



(10) 申请公布号 CN 103541018 A

(21) 申请号 201310472612.6

C40B 40/10 (2006.01)

(22) 申请日 2007.09.28

G06F 19/10 (2011.01)

(30) 优先权数据

60/849,035 2006. 10. 02 US

(62) 分案原申请数据

200780044613.3 2007.09.28

(71) 申请人 航道生物技术有限责任公司

地址 美国加利福尼亚州

(72)发明人 劳伦斯.霍罗威茨 拉梅什.R.巴特

阿伦·L·库尔茨曼

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 岑晓东

(51) Int. Cl.

C40B 50/06 (2006.01)

权利要求书1页 说明书25页

序列表84页 附图40页

(54) 发明名称

多样性合成肽和多肽文库的设计和构建

(57) 摘要

本发明涉及多样性合成肽和多肽文库的设计和构建。本发明涉及多样性肽和多肽文库的设计和构建。具体而言,本发明涉及使用多重相关参数作为过滤器创建数据集的分析数据库设计方法;还涉及通过定向多重合成寡核苷酸合成(directed multisyntheses oligonucleotide synthesis)来生成序列多样性的方法。本发明使得人们能够基于可个别直接限定的单一或多重相关关键参数,将大型、复杂的附注释的数据库简化成更简单的相关序列数据集。所述方法还使得人们能够基于这种方法,使用离散和简并寡核苷酸的多重合成群集来捕获多种多样的序列群集或其部分,来生成多样性文库。

通用文档库步骤

使用下列哪种软件将数据拟合到 Kahan-CRDR 模型	• 最小二乘法拟合 - 使用不同的拟合的算法将数据拟合到各个拟合变量上 - 可以使用符号拟合 - 可以包括多种线性组合中不存在的矩阵项。 - 有误差而没有拟合项 • 连续数据集中最频繁使用的软件用于拟合 - V_{10} V_{10} - V_{11} V_{11} - V_{12} V_{12} • 将多种线性拟合方程模型上 - 有缺陷的拟合函数可以忽略 - 非线性模型，添加非线性项，而原形拟合。 - 非线性，自然免除了非线性
• 量值 - 标准误差，拟合项 - CRDR 大小范围， CRDR 1-6 和 CRDR 2-13 - CRDR 3 主要通过 例如 V_{10} 时 V_{10} • 拟合 - 全部 CRDR 大小范围 例如 选择拟合 CRDR 1-7, CRDR 2-10 且 CRDR 3 个最重要的拟合方程 - 拟合数据 例如 K1 或 K3 • 使用软件—一项空白表格可以生成 CRDR 多样性拟合 - 为单独拟合表格与使用其设定下限(lower bound) - 为使用拟合的变量和表格使用其设定下限 - 可以创建单一或多个 CRDR，使用非连续拟合参数的多个变量来降低下限	

1. 一种产生肽或多肽序列的组合文库的方法,包括使用组合式寡核苷酸合成来将氨基酸侧链化学多样性引入所述肽或多肽序列的两个或更多个氨基酸位置。

2. 权利要求 1 的方法,其中所述氨基酸侧链化学多样性被设计为模拟在所述肽或多肽序列中天然存在的多样性。

3. 权利要求 1 或权利要求 2 的方法,其中所述文库是抗体文库。

4. 权利要求 3 的方法,其中所述抗体文库包含抗体重链可变域序列。

5. 权利要求 3 的方法,其中所述文库包含抗体轻链可变域序列。

6. 权利要求 3 的方法,其中所述文库是组合单链可变片段(scFv)文库。

7. 权利要求 3 的方法,其中所述抗体文库是 Fab、Fab' 或 F(ab')₂ 片段的文库。

8. 一种用于数据库多样性分析的方法,所述数据库包含以至少一个共享序列基序为特征的相关氨基酸序列,所述方法包括如下步骤:

(a) 对所述相关氨基酸序列进行比对;

(b) 通过对所述包含共享序列基序的相关氨基酸序列应用预定的两个或更多个过滤器的组合来创建第一数据集;

(c) 分析所述第一数据集在所述共享序列基序内的位置氨基酸使用频率;和

(d) 创建以在所述共享序列基序内的一个或多个氨基酸位置处的最小阈值氨基酸使用频率为特征的第二数据集。

多样性合成肽和多肽文库的设计和构建

[0001] 本申请是申请日为 2007 年 9 月 28 日、中国申请号为 200780044613.3、发明名称为“多样性合成肽和多肽文库的设计和构建”的发明申请的分案申请。

技术领域

[0002] 本发明涉及多样性肽和多肽文库的设计和构建。具体而言,本发明涉及使用多重相关参数作为过滤器的用于创建数据集的分析数据库设计方法;还涉及通过定向多重合成寡核苷酸合成(directed multisyntheses oligonucleotide synthesis)来生成序列多样性的方法。本发明使得人们能够基于可个别直接限定的单一或多重相关关键参数,将大型、复杂的附注释的数据库简化成更简单的相关序列数据集。所述方法还使得人们能够基于这种方法,使用离散和简并寡核苷酸的多重合成群集(multisynthetic collection)来捕获多种多样的序列群集或其部分,来生成多样性文库。

[0003] 发明背景

[0004] 基于肽或多肽的药物候选物的开发通常起始于筛选相关肽或多肽序列的文库。因此筛选治疗性抗体候选物的第一步通常是创建高多样性的抗体序列文库。

[0005] 有数种用于设计和构建多样性抗体文库的方法是本领域已知的。

[0006] 已经有人描述了丝状噬菌体-肽组合抗体文库可以通过改组重链和轻链基因来增加多样性(Kang 等, Proc. Natl. Acad. Sci. USA, 88:11120-11123, (1991)), 或通过易错聚合酶链式反应(PCR)向文库引入随机突变来增加多样性(Gram 等, Proc. Natl. Acad. Sci. USA, 89:3576-3580, (1992))。使用确定的框架作为生成抗体文库的基础已经由 Barbas 等, Proc. Natl. Acad. Sci. USA 89:4457-4461(1992)(随机化 CD3-H3); Barbas 等, Gene 137:57-62(2003)(对 V_k CDR3 的延伸随机化); 和 Hayanashi 等, Biotechniques 17:310(1994)(通过重叠序列延伸和 PCR 对抗体 CDR 区进行同时诱变)描述。其它文献报导了 CDR-H3 文库与单一 V_L 基因的组合(Nissim 等, EMBO J. 13:692-698(1994)), CDR-H3 文库与 V_L 基因的限定群集的组合(De Kruif 等, J. Mol. Biol. 248:97-105(1995))或 CDR-H3 文库与 V_L 基因的随机化库(randomized repertoire)的组合(Griffiths 等, EMBO J. 13:3245-3260(1994))。

[0007] 另见美国专利 Nos. 5667988; 6096551; 7067284, 它们描述了使用通用或随机化的免疫球蛋白轻链来产生抗体文库的方法。

[0008] Knappik 等, J. Mol. Biol. 296:57-86(2000)描述了一种用于设计和构建人抗体文库的不同构思,称为 HuCAL(Human Combinatorial Antibody Libraries; 人类组合抗体文库)。这种方法基于这样的发现,在免疫应答过程中被频繁使用的人 V_H 和 V_L 亚家族各自分别由一种共有框架所代表,这样就有七个重链的 HuCAL 共有基因和七个轻链的 HuCAL 共有基因,它们产生 49 种可能的组合。所有基因均通过全合成制备,同时考虑密码子选择、促进蛋白质聚集的不利残基和所有 CDR 侧翼的独特及通用限制位点等因素。所述方法导致生成包含 CDR 的模块化抗体,这些 CDR 可按照需要转化成不同抗体形式。HuCAL 抗体文库的设计和合成在美国专利 6300064; 6696248; 6706484; 和 6828422 中有描述。

[0009] 尽管有这些以及其它的进展,还是非常需要用于设计和构建高多样性(多)肽(如抗体)文库的新的有效方法。

发明内容

[0010] 本发明涉及多样性肽和多肽文库的设计和构建。

[0011] 在一个方面,本发明涉及用于数据库多样性分析的方法,所述数据库包含相关的氨基酸序列,所述相关的氨基酸序列以至少一个共享的序列基序为特征,所述方法包括步骤:

[0012] (a) 对所述相关的氨基酸序列进行比对;

[0013] (b) 通过对包含共享序列基序的相关氨基酸序列应用预定的两个或更多个过滤器(filter)的组合来创建第一数据集;

[0014] (c) 分析第一数据集在共享序列基序内的位置氨基酸使用频率(positional amino acid usage frequency);和

[0015] (d) 创建以共享序列基序内的一个或多个氨基酸位置处的最小阈值氨基酸使用频率(minimum threshold amino acid usage frequency)为特征的第二数据集。

[0016] 在步骤(d)中,可以为共享序列基序内的任何和全部氨基酸位置指定最小阈值氨基酸使用频率。

[0017] 在一个具体实施方案中,为共享序列基序内的大多数氨基酸位置指定最小阈值氨基酸使用频率。在另一个具体实施方案中,为共享序列基序内的全部氨基酸位置指定最小阈值氨基酸使用频率。在多种实施方案中,为共享序列基序内的具体氨基酸位置指定的最小阈值氨基酸使用频率可以是相同或不同的。

[0018] 在一个进一步的实施方案中,最小阈值氨基酸使用频率被设定为给共享序列基序内的大部分氨基酸位置提供最小总和氨基酸使用率(minimum sum amino acid usage)。

[0019] 在一个更进一步的实施方案中,最小阈值氨基酸使用频率被设定为给所述共享序列基序内的所有氨基酸位置提供最小总和氨基酸使用率。

[0020] 可以将最小总和氨基酸使用率设定为任何期望的水平,并且在具体的实施方案中是至少约40%,或至少约45%,或至少约50%,或至少约55%,或至少约60%,或至少约65%,或至少约70%,或至少约75%,或至少约80%,或至少约85%,或至少约90%。

[0021] 在另一个实施方案中,所述相关的氨基酸序列是抗体序列。

[0022] 在另一个实施方案中,所述相关的氨基酸序列包含抗体重链序列。

[0023] 在一个进一步的实施方案中,所述相关的氨基酸序列包含抗体轻链序列。

[0024] 如果所述相关的氨基酸序列是抗体序列,那么共享序列基序可以是,例如,CDR序列,诸如CDR1、CDR2或CDR3序列。

[0025] 对可用于本发明方法的步骤(b)中的过滤器的性质或数目没有限制。在一种具体的实施方案中,在抗体序列的情况下,预定的过滤器组合可以选自下组:(1)抗体重链或轻链的同种型;(2)CDR1、CDR2和CDR3序列中一种或多种的长度;(3)在CDR1、CDR2和CDR3序列中的一种或多种内的一个或多个预定位置处一个或多个预定氨基酸残基的存在;(4)框架类型;(5)抗体结合的抗原;(6)抗体亲和力;和(7)CDR序列之外的位置氨基酸残基(positional amino acid residues)。

[0026] 在一个进一步的实施方案中,抗体重链和 / 或轻链 CDR1、CDR2 和 CDR3 序列中的至少一个的大小是匹配的。例如,这个参数可以作为附加的过滤器与抗体重链和 / 或轻链序列的同种型组合。

[0027] 在多种实施方案中,位置氨基酸使用频率是至少约 3%,或至少约 5%,或至少约 10%,或至少约 15%;或设定在约 3% 至约 15%,或约 5% 至约 10%。

[0028] 在本发明的方法的另一个实施方案中,CDR 序列内的每个氨基酸以相同的位置氨基酸使用频率为特征。在一个供选择的实施方案中,位置氨基酸使用频率在所述 CDR 序列内的至少两个氨基酸残基处不同。

[0029] 在另一个实施方案中,预定的过滤器组合包括框架类型。

[0030] 在另一个实施方案中,对抗体重链和轻链序列二者进行分析。任选地,将抗体重链序列与预定的抗体轻链特征配对,或将抗体轻链序列与预定的抗体重链特征配对。

[0031] 在一个进一步的实施方案中,相关的抗体序列来自至少一种功能性抗体。

[0032] 在一个更进一步的实施方案中,本发明方法步骤 (b) 中应用的过滤器中的至少一个是与功能性抗体重链和 / 或轻链框架序列最相似的种系序列。

[0033] 所述功能性抗体可以与例如,但不限于,选自下组的多肽结合:细胞表面和可溶性受体、细胞因子、生长因子、酶;蛋白酶;和激素。因此,所述抗体可以与细胞因子结合,所述细胞因子诸如白细胞介素,例如 IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-10、IL-11、IL-12、IL-15、IL-17、IL-18、IL-23,以及它们各自的家族成员。或者,所述细胞因子可以例如选自下组:干扰素 α 、 β 和 γ (IFN- α 、 $-\beta$ 和 $-\gamma$),肿瘤坏死因子 α 和 β (TNF- α 和 $-\beta$),TWEAK, RANKL, Blys, RANTES, MCP-1, MIP-1 α , MIP-1 β , SDF-1, 集落刺激因子 (CSF), 粒细胞集落刺激因子 (G-CSF), 和粒细胞巨噬细胞集落刺激因子 (GM-CSF)。

[0034] 抗体结合的多肽也可以是生长因子,包括但不限于,神经生长因子 (NGF)、胰岛素样生长因子 1 (IGF-1)、表皮生长因子 (EGF)、血小板衍生生长因子 (PDGF)、血管内皮生长因子 (VEGF)、胎盘生长因子 (PLGF)、组织生长因子 α (TGF- α) 和组织生长因子 β (TGF- β)。

[0035] 在另一个实施方案中,功能性抗体与半抗原结合,所述半抗原例如 Dig、Bio、DNP 或 FITC。

[0036] 在本发明的方法的另一个实施方案中,相关氨基酸序列源自分泌性蛋白或胞外蛋白家族的成员,例如其可以是细胞因子。

[0037] 在一个特定的实施方案中,细胞因子是干扰素 α , 并且相关氨基酸序列是 IFN- α 亚型的序列。

[0038] 在一个具体实施方案中,本发明进一步包括合成相关氨基酸序列的物理文库的步骤,所述文库的设计是在鉴定的数据集的辅助下进行的。

[0039] 在这种方法的某些实施方案中,所述文库的合成是通过生成离散的多个确定的或简并的寡核苷酸从而仅产生确定的氨基酸来实现的。

[0040] 在一个进一步的实施方案中,产生的物理文库的多样性超过作为被鉴定的数据集的物理表示 (physical presentation) 的文库的多样性。这可能是例如下述事实所致的结果,即至少一个不满足最小阈值氨基酸使用频率的氨基酸也被合成以提供所述多样性。

[0041] 在一个更进一步的实施方案中,产生的物理文库的多样性低于作为鉴定的数据集的物理表示的文库的多样性。这可能是例如下述事实所致的结果,即并非所有满足最小阈

值氨基酸使用频率的氨基酸都被合成。

[0042] 在另一个实施方案中,数据集包含抗体重链和 / 或轻链序列,其可以包括一个或多个 CDR。

[0043] 在另一个实施方案中,所述 CDR 被克隆到框架序列支架中,任选地,所述框架序列可以是包含所述 CDR 的数据库中使用最频繁的框架序列。

[0044] 可以使用任何表达系统来表达所述物理文库,包括所有原核和真核表达系统。

[0045] 在一个特定的实施方案中,使用噬菌粒展示、mRNA 展示、微生物细胞展示、哺乳动物细胞展示、微珠展示技术 (microbead display technique)、抗体阵列或基于蛋白质-DNA 关联的展示 (display based on protein-DNA linkage) 来表达并展示所述物理文库。

[0046] 在本发明的另一个实施方案中,对文库筛选其成员的一种或多种化学和 / 或生物学性质。这些性质可以包括,但不限于,半衰期、效价 (potency)、效力 (efficacy)、结合亲和力和免疫原性。

[0047] 在另一个实施方案中,将氨基酸侧链多样性引入文库成员的一个或多个氨基酸位置处。

[0048] 在一个具体的实施方案中,所述氨基酸侧链多样性通过在所述一个或多个氨基酸位置提供具有至少两种不同侧链化学官能性 (side chain chemical functionalities) 的氨基酸残基来引入。

[0049] 在其它实施方案中,在每个氨基酸位置处呈现全部氨基酸化学的至少 30%,或至少 50%,或至少 55%,或至少 60%。

[0050] 优选地,所述氨基酸侧链多样性是通过使用组合式简并寡核苷酸合成来引入的。

[0051] 在另一个方面,本发明涉及一种产生肽或多肽序列组合文库的方法,包括使用组合寡核苷酸合成来将氨基酸侧链化学多样性引入肽或多肽序列的两个或更多个氨基酸位置。

[0052] 在一个实施方案中,氨基酸侧链化学多样性被设计为模拟在所述肽或多肽序列中天然存在的多样性。

[0053] 所述文库可以是任何类型的文库,包括,但不限于抗体文库。

[0054] 在一个特定的实施方案中,抗体文库包含抗体重链可变域序列。

[0055] 在另一个实施方案中,文库包含抗体轻链可变域序列。

[0056] 在另一个实施方案中,文库是组合单链可变区片段 (scFv) 文库。

[0057] 在一个进一步的实施方案中,抗体文库是 Fab、Fab' 或 F(ab')₂ 片段的文库。

[0058] 本发明涉及下述各项。

[0059] 1. 一种用于数据库多样性分析的方法,所述数据库包含以至少一个共享序列基序为特征的相关氨基酸序列,所述方法包括如下步骤:

[0060] (a) 对所述相关氨基酸序列进行比对;

[0061] (b) 通过对所述包含共享序列基序的相关氨基酸序列应用预定的两个或更多个过滤器的组合来创建第一数据集;

[0062] (c) 分析所述第一数据集在所述共享序列基序内的位置氨基酸使用频率;和

[0063] (d) 创建以在所述共享序列基序内的一个或多个氨基酸位置处的最小阈值氨基酸使用频率为特征的第二数据集。

[0064] 2. 项 1 的方法,其中在步骤 (d) 中最小阈值氨基酸使用频率被指定给所述共享序列基序内的大部分氨基酸位置。

[0065] 3. 项 1 的方法,其中在步骤 (d) 中最小阈值氨基酸使用频率被指定给所述共享序列基序内的全部氨基酸位置。

[0066] 4. 项 2 或项 3 的方法,其中所有指定给所述氨基酸位置的阈值氨基酸使用频率都是相同的。

[0067] 5. 项 2 或项 3 的方法,其中并非所有指定给所述氨基酸位置的阈值氨基酸使用频率都是相同的。

[0068] 6. 项 1 的方法,其中所述最小阈值氨基酸使用频率被设定为对所述共享序列基序内的大部分氨基酸位置提供最小总和氨基酸使用率。

[0069] 7. 项 6 的方法,其中所述最小阈值氨基酸使用频率被设定为对所述共享序列基序内的全部氨基酸位置提供最小总和氨基酸使用率。

[0070] 8. 项 7 的方法,其中所述最小总和氨基酸使用率是至少约 60%。

[0071] 9. 项 7 的方法,其中所述最小总和氨基酸使用率是至少约 65%。

[0072] 10. 项 7 的方法,其中所述最小总和氨基酸使用率是至少约 70%。

[0073] 11. 项 7 的方法,其中所述最小总和氨基酸使用率是至少约 75%。

[0074] 12. 项 7 的方法,其中所述最小总和氨基酸使用率是至少约 80%。

[0075] 13. 项 7 的方法,其中所述最小总和氨基酸使用率是至少约 85%。

[0076] 14. 项 7 的方法,其中所述最小总和氨基酸使用率是至少约 90%。

[0077] 15. 项 1 的方法,其中所述相关氨基酸序列是抗体序列。

[0078] 16. 项 15 的方法,其中所述相关氨基酸序列包含抗体重链序列。

[0079] 17. 项 15 的方法,其中所述相关氨基酸序列包含抗体轻链序列。

[0080] 18. 项 16 或项 17 的方法,其中所述共享序列基序是 CDR 序列。

[0081] 19. 项 18 的方法,其中所述共享序列基序选自下组:CDR1、CDR2 和 CDR3 序列。

[0082] 20. 项 19 的方法,其中在步骤 (b) 中,所述预定的过滤器的组合选自下组:(1) 所述抗体重链或轻链的同种型;(2) 所述 CDR1、CDR2 和 CDR3 序列中一种或多种的长度;(3) 在所述 CDR1、CDR2 和 CDR3 序列中一种或多种内的一个或多个预定位置处一个或多个预定氨基酸残基的存在;(4) 框架类型;(5) 所述抗体结合的抗原;(6) 所述抗体的亲和力;和 (7) 所述 CDR 序列之外的位置氨基酸残基。

[0083] 21. 项 20 的方法,其中抗体重链和 / 或轻链 CDR1、CDR2 和 CDR3 序列中的至少一种是大小匹配的。

[0084] 22. 项 21 的方法,其中一种附加的过滤器是所述抗体重链和 / 或轻链序列的同种型。

[0085] 23. 项 19 的方法,其中所述位置氨基酸残基使用频率是至少约 3%。

[0086] 24. 项 19 的方法,其中所述位置氨基酸残基使用频率是至少约 5%。

[0087] 25. 项 19 的方法,其中所述位置氨基酸残基使用频率是至少约 10%。

[0088] 26. 项 19 的方法,其中所述位置氨基酸残基使用频率是至少约 15%。

[0089] 27. 项 19 的方法,其中所述位置氨基酸残基使用频率是约 3% 至约 15%。

[0090] 28. 项 19 的方法,其中所述位置氨基酸残基使用频率是约 5% 至约 10%。

[0091] 29. 项 18 的方法,其中相同的位置氨基酸使用频率是所述 CDR 序列内的每个氨基酸的特征。

[0092] 30. 项 18 的方法,其中所述位置氨基酸使用频率在所述 CDR 序列内的至少两个氨基酸残基处是不同的。

[0093] 31. 项 20 的方法,其中所述预定的过滤器的组合包括框架的类型。

[0094] 32. 项 15 的方法,其中对抗体重链和轻链序列二者进行分析。

[0095] 33. 项 32 的方法,其中将所述抗体重链序列与预定的抗体轻链特征配对。

[0096] 34. 项 32 的方法,其中将所述抗体轻链序列与预定的抗体重链特征配对。

[0097] 35. 项 15 的方法,其中所述相关抗体序列来自至少一种功能性抗体。

[0098] 36. 项 35 的方法,其中步骤 (b) 中应用的所述过滤器之一是与所述功能性抗体的重链和 / 或轻链的框架序列最相似的种系序列。

[0099] 37. 项 35 的方法,其中所述功能性抗体与多肽结合,所述多肽选自下组:细胞表面和可溶性受体,细胞因子,生长因子,酶;蛋白酶;和激素。

[0100] 38. 项 37 的方法,其中所述多肽是细胞因子。

[0101] 39. 项 38 的方法,其中所述细胞因子是白细胞介素。

[0102] 40. 项 39 的方法,其中所述白细胞介素选自下组:IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-10、IL-11、IL-12、IL-15、IL-17、IL-18、IL-23,和它们各自的家族成员。

[0103] 41. 项 38 的方法,其中所述细胞因子选自下组:干扰素 α 、 β 和 γ (IFN- α 、 $-\beta$ 和 $-\gamma$), 肿瘤坏死因子 α 和 β (TNF- α 和 $-\beta$), TWEAK, TANKL, BLys, RANTES, MCP-1, MIP-1 α , MIP-1 β , SDF-1, 集落刺激因子 (CSF), 粒细胞集落刺激因子 (G-CSF), 和粒细胞巨噬细胞集落刺激因子 (GM-CSF)。

[0104] 42. 项 38 的方法,其中所述多肽是生长因子。

[0105] 43. 项 42 的方法,其中所述生长因子选自下组:神经生长因子 (NGF)、胰岛素样生长因子 1 (IGF-1)、表皮生长因子 (EGF)、血小板衍生生长因子 (PDGF)、血管内皮生长因子 (VEGF)、胎盘生长因子 (PLGF)、组织生长因子 α (TGF- α) 和组织生长因子 β (TGF- β)。

[0106] 44. 项 35 的方法,其中所述功能性抗体与半抗原结合。

[0107] 45. 项 44 的方法,其中所述半抗原选自下组:Dig、Bio、DNP 和 FITC。

[0108] 46. 项 1 的方法,其中所述相关氨基酸序列源自分泌性蛋白或胞外蛋白家族的成员。

[0109] 47. 项 46 的方法,其中所述相关氨基酸序列源自细胞因子家族成员。

[0110] 48. 项 47 的方法,其中所述细胞因子是干扰素 α 。

[0111] 49. 项 48 的方法,其中所述相关氨基酸序列是 IFN- α 亚型的序列。

[0112] 50. 项 1 的方法,进一步包括合成相关氨基酸序列的物理文库的步骤,所述文库的设计在所鉴定的数据集的辅助下进行。

[0113] 51. 项 50 的方法,其中所述文库的合成是通过生成离散的多个确定的或简并的寡核苷酸,从而仅产生确定的氨基酸而进行的。

[0114] 52. 项 50 的方法,其中产生的物理文库的多样性超过作为所鉴定的数据集的物理表示的文库的多样性。

[0115] 53. 项 52 的方法,其中还合成至少一个不满足最小阈值氨基酸使用频率的氨基酸

以提供所述多样性。

[0116] 54. 项 50 的方法,其中产生的物理文库的多样性低于作为所鉴定的数据集的物理表示的文库的多样性。

[0117] 55. 项 54 的方法,其中并非所有满足最小阈值氨基酸使用频率的氨基酸都被合成。

[0118] 56. 项 50-55 中任一项的方法,其中所述数据集包含抗体重链和 / 或轻链序列。

[0119] 57. 项 56 的方法,其中所述抗体重链和 / 或轻链序列包含一个或多个 CDR。

[0120] 58. 项 57 的方法,其中将所述 CDR 克隆至框架序列支架中。

[0121] 59. 项 58 的方法,其中所述框架序列是在包含所述 CDR 的数据库中最频繁使用的框架序列。

[0122] 60. 项 50 的方法,其中所述物理文库是使用原核或真核表达系统表达的。

[0123] 61. 项 50 的方法,其中所述物理文库是使用噬菌粒展示、mRNA 展示、微生物细胞展示、哺乳动物细胞展示、微珠展示技术、抗体阵列或基于蛋白质-DNA 关联的展示来表达并展示的。

[0124] 62. 项 50 的方法,其中针对所述文库筛选其成员的一种或多种化学和 / 或生物学性质。

[0125] 63. 项 62 的方法,其中所述生物学性质选自下组:半衰期、效价、效力、结合亲和力和免疫原性。

[0126] 64. 项 50 的方法,包括在一个或多个氨基酸位置处引入氨基酸侧链多样性。

[0127] 65. 项 64 的方法,其中所述氨基酸侧链多样性通过在所述一个或多个氨基酸位置处提供具有至少两种不同侧链化学官能性的氨基酸残基来引入。

[0128] 66. 项 65 的方法,其中在每个氨基酸位置呈现全部氨基酸化学的至少 30%。

[0129] 67. 项 65 的方法,其中在每个氨基酸位置呈现全部氨基酸化学的至少 50%。

[0130] 68. 项 65 的方法,其中所述侧链多样性通过使用组合式简并寡核苷酸合成来引入。

[0131] 69. 一种产生肽或多肽序列的组合文库的方法,包括使用组合式寡核苷酸合成来将氨基酸侧链化学多样性引入所述肽或多肽序列的两个或更多个氨基酸位置。

[0132] 70. 项 69 的方法,其中所述氨基酸侧链化学多样性被设计为模拟在所述肽或多肽序列中天然存在的多样性。

[0133] 71. 项 69 或项 70 的方法,其中所述文库是抗体文库。

[0134] 72. 项 71 的方法,其中所述抗体文库包含抗体重链可变域序列。

[0135] 73. 项 71 的方法,其中所述文库包含抗体轻链可变域序列。

[0136] 74. 项 71 的方法,其中所述文库是组合单链可变片段(scFv)文库。

[0137] 75. 项 71 的方法,其中所述抗体文库是 Fab、Fab' 或 F(ab')₂ 片段的文库。

[0138] 附图简述

[0139] 图 1 是对多样性人抗体文库的设计和构建中代表性步骤的总结。

[0140] 图 2:V_k CDR1、2 和 3 的频率分析;按照位置测定的绝对使用率。

[0141] 图 3:V_k I 轻链阈值分析。不报告低于 10% 使用率的单独氨基酸。

[0142] 图 4:V_k I 轻链阈值分析。不报告低于 5% 使用率的单独氨基酸。

[0143] 图 5:合成轻链 CDR1 多样性。按照出现的顺序,图 5 分别公开了 SEQ ID NOS4 和 5-20。

[0144] 图 6:V_H3 重链合成文库阈值分析;长度 10 个残基。为每个氨基酸位置单独设定 3%-10% 的阈值百分比使用率。

[0145] 图 7:用于合成如图 6 中所示设计的文库的寡核苷酸(按照出现的顺序,分别为 SEQ ID NOS21-116)。

[0146] 图 8:确定有效的(productive)抗 TNF- α 抗体重链(按照出现的顺序,分别为 SEQ ID NOS117-139)的种系来源。

[0147] 图 9:说明有效的(productive)抗 TNF- α 抗体重链种系来源的树形比对图。

[0148] 图 10:确定有效的(productive)抗 TNF- α 抗体轻链(按照出现的顺序,分别为 SEQ ID NOS140-159)的种系来源。

[0149] 图 11:说明有效的(productive)抗 TNF- α 抗体轻链种系来源的树形比对图。

[0150] 图 12:V_K1 轻链合成文库多样性。图 12 公开了 SEQ ID NO:160。

[0151] 图 13:V_H3CDR1 和 CDR2 的频率分析。

[0152] 图 14:CDR1 和 CDR2 阈值分析——第一部分。

[0153] 图 15:CDR1 和 CDR2 阈值分析——第二部分。

[0154] 图 16:V_H3 重链合成文库多样性。图 16 公开了 SEQ ID NO:161。

[0155] 图 17:基于抗洋地黄毒苷抗体 D2E7 设计 V_H3 重链合成文库多样性。图 17 公开了 SEQ ID NO:162。

[0156] 图 18:抗洋地黄毒苷抗体 Ig λ 轻链可变区(SEQ ID NO:1)和重链可变区(SEQ ID NO:2)序列。

[0157] 图 19:确定抗洋地黄毒苷抗体重链和轻链的种系来源(按照出现的顺序,分别为 SEQ ID NOS2、125、1 和 163)。

[0158] 图 20:对 λ 长度匹配的 C_L1 框架进行的半抗原分析。

[0159] 图 21:对 H3- 长度 8 氨基酸进行的半抗原分析。

[0160] 图 22:IFN- α 亚型的氨基酸残基 32-38(按照出现的顺序,分别为 SEQ ID NOS164-168、167、167、169、167、170、167 和 171)的比对。

[0161] 图 23:编码期望的 IFN- α 多样性的寡核苷酸设计(按照出现的顺序,分别为 SEQ ID NOS172、172 和 173-174)。

[0162] 图 24:按照侧链化学分类的氨基酸。

[0163] 图 25:编码具有化学探针的(chemically probed)多样性位置。

[0164] 图 26:包含具有化学探针的多样性的 CDR3。

[0165] 图 27:使用化学探针集(按照出现的顺序,分别为 SEQ ID NOS175 和 176)编码 CDR3 重链多样性。

[0166] 发明详述

[0167] A. 定义

[0168] 除非另有定义,本文使用的科技术语与本发明所属技术领域普通技术人员一般理解的意思相同。Singleton 等,Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley&Sons(New York, NY1994)为本领域技术人员提供了关于本申请中使用的许多

术语的综合指导。

[0169] 本领域技术人员将意识到许多与本文描述的那些方法和材料类似或等效的方法和材料能够用于实施本发明。事实上,本发明决不限于描述的方法和材料。就本发明而言,在下文定义以下术语。

[0170] 短语“共享序列基序 (shared sequence motif)”按照最广的含义用于本文,用于表示在两个或更多个肽或多肽序列之间共有的氨基酸残基模式 (pattern of amino acid residues)。序列基序可以通过多种模式发现算法 (pattern discovery algorithm) 轻易地鉴定,例如在本发明的详述中讨论的那些算法。

[0171] 在本发明的上下文中,术语“抗体”(Ab) 按照最广的含义使用,包括对特定抗原展现结合特异性的免疫球蛋白以及缺乏抗原特异性的免疫球蛋白和其它抗体样分子。例如,后一种类型的多肽被淋巴系统以低水平产生,而被骨髓瘤以增加的水平产生。在本申请中,术语“抗体”具体涵盖,但不限于,单克隆抗体、多克隆抗体和抗体片段。

[0172] “天然抗体”通常是大约 150000 道尔顿的异四聚体糖蛋白,由两个相同的轻链 (L) 和两个相同的重链 (H) 构成。每条轻链通过共价二硫键与重链连接,而不同免疫球蛋白同种型的重链之间的二硫键数目不同。每条重链和每条轻链还具有规则间隔排列的链间二硫桥。每条重链在一端具有一个可变域 (V_H),其后为多个恒定域。每条轻链在一端具有一可变域 (V_L),在其另一端具有一恒定域;轻链的恒定域与重链的第一恒定域并列,轻链的可变域与重链的可变域并列。认为特定氨基酸残基在轻链和重链可变域之间形成一界面,Chothia 等, *J. Mol. Biol.* 186:651 (1985); Novotny 和 Haber, *Proc. Natl. Acad. Sci. U. S. A.* 82:4592 (1985)。

[0173] 就抗体链而言,术语“可变的”用于指抗体链中在抗体之间存在广泛序列差异的部分,并且该部分参与每种特定抗体与其特定抗原的结合和特异性。这种可变性集中在轻链和重链可变域中称为超变区的三个区段。可变域中更加高度保守的部分称为框架区 (FR)。天然重链和轻链的可变域各包含四个 FR (分别为 FR1、FR2、FR3 和 FR4),主要采用由三个超变区连接的 β -折叠构型,这三个超变区形成连接 β -折叠结构且在某些情况下形成 β -折叠结构的一部分的环。每条链中的超变区通过 FR 紧密结合,并且与来自另一条链的超变区一起对抗体的抗原结合部位的形成起贡献 (参见 Kabat 等, *Sequences of Proteins of Immunological Interest*, 5th Ed. Public Health Service, National Institutes of Health, Bethesda, Md. (1991), 第 647-669 页)。恒定域不直接参与抗体与抗原的结合,但显示多种效应器功能 (effector functions),诸如抗体对抗体依赖性细胞毒性的参与。

[0174] 术语“超变区”当用于本文时指抗体中负责抗原结合的氨基酸残基。超变区包含来自“互补决定区”或“CDR”的氨基酸残基 (即,轻链可变域中的残基 30-36 (L1)、46-55 (L2) 和 86-96 (L3) 和重链可变域中的 30-35 (H1)、47-58 (H2) 和 93-101 (H3); MacCallum 等, *J. Mol. Biol.* 1996)。

[0175] 术语“框架区”指抗体可变区中本领域公认的、存在于变异性更高的 CDR 区之间的部分。这些框架区通常被称为框架 1-4 (FR1、FR2、FR3 和 FR4),并且在三维空间中为重链和轻链抗体可变区中存在的三个 CDR 的结合提供支架,从而使这些 CDR 能够形成抗原结合表面。

[0176] 依赖于它们重链恒定域的氨基酸序列,抗体可分为不同种类。主要有五种抗体:

IgA、IgD、IgE、IgG 和 IgM,而这些中的几种可以进一步分成亚类(同种型),例如,IgG1、IgG2、IgG3、IgG4、IgA 和 IgA2。

[0177] 对应于不同种类免疫球蛋白的重链恒定域分别称为 α 、 δ 、 ϵ 、 γ 和 μ 。

[0178] 来自任何脊椎动物物种的抗体的“轻链”可以基于它们恒定域的氨基酸序列归于两种明显不同的类型之一,这两种类型称为 kappa(κ) 和 lambda(λ)。

[0179] “抗体片段”包含全长抗体的一部分,通常是抗体的抗原结合域或可变域。抗体片段的实例包括,但不限于,Fab、Fab'、F(ab')₂、Dab 和 Fv 片段、线性抗体、单链抗体分子、双抗体和从抗体片段形成的多特异性抗体。

[0180] 术语“单克隆抗体”用于指单个 B 细胞克隆所合成的抗体分子。修饰语“单克隆”表示抗体的这样的性质,即它是从基本上均一的(homogeneous)抗体群体获得的,而不应解释为要求通过任何特定方法来产生所述抗体。因此,单克隆抗体可以通过由 Kohler 和 Milstein, Nature 256:495(1975); Eur. J. Immunol. 6:511(1976) 首先描述的杂交瘤法制备,通过重组 DNA 技术制备,或者也可以从噬菌体抗体文库分离。

[0181] 术语“多克隆抗体”用于指由 B 细胞群体合成的抗体分子群体。

[0182] “单链 Fv”或“sFv”抗体片段包含抗体的 V_H 和 V_L 结构域,其中这些结构域存在于一个多肽单链中。概括而言,Fv 多肽进一步包含 V_H 和 V_L 结构域之间的多肽接头,其使得 sFv 能够形成用于抗原结合的理想结构。关于 sFv 的综述参见 Plückthun in The Pharmacology of Monoclonal Antibodies, 第 113 卷 Rosenberg 和 Moore 编, Springer-Verlag, New York, 第 269-315 页(1994)。单链抗体记载在例如 W088/06630 和 W092/01047 中。

[0183] 如用于本文,术语“抗体结合区”指能与抗原结合的免疫球蛋白或抗体可变区的一个或多个部分。通常而言,抗体结合区是例如抗体轻链(VL)(或其可变区),抗体重链(VH)(或其可变区),重链 Fd 区,抗体轻链和重链(或它们的可变区)的组合诸如 Fab、F(ab')₂, 单结构域,或单链抗体(scFv),或全长抗体,例如 IgG(例如 IgG1、IgG2、IgG3 或 IgG4 亚型)、IgA1、IgA2、IgD、IgE 或 IgM 抗体。

[0184] 术语“阈值出现频率(threshold frequency of occurrence)”指本发明的一个标准,其要求本文文库中选用的序列衍生自己被确定为受到表达偏爱的(favored to be expressed)序列。依赖于最终目的,诸如需要的多样性程度、期望的文库大小,“阈值出现频率”可以设定为不同水平。

[0185] 术语“氨基酸”或“氨基酸残基”通常指具有本领域公认的定义的氨基酸,诸如选自下组的氨基酸:丙氨酸(Ala);精氨酸(Arg);天冬酰胺(Asn);天冬氨酸(Asp);半胱氨酸(Cys);谷氨酰胺(Gln);谷氨酸(Glu);甘氨酸(Gly);组氨酸(His);异亮氨酸(Ile);亮氨酸(Leu);赖氨酸(Lys);甲硫氨酸(Met);苯丙氨酸(Phe);脯氨酸(Pro);丝氨酸(Ser);苏氨酸(Thr);色氨酸(Trp);酪氨酸(Tyr);和缬氨酸(Val);但如有需要也可以使用经修饰的、合成的或罕见的氨基酸。因此,将 37CFR1.822(b)(4) 中列出的经修饰的和不常用的氨基酸具体地包括在本定义中,并且明确地通过引用并入它们。可将氨基酸细分为多个亚组。因此,可将氨基酸分为:具有非极性侧链的组(例如,Ala、Cys、Ile、Leu、Met、Phe、Pro、Val);具有带负电的侧链的组(例如,Asp、Glu);具有带正电的侧链的组(例如,Arg、His、Lys);或具有不带电的极性侧链的组(例如,Asn、Cys、Gln、Gly、His、Met、Phe、Ser、Thr、Trp 和 Tyr)。氨基酸也可以分成小氨基酸组(Gly、Ala),亲核氨基酸组(Ser、His、Thr、Cys),

疏水氨基酸组 (Val、Leu、Ile、Met、Pro)，芳族氨基酸组 (Phe、Tyr、Trp、Asp、Glu)，酰胺组 (Asp、Glu)，和碱性氨基酸组 (Lys、Arg) (参见，图 25)。

[0186] 术语“保守氨基酸残基”指对于在两个或更多个被比较的氨基酸序列中的给定残基位置而言，确定为以高频率出现的某种氨基酸残基，所述频率通常为至少 50% 或更高 (例如，约 60%、70%、80%、90%、95% 或更高)。

[0187] 术语“半保守氨基酸残基”指被确定为对于给定残基位置而言，在所比较的两个或更多个氨基酸序列之间以高频率出现的多种氨基酸残基 (amino acid residues determined to occur with a high frequency between two or more amino acid sequences compared for a given residue position)。当 2-3 种残基 (特别是 2 种残基) 合起来的出现频率为约 40% 几率或更高的频率 (例如，50%、60%、70%、80%、90% 或更高) 时，将所述残基确定为半保守的。

[0188] 术语“可变氨基酸残基”指被确定为对于给定残基位置而言，在所比较的两个或更多个序列之间以可变的频率出现的多种氨基酸残基 (amino acid residues determined to occur with a variable frequency between two or more amino acid sequences compared for a given residue position)。当在给定位置处出现许多残基时，将该残基位置确定为可变的。

[0189] 术语“可变性谱 (variability profile)”指对多肽序列内 (例如抗体的 CDR 内) 特定氨基酸位置上存在的氨基酸和它们各自的出现频率所做的编录 (cataloguing)。

[0190] 术语“多核苷酸”指核酸如 DNA 分子和 RNA 分子和它们的类似物 (analog) (例如，使用核苷酸类似物或使用核酸化学产生的 DNA 或 RNA)。如果希望，所述多核苷酸可以以合成方法制备，例如，使用本领域公认的核酸化学或酶促方法 (使用例如聚合酶)，并且如果希望的话对其进行修饰。典型的修饰包括甲基化、生物素化和其它本领域已知的修饰。此外，核酸分子可以是单链或双链的，并且在希望的情况下与可检测的部分相连接。

[0191] 除非另有指定，术语“诱变”指任何本领域公认的用于改变多核苷酸或多肽序列的技术。优选的诱变类型包括易错 PCR 诱变、饱和诱变 (saturation mutagenesis) 或其它定点诱变。

[0192] 术语“载体”用于指这样的 rDNA 分子，其能够在细胞中自主复制，并且可以与 DNA 区段 (例如基因或多核苷酸) 可操作地连接，从而引起该附接区段的复制。能够指导编码一种或多种多肽的基因表达的载体在本文称为“表达载体”。

[0193] 术语“引物”，用于本文时，指这样的多核苷酸，其或是从核酸限制消化反应纯化的，或是合成产生的，当其处于与核酸链互补的引物延伸产物的合成受到诱导的条件下时，能起到核酸合成起始点的作用。这些条件可以包括在合适的温度和 pH 下核苷酸和 DNA 聚合酶、逆转录酶等的存在。所述引物优选是单链的，但是也可以是双链形式的。引物必须长到足以在用于聚合的作用剂的存在下引发衍生产物的合成。引物的确切长度将依赖于许多因素，包括靶序列的复杂度、温度和引物的来源。引物通常含有约 15 至约 25 个核苷酸，但是也可以使用更短或更长的引物。较短的引物通常需要较低的温度来与模板形成稳定的复合物。

[0194] “噬菌体展示文库”是一种蛋白质表达文库，其将克隆的蛋白质序列的群集表达为与噬菌体外壳蛋白的融合物。因此，短语“噬菌体展示文库”在本文指噬菌体 (例如，丝状

噬菌体)的群集,其中所述噬菌体表达外来的(通常为异源的)蛋白。外来蛋白可自由地与噬菌体所接触的其它部分(moiety)相互作用(结合)。每个展示外来蛋白的噬菌体都是所述噬菌体文库中的一个“成员”。

[0195] “抗体噬菌体展示文库”指展示抗体或抗体片段的噬菌体展示文库。抗体文库包括噬菌体群体或编码这样的噬菌体群体的载体群集,或含有这样的噬菌体或载体群集的细胞。文库可以是单价的,平均每个噬菌体颗粒展示一个单链抗体或抗体片段,或者可以是多价的,平均每个病毒颗粒展示两个或更多个抗体或抗体片段。术语“抗体片段”包括,但不限于,单链 Fv(scFv)片段和 Fab 片段。优选的抗体文库平均包含超过 10^6 ,或超过 10^7 ,或超过 10^8 ,或超过 10^9 个不同的成员。

[0196] 术语“丝状噬菌体”指能够在其表面上展示异源多肽的病毒颗粒,包括,但不限于, f1、fd、Pf1 和 M13。丝状噬菌体可以包含选择性标记,诸如四环素(例如,“fd-tet”)。多种丝状噬菌体展示系统是本领域技术人员所熟知的(参见,例如, Zacher 等, Gene9:127-140(1980);Smith 等, Science228:1315-1317(1985);和 Parmley 和 Smith Gene73:305-318(1988))。

[0197] 术语“淘选(panning)”用于指在对携带与靶物具有高亲和力和特异性的化合物(诸如抗体)的噬菌体的鉴定和分离中的多轮筛选过程。

[0198] B. 详述

[0199] 用于实施本发明方法的技术是本领域熟知的,并且在标准实验室手册,包括例如 Ausubel 等, Current Protocols of Molecular Biology, John Wiley and Sons(1997); Molecular Cloning:A Laboratory Manual, Third Edition, J. Sambrook and D. W. Russell 编, Cold Spring Harbor, New York, USA, Cold Spring Harbor Laboratory Press, 2001; O' Brian 等, Antibody Phage Display, Methods and Protocols, Humana Press, 2001; Phage Display:A Laboratory Manual, C. F. Barbas III 等编, Cold Spring Harbor, New York, USA, Cold Spring Harbor Laboratory Press, 2001;和 Antibodies, G. Subramanian 编, Kluwer Academic, 2004 中有记载。例如,诱变可以使用定点诱变进行(Kunkel 等, Proc. Natl. Acad. Sci USA82:488-492(1985))。PCR 扩增方法在美国专利 4683192、4683202、4800159 和 4965188,和几种教科书包括“PCR Technology:Principles and Applications for DNA Amplification”, H. Erlich 编, Stockton Press, New York(1989);和“PCR Protocols:A Guide to Methods and Applications”, Innis 等编, Academic Press, San Diego, Calif. (1990) 中描述。

[0200] 关于使用 Kabat 数据库和 Kabat 规程进行的抗体序列分析的信息可以在例如 Johnson 等, The Kabat database and a bioinformatics example, Methods Mol Biol. 2004;248:11-25;和 Johnson 等, Preferred CDRH3 lengths for antibodies with defined specificities, Int Immunol. 1998, Dec;10(12):1801-5 中找到。

[0201] 关于使用 Chothia 规程进行的抗体序列分析的信息可以在例如 Chothia 等, Structural determinants in the sequences of immunoglobulin variable domain, J Mol Biol. 1998 May 1;278(2):457-79;Morea 等, Antibody structure, prediction and redesign, Biophys Chem. 1997;68(1-3):9-16.;Morea 等, Conformations of the third hypervariable region in the VH domain of immunoglobins; J Mol

Biol. 1998, 275(2):269-94; Al-Lazikani 等, Standard conformations for the canonical structures of immunoglobins, J Mol Biol. 1997, 273(4):927-48. Barre 等, Structural conservation of hypervariable regions in immunoglobins evolution, Nat Struct Biol. 1994, 1(12):915-20; Chothia 等, Structural repertoire of the human VH segments, J Mol Biol. 1992, 227(3):799-817 Conformations of immunoglobulin hypervariables, Nature. 1989, 342(6252):877-83; 和 Chothia 等, Review Canonical structures for the hypervariables of immunoglobins, J Mol Biol. 1987, 196(4):901-17) 中找到。

[0202] 1. in silico 多样性 (多) 肽文库设计

[0203] 根据本发明, 多样性 (多) 肽文库的设计首先是使用相关感兴趣 (多) 肽序列的数据库, 以及, 典型地, 鉴定由文库中各个成员所共享的序列基序。多种用于鉴定多肽中序列基序的计算机程序是本领域熟知的, 并可在线使用。因此, 例如, 序列基序可以使用 ELPH (一种用于在一组 DNA 或蛋白质序列中查找基序的通用型 Gibbs 采样器), MEME (多重 EM 基序引出 (Multiple EM for Motif Elicitation) 系统, 用来在多组相关 DNA 或蛋白质序列中发现高度保守区); PPSEARCH (可用于在 PROSITE 数据库 (EBI) 中搜索序列的基序或功能模式); emotif (一种研究系统, 其针对被比对序列的子集形成基序, 并且通过基序的特异性和其覆盖的供给序列的数目二者来对其所找到的基序进行评级 (Stanford Bioinformatics Group)); 等来鉴定。

[0204] 在接下来的步骤中, 将鉴定出的一个或多个序列基序相互进行比对, 并把它们细分成不同的数据集, 每个数据集的特征在于共享预定的参数组合, 其中所述参数是一个或多个被比对的序列基序的特征参数。例如, 这样的参数可以是长度、特定序列基序所属的亚家族、该序列来源的物种、生物学功能等。然后, 对于以给定的两种或更多种参数的组合为特征的数据集进行逐位氨基酸使用频率分析, 从而鉴定数据集内单独的氨基酸序列段 (stretches) 中的关键氨基酸使用率。

[0205] 序列基序的比对可以以本领域技术范围内的多种方式实现, 例如通过使用公众可以获得的计算机软件如 BLAST、BLAST-2、ALIGN 或 Megalign (DNASTAR) 软件。本领域技术人员能够确定用于测量比对的适当参数, 包括在被比较的序列的全长范围内实现最大比对所需的任何算法。

[0206] 氨基酸使用频率的确定可以基于数据集全部成员中给定位置 (保守氨基酸残基) 上的高度同一性 (通常为至少 50%) 的出现, 优选完全同一性的出现; 或者基于就给定的残基位置而言某种氨基酸残基在数据集的两个或更多个成员 (优选大多数成员) 中的出现。然后可以创建以一种或多种附加参数为特征的附加数据集, 所述附加数据集无需全部与序列相关。

[0207] 例如, 如果目的是设计多样性抗体文库, 可以使用预定参数的独特组合 (过滤器) 来分析 Kabat 数据库 (一种含有非冗余重排抗体序列的电子数据库) 中存在的抗体重链和轻链 CDR 序列的位置频率。Kabat 数据库含有在提交时进行了注释的抗体蛋白序列。可以将来自 Kabat 数据库的信息导入到可方便地应用过滤器的其它环境下, 例如, Microsoft Access 数据库, 并且可以将结果制表并使用任何其它软件 (包括, 例如 Excel) 进一步分析。

[0208] 本发明的方法允许使用很多种 (a wide array of) 参数 (过滤器) 和参数 (过滤器) 组合来同时过滤抗体重链和轻链序列。因此, 针对特定重链的多样性数据集的生成可

以与所选择的轻链限制条件联系起来。例如,用于抗体重链 CDR 序列分析的过滤器可以包括以下的一种或多种:(1)与特定轻链类型(例如 kappa(κ)或 lambda(λ))的配对;(2)CDR 大小(例如 CDR1=6 个残基;CDR2=13 个残基);和(3)CDR3 亚家族(例如 V_H1 相对于 V_H3)。在轻链中,全部 CDR 都可以是大小匹配的。例如,可以预先决定 CDR1=7 个,CDR2=10 个,CDR3=8 个氨基酸残基。此外/或者,可以基于轻链亚家族的类型(例如 $\kappa 1$ 或 $\kappa 3$ 亚家族)来对轻链进行过滤(细分)。

[0209] 因此,例如,重链多样性分析可以基于与 κ 轻链的配对来进行,但是也可进一步将所述分析限制到与 $V_K 3$ 亚家族轻链配对的那些重链序列,或限制到具有长度为 8 个氨基酸的 CDR3 的轻链,或它们的组合。

[0210] 用于抗体重链和/或轻链的共变分析(covariant analysis)的附加过滤器可以包括,但不限于,同种型、抗原类型、亲和力和/或与 CDR 或抗体链类型或亚型不相关的位置残基。

[0211] 此外,本发明使得基于“有效的”(productive)重链和轻链配对来设计主题文库(themed libraries)成为可能。由此能够对抗相同抗原的多种抗体,包括商业抗体,进行多样性分析来鉴定最有可能在人类治疗中成功的抗体。

[0212] 如果目标是设计主题抗体文库(themed antibody library),那么基于有效性的重链和轻链配对,针对所选择的抗原选择一种或多种有效的(productive)(例如商品化的)抗体。然后,确定重链和轻链二者的种系来源(germline origin),并且对相同类型的重链和轻链 CDR 序列(例如 V_H3 , $V_K 1$)施以上述类型的多变量分析以创建多样性数据集。优选地,所述分析应该仅基于大小匹配的 CDR 来进行。

[0213] 在本发明的方法中,在比对和应用过滤器之后进行位置分析,以确定先前创建的数据集内的单独氨基酸或氨基酸组的位置频率,并且生成多样性数据集,例如 CDR 多样性数据集。在针对每个感兴趣的氨基酸位置确定了绝对位置氨基酸使用率之后,可以降低氨基酸的百分比使用率及总和使用率的阈值,以适用于更大的多样性覆盖率。因此,例如,需要的总覆盖率可以设定为高于 80%,且不呈现低于 10% 的单独氨基酸(with no individual amino acid being represented below 10%)。

[0214] in silico 建模以使用来自任何相关来源的附加建模信息持续更新,例如,来自基因和蛋白质序列和三维数据库和/或先前测试的多肽(例如抗体)的结果,从而使 in silico 数据库在其预测能力方面变得更加精确。

[0215] 此外, in silico 子集可以用生物学测定的结果来加以补充,例如,结合亲和力(affinity)/亲合力(avidity)结果,先前测定的抗体的生物学活性。这样,可以将结构特征与在期待用途上的预期性能更加紧密地关联起来。

[0216] 设计 CDR 多样性数据集之后合成提供所需多样性的组合(简并)寡核苷酸序列的群集,并且将所述群集克隆至合适的模板背景上。

[0217] 2. 构建多样性(多)肽文库

[0218] 在如上所述创建组合位置多样性数据集之后,可以通过多重合成寡核苷酸合成(multisyntheses oligonucleotide synthesis)生成物理组合多样性集合(physical combinatorial diversity sets)。根据本发明,不使用诱变密码或混合密码子三聚体,取而代之的是生成离散的简并寡核苷酸群集(discrete degenerate oligonucleotide

collections), 可将所述群集定量地限缩 (restrict) 或放宽 (relax) 以物理地呈现 (physically represent) 通过前述分析和设计而产生的组合多样性集合。放宽标准有助于通过合成较少的寡核苷酸探针来捕获期望的多样性, 或者, 如果克隆群集的能力超过了通过多样性分析生成的预测群集, 有助于合理地放大多样性集合。此外, 物理组合多样性集合可以包括虚拟多样性集合 (virtual diversity sets) 中不存在的副产物, 同时具有或不具有附加规则集 (additional rule sets)。这种方法在组合抗体文库生成领域中帮助最大, 但是也可以合理地延伸至其它适合的应用中, 例如生成各种多肽类型的文库 (例如生长因子文库) 等。重要的是应注意所述物理文库不是必须要求含有在任何给定位置上包含通过如上所述设定阈值百分比使用率而鉴定的所有氨基酸的成员 (members comprising all amino acids at any given position that were identified by setting the threshold percentage usage as described above)。出于多种理由, 例如为了降低需要的寡核苷酸的数目, 在某个给定位置上省略特定的氨基酸可能是有利的。或者 / 并且, 可以通过合成在给定位置上具有不满足预定阈值频率使用率的氨基酸残基的成员, 来增加文库的覆盖率和多样性。所述两种方法可以结合起来, 即可以省略 *in silico* 多样性数据集中存在的特定氨基酸残基, 同时可以加入未呈现于 *in silico* 多样性数据集中的给定位置上的氨基酸残基。

[0219] 创建本文的肽或多肽文库的第一步是对氨基酸群集进行反向翻译, 以进行多重合成 (multiplexed synthesis) 来含有完整的位置群集 (positional collection)。反向翻译工具是本领域熟知且可通过商业途径获得的。例如, 基于 Java 的逆向翻译工具 Entelechon (DE) 使用适合的密码子用法将蛋白质翻译成核苷酸序列, 并且容许针对在特定生物中的表达来优化序列。在优选的实施方案中, 本发明的方法采用自动反向翻译算法, 该算法能够合成离散且简并的寡核苷酸集合来呈现通过 *in silico* 分析创建的多样性表。这种算法可包括或排除特定密码子, 并且甚至可包含非等摩尔简并度, 从而不仅更精确地实现数据集的多样性, 还更精确地实现相对分布。

[0220] 需要的寡核苷酸的数目可以通过选择简并碱基以每次同时编码多于一种常用氨基酸来加以限制。此外, 可以限制这样的简并碱基来避免感兴趣的物种的罕用密码子。例如, 如果在大肠杆菌中合成所述群集, 可以在反向翻译中限制使用大肠杆菌罕用的精氨酸密码子用法。此外, 已知不是所有的氨基酸均以相同的频率被使用。因此, 可以使用非等摩尔混合物来更精确地反映虚拟 (*in silico*) 多样性表的概貌 (profile)。

[0221] 当位置多样性需要合成过多的寡核苷酸时, 可以使用化学探针群集来任意地限定多样性。因此, 可以将氨基酸侧链化学捕获 (capture) 在氨基酸子集内, 例如小氨基酸、疏水氨基酸、芳族氨基酸、碱性氨基酸、酸性氨基酸、酰胺类氨基酸、亲核氨基酸等氨基酸可构成这些子集。正如实施例将会阐述的, 这种带化学探针的 (chemically probed) 多样性位置可使用比其它方式所需数目少得多的寡核苷酸来合成。具有化学探针的多样性 (chemically probed diversity) 覆盖大部分天然存在的多样性, 并且提供广泛的相互作用化学。

[0222] 当构建本发明的多样性抗体文库时, 可以按照期望向抗体序列 (例如 CDR) 中引入经修饰的氨基酸残基, 例如大多数多肽中使用的 20 种传统氨基酸之外的残基 (例如, 高半胱氨酸)。这可以使用本领域公认的技术来进行, 所述技术通常将终止密码子掺入

多核苷酸中希望存在修饰的氨基酸残基处。该技术随后提供修饰的 tRNA (所谓的阻抑 tRNA (suppressor RNA), 例如琥珀、乳白或赭石终止密码子的阻抑 tRNA), 该 tRNA 连接着待掺入所述多肽的修饰氨基酸 (参见, 例如, Köhrer 等, PNAS, 98, 14310-14315 (2001))。

[0223] 在优选的实施方案中, 上述步骤中的一步或多步是由计算机辅助的。在具体的实施方案中, 计算机辅助的步骤包括, 例如, 挖掘 Kabat 数据库, 以及任选地, 将结果相对于 Vbase 序列目录 (Tomlinson, I. M. 等, .VBASE Sequence Directory. Cambridge, U. K.: MRC Centre for Protein Engineering; 1995) 进行相互参照。本发明的方法适用于高通量方法, 所述高通量方法包含执行多个步骤的软件 (例如, 计算机可读的指令) 和硬件 (例如, 计算机、机器人和芯片)。

[0224] 用于生成本文的文库的寡核苷酸可以通过已知用于 DNA 合成的方法来合成。已知的合成方法包括亚磷酰胺化学法 (Beaucage 和 Caruthers, Tetrahedron Letts., 22(20):18591862(1981)), 其允许有效的寡核苷酸 (oligo) 制备, 特别是最常见的 4080bp 大小范围的寡核苷酸, 其中使用自动合成仪, 如例如 Needham-VanDevanter 等 Nucleic Acids Res., 12:61596168(1984) 中所述。此外, 寡核苷酸可以通过本领域熟知的三酯、亚磷酸酯和 H-磷酸酯方法。关于寡核苷酸合成方法的综述, 参见, 例如 "Oligonucleotide Synthesis: A Practical Approach", ed. M. J. Gait, JRL Press, New York, N. Y. (1990)。寡核苷酸也可以从多种商业来源订购, 例如, The Midland Certified Reagent Company (Midland, TX)、The Great American Gene Company (Salt Lake City, UT)、ExpressGen Inc. (Chicago, IL)、Operon Technologies Inc. (Alameda, Calif.)。

[0225] 如果文库是抗体文库, 在接下来的步骤中, 将多样性克隆到框架中以产生多样性抗体文库。

[0226] 框架支架 (framework scaffold) 可以通过本领域熟知的方法来选择。因此, 可以选择数据库中最经常使用的框架用作支架, 并且将多样性克隆到种系框架中。对于框架序列的选择, 将确定为响应于特定抗原而表达的所有可用框架支架的子集排成阵列 (arrayed)。通过确定在自然条件下 (in nature) 响应于给定的抗原类别表达频率最高的框架, 选择合适的框架受体。例如, 为了确定响应于蛋白质抗原而表达的优选受体框架, 在 Kabat 数据库搜索 "定向于蛋白质的 (protein-directed)" 框架。如果需要优选的受体序列来呈现 (present) 针对不同的抗原类别的 CDR 和 / 或特定物种的受体序列, 则相应地设置 Kabat 蛋白序列过滤器。因此, 为了确定作为针对蛋白质靶物的人用治疗剂使用的序列, 将过滤器设置为仅集中针对识别蛋白质 / 肽抗原的人抗体序列。这大大降低了数据集中的冗余和可能使结果产生偏差的序列信息。对于 V_H 、 V_K 和 / 或 V_L 基因可以按照类似方式进行这种分析。

[0227] 可以将多样性群集掺入靶特异性受体以生成变体群集, 用于抗体工程。

[0228] 生成的 CDR 多样性可以通过本领域已知的方法诸如聚合酶链式反应 (PCR) 掺入框架区。例如, 可以使用寡核苷酸作为延伸用引物。在这种方法中, 编码对应于确定区域如 CDR (或其部分) 的诱变盒 (mutagenic cassette) 的寡核苷酸彼此互补, 并且使用聚合酶 (例如 Taq 聚合酶) 能够将它们延伸形成大的基因盒 (例如, scFv)。

[0229] 在另一种方法中, 设计部分重叠的寡核苷酸。内部寡核苷酸与它们的互补链退火产生双链 DNA 分子, 这样分子具有单链延伸, 可用于进一步退火。其后可以将退火后的配

对混合在一起,利用 PCR 进行延伸并且连接形成全长双链分子。可以在合成基因的末端附近设计方便的限制位点用于克隆入合适载体中。在这种方法中,也可以直接引入简并核苷酸来替代所述寡核苷酸之一。在引物延伸反应过程中,通过由聚合酶辅助的酶促延伸,从来自另一条链的部分互补寡核苷酸合成互补链。在合成阶段掺入简并多核苷酸可使克隆更简单,例如,在对基因中多于一个结构域或限定区域进行诱变或工程改造的来产生多样性的情况下。

[0230] 无论使用什么方法,在转化成双链形式之后,可以通过标准技术将寡核苷酸连接至合适的表达载体中。依靠适当的载体,诸如合适的质粒,可以将质粒引入无细胞提取物,或适于抗体表达的原核细胞或真核细胞。

[0231] 在一种不同的方法中,可以将希望的编码序列克隆到噬菌体载体或具有丝状噬菌体复制起点的载体(其允许利用辅助噬菌体来增殖单链分子)中。可以将单链模板与一组呈现期望突变的简并寡核苷酸退火、延伸并连接,由此将每种类似链(analog strand)掺入一群分子中,可以将这群分子引入适当的宿主(参见,例如,Sayers, J. R. 等, *Nucleic Acids Res.* 16:791-802(1988))。

[0232] 适于产生本文的文库(如合成人抗体文库)的多种噬菌粒克隆系统是本领域已知的,并且已经由例如 Kang 等, *Proc. Natl. Acad. Sci., USA*, 88:4363-4366(1991); Barbas 等, *Proc. Natl. Acad. Sci., USA*, 88:7978-7982(1991); Zebedee 等, *Proc. Natl. Acad. Sci., USA*, 89:3175-3179(1992); Kang 等, *Proc. Natl. Acad. Sci., USA*, 88:11120-11123(1991); Barbas 等, *Proc. Natl. Acad. Sci., USA*, 89:4457-4461(1992); 和 Gram 等, *Proc. Natl. Acad. Sci., USA*, 89:3576-3580(1992)进行了描述。

[0233] 文库的大小将依赖于 CDR 长度和需要呈现的 CDR 多样性的量而变化。优选地,文库将设计为含有少于 10^{15} 、 10^{14} 、 10^{13} 、 10^{12} 、 10^{11} 、 10^{10} 、 10^9 、 10^8 、 10^7 个,并且更优选 10^6 个或更少的抗体或抗体片段。

[0234] 根据本发明构建的文库也可以使用本领域公认的技术附接于固相支撑体,如微芯片,并且优选使其形成阵列(arrayed)。

[0235] 根据本发明构建的文库可以使用任何本领域已知的方法表达,所述方法包括,但不限于,细菌表达系统,哺乳动物表达系统,和体外核糖体展示系统。

[0236] 在优选的实施方案中,本发明包括使用噬菌体载体表达本文的多样性文库。所述方法通常包括使用用于克隆和表达的丝状噬菌体(噬菌粒)表面表达载体系统。参见,例如, Kang 等, *Proc. Natl. Acad. Sci., USA*, 88:4363-4366(1991); Barbas 等, *Proc. Natl. Acad. Sci., USA*, 88:7978-7982(1991); Zebedee 等, *Proc. Natl. Acad. Sci., USA*, 89:3175-3179(1992); Kang 等, *Proc. Natl. Acad. Sci., USA*, 88:11120-11123(1991); Barbas 等, *Proc. Natl. Acad. Sci., USA*, 89:4457-4461(1992); Gram 等, *Proc. Natl. Acad. Sci., USA*, 89:3576-3580(1992); Brinkman 等, *J. Immunol. Methods* 182:41-50(1995); Ames 等, *J. Immunol. Methods* 184:177-186(1995); Kettleborough 等, *Eur. J. Immunol.* 24:952-958(1994); Persic 等, *Gene* 187:9-18(1997); Burton 等, *Advances in Immunology* 57:191-280(1994); 和美国专利 Nos. 5698426; 5233409; 5580717; 5427908;

5750753 ;5821047 ;5403484 ;5571698 ;5516637 ;5780225 ;5658727 ;5733743 ;5837500 ;5969108 ;6326155 ;5885793 ;6521404 ;6492160 ;6492123 ;6489123 ;6342588 ;6291650 ;6225447 ;6180336 ;6172197 ;6140471 ;5994519 ;6969108 ;5871907 ;和 5858657。

[0237] 用载体转化重组宿主细胞,培养所述细胞以使得引入的噬菌体基因和展示蛋白基因得以表达,对于噬菌体颗粒而言,使其得以装配并从宿主细胞释放。其后从宿主细胞培养基收获(收集)释放的噬菌体颗粒并且筛选期望的抗体结合性质。通常,对收获的颗粒“淘选”其与预先选择的抗原的结合。收集强结合的颗粒,并且克隆分离单种颗粒(individual species of particles),并且进一步筛选与所述抗原的结合。选择产生具有期望的抗原结合特异性的结合位点的噬菌体。

[0238] 需要强调的是,本发明的方法不限于任何用于抗体文库表达和展示的特特定技术。其它展示技术也是合适的,诸如核糖体或 mRNA 展示(Mattheakis 等, Proc. Natl. Acad. Sci. USA91:9022-9026(1994);Hanes 和 Pluckthun, Proc. Natl. Acad. Sci. USA94:4937-4942(1997));微生物细胞展示,诸如细菌展示(Georgiou 等, Nature Biotech. 15:29-34(1997)),或酵母细胞展示(Kieckhefer 等, Protein Eng. 10:1303-1310(1997));哺乳动物细胞上的展示;孢子展示;病毒展示,诸如逆转录病毒展示(Urban 等, Nucleic Acids Res. 33:e35(2005));基于蛋白质-DNA 关联的展示(Odegrip 等, Proc. Acad. Natl. Sci. USA101:2806-2810(2004);Reiersen 等, Nucleic Acids Res. 33:e10(2005))和微珠展示(Sepp 等, FEBS Lett. 532:455-458(2002))。

[0239] 在核糖体展示中,抗体和编码 mRNA 通过核糖体连接,在翻译结束时,使 mRNA 终止而不释放多肽。基于作为整体的三元复合物进行选择。

[0240] 在 mRNA 展示文库中,抗体和编码 mRNA 之间藉由作为衔接分子的嘌呤霉素建立共价键(Wilson 等, Proc. Natl. Acad. Sci. USA98:3750-3755(2001))。关于使用这种技术来展示抗体,参见,例如, Lipovsek 和 Pluckthun, J. Immunol. Methods. 290:51-67(2004)。

[0241] 微生物细胞展示技术包括酵母(诸如酿酒酵母(*Saccharomyces cerevisiae*))上的表面展示(Boder 和 Wittrup, Nat. Biotechnol. 15:553-557(1997))。因此,例如,可以藉由与位于酵母细胞壁上的 α -凝集素酵母粘附受体(α -agglutinin yeast adhesion receptor)的融合,将抗体展示在酿酒酵母的表面上。这种方法提供了通过流式细胞仪筛选全套抗体(repertoire)的可能性。通过用荧光标记的抗原和抗表位标签试剂蜡染色细胞,可以根据细胞表面上的抗原结合和抗体表达水平来分选酵母细胞。酵母展示平台还可以与噬菌体联用(参见,例如, Van den Beucken 等, FEBS Lett. 546:288-294(2003))。

[0242] 关于用于选择和筛选抗体文库的技术的综述,参见,例如, Hoogenboom, Nature Biotechnol. 23(9):1105-1116(2005)。

[0243] 将通过以下非限定性实施例对本发明进行说明。

实施例

[0244] 用于进行本发明的方法的技术是本领域熟知的并且在标准实验室手册中有所描述,包括,例如, Ausubel 等, Current Protocols of Molecular Biology, John Wiley and Sons(1997);Molecular Cloning:A Laboratory Manual, Third Edition, J. Sambrook 及 D. W. Russell 编, Cold Spring Harbor, New York, USA, Cold Spring Harbor Laboratory

Press, 2001; O' Brian 等, Antibody Phage Display, Methods and Protocols, Humana Press, 2001; Phage Display: A Laboratory Manual, C. F. Barbas III 等编, Cold Spring Harbor, New York, USA, Cold Spring Harbor Laboratory Press, 2001; 和 Antibodies, G. Subramanian 编, Kluwer Academic, 2004。例如, 诱变的进行可以使用定点诱变 (Kunkel 等, Proc. Natl. Acad. Sci. USA 82:488-492 (1985)); DNA Cloning, Vols. 1 and 2, (D. N. Glover, Ed. 1985); Oligonucleotide Synthesis (M. J. Gait 编 1984); PCR Handbook Current Protocols in Nucleic Acid Chemistry, Beaucage 编. John Wiley & Sons (1999) (Editor); Oxford Handbook of Nucleic Acid Structure, Neidle 编, Oxford Univ Press (1999); PCR Protocols: A Guide to Methods and Applications, Innis 等, Academic Press (1990); PCR Essential Techniques: Essential Techniques, Burke 编, John Wiley & Son Ltd (1996); The PCR Technique: RT-PCR, Siebert 编, Eaton Pub. Co. (1998); Antibody Engineering Protocols (Methods in Molecular Biology), 510, Paul, S., Humana Pr (1996); Antibody Engineering: A Practical Approach (Practical Approach Series, 169), McCafferty, Ed., Irl Pr (1996); Antibodies: A Laboratory Manual, Harlow 等, C. S. H. L. Press, Pub. (1999); Large-Scale Mammalian Cell Culture Technology, Lubiniecki, A. 编, Marcel Dekker, Pub., (1990)。Border 等, Yeast surface display for screening combinatorial polypeptide libraries, Nature Biotechnology, 15(6):553-7 (1997); Border 等, Yeast surface display for directed evolution of protein expression, affinity, and stability, Methods Enzymol., 328:430-44 (2000); 核糖体展示, 如 Pluckthun 等在美国专利 No. 6348315 中所述, 和 Profusion™, 如 Szostak 等在美国专利 Nos. 6258558; 6261804; 和 6214553 中所述; 和细菌周质表达, 如 US20040058403A1 中所述。

[0245] 有关使用 Kabat 规程的抗体序列分析的更详细内容可以参见, 例如, Johnson 等, The Kabat database and a bioinformatics example, Methods Mol Biol. 2004; 248:11-25; Johnson 等, Preferred CDRH3 lengths for antibodies with defined specificities, Int Immunol. 1998, Dec; 10(12):1801-5; Johnson 等, SEQHUNT. A program to screen aligned nucleotide and amino acid sequences, Methods Mol Biol. 1995; 51:1-15. 和 Wu 等, Length distribution of CDRH3 in antibodies; 和 Johnson 等, Proteins. 1993 May; 16(1):1-7. Review。

[0246] 有关使用 Chothia 规程的抗体序列分析的更详细内容可以参见, 例如, Chothia 等, Structural determinants in the sequences of immunoglobulins variable domain, J Mol Biol. 1998 May 1; 278(2):457-79; Morea 等, Antibody structure, prediction and redesign, Biophys Chem. 1997 Oct; 68(1-3):9-16.; Morea 等, Conformations of the third hypervariable region in the VH domain of immunoglobulins; J Mol Biol. 1998 Jan 16; 275(2):269-94; Al-Lazikani 等, Standard conformations for the canonical structures of immunoglobulins, J Mol Biol. 1997 Nov 7; 273(4):927-48. Barre 等, Structural conservation of hypervariable regions in immunoglobulins evolution, Nat Struct Biol. 1994 Dec; 1(12):915-20; Chothia 等, Structural repertoire of the human VH segments, J Mol Biol. 1992 Oct 5

;227(3):799-817 Conformations of immunoglobulin hypervariable regions, Nature. 1989Dec21-28;342(6252):877-83;和 Chothia 等, Review Canonical structures for the hypervariable regions of immunoglobulins, J Mol Biol. 1987Aug20;196(4):901-17)。

[0247] 有关 Chothia 分析的更详细内容在例如 Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Conformations of the third hypervariable region in the VH domain of immunoglobulins. J Mol Biol. 1998Jan16;275(2):269-94; Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. Structural repertoire of the human VH segments. J Mol Biol. 1992Oct5;227(3):799-817; Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR 等, Conformations of immunoglobulin hypervariable regions. Nature. 1989Dec21-28;342(6252):877-83; Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol. 1987Aug 20;196(4):901-17; 和 Chothia C, Lesk AM. The evolution of protein structures. Cold Spring Harb Symp Quant Biol. 1987;52:399-405 中有描述。

[0248] 有关 CDR 接触的考虑因素的更详细内容在例如 MacCallum RM, Martin AC, Thornton JM. Antibody-antigen interactions: contact analysis and binding site Topography. J Mol Biol. 1996Oct11;262(5):732-45 中有描述。

[0249] 有关本文涉及的抗体序列和数据库的更详细内容可见, 例如, Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. J Mol Biol. 1992Oct5;227(3):776-98; Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics. 2001Mar;17(3):282-3; [VBDB] www.mrc-cpe.cam.ac.uk/vbase-ok.php?menu=901; [KBTD] database.com; [BLST] www.ncbi.nlm.nih.gov/BLAST/; [CDHIT] bioinformatics.ljcrf.edu/cd-hi/; [EMBOSS] www.hgmp.mrc.ac.uk/Software/EMBOSS/; [PHYLIP] evolution.genetics.washington.edu/phylip.html; 和 [FASTA] fasta.bioch.virginia.edu。

[0250] 实施例 1

[0251] 抗体轻链 V_{κ} 1CDR1、2 和 3 序列的频率分析

[0252] 在第一步中, 从 Kabat 免疫学相关蛋白序列数据库 (Kabat Database of Sequences of Proteins of Immunological Interest) 收集了 2374 个人抗体 V_{κ} 1 轻链可变域序列。对于每个序列, 将基因序列翻译成相应的氨基酸序列, 并且根据 Kabat 编号系统按位置比对所述氨基酸序列。

[0253] 接下来, 对于所获得的 V_{κ} 1 轻链序列群集, 通过选择在位置 18-19 具有氨基酸“RV”的序列并且应用以下长度限制来进行过滤: CDR1=7 个氨基酸, CDR2=10 个氨基酸, 且 CDR3=8 个氨基酸。通过应用这些过滤器, 将最初 2374 个成员的群集减少至 771 个成员。

[0254] 通过仅使用含有从 CDR1 之前的“RV”基序直至整个 CDR3 序列的完整明确序列的成员 (entries), 将 V_{κ} 1 轻链可变域序列的数目进一步减少至 383。

[0255] 接下来, 比对这些序列, 将每个位置出现的氨基酸列表, 并且计算在每个位置上 20

种天然存在的氨基酸的分布,根据逐位的氨基酸绝对使用率 (absolution usage of amino acids by position) 来产生基于位置频率的 CDR 结构域多样性数据库。这个表格的结果示于图 2。

[0256] 通过对于任何给定位置仅报告 10% 以上的氨基酸使用率来进一步过滤图 2 中所列的数据集。结果列于图 3。为了评估指定的百分比使用率对于多样性的影响,通过仅包括在 5% 以上的氨基酸使用率来生成另一个数据集。结果列于图 4。通过比较图 3 和图 4 的数据集清楚可见,要求较低的氨基酸使用率百分比可实现更大的多样性覆盖。

[0257] 如图 5 中所示,为了编码图 4 中所列的轻链 CDR1 多样性,需要合成 128 种组合寡核苷酸或 16 种简并组合寡核苷酸。碱基无需是等摩尔的,并且可以加以调整来造成氨基酸使用率偏差以反映当前分析中发现的频率,甚至包括频率表中不包括的残基。或者 / 并且,可以省略频率表中包括的残基,例如以进一步减少合成所需的寡核苷酸的数目。

[0258] 实施例 2

[0259] V_H3 重链合成文库多样性的设计

[0260] 通过分析长度为 10 个氨基酸的 V_H3 重链多肽序列生成了表 6 中所示的数据,其中所述序列是基本上如实施例 1 中所述从 Kabat 抗体序列数据库获得的。如图 6 中所示,仅使用 96 种简并寡核苷酸,为各个氨基酸位置设置不同的阈值百分比使用率,可提供 3.3×10^5 的 CDR3 多样性,呈现除残基 97 外的全部位置的至少 75% 的位置覆盖率。因此,所述阈值百分比使用率第 93、94、100 和 101 位是 10%;第 95、96、98 和 99 位是 5%;第 97 位是 4%;第 100A 位是 3%。合成这种多样性需要的寡核苷酸序列示于图 7。

[0261] 实施例 3

[0262] 制备半合成抗体文库

[0263] 如先前所述,对 V_H CDR 多样性的分析和 V_H CDR 多样性的生成可以进行调整,以配合具体情况反映出与 κ 和 λ 轻链的有效且特异性的配对 (即导致抗体特异性结合靶抗原的配对) 的组成。这些合成 V_H 库 (repertoires) 并不需要排他性地与合成轻链库配对,而可以与淋巴细胞衍生的轻链的群集一起组合克隆。在实践中,将 κ 和 λ 轻链的群集分别克隆至噬菌体展示载体中,继而或者克隆单独重链可变区框架以便随后引入多样性,或者克隆预先多样化 (pre-diversified) 的可变区框架群集。在任一情况下,与轻链兼容配对的重链可变区 (the light chain compatibly paired heavy chain variable regions) 均可望更有效地与相应的轻链配对。

[0264] 实施例 4

[0265] 通过针对基础克隆产生变体文库来工程改造改进的抗体

[0266] 以类似于在种系受体框架 (germline acceptor framework) 上引入有效多样性 (productive diversity) 来从头产生免疫球蛋白库的方式,为特定的抗体或抗体的特定群集生成靶特异性诱变文库 (target specific mutagenesis library)。这样的文库可用于抗体工程作业,特别是在亲和力成熟领域中。从感兴趣的单克隆抗体出发,确定限定性特征 (defining characteristics),这些特征是在先前限定的本发明的多样性影响要素 (diversity influencing elements) 中捕捉 (capture) 的,所述要素例如种系框架起源、轻链类型以及轻链和重链 CDR 长度。在确定了这些特征或类似的特征之后,接下来的步骤是查阅 (refer to) 对应于这些参数的数据库序列。鉴定了相应的序列集合之后,进行类似

于前文所述的分析,以检查子集库(subset repertoire)的多样性,然后生成编码期望的多样性所必需的相应的多重简并寡核苷酸(multi-degenerate oligonucleotide)。然后将这些多重简并寡核苷酸作为单一或组合 CDR 群集予以克隆。由于使用多重 CDR 诱变更有可能发现协同性的改进(synergistic improvement),优选生成组合 CDR 诱变文库。使用来自上述分析的多重简并寡核苷酸,合理地生成抗体并根据与人类偏差和偏好相关的位置多样性重新多样化(re-diversify)抗体。重要的是应注意,在任意轻链 CDR 序列或重链 CDR1 或 CDR2 序列是从种系序列异化而来(diverge from the germline sequence)的情况下,相应的编码种系的寡核苷酸也要包括在组合 CDR 文库中。这样,通过将编码种系的寡核苷酸包括在内,为种系序列回交(backcrossing)以生成更有效(productive)的 CDR 组合提供了可能。

[0267] 这种“多样性再引入方案(diversity reincorporation scheme)”也可用于工程改造来自已有的合成抗体克隆的、重新多样化的抗体集合。由于根据本发明生成的合成文库的潜在多样性超过了目前可用的技术展示和选择全部成员的极限,非常有可能的是任何发现的靶特异性克隆仅代表在任何经通常筛选的文库中在 DNA 水平上存在且可访问的可能解决方案(possible solutions present and accessible at the DNA-level)中的一部分。因此,经过进行四轮淘选在本发明的文库中鉴定了一种抗 EFG 抗体之后,再将最初设计的多样性以组合方式重新引入(combinatorially reintroduce)克隆中以生成新的变体集合。然后通过针对 EGF 进行淘选并且逐轮增加结合和清洗的严紧度来重新筛选这些新的变体集合。净结果生成了 EGF 结合噬菌体的池(pools),其与原始淘选中存在的那些相比被富集至更高的相对于背景的水平。

[0268] 实施例 5

[0269] 细胞因子主题文库的设计

[0270] 为了生成用于发现新的抗细胞因子抗体的有效性文库(productive library),选择了一种有效性(productive)抗 TNF- α 抗体 HUMIRA[®] (阿达木单抗(adalimumab))作为基础主题(basic theme)。HUMIRA[®] (阿达木单抗)是一种重组人 IgG1 单克隆抗体,系使用导致具有源自人的重链和轻链可变区和人 IgG1: κ 恒定区的抗体的噬菌体展示技术所产生。

[0271] 为了确定亲本抗体 D2E7 的重链的种系来源,对框架区进行了分析。这通过掩蔽 D2E7 和人种系 VH 基因的 CDR 来实现。接下来,通过 BLAST 算法将 D2E7 的 FR1 和 FR3 之间的其余序列相对于全部人种系 VH 基因进行比对。如图 8 中所示,D2H7V_H 区显示出与 V_H3_3-09 的最大相似度。图 9 中所示的树形比对图显示了相同的结果。按照相似的方式,发现亲本抗体 D2E7 的轻链与 V_K1A20 最相似(图 10 和 11)。

[0272] 对实施例 1 中所述的对抗体轻链 V_K1CDR1、CDR2 和 CDR3 序列的频率分析进行修改,将阈值百分比使用率过滤器设置为 6%。如图 12 中所示,使用这个过滤器,对于除位置 91 外的全部氨基酸位置,总和使用率均超过 80%,其提供了 9×10^6 的文库多样性,并且这种多样性可由 30 种简并寡核苷酸来提供。

[0273] 接下来,从 Kabat 免疫学相关蛋白序列数据库收集了 5971 个人抗体重链可变域序列。对于每个序列,将基因序列翻译成相应的氨基酸序列,并且根据 Kabat 编号系统按位置

比对所述氨基酸序列。

[0274] 然后对该重链可变域群集施用以下过滤器：

[0275] 1. 在氨基酸位置 22-25 含有“CAAS”(SEQ ID NO:3) 的 V_H3 序列 (5971 个成员中的 1530 个)；

[0276] 2. 与 κ 轻链组合的序列, CDR1=6 个氨基酸并且 CDR2=13 个氨基酸 (1530 个成员中的 226 个)；

[0277] 3. 仅包括含有从 CDR1 之前的“CAAS”(SEQ ID NO:3) 直至整个 CDR2 序列的完整序列的成员 (226 个成员中的 180 个)。

[0278] 然后, 比对所述序列, 将每个位置出现的氨基酸列表, 并且计算在每个位置上 20 种天然存在的氨基酸的分布以基于逐位的氨基酸绝对使用率来产生基于位置频率的 CDR 结构域多样性数据库。该制表结果示于图 13。

[0279] 通过对于任何给定位置仅报告至少为 10% 的氨基酸使用率来进一步过滤图 13 中所列的数据集。如图 14 中所示, 使用这种过滤器, 在 CDR2 中, 第 52、52A、55 和 58 位的总和氨基酸覆盖率小于 75%。为了提供更大的覆盖率, 将要求的百分比使用率从 10% 降低至 5%。如图 15 中所示, 这种变化已经使得全部位置的总和氨基酸使用率升高至 75% 以上。

[0280] 对 CDR1 和 CDR2 二者应用 5% 使用率过滤器, 合成 CDR1 区需要 4 种简并寡核苷酸, CDR2 多样性可以由 28 种简并寡核苷酸编码 (见图 16)。由此, 使用总共 28 种简并寡核苷酸, 能够实现 1.5×10^8 的总体多样性, 提供超过 80% 的位置覆盖率。

[0281] 在下一步中, 从上述 5971 个人抗体重链可变域序列, 不考虑同种型, 汇编 (compile) 长度为 13 个氨基酸的 V_H3 序列。将每个位置所要求的百分比氨基酸使用率设定为 4%, 除了第 93、94 和 101 位氨基酸将阈值设定为 4% 使用率之外。结果示于图 17。通过设定这些阈值, 能够通过使用 384 种简并寡核苷酸来制备具有 7.5×10^9 的总体多样性的合成 V_H3 重链合成文库。可见, CDR3 区中的残基显示出与亲本抗体 D2E7 中相应残基的良好一致性。

[0282] 实施例 6

[0283] 半抗原主题抗体文库的设计

[0284] 这种方法的目的是为鉴定新的抗半抗原抗体设计有效性 (productive) 文库。

[0285] 该设计从抗洋地黄毒苷 (抗 DIG) 抗体开始 (Dorsam, H. 等, FEBS Lett. 414:7-13(1997))。这种抗体的 Ig λ 轻链可变区序列 (SEQ ID NO:1) 和重链可变区序列 (SEQ ID NO:2) 示于图 18。

[0286] 为确定这种亲本抗体的重链和轻链的种系来源进行了分析。如图 19 中所示, V_L-1g 与所述轻链最相似, 而 V_H3-23 与所述重链最相似, 因此, 将 CDR 置于这种环境中来创造有效性 (productive) 文库用于鉴定抗半抗原抗体。

[0287] 接下来, 如前文实施例中所述对轻链 CDR1 和 CDR2 序列分析了 λ 长度匹配的 V_L 框架残基。将对于每个位置要求的百分比氨基酸使用率设定为 6%, 从而低于 6% 的单独序列不被报告。如图 20 中所示, 这种过滤器为每个氨基酸位置提供了极好的覆盖率。对 H3 长度匹配 (8 个氨基酸) 的重链进行了类似的分析, 但是应用 6.25% 过滤器, 包括全部位置在内的总和氨基酸覆盖率在 75% 以上 (图 21)。

[0288] 实施例 7

[0289] 细胞因子 (IFN- α) 分析和文库创建

[0290] IFN 是具有抗病毒活性的细胞因子的通称,其中在病毒或双链核酸刺激下由白细胞或淋巴母细胞的那些细胞因子称作 IFN- α 。IFN- α 具有多种活性,包括抗病毒活性和细胞生长抑制活性,已经发现这些活性可用于治疗多种疾病,例如乙型和丙型肝炎感染,以及癌症。对来自多种 DNA 文库克隆的 IFN- α 基因序列的分析揭示了 IFN- α 以几种亚型存在。例如,对于 IFN- α 2 基因,已经鉴定了三种额外的类型 (α 2a、 α 2b 和 α 2c)。总体而言,存在超过 20 种目前已知的 IFN- α 亚型。其它已知的亚型包括,例如,IFN- α 1a、IFN- α 1b、IFN- α 4a、IFN- α 4b、IFN- α 5、IFN- α 6 等。已经证明了多数 IFN- α 亚型在它们的生物活性和其它生物学性质方面有差异。因此,基于 IFN- α 家族成员之间现存的天然多样性创建的文库可用于生成具有新的和改进的性质的 IFN- α 多肽,所述性质例如增加的效价、降低的免疫原性、增加的半衰期、改进的蛋白水解稳定性。

[0291] 作为创建多样性 IFN- α 文库的第一步,鉴定了 11 个长度为 189 个氨基酸的基因产物。将这些 IFN- α 多肽的第 32-38 位氨基酸残基相互比对,确定了残基使用频率,如图 22 中所示。当阈值百分比氨基酸使用率设定为 9% 时,使用 2 种简并寡核苷酸能够达到 100% 覆盖率 (见图 22 和 23)。如图 23 中所示,使用无简并设计,需要 40 种寡核苷酸来提供要求的覆盖率。

[0292] 一旦制成了文库,对期望的新性质的筛选可以通过本领域已知的方法来进行。因此,可以用标准生物学测定法测试增加的效价,例如通过生物淘选 (biopanning) 噬菌体展示 IFN- α 文库。可以例如通过针对 IFN- α 受体生物淘选噬菌体展示文库,或者通过将文库成员暴露于一种或多种血清蛋白酶,来鉴定具有增加的半衰期的成员。降低的免疫原性的测试可以通过例如鉴定文库中存在的与 MHC 分子显示最小结合的肽或多肽,或者通过直接测试完整蛋白质的 T 细胞表位呈递来进行。

[0293] 这些测试和多种其它测试是相关领域普通技术人员们熟知的。

[0294] 实施例 8

[0295] 具有化学探针的 (chemically probed) 的抗体群集

[0296] 本实施例显示使用基于化学原理设计的探针集来创建 CDR3 重链多样性。

[0297] 氨基酸可以分成七类,其特征分别在于小的、亲核的、疏水的、芳族的、酸性的、酰胺类的和碱性的侧链化学功能性 (图 24)。图 25 中左上图面显示七组中的每一组的氨基酸的单字母符号。选择了九种代表不同侧链化学的氨基酸 (A, S, H, L, P, Y, D, Q, R)。如图 25 的剩余部分所示,通过九种密码子或 2 种简并密码子可编码高亮显示的九种氨基酸,并且由此可捕获侧链化学多样性。(B=C、G 或 T;M=A 或 C;Y=C 或 T;D=A、G 或 T。)

[0298] 天然重链 CDR3 序列含有高度的化学多样性 (大约 60% 或更多)。已经确定,通过使用 128 种简并寡核苷酸的组合式简并寡核苷酸合成可生成类似的化学多样性。相应的简并寡核苷酸的设计示于图 27。如图 26 中所示,这种方法覆盖了大部分天然存在的多样性并且提供了广泛的相互作用化学。

[0299] 这种具有化学探针的多样性方法可以单独使用,或与任何本发明的其它方法组合使用,从而产生具有期望性质的组合文库。

[0300] 尽管在前述说明书中参考特定实施方案阐释了本发明,但是本发明不限于此。事实上,除了本文说明和描述的那些之外,根据前述说明书对本发明的各种修改对于本领域

那些技术人员将是显而易见的并且在所附权利要求书的范围之内。因此,尽管本发明是援引抗体文库进行示例说明的,但是可将其广泛地扩展至所有的肽和多肽文库。

[0301] 说明书全文中引用的全部参考文献通过提述明确并入本文。

[0001]

序列表

<110> 航道生物技术有限责任公司(SEA LANE BIOTECHNOLOGIES, LLC)

<120> 多样性合成肽和多肽文库的设计和构建

<130> SLN-0005. CN

<140>

<141>

<150> PCT/US2007/079849

<151> 2007-09-28

<150> 60/849,035

<151> 2006-10-02

<160> 176

<170> PatentIn version 3.5

<210> 1

<211> 111

<212> PRT

<213> 人(Homo sapiens)

<400> 1

Ser Tyr Glu Leu Thr Gln Pro Pro Ser Ala Ser Gly Thr Pro Gly Gln
1 5 10 15

Arg Val Thr Ile Ser Cys Ser Gly Ser Thr Ser Asn Ile Gly Ser Asn
20 25 30

Tyr Val Tyr Trp Tyr Gln His Leu Pro Gly Thr Ala Pro Lys Leu Leu
35 40 45

Ile Phe Arg Asn Ser Gln Arg Pro Ser Gly Val Pro Asp Arg Phe Ser
50 55 60

Gly Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile Ser Gly Leu Arg

[0002]

65	70	75	80
Ser Glu Asp Glu Ala Asp Tyr Tyr Cys Ala Ala Trp Asp Asp Ser Leu			
85	90	95	
Ser Gly Val Val Phe Gly Gly Gly Thr Lys Leu Thr Val Leu Gly			
100	105	110	
<210> 2			
<211> 116			
<212> PRT			
<213> 人			
<400> 2			
Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly			
1	5	10	15
Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr			
20	25	30	
Ala Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val			
35	40	45	
Ser Ala Ile Ser Gly Ser Gly Gly Ser Thr Tyr Tyr Ala Asp Ser Val			
50	55	60	
Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr			
65	70	75	80
Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys			
85	90	95	
Ala Lys Asp Lys Gly Ser Gly Trp Tyr Trp Gly Gln Gly Thr Leu Val			
100	105	110	

[0003]

Thr Val Ser Ser
115

<210> 3
<211> 4
<212> PRT
<213> 人

<400> 3
Cys Ala Ala Ser
1

<210> 4
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 对人工序列的描述: 合成的寡核苷酸

<400> 4
vrtmvttdke tcrymtggtc t 21

<210> 5
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 对人工序列的描述: 合成的寡核苷酸

<400> 5
sgtamttgkc tcgcatgggc t 21

<210> 6
<211> 21
<212> DNA
<213> 人工序列

[0004]

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 6

artamttgke tcgcattggc t

21

<210> 7

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 7

sgtngttgke tcgcattggc t

21

<210> 8

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 8

artngttgke tcgcattggc t

21

<210> 9

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 9

sgtamttwte tcgcattggc t

21

<210> 10

[0005]

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 10

artamtwtgc tcgcattggtc t

21

<210> 11

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 11

sgtmggttgc tcgcattggtc t

21

<210> 12

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 12

artmggttgc tcgcattggtc t

21

<210> 13

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 13

[0006]

sgtamttgkc tcatctgggc t	21
<210> 14	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 14	
artamttgkc tcatctgggc t	21
<210> 15	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 15	
sgtmgttgkc tcatctgggc t	21
<210> 16	
<211> 21	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 16	
artmgttgkc tcatctgggc t	21
<210> 17	
<211> 21	
<212> DNA	
<213> 人工序列	

[0007]

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 17

sgtamttwtc tcatctgggc t

21

<210> 18

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 18

artamttwtc tcatctgggc t

21

<210> 19

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 19

sgtmgttwtc tcatctgggc t

21

<210> 20

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 20

artmgttwtc tcatctgggc t

21

<210> 21

[0008]

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 21

accgcagttt actattgcgc acgtvtgskt aagkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 22

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 22

accgcagttt actattgcgc aaagvtgskt aagkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 23

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 23

accgcagttt actattgcgc acgtgrwskt aagkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 24

<211> 65

<212> DNA

[0009]

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 24

accgcagttt actattgcgc aaaggrwskt aagkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 25

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 25

accgcagttt actattgcgc acgtvtggmt aagkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 26

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 26

accgcagttt actattgcgc aaagvtggmt aagkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 27

<211> 65

<212> DNA

<213> 人工序列

[0010]

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 27

accgcagttt actattgcgc acgtgrwgmt aagkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 28

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 28

accgcagttt actattgcgc aaaggrwgmt aagkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 29

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 29

accgcagttt actattgcgc acgtvtgskt gdtkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 30

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

[0011]

<400> 30	
accgcagttt actattgcgc aaagvtgskt gdtkmtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 31	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 31	
accgcagttt actattgcgc acgtgrwskt gdtkmtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 32	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 32	
accgcagttt actattgcgc aaaggrwskt gdtkmtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 33	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 33	

[0012]

accgcagttt actattgcgc acgtvtggmt gdtkmtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 34	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 34	
accgcagttt actattgcgc aaagvtggmt gdtkmtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 35	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 35	
accgcagttt actattgcgc acgtgrwgmt gdtkmtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 36	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 36	
accgcagttt actattgcgc aaaggrwgmt gdtkmtbrtg stwtkgatta ctggggtcag	60

[0013]

ggcac	65
<210> 37	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<220>	
<221> 修饰的碱基	
<222> (32)..(32)	
<223> a, c, g, t, 未知或其它	
<400> 37	
accgcagttt actattgcgc acgtvtgskt cntkmtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 38	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<220>	
<221> 修饰的碱基	
<222> (32)..(32)	
<223> a, c, g, t, 未知或其它	
<400> 38	
accgcagttt actattgcgc aaagvtgskt cntkmtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 39	

[0014]

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 39

accgcagttt actattgcgc acgtgrwskt cntkmtbrtg stwtkgatta ctggggteag 60

ggcac 65

<210> 40

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 40

accgcagttt actattgcgc aaaggrwskt cntkmtbrtg stwtkgatta ctggggteag 60

ggcac 65

<210> 41

<211> 65

<212> DNA

<213> 人工序列

[0015]

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 41

accgcagttt actattgcgc acgtvtggmt cntkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 42

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 42

accgcagttt actattgcgc aaagvtggmt cntkmtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 43

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

[0016]

<220>		
<221> 修饰的碱基		
<222> (32)..(32)		
<223> a, c, g, t, 未知或其它		
<400> 43		
accgcagttt actattgcgc acgtgrwgmt entkmtbrtg stwtkgatta ctggggtcag	60	
ggcac	65	
<210> 44		
<211> 65		
<212> DNA		
<213> 人工序列		
<220>		
<223> 对人工序列的描述: 合成的寡核苷酸		
<220>		
<221> 修饰的碱基		
<222> (32)..(32)		
<223> a, c, g, t, 未知或其它		
<400> 44		
accgcagttt actattgcgc aaaggrwgmt entkmtbrtg stwtkgatta ctggggtcag	60	
ggcac	65	
<210> 45		
<211> 65		
<212> DNA		
<213> 人工序列		
<220>		
<223> 对人工序列的描述: 合成的寡核苷酸		
<400> 45		
accgcagttt actattgcgc acgtvtgskt aagsgtbrtg stwtkgatta ctggggtcag	60	
ggcac	65	

[0017]

<210> 46

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 46

accgcagttt actagtgcgc aaagvtgskt aagsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 47

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 47

accgcagttt actattgcgc acgtgrwskt aagsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 48

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 48

accgcagttt actattgcgc aaaggrwskt aagsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 49

[0018]

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 49

accgcagttt actattgcgc acgtvtggmt aagsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 50

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 50

accgcagttt actattgcgc aaagvtggmt aagsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 51

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 51

accgcagttt actattgcgc acgtgrwgmt aagsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 52

<211> 65

<212> DNA

[0019]

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 52

accgcagttt actattgcgc aaaggrwgmt aagsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 53

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 53

accgcagttt actattgcgc acgtvtgskt gdtsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 54

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 54

accgcagttt actattgcgc aaagvtgskt gtsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 55

<211> 65

<212> DNA

<213> 人工序列

[0020]

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 55

accgcagttt actattgcgc acgtgrwskt gdtsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 56

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 56

accgcagttt actattgcgc aaaggrwskt gdtsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 57

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 57

accgcagttt actattgcgc acgtvtggmt gdtsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 58

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

[0021]

<400> 58	
accgcagttt actattgcgc aaagvtggmt gdtsgtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 59	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 59	
accgcagttt actattgcgc acgtgrwgmt gdtsgtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 60	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 60	
accgcagttt actattgcgc aaaggrwgmt gdtsgtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 61	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	

[0022]

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

<400> 61

accgcagttt actattgcgc acgtvtgskt entsgtbrtg stwtkgatta ctggggtcag 60

ggcac

65

<210> 62

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

<400> 62

accgcagttt actattgcgc aaagvtgskt entsgrbrtg stwtkgatta ctggggtcag 60

ggcac

65

<210> 63

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

[0023]

<400> 63
accgcagttt actattgcgc acgtgrwskt entsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 64
<211> 65
<212> DNA
<213> 人工序列

<220>
<223> 对人工序列的描述: 合成的寡核苷酸

<220>
<221> 修饰的碱基
<222> (32).. (32)
<223> a, c, g, t, 未知或其它

<400> 64
accgcagttt actattgcgc aaaggrwskt entsgtbrtg stwtkgatta ctggggtcag 60

ggcac 65

<210> 65
<211> 65
<212> DNA
<213> 人工序列

<220>
<223> 对人工序列的描述: 合成的寡核苷酸

<220>
<221> 修饰的碱基
<222> (32).. (32)
<223> a, c, g, t, 未知或其它

<400> 65
accgcagttt actattgcgc acgtvtggmt entsgtbrtg stwtkgatta ctggggtcag 60

[0024]

ggcac	65
<210> 66	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述：合成的寡核苷酸	
<220>	
<221> 修饰的碱基	
<222> (32)..(32)	
<223> a, c, g, t, 未知或其它	
<400> 66	
accgcgttta ctattgcgca aagctgggmt cntsgtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 67	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述：合成的寡核苷酸	
<220>	
<221> 修饰的碱基	
<222> (32)..(32)	
<223> a, c, g, t, 未知或其它	
<400> 67	
accgcagttt actattgcgc acgtgrwgmt cntsgtbrtg stwtkgatta ctggggtcag	60
ggcac	65
<210> 68	

[0025]

<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<220>	
<221> 修饰的碱基	
<222> (32).. (32)	
<223> a, c, g, t, 未知或其它	
<400> 68	
accgcagttt actattgcgc aaaggrwgmt entsgtbrtg stwtkgatta ctggggteag	60
ggcac	65
<210> 69	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 69	
accgcagttt actattgcgc acgtvtgskt aagkmtbrtt mtwtkgatta ctggggteag	60
ggcac	65
<210> 70	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 70	
accgcagttt actattgcgc aaagvtgskt aagkmtbrtt mtwtkgatta ctggggteag	60

[0026]

ggcac	65
<210> 71	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述：合成的寡核苷酸	
<400> 71	
accgcagttt actattgcgc acgtgrwskt aagkmtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 72	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述：合成的寡核苷酸	
<400> 72	
accgcagttt actattgcgc aaaggrwskt aagkmtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 73	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述：合成的寡核苷酸	
<400> 73	
accgcagttt actattgcgc acgtvtggmt aagkmtbrtt mtwtkgatta ctggggtcag	60
ggcac	65

[0027]

<210> 74

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 74

accgcagttt actattgcgc aaagvtggmt aagkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 75

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 75

accgcagttt actattgcgc acgtgrwgmt aagkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 76

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 76

accgcagttt actattgcgc aaaggrwgmt aagkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 77

[0028]

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 77

accgcagttt actattgcgc acgtvtgskt gdtkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 78

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 78

accgcagttt actattgcgc aaagvtgskt gdtkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 79

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 79

accgcagttt actattgcgc acgtgrwskt gdtkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 80

<211> 65

<212> DNA

[0029]

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 80

accgcagttt actattgcgc aaaggrwskt gdtkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 81

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 81

accgcagttt actattgcgc acgtvtggmt gdtkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 82

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 82

accgcagttt actattgcgc aaagvtggmt gdtkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 83

<211> 65

<212> DNA

<213> 人工序列

[0030]

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 83

accgcagttt actattgcgc acgtgrwgmt gdtkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 84

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 84

accgcagttt actattgcgc aaaggrwgmt gdtkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 85

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 85

accgcagttt actattgcgc acgtvtgskt entkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 86

[0031]

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 86

accgcagttt actattgcgc aaagvtgskt cntkmtbrtt mtwtkgatta ctggggteag 60

ggcac 65

<210> 87

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 87

accgcagttt actattgcgc acgtgrwskt cntkmtbrtt mtwtkgatta ctggggteag 60

ggcac 65

<210> 88

<211> 65

<212> DNA

<213> 人工序列

[0032]

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 88

accgcagttt actattgcgc aaaggrwskt cntkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 89

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 89

accgcagttt actattgcgc acgtvtggmt cntkmtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 90

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

[0033]

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

<400> 90

accgcagttt actattgcgc aaagvtggmt entkmtbrtt mtwtkgatta ctggggtcag 60

ggcac

65

<210> 91

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

<400> 91

accgcagttt actattgcgc acgtgrwgmt entkmtbrtt mtwtkgatta ctggggtcag 60

ggcac

65

<210> 92

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

[0034]

<400> 92	
accgcagttt actattgcgc aaaggrwgmt cntkmtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 93	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 93	
accgcagttt actattgcgc acgtvtgskt aagsgtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 94	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 94	
accgcagttt actattgcgc aaagvtgskt aagsgtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 95	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 95	

[0035]

accgcagttt actattgcgc acgtgrwskt aagsgtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 96	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 96	
accgcagttt actattgcgc aaaggrwskt aagsgtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 97	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 97	
accgcagttt actattgcgc acgtvtggmt aagsgtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 98	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述: 合成的寡核苷酸	
<400> 98	
accgcagttt actattgcgc aaagvtggmt aagsgtbrtt mtwtkgatta ctggggtcag	60

[0036]

ggcac	65
<210> 99	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述：合成的寡核苷酸	
<400> 99	
accgcagttt actattgcgc acgtgrwgmt aagsgtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 100	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述：合成的寡核苷酸	
<400> 100	
accgcagttt actattgcgc aaaggrwgmt aagsgtbrtt mtwtkgatta ctggggtcag	60
ggcac	65
<210> 101	
<211> 65	
<212> DNA	
<213> 人工序列	
<220>	
<223> 对人工序列的描述：合成的寡核苷酸	
<400> 101	
accgcagttt actattgcgc acgtvtgskt gdtsgtbrtt mtwtkgatta ctggggtcag	60
ggcac	65

[0037]

<210> 102

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 102

accgcagttt actattgcgc aaagvtgskt gdtsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 103

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 103

accgcagttt actattgcgc acgtgrwskt gdtsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 104

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 104

accgcagttt actattgcgc aaaggrwskt gdtsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 105

[0038]

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 105

accgcagttt actattgcgc acgtvtggmt gdtsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 106

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 106

accgcagttt actattgcgc aaagvtggmt gdtsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 107

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 107

accgcagttt actattgcgc acgtgrwgmt gdtsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 108

<211> 65

<212> DNA

[0039]

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 108

accgcagttt actattgcgc aaaggrwgtt gdtsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 109

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

<400> 109

accgcagttt actattgcgc acgtvtgskt ontsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 110

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

[0040]

<400> 110
accgcagttt actattgcgc aaagvtgskt entsgrbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 111
<211> 65
<212> DNA
<213> 人工序列

<220>
<223> 对人工序列的描述：合成的寡核苷酸

<220>
<221> 修饰的碱基
<222> (32).. (32)
<223> a, c, g, t, 未知或其它

<400> 111
accgcagttt actattgcgc acgtgrwskt entsgtbrtt mtwtkgatta ctggggtcag 60

ggcac 65

<210> 112
<211> 65
<212> DNA
<213> 人工序列

<220>
<223> 对人工序列的描述：合成的寡核苷酸

<220>
<221> 修饰的碱基
<222> (32).. (32)
<223> a, c, g, t, 未知或其它

<400> 112
accgcagttt actattgcgc aaaggrwskt entsgtbrtt mtwtkgatta ctggggtcag 60

[0041]

ggcac

65

<210> 113

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

<400> 113

accgcagttt actattgcgc acgtvtggmt cntsgtbrtt mtwtkgatta ctggggtcag

60

ggcac

65

<210> 114

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32)..(32)

<223> a, c, g, t, 未知或其它

<400> 114

accgcagttt actattgcgc aaagvtggmt cntsgtbrtt mtwtkgatta ctggggtcag

60

ggcac

65

<210> 115

[0042]

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 115

accgcagttt actattgcgc acgtgrwgmt cntsgtbrtt mtwtkgatta ctggggteag 60

ggcac 65

<210> 116

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (32).. (32)

<223> a, c, g, t, 未知或其它

<400> 116

accgcagttt actattgcgc aaaggrwgmt cntsgtbrtt mtwtkgatta ctggggteag 60

ggcac 65

<210> 117

<211> 104

<212> PRT

<213> 人工序列

[0043]

<220>

<223> 对人工序列的描述：合成的多肽

<400> 117

Pro Ala Met Ala Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val
1 5 10 15

Gln Pro Gly Arg Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr
20 25 30

Phe Asp Asp Tyr Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly
35 40 45

Leu Glu Trp Val Ser Ala Ile Thr Trp Asn Ser Gly His Ile Asp Tyr
50 55 60

Ala Asp Ser Val Glu Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys
65 70 75 80

Asn Ser Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala
85 90 95

Val Tyr Tyr Cys Ala Lys Val Ser
100

<210> 118

<211> 99

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 118

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

[0044]

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Gly Ile Ser Trp Asn Ser Gly Ser Ile Gly Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Leu Tyr Tyr Cys
85 90 95

Ala Lys Asp

<210> 119

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 119

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Val Val Arg Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Gly Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

[0045]

Ser Gly Ile Asn Trp Asn Gly Gly Ser Thr Gly Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Leu Tyr His Cys
85 90 95

Ala Arg

<210> 120

<211> 99

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 120

Glu Val Gln Leu Val Glu Ser Gly Gly Val Val Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asp Asp Tyr
20 25 30

Thr Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Leu Ile Ser Trp Asp Gly Gly Ser Thr Tyr Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Ser Leu Tyr

[0046]

65	70	75	80
Leu Gln Met Asn Ser Leu Arg Thr Glu Asp Thr Ala Leu Tyr Tyr Cys			
	85	90	95
Ala Lys Asp			
<210> 121			
<211> 100			
<212> PRT			
<213> 人工序列			
<220>			
<223> 对人工序列的描述：合成的多肽			
<400> 121			
Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Lys Pro Gly Gly			
1 5 10 15			
Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Asn Ala			
20 25 30			
Trp Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val			
35 40 45			
Gly Arg Ile Lys Ser Lys Thr Asp Gly Gly Thr Thr Asp Tyr Ala Ala			
50 55 60			
Pro Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Asn Thr			
65 70 75 80			
Leu Tyr Leu Gln Met Asn Ser Leu Lys Thr Glu Asp Thr Ala Val Tyr			
85 90 95			

[0047]

Tyr Cys Thr Thr
100

<210> 122
<211> 100
<212> PRT
<213> 人工序列

<220>
<223> 对人工序列的描述: 合成的多肽

<400> 122
Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Thr Ala Ser Gly Phe Thr Phe Gly Asp Tyr
20 25 30

Ala Met Ser Trp Phe Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Gly Phe Ile Arg Ser Lys Ala Tyr Gly Gly Thr Thr Glu Tyr Thr Ala
50 55 60

Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Gly Ser Lys Ser Ile
65 70 75 80

Ala Tyr Leu Gln Met Asn Ser Leu Lys Thr Glu Asp Thr Ala Val Tyr
85 90 95

Tyr Cys Thr Arg
100

<210> 123
<211> 99
<212> PRT

[0048]

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 123

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Asp His
 20 25 30

Tyr Met Asp Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
 35 40 45

Gly Arg Thr Arg Asn Lys Ala Ser Tyr Thr Thr Glu Tyr Ala Ala Ser
 50 55 60

Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Asn Ser Leu
65 70 75 80

Tyr Leu Gln Met Asn Ser Leu Lys Thr Glu Asp Thr Ala Val Tyr Tyr
 85 90 95

Cys Ala Arg

<210> 124

<211> 100

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 124

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly

[0049]

1	5	10	15
Ser Leu Lys Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Gly Ser			
20	25	30	
Ala Met His Trp Val Arg Gln Ala Ser Gly Lys Gly Leu Glu Trp Val			
35	40	45	
Gly Arg Ile Arg Ser Lys Ala Asn Ser Tyr Ala Thr Ala Tyr Ala Ala			
50	55	60	
Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Asn Thr			
65	70	75	80
Ala Tyr Leu Gln Met Asn Ser Leu Lys Thr Glu Asp Thr Ala Val Tyr			
85	90	95	
Tyr Cys Thr Arg			
100			

<210> 125

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 125

Glu Val Gln Leu Leu Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly			
1	5	10	15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr			
20	25	30	

[0050]

Ala Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Ser Gly Ser Gly Gly Ser Thr Tyr Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Lys

<210> 126

<211> 97

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 126

Glu Val Gln Leu Val Glu Thr Gly Gly Gly Leu Ile Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Val Ser Ser Asn
20 25 30

Tyr Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Val Ile Tyr Ser Gly Gly Ser Thr Tyr Tyr Ala Asp Ser Val Lys
50 55 60

[0051]

Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr Leu
65 70 75 80

Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys Ala
85 90 95

Arg

<210> 127

<211> 97

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 127

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Val Ser Ser Asn
20 25 30

Tyr Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Val Ile Tyr Ser Gly Gly Ser Thr Tyr Tyr Ala Asp Ser Val Lys
50 55 60

Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr Leu
65 70 75 80

Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys Ala
85 90 95

[0052]

Arg

<210> 128

<211> 96

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 128

Glu Val Gln Leu Val Glu Ser Arg Gly Val Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Val Ser Ser Asn
20 25 30

Glu Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ser Ile Ser Gly Gly Ser Thr Tyr Tyr Ala Asp Ser Arg Lys Gly
50 55 60

Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu His Leu Gln
65 70 75 80

Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys Lys Lys
85 90 95

<210> 129

<211> 98

<212> PRT

<213> 人工序列

[0053]

<220>

<223> 对人工序列的描述：合成的多肽

<400> 129

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Tyr Val
35 40 45

Ser Ala Ile Ser Ser Asn Gly Gly Ser Thr Tyr Tyr Ala Asn Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
65 70 75 80

Leu Gln Met Gly Ser Leu Arg Ala Glu Asp Met Ala Val Tyr Tyr Cys
85 90 95

Ala Arg

<210> 130

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 130

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

[0054]

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
20 25 30

Trp Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Asn Ile Lys Gln Asp Gly Ser Glu Lys Tyr Tyr Val Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Arg

<210> 131

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 131

Gln Val Gln Leu Val Glu Ser Gly Gly Gly Val Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
20 25 30

Gly Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

[0055]

Ala Val Ile Ser Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Lys

<210> 132

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 132

Gln Val Gln Leu Val Glu Ser Gly Gly Gly Val Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
20 25 30

Gly Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Val Ile Ser Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr

[0056]

65	70	75	80
Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys			
	85	90	95

Ala Lys

<210> 133
<211> 98
<212> PRT
<213> 人工序列

<220>
<223> 对人工序列的描述: 合成的多肽

<400> 133
Gln Val Gln Leu Val Glu Ser Gly Gly Gly Val Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
20 25 30

Gly Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Val Ile Trp Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

[0057]

Ala Arg

<210> 134

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 134

Gln Val Gln Leu Val Glu Ser Gly Gly Gly Val Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Val Ile Ser Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Arg

<210> 135

<211> 98

<212> PRT

[0058]

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 135

Gln Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Lys Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Asp Tyr
 20 25 30

Tyr Met Ser Trp Ile Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
 35 40 45

Ser Tyr Ile Ser Ser Ser Gly Ser Thr Ile Tyr Tyr Ala Asp Ser Val
 50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
 85 90 95

Ala Arg

<210> 136

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 136

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Lys Pro Gly Gly

[0059]

1	5	10	15
Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr	20	25	30
Ser Met Asn Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val	35	40	45
Ser Ser Ile Ser Ser Ser Ser Ser Tyr Ile Tyr Tyr Ala Asp Ser Val	50	55	60
Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr	65	70	75
80			
Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys	85	90	95

Ala Arg

<210> 137

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 137

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly			
1	5	10	15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr		
20	25	30

[0060]

Ser Met Asn Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Tyr Ile Ser Ser Ser Ser Thr Ile Tyr Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Asp Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Arg

<210> 138

<211> 97

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 138

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
20 25 30

Asp Met His Trp Val Arg Gln Ala Thr Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Ala Ile Gly Thr Ala Gly Asp Thr Tyr Tyr Pro Gly Ser Val Lys
50 55 60

[0061]

Gly Arg Phe Thr Ile Ser Arg Glu Asn Ala Lys Asn Ser Leu Tyr Leu
65 70 75 80

Gln Met Asn Ser Leu Arg Ala Gly Asp Thr Ala Val Tyr Tyr Cys Ala
85 90 95

Arg

<210> 139

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 139

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
20 25 30

Trp Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Val Trp Val
35 40 45

Ser Arg Ile Asn Ser Asp Gly Ser Ser Thr Ser Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Thr Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

[0062]

Ala Arg

<210> 140

<211> 97

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 140

Ser Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val
1 5 10 15

Gly Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn
 20 25 30

Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu
 35 40 45

Ile Tyr Ala Ala Ser Thr Leu Gln Ser Gly Val Pro Ser Arg Phe Ser
 50 55 60

Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln
65 70 75 80

Pro Glu Asp Val Ala Thr Tyr Tyr Cys Gln Arg Tyr Asn Arg Ala Pro
 85 90 95

Tyr

<210> 141

[0063]

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 141

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

Asp	Arg	Val	Thr	Ile	Thr	Cys	Arg	Ala	Ser	Gln	Gly	Ile	Ser	Asn	Tyr
			20					25						30	

Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Gly	Lys	Val	Pro	Lys	Leu	Leu	Ile
			35				40						45		

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

Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Thr	Ile	Ser	Ser	Leu	Gln	Pro
			65				70						75		80

Glu	Asp	Val	Ala	Thr	Tyr	Tyr	Cys	Gln	Lys	Tyr	Asn	Ser	Ala	Pro	
							85						90		95

<210> 142

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 142

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

[0064]

Asp Arg Val Thr Ile Thr Cys Gln Ala Ser Gln Asp Ile Ser Asn Tyr
20 25 30

Leu Asn Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Asp Ala Ser Asn Leu Glu Thr Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Phe Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Ile Ala Thr Tyr Tyr Cys Gln Gln Tyr Asp Asn Leu Pro
85 90 95

<210> 143

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 143

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

Asp Arg Val Thr Ile Thr Cys Gln Ala Ser Gln Asp Ile Ser Asn Tyr
20 25 30

Leu Asn Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Asp Ala Ser Asn Leu Glu Thr Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

[0065]

Ser Gly Ser Gly Thr Asp Phe Thr Phe Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Ile Ala Thr Tyr Tyr Cys Gln Gln Tyr Asp Asn Leu Pro
85 90 95

<210> 144

<211> 94

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 144

Ala Ile Arg Met Thr Gln Ser Pro Phe Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Trp Ala Ser Gln Gly Ile Ser Ser Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Ala Lys Ala Pro Lys Leu Phe Ile
35 40 45

Tyr Tyr Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Ser
50 55 60

Gly Ser Gly Thr Asp Tyr Thr Leu Thr Ile Ser Ser Leu Gln Pro Glu
65 70 75 80

Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Tyr Tyr Ser Thr Pro
85 90

<210> 145

[0066]

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 145

Val Ile Trp Met Thr Gln Ser Pro Ser Leu Leu Ser Ala Ser Thr Gly
1 5 10 15

Asp Arg Val Thr Ile Ser Cys Arg Met Ser Gln Gly Ile Ser Ser Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Glu Leu Leu Ile
35 40 45

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

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Cys Leu Gln Ser
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Tyr Tyr Ser Phe Pro
85 90 95

<210> 146

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 146

Ala Ile Arg Met Thr Gln Ser Pro Ser Ser Phe Ser Ala Ser Thr Gly
1 5 10 15

[0067]

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Ser Ser Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

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

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Cys Leu Gln Ser
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr Pro
85 90 95

<210> 147

<211> 94

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 147

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

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Ser Ile Ser Ser Tyr
20 25 30

Leu Asn Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

[0068]

Ser Gly Ser Gly Thr Asp Phe Leu Thr Ile Ser Ser Leu Gln Pro Glu
65 70 75 80

Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Ser Tyr Ser Thr Pro
85 90

<210> 148

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 148

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

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Ser Ile Ser Ser Tyr
20 25 30

Leu Asn Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Ser Tyr Ser Thr Pro
85 90 95

<210> 149

[0069]

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 149

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

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Ser Ile Ser Ser Trp
 20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
 35 40 45

Tyr Asp Ala Ser Ser Leu Glu Ser Gly Val Pro Ser Arg Phe Ser Gly
 50 55 60

Ser Gly Ser Gly Thr Glu Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Asp Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Tyr Asn Ser Tyr Ser
 85 90 95

<210> 150

<211> 94

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 150

Ala Ile Gln Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

[0070]

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Ser Ser Ala
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Asp Ala Ser Ser Leu Glu Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Glu
65 70 75 80

Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Phe Asn Ser Tyr Pro
85 90

<210> 151

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 151

Ala Ile Gln Leu Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Ser Ser Ala
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Asp Ala Ser Ser Leu Glu Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

[0071]

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Phe Asn Ser Tyr Pro
85 90 95

<210> 152

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 152

Asp Ile Gln Leu Thr Gln Ser Pro Ser Phe Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Ser Ser Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

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

Ser Gly Ser Gly Thr Glu Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Leu Asn Ser Tyr Pro
85 90 95

<210> 153

[0072]

<211> 94

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 153

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

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Asp
20 25 30

Leu Gly Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Arg Leu Ile
35 40 45

Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Glu Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Leu Gln His Asn Ser Tyr Pro
85 90

<210> 154

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 154

Ala Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly
1 5 10 15

[0073]

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Arg Asn Asp
20 25 30

Leu Gly Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Leu Gln Asp Tyr Asn Tyr Pro
85 90 95

<210> 155

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 155

Asn Ile Gln Met Thr Gln Ser Pro Ser Ala Met Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Arg Gln Gly Ile Ser Asn Tyr
20 25 30

Leu Ala Trp Phe Gln Gln Lys Pro Gly Lys Val Pro Lys His Leu Ile
35 40 45

Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

[0074]

Ser Gly Ser Gly Thr Glu Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Leu Gln His Asn Ser Tyr Pro
85 90 95

<210> 156

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 156

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

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Ser Asn Tyr
20 25 30

Leu Ala Trp Phe Gln Gln Lys Pro Gly Lys Ala Pro Lys Ser Leu Ile
35 40 45

Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Tyr Asn Ser Tyr Pro
85 90 95

<210> 157

[0075]

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 157

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

Asp	Arg	Val	Thr	Ile	Thr	Cys	Arg	Ala	Ser	Gln	Gly	Ile	Ser	Ser	Trp
			20					25						30	

Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Glu	Lys	Ala	Pro	Lys	Ser	Leu	Ile
			35				40							45	

Tyr	Ala	Ala	Ser	Ser	Leu	Gln	Ser	Gly	Val	Pro	Ser	Arg	Phe	Ser	Gly
			50				55							60	

Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Thr	Ile	Ser	Ser	Leu	Gln	Pro
65						70				75					80

Glu	Asp	Phe	Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	Tyr	Asn	Ser	Tyr	Pro	
							85			90				95	

<210> 158

<211> 95

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 158

Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ser	Ser	Val	Ser	Ala	Ser	Val	Gly
1					5					10				15	

[0076]

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Ser Ser Trp
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Ala Ala Ser Ser Leu Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Ala Asn Ser Phe Pro
85 90 95

<210> 159

<211> 88

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的多肽

<400> 159

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Val Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Gly Ile Ser Ser Trp
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile
35 40 45

Tyr Gly Val Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe
50 55 60

[0077]

Thr Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Tyr
65 70 75 80

Cys Gln Gln Ala Asn Ser Phe Pro
85

<210> 160
<211> 25
<212> PRT
<213> 人工序列

<220>
<223> 对人工序列的描述: 合成的肽

<400> 160
Arg Asn Tyr Leu Ala Trp Tyr Leu Leu Ile Tyr Ala Ala Ser Thr Leu
1 5 10 15

Gln Gln Arg Tyr Asn Arg Ala Pro Tyr
20 25

<210> 161
<211> 19
<212> PRT
<213> 人工序列

<220>
<223> 对人工序列的描述: 合成的肽

<400> 161
Asp Asp Tyr Ala Met His Trp Val Ser Ala Ile Thr Trp Asn Ser Gly
1 5 10 15

His Ile Asp

[0078]

<210> 162

<211> 13

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的肽

<400> 162

Ala Lys Val Ser Tyr Leu Ser Thr Ala Ser Ser Leu Asp

1 5 10

<210> 163

<211> 98

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的多肽

<400> 163

Gln Ser Val Leu Thr Gln Pro Pro Ser Ala Ser Gly Thr Pro Gly Gln

1 5 10 15

Arg Val Thr Ile Ser Cys Ser Gly Ser Ser Ser Asn Ile Gly Ser Asn

20 25 30

Tyr Val Tyr Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu Leu

35 40 45

Ile Tyr Arg Asn Asn Gln Arg Pro Ser Gly Val Pro Asp Arg Phe Ser

50 55 60

Gly Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile Ser Gly Leu Arg

65 70 75 80

Ser Glu Asp Glu Ala Asp Tyr Tyr Cys Ala Ala Trp Asp Asp Ser Leu

85 90 95

[0079]

Ser Gly

<210> 164

<211> 14

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的肽

<400> 164

His	Ser	Leu	Gly	Ser	Arg	Arg	Thr	Leu	Met	Leu	Leu	Ala	Gln
1				5					10				

<210> 165

<211> 14

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的肽

<400> 165

His	Ser	Leu	Gly	His	Arg	Arg	Thr	Met	Met	Leu	Leu	Ala	Gln
1				5					10				

<210> 166

<211> 14

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的肽

<400> 166

His	Ser	Leu	Ser	Asn	Arg	Arg	Thr	Leu	Met	Ile	Met	Ala	Gln
1				5					10				

[0080]

<210> 167

<211> 14

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的肽

<400> 167

His Ser Leu Gly Asn Arg Arg Ala Leu Ile Leu Leu Ala Gln

1

5

10

<210> 168

<211> 14

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的肽

<400> 168

His Ser Leu Asn Asn Arg Arg Thr Leu Met Leu Met Ala Gln

1

5

10

<210> 169

<211> 14

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的肽

<400> 169

His Ser Leu Gly Asn Arg Arg Ala Leu Ile Leu Leu Gly Gln

1

5

10

<210> 170

<211> 14

<212> PRT

<213> 人工序列

[0081]

<220>

<223> 对人工序列的描述：合成的肽

<400> 170

His Ser Leu Arg Asn Arg Arg Ala Leu Ile Leu Leu Ala Gln

1 5 10

<210> 171

<211> 14

<212> PRT

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的肽

<400> 171

His Ser Leu Gly Asn Arg Arg Ala Leu Ile Leu Leu Ala Gln

1 5 10

<210> 172

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 172

cttvrtmatc gtcgtretmt g

21

<210> 173

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述：合成的寡核苷酸

<400> 173

cttratmatc gtcgtretmt g

21

[0082]

<210> 174

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<400> 174

cttvgtmatc gtcgtretmt g

21

<210> 175

<211> 33

<212> DNA

<213> 人工序列

<220>

<223> 对人工序列的描述: 合成的寡核苷酸

<220>

<221> 修饰的碱基

<222> (8).. (8)

<223> a, c, g 或 t

<220>

<221> 修饰的碱基

<222> (11).. (11)

<223> a, c, g 或 t

<220>

<221> 修饰的碱基

<222> (14).. (14)

<223> a, c, g 或 t

<220>

<221> 修饰的碱基

<222> (17).. (17)

<223> a, c, g 或 t

<220>

[0083]

<221> 修饰的碱基
<222> (20).. (20)
<223> a, c, g 或 t

<220>
<221> 修饰的碱基
<222> (23).. (23)
<223> a, c, g 或 t

<220>
<221> 修饰的碱基
<222> (26).. (26)
<223> a, c, g 或 t

<400> 175
getcgtbnkb nkbnkbnkbn kbnkbnkwtk gat

33

<210> 176
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> 对人工序列的描述：合成的寡核苷酸

<220>
<221> 修饰的碱基
<222> (8).. (8)
<223> a, c, g 或 t

<220>
<221> 修饰的碱基
<222> (11).. (11)
<223> a, c, g 或 t

<220>
<221> 修饰的碱基
<222> (14).. (14)
<223> a, c, g 或 t

<220>
<221> 修饰的碱基

[0084]

<222> (17)..(17)

<223> a, c, g 或 t

<220>

<221> 修饰的碱基

<222> (20)..(20)

<223> a, c, g 或 t

<220>

<221> 修饰的碱基

<222> (23)..(23)

<223> a, c, g 或 t

<220>

<221> 修饰的碱基

<222> (26)..(26)

<223> a, c, g 或 t

<400> 176

gctcgtnbb nbbnbbnbbn bbnbbnbwtk gat

33

通用抗体文库步骤

使用针对下列的独特过滤器分析 Kabat CDR 位置频率	
• 重链 - 轻链配对 例如 κ 或 λ - CDR 大小匹配 例如 CDR1=6 或 CDR2=13 - CDR3 亚家族过滤 例如 V_{H1} 对 V_{H3} • 轻链 - 全部 CDR 大小匹配 例如 选择仅含 CDR1=7, CDR2=10 且 CDR3=8 个氨基酸的序列条目 - 亚家族过滤 例如 K1 或 K3 • 位置分析—确定位置频率以生成 CDR 多样性集合 - 为单独氨基酸的%使用率设定下限(lower bound) - 为最常使用的氨基酸的总和%使用率设定下限 - 可以通过单一或多个 CDR、或者甚至通过单独的多个位置来限定下限	• 生成多样性集合 - 使用个别合成的寡核苷酸来捕捉整个组合位置多样性集合 - 可以使用简并合成 - 可以包括多样性集合中不存在的副产物，有或没有附加规则集 • 选择数据库中最高频使用的框架用作支架 - V_{H1} V_{H3} - V_{K1} V_{K3} - V_{L1} V_{L2} V_{L3} 将多样性克隆至种系框架上 • 可能的附加过滤器可以包括 - 抗原类型，诸如分泌性蛋白、病原体靶物，半抗原、自身免疫源等 • 重链同种型，诸如 IgG、IgM 等。

图 1

频率分析 V_k1 CDR1, 2 和 3

• 按位置计的绝对使用率

		30	31	32	33	34	35	36	46	47	48
#序列		383	383	383	383	383	383	383	383	383	383
阈值百分比使用率		0							N		L
A			3%	1%	2%	0%	0%	0%	0%	0%	0%
C			0%	0%	1%	0%	0%	0%	0%	0%	0%
D			4%	2%	8%	0%	0%	0%	0%	0%	0%
E			0%	0%	0%	0%	0%	0%	0%	0%	0%
F			0%	0%	8%	0%	0%	7%	1%	2%	0%
G			8%	2%	0%	7%	0%	0%	0%	0%	0%
H			0%	2%	1%	2%	0%	1%	0%	0%	0%
I			2%	3%	0%	0%	0%	0%	1%	0%	95%
K			2%	3%	0%	0%	0%	0%	0%	0%	0%
L			1%	0%	0%	0%	0%	1%	82%	97%	2%
M			0%	0%	0%	0%	0%	0%	0%	0%	2%
N			8%	34%	4%	42%	0%	0%	0%	0%	0%
P			0%	0%	0%	0%	0%	0%	1%	0%	0%
Q			0%	0%	0%	0%	0%	0%	0%	0%	0%
R			12%	5%	2%	0%	0%	0%	6%	0%	0%
S			54%	39%	6%	2%	0%	0%	4%	0%	0%
T			3%	6%	0%	2%	0%	0%	1%	0%	0%
V			1%	0%	0%	2%	0%	0%	4%	0%	1%
W			0%	0%	17%	0%	100%	0%	0%	0%	0%
Y			2%	1%	52%	0%	0%	92%	0%	0%	0%
呈现的% 总和百分比阈值			100%	100%	100%	100%	100%	100%	100%	100%	100%
80		CDR1								CDR 2	

图 2

49	50	51	52	53	54	55	89	90	91	92	93	94	95	96
383	383	383	383	383	383	383	383	383	383	383	383	383	383	383
							Q	G						
0%	55%	90%	1%	0%	0%	5%	0%	0%	5%	1%	2%	5%	1%	0%
0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%
0%	13%	0%	0%	0%	0%	1%	0%	0%	2%	11%	4%	1%	0%	1%
0%	1%	0%	0%	0%	0%	22%	0%	1%	0%	1%	1%	0%	0%	0%
4%	1%	0%	1%	0%	1%	0%	0%	0%	2%	2%	0%	9%	0%	7%
0%	9%	1%	0%	0%	0%	0%	1%	0%	1%	1%	2%	1%	0%	1%
1%	0%	0%	0%	1%	0%	6%	2%	2%	7%	7%	3%	1%	0%	2%
0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	1%	2%	2%	0%	5%
0%	10%	0%	0%	4%	0%	1%	0%	4%	0%	2%	0%	0%	0%	2%
0%	0%	0%	0%	1%	93%	1%	10%	0%	4%	1%	0%	10%	5%	26%
0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
0%	1%	0%	1%	16%	0%	0%	0%	0%	2%	28%	16%	2%	0%	0%
0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	84%	4%
0%	0%	0%	0%	0%	0%	61%	86%	92%	0%	1%	0%	0%	2%	3%
0%	1%	0%	0%	2%	3%	1%	0%	1%	1%	0%	1%	1%	1%	14%
7%	3%	1%	94%	48%	1%	0%	0%	0%	26%	6%	61%	3%	7%	1%
0%	3%	5%	2%	27%	2%	0%	0%	0%	5%	1%	7%	23%	0%	0%
0%	1%	2%	0%	0%	0%	0%	0%	0%	4%	1%	0%	4%	0%	3%
0%	1%	0%	0%	0%	1%	0%	0%	0%	1%	1%	0%	1%	0%	13%
88%	0%	0%	0%	0%	0%	0%	0%	0%	40%	36%	0%	33%	0%	18%
100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
CDR2							CDR3							

图2(续)

V_LK1 阈值分析: 第一部分

- 不报告低于 10%使用率的单独氨基酸
- 注意每个 CDRs 中至少一个位置总和和使用率低于 75%
 - 为了提供较大覆盖率可以降低阈值百分比使用率

#序列	30	31	32	33	34	35	36	46	47	48	49	50
阈值百分比使用率	383	383	383	383	383	383	383	383	383	383	383	383
A	10							N		L		
C					43%							55%
D												13%
E												
F												
G												
H												
I												
K										95%		10%
L				95%				82%	97%			
M												
N		34%			42%							
P												
Q												
R	12%											
S	54%	39%										
T												
V												
W			17%			100%						
Y			52%				92%				88%	
呈现的%	66%	73%	69%	95%	85%	100%	92%	82%	97%	95%	88%	76%
是否接受?	否	否	否	否								否
总和百分比阈值	80	CDR1					CDR2					否

图 3

[illegible]

图 3(续)

V_LK 阈值分析: 第二部分

- 不报告低于 5% 使用率的单独氨基酸
- 结果全部位置总和和使用率均为至少 75%
— 这可提供更大的多样性覆盖率

#序列	30	31	32	33	34	35	36	46	47	48	49	50
阈值百分比使用率	5	383	383	383	383	383	383	383	383	383	383	383
A					43%			N		L		55%
C												
D			8%									13%
E												
F			8%				7%					
G		8%			7%							9%
H												
I										95%		
K												10%
L				95%				82%	97%			
M												
N		8%	34%		42%							
P												
Q												
R		12%	5%					6%				
S		54%	39%	6%							7%	
T			6%									
V												
W			17%			100%						
Y			52%				92%				88%	
呈现的%		83%	85%	90%	95%	92%	99%	88%	97%	95%	95%	88%
是否接受?												
总和百分比阈值	75	CDR1					CDR2					

图 4

51	52	53	54	55	89	90	91	92	93	94	95	96
383	383	383	383	383	383	383	383	383	383	383	383	383
					Q	G						
90%				5%			5%					
								11%				
				22%						9%		7%
							7%	7%				
				6%								
					10%					10%		26%
		16%	93%					28%	16%			
				61%	86%	92%						14%
	94%	48%					26%	6%	61%		7%	
		7%							7%	23%		
												13%
							40%	36%		33%		18%
90%	94%	91%	93%	95%	97%	92%	79%	88%	84%	75%	91%	78%
CDR2				CDR3								

图 4(续)

合成轻链 CDR1 多样性

- 为了覆盖多样性，必须制备 128 种组合寡核苷酸来编码预期的多样性
- 或制备 16 种简并组合寡核苷酸

注意：碱基无需是等摩尔的，可以通过调整碱基使氨基酸使用率产生偏向以反映我们的分析中发现的频率，甚至可以包括频率表中不包括的残基。

简并设计

	S		A		M		T		G		K		C	T	C	G		A		T	G	T	C	T	#寡核苷酸										
	S		A		M		T		G		K					A		C																	
	G	R	T	T	A	G	T	T	T	W	T	G				A	T	C																	
	2				2				2				1		1		2				1		1		16										

K1_01	S	G	T	T	A	M	T	T	T	G	K	C	T	C	G	C	A	T	G	G	T	C	T	
K1_02	A	R	T	T	A	M	T	T	T	G	K	C	T	C	G	C	A	T	G	G	T	C	T	
K1_03	S	G	T	T	M	G	T	T	T	G	K	C	T	C	G	C	A	T	G	G	T	C	T	
K1_04	A	R	T	T	M	G	T	T	T	G	K	C	T	C	G	C	A	T	G	G	T	C	T	
K1_05	S	G	T	T	A	M	T	T	T	W	T	C	T	C	G	C	A	T	G	G	T	C	T	
K1_06	A	R	T	T	A	M	T	T	T	W	T	C	T	C	G	C	A	T	G	G	T	C	T	
K1_07	S	G	T	T	M	G	T	T	T	W	T	C	T	C	G	C	A	T	G	G	T	C	T	
K1_08	A	R	T	T	M	G	T	T	T	W	T	C	T	C	G	C	A	T	G	G	T	C	T	
K1_09	S	G	T	T	A	M	T	T	T	G	K	C	T	C	A	T	C	T	G	G	T	C	T	
K1_10	A	R	T	T	A	M	T	T	T	G	K	C	T	C	A	T	C	T	G	G	T	C	T	
K1_11	S	G	T	T	M	G	T	T	T	G	K	C	T	C	A	T	C	T	G	G	T	C	T	
K1_12	A	R	T	T	M	G	T	T	T	G	K	C	T	C	A	T	C	T	G	G	T	C	T	
K1_13	S	G	T	T	A	M	T	T	T	W	T	C	T	C	A	T	C	T	G	G	T	C	T	
K1_14	A	R	T	T	A	M	T	T	T	W	T	C	T	C	A	T	C	T	G	G	T	C	T	
K1_15	S	G	T	T	M	G	T	T	T	W	T	C	T	C	A	T	C	T	G	G	T	C	T	
K1_16	A	R	T	T	M	G	T	T	T	W	T	C	T	C	A	T	C	T	G	G	T	C	T	

图 5

V_H3 重链合成文库多样性长度 10

• CDR3 含有 3.3 × 10⁵ 的总体多样性

• 仅需要 96 种简并寡核苷酸

• 第 97 位省略了 SWY (各 4%)

• 第 100A 位省略了 G (6%)

VH3 长度 10	93	94	95	96	97	98	99	100	100A	101
#序列	68	68	68	68	68	68	67	68	68	68
阈值百分比使用率	10%	10%	5%	5%	4%	5%	5%	10%	3%	5%
A	91%			6%		9%		32%		
C										
D			25%	7%	6%	7%	9%			85%
E			12%							
F									32%	
G			10%	13%	13%	31%	19%	22%		
H							9%			
I									3%	
K		22%	6%		21%					
L				6%	6%				7%	
M			21%						38%	
N										
P					6%					
Q					4%					
R		63%		31%	4%	6%	6%			
S						12%		16%		
T										
V			6%	12%	9%					
W										
Y						12%	36%	12%		
%覆盖率	91%	85%	79%	75%	69%	76%	79%	82%	81%	85%
CDR3										
# 简并密码子	1	2	2	2	3	2	1	2	1	1
96 个寡核苷酸										
多样性	1	2	6	6	8	6	6	4	4	1
3.3E+0 5 总体 CDR3 多样性										

图 6

合成 V_H3 重链 CDR3 长度 10 多样性

名称	序列
H3 3 10 001a	ACCGCAGTTTACTATTGCGCACGTVTGSKTAAGKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 002a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTAAGKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 003a	ACCGCAGTTTACTATTGCGCACGTVTGSKTAAGKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 004a	ACCGCAGTTTACTATTGCGCAAAGGRTGSKTAAGKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 005a	ACCGCAGTTTACTATTGCGCACGTVTGSKTAAGKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 006a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTAAGKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 007a	ACCGCAGTTTACTATTGCGCACGTVTGSKTAAGKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 008a	ACCGCAGTTTACTATTGCGCAAAGGRTGSKTAAGKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 009a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 010a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 011a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 012a	ACCGCAGTTTACTATTGCGCAAAGGRTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 013a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 014a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 015a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 016a	ACCGCAGTTTACTATTGCGCAAAGGRTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 017a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 018a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 019a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 020a	ACCGCAGTTTACTATTGCGCAAAGGRTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 021a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 022a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 023a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 024a	ACCGCAGTTTACTATTGCGCAAAGGRTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 025a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 026a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 027a	ACCGCAGTTTACTATTGCGCACGTVTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC
H3 3 10 028a	ACCGCAGTTTACTATTGCGCAAAGGRTGSKTGTGDKMTBRTGSTWTKGATTACTGGGGTCAGGGGCAC

图 7

H3 3 10 029a	ACCGCAGTTTACTATTGCGCACGVTGGMTAAGSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 030a	ACCGCAGTTTACTATTGCGCAAAGVTGGMTAAGSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 031a	ACCGCAGTTTACTATTGCGCACGTRWGMTAAGSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 032a	ACCGCAGTTTACTATTGCGCAAAGRWGMTAAGSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 033a	ACCGCAGTTTACTATTGCGCACGVTGSKTGDTSGBRTGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 034a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTGSTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 035a	ACCGCAGTTTACTATTGCGCACGTRWSKTGDTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 036a	ACCGCAGTTTACTATTGCGCAAAGRWSKTGDTSGBRTGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 037a	ACCGCAGTTTACTATTGCGCACGVTGGMTGDTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 038a	ACCGCAGTTTACTATTGCGCAAAGVTGGMTGDTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 039a	ACCGCAGTTTACTATTGCTCACGTGRWGMTGDTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 040a	ACCGCAGTTTACTATTGCGCAAAGRWGMTGDTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 041a	ACCGCAGTTTACTATTGCGCACGIVTGSKTCNTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 042a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTCNTSGRBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 043a	ACCGCAGTTTACTATTGCGCACGTRWSKTCNTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 044a	ACCGCAGTTTACTATTGCGCAAAGRWSKTCNTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 045a	ACCGCAGTTTACTATTGCGCACGIVTGMTCNTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 046a	ACCGCGTTTACTATTGCGCAAAGCTGGMTCNTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 047a	ACCGCAGTTTACTATTGCGCACGTRWGMTCNTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 048a	ACCGCAGTTTACTATTGCGCAAAGRWGMTCNTSGTBRGTSWTGKATTACTGGGGTCAGGGCAC
H3 3 10 049a	ACCGCAGTTTACTATTGCGCACGVTGSKTAAGKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 050a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTAAGKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 051a	ACCGCAGTTTACTATTGCGCACGTRWSKTAAGKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 052a	ACCGCAGTTTACTATTGCGCAAAGRWSKTAAGKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 053a	ACCGCAGTTTACTATTGCGCACGVTGGMTAAGKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 054a	ACCGCAGTTTACTATTGCGCAAAGVTGGMTAAGKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 055a	ACCGCAGTTTACTATTGCGCACGTRWGMTAAGKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 056a	ACCGCAGTTTACTATTGCGCAAAGRWGMTAAGKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 057a	ACCGCAGTTTACTATTGCGCACGIVTGSKTGDTKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC
H3 3 10 058a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTGDTKMTBRTTMTWTGKATTACTGGGGTCAGGGCAC

图 7(续)

H3 3 10 059a	ACCGCAGTTTACTATTGCGCACGTGRWSKTDGDKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 060a	ACCGCAGTTTACTATTGCGCAAGGRWSKTDGDKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 061a	ACCGCAGTTTACTATTGCGCACGTGTGMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 062a	ACCGCAGTTTACTATTGCGCAAGVTGMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 063a	ACCGCAGTTTACTATTGCGCACGTGRWMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 064a	ACCGCAGTTTACTATTGCGCAAGGRWMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 065a	ACCGCAGTTTACTATTGCGCACGTGTGSKTCNTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 066a	ACCGCAGTTTACTATTGCGCAAGVTGSKTCNTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 067a	ACCGCAGTTTACTATTGCGCACGTGRWSKTCNTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 068a	ACCGCAGTTTACTATTGCGCAAGGRWSKTCNTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 069a	ACCGCAGTTTACTATTGCGCACGTGTGMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 070a	ACCGCAGTTTACTATTGCGCAAGVTGMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 071a	ACCGCAGTTTACTATTGCGCACGTGRWMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 072a	ACCGCAGTTTACTATTGCGCAAGGRWMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 073a	ACCGCAGTTTACTATTGCGCACGTGTGSKTAAGSGTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 074a	ACCGCAGTTTACTATTGCGCAAGVTGSKTAAGSGTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 075a	ACCGCAGTTTACTATTGCGCACGTGRWSKTAAGSGTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 076a	ACCGCAGTTTACTATTGCGCAAGGRWSKTAAGSGTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 077a	ACCGCAGTTTACTATTGCGCACGTGTGMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 078a	ACCGCAGTTTACTATTGCGCAAGVTGMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 079a	ACCGCAGTTTACTATTGCGCACGTGRWMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 080a	ACCGCAGTTTACTATTGCGCAAGGRWMDTGDTKMTBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 081a	ACCGCAGTTTACTATTGCGCACGTGTGSKTDGTSGBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 082a	ACCGCAGTTTACTATTGCGCAAGVTGSKTDGTSGBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 083a	ACCGCAGTTTACTATTGCGCACGTGRWSKTDGTSGBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 084a	ACCGCAGTTTACTATTGCGCAAGGRWSKTDGTSGBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 085a	ACCGCAGTTTACTATTGCGCACGTGTGMDTGDTSGBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 086a	ACCGCAGTTTACTATTGCGCAAGVTGMDTGDTSGBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 087a	ACCGCAGTTTACTATTGCGCACGTGRWMDTGDTSGBRTTMTWTKGATTTACTGGGGTCAGGGCAC
H3 3 10 088a	ACCGCAGTTTACTATTGCGCAAGGRWMDTGDTSGBRTTMTWTKGATTTACTGGGGTCAGGGCAC

图 7(续)

H3 3 10 089a	ACCGCAGTTTACTATTGCGCACGTVTGSKTCNTSGTBRTTMTWTKGATTACTGGGGTCAGGGCAC
H3 3 10 090a	ACCGCAGTTTACTATTGCGCAAAGVTGSKTCNTSGRBRTTMTWTKGATTACTGGGGTCAGGGCAC
H3 3 10 091a	ACCGCAGTTTACTATTGCGCACGTCGRWSKTCNTSGTBRTTMTWTKGATTACTGGGGTCAGGGCAC
H3 3 10 092a	ACCGCAGTTTACTATTGCGCAAAGGRWSKTCNTSGTBRTTMTWTKGATTACTGGGGTCAGGGCAC
H3 3 10 093a	ACCGCAGTTTACTATTGCGCACGTVTGGMTTCNTSGTBRTTMTWTKGATTACTGGGGTCAGGGCAC
H3 3 10 094a	ACCGCAGTTTACTATTGCGCAAAGVTGGMTTCNTSGTBRTTMTWTKGATTACTGGGGTCAGGGCAC
H3 3 10 095a	ACCGCAGTTTACTATTGCGCACGTCGRWGMTTCNTSGTBRTTMTWTKGATTACTGGGGTCAGGGCAC
H3 3 10 096a	ACCGCAGTTTACTATTGCGCAAAGGRWGMTTCNTSGTBRTTMTWTKGATTACTGGGGTCAGGGCAC

图 7(续)

确定重链来源

— V_H3:3-09 最相似

	33	40	50	60	70	80	90
D2E7	(33)	PAMAEVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSAITW			
VH3_3-09	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-20	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-43	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-15	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-49	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-72	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-73	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-23	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-53	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-66	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-d	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-64	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-07	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-30	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-30.5	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-33	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-30.3	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-11	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-21	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-48	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-13	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			
VH3_3-74	(1)	EVQLVESGGGLVQPGKSLRLSCAASGFTFE	DDYAMHW	MRQAPGKGLEWVSGISW			

图 8

100	110	120	130
NSGHIDYADSV	EGREFTISRDN	AKNSLYLQ	NSISRAEDTAMVYCAKVS
NSGSIYADSV	KGREFTISRDN	AKNSLYLQ	NSISRAEDTAMVYCAKD
NGGSTYADSV	KGREFTISRDN	AKNSLYLQ	NSISRAEDTAMVYCAR
DGGSTYADSV	KGREFTISRDN	SKNSLYLQ	NSISRAEDTAMVYCAKD
DGGIYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCTI
YGGIIEYTA	SVKGRFTISR	DGSKSIA	MYLQNSIKIEDTAMVYCTIR
SYIIEYTA	SVKGRFTISR	DGSKNSLYLQ	NSISRAEDTAMVYCAR
NSYATAYAA	SVKGRFTISR	DDSKNTA	MYLQNSIKIEDTAMVYCTIR
SGGSTYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAK
CGGSTYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAR
CGGSTYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAR
CGGSTYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAK
NGGSTYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAR
DGSEKMYV	DSVKGREFTISR	DNKNSLYLQ	NSISRAEDTAMVYCAR
DGSKMYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAK
DGSKMYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAK
DGSKMYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAR
DGSKMYADSV	KGREFTISRDN	SKNTLYLQ	NSISRAEDTAMVYCAR
SSSYIYADSV	KGREFTISRDN	AKNSLYLQ	NSISRAEDTAMVYCAR
SSSYIYADSV	KGREFTISRDN	AKNSLYLQ	NSISRAEDTAMVYCAR
SSSYIYADSV	KGREFTISRDN	AKNSLYLQ	NSISRAEDTAMVYCAR
AG-DIYIPG	SVKGRFTISR	ENAKNSLYLQ	NSISRAEDTAMVYCAR
DGSSISYADSV	KGREFTISRDN	AKNTLYLQ	NSISRAEDTAMVYCAR

图8(续)

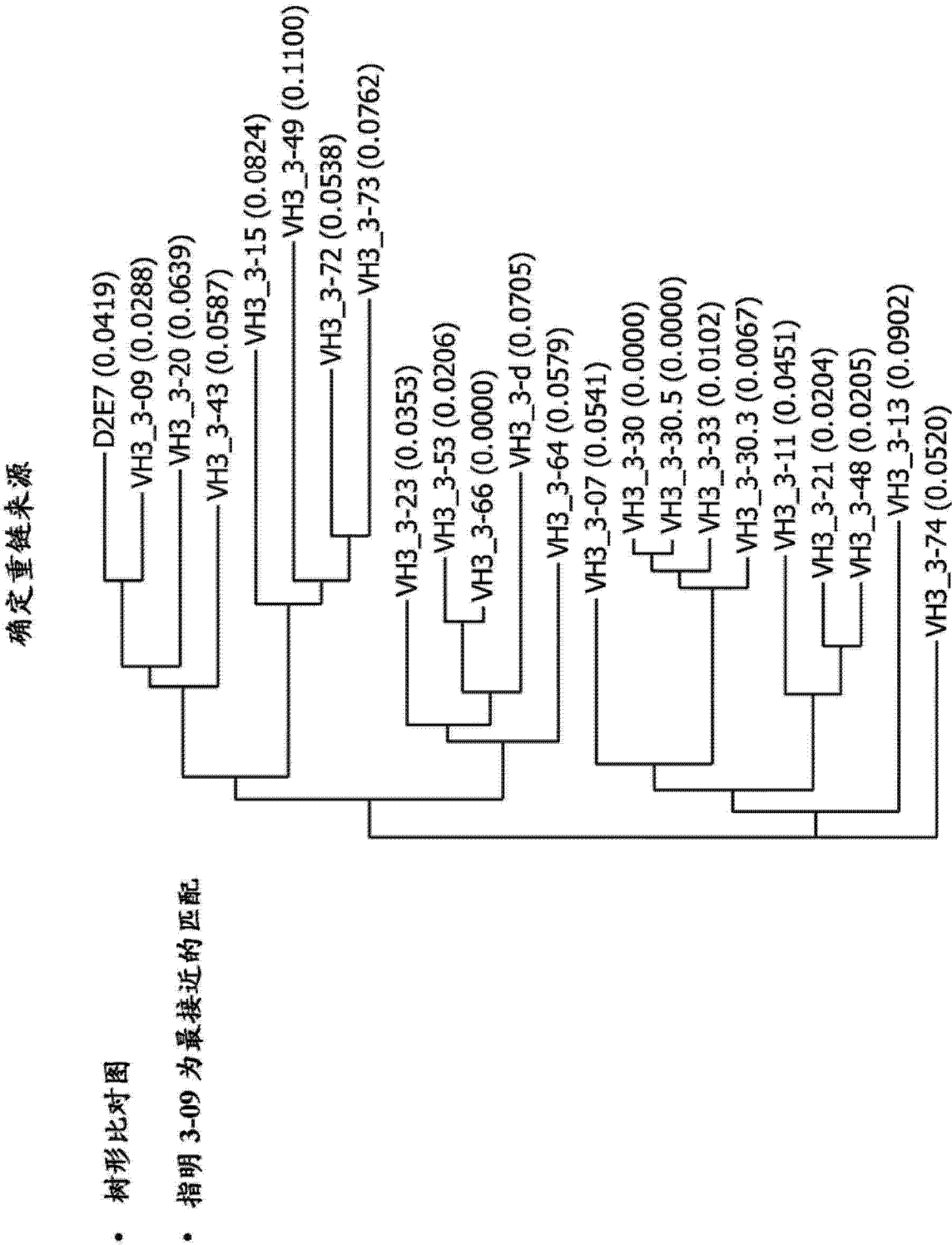


图 9

确定轻链来源

- A20 最相似

	172	180	190	200	210	220
D2E7(172)	SDIQMTQSPSSISASV	GDRVTITICRASQGI	RNYI	AWYQOKPGKAPKLI	YAAASTI	QS
VK1_A20	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SNYI	AWYQOKPGKAPKLI	YAAASTI
VK1_O18	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SNYI	AWYQOKPGKAPKLI	YAAASTI
VK1_O8	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SNYI	AWYQOKPGKAPKLI	YAAASTI
VK1_O8	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SNYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L23	(1)	-AIRMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L24	(1)	-VIMMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L9	(1)	-AIRMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_O12	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_O2	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L12	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L18	(1)	-AIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L4	(1)	-AIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L8	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_A30	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	RNDI	AWYQOKPGKAPKLI	YAAASTI
VK1_L11	(1)	-AIQMTQSPSSISASV	GDRVTITICRASQGI	RNDI	AWYQOKPGKAPKLI	YAAASTI
VK1_L14	(1)	-NIQMTQSPSSISASV	GDRVTITICRASQGI	SNYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L1	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L15	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L19	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI
VK1_L5	(1)	-DIQMTQSPSSISASV	GDRVTITICRASQGI	SSYI	AWYQOKPGKAPKLI	YAAASTI

图 10

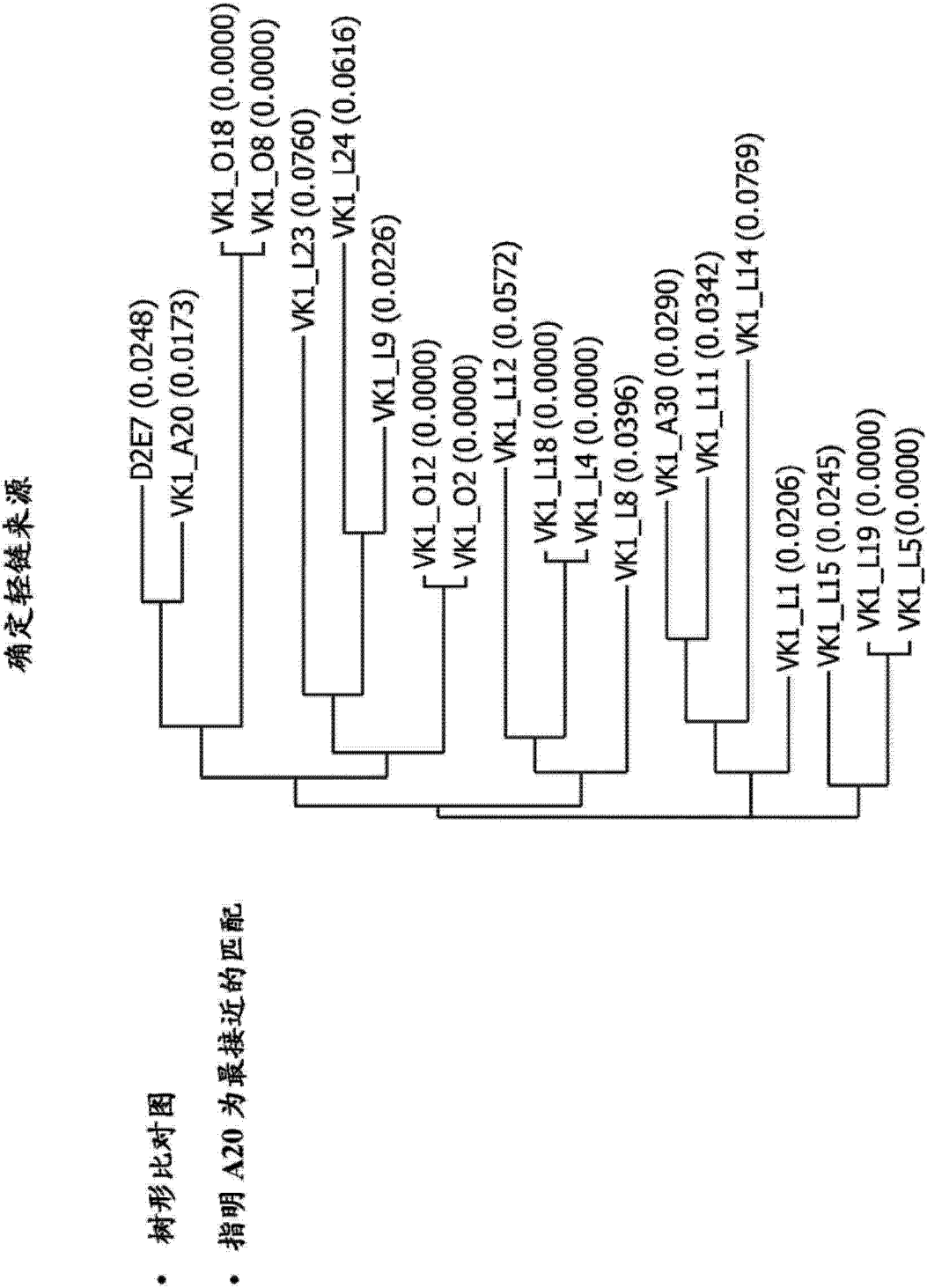


图 11

54	55	89	90	91	92	93	94	95	96	
383	383	383	383	383	383	383	383	383	383	
		Q	G							
				5%			5%			
					11%					
	22%						9%		7%	
	6%			7%	7%					
93%		10%					10%		26%	
					28%	16%				
								84%		
	61%	86%	92%							
									14%	
				26%	6%	61%		7%		
						7%	23%			
									13%	
				40%	36%		33%		18%	
93%	90%	97%	92%	78%	88%	84%	80%	91%	78%	
CDR2		CDR3								
1	1	1	1	1	2	1	3	1	2	总寡核苷酸
12个寡核苷酸										30
1	2	1	1	4	4	3	6	1	7	组合多样性
2016个CDR多样性										9.29E+06
L	Q	Q	R	Y	N	R	A	P	Y	

图 12(续)

频率分析 V_H3 CDR1 和 CDR2

• 按位置的绝对使用率

#序列	30	31	32	33	34	35	47	48	49	50
0	181	181	181	181	181	181	181	181	181	181
阈值百分比使用率										
A	0%	1%	0%	44%	0%	0%	0%	1%	40%	12%
C	0%	1%	2%	1%	0%	0%	5%	0%	0%	0%
D	4%	11%	1%	3%	0%	3%	0%	0%	0%	1%
E	1%	0%	0%	11%	0%	0%	0%	0%	1%	0%
F	0%	1%	12%	0%	0%	1%	0%	0%	0%	4%
G	1%	2%	0%	18%	0%	1%	0%	0%	1%	13%
H	0%	0%	4%	0%	0%	34%	0%	0%	0%	2%
I	1%	0%	0%	0%	6%	0%	0%	1%	0%	2%
K	4%	1%	0%	0%	0%	0%	0%	0%	0%	0%
L	0%	0%	1%	1%	2%	0%	0%	1%	0%	11%
M	0%	0%	0%	0%	89%	0%	0%	0%	0%	1%
N	2%	24%	1%	0%	0%	23%	0%	0%	0%	3%
P	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Q	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
R	10%	2%	0%	0%	1%	0%	1%	0%	0%	1%
S	76%	40%	3%	11%	1%	34%	0%	0%	58%	16%
T	1%	17%	2%	1%	0%	3%	0%	0%	0%	1%
V	0%	1%	1%	2%	3%	0%	0%	98%	0%	18%
W	0%	0%	0%	4%	0%	0%	94%	0%	0%	2%
Y	0%	0%	76%	4%	0%	2%	0%	0%	0%	14%
呈现的% 总和百分比阈值	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	CDR1									

图 13

51	52	52A	53	54	55	56	57	58
181	180	181	180	181	180	180	180	181
0%	0%	4%	1%	1%	5%	2%	0%	0%
0%	0%	0%	0%	0%	0%	0%	0%	0%
0%	0%	6%	32%	0%	9%	1%	0%	1%
0%	0%	2%	0%	1%	3%	2%	3%	1%
2%	0%	2%	0%	0%	0%	2%	0%	4%
0%	6%	19%	11%	78%	31%	0%	0%	4%
0%	0%	1%	1%	0%	0%	0%	0%	4%
88%	0%	1%	0%	0%	1%	4%	23%	0%
1%	7%	2%	0%	0%	0%	2%	28%	1%
3%	1%	0%	1%	0%	0%	2%	4%	0%
2%	0%	0%	0%	0%	0%	0%	0%	0%
0%	7%	6%	14%	1%	3%	28%	0%	5%
0%	0%	1%	0%	0%	0%	0%	0%	0%
0%	0%	4%	0%	0%	0%	1%	1%	0%
2%	5%	4%	3%	0%	7%	3%	2%	2%
1%	61%	20%	36%	20%	34%	22%	1%	9%
2%	3%	2%	2%	0%	5%	22%	37%	1%
1%	1%	0%	0%	0%	1%	1%	1%	0%
0%	7%	6%	0%	0%	0%	0%	0%	2%
0%	3%	23%	0%	0%	1%	10%	0%	66%
100%	100%	100%	100%	100%	100%	100%	100%	100%
CDR2								

图 13(续)

CDR1 和 CDR2 阈值分析: 第一部分

- 不报告低于 10%使用率的单独氨基酸
- 在 CDR2 中, 第 52、52A、55 和 58 位的总和氨基酸使用率低于 75%
 - 为了提供较大的覆盖率, 可以降低阈值百分比使用率

#序列	30	31	32	33	34	35	47	48	49	50
阈值百分比使用率	181	181	181	181	181	181	181	181	181	181
A	10			44%					40%	12%
C										
D		11%								
E				11%						
F			12%							
G				18%						13%
H						34%				
I										
K										
L										11%
M					89%					
N		24%				23%				
P										
Q										
R	10%									
S	76%	40%		11%		34%			58%	16%
T		17%								
V								98%		18%
W							94%			
Y			76%							14%
呈现的%	86%	93%	88%	85%	89%	91%	94%	98%	98%	85%
是否接受?										
总和百分比阈值	75	CDR1								

图 14

CDR1 和 CDR2 阈值分析: 第二部分

- 将百分比阈值变为 5%，使低于 5% 使用率的个体不被报告
- 这样全部位置的总和氨基酸使用率均超过 75%

		30	31	32	33	34	35	47	48	49	50	51
#序列		181	181	181	181	181	181	181	181	181	181	181
阈值百分比使用率	5											
A					44%					40%	12%	
C												
D			11%									
E					11%							
F				12%								
G					18%						13%	
H							34%					
I						6%						88%
K												
L												
M						89%						
N			24%				23%					
P												
Q												
R		10%										
S		76%	40%		11%		34%			58%	16%	
T			17%									
V									98%		18%	
W								94%				
Y				76%							14%	
呈现的%		86%	93%	88%	85%	94%	91%	99%	98%	98%	85%	88%
是否接受?												
总和百分比阈值	75	CDR1										
降低阈值可改善覆盖率												

图 15

52	52A	53	54	55	56	57	58
180	181	180	181	180	180	180	181
				5%			
	6%	32%		9%			
6%	19%	11%	78%	31%			
7%						23%	
						28%	
7%	6%	14%			28%		5%
5%				7%			
61%	20%	36%	20%	34%	22%		9%
				5%	22%	37%	
7%	6%						
	29%				10%		66%
93%	79%	93%	98%	91%	81%	88%	80%
CDR2							

图 15(续)

V_H3 重链合成文库多样性

- 组合的 CDR2 & CDR2 含有 1.5×10^8 总体多样性, 呈现>80%的位置覆盖率
- 需要 28 种简并寡核苷酸

VH3 κ 配对	30	31	32	33	34	35	47	48	49	50	51	52
#序列	181	181	181	181	181	181	181	181	181	181	181	180
A				44%					40%	12%		
C												
D												
E				11%								
F			12%									
G				18%						13%		
H						34%					88%	
I												7%
K												
L										11%		
M					89%							
N		24%				23%						7%
P												
Q												
R	10%											5%
S	76%	40%		11%		34%			58%	16%		61%
T		17%										
V												
W							94%					7%
Y			76%							14%		
%覆盖率	86%	82%	88%	85%	89%	91%	94%	98%	98%	85%	88%	87%
CDR1												
# 简并密码子	1	1	1	2	1	2	1	1	1	3	1	2
4 个寡核苷酸												
多样性	2	3	2	4	1	3	1	1	2	6	1	4
144 个 CDR1 多样性												
CDR2												
# 简并密码子	1	1	1	2	1	2	1	1	1	3	1	2
24 个寡核苷酸												
多样性	2	3	2	4	1	3	1	1	2	6	1	4
144 个 CDR1 多样性												
D2E7 匹配	D	D	Y	A	M	H	W	V	S	A	I	T
蓝色高亮残基是 D2E7 和合成库中均存在的												

图 16

52A	53	54	55	56	57	58
181	180	181	180	180	180	181
4%			5%			
0%						
6%	32%		9%			
19%	11%	78%	31%			
					23%	
					28%	
6%	14%		3%	28%		5%
			7%			
20%	36%	20%	34%	22%		9%
			5%	22%	37%	0%
6%						
23%				10%		66%
83%	93%	98%	94%	81%	88%	80%
2	1	1	2	1	1	1
总寡核苷酸						
28						
8	4	2	7	4	3	4
组合多样性						
1.03E+06 总体 CDR2 多样性						
1.49E+08						
W	N	S	G	H	I	D

图 16(续)

V_H3 重链合成文库多样性

- CDR3 含有 7.5×10^9 总体多样性, 呈现>80%的位置覆盖率
- 需要 384 种简并寡核苷酸

VH3 长度 I3		93	94	95	96	97	98	99	100	100A	100B	100C	100D	101
#序列		114	114	114	114	114	114	114	114	114	114	114	114	114
A		95%		14%		18%	5%	2%		9%	6%	17%		
C														
D				23%		11%	11%				5%			79%
E						2%								
F						1%		5%	4%		4%		43%	
G				23%	14%	12%	14%	14%	25%	7%	13%	7%		9%
H					3%							3%		
I					4%		4%	7%					1%	
K			56%			5%	12%							
L					17%			4%	4%	11%			24%	
M													8%	
N				8%	5%			4%		1%		2%		
P					9%				5%	8%	6%	13%		
Q														
R			26%		11%	5%			16%	6%	13%			
S				7%	11%	9%	10%	18%	11%	14%	15%	20%		
T				4%	0%		12%	9%		8%	6%	4%		
V					8%	6%	6%	4%	5%	2%	4%			
W								7%	7%	15%		4%		
Y						12%	6%	12%	4%	10%	13%	19%		
%覆盖率		95%	82%	79%	82%	83%	82%	87%	80%	90%	85%	90%	76%	88%
CDR3														
# 简并密码子	1	1	1	1	2	2	3	2	2	2	2	2	1	1
384 寡核苷酸														
多样性	1	2	6	10	10	10	9	11	8	11	10	9	4	2
7.5E+09 总体 CDR3 多样性														
D2E7 匹配	A	K	V	S	S	Y	L	S	T	A	S	S	L	D

蓝色高亮的残基是在 D2E7 和合成的所有组分中均存在的那些

图 17

省略的位置

• 残基 95 E (6%出现率)

• 残基 98 P (5%出现率)

• 残基 99 E (6%出现率)

• 残基 100D G (11%出现率)

添加的位置

• 用阴影小格和粗体文字标明

图 17(续)

抗洋地黄毒苷抗体

IG 人轻链可变区

重链可变区

抗半抗原主题分析

SYELTQPPSASGTPGQRVITISCSGSTSNIGSNYYWYQHLPGTAPKLLIFRNSQRPS

GVPDFSGSKSGTSASLAISGLRSEDEADYYCAAWDDSLSGVVFGGTKLTVLG

EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPKGLEWVSAISGSG

GSTYYADSVKGRFTISRDN SKNTLYLQMNSLRAEDTAVYYCAKDKGSGWYWGQGT

LVTVSS

图 18

确定重链和轻链的来源

- VL-1g 为最相似的轻链
- VH3-23 为最相似的重链

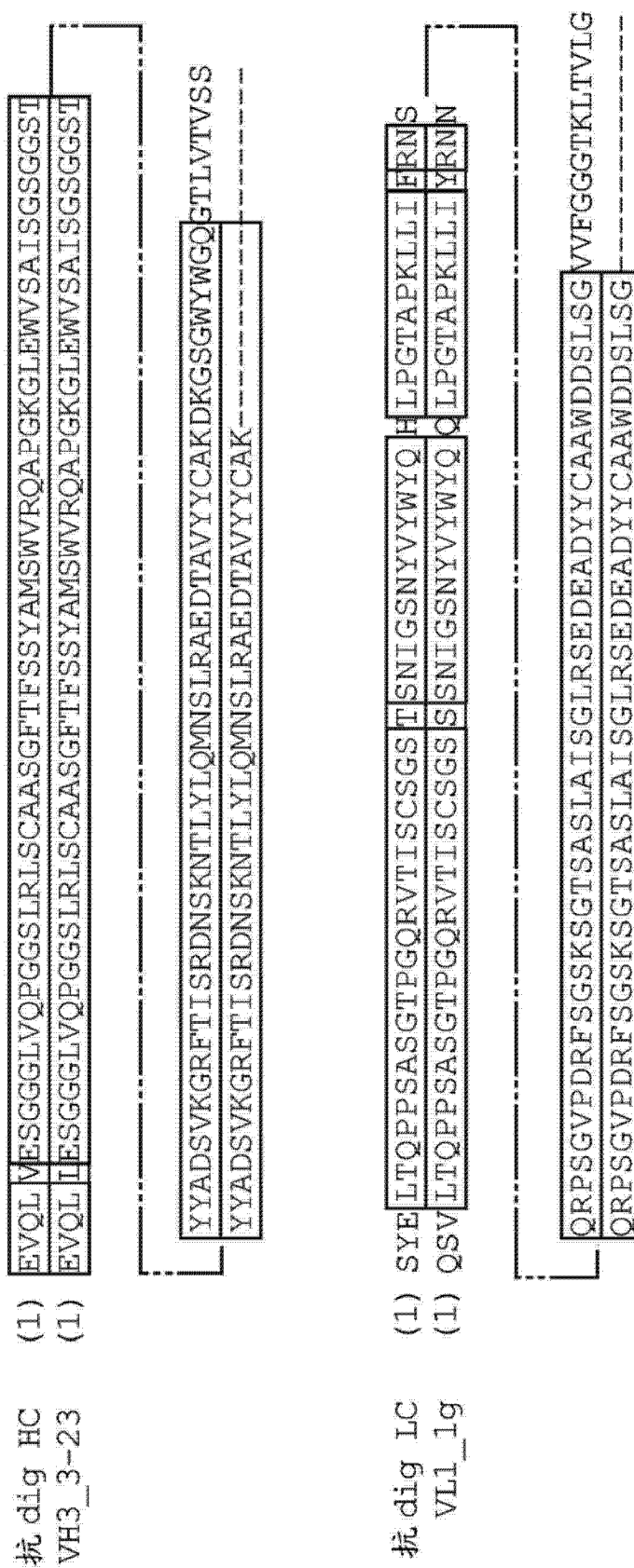


图 19

λ 长度匹配的 VL1 框架的半抗原分析

半抗原分析	27C	29	30	31	32	33	34	35	36	46	47	48	49	50	51
#序列	166	166	166	166	166	166	166	166	166	166	166	166	166	166	166
阈值百分比使用率	6														
A															
C															
D														24%	13%
E														12%	
F														7%	
G		92%													
H															
I	95%											96%			
K															
L										94%	95%				
M															
N			36%	87%			40%							11%	82%
P															
Q															
R														13%	
S			39%		7%		39%							18%	
T					29%										
V						95%									
W								100%							
Y					45%		13%		96%				87%		
呈现的%	95%	92%	75%	87%	81%	95%	92%	100%	96%	94%	95%	96%	87%	85%	95%
是否接受?															
总和百分比阈值	75	CDR1					CDR2								
	1	1	1	1	2	1	2	1	1	1	1	1	1	3	1
					4	寡核苷酸								3	寡核苷酸
		1	2	1	2		2	1	1	1	1	1	1	5	2
					8多样性									40多样性	

图 20

H3 长度 8 的半抗原分析

未配对的 VH3 分析		93	94	95	96	97	98	99	101
#序列		32	32	32	32	32	32	32	32
阈值百分比使用率		6.25							
A		91%		6%	6%	9%			9%
C									
D				47%				6%	56%
E					16%				
F				6%			6%	22%	
G				22%	6%	28%	16%		9%
H					9%				
I									
K			31%						6%
L		6%			6%	6%	9%	9%	
M						16%		16%	9%
N									
P								6%	
Q					13%		6%		
R			44%		19%				
S			9%				9%		
T							19%		
V					13%			13%	
W				6%			6%		
Y						22%	13%	6%	
呈现的%			84%	88%	88%	81%	84%	78%	91%
是否接受?									
总和百分比阈值		75	CDR3						

图 21

干扰素的第 32-38 个氨基酸

干扰素 α 实例		32	33	34	35	36	37	38
#序列		11	11	11	11	11	11	11
阈值百分比使用率	9							
A							64%	
C								
D			9%					
E								
F								
G			64%					
H				9%				
I								
K								
L		100%						91%
M								9%
N			9%	91%				
P								
Q								
R			9%		100%	100%		
S			9%					
T							36%	
V								
W								
Y								
%覆盖率		100%	100%	100%	100%	100%	100%	100%
#简并合成		1	2	1	1	1	1	1
2								
多样性		1	5	2	1	1	2	2
40								

1. 比对 11 个 IFN- α 序列的第 32-38 个氨基酸
2. 确定残基频率使用率

(30)	30	43
P01563(30)	HSIGSRRTIMLLAQ	
P05013(30)	HSIGHRRIMLLAQ	
P01569(30)	HSISNRRIMIMAAQ	
P32881(30)	HSIGNRRALILIAQ	
P01570(30)	HSINRRIMIMAAQ	
P05015(30)	HSIGNRRALILIAQ	
P01568(30)	HSIGNRRALILIAQ	
P01566(30)	HSIGNRRALILIGQ	
P01571(30)	HSIGNRRALILIAQ	
P01567(30)	HSIGNRRALILIAQ	
P05014(30)	HSIGNRRALILIAQ	
共有序列 (30)HSIGNRRALILIAQ		

图 22

146

按侧链化学分类的氨基酸

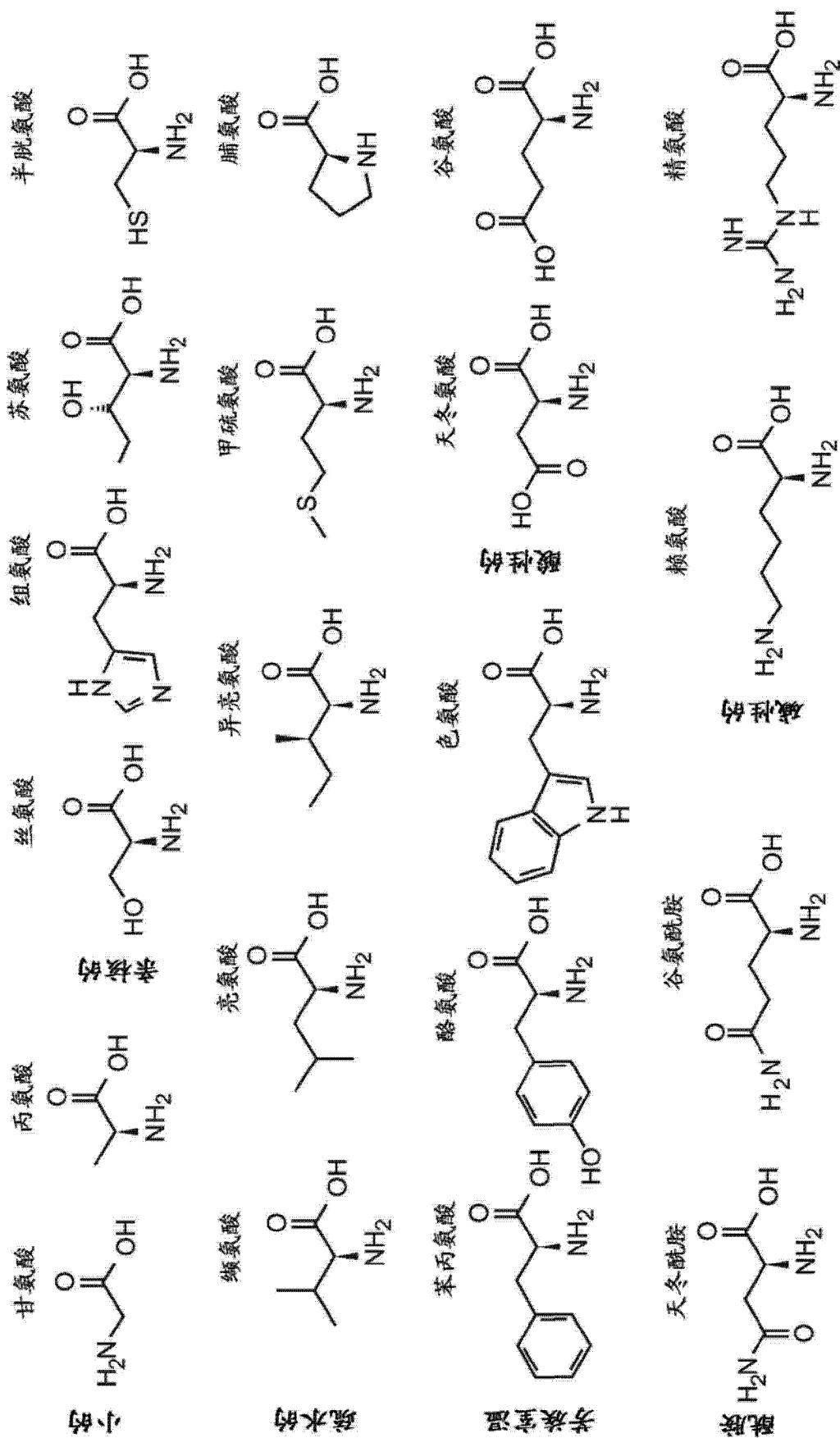


图 24

包含具有化学探针的多样性的 CDR

VH3 长度 11 化学探针	93	94	95	96	97	98	99	100	100A	100B	101
A	93%		8%	4%	16%	9%	8%	8%	10%		
C											
D			26%	4%	1%	3%	6%	0%	3%		86%
E											
F										48%	
G											
H			3%	1%	0%	0%	0%	11%	0%		
I										1%	
K											
L			4%	6%	3%	5%	5%	1%	1%	15%	
M										18%	
N											
P			0%	6%	13%	0%	4%	3%	8%		
Q			0%	1%	1%	3%	1%	1%	0%		
R		59%	14%	16%	11%	13%	8%	5%	5%		
S			1%	16%	10%	11%	20%	10%	10%		
T											
V											
W											
Y			1%	5%	8%	16%	9%	19%	24%		
呈现的 %	93%	59%	56%	60%	63%	59%	60%	58%	60%	81%	86%
CDR3											
寡核苷酸简并性	1	1	2	2	2	2	2	2	2	1	1
位置多样性	1	1	9	9	9	9	9	9	9	4	1
具有化学探针的位置											
具有化学探针的多样性											
• 提供广泛的相互作用用化学											
• 覆盖大部分天然存在的多样性											
寡核苷酸多样性											
2E+07											
128											
寡核苷酸多样性											

图 26

