

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200580028093.8

[51] Int. Cl.

A61K 38/00 (2006.01)

G01N 33/53 (2006.01)

C07H 21/02 (2006.01)

[43] 公开日 2007年8月15日

[11] 公开号 CN 101018559A

[22] 申请日 2005.6.17

[21] 申请号 200580028093.8

[30] 优先权

[32] 2004. 6. 17 [33] US [31] 10/871,602

[32] 2004. 9. 30 [33] US [31] 10/957,351

[86] 国际申请 PCT/US2005/021558 2005.6.17

[87] 国际公布 WO2006/009888 英 2006.1.26

[85] 进入国家阶段日期 2007.2.16

[71] 申请人 安姆根山景公司

地址 美国加利福尼亚州

[72] 发明人 W·P·C·斯特默

D·V·佩尔罗思 S·萨特亚尔

B·M·阿尔巴 A·巴克

A·N·杜瓜伊 Q·刘

J·西尔弗曼 R·史密斯

C·斯维默

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 权陆军 黄可峻

权利要求书4页 说明书111页 附图17页

[54] 发明名称

c-MET 激酶结合蛋白

[57] 摘要

提供了包含结合 c-MET 的单体域或其部分的多肽。

1. 包含结合 c-MET 的单体域的多肽，其中所述单体域：
是非天然发生的包含 30-50 个氨基酸的单体域；且
包含至少一个二硫键。

2. 权利要求 1 的多肽，其中所述单体域是 LDL 受体 A 类单体域。

3. 权利要求 1 的多肽，其中所述单体域是包含下述序列的 LDL 受体 A 类单体域：

EFXCXNGXCIPXXWXCDGXDDCGDXSDE,

其中 X 是任意的氨基酸。

4. 权利要求 1 的多肽，其中所述多肽包含至少一个且不超过 6 个结合 c-MET 的单体域。

5. 权利要求 1 的多肽，其中所述多肽包含至少 2 个结合 c-MET 的单体域。

6. 权利要求 1 的多肽，其中所述多肽还包含第 2 个单体域，其中所述第 2 个单体域具有对血液因子的结合特异性，从而当该多肽被注射进动物中时，与缺少血液因子-结合性单体域的多肽的血清半衰期相比，增加了该多肽的血清半衰期。

7. 权利要求 6 的多肽，其中所述血液因子是血清清蛋白、免疫球蛋白或红细胞。

8. 权利要求 6 的多肽，其中所述第 2 个单体域结合免疫球蛋白 (IgG)，且所述第 2 个单体域是包含选自下述的序列的 LDL 受体 A 类单体域：

CXSSGRCIPXXWVCDGXDCRDXSDE, 和

CXSSGRCIPXXWLCGXDCRDXSDE,

其中 X 是任意的氨基酸。

9. 权利要求 6 的多肽，其中所述第 2 个单体域结合免疫球蛋白 (IgG)，且所述第 2 个单体域是包含下述序列的 LDL 受体 A 类单体域：

[EQ]FXCRX[ST]XRC[IV]XXXW[ILV]CDGXDCXD[DN]SDE,

其中 X 是任意的氨基酸，且在方括号中的氨基酸是在单个位置的

替代氨基酸。

10. 权利要求 6 的多肽，其中所述第 2 个单体域包含 CHPTGQFRCRSSGRCVSPWVCDGDNDGNSDEENC。

11. 权利要求 1 的多肽，其中所述单体域各自都是 35 至 45 个氨基酸。

12. 权利要求 1 的多肽，其中每个单体域包含 3 个二硫键。

13. 权利要求 1 的多肽，其中所述单体域是非天然发生的 LDL-受体 A 类域。

14. 权利要求 1 的多肽，其中所述单体包含氨基酸序列，其中：
该序列中的至少 10%氨基酸是半胱氨酸；和/或
至少 25%氨基酸是非天然发生的氨基酸。

15. 鉴别结合 c-MET 的多肽的方法，该方法包含，
针对对 c-MET 的亲和力筛选多肽文库，和
选择包含至少一个结合 c-MET 的单体域的多肽，其中所述单体域：

具有 30-100 个氨基酸；
是非天然发生的单体域；
包含至少一个二硫键。

16. 权利要求 15 的方法，其中所述选择步骤包含选择降低 HGF-介导的细胞增殖和/或迁移的多肽。

17. 权利要求 15 的方法，还包含选择抑制动物中的肿瘤生长的多肽。

18. 权利要求 15 的方法，其中所述单体包含氨基酸序列，其中：
该序列中的至少 10%氨基酸是半胱氨酸；和/或
至少 25%氨基酸是非天然发生的氨基酸。

19. 权利要求 15 的方法，还包含，将选择的多肽中的单体域连接到第 2 个单体域上，以形成多聚体文库，每个多聚体包含至少 2 个单体域；

针对结合 c-MET 的能力筛选多聚体文库；和
选择结合 c-MET 的多聚体。

20. 权利要求 15 的方法，还包含，将选择的多肽中的单体域连接到第 2 个单体域上，以形成多聚体文库，每个多聚体包含至少 2 个

单体域；

针对结合除 c-MET 以外的靶分子的能力筛选多聚体文库；和选择结合靶分子的多聚体。

21. 权利要求 15 的方法，还包含突变至少一个单体域，从而提供包含突变的单体域的文库的步骤。

22. 权利要求 15 的方法，其中将单体域的文库表达为噬菌体展示、核糖体展示或细胞表面展示。

23. 权利要求 15 的方法，其中所述多肽包含至少 2 个单体域，且单体域通过接头相连。

24. 权利要求 15 的方法，其中所述单体域各自都是 35 至 45 个氨基酸。

25. 权利要求 15 的方法，其中每个单体域包含 3 个二硫键。

26. 权利要求 15 的方法，其中至少一个单体域是 LDL-受体 A 类域。

27. 权利要求 15 的方法，其中所述单体域包含氨基酸序列，其中：

该序列中的至少 10% 氨基酸是半胱氨酸；和/或至少 25% 氨基酸是非天然发生的氨基酸。

28. 编码权利要求 1 的多肽的多核苷酸。

29. 包含结合免疫球蛋白-G (IgG) 的单体域的多肽，其中所述单体域是包含选自下述的序列的 LDL 受体 A 类单体域：

CXSSGRCIPXXWVCDGXXDCRDXSDE,

CXSSGRCIPXXWLCDGXXDCRDXSDE, 和

[EQ]FXCRX[ST]XRC[IV]XXXW[ILV]CDGXXDCXD[DN]SDE

其中 X 是任意的氨基酸，且在方括号中的氨基酸是在单个位置的替代氨基酸；且

其中当该多肽被注射进动物中时，与缺少结合 IgG 的单体域的多肽的血清半衰期相比，该多肽具有增加的血清半衰期。

30. 权利要求 29 的多肽，其中所述单体域包含
CHPTGQFRRCSSGRCVSPTWVCDGDNDCGDNSDEENC SAPASEPPGSL。

31. 权利要求 29 的多肽，还包含第 2 个单体域，所述第 2 个单体域具有对除 IgG 以外的分子的结合特异性，其中该第 2 个单体域：

具有 30-100 个氨基酸；
是非天然发生的单体域；
包含至少一个二硫键。

32. 权利要求 31 的多肽，其中所述第 2 个单体域是非天然发生的 LDL-受体 A 类域。

33. 编码权利要求 29 的多肽的多核苷酸。

c-MET 激酶结合蛋白

相关申请的交叉参考

本申请是 2004 年 9 月 30 日提交的美国专利申请号 10/957,351 的部分继续申请，后者是 6 月 17 日提交的美国专利申请号 10/871,602 的部分继续申请，它们各自的公开内容为所有目的整体引作参考。

发明背景

肝细胞生长因子/分散因子(HGF/SF)是间充质-衍生的多效因子，其调节细胞生长、细胞运动性和各种类型的细胞的形态发生，并在胚胎发育和器官发生的过程中，介导上皮-间充质相互作用，后者负责形态发生的组织相互作用。尽管 HGF 最初被鉴别为肝细胞的有效促分裂原，但它还被鉴别为生血管生长因子。

Met 在 1980 年代首次被鉴别为癌基因，且是 HGF 的受体。发现原癌基因 c-MET 编码受体酪氨酸激酶。响应于 HGF 治疗，观察到一定范围的活性：受体的磷酸化，信号传导中间体 Gab-1/Grb2 的停靠，最终导致激酶的活化，所述激酶例如 PI3K、ERK1 和 2 和 AKT。这些活性有助于细胞生长、存活、迁移和新血管形成。

受体酪氨酸激酶 Met 和它的配体肝细胞生长因子/分散因子(HGF/SF)的不适当的表达或信号传导，与广泛种类的实体人肿瘤的进攻性表型和较差的临床预后有关。

4 种证据巩固了 c-MET 在癌症中起作用的情形：

首先，异位地超表达 HGF 和/或 Met 的小鼠和人细胞系在无胸腺的裸鼠中变成致瘤的和转移的。其次，在人肿瘤细胞中 Met 或 HGF 表达的下调，降低了它们的致瘤潜力。表达作为转基因的受体或配体的小鼠模型形成了各种类型的肿瘤和转移的肿瘤。第三，大量研究表明，HGF 和/或 Met 经常在癌、其它类型的人实体瘤和它们的转移灶中表达，且 HGF 和/或 Met 超表达或错表达(misexpression)经常与不良预后相关。第四，通过激活已经在散发和遗传形式的人肾乳头状癌中发现的突变，提供了使 Met 牵连于人癌症中的明确证据。

发明概述

本发明提供了多肽，其包含结合 c-MET 的单体域。在有些实施方案中，该单体域：

是由 30-50 个氨基酸组成的非天然发生的单体域；
包含至少一个二硫键；且任选地，
结合离子。

在有些实施方案中，单体域是 LDL 受体 A 类单体域。在有些实施方案中，单体域是包含下述序列的 LDL 受体 A 类单体域：

EFXCXNGXCIPXXWXCDGXDDCGDXSDE,

其中 X 是任意的氨基酸。

在有些实施方案中，该多肽包含至少一个且不超过 6 个结合 c-MET 的单体域。在有些实施方案中，该多肽包含至少 2 个结合 c-MET 的单体域。

在有些实施方案中，该多肽还包含第 2 个单体域，其中该第 2 个单体域具有对血液因子的结合特异性，从而当该多肽被注射进动物中时，与缺少血液因子-结合性单体域的多肽的血清半衰期相比，增加了该多肽的血清半衰期。在有些实施方案中，血液因子是血清清蛋白、免疫球蛋白或红细胞。在有些实施方案中，第 2 个单体域结合免疫球蛋白(IgG)，且第 2 个单体域是包含选自下述的序列的 LDL 受体 A 类单体域：

CXSSGRCIPXXWVCDGXXDCRDXSDE, 和
CXSSGRCIPXXWLCDGXXDCRDXSDE,

其中 X 是任意的氨基酸。

在有些实施方案中，第 2 个单体域结合免疫球蛋白(IgG)且第 2 个单体域是包含下述序列的 LDL 受体 A 类单体域：

[EQ]FXCRX[ST]XRC[IV]XXXW[ILV]CDGXXDCXD[DN]SDE,

其中 X 是任意的氨基酸，且在方括号中的氨基酸是在单个位置的替代氨基酸。在有些实施方案中，第 2 个单体域包含 CHPTGQFRCRSSGRCVSPTWVCDGDND CGDNSDEENC SAPASEP PGS L。在有些实施方案中，第 2 个单体域包含 CHPTGQFRCRSSGRCVSPTWVCDGDND CGDNSDEENC。

在有些实施方案中，至少一个单体域与 c-MET 的结合抑制 Met 的二聚化。在有些实施方案中，至少一个单体域结合 c-MET 的 Sema 域，从而阻止 Met 配体与 c-MET 的结合。

在有些实施方案中，该多肽包含至少一个且不超过 6 个单体域。在有些实施方案中，该多肽包含至少 2 个单体域，且单体域通过接头相连。在有些实施方案中，接头是肽接头。在有些实施方案中，接头是 4-12 个氨基酸长。

在有些实施方案中，单体域各自是 35-45 个氨基酸。

在有些实施方案中，每个单体域包含 2 个二硫键。在有些实施方案中，每个单体域包含 3 个二硫键。

在有些实施方案中，离子是金属离子。在有些实施方案中，离子是钙离子。

在有些实施方案中，至少一个单体域源自 LDL-受体 A 类域。在有些实施方案中，至少一个单体域源自 EGF-样域。

在有些实施方案中，单体包含这样的氨基酸序列，其中该序列中的至少 10% 氨基酸是半胱氨酸；和/或至少 25% 氨基酸是非天然发生的氨基酸。

本发明也提供了鉴别结合 c-MET 的多肽的方法。在有些实施方案中，该方法包含，

针对对 c-MET 的亲合力筛选多肽文库，和

选择包含至少一个结合 c-MET 的单体域的多肽，其中该单体域：是非天然发生的单体域；

包含至少一个二硫键；和

结合离子。

在有些实施方案中，选择的多肽包含单体域，所述单体域包含下述的任一个：

Cxxx[EQ]FxCxSTxRC[IV]xxxWxCDGDNDCEdXSDEx

Cxxxx[EQ]FECxSTxRC[IV]xxxWxCDGxNDCEDXSDEx

Cxxxx[EQ]FxCxSTxRC[ILV]PxxWxCDGxxDCEDXSDExx

Cxxx[EQ]FQCxSTxRC[IV]PxxWxCDGxNDCEDSSDExxC

Cxxxx[EQ]FxCxxxxxC[ILV]xxxxxxxxxxDCxDXSDEx

Cxxx [EQ] FxCxSTGRCxPxxWxCxGxNDCEDxSDEx
 Cxxxx [EQ] FxCxSTxRC [ILV] xxxWxCxxxxDCxDxSDxxxxxCx
 Cxxx [EQ] FxCxxxxxC [ILV] xxxWxCxGxNDCxDxSxExxxxC
 Cxxxx [EQ] FxCxSTxRC [ILV] PxxWxCxGxxDCxDxSDEx
 Cxxxx [EQ] FxCxxxxxC [ILV] xxxWxCxGxxDCxDxSDEx
 EFXCXNGXCIPXXWXCDGXDDCGDXSDE。

在有些实施方案中，选择步骤包含，选择减少 HGF-介导的细胞增殖和/或迁移的多肽。在有些实施方案中，该方法还包含，选择抑制动物中的肿瘤生长的多肽。

在有些实施方案中，单体域包含这样的氨基酸序列，其中该序列中的至少 10% 氨基酸是半胱氨酸；和/或至少 25% 氨基酸是非天然发生的氨基酸。

在有些实施方案中，该方法还包含，将选择的多肽中的单体域连接到第 2 个单体域上，以形成多聚体文库，每个多聚体包含至少 2 个单体域；

针对结合 c-MET 的能力筛选多聚体文库；和
选择结合 c-MET 的多聚体。

在有些实施方案中，该方法还包含，将选择的多肽中的单体域连接到第 2 个单体域上，以形成多聚体文库，每个多聚体包含至少 2 个单体域；

针对结合除 c-MET 以外的靶分子的能力筛选多聚体文库；和
选择结合靶分子的多聚体。

在有些实施方案中，该方法还包含下述步骤，突变至少一个单体域，从而提供包含突变的单体域的文库。

在有些实施方案中，将单体域的文库表达为噬菌体展示、核糖体展示或细胞表面展示。

在有些实施方案中，该多肽包含至少 2 个单体域，且单体域通过接头相连。在有些实施方案中，接头是肽接头。在有些实施方案中，接头是 4-12 个氨基酸长。

在有些实施方案中，单体域各自是 35-45 个氨基酸。

在有些实施方案中，每个单体域包含 2 个二硫键。在有些实施方案中，每个单体域包含 3 个二硫键。

在有些实施方案中，离子是金属离子。在有些实施方案中，离子是钙离子。

在有些实施方案中，至少一个单体域源自 LDL-受体 A 类域。在有些实施方案中，至少一个单体域源自 EGF-样域。

在有些实施方案中，单体域包含这样的氨基酸序列，其中该序列中的至少 10% 氨基酸是半胱氨酸；和/或至少 25% 氨基酸是非天然发生的氨基酸。

本发明也提供了编码包含结合 c-MET 的单体域的多肽的多核苷酸，其中该单体域：

是由 30-50 个氨基酸组成的非天然发生的单体域；
包含至少一个二硫键。

本发明也提供了：

多肽，其包含结合免疫球蛋白-G(IgG)的单体域，其中该单体域是包含选自下述的序列的 LDL 受体 A 类单体域：

CXSSGRCIPXXWVCDGXXDCRDXSDE,
CXSSGRCIPXXWLC DGXXDCRDXSDE, 和
[EQ]FXCRX[ST]XRC[IV]XXXW[ILV]CDGXXDCXD[DN]SDE

其中 X 是任意的氨基酸，且在方括号中的氨基酸是在单个位置的替代氨基酸；且

其中当该多肽被注射进动物中时，与缺少结合 IgG 的单体域的多肽的血清半衰期相比，该多肽具有增加的血清半衰期。

在有些实施方案中，该单体域包含 CHPTGQFRCRSSGRCVSPTWVCDGDNDCGDNSDEENC SAPASEP PGS L。在有些实施方案中，该单体域包含 CHPTGQFRCRSSGRCVSPTWVCDGDNDCGDNSDEENC。

在有些实施方案中，该多肽包含第 2 个单体域，所述第 2 个单体域具有对除 IgG 以外的分子的结合特异性，其中该第 2 个单体域：

具有 30-100 个氨基酸；
是非天然发生的单体域；
包含至少一个二硫键。

在有些实施方案中，第 2 个单体域是非天然发生的 LDL-受体 A

类域。

本发明也提供了编码如上所述的多肽的多核苷酸。

定义

除非另有说明，下面的定义替代本领域的那些。

“Met”也称作“c-MET”，指肝细胞生长因子/分散因子(HGF/SF)-结合受体酪氨酸激酶。响应于 HGF 治疗，观察到一定范围的活性：受体的磷酸化，信号传导中间体 Gab-1/Grb2 的停靠，最终导致激酶的活化，所述激酶例如 PI3K、ERK1 和 2 和 AKT。这些活性有助于细胞生长、存活、迁移和新血管形成。见，例如，Birchmeier 等, Mol. Cell Biol. 4: 915-925 (2003)。Met 的氨基酸序列是已知的，且展示在 SEQ ID NO: 1。见，例如，Park 等, Proc. Natl. Acad. Sci. USA 84(18): 6379 (1987)。

术语“单体域”或“单体”可互换地使用，且在本文中指在蛋白或多肽中发现的离散区域。在没有侧接的天然氨基酸序列存在的情况下，单体域在溶液中形成天然的三维结构。本发明的单体域经常结合靶分子。例如，形成结合于靶分子的三维结构的多肽是单体域。如本文所使用的，术语“单体域”不包括抗体的互补性决定区(CDR)。

术语“环”指单体域的一部分，其一般地通过单体域蛋白的支架结构的装配暴露于环境中，且参与靶结合。本发明提供了通过特定特征鉴别出的 3 类环，所述特征例如，形成二硫键的潜力、二级蛋白结构之间的桥接和分子动力学(即柔性)。3 类环序列是半胱氨酸-定义的环境序列、结构-定义的环境序列和 B-因子-定义的环境序列。

如本文所使用的，术语“半胱氨酸-定义的环境序列”指天然发生的单体域-编码序列的子序列，其在每个末端结合有半胱氨酸残基，所述残基相对于相同家族的至少一种其它天然发生的单体域是保守的。如下鉴别半胱氨酸-定义的环境序列：天然发生的单体域的多重序列比对，随后序列分析，以鉴别保守的半胱氨酸残基。每个相邻的保守的半胱氨酸残基对之间的序列是半胱氨酸-定义的环境序列。半胱氨酸-定义的环境序列不包括邻近每个末端的半胱氨酸残基。具有半胱氨酸-定义的环境序列的单体域包括 LDL 受体 A-域、EGF-样域、sushi 域、纤连蛋白 1 型域等。因而，例如，在共有序列 CX₆CX₄CX₆CX₅CX₈C

代表的 LDL 受体 A-域的情况下，其中 X_6 、 X_4 、 X_5 和 X_8 分别代表着包含指定数目的氨基酸的半胱氨酸-定义的环境序列。

如本文所使用的，术语“结构-定义的环境序列”指单体-域编码序列的子序列，其在每个末端结合各自形成二级结构的子序列。如 Frishman, D. 和 Argos, P. (1995) “Knowledge-based secondary structure assignment,” *Proteins*, 23(4): 566-79 (在环球网上也见 //hgmp.mrc.ac.uk/Registered/Option/stride.html) 所述，根据分配蛋白二级结构的算法 STRIDE，鉴别具有已知的三维结构的蛋白的二级结构。根据 Jones, D. T. (1999), “Protein secondary structure prediction based on position-specific scoring matrices,” *J. Mol. Biol.*, 292: 195-202 (也见 McGuffin, L. J., Bryson, K., Jones, D. T. (2000) “The PSIPRED protein structure prediction server,” *Bioinformatics*, 16: 404-405, 以及环球网上的 //bioinf.cs.ucl.ac.uk/psipred/) 所述的算法，鉴别具有未知的或未表征的三维结构的蛋白的二级结构。二级结构包括，例如，折叠、螺旋等。具有结构-定义的环境序列的单体域的实例是 C2 域、Ig 域、因子 5/8 C 域、纤连蛋白 3 型域等。

术语“B-因子-定义的环境序列”指单体-域编码序列的至少 3 个氨基酸残基的子序列，其中 B-因子-定义的环的 α 碳的 B-因子在整个单体域的 25% 最高 α 碳 B 因子中。一般地，子序列的平均 α 碳 B-因子是至少约 65。如本文所使用的，术语“B-因子” (或“温度因子”或“Debye-Waller 因子”) 源自 X-射线散射数据。B-因子是可以应用于每个原子或原子组的 X-射线散射术语的因子，其描述了电子密度扩散的程度。在本发明的实践中采用的 B-因子可以是各向同性的或各向异性的。术语“平均 α 碳 B-因子”指：

$$\frac{\sum_{i=1}^n \text{B-因子}_{C_{\alpha i}}}{n}$$

其中 n 对应着环中的残基数，且至少是 3，且 B-因子 $C_{\alpha i}$ 是环的氨基酸残基 i 的 α 碳的 B-因子。

术语“多聚体”在本文中用于指包含至少 2 个单体域的多肽。多聚体中的分开的单体域可以通过接头连接到一起。多聚体也称作组合的嵌合蛋白或重组的嵌合蛋白。

术语“家族”和“家族类”可互换地使用，指基于它们的氨基酸序列的相似性归类到一起的蛋白。这些类似的序列通常是保守的，因为它们对于蛋白的功能和/或蛋白的三维结构的维持是重要的。这样的家族的实例包括 LDL 受体 A-域家族、EGF-样家族等。另外，可以基于共有的序列基序，将结合相同靶分子的有关序列分成家族。

术语“配体”在本文中称作“靶分子”，包括非常多种物质和分子，其范围为简单的分子到复杂的靶。靶分子可以是蛋白、核酸、脂质、碳水化合物或能被多肽域识别的任何其它分子。例如，靶分子可以包括化合物(即，非生物的化合物，例如，有机分子、无机分子或具有有机和无机原子的分子，但是不包括多核苷酸和蛋白)、化合物的混合物、空间上定位的化合物的阵列、生物大分子、噬菌体肽展示文库、多核糖体肽展示文库、从生物材料(例如细菌、植物、真菌或动物(例如，哺乳动物)细胞或组织)制成的提取物、蛋白、毒素、肽激素、细胞、病毒等。其它靶分子包括，例如，整个细胞、整个组织、相关的或不相关的蛋白的混合物、病毒或细菌菌株的混合物等。通过包含在本文所述的筛选测定中，或通过增强或抑制特定的蛋白相互作用，也可以定义靶分子(即，选择性地抑制 2 种预定多肽之间的结合相互作用的试剂)。

术语“接头”在本文中用于指接合或连接 2 个或更多个离散的单独的单体域的部分或部分组。当在多聚体中连接到一起时，接头允许离散的单独的单体域保持分开。接头部分一般地是基本上线性的部分。合适的接头包括多肽、多核酸、肽核酸等。合适的接头也任选地包括被取代的亚烷基部分，其具有一个或多个整合在碳主链中的氧原子。一般地，接头的分子量小于约 2000 道尔顿。更一般地，接头的分子量小于约 1500 道尔顿，且通常小于约 1000 道尔顿。接头可以足够小，以允许离散的单独的单体域协作，例如，当多聚体中的每个离散的单独的单体域通过分开的结合位点结合相同的靶分子时。示例性的接头包括编码多肽、或氨基酸的多肽或其它非天然发生的部分的多核苷酸。接头可以是天然序列的一部分、其变体或合成的序列。接头可以包含，例如，天然发生的、非天然发生的氨基酸或二者的组合。

术语“分开的”在本文中用于指部分的性质，所述部分是独立的且甚至当与其它部分(包括，例如，其它单体域)复合时，仍保持独

立。单体域是蛋白中的分开的域，因为它具有独立的性质，其可以被识别，并与蛋白分离。例如，LDLR 中的 A-域的配体结合能力是独立的性质。分开的其它实例包括多聚体中的分开的单体域，甚至当通过接头在多聚体中复合或连接到一起时，其仍保持分开的独立域。分开的性质的另一个实例是多聚体中的分开的配体结合位点。

如本文所使用的，“定向进化”指在递归过程中产生、表达多核苷酸变体和筛选其活性(例如，具有结合活性的多肽)的过程。选择筛选中的一种或多种候选物，然后使用编码选择的候选物的多核苷酸，重复该过程，以产生新的变体。定向进化包含至少 2 轮的变化产生，且可以包括 3、4、5、10、20 或更多轮的变化产生和选择。通过本领域的技术人员已知的任何方法，包括，例如，易错 PCR、基因重组、化学诱变等，可以产生变化。

术语“改组”在本文中用于指不相同的序列之间的重组。在有些实施方案中，改组可以包括通过同源重组或通过非同源重组的交换，例如通过 cre/lox 和/或 flp/frt 系统。通过采用许多种不同的格式，可以进行改组，所述格式包括，例如，体外和体内改组格式、计算机改组格式、利用双链或单链模板的改组格式、基于引物的改组格式、基于核酸断裂的改组格式和寡核苷酸-介导的改组格式，它们都是基于不相同的序列之间的重组事件，且在本文下面更详细地描述或参考，以及其它类似的基于重组的格式。如本文所使用的术语“随机的”指通过机遇的或随机的方法构建的由 2 或更多个氨基酸组成的多核苷酸序列或氨基酸序列。随机的多核苷酸序列或氨基酸序列可以包括框架或支架基序，后者可以包含不变的序列。

如本文所使用的术语“伪随机的”指一组序列、多核苷酸或多肽，其具有有限的可变性，所以在有些位置的残基可变性的程度是有限的，但是任意的伪随机的位置都允许至少一定程度的残基变化。

术语“多肽”、“肽”和“蛋白”在本文中可互换地使用，指 2 或更多个氨基酸的氨基酸序列。

术语“氨基酸”指天然发生的和合成的氨基酸，以及氨基酸类似物和氨基酸模仿物，其以与天然发生的氨基酸类似的方式起作用。天然发生的氨基酸是由遗传密码编码的那些，以及以后修饰的那些氨基酸，例如，羟脯氨酸、 γ -羧基谷氨酸和 O-磷酸丝氨酸。氨基酸类似物

指具有与天然发生的氨基酸相同的基本化学结构的化合物，即，结合氢、羧基、氨基和 R 基团的 α 碳，例如，高丝氨酸、正亮氨酸、甲硫氨酸亚砷、甲硫氨酸甲基砷。这样的类似物具有修饰的 R 基团(例如，正亮氨酸)或修饰的肽主链，但是保留与天然发生的氨基酸相同的基本化学结构。“氨基酸模仿物”指具有与氨基酸的一般化学结构不同的结构的化合物，但是其以与天然发生的氨基酸类似的方式起作用。

“保守氨基酸置换”指具有类似侧链的残基的互换性。例如，一组具有脂族侧链的氨基酸是甘氨酸、丙氨酸、缬氨酸、亮氨酸和异亮氨酸；一组具有脂族-羟基侧链的氨基酸是丝氨酸和苏氨酸；一组具有含有酰胺的侧链的氨基酸是天冬酰胺和谷氨酰胺；一组具有芳族侧链的氨基酸是苯丙氨酸、酪氨酸和色氨酸；一组具有碱性侧链的氨基酸是赖氨酸、精氨酸和组氨酸；且一组具有含硫侧链的氨基酸是半胱氨酸和甲硫氨酸。优选的保守氨基酸置换组是：缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸和天冬酰胺-谷氨酰胺。

短语“核酸序列”指从 5'向 3'端阅读的脱氧核糖核苷酸或核糖核苷酸碱基的单链或双链聚合物或其类似物。

术语“编码”指编码 1 个或多个氨基酸的多核苷酸序列。该术语不需要起始或终止密码子。在多核苷酸序列提供的 6 个不同的读框中的任一个都可以编码氨基酸序列。

术语“启动子”指位于转录起始的上游和/或下游的区域或序列，其参与 RNA 聚合酶和其它蛋白的识别和结合，以启动转录。

“载体”指多核苷酸，当独立于宿主染色体时，其能在宿主生物中复制。载体的实例包括质粒。载体一般地具有复制起点。载体可以包含，例如，转录和翻译终止子、转录和翻译起始序列和用于调节特定核酸的表达的启动子。

当提及例如细胞、或核酸、蛋白或载体时使用的术语“重组的”指，已经通过异源核酸或蛋白的导入或天然核酸或蛋白的改变修饰的细胞、核酸、蛋白或载体，或指该细胞源自这样修饰的细胞。因而，例如，重组的细胞表达在天然的(非重组的)形式的细胞内未发现的基因，或表达否则会异常表达、表达不足或根本不表达的天然基因。

当提及单体或多聚体时，短语“特异性地(或选择性地)结合”多

肽指，可以确定蛋白的异质群体(例如，细胞或组织裂解物)和其它生物制剂中多肽的存在的结合反应。因而，在抗体结合测定中使用的标准条件或测定下，指定的单体或多聚体会超过背景(例如，超过背景 2X、5X、10X 或更多倍)地结合特定靶分子，且不会以显著的量结合存在于样品中的其它分子。

在 2 个或更多个核酸或多肽序列的上下文中，术语“相同的”或百分比“同一性”指相同的 2 个或更多个序列或子序列。“基本上相同的”指，当在对比窗或指定区域中对比和比对最大对应时，如使用下面的序列对比算法之一或通过手工比对和目检所测得的，具有指定百分比的相同氨基酸残基或核苷酸(即，在指定区域具有 60%同一性，任选地 65%、70%、75%、80%、85%、90%或 95%同一性，或当未指定时，在整个序列中)的 2 个或更多个核酸或多肽序列。任选地，在至少约 50 核苷酸长的区域，或更优选地在 100-500 或 1000 或更多个核苷酸或氨基酸长的区域，存在同一性或相当大的同一性。

如果 2 个序列没有以与在天然发生的序列中发现的相同方式相连接，则多核苷酸或氨基酸序列与第 2 个序列是“异源的”。例如，可操作地连接到异源编码序列上的启动子指，与任何天然发生的等位基因变体不同的编码序列。当提及多聚体时使用的术语“异源接头”指，多聚体包含不具有天然发现的相同的彼此关系的接头和单体(例如，它们形成非天然发生的融合蛋白)。

蛋白序列中的“非天然发生的氨基酸”指，在与天然发生的多肽相比对的对应区中以最低的最小的总和概率发生的氨基酸以外的任何氨基酸，其中对比窗是查询的单体域的长度，且当使用如本文所述的 BLAST 2.0，与 Genbank 的非冗余的(“nr”)数据库中的天然发生的序列相对比时。

通过对比窗中对比 2 个最佳比对的序列，确定“序列同一性百分比”，其中与参照序列(其不包含添加或缺失)相对比，对于 2 个序列的最佳比对，对比窗中的多核苷酸序列部分可以包含添加或缺失(即，缺口)。如下计算百分比：通过确定在两个序列中出现相同的核酸碱基或氨基酸残基的位置的数目，以产生匹配的位置的数目，将匹配的位置的数目除以对比窗中的位置总数，并将结果乘以 100，以产生序列同一性百分比。

在 2 个或更多个核酸或多肽序列的上下文中，术语“相同的”或百分比“同一性”指，当在对比窗或指定区域中对比和比对最大对应时，如使用下面的序列对比算法之一或通过手工比对和目检所测得的，相同的或具有指定的相同氨基酸残基或核苷酸百分比的 2 个或更多个序列或子序列。于是，这样的序列称作“基本上相同的”。该定义也指测试序列的互补体。任选地，在至少约 50 个氨基酸或核苷酸长的区域，或更优选地在 75-100 个氨基酸或核苷酸长的区域，存在同一性。

为了序列对比，一般地将一个序列作为参照序列，与其对比测试序列。当使用序列对比算法时，将测试和参照序列输入计算机，如果必要，指定子序列坐标，并指定序列算法程序参数。可以使用默认的程序参数，或可以指定替代参数。然后，基于程序参数，序列对比算法计算测试序列相对于参照序列的百分比序列同一性。

如本文所使用的，“对比窗”包括选自 20 至 600、通常约 50 至约 200、更通常约 100 至约 150 的邻接位置数的任一个的片段，其中在最佳比对 2 个序列后，可以对比一个序列和具有相同邻接位置数的参照序列。进行对比的序列比对方法，是本领域众所周知的。可以进行用于对比的最佳序列比对，例如，通过 Smith 和 Waterman (1970) *Adv. Appl. Math.* 2: 482c 的局部同源性算法，通过 Needleman 和 Wunsch (1970) *J. Mol. Biol.* 48: 443 的同源性比对算法，通过 Pearson 和 Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85: 2444 的搜索相似性方法，通过这些算法的计算机实现 (Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis. 中的 GAP、BESTFIT、FASTA 和 TFASTA)，或通过手工比对和目检 (见，例如，Ausubel 等，*Current Protocols in Molecular Biology* (1995 增刊))。

有用的算法的一个实例是 BLAST 2.0 算法，其分别记载在 Altschul 等 (1990) *J. Mol. Biol.* 215: 403-410。从国家生物技术信息中心 (National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov/>)，可以公开地得到用于执行 BLAST 分析的软件。该算法包含，首先通过鉴别查询序列中长为 W 的短字，鉴别高评分序列对 (HSP)，当与数据库序列中的相同长度的字相比对时，

所述 HSP 匹配或满足一些正值的阈分数 T 。 T 称作邻近字分数阈值 (Altschul 等, 同上)。这些原始的邻近字命中值(hit)作为启动搜索的种子, 以发现更长的含有它们的 HSP。字命中值沿着每个序列向两个方向延伸, 只要可以增加累积比对分数。对于核苷酸序列, 使用参数 M (一对匹配残基的奖励分数; 总是 >0) 和 N (错配残基的惩罚分数; 总是 <0), 计算累积分数。对于氨基酸序列, 使用评分矩阵计算累积分数。当下述情况时, 停止每个方向的字命中值的延伸: 累积比对分数从它的最大实现值下降了量 X ; 由于一个或多个负评分残基比对的累积, 累积分数达到 0 或以下; 或达到任一个序列的末端。BLAST 算法参数 W 、 T 和 X 决定了比对的灵敏度和速度。BLASTN 程序(对于核苷酸序列)使用的默认值是, 字长(W)为 11, 期望(E)为 10, $M=5$, $N=-4$, 且对比 2 条链。对于氨基酸序列, BLASTP 程序使用的默认值是, 字长为 3, 期望(E)为 10, 且 BLOSUM62 评分矩阵 (见 Henikoff 和 Henikoff (1989) Proc. Natl. Acad. Sci. USA 89: 10915) 比对(B)为 50, 期望(E)为 10, $M=5$, $N=-4$, 且对比 2 条链。

BLAST 算法也执行 2 个序列之间的相似性统计学分析(见, 例如, Karlin 和 Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-5787)。BLAST 算法提供的相似性的一种度量是最小总和概率 ($P(N)$), 其指示着 2 个核苷酸或氨基酸序列可随机发生匹配的可能性。例如, 如果测试核酸与参照核酸的对比中的最小总和概率小于约 0.2, 更优选地小于约 0.01, 且最优选地小于约 0.001, 则认为核酸与参照序列类似。

附图简述

图 1 示意地解释了来自许多种 LDL-受体 A 类-域的部分氨基酸序列的比对, 以证实保守的半胱氨酸。在共有序列上, 示意地解释了折叠的域的 3 个二硫键中的半胱氨酸的连接性。在共有序列中, 用星号指示其侧链有助于钙结合的残基。

图 2, 小图 A 示意地解释了 A-域的实例。小图 A 示意地解释了约 40 个氨基酸长的 A-域的保守氨基酸。C 指示保守的半胱氨酸残基, 且带有减号("-")的圆圈指示保守的带负电荷的氨基酸。带有“H”的圆圈指示保守的疏水残基。小图 B 示意地解释了通过接头相连的 2

个折叠的 A-域。小图 B 也指示了 2 个钙结合位点, 带有 Ca^{+2} 的黑圈, 和在每个折叠的 A-域中的 3 个二硫键, 共 6 个二硫键。

图 3 指示了由 LDL-受体家族的天然发生的成员识别的一些配体, 其包括抑制剂、蛋白酶、蛋白酶复合物、维生素-载体复合物、参与脂蛋白代谢的蛋白、非-人配体、抗生素、病毒和其它的。

图 4 示意地解释了一般方案, 其用于鉴别结合配体的单体域, 分离选择的单体域, 通过以各种组合连接选择的单体域, 产生选择的单体域的多聚体, 和筛选多聚体, 以鉴别包含超过 1 个结合配体的单体的多聚体。

图 5 是另一个选择策略(引导的选择)的示意图。从单体域文库中, 鉴别出了具有适当的结合性质的单体域。然后, 将鉴别的单体域连接到来自另一个单体域文库的单体域上, 以形成多聚体文库。筛选多聚体文库, 以鉴别同时结合靶的一对单体域。然后, 可以重复该过程, 直到在多聚体中得到最佳的结合性质。

图 6 描述了 A 域的比对。在图的顶部和底部, 小写字母(a-q)指示着保守的残基。

图 7 解释了各种可能的抗体-单体或多聚体构象。在有些实施方案中, 单体或多聚体替代抗体的 Fab 片段。

图 8 描述了本发明的多聚体可能的构象, 其包含至少一个结合半衰期延长性分子的单体域和结合 1 个或任选地 2 个或更多个靶分子的其它单体域。在图中, 2 个单体域结合 2 个第一靶分子。任选地, 2 个单体域可以结合 1 个第一靶分子上的不同位点(未画出)。

图 9 显示了 c-METFc、c-MET-特异性的单体(M26)和 c-MET-特异性的二聚体(RM12; RecM12)之间的关于它们阻断 HGF-诱导的血清-饥饿的 A549-SC 人肺腺癌细胞的增殖的相对能力的对比。

图 10 解释了猴子中结合 IgG 的单体的血清半衰期。

发明详述

I. 引言

本发明提供了结合 c-MET 的非天然发生的蛋白。一般地, 本发明的蛋白包含结合 c-MET 的域。可以如下容易地鉴别这些域: 使用许多种多肽支架, 以产生许多多肽变体, 然后选择结合 c-MET 的变

体。因此，本发明也提供了选择结合 c-MET 的蛋白。结合 c-MET 的蛋白可以用于，例如，治疗具有表达 c-MET 的实体瘤的个体。本发明的多肽也可以用于检测其中表达 Met 的组织，且可以用于将分子靶向那些组织。

c-MET 在它的静止单体状态是无活性的，且二聚体形成导致受体激活(经常甚至在没有配体结合的情况下)。成熟形式的受体由单独的胞外 α 链和更长的 β 链组成，所述 β 链包含胞外域的剩余部分、跨膜域和胞质尾。胞质尾含有近膜域、激酶域和信号传导中间体的停靠位点。 α 链和 β 链的前 212 个氨基酸，也称作 Sema 域 (Kong-Beltran, 等, *Cancer Cell* 6: 75-84 (2004)，对于与 HGF 的结合是足够的。 β 链的其它胞外部分由富含半胱氨酸的 C 域和罕见的免疫球蛋白域的 4 个重复序列组成。因此，在有些实施方案中，本发明的多肽包含至少一个单体域，其抑制 c-MET α 和 β 链的二聚化，和/或起拮抗剂的作用，以阻止 c-MET 的配体结合和/或激活 c-MET。

尽管本发明提供了包含单个域的多肽，但也可以合成和使用域的多聚体。在有些实施方案中，多聚体所有域都结合 c-MET。在有些这样的实施方案中，每个域是相同的，且结合 c-MET 的相同部分(即，“表位”)。例如，在有些实施方案中，单体域结合 c-MET 的 Sema 域。在其它实施方案中，至少多聚体中的有些域结合 c-MET 的不同部分。在其它实施方案中，至少多肽的有些域结合除 c-MET 以外的一种或多种分子(例如，血液因子，例如血清清蛋白、免疫球蛋白或红细胞)。

II. 单体

单体域可以是任意大小的多肽链。在有些实施方案中，单体域具有约 25 至约 500、约 30 至约 200、约 30 至约 100、约 35 至约 50、约 35 至约 100、约 90 至约 200、约 30 至约 250、约 30 至约 60、约 9 至约 150、约 100 至约 150、约 25 至约 50 或约 30 至约 150 个氨基酸。类似地，本发明的单体域可以包含，例如，约 30 至约 200 个氨基酸；约 25 至约 180 个氨基酸；约 40 至约 150 个氨基酸；约 50 至约 130 个氨基酸；或约 75 至约 125 个氨基酸。单体域可以一般地维持稳定的溶液构象，且经常是热稳定的，例如，在 95°C 稳定至少 10 分钟，

而不丧失结合亲和力。有时，单体域可以独立地折叠成稳定的构象。在一个实施方案中，稳定的构象是由离子（例如，金属或钙离子）稳定的。稳定的构象可以任选地含有二硫键（例如，至少 1、2 或 3 或更多个二硫键）。二硫键可以任选地在 2 个半胱氨酸残基之间形成。在有些实施方案中，单体域或单体域变体，与示例的序列基本上相同。

A. c-MET 粘合剂

在有些方面，本发明提供了结合 c-MET 多肽或其部分的单体域。多肽的部分可以是，例如，多肽的至少 5、10、15、20、30、50、100 或更多个邻接氨基酸。

产生了大量具有 A 域支架的 c-MET 结合序列。如在实施例 1 中详细描述，已经鉴别出了结合 c-MET 的单体域的 10 个家族（即，家族 1-10，或“Fam 1-10”）。基于这些家族产生的共有基序指示着 c-MET 粘合剂之间的共有氨基酸残基。从基序中省略侧接构成该基序的保守残基的序列，尽管假定构成 A-域结构的所有残基存在于基于下面的家族的任何结合域中。本领域的技术人员会明白，非共有（用“X”标记）的位置可以是任意的氨基酸。在有些实施方案中，从示例的来自相同家族或不同家族的 c-MET 粘合剂之一的类似位置处的氨基酸中，选择在“X”位置处的氨基酸。

家族 1 具有下面的共有基序：

Cxxx [EQ] FxCxSTxRC [IV] xxxWxCDGDNDCEDxSDEx

在实施例 1 中，显示了示例性的包含 c-MET 家族 1 基序的序列。c-MET 结合单体或多聚体包括在实施例 1 中示例的每个家族 1 序列。

家族 2 具有下面的基序：

Cxxxx [EQ] FECxSTxRC [IV] xxxWxCDGxNDCEDxSDEx

在实施例 1 中，显示了示例性的包含 c-MET 家族 2 基序的序列。c-MET 结合单体或多聚体包括在实施例 1 中示例的每个家族 2 序列。

家族 3 具有下面的基序：

Cxxxx [EQ] FxCxSTxRC [ILV] PxxWxCDGxxDCEDxSDExx

在实施例 1 中，显示了示例性的包含 c-MET 家族 3 基序的序列。c-MET 结合单体或多聚体包括在实施例 1 中示例的每个家族 3 序列。

家族 4 具有下面的基序：

Cxxx[EQ]FQCxSTxRC[IV]PxxWxCDGxNDCEDSSDExxC

在实施例中，显示了示例性的包含 c-MET 家族 4 基序的序列。
c-MET 结合单体或多聚体包括在实施例中示例的每个家族 4 序列。

家族 5 具有下面的基序：

Cxxxx[EQ]FxCxxxxxC[ILV]xxxxxxxxxxDCxDxSDEx

在实施例中，显示了示例性的包含 c-MET 家族 5 基序的序列。
c-MET 结合单体或多聚体包括在实施例中示例的每个家族 5 序列。

家族 6 具有下面的基序：

Cxxx[EQ]FxCxSTGRCxPxxWxCxGxNDCEDxSDEx

在实施例中，显示了示例性的包含 c-MET 家族 6 基序的序列。
c-MET 结合单体或多聚体包括在实施例中示例的每个家族 6 序列。

家族 7 具有下面的基序：

Cxxxx[EQ]FxCxSTxRC[ILV]xxxWxCxxxxDCxDxSDxxxxxCx

在实施例中，显示了示例性的包含 c-MET 家族 7 基序的序列。
c-MET 结合单体或多聚体包括在实施例中示例的每个家族 7 序列。

家族 8 具有下面的基序：

Cxxx[EQ]FxCxxxxxC[ILV]xxxWxCDGxNDCxDxSxExxxxC

在实施例中，显示了示例性的包含 c-MET 家族 8 基序的序列。
c-MET 结合单体或多聚体包括在实施例中示例的每个家族 8 序列。

家族 9 具有下面的基序：

Cxxxx[EQ]FxCxSTxRC[ILV]PxxWxCxGxxDCxDxSDEx

在实施例中，显示了示例性的包含 c-MET 家族 9 基序的序列。
c-MET 结合单体或多聚体包括在实施例中示例的每个家族 9 序列。

家族 10 具有下面的基序：

Cxxxx[EQ]FxCxxxxxC[ILV]xxxWxCDGxxDCxDxSDEx

其可以进一步缩写成：

EFXCXNGXCIPXXWXCDGXDDCGDXSDE

在实施例中，显示了示例性的包含 c-MET 家族 10 基序的序列。
c-MET 结合单体或多聚体包括在实施例中示例的每个家族 10 序列。

B. IgG 粘合剂和血清半衰期延长

本发明还提供了结合血液因子 (例如, 血清清蛋白、免疫球蛋白或红细胞) 的单体域。

在有些实施方案中, 单体域结合免疫球蛋白多肽或其一部分。

已经鉴别了结合免疫球蛋白的单体域的 2 个家族(即, A 域家族 2 和 3)。

家族 2 具有下面的基序:

[EQ]FXCRX[ST]XRC[IV]XXXW[ILV]CDGXXDCXD[DN]SDE

在实施例中, 显示了示例性的包含 IgG 家族 2 基序的序列。IgG 结合单体或多聚体包括在实施例中示例的每个家族 2 序列。

家族 3 具有下面的 2 个基序的任一个:

CXSSGRCIPXXWVCDGXXDCRDXSDE; 或

CXSSGRCIPXXWLCDGXXDCRDXSDE

在实施例中, 显示了示例性的包含 IgG 家族 3 基序的序列。IgG 结合单体或多聚体包括在实施例中示例的每个家族 3 序列。

在美国专利公布号 2005/0048512 中, 描述了结合红细胞(RBC)或血清清蛋白(CSA)的单体域, 且包括, 例如:

RBCA CRSSQFQCNSRIPCIPGRWRCDGDNDCCQDGSDETGCGDSHILPFSTPGPST
RBCB CPAGEFPCKNGQCLPVTWLCDGVNDCLDGSDEKGCGRPGPGATSAPAA
RBC11 CPPDEFPCCKNGQCI PQDWLCDGVNDCLDGSDEKDCGRPGPGATSAPAA
CSA-A8 CGAGQFPCKNGHCLPLNLLCDGVND CEDNSDEPSELCKALT

本发明提供了延长动物中的蛋白的血清半衰期的方法, 所述蛋白包括, 例如, 本发明的多聚体或目标蛋白。目标蛋白可以是具有治疗、预防或其它希望的功能性的任何蛋白。该方法包含, 首先提供已经被鉴别为特异性地结合半衰期延长剂的结合蛋白的单体域, 所述的半衰期延长剂例如血液-携带的分子或细胞, 例如血清清蛋白(例如, 人血清清蛋白)、IgG、红细胞等。然后, 将半衰期延长剂-结合单体共价地连接到具有对目标蛋白(例如, c-MET 或不同的靶)的结合亲和力的另一个单体域上。该复合物形成会导致半衰期延长, 从而保护多聚体和/或结合的蛋白免受蛋白水解性降解和/或多聚体和/或蛋白的其它去除, 并从而延长蛋白和/或多聚体的半衰期。本发明的该用途的一个变化包括, 将半衰期延长剂-结合单体共价地连接到目标蛋白上。目标蛋白可以包括单体域、单体域的多聚体或合成的药物。或者, 使

用上述的方法，可以产生结合免疫球蛋白或红细胞的单体，且可以将其用于半衰期延长。

半衰期延长剂-结合多聚体一般地是至少 2 个域、嵌合域或诱变的域（即，一个结合 Met 而一个结合血液-携带的分子或细胞）的多聚体。合适的域包括本文所述的所有域，进一步筛选和选择其对半衰期延长剂的结合。根据制备本文所述的多聚体的方法，使用例如预筛选了半衰期延长剂-结合活性的单体域，产生半衰期延长剂-结合多聚体。分子的血清半衰期可以延长至，例如，至少 1、2、3、4、5、10、20、30、40、50、60、70 80、90、100、150、200、250、400、500 或更多个小时。

C. 单体域的讨论

特别适用于实践本发明的单体域是包含二硫键的富含半胱氨酸的域。在本发明的实践中采用的富含半胱氨酸的域一般地不会形成 α 螺旋、 β 折叠或 β -桶结构。一般地，二硫键会促进域折叠成三维结构。通常，富含半胱氨酸的域具有至少 2 个二硫键，更一般地至少 3 个二硫键。在有些实施方案中，单体域中的至少 5、10、15 或 20% 氨基酸是半胱氨酸。

域可以具有任意数目的特征。例如，在有些实施方案中，域在动物（例如，人）中具有较低的免疫原性或没有。域可以具有小的尺寸。在有些实施方案中，域足够小，以致于能穿透皮肤或其它组织。域可以具有一定范围的体内半衰期或稳定性。

示例性的适用于实践本发明的单体域包括，例如，EGF-样域、三环域、纤连蛋白 I 型域、纤连蛋白 II 型域、纤连蛋白 III 型域、PAN 域、Gla 域、SRCR 域、Kunitz/Bovine 胰腺胰蛋白酶抑制剂域、Kazal-型丝氨酸蛋白酶抑制剂域、Trefoil(P-型)域、冯维勒布兰德因子 C 型域、过敏毒素-样域、CUB 域、甲状腺球蛋白 I 型重复序列、LDL-受体 A 类域、Sushi 域、Link 域、血小板反应蛋白 I 型域、免疫球蛋白-样域、C-型凝集素域、MAM 域、冯维勒布兰德因子 A 型域、生长调节素 B 域、WAP-型四二硫键核心域、F5/8C 型域、血红素结合蛋白域、SH2 域、SH3 域、层粘连蛋白-型 EGF 域、C2 域和本领域的普通技术人员已知的其它这样的域及其衍生物和/或变体。

在有些实施方案中，可以从含有通过 Simple Modular Architecture Research Tool (SMART) 见 Shultz 等, SMART: a web-based tool for the study of genetically mobile domains, (2000) Nucleic Acids Research 28(1): 231-234)或 CATH (见 Pearl 等, Assigning genomic sequence to CATH, (2000) Nucleic Acids Research 28(1): 277-282)等计算序列分析工具定义的 β -夹心或 β -桶三维结构的蛋白域家族中，选择合适的单体域(例如具有独立地折叠的能力的域，或具有一些有限的帮助的域)。

在另一个实施方案中，本发明的单体域包括除纤连蛋白 III 型域、anticalin 域和来自 CTLA-4 的 Ig-样域以外的域。在 2001 年 9 月 7 日公开的 Lipovsek 等的标题为“Protein scaffolds for antibody mimics and other binding proteins”的 WO 01/64942 中，1999 年 4 月 8 日公开的 Beste 等的标题为“Anticalins”的 WO 99/16873 中，和 2000 年 10 月 12 日公开的 Desmet 等的标题为“A polypeptide structure for use as a scaffold”的 WO 00/60070 中，描述了这些域的一些方面。

如上所述，单体域任选地是富含半胱氨酸的。合适的富含半胱氨酸的单体域包括，例如，LDL 受体 A 类域 (“A-域”)或 EGF 域。单体域也可以具有一簇带负电荷的残基。

单体域的其他特征可以包括，结合配体的能力或结合离子的能力(例如，LDL 受体 A-域对 Ca^{2+} 的结合)。结合离子来维持它们的二级结构的单体域包括，例如，A 域、EGF 域、EF 手(例如，发现存在于钙调蛋白和肌钙蛋白 C 中的那些)、钙粘着蛋白域、C-型凝集素、C2 域、膜联蛋白、Gla-域、Trombospondin 3 型域，它们都结合钙，和锌指(例如，C2H2 型 C3HC4 型(RING 指)、整合酶锌结合域、PHD 指、GATA 锌指、FYVE 锌指、B-框锌指)，其结合锌。无意限制本发明，认为离子-结合提供了二级结构的稳定性，同时提供足够的柔性，以允许许多依赖于一级序列的结合构象。

如本文所述，可以选择单体域结合靶的能力，所述靶是除同源的天然发生的域可以结合的靶以外的靶。因而，在有些实施方案中，本发明提供了不结合基本上相同的天然发生的域可以结合的靶或靶蛋白类或家族的单体域(和包含这样的单体的多聚体)。

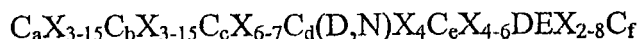
单体域的特征可以包括独立地折叠的能力和形成稳定的结构的

能力。因而，单体域的结构经常是保守的，尽管编码单体的多核苷酸序列无需是保守的。例如，在 A-域家族的成员中，A-域结构是保守的，而 A-域核酸序列不是。因而，例如，通过它的半胱氨酸残基和它对钙的亲和力，不一定通过它的核酸序列，将单体域归入 A-域。见，图 1 和 2。

具体地，A-域(有时称作“补体-型重复”或“LDL 受体 A 型或类域”)含有约 30-50 或 30-65 个氨基酸。在有些实施方案中，该域包含约 35-45 个氨基酸，且在有些情况下，约 40 个氨基酸。在 30-50 个氨基酸中，有约 6 个半胱氨酸残基。在 6 个半胱氨酸中，二硫键一般地出现在下面的半胱氨酸之间：C1 和 C3、C2 和 C5、C4 和 C6。该域的半胱氨酸残基被二硫键连接，以形成紧凑的、稳定的、功能上独立的部分。见，图 3。这些重复序列的簇形成配体结合域，且有差别的成簇可以影响关于配体结合的特异性。

在图 1 和 2 中描述了示例性的 A 域序列和共有序列。用于鉴别 A 域的一个一般的共有序列如下：C-[VILMA]-X₍₅₎-C-[DNH]-X₍₃₎-[DENQHT]-C-X_(3,4)-[STADE]-[DEH]-[DE]-X_(1,5)-C，其中在方括号中的残基指示着在一个位置处的可能的残基。“X_(#)”指示着残基的数目。这些残基可以是任意的氨基酸残基。含有 2 个数字的附加说明指可以占据该位置的氨基酸的范围(例如，“[DE]-X_(1,5)-C”指，氨基酸 DE 后面是 1、2、3、4 或 5 个残基，然后是 C)。该共有序列仅代表着从第 3 个半胱氨酸开始的 A 域的部分。第 2 个共有序列如下：C-X₍₃₋₁₅₎-C-X₍₄₋₁₅₎-C-X₍₆₋₇₎-C-[N,D]-X₍₃₎-[D,E,N,Q,H,S,T]-C-X₍₄₋₆₎-D-E-X₍₂₋₈₎-C。第 2 个共有序列预测了跨所有 6 个半胱氨酸残基的氨基酸残基。在有些实施方案中，A 域变体包含与任一种上述的序列基本上相同的序列。应当指出，为了本发明的目的，“LDL 受体 A 类”域无意指该域的来源或结合性质。

其它示例性的 A 域包括下面的序列：



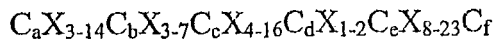
其中 C 是半胱氨酸，X_{n-m} 代表着 n 和 m 之间数目的独立选择的氨基酸，且(D,N)指该位置可以是 D 或 N；且其中 C_a-C_c、C_b-C_e 和 C_d-C_f 形成二硫键。

迄今为止，基于 cDNA 序列，鉴别了至少 190 个天然发生的人

A-域。见，例如，图 6。示例性的含有天然发生的 A-域的蛋白包括，例如，补体组分(例如，C6、C7、C8、C9 和因子 I)、丝氨酸蛋白酶(例如，肠肽酶、matriptase 和 corin)、跨膜蛋白(例如，ST7、LRP3、LRP5 和 LRP6)和内吞受体(例如，Sortilin-相关的受体、LDL-受体、VLDLR、LRP1、LRP2 和 ApoER2)。在本发明的实践中，可以容易地采用作为单体域和其变体的 A 域和 A 域变体。在下面的出版物及其中引用的文献中，可以发现 A 域的有关描述：Howell 和 Hertz, *The LDL receptor gene family: signaling functions during development*, (2001) *Current Opinion in Neurobiology*. 11: 74-81; Herz (2001), 同上；Krieger, *The “best” of cholesterols, the “worst” of cholesterols: A tale of two receptors*, (1998) *PNAS* 95: 4077-4080; Goldstein 和 Brown, *The Cholesterol Quartet*, (2001) *Science*. 292: 1310-1312; 和, Moestrup 和 Verroust, *Megalin-and Cubilin-Mediated Endocytosis of Protein-Bound Vitamins, Lipids, and Hormones in Polarized Epithelia*, (2001) *Ann. Rev. Nutr.* 21: 407-28。

许多其它的域类型也可以用于产生 c-MET-结合单体域。

示例性的 EGF 单体域包括序列：



其中 C 是半胱氨酸， X_{n-m} 代表着 n 和 m 之间数目的独立选择的氨基酸；且

其中 C_a-C_c 、 C_b-C_e 和 C_d-C_f 形成二硫键。

下述的每个域采用示例性的基序(即，支架)。某些位置标有 x，从而表示任意的氨基酸可以占据该位置。这些位置可以包括许多不同的氨基酸可能性，从而允许序列多样性和因而对不同靶分子的亲和力。在基序中使用方括号，表示在该位置中的可能的可替代氨基酸(例如，“[ekq]”表示 E、K 或 Q 可以在该位置)。在基序中使用圆括号，表示在该圆括号中的位置可能存在或不存在(例如，“([ekq])”表示该位置不存在，或 E、K 或 Q 可以在该位置)。当在圆括号中使用超过 1 个“x”时(例如，“(xx)”)，每个 x 代表着一个可能的位置。因而，“(xx)”表示在该位置可能存在 0、1 或 2 个氨基酸，其中每个氨基酸独立地选自任意的氨基酸。 α 代表着芳族的/疏水的氨基酸，例如，W、Y、F 或 L； β 代表着疏水的氨基酸，例如，V、I、L、A、M 或 F； χ 代

表着小的或极性的氨基酸，例如，G、A、S或T； δ 代表着带电荷的氨基酸，例如，K、R、E、Q或D； ϵ 代表着小氨基酸，例如，V、A、S或T；且 ϕ 代表着带负电荷的氨基酸，例如，D、E或N。

合适的域包括，例如血小板反应蛋白 I 型域、trefoil 域和甲状腺球蛋白域。

血小板反应蛋白 1 型 (“TSP1”) 域含有约 30-50 或 30-65 个氨基酸。在有些实施方案中，该域包含约 35-55 个氨基酸，且在有些情况下，约 50 个氨基酸。在 35-55 个氨基酸中，一般地有约 4 至约 6 个半胱氨酸残基。在 6 个半胱氨酸中，二硫键一般地出现在下面的半胱氨酸之间：C1 和 C5、C2 和 C6、C3 和 C4。该域的半胱氨酸残基被二硫键连接，以形成紧凑的、稳定的、功能上独立的部分，其包含扭曲的 β 链。这些重复序列的簇构成配体结合结构域，且有差别的成簇可以影响关于配体结合的特异性。

示例性的 TSP1 域序列和共有序列如下：

- (1) (xxxxxx)C₁xxx C₂xxxxx(x)xxxxx C₃xxxx(xxx)xxxxx C₄xxxxxx(x)xxx C₅(x)xxxx C₆;
- (2) (wxxWxx)C₁xxx C₂xx Gxx(x)xRxxx C₃xxxx(Pxx)xxxxx C₄xxxxxx(x)xxx C₅(x)xxxx C₆
- (3) (wxxWxx)C₁sxt C₂xx Gxx(x)xRxx C₃xxxx(Pxx)xxxxx C₄xxxxxx(x)xxx C₅(x)xxxx C₆
- (4)
(WxxWxx)C₁[Stnd][Vkaq][Tspl]C₂xx[Gq]xx(x)x[Re]x[Rktvm]x[C₃vldr]xxxx([Pq]xx)xxxxx[C₄ldae]xxxxxx(x)xxx C₅(x)xxxx C₆;
- (5)
(WxxWxx)C₁[Stnd][Vkaq][Tspl]C₂xx[Gq]xx(x)x[Re]x[Rktvm]x[C₃vldr]xxxx([Pq]xx)xxxxx[C₄ldae]xxxxxx(x)xxx C₅(x)xxxx C₆; 和
- (6)
C₁[nst][aegiklqrstv][adenpqrst]C₂[adetgs]xgx[ikqrstv]x[aqrst]x[almrtv]x C₃xxxxxxxxxx(xxxxxx)C₄xxxxxxxxxx(xx)C₅xxxx C₆

在有些实施方案中，血小板反应蛋白 1 型域变体包含与任一种上述序列基本上相同的序列。

迄今为止，基于 cDNA 序列，已经鉴别了至少 1677 个天然发生的血小板反应蛋白域。示例性的含有天然发生的血小板反应蛋白域的蛋白包括，例如，在补体途径中的蛋白(例如，备解素、C6、C7、C8A、C8B 和 C9)、胞外基质蛋白(例如，mindin、F-spondin、SCO-spondin)、环孢子表面蛋白 2 和疟虫属(*Plasmodium*)的 TRAP 蛋白。血小板反应蛋白 1 型域进一步记载在，例如，Roszmusz 等，BBRC 296: 156 (2002); Higgins 等，J Immunol. 155: 5777-85 (1995); Schultz-Cherry 等，J. Biol. Chem. 270: 7304-7310 (1995); Schultz-Cherry 等，J. Biol.

Chem. 269: 26783-8 (1994); Bork, FEBS Lett 327: 125-30 (1993);
和 Leung-Hagesteijn 等, Cell 71: 289-99 (1992)。

另一种示例性的适用于实践本发明的单体域是 trefoil 域。trefoil 单体域一般地是约 30-50 或 30-65 个氨基酸。在有些实施方案中, 该域包含约 35-55 个氨基酸, 且在有些情况下, 约 45 个氨基酸。在 35-55 个氨基酸中, 一般地存在约 6 个半胱氨酸残基。在 6 个半胱氨酸中, 二硫键一般地存在于下面的半胱氨酸之间: C1 和 C5、C2 和 C4、C3 和 C6。

迄今为止, 基于 cDNA 序列, 已经鉴别了至少 149 个天然发生的 trefoil 域。示例性的含有天然发生的 trefoil 域蛋白包括, 例如, 蛋白 pS2 (TFF1)、解痉肽 SP (TFF2)、肠 trefoil 因子 (TFF3)、肠 surcease-异麦芽糖酶和可能通过保护上皮参与抗微生物感染的防御的蛋白 (例如, 非洲爪蟾属 (*Xenopus*) xP1、xP4、皮肤粘蛋白 A.1 和 C.1。trefoil 域进一步记载在, 例如, Sands 和 Podolsky, *Annu. Rev. Physiol.* 58: 253-273 (1996); Carr 等, *PNAS USA* 91: 2206-2210 (1994); DeA 等, *PNAS USA* 91: 1084-1088 (1994); Hoffman 等, *Trends Biochem Sci* 18: 239-243 (1993)。

示例性的 trefoil 域序列和共有序列如下:

- (1) C₁(xx)xxxxxxxxC₂xx(x)xxxxxxxxC₃xxxxC₄C₅xxxx(x)xxxxC₆
- (2) C₁(xx)xxxxRxxC₂xx(x)xxxxxxxxC₃xxxxC₄C₅xxxx(x)xxxxC₆
- (3) C₁(xx)xxxpxRxnC₂gx(x)pxitxxxC₃xxxgC₄C₅fdxxx(x)xxxpwC₆f
- (4) C₁(xx)xxx[Pvae]xxRx[ndpm]C₂[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[qedk]C₃xx[krln][Gnk]C₄C₅[Fwy][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C₆[Fy]
- (5) C₁(xx)xxx[Pvae]xxRx[ndpm]C₂[Gaiy][ypfst]([de]x)[pskq]x[Ivap][Tsa]xx[keqd]C₃xx[krln][Gnk]C₄C₅[α][Dnrs][sdpnte]xx(x)xxx[pki][Weash]C₆[Fy]
- (6) C₁([dnps])[adiklnprstv][dfilmv][adenprst][adelprv][ehklnqrs][adegknsv][kqr][filqrtv][dnpps]C₂[agiy][flpsvy][dknpqs][adfglhp][aipv][st][aegkpqs][adegkpqs][deiknqt]C₃[adefknqrt][adegknqs][gn]C₄C₅[wyfh][deinrs][adgnpst][aefgqlrstw][giknsvmq]([afmprstv][degklns][afiqstv][iknpv]w)C₆

另一种示例性的适用于本发明的单体域是甲状腺球蛋白域。甲状腺球蛋白单体域一般地是约 30-85 或 30-80 个氨基酸。在有些实施方案中, 该域包含约 35-75 个氨基酸, 且在有些情况下, 约 65 个氨基酸。在 35-75 个氨基酸种, 一般地存在约 6 个半胱氨酸残基。在 6 个半胱氨酸中, 二硫键一般地存在于下面的半胱氨酸之间: C1 和 C2、

C3 和 C4、C5 和 C6。

迄今为止，基于 cDNA 序列，已经鉴别了至少 251 个天然发生的甲状腺球蛋白域。Tg 的 N-末端部分含有约 65 个氨基酸的域的 10 个重复，其称作 Tg 1 型重复 PUBMED: 3595599、PUBMED: 8797845。示例性的含有天然发生的甲状腺球蛋白域的蛋白包括，例如，HLA II 类相关的不变链、人胰腺癌标记蛋白、巢蛋白(触觉蛋白)、胰岛素-样生长因子结合蛋白(IGFBP)、saxiphilin、大麻哈鱼卵半胱氨酸蛋白酶抑制剂和 equistatin。Thyr-1 和相关的域属于 MEROPS 蛋白酶抑制剂家族 I31，异种集团 IX。甲状腺球蛋白域进一步记载在，例如，Molina 等, Eur. J Biochem. 240: 125-133 (1996); Guncar 等, EMBO J 18: 793-803 (1999); Chong 和 Speicher, DW 276: 5804-5813 (2001)。

示例性的甲状腺球蛋白域序列和共有序列如下：

- (1) C₁XXXXXXXXXXXXXXXX(XXXXXXXXXXX)XXXXXXXXXXXXC₂XXXXXXXXXXXXC₃X(X)X(XXX)XXXXC₄X
C₅XXXX(X)XXXXXXXXXXXXXXXX(XX)XC₆
- (2) C₁XXXXXXXXXXXXXXXX(XXXXXXXXXXX)XXXXXXXXXXYPXC₂XXXGXXXXXC₃X(X)X(XXX)XXXXC₄
WC₅VXXX(X)GXXXXGXXXXXXXX(XX)XC₆
- (3) C₁XXXXXXXXXXXXXXXX(XXXXXXXXXXX)XXXXXXXXXXYPXC₂XXXGXYXXXQC₃X(X)S(XXX)XXGX C₄WC₅
Vdxx(X)GXXXXGXXXXgxx(XX)XC₆
- (4) C₁[qerl]XXXXXXXXXXXXXXXX(XXXXXXXXXXX)XXXXXXXX[Yfhp]XPXC₂XXXGX[Yf]XX[vkrl]QC₃X(X)[S
a]XXX)XX[Gsa]XC₄[WYf]C₅V[Dnyfl]XX(X)GXXXX[Gdne]XXXXgxx(XX)XC₆
- (5) C₁[qerl]XXXXXXXXXXXXXXXX(XXXXXXXXXXX)XXXXXXXX[αhp]XPXC₂XXXGX[α]XX[vkrl]QC₃X(X)[sa]
XXX)XX[gas]XC₄[α]C₅V[Dnα]XX(X)GXXXX[φg]XXXXgxx(XX)XC₆

另一种可以用于本发明中的示例性的单体域是层粘连蛋白-EGF 域。层粘连蛋白-EGF 域一般地是约 30-85 或 30-80 个氨基酸。在有些实施方案中，该域包含约 45-65 个氨基酸，且在有些情况下，约 50 个氨基酸。在 45-65 个氨基酸中，一般地存在约 8 个半胱氨酸残基，其相互作用，以形成 4 个二硫键。层粘连蛋白是基膜的主要非胶原性组分，其介导细胞粘附、生长迁移和分化。它们由不同的、但是相关的 α、β 和 γ 链组成。这 3 条链形成由一条长臂和 3 条短球形臂组成的十字状分子。长臂由所有 3 条链促成的并通过链间二硫键交联的卷曲螺旋结构组成。

示例性的层粘连蛋白 EGF 域序列和共有序列如下：

- (1)
 $C_1x C_2xxxxxxx(xxx)xx C_3xxx(xxxxxx)xxxx C_4x C_5xxxxxxxx C_6xx C_7xxxxxxxx(xxxxx)xxx$
 $xx C_8$
- (2)
 $C_1x C_2xxxxxxx(xxx)xx C_3xxx(xxxxxx)xx gx C_4x C_5xxxxx Gxx C_6xx C_7xxxxxxxx(xxxxx)xx$
 $xxx C_8$
- (3)
 $C_1x C_2[ndh]xxxxx(xxx)xx C_3xxx(xxxxxx)xx gx C_4x C_5xxxxx Gxx C_6[denq]x C_7xx[gn][yf$
 $ht]xxx(xxxxx)xxxxx C_8$

在有些实施方案中，单体域是 Notch/LNR 单体域、DSL 单体域、Anato 单体域、整联蛋白 β 单体域和 Ca-EGF 单体域。

在有些实施方案中，Ca-EGF 单体域包含下面的序列：

$DxdEC_1xx(xx)xxxx C_2x(xx)xxxx C_3xNxxGxfx C_4x(xxx)x C_5xxgxxxxxxxx(xxxxx)xxx C_6$ 。

在有些实施方案中，Notch/LNR 单体域包含下面的序列：

$C_1xx(xx)xxx C_2xxxxxn Gx C_3xxx C_4nxxx C_5xx DGx DC_6$

在有些实施方案中，DSL 单体域包含下面的序列：

$C_1xxx Yygxx C_2xxf C_3xxxxdxxxhxxx C_4xxx Gxxx C_5xx GWx Gxx C_6$ 。

Anato 单体域包含下面的序列：

$C_1 C_2xdgxxxxx(x)xxxx C_3exrxxxxxx(x)xx C_4xxx fxx C_5 C_6$ 。

在有些实施方案中，整联蛋白 β 单体域包含下面的序列：

$C_1xx C_2xxxxpx C_3xw C_4xxxx fxxx(gx)xxxx RC_5dxxxx Lxxxxg C_6$ ；

且“x”是任意的氨基酸。

在有些实施方案中，Notch/LNR 单体域的 C_1-C_5 、 C_2-C_4 和 C_3-C_6 形成二硫键；且 DSL 单体域的 C_1-C_5 、 C_2-C_4 和 C_3-C_6 形成二硫键。

在有些实施方案中，Ca-EGF 单体域包含下面的序列：

$D[\beta][Dn]EC_1xx(xx)xxxx C_2[pdg](dx)xxxx C_3xNxxG[sgt][\alpha]x C_4x(xxx)x C_5xx[Gsn][\alpha s]x$
 $xxxxx(xxxxx)xxx C_6$ 。

在有些实施方案中，Notch/LNR 单体域包含下面的序列：

$C_1xx(x[\beta \alpha])xxx C_2x[\phi s]xxx[\phi][Gk]x C_3[nd]x[\phi sa] C_4[\phi s]xx[aeg] C_5x[\alpha] DGx DC_6$ 。

在有些实施方案中，DSL 单体域包含下面的序列：

$C_1xxx[\alpha][\alpha h][Gsna]xx C_2xx[\alpha] C_3x[pae]xx[Da]xx[\chi][Hrgk][\alpha k]x C_4[dnsg]xx Gxxx C_5xx$
 $G[\alpha]x Gxx C_6$ 。

在有些实施方案中，Anato 单体域包含下面的序列：

C₁C₂x[Dhtl][Ga]xxxx[plant](xx)xxxxC₃[esqdat]x[Rlps]xxxxxx([gepa]x)xxC₄xx[avfpt][Fqvy]xxC₅C₆。

在有些实施方案中，整联蛋白 β 单体域包含下面的序列：

C₁xxC₂[β]xx[ghds][Pk]xC₃[χ][α]C₄xxxx[α]xxx([Gr]xx)x[χ]xRC₅[Dnae]xxxxL[βk]xx[Gn]C₆；

α 选自：w、y、f 和 l；β 选自：v、l、l、a、m 和 f，χ 选自：g、a、s 和 t；δ 选自：k、r、e、q 和 d；ε 选自：v、a、s 和 t；且 φ 选自：d、e 和 n。

在有些实施方案中，Ca-EGF 单体域包含下面的序列：

D[vilf][Dn]EC₁xx(xx)xxxxC₂[pdg](dx)xxxxC₃xNxxG[sgt][fy]xC₄x(xxx)xC₅xx[Gsn][αs]xxxxxx(xxxxx)xxxC₆。

在有些实施方案中，Notch/LNR 单体域，包含下面的序列：

C₁xx(x[yiflv])xxxC₂x[dens]xxx[Nde][Gk]xC₃[nd]x[densa]C₄[Nsde]xx[aeg]C₅x[wyf]DGxDC₆。

在有些实施方案中，DSL 单体域包含下面的序列：

C₁xxx[Ywf][Yfh][Gasn]xxC₂xx[Fy]C₃x[pae]xx[Da]xx[glast][Hrgk][ykfw]xC₄[dsgn]xxGxxxC₅xxG[Wlfy]xGxxC₆。

在有些实施方案中，Anato 单体域包含下面的序列：

C₁C₂x[adehlt]gxxxxxxxx(x)[derst]C₃xxxxxxxxxx(xx[aersv])C₄xx[apvt][fmq][eklqrv][adehqrsk](x)C₅C₆。

在有些实施方案中，整联蛋白 β 单体域包含下面的序列：

C₁[aegkqrst][kreqd]C₂[il][aelqrv][vilas][dghs][kp]xC₃[gast][wy]C₄xxxx[fl]xxxx(xxxx[vilar]r)C₅[and][dilt][iklpqrv][adepts][aenq]l[iklv]x[adknr][gn]C₆。

一般地采用编码单体域的多核苷酸，来通过表达制备单体域。编码单体域的核酸可以源自许多种不同的来源。通过表达许多不同的编码天然发生的单体域、改变的单体域(即，单体域变体)或其组合的核酸，可以制备单体域文库。例如，可以设计文库，其中氨基酸的支架保持恒定(例如，LDL A 受体域、EGF 域)，而支架中的插入氨基酸包含随机地产生的氨基酸。

本发明提供了鉴别结合选定的或所需配体或配体混合物的单体域的方法。在有些实施方案中，鉴别或选择单体域的所需的性质(例如，结合亲和力)，然后使单体域形成多聚体。见，例如，图4。对于那些实施方案，可以使用导致选择具有所需的性质(例如，特异性的结合性质)的域的任何方法。例如，该方法可以包含，提供许多不同的核酸，每种核酸编码一种单体域；翻译许多不同的核酸，从而提供许多不同的单体域；筛选许多不同的单体域对所需配体或配体混合物的结合；和鉴别结合所需配体或配体混合物的许多不同的单体域的成员。

单体域可以是天然发生的或改变的(非天然的变体)。术语“天然发生的”在本文中用于指，可以在自然界发现的物体。例如，天然的单体域可以包括人单体域，或任选地，源自不同物种或来源的域，例如，哺乳动物、灵长类动物、啮齿类动物、鱼、鸟、爬行动物、植物等。天然发生的单体域可以通过许多方法得到，例如，通过基因组DNA或cDNA的PCR扩增。

本发明的单体域可以是天然发生的域或非天然发生的变体。在本发明的实践中采用的单体域文库可以含有天然发生的单体域、非天然发生的单体域变体或其组合。

单体域变体可以包括祖先域、嵌合域、随机化的域、突变的域等。例如，祖先域可以基于系统发生分析。嵌合域是这样的域，其中一个或多个区域被替换为来自相同家族的其它域的对应区域。例如，通过组合来自相同家族的多个相关域的环序列，以形成具有潜在地降低的免疫原性的新域，可以构建嵌合域。本领域的技术人员会认识到，通过组合来自相同家族的各种相关域的环区域，而不是生成随机的氨基酸序列，来构建修饰的结合域单体的免疫学益处。例如，通过组合在人LDL受体A类-域中天然地发生的环序列或甚至多个环序列，来构建可变域，得到的域可能含有新的结合性质，但是可能不含有任何免疫原性蛋白序列，因为所有暴露的环都是人来源的。在内源上下文中环氨基酸序列的组合，可以应用于本发明的所有单体构建体。因而，本发明提供了生成源自人蛋白的嵌合单体域的文库的方法，该方法包含：提供环序列，其与至少一个来自人蛋白的至少2种不同的天然发生的变体的每一个的环相对应，其中该环序列是多核苷酸或多肽

序列；和共价组合环序列，以生成至少 2 种不同的嵌合序列的文库，其中每个嵌合序列编码具有至少 2 个环的嵌合单体域。一般地，嵌合域具有至少 4 个环，且通常至少 6 个环。如上所述，本发明提供了通过特定的特征鉴别出的 3 类环，例如，形成二硫键键合的潜力、二级蛋白结构之间的桥接和分子动力学(即柔性)。3 类环序列是半胱氨酸-定义的环境序列、结构-定义的环境序列和 B-因子-定义的环境序列。

随机化的域是这样的域，其中一个或多个区域是随机化的。随机化可以基于完全随机化，或任选地，部分随机化，其基于序列多样性的天然分布。

本发明也提供了重组核酸，其编码一种或多种包含一个或多个结合 c-MET 的单体域的多肽。例如，可以选择多肽，以包含来自下列的非天然发生的域：EGF-样域、三环域、纤连蛋白 I 型域、纤连蛋白 II 型域、纤连蛋白 III 型域、PAN 域、Gla 域、SRCR 域、Kunitz/Bovine 胰腺胰蛋白酶抑制剂域、Kazal-型丝氨酸蛋白酶抑制剂域、Trefoil(P-型)域、冯维勒布兰德因子 C 型域、过敏毒素-样域、CUB 域、甲状腺球蛋白 I 型重复序列、LDL-受体 A 类域、Sushi 域、Link 域、血小板反应蛋白 I 型域、免疫球蛋白-样域、C-型凝集素域、MAM 域、冯维勒布兰德因子 A 型域、生长调节素 B 域、WAP-型四二硫键核心域、F5/8C 型域、血红素结合蛋白域、SH2 域、SH3 域、层粘连蛋白-型 EGF-样域、C2 域和其中的一种或多种的变体。在另一个实施方案中，天然发生的多肽编码在 Pfam 数据库和/或 SMART 数据库中发现的单体域。

本发明的所有组合物，包括通过本发明的方法生产的组合物，例如，单体域和/或免疫域，以及多聚体和其文库，可以任选地结合到亲和材料的基质上。亲和材料的实例包括珠、柱、固体支持物、微阵列、试剂-支持物的其它库等。

III. 多聚体

生成多聚体的方法，是本发明的一个特征。多聚体包含至少 2 个单体域。例如，本发明的多聚体可以包含 2 至约 10 个单体域，2 至约 8 个单体域，约 3 至约 10 个单体域，约 7 个单体域，约 6 个单体域，约 5 个单体域，或约 4 个单体域。在有些实施方案中，多聚体

包含 3 或至少 3 个单体域。在有些实施方案中，多聚体具有不超过 2、3、4、5、6、7 或 8 个单体域。考虑到单体域大小的可能范围，本发明的多聚体可以是，例如，小于 100 kD、小于 90 kD、小于 80 kD、小于 70 kD、小于 60 kD、小于 50 kD、小于 40 kD、小于 30 kD、小于 25 kD、小于 20 kD、小于 15 kD、小于 10 kD 或可以更小或更大。在有些情况下，已经针对与目标靶分子(例如，Met)的结合预选择单体域。

在有些实施方案中，每个单体域特异性地结合一个靶分子(例如，c-Met)。在有些这样的实施方案中，每个单体结合靶分子上的一个不同的位置(类似于表位)。结合相同靶分子的多个单体域会导致抗体亲抗原性效应，从而与每种单个的单体亲和力相比，产生提高的多聚体对靶分子的亲和力。在有些实施方案中，多聚体具有的抗体亲抗原性是单独的单体域的抗体亲抗原性的至少约 1.5、2、3、4、5、10、20、50、100、200、500 或 1000 倍。在有些实施方案中，多聚体的至少 1、2、3、4 或更多个(例如，所有)单体都结合离子，例如钙或另一种离子。多聚体可以包含单体域的许多组合。例如，在单个的多聚体中，选择的单体域可以相同或不同。另外，选择的单体域可以包含来自相同单体域家族的各种不同单体域，或来自不同域家族的各种单体域，或任选地，二者的组合。例如，单体域可以选自 c-MET 结合单体域的家族 1-10。在有些实施方案中，至少一个单体域选自 c-MET 结合单体域的家族 10。在实施例 10 中列出了示例性的 c-MET-结合二聚体(包含 2 种 c-MET-结合单体)。

在本发明的实践中生成的多聚体可以是下列的任一种：

(1) 同源多聚体(相同域的多聚体，即，A1-A1-A1-A1)；

(2) 相同域类别的不同域的异源多聚体，例如，A1-A2-A3-A4。例如，异源多聚体包括这样的多聚体，其中 A1、A2、A3 和 A4 是特定 LDL-受体 A 类域的不同非天然发生的变体，或其中有些 A1、A2、A3 和 A4 是 LDL-受体 A 类域的天然发生的变体。

(3) 来自不同单体域类别的域的异源多聚体，例如，A1-B2-A2-B1。例如，其中 A1 和 A2 是来自 LDL-受体 A 类的 2 个不同的单体域(天然发生的或非天然发生的)，且 B1 和 B2 是来自 EGF-样域类别的 2 个不同的单体域(天然发生的或非天然发生的)。

在另一个实施方案中，多聚体包含具有对不同靶分子(例如，血液因子，例如血清清蛋白、免疫球蛋白或红细胞)的特异性的单体域。例如，在有些实施方案中，本发明的多聚体包含 1、2、3 或更多个结合 Met 的单体域和至少一个结合第 2 种靶分子的单体域。示例性的靶分子包括，例如，延长多聚体的血清半衰期的血清分子(例如，免疫球蛋白或血清清蛋白)、EGFR 基因家族成员、VEGF 受体、PDGF 受体、其它受体酪氨酸激酶、整联蛋白、参与肿瘤发生的其它分子或肿瘤组织的标记。示例性的延长多聚体的血清半衰期的分子包括，例如，红细胞(即，红细胞)、IgG 和血清清蛋白，例如 HSA。示例性的多聚体包括来自 c-MET 结合单体域家族 10 的单体域，和来自免疫球蛋白结合单体域家族 2 或 3 的单体域。

在本发明的实践中采用的多聚体文库可以含有同源多聚体，相同单体类别的不同单体域(天然的或非天然的)的异源多聚体，或来自不同单体类别的单体域(天然的或非天然的)的异源多聚体，或其组合。

如本文所述的，也可以在含免疫域的异源多聚体(即，具有至少一个免疫域变体和 1 个单体域变体的多聚体)中，容易地采用单体域。因而，本发明的多聚体可以具有至少一个免疫域，例如，微型抗体(minibody)、单域抗体、单链可变片段(ScFv)或 Fab 片段；和至少一个单体域，例如，EGF-样域、三环域、纤连蛋白 I 型域、纤连蛋白 II 型域、纤连蛋白 III 型域、PAN 域、Gla 域、SRCR 域、Kunitz/Bovine 胰腺胰蛋白酶抑制剂域、Kazal-型丝氨酸蛋白酶抑制剂域、Trefoil(P-型)域、冯维勒布兰德因子 C 型域、过敏毒素-样域、CUB 域、甲状腺球蛋白 I 型重复序列、LDL-受体 A 类域、Sushi 域、Link 域、血小板反应蛋白 I 型域、免疫球蛋白-样域、C-型凝集素域、MAM 域、冯维勒布兰德因子 A 型域、生长调节素 B 域、WAP-型四二硫键核心域、F5/8C 型域、血红素结合蛋白域、SH2 域、SH3 域、层粘连蛋白-型 EGF-样域、C2 域或其变体。

在连接域以形成多聚体之前，无需选择域。另一方面，在连接进多聚体之前，可以针对结合靶分子的能力选择域。因而，例如，多聚体可以包含结合 1 个靶分子的 2 个域和结合第 2 种靶分子的第 3 个域。

本发明的多聚体可以具有下面的性质：多价的、多特异性的、单链、热稳定的、延长的血清和/或保存半衰期。而且，至少一个、超过1个或所有单体域可以结合离子(例如，金属离子或钙离子)，至少一个、超过1个或所有单体域可以源自 LDL 受体 A 域和/或 EGF-样域，至少一个、超过1个或所有单体域可以是非天然发生的，和/或至少一个、超过1个或所有单体域可以包含 1、2、3 或 4 个二硫键/单体域。在有些实施方案中，多聚体包含至少 2 个(或至少 3 个)单体域，其中至少一个单体域是非天然发生的单体域，且该单体域结合钙。在有些实施方案中，多聚体包含至少 4 个单体域，其中至少一个单体域是非天然发生的，且其中：

a. 每个单体域是 30-100 个氨基酸，且每个单体域包含至少一个二硫键；或

b. 每个单体域是 30-100 个氨基酸，且源自胞外蛋白；或

c. 每个单体域是 30-100 个氨基酸，且结合蛋白靶。

在有些实施方案中，多聚体包含至少 4 个单体域，其中至少一个单体域是非天然发生的，且其中：

a. 每个单体域是 35-100 个氨基酸；或

b. 每个域包含至少一个二硫键，且源自人蛋白和/或胞外蛋白。

在有些实施方案中，多聚体包含至少 2 个单体域，其中至少一个单体域是非天然发生的，且其中每个域是：

a. 25-50 个氨基酸长，且包含至少一个二硫键；或

b. 25-50 个氨基酸长，且源自胞外蛋白；或

c. 25-50 个氨基酸，且结合蛋白靶；或

d. 35-50 个氨基酸长。

在有些实施方案中，多聚体包含至少 2 个单体域，其中至少一个单体域是非天然发生的，且：

a. 每个单体域包含至少一个二硫键；或

b. 至少一个单体域源自胞外蛋白；或

c. 至少一个单体域结合靶蛋白。

鉴别的单体域和/或多聚体可以具有生物活性，这是指，包括至少对选定的或所需配体的特异性的结合亲和力，且在有些情况下，还包括下述能力：阻断其它化合物的结合，刺激或抑制代谢途径，起信

号或信使的作用，刺激或抑制细胞活性，等。可以产生单体域，以起受体(孤儿受体)的配体的作用，其中尚未鉴别出该受体的天然配体。可以生成这些孤儿配体，以阻断或激活它们结合的受体。

可以使用单一配体，或任选地可以使用多种配体，以选择单体域和/或多聚体。本发明的单体域可以结合单一配体或多种配体。本发明的多聚体可以具有单一配体的多个离散的结合位点，或任选地可以具有多种配体的多个结合位点。

在有些实施方案中，多聚体包含具有对不同蛋白的特异性的单体域。不同的蛋白可以是相关的或无关的。相关的蛋白的实例包括蛋白家族的成员或病毒的不同血清型。或者，多聚体的单体域可以靶向生理途径中的不同分子(例如，不同的凝血蛋白)。在其它实施方案中，单体域结合无关途径中的蛋白(例如，2个域结合血液因子、2个其它域结合炎症-相关的蛋白、且第5个结合血清清蛋白)。在另一个实施方案中，多聚体包含结合不同的目标病原体或污染物的单体域。这样的多聚体可以用作单个检测试剂，其能检测众多病原体或污染物中的任一种的可能性。

在有些实施方案中，本发明的多聚体结合相同或其他的多聚体，以形成聚集体。聚集可以由例如2个单体域上的疏水域的存在来介导，从而导致2个单体域之间的非共价相互作用的形成。或者，聚集可以由多聚体中的一个或多个单体域促进，其具有对另一种多聚体中的单体域的结合特异性。由于单体域或多聚体上的亲和肽的存在，也可以形成聚集体。聚集体可以含有比单个多聚体更多的靶分子结合域。

具有对细胞表面靶和第2种靶的亲力的多聚体可以提供增强的抗体亲抗原性效应。在有些情况下，在最优化(通过自装配)相互作用的间隔和效价方面，膜流动性可以比蛋白接头更灵活。在有些情况下，多聚体会结合2种不同的靶，其各自在不同的细胞上，或一种在细胞上，而另一种在具有多个结合位点的分子上。

在有些实施方案中，将本发明的单体或多聚体连接到另一种多肽上，以形成融合蛋白。本领域的任意多肽都可以用作融合配偶体，尽管若融合配偶体形成多聚体则是有用的。例如，本发明的单体或多聚体可以例如融合到下面的抗体位置或位置的组合：

1. 在 VH1 和/或 VL1 域的 N-末端, 任选地刚好在前导肽之后且在域起始之前(构架区 1);
2. 在 CH1 或 CL1 域的 N-末端, 从而替代 VH1 或 VL1 域;
3. 在重链的 N-末端, 任选地在 CH1 域之后且在铰链的半胱氨酸残基之前(Fc-融合体);
4. 在 CH3 域的 N-末端;
5. 在 CH3 域的 C-末端, 任选地通过短接头附着到最后的氨基酸残基上;
6. 在 CH2 域的 C-末端, 从而替代 CH3 域;
7. 在 CL1 或 CH1 域的 C-末端, 任选地在形成链间二硫键的半胱氨酸之后; 或
8. 在 VH1 或 VL1 域的 C-末端。见, 例如, 图 7。

在有些实施方案中, 可以将本发明的一个或多个单体或多聚体域连接到用作药物的分子(例如, 蛋白、核酸、有机小分子等)上。示例性的药物蛋白包括, 例如, 细胞因子、抗体、趋化因子、生长因子、白细胞介素、细胞表面蛋白、胞外域、细胞表面受体、细胞毒素等。示例性的小分子药物包括毒素或治疗剂。在有些实施方案中, 金属可以结合到本发明的多肽上。这可以用作例如造影剂, 例如, 用于 MRI。

在有些实施方案中, 选择单体或多聚体, 以结合组织-或疾病-特异性的靶蛋白。组织-特异性的蛋白是, 与动物的其它组织相比, 在一种或几种特定组织中排它地或以显著更高的水平表达的蛋白。由于 c-MET 在肝中以显著水平表达, 所以可以使用结合 Met 的单体域, 以将其它分子(包括其它单体域)靶向肝。这可以用于靶向肝-特异性的疾病, 例如, 通过将治疗性或毒性分子靶向肝。可以治疗的肝病的一个实例是肝细胞癌。类似地, 疾病-特异性的蛋白是, 与动物的其它无病的细胞或组织相比, 在一种或几种患病的细胞或组织中排它地或以显著更高的水平表达的蛋白。

在有些实施方案中, 将结合靶蛋白的单体或多聚体连接到药物蛋白或小分子上, 从而使得得到的复合物或融合体靶向特定的表达靶蛋白(例如, c-MET)的组织或疾病-相关的细胞。可以针对与靶蛋白的结合初步选择在这样的复合物或融合体中使用的单体或多聚体, 且随后可以通过针对其它细胞或组织的负选择进行选择(例如, 为了避免靶

向设定药物毒性的下限的骨髓或其它组织), 其中需要降低或消除其它非靶细胞或组织中的结合。通过使药物远离敏感的组织, 会升高治疗窗, 从而可以安全地施用更高的剂量。在另一个替代方案中, 通过将单体或多聚体文库注射进动物, 然后分离结合特定的目标组织或细胞的单体或多聚体, 可以在动物中进行体内淘选。

上述的融合蛋白也可以包括在药物蛋白和单体或多聚体之间的接头肽。可以采用肽接头序列, 以分开例如多肽组分一定的距离, 所述距离足以确保每个多肽折叠成它的二级和三级结构。通常, 可以使用标准的技术, 包括化学缀合, 来制备融合蛋白。通过标准的技术, 也可以在表达系统中将融合蛋白表达为重组蛋白。

根据本领域已知的任意方法, 可以生产本发明的多聚体或单体域。在有些实施方案中, 诱导包含编码多肽的 pET-衍生的质粒的大肠杆菌 (*E.coli*), 以表达蛋白。收获细菌后, 可以裂解它们, 并通过离心澄清。使用 Ni-NTA 琼脂糖洗脱, 可以纯化多肽, 并通过渗析再折叠。通过用碘乙酸使游离的巯基加帽, 可以中和错误折叠的蛋白。可以使用 Q 琼脂糖洗脱、丁基琼脂糖流通、SP 琼脂糖洗脱、DEAE 琼脂糖洗脱和/或 CM 琼脂糖洗脱, 来纯化多肽。也可以采用等价的阴离子和/或阳离子交换纯化步骤。

在有些实施方案中, 将包含本发明的单体或多聚体的多肽连接到它自身上(C-末端到 N-末端), 例如, 为了蛋白稳定性。

IV. 接头

可以通过接头连接单体域, 以形成多聚体。例如, 可以将接头置于多聚体中的每个分开的离散的单体域之间。

使用本领域已知的多种技术, 可以通过接头连接选择的单体域。例如, 通过限制酶切消化和重新连接, 通过基于 PCR 的、自引发重叠反应, 或其它重组方法, 可以实现编码选择的单体域的多核苷酸的组合装配。在鉴别单体的结合靶多聚体的能力之前, 或在已经选择单体的结合靶多聚体的能力之后, 可以将接头附着到单体上。

接头可以是天然发生的, 合成的, 或二者的组合。例如, 合成的接头可以是随机化的接头, 例如, 在序列和大小方面。在一个方面, 随机化的接头可以包含完全随机化的序列, 或任选地, 随机化的接

头，其可以基于天然的接头序列。接头可以包含，例如，非-多肽部分、多核苷酸、多肽等。

接头可以是刚性的，或柔性的，或二者的组合。接头柔性可以随接头和与接头相互作用的单体域的组成而变。接头连接2个选择的单体域，并将单体域维持作为分开的离散的单体域。接头可以允许分开的离散的单体域协作，而仍维持分开的性质，例如多聚体中的相同配体的多个分开的结合位点，或例如，多聚体中的不同配体的多个分开的结合位点。

对要连接2个或更多个单体域(即多肽链)的特定情况合适的接头的选择，可能依赖于多种参数，包括例如单体域的性质，多肽多聚体要结合的靶的结构和性质，和/或肽接头对蛋白水解和氧化的稳定性。

本发明提供了一旦已经鉴别出所需的单体域/变体就最优化接头的选择的方法。通常，可以如上所述，容易地制备和筛选具有在单体域组成方面固定的组成、但是接头组成和长度可变的多聚体的文库。

接头的更详细的讨论，可以参见，例如，美国专利公布号2005/0048512。

V. 鉴别具有对靶分子的亲和力的单体或多聚体

本领域的技术人员可以容易地鉴别具有所需的性质(例如，结合亲和力)的单体域。对于那些实施方案，可以使用导致具有所需的性质(例如，特异性的结合性质)的域的选择的任何方法。例如，该方法可以包含，提供许多不同的核酸，每种核酸编码一种单体域；翻译许多不同的核酸，从而提供许多不同的单体域；针对所需配体或配体混合物的结合筛选许多不同的单体域；和鉴别结合所需配体或配体混合物的许多不同的单体域的成员。

另外，可以使用任意的诱变方法，例如定点诱变和随机诱变(例如，化学诱变)，以生产单体域，例如，对于单体域文库。在有些实施方案中，采用易错PCR来生成变体。其它方法包括，通过比对许多天然发生的单体域中的保守氨基酸，比对许多天然发生的单体域；和，通过维持保守氨基酸和插入、缺失或改变保守氨基酸周围的氨基酸，以产生非天然发生的单体域，来设计非天然发生的单体域。在

一个实施方案中，保守氨基酸包含半胱氨酸。在另一个实施方案中，插入步骤使用随机的氨基酸，或任选地，插入步骤使用天然发生的单体域的部分。该部分理想地编码来自相同家族的域的环。使用合成的寡核苷酸，或通过改组，或通过基于限制酶的重组，来插入或交换氨基酸。本发明的人嵌合域可以用于需要最小的免疫原性的治疗用途。本发明提供了产生人嵌合域文库的方法。通过组合来自人单体域的不同变体的环序列，如上所述，可以构建人嵌合单体域文库。组合的环序列可以是序列-定义的环、结构-定义的环、B-因子-定义的环或其中的任意 2 种或更多种的组合。

或者，与环水平相比，通过在氨基酸水平修饰天然发生的人单体域，可以产生人嵌合域文库。在有些实施方案中，为了使免疫原性的潜力最小化，仅仅使用在来自相同人单体域家族的蛋白序列中天然发生的那些残基，来生成嵌合序列。这可以如下实现：提供来自相同单体域家族的至少 2 个人单体域的序列比对，鉴别在人单体域之间不同的人单体域序列对应位置中的氨基酸残基，产生 2 个或更多个人嵌合单体域，其中每个人嵌合单体域序列由在类型和位置上与来自相同单体域家族的 2 个或更多个人单体域的残基相对应的氨基酸残基组成。可以如下采用人嵌合单体域文库来鉴别结合目标靶的人嵌合单体域：针对与靶分子的结合筛选人嵌合单体域文库，和鉴别结合靶分子的人嵌合单体域。在最初的序列比对步骤中采用的合适的天然发生的人单体域序列包括与任一种本文所述的天然发生的单体域相对应的那些。

通过本领域的普通技术人员已知的方法，可以制备本发明的人单体变体文库的域(无论是通过改变环还是单个的氨基酸残基产生的)。特别适用于产生这些文库的方法是，如 WO01/23401 所述的裂解-库(split-pool)格式和三核苷酸合成格式。

在有些实施方案中，如下针对潜在免疫原性筛选本发明的单体域：

提供候选蛋白序列；

对比候选蛋白序列和人蛋白序列数据库；

鉴别与来自数据库的人蛋白序列的部分相对应的候选蛋白序列的部分；和

确定候选蛋白序列和来自数据库的人蛋白序列之间的对应程度。

通常，候选蛋白序列和来自数据库的一种或多种人蛋白序列之间的对应性程度越大，与具有很少的与来自数据库的任一种人蛋白序列的对应性的候选蛋白相对比所预测的免疫原性潜力就越低。适用于实践本发明的筛选候选蛋白的方法的人蛋白序列数据库，可以参见环球网 ncbi.nlm.nih.gov/blast/Blast.cgi (另外，可以使用下面的环球网网站来搜索短的、近乎准确的匹配：[cbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&LAYOUT=TwoWindows&AUTO_FORMAT=Semiauto&ALIGNMENTS=50&ALIGNMENT_VIEW=Pairwise&CLIENT=web&DATABASE=nr&DESCRIPTIONS=100&ENTREZ_QUERY=\(none\)&EXPECT=1000&FORMAT_OBJECT=Alignment&FORMAT_TYPE=HTML&NCBI_GI=on&PAGE=Nucleotides&PROGRAM=blastn&SERVICE=plain&SET_DEFAULTS.x=29&SET_DEFAULTS.y=6&SHOW_OVERVIEW=on&WORD_SIZE=7&END_OF_HTTPGET=Yes&SHOW_LINKOUT=yes](http://ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&LAYOUT=TwoWindows&AUTO_FORMAT=Semiauto&ALIGNMENTS=50&ALIGNMENT_VIEW=Pairwise&CLIENT=web&DATABASE=nr&DESCRIPTIONS=100&ENTREZ_QUERY=(none)&EXPECT=1000&FORMAT_OBJECT=Alignment&FORMAT_TYPE=HTML&NCBI_GI=on&PAGE=Nucleotides&PROGRAM=blastn&SERVICE=plain&SET_DEFAULTS.x=29&SET_DEFAULTS.y=6&SHOW_OVERVIEW=on&WORD_SIZE=7&END_OF_HTTPGET=Yes&SHOW_LINKOUT=yes))。该方法特别适用于确定嵌合蛋白(例如，嵌合单体域)中的交换(crossover)序列是否可能造成免疫原性事件。如果交换序列与在人蛋白序列数据库中发现的序列的部分相对应，则认为交换序列不太可能造成免疫原性事件。

关于来自数据库的人蛋白序列的部分的信息，可以用于设计人-样嵌合蛋白的蛋白文库。使用关于存在于天然发生的人蛋白中的“交换序列”的信息，可以产生这样的文库。术语“交换序列”在本文中指完全存在于至少一种天然发生的人蛋白中的序列，其中该序列的部分存在于两种或更多种天然发生的蛋白中。因而，后两种或更多种天然发生的蛋白的重组，会产生嵌合蛋白，其中该序列的嵌合部分实际上与在另一种天然发生的蛋白中发现的序列相对应。交换序列含有2个连续氨基酸残基位置的嵌合接头，其中第一个氨基酸位置被在第一种和第二种天然发生的人蛋白序列中发现、但是没有在第三种天然发生的人蛋白序列中发现的类型和位置相同的氨基酸残基占据。第二个氨基酸位置被在第二种和第三种天然发生的人蛋白序列中发现、但是没有在第一种天然发生的人蛋白序列中发现的类型和位置相同的氨基酸残基占据。换言之，“第二种”天然发生的人蛋白序列与交换序列在其中整体出现的天然发生的人蛋白相对应，如上所述。

在有些实施方案中，如下产生人-样嵌合蛋白的文库：鉴别与来

自相同蛋白家族的蛋白相对应的来自数据库的人蛋白序列；比对来自相同蛋白家族的人蛋白序列和参照蛋白序列；鉴别一组来自相同家族的不同人蛋白序列的子序列，其中每个子序列共有与至少一种源自不同的天然发生的人蛋白序列的其它子序列一致的区域；鉴别来自第一种、第二种和第三种子序列的嵌合接头，其中每个子序列源自不同的天然发生的人蛋白序列，且其中嵌合接头包含 2 个连续氨基酸残基位置，其中第一个氨基酸位置被第一种和第二种天然发生的人蛋白序列中共有、但是第三种天然发生的人蛋白序列没有的氨基酸残基占据，且第二个氨基酸位置被第二种和第三种天然发生的人蛋白序列共有的氨基酸残基占据，并产生人-样嵌合蛋白分子，其各自的序列对应着来自子序列组的两种或更多种子序列，且各自包含鉴别的嵌合接头中的一个或多个。

因而，例如，如果第一种天然发生的人蛋白序列是 A-B-C，第二种是 B-C-D-E，第三种是 D-E-F，则嵌合接头是 C-D。或者，如果第一种天然发生的人蛋白序列是 D-E-F-G，第二种是 B-C-D-E-F，第三种是 A-B-C-D，则嵌合接头是 D-E。可以以多种方式产生人-样嵌合蛋白分子。例如，可以将包含编码嵌合接头的序列的寡核苷酸与其序列与 2 种或更多种来自上述子序列组的子序列相对应的寡核苷酸重组，以产生人-样嵌合蛋白，和其文库。用于比对天然发生的人蛋白的参照序列是来自天然发生的人蛋白的相同家族的序列，或该家族的蛋白的嵌合体或其它变体。

也可以混合和/或重组(例如，通过使用化学地或酶促地生成的片段)编码天然发生的单体域的片段的核酸，以产生全长的修饰的单体域。通过操纵编码域或其片段的核酸，也可以重组片段和单体域。例如，可以连接编码单体域的片段的核酸构建体，以产生改变的单体域。

也可以如下产生改变的单体域：提供编码肽序列的保守的、随机的、伪随机的或定义的序列的合成寡核苷酸(例如，重叠寡核苷酸)的集合，然后通过连接，将其插入编码单体域的多核苷酸的预定位点。类似地，通过用定点诱变、随机的突变、伪随机的突变、定义的核心(kernal)突变、基于密码子的突变等，来突变单体域，可以扩展一个或多个单体域的序列多样性。可以在用于克隆和扩增的宿主

中，繁殖得到的核酸分子。在有些实施方案中，改组核酸。

本发明也提供了重组许多编码单体域的核酸和筛选得到的文库中结合所需配体或配体混合物等的单体域的方法。通过与编码中性序列（即，对结合具有非实质的功能作用）的多核苷酸序列改组，也可以回交选择的单体域核酸，例如，通过与野生型或与选择的序列基本上相同的天然发生的序列回交，以生成天然-样功能单体域。通常，在回交过程中，进行后续选择，以保留性质，例如，与配体的结合。

在有些实施方案中，通过改组制备单体文库。在这样的情况下，分离和改组单体域，以组合地重组编码单体域的核酸序列（重组可以发生在单体域之间或内部，或二者）。第一步包含，鉴别具有所需的性质（例如，对某种配体的亲和力）的单体域。在重组过程中，在维持保守氨基酸的同时，编码单体域的核酸序列可以重组，或重组并连接成多聚体。

本发明的显著优点是，可以使用已知的配体或未知的配体来选择单体域和/或多聚体。不需要现有的关于配体结构的信息，以分离目标单体域或目标多聚体。鉴别的单体域和/或多聚体可以具有生物活性，这是指，包括至少对选定的或所需配体的特异性的结合亲和力，且在有些情况下，还包括下述能力：阻断其它化合物的结合，刺激或抑制代谢途径，起信号或信使的作用，刺激或抑制细胞活性，等。可以产生单体域，以起受体（孤儿受体）的配体的作用，其中尚未鉴别出该受体的天然配体。可以生成这些孤儿配体，以阻断或激活它们结合的受体。

可以使用单一配体，或任选地可以使用多种配体，以选择单体域和/或多聚体。本发明的单体域可以结合单一配体或多种配体。本发明的多聚体可以具有单一配体的多个离散的结合位点，或任选地可以具有多种配体的多个结合位点。

本发明也包括通过本发明的方法生产的组合物。例如，本发明包括从包含通过本发明的方法生产的单体域的一个和/或多个文库选择或鉴别的单体域。

本发明也提供了单体域的文库和编码单体域的核酸的文库。文库可以包括，例如，约 100、250、500 或更多个编码单体域的核酸，或文库可以包括，例如，约 100、250、500 或更多个编码单体域的多肽。

文库可以包括含有相同半胱氨酸框的单体域，例如，A-域或 EGF-样域。

在有些实施方案中，通过重组来自相同单体域家族(例如，LDL受体 A 类域)的 2 种或更多种不同的序列，产生变体。或者，可以组合来自不同家族的 2 种或更多种不同的单体域，以形成多聚体。在有些实施方案中，从至少一种下述家族类别的单体或单体变体，形成多聚体：EGF-样域、三环域、纤连蛋白 I 型域、纤连蛋白 II 型域、纤连蛋白 III 型域、PAN 域、Gla 域、SRCR 域、Kunitz/Bovine 胰腺胰蛋白酶抑制剂域、Kazal-型丝氨酸蛋白酶抑制剂域、Trefoil(P-型)域、冯维勒布兰德因子 C 型域、过敏毒素-样域、CUB 域、甲状腺球蛋白 I 型重复序列、LDL-受体 A 类域、Sushi 域、Link 域、血小板反应蛋白 I 型域、免疫球蛋白-样域、C-型凝集素域、MAM 域、冯维勒布兰德因子 A 型域、生长调节素 B 域、WAP-型四二硫键核心域、F5/8C 型域、血红素结合蛋白域、SH2 域、SH3 域、层粘连蛋白-型 EGF-样域、C2 域和其衍生物。在另一个实施方案中，单体域和不同的单体域可以包括存在于 Pfam 数据库和/或 SMART 数据库中的一个或多个域。本发明也包括通过上述方法生成的文库，包含该文库的一种或多种成员的一种或多种细胞，以及包含该文库的一种或多种成员的一种或多种展示。

任选地，可以产生编码单体域的核酸字符串的数据组，例如，通过混合编码单体域的第一种字符串和编码不同单体域的一个或多个字符串，从而生成编码单体域的核酸字符串的数据组，包括本文所述的那些。在另一个实施方案中，单体域和不同的单体域可以包括存在于 Pfam 数据库和/或 SMART 数据库中的一个或多个域。该方法还可以包含，将编码单体域的第一种字符串和编码不同单体域的一个或多个第二种字符串输入计算机，并在计算机中产生多聚体字符串或其文库。

可以针对所需的性质筛选文库，例如，对所需配体或配体混合物的结合。例如，可以针对与已知或未知配体或配体混合物的结合显示和预筛选单体域文库的成员。然后，可以诱变(例如，重组、化学地改变等)或以其他方式改变单体域序列，并针对以提高的亲和力与配体或配体混合物的结合再次筛选新单体域。可以组合或连接选择的单

体域，以形成多聚体，然后可以针对对于配体或配体混合物的提高的亲和力或抗体亲抗原性或改变的特异性对其进行筛选。改变的特异性可以指，特异性被拓宽，例如，多种相关病毒的结合，或任选地，改变的特异性可以指，特异性被缩小，例如，在配体的特定区域内的结合。本领域的技术人员能认识到，存在许多可用于计算抗体亲抗原性的方法。见，例如，Mammen 等, *Angew Chem Int. Ed.* 37: 2754-2794 (1998); Muller 等, *Anal. Biochem.* 261: 149-158 (1998)。

VI. 结合 c-MET 的单体域的选择

通过筛选能结合 c-MET 的试剂，可以进行初步筛选，因为至少有些这样鉴别的试剂可能是 c-MET 调节剂(例如，拮抗剂或激动剂)。结合测定通常包含，使 c-MET 蛋白 (或其片段，例如包含 SEMA 域或 α 链的片段)接触一种或多种测试剂(即，本发明的单体或多聚体)，并允许足以使蛋白和测试剂形成结合复合物的时间。使用众多确立的分析技术中的任一种，可以检测形成的任何结合复合物。蛋白结合测定包括，但不限于，免疫组织化学结合测定、流式细胞仪或其它测定。在这样的测定中使用的 c-MET 蛋白可以是天然地表达的、克隆的或合成的。类似的方法可以用于鉴别结合 IgG 的单体域或多聚体。

可以进行本发明的筛选方法，作为体外或基于细胞的测定。可以在表达 c-MET 的任意细胞中，进行基于细胞的测定。基于细胞的测定可以包含含有 c-MET 受体的整个细胞或细胞级分，以筛选试剂结合或该试剂对 c-MET 的活性的调节。可以根据本发明的方法使用的示例性的细胞类型包括，例如，任意的哺乳动物细胞，以及真菌细胞，包括酵母和细菌细胞。细胞可以是原代细胞或肿瘤细胞或其它类型的无限增殖化细胞系。当然，可以在不内源地含有 c-MET 的细胞中表达 c-MET。

c-MET 活性测定也可以用于鉴别 c-MET 的调节剂(拮抗剂或激动剂)。在这些实施方案中，使一种或多种测试剂接触表达 c-MET 的细胞，然后测试 c-MET 的活性。示例性的 c-MET 活性包括 HGF-依赖性的或组成型的激酶活性。见，例如，Christensen 等, *Cancer Res.* 63: 7345-7355 (2003)。在其它实施方案中，也可以监控下游分子事件，

以确定信号传导活性。例如，c-MET 诱导细胞生长(增殖和存活)、细胞运动性、侵入和形态学变化。另外，除了别的以外，c-MET 间接介导 Gab-1、Akt、转录 3 的信号转导物和激活剂、磷脂酶 C γ 和病灶性粘着激酶的磷酸化。见，例如，Christensen 等，Cancer Res. 63: 7345-7355 (2003)。

在有些实施方案中，也使用活性测定来证实，鉴别的拮抗剂单体或多聚体(即，其与 HGF 竞争)缺少激动剂活性(即，在不存在 HGF 或另一种激动剂的情况下，它们不会激活 c-MET)。

可以进一步测试通过任意的前述筛选方法初步鉴别的试剂，以验证表观活性。可以用合适的动物模型进行这样的研究。这样的方法的基本格式包含，向用作人模型的动物施用在初步筛选过程中鉴别出的前导化合物，然后确定是否实际上调节 c-MET 和/或改善疾病或状况。在验证研究中使用的动物模型通常是任意种类的哺乳动物。合适的动物的特定实例包括，但不限于，灵长类动物、小鼠和大鼠。

通过多种方法，可以从域文库选择结合 c-MET 的单体域。例如，一种鉴别具有所需的性质(例如，结合 c-MET 或 IgG)的单体域的方法包含，翻译许多核酸，其中每种核酸都编码一种单体域，筛选由许多核酸编码的多肽，和鉴别例如结合所需配体或配体混合物的那些单体域，从而生成选择的单体域。通过本领域已知的方法(即，淘选、亲和色谱、FACS 分析)，可以测试由每种核酸表达的单体域结合配体的能力。

如上所述，单体域的选择可以基于与配体(例如 c-MET 或其片段)或其它靶分子(例如，脂质、碳水化合物、核酸等)的结合。在该方法中，可以任选地包含其它分子和靶，例如，离子，例如 Ca^{+2} 。

当基于它结合配体的能力选择本发明的单体域时，选择基础可以包括基于慢离解速率的选择，其通常是高亲和力的先兆。可以改变配体的效价，以控制选择的单体域的平均结合亲和力。配体可以以不同的密度结合于表面或基底，例如通过包含竞争化合物，通过稀释，或通过本领域的技术人员已知的其它方法。预定配体的高密度(效价)，可以用于富集具有相对低亲和力的单体域，而低密度(效价)可以优先地富集高亲和力单体域。

多种报告显示载体或系统可以用于表达编码本发明的单体域和/

或多聚体的核酸，并用于测试所需的活性。例如，噬菌体展示系统是这样的系统，其中将单体域表达为噬菌体表面上的融合蛋白 (Pharmacia, Milwaukee Wis.)。噬菌体展示可以包含，在丝状噬菌体的表面呈递编码单体域的多肽序列，一般地作为与噬菌体外壳蛋白的融合体。

通常，在这些方法中，每个噬菌体颗粒或细胞用作单个的文库成员，其展示单个种类的展示的多肽以及天然的噬菌体或细胞蛋白序列。将核酸克隆进噬菌体 DNA 中导致融合蛋白转录的位置，其部分由许多核酸编码。含有核酸分子的噬菌体在细胞中经历复制和转录。融合蛋白的前导序列指导着融合蛋白向噬菌体颗粒顶端的运输。因而，在用于通过上面和下面所述的方法进行检测和选择的噬菌体颗粒上，展示部分地由核酸编码的融合蛋白。例如，噬菌体文库可以与预定的配体（例如 c-MET 或其片段）一起温育，以便可以将呈递结合于配体的融合蛋白序列的噬菌体颗粒与不呈递结合于预定配体的多肽序列的那些有差别地区分开。例如，通过固定化预定的配体，可以提供分离。然后，回收并复制结合到固定化的配体上的噬菌体颗粒（即，文库成员），以扩增选择的噬菌体亚群，用于随后的亲和富集和噬菌体复制循环。几轮亲和富集和噬菌体复制循环后，分离这样选择的噬菌体文库成员，并测定编码展示的多肽序列的核苷酸序列，从而鉴别结合于预定配体的多肽的序列。这样的方法另外记载在 PCT 专利公开号 91/17271、91/18980 和 91/19818 和 93/08278 中。

其它展示系统的实例包括核糖体展示、核苷酸-连接的展示（见，例如，美国专利号 6,281,344、6,194,550、6,207,446、6,214,553 和 6,258,558）、多核糖体展示、细胞表面展示等。细胞表面展示包括多种细胞，例如，大肠杆菌、酵母和/或哺乳动物细胞。当将细胞用作展示时，将核酸（例如，通过 PCR 扩增随后消化得到的）导入细胞并翻译。任选地，通过例如注射，可以将编码本发明的单体域或多聚体的多肽导入细胞中。

可以针对所需的性质，例如所需配体（例如，c-MET）或配体混合物的结合，筛选本发明的单体和多聚体文库。例如，可以展示单体域文库的成员，并预筛选向已知或未知配体或配体混合物的结合。然后，可以诱变（例如，重组、化学地改变等）或以其他方式改变单体域

序列，并针对以提高的亲和力与配体或配体混合物的结合再次筛选新单体域。可以组合或连接选择的单体域，以形成多聚体，然后可以针对对配体或配体混合物的提高的亲和力或抗体亲抗原性或改变的特异性筛选它们。改变的特异性可以指，特异性被拓宽，例如，多种相关配体的结合，或任选地，改变的特异性可以指，特异性被缩小，例如，在配体的特定区域内的结合。本领域的技术人员能认识到，存在许多可用于计算抗体亲抗原性的方法。见，例如，Mammen 等, *Angew Chem Int. Ed.* 37: 2754-2794 (1998); Muller 等, *Anal. Biochem.* 261: 149-158 (1998)。

本领域的技术人员将认识到，可以重复(即，递归地进行)产生变化和针对所需的性质进行筛选的步骤，以最优化结果。例如，在噬菌体展示文库或其它类似的格式中，可以以相对较低的严格性进行文库的第一次筛选，从而选择尽可能多的与靶分子有关的颗粒。然后，可以分离选择的颗粒，并从颗粒分离编码单体或多聚体的多核苷酸。然后，可以从这些序列产生其它变化，并随后在更高的亲和力进行筛选。

本发明的所有组合物，例如，单体域以及多聚体和其文库，可以任选地结合到亲和材料的基质上。亲和材料的实例包括珠、柱、固体支持物、微阵列、试剂-支持物的其它库等。

当需要能结合相对较大的靶的多聚体时，可以通过“步移(walking)”选择方法产生它们。通过提供单体域文库和针对对第一种靶分子的亲和力筛选单体域文库，进行该方法。一旦鉴别出至少一种结合于靶的单体，就将该单体共价地连接到新文库或原始单体域文库的每种剩余成员。然后，筛选多聚体(二聚体)的该新文库中以增强的亲和力结合靶的多聚体，并可以鉴别以增强的亲和力结合靶的多聚体。“步移”单体选择方法提供了装配多聚体的方法，所述多聚体由给定接头长度限制时可以附加地或甚至协同地彼此作用的单体组成。当选择和装配能以高亲和力结合大靶蛋白的多聚体时，该步移技术是非常有用的。可以重复该步移方法，以添加更多的单体，从而生成包含 2、3、4、5、6、7、8 或更多个连接到一起的单体的多聚体。

在有些实施方案中，选择的多聚体包含超过 2 个域。可以以逐步的方式产生这样的多聚体。例如，其中单独地测试每个新域的添加，

并以连续的方式测试该域的作用。见，例如，图 5。在一个替代实施方案中，连接域，以形成包含超过 2 个域的多聚体，并选择结合，而无需更小的多聚体或可选择地每个域如何结合的现有知识。

本发明的方法也包括进化单体或多聚体的方法。可以将域内重组越过整个单体导入单体中，或通过取不同单体的部分，以形成新的重组单元。可以实现域间重组(例如，将不同单体重组进多聚体之内或之间)或模块重组(例如，多聚体内的多个单体)。也预期文库间重组。

进化单体或多聚体的方法可以包含，例如，任一个或所有的下述步骤：提供许多不同的核酸，其中每种核酸编码一种单体域；翻译许多不同的核酸，其提供许多不同的单体域；针对对所需配体(例如，c-MET)或配体混合物的结合筛选许多不同的单体域；和鉴别结合所需配体或配体混合物的许多不同的单体域的成员，其提供选择的单体域；用至少一个接头连接选择的单体域，以产生至少一个多聚体，其中该至少一个多聚体包含至少两个选择的单体域和至少一个接头；和，与选择的单体域相比，针对对所需配体或配体混合物的提高的亲和力或抗体亲抗原性或改变的特异性筛选至少一个多聚体。

可以将变化导入单体或多聚体。改良单体的实例包括域内重组，其中在将变化导入(例如，通过改组或其它重组方法)得到的扩增产物的条件下，分别扩增单体的 2 个或更多个(例如，3、4、5 或更多个)部分，从而合成单体的不同部分的变体的文库。通过将中间引物的 5' 末端定位在两种 PCR 片段共同具有的“中间”或“重叠”序列中，可以通过重叠 PCR 组合得到的“左”侧和“右”侧文库，以产生原始单体库的新变体。然后，可以针对所需的性质筛选这些新变体，例如，针对靶进行淘选，或筛选功能作用。可以选择与单体的任意片段相对应的“中间”引物，且其一般地基于支架或单体内的 1 个或多个共有氨基酸(例如，半胱氨酸，例如存在于 A 域中的那些)。

类似地，通过在单体水平导入变化，然后重组单体变体文库，可以生成多聚体。在更大的规模，可以重组具有所需的性质的多聚体(单个的或库)，以形成更长的多聚体。在有些情况下，将变化导入(一般地，合成地)单体或接头，以形成文库。这可以用例如结合 2 种不同的靶的 2 种不同的多聚体来实现，从而最终选择具有结合一种靶的部分和结合第二种靶的部分的多聚体。

通过在域之间插入不同长度和组成的接头，可以导入其它变化。这允许选择域之间的最佳接头。在有些实施方案中，接头的最佳长度和组成会允许域的最佳结合。在有些实施方案中，通过不同接头连接具有特定结合亲和力的域，并在结合测定中选择最佳接头。例如，针对所需的结合性质选择域，然后形成包含多种接头的文库。然后，可以筛选文库，以鉴别最佳的接头。或者，可以形成多聚体文库，其中未知域或接头对靶分子结合的作用。

本发明的方法也包括，通过提供许多单体域，产生一种或多种选择的多聚体。针对对所需配体或配体混合物的结合筛选许多单体域。鉴别结合所需配体或配体混合物的许多域的成员，从而提供具有所需的亲和力的域。用至少一种接头连接鉴别的域，以产生多聚体，其中每个多聚体包含至少 2 个选择的域和至少一个接头；和，与选择的域相比，针对对所需配体或配体混合物的提高的亲和力或抗体亲抗原性或改变的特异性筛选多聚体，从而鉴别一种或多种选择的多聚体。

在有些实施方案中，可以通过在基于重组酶的方法中组合 2 个或更多个文库或单体或多聚体，产生多聚体文库，其中每个文库成员包含一个重组位点(例如，lox 位点)。更大的分子上不同的文库成员库原则上包括更多具有所需的性质(例如更高的靶-结合亲和力和功能活性)的变体。当在噬菌体载体中构建文库时，可以将其转化进大肠杆菌中，通过大肠杆菌的转化效率，限制文库大小(10^9 - 10^{10})。可以采用重组酶/重组位点系统(例如，Cre-loxP 系统)和体内重组，以产生其大小不受大肠杆菌的转化效率限制的文库。

例如，可以使用 Cre-loxP 系统来产生具有 10^{10} 、 10^{11} 、 10^{12} 、 10^{13} 或更大的多样性的二聚体文库。在有些实施方案中，使用大肠杆菌作为一个首次用于实验的单体文库的宿主和携带第 2 个首次用于实验的单体文库的丝状噬菌体。在该情况下，文库大小仅受传染性噬菌体(携带 1 个文库)的数目和能受感染的大肠杆菌细胞(携带其它文库)的数目的限制。例如，用 $>10^{12}$ 噬菌体感染 10^{12} 大肠杆菌细胞(1 L, OD600=1)，会生成多达 10^{12} 的二聚体组合。

可以使用多种技术，包括上述的用于鉴别单体域的那些，完成多聚体的选择。其它选择方法包括，例如，基于与选择的单体域相比，提高的对配体的亲和力或抗体亲抗原性或改变的特异性的选择。例

如，选择可以基于与特定细胞类型或一组相关的细胞或蛋白类型(例如，不同的病毒血清型)的选择性结合。然后，可以如本发明所述，通过重组域，以及操纵各个单体域或接头域的氨基酸序列或编码这样的域的核苷酸序列，来最优化选择的性质，例如对配体的抗体亲抗原性。

通过展示多聚体，可以完成鉴别多聚体的一种方法。如同单体域一样，在多种展示系统上，例如，噬菌体展示、核糖体展示、多核糖体展示、核苷酸-连接的展示(见，例如，美国专利号 6,281,344、6,194,550、6,207,446、6,214,553 和 6,258,558)和/或细胞表面展示，如上所述，任选地表达或展示多聚体。细胞表面展示可以包括但不限于大肠杆菌、酵母或哺乳动物细胞。另外，可以针对对一种或多种配体的亲和力或抗体亲抗原性或改变的特异性淘选具有多个结合位点的多聚体的展示文库。

使用双杂交筛选测定，可以在酵母细胞中针对靶结合活性筛选单体或多聚体。在这类筛选中，将要筛选的单体或多聚体文库克隆进载体，后者指导文库的每个单体或多聚体和酵母转录激活剂片段(即，Gal4)之间的融合蛋白的形成。将编码“靶”蛋白的序列克隆进载体，这导致靶和 Gal4 蛋白的剩余部分(DNA 结合域)之间的融合蛋白的生成。第3种质粒含有在 Gal4 结合位点的 DNA 序列下游的报告基因。可以结合靶蛋白的单体会给它带来 Gal4 激活域，从而重构功能性的 Gal4 蛋白。结合到报告基因上游的结合位点上的该功能性的 Gal4 蛋白，导致报告基因的表达和作为靶结合蛋白的单体或多聚体的选择(见 Chien 等 (1991) *Proc. Natl. Acad. Sci. (USA)* 88: 9578; Fields S. 和 Song O. (1989) *Nature* 340: 245)。使用双杂交系统进行文库筛选，进一步记载在美国专利号 5,811,238 (也见 Silver S. C. 和 Hunt S. W. (1993) *Mol. Biol. Rep.* 17: 155; Durfee 等 (1993) *Genes Devel.* 7: 555; Yang 等 (1992) *Science* 257: 680; Luban 等 (1993) *Cell* 73: 1067; Hardy 等 (1992) *Genes Devel.* 6: 801; Bartel 等 (1993) *Biotechniques* 14: 920; 和 Vojtek 等 (1993) *Cell* 74: 205)。用于实现本发明的另一种有用的筛选系统是大肠杆菌/BCCP 相互作用筛选系统(Germino 等 (1993) *Proc. Nat. Acad. Sci. (U.S.A.)* 90: 993; Guarente L. (1993) *Proc. Nat. Acad. Sci. (U.S.A.)* 90: 1639)。

其它变化包括，使用多种结合化合物，从而可以同时针对具有不同结合特异性的配体或化合物的多重性筛选这些分子的单体域、多聚体或文库。可以在单个文库中伴随地筛选多种预定的配体或化合物，或先后筛选许多单体域或多聚体。在一种变化中，可以混合多种配体或化合物，其各自在分开的珠(或珠的亚群)上编码，并在合适的结合条件下，与这些分子的单体域、多聚体或文库一起温育。然后，通过亲和选择，可以使用包含多种配体或化合物的珠的集合，来分离选择的单体域、选择的多聚体或文库成员。通常，随后的亲和筛选循环可以包括相同的珠、其亚群或仅含有一种或两种单个的配体或化合物的珠的混合物。该方法提供有效的筛选，且与实验室自动化、批处理和高通量筛选方法相容。

在另一个实施方案中，可以同时针对结合多种配体的能力筛选多聚体，其中每种配体包含不同的标记。例如，可以用不同的荧光标记来标记每种配体，其同时接触多聚体或多聚体文库。然后，基于连接到所需的标记上的标记的存在，鉴别(例如，通过 FACS 分选)具有所需的亲和力的多聚体。

可以以许多不同的格式，同时针对多种配体筛选(即，淘选)单体域或多聚体(为了方便，在下面的讨论中称作“亲和试剂”)的文库。例如，可以在简单的混合物中，在阵列中，筛选多种配体，在细胞或组织(例如，细胞或组织提供许多可以被本发明的单体域或多聚体结合分子)上展示，和/或固定化。可以任选地在酵母或噬菌体展示系统上展示亲和试剂的文库。类似地，如果需要，可以在酵母或噬菌体展示系统中展示配体(例如，在 cDNA 文库中编码的)。

最初，针对多种配体淘选亲和试剂文库。任选地，针对配体淘选得到的“命中值(hit)”一次或多次，以富集得到的亲和试剂群体。

如果需要，可以测定各个亲和试剂和/或配体的同一性。在有些实施方案中，在噬菌体上展示亲和试剂。将在最初筛选中鉴别为结合的亲和试剂分成第一部分和第二部分。用第一部分感染细菌，从而根据使用的噬菌体的类型，产生噬菌斑或细菌菌落。固定化表达的噬菌体，然后用在如下所述选择的噬菌体中展示的配体探测。

将第二部分偶联到珠上或以其他方式固定化，并使含有至少一些原始混合物中的配体的噬菌体展示文库接触固定化的第二部分。随后

洗脱结合于第二部分的噬菌体，并接触上段所述的固定化的噬菌体。检测噬菌体-噬菌体相互作用(例如，使用对配体-表达噬菌体特异性的单克隆抗体)，且可以分离得到的噬菌体多核苷酸。

在有些实施方案中，测定亲和试剂-配体对的同一性。例如，当在噬菌体或酵母上展示亲和试剂和配体两者时，可以分离来自该对的DNA，并测序。在有些实施方案中，扩增对配体和亲和试剂特异性的多核苷酸。每个反应的扩增引物可以包括互补的5'序列，从而融合得到的扩增产物，从而形成杂合体多核苷酸，其包含编码亲和试剂的至少一部分和配体的至少一部分的多核苷酸。得到的杂合体可以用于探测亲和试剂或配体(例如，cDNA-编码的)多核苷酸文库，以鉴别亲和试剂和配体。

上述方法可以容易地与“步移”相组合，以同时产生和鉴别多种多聚体，其中的每一种都结合配体混合物中的配体。在这些实施方案中，针对多种配体淘选亲和试剂(单体域或多聚体)的第一个文库，并将洗脱的亲和试剂连接到亲和试剂的第一个或第2个文库，以形成多聚的亲和试剂(例如，包含2、3、4、5、6、7、8、9或更多个单体)的文库，随后针对多种配体淘选它们。可以重复该方法，以继续产生更大的多聚的亲和试剂。增加单体域的数目，可以导致增加的对特定靶的亲合力和抗体亲抗原性。例如，发明人已经发现，结合CD28的单体域的二聚体具有比二聚体更高的亲合力，后者又具有比单个的CD28-结合单体域自身更高的亲合力。当然，在每个阶段，任选地重复淘选，以富集重要的粘合剂。在有些情况下，通过在单体的末端插入重组位点(例如，lox位点)，并通过重组酶-介导的事件重组单体文库，可以促进步移。

可以进一步操纵上述方法选择的多聚体，例如，通过重组或改组选择的多聚体(重组可以发生在多聚体之间或内部或二者)、突变选择的多聚体等。这导致改变的多聚体，然后可以筛选和选择与选择的多聚体相比具有增强的性质的成员，从而生成选择的改变的多聚体。

考虑到本文的描述，显然可以遵循下面的方法。可以重组天然地或非天然地发生的单体域，或可以形成变体。任选地，在开始或以后针对在它们的目标宿主中不太可能是免疫原性的那些序列选择域。任选地，针对希望的亲合力淘选包含重组域的噬菌体文库。可以针对对

靶的 IC₅₀ 筛选由噬菌体表达的单体域或多聚体。可以选择异-或同-聚的多聚体。可以针对对任何靶的亲合力，包括，例如，异-或同-多聚的靶选择多肽。

通过上面和下面所述的方法生成的接头、多聚体或选择的多聚体是本发明的特征。提供了包含多聚体的文库，例如，包含约 100、250、500 或更多种通过本发明的方法生产或通过本发明的方法选择的成员的文库。在有些实施方案中，还包括一种或多种包含文库成员的细胞。重组多肽的文库也是本发明的特征。例如，包含约 100、250、500 或更多种不同的重组多肽的文库。

本发明的组合物可以结合到亲和材料的基质上，例如，重组多肽。亲和材料的实例包括，例如，珠、柱、固体支持物等。

VII. 治疗性和预防性治疗方法

本发明也包括，通过向受试者体内或离体(*ex vivo*)地施用一种或多种上述的本发明的核酸或多肽(或组合物，其包含药学上可接受的赋形剂和一种或多种这样的核酸或多肽)，治疗地或预防地治疗疾病或障碍的方法，所述受试者包括，例如哺乳动物，包括人、灵长类动物、小鼠、猪、母牛、山羊、兔、大鼠、豚鼠、仓鼠、马、羊；或非哺乳动物的脊椎动物，例如鸟(例如，鸡或鸭)、鱼，或无脊椎动物。

c-MET 拮抗剂，包括本发明的 c-MET-结合单体域或多聚体，可以用于治疗表达 c-MET 的人癌症。已知表达 c-MET 和/或它的配体 HGF 的人癌症的概要见 Birchmeier, C., Birchmeier, W., Gherardi, E. & Vande Woude, G. F. Met, metastasis, motility and more. *Nat Rev Mol Cell Biol* 4, 915-25 (2003)第 922 页的表 1。c-MET 拮抗剂在所有这些癌症中都具有治疗价值。更具体地，c-MET 拮抗剂可以用于满足下述疾病的明显未满足的医学需要：胰腺癌、间皮瘤、骨髓瘤、头和颈癌、肺(NSCLC)癌、卵巢癌、乳腺癌、前列腺癌、结肠癌、成胶质细胞瘤和骨肉瘤。其它示例性的癌症包括膀胱、乳房、颈的、结肠直管的、食管的、胃的、肾、肝、肺、鼻咽的、胆囊、前列腺或甲状腺癌、骨肉瘤、滑膜肉瘤、横纹肌肉瘤、MFH/纤维肉瘤、卡波西肉瘤、多发性骨髓瘤、淋巴瘤、成人 T-细胞白血病、成胶质细胞瘤、星形细胞瘤、黑素瘤和 Wilm 氏肿瘤。

可以治疗个体,例如,通过每周一次地静脉内注射由结合 c-MET 的本发明的单体域或多聚体组成的 c-MET 拮抗剂的可溶制剂,任选地与一种或多种其它的治疗实体(例如生物的或化疗的)相组合。

在本发明的一个方面,在离体方法中,从受试者得到或取出一个或多个目标细胞或细胞群体(例如,肿瘤细胞,肿瘤组织样品,器官细胞,血细胞,皮肤、肺、心脏、肌肉、脑、粘膜、肝、肠、脾、胃、淋巴系统、子宫颈、阴道、前列腺、口、舌等的细胞),并与在预防上或治疗上有效地治疗疾病、障碍或其它状况的量的本发明的选择的单体域和/或多聚体接触。然后,将接触过的细胞返回或送递到受试者的得到它们的部位或要治疗的受试者的另一个目标部位(例如,包括上面定义的那些)。如果需要,可以使用标准的和众所周知的移植技术,将接触过的细胞移植到受试者的目标组织、器官或系统部位(包括上面定义的所有的)上,或者例如,使用标准的送递或输液技术,送递到血液或淋巴系统。

本发明也提供了体内方法,其中使受试者的一个或多个目标细胞或细胞群体直接或间接地接触在预防上或治疗上有效地治疗疾病、障碍或其它状况的量的本发明的选择的单体域和/或多聚体。在直接接触/施用格式中,一般地通过多种格式中的任一种,包括局部施用,注射(例如,使用针或注射器),或疫苗或基因枪送递,推入组织、器官或皮肤部位,将选择的单体域和/或多聚体直接施用或转移给要治疗的细胞或目标组织部位(例如,肿瘤细胞,肿瘤组织样品,器官细胞,血细胞,皮肤、肺、心脏、肌肉、脑、粘膜、肝、肠、脾、胃、淋巴系统、子宫颈、阴道、前列腺、口、舌等的细胞)。可以通过下述方式送递选择的单体域和/或多聚体,例如,肌内地、皮内地、真皮下地(subdermally)、皮下地、经口地、腹膜内地、鞘内地、静脉内地或置于体腔内(包括,例如,在外科手术过程中)或通过吸入或阴道或直肠施用。

在体内间接接触/施用格式中,一般地通过接触本发明的多肽,或将其直接施用给一个或多个细胞或细胞群体,将选择的单体域和/或多聚体间接施用或转移给要治疗的细胞或目标组织部位,包括上述的那些(例如,皮肤细胞、器官系统、淋巴系统或血细胞系统等),从而促进治疗。例如,通过使血液或淋巴系统、皮肤或器官的细胞接触

足够量的选择的单体域和/或多聚体，从而将选择的单体域和/或多聚体送递到目标部位(例如，目标组织、器官或细胞或身体内的血液或淋巴系统)，并导致有效的预防性或治疗性治疗，可以治疗受试者身体内的肿瘤细胞。一般地，使用一种或多种上述的施用途径或模式，进行这样的接触、施用或转移。

在另一个方面，本发明提供了离体方法，其中从受试者得到或取出一个或多个目标细胞或目标细胞群体(例如，肿瘤细胞，肿瘤组织样品，器官细胞，血细胞，皮肤、肺、心脏、肌肉、脑、粘膜、肝、肠、脾、胃、淋巴系统、子宫颈、阴道、前列腺、口、舌等的细胞)，并通过使所述的一个或多个细胞或细胞群体接触包含本发明的核酸序列的多核苷酸构建体，来进行转化，所述核酸序列编码生物活性的目标多肽(例如，选择的单体域和/或多聚体)，后者能在预防上或治疗上有效地治疗疾病、障碍或其它状况。使一个或多个细胞或细胞群体接触足够量的多核苷酸构建体和控制所述核酸序列的表达的启动子，从而发生将多核苷酸构建体(和启动子)向细胞中的摄入，且本发明的靶核酸序列的充分表达，导致生成在预防上或治疗上有效地治疗疾病、障碍或状况的量的生物活性的多肽，其编码选择的单体域和/或多聚体。多核苷酸构建体可以包括控制本发明的核酸序列的表达的启动子序列(例如，CMV 启动子序列)，和/或，如果需要，一个或多个额外的核苷酸序列，其编码至少一种或多种本发明的另一种多肽、细胞因子、佐剂或共同刺激分子或其它目标多肽。

转染后，将转化的细胞返回、送递或转移到受试者的得到它们的组织部位或系统或受试者的要治疗的另一个部位(例如，肿瘤细胞，肿瘤组织样品，器官细胞，血细胞，皮肤、肺、心脏、肌肉、脑、粘膜、肝、肠、脾、胃、淋巴系统、子宫颈、阴道、前列腺、口、舌等的细胞)。如果需要，可以使用标准的和众所周知的移植技术，将细胞移植到受试者的目标组织、皮肤、器官或身体系统，或者使用标准的送递或输液技术，送递到血液或淋巴系统。一般地，使用一种或多种上述的施用途径或模式，进行转化的细胞的这样的送递、施用或转移。靶核酸的表达会天然地发生，或可以被诱导(如下面更详细地描述的)，并表达足够的和有效的量的编码的多肽，以治疗该部位或组织系统的疾病或状况。

在另一个方面，本发明提供了体内方法，其中通过使细胞或细胞群体接触(或使用一种或多种上述的施用途径或模式，向细胞或细胞群体施用或转移)多核苷酸构建体，所述构建体包含本发明的核酸序列，后者编码目标生物活性的多肽(例如，选择的单体域和/或多聚体)，其在预防上或治疗上有效地治疗疾病、障碍或其它状况，将受试者的一个或多个目标细胞或细胞群体(例如，包括上述的那些细胞和细胞系统和受试者)转化进受试者的身体。

可以将多核苷酸构建体直接施用或转移给患有疾病或障碍的细胞(例如，通过使用一种或多种上述的施用途径或模式，直接接触)。或者，可以如下将多核苷酸构建体间接施用或转移给患有疾病或障碍的细胞：首先，使用一种或多种上述的施用途径或模式，使无疾病的细胞或其它有疾病的细胞直接接触足够量的多核苷酸构建体，所述构建体包含编码生物活性的多肽的核酸序列和控制该核酸序列的表达的启动子，从而发生将多核苷酸构建体(和启动子)向细胞中的摄入，且本发明的核酸序列的充分表达，导致生成在预防上或治疗上有效地治疗疾病或障碍的量的生物活性的多肽，且其中多核苷酸构建体或得到的表达的多肽从受试者身体的最初的送递部位、系统、组织或器官天然地或自动地转移到受试者身体的患病的部位、组织、器官或系统(例如，通过血液或淋巴系统)。靶核酸的表达会天然地发生，或可以被诱导(如下面更详细地描述的)，从而表达的多肽的量是足够的且有效的，以治疗该部位或组织系统的疾病或状况。多核苷酸构建体可以包括控制核酸序列的表达的启动子序列(例如，CMV 启动子序列)，和/或，如果需要，一个或多个额外的核苷酸序列，其编码至少一种或多种本发明的另一种多肽、细胞因子、佐剂或共同刺激分子或其它目标多肽。

在每种如上所述的体内和离体治疗方法中，可以施用或送递包含赋形剂和本发明的多肽或核酸的组合物。在一个方面，以有效地治疗疾病或障碍的量，将包含药学上可接受的赋形剂和本发明的多肽或核酸的组合物施用或送递给如上所述的受试者。

在另一个方面，在每种如上所述的体内和离体治疗方法中，施用给细胞或受试者的多核苷酸的量可以是这样的量，其使得发生所述多核苷酸向受试者的一个或多个细胞中的摄入，且所述核酸序列的充分

表达, 导致生成有效地增强受试者的免疫反应的量的生物活性的多肽, 所述反应包括由免疫原(例如, 抗原)诱发的免疫反应。在另一个方面, 对于每种这样的方法, 施用给细胞或受试者的多肽的量可以是这样的量, 其足以增强受试者的免疫反应, 其包括由免疫原(例如, 抗原)诱发的免疫反应。

在另一个方面, 在体内或体内治疗方法中, 其中使用多核苷酸构建体(或包含多核苷酸构建体的组合物)向受试者送递生理活性的多肽, 通过使用可诱导的开-和关-基因表达系统, 可以诱导多核苷酸构建体的表达。这样的开-和关-基因表达系统的实例分别包括 Tet-On™ 基因表达系统和 Tet-Off™ 基因表达系统(关于每种这样的系统的详细描述, 见, 例如, Clontech Catalog 2000, 第 110-111 页)。其它可控制的或可诱导的开-和关-基因表达系统是本领域的普通技术人员已知的。利用这样的系统, 可以以精确的、可逆的和定量的方式, 调节多核苷酸构建体的靶核酸的表达。可以诱导靶核酸的基因表达, 例如, 在将稳定的转染的细胞(其含有包含靶核酸的多核苷酸构建体)送递或转移到或接触目标组织部位、器官或系统后。在延迟或精确地控制靶核酸的表达(例如, 以留出时间完成外科手术和/或外科手术后愈合; 以留出时间使包含靶核酸的多核苷酸构建体达到要治疗的部位、细胞、系统或组织; 以留出时间使含有用构建体转化的细胞的移植物整合进它已经剪接或附着在其上或其中的组织或器官, 等)是有利的治疗方法和格式中, 这样的系统是特别有益的。

VIII. 其它多聚体用途

本发明的多聚体的潜在应用是多样的, 且包括需要亲和试剂的任何用途。

在有些情况下, 选择会结合相同靶的一对单体或多聚体(即, 用于基于夹心的测定中)。为了选择匹配的单体或多聚体对, 2种不同的单体或多聚体一般地能同时结合靶蛋白。鉴别这样的对的一种方法包含下面的内容:

- (1) 固定化以前选择的结合靶蛋白的噬菌体或蛋白混合物;
- (2) 使靶蛋白接触固定化的噬菌体或蛋白, 并洗涤;
- (3) 使噬菌体或蛋白混合物接触结合的靶, 并洗涤; 和

(4) 洗脱结合的噬菌体或蛋白，而不洗脱固定化的噬菌体或蛋白。

本发明的多聚体或单体域的一种用途是，用于替代在检测或其它基于亲和的测定中的抗体或其它亲和试剂。因而，在有些实施方案中，针对结合混合物中的靶以外的组分的能力，选择单体域或多聚体。一般方法可以包括，在与测定条件非常相似的条件，包括在测定过程中模仿样品的组成，进行亲和选择。因而，一个选择步骤可包括，使单体域或多聚体接触不包含靶配体的混合物，和选择结合混合物的任何单体域或多聚体。因而，代表这测定中的样品(血清、血液、组织、细胞、尿、精液等)的混合物(缺少靶配体，其可以使用抗体、单体域或多聚体排除)，可以用作阻断剂。这样的减法可以用于，例如，生成药物蛋白，其结合它们的靶，但是不会结合其它的血清蛋白或非靶组织。

例如，本发明可以用于生成拮抗剂，其中选择的单体域或多聚体阻断2种蛋白(例如，Met的 α 和 β 链)之间和/或Met和HGF之间的相互作用。任选地，本发明可以生成激动剂。例如，结合2种不同的蛋白(例如，酶和底物)的多聚体可以增强蛋白功能，包括，例如，酶活性和/或底物转化。

在有些实施方案中，单体域用于配体抑制、配体清除或配体刺激。这些方法中可行的配体包括，例如，HGF。

如果希望抑制配体与受体的结合，选择这样的单体域，其在接触配体的受体的配体部分结合配体(例如，HGF)，或在接触配体的受体部分结合受体，从而阻止配体-受体相互作用。如果需要，可以将单体域任选地连接到半衰期延长剂上。

配体清除指调节体液中的可溶配体的半衰期。例如，大多数单体域，在没有半衰期延长剂的情况下，具有短的半衰期。因而，单体域与配体的结合，会减小配体的半衰期，从而通过从肾清除配体，来降低配体浓度，只要复合物不大于能穿过肾的最大尺寸(小于约50或40kD)。单体域结合的配体(例如，HGF)部分通常无关紧要，尽管它可以有益地在结合它的受体(例如，Met)的配体部分结合配体，从而进一步抑制配体的作用。该方法可以用于降低血流中的任何分子的浓度。

或者,可以使用多聚体来增加配体的半衰期,所述多聚体包含结合半衰期延长剂的第一个单体域和结合不会结合配体的受体的配体部分的第2个单体域。

在另一个实施方案中,可以使用多聚体来增加配体对受体的有效亲和力,所述多聚体包含结合配体的第一个单体域和结合受体的第2个单体域。

在另一个实施方案中,使用包含至少2个结合受体的单体的多聚体,以通过使二者结合多聚体,使2个受体接近,从而激活受体。

本发明的潜在用途的其它实例包括,单体域和其多聚体,其能药物结合(例如,为寻靶的放射性核苷酸(radionucleotide)结合、为延长药物半衰期的药物结合、为超剂量治疗和成瘾治疗的受控制的物质结合)、免疫功能调节(例如,通过结合诸如 CTLA-4 的受体的免疫原性封闭,通过结合诸如 CD80 的受体的免疫原性增强,或通过 Fc 型结合的补体激活)和专门的送递(例如,通过接头切割的缓慢释放、电转运域、二聚化域或与下述物质的特异性结合:细胞进入域,清除受体例如 FcR,经口送递受体例如用于跨粘膜运输的 pIgR,和血-脑转移受体例如运铁蛋白 R)。

在其它实施方案中,可以将单体或多聚体连接到检测标记(例如, Cy3、Cy5 等)或连接到报告基因产物(例如, CAT、萤光素酶、辣根过氧化物酶、碱性磷酸酶、GFP 等)。

本发明的结合 Met 的单体或多聚体也可以用于其中检测 Met 是有用的诊断和预测应用。例如, Met 的检测可以用于预测乳腺癌的预后,其中比正常组织更高丰度的 Met 指示着预后不良。见,例如,美国专利号 6,673,559。

IX. 进一步操纵单体域和/或多聚体核酸和多肽

如上所述,可以改变本发明的多肽。在本文和其中引用的文献中,描述了用于产生修饰的或改变的编码这些多肽的核酸序列的多种多样性产生方法。

本发明的另一个方面包括单体域、选择的单体域、多聚体和/或选择的多聚体编码核酸的克隆和表达。因而,使用本领域众所周知的表达系统,可以合成作为单个蛋白的多聚体域。描述在本文中有用的

分子生物学技术(包括载体、启动子和许多与表达核酸有关的其它主题的使用,所述主题例如单体域、选择的单体域、多聚体和/或选择的多聚体)的一般文件包括, Berger 和 Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, Calif. (Berger); Sambrook 等, Molecular Cloning--A Laboratory Manual (第 2 版), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 (“Sambrook”)和 Current Protocols in Molecular Biology, F. M. Ausubel 等,编, Current Protocols, a joint venture between Greene Publishing Associates, Inc.和 John Wiley & Sons, Inc., (supplemented through 1999) (“Ausubel”)。足以指导本领域的人员,通过体外扩增方法用于鉴别、分离和克隆单体域和多聚体编码核酸的技术的实例,包括聚合酶链反应(PCR)、连接酶链反应(LCR)、Q-复制酶扩增和其它 RNA 聚合酶介导的技术(例如, NASBA),见 Berger, Sambrook, 和 Ausubel,以及 Mullis 等, (1987) 美国专利号 4,683,202; PCR Protocols A Guide to Methods and Applications (Innis 等编) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (Oct. 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3, 81-94; (Kwoh 等 (1989) Proc. Natl. Acad. Sci. USA 86, 1173; Guatelli 等 (1990) Proc. Natl. Acad. Sci. USA 87, 1874; Lomell 等 (1989) J. Clin. Chem 35, 1826; Landegren 等, (1988) Science 241, 1077-1080; Van Brunt (1990) Biotechnology 8, 291-294; Wu 和 Wallace, (1989) Gene 4, 560; Barringer 等 (1990) Gene 89, 117,和 Sooknanan 和 Malek (1995) Biotechnology 13: 563-564。改良的体外克隆扩增的核酸的方法记载在 Wallace 等, 美国专利号 5,426,039。改良的通过 PCR 扩增核酸的方法,总结在 Cheng 等 (1994) Nature 369: 684-685 和其中的文献中,其中产生最多达 40 kb 的 PCR 扩增子。技术人员会明白,使用逆转录酶和聚合酶,可以将基本上任意的 RNA 转化成适于限制消化、PCR 扩增和测序的双链 DNA。见, Ausubel, Sambrook 和 Berger, 都同上。

本发明也涉及,本发明的载体向宿主细胞中的导入,和通过重组技术,生产本发明的单体域、选择的单体域、多聚体和/或选择的多

聚体。用本发明的载体，其可以是，例如，克隆载体或表达载体，基因工程改造(即，转导、转化或转染)宿主细胞。载体可以是，例如，质粒、病毒颗粒、噬菌体等的形式。可以在改进成适用于激活启动子、选择转化体、或扩增目标单体域、选择的单体域、多聚体和/或选择的多聚体基因的常规营养培养基中，培养工程改造的宿主细胞。培养条件，例如温度、pH等，是选择的用于表达的宿主细胞以前使用的那些，且对本领域的技术人员，且在本文引用的文献中是显而易见的，所述文献包括，例如，Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, 第3版, Wiley-Liss, New York 和其中引用的文献。

如上所述，也可以在非动物细胞中生产本发明的多肽，例如植物、酵母、真菌、细菌等。实际上，如到处指出的，噬菌体展示是特别相关的生产这样的多肽的技术。除了 Sambrook, Berger 和 Ausubel 以外，关于细胞培养的细节可以参见 Payne 等 (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, N.Y.; Gamborg 和 Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual*, Springer-Verlag (Berlin Heidelberg New York) 和 Atlas 和 Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL.

本发明也包括改变单体域、免疫域和/或多聚体来提高药理学性质、降低免疫原性或促进多聚体和/或单体域向细胞或组织中的运输(例如，通过血-脑屏障，或通过皮肤)。这些类型的改变包括多种修饰(例如，添加糖基团或糖基化)、添加 PEG、添加结合某些蛋白(例如，HSA 或其它血清蛋白)的蛋白域、添加指导向细胞内、外或穿过细胞运动或运输的蛋白片段或序列。也可以将其它组分添加到多聚体和/或单体域上，以操纵多聚体和/或单体域的性质。也可以添加多种组分，其包括，例如，结合已知受体的域(例如，结合 Fc 受体的 Fc-区蛋白域)，毒素或毒素的部分，可以任选地被切割掉以激活多聚体或单体域的前域(prodomain)，报告分子(例如，绿色荧光蛋白)，结合报告分子(例如用于放射疗法的放射性核素、生物素或抗生物素蛋白)的组分或修饰的组合。

X. 动物模型

本发明的另一个方面是，开发特定的非人动物模型，在其中测试单体或多聚体域的免疫原性。生产这样的非人动物模型的方法包含：向受体非人动物的至少一些细胞中，导入包含编码许多来自相同蛋白家族的人蛋白的基因的载体，其中基因各自可操作地连接到在至少一些导入了载体的细胞中有功能的启动子上，从而得到基因修饰的非人动物，其可以表达许多来自相同蛋白家族的人蛋白。

在本发明的实践中采用的合适的非人动物包括除人以外的所有脊椎动物(例如，小鼠、大鼠、兔、羊等)。一般地，蛋白家族的许多成员包括该家族的至少 2 个成员，且通常至少 10 个家族成员。在有些实施方案中，许多包括蛋白家族的所有已知的成员。示例性的可以使用的基因包括编码单体域的那些，例如，LDL 受体 A 类-域家族、EGF-样域家族，以及本文所述的其它域家族的成员。

本发明的非人动物模型可以用于筛选源自由该非人动物模型表达的相同蛋白家族的单体或多聚体域的免疫原性。本发明包括根据上述方法制备的非人动物模型，以及转基因非人动物，其体细胞和生殖细胞含有和表达编码许多来自相同蛋白家族的人蛋白(例如本文所述的单体域)的 DNA 分子，其中 DNA 分子已经导入处于胚胎期的转基因的非人动物，且其中 DNA 分子各自可操作地连接到至少一些在其中已经导入了 DNA 分子的细胞的启动子上。

下面描述了用于筛选 LDL 受体 A 类-域衍生的结合蛋白的小鼠模型的实例。使用 PCR，从人细胞扩增编码野生型人 LDL 受体 A 类-域单体的基因簇。仅用 3 个分开的 PCR 扩增反应，每个约 7 kb，就可以扩增几乎所有 200 种不同的 A-域。然后，根据上述的方法，使用这些片段，以产生转基因小鼠。转基因小鼠会将人 A-域识别为“自身的”，从而模仿人关于 A-域的“自身性(selfness)”。通过将 A-域-衍生的单体或多聚体注射进小鼠，然后分析产生的免疫反应(或反应的缺失)，可以在这些小鼠中测试单个的 A-域-衍生的单体或多聚体。测试小鼠，以确定它们是否已经发展了小鼠抗-人反应(MAHR)。当施用给人时，没有导致 MAHR 的产生的单体和多聚体可能是非免疫原性的。

历史上，使用在转基因小鼠中的 MAHR 试验来测试小鼠中的单个的蛋白，所述小鼠对该单个蛋白是转基因的。相反地，上述的方法提供了非人动物模型，其能将整个人蛋白家族识别为“自身的”，且其可以用于评价大量的变体蛋白，其中的每种能极大地改变结合活性和用途。

XI. 试剂盒

包含该方法所需组分(一般地，以未混合的形式)和用于容纳该组分的试剂盒组分(包装材料、关于使用组分和/或方法的说明书、一个或多个容器(反应管、柱等))的试剂盒，是本发明的一个特征。本发明的试剂盒可以含有多聚体文库，或单一类型的单体或多聚体。试剂盒也可以包含适用于促进靶分子结合的试剂，例如有利于检测的缓冲剂或试剂，包括可检测地标记的分子。校准配体与单体域等的结合的标准，也可以包含在本发明的试剂盒中。

本发明也提供了商业上有价值的结合测定和试剂盒，以实践测定。在有些本发明的测定中，采用一种或多种配体来检测单体域、免疫域和/或多聚体的结合。这样的测定是基于本领域已知的任意的方法的，例如，流式细胞仪、荧光显微镜术、等离子体共振等，以检测配体与单体域和/或多聚体的结合。

也提供了基于测定的试剂盒。试剂盒一般地包含容器和一种或多种配体。试剂盒任选地包含关于进行测定的指导，额外的检测试剂，缓冲剂，或关于使用这些组分中的任一种的说明书，等。或者，试剂盒可以包括用于表达本发明的单体域和/或多聚体的细胞、载体(例如，包含本发明的多肽的表达载体，分泌载体)。

在另一个方面，本发明提供了本文的任意的组合物、单体域、免疫域、多聚体、细胞、细胞培养物、装置、装置组分或试剂盒在实践本文的任意方法或测定中的用途，和/或任意的装置或试剂盒在实践本文的任意测定或方法中的用途，和/或本文的细胞、细胞培养物、组合物或其它特征作为治疗制剂的用途。也提供了作为用于本文所述的治疗的治疗制剂的本文的所有组分的生产。

XII. 集成系统

本发明提供了计算机、计算机可读介质和集成系统，后者包含与单体域、选择的单体域、多聚体和/或选择的多聚体和编码这样的多肽的核酸相对应的字符串。通过计算机重组方法，或通过标准的序列比对或字处理软件，可以操纵这些序列。

例如，可以在本文的集成系统中，检测和识别不同类型的相似性和各种严格性和字符串长度的考虑。例如，已经设计了许多同源性测定方法，以用于生物聚合物序列的对比分析，字处理的拼写检查，和从各个数据库的数据检索。随着对天然多核苷酸中的4种基本核碱基之间的双螺旋逐对补体相互作用的理解，也可以将模仿互补同源多核苷酸串的退火的模型用作序列比对或在与本文的序列相对应的字符串上一般地进行的其它操作(例如，字处理操作，包含序列或子序列字符串的图的构建，输出表，等)的基础。具有用于计算序列相似性的GO的软件包的实例是BLAST，通过输入与本文的序列相对应的字符串，可以使它适用于本发明。

BLAST记载在 Altschul 等, (1990) *J. Mol. Biol.* 215: 403-410。从国家生物技术信息中心(在环球网上可从 ncbi.nlm.nih.gov 得到)，可以公开地得到用于进行BLAST分析的软件。该算法包含，首先通过鉴别查询序列中长为W的短字，鉴别高评分序列对(HSP)，当与数据库序列中的相同长度的字相比时，所述HSP匹配或满足一些正值的阈分数T。T称作邻近字分数阈值(Altschul等，同上)。这些最初的邻近字命中值作为开始搜索的种子，以发现含有它们的更长的HSP。字命中值沿着每个序列向两个方向延伸，只要可以增加累积比对分数。对于核苷酸序列，使用参数M(一对匹配残基的奖励分数；总是 >0)和N(错配残基的惩罚分数；总是 <0)，计算累积分数。对于氨基酸序列，使用评分矩阵计算累积分数。当下述情况时，停止每个方向的字命中值的延伸：累积比对分数从它的最大实现值下降了量X；由于一个或多个负评分残基比对的累积，累积分数达到0或以下；或达到任一个序列的末端。BLAST算法参数W、T和X决定了比对的灵敏度和速度。BLASTN程序(对于核苷酸序列)使用的默认值是，字长(W)为11，期望(E)为10，截止为100，M=5，N=-4，且对比2条链。对于氨基酸序列，BLASTP程序使用的默认值是，字长(W)为3，期望(E)为10，和BLOSUM62评分矩阵(见Henikoff和Henikoff

(1989) Proc. Natl. Acad. Sci. USA 89: 10915).

有用的序列比对算法的另一个实例是 PILEUP。PILEUP 使用渐进的逐对比对，从一组相关的序列产生多个序列比对。也可以绘制显示用于产生比对的成簇关系的树。PILEUP 使用 Feng & Doolittle, (1987) J. Mol. Evol. 35: 351-360 的渐进比对方法的简化。使用的方法类似于 Higgins & Sharp, (1989) CABIOS 5: 151-153 所述的方法。程序可以比对，例如，最多达 300 个最大长度为 5,000 个字母的序列。多重比对方法从 2 个最相似的序列的逐对比对开始，从而生成 2 个比对的序列簇。然后，可以将该簇与下一个最相关的序列或比对的序列簇相比对。通过 2 个单独的序列的逐对比对的简单延长，可以比对 2 族序列。通过一系列的渐进的逐对比对，实现最终比对。该程序也可以用于绘制成簇关系的系统树(dendogram)或树表示。通过为序列对比区域指定特定的序列和它们的氨基酸或核苷酸坐标，运行该程序。例如，为了确定单体域家族中的保守氨基酸或对比家族中的单体域序列，比对本发明的序列或编码核酸，以提供结构-功能信息。

在一个方面，使用计算机系统与单体域相对应的字符串的“计算机”序列重组或改组。多种这样的方法记载在 Selifonov 和 Stemmer 于 1999 年 2 月 5 日提交的“Methods For Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics” (USSN 60/118854)和 Selifonov 和 Stemmer 于 1999 年 10 月 12 日提交的“Methods For Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics” (USSN 09/416,375)。简而言之，在遗传算法中使用遗传操作符来改变给定的序列，例如，通过模仿遗传事件，例如突变、重组、死亡等。也可以在计算机系统中进行最优化序列的多维分析，例如，如 ‘375 申请所述。

数字系统也可以指导寡核苷酸合成仪合成寡核苷酸，例如，用于基因重构或重组，或从商业来源订购寡核苷酸(例如，通过打印适当的订货单或通过因特网上连接到订货单)。

数字系统也可以包括控制核酸合成的输出元件(例如，基于重组体的序列或比对，例如，本文所述的重组的单体域)，即，本发明的集成系统任选地包括寡核苷酸合成仪或寡核苷酸合成控制器。系统可

以包括发生在比对下游的其它操作,或使用与本文的序列相对应的字符串进行的其它操作,例如,如上面参考测定所指出的。

实施例

下面的实施例用于解释而不是限制要求保护的发明。

实施例 1

本实施例描述了单体域的选择和多聚体的生成。

从多种人和/或非人序列的任一种,可以衍生出从选择的单体域和方法鉴别单体域和生成多聚体的原料。例如,为了生产具有对所需配体或配体混合物的特异性结合的选择的单体域,从结合某些配体的单体域家族选择一个或多个单体域基因。通过基因组 DNA 或 cDNA 的 PCR 扩增,可以得到编码一个或多个单体域基因的核酸序列,或任选地,可以使用重叠寡核苷酸合成地生产所述核酸序列。

最常见地,然后将这些序列克隆进用于表达和筛选的细胞表面展示格式(即,细菌、酵母或哺乳动物(COS)细胞表面展示;噬菌体展示)。将重组序列转染(转导或转化)进适当的宿主细胞,它们在其中表达,并展示在细胞表面上。例如,可以用标记的(例如,荧光标记的)所需配体对细胞染色。通过流式细胞仪,分选染色的细胞,并从阳性细胞回收(例如,通过质粒分离、PCR 或扩增和克隆)选择的单体域编码基因。可以重复染色和分选过程多次(例如,使用递减浓度的所需配体,直到得到所需的富集水平)。或者,可以采用本领域已知可用于鉴别结合所需配体或配体混合物的细胞的任何筛选或检测方法。

根据本文或引用的文献所述的方法中的任一种,可以任选地重组从所需配体或配体混合物结合细胞回收的选择的单体域编码基因。然后,通过相同的或不同的方法,筛选在该多样化循环中生成的重组序列,以鉴别具有提高的对所需的或靶配体的亲和力的重组基因。任选地重复多样化和选择过程,直到得到所需的亲和力。

通过接头序列,可以将通过该方法选择的选择的单体域核酸连接到一起,以生成多聚体,例如,通过 DNA 连接,或任选地,基于 PCR 的自引发的重叠反应,组合装配编码选择的单体域的核酸序列。然后,将编码多聚体的核酸序列克隆进用于表达和筛选的细胞表面展示格式(即,细菌、酵母或哺乳动物(COS)细胞表面展示;噬菌体展示)。

将重组序列转染(转导或转化)进适当的宿主细胞,它们在其中表达,并展示在细胞表面上。例如,可以用标记的(例如,荧光标记的)所需配体或配体混合物对细胞染色。通过流式细胞仪,分选染色的细胞,并从阳性细胞回收(例如,通过 PCR 或扩增和克隆)选择的多聚体编码基因。阳性细胞包含与选择的单体域相比,具有提高的对所需配体或配体混合物的亲和力或抗体亲抗原性或改变的特异性的多聚体。可以重复染色和分选过程多次(例如,使用递减浓度的所需配体或配体混合物,直到得到所需的富集水平)。或者,可以采用本领域已知可用于鉴别结合所需配体或配体混合物的细胞的任何筛选或检测方法。

根据本文或引用的文献所述的方法中的任一种,可以任选地重组从所需配体或配体混合物结合细胞回收的选择的多聚体编码基因。然后,通过相同的或不同的方法,筛选在该多样化循环中生成的重组序列,以鉴别具有提高的对所需的或靶配体的亲和力或抗体亲抗原性或改变的特异性的重组基因。任选地重复多样化和选择过程,直到得到所需的亲和力或抗体亲抗原性或改变的特异性。

实施例 2

本实施例描述了生成更大多样性的文库的体内蛋白内重组

通过它的相容的 *loxP* 位点,以 Cre-依赖性的方式,将由直向同源 *loxP* 位点侧接的单体-编码质粒载体(pCK-衍生的载体;见下面)与噬菌体载体重组。使用对重组的构建体特异性的引物,通过 PCR 检测重组的噬菌体载体。DNA 测序表明,产生了正确的重组产物。

试剂和实验方法

pCK-cre-lox-单体-loxP。该载体具有 2 个特别相关的特征。首先,它携带在 P_{lac} 控制下的 *cre* 基因,其编码位点特异性的 DNA 重组酶 Cre。用 *cre*-特异性的引物,从 p705-*cre* (购自 GeneBridges) PCR-扩增 *cre*, 其在 PCR 产物的末端整合 *Xba*I (5')和 *Sfi*I (3')。用 *Xba*I 和 *Sfi*I 消化该产物,并克隆进 pCK (pCK110919-HC-Bla (pACYC ori) 的 *bla*⁻, Cm^R 衍生物) 的相同位点,以生成 pCK-*cre*。

第二个特征是首次用于实验的 A 域文库,其由 2 个直向同源 *loxP*

位点 *loxP*(野生型)和 *loxP*(FAS)侧接, 所述位点是 Cre 催化的位点特异性的 DNA 重组所必需的。见, 例如, Siegel, R. W.,等 *FEBS Letters* 505: 467-473 (2001)。这些位点很少彼此重组。先后将 *loxP* 位点构建进 pCK-cre。将携带 *loxP*(WT)和 *EcoRI* 和 *HindIII*-相容的允许连接到消化的 *EcoRI* 和 *HindIII*-消化的 pCK 上的突出端的 5'-磷酸化的寡核苷酸 *loxP*(K)和 *loxP*(K_rc)在一起杂交, 并在标准的连接反应中连接到 pCK-cre 上(T4 连接酶; 在 16°C 过夜)。

用 *EcoRI* 和 *SphI* 消化得到的质粒, 并连接到杂交的、5'-磷酸化的寡核苷酸(oligos) *loxP*(L)和 *loxP* (L_rc)上, 其携带 *loxP*(FAS)和 *EcoRI* 和 *SphI*-相容的突出端。为了准备文库构建, 根据 Qiagen 的规程, 进行 pCK-cre-*loxP*(wt)-*loxP*(FAS)的大规模纯化(Qiagen MAXI prep)。对 Qiagen-纯化的质粒进行 CsCl 梯度离心, 以用于进一步纯化。然后, 用 *SphI* 和 *BglII* 消化该构建体, 并连接到消化的首次用于实验的 A 域文库插入片段上, 后者通过预先存在的 A 域文库的库的 PCR-扩增得到。通过设计, *loxP* 位点和单体符合读框, 其产生具有 *loxP*-编码的接头的单体。在如下详述的体内重组方法中, 使用该文库。

fUSE5HA-单体-*loxP*-*loxP* 载体。该载体是来自 George Smith 的实验室(University of Missouri)的 fUSE5 的衍生物。随后, 将它修饰成携带用于免疫检测测定的 HA 标签。先后将 *loxP* 位点构建进 fUSE5HA。将携带 *loxP*(WT)、一串终止密码子和 *XmaI* 和 *SfiI*-相容的突出端的 5'-磷酸化的寡核苷酸 *loxP*(I)和 *loxP*(I)_rc 在一起杂交, 并在标准的连接反应中连接到 *XmaI*-和 *SfiI*-消化的 fUSE5HA 上(New England Biolabs T4 连接酶; 在 16°C 过夜)。

接着, 用 *XmaI* 和 *SphI* 消化得到的噬菌体载体, 并连接到杂交的寡核苷酸 *loxP*(J)和 *loxP*(J)_rc 上, 后者携带 *loxP*(FAS)和与 *XmaI* 和 *SphI* 相容的突出端。用 *XmaI/SfiI* 消化该构建体, 然后连接到预切割的(*XmaI/SfiI*) 首次用于实验的 A 域文库插入片段(PCR 产物)上。终止密码子位于 *loxP* 位点之间, 阻止 *gIII* 的表达, 并从而阻止感染性噬菌体的生成。

随后, 将连接的载体/文库转化进大肠杆菌宿主, 其携带 *gIII*-表达质粒, 所述质粒允许拯救 fUSE5HA-单体-*loxP*-*loxP* 噬菌体, 如下所

详述。

pCK-gIII. 该质粒携带 *gIII*，后者在它的天然启动子的控制下。通过用引物 *gIII* 启动子_ *EcoRI* 和 *gIII* 启动子_ *HinDIII* 从 VCSM13 辅助噬菌体(Stratagene)PCR-扩增 *gIII* 和它的启动子，来构建它。用 *EcoRI* 和 *HinDIII* 消化该产物，并克隆进 pCK110919-HC-Bla 的相同位点。由于 *gIII* 是在它自己的启动子的控制下的，所以 *gIII* 表达可能是组成型的。将 pCK-*gIII* 转化进大肠杆菌 EC100 (Epicentre)。

体内重组操作。总的来说，该操作包含下面的关键步骤：a) 用表达来自质粒的 *gIII* 的大肠杆菌宿主，产生 fUSE5HA-单体-lox-lox 文库的感染性（即，拯救）；b) 克隆第 2 个文库(pCK)，并转化进 F⁺ TG1 大肠杆菌；c) 用拯救的 fUSE5HA-单体-lox-lox 噬菌体文库，感染携带第 2 个文库的培养物。

a. 噬菌体载体的拯救。通过标准规程，制备携带 pCK-*gIII* 的电感受态(electrocompetent)细胞。这些细胞具有 $4 \times 10^8 / \mu\text{g}$ DNA 的转化频率，且用 fUSE5HA-lox-lox 载体和首次用于实验的 A 域文库插入片段的大规模连接(约 5 μg 载体 DNA)电穿孔。以约 70 μL 细胞/杯完成个别的电穿孔 (100 ng DNA/电穿孔)后，加入 930 μL 温 SOC 培养基，并在摇动下，在 37 $^{\circ}\text{C}$ ，使细胞恢复 1 小时。接着，加入四环素至 0.2 $\mu\text{g}/\text{mL}$ 的终浓度，并在 37 $^{\circ}\text{C}$ ，摇动细胞约 45 分钟。取出该培养物的等分试样，10 倍系列稀释，并铺平板，以测定得到的文库大小(1.8×10^7)。将剩余的培养物稀释至 2x 500 mL 2x YT (具有 20 $\mu\text{g}/\text{mL}$ 氯霉素和 20 $\mu\text{g}/\text{mL}$ 四环素，以分别选择 pCK-*gIII* 和基于 fUSE5HA 的载体)，并在 30 $^{\circ}\text{C}$ 生长过夜。

使用标准的 PEG/NaCl 沉淀规程，收获拯救的噬菌体。滴度是大约 1×10^{12} 转导单位/mL。

b. 克隆第 2 个文库和转化进大肠杆菌宿主。将连接的 pCK/首次用于实验的 A 域文库电穿孔进细菌 F⁺ 宿主，期望的文库大小为约 10^8 。在 37 $^{\circ}\text{C}$ ，在摇动下，经历 1 小时长的恢复时间段后，在 2x YT (加 20 $\mu\text{g}/\text{mL}$ 氯霉素)中将电穿孔的细胞稀释至 OD₆₀₀ 约 0.05，并在 37 $^{\circ}\text{C}$ 生长至对数中期，然后用 fUSE5HA-单体-lox-lox 感染。

c. 用拯救的 fUSE5HA-单体-lox-lox 噬菌体文库感染携带第 2 个文库的培养物。为了使重组体的产生最大化，培养物中高感染率

(>50%)的大肠杆菌是需要的。大肠杆菌的感染性依赖于许多因素，包括 F 菌毛的表达和生长条件。大肠杆菌背景 TG1 (携带 F')和 K91 (Hfr 菌株) 是重组系统的宿主。

寡核苷酸

loxP(K)

[P-5' agcttataacttcgtatagaaaggtatatacgaagttatagatctcgtgctgcatgcggtgcg]

loxP(K_rc)

[P-5' aattcgcaccgcatgcagcagagatctataacttcgtatatacctttctatacgaagtataagct]

loxP(L)

[P-5' ataacttcgtatagcatacattatacgaagttatcgag]

loxP (L_rc)

[P-5' ctcgataacttcgtataatgtatgctatacgaagttatg]

loxP(I)

[P5' ccgggagcagggcatgctaagtgagtaataagtgagtaaataacttcgtatatacctttctatacgaagttatcgtctg]

loxP(I)_rc

[P-5' acgataacttcgtatagaaaggtatatacgaagttattactcacttattactcacttagcatgccectgctc]

loxP(J)

[5' ccgggaccagtggcctctggggccataacttcgtatagcatacattatacgaagttatg]

loxP(J)_rc

[5' cataacttcgtataatgtatgctatacgaagttatggccccagaggccactggtc]

gIII 启动子 _EcoRI

[5' atggcgaattctcattgtcggcgcaactat

gIII 启动子 _HinDIII

[5' gataagcttcattaagactcctlattacgcag]

实施例 6

本实施例描述了基于 EGF 的单体文库的构建。

CaEGF 域文库 E3 编码具有下面模式的 36-43 个氨基酸的蛋白域：

X(5)C1-X(4/6)-C2-X(4,5)-C3-X(8)-C4-X(1)-C5-X(8/12)-C6

下表描述了每个位置，该氨基酸在基于人钙结合 EGF 域的天然

的多样性的文库中编码:

X(5)					X(4,6)						X(4,5)					X(8)								X(6,12)											
X1	X2	X3	X4	X5	X1	X2	X3	X4	X5	X6	X1	X2	X3	X4	X5	X1	X2	X3	X4	X5	X6	X7	X8	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
V	D	V	N	E	V	S	S	F			V	S	G	S		V	N	T	V	G	S	T	S	V	Q	G	Y	T	G	V	S				
T	I	D			S	P	G	A	D	A	T	R	D	R	I	Q	D	S	Q		G	F	R	S	P	P	F	S	L	T	R	A	G	K	A
I					L	E	N	S	G	I	S	N	N	Q	K	L	L	P						R	L	E		A	P	Q	N	D	R	N	D
K					A	L	D	H	L		R	H	S	N	N	K	I	L						Q	A	A		E	Q	P	L	N	S	R	E
E					D	Q	I	N	P		Q		L	R		I	E							N	H	K		K	R	L	I	T		S	G
A					E	V	V	R	T		P		K	B		E	A							M	N	T		L	K	H			T	K	
					G			S	V		N		I	T										L	R			M	I	A				N	
					K						M		H											K	S			P	E	E				R	
					N						L		A											I	T			Q	A	G				S	
					R						K		G											H				V	D	P				T	
					T						I		M											G	E			G	Q						
											M		G											E				N							
											G		E											D				S							
											E		D																						
											A																								

通过如 Stemmer 等, Gene 164: 49-53 (1995)所述的装配 PCR, 生成编码单体的钙结合 EGF 域的 DNA 序列文库 E3。在该 PCR 反应中使用的寡核苷酸为 2 组, 1 和 2。它们是:

组 1:

- 5'-AAAAGGCCTCGAGGGCCTGGGTGGCAATGGT-3'
- 5'-CCTGAACCACCACAKHKACCGYKSNBGCACGGAYYCGRCRMACATTCATYAAYATCTDYACCATTGCCACCC-3'
- 5'-CCTGAACCACCACAKNTGSCGYGYKMHSGCACGGAYYCGRCRMACATTCATYAAYATCTDYACCATTGCCACCC-3'
- 5'-CCTGAACCACCACAKHKACCGYKSNBGCARBAAYBCGVVAHYCWSKBYACATTCATYAAYATCTDYACCATTGCCACCC-3'
- 5'-CCTGAACCACCACAKNTGSCGYGYKMHSGCAARBAAYBCGVVAHYCWSKBYACATTCATYAAYATCTDYACCATTGCCACCC-3'
- 5'-TGAATTTTCTGTATGAGGTTTTGCTAAACAACCTTCAACAGTTTCGGCCCAGAGGCCCTGGAGCCACCTGAACCACCACA-3'

组 2:

- 5'-ACGGTGCCTACCCGTATGATGTTCCGGATTATGCCCCGGGTGGCAATGGT-3'
- 5'-CCTGAACCACCACAGHKTDACCAGHAWAGCCTKSCRS GCASHBACAKYKAWAGCYACCCDSTRWATYTWBACCATTGCCACCC-3'
- 5'-CCTGAACCACCACAKBYKBTCKYCYGKYCBSABYCNCGDBAWAGCCTKBGBKGCASHBACAKYKAWAGCYACCCDSTRWATYTWBACCATTGCCACCC-3'
- 5'-AAAAGGCCCCAGAGGCCCTGAACCACCACA-3'

其中 R=A/G, Y=C/T, M=A/C, K=G/T, S=C/G, W=A/T, B=C/G/T, D=A/G/T, H=A/C/T, V=A/C/G, 且 N=A/C/G/T.

组 1 和 2 寡核苷酸的分开的 PCR 后, 用 BpmI 消化组 1 PCR 片

段,并用 BsrDI 消化组 2 PCR 片段。使用 Qiagen Qiaquick 柱纯化消化产物,然后连接到一起。然后,使用 2 种引物,在 PCR 中扩增连接的 DNA。它们是:

5'-AAAAGGCCTCGAGGGCCTGGGTGGCAATGGT-3'
5'-AAAAGGCCCCAGAGGCCCTGAACCACCACA-3'

用 Qiagen Qiaquick 柱纯化 PCR 产物,并用 *Sfi*I 消化。用 Qiagen Qiaquick 柱纯化消化产物。将 DNA 片段连接进噬菌体展示载体 fuse5-HA(G4S)₄ (携带符合读框的 HA-表位和甘氨酸、丝氨酸柔性的接头的 fuse5 的衍生物)的 *Sfi*I 限制位点。将连接混合物电穿孔进 TransforMax™ EC100™ 电感受态大肠杆菌细胞。转化的大肠杆菌细胞在含有 20μg/ml 四环素的 2x YT 培养基中在 37℃ 生长过夜。得到的文库含有 2X10⁹ 独立克隆。通过 PEG-沉淀,从培养基纯化噬菌体颗粒。噬菌体的滴度是 1.3X10¹²/ml。测定 24 个单个的克隆的序列,且它们与文库设计相一致。

实施例 3

本实施例描述了基于 EGF 的单体文库的构建。

重组可以用于域内最优化。例如,可以使用 PCR 重叠反应,其重组单一域中彼此相关的 2 个或更多个片段。人们可以以与所述的相同的方式,使用 2、3、4、5 或更多个片段重叠反应。该重组过程具有许多应用。一个应用是,重组数百个以前选择的没有序列信息的克隆的大库。每个重叠进行工作只需要存在于每个克隆的相同位置处的(相对地)恒定的序列的一个已知区域(固定位点方法)。对于 A 域,一般地,这些克隆已经源自这样的文库,其中分布在所有 5 个半胱氨酸间片段上的 20-25 个氨基酸是随机化的。通过基于随机断裂的标准的 DNA 重组 (例如,Stemmer, Nature 370: 389-391 (1994)) 和基于 DNA 序列同源性的重新组装,也可以在一组序列-相关的单体域上进行域内重组方法,其在所有要重组的克隆中不需要固定的重叠位点。

该方法的另一个应用是,生成多个分开的首次用于实验的(指未淘选的)文库,在其中的每一个中,仅一个半胱氨酸间环是随机化的,以随机化每个文库中的不同环。针对靶分别淘选这些文库后,再重组

选择的克隆。从每个淘选的文库中，通过 PCR 仅扩增随机化的片段，然后将多个随机化的片段组合进单个域，从而生成改组的文库，针对增加的效力对所述文库进行淘选和/或筛选。该过程也可以用于改组小量已知序列的克隆。

任意的共有序列可以用作交换点。对于 A 域或其它含有半胱氨酸的单体，半胱氨酸残基是合乎逻辑的交换位置。但是，存在其它确定最佳交换点的方式，例如计算机建模。或者，具有最高熵或最少数目的分子内接触的残基，也可以是好的交换点。

下面描述了产生包含具有随机化的半胱氨酸间环的蛋白的文库的示例性的方法。在该实例中，与上述的分开的环境、分开的文库方法不同，在相同文库中同时随机化多个半胱氨酸间环。

构建了 A 域 NNK 文库，其编码具有下面模式的 39-45 个氨基酸的蛋白域：

C1-X(4,6)-E1-F-R1-C2-A-X(2,4)-G1-R2-C3-I-P-S1-S2-W-V-C4-D1-G2-E2-D2-D3-C5-G3-D4-G4-S3-D5-E3-X(4,6)-C6;

其中，

C1-C6: 半胱氨酸；

X(n): 在每个位置具有任意残基的 n 个氨基酸的序列；

E1-E3: 谷氨酰胺；

F: 苯丙氨酸；

R1-R2: 精氨酸；

A: 丙氨酸；

G1-G4: 甘氨酸；

I: 异亮氨酸；

P: 脯氨酸；

S1-S3: 丝氨酸；

W: 色氨酸；

V: 缬氨酸；

D1-D5: 天冬氨酸；且

C1-C3、C2-C5 & C4-C6 形成二硫键。

通过如 Stemmer 等, Gene 164: 49-53 (1995)所述的装配 PCR, 生

使用 4 个寡核苷酸库的混合物，每库含有 400 pmol DNA，通过 PCR 扩增的 10 个循环的初始循环，构建文库。库 1 含有寡核苷酸 1-9，库 2 含有 10-17，库 3 仅含有 18，且库 4 含有 19-27。使用库 1 和 4，通过 PCR 的额外 8 个循环，得到完全装配的文库。用 *XmaI* 和 *SfiI* 消化文库片段。将 DNA 片段连接进噬菌体展示载体 fuse5-HA（携带符合读框的 HA-表位的 fuse5 的衍生物）的对应的限制位点。将连接混合物电穿孔进 TransforMax™ EC100™ 电感受态大肠杆菌细胞，从而产生 2×10^9 个单个克隆的文库。转化的大肠杆菌细胞在含有 20 μg/ml 四环素的 2x YT 培养基中在 37°C 生长过夜。通过 PEG-沉淀，从培养基纯化噬菌体颗粒，并测得 1.1×10^{13} /ml 的滴度。测定 24 个克隆的序列，且它们与文库设计的预期相一致。

实施例 4

本实施例描述了通过最优化单体和/或接头与靶的结合，最优化多聚体。

最优化结合靶的多聚体的一种方法包含单体、多聚体和接头的最优化。首先，针对与靶(例如，Met)的结合淘选单体文库。但是，有些单体可能结合在靶上彼此远离的位置，从而使得结合这些位点的域不能被接头肽连接。因此，在最优化单体之前，从这些单体生成和筛选同-或异三聚体的大文库是有用的。可以筛选这些三聚体文库，例如，在噬菌体上(一般地，对于从大单体库生成的异三聚体)，或分别制备和测定(例如，对于同三聚体)。通过该方法，鉴别出最佳的三聚体。测定可以包括与靶的结合测定或在功能性基于蛋白或细胞的测定中的多聚体的激动剂或拮抗剂效力测定。

然后，最优化单个最佳三聚体的单体域，作为第 2 步。同多聚体最容易最优化，因为仅存在一个域序列，尽管也可以合成异多聚体。对于同多聚体，与单体相比，多聚体的结合的增加是抗体亲抗原性效应。

最优化域序列自身(例如，通过重组或 NNK 随机化)和噬菌体淘选后，使用改良的单体来构建具有接头文库的二聚体。可以形成接头文库，例如，从具有 NNK 组成和/或可变序列长度的接头。

淘选该接头文库后，将最佳克隆(例如，通过抑制或其它功能测

定中的效力所测得的)转化成多聚体,其由多个(例如,2,3,4,5,6,7,8,等)序列-最优化的域和长度-和序列-最优化的接头组成。

实施例 5

本实施例描述了 A 域的结构分析。

实际上,如同所有的蛋白一样,仅仅 A-域的所有表面中的小部分参与结合单个靶。基于该域的溶液结构,可以鉴别邻近的残基位置,所述位置可能能在与给定的靶的结合中协同作用。在本文中,这样的邻近的残基组称作结构类。作为实例,已经通过检查 A-域结构,鉴别出了 4 个这样的类别,命名为顶、底、环 1 和环 2。通过设计仅仅允许给定类中的多样性的文库,可以显著减小文库允许的理论序列间隙,从而允许物理文库更好地覆盖理论间隙。而且,在无重叠类别(例如顶和底类别)的情况下,可以将针对不同靶选择的半-域序列合并成单个序列,其能同时或替代地结合选择的靶。在任一种情况下,生成仅占据半域的结合位点,允许生成这样的分子,其大小是一半,且具有一半的免疫原表位数,从而降低免疫原性的危险。

A-域位置的结构分类

下面显示了标准的 A-域序列,将高多样性位置表示为 X。用星号表示属于顶、底、环 1 或环 2 类别的位置。

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	C	X	X	X	X	F	X	C	X	X	X	X	C	I	X	X	X	W	X
顶	.	*	.	.	*	*	.	*	*	.	*	*	.
底	*	*	.	*	*	.	.	*
环1	.	*	*	*	*	*	*	*	.	.
环2	*	*	*	*

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	C	D	G	X	X	D	C	X	D	X	S	D	E	X	X	C
顶	.	*	*	*	*	*	.	*	.	*
底	*	.	*	.	.	*	*	.
环1
环2	*	*	*

实施例 6

本实施例描述了结合 c-MET(也称作 HGFR)的单体或多聚体的筛

选。

在固体支持物上(例如 Nunc Maxisorp 平板)或在溶液中(例如 Dynal 链霉抗生物素蛋白或 A 蛋白珠),通过几个循环,淘选噬菌体文库。选择输出噬菌体库,其具有(a)最高频率的结合 c-MET 的单个噬菌体克隆,和(b)结合-阳性的噬菌体克隆中的高序列多样性,以用于蛋白筛选。

I. 循环 1 (Maxisorp 平板或 Dynal 珠)

1. 包被靶

A. 包被平板: 使用 100 uL/孔的在 TBS[pH 7.5]/2 mM CaCl₂ 中稀释的 5µg/mL c-MET-ECD, 用 c-MET 胞外域(ECD)/Fc 嵌合体(0.5µg/孔)直接包被 6 孔/文库。当使用 c-MET-ECD/Fc 融合体(R & D Systems; 无载体)作为靶时, 在室温, 在摇动下, 用 A 蛋白预包被平板 1 小时。当使用生物素化的形式的 c-MET ECD/Fc 时, 在室温, 在摇动下, 用链霉抗生物素蛋白预包被平板 1 小时。另外, 仅用 TBS[pH 7.5]/2 mM CaCl₂ 包被一个阴性对照孔/文库。完成预-包被后, 加入 c-MET-Fc (+/-生物素), 并在室温, 在摇动下, 温育平板 1.5 小时。

B. 包被珠: 将 20 uL Dynal 链霉抗生物素蛋白 (M-280; Dynal ASA)或 Dynal A 蛋白珠 (Dynal ASA)分别与溶于 500 uL TBS[pH 7.5]/2 mM CaCl₂ 中的 5µg 生物素化的 c-MET/Fc 或未生物素化的 c-MET/Fc 一起温育, 并在室温, 在微量离心管中旋转 1 小时。作为阴性对照, 在 500 uL TBS[pH 7.5]/2 mM CaCl₂ 中温育没有靶的 20 uL Dynal 链霉抗生物素蛋白或 A 蛋白珠, 并在室温旋转 1 小时。应当指出, 用 TBS [pH 7.5]/2 mM CaCl₂ 洗涤 Dynal 珠至少 2 次, 然后加入靶, 并大量包被珠。

2. 封闭

A. 封闭平板: 取出包被溶液, 并用 200 uL/孔的 TBS[pH 7.5]/2 mM CaCl₂ 洗涤孔一次。加入 250 µl/孔的溶于 TBS[pH 7.5]/2 mM CaCl₂ 中的 1% BSA (无蛋白酶), 并在室温, 在摇动下, 温育 1 小时。替代试剂(例如酪蛋白或乳)可以用于封闭。

B. 封闭珠: 取出包被溶液, 并用 TBS [pH 7.5]/2 mM CaCl₂ 洗涤珠 2 次。加入 500 µl 溶于 TBS[pH 7.5]/2 mM CaCl₂ 中的 1% BSA (无

蛋白酶), 并在室温旋转 1 小时。如上所述, 可以使用替代的封闭试剂。

3. 洗涤

A. 洗涤平板: 用 200 μL /孔的 TBS[pH 7.5]/2 mM CaCl_2 洗涤孔 3 次, 以去除多余的靶。

B. 洗涤珠: 用 1000 μL TBS[pH 7.5]/2 mM CaCl_2 洗涤珠 3 次, 以去除多余的靶。在每次洗涤后, 使珠收集到磁体上几分钟, 以避免珠损失。

4. 噬菌体添加

A. 向平板添加噬菌体: 将约 1000 文库等同物(A1 域首次用于实验的噬菌体文库)加入噬菌体添加缓冲液(1%脱脂奶粉/0.2% BSA (无蛋白酶), 或其它适当的封闭剂, 溶于 TBS [pH 7.5]/2 mM CaCl_2 中), 并在室温, 在摇动下, 温育 2 小时。在循环 2-3, 将 100 μL 收获的全部噬菌体加入在噬菌体添加缓冲液中稀释的 7 个孔(6 靶+1 阴性对照)。

B. 向珠添加噬菌体: 将约 1000 文库等同物(A1 域首次用于实验的噬菌体文库)加入溶于 TBS [pH 7.5]/2 mM CaCl_2 中的 500 μL 1%脱脂奶粉+100 μL 1% BSA (无蛋白酶), 并在室温, 在旋转下, 温育 2 小时。在循环 2-3, 将 100 μL 收获的全部噬菌体加给珠。

5. 洗涤

A. 洗涤平板: 经 10 分钟时间段, 用 200 μL /孔的 TBS [pH 7.5]/2 mM CaCl_2 /0.1% Tween-20 洗涤平板 8-12 次。

B. 洗涤珠: 经 30-45 分钟时间段, 用 800 μL TBS [pH 7.5]/2 mM CaCl_2 /0.1% Tween-20 洗涤珠 8-12 次。通过将洗涤缓冲液直接分配到收集的珠上, 或通过上下移液(不通过涡旋), 促进珠重悬浮。或者, 可以将 KingFisher 装置 (Thermo LabSystems)或等同物用于珠洗涤。

严格洗涤的条件(可选)

- a. 800 μL TBS [pH 7.5]/2 mM CaCl_2 /0.1% Tween-20, 在 37 $^{\circ}\text{C}$;
- b. 800 μL TBS [450 mM NaCl, pH 7.5]/2 mM CaCl_2 /0.1% Tween-20, 在室温;
- c. 通常洗涤珠 6-8 次, 然后在室温或 37 $^{\circ}\text{C}$, 加入 1 μg 未标记的

c-MET-ECD 1 小时。保留在该洗涤后保持结合的噬菌体，以用于洗脱/感染；

d. 1%乳/0.2% BSA/有或没有 1 M 脲/ 37°C(高严格性)。

6. 竞争(可选):

A. 平板上的竞争: 在室温，在摇动下，将噬菌体与 100 uL/孔的溶于 TBS [pH 7.5]/2 mM CaCl₂ 中的 50µg/mL (5µg/孔) HGF (c-MET 配体)一起温育 1 小时。保留 HGF 洗脱物，以用于感染 BlueKan K91 大肠杆菌。

B. 珠上的竞争: 在室温，在摇动下，将噬菌体与溶于 500 uL TBS [pH 7.5]/2 mM CaCl₂ 中的 10µg HGF 一起温育 1 小时。保留 HGF 洗脱物，以用于感染 BlueKan K91 大肠杆菌。

7. 噬菌体洗脱

A. 平板的洗脱: 加入 100 uL/孔的溶于 TBS [pH 7.5]/2 mM CaCl₂ 中的 10 mg/mL 胰蛋白酶，并在摇动下，在 37°C，温育平板 30 分钟。

B. 珠的洗脱: 给珠添加 100 uL 10 mg/ml 胰蛋白酶 TBS [pH 7.5]/2 mM CaCl₂，然后在摇动下，在 37°C，温育(在 Eppendorf rack 中)30 分钟。

C. 替代洗脱/感染: 将 200 µl OD₆₀₀ 约 0.5 的对数期 BlueKan K91 大肠杆菌细胞加入每个孔(对于平板)或加给吸气的珠。在无摇动下，使感染在 37°C 进行 30 分钟。接着，合并 200 uL 体积，并加入到约 3 mL 2x YT/0.2µg/mL 四环素，且在 37°C 摇动 15 分钟。

8. 感染: (对于平板和珠规程，是相同的)

使适当体积的对数期 BlueKan K91 大肠杆菌(在 2x YT/40µg/mL 卡那霉素中)生长至 OD₆₀₀ 约 0.5-0.6。当培养物达到 OD₆₀₀ 时，将它置于冰上备用，尽管通常使在冰上的时间最小化。

A. 在 50 mL 无菌的锥形管中，将洗脱的噬菌体与 5 mL 对数期 BlueKan K91 大肠杆菌培养物相混合，并在无摇动下，在 37°C 温育 25 分钟。给无菌的锥形管覆盖 AirPore 带(Qiagen)，以促进通气。

B. 加入四环素至 0.2µg/mL 的终浓度，并在 37°C 摇动 15 分钟。

C. 取 10 uL 等分试样用于滴定，并在 2x YT 中系列稀释 10 倍(10⁻¹ 至 10⁻⁶)，以 8 uL/稀释斑点铺平板到 2x YT/20µg/ml 四环素平板上，并在 30°C 或 37°C 温育过夜。铺平板剩余体积的 10⁻²-10⁻⁴ 稀释液，以

得到单个菌落，用于随后的噬菌体 ELISA。

D. 将感染的 5 mL 培养物稀释约 10 倍，得到 50 mL 2x YT/20 μ g/mL 四环素，并在摇动下，在 30 $^{\circ}$ C 温育过夜到饱和。

9. 在当前的淘选循环中，使用滴定输入噬菌体(对于平板和珠规程，是相同的)

A. 在 2x YT 中，制备收获的噬菌体的 100 倍系列稀释液(10^{-4} 至 10^{-10})。

B. 将 100 μ L/孔的 OD₆₀₀ 0.5-0.6 的对数期 BlueKan K91 大肠杆菌培养物加入 96-孔聚丙烯平板的 6 个孔中。

C. 将 10 μ L 稀释的噬菌体加入含有 100 μ L BlueKan K91 大肠杆菌的孔中。

D. 在无摇动下，在 37 $^{\circ}$ C，温育噬菌体/细胞混合物 25 分钟，并给平板覆盖 AirPore 带(Qiagen)，以促进通气。

E. 加入四环素至 0.2 μ g/mL 的终浓度，并在 37 $^{\circ}$ C 摇动平板 15 分钟。

F. 将 8 μ L 每种稀释液(10^{-4} 至 10^{-10})铺平板到干燥的 2x YT 琼脂/20 μ g/mL 四环素平板上。

G. 在 30 $^{\circ}$ C 或 37 $^{\circ}$ C 温育平板过夜。

10. 收获噬菌体 (对于平板和珠规程是相同的)

A. 在一次性 50 mL 管中，在 7000 rpm，离心过夜培养物 25 分钟，以沉淀细胞。

B. 通过向培养上清液加入 1/5 体积的 20% PEG/15% NaCl 原液，进行标准的 PEG/NaCl 噬菌体-沉淀操作。通过重复反转，并在冰上温育 45 分钟至 1 小时，来混合均匀。

C. 在 7000 rpm 离心培养物 40 分钟，以沉淀噬菌体，并抛弃上清液。

D. 将噬菌体沉淀重新悬浮在 1 mL TBS [pH 7.5]/2 mM CaCl₂ 中，转移到微量离心管，并在 13K rpm 离心至少 2 分钟，以沉淀不溶材料。

E. 将上清液转移到新管，加入 1/5 体积的 PEG/NaCl，混合，并在冰上温育约 5 分钟。

F. 然后，在 13000 rpm 离心混合物至少 2 分钟，并取出上清液。

将沉淀的、纯化的噬菌体重新悬浮于最多 1 mL TBS [pH 7.5]/2 mM CaCl₂ 中，并在 4℃ 保藏。

II. 循环 2 和循环 3 淘选

第 2 个和第 3 个循环淘选条件通常与上述的循环 I 相同，例外是，对于每个后续循环将包被的靶(即 c-MET-ECD)量降低 2 至 4 倍，且在淘选每个后续循环中，洗涤平板(或珠)额外的 2-4 次。

III. 任选的域内重组

以下面的方法，重组噬菌体展示-选择的噬菌体库的单体序列。该方法产生杂合单体，其源自给定库中的起始单体集合的混合的一半。对于基于 A1-域的噬菌体文库，使用引物对 SHF1 (ATTATGCCCCGGGTCTGGAGGCGTC)/SHB 重叠 (CGCCGTCGCAA) 和 SHF 重叠 (TTGCGACGGCG)/B3 (TCGGCCCCAGAGGCCTGCAATG) 来 PCR-扩增 2 个一半的单体。用 LA Taq 聚合酶 (Takara)，将两个一半融合到一起。接着，通过引物 SHF2 (CCGGATTATGCCCCGGGTCTGGA) 和 SHB4 (AACAGTTTCGGCCCCAGAGGCCTGC)，扩增融合的杂合体编码序列。用 *Sfi*I (NEB) 消化纯化的 PCR 产物，并与 *Sfi*I-消化的 fUSE5HA 噬菌体载体连接，以产生重组的单体文库。针对 c-MET ECD/Fc，淘选重组文库至少 2 个额外的循环，并如下所述筛选。表征重组的单体的数据见表 1 和 2。

IV. 淘选输出的分析(对于平板和珠规程是相同的)

噬菌体 ELISA: 对于每个要分析的输出“噬菌体库”(一般地，循环 2、3 和 4，如果适用)，将独立的克隆接种进在 Costar 96-孔聚丙烯深孔平板中生长的 1 mL (2x YT/20µg/mL 四环素) 培养物中。放入接种尖，在 37℃ 摇动平板过夜。通过在 3600 rpm 离心 15 分钟，沉淀细胞。保留培养物上清液，且如下所述进行 ELISA。

将未生物素化的 c-MET ECD/Fc (0.1µg/孔) 直接包被到 Nunc Maxisorp 平板上。但是，生物素化的 c-MET ECD/Fc，96-孔 Nunc Maxisorp 平板应当首先包被 50 µL/孔的在 TBS [pH 7.5]/2 mM CaCl₂

中稀释的 50 μ g/mL (2.5 μ g/孔)链霉抗生物素蛋白。在摇动下, 在 37 $^{\circ}$ C 温育平板 1 小时。用 200 μ L/孔的 TBS [pH 7.5]/2 mM CaCl₂ 洗涤平板 3 次。用 200 μ L/孔的 1% BSA (级分 V)封闭孔, 并在摇动下, 在室温温育覆盖的平板 1 小时。用 TBS [pH 7.5]/2 mM CaCl₂ 洗涤平板 3 次。接着, 用 100 μ L/孔的在 TBS [pH 7.5]/2 mM CaCl₂ 中稀释的 1 μ g/mL (0.1 μ g/孔) 生物素化的 c-MET-ECD, 或只用 100 μ L/孔的缓冲液(阴性对照), 包被 96-孔 Maxisorp 平板。在摇动下, 在室温温育平板 1 小时。用 TBS [pH 7.5]/2 mM CaCl₂ 洗涤平板 3 次。接着, 在 70 μ l 1% 乳/0.2% BSA/[pH 7.5]/2 mM CaCl₂/0.02% Tween-20 存在下, 将 30 μ l 每种噬菌体上清液加入孔中。在摇动下, 在室温温育覆盖的平板 1.5 小时。

用 TBS [pH 7.5]/2 mM CaCl₂/0.02% Tween-20 洗涤平板 4 次。接着, 加入 100 μ L/孔的在 TBS [pH 7.5]/2 mM CaCl₂+0.02% Tween-20 中 1: 5000 稀释的 α -M13-HRP 单克隆抗体(Amersham Pharmacia)。在摇动下, 在 4 $^{\circ}$ C 温育平板 1 小时。用冷 TBS [pH 7.5]/2 mM CaCl₂/0.02% Tween-20 洗涤平板 3 次。加入 100 μ L/孔的 1: 1 稀释的 TMB/H₂O₂ 混合物 (Pierce), 用于 ELISA 显影。

允许反应变成蓝色, 直到最强的 OD₆₅₀ 信号达到约 1.0。用 100 μ L/孔的 2N H₂SO₄ 停止反应, 且阳性孔的颜色从蓝色变成黄色。一旦停止反应, 就使用 SoftMaxPro 软件, 在 ELISA 平板读数仪上读取 OD₄₅₀。

如果噬菌体 ELISA-阳性的噬菌体库具有(a)高频率的结合 c-Met ECD/Fc 的单个噬菌体克隆, 和(b) 结合-阳性的噬菌体克隆中的高序列多样性, 就选择它们用于亚克隆进表达载体。选择满足这些标准的库, 用于下述方法中的蛋白筛选。为了将来自给定噬菌体库的单体或多聚体序列亚克隆进表达载体 pEve, 通过如下的 25 个 PCR 循环, 扩增大约 10⁸-10¹⁰ 噬菌体:

PCR 方法

0.5 - 1 μ L 纯化的噬菌体

5 μ L 10X 缓冲液

8 μ L 2.5 mM dNTP

5 μ L 10 μ M VS-For 引物 (5'-ATCATCTGGCCGGTCCGGCCTACCCGTATGATGTTCCGGA-3')

5 μ L 10 μ M EveNut 引物 (5'-AAAAGGCCCCAGAGGCCTTCTGCAATGAC-3')

26 uL H₂O

0.5 uL LA Taq 聚合酶 (1 单位) (Takara)

循环: 25X[94°C/10 秒-45°C/30 秒-72°C/30 秒]

将 PCR 产物在 3% 琼脂糖凝胶上电泳, 进行分析。用 QIAquick 旋转柱(Qiagen)纯化单体或多聚体产物(大约 200 bp), 用 *Sfi* I (NEB) 消化, 再次用 QIAquick 柱纯化, 然后使用 T4 DNA 连接酶 (NEB) 连接到 *Sfi* I 消化的载体 pEve 上。将连接体转化进电感受态 BL21 (DE3) 大肠杆菌, 并铺平板到含有 40µg/mL 卡那霉素的 2x YT 平板上。过夜生长后, 将大约 6000 单个克隆接种进 2x YT/卡那霉素, 并生长过夜。在平板上也包含阳性和阴性对照。

V. 筛选 1 mL 细胞裂解物中的数千单体蛋白

1 mL 加热的裂解物的蛋白生产(第 1 天): 将单个克隆接种进含有 400 uL/孔的 2x YT/40µg/mL 卡那霉素的 96-孔 Costar 深孔平板的孔中。使培养物生长过夜(将接种尖保留在孔中), 同时在 37°C、300 rpm 摇动。该方法允许在细胞裂解物水平筛选数千单个的、部分地纯化的单体。

(第 2 天) 将 100 µl 过夜培养物接种进含有 1 mL/孔的 2x YT/40µg/mL 卡那霉素 +1 mM CaCl₂ 的新 96-孔 Costar 深孔平板。(向剩余的过夜培养物添加 25% 最终甘油浓度来使其归档, 然后在 -80°C 保藏备用)。给平板覆盖 AirPore 带(Qiagen), 并在 375 rpm 摇动下, 在 37°C 生长培养物, 直到达到约 0.8 至 1.0 的 OD₆₀₀。一旦达到所需的 OD₆₀₀, 就用 1 mM IPTG 诱导培养物 3 小时, 同时在 37°C、375 rpm 摇动。然后, 在 4°C、3600 rpm, 离心含有诱导的培养物的平板 15 分钟, 以沉淀细胞。取出上清液, 并抛弃, 将剩余的细胞沉淀重新悬浮在 100 µl TBS [pH 7.5]/1 mM CaCl₂ 中。将重新悬浮的细胞从 96-孔深孔平板转移至 96-孔聚丙烯 PCR 平板, 并在 PCR 仪中在 65°C 加热 5 分钟。然后, 在 4°C、3600 rpm, 离心加热/裂解的细胞 15 分钟。离心后, 完成蛋白生产, 并通过结合 ELISA 和/或竞争 AlphaScreen 测定, 将加热的裂解物准备用于初步筛选中的表征。

C-Met ECD/Fc 蛋白 ELISA: 用 100 uL/孔的在 TBS[pH 7.5]/1 mM CaCl₂ 中稀释的 1µg/mL (0.1µg/孔) c-MET ECD/Fc (R&D

Systems), 包被 96-孔 Maxisorp 平板, 然后在摇动下, 将平板在 4℃ 温育过夜或在室温(RT)温育 1.5 小时。排空孔, 然后用 200 uL/孔的 1% BSA (级分 V)/TBS[pH7.5]/1 mM CaCl₂ 封闭。在摇动下, 在室温温育覆盖的平板 1 小时。用 TBS [pH 7.5]/1 mM CaCl₂ 洗涤平板 3 次。向平板加入 100 uL/孔的在 TBS [pH 7.5]/1 mM CaCl₂/0.1% BSA/0.02% Tween-20 中稀释的单体蛋白。将来自 1 mL 加热的裂解物制剂的蛋白加入孔中, 作为 1: 10 稀释的单点浓度。在摇动下, 在室温温育覆盖的平板 1.5 小时。用 TBS [pH 7.5]/1 mM CaCl₂/0.02% Tween-20 洗涤平板 3 次。加入 100 uL/孔的在 TBS [pH 7.5]/1 mM CaCl₂/0.1% BSA/ 0.02% Tween-20 中 1: 2000 稀释的抗-HA-HRP 检测抗体 (Roche)。在摇动下, 在室温温育覆盖的平板 1 小时。用 TBS [pH 7.5]/1 mM CaCl₂/0.02% Tween-20 洗涤平板 3 次。加入 100 uL/孔的 1: 1 稀释的 TMB/H₂SO₄ 混合物。允许颜色变成蓝色, 直到最强的 OD₆₅₀ 信号达到约 1.0。用 100 uL/孔的 2N H₂SO₄ 停止反应。一旦停止, 就在 ELISA 平板读数仪上在 OD₄₅₀ 读取平板。

AlphaScreen c-Met/Fc-生物素化的(bn)HGF 均质竞争测定: 在如下 AlphaScreen 缓冲液中稀释所有测定组分: 40 mM HEPES [pH 7.4] w/NaOH、1 mM CaCl₂、0.1% BSA (w/v)、0.05% Tween-20、100 mM NaCl。向白色的 384-孔、减少的体积的 Greiner 微量滴定测定平板进行 3 次添加, 在添加之间没有温育时间。首先, 以 2 uL/孔, 将单体或未标记的重组的人 HGF (rhHGF) (作为阳性对照) 加入平板。以单一浓度(未稀释的[即 1: 4 最终测定稀度]或最高达 1: 100 稀度[1: 400 最终测定稀度]), 将来自 1 mL 加热的裂解物制剂的单体加入孔中。作为阳性对照, 代替单体蛋白, 将 2 uL/孔的未标记的 rhHGF (PeproTech)加入平板, 作为 12 点浓度曲线, 所述曲线从 400 nM (即 100 nM 最终测定浓度)开始, 然后为 1: 4 系列稀释液, 最后为仅缓冲液的点。其次, 将 4 uL/孔的 0.6 nM (即 0.3 nM 最终测定浓度)c-MET ECD/Fc 加入平板。应当指出, 在减弱的或绿色过滤的光线中完成剩余测定, 因为 AlphaScreen 珠是光线敏感的。第三, 向平板加入 2 uL/孔的 1 nM (即 0.25 nM 最终测定浓度)bn-HGF 和均稀释至 40µg/mL(即, 10µg/mL 最终测定浓度)的 AlphaScreen 链霉抗生物素蛋白“供体珠”和 A 蛋白“受体珠”(PerkinElmer)的混合物。然

后，用顶盖密封覆盖测定平板，并在 800 rpm 旋转约 30 秒。然后，在室温、在黑暗中，温育平板过夜，且次日在 Fusion 平板读数仪 (PerkinElmer) 上读数。

VI. 噬菌体展示-选择的单体的多聚化和重组

以下述的方式，多聚化已经亚克隆进 pEve 的单体(pEve/单体)。用 BsrDI 或 BpmI (NEB)，消化 pEve/单体质粒(单个地或在库中)。从 1%琼脂糖凝胶分离约 1.1 kb BsrDI 和约 2.9 BpmI 片段，并用 Qiagen QIAquick 旋转柱纯化。使用 T4 DNA 连接酶 (NEB)，连接 2 种片段中的每一种的库；随后，用 Qiagen QIAquick 旋转柱，纯化连接体。使用在上面的噬菌体亚克隆部分所述的引物 VS-For 和 EveNut，从连接体 PCR-扩增多聚体编码序列。纯化 PCR 产物，并用 SfiI (NEB) 消化，随后用 pEve 连接，并转化 BL21 (DE3) 大肠杆菌。该方法生成包含起始单体的不同组合的二聚体。该方法也可以用于产生其它的多聚体，例如三聚体。当制备三聚体时，pEve/二聚体(例如在上面的实施例中和 pEve/单体(起始集合)的库是原料。如上所述处理它们。也可以使用与下述的制备“步移文库”的方法类似的分子生物学方法，产生多聚体。在所有情况下，如上所述表达、纯化和筛选蛋白。

通过连接噬菌体展示-选择的单体(即选择的单体)和首次用于实验的单体文库的全部代表，产生称作“步移文库”的其它文库。以下面的方式，构建这些文库。在 2 个分开的反应中，使用 PCR 进行扩增： a) 选择的单体的编码序列，其具有 pETF (ACCCGTATGATGTTCCGGATTA)/pETB2r (GATGTATTCGGCCCCAGA GGCCTGCAATGAC); 和 b) 单体文库的首次用于实验的单体的编码序列，其具有 21new1 (GAAATTCACCTCGAAAGCAA)/23 (ATGGGTTCCTATTGGGCT)。从 3%琼脂糖凝胶分离约 200 bp 产物，并用 Qiagen QIAquick 旋转柱纯化。在分开的反应中，用 BsrDI 或 BpmI (NEB) 消化来自上面的(a)和(b)的每种产物。BpmI-消化的单体具有可以连接到 BsrDI-消化的单体上的突出端。使用 T4 DNA 连接酶 (NEB)，将纯化的消化产物彼此连接。连接 BsrDI-切割的首次用于实验的单体与 BpmI-切割的选择的单体，产生了步移二聚体文

库, 其包含融合到 C-末端选择的单体上的 N-末端首次用于实验的单体。连接 BpmI-切割的首次用于实验的单体与 BsrDI-切割的选择的单体, 产生了步移二聚体文库, 其包含融合到融合到 N-末端选择的单体上的 C-末端首次用于实验的单体。使用引物 pETF/pETB2r, 从连接体 PCR-扩增连接的二聚体编码序列, 并用 *SfiI*、然后用 *XmaI* 消化纯化产物。将消化的产物连接到噬菌体载体 fUSE5HA, 以用于产生噬菌体展示二聚体“步移文库”, 一般地具有 10^8 - 10^9 种独特成员。可以以类似的方式, 产生三聚体(或更大的多聚体)“步移文库”, 例外是, 原料是二聚体(或更大的)和首次用于实验的单体。针对 c-MET ECD/Fc 淘选步移文库, 并如上所述进行筛选。

VII. 在结合和竞争测定中表征纯化的单体

一旦在加热的蛋白裂解物水平表征了蛋白, 就选择最佳的单体, 以用于进一步表征。制备单个克隆的更大规模的培养物, 并通过 Ni-NTA 树脂, 纯化携带 6His 标签的单体。在结合 ELISA 和 AlphaScreen 竞争测定中, 测定这些镍-纯化的单体。表征纯化的单体的蛋白序列数据和生化数据见表 1 和 2。

蛋白纯化, 500 mL 培养物, 对于 NiNTA: (第 1 天) 在 15 mL 含有 3 mL 2x YT+40 μ g/mL 卡那霉素的培养物管中, 接种适当的“初步命中孔”归档的甘油原液。在 300 rpm、37 $^{\circ}$ C, 摇动培养物过夜。

(第 2 天) 将 2 mL 过夜培养物接种进含有 500 mL 2x YT+40 μ g/mL 卡那霉素的 1 L 锥形摇瓶中。在摇动下, 在 375 rpm、37 $^{\circ}$ C, 使培养物生长, 直到达到约 0.8-1.0 的 OD₆₀₀。一旦达到所需的 OD₆₀₀, 就用 1 mM 终浓度的 IPTG 诱导培养物 3 小时, 同时在 375 rpm 摇动。诱导 3 小时后, 将 500 mL 培养物转移到清洁的/高压灭菌的 Sorvall 管, 并在 8000 rpm、4 $^{\circ}$ C 离心 8 分钟, 以沉淀细胞。

一旦细胞被沉淀, 就取出并抛弃上清液, 向每个管中加入 20 mL 超声处理缓冲液 (10% 蔗糖/20 mM Tris [pH 7.5]/150 mM NaCl/0.2 mM CaCl₂)。用 10 mL 血清学移液管, 将沉淀重新悬浮在超声处理缓冲液中, 直到没有可看到的块, 然后将重新悬浮的细胞(约 30 mL)转移进 35 mL Oakridge 管, 并在约 16 功率输出超声处理 8 分钟。超声处理后, 将含有超声处理的细胞的温 Oakridge 管置于冰/水浴上约 10

分钟，以进行冷却。一旦冷却，就在 18,000 rpm、4℃，离心管 30 分钟，以沉淀裂解的细胞。

在离心含有裂解的细胞的管的同时，用 Milli-Q 水洗涤 NiNTA 树脂(Qiagen)，以去除乙醇。使用 3 mL 1: 1 稀释的 NiNTA 树脂/蛋白(即实际上，使用 1.5 mL 树脂/蛋白)。将每种 3 mL 树脂/水混合物加入适当地标记的(用蛋白 ID)清洁的 50 mL 螺帽管。沉淀超声处理的细胞后，取出蛋白上清液，并将其加入含有 1.5 mL 洗涤的 NiNTA 树脂的 50 mL 管。通过在室温轻轻摇动 0.5 小时，使蛋白结合 NiNTA 树脂。与 NiNTA 树脂温育后，在约 1500 rpm 离心 50 mL 管 10 分钟，使 NiNTA 结合蛋白。轻轻倒出上清液，并抛弃。

如下将 NiNTA 树脂+结合的蛋白转移至适当地标记的 15 mL Clontech 柱：将 1 mL NiNTA 洗涤缓冲液(20 mM Tris [pH 7.5]、200 mM NaCl、0.1 mM CaCl₂、20 mM 咪唑)加入到含有树脂的 50 mL 管，回荡以进行重新悬浮，然后将混合物移液入已经封固到多支管真空装置上的柱中。用至少 10 柱体积(15 mL)的 NiNTA 洗涤缓冲液，洗涤 NiNTA 树脂+结合的蛋白。将含有 NiNTA 树脂+结合的和洗涤的蛋白的 15 mL 柱转移至清洁的 15 mL 螺帽收集管。将 4 mL Ni 洗脱缓冲液(20 mM Tris [pH 7.5]、200 mM NaCl、0.1 mM CaCl₂、200 mM 咪唑)加入每个柱，以将蛋白洗脱进 15 mL 收集管。然后，通过重力使其进行洗脱。

使用 18.5 号针和 5 mL 注射器来装载盒，将洗脱的蛋白转移至 slide-A-lyzer 盒(适当的 MW 截止—对于单体，使用 3.5 kDa 截止；且对于二聚体和三聚体，使用 10 kDa 截止)。将含有洗脱的蛋白的 Slide-A-lyzer 置于含有氧化还原剂的过夜透析缓冲液(20 mM Tris [pH 7.5]、100 mM NaCl、1 mM CaCl₂、1 mM 2-巯基乙醇、0.25 mM 2-羟乙基二硫化物)。

(第 3 天) 将含有过夜透析的蛋白的 Slide-A-lyzer 盒转移进没有氧化还原剂的透析缓冲液(20 mM Tris [pH 7.5]、100 mM NaCl、1 mM CaCl₂)中。透析 3 小时后，将 slide-A-lyzer 盒转移进没有氧化还原剂的新 TBS/CaCl₂ 中另外 3 小时。第 2 次透析变化后，使用 18.5 号针和 5 mL 注射器，从 slide-A-lyzer 盒取出蛋白，并通过使用 0.2 微米注射器过滤器过滤，将蛋白转移进适当地标记的 15 mL 聚丙烯

管。

通过 Q-琼脂糖阴离子交换,进一步纯化抗-c-MET NiNTA 纯化的蛋白,其在 AlphaScreen 竞争测定中被选择为“最佳抑制剂”,以取出污染物。Q-琼脂糖纯化:将 1 mL Q-琼脂糖高流速树脂(Amersham Biosciences)加入 15 mL Clontech 柱。用 15 柱体积(或 15 mL)的 20 mM Tris [pH 7.5]、50 mM NaCl、1 mM CaCl₂ 平衡树脂。将 2 mL (约 5 mg)过滤的 NiNTA-纯化的蛋白加入树脂,并通过重力,使蛋白结合树脂。收集向 96-孔平板的第一个柱的流通物(flow-through)。将装载蛋白的柱转移至 15 mL 收集管,并用 10 柱体积(或 10 mL) 20 mM Tris [pH 7.5]、50 mM NaCl、1 mM CaCl₂ 洗涤树脂/结合的蛋白。一旦洗涤,就开始蛋白的 NaCl 梯度洗脱。以如下的梯度,改变 NaCl 浓度: 100 mM、150 mM、200 mM、250 mM、300 mM、350 mM、400 mM、500 mM 和最后的 1 M NaCl,至基础的 20 mM Tris [pH 7.5]、1 mM CaCl₂。在 96-孔深孔聚丙烯平板中收集级分--2 mL/级分,以 1 mL 递增。通过 Bradford 测试含有蛋白的级分,并通过 SDS PAGE 进行分析。以下面的变化,在如上所述的结合 ELISA 和竞争测定中测试级分。将来自 500 mL NiNTA 纯化的制剂或 NiNTA+Q-琼脂糖纯化的制剂的蛋白加入平板,作为 12 点浓度曲线,所述曲线从 1: 5 至 1: 100 的第一稀释液开始,然后为 1: 4 系列稀释液,最后为仅缓冲液的点。表征纯化的单体的蛋白序列数据和生化数据见表 1 和 2。

表 1 抗-c-MET Avimer (M)和重组的(Rec) 单体序列。应当指出,从重组的单体文库分离的有些蛋白是二聚体。

蛋白 ID	构建体	序列
M01	C128	CPSDEFKCHSTGRCLPVEWLCVGDVNDCEDESSDEASCSAPASEPPGSLSLQ
M02	C129	CQSNEFTCOSTNRCLPLPWWCDGDNDCEDSSDEANCGORTSLQ
M03	C130	CEANEFRCRSTGRCSISQWRCDDDDCEDSSDEANCKPPTSLQ
M04	C131	CLSNEFRCSSTGRCLPRPWVCDGDNDCEDGSDDEAPICGRPGGATSAPAAASLQ
M05	C132	CNTTQFSCRSTNRCIPLDWQCQDGVTDCEDESSDEASCSAPASEPPGSLSLQ
M06	C133	CPPDEFPCRSTERCIPLAWVCDGDNDCEDSSDEAGCTTPEPTSLQ
M07	C134	CXATQFRCPRTRLCIPPTWLCDDGDNDCEDGSDDEANCTASESKPLQSLQ
M08	C135	COSSEFTCKSTERCIPLTWVCDGDNDCEDGSDDEENCSQDPEFHKVSLSQ
M09	C136	CLASEFTCHSTGRICIPETWVCDGVNDCEDESSDEADCGRPGGATSAPAAASLQ
M10	C137	CQPDEFNCNSTGRICIPDWVCDGVDDCEDGSDDETCGSDPEFHKVSLSQ
M11	C237	CLASEFTCHSTGRICIPETWVCDGVNDCEDESSDEADNAHCGRPGGATSAPAAASLQ
M12	C238	CESSQFTCNSTKRCIPLAWVCDGDDDCEDGSDDEKSCAAPTSLQ
M13	C239	CLADEFQCHSTKRCVPIRHWLCVGDVNDCEDESSDEKSCSDPEFHKVSLSQ
M14	C240	CAPNEFTCSSTGRCLPRAWVCDGVDDCEDGSDDETCGATVHTSLQ
M15	C241	CAPDEFPCRSTGRCVPLTWLCDDGDNDCEDGSDDEASATCGRPGGATSAPAAASLQ
M16	C377	CAPSEFTCNSTGRICIPQEWVCDGDNDCEDSSDEAPDLCAAAAPTSLQ
M17	C378	CRANEFQCHSTGRICIPQWVCDGDNDCEDGSDDEAGCAASGPTSLQ
M18	C379	CESNEFQCOSTSRCIPLTWRCDDGDNDCEDGSDDEANCTAAVHTSLQ
M19	C380	CESSEFRCRSTGRICIPGWVCDGDNDCEDSSDETCGAPASEPPGSLSLQ
M20	C381	CEADEFRCRSTGRCSISVWRCDDGDNDCEDSSDEESCESTAPTSLQ
M21	C382	CVSNEFTCRSTKRCVPOEWVCDGVNDCEDESSDETCGPKHTSLQ
M22	C383	CPSGEFQCRSTNRCIPLWVCDGDNDCEDGSDDEESCTPPTSLQ
M23	C384	CAANEFQCHSTGRCIPLSWVCDGVNDCEDESSDETCRAPTSEPKGVSLSQ
M24	C385	CPPDEFPCRSTGRICIPRAWLCHGDNDCEDSSDEKDKCKPHTSLQ
M25	C386	CESGEFQCHSTGRICIPASWVCDGDNDCEDGSDDESQLCTAHTSLQ
M26	C387	CQSFTFECHSTGRICIPASWVCDGDNDCEDSSDESPANCAPTAHTSLQ
M27	C388	CVASEFTCRSTGRICIPESWRCDDGDNDCEDSSDESPDLCSAPASEPPGSLQ
M28	C389	COSSEFQCHSTGRICIPENWVCDGDDDCEDSSDEKSCSTSAAPTSLQ
M29	C390	CQAGQFECRSTNRCIPQDWVCDGVNDCEDESSDEESCTSPARTSLQ
M30	C391	CQAGQFECRNTNRCIPQDWVCDGVNDCEDESSDEESCTSPARTSLQ
M31	C392	CLPSEFQCKSTNRCIPLQWVCDGDNDCEDGSDDETCNSAPASEPPGSLSLQ
M32	C393	CQPNFQCHSTGRICIPASWVCDGDNDCEDGSDDESQLCTAHTSLQ
M33	C394	CAADEFOCNSTGRICIPVSWVCDGVNDCEDESSDEAGCATSGPTSLQ
M34	C395	CPSSQFTCHSTRRCIPQTWVCDGDNDCEDGSDDETCGPHHTSLQKASSGGG
M35	C396	CLANEFRCNSTGRICIPRAWLCHGDNDCEDSSDEKDKCKQHTSLQ
M35	C396	CSSDEFQCSSTSRCIPREWVCDGVNDCEDESSDEALACTSTAHTSLQ
M36	C397	CESNEFQCHSTSRCIPLTWRCDDGDNDCEDGSDDEANCTAAVHTSLQ
M37	C398	CGANEFQCHSTNRCIPQSWVCDGVNDCEDESSDESPVLCATTVHTSLQ
M38	C399	CVSNEFTCRSTKRCVPOEWVCDGVNDCEDESSDETCGPKHTSLQ
M39	C400	CVSNEFFQCOSTDRICPRAWVCDGDNDCEDGSDDEKDCSAPASEPPGSLSLQ
M40	C401	CLPSEFQCKSTNRCIPLQWVCDGDNDCEDGSDDETCNSAPASEPPGSLSLQ
RecM01 (二聚体)	C409	CPAGQFTCRSTNRCIPLQWVCDGDNDCEDSSDESPAICATTGPTSLQKASAAYPYDVPDYAPGLEASGGG
RecM02	C410	CESNEFQCHSTSRCIPLQWVCDGDNDCEDSSDEASCGCPGGATSAPAAASLQ
RecM03 (二聚体)	C411	CHAPTQFECRSTGRICIPLTWVCDGDNDCEDGSDDEKDCGDSHILPFSTPGPSTSLQKASAAYPYDVPDYAPGLEASG
RecM04	C412	CDPIAEFKCHSTGRICIPLDWLCDDGDNDCEDSSDESPAHCAPASEPPGSLSLQ
RecM05	C413	CBASEFTCRSTNRCIPVDWVCDGVNDCEDESSDESDICAPASEPPGSLSLQ
RecM06 (二聚体)	C414	CHPTAEFQCHSTGRICIPVDWVCDGDNDCEDSSDEKDCGDSHILPFSTPGPSTSLQ
RecM07	C415	CQASDQFECKSTGRICIPAWVCDGDNDCEDGSDDESPAICGRPGGLEASGGG
RecM08	C416	CRANEFQCHSTGRICIPASWVCDGDNDCEDGSDDEASCGRPGGTSAPAAASLQ
RecM09	C417	CAADEFOCNSTGRICIPVNWVCDGDNDCEDSSDEENCSAPASEPPCSLSLQ
RecM10 (二聚体)	C418	CQSFTFECHSTGRICIPVDWVCDGDNDCEDSSDESPAICAPASEPPGSLSLQ
RecM11 (二聚体)	C419	CESNEFQCHSTGRICIPASWVCDGDNDCEDGSDDESPAICGRPLGLGATSAPAAASLQ
RecM12 (二聚体)	C420	CLANEFQCHSTGRICIPVSWVCDGDNDCEDSSDEKDCGDSHILPFSTPGPSTSLQ
RecM13 (二聚体)	C421	CESNEFQCHSTSRCIPLTWRCDDGDNDCEDSSDEKSCSAPASEPPGLEASGGG
RecM14	C422	CFASEFTCRSTGRCSISQWVCDGDNDCEDSSDESPAICATTGPTSLQ

表 2 抗-c-MET 单体和重组的单体结合 Kd 和生化 IC50 数据。应当指出，从重组的单体文库分离的有些蛋白是二聚体(在表 1 中指出)。空白条目指数据未获得。

蛋白 ID	构建体 #	AlphaScreen IC50 NINTA-纯的 (nM)	结合 ELISA Kd NINTA-纯的 (nM)	AlphaScreen IC50 Q-纯的 (nM)	结合 ELISA Kd Q-纯的 (nM)
M01	C128	120	105		
M02	C129	196	172		
M03	C130	129	58	328	
M04	C131	197	251		
M05	C132	392	69		
M06	C133	27	40	19	
M07	C134	161	121		
M08	C135	80	130		
M09	C136	81	185		
M10	C137	256	439		
M11	C237	41	NA	386	
M12	C238	79	NA		
M13	C239	247	NA		
M14	C240	293	NA		
M15	C241	81	NA		
M16	C377	15	46	27	
M17	C378	36	79	165	
M18	C379	58	113		
M19	C380	35	82	111	
M20	C381	22	158	186	
M21	C382	83	116		
M22	C383	34	66	57	
M23	C384	43	138		
M24	C385	57	77		
M25	C386	11	74	29	
M26	C387	13	100	16	5.8
M27	C388	42	93	105	
M28	C389	30	31	152	
M29	C390	23	43	162	
M30	C391	34	73	161	
M31	C392	19	99	145	
M32	C393	20	232	69	
M33	C394	42	157	185	
M34 (二聚体)	C395	3	556	2	
M35	C396	28	93	290	
M36	C397	52	79	213	
M37	C398	28	85	232	
M38	C399	50	95	881	
M39	C400	35	86	85	
M40	C401	21	240	97	
RecM01	C409	2	2.4	1	0.3
RecM02	C410	2	20.3	2	344.3
RecM03	C411	2	2.7	3	0.2
RecM04	C412	24	16.4	94	126.2
RecM05	C413	15	9.7	23	2.0
RecM06	C414	1	0.4	2	0.7
RecM07	C415	1	3.5	13	5.4
RecM08	C416	8	39.3	16	24.6
RecM09	C417	19	42.2	52	51.6
RecM10	C418	1	2.0	2	3.9
RecM11	C419	1	1.1	1	0.3
RecM12	C420	1	3.8	0.33	0.3
RecM13	C421	3	2.5	2	0.2
RecM14	C422	11	7.9	15	44.6
对照	rhHGF (R&D)	0.152			

实施例 7

本实施例描述了证实 c-MET-结合单体对 HGF-诱导的细胞增殖的抑制的实验。

HGF 是上皮细胞增殖的有效刺激物。本领域已经充分确立了

A549 人肺腺癌细胞在用于测定 HGF 和/或 c-MET 抑制剂的功效率的 HGF-诱导的增殖的测定中的用途。为了这些实验的目的，通过有限稀释，衍生出称作 A549-SC 的 A549 细胞系的单细胞克隆。在它的在 HGF 存在下的强细胞分散反应的基础上，选择 A549-SC 克隆。

以 100 μl 无血清的 F-12 培养基/孔，将 A549-SC 细胞铺平板到胶原-包被的 96 孔平板 (1×10^4 细胞/孔)上，然后，在 37 $^{\circ}\text{C}$ 、5% CO_2 温育 48 小时。48 小时后，从孔取出培养基，并替换为单体的稀释液，其体积为 50 μl /孔无血清的 F-12 培养基。在 37 $^{\circ}\text{C}$ 、5% CO_2 温育 1 小时后，加入 50 μl 补加了 40 ng/ml 重组人 HGF 的无血清的 F-12 培养基，以产生 20 ng/ml HGF 的终浓度，即 HGF 的 EC50。在 37 $^{\circ}\text{C}$ 、5% CO_2 温育平板另外 48 小时，然后每孔用 2 μCi 氚标记的甲基胸苷脉冲另外 15 小时。脉冲后，取出培养基，并替换为 200 μl 0.05% 胰蛋白酶/孔，并在 37 $^{\circ}\text{C}$ 温育平板 5 分钟。然后，使用 Tomtec Harvester 96，将标记的细胞收获到玻璃纤维滤器。然后，通过闪烁计数，测量掺合的标记。

将人 c-MET 的胞外域与免疫球蛋白 Fc 域的重组融合体(c-METFc)用作这些实验中的阳性对照(R&D Systems)。将 c-METFc 的滴定剂(titration)与重组人 HGF 混合至 20 ng/ml HGF 的终浓度，并在 37 $^{\circ}\text{C}$ 温育 1 小时。然后，将该 c-METFc 和 HGF 的混合物加给 96 孔平板中的血清-饥饿的 A549-SC 细胞。然后，以与用单体或多聚体处理的那些相同的方式，加工这些细胞。

图 9 显示了 c-METFc、c-MET-特异性的单体 (M26)和 c-MET-特异性的二聚体 (RM12; RecM12)之间关于它们阻断血清-饥饿的 A549-SC 人肺腺癌细胞的 HGF-诱导的增殖的相对能力的对比。

二聚体 RM12 的 IC50 是 0.32 nM。c-METFc 的 IC50 是 1.73 nM。(对于所有数据点，n=3) 单体 M26 在该基于细胞的测定中，几乎没有表现出可检测的抑制活性。

该测定提供了使用人细胞在体外生物测定中针对抗-c-MET 活性筛选单体或多聚体的方法。通过测定测试的多聚体或单体的 IC50 值，可以在它们的生物学活性的基础上，鉴别和排列最佳的分子。

实施例 8

本实施例描述了证实单体与 c-MET-表达人细胞系的结合的实验。

构建了包含流感血凝素(HA) 附加表位的单体。这使得单体可以用作最初的流式细胞仪检测试剂，而使荧光标记的抗-HA 第二抗体用作第二种检测试剂。

测试通过针对 c-MET 的淘选选择的 15 种单体结合 A549 人肺腺癌细胞(一种 c-MET-表达细胞系)的能力。Jurkat T 细胞用作 c-MET-阴性对照细胞系。

使用溶于磷酸缓冲盐水(pH 7.4)中的 10 mM EDTA，从组织培养平板收获贴壁的 A549 细胞。通过离心，从培养基取出 Jurkat T 细胞。为了测定单体结合，在冰上，用溶于 100 μ l 流式细胞仪染色缓冲液(“FACS 缓冲液:” PBS pH 7.4、5%胎牛血清、0.01% 叠氮化钠)中的 10 μ M c-MET 单体对 2.5×10^5 细胞染色 30 分钟。用 4 ml 冰冷的 FACS 缓冲液洗涤细胞一次，然后重新悬浮在 100 μ l FACS 缓冲液+ 0.2 μ g FITC-缀合的抗-HA 单克隆抗体(Santa Cruz Biotechnology)中，并在冰上温育 30 分钟。用 4 ml 冰冷的 FACS 缓冲液洗涤细胞一次，然后重新悬浮在 200 μ l FACS 缓冲液中，并使用 FACSCalibur 流式细胞仪(BD Biosciences)进行分析。使用 CellQuest Pro (BD Biosciences)，收集和分析数据。测定 A549 和 Jurkat T 细胞的几何平均荧光，并针对被单独的 FITC-缀合的抗-HA 单克隆抗体染色的细胞系的几何平均荧光进行标准化。

下面解释了 c-MET-特异性的单体与 c-MET 阳性的 A549 细胞(而不是 c-MET 阴性的 Jurkat T 细胞)的优先结合。

单体#	几何平均荧光(荧光素通道) vs. 单独的第二抗体	
	A549	Jurkat
单独的第二抗体	1	1
1	1.19	0.97
2	1.69	1.01
3	1.60	1.05
4	1.77	1.05
5	1.72	1.11
6	1.72	1.09
7	1.70	1.11
8	1.46	1.05
9	1.48	1.05
10	1.56	1.04
11	2.03	1.07
12	2.42	1.12
13	2.41	1.10
14	2.80	1.13
15	2.22	1.12

这些数据表明，抗-c-MET 单体结合 A549 人肺腺癌（一种 c-MET 阳性的细胞系），但是不结合 Jurkat T 细胞（一种 c-MET 阴性的细胞系）。该基于流式细胞仪的方法可以用于证实特定的单体在其它细胞表面蛋白的背景下与靶的结合。除了证实单体结合天然的 c-MET 以外，该方法也表明，单体会表现出微弱的或不表现出与细胞的非特异性结合。

实施例 9

本实施例描述了设计用于表明 HGF-诱导的细胞分散的单体抑制的实验。

HGF 被鉴别为“分散因子”，其在上皮细胞中诱导能动的表型。添加 HGF 后，上皮细胞簇分裂开，且细胞彼此远离地迁移，或分散。基于形成紧密簇，其在添加重组人 HGF 后，经 24 小时分散，通

过有限稀释，分离 A549 人肺腺癌的单细胞克隆(称作 A549-SC)。在所有后续实验中，使用该克隆。

以 25 细胞/孔，将 A549-SC 铺平板到 96 孔平板中的补加了 10% FBS 的 F-12 培养基中。培养细胞，直到可以看到 20 至 30 个细胞的簇，大约 4 天。

4 天后，从细胞取出培养基，并替换为单体稀释液，其体积为 50 μ l/孔无血清的 F-12 培养基。另外，将人 c-MET 的胞外域与免疫球蛋白 Fc 域的重组融合蛋白(c-METFc)用作这些实验的阳性对照(R&D Systems)。在 37 $^{\circ}$ C、5% CO₂ 温育 1 小时后，加入 50 μ l/孔的在无血清的 F-12 培养基中的 40 ng/ml 重组的 HGF，以产生 20 ng/ml HGF 的终浓度。也包括缺少 HGF 的对照细胞。然后，在 37 $^{\circ}$ C、5% CO₂ 温育平板 24 小时。24 小时后，从平板取出培养基，用 100% 甲醇在室温固定细胞 15 分钟，然后，用溶于 30%乙醇中的 0.2%结晶紫，在室温染色 1 小时。用磷酸缓冲盐水洗涤染色的细胞，然后照像。

20 ng/ml (大约 EC50)的 c-MET 配体诱导 A549-SC 细胞的细胞分散反应；如预期的，无关的特异性的单体(阴性对照)不抑制该分散反应。相反地，0.5 μ M c-MET-Fc(阳性对照)和 1 μ M 抗-c-MET 单体 Avimer 部分地逆转 HGF-诱导的分散反应。这些数据说明，抗-c-MET 单体可以抑制分散反应，至少到与可比较浓度的阳性对照抑制剂 c-MET-Fc 类似的程度。

c-MET 结合单体和二聚体

下面提供了通过序列同源性鉴别、分组的 c-MET 单体的总结。存在 10 个家族，其中相同家族的成员具有相关的序列。

信息可以总结如下。方括号(“[]”)中的序列表示在单个位置的替代氨基酸。

所有 10 个家族的基序：

家族共有序列(句点(“.”)表示任意的氨基酸；间隔仅仅是为了比对目的。一行包括一个连续的多肽)：

Fam 1	c...	[eq]f.c.	st.r	c[iv] ...	w.cdgndnced.sde.
Fam 2	c....	[eq]fec.	st.r	c[iv] ...	w.cdg.ndnced.sde.
Fam 3	c....	[eq]f.c.	st.r	c[ilv]p..	w.cdg..dnced.sde..
Fam 4	c...	[eq]fqc.	st.r	c[iv] p..	w.cdg.ndncedssde..c
Fam 5	c....	[eq]f.c.	c[ilv]...dc.d.sde.
Fam 6	c...	[eq]f.c.	stgr	c. p..	w.c.g.ndnced.sde.
Fam 7	c....	[eq]f.c.	st.r	c[ilv]...	w.c....dc.d.sd.....c.
Fam 8	c...	[eq]f.c.	c[ilv]...	w.cdg.ndc.d.s.e....c
Fam 9	c....	[eq]f.c.	st.r	c[ilv]p..	w.c.g..dc.d.sde.
Fam 10	c....	[eq]f.c.	c[ilv]...	w.cdg..dc.d.sde.

天然 A-域:

c(.)... . f.c. ...(.) c[ilv]... ..cd...dc.d.sde.(.....)c

A1 文库:

c..(.)	.	f.c.	ccdg..dc.d.sde..(.)c
a	e			I	p	ll	dp
p	q			l	s	wr	en
s	k			v		v	vd

基于家族 10 比对, 本发明提供了包含非天然发生的单体域的多肽, 该单体域结合 c-MET, 且具有紧挨在 A 域支架的第三个半胱氨酸之前的序列 GR 或 KR。

每个 c-MET-结合家族的细节如下。插入破折号(“-”)是为了比对目的, 而不表示蛋白中的位置。

Fam1

```

CAPSEFTCNSTGRCIPQEWVCDGDNDCEDESSEAPDLCASAAPT
CAPSEFTCNSTGRCIPQEWVCDGDNDCEDESSEAPDLCASAAPT
CAPSEFTCNSTGRCIPQEWVCDGDNDCEDESSEAPDLCASAAPT
CAPSEFTCNSTGRCIPQEWVCDGDNDCEDESSEAPDLCASAAPT
CAPSEFTCNSTGRCIPQEWVCDGDNDCEDESSEAPDLCASAAPT
CAPSEFTCNSTGRCIPQEWVCDGDNDCEDESSEAPDLCASAAPT
CAPSQFTCNSTGRCIPQEWVCDGDNDCEDESSEAPDLCAIAAPT
CLANEFTCRSTGRCIPQTVVCDGDNDCEDESSEAPDLCASAAPT
CESNEFQCRSTNRCIPLQWVCDGDNDCEDESSEAPDLCASAAPT
CESNEFQCSSTGRCIPQAWVCDGDNDCEDESSEAPDLCASAAPT
CRANEFQCHSTGRCIPASWLCGDNDCEDESSEAPDLCASAAPT
CEPNEFQCRSTGRCISLAWVCDGDNDCEDESSEAPALCKASVPT
CPASEFTCRSTGRCISQGWVCDGDNDCEDESSEAPALCATTTGPT
CPAGQFTCRSTNRCIPLQWVCDGDNDCEDESSEAPALCATTTGPT
CPASQFTCRSTDRCIPLAWVCDGDNDCEDESSEAPALCATTTGPT
CQASQFTCRSTGRCIPLDWVCDGDNDCEDESSEAPALCAAPAPT
CESNEFQCRSTGRCVPLSWVCDGDNDCEDESSEAPALCKTPGHT
CESNEFQCHSTGRCIPQAWLWVCDGDNDCEDESSEAPALCKTPGHT
CRSNEFTCRSTERCIPLGWVCDGDNDCEDESSEAPXIRKTPGHT
CPANEFKCHSTGRCISLAWVCDGDNDCEDESSEAPALCKTPGHT

```

Fam2

```

CQSFTEFECHSTGRCIPLQWVCDGDNDCEDESSEAP-----ATCATPGHT

```

CQSFTFEFCHSTGRCIPASWLCDGDNDCEDSSDESP-----ANCAATPAHT
 CQSFTFEFCHSTGRCIPVEWLCDGDNDCEDSSDEAP-----AICKTPGHT
 CQSFTFEFCHSTGRCIPASWLCDGDNDCEDSSDE-----EGCEAAAPT
 CQSFTFEFCHSTGRCIPVDWLCDGDNDCEDSSDE-----KCKQ--HT
 CQSFTFEFCHSTGRCIPVDWLCDGDNDCEDSSDE-----KCKQ--HT
 CQSFTFEFCHSTGRCIPRTWLCDGDNDCEDSSDE-----KCKQ--HT
 CQSFTFEFCHSTGRCIPVDWLCDGDNDCEDGSDE-----KSCPA--HT
 CQSFTFEFCHSTGRCIPVDWLCDGDNDCEDSSDE-----KNCQP--PT
 CHPTAEFECNSTGRCIPVDWLCDGDNDCEDSSDE-----KNCKA--HT
 CHPIAEFECNSTSRCIPRTWLCDGDNDCEDSSDE-----ANCQP--PT
 CHPTAEFECNSTGRCVSADWLCDGDNDCEDGSDESP-----ALCK--APT
 CHPTAEFECNSTGRCVSADWLCDGDNDCEDGSDESP-----ALCK--APT
 CHPTAEFECNSTGRCVSADWLCDGDNDCEDGSDESS-----APCETTGP
 CHPTSEFECRSTARCIPLTWVCDGDNDCEDSSDEK-----HCQPP--T
 CHPTSEFECRSTARCIPLTWVCDGDNDCEDSSDEAP-----AICKTPGHT
 CHAPTQFECHRSTNRCIPLQWVCDGDNDCEDSSDE-----TGCAK--PT
 CHPTQFECHRSTGRCIPLQWVCDGDNDCEDSSDE-----TGCAK--PT
 CHAPTQFECHRSTGRCIPLQWVCDGDNDCEDSSDES-----LATCQ--HT
 CNAPNQFECHRSTSRCIPLGWVCDGVNDCEDSSDE-----TDCQE--PT
 CHAPTQFECHRSTGRCIPRDWVCDGDNDCEDSSDEA-----SCGAPG--PT
 CQASDQFECKSTGRCIPLAWRCDGDNDCEDGSDESPAICGRPGLEASG--GS
 CQASDQFECKSTGRCIPLAWRCDGVNDCEDGSDE-----AGCAASG--PT
 CQASDQFECKSTGRCIPLDWLCDGVNDCEDSSDE-----ALENCA-QHT

Fam3

CG-SSEFQCHSTGRCIPENWVCDGDDDCEDSSDEK--SCTSAAPT-----
 CG-SSEFQCHSTGRCIPENWVCDGDDDCEDSSDEK--SCTSAAPT-----
 CG-SSEFQCHSTGRCIPENWVCDGDDDCDDSSDEK--SCTSAAPT-----
 CE-SNEFQCRSTGRCIPRTWVCDGDNDCEDSSDEK--SCTTPAPT-----
 CE-SNEFQCRSTGRCVPVAWVCDGDNDCEDSSDET--GCKAPT-----
 CE-SNEFQCRSTGRCVPVAWVCDGDNDCEDSSDET--GCAKPT-----
 CE-SNEFQCRSTGRCVPVAWVCDGDNDCEDSSDEK--NCKAHT-----
 CE-SNEFQCRSTGRCVPVAWVCDGDNDCEDSSDEK--NCKAPT-----
 CE-SNEFQCRSTGRCVPVAWVCDGDNDCEDSSDEK--DCSAPASEPPGSL-----
 CE-SNEFQCRSTGRCVPVAWVCDGDNDCEDSSDEA--NCGDSHILPFSTPGPST-
 CE-SNEFQCRSTGRCVPVAWVCDGDNDCEDSSDEK--DCGDSHILPFSTPGPST-
 CE-SNEFQCRSTGRCIPVSWVCDGDNDCEDSSDEA--SCGDSHILPFSTPGPST-
 CE-SNEFQCRSTGRCVPVAWVCDGDNDCEDSSDEA--SCG-----APGPT--
 CE-ASEFTCRSTNRCIPVDWVCDGDNDCEDSSDEK--GCGDSHILPFSTPGPST-
 CE-ASEFTCRSTNRCIPVDWVCDGDNDCEDSSDEK--GCGDSHILPFSTPGPST-
 CE-ASEFTCRSTNRCIPQDWVCDGDNDCEDSSDEK--GCGDSHILPFSTPGPST-
 CE-ASEFTCRSTNRCIPLQWVCDGDNDCEDSSDEA--NCGDSHILPFSTPGPST-
 CP-AGQFTCRSTNRCIPLQWVCDGDNDCEDSSDEA--NCGDSHILPFSTPGPST-
 CE-ASEFTCRSTNRCIPANWVCDGDNDCEDSSDEA--NCGDSHILPFSTPGPSX-
 CE-PSQFTCRSTSRCIPRTWLCDGDNDCEDSSDEAPALCGDSHILPFSTLGPST-
 CL-SSEFTCKSTNRCIPRAWVCDGDNDCEDSSDEAPALCGDSHILPFSTPGPST-
 CX-XSQFXCRSTGRCIPAEWVCDGDNDCEDSSDEA--SRGDSHILPFSTPGPST-
 CQ-ADQFQCRSTSRCIPAPWVCDGVNDCEDGSDET--SCGDSHILPFSTPGPST-
 CR-ADQFQCRSTNRCLPGPWVCDGVNDCEDGSDET--GCGDSHILPFSTPGPST-
 CQ-TGEFRCRSTDRCI PAEWVCDGSDCEDGSDET--NCGDSHILPFSTPGPST-
 CA-SNEFRCRSTGRCIPQRWVCDGDNDCEDGSDET--NCGDSHILPFNTPGPIT-
 CQSFTFEFCHSTGRCIPVDWLCDGDNDCEDSSDEK--GCGDSHILPFSTPGPST-
 CQSFTFEFCHSTGRCI PAEWVCDGDNDCEDSSDEK--GCGDSHILPFSTPGPST-
 CHPTAEFECNSTGRCIPVDWLCDGDNDCEDSSDEK--GCGDSHILPFSTPGPST-
 CHAPTQFECHRSTGRCIPLTWVCDGDNDCEDGSDEK--DCGDSHILPFSTPGPST-
 CX-PSEFTCKSTGRCIPLDWVCDGDNDCEDSSDEK--GCGDPHILPFSTPGPST-
 CA-ADEFQCNSTGRCIPVSWVCDGDNDCEDSSDEK--GCGDPHILPFSTPGPST-
 CL-ANEFTCRSTGRCIPLQWVCDGDNDCEDSSDEK--GCGDSHILP----GLEAS
 CE-ASEFTCRSTNRCIPLQWVCDGDNDCEDSSDEK--GCGDSHILPFSTPGLEAS

CA-SSEFRCRSTGRCIPQRWVCDGDNDCEDESSEK--GCGDSHILPFSTPGLEAS
 CA-SSEFRCRSTGRCIPQRWVCDGDNDCEDESGDET--NCGDSHILPFSTPGLEAS
 CR-SNEFTCRSTGRCIPRTWVCDGDNDCEDESGDESPAICGDSHILPFSTPGLEAS
 CE-SNEFQCQSTSRCIPLTWRCGDNDCEDESSEDER--SCKPPT-----
 CE-SNEFQCQSTSRCIPLTWRCGDNDCEDESSEK--SCSAPASEPPGLEAS---
 CE-SNEFQCQSTSRCIPLDWVCDGDNDCEDESSEDEA--SCGAPGPT-----
 CE-SNEFQCQSTSRCIPLTWRCGDNDCEDESSEDEA--SCGAPGPT-----
 CG-SDEFQCKSTSRCIPLTWRCGDSDCEDSSDEA--NCGRPGLEASGGS-----
 CQ-SNEFTCQSTNRCLPLPWVCDGDNDCEDESSEDEA--NCGQRT-----
 CA-ADEFQCNSTGRCIPAAWVCDGDNDCEDESSEDEA--SCGAPGPT-----
 CE-ADEFQCRSTNRCIPLQWVCDGDNDCEDESSEDEA--SCGAPGPT-----
 CA-ADEFQCRSTNRCIPLQWVCDGDNDCEDESSEDEA--NC-TP-PT-----

Fam4

CAPNEFQCSSTSRICIPQRWLCGDNDCEDESSEANCAK--HT
 CAPNEFQCSSTSRICIPQRWLCGDNDCEDESSEANCAK--HT
 CVSSEFQCHSTGRCIPRTWLCGDNDCEDESSEANCAK--HT
 CAPSEFQCQSTKRCIPEGWLCGVNDCEDSSDEKGEAPVRT
 CAPSEFQCQSTKRCIPEGWLCGVNDCEDSSDEKGEAPVRT
 CVPSEFQCQSTGRCIPRTWLCGDNDCEDESSEADCKAPGHT
 CPSDQFQCSSTNRICIPRSWLCGVNDCEDSSDEADCAA--PT
 CPSDQFQCSSTNRICIPRSWLCGVNDCEDSSDEADCAA--PT
 CESNQFQCHSTGRCVPSWRCDGVNDCEDSSDETDCAP--PT
 CLSNQFQCNSTNRICIPQHWLCGDNDCEDESSEKGAATGHT
 CGADQFQCQSTNRICVPQRWVCDGDNDCEDESSEKDKCATART

Fam5

CPX-NEFRCG-NGRCLPLRLRCXXENDCGDXSDE--E-----NCSAPASEPPG
 CPA-GEFQCK-NGRCLPPAWLCGDNDCEDESSE--T-----GCSAPASEPPG
 CQA-DQFPCS-NGHCVQTLVXDGVPCQDDSDSDE--T-----NCSAPASEPPG
 CLA-DEFPCCHSTGRCIPAAWLCGDNDCEDESGSDE--T-----NCSAPASEPPG
 CAA-DEFQCQSTGRCIPVRWLCGDNDCEDESGSDE--T-----SCSAPASEPPG
 CLA-NQFQCRSTGRCISRDWVCDGVNDCEDSGSDE--T-----SCSAPASEPPG
 CAA-DQFQCRSTGRCIPRTWLCGDVNDCEDESGSDEPLA-----LCSAPASEPPG
 CNT-TQFSCRSTNRCIPLDWQCDGVTDCEDNSDE--A-----SCSAPASEPPG
 CLP-SEFQCKSTNRICIPQAWLYDGVNDCEDSGSDE--T-----NCSAPASEPPG
 CLP-SEFQCKSTNRICIPQAWLYDGVNDCEDSGSDE--T-----SCSAPASEPPG
 CLP-SQFQCNSTNRICIPLAWLYDGVNDCEDSSDE--X-----SCSAPASEPPG
 CXP-SQFTCHSTDRCIPLWLCGDNDCEDESSE--T-----GCSAPASEPPG
 CEP-NQFTCHSTSRICIPQPWRCGDVNDCEDESGSDEALA-----TCSAPASEPPG
 CEP-NQFTCHSTSRICIPQPWRCGDVNDCEDESGSDEALA-----TCSAPASEPPG
 CES-NEFQCQSTSRICIPASWLCGDVNDCEDESGSDE--T-----NCSAPASEPPG
 CGS-DEFQCKSTRRCIPLNWLCGDVNDCEDESSEPPA-----TCSAPASEPPG
 CGS-DEFQCKSTRRCIPLNWLCGDVNDCEDESSEPPA-----TCSAPASEPPG
 CGS-DEFQCKSTRRCIPLNWLCGDVNDCEDESSEPPA-----TCSAPASEPPG
 CPS-DEFQCNSTGRCISLTWLCGDVNDCEDESGSDE--K-----SCSAPASEPPG
 CPS-DEFQCNSTGRCISLTWLCGDVNDCEDESRDE--K-----SCSAPASEPPG
 CQSFTTEFECHSTGRCIPVDWLCGDNDCEDESSESPA-----ICSAPASEPPG
 CQSFTTEFECHSTGRCIPVDWLCGDNDCEDESSESPA-----ICSAPASEPPG
 CQSFTTEFECHSTGRCIPVDWLCGDNDCEDESSESPA-----HCSAPASEPPG
 CHPTAEFECHSTGRCIPVAWLCGDNDCEDESSESPA-----HCSAPASEPPG
 CDPIAEFKCHSTGRCIPLDWLCGDVNDCEDESSESPA-----HCSAPASEPPG
 CPS-DEFKCHSTGRCLPVEWLCGDVNDCEDESGSDE--A-----SCSAPASEPPG
 CPP-NEFQCHSTGRCISRDWLCGDVNDCEDESSEAPD-----LCGAPASEPPG
 CVA-SEFTCRSTGRCIPESWRCDGDNDCEDESSESPD-----LCSAPASEPPG
 CGA-SEFQCRSTGRCLPQHWRCGDNDCEDESSEED-----CSAPASEPPG
 CQA-SEFTCHSTGRCLPRAWLCGDNDCEDESGSDEED-----CSAPASEPPG

CES-SEFRCRSTGRICIPGGWLCDGDNDCEDSSDETD-----CSAPASEPPG
 CAA-DEFQCNSTGRICIPVSWVCDGDNDCEDSSDESPD-----LCSAPASEPPG
 CAA-DEFQCNSTGRICIPVSWVCDGDNDCEDSSDESSA-----HCSAPASEPPG
 CAA-DEFQCNSTGRICIPVSWVCDGDNDCEDSSDEKD-----CSAPASEPPG
 CSS-DEFQCSSTSRICIPREWVCDGDNDCEDSSDEKD-----CGAPASEPPG
 CLA-NEFTCRSTERCIPLGWVXXGDNDCEDSSDE--E-----NCSASASEPPC
 CLX-NEFTCRSTNRCIPLQWVCXGXNDCEDSSDE--K-----NCSAPASEPPG
 CLA-NEFTCRSTNRCIPLQWVCDGDNDCEDSSDE--A-----GCSAPASEPPG
 CLS-NEFTCRSTKRCLPRQWVCDGDNDCEDSSDE--E-----DCSAPASEPPG
 CGS-NQFTCRSTKRKITATWVCDGDNDCEDSSDE--T-----DCSAPASEPPG
 CQA-NEFTCRSTSRICIPATWVCDGDNDCEDSSDE--K-----DCSAPASEPPG
 CES-NEFQCQSTGRICISRWDVCDGDNDCEDSSDE--E-----DCSAPASEPPG
 CES-NEFQCQSTSRICILTWRCDGDNDCEDSSDEAPE-----HCSAPASEPPG
 CES-NEFQCQSTSRICIPREWVCDGDNDCEDSSDESSA-----HCSAPASEPPG
 CES-NEFTCRSTGRICIPASWLCDGDNDCEDSSDESSA-----HCSAPASEPPG
 CRS-NEFTCRSTERCIPLGWVCDGDNDCEDSSDESSA-----HCSAPASEPPG
 CXSFTEFECRSTGXCIPLTWVCDGDNDCEDSSDE--E-----XCSAPASEPPG
 CEA-SEFTCRSTNRCIPLDWVCDGDNDCEDSSDEKS-----CSAPASEPPG
 CEA-SEFTCRSTNRCIPVDWVCDGVNDCEDSSDESSD-----ICSAPASEPPG
 CVP-SEFQCRSTNRCIPLDWVCDGDNDCEDSSDEKS-----CSAPASEPPG
 CVS-GEFTCRSTNRCIPVDWVCDGDNDCEDSSDEKD-----CGAPASEPPG
 CEP-SQFTCRSTNRCIPQEWVCDGDNDCEDGSDE--K-----SCSAPASEPPG
 CEP-SQFPCHSTNRCIPLAWVCDGDNDCEDSSDE--K-----NCSAPASEPPG
 CES-SQFTCNSTKRICIPAWVCDGDDDCEDGSDE--K-----SCEAPAHT---
 CQP-SQFTCHSTDRICIPLEWLCDGDNDCEDSSDE--K-----NCKAHT-----
 CQP-SQFTCHSTDRICIPLEWLCDGDNDCEDSSDE--K-----NCKAHT-----
 CLP-SQFTCHSTDRICIPLEWLCDGDNDCEDSSDE--K-----NCKAHT-----
 CQP-DQFTCHSTDRICIPLEWLCDGDNDCEDSSDE--K-----NCKAHT-----
 CCP-NQFTCHSTDRICIPLEWLCDGDNDCEDSSDE--K-----NCKAHT-----
 CQP-SQFTCXRTDRICIPLEWLCDGDNDCEDSSDE--K-----NCKAXT-----

Fam6

CAADEFQCNSTGRICIPVSWVCDGVNDCEDSSDEAGC-ATSGPT--
 CAADEFQCNSTGRICIPVSWVCDGVNDCEDSSDEAGC-ATSGPT--
 CAADEFQCNSTGRICIPVSWVCDGVNDCEDSSDEAGC-ATSGPT--
 CAADQFQCHSTGRICIPVSWVCDGVNDCEDSSDEAGC-ATSGPT--
 CAADEFQCNSTGRICIPVSWVCDGVNDCEDSSDETDC-APH--T--
 CAADEFQCNSTGRICIPVSWVCDGDNDCEDGSDESPA-LCKAPT--
 CAADEFQCNSTGRICIPQEWVCDGVNDCEDSSDESPA-LCKAPT--
 CAADEFQCNSTGRICIPVSWVCDGDNDCEDSSDEES---CETPT--
 CAASEFQCRSTGRICIPVEWXCDGDNDCEDSSDETG---CKXPT--
 CESDEFQCHSTGRICIPLDWVCDGDNDCEDSSDE--K-DCKQHT--
 CESDEFQCHSTGRICIPLDWVCDGDNDCEDSSDE--K-DCKQHT--
 CESNEFQCHSTGRICIPLQWVCDGDNDCEDSSDE--K-DCKXHT--
 CVS-NEFQCHSTGRICIPREWRCGDNDCEDSSDE--K-DCKQHT--
 CESNEFQCHSTGRICIPASWLCDGDNDCEDSSDE--K-DCKQHT--
 CRANEFQCHSTGRICIPASWLCDGDNDCEDSSDE--K-DCKQHT--
 CLANEFTCRSTGRICIPASWLCDGDNDCEDSSDE--K-DCKQHT--
 CEASEFQCNSTGRICIPRQWLCDGDNDCEDSSDE--K-DCKQHT--
 CAASEFQCNSTGRICIPASWLCDGDNDCEDSSDESLA-TCQQHT--
 CPPDEFRCNSTGRICIPRAWLCDGDNDCEDSSDE--K-DCKQHT--
 CPPDEFRCNSTGRICIPRAWLCDGDNDCEDSSDE--K-DCKQHT--
 CPPDEFRCNSTGRICIPRAWLCHGDNDCEDSSDE--K-DCKQHT--
 CPPDEFRCNSTGRICIPRAWLCHGDNDCEDSSDE--K-DCKKHT--
 CPPDEFRCNSTGRICIPRAWLCHGDNDCEDSSDE--K-DCKPHT--
 CLANEFTCRSTGRICIPRAWLCDGDNDCEDSSDE--K-DCKQHT--
 CQTGEFRCNSTGRICIPRAWVCDGDNDCEDSSDE--K-DCKQHT--
 CRADEFQCRSTGRICIPGAWRCDGDNDCEDSSDE--K-DCKQHT--
 CAADEFQCNSTGRICIPVSWVCDGDNDCEDSSDE--K-DCKQHT--
 CAADEFQCNSTGRICIPLQWVCDGDNDCEDSSDE--K-DCKQHT--

CAADEFQCNSTGRICIPVSWVCDGDNDCEDESSDE--K-NCKAHT--
 CAADEFQCNSTGRCPAEWVCDGDNDCEDESSDE--K-NCKAHT--
 CAPSEFTCNSTGRICIPQEWVCDGDNDCEDESSDE--K-DCKQHT--
 CQPNEFQCHSTGRICIPASWVCDGDNDCEDESSDESPA-NCATPHT
 CVPNEFQCNSTGRICIPQAWVCDGVNDCEDESSDESSA-LCSEPT--
 CEPDEFQCRSTGRICIPLEWVCDGDNDCEDESSDET---GCAKPT--
 CPPDEFRCNSTGRICIPLAWVCDGDNDCEDESSDET---NCQPPT--
 CAAGEFQCNSTGRICIPAAWVCDGDNDCEDESSDEEGC-GAAEPT--
 CQLDQFRCRSTGRICIPQAWVCDGDNDCEDESSDEEGC-GAAEPT--

Fam7

CP-ADQFTCRSTDRICIPGDWVCDVNDCEDEGSDEK--NCLER-----T---
 CP-ADQFTCRSTDRICIPGDWVCDVNDCEDEGSDEK--NCLER-----T---
 CG-SDQFQCRSTDRICIPRTWVCDGDNDCEDEGSDEK--DCTRS-----VPT-
 CQ-SGQFQCKSTGRICIPRTWVCDGDNDCEDESSDEK--NCQPP-----T---
 CQ-SGQFQCKSTGRICIPRTWVCDGDNDCEDESSDEK--NCQPP-----T---
 CA-SDQFQCRSTGRICIPQHWVCDGDNDCEDEGSDEK--NCGPPGPSAISTAAG
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CR-ANEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CL-ANQFPCSTGRICIPASWVCDGDNDCEDEGSDE-SHLCTA-----HT---
 CR-ANEFPCSTGRICIPASWVCDGDNDCEDEGSDE-SHLCTA-----HT---
 CL-SNEFPCRSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CE-SGEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CE-SGEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CE-SGEFQCHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CEPSGQFECHSTGRICIPASWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CL-ADEFQCHSTGRICIPQAWVCDGDNDCEDEGSDE-SQLCTA-----HT---
 CE-ASEFTCRSTDRICIPVHWVCDGVNDCEDEGSDE-AQVCTE-----HT---
 CL-ADEFRCSTNRCIPLDWVCDGVNDCEDEGSDE-AQVCTE-----HT---
 CP-AGQFTCRSTNRCIPLQWVCDGVNDCEDESSDE-SQHCPP-----HT---
 CP-AGQFTCRSTNRCIPLQWVCHGVNDCEDESSDE-SQHCPP-----HT---
 CA-SDEFTRCRTRRCIPQTWVCDGDNDCEDEGSDE-TD-CPP-----HT---
 CP-SSQFTCHSTRRCIPLQWVCDGDNDCEDEGSDE-TD-CPP-----HT---
 CV-SNEFTCRSTKRCVPQEWVCDGVNDCEDEGSDE-TG-CPK-----HT---
 CV-SNEFTCRSTKRCVPQEWVCDGVNDCEDEGSDE-TG-CPK-----HT---
 CV-SNEFTCRSTKRCVPQEWVCDGVNDCEDEGSDE-TG-CPK-----HT---
 CV-SNEFTCRSTKRCVPQEWVCDGVNDCEDEGSDE-TG-CPK-----HT---
 CQ-ANQFKCRSTSRCIPLAWVCDGDNDCEDEGSDE-EG-CKP-----HT---
 CA-SGQFQCRSTGRCLPLPWVCDGDNDCEDEGSDEAPAICEK-----H--T-
 CA-SSEFQCKSTERCLPLEWVCDGVNDCEDEGSDEAPAICTT-----PGPT-
 CP-PSQFQCRSTGRCIPLHWVCDGVNDCEDEGSDEPPEPCTA-----TVPT-
 CQ-PNQFQCHSTGRCLPLDWVCDGVNDCEDEGSDESSAPCET-----TGPT-
 CE-SSQFQCRSTGRCLPPDWVCDGVNDCEDEGSDEAG--CQP-----HR---
 CE-ASEFQCRSTKRCLPRHWVCDGDNDCEDEGSDEKS--CPA-----HT---
 CE-ASEFQCRSTKRCLPRHWVCDGDNDCEDEGSDEKS--CPA-----HT---
 CE-ASEFQCRSTKRCLPRHWVCDGDNDCEDEGSDEKS--CPL-----H---
 CR-SGQFQCRSTNRCISRTWVCDGDNDCEDEGSDEASAIKES-----SEHT-
 CP-PDEFRCSTNRCISRTWVCDGDNDCEDEGSDEASAIKES-----SEHT-
 CE-SNEFQCRSTSRCIPLTWVCDGVNDCEDEGSDE--ANCTA-----AVHT-
 CE-SNEFQCRSTSRCIPLTWVCDGVNDCEDEGSDE--ANCTA-----AVHT-
 CE-SNEFQCRSTSRCIPLTWVCDGVNDCEDEGSDE--ANCTA-----AVHT-

CPP-DEFTCRSTERCIPLAWVCDGDNDCEDESSDEAG--CT-----TPEPT-
 CPP-DEFTCRSTERCIPLAWVCDGDNDCEDESSDEAG--CT-----TPEPT-
 CPP-DEFTCRSTERCIPLAWVCDGDNDCEDESSDEAG--CT-----TPEPT-
 CPP-DEFTCRSTERCIPLAWVCDGDNDCEDESSDEAG--CT-----TPEPT-
 CPP-DEFTCRSTERCIPLAWVCDGDNDCEDESSDEAG--CT-----TPEPT-
 CPP-DEFTCRSTERCIPLAWVCHGDNDCEDESSDEAG--CT-----TPEPT-
 CAP-SEFTCRSTGRICIPRTWLCDGDNDCEDESSDEAPALCT-----TPVPA-
 CPA-SEFQCHSTGRICIPASWLCDGDNDCEDESSDEAG--CT-----TPEPT-
 CES-GEFQCHSTGRICIPASWLCDGDNDCEDESSDEAG--CT-----TPEPT-
 CAS-XEFKC-SSGRCLPPSWLCDGXXDCEDGSDEAN--CT-----XPVPT-
 CPS-GEFQCRSTNRICIPETWLCDGEDDCGDSSDESALCGRPG--PGATSAPAA
 CVS-GEFTCRSTNRICIPVDWLCDGDNDCEDESSDEPPALCGRPG--PGATSAPAA
 CES-SEFQCRSTDRCLPVTWVCDGDNDCEDESSDES--NCGRPG--PGATSAXAA
 CHAPTQFECRSTGRICIPAAWVCDGDNDCEDESSDESPAICGRPG--LGXTSAPAA
 CHAPTQFECRSTGRICIPVSWVCDGVNDCEDESSDESPAICGRPG--LGATSAPAA
 CRA-NQFQCHSTGRICIPXSWLCDGVNDCEDESSDESPAICGRPG--LGATSAPAA
 CEA-SEFTCRSTDRCLPVSWVCDGVNDCEDESSDESPAICGRPG--LGATSAPA-
 CLS-NEFRCSSTGRCLPRPVWVCDGDNDCEDESSDEAPAICGRPG--PGATSAPAA
 CLA-SEFTCHSTGRICIPETWVCDGVNDCEDESSDE--ADCGRPG--PGATSAPAA
 CLA-SEFTCHSTGRICIPETWVCDGVNDCEDESSDEDNAHCGRPG--PGATSAPAA
 CPP-NEFTCQSTDRCLPADWVCDGVNDCEDESSDE-ADCGRPG--PGATSAPAA
 CDPIAEFKCHSTGRICIPLDWLCDGDNDCEDESSDEAS--CGRPG--PGGTSAPAA
 CQSFTEFECHSTGRICIPVDWLCDGDNDCEDESSDEAS--CGRPG--PGGTSAPAA
 CR-ANEFQCHSTGRICIPASWLCDGDNDCEDESSDEAS--CGRPG--PGGTSAPAA
 CA-PDEFPCRSTGRICVPLTWLCDGDNDCEDESSDEASATCGRPG--PGATSAPAA
 CRA-NEFQCHSTGRICIPETWLCDGDNDCEDESSDE--ESC-----TPPT-
 CRA-NEFQCHSTGRICIPASWLCDGDNDCEDESSDELEIC-----PQPT-
 CRA-NEFQCHSTGRICIPASWLCDGDNDCEDESSDE--KDC-----KTPGPT-
 CRA-NEFQCHSTGRICIPASWLCDGDNDCEDESSDE--KGCSDSHILPFXTPGPST
 CRA-NEFQCHSTGRICIPASWLCDGDNDCEDESSDE--TGC-----AKPT-
 CRA-NEFQCHSTGRICIPASWLCDGDNDCEDESSDE--TGC-----AKPT-
 CRA-NEFQCHSTGRICIPASWLCDGDNDCEDESSDE--TGC-----AKPX-
 CRA-NEFQCHSTGRICIPQTWLCDGDNDCEDESSDE--AGC-----AASGPT-
 CEA-NEFQCRSTGRICIPLNWLCDGDNDCEDESSDE--TNCG-----TPGPT-
 CEA-SEFTCRSTDRICIPLEWLCDGDNDCEDESSDEAN--CG-----AAART-
 CQS-SEFTCKSTNRICIPALWLCDGVNDCEDESSDEAN--CT-----SPERT-
 CRS-SEFTCRSTSRICIPENWLCDGVNDCEDESSDETG--CG-----TSAPT-
 CRS-SEFTCRSTSRICIPENWLCDGVNDCEDESSDETG--CG-----TSAPT-

Fam10

-CQA-GQFECRSTNRICIPQDWVCDGVNDCEDESSDEE----SCTSPART
 -CQA-GQFECRSTNRICIPQDWVCDGVNDCEDESSDEE----SCTSPART
 -CQA-GQFECRSTNRICIPQDWVCDGVNDCEDESSDEE----SCTSPART
 -CQA-GQFECRSTNRICIPQDWVCDGVNDCEDESSDEE----SCTSPART
 -CQA-GQFQCRSTNRICIPQDWVCDGVXDCEDSSDEE----RCTSPART
 -CPA-GQFQCRSTNRICIPQDWVCDGVNDCEDESSDEE----SCTSPART
 -CEA-NQFRCKSTSRICIPQNWLCDGVNDCEDESSDEE----NCTRAPT
 -CEA-DEFRCRSTGRICISVDWRCDGVSDCEDSSDEE----SCESTAPT
 -CEA-GEFRCKSTDRICIPLAWRCDGVNDCEDESSDEA----SCKSSAHT
 -CLA-NEFTCRSTGRICIPRTWRCDGVNDCEDESSDEA----NCKKPT--
 -CLA-NEFTCRSTGRICIPRTWRCDGVNDCEDESSDEA----NCKKPT--
 -CEA-NEFRCKSTGRICISQWRCDGDDDCEDSSDEA----NCKPPT--
 -CLP-SEFPCS-NGRCVPRPVWVCDGDDDCEDNSDEA----GCPKPT--
 -CEP-GEFPCSSTGRICVPAWHCDGVNDCEDESSDEA----GCQKRT--
 -CQP-DEFRCRNTDICIPQRWVCDGDNDCEDESSDEA----DCQQPT--
 -CQA-DEFRCRGN-GRICIPQRWVCDGDDDCGDGSDXX----DCXTPT--
 -CLA-DEFRCXSNNRCLPLDWVCDGDNDCEDESSDEK----DCAXPT--
 -CPP-DQFPC-DNGDCLPQPWVCDGEXDCPDSSDE----ASCTTSVHT
 -CAA-DQFKC-DNGRCVQNWRCVCDGEXDCGDNSSDE----ENCTTPT--
 SQPI-GQFKC-GNGNCVPRTWRCVNDGCPDSSDE----TDCPTPT--

```

-CEA-GQFRC-NNGNCXPQHWLXCXGEXDCEDNSDE----AXCEKPT--
-CAP-DXFXC-XNGKCLPLDWVCDGEDDCGDNSE----TXCQ-----
-CAS-NQFTC-NNGHCLPQHWRCGEGDDCGDNSE----ASCQPPT--
-CQA-DEFKC-GNGRCLPEAWVCDGEDDCGDNSE----ADCQAPT--
-CQA-DEFRC-GNGRCISPTWVCDGEXDCGDDSE----ANCAATERT
-CQP-GEFRC-RNGKCIPTWLXXGXDDCGDNSE----ADCAATAPT
-CPP-DEFKC-GNGHCISQTLWCDGEXDCGDNSE----ESCA--APT
-CPS-GEFRC-SNGSCIPQEWGCDGXNDCGDDSE----KNCAAAGPT
-CPS-GEFRCQSSNTCIPLNWLCDGEDDCGDDSE----KNCEASVPT
-CLS-GEFRC-SNGNCLPADWLCDGEDDCGDNSE----TSCAASEPT
-CQP-GEFTC-NNGNCLPLEWVCDGENDCGDSSDE----ENCGGSEHT
-CQS-DQFRC-SNGRCIPVEWVCDGEDDCLDGSDEP---QVCGTTAPT
-CPP-DEFRC-SNGRCLPQPWVCDGEDDCGDSDE----TSCAATAPT
-CAS-NQFRC-RNGRCIPLPWVCDGEDDCQDNSE----ASCAAPAPT
-CAS-NQFRC-RNGRCIPLPWVCDGEDDCQDNSE----ASCAAPAPT
-CVA-DEFPCGN-GNCIPLPWRCDGDDDCGDNSE----TDCESSXPT
-CPP-DEFPCNSGICIPRSWRCGEGDDCGDNSE----EDCTSAGHT
-CAP-NEFPCGN-GRCIPATWLCDGDNDCGDNSE----EGCGGSART
-CPP-SEFPCGN-GSCVPQAWVCDGDPDCPDNSDE----EGCTGTGPT
-CPP-DEFRCNN-GKCIPLSWRCGDDDCQDSSDE----AGCT--ERT
-CXP-GEFQC-NNGRCIPATWLCDGDDDCGDNSE----TGCTEHT--
-CXP-GEFQC-NNGRCIPATWLCDGDDDCGDNSE----TGCTEHT--
-CQS-NEFQC-NNGRCISVTWLCDGDDDCGDSSE----TDCTSAVPT
-CPS-SEFQCRNNTKCIPLNWLCDGEDDCGDSSE----TDCTTHT--
-CVP-GEFRCHDSGTCVPLAXLXCXGDNDCGDNSE----ASCESSEPT
-CAP-GQFRCKN-GRCVPLSWVCDGDDDCEDDSDE----ANCESSEPT
-CAA-DQFRCSS-GRCVPLTWLCDGDDDCADDSDE----KDCESTAHT
-CAA-DEFQCNSTGRICIPVSWVCDGEDDCRDDSE----ENCRSSEPT
-CLA-GEFRCNS-GRCIPEHWRCGEGDDCLDSSDE----KDCTTSEPT
-CX-AXQFTC-DNGQCLPQNWVCDGENDCPDXSDE----KNCAPHT--
-CX-SSXFRC-XNGXCLPLXWVCDGENDCGDXSDE----XXC-----
-CV-ADQFRC-DNGRCLSREWVCDGVNDCQDGSDE----TNCQERT--
-CA-AGEFRCRDSGRCLPQHWLCDGENDCADGSDE----TNCTQHT--
-CX-PSEFTC-SSGQCIPEWVCDGXNDCGDDSE----TNCETRT--
-CV-ANEFKC-GSGKCIPEWVCDGDNDCGDSDE----ASCAQPT--
-CG-ANEFKC-SSGSCIPQEWRCGENDCGDNSE--LAPCKEPT--
-CR-ADEFKC-GNGHCIPGQWLCDGENDCQDGSDE----KSCEQPT--
-CL-PNQFQCQSSGRCIPLNWLCDGDDDCGDDSE----TSCKAPT--
-CP-ASEFQCGN-GRCISEHWLCDGDNDCGDNSE----TSCAPVPT
-CQ-ADEFQCRNTEKCIPLNWLCDGDNDCGDDSE----TSCATPT--
-CVA-SEFTCKDTRCIPLHWVCDGVDDCGDNSEAD----CETSVHT
-CEA-NEFRCQSTDRCIPASWVCDGVDDCEDGSDEKS----CTTSVHT
-CEA-SEFTCNSTGRCLPLTWVCDGVNDCEDGSDEKS----CTTSVRT
-CAP-NEFTCSSTGRCLPRAWVCDGVDDCEDGSDETS----CGATVHT
-CGA-NEFTCQSTNRCIPQSWVCDGVNDCEDGSDESPV--LCATTVHT
-CQP-DEFRCRSTGRCLPQEWLCDGVNDCEDSSDEAD----CGTSAHT
-CAP-GEFPCRSTGRCIPTWVCDGVNDCEDSSDEKS----CATAEHT

```

Fam 10 单体域可以进一步分成亚家族(称作“10A”, “10B”等)。

下面列出了各种亚家族的共有基序:

- 10A CxxxEFQCNnGRCI PxxWLCDGDdDCGDxSDETxC
- 10B CPPxEFPCxNGxCIPxxWxCDGDxDCxDNSDEEGCT
- 10C CxAgEFrCxGRCi PLxWxCDGdDDCgDxSDExdCESS

10D CpsGEFRCSNGxCIPqxWlCDGeDDCGDxSDExxCA

10E CxADEFKCGNGrCIpxxWvCDGexDCGDdSDExxSC

产生了所有 **Fam 10** 亚家族的几个共有序列:

10A-E CpaxEFxCxNGrCIPxxWxCDGddDCGDxSDExxC

10A-E CxxxEFxCxNGxCIPxxWxCDGxdDCGDxSDExxC

因而, 本发明包括具有 A 域支架且包含下述序列的 **c-MET**-结合单体域:

EFXCXNGXCIPXXWXCDGXDDCGDXSDE

下面提供了 **c-MET** 结合二聚体, 即, 包含 2 个单体域的多肽, 其中的每个都结合 **c-MET**。在每个二聚体家族后面的表表示基于家族成员的比对的共有基序。应当指出, 下面的“**Fam**”名称指二聚体家族, 其与上面列出的单体家族不同。

Fam 1

CQASDQFECKSTGRCIPLAWRCGDNDCEDEGSDESPAICG-----RPGLEASGGSCRAN-
EFQCHSTGRCI PASWLCGDNDCEDEGSDE-AS-CGRPGPGGTS---APAA

CRAN-
EFQCHSTGRCI PASWLCGDNDCEDESSDEAPDLCASAAPTSLQASGLEASGGSCHAPTQFECRSTGRCI PAAWV
DGDNDCEDEGSDESPAICGRPGLGXTSA--PAA

CESG-EFQCHSTGRCI PASWLCGDNDCEDEGSDES-QLCT-----AHTCAPG-
EFQCHSTGRCI PASWLRDGDNDCEDEGSDES-XLCTA-HX-----

CRSN-EFTCRSTGRCI PRTWVCDGDNDCEDEGSDESPAICGD SHILPFSTPGLASGGSCP-
AGQFTCRSTNRCI PLQWVCDGDNDCEDESSDEAN---CGD SHILPFSTPGPST

CLAN-EFTCRSTGRCI PLQWVCDGDNDCEDESSDEK--GCGD SHILP----GLEASXGSCX-
XSQFXCRSTGRCI PAEWVCDGDNDCEDESSDEAS--CGD SHILPFSTPGPST

CASS-EFRCRSTGRCI PQRWVCDGDNDCEDEGSDET--NCGD SHILPFSTPGLASGGSCQ-
TGEFRCRSTDRCI PAEWVCDGSDCEDEGSDETN--CGD SHILPFSTPGPST

CASS-EFRCRSTGRCI PQRWVCDGDNDCEDESSDEK--GCGD SHILPFSTPGLASGGSCA-
ADQFQCRSTGRCI PRTWLCGDVNDCEDEGSDEPLALCSAPASEP-----PGSL

CEAS-EFTCRSTNRCI PLQWVCDGDNDCEDESSDEK--GCGD SHILPFSTPGLASGGSCG-
SNQFTCRSTKRCI TATWVCDGDNDCEDESSDE-TD-CSAPASEP-----PGSL

CGSD-EFQCKSTSRICIPLTWRCDGSDCEDSSDEA--NCGR-----PGLEASGGSCQ-
SGQFQCKSTGRICIPRTWVCDGDNDCEDSSDEK-N-CQ-P-----PT

CESN-EFQCQSTSRICIPLTWRCDGDNDCEDSSDEK--SCSAPASEP---PGLEASGGSCP-
ASEFTCRSTGRICISQGWVCDGDNDCEDSSDESPAICATTG-----PT

下面的共有序列包括问号(“?”)。它们表示可以存在或不存在的
位置。

分 匹 预
数 配 期 基
的 的 序

1
2 6 10⁵⁶ c.[as]..?[eq]f.c.st.rcip..w.cdgd.dced.sde..??.c..??.??.??.??.??.?
6 56 .??.??.??.??.[st]c..??.?[eq]f.c.st.rci...w[ilv].dgdndced.sde..??.
6 ...??.??.??.??.??.??.??.??.??.?

1
1 5 10⁶⁰ c.[as]..?[eq]f.c.st.rcip..w.cdgd.dced.sde..??.c..??.??.??.??.??.?
2 5 60 .??.??.??.??.ggsc..??.?[eq]f.c.st.rci...w[ilv].dgdndced.sde..??.??.
7 ??.??.??.??.??.??.??.??.??.?

9
5 4 10⁶⁴ c.[as]..?[eq]f.c[kr]st.rcip..w.cdgd.dced.sde..??.cg..??.??.??.??.?
3 64 ??.??.??.??.??.ggsc..??.?[eq]f.c.st.rci...w[ilv].dgdndced.sde..??.c
...??.??.??.??.??.??.??.??.??.?

7
7 3 10⁷⁰ c[ekqr][as]..?[eq]f[eq]c.stgrcip.[as]w.cdgdndced.sde[as].??.[ilv]c.
1 70 .??.??.??.??.??.??.??.??.??.??.??.??.[st]c..??.?[eq]f[eq]c.stgrcipa[as]w
[ilv].dgdndcedgsde[as].??.??.??.??.??.??.??.??.??.??.??.??.?

1
0 2 10¹⁵¹ c.s[dn].?ef.c[kr]st.rcip.tw.cdgd.dced.sde[as].??.cg..??.??.??.??.??.?
5 2 151 ??.??.??.??.??.ggsc..?[as].?qf.c.st.rcip..wvcdgdndcedssde..??.c...?
2 .??.??.??.??.??.??.??.??.??.?

6
7 1 10¹⁹⁶ cgsd.?efqckstsrincipltwrcdgdsdcddssdea.?.?ncg.?.??.??.??.??.??.??.?
6 196 .??.??.ggscq.?s.?qfqc.stgriciprtwvcdgdndcedssdek.?.?cqp.?.??.??.??.?
.??.??.??.??.?

Fam 2

CXAXQFTCD-NGQCLPQNWVCDGENDCPDXSDEKN--C--
APHTCPSGEFQCRSTNRCIPETWLCGDNDCEDGSDEESCTPPT

CQPGEFTCN-NGNCLPLEWVCDGENDCGDSSDEEN--
CGGSEHTCPSGEFQCRSTNRCIPETWLCGDNDCEDGSDEESCTPPT

CLAGEFRCN-SGRICIPHWRCGDGEDDCLDSSDEKD--
CTTSEPTCPSGEFQCRSTNRCIPETWLCGDNDCEDGSDEESCTPPT

CPSGEFRC-SNGSCIPQEWGCDGXNDCGDDSSDEKN--
CAAAGPTCPSGEFQCRSTNRCIPKTWLCGDNDCEDGSDEESCTPPT

CPSGEFRCQSSNTCIPLNWLCDGEDDCGDDSDSEKN--
CEASVPTCPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CASNQFRCR-NGRCIPLPWVCDGEDDCQDNSDEAS--
CAAPAPTCPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CASNQFRCR-NGRCIPLPWVCDGEDDCQDNSDEAS--
CAAPAPTCPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CASNQFTCN-NGHCLPQHWRCDEDDCGDNSDEAS--CQP--
PTCPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTXTT

CQADEFRCG-NGRCISPTWVCDGEXDCGDDSDSEAN--
CATTERTCPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CPPDEFKCG-NGHCISQTWLCDGEXDCGDNSDEES--CAAP--
TCPSGEFQCRXTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CPPDEFRC-NGRCLPQPWVCDGEDDCGDDSDSEAN--
CATTAPTCPGSEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CVANEFKCG-SGKCIPTWVCDGDNDCEDEGSDEAS--CAQPT--
CPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CGANEFKCS-SGSCIPQEWRCDEGDCGNSDES LAPCKEPT--
CPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CRADEFKCG-NGHCIPGQWLCDGENDCQDGSDEKS--CEQPT--
CPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

CVPGEFRCHDSGTCVPLAXLXGDNDCGNSDEAS--
CESSEPTCPSGEFQCRSTNRCIPETWLCDGDNDCEDEGSDEESCTPPT

分 预
匹 期
数 配 基
的 序

12	15	10 ⁻¹⁷	c...[eq]f.c.?.?.c[ilv]....c.g..dc.d.sde...?.?.?.?.?.?.?.?.[dn][dn]..[de]...[de]e.....
32	82	10 ⁻⁶³	c...[eq]f.c.?.?.c[ilv]....c.g..dc.d.sde...?.?c.?.?.?.?.?.?cp sgefqcr.tnrcip[ekq]twlcdgdndcedgsdeesct..t
31	85	10 ⁻⁶⁶	c...[eq]f.c.?.?.c[ilv]...w.cdg..dc.d.sde...?.?c.?.?.?.?.?.?cp sgefqcr.tnrcip[ekq]twlcdgdndcedgsdeesct..t
30	45	10 ⁻⁶⁹	c...[eq]f.c.?.?.c[ilv]...w.cdg..dc.d.sde...?.?c.?.?.?.?.?.?cp sgefqcr.tnrcip[ekq]twlcdgdndcedgsdeesctppt
28	49	10 ⁻⁷⁰	c...[eq]f.c.?.?.c[ilv]...w.cdge.dc.d.sde...?.?c.?.?.?.?.?.?cp sgefqcr.tnrcipetwlcdgdndcedgsdeesctppt
26	53	10 ⁻⁷²	c...[eq]f.c.?.?.c[ilv]p..w.cdg.[dn]dc.d.sde...?.?c.?.?.?.?.?.? ?cpsgefqcrstnrcip[ekq]twlcdgdndcedgsdeesctppt
24	35	10 ⁻⁷⁴	c...[eq]f.c.?.?.c[ilv]p..w.cdge[dn]dc.d.sde...?.?c.?.?.?.?.?.? ?cpsgefqcrstnrcipetwlcdgdndcedgsdeesctppt

21 95	8	10^{-75}	c...[eq]f.c.?.?.g.c[ilv]p..w.cdge[dn]dc.d.sde...?.?.?.?.?. ?cpsgefqrstnrcipetwlcgdndcedgsdeesctppt
16 87	6	10^{-77}	c...[eq]f.c.?.?ng.c[ilv]p..w[ilv]cdge[dn]dc.d.sde...?.?.?.?.?. ?.?.?cpsgefqrstnrcipetwlcgdndcedgsdeesctppt
14 14	5	10^{-78}	c...[eq]f.c.?.?ng.c[ilv]p..wcdge[dn]dc.d.sde...?.?.?.?.?.?. ?cpsgefqrstnrcipetwlcgdndcedgsdeesctppt
11 66	4	10^{-80}	c..[dn][eq]f[kr]c.?.?ng.c[ilv]p..w[ilv]cdge[dn]dc.d.sde.s.?.?.c. ?.?.?.?.?.?cpsgefqrstnrcipetwlcgdndcedgsdeesctppt

Fam 3

CPSG-EFQCRSTNRCIPETWLCDGE-DDCGDSSDESALCGRPGPATSAPAACP-SGEFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CPSG-EFQCRSTNRCIPETWLCDGD--NDCEDGSDE-----ESCTPPTCP-PGFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CQSFEFECHSTGRICIPASWLCDGD--NDCEDSSDEE-----GCEAAAPTCP-SGFRCRXTXRCIPXTWLCDGDNDCEDGSXEESCTPPT

CRAN-EFQCHSTGRICIPASWLCDGD--NDCEDGSDE-----SQLCTAHTCP-SGFQCRSTNRCIPETWLCDGDNDCEDGSDEES-CTPPE

CXPG-EFQCNNGR-CIPATWLCDGD--DDCGDNSDET-----GCT--EHTCP-SGFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CXPG-EFQCNNGR-CIPATWLCDGD--DDCGDNSDET-----GCT--EHTCP-SGFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CQSN-EFQCNNGR-CISVTWLCDGD--DDCGDSSDET-----DCTSAVPTCP-SGFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CPSS-EFQCRNNTKTCIPRNWLCDGE--DDCGDSSDET-----DCT--THTCP-SGFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CLPS-EFPC-SNGRCVPRPWCDGD--DDCEDNSDEA-----GCP--KPTCP-SGFQCRSTNRCIPXTWLCDGDNDCEDGSDEESCTPPT

CPPS-EFPC-GNGSCVPQAWVCDGD--PDCPDNSDEE-----GCTGTGPTCP-SGFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CLPN-QFQCQSSGRICIPLNWLCDGD--DDCGDSSDET-----SCK--APTCP-SGFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CQAD-EFQCRNTEKCLPLNWLCDGD--NDCGDDSDDET-----SCA--TPTCP-SGFQCRSTNRCIPETWLCDGDNDCEDGSDEESCTPPT

CQPD-EFRCRNTDICIPQRWVCDGD--NDCEDSSDEADCQQPTCR-ANEFQCHSTGRICIPETWLCDGDNDCEDGSDEES-----CTPPT-

CQSDQFRCSN-GRCIPVEWVCDGEDDCLDGSDEP-
 QVCQTAPTCAAEFQCNSTGRICIPVSWVCDGVNDCEDSSDEAG--CATSGPT

CQAEDEFKCGN-GRCLPEAWVCDGEDDCGDNSDE----
 ADCQAPTCAAEFQCNSTGRICIPVSWVCDGXNDCEDSSDEAG--CATSGPT

CPPDEFPCSNNGICIPRSWRCDGEDDCGDNSDEE-D-
 CTSAGHTCAPSEFTCNSTGRICIPQEWVCDGDNDCEDSSDEAPDLCASAAPT

CQPGEFRCRN-GKCIPTWLXXGXDDCGDNSDE--
 ADCATTAPTCPPDEFCTCRSTERCIPLAWVCDGDNDCEDSSDEAG--CTTPEPT

CLSGEFRCNSN-GNCLPADWLCDGEDDCGDNSDE--
 TSCAASEPTCPPDEFCTCRSTERCIPLAWVCDGDNDCEDSSDEAG--CTTPEPT

CGSSEFQCHSTGRICIPENWVCDGDDDCEDSSDE--
 KSCTSAAPTCPPDEFCTCRSTERCIPLAWVCDGDNDCEDSSDEAG--CTTPEPT

CAADQFKCDN-GRCVPQNWRCDGEXDCGDNSDE--ENCTT--
 PTCPPDEFCTCRSTERCIPLAWVCDGDNDCEDSSDEAG--CTTPEPT

分 匹 数 配	预 期 基序 的
18 79	10^{-55} c...[eq]f.c..?.?g.c[ilv]p..w...g..dc.d.sde.?.?.?.?.?.?.?.tc...e f.c.st.rcip..w[ilv]cdg.ndcedssdea..?.?c[ast][st]..pt
17 76	10^{-59} c...[eq]f.c..?.?g.c[ilv]p..w.cdg[de].dc.d.sde.?.?.?.?.?.?.?.tc. ..ef.c.st.rcip..w[ilv]cdg.ndcedssdea..?.?c[ast][st]..pt
16 28	10^{-63} c...[eq]f[kqr]c..?.?g.c[ilv]p..w.cdg[de].dc.d.sde.?.?.?.?.?.?.?. .tc...ef.c.st.rcip.[as]w[ilv]cdg.ndcedssdeag.?.?c[ast]t..pt
14 51	10^{-65} c...[eq]f[kqr]c..?.?g.c[ilv]p..w...g..dc.d.sde.?.?.?.?.?.?c.[ast].? ?.tc...ef.c.st.rcip.[as]w[ilv]cdgdndcedssdeag.?.?cttpept
12 76	10^{-69} c...[eq]f[kqr]c..?.?g.c[ilv]p..w.cdg[de].dc.d.sde.?.?.?.?.?.?c.[ast] .?.?.tc...ef.c.st.rcip.[as]w[ilv]cdgdndcedssdeag.?.?cttpept
11 20	10^{-77} c...[eq]f[kqr]c..?.?g.c[ilv]p..w...g..dc.d.sde.?.?.?.?.?.?c[ast][ast] .?.?.?ptcppdefctcrsterciplawvcdgdndcedssdeag.?.?cttpept
89 8	10^{-83} c.[as].[eq]f[kqr]c..?.?g.c[ilv]p.[dn]w.cdg[de].dc.d.sde.?.?.?.?.?.?c [ast][ast].?.?ptcppdefctcrsterciplawvcdgdndcedssdeag.?.?cttpept
62 2	10^{-86} c.[as].[eq]f[kqr]c..?.?g.c[iv]p[eq]nw.cdg[de].dc.d.sde.?.?.?.?.?.?c [st].?.?ptcppdefctcrsterciplawvcdgdndcedssdeag.?.?cttpept

实施例 11





通过实施例 7 所述的方法，基本上鉴别了结合人 IgG 和/或来自其它物种的 IgG、包括猕猴 IgG(共同称作 IgG)的单体或多聚体。

鉴别了下面的 IgG-结合单体。在每个二聚体家族后面的表表示基

于家族成员的比对的共有基序。




Fam1

```
CASGQFQCRSTSICVPMWWRCDGVPDCPDNSDEK--SCEPP----T-----
CASGQFQCRSTSICVPMWWRCDGVPDCVDNSDET--SCTST----VHT-----
CASGQFQCRSTSICVPMWWRCDGVPDCADGSDEK--DCQOH----T-----
CASGQFQCRSTSICVPMWWRCDGVNDCGDSDEA--DCGRPGPGATSAPAA--
CASGQFQCRSTSICVPMWWRCDGVPDCLDSSDEK--SCNAP----ASEPPGSL
CASGQFQCRSTSICVPMWWRCDGVPDCRDGSDEAPAHCSAP----ASEPPGSL
CASGQFQCRSTSICVPMWWRCDGVPDCRDGSDEP-EQCTPP----T-----
CLSSQFRCRDTGICVPQWWVCDGVPDCGDSDEK--CGRT----GHT-----
CLSSQFRCRDTGICVPQWWVCDGVPDCRDGSDEAAV-CGRP----GHT-----
CLSSQFRCRDTGICVPQWWVCDGVPDCRDGSDEAPAHCSAP----ASEPPGSL
```

分数	匹配的	预期的	基序
117	5	10 ⁻²⁸	 c.s.qf[kqr]cr.t.icvp.ww.cdgvpdc.d.sde...??.?c....??.?..?..?..? ?.?.?.?.?.?
109	6	10 ⁻²⁹	 c.s.qf[kqr]cr.t.icvp.ww.cdgvpdc.d.sde...??.?c....??.?..?..?..? ?.?.?.?.?.?
942	7	10 ⁻³³	 casgqfqrstsicvp.ww.cdgvpdc.d.sde...??.?c....??.?..?..?..?..? ?.?.?.?.?
866	6	10 ⁻³⁶	 casgqfqrstsicvpmwvrddgvpdc.d.sde...??.?c....??.?..?..?..?..? ?.?.?.?.?

Fam2

```
CGAS-EFTCRSSSRCIPQAWVCDGENDCRDNDNSDE--ADCSAPASEPPGSL
CRSN-EFTCRSSSERCIPLAWVCDGDNDCRDSDSDE--ANCSAPASEPPGSL
CVSN-EFQCRGTRRCIPRTWLC DGLPDCGDNSEAPANCSAPASEPPGSL
CHPTGQFRCRSSGRCVSPPTWVCDGDNDGDNSE--ENCSAPASEPPGSL
CQAC-EFQC-GNGRCISPAWVCDGENDCRDGSDE--ANCSAPASEPPGSL
```




分数	匹配的	预期的	基序
683	5	10 ⁻³⁴	 c....?[eq]f.c?...rc[iv]..[ast]w[ilv]cdg..dc.d.sde...?.?.[dn]c sapaseppgsl
584	4	10 ⁻³⁶	 c....?[eq]f.c?...rc[iv]..[ast]wvcdg[de]ndc.d.sde...?.?.[dn]cs apaseppgsl
480	3	10 ⁻⁴¹	 c.[as]...?ef.c?...rci...awvcdg[de]ndcrd.sde...?a[dn]csapasepp gsl

总结家族 2 IgG 结合 A 域单体的基序如下：

[eq]fxcrx[st]xrc[iv]xxxw[ilv]cdgxxdcxd[dn]sde

Fam3

CPFSQFTCKSNDKCIPVHWLDCGDNDCGDSSDE--ANCGRPGPGATSAPAA
 CPSGEFFPCRSSGRCIPLAWLDCGDNDCRDNDSEFPALCGRPGPGATSAPAA
 CAPSEFQCRSSGRCIPLPWVCDGEDDCRDGSDES-AVCGAPAP--T-----
 CQASEFTCKSSGRICIPQEWLDCGEDDCRDSSDE--KNCQQPT-----
 CLSSEFQCQSSGRCIPLAWVCDGDNDCRDSSDE--KSCKPRT-----

分数	匹配	预期的	基序
526	5	10 ⁻²⁴	 c...[eq]f.c[kqr]s..[kr]cip..w[ilv]cdg[de][dn]dc.d.sde.?.?.c.....? ?.?.?.?.?.?.?.?
476	4	10 ⁻²⁸	 c...ef.c[kqr]ssgrcip..w[ilv]cdg[de][dn]dcrd.sde.?.?.c.....?.?.?.? ?.?.?.?.?
375	3	10 ⁻³⁰	 c..sef.c[kqr]ssgrcip..w[ilv]cdg[de][dn]dcrd.sde.?.?.c...[ast].??. ?.?.?.?.?.?.?





总结家族 3 IgG 结合 A 域单体的 2 个基序如下:

CXSSGRICIPXXWVCDGXXDCRDXSDE
 CXSSGRICIPXXWVCDGXXDCRDXSDE

基于家族 3 比对, 本发明提供了包含非天然发生的单体域的多肽, 该单体域结合 IgG, 且具有紧挨在 A 域支架的第三个半胱氨酸之前的序列 SSGR。

Fam4

CPANEFQCSNGRRCISPAWLCDGENDCVDGSDE--KGCTPRT
 CPPSEFQCGNGRRCISPAWLCDGDND CVDGSDE--TNCTTSGPT
 CPPGEFQCGNGRRCISAGWVCDGENDCVDDSDSDE--KDCPART
 CGSGEFQCSNGRRCISLGWVCDGEDDCPDGSDE--TNCGDHILPFSTPGPST
 CPADEFTCGNGRRCISPAWVCDGEPDCRDGSDE-AAVCETHT
 CPSNEFTCGNGRRCISLAWLCDGEPDCRDSSDES LAICSQDPEFHKV

分数	匹配	预期的	基序
630	6	10 ⁻²⁴	 c...ef.c.ngrcis..w[ilv]cdg[de].dc.d.sde.?.?.c..
546	5	10 ⁻²⁵	 cp..ef.c.ngrcis..w[ilv]cdg[de].dc.d.sde.?.?.c..
452	4	10 ⁻²⁷	 cp..ef.c.gngrcis..w[ilv]cdg[de].dc.d.sde.?.?.c..
367	3	10 ⁻²⁹	 cp..efqc.ngrcis..w[ilv]cdg[de]ndc vd.sde.?.?.c..

实施例 10

本实施例解释了 IgG-结合单体对来自各个动物物种的 IgG 的结

合亲和力。

域	物种的IgG亲和力 (nM)			
	人	猕猴	小鼠	大鼠
IgM02	10.7	8.7	23.8	47.7
Ig156	95.5	586	1326	1513
RM09	3900	8900	5300	5750
RM15	812	5630	6300	>10000
502	336	1850	>10000	>10000

表：物种的 IgG-结合域的亲和力

将来自标示物种的 0.2 μg 完整 IgG 级分固定化在 96-孔 Maxisorp 平板 (Nunc) 的一式两份孔中，并用 1% BSA 封闭。然后，加入纯化的域的系列稀释液，并使用标准的 ELISA 方法，通过 HRP-缀合的、高亲和力抗-HA 第二抗体，定量结合的蛋白的量。使用非线性最佳拟合算法，将数据拟合 1: 1 结合模型，以确定 K_D (亲和力)。

实施例 11

本实施例描述了用于解释 IgG-结合 Avimer 域 Ig-M02 的存在赋予多聚体的药物代谢动力学半衰期的实验。

Avimer 构建体 C242 是 Avimer 域的三聚体(约 15 kDa)。N-末端 Avimer 域是 Ig-M02。给三只猕猴注射单次 1 mg/kg 剂量的已经用 ^{125}I 示踪标记的 Avimer C242。猴子 1 和 2 接受静脉内给药；猴子 3 接受肌肉内给药。在图 10 标示的时间，最多至 288 小时，得到血清样品，并评估 ^{125}I cpm。

在该实验中观察到的末端血清半衰期是约 53 小时，其比速增长地按比例放大到预测的在人中的约 106 小时半衰期。在类似的小鼠实验中，观察到 7-9 小时末端半衰期，与啮齿动物的更小的大小相一致。而且，因为肌肉注射的和静脉内注射的动物的绘图会聚，从注射后约 12 小时开始变得几乎相同，所以我们推断单体在体内表现出高生物利用率。

53 小时比为该大小的蛋白所预期的明显更长。例如，20 kDa 细胞因子白细胞介素-6 在绒猴中的血清半衰期是 4-6 小时 (Ryffel, B. 等 Blood 83, 2093-102 (1994))。因而，IgG-结合 Ig-M02 域赋予足够长的

半衰期，以允许在人受试者中每周至少给药一次。

尽管为了清楚和理解的目的，已经比较详细地描述了前面的发明，但本领域的技术人员在阅读本公开内容后会明白，可以在不脱离本发明的真实范围的情况下在形式和细节上进行各种变化。例如，可以以各种组合，使用上述的所有技术、方法、组合物、装置和系统。在本申请中引用的所有出版物、专利、专利申请或其它文件，都为所有目的整体引作参考，其程度与单独地指明将每篇单个的出版物、专利、专利申请或其它文件为所有目的整体引作参考相同。

序列表

SEQ ID NO:1 - c-MET 氨基酸序列

```

1 mkapavlapg ilvllftlvq rsngeckeal aksemnvnmk yqlpnftaet piqnvilheh
61 hiflgatnyi yvlneedlqk vaeyktgpvl ehpdcfpcqd csskanlsgg vwkdinmal
121 vvdtyyddql iscgsvnrgt cqrhvfphnh tadiqsevhc ifspqieeps qcpdcvvsal
181 gakvlssvkd rfinffvgnt inssyfpdhp lhsisvrrlk etkdgfmflt dqsyidvlpe
241 frdsypikyv hafesnnfiy fltvqretld aqtfhtriir fcsinsglhs ymemplecil
301 tekrrkrstk kevfnilqaa yvskpgaqla rqigaslndd ilfgvfaqsk pdsaepmdrs
361 amcafpikyv ndffnkivnk nnvrclqhfy gpnhehcfnr tllrnssgce arrdeyrtef
421 ttalqrvdlf mgqfsevlit sistfikgdl tianlgtseg rfmqvvsrs gpstphvnfl
481 ldshpvspev ivehtlnqng ytlvitgkki tkiplnglgc rhfqscsqcl sappfvqcgw
541 chdkcvrsee clsgtwtqqi clpaiykvfp nsapleggtr lticgwdfgf rrnnkfdlkk
601 trvllgnesc tlstsestmn tlkctvgpam nkhnfnmsiii snghgttqys tfsyvdpvit
661 sispykypma ggtiltltgn ylnsgnsrhi siggktctlk svnsilecy tpaqtistef
721 avklkidlan retsifsyre dpivyeihpt ksfiggsti tgvgnlnsv svprmvinvh
781 eagrnftvac qhrsenseiic cttpslqqln lqlplktkaf fmlgdilsky fdliyvhnpv
841 fkpfekevpmi smgnenvlei kgndidpeav kgevlkvgnk scenihlhse avlctvpndl
901 lklnselnie wkqaisstvl gkvivqpdqn ftgliagvvs istalllllg fflwlkkrkq
961 ikdlgselvr ydarvhtphl drlvsarsvs pttemvsnes vdyratfped qfpnssqngs
1021 crqvqypltd mspiltsgds disspllqnt vhidlsalnp elvqavqhv igpsslivhf
1081 nevigrgghfg cvyhgltldn dgkkihcavk slnritdige vsqfltegi mkdfshpnl
1141 sllgiclrse gsplvvlpym khgdlnfir nethnptvkd ligfglvak gmkyaskkf
1201 vhrdlaarnc mldekftvkv adfglardmy dkeyysvhnk tgaklpvkw aleslqtqkf
1261 ttksdvwsfg vllwelmrtrg appypdvntf ditvyllqgr rllqpeycpd plyevmlkcw
1321 hpkaemrpsf selvsrisai fstfigehyv hvnatyvvnk cvapypls sednaddevd
1381 trpasfwets

```

LDL-受体A类域

LRP1_人 C.EPYQFRCKNNR.....CVPGRWQ.CDYDNDCGDNSDEES.....C
 LRP1_人 C.LPSQFKCTNTR.....CIPGIFR.CNGQDNCGDGEDERD.....C
 LDLR_人 C.SQDEFRCHDGK.....CISRQFV.CDSRDCLDGSDEAS.....C
 LRP2_人 C.SSSAFTCGHGE.....CIPAHRW.CDKRNDKVDGSDEHN.....C
 LRP2_人 C.SSSEFQCASGR.....CIPQHWY.CDQETDCFDASDEPAS.....C
 CORI_人 CHSQGLVECRNGQ.....CIPSTFQ.CDGDEDCKDGSDEEN.....C
 MAT_人 C.PAQTFRCSNGK.....CLSKSQQ.CNGKDDCGDGSDEAS.....C
 CO8B_人 C...EGFVCAQTGR.....CVNRRLL.CNGDNDCGDQSDAN.....C
 MAT_人 C.TKHTYRCLNGL.....CLSKGNPECDGKEDCSDGSDEKD.....C
 LDVR_人 CLGPGKFKCRSGE.....CIDISKV.CNQEQDCRDWSDEPLKE..C
 APOER2_人 C.PAEKLSCGPTSHK...CVPASWR.CDGEKDCGGADEAG.....C
 SORL_人 CTHFMDFVCKNRQQ.....CLFHSMV.CDGIQCRDGSDEDAAFAGC
 ST7_人 C.AYNQFQCLSRFTKVYTCLPESLK.CDGNIDCLDLGDEID.....C
 共有序列 C.1234F6C12G4.....CI23456.CDG34DC1D3SDE78.....C

图 1

被天然发生的LDL-受体家族识别的配体

配体	受体家族
蛋白酶	参与脂蛋白代谢的蛋白
因子 IXa	apoB100
前-uPA	apoE
t-PA	apoJ(簇连蛋白)
纤溶酶原	apoH(β_2 -糖蛋白 I)
MMP-9	Lp(a)
抑制剂/	磷脂酶
α_2 -巨球蛋白	脂蛋白脂酶
PAI-1	IDL
TFPI	VLDL
胰腺胰蛋白酶抑制剂	β -VLDL
复合物	非人的
蛋白酶/	假单胞菌外毒素 A
α_1 -抗胰蛋白酶	环孢菌素 A
C-蛋白抑制剂	天花粉蛋白
蛋白酶连接蛋白-1	蓖麻毒素蛋白 A
抗凝血酶	皂草素
CI-抑制剂	抗毒素
凝血酶/肝素辅因子 II	庆大霉素
组织蛋白酶 G/ α_1 -抗胰凝乳蛋白酶	多粘菌素 B
维生素-载体复合物	病毒
维生素 D-bp, 维生素 D	HRV2(Rhino)
视黄醇-bp, 维生素 A	HCV(Flavi)
钴胺传递蛋白, 维生素 B12	BVDV(Flavi)
各种的	参与脂蛋白代谢的蛋白
清蛋白	apoB100
运甲状腺素蛋白	apoE
β -淀粉状蛋白前体蛋白	apoJ(簇连蛋白)
RAP	apoH(β_2 -糖蛋白 I)
补体 C3	Lp(a)
乳铁蛋白	磷脂酶
甲状腺球蛋白	脂蛋白脂酶
血小板反应蛋白	IDL
saposin前体	VLDL
reelin	β -VLDL
胰岛素	非人的
甲状旁腺激素 (PTH)	假单胞菌外毒素 A
抑蛋白酶肽	环孢菌素 A
α -淀粉酶	天花粉蛋白
CIq	蓖麻毒素蛋白 A
α_1 -微球蛋白	皂草素
β_2 -微球蛋白	抗毒素
添味剂-结合蛋白	庆大霉素
表皮生长因子	多粘菌素 B
催乳素	病毒
溶菌酶	HRV2(Rhino)
结缔组织生长因子 (CTGF)	HCV(Flavi)
细胞色素 C	BVDV(Flavi)
精囊分泌蛋白 II	病
克拉拉细胞分泌蛋白 (CCSP)	病毒
cubulin	HRV2(Rhino)
因子 VIII	HCV(Flavi)
	BVDV(Flavi)

图 3

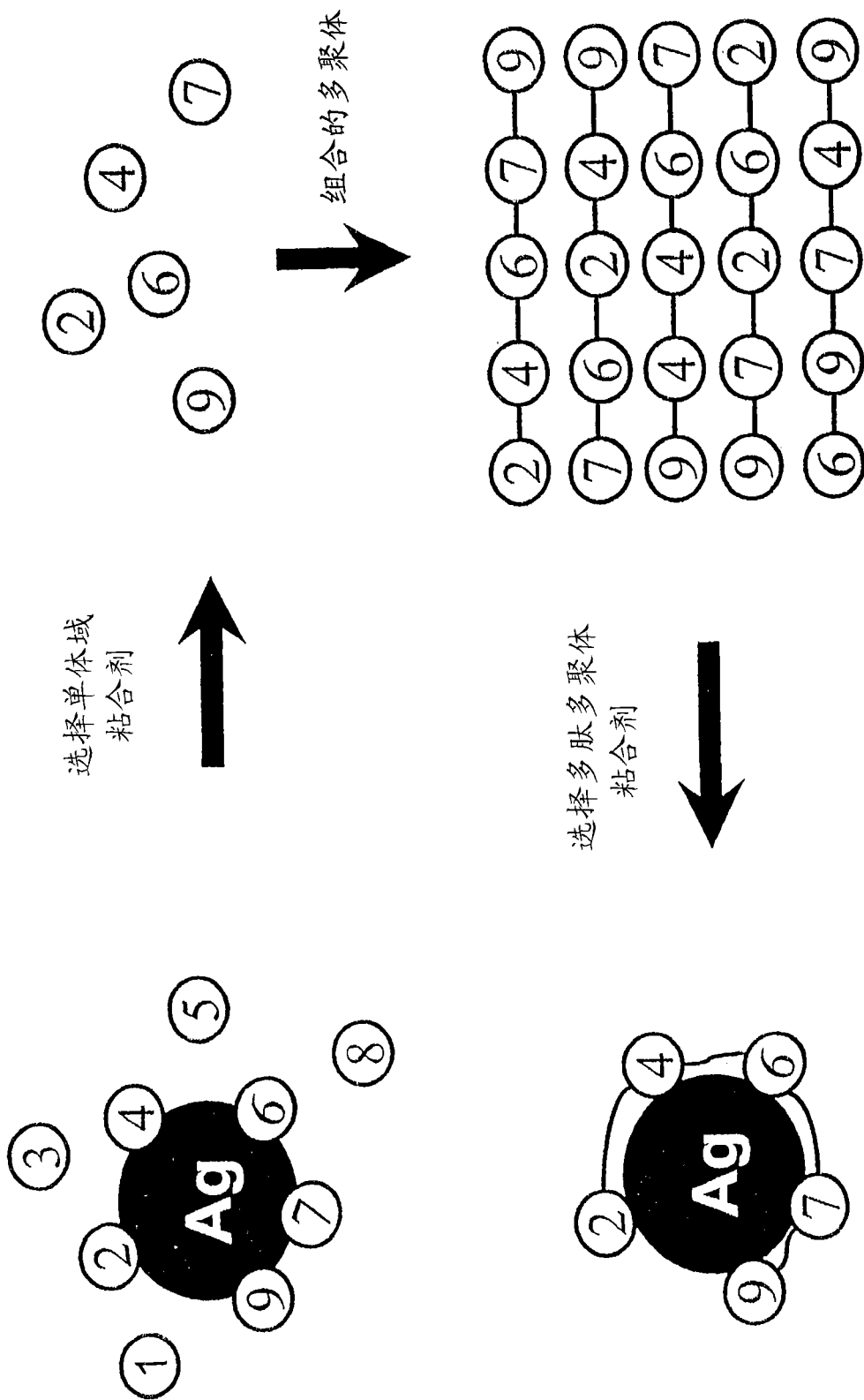


图 4

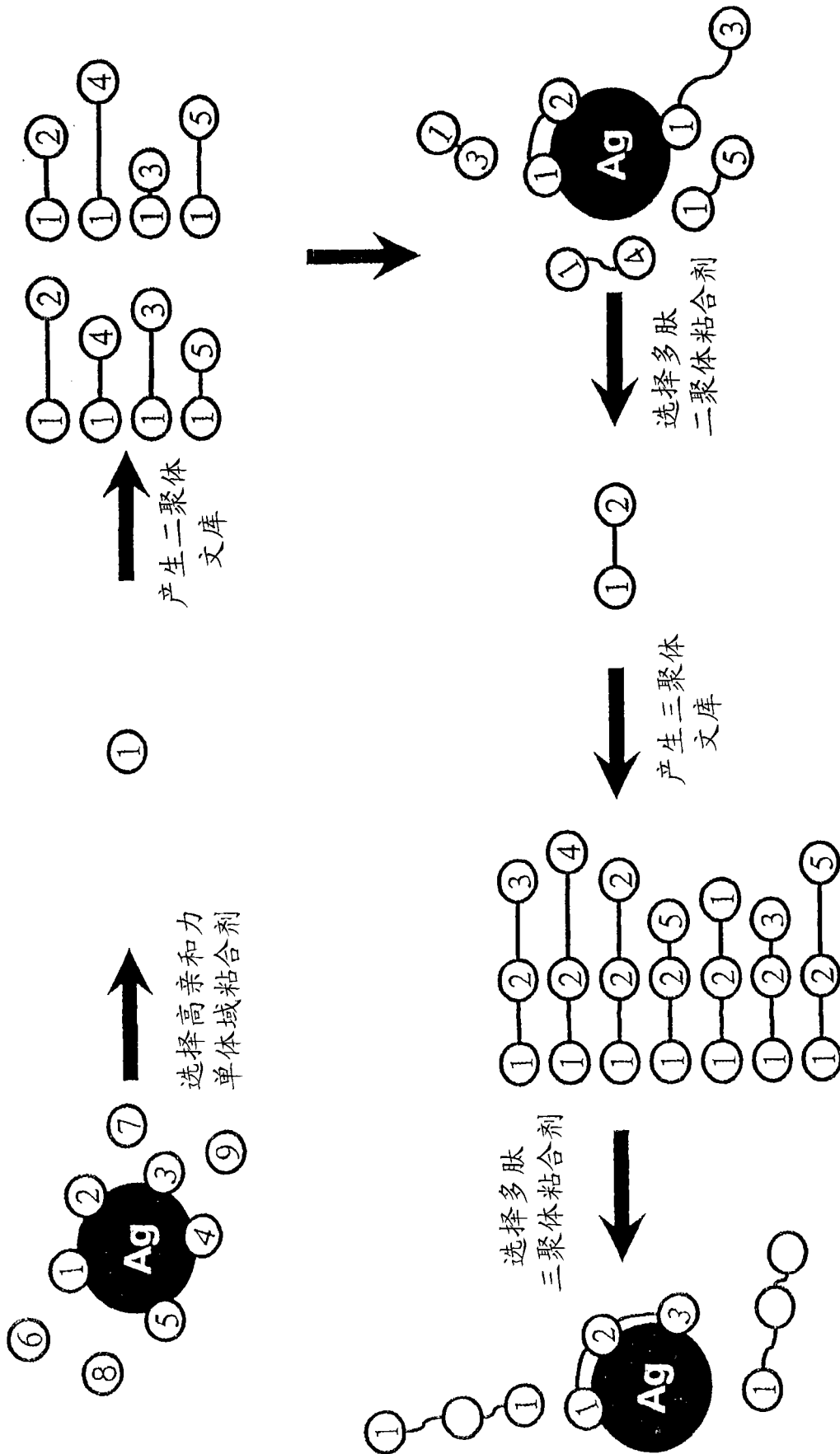


图 5

a	b c	de	f ghi	jk lm	nop	q
IDD_人NPGQFACRSGTIQ.....	CIPLPWQ.CD	GWATCEDE.....	SDEAN.....C	
LRP3_人QADEFRCDNGK.....	CLPGPWQ.CNTV	DECGDG.....	SDEGN.....C	
LRP3_人PGGTFPCSGARSTR.....	CLPVERR.CDGLQD	CGDG.....	SDEAG.....C	
LRP3_人LPWEQPCGSSSDSGSLGDQGC	FCFSEPQR.CDGWWHC	CASG.....	RDEQG.....C	
LRP3_人PPDQYPCGGGSL.....	CYTPADR.CNNQKSC	CPDG.....	ADEKN.....C	
LRP3_人QPGTFHCGTNL.....	CIFETWR.CDGEDEC	QDG.....	SDEHG.....C	
LRP5_人SPDQFACATGEID.....	CIPGAWR.CDGFPEC	DDQ.....	SDEEG.....C	
LRP5_人SAAQFPCARGQ.....	CVDLRLR.CDGEADC	QDR.....	SDEVD.....C	
LRP5_人LPNQFRCASGQ.....	CVLIKQO.CDSFPD	CIDG.....	SDELM.....C	
LRP6_人SPQQFTCFTEID.....	CIPVAWR.CDGFTE	CEDH.....	SDELN.....C	
LRP6_人SESQFQCASGQ.....	CIDGALR.CNGDANC	QDK.....	SDEKN.....C	
LRP6_人LIDQFRCANGQ.....	CIGKHKK.CDHNVD	CSDK.....	SDELD.....C	
ST7_人ACDQFRCGNGK.....	CIPEAWK.CNNMDE	CGDS.....	SDEEI.....C	
ST7_人AYNQFQCLSRFTKVT.....	CLPESLK.CDGNID	CLDL.....	GDEID.....C	
ST7_人LPWEIPCNGWNG.....	CYTEQOR.CDGYWH	CPNG.....	RDETN.....C	
ST7_人QKEEFPCSRNGV.....	CYPRSDR.CNYQNH	CPNG.....	SDEKN.....C	
ST7_人QPGNFHCKNNR.....	CVFESWV.CDSQDD	CGDG.....	SDEEN.....C	
CORI_人GRGENFLCASGI.....	CIPGKLQ.CNGYND	CDDW.....	SDEAH.....C	
CORI_人SENLFHCHTGK.....	CLNYSLV.CDGYDD	CGDL.....	SDEQN.....C	
CORI_人NPTTEHRCGDGR.....	CIAMEWV.CDGDHD	CVDK.....	SDEVN.....C	
CORI_人HSQGLVECRNGQ.....	CIPSTFQ.CDGEDC	CKDG.....	SDEEN.....C	
CORI_人SPSHFKCRSGQ.....	CVLASRR.CDQAD	CDDD.....	SDEEN.....C	
CORI_人KERDLWECPSNKQ.....	CLKHTVI.CDGFPC	PDY.....	MDEKN.....C	
CORI_人QDDELECANHA.....	CVSRDLW.CDGEAD	CSDS.....	SDEWD.....C	
TMS2_人SNSGIECDSSGT.....	CINPSNW.CDGVSH	CPGG.....	EDENR.....C	
TMS3_人SGKYRCRSSFK.....	CIELIAR.CDGVSD	CKDG.....	EDEYR.....C	
MAT_人PGQFTCRTGR.....	CIRKELR.CDGWAD	CTDH.....	SDELN.....C	
MAT_人DAGHQFTCKNKF.....	CKPLFWV.CDSVND	CGDN.....	SDEQG.....C	
MAT_人PAQTFRCSNGK.....	CLSKSQO.CNGKDD	CGDG.....	SDEAS.....C	

图 6A

MAT_人	C.....TKHTYRCLNGL.....CLSKGNPECDGKEDCSDG.....SDEKD.....C
ENTK_人	C.....LPGSSPCTDALT.....CIKADLF.CDGEVNCPPDG.....SDEDNKM.....C
ENTK_人	C.....KADHFQCKNGE.....CVPLVNL.CDGHHLHCEDG.....SDEAD.....C
HAIL_人	C.....QPTQFRCSNGC.....CIDSFLE.CDDTPNCPDA.....SDEAA.....C
CFAI_人	C.YTQKADSPMDDFFQCXNGK.....YISQMK.A.CDGINDCGDQ.....SDEL.....C
CFAI_人	C.....QGKGFHCKSGV.....CIPSQYQ.CNGEVDCITG.....EDEVG.....C
CO6_人	C.....KNKFRCDSGR.....CIARKLE.CNGENDCGDN.....SBERD.....C
CO7_人	C.....GERFRFCFSQ.....CISKSLV.CNGSDSCDEDS.....ADEDR.....C
CO8A_人	C.....GQDFQCKETGR.....CLKRHLV.CNGDQDCLDG.....SDEDD.....C
CO8B_人	C.....EGFVCAQTGR.....CVNRRLL.CNGDNDCGDQ.....SDEAN.....C
CO9_人	C.....GNDFQCSTGR.....CIKMRLR.CNGDNDCGDF.....SDEDD.....C
PERL_人	C.....TEAEFACHSYNE.....CVALEYR.CDRRPPCDRM.....SDELN.....C
PERL_人	C.....GPQEAACRNHG.....CIPRDYL.CDGQEDCEDG.....SDELD.....C
PERL_人	C.....EPNEFFPCGNHG.....CALKLWR.CDGFDFCEDR.....TDEAN.....C
PERL_人	C.....GPTQFRCVSTNM.....CIPASFH.CDEESDCPDR.....SDEFG.....C
SORL_人	C.....LRNQYRCSNGN.....CINSIWW.CDFDNDCGDM.....SDERN.....C
SORL_人	C.....DLDTQFRCQESGT.....CIPLSYK.CDLEDDCGDN.....SDESH.....C
SORL_人	C.....RSDEYNCSSGM.....CIRSSWV.CDGDNDCRDW.....SDEAN.....C
SORL_人	C.....EASNFQCRNGH.....CIPQWA.CDGDTDCCDG.....SDEDPVN.....C
SORL_人	C.....NGFRCPNGT.....CIPSSKH.CDGLRDCSDG.....SDEQH.....C
SORL_人	C.....THPMDFVCKNRQQ.....CLFHSMV.CDGIQCRDG.....SDEDAAFAG...C
SORL_人	C.....DEFGQCQNGV.....CISLIWK.CDGMDDCCGDY.....SDEAN.....C
SORL_人	C.....SRYFQFRCENGH.....CIPNRWK.CDRENDCCGDW.....SDEKD.....C
SORL_人	C.....LPNYRCSSGT.....CVMDTWV.CDGYRDCADG.....SDEEA.....C
SORL_人	C.....DRFEFECHQPKT.....CIPNWKR.CDGHQDCQDG.....RDEAN.....C
SORL_人	C.....MSREFQCEGGEA.....CIVLSEK.CDGFLLDCSDE.....SDEKA.....C
AP0ER2_人	C.....EKDQFQCRNER.....CIPSVWR.CDEDDDCLDH.....SDEDD.....C
AP0ER2_人	C.....ADSDFTCDNGH.....CIHERWK.CDGEEECPDG.....SDESEAT...C
AP0ER2_人	C.....PAEKLSCGPTSHK.....CVPASWR.CDGEKDCEGG.....ADEAG.....C
AP0ER2_人	C.....APHEFQCGNRS.....CLAAVFV.CDGDDDCCGDG.....SBERG.....C
AP0ER2_人	C.....GPREFRCCGGGGA.....CIPERWV.CDRQFDCEDR.....SDEAAEL...C
AP0ER2_人	C.....ATVSQFACRSGE.....CVHLGWR.CDGDRCCKDK.....SDEAD.....C
AP0ER2_人	C.....RGDEFQCGDGT.....CVLAIKH.CNQEQDCPDG.....SDEAG.....C

图 6B

LDLR_人	C.....ERNEFQCQDGK.....CISYKMW.CDGSAEQDGG.....SDESQET.....C
LDLR_人	C.....KSGDFSCGGRVNR.....CIPQFWR.CDGQVDCDNG.....SDEQG.....C
LDLR_人	C.....SQDEFRCHDGK.....CISRQFV.CDSDRDCLDG.....SDEAS.....C
LDLR_人	C.....GPASFQCNSSST.....CIPQLWA.CDNPDPCEDG.....SDEWFOR.....C
LDLR_人	C.....SAFEHCLSGE.....CIHSSWR.CDGGPCKDK.....SDEEN.....C
LDLR_人	C.....RPDEFQCSJDN.....CIHGSRO.CDREYDCKDM.....SDEVG.....C
LDLR_人	C.....EGPNKFKCHSGE.....CITLDKV.CNMARDCRDW.....SDEPIKE.....C
LDVR_人	C.....EPSQFOCTNGR.....CITLLWK.CDGEDDCVDG.....SDEKN.....C
LDVR_人	C.....AESDFVCNNGQ.....CVPSRWK.CDGDPCEDG.....SDESPEQ.....C
LDVR_人	C.....RIHEISGAHSTQ.....CIPVSWR.CDGENDCDSG.....EDEEN.....C
LDVR_人	C.....SPDEFTCSSGR.....CISRNFV.CNGQDDCSDG.....SDELD.....C
LDVR_人	C.....GAHEFQCSTSS.....CIPISWV.CDDADDCSDQ.....SDESLEQ.....C
LDVR_人	C.....PASEIQCSGE.....CIHKKWR.CDGDPPCKDG.....SDEVN.....C
LDVR_人	C.....RPDQFECEGDS.....CIHGSRO.CNGIRDVCVG.....SDEVN.....C
LDVR_人	C.....LPGKFKCRSGE.....CIDISKV.CNQEQDCRDW.....SDEPLKE.....C
LRP1_人	C.....SPKQFACRDQIT.....CISKGWR.CDGERDCPDG.....SDEAPEI.....C
LRP1_人	C.....QPNEHNCIGTEL.....CVPMURL.CNGVQDCMDG.....SDEGPH.....C
LRP1_人	C.....QPGEFACANSR.....CIQERWK.CDGDNDCLDN.....SDEAPAL.....C
LRP1_人	C.....PSDRFKCENNR.....CIPNRWL.CDGDNDCCGNS.....EDSNAT.....C
LRP1_人	C.....PPNQFSCASGR.....CIPISWT.CDLDDDCGDR.....SDESAS.....C
LRP1_人	C.....FPLTQFTCNNGR.....CININWR.CDNDNDCCGN.....SDEAG.....C
LRP1_人	C.....SSTQFKCNSGR.....CIPRHWT.CDGDNDCCGY.....SDETHAN.....C
LRP1_人	C.....HTDEFQCRIDGL.....CIPLRWR.CDGDIDCMDG.....SDEKS.....C
LRP1_人	C.....DPSVKFGCKDSAR.....CISKAWV.CDGDNDCCDN.....SDEEN.....C
LRP1_人	C.....RPPSHPCANNTSV.....CLPPDKL.CDGNDDCCGDG.....SDEGEL.....C
LRP1_人	C.....RAQDEFECANGE.....CINFSLT.CDGVPHCKDK.....SDEKPSY.....C
LRP1_人	C.....KKTFRQCSNGR.....CVSNMLW.CNGADDCCGDG.....SDEIP.....C
LRP1_人	C.....GVGEFRCRDGT.....CIGNSSR.CNQFVDCEDA.....SDEMN.....C
LRP1_人	CSSYFRLGVKGVLPQCERTSL.....CYAPSWV.CDGANDCGDY.....SDERD.....C
LRP1_人	C.....PLNYFACPSGR.....CIPMSWT.CDKEDDCEHG.....EETH.....C
LRP1_人	C.....SEAQFECQNR.....CISKQWL.CDGSDDCCGDG.....SDEAAH.....C
LRP1_人	C.....GPSFSFCPGTHV.....CVPERWL.CDGDKDCADG.....ADESIAAG.....C
LRP1_人	C.....DDREFMCQNRQ.....CIPKHFV.CDHDRDCADG.....SDESPE.....C

图 6C

LRP1_人	C.....GPSEFRCANGR.....CLSSRQWECGDGENDCHDQ.....SDEAPKNPH..C
LRP1_人	C.....NASSQFLCSSGR.....CVAEALL.CNGQDDCCGDS.....SDERG.....C
LRP1_人	C.....TASQFVCKNDK.....CIPFWWK.CDTEDDCCGDH.....SDEPPD.....C
LRP1_人	C.....RPGQFQCSTGI.....CTNPAFI.CDGDNDCCQDN.....SDEAN.....C
LRP1_人	C.....LPSQFKCTNTR.....CIPGIFR.CNGQDNCGDG.....EDERD.....C
LRP1_人	C.....APNQFQCSITKR.....CIPRVWV.CDRDNDCCVDG.....SDEPAN.....C
LRP1_人	C.....GVDEFRCCKDSGR.....CIPARWK.CDGEDDCCGG.....SDEPKEE.....C
LRP1_人	C.....EPYQFRCKNNR.....CVPGRWQ.CDYDNDCCGDN.....SDEES.....C
LRP1_人	C.....SESEFSCANGR.....CIAGRWK.CDGDHDCADG.....SDEKD.....C
LRP1_人	C.....DMDQFQCKSGH.....CIPLRWR.CDADADCCMG.....SDEEA.....C
LRP1_人	C.....PLDEFQCNTL.....CKPLAWK.CDGEDDCCGDN.....SDENPEE.....C
LRP1_人	C.....PPNRPFRCCKNDRV.....CLWIGRQ.CDGTDNCGDG.....TDEED.....C
LRP1_人	C.....KDKKEFLCRNQR.....CLSSSLR.CNMFDDCCGG.....SDEED.....C
LRP2_人	C.....DSAHFRCSGSH.....CIPADWR.CDGTKDCSDD.....ADEIG.....C
LRP2_人	C.....QQGYFKCQSEGO.....CIPSSWV.CDQDQDCDDG.....SDEKQD.....C
LRP2_人	C.....SSHQITCSNGO.....CIPSEYR.CDHVRDCPDG.....ADEND.....C
LRP2_人	C.....EQLTCDNGA.....CYNTSQK.CDWKVDCCRS.....SDEIN.....C
LRP2_人	C.....LHNEFSCNGE.....CIPRAYV.CDHDNDCCQDG.....SDEHA.....C
LRP2_人	C.....GGYFTCPSPGR.....CIYQNVV.CDGEDDCKDN.....GDEDG.....C
LRP2_人	C.....SPREWSCPESGR.....CISIIYKV.CDGILDCPGR.....EDENNTSTGKYC
LRP2_人	C.....GLFSFPCKNGR.....CVPNYYL.CDGVDDCHDN.....SDEQL.....C
LRP2_人	C.....SSSAFTCGHE.....CIPAHWR.CDKRNDCCVDG.....SDEHN.....C
LRP2_人	C.....LDTQYTCDNHQ.....CISKWV.CDNDNDCCGG.....SDEKN.....C
LRP2_人	C.....QPSQFNCPNHR.....CIDLSFV.CDGDKDCVDG.....SDEVG.....C
LRP2_人	C.....TASQFKCASGDK.....CIGVTNR.CDGVFDCSDN.....SDEAG.....C
LRP2_人	C.....HSDEFQCEDEGI.....CIPNFWE.CDGHPCCLYG.....SDEHNA.....C
LRP2_人	C.....PSSYFHCDNGN.....CIHRAWL.CDRDNDCCGDM.....SDEKD.....C
LRP2_人	C.....PSWQWQCLGHNI.....CVNLSVV.CDGIFFDCPNG.....TDESPJL.....C
LRP2_人	C.....GASSFTCSNGR.....CISEWK.CDNDNDCCGG.....SDEMESV.....C
LRP2_人	C.....SPTAFTCANGR.....CVQYSYR.CDIYNDCCGG.....SDEAG.....C
LRP2_人	C.....NATTEFMCNRR.....CIPREFI.CNGVDNCHDNT.....SDEKN.....C
LRP2_人	C.....QSGYTKCHNSNI.....CIPRVYL.CDGDNDCCGDN.....SDENPTY.....C
LRP2_人	C.....SSSEFQCASGR.....CIPQHWY.CDQETDCFDA.....SDEPAS.....C

图 6D

LRP2_人 C.....LADEFKCDGGR.....CIPSEWI.CDGDNDCCGDM.....SDEDKRHQ....C
LRP2_人 C.....SDSEFLCVNDRPPDRR.....CIPQSWV.CDGDVDTDGD.....YDENQN.....C
LRP2_人 C.....SENEFTCGYGL.....CIPKIFR.CDRHNDCCGY.....SDERG.....C
LRP2_人 C.....QQNQFTCQNGR.....CISKTFV.CDEDNDCGDG.....SDELMHL.....C
LRP2_人 C.....PPEFKCDNGR.....CIEMMKL.CNHLDDCLDN.....SDEKG.....C
LRP2_人 C.....SSTQFLCANNEK.....CPIWVK.CDGQKDCSDG.....SDELAL.....C
LRP2_人 C.....RIGQFQCSNGN.....CTSPQTL.CNAHQNCPDG.....SDEDRLL.....C
LRP2_人 C.....DSNEWQCANKR.....CIPESWQ.CDTFNDCCDN.....SDESSH.....C
LRP2_人 C.....RPGQFRCANGR.....CIPQAWK.CDVNDCCGDH.....SDEPIEE.....C
LRP2_人 C.....DNFTEFSCKTYR.....CIPKWAV.CNGVDDCKDN.....SDEQG.....C
LRP2_人 C.....HPVGDFRCKNHH.....CIPLRWQ.CDGQNDCCGN.....SDEEN.....C
LRP2_人 C.....TESEFRCVNQQ.....CIPSRWI.CDHYNDCGDN.....SDERD.....C
LRP2_人 C.....HPEYFQCTSGH.....CVHSELK.CDGSADCLDA.....SDEAD.....C
LRP2_人 C.....QATMFECKNHV.....CIPPYWK.CDGDDDCGDG.....SDEELHL.....C
LRP2_人 C.....NSPNRFRCDNRR.....CIYSHEV.CNGVDDCCGDG.....TDETEEH.....C
LRP2_人 C.....TEYEYKCGNGH.....CIPHNV.CDDADDCGDW.....SDELG.....C
LR1B_人 C.....DPGEFLCHDHVT.....CVSQSWL.CDGDPPDCPD.....SDESLDT.....C
LR1B_人 C.....PLNHIACLGTNK.....CVHLSQL.CNGVLDCPDG.....YDEGVH.....C
LR1B_人 C.....KAGEFRCKNRH.....CQARWK.CDGDDDCLDG.....SDEDSVN.....C
LR1B_人 C.....PDDQFKCQNNR.....CIPKRWL.CDGANDCGSN.....EDESNOT.....C
LR1B_人 C.....QVDQFSCNGR.....CIPRAWL.CDREDDCCGQ.....TDEMAS.....C
LR1B_人 C.....EPLTQFVCKSGR.....CISSKWH.CDSDDDCGDG.....SDEVG.....C
LR1B_人 C.....FDNQFRCSSGR.....CIPGWA.CDGDNDCCGDF.....SDEAQIN.....C
LR1B_人 C.....NGNEFQCHPDGN.....CVPDLWR.CDGEKDCEDG.....SDEKG.....C
LR1B_人 C.....DHKTKFSCWSTGR.....CINKAWV.CDGDIDCEDQ.....SDEDD.....C
LR1B_人 C.....GPPKHPCANDTSV.....CLQPEKL.CNGKKDCPDG.....SDEGYL.....C
LR1B_人 C.....NAYSEFECNGE.....CIDYQLT.CDGI PHCKDK.....SDEKLLY.....C
LR1B_人 C.....RRGFKPCYNRR.....CIPHGKL.CDGENDCGDN.....SDELD.....C
LR1B_人 C.....ATVEFRCADGT.....CIPRSAR.CNQNICADA.....SDEKN.....C
LR1B_人 CTHFYKLGVTGTGIRCNSTSL.....CVLPTWI.CDGSNDCCGY.....SDELK.....C
LR1B_人 C.....EENYFSCPSGR.....CILNTWI.CDGQKDCEDG.....RDEFH.....C
LR1B_人 C.....SWNQFACSAQK.....CISKHWI.CDGEDDCGDG.....LDESISI.....C
LR1B_人 C.....AADMFSCQGSRA.....CVPRHVL.CDGERDCPDG.....SDELSTAG.....C

图 6E

```

LR1B_人 DENAFMCHNKV.....CIPKQFV.CDHDDDCGDG.....SDESPQ.....C
LR1B_人 GTEEFSCADGR.....CLLNTQWQCDGDFDCPDH.....SDEAPLNPK...C
LR1B_人 NSSFFMCKNGR.....CIPSGGL.CDNKDDCGDG.....SDERN.....C
LR1B_人 TASQFRCKTDK.....CIPFWWK.CDTVDDCGDG.....SDEPDD.....C
LR1B_人 QPGRFQCGTGL.....CALPAFI.CDGENDCGDN.....SDELN.....C
LR1B_人 LSGQFKCTKNQK.....CIPVNLR.CNGQDDCGDE.....EDERD.....C
LR1B_人 SPDYFQCKTKKH.....CISKLWV.CDEDPCADA.....SDEAN.....C
LR1B_人 GPHEFQCKNNN.....CIPDHRW.CDSQNDCCSDN.....SDEEN.....C
LR1B_人 TLKDFLCANGD.....CVSSRFW.CDGDFDCADG.....SDERN.....C
LR1B_人 SKDQFRCSNGQ.....CIPAKWK.CDGHEDCYK.....EDEKS.....C
LR1B_人 SSREYICASDG.....CISASLK.CNGEYDCADG.....SDEMD.....C
LR1B_人 KEDQFRCKNAH.....CIPIRWL.CDGIHDCVDG.....SDEEN.....C
LR1B_人 RADEFLCNNSL.....CKLHFVW.CDGEDDCGDN.....SDEAPDM.....C
LR1B_人 PSTRPHRCRNRI.....CLQSEQM.CNGIDECGDN.....SDEDH.....C
LR1B_人 KKDEFACSNKK.....CIPMDLQ.CDRLLDDCGDG.....SDEQG.....C

075851 ..AEGEALCOENGH.....CVPHGWL.CDNQDDCGDG.....SDEEGE.....C
075851 ..GEGQMTCSGH.....CLPLALL.CDRQDDCGDG.....TDEPSYP.....C
075851 ..PQGLLACADGR.....CLPPALL.CDGHFDC LDA.....ADEES.....C
075851 ..VPGEVSCVDGT.....CLGAIQL.CDGVWDCPDG.....ADEGPGH.....C
ENSP00000262089
= 075851 ..GPFEFRCSGE.....CTPRGWR.CDQEEDCADG.....SDERG.....C
ENSP00000262089
075851 ..APHHAPCARGPH.....CVSPEQL.CDGVRQCPDG.....SDEGPDA.....C
075851 ..PGLFPCGVAPGL.....CLTPEQL.CDGIPDCPQG.....EDEL.....C
075851 ..PEYTCPNGT.....CIGFQLV.CDGQPDCCRPGQVGPSPPEEQ.....C
075851 ..EPGVGLRCASGE.....CVLRGGP.CDGVLDCEDEG.....SDEEG.....C
ENSP00000262089
075851 ..GPGQTPCEVLG.....CVEQAQV.CDGRREDCLDG.....SDERH.....C
075851 ..SPSQLSCGSGE.....CLSAERR.CDLRPDCQDG.....SDEDEG.....C
C18ORF1 ..KFTCTSGK.....CLYGLSLVCNQNDCCGDN.....SDEEN.....C
AAH07083/Q9NPF0

```

图 6F

AAH07083/Q9NPF0
 C.....PPTKFQCR^TSGL.....CVPLTWR.CDRDLDCSDG.....SDEEE.....C
 C.....LAGELRCTLSD^D.....CIPLTWR.CDGHFDCPPS.....SDELG.....C
 Q9HBX9 C.....SLGYFPCGNITK.....CLPQLLH.CNGVDDCGNQ.....ADEDN.....C
 Q9BY79/Q96DQ9 C.....AHDEFRCQLI.....CLLPDSV.CDGFANCADG.....SDETN.....C
 Q9BY79/Q96DQ9 C.....GPSELSCQAGG.....CKGVQWM.CDMWRDCTDG.....SDDN.....C
 BAB55257 =
 ENSP00000239367
 C.....SRYHFFCDDGC.....CIDITIA.CDGVQCCPDG.....SDEF.....C
 C.....PGEFLCSVNGL.....CVPA.....CDGVKDCPNG.....LDERN.....C
 O95518 =
 ENSP00000255793
 ENSP00000255793
 C.....RATFQCKEDST.....CISLPKV.CDGQPDCLNG.....SDEEQ.....C
 ENSP00000255793
 C.....GTFTFQCEDRS.....CVKKNPQCDCGRPDCRDG.....SDEEH.....C
 Q8WXD0 C QKGYFPCGNLTK CLPRAFH CDGKDDCCNG ADEEN C
 Q8NBJ0 C STARYHCKNGL CIDKFSI CDGQNNCQDN SDEES C
 Q8NBJ0 C GPTFFPCASGIH CIIGRFR CNGFEDCPDG SDEEN C
 Q8NBJ0 C NIPGNFMCNNGR CIPGAWQ CDGLPDCFDK SDEKE C
 MEGF7 C ALDQFLCWNGR CIGQRKL CNGVNDCCGN SDESPQN C
 MEGF7 C EEDEFPCQNGY CIRSLWH CDGDNDCCGN SDEQ C
 MEGF7 C RSGEFMCD^SGL CINAGWR CDGDADCDDQ SDERN C
 MEGF7 C TAEQFRCHSGR CVRLSWR CDGEDDCADN SDEEN C
 MEGF7 C SPLDFHCDNGK CIRRSWV CDGDNDCEDD SDEQD C
 MEGF7 C NLEEFQCA^YGR CILDIYH CDGDDDCGDW SDESD C
 MEGF7 C SDKEFRCS^DGS CIAEH^WY CDGDTDC^KDG SDEEN C
 MEGF7 C GRSHFTCAVSALGECT CIPAQWQ CDGDNDCCGDH SDEDG C
 CAD61944 C LQEEFQCLNHR CVSAVQR CDGVDACGDG SDEAG C
 CAD61944 C PPGHFP^CGAAGTSGATA CYLPADR CNYQTFCADG ADERR C
 CAD61944 C QPGNFR^CRDEK CVYETWV CDGQPD^CCADG SDEWD C
 ENSG00000181006
 C PEITDFLC^RDKK CIA^SHLL CDYK^PDCSDR SDEAH C

ENSP00000320248	C	NNRTFKCGNDI	CFRKQNAKCDGTVDCPDG	SDEEG	C
ENSP00000277547	C	PPGHHHCQNKV	CVEPQQL CDGEDNCGDL	SDENPLT	C
ENSP00000320022	C	KQGHLAGGDL	CVPPEQL CDFEEQCAGG	EDEQA	C
ENSP00000313222	C	PGNSFSCGNSQ	CVTKVNPECDDQEDCSDG	SDEAH	C

图 6H

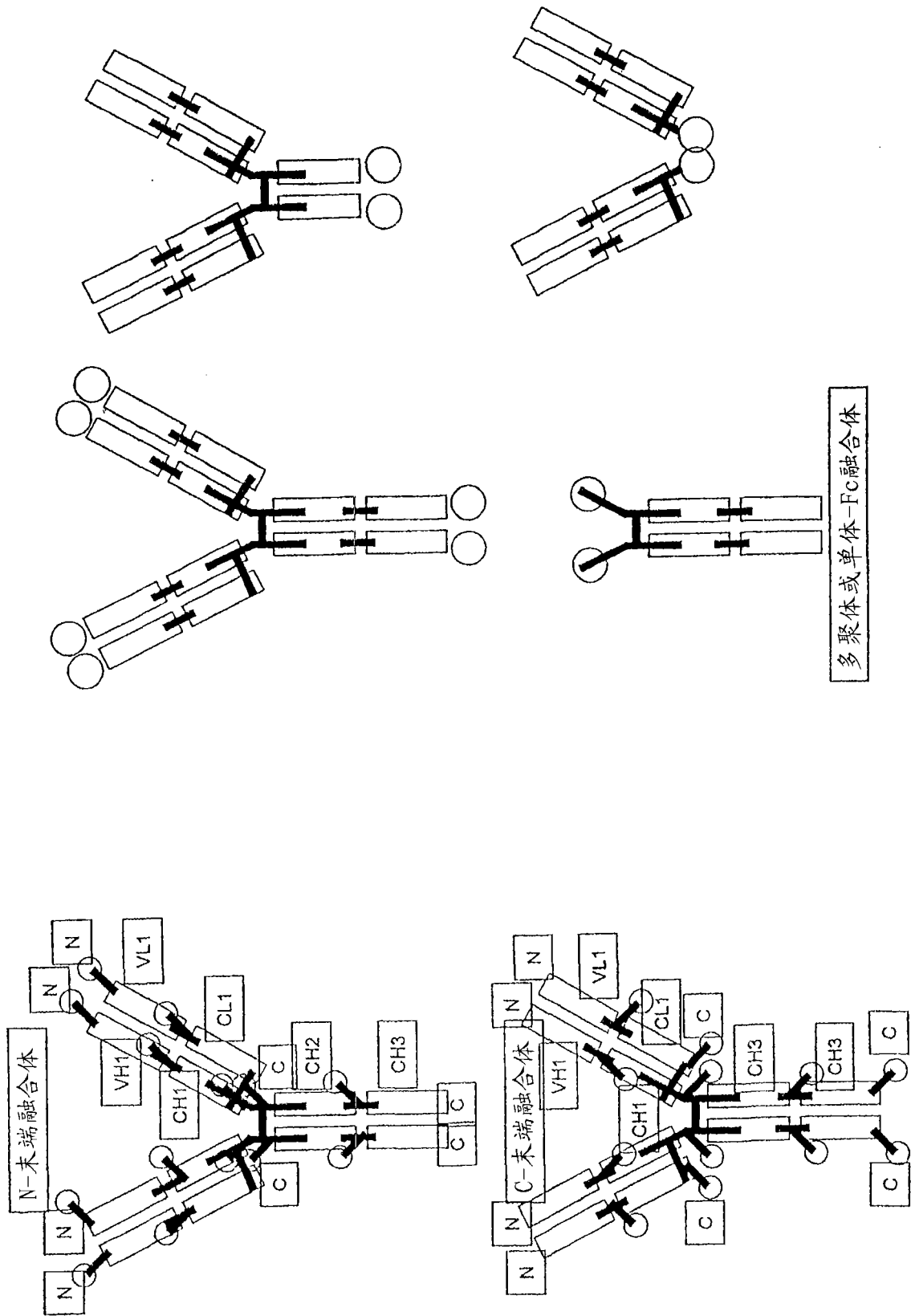


图 7

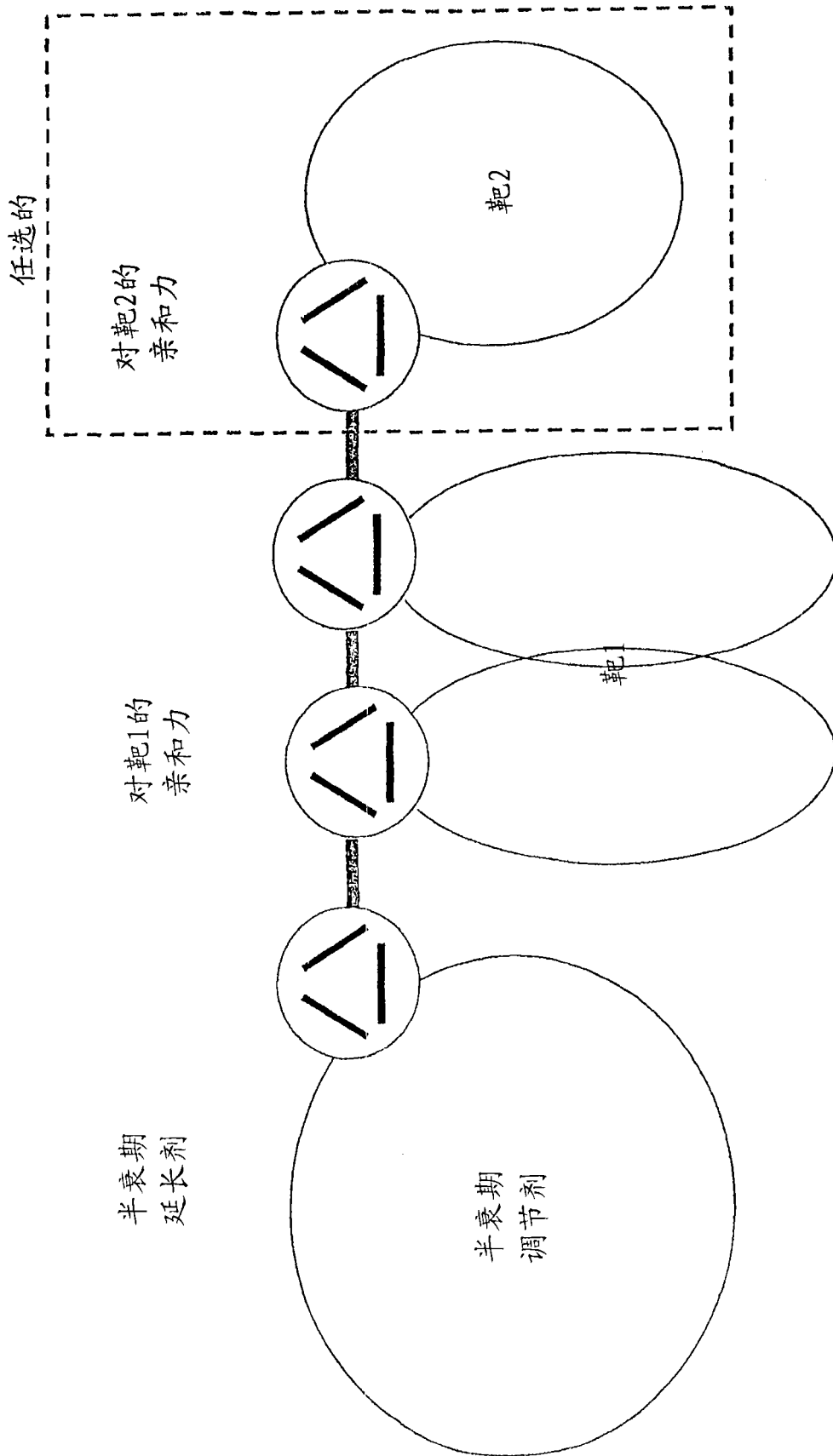


图 8

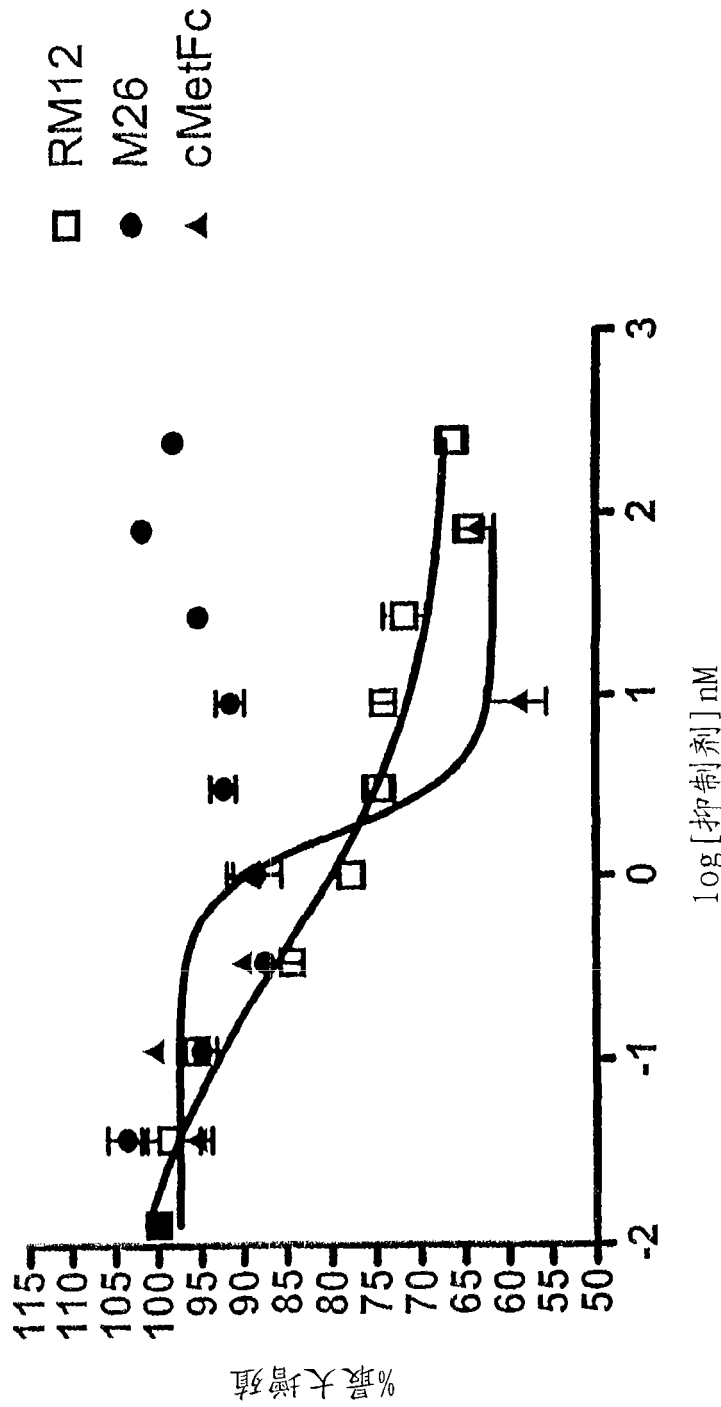


图 9

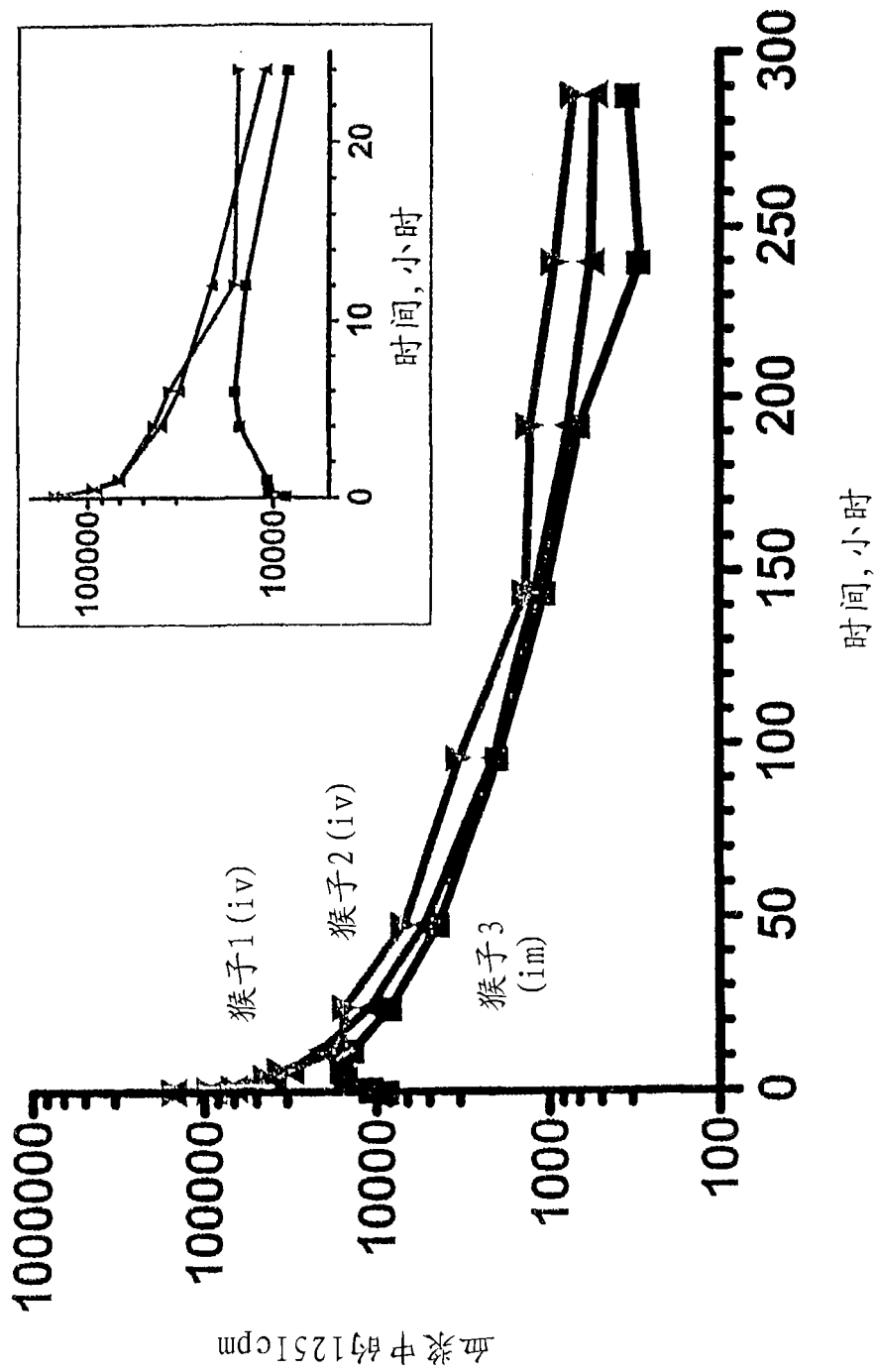


图 10

专利名称(译)	c - MET激酶结合蛋白		
公开(公告)号	CN101018559A	公开(公告)日	2007-08-15
申请号	CN200580028093.8	申请日	2005-06-17
[标]申请(专利权)人(译)	安姆根山景公司		
申请(专利权)人(译)	安姆根山景公司		
当前申请(专利权)人(译)	安姆根山景公司		
[标]发明人	A巴克 Q刘 J西尔弗曼 R史密斯		
发明人	W·P·C·斯特默 D·V·佩尔罗思 S·萨特亚尔 B·M·阿尔巴 A·巴克 A·N·杜瓜伊 Q·刘 J·西尔弗曼 R·史密斯 C·斯维默		
IPC分类号	A61K38/00 G01N33/53 C07H21/02		
CPC分类号	G01N2500/00 A61K38/00 G01N33/574 C07K14/47		
优先权	10/871602 2004-06-17 US 10/957351 2004-09-30 US		
外部链接	Espacenet SIPO		

摘要(译)

提供了包含结合c - MET的单体域或其部分的多肽。

n

$$\left(\sum_{i=1}^n B\text{-因子 } C_{ai} \right) / n$$