 3' 非翻译序列 < amp < DHFR 盒 < Cneo 盒 (新霉素磷酸转移酶) |

[0422] * > 表示方向 5' 到 3'

[0423] 实验程序

[0424] 材料和方法

[0425] 对产生艾杜糖醛酸 2-硫酸酯酶和 N-乙酰半乳糖胺 6-硫酸酯酶的 HT-1080 细胞的转染

[0426] 采集 HT-1080 细胞, 以得到对于每次电穿孔有 $9-12 \times 10^6$ 的细胞。用两种质粒重复转染两次: 一种是待测试 (FGE) 的, 而另一种作对照; 在此情形中对照质粒包含相对于 CMV 启动子以反向克隆的 FGEcDNA。细胞以约 1000RPM 离心 5 分钟。细胞以 16×10^6 个细胞 / mL 悬浮在 1X PBS 中。将 100 μ g 的质粒 DNA 加到电穿孔小杯的底部, 将 750 μ L 的细胞悬浮液 (12×10^6 个细胞) 加入小杯中的 DNA 溶液中。细胞和 DNA 用塑料移液管轻轻混合, 小心不要产生泡沫。细胞以 450V, 250 μ F (BioRad Gene Pulser) 进行电穿孔。记录时间常数。

[0427] 让电穿孔的细胞静置 10-30 分钟。然后加入 1. 25mL 的 DMEM/10% 小牛血清到每一个小杯, 混合, 并将所有细胞转移到含有 20mL DMEM/10 的新鲜的 T75 烧瓶。24 小时后, 烧瓶中再注入 20mL DMEM/10 以除去死亡的细胞。转染后 48-72 小时, 收集培养基样本, 从完全

一样的 T75 烧瓶采集细胞。

[0428] 培养基制备

[0429] 1L DMEM/10(含有:23ml 的 2mM L- 谷氨酸,115mL 小牛血清)

[0430] 细胞在不含氨甲蝶呤 (MTX) 的培养基中被转染。24 小时后,细胞用含有适量 MTX(36F = 1.0 μ M MTX,30C6 = 0.1M MTX) 的培养基再进行培养。再培养后 24、48 和 72 小时,收获培养基和收集细胞。

[0431] 活性分析

[0432] 艾杜糖醛酸-2-硫酸酯酶 (I2S)

[0433] NAP5 脱盐柱 (Amersham Pharmacia Biotech AB,Uppsala,Sweden) 用透析缓冲液 (5mM 醋酸钠,5mM tris, pH 7.0) 平衡。含有 I2S 的样本被加到柱上,并让其进入柱床。样本以 1mL 的透析缓冲液洗脱。脱盐的样本用反应缓冲液 (5mM 醋酸钠,0.5mg/L BSA,0.1% TritonX-100,pH 4.5) 进一步稀释到约 100ng/mL I2S。将 10 μ L 的每一 I2S 样本加入 96-孔的 Fluormetric Plate (Perkin Elmer,Norwalk,CT) 的顶上一排,并在 37°C 预培育 15 分钟。底物通过将 4- 甲基 - 纈形基硫酸酯 (Fluka,Buchs,Switzerland) 溶解在底物缓冲液 (5mM 醋酸钠,0.5mg/mL BSA,pH 4.5) 中至最终浓度为 1.5mg/mL 而制备。100 μ L 的底物被加至含有 I2S 样本的每一孔中,并将平板在 37°C 于黑暗中培育 1 小时。培育后,将 190 μ L 的终止缓冲液 (332.5mM 甘氨酸,207.5mM 碳酸钠,pH 10.7) 加至含有样本的每一孔中。4- 甲基伞形酮 (4-MUF, Sigma, St. Louis, MO) 储液在试剂级水中制备成最终浓度为 1 μ M 以作为产物标准。150 μ L 的 1 μ M 4-MUF 储液和 150 μ L 终止缓冲液加至平板中的顶上一排。150 μ L 的终止缓冲液加至 96 孔平板中剩下的每一孔中。从平板的每一列的顶上一排至最后一排制备两倍的系列稀释液。平板在 Fusion Universal Microplate Analyzer (Packard,Meriden,CT) 上以 330nm 的激发过滤波长和 440nm 的发射过滤波长进行阅读。得到 4-MUF 储液的微摩尔数量对荧光的标准曲线,而未知样本具有由此曲线推断的荧光。结果以“单位/mL”(一个单位的活性等于在 37°C 下每分钟产生 1 μ mole 的 4-MUF) 显示。

[0434] N- 乙酰半乳糖胺 6- 硫酸酯酶 (GALNS)

[0435] GALNS 活性分析利用了荧光底物 4- 甲基纈形基 - β -D- 吡喃半乳糖苷 -6- 硫酸酯 (Toronto Research Chemicals Inc., 目录号 M33448)。分析包括两个步骤。在第一步骤中,75 μ L 的在反应缓冲液 (0.1M 醋酸钠,0.1M 氯化钠, pH 4.3) 中制备的 1.3mM 底物与 10 μ L 的培养基 / 蛋白质样本或其相应的稀释液在 37°C 培育 4 小时。反应通过加入 5 μ L 的 2M 磷酸二氢钠以抑制 GALNS 活性而被终止。加入约 500U 的来自米曲霉 (*Aspergillus oryzae*) 的 β - 半乳糖苷酶 (Sigma, 目录号 G5160) 后,反应混合物在 37°C 再培育 1 小时以释放底物的荧光部分。第二反应通过加入 910 μ L 的终止溶液 (1% 甘氨酸,1% 碳酸钠, pH 10.7) 而被终止。所得混合物的荧光通过使用 359nm 的测量波长和 445nm 的参考波长,并以 4- 甲基伞形酮 (钠盐, Sigma, 目录号 M1508) 用作参考标准来测量。一个活性单位对应每小时释放的 4- 甲基伞形酮的 nmole 数量。

[0436] 免疫分析 (ELISA)

[0437] 艾杜糖醛酸-2-硫酸酯酶 (I2S)

[0438] 96 孔平底平板用在 50nM 碳酸氢钠 (pH 9.6) 中稀释到 10 μ g/mL 的小鼠单克隆抗-I2S 抗体于 37°C 包被 1 小时。抗纯化的、重组产生的全长人 I2S 多肽的小鼠单克隆

抗-I2S 抗体在与 Maine Biotechnology Services, Inc. (Portland, ME) 的合同下应用标准杂交瘤产生技术来制备。平板用含有 0.1% Tween-20 的 1X PBS 清洗 3 次和用清洗缓冲液中的 2% BSA 在 37°C 封闭 1 小时。含 2% BSA 的清洗缓冲液用于稀释样本和标准物。I2S 标准物被稀释到 100ng/mL 至 1.56ng/mL, 并如此进行应用。除去封闭缓冲液后, 样本和标准被施加到平板上, 并在 37°C 培育 1 小时。将探测抗体, 即缀合辣根过氧化物酶的小鼠抗-I2S 抗体, 在含 2% BSA 的清洗缓冲液中稀释到 0.15 μ g/mL。平板清洗 3 次, 将探测抗体加至平板, 并在 37°C 培育 30 分钟。为使平板显色, 制备 TMB 底物 (Bio-Rad, Hercules, CA)。平板清洗 3 次, 将 100 μ L 的底物加至每一孔中, 并在 37°C 培育 15 分钟。反应用 2N 硫酸 (100 μ L/孔) 终止, 并将平板在微量滴定平板读取器上于 450nm, 并用 655nm 作为参考波长进行阅读。

[0439] N-乙酰半乳糖胺 6-硫酸酯酶 (GALNS)

[0440] 两种小鼠单克隆抗-GALNS 抗体提供了 GALNS ELISA 的基础。抗纯化的、重组产生的全长人 GALNS 多肽的小鼠单克隆抗-GALNS 抗体也在与 Maine Biotechnology Services, Inc. (Portland, ME) 的合同下应用标准杂交瘤产生技术来制备。为捕获 GALNS, 第一抗体被用于在包被缓冲液 (50mM 碳酸氢钠, pH 9.6) 中对 F96 MaxiSorpNunc-Immuno Plate (Nalge Nunc, 目录号 442404) 进行包被。在 37°C 培育 1 小时和用清洗缓冲液清洗后, 平板用封闭缓冲液 (PBS, 0.05% Tween-20, 2% BSA) 在 37°C 封闭 1 小时。然后将实验和对照样本与 GALNS 标准物一起装载到平板上, 并进一步在 37°C 培育 1 小时。在用清洗缓冲液清洗后, 抗体, 即缀合 HRP 的探测抗体被加到封闭缓冲液中, 之后在 37°C 培育 30 分钟。再次清洗平板后, 加入 Bio-Rad TMB 底物试剂, 并培育 15 分钟。然后加入 2N 硫酸以终止反应, 结果通过应用 Molecular Device 平板读取器在 450nm 波长以分光光度测量法进行评分。

[0441] 讨论

[0442] FGE 对硫酸酯酶活性的影响

[0443] GALNS

[0444] 相对于对照水平观察到大约 50 倍的总 GALNS 活性的增长 (图 5)。此增长的活性的水平在全部三个培养基取样时间点都被观察到。此外, GALNS 活性随时间线性积累, 在 24 和 48 小时之间具有 4 倍的增长, 在 48 和 72 小时之间的时间点有 2 倍的增长。

[0445] I2S

[0446] 虽然是更小的绝对数量级, 但对于总 I2S 活性, 也观察到类似效应, 其中相对于对照水平观察到大约 5 倍的总 I2S 活性增长。此增长的活性的水平在实验延续过程中维持不变。I2S 活性在培养基中随时间线性积累, 类似于用 GALNS 所见到的结果 (在 24 和 48 小时之间为 2.3 倍, 在 48 和 72 小时之间为 1.8 倍)。

[0447] FGE 对硫酸酯酶比活性的影响

[0448] GALNS

[0449] FGE 在 36F 细胞中的表达以相对于对照水平的 40-60 倍增强了 GALNS 的表观比活性 (由 ELISA 测得的酶活性对总酶的比率) (图 6)。此比活性的增长在研究中的三个时间点都维持不变, 并似乎在转染后积累的三天中一直增长。

[0450] I2S

[0451] 用 I2S 见到类似效应, 其中相对于对照值 (0.5-0.7U/mg) 观察到比活性的 6-7 倍的增长 (3-5U/mg)。

[0452] GALNS(图7)和I2S的ELISA值都没有受到FGE的转染的明显影响。这表明FGE的表达没有影响到参与硫酸酯酶产生的翻译和分泌途径。

[0453] 总之,两种硫酸酯酶的所有这些结果表明,FGE表达在过表达GALNS和I2S的细胞系中显著地增长了硫酸酯酶的比活性。

[0454] FGE(SUMF1)和其它硫酸酯酶基因的共表达

[0455] 为测试FGE(SUMF1)在正常细胞中对另外的硫酸酯酶活性的效应,我们在共转染或没有共转染FGE(SUMF1)cDNA的不同细胞系中过表达ARSA(SEQ ID NO:14),ARSC(SEQ ID NO:18)和ARSE(SEQ ID NO:22)cDNA,并测量硫酸酯酶活性。硫酸酯酶cDNA在Cos-7细胞中的过表达引起硫酸酯酶活性的适度增长,而当硫酸酯酶基因和FGE(SUMF1)基因共表达时,观察到惊人的协同性增长(20到50倍)。虽然更弱但类似的效应在三种另外的细胞系即HepG2、LE293和U20S中被观察到。多重硫酸酯酶cDNA的同时过表达引起每一种专一硫酸酯酶活性相对于单一硫酸酯酶过表达时更少的增长,表明存在不同硫酸酯酶对修饰机制的竞争。

[0456] 为测试FGE(SUMF1)基因在进化过程中的功能保守性,我们在共转染或没有共转染多种硫酸酯酶缺乏症cDNA的多种细胞系中过表达ARSA、ARSC和ARSE cDNA,并测量硫酸酯酶活性。鼠和果蝇的FGE(SUMF1)基因对全部三种人硫酸酯酶都有活性,果蝇FGE(SUMF1)效率更低。这些数据证明FGE(SUMF1)在进化过程中的高度功能保守性,暗示了对细胞功能和生存的显著的生物重要性。通过应用FGE2(SUMF2)基因观察到虽然弱得多但类似和一致的效应,暗示由此基因编码的蛋白质也具有硫酸酯酶修饰活性。这些数据证明FGE(SUMF1)所编码的蛋白质的数量是对硫酸酯酶活性的限制性因素,这是一个对于大规模生产用于酶替代治疗的活性硫酸酯酶具有重要暗示的发现。

[0457] 实施例4:

[0458] 通过应用微细胞介导的染色体转移的功能互补来鉴定多种硫酸酯酶缺乏症中突变的基因

[0459] 在应用微细胞介导的染色体转移以功能互补反式进行的单独的实验中,我们证实了多种硫酸酯酶缺乏症中突变的基因是FGE。我们的发现提供了对于影响某一在关系较疏远的生物体中的整个蛋白质家族的新生物机制的进一步见解。除了鉴别罕见遗传病的分子基础之外,我们的数据进一步证实了FGE基因产物对硫酸酯酶活性的有力的增强效应。后一发现对于至少8种由硫酸酯酶缺乏引起的人类疾病的治疗具有直接的临床意义。

[0460] 染色体3p26的多种硫酸酯酶缺乏症基因图谱

[0461] 为鉴别多种硫酸酯酶缺乏症中突变的基因的染色体位置,我们尝试以通过微细胞介导的染色体转移的功能互补来补救缺陷性的硫酸酯酶。含有单独的用显性可选择标记HyTK标记的正常人染色体的一组人/小鼠杂交细胞系,被用作人染色体供体的来源,并与来自多种硫酸酯酶缺乏症患者的永生化细胞系融合。所有22条人常染色体被一条条地转移到患者细胞系,用潮霉素选出杂交体。在22个转移实验的每一个中,挑选出大约25个存活的克隆。这些克隆单独生长,并进行采集以用于随后的酶测试。对于大约440个(20×22)克隆的每一个,测试芳基硫酸酯酶A(ARSA)(SEQ ID NO:15),芳基硫酸酯酶B(ARSB)(SEQ ID NO:17)和芳基硫酸酯酶C(ARSC)(SEQ ID NO:19)活性。此分析清晰地表明源自染色体3转移的数个克隆的硫酸酯酶活性比其它所有克隆都显著地高。当分析源自染色体3转移的

每一单独克隆的活性时,观察到惊人的可变性。为核实是否每一克隆都具有来自供体细胞系的完整的人染色体 3,我们应用一组 23 个染色体 3 多态性遗传标记,它们沿染色体的长度均匀分布,并于先前基于在供体和患者细胞系之间具有不同等位基因而挑选出来。这允许我们检测供体染色体的存在和鉴别由于偶然染色体断裂而造成的特异区域可能的丢失。每一具有高酶活性的克隆保持了完整的来自供体细胞系的染色体 3,然而根据来自供体细胞系的染色体 3 等位基因的缺失,低活性的克隆似乎已没有了完整的染色体。后者的克隆很可能保持了含有可选择标记基因的供体染色体的小区域,这些标记基因使得这些克隆能够在含有潮霉素的培养基中存活。这些数据表明正常的人染色体 3 能对在多种硫酸酯酶缺乏症患者细胞系中所观察到的缺陷进行补充。

[0462] 为确定含有对于补充活性起着作用的基因的特异染色体区域,我们应用被发现已丢失染色体的多个部分的 Neo^r- 标记的染色体 3 杂合体。此外,我们进行了 HyTK^r- 标记的人染色体 3 的经辐射微细胞介导的染色体转移。通过使用一组 31 个跨越整个染色体的多态性微卫星标记,对 115 个染色体 3 被辐射的杂合体进行硫酸酯酶活性测试和基因型的确定。显示高酶活性的所有克隆似乎已保存了染色体 3p26。应用来自此区域的另外标记的更高分辨率的分析绘制出了在标记 D3S3630 和 D3S2397 之间的补充基因的推定位置。

[0463] 多种硫酸酯酶缺乏症中突变的基因的鉴别

[0464] 我们研究了来自涉及多种硫酸酯酶缺乏症患者中的突变的 3p26 基因组区域的基因。每一包括拼接点的外显子进行 PCR 扩增,和通过直接测序进行分析。突变分析在 12 位不相关的被影响的个体中完成;5 位是前述的 MSD 患者,而 7 位是未发表的病例。从我们的多种硫酸酯酶缺乏症群体中,在已表达序列标志 (EST) AK075459 (SEQ ID NOs :4,5) 中鉴定了数个突变,该已表达序列标志对应于未知功能的基因,这强烈暗示了这是涉及多种硫酸酯酶缺乏症的基因。发现每一个突变在 100 个对照个体中不存在,因此排除了序列多态性的存在。另外的确证性突变分析是对于反转录的患者 RNA 来进行的,特别是在基因组 DNA 分析显示突变存在于拼接位点或其附近而有可能影响拼接的那些情况下。移码、无义、拼接和错义突变也被鉴别,暗示此病由于功能机制的丧失而引起,正如对隐性疾病所预言的。这也与以下的观察相一致,即几乎所有的错义突变影响在整个进化中高度保守的氨基酸(参见下面)。

[0465] 表 8:另外的被鉴别的多种硫酸酯酶缺乏症突变

[0466]

| 病例 | 文献 | 表型 | 外显子 | 核苷酸改变 | 氨基酸改变 |
|-----------|---------------------|-------|-----|--------------------|---------|
| 1. BA426 | Conary et al, 1988 | 中度 | 3 | 463T>C | S155P |
| | | | 3 | 463T>C | S155P |
| 2. BA428 | Burch et al, 1986 | 严重初生 | 5 | 661delG | 移码 |
| 3. BA431 | Zenger et al, 1989 | 中度 | 1 | 2T>G | M1R |
| | | | 2 | 276delC | 移码 |
| 4. BA799 | Burk et al, 1981 | 轻微-中度 | 3 | 463T>C | S155P |
| | | | 3 | 463T>C | S155P |
| 5. BA806 | 未发表 | 严重初生 | 9 | 1045T>C | R349W |
| 6. BA807 | Schmidt et al, 1995 | 未知 | 3 | c519+4delGTAA ex 3 | 缺失 |
| | | | 9 | 1076C>A | S359X |
| 7. BA809 | Couchot et al, 1974 | 轻微-中度 | 1 | 1A>G | M1V |
| | | | 9 | 1042G>C | A348P |
| 8. BA810 | 未发表 | 严重 | 8 | 1006T>C | C336R |
| | | | 9 | 1046G>A | R349Q |
| 9. BA811 | 未发表 | 严重初生 | 3 | c519+4delGTAA ex 3 | 缺失 |
| | | | 8 | 979C>T | R327X |
| 10. BA815 | 未发表 | 中度 | 5 | c.603-6delC | ex 6 缺失 |
| | | | 6 | 836C>T | A279V |
| 11. BA919 | 未发表 | 轻微-中度 | 9 | 1033C>T | R345C |
| | | | 9 | 1033C>T | R345C |
| 12. BA920 | 未发表 | 中度 | 5 | 653G>A | C218Y |
| | | | 9 | 1033C>T | R345C |

[0467] 在每一位被测试的多种硫酸酯酶缺乏症患者中鉴定突变,因此排除了座位点异质性。在所鉴定的突变的类型和患者中所报道的表型的严重程度之间没有观察到明显的相关性,暗示临床上的可变性不是由等位基因异质性引起的。在三种情况下,发现不同患者(表6中的病例1和4,病例6和9,和病例11和12)携带相同突变。这些患者中的两位(病例11和12)来自 Sicily 中的同一城镇,这暗示单元型分析(haplotype analysis)所真正确证的建立者效应的存在。令人惊讶的是,发现大多数患者是双重杂合子,携带不同的等位突变,而只有一小部分是纯合子。虽然与父母所报告的不存在近亲关系相一致,但这是关于非常罕见的隐性疾病例如多种硫酸酯酶缺乏症的无法预料的发现。

[0468] FGE 基因和蛋白质

[0469] 人 FGE (也在本文中可互换性地用作 SUMF1) cDNA (SEQ ID NO :1) 的共有 cDNA 序列组装自数个已表达序列标志 (EST) 克隆,并部分组装自对应的基因组序列。此基因包含 9 个外显子,并跨越大约 105kb (参见实施例 1)。序列比较也鉴别了位于人染色体 7 上的 FGE 基因平行进化同源物的存在,我们将其指定为 FGE2 (也在本文中可互换性地用作 SUMF2) (SEQ ID NOs :45, 46)。

[0470] 硫酸酯酶缺乏的功能补充

[0471] 来自两位多种硫酸酯酶缺乏症患者（表 8 中的病例 1 和 12, 在他们中我们鉴别了 FGE (SUMF1) 基因的突变（细胞系 BA426 和 BA920））的成纤维细胞用含有 FGE (SUMF1) cDNA 的野生型和两个突变形式 (R327X 和 $\Delta ex3$) 的 HSV 病毒感染。ARSA、ARSB 和 ARSC 活性在感染后 72 小时进行测试。野生型 FGE (SUMF1) cDNA 的表达导致全部三种活性的功能补充, 而突变体 FGE (SUMF1) cDNA 则没有（表 9）。这些数据提供了对鉴别 FGE (SUMF1) 作为多种硫酸酯酶缺乏症基因的决定性的证据, 且它们证明了患者中发现的突变的相关性。疾病相关的突变引起硫酸酯酶缺乏, 因此证明 FGE (SUMF1) 是硫酸酯酶活性的必需因素。

[0472] 表 9 :硫酸酯酶缺乏的功能补充

[0473]

| 受体 MSD 细胞系 | 构建体 | ARSA ⁽¹⁾ | ARSB ⁽¹⁾ | ARSC ⁽¹⁾ |
|---------------|---------------------|---------------------|---------------------|---------------------|
| BA426 | HSV 扩增子 | 24.0 | 22.5 | 0.15 |
| | SUMF1- $\Delta ex3$ | 42.0 | 23.8 | 0.29 |
| | SUMF1-R327X | 33.6 | 24.2 | 0.16 |
| | SUMF1 | 119.5 (4.9 x) | 37.8 (1.7 x) | 0.62(4.1 x) |
| BA920 | HSV 扩增子 | 16.6 | 11.3 | 0.15 |
| | SUMF1- $\Delta ex3$ | 17.2 | 14.4 | 0.07 |
| | SUMF1-R327X | 36.0 | 13.5 | 0.13 |
| | SUMF1 | 66.5 (4.0 x) | 21.6 (1.9 x) | 0.42(2.8 x) |
| 对照范围 | | 123.7-394.6 | 50.6-60.7 | 1.80-1.58 |

[0474] ⁽¹⁾ 所有酶活性均以释放的 4- 甲基伞形酮的 nmole 数 \cdot mg 蛋白质⁻¹ \cdot 3hrs 表示

[0475] 多种硫酸酯酶缺乏症细胞系 BA426 和 BA920 单独用 HSV 扩增子感染, 和用携带突变体或野生型 SUMF1 cDNA 的构建体感染。在用野生型 SUMF1 基因感染的成纤维细胞中的单一的芳基硫酸酯酶活性, 相对于只用载体感染的细胞的活性的增加在括号中指出。标明了在未感染的对照成纤维细胞中测得的活性。

[0476] 多种硫酸酯酶缺乏症的分子基础

[0477] 基于这样的假设即疾病基因应能补偿患者细胞系中的酶缺乏, 我们完成了微细胞介导的染色体转移, 转移到来自多种硫酸酯酶缺乏症患者的永生化的细胞系。此技术已成功用于其预期功能能在细胞系中评估（例如通过测量酶活性或通过探测形态学特征）的基因的鉴别。为说明酶活性随机变化性的问题, 我们在互补分析中测量了三种不同硫酸酯酶 (ARSA, ARSB 和 ARSC) 的活性。染色体转移的结果清晰地指明了补充性基因在染色体 3 上的位置。亚区域的图谱绘制通过产生染色体 3 的放射杂合组而实现。单独的杂交体克隆同时在基因组水平（通过在供体和受体细胞系之间显示不同等位基因的 31 个微卫星标记的分型）和功能水平（通过测试硫酸酯酶活性）被表征。对 130 个这类杂交体的分析导致对染色体 3p26 补充区域的图谱绘制。

[0478] 一旦关键的基因组区域被确定, FGE (SUMF1) 基因也通过患者 DNA 中的突变分析而鉴别。突变在所有测试患者中被发现, 显示出单一基因涉及多种硫酸酯酶缺乏症。所发现

的突变是不同的类型,其中的主要部分(例如拼接位点,起始位点,无义,移码)估计会导致所编码蛋白质的功能丧失,如对隐性疾病所预计的。大多数错义突变影响对应于进化过程中高度保守的氨基酸的密码子,暗示这些突变也导致功能的丧失。在突变类型和表现型的严重程度之间没有相关性,表明后者是由于不相关的因素造成的。对于罕见的遗传疾病出乎意料的是,许多患者被发现是杂合子,携带两种不同的突变。然而,对源自 Sicily 中的一个小镇的一个突变的奠基者效应被鉴别。

[0479] FGE(SUMF1) 基因功能

[0480] 对 FGE(SUMF1) 基因作为“互补因子”的鉴别,通过插入病毒载体中的外源性 FGE(SUMF1) cDNA 在两种不同患者细胞系的表达挽救了四种不同硫酸酯酶的酶性缺乏这一点而肯定地被证明。在每一病例中,所有硫酸酯酶活性的一致的和部分的(相对于以空载体转染的对照患者细胞系的)恢复被观察到。平均地,酶活性的增加范围在 1.7 到 4.9 倍之间,达到在正常细胞系中所观察到的大约一半的水平。酶活性与在每一实验中所用的病毒颗粒的数目相关,也与(通过标记蛋白质(GFP)分析所测试的)感染效率相关。在相同实验中,含有携带在患者中发现的两种突变(R327X 和 $\Delta ex3$)的 FGE(SUMF1) cDNA 的载体被应用,而没有显著的酶活性增加被观察到,因此证明这些突变的功能相关性。

[0481] 如本文中其它部分所提到的, Schmidt 等首先发现硫酸酯酶经历高度保守的半胱氨酸到 C_α-甲酰甘氨酸的翻译后修饰,其被发现位于大多数硫酸酯酶的活性位点。它们也显示出,此修饰在多种硫酸酯酶缺乏症中有缺陷(Schmidt, B. 等, Cell, 1995, 82: 271-278)。我们的突变性和功能性的数据为以下这点提供了有力的证据,即 FGE(SUMF1) 对此修饰负责。

[0482] FGE(SUMF1) 基因在所有被分析的远距离相关物种(从细菌到人)中显示非常高程度的序列保守性。我们提供了证据,即人 FGE(SUMF1) 基因的果蝇同源物能激活过表达的人硫酸酯酶,证明所观察的远距离相关物种的 FGE(SUMF1) 基因的高水平的序列类似性与惊人的功能保守性相关。值得注意的例外是酵母,其似乎缺失了 FGE(SUMF1) 基因以及任何的硫酸酯酶编码基因,说明硫酸酯酶功能不为此生物所需,也暗示对 FGE(SUMF1) 和硫酸酯酶基因的进化的互惠性影响的存在。

[0483] 有趣的是,在所有被分析的脊椎动物包括人的基因组中存在两个同源性的基因 FGE(SUMF1) 和 FGE2(SUMF2)。如系统发生树上明显的, FGE2(SUMF2) 基因似乎已独立地从 FGE(SUMF1) 基因进化。在我们的分析中, FGE2(SUMF2) 基因也能激活硫酸酯酶,然而它以相对于 FGE(SUMF1) 基因效率低很多的方式完成。这可解释在多种硫酸酯酶缺乏症患者中发现的残余硫酸酯酶活性,并暗示完全的硫酸酯酶缺乏将是致命的。此时,我们不能排除 FGE2(SUMF2) 基因具有另外而仍然未知的功能的可能性。

[0484] 在因硫酸酯酶缺乏导致的疾病的治疗方面的影响

[0485] 相对于只过表达单一硫酸酯酶的细胞,硫酸酯酶活性的强增长(高至 50 倍)在 FGE(SUMF1) cDNA 与 ARSA, ARSC 或者 ARSE cDNAs 一起过表达的细胞中被观察到。在所有细胞系中,显著的协同效应被发现,表明 FGE(SUMF1) 是硫酸酯酶活性的限制性因子。然而,可变性在不同硫酸酯酶中被观察到,可能是由于 FGE(SUMF1)-编码蛋白质与多种硫酸酯酶的不同亲和性。可变性也在可能具有内源性甲酰甘氨酸生成酶的不同水平的不同细胞系之间观察到。与这些观察相一致,我们发现多种硫酸酯酶缺乏症基因的表达在不同组织间变化,

在肾和肝中具有显著的高水平。这可具有重要的意义,因为具有 FGE (SUMF1) 基因低表达水平的组织不太能够有效修饰外源性递送的硫酸酯酶蛋白质(参见下面)。这些数据合在一起暗示着 FGE (SUMF1) 基因的功能已进化到实现双重调节系统,每一硫酸酯酶被两个机制同时控制,一个是负责每一结构性硫酸酯酶基因的 mRNA 水平的单独机制,一个是所有硫酸酯酶共享的普遍机制。此外, FGE2 (SUMF2) 提供了对硫酸酯酶修饰的部分冗余。

[0486] 这些数据对将在酶替换治疗中应用的活性硫酸酯酶的大量生产具有深远的意义。酶替换研究已在硫酸酯酶缺乏的动物模型上报道,例如,粘多糖病 VI 的猫科动物模型,并被证明是在防止和治愈数个症状中有效。针对两种源于硫酸酯酶缺乏的先天疾病 MPSII (Hunter 综合症) 和 MPSVI (Maroteaux-Lamy 综合症) 的人的治疗性试验目前正在进行,并将很快扩展到大量的患者。

[0487] 实施例 5:

[0488] 以 FGE- 激活的 GALNS 对 Morquio 病 MPS IVA 的酶替换疗法

[0489] Morquio 患者中主要的骨病理学原因是硫酸角质素 (KS) 在骺盘(生长平板)软骨细胞中因缺乏溶酶体硫酸酯酶 (GALNS) 而积累。体内研究的基本目标是确定是否静脉内 (IV) 施用的 FGE- 激活的 GALNS 能渗透生长平板的软骨细胞以及其它正常小鼠中的合适的细胞类型。尽管一般缺乏骨畸形, GALNS 缺乏的小鼠模型 (Morquio Knock-In-MKI, S. Tomatsu, St. Louis University, MO) 也被用于证明重复施用的 FGE- 激活的 GALNS 的体内生化活性。小鼠模型中的骨病理的缺乏反映了这样的事实,即骨 KS 在啮齿动物中或剧烈减少或者缺乏 (Venn G, & Mason RM., Biochem J., 1985, 228 :443-450)。然而,这些小鼠的确显示了 GAG 和其它细胞异常在多种器官和组织中的可探测的积累。因此,研究的总体目标是证明 FGE- 激活的 GALNS 渗透进了生长平板(生物分布研究)和显示涉及积累的 GAG 在受影响组织中的清除的功能性 GALNS 酶活性(药物动力学研究)。

[0490] 这些研究的结果证明 IV 注射的 FGE- 激活的 GALNS 被生长平板的软骨细胞内化,虽然相对于其它组织是相对较低的水平。此外, FGE- 激活的 GALNS 在 MKI 小鼠中持续 16 周的注射有效清除了积累的 GAG,也减少了在所有被检测的软组织中溶酶体生物标记染色。总之,实验成功证明了 GALNS 到生长平板软骨细胞的转运,也证明了关于 GAG 在多种组织中的清除方面的生化活性。

[0491] 生物分布研究

[0492] 四周大的 ICR(正常)小鼠被给予单次 5mg/kg FGE- 激活的 GALNS 的 IV 注射。肝,股骨,心脏,肾和脾在注射后两小时被收集,并为组织学检测作准备。抗-人 GALNS 的单克隆抗体被用于探测注射的 GALNS 在多种组织中的存在。与载体对照相比, GALNS 在所有被检测的组织中被探测。此外,应用辣根过氧化物酶报告系统, GALNS 很容易地在所有被检测的组织中被观察到,但骨例外。对 GALNS 在生长平板中吸收的证明需要更灵敏的荧光素-异硫氰酸酯 (FITC) 报告系统的应用,说明了虽然 GALNS 渗透进生长平板,但相对于软组织的细胞,它不那么容易渗透至生长平板软骨细胞。尽管需要更灵敏的荧光探测方法,与载体对照相比, GALNS 到骨生长平板软骨细胞的转运在所有被检测的生长平板切片中被观察到。

[0493] MKI 小鼠中的药物动力学研究

[0494] 四周大的 MKI 或野生型小鼠每周性地被给予 IV 注射(每组 n = 8)直到 20 周龄。每一次每周性的注射由 2mg/kg FGE- 激活的 GALNS 或者载体对照(对野生型小鼠没有注

射) 所组成。所有小鼠在 20 周龄时被杀死用于组织学检测,并用下面的方法染色:苏木精和曙红用于细胞形态学,爱茜蓝用于 GAG 的探测。

[0495] 积累的 GAG 的清除通过在所有被检测的软组织(肝、心脏、肾和脾)中减少的或缺乏的爱茜蓝染色而证明。这只在 GALNS 注射的小鼠中被观察到,虽然 MKI 小鼠中的生长平板如正常骨形态学所证实的是正常发挥功能,但存在被观察到的更细微的细胞异常(包括没有明显病理效应的软骨细胞空泡化)。生长平板肥大和增生区域的空泡化的软骨细胞没有受 GALNS 施用影响。这与生长平板的钙化区中的软骨细胞相反,在那种情形中,空泡化的减弱在 GALNS 注射的小鼠中被观察到。一般地,由于已知的 KS 在小鼠的生长平板中的缺乏,软骨细胞的空泡化和推测的非-KS GAG 在 MKI 小鼠的生长平板中的积累是令人惊讶和出乎意料的。这些特别的观察结果很可能反映了这样的事实,即在敲入(knock-in)小鼠中,存在突变体 GALNS 的高水平(与敲除小鼠相反,其中不存在残余的突变体 GALNS,不存在生长平板软骨细胞空泡化,也不存在 GAG 积累——Tomatsu S. 等, Human Molecular Genetics, 2003, 12 :3349-3358)。生长平板中的空泡化现象也许表明对表达突变体 GALNS 的细胞的亚类的第二效应。但是,持续 16 周的酶注射有力证明了多种组织的 FGE- 激活的 GALNS 运送和体内酶活性的事实。

[0496] 附图详述

[0497] 图 1 :P23 在来自牛睾丸微粒体的可溶性抽提物缺乏(A)或存在(B)条件下培育后的 MALDI-TOF 质谱

[0498] 6pmol 的 P23 在标准条件下 37°C 培育 10 分钟,其中缺乏或存在 1 μ l 的微粒体抽提物。按实验程序所述为 MALDI-TOF 质谱制备样本。P23(2526. 28) 和其 FGly 衍生物(2508. 29) 的单同位素质量 MH^+ 被指出。

[0499] 图 2 :源自人 FGE 和 PFAM-DUF323 种子的 21 种蛋白质的比对的系统发生树

[0500] 分支处的数目指出了系统发生距离。蛋白质通过它们的 TrEMBL ID 数目和物种名称被指定。hFGE- 人 FGE。右上部:系统发生距离的标尺。星号表示基因已进一步地被研究。顶部的 7 种基因是 FGE 基因家族的部分。

[0501] 图 3 :人和鼠 FGE 基因位点的组织

[0502] 外显子被显示成以暗盒(人位点)和亮盒(鼠位点)来按比例表示。右下方角落的横杆显示标尺。外显子之间的线显示内含子(没有按比例)。内含子线上的数字表示以 kb 表示的内含子的大小。

[0503] 图 4 :显示 FGE 表达质粒 pXMG. 1. 3 图谱的图

[0504] 图 5 :描述在以 FGE 表达质粒瞬时转染的 36F 细胞中的 N- 乙酰半乳糖胺 6- 硫酸酯酶活性的柱状图

[0505] 在不含氨甲蝶呤(MTX)的培养基中以带有反向的 FGE cDNA 的对照质粒 pXMG. 1. 2, 或者 FGE 表达质粒 pXMG. 1. 3 转染细胞。24 小时后,细胞被含有 1.0 μ M MTX 的培养基再培养。再培养 24, 48 和 72 小时后,采集培养基并收集细胞。N- 乙酰半乳糖胺 6- 硫酸酯酶活性通过活性分析被测定。显示的每一值是两次分别转染的平均值,标准方差以误差标尺指出。

[0506] 图 6 :描述在以 FGE 表达质粒瞬时转染的 36F 细胞中的 N- 乙酰半乳糖胺 6- 硫酸酯酶比活性的柱状图

[0507] 在不含氨甲蝶呤 (MTX) 的培养基中以带有反向的 FGE cDNA 的对照质粒 pXMG. 1. 2, 或者 FGE 表达质粒 pXMG. 1. 3 转染细胞。24 小时后, 细胞被含有 $1.0 \mu\text{M}$ MTX 的培养基再培养。再培养 24, 48 和 72 小时后, 采集培养基并收集细胞。N- 乙酰半乳糖胺 6- 硫酸酯酶比活性通过活性分析和 ELISA 被测定, 并以 N- 乙酰半乳糖胺 6- 硫酸酯酶活性 / 每毫克 ELISA- 反应性的 N- 乙酰半乳糖胺 6- 硫酸酯酶的比率表示。每个显示的值是两次分别转染的平均值。

[0508] 图 7 : 描述在以 FGE 表达质粒瞬时转染的 36F 细胞中的 N- 乙酰半乳糖胺 6- 硫酸酯酶产生的柱状图

[0509] 在不含氨甲蝶呤 (MTX) 的培养基中以带有反向的 FGE cDNA 的对照质粒 pXMG. 1. 2, 或者 FGE 表达质粒 pXMG. 1. 3 转染细胞。24 小时后, 细胞被含有 $1.0 \mu\text{M}$ MTX 的培养基再培养。再培养 24, 48 和 72 小时后, 采集培养基并收集细胞。N- 乙酰半乳糖胺 6- 硫酸酯酶总蛋白通过 ELISA 测定。显示的每一值是两次分别转染的平均值, 标准方差以误差标尺指出。

[0510] 图 8 : 描述在以 FGE 表达质粒瞬时转染的 30C6 细胞中的艾杜糖醛酸 -2- 硫酸酯酶活性

[0511] 在不含氨甲蝶呤 (MTX) 的培养基中以带有反向的 FGE cDNA 的对照质粒 pXMG. 1. 2, 或者 FGE 表达质粒 pXMG. 1. 3 转染细胞。24 小时后, 细胞被含有 $1.0 \mu\text{M}$ MTX 的培养基再培养。再培养 24, 48 和 72 小时后, 采集培养基并收集细胞。艾杜糖醛酸 -2- 硫酸酯酶活性通过活性分析测定。显示的每一值是两次分别转染的平均值。

[0512] 图 9 : 描述体现本发明特征的试剂盒

[0513] 本文中公开的所有参考文献都以全文被整合而作为参考。权利要求出现在下面, 并在其后给出序列表。

[0001]

序列表

- <110> Transkaryotic Therapies, Inc.
 von Figura, Kurt
 Schmidt, Bernhard
 Dierks, Thomas
 Heartlein, Michael W.
 Cosma, Maria P.
 Ballabio, Andrea
- <120> 使用甲酰-甘氨酸生成酶 (FGE) 对多种硫酸酶缺乏症和其它病症进行诊断和治疗
- <130> 0403W0
- <150> US 60/447,747
 <151> 2003-02-11
- <160> 95
- <170> PatentIn version 3.2
- <210> 1
 <211> 1180
 <212> DNA
 <213> 智人
- <220>
 <221> CDS
 <222> (20)..(1141)
 <223> FGE cDNA
- <400> 1
 acatggcccg cgggacaac atg gct gcg ccc gca cta ggg ctg gtg tgt gga 52
 Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly
 1 5 10
- cgt tgc cct gag ctg ggt ctc gtc ctc ttg ctg ctg ctg ctc tcg ctg 100
 Arg Cys Pro Glu Leu Gly Leu Val Leu Leu Leu Leu Leu Ser Leu
 15 20 25
- ctg tgt gga gcg gca ggg agc cag gag gcc ggg acc ggt gcg ggc gcg 148
 Leu Cys Gly Ala Ala Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala
 30 35 40

[0002]

| | |
|---|-----|
| ggg tcc ctt gcg ggt tct tgc ggc tgc ggc acg ccc cag cgg cct ggc Gly Ser Leu Ala Gly Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly 45 50 55 | 196 |
| gcc cat ggc agt tcg gca gcc gct cac cga tac tcg cgg gag gct aac Ala His Gly Ser Ser Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn 60 65 70 75 | 244 |
| gct ccg ggc ccc gta ccc gga gag cgg caa ctc gcg cac tca aag atg Ala Pro Gly Pro Val Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met 80 85 90 | 292 |
| gtc ccc atc cct gct gga gta ttt aca atg ggc aca gat gat cct cag Val Pro Ile Pro Ala Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln 95 100 105 | 340 |
| ata aag cag gat ggg gaa gca cct gcg agg aga gtt act att gat gcc Ile Lys Gln Asp Gly Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala 110 115 120 | 388 |
| ttt tac atg gat gcc tat gaa gtc agt aat act gaa ttt gag aag ttt Phe Tyr Met Asp Ala Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe 125 130 135 | 436 |
| gtg aac tca act ggc tat ttg aca gag gct gag aag ttt ggc gac tcc Val Asn Ser Thr Gly Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser 140 145 150 155 | 484 |
| ttt gtc ttt gaa ggc atg ttg agt gag caa gtg aag acc aat att caa Phe Val Phe Glu Gly Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln 160 165 170 | 532 |
| cag gca gtt gca gct gct ccc tgg tgg tta cct gtg aaa ggc gct aac Gln Ala Val Ala Ala Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn 175 180 185 | 580 |
| tgg aga cac cca gaa ggg cct gac tct act att ctg cac agg ccg gat Trp Arg His Pro Glu Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp 190 195 200 | 628 |
| cat cca gtt ctc cat gtg tcc tgg aat gat gcg gtt gcc tac tgc act His Pro Val Leu His Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr 205 210 215 | 676 |

[0003]

<210> 2
 <211> 374
 <212> PRT
 <213> 智人

<400> 2

Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly Arg Cys Pro Glu Leu
 1 5 10 15

Gly Leu Val Leu Leu Leu Leu Leu Leu Ser Leu Leu Cys Gly Ala Ala
 20 25 30

Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly
 35 40 45

Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser
 50 55 60

Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val
 65 70 75 80

Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala
 85 90 95

Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly
 100 105 110

Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Phe Tyr Met Asp Ala
 115 120 125

Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly
 130 135 140

Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly

[0005]

| | | | |
|---|-----|-----|-----|
| 145 | 150 | 155 | 160 |
| Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala | | | |
| | 165 | 170 | 175 |
| Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu | | | |
| | 180 | 185 | 190 |
| Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His | | | |
| | 195 | 200 | 205 |
| Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg | | | |
| | 210 | 215 | 220 |
| Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His | | | |
| 225 | 230 | 235 | 240 |
| Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His | | | |
| | 245 | 250 | 255 |
| Tyr Ala Asn Ile Trp Gln Gly Glu Phe Pro Val Thr Asn Thr Gly Glu | | | |
| | 260 | 265 | 270 |
| Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly | | | |
| | 275 | 280 | 285 |
| Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp | | | |
| | 290 | 295 | 300 |
| Trp Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly | | | |
| 305 | 310 | 315 | 320 |
| Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys | | | |

[0006]

| | | | |
|---|-----|-----|-----|
| | 325 | 330 | 335 |
| His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn | | | |
| | 340 | 345 | 350 |

| | | | |
|---|-----|-----|-----|
| Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp | | | |
| | 355 | 360 | 365 |

| | |
|-------------------------|-----|
| Arg Leu Pro Thr Met Asp | |
| | 370 |

<210> 3

<211> 1122

<212> DNA

<213> 智人

<400> 3

```

atggctgcgc ccgcaactagg gctgggtgtgt ggacgttgcc ctgagctggg tctcgtcctc      60
ttgctgctgc tgctctcgtt gctgtgtgga gcggcagggg gccaggaggc cgggaccggt      120
gcgggcgcgg ggtcccttgc gggttcttgc ggctgcggca cggcccagcg gcctggcgcc      180
catggcagtt cggcagccgc tcaccgatac tcgcgggagg ctaacgctcc gggccccgta      240
cccggagagc ggcaactcgc gcaactcaaag atgggtccca tccttgctgg agtatttaca      300
atgggcacag atgacacctc gataaagcag gatggggaag cacctgagag gagagttact      360
attgatgcct ttacatgga tgcctatgaa gtcagtaata ctgaatttga gaagtttgtg      420
aactcaactg gctatttgac agaggctgag aagtttggcg actcctttgt ctttgaaggc      480
atgttgagtg agcaagtga gaccaatatt caacaggcag ttgcagctgc tccctgggtg      540
ttacctgtga aaggcgttaa ctggagacac ccagaagggc ctgactctac tattctgcac      600
aggccggatc atccagttct ccatgtgtcc tggaatgatg cggttgccca ctgcacttgg      660
gcaggaagc ggctgcccac ggaagctgag tgggaataca gctgtcgagg aggcctgcat      720

```

[0007]

aatagacttt tcccctgggg caacaaactg cagcccaaag gccagcatta tgccaacatt 780
 tggcagggcg agtttccggt gaccaacact ggtgaggatg gcttccaagg aactggcct 840
 gttgatgcct tccctcccaa tggttatggc ttatacaaca tagtggggaa cgcattggaa 900
 tggacttcag actggtggac tgttcatcat tctgttgaag aaacgcttaa cccaaaaggt 960
 cccccttctg ggaaagaccg agtgaagaaa ggtggatcct acatgtgcca taggtcttat 1020
 tgttacaggt atcgtgtgc tgcctggagc cagaacacac ctgatagctc tgcttcgaat 1080
 ctgggattcc gctgtgcagc cgaccgcctg cccacatgg ac 1122

<210> 4

<211> 2130

<212> DNA

<213> 智人

<400> 4

acatggcccg cgggacaaca tggctgcgcc cgcactaggg ctggtgtgtg gacgttggcc 60
 tgagctgggt ctctctctt tctgtctgt gctctcctg ctgtgtggag cggcagggag 120
 ccaggaggcc gggaccgggt cgggcgcggg gtcccttgcg ggtcttgcg gctgcggcac 180
 gccccagcgg cctggcgcgc atggcagttc ggcagccgct caccgatact cgcggggagc 240
 taacgctccg ggccccgtac ccggagagcg gcaactcgcg cactcaaaga tggccccat 300
 ccctgctgga gtatttaca tgggcacaga tgatcctcag ataaagcagg atggggaagc 360
 acctgcgagg agagttacta ttgatgccct ttacatggat gcctatgaag tcagtaatac 420
 tgaatttgag aagtttgtga actcaactgg ctatttgaca gaggctgaga agtttggcga 480
 ctcccttctc tttgaaggca tgttgagtga gcaagtgaag accaatattc aacaggcagt 540
 tgcagctgct cctggtggt tacctgtgaa aggcgctaac tggagacacc cagaagggcc 600
 tgactctact attctgcaca ggccggatca tccagttctc catgtgtcct ggaatgatgc 660

[0008]

| | |
|---|------|
| ggttgccctac tgcacttggg caggggaagcg gctgcccacg gaagctgagt gggaatacag | 720 |
| ctgtcgagga ggccctgcata atagactttt cccctggggc aacaaactgc agcccaaagg | 780 |
| ccagcattat gccaacattt ggcagggcga ttttccggtg accaactg gtgaggatgg | 840 |
| cttccaagga actgcgcctg ttgatgcctt ccctccaat gtttatggct tataacaacat | 900 |
| agtggggaac gcatgggaat ggacttcaga ctggtggact gttcatcatt ctgttgaaga | 960 |
| aacgcttaac ccaaaaggtc ccccttctgg gaaagaccga gtgaagaaag gtggatccta | 1020 |
| catgtgceat aggtcttatt gttacagga tcgctgtgct gctcggagcc agaacacacc | 1080 |
| tgatagctct gcttgaatc tgggattccg ctgtgcagcc gaccgectgc ccaccatgga | 1140 |
| ctgacaacca agggtagtct tccccagtc aaggagcagt cgtgtctgac ctacattggg | 1200 |
| ctttcctcag aactttgaac gatcccatgc aaagaattcc caccctgagg tgggttacat | 1260 |
| acctgccc aa tggccaaagg aaccgccttg tgagaccaa ttgctgacct gggtcagtgc | 1320 |
| atgtgcttta tgggtgtggg catctttgga gatcatcacc atatcttact tttgagagtc | 1380 |
| tttaaagagg aaggggagtg gaggaacce tgagctaggc ttcaggagc cgcaccca | 1440 |
| cgcaggctct gccacagggg ttagaccca ggtccgacgc ttgaccttc tgggcctcaa | 1500 |
| gtgccctccc ctatcaaatg aaggaatgga cagcatgacc tctgggtgtc tctccaactc | 1560 |
| accagttcta aaaagggtat cagattctat tgtgacttca tagaatttat gatagattat | 1620 |
| tttttagcta tttttccat gtgtgaacct tgagtatac taatcatgta aagtaagagt | 1680 |
| tctcttatgt attatgttcg gaagaggggt gtggtgactc ctttatattc gtactgcact | 1740 |
| ttgttttcc aaggaaatca gtgtctttta cgttggtatg atgaatcca catggggccg | 1800 |
| gtgatggtat gctgaagttc agccgttgaa cacataggaa tgtctgtggg gtgactctac | 1860 |
| tgtgctttat cttttaacat taagtgcctt tggttcagag gggcagtcac aagctctgtt | 1920 |
| tccccctctc cccaaagcct tcagcgaacg tgaaatgtgc gctaaacggg gaaacctgtt | 1980 |

[0009]

taattctaga tatagggaaa aaggaacgag gaccttgaat gagctatatt cagggtatcc 2040
 ggtattttgt aatagggaaat aggaaacctt gttggctgtg gaatatccga tgcttgaat 2100
 catgcactgt gttgaataaa cgtatctgct 2130

<210> 5

<211> 374

<212> PRT

<213> 智人

<400> 5

Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly Arg Cys Pro Glu Leu
 1 5 10 15

Gly Leu Val Leu Leu Leu Leu Leu Leu Ser Leu Leu Cys Gly Ala Ala
 20 25 30

Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly
 35 40 45

Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser
 50 55 60

Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val
 65 70 75 80

Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala
 85 90 95

Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly
 100 105 110

Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Leu Tyr Met Asp Ala
 115 120 125

[0010]

Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly
 130 135 140

Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly
 145 150 155 160

Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala
 165 170 175

Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu
 180 185 190

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

Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg
 210 215 220

Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His
 225 230 235 240

Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His
 245 250 255

Tyr Ala Asn Ile Trp Gln Gly Asp Phe Pro Val Thr Asn Thr Gly Glu
 260 265 270

Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly
 275 280 285

Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp
 290 295 300

[0011]

Trp Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly
305 310 315 320

Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys
325 330 335

His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn
340 345 350

Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp
355 360 365

Arg Leu Pro Thr Met Asp
370

<210> 6

<211> 2297

<212> DNA

<213> 智人

<400> 6

cggctgtggt ggcagtcctt catgggttcc cgacgaggag gtctctgtgg ctgcggcggc 60

tgctaactgc gccacctgct gcagcctgtc cccgccgctc tgaagcggcc gcgtcgaagc 120

cgaaatgccg ccaccccgga cggccgagg ccttctctgg ctgggtctgg ttctgagctc 180

cgctcgcgtc gccctcggat ccgaaacgca ggccaactcg accacagatg ctctgaacgt 240

tcttctcatc atcgtggatg acctgcgccc ctccctgggc tgttatgggg ataagctggt 300

gaggtcccca aatattgacc aactggcatc ccacagcctc ctctccaga atgcctttgc 360

gcagcaagca gtgtgcgcc cgagcccggt tcttttcctc actggcagga gacctgacac 420

caccgcctg tacgacttea actcctactg gaggtgcac gctggaaact tctccacat 480

[0012]

| | |
|---|------|
| ccccagtac ttcaaggaga atggctatgt gaccatgtcg gtgggaaaag tctttcaccc | 540 |
| tgggatatct tctaaccata cggatgattc tccgtatagc tggcttttc caccttatca | 600 |
| tccttcctct gagaagtatg aaaacactaa gacatgtcga gggccagatg gagaactcca | 660 |
| tgccaacctg ctttgcctg tggatgtgct ggatgttccc gagggcacct tgcctgacaa | 720 |
| acagagcact gagcaagcca tacagtigtg gaaaagatg aaaacgtcag ccagtccttt | 780 |
| cttcctggcc gttgggtatc ataagccaca catccccttc agatacccca aggaatttca | 840 |
| gaagttgtat cccttggaga acatcacctt ggccccgat cccgaggtcc ctgatggcct | 900 |
| accccctgtg gcctacaacc cctggatgga catcaggcaa cgggaagacg tccaagcctt | 960 |
| aaacatcagt gtgccgatg gtccaattcc tgtggacttt cagcggaaaa tccgccagag | 1020 |
| ctactttgcc tctgtgcat atttgatac acaggtcggc cgcctcttga gtgctttgga | 1080 |
| cgatcttcag ctggccaaca gcacatcat tgcatttacc tcggatcatg ggtgggctct | 1140 |
| aggtgaacat ggagaatggg ccaatacag caattttgat gttgctacc atgttcccct | 1200 |
| gatattctat gttcctggaa ggacggcttc acttccggag gcaggcgaga agcttttccc | 1260 |
| ttacctgac ccttttgatt ccgctcaca gttgatggag ccaggcaggc aatccatgga | 1320 |
| ccttgggaa ctigtgtctc ttttccac gctggctgga cttgcaggac tgcaggttcc | 1380 |
| acctcgtgc cccgttctt catctcagt tgagctgtgc agagaaggca agaaccttct | 1440 |
| gaagcatttt cgattcctg acttgaaga ggatccgtac ctccctggta atcccctga | 1500 |
| actgattgcc tatagccagt atccccgcc ttcagacatc cctcagtgga attctgacaa | 1560 |
| gccgagttta aaagatataa agatcatggg ctattccata cgcacatag actataggta | 1620 |
| tactgtgtgg gttggcttca atcctgatga atttctagct aacttttctg acatccatgc | 1680 |
| aggggaactg tattttgtgg attctgacct attgcaggat cacaatatgt ataatgattc | 1740 |
| ccaagggtga gatcttttcc agttgtgat gccttgagt ttgccaacca tggatggcaa | 1800 |

[0013]

atgtgatgtg ctcccttcca gctggtgaga ggaggagtta gagctggtcg ttttgtgatt 1860
 acccataata ttggaagcag cctgagggct agttaatcca aacatgcatc aacaatttgg 1920
 cctgagaata tgtaacagcc aaaccttttc gtttagtctt tattaaaatt tataattggt 1980
 aattggacca gttttttttt taatttcctt ctttttaaaa cagttacggc ttatttactg 2040
 aataaataca aagcaaaca actcaagtta tgtcacacct ttggatacga agaccataca 2100
 taataacca acataacatt atacacaaag aatactttca ttatttgtgg aatttagtgc 2160
 atttcaaaaa gtaatcatat atcaaactag gcaccacact aagttcctga ttattttggt 2220
 tataatttaa taatatactt tatgagccct atatattcaa aatattatgt taacatgtaa 2280
 tccatgttcc tttttccc 2297

<210> 7

<211> 550

<212> PRT

<213> 智人

<400> 7

Met Pro Pro Pro Arg Thr Gly Arg Gly Leu Leu Trp Leu Gly Leu Val
 1 5 10 15

Leu Ser Ser Val Cys Val Ala Leu Gly Ser Glu Thr Gln Ala Asn Ser
 20 25 30

Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp Asp Leu Arg
 35 40 45

Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile
 50 55 60

Asp Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln
 65 70 75 80

[0014]

Gln Ala Val Cys Ala Pro Ser Arg Val Ser Phe Leu Thr Gly Arg Arg
 85 90 95

Pro Asp Thr Thr Arg Leu Tyr Asp Phe Asn Ser Tyr Trp Arg Val His
 100 105 110

Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu Asn Gly Tyr
 115 120 125

Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn
 130 135 140

His Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro
 145 150 155 160

Ser Ser Glu Lys Tyr Glu Asn Thr Lys Thr Cys Arg Gly Pro Asp Gly
 165 170 175

Glu Leu His Ala Asn Leu Leu Cys Pro Val Asp Val Leu Asp Val Pro
 180 185 190

Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala Ile Gln Leu
 195 200 205

Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly
 210 215 220

Tyr His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys
 225 230 235 240

Leu Tyr Pro Leu Glu Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro
 245 250 255

[0015]

Asp Gly Leu Pro Pro Val Ala Tyr Asn Pro Trp Met Asp Ile Arg Gln
 260 265 270

Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr Gly Pro Ile
 275 280 285

Pro Val Asp Phe Gln Arg Lys Ile Arg Gln Ser Tyr Phe Ala Ser Val
 290 295 300

Ser Tyr Leu Asp Thr Gln Val Gly Arg Leu Leu Ser Ala Leu Asp Asp
 305 310 315 320

Leu Gln Leu Ala Asn Ser Thr Ile Ile Ala Phe Thr Ser Asp His Gly
 325 330 335

Trp Ala Leu Gly Glu His Gly Glu Trp Ala Lys Tyr Ser Asn Phe Asp
 340 345 350

Val Ala Thr His Val Pro Leu Ile Phe Tyr Val Pro Gly Arg Thr Ala
 355 360 365

Ser Leu Pro Glu Ala Gly Glu Lys Leu Phe Pro Tyr Leu Asp Pro Phe
 370 375 380

Asp Ser Ala Ser Gln Leu Met Glu Pro Gly Arg Gln Ser Met Asp Leu
 385 390 395 400

Val Glu Leu Val Ser Leu Phe Pro Thr Leu Ala Gly Leu Ala Gly Leu
 405 410 415

Gln Val Pro Pro Arg Cys Pro Val Pro Ser Phe His Val Glu Leu Cys
 420 425 430

[0016]

Arg Glu Gly Lys Asn Leu Leu Lys His Phe Arg Phe Arg Asp Leu Glu
 435 440 445

Glu Asp Pro Tyr Leu Pro Gly Asn Pro Arg Glu Leu Ile Ala Tyr Ser
 450 455 460

Gln Tyr Pro Arg Pro Ser Asp Ile Pro Gln Trp Asn Ser Asp Lys Pro
 465 470 475 480

Ser Leu Lys Asp Ile Lys Ile Met Gly Tyr Ser Ile Arg Thr Ile Asp
 485 490 495

Tyr Arg Tyr Thr Val Trp Val Gly Phe Asn Pro Asp Glu Phe Leu Ala
 500 505 510

Asn Phe Ser Asp Ile His Ala Gly Glu Leu Tyr Phe Val Asp Ser Asp
 515 520 525

Pro Leu Gln Asp His Asn Met Tyr Asn Asp Ser Gln Gly Gly Asp Leu
 530 535 540

Phe Gln Leu Leu Met Pro
 545 550

<210> 8

<211> 2657

<212> DNA

<213> 智人

<400> 8

gaattccggg ccatgagctg ccccgtagcc gctgctgcg cgctgctgct agtcctggg 60

ctctgccggg cgcgtccccg gaacgcactg ctgctcctcg cggatgacgg aggctttgag 120

[0017]

| | |
|---|------|
| agtggcgctg acaacaacag cgccatcgcc accccgcacc tggacgcctt ggcccgcgc | 180 |
| agcctcctct ttcgcaatgc cttcacctcg gtcagcagct gctctcccag ccgcgccagc | 240 |
| ctcctcactg gcctgcccc a gcatcagaat gggatgtacg ggctgcacca ggacgtgcac | 300 |
| cacttcaact ccttcgacaa ggtgctggagc ctgcccgtgc tgctcagcca agctgggtgtg | 360 |
| cgcacaggca tcacgggaa gaagcacgtg gggccggaga ccgtgtacc gtttgacttt | 420 |
| gcgtacacgg aggagaatgg ctccgtcctc cagggtggggc ggaacatcac tagaattaag | 480 |
| ctgctcgtcc ggaaattcct gcagactcag gatgaccggc ctttcttct ctacgtcgcc | 540 |
| ttccacgacc cccaccgctg tgggcactcc cagccccagt acggaacctt ctgtgagaag | 600 |
| tttgcaacg gagagagcgg catgggtcgt atcccagact ggacccccca ggccctacgac | 660 |
| ccactggacg tgctgggtcc ttacttcgtc cccaacacc cggcagccc agccgacctg | 720 |
| gccgctcagt acaccaccgt cggccgatg gaccaaggag ttggactggt gctccaggag | 780 |
| ctgctgacg ccggtgtcct gaacgacaca ctggtgatct tcacgtccga caacgggatc | 840 |
| ccctcccca gggcaggac caacctgtac tggccgggca ctgctgaacc cttactggtg | 900 |
| tcacccccg agcaccmeta acgctggggc caagtcagc aggcctacgt gagcctccta | 960 |
| gacctcacgc ccaccatctt ggattggttc tgatcccgt accccagcta cgccatctt | 1020 |
| ggctcgaaga ccatccact cactggccgg tccctcctgc cggcgtgga ggccgagccc | 1080 |
| ctctgggcca ccgtctttgg cagccagagc caccacgagg tcaccatgct ctacccatg | 1140 |
| cgctccgtgc agcaccggca cttccgctc gtgcacaacc tcaactcaa gatgccctt | 1200 |
| cccatcgacc aggacttcta cgtctcacc accttccagg acctcctgaa ccgcaccaca | 1260 |
| gctggtcagc ccacgggctg gtacaaggac ctccgtcatt actactaccg ggcgcgtgg | 1320 |
| gagctctacg accggagccg ggacccccac gagaccaga acctggccac cgacccgcgc | 1380 |
| tttctcagc ttctggagat gcttcgggac cagctggcca agtggcagtg ggagaccac | 1440 |

[0018]

| | |
|--|------|
| gacccctggg tgtgcgcccc cgacggcgtc ctggaggaga agctctctcc ccagtgccag | 1500 |
| cccctccaca atgagctgtg accatcccag gaggcctgtg cacacatccc aggcattgtcc | 1560 |
| cagacacatc ccacacgtgt ccgtgtggcc ggccagcctg gggagtagtg gcaacagccc | 1620 |
| ttcctccac actcccatcc aaggagggtt ctcccttctt gtggggtcac tcttgccatt | 1680 |
| gcctggaggg ggaccagagc atgtgaccag agcatgtgcc cagccccctcc accaccaggg | 1740 |
| gcactgccgt catggcaggg gacacagtgt tccttgtgtc tgaacctgt cccagcacgg | 1800 |
| gaattctaga catacgtggt ctgcggacag ggcagcggcc ccagcccatg acaaggaggt | 1860 |
| cttgttttct ggcttggtt ggggacctgc aaatgggagg cctgaggccc tcttcaggct | 1920 |
| ttggcagcca cagatacttc tgaacccttc acagagagca ggcaggggct tcggtgccgc | 1980 |
| gtgggcagta cgcaggctcc accgacactc acctgggagc acggcgctg gctcttacca | 2040 |
| gcgtctggcc tagaggaagc ctttgagcga cctttgggca gtttctgct tcttctgttt | 2100 |
| tgcccatggt caagtccctg ttccccaggc aggtttcagc tgattggcag caggctccct | 2160 |
| gagtgatgag cttgaacctg tgggttttct gggcagaagc ttatctttt tgagagtgtc | 2220 |
| cgaagatgaa ggcattggcga tgcccgtcct ctggcttggg ttaattcttc ggtgacactg | 2280 |
| gcattgctgg gtggtgatgc ccgtcctctg gcttgggtta attcttcggt gacactggcg | 2340 |
| ttgctgggtg gcaatgcccg tcctctggct tgggttaatt cttcggtgac actggcgttg | 2400 |
| ctgggtggcg atgcccgctc tctggcttgg gtttaattctt ggatgacgtc ggcgttgctg | 2460 |
| ggagaatgtg ccgttcctgc cctgcctcca cccacctcg ggcagaaagc ccggcctgga | 2520 |
| caccctcgg cctggacacc cctcgaagga gaggcgcctt ccttgagtag gtgggctccc | 2580 |
| cttgcccttc cctccctatc actccatact ggggtgggct ggaggaggcc acaggccagc | 2640 |
| tattgtaaaa gcttttt | 2657 |

<210> 9

[0019]

<211> 502

<212> PRT

<213> 智人

<400> 9

Met Ser Cys Pro Val Pro Ala Cys Cys Ala Leu Leu Leu Val Leu Gly
1 5 10 15

Leu Cys Arg Ala Arg Pro Arg Asn Ala Leu Leu Leu Leu Ala Asp Asp
 20 25 30

Gly Gly Phe Glu Ser Gly Ala Tyr Asn Asn Ser Ala Ile Ala Thr Pro
 35 40 45

His Leu Asp Ala Leu Ala Arg Arg Ser Leu Leu Phe Arg Asn Ala Phe
50 55 60

Thr Ser Val Ser Ser Cys Ser Pro Ser Arg Ala Ser Leu Leu Thr Gly
65 70 75 80

Leu Pro Gln His Gln Asn Gly Met Tyr Gly Leu His Gln Asp Val His
 85 90 95

His Phe Asn Ser Phe Asp Lys Val Arg Ser Leu Pro Leu Leu Leu Ser
 100 105 110

Gln Ala Gly Val Arg Thr Gly Ile Ile Gly Lys Lys His Val Gly Pro
 115 120 125

Glu Thr Val Tyr Pro Phe Asp Phe Ala Tyr Thr Glu Glu Asn Gly Ser
130 135 140

Val Leu Gln Val Gly Arg Asn Ile Thr Arg Ile Lys Leu Leu Val Arg
145 150 155 160

[0020]

Lys Phe Leu Gln Thr Gln Asp Asp Arg Pro Phe Phe Leu Tyr Val Ala
 165 170 175

Phe His Asp Pro His Arg Cys Gly His Ser Gln Pro Gln Tyr Gly Thr
 180 185 190

Phe Cys Glu Lys Phe Gly Asn Gly Glu Ser Gly Met Gly Arg Ile Pro
 195 200 205

Asp Trp Thr Pro Gln Ala Tyr Asp Pro Leu Asp Val Leu Val Pro Tyr
 210 215 220

Phe Val Pro Asn Thr Pro Ala Ala Arg Ala Asp Leu Ala Ala Gln Tyr
 225 230 235 240

Thr Thr Val Gly Arg Met Asp Gln Gly Val Gly Leu Val Leu Gln Glu
 245 250 255

Leu Arg Asp Ala Gly Val Leu Asn Asp Thr Leu Val Ile Phe Thr Ser
 260 265 270

Asp Asn Gly Ile Pro Phe Pro Ser Gly Arg Thr Asn Leu Tyr Trp Pro
 275 280 285

Gly Thr Ala Glu Pro Leu Leu Val Ser Ser Pro Glu His Pro Lys Arg
 290 295 300

Trp Gly Gln Val Ser Glu Ala Tyr Val Ser Leu Leu Asp Leu Thr Pro
 305 310 315 320

Thr Ile Leu Asp Trp Phe Ser Ile Pro Tyr Pro Ser Tyr Ala Ile Phe
 325 330 335

[0021]

Gly Ser Lys Thr Ile His Leu Thr Gly Arg Ser Leu Leu Pro Ala Leu
 340 345 350

Glu Ala Glu Pro Leu Trp Ala Thr Val Phe Gly Ser Gln Ser His His
 355 360 365

Glu Val Thr Met Ser Tyr Pro Met Arg Ser Val Gln His Arg His Phe
 370 375 380

Arg Leu Val His Asn Leu Asn Phe Lys Met Pro Phe Pro Ile Asp Gln
 385 390 395 400

Asp Phe Tyr Val Ser Pro Thr Phe Gln Asp Leu Leu Asn Arg Thr Thr
 405 410 415

Ala Gly Gln Pro Thr Gly Trp Tyr Lys Asp Leu Arg His Tyr Tyr Tyr
 420 425 430

Arg Ala Arg Trp Glu Leu Tyr Asp Arg Ser Arg Asp Pro His Glu Thr
 435 440 445

Gln Asn Leu Ala Thr Asp Pro Arg Phe Ala Gln Leu Leu Glu Met Leu
 450 455 460

Arg Asp Gln Leu Ala Lys Trp Gln Trp Glu Thr His Asp Pro Trp Val
 465 470 475 480

Cys Ala Pro Asp Gly Val Leu Glu Glu Lys Leu Ser Pro Gln Cys Gln
 485 490 495

Pro Leu His Asn Glu Leu
 500

[0022]

<210> 10
 <211> 1014
 <212> DNA
 <213> 智人

<400> 10
 cgtgcctgta atcccagcag ctactcactc aggaggctga ggcaggagaa tctcttgaac 60
 ccggaaggca gaggttgacg tgagccaaga tcgcgccact gaactccagc ctgggtgaca 120
 gagtgagaact gtctcagaac agcaacaaca aatgcccgc tctgctggg tccagaagag 180
 cttgaataac tgcatgttct ttttctcaat tttcatttcc cagaactggg cacctccggg 240
 ctgtgaaaag ttagggaagt gtctgacacc tccagaatcc attccaaga agtgcctctg 300
 gtcccactag cacctgcgca gactcaggcc aggccataga tctccagttg gccctgcaag 360
 tgccctggagg aaggatggct ctggcctcgg tcctcccca accctgcca agccagacag 420
 acagcacctg cagacgcagg gggactgcac aattccacct gcccaggacc tgaccctggc 480
 gtgtgcttgg ccctcctcct cccccacggc gcctcagatt tcaggacct cctcctcgcc 540
 cacggcgctt cagacctcag gaccctgccg tctcagcct ttgtgaacce caaatatctg 600
 agaccagtct cagtttattt tgccaagggt aaggatgcac ctgtgacagc ctcaggaggt 660
 cctgacaaca ggtgcccagag gtggctgggg atacagttg cttttataca tcttagggag 720
 acacaagatc agtatgtgta tggcgtacat tggttcagtc agccttccac tgaatacacg 780
 attgagtctg gcccagtga tccgcatttt tatgtaaaca gtaaggaac ggggcaatca 840
 tataagcggt tgtctcaggg gagccccaga gggatgactt ccagttccgt ctgtcctttg 900
 tccacaagga atttcctgg gcgctaatta tgaggaggc gtgtagcttc ttatcattgt 960
 agctatgtta tttagaaata aacgggagg caggtttgcc taattcccag gttg 1014

<210> 11

[0023]

Gly Ser Pro Asn Cys His Phe Gly Pro Tyr Asp Asn Lys Ala Arg Pro
 165 170 175

Asn Ile Pro Val Tyr Arg Asp Trp Glu Met Val Gly Arg Tyr Tyr Glu
 180 185 190

Glu Phe Pro Ile Asn Leu Lys Thr Gly Glu Ala Asn Leu Thr Gln Ile
 195 200 205

Tyr Leu Gln Glu Ala Leu Asp Phe Ile Lys Arg Gln Ala Arg His His
 210 215 220

Pro Phe Phe Leu Tyr Trp Ala Val Asp Ala Thr His Ala Pro Val Tyr
 225 230 235 240

Ala Ser Lys Pro Phe Leu Gly Thr Ser Gln Arg Gly Arg Tyr Gly Asp
 245 250 255

Ala Val Arg Glu Ile Asp Asp Ser Ile Gly Lys Ile Leu Glu Leu Leu
 260 265 270

Gln Asp Leu His Val Ala Asp Asn Thr Phe Val Phe Phe Thr Ser Asp
 275 280 285

Asn Gly Ala Ala Leu Ile Ser Ala Pro Glu Gln Gly Gly Ser Asn Gly
 290 295 300

Pro Phe Leu Cys Gly Lys Gln Thr Thr Phe Glu Gly Gly Met Arg Glu
 305 310 315 320

Pro Ala Leu Ala Trp Trp Pro Gly His Val Thr Ala Gly Gln Val Ser
 325 330 335

[0025]

His Gln Leu Gly Ser Ile Met Asp Leu Phe Thr Thr Ser Leu Ala Leu
 340 345 350

Ala Gly Leu Thr Pro Pro Ser Asp Arg Ala Ile Asp Gly Leu Asn Leu
 355 360 365

Leu Pro Thr Leu Leu Gln Gly Arg Leu Met Asp Arg Pro Ile Phe Tyr
 370 375 380

Tyr Arg Gly Asp Thr Leu Met Ala Ala Thr Leu Gly Gln His Lys Ala
 385 390 395 400

His Phe Trp Thr Trp Thr Asn Ser Trp Glu Asn Phe Arg Gln Gly Ile
 405 410 415

Asp Phe Cys Pro Gly Gln Asn Val Ser Gly Val Thr Thr His Asn Leu
 420 425 430

Glu Asp His Thr Lys Leu Pro Leu Ile Phe His Leu Gly Arg Asp Pro
 435 440 445

Gly Glu Arg Phe Pro Leu Ser Phe Ala Ser Ala Glu Tyr Gln Glu Ala
 450 455 460

Leu Ser Arg Ile Thr Ser Val Val Gln Gln His Gln Glu Ala Leu Val
 465 470 475 480

Pro Ala Gln Pro Gln Leu Asn Val Cys Asn Trp Ala Val Met Asn Trp
 485 490 495

Ala Pro Pro Gly Cys Glu Lys Leu Gly Lys Cys Leu Thr Pro Pro Glu
 500 505 510

[0026]

Ser Ile Pro Lys Lys Cys Leu Trp Ser His
 515 520

<210> 12

<211> 2379

<212> DNA

<213> 智人

<400> 12

```

ggaattccgg tcggcctctc gcccttcage tacctgtgcg tcctccgtc cgtcccgtc   60
ccggggtcac cccggagcct gtccgctatg cggctcctgc ctctagcccc aggtcggctc   120
cggcggggca gccccgcca cctgcctcc tgcagcccag cgctgctact gctgggtctg   180
ggcggctgcc tgggggtctt cgggggtgct gcggaaccc ggaggccaa cgtgggtctg   240
ctcctcacgg acgaccagga cgaagtgctc ggcgcatga caccactaaa gaaaacaaa   300
gctctcatcg gagagatggg gatgactttt tccagtgtt atgtccaag tgctctctgc   360
tgccccagca gagccagtat cctgacagga aagtaccac ataatcatca cgttgtgaac   420
aacactctgg aggggaactg cagtagtaag tcctggcaga agatccaaga accaaatact   480
ttcccagcaa ttctcagatc aatgtgtggt tatcagacct ttttgcagg gaaatatta   540
aatgagtacg gagccccaga tgcaggtgga ctagaacacg ttcctctggg ttggagttac   600
tggtatgcct tggaaaagaa ttctaagtat tataattaca cctgtctat caatgggaag   660
gcacggaagc atggtgaaaa ctatagtgtg gactacctga cagatgtttt ggctaatgtc   720
tccttgact ttctggacta caagtccaac tttgagcct tcttcatgat gatgccact   780
ccagcgcctc attgccttg gacagctgca cctcagtacc agaaggcttt ccagaatgtc   840
tttcaccaa gaaacaagaa cttcaacatc catggaacga acaagcactg gtaattagg   900
caagccaaga ctccaatgac taattcttca atacagtttt tagataatgc atttaggaaa   960

```

[0027]

| | |
|--|------|
| aggtggcaaa ctctcctctc agttgatgac cttgtggaga aactgggtcaa gaggetggag | 1020 |
| ttcactgggg agctcaacaa cacttacatc ttctatacct cagacaatgg ctatcacaca | 1080 |
| ggacagtttt ccttgccaat agacaagaga cagctgtatg agtttgatat caaagttcca | 1140 |
| ctgttggttc gaggacctgg gatcaaacca aatcagacaa gcaagatgct ggttgccaac | 1200 |
| attgacttgg gtcctactat tttggacatt gctggctacg acctaaataa gacacagatg | 1260 |
| gatgggatgt ccttattgcc cattttgaga ggtgccagta acttgacctg gcgatcagat | 1320 |
| gtcctgggtg aataccaagg agaaggccgt aacgtcactg acccaacatg cccttcctg | 1380 |
| agtcctggcg tatctcaatg ctcccagac tgtgtatgtg aagatgctta taacaatacc | 1440 |
| tatgcctgtg tgaggacaat gtcagcattg tggaatttgc agtattgcca gtttgatgac | 1500 |
| caggaggtgt ttgtagaagt ctataatctg actgcagacc cagaccagat cactaacatt | 1560 |
| gctaaaacca tagaccaga gcttttagga aagatgaact atcggttaat gatgttacag | 1620 |
| tcctgttctg ggccaacctg tcgactcca ggggtttttg accccggata caggtttgac | 1680 |
| ccccgtctca tgttcagcaa tcgcgccagt gtcaggactc gaagattttc caaacatctt | 1740 |
| ctgtagcgac ctcacacagc ctctgcagat ggatccctgc acgcctcttt ctgatgaagt | 1800 |
| gattgtagta ggtgtctgta gctagtcttc aagaccacac ctggaagagt ttctgggctg | 1860 |
| gctttaagtc ctgtttgaaa aagcaacca gtcagctgac ttctctgtgc aatgtgttaa | 1920 |
| actgtgaact ctgcccattg gtcaggagtg gctgtctctg gtctcttctt ttagctgaca | 1980 |
| aggacactcc tgaggctttt gttctcactg tatttttttt atcctggggc cacagttctt | 2040 |
| gattattcct cttgtggtta aagactgaat ttgtaaacc attcagataa atggcagtac | 2100 |
| tttaggacac acacaaacac acagatacac cttttgatat gtaagcttga cctaaagtca | 2160 |
| aaggacctgt gtagcatttc agattgagca ctccactatc aaaaatacta acatcacatg | 2220 |
| gcttgaagag taaccatcag agctgaatca tccaagtaag aacaagtacc attgttgatt | 2280 |

[0028]

gataagtaga gatacatgtt ttatgatgtt catcacagtg tggtaagggtt gcaaattcaa 2340

aacatgtcac ccaagctctg ttcattgttt tgtgaattc 2379

<210> 13

<211> 552

<212> PRT

<213> 智人

<400> 13

Met Arg Leu Leu Pro Leu Ala Pro Gly Arg Leu Arg Arg Gly Ser Pro
1 5 10 15

Arg His Leu Pro Ser Cys Ser Pro Ala Leu Leu Leu Leu Val Leu Gly
 20 25 30

Gly Cys Leu Gly Val Phe Gly Val Ala Ala Gly Thr Arg Arg Pro Asn
 35 40 45

Val Val Leu Leu Leu Thr Asp Asp Gln Asp Glu Val Leu Gly Gly Met
 50 55 60

Thr Pro Leu Lys Lys Thr Lys Ala Leu Ile Gly Glu Met Gly Met Thr
65 70 75 80

Phe Ser Ser Ala Tyr Val Pro Ser Ala Leu Cys Cys Pro Ser Arg Ala
 85 90 95

Ser Ile Leu Thr Gly Lys Tyr Pro His Asn His His Val Val Asn Asn
 100 105 110

Thr Leu Glu Gly Asn Cys Ser Ser Lys Ser Trp Gln Lys Ile Gln Glu
 115 120 125

[0029]

Pro Asn Thr Phe Pro Ala Ile Leu Arg Ser Met Cys Gly Tyr Gln Thr
 130 135 140

Phe Phe Ala Gly Lys Tyr Leu Asn Glu Tyr Gly Ala Pro Asp Ala Gly
 145 150 155 160

Gly Leu Glu His Val Pro Leu Gly Trp Ser Tyr Trp Tyr Ala Leu Glu
 165 170 175

Lys Asn Ser Lys Tyr Tyr Asn Tyr Thr Leu Ser Ile Asn Gly Lys Ala
 180 185 190

Arg Lys His Gly Glu Asn Tyr Ser Val Asp Tyr Leu Thr Asp Val Leu
 195 200 205

Ala Asn Val Ser Leu Asp Phe Leu Asp Tyr Lys Ser Asn Phe Glu Pro
 210 215 220

Phe Phe Met Met Ile Ala Thr Pro Ala Pro His Ser Pro Trp Thr Ala
 225 230 235 240

Ala Pro Gln Tyr Gln Lys Ala Phe Gln Asn Val Phe Ala Pro Arg Asn
 245 250 255

Lys Asn Phe Asn Ile His Gly Thr Asn Lys His Trp Leu Ile Arg Gln
 260 265 270

Ala Lys Thr Pro Met Thr Asn Ser Ser Ile Gln Phe Leu Asp Asn Ala
 275 280 285

Phe Arg Lys Arg Trp Gln Thr Leu Leu Ser Val Asp Asp Leu Val Glu
 290 295 300

[0030]

Lys Leu Val Lys Arg Leu Glu Phe Thr Gly Glu Leu Asn Asn Thr Tyr
 305 310 315 320

Ile Phe Tyr Thr Ser Asp Asn Gly Tyr His Thr Gly Gln Phe Ser Leu
 325 330 335

Pro Ile Asp Lys Arg Gln Leu Tyr Glu Phe Asp Ile Lys Val Pro Leu
 340 345 350

Leu Val Arg Gly Pro Gly Ile Lys Pro Asn Gln Thr Ser Lys Met Leu
 355 360 365

Val Ala Asn Ile Asp Leu Gly Pro Thr Ile Leu Asp Ile Ala Gly Tyr
 370 375 380

Asp Leu Asn Lys Thr Gln Met Asp Gly Met Ser Leu Leu Pro Ile Leu
 385 390 395 400

Arg Gly Ala Ser Asn Leu Thr Trp Arg Ser Asp Val Leu Val Glu Tyr
 405 410 415

Gln Gly Glu Gly Arg Asn Val Thr Asp Pro Thr Cys Pro Ser Leu Ser
 420 425 430

Pro Gly Val Ser Gln Cys Phe Pro Asp Cys Val Cys Glu Asp Ala Tyr
 435 440 445

Asn Asn Thr Tyr Ala Cys Val Arg Thr Met Ser Ala Leu Trp Asn Leu
 450 455 460

Gln Tyr Cys Glu Phe Asp Asp Gln Glu Val Phe Val Glu Val Tyr Asn
 465 470 475 480

[0031]

Leu Thr Ala Asp Pro Asp Gln Ile Thr Asn Ile Ala Lys Thr Ile Asp
 485 490 495

Pro Glu Leu Leu Gly Lys Met Asn Tyr Arg Leu Met Met Leu Gln Ser
 500 505 510

Cys Ser Gly Pro Thr Cys Arg Thr Pro Gly Val Phe Asp Pro Gly Tyr
 515 520 525

Arg Phe Asp Pro Arg Leu Met Phe Ser Asn Arg Gly Ser Val Arg Thr
 530 535 540

Arg Arg Phe Ser Lys His Leu Leu
 545 550

<210> 14

<211> 2022

<212> DNA

<213> 智人

<400> 14

ccggtaccgg ctctcctggt gctccctcta gcgccttccc cccggcccga ctgcctggtc 60
 agegccaaagt gacttacgcc cccgacctg agccccgacc gctaggcgag gaggatcaga 120
 tctccgctcg agaatctgaa ggtgccctgg tcttggagga gttccgtccc agcctcgcg 180
 tctcccggta ctgctgccc cggccctctg gagcttcagg aggcggccgt cagggtcggg 240
 gagtatttgg gtccggggtc tcaggaagg gcggcgctg ggtctgcggt atcggaaaga 300
 gcctgctgga gccaaagtagc cctccctctc ttgggacaga ccctcggtc ccatgtccat 360
 gggggcaccg cggtccctcc tctggcctt ggctgctggc ctggccgttg cccgtccgcc 420
 caacatcgtg ctgatcttgg ccgacgacct cggctatggg gacctgggct gctatgggca 480
 ccccagctct accactccca acctggacca gctggcggcg ggagggtgc ggttcacaga 540

[0032]

| | |
|---|------|
| cttctacgtg cctgtgtctc tgtgcacacc ctctagggcc gccctcctga cgggccggct | 600 |
| cccggttcgg atgggcatgt accctggcgt cctggtgccc agctcccggg ggggcctgcc | 660 |
| cctggaggag gtgaccgtgg ccgaagtcc tggctgcccga ggctacctca caggaatggc | 720 |
| cggcaagtgg caccttgggg tggggcctga gggggccttc ctgcccccc atcagggtt | 780 |
| ccatcgattt ctaggcatcc cgtactccca cgaccagggc ccctgccaga acctgacctg | 840 |
| cttcccgcg gccactcctt gcgacggtag ctgtgaccag ggcttggctc ccatcccact | 900 |
| gttggccaac ctgtccgtgg aggcgcagcc cccttggctg cccggactag aggcccgcta | 960 |
| catggctttc gcccatgacc tcatggccga cgcccagcgc caggatcgcc ctttcttct | 1020 |
| gtactatgcc tctcaccaca cccactacc tcagttcagt gggcagagct ttgcagagcg | 1080 |
| ttcaggccgc gggccatttg gggactccct gatggagctg gatgcagctg tggggaccct | 1140 |
| gatgacagcc ataggggacc tggggctgct tgaagagacg ctggatcatct tcaactgcaga | 1200 |
| caatggacct gagaccatgc gtatgtccc aggcggctgc tccggtctct tgcggtgtgg | 1260 |
| aaaggaacg acctacgagg gcggtgtccg agagcctgcc ttggccttct ggccaggta | 1320 |
| tatcgctccc ggctgaccc acgagctggc cagctccctg gacctgctgc ctaccctggc | 1380 |
| agccctggct ggggccccac tgcccattgt caccttggat ggctttgacc tcagcccct | 1440 |
| gctgctgggc acaggcaaga gccctggca gtctctcttc ttctaccct cctaccaga | 1500 |
| cgaggtccgt ggggtttttg ctgtgcggac tggaaagtac aaggctcact tcttaccca | 1560 |
| gggctctgcc cacagtgata ccaactgcaga ccctgcctgc cacgcctcca gctctctgac | 1620 |
| tgctcatgag cccccgtgc tctatgacct gtccaaggac cctggtgaga actacaacct | 1680 |
| gctggggggt gtggccgggg ccaccccaga ggtgctgcaa gccctgaaac agcttcagct | 1740 |
| gctcaaggcc caglttagacg cagctgtgac cttcgcccc agccagggtg cccggggcga | 1800 |
| ggaccccgcc ctgcagatct gctgtcatcc tggctgeacc cccgcccag cttgctgcca | 1860 |

[0033]

ttgccagat ccccatgcct gagggccct cggctggcct gggcatgtga tggetcctca 1920
 ctgggagcct ggggggagg ctcaggtgtc tggaggggt ttgtgcctga taacgtaata 1980
 acaccagtgg agacttgac atctgaaaaa aaaaaaaaaa aa 2022

<210> 15

<211> 507

<212> PRT

<213> 智人

<400> 15

Met Gly Ala Pro Arg Ser Leu Leu Leu Ala Leu Ala Ala Gly Leu Ala
 1 5 10 15

Val Ala Arg Pro Pro Asn Ile Val Leu Ile Phe Ala Asp Asp Leu Gly
 20 25 30

Tyr Gly Asp Leu Gly Cys Tyr Gly His Pro Ser Ser Thr Thr Pro Asn
 35 40 45

Leu Asp Gln Leu Ala Ala Gly Gly Leu Arg Phe Thr Asp Phe Tyr Val
 50 55 60

Pro Val Ser Leu Cys Thr Pro Ser Arg Ala Ala Leu Leu Thr Gly Arg
 65 70 75 80

Leu Pro Val Arg Met Gly Met Tyr Pro Gly Val Leu Val Pro Ser Ser
 85 90 95

Arg Gly Gly Leu Pro Leu Glu Glu Val Thr Val Ala Glu Val Leu Ala
 100 105 110

Ala Arg Gly Tyr Leu Thr Gly Met Ala Gly Lys Trp His Leu Gly Val
 115 120 125

[0034]

Gly Pro Glu Gly Ala Phe Leu Pro Pro His Gln Gly Phe His Arg Phe
 130 135 140

Leu Gly Ile Pro Tyr Ser His Asp Gln Gly Pro Cys Gln Asn Leu Thr
 145 150 155 160

Cys Phe Pro Pro Ala Thr Pro Cys Asp Gly Gly Cys Asp Gln Gly Leu
 165 170 175

Val Pro Ile Pro Leu Leu Ala Asn Leu Ser Val Glu Ala Gln Pro Pro
 180 185 190

Trp Leu Pro Gly Leu Glu Ala Arg Tyr Met Ala Phe Ala His Asp Leu
 195 200 205

Met Ala Asp Ala Gln Arg Gln Asp Arg Pro Phe Phe Leu Tyr Tyr Ala
 210 215 220

Ser His His Thr His Tyr Pro Gln Phe Ser Gly Gln Ser Phe Ala Glu
 225 230 235 240

Arg Ser Gly Arg Gly Pro Phe Gly Asp Ser Leu Met Glu Leu Asp Ala
 245 250 255

Ala Val Gly Thr Leu Met Thr Ala Ile Gly Asp Leu Gly Leu Leu Glu
 260 265 270

Glu Thr Leu Val Ile Phe Thr Ala Asp Asn Gly Pro Glu Thr Met Arg
 275 280 285

Met Ser Arg Gly Gly Cys Ser Gly Leu Leu Arg Cys Gly Lys Gly Thr
 290 295 300

[0035]

Thr Tyr Glu Gly Gly Val Arg Glu Pro Ala Leu Ala Phe Trp Pro Gly
305 310 315 320

His Ile Ala Pro Gly Val Thr His Glu Leu Ala Ser Ser Leu Asp Leu
325 330 335

Leu Pro Thr Leu Ala Ala Leu Ala Gly Ala Pro Leu Pro Asn Val Thr
340 345 350

Leu Asp Gly Phe Asp Leu Ser Pro Leu Leu Leu Gly Thr Gly Lys Ser
355 360 365

Pro Arg Gln Ser Leu Phe Phe Tyr Pro Ser Tyr Pro Asp Glu Val Arg
370 375 380

Gly Val Phe Ala Val Arg Thr Gly Lys Tyr Lys Ala His Phe Phe Thr
385 390 395 400

Gln Gly Ser Ala His Ser Asp Thr Thr Ala Asp Pro Ala Cys His Ala
405 410 415

Ser Ser Ser Leu Thr Ala His Glu Pro Pro Leu Leu Tyr Asp Leu Ser
420 425 430

Lys Asp Pro Gly Glu Asn Tyr Asn Leu Leu Gly Gly Val Ala Gly Ala
435 440 445

Thr Pro Glu Val Leu Gln Ala Leu Lys Gln Leu Gln Leu Leu Lys Ala
450 455 460

Gln Leu Asp Ala Ala Val Thr Phe Gly Pro Ser Gln Val Ala Arg Gly
465 470 475 480

[0036]

Glu Asp Pro Ala Leu Gln Ile Cys Cys His Pro Gly Cys Thr Pro Arg
 485 490 495

Pro Ala Cys Cys His Cys Pro Asp Pro His Ala
 500 505

<210> 16

<211> 2228

<212> DNA

<213> 智人

<400> 16

```

acaaggatgg gtccgcgagg cgcggcgagc ttgccccgag gccccggacc tcggcggctg      60
ctcctccccg tcgtcctccc gctgctgctg ctgctgttgt tggcgccgcc gggctcgggc      120
gccggggcca gccggccgcc ccacctggtc ttcttctgctg cagacgacct aggctggaac      180
gacgtcggtt tccacggctc ccgcatccgc acgccgcacc tggacgcgct ggcgccggc      240
ggggtgctcc tggacaacta ctacacgcag ccgctgtgca cgcgctcgcg gagccagctg      300
ctcactggcc gctaccagat ccgtacaggt ttacagcacc aaataatctg gccctgtcag      360
cccagctgtg ttctctgga tgaaaaactc ctgccccagc tcctaaaaga agcaggttat      420
actaccata tggtcggaaa atggcacctg ggaatgtacc ggaaagaatg cttccaacc      480
cgccgaggat ttgataccta ctttgatata ctctgggta gtgaagatta ttattcccat      540
gaacgctgta cattaattga cgctctgaat gtcacacgat gtgctcttga ttttcgagat      600
ggcgaagaag ttgcaacagg atataaaaat atgtattcaa caaacatatt caccaaaagg      660
gctatagccc tcataactaa ccatccacca gagaagcctc tgtttcteta ccttgctctc      720
cagtctgtgc atgagcccct tcaggtccct gaggaatact tgaagccata tgactttatc      780
caagacaaga acaggcatca ctatgcagga atgggtgtccc ttatggatga agcagtagga      840

```

[0037]

| | |
|--|------|
| aatgtcactg cagctttaa aagcagtggg ctctggaaca acacgggtgtt catcttttct | 900 |
| acagataacg gagggcagac tttggcaggg ggtaataact ggccccttcg aggaagaaaa | 960 |
| tggagcctgt gggaaggagg cgtccgaggg gtgggctttg tggcaagccc ctigtctgaag | 1020 |
| cagaagggcg tgaagaaccg ggagctcatc cacatctctg actggctgcc aacactcgtg | 1080 |
| aagctggcca ggggacacac caatggcaca aagcctctgg atggcttcga cgtgtggaaa | 1140 |
| accatcagtg aaggaagccc atccccaga attgagctgc tgcataatat tgacccaaac | 1200 |
| ttcgtggact cttcacctg tcccaggaac agcatggctc cagcaaagga tgactcttct | 1260 |
| cttcagaat attcagcctt taacacatct gtccatgctg caattagaca tggaaattgg | 1320 |
| aaactcctca cgggctaccc aggctgtggt tactggttcc ctccacgctc tcaatacaat | 1380 |
| gtttctgaga taccctcatc agaccacca accaagaccc tctggctctt tgatattgat | 1440 |
| cgggaccctg aagaaagaca tgacctgtcc agagaatata ctacatcgt cacaaagctc | 1500 |
| ctgtcccgcc tacagttcta ccataaacac tcagtccccg tgtacttccc tgcacaggac | 1560 |
| ccccgctgtg atccaaggc cactgggggtg tggggcctt ggatgtagga tttcaggag | 1620 |
| gctagaaaac ctttcaattg gaagttggac ctcaggcctt ttctcagac tcttgtctca | 1680 |
| tttgttatcc caacctgggt tcaactggcc cttctcttgc tcttaaacca caccgaggtg | 1740 |
| tctaatttca acccctaattg catttaagaa gctgataaaa tctgcaacac tcctgtctgt | 1800 |
| ggctggagca tgtgtctaga ggtgggggtg gctgggttta tcccccttc ctaagccttg | 1860 |
| ggacagctgg gaacttaact tgaatatagga agttctcact gaatcctgga ggctggaaca | 1920 |
| gctggctctt ttagactcac aagtcagacg ttcgattccc ctctgccaat agccagttt | 1980 |
| attggagtga atcacattc ttacgcaaat gaaggagca gacagtgatt aatggtctg | 2040 |
| ttggccaagg cltctccctg tcggtgaagg atcatgttca ggcactcaa gtgaaccacc | 2100 |
| cctcttgggt cacccttac tcaactatct catcacagag cataaggccc atttgtgtg | 2160 |

[0038]

tcagggtcaac agcaaaatgg cctgcacat gactgtggct tttaaataa agaaatgtgt 2220

ttttatcg 2228

<210> 17
 <211> 533
 <212> PRT
 <213> 智人

<400> 17

Met Gly Pro Arg Gly Ala Ala Ser Leu Pro Arg Gly Pro Gly Pro Arg
 1 5 10 15

Arg Leu Leu Leu Pro Val Val Leu Pro Leu Leu Leu Leu Leu Leu
 20 25 30

Ala Pro Pro Gly Ser Gly Ala Gly Ala Ser Arg Pro Pro His Leu Val
 35 40 45

Phe Leu Leu Ala Asp Asp Leu Gly Trp Asn Asp Val Gly Phe His Gly
 50 55 60

Ser Arg Ile Arg Thr Pro His Leu Asp Ala Leu Ala Ala Gly Gly Val
 65 70 75 80

Leu Leu Asp Asn Tyr Tyr Thr Gln Pro Leu Cys Thr Pro Ser Arg Ser
 85 90 95

Gln Leu Leu Thr Gly Arg Tyr Gln Ile Arg Thr Gly Leu Gln His Gln
 100 105 110

Ile Ile Trp Pro Cys Gln Pro Ser Cys Val Pro Leu Asp Glu Lys Leu
 115 120 125

[0039]

Leu Pro Gln Leu Leu Lys Glu Ala Gly Tyr Thr Thr His Met Val Gly
 130 135 140

Lys Trp His Leu Gly Met Tyr Arg Lys Glu Cys Leu Pro Thr Arg Arg
 145 150 155 160

Gly Phe Asp Thr Tyr Phe Gly Tyr Leu Leu Gly Ser Glu Asp Tyr Tyr
 165 170 175

Ser His Glu Arg Cys Thr Leu Ile Asp Ala Leu Asn Val Thr Arg Cys
 180 185 190

Ala Leu Asp Phe Arg Asp Gly Glu Glu Val Ala Thr Gly Tyr Lys Asn
 195 200 205

Met Tyr Ser Thr Asn Ile Phe Thr Lys Arg Ala Ile Ala Leu Ile Thr
 210 215 220

Asn His Pro Pro Glu Lys Pro Leu Phe Leu Tyr Leu Ala Leu Gln Ser
 225 230 235 240

Val His Glu Pro Leu Gln Val Pro Glu Glu Tyr Leu Lys Pro Tyr Asp
 245 250 255

Phe Ile Gln Asp Lys Asn Arg His His Tyr Ala Gly Met Val Ser Leu
 260 265 270

Met Asp Glu Ala Val Gly Asn Val Thr Ala Ala Leu Lys Ser Ser Gly
 275 280 285

Leu Trp Asn Asn Thr Val Phe Ile Phe Ser Thr Asp Asn Gly Gly Gln
 290 295 300

[0040]

Thr Leu Ala Gly Gly Asn Asn Trp Pro Leu Arg Gly Arg Lys Trp Ser
305 310 315 320

Leu Trp Glu Gly Gly Val Arg Gly Val Gly Phe Val Ala Ser Pro Leu
325 330 335

Leu Lys Gln Lys Gly Val Lys Asn Arg Glu Leu Ile His Ile Ser Asp
340 345 350

Trp Leu Pro Thr Leu Val Lys Leu Ala Arg Gly His Thr Asn Gly Thr
355 360 365

Lys Pro Leu Asp Gly Phe Asp Val Trp Lys Thr Ile Ser Glu Gly Ser
370 375 380

Pro Ser Pro Arg Ile Glu Leu Leu His Asn Ile Asp Pro Asn Phe Val
385 390 395 400

Asp Ser Ser Pro Cys Pro Arg Asn Ser Met Ala Pro Ala Lys Asp Asp
405 410 415

Ser Ser Leu Pro Glu Tyr Ser Ala Phe Asn Thr Ser Val His Ala Ala
420 425 430

Ile Arg His Gly Asn Trp Lys Leu Leu Thr Gly Tyr Pro Gly Cys Gly
435 440 445

Tyr Trp Phe Pro Pro Pro Ser Gln Tyr Asn Val Ser Glu Ile Pro Ser
450 455 460

Ser Asp Pro Pro Thr Lys Thr Leu Trp Leu Phe Asp Ile Asp Arg Asp
465 470 475 480

[0041]

Pro Glu Glu Arg His Asp Leu Ser Arg Glu Tyr Pro His Ile Val Thr
 485 490 495

Lys Leu Leu Ser Arg Leu Gln Phe Tyr His Lys His Ser Val Pro Val
 500 505 510

Tyr Phe Pro Ala Gln Asp Pro Arg Cys Asp Pro Lys Ala Thr Gly Val
 515 520 525

Trp Gly Pro Trp Met
 530

<210> 18

<211> 2401

<212> DNA

<213> 智人

<400> 18

gcctccagca gctgacggga ccagctgta gtgaggttgc agtgattgag taggattggc 60
 ctgcttcaaa gcagaggttt ctcattggaa tatgcttatt aaactccac tggcagcaa 120
 accatgaaca gaggatgaac aagtgaagt gcaatctct ccatcacagc tcagttcccc 180
 aacaacagga tcacaagctg gagatgcctt taaggaagat gaagatcctt ttctctctac 240
 tgttctttct gtgggaagcc gagagccacg cagcatcaag gccgaacatc atccttggtga 300
 tggctgacga cctcggcatt ggagatcctg ggtgctatgg gaacaaaact atcaggactc 360
 ccaatatcga ccggttgccc agtgggggag tgaactcac tcagcacctg gcagcatcac 420
 cgctgtgcac accaagcagg gcagccttca tgactggccg gtaccctgtc cgatcaggaa 480
 tggcatcttg gtcccgaact ggagttttcc tcttcacagc ctcttcggga ggacttccca 540
 ccgatgagat tacctttgct aagcttciga aggatcaagg ttattcaaca gcactgatag 600
 ggaaatggca ccttgggatg agctgtcaca gcaagactga cttctgtcac caccctttac 660

[0042]

| | |
|--|------|
| atcacggctt caattatttc tatgggatct ctttgaccaa tctgagagac tgcaagcccg | 720 |
| gagagggcag tgtcttcacc acgggcttca agaggtggt cttcctcccc ctgcagatcg | 780 |
| tcggggtcac cctccttacc cttgctgcac tcaattgtct ggggctactc cacgtgcctc | 840 |
| taggcgtttt tttcagcctt ctcttcttag cagccctaatt cctgaccctt ttcttgggct | 900 |
| tccttcatta cttccggccc ctgaactgct tcatgatgag gaactacgag atcattcagc | 960 |
| ageccatgtc ctatgacaat ctcacccaga ggctaacggt ggaggcggcc cagttcatac | 1020 |
| agcggaacac tgagactccg ttctgtcttg tcttgtccta cctccacgtg cacacagccc | 1080 |
| tgttctccag caaagacttt gctggcaaaa gtcaacacgg agtctacggg gatgctgttg | 1140 |
| aggaaatgga ctggagtgtg gggcagatct tgaaccttct ggatgagctg agattggcta | 1200 |
| atgataccct catctacttc acatcggacc agggagcaca tntagaggag gtgtcttcca | 1260 |
| aaggagaaat tcatggcggga agtaatggga tctataaagg aggaaaagca aacaactggg | 1320 |
| aaggaggtat ccgggttcca ggcatccttc gttggcccag ggtgatacag gctggccaga | 1380 |
| agattgatga gccactagc aacatggaca tatttcttac agtagccaag ctggctggag | 1440 |
| ctcccttgcc tgaggacagg atcattgatg gacgtgatct gatgccctg cttgaaggaa | 1500 |
| aaagccaacg ctccgatcat gaggttctct tccattactg caacgcctac ttaaagtctg | 1560 |
| tgcgctggca ccctcagaac agcacatcca tctggaagge cttttcttc accccaact | 1620 |
| tcaaccccgt gggttccaac ggatgctttg ccacacacgt gtgcttctgt ttcgggagtt | 1680 |
| atgtcaccca tcacgacca cctttactct ttgatattc caaagatccc agagagagaa | 1740 |
| accactaac tccagatcc gagccccgt tttatgaaat cctcaaagtc atgcaggaag | 1800 |
| ctgcggacag acacaccag accctgccag aggtgcccga tcagttttca tggaacaact | 1860 |
| ttctttggaa gccttggett cagctgtgct gtcttccac cggcctgtct tgccagtgtg | 1920 |
| atagagaaaa acaggataag agactgagcc gctagcagcg cctggggacc agacagacgc | 1980 |

[0043]

atgtggcaaa gctcaccatc ttcactacaa acacgcctga gagtggcact ggggaaacat 2040
 aactccatct acaccttga tttggactga ttctccattt tatcacctga aggcttgggc 2100
 cagagctcaa cagctactca actggagggg tgaggggat aaggtctgta gtatacagac 2160
 aggaagatgg taggtttatg ccttctgtgg ccagagtctt ggactcatgg aaatagaatg 2220
 aatagagggg cattcacaag gcacaccagt gcaagcagat gacaaaaagg tgcagaaggc 2280
 aatcttaaaa cagaaagggtg caggagggtac cttactcac ccctcagcaa atacctatgt 2340
 caacagtata agttaccatt tactctataa tctgcagtga tgcaataacc agcataataa 2400
 a 2401

<210> 19
 <211> 583
 <212> PRT
 <213> 智人

<400> 19

Met Pro Leu Arg Lys Met Lys Ile Pro Phe Leu Leu Leu Phe Phe Leu
 1 5 10 15

Trp Glu Ala Glu Ser His Ala Ala Ser Arg Pro Asn Ile Ile Leu Val
 20 25 30

Met Ala Asp Asp Leu Gly Ile Gly Asp Pro Gly Cys Tyr Gly Asn Lys
 35 40 45

Thr Ile Arg Thr Pro Asn Ile Asp Arg Leu Ala Ser Gly Gly Val Lys
 50 55 60

Leu Thr Gln His Leu Ala Ala Ser Pro Leu Cys Thr Pro Ser Arg Ala
 65 70 75 80

[0044]

Ala Phe Met Thr Gly Arg Tyr Pro Val Arg Ser Gly Met Ala Ser Trp
85 90 95

Ser Arg Thr Gly Val Phe Leu Phe Thr Ala Ser Ser Gly Gly Leu Pro
100 105 110

Thr Asp Glu Ile Thr Phe Ala Lys Leu Leu Lys Asp Gln Gly Tyr Ser
115 120 125

Thr Ala Leu Ile Gly Lys Trp His Leu Gly Met Ser Cys His Ser Lys
130 135 140

Thr Asp Phe Cys His His Pro Leu His His Gly Phe Asn Tyr Phe Tyr
145 150 155 160

Gly Ile Ser Leu Thr Asn Leu Arg Asp Cys Lys Pro Gly Glu Gly Ser
165 170 175

Val Phe Thr Thr Gly Phe Lys Arg Leu Val Phe Leu Pro Leu Gln Ile
180 185 190

Val Gly Val Thr Leu Leu Thr Leu Ala Ala Leu Asn Cys Leu Gly Leu
195 200 205

Leu His Val Pro Leu Gly Val Phe Phe Ser Leu Leu Phe Leu Ala Ala
210 215 220

Leu Ile Leu Thr Leu Phe Leu Gly Phe Leu His Tyr Phe Arg Pro Leu
225 230 235 240

Asn Cys Phe Met Met Arg Asn Tyr Glu Ile Ile Gln Gln Pro Met Ser
245 250 255

[0045]

Tyr Asp Asn Leu Thr Gln Arg Leu Thr Val Glu Ala Ala Gln Phe Ile
 260 265 270

Gln Arg Asn Thr Glu Thr Pro Phe Leu Leu Val Leu Ser Tyr Leu His
 275 280 285

Val His Thr Ala Leu Phe Ser Ser Lys Asp Phe Ala Gly Lys Ser Gln
 290 295 300

His Gly Val Tyr Gly Asp Ala Val Glu Glu Met Asp Trp Ser Val Gly
 305 310 315 320

Gln Ile Leu Asn Leu Leu Asp Glu Leu Arg Leu Ala Asn Asp Thr Leu
 325 330 335

Ile Tyr Phe Thr Ser Asp Gln Gly Ala His Val Glu Glu Val Ser Ser
 340 345 350

Lys Gly Glu Ile His Gly Gly Ser Asn Gly Ile Tyr Lys Gly Gly Lys
 355 360 365

Ala Asn Asn Trp Glu Gly Gly Ile Arg Val Pro Gly Ile Leu Arg Trp
 370 375 380

Pro Arg Val Ile Gln Ala Gly Gln Lys Ile Asp Glu Pro Thr Ser Asn
 385 390 395 400

Met Asp Ile Phe Pro Thr Val Ala Lys Leu Ala Gly Ala Pro Leu Pro
 405 410 415

Glu Asp Arg Ile Ile Asp Gly Arg Asp Leu Met Pro Leu Leu Glu Gly
 420 425 430

[0046]

Lys Ser Gln Arg Ser Asp His Glu Phe Leu Phe His Tyr Cys Asn Ala
 435 440 445

Tyr Leu Asn Ala Val Arg Trp His Pro Gln Asn Ser Thr Ser Ile Trp
 450 455 460

Lys Ala Phe Phe Phe Thr Pro Asn Phe Asn Pro Val Gly Ser Asn Gly
 465 470 475 480

Cys Phe Ala Thr His Val Cys Phe Cys Phe Gly Ser Tyr Val Thr His
 485 490 495

His Asp Pro Pro Leu Leu Phe Asp Ile Ser Lys Asp Pro Arg Glu Arg
 500 505 510

Asn Pro Leu Thr Pro Ala Ser Glu Pro Arg Phe Tyr Glu Ile Leu Lys
 515 520 525

Val Met Gln Glu Ala Ala Asp Arg His Thr Gln Thr Leu Pro Glu Val
 530 535 540

Pro Asp Gln Phe Ser Trp Asn Asn Phe Leu Trp Lys Pro Trp Leu Gln
 545 550 555 560

Leu Cys Cys Pro Ser Thr Gly Leu Ser Cys Gln Cys Asp Arg Glu Lys
 565 570 575

Gln Asp Lys Arg Leu Ser Arg
 580

<210> 20
 <211> 1945
 <212> DNA
 <213> 智人

[0047]

| | |
|--|------|
| <400> 20 | |
| ggaagccttg gcactagcgg cgcccgggcg cggagtgcgc agggcaaggt cctgcgctct | 60 |
| gggccagcgc tcggccatgc gatccgccgc ggggagggga cgcgccgcgc ccgccccag | 120 |
| ggactctttg ccggtgctac tgtttttatg cttgcttctg aagacgtgtg aacctaaaac | 180 |
| tgcaaatgcc tttaaaccaa atactctact gatcatggcg gatgatctag gcactgggga | 240 |
| tctcggttgc tacgggaaca atacactgag aacgccgaat attgaccagc ttgcagagga | 300 |
| aggtgtgagg ctactcagc acctggcgcg cgcgccgctc tgcaccccaa gccgagctgc | 360 |
| attcctcaca gggagacatt ccttcagatc aggcattggac gccagcaatg gataccgggc | 420 |
| ccttcagtgg aacgcaggct cagggtgact ccctgagaac gaaaccactt ttgcaagaat | 480 |
| cttcagcag catggctatg caaccggcct cataggaaaa tggcaccagg gtgtgaattg | 540 |
| tgcattcccgc ggggatcact gccaccacc cctgaaccac ggatttgact atttctacgg | 600 |
| catgcccttc acgctcaca acgactgtga cccaggcagg ccccccgaag tggacgccgc | 660 |
| cctgagggcg cagctctggg gttacacca gttcctggcg ctggggattc tcacctggc | 720 |
| tgccggccag acctgcggtt tcttctctgt ctccgcgaga gcagtcaccg gcatggccgg | 780 |
| cgtgggctgc ctgtttttca tctcttgta ctctccttc gggtttgtgc gacgctggaa | 840 |
| ctgtatctg atgagaaacc atgacgtcac ggagcaacc atggttctgg agaaaacagc | 900 |
| gagtcttatg ctaaaggaag ctgtttccta tattgaaaga cacaagcatg ggccatttct | 960 |
| cctcttctt tctttgctgc atgtgcacat tccccttgatg accacgagtg cattcctggg | 1020 |
| gaaaagtcag catggcttat atggtgataa tgtggaggag atggactggc tcataggtaa | 1080 |
| ggttcttaat gccatcgaag acaatggttt aaagaactca acattcacgt atttcacctc | 1140 |
| tgaccatgga ggacatttag aggcaagaga tggacacagc cagttagggg gatggaacgg | 1200 |
| aatttacaaa ggtgggaagg gcatgggagg atgggaaggt gggatccgag tgcccgggat | 1260 |

[0048]

cttccactgg ccggggggtgc tcccggccgg ccgagtgatt ggagagccca cgagcctgat 1320
 ggacgtgttc cctactgtgg tccagctggt ggggtggcgag gtgccccagg acagggtgat 1380
 tgatggccac agcctgggtac ccttgctgca gggagctgag gcacgctcgg cacatgagtt 1440
 cctgtttcat tactgtgggc agcatcttca cgcagcacgc tggcaccaga aggacagtgg 1500
 aagcgtctgg aaggttcatt acacgacccc gcagttccac cccgaggagc ggggcctgct 1560
 aacggccgag gcgtctgccc atgctgaatg gggaggcgtg acccatcaca gaccccttt 1620
 gctctttgac ctctccaggg acccctccga ggcacggccc ctgacccccg actccgagcc 1680
 cctgtaccac gccgtgatag caaggtagg tgccgcggtg tcggagcatc ggcagaccct 1740
 gagtcctgtg ccccagcagt tttccatgag caacatcctg tggaagccgt ggctgcagcc 1800
 gtgctgcgga catttcccgt tctgttcatg ccacaggat ggggatggca ccccctgaat 1860
 gccaggactg tgagagagga tccaggagag cctgactgcg ttgcaaaca aatttccaa 1920
 gcttggttct atcttcagtc cggaa 1945

<210> 21
 <211> 593
 <212> PRT
 <213> 智人

<400> 21

Met Arg Ser Ala Ala Arg Arg Gly Arg Ala Ala Pro Ala Ala Arg Asp
 1 5 10 15

Ser Leu Pro Val Leu Leu Phe Leu Cys Leu Leu Leu Lys Thr Cys Glu
 20 25 30

Pro Lys Thr Ala Asn Ala Phe Lys Pro Asn Ile Leu Leu Ile Met Ala
 35 40 45

[0049]

Asp Asp Leu Gly Thr Gly Asp Leu Gly Cys Tyr Gly Asn Asn Thr Leu
50 55 60

Arg Thr Pro Asn Ile Asp Gln Leu Ala Glu Glu Gly Val Arg Leu Thr
65 70 75 80

Gln His Leu Ala Ala Ala Pro Leu Cys Thr Pro Ser Arg Ala Ala Phe
85 90 95

Leu Thr Gly Arg His Ser Phe Arg Ser Gly Met Asp Ala Ser Asn Gly
100 105 110

Tyr Arg Ala Leu Gln Trp Asn Ala Gly Ser Gly Gly Leu Pro Glu Asn
115 120 125

Glu Thr Thr Phe Ala Arg Ile Leu Gln Gln His Gly Tyr Ala Thr Gly
130 135 140

Leu Ile Gly Lys Trp His Gln Gly Val Asn Cys Ala Ser Arg Gly Asp
145 150 155 160

His Cys His His Pro Leu Asn His Gly Phe Asp Tyr Phe Tyr Gly Met
165 170 175

Pro Phe Thr Leu Thr Asn Asp Cys Asp Pro Gly Arg Pro Pro Glu Val
180 185 190

Asp Ala Ala Leu Arg Ala Gln Leu Trp Gly Tyr Thr Gln Phe Leu Ala
195 200 205

Leu Gly Ile Leu Thr Leu Ala Ala Gly Gln Thr Cys Gly Phe Phe Ser
210 215 220

[0050]

Val Leu Pro Ala Gly Arg Val Ile Gly Glu Pro Thr Ser Leu Met Asp
 405 410 415

Val Phe Pro Thr Val Val Gln Leu Val Gly Gly Glu Val Pro Gln Asp
 420 425 430

Arg Val Ile Asp Gly His Ser Leu Val Pro Leu Leu Gln Gly Ala Glu
 435 440 445

Ala Arg Ser Ala His Glu Phe Leu Phe His Tyr Cys Gly Gln His Leu
 450 455 460

His Ala Ala Arg Trp His Gln Lys Asp Ser Gly Ser Val Trp Lys Val
 465 470 475 480

His Tyr Thr Thr Pro Gln Phe His Pro Glu Glu Arg Gly Leu Leu Thr
 485 490 495

Ala Glu Ala Ser Ala His Ala Glu Trp Gly Gly Val Thr His His Arg
 500 505 510

Pro Pro Leu Leu Phe Asp Leu Ser Arg Asp Pro Ser Glu Ala Arg Pro
 515 520 525

Leu Thr Pro Asp Ser Glu Pro Leu Tyr His Ala Val Ile Ala Arg Val
 530 535 540

Gly Ala Ala Val Ser Glu His Arg Gln Thr Leu Ser Pro Val Pro Gln
 545 550 555 560

Gln Phe Ser Met Ser Asn Ile Leu Trp Lys Pro Trp Leu Gln Pro Cys
 565 570 575

[0052]

Cys Gly His Phe Pro Phe Cys Ser Cys His Glu Asp Gly Asp Gly Thr
 580 585 590

Pro

<210> 22

<211> 1858

<212> DNA

<213> 智人

<400> 22

```

ccttcctctt cttgatcggg gattcaggaa ggagcccagg agcagaggaa gtagagagag      60
agacaacatg ttacatctgc accattcttg tttgtgtttc aggagctggc tgccagcgat      120
gctcgtgta ctgctaagtt tggcaccatc agcttccagc gacatttccg cctcccgacc      180
gaacatcctt cttctgatgg cggacgacct tggcattggg gacattggct gctatggcaa      240
caacaccatg aggactcega atattgaccg ctttcagag gacggcgtga agctgaccca      300
acacatctct gccgcatctt tgtgcacccc aagcagagcc gccttctca cgggcagata      360
ccctgtgcca tcagggatgg tttccagcat tggttaccgt gttcttcagt ggaccggagc      420
atctggaggt ctccaacaa atgagacaac ttttgcaaaa atactgaaag agaaaggcta      480
tgccactgga ctattggaa aatggcatct gggctctaac tgtgagtcag ccagtgatca      540
ttgccaccac cctctccatc atggctttga gcatttctac ggaatgcctt tctccttgat      600
gggtgattgc gcccgctggg aactctcaga gaagcgtgtc aacctggaac aaaaactcaa      660
cttctcttc caagtcctgg ctttggttgc cctcacactg gtagcaggga agctcacaca      720
cctgataccc gtctcgtgga tgccggtcat ctggtcagcc ctttcggccg tctctctcct      780
cgcaagetcc tattttgtgg gtgctctgat tgtccatgcc gattgctttc tgatgagaaa      840
ccacaccatc acggagcagc ccatgtgctt ccaagaacg acaccctta ttctgcagga      900

```

[0053]

ggttgcgtcc ttctcaaaa ggaataagca tgggccttcc ctctctttg tttcctttct 960
 acacgttcac atccctctta tcactatgga gaacttcctc gggaagagtc tccacgggct 1020
 gtatggggac aacgtagagg agatggactg gatggtagga cggatccttg acactttgga 1080
 cgtggagggt ttgagcaaca gcacctcat ttattttacg tcggatcacg gcggttcctt 1140
 agagaatcaa ctggaaaca cccagtatgg tggtggaat ggaattata aaggtgggaa 1200
 gggcatggga ggatgggaag gtgggatccg cgtgccggg atcttccgct ggcccggggt 1260
 gctcccggcc ggccgagtga ttggcgagcc cacgagtctg atggacgtgt tccccaccgt 1320
 ggtccggctg gcggcgggcg aggtgcccc aagacagagt attgacggcc aagaccttct 1380
 gcccttgctc ctggggacag cccaacactc agaccacgag ttctgatgc attattgtga 1440
 gaggtttctg cacgcagcca ggtggcatca acgggacaga ggaacaatgt ggaaagtcca 1500
 ctttgtagc cctgtgttcc agccagaggg agccggtgcc tgctatggaa gaaaggtctg 1560
 cccgtgcttt ggggaaaaag tagtccacca cgatccacct ttgctctttg acctctcaag 1620
 agaccttct gagaccaca tcctcacacc agcctcagag cccgtgttct atcaggtgat 1680
 ggaacgagtc cagcaggcgg tgtgggaaca ccagcggaca ctcagcccag ttctctgca 1740
 gctggacagg ctgggcaaca tctggagacc gtggctgcag ccttgctgtg gcccgttccc 1800
 cctctgctgg tgccttaggg aagatgacc acaataaatg tctgcagtga aaagctgg 1858

<210> 23
 <211> 589
 <212> PRT
 <213> 智人

<400> 23

Met Leu His Leu His His Ser Cys Leu Cys Phe Arg Ser Trp Leu Pro
 1 5 10 15

[0054]

Ala Met Leu Ala Val Leu Leu Ser Leu Ala Pro Ser Ala Ser Ser Asp
 20 25 30

Ile Ser Ala Ser Arg Pro Asn Ile Leu Leu Leu Met Ala Asp Asp Leu
 35 40 45

Gly Ile Gly Asp Ile Gly Cys Tyr Gly Asn Asn Thr Met Arg Thr Pro
 50 55 60

Asn Ile Asp Arg Leu Ala Glu Asp Gly Val Lys Leu Thr Gln His Ile
 65 70 75 80

Ser Ala Ala Ser Leu Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly
 85 90 95

Arg Tyr Pro Val Arg Ser Gly Met Val Ser Ser Ile Gly Tyr Arg Val
 100 105 110

Leu Gln Trp Thr Gly Ala Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr
 115 120 125

Phe Ala Lys Ile Leu Lys Glu Lys Gly Tyr Ala Thr Gly Leu Ile Gly
 130 135 140

Lys Trp His Leu Gly Leu Asn Cys Glu Ser Ala Ser Asp His Cys His
 145 150 155 160

His Pro Leu His His Gly Phe Glu His Phe Tyr Gly Met Pro Phe Ser
 165 170 175

Leu Met Gly Asp Cys Ala Arg Trp Glu Leu Ser Glu Lys Arg Val Asn
 180 185 190

[0055]

Leu Glu Gln Lys Leu Asn Phe Leu Phe Gln Val Leu Ala Leu Val Ala
 195 200 205

Leu Thr Leu Val Ala Gly Lys Leu Thr His Leu Ile Pro Val Ser Trp
 210 215 220

Met Pro Val Ile Trp Ser Ala Leu Ser Ala Val Leu Leu Leu Ala Ser
 225 230 235 240

Ser Tyr Phe Val Gly Ala Leu Ile Val His Ala Asp Cys Phe Leu Met
 245 250 255

Arg Asn His Thr Ile Thr Glu Gln Pro Met Cys Phe Gln Arg Thr Thr
 260 265 270

Pro Leu Ile Leu Gln Glu Val Ala Ser Phe Leu Lys Arg Asn Lys His
 275 280 285

Gly Pro Phe Leu Leu Phe Val Ser Phe Leu His Val His Ile Pro Leu
 290 295 300

Ile Thr Met Glu Asn Phe Leu Gly Lys Ser Leu His Gly Leu Tyr Gly
 305 310 315 320

Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Arg Ile Leu Asp Thr
 325 330 335

Leu Asp Val Glu Gly Leu Ser Asn Ser Thr Leu Ile Tyr Phe Thr Ser
 340 345 350

Asp His Gly Gly Ser Leu Glu Asn Gln Leu Gly Asn Thr Gln Tyr Gly
 355 360 365

[0056]

Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu
370 375 380

Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Gly Val Leu Pro
385 390 395 400

Ala Gly Arg Val Ile Gly Glu Pro Thr Ser Leu Met Asp Val Phe Pro
405 410 415

Thr Val Val Arg Leu Ala Gly Gly Glu Val Pro Gln Asp Arg Val Ile
420 425 430

Asp Gly Gln Asp Leu Leu Pro Leu Leu Leu Gly Thr Ala Gln His Ser
435 440 445

Asp His Glu Phe Leu Met His Tyr Cys Glu Arg Phe Leu His Ala Ala
450 455 460

Arg Trp His Gln Arg Asp Arg Gly Thr Met Trp Lys Val His Phe Val
465 470 475 480

Thr Pro Val Phe Gln Pro Glu Gly Ala Gly Ala Cys Tyr Gly Arg Lys
485 490 495

Val Cys Pro Cys Phe Gly Glu Lys Val Val His His Asp Pro Pro Leu
500 505 510

Leu Phe Asp Leu Ser Arg Asp Pro Ser Glu Thr His Ile Leu Thr Pro
515 520 525

Ala Ser Glu Pro Val Phe Tyr Gln Val Met Glu Arg Val Gln Gln Ala
530 535 540

[0057]

| | |
|--|------|
| ccacacgtcc cctttatact gggactgcct cctcatgcgg gggcacgaga tcacggagca | 840 |
| gccccatgaag gctgaacgag ctggatccat tatggtgaag gaagcgattt cctttttaga | 900 |
| aaggcacagt aaggaaactt tccttctctt tttctccttt cttcacgtgc acacacctct | 960 |
| ccccaccacg gacgatttca ctggcaccag caagcatggc ttgtatgggg ataalgtaga | 1020 |
| agagatggac tccatggtgg gcaagattct tgatgctatc gatgattttg gcctaaggaa | 1080 |
| caacaccctt gtctacttta catcagatca cggagggcat ttggaagcta ggcgagggca | 1140 |
| tgccccactt ggtggatgga atggaatata caaagggtgga aaaggcatgg ggggctggga | 1200 |
| aggtggaatc cgcgtcccag gaattgtccg atggcctgga aaggtaccag ctggacggtt | 1260 |
| gattaaggaa cctacaagt taatggatat ttaccaact gtcgcatcag tgtcaggagg | 1320 |
| aagtctccct caggacaggg tcattgacgg ccgagacctc atgcccttgc tgcagggcaa | 1380 |
| cgtcaggcac tcggagcatg aatttctttt cactactgt ggctcctacc tgcacgccgt | 1440 |
| gcgggtgatc cccaaggacg acagtgggtc agtttgaag gctcactatg tgacccccgt | 1500 |
| attccagcca ccagcttctg gtggctgcta tgtcacctca ttatgcagat gtttcggaga | 1560 |
| acaggttacc taccacaacc cccctctgct ettcgatctc tccagggacc cctcagagtc | 1620 |
| cacaccctg acacctgcca cagagcccct ctatgatttt gtgattaana aggtggccaa | 1680 |
| cgccctgaag gaacaccagg aaaccatcgt gcctgtgacc taccaactct cagaactgaa | 1740 |
| tcagggcagg acgtggctga agccttgctg tgggggttcc ccattttgtc tgtgtgacaa | 1800 |
| ggaagaggaa gtctctcagc ctcgggttcc taacgagaag agataattac aatcaggcta | 1860 |
| ccagaggaag cctttggtcc taacgagaag agataattac aatcaggcta ccaaaggaag | 1920 |
| cactaacttt ggtgctttca agttggcaag gagtgcattt aatagtcaat aaattcatct | 1980 |
| accattccag attatt | 1996 |

<210> 25

[0059]

<211> 591
 <212> PRT
 <213> 智人

<400> 25

Met Arg Pro Arg Arg Pro Leu Val Phe Met Ser Leu Val Cys Ala Leu
 1 5 10 15

Leu Asn Thr Trp Pro Gly His Thr Gly Cys Met Thr Thr Arg Pro Asn
 20 25 30

Ile Val Leu Ile Met Val Asp Asp Leu Gly Ile Gly Asp Leu Gly Cys
 35 40 45

Tyr Gly Asn Asp Thr Met Arg Thr Pro His Ile Asp Arg Leu Ala Arg
 50 55 60

Glu Gly Val Arg Leu Thr Gln His Ile Ser Ala Ala Ser Leu Cys Ser
 65 70 75 80

Pro Ser Arg Ser Ala Phe Leu Thr Gly Arg Tyr Pro Ile Arg Ser Gly
 85 90 95

Met Val Ser Ser Gly Asn Arg Arg Val Ile Gln Asn Leu Ala Val Pro
 100 105 110

Ala Gly Leu Pro Leu Asn Glu Thr Thr Leu Ala Ala Leu Leu Lys Lys
 115 120 125

Gln Gly Tyr Ser Thr Gly Leu Ile Gly Lys Trp His Gln Gly Leu Asn
 130 135 140

Cys Asp Ser Arg Ser Asp Gln Cys His His Pro Tyr Asn Tyr Gly Phe
 145 150 155 160

[0060]

Asp Tyr Tyr Tyr Gly Met Pro Phe Thr Leu Val Asp Ser Cys Trp Pro
 165 170 175

Asp Pro Ser Arg Asn Thr Glu Leu Ala Phe Glu Ser Gln Leu Trp Leu
 180 185 190

Cys Val Gln Leu Val Ala Ile Ala Ile Leu Thr Leu Thr Phe Gly Lys
 195 200 205

Leu Ser Gly Trp Val Ser Val Pro Trp Leu Leu Ile Phe Ser Met Ile
 210 215 220

Leu Phe Ile Phe Leu Leu Gly Tyr Ala Trp Phe Ser Ser His Thr Ser
 225 230 235 240

Pro Leu Tyr Trp Asp Cys Leu Leu Met Arg Gly His Glu Ile Thr Glu
 245 250 255

Gln Pro Met Lys Ala Glu Arg Ala Gly Ser Ile Met Val Lys Glu Ala
 260 265 270

Ile Ser Phe Leu Glu Arg His Ser Lys Glu Thr Phe Leu Leu Phe Phe
 275 280 285

Ser Phe Leu His Val His Thr Pro Leu Pro Thr Thr Asp Asp Phe Thr
 290 295 300

Gly Thr Ser Lys His Gly Leu Tyr Gly Asp Asn Val Glu Glu Met Asp
 305 310 315 320

Ser Met Val Gly Lys Ile Leu Asp Ala Ile Asp Asp Phe Gly Leu Arg
 325 330 335

[0061]

Asn Asn Thr Leu Val Tyr Phe Thr Ser Asp His Gly Gly His Leu Glu
 340 345 350

Ala Arg Arg Gly His Ala Gln Leu Gly Gly Trp Asn Gly Ile Tyr Lys
 355 360 365

Gly Gly Lys Gly Met Gly Gly Trp Glu Gly Gly Ile Arg Val Pro Gly
 370 375 380

Ile Val Arg Trp Pro Gly Lys Val Pro Ala Gly Arg Leu Ile Lys Glu
 385 390 395 400

Pro Thr Ser Leu Met Asp Ile Leu Pro Thr Val Ala Ser Val Ser Gly
 405 410 415

Gly Ser Leu Pro Gln Asp Arg Val Ile Asp Gly Arg Asp Leu Met Pro
 420 425 430

Leu Leu Gln Gly Asn Val Arg His Ser Glu His Glu Phe Leu Phe His
 435 440 445

Tyr Cys Gly Ser Tyr Leu His Ala Val Arg Trp Ile Pro Lys Asp Asp
 450 455 460

Ser Gly Ser Val Trp Lys Ala His Tyr Val Thr Pro Val Phe Gln Pro
 465 470 475 480

Pro Ala Ser Gly Gly Cys Tyr Val Thr Ser Leu Cys Arg Cys Phe Gly
 485 490 495

Glu Gln Val Thr Tyr His Asn Pro Pro Leu Leu Phe Asp Leu Ser Arg
 500 505 510

[0062]

Asp Pro Ser Glu Ser Thr Pro Leu Thr Pro Ala Thr Glu Pro Leu Tyr
 515 520 525

Asp Phe Val Ile Lys Lys Val Ala Asn Ala Leu Lys Glu His Gln Glu
 530 535 540

Thr Ile Val Pro Val Thr Tyr Gln Leu Ser Glu Leu Asn Gln Gly Arg
 545 550 555 560

Thr Trp Leu Lys Pro Cys Cys Gly Val Phe Pro Phe Cys Leu Cys Asp
 565 570 575

Lys Glu Glu Glu Val Ser Gln Pro Arg Gly Pro Asn Glu Lys Arg
 580 585 590

<210> 26

<211> 1578

<212> DNA

<213> 智人

<400> 26

atgggctggc tttttctaaa ggttttggtg gcgggagtga gtttctcagg atttctttat 60
 cctcttgtgg atttttgcat cagtgggaaa acaagaggac agaagccaaa ctttgtgatt 120
 attttgccg atgacatggg gtggggtgac ctgggagcaa actgggcaga aacaaaggac 180
 actgccaacc ttgataagat ggcttcggag ggaatgaggt ttgtggattt ccatgcagct 240
 gctccacct gctcacctc cgggcttcc ttgctcaccg gccggcttgg ccttcgcaat 300
 ggagtcacac gcaacttgc agtcacttct gtgggaggcc ttccgctcaa cgagaccacc 360
 ttggcagagg tgctgcagca ggcgggttac gtcactggga taataggcaa atggcatctt 420
 ggacaccacg gctcttatca cccaacttc cgtggttttg attactactt tggaatccca 480

[0063]

| | |
|--|------|
| tatagccatg atatgggctg tactgatact ccaggctaca accacctcc ttgtccagcg | 540 |
| tgccacagg gtgatggacc atcaaggaac cttcaaagag actgttacac tgacgtggcc | 600 |
| ctccctcttt atgaaaacct caacattgtg gagcagccgg tgaacttgag cagccttgcc | 660 |
| cagaagtatg ctgagaaagc aaccagttc atccagcgtg caagcaccag cgggagggcc | 720 |
| ttcctgctct atgtggctct ggcccacatg cacgtgcct tacctgtgac tcagctacca | 780 |
| gcagcgccac ggggcagaag cctgtatggt gcagggctct gggagatgga cagtctggtg | 840 |
| ggccagatca aggacaaagt tgaccacaca gtgaaggaaa acacattcct ctggtttaca | 900 |
| ggagacaatg gcccgtgggc tcagaagtgt gagctagcgg gcagtgtggg tcccttcaact | 960 |
| ggatthttggc aaactcgtca agggggaagt ccagccaagc agacgacctg ggaaggaggg | 1020 |
| caccgggtcc cagcactggc tiactggcct ggcagagttc cagttaatgt caccagcact | 1080 |
| gccttgttaa gcgtgctgga catthttcca actgtggtag ccctggccca ggccagctta | 1140 |
| cctcaaggac ggcgctttga tgggtggtgac gtctccgagg tgctctttgg ccggtcacag | 1200 |
| cctgggcaca ggggtgctgtt ccaccccaac agcggggcag ctggagagtt tggagccctg | 1260 |
| cagactgtcc gcctggagcg ttacaaggcc ttctacatta ccggtggagc cagggcgtgt | 1320 |
| gatgggagca cggggcctga gctgcagcat aagthtctc tgatthtcaa cctggaagac | 1380 |
| gataccgcag aagctgtgcc cctagaaaga ggtggtgctg agtaccaggc tgtgctgccc | 1440 |
| gaggtcagaa aggttcttgc agacgtcctc caagacattg ccaacgacaa catctccage | 1500 |
| gcagattaca ctcaggacc tttagtaact ccctgctgta atccctacca aattgcctgc | 1560 |
| cgctgtcaag ccgcataa | 1578 |

<210> 27

<211> 525

<212> PRT

<213> 智人

[0064]

<400> 27

Met Gly Trp Leu Phe Leu Lys Val Leu Leu Ala Gly Val Ser Phe Ser
 1 5 10 15

Gly Phe Leu Tyr Pro Leu Val Asp Phe Cys Ile Ser Gly Lys Thr Arg
 20 25 30

Gly Gln Lys Pro Asn Phe Val Ile Ile Leu Ala Asp Asp Met Gly Trp
 35 40 45

Gly Asp Leu Gly Ala Asn Trp Ala Glu Thr Lys Asp Thr Ala Asn Leu
 50 55 60

Asp Lys Met Ala Ser Glu Gly Met Arg Phe Val Asp Phe His Ala Ala
 65 70 75 80

Ala Ser Thr Cys Ser Pro Ser Arg Ala Ser Leu Leu Thr Gly Arg Leu
 85 90 95

Gly Leu Arg Asn Gly Val Thr Arg Asn Phe Ala Val Thr Ser Val Gly
 100 105 110

Gly Leu Pro Leu Asn Glu Thr Thr Leu Ala Glu Val Leu Gln Gln Ala
 115 120 125

Gly Tyr Val Thr Gly Ile Ile Gly Lys Trp His Leu Gly His His Gly
 130 135 140

Ser Tyr His Pro Asn Phe Arg Gly Phe Asp Tyr Tyr Phe Gly Ile Pro
 145 150 155 160

Tyr Ser His Asp Met Gly Cys Thr Asp Thr Pro Gly Tyr Asn His Pro
 165 170 175

[0065]

Pro Cys Pro Ala Cys Pro Gln Gly Asp Gly Pro Ser Arg Asn Leu Gln
 180 185 190

Arg Asp Cys Tyr Thr Asp Val Ala Leu Pro Leu Tyr Glu Asn Leu Asn
 195 200 205

Ile Val Glu Gln Pro Val Asn Leu Ser Ser Leu Ala Gln Lys Tyr Ala
 210 215 220

Glu Lys Ala Thr Gln Phe Ile Gln Arg Ala Ser Thr Ser Gly Arg Pro
 225 230 235 240

Phe Leu Leu Tyr Val Ala Leu Ala His Met His Val Pro Leu Pro Val
 245 250 255

Thr Gln Leu Pro Ala Ala Pro Arg Gly Arg Ser Leu Tyr Gly Ala Gly
 260 265 270

Leu Trp Glu Met Asp Ser Leu Val Gly Gln Ile Lys Asp Lys Val Asp
 275 280 285

His Thr Val Lys Glu Asn Thr Phe Leu Trp Phe Thr Gly Asp Asn Gly
 290 295 300

Pro Trp Ala Gln Lys Cys Glu Leu Ala Gly Ser Val Gly Pro Phe Thr
 305 310 315 320

Gly Phe Trp Gln Thr Arg Gln Gly Gly Ser Pro Ala Lys Gln Thr Thr
 325 330 335

Trp Glu Gly Gly His Arg Val Pro Ala Leu Ala Tyr Trp Pro Gly Arg
 340 345 350

[0066]

Val Pro Val Asn Val Thr Ser Thr Ala Leu Leu Ser Val Leu Asp Ile
 355 360 365

Phe Pro Thr Val Val Ala Leu Ala Gln Ala Ser Leu Pro Gln Gly Arg
 370 375 380

Arg Phe Asp Gly Val Asp Val Ser Glu Val Leu Phe Gly Arg Ser Gln
 385 390 395 400

Pro Gly His Arg Val Leu Phe His Pro Asn Ser Gly Ala Ala Gly Glu
 405 410 415

Phe Gly Ala Leu Gln Thr Val Arg Leu Glu Arg Tyr Lys Ala Phe Tyr
 420 425 430

Ile Thr Gly Gly Ala Arg Ala Cys Asp Gly Ser Thr Gly Pro Glu Leu
 435 440 445

Gln His Lys Phe Pro Leu Ile Phe Asn Leu Glu Asp Asp Thr Ala Glu
 450 455 460

Ala Val Pro Leu Glu Arg Gly Gly Ala Glu Tyr Gln Ala Val Leu Pro
 465 470 475 480

Glu Val Arg Lys Val Leu Ala Asp Val Leu Gln Asp Ile Ala Asn Asp
 485 490 495

Asn Ile Ser Ser Ala Asp Tyr Thr Gln Asp Pro Ser Val Thr Pro Cys
 500 505 510

Cys Asn Pro Tyr Gln Ile Ala Cys Arg Cys Gln Ala Ala
 515 520 525

[0067]

| | | |
|------------|--|------|
| <210> | 28 | |
| <211> | 4669 | |
| <212> | DNA | |
| <213> | 智人 | |
| <400> | 28 | |
| cgcagaccgt | cgctaatgaa tcttggggcc ggtgtcgggc cggggcggct tgatcgcaa | 60 |
| ctaggaaacc | ccaggcgcag aggccaggag cgagggcagc gaggatcaga ggccaggcct | 120 |
| tcccggctgc | cggcgctcct cggaggtcag ggcagatgag gaacatgact ctcccccttc | 180 |
| ggaggaggaa | ggaagtcccc ctgccacctt atctctgctc ctctgcctcc tccctgttcc | 240 |
| cagagctttt | tctctagaga agattttgaa ggcggctttt gtgctgacgg ccaccacca | 300 |
| tcacttaaag | aagataaact tggcaaatga catgcaggtt cttcaaggca gaataattgc | 360 |
| agaaaatctt | caaaggacct tatctgcaga tgttctgaat acctctgaga atagagattg | 420 |
| attattcaac | caggatacct aattcaagaa ctccagaaat caggagacgg agacattttg | 480 |
| tcagttttgc | aacattggac caaatacaat gaagtattct tgctgtgctc tggttttggc | 540 |
| tgtcctgggc | acagaattgc tgggaagcct ctgttcgact gtcagatccc cgaggttcag | 600 |
| aggacggata | cagcaggaac gaaaaaacat cggacceaac attattcttg tgcttaccga | 660 |
| tgatcaagat | gtggagctgg ggtccctgca agtcatgaac aaaacgagaa agattatgga | 720 |
| acatgggggg | gccaccttca tcaatgcctt tgtgactaca cccatgtgct gcccgtcacg | 780 |
| gtctccatg | ctcaccggga agtatgtgca caatcacaat gtctacacca acaacgagaa | 840 |
| ctgtctttcc | ccctcgtggc aggccatgca tgagcctcgg acttttctg tatacttaa | 900 |
| caacactggc | tacagaacag ccttttttgg aaaatacctc aatgaatata atggcagcta | 960 |
| catccccct | gggtggcgag aatggcttgg attaatacag aattctcgtc ttataatta | 1020 |
| cactgtttgt | cgcaatggca tcaaagaaa gcatggattt gattatgcaa aggactactt | 1080 |

[0068]

| | |
|--|------|
| cacagactta atcactaacg agagcattaa ttacttcaaa atgtctaaga gaatgtatcc | 1140 |
| ccataggccc gttatgatgg tgatcagcca cgctgcgccc cacggccccg aggactcagc | 1200 |
| cccacagttt tctaaactgt accccaatgc ttccaacac ataactccta gttataacta | 1260 |
| tgcaccaa atggataaac actggattat gcagtacaca ggaccaatgc tgcccatcca | 1320 |
| catggaattt acaaacattc tacagcgcaa aaggctccag actttgatgt cagtggatga | 1380 |
| ttctgtggag aggctgtata acatgctcgt ggagacgggg gagctggaga ataactacat | 1440 |
| catttacacc gccgaccatg gttaccatat tgggcagttt ggactggatca aggggaaatc | 1500 |
| catgccatat gactttgata ttcgtgtgcc ttttttatt cgtggtccaa gtgtagaacc | 1560 |
| aggatcaata gtcccacaga tcgttctcaa cattgacttg gccccacga tcctggatat | 1620 |
| tgctgggctc gacacacctc ctgatgtgga cggcaagtct gtcctcaaac ttctggacct | 1680 |
| agaaaagcca ggtaacaggt ttcgaacaaa caagaaggcc aaaatttggc gtgatacatt | 1740 |
| cctagtggaa agaggcaaat ttctacgtaa gaaggaagaa tccagcaaga atatccaaca | 1800 |
| gtcaaatcac ttgcccaaat atgaacgggt caaagaacta tgccagcagg ccaggtacca | 1860 |
| gacagcctgt gaacaaccgg ggcagaagtg gcaatgcatt gaggatacat ctggcaagct | 1920 |
| tcgaattcac aagtgtaaag gaccagtgca cctgctcaca gtccggcaga gcacgcggaa | 1980 |
| cctctacgct cgcggcttcc atgacaaaga caaagagtgc agttgtaggg agtctggta | 2040 |
| ccgtgccagc agaagccaaa gaaagagtca acggcaattc ttgagaaacc aggggactcc | 2100 |
| aaagtacaag ccagatttg tccatactcg gcagacacgt tccttgtccg tcgaatttga | 2160 |
| agggtaaata tatgacataa atctggaaga agaagaagaa ttgcaagtgt tgcaaccaag | 2220 |
| aaacattgct aagcgtcatg atgaaggcca caaggggcca agagatctcc aggcttccag | 2280 |
| tggtggcaac aggggcagga tgctggcaga tagcagcaac gccgtgggcc cacctaccac | 2340 |
| tgtecgagtg acacacaagt gttttattct tccaatgac tctatccatt gtgagagaga | 2400 |

[0069]

| | |
|--|------|
| actgtaccaa tggccagag cgtggaagga ccataaggca tacattgaca aagagattga | 2460 |
| agctctgcaa gataaaatta agaatttaag agaagtgaga ggacatctga agagaaggaa | 2520 |
| gcctgaggaa tntagctgca gtaaacaag ctattacaat aaagagaaag gtgtaaaaaa | 2580 |
| gcaagagaaa ttaaagagcc atcttcaccc attcaaggag gctgctcagg aagtagatag | 2640 |
| caaactgcaa cttttcaagg agaacaaccg taggaggaag aaggagagga aggagaagag | 2700 |
| acggcagagg aagggggaag agtgcagcct gcctggcctc acttgcttca cgcatgacaa | 2760 |
| caaccactgg cagacagccc cgttctgga cctgggatct ttctgtgctt gcacgagttc | 2820 |
| taacaataac acctactggt gtttgcgtac agttaatgag acgcataatt ttcttttctg | 2880 |
| tgagtttgct actggctttt tggagtattt tgatatgaat acagatcctt atcagctcac | 2940 |
| aaatacagtg cacacggtag aacgaggcat tttgaatcag ctacacgtac aactaatgga | 3000 |
| gctcagaagc tgtcaaggat ataagcagtg caaccaaga cctaagaatc ttgatgttg | 3060 |
| aaataaagat ggaggaagct atgacctaca cagaggacag ttatgggatg gatgggaagg | 3120 |
| ttaatcagcc cegtctcact gcagacatca actggcaagg cctagaggag ctacacagtg | 3180 |
| tgaatgaaaa catctatgag tacagacaaa actacagact tagtctggtg gactggacta | 3240 |
| attaactgaa ggatttagat agagtatttg cactgctgaa gagtcactat gagcaaaata | 3300 |
| aaacaaataa gactcaaact gctcaaagtg acgggttctt ggttgtctct gctgagcacg | 3360 |
| ctgtgtcaat ggagatggcc tctgtgact cagatgaaga cccaaggcat aaggttgga | 3420 |
| aaacacctca ttgaccttg ccagctgacc ttcaaaccct gcatttgaac cgaccaacat | 3480 |
| taagtccaga gagtaaactt gaatggaata acgacattcc agaagttaat catttgaatt | 3540 |
| ctgaacactg gagaaaaacc gaaaaatgga cggggcatga agagactaat catctggaaa | 3600 |
| ccgatttcag tggcgalggc atgacagagc tagagctcgg gccagcccc aggctgcagc | 3660 |
| ccattcacag gcacccgaaa gaacttcccc agtatggtgg tcctggaaag gacatTTTTTg | 3720 |

[0070]

aagatcaact atatcttct gtgcattccg atggaatttc agttcatcag atgttcacca 3780
tggccaccgc agaaccaccga agtaattcca gcatagcggg gaagatgttg accaaggtgg 3840
agaagaatca cgaaaaggag aagtcacagc acctagaagg cagcacctcc tcttcactct 3900
cctctgatta gatgaaactg ttaccttacc ctaaacacag tatttctttt taactttttt 3960
atttgtaaac taataaaggt aatcacagcc accaacattc caagctacce tgggtacctt 4020
tgtgcagtag aagctagtga gcatgtgagc aagcgggtg cacaaggaga ctcategta 4080
taatttacta tctgccaaga gtagaaaga aggctgggga tatttgggtt ggcttggttt 4140
tgattttttg cttgtttggt tgtttgtac taaaacagta ttatcttttg aatatcgtag 4200
ggacataagt atatacatgt tatccaatca agatggctag aatggtgcct ttctgagtg 4260
ctaaaacttg acaccctgg taaatctttc aacacacttc cactgcctgc gtaatgaagt 4320
tttgattcat ttttaaccac tggaaatfff caatgccgtc atttcagtt agatgatfff 4380
gcactttgag attaaaatgc catgtctatt tgattagtct tattttttta tttttacagg 4440
cttatcagtc tcaactgttg ctgcatgtg gacaaagtc aataaacccc caaggacgac 4500
acacagtatg gatcacatat tgtttgacat taagcttttg ccagaaaatg ttgcatgtgt 4560
tttacctcga cttgctaaaa tcgattagca gaaaggcatg gctaataatg ttggtggtga 4620
aaataaataa ataagtaaat gaaaaaaaa aaaaaaaaa aaaaaaaaa 4669

<210> 29

<211> 871

<212> PRT

<213> 智人

<400> 29

Met Lys Tyr Ser Cys Cys Ala Leu Val Leu Ala Val Leu Gly Thr Glu
1 5 10 15

[0071]

Leu Leu Gly Ser Leu Cys Ser Thr Val Arg Ser Pro Arg Phe Arg Gly
 20 25 30

Arg Ile Gln Gln Glu Arg Lys Asn Ile Arg Pro Asn Ile Ile Leu Val
 35 40 45

Leu Thr Asp Asp Gln Asp Val Glu Leu Gly Ser Leu Gln Val Met Asn
 50 55 60

Lys Thr Arg Lys Ile Met Glu His Gly Gly Ala Thr Phe Ile Asn Ala
 65 70 75 80

Phe Val Thr Thr Pro Met Cys Cys Pro Ser Arg Ser Ser Met Leu Thr
 85 90 95

Gly Lys Tyr Val His Asn His Asn Val Tyr Thr Asn Asn Glu Asn Cys
 100 105 110

Ser Ser Pro Ser Trp Gln Ala Met His Glu Pro Arg Thr Phe Ala Val
 115 120 125

Tyr Leu Asn Asn Thr Gly Tyr Arg Thr Ala Phe Phe Gly Lys Tyr Leu
 130 135 140

Asn Glu Tyr Asn Gly Ser Tyr Ile Pro Pro Gly Trp Arg Glu Trp Leu
 145 150 155 160

Gly Leu Ile Lys Asn Ser Arg Phe Tyr Asn Tyr Thr Val Cys Arg Asn
 165 170 175

Gly Ile Lys Glu Lys His Gly Phe Asp Tyr Ala Lys Asp Tyr Phe Thr
 180 185 190

[0072]

Asp Leu Ile Thr Asn Glu Ser Ile Asn Tyr Phe Lys Met Ser Lys Arg
 195 200 205

Met Tyr Pro His Arg Pro Val Met Met Val Ile Ser His Ala Ala Pro
 210 215 220

His Gly Pro Glu Asp Ser Ala Pro Gln Phe Ser Lys Leu Tyr Pro Asn
 225 230 235 240

Ala Ser Gln His Ile Thr Pro Ser Tyr Asn Tyr Ala Pro Asn Met Asp
 245 250 255

Lys His Trp Ile Met Gln Tyr Thr Gly Pro Met Leu Pro Ile His Met
 260 265 270

Glu Phe Thr Asn Ile Leu Gln Arg Lys Arg Leu Gln Thr Leu Met Ser
 275 280 285

Val Asp Asp Ser Val Glu Arg Leu Tyr Asn Met Leu Val Glu Thr Gly
 290 295 300

Glu Leu Glu Asn Thr Tyr Ile Ile Tyr Thr Ala Asp His Gly Tyr His
 305 310 315 320

Ile Gly Gln Phe Gly Leu Val Lys Gly Lys Ser Met Pro Tyr Asp Phe
 325 330 335

Asp Ile Arg Val Pro Phe Phe Ile Arg Gly Pro Ser Val Glu Pro Gly
 340 345 350

Ser Ile Val Pro Gln Ile Val Leu Asn Ile Asp Leu Ala Pro Thr Ile
 355 360 365

[0073]

Leu Asp Ile Ala Gly Leu Asp Thr Pro Pro Asp Val Asp Gly Lys Ser
370 375 380

Val Leu Lys Leu Leu Asp Pro Glu Lys Pro Gly Asn Arg Phe Arg Thr
385 390 395 400

Asn Lys Lys Ala Lys Ile Trp Arg Asp Thr Phe Leu Val Glu Arg Gly
405 410 415

Lys Phe Leu Arg Lys Lys Glu Glu Ser Ser Lys Asn Ile Gln Gln Ser
420 425 430

Asn His Leu Pro Lys Tyr Glu Arg Val Lys Glu Leu Cys Gln Gln Ala
435 440 445

Arg Tyr Gln Thr Ala Cys Glu Gln Pro Gly Gln Lys Trp Gln Cys Ile
450 455 460

Glu Asp Thr Ser Gly Lys Leu Arg Ile His Lys Cys Lys Gly Pro Ser
465 470 475 480

Asp Leu Leu Thr Val Arg Gln Ser Thr Arg Asn Leu Tyr Ala Arg Gly
485 490 495

Phe His Asp Lys Asp Lys Glu Cys Ser Cys Arg Glu Ser Gly Tyr Arg
500 505 510

Ala Ser Arg Ser Gln Arg Lys Ser Gln Arg Gln Phe Leu Arg Asn Gln
515 520 525

Gly Thr Pro Lys Tyr Lys Pro Arg Phe Val His Thr Arg Gln Thr Arg
530 535 540

[0074]

Ser Leu Ser Val Glu Phe Glu Gly Glu Ile Tyr Asp Ile Asn Leu Glu
545 550 555 560

Glu Glu Glu Glu Leu Gln Val Leu Gln Pro Arg Asn Ile Ala Lys Arg
565 570 575

His Asp Glu Gly His Lys Gly Pro Arg Asp Leu Gln Ala Ser Ser Gly
580 585 590

Gly Asn Arg Gly Arg Met Leu Ala Asp Ser Ser Asn Ala Val Gly Pro
595 600 605

Pro Thr Thr Val Arg Val Thr His Lys Cys Phe Ile Leu Pro Asn Asp
610 615 620

Ser Ile His Cys Glu Arg Glu Leu Tyr Gln Ser Ala Arg Ala Trp Lys
625 630 635 640

Asp His Lys Ala Tyr Ile Asp Lys Glu Ile Glu Ala Leu Gln Asp Lys
645 650 655

Ile Lys Asn Leu Arg Glu Val Arg Gly His Leu Lys Arg Arg Lys Pro
660 665 670

Glu Glu Cys Ser Cys Ser Lys Gln Ser Tyr Tyr Asn Lys Glu Lys Gly
675 680 685

Val Lys Lys Gln Glu Lys Leu Lys Ser His Leu His Pro Phe Lys Glu
690 695 700

Ala Ala Gln Glu Val Asp Ser Lys Leu Gln Leu Phe Lys Glu Asn Asn
705 710 715 720

[0075]

Arg Arg Arg Lys Lys Glu Arg Lys Glu Lys Arg Arg Gln Arg Lys Gly
725 730 735

Glu Glu Cys Ser Leu Pro Gly Leu Thr Cys Phe Thr His Asp Asn Asn
740 745 750

His Trp Gln Thr Ala Pro Phe Trp Asn Leu Gly Ser Phe Cys Ala Cys
755 760 765

Thr Ser Ser Asn Asn Asn Thr Tyr Trp Cys Leu Arg Thr Val Asn Glu
770 775 780

Thr His Asn Phe Leu Phe Cys Glu Phe Ala Thr Gly Phe Leu Glu Tyr
785 790 795 800

Phe Asp Met Asn Thr Asp Pro Tyr Gln Leu Thr Asn Thr Val His Thr
805 810 815

Val Glu Arg Gly Ile Leu Asn Gln Leu His Val Gln Leu Met Glu Leu
820 825 830

Arg Ser Cys Gln Gly Tyr Lys Gln Cys Asn Pro Arg Pro Lys Asn Leu
835 840 845

Asp Val Gly Asn Lys Asp Gly Gly Ser Tyr Asp Leu His Arg Gly Gln
850 855 860

Leu Trp Asp Gly Trp Glu Gly
865 870

<210> 30
<211> 4279
<212> DNA
<213> 智人

[0076]

<400> 30
 gggccatttc tggacaacag ctgctatfff cacttgagcc caagttaatt tctcggggag 60
 ttctcgggcg cgcacaggca gctcggtttg ccttgcgatt gagctgcggg tcgcgcccg 120
 cgccggcctc tccaatggca aatgtgtgtg gctggaggcg agcgcgaggc ttccggcaaa 180
 ggcagtcgag tgtttgcaga ccggggcgag tcctgtgaaa gcagataaaa gaaaacattt 240
 attaacgtgt cattacgagg ggagcggccg gccggggctg tcgactccc cgcggaacat 300
 ttggctcctc ccagctccta gagaggagaa gaagaaagcg gaaaagaggc agattcacgt 360
 cgtttccagc caagtggacc tgatcgatgg cctcctgaa ttatcacga tatttgattt 420
 attagcgatg cccctgggtt tgtgtgttac gcacacacac gtgcacacaa ggctctggct 480
 cgcttccctc cctcgtttcc agctcctggg cgaatccac atctgtttca actctccgcc 540
 gagggcgagc aggagcgaga gtgtgtcgaa tctgcgagtg aagagggacg agggaaaaga 600
 aacaaagcca cagacgaac ttgagactcc cgcattccaa aagaagcacc agatcagcaa 660
 aaaaagaaga tgggcccccc gagcctcgtg ctgtgcttgc tgcgcgaac tgtgttctcc 720
 ctgctgggtg gaagctggc cttcctgtcg caccaccgcc tgaagggcag gtttcagagg 780
 gaccgcagga acatccgcc caacatcctc ctggtgctga cggacgacca ggatgtggag 840
 ctgggttcca tgcaggtgat gaacaagacc cggcgcatca tggagcaggc cggggcgcac 900
 ttcatcaacg cttcgtgac cacaccatg tgctgccctc cacgctcctc cctcctcacc 960
 ggcaagtacg tccacaacca caacacctac accaacaatg agaactgctc ctcgccctcc 1020
 tggcaggcac agcacgagag ccgcacctt gccgtgtacc tcaatagcac tggctaccgg 1080
 acagctttct tcgggaagta tcttaatgaa tacaacggct cctacgtgcc acccggctgg 1140
 aaggagtggg tcggactcct taaaaactcc cgcctttata actacacgct gtgtcggaac 1200
 ggggtgaaag agaagcacgg ctccgactac tccaaggatt acctcacaga cctcctcacc 1260

[0077]

| | |
|---|------|
| aatgacagcg tgagcttctt ccgcacgtcc aagaagatgt acccgcacag gccagtcctc | 1320 |
| atggtcatca gccatgcagc cccccacggc cctgaggatt cagccccaca atattcacgc | 1380 |
| ctcttcccaa acgcatctca gcacatcacg ccgagctaca actacgcgcc caaccggac | 1440 |
| aaacactgga tcatgcgcta cacggggccc atgaagccca tccacatgga attcaccaac | 1500 |
| atgctccagc ggaagcgctt gcagaccctc atgtcgggtg acgactccat ggagacgatt | 1560 |
| tacaacatgc tggttgagac gggcgagctg gacaacacgt acatcgtata caccgccgac | 1620 |
| cacggttacc acatcggcca gtttggcctg gtgaaaggga aatccatgcc atatgagttt | 1680 |
| gacatcaggg tcccgttcta cgtgaggggc cccaacgtgg aagccggctg tctgaatccc | 1740 |
| cacatcgtcc tcaacattga cctggcccc accatcctgg acattgcagg cctggacata | 1800 |
| cctgcggata tggacgggaa atccatcctc aagctgctgg acacggagcg gccggtgaat | 1860 |
| cggtttcact tgaaaaagaa gatgagggtc tggcgggact ctttcttggg ggagagaggc | 1920 |
| aagctgctac acaagagaga caatgacaag gtggacgccc aggaggagaa ctttctgccc | 1980 |
| aagtaccagc gigtgaagga cctgtgtcag cgtgctgagt accagacggc gtgtgagcag | 2040 |
| ctgggacaga agtggcagtg tgtggaggac gccacgggga agctgaagct gcataagtgc | 2100 |
| aagggcccca tgcggctggg cggcagcaga gcctctcca acctcgtgcc caagtactac | 2160 |
| gggcagggca gcgaggcctg cacctgtgac agcggggact acaagctcag cctggccgga | 2220 |
| cgccgaaaa aactcttcaa gaagaagtac aaggccagct atgtccgcag tcgctccatc | 2280 |
| cgctcagtgg ccatcgaggt ggacggcagg gtgtaccacg taggcctggg tgatgccgcc | 2340 |
| cagccccgaa acctcaccaa gcggcactgg ccaggggccc ctgaggacca agatgacaag | 2400 |
| gatggtgggg acttcagtgg cactggaggc cttcccact actcagccgc caacccatt | 2460 |
| aaagtacac atcggtgcta catcctagag aacgacacag tccagtgtga cctggacctg | 2520 |
| tacaagtccc tgcaggcctg gaaagaccac aagctgcaca tcgaccacga gattgaaacc | 2580 |

[0078]

| | |
|---|------|
| ctgcagaaca aaattaagaa cctgagggaa gtccgaggtc acctgaagaa aaagcggcca | 2640 |
| gaagaatgtg actgtcacia aatcagctac cacaccacgc acaaaggccg cctcaagcac | 2700 |
| agaggctcca gctgcatcc tttcaggaag ggcctgcaag agaaggacia ggtgtggctg | 2760 |
| ttgcgggagc agaagcgcaa gaagaaactc cgcaagctgc tcaagcgctt gcagaacaac | 2820 |
| gacacgtgca gcatgccagg cctcacgtgc ttcaccacgc acaaccagca ctggcagacg | 2880 |
| gcgcttttct ggacactggg gcctttctgt gcctgcacca gcgccaacia taacacgtac | 2940 |
| tggtgcatga ggaccatcaa tgagactcac aatttcctct tctgtgaatt tgcaactggc | 3000 |
| ttcctagagt actttgatct caacacagac ccctaccagc tgatgaatgc agtgaacaca | 3060 |
| ctggacaggg atgtcctcaa ccagctacac gtacagctca tggagctgag gagctgcaag | 3120 |
| ggttacaagc agtgtaacct ccggactcga aacatggacc tgggacttaa agatggagga | 3180 |
| agctatgagc aatacaggca gtttcagcgt cgaaagtggc cagaaatgaa gagaccttct | 3240 |
| tccaaatcac tgggacaact gtgggaaggc tgggaaggtt aagaaciac agaggtggac | 3300 |
| ctccaaaaac atagaggcat cacctgactg cacaggcaat gaaaaacct gtgggtgatt | 3360 |
| tccagcagac ctgtgctatt ggccaggagg cctgagaaaag caagcacgca ctctcagta | 3420 |
| acatgacaga ttctggagga taaccagcag gacagagat aacttcagga agtccatctt | 3480 |
| tgcccctgct ttgcttttg attatacctc accagctgca caaatgcat ttttctgtat | 3540 |
| caaaaagtca cactaaccc tccccagaa gctcaciaag gaaaacggag agagcagcgc | 3600 |
| agagagatct ctttggaat ttctccaag ggcgaaagtc attggaattt taaatcata | 3660 |
| ggggaaaagc agtctgttc taaatcctct tattcttttg gttgtcaca aagaaggac | 3720 |
| taagaagcag gacagaggca acgtggagag gctgaaaaca gtgcagagac gtttgacaat | 3780 |
| gagtcagtag cacaaaagag atgacattta cctagcatat aaacctggt tgcctctgaa | 3840 |
| gaaactgctt tcatgtata tatgtgacta ttacatgta atcaacatgg gaacttttag | 3900 |

[0079]

gggaacctaa taagaaatcc caattttcag gagtgggtgt gtcaataaac gctctgtggc 3960
 cagtgtaaaa gaaaaaaaaa aaaaattgtg gacatttctg ttcctgtcca gataaccattt 4020
 ctctagtagt ttctttgta tgtcccagaa ctgatgtttt ttttttaagg tactgaaaag 4080
 aaatgaagtt gatgtatgtc ccaagttttg atgaaactgt atttgtaaaa aaaattttgt 4140
 agtttaagta ttgtcataca gtgttcaaaa cccagccaa tgaccagcag ttggtatgaa 4200
 gaaccttga cattttgtaa aaggccattt cttggggaaa aaaaaaaaaa aaaaaaaaaa 4260
 aaaaaaaaaa aaaaaaaaaa 4279

<210> 31

<211> 870

<212> PRT

<213> 智人

<400> 31

Met Gly Pro Pro Ser Leu Val Leu Cys Leu Leu Ser Ala Thr Val Phe
 1 5 10 15

Ser Leu Leu Gly Gly Ser Ser Ala Phe Leu Ser His His Arg Leu Lys
 20 25 30

Gly Arg Phe Gln Arg Asp Arg Arg Asn Ile Arg Pro Asn Ile Ile Leu
 35 40 45

Val Leu Thr Asp Asp Gln Asp Val Glu Leu Gly Ser Met Gln Val Met
 50 55 60

Asn Lys Thr Arg Arg Ile Met Glu Gln Gly Gly Ala His Phe Ile Asn
 65 70 75 80

Ala Phe Val Thr Thr Pro Met Cys Cys Pro Ser Arg Ser Ser Ile Leu
 85 90 95

[0080]

Thr Gly Lys Tyr Val His Asn His Asn Thr Tyr Thr Asn Asn Glu Asn
 100 105 110

Cys Ser Ser Pro Ser Trp Gln Ala Gln His Glu Ser Arg Thr Phe Ala
 115 120 125

Val Tyr Leu Asn Ser Thr Gly Tyr Arg Thr Ala Phe Phe Gly Lys Tyr
 130 135 140

Leu Asn Glu Tyr Asn Gly Ser Tyr Val Pro Pro Gly Trp Lys Glu Trp
 145 150 155 160

Val Gly Leu Leu Lys Asn Ser Arg Phe Tyr Asn Tyr Thr Leu Cys Arg
 165 170 175

Asn Gly Val Lys Glu Lys His Gly Ser Asp Tyr Ser Lys Asp Tyr Leu
 180 185 190

Thr Asp Leu Ile Thr Asn Asp Ser Val Ser Phe Phe Arg Thr Ser Lys
 195 200 205

Lys Met Tyr Pro His Arg Pro Val Leu Met Val Ile Ser His Ala Ala
 210 215 220

Pro His Gly Pro Glu Asp Ser Ala Pro Gln Tyr Ser Arg Leu Phe Pro
 225 230 235 240

Asn Ala Ser Gln His Ile Thr Pro Ser Tyr Asn Tyr Ala Pro Asn Pro
 245 250 255

Asp Lys His Trp Ile Met Arg Tyr Thr Gly Pro Met Lys Pro Ile His
 260 265 270

[0081]

Met Glu Phe Thr Asn Met Leu Gln Arg Lys Arg Leu Gln Thr Leu Met
 275 280 285

Ser Val Asp Asp Ser Met Glu Thr Ile Tyr Asn Met Leu Val Glu Thr
 290 295 300

Gly Glu Leu Asp Asn Thr Tyr Ile Val Tyr Thr Ala Asp His Gly Tyr
 305 310 315 320

His Ile Gly Gln Phe Gly Leu Val Lys Gly Lys Ser Met Pro Tyr Glu
 325 330 335

Phe Asp Ile Arg Val Pro Phe Tyr Val Arg Gly Pro Asn Val Glu Ala
 340 345 350

Gly Cys Leu Asn Pro His Ile Val Leu Asn Ile Asp Leu Ala Pro Thr
 355 360 365

Ile Leu Asp Ile Ala Gly Leu Asp Ile Pro Ala Asp Met Asp Gly Lys
 370 375 380

Ser Ile Leu Lys Leu Leu Asp Thr Glu Arg Pro Val Asn Arg Phe His
 385 390 395 400

Leu Lys Lys Lys Met Arg Val Trp Arg Asp Ser Phe Leu Val Glu Arg
 405 410 415

Gly Lys Leu Leu His Lys Arg Asp Asn Asp Lys Val Asp Ala Gln Glu
 420 425 430

Glu Asn Phe Leu Pro Lys Tyr Gln Arg Val Lys Asp Leu Cys Gln Arg
 435 440 445

[0082]

Ala Glu Tyr Gln Thr Ala Cys Glu Gln Leu Gly Gln Lys Trp Gln Cys
450 455 460

Val Glu Asp Ala Thr Gly Lys Leu Lys Leu His Lys Cys Lys Gly Pro
465 470 475 480

Met Arg Leu Gly Gly Ser Arg Ala Leu Ser Asn Leu Val Pro Lys Tyr
485 490 495

Tyr Gly Gln Gly Ser Glu Ala Cys Thr Cys Asp Ser Gly Asp Tyr Lys
500 505 510

Leu Ser Leu Ala Gly Arg Arg Lys Lys Leu Phe Lys Lys Lys Tyr Lys
515 520 525

Ala Ser Tyr Val Arg Ser Arg Ser Ile Arg Ser Val Ala Ile Glu Val
530 535 540

Asp Gly Arg Val Tyr His Val Gly Leu Gly Asp Ala Ala Gln Pro Arg
545 550 555 560

Asn Leu Thr Lys Arg His Trp Pro Gly Ala Pro Glu Asp Gln Asp Asp
565 570 575

Lys Asp Gly Gly Asp Phe Ser Gly Thr Gly Gly Leu Pro Asp Tyr Ser
580 585 590

Ala Ala Asn Pro Ile Lys Val Thr His Arg Cys Tyr Ile Leu Glu Asn
595 600 605

Asp Thr Val Gln Cys Asp Leu Asp Leu Tyr Lys Ser Leu Gln Ala Trp
610 615 620

[0083]

Asp Val Leu Asn Gln Leu His Val Gln Leu Met Glu Leu Arg Ser Cys
 805 810 815

Lys Gly Tyr Lys Gln Cys Asn Pro Arg Thr Arg Asn Met Asp Leu Gly
 820 825 830

Leu Lys Asp Gly Gly Ser Tyr Glu Gln Tyr Arg Gln Phe Gln Arg Arg
 835 840 845

Lys Trp Pro Glu Met Lys Arg Pro Ser Ser Lys Ser Leu Gly Gln Leu
 850 855 860

Trp Glu Gly Trp Glu Gly
 865 870

<210> 32
 <211> 6
 <212> PRT
 <213> 智人

<220>
 <221> 变体
 <222> (1).. (1)
 <223> Leu 或 Val

<220>
 <221> misc_feature
 <222> (1).. (3)
 <223> Xaa 可以是任何天然氨基酸

<220>
 <221> 变体
 <222> (2).. (2)
 <223> Cys 或 Ser

<220>

[0085]

<221> 变体
 <222> (3)..(3)
 <223> 任何氨基酸

<400> 32

Xaa Xaa Xaa Pro Ser Arg
 1 5

<210> 33
 <211> 23
 <212> PRT
 <213> 人工的

<220>
 <223> 来自人芳基硫酸酯酶 A 的序列

<220>
 <221> 肽
 <222> (1)..(23)
 <223> 合成的 FGly 形成底物；来自人芳基硫酸酯酶 A 的一级序列

<400> 33

Met Thr Asp Phe Tyr Val Pro Val Ser Leu Cys Thr Pro Ser Arg Ala
 1 5 10 15

Ala Leu Leu Thr Gly Arg Ser
 20

<210> 34
 <211> 16
 <212> PRT
 <213> 人工的

<220>
 <223> ASA 65-80 肽的变体，其中对于 FGly 形成关键的残基 Cys69、Pro71 和 Arg73 被混
 杂

[0086]

<220>

<221> MISC_FEATURE

<222> (1)..(16)

<223> 混杂寡肽

<400> 34

Pro Val Ser Leu Pro Thr Arg Ser Cys Ala Ala Leu Leu Thr Gly Arg
1 5 10 15

<210> 35

<211> 16

<212> PRT

<213> 人工的

<220>

<223> ASA65-80 肽的变体, 其中 Cys69 被丝氨酸替换

<220>

<221> MISC_FEATURE

<222> (1)..(16)

<223> Ser69 寡肽

<400> 35

Pro Val Ser Leu Ser Thr Pro Ser Arg Ala Ala Leu Leu Thr Gly Arg
1 5 10 15

<210> 36

<211> 19

<212> DNA

<213> 人工的

<220>

<223> 人 FGE 特异性 PCR 引物

<220>

<221> misc_feature

<222> (1)..(19)

<223> 人 FGE 特异性 PCR 引物 1199nc

[0087]

| | |
|--------------------------------|----|
| <400> 36 | |
| ccaatgtagg tcagacacg | 19 |
| <210> 37 | |
| <211> 16 | |
| <212> DNA | |
| <213> 人工的 | |
| <220> | |
| <223> 人 FGE 特异性 PCR 引物 | |
| <220> | |
| <221> misc_feature | |
| <222> (1)..(16) | |
| <223> 人 FGE 特异性正向 PCR 引物 1c | |
| <400> 37 | |
| acatggcccg cgggac | 16 |
| <210> 38 | |
| <211> 19 | |
| <212> DNA | |
| <213> 人工的 | |
| <220> | |
| <223> 人 FGE 特异性 PCR 引物 | |
| <220> | |
| <221> misc_feature | |
| <222> (1)..(19) | |
| <223> 人 FGE 特异性反向 PCR 引物 1182c | |
| <400> 38 | |
| cgactgctcc ttggactgg | 19 |
| <210> 39 | |
| <211> 24 | |
| <212> DNA | |

[0088]

- <213> 人工的
- <220>
- <223> 人 FGE 特异性 PCR 引物
- <220>
- <221> misc_feature
- <222> (1)..(24)
- <223> 含有 EcoRI 的人 5' -FGE 特异性 PCR 引物
- <400> 39
- ggaattcggg acaacatggc tgcg 24
- <210> 40
- <211> 54
- <212> DNA
- <213> 人工的
- <220>
- <223> HA 特异性引物
- <220>
- <221> misc_feature
- <222> (1)..(54)
- <223> HA 特异性引物
- <400> 40
- cccaagctta tgcgtagtca ggcacatcat acggatagtc catggtgggc aggc 54
- <210> 41
- <211> 57
- <212> DNA
- <213> 人工的
- <220>
- <223> c-myc -特异性引物
- <220>
- <221> misc_feature

[0089]

- <222> (1)..(57)
 <223> c-myc -特异性引物
- <400> 41
 cccaagctta caggtcttct tcagaaatca gcttttggtc gtccatggtg ggcaggc 57
- <210> 42
 <211> 54
 <212> DNA
 <213> 人工的
- <220>
 <223> RGS-His6 - 特异性引物
- <220>
 <221> misc_feature
 <222> (1)..(54)
 <223> RGS-His6 - 特异性引物
- <400> 42
 cccaagctta gtgatggtga tggatgatgcg atcctctgtc catggtgggc aggc 54
- <210> 43
 <211> 15
 <212> PRT
 <213> 人工的
- <220>
 <223> 人 FGE 制备物的胰蛋白酶寡肽
- <220>
 <221> MISC_FEATURE
 <222> (1)..(15)
 <223> 人 FGE 制备物的胰蛋白酶寡肽
- <400> 43
 Ser Gln Asn Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg
 1 5 10 15

[0090]

<210> 44
 <211> 19
 <212> PRT
 <213> 人工的

<220>
 <223> 人 FGE 制备物的胰蛋白酶寡肽

<220>
 <221> MISC_FEATURE
 <222> (1)..(19)
 <223> 人 FGE 制备物的胰蛋白酶寡肽

<400> 44

Met Val Pro Ile Pro Ala Gly Val Phe Thr Met Gly Thr Asp Asp Pro
 1 5 10 15

Gln Ile Lys

<210> 45
 <211> 906
 <212> DNA
 <213> 智人

<400> 45
 atggcccggc atgggttacc gctgctgccc ctgctgtcgc tctggtcgg cgcgtggctc 60
 aagctaggaa atggacaggc tactagcatg gtccaactgc aggggtgggag attcctgatg 120
 ggaacaaatt ctccagacag cagagatggt gaagggcctg tgcgggaggc gacagtgaaa 180
 ccctttgcca tcgacatatt tcctgtcacc aacaaagatt tcagggattt tgcagggag 240
 aaaaagtatc ggacagaagc tgagatgttt ggatggagct ttgtctttga ggactttgtc 300
 tctgatgagc tgagaaacia agccaccag ccaatgaagt ctgtactctg gtggcittcca 360
 gtggaaaagg cattttggag gcagcctgca gtcctggct ctggcatccg agagagactg 420

[0091]

gagcaccag tgttacagt gagctggaat gacgcccgtg cctactgtgc ttggcgggga 480
 aaacgactgc ccacggagga agagtgggag ttgcccgcc gaggggctt gaagggtaa 540
 gttlacccat gggggaactg gticcagcca aaccgcacca acctgtggca gggaaagtc 600
 cccaaggag acaaagctga ggatggcttc catggagtct cccagtga tgctttccc 660
 gccagaaca actacgggct ctatgacctc ctggggaac tgtgggagtg gacagcatca 720
 ccgtaccagg ctgctgagca ggacatgcgc gtcctccgg ggcacatctg gatcgacaca 780
 gctgatggct ctgccaatca ccgggcccgg gtcaccacca ggatgggcaa cactccagat 840
 tcagcctcag acaacctgg ttccgctgt gctgcagac caggccggcc gccaggggag 900
 ctgtaa 906

<210> 46
 <211> 301
 <212> PRT
 <213> 智人

<400> 46

Met Ala Arg His Gly Leu Pro Leu Leu Pro Leu Leu Ser Leu Leu Val
 1 5 10 15

Gly Ala Trp Leu Lys Leu Gly Asn Gly Gln Ala Thr Ser Met Val Gln
 20 25 30

Leu Gln Gly Gly Arg Phe Leu Met Gly Thr Asn Ser Pro Asp Ser Arg
 35 40 45

Asp Gly Glu Gly Pro Val Arg Glu Ala Thr Val Lys Pro Phe Ala Ile
 50 55 60

Asp Ile Phe Pro Val Thr Asn Lys Asp Phe Arg Asp Phe Val Arg Glu

[0092]

| | | | |
|---|-----|-----|-----|
| 65 | 70 | 75 | 80 |
| Lys Lys Tyr Arg Thr Glu Ala Glu Met Phe Gly Trp Ser Phe Val Phe | | | |
| | 85 | 90 | 95 |
| Glu Asp Phe Val Ser Asp Glu Leu Arg Asn Lys Ala Thr Gln Pro Met | | | |
| | 100 | 105 | 110 |
| Lys Ser Val Leu Trp Trp Leu Pro Val Glu Lys Ala Phe Trp Arg Gln | | | |
| | 115 | 120 | 125 |
| Pro Ala Gly Pro Gly Ser Gly Ile Arg Glu Arg Leu Glu His Pro Val | | | |
| | 130 | 135 | 140 |
| Leu His Val Ser Trp Asn Asp Ala Arg Ala Tyr Cys Ala Trp Arg Gly | | | |
| 145 | 150 | 155 | 160 |
| Lys Arg Leu Pro Thr Glu Glu Glu Trp Glu Phe Ala Ala Arg Gly Gly | | | |
| | 165 | 170 | 175 |
| Leu Lys Gly Gln Val Tyr Pro Trp Gly Asn Trp Phe Gln Pro Asn Arg | | | |
| | 180 | 185 | 190 |
| Thr Asn Leu Trp Gln Gly Lys Phe Pro Lys Gly Asp Lys Ala Glu Asp | | | |
| | 195 | 200 | 205 |
| Gly Phe His Gly Val Ser Pro Val Asn Ala Phe Pro Ala Gln Asn Asn | | | |
| | 210 | 215 | 220 |
| Tyr Gly Leu Tyr Asp Leu Leu Gly Asn Val Trp Glu Trp Thr Ala Ser | | | |
| 225 | 230 | 235 | 240 |
| Pro Tyr Gln Ala Ala Glu Gln Asp Met Arg Val Leu Arg Gly Ala Ser | | | |

[0093]

| | | | |
|---|-----|-----|-----|
| | 245 | 250 | 255 |
| Trp Ile Asp Thr Ala Asp Gly Ser Ala Asn His Arg Ala Arg Val Thr | | | |
| | 260 | 265 | 270 |
| Thr Arg Met Gly Asn Thr Pro Asp Ser Ala Ser Asp Asn Leu Gly Phe | | | |
| | 275 | 280 | 285 |
| Arg Cys Ala Ala Asp Ala Gly Arg Pro Pro Gly Glu Leu | | | |
| | 290 | 295 | 300 |

<210> 47

<211> 927

<212> DNA

<213> 小家鼠

<400> 47

```

atgcgctctg agttctggtt ccccagcatg ggttccttgc tcctccggt gttgctgctg      60
aggctcctgt cctgccccag gcttcagcta ggacatgccc aggatcctgc catggtgcat      120
ctgccaggtg gccggtttct gatggggaca gacgctccag atggcagaga cggtaaggg      180
cctgccccggg aagtgcagt aaaacccttt gccatcgaca tatttccagt caccaataaa      240
gacttcaggg agttgtcag ggagaagaag taccagactg aagccgagc attcgggtgg      300
agcttcgtct ttgaggattt tgtctccct gagctcagaa agcaagaaa tctgatgccg      360
gctgttcact ggtggcagcc agtgccaaag gcattttgga ggcagcctgc aggtcccggc      420
tctggcatcc gagagaaact ggagcttccc gtggtacacg tgagctggaa cgacgtggt      480
gcttactgcg catggcgggg gagacgcttg cccacagaag aggagtggga gtttgcagcc      540
cgagggggct tgaaggtca ggtttatcca tgggggaacc gttccagcc aaaccgcacc      600
aacttatggc agggaaagtt ccccaaaggt gacaaagctg aagatggitt tcatggactg      660
tcaccagtga acgctttccc cccacagaac aactacggac tgtatgacct catgggcaat      720

```

[0094]

Arg Lys Gln Glu Asn Leu Met Pro Ala Val His Trp Trp Gln Pro Val
 115 120 125

Pro Lys Ala Phe Trp Arg Gln Pro Ala Gly Pro Gly Ser Gly Ile Arg
 130 135 140

Glu Lys Leu Glu Leu Pro Val Val His Val Ser Trp Asn Asp Ala Gly
 145 150 155 160

Ala Tyr Cys Ala Trp Arg Gly Arg Arg Leu Pro Thr Glu Glu Glu Trp
 165 170 175

Glu Phe Ala Ala Arg Gly Gly Leu Lys Gly Gln Val Tyr Pro Trp Gly
 180 185 190

Asn Arg Phe Gln Pro Asn Arg Thr Asn Leu Trp Gln Gly Lys Phe Pro
 195 200 205

Lys Gly Asp Lys Ala Glu Asp Gly Phe His Gly Leu Ser Pro Val Asn
 210 215 220

Ala Phe Pro Pro Gln Asn Asn Tyr Gly Leu Tyr Asp Leu Met Gly Asn
 225 230 235 240

Val Trp Glu Trp Thr Ala Ser Thr Tyr Gln Pro Ala Gly Gln Asp Met
 245 250 255

Arg Val Leu Arg Gly Ala Ser Trp Ile Asp Thr Ala Asp Gly Ser Ala
 260 265 270

Asn His Arg Ala Arg Val Thr Thr Arg Met Gly Asn Thr Pro Asp Ser
 275 280 285

[0096]

Ala Ser Asp Asn Leu Gly Phe Arg Cys Ala Ser Ser Ala Gly Arg Pro
 290 295 300

Lys Glu Asp Leu
 305

<210> 49

<211> 855

<212> DNA

<213> 小家鼠

<400> 49

atggtcccca ttctgctgg agtattcaca atgggcactg atgacctca gatcaggcag 60
 gatggagaag cccctgccag gagagtcact gttgatggct ttacatgga cgcctatgaa 120
 gtcagcaatg cggattttga gaagtttgg aactcgactg gctatttgac agaggctgag 180
 aagtttgag actctttcgt cttgaaggc atgttgagcg agcaagtga aacgcatatc 240
 caccaggcag ttgcagctgc tccatggtgg ttgcctgtca agggagctaa ttggagacac 300
 ccagagggtc cggactccag tattctgcac aggtcaaadc atccggttct ccatgtttcc 360
 tggaacgatg ctgttgccca ctgcacatgg gcgggcaaga ggttcctac tgaggcagag 420
 tgggaataca gctgtagagg aggcctgcag aacaggcttt tcccctgggg caacaaactg 480
 cagcccaaag gacagcatta tgccaacatc tggcagggca agtttctgt gagcaacact 540
 ggcgaggatg gcttccaagg aactgcccc gttgatgcct ttctcccaa tggctatggc 600
 ttatacaaca tagtggggaa tgtgtgggag tggacctcag actggtggac tgttaccat 660
 tctgttgagg aaacgttcaa cccaaagggt cccacttctg ggaaagaccg agtgaagaag 720
 ggtggatcct acatgtgcca taagtcctat tgctataggt accgctgtgc agctcgaagc 780
 cagaacacac cagatagctc tgcattcaac ctgggattcc gatgtgcagc cgaccacctg 840
 cccaccgcag actga 855

[0097]

Cys Arg Gly Gly Leu Gln Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu
145 150 155 160

Gln Pro Lys Gly Gln His Tyr Ala Asn Ile Trp Gln Gly Lys Phe Pro
165 170 175

Val Ser Asn Thr Gly Glu Asp Gly Phe Gln Gly Thr Ala Pro Val Asp
180 185 190

Ala Phe Pro Pro Asn Gly Tyr Gly Leu Tyr Asn Ile Val Gly Asn Val
195 200 205

Trp Glu Trp Thr Ser Asp Trp Trp Thr Val His His Ser Val Glu Glu
210 215 220

Thr Phe Asn Pro Lys Gly Pro Thr Ser Gly Lys Asp Arg Val Lys Lys
225 230 235 240

Gly Gly Ser Tyr Met Cys His Lys Ser Tyr Cys Tyr Arg Tyr Arg Cys
245 250 255

Ala Ala Arg Ser Gln Asn Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly
260 265 270

Phe Arg Cys Ala Ala Asp His Leu Pro Thr Ala Asp
275 280

<210> 51

<211> 1011

<212> DNA

<213> 黄果蝇

<400> 51

atgacaacaa ttatattagt cctctttatt tggatagttt tattcaatga cgtatccagc 60

[0099]

| | |
|--|------|
| gactgtggct gccaaaagct cgaccggaag gccccggata tgccgtccat ttccggacaa | 120 |
| gtgtgccagc aacgagcaca ggggtcacac agccactacc gggattacta tggcgaactg | 180 |
| gagccaaata ttgcggacat gtcactgctt ccgggaggca cggtttacat gggactgac | 240 |
| aaaccgcact ttccggccga ccgcgaggct ccggaacggc agtgaagct gaatgacttc | 300 |
| tacatcgaca agtatgaggt ttccaacgaa gcctttgcga agtttgttct gcacactaac | 360 |
| tacaccacgg aggctgagcg atatggcgac agttttctgt ttaagagcct titgagccca | 420 |
| ttggagcaga agaacctaga ggacttccga gtggcgagcg ctgtctggtg gtacaaagtg | 480 |
| gccggcgtga actggcgaca tccaatggc gtggacagcg atatagacca cttaggccga | 540 |
| cacccgtag tgcacgtatc gtggcgcgac gctgtggagt actgtaagtg ggccggcaag | 600 |
| cggttgcccga gcgaggcgga gtgggaggcg gcttgcaggg gcggcaagga gcgcaaactg | 660 |
| tttcctggg gcaacaagct gatgccaagg aatgaacatt ggctgaacat ctggcaggga | 720 |
| gactttcccg atggcaacct ggctgaagat gggtttgagt acaccagccc cgtggatgcc | 780 |
| ttccgacaga atatttacga cctgcacaac atgggtggca acgtctggga gtggacggca | 840 |
| gatctgtggg acgtaaata cgtagcgat aatccaaatc gggtaagaa gggcggttct | 900 |
| tatctgtgtc acaagtccta ctgtacagg tacaggtgcg cggcacgctc gcagaacaca | 960 |
| gaagacagtt cagccggtaa cctgggtttt cgggtcgcca agaatgcgtg a | 1011 |

<210> 52

<211> 336

<212> PRT

<213> 黄果蝇

<400> 52

Met Thr Thr Ile Ile Leu Val Leu Phe Ile Trp Ile Val Leu Phe Asn
 1 5 10 15

[0100]

Asp Val Ser Ser Asp Cys Gly Cys Gln Lys Leu Asp Arg Lys Ala Pro
 20 25 30

Asp Met Pro Ser Ile Ser Gly Gln Val Cys Gln Gln Arg Ala Gln Gly
 35 40 45

Ala His Ser His Tyr Arg Asp Tyr Tyr Gly Glu Leu Glu Pro Asn Ile
 50 55 60

Ala Asp Met Ser Leu Leu Pro Gly Gly Thr Val Tyr Met Gly Thr Asp
 65 70 75 80

Lys Pro His Phe Pro Ala Asp Arg Glu Ala Pro Glu Arg Gln Val Lys
 85 90 95

Leu Asn Asp Phe Tyr Ile Asp Lys Tyr Glu Val Ser Asn Glu Ala Phe
 100 105 110

Ala Lys Phe Val Leu His Thr Asn Tyr Thr Thr Glu Ala Glu Arg Tyr
 115 120 125

Gly Asp Ser Phe Leu Phe Lys Ser Leu Leu Ser Pro Leu Glu Gln Lys
 130 135 140

Asn Leu Glu Asp Phe Arg Val Ala Ser Ala Val Trp Trp Tyr Lys Val
 145 150 155 160

Ala Gly Val Asn Trp Arg His Pro Asn Gly Val Asp Ser Asp Ile Asp
 165 170 175

His Leu Gly Arg His Pro Val Val His Val Ser Trp Arg Asp Ala Val
 180 185 190

[0101]

Glu Tyr Cys Lys Trp Ala Gly Lys Arg Leu Pro Ser Glu Ala Glu Trp
 195 200 205

Glu Ala Ala Cys Arg Gly Gly Lys Glu Arg Lys Leu Phe Pro Trp Gly
 210 215 220

Asn Lys Leu Met Pro Arg Asn Glu His Trp Leu Asn Ile Trp Gln Gly
 225 230 235 240

Asp Phe Pro Asp Gly Asn Leu Ala Glu Asp Gly Phe Glu Tyr Thr Ser
 245 250 255

Pro Val Asp Ala Phe Arg Gln Asn Ile Tyr Asp Leu His Asn Met Val
 260 265 270

Gly Asn Val Trp Glu Trp Thr Ala Asp Leu Trp Asp Val Asn Asp Val
 275 280 285

Ser Asp Asn Pro Asn Arg Val Lys Lys Gly Gly Ser Tyr Leu Cys His
 290 295 300

Lys Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn Thr
 305 310 315 320

Glu Asp Ser Ser Ala Gly Asn Leu Gly Phe Arg Cys Ala Lys Asn Ala
 325 330 335

<210> 53

<211> 870

<212> DNA

<213> *Anopheles gambiae*

<400> 53

cgggagagct tgctcgatct ggtggaacat tccaagcggc tcgaagacat ggccttacc 60

[0102]

| | |
|--|-----|
| ccaggaggtg aatatgtaat cggcacaat gaacctatct tcgtcaagga tcgcgaatca | 120 |
| ccggccccggc ccgcgacgat ccgcgacttt tacctcgacc agtacgaagt ctccaacgca | 180 |
| cagttcaagg cattcgtcga ccagacgggc tacgtcacgg aggcggaaaa gtttggcgac | 240 |
| agcttcgtct tccagcagct gtcagcgaa ccggtgcgcc agcagtacga agatttccgc | 300 |
| gtggcggcgg cgccttggtg gtacaagga cgtggagcct cctggcagca tccggaaggt | 360 |
| gatgtgtcac gtgatataag cgaccgattg gaccatccgg tggtgacgt gtcttggaac | 420 |
| gatgcggtcg cgtactgcgc ctggaaaggg aagcgcctgc cgacggaagc ggaatgggaa | 480 |
| gcggcctgcc gggcggtcga caagcagaag ctgttcccct gggtaacaa gctgatgcc | 540 |
| aaggagcagc acatgatgaa catatggcag ggcgagttcc cggacagcaa tctgaaggag | 600 |
| gatggctacg agaccacctg cccggtgacg tccttcgcc agaaccctt cgagctgtac | 660 |
| aacatcgttg gcaacgtgtg ggagtggacg gcggatcttt gggacgcgaa ggatgcggcc | 720 |
| atcgagcgca agccgggcag cgatccaccg aatcgggtga aaaagggtgg ctcatacctg | 780 |
| tgtcacgaat cgtactgcta tcgctatcgc tgtcggctc gatcgagaa caccgaggac | 840 |
| agttcggcgg gcaatctggg cttccggtgc | 870 |

<210> 54

<211> 290

<212> PRT

<213> *Anopheles gambiae*

<400> 54

Pro Glu Ser Leu Leu Asp Leu Val Glu His Ser Lys Arg Phe Glu Asp
1 5 10 15

Met Ser Leu Ile Pro Gly Gly Glu Tyr Val Ile Gly Thr Asn Glu Pro
 20 25 30

[0103]

Ile Phe Val Lys Asp Arg Glu Ser Pro Ala Arg Pro Ala Thr Ile Arg
 35 40 45

Asp Phe Tyr Leu Asp Gln Tyr Glu Val Ser Asn Ala Gln Phe Lys Ala
 50 55 60

Phe Val Asp Gln Thr Gly Tyr Val Thr Glu Ala Glu Lys Phe Gly Asp
 65 70 75 80

Ser Phe Val Phe Gln Gln Leu Leu Ser Glu Pro Val Arg Gln Gln Tyr
 85 90 95

Glu Asp Phe Arg Val Ala Ala Ala Pro Trp Trp Tyr Lys Val Arg Gly
 100 105 110

Ala Ser Trp Gln His Pro Glu Gly Asp Val Ser Arg Asp Ile Ser Asp
 115 120 125

Arg Leu Asp His Pro Val Val His Val Ser Trp Asn Asp Ala Val Ala
 130 135 140

Tyr Cys Ala Trp Lys Gly Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu
 145 150 155 160

Ala Ala Cys Arg Gly Gly Arg Lys Gln Lys Leu Phe Pro Trp Gly Asn
 165 170 175

Lys Leu Met Pro Lys Glu Gln His Met Met Asn Ile Trp Gln Gly Glu
 180 185 190

Phe Pro Asp Ser Asn Leu Lys Glu Asp Gly Tyr Glu Thr Thr Cys Pro
 195 200 205

[0104]

Val Thr Ser Phe Arg Gln Asn Pro Phe Glu Leu Tyr Asn Ile Val Gly
 210 215 220

Asn Val Trp Glu Trp Thr Ala Asp Leu Trp Asp Ala Lys Asp Ala Ala
 225 230 235 240

Ile Glu Arg Lys Pro Gly Ser Asp Pro Pro Asn Arg Val Lys Lys Gly
 245 250 255

Gly Ser Tyr Leu Cys His Glu Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala
 260 265 270

Ala Arg Ser Gln Asn Thr Glu Asp Ser Ser Ala Gly Asn Leu Gly Phe
 275 280 285

Arg Cys
 290

<210> 55

<211> 945

<212> DNA

<213> 天蓝色链霉菌

<400> 55

gtggccgtgg ccgccccgtc ccccggggcc gccgcggagc cggggcccgc cgcccgtccg 60

cgctcgaccc gcggacaggt ggcctgccc ggcggtgagt tcgcgatggg ggacgccttc 120

ggggagggat atccggccga cggcgagaca cccgtgcaca cggtgccct gcggcccttc 180

cacatcgacg agaccgccgt caccaacgcc cggttcgccg ccttcgtcaa ggcgaccggc 240

catgtgaccg acgccgaacg cttcggtccc tcggccgtct tccacctggt cgtcgccgcc 300

ccggacgccg acgtcctcgg cagcggccc ggcgccccct ggtggatcaa cgtgcggggc 360

gcccactggc gccgccccga gggcgcccgc tccgacatca ccggccggcc gaaccatccg 420

[0105]

gtcgtccaag tctcctggaa cgatgccacc gcctacgcgc ggtgggccgg caagcgctg 480
cccaccgagg ccgaatggga gtacgccgcc cgcgggggac tggccggccg ccgctacgcc 540
tggggcgacg agctgacccc gggcggccgg tggcgtgca acatctggca gggccgcttc 600
ccgcacgtca acacggccga ggacgggcac ctgagcaccg caccggtcaa gtcctaccgg 660
cccaacggcc acggcctgtg gaacaccgcg ggcaacgtgt gggaatggtg ctccgactgg 720
ttctcgccca cctactacgc cgaatcacc accgtcgacc cgcacggccc cgggaccggg 780
gcggcacggg tgctgcggg cggtcctac ctgtgccacg actcctactg caaccgetac 840
cgggtcgccg cccgtcctc caacacccc gactcctgt cggcaacct cggattccgc 900
tgcgccaacg acgggacct cacgtccgga tcagccgctg agtga 945

<210> 56

<211> 314

<212> PRT

<213> 天蓝色链霉菌

<400> 56

Met Ala Val Ala Ala Pro Ser Pro Ala Ala Ala Ala Glu Pro Gly Pro
1 5 10 15

Ala Ala Arg Pro Arg Ser Thr Arg Gly Gln Val Arg Leu Pro Gly Gly
 20 25 30

Glu Phe Ala Met Gly Asp Ala Phe Gly Glu Gly Tyr Pro Ala Asp Gly
 35 40 45

Glu Thr Pro Val His Thr Val Arg Leu Arg Pro Phe His Ile Asp Glu
 50 55 60

Thr Ala Val Thr Asn Ala Arg Phe Ala Ala Phe Val Lys Ala Thr Gly

[0106]

| 65 | 70 | 75 | 80 |
|-----------------|---------------------|-----------------|-----------------|
| His Val Thr Asp | Ala Glu Arg Phe Gly | Ser Ser Ala Val | Phe His Leu |
| | 85 | 90 | 95 |
| Val Val Ala Ala | Pro Asp Ala Asp | Val Leu Gly Ser | Ala Ala Gly Ala |
| | 100 | 105 | 110 |
| Pro Trp Trp Ile | Asn Val Arg Gly | Ala His Trp Arg | Arg Pro Glu Gly |
| | 115 | 120 | 125 |
| Ala Arg Ser Asp | Ile Thr Gly Arg | Pro Asn His Pro | Val Val His Val |
| | 130 | 135 | 140 |
| Ser Trp Asn Asp | Ala Thr Ala Tyr | Ala Arg Trp Ala | Gly Lys Arg Leu |
| 145 | 150 | 155 | 160 |
| Pro Thr Glu Ala | Glu Trp Glu Tyr | Ala Ala Arg Gly | Gly Leu Ala Gly |
| | 165 | 170 | 175 |
| Arg Arg Tyr Ala | Trp Gly Asp Glu | Leu Thr Pro Gly | Gly Arg Trp Arg |
| | 180 | 185 | 190 |
| Cys Asn Ile Trp | Gln Gly Arg Phe | Pro His Val Asn | Thr Ala Glu Asp |
| | 195 | 200 | 205 |
| Gly His Leu Ser | Thr Ala Pro Val | Lys Ser Tyr Arg | Pro Asn Gly His |
| | 210 | 215 | 220 |
| Gly Leu Trp Asn | Thr Ala Gly Asn | Val Trp Glu Trp | Cys Ser Asp Trp |
| 225 | 230 | 235 | 240 |
| Phe Ser Pro Thr | Tyr Tyr Ala Glu | Ser Pro Thr Val | Asp Pro His Gly |

[0107]

| | | | |
|---|-----|-----|-----|
| | 245 | 250 | 255 |
| Pro Gly Thr Gly Ala Ala Arg Val Leu Arg Gly Gly Ser Tyr Leu Cys | | | |
| | 260 | 265 | 270 |
| His Asp Ser Tyr Cys Asn Arg Tyr Arg Val Ala Ala Arg Ser Ser Asn | | | |
| | 275 | 280 | 285 |
| Thr Pro Asp Ser Ser Ser Gly Asn Leu Gly Phe Arg Cys Ala Asn Asp | | | |
| | 290 | 295 | 300 |
| Ala Asp Leu Thr Ser Gly Ser Ala Ala Glu | | | |
| | 305 | 310 | |

<210> 57

<211> 1005

<212> DNA

<213> *Corynebacterium efficiens*

<400> 57

```

gtggttcgcc atcgactggg ccaccggccc tgcacactga ggattacgtc catgagtaac      60
tgctgctccc cgtaagcgc acaatggcgt accactaccc gggatttata agatcctgtc      120
aatcccacca ctccatgcaa cccggaacaa tcccgcgatg ctgtgacact gccgggtgga      180
gctttccaca tgggcgatca tcacggggag gggatccggg cggacgggga ggggccagta      240
catgaggttc acctegcccc ctteggcatt aatgtcacca cggtcacgaa tgccgagttc      300
ggacgattta ttgaagccac agggatatac acgacagcgg aacgctacgg tgtctcggct      360
gtattctacg cagcgttcca agggcaacgc gctgacattc ttcgccaggt tcccggcgtg      420
ccctgggtgc tggcggtcaa gggtgcaaac tggcagcgtc ccaacggccc cggatccacc      480
ctggacgggc ttgaggacca ccccgctggt cacgtttcct gggatgatgc ogttgcctac      540
tgcacctggg ctggcggteg tctgccacc gaagccgagt gggaatacgc cccccggggt      600

```

[0108]

ggactgcagg ggcacagata tgcctggggg gataacctcg cctagacgg gaggtggaac 660
 tgcaatatct ggcagggggg cttcccatg gagaacaccg ccgcgatgg ttacctcacc 720
 actgcaccgg tgaagaccta cacgcccaat ggatacggtc tgtggcagat ggcagggaat 780
 gtatgggaat ggtgccagga ctggtttgat gcggagtact actcccgtgc ttcctccatc 840
 aaccgcggg gaccggatac cggtgccgc cgggtgatgc gcggaggctc gtatctctgc 900
 catgattcct actgcaacag ataccgggtg gccgccgca attcgaacac cccgattcc 960
 acctcgggga ataccggtt cgggtgcgtt ttcgatagtc cttga 1005

<210> 58

<211> 334

<212> PRT

<213> *Corynebacterium efficiens*

<400> 58

Met Val Arg His Arg Leu Gly His Arg Pro Cys Thr Leu Arg Ile Thr
 1 5 10 15

Ser Met Ser Asn Cys Cys Ser Pro Ser Ser Ala Gln Trp Arg Thr Thr
 20 25 30

Thr Arg Asp Leu Ser Asp Pro Val Asn Pro Thr Thr Pro Cys Asn Pro
 35 40 45

Glu Gln Ser Arg Asp Ala Val Thr Leu Pro Gly Gly Ala Phe His Met
 50 55 60

Gly Asp His His Gly Glu Gly Tyr Pro Ala Asp Gly Glu Gly Pro Val
 65 70 75 80

His Glu Val His Leu Ala Pro Phe Gly Ile Asn Val Thr Thr Val Thr

[0109]

| 85 | 90 | 95 |
|--|-----|---------|
| Asn Ala Glu Phe Gly Arg Phe Ile Glu Ala Thr Gly Tyr Thr Thr Thr 100 | 105 | 110 |
| Ala Glu Arg Tyr Gly Val Ser Ala Val Phe Tyr Ala Ala Phe Gln Gly 115 | 120 | 125 |
| Gln Arg Ala Asp Ile Leu Arg Gln Val Pro Gly Val Pro Trp Trp Leu 130 | 135 | 140 |
| Ala Val Lys Gly Ala Asn Trp Gln Arg Pro Asn Gly Pro Gly Ser Thr 145 | 150 | 155 160 |
| Leu Asp Gly Leu Glu Asp His Pro Val Val His Val Ser Trp Asp Asp 165 | 170 | 175 |
| Ala Val Ala Tyr Cys Thr Trp Ala Gly Gly Arg Leu Pro Thr Glu Ala 180 | 185 | 190 |
| Glu Trp Glu Tyr Ala Ala Arg Gly Gly Leu Gln Gly Ala Arg Tyr Ala 195 | 200 | 205 |
| Trp Gly Asp Asn Leu Ala Leu Asp Gly Arg Trp Asn Cys Asn Ile Trp 210 | 215 | 220 |
| Gln Gly Gly Phe Pro Met Glu Asn Thr Ala Ala Asp Gly Tyr Leu Thr 225 | 230 | 235 240 |
| Thr Ala Pro Val Lys Thr Tyr Thr Pro Asn Gly Tyr Gly Leu Trp Gln 245 | 250 | 255 |
| Met Ala Gly Asn Val Trp Glu Trp Cys Gln Asp Trp Phe Asp Ala Glu | | |

[0110]

| | | | | | |
|---|-----|-----|-----|-----|-----|
| | 260 | | 265 | | 270 |
| Tyr Tyr Ser Arg Ala Ser Ser Ile Asn Pro Arg Gly Pro Asp Thr Gly | | | | | |
| | 275 | | 280 | | 285 |
| Ala Arg Arg Val Met Arg Gly Gly Ser Tyr Leu Cys His Asp Ser Tyr | | | | | |
| | 290 | | 295 | | 300 |
| Cys Asn Arg Tyr Arg Val Ala Ala Arg Asn Ser Asn Thr Pro Asp Ser | | | | | |
| 305 | | 310 | | 315 | 320 |
| Thr Ser Gly Asn Thr Gly Phe Arg Cys Val Phe Asp Ser Pro | | | | | |
| | 325 | | 330 | | |

<210> 59

<211> 1017

<212> DNA

<213> Novosphingobium aromaticivorans

<400> 59

```

atggcgcaac cattccgac gacggcggcc agtcgtacaa gtattgaacg ccatctcgaa    60
cccaattgca ggagcacgtc gcgaatggtc gaacgccccg gcatgcgcct gatcgaagcc    120
ggcactttca ccatgggctc ggaagccttc taccggagg aagcgcctct tcgccgggtg    180
aaggtagaca gcttctggat cgatgaagcg ccggtgacga acgcacagtt cgcgcattc    240
gtggaggcca cgggatacgt cactgtggcc gagatcgagc cggatcccaa ggactacccc    300
ggcatgctcc cgggcatgga ccgcgcggga tcgctgggtg tccagaaaac agcagggccg    360
gtcgacatgg cggatgcgtc caactgggtg cactttacct ttggcgctg ctggaagcat    420
ccacttgac cgggcagttc catcgatggg atcgaggacc atcccgtcgt tcacgtcgcc    480
tatgccgatg ccgaggccta tgccaaatgg gcgggcaagg atctgccgac cgaagccgag    540
ttcgaatatg ctgcgcgagg cgggttggaac ggttccgaat tttcctgggg agacgaactc    600

```

[0111]

gcacctgaag gccggatgat ggccaactac tggcaaggcc tgtttccctt cgccaaccag 660
 tgcctcgatg gctgggaacg gacatcgccc gtccgcaact tcccgcccaa cggctatggt 720
 ctttacgaca tgatcgggaa cacgtgggag tggacctgcg attggtgggc cgacaagccg 780
 ctgactccgc aaaggaaatc ggcatgctgc gcgatcagca atccgcgcgg cggcaagctc 840
 aaggacagct tcgaccgctc gcaacccgca atgcgcatcg gccggaaggt cataaagggc 900
 ggttcgcacc tgtgtgcggc caattactgc cagcgtatc gccccgcagc acgcatcct 960
 gaaatggttg ataccgcgac gacgcacatc ggcttcaggt gtgtggtgcg gccctga 1017

<210> 60

<211> 338

<212> PRT

<213> *Novosphingobium aromaticivorans*

<400> 60

Met Ala Gln Pro Phe Arg Ser Thr Ala Ala Ser Arg Thr Ser Ile Glu
 1 5 10 15

Arg His Leu Glu Pro Asn Cys Arg Ser Thr Ser Arg Met Val Glu Arg
 20 25 30

Pro Gly Met Arg Leu Ile Glu Gly Gly Thr Phe Thr Met Gly Ser Glu
 35 40 45

Ala Phe Tyr Pro Glu Glu Ala Pro Leu Arg Arg Val Lys Val Asp Ser
 50 55 60

Phe Trp Ile Asp Glu Ala Pro Val Thr Asn Ala Gln Phe Ala Ala Phe
 65 70 75 80

Val Glu Ala Thr Gly Tyr Val Thr Val Ala Glu Ile Glu Pro Asp Pro

[0112]

| | 85 | 90 | 95 |
|---|-----|-----|-----|
| Lys Asp Tyr Pro Gly Met Leu Pro Gly Met Asp Arg Ala Gly Ser Leu | 100 | 105 | 110 |
| Val Phe Gln Lys Thr Ala Gly Pro Val Asp Met Ala Asp Ala Ser Asn | 115 | 120 | 125 |
| Trp Trp His Phe Thr Phe Gly Ala Cys Trp Lys His Pro Leu Gly Pro | 130 | 135 | 140 |
| Gly Ser Ser Ile Asp Gly Ile Glu Asp His Pro Val Val His Val Ala | 145 | 150 | 155 |
| Tyr Ala Asp Ala Glu Ala Tyr Ala Lys Trp Ala Gly Lys Asp Leu Pro | 165 | 170 | 175 |
| Thr Glu Ala Glu Phe Glu Tyr Ala Ala Arg Gly Gly Leu Asp Gly Ser | 180 | 185 | 190 |
| Glu Phe Ser Trp Gly Asp Glu Leu Ala Pro Glu Gly Arg Met Met Ala | 195 | 200 | 205 |
| Asn Tyr Trp Gln Gly Leu Phe Pro Phe Ala Asn Gln Cys Leu Asp Gly | 210 | 215 | 220 |
| Trp Glu Arg Thr Ser Pro Val Arg Asn Phe Pro Pro Asn Gly Tyr Gly | 225 | 230 | 235 |
| Leu Tyr Asp Met Ile Gly Asn Thr Trp Glu Trp Thr Cys Asp Trp Trp | 245 | 250 | 255 |
| Ala Asp Lys Pro Leu Thr Pro Gln Arg Lys Ser Ala Cys Cys Ala Ile | | | |

[0113]

| | | | |
|---|-----|-----|-----|
| | 260 | 265 | 270 |
| Ser Asn Pro Arg Gly Gly Lys Leu Lys Asp Ser Phe Asp Pro Ser Gln | | | |
| | 275 | 280 | 285 |
| Pro Ala Met Arg Ile Gly Arg Lys Val Ile Lys Gly Gly Ser His Leu | | | |
| | 290 | 295 | 300 |
| Cys Ala Ala Asn Tyr Cys Gln Arg Tyr Arg Pro Ala Ala Arg His Pro | | | |
| 305 | 310 | 315 | 320 |
| Glu Met Val Asp Thr Ala Thr Thr His Ile Gly Phe Arg Cys Val Val | | | |
| | 325 | 330 | 335 |

Arg Pro

<210> 61

<211> 1119

<212> DNA

<213> Mesorhizobium loti

<400> 61

```

atgggccccac gagtgcgagg tcaaaaaccg catgaaaggc gacgcggcca tgttcgacat    60
tgccgggaag ttctagccga tagcgggtgg gcggtgatg gagatgagca cgccgtgtca    120
tttcgggatc ttctgatgaa cgcccctgcc gaagtcttcg agcgcgctgc agccgaacgg    180
tcgtaccccg gaatggtctg gatccccggc ggtaccttcc tgatgggctc agacaaccac    240
tatccggagg aggcaccggc ccaccgggtc agggctgacg gcttctggat ggacaaattc    300
accgtctcca accgcgactt cgaacgcttc gttgcggcga caggacatgt cactcttgcc    360
gagaaaccgg ccaatcccga cgactatccc ggtgccttac ccgatctgct ggctccgtcc    420
tcgatgatgt tcaggaagcc ggccggccct gtcgacctg gcaatcacta caattggtgg    480

```

[0114]

gtctatgtcc gggcgccaa ctggcgccat ccacgcgggc cggcaagtac aatcaagaag 540
 gttgcagatc atccggtcgt gcatgtggcc tacgaggatg tcgtggccta tgccaactgg 600
 gcaggcaagg aacttcccac cgaggccgag tgggaattcg cggcgcgagg cggcctcgat 660
 gccgccgaat acgtctgggg caacgagctt acgccggccg ggaagcacat ggccaacatc 720
 tggaaggag actttcccta ccggaatact gtcgacgacg gttacgaata tacggcccca 780
 gtaggctcgt tcccgccaa cgactacggt ctctacgaca tggccggcaa tgtctggcaa 840
 tggacgaccg actggtacca ggaccacaag gcgatcgaca gcccggtgctg caccgctgtc 900
 aatccgcgtg gggccatcg cgaagcgagc tatgacaccc ggctacctga cgtaagatc 960
 cctcgcaagg tcaccaaggg tggtcccat ctgtgcgcgc cgaactactg tcggcgetac 1020
 cggcccgcgg cggaatggc gcaaccgctc gacactgcaa tctcccatct cggctttcgc 1080
 tgcacgtgc gaaggaaaat ggaattgaac ggcagtaa 1119

<210> 62

<211> 372

<212> PRT

<213> Mesorhizobium loti

<400> 62

Met Gly Pro Arg Gly Arg Gly Gln Lys Pro His Glu Arg Arg Arg Gly
 1 5 10 15

His Val Arg His Cys Arg Glu Val Leu Ala Asp Ser Gly Trp Ala Ala
 20 25 30

Asp Gly Asp Glu His Ala Val Ser Phe Arg Asp Leu Ser Met Asn Ala
 35 40 45

Pro Ala Glu Val Phe Glu Arg Ala Ala Ala Glu Arg Ser Tyr Pro Gly

[0115]

| 50 | 55 | 60 |
|--|-----|---------|
| Met Val Trp Ile Pro Gly Gly Thr Phe Leu Met Gly Ser Asp Asn His 65 | 70 | 75 80 |
| Tyr Pro Glu Glu Ala Pro Ala His Arg Val Arg Val Asp Gly Phe Trp 85 | 90 | 95 |
| Met Asp Lys Phe Thr Val Ser Asn Arg Asp Phe Glu Arg Phe Val Ala 100 | 105 | 110 |
| Ala Thr Gly His Val Thr Leu Ala Glu Lys Pro Ala Asn Pro Asp Asp 115 | 120 | 125 |
| Tyr Pro Gly Ala Leu Pro Asp Leu Leu Ala Pro Ser Ser Met Met Phe 130 | 135 | 140 |
| Arg Lys Pro Ala Gly Pro Val Asp Leu Gly Asn His Tyr Asn Trp Trp 145 | 150 | 155 160 |
| Val Tyr Val Arg Gly Ala Asn Trp Arg His Pro Arg Gly Pro Ala Ser 165 | 170 | 175 |
| Thr Ile Lys Lys Val Ala Asp His Pro Val Val His Val Ala Tyr Glu 180 | 185 | 190 |
| Asp Val Val Ala Tyr Ala Asn Trp Ala Gly Lys Glu Leu Pro Thr Glu 195 | 200 | 205 |
| Ala Glu Trp Glu Phe Ala Ala Arg Gly Gly Leu Asp Ala Ala Glu Tyr 210 | 215 | 220 |
| Val Trp Gly Asn Glu Leu Thr Pro Ala Gly Lys His Met Ala Asn Ile | | |

[0116]

| | | | |
|---|-----------------------|-----|-----|
| 225 | 230 | 235 | 240 |
| Trp Gln Gly Asp Phe Pro Tyr Arg Asn Thr Val Asp Asp Gly Tyr Glu | | | |
| | 245 | 250 | 255 |
| Tyr Thr Ala Pro Val Gly Ser Phe Pro Ala Asn Asp Tyr Gly Leu Tyr | | | |
| | 260 | 265 | 270 |
| Asp Met Ala Gly Asn Val Trp Gln Trp Thr Thr Asp Trp Tyr Gln Asp | | | |
| | 275 | 280 | 285 |
| His Lys Ala Ile Asp Ser Pro Cys Cys Thr Ala Val Asn Pro Arg Gly | | | |
| | 290 | 295 | 300 |
| Gly His Arg Glu Ala Ser Tyr Asp Thr Arg Leu Pro Asp Val Lys Ile | | | |
| 305 | 310 | 315 | 320 |
| Pro Arg Lys Val Thr Lys Gly Gly Ser His Leu Cys Ala Pro Asn Tyr | | | |
| | 325 | 330 | 335 |
| Cys Arg Arg Tyr Arg Pro Ala Ala Arg Met Ala Gln Pro Val Asp Thr | | | |
| | 340 | 345 | 350 |
| Ala Ile Ser His Leu Gly Phe Arg Cys Ile Val Arg Arg Lys Met Glu | | | |
| | 355 | 360 | 365 |
| Leu Asn Ala Gln | | | |
| | 370 | | |
| <210> | 63 | | |
| <211> | 1251 | | |
| <212> | DNA | | |
| <213> | Burkholderia fungorum | | |

[0117]

<400> 63
 atgaagagtg aaagagatcg agagcccgca aagtcgtccc gctcgaacgg gtcggtcgca 60
 gcaacccaaa cgcgcgccgg tcgcgtgcgc aaactaatgt tgtggggcgc cctgctcgtc 120
 atactgcccc cctgtgtcgg cgcgcggtc agttgggcct tcacgccgca cgcacccgct 180
 caccgcgaaa tcgttttcgg cgacggcacg catggctcgc tcggcatggc gtgggtgcc 240
 ggcgccagcagtc tcctcatggg cagcgacgcc aaacaggcgc aaccgaacga acgccccgcg 300
 cacaaggtea aggtgcacgg cttctggatg gaccgccatc acgtgaccaa cgccgaattc 360
 cgccgcttcg tcgaagcgac cggctacgtc accacggccg agaagaaacc cgactgggag 420
 accctgaaag tccagttgcc gcccggcacg ccgcgccgc ccgagagcgc gatggtggcg 480
 ggtgcaatgg tgttcgtcgg caccagccgt cccgtgcgc tagacgacta ttcgcagtgg 540
 tggcgctatg tgcctggcgc taactggcgt catccagccg gcctgagag caacatcacc 600
 ggtaaagatg atcaccccggt ggttcaagtg tcctacgaag atgcgcaggc ttatgcgaaa 660
 tgggccggca agcgtctgcc gaccgaagcc gaatgggaat tcgccgcgcg cggcggcctc 720
 gaacaggcca cgtatgcgtg gggcgatcag ttctctccca acggcaaaca gatggccaac 780
 gtctggcagg gccagcagcc gcagtctttc cccgttgta acccgaaagc ggggtggcgcg 840
 ctcggtacaa gtccggtggg tactttccc gccaacggct acggcctttc cgacatgacc 900
 ggcaacgcct ggcatgggt tgccgactgg tatcgcgcgg atcagttcag gcgtgaggcg 960
 gtaagcacca gcgcgatcga caatccggtg ggcccagcgc agtcgtggga ccccgcagac 1020
 cagggcgtgc ccgtcaacgc gcccaagcgt gtcacacgcg gcggttcggt cctctgcaac 1080
 gaaatctatt gcctgagcta ccggcccagc gcgagacgcg gcaccgatcc ctacaacagc 1140
 atgtcgcacc tgggcttccg gctggtgatg gacgaagaca cctggaaaga agccggtgct 1200
 gccagcctt cggcgaagc tgccggcgcg cctggaacct ctggcggcta g 1251

[0118]

<210> 64
 <211> 416
 <212> PRT
 <213> Burkholderia fungorum

<400> 64

Met Lys Ser Glu Arg Asp Arg Glu Pro Ala Lys Ser Ser Arg Ser Asn
 1 5 10 15

Gly Ser Val Ala Ala Thr Gln Thr Arg Ala Gly Arg Val Arg Lys Leu
 20 25 30

Met Leu Trp Gly Ala Leu Leu Val Ile Leu Pro Ala Cys Val Gly Ala
 35 40 45

Ala Val Ser Trp Ala Phe Thr Pro His Ala Pro Ala His Pro Gln Ile
 50 55 60

Val Phe Gly Asp Gly Thr His Gly Pro Leu Gly Met Ala Trp Val Pro
 65 70 75 80

Gly Gly Gln Phe Leu Met Gly Ser Asp Ala Lys Gln Ala Gln Pro Asn
 85 90 95

Glu Arg Pro Ala His Lys Val Lys Val His Gly Phe Trp Met Asp Arg
 100 105 110

His His Val Thr Asn Ala Glu Phe Arg Arg Phe Val Glu Ala Thr Gly
 115 120 125

Tyr Val Thr Thr Ala Glu Lys Lys Pro Asp Trp Glu Thr Leu Lys Val
 130 135 140

Gln Leu Pro Pro Gly Thr Pro Arg Pro Pro Glu Ser Ala Met Val Ala

[0119]

| | | | | | | |
|---|-----|-----|-----|-----|-----|-----|
| 145 | | 150 | | 155 | | 160 |
| Gly Ala Met Val Phe Val Gly Thr Ser Arg Pro Val Pro Leu Asp Asp | | | | | | |
| | 165 | | 170 | | 175 | |
| Tyr Ser Gln Trp Trp Arg Tyr Val Pro Gly Ala Asn Trp Arg His Pro | | | | | | |
| | 180 | | 185 | | 190 | |
| Ala Gly Pro Glu Ser Asn Ile Ile Gly Lys Asp Asp His Pro Val Val | | | | | | |
| | 195 | | 200 | | 205 | |
| Gln Val Ser Tyr Glu Asp Ala Gln Ala Tyr Ala Lys Trp Ala Gly Lys | | | | | | |
| | 210 | | 215 | | 220 | |
| Arg Leu Pro Thr Glu Ala Glu Trp Glu Phe Ala Ala Arg Gly Gly Leu | | | | | | |
| 225 | | 230 | | 235 | | 240 |
| Glu Gln Ala Thr Tyr Ala Trp Gly Asp Gln Phe Ser Pro Asn Gly Lys | | | | | | |
| | 245 | | 250 | | 255 | |
| Gln Met Ala Asn Val Trp Gln Gly Gln Gln Pro Gln Ser Phe Pro Val | | | | | | |
| | 260 | | 265 | | 270 | |
| Val Asn Pro Lys Ala Gly Gly Ala Leu Gly Thr Ser Pro Val Gly Thr | | | | | | |
| | 275 | | 280 | | 285 | |
| Phe Pro Ala Asn Gly Tyr Gly Leu Ser Asp Met Thr Gly Asn Ala Trp | | | | | | |
| | 290 | | 295 | | 300 | |
| Gln Trp Val Ala Asp Trp Tyr Arg Ala Asp Gln Phe Arg Arg Glu Ala | | | | | | |
| 305 | | 310 | | 315 | | 320 |
| Val Ser Thr Ser Ala Ile Asp Asn Pro Val Gly Pro Ser Glu Ser Trp | | | | | | |

[0120]

| | | | |
|---|-----|-----|-----|
| | 325 | 330 | 335 |
| Asp Pro Ala Asp Gln Gly Val Pro Val Asn Ala Pro Lys Arg Val Thr | 340 | 345 | 350 |
| Arg Gly Gly Ser Phe Leu Cys Asn Glu Ile Tyr Cys Leu Ser Tyr Arg | 355 | 360 | 365 |
| Pro Ser Ala Arg Arg Gly Thr Asp Pro Tyr Asn Ser Met Ser His Leu | 370 | 375 | 380 |
| Gly Phe Arg Leu Val Met Asp Glu Asp Thr Trp Lys Glu Ala Gly Ala | 385 | 390 | 395 |
| | | | 400 |
| Arg Gln Ala Ser Ala Lys Ala Ala Gly Ala Pro Gly Thr Pro Gly Gly | 405 | 410 | 415 |

<210> 65

<211> 912

<212> DNA

<213> Sinorhizobium meliloti

<400> 65

```

atggtctgga ttcccggagc gacctcatg atggggtcga acgaccatta cccggaggaa      60
gcgccgtgc atccggtaac cgtcgacgga ttctggatcg atgtgacacc ggtaacgaac      120
cgccagtttc tcgaattcgt aatgcgacg gggcatgtga ccttcgcgga aagaaagccg      180
cgcgccgaag actatccggg cgctccgcca tccaatctaa gggccggttc gctcgtcttc      240
acacccccga agcgaccgct gcaggaacg gatatatcgc agtgggtgat attcacgctg      300
ggtgccaaact ggcggcaccc gctcgggccc aagagcagca tcggagcgat tctggatcat      360
cgggtcgtcc atgtcgctta cagcgacgca aaggcctatg ccgaatgggc cggcaaggac      420
ctcccgaccg agaccgagtg ggagctggcg gcccgggcgc gcctcgatgg ggctgaattt      480

```

[0121]

tcctggggcg gcgagcttgc gccgggcgga aatcacatgg ccaatacttg gcaggggaagt 540
 ttcccggtcg agaattctat ggacgatggt ttcgcgcgaa catcgccggt cagattttac 600
 ccgccgaacg gctacggcct ctacgacatg atcggcaatg tgtgggagtg gaccacggat 660
 tactggtccg tgcgccaccc ggaagcggcc gccaaagcctt gctgcattcc gagcaatccc 720
 cgcaatgccg atgccgatgc gaggatcgc cgggcggcga gcgtgaaagt tccgcgccgg 780
 gtgctcaagg gtggatcgca tctctgcgcg ccgaactact gccggcggta ccgccttgcg 840
 gcgaggcacg cccaggaaat cgacacgacg accagccatg tcggtttccg atgtgtcagg 900
 cgcgttcgat aa 912

<210> 66
 <211> 303
 <212> PRT
 <213> *Sinorhizobium meliloti*

<400> 66

Met Val Trp Val Pro Gly Ala Thr Phe Met Met Gly Ser Asn Asp His
 1 5 10 15

Tyr Pro Glu Glu Ala Pro Val His Pro Val Thr Val Asp Gly Phe Trp
 20 25 30

Ile Asp Val Thr Pro Val Thr Asn Arg Gln Phe Leu Glu Phe Val Asn
 35 40 45

Ala Thr Gly His Val Thr Phe Ala Glu Arg Lys Pro Arg Ala Glu Asp
 50 55 60

Tyr Pro Gly Ala Pro Pro Ser Asn Leu Arg Ala Gly Ser Leu Val Phe
 65 70 75 80

[0122]

Thr Pro Pro Lys Arg Pro Leu Gln Gly Thr Asp Ile Ser Gln Trp Trp
85 90 95

Ile Phe Thr Leu Gly Ala Asn Trp Arg His Pro Leu Gly Arg Lys Ser
100 105 110

Ser Ile Gly Ala Ile Leu Asp His Pro Val Val His Val Ala Tyr Ser
115 120 125

Asp Ala Lys Ala Tyr Ala Glu Trp Ala Gly Lys Asp Leu Pro Thr Glu
130 135 140

Thr Glu Trp Glu Leu Ala Ala Arg Gly Gly Leu Asp Gly Ala Glu Phe
145 150 155 160

Ser Trp Gly Gly Glu Leu Ala Pro Gly Gly Asn His Met Ala Asn Thr
165 170 175

Trp Gln Gly Ser Phe Pro Val Glu Asn Ser Met Asp Asp Gly Phe Ala
180 185 190

Arg Thr Ser Pro Val Arg Phe Tyr Pro Pro Asn Gly Tyr Gly Leu Tyr
195 200 205

Asp Met Ile Gly Asn Val Trp Glu Trp Thr Thr Asp Tyr Trp Ser Val
210 215 220

Arg His Pro Glu Ala Ala Ala Lys Pro Cys Cys Ile Pro Ser Asn Pro
225 230 235 240

Arg Asn Ala Asp Ala Asp Ala Ser Ile Asp Pro Ala Ala Ser Val Lys
245 250 255

[0123]

Val Pro Arg Arg Val Leu Lys Gly Gly Ser His Leu Cys Ala Pro Asn
 260 265 270

Tyr Cys Arg Arg Tyr Arg Pro Ala Ala Arg His Ala Gln Glu Ile Asp
 275 280 285

Thr Thr Thr Ser His Val Gly Phe Arg Cys Val Arg Arg Val Arg
 290 295 300

<210> 67

<211> 1065

<212> DNA

<213> 微颤蓝细菌属物种

<400> 67

atgaaataca ttttttagt tcttttctta tgggccttga cccgatgtac cggaaagtat 60
 gaggacaaga gagtggaaac tgatacttcc agaccaaaag ccgaagcgtc agatataaaa 120
 gttcccgaag gaatggctta tattcccgcg ggccagtaca tgatgggagg taaatcagac 180
 caggcttata aggatgaata tccccgcat aacgtgaagg tttcggcttt ttatatggac 240
 cttacagaag tgaccaatgc ggagtttaag cggttttag acgaaacggg ctacgtgacc 300
 attgctgaga aagatattga ctgggaagag ttaaagtctc aggtgccaca gggtaacccg 360
 aagcctcctg attctgtgct tcaggcaggt tcaactggtt tcaagcagac agatgaaccc 420
 gtttctctcc aggattatc acagtggtag gaatggacta tcggagccaa ctggcgaat 480
 ccggagggtc caggtagtag gattgaggat cgtatggatc atccggtggt acacgtttcc 540
 tttgaagatg tccaagcgtg tgcggattgg gccggtlaagc gctgcctac tgaggcagaa 600
 tggaatggg ccgcatggg aggccaaaat gacgtgaaat atccatgggg aatgaatcg 660
 gtcgaacaag catccgataa agcaaaacttt tggcagggga atttccaca tcaaaactat 720
 gccctgatg gattcgaacg caccgccct gtacgtcct tcccagcga tgggtacggc 780

[0124]

ctatatgata tggctggcaa tgtgtgggaa tggtgccagg ataagtatga tgtcaatgct 840
 tatgaaagct ataagcaaaa aggactgaca gaagacccca cgggttctga gcactacaac 900
 gaccctaggg aaccgtatac tcctaagcat gtgatcagag ggggttcttt cctatgcaat 960
 gacagctact gtagtgggta tcgtgtttca cgtcgatga gttccagtag agattcaggt 1020
 tttaatcata cgggattcag gtgtgtgaaa gatgtaaag gatag 1065

<210> 68

<211> 354

<212> PRT

<213> 微颤蓝细菌属物种

<400> 68

Met Lys Tyr Ile Phe Leu Val Leu Phe Leu Trp Ala Leu Thr Arg Cys
 1 5 10 15

Thr Gly Lys Tyr Glu Asp Lys Arg Val Glu Thr Asp Thr Ser Arg Pro
 20 25 30

Lys Ala Glu Ala Ser Asp Ile Lys Val Pro Glu Gly Met Ala Tyr Ile
 35 40 45

Pro Ala Gly Gln Tyr Met Met Gly Gly Lys Ser Asp Gln Ala Tyr Lys
 50 55 60

Asp Glu Tyr Pro Arg His Asn Val Lys Val Ser Ala Phe Tyr Met Asp
 65 70 75 80

Leu Thr Glu Val Thr Asn Ala Glu Phe Lys Arg Phe Val Asp Glu Thr
 85 90 95

Gly Tyr Val Thr Ile Ala Glu Lys Asp Ile Asp Trp Glu Glu Leu Lys

[0125]

| | 100 | 105 | 110 |
|---|-----|-----|-----|
| Ser Gln Val Pro Gln Gly Thr Pro Lys Pro Pro Asp Ser Val Leu Gln | 115 | 120 | 125 |
| Ala Gly Ser Leu Val Phe Lys Gln Thr Asp Glu Pro Val Ser Leu Gln | 130 | 135 | 140 |
| Asp Tyr Ser Gln Trp Trp Glu Trp Thr Ile Gly Ala Asn Trp Arg Asn | 145 | 150 | 160 |
| Pro Glu Gly Pro Gly Ser Thr Ile Glu Asp Arg Met Asp His Pro Val | 165 | 170 | 175 |
| Val His Val Ser Phe Glu Asp Val Gln Ala Tyr Ala Asp Trp Ala Gly | 180 | 185 | 190 |
| Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Trp Ala Ala Met Gly Gly | 195 | 200 | 205 |
| Gln Asn Asp Val Lys Tyr Pro Trp Gly Asn Glu Ser Val Glu Gln Ala | 210 | 215 | 220 |
| Ser Asp Lys Ala Asn Phe Trp Gln Gly Asn Phe Pro His Gln Asn Tyr | 225 | 230 | 240 |
| Ala Leu Asp Gly Phe Glu Arg Thr Ala Pro Val Arg Ser Phe Pro Ala | 245 | 250 | 255 |
| Asn Gly Tyr Gly Leu Tyr Asp Met Ala Gly Asn Val Trp Glu Trp Cys | 260 | 265 | 270 |
| Gln Asp Lys Tyr Asp Val Asn Ala Tyr Glu Ser Tyr Lys Gln Lys Gly | | | |

[0126]

gagcagccca agggcaagct catggccaat acctggcagg gtcagttccc ttatcgcaat 540
 gcggcgaagg atggttttac cggtacatcg cccgtgggtt gcttcccggc caacggcttt 600
 ggctgtttcg atgccggcgg caatgtctgg gagctgactc gcacgggcta tcggccaggc 660
 catgacgcac agcgcgacgc caagctcgac ccctcaggcc cggccctgag tgacagcttc 720
 gaccggcag accccggcgt gccggtggcg gtaatcaaag gcggctcgca cctgtgttcg 780
 gcgaccgct gcatgcgcta ccgcccctcg gcacgccagc cgcagccggt gttcatgacg 840
 acctcgacg tgggtttcag aacgattcgg caatga 876

<210> 70
 <211> 291
 <212> PRT
 <213> 恶臭假单胞菌 KT2440

<400> 70

Met Val His Val Pro Gly Gly Glu Phe Ser Phe Gly Ser Ser Arg Phe
 1 5 10 15

Tyr Asp Glu Glu Gly Pro Pro His Pro Ala Lys Val Ser Gly Phe Trp
 20 25 30

Ile Asp Val His Pro Val Thr Asn Ala Gln Phe Ala Arg Phe Val Lys
 35 40 45

Ala Thr Gly Tyr Val Thr His Ala Glu Arg Gly Thr Arg Val Glu Asp
 50 55 60

Asp Pro Ala Leu Pro Asp Ala Leu Arg Ile Pro Gly Ala Met Val Phe
 65 70 75 80

His Gln Gly Ala Asp Val Leu Gly Pro Gly Trp Gln Phe Val Pro Gly

[0128]

| | 85 | 90 | 95 |
|---|-----|-----|-----|
| Ala Asn Trp Arg His Pro Gln Gly Pro Gly Ser Ser Leu Ala Gly Leu | 100 | 105 | 110 |
| Asp Asn His Pro Val Val Gln Ile Ala Leu Glu Asp Ala Gln Ala Tyr | 115 | 120 | 125 |
| Ala Arg Trp Ala Gly Arg Glu Leu Pro Ser Glu Ala Gln Leu Glu Tyr | 130 | 135 | 140 |
| Ala Met Arg Gly Gly Leu Thr Asp Ala Asp Phe Ser Trp Gly Thr Thr | 145 | 150 | 155 |
| Glu Gln Pro Lys Gly Lys Leu Met Ala Asn Thr Trp Gln Gly Gln Phe | 165 | 170 | 175 |
| Pro Tyr Arg Asn Ala Ala Lys Asp Gly Phe Thr Gly Thr Ser Pro Val | 180 | 185 | 190 |
| Gly Cys Phe Pro Ala Asn Gly Phe Gly Leu Phe Asp Ala Gly Gly Asn | 195 | 200 | 205 |
| Val Trp Glu Leu Thr Arg Thr Gly Tyr Arg Pro Gly His Asp Ala Gln | 210 | 215 | 220 |
| Arg Asp Ala Lys Leu Asp Pro Ser Gly Pro Ala Leu Ser Asp Ser Phe | 225 | 230 | 235 |
| Asp Pro Ala Asp Pro Gly Val Pro Val Ala Val Ile Lys Gly Gly Ser | 245 | 250 | 255 |
| His Leu Cys Ser Ala Asp Arg Cys Met Arg Tyr Arg Pro Ser Ala Arg | | | |

[0129]

| | 260 | 265 | 270 | |
|-------|---|-----|-----|-----|
| | Gln Pro Gln Pro Val Phe Met Thr Thr Ser His Val Gly Phe Arg Thr | | | |
| | 275 | 280 | 285 | |
| | Ile Arg Gln | | | |
| | 290 | | | |
| <210> | 71 | | | |
| <211> | 780 | | | |
| <212> | DNA | | | |
| <213> | <i>Ralstonia metallidurans</i> | | | |
| <400> | 71 | | | |
| | atggtcgcgg gcgggatggt gttcgtcggc accaacagcc cggtgccgct gcgcgaatac | | | 60 |
| | tggcgctggt ggcgcttcgt acctggcgcg gactggcgtc acccgaccgg cccgggcagt | | | 120 |
| | tccatcgaag gcaaggacaa tcatcccgtc gtgcaggtct cgtatgaaga cgcgcaggcg | | | 180 |
| | tacgccaagt gggccggcaa gcgtctgccc accgaggccg agtgggagtt tgccgccgt | | | 240 |
| | ggcggcctgg agcaggccac ctacgcctgg ggtgacaagt tcgcgccgga tggccggcag | | | 300 |
| | atggcgaatg tctggcaggg ccagcaggtg cagccgttcc cggtggtcag cgccaaggcg | | | 360 |
| | ggcggcgcgg ctggcaccag tgctgtcggc acgttcccgg gcaatggcta tgggctctat | | | 420 |
| | gacatgaccg gcaacgcctg gcagtgggtg gccgactggt atcgcgcgga ccagttccgc | | | 480 |
| | cgcgaagcca cgggtggcggc agtgctgcag aatccgaccg gcccgccga ttcgtgggac | | | 540 |
| | ccgaccgaac ctggcgtgcc ggtgtcggcg cccaagcggg tcacgcgcgg tggtcgttc | | | 600 |
| | ctctgcaacg aggacttctg cctcagctac cgcccagtg cccggcgcgg taccgaccg | | | 660 |
| | tacaccagca tgtcgcacct aggttccgg ctcgtgatgg atgaccccg ttgggcagaa | | | 720 |
| | gttcgcaagc agccagccgt ggcaatggcc gcgggcgggc agcagaacgt gcagaaataa | | | 780 |

[0130]

<210> 72
 <211> 259
 <212> PRT
 <213> *Ralstonia metallidurans*

 <400> 72

 Met Val Ala Gly Gly Met Val Phe Val Gly Thr Asn Ser Pro Val Pro
 1 5 10 15

 Leu Arg Glu Tyr Trp Arg Trp Trp Arg Phe Val Pro Gly Ala Asp Trp
 20 25 30

 Arg His Pro Thr Gly Pro Gly Ser Ser Ile Glu Gly Lys Asp Asn His
 35 40 45

 Pro Val Val Gln Val Ser Tyr Glu Asp Ala Gln Ala Tyr Ala Lys Trp
 50 55 60

 Ala Gly Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Phe Ala Ala Arg
 65 70 75 80

 Gly Gly Leu Glu Gln Ala Thr Tyr Ala Trp Gly Asp Lys Phe Ala Pro
 85 90 95

 Asp Gly Arg Gln Met Ala Asn Val Trp Gln Gly Gln Gln Val Gln Pro
 100 105 110

 Phe Pro Val Val Ser Ala Lys Ala Gly Gly Ala Ala Gly Thr Ser Ala
 115 120 125

 Val Gly Thr Phe Pro Gly Asn Gly Tyr Gly Leu Tyr Asp Met Thr Gly
 130 135 140

 Asn Ala Trp Gln Trp Val Ala Asp Trp Tyr Arg Ala Asp Gln Phe Arg

[0131]

| | | | | |
|---|-----|-----|-----|-----|
| 145 | 150 | 155 | 160 | |
| Arg Glu Ala Thr Val Ala Ala Val Leu Gln Asn Pro Thr Gly Pro Ala | 165 | 170 | 175 | |
| Asp Ser Trp Asp Pro Thr Glu Pro Gly Val Pro Val Ser Ala Pro Lys | 180 | 185 | 190 | |
| Arg Val Thr Arg Gly Gly Ser Phe Leu Cys Asn Glu Asp Phe Cys Leu | 195 | 200 | 205 | |
| Ser Tyr Arg Pro Ser Ala Arg Arg Gly Thr Asp Pro Tyr Thr Ser Met | 210 | 215 | 220 | |
| Ser His Leu Gly Phe Arg Leu Val Met Asp Asp Ala Arg Trp Ala Glu | 225 | 230 | 235 | 240 |
| Val Arg Lys Gln Pro Ala Val Ala Met Ala Ala Gly Gly Gln Gln Asn | 245 | 250 | 255 | |
| Val Gln Lys | | | | |
| <210> 73 | | | | |
| <211> 876 | | | | |
| <212> DNA | | | | |
| <213> <i>Prochlorococcus marinus</i> | | | | |
| <400> 73 | | | | |
| gtgaccacat cttgccagt agagatgta accatcccc cagggtcta tcgagttggc | | | 60 | |
| tgtgatcgt gctatccgga tggttcagtt cgctgctatc cggaggaaac acccgcgca | | | 120 | |
| gaagtgcagc ttgactcatt ccagatcgac gtagggccag tcaccaatgc ccagttccga | | | 180 | |
| gctttcgta gcgccagca gcatecaca gtctcggagc taccacctga tccaagctc | | | 240 | |

[0132]

tatcccgatc tagcgccga ggaacgcatc cctgaatcag ttgtctttca accgctcca 300
gcaacggtgg atcgagcaa acccttgagc tggtagacc tcattggctgg ggctgattgg 360
cgtcatcccc aaggaccga aagcacgac gatggccttg atgacaccc tgtcgtgcat 420
gtcgcctatg ccgacgcat cgcctatgcc cattgggctg gcaagcgtct ccctctgct 480
gaagagtggg aagtagccg ccgcggggt cttgtcgatg cccaatagc ctgggggaat 540
gaactcactc ccaataaccg ctggatggcg aacatctggc aaggtccttt cccttggcac 600
aacgaggagc tagacggctg gttctggacc tcgcccgttg gcagcttcc tgccaacggc 660
tatggactct tggatgttg cggcaatgig tgggaatgga ccaactctgt ttatcccgig 720
gcgtcaggcc accaggaacg gcgaactatc aaaggcggat cgtttctctg cgcagataat 780
tactgcgtac gttatcgacc ctctgcacta caaggccaga cagtagacac tgccacctgt 840
cacatgggct ttcgctgtgc aaaaggaggg ccttga 876

<210> 74

<211> 291

<212> PRT

<213> *Prochlorococcus marinus*

<400> 74

Met Thr Thr Ser Leu Pro Val Glu Met Val Thr Ile Pro Ala Gly Leu
1 5 10 15

Tyr Arg Val Gly Cys Asp Arg Cys Tyr Pro Asp Gly Ser Val Arg Cys
20 25 30

Tyr Pro Glu Glu Thr Pro Ala Arg Glu Val Gln Leu Asp Ser Phe Gln
35 40 45

Ile Asp Val Gly Pro Val Thr Asn Ala Gln Phe Arg Ala Phe Val Ser

[0133]

| 50 | 55 | 60 |
|--|-----|---------|
| Ala Thr Gln His Leu Thr Val Ser Glu Leu Pro Pro Asp Pro Thr Leu 65 | 70 | 75 80 |
| Tyr Pro Asp Leu Ala Pro Glu Glu Arg Ile Pro Glu Ser Val Val Phe 85 | 90 | 95 |
| Gln Pro Pro Pro Ala Thr Val Asp Arg Ser Lys Pro Leu Ser Trp Trp 100 | 105 | 110 |
| Thr Leu Met Ala Gly Ala Asp Trp Arg His Pro Gln Gly Pro Glu Ser 115 | 120 | 125 |
| Thr Ile Asp Gly Leu Asp Asp His Pro Val Val His Val Ala Tyr Ala 130 | 135 | 140 |
| Asp Ala Ile Ala Tyr Ala His Trp Ala Gly Lys Arg Leu Pro Ser Ala 145 | 150 | 155 160 |
| Glu Glu Trp Glu Val Ala Ala Arg Gly Gly Leu Val Asp Ala Gln Tyr 165 | 170 | 175 |
| Ala Trp Gly Asn Glu Leu Thr Pro Asn Asn Arg Trp Met Ala Asn Ile 180 | 185 | 190 |
| Trp Gln Gly Pro Phe Pro Trp His Asn Glu Glu Leu Asp Gly Trp Phe 195 | 200 | 205 |
| Trp Thr Ser Pro Val Gly Ser Phe Pro Ala Asn Gly Tyr Gly Leu Leu 210 | 215 | 220 |
| Asp Val Cys Gly Asn Val Trp Glu Trp Thr Asn Ser Val Tyr Pro Val | | |

[0134]

| | | | |
|---|-----|-----|-----|
| 225 | 230 | 235 | 240 |
| Ala Ser Gly His Gln Glu Arg Arg Thr Ile Lys Gly Gly Ser Phe Leu | | | |
| | 245 | 250 | 255 |
| Cys Ala Asp Asn Tyr Cys Val Arg Tyr Arg Pro Ser Ala Leu Gln Gly | | | |
| | 260 | 265 | 270 |
| Gln Thr Val Asp Thr Ala Thr Cys His Met Gly Phe Arg Cys Ala Lys | | | |
| | 275 | 280 | 285 |
| Gly Gly Pro | | | |
| 290 | | | |

<210> 75

<211> 1017

<212> DNA

<213> 新月柄杆菌 CB15

<400> 75

```

ttgggaaaac tgacggcgcct tcccgtcctg atgcttctgg cgctggccgg ctgcggccag    60
ccggcgccca agccttgccct ggccggacctg ccggttccag atccccagaa ccgcacggcg    120
ggtatggttc ggctggcggg ccggcgacttc cagatgggcg ctgcgccgct gctccggag    180
gagggaccgc cccagacggt cacggtcccg ccgcttctgga tcgatcagac agaggtcacc    240
aacgccgcct tcgcgcgggt cgtcgaggcc acgggttata gcaccgtggc cgagcgaccg    300
ctcgaccccc cgcgctacgc ccacgtaccg gcggcgcagc ggcgtccggc ctgctcgtc    360
ttcgtggggg cgaagggggc gaggtcggac gatecttccc aatggtggca ggtgatcccc    420
ggcgccgact ggccgcatcc cgaaggtccc ggctcgaaca tccggggcag ggacgcctgg    480
ccggtggtgc atatcgctg ggaggacgcc atggcctacg cccgctggtt gggccgtgac    540
ctgccacag aggcogaatg ggagtaagcc gcgcgcggcg ggctggttgg caagcgtac    600

```

[0135]

acctggggcg accaggetca ggatcctgca aagccgcgcg ccaatacttg gcaaggcgtg 660
 ttcccggccc aggaccttgg caatgacggc ttcaaggcca agcccgcgcc ggtcggctgc 720
 ttcccggcca acggctatgg cctgcgcgac atggccggca atgtctggga gtggaccgcg 780
 gactggttca agccgggcct ggatccggtc agcgtcctcg aaaccggcgg gccgcccag 840
 gcccgcgcg tggatcccga ggaccgaac acgcccgaac acgtcgtgaa gggcggttcg 900
 ttctgtgctg ccgacgacta ctgcttccgc tatcgacctg cggcgcgaac gccggggccg 960
 ccggacagcg gcgcatcgca tgcgggttc cgcaccgtgc tccgcgccga gcgctga 1017

<210> 76

<211> 338

<212> PRT

<213> 新月柄杆菌 CB15

<400> 76

Met Gly Lys Leu Thr Ala Leu Pro Val Leu Met Leu Leu Ala Leu Ala
 1 5 10 15

Gly Cys Gly Gln Pro Ala Pro Lys Ala Cys Leu Ala Asp Leu Pro Val
 20 25 30

Pro Asp Pro Gln Asn Arg Thr Ala Gly Met Val Arg Leu Ala Gly Gly
 35 40 45

Asp Phe Gln Met Gly Ala Ala Pro Leu Arg Pro Glu Glu Gly Pro Pro
 50 55 60

Gln Thr Val Thr Val Pro Pro Phe Trp Ile Asp Gln Thr Glu Val Thr
 65 70 75 80

Asn Ala Ala Phe Ala Arg Phe Val Glu Ala Thr Gly Tyr Arg Thr Val

[0136]

| | 85 | 90 | 95 |
|---|-----|-----|-----|
| Ala Glu Arg Pro Leu Asp Pro Ala Arg Tyr Ala His Val Pro Ala Ala | 100 | 105 | 110 |
| Gln Arg Arg Pro Ala Ser Leu Val Phe Val Gly Ala Lys Gly Ala Arg | 115 | 120 | 125 |
| Ser Asp Asp Pro Ser Gln Trp Trp Gln Val Ile Pro Gly Ala Asp Trp | 130 | 135 | 140 |
| Arg His Pro Glu Gly Pro Gly Ser Asn Ile Arg Gly Arg Asp Ala Trp | 145 | 150 | 155 |
| Pro Val Val His Ile Ala Trp Glu Asp Ala Met Ala Tyr Ala Arg Trp | 165 | 170 | 175 |
| Leu Gly Arg Asp Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ala Ala Arg | 180 | 185 | 190 |
| Gly Gly Leu Val Gly Lys Arg Tyr Thr Trp Gly Asp Gln Ala Gln Asp | 195 | 200 | 205 |
| Pro Ala Lys Pro Arg Ala Asn Thr Trp Gln Gly Val Phe Pro Ala Gln | 210 | 215 | 220 |
| Asp Leu Gly Asn Asp Gly Phe Lys Ala Lys Pro Ala Pro Val Gly Cys | 225 | 230 | 235 |
| Phe Pro Pro Asn Gly Tyr Gly Leu Arg Asp Met Ala Gly Asn Val Trp | 245 | 250 | 255 |
| Glu Trp Thr Arg Asp Trp Phe Lys Pro Gly Leu Asp Pro Val Ser Val | | | |

[0137]

| | | | | |
|---|-----|-----|-----|--|
| | 260 | 265 | 270 | |
| Leu Glu Thr Gly Gly Pro Pro Glu Ala Arg Ala Leu Asp Pro Glu Asp | | | | |
| | 275 | 280 | 285 | |
| Pro Asn Thr Pro Lys His Val Val Lys Gly Gly Ser Phe Leu Cys Ala | | | | |
| | 290 | 295 | 300 | |
| Asp Asp Tyr Cys Phe Arg Tyr Arg Pro Ala Ala Arg Thr Pro Gly Pro | | | | |
| 305 | 310 | 315 | 320 | |
| Pro Asp Ser Gly Ala Ser His Val Gly Phe Arg Thr Val Leu Arg Ala | | | | |
| | 325 | 330 | 335 | |

Glu Arg

<210> 77

<211> 900

<212> DNA

<213> 结核分枝杆菌 H37Rv

<400> 77

gtgctgaccg agttggttga cctgcccggc ggatcgttcc gcatgggctc gacgcgcttc 60

taccccgaag aagcgcgat tcataccgtg accgtgcgcg cctttgcggt agagcgacac 120

ccggtgacca acgcgaatt tgccgaattc gtctccgga caggctatgt gacggttgca 180

gaacaacccc ttgaccccgg gctctaccca ggagtggacg cagcagacct gtgtcccgtt 240

gcgatggtgt tttgtccgac ggcccggccg gtcgacctgc gtgactggcg gcaatggtgg 300

gactgggtac ctggcgctg ctggcgccat ccgtttggcc gggacagcga tctcggcag 360

cgagccggcc acccggtcgt acaggtggcc tatccggacg ccgtggccta cgcacgatgg 420

gctggtcgac gcctaccgac cgaggccgag tgggagtagc cggcccgtgg cggaaccacg 480

[0138]

gcaacctatg cgtggggcga ccaggagaag ccggggggca tgctcatggc gaacacctgg 540
 cagggccggt ttccttaccg caacgacggt gcattgggct gggtgggaac ctccccggtg 600
 ggcaggtttc cggccaacgg gtttgcttg ctgacatga tcggaaacgt ttgggagtgg 660
 accaccaccg agttctatcc acaccatcgc atcgatccac cctcgacggc ctgctgcgca 720
 ccggtcaagc tcgctacagc cgccgaccg acgatcagcc agaccctcaa gggcggctcg 780
 cacctgtgcg cgccggagta ctgccaccgc taccgcccg cggcgcgctc gccgcagtcg 840
 caggacaccg cgaccacca tctcgggttc cgggtcgtgg ccgaccgggt gtccgggtag 900

<210> 78

<211> 299

<212> PRT

<213> 结核分枝杆菌 H37Rv

<400> 78

Met Leu Thr Glu Leu Val Asp Leu Pro Gly Gly Ser Phe Arg Met Gly
 1 5 10 15

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

Arg Ala Phe Ala Val Glu Arg His Pro Val Thr Asn Ala Gln Phe Ala
 35 40 45

Glu Phe Val Ser Ala Thr Gly Tyr Val Thr Val Ala Glu Gln Pro Leu
 50 55 60

Asp Pro Gly Leu Tyr Pro Gly Val Asp Ala Ala Asp Leu Cys Pro Gly
 65 70 75 80

Ala Met Val Phe Cys Pro Thr Ala Gly Pro Val Asp Leu Arg Asp Trp

[0139]

| 85 | 90 | 95 |
|--|-----|-----|
| Arg Gln Trp Trp Asp Trp Val Pro Gly Ala Cys Trp Arg His Pro Phe 100 | 105 | 110 |
| Gly Arg Asp Ser Asp Ile Ala Asp Arg Ala Gly His Pro Val Val Gln 115 | 120 | 125 |
| Val Ala Tyr Pro Asp Ala Val Ala Tyr Ala Arg Trp Ala Gly Arg Arg 130 | 135 | 140 |
| Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ala Ala Arg Gly Gly Thr Thr 145 | 150 | 155 |
| Ala Thr Tyr Ala Trp Gly Asp Gln Glu Lys Pro Gly Gly Met Leu Met 165 | 170 | 175 |
| Ala Asn Thr Trp Gln Gly Arg Phe Pro Tyr Arg Asn Asp Gly Ala Leu 180 | 185 | 190 |
| Gly Trp Val Gly Thr Ser Pro Val Gly Arg Phe Pro Ala Asn Gly Phe 195 | 200 | 205 |
| Gly Leu Leu Asp Met Ile Gly Asn Val Trp Glu Trp Thr Thr Thr Glu 210 | 215 | 220 |
| Phe Tyr Pro His His Arg Ile Asp Pro Pro Ser Thr Ala Cys Cys Ala 225 | 230 | 235 |
| Pro Val Lys Leu Ala Thr Ala Ala Asp Pro Thr Ile Ser Gln Thr Leu 245 | 250 | 255 |
| Lys Gly Gly Ser His Leu Cys Ala Pro Glu Tyr Cys His Arg Tyr Arg | | |

[0140]

260

265

270

Pro Ala Ala Arg Ser Pro Gln Ser Gln Asp Thr Ala Thr Thr His Ile
 275 280 285

Gly Phe Arg Cys Val Ala Asp Pro Val Ser Gly
 290 295

<210> 79

<211> 7

<212> PRT

<213> 人工的

<220>

<223> 原核生物和原核生物中保守的结构域

<220>

<221> DOMAIN

<222> (1).. (7)

<223> 保守结构域

<220>

<221> MISC_FEATURE

<222> (3).. (4)

<223> 任何氨基酸

<220>

<221> MISC_FEATURE

<222> (6).. (6)

<223> 任何氨基酸

<220>

<221> MISC_FEATURE

<222> (6).. (6)

<223> Gly 或 Ala

<400> 79

Arg Val Xaa Xaa Gly Xaa Ser
 1 5

[0141]

<210> 80
 <211> 630
 <212> DNA
 <213> *Oncorhynchus mykiss*

<400> 80
 tcaggtggct gctgccccct ggtggttgcc tgtcagagga gcagactgga ggcaccctga 60
 gggccccgac tccagcatca cagacaggct ggaccaccct gtgctgcatg tgtcatggca 120
 ggacgctgtg gcctactgct cctgggccta caagagacta cccacagagg ctgagtggga 180
 gtacgcctgc agagggggcc tacaggagag actttaccgc tgggggaaca aactgaaacc 240
 taaaggacag cactacgcca acctctggca gggaaagttc cccacacaca actcagaaga 300
 ggacgggtac actaaaacct caccagtga gtcatttctt gcaaatggct atggcctgta 360
 caacatggta gggaatgcat gggagtggac atctgactgg tggactgtac accacaccac 420
 agatgaacag cacaaccggc caggtccacc atcaggcaca gaccgagtga agaaaggagg 480
 ctctacatg tgccataagt catactgta caggtacagg tgtgcagcac ggagtcagaa 540
 caccctgac agctctgect ctaacctagg gttccgctgt gtctcccagg agcagccgta 600
 acctttcacc ctgaccctg acatgggtag 630

<210> 81
 <211> 655
 <212> DNA
 <213> 斑马鱼

<220>
 <221> misc_feature
 <222> (590)..(590)
 <223> n 是 a、c、g 或 t

<220>
 <221> misc_feature

[0142]

<222> (626)..(626)
 <223> n 是 a、c、g 或 t

 <400> 81
 caaatggttt tatttacata aaaaaatcct cttagtttga agtgtaagac agtgagatta 60

 gtgatgtttg aggttatgga tcaacatcag aggcgcagcg gaagcccaag ttcgaggctg 120

 aactgtccgg tgtgttctga ctgcgagcgg cacacctgta tctgtagcag taagacttgt 180

 ggcacatgta ggatcctcct ttcttgactc tgtctgtccc tgattctggt ccctttgggt 240

 taaacttgtc ttctgcagtg tgatgcacag tccaccagtc tgccgtccac tcccacgcat 300

 ttcccacat gtcatacagg ccaaagccat tgggaggaaa agacatcacc ggggatgtgt 360

 tggcatagcc gtctctgca gtgttgtgat tagggaaatc tccctgccac aggttagcat 420

 agtgctgccc tcttggcatt aatttatttc cccatgggta catcctgtcc tgtagtcctc 480

 ctctacagge caactcccat tcagcttctg taggaagtct gcgtttgcc cattgacagt 540

 acgcccgtgc atcatcccat gaaacatgca gagcagggtg attcattctn gtgtgtatgg 600

 ttgaatctgg tcctttctgg tglctncagt ctgcaccttt cactggtgac cacca 655

 <210> 82
 <211> 773
 <212> DNA
 <213> *Oryzias latipes*

 <220>
 <221> misc_feature
 <222> (690)..(690)
 <223> n 是 a、c、g 或 t

 <220>
 <221> misc_feature
 <222> (755)..(755)
 <223> n 是 a、c、g 或 t

 <400> 82

[0143]

| | |
|---|-----|
| tctccttttt tccataaata acattagagt ccttacattc tgcctttaca tacattgtca | 60 |
| gagacagtac aaaaaatctg cctttgtaa attagagtta caaaaatata ttttagattt | 120 |
| gacttcttca gaattgtcgg tggcagcaaa agaatcggat tgatctcatg acaagagcgt | 180 |
| gagccagaag ttcttggatc aaactgattt gtttctgtca tcgtttctgt tcagcagcac | 240 |
| agcgaaaacc aagattggaa gcggagctgt ctggagtgtt ttggcttcga gcagcacatc | 300 |
| tgtacctgta acaataagac ttgtggcaca tgtacgagcc tcctttcttc accttatctg | 360 |
| tgcttgacgg aggaccctt gggttgtgct gatggctctgt tgtgtggtgc acgctccacc | 420 |
| agtctgaggt ccaactccat gcgttcccca ccatgtcata cagaccaaaa gcattgcctg | 480 |
| ggaaggacat caccggggag gttttagtgt agccatcctc tgcagagttg tgtgctggga | 540 |
| attccccctg ccagaggttg gcgtaatget gtccctttgg gtttagcttg tttccccagg | 600 |
| ggtagagtct gtccctcagg ccgccctgc aggcaacctc ccaactctgcc tcagtgggaa | 660 |
| gtctcttggt gaccagagg cagtaagccn aggcatcatt ccagaaacc tgaacgacgg | 720 |
| atgatccatc ctgtctgtga tglggagtc tggancttca gggtgcttcc agt | 773 |

<210> 83

<211> 566

<212> DNA

<213> 非洲爪蟾

<220>

<221> misc_feature

<222> (6)..(6)

<223> n 是 a、c、g 或 t

<220>

<221> misc_feature

<222> (47)..(47)

<223> n 是 a、c、g 或 t

<220>

[0144]

| | | |
|-------|--|-----|
| <221> | misc_feature | |
| <222> | (81)..(81) | |
| <223> | n 是 a、c、g 或 t | |
| <400> | 83 | |
| | atatgnaact aaaggtaatg taattggaat gätggatttc acaaggntg agagtccct | 60 |
| | attgctcctg cttgtcgtgt nacaggtcac ggagccggcg ccacacagcg aaatcccagg | 120 |
| | ttggaggccg agctgtcggg tgtattctga cttcgagcag cacagcgata cctgtagcaa | 180 |
| | taggactcat ggcacatgta ggagcctcct ttcttctactc tatcatttcc cgtagaaggt | 240 |
| | ccittcgggt tgtgaacctc atctgctgta tgatgagtgt cccaccaatc agatgtccac | 300 |
| | tccaagcat ttcccacat gttatataga ccataacatc ttgctgggaa agcagttaca | 360 |
| | ggtgaagtct gcacataacc atcctctcca gtgttttggg ttgaaaatc cccctgccag | 420 |
| | acattcgcac aatgttgctc ctttggttcc agcttgttcc cccatggaaa aatcctgttc | 480 |
| | tcaagtcctc cgcggcagge gtattccac tcagcttcag ttggaaggcg tttacctgcc | 540 |
| | caggtgcaga aagcagaagc atcatt | 566 |
| <210> | 84 | |
| <211> | 647 | |
| <212> | DNA | |
| <213> | <i>Silurana tropicalis</i> | |
| <400> | 84 | |
| | gccgcttttt tttttttttt tttttttttt catcacaanaa ataattttat taataaana | 60 |
| | ggattttggt ttcatctta ttatgaagga caaggaatgt cattgaaatt tttgtttca | 120 |
| | caaggtcttg ggagttcctt cctgctcagg tcattttgca gtggtcacgg agccgacgc | 180 |
| | acgcagcggg atcccaggtt agaggccgag ctgtcaggtg tattctgact tcgagcagca | 240 |
| | cagcgatacc tgtagcagta ggactcatgg cacatgtatg agccctcttt tttcaccttg | 300 |
| | tcctttcccg taaaaggacc tttcgggttg taagtctcat ctgctgtatg atgagtgtcc | 360 |

[0145]

| | |
|---|-----|
| caccaatcgg atgtccactc ccaagcattt cccacatgt tatataggct ataaccattg | 420 |
| gctgggaaag cggttacagg tgaagtctgc acatagecgt cctctccagt gttttgggtt | 480 |
| ggaaattccc cctgccagac attcgcataa tgttctcct ttggttcag cttgttcccc | 540 |
| cacggaaaaa gcctgttctc aagtccccca cgggaggcat attcccactc agcttctgtc | 600 |
| ggaaggcgt taccgcccc ggtgcagaag gcagaagcat cgtcca | 647 |
| <210> 85 | |
| <211> 636 | |
| <212> DNA | |
| <213> 加拿大鲑鱼 | |
| <400> 85 | |
| atagacattt tttaaatatt ttacaacaaa atatattcca taaatatcca catgtcatgc | 60 |
| ggtaatcctg catttcatga agaacactga catcactggc tgtatgaaga ggtgcacttg | 120 |
| atgtgttcg cctggcgggc aagataggca gagttagcac cctagactag agccaatggc | 180 |
| gaatgtaca aaaagggaaa agtcagacta cccatgtcag ggtcaagggt aaaaggttac | 240 |
| ggctgctcct gggagacaca gcggaacct aggttagagg cagagctgtc aggggtgttc | 300 |
| tgactccgtg ctgcacacct gtacctgtaa cagtatgact tatggcacat gtaggagcct | 360 |
| ccittcttca ctcggtctgt gccigtatgt ggacctgccg ggttgtgccg ttcactctgt | 420 |
| gtgtgggtgta cagtccacca gtcagatgtc cactcccatg cattccctac catgtttgtac | 480 |
| aggccatagc catttgcagg aatgacttc actggtgagg ttttgggtgta cccgtcctct | 540 |
| tctgagttgt gtgtggggaa ctttccttgc cagaggttgg cgtagtgtcgc tcctttaggt | 600 |
| ttcagtttgt tccccacgg gtaaagtctg tctgt | 636 |
| <210> 86 | |
| <211> 415 | |
| <212> DNA | |
| <213> 野猪 | |

[0146]

| | | |
|---|-----|--|
| <400> 86 | | |
| agtttctgt gaccaacacc ggagaggatg gcttccgagg aactgcgcct gttgatgcct | 60 | |
| ttcttcccaa tggttatggc cttacaata tagtagggaa cgctgggaa tggacctcag | 120 | |
| actggtggac cattcacat gctgctgaag aaacaattaa cccatcaagt tcttctgct | 180 | |
| gcaccgaata acagagccgc cactacgtga tgaaagcaga gaaaggcccc ctttctggga | 240 | |
| aagaccgggt gaagaaagg ggatcctata tgtgccataa gtcctactgc tacaggtacc | 300 | |
| gctgtgctgc tcgaagccag aacacgccgg acagctcggc ttcaaatctg gggttccgct | 360 | |
| gtgcagctga ccaccagccc accacaggct gagtcaggaa gagtcttccc gaatc | 415 | |
| <210> 87 | | |
| <211> 595 | | |
| <212> DNA | | |
| <213> 黄牛 | | |
| <400> 87 | | |
| ccacgcgtcc gggggcaaca aactgcagcc gaaaggccag cattatagcc aacatcttg | 60 | |
| caaggcgagt ttctgtgac caacaccggg gaggacggct tccgaggac cgcgctgtt | 120 | |
| gacgccttc ctccaatgg ttattggctt atacaatata gtagggaacg cctgggagt | 180 | |
| gacttcagac tgggtgactg ttcaccattc tgctgaagaa acgattaacc caaaaggccc | 240 | |
| cccttctggg aaagaccggg tgaagaaagg tggatcctac atgtgccata aatcctattg | 300 | |
| ctacaggtat cgctgtgctg ctccaagcca gaacacacc gacagctctg cttcgaatct | 360 | |
| gggattcgt tgtgcagctg accacctgcc caccacaggc taagagccaa aaagagcctt | 420 | |
| cccgaaccg agaagtcgtg tctactctgc acgcgcttc cctcagaagg ctgaacaacc | 480 | |
| tgctgtgaag aattcccacc ccaaggtggg ttacatacct tgcccagtgg ccaaaggacc | 540 | |
| tatggcaaga ccaaattgct gagctgatca gcatgtgcgc tttattgggg gatgg | 595 | |

[0147]

<210> 88
 <211> 1611
 <212> DNA
 <213> 智人

<220>
 <221> CDS
 <222> (1)..(1608)
 <223> hSULF3

<400> 88
 atg cta ctg ctg tgg gtg tcg gtg gtc gca gcc ttg gcg ctg gcg gta 48
 Met Leu Leu Leu Trp Val Ser Val Val Ala Ala Leu Ala Leu Ala Val
 1 5 10 15

ctg gcc ccc gga gca ggg gag cag agg cgg aga gca gcc aaa gcg ccc 96
 Leu Ala Pro Gly Ala Gly Glu Gln Arg Arg Arg Ala Ala Lys Ala Pro
 20 25 30

aat gtg gtg ctg gtc gtg agc gac tcc ttc gat gga agg tta aca ttt 144
 Asn Val Val Leu Val Val Ser Asp Ser Phe Asp Gly Arg Leu Thr Phe
 35 40 45

cat cca gga agt cag gta gtg aaa ctt cct ttt atc aac ttt atg aag 192
 His Pro Gly Ser Gln Val Val Lys Leu Pro Phe Ile Asn Phe Met Lys
 50 55 60

aca cgt ggg act tcc ttt ctg aat gcc tac aca aac tct cca att tgt 240
 Thr Arg Gly Thr Ser Phe Leu Asn Ala Tyr Thr Asn Ser Pro Ile Cys
 65 70 75 80

tgc cca tca cgc gca gca atg tgg agt ggc ctc ttc act cac tta aca 288
 Cys Pro Ser Arg Ala Ala Met Trp Ser Gly Leu Phe Thr His Leu Thr
 85 90 95

gaa tct tgg aat aat ttt aag ggt cta gat cca aat tat aca aca tgg 336
 Glu Ser Trp Asn Asn Phe Lys Gly Leu Asp Pro Asn Tyr Thr Thr Trp
 100 105 110

atg gat gtc atg gag agg cat ggc tac cga aca cag aaa ttt ggg aaa 384
 Met Asp Val Met Glu Arg His Gly Tyr Arg Thr Gln Lys Phe Gly Lys
 115 120 125

[0148]

| | |
|---|-----|
| ctg gac tat act tca gga cat cac tcc att agt aat cgt gtg gaa gcg Leu Asp Tyr Thr Ser Gly His His Ser Ile Ser Asn Arg Val Glu Ala 130 135 140 | 432 |
| tgg aca aga gat gtt gct ttc tta ctc aga caa gaa ggc agg ccc atg Trp Thr Arg Asp Val Ala Phe Leu Leu Arg Gln Glu Gly Arg Pro Met 145 150 155 160 | 480 |
| gtt aat ctt atc cgt aac agg act aaa gtc aga gtg atg gaa agg gat Val Asn Leu Ile Arg Asn Arg Thr Lys Val Arg Val Met Glu Arg Asp 165 170 175 | 528 |
| tgg cag aat aca gac aaa gca gta aac tgg tta aga aag gaa gca att Trp Gln Asn Thr Asp Lys Ala Val Asn Trp Leu Arg Lys Glu Ala Ile 180 185 190 | 576 |
| aat tac act gaa cca ttt gtt att tac ttg gga tta aat tta cca cac Asn Tyr Thr Glu Pro Phe Val Ile Tyr Leu Gly Leu Asn Leu Pro His 195 200 205 | 624 |
| cct tac cct tca cca tct tct gga gaa aat ttt gga tct tca aca ttt Pro Tyr Pro Ser Pro Ser Ser Gly Glu Asn Phe Gly Ser Ser Thr Phe 210 215 220 | 672 |
| cac aca tct ctt tat tgg ctt gaa aaa gtg tct cat gat gcc atc aaa His Thr Ser Leu Tyr Trp Leu Glu Lys Val Ser His Asp Ala Ile Lys 225 230 235 240 | 720 |
| atc cca aag tgg tca cct ttg tca gaa atg cac cct gta gat tat tac Ile Pro Lys Trp Ser Pro Leu Ser Glu Met His Pro Val Asp Tyr Tyr 245 250 255 | 768 |
| tct tct tat aca aaa aac tgc act gga aga ttt aca aaa aaa gaa att Ser Ser Tyr Thr Lys Asn Cys Thr Gly Arg Phe Thr Lys Lys Glu Ile 260 265 270 | 816 |
| aag aat att aga gca ttt tat tat gct atg tgt gct gag aca gat gcc Lys Asn Ile Arg Ala Phe Tyr Tyr Ala Met Cys Ala Glu Thr Asp Ala 275 280 285 | 864 |
| atg ctt ggt gaa att att ttg gcc ctt cat caa tta gat ctt ctt cag Met Leu Gly Glu Ile Ile Leu Ala Leu His Gln Leu Asp Leu Leu Gln 290 295 300 | 912 |

[0149]

cac cag tat aat aaa gag cag ttt atc aag tgg aaa caa agt ata gga 1488
 His Gln Tyr Asn Lys Glu Gln Phe Ile Lys Trp Lys Gln Ser Ile Gly
 485 490 495

cag aat tat tca aac gtt ata gca aat ctt agg tgg cac caa gac tgg 1536
 Gln Asn Tyr Ser Asn Val Ile Ala Asn Leu Arg Trp His Gln Asp Trp
 500 505 510

cag aag gaa cca agg aag tat gaa aat gca att gat cag tgg ctt aaa 1584
 Gln Lys Glu Pro Arg Lys Tyr Glu Asn Ala Ile Asp Gln Trp Leu Lys
 515 520 525

acc cat atg aat cca aga gca gtt tga 1611
 Thr His Met Asn Pro Arg Ala Val
 530 535

<210> 89

<211> 536

<212> PRT

<213> 智人

<400> 89

Met Leu Leu Leu Trp Val Ser Val Val Ala Ala Leu Ala Leu Ala Val
 1 5 10 15

Leu Ala Pro Gly Ala Gly Glu Gln Arg Arg Arg Ala Ala Lys Ala Pro
 20 25 30

Asn Val Val Leu Val Val Ser Asp Ser Phe Asp Gly Arg Leu Thr Phe
 35 40 45

His Pro Gly Ser Gln Val Val Lys Leu Pro Phe Ile Asn Phe Met Lys
 50 55 60

Thr Arg Gly Thr Ser Phe Leu Asn Ala Tyr Thr Asn Ser Pro Ile Cys
 65 70 75 80

[0151]

Ser Ser Tyr Thr Lys Asn Cys Thr Gly Arg Phe Thr Lys Lys Glu Ile
 260 265 270

Lys Asn Ile Arg Ala Phe Tyr Tyr Ala Met Cys Ala Glu Thr Asp Ala
 275 280 285

Met Leu Gly Glu Ile Ile Leu Ala Leu His Gln Leu Asp Leu Leu Gln
 290 295 300

Lys Thr Ile Val Ile Tyr Ser Ser Asp His Gly Glu Leu Ala Met Glu
 305 310 315 320

His Arg Gln Phe Tyr Lys Met Ser Met Tyr Glu Ala Ser Ala His Val
 325 330 335

Pro Leu Leu Met Met Gly Pro Gly Ile Lys Ala Gly Leu Gln Val Ser
 340 345 350

Asn Val Val Ser Leu Val Asp Ile Tyr Pro Thr Met Leu Asp Ile Ala
 355 360 365

Gly Ile Pro Leu Pro Gln Asn Leu Ser Gly Tyr Ser Leu Leu Pro Leu
 370 375 380

Ser Ser Glu Thr Phe Lys Asn Glu His Lys Val Lys Asn Leu His Pro
 385 390 395 400

Pro Trp Ile Leu Ser Glu Phe His Gly Cys Asn Val Asn Ala Ser Thr
 405 410 415

Tyr Met Leu Arg Thr Asn His Trp Lys Tyr Ile Ala Tyr Ser Asp Gly
 420 425 430

[0153]

Ala Ser Ile Leu Pro Gln Leu Phe Asp Leu Ser Ser Asp Pro Asp Glu
 435 440 445

Leu Thr Asn Val Ala Val Lys Phe Pro Glu Ile Thr Tyr Ser Leu Asp
 450 455 460

Gln Lys Leu His Ser Ile Ile Asn Tyr Pro Lys Val Ser Ala Ser Val
 465 470 475 480

His Gln Tyr Asn Lys Glu Gln Phe Ile Lys Trp Lys Gln Ser Ile Gly
 485 490 495

Gln Asn Tyr Ser Asn Val Ile Ala Asn Leu Arg Trp His Gln Asp Trp
 500 505 510

Gln Lys Glu Pro Arg Lys Tyr Glu Asn Ala Ile Asp Gln Trp Leu Lys
 515 520 525

Thr His Met Asn Pro Arg Ala Val
 530 535

<210> 90
 <211> 1722
 <212> DNA
 <213> 智人

<220>
 <221> CDS
 <222> (1)..(1719)
 <223> hSULF4

<400> 90

atg ggg gcg ctg gca gga ttc tgg atc ctc tgc ctc ctc act tat ggt
 Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu Leu Thr Tyr Gly
 1 5 10 15

48

[0154]

| | |
|---|-----|
| tac ctg tcc tgg ggc cag gcc tta gaa gag gag gaa gaa ggg gcc tta | 96 |
| Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu Glu Gly Ala Leu | |
| 20 25 30 | |
| cta gct caa gct gga gag aaa cta gag ccc agc aca act tcc acc tcc | 144 |
| Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro Ser Thr Thr Ser Thr Ser | |
| 35 40 45 | |
| cag ccc cat ctc att ttc atc cta gcg gat gat cag gga ttt aga gat | 192 |
| Gln Pro His Leu Ile Phe Ile Leu Ala Asp Asp Gln Gly Phe Arg Asp | |
| 50 55 60 | |
| gtg ggt tac cac gga tct gag att aaa aca cct act ctt gac aag ctc | 240 |
| Val Gly Tyr His Gly Ser Glu Ile Lys Thr Pro Thr Leu Asp Lys Leu | |
| 65 70 75 80 | |
| gct gcc gaa gga gtt aaa ctg gag aac tac tat gtc cag cct att tgc | 288 |
| Ala Ala Glu Gly Val Lys Leu Glu Asn Tyr Tyr Val Gln Pro Ile Cys | |
| 85 90 95 | |
| aca cca tcc agg agt cag ttt att act gga aag tat cag ata cac acc | 336 |
| Thr Pro Ser Arg Ser Gln Phe Ile Thr Gly Lys Tyr Gln Ile His Thr | |
| 100 105 110 | |
| gga ctt caa cat tct atc ata aga cct acc caa ccc aac tgt tta cct | 384 |
| Gly Leu Gln His Ser Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro | |
| 115 120 125 | |
| ctg gac aat gcc acc cta cct cag aaa ctg aag gag gtt gga tat tca | 432 |
| Leu Asp Asn Ala Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser | |
| 130 135 140 | |
| acg cat atg gtc gga aaa tgg cac ttg ggt ttt tac aga aaa gaa tgc | 480 |
| Thr His Met Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys | |
| 145 150 155 160 | |
| atg ccc acc aga aga gga ttt gat acc ttt ttt ggt tcc ctt ttg gga | 528 |
| Met Pro Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly | |
| 165 170 175 | |
| agt ggg gat tac tat aca cac tac aaa tgt gac agt cct ggg atg tgt | 576 |
| Ser Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys | |
| 180 185 190 | |

[0155]

| | |
|---|------|
| ggc tat gac ttg tat gaa aac gac aat gct gcc tgg gac tat gac aat Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp Asn 195 200 205 | 624 |
| ggc ata tac tcc aca cag atg tac act cag aga gta cag caa atc tta Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln Ile Leu 210 215 220 | 672 |
| gct tcc cat aac ccc aca aag cct ata ttt tta tat att gcc tat caa Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln 225 230 235 240 | 720 |
| gct gtt cat tca cca ctg caa gct cct ggc agg tat ttc gaa cac tac Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe Glu His Tyr 245 250 255 | 768 |
| cga tcc att atc aac ata aac agg agg aga tat gct gcc atg ctt tcc Arg Ser Ile Ile Asn Ile Asn Arg Arg Arg Tyr Ala Ala Met Leu Ser 260 265 270 | 816 |
| tgc tta gat gaa gca atc aac aac gtg aca ttg gct cta aag act tat Cys Leu Asp Glu Ala Ile Asn Asn Val Thr Leu Ala Leu Lys Thr Tyr 275 280 285 | 864 |
| ggc ttc tat aac aac agc att atc att tac tct tca gat aat ggt ggc Gly Phe Tyr Asn Asn Ser Ile Ile Ile Tyr Ser Ser Asp Asn Gly Gly 290 295 300 | 912 |
| cag cct acg gca gga ggg agt aac tgg cct ctc aga ggt agc aaa gga Gln Pro Thr Ala Gly Gly Ser Asn Trp Pro Leu Arg Gly Ser Lys Gly 305 310 315 320 | 960 |
| aca tat tgg gaa gga ggg atc cgg gct gta ggc ttt gtg cat agc cca Thr Tyr Trp Glu Gly Gly Ile Arg Ala Val Gly Phe Val His Ser Pro 325 330 335 | 1008 |
| ctt ctg aaa aac aag gga aca gtg tgt aag gaa ctt gtg cac atc act Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val His Ile Thr 340 345 350 | 1056 |
| gac tgg tac ccc act ctc att tca ctg gct gaa gga cag att gat gag Asp Trp Tyr Pro Thr Leu Ile Ser Leu Ala Glu Gly Gln Ile Asp Glu 355 360 365 | 1104 |

[0156]

| | |
|---|------|
| gac att caa cta gat ggc tat gat atc tgg gag acc ata agt gag ggt Asp Ile Gln Leu Asp Gly Tyr Asp Ile Trp Glu Thr Ile Ser Glu Gly 370 375 380 | 1152 |
| ctt cgc tca ccc cga gta gat att ttg cat aac att gac ccc ata tac Leu Arg Ser Pro Arg Val Asp Ile Leu His Asn Ile Asp Pro Ile Tyr 385 390 395 400 | 1200 |
| acc aag gca aaa aat ggc tcc tgg gca gca ggc tat ggg atc tgg aac Thr Lys Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr Gly Ile Trp Asn 405 410 415 | 1248 |
| act gca atc cag tca gcc atc aga gtg cag cac tgg aaa ttg ctt aca Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr 420 425 430 | 1296 |
| gga aat cct ggc tac agc gac tgg gtc ccc cct cag tct ttc agc aac Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln Ser Phe Ser Asn 435 440 445 | 1344 |
| ctg gga ccg aac cgg tgg cac aat gaa cgg atc acc ttg tca act ggc Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr Leu Ser Thr Gly 450 455 460 | 1392 |
| aaa agt gta tgg ctt ttc aac atc aca gcc gac cca tat gag agg gtg Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro Tyr Glu Arg Val 465 470 475 480 | 1440 |
| gac cta tct aac agg tat cca gga atc gtg aag aag ctc cta cgg agg Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys Leu Leu Arg Arg 485 490 495 | 1488 |
| ctc tca cag ttc aac aaa act gca gtg ccg gtc agg tat ccc ccc aaa Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg Tyr Pro Pro Lys 500 505 510 | 1536 |
| gac ccc aga agt aac cct agg ctc aat gga ggg gtc tgg gga cca tgg Asp Pro Arg Ser Asn Pro Arg Leu Asn Gly Gly Val Trp Gly Pro Trp 515 520 525 | 1584 |
| tat aaa gag gaa acc aag aaa aag aag cca agc aaa aat cag gct gag Tyr Lys Glu Glu Thr Lys Lys Lys Lys Pro Ser Lys Asn Gln Ala Glu 530 535 540 | 1632 |

[0157]

aaa aag caa aag aaa agc aaa aaa aag aag aag aaa cag cag aaa gca 1680
 Lys Lys Gln Lys Lys Ser Lys Lys Lys Lys Lys Lys Gln Gln Lys Ala
 545 550 555 560

gtc tca ggt tca act tgc cat tca ggt gtt act tgt gga taa 1722
 Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys Gly
 565 570

<210> 91
 <211> 573
 <212> PRT
 <213> 智人

<400> 91

Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu Leu Thr Tyr Gly
 1 5 10 15

Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu Glu Glu Gly Ala Leu
 20 25 30

Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro Ser Thr Thr Ser Thr Ser
 35 40 45

Gln Pro His Leu Ile Phe Ile Leu Ala Asp Asp Gln Gly Phe Arg Asp
 50 55 60

Val Gly Tyr His Gly Ser Glu Ile Lys Thr Pro Thr Leu Asp Lys Leu
 65 70 75 80

Ala Ala Glu Gly Val Lys Leu Glu Asn Tyr Tyr Val Gln Pro Ile Cys
 85 90 95

Thr Pro Ser Arg Ser Gln Phe Ile Thr Gly Lys Tyr Gln Ile His Thr
 100 105 110

[0158]

Gly Leu Gln His Ser Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro
 115 120 125

Leu Asp Asn Ala Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser
 130 135 140

Thr His Met Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys
 145 150 155 160

Met Pro Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly
 165 170 175

Ser Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys
 180 185 190

Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp Asn
 195 200 205

Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln Ile Leu
 210 215 220

Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln
 225 230 235 240

Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe Glu His Tyr
 245 250 255

Arg Ser Ile Ile Asn Ile Asn Arg Arg Arg Tyr Ala Ala Met Leu Ser
 260 265 270

Cys Leu Asp Glu Ala Ile Asn Asn Val Thr Leu Ala Leu Lys Thr Tyr
 275 280 285

[0159]

Gly Phe Tyr Asn Asn Ser Ile Ile Ile Tyr Ser Ser Asp Asn Gly Gly
 290 295 300

Gln Pro Thr Ala Gly Gly Ser Asn Trp Pro Leu Arg Gly Ser Lys Gly
 305 310 315 320

Thr Tyr Trp Glu Gly Gly Ile Arg Ala Val Gly Phe Val His Ser Pro
 325 330 335

Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val His Ile Thr
 340 345 350

Asp Trp Tyr Pro Thr Leu Ile Ser Leu Ala Glu Gly Gln Ile Asp Glu
 355 360 365

Asp Ile Gln Leu Asp Gly Tyr Asp Ile Trp Glu Thr Ile Ser Glu Gly
 370 375 380

Leu Arg Ser Pro Arg Val Asp Ile Leu His Asn Ile Asp Pro Ile Tyr
 385 390 395 400

Thr Lys Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr Gly Ile Trp Asn
 405 410 415

Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr
 420 425 430

Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln Ser Phe Ser Asn
 435 440 445

Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr Leu Ser Thr Gly
 450 455 460

[0160]

Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro Tyr Glu Arg Val
 465 470 475 480

Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys Leu Leu Arg Arg
 485 490 495

Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg Tyr Pro Pro Lys
 500 505 510

Asp Pro Arg Ser Asn Pro Arg Leu Asn Gly Gly Val Trp Gly Pro Trp
 515 520 525

Tyr Lys Glu Glu Thr Lys Lys Lys Lys Pro Ser Lys Asn Gln Ala Glu
 530 535 540

Lys Lys Gln Lys Lys Ser Lys Lys Lys Lys Lys Lys Lys Gln Gln Lys Ala
 545 550 555 560

Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys Gly
 565 570

- <210> 92
- <211> 1710
- <212> DNA
- <213> 智人

- <220>
- <221> CDS
- <222> (1)..(1707)
- <223> hSULF5

<400> 92
 atg cac acc ctc act ggc ttc tcc ctg gtc agc ctg ctc agc ttc ggc 48
 Met His Thr Leu Thr Gly Phe Ser Leu Val Ser Leu Leu Ser Phe Gly
 1 5 10 15

[0161]

| | |
|---|-----|
| tac ctg tcc tgg gac tgg gcc aag ccg agc ttc gtg gcc gac ggg ccc Tyr Leu Ser Trp Asp Trp Ala Lys Pro Ser Phe Val Ala Asp Gly Pro 20 25 30 | 96 |
| ggg gag gct ggc gag cag ccc tcg gcc gct ccg ccc cag cct ccc cac Gly Glu Ala Gly Glu Gln Pro Ser Ala Ala Pro Pro Gln Pro Pro His 35 40 45 | 144 |
| atc atc ttc atc ctc acg gac gac caa ggc tac cac gac gtg ggc tac Ile Ile Phe Ile Leu Thr Asp Asp Gln Gly Tyr His Asp Val Gly Tyr 50 55 60 | 192 |
| cat ggt tca gat atc gag acc cct acg ctg gac agg ctg gcg gcc aag His Gly Ser Asp Ile Glu Thr Pro Thr Leu Asp Arg Leu Ala Ala Lys 65 70 75 80 | 240 |
| ggg gtc aag ttg gag aat tat tac atc cag ccc atc tgc acg cct tcg Gly Val Lys Leu Glu Asn Tyr Tyr Ile Gln Pro Ile Cys Thr Pro Ser 85 90 95 | 288 |
| cgg agc cag ctc ctc act ggc agg tac cag atc cac aca gga ctc cag Arg Ser Gln Leu Leu Thr Gly Arg Tyr Gln Ile His Thr Gly Leu Gln 100 105 110 | 336 |
| cat tcc atc atc cgc cca cag cag ccc aac tgc ctg ccc ctg gac cag His Ser Ile Ile Arg Pro Gln Gln Pro Asn Cys Leu Pro Leu Asp Gln 115 120 125 | 384 |
| gtg aca ctg cca cag aag ctg cag gag gca ggt tat tcc acc cat atg Val Thr Leu Pro Gln Lys Leu Gln Glu Ala Gly Tyr Ser Thr His Met 130 135 140 | 432 |
| gtg ggc aag tgg cac ctg ggc ttc tac cgg aag gag tgt ctg ccc acc Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys Leu Pro Thr 145 150 155 160 | 480 |
| cgt cgg ggc ttc gac acc ttc ctg ggc tcg ctc acg ggc aat gtg gac Arg Arg Gly Phe Asp Thr Phe Leu Gly Ser Leu Thr Gly Asn Val Asp 165 170 175 | 528 |
| tat tac acc tat gac aac tgt gat ggc cca ggc gtg tgc ggc ttc gac Tyr Tyr Thr Tyr Asp Asn Cys Asp Gly Pro Gly Val Cys Gly Phe Asp 180 185 190 | 576 |

[0162]

| | |
|---|------|
| ctg cac gag ggt gag aat gtg gcc tgg ggg ctc agc ggc cag tac tcc Leu His Glu Gly Glu Asn Val Ala Trp Gly Leu Ser Gly Gln Tyr Ser 195 200 205 | 624 |
| act atg ctt tac gcc cag cgc gcc agc cat atc ctg gcc agc cac agc Thr Met Leu Tyr Ala Gln Arg Ala Ser His Ile Leu Ala Ser His Ser 210 215 220 | 672 |
| cct cag cgt ccc ctc ttc ctc tat gtg gcc ttc cag gca gta cac aca Pro Gln Arg Pro Leu Phe Leu Tyr Val Ala Phe Gln Ala Val His Thr 225 230 235 240 | 720 |
| ccc ctg cag tcc cct cgt gag tac ctg tac cgc tac cgc acc atg ggc Pro Leu Gln Ser Pro Arg Glu Tyr Leu Tyr Arg Tyr Arg Thr Met Gly 245 250 255 | 768 |
| aat gtg gcc cgg cgg aag tac gcg gcc atg gtg acc tgc atg gat gag Asn Val Ala Arg Arg Lys Tyr Ala Ala Met Val Thr Cys Met Asp Glu 260 265 270 | 816 |
| gct gtg cgc aac atc acc tgg gcc ctc aag cgc tac ggt ttc tac aac Ala Val Arg Asn Ile Thr Trp Ala Leu Lys Arg Tyr Gly Phe Tyr Asn 275 280 285 | 864 |
| aac agt gtc atc atc ttc tcc agt gac aat ggt ggc cag act ttc tcg Asn Ser Val Ile Ile Phe Ser Ser Asp Asn Gly Gly Gln Thr Phe Ser 290 295 300 | 912 |
| ggg ggc agc aac tgg ccg ctc cga gga cgc aag ggc act tat tgg gaa Gly Gly Ser Asn Trp Pro Leu Arg Gly Arg Lys Gly Thr Tyr Trp Glu 305 310 315 320 | 960 |
| ggt ggc gtg cgg ggc cta ggc ttt gtc cac agt ccc ctg ctc aag cga Gly Gly Val Arg Gly Leu Gly Phe Val His Ser Pro Leu Leu Lys Arg 325 330 335 | 1008 |
| aag caa cgg aca agc cgg gca ctg atg cac atc act gac tgg tac ccg Lys Gln Arg Thr Ser Arg Ala Leu Met His Ile Thr Asp Trp Tyr Pro 340 345 350 | 1056 |
| acc ctg gtg ggt ctg gca ggt ggt acc acc tca gca gcc gat ggg cta Thr Leu Val Gly Leu Ala Gly Gly Thr Thr Ser Ala Ala Asp Gly Leu 355 360 365 | 1104 |

[0163]

| | |
|---|------|
| gat ggc tac gac gtg tgg ccg gcc atc agc gag ggc cgg gcc tca cca Asp Gly Tyr Asp Val Trp Pro Ala Ile Ser Glu Gly Arg Ala Ser Pro 370 375 380 | 1152 |
| cgc acg gag atc ctg cac aac att gac cca ctc tac aac cat gcc cag Arg Thr Glu Ile Leu His Asn Ile Asp Pro Leu Tyr Asn His Ala Gln 385 390 395 400 | 1200 |
| cat ggc tcc ctg gag ggc ggc ttt ggc atc tgg aac acc gcc gtg cag His Gly Ser Leu Glu Gly Gly Phe Gly Ile Trp Asn Thr Ala Val Gln 405 410 415 | 1248 |
| gct gcc atc cgc gtg ggt gag tgg aag ctg ctg aca gga gac ccc ggc Ala Ala Ile Arg Val Gly Glu Trp Lys Leu Leu Thr Gly Asp Pro Gly 420 425 430 | 1296 |
| tat ggc gat tgg atc cca ccg cag aca ctg gcc acc ttc ccg ggt agc Tyr Gly Asp Trp Ile Pro Pro Gln Thr Leu Ala Thr Phe Pro Gly Ser 435 440 445 | 1344 |
| tgg tgg aac ctg gaa cga atg gcc agt gtc cgc cag gcc gtg tgg ctc Trp Trp Asn Leu Glu Arg Met Ala Ser Val Arg Gln Ala Val Trp Leu 450 455 460 | 1392 |
| ttc aac atc agt gct gac cct tat gaa cgg gag gac ctg gct ggc cag Phe Asn Ile Ser Ala Asp Pro Tyr Glu Arg Glu Asp Leu Ala Gly Gln 465 470 475 480 | 1440 |
| cgg cct gat gtg gtc cgc acc ctg ctg gct cgc ctg gcc gaa tat aac Arg Pro Asp Val Val Arg Thr Leu Leu Ala Arg Leu Ala Glu Tyr Asn 485 490 495 | 1488 |
| cgc aca gcc atc ccg gta cgc tac cca gct gag aac ccc cgg gct cat Arg Thr Ala Ile Pro Val Arg Tyr Pro Ala Glu Asn Pro Arg Ala His 500 505 510 | 1536 |
| cct gac ttt aat ggg ggt gct tgg ggg ccc tgg gcc agt gat gag gaa Pro Asp Phe Asn Gly Gly Ala Trp Gly Pro Trp Ala Ser Asp Glu Glu 515 520 525 | 1584 |
| gag gag gaa gag gaa ggg agg gct cga agc ttc tcc cgg ggt cgt cgc Glu Glu Glu Glu Glu Gly Arg Ala Arg Ser Phe Ser Arg Gly Arg Arg 530 535 540 | 1632 |

[0164]

aag aaa aaa tgc aag att tgc aag ctt cga tcc ttt ttc cgt aaa ctc 1680
 Lys Lys Lys Cys Lys Ile Cys Lys Leu Arg Ser Phe Phe Arg Lys Leu
 545 550 555 560

aac acc agg cta atg tcc caa cgg atc tga 1710
 Asn Thr Arg Leu Met Ser Gln Arg Ile
 565

<210> 93
 <211> 569
 <212> PRT
 <213> 智人

<400> 93

Met His Thr Leu Thr Gly Phe Ser Leu Val Ser Leu Leu Ser Phe Gly
 1 5 10 15

Tyr Leu Ser Trp Asp Trp Ala Lys Pro Ser Phe Val Ala Asp Gly Pro
 20 25 30

Gly Glu Ala Gly Glu Gln Pro Ser Ala Ala Pro Pro Gln Pro Pro His
 35 40 45

Ile Ile Phe Ile Leu Thr Asp Asp Gln Gly Tyr His Asp Val Gly Tyr
 50 55 60

His Gly Ser Asp Ile Glu Thr Pro Thr Leu Asp Arg Leu Ala Ala Lys
 65 70 75 80

Gly Val Lys Leu Glu Asn Tyr Tyr Ile Gln Pro Ile Cys Thr Pro Ser
 85 90 95

Arg Ser Gln Leu Leu Thr Gly Arg Tyr Gln Ile His Thr Gly Leu Gln
 100 105 110

[0165]

His Ser Ile Ile Arg Pro Gln Gln Pro Asn Cys Leu Pro Leu Asp Gln
 115 120 125

Val Thr Leu Pro Gln Lys Leu Gln Glu Ala Gly Tyr Ser Thr His Met
 130 135 140

Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys Leu Pro Thr
 145 150 155 160

Arg Arg Gly Phe Asp Thr Phe Leu Gly Ser Leu Thr Gly Asn Val Asp
 165 170 175

Tyr Tyr Thr Tyr Asp Asn Cys Asp Gly Pro Gly Val Cys Gly Phe Asp
 180 185 190

Leu His Glu Gly Glu Asn Val Ala Trp Gly Leu Ser Gly Gln Tyr Ser
 195 200 205

Thr Met Leu Tyr Ala Gln Arg Ala Ser His Ile Leu Ala Ser His Ser
 210 215 220

Pro Gln Arg Pro Leu Phe Leu Tyr Val Ala Phe Gln Ala Val His Thr
 225 230 235 240

Pro Leu Gln Ser Pro Arg Glu Tyr Leu Tyr Arg Tyr Arg Thr Met Gly
 245 250 255

Asn Val Ala Arg Arg Lys Tyr Ala Ala Met Val Thr Cys Met Asp Glu
 260 265 270

Ala Val Arg Asn Ile Thr Trp Ala Leu Lys Arg Tyr Gly Phe Tyr Asn
 275 280 285

[0166]

Asn Ser Val Ile Ile Phe Ser Ser Asp Asn Gly Gly Gln Thr Phe Ser
 290 295 300

Gly Gly Ser Asn Trp Pro Leu Arg Gly Arg Lys Gly Thr Tyr Trp Glu
 305 310 315 320

Gly Gly Val Arg Gly Leu Gly Phe Val His Ser Pro Leu Leu Lys Arg
 325 330 335

Lys Gln Arg Thr Ser Arg Ala Leu Met His Ile Thr Asp Trp Tyr Pro
 340 345 350

Thr Leu Val Gly Leu Ala Gly Gly Thr Thr Ser Ala Ala Asp Gly Leu
 355 360 365

Asp Gly Tyr Asp Val Trp Pro Ala Ile Ser Glu Gly Arg Ala Ser Pro
 370 375 380

Arg Thr Glu Ile Leu His Asn Ile Asp Pro Leu Tyr Asn His Ala Gln
 385 390 395 400

His Gly Ser Leu Glu Gly Gly Phe Gly Ile Trp Asn Thr Ala Val Gln
 405 410 415

Ala Ala Ile Arg Val Gly Glu Trp Lys Leu Leu Thr Gly Asp Pro Gly
 420 425 430

Tyr Gly Asp Trp Ile Pro Pro Gln Thr Leu Ala Thr Phe Pro Gly Ser
 435 440 445

Trp Trp Asn Leu Glu Arg Met Ala Ser Val Arg Gln Ala Val Trp Leu
 450 455 460

[0167]

Phe Asn Ile Ser Ala Asp Pro Tyr Glu Arg Glu Asp Leu Ala Gly Gln
465 470 475 480

Arg Pro Asp Val Val Arg Thr Leu Leu Ala Arg Leu Ala Glu Tyr Asn
485 490 495

Arg Thr Ala Ile Pro Val Arg Tyr Pro Ala Glu Asn Pro Arg Ala His
500 505 510

Pro Asp Phe Asn Gly Gly Ala Trp Gly Pro Trp Ala Ser Asp Glu Glu
515 520 525

Glu Glu Glu Glu Glu Gly Arg Ala Arg Ser Phe Ser Arg Gly Arg Arg
530 535 540

Lys Lys Lys Cys Lys Ile Cys Lys Leu Arg Ser Phe Phe Arg Lys Leu
545 550 555 560

Asn Thr Arg Leu Met Ser Gln Arg Ile
565

<210> 94
<211> 2067
<212> DNA
<213> 智人

<220>
<221> CDS
<222> (1)..(2064)
<223> hSULF6

<400> 94

atg cta att tca gga aga gaa gag aac caa ata gac ata tcc aag acc
Met Leu Ile Ser Gly Arg Glu Glu Asn Gln Ile Asp Ile Ser Lys Thr
1 5 10 15

48

[0168]

| | |
|---|-----|
| aca gag gta gat tgt ttt gtg gtt gaa tta gga agt cta cac aat cct | 96 |
| Thr Glu Val Asp Cys Phe Val Val Glu Leu Gly Ser Leu His Asn Pro | |
| 20 25 30 | |
| aca cgg aac cca cag cga att ttc acc aag cac gtg gcc acc aag tca | 144 |
| Thr Arg Asn Pro Gln Arg Ile Phe Thr Lys His Val Ala Thr Lys Ser | |
| 35 40 45 | |
| tcc agc tcc aaa tgt cag ctg gac caa ggt gga aaa agc ctg gtc cag | 192 |
| Ser Ser Ser Lys Cys Gln Leu Asp Gln Gly Gly Lys Ser Leu Val Gln | |
| 50 55 60 | |
| tgc att tta ccc aga tct tca aag ctc ctc tca ccc ttg tgt ctc ccc | 240 |
| Cys Ile Leu Pro Arg Ser Ser Lys Leu Leu Ser Pro Leu Cys Leu Pro | |
| 65 70 75 80 | |
| cat ccg tgt gga gct tta ctt ctg tat aga tcc tca gga atc gcc tct | 288 |
| His Pro Cys Gly Ala Leu Leu Leu Tyr Arg Ser Ser Gly Ile Ala Ser | |
| 85 90 95 | |
| gct ctt gct gcc ttt aca gac tcc ctc tct agg agc tgc tgg ctg tca | 336 |
| Ala Leu Ala Ala Phe Thr Asp Ser Leu Ser Arg Ser Cys Trp Leu Ser | |
| 100 105 110 | |
| gtg tcc ctg tgc tgt ttg ttt tgc ggt gtt gat ggc aca ttt atg aca | 384 |
| Val Ser Leu Cys Cys Leu Phe Cys Gly Val Asp Gly Thr Phe Met Thr | |
| 115 120 125 | |
| aga aac gcc aga ccc aac att gtc ctg ctg atg gca gat gac ctt gga | 432 |
| Arg Asn Ala Arg Pro Asn Ile Val Leu Leu Met Ala Asp Asp Leu Gly | |
| 130 135 140 | |
| gtg ggg gat ttg tgc tgc tac ggt aat aac tca gtg agc aca cct aat | 480 |
| Val Gly Asp Leu Cys Cys Tyr Gly Asn Asn Ser Val Ser Thr Pro Asn | |
| 145 150 155 160 | |
| att gac cgc ctg gca agt gaa gga gtg agg ctt acc cag cat ctc gca | 528 |
| Ile Asp Arg Leu Ala Ser Glu Gly Val Arg Leu Thr Gln His Leu Ala | |
| 165 170 175 | |
| gct gct tcc atg tgc acc cca agt cgg gct gcc ttc ctg acc ggc cgg | 576 |
| Ala Ala Ser Met Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly Arg | |
| 180 185 190 | |

[0169]

| | |
|---|------|
| tac ccc atc aga tca ggg atg gtg tct gcc tac aac ctg aac cgt gcc Tyr Pro Ile Arg Ser Gly Met Val Ser Ala Tyr Asn Leu Asn Arg Ala 195 200 205 | 624 |
| ttc acg tgg ctt ggt ggg tca ggt ggt ctt ccc acc aat gaa acg act Phe Thr Trp Leu Gly Gly Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr 210 215 220 | 672 |
| ttt gcc aag ctg ctg cag cac cgt ggc tac cgc acg gga ctc ata ggc Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly 225 230 235 240 | 720 |
| aaa tgg cac ctg ggt ttg agc tgc gcc tct cgg aat gat cac tgt tac Lys Trp His Leu Gly Leu Ser Cys Ala Ser Arg Asn Asp His Cys Tyr 245 250 255 | 768 |
| cac ccg ctc aac cat ggt ttt cac tac ttt tac ggg gtg cct ttt gga His Pro Leu Asn His Gly Phe His Tyr Phe Tyr Gly Val Pro Phe Gly 260 265 270 | 816 |
| ctt tta agc gac tgc cag gca tcc aag aca cca gaa ctg cac cgc tgg Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp 275 280 285 | 864 |
| ctc agg atc aaa ctg tgg atc tcc acg gta gcc ctt gcc ctg gtt cct Leu Arg Ile Lys Leu Trp Ile Ser Thr Val Ala Leu Ala Leu Val Pro 290 295 300 | 912 |
| ttt ctg ctt ctc att ccc aag ttc gcc cgc tgg ttc tca gtg cca tgg Phe Leu Leu Leu Ile Pro Lys Phe Ala Arg Trp Phe Ser Val Pro Trp 305 310 315 320 | 960 |
| aag gtc atc ttt gtc ttt gct ctc ctc gcc ttt ctg ttt ttc act tcc Lys Val Ile Phe Val Phe Ala Leu Leu Ala Phe Leu Phe Phe Thr Ser 325 330 335 | 1008 |
| tgg tac tct agt tat gga ttt act cga cgt tgg aat tgc atc ctt atg Trp Tyr Ser Ser Tyr Gly Phe Thr Arg Arg Trp Asn Cys Ile Leu Met 340 345 350 | 1056 |
| agg aac cat gaa att atc cag cag cca atg aaa gag gag aaa gta gct Arg Asn His Glu Ile Ile Gln Gln Pro Met Lys Glu Glu Lys Val Ala 355 360 365 | 1104 |

[0170]

| | |
|---|------|
| tcc ctc atg ctg aag gag gca ctt gct ttc att gaa agg tac aaa agg | 1152 |
| Ser Leu Met Leu Lys Glu Ala Leu Ala Phe Ile Glu Arg Tyr Lys Arg | |
| 370 375 380 | |
| | |
| gaa cct ttt ctc ctc ttt ttt tcc ttc ctg cac gta cat act cca ctc | 1200 |
| Glu Pro Phe Leu Leu Phe Phe Ser Phe Leu His Val His Thr Pro Leu | |
| 385 390 395 400 | |
| | |
| atc tcc aaa aag aag ttt gtt ggg cgc agt aaa tat ggc agg tat ggg | 1248 |
| Ile Ser Lys Lys Lys Phe Val Gly Arg Ser Lys Tyr Gly Arg Tyr Gly | |
| 405 410 415 | |
| | |
| gac aat gta gaa gaa atg gat tgg atg gtg ggt aaa atc ctg gat gcc | 1296 |
| Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Lys Ile Leu Asp Ala | |
| 420 425 430 | |
| | |
| ctg gac cag gag cgc ctg gcc aac cac acc ttg gtg tac ttc acc tct | 1344 |
| Leu Asp Gln Glu Arg Leu Ala Asn His Thr Leu Val Tyr Phe Thr Ser | |
| 435 440 445 | |
| | |
| gac aac ggg ggc cac ctg gag ccc ctg gac ggg gct gtt cag ctg ggt | 1392 |
| Asp Asn Gly Gly His Leu Glu Pro Leu Asp Gly Ala Val Gln Leu Gly | |
| 450 455 460 | |
| | |
| ggc tgg aac ggg atc tac aaa ggt ggc aaa gga atg gga gga tgg gaa | 1440 |
| Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu | |
| 465 470 475 480 | |
| | |
| gga ggt atc cgt gtg cca ggg ata ttc cgg tgg ccg tca gtc ttg gag | 1488 |
| Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Ser Val Leu Glu | |
| 485 490 495 | |
| | |
| gct ggg aga gtg atc aat gag ccc acc agc tta atg gac atc tat ccg | 1536 |
| Ala Gly Arg Val Ile Asn Glu Pro Thr Ser Leu Met Asp Ile Tyr Pro | |
| 500 505 510 | |
| | |
| acg ctg tct tat ata ggc gga ggg atc ttg tcc cag gac aga gtg att | 1584 |
| Thr Leu Ser Tyr Ile Gly Gly Gly Ile Leu Ser Gln Asp Arg Val Ile | |
| 515 520 525 | |
| | |
| gac ggc cag aac cta atg ccc ctg ctg gaa gga agg gcg tcc cac tcc | 1632 |
| Asp Gly Gln Asn Leu Met Pro Leu Leu Glu Gly Arg Ala Ser His Ser | |
| 530 535 540 | |

[0171]

<400> 95

Met Leu Ile Ser Gly Arg Glu Glu Asn Gln Ile Asp Ile Ser Lys Thr
 1 5 10 15

Thr Glu Val Asp Cys Phe Val Val Glu Leu Gly Ser Leu His Asn Pro
 20 25 30

Thr Arg Asn Pro Gln Arg Ile Phe Thr Lys His Val Ala Thr Lys Ser
 35 40 45

Ser Ser Ser Lys Cys Gln Leu Asp Gln Gly Gly Lys Ser Leu Val Gln
 50 55 60

Cys Ile Leu Pro Arg Ser Ser Lys Leu Leu Ser Pro Leu Cys Leu Pro
 65 70 75 80

His Pro Cys Gly Ala Leu Leu Leu Tyr Arg Ser Ser Gly Ile Ala Ser
 85 90 95

Ala Leu Ala Ala Phe Thr Asp Ser Leu Ser Arg Ser Cys Trp Leu Ser
 100 105 110

Val Ser Leu Cys Cys Leu Phe Cys Gly Val Asp Gly Thr Phe Met Thr
 115 120 125

Arg Asn Ala Arg Pro Asn Ile Val Leu Leu Met Ala Asp Asp Leu Gly
 130 135 140

Val Gly Asp Leu Cys Cys Tyr Gly Asn Asn Ser Val Ser Thr Pro Asn
 145 150 155 160

Ile Asp Arg Leu Ala Ser Glu Gly Val Arg Leu Thr Gln His Leu Ala
 165 170 175

[0173]

Ala Ala Ser Met Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly Arg
 180 185 190

Tyr Pro Ile Arg Ser Gly Met Val Ser Ala Tyr Asn Leu Asn Arg Ala
 195 200 205

Phe Thr Trp Leu Gly Gly Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr
 210 215 220

Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly
 225 230 235 240

Lys Trp His Leu Gly Leu Ser Cys Ala Ser Arg Asn Asp His Cys Tyr
 245 250 255

His Pro Leu Asn His Gly Phe His Tyr Phe Tyr Gly Val Pro Phe Gly
 260 265 270

Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp
 275 280 285

Leu Arg Ile Lys Leu Trp Ile Ser Thr Val Ala Leu Ala Leu Val Pro
 290 295 300

Phe Leu Leu Leu Ile Pro Lys Phe Ala Arg Trp Phe Ser Val Pro Trp
 305 310 315 320

Lys Val Ile Phe Val Phe Ala Leu Leu Ala Phe Leu Phe Phe Thr Ser
 325 330 335

Trp Tyr Ser Ser Tyr Gly Phe Thr Arg Arg Trp Asn Cys Ile Leu Met
 340 345 350

[0174]

Arg Asn His Glu Ile Ile Gln Gln Pro Met Lys Glu Glu Lys Val Ala
 355 360 365

Ser Leu Met Leu Lys Glu Ala Leu Ala Phe Ile Glu Arg Tyr Lys Arg
 370 375 380

Glu Pro Phe Leu Leu Phe Phe Ser Phe Leu His Val His Thr Pro Leu
 385 390 395 400

Ile Ser Lys Lys Lys Phe Val Gly Arg Ser Lys Tyr Gly Arg Tyr Gly
 405 410 415

Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Lys Ile Leu Asp Ala
 420 425 430

Leu Asp Gln Glu Arg Leu Ala Asn His Thr Leu Val Tyr Phe Thr Ser
 435 440 445

Asp Asn Gly Gly His Leu Glu Pro Leu Asp Gly Ala Val Gln Leu Gly
 450 455 460

Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu
 465 470 475 480

Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Ser Val Leu Glu
 485 490 495

Ala Gly Arg Val Ile Asn Glu Pro Thr Ser Leu Met Asp Ile Tyr Pro
 500 505 510

Thr Leu Ser Tyr Ile Gly Gly Gly Ile Leu Ser Gln Asp Arg Val Ile
 515 520 525

[0175]

Asp Gly Gln Asn Leu Met Pro Leu Leu Glu Gly Arg Ala Ser His Ser
530 535 540

Asp His Glu Phe Leu Phe His Tyr Cys Gly Val Tyr Leu His Thr Val
545 550 555 560

Arg Trp His Gln Lys Asp Cys Ala Thr Val Trp Lys Ala His Tyr Val
565 570 575

Thr Pro Lys Phe Tyr Pro Glu Gly Thr Gly Ala Cys Tyr Gly Ser Gly
580 585 590

Ile Cys Ser Cys Ser Gly Asp Val Thr Tyr His Asp Pro Pro Leu Leu
595 600 605

Phe Asp Ile Ser Arg Asp Pro Ser Glu Ala Leu Pro Leu Asn Pro Asp
610 615 620

Asn Glu Pro Leu Phe Asp Ser Val Ile Lys Lys Met Glu Ala Ala Ile
625 630 635 640

Arg Glu His Arg Arg Thr Leu Thr Pro Val Pro Gln Gln Phe Ser Val
645 650 655

Phe Asn Thr Ile Trp Lys Pro Trp Leu Gln Pro Cys Cys Gly Thr Phe
660 665 670

Pro Phe Cys Gly Cys Asp Lys Glu Asp Asp Ile Leu Pro Met Ala Pro
675 680 685

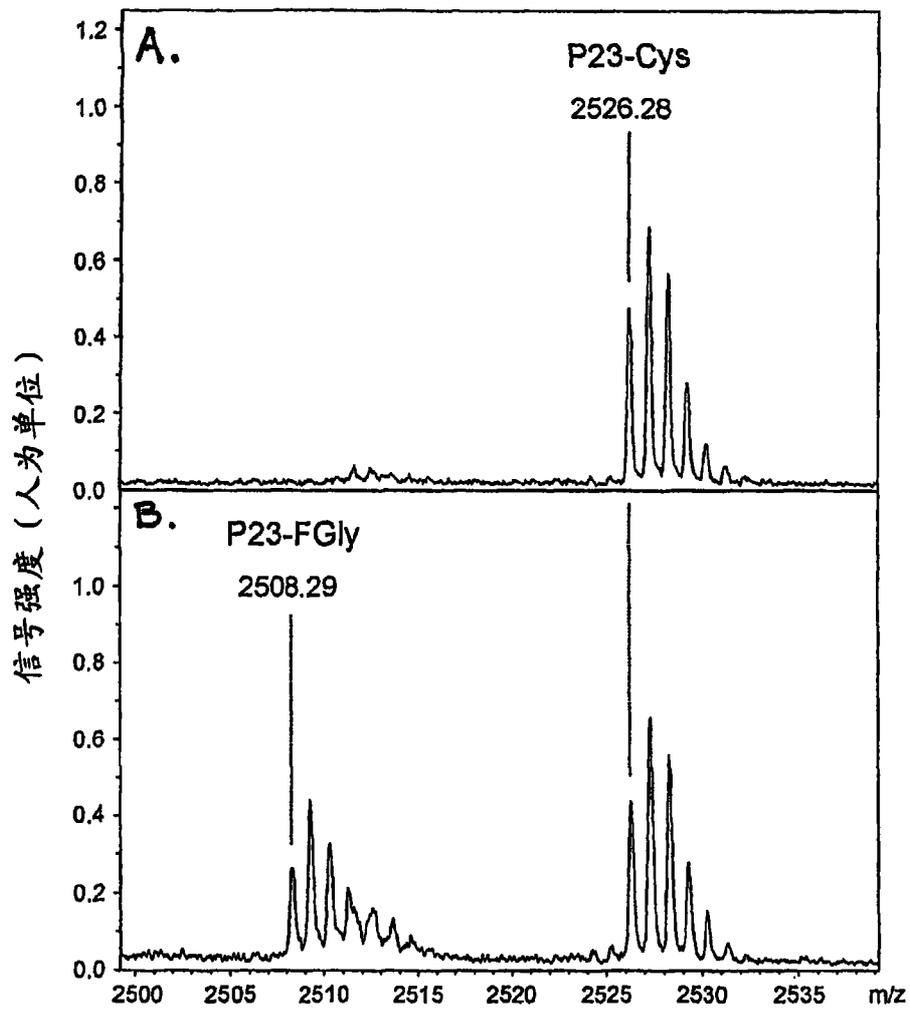


图 1

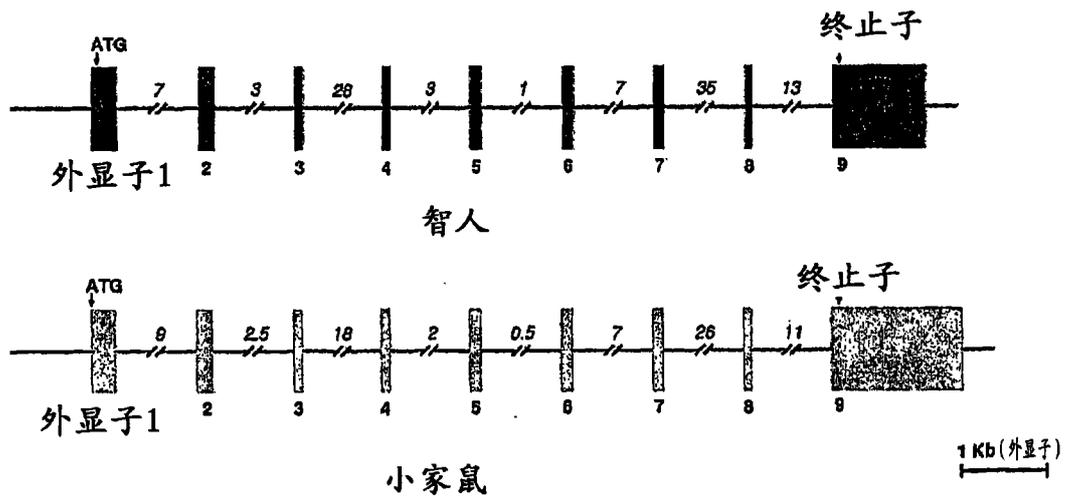


图 3

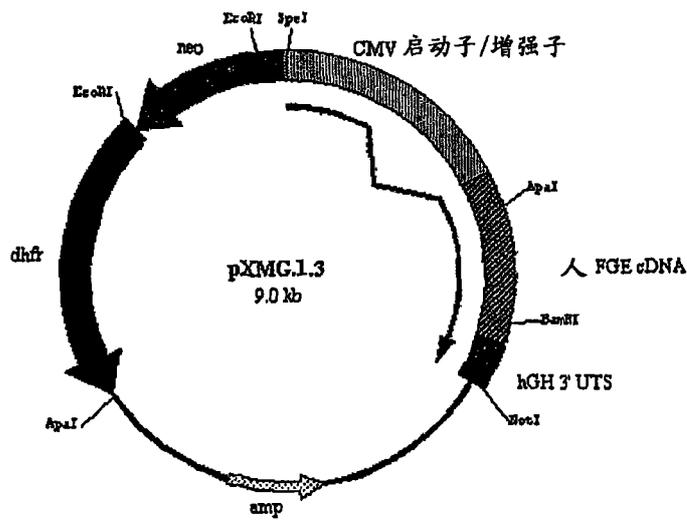


图 4

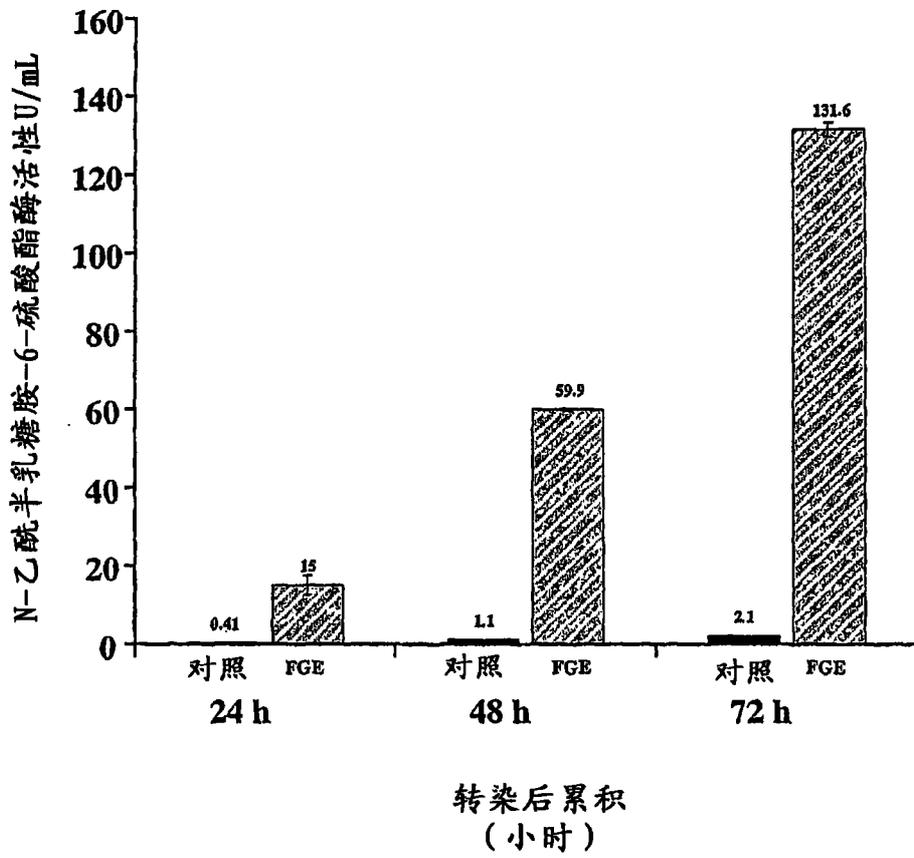


图 5

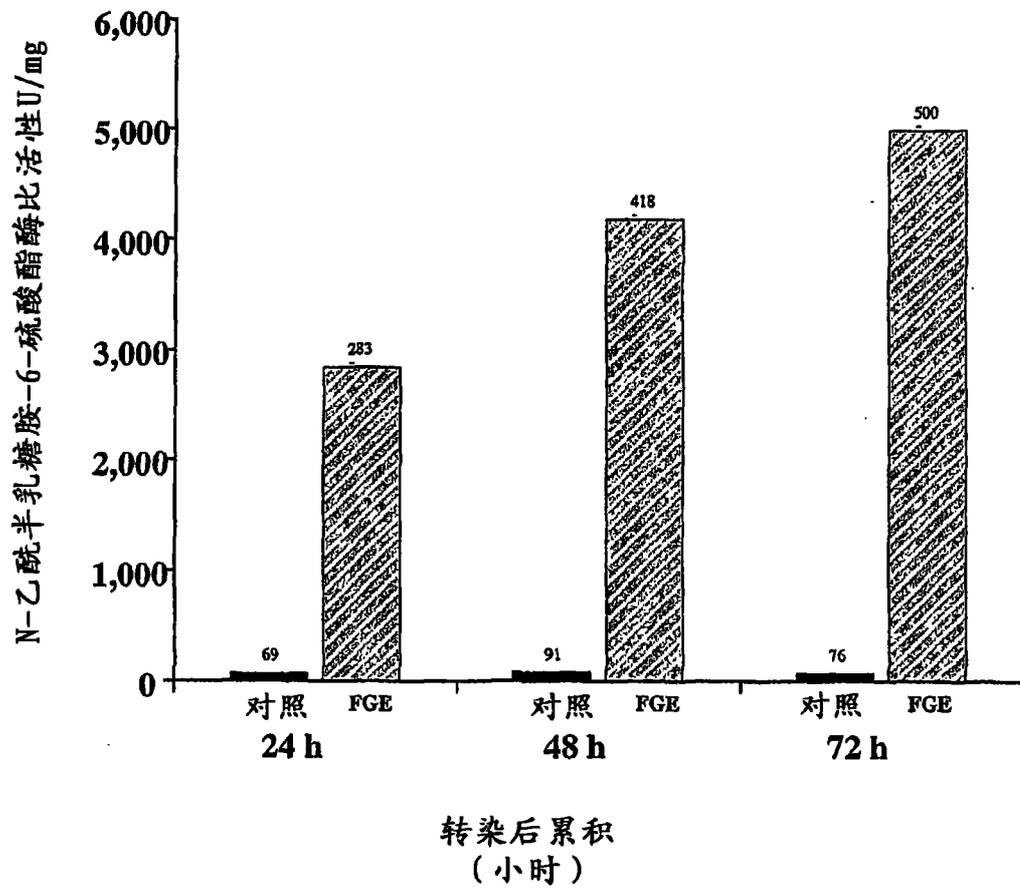


图 6

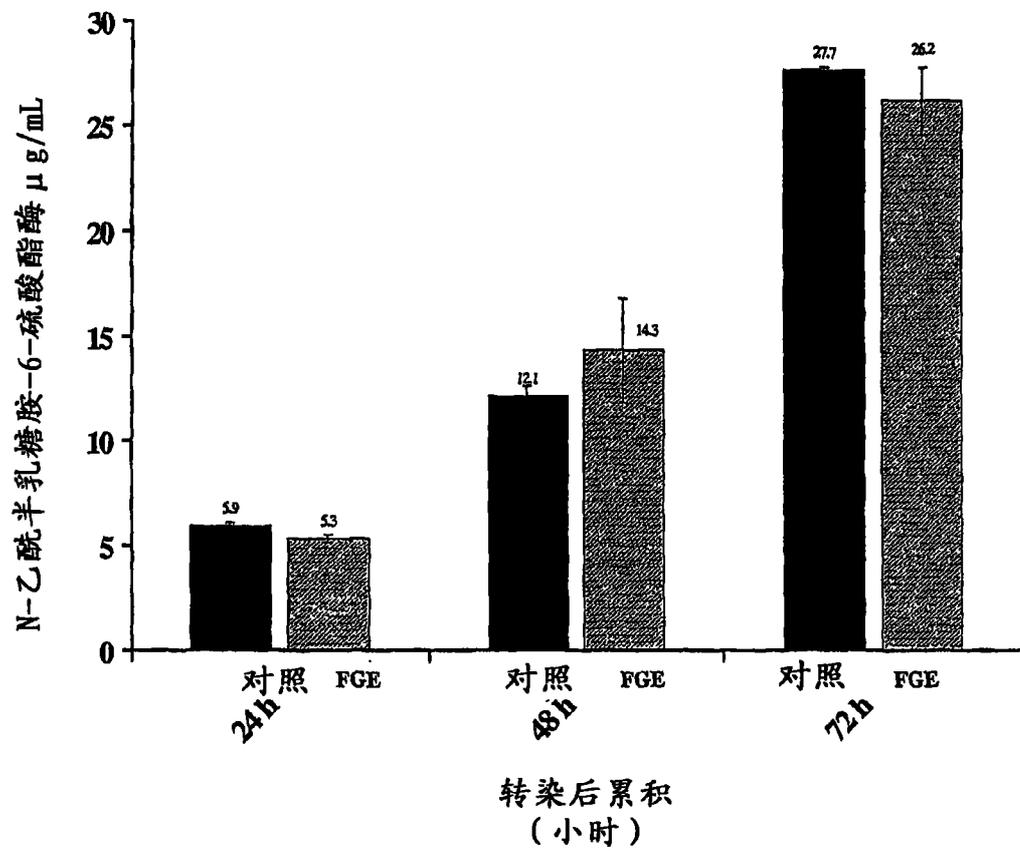


图 7

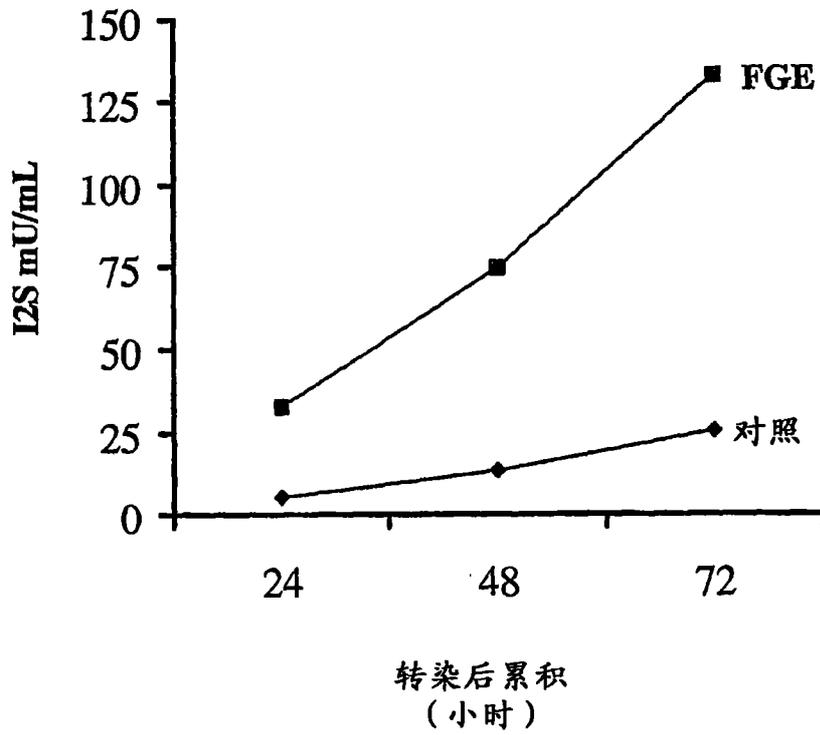


图 8

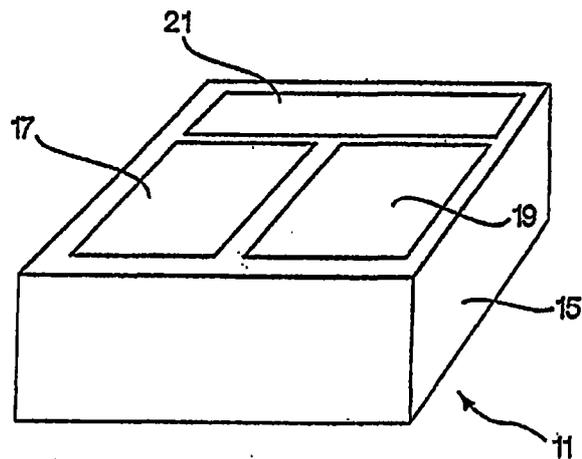


图 9