



(12)发明专利



(10)授权公告号 CN 107430130 B

(45)授权公告日 2020.07.24

(21)申请号 201580063963.9

(22)申请日 2015.11.24

(65)同一申请的已公布的文献号

申请公布号 CN 107430130 A

(43)申请公布日 2017.12.01

(30)优先权数据

1420852.4 2014.11.24 GB

(85)PCT国际申请进入国家阶段日

2017.05.24

(86)PCT国际申请的申请数据

PCT/GB2015/053573 2015.11.24

(87)PCT国际申请的公布数据

W02016/083793 EN 2016.06.02

(73)专利权人 吉尼维拉基公司

地址 匈牙利皮利什森特拉斯洛

(72)发明人 乔鲍·耶奈伊

(74)专利代理机构 广州华进联合专利商标代理

有限公司 44224

代理人 刘培培

(51)Int.Cl.

G01N 33/68(2006.01)

C12N 15/10(2006.01)

(56)对比文件

Sara D'Angelo et al..From deep sequencing to actual clones.《Protein Engineering, Design and Selection》.2014,第27卷(第10期),

Sara D'Angelo et al..The antibody mining toolbox.《MABS》.2013,第6卷(第1期),

Gustav N. Sundell and Ylva

Ivarsson.Interaction Analysis through Proteomic Phage Display.《BioMed Research International》.2014,第11卷(第5期),

Amita Gupta et al..A Novel Helper Phage Enabling Construction of Genome-Scale ORF-Enriched Phage Display Libraries.《PLOS》.2013,第8卷(第9期),

Ylva Ivarsson et al..Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes.《PNAS》.2014,第111卷(第7期),

U. Ravn et al..By-passing in vitro screening-next generation sequencing technologies applied to antibody display and in silico candidate selection.《Nucleic Acids Research》.2010,第38卷(第21期),

Wei Li.ORF phage display to identify cellular proteins with different functions.《Methods》.2012,第58卷(第1期),

Ulla Ravn et al..Deep sequencing of phage display libraries to support antibody discovery.《Methods》.2013,第60卷(第1期),

H Benjamin Larman et al..Autoantigen discovery with a synthetic human peptidome.《Nature Biotechnology》.2011,第29卷(第6期),

审查员 杨玉路

权利要求书2页 说明书25页  
序列表7页 附图1页

(54)发明名称

方法

(57)摘要

本发明一般涉及用于检测结合相互作用,特别是蛋白质-蛋白质相互作用的方法和试剂盒,

特别涉及用于标记、分析、检测和测量蛋白质-蛋白质相互作用的高通量方法。

1. 一种用于确定结合剂和靶之间的结合相互作用的方法,所述方法包括:
  - a) 使结合剂文库与靶接触以允许形成结合剂/靶复合物,其中所述结合剂文库的每个成员与单一核苷酸序列相关联;其中与所述结合剂文库的每个成员相关联的所述单一核苷酸序列与所述结合剂文库的所述成员的结合特征相关联;并且其中所述靶与单一核苷酸序列不关联;
  - b) 将所述复合物分离于区室中,使得一个区室存在单个复合物;
  - c) 将与结合剂相关联的单一核苷酸序列连接在复合物中以形成连接的核苷酸序列,其中在连接步骤中使分离的复合物保持分离;
  - d) 从所述连接的核苷酸序列识别复合物中存在的结合剂。
2. 一种确定结合剂和靶之间的结合相互作用的方法,所述方法包括:
  - a) 使结合剂文库与靶接触以允许形成结合剂/靶复合物,其中所述结合剂文库的每个成员与单一核苷酸序列相关联并且其中所述靶与单一核苷酸序列相关联;
  - b) 将所述复合物分离于区室中,使得一个区室存在单个复合物;
  - c) 将与结合剂和靶相关联的单一核苷酸序列连接在复合物中以形成连接的核苷酸序列,其中在连接步骤中使分离的复合物保持分离;
  - d) 从所述连接的核苷酸序列识别复合物中存在的结合剂;
  - e) 使用连接的核苷酸序列以使所述结合剂文库的每个成员的单一核苷酸序列与所述成员的结合特征关联。
3. 根据权利要求1或权利要求2所述的方法,其特征在于,所述结合剂文库包含抗体文库。
4. 根据权利要求3所述的方法,其特征在于,所述抗体文库包含抗体展示文库或抗体文库,其中每个抗体用所述单一核苷酸序列标记。
5. 根据权利要求1或2所述的方法,其特征在于,所述靶包含蛋白质。
6. 根据权利要求1或2所述的方法,其特征在于,所述靶包含蛋白质展示文库,其中所述文库的每个成员与单一核苷酸序列相关联。
7. 根据权利要求6所述的方法,其特征在于,所述蛋白质展示文库为cDNA噬菌体展示文库。
8. 根据权利要求5所述的方法,其特征在于,所述蛋白质处于蛋白质混合物内。
9. 根据权利要求8所述的方法,其特征在于,所述蛋白质混合物为富集的蛋白质混合物。
10. 根据权利要求8所述的方法,其特征在于,所述蛋白质混合物富含磷蛋白、膜蛋白和/或天然或人造修饰的蛋白质。
11. 根据权利要求1或2所述的方法,其特征在于,在使用前从所述结合剂文库中除去结合至其它结合剂的结合剂。
12. 根据权利要求1或2所述的方法,其特征在于,所述结合剂文库在使用前富集。
13. 根据权利要求1或2所述的方法,其特征在于,所述区室位于固体表面上。
14. 根据权利要求1或2所述的方法,其特征在于,所述区室为乳液微滴,扩散受限或分离的区室。
15. 根据权利要求1所述的方法,其特征在于,与所述复合物中的结合剂相关联的核苷

酸序列接合至与所述复合体中的结合剂相关联的另一核苷酸序列。

16. 根据权利要求2所述的方法, 其特征在于, 与所述复合体中的结合剂相关联的核苷酸序列接合至与所述结合剂/靶复合体内的靶相关联的核苷酸序列。

17. 根据权利要求1或2所述的方法, 其特征在于, 从所述连接的核苷酸序列识别所述复合体内存在的结合剂和/或靶。

18. 根据权利要求1或2所述的方法, 其特征在于, 通过对所述连接的核苷酸序列进行测序来识别存在于复合体内的 (i) 结合剂或 (ii) 结合剂和靶。

19. 根据权利要求1或2所述的方法, 其特征在于, 在分离后固定所述结合剂/靶复合体。

20. 根据权利要求1或2所述的方法, 其特征在于, 在识别所述结合剂和/或靶之前, 组合来自一种以上的复合体的连接的核苷酸序列。

21. 根据权利要求1或2所述的方法, 其特征在于, 在存在和/或不存在化合物的情况下, 实施使所述结合剂文库与靶接触的步骤(a)。

22. 根据权利要求5所述的方法, 其特征在于, 所述蛋白质与蛋白质混合物内的其它蛋白质交联。

23. 根据权利要求1或2所述的方法, 通过稀释、特异性结合进行分离, 或利用所述复合体的物理和/或化学性质进行分离。

24. 根据权利要求3所述的方法, 其特征在于, 与抗体噬菌体文库的每个成员相关联的单一核苷酸序列为噬菌体内的编码所述抗体的CDR的核苷酸序列。

25. 根据权利要求7所述的方法, 其特征在于, 与cDNA噬菌体展示文库的每个成员相关联的单一核苷酸序列为噬菌体内的编码所展示的蛋白质的序列。

26. 根据权利要求15或16所述的方法, 其特征在于, 接合所述核苷酸序列包括:

i. 使用至少两对PCR引物扩增与所述结合剂或结合剂和靶相关联的核苷酸序列, 以产生至少两组扩增子, 其中设计引物使得由第一组引物产生的扩增子包含与由第二组引物产生的扩增子中的序列互补的序列;

ii. 使扩增子组退火;

iii. 进行扩增反应以产生连接的核苷酸序列。

27. 根据权利要求26所述的方法, 其特征在于, 使用乳液PCR扩增实施所述扩增。

28. 根据权利要求26所述的方法, 其特征在于, 步骤i-iii同时进行。

29. 根据权利要求1或2所述的方法, 其特征在于, 重复所述方法, 并且改变所述靶或结合剂文库的浓度。

30. 一种用于实施前述任一项权利要求所述的方法的试剂盒, 所述试剂盒包括:

a) 结合剂文库, 其中所述结合剂文库的每个成员具有预定的蛋白质结合特征且与已知的单一核苷酸序列相关联; 和

b) 用于接合与结合剂相关联的核苷酸序列的至少两对引物组; 以及任选的使用说明书。

31. 根据权利要求30所述的试剂盒, 其特征在于, 所述试剂盒还包含蛋白质展示文库, 其中所述文库的每个成员与单一核苷酸序列相关联。

## 方法

### 技术领域

[0001] 本发明一般涉及用于检测结合相互作用,特别是蛋白质-蛋白质相互作用,的方法和试剂盒,特别涉及用于标记、分析、检测和测量蛋白质-蛋白质相互作用的高通量方法。

### 背景技术

[0002] 细胞结构由其复合体,实际制造细胞的分子机器来限定。传统上,细胞生物基于其作为催化剂、信号分子或细胞和微生物的构件(building blocks)时各自的作用来识别蛋白质。目前,我们见证了后基因组观点的出现,后基因组观点扩展了蛋白质的作用,将其作为蛋白质-蛋白质相互作用网中的元素,以及功能模块中的“环境”或“细胞”功能。

[0003] 复杂的蛋白质-蛋白质网的定性和定量表征以及主要细胞类型特异性相互作用蛋白质的识别对于理解多种人类疾病如癌症、自身免疫性疾病和其它疾病中的蛋白质-蛋白质相互作用的生理过程和改变至关重要。蛋白质-蛋白质网的详细见解以及疾病相关差异的识别可能会使特定药物的合理设计和开发引向新途径。细胞或组织中蛋白质-蛋白质相互作用的模式也可用作分子诊断的工具。

[0004] 蛋白质参与代表了很多细胞生物学和功能的机理基础的复合体相互作用。这些蛋白质-蛋白质相互作用被组织成精美复杂的网。蛋白质-蛋白质相互作用网的架构被认为是无标度的,大多数蛋白质仅具有一个或两个连接,且具有相对较少的“中枢”,该“中枢”具有数十个,数百个或更多个连接。相互作用网是高度动态的,允许相互作用组根据,例如外部刺激甚至发展过程而快速更改。核心蛋白之间和两个或更多个模块蛋白之间的相互作用可能由结构域-结构域相互作用介导。附着蛋白内和附着蛋白之间的相互作用不太可能以这种方式发生。尽管蛋白质复合体和相互作用对生物过程的调节和执行有所贡献,但在结构和功能方面,能被充分理解的复合体相对较少。

[0005] 在通过实验获得细胞相互作用的动力学常数上的尝试较少。这些定量参数将有助于基于微分方程的细胞过程动力学模型的开发。这些模型对于了解药物作用是必要的,并将促进许多复杂疾病的新药物的发现。定量多尺度模型的开发可以提供对细胞水平的药物治疗作用和副作用的理论认识。

[0006] 术语“抽样”用于仅询问群体子集的实验设计。代表性抽样在蛋白质相互作用数据集的产生中并不常见,其中抽样往往以生物学优先为指导。“覆盖率”总结实际测试了可能的相互作用的总集的哪一部分。根据当前技术,对“相互作用组”(例如,在所研究的条件下,在细胞中发生的所有物理相互作用的集合)进行推理是无效的。

[0007] 设计若干方法来研究蛋白质-蛋白质相互作用,包括用于选择并检测结合另一种蛋白的蛋白的物理方法,如蛋白质亲和层析、亲和印迹、免疫沉淀(包括2D凝胶电泳和质谱)、交联;基于文库的方法:蛋白质探针、噬菌体展示、双杂交体系;基于文库的其它方法和遗传方法:基因外抑制因子、合成致死作用、过度产生表型、野生型蛋白质的过度产生和突变蛋白质的过度产生;和不连接的非互补。

[0008] 这些方法中有许多不适合高通量蛋白质相互作用分析。最有前景的高通量技术可

通过开发如下技术而获得：肽和蛋白质文库筛选技术（例如，酵母双杂交策略），肽和蛋白质文库筛选技术是识别和克隆与感兴趣的蛋白质相互作用的蛋白质的基因的方法；双杂交阵列，其中以菌落阵列方式实施大规模实验，其中每个酵母菌落表达定义的成对的“诱饵”和“猎物”蛋白，这可以自动化方式对报道基因的活性进行评分-表明相互作用；噬菌体展示（其中在蛋白质文库中淘选“诱饵”蛋白质）以及亲和纯化/质谱（AP-MS）特别是用于限定细胞中所有复合体（“复合体组”）及其组成蛋白质；和串联亲和纯化（TAP）。TAP根据其在该复合体的各种形式中出现的频率显示出作为核心、模块或附着蛋白质的相互作用的蛋白质。

[0009] 所有这些方法都具有与使用这些技术获得的信息的可靠性、完整性和容易性相关的优点和缺点。理想的方法以有时间和成本效益的方式捕获相互作用组信息，实现随机抽样和高冗余度抽样。理想的方法基于动态的、原始的细胞环境，基于天然的蛋白质-蛋白质相互作用提供全面、覆盖率足够大的甚至较大的多单元蛋白质复合体的定量的相互作用数据。理想的方法抑制随机变量的影响，例如非特异性检测、意外相互作用的蛋白质。理想的方法还减少变量的影响，变量是原始蛋白质-蛋白质相互作用以外的检测原理中涉及的任何结合事件相关变量。

[0010] 双杂交筛选，特别是基于阵列的技术，可以实现大规模的相互作用组信息生成。然而，由于其由两部分构成，成对检测，缺乏基于原始环境的动态信息，人造结合剂（杂交蛋白）和酵母细胞环境限制性原理（例如相比原始宿主，翻译后修饰是扭曲的），因而存在较多缺点。几乎所有这些缺点都已通过各种方式部分解决。然而，尚未设计出组合所有这些所需特征的方法。

[0011] 基于亲和力的方法，特别是使用质谱分析法作为检测原理的方法，部分在正确的细胞环境中产生大量的半定量的相互作用组数据。然而，该方法受到随机和结合（亲和力）相关变量的影响。该方法检测意外的非特异性结合事件。为了生成随机抽样的，高覆盖率、全面的数据集将需要大量的时间和费用，这会损害其潜在利益来检测相互作用组的动态性质。这些问题中的一些已经得到解决，特别是使用串联亲和纯化（TAP），其中意外的非特异性结合事件减少到最小，但是以牺牲不太可靠的蛋白质-蛋白质复合体回收为代价。

[0012] 这些技术大大加快了蛋白质-蛋白质相互作用（PPI）的产生。在对相互作用组进行开创性研究之后，进行了若干大规模研究，从而产生了一些高质量的成对蛋白质-蛋白质相互作用的数据集。例如，过滤的酵母相互作用组（FYI）为不同数据集的交集，包括Y2H数据、AP-MS数据、计算机预测，慕尼黑蛋白质序列信息中心（Munich Information Centre for Protein Sequences）物理相互作用和文献报道的蛋白质复合体。

[0013] 由于现有的方法学方法不能完全满足蛋白质-蛋白质相互作用和相互作用组研究的需要，因此需要分析和表征复杂的蛋白质-蛋白质网的新方法。

[0014] 本发明提供用于检测结合相互作用，特别是细胞水平的蛋白质-蛋白质相互作用的方法和试剂盒。所述方法和试剂盒可用于同时检测优选在细胞的原始环境中的复合体蛋白质网的所有相互作用蛋白质或其子集。该方法和试剂盒基于动态的，原始的细胞环境，基于天然的蛋白质-蛋白质相互作用，提供全面、覆盖率足够大的甚至较大的多单元蛋白质复合体的定量和潜在的动力学相互作用数据。

[0015] 本发明可使用抗体展示技术检测蛋白质-蛋白质相互作用，使用多种抗体噬菌体作为结合剂。本发明也可使用适体技术检测蛋白质-蛋白质相互作用，使用多种适体作为结

合剂。多种结合剂的复杂性在广泛范围内变化,可以为几种结合剂之间至数万或数十万,数百万或数千万或数亿或数亿种结合剂。为了从适用于本发明方法的高复杂性结合剂获得低复杂度的结合剂,设计了一种复杂性降低方法(富集)。

[0016] 可以识别和监测目标分子之间更详细的相互作用。例如可以检测到蛋白质-蛋白质相互作用。在结合剂/靶复合体内存在两种或更多种结合剂可以指示复合体内可能存在两个或多个靶。这表明两个或更多个靶可能彼此相互作用或相互绑定。如果已知特异性结合剂的可识别部分,例如蛋白质或核酸序列,则可以识别该靶。该方法可以使用高度平行PCR扩增,通过连接绑定的展示抗体的噬菌体的可识别核酸序列,即具有预定结合特征(例如,具有已知的表位序列或已知结合至特定分子)的那些核酸序列来实施。该方法可优选通过乳液PCR进行。该方法可以在低蛋白质复合体浓度下进行,优选在区室中进行。可以通过高度平行PCR扩增,优选使用复杂度降低的结合检测剂来检测靶之间的相互作用,例如蛋白质-蛋白质相互作用。通过连接的可识别序列的测序获得,优选通过高度平行DNA测序或通过其它序列检测手段获得靶-靶,例如,蛋白质-蛋白质,相互作用信息。输入材料(例如,靶)的量的改变可用于收集配体结合动力学数据。此外,该方法可以在存在和不存化化合物的情况下进行,以确定化合物对靶相互作用是否具有任何影响,以及该作用是激动的还是拮抗的。

[0017] 本发明还可以使用蛋白质展示技术,展示生物体的蛋白质片段并确定多种展示的抗体的结合特征,每种抗体具有单一的可识别的序列信息,并且每个展示的蛋白质片段具有可识别的序列信息。优选地,所展示的蛋白质片段的可识别序列信息是编码所展示的氨基酸序列的序列。可以由每个抗体-蛋白质复合体的可识别序列信息确定绑定的抗体的身份。任选地,可以识别每个抗体-蛋白质复合体内的绑定的蛋白质片段的身份。任选地,可以由每种抗体-蛋白质复合体的连接的可识别序列信息确定绑定的抗体的身份和绑定的蛋白质片段的身份。也可以使用不同量的靶,例如蛋白质和结合剂,比如,展示的蛋白质或展示的抗体来确定结合动力学特征。

[0018] 本发明的方法和组合物也可用于识别可以使这种蛋白质-蛋白质相互作用激动或拮抗的化合物。本发明提供以拮抗(破坏性的)或激动(促进性的)化合物检测结合相互作用的方法和试剂盒。本发明提供用于同时检测复合体蛋白质网中,优选在细胞的原始环境中,拮抗和/或激动化化合物的结合相互作用的方法和试剂盒。该方法和试剂盒提供基于细胞原始环境的基于天然蛋白质-蛋白质相互作用的数据,该数据是全面的,并且即使对于较大的多单元蛋白质复合体也具有覆盖率足够大的定量和潜在的动力学相互作用数据。

## 发明内容

[0019] 本发明提供一种用于确定结合剂和靶之间的结合相互作用的方法,其包括:

[0020] a) 使结合剂文库与靶接触以允许形成结合剂/靶复合体,其中所述结合剂文库的每个成员与单一核苷酸序列相关联;

[0021] b) 分离所述结合剂/靶复合体;

[0022] c) 将与结合剂相关的核苷酸序列连接在结合剂/靶复合体中以形成连接的核苷酸序列;

[0023] d) 由连接的核苷酸序列识别所述复合体中存在的结合剂。

[0024] 本发明描述分析和表征复合体结合相互作用,特别是蛋白质-蛋白质网或相互作用组的方法。该方法基于结合剂及任选的结合剂的靶(例如,蛋白质)的相关共定位的识别,其中关于结合剂和任选的结合剂的靶的共定位(优选在多个区室中)的信息成对地连接和翻译为核苷酸。也可以由核苷酸序列确定结合剂和任选的靶的身份。该信息可以通过测序显示。

[0025] 本发明还描述分析和表征拮抗(破坏性的)或激动(促进性的)化合物对靶分子相互作用的影响的方法。该方法基于在存在和不存在化合物的情况下识别结合剂及其靶,例如蛋白质。通过绑定的靶特异性结合剂的单一识别序列(其随后翻译成核苷酸)的成对连锁而,优选在多个区室中,进行结合剂与其靶之间形成的复合体的检测以及结合剂的识别。可以通过测序来显示复合体的量的变化以及结合剂和所涉及的任选的靶的身份。

[0026] 结合剂优选为抗体或适体。

[0027] 优选地,所述结合剂为抗体展示文库或抗体文库,其中每个抗体用所述单一核苷酸序列标记。

[0028] 优选地,所述靶也与单一核苷酸序列相关联。

[0029] 与复合体中的结合剂相关联的核苷酸序列可以连接至与复合体中的第二结合剂相关联的第二核苷酸序列。该方法可以用于识别结合单个靶的多种结合剂。例如,当靶为蛋白质时,该方法可以识别结合至蛋白质上不同表位的抗体。可替换地,靶可以为蛋白质复合体,并且该方法可以识别结合至复合体内的不同蛋白质的多种结合剂。例如,与结合剂/靶复合体中的一种结合剂相关联的核苷酸序列可以连接至与结合剂/靶复合体中的第二结合剂相关的核苷酸序列。一旦(从所连接的序列)得知结合剂的身份,则可以识别靶的组分,并且因此识别靶中自然相互作用的蛋白质。例如,如果结合剂是具有已知结合特征的抗体,则可以识别由抗体绑定的蛋白质。因此,可以识别靶内的蛋白质的身份。这可以检测和识别样品内的蛋白质-蛋白质相互作用。此外,一旦识别了蛋白质-蛋白质相互作用,该方法可用于监测化合物对相互作用的影响。

[0030] 可替换地,与结合剂/靶复合体中的结合剂相关联的核苷酸序列可以连接至与结合剂/靶复合体内的靶相关联的核苷酸序列。该方法可用于识别哪个结合剂与哪个靶相互作用。例如,该方法可以用于识别结合剂文库的哪些成员可以与已知靶形成复合体。该信息可用于表征结合剂文库的成员以获得结合特征信息。

[0031] 优选地,所述随机配对的、连接的核酸产物的产生包括使用至少两对PCR引物来扩增相同或不不同的扩增子;其中5'端的PCR引物具有序列标签,其中用带标签的引物进行扩增产生随机配对的、连接的核酸产物。更优选地,扩增为乳液PCR扩增,并且所述扩增子和随机配对的、连接的核酸产物的产生是并行进程。

[0032] 优选地,所述接合的扩增产物的所述测序为高度平行的测序方法。

[0033] 本发明的方法可用于研究化合物对结合剂与靶之间的相互作用或两种或多种靶分子之间相互作用的影响。可以在存在和不存在化合物的情况下实施使结合剂与靶接触的步骤,并且比较结果以确定化合物是否影响结合剂与靶之间或靶分子之间的结合相互作用。该方法可用于识别可用于治疗医学疾病和病症的潜在药剂。

[0034] 本发明还提供用于实施本发明的方法的试剂盒,其包括:

[0035] (i) 结合剂文库,其中所述结合剂文库中的每个成员与单一核苷酸序列相关联;和

[0036] (ii) 至少两对引物组,其用于连接与结合剂相关的核苷酸序列;以及任选的使用说明书。所述试剂盒还可以包含蛋白质展示文库,其中所述文库的每个成员与单一核苷酸序列相关联。

## 附图说明

[0037] 图1描述了使用本发明的方法和组合物检测蛋白质-蛋白质相互作用的试验的一般原理。具有预定的结合特征信息的抗体文库的噬菌体用于显示相互作用蛋白质的结合信息。通过PCR二聚化在多个区室中确定结合信息。噬菌体在该过程中裂解,释放出其单一的DNA。使用通用引物扩增这些单一的序列并进行二聚化。通过下一代测序(NGS)对编码结合信息的二聚产物进行测序,并且基于结合特征(包括其识别的结合靶的身份)已知的特异性噬菌体的检测来确定绑定的蛋白质身份。

[0038] 图2描述了用于表征结合剂的试验的一般原理-使用本发明的方法和组合物将结合剂核酸标记身份与其结合特征信息(包括其识别的结合靶的身份)相关联。展示cDNA和展示抗体的噬菌体彼此绑定。通过PCR二聚化在多个区室中测定结合信息。噬菌体在该过程中裂解,释放出其单一DNA。使用通用引物扩增这些单一序列并进行二聚化。通过下一代测序(NGS)对编码结合信息的二聚产物进行测序,并且通过数据库搜索从所述序列确定cDNA的蛋白质身份,并与结合剂核酸标记身份相关联。

## 具体实施方式

[0039] 理想的方法是以有时间和成本效益的方式从相互作用组捕获信息,从而能够进行随机抽样和高冗余抽样。该方法甚至为较大的多单元蛋白质复合体提供全面覆盖的定量相互作用数据。该数据在原始细胞环境中获得的,因此可以测量天然的蛋白质-蛋白质相互作用,并可用于检测动态的相互作用。该方法抑制随机变量的影响,比如检测意外的非特异性相互作用的蛋白质。该方法还减少变量影响,变量影响是除了原始蛋白质-蛋白质相互作用以外的检测原理中涉及的任何结合事件相关影响,例如自身结合或特异性结合。

[0040] 图1中总结了本发明的一个实施例并且在下面详细描述试验体系的不同组成部分。图2中总结了本发明的另一实施例并且在下面详细描述该体系的另外的组成部分。

[0041] 本发明提供了一种用于确定结合剂和靶之间的结合相互作用的方法,包括:

[0042] a) 使结合剂文库与靶接触以允许形成结合剂/靶复合体,其中所述结合剂文库的每个成员与单一核苷酸序列相关联;

[0043] b) 分离所述结合剂/靶复合体;

[0044] c) 将与结合剂相关的核苷酸序列连接在结合剂/靶复合体中以形成连接的核苷酸序列;

[0045] d) 从连接的核苷酸序列中识别出复合体中存在的结合剂。

[0046] 该方法可以在间接体内、体内或体外进行。

[0047] 结合剂

[0048] 优选地,结合剂为抗体、适体,或基于工程化蛋白质支架。可替换地,该结合剂可以为化合物。该结合剂可以为抗体展示文库或抗体文库的成员,其中每个抗体用单一核苷酸序列标记。该方法可以使用展示的抗体试剂作为结合剂,其中结合特征,例如结合剂结合的



靶是已知的,并且确定与多种展示的抗体试剂相关联的单一核苷酸序列,并且结合特征和单一的核苷酸序列彼此相关。因此,本发明提供确定结合特征并使其与多种展示的抗体试剂的可识别的单一核苷酸序列关联的方法。这提供绑定特征信息。

[0049] 本发明中使用的结合剂可以为抗体。本文所用的术语“抗体”是指免疫球蛋白分子和免疫球蛋白分子的免疫活性部分,免疫球蛋白分子即含有特异性结合抗原的抗原结合位点的分子(无论是天然的还是部分或完全合成产生的)。术语“抗体”包括抗体的抗体片段、衍生物、功能等同物和同源物、人源化抗体,包括包含免疫球蛋白结合结构域的任何多肽,无论是天然还是全部或部分合成,以及具有结合结构域的作为抗体结合结构域与抗体结合结构域同源的任何多肽或蛋白质。因此,包括与另一多肽融合的包含免疫球蛋白结合结构域的嵌合分子或等同物。EP-A-0120694和EP-A-0125023中描述了嵌合抗体的克隆和表达。抗体的实例为免疫球蛋白亚型(例如IgG, IgE, IgM, IgD和IgA)及其亚型亚类;包含抗原结合结构域的片段如Fab、scFv、Fv、dAb、Fd;和双抗体。抗体可以为多克隆或单克隆。

[0050] 互补决定区(CDR)是免疫球蛋白(抗体)中的可变链的一部分,由B细胞产生,其中这些分子与其特异性抗原结合。作为分子的最可变部分,CDR对免疫球蛋白产生的抗原特异性的多样性至关重要。在免疫球蛋白的可变结构域的氨基酸序列上有三个不连续排列的CDR(CDR1、CDR2和CDR3)。由于免疫球蛋白通常由两个可变结构域(在两条不同的多肽链,重链和轻链上)组成,每个抗原受体有6个CDR可共同与抗原接触。

[0051] 已经表明,整个抗体的片段可以执行结合抗原的功能。结合片段的实例是(i)由VL、VH、CL和CH1结构域组成的Fab片段;(ii)由VH和CH1结构域组成的Fd片段;(iii)由单抗体的VL和VH结构域组成的Fv片段;(iv)由VH结构域组成的dAb片段(Ward, E.S. et al., Nature 341:544-546 (1989));(v)分离的CDR区;(vi)F(ab')<sub>2</sub>片段,包含两个连接的Fab片段的二价片段;(vii)单链Fv分子(scFv),其中VH结构域和VL结构域通过肽接头连接,其允许两个结构域相关联以形成抗原结合位点(Bird et al., Science 242:423-426 (1988); Huston et al., PNAS USA 85:5879-5883 (1988));(viii)双特异性单链Fv二聚体(PCT/US92/09965)和(ix)“双抗体”,通过基因融合构建的多价或多特异性片段(W094/13804; P.Hollinger et al., Proc.Natl.Acad.Sci.USA 90:6444-6448 (1993))。

[0052] “抗原结合结构域”是抗体的一部分,其包含与抗原部分或全部特异性结合并互补的区域。当抗原较大时,抗体可能仅与抗原的特定部分结合,该部分称为表位。抗原结合结构域可以由一个或多个抗体可变结构域提供。抗原结合结构域可以包含抗体轻链可变区(VL)和抗体重链可变区(VH)。

[0053] 可替换地,结合剂可以基于工程化的蛋白质支架。蛋白质支架衍生已修饰以提供感兴趣的靶分子的结合位点的自稳定的、可溶的、天然的蛋白质结构。工程化蛋白质支架的实例包括但不限于亲和体,亲和体基于在其两个 $\alpha$ -螺旋上提供结合界面的葡萄球菌蛋白A的Z-结构域(Nygren, P.A. (2008) .FEBS J 275 (11):2668-76); anticalin, 衍生自脂质运载蛋白,其在 $\beta$ 桶折叠的开放端掺入小配体的结合位点(Skerra, A. (2008) FEBS J 275 (11):2677-83); 纳米体和DARPin。工程化的蛋白质支架通常靶向结合与抗体相同的抗原蛋白。短肽也可以用于结合靶蛋白。Phylomers是衍生自细菌基因组的天然结构化肽。此肽代表蛋白质结构折叠的多样化阵列,且可用于抑制/破坏体内的蛋白质-蛋白质相互作用(Watt, P.M. (2006) .Nat Biotechnol 24 (2):177-83)]。

[0054] 可替换地,该结合剂可以是适体。适体是通过形状互补和非共价化学键的组合以高亲和力和特异性来识别靶分子的合成寡核苷酸(DNA或RNA)(Blank&Blind,Current Opin.Chem.Biol.,2005,9:336-342)。这些人造配体在体外非常容易获得,并且可以开发以识别范围从纯离子(例如 $Pb^{2+}$ ,Liu&Lu,2003.J Am Chem Soc.,125,6642-6643)至核苷酸、小分子、蛋白质、病毒和细胞直到整个生物体(Menger et al.,2006.Handbook of Experimental Pharmacology,359-373)不等的各种各样的不同分子类别。已经通过众所周知的SELEX方法(Ellington&Szostak,1990.Nature,346,818-822)选择高结合亲和性适体,用于检测低分子量分子如茶碱(Jenison et al.,1994.Science,263,1425-1429)、L-精氨酸(Geiger et al.,1996.Nucl.Acids Res.,24,1029-1036)、默诺霉素(Schuerer et al.,2001.Bioorg.Med.Chem.,92,2557-2563)、17 $\beta$ -雌二醇(Kim et al.,2007.Biosens.Bioelectron.,22,2525-2531),还用于检测较大分子如凝血酶(凝血酶结合适体:5'-GGTTGGTGTGGTTGG-3')(Baldrich et al.,Anal Chem.2004,76,23,7053-63),霍乱毒素或HIV-1tat蛋白等(评论参见Tombelli et al.,2007,Biomolec Eng.,24,191-20)。上述一些适体已经用于微板上或生物传感器换能器(QCM,SPR)的表面上的ELISA样试验。还开发了适体修饰的AuNP比色体系用于在基于夹心的试验中确定蛋白质PDGF(Huang et al.,2005,77,5735-5741)。

[0055] 该结合剂可以为文库的一部分,例如展示的结合剂文库,例如细菌展示,mRNA展示,噬菌体展示,适体、核糖体展示或酵母展示文库。优选地,展示的结合剂文库为抗体噬菌体展示文库。文库应足够大,因此文库由预期结合至该靶样本中至少75%的感兴趣靶的多种结合成员构成。更优选地,文库设计为结合样品中至少80%,85%,90%,95%,97.5%或99%的感兴趣的靶。例如,该结合剂文库包含多个结合成员,该多个结合成员结合至覆盖95%或更多样品内的预期的或想要的蛋白质的蛋白质或肽序列。文献中公布了这些文库。文库的每个成员具有可检测的核酸身份标记,该核酸身份标记优选为对于文库的一个成员是单一的。优选地,连接单一核酸身份标记。“连接”是指连接过程具有在合适的试验条件下基于这些核酸身份标记的共定位形成随机多聚体核酸产物的潜力。优选地,多聚体产物为二聚体。合适的试验条件包括例如通过在脂质乳剂中热处理,优选在分离的区室中来分解噬菌体颗粒,并且单一序列的特异性共有扩增产生可连接的扩增子。可连接的扩增子的接合(例如,结合展示特异性核酸结构域)形成连接的身份标记,连接的身份标记编码身份标记的共定位信息。优选地,单一序列是结合展示的特异性核酸结构域,例如,编码一个或多个CDR区的序列。接合反应可以基于扩增或涉及其它技术。基于接合的扩增可以利用两个或更多个扩增引物对,扩增引物对具有相同的结合能力,但是具有互补的5'标记或二聚体接头序列,互补的5'标签或二聚体接头序列导致聚合酶可延伸的核酸双链体的形成。该标签或二聚体接头序列是指由一个引物对扩增的序列将与由第二引物对扩增的序列杂交。因此连接身份标记。

[0056] 结合剂文库的每个成员与单一核苷酸序列相关联,单一核苷酸序列可用于识别结合剂。本文所用的“相关联”是指可以通过在该方法中产生的连接序列内的核酸序列的存在来检测结合剂在复合体中的存在。核苷酸序列可以作为标记附着于结合剂,作为结合剂本身的一部分,例如,适体,或存在于结合剂内,例如噬菌体内的核酸。例如,文库的每个成员可以用单一核苷酸序列标记。本文所用的“标记的”是指附接至文库成员的核苷酸序列。本

领域已知将核苷酸附接至结合剂如抗体或化合物的方法。可替换地,如果结合剂文库为如上所述的展示文库,则单一核苷酸序列可以是编码一个或多个CDR区或展示的结合结构域的序列。例如,可以通过将编码要展示的氨基酸序列的序列插入至噬菌体中,在已知位置处,来产生展示文库。然后可以使用会扩增插入序列的通用引物,从而识别结合序列。或者,如果结合剂为适体,则适体本身可以是单一核苷酸序列。

[0057] 该核苷酸序列为寡核苷酸,可以包含单链或双链的RNA或DNA。用于标记结合剂或靶的核苷酸长度通常为5-150个碱基,例如长度为10-40个,或20-30个碱基。形成核酸的核苷酸可以在化学上修饰以增加分子的稳定性,改善其生物利用度或在该核苷酸赋予其额外的活性。例如,嘧啶碱基可以在6位或8位进行修饰,5位嘌呤碱基用CH<sub>3</sub>或卤素例如I,Br或Cl修饰。修饰或嘧啶碱基还包括2个2NH<sub>3</sub>、O<sup>6</sup>-CH<sub>3</sub>、N<sup>6</sup>-CH<sub>3</sub>和N<sup>2</sup>-CH<sub>3</sub>。2'位的修饰为糖修饰,通常包括NH<sub>2</sub>,F或OCH<sub>3</sub>基团。修改也可以包括3'和5'修改,如加帽。

[0058] 可以使用修饰的核苷酸,比如吗啉代核苷酸,锁核酸(LNA)和肽核酸(PNA)。吗啉代寡核苷酸由不同的吗啉代亚基组装而成,其中每个都含有连接至6元吗啉环的四种遗传碱基(腺嘌呤,胞嘧啶,鸟嘌呤和胸腺嘧啶)之一。亚基通过非离子型磷酸二亚胺亚基间连锁接合,得到吗啉代寡核苷酸。LNA单体的特征在于呋喃糖环构象受连接2'-O位置至4'-C位置的亚甲基接头的限制。PNA为DNA的类似物,其中骨架为假肽而不是糖。

[0059] 优选地,该结合剂能够(优选以不同的表观亲和力)检测一种以上的靶。可替换地,该结合剂能够(优选以不同的表观亲和力)使用不同表位或结合位点检测单个靶。

[0060] 可以预先确定抗体噬菌体文库的成员的结合特征。例如,可以确定哪些表位被编码并表达噬菌体的结合剂(抗体)的CDR绑定。该信息可以与编码CDR的单一核苷酸序列相关联。因此,可以从单一核苷酸序列的序列识别出由噬菌体表达的抗体绑定的表位。一旦存在于绑定的靶中的表位序列是已知的,则可能识别绑定的蛋白质或蛋白质组。

[0061] 可以使用标记的表位的单一核苷酸序列或标记表位文库的单一核苷酸序列来确定抗体噬菌体文库的成员的结合特征。

[0062] 靶

[0063] 本文所用的“靶”是与结合剂形成复合体的分子或分子组。该复合体通常在感兴趣的生物体的正常生理条件下形成。

[0064] 优选地,靶包含蛋白质。更优选地,靶为蛋白质样品的一部分。蛋白质样品可以包含蛋白质展示文库,优选其中所述文库的每个成员与单一核苷酸序列相关联。优选地,所述蛋白质展示文库为cDNA噬菌体展示文库。任选地,所述靶可以交联至多个靶(例如,蛋白质样品)内的其它靶。例如,样品内的蛋白质可以与样品中的一种或多种其它蛋白质交联。

[0065] 所述靶可以为已知靶。可以识别与靶形成复合体的结合剂,包括与靶相互作用的化合物。可替换地,靶可能是未知的,并且本发明的方法用于识别彼此相互作用的靶或多个靶分子。

[0066] 靶可以与单一的核苷酸序列相关联。“相关联”是指通过该方法产生的连接的序列内核酸序列的存在可以检测结合剂/靶复合体内靶的存在。核苷酸序列可以作为标记附着于靶上,或者存在于靶内,例如,噬菌体内的核酸。可替换地,该核苷酸序列可以是已知结合至靶的适体的一部分。可以将结合剂/靶复合体与适体接触,以通过单一核苷酸序列(包括适体)的连锁来识别存在的靶。

[0067] 本发明的试验可以应用于任何蛋白质样品。蛋白质可以衍生自任何生物试样,包括但不限于组织、细胞试样、体液、细胞培养物或任何其它含有蛋白质复合体的材料。体液样品包括血液、唾液、尿液、脑脊髓液或血清。可替换地,所述样品可以通过重组表达方法产生。可以使用本领域已知的标准方法由试样制备蛋白质。所述试样可以在提取前进行化学处理,例如可以使用不同的固定剂化学品或交联剂(例如BS3-(双(磺基琥珀酰亚胺基)辛二酸酯),蛋白质样品可以交联或不交联。可替换地,可以通过体外转录翻译体系或通过重组表达体系生产蛋白质。根据实验目的和所研究的蛋白质-蛋白质相互作用的类型,可以分析以变性或非变性形式和/或交联或非交联形式的蛋白质。蛋白质样品可以分析多个条件以收集关于多种蛋白质-蛋白质相互作用的定量结合特征的信息,例如可以改变结合剂的浓度或数量以确定解离常数和其动力学参数。

[0068] 可以预选蛋白质混合物。例如,蛋白质混合物可以是特异性蛋白质的富集,例如,来自特定细胞位置、来自特定细胞类型、相似大小或静电电荷的蛋白质的蛋白质,具有相似结合特征的蛋白质,类似的序列特征或类似功能,例如酶(Current Protocols in Molecular Biology (2006) 20.0.1-20.0.6 CHAPTER 20 Analysis of Protein Interactions)。优选地,特异性蛋白质包含磷蛋白、膜蛋白或天然的翻译后人工修饰的蛋白。该方法的蛋白质混合物中的蛋白质可以是变性的或非变性的和/或交联的或不交联的。

[0069] 该蛋白质可以为蛋白质展示文库的形式。实例包括细菌展示、mRNA展示、噬菌体展示和核糖体展示和酵母展示文库。优选地,该蛋白质展示文库为蛋白质噬菌体展示文库,更优选为cDNA噬菌体展示文库。该文库应该足够大,使得其由多种肽或蛋白质成员组成,其中预期在该样品中通过所述方法检测到覆盖至少70%的蛋白质。更优选地该文库足够大以覆盖样品中75%、80%、85%、90%、95%、97.5%、99%或更高的蛋白质或肽。优选地,该展示文库覆盖任何合适的生物实体,例如组织样品或整个生物体,例如,覆盖任何合适的生物实体的95%或更高的蛋白质。文献中公开了这些文库(Danner S, Belasco JG. T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. Proc Natl Acad Sci USA. 2001 Nov 6; 98(23):12954-9. Epub 2001 Oct 23. PubMed PMID:11606722; PubMed Central PMCID:PMC60806.)。文库的每个成员与单一核苷酸序列相关联,即每个成员具有单一的可检测的核酸身份标记。优选地,连接单一核酸身份标记。“连接”是指连接过程具有在合适的试验条件下基于这些核酸身份标记的共定位形成随机多聚体核酸产物的潜力。优选地,多聚体产物为二聚体。合适的试验条件包括例如通过在脂质乳剂中进行热处理来(优选在分离的区室中)分解噬菌体颗粒,并且单一序列的特异性共有扩增产生可连接的扩增子。可连接的扩增子的接合,例如,结合的展示特异性核酸结构域,形成连接的身份标记,编码身份标记的共定位信息。接合反应可以基于扩增或涉及其它技术。基于扩增的连接可以利用具有相同结合能力的两个或更多个扩增引物对,但是具有互补的5'标记或二聚体接头序列,其导致聚合酶可延伸的核酸双链体的形成。标记或二聚体接头序列意味着由一个引物对扩增的序列将与由第二引物对扩增的序列杂交。身份标记因此变得联系起来。

[0070] 在其生物背景中,身份标记,即用于结合剂文库和靶文库(比如蛋白质展示文库和抗体文库)的相关联的单一核苷酸序列,可能不同,因此扩增和接合过程基于两种不同的引物对,例如,一个引物对扩增靶序列,比如基于cDNA的身份标记,并且第二引物对扩增用作

身份标记的结合剂特异性核苷酸序列。连接不同的标记使得可以将结合剂特定信息与靶信息(例如,由展示的cDNA编码的蛋白质)关联。图2展示了这个过程的一个例子。

[0071] 一种结合剂,优选展示的抗体噬菌体,可以识别特异性靶,例如,展示的蛋白质靶和相应的蛋白质。这被称为特异性。可替换地,多种结合剂可以识别一个靶,比如特异性靶,例如,展示的蛋白质和相应的蛋白质。这被称为冗余。类似地,基于因例如蛋白质构象或蛋白质序列引起的靶构象的相似性,一种结合剂可以识别一种以上的靶,例如蛋白质种类。这种现象称为交叉反应。此外,靶蛋白的结合剂识别基于蛋白质或其蛋白质序列的构象。这被称为其反应性。结合剂(比如展示的结合剂)的蛋白结合亲和力可以从测序数据集的定量信息计算出来。展示的结合剂的成员的预定结合特征可以包括具有特异性的反应性和交叉反应性以及具有计算的亲和力的冗余性。

[0072] 可以使用这些措施来计算检测到的靶-靶相互作用,例如,蛋白质-蛋白质相互作用。通过基于反应性、特异性和冗余度的输入以及连接的身份标记的身份和丰度进行计算,可以看透特异性靶-靶相互作用,例如蛋白质-蛋白质相互作用。类似地,使用这些措施可以减少由交叉反应性引起的计算的不确定性,其中考虑展示的抗体试剂的冗余性和亲和力。可以使用不同浓度的结合剂和/或靶来计算多种相互作用(例如蛋白质-蛋白质相互作用)的定量参数。因此优选地,使用不同浓度的结合剂和/或靶实施该方法。

[0073] 检测的定量性质使得可以通过身份标记的非特异性共定位或相同身份标记的自连接来确定和计算背景、非信息性排序读数。然而,自我链接标记的检测包含有关数据集质量的信息。

[0074] 为了实现足够的覆盖,可以使用富集的库。可以富集结合剂文库和/或靶文库。可以富集展示的靶,比如蛋白质,以涵盖实验环境中的所有潜在的结合伴侣。类似地,可以选择展示的结合剂以在实验环境中使结合特异性朝向可检测靶富集。例如,该文库可能限制为感兴趣的特定靶。这可以通过基于淘选的选择来进行,例如,具有受控的复合度的靶样品固定在固体表面上并与结合剂文库接触,通过洗涤和随后洗脱绑定剂来选择结合剂文库的绑定剂。类似地,将具有受控的复合度的靶样品固定在固体表面上并与靶文库接触,通过洗涤和随后洗脱绑定的靶来选择靶文库的绑定靶。

[0075] 优选地,富集意味着自结合的展示的结合剂的去除,即,在试验条件下结合至其它结合剂的结合剂。更优选地,富集意味着降低所展示的结合剂文库的复合度,但仍然确保其具有高覆盖率以检测实验环境中的靶。减少复合度可能涉及降低文库内的成员的总数,消除接合至非靶蛋白质的成员,或只选择结合至感兴趣的靶的那些成员。

[0076] 可以通过获得具有感兴趣的成分的蛋白质混合物来富集结合剂文库,例如,研究已知的蛋白质-蛋白质相互作用并动态地验证其相互作用或测试激动化合物或拮抗化合物的作用。例如,将感兴趣的靶,例如,蛋白质混合物固定在固体表面上允许结合剂文库的成员结合至期望的靶,例如,蛋白。可以通过稀释或其它方式将绑定至感兴趣的靶的结合剂与未绑定的结合剂分离。在本发明方法中,如上所述,结合剂文库的绑定成员可以洗脱并用作复合度降低的文库。

[0077] 使用富集的结合剂文库通常是有利的,因为分离方法(例如基于微阵列或乳液的分离)在一个区室内产生单个复合体的能力有限,从而可以获得连接的可识别序列。富集展示剂文库的复合性降低可以直接转换为需要进行的分离次数,以获得基于分离的随机配

对的接合的扩增产物。

[0078] 结合剂的结合

[0079] 本发明的方法优选在生理条件下进行,使得可以在其原始环境中检测相互作用,即在与存在于细胞中的相同条件下。这提供了有关自然发生的结合相互作用的信息。

[0080] 将结合剂文库与靶接触的步骤通常在已知缓冲体系(例如,已经用于研究蛋白质-蛋白质相互作用(例如TBST-缓冲液)的缓冲体系中)中进行。取决于亲和力,该反应可以在室温或4℃下进行。为了获得可重现信号,确定最佳时间、最佳温度和其它试验条件,包括结合、洗涤和检测步骤。可由本领域技术人员确定最佳条件。

[0081] 蛋白质复合体与绑定的展示的抗体试剂分离

[0082] 需要分离结合剂/靶复合体,即在连接标记序列(即,相关联的单一核苷酸序列)之前与其它复合体分离。分离通过本领域已知的方法进行。例如,分离可以通过稀释、特异性结合或通过物理和/或化学性质来进行分离。优选地,将复合体分离成区室,例如乳液液滴、微孔等,优选扩散受限或分离的区室。

[0083] 优选地,所述分离包括固体表面结合、在其它物质中稀释或相分离,或提供扩散受限或分离的区室中的任何一种或多种。该分离限制了区室中未绑定的结合剂的数目,优选地平均为一个。例如,区室中的未绑定的结合剂的平均数为一个。该区室可以为乳液内的单个液滴,或者诸如微腔的单个物理室。可以根据物理或化学特征分离该复合体。优选地,所述稀释为有限的稀释。

[0084] 分隔(例如有效分离或大量反应分离)是基于单个未绑定噬菌体的基于泊松分布的分离;例如,乳液和微阵列是最知名的现有技术的方法。

[0085] 在进一步分析之前将结合剂/靶复合体充分分离将提供产生的核酸标记对的情形,该情形基于结合剂的共定位。如果复合体没有充分分离,不同复合体的成员的核酸标记将会被连接,从而提供虚假信息。分离降低核酸身份标记的非特异性共定位的量,并能够识别特异性结合伴侣,特别是当研究复合体蛋白质混合物时。例如,每个区室平均可能导致单个未绑定的结合剂的分离,其中连接将仅提供自连接的核苷酸序列,从而降低结合剂文库成员之间随机连接的可能性。由于任何试剂的分配都是基于泊松分配,所以可以计算出实现试剂的合适分离的必要措施。这优选会导致在一个区室内有单个复合体。优选地,使用乳剂,其可用于扩增和连接特异性核酸结构域以形成连接的身份标记。是本领域技术人员周知乳液扩增方法,例如Schütze et al.,Anal.Biochem.2011 March 1;410(1):155-7。

[0086] 用于固定化的固体表面

[0087] 任选地,在分离后固定结合剂/靶复合体。例如,结合剂/靶复合体可以捕获于表面上,例如作为阵列的一部分。这有助于维持复合体之间的分离。优选地,例如在连接步骤期间维持复合体的分离。

[0088] 结合剂/靶复合体任选地固定在固体支持物表面上,包括但不限于膜,例如,聚偏二氟乙烯(PVDF)或硝化纤维素、塑料表面(例如,聚苯乙烯),或者可以共价偶联至合适的珠粒上(例如环氧活化的珠粒)。通过蛋白质的标准方法("Antibodies,a Laboratory Manual."Harlow,E.,and Lane,D.,eds.Cold Spring Harbor Press,Cold Spring Harbor N.Y.,1988)或通过包括抗体的特异性结合或其它特异性结合相互作用,例如生物素-抗生物素蛋白结合或偶联至固体表面。

[0089] 固定化可以包括(a)获得具有通用的已识别的蛋白质复合体结合能力的固体支持物,即固体支持物能够结合所述结合剂文库或靶分子的所有成员。已识别的蛋白质复合体的量和固体支持物上可用的结合位点的数目应该是平衡的,以便在绑定的已识别的蛋白质复合体之间实现充分的分离。优选地,所述固体支持物包括膜、塑料表面或珠粒。更优选地,固体支持物为珠粒,并且实现了分离,其中平均一个结合剂/靶复合体绑定至一个珠粒。更优选地,由抗噬菌体抗体提供通用的已识别的蛋白质复合体的结合能力。优选地,分离包括反应室中的物理分离或乳液内液滴的物理-化学分离。

[0090] 优选地,所述随机配对连接的核酸产物的产生包括使用至少两对PCR引物来扩增相同或不不同的扩增子;其中5'端的PCR引物具有序列标签,其中用带标签的引物进行扩增导致随机配对的连接的核酸产物。更优选地,扩增为乳液PCR扩增,并且所述扩增子和随机配对连接的核酸产物的产生为平行进程。优选地,所述接合的扩增产物的所述测序是高度平行的测序方法。分离允许连接的身份标记的特异性形成,其中共定位在分离的或扩散受限的区室中(例如在乳液中的固体表面上)的分子具有特异性连接仅有的绑定的结合剂的身份标记序列的倾向。然而,也可以连接未绑定但意外共定位的结合剂的身份标记。类似地,可以连接未绑定在复合体内的结合剂的身份标记。该连接过程可以提供具有两个相同身份标签的连接的身份标记,例如。其中在区室中仅存在一种结合剂。类似地,在识别靶的结合剂有一种以上的情况下,可以以相同的结合特异性产生不同的连接的身份标记。该连接步骤接合存在的结合剂文库的一个或多个成员的身份标记。此外,该连接步骤可以将结合剂文库的成员的身份标记接合至与靶相关的核酸序列。该连接过程不依赖于由于其紧密接近(例如,通过彼此杂交)而彼此相互作用的核酸。通过使用连接过程(例如,通过本文所述的扩增方法)将序列接合在一起。使用这些方法,而不是依靠紧密的物理接近连接试验,例如US7306904中的方法,允许并行检测多个相互作用。该标记不需要紧密接近以进行连接,从而进行检测。该标记只需要处于相同的区室内。

[0091] 核苷酸序列的连锁

[0092] 与复合体成员相关联的单一核苷酸序列可以通过以下方法连接,该方法包括:

[0093] (i) 使用至少两对PCR引物扩增与结合剂相关联的核苷酸序列,以及任选地如果存在与靶相关联的序列,以产生至少两组扩增子,其中设计引物成使第一组扩增子包含与第二组扩增子中的序列互补的序列;

[0094] (ii) 使至少两组扩增子退火;以及

[0095] (iii) 进行扩增反应以产生连接的核苷酸序列

[0096] 步骤i-iii可以依序或同时进行。

[0097] 每对引物包括正向引物和反向引物。这些引物的序列设计成能够对可识别的序列进行扩增。优选地,这些引物为通用引物,即,该引物结合文库中的所有可识别序列,例如,结合剂文库或靶文库的所有成员。对于文库的一个成员,引物之间的扩增的序列是单一的,能够识别。优选地,将成对的PCR引物设计成产生二聚体连接的核酸序列,否则产生多聚体连接的核酸序列。这通过使用至少两个引物的5'端的二聚体接头序列扩增二聚体或多聚体的成员扩增子来实现。因此,这些扩增子可以在其3'端形成部分重叠且聚合酶可延伸的杂交产物(例如,如图2所示)。

[0098] 优选地,该扩增为乳液PCR扩增,并且适当分离与结合剂相关的核苷酸序列或结合



剂和靶。

[0099] 任选地,在识别连接的结合剂和/或存在的靶之前,组合来自一种以上的复合体的连接的核苷酸序列。可以例如通过测序连接的核苷酸序列来确定结合剂和/或靶的身份。这可以使用高度并行的体系进行。可以组合连接的序列,从而可以进行单个反应以识别所有连接的序列。例如,所有连接的核苷酸可以在单个反应中进行测序。可以定量地确定连接的序列,以测量连接的序列的相对丰度。

[0100] 确定所述连接的核酸产物的信息

[0101] 基于结合剂的预定的结合特征,可以从连接的身份标签推断共定位信息。多个连接的身份标签,即,测序读数计数的形式的与结合剂和任选的靶相关联的单一核苷酸序列,提供了结合剂/靶复合体成员的身份及其相对丰度的信息。其它信息也是可推断的,其中会考虑所有连接的身份标签的信息。实例包括但不限于比较结合剂的相对亲和力与预定亲和力、比较针对不同结合剂计算的相对丰度、确定绑定的和未绑定的靶或蛋白质比率。

[0102] 基于抗体结合剂的CDR区(CDR1,CDR2和CDR3)的多聚体PCR连接,可以使用NGS测序确定所有CDR区的序列。由于结合剂的预定结合特征优选地基于单个CDR区序列身份,在一个优选的实施例中,抗体结合剂的完整序列信息可以与其预定的结合特征相关联。

[0103] 确定化合物对多种蛋白质-蛋白质相互作用的影响。

[0104] 本发明的方法可以包括在存在或不存在化合物的情况下使结合剂文库与靶接触的步骤;以确定所述化合物是否影响结合相互作用。该方法可用于确定化合物或其它化学部分对多种结合相互作用,例如蛋白质-蛋白质相互作用,的影响。优选地,该方法用于检测可以促进或破坏某些蛋白质-蛋白质或其它可检测的相互作用的化合物或化学部分,其中所述化合物或化学部分用作药物或消除或抑制这种情况的病理后果。优选地,这些药物可用于治疗不同的疾病,包括但不限于癌症、感染性疾病、自身免疫性疾病等。

[0105] 本文所用的化合物是指通过共价键连接的两个或更多个原子。化学部分是形成官能团的化合物的一部分。该化合物可以为已知的药剂。

[0106] 本发明的方法可以包括在存在或不存在化合物的情况下使结合剂文库与靶接触的步骤;以确定所述化合物是否影响结合相互作用。

[0107] 在另一实施例中,本发明提供了确定相互作用组数据的方法。在多个区室中,结合剂(例如展示的抗体试剂)根据其结合特征共定位,即,两种或更多种结合剂因其结合相同的靶而存在于一个区室内,或结合至其自身相互作用或相互绑定靶。连接其可识别序列,即,单一核苷酸序列,从而以连接身份形式携带共定位信息。在结合特征信息和共定位信息的基础上,可以确定蛋白质-蛋白质的相互作用和蛋白质的身份信息。

[0108] 本发明还描述了一种用于确定获得的试样中蛋白质-蛋白质相互作用的方法,所述方法包括:

[0109] (a) 获得蛋白质混合物;

[0110] (b) 使结合剂文库与所述蛋白质混合物接触,从而形成结合剂-蛋白质复合体,其中结合剂文库的每个成员与单一核苷酸序列相关联;

[0111] (c) 分离所述结合剂-蛋白质复合体;

[0112] (d) 任选地将所述结合剂蛋白质复合体固定在固体表面上;

[0113] (e) 检测所述结合剂-蛋白质复合体内的所述结合剂的所述单一核苷酸序列,并连



接所述结合剂的所述单一核苷酸序列,从而基于所述分离提供的共定位提供连接的核酸产物;

[0114] (f) 任选地组合所述连接的核酸产物;以及

[0115] (g) 对所述连接的核酸产物进行测序,所述连接的核酸产物对应所述展示的抗体试剂的蛋白质结合特征。连接的核酸产物的序列用于推断所述试样的所述蛋白质混合物中蛋白质-蛋白质相互作用的存在。单一核苷酸序列能够识别存在的抗体试剂,以及所述展示的抗体试剂的相应的蛋白结合特征。

[0116] 可以通过统计学方法验证蛋白质-蛋白质相互作用数据,包括确定和扣除蛋白质-蛋白质相互作用水平的背景,蛋白质的表观相对亲和力和相对丰度的确定以及蛋白质-蛋白质相互作用。

[0117] 本发明还涉及用于确定化合物对试样中存在的蛋白质-蛋白质相互作用的影响的方法,所述方法包括:

[0118] (a) 在存在和不存在化合物的情况下获得蛋白质混合物;

[0119] (b) 使结合剂文库与所述蛋白质混合物接触,从而形成结合剂-蛋白质复合体,其中所述结合剂文库的每个成员与单一核苷酸序列相关联;

[0120] (c) 分离所述结合剂-蛋白质复合体;

[0121] (d) 任选地将所述结合剂蛋白质复合体固定在固体表面上;

[0122] (e) 检测与在所述结合剂-蛋白质复合体内的所述结合剂相关联的所述单一核苷酸序列,并连接与所述结合剂相关联的所述单一核苷酸序列,以提供基于由所述分离提供的共定位的随机配对连接的核酸产物;

[0123] (f) 任选地组合所述连接的核酸产物;和

[0124] (g) 对所述连接的核酸产物进行测序;

[0125] (h) 比较在存在和不存在化合物的情况下获得的信息;

[0126] 结合剂的单一核苷酸序列使得存在的结合剂能够被识别,因此使用所述结合剂的预定蛋白结合特征来推断所述试样中蛋白质混合物的蛋白质-蛋白质相互作用的存在。

[0127] 优选地,化合物的影响的测定使用高通量实验设备来确定。

[0128] 在本发明的另一个实施例中,提供一种用于确定结合剂文库的成员的蛋白质结合特征的方法,所述方法包括:

[0129] (a) 获得展示的蛋白质文库,其中每个成员与单一核苷酸序列相关联;

[0130] (b) 使结合剂文库与所述展示的蛋白质文库接触,以形成结合剂/蛋白质复合体,其中所述结合剂文库的每个成员与单一核苷酸序列相关联;

[0131] (c) 分离所述结合剂/蛋白质复合体;

[0132] (d) 任选地将所述分离的结合剂/蛋白质复合体固定在固体表面上;

[0133] (e) 任选地在保持复合体分离的同时,连接与结合剂和蛋白质相关联的单一核苷酸序列以产生连接的核酸产物;

[0134] (f) 任选地组合连接的核酸产物;

[0135] (g) 确定连接的核酸产物的序列。

[0136] 可以从连接的核酸产物的序列内的信息确定结合剂文库成员的蛋白质结合特征。与展示的试剂文库的成员和蛋白质相关联的连接的单一核苷酸序列的检测指示识别并结

合展示的蛋白质文库的某些成员。该序列可以提供关于文库哪些成员结合哪些蛋白质的信息。

[0137] 可以计算多种所述结合剂的蛋白质结合特征。可以组合所述结合剂文库的所有成员的所有结合特征信息作为结合信息。

[0138] 基于由所述分离提供的共定位,与所述结合剂和所述蛋白质相关联的所述单一核苷酸序列的连接提供随机配对的连接的核酸产物。

[0139] 优选地,在多个测量中使用不同浓度的所述结合剂文库和/或靶(例如蛋白质样品),从而能够计算定量结合信息,包括所述多种蛋白质-蛋白质相互作用的解离常数。本领域技术人员已知计算这些措施的方法。

[0140] 本发明还涉及用于实施本发明的方法的试剂盒。所述试剂盒包括:

[0141] a) 结合剂文库,其中所述结合剂文库的每个成员与单一核苷酸序列相关联;以及

[0142] b) 用于连接与结合剂相关联的核苷酸序列的至少两对引物组;以及任选的使用说明。

[0143] 所述试剂盒还包括用于检测蛋白质-蛋白质相互作用的装置,其中提供试剂和任选的材料以实施以下步骤中任何一个或多个:分离、固定、检测单一核苷酸序列,核苷酸序列的连接和/或检测连接的核苷酸序列以获得所述共定位信息。所述试剂盒还可以包括用于实施本发明的方法和利用所述试剂盒的说明书。

[0144] 本发明的试剂盒还可以包含蛋白质展示文库,其中所述文库的每个成员与单一核苷酸序列相关联。

[0145] 本发明方法与现有方法的比较

[0146] 本发明的方法是以有时间和成本效益的方式捕获相互作用组的信息的新方法,从而能够进行随机抽样和高冗余抽样。该方法提供基于动态的、原始的细胞环境(例如,生理的),基于天然蛋白质-蛋白质相互作用,以及全面覆盖的甚至较大的多单位蛋白质复合体的定量相互作用数据。该方法抑制随机变量(比如,检测非特异性、意外相互作用的蛋白质)的影响。该方法还降低变量(比如,除了原始的蛋白质-蛋白质相互作用之外的涉及检测原理的任何结合事件相关的变量)的影响。该方法适用于不仅与诱饵蛋白相互作用,而且还与DNA、RNA和化学化合物相互作用的蛋白质的体外检测。

[0147] 使用基于PCR的邻位连接试验(LGA)的检测来确定靶的相对表达。由于蛋白质和mRNA表达谱不相同,相对于生物学过程,所观察到的差异中可能是重要的。该试验能够在给定实验中检测若干靶。然而,由于需要制备大量人为标记的特异性抗体,因此在相互作用组的水平上是不可行的。没有高度特异性的抗体,交叉反应性降低了邻位连接试验清楚地区分特异性和非特异性相互作用的能力。数量非常大的靶的并行检测成本高昂且麻烦。

[0148] 已经开发了多路复用形式的LGA。在该试验中,固定在固体支持物上的抗体用作捕获试剂以从蛋白质的复合体混合物局部富集抗原。洗涤后,添加成对的邻位连接试验(PLA)探针。接着进一步洗涤并连接邻位带来的寡核苷酸。LGA可以基于对三个结合事件的需要提供较高的特异性。结合使用PCR扩增,LGA能够具有高特异性和灵敏度,以及广泛动态范围的蛋白质定量。该方法与下一代测序(NGS)相结合,以数字记录蛋白质丰度的模式,并用于验证36种蛋白质分析物的同时检测。

[0149] LGA的变形被描述为用于检测复合体靶结构(比如微泡)的极端敏感和特异性试验

(4PLA), 其中首先通过固定化抗体捕获靶, 然后通过使用具有附着的DNA链的四种其它抗体进行检测。通过五种抗体的同时结合以产生可扩增的报道子的要求导致特异性和灵敏度增加。

[0150] 使用邻位连接的所有基于邻位连接的试验类型由于非常频繁的空间限制而需要实验验证。

[0151] 在本发明的方法的情况下, 低特异性不会成为问题, 甚至可以用作验证信息。本发明的方法使用共定位和区室化, 其中扩增的身份标记在区室中自由扩散, 使得能够实现更宽松的空间条件。

[0152] 基于双重表达重组酶 (DERB) 的目的载体单独编码用于插入感兴趣的蛋白质开放阅读框 (ORF) 的两组重组酶可识别序列, 在具有用于检测相互作用的ORF的框架中的两组启动子和报道子标签。将载体引入活细胞 (原核的和真核的) 能够通过荧光共振能量转移 (FRET) 或双分子荧光互补 (BiFC) 检测蛋白质相互作用。DERB平台通过引入基于重组酶的克隆和经由原理证明实验验证的可兼容接受载体以及未知相互作用的识别, 显示出优于当前商业化体系的优点。该体系需要大量筛选的相互作用, 因此需要大量的尝试和成本, 仅适用于相互作用组水平的机器人体系, 并且使用人工测试条件 (融合蛋白和人造启动子)。

[0153] 酵母双杂交 (Y2H) 筛选是蛋白质片段互补试验或PCA的具体实施, 其中蛋白质-蛋白质相互作用的识别基于两个蛋白质片段, 每个蛋白质片段共价连接至第三蛋白质的不完整片段 (例如DHFR, 其充当报道子)。蛋白质之间的相互作用使得报道蛋白的片段足够密切接近, 使得其能够形成可以测量其活性的功能性报道蛋白。这个原理可以应用于许多不同的报道蛋白, 比如酵母双杂交筛选使用GAL4转录因子。酵母双杂交筛选研究酵母核内人工融合蛋白之间的相互作用。该方法假阳性比率高, 因此有必要以其它方式验证识别的相互作用。该方法不在细胞环境中进行, 缺乏自然环境限制了其在由蛋白质的细胞环境特异性修饰所决定的相互作用或在低亲和力相互作用的情况下的用途。在相互作用组水平, 该方法需要进一步优化并使用阵列, 以使构建非常大的互作用组数据集成为可能。这涉及高成本, 但仍然不能克服该试验类型的所有限制。

[0154] 可以使用双诱饵体系, 通过立即选择来提高文库筛选的准确性, 以消除假阳性。

[0155] 已经描述了卷曲螺旋介导的异源二聚化功能相互作用阱试验, 其中卷曲螺旋异二聚化结构域被替代为模块化的蛋白质结合结构域。这可用于验证功能相关的蛋白质-蛋白质相互作用, 将酶引向特定底物, 以及筛选功能上重要的相互作用伴侣的融合文库。

[0156] 针对已知的Y2H筛选限制, 开发了基于哺乳动物细胞的两种杂交 (M2H) 体系。该M2H体系在通过将感兴趣的每个蛋白质对分别融合至DNA结合和反式转录激活结构域来研究相互作用上与酵母双杂交体系相似。与基于酵母的试验相比, 基于哺乳动物细胞的双杂交技术具有许多优点, 并解决一些已知问题。由于酵母缺乏参与翻译后修饰的关键蛋白质, 因此不能测定基于这些蛋白质的相互作用。此外, 可以使用若干种不同的哺乳动物细胞环境来提供细胞环境下的特异性相互作用组数据。然而, 由于需要处理非常大量的哺乳动物细胞培养试样, 因此难以产生大数据集, 所以无法实现相互作用的相互作用组水平。

[0157] 在Y2H筛选中检测蛋白质-蛋白质相互作用的变形是使用特定的基于PCR的测序方法, 称为Stitch-seq。Stitch-seq为PCR拼接, 其在相同的PCR扩增子上放置编码相互作用蛋白质的成对的序列。PCR拼接由两轮PCR组成。在第一轮中, 分别用DB-和AD载体特异性上游

引物扩增X和Y(存在于Y2H DB-X和AD-Y载体上)。第一轮扩增子作为模板用于产生由通过82-bp接头序列连接的X和Y ORF组成的连锁PCR产物。将PCR产物合并并通过下一代DNA测序进行测序以产生拼接的IST(sIST)。Stitch-seq已经移除了一些Y2H方案的瓶颈,但仍然没有解决与关键步骤相关联的问题。

[0158] 描述对PCA进行以下改进的专利:1) 报道基因(及检测其表达的方法),其能够轻易分析大型文库(大小 $>10^7$ )且可以轻易地“调整”、修饰和/或监测其选择性,2) 多重相互作用的同时和独立测量的方法(通过判断不同报道基因的表达),以及3) 使用基于噬菌粒的体系构建文库,基于噬菌粒的体系提供a) 用于执行文库对比文库实验的可自动化的有效方法,和b) 简化来自原核PCA中进行的任何筛选/选择的阳性候选物的分析的方法。使用基于噬菌粒的技术来筛选文库对比文库实验涉及跨文库,例如,用噬菌体的猎物文库感染细胞的诱饵文库(在噬菌体上使用过量的细胞,以确保每个细胞平均只被一个噬菌体感染),并寻找报道基因的激活表达。这是朝向大规模相互作用组扫描的重要一步,但它既不是动态的也不是基于细胞环境。

[0159] 对于在酵母中进行的文库与文库实验,使用从起始单倍体细胞中携带DNA的二倍体 $\alpha$ 细胞的形成。因此,携带猎物杂种文库的细胞可以与携带测试诱饵杂交体的 $\alpha$ 细胞配对。虽然这消除了转换效率问题,但它并没有解决理想的交互式扫描的其他要求。

[0160] 在与疏水核心相邻的位置半随机化的两个亮氨酸拉链文库与酶小鼠二氢叶酸还原酶(mDHFR)的两个设计片段中的任一个遗传融合,并共转化到大肠杆菌中。需要文库多肽之间的相互作用来重构mDHFR的酶活性,从而允许细菌生长。然而,该策略受到可以在细菌细胞中实现的转化效率的限制。

[0161] 还设计使GST下拉试验适应96孔滤板格式。使用多孔滤板使得可以以比传统的单管试验更少的试剂和更有效的样品处理在更短的时间内分析更多的样品。这种试验类型解决了一些造成技术瓶颈的问题;然而,使用该体系生成所需的非常大的数据集是不可行的。

[0162] 串联亲和纯化(TAP)方法,可以看作是更具体的共免疫沉淀版本,能够高通量识别蛋白质相互作用。可以将该方法的准确性与小规模共免疫沉淀实验进行比较,并在正确的细胞环境中检测相互作用。然而,该方法需要蛋白质纯化的两个连续步骤,因此不能轻易地检测瞬时蛋白质-蛋白质相互作用。TAP方法将TAP标签的融合应用于正在研究的蛋白质的C端。TAP标签由来自N-末端的钙调蛋白结合肽(CBP),随后是烟草蚀斑病毒蛋白酶(TEV蛋白酶)切割位点和蛋白质A组成。该方法能够真实定量测定正确的细胞环境中的蛋白质伴侣,但是,在相互作用组的水平上,该方法需要大量的尝试,并且会花费大量成本从而以所使用的构建体涵盖整个蛋白质组。

[0163] 针对这个问题,开发了无标签策略,使用多种分离技术,包括非变性高性能多肽离子交换层析,蔗糖梯度离心和等电聚焦,进行可溶性人蛋白质相互作用组的系统化,高度广泛的生化分馏。该方法需要验证和统计分析以产生可靠的数据集,还需要大量的样品材料,并且其成本限制了该方法在随机抽样和高抽样冗余度上的使用。

[0164] 蛋白质微阵列引入了一种新方法来识别和表征蛋白质相互作用,提供了在单个实验中快速识别成千上万个蛋白质之间的新相互作用的能力。由于阵列上每个蛋白质的位置和身份是已知的,所以可以从蛋白质阵列的迭代探测中快速发展相互作用图谱。因为在一天内进行蛋白质微阵列实验,并且在千上万个其它蛋白质的环境中评估相互作用,微阵列

上的相互作用分布图可以极大地加速发现新蛋白质相互作用的速率。此外,蛋白质微阵列实验的体外性质允许控制影响蛋白质相互作用的探测条件,例如蛋白质浓度、翻译后修饰和辅因子的存在,这可能无法用其它方法如酵母双杂交筛选。然而,经典的一个时间对一个探针的方法不适合较大的相互作用水平的实验。

[0165] 使用基于蛋白质阵列的检测版本,其中细胞蛋白裂解物或合成肽混合物应用于具有固定的诱饵蛋白/肽的蛋白质阵列。非特异性蛋白质/肽在各种严格条件下洗去,只有与诱饵蛋白质/肽特异性相互作用的蛋白质保留在芯片上。最后,通过SELDI-TOF质谱分析捕获的相互作用的蛋白质/肽复合体,并通过其预测的区别性质量确认其身份。这是非常有前景的方法,但是SELDI-TOF的蛋白质测序受到若干因素(蛋白质的量、分离、翻译后修饰)和相互作用缺乏自然环境的限制。

[0166] 蛋白质混合物也可以固定在固体支持物上,并在合适的结合条件下与多个未标记的蛋白质-蛋白质相互作用结构域接触。在存在至少一个标记的选择的蛋白质-蛋白质相互作用结构域(标记的蛋白质-蛋白质相互作用结构域不同于未标记的蛋白质-蛋白质相互作用结构域)的情况下,测量标记的蛋白质-蛋白质相互作用结构域的结合。这种方法是相互作用结构域特定的,这限制了其应用。

[0167] 结合下一代测序(NGS)的无细胞展示技术可以提高相互作用组数据集的覆盖面和可靠性。完全无细胞的方法提供了高通量且较大的检测空间,不使用克隆来测试相互作用。NGS提供的定量信息减少了假阳性的量。该方法适用于不仅与诱饵蛋白相互作用,而且还与DNA、RNA和化学化合物相互作用的蛋白质的体外检测。该方法采用cDNA文库(从细胞和组织中提取)进行完全体外处理,并对靶蛋白进行选择,以获得用于NGS的选定cDNA序列。使用该方法的选择在无细胞条件下进行,NGS的随后测序不受使用任何种类的细胞的克隆步骤的限制。这种方法将一个在时间上的方法应用于细胞环境外,限制了其在生成相互作用组水平的大型数据集的能力。

[0168] 另一种无细胞试验使用所谓的内含肽(intein),内含肽是能够在体外引导蛋白质反式剪接的肽序列。内含肽为在蛋白质剪接期间从蛋白质前体中切下的蛋白质前体中的插入蛋白质序列。提供两种杂交融合构建体,其中一种具有第一测试试剂和N端内含肽片段或N-内含肽,另一种具有第二测试试剂和C-端内含肽片段或C-内含肽。此外,一种或两种融合构建体可以具有在融合构建体的翻译-剪接后经历可检测的变化的报道子。通量和无细胞环境特征都是该方法的显著的缺点。

[0169] 用于研究蛋白质DNA或蛋白质-蛋白质相互作用的另一种方法是噬菌体展示的方法。蛋白质展示在编码展示的蛋白质的DNA的丝状噬菌体(例如M13)的表面上。感兴趣的目标蛋白质或DNA序列固定在固体支持物上,并用于亲和力-富集用于结合至靶的候选物的噬菌体展示的蛋白质文库。该方法已经用于识别和表征蛋白质-DNA和蛋白质-蛋白质相互作用。噬菌体展示是一个富集过程,需要多个周期推测蛋白质-蛋白质相互作用数据。富集在体外进行,偏倚相互作用并且有利于检测中的高亲和力相互作用。某些蛋白质(特别是较大的蛋白质)不太适合通过噬菌体展示进行分析。基于细菌的ORF噬菌体展示的主要缺点是噬菌体表面上展示的蛋白质缺乏适当的翻译后修饰,如糖基化。

[0170] 如上所述,本领域所使用的噬菌体展示仅限于再现自然结合事件。然而,表达为噬菌体展示的抗体谱在提供用于治疗性抗体的药理学引导或用于诊断试验的检测抗体方

面非常成功。本发明在后一种情况下应用噬菌体展示技术,其中较大的噬菌体抗体文库包含几乎任何可识别的覆盖多种特异性的靶。

[0171] 噬菌体展示的变形涉及用于选择和识别相互作用伴侣的方法。靶分子(配体)固定在固相载体的表面上,使得其位于二维等级可寻址的位置并与蛋白质展示病毒接触。通过检测和确定固定的配体和相互作用伴侣之间的结合位置来识别相互作用伴侣。所描述的优选的检测方法为表面等离子体共振(SPR)。

[0172] 在尝试分离/识别相互作用的蛋白质之前,化学交联常用于将蛋白质相互作用“稳固”于合适位置。用于该应用的常用交联剂包括不可裂解的[NHS-酯]交联剂、[双磺基琥珀酰亚胺辛二酸酯](BS3)、BS3的可切割版、[二硫代双(磺基琥珀酰亚基丙酸酯)](DTSSP)和在ChIP试验中流行用于稳固相互作用的[亚氨酯]交联剂[二甲基二硫代双丙酰亚胺酯(DTBP)]。

[0173] 已经开发了用于识别使用基因编码肽的噬菌体展示文库特异性结合预测的转录调节元件的蛋白质的技术,该蛋白质结合包含感兴趣的DNA基序序列的表面固定的双链DNA。在对特定的DNA-蛋白质相互作用进行富集后,扩增绑定的噬菌体,并对来自富集噬菌体的插入片段进行测序,从而使用标记和与DNA微阵列杂交来确定相互作用的蛋白质。

[0174] 据报道,使用chemFET阵列可测量一种或多种分析物。阵列可以包括提供关于感兴趣的化学过程的相关信息,包括抗体与抗原的结合,的各种化学物质中的任何一种。在一些方面,除了仅检测分析物的存在之外,测量一种或多种分析物的水平或浓度的能力提供与化学过程有关的有价值的信息。

[0175] 从上述描述可以看出,本发明提供了用于检测和表征蛋白质-蛋白质相互作用以及选择能够调节蛋白质-蛋白质相互作用的化合物的功能强大且通用的体外体系。该体系可以非常方便地使用,并且可以轻易地适应高通量筛选程序。

[0176] 本说明书中提及的所有出版物和专利申请都表明本发明所属领域的技术人员的水平。所有出版物和专利申请通过引用并入本文,其程度如同具体和单独地指定将每个单独的出版物或专利申请通过引用并入。

[0177] 虽然为了清楚理解的目的已经通过说明和实例的方式对本发明进行了详细的描述,但显然可以在所附权利要求的范围内实施某些改变和修改。

[0178] 实施例1

[0179] 噬菌体裂解和液滴数字PCR检测

[0180] pBluescript II SK(+)噬菌粒载体(Agilent,212205)f1起源于(+)取向,宿主菌株中的Sac<sup>+</sup>→Kpn多接头取向:XL1-Blue MRF'已用于产生pBluescript II SK(+)噬菌体且购买M13K07辅助噬菌体(M13K07 Helper Phage,NEB,N0315S)。对噬菌体进行滴定以确定感染性噬菌体的数量。将噬菌体连续稀释10倍,以实现每区室稀释液低于单个噬菌体。使用QX200 Droplet Digital PCR(ddPCR<sup>TM</sup>)System对稀释液进行数字PCR。简而言之,使用5'CTCAAGTCGGTGACGGTGAT3'(M13K07特异性正向),5'GACAAAAGGGCGACATTCAA3'(M13K07特异性反向)和/或5'TCTTGATCCGGCAAACAAAC3'(pBluescript II SK(+)特异性正向),5'TTTTCTGCGCGTAATCTGCT3'(pBluescript II SK(+)特异性反向)以及探针5'CTGGTAGCGGTGGTTTTT3'(pBluescript II SK(+)特异性探针FAM-MGB标记),5'CCGTCAATATTTACCTTCCC3'(M13K07特异性探针VIC-MGB标记)扩增区室化的噬菌体,将扩增

记录在两个不同的通道并且根据制造商的方案进行微滴产生,PCR和检测。

[0181] 检测到有效的噬菌体裂解和单噬菌体检测的计数泊松分布,指示单噬菌体检测灵敏度。

[0182] 实施例2

[0183] 二聚化PCR

[0184] pBluescript II SK (+) 噬菌粒载体 (Agilent, 212205) f1起源于 (+) 取向, 宿主菌株中的 Sac<sup>+</sup>→Kpn 多接头取向: XL1-Blue MRF' 已用于产生 pBluescript II SK (+) 噬菌体且购买 M13K07 辅助噬菌体 (M13K07 Helper Phage, NEB, N0315S)。根据制造商的说明书, 使用 EZ-Link Sulfo-NHS-Biotin (Thermo, 21326) 用 20 摩尔过量的生物素标记噬菌体, 并使用 (Dong D, Sutaria S, Hwangbo JY, Chen P. A simple and rapid method to isolate purer M13 phage by isoelectric precipitation. Appl Microbiol Biotechnol. 2013 Sep; 97 (18): 8023-9.) 所述的等电点沉淀。将等量的纯化噬菌体混合, 从而提供浓度为 10e+6/ml 的噬菌体, 并与抗生物素蛋白 (A9275-1MG, Sigma-Aldrich) 以摩尔当量组合或用作混合物, 并进行数字 PCR (QX200 Droplet Digital PCR (ddPCR<sup>TM</sup>) System)。简而言之, 使用 5' TAACGTGGGAATGGTGCTTCCTCAAGTCGGTGACGGTGAT3' (M13K07 特异性正向), 5' GACAAAAGGGCGACATTCAA3' (M13K07 特异性反向) 和 5' GAAGCACCATTCCCACGTTATCTTGATCCGGC AAACAAAC3' (pBluescript II SK (+) 特异性正向), 5' TTTTCTGCGCGTAATCTGCT3' (pBluescript II SK (+) 特异性反向) 以及探针 5' CTGGTAGCGGTGGTTTTT3' (pBluescript II SK (+) 特异性探针 FAM-MGB 标记), 5' CCGTCAATATTTACCTTCCC3' (M13K07 特异性探针 VIC-MGB 标记) 扩增区室化的噬菌体, 扩增记录在两种不同的通道并且根据制造商的方案进行微滴产生, PCR 和检测。在存在抗生物素蛋白情况下, 由于生物素-抗生物素蛋白结合, 预期“连锁”计数增加。为了检测二聚化 PCR 产物, 根据制造商推荐提取扩增的 DNA, 并使用如下引物在实时 PCR 仪器中进行 PCR: 5' TTTTCTGCGCGTAATCTGCT3' (pBluescript II SK (+) 特异性反向), 5' GACAAAAGGGCGACATTCAA3' (M13K07 特异性反向) 和探针 5' CTGGTAGCGGTGGTTTTT3' (pBluescript II SK (+) 特异性探针 FAM-MGB 标记), 5' CCGTCAATATTTACCTTCCC3' (M13K07 特异性探针 VIC-MGB 标记)。只有正确的二聚化产物是可扩增的, 并且由两个探针产生预期的放大信号。

[0185] 实施例3

[0186] 基于区室的蛋白质/蛋白质复合体的识别

[0187] (a) 通过标准方法 (包括任何分离方法) 实现蛋白质复合体的提取, 这提供甚至部分保留的蛋白质复合体和部分纯化蛋白质复合体的方法 (例如, 根据其翻译后修饰或其它方法) 提供蛋白质复合体的提取来实现。

[0188] (b) 组合抗体文库噬菌体与蛋白质复合体, 并进行稀释和区室化, 以使未绑定的噬菌体实现的单噬菌体水平的分离。区室化 (例如大量反应的有效分离) 是在基于泊松分布的单个未绑定噬菌体的分离的基础上; 例如, 乳液和微阵列是最知名的现有技术的方法。单一的要求是噬菌体和蛋白质复合体的最小非特异性共定位。使噬菌体有效热裂解, 随后进行连接 PCR 过程, 以连接核苷酸序列。

[0189] (c) (使用能进行二聚化的通用引物) 扩增并连接编码 (基于噬菌体文库中的抗体基因的任何 CDR 区域, 优选基于 CDR3) 每区室中的噬菌体 (绑定的和未绑定的) 特异性 DNA 片

段的抗体(二聚化PCR的例子见实施例2和4)

[0190] (d)生成的二聚化的,甚至未纯化的PCR扩增子优选组合使用从乳液中提取DNA或物理消除微阵列腔或其它装置中的反应。

[0191] (e)优选通过下一代DNA测序以高度平行和定量的方式显示多个PCR二聚体扩增子的连锁信息。

[0192] (f)单个噬菌体的预定结合特征信息(见实施例5和7)和绑定的噬菌体的连锁信息用于在统计学基础上计算相互作用组,包括在扣除背景的连锁信息的基础上确定显著的相互作用组(去除混杂或随机或偶然的相互作用);在不同噬菌体的冗余连锁信息的基础上确认并过滤相互作用,并且可能通过其已知解离常数信息对相同结合特征信息进行加权;在确认和过滤,可能加权的连锁信息的基础上确定二聚体和多聚体相互作用;在定量连锁信息的基础上测定蛋白质和蛋白质复合体的相对丰度;在具有相同结合特征信息的若干噬菌体检测的冗余测量的基础上,确认,计算二聚体和多聚体相互作用的统计学误差以及蛋白质和蛋白质复合体的相对丰度。

[0193] 实施例4

[0194] 基于区室的对照蛋白/蛋白质复合体的识别

[0195] 展示抗体的M13噬菌体购自Source BioScience(6001\_hDAb),包括噬菌粒抗体文库( $\sim 3 \times 10^9$ )(Dudgeon K,Famm K,Christ D.Sequence determinants of protein aggregation in human VH domains.Protein Eng Des Sel.2008 Oct 28)KM13辅助噬菌体,TG1Tr细菌菌株和抗 $\beta$ -半乳糖苷酶和展示抗牛泛素抗体的噬菌体。通过NGS(ThermoFisher Scientific PGM,Ion Xpress<sup>TM</sup> Plus Fragment Library Kit,4471269)对对照噬菌体进行测序,并通过ELISA针对相应的抗原来验证其结合亲和力。该抗原为来自LifeSensors(#S1280)的、 $\beta$ -半乳糖苷酶生物素标记(G5025)(b-BGAL)的单生物素化泛素(b-UBI)。蛋清抗生物素蛋白(A9275)也来自Sigma-Aldrich。在测序结果的基础上,设计通用引物:通用正向引物-CCAAGAACACGCTGTATCTGCA;能够进行二聚化的反向通用引物-TGCGCATCCATTGTAGAGGTGAGACGGTGACCAGGGTTCC和ACCTCTACAATGGATGCGCAGAGACGGTGACCAGGGTTCC。为了检测二聚体产物,设计二聚体特异性实时PCR反应:正向-AGTTGGAGTCTTGGGGTCAGG,反向AGGTGGGTCGATGTTTGACTACTG和探针-FAM TCTCACCTCTACAATGGAT MGB。

[0196] 还设计了对照噬菌体特异性探针:抗 $\beta$ -半乳糖苷酶-FAM GCTAGGGCTATGTATCC MGB;

[0197] 抗牛泛素-VIC TGGGTCGATGTTTGACTAC MGB。

[0198] 根据说明书,扩增对照噬菌体(抗 $\beta$ -半乳糖苷酶 $3.8 \times 10^{12}/\text{ml} = 6.48\text{nM}$ ,抗牛泛素 $4.0 \times 10^{12}/\text{ml} = 6.7\text{nM}$ )。合并抗生物素蛋白(36nM)(或省去),b-UBI(72nM)和b-BGAL(单体为72nM),并在室温下孵育1小时形成复合体。将10倍稀释的复合体合并,并与1.5nM对照噬菌体孵育过夜。将绑定噬菌体的复合体稀释 $2 \times 10^6$ 次,且根据(QX200Droplet Digital PCR(ddPCR<sup>TM</sup>)System)的方案生成乳液微滴,并使用PCR条件扩增:对于探针,使用ddPCRSupermix(无dUTP)(186-3023),正向通用引物浓度为800nM,能进行二聚化的反向通用引物浓度为50nM。在一些情况下,加入250nM对照噬菌体特异性探针。扩增的微滴根据制造商方案用氯仿提取以回收扩增的二聚体产物。

[0199] 通过二聚体特异性实时PCR成功检测二聚体,证明其正确的二聚化结构。如果在乳



液PCR反应中包括特异性对照探针,在存在抗生物素蛋白的情况下,由于对照噬菌体和抗生物素蛋白/抗原复合体结合,检测到“连锁”增加,表明抗B-半乳糖苷酶和抗牛泛素噬菌体的检测以高于仅为偶然的速率在相同的微滴中定位。

[0200] 实施例5

[0201] 预先确定抗体噬菌体文库的结合特征信息

[0202] (a) 合并抗体(要确定结合特征信息的噬菌体文库)和cDNA文库噬菌体(由将要确定抗体噬菌体文库的结合特征信息的那些cDNA构成),并将抗体cDNA噬菌体复合体稀释并区室化,以实现未绑定的噬菌体的单噬菌体分离水平-进一步的细节参见实施例3部分(b),

[0203] (b) (使用能进行二聚化的通用引物)扩增并连接编码每区室中的抗体(基于噬菌体文库中的抗体基因的任何CDR区域,优选基于CDR3)的噬菌体DNA片段和cDNA的噬菌体DNA片段(绑定的和未绑定的)(二聚化PCR的例子见实施例2和4)。

[0204] (c) 生成的二聚化的PCR扩增子优选组合使用从乳液中提取DNA或物理消除微阵列腔的反应。

[0205] (d) 优选通过下一代DNA测序以高度平行和定量的方式显示多个PCR二聚体扩增子的连锁信息。

[0206] (e) 使用绑定的噬菌体的连锁信息计算抗体噬菌体文库对cDNA噬菌体文库的结合特征信息,包括在扣除背景的连锁信息(去除混杂或随机或偶然的相互作用)的基础上确定显著的相互作用;基于在统计学上显著的相互作用识别重要的抗体-cDNA结合;确定每个检测到的抗体噬菌体(包括检测到的cDNA片段、推断的检测到的蛋白质)的结合特征信息;在具有相同结合特征信息的若干噬菌体检测的冗余测量的基础上,确认、计算结合特征信息的统计误差。

[0207] 实施例6

[0208] 抗体噬菌体文库的富集

[0209] (a) 通过可分离的装置(用于分离未绑定的抗体噬菌体),优选在微珠上固定cDNA文库噬菌体(由需要通过富集的抗体噬菌体文库检测的那些cDNAs构成)

[0210] (b) 将抗体文库噬菌体(将富集的噬菌体文库)与固定化的cDNA噬菌体合并以实现绑定和未绑定噬菌体的分离

[0211] (c) 优选通过洗涤去除未绑定的抗体噬菌体

[0212] (d) 洗脱绑定的抗体噬菌体并任选地通过合适的装置扩增以得到高滴度制剂

[0213] (e) 任选地,对富集的抗体噬菌体的高滴度制剂进行下一轮富集

[0214] (f) 任选地,使用实施例5中描述的方法,针对cDNA噬菌体文库验证洗脱的绑定的噬菌体。

[0215] 实施例7

[0216] 预先确定富集的抗体噬菌体文库的结合特征信息

[0217] 展示抗体的M13噬菌体购自Source BioScience(6001\_hDAb),包括噬菌粒抗体文库( $\sim 3 \times 10^9$ )(Dudgeon K,Famm K,Christ D.Sequence determinants of protein aggregation in human VH domains.Protein Eng Des Sel.2008 Oct 28.)KM13辅助噬菌体和TG1Tr细菌菌株。扩增该文库,用KM13辅助噬菌体感染,根据方案(Lee CM,Iorno N,Sierro F,Christ D.Selection of human antibody fragments by phage display.Nat

Protoc.2007;2(11):3001-8.)收集噬菌体。PhD12噬菌体展示肽文库(PhD12 Phage Display Peptide Library,E8110S)和大肠杆菌ER2738宿主(E.coli ER2738)菌株购自New England Biolabs。在LB/IPTG/Xgal平板上培养PhD12文库,挑取50个噬菌斑并合并(抗原诱饵文库)。抗原诱饵文库被吸收在微量滴定板上,并根据Source BioScience(6001\_hDAb)文库方案,使用全部Source BioScience(6001\_hDAb)文库进行淘选。将总共612个克隆放在LB/氨苄青霉素平板上,扩增,用KM13辅助噬菌体感染并收集(富集的抗体文库)。

[0218] 在噬菌体序列的基础上,设计通用引物:Source BioScience(6001\_hDAb)文库特异性正向通用引物-CCAAGAACACGCTGTATCTGCA;能进行二聚化的Source BioScience(6001\_hDAb)文库特异性反向通用引物-TGCGCATCCATTGTAGAGGTGAGACGGTGACCAGGGTTCC-和PhD12噬菌体展示肽文库特异性正向通用引物-CGCAATTCCTTTAGTGGTACCTTT;能进行二聚化的PhD12噬菌体展示肽文库特异性反向通用引物-ACCTCTACAATGGATGCGCATCTGTATGGGATTTTGCTAAACAACCT。

[0219] 为了检测二聚化产物,设计了二聚体特异性实时PCR反应:正向- $\text{CGGACTGTTGAAAGTTGTTTAGCA}$ ,反向 $\text{GGTCACCGTCTCACCTCTAC}$ 和探针-VIC-CATACAGATGCGCATCC-MGB。

[0220] 将 $10^{12}$ 抗原诱饵文库和富集的抗体文库噬菌体合并,并在室温下孵育过夜。将噬菌体复合体稀释 $2 \times 10^6$ 倍,根据(QX200 Droplet Digital PCR(ddPCR™)System)方案生成乳液微滴,并使用PCR条件扩增:用于探针的ddPCR Supermix(无dUTP)(186-3023),正向通用引物浓度为800nM,能进行二聚化的反向通用引物浓度为50nM。扩增的微滴根据制造商方案用氯仿提取以回收扩增的二聚体产物。扩增的二聚体产物用NGS测序,并且检测到特异性诱饵文库和富集的抗体文库二聚化产物,表明抗原诱饵文库的成员和对冲抗体文库之间基于特异性序列的相互作用。

[0221] 实施例8

[0222] 确定抗体噬菌体文库的成员的定量结合信息

[0223] (a)应用实施例3中的方法,并修改步骤b,在平衡条件下使用几种量化的蛋白质复合体,产生几次平行确定

[0224] (b)在平行确定中获得的定量信息的基础上,可以构建多种蛋白质-噬菌体相互作用和解离常数的定量结合曲线,并可计算结合能力信息。

[0225] 实施例9

[0226] 发明的化学计量法

[0227] 对于在大肠杆菌的细胞体积中结合化学计量1nM的蛋白质约为1个分子/细胞和2,000个分子/哺乳动物(HeLa)细胞,并且信号蛋白的特征浓度(在此作为例子)为10nM-1 $\mu$ M。此外,由于噬菌体展示抗体的解离常数(Kd)在10nM和低至0.1nM的范围内,解离率为 $10^{(-3)}$ 至 $10 \cdot 10^{(-4)} \text{ s}^{-1}$ ,并且由于可以常规选择这些噬菌体,对于大多数蛋白质/表位,预期结合化学计量是饱和的,并且解离率提供足够的时间来使复合体区室化而不会提早解离。

[0228] 在细菌、酵母和哺乳动物细胞中每立方微米(即1fL)存在 $2-4 \times 10^6$ 个蛋白质(Bioessays.2013 Dec;35(12):1050-5.),对于5000个真核细胞的体积(10000fL),存在 $10^{10}$ 个蛋白,噬菌体最大浓度约为 $10^{16}/\text{ml}$ →在10000fL体积中: $10^{11}$ 噬菌体。

[0229] 在10(+4-5)范围内,所需的相互作用/噬菌体多重性的相互作用组复杂性约为10,

其对应于 $10^5$ 个个体噬菌体,即,(在 $10^{16}$ /ml中)每个噬菌体为0.1nM,并且每10000fL超过 $10^6$ 个(总蛋白质的0.01%)的每种蛋白质的浓度大于1nM,且Kd平均为0.1-1nM的抗体噬菌体(HuCAL GOLD亚纳摩尔概率:30%)(J.Mol.Biol.(2008) 376,1182-1200)可提供50-5%的饱和度。这对应于25-0.25%的共定位饱和度(共定位饱和度是指具有有区别的特异性的两种绑定的噬菌体位于相同的区室中)。

[0230] 在HiSEQ 2500 NGS仪器共定位率为0.25%的情况下,3亿次读数(10-300Gb,250bp读数)对应于异二聚体PCR产物的最小数量为 $7 \times 10^5$ ,这意味着最小的二元相互作用测序覆盖率:375。

[0231] 完全随机的、原始的、高覆盖度的抗体噬菌体文库复杂度高达 $10^{(+13)}$ 个个体噬菌体克隆,可控的乳液PCR或微阵列区室的数量范围为 $10^5$ - $10^8$ (NGS芯片区室的当前数量)这些不成比例的数字需要通过降低抗体噬菌体文库的复杂性来融合并增加针对靶向的蛋白(靶向部分或全部相互作用组)具有结合能力的噬菌体的丰度。为了减少复杂性,设计特异性选择过程的-使用文库(抗体)选择文库(cDNA)方法从完全随机化的文库选择噬菌体导致低复杂性、亲和力富集的、本源的通用噬菌体文库,此外,可以通过在选择过程中检测抗体-蛋白质的结合来监测该过程,或甚至可以使用例如不同量的展示输入的蛋白质的噬菌体来提取结合动力学信息。

[0232] 也可以通过使用自下而上(混合具有已知结合特征的噬菌体并添加背景噬菌体)逐渐构建越来越多复合体的文库来生成文库;这些是针对特定任务定制的特定库,以及自上而下的方法(例如,通过降低选择相互作用的噬菌体的复杂性)

[0233] 实施例10

[0234] 所获信息的统计评估

[0235] (a) 通过标准方法实现蛋白质复合体的提取,并结合抗体文库噬菌体,并将形成的复合体稀释并区室化,以实现未绑定的噬菌体的单噬菌体分离水平。

[0236] (b) 编码每个区室中的噬菌体(绑定的和未绑定的)的特异性DNA片段的抗体通过PCR扩增并连接在一起(优选通过有限数量的扩增循环)。

[0237] (c) 通过下一代DNA测序以高度平行和定量的方式显示该连锁信息。

[0238] (d) 使用个体噬菌体的预定结合特征信息和绑定的噬菌体的连锁信息来计算相互作用组。

[0239] 该方法在基于区室的蛋白质/蛋白质复合体的识别的基础上进行,其中每个区室的单个蛋白质复合体的蛋白质-抗体身份转化为DNA。使用所有连接的DNA片段的识别来定量确定相互作用,然而意外捕获在相同区室中的未绑定的噬菌体可以带来背景。这种背景可以通过简单的统计手段来处理,因为该背景为随机事件,可以根据具体事件进行区分。在区室化期间,结合剂的分布由泊松分布控制,从而计算每种结合剂的发生情况(通过在基于有限区室的PCR扩增后通过NGS来确定结合剂的相对丰度),如果区室数量是已知的,则可以计算蛋白质/蛋白质复合体的背景检测。使用多个特异性结合事件识别抗体的确切靶蛋白,如在多蛋白质复合体中,链接信息由共定位产生,并且表示直接结合。对于每个蛋白质/蛋白质复合体,计算背景检测,蛋白质/蛋白质复合体的任何检测变化是由于真实的结合效应,其可以通过除去计算出的蛋白质/蛋白质复合体的背景检测进行简单扣除(或泊松校正扣除,因为蛋白质/蛋白质复合体绑定结合剂的减法改变了结合剂的总数)来计算。如果使

用在平衡条件下不同组合的反应抗体和蛋白质分析物,可以构建Scatchard图或其它结合动力学计算,以计算抗体蛋白质相互作用的Kds或其它参数。也可以使用不同浓度的相互作用蛋白(改变实验条件或使用尖峰分析)来计算所有或多个相互作用的相互作用组的内部动力学数据。

- [0001] 序列表
- [0002] <110> 吉尼维拉基公司
- [0003] 黛博拉·哈特
- [0004] <120> 方法
- [0005] <130> P61471W0
- [0006] <150> GB 1420852.4
- [0007] <151> 2014-11-24
- [0008] <160> 31
- [0009] <170> PatentIn version 3.5
- [0010] <210> 1
- [0011] <211> 20
- [0012] <212> DNA
- [0013] <213> Artificial Sequence
- [0014] <220>
- [0015] <223> M13K07 Specific forward
- [0016] <400> 1
- [0017] ctcaagtcgg tgacggtgat 20
- [0018] <210> 2
- [0019] <211> 20
- [0020] <212> DNA
- [0021] <213> Artificial Sequence
- [0022] <220>
- [0023] <223> M13K07 Specific Reverse
- [0024] <400> 2
- [0025] gacaaaaggg cgacattcaa 20
- [0026] <210> 3
- [0027] <211> 20
- [0028] <212> DNA
- [0029] <213> Artificial Sequence
- [0030] <220>
- [0031] <223> pBluescript II SK(+) specific forward
- [0032] <400> 3
- [0033] tcttgatccg gcaaacaac 20
- [0034] <210> 4
- [0035] <211> 20
- [0036] <212> DNA
- [0037] <213> Artificial Sequence
- [0038] <220>

[0039]	<223> pBluescript II SK(+) specific reverse
[0040]	<400> 4
[0041]	ttttctgcgc gtaatctgct 20
[0042]	<210> 5
[0043]	<211> 18
[0044]	<212> DNA
[0045]	<213> Artificial Sequence
[0046]	<220>
[0047]	<223> pBluescript II SK(+) specific probe FAM-MGB labeled
[0048]	<400> 5
[0049]	ctggtagcgg tggttttt 18
[0050]	<210> 6
[0051]	<211> 20
[0052]	<212> DNA
[0053]	<213> Artificial Sequence
[0054]	<220>
[0055]	<223> M13K07 specific probe VIC-MGB labeled
[0056]	<400> 6
[0057]	ccgtcaatat ttaccttccc 20
[0058]	<210> 7
[0059]	<211> 40
[0060]	<212> DNA
[0061]	<213> Artificial Sequence
[0062]	<220>
[0063]	<223> M13K07 specific forward Primer
[0064]	<400> 7
[0065]	taacgtggga atggtgcttc ctcaagtcgg tgacggtgat 40
[0066]	<210> 8
[0067]	<211> 20
[0068]	<212> DNA
[0069]	<213> Artificial Sequence
[0070]	<220>
[0071]	<223> M13K07 specific reverse primer
[0072]	<400> 8
[0073]	gacaaaaggc cgacattcaa 20
[0074]	<210> 9
[0075]	<211> 40
[0076]	<212> DNA
[0077]	<213> Artificial Sequence

[0078] <220>  
[0079] <223> pBluescript II SK(+) specific forward  
[0080] <400> 9  
[0081] gaagcaccat tcccacgtta tcttgatccg gcaaacaac 40  
[0082] <210> 10  
[0083] <211> 20  
[0084] <212> DNA  
[0085] <213> Artificial Sequence  
[0086] <220>  
[0087] <223> pBluescript II SK(+) specific reverse  
[0088] <400> 10  
[0089] ttttctgcgc gtaatctgct 20  
[0090] <210> 11  
[0091] <211> 18  
[0092] <212> DNA  
[0093] <213> Artificial Sequence  
[0094] <220>  
[0095] <223> pBluescript II SK(+) specific probe FAM-MGB labeled  
[0096] <400> 11  
[0097] ctggtagcgg tggttttt 18  
[0098] <210> 12  
[0099] <211> 20  
[0100] <212> DNA  
[0101] <213> Artificial Sequence  
[0102] <220>  
[0103] <223> M13K07 specific probe VIC-MGB labeled  
[0104] <400> 12  
[0105] ccgtcaatat ttaccttccc 20  
[0106] <210> 13  
[0107] <211> 20  
[0108] <212> DNA  
[0109] <213> Artificial Sequence  
[0110] <220>  
[0111] <223> pBluescript II SK(+) specific reverse  
[0112] <400> 13  
[0113] ttttctgcgc gtaatctgct 20  
[0114] <210> 14  
[0115] <211> 20  
[0116] <212> DNA

---

[0117]	<213> Artificial Sequence
[0118]	<220>
[0119]	<223> M13K07 specific reverse
[0120]	<400> 14
[0121]	gacaaaaggg cgacattcaa 20
[0122]	<210> 15
[0123]	<211> 18
[0124]	<212> DNA
[0125]	<213> Artificial Sequence
[0126]	<220>
[0127]	<223> pBluescript II SK(+) specific probe FAM-MGB labeled
[0128]	<400> 15
[0129]	ctggtagcgg tggttttt 18
[0130]	<210> 16
[0131]	<211> 20
[0132]	<212> DNA
[0133]	<213> Artificial Sequence
[0134]	<220>
[0135]	<223> M13K07 specific probe VIC-MGB labeled
[0136]	<400> 16
[0137]	ccgtcaatat ttaccttccc 20
[0138]	<210> 17
[0139]	<211> 22
[0140]	<212> DNA
[0141]	<213> Artificial Sequence
[0142]	<220>
[0143]	<223> General Forward Primer
[0144]	<400> 17
[0145]	ccaagaacac gctgtatctg ca 22
[0146]	<210> 18
[0147]	<211> 40
[0148]	<212> DNA
[0149]	<213> Artificial Sequence
[0150]	<220>
[0151]	<223> Dimerisation capable general reverse primer
[0152]	<400> 18
[0153]	tgcgcatcca ttgtagaggt gagacggtga ccagggttcc 40
[0154]	<210> 19
[0155]	<211> 40



[0156] <212> DNA  
[0157] <213> Artificial Sequence  
[0158] <220>  
[0159] <223> Dimerisation capable general reverse primer  
[0160] <400> 19  
[0161] acctctacaa tggatgcgca gagacggtga ccagggttcc 40  
[0162] <210> 20  
[0163] <211> 21  
[0164] <212> DNA  
[0165] <213> Artificial Sequence  
[0166] <220>  
[0167] <223> Dimer specific real-time forward primer  
[0168] <400> 20  
[0169] agttggagtc ttggggtcag g 21  
[0170] <210> 21  
[0171] <211> 24  
[0172] <212> DNA  
[0173] <213> Artificial Sequence  
[0174] <220>  
[0175] <223> Dimer specific real-time reverse primer  
[0176] <400> 21  
[0177] aggtgggtcg atgtttgact actg 24  
[0178] <210> 22  
[0179] <211> 19  
[0180] <212> DNA  
[0181] <213> Artificial Sequence  
[0182] <220>  
[0183] <223> Dimer specific real-time PCR Probe  
[0184] <400> 22  
[0185] tctcacctct acaatggat 19  
[0186] <210> 23  
[0187] <211> 17  
[0188] <212> DNA  
[0189] <213> Artificial Sequence  
[0190] <220>  
[0191] <223> Anti-beta-galactosidase probe  
[0192] <400> 23  
[0193] gctagggcta tgtatcc 17  
[0194] <210> 24

[0195]	<211> 19
[0196]	<212> DNA
[0197]	<213> Artificial Sequence
[0198]	<220>
[0199]	<223> Anti-bovine ubiquitin PProbe
[0200]	<400> 24
[0201]	tgggtcgatg tttgactac 19
[0202]	<210> 25
[0203]	<211> 22
[0204]	<212> DNA
[0205]	<213> Artificial Sequence
[0206]	<220>
[0207]	<223> General forward Source BioScience (6001_hDAb) library specific
[0208]	primer
[0209]	<400> 25
[0210]	ccaagaacac gctgtatctg ca 22
[0211]	<210> 26
[0212]	<211> 40
[0213]	<212> DNA
[0214]	<213> Artificial Sequence
[0215]	<220>
[0216]	<223> Dimerisation capable general reverse primer Source BioScience
[0217]	(6001_hDAb) library specific primer
[0218]	<400> 26
[0219]	tgcgcattcca ttgttagaggt gagacggtga ccagggttcc 40
[0220]	<210> 27
[0221]	<211> 24
[0222]	<212> DNA
[0223]	<213> Artificial Sequence
[0224]	<220>
[0225]	<223> General forward PhD12 Phage Display Peptide Library specific
[0226]	primer
[0227]	<400> 27
[0228]	cgcaattcct ttagtggtac cttt 24
[0229]	<210> 28
[0230]	<211> 46
[0231]	<212> DNA
[0232]	<213> Artificial Sequence
[0233]	<220>

[0234]	<223> Dimerisation capable general reverse primer PhD12 Phage Display
[0235]	Peptide Library specific
[0236]	<400> 28
[0237]	acctctacaa tggatgcgca tctgtatggg attttgctaa acaact 46
[0238]	<210> 29
[0239]	<211> 24
[0240]	<212> DNA
[0241]	<213> Artificial Sequence
[0242]	<220>
[0243]	<223> Dimer specific real-time forward primer
[0244]	<400> 29
[0245]	cggactgttg aaagttgttt agca 24
[0246]	<210> 30
[0247]	<211> 20
[0248]	<212> DNA
[0249]	<213> Artificial Sequence
[0250]	<220>
[0251]	<223> Dimer specific real-time reverse primer
[0252]	<400> 30
[0253]	ggtcaccgtc tcacctctac 20
[0254]	<210> 31
[0255]	<211> 17
[0256]	<212> DNA
[0257]	<213> Artificial Sequence
[0258]	<220>
[0259]	<223> Dimer specific real-time probe
[0260]	<400> 31
[0261]	catacagatg cgcatcc 17

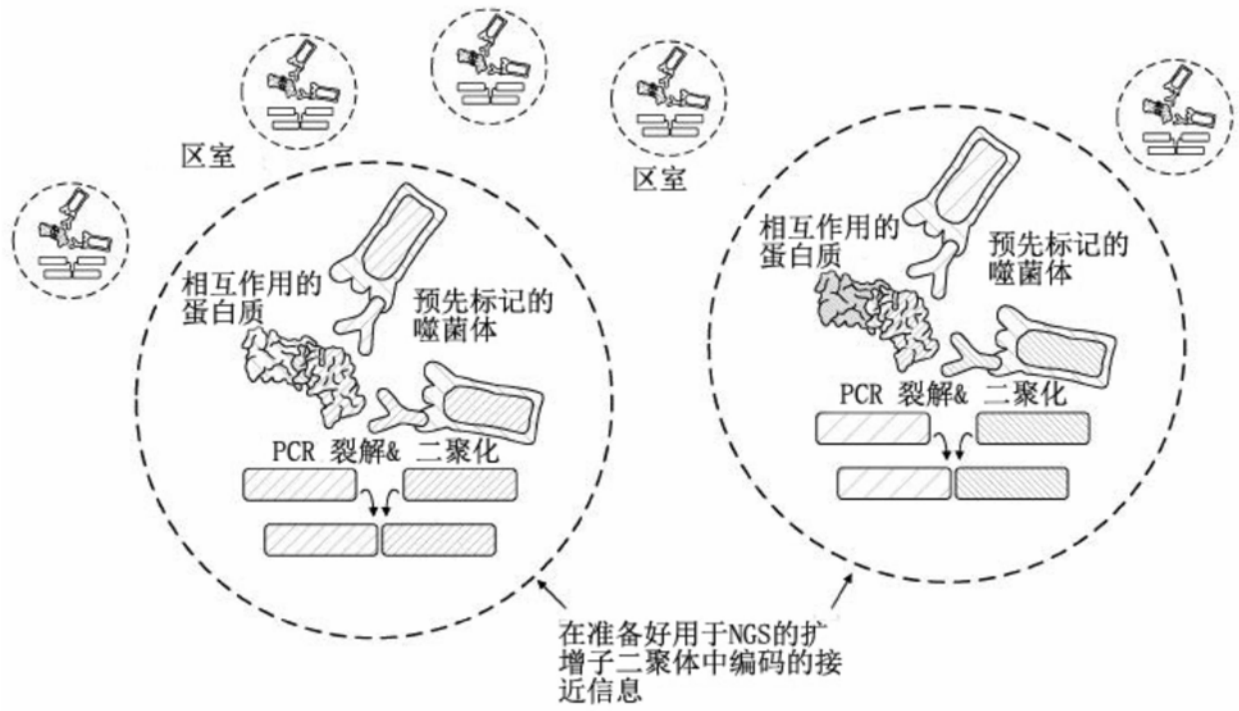


图1

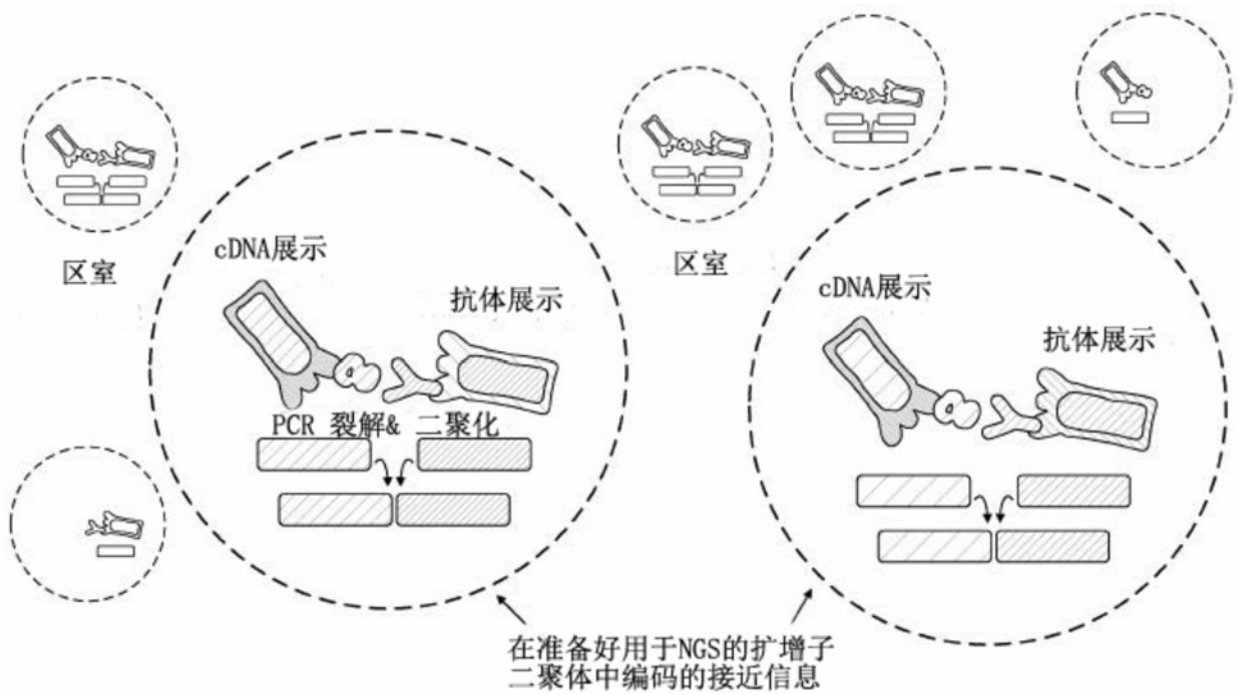


图2