(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 102421789 B
(45)授权公告日 2015.02.25

(21)申请号 201080020628.8
(22)申请日 2010.04.14
(30)优先权数据
61/169,636 2009.04.15 US
(85)PCT国际申请进入国家阶段日
2011.11.14
(86)PCT国际申请的申请数据
PCT/KR2010/002318 2010.04.14
(87)PCT国际申请的公布数据
WO2010/120121 KO 2010.10.21
(73)专利权人 浦项工科大学校产学协力团
地址 韩国庆尚北道
(72)发明人 崔允燮 郑埈昊 俞知镐 曹炫寿 尹秀珉 金耿禄 柳成浩 金相旭
(74)专利代理机构 上海一平知识产权代理有限公司 31266
代理人 雷芳 祝莲君
(51)Int.Cl.
C07K 1/107(2006.01)
C07K 14/485(2006.01)
G06F 19/18(2011.01)
C12N 15/12(2006.01)
C12N 15/63(2006.01)
C12N 1/21(2006.01) (续)

(56)对比文件
CN 1472336 A,2004.02.04, 全文.
CN 1730665 A,2006.02.08, 全文.
CN 1826129 A,2006.08.30, 全文.
US 2008/0015194 A1,2008.01.17, 全文.
CN 101205555 A,2008.06.25, 全文.
E.FIEDLER ET AL.AFFILINTM MOLECULES Novel Ligands for Bioseparation.《Food and Bioproducts Processing》.2006, 第3-8页.
叶德举等.先导化合物的发现——整合计算机虚拟筛选、化学合成和生物测试方法.《化学进展》.2007, 第19卷(第12期), 第1939-1946页.
Martin Offterdinger et al.Imaging Phosphorylation Dynamics of the Epidermal Growth.《THE JOURNALOF BIOLOGICAL CHEMISTRY》.2004, 第279卷(第35期), 第36972-36981页.
李洪林等.虚拟筛选与新药发现.《生命科学》.2005, 第17卷(第2期), 第125-131页.
MichaelT.Stumpp et al.DARPins:A new generation of protein therapeutics.《Drug Discovery Today》.2008, 第13卷(第15/16期), 第695-701页.

审查员 夏文静

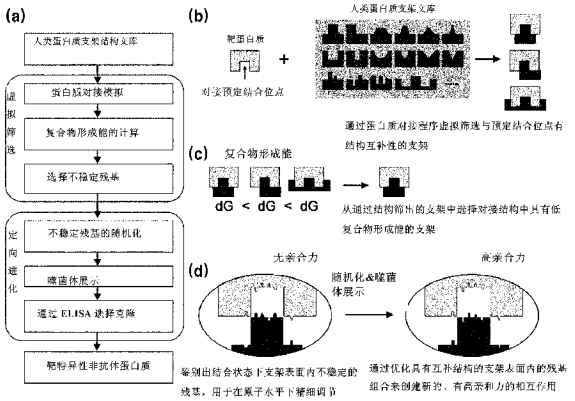
权利要求书1页 说明书12页
序列表5页 附图7页

(54)发明名称

靶特异性非抗体蛋白质及其制备方法

(57)摘要

本发明涉及一种制备靶特异性非抗体蛋白质的方法,更具体地,涉及一种制备靶特异性非抗体蛋白质的方法,包括步骤:从非抗体蛋白质文库中选择具有与靶蛋白质的靶位点有结构互补性的非抗体蛋白质;计算所选择的非抗体蛋白质与靶蛋白质之间的结合能;从所选择的非抗体蛋白质中选择具有理想结合能的非抗体蛋白质;从选择的非抗体蛋白质与靶蛋白质的界面氨基酸残基中选择具有高结合能的氨基酸残基;用具有低结合能的氨基酸残基替代选定的氨基酸残基。另外,本发明涉及用这种方法制备的、与EGFR(表皮生长因子受体)结构域2特异性结合的靶特异性的非抗体蛋白质,以及含有该靶特异性非抗体蛋白质的癌症治疗组合物。



[转续页]

CN 102421789 B

[接上页]

(51) Int.Cl.

C12N 1/15(2006.01)

C12N 1/19(2006.01)

C12N 5/10(2006.01)

A61K 38/18(2006.01)

A61P 35/00(2006.01)

1. 一种制备靶特异性非抗体蛋白质的方法，其特征在于，包括以下步骤：

(a) 在非抗体蛋白质文库中，选择与靶蛋白质的靶位点具有结构互补性的非抗体蛋白质；

(b) 计算所选择的非抗体蛋白质和靶蛋白质之间的结合能；

(c) 从所选择的非抗体蛋白质中，选择具有最低结合能的非抗体蛋白质；

(d) 从所选择的非抗体蛋白质与靶蛋白质的界面氨基酸残基中，选择具有高结合能的氨基酸残基；和

(e) 用具有较低结合能的氨基酸残基替代步骤 (d) 中选择的氨基酸残基，

其中，步骤 (a) 中的非抗体蛋白质文库包括非抗体蛋白质，所述非抗体蛋白质分子量为 10-40kDa，选自人类蛋白质，并能够形成单体、同源二聚体和同源三聚体中的任何一种或多种，

步骤 (a) 中的靶蛋白是 EGFR（表皮生长因子受体）结构域 2，

其中，步骤 (a) 中的文库是用包含以下步骤的方法构建的：

(i) 从人类蛋白质中选择具有 10-40kDa 分子量的蛋白质；

(ii) 从步骤 (i) 中选择的蛋白质中排除膜蛋白质和抗体蛋白质；

(iii) 从步骤 (ii) 中剩余的蛋白质中选择具有 10 或更少种已知相互作用的蛋白质；和

(iv) 从步骤 (iii) 中选择的蛋白质中选择能形成单体、同源二聚体和同源三聚体中的任一或多种的蛋白质，

其中，步骤 (a) 中用对接模拟进行选择，

其中，步骤 (e) 是用噬菌体展示和生物筛选 (bio-panning) 技术进行的。

2. 如权利要求 1 所述的方法，其特征在于，步骤 (a) 和步骤 (b) 依次进行或同时进行。

靶特异性非抗体蛋白质及其制备方法

技术领域

[0001] 本发明涉及一种制备靶特异性非抗体蛋白质的方法，更具体地，涉及一种制备靶特异性非抗体蛋白质的方法，包括步骤：从非抗体蛋白质文库中选择与靶蛋白质的靶位点有结构互补性的非抗体蛋白质；计算所选择的非抗体蛋白质与靶蛋白质之间的结合能；在所选择的非抗体蛋白质中选择具有理想结合能的非抗体蛋白质；在所选择的非抗体蛋白质与靶蛋白质的界面氨基酸残基中选择具有高结合能的氨基酸残基；以及，用具有低结合能的氨基酸残基替代所选择的氨基酸残基。另外，本发明涉及用这种方法制备的、与 EGFR（表皮生长因子受体）结构域 2 特异性结合的靶特异性非抗体蛋白质，以及含有该靶特异性非抗体蛋白质的癌症治疗组合物。

背景技术

[0002] 抗体通常被用做新的治疗蛋白药物。抗体是由免疫系统中称作 B 细胞的血液白细胞受到抗原刺激后产生的蛋白质。当抗体遇到抗原时，它们经由细胞上的受体来识别并结合抗原。其特点在于，抗体特异性识别并紧密结合某特定蛋白质。由于具有结合致病（如癌症）蛋白质这一特点，抗体已被用于预防恶性蛋白质间的相互作用的目的。

[0003] 但是，治疗性抗体的分子尺寸大，难以大量生产，而且它们的生产工艺需要长时间加工和高昂的成本。另一种选择是，目前已在积极研究其它的用于病人定制和靶向治疗的非抗体蛋白质支架。非抗体蛋白质是一类用于病人定制和靶向治疗的新型药物，作为克服已经引起目前药物市场兴趣的治疗性抗体的局限性的另一种选择，最近开始引起关注。像治疗性抗体一样，非抗体蛋白质旨在通过选择性识别特定靶分子并与其紧密结合以抑制该靶分子活性的方式来治疗各种疾病，如癌症和自身免疫性疾病，或者防止疾病发展。

[0004] 但是，和传统抗体一样，现有非抗体蛋白质的制备也是依赖单一的支架而不管靶分子的结构。在三级结构基础上，蛋白质相互结合，就像乐高积木那样互相扣合在一起。所以，理想的做法是针对特定靶标使用定制的蛋白质支架。但是，还不存在可以找到具有结构互补性、能特异性结合靶标的非抗体蛋白质的技术。问题在于，找到用于疾病治疗的靶特异性非抗体蛋白质是困难的。由于蛋白质结构和类型的多样性，在实验中不可能鉴定哪种蛋白质特异性结合于特定的靶标。

[0005] 对于基于单一支架的现有抗体和非抗体蛋白质来说，它们的靶位点结合区是偶然决定的，所以无法预先设计结合区。另外，与已知的非抗体蛋白质和抗体结合的靶位点具有结构限制性。要靠近凹形靶位点基本上是不可能的。不便的是，为了从靶分子结合蛋白质中找到能结合目的靶分子区的蛋白质，要进行长期的筛选候选蛋白质的复杂过程，如表位定位过程。

[0006] 为了克服现有抗体和非抗体蛋白质的局限性，迫切需要开发一项技术，用于选择针对靶标的优化支架，以设计结合该靶分子的目的靶位点的非抗体蛋白质。用于设计和制备与靶蛋白质目的位点有紧密结合能力和特异性的非抗体蛋白质的技术是一项适用广泛范围的靶蛋白质的平台技术，这项技术在新型抗体药物或适体领域尚未得到认识。所以，业

界也迫切需要开展分阶段的病人定制和靶向治疗。

[0007] 本发明人致力于开发一种制备靶特异性非抗体蛋白质的方法。结果他们发现：根据具有预定三级结构的蛋白质结构文库，虚拟筛选一给定靶蛋白质的目的靶位点，以选择具有最高结构互补性的最佳蛋白质支架，以及，利用噬菌体展示技术和生物筛选技术(bio-panning)对干扰靶分子结合的氨基酸进行随机化以制备能最佳结合预定靶蛋白质特定位点的靶特异性非抗体蛋白质。他们进一步制备了靶向抗癌治疗靶EGFR(表皮生长因子受体)结构域2的非抗体蛋白质，并且检测了它与该靶标的结合力，从而完成了本发明。

发明内容

[0008] 本发明的一个目的是提供一种制备靶特异性非抗体蛋白质的方法，其包括以下步骤：(a)从非抗体蛋白质文库中选择与靶蛋白质的靶位点有结构互补性的非抗体蛋白质；(b)计算所选择的非抗体蛋白质与靶蛋白质之间的结合能；(c)从所选择的非抗体蛋白质中选择理想结合能的非抗体蛋白质；(d)从所选择的非抗体蛋白质与靶蛋白质的界面氨基酸残基中选择具有高结合能的氨基酸残基；以及，(e)用具有低结合能的氨基酸残基替代步骤(d)中选定的氨基酸残基。

[0009] 本发明的另一目的是提供一种能特异性结合EGFR结构域2的靶特异性非抗体蛋白质。

[0010] 本发明的另一目的是提供一种编码特异性结合EGFR结构域2的靶特异性非抗体蛋白质的核酸、含有该核酸的载体或用该载体转化的转化子。

[0011] 本发明的另一目的是提供一种含有特异性结合EGFR结构域2的靶特异性非抗体蛋白质的癌症治疗组合物。

附图说明

[0012] 图1是用于设计本发明的靶特异性支架的PELEX(运用实验室进化与大量计算的蛋白质工程)方法的概况，其中：

[0013] (a)是PELEX二步法的流程图，包括步骤：根据人类蛋白质支架文库进行的对接模拟(docking simulation)和复合物形成能计算的虚拟筛选步骤和优化所筛选支架与预定结合位点间结合亲和力的定向进化步骤，

[0014] (b)是根据人类蛋白质支架文库和对接模拟结果选择与靶蛋白质预定结合位点有互补形状的蛋白质支架的流程图，其中大规模的蛋白质支架文库包括所有已知的人类蛋白质结构，

[0015] (c)是选择展示理想结合能的支架供定向进化步骤作进一步优化的流程图，该定向进化步骤是在使用所筛选的对接结构计算复合物形成能后进行，以及

[0016] (d)是通过序列随机化和噬菌体展示技术进行定向进化来精筛靶特异性支架的流程图，其中将不理想的界面残基随机化，用于优化靶和支架间的亲和力；

[0017] 图2显示了EGFR的氨基酸序列，其中用下划线标出的氨基酸是EGFR结构域2的第166-309个氨基酸，方框内的28个氨基酸被选作界面残基候选氨基酸；

[0018] 图3显示了EGFR活化机制和设计用来抑制EGFR活化的支架活化机制，其中：

[0019] (a)显示了在自我抑制状态下结构域2与结构域4之间发生分子内相互作用使

EGFR 失活，结构域 2 的二聚化臂被关闭，

[0020] (b) 显示了 EGF 结合使结构域 2 与结构域 4 之间的分子内相互作用消失并促使大结构域重排，暴露结构域 2 来激活 EGFR，然后 EGFR 分子通过结构域 2 的每个二聚化臂间的相互作用形成同源二聚体，以及

[0021] (c) 显示了通过所设计的支架与 EGFR 结构域 2 之间的结合来防止同源二聚体形成从而抑制 EGFR 活性的机制，其中该结合蛋白只识别活化的 EGFR，但不识别自我抑制的 EGFR ；

[0022] 图 4 显示了 1OZJ 的结构和能量，其中 ：

[0023] (a) 显示了 1OZJ 的结构 (B 链)，

[0024] (b) 显示了 1OZJ-EGFR 的对接结构，其中深色链表示支架，浅色链表示 EGFR 的结构域 1、2、3 和 4，EGFR 的界面残基和支架用三维方式示出 ((a) 中的支架结构与 (b) 中的显示方向一致)，以及

[0025] (c) 显示了 1OZJ 的与 EGFR 形式复合物的界面残基的能量贡献值，其中选出的供定向进化做进一步精筛的不理想残基用斜纹条表示，它们是进一步随机化发生突变的氨基酸残基 ；

[0026] 图 5 是 ELISA 结果，显示了 1OZJ 的不理想残基发生突变后的结合亲和力，其中横轴代表克隆编号，纵轴代表 ELISA 分析得到的光密度，1OZJ 克隆的序列随机化结果如表 2 所示 ；

[0027] 图 6 显示了 1OZJ 支架随机化突变前的野生型支架的氨基酸序列 (左) 和 DNA 序列 (右)，其中，用下划线标示的序列是要进行突变的不理想残基 ；

[0028] 图 7 显示了 1OZJ 支架的 6 号克隆的氨基酸序列 (左) 和 DNA 序列 (右)，其中，用下划线标示的序列是要进行突变的不理想残基 ；

[0029] 图 8 显示了 1OZJ 支架的 7 号克隆的氨基酸序列 (左) 和 DNA 序列 (右)，其中，用下划线标示的序列是要进行突变的不理想残基 ；以及

[0030] 图 9 显示了 1OZJ 支架的 9 号克隆的氨基酸序列 (左) 和 DNA 序列 (右)，其中，用下划线标示的序列是要进行突变的不理想残基。

具体实施方式

[0031] 为了实现上述目的，本发明的一个实施例涉及制备靶特异性非抗体蛋白质的方法，包括步骤 ：(a) 从非抗体蛋白质文库中选择与靶蛋白质的靶位点有结构互补性的非抗体蛋白质 ；(b) 计算所选择的非抗体蛋白质与靶蛋白质之间的结合能 ；(c) 从所选择的非抗体蛋白质中选择具有理想结合能的非抗体蛋白质 ；(d) 从所选择的非抗体蛋白质与靶蛋白质的界面氨基酸残基中选择具有高结合能的氨基酸残基 ；以及，(e) 用具有低结合能的氨基酸残基替代步骤 (d) 中选定的氨基酸残基。

[0032] 本发明制备靶特异性非抗体蛋白质的方法的概况如图 1a 所示。本发明人定义该制备靶特异性非抗体蛋白质的方法为 PELEX(运用实验室进化与大量计算的蛋白质工程)，该方法包括步骤 ：通过本发明人开发的对接模拟程序进行虚拟筛选并计算结合能 ；进行定向进化，该定向进化是通过噬菌体展示和生物淘洗技术进行的残基随机化来完成。

[0033] 如本文所用，术语“非抗体蛋白质”是指分子量为 10-40kDa、比现有治疗性抗体分

子量低 1/10 的蛋白质，并且是能轻易透过靶组织并能在细菌等中大量生产的非抗体蛋白质。该术语还指一种通过选择性识别特定靶分子并与其紧密结合以抑制这些靶分子活性的用于治疗各类疾病（如癌症和自身免疫性疾病）或防止这些疾病发展的蛋白质。在本文中，“非抗体蛋白质”与“支架”、“支架蛋白质”或“结合蛋白质”可互换使用。

[0034] 如本文所用，术语“非抗体蛋白质文库”是指包括所有自然界中蛋白质结构的蛋白质支架的集合，其三级结构已被人类揭示，可将这个集合制备后存在数据库中。对于非抗体蛋白质文库的制备，可以使用但不限于包括人类蛋白质全部三级结构的数据库如 PDB（蛋白质数据库）和 SCOP（蛋白质结构分类体系）。也可以使用其他各种包括人类蛋白质全部三级结构的数据库来选择用于制备非抗体蛋白质文库的蛋白质。这些数据库根据人类蛋白质的每个三级结构被分成许多组。非抗体蛋白质文库的具体制备方法如下所述。首先，为利于穿透组织，从人类蛋白质中只选择具有 10-40kDa 有限分子量的蛋白质。为此，从人类蛋白质的每个结构类别中选择分子量为 10-40kDa 的 5 个或 5 个以内有代表性的蛋白质。这样就包括了所有类型的人类蛋白质，也缩小了电子文库的大小，从而加快检索速度。然后，为了从人类蛋白质中排除不同于非抗体蛋白质的蛋白质，还可以包括去除膜蛋白和抗体蛋白的步骤。使用具有已知结构的蛋白质数据库（但不限于此），用以排除膜蛋白和抗体蛋白，例如，可以使用 PDBTM（跨膜蛋白质的蛋白质数据库）等数据库进行关键词检索。之后，为避免随机结合，还可以包括选择已知相互作用数量仅为 10 或少于 10 的蛋白质的步骤。关于这些相互作用的信息也可从提供蛋白质 - 蛋白质相互作用信息的各类数据库中获得。例如，可利用 HPRD（人类蛋白质参考数据库）计算蛋白质 - 蛋白质相互作用的数量。为了排除具有大量相互作用的蛋白质，如同源四聚体和同源六聚体，还可以包括选出能形成单体、同源二聚体和同源三聚体中的任意一个或多个的蛋白质的步骤。可以使用通常用来确定蛋白质间结合情况的已知数据库（但不限于此），例如，SWISS-PROT 数据库。利用上述所有步骤制备的非抗体蛋白质文库可以包括 1000-2000 个结构已知的非抗体蛋白质。根据本发明的一个实施例，制备了一个非抗体蛋白质文库，所述文库包括 1261 个非抗体可溶性蛋白质，所述蛋白质分子量低、随机结合的风险低，虚拟筛选如对接模拟操作就在这个文库基础上进行。

[0035] 优选地，步骤 (a) 从非抗体蛋白质文库中选择与靶蛋白质的靶位点有结构互补性的非抗体蛋白质和步骤 (b) 计算所选择的非抗体蛋白质与靶蛋白质之间的结合能，可以按顺序进行或同时进行。

[0036] 如本文所用，术语“靶蛋白质”是指与非抗体蛋白质结合的所需要的蛋白质，根据所需要的应用，通过与非抗体蛋白质结合而不同。它可以选自（但不限于此）：致癌蛋白质、免疫性疾病的致病蛋白质和感染的外源蛋白质，例如 EGFR（表面生长因子受体）。

[0037] 如本文所用，术语“靶位点”是指所述靶蛋白三级结构内的、直接或间接参与病原蛋白质相互作用与结合的区域，所述非抗体蛋白质与所述对应区域紧密结合以防止恶性蛋白质的相互作用。靶位点例如可以是 EGFR 结构域 2。进一步，与本发明的非抗体蛋白质结合的靶蛋白质的结合位点可以有各种表面，如凹面，所以非抗体蛋白质的制备不限于与其结合的靶蛋白质的三级结构。

[0038] 如本文所用，术语“EGFR（表面生长因子受体）”是一个代表性的生长因子受体，在乳腺癌、肺癌、结肠癌、肾癌、胆囊癌、头颈癌、卵巢癌、前列腺癌、宫颈癌和胃癌等各种癌症

细胞的表面过量表达。它是一种 EGF 受体，通过与生长因子 EGF 结合而被激活，从而在癌细胞增殖、入侵、血管生成或转移方面扮演重要角色。直到最近，各种抑制 EGFR 以选择性攻击癌细胞的 EGFR 靶向制剂一直在积极研究当中，但还没有开展如本发明中以 EGFR 结构域 2 为靶的靶向 EGFR 结构域 2 的研究。

[0039] EGFR 由四个结构域组成。它与配体 EGF 结合，引起构象变化，暴露的结构域 2 介导同源二聚体形成，导致 EGFR 活化。所以，与本发明中用作靶的 EGFR 结构域 2 紧密结合的非抗体蛋白质特异性地结合到因结合配体 EGF 而外露的结构域 2 上以防止同源二聚体形成，从而特异性地抑制 EGFR 活性（图 3）。简要地说，由于非抗体蛋白质仅选择性地抑制活化的 EGFR，所以它会选择性地攻击正常细胞外的癌细胞，从而发挥癌症治疗制剂的作用。

[0040] 从非抗体蛋白质文库中选择与靶蛋白质的靶位点有结构互补性的非抗体蛋白质的步骤 (a) 是用超级计算机检查制备好的非抗体蛋白质文库中的多个非抗体蛋白质或支架是否与靶蛋白质有结构互补性的步骤，该步骤可通过使用对接模拟程序进行虚拟结合来完成。这个步骤是要通过靶与支架文库中各种不同支架间的虚拟结合来制备各种对接复合物，这种对接模拟操作可通过已知的程序（但不限于此）完成，如 ZDOCK、PIPER、ClusPro、ICM-DISCO、RosettaDock、PatchDock 和 MolFit。在本发明的一个实施例中，使用 PatchDock 来选择具有结构互补性的非抗体蛋白质。通过结构互补性结合单个靶和单个非抗体蛋白质的方法有很多种，PatchDock 程序按照互补性将对接结构排名。这时候，从各非抗体蛋白质的各种对接结构中选择了排在前 10 位的对接结构，供下一分析步骤使用。优选地，所述对接模拟步骤是要选择与靶蛋白质中的预定数量或多于预定数量的靶氨基酸残基结合的非抗体蛋白质。更优选地，如果靶蛋白质的靶位点是 EGFR 结构域 2，靶与非抗体蛋白质之间的界面残基数量可以是 28，该预定数量是 10。也就是说，当 28 个界面残基中有 10 个或 10 个以上参与了结合，就认为该对接结构的支架是用于结合 EGFR 结构域 2 的理想结构，因此可选择形成该对接结构的非抗体蛋白质。界面残基的数量可根据靶蛋白质或靶位点的类型变化而变化。

[0041] 步骤 (b) 是计算步骤 (a) 中选定的非抗体蛋白质与靶蛋白质之间的结合能，并且是根据对接模拟结果，选择能在预定水平或更高水平与靶蛋白质的目的区域结合的非抗体蛋白的步骤。在这方面，当参与结合的靶残基数量超过预定数量时，发生的结合较为理想。

[0042] 非抗体蛋白质与靶之间的结合能可根据以下方程式 1 计算 :1、结合能可按照各种已知程序计算，例如但不限于，EGAD 或 RossettaDesign 等。根据本发明的一个实施例，使用 EGAD 来计算结合能。

[0043] 方程式 1

[0044] $\Delta G_{结合} = \Delta G_{复合物} - \Delta G_{自由}$

[0045] 结合能 ($G_{结合}$) 是靶蛋白质与非抗体蛋白质之间结合所需的能量，定义为 $G_{复合物}$ 与 $G_{自由}$ 之间的能量差，$G_{复合物}$ 是靶蛋白质与非抗体蛋白质的复合物的能量，$G_{自由}$ 是结合前靶蛋白质与非抗体蛋白质各自的能量。非抗体蛋白质按最低结合能到最高结合能的顺序排列。

[0046] 步骤 (c) 是从所选择的非抗体蛋白质中选择具有理想结合能的非抗体蛋白质的步骤。当结合能较低时，靶标与非抗体蛋白质之间较易结合，发生更紧密和更特异性结合的可能性增加。在这方面，最好预先确定所需数量，最好是只排列与预定数量一样多的非抗体蛋白质。

[0047] 步骤 (d) 是从所选择的非抗体蛋白质与靶蛋白质的界面氨基酸残基中选择具有高结合能的氨基酸残基的步骤，计算每个界面氨基酸残基的结合能以检查各残基的能量贡献。选择具有不理想结合能的残基用于进一步的随机化。为便于随机化，可通过与排名靠前的支架对比，将所含的不理想残基数量较多的支架排除。根据本发明的一个实施例，从选定的非抗体蛋白质 -EGFR 结构域 2 对接复合物中选择了 1OZJ 作为结构互补性高和结合能理想的支架，并计算每一个参与所选定的 1OZJ 的结合过程的残基的结合能。结果发现了 44Lys、107Glu、110Glu、114Asn 和 118Asp 这 5 个具有高结合能的界面残基，将其选定为不理想残基 (图 4c，表 2)

[0048] 步骤 (e) 是用具有低结合能的氨基酸残基替代步骤 (d) 中选定的不理想氨基酸残基的步骤，该步骤通过定向进化来进行。即，靶蛋白质和非抗体蛋白质的不理想界面残基具有阻碍结合的高结合能，所以用具有低结合能的氨基酸残基替换这些不理想的界面残基，从而达到预期的结合。氨基酸残基的替换过程可通过已知的分子生物学方法 (但不限于此) 进行，该方法优选为噬菌体展示技术和生物筛选技术。根据本发明的一个实施例，算出了参与所选支架 1OZJ 结合过程的各残基的结合能。为替换所发现的具有高结合能的五个氨基酸残基 44Lys、107Glu、110Glu、114Asn 和 118Asp，构建了噬菌体展示随机文库，然后进行 5 轮淘洗，以便根据噬菌体 ELISA 选出比野生型 1OZJ 支架更强的克隆 (图 5)。

[0049] 本发明制备了人类支架文库，所述文库由 1261 个分子量为 10-40kDa、所有已知人类蛋白质中随机结合风险低的非抗体可溶性蛋白质组成。然后，为筛选与 EGFR 结构域 2 互补的支架，使用 PatchDock 程序进行大规模的对接模拟操作，从而选择排名前 10 的对接模型。在这方面，当 EGFR 结构域 2 的 28 个残基中有 10 个或 10 个以上参与了该结合过程，这些支架就视为稳定结构。结果找到了具有与 EGFR 结构域 2 的靶结合位点互补的形状的蛋白质支架。再用 EGAD 计算该选定支架的各残基的能量贡献值以便选择具有最理想的复合物形成能的支架 1OZJ 和 5 个具有不理想结合能的界面残基 (表 1 和图 4)。之后，通过构建噬菌体展示文库并进行生物筛选，将这 5 个残基替换为具有理想结合能的残基，然后利用噬菌体 ELISA 检查与 EGFR 的结合情况来选出 15 个更强的克隆 (表 2 和图 5)。这些克隆中，挑出最强的克隆 6、7 和 9，然后进行序列分析以制备靶特异性强、富含利用替换方式获得的目的序列的非抗体蛋白质 (图 7-9)。

[0050] 本发明的另一实施例涉及特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质。

[0051] 所述特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质与上述的蛋白质一样，优选地，可通过所述的用于制备靶特异性非抗体蛋白质的方法制备。优选地，该特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质可以具有如 SEQ ID NO.3、4 或 5 所示的氨基酸序列。根据本发明的一个实施例，对选作特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质的 1OZJ 进行噬菌体 ELISA，所得的克隆 6、7 和 9 与 EGFR 的结合力增强，对这些克隆进行序列分析。分析结果显示，它们分别具有如 SEQ ID NO.3、4 和 5 所示的序列 (表 2，图 5、7-9)

[0052] 本发明的另一实施例涉及编码特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质的核酸。优选地，该核酸可具有如 SEQ ID NO.7、8 或 9 所示的核苷酸序列。在本发明的一个实施例中，对选作特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质的 1OZJ 进行噬菌体 ELISA，所得的克隆 6、7 和 9 与 EGFR 的结合力增强，对这些克隆进行序列分析。分析结果显示，它们分别具有如 SEQ ID NO.7、8 和 9 所示的序列 (表 2，图 5、7-9)。

[0053] 本发明的另一实施例涉及包含编码特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质的核酸的载体。本发明的载体例如包括但不限于质粒载体、柯斯载体、噬菌体载体和病毒载体。合适的表达载体例如包括表达调控元件，如启动子、操纵子、起始密码子、终止密码子、多聚腺苷化信号和增强子，以及膜靶向或分泌信号序列，或先导序列。该载体的启动子可以是组成型或诱导型的。该表达载体还包括选择标记，用于选择含有该载体的宿主细胞。

[0054] 本发明的另一实施例涉及一种转化子，所述转化子是用包含编码特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质的核酸的载体转化的。

[0055] 转化方法包括将核酸导入有机体、细胞、组织或器官的任一方法，可运用本领域已知的对应宿主细胞的优选标准技术进行。该转化方法例如包括但不限于电穿孔法、原生质体融合法、磷酸钙 ($CaPO_4$) 沉淀法、氯化钙 ($CaCl_2$) 沉淀法、碳化硅纤维震荡法、农杆菌介导转化法、PEG 法、硫酸葡聚糖法、脂质体法。依宿主细胞不同，表达和修饰量不同，根据其目的选择和使用最优的宿主细胞。宿主细胞例如包括但不限于原核细胞，如大肠杆菌 (Escherichia coli)、枯草杆菌 (Bacillus subtilis)、链霉菌 (Streptomyces)、假单胞菌 (Pseudomonas)、奇异变形杆菌 (Prometeus mirabilis) 和葡萄球菌 (Staphylococcus)。进一步，可以使用低等真核细胞，如真菌（如曲霉 (Aspergillis)）和酵母菌（如巴斯得毕赤酵母菌 (Pichia pastoris)、酿酒酵母菌 (Saccharomyces cerevisiae)、裂殖酵母菌 (Schizosaccharomyces)、链孢霉 (Neurospora crassa)），以及来源于高等真核细胞的细胞，如昆虫细胞、植物细胞和哺乳动物细胞作为宿主细胞。

[0056] 本发明的另一实施例涉及治疗或预防癌症的组合物，包括特异性结合 EGFR 结构域 2 的靶特异性非抗体蛋白质。

[0057] 本发明治疗或预防癌症的组合物适用 EGFR 活化引起的所有癌症类型，并且适用患有癌症的任何动物。所述动物包括但不限于家畜，如牛、猪、羊、马、狗和猫，以及人和灵长类动物。

[0058] 如本文所用，术语“预防”是指通过施用包含本发明的非抗体蛋白质的组合物使癌症发生受到阻碍或抑制的所有行为。术语“治疗”是指通过施用包含本发明的非抗体蛋白质的组合物使癌症往更好方面转变或有所改善的所有行为。如果将 EGFR 结构域 2- 特异性非抗体蛋白质用作治疗性组合物，则将其与现有治疗剂直接连接（如共价键）或通过接头等间接连接，以便以非抗体蛋白质 - 治疗剂复合物形式施用到体内，用于癌症的预防或治疗。要使用的治疗剂包括化学治疗剂、免疫治疗剂、细胞因子、趋化因子、抗病毒剂、生物制剂、酶抑制剂等。包括本发明的 EGFR 结构域 2 特异性非抗体蛋白质的癌症治疗剂可进一步包括药学上可接受的载体，可以与该载体一起制成制剂。

[0059] 如本文所用，术语“药学上可接受的载体”是指不会给生物活性和活性成分的性质造成损害的载体或稀释剂。对液态制剂来说，药学上可接受的载体应当无菌且适于生命体。例如，药学上可接受的载体可包括：生理盐水、无菌水、Ringer 溶液、缓冲生理盐水、白蛋白注射液、葡萄糖溶液、麦芽 - 糊精溶液、甘油、乙醇或以上一种或多种成分的组合。如果有必要，可以加入其它常用添加剂，如抗氧化剂、缓冲液、抑菌剂等。为了将本组合物制成注射剂，如水性溶液、混悬剂、乳剂、丸剂、胶囊剂、颗粒剂或片剂，还可进一步加入稀释剂、分散剂、表面活性剂、黏合剂或润滑剂。

[0060] 所述包含EGFR结构域2-特异性非抗体蛋白质和药学上可接受的载体的抗癌组合物适用于任何含有该抗癌组合物作为活性成分的制剂，可制成口服制剂或非经肠道给药制剂。本发明的药物制剂可包括那些适于口腔、直肠、局部（包括颊部和舌下）、透皮、阴道或非经肠道（包括肌内、皮下和静脉内）给药的制剂或适于吸入或吹入给药的形式。

[0061] 含有EGFR结构域2-特异性非抗体蛋白质作为活性成分的口服制剂包括片剂、锭剂、含片、水溶性或油性混悬剂、粉剂或颗粒剂、乳剂、硬胶囊或软胶囊剂、糖浆剂或酏剂等。为了制备片剂或胶囊剂型的制剂，所述组合物可进一步包括黏合剂如乳糖、蔗糖、山梨糖醇、甘露醇、淀粉、支链淀粉、纤维素或明胶；赋形剂如磷酸氢钙；崩解剂如玉米淀粉或土豆淀粉；以及润滑剂如硬脂酸镁、硬脂酸钙、硬脂富马酸钠或聚乙二醇蜡。对胶囊剂来说，除上述化合物外，可进一步使用液态载体如脂。

[0062] 含有本发明组合物作为活性成分的非经肠道给药制剂可制成用于皮下、静脉内或肌内途径的注射剂、栓剂，或经由呼吸道吸入的喷雾剂，如气雾剂。将本发明的组合物与稳定剂或缓冲剂一起溶解或混悬到水中，然后将该溶液或混悬液装到安瓿或西林瓶中即可制成注射制剂。栓剂通常由栓基基质如可可脂或其它甘油脂或者由治疗性通便剂制成。对于喷雾剂来说，如气雾剂，可以将用于喷射水分散乳剂或湿粉的抛射剂与添加剂一并使用。

[0063] 下面将结合以下实施例对本发明做更详细地描述。但是，这些实施例仅用作阐述目的，本发明并非要受这些实施例的限制。

[0064] 实施例1：构建结构确定的人类支架文库

[0065] 为了构建包含几乎每个已知人类蛋白质的人类蛋白质支架文库，从根据SCOP（结构分类蛋白质）数据库中人类蛋白质的3D蛋白质结构进行分类的各个组中调取5个有代表性的结构。为了利于穿透组织，选择5个或5个以内具有10-40kDa的有限分子量的蛋白质，以便包括人类蛋白质的所有类型并降低该文库的大小。然后，本发明人利用PDBTM数据库中的注释排除膜蛋白结构，同时通过关键字检索将抗体蛋白质排除。之后，为避免随机结合，只选择具有已知相互作用数量为10或10以内的蛋白质。用HPRD（人类蛋白质参考数据库）计算蛋白质-蛋白质相互作用的数量。根据SWISS-PROT数据库的“SUBNIT”部分，排除具有大量相互作用如同源四聚体和同源六聚体的同源多聚体，选择形成单体、同源二聚体和同源三聚体的蛋白质。

[0066] 最后构建了由1261个低分子量、随机结合风险低的非抗体可溶性蛋白质组成的人类蛋白质支架文库。

[0067] 实施例2：基于结构的虚拟筛选

[0068] 基于实施例1中构建的人类蛋白质支架文库，为了筛选具有与EGFR结构域2有互补形状的支架，通过大量的与EGFR的对接模拟程序来进行虚拟筛选（图1b）。该方法如下：蛋白质对接模拟程序的最初目的是要预测现有的蛋白质-蛋白质相互作用，但在本发明中用它来寻找两个蛋白质间的结构互补性，用于创造新的蛋白质-蛋白质相互作用。

[0069] 为筛选具有与EGFR结构域2有互补结构的支架，用PatchDock程序(Schneidman-Duhovny,D.et al.Proteins 52,107-112(2003),Schneidman-Duhovny,D.et al.Nucleic Acids Res 33,W363-367(2005))将人类EGFR胞外区域的A链(PDB id:1IVO,Ogiso H et.al,Cell.Sep 20;110(6):775-87,2002)与文库中每个支架之间进行对接模拟。然后，从支架-EGFR对接结果中产生排名前10的对接模型，分析这些模型的结合模式以

找出 EGFR 中参与复合物形成的残基，并寻找 EGFR 结构域 2 的对接结构。预先设定了 EGFR 结构域 2 中作为靶位点的 28 个界面残基。当 28 个界面残基中有 10 个或 10 个以上参与了复合物形成，就视为该对接复合物的支架是用于结合 EGFR 结构域 2（如 SEQ ID NO.1 所示的 EGFR 氨基酸序列中的 166-309 序列，图 2）的理想结构：229、230、239、242、244、245、246、248、249、250、251、252、253、262、263、264、265、275、278、279、280、282、283、284、285、286、303、304（如 SEQ ID NO.1 所示的 EGFR 氨基酸序列中的 28 个预定氨基酸序列，图 2）。

[0070] 在本发明中，两条链之间一经形成对接复合物后其 SASA（溶剂可及表面积）的变化大于1Å^{2}的残基就视为是那些参与了该复合物形成的残基。使用 Naccess（http://www.bioinf.manchester.ac.kr）计算每个残基的 SASA，从而找出参与该复合物形成的特定残基。

[0071] 结果找到了与 EGFR 结构域 2 的靶结合位点有互补形状的蛋白质支架。

[0072] 实施例 3：计算复合物形成能并选择能量理想的蛋白质支架

[0073] 从选定的涉及 EGFR 结构域 2 结合的支架 -EGFR 对接复合物中，筛出若干表现出最理想的复合物形成能的支架供进一步精筛（图 1c）。

[0074] 用 EGAD（http://egad.berkeley.edu/EGAD_manual/index.html）算出支架 -EGFR 对接复合物的复合物形成能和每个支架界面残基的能量贡献值。将 PatchDock 产生的对接结构用作模板，在 JOBTYPE 字段中输入 complex_formation_energy 命令，来计算结合能。其它所有选项均使用缺省值。最后得到了 pseudo_DELTA_G 复合物形成值，将其作为复合物形成能，然后收集显示为 level_1 的界面残基的 dG 值来选择不理想的残基。这时候排除间接参与结合的显示为 level_2 的残基。这个过程当中，除了考虑能量是否理想外，还考虑了另外两个标准。第一，通过人工检查将不具有合适结构的支架滤除。第二，因便于构建随机序列文库用于定向进化，优先考虑一级序列中能很好成簇的不理想残基的数量较少的支架。

[0075] 用下面的方程式 1 计算结合能。

[0076] [方程式 1]

[0077] $\Delta G_{结合} = \Delta G_{复合物} - \Delta G_{自由}$

[0078] 结合能（$G_{结合}$）是靶蛋白质与非抗体蛋白质之间结合所需的能量，计算靶蛋白质与非抗体蛋白质的复合物的能量（$G_{复合物}$）与结合前靶蛋白质与非抗体蛋白质各自的能量（$G_{自由}$）之间的能量差以确定复合物形成的稳定性。

[0079] 结果，选出 1OZJ（Smad3-MH1 的晶体结构）的 B 链（图 4a）作为具有最理想复合物形成能的支架（表 1）。利用对接模型算出 1OZJ 与 EGFR 之间的复合物形成能（图 4b）。该支架用深色表示，其结构以同样方式在图 4a 中示出。1OZJ 与 EGFR 的界面残基用三维方式示出。

[0080] 表 1

[0081]

PDB id	链	PatchDock排名	ddG(kcal/mol)
2A5E	-	9	36.86
1KZY	D	4	62.75
1ZOP	B	8	76.26
1OZJ	**B**	**9**	**84.69**
4FAP	B	1	86.45
1AUT	L	9	88.62
1RK9	A	10	94.61
1XMJ	A	4	110.12
1SI3	A	9	111.94
1SI3	A	1	114.67

[0082] 然后，检查所选支架的各个界面残基的能量贡献值，选出不利于形成复合物的氨基酸残基用于序列随机化。图 4c 显示了 1OZJ 的各界面残基的能量贡献值。

[0083] 结果，选出了 5 个用于序列随机化的 1OZJ 残基，用斜纹条表示 (44Lys、107Glu、110Glu、114Asn、118Asp，图 4c)。

[0084] 实施例 4 ：构建随机文库

[0085] 用 Genscript(Piscataway，NJ) 合成实施 3 中选出的 EGFR- 结合支架 (1OZJ) 的 DNA(SEQ ID NO.6，图 6)，使用设计用于特定结合区随机化的 NNK 引物来进行随机化。用 SfiI(Roche，Indianapolis，IN) 酶切随机化后的支架，连接到噬粒载体 pComb3X 中，然后将其转化到电转感受态细胞 ER2738(New England Biolabs，Beverly，MA) 中构建随机文库。

[0086] 最后通过构建随机文库将所选的 1OZJ 的 5 条不理想的界面残基完全随机化。所选界面残基总共产生 4×10^8 种组合。然后根据以下噬菌体展示技术从突变子中筛选强结合蛋白质。

[0087] 实施例 5 ：生物筛选、噬菌体展示和噬菌体 ELISA

[0088] 5.1 生物筛选和噬菌体展示

[0089] 为富集 EGFR 特异性结合蛋白，进行 5 轮作为亲和力选择技术的生物筛选操作。

[0090] 本发明人按照制造商的操作说明将 Dynabead M-270 Epoxy(Dynal，Invitrogen，Carlsbad，CA) 包被并封闭。用 1.5μg EGFR(Sigma，St.Louis，MO) 在 PBS(137mM 氯化钠、10mM 磷酸、2.7mM 氯化钾，pH 7.4) 中包被 5X 10^6 个 Dynabeads 磁珠，每轮筛选用 PBS 作母液。将 EGFR 包被的 Dynabeads 磁珠与 500μl 随机化的噬菌体文库一起在室温下旋转培养 2 小时。为除去未结合的噬菌体，在第一轮筛选步骤后，用含 0.05% Tween 20 的 PBS(v/v) 洗涤 Dynabeads 磁珠一次，第二轮和第三轮筛选步骤后洗涤三次，余下的两轮筛选结束后洗涤五次。与 EGFR 包被的 Dynabeads 磁珠结合的噬菌体用 30μl、0.1M 甘氨酸 -HCl(pH2.2) 洗脱两次，然后加 2M Tris-HCl(pH9.1) 中和。用洗脱后的噬菌体转染新鲜培养的 ER2738，培养过夜用于下一步的生物筛选步骤。

[0091] 5.2 噬菌体 ELISA

[0092] 从最后一轮的产出滴定板上随机挑选单个的克隆，在 1ml 超级培养液中温育 (3%

胰蛋白胨、2%酵母提取物、1% 3-[N-吗啉基]丙磺酸[MOPS]，pH7.0)。37℃温育过夜后，用上清液进行噬菌体ELISA(酶联免疫吸附试验)。用4μg/ml含EGFR的PBS包被微量滴定板孔，4℃过夜，然后用含3% BSA的PBS37℃封闭1小时。培养含噬菌体上清的平板，37℃培育2小时，之后用含0.05% Tween-20的PBS(PBST)洗涤三次。然后加入经封闭缓冲液(1 ： 2000)稀释的HRP(辣根过氧化物酶)偶联的抗M13抗体(Sigma)37℃培养1小时。往每个孔内加50μlTMB基质溶液(Pierce，Rockford，IL)，在650nm测定OD。

[0093] 筛选后，发现来自1OZJ支架的克隆与EGFR的结合能力大大增强(图5)。野生型1OZJ未显示对EGFR有亲和力，但将实施例3中选择的5个氨基酸随机化后，15或更多个1OZJ突变克隆显示出与EGFR有很强的结合力(图5和表2)。用分子生物学方法分析最强克隆6、7和9的氨基酸和DNA序列。最后发现克隆6具有如SEQ ID NO.3所示的氨基酸序列和SEQ ID NO.7所示的DNA序列(图7)，发现克隆7具有如SEQ ID NO.4所示的氨基酸序列和SEQ ID NO.9所示的DNA序列(图8)，克隆9具有如SEQ ID NO.5所示的氨基酸序列和SEQ ID NO.8所示的DNA序列(图9)。

[0094] 表2

[0095]

	残基					
克隆	44	107	110	114	118	OD_{650nm}
WT	K	E	E	N	D	0.086
6	S	G	S	N	L	1.826
9	A	Q	I	L	P	1.748
7	L	A	A	M	C	1.741
8	G	P	N	T	P	1.18
54	C	I	D	P	R	1.156
24	S	A	D	V	E	1.144
21	T	L	A	A	M	1.089
35	T	L	A	A	M	1.041
20	S	T	T	Q	G	1.028
53	A	S	S	S	I	1.02
34	S	T	T	Q	G	0.987
55	I	K	R	P	D	0.964
31	G	H	P	K	W	0.959
18	S	A	D	V	E	0.927
43	R	H	N	G	A	0.919

[0096] 表2和图5的结果显示支架蛋白质和EGFR之间真正地产生了新的结合，它们之间除具有互补结构和理想的结合能之外无其他关联。此外，结果表明用本发明的虚拟筛选和生物筛选步骤制备靶特异性非抗体蛋白质的方法能用于制备优化后对预定靶位点有很强的结合力和特异性的结合蛋白质。

[0097] 本发明的效果

[0098] 本发明制备靶特异性非抗体蛋白质的方法能用来提供平台技术，开发适用于各种还没有被现有抗体药物、蛋白质药物、适体等涉及的靶标新型生物药。所以，本发明的方法能用来开发连续的病人定制和靶向治疗，预期会有巨大的社会和经济影响。另外，本发明的

抑制EGFR结构域2的靶特异性蛋白质特异性地抑制作为癌症治疗靶的EGFR活性来选择性地攻击正常细胞外的癌细胞，从而提供了有效的癌症治疗效果。

[0001]

序列表

<110> 浦项工科大学校产学协力团

<120> 靶特异性非抗体蛋白质及其制备方法

<130> OPA10039/PCT

<150> US61/169,636
<151> 2009-04-15

<160> 9

<170> KopatentIn 1.71

<210> 1
<211> 420
<212> PRT
<213> 智人 (Homo sapiens)

<220>
<221> DOMAIN
<222> (166)..(309)
<223> EGFR 的结构域2

<400> 1

```
Met Glu Glu Lys Lys Val Cys Gln Gly Thr Ser Asn Lys Leu Thr Gln
  1               5                  10                  15
Leu Gly Thr Phe Glu Asp His Phe Leu Ser Leu Gln Arg Met Phe Asn
             20                  25                  30
Asn Cys Glu Val Val Leu Gly Asn Leu Glu Ile Thr Tyr Val Gln Arg
         35                  40                  45
Asn Tyr Asp Leu Ser Phe Leu Lys Thr Ile Gln Glu Val Ala Gly Tyr
     50                  55                  60
Val Leu Ile Ala Leu Asn Thr Val Glu Arg Ile Pro Leu Glu Asn Leu
 65                  70                  75                  80
Gln Ile Ile Arg Gly Asn Met Tyr Tyr Glu Asn Ser Tyr Ala Leu Ala
                 85                  90                  95
Val Leu Ser Asn Tyr Asp Ala Asn Lys Thr Gly Leu Lys Glu Leu Pro
            100                 105                 110
Met Arg Asn Leu Gln Glu Ile Leu His Gly Ala Val Arg Phe Ser Asn
        115                 120                 125
Asn Pro Ala Leu Cys Asn Val Glu Ser Ile Gln Trp Arg Asp Ile Val
    130                 135                 140
Ser Ser Asp Phe Leu Ser Asn Met Ser Met Asp Phe Gln Asn His Leu
145                 150                 155                 160
Gly Ser Cys Gln Lys Cys Asp Pro Ser Cys Pro Asn Gly Ser Cys Trp
                165                 170                 175
Gly Ala Gly Glu Glu Asn Cys Gln Lys Leu Thr Lys Ile Ile Cys Ala
            180                 185                 190
Gln Gln Cys Ser Gly Arg Cys Arg Gly Lys Ser Pro Ser Asp Cys Cys
        195                 200                 205
His Asn Gln Cys Ala Ala Gly Cys Thr Gly Pro Arg Glu Ser Asp Cys
```

[0002]

210 215 220
Leu Val Cys Arg Lys Phe Arg Asp Glu Ala Thr Cys Lys Asp Thr Cys
225 230 235 240
Pro Pro Leu Met Leu Tyr Asn Pro Thr Thr Tyr Gln Met Asp Val Asn
245 250 255
Pro Glu Gly Lys Tyr Ser Phe Gly Ala Thr Cys Val Lys Lys Cys Pro
260 265 270
Arg Asn Tyr Val Val Thr Asp His Gly Ser Cys Val Arg Ala Cys Gly
275 280 285
Ala Asp Ser Tyr Glu Met Glu Glu Asp Gly Val Arg Lys Cys Lys Lys
290 295 300
Cys Glu Gly Pro Cys Arg Lys Val Cys Asn Gly Ile Gly Ile Gly Glu
305 310 315 320
Phe Lys Asp Ser Leu Ser Ile Asn Ala Thr Asn Ile Lys His Phe Lys
325 330 335
Asn Cys Thr Ser Ile Ser Gly Asp Leu His Ile Leu Pro Val Ala Phe
340 345 350
Arg Gly Asp Ser Phe Thr His Thr Pro Pro Leu Asp Pro Gln Glu Leu
355 360 365
Asp Ile Leu Lys Thr Val Lys Glu Ile Thr Gly Phe Leu Leu Ile Gln
370 375 380
Ala Trp Pro Glu Asn Arg Thr Asp Leu His Ala Phe Glu Asn Leu Glu
385 390 395 400
Ile Ile Arg Gly Arg Thr Lys Gln His Gly Gln Phe Ser Leu Ala Val
405 410 415
Val Ser Leu Asn
420

<210> 2
<211> 144
<212> PRT
<213> 智人 (Homo sapiens)

<400> 2
Met Ser Ser Ile Leu Pro Phe Thr Pro Pro Ile Val Lys Arg Leu Leu
1 5 10 15
Gly Trp Lys Lys Gly Glu Gln Asn Gly Gln Glu Glu Lys Trp Cys Glu
20 25 30
Lys Ala Val Lys Ser Leu Val Lys Lys Leu Lys Lys Thr Gly Gln Leu
35 40 45
Asp Glu Leu Glu Lys Ala Ile Thr Thr Gln Asn Val Asn Thr Lys Cys
50 55 60
Ile Thr Ile Pro Arg Ser Leu Asp Gly Arg Leu Gln Val Ser His Arg
65 70 75 80
Lys Gly Leu Pro His Val Ile Tyr Cys Arg Leu Trp Arg Trp Pro Asp
85 90 95
Leu His Ser His His Glu Leu Arg Ala Met Glu Leu Cys Glu Phe Ala
100 105 110
Phe Asn Met Lys Lys Asp Glu Val Cys Val Asn Pro Tyr His Tyr Gln
115 120 125
Arg Val Glu Thr Pro Val Leu Pro Pro Val Leu Val Pro Arg His Thr
130 135 140

[0003]

<210> 3
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> 10ZJ克隆6

<400> 3
Met Ser Ser Ile Leu Pro Phe Thr Pro Pro Ile Val Lys Arg Leu Leu
1 5 10 15
Gly Trp Lys Lys Gly Glu Gln Asn Gly Gln Glu Glu Lys Trp Cys Glu
20 25 30
Lys Ala Val Lys Ser Leu Val Lys Lys Leu Lys Ser Thr Gly Gln Leu
35 40 45
Asp Glu Leu Glu Lys Ala Ile Thr Thr Gln Asn Val Asn Thr Lys Cys
50 55 60
Ile Thr Ile Pro Arg Ser Leu Asp Gly Arg Leu Gln Val Ser His Arg
65 70 75 80
Lys Gly Leu Pro His Val Ile Tyr Cys Arg Leu Trp Arg Trp Pro Asp
85 90 95
Leu His Ser His His Glu Leu Arg Ala Met Gly Leu Cys Ser Phe Ala
100 105 110
Phe Asn Met Lys Lys Leu Glu Val Cys Val Asn Pro Tyr His Tyr Gln
115 120 125
Arg Val Glu Thr Pro Val Leu Pro Pro Val Leu Val Pro Arg His Thr
130 135 140

<210> 4
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> 10ZJ 克隆7

<400> 4
Met Ser Ser Ile Leu Pro Phe Thr Pro Pro Ile Val Lys Arg Leu Leu
1 5 10 15
Gly Trp Lys Lys Gly Glu Gln Asn Gly Gln Glu Glu Lys Trp Cys Glu
20 25 30
Lys Ala Val Lys Ser Leu Val Lys Lys Leu Lys Leu Thr Gly Gln Leu
35 40 45
Asp Glu Leu Glu Lys Ala Ile Thr Thr Gln Asn Val Asn Thr Lys Cys
50 55 60
Ile Thr Ile Pro Arg Ser Leu Asp Gly Arg Leu Gln Val Ser His Arg
65 70 75 80
Lys Gly Leu Pro His Val Ile Tyr Cys Arg Leu Trp Arg Trp Pro Asp
85 90 95
Leu His Ser His His Glu Leu Arg Ala Met Ala Leu Cys Ala Phe Ala
100 105 110

[0004]

Phe Met Met Lys Lys Cys Glu Val Cys Val Asn Pro Tyr His Tyr Gln
115 120 125
Arg Val Glu Thr Pro Val Leu Pro Pro Val Leu Val Pro Arg His Thr
130 135 140

<210> 5
<211> 144
<212> PRT
<213> 人工序列

<220>
<223> 10ZJ 克隆9

<400> 5
Met Ser Ser Ile Leu Pro Phe Thr Pro Pro Ile Val Lys Arg Leu Leu
1 5 10 15
Gly Trp Lys Lys Gly Glu Gln Asn Gly Gln Glu Glu Lys Trp Cys Glu
20 25 30
Lys Ala Val Lys Ser Leu Val Lys Lys Leu Lys Ala Thr Gly Gln Leu
35 40 45
Asp Glu Leu Glu Lys Ala Ile Thr Thr Gln Asn Val Asn Thr Lys Cys
50 55 60
Ile Thr Ile Pro Arg Ser Leu Asp Gly Arg Leu Gln Val Ser His Arg
65 70 75 80
Lys Gly Leu Pro His Val Ile Tyr Cys Arg Leu Trp Arg Trp Pro Asp
85 90 95
Leu His Ser His His Glu Leu Arg Ala Met Gln Leu Cys Ile Phe Ala
100 105 110
Phe Leu Met Lys Lys Pro Glu Val Cys Val Asn Pro Tyr His Tyr Gln
115 120 125
Arg Val Glu Thr Pro Val Leu Pro Pro Val Leu Val Pro Arg His Thr
130 135 140

<210> 6
<211> 432
<212> DNA
<213> 智人（Homo sapiens）

<400> 6
atgtcgtcca tcctgccttt cactcccccg atcgtgaagc gcctgctggg ctggaagaag 60
ggcgagcaga acgggcagga ggagaaatgg tgcgagaagg cggtcaagag cctggtcaag 120
aaactcaaga agacggggca gctggacgag ctggagaagg ccatcaccac gcagaacgtc 180
aacaccaagt gcatcaccat ccccaggtcc ctggatggcc ggttgcaggt gtcccatcgg 240
aaggggctcc ctcatgtcat ctactgccgc ctgtggcgat ggccagacct gcacagccac 300
cacgagctac gggccatgga gctgtgtgag ttcgccttca atatgaagaa ggacgaggtc 360
tgcgtgaatc cctaccacta ccagagagta gagacaccag ttctacctcc tgtgttggtg 420
ccacgccaca ca 432

<210> 7
<211> 432
<212> DNA

[0005]

<213> 人工序列

<220>
<223> 10ZJ 克隆6

<400> 7
atgtcgtcca tcctgccttt cactcccccg atcgtgaagc gcctgctggg ctggaagaag 60
ggcgagcaga acgggcagga ggagaaatgg tgcgagaagg cggtcaagag cctggtcaag 120
aaactcaaga gcacggggca gctggacgag ctggagaagg ccatcaccac gcagaacgtc 180
aacaccaagt gcatcaccat ccccaggtcc ctggatggcc ggttgcaggt gtcccatcgg 240
aaggggctcc ctcatgtcat ctactgccgc ctgtggcgat ggccagacct gcacagccac 300
cacgagctac gggccatggg tctgtgttcc ttcgccttca atatgaagaa gttggaggtc 360
tgcgtgaatc cctaccacta ccagagagta gagacaccag ttctacctcc tgtgttggtg 420
ccacgccaca ca 432

<210> 8
<211> 432
<212> DNA
<213> 人工序列

<220>
<223> 10ZJ 克隆7

<400> 8
atgtcgtcca tcctgccttt cactcccccg atcgtgaagc gcctgctggg ctggaagaag 60
ggcgagcaga acgggcagga ggagaaatgg tgcgagaagg cggtcaagag cctggtcaag 120
aaactcaagc ttacggggca gctggacgag ctggagaagg ccatcaccac gcagaacgtc 180
aacaccaagt gcatcaccat ccccaggtcc ctggatggcc ggttgcaggt gtcccatcgg 240
aaggggctcc ctcatgtcat ctactgccgc ctgtggcgat ggccagacct gcacagccac 300
cacgagctac gggccatggg tctgtgtgcg ttcgccttca tgatgaagaa gtgtgaggtc 360
tgcgtgaatc cctaccacta ccagagagta gagacaccag ttctacctcc tgtgttggtg 420
ccacgccaca ca 432

<210> 9
<211> 432
<212> DNA
<213> 人工序列

<220>
<223> 10ZJ 克隆9

<400> 9
atgtcgtcca tcctgccttt cactcccccg atcgtgaagc gcctgctggg ctggaagaag 60
ggcgagcaga acgggcagga ggagaaatgg tgcgagaagg cggtcaagag cctggtcaag 120
aaactcaagg cgacggggca gctggacgag ctggagaagg ccatcaccac gcagaacgtc 180
aacaccaagt gcatcaccat ccccaggtcc ctggatggcc ggttgcaggt gtcccatcgg 240
aaggggctcc ctcatgtcat ctactgccgc ctgtggcgat ggccagacct gcacagccac 300
cacgagctac gggccatgca gctgtgtatt ttcgccttcc tgatgaagaa gccggaggtc 360
tgcgtgaatc cctaccacta ccagagagta gagacaccag ttctacctcc tgtgttggtg 420
ccacgccaca ca 432

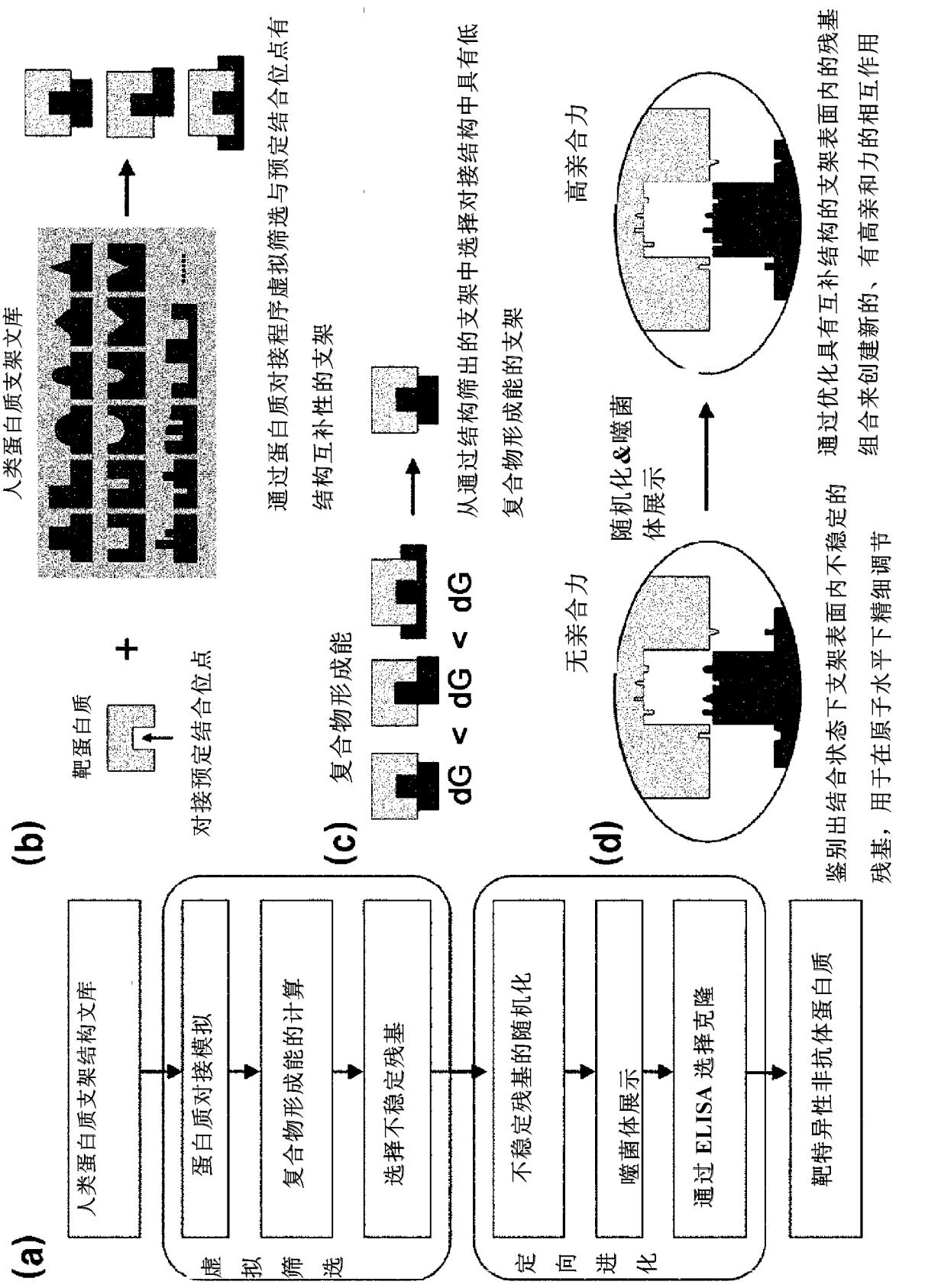

(a)
人类蛋白质支架结构文库
虚拟筛选
蛋白质对接模拟
复合物形成能的计算
选择不稳定残基
定向进化
不稳定残基的随机化
噬菌体展示
通过 ELISA 选择克隆
靶特异性非抗体蛋白质
(b)
人类蛋白质支架文库
靶蛋白质
对接预定结合位点
通过蛋白质对接程序虚拟筛选与预定结合位点有结构互补性的支架
(c)
复合物形成能
dG < dG < dG
从通过结构筛出的支架中选择对接结构中具有低复合物形成能的支架
(d)
无亲合力
随机化&噬菌体展示
高亲合力
鉴别出结合状态下支架表面内不稳定的残基，用于在原子水平下精细调节
通过优化具有互补结构的支架表面内的残基组合来创建新的、有高亲和力的相互作用

图 1

```
        10         20         30         40         50         60
MEEKKVCQGT SNKLTQLGTF EDHFLSLQRM FNNCEVVLGN LEITYVQRNY DLSFLKTIQE

        70         80         90        100        110        120
VAGYVLIALN TVERIPLENL QIIRGNMYYE NSYALAVLSN YDANKTGLKE LPMRNLQEIL

       130        140        150        160        170        180
HGAVRFSNNP ALCNVESIQW RDIVSSDFLS NMSMDFQNHL GSCQKCDPSC PNGSCWGAGE
```

下划线： 结构域 II （166-309）

```
       190        200        210        220        230        240
ENCQKLTKII CAQQCSGRCR GKSPSDCCHN QCAAGCTGPR ESDCLVCRKF RDEATCKDTC

       250        260        270        280        290        300
PPLMLYNPTT YQMDVNPEGK YSFGATCVKK CPRNYVVTDH GSCVRACGAD SYEMEEDGVR

       310        320        330        340        350        360
KCKKCEGPCR KVCNGIGIGE FKDSLSINAT NIKHFKNCTS ISGDLHILPV AFRGDSFTHT

       370        380        390        400        410        420
PPLDPQELDI LKTVKEITGF LLIQAWPENR TDLHAFENLE IIRGRTKQHG QFSLAVVSLN
.........
```

图 2

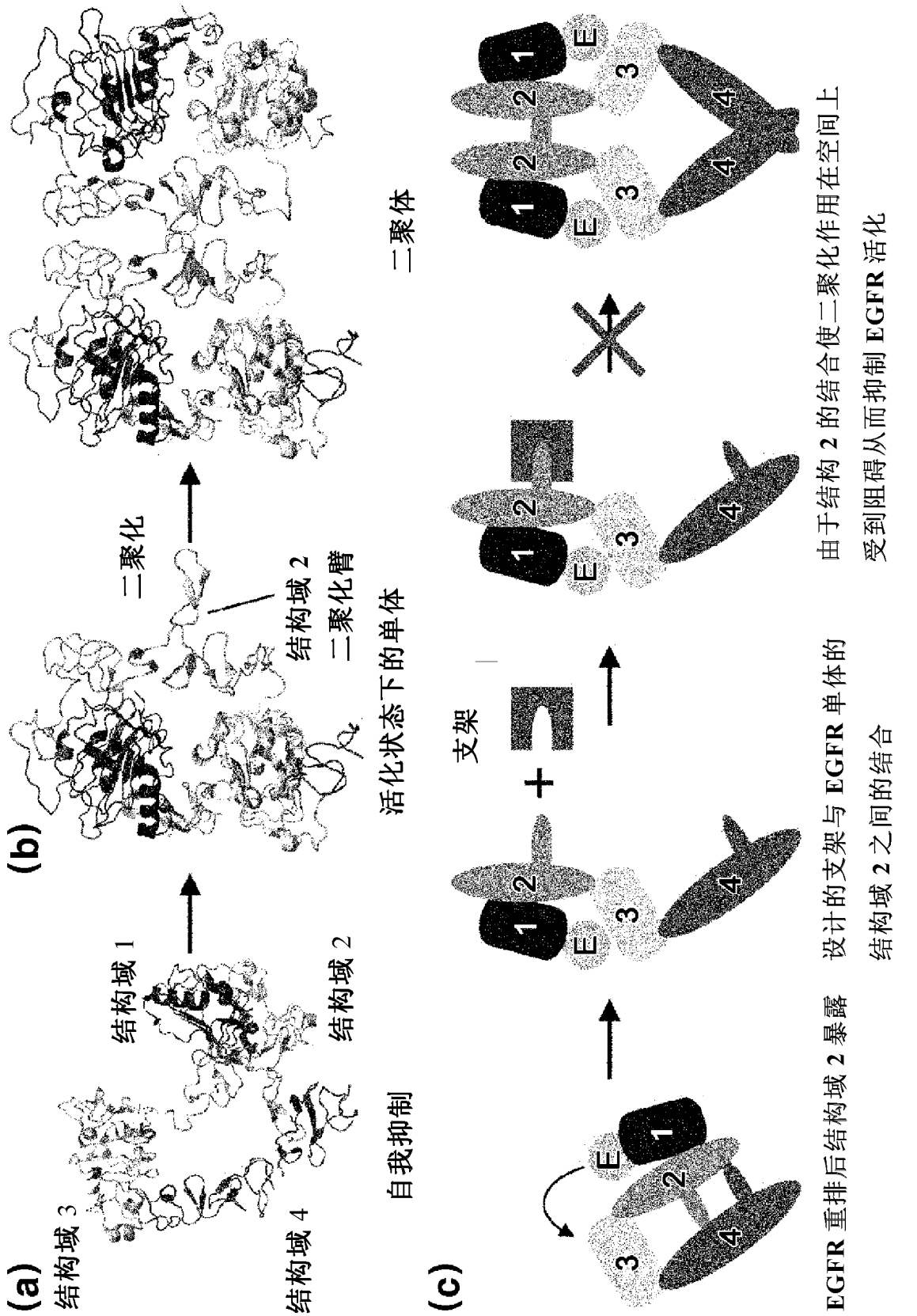

(a)
结构域 3
结构域 1
结构域 4
结构域 2
自我抑制
(b)
二聚化
结构域 2
二聚化臂
活化状态下的单体
二聚体
(c)
支架
EGFR 重排后结构域 2 暴露
设计的支架与 EGFR 单体的
结构域 2 之间的结合
由于结构 2 的结合使二聚化作用在空间上
受到阻碍从而抑制 EGFR 活化

图 3

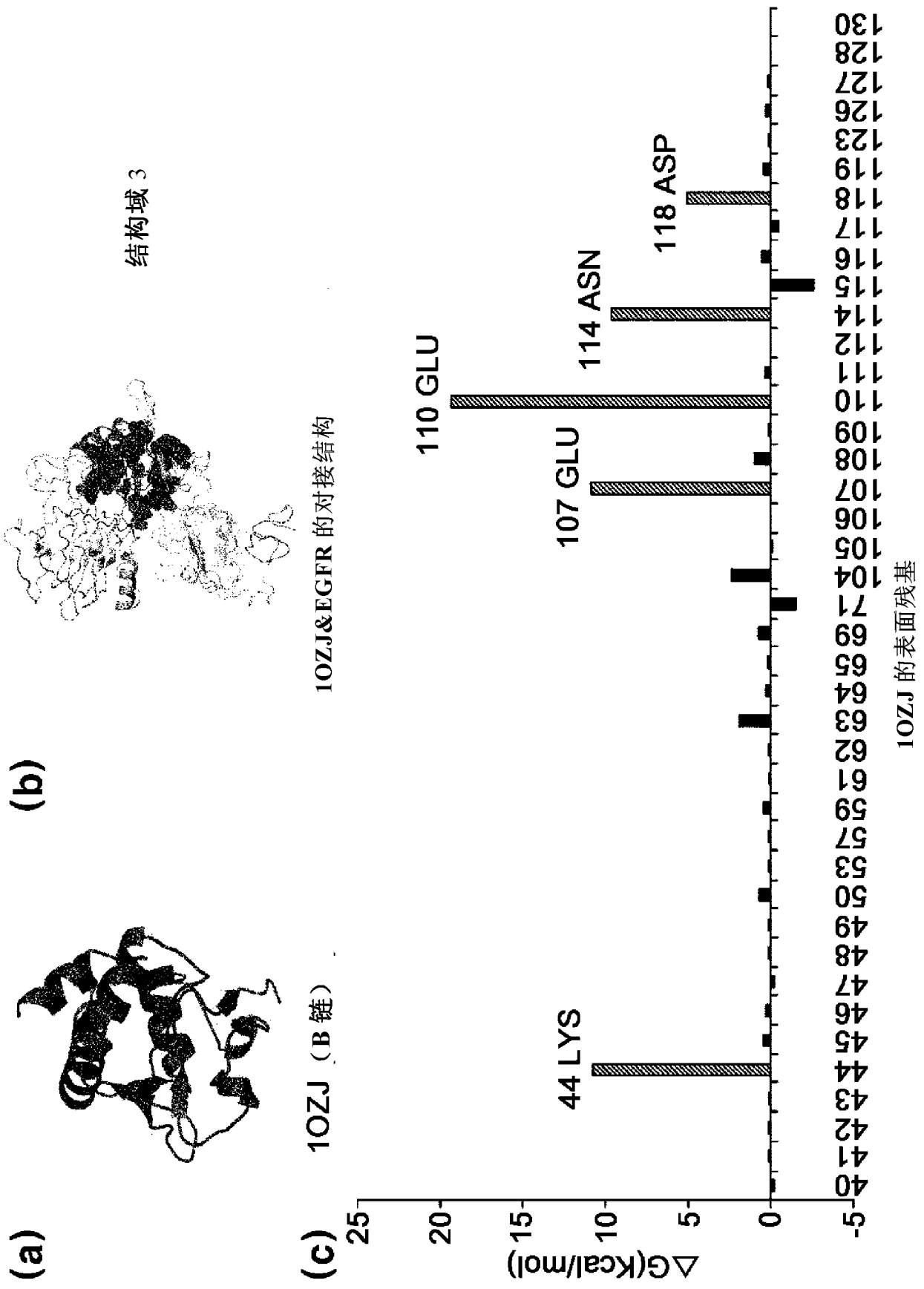

(a)
(b)
结构域 3
1OZJ&EGFR 的对接结构
(c)
1OZJ（B 链）
44 LYS
107 GLU
110 GLU
114 ASN
118 ASP
△G(Kcal/mol)
1OZJ 的表面残基

图 4

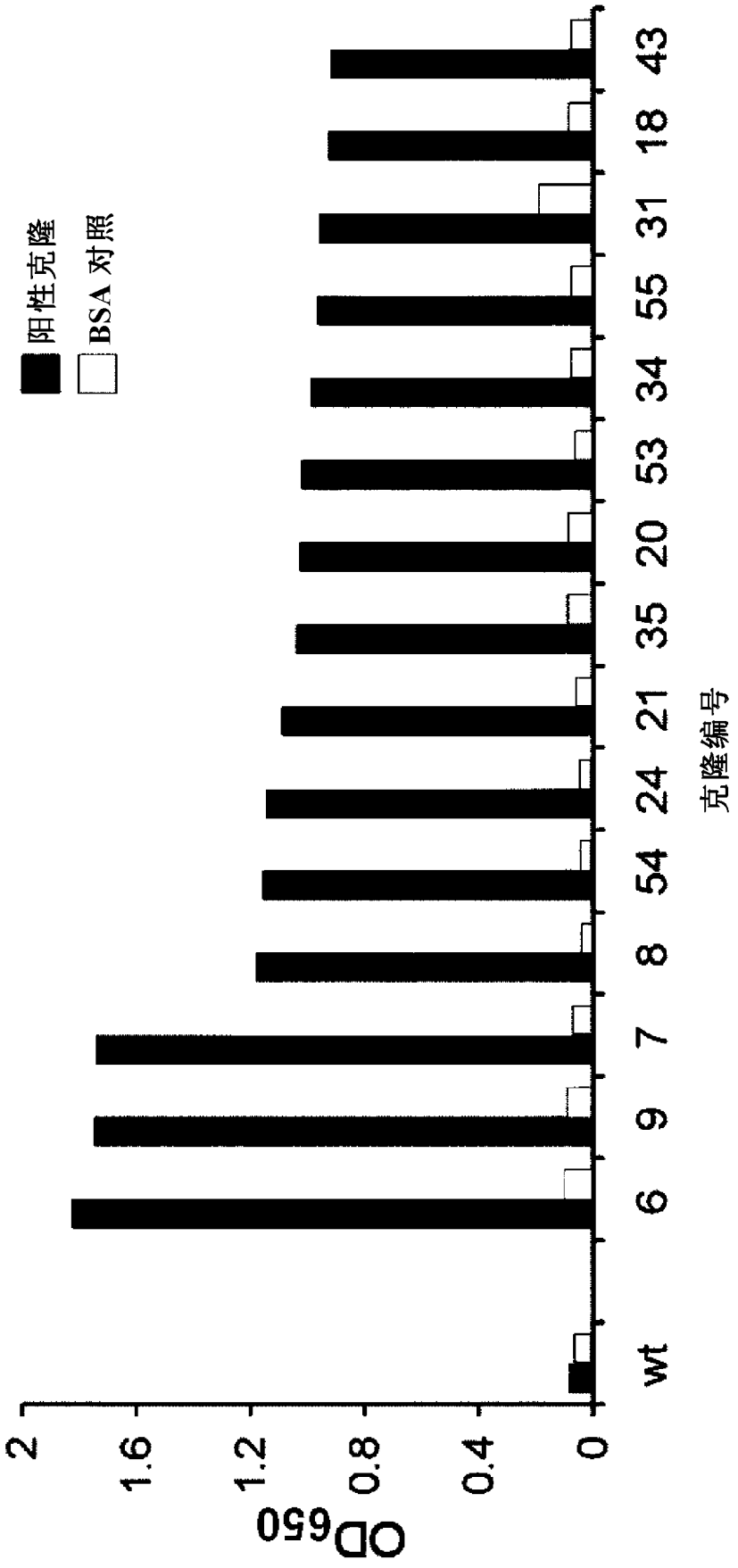

阳性克隆
BSA 对照
OD650
2
1.6
1.2
0.8
0.4
0
wt
6
9
7
8
54
24
21
35
20
53
34
55
31
18
43
克隆编号

图 5

```
MSSILPFTPP  IVKRLLGWKK  GEQNGQEEKW  CEKAVKSLVK  KLKKTGQLDE  LEKAITTQNV
NTKCITIPRS  LDGRLQVSHR  KGLPHVIYCR  LWRWPDLHSH  HELRAMELCE  FAFNMKKDEV
CVNPYHYQRV  ETPVLPPVLV  PRHT

ATGTCGTCCA  TCCTGCCTTT  CACTCCCCCG  ATCGTGAAGC  GCCTGCTGGG  CTGGAAGAAG
GGCGAGCAGA  ACGGGCAGGA  GGAGAAATGG  TGCGAGAAGG  CGGTCAAGAG  CCTGGTCAAG
AAACTCAAGA  AGACGGGGCA  GCTGGACGAG  CTGGAGAAGG  CCATCACCAC  GCAGAACGTC
AACACCAAGT  GCATCACCAT  CCCCAGGTCC  CTGGATGGCC  GGTTGCAGGT  GTCCCATCGG
AAGGGGCTCC  CTCATGTCAT  CTACTGCCGC  CTGTGGCGAT  GGCCAGACCT  GCACAGCCAC
CACGAGCTAC  GGGCCATGGA  GCTGTGTGAG  TTCGCCTTCA  ATATGAAGAA  GGACGAGGTC
TGCGTGAATC  CCTACCACTA  CCAGAGAGTA  GAGACACCAG  TTCTACCTCC  TGTGTTGGTG
CCACGCCACA  CA
```

图 6

```
MSSILPFTPP  IVKRLLGWKK  GEQNGQEEKW  CEKAVKSLVK  KLKSTGQLDE  LEKAITTQNV
NTKCITIPRS  LDGRLQVSHR  KGLPHVIYCR  LWRWPDLHSH  HELRAMGLCS  FAFNMKKLEV
CVNPYHYQRV  ETPVLPPVLV  PRHT

ATGTCGTCCA  TCCTGCCTTT  CACTCCCCCG  ATCGTGAAGC  GCCTGCTGGG  CTGGAAGAAG
CGCGAGCAGA  ACGGGCAGGA  CGAGAAATGG  TGCGAGAAGG  CGGTCAAGAG  CCTCGTCAAG
AAACTCAAGA  GCACGGGGCA  GCTGGACGAG  CTGGAGAAGG  CCATCACCAC  GCAGAACGTC
AACACCAAGT  GCATCACCAT  CCCCAGGTCC  CTGGATGGCC  GGTTGCAGGT  GTCCCATCGG
AAGGGGCTCC  CTCATGTCAT  CTACTGCCGC  CTGTGGCGAT  GGCCAGACCT  GCACAGCCAC
CACGAGCTAC  GGGCCATGGG  TCTGTGTTCC  TTCGCCTTCA  ATATGAAGAA  GTTGGAGGTC
TGCGTGAATC  CCTACCACTA  CCAGAGAGTA  GAGACACCAG  TTCTACCTCC  TGTGTTGGTG
CCACGCCACA  CA
```

图 7

```
MSSILPFTPP  IVKRLLGWKK  GEQNGQEEKW  CEKAVKSLVK  KLKLTGQLDE  LEKAITTQNV
NTKCITIPRS  LDGRLQVSHR  KGLPHVIYCR  LWRWPDLHSH  HELRAMALCA  FAFMMKKCEV
CVNPYHYQRV  ETPVLPPVLV  PRHT

ATGTCGTCCA  TCCTGCCTTT  CACTCCCCCG  ATCGTGAAGC  GCCTGCTGGG  CTGGAAGAAG
GGCGAGCAGA  ACGGGCAGGA  GGAGAAATGG  TGCGAGAAGG  CGGTCAAGAG  CCTGGTCAAG
AAACTCAAGC  TTACGGGGCA  GCTGGACGAG  CTGGAGAAGG  CCATCACCAC  GCAGAACGTC
AACACCAAGT  GCATCACCAT  CCCCAGGTCC  CTGGATGGCC  GGTTGCAGGT  GTCCCATCGG
AAGGGGCTCC  CTCATGTCAT  CTACTGCCGC  CTGTGGCGAT  GGCCAGACCT  GCACAGCCAC
CACGAGCTAC  GGGCCATGGG  TCTGTGTGCG  TTCGCCTTCA  TGATGAAGAA  GTGTGTTGGTG
TGCCGTGAATC  CCTACCACTA  CCAGAGAGTA  GAGACACCAG  TTCTACCTCC  TGTGTTGGTG
CCACGCCACA  CA

MSSILPFTPP  IVKRLLGWKK  GEQNGQEEKW  CEKAVKSLVK  KLKATGQLDE  LEKAITTQNV
NTKCITIPRS  LDGRLQVSHR  KGLPHVIYCR  LWRWPDLHSH  HELRAMQLCI  FAFLMKKPEV
CVNPYHYQRV  ETPVLPPVLV  PRHT

ATGTCGTCCA  TCCTGCCTTT  CACTCCCCCG  ATCGTGAAGC  GCCTGCTGGG  CTGGAAGAAG
GGCGAGCAGA  ACGGGCAGGA  GGAGAAATGG  TGCGAGAAGG  CGGTCAAGAG  CCTGGTCAAG
AAACTCAAGG  CGACGGGGCA  GCTGGACGAG  CTGGAGAAGG  CCATCACCAC  GCAGAACGTC
AACACCAAGT  GCATCACCAT  CCCCAGGTCC  CTGGATGGCC  GGTTGCAGGT  GTCCCATCGG
AAGGGGCTCC  CTCATGTCAT  CTACTGCCGC  CTGTGGCGAT  GGCCAGACCT  GCACAGCCAC
CACGAGCTAC  GGGCCATGCA  GCTGTGTATT  TTCGCCTTCC  TGATGAAGAA  GCCGGAGGTC
TGCCGTGAATC  CCTACCACTA  CCAGAGAGTA  GAGACACCAG  TTCTACCTCC  TGTGTTGGTG
CCACGCCACA  CA
```

图 8

图 9