

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G01N 33/532

C12Q 1/37 C12N 15/10

C07K 16/00

[12] 发明专利申请公开说明书

[21] 申请号 00805178. X

[43] 公开日 2002 年 4 月 10 日

[11] 公开号 CN 1344370A

[22] 申请日 2000.3.17 [21] 申请号 00805178. X

[30] 优先权

[32]1999.3.23 [33]GB [31]9906551.8

[32]1999.3.29 [33]GB [31]9907057.5

[32]1999.4.6 [33]GB [31]9907641.6

[32]1999.6.28 [33]GB [31]9914874.4

[32]1999.7.2 [33]GB [31]9915363.7

[32]1999.7.6 [33]GB [31]9915677.0

[32]1999.7.14 [33]GB [31]9916511.0

[32]1999.8.31 [33]GB [31]9920503.1

[32]1999.9.21 [33]GB [31]9922285.3

[86] 国际申请 PCT/GB00/01015 2000.3.17

[87] 国际公布 WO00/57183 英 2000.9.28

[85] 进入国家阶段日期 2001.9.18

[71] 申请人 拜奥维森有限公司

地址 英国阿伯丁

[72] 发明人 F·J·卡尔

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 刘 玥

权利要求书 5 页 说明书 49 页 附图页数 0 页

[54] 发明名称 蛋白质的分离和分析

[57] 摘要

提供用于蛋白鉴定和/或测序的新方法。这些方法特别适合于筛选 抗体文库,在优选实施方案中,利用质谱技术进行直接或间接测序。

I S S N 1 0 0 8 - 4 2 7 4

权 利 要 求 书

5 1. 一种蛋白鉴定、筛选和/或测序方法，包括提供一种个别蛋白文库，所述个别蛋白中的一种或更多种可以与目标靶结合，其中每个个别蛋白在其序列中包括一个“条形码”序列，所述序列可以用来在所述文库中鉴定每种个别蛋白。

2. 权利要求 1 中要求保护的方法，其中所述个别“条形码”序列由插入所述文库中所述个别蛋白的编码基因中的一个或多个核酸序列编码。

10 3. 权利要求 2 中要求保护的方法，其中一个或多个所述“条形码”序列邻接一个或多个内切蛋白酶消化的识别位点。

4. 权利要求 3 中要求保护的方法，其中所述识别位点是内切蛋白酶例如肠激酶或因子 Xa 的位点。

15 5. 权利要求 3 或权利要求 4 中要求保护的方法，其中使所述蛋白文库与一种或更多种靶部分例如靶蛋白接触/相连。

6. 权利要求 5 中要求保护的方法，其中所述蛋白和一种或更多种靶部分在溶液中结合。

20 7. 权利要求 5 或权利要求 6 中要求保护的方法，其中在结合后分离所述蛋白/靶部分复合物，然后用内切蛋白酶消化，以释放一种或种个所述“条形码”序列。

8. 权利要求 7 中要求保护的方法，其中一种或多种所述释放的“条形码”序列用来设计一种或更多种合成寡核苷酸例如引物，以回收或扩增一种或更多种与所述靶部分结合的蛋白的编码基因。

25 9. 权利要求 8 中要求保护的方法，其中用质谱法测定任何释放的“条形码”序列的质量，所述质量进而可以鉴定所释放的一种或更多种序列，或其中用质谱法直接测定任何释放的“条形码”的序列。

10. 权利要求 1-9 的任一项中要求保护的方法，其中所述蛋白文库是抗体文库。

11. 权利要求 10 中要求保护的方法, 其中所述蛋白文库是抗体结构域文库, 例如包括抗体可变区的重组抗体结构域例如 Fv 的文库。

12. 权利要求 11 中要求保护的方法, 其中所述文库包括由两条链(重链和轻链衍生链, VH 和 VL)组成的 Fv。

5 13. 权利要求 12 中要求保护的方法, 其中所述 VH 和 VL 链各自具有其自身的“条形码”序列。

14. 权利要求 10-13 的任一项中要求保护的方法, 其中所述“条形码”序列位于所述 Fv 序列的 C 末端。

15. 在权利要求 1-14 的任一项中限定的蛋白文库。

10 16. 一种筛选蛋白文库的方法, 包括根据一种或更多种所需特性筛选所述文库, 然后将其去复制, 以鉴定所述文库中具有所述所需特性的一种或更多种个别蛋白。

17. 权利要求 16 中要求保护的方法, 其中根据与靶部分的结合筛选所述文库。

15 18. 权利要求 17 中要求保护的方法, 其中通过质谱法、特别是基质辅助激光解附/电离飞行时间(MALDI-ToF)质谱检测结合。

19. 权利要求 16 中要求保护的方法, 其中根据特定生物活性筛选所述文库。

20 20. 权利要求 17 或权利要求 18 中要求保护的方法, 其中所述靶是复杂的混合物, 例如多种分子的混合物、全细胞或细胞膜。

21. 一种蛋白鉴定和/或测序方法, 包括提供一种个别蛋白的文库, 所述个别蛋白中的一种或更多种可以与目标靶结合, 其中每种个别蛋白与其基因一起结合于一个“结合部分”。

25 22. 权利要求 21 中要求保护的方法, 其中使所述蛋白文库在与所述“结合部分”接触之前或者之后, 与所述目标靶接触。

23. 权利要求 21 或 22 中要求保护的方法, 其中在根据与所述靶结合进行筛选后, 将所述文库去复制, 以鉴定具有所需特性的一种或更多种蛋白, 即与所述靶结合的蛋白。

24. 权利要求 21-23 的任一项中要求保护的方法，其中所述“结合部分”是一种粒子。

25. 权利要求 24 中要求保护的方法，其中所述粒子是胶乳珠粒。

5 26. 权利要求 21-23 的任一项中要求保护的方法，其中所述“结合部分”是一种蛋白或蛋白复合物。

27. 权利要求 26 中要求保护的方法，其中所述“结合部分”是抗生物素蛋白或链霉抗生物素蛋白，并且所述文库中的每种蛋白及其相关基因被生物素酰化。

10 28. 权利要求 21 或权利要求 22 中要求保护的方法，其中所述“结合部分”是能够与所述蛋白和基因两者均结合的双特异性结合分子。

29. 权利要求 21-23 的任一项中要求保护的方法，其中所述“结合部分”是活细胞或细胞的病毒，例如细菌或噬菌体。

30. 权利要求 21-29 的任一项中要求保护的方法，其中将改变所述文库中蛋白的特性的一种分子或其它分子结合于所述“结合部分”。

15 31. 权利要求 21-30 的任一项中要求保护的方法，其中将所述文库中的蛋白的编码基因在合成所述个别蛋白之前连接于所述“结合部分”。

32. 权利要求 21-31 的任一项中要求保护的方法，其中所述蛋白文库是抗体蛋白文库，例如抗体结构域诸如 Fv 的文库。

20 33. 一种蛋白鉴定和/或筛选方法，包括提供一种个别蛋白的文库，所述蛋白中的一种或更多种可以与目标靶结合，其中将每种个别蛋白连接于一种个别“编码部分”。

34. 权利要求 33 中要求保护的方法，其中所述“编码部分”是具有特有标识符“编码”的粒子。

25 35. 权利要求 34 中要求保护的方法，其中所述“编码”是可测量信号的不同比率，所述可测量信号例如为荧光标记、化学发光标记或放射性标记，或者是一种物理特征，例如特有标记。

36. 一种分析蛋白混合物的方法，包括：

- (iii) 消化或切割所述蛋白混合物;
- (iv) 将所产生的肽分级分离; 以及
- (v) 根据所产生的肽的质量和/或序列对所产生的肽进行分析。

5 37. 权利要求 36 中要求保护的方法, 其中所述步骤(ii)中的分级分离采用蛋白结合剂文库进行。

38. 权利要求 36 中要求保护的方法, 其中作为所述分级分离步骤的一部分, 对所述产生的肽进行物理分级分离和/或化学标记。

39. 权利要求 36 中要求保护的方法, 其中作为所述分级分离步骤的一部分, 对所述产生的肽进行一个或更多个氨基酸的添加。

10 40. 权利要求 37-39 的任一项中要求保护的方法, 其中所述蛋白结合剂文库是抗体或抗体片段的文库。

41. 权利要求 37-39 的任一项中要求保护的方法, 其中所述蛋白结合剂是主要组织相容性蛋白、 T 细胞受体和参与蛋白-蛋白结合相互作用的天然蛋白或蛋白结构域, 例如 SH1 结构域。

15 42. 权利要求 40 或权利要求 41 中要求保护的方法, 其中根据与得自受分析的蛋白混合物或相关蛋白混合物的一种或更多种蛋白或肽的结合, 对所述蛋白结合剂文库进行预选择。

43. 权利要求 42 中要求保护的方法, 其中所述蛋白混合物得自标准化重组基因文库。

20 44. 权利要求 36-43 的任一项中要求保护的方法, 其中最初在消化或切割之前, 或者通过所述 N 末端或 C 末端, 或者通过特定氨基酸, 或者通过氨基酸的特定序列, 将所述蛋白混合物结合于固相。

25 45. 权利要求 36-43 的任一项中要求保护的方法, 其中将在所述蛋白中发现的特定氨基酸或经修饰的氨基酸衍生化, 然后结合于固相, 这种结合发生在所述蛋白混合物的消化或切割之前或之后。

46. 权利要求 45 中要求保护的方法, 其中在与抗生物素蛋白或链霉抗生物素蛋白结合之前, 用生物素将所述特定的或经修饰的氨基酸衍生化。

47. 权利要求 45 中要求保护的方法, 其中在与配体特异性亲和试剂结合之前, 用配体将特定的或经修饰的氨基酸衍生化。

5 48. 权利要求 36-43 的任一项中要求保护的方法, 其中用修饰特异性亲和试剂将所述蛋白中发现的特定的天然修饰的氨基酸结合于固相, 这种结合发生在所述蛋白混合物的消化或切割之前或之后。

49. 权利要求 45-48 的任一项中要求保护的方法, 其中进行一个以上的消化/切割和衍生化的循环。

50. 权利要求 49 中要求保护的方法, 其中在每个消化或切割循环之后进行质量分析。

10 51. 权利要求 36-50 的任一项中要求保护的方法, 其中在用蛋白结合剂分级分离之前或之后, 采用物理方法例如 HPLC 对在消化/切割后释放的肽进行分级分离。

说明书

蛋白质的分离和分析

5 本发明涉及尤其是通过质量分析来分离和分析蛋白质。本发明尤其可应用于分离诸如抗体之类的结合蛋白。本发明也提供对蛋白质或蛋白质片段的修饰，以便有助于质量分析和或分离由基因文库成员编码的特定蛋白质。

10 关于蛋白质分离，本发明提供了借助与特定靶的结合而从这类蛋白质的复杂混合物中分离特定蛋白的新方法。具体地说，本发明提供借助于与特定靶抗原的结合而从基因文库衍生的这类抗体结构域混合物中分离特异性抗体结构域的方法。对于蛋白质的分析，本发明提供用于分析复杂的蛋白质混合物、尤其是用于比较两种或更多种不同样品之间的蛋白质的新方法。

15 对于借助于与特定靶的结合从复杂混合物中分离蛋白并且在该蛋白的身份或氨基酸序列预先未知的情况下，分离足够的与所述靶结合的蛋白以对该蛋白进行直接表征通常是非常困难的。为了从天然的、合成的或半合成的蛋白的大文库中选择目的蛋白，已经开发了“蛋白展示”法，用该方法产生在物理上与其基因相连接的重组蛋白，使得所述蛋白的回收使得随后可以快速回收所述基因。这类方法包括“体内”展示法诸如在噬菌体(“噬菌体展示”)、细菌和酵母上的展示，还包括“体外”展示法诸如在核糖体上的展示(“核糖体展示”)。可以对所回收的基因进行测序，以便确定所回收蛋白的身份，或可以用所回收的基因再生所述回收的蛋白。如果对基因文库进行蛋白展示
20 法，藉此根据特定特性例如与抗原结合(对于抗体可变区)选择蛋白，那么在每轮选择后，所回收的基因将富集编码显示这类特定特性的蛋白的基因。目前的“体内”展示法的缺点包括：对所展示的功能蛋白量的限制(噬菌体展示通常限于小于 40kDa 的多肽)，通常需要将重组
25

蛋白与宿主蛋白融合(这可能干扰重组蛋白的功能或结合), 以及不能改变每个展示粒子所展示的蛋白数; 后者也是“体外”展示法诸如核糖体展示的问题。另外, 因为展示粒子的体积小, 对于具有特定特性例如与抗原结合的蛋白的选择方法是有限的, 使得不容易使用诸如荧光激活细胞分类(FACS)的方法。因此, 仍需要发展新方法, 以改进从复杂混合物中分离蛋白、特别是改进从复杂的抗体可变区(Fv)混合物中分离 Fv。亦即本发明提供从复杂混合物中分离蛋白的改进方法。具体地说, 本发明将由基因文库产生的蛋白文库的应用与质谱法改进、特别是基质辅助激光解附/电离飞行时间(MALDI-ToF)质谱的改进和通过串联质谱(MS-MS)对 ToF 分离的肽直接测序的能力、最近将 ToF 和 MS/MS 结合进一个装置(Q-ToF)的能力以及将 HPLC 和电喷雾(ES)串联质谱结合的能力相结合。本发明也包括从复杂的蛋白混合物中筛选个别蛋白的新方法, 用所述方法不用将这些蛋白“展示”, 即在所述蛋白与所述靶结合期间或之后将这些蛋白与其相应的基因结合。本发明也包括从复杂的蛋白混合物中筛选个体蛋白的新方法, 用所述方法所述蛋白以及所述靶均不用“展示”, 即与任何其它分子或结构结合。本发明也包括从复杂的蛋白混合物中筛选个体蛋白的新方法, 用所述方法, 所述蛋白及其相应的基因通过加入或包括一个“结合部分(associating moiety)”连接在一起, 从而在加入所述“结合部分”或者之前或者之后所述蛋白与所述靶结合。

因此, 在第一方面, 本发明提供一种蛋白鉴定、筛选和/或测序的方法, 包括提供个别蛋白的文库, 所述个别蛋白中的一种或更多种可以与感兴趣的靶结合, 其中每种个别蛋白在其序列中包括一个“条形码”序列, 该序列可以用来在该文库中鉴定每种个别蛋白。

本发明的该方面可供用于蛋白、尤其是重组抗体结构域(例如 Fv)的文库, 从而该文库的各个蛋白成员在其氨基酸序列中包括一段序列(“条形码”), 可以随后对该段序列测序, 以便鉴别哪种蛋白与所述特定靶(或在 Fv 的情况下为“抗原”)结合。该实施方案将尤其可应用

于 Fv 衍生自人类基因的情况，从而所选择的 Fv 可能适合于人类治疗或诊断用途。在这种特定应用中，由免疫球蛋白 cDNA 库，例如衍生自人类外周血 B 细胞的库，或例如用在一个或多个位置具有半随机化的(“组合”) CDR (互补决定区) 的人可变区合成产生的库，产生广泛的 Fv 基因文库。如果构建该基因文库的方式使得在 Fv 编码区内或该区的末端包括一个随机(或半随机)基因序列，则这样一种随机/半随机基因序列将产生一种与个别 Fv 结合的随机/半随机肽序列。采用标准方法诸如寡核苷酸引发/DNA 聚合酶延伸或 PCR，构建这样一种随机/半随机基因序列，由此用一种随机/半随机合成寡核苷酸序列作为在 Fv 基因文库构建过程中用来扩增免疫球蛋白基因片段的一对引物之一。如果该 Fv 文库的成员包括两条链(即重链和轻链衍生链(VH 和 VL))，而不是单链(通过肽接头连接的 VH 和 VL)，那么可以将各个条形码与每条链连接(或可以仅与其中一条链连接)。当构建该文库时，所得到的 Fv 各包括一个或多个“肽条形码”，所述“肽条形码”对于该复杂文库内的特定 Fv 是特有的，或对于该复杂文库内的一小组 Fv 是特有的。最好是，所述肽条形码位于单链 Fv 区的 C 末端或者 VH 或 VL 或者这两者的 C 末端，并且包括在其本身和 Fv 区之间邻接的一个或多个蛋白酶敏感位点，例如肠激酶位点(在 Asp-Asp-Asp-Asp-Lys 之后切割)、因子 Xa 位点(在 Ile-Glu/Asp-Gly-Arg 之后切割)或其它内肽酶位点。如果由合适的基因文库产生这类 Fv 的混合物，那么将该混合物与靶抗原(或多种抗原，例如细胞上的抗原)混合，其中通常将所述抗原固定化。这导致特异性 Fv 与靶抗原结合，而非结合物(或弱结合物，取决于洗涤的严格性)被洗去。在洗去多余的抗体后，则通常通过采用用以切割所引入的蛋白酶敏感位点的内切蛋白酶消化，从所述 Fv 中释放出保留的抗原/Fv 复合物。然后将这种释放的条形码标记肽或者直接，或如有需要在借助使得所述肽被捕获到固相的特定氨基酸或氨基酸序列进行捕获后，进行质量分析/质谱测序，所述特定氨基酸或氨基酸序列例如为：可以被生物素化用以随后被捕获到固定化抗生物素

蛋白或链霉抗生物素蛋白上的半胱氨酸残基。另一方面，可以采用任何其它方法，以测定或者在 Fv 内或者释放后所述肽条形码的序列，包括用以序列特异性方式与所述条形码结合的特异性配体。在测定得自结合 Fv 的条形码序列(或部分序列)后，产生相应的合成寡核苷酸，并且用来从该文库中特异性地扩增或富集特定 Fv 基因。然后这些特定的或被富集的 Fv (或 VH 和 VL)基因被进一步用来产生相应的 Fv，然后或者单独地，或者作为所分离 Fv 小库的部分，再测试所述 Fv 的抗原结合。最后，用该方法，可以产生具有所需抗原结合特性的特定 Fv，并且如果所述 Fv 来自人类来源，则具有潜在的临床用途。该方面也包
5 括应用与个别蛋白或 Fv 结合的多个条形码，例如在 Fv 的 C 末端的两个相邻的条形码，由此通过蛋白酶消化，或者同时从每种 Fv 释放两种肽，以便增强对与所述靶结合的 Fv 的识别(identity)，或者顺序释放两种肽，从而在连接多轮消化中使用不同的蛋白酶，以提供一种不同的方式，以随后扩增对应于与该靶结合的 Fv 的 Fv 基因。该方面也包
10 括应用同时进行分析的多种条形码，以便增加全部条形码序列的多样性，以提供个别蛋白的特定编码。该方面也包括在个别蛋白内应用条形码，例如在 Fv 的一个或更多个 CDR 位置中应用条形码。该方面也包
15 括应用也可能消化所述蛋白:靶混合物的蛋白组分的蛋白酶或以及任何用于固定化所述靶的蛋白媒介物，条件是在其它肽的背景下仍可以检测到从结合的受试蛋白释放的条形码肽，并且对其进行测序。在
20 该方面的优选形式中，在轻链 C 末端提供一个条形码区，形成一个可溶性 Fab 片段，由此 VH 和 VL 由同一表达盒或顺反子编码，使得使用所述条形码序列既可以选择 VH 基因，又可以选择 VL 基因。这样
25 一种 Fab 片段可以方便地用多种表达系统产生，例如 M13 噬菌体系统，其中通过引入分泌前导序列，Fab 的重链和轻链被分泌到宿主细菌的壁膜间隙中，然后从该间隙中收获所述重链和轻链。首先通过在合成的寡核苷酸混合物中克隆，制备具有符合读框的条形码的载体系统。对于两个相邻条形码的形成，这可以通过顺序克隆或寡核苷酸诱

变来方便地进行, 由此制备含有混合的第一条形码的汇集的 M13 重组体, 作为随后克隆第二个符合读框的条形码的模板。最好是, 设计条形码, 使得所编码的蛋白含有邻接这两种条形码并且位于这两种条形码之间的内切核酸酶位点, 并且由此与所述条形码之一相邻的一个“间隔”区产生包括该条形码、其分子量大于另一条形码的肽。通过合适的设计条形码并且以该方式应用多个条形码, 提供一种选择, 用于简单地分析内切核酸酶释放的肽的质量, 例如通过 MALDI-ToF 进行分析, 由此可以推导出(或接近推导出)所述肽的序列, 使得可以设计合成的寡核苷酸来分离(或富集)具有通过MALDI-ToF分析检测到的条形码的特定蛋白。这些序列的这种推导可以通过序列设计来完成, 由此使特定氨基酸仅存在于沿所述肽的一个或两个位置。例如, 当用 20 种天然氨基酸中的 17 种(因此为指定的 A-Q)设计所述肽时, 可以设计所述序列, 对于沿所肽序列的每个位置的三个氨基酸中的任一种有如下多种选择:

15

aa 位置:	1	2	3	4	5	6	7	8
氨基酸选择:	A	C	E	G	I	K	M	O
	B	D	F	H	J	L	N	P
	C	E	G	I	K	M	O	Q

20

这种设计理论上能产生 6561 种不同的肽序列条形码。如果也在同一基础上设计一种具有一个间隔区的相邻条形码, 则这将产生另外 6561 种不同的条形码。总之, 这将产生 4×10^7 种个条形码序列, 足以独特标记这种大小的蛋白质文库的大多数成员。应用添加的相邻条形码或基于例如在该序列中的任一位置应用两种特定氨基酸的较长的条形码(因此用 19 中氨基酸产生 262,144 种不同的条形码序列)将增加所提供的条形码的多样性。实际上, 在设计混合的合成寡核苷酸过程中, 通过合理选择该序列内每一位置的密码子, 降低密码子冗余性。

用于 MS/MS 测序的 8 个氨基酸条形码肽的一种寡核苷酸设计如下:

密码子 -	NAC	NCC	NGG	NTG	TKC	VAG	GNV	CNT
氨基酸 -	N	T	R	L	F	Q	D	H
	D	P	G	M	C	E	V	L
	H	A	W	V		K	A	P
	Y	S					G	R

其中密码子 N = A、C、G 或 T
K = G 或 T
V = A、C 或 G

$4 \times 4 \times 3 \times 3 \times 2 \times 3 \times 4 \times 4 = 13824$ 种条形码序列

5 也可以通过不连续的寡核苷酸合成掺入特定的密码子, 从而将特定的密码子顺序加至分开的先前合成的寡核苷酸混合物(“密码子诱变”)。一旦通过 MALDI-ToF 或 MS/MS 推导出候选条形码序列, 并且单个条形码的多样性小于该文库的多样性, 则可以用相应的特异性寡核苷酸(或如果在密码子选择中有冗余性, 则用寡核苷酸混合物)作为 PCR 引物, 并结合根据该蛋白或载体系统设计的相对的引物, 来富集可从中检测到所述条形码的蛋白的编码基因。当用相邻条形码时, 由所述第一引物产生的在一个基因片段内嵌套的第二引物则可以用来富集所检测的实际蛋白的编码基因。如有需要, 上述方法可以加入三
10 种或更多种条形码, 以便增加寡核苷酸指导的编码所需蛋白的特定基因富集的特异性。本领域技术人员会理解, 本发明的该第一方面可以包括多种变化的方法, 其主要原理是: 通过对与该特定蛋白结合或由该特定蛋白编码的一种或更多种肽进行质量分析或序列分析, 从这种蛋白文库中回收特定蛋白, 因此该方面在分离编码仅确定或推定肽序列的所需蛋白的基因方面具有广泛的应用。

20 本领域技术人员会理解, 在本发明范围内有所述第一方面的多种

变化形式。例如，人们会理解，肽条形码可以被掺入多对或多组蛋白中，然后让所述蛋白结合，通过鉴定来自参与结合的每种蛋白质的条形码，确定哪些蛋白相互结合。作为从复杂混合物中分离蛋白质的一种替代方法，可以用本发明的方法检测这些复杂混合物中表现出一定结合特性诸如与其它大分子(例如 DNA)结合的蛋白。本发明的方法包括将肽条形码加入蛋白质的多种方式，包括其中在编码该蛋白的基因片段内编码条形码的方法。然而，可以通过肽的直接连接，将条形码加入这种蛋白，或加入任何其它合适的分子混合物。例如，可以用多种化学方法或光化学方法，将特定肽加至特定抗体或蛋白质。这样一种方法的一个应用是用一种条形码标记一种复杂的蛋白混合物(或例如根据不同的蛋白特异性选择条形码)，而用另一种条形码标记另一复杂的蛋白混合物，例如以差别条形码标记(barcode)来自两种不同样品的蛋白，然后将其混合。本领域技术人员会理解，将肽条形码加至蛋白或其它分子的原理也可以应用于非肽条形码，由此可以用质量或测序法直接鉴定(或大致鉴定)这类条形码。因此，所述条形码可以包括与蛋白或其它分子(包括核酸)连接的核酸条形码。如同肽条形码一样，可以通过质谱法分析这类核酸条形码，以提供对质量的准确估计。可以用限制性酶代替蛋白酶，从所述蛋白中释放出这种条形码。

本领域技术人员会理解，在本发明范围内有所述第一方面的蛋白分离以外的多种应用。例如，可以根据质量或序列来分析条形码，分析与活有机体内特定器官相关的蛋白质或其它配体在该有机体内的分布。在肽或蛋白与其它分子的结合特异性的分析中，可以构建条形码作为所述肽或蛋白结合区的部分，以便通过对条形码的质量或序列分析来分析特异性。例如，可以在 MHC 分子的已知锚定残基周围构建混合肽条形码序列，然后通过所述条形码的洗脱和质量或序列分析，测定与特定 MHC 分子结合的肽谱。

第二方面，本发明提供一种筛选蛋白质文库的方法，包括根据一种或更多种所需特性筛选所述文库，然后去复制(dereplication)以鉴定

该文库中具有所需特性的一种或更多种个别蛋白。

本发明的该方面可供用于蛋白质文库，尤其是诸如 Fv 之类的重组抗体文库，藉此分离与特定靶结合的所述文库的个别成员，从而分别筛选来自该文库的蛋白质库，然后对阳性库进行一轮或更多轮去复制，直至鉴定出该文库中与所述靶结合的个别蛋白。具体地说，该方面涉及不用展示系统而筛选蛋白质文库，即其中所述蛋白与相应的基因没有物理上的连接。在该方面，根据与所述靶的结合筛选多个蛋白库，由此或者标记所述靶以指示哪个(哪些)库含有结合的蛋白，或者其中不经标记而检测所述靶。一种特别优选的方法是在溶液中筛选蛋白库，而不将蛋白质与其它部分(其可能影响蛋白与其靶的结合)进行任何融合或连接，然后在质量分析、尤其是通过 MALDI-ToF 进行质量分析之前沉淀所述总蛋白库(与任何连接的靶一起)，以便根据代表所述靶并因此指示所述靶是否结合的电离峰的“指纹”进行筛选。一旦鉴定出一个或更多个阳性结合库，则可以将这些库去复制，或者以降低所述库的复杂性，或者以分离出个别蛋白，以筛选与所述靶的结合物。在实践中，装配蛋白库的一个特别优选的方式是首先装配编码这些蛋白质的基因库。如果将基因克隆到例如质粒或噬菌体载体中，则可以通过将各个细菌菌落或噬菌斑混合在一起合并这些基因，或更方便的是通过在分开的琼脂平板上铺平板(以例如每个平板 1000 个菌落/噬菌斑的密度)对菌落/噬菌斑进行分离，并且将菌落/噬菌斑从这些平板上刮/洗脱为一种混合物，合并这些基因，然后用该混合物或者通过细菌/噬菌体表达或者通过体外转录/翻译合成所述蛋白。以相似的方式，可以用其它微生物或体外合成系统来合成蛋白质。该方面也包括应用复杂的靶，例如多种分子的混合物、全细胞或细胞膜，由此所述分子靶产生成为与所述复杂靶内的特定分子靶结合的特征性的质量分析“指纹”。当所述靶为蛋白质时，该方面也包括应用蛋白酶消化所述靶，以便产生指示所述靶的肽质量指纹，并且当所述蛋白酶也消化来自该文库的所述蛋白质时，甚至在衍生自该文库的其它肽的背景中

5 仍可以检测到所述指纹。该方面也包括多种不同类型的“靶”和除与靶结合以外的筛选蛋白质库或个别蛋白的标准。例如，该方面包括应用生物测定系统作为选择蛋白质的标准，例如其中根据刺激或抑制一种生物活性来选择蛋白质。结合测定的其它形式将包括对配体与其受体结合的抑制以及选择与靶上某些位置结合的蛋白质，其中所述靶可以例如是一种分子、细胞或组织切片。

第三方面，本发明提供一种蛋白鉴定和/或测序的方法，包括提供一种个别蛋白的文库，一种或更多种所述个别蛋白可以与感兴趣的靶结合，其中每种个别蛋白与其基因一起结合于一个“结合部分”。

10 本发明的该方面可供用于蛋白质文库，尤其是诸如 Fv 之类的重组抗体文库，从而通过加入或包括一个“结合部分”将所述蛋白质及其相应的基因结合在一起，由此让所述蛋白质或 Fv 在加入所述“结合部分”之前或之后与所述靶结合。所述结合部分可以用来使得通过所结合的相应基因，例如通过 PCR 扩增(或其它扩增方式，例如通过细菌转化、或通过直接测序和随后通过该序列进行再生)，能够再生所述蛋白或 Fv。如果所述蛋白或 Fv 作为具有相应基因库的库产生，则与所述靶结合(或者如果需要，则不结合)的蛋白或 Fv 连接的基因可用作再生个别或较小的蛋白库或 Fv 库的基础，以便重复筛选，以鉴定出与
15 所述靶结合的特定蛋白或 Fv (通过所述相应的基因)。

20 该第三方面的一种具体方式是所述结合部分是一种粒子，并且因此重组蛋白及其相应的基因被共固定化至粒子上，从而个别粒子的回收可供用于鉴定编码所述重组蛋白的一种或多种基因。这种形式具体涉及用来将编码所述重组蛋白的基因共固定化至固定化其相应蛋白的同一粒子上的方法，使得在选择所述重组蛋白时，也将选择出相应的
25 基因，使得可以鉴定被选择的蛋白的身份(通过对该基因测序)，或使得可以由该基因产生其它重组蛋白。本发明的该方法包括通常通过控制该粒子上所述重组蛋白与之结合的部分的数目，保证对该粒子上所展示的蛋白的量的控制。本发明包括保证将其它分子(包括其它蛋白或

蛋白链，并且包括所述重组蛋白与之结合(例如抗原)与所述重组蛋白共展示在所述粒子上。

在本发明的第三方面的基本操作中，提供一种基因阵列或基因混合物阵列，由此用例如体外转录和翻译或噬菌体展示等方法来合成重组蛋白，这类蛋白的实例为抗体可变区(Fv)。随后，将基因和重组蛋白共固定化至粒子上，将一种或更多种配体与或者为 DNA 或者为 mRNA 的所述基因连接，因此这类配体与所述粒子表面上的“受体”结合，或者因此这类配体与所述粒子表面反应，产生共价连接或离子连接。或者，所述基因通过形成与天然 DNA 或 RNA 反应性基团的一种或更多种共价键或离子键，而直接固定化至所述粒子上。得到的由所述基因编码的重组蛋白可以具有结合的一个或更多个配体(这类配体是所述蛋白上可以达到固定化的部分)，例如蛋白序列标记(由所述基因编码)或生物素基团(用生物素酰基赖氨酸通过体外转录和翻译而掺入)，使得它们也可以与所述粒子表面上的“受体”结合，或者由此这类配体与所述粒子表面反应，产生共价连接或离子连接。所述基因和蛋白上的配体可以是相同或不同的配体，可以固定化至相同或不同的受体上。关于本发明该方面的有用操作，或者作为 DNA、mRNA 或者在活微生物例如噬菌体中的基因或基因库被分布到阵列(或多个反应容器等)中，并且重组蛋白在这样的阵列中产生(例如通过体外转录和翻译或通过噬菌体的生长)。含有所述基因的主阵列可以用作用以产生所述重组蛋白的材料源，藉此基因或蛋白样品被分配到服务器阵列中，使得每个基因库或蛋白库的阵列位置得以存贮。或者在该过程之前、期间或者之后，将一种或更多种粒子引入该阵列中的每个位置中，提供基因和蛋白可以与之结合的受体。在本发明的一个变体，首先将所述基因连接于所述粒子，然后从这些基因直接产生蛋白，使得这些重组蛋白随后被固定化到至同一粒子上。在重组蛋白连接至所述粒子之前或者之后，所述蛋白可以任选地经过修饰，例如用其它激酶磷酸化或被其它蛋白结合。在第三方面的一个变体中，所述阵列包括

其中分隔了基因或活微生物的小滴(通常通过在蛋白合成之前产生所述小滴, 因此将所述基因布置在所述小滴中), 例如乳液包油小滴或脂质体。蛋白质在所述小滴中产生, 然后将包括所述基因在内的这些物质共连接至所述粒子。在小滴的情况下, 所述基因和蛋白共连接的粒子可以或者被引入所述小滴中, 或者所述粒子可以是小滴本身。例如, 在脂质体的情况下, 可以产生具有与脂质体膜结合的亲脂标记的蛋白, 尤其是在这导致所述蛋白在脂质体外表面上“展示”的情况下。一个相关的实例是使用 mRNA 体外翻译的情况, 其中微粒体膜可以被引入到该反应中, 从而具有亲脂标记的蛋白可以整合到这类膜中, 随后这类膜可以分散成为小粒子。

如果希望然后随所述粒子合并用于选择过程, 则从阵列中回收粒子并将其混合; 然后对粒子上的重组蛋白进行选择, 通常通过将其暴露于与粒子上的选定蛋白结合的靶进行选择。某些重组蛋白也可以在该阶段进行修饰。携带选定或经修饰的蛋白的粒子然后通过多种方法回收, 例如, 如果用荧光标记来标记靶, 则可以用 FACS 来分离出具有(或不具有)所述靶的粒子。在本发明的第一个主要方面, 然后可以通过对共固定化 DNA 或 mRNA 进行 PCR 扩增, 回收在这类选定粒子上的编码重组蛋白的基因。

有许多类型的可以用于本发明的第三方面连接蛋白与其相应基因的“结合部分”。使用的粒子包括胶乳粒子和磁性粒子、以及在其上直接合成合成的寡核苷酸的粒子。这类粒子通常装备有所合成的多肽可以结合的“受体”。其它结合部分可以是单个分子或分子复合体, 它们可以用作将基因分子与所合成的蛋白连接的桥。例如, 基因分子和所合成的蛋白均可以包括生物素基团, 然后通过加入链霉抗生物素蛋白将其交联, 因此链霉抗生物素蛋白可用作结合部分。以相似的方式, 可以采用例如双特异性结合试剂例如双特异性抗体(与序列标记和配体两者均结合)或抗体-链霉抗生物素蛋白缀合物(因此所述抗体或者结合于蛋白或者结合于基因上的配体, 而链霉抗生物素蛋白结合

于蛋白或基因上的生物素，无论其是否是无配体的)，将蛋白上的序列标记和基因分子上的配体交联。其它结合部分可以是细菌或噬菌体，由此所合成的多肽结合于所述细菌或噬菌体上的特异性配体。例如 M13 表达系统可以用来在大肠杆菌中产生特异性抗体的 Fv 片段，然后所述片段可以结合于 M13 自身上的特定蛋白抗原，尤其是在蛋白抗原展示在与衣壳蛋白融合的所述噬菌体头部上的情况下。通过测试已经结合 Fv 的 M13 噬菌体，可以通过对 M13 编码的 Fv 基因测序，确定编码特异性 Fv 的基因。同样，M13 表达系统可以用来产生与 M13 自身上展示的特定蛋白结合的蛋白。在所有情况下，第三方面的特有特征是：重组蛋白分子可以在合成后，通过一个结合部分与相应的基因连接。这样的合成后的连接尤其保证不受阻碍地合成蛋白分子，而例如不需要蛋白作为具有其它蛋白分子的融合体来合成，所述其它蛋白分子可能改变所述蛋白的构象或干扰其识别或功能。

本发明包括在与结合部分连接之前产生重组蛋白的几种方法。这些方法尤其包括通过体外转录和翻译的蛋白合成、以及由质粒或噬菌体指导的在细菌中的蛋白合成。在后一种情况下，本发明提供的优点是：所产生的蛋白不需要与噬菌体蛋白融合，因为本发明中的所产生的蛋白随后被固定化至分离的粒子上。与目前蛋白质的噬菌体展示方法(其中这类蛋白与表面噬菌体蛋白或可以达到表面的蛋白融合)比较，本发明的第三方面将或者需要噬菌体的裂解，或者需要重组蛋白从噬菌体头部分泌或渗漏，以便保证其随后固定化至所述粒子上。因此，产生蛋白的其它体内方法，例如在细菌、酵母或甚至哺乳动物细胞中的表达，可以用于所述第三方面，因此具有比单一展示法更为通用的优点。因此，重组蛋白在固定化之前，可以被特定宿主修饰，例如被哺乳动物细胞糖基化。本发明的一个特别有用的方面是能够控制结合部分上的重组蛋白分子数目，尤其是当所述结合部分是粒子时，通过控制所述粒子上“受体”的分子数目来加以控制。因此在抗体可变区的情况下，单种抗体或抗体库的效价可以根据选择标准而变化。

再一种可供选择的结合部分可以是活细胞本身，由此重组蛋白被连接至活细胞表面或附近的配体，例如带有所述表达质粒的细菌或哺乳动物细胞的细胞表面标记，或者由此当分泌时，所述蛋白则结合于从中表达所述蛋白的细胞。所述蛋白然后可以与靶反应，然后可以分离出带有结合所述靶的特异性 Fv 的表达盒的细胞。

5 本文的第三方面提供了一种特别有用的方法，用于选择与靶结合的重组蛋白，或用于选择通过特定处理例如通过用细胞或组织裂解液处理而被修饰的重组蛋白。因此该方法将证明尤其可用于重组蛋白的分子进化，藉此连续多轮选择确保仅回收具有严格特性(例如与靶高亲和性地结合)的蛋白。该方法也包括对选定基因进行连续多轮诱变，以将进化选择的多样性最大化。有许多可以基于本发明的第三方面使用、但属于本发明范围内的变化方式，这对本领域技术人员是显而易见的。例如，用来捕获所述基因和重组蛋白的结合部分、尤其是粒子，可以将另一种多肽链结合至粒子本身，由此当蛋白-蛋白结合发生时，所述重组蛋白不是被所述粒子直接捕获，而是被已经在所述粒子上的多肽链捕获。然后，可以用重组蛋白上的适当标记或配体为检测蛋白-蛋白结合事件提供工具。可以以同一方式将合成寡核苷酸结合到粒子上，所述合成寡核苷酸随后用来退火至所述基因上，用作捕获所述基因到粒子上的工具。

15 20 第四方面，本发明提供一种蛋白鉴定和/或测序的方法，包括提供一种个别蛋白的文库，一种或更多种所述个别蛋白可以结合于目标靶，其中每种个别蛋白连接于一种单独的“编码部分”。

在本发明的该方面，由基因文库合成的重组蛋白随后被连接于通过一种编码方法或其它编码方法可区别的“编码部分(coding moiety)”，例如粒子，所述区别方式使得所述编码与连接于所述粒子的重组蛋白的编码基因的身份相关。当将重组蛋白固定化至所编码的粒子上时，所述重组蛋白可以具有一种或更多种相连接的配体，例如蛋白序列标记或生物素基团，使得它们可以结合于所编码的粒子表面

的“受体”上，或由此所述配体与所述粒子表面反应，以产生共价连接或离子连接。在本发明该方面的操作中，在大阵列中合成或分隔重组蛋白或蛋白库。然后将具有独特编码的粒子引入该阵列中的每个位置中。这类编码包括例如可测量的信号部分(例如荧光标记、化学发光标记或放射性标记)的不同比率或区别粒子的不同物理特征，例如不同形状或标志，例如一种蚀刻在所述粒子中的一种编码或特有标记。在每种情况下，各种所编码的粒子可以相互区分。用于本发明的粒子包括任何具有可以连接蛋白的特性的这类粒子、复合体或分子。在合并粒子并且让所编码粒子上的蛋白混合物与特定靶结合后，则可以测定选定粒子的编码，以便确定其原始阵列位置，并且因此确定所选定重组蛋白的编码基因的阵列位置。作为本发明这些方面的一个变体，可以采用例如 MALDI-TOF (质谱法)的方法或采用标记抗体以鉴定已知蛋白，鉴定粒子上的选定蛋白。本发明第四方面的操作和范围将共享本发明上述第三方面的许多方面和范围。

35. 第五方面，本发明提供一种分析蛋白混合物的方法，包括：

- (i) 消化或切割所述蛋白混合物；
- (ii) 分级分离所产生的肽；以及

通过其质量和/或序列来分析所产生的肽。

本发明该方面涉及分析蛋白混合物的方法。具体地说，本发明涉及比较不同细胞和组织之间的蛋白的方法。本发明涉及以下步骤的组合：消化或切割蛋白混合物、采用蛋白结合试剂文库分级分离肽以及随后分析肽分级分离部分的质量或序列。除用蛋白结合试剂分级分离外，本发明还包括任选的蛋白或肽片段的物理分级分离。目前分析整个复杂的蛋白混合物(例如哺乳动物细胞或组织中的蛋白混合物)的方法，需要采用例如双向(2D)凝胶电泳的技术分离蛋白。对于这一技术，细胞蛋白通常在一向基于电荷进行分离，而在另一向基于大小进行分离。可以或者参考已知蛋白的电泳迁移图形，或者通过从电泳分离的斑点中洗脱所述蛋白并且通过例如质谱法和核磁共振的方法进行分

析，来鉴定蛋白质。然而，2D 蛋白凝胶法的限制包括对来自细胞的蛋白质的分辨率和检测有限(通常仅清楚地检测到 5000 种细胞蛋白)、对所分离蛋白鉴定的限制(例如质谱法通常需要 100 费摩尔或更多的蛋白用于鉴定)、该技术的专业化性质和难以使该技术自动化以便达到非常高的蛋白分析通量。因此需要优越的方法来分析整个复杂的蛋白混合物，尤其是采用不用凝胶电泳的方法和易于自动化的方法进行分析。

第五方面的核心是：将蛋白或者消化或者切割为较小的肽片段，然后用蛋白亲和试剂文库进行分级分离，然后进行质量分析，尤其是通过质谱法分析。除用蛋白亲和试剂进行分级分离外，可以任选地对蛋白或肽片段进行物理分级分离，并且也可以将其与一种或更多种“化学标记”缀合，以有助于分级分离。

第五方面的主要方面可供用于采用蛋白酶或化学方法切割蛋白、分级分离由此产生的肽混合物，并且随后进行质量分析。采用蛋白亲和试剂、尤其是重组抗体片段的文库，来完成肽的分级分离。该方法可选地包括采用物理方法或特异性亲和试剂例如抗体、或者固相或反应性化学基团分离肽或肽混合物，来对蛋白或肽进行另外的分级分离，以用于随后的质量分析。用蛋白亲和试剂从肽混合物中回收个别肽或多组肽，以进行随后的质量分析。或者，或另外，可以用蛋白亲和试剂从该混合物中去除肽，由此随后对所述混合物本身进行质量分析。蛋白亲和试剂或者可以借助肽中特定序列或结构来结合，或者借助特定化学基团(或者作为所述肽的天然组分，或者作为在切割之前或之后加入所述肽的化学标记)来结合。

关于对较大的肽混合物的分析，可以使用多组蛋白亲和试剂，例如由抗体 Fv 片段(包括单链 Fv)的重组文库提供的蛋白亲和试剂，以便分离多个亚组的肽以进行随后的分析。这样的多组 Fv 将包括各种各样的肽特异性，所述肽特异性可以例如通过将抗体文库预吸附至目的肽样品上而达到，或者通过用目的肽样品免疫动物并且由所述动物的 B

细胞产生重组 Fv 文库而达到。另一方面，可以由免疫动物制备多克隆抗血清或多组单克隆抗体，将其用来分级分离肽。然后用个别的选定抗体或选定抗体混合物，从试验样品中分离(或去除)特定亚组的肽。对多种肽的随后质量分析可以便于检测出试验样品之间的特定蛋白的差异。

5

产生抗一种混合物中所有肽的重组 Fv 或抗体是困难的，并且这高度取决于混合物中肽的数目和个别肽以适当亲和性与抗体结合的容易性(“抗原性”)。对于非常大的肽混合物，一个限制是冗余度，由此具有相同肽特异性的抗体重复出现，而针对其它肽特异性的抗体出现不足或缺乏。如果来自特定蛋白的肽没有被抗体结合，这可能使得未对特定蛋白进行质量分析。因此，一种特别有用的方法是：通过在切割之前将蛋白通过其 N 末端和/或 C 末端预吸附至固相，或者通过将 N 末端和/或 C 末端化学标记以用于切割后的随后分离，从蛋白中分离 N 末端肽或 C 末端肽。原则上，这则应该导致回收代表来自该样品的所有蛋白的所有 N 末端和/或 C 末端肽。蛋白质中 N 末端氨基和 C 末端羧基与内部氨基和羧基相比的差别反应性质，大大有利于这种 N 末端和/或 C 末端肽的分离。采用或者识别特定肽序列或者识别所述肽上化学标记的其它亲和试剂，可以对这类分离的 N 末端肽和/或 C 末端肽进行进一步分级分离，或者通过物理方法例如 HPLC 进行进一步的分级分离。然后在质量分析之前，用蛋白亲和试剂对这类分离的 N 末端肽和/或 C 末端肽进行分级分离。此外，本发明可供用于将不同的化学标记顺序缀合于所述蛋白/肽混合物，尤其是通过特异性切割所述蛋白/肽而顺序暴露 N 末端或 C 末端、并且因此所述 N 末端或 C 末端(或两者)在该末端暴露时与特定化学标记缀合。因此，本发明的该方面可供用于一系列具有多种于末端引入缀合的化学标记的蛋白部分，这样的部分用与所述标记结合的亲和试剂进行分离。作为替代于蛋白分子末端进行化学标记的一个特别有用的方法，也可以将化学标记特异性地连接于非末端氨基酸，使得可以通过内部化学标记分离内部肽。特有

10

15

20

25

的化学适用于将配体连接至几种特定的氨基酸，例如赖氨酸的 ϵ -氨基、半胱氨酸的巯基以及天冬氨酸和谷氨酸的羧基。借助非末端标记分离肽的一个优点是：可以对较大的肽进行选择，所述较大的肽更可能含有连接标记的特定氨基酸，因此在质量分析期间利用超过具有较大的背景噪音的低分子量质量的质量来分离肽。另一个优点是：已经可以获得将化学标记引入到蛋白或肽中的特定氨基酸上的试剂阵列，尤其是提供生物素标记的试剂阵列。

第五方面的另一实施方案提供了顺序循环：用蛋白酶或化学方法进行蛋白切割，在连续的蛋白切割步骤期间或之后用蛋白亲和试剂进行分级分离以及随后进行质量分析。在这种情况下，顺序的切割循环有助于蛋白混合物的分析，从而用蛋白亲和试剂对所述范围的蛋白和肽进行分级分离，并且在每个切割循环后进行分析。该方法也可以包括切割循环之间的化学标记循环以增加质量，或包括除去侧基例如糖基团以便降低质量的步骤。如果此外在每次切割循环(或者加上或不加化学标记、切割或其它修饰)结束时测定在该范围的蛋白片段的质量，则每个循环将获得一定范围的质量分布。采用合适的一系列质量修饰循环，一种蛋白或混合物的结果将是在连续循环时改变的蛋白/肽片段的质谱；这些改变的模式将提供该混合物中特定蛋白/肽的“指纹”。在加上或不加化学标记、切割或其它修饰的特定切割循环之后具有一定质量的特定蛋白/肽片段的出现和消失，将为所述片段序列的鉴定提供了指纹，尤其是为通过参考这类指纹的数据库进行鉴定提供了指纹。因此，对来自不同相关样品的蛋白/肽片段谱的比较，可供用于鉴定这些样品之间的蛋白/肽片段差异。在本发明该方面特别有用的是特异性地识别两个氨基酸并且因此切割该蛋白的蛋白酶。这类蛋白酶的一个实例是激素原转化酶，它在双碱性(dibasic)氨基酸对之间切割。因此，本发明的第五方面提供了采用蛋白消化或切割、用蛋白亲和试剂进行分级分离和质量分析的组合来分析蛋白混合物的新方法。

在第五方面的一个相关方面，在切割之前对蛋白进行分级分离。

对于大的蛋白混合物，特别是直接从全细胞或组织中分离的蛋白混合物，可能希望对蛋白进行预分级分离，以便降低进行随后切割、肽分级分离和质量分析的混合物的复杂性。尽管主要使用直接结合所述蛋白/肽中序列或结构的蛋白亲和试剂，但一种替代方法或一种补充方法是应用化学标记文库，以提供由一组蛋白亲和试剂结合的部分。预分级分离的更常规的方法包括应用或者单向或者双向凝胶电泳，其中分离所凝胶切片，然后对其中的蛋白进行切割和质量分析。其它预分级分离方法包括根据天然修饰例如磷酸化、糖基化、蛋白-蛋白(或肽)相互作用来分离蛋白；或者，可以对膜蛋白或者对来自细胞内特定区室的蛋白进行预分级分离。另一重要的预分级分离方法是：用亲和试剂例如抗体结合并除去高丰度蛋白，从混合物中除去这类蛋白。作为预分级分离的一种替代方法，在切割后产生的肽可以用这些方法中的许多方法来分级分离，也包括利用 HPLC 的大小/电荷分级分离法。这类方法对于分级分离已经通过应用蛋白亲和试剂从混合物选择出的肽特别有用。特别是，可以将 HPLC 与质量分析相连，使得直接对来自 HPLC 分离的肽部分进行质量分析。也可以根据天然修饰，采用例如结合肽中磷酸化氨基酸的抗体，对切割后产生的肽进行分级分离。通过采用蛋白亲和试剂例如单克隆/多克隆抗体来分离用于随后切割和质量分析的特定蛋白，可以完成蛋白的预分级分离。对于较大蛋白混合物的这类分析，优选抗体文库，例如由重组 Fv 文库提供的抗体文库，以便分离多个亚组的蛋白或多个亚组的切割肽，以用于随后的质量分析。这类抗体文库将包括各种各样的蛋白或肽特异性，但也可以预富集与特定目的样品中的目的蛋白/肽结合的抗体文库。对于肽，这最好是通过测试个别 Fv 与该样品中的一种肽或少数肽的选择性结合来完成。或者，通过将抗体文库预吸附至混合目的蛋白/肽样品上，然后用个别选定抗体或选定抗体的混合物分离多组蛋白或肽，可以完成预富集。用蛋白亲和试剂进行分级分离，为一定范围的不同蛋白/肽部分提供质谱，因此有助于检测样品之间特定蛋白的差异。

应用化学标记的再一优点是：随后用亲和试剂对肽进行分级分离，可以大大减少来自一种蛋白分子的选定肽的数目，因此从质量分子中去除分子的其余肽。一种尤其方便的选择性化学标记的方法是：标记混合物中该蛋白分子的 N 末端和 C 末端中任一个(或者这两者)，
5 然后用合适的选择性试剂例如氨基酸特异性或序列特异性蛋白酶(例如内肽酶 Arg-C)或切割试剂(例如酸性 pH，以于 Asp-Pro 切割)消化或切割所述蛋白分子。然后采用亲和试剂，可以分离出来自所述原始蛋白的 N 末端肽或 C 末端肽(或这两者)，并且弃去所有内部肽。因此这种复杂性的降低足以满足用质量分析，尤其是采用 HPLC 与串联质谱
10 仪偶联分析整个肽，以从混合物鉴定所述个别肽。

另一方面，可以仅在例如用双碱性切割剂激素原转化酶消化/切割后进行化学标记。这将确保仅在原始蛋白的一种或多种内部位点进行标记。如果然后用一种不同的酶或切割试剂对所述蛋白混合物进行第二个消化/切割步骤，则当在原始蛋白中存在切割位点时，将减小所标记的肽的大小。然后可以用蛋白亲和试剂对所标记的肽进行分级分离，并进行质量分析。
15

在第五方面的另一实施方案中，让一种蛋白混合物经过多个标记、消化/切割和质量分析循环，，由此仅对通过应用亲和试剂与特定化学标记的结合所产生的一等份混合物产物进行蛋白亲和试剂分级分离和质量分析，然后用一种不同的化学标记对主混合物进行标记以及消化/切割。这顺序提供了一定范围的不同片段。对该方法的另一种变化涉及与上述相同的初始步骤，但在切割后暴露了新的 N 末端和 C 末端，然后可任选地用一种不同的化学物质标记这些新末端之一(或两者)，因此这标记了原始蛋白的内部位点。如果需要，可以用一种不同的蛋白酶或切割试剂将该过程重复一次或多次，每次将一种不同的化学标记加至所述 N 末端或 C 末端。在该方法的一种形式中，首先用两种不同的化学基团分别在 N 末端和 C 末端对整个蛋白混合物进行标记，然后用一种蛋白酶切割，例如用在特定氨基酸邻近特异性切割的
20
25

蛋白酶切割，然后用另外两种不同的化学基团于新 N 末端和 C 末端进行标记。这将产生各自于末端带有化学标记的肽混合物。由于所述 N 末端肽和 C 末端肽将具有一种特定标记，因此此后用合适的亲和试剂从混合物中分离出所述肽。没有原始 N 末端标记或 C 末端标记的内部肽可以利用其特定标记加以分离。然后可以重复消化和标记过程，以产生带有标记的其它肽。采用特定标记的亲和试剂的特定组合，则可以分离出来自所述原始蛋白的 N 末端肽或 C 末端肽或特定内部肽，并且弃去选定的肽，以达到复杂性的降低。在将化学标记加至肽中的两种或更多种氨基酸侧链基团时，顺序应用亲和标记可以分离出含有特定氨基酸组合的肽的部分。例如，如果将平均长度为 20 个氨基酸的肽混合物分别于赖氨酸和苯丙氨酸标记，并且该混合物包含 25% 的不包括赖氨酸和苯丙氨酸的肽、25% 的仅包括赖氨酸的肽、25% 的具有苯丙氨酸的肽以及 25% 的既包括赖氨酸又包括苯丙氨酸的肽，那么分开应用或顺序应用或者赖氨酸特异性或者苯丙氨酸特异性的亲和试剂，将导致将肽分级分离为 4 个相等的部分。实际上，这样一种分级分离方案将有利于较大肽与亲和试剂的结合，因为这些肽更可能含有一种或更多种标记的特定氨基酸。这将不利于分析非常小的肽，例如分子量小于 1000 道尔顿的那些肽，这些肽当经过质谱分析时，将更可能与由片段化的肽和其它小分子产生的背景噪音重合。

当需要分析复杂的混合物，例如哺乳动物细胞或组织中的蛋白混合物时，本发明提供了一种主要方法，其中在切割之前或者之后用蛋白亲和试剂对蛋白进行分级分离，然后对所述肽进行质量分析。对复杂的蛋白或肽混合物的分级分离需要相应的复杂的蛋白亲和试剂混合物，并且可以用一种或更多种另外的可以识别作为分级分离基础的所述蛋白/肽特征的亲和试剂来辅助进行。当切割在分级分离之前进行时，本发明中最常用的方法是用例如胰蛋白酶或 V8 (Glu-C) 蛋白酶的蛋白酶切割整个蛋白混合物，然后对某些肽进行选择性地分离和质量分析。

通常是，一般通过在切割之前将化学标记加至所述蛋白的 N 末端和/或 C 末端，并且采用一种分离具有所述化学标记的肽的亲试剂，分离来自所述肽混合物的 N 末端肽或 C 末端肽(或这两者)。或者，可以采用已经选择用于与特定蛋白中的特定肽结合的亲和试剂，分离特定肽(N/C 末端的或其它的)；这些将从所述混合物中选出那些肽。对于更加复杂的蛋白混合物，在质量分析之前，最好使用基于大小、电荷或疏水性的再一个分级分离步骤，例如 HPLC 分级分离，尤其是当该步骤可以与质量分析相连时。然后对肽的选择性分离可供用于对衍生自替代蛋白混合物的特定肽的相对量(涉及其相应混合物中所述蛋白的相对水平)以及在某些情况下对所述肽的修饰进行比较分析。

关于 N 末端肽或 C 末端肽的分级分离，蛋白亲和试剂的制备和应用是本发明的一个重要方面，蛋白的 N 末端或 C 末端的标记是另一重要方面。关于典型的来自哺乳动物细胞或组织或者许多活有机体的蛋白混合物，将修饰(例如通过甲基化)这些蛋白的几个 N 末端(和某些 C 末端)，使得可以阻止将化学标记加至所述末端。另外，对于典型的来自哺乳动物细胞或组织或许多活有机体的蛋白混合物，所述蛋白将以不同的相对丰度水平存在，通常肯定包括高丰度蛋白。当将来自哺乳动物细胞或组织或其它活有机体的蛋白混合物用于蛋白亲和试剂的初次选择时，这类高丰度蛋白可能决定亲和试剂的选择，并且可能在用于质量分析的最终肽混合物中占优势。对这两个问题的解决方法是用人工混合蛋白源分离所述亲和试剂。通常，这将是一种基因表达系统，藉此用基因(通常是 cDNA)文库来产生无 N 末端或 C 末端修饰的蛋白。另外，基因表达系统的应用使得基因文库可以被“标准化”，以减少或除去文库内的高丰度基因。这通常通过在构建该文库之前使所述 DNA (或 RNA)自退火来完成。因此，本发明中的一种常用方法是通过基因文库(通常被标准化的)的表达产生蛋白，产生无显著 N 末端或 C 末端修饰的蛋白，当被标准化时，产生无特定蛋白占优势的蛋白混合物。与基因文库一起使用的一种典型的表达系统是采用真核生物

核糖体制剂的体外转录和翻译；这也提供了将经修饰的氨基酸掺入所表达蛋白中的可能性。所表达的蛋白混合物然后可以直接用于N末端或C末端标记。在N末端氨基或C末端羧基未被修饰或防止其进行随后的化学标记的情况下，也可以使用其它表达系统。在发生修饰时，
5 在某些情况下可以或者用诸如组蛋白脱乙酰酶之类的酶或者诸如除去N末端甲硫氨酸的有限溴化氰切割之类的化学法，除去N末端修饰。

在产生无N/C末端修饰的蛋白混合物后，则可以将化学标记加至N/C末端氨基。对于N末端，最初可以用例如柠康酐或乙酰亚胺酸甲酯(methyl acetimidate)之类的试剂封闭赖氨酸的 ϵ -氨基，以便仅允许N
10 末端氨基反应。或者，可以通过将经修饰的赖氨酸加入表达系统例如体外转录/翻译中，从而例如可以将生物素修饰的赖氨酸代替赖氨酸直接掺入，封闭赖氨酸的 ϵ -氨基。然后，例如用可以得到亲和试剂的特定分子的异硫氰酸酯，将化学标记选择性地加至蛋白的N末端。一种这样的实例是荧光素，通过所述蛋白与异硫氰酸荧光素反应掺入荧光
15 素，使得随后可以用抗荧光素抗体进行纯化。或者，可以将多羧酸螯合剂作为异硫氰酸酯掺入，使得随后可以用特定金属进行纯化。一旦标记了混合物中蛋白的N末端和/或C末端，则可以或者用化学法或者用酶法，用例如胰蛋白酶的蛋白酶或另一种切割剂，对所述蛋白进行全面的特异性切割。这样的切割由此从每种蛋白释放出单个标记的末
20 端肽片段，然后可以用一种适当亲和试剂(例如对所述化学标记特异性的抗体)从未标记肽混合物中纯化这种片段的集合。如果需要，可以增加化学标记的大小，以产生更大的质量用于分析；这对于由非常靠近所述化学标记的切割而产生的肽片段是有用的，否则所述肽片段是如此之小以致在质量分析中被分析为较低分子量“噪音”。。化学标记
25 可能例如包括一段在核酸合成期间引入的通过反应基与所述肽连接的核酸。这种核酸分子也可用于通过将所述核酸退火至互补序列来分离所标记的肽。

在化学标记和分离后，可以用所回收的N/C末端肽混合物作为

“饵”，从直接得自哺乳动物细胞或组织或其它活有机体的蛋白中分离与这些相同肽结合的蛋白亲和试剂。这类亲和试剂通常衍生自作为粒子一部分展示的重组 Fv 文库，其中所述粒子含有编码所述抗体的相应基因。这类粒子的实例是核糖体展示粒子或噬菌体展示粒子，在每
5 种情况下，可以拯救来自选定抗体的基因，以便繁殖那些特异性抗体。作为一种可供选择的方法，可以用 N/C 末端肽混合物筛选抗体(例如重组单链抗体或 Fab、Fv)的大阵列，并且可以通过相应的基因回收显示出与所述肽结合的抗体。作为另一种可供选择的方法，通过适当地免疫动物，可以用 N 末端肽和/或 C 末端肽直接产生多克隆或单克隆抗
10 体。采用这些方法，选择蛋白亲和试剂文库，然后所述文库可以用于分析蛋白混合物，例如来自哺乳动物细胞或组织或其它活有机体的蛋白混合物。这种分析可能涉及或者用所述亲和试剂文库从得自哺乳动物细胞或组织或其它活有机体的蛋白中选择出 N/C 末端肽，或者用个别亲和试剂选择出个别肽。然后可以对所选定的肽进行质量分析，通常采用 MALDI-ToF (基质辅助激光解附/电离飞行时间)分析，其中所述
15 个别肽给出独特的荷质比，该比率则可以用来鉴定所述肽的氨基酸组分。如果可以分离出所述肽，则随后可以用 MS-MS (双质谱法)肽测序来鉴定所述肽。另一方面，新一代四极-ToF LC-MS-MS (“Q-ToF”)仪可供用于在同一仪器中进行顺序的 MALDI-ToF 和 MS-MS。事实上，或者单独或者在混合物中的蛋白亲和试剂可以或者间接或者直接固定化至插入 MALDI-ToF 仪中的解附芯片上，随后可以通过亲和试剂将肽结合至芯片上。这样，可以在一个芯片上分析由多种亲和试剂在不同位点上吸附的多种肽部分。应用重组蛋白作为“饵”来分离蛋白亲和
20 试剂，也提供了将其它标记连接至那些蛋白、从而由所述基因序列编码所述标记的前景；例如，可以例如通过 PCR 介导的掺入，将 C 末端多聚组氨酸标记(使得随后可以用镍整合剂纯化所标记的片段)加入所述基因序列中。

应用重组蛋白作为“饵”来分离蛋白亲和试剂，也提供了本发明

第五方面用于采用由所述重组蛋白编码的标记来特异性地分离肽的另一种常用的方法。采用编码这类标记并且设计用以将这类标记加入所产生的表达蛋白中的特定 PCR 引物，这类标记可以方便地在基因文库构建期间掺入基因(通常为 cDNA)文库成员中，或掺入单个克隆或多组克隆中。最好是，这类标记将在所有读框中掺入所表达的蛋白中，以便产生生产性标记的蛋白。这类标记最好通过 PCR 反应的下游引物掺入，通常的结果是朝所表达的蛋白的 C 末端产生所述标记(尽管在某些克隆中，上游终止密码子可能阻止这种产生)。然而，也可以在 N 末端或者在 N 末端和 C 末端两者中掺入标记。

对于从肽混合物中分离特定肽，可以合成(或通过重组 DNA)产生所述肽序列，然后如上所述，将其用作“饵”来捕获特定的蛋白亲和试剂。这些亲和试剂然后可以用来从得自例如哺乳动物细胞或组织或其它活有机体的经切割的蛋白混合物中分离这些同样的肽。

作为选择性分级分离 N 末端肽或 C 末端肽或特定内部肽的一种可供选择的方法，诸如包括磷酸化氨基酸的肽之类的修饰肽，可以用选择性地与磷酸化氨基酸(酪氨酸、苏氨酸或丝氨酸或其组合)结合的抗体来分离，或者采用固定化 Fe^{3+} 来捕获带负电荷的肽。同样，在肽的修饰被进一步衍生化以有助于分离的某些情况下，可以分离出通过糖基化和其它修饰来修饰的肽。例如，可以容易地通过高碘酸盐反应来修饰糖类，作为加入例如荧光素的化学标记的中间体。本发明的一个特别重要的方面是对选择性修饰的肽进行分级分离，由此在切割之前借助将这类肽差别暴露在原始蛋白环境内的标记，选择性地标记这类肽。例如，通过用优先与特定氨基酸基团反应的标记剂处理细胞，可以例如用生物素选择性地标记活细胞上的表面暴露蛋白。用于在通过细胞内其它刺激物天然标记的蛋白中达到这样的标记的一种间接方法是应用这类刺激物实现所述蛋白的标记。例如，通过将受体的配体加入那些细胞中，可以潜在地标记(例如磷酸化)细胞内的与受体结合的酪氨酸激酶分子。在修饰后，通过切割从蛋白中释放出肽，然后直接

对其进行质量分析，或者在进行质量分析之前，用蛋白亲和试剂如上所述进行分级分离。

最好用质谱法进行本发明对蛋白和肽的质量分析。特别是，MALDI-ToF 分析具有非常准确地测量各种肽的具体质荷比的能力。该方法具有同时分析数千种肽的能力。当个别肽(和蛋白)大于 4kD 时，对其的分辨率则变得较差，使得必需将蛋白切割为肽片段以提供精细的分辨率。最近将液相色谱分离法(例如 HPLC)与串联质谱法连接的方法已经使得可以对包含至多 200 种蛋白的蛋白混合物进行质谱分析。当在蛋白酶消化后分析这类蛋白时，如果假定每种蛋白平均 10 种肽，则该方法可以分析至多 2000 种肽。因此采用本发明的方法，例如仅分析标记的 N 末端肽，则任何一次都可以分析衍生自至多 2000 种蛋白的至多 2000 种 N 末端肽。由于这对于全面分析来自细胞的哺乳动物蛋白(通常 50,000 种/细胞)不够灵敏，因此必须将肽分隔为至少 25 个部分，以便对所有这些部分进行分析。这种进一步的分级分离可以或者采用预先选定的蛋白亲和试剂文库直接完成，或者在连续蛋白消化/切割步骤后利用试剂标记内部末端，然后根据其标记用特定蛋白亲和试剂分级分离肽而完成。作为标准质谱法的一种替代方法，可以用 MALDI-ToF 来产生蛋白质量分布图，可以比较来自不同细胞的蛋白混合物的所述分布图。

化学标记通常是可以通常于 N 末端或 C 末端共价连接于蛋白的部分。对 N 末端的化学标记通常于末端氨基进行。如果必需避免标记赖氨酸的 ϵ -氨基，则可以最初采用例如柠康酐或乙酰亚胺酸甲酯的试剂封闭这些氨基。末端氨基可与种类繁多的化学试剂反应，通常采用异硫氰酸酯进行反应。因此，可以常用的抗体所识别的配体例如二硝基酚和荧光素将这些标记连接于 N 末端，随后可以采用抗体亲和试剂进行分级分离。例如，常用的 Edman 试剂异硫氰酸苯酯可以用来特异性地连接至蛋白的 N 末端，并且必要时可以用可供用于随后与亲和试剂结合的部分将其衍生化。对于 C 末端的化学标记，通常采用基于碳二

亚胺活化的方法来引入被亲和试剂结合的配体。另一方面，已经描述了采用反向蛋白水解将各个部分加入蛋白的 C 末端，从而某些蛋白酶例如羧肽酶 Y 和赖氨酰内肽酶可以以反向工作，以加入化学标记，通常通过或者作为具有与亲和试剂结合的标记的衍生化氨基酸的氨基酸，或者通过随后可以被亲和试剂特异性结合的氨基酸的天然序列来进行。人们会认识到，也可以化学标记各种各样的内部氨基酸，包括通过 ϵ -氨基标记 Lys、通过羧基标记 Glu/Asp、通过巯基标记 Cys、通过羟基标记 Ser/Thr 以及通过羟苯基标记 Tyr。已经描述了大部分其它氨基酸的特异性衍生化。人们公认可以利用翻译后蛋白修饰来加入化学标记，尤其是利用糖基化，其中通常用高碘酸将糖残基氧化为甲醛基团，然后甲醛基团可以与含胺分子反应。可以用来加入化学标记的其它修饰包括脂化(lipidation)、磷酸化和金属离子的加入。人们会认识到，本领域中有许多用于在蛋白分子或肽内的特定位点引入一种或更多种化学标记的方法。

用于第五方面的蛋白亲和试剂通常是单克隆抗体。对于蛋白或肽内的特定序列或结构，采用通常为 Fab、Fv 或单链 Fv 形式的重组抗体结合位点文库，其中所述抗体结合位点通常采用例如噬菌体或核糖体复合体来“展示”，使得可以回收编码特定抗体结合位点的基因。关于在本发明中的应用，可以将抗体结合位点文库分为多个组，例如通过将噬菌斑挑出并且形成阵列，或者将载体中的基因挑出并且形成阵列以进行核糖体展示。这类库通常含有数种蛋白或肽的抗体结合位点，使得可以用所述库进行分级分离。另一方面，可以将需要抗体亲和试剂文库的蛋白或肽混合物固定化，并且用作预选择合适亲和试剂的靶，然后将所述亲和试剂分为多个库，或作为单个试剂使用。对于化学标记，可以用个别单克隆抗体来特异性地与个别标记结合，以便达到随后的分级分离。

本发明的第五方面包括应用非单克隆抗体的蛋白亲和试剂，其中这类试剂可以便于在质量分析之前将肽或蛋白分级分离。这类亲和试

剂可以包括选择性地结合某些肽例如主要组织相容性蛋白和 T 细胞受体的免疫分子。其它蛋白亲和试剂可以包括通常参与蛋白-蛋白结合相互作用的蛋白结构域, 例如 SH1 结构域。本发明包括将包括在混合物中的肽环化的概念, 尤其是当通过例如在还原性条件下连接半胱氨酸残基连接至固相时。对此的一种方法可以是在固定化肽暴露的 N 末端或 C 末端上加入额外的半胱氨酸残基, 例如对于 C 末端固定化肽, 采用肽合成的标准条件, 或采用反向蛋白水解, 由此使用某些蛋白酶例如羧肽酶 Y 和赖氨酰内肽酶。第五方面也包括将蛋白或肽进一步分级分离的方法, 即通过通常于 N 末端加入形成蛋白酶识别序列一部分的氨基酸, 所述蛋白酶在两个或多个氨基酸识别序列特异性切割, 而由起始蛋白或肽提供所述蛋白酶识别位点中的一个或更多个末端氨基酸。这样, 借助新产生的序列, 仅对加入新氨基酸的蛋白或肽部分进行蛋白酶切割。这样, 可以用通常位于 N 末端的额外氨基酸标记蛋白或肽, 在由此标记的混合物部分中产生特定蛋白酶切割位点。然后可以例如通过所述新末端, 例如采用一种经标记的末端氨基酸, 或通过一种化学标记加至所述末端, 由此可以用一种亲和试剂来将所标记的部分固定化, 从而将所述蛋白或肽固定化。在除去非固定化的未标记分子后, 可以用所述特定蛋白酶切割所述蛋白或肽, 所述蛋白酶则仅在已经通过合成衍生的氨基酸和原始蛋白或肽衍生的氨基酸组合产生的切割位点进行切割。然后可以用蛋白亲和试剂对经切割的蛋白进行分级分离, 并且进行质量分析(或在质量分析之前进一步加工), 由此提供所述肽混合物的一个亚组。通过采用将特定氨基酸平行合成至已暴露末端, 然后固定化并进行切割, 可以根据末端氨基酸对大的蛋白或肽混合物进行分级分离。蛋白酶识别位点的一个实例是 ile, glu, gly, arg, 该识别位点在 gly 和 arg 之间被因子 Xa 切割。可以将序列 ile, glu, gly 合成到蛋白或肽的 N 末端, 因此, 如果该蛋白或肽序列中的相邻氨基酸是 arg, 则将产生所述切割位点, 并且可以被因子 Xa 切割。蛋白酶切割位点的其它实例是 asp, asp, asp, asp, lys, 由肠激酶在 asp 和

lys 之间切割; pro, gly, ala, ala, his, tyr, 由基因酶(genease) I 在 his 和 tyr 之间切割; leu, val, pro, arg, gly, ser, 由凝血酶在 arg 和 gly 之间切割。部分序列 asp, asp, asp, asp 的 N 末端加入, 可以用来鉴定具有 N 末端 lys (由肠激酶切割)的蛋白或肽, pro, gly, ala, ala, his 的 N 末端加入可以用来以鉴定具有 N 末端 tyr (由基因酶切割)的蛋白/肽, leu, val, pro, arg 的 N 末端加入可以用来鉴定 N 末端 gly, ser; 或 leu, val, pro, arg, gly 的 N 末端加入可以用来鉴定 N 末端 ser (由凝血酶切割)。诸如 MMP (基质金属蛋白酶)之类的具有特定识别位点的其它蛋白酶可以用来分级分离具有其它 N 末端氨基酸的蛋白。不同的蛋白酶识别位点因此可以与所述蛋白酶结合使用, 以根据 N 末端氨基酸分级分离蛋白或肽。作为一种替代方法, 将一个或多个氨基酸加至肽的游离 N 末端, 可以用来产生被亲和试剂结合的位点, 包括这样一种位点取决于来自所述肽的一个或多个 N 末端氨基酸的情况。因此, 通过将氨基酸加至 N 末端, 以取决于 N 末端氨基酸的方式产生蛋白酶消化位点或亲和试剂结合的位点, 可以区别不同的肽或多组肽。当蛋白用作原材料, 尤其是对于来自哺乳动物细胞的蛋白, 因此所述 N 末端蛋白是甲硫氨酸时, 必要时这可以通过例如甲酰化或者通过特异性除去末端甲酰甲硫氨酸的细菌蛋白酶的切割来除去。

蛋白亲和试剂是本发明第五方面的一个重要方面, 并且可以用于广泛地分级分离多组蛋白/肽, 或用于特异性地分级分离个别蛋白/肽。关于分级分离, 首先必需制备与一个特定部分的肽或特定肽结合并且与其它部分/肽不结合的多个部分的蛋白亲和试剂或单个蛋白亲和试剂。一种便利的方法是在分离蛋白亲和试剂之前分级分离所述蛋白或肽。在抗体作为蛋白亲和试剂的情况下, 则可以用这种蛋白/肽或者结合来自文库的展示的抗体, 或者可以用来免疫动物用以产生抗血清。当使用重组抗体结合位点文库例如单链 Fv 文库时, 可以在与蛋白/肽部分结合后回收编码这些重组抗体结合位点的基因, 提供可复制的亲亲和试剂源, 用于随后分离特定的蛋白/肽部分。另一方面, 可以根据结

合特异性筛选个别单链 Fv，例如通过用 MALDI-ToF 分析肽结合。在这种情况下，与来自大蛋白混合物的一种肽结合的单链 Fv 作为基因克隆保留(实际上也保留结合至多 3 种肽的那些单链 Fv)，用于随后单独应用或在 Fv 混合物中应用，以从该混合物中分离一个蛋白/肽部分。

5 人们会认识到，来自蛋白的游离 N 末端通常是用于分离高度特异性抗体的良好的靶，因此来自蛋白的 N 末端肽的捕获和释放将特别有利于随后的抗体分离。某些 Fv 可以用于从混合物除去高丰度蛋白或肽。人们会认识到，单链 Fv 结合的保留和表征也可以提供通过消除具有与其它 Fv 相同的特异性的 Fv 来降低丰余性的一种手段。

10 本发明第五方面的各种实施方案包括以下的组合：蛋白消化/切割、用蛋白亲和试剂进行分级分离和质量分析、以及采用针对所述蛋白或肽中特定序列或结构的亲和标记的可任选的分级分离步骤、和化学标记并根据这些标记分级分离的可任选的步骤。不同的方面包括如下这些步骤的不同顺序：

- 15
- 1 - 重复消化/切割循环和质量分析
 - 2 - 消化/切割、用蛋白亲和试剂分级分离、质量分析
 - 3 - 用蛋白亲和试剂分级分离、消化/切割、质量分析
 - 4 - 末端化学标记、消化/切割、用亲和试剂分级分离、质量分析
 - 5 - 同 3 一样，但还有标记、消化/切割、分级分离的额外的循环
 - 20 6 - 同 3 一样，但还有重复的标记、消化/切割循环和质量分析

本发明第五方面应该被认为包括这些和相关的蛋白/肽加工步骤以及降低蛋白混合物复杂性的核心目的，以便达到对所产生的蛋白/肽步骤的质量分析。

25 目前用于本发明操作的常用方法涉及消化蛋白(或者是天然的，或者由 cDNA 文库编码)混合物的 N 末端和/或 C 末端、用蛋白酶切割、将所述 N 末端和/或 C 末端肽片段固定化、以及释放所述肽和对所述肽进行质量分析。另一方面，通过在用序列特异性蛋白酶切割之前加入氨基酸，可以对 N 末端或 C 末端进行修饰。在质量分析之前，用所述

肽来结合蛋白亲和试剂例如抗体，其中已经预选择出这些抗体来分级分离所述肽，或这些抗体本身作为亲和试剂保留。可以例如根据大小对蛋白混合物进行预分级分离，或可以由通过分隔克隆预分级分离的 cDNA 文库产生蛋白混合物。然后用所保留的蛋白亲和试剂来分析复杂的蛋白样品，从而用所述抗体来结合肽，然后对所述肽进行质量分析。

人们会认识到，本文描述的蛋白的消化/切割、分级分离和质量分析的相同原理中的许多原理也可以应用于其它聚合分子，例如 DNA 或 RNA。在 DNA 或 RNA 的情况下，5'端和 3'端的游离磷酸基团和羧基分别提供一种高度特异性地加入化学标记或与固相直接连接的一种工具。序列特异性限制性酶或修饰酶可供用于切割或修饰 DNA 分子。有用的 DNA 或 RNA 的亲和试剂是核酸本身，它们可以与互补 DNA 或 RNA 序列特异性地杂交，在杂交之前或之后将其连接于固相。用这类方法，可以对复杂的核酸混合物进行分级分离，然后进行质量分析，尤其是采用质谱法进行质量分析。

通过以下实施例描述本发明，所述实施例不应被认为对范围进行限制；

实施例 1

采用一对修饰的单链抗体(scAb)基因进行本实施例中描述的实验。制备两种修饰的 scAb，它们由分别融合于或者来自 c-jun 基因或者来自 c-fos 基因的 b-拉链结构域的 N 末端附加表位、重链可变区(VH)、14 个氨基酸的接头(EGKSSGSGSESKVD)、轻链可变区(VL)组成。

将这些构建体克隆到载体 pET 5c (Rosenberg AH 等, Gene, 56:125-135, 1987)中，该载体提供一个 T7 启动子，后接来自 T7 基因 10 的核糖体结合位点。将所述 scAb 构建体插入该载体的 NdeI 位点，使得编码所述附加表位的序列位于 T7 基因 10 的第一个 ATG 之后。第

一种构建体包括抗铜绿假单胞菌(*Pseudomonas aeruginosa*)的 scAb (Molloy P.等, *Journal of Applied Bacteriology*, 78:359-365, 1995), 同时在 N 末端加入 FLAG 表位(MDYKDDDK) (Knappik A 和 Pluckthun A, *BioTechniques*, 17, 754-761, 1994), 在所述蛋白的 C 末端区加入 c-fos 的 b-拉链结构域(Abate, C 等 *Proc. Natl. Acad. Sci. USA*. 87:1032-1036, 1990)。第二种构建体包含由抗胎儿抗原抗体 340 (Durrant LG 等 *Prenatal Diagnosis*, 14:131-140, 1994)构建的 ScAb, 同时在该蛋白 N 末端具有多聚组氨酸标记、在该蛋白 C 末端区具有 c-jun 的 b-拉链结构域(Abate C.等, 出处同上)。

如下所述构建抗铜绿假单胞菌(α -Ps-fos) scAb 和 340-jun scAb:

用分别引入 5' FLAG 表位序列以及除去 3' 终止密码子的引物 RD

5' FLAG: 5'gcgatcccatatggactacaaagacgatgacgacaaacaggtgcagctgcag3' (Genosys Biotechnologies Europe Ltd, Cambridge, UK.)和 RD 3':

5'gcaattcgtggtggtggtggtggtggtgactctcc3' (Genosys), 扩增载体 pPM1His

(Molloy P.等, 出处同上)中的 α -Ps scAb 的 DNA。反应混合物包括 0.1

μ g 模板 DNA、2.6 单位 Expand™高保真性 PCR 酶混合物(Boehringer Mannheim, Lewes, UK.)、Expand HF 缓冲液(Boehringer Mannheim)、

1.5 mM MgCl₂、200 μ M 脱氧核苷酸三磷酸(dNTP) (Life Technologies,

Paisley, UK)和每种引物 25 皮摩尔。循环是 96 °C 5 分钟、然后是[95

°C 1 分钟、50 °C 1 分钟、72 °C 1 分钟] × 5、[95 °C 45 秒、50 °C 1

分钟、72 °C 1 分钟 30 秒] × 8、[95 °C 45 秒、50 °C 1 分钟、72 °C 2

分钟] × 5、以 72 °C 5 分钟结束。用 BamHI 和 EcoRI 切割所获得的 1123

bp 产物, 并将其克隆到载体 pUC19 (Boehringer Mannheim)中。采用

Thermo Sequenase 放射性标记终止子循环测序试剂盒与[³³P]二脱氧核

苷酸(Amersham Life Science, Amersham, UK), 确证该 DNA 序列。将该

构建体作为 NdeI-EcoRI 片段克隆到 pET5c 载体(Promega UK Ltd,

Southampton, UK.) (参见 *Molecular Cloning, A Laboratory Manual* Sambrook J, Fritsch EF, Maniatis T. 编著, Cold Spring Harbor Laboratory

Press 1989, New York, USA). 用 Wizard® Plus SV Minipreps DNA 纯化系统(Promega UK Ltd), 或者对于较大规模, 采用 Qiagen Plasmid Midi 试剂盒(Qiagen Ltd, Crawley, UK.), 制备质粒 DNA. 所产生的新质粒命名为 pET5c FLAG-αPs scAb.

5 通过以下重叠寡核苷酸的 PCR 装配 fos 盒:

Fos1for 5'-atggaattcctcgagaccgacaccctacaggcggaaaccgaccagctgga
Fos80rev 5'-tggcgatttcgggttcgagcgcggattttcgtcttcagctggtcggt
Fos 71for 5'-aaaccgaaatcgcaacctgctgaaagaaaaagaaagctggagttcatc
Fos 155rev 5'ggaagcttgaattccgcccggacgggtgtgccccaggatgaactccagctt

上述寡核苷酸以各 1 pmol 包括在反应混合物中, 采用高保真性聚合酶和如先前所述的反应组分, 用 10 pmol 引物 Fos1fS: 5'-atggaattcctcgagacc 和 Fos 155rS 5'-ggaagcttgaattccgcc 驱动该反应. 所产生的 155bp 产物用 EcoRI 消化, 将其纯化并克隆到 EcoRI 切割的 pUC19 中, 用于采用标准方法进行序列分析(参见 *Molecular Cloning, A Laboratory Manual* 出处同上). 通过取代现有的携带人恒定区结构域的 320bp XhoI-EcoRI 片段, 将 Fos 盒作为 XhoI-EcoRI 片段亚克隆到 pET5c FLAG-αPs scAb 质粒中.

15 通过用 340 抗体的 VH 和 VK 来代替 ppM1His 中的 α-Ps VH 和 VK, 产生 340 scAb. 用引物 5'cagctgcaggagtctggggaggcttag3' (Genosys) 和 5'tcagtagacggtgaccgaggttccttgacccagta3' (Genosys), 扩增 340 VH. 反应混合物包括 0.1 μg 模板 DNA、2.6 单位 Expand™高保真性 PCR 酶混合物、Expand HF 缓冲液、1.5 mM MgCl₂、200 μM dNTP 和每
20 种引物 25 pmole. 循环是 96 °C 5 分钟、然后是[95 °C 1 分钟、50 °C 1 分钟、72 °C 1 分钟] × 5、[95 °C 45 秒、50 °C 1 分钟、72 °C 1 分钟 30 秒] × 8、[95 °C 45 秒、50 °C 1 分钟、72 °C 2 分钟] × 5、以 72 °C 5 分钟结束. 用 PstI 和 BstEII 切割所获得的 357 bp 产物, 并将其克隆到 PstI 和 BstEII 切割的 pPM1His (参见 *Molecular Cloning, A Laboratory Manual* 出处同上)中. 同样, 用引物 5'gtgacattgagctcacacagtctcct3'和
25 5'cagcccgtttatctcgagcttggtccg3' (Genosys), 扩增 340 VK. 用 SstI 和 XhoI

切割 339bp 产物并克隆进 *Sst*I 和 *Xho*I 切割修饰的 pPM1His (在上面产生)中。如前所述, 采用 Thermo Sequenase 放射性标记终止子循环测序试剂盒与 [³³P]二脱氧核苷酸, 确证该 DNA 序列。用分别于 5'端引入 6 个组氨酸残基和除去 3'终止密码子的引物 RD 5' HIS:

5 5'gcgatcccatatgcaccatcatcaccatcaccaggtgcagctgcag3' (Genosys)和 RD 3' (在以上给出的), 扩增载体 pPM1His 中的 340 scAb 的 DNA。用于扩增的试剂和条件与用于α-Ps 构建体的完全一样。用 BamHI 和 EcoRI 切割所获得的 1114 bp 产物, 并将其克隆到载体 pUC19 (参见 *Molecular Cloning, A Laboratory Manual* 出处同上)中。如上所述确证所述 DNA
10 序列, 将该构建体作为 NdeI-EcoRI 片段克隆到 pET5c 载体中, 产生质粒 pEt5c HIS 340 scAb。

通过以下重叠寡核苷酸的 PCR 装配 jun 盒:

Jun1for 5'-atgagaattctcgagcgtatcgctcgtctggaagaaaagttaaaccct
Jun 85rev 5'-tagcgggtggaagccagttcggagttctgagcttcagggtttaactttt
Jun 71for 5'-tggctccaccgctaacaatgctgcgtgaacaggttgcagctgaaacag
Jun 146rev 5'-catgccaattcgtggttcataacttctgttcagctgagcaacc

上述寡核苷酸以各 1 pmol 包括在反应混合物中, 采用高保真性聚合酶和如先前所述的反应组分, 用 10 pmol 引物 Jun 1for-S: 5'-
15 atgagaattctcgagcg 和 Jun 146rev-S: 5'-catgccaattcgtggttc 驱动该反应。所产生的 146bp 产物用 EcoRI 消化, 将其纯化并克隆到 EcoRI 切割的 pUC19 中, 用于采用标准方法进行序列分析(参见 *Molecular Cloning, A Laboratory Manual* 出处同上)。通过取代现有的携带人恒定区结构域的
20 320bp *Xho*I-EcoRI 片段, 将 Jun 盒作为 *Xho*I-EcoRI 片段亚克隆到 pEt5c HIS 340 scAb 质粒中。

质粒 his-340-jun 和 FLAG-αPs-fos 用作采用生物素化引物 BioT7: 5'-agatctcgatccccgcaaatta 和引物 petrev:-5'-aaataggcgtatcacgaggcc 的 PCR
25 的模板。引物由 GenoSys (Cambridge, UK)供应, 并以 1pmol 的浓度用于该反应中。组分和 PCR 条件与前述相同。 his-340-jun 反应产物是 992bp, 而 FLAG-αPs-fos 反应产物是 1002bp。用旋转纯化柱(Qiagen,

Crawley, UK)纯化产物, 并将其稀释至 100ng/ μ l 的浓度。通过 260nm 下的 UV 吸光度定量。使 500ng 生物素标记的 DNA 与 10 μ l 链霉抗生物素蛋白包被的磁性粒子(Bangs labs, Fishers, USA)反应。反应在硅化微量离心管中的 500 μ l PBS 1% (w/v) BSA 体积中于室温进行 10 分钟。在结合后, 用磁铁(Dynal, Bromborough, UK)收获粒子, 将其用 PBS 1% BSA 洗涤 3 次。

在最后一次洗涤后, 通过加入补充有生物素酰基赖氨酸 tRNA (Promega)的 25 μ l T7 Quick 偶联转录翻译混合物(Promega, Southampton, UK), 起始体外翻译反应。翻译反应于 30 $^{\circ}$ C 进行 60 分钟, 然后置于冰上。用磁铁收获粒子, 并且采用含有 1% BSA 的冰冷 PBS 洗涤。

在某些实验中, 将非磁性链霉抗生物素蛋白粒子用于 IVTT 反应 (Bangs Labs, Fishers, USA)中。在这种情况下, 在洗涤循环期间通过离心回收粒子。

在某些实验中, 在 IVTT 反应中使用着色的磁性和非磁性链霉抗生物素蛋白粒子(Bangs Labs)。

在某些实验中, 用抗体检测结合于粒子上的反应产物的工程化到每种模型基因构建体中的或者 Flag 或者 his6 表位。将抗体加入在 PBS 中稀释的洗涤过的粒子中。在温和混合下于 4 $^{\circ}$ C 温育 60 分钟。将第二种试剂(抗小鼠-HRP 缓合物)以 PBS 中的推荐稀释液加入, 并且于 4 $^{\circ}$ C 再温育 30 分钟。在用生色底物显色之前, 用 200 μ l PBS 将粒子洗涤 3 次。反应物于 492nm 下读数。

用结合于粒子表面的 IVTT 蛋白进行蛋白-蛋白结合反应。在这类实验中, 通过蛋白介导的(fos:jun)与磁性粒子表面的结合, “捕获”非磁性链霉抗生物素蛋白粒子。将带有 fos IVTT 产物的磁性粒子与表面结合有 jun 的磁性粒子温和地混合。反应在 100 μ l PBS, BSA 中进行, 让其于室温进行 30 分钟。在一个阴性对照反应中, 将表面结合有 Sca 蛋白(α -Ps scAb; Molloy P 等, 出处同上)与包被有 fos IVTT 产物的磁性粒子混合。温育后, 用磁铁捕获所述粒子, 将其用 PBS, 1% BSA 洗涤

6 次。

采用 PCR 和 DNA 测序确证所捕获的靶蛋白基因的存在。为了检测 jun 模型基因，在 PCR 测定中使用 jun 特异性引物 Jun 1for-S 和 Jun146rev-。通过将 10% (v/v) 粒子加入 PCR 混合物中启动该测定。组分和反应条件如前所述。通过凝胶电泳检测 146bp jun 特异性产物。为了测定 fos 模型基因，在 PCR 测定中使用引物 Fos1fS 和 Fos 155rS。反应条件和 155bp fos 特异性产物的检测如上所述。为了测定阴性对照蛋白，用引物 Seq1scab 5'agatccctactataggtta 和 Seq2scab: 5'-ggtgagctcgatgatcc 来检测 α -Ps scAb 蛋白基因中的 115bp 产物。

在上述实验中，在与 fos 磁性粒子相互作用后不能检测到 α -Ps scAb PCR 产物的条件下，在用 fos 磁性粒子捕获后检测 Jun PCR 产物。

实施例 2

在该实施例中，产生包括独特肽“条形码”的单链抗体文库。按照标准方法制备人外周血淋巴细胞 RNA。简而言之，由取自 16 位正常健康供体的 10ml 肝素化血液制备淋巴细胞。在采用 Lymphoprep 介质(Sigma, Poole, UK)进行密度梯度离心程序后，收集淋巴细胞。用 QuickPrep 系统和由供应商(Pharmacia, St Albans, UK)提供的说明制备 RNA。用 cDNA 合成试剂盒(Pharmacia, St Albans, UK)和随机六聚体引物和供应商建议的条件，进行 cDNA 的合成。用设计用来将 Vh 和 V1 所有组成成分最大化的引物组，在分开的 PCR 混合物中由 cDNA 扩增免疫球蛋白重链可变区(Vh)和轻链可变区(V1)。引物组如先前所述(Marks J.D.等 1991, Eur. J. Immunol. 21:985)。用 2.6 单位 Expand™高保真性 PCR 酶混合物(Boehringer Mannheim, Lewes, UK.)、Expand HF 缓冲液(Boehringer)、1.5 mM MgCl₂、200 μ M 脱氧核苷酸三磷酸(dNTP)(Life Technologies, Paisley, UK)和每种引物库 25 pmole 进行 Vh 和 V1 PCR 反应。循环是 96 $^{\circ}$ C 5 分钟、然后是[95 $^{\circ}$ C 1 分钟、50 $^{\circ}$ C 1 分钟、72 $^{\circ}$ C 1 分钟] \times 5、[95 $^{\circ}$ C 45 秒、50 $^{\circ}$ C 1 分钟、72 $^{\circ}$ C 1 分钟 30 秒]

× 8、[95 °C 45 秒、 50 °C 1 分钟、 72 °C 2 分钟] × 5、以 72 °C 5 分钟结束。

5 在一个单独的 PCR 中，用先前详细描述引物(Marks, J. D, *Antibody Engineering*, Borrebaek C.A.K.编著, New York O.U.P., 1995), 由所克隆的模板 pSW1-ScFvD1.3 (McCafferty 等, 1990, *Nature* 348: 522-554)扩增形式(Gly₄Ser)₃的接头片段(Huston J.S.等, 1988, *PNAS*, 85: 5879-5883)。将 93bp 接头片段产物与 Vh 和 V1 PCR 产物的等摩尔混合物一起退火。采用 Vaughan 等详述的侧翼引物 HuVHBACKsfi 和 HuFORNot (Vaughan T.J.等 1996, *Nature Biotech.* 14: 309-314), 在“穿入(pull through)”反应中进一步扩增该混合物。将用于穿入反应中的所有片段在包括在该反应之前进行纯化，使其不含其起始引物。用 Promega 的 Wizard PCR Preps 系统(Promega, Southampton UK)进行纯化。

15 如上所述，采用标准条件用限制性酶 SfiI 和 NotI (Boehringer)消化形式 Vh-接头-V1 的装配重叠群，并且将其纯化。将所纯化的片段与一种双链合成寡核苷酸连接物混合物一起退火，所述混合物设计用以引入一个 V8 蛋白酶切割位点，该切割位点与 V1 基因 C 末端符合读框的一段随机序列重叠。通过一对合成寡核苷酸库 - 形式 5'-
ggccgcgaggaagaggaa[(atg)/(can)/(agn)/(aan)/(gan)/(ttn)]₂gc-3'和 5'-
20 ggccgc[(naa)/(ntc)/(ngt)/(nct)/(nag)/(cat)]₂ctcctctcctcgc-3'的退火，产生这种 V8/独特序列条形码。该接头具有 NotI 匹配末端(下划线)，因此便于将完整的单链抗体-V8/独特序列条形码片段插入 SfiI-NotI 制备的 pCANTAB 5 (Pharmacia)噬菌粒载体中。

25 所述独特序列条形码设计用来避免引入终止密码子，以及进一步偏向排除具有两种以上可变密码子的编码残基。采用该策略，将鉴定给定解码的肽序列所需要的具体寡核苷酸数目最小化。总之，所述独特序列条形码能够编码 20 种氨基酸中的 11 种。除 V8 肽酶切割位点(一串 4 个谷氨酸残基)外，所述序列条形码长 12 个密码子。因此从 11 种

氨基酸(其中 10 种由两种密码子中任一种编码)的所有组成成分, 能够编码 $11^{12/2} = \text{约 } 1.5 \times 10^{12}$ 种不同的肽。

5 将具有 SfiI 和 NotI 制备的末端的装配 scfv 片段(Vh-接头-V1)退火并连接至所述 NotI 序列-条形码连接物, 并进行再纯化。对于通过噬菌体展示表达人 scfv 文库的实验, 将整个片段连接到 SfiI-NotI 制备的 pCANTAB 5 (Pharmacia)噬菌粒载体中, 并且转化到感受态 TG1 大肠杆菌中。

10 对于采用体外转录和翻译(IVTT)的其它实验, 将装配的 scfv 文库亚克隆到 SfiI NotI 制备的 pCANTAB5-T7 中。该载体与市售的 pCANTAB5 相同, 只是其经过修饰, 以包括于位置 2235 的 HindIII 位点插入的 T7 启动子序列(ttaatacgactcactata)。通过将序列为 5' - agctaatacgactcactata 的双链合成 DNA 接头连接到 HindIII 切割并且去磷酸化的 pCANTAB5 中, 完成该修饰。用诊断性 PCR 选择含有 T7 启动子的重组克隆。

15 在连接和转化到感受态 TG1 大肠杆菌中后, 让细胞在 1ml SOC 培养基中生长 1 小时, 然后平板接种到含有 100ug/ml 氨苄青霉素的 TYE 培养基上。将菌落刮下, 接种到含有氨苄青霉素的 5ml 2x TY 肉汤中。用所述培养的文库制备用于 IVTT 反应的 DNA。

20 在体外翻译反应中使用 pCANTAB5-T7 Scfv 文库 DNA。采用 50 μ l 总体积中的 T7 Quick 偶联转录翻译混合物(Promega, Southampton, UK)和 10 μ g pCANTAB5-T7 Scfv 文库 DNA 进行 IVTT。翻译反应于 30 $^{\circ}$ C 进行 90 分钟, 然后置于冰上。在某些实验中, 用 35 S-甲硫氨酸掺入测定, 根据翻译产物的存在监测反应。反应物在用于结合测定和筛选测定之前贮存于 -70 $^{\circ}$ C。

25 该单链抗体文库用于与重组人 p53 蛋白(Oncogene Research Products-Calbiochem, Nottingham, UK)的结合反应中。将 IVTT 混合物在 PBS 中稀释 10 倍, 并用于与 96 孔微量培养板中固定化的人重组 p53 蛋白的结合测定中。将 p53 蛋白以在磷酸缓冲液中的 100 μ g/ml 的浓度

于 4 °C 保温过夜而将其固定化。用 PBS 0.5% (w/v) BSA 洗涤板，并将稀释的 IVTT 混合物加入测试孔和对照孔中进行结合。结合反应于 37 °C 进行 90 分钟。用 PBS-T (PBS + 0.05% v/v 吐温 20) 将板洗涤 3 次，然后进行 V8 蛋白酶消化 (Takara, Wokingham, UK)。从上清液中收集蛋白片段，并且进行大小分级分离，以排除 V8 蛋白酶和其它大片段，然后通过 MALDI-tof 进行分析。

MALDI-tof 片段分析鉴定出许多肽片段。用所述肽序列设计一组相应的合成寡核苷酸。将所述寡核苷酸用于基于 PCR 的单链文库筛选中。用 Pfu turbo (Stratagene Europe) DNA 聚合酶合成所述人单链抗体文库 DNA 中成员的互补链。在 15 轮热循环后，对产物进行 DpnI 消化。该步骤中排除混合物中的亲代质粒分子，以确保仅增殖新合成的引发产物。将 1 μl 该反应物转化到 TG1 感受态细胞中，然后平板接种到含 100 μg/ml 氨苄青霉素的 LB 平板上。按照标准方法，挑出单个克隆，放大培养并且制备 DNA。所述 DNA 直接用于包括 IVTT、抗原结合、V8 蛋白酶消化、MALDI-tof 片段分析的第二轮筛选中。2 轮筛选后，分离出 6 种结合重组 p53 的 scFv。

实施例 3

用 Fab 表达载体 pC5A8-03 进行本实施例中描述的实验，该载体的构建如下。载体 pC5A8-01 基于载体 pLITMUS28 (New England Biolabs, MA, USA)，后者提供一种诱导型 lac 启动子和一个 M13 复制起点。由针对 CD4 的人源化单克隆抗体 5A8 (Reimann KA 等, *Acids research and human retroviruses* 13, 11: 第 933 页, 1997) 的编码重链可变区 (VH) 和第一恒定区 (CH1) 以及 κ 链的可变区 (VK) 和恒定区 (CK) 的两种 DNA 片段，装配抗体的 Fab 区。如下所述，将这些片段融合于 pelB 前导序列 (Lei S-P 等 *Journal of Bacteriology*, 169: 9: 4379-4383, 1987)，并且插入 pLITMUS28 的 *Bgl*III 和 *Bst*98I 限制位点之间。所有以下分子生物学方法是本领域技术人员所熟悉的，并且可以在 *Molecular Cloning*,

A Laboratory Manual, Sambrook J., Fritsch EF.和 Maniatis T.编著, Cold Spring Harbor Laboratory Press 1989, New York, USA 中找到。所有寡核苷酸均由 Genosys Biotechnologies Europe Ltd, Cambridge, UK 合成。除非另有说明, 否则所有的限制性内切核酸酶均购自 Life Technologies, Paisley, UK。所有的聚合酶链式反应均采用 pfu DNA 聚合酶(Promega, Southampton, UK)进行。

为了装配所述轻链片段, 用 Hybaid Touchdown 热循环仪, 采用聚合酶链式反应, 从含有抗铜绿假单胞菌的单链抗体片段(scAb)的克隆 pPM1-HIS (Molloy, P.等, Journal of Applied Bacteriology, 78: 359-365, 1995)中, 扩增 pel B 前导序列。这一初始反应用寡核苷酸 OL001 和 OL002 进行, 寡核苷酸 OL001 编码一个 *Bgl*III 限制位点、pelB 前导序列的 N 末端残基和 SD 序列, 而寡核苷酸 OL002 编码 pel B 前导序列的 C 末端和来自 pDIVKV3 (文献(ref))的 κ 轻链的 N 末端残基。用 Wizard® PCR 纯化试剂盒(Promega UK Ltd, Southampton, UK), 从 NuSieve GTG 琼脂糖(Flowgen, Lichfield, UK)中纯化该反应的产物, 将其变性, 并且与编码所述 κ 轻链的可变区和恒定区的接点的 OL004 结合使用, 以用标准方案, 通过 PCR 从克隆 pDIVKV3 扩增所述 κ 轻链的可变区。用编码所述可变区 C 末端残基和所述恒定区 N 末端残基的 OL003 和编码所述 κ 轻链恒定区的 C 末端残基以及限制性酶切位点 *Eco*RI 的 OL005, 通过 PCR 从克隆 pPM1-HIS (Molloy 等人)扩增所述 κ 轻链的恒定区。随后用 OL001 和 OL005 通过重叠 PCR 扩增这两种片段, 用 *Bgl*III 和 *Eco*RI 消化, 并将其克隆到 pLITMUS28 中, 以产生 pC5A8-01。

通过用编码一个 *Eco*RI 位点和 SD 序列和 OL006 和编码 pel B 前导序列的 C 末端残基和来自 pDIVKV4 的重链 N 末端残基的 OL007, 从所述装配的轻链扩增 pel B 前导序列, 装配所述重链。使用该反应的产物以及编码所述重链可变区和恒定区接点的 OL009, 从克隆 pDIVHV4 扩增 IgG1 重链的可变区。用编码所重链可变区的 C 末端残

基和所述恒定链 N 末端(OL008)以及外显子 1 的 C 末端残基和 *SstI* 限制位点(OL010)的 OL008 和 OL010, 通过 PCR 从克隆 pSVgptHuIgG1 扩增所述重链恒定区的外显子 1。用 OL006 和 OL010, 通过重叠 PCR 扩增这些反应的产物, 用 *EcoRI* 和 *SstI* 消化, 并克隆到含有所述轻链片段的 pLITMUS28 中, 以产生 pC5A8-02。

用包括限制位点 *EcoI* CRI 和 *Bst* 98I 的 OL011 和 OL012, 加入 CH1 的 C 末端残基和 C 末端 FLAG 标记序列(DYKDDDDK) (Knappik A. 和 Pluckthun A. *Biotechniques*, 17; 754-761, 1990), 产生 pC5A8-03。或者, 这些标记可以包括 6HIS 标记或 MS 标记(参见实施例)。

以下列出用于产生 pC5A8-0、pC5A8-02 和 pC5A8-03 的寡核苷酸:

OL001: 5' GGGCAGATCTTTAACTTTAAGAAGGAGATATACATATGAAATACCTATTGCCTACGG 3'
OL002: 5' GGGTCTGGGTCATAACGATATCGGCCATGCTGGTTGGGCAGC 3'
OL003: 5' GGTACCAAACCTGGAGATCAAACGGACTGTGGCTGCACCATCT 3'
OL004: 5' AGATGGTGACGCCACAGTCCGTTTGATCTCCAGTTTGGTACC 3'
OL005: 5' GATCGAATTCCTAACACTCTCCGCGTTGAAGCTCTTTG 3'
OL006: 5' GATCGAATTCCTAACCTTTAAGAAGGAGATATACATATG 3'
OL007: 5' GGACTGAACCAAGTTGGACTTCGGCCATCGCTGGTTGGGCAGC 3'
OL008: 5' ACCCTGGTTACCGTCTCCTCAGCCTCCACCAAGGCCCATC 3'
OL009: 5' GATGGCCCTTGGTGGAGGCTGAGGAGACGGTAACCAAGGGTAC 3'
OL010: 5' GATCGAGCTCTGCTTTCTTGTCACCTTGGTGTTC 3'
OL011: 5' CCCAAATCTTGGCTGCAGACTACAAAGACGACGACAAATAGCTCGAGC 3'
OL012: 5' TTAAGCTCGAGCTATTGTGCTGCTGCTTTGTAGTCTGCAGCGCAAGATTGGG 3'

通过 ELISA 证明产生功能性 Fab。总之, 将上述载体转移到大肠杆菌菌株 DH5 α 中并在 100 μ g/ml 氨苄青霉素和 1% 葡萄糖存在下于 37 $^{\circ}$ C 生长直至达到 0.5 的 OD₆₀₀。通过在缺乏葡萄糖的情况下加入 1mM 异丙硫基- β -D-半乳糖苷(IPTG)诱导蛋白产生。用 30mM Tris HCl、20%蔗糖 pH 8.0、1mM EDTA, 然后用 5mM MgSO₄, 通过渗透休克释放周质部分(Molloy, P.等, *Journal of Applied Bacteriology*, 78: 359-365, 1995), 并将其直接加入已经在潮湿室中预先用磷酸缓冲盐溶液(PBS) pH 7.4 中的 1 μ g/ml 浓度的可溶性人 CD4 (Intracel Corp., Issaquah, WA) 于室温包被过夜的 Immulon 4 ELISA 板(Dynex,)中。或者, 可以通过细胞裂解或者通过加入 1mM EDTA 释放周质部分。通过在加入可溶性 Fab 之前, 将板与含有 0.05% 吐温 20、2% 牛血清白蛋白(BSA)和 0.05% 硫柳汞(Sigma)的 PBS 于室温温育 1 小时, 降低非特异性结合。用山羊

抗人 IgG Fab 特异性辣根过氧化物酶缀合物(Sigma, UK)检测抗 CD4 特异性 Fab , 所述辣根过氧化物酶缀合物本身用磷酸/柠檬酸缓冲液 pH 5.0 中的 5,5'-四甲基联苯胺二盐酸盐(TMB) (Sigma, UK)和过氧化氢检测。30 分钟后用 0.2N 硫酸终止显色反应, 于 450 nm 监测吸光度。或者用 ABTS/柠檬酸盐(Sigma, UK)进行检测。

为了产生 CDR3 序列的文库, 通过寡核苷酸定向诱变(Kunkel TA. Proc. Natl. Acad. Sci. USA: 488-492 (1985)和 Current Protocols in Molecular Biology, Ausubel FM, Brent R., Kingston RE. Moore DD., Seidman JG., Smith JA., Struhl K. 编著, John Wiley & Sons, Inc.), 用以下所列的寡核苷酸, 将独特限制位点引入载体 pC5A8-03 中。通过用合适的限制性酶消化, 证实分别在所述κ 轻链和重链中存在 AatII 和 HindIII (位于 LCDR3 的 5'和 3')以及 BssHII 和 SmaI (位于 HCDR3 的 5'和 3')限制位点。各含有一个额外限制位点的这些质粒被命名为 pC5A8-04 至 pC5A8-07。

OL013; 5' GAAGACGTCGCTGTTTAC 3'
OL014; 5' GGTACCAAGCTTGAGATC 3'
OL015; 5' CTACTGCGCGGTGAAAAAG 3'
OL016; 5' GGGTCAGGGGACCCTGG 3'

在用 AatII 和 HindIII 消化 pC5A8-07 后, 采用携带锚定残基(aa 83-88 和 aa 97-103)和于其 3'端包含 HindIII 内切核酸酶位点的 10 个核苷酸的回文序列的简并寡核苷酸混合物, 将所述κ 轻链可变区 CDR3 中的高度可变残基随机化。这些寡核苷酸于其 3'端杂交, 然后用作 DNA 聚合酶的底物, 导致产生双链同双链体, 将该双链体用所述两种限制性酶消化, 并采用标准方案将其克隆到所述消化的载体中(参见 Current Protocols in Molecular Biology, Ausubel FM, Brent R., Kingston RE. Moore DD., Seidman JG., Smith JA., Struhl K. 编著, John Wiley & Sons, Inc.)。制备所述寡核苷酸, 使得通过在所述寡核苷酸合成(Geneosys, Cambridge, UK)的每个步骤包括等浓度的每种寡核苷酸, 将残基 91、92、93、94、95、95A、95B 和 96 随机化。

可以使用另外两种标记。在这种情况下, 所述标记必须在长度上相差至少两个残基, 以便在用蛋白酶例如因子 Xa 取出标记 1 和标记 2 后进行区别。设计于 3' 端具有一个回文序列的寡核苷酸, 所述回文序列包括 *XhoI* 的限制性内切核酸酶位点。所述寡核苷酸于其 3' 端杂交, 然后用作 DNA 聚合酶的底物, 导致产生双链的同双链体, 将该双链体用两种限制性酶 *PstI* 和 *XhoI* 消化, 并采用标准方案将其克隆到所述消化的载体中(参见 *Current Protocols in Molecular Biology*, Ausubel FM, Brent R., Kingston RE, Moore DD., Seidman JG., Smith JA., Struhl K. 编著, John Wiley & Sons, Inc.)。

作为一个实例, 用寡核苷酸 5' NAC NCC NGG NTG TKC VAG GNV CNT 3' 产生 8 个残基的标记。如果由于加入蛋白酶因子 Xa 的位点(以斜体显示)导致也包括一个 8 个残基的第二标记, 则将该标记的长度增至 11 个残基。这使得所述标记在将其取出并进行质谱分析后, 可以被鉴定为标记 1 或标记 2。

单一标记

正向寡核苷酸: 5' GCG CTG CAG *GAY GGN CGN* NAC NCC NGG NTG TKC VAG GNV CNT TAG CTC GAG CTA 3'

反向寡核苷酸: 5' TAG CTC GAG CTA ANG BNC CTB GMA CAN CCN GGN GTN CCG CCC GTC CTG CAG CGC 3'

双标记

正向寡核苷酸: 5' GCG CTG CAG *GAY GGN CGN* NAC NCC NGG NTG TKC VAG GNV CNT *GAY GGN CGN* NAC NCC NGG NTG TKC VAG GNV CNT TAG CTC GAG CTA 3'

反向寡核苷酸: 5' TAG CTC GAG CTA ANG BNC CTB GMA CAN CCN GGN GTN CCG CCC GTC ANG BNC CTB GMA CAN CCN GGN GTN CCG CCC GTC CTG CAG CGC 3'

实施例 4

为了选择高亲和性结合剂, 通过电穿孔(Bio-Rad)将原始文库转移到大肠杆菌 DH5 α 中, 平板接种到含有 100 μ g/ml 氨苄青霉素和 1% 葡萄糖的 LB 琼脂上, 并且于 37 $^{\circ}$ C 温育过夜。收获转化细胞, 并用来接

种一批新鲜的含有 100 μ g/ml 氨苄青霉素的 L 肉汤。应该保留该文库的
5 剩余部分并且将其贮存于-70 $^{\circ}$ C, 如随后描述用作拯救高亲和性克隆的
原材料。在加入异丙硫基- β -D 半乳糖苷(IPTG)至 0.1mM 终浓度之前,
将新接种的培养物于 37 $^{\circ}$ C 温育 2 小时。然后将所述培养物于 37 $^{\circ}$ C 再
温育 3 小时。

将产生可溶性 Fab 文库的 100ml 细菌培养物于 4 $^{\circ}$ C 以 4000 rpm 离
心 20 分钟, 将产生的沉淀重悬于含有 1 mM EDTA 的磷酸缓冲盐溶液
中。在冰上搅拌 5-20 分钟后, EDTA 使外膜预透化, 使得渗漏出周质
内容物。然后通过离心将上清液澄清, 将该上清液用于随后的步骤。

10 也可以利用释放周质内容物的其它可选方案(Molloy, P.等, Journal of
Applied Bacteriology, 78: 359-365, 1995 和 Molecular Cloning, A
Laboratory Manual, Sambrook J., Fritsch EF.和 Maniatis T.编著, Cold
Spring Harbor Laboratory Press 1989, New York, USA)。

将含有所述 Fab 文库的周质提取物等份加入到已经在潮湿室中用
15 磷酸缓冲盐溶液(PBS) pH 7.4 中的 1 μ g/ml 浓度的可溶性人 CD4
(Intracel Corp., Issaquah, WA)于室温包被过夜的 Nunc-immunotube 中。
通过在加入可溶性 Fab 之前, 将所述管与含有 0.05%吐温 20、2%牛
血清白蛋白(BSA)和 0.05%硫柳汞(Sigma)的 PBS 于室温温育 1 小时,
降低非特异性结合。在让所述 Fab 与所述 CD4 抗原于室温结合 1 小时
20 后, 通过用 PBS, 0.05%吐温 20 将管洗涤 20 次去除未结合的 Fab。

为了鉴定保持结合的那些 Fab 的氨基酸序列, 采用标准方案用因
子 Xa 取出质量标记。然后通过 MALDI-TOF (MS/MS)光谱测定法分析
所述质量标记, 其中测定每种标记的的分子量, 然后通过分析第二次
电离事件获得序列信息。通过综合该信息, 可以确定所述标记的氨基
25 酸序列。

在某些情况下, 可能必需通过洗脱结合的 Fab、中和并通过用按
照生产商的说明制备的 sepharose-抗 Ck 柱(Pierce Warriner, Cheshire,
UK)进行亲和纯化、从其它大肠杆菌蛋白中纯化所述 Fab, 来提高蛋

白酶切割效率。然后可以用因子 Xa 从结合的 Fab 取出这些质量标记。

在鉴定质量标记后，产生另外两种寡核苷酸。所述 3'寡核苷酸编码所述质量标记的序列，而 5'寡核苷酸是于所述 Fab 的 N 末端编码所述序列的 OL001。

正链: 5' GG GCA GAT CTT TAA CTT TAA GAA GGA GAT ATA CAT
ATG AAA TAC CTA TTG CCT ACG G 3'

负链: 5' TAG CTC GAG CTA ANG BNC CTB GMA CAN CCN GGN GTN CCG CCC
GTC ANG BNC CTB GMA CAN CCN GGN GTN CCG CCC GTC CTG CAG CGC 3'

5

通过将 10 μ l 所述大肠杆菌文库加入含有上述寡核苷酸的 PCR 反应中，拯救含有高亲和性结合剂的克隆。该反应所需的条件可以根据所使用的寡核苷酸而变化。在扩增后，对 PCR 产物测序，随后从低熔点琼脂糖中纯化，用 *Aat*II 和 *San*DI 消化，*Aat*II 位点存在于所述 κ 轻链 CDR3 的 N 末端，而 *San*DI 位点存在于载体 pC5A8-07 中的重链 CDR3 的 C 末端；用标准方案(参见 Molecular Cloning, A Laboratory Manual, Sambrook J., Fritsch EF.和 Maniatis T.编著, Cold Spring Harbor Laboratory Press 1989, New York, USA), 将所述产物转移到已经用相同的限制性内切核酸酶消化的载体 pC5A8-07 中。采用标准方案，通过电穿孔将所产生的质粒转移到大肠杆菌 DH5 α 中，并贮存于-70 $^{\circ}$ C。或者，可以用多种替代的限制性内切核酸酶消化所述 PCR 反应产物，并将其转移到用于 Fab 表达的替代载体中。

10

15

在某些情况下，在最初一轮淘选后可以存在多种质量标记。在这种情况下，用 3'寡核苷酸混合物从所述贮存的文库中扩增一个克隆文库。然后可以对该有限文库进行另外的多轮淘选，可以 MALDI-TOF 对结合的克隆进行再分析，并且用内部标记的序列产生一种有限的 PCR 引物的所有组分成分。

20

为了证实选定的抗 CD4 特异性 Fab 的亲和性，应该如上所述制备周质提取物，并且立即用于 CD4 特异性 ELISA 中。表观亲和性是真实亲和性和所述 Fab 浓度的综合，因此应该通过在同一提取物上进行另

25

外的捕获 ELISA 来确定所述 Fab 的浓度,其中针对 FLAG 标记或人 Ck 结构域作标准浓度曲线(McGregor DP., Molloy PE., Cunningham C.和 Harris WJ. *Molecular Immunology* 31, 219-116, 1994).

5 实施例 5

在该实施例中,用化学标记于 N 末端修饰人 p53 蛋白,用蛋白酶切割,然后用标记特异性单克隆抗体回收化学标记的肽,通过 MALDI-ToF 分析所述肽。p53 蛋白是来自 Borek Vojisek 博士 (University of Brno, Czech Republic)的赠品。按照 Mikolajczyk 等, 10 *Bioconjugate Chem.*, 第 7 卷(1996) 第 150-158 页,用碳酸(甲基磺酰)乙酯的琥珀酰亚胺酯处理 100ug p53 蛋白,以封闭赖氨酸侧链。将所述被封闭的蛋白以 1mg/ml 溶于 0.1M 碳酸氢钠缓冲液 pH 8.5,将 NHS-SS-生物素(Pierce, Chester, UK)加至最终 100ug/ml。该反应于室温进行 6 小时,并用乙醇胺终止反应。然后使该蛋白混合物在 PBS 中通过 15 Sephadex G25 柱(Pharmacia, Milton Keynes, UK),采用对洗出液的 A280 测量收集空隙容积。然后将含有 2ug p53 的 40ul 洗出液热变性(95 °C 5 分钟)、冷却至 37c,然后加入 1ug 内切蛋白酶 Arg-C (得自溶组织梭菌(*C. histolyticum*), Calbiochem, Nottingham, UK),并且将混合物于 37c 温育 1 小时。然后加入 10ul PBS 中的链霉抗生物素蛋白-琼脂糖 20 (Sigma, Poole, UK),并且将混合物震荡 10 分钟。将琼脂糖以 16000g 沉淀 1 分钟,并且在 TSO 缓冲液(75mM Tris.HCl, 200 mM NaCl, 0.5% N-辛基糖苷, pH 8)中洗涤 3 次,然后在 TSMK (10mM Tris.HCl, 200 mM NaCl, 5mM 2-巯基乙醇, pH 8)中洗涤 3 次。最后,将 10ul α -氰基-4-羧基肉桂酸在 1%三氟乙酸/乙腈水溶液(1:1 v/v)中的饱和溶液加入洗涤过的 25 珠粒中,并将 1ul 该物质上样至质谱仪芯片上。用 Perceptive Biosystems Voyager-DE STR Biospectrometry Workstation (Perceptive Biosystems)进行该分析。或者加入得自 200 激光单脉冲的波谱而收集质谱。

结果显示出一个对应于 65 个氨基酸的 N 末端 Arg-C 内切蛋白酶片段的主峰，没有显著水平的其它 p53 Arg-C 峰。

实施例 6

5 重复实施例 5 的方法，只是用所述 N 末端生物素标记的肽，从单链 Fv 的噬菌体展示文库分离一种单链 Fv 抗体片段。随后，用该单链 Fv 从通过 MALDI-ToF 证实的所述受试蛋白的蛋白酶消化物中分离 N 末端肽片段。按照 Harlow 和 Lane, "Antibodies" (1988) (Cold Spring Harbor Publications), 将按照实施例 4 制备的正常人脑的提取物与 KLH
10 缀合，并且用来免疫 2 只 BalbC 小鼠。经腹膜内给予 2 剂，2 剂之间间隔 4 周。在第二次接种后 3-4 天，处死小鼠并且通过解剖取出脾脏。用 QuickPrep™ mRNA 纯化试剂盒(Pharmacia), 按照生产商的说明，启动脾 mRNA 的制备。

用 Pharmacia 重组噬菌体抗体系统(Pharmacia)产生小鼠单链 Fv (ScFv)文库。用 M-MuLV 反转录酶和随机六聚体引物，由所述 mRNA 产生第一链 cDNA。然后用互补于所述抗体可变区侧翼的保守序列的特异性重链引物和轻链引物，扩增抗体的重链和轻链基因。在琼脂糖凝胶电泳后，分别纯化所产生的分别为 340 个碱基对和 325 个碱基对的重链 DNA 和轻链 DNA 的产物。然后用 DNA 接头-引物混合物产生
20 通过(Gly4Ser)3 肽与 VL 区连接的 VH 区，将这些产物装配成单一 ScFc 构建体。用设计分别于 5'端和 3'端插入 Sfi I 和 Not I 位点的引物，扩增所装配的 ScFv，产生一种 800 bp 产物。将该片段纯化，用 Sfi I 和 Not I 顺序消化，然后再纯化。然后将片段连接到 Sfi I 和 Not I 切割的 pCANTAB 5 噬菌粒载体中。pCANTAB 5 含有编码噬菌体基因 3 蛋白 (g3p)的基因，将该 ScFv 邻近 g3 信号序列插入，使得它将作为 g3p 融合蛋白表达。用 pCantab 5/ScFv 噬菌粒转化感受态大肠杆菌 TG1 细胞，随后用 M13KO7 辅助噬菌体感染。所产生的重组噬菌体含有编码
25 所述 ScFv 基因的 DNA，并且在其尖端上展示一个或更多个拷贝的作

为融合蛋白的重组抗体。

通过淘选选择或富集与所述肽结合的噬菌体展示的 ScFv。简而言之，将来自实施例 1 的生物素酰化和经蛋白酶处理的 p53 制备物应用于链霉抗生物素蛋白包被的玻片(Radius Biosciences, Waltham, USA)上，将玻片在 PBS 中洗涤 4 次。用 2%脱脂乳粉的 PBS 溶液封闭后，应用所述噬菌体制备物并温育 1 小时。用 TBS/0.05%吐温 20 洗涤 10 次后，用辣根过氧化物酶缀合的抗 M13 抗体检测肽反应性重组噬菌体，并用邻苯二胺显色底物显示。随后用 0.1M 甘氨酸.Hcl pH 2.2 和 1mg/ml BSA 稀释这些噬菌体，并且用 2M Tris 碱中和。在于 25ml J 肉汤中生长的 JM103 中扩增经稀释的噬菌体。进行另外 2 轮淘选，最后分离出 10 个单个噬菌斑，将其合并并进一步纯化。将 1 等份 10^{10} 个扩增的噬菌体于 4c 与 TSO 缓冲液中的 0.1ug 生物素酰化和内切蛋白酶 Arg-C 消化的 p53 一起温育 2 小时。2 小时后，加入 TSO 中的 0.5ug 抗 M13 (Pharmacia)并且温育 1 小时，此后加入 5ul A 蛋白/G 琼脂糖 (Sigma)，在旋转下将混合物再温育 0.5 小时。然后将琼脂糖珠粒沉淀，如以上实施例 1 中所述进行洗涤，并且通过质谱法分析。

结果显示出与实施例 1 中相同的对应于 65 个氨基酸的 N 末端 Arg-C 内切蛋白酶片段的主峰。

20 实施例 7

在该实施例中，用编码多聚组氨酸标记的合成寡核苷酸引发编码受试蛋白的基因片段。使所述 cDNA 通过体外转录和翻译(IVTT)表达，然后用镍螯合柱分离标记的肽片段。然后用这些片段分离单链 Fv 抗体片段。随后，用所述单链 Fv 从经质谱法证实的所述受试蛋白的蛋白酶消化物分离出一种肽片段。

实施例 8

用来自细胞的总蛋白制备物重复实施例 6 的方法，并且用所述化

学标记的肽分离单链 Fv 抗体片段集合。随后，用 12 种这些单链 Fv 的混合物从所述受试蛋白的蛋白酶消化物中分离肽片段，并且通过质谱法分析。

专利名称(译)	蛋白质的分离和分析		
公开(公告)号	CN1344370A	公开(公告)日	2002-04-10
申请号	CN00805178.X	申请日	2000-03-17
[标]发明人	FJ卡尔		
发明人	F·J·卡尔		
IPC分类号	G01N27/62 C07K1/04 C07K16/12 C07K16/18 C07K16/28 C12N15/02 C12N15/10 C12Q1/37 C40B30/04 G01N33/53 G01N33/532 G01N33/545 G01N33/554 G01N33/566 G01N33/68 C07K16/00		
CPC分类号	C07K2317/34 C07K2319/00 C40B30/04 G01N33/6845 C07K16/1214 C12Q1/37 C07K16/18 C07K2317/55 C07K16/2812 G01N33/6851 G01N33/532 C07K2317/622 G01N33/6848 C07K1/047 G01N33/6842 C12N15/1065		
代理人(译)	刘玥		
优先权	1999016511 1999-07-14 GB 1999007057 1999-03-29 GB 1999020503 1999-08-31 GB 1999014874 1999-06-28 GB 1999015363 1999-07-02 GB 1999022285 1999-09-21 GB 1999007641 1999-04-06 GB 1999006551 1999-03-23 GB 1999015677 1999-07-06 GB		
外部链接	Espacenet	SIPO	

摘要(译)

提供用于蛋白鉴定和/或测序的新方法。这些方法特别适合于筛选抗体文库,在优选实施方案中,利用质谱技术进行直接或间接测序。

	1	2	3	4	5	6	7	8
肽	A	C	E	G	I	K	M	O
	B	D	F	H	J	L	N	P
	C	E	G	I	K	M	O	Q