



(12) 发明专利申请

(10) 申请公布号 CN 103890246 A

(43) 申请公布日 2014. 06. 25

(21) 申请号	201280047892. X	C40B 40/08 (2006. 01)
(22) 申请日	2012. 08. 17	C40B 40/02 (2006. 01)
(30) 优先权数据		C40B 30/04 (2006. 01)
	2011903298 2011. 08. 18 AU	C40B 50/06 (2006. 01)
		C07K 16/44 (2006. 01)
(85) PCT国际申请进入国家阶段日		C12N 15/13 (2006. 01)
	2014. 03. 28	A61K 39/395 (2006. 01)
(86) PCT国际申请的申请数据		A61K 31/7088 (2006. 01)
	PCT/AU2012/000970 2012. 08. 17	
(87) PCT国际申请的公布数据		
	W02013/023251 EN 2013. 02. 21	
(71) 申请人	亲和生物科学公司	
	地址 澳大利亚维多利亚州	
(72) 发明人	马修·大卫·贝亚斯利	
	基思·菲利普·尼文	
	本·罗斯·基费尔	
(74) 专利代理机构	北京集佳知识产权代理有限公司	
	11227	
	代理人 彭鲲鹏 卢蓓	
(51) Int. Cl.		
	C40B 40/10 (2006. 01)	

权利要求书3页 说明书48页
序列表37页 附图14页

(54) 发明名称
可溶性多肽
(57) 摘要

本发明一般地涉及表现出高稳定性和溶解度的多肽（例如抗体分子）。特别地，本发明涉及包含在还原性或细胞内环境中表现出可溶性表达和折叠之成对V_L和V_H结构域的多肽。本发明还涉及编码所述多肽的多核苷酸，涉及所述多肽或多核苷酸的文库，以及涉及在研究、诊断和治疗应用中
使用所述多肽的方法。

1. 包含多种不同多肽的多肽文库,所述多肽包含:

i) 抗体重链可变区 (V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区;以及

ii) 抗体轻链可变区 (V_L),其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40(如 SEQ ID NO:18 所示)、IGLV1-44(如 SEQ ID NO:21 所示)、IGLV1-47(如 SEQ ID NO:24 所示)、IGLV1-51(如 SEQ ID NO:15 所示)、IGLV3-1(如 SEQ ID NO:6 所示)、IGLV3-19(如 SEQ ID NO:27 所示)、IGLV3-21(如 SEQ ID NO:9 所示)、IGLV6-57(如 SEQ ID NO:12 所示);其中所述 V_H 和所述 V_L 能够形成抗原结合位点;并且

其中所述多肽的至少两种在所述 V_H 和 / 或 V_L 可变区中一个或更多个互补性决定区 (CDR) 中存在的氨基酸序列上彼此不同。

2. 权利要求 1 所述的文库,其中在所述 V_H 和 / 或 V_L 可变结构域的一个或更多个 CDR 中的氨基酸序列是随机或半随机的或者来自人抗体。

3. 构建多肽文库的方法,所述方法包括制备多种不同的多肽,所述多肽包含:

i) 抗体重链可变区 (V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区;以及

ii) 抗体轻链可变区 (V_L),其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40(如 SEQ ID NO:18 所示)、IGLV1-44(如 SEQ ID NO:21 所示)、IGLV1-47(如 SEQ ID NO:24 所示)、IGLV1-51(如 SEQ ID NO:15 所示)、IGLV3-1(如 SEQ ID NO:6 所示)、IGLV3-19(如 SEQ ID NO:27 所示)、IGLV3-21(如 SEQ ID NO:9 所示)、IGLV6-57(如 SEQ ID NO:12 所示);其中所述 V_H 和所述 V_L 能够形成抗原结合位点;并且

其中所述多肽的至少两种在所述 V_H 和 / 或 V_L 可变区中一个或更多个 CDR 中存在的氨基酸序列上彼此不同。

4. 多核苷酸文库,其包含多种不同的多核苷酸,其中每种多核苷酸编码多肽,所述多肽包含:

i) 抗体重链可变区 (V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区;以及

ii) 抗体轻链可变区 (V_L),其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40(如 SEQ ID NO:18 所示)、IGLV1-44(如 SEQ ID NO:21 所示)、IGLV1-47(如 SEQ ID NO:24 所示)、IGLV1-51(如 SEQ ID NO:15 所示)、IGLV3-1(如 SEQ ID NO:6 所示)、IGLV3-19(如 SEQ ID NO:27 所示)、IGLV3-21(如 SEQ ID NO:9 所示)、IGLV6-57(如 SEQ ID NO:12 所示);其中所述 V_H 和所述 V_L 能够形成抗原结合位点;并且

其中所述多核苷酸的至少两种彼此的不同在于编码下述多肽:所述多肽在所述 V_H 和 / 或 V_L 可变区中包含一个或更多个不同的 CDR。

5. 权利要求 4 所述的文库,其中所述多核苷酸编码的所述 V_H 和 / 或 V_L 可变结构域的一个或更多个 CDR 中的氨基酸序列是随机或半随机的或者来自人抗体。

6. 构建多核苷酸文库的方法,所述方法包括制备编码多肽的多种不同多核苷酸,所述多肽包含:

i) 抗体重链可变区 (V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区;以及

ii) 抗体轻链可变区 (V_L), 其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40 (如 SEQ ID NO:18 所示)、IGLV1-44 (如 SEQ ID NO:21 所示)、IGLV1-47 (如 SEQ ID NO:24 所示)、IGLV1-51 (如 SEQ ID NO:15 所示)、IGLV3-1 (如 SEQ ID NO:6 所示)、IGLV3-19 (如 SEQ ID NO:27 所示)、IGLV3-21 (如 SEQ ID NO:9 所示)、IGLV6-57 (如 SEQ ID NO:12 所示); 其中所述 V_H 和所述 V_L 能够形成抗原结合位点; 并且

其中所述多核苷酸的至少两种彼此的不同在于编码下述多肽: 所述多肽在所述 V_H 和 / 或 V_L 可变区中包含一个或更多个不同的 CDR。

7. 分离的和 / 或重组的多肽, 所述多肽包含:

i) 抗体重链可变区 (V_H), 其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区; 以及

ii) 抗体轻链可变区 (V_L), 其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40 (如 SEQ ID NO:18 所示)、IGLV1-44 (如 SEQ ID NO:21 所示)、IGLV1-47 (如 SEQ ID NO:24 所示)、IGLV1-51 (如 SEQ ID NO:15 所示)、IGLV3-1 (如 SEQ ID NO:6 所示)、IGLV3-19 (如 SEQ ID NO:27 所示)、IGLV3-21 (如 SEQ ID NO:9 所示)、IGLV6-57 (如 SEQ ID NO:12 所示); 其中所述 V_H 和所述 V_L 能够形成抗原结合位点。

8. 权利要求 7 所述的多肽, 其中所述 V_L 包含与如 SEQ ID NO.6 所示 IGLV3-1 的支架区具有至少 90% 同一性的支架区。

9. 权利要求 7 或权利要求 8 所述的多肽, 其是可变片段 (F_v)。

10. 权利要求 7 至 9 中任一项所述的多肽, 其是 Fab 片段、Fab' 片段、 $F(ab')$ 片段、scFv、双抗体、三抗体、四抗体或更高阶的多肽复合物。

11. 权利要求 10 所述的多肽, 其中所述多肽是 scFv 并且所述 V_H 和所述 V_L 通过肽接头连接在一起。

12. 权利要求 7 至 11 中任一项所述的多肽, 其中所述 V_H 和 / 或 V_L 可变区的支架区与任一给定序列的支架区具有至少 95%、96%、97%、98% 或 99% 的同一性。

13. 权利要求 7 至 12 中任一项所述的多肽, 其在还原性条件下可溶。

14. 权利要求 7 至 13 中任一项所述的多肽, 当在还原性条件下产生时所述多肽可溶并且能够稳定地形成抗原结合位点。

15. 权利要求 7 至 14 中任一项所述的多肽, 其与化合物缀合。

16. 权利要求 15 所述的多肽, 其中所述化合物选自: 放射性同位素、可检测的标签、治疗性化合物、胶体、毒素、核酸、肽、蛋白质、提高所述多肽在对象中之半衰期的化合物、及其混合物。

17. 分离的和 / 或外源的多核苷酸, 其编码权利要求 7 至 16 中任一项所述的多肽或其重链或轻链可变区。

18. 载体, 其包含权利要求 17 所述的多核苷酸。

19. 宿主细胞, 其包含权利要求 7 至 16 中任一项所述的多肽、权利要求 17 所述的多核苷酸或权利要求 18 所述的载体。

20. 筛选与靶分子相结合之多肽的方法, 所述方法包括将权利要求 7 至 16 中任一项所述的多肽或权利要求 1、2、4 和 5 中任一项所述的文库与靶分子相接触, 以及确定权利要求 7 至 16 中任一项所述的多肽是否与所述靶分子相结合。

21. 权利要求 20 所述的方法,其中编码权利要求 7 至 16 中任一项所述之多肽的多核苷酸在宿主细胞中或在无细胞表达系统中表达以产生权利要求 7 至 16 中任一项所述的多肽。

22. 权利要求 21 所述的方法,其中所述多核苷酸在宿主细胞的胞质和 / 或周质中表达。

23. 权利要求 22 所述的方法,其中所述宿主细胞是细菌细胞、酵母细胞或哺乳动物细胞。

24. 权利要求 23 所述的方法,其中所述宿主细胞是细菌细胞,并且所述方法包括:

a) 培养包含编码权利要求 7 至 16 中任一项所述多肽之多核苷酸的细菌细胞,从而产生所述多肽,

b) 透化所述细菌细胞,其中所述多核苷酸和所述多肽保留在经透化的细菌细胞内,

c) 将所述经透化的细菌细胞与靶分子相接触,从而使所述靶分子扩散到所述经透化的细菌细胞中,以及

d) 确定权利要求 7 至 16 中任一项所述的多肽是否与所述靶分子相结合。

25. 权利要求 21 所述的方法,其中所述无细胞表达系统包括核糖体展示系统、mRNA 展示系统或顺式展示系统。

26. 宿主细胞文库,其包含多种宿主细胞,所述宿主细胞包含权利要求 7 至 16 中任一项所述的多肽,其中至少一种宿主细胞包含这样的多肽,其与文库中另一宿主细胞中存在的多肽在 V_H 和 / 或 V_L 可变结构域中一个或更多个 CDR 中存在的氨基酸序列上不同。

27. 组合物,其包含权利要求 7 至 16 中任一项所述的多肽、权利要求 17 所述的多核苷酸和 / 或权利要求 18 所述的载体,以及可药用载剂。

28. 试剂盒,其包含权利要求 7 至 16 中任一项所述的多肽、权利要求 17 所述的多核苷酸和 / 或权利要求 18 所述的载体,以及能够透化细菌细胞的试剂。

29. 权利要求 7 至 16 中任一项所述多肽在治疗或诊断应用中的用途。

可溶性多肽

技术领域

[0001] 本发明一般地涉及表现出高稳定性和溶解度的多肽（例如抗体分子）。特别地，本发明涉及包含在还原性或细胞内环境中表现出可溶性表达和折叠的成对 V_L 和 V_H 结构域的多肽。本发明还涉及编码所述多肽的多核苷酸，涉及所述多肽或多核苷酸的文库，以及涉及在研究、诊断和治疗应用中使用所述多肽的方法。例如所述多肽可用于筛选方法以鉴定与特定靶分子相结合的多肽。

背景技术

[0002] 脊椎动物抗体谱 (antibody repertoire) 通过对两种免疫球蛋白 (Ig) 折叠之异二聚体的祖先基因 (ancestral gene) 进行复制和多样化而形成。由免疫系统产生的多样性不仅依赖于 Ig 基因的种系基因 (germline gene) 家族，而且来自 B 细胞和 T 细胞发育过程中亚结构域外显子的体内重组以形成在外显子边界处具有额外多样性的众多独特的谱系，这发生于 Ig 蛋白的表面暴露环 (surface-exposed loop)。这一重组过程称为 V(D)J 重组，在两个可变轻 (V_L) 外显子和三个可变重 (V_H) 外显子分别重组以形成抗体之轻链和重链的 N 端抗原结合结构域之后这样称呼。然而，由于经复制的基因偏离其祖先配对 (ancestral pair)，突变的累积效应导致了可变结构域的异二聚体单元之间不那么完美的界面契合 (interfacial fit)。选择压力不是对任一基因施加，而是对作为整体的家族施加。这样，最大的多样性（其对于免疫系统来说是好事）可导致对于个体家族成员来说不那么理想的折叠稳定性。此外，该结合结构域本身可具有不同的折叠稳定性。由众多偏离的亚单位形成功能性异二聚体的需要通过在结构域的 β -片层之间存在保守二硫键来补偿。然而，界面可仍然是不稳定的契合，需要 ER 中的折叠检查点。

[0003] 通过抗体可变结构域施加的蛋白质契合，作为“共有序列”方法的结果，一些配对具有低折叠稳定性并且倾向于在细菌 / 哺乳动物宿主中较差地表达，以及倾向聚集。此外，在几乎所有情况下，对于将在 V_L 和 V_H 结构域内形成片层间二硫键有总的需要。这就需要对于在细菌宿主（例如大肠杆菌）中表达抗体文库来说，抗体在细胞的周质（具有二硫化物分子伴侣的氧化性空间）中表达，并且经常作为 V_L 和 V_H 结构域之间的融合（单链抗体；scFv）。然而，输出至周质需要通过内膜排泄，这在对于抗体之高表达所期望的水平下饱和，导致远低于胞质表达的产率。

[0004] 除了在大肠杆菌胞质中较便宜地产生 scFv 抗体的优势之外，能够在还原性环境中折叠的抗体支架也将能够用作哺乳动物胞质中的亲和试剂。这将能够扩展抗体作为科学试剂在胞质或细胞核中用于成像或阻断蛋白质功能以及类似地在治疗和诊断中的用途。

[0005] 由于几乎所有哺乳动物抗体在胞质中是不可溶的，一些团队已经寻找折叠形成稳定异二聚体的罕见基因组合以用作支架从而构建进一步的多样性。所采用的寻找胞质可溶性抗体的方法是对抗体克隆在胞质中稳定表达的偶然事件的观察 (Tavladoraki 等, 1999; Vaccaro 等, 2006)，其可以形成细胞内抗体（“胞内抗体”）支架的基础，或者作为替代地，采用进化方法来使 scFv 基因向稳定性进化，无论体内 (Martineau 等, 1998; Visintin 等,

1999 ;Auf der Maur 等,2002 ;Fisher 和 DeLisa,2009) 或体外 (Contreras-Martinez 和 Delisa,2007 ;Jermutus L.,等 2001)。此外,已经证明单结构域抗体(其中仅单个的、未成对的、可变结构域与靶抗原结合)在胞质中是可溶且稳定的。已经描述了当在胞质中表达两种骆驼科(camelid)单结构域抗体时,所述抗体是折叠且可溶的(Kirchhofer A1.,等,2010 ;Saerens 等,2005)。

[0006] 用于在细菌胞质溶胶中产生细胞内抗体的另一策略是使用大肠杆菌突变体,所述大肠杆菌突变体具有将胞质中蛋白质的氧化还原状态从还原改变为氧化的突变。这产生了在大肠杆菌胞质中折叠的且部分地和/或完全地氧化的 scFv(He 等,1995 ;Jurado P.,等,2002)。

[0007] 使用酵母双杂交(Y2H)系统来从 scFv 文库体内筛选与抗原相结合之 scFv 的两个团队编译了用于其可溶性克隆的序列。第一团队(Tse 等,2002)发现 VH3 进化枝(clade)与进化枝 VL κ 1 和 4 配对。通过对多种可溶性 scFv 进行比对,他们编译了用于可溶性 V_L 和 V_H 基因的共有序列,其几乎准确地匹配家族 VH3 和 VL κ 1 的 Morphosys HuCAL™ 文库编译的家族共有序列(Knappik 等,2000)。使用 Y2H 的第二团队在 W003 / 097697 中报道了其可溶性 scFv 是与 VH3、VH1a 或 VH1b 进化枝之成员最密切地相关的序列,其与 VL κ 1 或 VL λ 1 或 VL λ 3 进化枝最密切相关的成员相组合。然而,其最优的构造是与 VH1b 配对的 VL λ 3。然而,关键要注意的是,没有一个报道的序列与最接近的同源免疫球蛋白基因之种系序列的翻译准确地匹配,在整个序列中具有多个突变。推测这是由于这两个团队使用预筛选的噬菌体文库以在酵母转化的限制步骤之前富集抗原结合克隆。然而,这意味着在每个基因中,一个或更多个突变可对胞质中的 scFv 折叠赋予稳定化作用。

[0008] 目前为止,还没有发表关于具有准确鉴别对应于 VL 和 VH 基因之人种系氨基酸序列的细胞内抗体的报道。这样的抗体将会是有助于建造多样性的支架,因为这允许胞质表达的高产率,将以氧化的形式提供较高的稳定性,将提供更高的结构稳定性以确保环多样性的更大耐受性,以及将包含完全天然的序列,通过产生全抗体来产生较低的患者排斥。

[0009] 我们在此报道了应用先前描述于 W02011 / 075761 的蛋白质展示方法来筛选人 scFv 文库以及分离与人种系序列具有相同骨架区的可溶性 scFv 基因。此外,我们证明了显著的热稳定性和移植到 scFv 支架上之 CDR3 的耐受性。

[0010] 发明概述

[0011] 在一个方面,本发明提供了包含多种不同多肽的多肽文库,其包含:

[0012] i) 抗体重链可变区(V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区(scaffold region)具有至少 90%同一性的支架区;以及

[0013] ii) 抗体轻链可变区(V_L),其包含与以下任一种的支架区具有至少 90%同一性的支架区:IGLV1-40(如 SEQ ID NO:18 所示)、IGLV1-44(如 SEQ ID NO:21 所示)、IGLV1-47(如 SEQ ID NO:24 所示)、IGLV1-51(如 SEQ ID NO:15 所示)、IGLV3-1(如 SEQ ID NO:6 所示)、IGLV3-19(如 SEQ ID NO:27 所示)、IGLV3-21(如 SEQ ID NO:9 所示)、IGLV6-57(如 SEQ ID NO:12 所示);

[0014] 其中所述 V_H 和所述 V_L 能够形成抗原结合位点(antigen-binding site);并且

[0015] 其中至少两种多肽在所述 V_H 和/或 V_L 可变区中一个或更多个互补性决定区(CDR)中存在的氨基酸序列上彼此不同。

[0016] 优选地,所述 V_H 和 / 或 V_L 可变结构域的一个或更多个 CDR 中氨基酸的序列是随机或半随机的,或来自人抗体。

[0017] 在另一方面中,本发明提供了构建多肽文库的方法,所述方法包括制备多种不同的多肽,所述多肽包含:

[0018] i) 抗体重链可变区 (V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区;以及

[0019] ii) 抗体轻链可变区 (V_L),其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40 (如 SEQ ID NO:18 所示)、IGLV1-44 (如 SEQ ID NO:21 所示)、IGLV1-47 (如 SEQ ID NO:24 所示)、IGLV1-51 (如 SEQ ID NO:15 所示)、IGLV3-1 (如 SEQ ID NO:6 所示)、IGLV3-19 (如 SEQ ID NO:27 所示)、IGLV3-21 (如 SEQ ID NO:9 所示)、IGLV6-57 (如 SEQ ID NO:12 所示);其中所述 V_H 和所述 V_L 能够形成抗原结合位点;并且

[0020] 其中至少两种多肽在所述 V_H 和 / 或 V_L 可变区中一个或更多个 CDR 中存在的氨基酸序列上彼此不同。

[0021] 在另一方面中,本发明提供了包含多种不同多核苷酸的多核苷酸文库,

[0022] 其中每种多核苷酸编码多肽,所述多肽包含:

[0023] i) 抗体重链可变区 (V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区;以及

[0024] ii) 抗体轻链可变区 (V_L),其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40 (如 SEQ ID NO:18 所示)、IGLV1-44 (如 SEQ ID NO:21 所示)、IGLV1-47 (如 SEQ ID NO:24 所示)、IGLV1-51 (如 SEQ ID NO:15 所示)、IGLV3-1 (如 SEQ ID NO:6 所示)、IGLV3-19 (如 SEQ ID NO:27 所示)、IGLV3-21 (如 SEQ ID NO:9 所示)、IGLV6-57 (如 SEQ ID NO:12 所示);

[0025] 其中所述 V_H 和所述 V_L 能够形成抗原结合位点;并且

[0026] 其中至少两种多核苷酸彼此的不同在于编码下述多肽:所述多肽在所述 V_H 和 / 或 V_L 可变区中包含一个或更多个不同的 CDR。

[0027] 优选地,所述多核苷酸编码的所述 V_H 和 / 或 V_L 可变结构域的一个或更多个 CDR 中的氨基酸序列是随机或半随机的,或者来自人抗体。

[0028] 在另一方面中,本发明提供了构建多核苷酸文库的方法,所述方法包括制备编码多肽的多种不同多核苷酸,所述多肽包含:

[0029] i) 抗体重链可变区 (V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区;以及

[0030] ii) 抗体轻链可变区 (V_L),其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40 (如 SEQ ID NO:18 所示)、IGLV1-44 (如 SEQ ID NO:21 所示)、IGLV1-47 (如 SEQ ID NO:24 所示)、IGLV1-51 (如 SEQ ID NO:15 所示)、IGLV3-1 (如 SEQ ID NO:6 所示)、IGLV3-19 (如 SEQ ID NO:27 所示)、IGLV3-21 (如 SEQ ID NO:9 所示)、IGLV6-57 (如 SEQ ID NO:12 所示);

[0031] 其中所述 V_H 和所述 V_L 能够形成抗原结合位点;并且

[0032] 其中至少两种多核苷酸彼此的不同在于编码下述多肽:所述多肽在所述 V_H 和 / 或 V_L 可变区中包含一个或更多个不同 CDR 的多肽。

[0033] 在另一方面中,本发明提供分离的和 / 或重组的多肽,所述多肽包含:

[0034] i) 抗体重链可变区 (V_H),其包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区;以及

[0035] ii) 抗体轻链可变区 (V_L),其包含与以下任一种支架区具有至少 90% 同一性的支架区: IGLV1-40 (如 SEQ ID NO:18 所示)、IGLV1-44 (如 SEQ ID NO:21 所示)、IGLV1-47 (如 SEQ ID NO:24 所示)、IGLV1-51 (如 SEQ ID NO:15 所示)、IGLV3-1 (如 SEQ ID NO:6 所示)、IGLV3-19 (如 SEQ ID NO:27 所示)、IGLV3-21 (如 SEQ ID NO:9 所示)、IGLV6-57 (如 SEQ ID NO:12 所示);

[0036] 其中所述 V_H 和所述 V_L 能够形成抗原结合位点。

[0037] 所述 V_L 优选地包含与如 SEQ ID NO6 所示 IGLV3-1 的支架区具有至少 90% 同一性的支架区。

[0038] 优选地,所述多肽是可变片段 (Fv),例如 Fab 片段、Fab' 片段、F(ab') 片段、scFv、双抗体 (diabody)、三抗体 (triabody)、四抗体 (tetrabody) 或更高阶的 (higher order) 多肽复合物。更优选地,所述多肽是 scFv 并且所述 V_H 和 V_L 通过肽接头连接在一起。

[0039] 优选地,本发明的多肽中 V_H 和 / 或 V_L 可变区的支架区与任一给定序列的支架区具有至少 95%、96%、97%、98% 或 99% 的同一性。

[0040] 本发明的多肽优选地在还原性条件下可溶。此外,本发明的多肽优选地在还原性条件下产生时可溶并且能够稳定地形成抗原结合位点。

[0041] 在另一个优选的实施方案中,本发明的多肽与化合物缀合。所述化合物可以选自:放射性同位素、可检测的标签、治疗性化合物、胶体、毒素、核酸、肽、蛋白质、提高所述多肽在对象中之半衰期的化合物、及其混合物。

[0042] 在另一方面中,本发明提供了分离的或外源的多核苷酸,其编码本发明的多肽或其重链或轻链可变区之多肽。

[0043] 在另一方面中,本发明提供了包含本发明之多核苷酸的载体 (vector)。

[0044] 在另一方面中,本发明提供了包含本发明之多肽、本发明之多核苷酸或本发明之载体的宿主细胞。

[0045] 在另一方面中,本发明提供了用于筛选与靶分子相结合之多肽的方法,所述方法包括将本发明的多肽与靶分子相接触,并且确定所述多肽是否与所述靶分子相结合。在这样的一些方法中,优选如果编码所述多肽的多核苷酸在宿主细胞中或在无细胞表达系统中表达以产生所述多肽。当在细胞内表达所述多肽时,这种表达可发生在宿主细胞 (例如细菌细胞、酵母细胞或哺乳动物) 的胞质和 / 或周质 (periplasm) 中。

[0046] 在一个优选的实施方案中,所述宿主细胞是细菌细胞并且所述方法包括:

[0047] a) 培养包含编码本发明多肽之多核苷酸的细菌细胞,从而产生所述多肽,

[0048] b) 透化所述细菌细胞,其中所述多核苷酸和所述多肽保留在经透化细菌细胞内,

[0049] c) 将经透化的细菌细胞与靶分子相接触,从而使所述靶分子扩散到所述经透化的细菌细胞中,以及

[0050] d) 确定本发明的多肽是否与所述靶分子相结合。

[0051] 本发明的筛选方法可以使用本文中所描述的任意多肽进行。优选地,本发明的筛选方法包括筛选本发明的文库。这样,所述筛选方法可包括表达本发明的多肽或多核苷酸

文库以及在这些文库中鉴定与靶分子相结合的多肽。优选地,这样的筛选方法在还原性条件下进行。例如,这样的方法可以在宿主细胞中进行。在一个优选的实施方案中,这样的方法在宿主细胞的胞质中进行。优选地,所述宿主细胞是细菌细胞,例如革兰氏阴性细菌细胞。在一个优选的实施方案中,所述细菌是大肠杆菌细胞。

[0052] 在另一方面中,本发明提供包含多种宿主细胞的宿主细胞文库,所述宿主细胞包含本发明的多肽,其中至少一种宿主细胞包含这样的多肽,其与文库中另一种宿主细胞中存在的多肽在 V_H 和 / 或 V_L 可变结构域中一个或多个 CDR 中存在的氨基酸序列不同。在本发明的宿主细胞文库中的一种或更多种宿主细胞可以包含一种或更多种编码本发明之多肽的多核苷酸。例如,在所述宿主细胞文库中的宿主细胞可以含有一种编码 V_H 的多核苷酸和另一种编码 V_L 的多核苷酸。

[0053] 在另一方面中,本发明提供组合物,其包含本发明之多肽、多核苷酸和 / 或载体、以及可药用载剂 (pharmaceutically acceptable carrier)。

[0054] 在另一方面中,本发明提供试剂盒 (kit),其包含本发明之多肽、本发明之多核苷酸和 / 或本发明、以及能够透化细菌细胞的试剂 (agent)。

[0055] 在另一方面中,本发明提供了本发明之多肽在治疗或诊断应用中的用途。

[0056] 如将是显而易见的那样,根据情况作适当变动后,本发明的一个方面中优选的特性和特征可适用于本发明的任何其他方面。

[0057] 贯穿本说明书的词“包括”或变体例如“包含”或“含有”将被理解为意指包括所陈述的要素、整数或步骤或者要素、整数或步骤的组,但不排除任何其他元素、整数或步骤或者要素、整数或步骤的组。

[0058] 在下文中通过以下非限制性实施例并参照附图的方式描述本发明。

[0059] 附图简述

[0060] 图 1 示出良好表达的可溶性 scFv 克隆的典型外观 (1A, 以及插图), 以及良好表达但不可溶性 scFv 克隆 (1B, 以及插图)。

[0061] 图 2 示出所选择的与 VL 基因 IGLV3-1、IGLV3-21 和 IGLV6-57 具有高度相似性 (similarity) 或总同一性 (total identity) 的可溶性克隆的多重比对 (multiple alignment)。

[0062] 图 3 展示出两个克隆 (一个 IGLV3-1 和一个 IGLV3-21) 在提高的温度下表达的行为。

[0063] 图 4 展示出当在 25°C 下于大肠杆菌胞质溶胶中表达时 IGLV3-1 克隆溶解性。scFv : :127 : :FLAG 融合蛋白完整地可溶性 (S) 级分中。

[0064] 图 5 展示出具有用 J1 或 J2 替换 λ J 区之初始克隆 (#8.93) 的热稳定性行为。

[0065] 图 6A 展示出具有多样化 IGLV3-1CDR3 的 4 个独立克隆的溶解性和高表达。

[0066] 图 6B 展示出具有多样化 IGHV3-23CDR3 克隆之完整群体的样品。

[0067] 图 7 举例说明在本文中描述的优选可变区中的示例性 CDR (用粗体和 / 或下划线表示)。

[0068] 图 8 举例说明编码具有可变 CDR3 区之 IGLV3-1 : :IGHV3-23 支架的多核苷酸序列以及响应的经翻译氨基酸序列的实例。CDR 用下划线和粗体表示。肽接头序列用斜体表示。

[0069] 图 9 示出 SNAP 配体标记的 IGLV3-1 : :IGHV3-23scFv 文库, 展示出可溶性文库成

员的高频率。

[0070] 图 10 示出通过 RED 筛选分离结合 mAG 的 scFv。克隆 34 对于 mAG 结合是阳性的。克隆 25 是阴性的。

[0071] 图 11 举例说明具有通过 RED 筛选分离之完整 α -mAG scFv 的 IGLV3-1 : : IGHV3-23scFv 支架,其作为存在于可溶性级分 (S) 中的 C-端 His6FLAG 融合蛋白而克隆,且在不可溶性级分 (P) 中没有蛋白质。使用 α -FLAG 单克隆抗体进行检测。

[0072] 图 12 示出与 IMAC Ni 琼脂糖结合的 α -mAG scFv His6FLAG 对 mAG 的结合。

[0073] 图 13 展示出 α -mAG scFv 与 mAG 相互作用的特异性 (通过从大肠杆菌裂解物“拉下 (pull-down)”未经纯化的 mAG)。 α -mAG scFv His6FLAG 与 IMAC Ni-琼脂糖树脂相结合,添加在总大肠杆菌细胞裂解物中的 mAG (泳道 6 和 7) 导致具有 mAG 预期大小 (约 26KD) 之蛋白质的结合。

[0074] 图 14 示出来自使用封装裂解缺陷型噬菌体展示 gpD : : α -mAG scFv 融合蛋白的“掺杂的 (doped)”mAG 文库筛选的 FACS 阶段的抓屏 (screen-grab) (上图)。含有经封装噬菌体的 mAG 阳性细胞在右侧门 (gate) 中。使用 RED 方法诱导从 FACS 筛选回收的噬菌体用于复制和 gpD : : α -mAG 表达以及用 mAG 标记 (下图)。

[0075] 序列表要点

[0076] SEQ ID NO :1- 编码 IGHV3-23 的多核苷酸序列 (NCBI Ref. NT_026437. 12)。

[0077] SEQ ID NO :2- 编码 IGHV3-23 的多核苷酸序列,排除内含子。

[0078] SEQ ID NO :3-IGHV3-23 的氨基酸序列

[0079] SEQ ID NO :4- 编码 IGLV3-1 的多核苷酸序列 (NCBI Ref. NT_011520. 12)。

[0080] SEQ ID NO :5- 编码 IGLV3-1 的多核苷酸序列,排除内含子。

[0081] SEQ ID NO :6-IGLV3-1 的氨基酸序列

[0082] SEQ ID NO :7- 编码 IGLV3-21 的多核苷酸序列 (NCBI Ref. NT_011520. 12)。

[0083] SEQ ID NO :8- 编码 IGLV3-21 的多核苷酸序列,排除内含子。

[0084] SEQ ID NO :9-IG Ly3-21 的氨基酸序列

[0085] SEQ ID NO :10- 编码 IGLV6-57 的多核苷酸序列 (NCBI 参考 :NW_001838745. 1)。

[0086] SEQ ID NO :11- 编码 IGLV6-57 的多核苷酸序列,排除内含子。

[0087] SEQ ID NO :12-IGLV6-57 的氨基酸序列

[0088] SEQ ID NO :13- 编码 IGLV1-51 的多核苷酸序列 (NCBI 参考序列 :NT_011520. 12)

[0089] SEQ ID NO :14- 编码 IGLV1-51 的多核苷酸序列,排除内含子。

[0090] SEQ ID NO :15-IGLV1-51 的氨基酸序列

[0091] SEQ ID NO :16- 编码 IGLV1-40 的多核苷酸序列 (NCBI 参考序列 :NT_011520. 12)

[0092] SEQ ID NO :17- 编码 IGLV1-40 的多核苷酸序列,排除内含子。

[0093] SEQ ID NO :18-IGLV1-40 的氨基酸序列

[0094] SEQ ID NO :19- 编码 IGLV1-44 的多核苷酸序列 (NCBI 参考序列 :NT_011520. 12)。

[0095] SEQ ID NO :20- 编码 IGLV1-44 的多核苷酸序列,排除内含子。

[0096] SEQ ID NO :21-IGLV1-44 的氨基酸序列

[0097] SEQ ID NO :22- 编码 IGLV1-47 的多核苷酸序列 (NCBI 参考序列 :NT_011520. 12)

[0098] SEQ ID NO :23- 编码 IGLV1-47 的多核苷酸序列,排除内含子。

- [0099] SEQ ID NO :24-IGLV1-47 的氨基酸序列
- [0100] SEQ ID NO :25- 编码 IGLV3-19 的多核苷酸序列 (NCBI 参考序列 :NT_011520.12)
- [0101] SEQ ID NO :26- 编码 IGLV3-19 的多核苷酸序列,排除内含子。
- [0102] SEQ ID NO :27-IGLV3-19 的氨基酸序列
- [0103] SEQ ID NO :28- 优选的肽接头
- [0104] SEQ ID NO :29-CDR 变体序列
- [0105] SEQ ID NO :30- 作为替代的 CDR 变体序列
- [0106] SEQ ID NO :31- 引物 HVK1F1
- [0107] SEQ ID NO :32- 引物 HVK1F2
- [0108] SEQ ID NO :33- 引物 HVK2F
- [0109] SEQ ID NO :34- 引物 HVK3F
- [0110] SEQ ID NO :35- 引物 HVK4F
- [0111] SEQ ID NO :36- 引物 HVK5F
- [0112] SEQ ID NO :37- 引物 HVK6F
- [0113] SEQ ID NO :38- 引物 HVKCL R
- [0114] SEQ ID NO :39- 引物 HVL1F1
- [0115] SEQ ID NO :40- 引物 HVL1F2
- [0116] SEQ ID NO :41- 引物 HVL2F
- [0117] SEQ ID NO :42- 引物 HVL3F1
- [0118] SEQ ID NO :43- 引物 HVL3F2
- [0119] SEQ ID NO :44- 引物 HVL4F1
- [0120] SEQ ID NO :45- 引物 HVL4F2
- [0121] SEQ ID NO :46- 引物 HVL5F
- [0122] SEQ ID NO :47- 引物 HVL6F
- [0123] SEQ ID NO :48- 引物 HVL7 / 8F
- [0124] SEQ ID NO :49- 引物 HVL9 / 10F
- [0125] SEQ ID NO :50- 引物 01115HVLCL R
- [0126] SEQ ID NO :51- 引物 01116HVLCL R2
- [0127] SEQ ID NO :52- 引物 HVK12F1
- [0128] SEQ ID NO :53- 引物 HVK12F2
- [0129] SEQ ID NO :54- 引物 HVK22F
- [0130] SEQ ID NO :55- 引物 HVK32F
- [0131] SEQ ID NO :56- 引物 HVK42F
- [0132] SEQ ID NO :57- 引物 HVK52F
- [0133] SEQ ID NO :58- 引物 HVK62F
- [0134] SEQ ID NO :59- 引物 HVKCL2R
- [0135] SEQ ID NO :60- 引物 HVL12F1
- [0136] SEQ ID NO :61- 引物 HVL12F2
- [0137] SEQ ID NO :62- 引物 HVL22F

- [0138] SEQ ID NO :63- 引物 HVL32F1
- [0139] SEQ ID NO :64- 引物 HVL32F2
- [0140] SEQ ID NO :65- 引物 HVL42F1
- [0141] SEQ ID NO :66- 引物 HVL42F2
- [0142] SEQ ID NO :67- 引物 HVL52F
- [0143] SEQ ID NO :68- 引物 HVL62F
- [0144] SEQ ID NO :69- 引物 HVL7 / 82F
- [0145] SEQ ID NO :70- 引物 HVL9 / 102F
- [0146] SEQ ID NO :71- 引物 HVLCL2R
- [0147] SEQ ID NO :72- λ J 区 J1
- [0148] SEQ ID NO :73- λ J 区 J2
- [0149] SEQ ID NO :74- λ J 区 J3
- [0150] SEQ ID NO :75- λ J 区 J4
- [0151] SEQ ID NO :76- λ J 区 J5
- [0152] SEQ ID NO :77- λ J 区 J6
- [0153] SEQ ID NO :78- λ J 区 J7
- [0154] SEQ ID NO :79- 杂合 J 区序列
- [0155] SEQ ID NO :80-PCR 引物
- [0156] SEQ ID NO :81- 翻译的序列
- [0157] SEQ ID NO :82-PCR 引物
- [0158] SEQ ID NO :83- 翻译的序列
- [0159] SEQ ID NO :84- 编码具有可变 CDR3 区之 IGLV3-1 : IGHV3-23 支架的多核苷酸序列
- [0160] SEQ ID NO :85- 由 SEQ ID NO :84 之多核苷酸编码的氨基酸序列
- [0161] SEQ ID NO :86- 模板 CDR3 序列
- [0162] SEQ ID NO :87- 作为替代的模板 CDR3 序列
- [0163] SEQ ID NO :88-IGLV3-1 的骨架序列和 IGHV3-23 的 J 区
- [0164] SEQ ID NO :89- 间插序列
- [0165] SEQ ID NO :90- 简并引物 1
- [0166] SEQ ID NO :91- 简并引物 2
- [0167] SEQ ID NO :92-CDR3 环 L1
- [0168] SEQ ID NO :93-CDR3 环 H1
- [0169] SEQ ID NO :94-CDR3 环 L2
- [0170] SEQ ID NO :95-CDR3 环 H2
- [0171] SEQ ID NO :96-CDR3 环 L3
- [0172] SEQ ID NO :97-CDR3 环 H3
- [0173] SEQ ID NO :98-CDR3 环 L4
- [0174] SEQ ID NO :99-CDR3 环 H4
- [0175] SEQ ID NO :100-CDR3 环 L5

- [0176] SEQ ID NO:101-CDR3 环 H5
[0177] SEQ ID NO:102-CDR3 环 L6
[0178] SEQ ID NO:103-CDR3 环 H6
[0179] SEQ ID NO:104-CDR3 环 L8
[0180] SEQ ID NO:105-CDR3 环 H8
[0181] SEQ ID NO:106-CDR3 环 L9
[0182] SEQ ID NO:107-CDR3 环 H9
[0183] SEQ ID NO:108-CDR3 环 L10
[0184] SEQ ID NO:109-CDR3 环 H10
[0185] SEQ ID NO:110-mAG-BioHis6 蛋白
[0186] SEQ ID NO:111-Anti-mAG-BioHis6scFv 序列
[0187] SEQ ID NO:112-gpD: : α -mAG scFv 融合构建体多核苷酸序列
[0188] SEQ ID NO:113-gpD: : α -mAG scFv 融合蛋白序列
[0189] SEQ ID NO:114-野生型人 IGLV3-1
[0190] SEQ ID NO:115-可溶性克隆 8.93
[0191] SEQ ID NO:116-可溶性克隆 8.184
[0192] SEQ ID NO:117-可溶性克隆 8.174
[0193] SEQ ID NO:118-可溶性人 IGLV3-21 克隆 8.186
[0194] SEQ ID NO:119-可溶性人 IGLV3-21 克隆 8.39
[0195] SEQ ID NO:120-野生型人 IGLV3-21
[0196] SEQ ID NO:121-可溶性人 IGLV3-21 克隆 9.19
[0197] SEQ ID NO:122-野生型人 IGLV6-57
[0198] SEQ ID NO:123-可溶性克隆 16.26
[0199] SEQ ID NO:124-可溶性克隆 16.1
[0200] SEQ ID NO:125-可溶性克隆 16.121

[0201] 发明详述

[0202] 一般技术和定义

[0203] 除非另外特别地定义,本文中所使用的所有技术和科学术语应当采用与本技术领域(例如,在蛋白质化学、生物化学、细胞培养、分子遗传学、微生物学和免疫学中)普通技术人员通常理解的含义。

[0204] 除非另有说明,在本发明中使用的重组蛋白质、细胞培养和免疫学技术是本领域技术人员公知的标准操作。这样的技术通过例如以下来源的文献描述和阐明: J.Perbal, A Practical Guide to Molecular Cloning, John Wiley and Sons(1984); J.Sambrook 等, Molecular Cloning: A Laboratory Manual, 第3版, Cold Spring Harbour Laboratory Press(2001); R.Scopes, Protein Purification-Principals and Practice, 第3版, Springer (1994); T.A.Brown(编辑), Essential Molecular Biology: A Practical Approach, 卷1和2, IRL Press(1991); D.M.Glover 和 B.D.Hames(编辑), DNA Cloning: A Practical Approach, 卷1-4, IRL Press(1995和1996); 以及 EM.Ausubel 等(编辑), Current Protocols in Molecular Biology, Greene Pub.Associates and

Wiley-Interscience(1988,包括到目前为止所有的更新);Ed Harlow 和 David Lane(编辑), *Antibodies: A Laboratory Manual*, Cold Spring Harbour Laboratory, (1988);以及 J.E. Coligan 等(编辑), *Current Protocols in Immunology*, John Wiley&Sons(包括到目前为止所有的更新)。

[0205] 术语“多肽”、“蛋白质”、“肽”在本文中通常可互换使用。本文中使用的术语“外源性多肽”是指由外源性多核苷酸编码的多肽。本文中使用的术语“外源性多核苷酸”是指多核苷酸序列对于其所引入的细胞来说是外来的,或者序列与其所引入的细胞中的序列是同源的,但是处于宿主细胞核酸中通常不出现所述多核苷酸的位置。

[0206] 本文中使用的术语“抗体”、“抗体分子”包括多克隆抗体、单克隆抗体、双特异性抗体、双抗体、三抗体、多抗体(multibody)、异缀合(heteroconjugate)抗体、嵌合抗体(包括完整分子及其片段,例如 Fab、F(ab')₂、Fv 和 scFv 以及其他抗体样分子)。技术人员将会知晓,抗体通常被认为是包含由多个多肽链构成之可变区(例如轻链可变区(V_L)和重链可变区(V_H))的蛋白质。抗体还可以包含恒定结构域,其可以设置成恒定区或恒定片段或可结晶片段(fragment crystallisable, Fc)。抗体可以与一个或几个密切相关的抗原特异性地结合。全长抗体通常包含共价连接的两个重链(约 50 至 70kD)和两个轻链(各约 23kD)。轻链通常包含可变区和恒定结构域,并且在哺乳动物中是 κ 轻链或 λ 轻链。重链通常包含可变区和通过铰链区与另外的恒定结构域相连的一个或两个恒定结构域。哺乳动物的重链是下列类型之一: α 、 δ 、 ϵ 、 γ 或 μ 。每个轻链还与重链之一共价相连。例如,两个重链以及重链与轻链可以通过链间二硫键和/或通过非共价相互作用连在一起。链间二硫键(如果存在)的数目可根据抗体的不同类型而变化。每个链具有 N 端可变区(V_H或 V_L,其中每个长度为约 110 个氨基酸)以及在 C 端的一个或更多个恒定结构域。轻链的恒定结构域(长 C_L,其长度为约 110 个氨基酸)经常与重链的第一恒定结构域(C_H,其长度为约 330 至 340 个氨基酸)对齐并且通过二硫键与其键合。轻链可变区经常与重链可变区对齐。抗体重链可以包含 2 个或更多个另外的 C_H 结构域(例如 C_H2、C_H3 等)并且可包含可在 C_H1 和 C_m 恒定结构域之间确认的铰链区。抗体可以是任意类型(例如, IgG、IgE、IgM、IgD、IgA 和 IgY)、类别(例如, IgG₁、IgG₂、IgG₃、IgG₄、IgA₁ 和 IgA₂)或亚类。优选地,所述抗体是鼠(小鼠或大鼠)抗体或灵长类(优选人)抗体。术语“抗体”还包括人源化抗体、灵长化(primatized)抗体、人抗体和嵌合抗体。

[0207] 本文中使用的术语“可变区”是指如本文中所定义之抗体的轻链和重链的一部分,其包含 CDR(即, CDR1、CDR2 和 CDR3)和骨架区(framework region, FR)的氨基酸序列。V_H是指重链的可变区。V_L是指轻链的可变区。

[0208] 本文中使用的术语“支架区”是指除了 CDR 残基之外的所有可变区残基。

[0209] 本文中使用的术语“骨架区”(FR)将被理解为意指除了 CDR 残基之外的可变区残基的连续序列。这样,所有 FR 一起构成“支架区”。天然抗体的每个可变区通常具有四个 FR,确定为 FR1、FR2、FR3 和 FR4。如果 CDR 是根据 Kabat 定义的,示例性轻链 FR(LCFR)残基定位于约残基 1 至 23(LCFR1)、35 至 49(LCFR2)、57 至 88(LCFR3)和 98 至 107(LCFR4)。注意 λ LCFR1 不包含残基 10,其包含于 κ LCFR1。示例性重链 FR(HCFR)残基定位于约残基 1 至 30(HCFR1)、36 至 49(HCFR2)、66 至 94(HCFR3)和 103 至 113(HCFR4)。

[0210] 本文中使用的术语“互补性决定区”(CDR,即 CDR1、CDR2 和 CDR3 或高变区)是指

免疫球蛋白可变区的氨基酸残基,其存在对于抗原结合是必要的。每个可变区通常具有被确定为 CDR1、CDR2 和 CDR3 的三个 CDR 区。每个 CDR 可以包含来自如由 Kabat (1987 和 / 或 1991) 定义的“互补性决定区”的氨基酸残基。例如,在重链可变区中,CDRH1 在残基 31 至 35 之间,CDRH2 在残基 50 至 65 之间并且 CDRH3 在残基 95 至 102 之间。在轻链中,CDRL1 在残基 24 至 34 之间,CDRL2 在残基 50 至 56 之间并且 CDRL3 在残基 89 至 97 之间。这些 CDR 还可包含大量插入,例如 Kabat (1987 和 / 或 1991) 中所述。

[0211] 本文中使用的术语“恒定区”(CR 或可结晶片段或 Fc) 是指抗体的一部分,其包含至少一个恒定结构域并且其通常(尽管不是必须)被糖基化并且其与一个或多个受体和 / 或补体级联的组分(例如赋予效应子功能)相结合。重链恒定区可选自以下五种亚型中的任一种: α 、 δ 、 ϵ 、 γ 或 μ 。此外,多种亚类(例如重链的 IgG 亚类)的重链负责不同的效应子功能,并且因此通过选择期望的重链恒定区,可产生具有期望效应子功能的蛋白质。优选的重链恒定区是 $\gamma 1$ (IgG1)、 $\gamma 2$ (IgG2) 和 $\gamma 3$ (IgG3)。

[0212] “恒定结构域”是抗体中的结构域,其序列在相同类型(例如, IgG 或 IgM 或 IgE) 的抗体中高度相似。抗体的恒定区通常包含多个恒定结构域,例如 γ 、 α 和 δ 重链的恒定区包含三个恒定结构域并且 γ 、 α 和 δ 重链的 Fc 包含两个恒定结构域。 μ 和 ϵ 重链的恒定区包含四个恒定结构域并且 Fc 区包含两个恒定结构域。

[0213] 本文中使用的术语“Fv”应当被理解为任意蛋白质,无论包含多种多肽还是单一多肽,其中 V_L 和 V_H 缔合并形成具有抗原结合位点(即能够特异性地与抗原相结合)的复合物。 V_H 和 V_L 可以以单一多肽链或以不同多肽链形成抗原结合位点。此外本发明的 Fv(以及本发明的任意多肽)可以具有多个抗原结合位点,所述多个抗原结合位点可以或不可以与相同的抗原结合。术语“Fv”应当理解为包括直接衍生于抗体的片段以及对应于使用重组方式产生的此类片段的蛋白质。在一些优选的实施方案中,所述 V_H 不与重链恒定区 (C_H)1 相连和 / 或所述 V_L 不与轻链恒定区 (C_L) 相连。含有多肽或蛋白质的示例性 Fv 包括 Fab 片段、Fab' 片段、F(ab') 片段、scFv、双抗体、三抗体、四抗体、或更高阶的复合物,或者与恒定区或其结构域(例如 C_H2 或 C_H3 结构域)相连的上述任一种。“Fab 片段”由抗体的单价抗原结合片段组成,并且可以例如通过用木瓜蛋白酶消化整个抗体以产生完整轻链和重链的一部分组成的片段来产生或者可以通过重组方式产生。可以例如通过用胃蛋白酶处理整个免疫球蛋白,随后还原以产生由完整轻链和重链的一部分组成的分子来获得抗体的“Fab' 片段”。以这种方式处理每个抗体获得的两个 Fab' 片段。Fab' 片段也可通过重组方式产生。抗体的“F(ab')₂ 片段”由两个 Fab' 片段通过两个二硫键连在一起形成的二聚体组成,并且通过用胃蛋白酶处理整个抗体分子获得,无需后续的还原。“Fab₂”片段是重组片段,其包含使用例如亮氨酸拉链或 C_H3 结构域连接的两个 Fab 片段。

[0214] “单链 Fv”或“scFv”是含有免疫球蛋白之可变区片段(Fv)的重组分子,其中轻链的可变区和重链的可变区通过合适的柔性多肽接头共价连接。下文提供了落入该术语范围内的含有多肽之示例性 Fv 的详细讨论。

[0215] 本文中使用的术语“抗原结合位点”应当被理解为意指由能够与抗原特异性结合的多肽形成的结构。抗原结合位点不需要是连续的氨基酸,甚至不需要是在单一多肽链中。例如,由两条不同多肽链产生的 Fv 中,抗原结合位点由一些列与抗原相互作用的 V_L 和 V_H 之区域构成,并且其通常(但不总是)在每个可变区的一个或多个 CDR 中。

[0216] 本文中被命名为 CDR 和 FR 的任意氨基酸位置根据 Kabat (1987 和 1991) 定义。技术人员将能够容易地在本发明的进行中使用其他编号系统,例如 Chothia 和 Lesk (1987 和 / 或 1989) 和 / 或 Al-Lazikani 等 (1997) 的高变换编号系统。

[0217] 技术人员将会知晓,“二硫键”是通过巯基偶联形成的共价键。所述键也被成为 SS- 键或二硫桥 (disulfide bridge)。在多肽中,二硫键通常发生在两个半胱氨酸残基的巯基之间。

[0218] 技术人员还将会知晓,术语“非还原性条件”包括足以使蛋白质中的巯基 (-SH) 基团氧化的条件,例如容许二硫键形成。因此,术语“还原性条件”包括不足以使蛋白质中的巯基 (-SH) 基团氧化的条件,例如不容许二硫键形成。

[0219] 本文中使用的术语“抗原”应当被理解为意指主要针对其可使抗体应答出现的任意的物质组合物。示例性抗原包括蛋白质、肽、多肽、碳水化合物、磷酸基团、磷光肽 (phosphor-peptide) 或多肽,糖基化肽或肽等。

[0220] 本文中对可变区及其一部分、免疫球蛋白、抗体及其片段的描述和定义可以进一步通过 Kabat (1987 和 / 或 1991)、Bork 等 (1994) 和 / 或 Chothia 和 Lesk (1987 和 1989) 或 Al-Lazikani 等 (1997) 中的讨论来阐明,。

[0221] 本文中使用的术语“缀合”、“缀合的”或其变化形式广泛地用于指可用于本文中公开的方法中的化合物与其他试剂之间任意形式的共价键或非共价键缔合。

[0222] 术语“和 / 或” (例如“X 和 / 或 Y”) 应当被理解为意指“X 和 Y”或“X 或 Y”并且应当对两种含义或对含义之一提供清楚的支持。

[0223] 本文中使用的术语“约”是指指定值 + / -5% 的范围。

[0224] 如将通过以下描述理解的,本发明人应用了蛋白质展示方法以鉴别可在细胞的胞质中以可溶性形式表达并且展示出出乎意料水平的可溶性、热稳定性、以及对 CDR 多样化之耐受的多肽 (例如抗体)。本发明人进一步证明,仅使用抗体可变结构域之骨架区中的种系序列,人免疫球蛋白谱 (immunoglobulin repertoire) 具有用于胞质可溶性和稳定性的潜力。

[0225] 保留封装展示 (Retained Encapsulated Display, RED):

[0226] 本发明人已经鉴别了可以在细胞的胞质中以可溶性形式表达并且使用保留封装展示 (RED) 的方法展示出出乎意料水平的可溶性、热稳定性、以及对 CDR 多样化之耐受的多肽。RED 是用于革兰氏阴性细菌的蛋白质展示平台,这描述于 W02011 / 075761 (其内容通过引用整体并入本文)。在 RED 中待展示的蛋白质在细胞的周质或胞质中表达。然后用去污剂或有机溶剂透化细胞膜,同时完整的留下细胞壁。展示蛋白质通过细胞壁被保留,通过与蛋白质融合将其分子量提高至高于细胞壁的孔隙率限制 (例如与四聚体单体融合),或者通过与蛋白质结构域融合 (所述蛋白质结构域与 DNA、细胞壁自身或这二者结合)。通过在经透化细胞的细胞壁内共保留质粒和基因组 DNA 提供对于展示系统所需的表型 - 基因型连锁。

[0227] 多肽:

[0228] 人抗体谱含有功能性和假基因可变区 (由 Lefranc 总结, 2000)。这些可以作为外显子从非免疫谱系中基因组 DNA 或者从免疫细胞 (其已经经历了 V(D)J 重排) 来源的 mRNA 克隆,以准备可以用于表达抗体的遗传构建体。在这样的过程期间,轻链和重链的可变结构

域可以作为单体 scFv 或以形成二价或更高价的排列被克隆。还可以从可变结构域的下游克隆恒定区以产生 Fab 或全长抗体。

[0229] 在所有形式中,为了获得正确折叠并且在抗体的产生期间维持稳定性和可溶性,编码抗体的遗传构建体必须几乎总是在使得结构域内二硫键可以在 β 片层之间形成的条件下(即在非还原性条件下)表达。这样,在哺乳动物中,抗体插入内质网(ER)和高尔基体用于分泌或膜插入。如果在细菌宿主(例如大肠杆菌)中表达,其必须引导至二硫键伴侣 DsbA、B 和 C 存在的周质空间。如果抗体在非氧化性环境中(例如在细胞的胞质中),稳定化二硫键的缺乏导致错误折叠和降解,或者如果在大肠杆菌胞质中以高水平表达,作为亚细胞包涵体(inclusion body)而聚集。

[0230] 本发明人已经鉴别了包含抗体可变区支架的多肽,其甚至在多肽于非氧化(还原)性环境中表达时能够形成抗原结合位点。

[0231] 因此,本发明提供了分离的和/或重组的多肽,所述多肽包含通过肽接头与免疫球蛋白可变结构域之 $V_L \lambda$ 1、3 或 6 家族的抗体轻链可变区(V_L)相连的免疫球蛋白可变结构域之 V_H 3 家族的抗体重链可变区(V_H),其中所述 V_H 和所述 V_L 能够形成抗原结合位点。

[0232] 本发明的多肽可以以任意已知形式的抗体或抗体片段来提供。因而,本发明的多肽可以是:(i) 抗体;(ii) 单一结构域抗体;(iii) 单链 Fv(scFv);(iv) 双抗体、三抗体、四抗体;(v) 融合蛋白,其包含(ii)-(iv)中任一种和抗体的 Fc 结构域或其结构域;(vi) 融合蛋白,其包含(ii)-(iv)中任一种和能够与免疫效应细胞或任意其他已知形式的抗体相结合的蛋白质。

[0233] 优选地,本发明的多肽是 Fv。例如,本发明的多肽优选地是 Fab 片段、Fab' 片段、F(ab') 片段、scFv、双抗体、三抗体、四抗体或更高阶的多肽复合物。

[0234] 最优选地,本发明的多肽是 scFv。scFv 包含在单多肽链中的 V_H 和 V_L 区。优选地,所述多肽链还在 V_H 和 V_L 之间还包含多肽接头,其使得 scFv 能够形成期望的用于抗原结合的结构(即,用于所述单多肽链的 V_H 和 V_L 彼此相互缔合以形成抗原结合位点)。这与本发明的双抗体或更高阶的多聚物截然不同,其中来自不同多肽链的可变区彼此相互缔合或结合。所述肽接头可以包含 12 个或更多个氨基酸残基。例如,所述肽接头可以包含 12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30 个氨基酸或更多个。优选地,所述肽接头包含超过 12 个氨基酸残基,且 $(Gly_4Ser)_3$ (即 GGGSGGGSGGGGS(SEQ ID NO:28))是用于 scFv 的更偏好的接头之一。另一些合适的多肽在本领域中是已知的。本发明的多肽优选地包含免疫球蛋白可变区之 V_H 3 家族的 V_H 的支架区和/或免疫球蛋白可变结构域之 $V_L \lambda$ 1、3 或 6 家族的 V_L 的支架区。这样,本发明的多肽优选地包含本文中公开的任意可变区的所有氨基酸残基,不包括 CDR 残基。CDR 残基可以由本领域技术人员容易地鉴定,参考 Kabat(1987 和/或 1991)、Bork 等(1994)和/或 Chothia 和 Lesk(1987 和 1989)或 Al-Lazikani 等(1997)中的讨论。这样本发明的多肽可包含本文中公开的任意可变区的所有 FR。所述多肽还可以包含一个或更多个本文中公开的可变区之 CDR。所述多肽还可以包含一个或更多个未在本文中公开的可变区中存在的 CDR。因此,一个或更多个来自不同来源的 CDR 可以插入到本文中公开的可变区的支架区。这种可能性的进一步讨论包含于下文中。

[0235] 在一个优选的实施方案中,本发明的 scFv 包含通过肽接头与免疫球蛋白可变结

构域之 V_L λ 1、3 或 6 家族的 V_L 相连接的免疫球蛋白可变结构域之 V_H 3 家族的 V_H 的支架区。在更优选的一些实施方案中,本发明的 scFv 包含 IGHV3-23 的支架区和 IGLV1-40、IGLV1-44、IGLV1-47、IGLV1-51、IGLV3-1、IGLV3-19、IGLV3-21 和 IGLV6-57 中任一种的支架区。最优选地,本发明的 scFv 包含 IGHV3-23 的支架区和 IGLV3-1 的支架区。

[0236] 本发明的多肽可以用其与参考序列的百分比同一性来定义。这一百分比同一性可以通过本领域中已知的任意合适的方法计算。用于比较经比对序列的数种算法是已知的,并且可以用于确定本发明之多肽与参考序列的百分比同一性。例如,氨基酸和多核苷酸序列可以手动比较或通过利用基于计算机的序列比较和鉴定工具进行比较,其使用例如 BLAST (Basic Local Alignment Search Tool ;Altschul 等,1993) 的算法;还参见 www.ncbi.nlm.nih.gov/BLAST/),比对的 Clustal 方法 (Higgins and Sharp,1989) 等,其中用于每个特定序列比较的合适的参数可按照本领域技术人员将会理解的方式来选择。

[0237] 优选地,本发明的多肽是分离的和 / 或重组的多肽。本文中使用的术语“分离的”或“纯化的”旨在意指这样的多肽,其通常已经从脂质、核酸、其他多肽和肽、以及在其天然状态中与其相关的污染分子分离。优选地,所述分离的多肽至少 60% 没有、优选地至少 75% 没有、以及更优选地至少 90% 没有与其天然相关的组分。

[0238] 在多肽的上下文中,术语“重组”是指这样的多肽,当通过细胞或在无细胞表达系统中产生时,所述多肽与天然状态相比具有改变的量或改变的速率。在一个实施方案中,所述细胞是不天然产生所述多肽的细胞。然而,所述细胞可以是包含非内源性基因(其导致改变的,优选提高量的待产生之肽)的细胞。如本文中所描述的重组多肽包括未从转基因(重组)细胞或无细胞表达系统(所述多肽在其中产生)的其他组分分离的多肽,并且在这样的细胞或无细胞体系中产生的多肽随后被纯化以除去至少一些其他组分。

[0239] 本发明的多肽优选地包含来自鼠(小鼠或大鼠)抗体或灵长类(优选人)抗体的氨基酸序列。因此在本发明的多肽中包含的可变区和 / 或支架区可以是鼠(小鼠或大鼠)或灵长类(优选人)可变区和 / 或支架区。

[0240] 优选地,本发明的多肽是可溶性的。用于确定多肽之可溶性的方法是本领域中公知的,例如 J. Sambrook 等, *Molecular Cloning :A Laboratory Manual*, 3rd 版, Cold Spring Harbour Laboratory Press (2001) 中的描述。多肽如果(例如)不能从经裂解的和 / 或经透化的细胞级分中通过物理分离(例如通过离心)分离,则所述多肽可以被确定为是可溶性的。此外,如果本发明的多肽在细胞的胞质中不形成包涵体,则所述多肽可以被确定为是可溶性的。因此,如果当多肽在宿主细胞中表达时,所述多肽保留在通过任何合适的机械、去污剂和 / 或酶方法裂解宿主细胞后产生的可溶性级分中,则所述多肽可以被认为可溶性的。合适的机械方法包括例如使用声处理。合适的去污剂方法包括例如使用 n -辛基- β -D-硫代葡萄糖苷(8TGP)。合适的酶方法包括例如使用溶菌酶。优选地,本发明的多肽可在细胞裂解物的可溶性级分中保留的水平为至少 25%,例如至少 50%、至少 75%、至少 90%、至少 95%、或至少 95%。

[0241] 本发明的多肽优选地能够稳定地形成抗原结合位点。因此,所述多肽优选地能够在足以允许检测多肽抗原复合物的水平上与靶抗原相结合。这样的检测可以在合适的实验条件下进行,例如在温度为至少 5°C、至少 10°C、至少 15°C、至少 20°C、至少 25°C、至少 30°C、至少 35°C、至少 40°C、至少 45°C 或至少 50°C。

[0242] 缀合物

[0243] 本发明的多肽可以使用本领域中已知的任意方法与一个或更多种化合物缀合。多肽可以与之缀合的化合物实例选自：放射性同位素、可检测的标签、治疗性化合物、胶体、毒素、核酸、肽、蛋白质、提高所述多肽在对象中之半衰期的化合物及其混合物。示例性治疗剂包括但不限于抗血管发生剂 (anti-angiogenic agent)、抗新血管形成剂 (anti-neovascularization) 和 / 或其他血管形成剂 (vascularization agent)、抗增殖剂、促凋亡剂、化疗剂或治疗性核酸。

[0244] 毒素包括对细胞有害的 (例如杀死细胞) 任意试剂。多这些类药物的描述是本领域中已知的, 并且其作用的机理参见 Goodman 等, Goodman 和 Gilman' s The Pharmacological Basis of Therapeutics, 第 8 版, Macmillan Publishing Co., 1990。与制备免疫球蛋白 - 免疫毒素缀合物相关的另一些技术在例如 Vitetta(1993) 和 US5, 194, 594 中提供。示例性毒素包括白喉 A 链、白喉毒素的非结合活性片段、外毒素 A 链 (来自铜绿假单胞菌 (*Pseudomonas aeruginosa*))、蓖麻毒蛋白 (ricin) A 链、相思豆毒蛋白 (abrin) A 链、蒴莲根毒蛋白 (modeccin) A 链、 α -帚曲霉素 (alfa-sarcin)、油桐 (*Aleurites fordii*) 蛋白质、香石竹毒蛋白 (dianthin protein)、美洲商陆 (*Phytolaca americana*) 蛋白质 (PAPI、PAPII 和 PAP-S), 苦瓜 (*Momordica charantia*) 抑制因子、麻疯树毒蛋白 (curcin)、巴豆毒蛋白 (croton)、肥皂草 (*Saponaire officinalis*) 抑制因子、多花白树毒蛋白 (gelonin)、米托菌素 (mitogellin)、局限曲菌素 (restrictocin)、酚霉素 (phenomycin)、依诺霉素 (enomycin) 和单端孢霉烯类 (tricothecenes)。参见例如 W093 / 21232。

[0245] 用于形成包含本发明之多肽的免疫缀合物的合适的化疗剂包括: auristatin 和美登素 (maytansine)、泰素 (taxol)、细胞松弛素 B (cytochalasin B)、短杆菌肽 D (gramicidin D)、溴化乙锭、吐根碱 (emetine)、丝裂霉素、依托泊苷、替尼泊苷、长春新碱、长春碱、秋水仙碱、多柔比星、柔红霉素、二羟基蒽醌二酮 (dihydroxy anthracin dione)、米托蒽醌 (mitoxantrone)、光神霉素、放线菌素 D、1-去氢睾酮 (1-de-hydrotestosterone)、糖皮质激素、普鲁卡因、丁卡因、利多卡因、普萘洛尔和嘌呤霉素、抗代谢物类 (例如甲氨蝶呤、6-巯基嘌呤、6-硫鸟嘌呤、阿糖胞苷、氟达拉滨、5-氟尿嘧啶、达卡巴嗪、羟基脲、天冬酰胺酶、吉西他滨、克拉屈滨)、烷化剂 (例如氮芥、塞替派、苯丁酸氮芥、美法仑、卡莫司汀 (BSNU)、洛莫司汀 (CCNU)、环磷酰胺、白消安、二溴甘露醇、链脲霉素、达卡巴嗪 (DTIC)、丙卡巴肼)、丝裂霉素 C、顺铂和其他铂衍生物例如卡铂)、抗生素类 (例如更生霉素 (以前称为放线菌素)、博来霉素、柔红霉素 (以前称为道诺霉素)、多柔比星、伊达比星、光神霉素、丝裂霉素、米托蒽醌、普卡霉素 (plicamycin)、氨基霉素 (AMC))。

[0246] 合适的血管发生抑制剂 (抗血管发生剂) 的实例包括但不限于: 尿激酶抑制剂、基质金属蛋白酶抑制剂 (例如马立马司他 (marimastat)、新伐司他 (neovastat)、BAY12-9566、AG3340、BMS-275291 和类似药剂)、内皮细胞迁移和增殖的抑制剂 (例如 TNP-470、角鲨胺 (squalamine)、2 甲氧基雌二醇、康普瑞汀 (compretastatins)、内皮抑素 (endostatin)、血管抑素、青霉胺、SCH66336 (Schering-Plough Corp, Madison, NJ)、R115777 (Janssen Pharmaceutica, Inc, Titusville, NJ) 和类似药剂)、血管发生性生长因子拮抗剂 (例如 ZD6474、SU6668、针对血管发生剂的抗体和 / 或其受体 (例如 VEGF、bFGF 和

血管生成素-1)、沙立度胺、沙立度胺类似物(例如 CC5013)、Sugen5416、SU5402、抗血管发生核酶(例如 angiozyme)、干扰素 α (例如干扰素 α 2a)、苏拉明(suramin)和类似药剂)、VEGF-R 激酶抑制剂和其他抗血管发生的酪氨酸激酶抑制剂(例如 SU011248)、内皮细胞特异性整合素/生存信号的抑制剂(如 vitaxin 和类似药剂)、铜拮抗剂/螯合剂(如四硫钼酸盐(tetrathiomolybdate)、卡托普利和类似药剂)、羧胺三唑(carboxyamido-triazole, CAI)、ABT-627、CM101、白介素-12(IL-12)、IM862、PNU145156E 以及抑制血管发生的核苷酸分子(例如反义-VEGF-cDNA、编码血管抑素的 cDNA、编码 p53 的 cDNA 和编码缺陷型 VEGF 受体-2 的 cDNA)和类似药剂。血管发生、新血管形成和/或其他血管形成之抑制剂的另一些实例是抗血管发生的肝素衍生物和相关分子(例如,肝素酶 III(heparinase III))、替莫唑胺、NK4、巨噬细胞迁移抑制因子(MIF)、环氧合酶-2 抑制剂、低氧诱导因子 1 的抑制剂、抗血管发生的大豆异黄酮类、奥替普拉(oltipraz)、烟曲霉素及其类似物、生长抑素类似物、多硫酸戊聚糖(pentosan polysulfate)、替可加兰钠(tecogalan sodium)、达肝素(dalteparin)、肿瘤抑素(tumstatin)、血小板反应蛋白(trombospondin)、NM-3、卡贝塔汀(combrestatin)、血管能抑素(canstatin)、阿瓦斯汀(avastatin)、针对其他相关靶标的抗体(例如抗 α -v / β -3 整合素和抗高分子激肽原(kininostatin)mAb)和类似药剂。

[0247] 在一个实例中,如本文所描述的根据任意实施方案的多肽与其他多肽缀合或连接,所述其他多肽包括本发明的其他多肽或者包含免疫球蛋白可变区的多肽,例如免疫球蛋白或从其衍生的多肽(例如,如本文中所述的多肽)。不排除其他蛋白质。另外的蛋白质对技术人员来说将是显而易见的,包括例如免疫调节剂或延长半衰期的蛋白质或多肽或者与血清白蛋白相结合的其他蛋白质等。

[0248] 示例性的免疫调节剂包括细胞因子和趋化因子。术语“细胞因子”是从一个细胞群释放的、在另一细胞发挥作用的、作为细胞间介质的蛋白质或肽的通用术语。细胞因子的实例包括淋巴因子、单核因子、生长因子和传统的多肽激素。细胞因子中包括生长激素,例如人生长激素、N-甲硫氨酰基人生长激素、和牛生长激素;甲状旁腺激素、甲状腺素、胰岛素、胰岛素原、松弛素、松弛素原、糖蛋白激素例如促卵泡激素(FSH)、促甲状腺激素(TSH)和促黄体激素(LH)、肝生长因子;前列腺素、成纤维细胞生长因子、催乳素、胎盘催乳激素、OB 蛋白、肿瘤坏死因子- α 和肿瘤坏死因子- β , 缪勒氏抑制物质(mullerian-inhibiting substance)、促性腺激素相关肽、抑制素(inhibin)、激活蛋白(activin)、血管内皮生长因子、整合素、血小板生成素(TPO)、神经生长因子例如 NGF-B、血小板生长因子、转化生长因子(TGF)例如 TGF- α 和 TGF- β 、胰岛素样生长因子-I 或胰岛素样生长因子-II、促红细胞生成素(EPO)、骨诱导因子、干扰素如干扰素- α 、干扰素- β 或干扰素- γ ;集落刺激因子(CSF)例如巨噬细胞-CSF(M-CSF)、粒细胞-巨噬细胞 CSF(GM-CSF) 和粒细胞-CSF(G-CSF), 白介素(IL)例如 IL-1、IL-1 α 、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12、IL-13、IL-14、IL-15、IL-16、IL-17、IL-18、IL-21 和 LIF。

[0249] 趋化因子通常作为化学引诱物来将免疫效应细胞募集至趋化因子表达的部位。趋化因子包括但不限于 RANTES、MCAF、MIP1- α 或 MIP1- β 。技术人员将意识到已知某些细胞因子具有化学引诱作用并且可以归类为术语趋化因子。

[0250] 示例性血清白蛋白结合肽或蛋白质描述于 US20060228364 或 US20080260757。

[0251] 多种放射性核素可用于放射性缀合蛋白的产生。实例包括但不限于低能量辐射原

子核（例如适合用于诊断目的），例如 ^{13}C 、 ^{15}N 、 ^2H 、 ^{125}I 、 ^{123}I 、 ^{99}Tc 、 ^{43}K 、 ^{52}Fe 、 ^{67}Ga 、 ^{68}Ga 、 ^{111}In 等。优选地，放射性核素是发射 γ 、光子或正电子的放射性核素，具有半衰期适合于在施用和定位至成像部位之间的时间流逝之后允许放射性或检测。本发明还涵盖高能放射性核（例如用于治疗目的），例如 ^{125}I 、 ^{131}I 、 ^{123}I 、 ^{111}In 、 ^{105}Rh 、 ^{153}Sm 、 ^{67}Cu 、 ^{67}Ga 、 ^{166}Ho 、 ^{177}Lu 、 ^{186}Re 和 ^{188}Re 。这些同位素通常产生具有短的路径长度的高能 α -或 β -粒子。这样的放射性核素杀死与其非常靠近的细胞，例如缀合物已经附着或者进入的肿瘤细胞。它们对未定位的细胞影响很小的或没有影响并且基本上没有免疫原性。或者，高能同位素可以通过其他稳定同位素的热辐射产生，例如在硼中子俘获治疗中（Guan 等，1998）。

[0252] 在另一个实施方案中，所述多肽与用于在细胞中预靶标的“受体”（例如链霉亲和素）缀合，其中缀合物施用于患者，随后通过使用清除剂从循环移除未结合的缀合物并且随后施用与治疗剂（例如放射性核素）缀合的“配体”（例如亲和素）。

[0253] 本发明的蛋白质可以被修饰以含有另外的非蛋白质部分，其在本领域中是已知的并且可容易地获得。优选地，适合用于蛋白质之衍生的部分是水溶聚合物。水溶性聚合物的非限制性实例包括但不限于聚乙二醇（PEG）、聚乙烯醇（PVA）、乙二醇 / 丙二醇的共聚物、羧甲基纤维素、葡聚糖、聚乙烯醇、聚乙烯吡咯烷酮、聚-1,3-二氧戊环（dioxolane）、聚-1,3,6-三噁烷、乙烯 / 马来酸酐共聚物、聚氨基酸（无论是均聚物或随机共聚物）、以及聚（n-乙烯吡咯烷酮）聚乙二醇、丙二醇（PPG）均聚物、聚环氧丙烷 / 环氧乙烷共聚物、聚氧乙烯化多元醇（例如，甘油、POG）、聚乙烯醇、及其混合物。聚乙二醇丙醛由于其在水中的稳定性可以在生产中具有优势。

[0254] 聚合物分子通常特征为具有例如约 2 个至约 1000 个或约 2 个至约 300 个重复单元。

[0255] 例如，水可溶性聚合物（包括但不限于 PEG、聚（氧化乙烯）（PEO）、聚氧乙烯（POE）、聚乙烯醇、羟乙基纤维素或葡聚糖）通常与蛋白质缀合以提高蛋白质的稳定性或尺寸等。

[0256] PEG、PEO 或 POE 是指环氧乙烷的寡聚物或聚合物。在聚乙二醇的情况下，这些寡聚物或聚合物通过例如环氧乙烷的阴离子开环聚合产生，这是与通过氢氧根离子在环氧化物环上的亲核攻击。用于蛋白质修饰的 PEG 之更加有用的形式之一是单甲氧基 PEG（mPEG）。

[0257] 用于与本发明的多肽缀合的特别优选的化合物列在表 1 中。

[0258] 表 1：用于缀合的优选化合物

[0259]

基团	详情
放射性同位素(无论直接或间接)	^{123}I、^{125}I、^{130}I、^{133}I、^{135}I、^{47}Sc、^{72}As、^{72}Sc、^{90}Y、^{88}Y、^{97}Ru、^{100}Pd、$^{101\text{m}}\text{Rh}$、$^{101\text{m}}\text{Rh}$、^{119}Sb、^{128}Ba、^{197}Hg、^{211}At、^{212}Bi、^{153}Sm、^{169}Eu、^{212}Pb、^{109}Pd、^{111}In、^{67}Ga、^{68}Ga、^{67}Cu、^{75}Br、^{76}Br、^{77}Br、$^{99\text{m}}\text{Tc}$、^{11}C、^{13}N、^{15}O、^{18}F、^{188}Re、^{203}Pb、^{64}Cu、^{105}Rh、^{198}Au、^{199}Ag 或 ^{177}Lu
半衰期延长剂	- 聚乙二醇 - 甘油 - 葡萄糖
荧光探针	- 藻红蛋白 (phycoerythrin, PE) - 别藻蓝蛋白 (allophycocyanin, APC) - Alexa Fluor 488 - Cy5.5
生物制剂	- 荧光蛋白例如海肾萤光素酶, GFP - 免疫调节剂 - 毒素 - 免疫球蛋白 - 半衰期延长剂例如白蛋白
化疗剂	- 泰素 5-氟尿嘧啶 - 多柔比星 - 伊达比星

[0260] 在本发明的一个实例中,间隔区 (spacer) 部分包括在其所缀合的化合物和多肽之间。本发明的间隔区部分可以是可切割的或不可切割的。例如,间隔区部分可以是氧化还原可切割的间隔区部分,使得所述间隔区部分在具有较低的氧化还原电位的环境(例如在胞质和其他具有更高浓度的带有游离巯基之分子的区域)中是可切割的。由于氧化还原电位的变化而可以被切割的隔离区部分的实例包括含有二硫化物的那些。切割刺激可以根据缀合蛋白的细胞内摄取来提供,其中胞质的较低氧化还原电位有利于切割间隔区部分。

[0261] 在另一个实施方案中,pH 降低导致间隔区的切割,从而将化合物释放到靶细胞中。pH 降低牵涉到许多生理和病理过程,如内体运输、肿瘤生长、炎症和心肌缺血。pH 从生理的 7.4 下降至内体中的 5 至 6 或溶酶体中的 4 至 5。可以用于靶向癌细胞之溶酶体或内体的酸敏感性间隔区部分的实例包括具有酸可切割键的那些,例如见于缩醛、缩酮、原酸酯类、脞类、三苯甲基类、顺-乌头基类 (cis-aconyls)、或硫代氨基甲酰基类 (thiocarbamoyls) 中的那些(参见实例美国专利 No. 4,569,789、4,631,190、5,306,809 和 5,665,358)。

[0262] 可切割的间隔区部分可以对生物学上供应的与特定靶细胞相关的切割剂(例如溶酶体或肿瘤相关酶)敏感。可以被酶促切割的连接部分的实例包括但不限于肽或脂。示例性酶可切割的连接部分包括对肿瘤相关蛋白酶(例如组织蛋白酶 B (cathepsin B) 或纤溶酶 (plasmin)) 敏感的那些。组织蛋白酶 B 可切割的位点包括二肽序列缬氨酸-瓜氨酸、

苯丙氨酸-赖氨酸和/或缬氨酸-丙氨酸。

[0263] 蛋白质复合物

[0264] 本发明的多肽优选地与一个或更多个化合物缀合,这使其特别适合用于本文中提及的 RED 测定。例如,所述多肽可以与至少第二多肽(下文中称为“第二多肽”)相缔合以形成蛋白质复合物,所述蛋白质复合物具有使得所述蛋白质复合物保留在经透化细菌细胞内的分子量。所述多肽可以与所述第二多肽相缔合,通过例如共价键(例如二硫桥)或通过非共价缔合。“非共价键缔合”是指不涉及原子间键的分子相互作用。例如,非共价相互作用涉及离子键、氢键、疏水相互作用、以及范德华力。非共价力可用于在蛋白质或蛋白质复合物中将分离的多肽链连在一起。因此,所述多肽和第二多肽可以作为分离的多肽由相同或不同的载体表达,或者所述多肽中的一种或两种都可以由已经整合到细菌细胞基因组中的编码所述多肽的 DNA 表达。

[0265] 或者,在蛋白质复合物中相缔合的所述多肽和第二多肽可以是融合蛋白。本文中使用的“融合蛋白”是指杂合蛋白,其由两种或更多种多肽或其片段组成,通过表达编码两种多肽中每一种之至少一部分的多核苷酸而产生。

[0266] 蛋白质复合物通过分子大小保留在经透化细菌细胞中

[0267] 所述第二多肽可以是具有足够的分子大小(即足够的分子量或分子半径)的任意多肽,使得至少一些与用于筛选期望活性的多肽形成的复合物不能从经透化的细菌细胞扩散。因此,所述蛋白质复合物在所述细胞的透化之后保留在细菌细胞内。本领域技术人员将意识到第二多肽的性质(包括其分子量以及其是否是球状或杆状(丝状)蛋白)将确定其阻止或抑制蛋白质复合物通过细菌细胞壁扩散。在一个实施方案中,所述第二多肽的分子量为至少约 30kDa 或至少约 40、50、60、70、80、90、100、120、130、140、150 或更大的 kDa。在一个实施方案中,所述第二多肽为至少约 120kDa。

[0268] 在一个实施方案中,所述第二多肽形成具有大于经透化细菌细胞排除孔大小之分子大小的多聚体。本文中使用的术语“多聚体”及其语法上的变化形式是指两个或更多个不同分子之间形成多聚体复合物。所述多聚体可以包含例如两个或更多个相同蛋白质的分子(即同多聚体)或者两个或更多个不同或非同一蛋白质的混合物(即异多聚体)。适用于在本发明的方法中形成多聚体的蛋白质包括形成二聚体、三聚体、四聚体、五聚体、六聚体以及包含七个或更多个亚单位的更高阶多聚体的那些。

[0269] 多聚体蛋白质包括同二聚体,例如 PDGF 受体 α 和 β 同种型、促红细胞生成素受体、MPL 和 G-CSF 受体;异二聚体,其每个亚单位具有配体结合和效应子结构域,例如 PDGF 受体 $\alpha\beta$ 同种型;以及具有带完全不同功能的组成亚单位的多聚体,例如 IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、和 GM-CSF 受体。可以在本发明的方法中使用的其他多聚体蛋白质的非限制性实例包括参与 DNA 的合成和复制的因子,例如参与产生 mRNA 的 DNA 聚合酶蛋白,例如 TFIID 和 TFIIH;细胞,核及其他膜相关的蛋白质,例如激素和其他信号转导受体、主动转运蛋白和离子通道,血液中的多聚体蛋白,包括血红蛋白、纤维蛋白原和冯维勒布兰德因子(von Willabrand's factor);在细胞内形成结构的蛋白质,例如肌动蛋白、肌球蛋白、以及微管蛋白和其他细胞骨架蛋白;在细胞外环境中形成结构的蛋白质,例如胶原蛋白、弹性蛋白和纤连蛋白;参与细胞内和细胞外转运的蛋白质,例如驱动蛋白(kinesin)和动力蛋白(dynein),SNARE 家族的蛋白质(可溶性 NSF 附着蛋白受体)和网

格蛋白 (clathrin) ;有助于调节染色质结构的蛋白质,例如组蛋白和精蛋白、Swi3p、Rsc8p 和 moira ;多聚体转录因子例如 Fos、Jun 和 CBTF (CCAAT 盒转录因子) ;多聚体酶,例如乙酰胆碱酯酶和乙醇脱氢酶 ;伴侣蛋白例如 GroE、Gro EL (伴侣蛋白 60) 和 Gro ES (伴侣蛋白 10) ;抗毒素,例如蛇毒、肉毒中毒 (botulism) 毒素、链球菌 (Streptococcus) 超抗原 ;溶素 (lysin) (来自噬菌体和病毒的酶) ;以及大多数变构蛋白质。在一个实施方案中,多聚体蛋白是大肠杆菌蛋白质。形成多聚体的大肠杆菌蛋白质的非限制性实例包括 L- 鼠李糖异构酶 (RhnA, 例如 NCBI 登录 CAA43002)、 β - 半乳糖苷酶 (β -gal, 例如 NCBI 登录 YP001461520)、甜菜碱醛脱氢酶 (BetB, 例如 NCBI 登录 AAA23506)、谷氨酸-5- 激酶 (G5K, 例如 NCBI 登录 AAB08662)、谷胱甘肽合成酶 (GshB, 例如 NCBI 登录 AP_003504)、以及中链醛脱氢酶 (YdcW, 例如 NCBI 登录 AP_002067)。

[0270] 在一个实施方案中,所述多肽具有足以使所述多肽保留在细菌细胞壁内的分子大小。因此,本领域的技术人员将意识到,为了将所述多肽保留在经透化的细菌细胞内,这样的多肽不一定与第二蛋白质相缔合。

[0271] DNA 结合蛋白

[0272] 本发明人已经发现, DNA 在透化后保留在细菌细胞内。因此,在一个实施方案中,所述多肽与 DNA 结合蛋白缔合以形成蛋白质复合物,所述蛋白质复合物结合 DNA 并且在保留在细菌细胞内部。本文中使用的“DNA 结合蛋白”是指包含 DNA 结合结构域的任意蛋白质,所述 DNA 结合结构域包含至少一个识别双链或单链 DNA 的模体 (motif)。如本领域技术人员已知的, DNA 结合结构域包括螺旋-转角-螺旋、锌指、亮氨酸拉链、翼状螺旋 (winged helix)、翼状螺旋转角螺旋、螺旋-环-螺旋、识别 DNA 的免疫球蛋白折叠、或 B3 结构域。多肽与 DNA 结合蛋白的缔合有利地提供了提高的 DNA (例如编码本发明筛选方法中之多肽的质粒) 回收。

[0273] DNA 结合蛋白的实例包括细菌感受态蛋白,例如但不限于 :大肠杆菌 DNA 结合蛋白、淋病奈瑟菌 (*Neisseria gonorrhoeae*) DNA 结合蛋白 (例如 ComE)、腺病毒 E2 蛋白、AraC 转录因子、碱性螺旋-环-螺旋转录因子、碱性亮氨酸拉链转录因子、丁酸盐响应因子、着丝粒蛋白 B、COUP 转录因子、早期生长反应转录因子、G 盒 (G-box) 结合因子、GATA 转录因子、HMGA 蛋白、同源结构域蛋白、I- κ B 蛋白、整合宿主因子、干扰素调节因子、干扰素刺激基因因子 3、Kruppel 样转录因子、亮氨酸响应调节蛋白、基质附着区结合蛋白、甲基-CpG 结合蛋白、MutS 同源物 2 蛋白、髓性-淋巴性白血病蛋白、NF- κ B、NF1 转录因子、核呼吸因子 (nuclear respiratory factor)、癌基因蛋白 p55、起点识别复合物、成对盒转录因子、POU 结构域因子、原癌基因因子、Rad51 重组酶、Rad52DNA 修复和重组蛋白、复制蛋白 A、复制蛋白 C、视网膜母细胞瘤蛋白、Smad 蛋白、SOX 转录因子、T 盒结构域蛋白、TCF 转录因子、端粒结合蛋白、To11 样受体 9、反式激活蛋白、以及翼状螺旋转录因子。在一个实施方案中, DNA 结合蛋白是大肠杆菌 DNA 结合蛋白。在另一个实施方案中, DNA 结合蛋白是淋病奈瑟菌蛋白,例如 ComE。

[0274] 细胞壁结合蛋白

[0275] 所述多肽可以与细菌细胞壁结合蛋白相缔合。技术人员将理解,细胞壁结合蛋白的选择将取决于宿主细胞种类,因为不同的细菌具有不同的细胞壁组成。尽管细菌具有由肽聚糖 (PG) 构成的细胞壁,但是物种之间的化学修饰可影响跨物种结合。技术人员将能够

容易地确定适合用于在特定细菌种类中的细胞壁结合蛋白。

[0276] 细菌细胞壁结合蛋白包括已知具有结构域结构的蛋白质，从而使得天然结构中所述多肽链的一部分能够识别并结合细菌细胞壁上的特定分子或分子构象。因此，术语“细菌细胞壁结合蛋白”包括作为与细菌细胞壁特异性结合的蛋白质之一部分的蛋白质结构域。细菌细胞壁结合蛋白的实例包括由噬菌体编码的细胞壁水解酶、细菌的细胞壁水解酶和不同的自溶素 (autolysin)。还涵盖由噬菌体和其他病毒的 DNA 编码的受体分子。当细菌细胞壁结合蛋白来自噬菌体来源的水解酶（其能够与细菌特异性结合）时，所述细菌细胞壁结合蛋白保持其结合能力，但是优选地不具有显著的水解活性。

[0277] 在一个实施方案中，细胞壁结合蛋白与大肠杆菌的细胞壁非共价结合。例如，对于大肠杆菌宿主细胞，有内源 PG 结合蛋白，所述 PG 结合蛋白具有发生在 PAL、OmpA、YiaD、YfiB 和 MotB 中的保守的约 100 个氨基酸的 PG 结合结构域 (Parsons 等, 2006)。然而，已经表明，来自另一些生物体的蛋白质在大肠杆菌中很好的表达并且与细胞壁高亲和力和地结合，例如，来自假单胞菌 (*Pseudomonas*) ϕ KZ 噬菌体 (KzPG) 的约 70 个氨基酸的 PG 结合结构域 (Briers 等, 2009)。因此，可以将来自于结合 PG 之蛋白质的 PG 结合结构域用作本发明方法中的细菌细胞壁结合蛋白。

[0278] 在一个示例性实施方案中，所述 PG 结合结构域可以与本发明的多肽融合，并且在细菌细胞的胞质溶胶中表达。在膜透化之后，PG 结合结构域与细胞壁结合，导致目的多肽保留在经透化细胞内。本领域技术人员将理解，为了潜在地进一步增强目的多肽在细胞内的保留，多肽可以与细菌细胞壁结合蛋白之外的 DNA 结合蛋白缔合。

[0279] 或者，所述多肽可以与能够与细菌细胞壁共价连接的蛋白质缔合。优选地，所述蛋白质包含周质靶向信号。因此，所述多肽在细菌细胞的胞质溶胶中表达，但是靶向周质，其中所述多肽在膜透化之前与细胞壁连接。

[0280] 通过非限制性实例，共价附着到细胞壁上的细菌细胞壁结合蛋白可以是能够与细胞壁结合的脂蛋白，并且其缺少用于外膜附着所必需的功能性 N 端信号序列。例如，脂蛋白可以是大肠杆菌 LPP。LPP 是一种丰富的大肠杆菌蛋白质，其形成三聚体的卷曲螺旋。在其天然形式中，一端通过脂化束缚于外膜，另一端通过 C 端赖氨酸与细胞壁共价结合。脂蛋白还可以包含使脂蛋白靶向周质的序列，例如 OmpF 周质靶向序列。在一个实施方案中，脂蛋白是缺少用于外膜附着所必需的功能性 N 端信号序列的大肠杆菌脂蛋白。

[0281] 在本说明书的教导下，本领域技术人员将能够鉴定或设计与细胞壁共价连接并且适用于本发明方法的蛋白质。

[0282] 在本发明的一个实施方案中，本发明的多肽是融合多肽，其包含 KzPG 结构域和选自间隔区、SNAP 和 / 或 DBP 的一个或更多个其他结构域。在一个特别的实施方案中，所述融合多肽包含一个或更多个间隔区以及 KzPG、SNAP 和 DBP 结构域。

[0283] 多核苷酸

[0284] 本发明还提供了编码本发明之多肽的多核苷酸。优选地，所述核苷酸是分离的或重组的多核苷酸。

[0285] 术语“分离的多核苷酸”旨在意指通常已经从在其天然状态中与其缔合或连接的多核苷酸序列分离的多核苷酸序列。优选地，所述分离的多核苷酸至少 60% 不含、至少 75% 不含、并且更优选至少 90% 不含与其天然缔合的其他组分。此外，本文中术语“多核苷

酸”与术语“核酸分子”、“基因”和“mRNA”可互换使用。

[0286] 术语“重组”在多核苷酸的上下文中是指与其天然状态相比以改变的量存在于细胞中或无细胞表达系统中的多核苷酸。在一个实施方案中,所述细胞是不天然包含所述多核苷酸的细胞。然而,所述细胞可以是包含非内源性多核苷酸导致所编码多肽之量改变(优选提高)的细胞。本发明的重组多核苷酸包括已经从其所存在于的转基因(重组)细胞或无细胞表达系统中分离的多核苷酸,并且在这样的细胞或无细胞表达系统中产生的多核苷酸随后被纯化,以除去至少一些其他组分。

[0287] “多核苷酸”是指寡核苷酸、多核苷酸或其任意片段。其可以是基因组或人工来源的 DNA 或 RNA,双链或单链,并且与碳水化合物、脂质、蛋白质或其他材料组合以实施本文中所限定的特定活性。

[0288] 可以使用本技术领域中的标准方法来分离编码包含可变区之多肽的 DNA。例如,引物可以设计为与在目的区侧翼的可变区内的保守区退火,并且这些引物可以随后用于扩增插入的核酸,例如通过 PCR。合适的方法和/或引物在本领域中是已知的和/或描述于例如 Borrebaeck(编辑),1995 和/或 Froyen 等,1995。用于这样的扩增方法的模板 DNA 的合适来源可以来自例如杂交瘤、转染瘤和/或表达包含可变区之蛋白质的细胞(例如本文中的描述)。

[0289] 本发明的多核苷酸可以编码本发明的完整多肽。或者,所述多核苷酸可以编码本发明之多肽的重链或轻链。因此,各自编码重链或轻链之一的两种多肽可以在单个细胞中产生和表达以产生本发明的多肽。

[0290] 优选地,所述多核苷酸编码可变区的支架区以及一个或更多个 CDR。最优选地,本发明的多核苷酸编码可变区的支架区和所有三个 CDR。本发明的多核苷酸可以被诱变以在 CDR 的氨基酸序列中以及还可能在支架区的氨基酸序列中产生变化。本领域技术人员将会知晓用于该目的合适方法。

[0291] 本发明的多核苷酸还可编码蛋白质缀合物,其能够或不能够如本文中所述与本发明的多肽缀合。

[0292] 多肽产生

[0293] 本文中公开的多肽可以通过本领域中已知的任意方法合成,例如通过重组多肽的产生和回收以及通过多肽的化学合成。因此,本发明还提供了用于产生本发明之多肽的方法。

[0294] 本发明的多肽可以在还原性或非还原性条件下产生。优选地,本发明的多肽在还原性条件下(例如在宿主细胞的胞质中)产生。

[0295] 在重组多肽的情况中,编码所述重组多肽的核酸优选地放置于一个或更多个表达载体中,其随后转染到不另外产生免疫球蛋白的宿主细胞中,所述宿主细胞例如大肠杆菌细胞、酵母细胞、昆虫细胞或哺乳动物细胞,例如猿 COS 细胞,中国仓鼠卵巢(CHO)细胞或骨髓瘤细胞,以在重组宿主细胞中获得蛋白质的合成。关于编码免疫球蛋白之 DNA 在细菌中重组表达的综述文章包括 Skerra 等,(1993) 和 Plückthun(1992)。实现这些目的分子克隆技术是本领域中已知的并且描述于例如 Ausubel 或 Sambrook。很多种克隆和体外扩增方法适合用于重组核酸的构建。产生重组免疫球蛋白的方法也是本领域中已知的。参见 US4,816,567、US5225539、US6054297、US7566771 或 US5585089。

[0296] 分离之后,编码本发明之多肽的核酸优选地插入到表达构建体或可复制载体中用于进一步克隆(DNA的扩增)或用于在无细胞系统或在细胞中表达。优选地,所述核酸与启动子可操作地相连。

[0297] 如本文所用的,术语“启动子”采用其最宽泛的含义并且包括基因组的转录调节序列,包括TATA盒或起始元件,其是准确的转录起始所需要的,带有或不带有额外的改变核酸表达的调节元件(例如上游激活序列、转录因子结合位点、增强子和沉默子),例如响应于发育和/或外部刺激,或者以组织特异方式。在本上下文中,术语“启动子”也用于描述重组体、合成或融合核酸、或衍生物,其赋予、激活或增强与其可操作地相连之核酸的表达。优选的启动子可以包含额外的一个或更多个调节元件的拷贝以进一步增强所述核苷酸的表达和/或改变所述核苷酸的空间表达和/或时间表达。

[0298] 本文中使用的术语“可操作地相连”意指安置与核酸有关的启动子,从而使所述核酸的表达被所述启动子控制。

[0299] 本发明还考虑无细胞表达系统。例如,编码本发明之多肽的核酸可以与合适的启动子(例如T7启动子)可操作地相连,并且所产生的表达构建体暴露于足以转录和翻译的条件。已经描述用于体外表达或无细胞表达的典型表达载体,并且其包括但不限于TNT T7和TNT T3系统(Promega)、pEXP1-DEST和pEXP2-DEST载体(Invitrogen)。

[0300] 用于在细胞中表达的很多载体是可得的。载体组分通常包括但不限于下列的一种或更多种:信号序列、编码本发明之多肽的序列(例如,源自本发明提供的信息)、增强子元件、启动子和转录终止序列。技术人员将知晓用于表达蛋白质的合适的序列。例如,示例性的信号序列包括原核分泌信号(例如,peIB、碱性磷酸酶、青霉素酶、Ipp或热稳定的肠毒素II)、酵母分泌信号(例如,转化酶前导序列、 α 因子前导序列、或酸性磷酸酶前导序列)或哺乳动物分泌信号(例如,单纯疱疹病毒的gD信号)

[0301] 在一个优选的实施方案中,将编码本发明之多肽的多核苷酸插入到特别适合用于在本文中所述RED系统中表达的载体。因此,所述载体可以特别地适合用于在细菌细胞内表达。例如,所述载体可以包含用于将编码本发明之多肽插入到所述载体中的位点,以及编码第二多肽的开放阅读框,所述第二多肽与本发明的多肽缔合以形成可以保留在内部或可以附着至经透化细菌细胞之细胞壁的蛋白质复合物。合适的载体描述于W02011 / 075761中。

[0302] 优选地,所述载体还能够在细菌细胞内独立于宿主基因组地复制。合适的载体包括质粒、病毒和黏粒以及线性DNA元件,例如大肠杆菌的线性噬菌体N15和/或复制独立于细菌细胞基因组的染色体外DNA。

[0303] 技术人员将能够容易地确定适合用于在本发明的方法中表达多肽的细菌菌株。本领域技术人员将理解适合用于本发明之方法的革兰氏阴性细菌,包括例如沙门菌属(*Salmonella*)、大肠杆菌、志贺菌属(*Shigella*)、弯曲杆菌属(*Campylobacter*)、梭杆菌属(*Fusobacterium*)、鲍特杆菌属(*Bordetella*)、巴斯德菌属(*Pasteurella*)、放线杆菌属(*Actinobacillus*)、嗜血杆菌属(*Haemophilus*)和嗜组织菌属(*Histophilus*)。在一个优选的实施方案中,所述革兰氏阴性细菌是大肠杆菌。

[0304] 可以包含在本发明之载体中的示例性启动子包括在原核生物中有活性的那些(例如,phoA启动子, β -内酰胺酶和乳糖启动子系统,碱性磷酸酶、色氨酸(trp)启动

子系统和杂合启动子如 tac 启动子)。这些启动子对于在包括真细菌的原核生物中的表达是有用的,所述真细菌例如革兰氏阴性和革兰氏阳性生物,例如肠杆菌科如埃希氏菌属 (*Escherichia*) (例如大肠杆菌)、肠杆菌属 (*Enterobacter*)、欧文菌属 (*Erwinia*)、克雷伯菌属 (*Klebsiella*)、变形杆菌属 (*Proteus*)、沙门菌属 (*Salmonella*) 例如鼠伤寒沙门菌 (*Salmonella typhimurium*)、沙雷菌属 (*Serratia*) 例如黏质沙雷菌 (*Serratia marcescens*)、和志贺菌属 (*Shigella*)、以及芽孢杆菌属 (*Bacilli*) 如枯草芽孢杆菌 (*B. subtilis*) 和地衣芽孢杆菌 (*B. licheniformis*)、假单胞菌 (*Pseudomonas*) 例如铜绿假单胞菌 (*P. aeruginosa*)、以及链霉菌属 (*Streptomyces*)。优选地,所述宿主是大肠杆菌。一个优选的大肠杆菌克隆宿主是大肠杆菌 294 (ATCC31, 446), 尽管其他菌株例如大肠杆菌 B、大肠杆菌 X1776 (ATCC31, 537) 和大肠杆菌 W3110 (ATCC27, 325)、DH5 α 或 DH10B 也是合适的。

[0305] 在哺乳动物细胞中有活性的示例性启动子包括巨细胞病毒立即早期启动子 (CMV-IE)、人延伸因子 1- α 启动子 (EF1), 小核 RNA 启动子 (U1a 和 U1b)、 α -肌球蛋白重链启动子、猿病毒 40 启动子 (SV40)、劳斯肉瘤病毒启动子 (RSV)、腺病毒主要晚期启动子、p-肌动蛋白启动子;杂合调节元件,包括 CMV 增强子 / β -肌动蛋白启动子或免疫球蛋白启动子或其活性片段。可用的哺乳动物宿主细胞系的实例是用 SV40 转化的猴肾 CV1 系 (COS-7, ATCC CRL1651); 人胚肾系 (用于在悬浮培养物中生长的 293 或亚克隆的 293 细胞; 幼仓鼠肾细胞 (BHK, ATCC CCL10) 或中国仓鼠卵巢细胞 (CHO)。

[0306] 适合用于在酵母细胞 (例如选自以下的酵母细胞:巴斯德毕赤酵母 (*Pichia pastoris*)、酿酒酵母 (*Saccharomyces cerevisiae*) 和粟酒裂殖酵母 (*S. pombe*)) 中表达的典型启动子包括但不限于:ADH1 启动子、GAL1 启动子、GAL4 启动子、CUP1 启动子、PH05 启动子、nmt 启动子、RPR1 启动子或 TEF1 启动子。

[0307] 适合用于在昆虫细胞中表达的典型启动子包括但不限于 OPEI2 启动子、从桑蚕 (*Bombyx mori*) 中分离的昆虫肌动蛋白启动子、果蝇属 (*Drosophila* sp.) dsh 启动子 (Marsh 等, 2000) 和可诱导的金属硫蛋白启动子。用于重组蛋白表达的优选的昆虫细胞包括选自以下的昆虫细胞:BT1-TN-581-4 细胞和草地贪夜蛾 (*Spodoptera frugiperda*) 细胞 (例如, sf19 细胞、sr21 细胞)。适合用于表达核酸片段的昆虫包括但不限于果蝇属。还考虑使用草地贪夜蛾。

[0308] 用于将分离的核酸分子或包含所述核酸分子的核酸构建体引入到细胞中用于表达的方式是本领域技术人员已知的。用于给定细胞的技术取决于已知的成功技术。用于将重组 DNA 引入到细胞中的方式包括显微注射、由 DEAE-葡聚糖介导的转染、通过脂质体介导的转染 (例如通过使用 lipofectamine (Gibco, MD, USA) 和 / 或 cellfectin (Gibco, MD, USA))、PEG 介导的 DNA 摄取、电穿孔和微粒轰击 (microparticle bombardment) (例如通过使用包被 DNA 的钨或金颗粒 (Agracetus Inc., WI, USA) 等。

[0309] 用于产生本发明之蛋白质的宿主细胞可以在多种培养基中培养, 取决于所用的细胞类型。可市购的培养基例如 Ham's F10 (Sigma)、最低极限培养基 ((MEM), Sigma)、RPMI-1640 (Sigma) 和 Dulbecco 改良的 Eagle 培养基 ((DMEM), Sigma) 适合用于培养哺乳动物细胞。用于培养本文中所讨论的其他细胞类型的培养基是本领域中已知的。

[0310] 支架序列:

[0311] 目前为止,胞内抗体 (intrabody) (抗体分子,其序列已经被改造或演化为更高稳定性,从而使得其可以在胞质中高效地折叠)之结构区的序列(即非 CDR 序列)与同源种系基因组序列显著地不同。如本文中所公开的,该发明人已经筛选并确定与同源种系基因组序列一致或密切相关的人抗体可变区序列,并且其在非氧化性环境中表达时允许正确折叠和提高的稳定性。用于在本发明中使用的优选序列如下文所述。对于任何本文中所描述的可变区序列,应理解的是,本领域技术人员将能够鉴定 CDR(例如其中很多已经在 NCBI 数据库中鉴定)和保留的支架区。在本文中描述的每个可变区中 CDR 的具体实例在图 7 中显示。

[0312] IGHV3-23

[0313] 在一个优选的实施方案中,本发明的多肽包含免疫球蛋白可变结构域之 V_H3 家族的重链可变区 (V_H)。优选地,所述 V_H 为 IGHV3-23 (SEQ ID NO :3)。

[0314] IGHV3-23,也称作 DP-47,属于人 Ig 可变结构域的 V_H3 家族。 V_H3 家族占 V_H 基因的功能性成员的 43% (22 / 51) 并且 IGHV3-23 在 VH 谱中已被引用最高度表达的基因 (Stewart 等,1993)。还发现其在 B 细胞中以高频率产生 Ig 重排 (Brezinschek 等,1997)。由于其在天然 Ig 谱中的高频率,其也频繁地从人 V 区的噬菌体展示文库中分离 (Griffiths 等,1994)。其也被用作合成文库中的支架伴侣 (Jirholt 等,1998 ;Pini 等,1998 ;Soderlind 等,2000 ;Ge 等,2010)。IGHV3-23 在本发明人的研究中还被选作重链可变区伴侣以鉴别稳定的、可溶性抗体可变区支架。

[0315] 优选地,本发明的多肽包含抗体重链可变区 (V_H),其包含与 SEQ ID NO :3 之 IGHV3-23 的支架区具有至少 90% 同一性的支架区。例如,所述支架区可以与如 SEQ ID NO :3 所示 IGHV3-23 的支架区具有至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或 100% 的同一性。优选地,本发明的多肽包含抗体重链可变区 (V_H),其包含与如 SEQ ID NO :3 所示 IGHV3-23 的支架区具有至少 96% 同一性的支架区。所述支架区包含所有排除 CDR 残基的可变区残基。因此,本发明的多肽可以包含支架区,其包含 SEQ ID NO :3 的第 1-25、33-51、60-98 位氨基酸 (或者支架区,其氨基酸序列与这些氨基酸的序列具有至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或 100% 的同一性)。或者,本发明的多肽可以包含支架区,其包含 SEQ ID NO :3 的第 1-25 和 33-98 位氨基酸 (或者支架区,其氨基酸序列与这些氨基酸的序列具有至少至少 90%、95%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或 100% 的同一性)。在另一个替代方案中,本发明的多肽可以包含支架区,其包含 SEQ ID NO :3 的第 1-51 和 60-98 位氨基酸 (或者支架区,其氨基酸序列与这些氨基酸的序列具有至少至少 90%、95%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或 100% 的同一性)。

[0316] 本发明的多肽可以包含任意 CDR 序列。因此,本发明的多肽可以包含 IGHV3-23 的 CDR 序列 (即,SEQ ID NO :3 第 26-32、52-59 位氨基酸)。或者,本发明的多肽可以包含任意其他 CDR 序列。因此, V_H3 可变区的支架区可以作为可插入任何给定 CDR 序列的模板。可以随机产生 CDR 序列。作为替代或补充,所述 CDR 序列可以是半随机产生的 (随机的给 CDR 中的每个特定氨基酸位置分配一个选自所有可能氨基酸之子集的氨基酸残基,已知所述子集在 CDR 中的给定氨基酸位置是必须的或特别偏好的)。

[0317] 或者,所述 CDR 序列可以来自另一个抗体。因此,来自例如人抗体的 CDR 可以移植

到 V_H3 可变结构域支架。应理解,本领域技术人员可使用多种方法以确保如本文中所限定的支架区包含取自人抗体的一个或更多个 CDR 序列。优选地,这样的方法包括将一个或更多个 CDR 编码序列克隆到编码本发明之多肽的多核苷酸中,如在下文中更详细的描述。所述 CDR 编码序列可以通过靶向或随机诱变而另外地变化,以提供多种包含多种不同不同 CDR 序列的多肽。这样的方法可以应用在 CDR1、CDR2 和 CDR3 中的任一个或其组合。

[0318] 可以以任意组合将任意一个或更多个 CDR (CDR1、CDR2 和 CDR3) 的序列引入本文中所描述的可变结构域支架。优选地,至少将 CDR3 的序列引入本文中所描述的 V_H3 可变结构域支架。

[0319] 此外,引入本文所述的 V_H3 可变结构域支架的 CDR 序列的长度可以变化。例如,可将 15、14、13、12、11、10、9、8、7、6、5、4、3、2 或 1 个氨基酸的 CDR3 序列插入所述支架。本发明人发现长度少于 12 个(例如 11、10、9、8、7、6、5、4、3、2 或 1 个,并且最优选 7 个)氨基酸的缩短的 CDR3 序列表现出增强的稳定性。

[0320] IGHV3-23 的 IGK 和 IGL 轻链伴侣

[0321] 本发明的发明人,仅通过应用 scFv 融合在其 RED 平台中可溶的标准(与先前进行的在体内将抗体结合抗原靶标的功能性筛选相反并且其因此筛选从其各自的种系序列充分地突变的抗体)能够筛选在 V 区内还没有突变的天然轻链。因此,他们能够鉴别在 IGHV3-23scFv 融合之后赋予可溶性的种系序列。这具有显著的益处,确保 V_L 和 V_H 结构域的构建的人工支架文库在序列上与丰富的人抗体蛋白质是一致的,从而最小化免疫识别并拒绝延长任何衍生物的暴露。

[0322] 因此,本发明的多肽优选地包含与人种系 IGHV3-23 序列组合的免疫球蛋白可变结构域之 $V_L \lambda$ 1、3 或 6 家族的抗体轻链可变区 (V_L)。优选的 $V_L \lambda$ 1、3 或 6 家族成员包括 IGLV1-40(如 SEQ ID NO:18 所示)、IGLV1-44(如 SEQ ID NO:21 所示)、IGLV1-47(如 SEQ ID NO:24 所示)、IGLV1-51(如 SEQ ID NO:15 所示)、IGLV3-1(如 SEQ ID NO:6 所示)、IGLV3-19(如 SEQ ID NO:27 所示)、IGLV3-21(如 SEQ ID NO:9 所示)和 IGLV6-57(如 SEQ ID NO:12 所示)。因此,本发明的多肽优选地包含抗体轻链可变区 (V_L),其包含与以下任一支架区具有至少 90% 的同一性的支架区: IGLV1-40(如 SEQ ID NO:18 所示)、IGLV1-44(如 SEQ ID NO:21 所示)、IGLV1-47(如 SEQ ID NO:24 所示)、IGLV1-51(如 SEQ ID NO:15 所示)、IGLV3-1(如 SEQ ID NO:6 所示)、IGLV3-19(如 SEQ ID NO:27 所示)、IGLV3-21(如 SEQ ID NO:9 所示)和 IGLV6-57(如 SEQ ID NO:12 所示)。例如,所述支架区与以下任一支架区有至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或 100% 同一性: IGLV1-40(如 SEQ ID NO:18 所示)、IGLV1-44(如 SEQ ID NO:21 所示)、IGLV1-47(如 SEQ ID NO:24 所示)、IGLV1-51(如 SEQ ID NO:15 所示)、IGLV3-1(如 SEQ ID NO:6 所示)、IGLV3-19(如 SEQ ID NO:27 所示)、IGLV3-21(如 SEQ ID NO:9 所示)和 IGLV6-57(如 SEQ ID NO:12 所示)。

[0323] 最优选地,IGHV3-23 的 V_L 伴侣是 IGLV3-1(SEQ ID NO:6)。因此,本发明的多肽优选地包含抗体轻链可变区 (V_L),其包含与 IGLV3-1(SEQ ID NO:6) 的支架区至少 90% 同一性的支架区。例如,所述支架区与 IGLV3-1(SEQ ID NO:6) 的支架区有至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或 100% 同一性。IGLV3-1 的支架区可以包含 SEQ ID NO:6 的氨基酸 1-23、32-48、56-89。从而,本发明的多肽可包含抗体轻链可变区 (V_L),其包含 SEQ

ID NO:6 的氨基酸 1-23、32-48、56-89 (或者其氨基酸序列与这些氨基酸序列有至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或 100% 同一性)。

[0324] 编码具有可变 CDR3 区之 IGLV3-1 : IGHV3-23 支架的多核苷酸序列的优选实例和对应的经翻译氨基酸序列在图 8 中举例说明。

[0325] 优选地,本发明的多肽包含 V_L 可变结构域的支架区 (例如本发明的多肽可以包含 IGHV3-1 的支架区,例如 SEQ ID NO:6 的氨基酸 1-23、32-48、56-89)。再者,本发明的多肽可以包含任意 CDR 序列。因此本发明的多肽可以包含以下任意的氨基酸序列: IGLV1-40、IGLV1-44、IGLV1-47、IGLV1-51、IGLV3-1、IGLV3-19、IGLV3-21 和 / 或 IGLV6-57。或者,本发明的多肽可以包含任意其他 CDR 序列。因此, V_L 可变结构域的支架区可以作为可插入任意给定 CDR 的模板,如以上描述的关于 V_H3 可变结构域支架。因此,所述 CDR 序列可以是随机产生的。作为替代或补充,所述 CDR 序列可以是半随机产生的 (通过随机的将选自所有可能氨基酸之子集的氨基酸残基分配到 CDR 中的每个特定氨基酸位置,已知所述子集在 CDR 中的给定氨基酸位置是特别偏好的)。

[0326] 或者,所述 CDR 序列可以来自其他抗体。因此,可以将来自例如人抗体的 CDR 移植到 V_L 可变结构域支架上。本领域的技术人员将意识到多种不同的方法是可用的以确保如本文所限定的支架区包含一个或更多个取自人抗体的 CDR 序列。此外,人抗体的 CDR 序列在插入本文中所描述的 V_L 可变结构域支架之前可以被随机的诱变。

[0327] 可以将 CDR1、CDR2 和 CDR3 之序列的一个或更多个以任意组合插入本文中所描述的 V_L 可变结构域支架。优选地,至少将 CDR3 的序列插入本文中所描述的 V_L 可变结构域支架。

[0328] 此外,插入到本文所述的 V_L 可变结构域支架的 CDR 序列中的长度可以变化。例如,可将 15、14、13、12、11、10、9、8、7、6、5、4、3、2 或 1 个氨基酸的 CDR3 序列插入所述支架。本发明人发现长度少于 12 个 (例如 11、10、9、8、7、6、5、4、3、2 或 1 个,并且最优选 7 个) 氨基酸的缩短的 CDR3 序列表现出增强的稳定性。

[0329] 合成的多肽文库

[0330] 可以在多种蛋白质展示平台中克隆并表达多肽序列的文库以选择针对期望靶标的亲和蛋白。因此,本发明提供了包含多种本发明之多肽的文库。在一个优选的实施方案中,所述文库可以通过鉴别“亲本”多肽或多核苷酸序列并改变所述序列来产生多种变体序列以形成文库。可以通过合适的方法例如通过位点定向诱变或随机诱变进行改变。文库构建的合适的方法是本领域中已知的。

[0331] 如以上表明的,可以直接地从生物来源克隆可变结构域,使得结构序列和 CDR 二者都如通过 V(D)J 重组形成而存在。或者,所述抗体可以部分地区或全部地用 CDR 和从头合成组装的结构区合成。例如,单个人工支架代表一对通常表达的或特别稳定的,可以用于在宿主生物体中最优表达的抗体基因。甚至可以在单个反应中使用重叠寡核苷酸组装完整的支架和 CDR,例如由 Ge 等 (2010) 所描述。

[0332] 用于在抗原结合 CDR 中构建多样性的方法已经被现有技术完全地描述了。它们包括从 mRNA、从天然的或预免疫的免疫细胞获得 CDR;通过分析经核对的抗体序列来设计并合成 CDR;用基于经核对的抗体序列的加权氨基酸分布来设计与合成 CDR;采用随机的、无偏倚的 CDR 区。

[0333] 每个 Ig 结构域 (V_L 和 V_H) 有三个 CDR 区 (CDR1、CDR2 和 CDR3), 其长度可变并且与抗原的界面接触有不同的频率。对 V_L 和 V_H 二者, 在体内最多的变体 CDR 是 CDR3, 其环由 V-J 结构域 (V_L) 或 V-D-J 结构域 (V_H) 的外显子连接之间的重组形成。这代表了天然的免疫系统。然而, 一旦 B 细胞被刺激用于扩增, 随后也经常发生体细胞超突变以使 CDR1 和 2 多样化。

[0334] 然而, 用于将可变结构域的经典 scFv 文库构建到单个或数个 V_L 和 V_H 支架中是常见的, 由于 CDR 多样性受限于 CDR3, 用氨基酸组合物和环长度变化计算靶结合。

[0335] 在由本发明描述的稳定多肽的实例中, 这允许 scFv 的整体 (而不是天然变化的 CDR3 环区) 变得与同源抗体基因的种系序列相同或相近。这允许筛选酷似人天然抗体谱的亲本蛋白, 从而将可以启动患者免疫识别的经改造支架的序列差异最小化。由此, 在本发明中, 多肽文库可以包含在 CDR3 序列中彼此不同的多肽。

[0336] 可以利用单一支架构建人工抗体文库或者其可以由多种支架构成。因此, 本发明的文库可以包含一个或更多个具有本文中公开的重链和轻链可变区支架的特定组合的多肽以及一个或更多个具有本文中公开的重链和轻链可变区支架的另一不同组合的多肽。

[0337] 例如, 所述文库可以包含:

[0338] - 一个或更多个包含抗体重链可变区 (V_H) 和抗体轻链可变区 (V_L) 的多肽, 所述抗体重链可变区 (V_H) 包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区, 所述抗体轻链可变区 (V_L) 包含与 IGLV3-1 (如 SEQ ID NO:6 所示) 的支架区具有至少 90% 同一性的支架区, 其中所述 V_H 和所述 V_L 能够形成抗原结合位点, 以及

[0339] - 一个或更多个包含抗体重链可变区 (V_H) 和抗体轻链可变区 (V_L) 的多肽, 所述抗体重链可变区 (V_H) 包含与如 SEQ ID NO:3 所示 IGHV3-23 的支架区具有至少 90% 同一性的支架区, 所述抗体轻链可变区 (V_L) 包含与以下任意一个或更多个的支架区具有至少 90% 同一性的支架区: IGLV1-40 (如 SEQ ID NO:18 所示)、IGLV1-44 (如 SEQ ID NO:21 所示)、IGLV1-47 (人 SEQ ID NO:24 所示)、IGLV1-51 (如 SEQ ID NO:15 所示)、IGLV3-19 (如 SEQ ID NO:27 所示)、IGLV3-21 (如 SEQ ID NO:9 所示)、IGLV6-57 (如 SEQ ID NO:12 所示), 其中所述 V_H 和所述 V_L 能够形成抗原结合位点。

[0340] 由此, 本发明的文库可以包含一个或更多个多肽, 其具有本文中公开的重链和轻链可变区骨架的任意重组。

[0341] 多种支架可以代表两大类人 Ig 基因, 即与 κ 和 λ 轻链类配对的重链, 或者它们可以是单一轻链支架与多种重链的混合物, 或者轻链支架与单一重链的混合物。多个支架还可以从最稳的代表不同亚类的单一成员产生, 或者可以是仅仅最稳定支架 (属于任意类) 的组合。

[0342] 在本发明的例子中, 多肽文库优选地由支架区构成, 所述支架区与人 IGHV3-23 基因相同或近似相同 (例如至少 90、95、96、97、98 或 99% 相同), 与人基因的支架区 (IGLV1-40、IGLV1-44、IGLV1-47、IGLV1-51、IGLV3-1、IGLV3-19、IGLV3-21 或 IGLV6-53) 相同或近乎相同 (例如至少 90、95、96、97、98 或 99% 相同) 的序列可操作地相连接。所述文库可以以经验证的模式 (例如 scFv) 构成重链和轻链基因的单一支架配对或者可以成为 IGHV3-23 基因与前述轻链基因之一或更多种的支架对。本发明人证实从这些支架构建的文库在大肠杆菌胞质中具有优越的和期望的稳定性和可溶性。实际上, 它是优越的胞内抗

体文库, 其与其同源种系基因同源的理想序列的不同仅存在于 CDR 环区。

[0343] 本领域技术人员将会意识到, 用于筛选具有作为 VH 基因 (IGHV3-23) 之伴侣的胞质可溶性多肽的方法, 也可以用于筛选其他可变区 (无论 VL 还是 VH) 的可溶性伴侣。此外, 筛选胞质可溶性多肽的方法可以使用单一支架的变体迭代 (iterate) 以发现提高其稳定性和可溶性的突变。例如, 已经被鉴定的任意支架对 (IGHV3-23 与 IGLV1-40、IGLV1-44、IGLV1-47、IGLV1-51、IGLV3-1、IGLV3-19、IGLV3-21 或 IGLV6-57) 可以用作在单一支架上变体之其他文库的模板, 意图在亲本支架具有差的可溶性之温度下进行胞质筛选。

[0344] 本领域有经验的从业人员还将意识到, 多肽文库可以由前述的以低于 100% 丰度存在的多肽构成。由本发明描述的 50% 多肽, 或者由本发明描述的 25% 多肽, 或者由本发明描述的 10% 多肽构成的文库, 将仍然起作用以产生具有期望稳定性特性的亲和蛋白。因此, 本发明的多肽文库可以包含除了本发明的那些之外的多肽。

[0345] 此外, 尽管本发明人已经出人意料的发现用于描述的 VH 和 VL 结构域的在序列上与人种系基因相同或近乎相同的 scFv 支架在大肠杆菌胞质中具有优越的和期望的稳定性和可溶性, 本领域有经验的从业人员将意识到, 可以获得的这些序列比已报道的更有多态性, 但仍作为有期望的可溶性特性的亲和蛋白。因此, 本发明提供具有序列与本公开的支架区序列至多 10% 或 5% 相异, 并且仍然起作用以产生具有期望可溶性和 / 或稳定性特性的亲和蛋白。因此, 本发明的多肽可以包含与本文中公开的任意支架区序列具有至少 90%、95%、96%、97%、98% 或 99% 同一性的支架区序列。

[0346] 本发明提供了多肽文库和多核苷酸文库 (例如, DNA 文库) 二者。DNA 文库是含有编码所述多肽之 DNA 插入片段 (DNA 片段) 的重组载体的集合。DNA 插入片段的来源可以是基因组、cDNA、合成或半合成。

[0347] 编码本发明之多肽的 DNA 文库的克隆和构建可以使用本领域已知的方法进行。例如, Lutz 和 Patrick (2004) 已经回顾了产生文库变异性的方法和用于在蛋白质工程中基因重组的策略。为了筛选展示的多肽变体, 可以采用用于表面展示文库并且适用于本发明方法的策略 (Becker 等, 2004; Kenrick 等, 2007; Miller 等, 2006; Daugherty 等, 2000)。

[0348] 可以将核酸文库引入到大量的细菌细胞中, 导致在每个细菌细胞中表达文库的成员。除了被表达之外, 所述多肽被保留在经透化的细菌细胞中或者附着至细胞壁, 以评价其功能或特征。多肽的核酸文库, 例如, 可以通过多种方法包括通过突变 (例如点突变、缺失或插入或者通过重组事件) 来引入。用于变体之文库的产生方法是本领域中已知的, 并且包括易错 PCR、DNA 修复受损的细菌中 DNA 的合成、以及 DNA 的化学修饰。用于通过重组产生文库的方法是本领域中已知的并且包括基因洗牌 (gene shuffling)、在高度重组发生的细菌中组装 DNA、合成核酸文库组装等, 或者其任意组合。以这种方式, 编码多肽之多核苷酸的文库可以引入到大量的细菌细胞中, 导致在每个细菌细胞中文库的一个或更多个成员的表达。

[0349] 在一些实施方案中, 文库包含多肽的两个或更多个变体, 其中每个变体包含具有在氨基酸序列中 (例如在 CDR 序列中) 微小改变的独特多肽。文库可以具有至少 2、至少 5、至少 10、至少 50、至少 100、至少 1000、至少 10,000、至少 100,000、至少 1,000,000、至少 10^7 或更多个成员。

[0350] 筛选方法

[0351] 本文中公开的多肽（或抗体）文库可以用于筛选与靶分子结合的多肽。应理解，可以在每个表达不同多肽或多肽变体的细胞之文库的情况中，或者在表达单一多肽的细胞之单一类型的情况中筛选或选择多肽。术语“靶分子”是指与多肽结合或由多肽修饰的分子并且可以是例如抗原、酶、抗体、受体等。因此“靶分子”可以用于指底物例如酶底物或被评价用于结合的分子（例如，配体、表位、抗原、多聚化伴侣例如均二聚体伴侣或杂二聚体伴侣等，或其任意组合）。

[0352] 因此，本发明提供了筛选与靶多肽结合之多肽的方法，所述方法包括将本发明的多肽与靶多肽相接处，并且确定本发明的多肽是否与靶多肽结合。优选地，在这样的方法中使用多种本发明的多肽。

[0353] 许多合适的筛选方法是本领域中已知的，其可以根据本发明使用。

[0354] 例如，所述方法可以包括蛋白质展示方法。最早的蛋白质展示方法是噬菌体展示 (Smith, 1985)，在所述方法中，目的蛋白质与噬菌体的一种外壳蛋白 (outer-coat protein) 融合，从而其可以与野生型的蛋白质拷贝一起存在。例如，可以采用基于使用与 pIII 蛋白融合的 M13 丝状噬菌体的展示平台。

[0355] 其他合适的展示方法包括“体外”展示方法，其中使用细胞翻译提取来表达蛋白质，并通过物理连接实现多肽与编码核酸之间的偶联（例如核糖体展示、mRNA 展示）或通过附着到共用骨架上或膜内的封装，例如，在体外区室化 (in vitro compartmentalization, IVC) 中，其中 mRNA 在微团悬液中翻译，所述微团悬液还包含用于 mRNA 和蛋白质二者的微珠（磁性或琼脂糖）捕获系统。

[0356] 另一个合适的多肽展示的方法是微生物表面展示，其包括将表达的多肽靶向定位到微生物细胞外部，所述微生物为革兰氏阴性或革兰氏阳性真细菌或酵母。蛋白质与将其连接到细胞表面的锚定结构域融合。锚定结构域可具有指导脂化或与细胞壁共价连接的模体，或其可与暴露的环区内的内在膜蛋白融合。

[0357] 本申请证明本发明的多肽和多核苷酸用于包含无细胞表达系统的筛选方法时是特别有效的。在这样的表达系统中使用本发明的多肽和多核苷酸极大加快了多肽的筛选过程，表现出靶多肽的高表达、高溶解度和高亲和力。本发明的多肽（特别在还原性条件下）展示出高溶解度、稳定性和表达所产生的益处。

[0358] 因此，本发明的多肽和多核苷酸特别适合用于这样的筛选方法，所述筛选方法包括本文中所描述的任意无细胞或体外表达系统或本领域中已知的其他方法。例如，本发明的多肽和多核苷酸序列特别适合用于包括核糖体展示、mRNA 展示、顺式展示 (cis-display)（其中表达的多肽保持与其编码多核苷酸序列相缀合）或本领域中已知的其他此类方法的筛选方法。

[0359] 此外，本申请表明本发明的多肽和多核苷酸在用于基于蛋白质展示方法的筛选方法时是特别有效的。例如，本发明的多肽和多核苷酸在用于包括噬菌体展示（例如，裂解性 λ 噬菌体、M13 丝状噬菌体、裂解缺陷型噬菌体和其他本领域中已知的方法）的筛选方法时是特别有效的。在一个实例中，本发明的多肽和多核苷酸在用于包括 λ 噬菌体的筛选方法时出人意料地有效。

[0360] 此外，本发明的多肽和多核苷酸可以用于基于裂解缺陷型噬菌体（例如，如描述于国际专利申请号 PCT / AU2012 / 000761，其内容通过引用全文并入本文）与本文中和

WO2011 / 075761 中所描述的 RED 系统相组合的筛选方法。

[0361] 试剂盒

[0362] 用于实施本发明方法的必要组分可以以试剂盒的形式方便地提供。如将被本领域技术人员理解的,在试剂盒中的多种组分可以以单个容器或等等分量供应,或可以将溶液组分以不同的组合并且在不同浓度组合来实现本发明方法的最佳性能。确定试剂盒的哪些组分可以组合使得所述组分在使用前保持处于稳定的形式在本领域技术人员的知识范围内。

[0363] 本发明的试剂盒通常将以最小量含有载体,其包含用于将编码本发明之多肽的多核苷酸插入载体的位点,和编码第二多肽的开放阅读框,所述第二多肽与第一多肽缔合以形成可以保留在经透化细菌细胞内的或者可以附着至其细胞壁的蛋白质复合物。优选地,所述试剂盒还含有用于透化细菌细胞的试剂。在一个实施方案中,所述试剂盒还包含细菌细胞,优选革兰氏阴性细菌。其他另外的组分,或者由终端使用者提供的其他组分(如果需要)也可以包括在试剂盒中。

[0364] 用途

[0365] 本发明的多肽可用于多种应用(包括研究、诊断和治疗应用)。依赖于所述多肽所结合的抗原,所述多肽可用于将化合物递送至细胞(例如杀死细胞或阻止生长)和/或用于成像和/或用于体外测定。在一个实例中,所述多肽可用于成像和将细胞毒剂递送至细胞,即所述多肽与可检测的标签和细胞毒剂相缀合或者组合物包含蛋白质混合物的组合物,其中一些蛋白质与细胞毒剂缀合并且一些蛋白质与可检测标签缀合。

[0366] 本文中所描述的多肽还可以充当抑制剂来抑制(其可以降低或阻止)(a)(例如配体、抑制剂)与受体结合,(b)受体信号功能,和/或(c)刺激功能。充当受体功能之抑制剂的多肽可以直接地或间接地(例如通过引起构象改变)阻断配体结合。鉴于本文中所描述的优选多肽的稳定性和大小,本文中所描述的多肽可以特别地适合于涉及在宿主细胞内发生结合相互作用的应用。

[0367] 药物组合物和治疗方法

[0368] 本发明的多肽(同义词:活性成分)可用于肠胃外、表面、经口或局部施用、气雾剂施用、或用于预防性或治疗性治疗的透皮施用。所述药物组合物可以以多种依赖于施用方法的单位剂型来施用。例如,适合用于经口施用的单位剂型包括散剂、片剂、丸剂、胶囊剂和锭剂或通过肠胃外施用。应当意识到,本发明的药物组合物经口施用时应避免被消化。这通常通过将蛋白质与组合物复合以使其抗酸性和抗酶水解,或通过将化合物包装在合适的抗性载体(例如脂质体)中。保护蛋白质免受被消化的方式是本领域中已知的。

[0369] 通常地,多肽的治疗有效量将被配制成组合物用于向对象施用。短语“治疗有效量”是指量足以促进、诱导和/或增强在对象中的治疗或其他治疗效果。如将是显而易见的,本发明的蛋白质在这些制剂中的浓度可以很大的不同,并且将主要基于根据所选特定施用方式和患者需求的流体体积、黏度、体重等进行选择。取决于疾病的类型和严重程度,治疗有效量可以是约 $1\mu\text{g}/\text{kg}$ 至 $15\text{mg}/\text{kg}$ (例如 $0.1\text{--}20\text{mg}/\text{kg}$) 的分子,或者例如通过一次或更多次分开的施用,或者通过持续的输注。典型的日剂量可以为约 $1\mu\text{g}/\text{kg}$ 至 $100\text{mg}/\text{kg}$ 或更多。待施用于患者的蛋白质的示例性剂量范围为患者体重的约 0.1 至约 $10\text{mg}/\text{kg}$ 。对于重复施用持续数天或更长的,取决于条件,持续治疗直到发生疾病症状

的期望抑制。示例性的施用方案包括施用约 4mg / kg 的初始加载剂量,之后每周维持约 2mg / kg 蛋白质的剂量。其他剂量方案可以是有用的。该治疗的进展易于通过常规技术和测定进行监测。

[0370] 或者,本发明的多肽以浓缩剂量配制,其在施用于患者前稀释到治疗有效剂量。

[0371] 本发明的药物组合物对于肠胃外施用是特别有的,例如配制成用于通过静脉内、肌肉内、皮下、经皮或其他这样的途径(包括蠕动施用和直接滴注到肿瘤或疾病部位中(腔内施用))注射。用于施用的组合物通常将包含本发明之蛋白质溶于可药用载剂(优选地水载剂)的溶液。可使用多种水载剂,例如缓冲盐水等。其他示例性载剂包括水、盐水、林格液(Ringer's solution)、葡萄糖溶液和 5% 人血清白蛋白。也可以使用非水载剂例如混合油和油酸乙酯。还可以使用脂质体作为载剂。所述载剂可以含有少量的添加剂(例如缓冲剂或防腐剂)以增强等张性和化学稳定性。所述组合物可以含有所需的可药用的辅助物质以接近生理条件,例如 pH 调节和缓冲剂、毒性调节剂等,例如,乙酸钠、氯化钠、氯化钾、氯化钙、乳酸钠等。

[0372] 用于制备药物组合物的技术是本领域中公知的,如由 Remington's Pharmaceutical Sciences, 16th 版 Mack Publishing Company, 1980 示例。

[0373] WO2002 / 080967 描述用于施用包含蛋白质的雾化组合物用于治疗例如哮喘的组合物和方法,其也适用于施用本发明的蛋白质。

[0374] 本发明化合物的合适的剂量将根据特定的蛋白质、待诊断 / 治疗 / 预防的病症和 / 或被治疗的对象而变化。确定合适的剂量在技术人员的能力之内,例如通过开始用亚最佳剂量和递增地改变剂量以确定最佳的或有用的剂量。或者,使用来自细胞培养测定或动物研究的数据,以确定用于治疗 / 预防的合适剂量,其中合适的剂量在循环浓度范围内,所述循环浓度包括具有很少或没有毒性之活性化合物的 ED50。所述剂量可以在此范围内变化,根据所使用的剂型和施用的途径。可以从细胞培养测定初步地估计治疗 / 预防有效剂量。可以在动物模型中配制剂量以实现循环血浆溶度范围包括如在细胞培养中确定的 IC50(即化合物的浓度实现症状的半最大化抑制)。这样的信息可以用于更精确地确定在人的有用剂量。可以通过例如高效液相色谱法测量血浆中的水平。

[0375] 本发明的多肽可以在药物组合制剂或作为联合治疗的剂量方案中与第二化合物组合。药物组合制剂或剂量方案的第二化合物优选地具有与蛋白质之组合的互补活性,使得它们不会不利地互相影响。

[0376] 所述第二化合物可以是化疗剂、细胞毒剂、细胞因子、生长抑制剂、抗激素剂和 / 或心脏保护剂。这样的分子以有效用于预期目的的量合适地存在于组合中。含有本发明之蛋白质的药物组合物还可以具有治疗有效量的化学治疗剂,例如微管蛋白形成抑制剂、拓扑异构酶抑制剂或 DNA 结合剂。还可以使用药物“缓慢释放”胶囊或组合物。缓慢释放制剂一般设计为长时间地给出恒定的药物水平并且可以用于递送本发明的化合物。

[0377] 本发明还提供了在对象中治疗或预防病症的方法,所述方法包括向有此需要的对象施用治疗有效量的本发明之蛋白质。

[0378] 本文中使用的术语“预防”在预防病症的情况中包括施用本文中所描述之量的蛋白质足以阻止或阻碍特定疾病或病症的至少一种症状的发生。

[0379] 本文中使用的“治疗”包括施用治疗有效量的本文汇中描述之抑制剂和 / 或药剂

足以降低或消除特定疾病或病症的至少一种症状。

[0380] 本文中使用的术语“对象”应当被理解为意指任何动物（包括人），优选哺乳动物。示例性对象包括但不限于人、灵长类、家畜（例如绵羊、牛、马、驴、猪）、伴侣动物（例如狗、猫）、实验室试验动物（例如小鼠、兔、大鼠、豚鼠、仓鼠）、圈养野生动物（例如狐狸、鹿）。优选地，所述哺乳动物是人或灵长类。更优选地，所述哺乳动物是人。

[0381] 本文中使用的“病症”是正常功能的干扰或中断，并且不限于任何特定的条件，并且将包括疾病或紊乱。在一个实例中，所述病症是癌症或免疫病理疾病。

[0382] 示例性的癌症包括但不限于癌、淋巴瘤、胚细胞瘤、肉瘤和白血病或淋巴恶性肿瘤。此类癌症的更具体的实例包括鳞状细胞癌（例如上皮鳞状细胞癌）、肺癌（包括小细胞肺癌、非小细胞肺癌、肺的腺癌和肺的鳞状癌）、腹膜的癌症、肝细胞癌、胃癌（包括胃肠癌）、胰腺癌、成胶质细胞瘤、子宫颈癌、卵巢癌、肝癌、膀胱癌、肝细胞瘤、乳腺癌、结肠癌、直肠癌、结直肠癌、子宫内膜或子宫癌、唾液腺癌、肾癌、前列腺癌、外阴癌、甲状腺癌、肝癌、肛门癌、阴茎癌、以及头颈癌。

[0383] 免疫病理学是具有免疫病因之疾病的研究并且免疫学疾病是任何由免疫球蛋白与抗原的反应引起的病症。因此，“免疫病理疾病”可以被定义为从对象的免疫系统与抗原的反应产生的疾病。免疫病理疾病包括自身免疫疾病和超敏应答（例如 I 型：过敏反应（anaphylaxis）、麻疹（hives）、食物过敏、哮喘，II 型：自身免疫性溶血性贫血、输血反应，III 型：血清病、坏死性脉管炎（necrotizing vasculitis）、肾小球肾炎、类风湿关节炎、狼疮，IV 型：接触性皮炎，移植排斥）。自身免疫性疾病包括风湿性疾病（如，例如，类风湿关节炎、干燥综合征（Sjogren's syndrome）、硬皮病、狼疮（例如 SLE 和狼疮肾炎）、多肌炎 / 皮肌炎、冷球蛋白血症、抗磷脂抗体综合征和银屑病关节炎）、骨关节炎、自身免疫性胃肠和肝疾病（例如，炎症肠病（例如，溃疡性结肠炎和克罗恩病）、自身免疫性胃炎和恶性贫血、自身免疫性肝炎、原发性胆汁性肝硬化、原发性硬化性胆管炎、和乳糜泻（celiac disease）、脉管炎（例如，ANCA 相关脉管炎（包括 Churg-Strauss 脉管炎、Wegener 肉芽肿病和多动脉炎），自身免疫性神经病（例如，多发性硬化症、眼阵挛 - 肌阵挛综合征（opsoclonus myoclonus syndrome）、重症肌无力、视神经脊髓炎、和自身免疫性多发性神经病（autoimmune polyneuropathies）、肾疾病（例如，肾小球肾炎，Goodpasture 综合征和 Berger 病），自身免疫性皮肤病学疾病（例如银屑病、荨麻疹、麻疹、寻常性天疱疮（pemphigus vulgaris）、大疱性类天疱疮和皮肤红斑狼疮）、血液学疾病（例如，血小板减少性紫癜、血栓性血小板减少性紫癜、输血后紫癜、和自身免疫性溶血性贫血）、动脉粥样硬化、葡萄膜炎（uveitis）、自身免疫性听觉疾病（例如内耳疾病和听力损失），白塞病（Behcet's disease），雷诺综合征（Raynaud's syndrome），器官移植和自身免疫性内分泌疾病（例如，糖尿病相关自身免疫病，例如胰岛素依赖性糖尿病（insulin-dependent diabetes mellitus, IDDM），爱迪生病（Addison's disease）和自身免疫性甲状腺病（例如，格雷夫斯病（Graves' disease）和甲状腺炎）。更优选的此类疾病包括例如类风湿性关节炎、溃疡性结肠炎、ANCA 相关血管炎、狼疮、多发性硬化、干燥综合征、格雷夫斯病、IDDM、恶性贫血、甲状腺炎和肾小球肾炎。

[0384] 在另一个实施方案中，所述疾病是炎性疾病。炎症是身体组织对发炎或损伤的保护性应答并且可以是急性的或慢性的。因此，炎性疾病包括涉及以下的疾病：中性粒细胞、

单核细胞、肥大细胞、嗜碱性粒细胞、嗜酸性粒细胞、巨噬细胞,其中发生细胞因子释放、组胺释放、氧化爆发 (oxidative burst)、吞噬作用、其他颗粒酶释放和趋化作用。超敏应答 (根据如上所定义的免疫病理疾病) 还可以被视为炎性疾病 (急性或慢性), 因为其常常涉及补体激活和招募 / 浸润多种白细胞例如嗜中性粒细胞、肥大细胞、嗜碱性粒细胞等。

[0385] 本发明的组合物将以与剂型匹配的方式并且以治疗 / 预防有效的量施用。以多种方式 (例如通过摄取或注射或吸入) 容易地施用制剂。

[0386] 另一些治疗方案可以与本发明的多肽联合施用。联合治疗可以以同时或依次的方案施用。当依次施用时, 所述组合可以在两次或更多次施用中施用。联合施用包括共同施用, 使用分开的制剂或单一的药物制剂并且以任一顺序持续施用, 其中优选地有两种 (或所有) 活性剂同时发挥其生物学活性的时间段。

[0387] 现在将参照下列非限制性实施例来进一步描述本发明。

实施例

[0388] 实施例 1. 将人 VL 亚文库克隆到 IGHV3-23 展示载体中

[0389] 筛选用于将会在大肠杆菌胞质中充分表达并可溶的 IGHV3-23 的人轻链伴侣, 本发明人克隆了所有 10 个 λ 和 5 个 κ 功能性轻链家族作为与 IGHV3-23 融合的 scFv。所述 scFv 文库被克隆到阿拉伯糖可诱导的表达构建体中并且与赋予细胞壁结合的下游结构域 (肽聚糖 (PG) 结合结构域)、表达报道子结构域 (SNAP ;New England Biolabs) 和 DNA 结合结构域 (DBP) 以该次序进一步融合。当外部和内部细菌宿主细胞膜通过去污剂或有机溶剂透化时, 这些下游结构域能够通过将融合蛋白锚定到 DNA 和细胞壁二者保持 scFv 部分。

[0390] 可以从来自人外周血单核细胞 (PBMC) 制备的 cDNA 扩增 λ 和 κ 轻链家族。分别地在 7 和 11 次 PCR 反应中扩增 κ 和 λ 轻链亚家族。每个亚文库独立地第一次筛选以广泛地表征含有明显可溶性成员之文库的百分比。随后作为单个克隆筛选含有可评估的可溶性成员百分比 (>1%) 的亚文库。

[0391] 寡核苷酸引物基于由 Hust 和 Dubel (2010) 描述的序列, 具有末端修饰用于通过 BsmBI 克隆。对初始设计的针对轻链的 C1 恒定结构域的反向引物进行进一步的改变。这被认为包含非必要的序列和经设计针对轻链之 J 区域的简并引物。用于扩增轻链区的寡核苷酸序列列于表 2 中。

[0392] 表 2. 用于扩增轻链区的寡核苷酸序列

[0393]

κ 第一轮**寡核苷酸序列**

HVK1 F1	GAC ATC CAG ATG ACC CAG TCT CC (SEQ ID NO: 31)
HVK1 F2	GMC ATC CRG WTG ACC CAG TCT CC (SEQ ID NO: 32)
HVK2 F	GAT RTT GTG ATG ACY CAG WCT CC (SEQ ID NO: 33)
HVK3 F	GAA ATW GTG WTG ACR CAG TCT CC (SEQ ID NO: 34)
HVK4 F	GAC ATC GTG ATG ACC CAG TCT CC (SEQ ID NO: 35)
HVK5 F	GAA ACG ACA CTC ACG CAG TCT CC (SEQ ID NO: 36)
HVK6 F	GAW RTT GTG MTG ACW CAG TCT CC (SEQ ID NO: 37)
HVKCL R	ACA CTC TCC CCT GTT GAA GCT CTT (SEQ ID NO: 38)

λ 第一轮

HVL1 F1	CAG TCT GTG CTG ACT CAG CCA CC (SEQ ID NO: 39)
HVL1 F2	CAG TCT GTG YTG ACG CAG CCG CC (SEQ ID NO: 40)
HVL2 F	CAG TCT GCC CTG ACT CAG CCT (SEQ ID NO: 41)
HVL3 F1	TCC TAT GWG CTG ACW CAG CCA CC (SEQ ID NO: 42)
HVL3 F2	TCT TCT GAG CTG ACT CAG GAC CC (SEQ ID NO: 43)
HVL4 F1	CTG CCT GTG CTG ACT CAG CCC (SEQ ID NO: 44)
HVL4 F2	CAG CYT GTG CTG ACT CAA TCR YC (SEQ ID NO: 45)
HVL5 F	CAG SCT GTG CTG ACT CAG CC (SEQ ID NO: 46)
HVL6 F	AAT TTT ATG CTG ACT CAG CCC CA (SEQ ID NO: 47)
HVL7/8 F	CAG RCT GTG GTG ACY CAG GAG CC (SEQ ID NO: 48)
HVL9/10 F	CAG SCW GKG CTG ACT CAG CCA CC (SEQ ID NO: 49)
01115 HVLCL R	TGA ACA TTC TGT AGG GGC CAC TG (SEQ ID NO: 50)
01116 HVLCL R2	TGA ACA TTC CGT AGG GGC AAC TG (SEQ ID NO: 51)

κ 第二轮

[0394]

HVK1 2F1	ATCTAGAATG GGA GAC GGT GAC ATC CAG ATG ACC CAG TCT CC (SEQ ID NO: 52)
HVK1 2F2	ATCTAGAATG GGA GAC GGT GMC ATC CRG WTG ACC CAG TCT CC (SEQ ID NO: 53)
HVK2 2F	ATCTAGAATG GGA GAC GGT GAT RTT GTG ATG ACY CAG WCT CC (SEQ ID NO: 54)
HVK3 2F	ATCTAGAATG GGA GAC GGT GAA ATW GTG WTG ACR CAG TCT CC (SEQ ID NO: 55)
HVK4 2F	ATCTAGAATG GGA GAC GGT GAC ATC GTG ATG ACC CAG TCT CC (SEQ ID NO: 56)
HVK5 2F	ATCTAGAATG GGA GAC GGT GAA ACG ACA CTC ACG CAG TCT CC (SEQ ID NO: 57)
HVK6 2F	ATCTAGAATG GGA GAC GGT GAW RTT GTG MTG ACW CAG TCT CC (SEQ ID NO: 58)
HVKCL 2R	GATCAG GGT CTG AGA CGA TTT RAT HTC CAS YYK KGT CCC HBS GCC RAA VGT (SEQ ID NO: 59)

λ 第二轮

HVL1 2F1	ATCTAGAATG GGA GAC GGT CAG TCT GTG CTG ACT CAG CCA CC (SEQ ID NO: 60)
HVL1 2F2	ATCTAGAATG GGA GAC GGT CAG TCT GTG YTG ACG CAG CCG CC (SEQ ID NO: 61)
HVL2 2F	ATCTAGAATG GGA GAC GGT CAG TCT GCC CTG ACT CAG CCT (SEQ ID NO: 62)
HVL3 2F1	ATCTAGAATG GGA GAC GGT TCC TAT GWG CTG ACW CAG CCA CC (SEQ ID NO: 63)
HVL3 2F2	ATCTAGAATG GGA GAC GGT TCT TCT GAG CTG ACT CAG GAC CC (SEQ ID NO: 64)
HVL4 2F1	ATCTAGAATG GGA GAC GGT CTG CCT GTG CTG ACT CAG CCC (SEQ ID NO: 65)
HVL4 2F2	ATCTAGAATG GGA GAC GGT CAG CYT GTG CTG ACT CAA TCR YC (SEQ ID NO: 66)
HVL5 2F	ATCTAGAATG GGA GAC GGT CAG SCT GTG CTG ACT CAG CC (SEQ ID NO: 67)
HVL6 2F	ATCTAGAATG GGA GAC GGT AAT TTT ATG CTG ACT CAG CCC CA (SEQ ID NO: 68)
HVL7/8 2F	ATCTAGAATG GGA GAC GGT CAG RCT GTG GTG ACY CAG GAG CC (SEQ ID NO: 69)
HVL9/10 2F	ATCTAGAATG GGA GAC GGT CAG SCW GKG CTG ACT CAG CCA CC (SEQ ID NO: 70)
HVLCL 2R	GATCAG GGT CTG AGA CGA RRY GRT SAS CTB SGT BCC HBY DCC RAA BAC (SEQ ID NO: 71)

[0395] 使用 Vent DNA 聚合酶 (New England Biolabs) 在两轮 PCR 中扩增用于 λ 和 κ 轻链亚文库的 VL 基因。使用 BsmBI (New England Biolabs) 将每个亚文库克隆到 RED 展示载体中。估计每个文库生产约 20 至 40,000 个集落。

[0396] 实施例 2. 使用保留封装展示 (RED) 筛选人 scFv 融合

[0397] 作为可溶性的初始筛选,刮取每个文库平板并悬于 10mL 的 LB / 甘油 (10%) 中。悬液的一部分 (约 50uL) 在 1mL 的 LB 培养基 (每升 10g 蛋白胨、5g 酵母提取物、10g NaCl)

中于 37℃ 生长 1 小时并且随后用 0.2% 阿拉伯糖诱导并在 25℃ 进一步生长 2 小时。在这一点,通过将细菌沉淀在 LB 培养基中 0.5% 正辛基-β-D-硫代葡糖苷(8TGP) 中在 25℃ 下重悬 10 分钟以对细胞进行透化。经透化细胞通过在 LB 培养基中沉淀并重悬洗涤一次,在通过添加 SNAP-表面 488 试剂(S9124S;New England Biolabs) 标签经诱导的 scFv 融合蛋白并且在 25℃ 孵育 20 分钟之前。经标记的细胞随后通过在 PBS 中沉淀和重悬而再次洗涤,然后安放样品用于通过荧光显微术观察。

[0398] 显微镜检查显示尽管所有文库在每个视野内有一些细胞,其表现为良好表达且可溶,仅发现代表 Vλ 1、Vλ 3 和 Vλ 6 进化枝的亚文库具有 >1% 的细胞具有可溶性形态。因此除了 Vλ 1、Vλ 3 和 Vλ 6 以外的所有亚文库被认为是不具有足够高频率的可溶性克隆并且不再继续筛选。

[0399] 将亚文库 Vλ 1、Vλ 3 和 Vλ 6 稀释涂板以产生单个克隆并单独地筛选可溶性。因此,将亚文库 Vλ 1、Vλ 3 和 Vλ 6 稀释涂板允许清楚地挑选单个集落,随后对其进行诱导用于表达并如上所述准备用于显微术。

[0400] 图 1 表明,良好表达的可溶性 scFv 克隆的典型外观(1A 和插图),伴随良好表达的但不溶的 cFv 克隆(1B 和插图)。所述细胞在如上所述的透化之后用 SNAP 荧光团标记。不溶性克隆的独特聚集与可溶性克隆的更分散和周边定位有所不同。

[0401] 然后经证明可溶性表达的 scFv 克隆在 37℃ 于 100 μg / mL 氨苄青霉素选择下过夜生长,使用标准方法(Sambrook 等, Molecular Cloning: A Laboratory Manual, 第 3 版, Cold Spring Harbour Laboratory Press(2001)) 制备质粒并且用所述表达构建体的上游启动子区中的引物测序。然后使用 NCBI BLAST 算法针对人基因组 GenBank 数据库分析测序文件。

[0402] 图 2 代表经选择可溶性克隆的多重比对,所述经选择可溶性克隆与 VL 基因 IGLV3-1、IGLV3-21 和 IGLV6-57 高度相似或完全一致。使用 CLUSTAL X 准备所述多重比对。在比对之上的可溶性克隆(通过分离编号列出)的图像对应于下面的比对序列。

[0403] 克隆通过分离和测序检查。

[0404] 筛选 779 个可溶性个体的亚文库成员产生 IGLV1-40(SEQ ID NO:18) 的 11 个克隆, IGLV1-44(SEQ ID NO:21) 的 2 个克隆, IGLV1-47(SEQ ID NO:24) 的 3 个克隆, IGLV1-51(SEQ ID NO:15) 的 3 个克隆, IGLV3-1(SEQ ID NO:6) 的 25 个克隆, IGLV3-19(SEQ ID NO:27) 的 2 个克隆, IGLV3-21(SEQ ID NO:9) 的 4 个克隆, IGLV6-57(SEQ ID NO:12) 的 18 个克隆。可溶性 scFv 克隆之序列的分析显示对于成员 IGLV1-40、IGLV1-51、IGLV3-1 和 IGLV3-19 有明显的天然(naive) 序列,其在体内 B 细胞呈递期间未经亲和力选择或成熟。此外,某些 IGLV6-57 克隆有高可溶性,与种系序列的翻译比仅具有 1 个(2 个克隆)或 2 个(1 个克隆)氨基酸改变,总体的同一性分别为 99% 和 98%。

[0405] 因此,与 Tse 等,(2002)(其发现包含 V_H3 结构域的可溶性抗体与 VL κ 1 和 4 伴侣完全配对)在酵母的胞质中先前筛选可溶且稳定之人抗体的结果相反,本发明人发现 VL κ 亚文库作为一类具有差的表观可溶性,伴随 >99% 的 VL κ 亚文库克隆特异性地与 IGHV3-23 结构域配对,在大肠杆菌中较差地表达或者显示错误折叠的信号。Tissot 等(W003 / 097697)还对可溶性人 scFv 进行了 Y2H 筛选,并且报道其可溶性 scFv 是与 VH1a、VH1b 或 VH3 之成员最密切相关的序列,与序列组合的进化枝与 VL κ 1 或 VL λ 1 或 VL λ 3 进化枝的成员最密

切地相关。然而,其最优构造是与 VH1b 配对的 VL λ 3。

[0406] 然而,Tse 等 (2002) 和 Tissot(W003 / 097697) 二者都应用功能性筛选 (即抗体与抗原靶标体内结合) 作为可溶性的进一步要求,其具有阳性 Y2H 信号的产出抗体需要以下二者:1) 可溶性,和 2) 靶标结合,并且由此根据需要基本上从其种系序列突变。

[0407] 在与 IGHV3-23 结构域可溶性融合的筛选中分离的 VL 成员的大多数是 IGLV3-1,也称作 DPL23。虽然一些克隆具有众多突变,显著数目与 IGLV3-1 种系 V 序列 (SEQ ID NO:4) 相同,表明当所述种系序列与 IGHV3-23 结伴时在胞质中固有的稳定和溶解度。

[0408] IGLV3-1 在人免疫系统中有适度的高表达,占 λ 轻链的 15% (Knappik 等,2000),但不是最丰富的经表达 λ 成员 (DPL11)。在已发表的文献中未表征,缺乏任何特定引用,并且没有报道具有高度一致的结构。尽管先前已经构建使用 IGHV3-23 的人工 scFv 支架文库,主要基于其体内的相对表达水平选择 VL 伴侣,即高表达的 DPK22 (Pini 等,1998;和 Ge 等,2010)、DPL3 (亦称 IGLV1-47) (Kobayashi 等,1997;Soderlind et al,2000) 和 DPL16 (亦称 IGLV3-19) (Viti 等,2000)。

[0409] Ewert 等 (2003) 在 Morphosys HuCAL™ 文库上进行的人可变结构域谱的热稳定性的总体分析是仅有的已发表报道。在其题为“Biophysical Properties of Human Antibody Variable Domains”的文章中,作者检测了当在大肠杆菌周质中表达时,单个结构域的稳定性和结构域对 (V_L : : V_H) 的稳定性二者。

[0410] 已表明,与 IGHV3-23 相关的 VH3 共有序列对于重链可变区的热稳定性和可溶性最稳定,而 V_{κ} 3 共有序列是最稳定的轻链可变区。

[0411] 产生最稳定配对的 V_H : : V_L 组合是 H3: : κ 3、H1b: : κ 3、H5: : κ 3 和 H3: : κ 1 之间形成的那些。值得注意的是,没有一个最稳定的配对包括 VL3 家族。此外,使用 V_H 3 伴侣 (IGHV3-23) 构建我们的 scFv 文库, (以其自身) 不足以在大肠杆菌胞质中表达时赋予融合蛋白稳定性,由于在大多数亚文库中绝大多数的克隆是错误折叠或差表达的。

[0412] 总之,在现有技术的基础上,使用结构域 IGHV3-23 和 IGLV3-1 的组合作为 scFv 融合不能预测在还原性环境 (例如大肠杆菌胞质) 中表达时具有增强的稳定性和可溶性。

[0413] 实施例 3. scFv 克隆的热稳定性测试

[0414] 在 25°C 诱导并表达的 VL 亚文库的初始筛选之后,对所述 scFv 克隆进行进一步的筛选以对克隆和家族的热稳定性进行评级 (grade)。

[0415] 每个克隆在透化和用 SNAP 标记之前,在 26°C、28°C、30°C、32°C、34°C、36°C 和 38°C 的温度下诱导 90 分钟,如实施例 2 的描述。使用荧光显微术对克隆的可溶性进行评分,如实施例 2 的描述。

[0416] 图 3 展示出两个克隆 (一个 IGLV3-1 和一个 IGLV3-21) 的行为,在提高的温度下表达。所述 scFv 融合蛋白保持可溶,直到至少用于 IGLV3-1 克隆的 36°C,但是 IGLV3-21 克隆在 32°C 至 34°C 显示错误折叠的迹象。

[0417] 这一表达温度梯度表明,与 IGLV3-1 和 IGLV6-57 相关的或一致的 scFv 克隆被判定为具有最佳可溶性的类别,尽管其他 λ 基因的个体克隆也被表现出不同程度的可溶性。图 4 展示出从筛选分离的每类 VL 结构域的代表性克隆的可溶性。

[0418] 通过显微术 scFv 之表观可溶性不是人为的,这通过将所述 scFv 片段、以及来自人 Titin 的下游 127 结构域亚克隆到具有 C 端 FLAG 表位的表达构建体中进行确认。scFv: :

127 : :FLAG 融合蛋白用阿拉伯糖在范围为 26℃至 38℃的温度下诱导并且使用超声处理裂解大肠杆菌。通过离心(14K1 分钟)从不溶的碎片和蛋白质聚集体分离可溶性蛋白质。

[0419] 图 4 展示出 IGLV3-1 在大肠杆菌中于 25℃表达时优异的可溶性。scFv : :127 : :FLAG 融合蛋白完全在可溶性的级分中。其未显示出总蛋白中的小级分的 N 端切割,但是这在蛋白质在非变性条件下提取时消失,表明其由 8TGP 洗涤剂的透化释放的周质蛋白酶的相互作用引起。

[0420] 因此,由于来自大肠杆菌中可溶性筛选的 IGLV3-1 的高频恢复,其还表征用于可溶性 scFv 文库的示例性支架的必要特性,在胞质中于接近 37℃的温度下稳定并且耐受多样化的 CDR3 环。在大肠杆菌胞质中在温度范围 28℃至 38℃下测试 IGLV3-1 的热稳定性。发现当与轻链 J1 和 J2 区、以及使用简并寡核苷酸作为引物在 PBMC cDNA 的 PCR 期间形成的 J 区偶联时,直至 36℃都高度可溶。在 36℃及以上,其表现一定程度的错误折叠。这通过免疫荧光和通过 FLAG- 标记的 scFv 的 Western 印记二者进行确认。

[0421] 实施例 4. IGLV3-1J 结构域交换

[0422] 用于扩增 VL 结构域的在表 2 中列出的简并寡核苷酸序列不得不将 7 个不同的 λ J 区引入人基因组(表 3)。由此,通过筛选分离的克隆具有杂交 λ J 区,其代表可能降低其折叠稳定性的非典型序列。

[0423] 表 3. 人 λ J 区

[0424]

[0425]	λ J 区	氨基酸序列
	J1	VFGTGTKVTVs (SEQ ID NO: 72)
	J2	VFGGGTKLTVs (SEQ ID NO: 73)
	J3	VFGGGTKLTVs (SEQ ID NO: 74)
	J4	VFGGGTQLIIs (SEQ ID NO: 75)
	J5	VFGEGTELTVs (SEQ ID NO: 76)
	J6	VFGSGTKVTVs (SEQ ID NO: 77)
	J7	VFGGGTQLTAs (SEQ ID NO: 78)

[0426] 对具有骨架区之种系序列的最稳定 IGLV3-1 克隆的 J 区进行比较显示出与 J 区 1 和 2 / 3 最高的相似性。因此,用种系 λ J1 或 J2 / 3 序列(表 3)替代杂交 J 区(“VFGTGTKLIIs” (SEQ ID NO: 79))以检测 IGLV3-1 支架的热稳定性会不会被进一步提高。在温度 30℃、32℃、34℃、36℃和 38℃之间测试变体。主观上地,感觉受测试克隆的 λ J1 比 J2 / 3 或者初始杂交 J 区给出稍微更好的折叠和可溶性。

[0427] 图 5 表明具有 λ J 区替换为 J1 或 J2 的初始克隆(#8. 93)的热稳定性行为。

[0428] 实施例 5. IGLV3-1 和 IGHV3-23 对 CDR3 多样化的耐受

[0429] 对于可用作亲和文库骨架的 scFv,其需要耐受在 CDR3 区的替换。对于在还原性环境(例如大肠杆菌胞质溶胶)中表达的 scFv(其中缺乏稳定对结构域内二硫键的稳定),这是尤其真实的。

[0430] 为测试优选的 scFv 支架 (IGLV3-1 : IGHV3-23) 的稳定性,每个结构域的 CDR3 区被分别地多样化。因此,测试 IGLV3-1 和 IGHV3-23 基因二者对 CDR3 多样化的耐受。图 2 示出两个序列的围绕 CDR3 的区,以及假想的多样化。IGLV3-1 CDR3 的 2 个氨基酸被“-NNNGGNNN-”(SEQ ID NO:29) 区替代(其中 N 是除 Trp、Gln、Lys、Glu、Met 以外的氨基酸)。类似地,12 个氨基酸的 IGHV3-23 CDR3 被“-NNNGGNNN-”(SEQ ID NO:29) 区替代。分别地测试每个结构域的可溶性和作为克隆的混合文库的表达。此外,还测试随机挑选的单个克隆并对其进行测序以确定期望的多样性。

[0431] IGLV3-1 CDR3 从 91 位残基起被替换,通过使用(反向)寡核苷酸序列的 PCR 修饰 IGLV3-1 结构域:

[0432] GATCAGGGTCTGAGACGAGACCGTCACTTTCGTACCGGTGCCGAACACCACAGTANNANNANNTCCGCC ANNANNANNGTCCACGCCTGACAGTAATAGTCAGC (SEQ ID NO:80)

[0433] 该序列的 (Fev / comp) 翻译给出(其中“N”是除 Trp、Gln、Lys、Glu、Met 以外的氨基酸):

[0434] ...ADYYCQAWD(91)NNNGGNNN TVVFGTGTKVTVS S (SEQ ID NO:81)

[0435] IGLV3-1CDR3 替代是轰动性的成功。具有在 30°C 下出现具有蛋白质诱导的群体是具有很好表达的可溶性克隆。单独分析了 40 个克隆并且 36 个被评定为具有优秀的可溶性和表达。对 4 个失败的克隆和 16 个其他具有良好可溶性和表达之克隆的整个 vL 结构域进行测序。确认 4 个失败的克隆由于在用于扩增基因的长寡核苷酸引物中框移(frameshift)而失败。具有正确阅读框的所有其他受测试的克隆具有氨基酸的随机混合物,并且表现出种系 IGLV3-1 骨架非常耐受 CDR3 多样化。

[0436] 因此,对于 IGLV3-1,多样化的 CDR3 文库于 30°C 在大肠杆菌胞质中表达时的可溶性出人意料的高。约 90% 的克隆是可溶性的且具有高表达。10% 的克隆具有低表达或不表达,或者错误折叠,进行测序,显示出框移,主要在必要的约 100 碱基长的反向引物区。在使用长寡核苷酸构建合成文库时,碱基缺失是常见的错误,并且其他团体已经发展了基于抗生素选择以富集框内(in-frame)等位基因的预筛选策略。(例如 Ge 等,2010)

[0437] VH 结构域的 CDR3 区,IGHV3-23 使用反向高核苷酸替代,类似于以上描述的方法,从残基 98 起通过使用(反向)寡核苷酸序列的 PCR 修饰 IGHV3-23 结构域:

[0438] GATCAGGGTCTGAGACCCGCTGCTCACGGTAACCATGGTACCTTGACCCCA AATATCAAACGCANNAN NANNGCCANNANNANNTTTCGCACAGTAGTAAA CAGC (SEQ ID NO:82)

[0439] 该序列的 (Fev / comp) 翻译给出(其中“N”是除 Trp、Gln、Lys、Glu、Met 以外的氨基酸):

[0440] VYYCAK(98)NNNGNNN AFDIWGQGTMTV (SEQ ID NO:83)

[0441] 经证明 IGHV3-23 与 IGLV3-1 结构域对 CDR3 多样化刚好一样强。80% 受测试克隆显示出可溶性、高表达,而 20% 差表达的在测序后可以解释为是由于在骨架中的保守错配,或更常见地,在长的寡核苷酸引物的区域中单碱基对的缺失,从而改变蛋白质翻译的框。

[0442] 因此,对于 IGHV3-23,多样化的 CDR3 文库于 30°C 在大肠杆菌胞质中表达时的可溶性再次出人意料地高(约 80%)。另外,融合蛋白的框移是很多阴性结果的原因。将 CDR3 环从 12 个氨基酸缩短至 7 个与亲代克隆相比也提高了该文库的可溶性。

[0443] 图 6A 展示出具有经多样化 IGLV3-1CDR3 的 4 个独立克隆的可溶性和高表达。图 6B

展示出具有经多样化 IGHV3-23CDR3 之克隆的整个群体的样本。因此,概括地说,IGLV3-1 : IGHV3-23 骨架是构建亲和文库的示例性支架,即与人种序列有同一性并且保持与具有多样化序列的 CDR3 环之替代的强可溶性。此外,将所述支架与 RED 蛋白质展示方法在大肠杆菌胞质中组合能够同时筛选亲和力蛋白稳定性、表达二者和与靶分子的结合。所述支架在大肠杆菌胞质的还原性环境中是高度稳定和可溶性的,所述胞质中缺乏稳定的结构域内二硫键,其对于几乎所有其他 scFv 蛋白的折叠和稳定性是基本要求。该支架将能够在大肠杆菌胞质中低成本地产生亲和力试剂,用于研究、治疗或诊断用途,以及这样的试剂在哺乳动物细胞的胞质中用于靶向内源形蛋白质的用途。

[0444] 实施例 6. 构建多样化的 IGLV3-1 : IGHV3-23scFv 文库

[0445] 多样化 IGLV3-1 : IGHV3-23 支架,使用实施例 5 描述的策略将氨基酸序列 ‘NNNGNNN’ (SEQ ID NO :86) 和 ‘NNNGNNN’ (SEQ ID NO :87) 分别引入 VL 和 VH 的 CDR3 区。

[0446] 多样性的引入首先通过创建由如下的 IGLV3-1 之骨架序列和 IGHV3-23 之 J 区组成的基本支架 :

[0447] 骨架序列 :

[0448] ATG GGA GAC GGT CAG TCT GTG CTG ACT CAG CCA CCCTCAGTGTCCGTGTCCCCAGGA CAGACAGCCAGCATCACCTGCTCTGGAGATAAATTGGGGGATAAATATGCTTGCTGGTATCAGCAGAAGCCAGGCC AGTCCCCCTGTGCTGGTCATCTATCAAGATAGCAAGCGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAACT CTGGGAACACAGCCACTCTGACCATCAGCGGGACCCAGGCTATGGATGAGGCTGACTATTACTGTCAGGCG TGG GACTgagacctagacggtctct gcg TTT GAT ATT TGG GGT CAA GGT ACC ATG GTT ACCGTG AGC AGC TGC TCT CaG ACC (SEQ ID NO :88).

[0449] 用 PG 和 DNA 结合结构域元件通过侧翼 BsmBI 位点将所述骨架克隆到 RED 胞质表达载体中。间插序列 (VL J 区和 IGHV3-23 骨架) (SEQ ID NO :89) 在独立的质粒上编码,其充当使用简并引物 (SEQ ID NO :90 和 91) 之 PCR 的模板,所述简并引物分别包含 VL 和 VH 区在 5' 和 3' 端二者的 CDR3 多样性。这些引物序列具有末端 BsaI 限制位点,其能够将 PCR 产物无缝地克隆到支架中的合适取向的 BsaI 位点中。

[0450] 间插序列 :

[0451] ACT GTG GTG TTC GGC acc ggt acg aaa gtg acT gtc TCA TCT CAG ACC

[0452] GGTGGTTCTGGTGGTGGTGGTTCTGGCGGCGGCGGCTCCGGTGGTGGTGGATCCGAAGTCCAACTGCTG GAGTCCGGCGGTGGCCTGGTGACGCCAGGTGGCAGCCTGCGCCTGAGCTGCGCCGCATCCGGTTTTACTTTTCAGCAG CTACGCGATGTCGTGGGTGCGCCAGGCACCGGGCAAGGGCCTGGAGTGGGTGTCAGCGCCATCAGCGGTAGCGGCGGTT CTACGTATTATGCGGACAGCGTCAAGGGCCGTTTACCATCAGCCGTGACAATTCCAAAAACACCCTGTACTTGCAG ATGAACAGCTTGCGTGCGGAAGATACGGCTGTTTACTACTGTGCGAAA (SEQ ID NO :89)

[0453] 简并引物 1 :

[0454] gatcag ggtctca ggac NNT NNT NNT ggc gga NNT NNT NNT ACT GTG GTG TTCGGC acc ggt acg aaa gtg (SEQ ID NO :90)

[0455] 简并引物 2 :

[0456] GATCAGGGTCTCAACGCANNANNANNGCCANNANNANNTTTCGCACAGTAGTAAACAGCCGTATCTTC (SEQ ID NO :91)

[0457] 用 BsaI 切割 10 μ g 的基本支架载体。使用 Sureclean (Bioline) 将经切割载体沉

淀。使用引物 SEQ ID No :90 和 91 从核心骨架作为模板的 PCR 产生含有 CDR3 多样性区的插入片段。凝胶纯化 2 μ g 的插入片段 PCR, 然后用 BsaI 消化。使用 Sureclean 沉淀 PCR 消化产物。使用 T4DNA 连接酶连接等摩尔量的经消化载体和 PCR 插入片段。将连接热灭活并顺次电穿孔进入银大肠杆菌 (Argentum E. coli) 细胞 (Alchemy Bio)。将经电穿孔细胞涂到 15cm LB+ 羧苄青霉素 (40 μ g / mL)+ 葡萄糖 (0.1%) 琼脂平板上。总文库大小 $>1 \times 10^8$ 个独立克隆。

[0458] 文库构建的质量通过经多样化 scFv 的表达评估。如前所指出的, 可溶性、部分可溶性或不可溶性融合伴侣的表达可通过在 RED 展示系统中使用肽聚糖 (PG) 结合结构域和发光表达报道子例如 SNAP (New England Biolabs) 通过 scFv 的外观直接评估。可溶性融合蛋白显著地均匀地分布在细胞之周界周围, 由于其自由的扩散, 一旦膜被透化则与细胞壁结合 (例如如图 1A)。比较而言, 不溶性融合蛋白形成浓密着色的聚集体, 其不迁移至细胞壁 (例如如图 1B)。部分可溶性融合具有各自的一些特性。我们之前已经发现在以上所述的融合蛋白的外观与在 Western 印记中可溶性 / 不溶级分中出现的量之间有极佳的相关性。

[0459] 通过这一经验标准, 我们多样化的 scFv 文库由约 90% 可溶性并且良好表达的成员组成, 其表明 IGLV3-1 : IGHV3-23 支架对插入的 CDR3 多样性的耐受。图 9 是经表达文库之样品的 SNAP 标记图像。

[0460] 为了进一步确认文库由随机的 V_L 和 V_H CDR3 区构成, 对 10 个独立的克隆进行测序。测序显示 CDR3 环 (在表 4 中显示) 的组成是期望的, 通过用于简并性的“NNT”核酸三联体产生密码子多样性, 即不存在终止密码子和氨基酸 W、Q、M、K 和 E。

[0461] 表 4. 在随机文库中样本 CDR3 环序列

[0462]

克隆	VL CDR3	VH CDR3
1	PFGGGGYV (SEQ ID NO :92)	PPHGAPA (SEQ ID NO :93)
2	LCIGGVAS (SEQ ID NO :94)	HNSGNNF (SEQ ID NO :95)
3	FVSGGIST (SEQ ID NO :96)	FNFGNAY (SEQ ID NO :97)
4	INSGGASF (SEQ ID NO :98)	XXXGTNY (SEQ ID NO :99)
5	SRAGGCNG (SEQ ID NO :100)	FDYGHCI (SEQ ID NO :101)
6	TNRGGVCA (SEQ ID NO :102)	TAAGVPF (SEQ ID NO :103)
7	混合的克隆	混合的克隆
8	FSTGGCAF (SEQ ID NO :104)	AICGATA (SEQ ID NO :105)
9	FXGGGDGT (SEQ ID NO :106)	PYRGSFF (SEQ ID NO :107)
10	IIPGGLYA (SEQ ID NO :108)	PVIGSNT (SEQ ID NO :109)

[0463] 实施例 7. 筛选与 mAG1 靶标结合的 IGV3-1 : IGVH3-23 文库

[0464] 针对与靶蛋白质 mAG 结合的克隆对多样化文库进行筛选。Azami-绿色 (AG) 是维多利亚水母 (*Aequorea victoria*) 绿色荧光蛋白 (GFP) 的远直向同源物 (distal ortholog)。尽管序列一致性低 (5%), 相似的绿色荧光具有在 492 nm 的吸收峰和在 510 nm 的发射峰。单体形式 (mAG) 由 Karasawa 等 (2003) 报道并且为了在大肠杆菌中最优地表达而通过 DNA2.0 (USA) 重编码。包含 C 端大肠杆菌 BirA 生物素化模体和 6×His 标签以辅助纯化和 mAG 基质附着。mAG-BioHis6 蛋白的氨基酸序列如 SEQ ID NO:110 所列。

[0465] mAG BioHis6 蛋白序列:

[0466] MVSVIKPEMKIKLCMRGTVNGHNFVIEGEGKGNPYEGTOILDLNVTEGAPLPFAYDILTTVFQYGNRAF
TKYPADIQDYFKQTFPEGYHWERSMTYEDOGICTATSNISMRGDCFFYDIRFDGTNFPNGPVMQKKT
LKWEPSTEK
MYVEDGVLKGDVNMRLLLLEGGGHYRCDFKTTYKAKKEVRLPDAHKIDHIEILKHKDYNKV
KLYENAVARYSMLPS
QAKSGGLNDIFEAQKIEWHEDTGGSHHHHHH (SEQ IDNO:110)

[0467] 诱导多样化文库的 10^{10} 细胞 (代表约 100 倍冗余) 用于 RED 展示, 如实施例 2 和 WO 2011 / 075761 中所述。经透化细胞悬于 50 mL PBS 中并且用经纯化的 mAG 标记, 所述 mAG 已与 MACS 链霉亲和素缀合的微珠预结合 (130-048-102, Miltenyi Biotec)。细胞和微珠在 4°C 轻轻地搅拌过夜。随后将其加载到 3×LS 柱上 (130-042-401, Miltenyi Biotec), 所述 3×LS 柱被固定在磁性支持物上。每个柱用 50mL 的 PBS 洗涤。细胞用 10mL PBS 洗脱, 汇集并沉淀 (pelleted)。通过碱裂解从细胞沉淀中分离在 RED 展示载体中编码文库的质粒 DNA。然后将质粒电穿孔返回到银细胞 (*Argentum cell*) 中并重复诱导、结合和柱纯化。亲和筛选四次重复之后, 通过荧光显微镜观察到低丰度的 RED 经透化细胞与 mAG 蛋白结合。在第五次重复通过 FACS 分选 mAG 结合的经透化细胞。用于 FACS 的细胞使用 SNAP 配体标签, 以对融合蛋白表达和 mAG 进行归一化。在从 2.46×10^8 总事件中收集 4,428mAG 次阳性事件期间, 细胞群体的 FACS 示出丰度大约在 10^5 个细胞中一次结合事件。来自作为 FACS 产出的 mAG 阳性细胞的 scFv 通过使用在 scFv 序列侧翼的寡核苷酸引物的 PCR 回收并且产物被再克隆回到 RED 展示载体中。分析对 mAG 结合阳性的细胞的最终筛选产出, 示出克隆的约 40% (23 / 60) 为 mAG1 阳性。因此, FACS 阶段能够从文库背景中以约 10^5 倍富集阳性细胞。

[0468] 图 10 示出将 mAG 与 RED 经透化细胞相结合用于克隆, 对于 mAG 结合为阴性 (克隆 25) 或阳性 (克隆 34)。

[0469] 对 20 个 mAG 阳性克隆进行测序并且发现所有 20 个一致。结合 mAG 的 IGLV3-1 : IGHV3-23 克隆的蛋白质序列在以下 SEQ ID NO:111 中列出 (CDR3 序列加粗且放大的字体表示, 并且肽接头用下划线表示)。发现 VL CDR3 为 'FNLGGCGD' 并且 VH CDR3 为 'HIDGPVA', 其符合所设计的多样性。

[0470] 抗 m AG 结合 scFv:

[0471] Anti-mAG binding scFv:

[0472] MGDGQSVLTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRP

[0473] SGIPERFSGSNSGNTATLTISGTQAMDEADYYCQAWDFNLGGCGD**TVVFGTGT**KVTVSSQTGGSGGGGS
GGGGSGGGGSEVQLLESGGGLVQPGGSLRLSCAASGFTFSYAMSWVRQAPGKGLEWWSAISGSGGSTYYADSVKGR
FTISRDNKNTLYLQMNSLRAEDTAVYYCAKHIDGPVAAFDIWGQGTMVTVSSSSQTSILVA (SEQ ID NO:111)

[0474] 为了确定 α -mAG scFv 的性质,将该基因与 C 端 6×His 和 FLAG 表位标签克隆到表达载体中。用阿拉伯糖诱导 scFv 表达并且细胞用 0.5% 8TGP 透化以将可溶性 scFv 释放到上清液中。将不可溶性细胞物质沉淀并且两种提取物的样品都与 SDS-PAGE 加载染料一起煮沸并在 15% SDS-PAGE 凝胶上电泳。将分开的蛋白质转移至硝酸纤维素膜并用 α -FLAG 小鼠单克隆抗体探测。图 11 展示出 α -mAG scFv 几乎完全在可溶性级分中。

[0475] 为了证明 α -mAG scFv 对 mAG 蛋白是特异性的而不仅是“黏着”抗体, α -mAG scFv 经透化细胞用结构上和功能上与 mAG 相关的蛋白质 EGFP 进行标记。这些细胞,尽管结合 mAG,但不结合 EGFP(数据未示出)。为了进一步评价 α -mAG scFv 的特异性,还将 α -mAG scFvHis6-FLAG 蛋白与 IMAC Ni 琼脂糖树脂结合。“干净”mAG 蛋白质(移除 His6 和 FLAG 标签)的粗制细胞裂解物与树脂混合。洗去游离的蛋白质。在图 12 中的荧光显微术成像表明与 α -mAG scFv 连接的树脂珠结合 mAG,而对照珠没有。结合的蛋白质用咪唑洗脱并在 SDS-PAGE 凝胶上电泳。凝胶的考马斯染色(图 13)证实在 α -mAG scFv 样品中的一个条带为 mAG 蛋白质的正确大小,并且没有对 mAG 细胞裂解物明显特异的其他条带。

[0476] 因此,本发明可以用于成功地产生含有随机 CDR3 环之 scFv 多肽的文库并且对其进行筛选以鉴定显示特异性结合活性的 scFv。

[0477] 实施例 8. 使用 α -mAG IGLV3-1 : :IGHV3-23scFv 的 λ 噬菌体展示

[0478] 为了证实在胞质的还原性环境中表现出增强的稳定性并产生折叠的支架的效用,克隆 α -mAG IGLV3-1 : :IGHV3-23scFv 作为与 λ 噬菌体 gpD 衣壳蛋白的 C 端融合。 λ 噬菌体作为用于蛋白质展示的示例性载体已经被报道了很久,其具有多种优于丝状噬菌体的优势。 λ 衣壳蛋白(gpD)在每个噬菌体头上存在约 400 个拷贝,并且是强的并且耐受展示伴侣,允许每衣壳加载的 gpD>80% 成为重组融合蛋白,同时维持感染生存能力(Vaccaro 等, 2006)。此外,其耐受与其 N 或 C 末端的融合。

[0479] 因此, λ 噬菌体或等价的包装载体与标称的丝状噬菌体的单分子展示相比具有文库蛋白质的多价展示。该多价展示可导致从结合溶液中捕获噬菌体的显著效率,至多近 100% 捕获(Mikawa 等,1996)。此外通过可市购的试剂盒可以促进 λ 噬菌体文库的组装,所述试剂盒能高效地包装 λ (至多 2×10^9 pfu / μ g)。

[0480] 然而, λ 噬菌体没有享有使用丝状噬菌体用于抗体展示的普及性,则是由于单一的事实: λ 和相关噬菌体例如 P2 / P4、P22、T7 和 T4 具有溶菌的生活方式,这通过其在胞质中组装而引起。由于绝大多数抗体支架没有氧化的结构域间二硫桥无法有效地折叠,这很大程度上妨碍了使用 λ 噬菌体用于抗体展示。

[0481] 我们鉴定了多个用于形成胞质稳定的 scFv 支架之 IGHV3-23 结构域的 IGLV 伴侣,这使得我们能够展示出用于抗体筛选的 λ 展示的示例性应用。将 α -mAG scFv 作为 C 端融合克隆至与 λ gpD 衣壳蛋白,在阿拉伯糖可诱导的 araBAD 启动子和 araC 调节子的控制下表达融合蛋白。与其他 λ 展示平台类似地将该单元克隆到 λ 噬菌体基因组中(Mikawa 等,1996;Sternberg 和 Hoess,1995;Minenkova 等,2003),具有显著不同所用的 λ 基因组为 cI857gpD⁺RS⁻ 遗传起源的。RS 基因(其构成对细胞裂解所必要的 λ 内溶素(R)和孔蛋白(porin)(S)基因)的缺失描述于国际专利申请号 PCT / AU2012 / 000761,在 λ 类展示(lambdoid display)中使用裂解缺陷噬菌体。用于 λ 类展示的裂解缺陷噬菌体载体使得感染性噬菌体颗粒在胞质内包装。这些颗粒在细胞内持续聚集,其衣壳融合蛋白以高密度

结合至其表面,直到其生长被研究者暂停,通过处理宿主细胞用于胞质 RED 展示。所产生的制备物可以由此通过 FACS 筛选融合蛋白抗原结合。为了释放封装在经透化细胞(其已经通过 FACS 对抗原结合阳性分选)内的噬菌体颗粒,仅需要添加溶菌酶。对于该任务,可以购买商业的高活性溶菌酶制剂(例如 Epicentre 的 Ready-Lyse)。为了完成亲和选择克隆的回收,可以将感染性的噬菌体颗粒感染到大肠杆菌细胞中并作为溶源体而生长。因此,本领域的从业者应理解,与胞质稳定的人抗体支架相结合地使用裂解缺陷的噬菌体使得能够以无噬菌体形式高频率地捕获多价文库克隆,最终的筛选通过 FACS 进行。重要的是,在筛选形式中发生的这种改变没有任何对重构型文库表达构建体的需要。因此,这是具有双能力的筛选系统,具有低克隆选择性的高度平行性筛选(无噬菌体淘选)和具有高克隆选择性但低通量的筛选(经封装噬菌体的 FACS)。

[0482] 为了证明使用本发明之多肽的 λ 噬菌体展示的益处,将模型 α -mAGscFv 融合(作为本发明多肽的很多合适的实例之一)克隆为与 λ 衣壳蛋白 gpD 基因的 C 端融合。

[0483] 所用的 gpD : :s-mAG scFv 融合构建体的 DNA 序列为 :

[0484] ATGACGAGCAAAGAAACCTTTACCCATTACCAGCCGACAGGGCAACAGTGACCCGGCTCATACCGCAACCGCGCCCGGCGATTGAGTGCGAAAGCGCCTGCAATGACCCCGCTGATGCTGGACACCTCCAGCCGTAAGCTGGTTGCGTGGGATGGCACCACCGACGGTGCTGCCGTTGGCATTCTTGCGGTTGCTGCTGACCAGACCAGCACCACGCTGACGTTCTACAAGTCCGGCACGTTCCGTTATGAGGATGTGCTCTGGCCGGAGGCTGCCAGCGACGAGACGAAAAAACGGACCGCGTTTGCCGAACGGCAATCAGCATCGTTGGAGGTAGCGGCGGATCGGATGACGACGATAAGTCTAGAAATGGCGGAGACGGTCAGTCTGTGCTGACTCAGCCACCTCAGTGTCGGTGTCGCCAGGACAGACAGCCAGCATCACCTGCTCTGAGATAAATTGGGGGATAAATATGCTTGCTGGTATCAGCAGAAGCCAGGCCAGTCCCCTGTGCTGGTCATCTATCAAGATAGCAAGCGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAACTCTGGGAACACAGCCACTCTGACCATCAGCGGGACCCAGGCTATGGATGAGGCTGACTATTACTGTGAGGCGTGGGACTTTAATCTTGGCGGATGTGGTGATACTGTGGTGTTCCGGACCGGTACGAAAGTGACTGTCTCATCTCAGACCGGTGGTTCTGGTGGCGGTGGTTCTGGCGGCGGCGGCTCCGGTGGTGGTGGATCCGAAGTCCAACTGCTGGAGTCCGGTGGTGGCCTGGTGCAGCCAGGTGGCAGCCTGCGCCTGAGCTGCGCCGATCCGGTTTTACTTTTACGAGCTACGCGATGTCGTGGGTGCGCCAGGCACCGGGCAAGGGCC TGGAGTGGGTGAGCGCCATCAGCGGTAGCGGCGGTTCTACGTATTATGCGGACAGCGTCAAGGGCCGTTTCACCATCAGCCGTGACAATTCCAAAAACACCCTGTACTTGCAGATGAACAGCTTGCGTGCGGAAGATACGGCTGTTTACTACTGTGCGAAACATATTGATGGCCCTGTTGCTGCGTTTGATATTTGGGGTCAAGGTACCATGGTTACCGTGAGCAACTCGACGATTACAAGGACGATGATGACAAATAA (SEQ ID NO :112)

[0485] 所用的 gpD : : α -mAG scFv 融合蛋白的序列为 :

[0486] MTSKETFTHYQPQGNSDPAHTATAPGGLSAKAPAMTPLMLDTSSRKLVAWDGTTDGAAVGILAVAADQSTTTLTFYKSGTFRYEDVLWPEAASDETKKRTAFAGTAISIVGGSGGSDDDDKSRNGDGQSVLTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRPSGIPERFSGSNSGNTATLTISGTQAMDEADYYCQAWDFNLGGCGDVTVFVGTGKVTVSSQTGGSGGGGSGGGGSGGGGSEVQLLESGLLVQPGGSLRLSCAASGFTFSSYAMSWVRQAPGKGLEWVSAISGSGGSTYYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAVYYCAKHIDGPVAAFIDIWGQGTMTVTVSNSSDYKDDDDK* (SEQ ID NO :113)

[0487] 然后将 gpD : : α -mAG scFv 融合克隆到 λ 展示载体中,在 araBAD 启动子的控制下。通过在 42°C 加热溶源体克隆 15 分钟诱导宿主细胞用于 λ 封装(λ 遗传背景是 cI857gpD⁺RS⁻,具有温度敏感的 cI 抑制子)。温度诱导后立即用 0.2% 阿拉伯糖诱导融合蛋

白。培养物在 32℃ 进一步通气培养 75 分钟。将细胞沉淀并在三分之一培养物体积的 LB 培养基 +0.5% 8TGP 中重悬,并且在 25℃ 孵育 10 分钟以通过用于筛选的 RED 方法透化细胞。为了释放噬菌体,将万分之一培养物体积的 Ready-Lyse (Epicentre) 添加到悬液中。添加一滴氯仿以灭活任何活细胞并且释放的噬菌体颗粒用溶原体形成单位 (cfu / mL) 滴定。制备两种噬菌体储液,一种带有具有经克隆 gpD : : α -mAG scFv 融合的构建体,另一种为空构建体。将 gpD : : α -mAG scFv 融合稀释到 10^9 个空构建体中 1 个克隆,以模拟仅有几个阳性克隆的起始 scFv 文库密度。随后针对生物素化 mAG 结合链霉亲和素珠支持物淘选此“掺杂的”文库。根据本领域从业人员已知的一般用于噬菌体淘选的方法进行淘选。进行两轮淘选,最后一轮回收宿主大肠杆菌菌种中作为溶源体。第三轮的筛选通过 FACS 进行。溶源体细胞按照以上描述的用于噬菌体之加热诱导伴随 gpD : : α -mAG scFv 融合的阿拉伯糖诱导进行处理。然而,替代用溶菌酶处理释放噬菌体颗粒,经透化的细胞改为与 mAG 蛋白孵育。将经透化的细胞洗涤一次,在 TBS+10mM MgSO₄ 中重悬并且随后通过 FACS 对 mAG 结合 (即 mAG 阳性细胞将被标记为绿色) 进行分选。图 14(上) 示出 FACS 分选在操作中的截屏,表明 mAG 阳性细胞的发生率。FACS 后的 mAG 阳性细胞的最终发生率通过荧光显微术 (图 14, 下) 评估为 20%。

[0488] 因此,已经证明与本发明的稳定且可溶性 scFv 支架相结合的 λ 衣壳展示可有力地相对高起始稀释度 (10^9 中 1 个) 分离结合克隆。此外,当与裂解缺陷噬菌体组合并通过 RED 教导的方法处理时,能够进一步放大有益性质以包括 FACS 筛选的能力,而无需再克隆文库成员。

[0489] 这些方法的组合极大地加快了用于筛选具有理想性质 (高表达、高可溶性和高亲和性) 之抗体克隆的过程。

[0490] 本领域技术人员将理解,可以对在特定实施方案中示出的本发明进行多种改变和 / 或修改而不偏离广泛描述的本发明的范围。因此,在所有方面中,认为呈现的实施方案是说明性的而不是限制性的。

[0491] 本文中讨论和 / 或引用的所有出版物整体并入本文。

[0492] 本说明书中包含的文件、行为、材料、设备、文献等的任何讨论仅是为了提供本发明内容。不应当将其认为是承认这些事件中的任一种或全部形成了现有技术基础的一部分,或者是本发明相关领域中的一般常识,如同其存在于本申请的每一条权利要求的优先权日之前。

[0493] 参考文献

[0494] Al-Lazikani et al. (1997) J Mol Biol 273 :927-948.

[0495] Altschul et al. (1993) J. Mol. Biol. 215 :403-410.

[0496] Aufder Maur et al. (2002) JBC 277 :45075-45085.

[0497] Becker et al. (2004) Curr. Opin. Biotech. 15 :323-329.

[0498] Bork et al. (1994) J. Mol. Biol. 242 :309-320.

[0499] Borrebaeck(ed) (1995) Antibody Engineering. Oxford University Press.

[0500] Brezinschek et al. (1997) J. Clin. Invest. 99 :2488-2501.

[0501] Briers et al. (2009) Biochem. Biophys. Res. Comm. 383 :187-191.

[0502] Chen W et al. (2008) JMB 382 :779-789.

- [0503] Chothia and Lesk(1987)J. Mol Biol. 196 :901-917.
- [0504] Chothia et al. (1989)Nature342 :877-883.
- [0505] Contreras-Martinez and DeLisa(2007)JMB372 :513-524.
- [0506] Daugherty et al.(2000)J. Immunol. Methods243 :211-227.
- [0507] Ewert et al. (2003)JMB325 :531-553.
- [0508] Fisher and Delisa(2009)JMB385 :299-311.
- [0509] Froyen et al. (1995)Mol. Immunol. 37 :515-521.
- [0510] Ge et al. (2010)Biotech Bioeng106,347-57.
- [0511] Griffiths et al. (1994)EMBO J13 :3245-3260.
- [0512] Guan et al. (1998)Proc. Natl. Acad. Sci. USA, 95 :13206-10.
- [0513] He et al. (1995)NAR23 :4009-4010.
- [0514] Higgins and Sharp(1989)CABIOS. 5 :151-153.
- [0515] Hust and Dubel(2010)Antibody Eng. Chapter5 :Antibody Engineering, Vol1 ; Springer.
- [0516] Jermutus et al. (2001)PNAS98 :75-80.
- [0517] Jirholt et al. (1998)Gene215,471-476.
- [0518] Jurado et al. (2002)JMB320 :1-10.
- [0519] Kabat(1987and1991)Sequences of Proteins of Immunological Interest. National Institutes of Health.
- [0520] Karasawa et al. (2003)JBC278 :34167-34171.
- [0521] Kenrick et al. (2007)Curr. Prot. Cvt. 4. 6. 1-4. 6. 27.
- [0522] Kirchhofer et al. (2010)Nat. Struct. Mol. Bio. 17 :133-139.
- [0523] Knappik et al. (2000)JMB296 :57-86.
- [0524] Kobayashi et al. (1997)Biotechniques23 :500-503.
- [0525] Lefranc(2000)Curr. Prot. Imm. 1-37.
- [0526] Lutz and Patrick(2004)Curr. Opin. Biot. 15 :291-297.
- [0527] Marsh et al. (2000)Hum. Mol. Genet. 9 :13-25.
- [0528] Martineau et al. (1998)JMB280 :117-127.
- [0529] Mikawa et al. (1996)JMB262 :21-30.
- [0530] Miller et al. (2006)Nat. Meth. 3 :561-570.
- [0531] Minenkova et al. (2003)Int J Can106 :534-544.
- [0532] Parsons et. al. (2006)Biochem. 45 :2122-2128.
- [0533] Pini et al. (1998)JBC273 :21769-21776.
- [0534] Plückthun(1992)Immunol. Revs. , 130 :151-188.
- [0535] Saerens et al. (2005)JMB352 :597-607.
- [0536] Skerra et al(1993)Curr. Opinion in Immunol. 5 :256-262.
- [0537] Smith(1985)Science228 :1315-1317.
- [0538] Soderlind et al. (2000)Nat. Biotech. 18 :852-856.
- [0539] Sternberg and Hoess(1995)PNAS92 :1609-1613.

- [0540] Stewart et al. (1993) J. Exp. Med. 177 :409-418.
- [0541] Tavladoraki et al. (1999) Eur. J. Biochem. 262 :617-624.
- [0542] Tse et al. (2002) JMB 317 :85-94.
- [0543] Vaccaro et al. (2006) J. Imm. Methods. 310 :149-158.
- [0544] Visintin et al. (1999) PNAS 96 :11723-11728.
- [0545] Vitetta et al. (1993) Immunol. Today 14 :252-259.
- [0546] Viti et al. (2000) Meth. Enzy. 326 :480-505.

[0001]

序列表

<110> Affinity Biosciences Pty Ltd
Matthew, Beasley

<120> 可溶性抗体

<130> 512188

<160> 125

<170> PatentIn version 3.5

<210> 1

<211> 456

<212> DNA

<213> 人

<400> 1

```
tcttttcgcac agtaatatatc ggccgtgtcc tcggctctca ggcgtgiteat ttgcagatac      60
agcgtgtttct tggaattgtc tcctggagatg gtgaaccggc ccttcacgga gtcctgcgtag      120
tatgtgtctac caccactacc actaatagct gagaccact ccagccctt ccctggagcc      180
tggcggaccc agctcatggc atagctgcta aagggtgaatc cagaggctgc acaggagagt      240
ctcagggacc ccccaggtcg taccaagcct ccccagact ccaacagctg caccacacac      300
tggacacctg caaacaaaaa gaaacctgg tcagaaactg ccacacgtat ccactgtttc      360
tctcactctt atccaticac actcaatttt tctatttctc catgaattac cttttaaaat      420
agccacaaga aaaagccagc tcagcccaaa ctccat      456
```

<210> 2

<211> 296

<212> DNA

<213> 人

<400> 2

```
gaggtgcagc tgttggagtc tgggggaggc ttggtacagc ctggggggtc cctgagaactc      60
tcctgtgcag cctctggatt cacccttagc agctatgcc aagagctgggt ccgccaggtc      120
ccaggaaggg ggcctggagtg ggtctcagct attagtggta gtggtggtag cacatactac      180
gcagactccg tgaaggcccg gttcaccatc tccagagaca attccaagaa cacgctgtat      240
ctgcaaatga acagcctgag agccgaggac acggccgtat attactgtgc gaaaga      296
```

<210> 3

<211> 98

<212> PRT

<213> 人

<400> 3

```
Glu Val Gln Leu Leu Glu Ser 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
```

[0002]

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> 4

<211> 481

<212> DNA

<213> 人

<400> 4

atgacctgc tgcaggtaga tgggctcggc ggggctgaaa tccccccaca cagtgtcat 60

gtgtcacac tgccttaggg ctcttcac cctggatctg tgtccaggec aggcacgtgg 120

gaagatttac ttggagtcca gctctcagt ttcaagcctt ttctctcccg tttctctcc 180

tgtaggatcc gtggcctcct atgagctgac tcagccaccc tcagtgtccg tgtccccagg 240

acagacagcc agcatcacct gctctggaga taaatgggg gataaatatg ctgtctggta 300

tcagcagaag ccaggecagt cccctgtgct ggtcacttat caagatagca agcgccctc 360

agggatccct gagcgattct ctggctcaa ctctgggaac acagccactc tgaccatcag 420

cgggacccag gctatggatg aggtgacta ttactgtcag gcgtgggaca gcagcactgc 480

a 481

<210> 5

<211> 285

<212> DNA

<213> 人

<400> 5

tcctatgagc tgactcagcc acctcagtg tccgtgtccc caggacagac agccagcatc 60

acctgctctg gagataaatt gggggataaa tatgcttctt ggtatcagca gaagccagge 120

cagtcacctg tgctggtcat ctatcaagat agcaagcggc cctcagggat ccctgagcga 180

ttctcttggt ccaactctgg gaacacagcc acctgacca tcagcgggac ccaggctatg 240

gatgaggtct actattactg tcaggcgtgg gacagcagca ctgca 285

<210> 6

<211> 95

<212> PRT

<213> 人

<400> 6

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

Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly Asp Lys Tyr Ala
20 25 30

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

Gln Asp Ser Lys Arg Pro Ser Gly Ile Pro Glu Arg Phe Ser Gly Ser

[0003]

50	55	60
Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly Thr Gln Ala Met		
65	70	75 80
Asp Glu Ala Asp Tyr Tyr Cys Gln Ala Trp Asp Ser Ser Thr Ala		
	85	90 95
<210> 7		
<211> 781		
<212> DNA		
<213> 人		
<400> 7		
atggcctgga ccgtttcct cctcgccctc ctctctcact gcacaggatga tccccccagg		60
gtctcaccaa cctgccccagc ccaaggggtc tgggtccagc gtgtccttga ttctgagctc		120
aggaggccce ttctgtgtgt gggcaggatg ctcatgaccc tgctgcaggg tgggaggctg		180
gtggggctga actccccca aactgtgctc aaaggcttgt gagagcctga gggactgcac		240
ctgccaggag agagtagtga gttttcagtt caaagtctcc atacaacagg aaagtcatgg		300
gccactgggg ctggggctga ttgcagggga taccctgagg gtccacagac tctctggagc		360
ttgtctggga cagcagggca agggatttca taagaagcat ctttcacctg caagccaacc		420
tctctcttat ttattttatt atttttttat ttattttatt atttttttt atctttgcag		480
gtctctgtac ctcttatgtg ctgactcagc caccctcggt gtcagtggcc ccaggacaga		540
cggccaggat tacctgtggg gaaacaaca ttggaagtaa aagtgtgcac tggatccagc		600
agaagccagg ccaggccctt gtctgtgtcg tctatgatga tagcgaccgg cctcaggga		660
tccttgagcg attctctggc tccaactctg ggaacacggc caccctgaac atcagcaggg		720
tcgaagccgg ggaataggcc gactattact gtcagggtgt ggatagtagt agtgatcatc		780
c		781
<210> 8		
<211> 290		
<212> DNA		
<213> 人		
<400> 8		
tcctatgtgc tgactcagcc accctcggtg tcagtggccc caggacagac ggccaggatt		60
acctgtgggg gaaacaacat tggaaagtaa agtgtgcact ggtaccagca gaagccagge		120
caggcccttg tgctgtgtgt ctatgatgat agcgaccggc cctcagggat cctgagcga		180
ttctctggct ccaactctgg gaacacggcc accctgacca tcagcagggt cgaagccggg		240
gatgagggcg actattactg tcagggtgtg gatagtagta gtgatcatcc		290
<210> 9		
<211> 97		
<212> PRT		
<213> 人		
<400> 9		
Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val Ala Pro Gly Gln		
1	5	10 15
Thr Ala Arg Ile Thr Cys Gly Gly Asn Asn Ile Gly Ser Lys Ser Val		

[0004]

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

Pro

<210> 10
 <211> 478
 <212> DNA
 <213> 人

<400> 10
 atggcctggg ctccactact tctcaccctc ctgcctcact gcacaggtgg ctgcctgcaa 60
 ggaattcagg gagcgttccct ggatgtcacc tgggcigatg atctgttccct cctgcctggg 120
 aaccagtctt catctctccc cactgatctc tgtgttgctc tcttcttgca ggttcttggg 180
 ccaattttat gctgactcag cccactctg tgcggagtc tccggggaag acggtaacca 240
 tctctgcac cggcagcagt ggcagcattg ccagcaacta tgtcagtggtg taccagcagc 300
 gcccgggcag tgccccacc actgtgatct atgaggataa ccaaagacc cctggggctc 360
 ctgatcggtt ctctggctcc atcgacagct cctccaaactc tgcctccctc accatctctg 420
 gactgaagac tgaggacgag gctgactact actgtcagtc ttatgatagc agcaatca 478

<210> 11
 <211> 296
 <212> DNA
 <213> 人

<400> 11
 aattttatgc tgactcagcc ccactctgtg tcggagtctc cggggaagac ggtaaccatc 60
 tcttgaccg gcagcagtggt cagcattgcc agcaactatg tgcagtggta ccagcagcgc 120
 ccgggcagtg cccccaccac tgtgatctat gaggataacc aaagaccctc tggggctccct 180
 gatcggttct ctggctccat cgacagctcc tccaactctg cctccctcac catctctgga 240
 ctgaagactg aggacgagcc tgactactac tgcagctctt atgatagcag caatca 296

<210> 12
 <211> 98
 <212> PRT
 <213> 人

<400> 12
 Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu Ser Pro Gly Lys
 1 5 10 15
 Thr Val Thr Ile Ser Cys Thr Gly Ser Ser Gly Ser Ile Ala Ser Asn

[0005]

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

Ser Asn

<210> 13
 <211> 462
 <212> DNA
 <213> 人

<400> 13
 atgacctgct cccctctect cctcacccctt ctcatcact gcacaggtgc ccagacacag 60
 ggtcaggsga ggggtccagg aagcccatga ggccctgctt tctccttctc tctctagacc 120
 aagaatcacc gtgtctgtgt ctctctctgt tccagggtcc tgggccagct ctgtgttgac 180
 gcagccgcce tcagtgtctg cggccccagg acagaaggtc accatctcct gctctggaag 240
 cagctccaac attgggaata attatgtatc ctggtaccag cagctcccag gaacagcccc 300
 caaactcttc atttatgaca ataataagcg acctcaggg attcttgacc gattctctgg 360
 ctccaagict ggcacgtcag ccacctggg catcaccgga ctccagactg gggacgaggc 420
 cgattattac tgcggaacat gggatagcag cctgagtgtc gg 462

<210> 14
 <211> 296
 <212> DNA
 <213> 人

<400> 14
 cagtctgtgt tgaagcagcc gccctcagtg tctggggccc caggacagaa ggtcaccatc 60
 tctgtctctg gaagcagctc caacattggg aataattatg tatectggta ccagcagctc 120
 ccaggaaacag cccccaaact cctcatttat gacaataata agcgaccctc agggattcct 180
 gaccgattct ctggtccaa gtctggcagc tcagccacc tgggcatcac cgaactccag 240
 acitgggagc aggccgatta ttactcgga acatgggata gcagcctgag tgctgg 296

<210> 15
 <211> 99
 <212> PRT
 <213> 人

<400> 15
 Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Ala Ala Pro Gly Gln
 1 5 10 15

Lys Val Thr Ile Ser Cys Ser Gly Ser Ser Ser Asn Ile Gly Asn Asn

[0006]

	20	25	30
	Tyr Val Ser Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu Leu		
	35	40	45
	Ile Tyr Asp Asn Asn Lys Arg Pro Ser Gly Ile Pro Asp Arg Phe Ser		
	50	55	60
	Gly Ser Lys Ser Gly Thr Ser Ala Thr Leu Gly Ile Thr Gly Leu Gln		
	65	70	75 80
	Thr Gly Asp Glu Ala Asp Tyr Tyr Cys Gly Thr Trp Asp Ser Ser Leu		
	85	90	95
	Ser Ala Gly		
<210>	16		
<211>	466		
<212>	DNA		
<213>	人		
<400>	16		
	atggcctggt ctctctctct cctcactctc ctgcctcact gcacaggatga ctggatacag	60	
	gtccagggga ggggccttgg gaagcctatg gattcttctgt ttctctctgtt gtctctagaa	120	
	gccgaataat gatgcctgtg tctctccac ttccagggtc ctgggccag tctgtgtga	180	
	cgcagccgcc ctccagtgtt ggggccccag gccagagggt caccatctec tgcactggga	240	
	gcagctccaa catcggggca ggttaigtg tacacttgga ccagcagett ccaggaacag	300	
	ccccaaaact cctcctctat ggtaacagca atcgccctc aggggtccct gaccgattct	360	
	ctggtccaa gtctggcacc tcagctccc tggccatcac tgggtccag gctgaggatg	420	
	aggctgatta ttactgccag tcctatgaca gcagctgag tggttc	466	
<210>	17		
<211>	299		
<212>	DNA		
<213>	人		
<400>	17		
	cagtctgtgc tgacgcagcc gccctcagt tctggggccc cagggcagag ggtcaccatc	60	
	ttctgcactg ggagcagctc caacatcgga gcaggttatg atgtacactg gtaccagcag	120	
	cttcaggaa cagccccaa actctcctc tatggtaaca gcaatcgcc ctccagggtc	180	
	cctgaccgat tctctggctc caagtctggc acctcagcct ccttgccat cactgggtc	240	
	caggctgagg atgaggctga ttattactgc cagtctatg acagcagcct gactggttc	299	
<210>	18		
<211>	100		
<212>	PRT		
<213>	人		
<400>	18		
	Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Gly Ala Pro Gly Gln		
	1 5 10 15		
	Arg Val Thr Ile Ser Cys Thr Gly Ser Ser Ser Asn Ile Gly Ala Gly		

[0007]

20	25	30
Tyr Asp Val His Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu 35 40 45		
Leu Ile Tyr Gly Asn Ser Asn Arg Pro Ser Gly Val Pro Asp Arg Phe 50 55 60		
Ser Gly Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile Thr Gly Leu 65 70 75 80		
Gln Ala Glu Asp Glu Ala Asp Tyr Tyr Cys Gln Ser Tyr Asp Ser Ser 85 90 95		
Leu Ser Gly Ser 100		
<210> 19 <211> 468 <212> DNA <213> 人		
<400> 19 atggccagct tccctctcct cctcaccctc ctcactcaact gtgcaggtga caggatgggg 60 accaagaaag gggccctggg aagcccatgg ggccctgctt tctcctcttg tctccttttg 120 tctctgttca atcaccatgt ctgtgtctct ctcacttcca gggtcctggg ccagttctgt 180 gttgactcag ccaccctcag cgtctgggac ccccgggcag agggtcacca tctcttgctc 240 tggaagcagc tccaacatcg gaagtaatac tgtaaacctg taccagcagc tcccaggaac 300 ggccccaaaa ctctctatct atagtaataa tcagcggccc tcaggggtcc ctgaccgatt 360 ctcttgctcc aagtctggca cctcagcctc cctggccatc agtgggtcc agtctgagga 420 tgaggctgat tattactgtg cagcatggga tgacagcctg aatggctc 468		
<210> 20 <211> 296 <212> DNA <213> 人		
<400> 20 cagtctgtgc tgactcagcc accctcagcg tetgggaccc ccgggcagag ggtcaccatc 60 tcttggtcig gaagcagcic caacatcgga agtaatacig taaaciggtc ccagcagctc 120 ccaggaacgg cccccaaact cctcatctat agtaataatc agcggccctc aggggtccct 180 gaccgatctc ctggtccaa gtctggcacc tcagcctccc tgccatcag tgggtccag 240 tctgaggatg aggtgatta ttactgtgca gcatgggatg acagcctgaa tggctc 296		
<210> 21 <211> 99 <212> PRT <213> 人		
<400> 21 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		

[0008]

	20		25		30
Thr Val Asn Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu Leu					
35		40		45	
Ile Tyr Ser 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 Gln					
65		70		75	80
Ser Glu Asp Glu Ala Asp Tyr Tyr Cys Ala Ala Trp Asp Asp Ser Leu					
	85		90		95

Asn Gly Pro

<210> 22
 <211> 468
 <212> DNA
 <213> 人

<400> 22
 atggcgggt tccctctctt cctcaccctt ctcactcact gtgcaggta caggatgggg 60
 accaagagag gggccctggg aagcccatgg ggcctgctt tctcctcttg tctcctttcg 120
 tctcttgtca atcaccatgt ctgtgtctct ctcacttcca gggctctggg ccagatctgt 180
 gctgactcag ccaccctcag cgtctgggac ccccgggcag agggtcacca tctcttgttc 240
 tgggaagcag tccaacatcg gaagtaatta tgtatactgg taccagcagc tcccaggaac 300
 ggcccccaaa ctctctatct atagtaataa tcagcggccc tcaggggtcc ctgaccgatt 360
 ctctggctcc aagtctggca cctcagcctc cctggccatc agtgggctcc ggtccgagga 420
 tgaggctgat tattactgtg cagcatggga tgacagcctg agtgggtcc 468

<210> 23
 <211> 296
 <212> DNA
 <213> 人

<400> 23
 cagtctgtgc tgactcagcc accctcagcg tctgggaccc ccgggcagag ggtaaccatc 60
 tcttggtctg gaagcagctc caacatcgga agtaattatg tatactggta ccagcagctc 120
 ccaggaaagg cccccaaact cctcatctat agtaataatc agcggcctc aggggtccct 180
 gaccgattct ctggctccaa gtctggcacc tcagcctccc tggccatcag tgggtccgg 240
 tccgaggaag aggtgatta ttactgtgca gcatgggatg acagcctgag tggctcc 296

<210> 24
 <211> 99
 <212> PRT
 <213> 人

<400> 24
 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

[0009]

	20	25	30
Tyr Val Tyr Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu Leu	35	40	45
Ile Tyr Ser 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
Ser Glu Asp Glu Ala Asp Tyr Tyr Cys Ala Ala Trp Asp Asp Ser Leu	85	90	95
Ser Gly Pro			

<210> 25
 <211> 493
 <212> DNA
 <213> 人

<400> 25
 atggcctgga cccctctctg gctcaactctc ctcaactcttt gcatagggtgc tgccctcccag 60
 ggetcaaccc catattatca tgctagctgt gccaacctgg cctgagctt cggetcaaca 120
 cagggagtag tgtagggtgt gggactctag gcgtgaaacc cttatcctca cctcttctgt 180
 cctcttttgc aggttctgtg gtittctctg agctgactca ggaccttget gigtctgtgg 240
 ccttgggaca gacagtcagg atcacatgcc aaggagacag cctcagaagc tattatgcaa 300
 gctgttacea gcagaagcca ggacaggccc ctgtacttgt catctatggt aaaaacaacc 360
 ggccctcagg gatccagac cgattctctg gctccagctc aggaacaca gcttcttga 420
 ccatcactgg ggctcaggcg gaagatgagg ctgactatta ctgtaactec cgggacagca 480
 gtggtaacca tct 493

<210> 26
 <211> 290
 <212> DNA
 <213> 人

<400> 26
 tcttctgagc tgactcagga ccttctgtgt tctgtggcct tgggacagac agtcaggatc 60
 acatgccaaag gagacagcct cagaagctat tatgcaagct ggtaccagca gaagccagga 120
 caggcccttg tacttgatcat ctatggtaaa aacaaccggc cctcagggat ccagaccga 180
 ttctctggct ccagctcagg aaacacagct tccttgacca tcactggggc tcaggcggaa 240
 gatgaggctg actattactg taactcccgg gacagcagtg gtaaccatct 290

<210> 27
 <211> 97
 <212> PRT
 <213> 人

<400> 27

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

[0010]

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

Leu

<210> 28
 <211> 15
 <212> PRT
 <213> 人工序列

<220>
 <223> 肽接头

<400> 28

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
 1 5 10 15

<210> 29
 <211> 8
 <212> PRT
 <213> 人工序列

<220>
 <223> CDR可变序列1

<220>
 <221> VARIANT
 <222> (1)..(3)
 <223> X是除了Trp、Gln、Lys、Glu、Met之外的任意氨基酸

<220>
 <221> VARIANT
 <222> (6)..(8)
 <223> X是除了Trp、Gln、Lys、Glu、Met之外的任意氨基酸

<400> 29

Xaa Xaa Xaa Gly Gly Xaa Xaa Xaa
 1 5

<210> 30
 <211> 7
 <212> PRT
 <213> 人工序列

<220>
 <223> CDR变体序列

<220>
 <221> VARIANT

[0011]

			
<222>	(1)...(3)		
<223>	X是除了Trp、Gln、Lys、Glu、Met之外的任意氨基酸		
<220>			
<221>	VARIANT		
<222>	(5)...(7)		
<223>	X是除了Trp、Gln、Lys、Glu、Met之外的任意氨基酸		
<400>	30		
Xaa Xaa Xaa Gly Xaa Xaa Xaa			
1	5		
<210>	31		
<211>	23		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	HVK1F1		
<400>	31		
gacatccaga tgaccagtc tcc		23	
<210>	32		
<211>	23		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	HVK1 F2		
<400>	32		
gmcatccrgw tgaccagtc tcc		23	
<210>	33		
<211>	23		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	HVK2 F		
<400>	33		
gatrttgtga tgacycagwc tcc		23	
<210>	34		
<211>	23		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	HVK 3F		
<400>	34		
gaaatwgtgw tgaccagtc tcc		23	
<210>	35		
<211>	23		
<212>	DNA		
<213>	人工序列		
<220>			
<223>	GAC ATC GTG ATG ACC CAG TCT CC		
<400>	35		
gacatcgtga tgaccagtc tcc		23	
<210>	36		

[0012]

<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVK5 F	
<400>	36	
	gaaacgacac tcacgcagtc tcc	23
<210>	37	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVK6 F	
<400>	37	
	gawrttgtgm tgacwcagtc tcc	23
<210>	38	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVKCL R	
<400>	38	
	acactctccc ctggtgaagc tctt	24
<210>	39	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL1 F1	
<400>	39	
	cagtctgtgc tgactcagcc acc	23
<210>	40	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVLI F2	
<400>	40	
	cagtctgtgy tgacgcagcc gcc	23
<210>	41	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL2F	
<400>	41	
	cagtctgecc tgactcagcc t	21
<210>	42	
<211>	23	
<212>	DNA	
<213>	人工序列	

[0013]

<220>		
<223>	HVL3 F1	
<400>	42	
	tcctatgwgc tgacwcagcc acc	23
<210>	43	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL3 F2	
<400>	43	
	tcttetgagc tgactcagga ccc	23
<210>	44	
<211>	21	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL4 F1	
<400>	44	
	ctgcctgtgc tgactcagcc c	21
<210>	45	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL4 F2	
<400>	45	
	cagcytgtgc tgactcaatc ryc	23
<210>	46	
<211>	20	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL5F	
<400>	46	
	cagsetgtgc tgactcagcc	20
<210>	47	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL6 F	
<400>	47	
	aattttatgc tgactcagcc cca	23
<210>	48	
<211>	23	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL7/8F	

[0014]

<400> 48 cagrcctgtgg tgacycagga gcc	23
<210> 49 <211> 23 <212> DNA <213> 人工序列	
<220> <223> HVL9/10F	
<400> 49 cagscwggkgc tgaetcagcc acc	23
<210> 50 <211> 23 <212> DNA <213> 人工序列	
<220> <223> 01115 HVLCL R	
<400> 50 tgaacattct gtaggggcca ctg	23
<210> 51 <211> 23 <212> DNA <213> 人工序列	
<220> <223> 01116 HVLCL R2	
<400> 51 tgaacattcc gtaggggcaa ctg	23
<210> 52 <211> 42 <212> DNA <213> 人工序列	
<220> <223> HVK1 2F1	
<400> 52 atctagaatg ggagacgggtg acatccagat gacccagtct cc	42
<210> 53 <211> 42 <212> DNA <213> 人工序列	
<220> <223> HVK1 2F2	
<400> 53 atctagaatg ggagacgggtg mcatccrgwt gacccagtct cc	42
<210> 54 <211> 42 <212> DNA <213> 人工序列	
<220> <223> HVK2 2F	
<400> 54 atctagaatg ggagacgggtg mcatccrgwt gacccagtct cc	42

[0015]

<210>	55	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVK3 2F	
<400>	55	
	atctagaatg ggagacggtg aaatwgtgwt gacrcagtct cc	42
<210>	56	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVK4 2F	
<400>	56	
	atctagaatg ggagacggtg acatcgtgat gacccagtct cc	42
<210>	57	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVK5 2F	
<400>	57	
	atctagaatg ggagacggtg aaacgacact caccagtct cc	42
<210>	58	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVK6 2F	
<400>	58	
	atctagaatg ggagacggtg awrttgtgmt gacwcagtct cc	42
<210>	59	
<211>	51	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVKCL 2R	
<400>	59	
	gacgagggtc tgagacgatt trathccas yykkgtecch bsgeccraav t	51
<210>	60	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL1 2F1	
<400>	60	
	atctagaatg ggagacggtc agtctgtgct gactcagcca cc	42
<210>	61	

[0016]

<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL1 2F2	
<400>	61	
atctagaatg ggagacggtc agtctgigyt gacgcagccg cc		42
<210>	62	
<211>	40	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL2 2F	
<400>	62	
atctagaatg ggagacggtc agtctgccct gactcagcct		40
<210>	63	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL3 2F1	
<400>	63	
atctagaatg ggagacggtt cctatgwget gacwcageca cc		42
<210>	64	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL3 2F2	
<400>	64	
atctagaatg ggagacggtt cttctgagct gactcaggac cc		42
<210>	65	
<211>	40	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL4 2F1	
<400>	65	
atctagaatg ggagacggtc tgcctgigct gactcagccc		40
<210>	66	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL4 2F2	
<400>	66	
atctagaatg ggagacggtc agcytgtgct gactcaater yc		42
<210>	67	
<211>	39	
<212>	DNA	
<213>	人工序列	

[0017]

<220>		
<223>	HVL5 2F	
<400>	67	
	atctagaatg ggagacggtc agsetgtgct gactcagcc	39
<210>	68	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL6 2F	
<400>	68	
	atctagaatg ggagacggta attttatgct gactcagccc ca	42
<210>	69	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL7/8 2F	
<400>	69	
	atctagaatg ggagacggtc agrctgtggt gacycaggag cc	42
<210>	70	
<211>	42	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVL9/10 2F	
<400>	70	
	atctagaatg ggagacggtc agscwkgct gactcagcca cc	42
<210>	71	
<211>	48	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	HVLCL 2R	
<400>	71	
	gactcagggtc tgagacgarr ygrtsasetb sgtbechbyd ccraabac	48
<210>	72	
<211>	11	
<212>	PRT	
<213>	人	
<400>	72	
	Val Phe Gly Thr Gly Thr Lys Val Thr Val Ser	
	1 5 10	
<210>	73	
<211>	11	
<212>	PRT	
<213>	人	
<400>	73	
	Val Phe Gly Gly Gly Thr Lys Leu Thr Val Ser	

[0018]

1 5 10

<210> 74
<211> 11
<212> PRT
<213> 人

<400> 74

Val Phe Gly Gly Gly Thr Lys Leu Thr Val Ser
1 5 10

<210> 75
<211> 11
<212> PRT
<213> 人

<400> 75

Val Phe Gly Gly Gly Thr Gln Leu Ile Ile Ser
1 5 10

<210> 76
<211> 11
<212> PRT
<213> 人

<400> 76

Val Phe Gly Glu Gly Thr Glu Leu Thr Val Ser
1 5 10

<210> 77
<211> 11
<212> PRT
<213> 人

<400> 77

Val Phe Gly Ser Gly Thr Lys Val Thr Val Ser
1 5 10

<210> 78
<211> 11
<212> PRT
<213> 人

<400> 78

Val Phe Gly Gly Gly Thr Gln Leu Thr Ala Ser
1 5 10

<210> 79
<211> 11
<212> PRT
<213> 人

<400> 79

Val Phe Gly Thr Gly Thr Lys Leu Ile Ile Ser
1 5 10

<210> 80
<211> 105
<212> DNA
<213> 人工序列

<220>

[0019]

<223> PCR引物

 <220>
 <221> misc_feature
 <222> (56)..(57)
 <223> n是a、c、g或t

 <220>
 <221> misc_feature
 <222> (59)..(60)
 <223> n是a、c、g或t

 <220>
 <221> misc_feature
 <222> (62)..(63)
 <223> n是a、c、g或t

 <220>
 <221> misc_feature
 <222> (71)..(72)
 <223> n是a、c、g或t

 <220>
 <221> misc_feature
 <222> (74)..(75)
 <223> n是a、c、g或t

 <220>
 <221> misc_feature
 <222> (77)..(78)
 <223> n是a、c、g或t

 <400> 80
 gatcagggtc tgagacgaga ccgtcacttt cgtaccggtg ccgaacacca cagtannann 60
 amtccgccca nnannamgt cccacgcctg acagtaatag tcage 105

 <210> 81
 <211> 31
 <212> PRT
 <213> 人

 <220>
 <221> VARIANT
 <222> (10)..(12)
 <223> X是除了Trp、Gln、Lys、Glu、Met之外的任意氨基酸

 <220>
 <221> VARIANT
 <222> (15)..(17)
 <223> X是除了Trp、Gln、Lys、Glu、Met之外的任意氨基酸

 <400> 81
 Ala Asp Tyr Tyr Cys Gln Ala Trp Asp Xaa Xaa Xaa Gly Gly Xaa Xaa
 1 5 10 15
 Xaa Thr Val Val Phe Gly Thr Gly Thr Lys Val Thr Val Ser Ser
 20 25 30

 <210> 82
 <211> 105
 <212> DNA
 <213> 人

 <220>
 <221> misc_feature
 <222> (65)..(66)
 <223> n是a、c、g或t

[0020]

```

<220>
<221> misc_feature
<222> (68)..(69)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (71)..(72)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (77)..(78)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (80)..(81)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (83)..(84)
<223> n是a、c、g或t

<400> 82
gatcagggtc tgagaccgcg tgcacacgtt aaccatggta ccttgacccc aaatatcaaa      60
cgccannanna nngccannan nanntttgcg acagtagtaa acagc                                105

<210> 83
<211> 25
<212> PRT
<213> 人

<220>
<221> VARIANT
<222> (7)..(9)
<223> X是除了Trp、Gln、Lys、Glu、Met之外的任意氨基酸

<220>
<221> VARIANT
<222> (11)..(13)
<223> X是除了Trp、Gln、Lys、Glu、Met之外的任意氨基酸

<400> 83
Val Tyr Tyr Cys Ala Lys Xaa Xaa Xaa Gly Xaa Xaa Xaa Ala Phe Asp
1          5          10          15

Ile Trp Gly Gln Gly Thr Met Val Thr
          20          25

<210> 84
<211> 771
<212> DNA
<213> 人

<220>
<221> misc_feature
<222> (286)..(287)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (289)..(290)
<223> n是a、c、g或t

<220>

```

[0021]

```

<221> misc_feature
<222> (292)..(293)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (301)..(302)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (304)..(305)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (307)..(308)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (706)..(707)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (709)..(710)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (712)..(713)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (718)..(719)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (721)..(722)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (724)..(725)
<223> n是a、c、g或t

<400> 84
atgggagacg gtcagtctgt gctgactcag ccaccctcag tgtccgtgtc cccaggacag 60
acagccagca tcacctgtctc tggagataaa ttgggggata aatatgett gctgtatcag 120
cagaagccag gccagtcctc tgtgctggtc atctatcaag atagcaagcg gccctcaggg 180
atccctgagc gattctcttg ctccaactct gggaacacag ccactctgac catcagcggg 240
accaggcta tggatgaggc tgactattac tgcaggcgt gggacnntnn tnntggaggt 300
nntnntnnta ctgtggtgtt cggcacgggc accaagctca tcatttcgtc tcagaccggt 360
ggttctgggt gtgggtggtt tggcggcggc ggtccgggtg gtggtggatc cgaagtecaa 420
ctgctggagt ccggcgggtg cctgggtcag ccagggtgca gccctgcct gagctgcgcc 480
gcatccggtt ttacttcag cagctacgcg atgtcgiggg tgcgccaggc accgggcaag 540
ggcctggagt gggtcagcgc catcagcggg agcggcggtt ctacgtatta tgcggacagc 600
gtcaagggcc gtttcacat cagccgtgac aattccaaaa acaccctgta cttgcagatg 660
aacagettgc gtgcggaaga tacggtgtt tactactgtg cgaaamntnn tnntggamti 720
nntnntgect ttgatatttg ggtcaaggt accatgggta ccgtgagcag c 771

```

[0022]

```

<210> 85
<211> 257
<212> PRT
<213> 人

<220>
<221> misc_feature
<222> (96)..(98)
<223> Xaa可以是任意天然氨基酸

<220>
<221> misc_feature
<222> (101)..(103)
<223> Xaa可以是任意天然氨基酸

<220>
<221> misc_feature
<222> (236)..(238)
<223> Xaa可以是任意天然氨基酸

<220>
<221> misc_feature
<222> (240)..(242)
<223> Xaa可以是任意天然氨基酸

<400> 85
Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val
1          5          10          15

Ser Pro Gly Gln Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly
20          25          30

Asp Lys Tyr Ala Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val
35          40          45

Leu Val Ile Tyr Gln Asp Ser Lys Arg Pro Ser Gly Ile Pro Glu Arg
50          55          60

Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly
65          70          75          80

Thr Gln Ala Met Asp Glu Ala Asp Tyr Tyr Cys Gln Ala Trp Asp Xaa
85          90          95

Xaa Xaa Gly Gly Xaa Xaa Xaa Thr Val Val Phe Gly Thr Gly Thr Lys
100         105         110

Val Thr Val Ser Ser Gln Thr Gly Gly Ser Gly Gly Gly Ser Gly
115         120         125

Gly Gly Gly Ser Gly Gly Gly Gly Ser Glu Val Gln Leu Leu Glu Ser
130         135         140

Gly Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala
145         150         155         160

Ala Ser Gly Phe Thr Phe Ser Ser Tyr Ala Met Ser Trp Val Arg Gln
165         170         175

Ala Pro Gly Lys Gly Leu Glu Trp Val Ser Ala Ile Ser Gly Ser Gly

```

[0023]

	180	185	190
	Gly Ser Thr Tyr Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser		
	195	200	205
	Arg Asp Asn Ser Lys Asn Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg		
	210	215	220
	Ala Glu Asp Thr Ala Val Tyr Tyr Cys Ala Lys Xaa Xaa Xaa Gly Xaa		
	225	230	235 240
	Xaa Xaa Ala Phe Asp Ile Trp Gly Gln Gly Thr Met Val Thr Val Ser		
	245	250	255
	Ser		
<210>	86		
<211>	8		
<212>	PRT		
<213>	人		
<220>			
<221>	misc_feature		
<222>	(1)..(3)		
<223>	Xaa可以是任意天然氨基酸		
<220>			
<221>	misc_feature		
<222>	(6)..(8)		
<223>	Xaa可以是任意天然氨基酸		
<400>	86		
	Xaa Xaa Xaa Gly Gly Xaa Xaa Xaa		
	1 5		
<210>	87		
<211>	7		
<212>	PRT		
<213>	人		
<220>			
<221>	misc_feature		
<222>	(1)..(3)		
<223>	Xaa可以是任意天然氨基酸		
<220>			
<221>	misc_feature		
<222>	(5)..(7)		
<223>	Xaa可以是任意天然氨基酸		
<400>	87		
	Xaa Xaa Xaa Gly Xaa Xaa Xaa		
	1 5		
<210>	88		
<211>	361		
<212>	DNA		
<213>	人		
<400>	88		
	atgggagacg gtcagtctgt gctgactcag ccaccctcag tgtccgtgtc ccaggacag		60

[0024]

acagccagca tcacctgctc tggagataaa ttgggggata aatatgcttg ctggtatcag	120
cagaagccag gccagtcgcc tgtgtgtggtc atctatcaag atagcaagcg gccctcaggg	180
atccctgagc gattctctgg ctccaactct gggaacacag ccactctgac catcagcggg	240
acccaggeta tggatgagcg tgactattac tgtcaggcgt gggactgaga cctagacggt	300
ctctgcgtti gatatttggg gtcaaggtae catggttacc gtgagcagct cgtctcagac	360
c	361
<210> 89	
<211> 396	
<212> DNA	
<213> 人	
<400> 89	
actgtggtgt tcggcaccgg tacgaaagtg actgtctcat ctacagccgg tggttctggt	60
gggtgtggtt ctggcggcgg cggctccggt ggtggtggat ccgaagtcga actgtctggag	120
tccggcgggt gcctgggtgca gccaggtggc agcctgcgcc tgagctgcgc cgcattccgt	180
tttactttca gcagctacgc gatgtcgtgg gtgcgccagg caccgggcaa gggectggag	240
tgggtcagcg ccattcagcg tagcggcgggt tctacgtatt atgcggacag cgtcaagggc	300
cgtttcacca tcagccgtga caattccaaa aacacctgt acttcagat gaacagcttg	360
cgtgcggaag atacggctgt ttactactgt gcgaaa	396
<210> 90	
<211> 71	
<212> DNA	
<213> 人	
<220>	
<221> misc_feature	
<222> (18)..(19)	
<223> n是a、c、g或t	
<220>	
<221> misc_feature	
<222> (21)..(22)	
<223> n是a、c、g或t	
<220>	
<221> misc_feature	
<222> (24)..(25)	
<223> n是a、c、g或t	
<220>	
<221> misc_feature	
<222> (33)..(34)	
<223> n是a、c、g或t	
<220>	
<221> misc_feature	
<222> (36)..(37)	
<223> n是a、c、g或t	
<220>	
<221> misc_feature	
<222> (39)..(40)	
<223> n是a、c、g或t	
<400> 90	
gatacgggtc tcaggacnnt nntnntggcg ganntnntnn tactgtggtg ttgggcacgg	60
gtacgaaagt g	71

[0025]

<210> 91
<211> 68
<212> DNA
<213> 人

<220>
<221> misc_feature
<222> (19)..(20)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (22)..(23)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (25)..(26)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (31)..(32)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (34)..(35)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (37)..(38)
<223> n是a、c、g或t

<400> 91
gatacagggtc tcaacgcann annanngcca nmannanttt tcgcacagta gtaaacagcc 60
gtatcttc 68

<210> 92
<211> 8
<212> PRT
<213> 人

<400> 92
Pro Phe Gly Gly Gly Tyr Val
1 5

<210> 93
<211> 7
<212> PRT
<213> 人

<400> 93
Pro Pro His Gly Ala Pro Ala
1 5

<210> 94
<211> 8
<212> PRT
<213> 人

<400> 94
Leu Cys Ile Gly Gly Val Ala Ser
1 5

[0026]

<210> 95
<211> 7
<212> PRT
<213> 人

<400> 95

His Asn Ser Gly Asn Asn Phe
1 5

<210> 96
<211> 8
<212> PRT
<213> 人

<400> 96

Phe Val Ser Gly Gly Ile Ser Thr
1 5

<210> 97
<211> 7
<212> PRT
<213> 人

<400> 97

Phe Asn Phe Gly Asn Ala Tyr
1 5

<210> 98
<211> 8
<212> PRT
<213> 人

<400> 98

Ile Asn Ser Gly Gly Ala Ser Phe
1 5

<210> 99
<211> 7
<212> PRT
<213> 人

<220>
<221> misc_feature
<222> (1)..(3)
<223> Xaa可以是任意天然氨基酸

<400> 99

Xaa Xaa Xaa Gly Thr Asn Tyr
1 5

<210> 100
<211> 8
<212> PRT
<213> 人

<400> 100

Ser Arg Ala Gly Gly Cys Asn Gly
1 5

<210> 101
<211> 7

[0027]

<212> PRT

<213> 人

<400> 101

Phe Asp Tyr Gly His Cys Ile
1 5

<210> 102

<211> 8

<212> PRT

<213> 人

<400> 102

Thr Asn Arg Gly Gly Val Cys Ala
1 5

<210> 103

<211> 7

<212> PRT

<213> 人

<400> 103

Thr Ala Ala Gly Val Pro Phe
1 5

<210> 104

<211> 8

<212> PRT

<213> 人

<400> 104

Phe Ser Thr Gly Gly Cys Ala Phe
1 5

<210> 105

<211> 7

<212> PRT

<213> 人

<400> 105

Ala Ile Cys Gly Ala Thr Ala
1 5

<210> 106

<211> 8

<212> PRT

<213> 人

<220>

<221> misc_feature

<222> (2)..(2)

<223> Xaa可以是任意天然氨基酸

<400> 106

Phe Xaa Gly Gly Gly Asp Gly Thr
1 5

<210> 107

<211> 7

<212> PRT

<213> 人

[0028]

<400> 107

Pro Tyr Arg Gly Ser Phe Phe
1 5

<210> 108

<211> 8

<212> PRT

<213> 人

<400> 108

Ile Ile Pro Gly Gly Leu Tyr Ala
1 5

<210> 109

<211> 7

<212> PRT

<213> 人

<400> 109

Pro Val Ile Gly Ser Asn Thr
1 5

<210> 110

<211> 254

<212> PRT

<213> 维多利亚水母

<400> 110

Met Val Ser Val Ile Lys Pro Glu Met Lys Ile Lys Leu Cys Met Arg
1 5 10 15Gly Thr Val Asn Gly His Asn Phe Val Ile Glu Gly Glu Gly Lys Gly
20 25 30Asn Pro Tyr Glu Gly Thr Gln Ile Leu Asp Leu Asn Val Thr Glu Gly
35 40 45Ala Pro Leu Pro Phe Ala Tyr Asp Ile Leu Thr Thr Val Phe Gln Tyr
50 55 60Gly Asn Arg Ala Phe Thr Lys Tyr Pro Ala Asp Ile Gln Asp Tyr Phe
65 70 75 80Lys Gln Thr Phe Pro Glu Gly Tyr His Trp Glu Arg Ser Met Thr Tyr
85 90 95Glu Asp Gln Gly Ile Cys Thr Ala Thr Ser Asn Ile Ser Met Arg Gly
100 105 110Asp Cys Phe Phe Tyr Asp Ile Arg Phe Asp Gly Thr Asn Phe Pro Pro
115 120 125Asn Gly Pro Val Met Gln Lys Lys Thr Leu Lys Trp Glu Pro Ser Thr
130 135 140Glu Lys Met Tyr Val Glu Asp Gly Val Leu Lys Gly Asp Val Asn Met
145 150 155 160

[0029]

Arg Leu Leu Leu Glu Gly Gly Gly His Tyr Arg Cys Asp Phe Lys Thr
 165 170 175
 Thr Tyr Lys Ala Lys Lys Glu Val Arg Leu Pro Asp Ala His Lys Ile
 180 185 190
 Asp His Arg Ile Glu Ile Leu Lys His Asp Lys Asp Tyr Asn Lys Val
 195 200 205
 Lys Leu Tyr Glu Asn Ala Val Ala Arg Tyr Ser Met Leu Pro Ser Gln
 210 215 220
 Ala Lys Ser Gly Gly Leu Asn Asp Ile Phe Glu Ala Gln Lys Ile Glu
 225 230 235 240
 Trp His Glu Asp Thr Gly Gly Ser His His His His His His
 245 250
 <210> 111
 <211> 266
 <212> PRT
 <213> 人
 <400> 111
 Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val
 1 5 10 15
 Ser Pro Gly Gln Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly
 20 25 30
 Asp Lys Tyr Ala Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val
 35 40 45
 Leu Val Ile Tyr Gln Asp Ser Lys Arg Pro Ser Gly Ile Pro Glu Arg
 50 55 60
 Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly
 65 70 75 80
 Thr Gln Ala Met Asp Glu Ala Asp Tyr Tyr Cys Gln Ala Trp Asp Phe
 85 90 95
 Asn Leu Gly Gly Cys Gly Asp Thr Val Val Phe Gly Thr Gly Thr Lys
 100 105 110
 Val Thr Val Ser Ser Gln Thr Gly Gly Ser Gly Gly Gly Ser Gly
 115 120 125
 Gly Gly Gly Ser Gly Gly Gly Gly Ser Glu Val Gln Leu Leu Glu Ser
 130 135 140
 Gly Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala
 145 150 155 160
 Ala Ser Gly Phe Thr Phe Ser Ser Tyr Ala Met Ser Trp Val Arg Gln
 165 170 175

[0030]

Ala Pro Gly Lys Gly Leu Glu Trp Val Ser Ala Ile Ser Gly Ser Gly
 180 185 190

Gly Ser Thr Tyr Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser
 195 200 205

Arg Asp Asn Ser Lys Asn Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg
 210 215 220

Ala Glu Asp Thr Ala Val Tyr Tyr Cys Ala Lys His Ile Asp Gly Pro
 225 230 235 240

Val Ala Ala Phe Asp Ile Trp Gly Gln Gly Thr Met Val Thr Val Ser
 245 250 255

Ser Ser Ser Gln Thr Ser Ile Leu Val Ala
 260 265

<210> 112
 <211> 1176
 <212> DNA
 <213> 人工序列

<220>
 <223> 融合构建体

<400> 112
 atgacgagca aagaaacctt taccattac cagccgcagg gcaacagtga cccggctcat 60
 accgcaaccg cgcccgccgg attgagtgcg aaagcgccctg caatgacccc gctgatgctg 120
 gacacctcca gccgttaagct ggttgcgtgg gatggcacca ccgacgggtgc tgccgttggc 180
 attcttgcgg ttgctgctga ccagaccagc accacgctga cgttctacaa gtccggcacg 240
 ttccgttatg aggatgtgct ctggccggag gctgccagcg acgagacgaa aaaaeggacc 300
 gcgtttgccg gaacggcaat cagcatcggt ggaggtagcg gcggatcgga tgacgacgat 360
 aagtctagaa atggcggaga cggtcagttc gtgctgactc agccaccctc agtgtccgtg 420
 tccccaggac agacagccag catcacctgc tctggagata aattggggga taaatatgct 480
 tgctgtatc agcagaagcc aggccagtcc cctgtgctgg tcattctatca agatagcaag 540
 cggccctcag ggatccctga gcgattctct ggctccaact ctgggaacac agccactctg 600
 accatcagcg ggaccaggc tatggatgag gctgactatt actgtcagcg gtgggacttt 660
 aatcttggcg gatgtggtga tactgtggtg ttccggcaccg gtacgaaagt gactgtctca 720
 tctcagaccg gtggttctgg tggcgggtgt tctggcggcg gcggctccgg tgggtggtgga 780
 tccgaagtcc aactgctgga gtccgggtgt ggcttggtc agccaggtgg cagcctgcgc 840
 ctgagctcgc ccgcatccgg ttctacttct agcagctacg cgatgtcgtg ggtgcgccag 900
 gcaccgggca agggcctgga gtgggtcagc gccatcagcg gtagcggcgg ttctacgtat 960
 tatgcggaca gcgtcaaggc ccgtttcacc atcagccgtg acaattccaa aaacaccctg 1020
 tacttgacga tgaacagctt gcgtgcggaa gatacggctg ttactactg tgcgaaacat 1080
 attgatggcc ctgttgcctc gtttgataat tggggtaag gtaccatggt taccgtgagc 1140
 aactcagcgc attacaagga cgatgatgac aaataa 1176

<210> 113

[0031]

```

<211> 391
<212> PRT
<213> 人工序列

<220>
<223> 融合蛋白

<400> 113
Met Thr Ser Lys Glu Thr Phe Thr His Tyr Gln Pro Gln Gly Asn Ser
1          5          10          15

Asp Pro Ala His Thr Ala Thr Ala Pro Gly Gly Leu Ser Ala Lys Ala
20          25          30

Pro Ala Met Thr Pro Leu Met Leu Asp Thr Ser Ser Arg Lys Leu Val
35          40          45

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

Ala Ala Asp Gln Thr Ser Thr Thr Leu Thr Phe Tyr Lys Ser Gly Thr
65          70          75          80

Phe Arg Tyr Glu Asp Val Leu Trp Pro Glu Ala Ala Ser Asp Glu Thr
85          90          95

Lys Lys Arg Thr Ala Phe Ala Gly Thr Ala Ile Ser Ile Val Gly Gly
100         105         110

Ser Gly Gly Ser Asp Asp Asp Asp Lys Ser Arg Asn Gly Gly Asp Gly
115         120         125

Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val Ser Pro Gly Gln
130         135         140

Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly Asp Lys Tyr Ala
145         150         155         160

Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val Leu Val Ile Tyr
165         170         175

Gln Asp Ser Lys Arg Pro Ser Gly Ile Pro Glu Arg Phe Ser Gly Ser
180         185         190

Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly Thr Gln Ala Met
195         200         205

Asp Glu Ala Asp Tyr Tyr Cys Gln Ala Trp Asp Phe Asn Leu Gly Gly
210         215         220

Cys Gly Asp Thr Val Val Phe Gly Thr Gly Thr Lys Val Thr Val Ser
225         230         235         240

Ser Gln Thr Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
245         250         255

Gly Gly Gly Gly Ser Glu Val Gln Leu Leu Glu Ser Gly Gly Gly Leu
260         265         270

```

[0032]

Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe
 275 280 285
 Thr Phe Ser Ser Tyr Ala Met Ser Trp Val Arg Gln Ala Pro Gly Lys
 290 295 300
 Gly Leu Glu Trp Val Ser Ala Ile Ser Gly Ser Gly Gly Ser Thr Tyr
 305 310 315 320
 Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser
 325 330 335
 Lys Asn Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr
 340 345 350
 Ala Val Tyr Tyr Cys Ala Lys His Ile Asp Gly Pro Val Ala Ala Phe
 355 360 365
 Asp Ile Trp Gly Gln Gly Thr Met Val Thr Val Ser Asn Ser Ser Asp
 370 375 380
 Tyr Lys Asp Asp Asp Asp Lys
 385 390
 <210> 114
 <211> 89
 <212> PRT
 <213> 人
 <400> 114
 Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val Ser Pro Gly Gln
 1 5 10 15
 Thr Ala Ser Ile Thr Cys Ser Gly Asp Lys Leu Gly Asp Lys Tyr Ala
 20 25 30
 Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val Leu Val Ile Tyr
 35 40 45
 Gln Asp Ser Lys Arg Pro Ser Gly Ile Pro Glu Arg Phe Ser Gly Ser
 50 55 60
 Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly Thr Gln Ala Met
 65 70 75 80
 Asp Glu Ala Asp Tyr Tyr Cys Gln Ala
 85
 <210> 115
 <211> 93
 <212> PRT
 <213> 人
 <400> 115
 Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val
 1 5 10 15

[0033]

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

<210> 116
 <211> 93
 <212> PRT
 <213> 人

<400> 116

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

<210> 117
 <211> 93
 <212> PRT
 <213> 人

<400> 117

Met Gly Asp Gly Gln Ser Val Leu Thr Gln Pro Pro Ser Val Ser Val
 1 5 10 15
 Ser Pro Gly Gln Thr Ala Thr Ile Thr Cys Ser Gly Asp Lys Leu Gly
 20 25 30
 Asp Lys Phe Ala Cys Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Val
 35 40 45
 Leu Val Ile Tyr Gln Asp Tyr Gln Arg Pro Ser Gly Ile Pro Glu Arg
 50 55 60

[0034]

Phe Ser Gly Ser Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Gly
65 70 75 80

Thr Gln Val Met Asp Glu Ala Asp Tyr Tyr Cys Gln Ala
85 90

<210> 118
<211> 87
<212> PRT
<213> 人

<220>
<221> misc_feature
<222> (41)..(41)
<223> Xaa可以是任意天然氨基酸

<220>
<221> misc_feature
<222> (61)..(61)
<223> Xaa可以是任意天然氨基酸

<400> 118

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

Ala Pro Glu Lys Thr Ala Thr Ile Ala Cys Gly Gly Asn Arg Ile Gly
20 25 30

Ser Lys Ser Val His Trp Tyr Gln Xaa Lys Pro Gly Gln Ala Pro Val
35 40 45

Leu Val Ile Tyr Asn Asp Asn Asp Arg Pro Ser Gly Xaa Pro Glu Arg
50 55 60

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

Val Glu Val Gly Asp Glu Ala
85

<210> 119
<211> 87
<212> PRT
<213> 人

<400> 119

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

Ala Pro Gly Lys Thr Ala Thr Ile Ala Cys Gly Gly Asn Arg Ile Gly
20 25 30

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

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

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

[0035]

Val Glu Val Gly Asp Glu Ala
85

<210> 120
<211> 83
<212> PRT
<213> 人

<400> 120

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

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

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

Tyr Asp Ser Asp Arg Pro Ser Gly Ile Pro Glu Arg Phe Ser Gly Ser
50 55 60

Asn Ser Gly Asn Thr Ala Thr Leu Thr Ile Ser Arg Val Glu Ala Gly
65 70 75 80

Asp Glu Ala

<210> 121
<211> 87
<212> PRT
<213> 人

<400> 121

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

Ala Pro Gly Lys Thr Ala Thr Ile Thr Cys Gly Gly Asn Asn Ile Glu
20 25 30

Asp Lys Ser Val His Trp Tyr Gln Gln Lys Pro Gly Gln Ala Pro Val
35 40 45

Leu Val Ile Tyr Tyr Asp Thr Asp Arg Pro Ser Gly Ile Pro Glu Arg
50 55 60

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

Val Glu Ala Gly Asp Glu Ala
85

<210> 122
<211> 95
<212> PRT
<213> 人

<400> 122

[0036]

Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu Ser Pro Gly Lys
 1 5 10 15
 Thr Val Thr Ile Ser Cys Thr Gly Ser Ser Gly Ser Ile Ala Ser Asn
 20 25 30
 Tyr Val Gln Trp Tyr Gln Gln Arg Pro Gly Ser Ala Pro Thr Thr Val
 35 40 45
 Ile Tyr Glu Asp Asn Gln Arg Pro Ser Gly Val Pro Asp Arg Phe Ser
 50 55 60
 Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu Thr Ile Ser Gly
 65 70 75 80
 Leu Lys Thr Glu Asp Glu Ala Asp Tyr Tyr Cys Gln Ser Tyr Asp
 85 90 95
 <210> 123
 <211> 99
 <212> PRT
 <213> 人
 <400> 123
 Met Gly Asp Gly Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu
 1 5 10 15
 Ser Pro Gly Lys Thr Val Thr Ile Pro Cys Thr Gly Ser Ser Gly Ser
 20 25 30
 Ile Ala Ser Asn Tyr Val Gln Trp Tyr Gln Gln Arg Pro Gly Ser Ala
 35 40 45
 Pro Thr Thr Val Ile Tyr Glu Asp Lys Gln Lys Pro Ser Gly Val Pro
 50 55 60
 Asp Arg Phe Ser Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu
 65 70 75 80
 Thr Ile Ser Gly Leu Lys Thr Glu Asp Glu Ala Asp Tyr Tyr Cys Gln
 85 90 95
 Ser Tyr Asp
 <210> 124
 <211> 99
 <212> PRT
 <213> 人
 <400> 124
 Met Gly Asp Gly Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu
 1 5 10 15
 Ser Pro Gly Lys Thr Val Thr Ile Ser Cys Thr Gly Ser Ser Gly Ser
 20 25 30
 Ile Ala Ser Asn Ser Val Gln Arg Tyr Gln Gln Arg Pro Gly Ser Ala

[0037]

35	40	45	
Pro Thr Thr Val Ile Tyr Glu Asp Asn Arg Arg Pro Ser Gly Val Pro			
50	55	60	
Asp Arg Phe Ser Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu			
65	70	75	80
Ile Ile Ser Gly Leu Gln Thr Glu Asp Glu Ala Asp Tyr Tyr Cys Gln			
	85	90	95
Ser Tyr Asp			
<210> 125			
<211> 99			
<212> PRT			
<213> 人			
<400> 125			
Met Gly Asp Gly Asn Phe Met Leu Thr Gln Pro His Ser Val Ser Glu			
1	5	10	15
Ser Pro Gly Lys Thr Val Thr Ile Ser Cys Thr Gly Arg Ser Gly Ser			
	20	25	30
Ile Ala Asp Asn Tyr Val Gln Trp Tyr Gln Gln Arg Pro Gly Ser Val			
	35	40	45
Pro Thr Ser Val Ile Phe Glu Asp Asn Gln Arg Pro Ser Gly Val Pro			
50	55	60	
Glu Arg Phe Ser Gly Ser Ile Asp Ser Ser Ser Asn Ser Ala Ser Leu			
65	70	75	80
Thr Ile Ser Gly Leu Met Thr Glu Asp Glu Ala Asp Tyr His Cys Gln			
	85	90	95
Ser Tyr Asp			

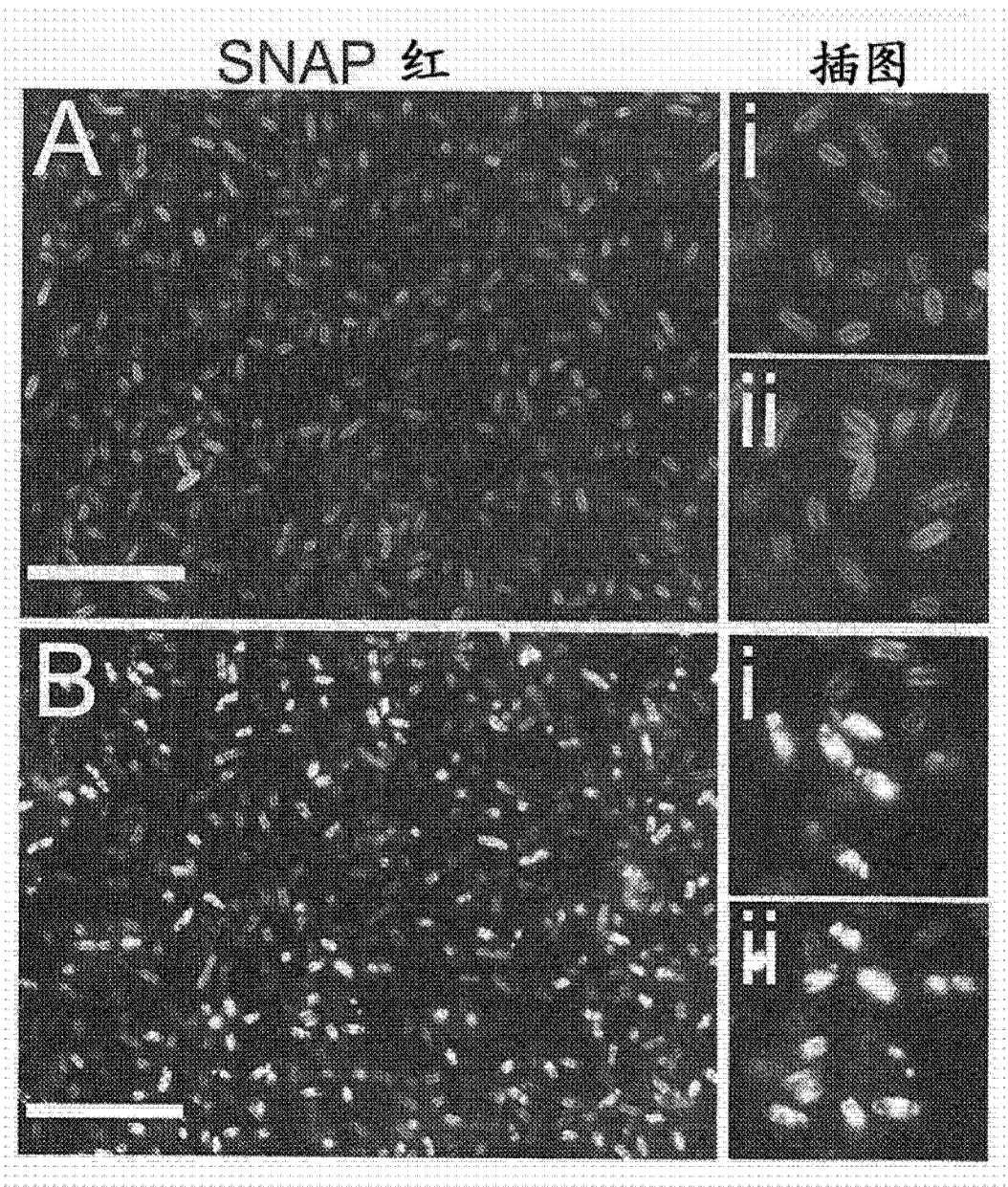


图 1

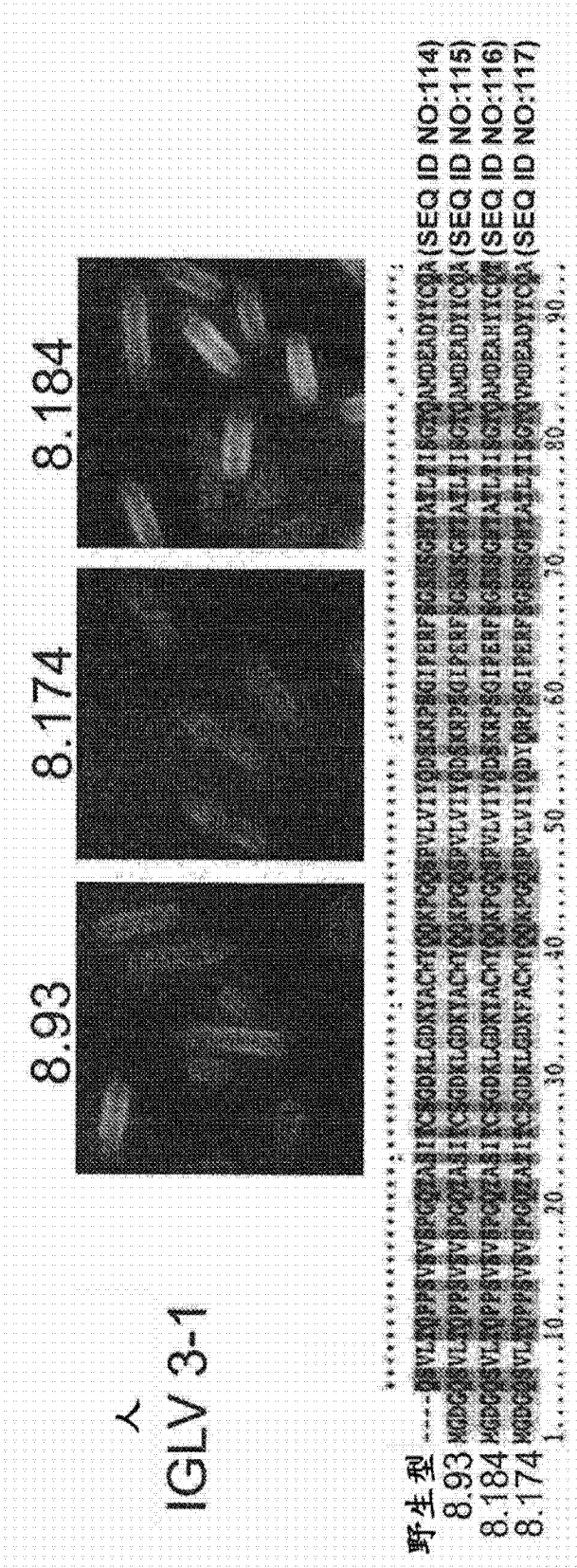


图 2

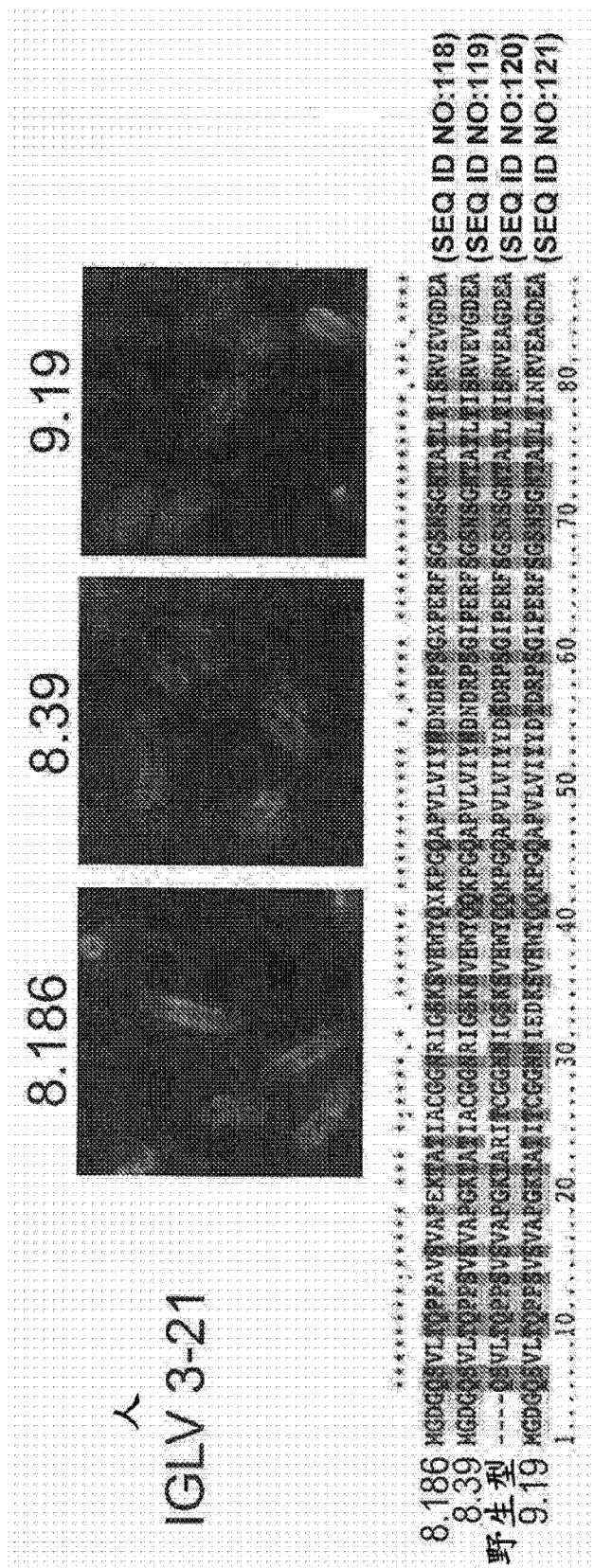


图 2(续)

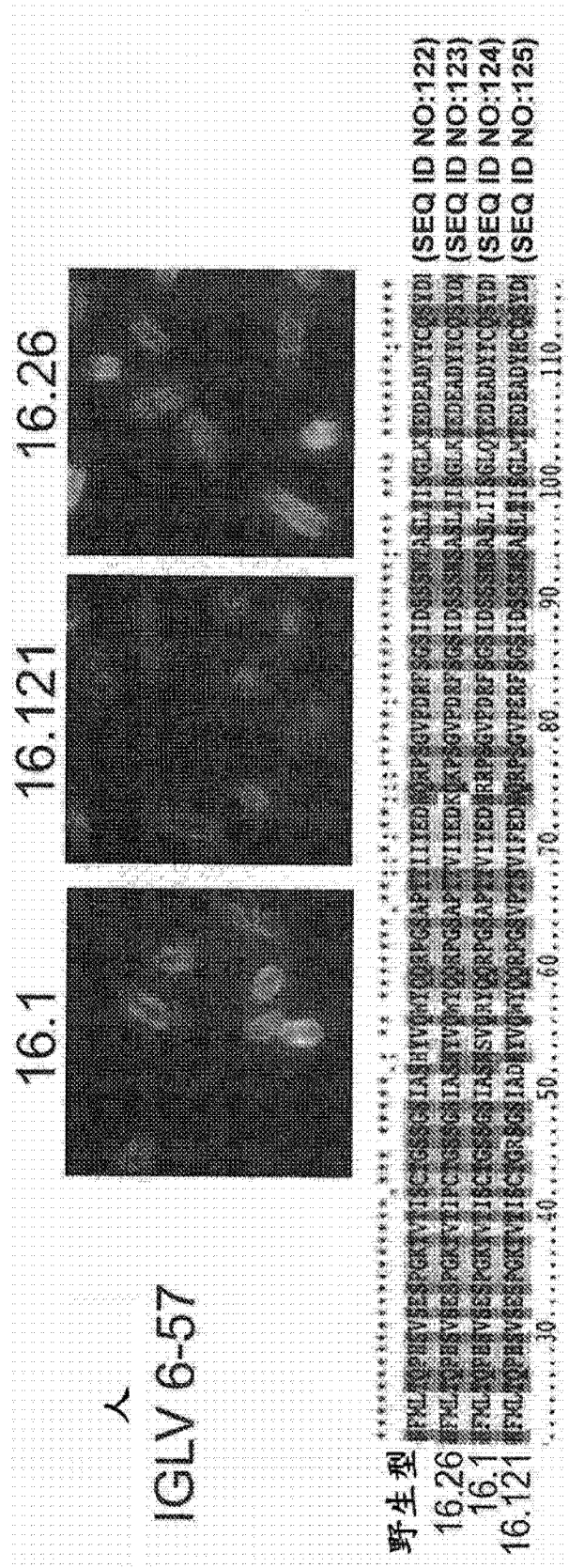


图2(续)

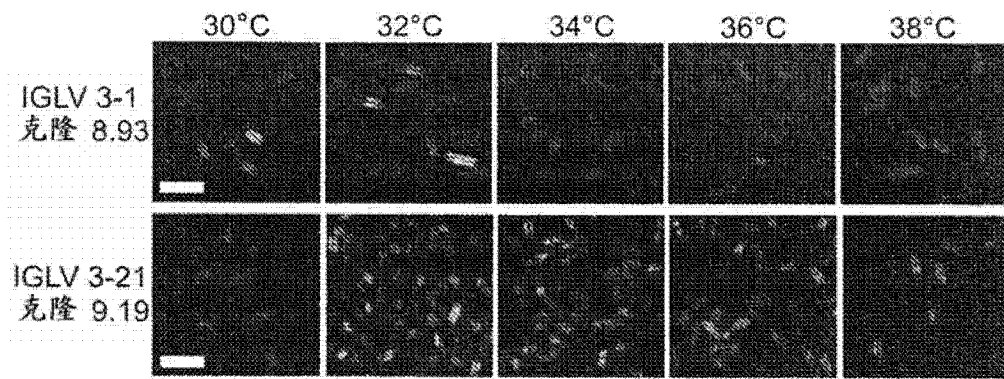


图 3

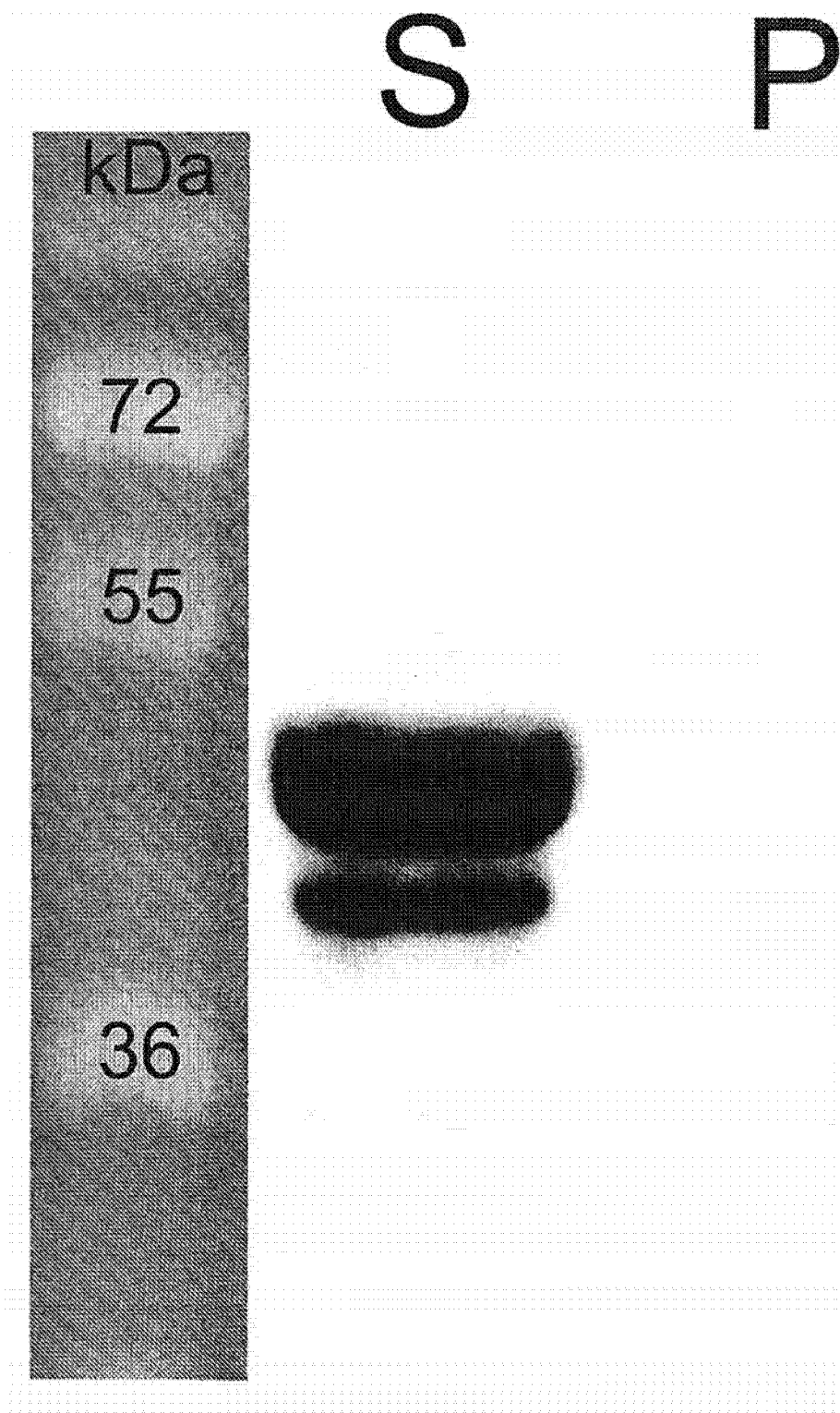


图 4

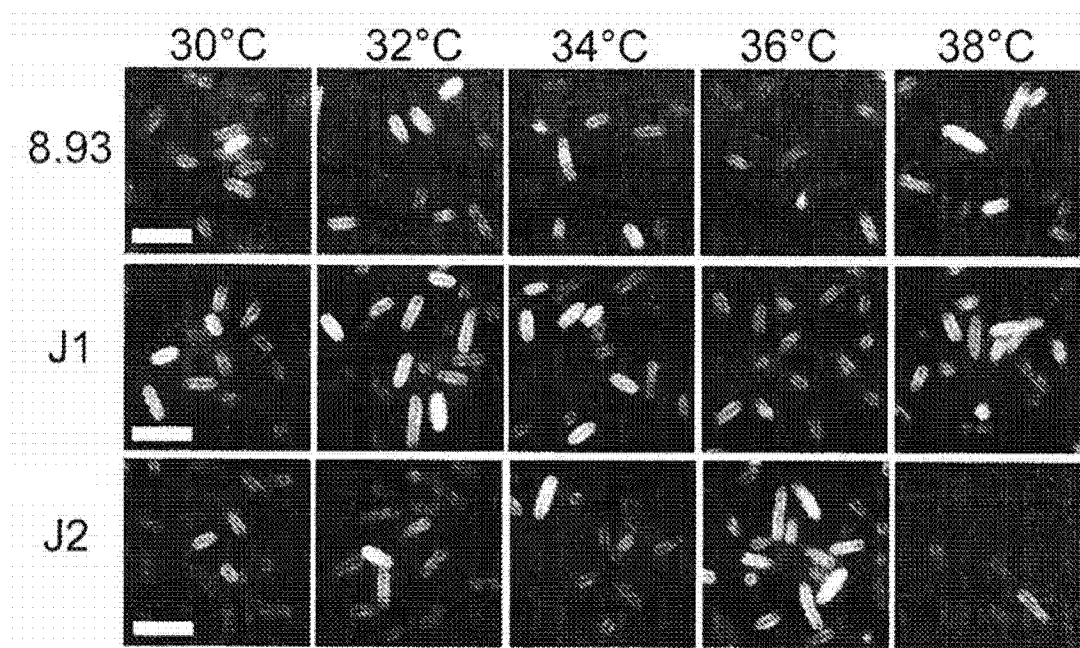


图 5

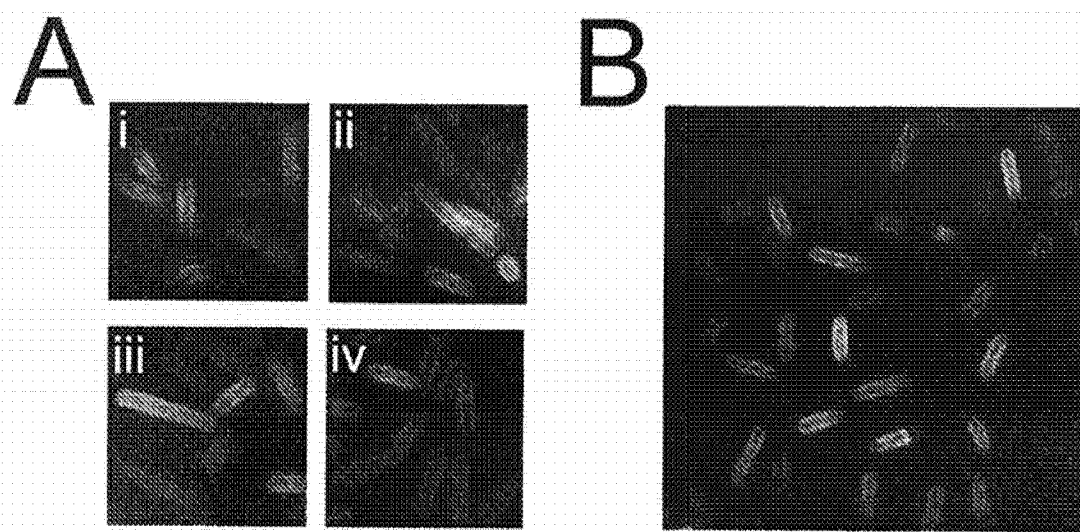


图 6

IGLV1-51 (aka DPL5) (SEQ ID NO: 15)
 QSVLTQPPSVSAAPGGKVTISCS**GSSSNIGNNY**VSWYQQLPGTAPKLLIYDNNKRPSGIPDRFSGSKSGTSATLGTGLQTGDEADYYCGT**WDSLSL SAG**

IGLV1-40 (aka DPL6) (SEQ ID NO: 18)
 QSVLTQPPSVSGAPGQRTVITISCT**GSSSNIGAGYD**VHWHYQQLPGTAPKLLIY**GNSNRPS**GVDPDRFSGSKSGTSASLITGLOAEDEADYYCQ**YDSSLSGS**

IGLV1-44 (aka DPL2) (SEQ ID NO: 21)
 QSVLTQPPSASGTPGQRTVITISCS**GSSSNIGSNT**VNWHYQQLPGTAPKLLIY**SNNRPS**GVDPDRFSGSKSGTSASLISGLQSEDEADYYCAA**WDDSLNGP**

IGV1-47 (aka DPL3) (SEQ ID NO: 24)
 QSVLTQPPSASGTPGQRTVITISCS**GSSSNIGSNY**VVWHYQQLPGTAPKLLIY**SNNRPS**GVDPDRFSGSKSGTSASLISGLRSEDEADYYCAA**WDDSLSGP**

IGLV3-1 (aka DPL23) (SEQ ID NO: 6)
 QSVLTQPPSVSVSPGQTASITCS**GDKLGDKYAC**WYQKPGQSPVLVIY**QDSKRPS**GIPERFSGSNSGNTATLTISGTQAMDEADYYCQA**WDSSTA**

IGLV3-19 (SEQ ID NO: 27)
 SSELTDPAVSVALGQTVRITCQ**GDLSRSYY**ASWYQKPGQAPVLVIY**GKNNRPS**GIPDRFSGSSSGNTASLTITGAOAEDEADYYCNS**RDSSGNHL**

IGLV3-21 (SEQ ID NO: 9)
 QSVLTQPPSVSVAPGQTARITCG**GNNIGSKS**VHWHYQKPGQAPVLWY**DDSDRPS**GIPERFSGSNSGNTATLTISRVEAGDEADYYCQV**WDSSTDHP**

IGLV6-57 (SEQ ID NO: 12)
 NFMLTQPHSVSES PGKTVTISCT**GSSGSIASNY**VQWYQKRPGSAPTTVIY**EDNQRPS**GVDPDRFSGSIDSSSNSASLTISGLKTEDEADYYCQ**YDSSN**

IGHV3-23 (aka DP47) (SEQ ID NO: 3)
 EVQLLESGGGLVQPGGSLRLSCAAS**GFTFSSY**AMSWVRQAPGKGLWVSA**SGSGGSTY**YADSVKGRFTISRDNKNTLYLQMNSLR AEDTAVYYCAK

图 7

具有可变 CDR3 区的 IGLV3-1::IGHV3-23 支架 (SEQ ID NO: 84)

```

ATG GGA GAC GGT CAG TCT GTG CTG ACT CAG CCA CCC
TCAGTGTCCGTGTCCCCAGGACAGACAGCCAGCATCACCTGCTCTGGAGAT
AAATTGGGGGATAAATATGCTTGCTGGTATCAGCAGAAGCCAGGCCAGTCC
CCTGTGCTGGTCATCTATCAAGATAGCAAGCGGCCCTCAGGGATCCCTGAG
CGATTCTCTGGCTCCAACCTCTGGGAACACAGCCACTCTGACCATCAGCGGG
ACCCAGGCTATGGATGAGGCTGACTATTACTGTCAGGCGTGGGAC
NNTNNTNNTGGAGGTNNTNNTNNT
ACTGTGGTGTTTCGGCACGGGCACCAAGCTCATCATTTTCGTCT
CAGACCGGTGGTCTGGTGGTGGTGGTCTGGCGGCGGCGGCTCCGGTGGT
GGTGGATCC
GAAGTCCAACCTGCTGGAGTCCGGCGGTGGCCTGGTGCAGCCAGGTGGCAGC
CTGCGCCTGAGCTGCGCCGCATCCGGTTTTACTTTTCAGCAGCTACGCGATGT
CGTGGGTGCGCCAGGCACCGGGCAAGGGCCTGGAGTGGGTTCAGCGCCATC
AGCGGTAGCGGCGGTTCTACGTATTATGCGGACAGCGTCAAGGGCCGTTTC
ACCATCAGCCGTGACAATTCAAAAACACCCTGTACTTGCAGATGAACAGC
TTGCGTGCGGAAGATACGGCTGTTTACTACTGTGCGAAA
NNTNNTNNTGGANNTNNTNNT
GCCTTIGATATTTGGGGTCAAGGTACCATGGTTACCGTGAGC AGC

```

翻译 (SEQ ID NO: 85):

```

MGDGSVLTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRPSG
IPERFSGSNSGNTATLTISGTQAMDEADYYCQAWD NNNGNNN TVVFGTGKVTVSS QTGGSGGGG
SGGGSGGGG EVQLLESGGGLVQPGSLRLSCAASGFTFSSYAMSWVRQAPGKLEWVS
AISGSGGSTYYADSVKGRFTISRDN SKNTLYLQMNSLR AEDTAVYYCAK NNNGNNN
AFDIWGQGTMTVTVSS

```

图 8

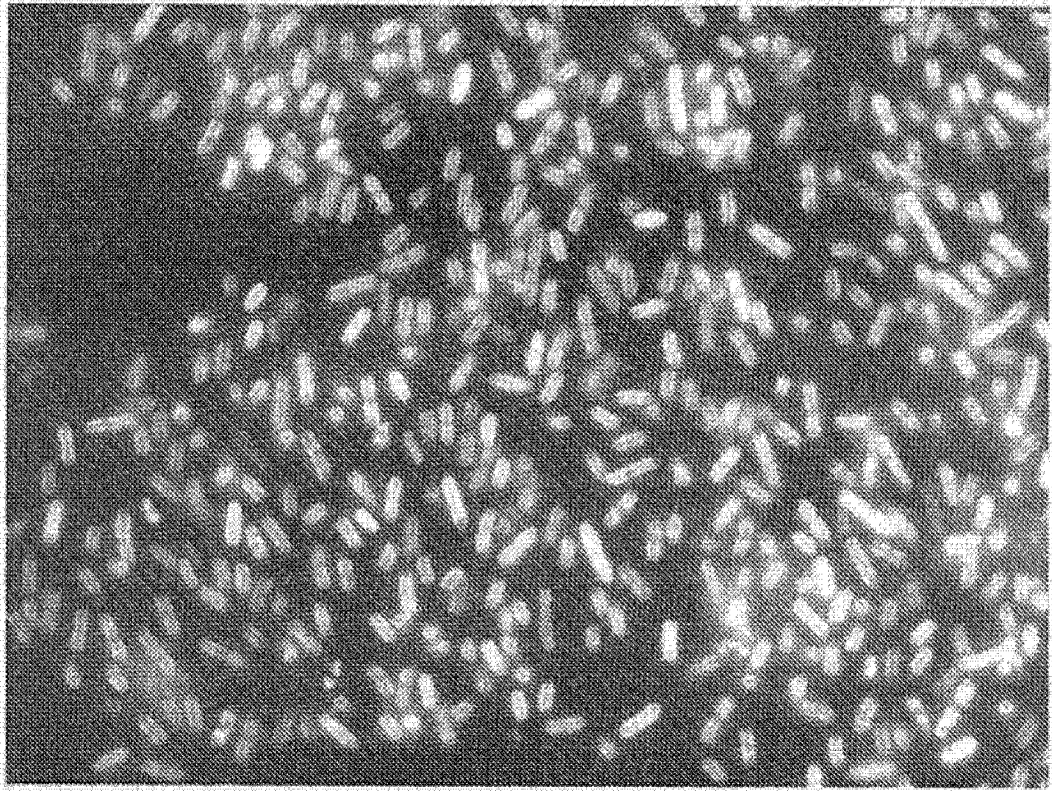


图 9

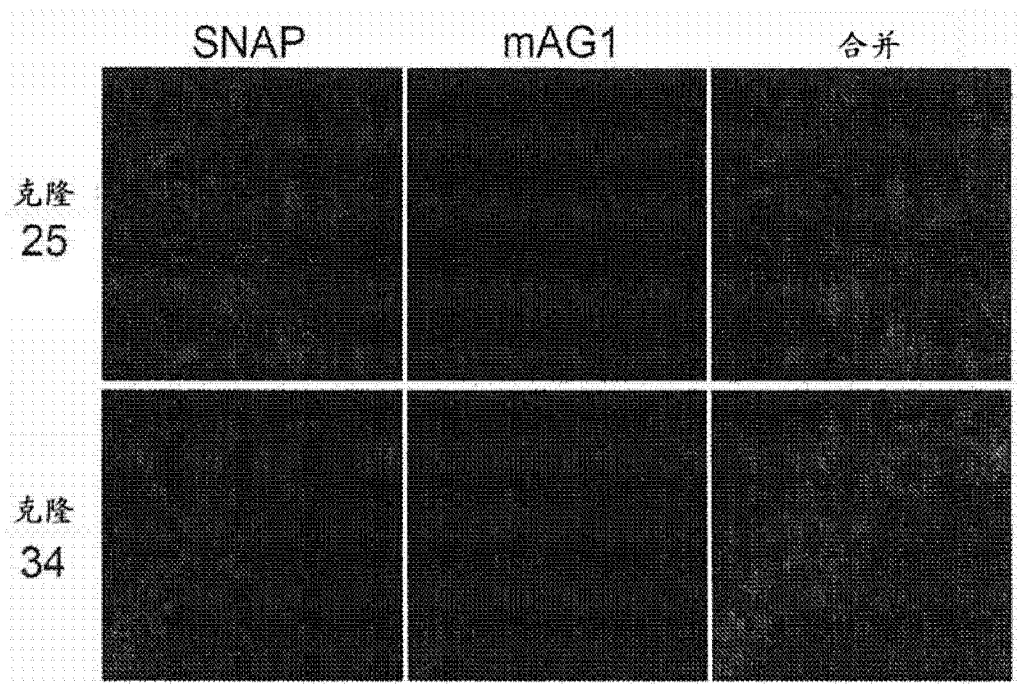


图 10

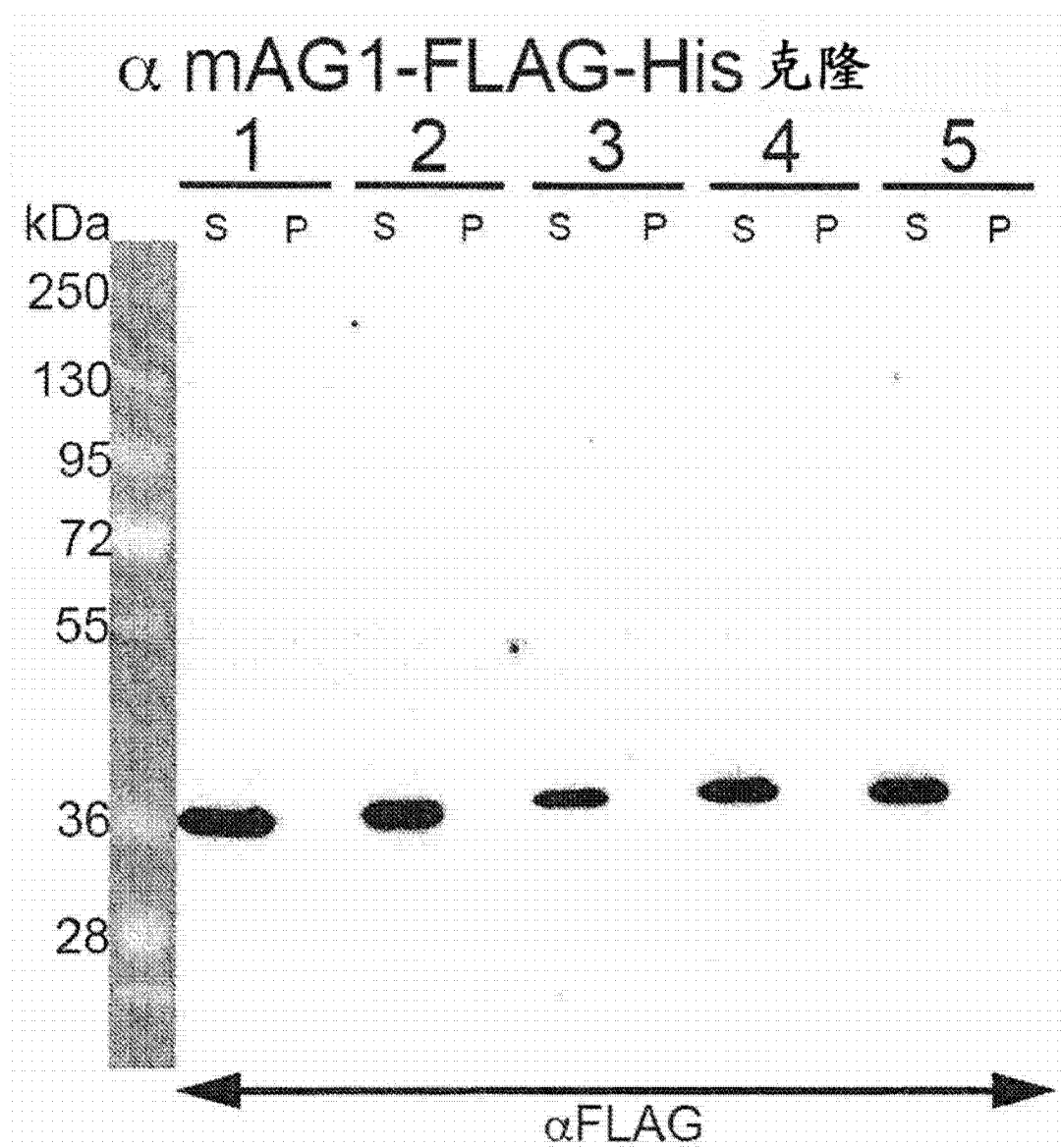


图 11

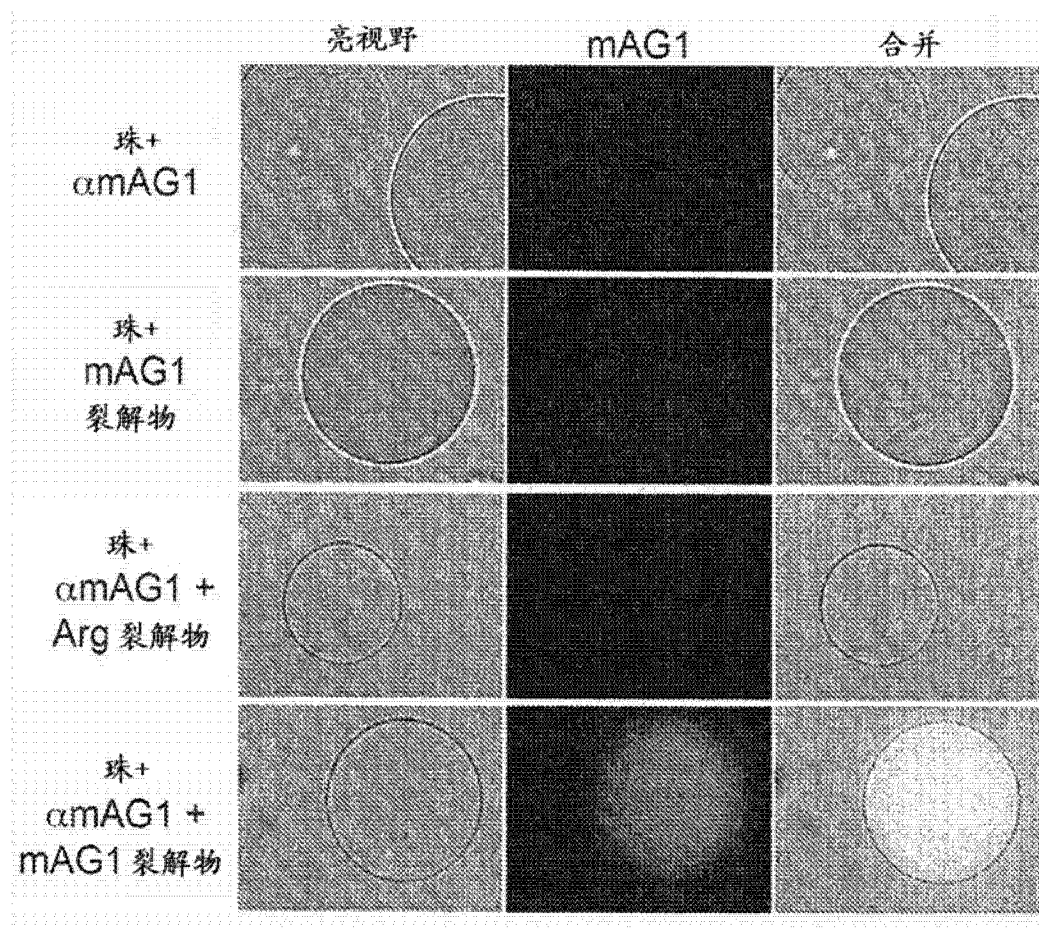


图 12

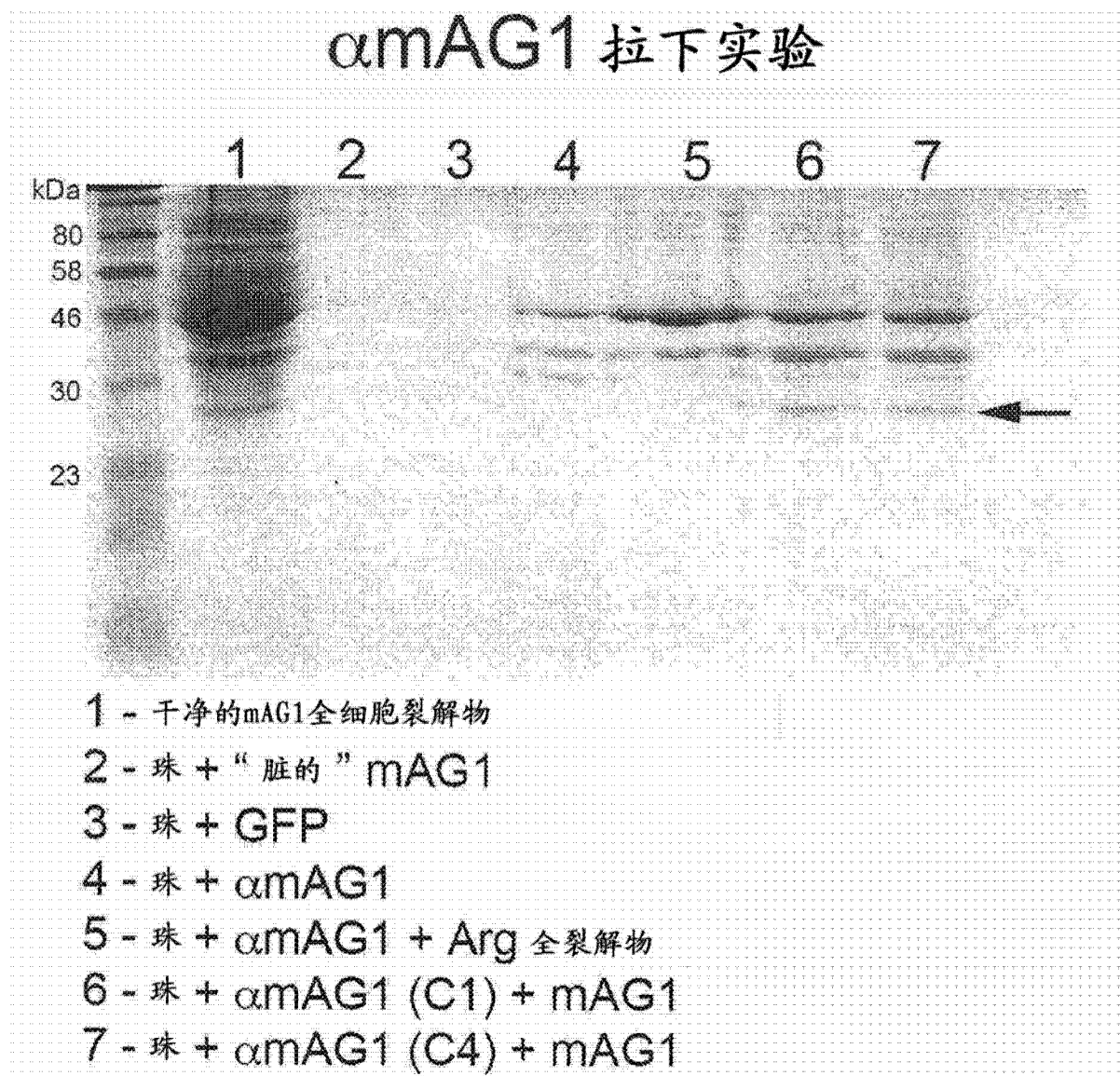


图 13

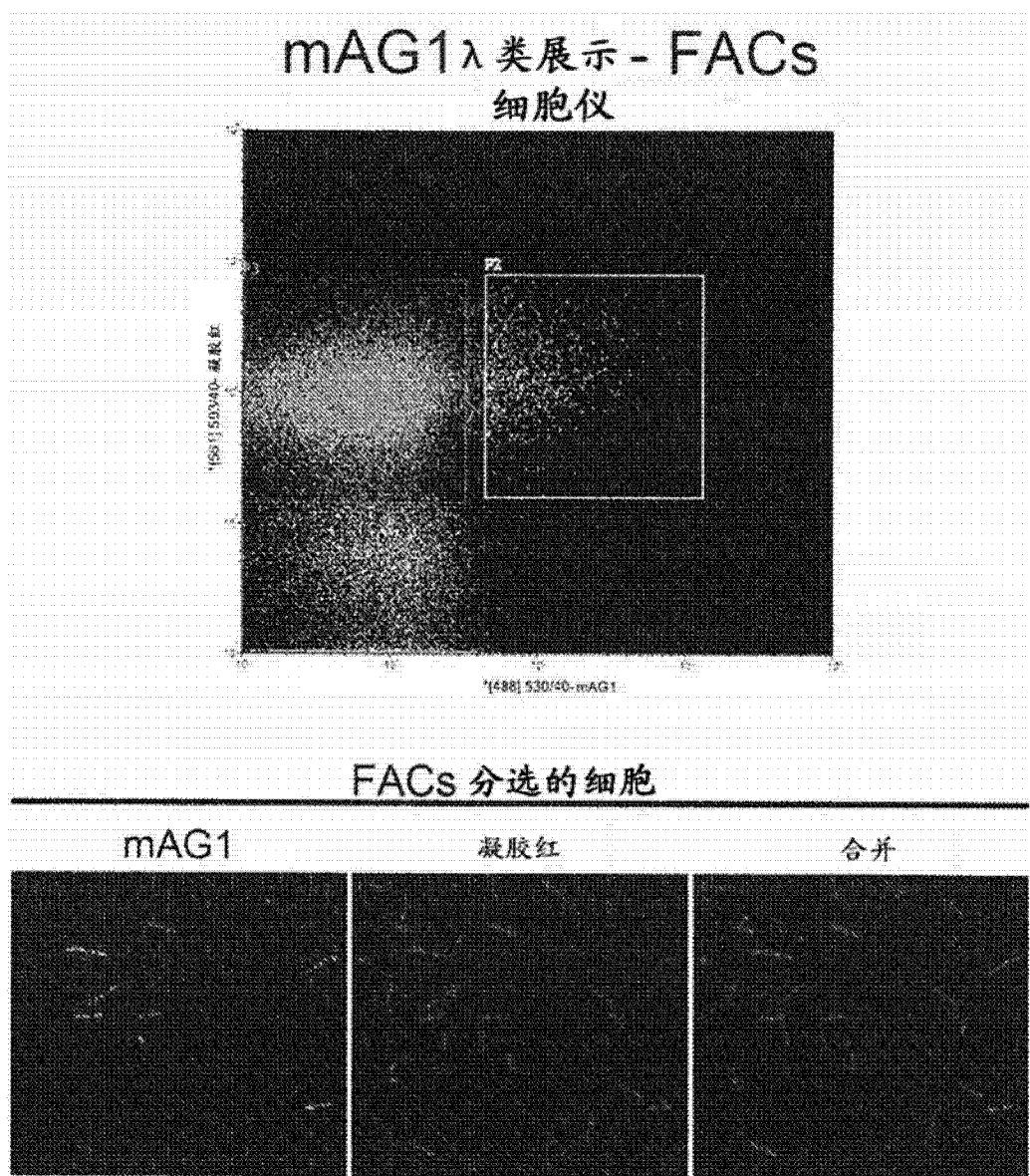


图 14