

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

C12N 15/10

C12N 15/66 C12N 15/62

C07K 1/13 C07K 16/00

C12Q 1/68 G01N 33/68



[12] 发明专利申请公开说明书

[21] 申请号 01815706.8

[43] 公开日 2004 年 5 月 5 日

[11] 公开号 CN 1494589A

[22] 申请日 2001.8.17 [21] 申请号 01815706.8

[30] 优先权

[32] 2000.8.17 [33] GB [31] 0020357.0

[32] 2000.11.14 [33] US [31] 60/247,995

[86] 国际申请 PCT/GB2001/003693 2001.8.17

[87] 国际公布 WO02/27327 英 2002.4.4

[85] 进入国家阶段日期 2003.3.14

[71] 申请人 森斯普罗特米克有限公司

地址 英国剑桥

[72] 发明人 J·M·布莱克伯恩

M·A·马尔德 M·萨马德尔

R·科兹洛斯基

[74] 专利代理机构 中国国际贸易促进委员会专利
商标事务所

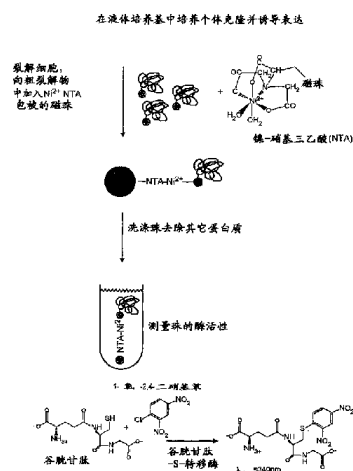
代理人 程泳

权利要求书 4 页 说明书 26 页 附图 4 页

[54] 发明名称 方法

[57] 摘要

本发明涉及生产如下蛋白质的新方法：其一个或多个结构域是全长且正确折叠的，在 N 端或 C 端附加有一种或多种标记部分，并且涉及含有这些蛋白质的阵列，以及阵列中的这些蛋白质在快速筛选中的用途。



1. 一种生产一种或多种蛋白质的方法，这些蛋白质中一个或多个结构域是全长且正确折叠的，在 N 端或 C 端均附加有一种或多种标记部分，该方法包括：

(a) 产生一种或多种 DNA 分子，其含有编码该蛋白质的开放阅读框以及 5' 和/或 3' 非翻译区；

(b) 在向子 DNA 分子中统计学意义上掺入 α -S-dNTP 及 dNTP 的条件下扩增该 DNA 分子；

(c) 特别保护该 DNA 分子的 5' 或 3' 端免于核酸酶消化；

(d) 在可以除去该开放阅读框的 5' 或 3' 非翻译区，包括起始或终止密码子，的条件下，首先用 5'-3' 或 3'-5' 核酸酶处理该 DNA 分子，产生一组嵌套缺失，随后用单链核酸酶处理；

(e) 将步骤 (d) 产生的片段克隆到含有一种或多种 5' 或 3' 标记部分的编码序列的表达载体中；

(f) 表达该编码蛋白质。

2. 如权利要求 1 所述的方法，其中该 DNA 分子的扩增在统计学意义上掺入一种 α -S-dNTP。

3. 如权利要求 2 所述的方法，其中一种 α -S-dNTP 是 α -S-dTTP 或 α -S-dATP。

4. 如权利要求 1-3 中任一项所述的方法，其中该核酸酶是外切核酸酶 III 或 λ 外切核酸酶。

5. 如权利要求 1-4 任一项所述的方法，其中该单链核酸酶是绿豆核酸酶或 T4 DNA 聚合酶。

6. 如权利要求 1-5 中任一项所述的方法，其中标记部分可以证实所述开放阅读框的表达。

7. 如权利要求 1-5 中任一项所述的方法，其中标记部分可证实所述开放阅读框的折叠。

8. 如权利要求 1-7 中任一项所述的方法，其中标记部分编码绿色

荧光蛋白。

9. 如权利要求 1-7 中任一项所述的方法，其中标记部分是一种肽序列，例如六组氨酸标签，一种完整蛋白质或蛋白质结构域，例如麦芽糖结合蛋白域。

10. 如权利要求 9 所述的方法，其中该标签允许纯化阵列中的各种蛋白质。

11. 如权利要求 1-10 中任一项所述的方法，其中插入标签，替换每种蛋白质的起始或终止密码子。

12. 如权利要求 1-10 中任一项所述的方法，其中该标签符合阅读框地插入紧接对于折叠和功能并不重要的每种蛋白质末端的区域中。

13. 如权利要求 1-10 中任一项所述的方法，其中该标签符合阅读框地插入开放阅读框内，但在对于折叠和功能并不重要的特异结构域边界之外的区域内。

14. 如权利要求 1-13 中任一项所述的方法，其中步骤(a)中 DNA 分子的扩增利用非校读聚合酶，例如 Taq 聚合酶或 DNA 聚合酶 I Klenow 片段。

15. 如权利要求 1-14 中任一项所述的方法，其中 α -S-dNTP 与 dNTP 之比为 1:1-1:3。

16. 如权利要求 1-15 中任一项所述的方法，其中该 5'-3'或 3'-5'核酸酶不能水解 α -S-磷酸二酯键。

17. 如权利要求 1-16 中任一项所述的方法，其中该 DNA 分子是通过反转录由 mRNA 序列产生的 cDNA。

18. 如权利要求 1-17 中任一项所述的方法，其中对多种 DNA 分子平行施行该方法。

19. 如权利要求 1-18 中任一项所述的方法，其中对一批中的 DNA 分子群体单独施行该方法。

20. 利用权利要求 1-19 中任一项所述的方法生产的经标记的蛋白质文库。

21. 一种生产蛋白质阵列的方法，该方法包括：

- (a) 克隆分离权利要求 20 的文库的每一个成员;
- (b) 以空间分离的形式表达各种经标记的蛋白质;
- (c) 借助标记部分纯化每种经标记的蛋白质;
- (d) 将每种蛋白质沉积于空间确定的阵列上。

22. 利用如权利要求 1-19 中任一项所述的方法制备, 或者利用权利要求 21 所述的方法生产的一种含蛋白质阵列。

23. 如权利要求 22 所述的阵列, 其中该阵列的成分例如固定于固体表面。

24. 如权利要求 22 或 23 所述的阵列, 其中借助标签部分固定每种蛋白质。

25. 一种根据生物活性筛选一种或多种化合物的方法, 包括下列步骤: 使一种或多种该化合物接触如权利要求 22-24 中任一项所述的蛋白质阵列, 测定一种或多种该化合物与阵列中蛋白质的结合。

26. 一种根据特定蛋白质-蛋白质相互作用筛选一种或多种蛋白质的方法, 包括下列步骤: 使一种或多种该蛋白质, 例如细胞表面受体, 接触如权利要求 22-24 中任一项所述的阵列, 测定一种或多种该特定蛋白质与阵列蛋白质的结合。

27. 一种根据特定蛋白质-核酸相互作用筛选一种或多种蛋白质的方法, 包括下列步骤: 使一种或多种该核酸探针接触如权利要求 22-24 中任一项所述的阵列, 测定该探针与阵列中蛋白质的结合;

28. 如权利要求 22-24 中任一项所述的阵列在快速筛选化合物、蛋白质或核酸中的用途。

29. 如权利要求 22-24 中任一项所述的阵列在筛选可识别阵列中每种蛋白质的分子中的用途, 其中该分子优选地是抗体。

30. 一种生产抗体阵列的方法, 包括使如权利要求 22-24 中任一项所述的蛋白质阵列接触一种抗体文库, 使得蛋白质阵列中的一种或多种蛋白质结合该抗体文库中的至少一种抗体, 除去所有未结合的抗体, 固定与蛋白质阵列中的蛋白质结合的那些抗体。

31. 一种筛查蛋白质功能或丰度的方法, 包括使如权利要求 30 所

述的抗体阵列接触一种或多种蛋白质的混合物的步骤。

32. 如权利要求 25-27 或 30 中任一项所述的方法，该方法也包括首先产生如权利要求 22-24 中任一项所述的蛋白质阵列的步骤。

33. 如权利要求 21 所述的方法，其中阵列中的蛋白质在一步中纯化并固定。

34. 利用此处所述方法生产的经标记的蛋白质的用途。

35. 利用此处所述方法生产的经标记的蛋白质在分析所表达的蛋白质与其它蛋白质之间相互作用中的用途。

36. 利用此处所述方法生产的经标记的蛋白质在固定于亲和柱/基质上的用途，例如，允许通过亲和层析纯化：a) 相互作用的蛋白质，b) DNA，或 c) 化学化合物。

37. 利用此处所述方法生产的经标记的蛋白质作为诊断工具在通过亲和纯化固定中进行抗体筛查（ELISA 测定）中的用途。

38. 利用此处所述方法生产的经标记的蛋白质作为 cDNA 微阵列探针的用途。

39. 如权利要求 38 所述的用途，用于 DNA 结合蛋白的鉴定。

40. 利用此处所述方法生产的经标记的蛋白质在阐明“蛋白质组”中蛋白质身份中的用途。

41. 如权利要求 40 所述的用途，其中对通过本发明的方法修饰的源文库或原材料的表达蛋白质成分进行质谱分析。

方法

发明领域

本发明涉及生产如下蛋白质的新方法：其一个或多个结构域是全长且正确折叠的，在 N 端或 C 端附加有一种或多种标记部分，并且涉及含有这些蛋白质的阵列，以及这些阵列在快速筛选中的用途。

发明背景

基因组作图计划给治疗靶标的发现方法和药物发现方法带来了革命。在鉴定新的治疗靶标后，现有的组合化学文库的高通量筛选将提出对这些靶标有活性的多种潜在先导化合物。对于通过较早期临床试验研究所有先导化合物这显然是不经济的；然而目前还没有用于评价这些先导化合物对生物中所有蛋白质的可能活性谱的快速方法。如果有的话，这种方法允许在早期估计所有先导化合物的可能的毒理学谱，此信息将显著加快决定研究及不研究哪种先导化合物的过程。

在制药工业中另外还需要鉴定现有药物（已经上市的或仍在研制中的）的所有靶标，从而确定其作用机制。这种信息的获得将大大加快新药获管理机构批准的过程，因为越来越清楚，管理机构现在认为对作用机制的了解极为重要。另外，这类信息也能用于设计改进的第二代药物。这是因为大多数药物至少有较小的副作用，这或许是由于药物或其代谢产物与不希望的靶标结合引起的；所有这些靶蛋白都需要鉴定，以确定设计改良药物所需的标准。然而目前还没有获得这种信息的简单方法，价值数百万美元的大量潜在药物仅仅由于缺乏对作用靶点的了解而遭放弃。

愈加认为蛋白质-蛋白质相互作用在负责细胞对内部和外部应激的反应方面非常重要。因此，特异蛋白质-蛋白质相互作用是药物介导干预传染病和其它病症的可能靶点。目前，酵母双杂交试验是测定

蛋白质-蛋白质相互作用的唯一可靠方法，但这种体内测定甚至在非高通量形式时也不易与蛋白质-蛋白质相互作用的特异激动剂或拮抗剂的鉴定相容。功能蛋白质组表达阵列，或“蛋白质组芯片”，允许以体外形式测定蛋白质-蛋白质相互作用的特异性和任何药物介导的作用的特异性。因此具有巨大的潜能，因为它们给该研究领域真正带来革命。

产生功能蛋白质组阵列的一种方法是分别克隆、表达、纯化和固定在特定蛋白质组中表达的所有蛋白质。但是，开始时需要考虑的一个重要问题是目的基因组的绝对大小，以及完整基因组序列数据的获得。

作为这一点的说明，一个典型细菌基因组约为 5Mbp，少量现已经完全测序（例如，幽门螺杆菌（*Helicobacter pylori*）、大肠杆菌和结核分枝杆菌（*Mycobacterium tuberculosis*））；真菌基因组一般约为 40Mbp，哺乳动物基因组约为 3Gbp，植物基因组约为 10Gbp。目前估计，人类基因组序列将在 2003 年前后完成，尽管有多少信息将要公开还值得怀疑。希望在短期限内获得典型模式生物之外的任何生物的基因组显然是完全不切实际的，然而从功能蛋白质组学的前景来看，模式生物只有有限的价值。因此，虽然原则上在后 4 年内有可能设计并合成用来从 cDNA 文库中克隆人类基因组中全部约 100 000 种基因的引物，但实际上这将是非常昂贵（仅引物的成本就将达到数百万美元）并且非常艰苦的过程，即使必要的序列数据库可以获得。

但是，那些无法获得完整序列数据的药物相关生物又怎么办呢？功能蛋白质组学不能简单地忽略它们，那么替代方法是什么呢？表达 cDNA 文库原则上能与非特异固定一起用来产生蛋白质阵列，但这种技术明显受限于由于蛋白质折叠遭破坏，非特异固定常常导致功能丧失的事实。另外，所有宿主细胞蛋白质都可以固定，最好则显著降低信噪比，最差则导致阳性结果混乱。因此，能生产功能蛋白质组阵列（其中多种蛋白质特异固定并通过共有基序或标签纯化，而不影响其功能，并且不需要知道其完整基因组序列）是功能蛋白质组学领域的

一个巨大进步。

发明者现在发展了一种新方法，解决了上述问题，它提供的方法可使蛋白质组中的每种蛋白质在蛋白质内的特定位置附加一种通用标记，而不需要预先知道相应基因的 DNA 序列。然后能用这种“标签”使下游固定和纯化方法具有通用性和特异性，随后能产生空间确定的阵列，其展示来自特定蛋白质组的数千种蛋白质。

此处的一个重要问题涉及“标签”的精确定位。如果标签符合阅读框地插入任何基因的随机、不确定的位置，产生的经标记的蛋白质可能以一种不确定的方式截短，在大多数情况下正确折叠，因而功能遭到破坏。通常发现，全长蛋白质在 N 端和 C 端之一（或两端）处含有短多肽延伸，它们可被截短但不影响折叠或功能。然而，如果这种截短去除任何 N 端或 C 端延伸，并与结构域边界交叉，蛋白质的折叠和功能通常受损。此处所述的方法允许将标签精确插入位于每种蛋白质 N 端或 C 端的正确阅读框内，或者插入靠近任一端且对于蛋白质的折叠和功能并不重要的区域内，使得各种经标记的蛋白质在特异固定于阵列中时能够正确折叠，从而保留功能。对于不同结构域有不同功能的多结构域蛋白质，此处所述的方法也允许将标签插入整个编码序列之内，但位于特定结构域边界之外，使得各个标记结构域在特异固定于阵列中时正确折叠，从而保留功能。

由于阵列中的每种蛋白质都具有完全的功能，因此能直接筛查这些阵列，以确定药靶和其它生物学相关分子。阵列的空间确定可将每种蛋白质的表型与其基因型直接相关，从而可以确定“命中（hit）”。

发明内容

因此，第一方面，本发明提供一种生产一种或多种蛋白质的方法，这些蛋白质中一个或多个结构域是全长且正确折叠的，在 N 端或 C 端均附加有一种或多种标记部分，该方法包括：

（a）产生一种或多种 DNA 分子，其含有编码该蛋白质的开放阅读框以及 5' 和/或 3' 非翻译区；

(b) 在向子 DNA(daughter DNA)分子中统计学意义上掺入 α -S-dNTP 及 dNTP 的条件下扩增该 DNA 分子;

(c) 特别保护该 DNA 分子的 5'或 3'端免于核酸酶消化;

(d) 在可以除去该开放阅读框的 5'或 3'非翻译区(包括起始或终止密码子)的条件下,首先用 5'-3'或 3'-5'核酸酶处理该 DNA 分子,产生一组嵌套缺失(nested deletion),随后用单链核酸酶处理;

(e) 将步骤(d)产生的片段克隆到含有一种或多种 5'或 3'标记部分的编码序列的表达载体中;

(f) 表达该编码蛋白质。

优选地, DNA 分子的扩增在统计学上掺入一种 α -S-dNTP, 更优选地掺入 α -S-dTTP 或 α -S-dATP。

标记部分可以是一种肽序列, 例如六组氨酸标签, 抗体表位或拟生物素, 或者实际上是一种完整蛋白质, 或蛋白质结构域, 例如麦芽糖结合蛋白域。标记部分本身可以被翻译后修饰, 例如通过添加生物素或脂类分子。在一个优选实施方案中, 标记部分也允许纯化“标记的”蛋白质。

因此, 本发明的方法可以不依靠对多种基因序列的任何了解, 而成批的(in one pot)特异修饰 cDNA 文库的每一个成员。事实上, 它依靠核酸酶对每种 cDNA 的非渐进性截短, 使得每种 cDNA 的 5'或 3'非翻译区被去除。然后将编码已知标记部分的其它已知 DNA 序列附加到产生的每种 cDNA 的嵌套缺失上。如果标记部分与单个 cDNA 处于同一阅读框内, 并且它之前不是任何阅读框内终止密码子, 则通过本发明的方法产生的每种遗传修饰 cDNA 将编码一种含有一个共同部分(例如与 N 端或 C 端融合的多肽“标签”)的蛋白质。对正确折叠的经标记的蛋白质进行筛选可以去掉与结构域边界交叉并且影响蛋白质折叠(因此影响功能)的所有截短体, 以及与标签融合的所有阅读框外融合体。

由于 cDNA 文库的每个成员都以相同方式修饰, 最终结果是, cDNA 文库编码的每一种蛋白质都在 N 端或 C 端附加有一个共同部分。

由 cDNA 文库表达的蛋白质通常被“标记”，这易于鉴定和分离。待其纯化后，例如，即可附着于微阵列上。借助于标签本身，或者借助于先与蛋白质结合的另一部分，能够实现这种附着。

通过此处所述方法形成的阵列构成了本发明的第二方面。这些阵列含有通常固定于固体载体上的“标记”蛋白质表达文库。熟练技术人员应当理解，有多种固体载体常用于阵列领域，任何一种这类“基质”都能用于生产本发明的阵列。

如此处所述，术语“蛋白质阵列”是指一种或多种蛋白质部分以一种模式在一种表面上的空间确定的排列。优选地，蛋白质部分直接或间接附着于该表面。这种附着可以是非特异的（例如，物理吸附于表面，或形成非特异的共价相互作用）。在一个优选实施方案中，利用此处所述的方法使蛋白质部分通过与每个蛋白质连接的共同标记部分附着于表面。

在另一个实施方案中，蛋白质部分可掺入被表面限定的小泡或脂质体中。

因此，例如，该模型的每个位置都可含有下列物质的一个或多个拷贝：

a) 一种蛋白质类型的样品（形式为单体、二聚体、三聚体、四聚体或更高的多聚体）；

b) 与相互作用分子结合的一种蛋白质类型的样品（例如 DNA、抗体、其它蛋白质）；

c) 与合成分子结合的一种蛋白质类型的样品（例如肽、化学化合物）；或

d) 位于阵列模型每个位置上的 2-100 种不同标记蛋白部分的混合物。

支持阵列的表面可通过如化学处理包被/衍生化。合适的表面的例子包括载玻片、聚丙烯、聚苯乙烯、硅、金或金属支持物，或由例如硝酸纤维素、PVDF、尼龙或磷酸纤维素制成的膜。

如此处所述，本发明的方法可将特定蛋白质组内的所有蛋白质在

N 端或 C 端特异添加标签。一些蛋白质可能不耐受 N 端延伸，而其它氨基酸可能不耐受 C 端延伸，但是绝大多数蛋白质都可耐受一种或其它这种延伸。然而，现有的文库克隆方法完全不能解决这一问题，因为它们将基因克隆为全长、未修饰的 cDNA，或与某些蛋白质配偶体融合的随机并且几乎必然截短的融合体。与后者相比，本方法可以使标签位置定向于位于或靠近 cDNA 产物 N 端或 C 端残基的序列，例如，使得与希望的肽配偶体的融合不影响 cDNA 产物的折叠或功能。与前者相比，如此处所述将蛋白质固定于阵列中的方法是通过特异的而不是非特异的相互作用，这些特异相互作用是添加到每个 cDNA 末端的标签的功能。另外，此处所述的方法也能用来筛选纯化、固定的在非细菌宿主生物中表达的蛋白质，通过正确折叠和翻译后修饰有助于保持功能，而现有方法如噬菌体展示或 λ -cDNA 表达文库只限于细菌宿主，发现大多数真核蛋白质在其中以非功能形式合成，这是由于错误的折叠或不正确的翻译后修饰。

本发明的方法具有广泛的潜在体外用途，大体可分为三个主要领域：蛋白质-配体相互作用的研究、蛋白质-蛋白质相互作用的研究、蛋白质-DNA 相互作用的研究。

蛋白质-配体相互作用

此处所述的方法可以快速分析特定新化学实体与特定蛋白质组中所有蛋白质的相互作用。这能通过以不同严格性用 NCE 探查适当的蛋白质组阵列简单地实现，其中可考虑反相高通量筛选。这种筛选的结果可直接用于多种情况，其中一些在以下描述。

对细胞或完整生物检测化合物文库的高通量筛选程序通常用来鉴定先导化合物，这可以在筛选前不知道靶标的情况下引起表型改变。然而，随后鉴定主要靶标是一个非常艰苦的过程。本发明的方法能直接用于这类问题，因为能产生有关物种的功能蛋白质组阵列，然后用先导化合物筛查该阵列，确定其靶向蛋白质组中的哪种蛋白质。这种总体平行的鉴定蛋白质-配体相互作用的方法将大大加速并简化 NCE 主要靶标的确定，也可鉴定也许重要的较弱的二级相互作用。另外，

这些方法还能直接用于种间交叉反应性的问题，例如，可以快速评价潜在抗真菌化合物与人类蛋白质组中所有蛋白质的相互作用；这类信息可证明在随后先导化合物的任何优化中非常有用。

高通量筛选方法现在可以快速鉴定能与以前被确定为潜在治疗靶标的特定蛋白质结合的小分子。然而，这些方法不能解决任何一种特定相互作用选择性如何的问题，而这方面的了解对于决定是否研究特定先导化合物可能是至关重要的；大家认为，导向一种蛋白质的化合物可能比导向大量有关或无关蛋白质的化合物显示更低的副作用。

已经成功通过三期临床试验、但由于其主要作用机制未知仍未能获得管理机构批准的化合物有大量的例子。抗抑郁药米安色林和曲唑酮（trazadone）和 Pfizer 的抗关节炎药替尼达帕就是这样的例子，它们都有上亿美元的投资没有回报。此处所述的方法能使这些失败的药物复活，因为如果能发现这些药物的主要靶标，随后证实其作用机制，则极其宝贵的临床实验数据将获得管理机构批准。

现有的所有药物都有副作用，程度或大或小，一个例子是具有吸引力的抗精神分裂药物氯氮平。如果能够确定这些副作用的分子来源，将可大大促进氨基酸主要作用优化而副作用最小的下一代药物的设计。此处所述的方法也能直接用于这些问题，因为在生成一种化合物与蛋白质组中所有蛋白质的相互作用谱时，可确定异常的二级相互作用，随后能评价它们是否与已知的副作用有关。

本发明的方法也能通过用普通抑制剂筛查蛋白质组阵列来鉴定蛋白质家族，如丝氨酸蛋白酶。然后可以研制展示如所有人丝氨酸蛋白酶或所有激酶或所有 p450 酶的生物芯片，用于更集中地筛选先导化合物。例如，p450 生物芯片可用于评价一种特定先导化合物是否能被代谢，因为 p450 介导的羟基化通常是该过程的第一步，被认为是药物反应中患者-患者间差异的主要原因之一；实际上现在药物设计的目的之一是产生首先不被代谢的化合物，因此 p450 芯片具有重要的潜在用途。

蛋白质-蛋白质相互作用

蛋白质-蛋白质相互作用和多蛋白质复合物在细胞生物学中至关重要。例如，信号传导途径一般由细胞表面受体与外部配体之间的相互作用启动，随后是蛋白质-蛋白质相互作用的级联，最终导致特定基因的激活。蛋白质-蛋白质相互作用可能依赖于特异配体的存在，或者可能被特异配体阻断，而一些多蛋白质复合物只以依赖配体的形式形成。

利用双杂交技术已经鉴定了上千种新的蛋白质-蛋白质相互作用。此处所述的方法克服了这些方法的限制，能用多种经标记的蛋白质筛查蛋白质组阵列，不仅鉴定各种相互作用的作用配偶体而且鉴定其相对强度。这些方法也能用于鉴定多蛋白质复合物的组分，甚至在装备依赖配体时。

这些方法在说明蛋白质-蛋白质相互作用中的用途的一个实例是确定与疾病状态有关的特定细胞表面受体的胞质域的信号配偶体；这些信号配偶体的鉴定与制药前景直接相关，因为这些蛋白质-蛋白质相互作用可能恰好代表可能的治疗靶标。

蛋白质-DNA 相互作用

估计在人类基因组的所有基因中，大约 10% 编码转录因子，而现在只鉴定了其中一小部分。特异转录因子与 DNA 增强子元件的结合（通常是响应外部刺激）是形成增强体（enhanceosome）复合物的必要条件，该复合物随后开启基因表达。关于给药原则上能影响基因表达有不同的观点：药物可阻断蛋白质或小分子与细胞表面受体的结合，从而阻断信号级联；药物可阻断蛋白质-蛋白质相互作用或抑制信号级联内的酶活性；或者另外，药物可阻断增强体复合物内特异蛋白质-DNA 或蛋白质-蛋白质相互作用的形成。一个例子是，转录因子 NF- κ B 参与如同免疫和炎症反应、肢发育、脓毒性休克、哮喘和 HIV 前肽产生一样多样化的细胞过程。NF- κ B 激活中的大多数胞内信号级联是所有这些反应共有的，因此不是可能的干预靶标。因此，反应之间的差异在于最初的配体-受体相互作用或特异增强体复合物的形成。周知 NF- κ B 可结合至少 14 种不同的增强子元件，因此增强体复合物是可能

的治疗靶标。然而，描述各种增强体复合物需要知道有关 DNA 结合蛋白的数量以及彼此的蛋白质-蛋白质相互作用。能用本方法直接解决这两个问题。为了鉴定新的 DNA 结合蛋白，能用特异性 DNA 探针筛查蛋白质组阵列。此外，也能用特定转录因子的反式激活域筛查蛋白质组阵列，以鉴定可与之相互作用的其它蛋白质。这些筛查的交叉相关性可鉴定特异增强体复合物的新成分。

用本发明的方法产生的蛋白质阵列也可选择识别阵列中展示的各种蛋白质的分子。在一个优选实施方案中，选择的分子是抗体或抗体样蛋白质，将在噬菌体或核糖体上展示，或将与编码 mRNA 共价连接。

因此，噬菌体展示抗体文库能应用于阵列中的每种固定蛋白质，并通过洗涤除去未结合的抗体。然后按照正常程序回收选择的噬菌体，用来感染细菌。噬菌体感染的细菌能产生展示所选抗体的噬菌体颗粒，用于后几轮筛选，或者它们能产生可溶性抗体片段，以备直接应用。术语“抗体”或“抗体片段”在此是指来源于小鼠、人、骆驼或其它生物的单链 Fvs、FAB 片段、轻链或重链片段。

在一个优选实施方案中，蛋白质阵列可以是微孔形式，使得在筛选步骤后，能通过向每孔中添加适当细菌细胞（这些细胞将被筛选的噬菌体颗粒感染）回收噬菌体颗粒。然后向每孔加入生长培养基，使感染的细菌生长并表达抗体片段，而选择的抗体片段与阵列中的每种固定蛋白质保持物理分离。希望时，能在下几轮筛选中使用感染的细菌产生的新噬菌体颗粒。这些方法现在已经是筛选纯化或固定蛋白质的多克隆抗体或单克隆抗体片段的常规方法。实际上，通过总体平行的方式，同时利用标准体外抗体筛选方法，最初的蛋白质阵列可产生针对上千种正确折叠的蛋白质的多克隆抗体或单克隆抗体片段。

从原初阵列的每个孔中选择的可溶性表达的抗体片段本身能固定于一个新的空间确定的阵列中，使得能在新阵列的每个位置上筛选针对固定于原初阵列上一个确定位置的蛋白质的抗体片段。这样产生的抗体阵列在每个位置上含有多克隆或单克隆抗体片段，这取决于在固定可溶性抗体片段前进行筛选的轮数。

这些抗体阵列具有许多可能的用途，包括从粗细胞或组织裂解液中捕获多种蛋白质，用于相关蛋白质组的差异表达监测。此外，也可以根据配体结合功能直接筛选抗体捕获的蛋白质。一般来说，任何一种单克隆抗体可结合靶蛋白，以致阻断其功能，但是另一种单克隆抗体可能结合但不阻断功能。以一种总体平行的方法评价蛋白质组中所有蛋白质的所有单克隆抗体结合但不影响功能的能力显然是不切实际的。然而，蛋白质组中所有蛋白质的一组多克隆抗体可能含有具有希望的结合但不影响功能的能力的抗体，另外，也可能含有识别特定蛋白质的所有翻译后修饰的抗体。因此，如述产生的多克隆而不是单克隆抗体阵列通常有利于根据功能直接筛选捕获的蛋白质。

与原初蛋白质阵列相比，用此处所述方法产生的抗体阵列具有固定于阵列上的所有蛋白质在类似条件下都稳定的优点。从粗细胞或组织裂解液中捕获的蛋白质不是重组的，而是天然表达的。而且，能在从粗细胞或组织裂解液中捕获后，根据功能或配体结合等直接筛选捕获的蛋白质，这应当有助于功能的保持。

因此，在其它方面，本发明提供：

(i) 一种根据生物活性筛选一种或多种化合物的方法，包括下列步骤：使一种或多种该化合物接触如此处所述的蛋白质阵列，测定一种或多种该化合物与阵列中蛋白质的结合；

(ii) 一种根据特定蛋白质-蛋白质相互作用筛选一种或多种蛋白质的方法，包括下列步骤：使一种或多种该蛋白质（例如细胞表面受体）接触如此处所述的阵列，测定一种或多种特定蛋白质与阵列蛋白质的结合；

(iii) 一种根据特异蛋白质-核酸相互作用筛选一种或多种蛋白质的方法，包括下列步骤：使一种或多种核酸探针接触如此处所述的阵列，测定该探针与阵列中蛋白质的结合；

(iv) 如此处所述的阵列在快速筛选化合物、蛋白质或核酸中的用途；

(v) 如此处所述的阵列在筛选可识别阵列中每种蛋白质的分子中

的用途，其中该分子优选地是抗体；

(vi) 一种生产抗体阵列的方法，包括使此处所述的蛋白质阵列接触一种抗体文库，使得蛋白质阵列中的一种或多种蛋白质至少结合抗体文库中的一种抗体，除去所有未结合的抗体，与蛋白质结合的抗体固定于蛋白质阵列中；和

(vii) 一种筛查蛋白质功能或丰度的方法，包括使此处所述的抗体阵列接触一种或多种蛋白质的混合物的步骤。

方法(i)、(ii)、(iii)和(iv)也可包括首先根据本发明的一种或多种方法生产阵列的步骤。

用此处所述方法生产的蛋白质的用途构成了本发明的其它方面。熟练技术人员应当理解，本领域周知可应用修饰蛋白质的多种用途，

因此，在其它方面，本发明提供：

(i) 用本发明的方法生产的经标记的蛋白质在多种表达宿主（即细菌、酵母、哺乳动物细胞）中的表达（例如，见：Walker EA, Clark AM, Hewison M, Ride JP, Stewart PM., 人 1 型 11- β -羟基类固醇脱氢酶的催化域的功能表达、表征和纯化。J Biol Chem 2001 Jun 15;276(24):21343-50; Cai J, Daoud R, Georges E, Gros P., 多药耐药性蛋白 1 在 *Pichia pastoris* 中的功能表达。Biochemistry 2001 Jul 17;40(28):8307-16; Hara H, Yoshimura H, Uchida S, Toyoda Y, Aoki M, Sakai Y, Morimoto S, Shiokawa K., 人 hepassocin——一种具有肝细胞有丝分裂原活性的肝特异性蛋白质——的 cDNA 的分子克隆和功能表达分析(1)。Biochim Biophys Acta 2001 Jul 30;1520(1):45-53)

(ii) 用此处所述方法生产的经标记的蛋白质的用途。

(iii) 用此处所述方法生产的经标记的蛋白质的用途，通过将修饰 DNA 分子克隆到酵母双杂交表达载体中，分析表达的蛋白质与酵母双杂交系统中其它蛋白质的相互作用（例如，见：Staudinger J, Perry M, Elledge SJ, Olson EN., 利用双杂交系统，酵母中脊椎动物螺旋-环-螺旋蛋白质之间的相互作用。J. Biol Chem 1993 Mar 5;268(7):4608-11, Vojtek AB, Hollenberg SM, Cooper JA., 哺乳动物

Ras 与丝氨酸/苏氨酸激酶 Raf 直接相互作用。Cell 1993 Jul 16;74(1):205-14)。

(iv) 用此处所述方法生产的经标记的蛋白质的用途，用于固定于亲和柱/基质上，例如，允许通过亲和层析纯化：a) 相互作用的蛋白质，b) DNA，或 c) 化学化合物。(例如，见：Rhodes N, Gilmer TM, Lansing TJ., 来自瞬时转染的哺乳动物细胞的活性重组 atm 蛋白的表达与纯化。Protein Expr Purif 2001 Aug;22(3):462-6; Zwicker N, Adelhelm K, Thiericke R, Grabley S, Hanel F., Strep-tag II 用于人 c-Myc 活性 bHLHzip 域的一步亲和纯化。Biotechniques 1999 Aug;27(2):368-75, Giuliani CD, Iemma MR, Bondioli AC, Souza DH, Ferreira LL, Amaral AC, Salvini TF, Selistre-de-Araujo HS., 活性重组赖氨酸 49 磷脂酶 A(2) myotoxin 作为融合蛋白在细菌中的表达。Toxicon 2001 Oct;39(10):1595-600)

(v) 用此处所述方法生产的经标记的蛋白质作为诊断工具在通过亲和纯化固定中进行抗体检测 (ELISA 测定) 的用途 (例如，见：Doellgast GJ, Triscott MX, Beard GA, Bottoms JD, Cheng T, Roh BH, Roman MG, Hall PA, Brown JE., 灵敏的酶联免疫吸附测定，用于通过酶联凝集测定利用信号扩增检测肉毒杆菌神经毒素 A、B、E。J Clin Microbiol 1993 Sep;31(9):2402-9)

(vi) 用此处所述方法生产的经标记的蛋白质作为 cDNA 微阵列探针鉴定 DNA 结合蛋白的用途(例如，见：DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM., cDNA 微阵列在分析人癌症中基因表达模式中的用途。Nat Genet 1996 Dec;14(4):457-60)

(vii) 用此处所述方法生产的经标记的蛋白质的用途，通过质谱分析用本发明的方法修饰的源文库或原材料的表达蛋白质成分，阐明“蛋白质组”中蛋白质的身份 (例如，见：Bordini E, Hamdan M., 利用矩阵-辅助激光解吸/电离时间飞行和电喷质谱法研究某些共价和非共价复合物。Rapid Commun Mass Spectrom 1999;13(12):1143-51)。

本发明每个方面的优选特征在细节上加以必要的修改后适用于其它任一方面。

现在将参照下列实施例描述本发明，这些实施例绝不应视为限制本发明的范围。

附图说明

图 1a: 显示载体 pMM106H 的构建;

图 1b: 显示标记之前一个示例基因 (GST) 的 PCR 扩增和外切核酸酶消化的细节;

图 1c: 显示特异连接和 PCR 扩增以导入标签的细节;

图 1d: 显示 GST 催化的谷胱甘肽与 1-氯-2,4-二硝基苯的反应。

实施例 1

(a) 载体构建 (见图 1a)

发明者构建了一种载体 pMM106H，其来源于 pUC19，含有一个强杂合启动子 (Ptrc)，用来引导克隆到紧接启动子序列下游 NcoI 位点的基因的表达。发明者在 NcoI 位点与下游 HpaI 位点之间插入一个 676bp 的无义 DNA 序列作为填充片段。HpaI 是一种平端切割酶，用来切割载体，如果阅读框位于平端的第一个碱基上，下游 DNA 编码一种聚天冬酰胺六组氨酸肽。六组氨酸标签后是一个琥珀终止密码子 (TAG)，之后是编码维多利亚水母 (*Aequorea Victoria*) 绿色荧光蛋白 (GFP) 的基因。pMM106H 的结构通过测序得到证实。

只有在克隆期间在 HpaI 位点处产生正确阅读框时，作为 NcoI/平端片段克隆到 pMM106H 中的基因才与 His-标签和 GFP 融合。GFP 在此用作有利于目视筛选表达 His 标签的克隆的报道基因，也是融合蛋白正确折叠的标志，因为只有在折叠成正确构象后 GFP 才有活性。

琥珀终止密码子将产生少量全长融合蛋白，显示为绿色菌落，但是大多数融合蛋白在 His 标签后立即终止，能用于随后的固定和酶测定。应当理解，可溶性表达使细胞具有某些可见表型的蛋白质的大量

不同肽能替代 GFP 作为经标记的蛋白质表达和折叠的标志。包括但不限于：氯霉素乙酰转移酶、 β -半乳糖苷酶、 β -半乳糖苷酶的 lacZ 片段和能抑制转录的蛋白质，如 λ -CI 阻抑蛋白。

下述方法采用的模板为 pSGTN。该质粒如下构建：首先在标准条件下利用引物'GSTfwd2'（5'-ATG CTG CAG ACG TCA ACA GTA TCC ATG GCC CCT ATA CTA GG-3'）和'GSTHindIII'（5'-GCG AGG AAG CTT GTC AAT CAG TCA CGA TGA ATT CCC G -3'）从 pGEX-2T（Pharmacia）中 PCR 扩增日本裂体吸虫（*Schistosoma japonicum*）谷胱甘肽 S 转移酶（GST）基因。这两条引物在 GST 的终止密码子处导入一个 NcoI 限制位点，使 GST 的第二个残基从丝氨酸突变为丙氨酸，并在 GST 基因 3' 的多克隆位点中导入一个终止密码子，随后是一个 HindIII 限制位点。然后在标准条件下将 PCR 产物作为 NcoI/HindIII 片段克隆到预先用 NcoI/HindIII 消化的 pTrcHisA（Invitrogen）中，产生 pGSTN。

（b）标记前基因的 PCR 扩增和外切核酸酶消化（见图 1b）

发明者使用定制设计的载体特异引物'STforwar'（5'-ATG CTG ACG TCA TGA GGC CCA TGG GGC CCG GAT AAC AAT TTC ACA CAG G-3'）和'STreverse'（5'-GCG GAT CCT TGC GGC CGC CAG GCA AAT TCT GTT T-3'）通过聚合酶链反应从构建体 pGSTN 中扩增 GST 基因，这两条引物分别结合载体起点上游 156bp 和终止密码子下游 84bp。在 4 个分开的 100 μ l 反应中进行 30 个循环的 PCR（94 $^{\circ}$ C 1min；57 $^{\circ}$ C 1min；72 $^{\circ}$ C 2min）。每个 PCR 反应含有~20ng 模板 DNA、各 50pmol 引物和 2.5 单位 Pwo 聚合酶。每个 PCR 反应都在标准缓冲液（10mM Tris.HCl pH8.8, 25mM KCl, 5mM (NH₄)₂SO₄, 2mM MgSO₄, 10% DMSO）中进行。4 个 PCR 反应也都含有非标准脱氧核苷酸三磷酸混合物，如下：

反应 1) 200 μ M dATP, 200 μ M dTTP, 200 μ M dCTP, 150 μ M dGTP, 50 μ M α -S-dGTP;

反应 2) 200 μ M dATP, 200 μ M dTTP, 200 μ M dGTP, 150 μ M dCTP, 50 μ M α -S-dCTP;

反应 3) 200 μ M dATP, 200 μ M dGTP, 200 μ M dCTP, 150 μ M dTTP, 50 μ M α -S-dTTP;

反应 4) 200 μ M dGTP, 200 μ M dTTP, 200 μ M dCTP, 150 μ M dATP, 50 μ M α -S-dATP.

在 α -S-dNTP 存在下模板 DNA 的扩增当然可以利用引物延伸反应进行, 其采用多种不同的 DNA 聚合酶, 包括缺乏 3'-5'外切核酸酶活性的热稳定聚合酶, 如 Taq 聚合酶, 和热不稳定聚合酶, 如 T4 DNA 聚合酶或 DNA 聚合酶 I 的 Klenow 片段。

在每种特异 PCR 混合物中含有一种 α -硫代脱氧核苷酸三磷酸使有关 α -S-dNTP 随机但统计学地掺入特异 PCR 终产物中。这些修饰核苷酸不是外切核酸酶 III 的底物, 其用来终止酶对核苷酸的逐渐去除。然后合并 4 种 PCR 混合物, 用 QIAquick PCR cleanup 试剂盒 (Qiagen) 在标准条件下纯化, 消化, 用限制酶 AatII 完成。然后凝胶纯化产生的~1000bp PCR 产物。用 AatII 限制酶切产生 3'-突出端, 它对外切核酸酶 III 活性有抗性, 因此保护 PCR 产物的 5'端免于降解。

可以设想特异保护 PCR 产物一端免受外切核酸酶消化的备选方法, 包括但不限于下列。产生 4 个或 4 个以上碱基的 3'-突出端的任何限制酶能代替 AatII, 在 PCR 引物的设计中加入必要的位点。产生 5'-突出端的任何限制酶也能代替 AatII, 在 PCR 引物的设计中加入必要的位点; 在这种情况下, 产生 5'-突出端后进行 DNA-聚合酶-介导的补平反应, 其中相关 α -硫代-dNTPs 代替 dNTPs, 保护 PCR 产物的新 3'

端免受外切核酸酶消化。

然后在 150 μ l 反应体积中, 10-15 μ g 消化的 PCR 产物与 75 单位外切核酸酶 III/ μ g DNA 在 37 $^{\circ}$ C 下温育 30 分钟。Exo III 消化在标准反应缓冲液 (66mM Tris.HCl pH8.0, 6.6mM MgCl₂, 5mM DTT, 50 μ g/ml 牛血清白蛋白) 中进行。这些条件确保 Exo III 消化完成。然后加热到 75 $^{\circ}$ C 15 分钟使酶灭活。Exo III 消化产物是从 PCR 产物 3'-端的一组嵌套缺失。

外切核酸酶 III 是一种非渐进性 3'-5' 外切核酸酶, 它不能水解含有 α -硫的核苷酸, 因此在本方案中, 每次 Exo III 到达 α -硫代-脱氧核苷酸碱基时, PCR 产物的凹缺 3'-端的渐进性截短即停止。最终结果是一组嵌套缺失, 是在较早阶段随机掺入各种 α -S-dNTP 的结果。在最初的 PCR 扩增中采用的 α -S-dNTP 与 dNTP 之比根据经验确定, 使嵌套缺失的外部横跨中心大小为 400bp 的窗口, 比最初的全长 PCR 产物约短 100bp。

可通过改变 α -S-dNTP 与正常 dNTP 之比控制获得的截短的大小范围。当该方法用于真核 cDNA 时这是重要的, 因为这些 cDNA 含有长度可变的 3' 非翻译区, 最常见的 3'-UTR 长度为 200-300bp。由于 4 种 α -S-dNTP 中的每一种的相对掺入效率因聚合酶种类而不同, 因此希望采用对 4 种碱基中的每一种并对特定聚合酶优化的 α -S-dNTP 与 dNTP 比例。采用的外消旋 α -S-dNTP 与 dNTP 的摩尔比一般为 1:1-1:3。

(c) 单链区的去除和克隆的准备 (见图 1c)

在上一步骤中通过外切核酸酶 III 消化产生的嵌套缺失体通过乙醇沉淀纯化, 重悬浮于 1 $\times$ 绿豆核酸酶缓冲液 (50mM 乙酸钠 pH5.0, 30mM NaCl, 1mM ZnSO₄) 中。在 100 μ l 反应中, 消化的 DNA 用 (2 单位/ μ g) 30 单位绿豆核酸酶 30 $^{\circ}$ C 处理 30 分钟。该步骤除去 5'-和 3'-突出端, 产生平端产物。加入 EDTA 至终浓度为 5mM 终止反应。消化的产物用 QIAquick PCR 纯化试剂盒 (Qiagen) 纯化, 用 NcoI 消化, 以 100bp DNA 阶梯作为标准, 在 1% 琼脂糖/TBE 凝胶上分离。用 QIAquick 凝胶提取试剂盒 (Qiagen) 从琼脂糖中提取大小为

800-1000bp 的产物。显然,也能用其它单链核酸酶如 S1 核酸酶从 3'-5' 外切核酸酶产生的嵌套 3'-缺失体上切除 5'-突出端。

(d) 嵌套缺失体制备的差异

产生嵌套缺失的其它大量标准分子生物学方法与最早的方法明显的差异。包括但不限于:使用任何一种 3'-5' 外切核酸酶、任何一种 5'-3' 外切核酸酶或从线性 DNA 片段末端逐渐截短的任何一种内切核酸酶。例如,能利用同上的反向引物,而用可结合 GST 基因起点约 2kb 上游的正向引物,进行最初的 PCR 扩增。这将产生 GST 基因侧翼为 5'-端>2kb 而 3'-端只有 84bp 的片段。纯化的 PCR 片段然后可用 Bal31 核酸酶处理,该酶从 5'-端和 3'-端逐步降解线性双链 DNA。该酶是非进性的, DNA 的降解速度取决于反应时间和温度,以及 DNA 的碱基组成。由于 PCR 产物中 GST 基因 3'-端的侧翼区明显短于 5'-端,在从另一端到达起始密码子之前发生可达到及超过终止密码子的降解。时程实验可以确定从 PCR 产物 3'-端除去可达 400bp 的最佳反应条件。然后可使产生的嵌套缺失体成为平端,用特殊限制酶消化,切下原始载体在基因 5'-端编码的剩余的任何单链区,直接克隆到标签载体中。此外,也能用 λ 外切核酸酶产生嵌套 5'-缺失。该酶的优选底物是 5'磷酸化双链 DNA,因此含有 5'羟基端能容易地保护 DNA 底物的一端。

5'-3' 外切核酸酶产生的嵌套 5'-缺失体的单链 3' 突出端能用多种不同的酶去除,包括 T4 DNA 聚合酶或单链 DNA 特异核酸酶,如 RNase T 或外切核酸酶 T 或绿豆核酸酶。

(e) 修饰产物的克隆和分析(见图 1c)

用限制酶 NcoI 和 HpaI 消化载体 pMM106H (3 μ g), 凝胶纯化 2870bp 骨架片段。然后在标准条件下将载体 DNA 和如上所述制备的限制酶切产物连接在一起,用连接混合物转化大肠杆菌 DH5 α 细胞,然后回收,平板接种于含 100 μ g/ml 羧苄青霉素的 LB 平板上。

对上一步获得的全套缺失体施行这一克隆程序。然而,只有那些切除了 GST 基因的终止密码子并且阅读框内密码子之后立即终止的缺失体才能通过该方法在克隆步骤后与六组氨酸标签和 GFP 产生阅

读框内融合；用该方法产生的其它所有缺失产物只能与六组氨酸标签和 GFP 产生阅读框外融合，或者由于在 GST 终止密码子处翻译终止而产生未融合的 GST 蛋白。这是因为缺失产物平端与载体平端的连接产生一种基因融合，其中下游载体 DNA 的翻译阅读框由 GST 编码区原始阅读框说明。如果缺失产物终止于不完全密码子，新添加的六组氨酸编码序列将与 GST 基因处于阅读框外，但如果缺失产物保留 GST 终止密码子，则不产生 GST 与六组氨酸标签的翻译融合。因此，只有上述全过程产生的六组氨酸-(和 GFP-)标记蛋白必须是 GST 与聚天冬酰胺、六组氨酸标签的融合体。没有必要是绝对全长的克隆，然而，将筛选它们正确折叠和保持酶活性的能力，以用于其它步骤。

转化的菌落在紫外线 (365nm) 下显示，肉眼选择发绿色荧光的 30 个菌落 (约占总数的 10%) 进行下一步分析。这些菌落复制平板接种，在标准条件下用抗-His-标签和抗-GST 抗体通过菌落 Western 印迹法分析。抗-His-标签抗体只结合表达六组氨酸标记蛋白的菌落，因此 Western 印迹分析得到关于表达六组氨酸标签阅读框内融合体的菌落数的直接信息。另一方面，抗-GST 抗体结合于 GST 蛋白 C 端附近，因此只能识别表达全长或接近全长 GST 蛋白的菌落。发明者鉴定了 19 个菌落 (绿色荧光菌落的 63%)，它们含有可被抗-His 标签和抗-GST 抗体阳性识别的蛋白质。其中 12 个菌落的 DNA 进行扩增、纯化并测序。测序数据证实存在与全长 GST 融合的两个正确阅读框内融合体，10 个菌落在 GST 基因中有短的截短，但仍与六组氨酸标签在阅读框内。因此，发明者经过总过程获得的全长 GST 菌落的分离率约为 17% (占绿色荧光菌落总数)，而预计保持活性的全长或接近全长 GST 克隆的分离率约为 63% (占绿色荧光菌落总数)。

(f) 经标记的蛋白质的固定和功能分析 (见图 1d)

用通过上述方法产生的全长、六组氨酸标记的 GST 质粒之一转化大肠杆菌 DH5 α 细胞。一个羧苄青霉素抗性菌落在 10ml 液体培养液中生长至对数中期，然后补加 100 μ M IPTG 诱导六组氨酸标记的 GST 表达。再生长 4 小时后，收获细胞，通过冻融/溶菌酶裂解。对粗裂解

液进行 SDS-PAGE, 显示有过量表达的预期大小 (27kDa) 的一种蛋白质, 约占总可溶性蛋白的 20%, 以及少量 54kDa GST-六组氨酸-GFP 融合蛋白, 这是由于琥珀抑制产生的。粗裂解液 (500 μ l; 100 μ g) 然后与镍-NTA 磁珠 (50 μ l; 结合能力为 15 μ g 六组氨酸标记的蛋白质) 混合, 在磁场下通过沉淀回收磁珠。弃去上清液, 洗涤磁珠, 然后重悬浮于含有各 1mM 谷胱甘肽和 1-氯-2,4-二硝基苯的谷胱甘肽 S 转移酶测定缓冲液中。在室温下 30 分钟后, 通过测量 340nm 的吸光度, 采集终点测定数据; 该波长对应于 GST-催化反应产物的 $\lambda_{\max}$ 。

作为对照, 含有亲本载体 (pMM106H) 或含有编码无关 His-标记蛋白 (丙氨酸消旋酶) 的质粒的 DH5 α 培养物平行生长、诱导、收获、裂解并测定。只检测与含有 His-标记 GST 的粗裂解液混合的磁珠上的 GST 活性, 清楚地证明, 观察到的 GST 活性是由于固定的 His-标记 GST, 该蛋白质在特异固定后保留活性。

在完成酶测定后, 通过加入含有 100mM 咪唑的缓冲液从磁珠上洗脱蛋白质, 通过 SDS-PAGE 分析。这表明, 得到阳性活性测定结果的样品含有预期为谷胱甘肽 S 转移酶大小 (27kDa) 的一种固定的蛋白质, 从而证实在磁珠上观察到的活性仅仅是由于这种重组 His-经标记的蛋白质。

实施例 2

(a) 载体构建

发明者构建了第二种载体 pMM111, 它与 pMM106H (见实施例 1) 基本相同, 除了 676bp NcoI/HpaI 无义 DNA 填充片段被替换为来自大肠杆菌 *gdhA* 基因的 300bp NcoI/HpaI 片段; HpaI 克隆位点被替换为 SmaI 位点, 其位置使得下游六组氨酸标签与 *gdhA* 基因在阅读框外 2 个核苷酸; GFP 基因的 ATG 起始密码子被替换为丙氨酸密码子 (GCG)。该载体已经被消化, 使得克隆到 SmaI 位点内的插入片段必须在 3'端含有密码子的第一个核苷酸, 与六组氨酸标签和 GFP 位于阅读框内。pMM111 的构建通过测序得到证实。

(b) 导入标签的改进方法

发明者然后进行与实施例 1 所述基本相同的方法，不同之处在于下列改变：第一，在最初的 PCR 扩增即部分 (b) 反应 3 中只添加 α -S-dTTP。第二，终产物克隆到载体 pMM111 的 NcoI - SmaI 位点。

该方法在理论上有优于实施例 1 所述的几个优点。主要是来自与一种 α -硫代-dNTP 掺入最初 PCR 产物中的统计学结合。因此，在外切核酸酶 III 完全消化后，嵌套 3'-凹缺的缺失体全都终止于 3'-胸苷碱基而不是 4 种核苷酸中的任一种。如果 3'-T 属于 GST 基因第一个框内终止密码子，或者在它之前，并且位于与第一个框内终止密码子的“T”相同的阅读框内，则这些片段克隆到 pMM111 的 SmaI 位点内只能产生与六组氨酸标签和 GFP 的阅读框内融合（这是因为 SmaI 消化在标签的编码序列之前产生 2 个核苷酸的缺口）。

在统计学上，用这种改进方法进行外切核酸酶水解可比实施例 1 少产生 4 倍的嵌套缺失体。然而，由于所有 3 种可能的终止密码子都含有“T”作为其第一个碱基，它们全都存在于缺失体组中，因此构成多 4 倍的全套缺失体。假定任一特定“T”可能与终止密码子的第一个“T”位于同一个阅读框内，用这种改进方法产生的所有克隆的 33% 应当是与载体编码的 His 标签的框内融合体，但影响折叠（从而影响功能）的缺失体产生“白色”菌落（由于实施例 1 所述的原因）。发明者发现，按照此改进方法，“绿色”群体内精确、全长克隆的比例显著高于按照实施例 1 所述方法发现的。对于“起始”密码子相同，因为所有已知的起始密码子（ATG, GTG, TTG, ATT, CTG）都在第二个位置含有“T”。

该改进方法的另一个优点是，能向最初 PCR 扩增使用的正向引物（例如，正向-A 5'-AAA AAA AAA AAA GAT CGA TCT CAT GAC GGA TAA CAA TTT CAC ACA GG-3'）的 5'-端掺入一个 polyA 尾。在以 3:1 的 dTTP: α -S-dTTP 比扩增中，至少一个 α -S-dTTP 残基非常可能在互补链的末端、PCR 产物的 5'端掺入。掺入的这些核苷酸耐受外切核酸酶 III 消化，因此在特别保护 PCR 产物末端不受降解时不需

要酶促步骤。

实施例 3

(a) 利用六组氨酸标签对第二种蛋白质的修饰

按照实施例 1 对谷胱甘肽-S-转移酶所述的方法, 发明者证明该方法不依赖于所操作的基因的序列。

从编码人转录因子 NF- κ B p50 的质粒开始, 除非另外说明, 严格按照实施例 1 所述的方法, 发明者能够证明 NF- κ B p50 的修饰, 使得第一个框内终止密码子切除, 而被替换为与编码聚天冬酰胺、六组氨酸标签和 GFP 的 DNA 融合的框内融合体(当琥珀终止密码子被抑制时)。当用远紫外线(365nm)激发时, 进一步表征发绿色荧光的克隆。使用抗-His-标签抗体的菌落 Western 印迹分析允许鉴定表达六组氨酸经标记的蛋白质的克隆。这些克隆的可溶性蛋白质裂解液通过 SDS-聚丙烯酰胺凝胶电泳分析, 用抗-His 标签抗体杂交。在大约 65kDa M_r (对应于与六组氨酸标签和 GFP 翻译融合的 NF- κ B p50)和约 38kDa M_r (NF- κ B p50-His 标签)处观察到免疫反应性信号。另外, 在约 27kDa M_r 处也有信号, 这可能是对应于组氨酸标记 GFP 蛋白的降解产物。测序数据证实, 有几个克隆编码全长或接近全长的 NF- κ B 与六组氨酸标签的正确框内融合体。在一个实验中, 根据绿色荧光筛选了 190 个菌落。总共 38 个菌落(占筛查克隆总数的 20%)在用远紫外线(365nm)激发时发绿色荧光。使用抗-His 标签抗体的菌落 Western 印迹分析表明, 38 个克隆中有 29 个表达六组氨酸标签。测序数据证实, 其中 18 个克隆是全长或者接近全长的 NF- κ B p50 与六组氨酸标签的框内融合体; 其中 7 个克隆是绝对全长的、组氨酸标记的 NF- κ B p50 基因, 其余 11 个组氨酸标记克隆含有 4-1 个氨基酸残基的短截短。该实验清楚地证明了具有可指示蛋白质表达和适当折叠的报道系统的优点。大约 50%的发绿色荧光的克隆是与组氨酸标签融合的全长框内融合体, 或者含有不与域边界交叉、与组氨酸标签框内融合的少量截短。

(b) 六组氨酸标记的 NF- κ B p50 的固定和功能分析

用通过上述方法产生的全长、六组氨酸标记的 NF- κ B 质粒之一转化大肠杆菌 DH5 α 细胞。一个羧苄青霉素抗性菌落在 10ml 液体培养液中生长至对数中期，然后补加 100 μ M IPTG 诱导六组氨酸标记的 NF- κ B p50 表达。再生长 4 小时后，收获细胞，通过超声处理裂解。对粗裂解液进行 SDS-PAGE，在预期大小（38kDa）处显示一种过量表达的蛋白质，约占总可溶性蛋白质的 5%。

κ B 基序

5'-CGT ATG TTG TGG GGA ATT CCC AGC GGA TAA C-3'
3'-GCA TAC AAC ACC CCT TAA GGG TCG CCT ATT G-5'

NF- κ B P50 结合位点

在标准条件下，利用 3'-端转移酶用洋地黄毒苷在 3'-碱基处标记含有 NF- κ B p50 的回文结合位点的双链寡核苷酸 κ B 基序”。

采用溶菌酶/冻融法在含有 5mM β -巯基乙醇的 PBS（磷酸缓冲液 pH7.5）中制备蛋白质裂解液。将每个克隆 200 μ l 可溶性蛋白质裂解液加到 Ni-NTA 包被的微孔中，在室温下温育 45 分钟。在温育结束时，用 PBST（含有 0.02% Triton X-100 的 PBS）洗孔三次，除去所有未结合的蛋白质。用含有 5mM β -巯基乙醇的 DNA 结合缓冲液（10mM Tris.HCl pH7.4, 75mM KCl）洗孔三次，浸泡时间为 1 分钟。向孔中 200 μ l 含 1 μ g poly(dI-dC) 非特异性 DNA 的 DNA 结合缓冲液中加入 3' 洋地黄毒苷标记的 κ B 基序（2 pmol）。再温育 30 分钟后，用含有 0.02% Triton X-100 的 10mM Tris.HCl pH7.4, 25mM KCl 洗孔三次，除去未结合的 DNA。用含有 0.2% 牛血清白蛋白的“抗体稀释缓冲液”（10mM Tris.HCl pH7.4, 25mM KCl）将抗-洋地黄毒苷抗体-碱性磷酸酶偶联物稀释为 150mU/ml。然后将稀释的抗体（200 μ l）加到微孔中。在室温下 30 分钟后，用含有 0.02% Triton X-100 的“抗体稀释缓冲液”（3 $\times$ 350 μ l）洗涤微孔，除去未结合的抗体。然后向孔中加入 200 μ l 含 250 μ M 对硝基苯磷酸酯（pNPP）——碱性磷酸酶的一种底物——的缓冲液（100mM Tris.HCl pH9.5, 100mM NaCl, 50mM MgCl₂），在室

温下反应过夜，之后在 405nm 下测量每孔的黄色显色（对应于产物对硝基苯酚的形成）。底物 pNPP 的背景水解率很低，因此，根据孔中黄色的出现可立即判断阳性测定结果。作为试验对照，发明者减掉粗裂解液，或标记的寡核苷酸，或抗体，或加入 20 倍过量的未标记双链 oligo，或在同一载体背景中，用来自表达六组氨酸标记 GST 的 DH5 α 细胞的等量粗细胞裂解液代替含六组氨酸标记 NF- κ B p50 的粗裂解液。

在该试验中，NF- κ B p50 首先经由特异结合位点结合标记的寡核苷酸。蛋白质-DNA 复合物然后通过六组氨酸标签固定于微孔中，然后洗掉其它所有蛋白质（包括标记 oligo 与粗裂解液中存在的其它 DNA 结合蛋白的复合物）以及任何未结合的、标记的 oligo。由于抗体偶联物可识别 oligo 上的标记物，而不是六组氨酸标记的蛋白质，如果在 NF- κ B p50 通过标签固定后保持 NF- κ B p50-DNA 相互作用，则试验中只能观察到阳性信号；如果该相互作用未保持，oligo 将在洗涤步骤中丢失，因此不会观察到颜色变化。

发明者发现，只有在含有六组氨酸标记的 NF- κ B p50 粗裂解液和洋地黄毒苷标记的寡核苷酸并且加入抗-洋地黄毒苷标记的抗体-碱性磷酸酶偶联物的微孔中才检测到黄色产物。这证明，观察到的颜色改变特别是由于固定的 NF- κ B p50-寡核苷酸复合物，而且 NF- κ B p50 在特异固定后保留活性。

实施例 4

（a）从 10 种基因的集合中鉴定一种蛋白质

发明者采用基本如实施例 1 所述的方法，不同之处在于下表列出的 10 种不同基因的集合。发明者制备了产生的特异修饰蛋白质的阵列，使得阵列中的每个位置对应于一种通过该方法附加的标签固定的重组蛋白。发明者然后通过功能试验筛查该阵列，成功鉴定了该集合中的各种蛋白质成分。

表 1. 集合中 10 种基因的大小和功能

基因	大小	来源与功能
谷胱甘肽 S 转移酶	950bp	细菌; 解毒
NF- κ B p50	1165 bp	人; 转录因子
麦芽糖结合蛋白	1325 bp	细菌; 糖类转运
丙氨酸消旋酶	1342 bp	细菌; 细胞壁生物合成
活化 T 细胞的核因子 (NFAT)	1087 bp	鼠; 转录因子
吲哚甘油磷酸合酶	1528 bp	细菌; 氨基酸生物合成
磷酸核糖氨基苯甲酸异构酶	920 bp	细菌; 氨基酸生物合成
色氨酸合酶 (α -亚基)	1122 bp	细菌; 氨基酸生物合成
胰凝乳蛋白酶抑制剂 2	389 bp	大麦; 丝氨酸蛋白酶抑制剂
β -内酰胺酶	1040 bp	细菌; 抗生素抗性

开始时, 所有 10 种基因均亚克隆到同一种 pTrcHisA 载体骨架中, 因为该载体模拟 cDNA 文库遇到的情况。实施例 1 所述的引物 “STforward” 和 “STreverse” 设计为用于扩增 pTrcHisA 载体骨架内编码基因的通用引物。

设计引物 “STforward”, 使其编码如下的多个限制位点:

5' -ATG CTG ACG TCA TGA GGC CCA TGG GGC CCG GAT AAC AAT TTC ACA CAG
G-3'

Aat II Bsp HI Sfi I

于是, 为了外切核酸酶保护目的能用限制酶 AatII 或 SfiI 中的任何一种产生 3'-突出端。为了在修饰过程结束时定向克隆, 在该实施例中发明者选用 BspHI, 因为尽管在统计学上它在文库内更频繁地酶切, 但它产生与此处所用标签载体 pMM106H 中的 NcoI 克隆位点相容的粘端, 并且不能酶切该集合中的所有 11 种基因。显然, 原则上能使用任何引物编码的限制位点, 只要标签载体在启动子下游含有一个相当的克隆位点; 在这方面 SfiI 在较大文库格式上具有明显的优点, 因为它含有一条 8bp 的识别序列, 因此 SfiI 位点在特定基因内随机出现的

频率 (6.5×10^4 分之一) 比 6bp 识别序列如 BspHI (4096 分之一) 更低。

标签载体 pMM106H 是一种 “ATG” 载体, 即为了表达天然蛋白质, 5'-克隆位点 (NcoI) 与位于核糖体结合位点 (RBS) 下游的 ATG 起始密码子重叠。然而, 在此处所述的方法中, 发明者不依靠在起始密码子处含有一个共同限制位点的克隆基因。而是只依靠载体编码的启动子起始转录产生 mRNA, 翻译起始的必要信号由克隆的基因本身提供。因此在本实施例中, 原始集合中的所有基因都在 RBS 之后紧接着含有一个起始密码子, 而无论在 ATG 处是否存在克隆位点。由于引物 “STforward” 结合于最初所有 11 个克隆的 RBS 上游, 随后用编码限制位点的任何引物进行修饰后克隆将向标签载体中导入新修饰的基因以及原始 RBS 和 ATG, 于是确保翻译起始。在 cDNA 文库形式中, 用于所有全长 cDNA 的一种类似情况含有其自身 5'-非翻译区 (UTR), 其中含有真核翻译起始信号。获得适当翻译起始所必需的是向真核载体中克隆每种修饰的 cDNA 以及其 5'-UTR, 产生转录起始信号, 因此与本实施例中使用的相当的一组通用 PCR 引物能以不依赖于序列的方式用于修饰 cDNA 文库的每一个成员。

实验方法如实施例 1 所述进行, 修改如下。等摩尔的一组所有 10 种基因用作使用引物 “STforward” 和 “STreverse” 的初始 PCR 扩增的模板, 之后用 AatII 消化片段以保护 5'端。外切核酸酶 III 和绿豆核酸酶处理的片段完全如实施例 1 所述产生, 然后用 BspHI 消化, 其限制酶切正向 PCR 引物结合位点内的独特片段, 产生一个用于克隆到载体 pMM106H 中的粘端。凝胶纯化获得的片段, 与载体连接。转化的细胞在紫外线 (365nm) 下显示, 肉眼选择发绿色荧光的菌落进行 Western 印迹分析。约占总数 2% 的转化菌落发绿色荧光。其中, 103 个 (42%) 表达可由抗-His 标签抗体识别的蛋白质。这些菌落分别接种于 96 深孔平板中的 1.5ml 液体培养基中, 生长过夜。离心收集细胞, 通过冻融/溶菌酶裂解。然后将各个粗裂解液加至镍-NTA 包被的 96 孔平板的各个孔中, 洗去未结合的蛋白质, 剩下固定于每个孔中的分离

的 His 标记重组蛋白。然后利用实施例 1 和 3 所述的试验测定固定的蛋白质的 GST 或 NF- κ B 活性，在所有情况下都根据绿色或黄色的出现分别确定含有阳性“命中”的孔。

在第一个试验中，发现阵列中有三种蛋白质显示 GST 活性，命中率约为 3%。相应质粒的测序显示，所有三种都编码 GST 基因与六组氨酸标签和 GFP 基因的框内融合体；在这三种之中，有两种为绝对全长 GST，一种略微截短，但显然不影响活性。

在第二个试验中，阵列中有三种蛋白质显示阳性“ κ B 基序”DNA 结合活性。对阳性克隆的进一步表征显示，两种克隆是 NF- κ B p50 基因与六组氨酸标签的框内融合体，其中之一接近全长（截短一个氨基酸），而另一个截短更多，但含有 NF- κ B p50 的完整 DNA 结合域。有趣的是，由于该试验是设计用于结合同源 DNA 序列，DNA 结合域仍然完整、折叠并具有功能的截短体在该试验中将为阳性。发现第三个克隆是鼠转录因子 NFAT 的 DNA 结合域与 His 标签的框内融合体。该试验使用的 3'-洋地黄毒苷标记的“ κ B 基序”含有一个 NF- κ B p50 的特异性、高亲和力（ K_d 约为 pM）结合位点，但 NFAT 的 DNA 结合域也能以接近 nM 的亲和力特异识别这同一个结合位点。因此该结果证明，该方法生产的阵列的功能筛查既能鉴定特异相互作用也能鉴定特异且生物学相关的较弱相互作用。

在下一个实验中，按照该实施例的方法制备含有约 340 种 His 标记的蛋白质的阵列。通过 GST 活性测定分析阵列显示，阵列中的所有蛋白质 8% 具有强 GST 活性。另外，也利用基因特异的引物对一组 340 个编码质粒 DNA 进行 PCR 分析，显示每种基因位于 His 标记的集合中。因此这些数据进一步证实本实施例的方法是不依赖于序列的，适用于不同基因的集合。

总之，发明者使用这些实施例所述的方法生产微孔形式的功能蛋白质阵列，并且利用这些阵列，根据特异蛋白质-配体相互作用（GST 活性测定）或特异蛋白质-DNA 相互作用（NF- κ B 结合测定），从一组集合中成功鉴定了三种不同的蛋白质。

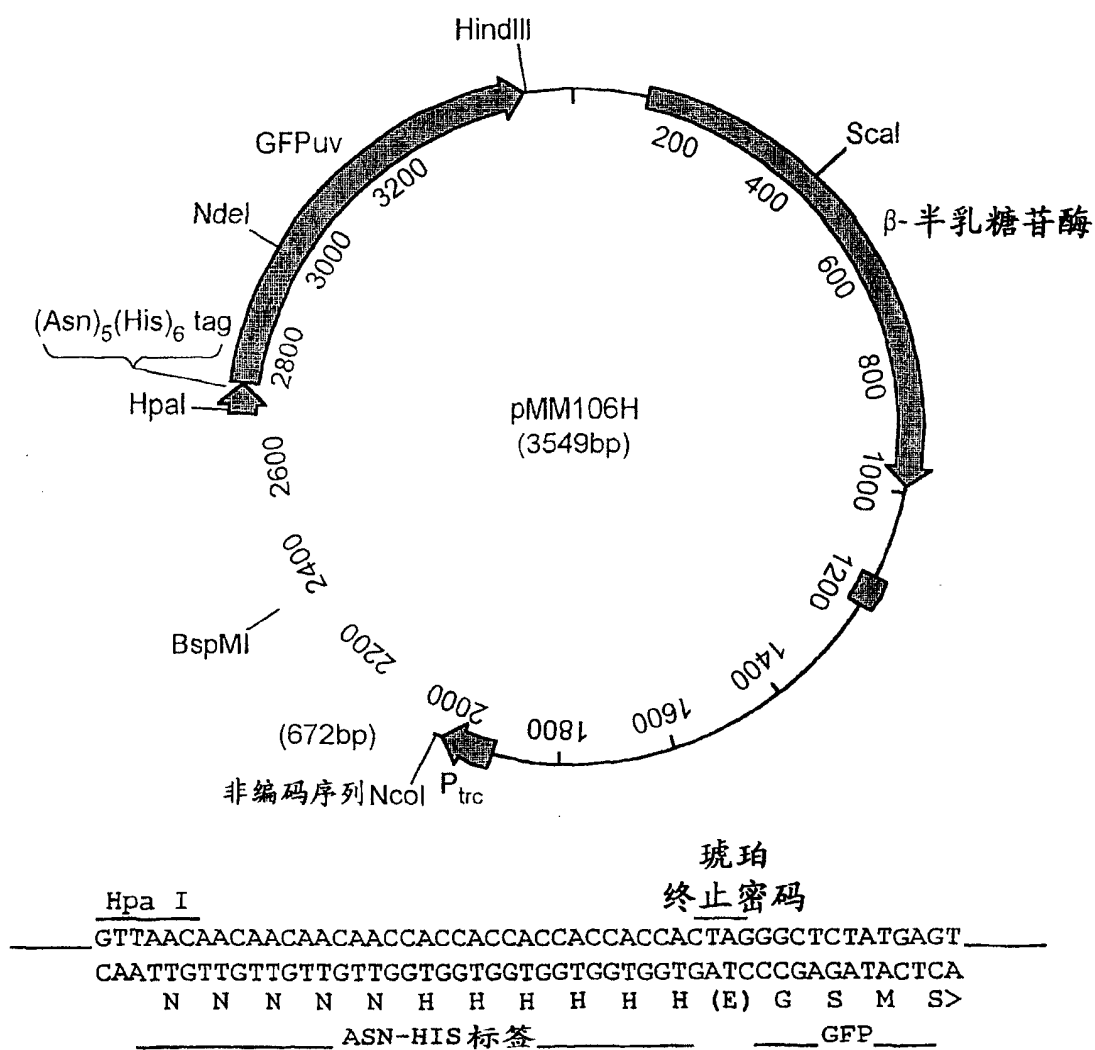
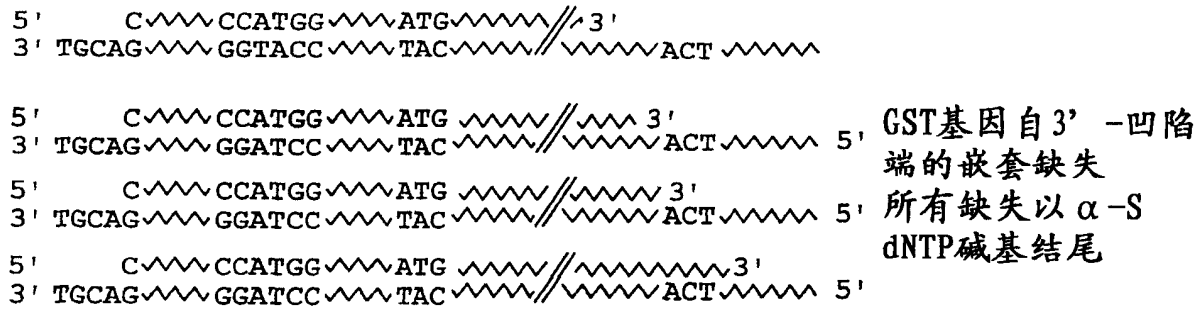
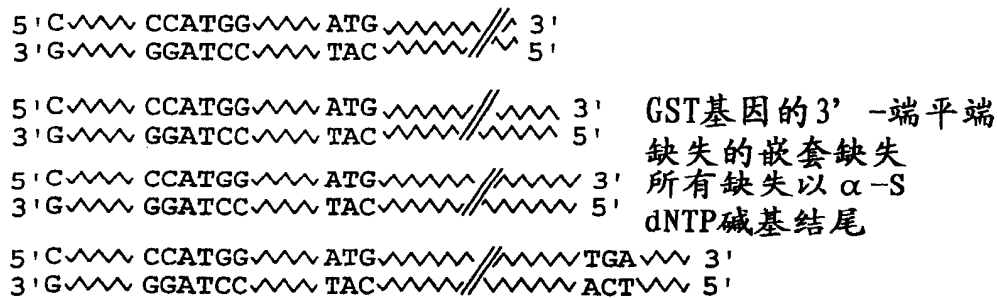


图 1a



↓
 绿豆核酸酶



用NcoI消化，将片段克隆入
 pMM106H的NcoI/HpaI位点，以
 加上对GST3' -端缺失嵌套有
 特异性的标签。在紫外光下观
 察荧光绿的克隆，以筛选正
 确折叠的与His标签符合读框
 地融合克隆。

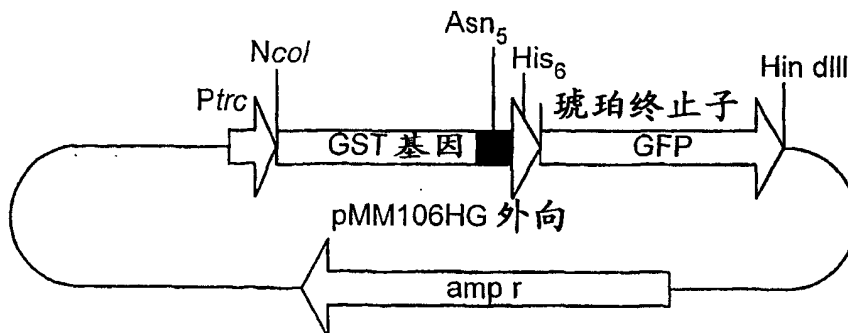


图1C

在液体培养基中培养个体克隆并诱导表达

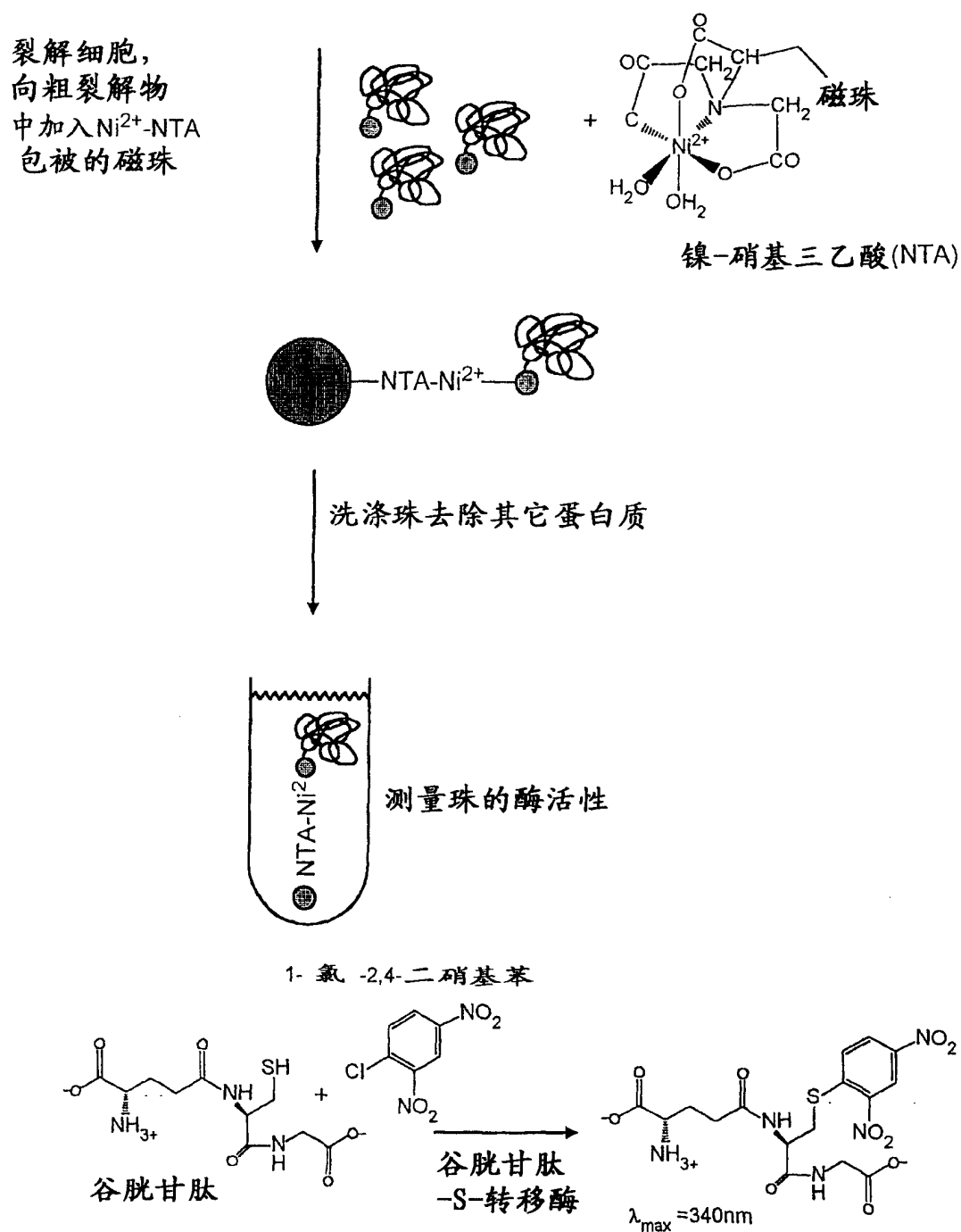


图1d