



(10) 授权公告号 CN 110249082 B

(45) 授权公告日 2023. 07. 07

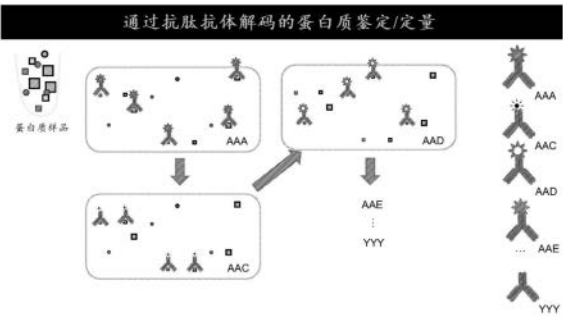
(21) 申请号 201780085386.2
(22) 申请日 2017.12.01
(65) 同一申请的已公布的文献号
申请公布号 CN 110249082 A
(43) 申请公布日 2019.09.17
(30) 优先权数据
62/429,063 2016.12.01 US
62/500,455 2017.05.02 US
(85) PCT国际申请进入国家阶段日
2019.08.01
(86) PCT国际申请的申请数据
PCT/US2017/064322 2017.12.01
(87) PCT国际申请的公布数据
W02018/102759 EN 2018.06.07
(73) 专利权人 诺迪勒思附属公司
地址 美国华盛顿州

(72) 发明人 帕拉格·马利克
(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262
专利代理师 王玮玮 郑霞
(51) Int.Cl.
C40B 30/04 (2006.01)
G01N 33/48 (2006.01)
G01N 33/53 (2006.01)
G01N 33/68 (2006.01)
(56) 对比文件
US 2007218503 A1,2007.09.20
US 2003054408 A1,2003.03.20
US 2003040020 A1,2003.02.27
US 2004067599 A1,2004.04.08
CN 1938430 A,2007.03.28
审查员 程鑫

权利要求书2页 说明书28页
序列表13页 附图19页

(54) 发明名称
测定蛋白质的方法

(57) 摘要
本文提供了用于鉴定样品中的蛋白质的方法和系统。获取抗体小组,其中没有一种抗体对单个蛋白质或蛋白质家族具有特异性。另外,确定该小组中的抗体的结合性质。此外,将蛋白质反复暴露于抗体小组。另外,确定结合该蛋白质的抗体组。基于抗体的已知结合性质,使用一种或多种解卷积方法将该抗体组与蛋白质序列匹配,从而确定蛋白质的身份。



1. 一种确定多种蛋白质的特征的方法,所述方法包括:

(a) 将一种或多种流体施加至基底,其中所述多种蛋白质的至少部分缀合至所述基底,使得结合的所述多种蛋白质的所述至少部分的每一个具有独特的、光学可分辨的空间地址,其中所述一种或多种流体包含一系列多组n种亲和试剂并且被配置成使得所述多组n种亲和试剂中的一种或多种亲和试剂结合所述多种蛋白质的所述至少部分的一个或多个表位,其中所述多组n种亲和试剂中的每种亲和试剂包含已知的非特异性程度,并且所述多组n种亲和试剂中的一种或多种亲和试剂与可识别标签连接;

(b) 鉴定所述基底上具有一个或多个信号的一个或多个独特空间地址,其中所述一个或多个信号源自与结合至所述多种蛋白质的所述至少部分的一个或多个表位的亲和试剂连接的所述可识别标签;和

(c) 确定具有鉴定的独特空间地址的所述多种蛋白质的所述至少部分的特征,其中具有鉴定的独特空间地址的所述多种蛋白质的所述至少部分包含与亲和试剂结合的一个或多个表位,该亲和试剂与具有一个或多个信号的可识别标签连接。

2. 根据权利要求1所述的方法,其中所述多组n种亲和试剂中的每种亲和试剂对单独的蛋白质或蛋白质家族具有特异性。

3. 根据权利要求1所述的方法,其中所述多组n种亲和试剂中的每种亲和试剂识别存在于超过一种所述蛋白质中的一个或多个表位的家族。

4. 根据权利要求1所述的方法,其中所述多种蛋白质的所述至少部分的表位是构象的或线性的。

5. 根据权利要求1所述的方法,其中所述多组n种亲和试剂包含对相应表位具有特异性的连续或非连续氨基酸。

6. 根据权利要求5所述的方法,其进一步包括:

基于所述至少部分内的所确定的一个或多个表位,确定对应于阈值准确度的所述多种蛋白质的所述至少部分的身份。

7. 根据权利要求1-6中任一项所述的方法,其中所述多组n种亲和试剂包含超过100种亲和试剂。

8. 根据权利要求1-6中任一项所述的方法,其中所述方法进一步包括使用结合单个蛋白质或单个蛋白质同种型的至少一种亲和试剂。

9. 根据权利要求1-6中任一项所述的方法,其进一步包括:

基于所述亲和试剂的结合模式,以阈值准确度确定所述多种蛋白质的所述至少部分的身份。

10. 根据权利要求1-6中任一项所述的方法,其中所述基底是流动池。

11. 根据权利要求1-6中任一项所述的方法,其中使用可光活化的连接体将所述多种蛋白质的所述至少部分缀合至所述基底。

12. 根据权利要求1-6中任一项所述的方法,其中使用可光切割的连接体将所述多种蛋白质的所述至少部分缀合至所述基底。

13. 根据权利要求1-6中任一项所述的方法,其中所述多组n种亲和试剂的至少一种亲和试剂被修饰成与可识别标签缀合。

14. 根据权利要求1-6中任一项所述的方法,其中可识别标签是荧光标签。

15. 根据权利要求1-6中任一项所述的方法,其中可识别标签是核酸条形码。
16. 根据权利要求1-6中任一项所述的方法,其中对由蛋白质的被鉴定的部分占据的空间地址的数目进行计数以定量样品中该蛋白质的水平。
17. 根据权利要求1-6中任一项所述的方法,其中使用解卷积软件确定所述多种蛋白质的所述至少部分的身份。
18. 根据权利要求1-6中任一项所述的方法,其中通过将独特空间地址相关的表位的组合解码来确定所述多种蛋白质的所述至少部分的身份。
19. 根据权利要求1-6中任一项所述的方法,其进一步包括:在将所述多种蛋白质的所述至少部分缀合至所述基底之前,使所述多种蛋白质变性。
20. 根据权利要求1-6中任一项所述的方法,其中连接至基底的所述多种蛋白质的所述至少部分存在于多种蛋白质的混合物中。
21. 根据权利要求1-6中任一项所述的方法,其中所述方法用于鉴定多种蛋白质。
22. 根据权利要求1-6中任一项所述的方法,其中将每种单独的亲和试剂的已知结合非特异性程度和所述亲和试剂在所述独特空间地址处的结合模式输入贝叶斯推理算法以鉴定单独的蛋白质。

测定蛋白质的方法

[0001] 交叉引用

[0002] 本申请要求于2016年12月1日提交的美国临时申请号62/429,063和于2017年5月2日提交的美国临时申请号62/500,455的权益,以上申请通过引用并入本文。

[0003] 发明背景

[0004] 用于蛋白质鉴定的当前技术通常依赖于高度特异和灵敏的抗体的结合和随后读出,或者依赖于来自质谱仪的肽读取数据(通常大约12-30AA长)。

发明内容

[0005] 本公开内容提供了用于测定蛋白质的方法和系统。在一些实施方案中,本公开内容提供了这样的方法,其中混合物中蛋白质的身份,即其序列,从可能高度不完整和/或对具体蛋白质无特异性的一系列测量中推断。本文所述的方法和系统还可用于表征和/或鉴定生物聚合物,包括蛋白质。另外,本文所述的方法和系统可用于比依赖于来自质谱仪的数据的蛋白质鉴定技术更快地鉴定蛋白质。在一些实例中,本文所述的方法和系统可用于以至少50%的准确度鉴定至少400种不同的蛋白质,且比依赖于来自质谱仪的数据的蛋白质鉴定技术快至少10%。在一些实例中,本文所述的方法和系统可用于以至少50%的准确度鉴定至少1000种不同的蛋白质,且比依赖于来自质谱仪的数据的蛋白质鉴定技术快至少10%。

[0006] 本发明的一个方面提供了一种确定蛋白质特征的方法。所述方法包括获得基底,其中一种或多种蛋白质的部分缀合至所述基底,使得每个单独的蛋白质部分具有独特的、可分辨的空间地址。在一些情况下,每个单独的蛋白质部分可具有独特的、光学可分辨的空间地址。所述方法进一步包括将含有第一至第n组的一种或多种亲和试剂的流体施加至所述基底。在一些实施方案中,所述亲和试剂可含有可识别标签或与之偶联。在将所述第一至第n组的一种或多种亲和试剂施加至所述基底后,所述方法包括执行以下步骤:观察所述亲和试剂或可识别标签;鉴定所述基底上具有一个或多个观察到的信号的一个或多个独特空间地址;以及确定具有鉴定的独特空间地址的所述一种或多种蛋白质的每个部分含有与所述一个或多个观察到的信号相关的一个或多个表位。在一些情况下,所述一种或多种蛋白质的每个缀合部分与所述基底上的独特空间地址相关联。在一些情况下,所述第一至第n组的一种或多种亲和试剂中的每种亲和试剂对单独的蛋白质或蛋白质家族不具有特异性。在一些情况下,所述亲和试剂的结合表位是未知的,或对单独的蛋白质或蛋白质家族不具有特异性。

[0007] 在一些情况下,本公开内容的方法还可与具有结合在单个位置的多种蛋白质的基底一起使用,其中在单个位置处的所述蛋白质中至少约50%、60%、70%、80%、90%或多于90%包含共同的氨基酸序列。在一些情况下,本公开内容的方法还可与具有结合在单个位置的多种蛋白质的基底一起使用,其中在单个位置处的所述蛋白质中至少50%、60%、70%、80%、90%或多于90%包含至少95%的氨基酸序列同一性。

[0008] 在一些实施方案中,所述一种或多种蛋白质可包含一种单蛋白质分子。在一些实

施方案中,所述一种或多种蛋白质可包含大批蛋白质。在一些实施方案中,所述一种或多种蛋白质可包含缀合在所述基底上的同一独特空间地址的多个相同蛋白质。

[0009] 在一些实施方案中,所述第一至第n组的一种或多种亲和试剂中的每种亲和试剂识别存在于超过一种蛋白质中的一个或多个表位的家族。在一些实施方案中,所述方法进一步包括基于所述部分内的所确定的一个或多个表位,以阈值准确度确定所述一种或多种蛋白质的所述部分的身份。在一些情况下,所述第一至第n组的一种或多种亲和试剂包含超过100种亲和试剂。在一些实施方案中,所述方法进一步包括使用结合单个蛋白质或单个蛋白质同种型的亲和试剂。

[0010] 在一些实施方案中,所述方法进一步包括基于所述亲和试剂的结合模式,以阈值准确度确定所述一种或多种蛋白质的所述部分的身份。在一些情况下,所述基底是流动池。在一些情况下,使用可光活化的连接体将所述一种或多种蛋白质的所述部分缀合至所述基底。在一些情况下,使用可光切割的连接体将所述一种或多种蛋白质的所述部分缀合至所述基底。

[0011] 在一些情况下,所述至少一组亲和试剂的至少一部分被修饰成与可识别标签缀合。在一些情况下,所述可识别标签是荧光标签。在一些情况下,所述可识别标签是磁性标签。在一些情况下,所述可识别标签是核酸条形码。在一些情况下,可识别标签是亲和标签(例如,Biotin、Flag、myc)。在一些情况下,对由蛋白质的鉴定部分占据的空间地址的数目进行计数以定量样品中该蛋白质的水平。在一些情况下,使用解卷积软件确定所述一种或多种蛋白质的所述部分的身份。在一些情况下,通过将具有独特空间地址相关的表位的组合解码来确定所述一种或多种蛋白质的所述部分的身份。在一些情况下,所述方法进一步包括在将所述一种或多种蛋白质的所述部分缀合至所述基底之前,使所述一种或多种蛋白质变性。在一些情况下,所述缀合至基底的一种或多种蛋白质的部分存在于多种蛋白质的复杂混合物中。在一些情况下,所述方法用于鉴定多种蛋白质。

[0012] 本发明的另一方面提供了一种鉴定蛋白质的方法,其包括:获取亲和试剂小组,其中没有一种亲和试剂对单个蛋白质或蛋白质家族具有特异性,确定所述小组中的所述抗体的结合性质,将所述蛋白质反复暴露于所述抗体小组,确定结合所述蛋白质的抗体组,以及基于所述抗体的已知结合性质,使用一种或多种解卷积方法将所述抗体组与蛋白质序列匹配,从而确定所述蛋白质的身份。在一些情况下,待鉴定的蛋白质在含有多种不同蛋白质的样品中得到鉴定。在一些情况下,所述方法能够同时鉴定单个样品中的多种蛋白质。

[0013] 本发明的另一方面提供了一种鉴定蛋白质的方法。所述方法包括获取抗体小组,其中没有一种抗体对单个蛋白质或蛋白质家族具有特异性,确定所述小组中的所述抗体的结合性质,将所述蛋白质反复暴露于所述抗体小组,确定不结合所述蛋白质的所述抗体的组,以及基于所述抗体的已知结合性质,使用一种或多种解卷积方法将所述抗体组与蛋白质序列匹配,从而确定所述蛋白质的身份。

[0014] 本发明的另一方面提供了一种使用m种亲和试剂独特地鉴定和定量蛋白质混合物中的n种蛋白质的方法,其中n大于m,并且n和m是大于1的正整数,并且其中所述蛋白质未通过固有性质分开。在一些情况下,n大约是m的5、10、20、50、100、500、1,000、5,000或10,000倍。

[0015] 本发明的另一方面提供了一种使用m种结合试剂独特地鉴定和定量蛋白质混合物

中的 n 种蛋白质的方法,其中 n 大于 m ,并且其中所述蛋白质是随机排列的。在一些情况下,所述蛋白质尚未通过基于大小或基于电荷的分离方法分开。

[0016] 本发明的另一方面提供了一种使用 m 种亲和试剂独特地鉴定和定量蛋白质分子混合物中的 n 种单蛋白质分子的方法。所述方法进一步包括, n 大于 m ,以及所述单蛋白质分子缀合至基底并在空间上分开,使得每个单独的蛋白质分子具有独特的、光学可分辨的空间地址。

[0017] 本发明的另一方面提供了一种从 n 种可能蛋白质的池中以高于阈值量的确定性鉴定未知单蛋白质分子的方法。所述方法包括使用亲和试剂小组,其中所述小组中亲和试剂的种数为 m ,并且其中 m 小于 n 的十分之一。

[0018] 本发明的另一方面提供了一种选择 m 种亲和试剂的小组的方法,所述 m 种亲和试剂的小组能够鉴定选自 n 种可能蛋白质的池的未知蛋白质,其中 m 小于 $n-1$ 。

[0019] 本发明的另一方面提供了一种选择 m 种亲和试剂的小组的方法,所述 m 种亲和试剂的小组能够鉴定选自 n 种可能蛋白质的池的未知蛋白质,其中 m 小于 n 的十分之一。

[0020] 本发明的另一方面提供了一种选择少于4000种亲和试剂的小组的方法,使得所述少于4000种亲和试剂的小组能够独特地鉴定20,000种不同蛋白质中的每一种。

[0021] 本发明的另一方面提供了一种使用 m 种结合试剂独特地鉴定和定量蛋白质混合物中的 n 种蛋白质的方法,其中 m 小于 $n-1$,并且其中通过所述 m 种结合试剂的亚组的独特结合概况鉴定每种蛋白质。

[0022] 在一些情况下,所述方法能够从人蛋白质样品中鉴定人蛋白质组中超过20%的蛋白质,其中所述蛋白质在过程中基本上不被破坏。在一些情况下,所述方法能够鉴定具有可用蛋白质序列数据库的任何生物体(例如酵母、大肠杆菌(*E.coli*)、秀丽隐杆线虫(*C.elegans*))的蛋白质组中超过20%的蛋白质。在一些情况下,可通过基因组、外显子组和/或转录组测序生成蛋白质序列数据库。在一些情况下,所述方法不需要超过4000种亲和试剂。在一些情况下,所述方法不需要超过100mg的蛋白质样品。

[0023] 本发明的另一方面提供了一种独特地鉴定单蛋白质分子的方法。所述方法包括获得亲和试剂小组,将所述单蛋白质分子暴露于所述小组中的每种亲和试剂,确定每种亲和试剂是否结合所述单蛋白质分子,以及使用所收集的结合数据来确定所述单蛋白质分子的身份。另外,在一些实施方案中,所述单蛋白质分子的身份不能通过所述亲和试剂小组中任何单独的亲和试剂的结合数据来确定。在一些情况下,可使用具有重叠结合特征的亲和试剂来富集对任何特定靶标的亲和力。

[0024] 本发明的另一方面提供了一种确定蛋白质特征的方法。所述方法包括将一种或多种蛋白质的部分缀合至基底,其中所述一种或多种蛋白质的每个缀合部分与所述基底上的独特空间地址相关联。在一些实例中,独特空间地址可以是与蛋白质的特定部分相关联的空间地址。所述方法还包括将所述第一至第 n 组的一种或多种亲和试剂施加至所述基底,其中所述第一至第 n 组的一种或多种亲和试剂中的每种亲和试剂识别长度在一至十个残基之间的表位,并且其中所述第一至第 n 组的一种或多种亲和试剂中的每种亲和试剂与可识别标签连接。另外,所述方法包括在将所述第一至第 n 组的一种或多种亲和试剂各自施加至所述基底后,执行以下步骤:观察所述可识别标签;鉴定所述基底上具有一个或多个观察到的信号的一个或多个独特空间地址;以及确定具有鉴定的独特空间地址的所述一种或多种蛋

白质的每个部分含有与所述一个或多个观察到的信号相关的一个或多个表位。

[0025] 本发明的另一方面提供了一种确定蛋白质特征的方法。所述方法包括将一种或多种蛋白质的部分缀合至基底,其中所述一种或多种蛋白质的每个缀合部分与所述基底上的独特空间地址相关联。所述方法还包括将第一至第n组的一种或多种亲和试剂施加至所述基底,其中所述第一至第n组的一种或多种亲和试剂中的每种亲和试剂识别存在于一种或多种蛋白质中的一个或多个表位的家族,并且其中所述第一至第n组的一种或多种亲和试剂中的每种亲和试剂与可识别标签连接。此外,所述方法包括在将所述第一至第n组的一种或多种亲和试剂各自施加至所述基底后,执行以下步骤:观察所述可识别标签;鉴定所述基底上具有观察到的信号的一个或多个独特空间地址;以及确定具有鉴定的独特空间地址的所述一种或多种蛋白质的每个部分含有表位。

[0026] 本发明的另一方面提供了一种鉴定蛋白质的方法,所述方法包括:获取具有已知非特异性程度的亲和试剂小组,确定所述小组中的所述亲和试剂的结合性质,将所述蛋白质反复暴露于所述亲和试剂小组,确定结合所述蛋白质的亲和试剂组,以及基于所述亲和试剂的已知结合性质,使用一种或多种解卷积方法将所述亲和试剂组与蛋白质序列匹配,从而确定所述蛋白质的身份。

[0027] 另外,本发明的另一方面提供了一种鉴定蛋白质的方法,所述方法包括获取具有已知非特异性程度的亲和试剂小组,确定所述小组中的所述亲和试剂的结合性质,将所述蛋白质反复暴露于所述亲和试剂小组,确定不结合所述蛋白质的亲和试剂组,以及基于所述亲和试剂的已知结合性质,使用一种或多种解卷积方法将所述亲和试剂组与蛋白质序列匹配,从而确定所述蛋白质的身份。

[0028] 援引并入

[0029] 本说明书中所提到的所有出版物、专利和专利申请均通过引用并入本文,其程度如同特别地且单独地指出每个单独的出版物、专利或专利申请通过引用而并入。

附图说明

[0030] 本发明的新颖特征在所附的权利要求书中具体阐述。通过参考以下对利用本发明原理的说明性实施方案加以阐述的详细描述和附图,将会获得对本发明的特征和优点的更好的理解,在这些附图中:

[0031] 图1示出了根据一些实施方案,通过抗肽抗体解码进行蛋白质定量的第一示意图;

[0032] 图2示出了根据一些实施方案,通过抗肽抗体解码进行蛋白质定量的第二示意图;

[0033] 图3示出了根据一些实施方案的流动池缀合;

[0034] 图4示出了根据一些实施方案,流动池上的独特空间地址的网格;

[0035] 图5示出了根据一些实施方案,将蛋白质解构为可与解码(d-code)抗体匹配的多组肽;

[0036] 图6示出了根据一些实施方案,通过抗肽抗体解码进行蛋白质鉴定/定量的示意图;

[0037] 图7示出了根据一些实施方案,第一组抗肽抗体的观察结果;

[0038] 图8示出了根据一些实施方案,第二组抗肽抗体的观察结果;

[0039] 图9示出了根据一些实施方案,第三组抗肽抗体的观察结果;

- [0040] 图10示出了根据一些实施方案,抗体测量数据的计算解码;
- [0041] 图11示出了根据一些实施方案的蛋白质组定量;
- [0042] 图12示出了根据一些实施方案,异常列表的实例;
- [0043] 图13示出了根据实施方案,定量可能需要的3-聚体解码抗体采样的覆盖度;
- [0044] 图14示出了计算机控制系统,其被编程或以其他方式配置用于实现本文提供的方法;以及
- [0045] 图15示出了根据本文中的实施方案,3-聚体解码探针的数目对蛋白质组的可识别性-覆盖度曲线的影响的实例。
- [0046] 图16A示出了根据本文中的实施方案,显示出缀合在基底上的单蛋白质分子的图像。
- [0047] 图16B示出了根据本文中的实施方案,显示出图16A的指示区域的放大部分的图像,其中缀合的蛋白质由箭头指示。
- [0048] 图17示出了根据本文中的实施方案的蛋白质鉴定。
- [0049] 图18示出了根据本文中的实施方案,鉴定蛋白质的示意图。

具体实施方式

[0050] 在一些实例中,方法可包括三个方面:1)可寻址的基底,蛋白质和/或蛋白质片段可以缀合至其中;2)一组亲和试剂,例如其中每种亲和试剂可以以不同的特异性与肽结合;以及3)软件,其能够使用关于亲和试剂的结合特征的先验知识、亲和试剂在基底中的每个地址处的特定结合模式,和/或混合物(例如人蛋白质组)中蛋白质的可能序列的数据库的组合来推断基底中精确空间地址处蛋白质的身份。在一些实例中,精确空间地址可以是独特空间地址。

[0051] 样品

[0052] 样品可以是含有蛋白质的任何生物样品。样品可取自组织或细胞,或者取自组织或细胞的环境。在一些实例中,样品可以是组织活检物、血液、血浆、细胞外液、培养的细胞、培养基、废弃组织、植物物质、合成蛋白质、古菌样品、细菌样品和/或病毒样品、真菌组织、古菌或原生动物。在一些实例中,蛋白质在样品制备期间从其主要来源(细胞、组织、体液如血液、环境样品等)中分离。蛋白质可以或可能不从其主要来源纯化。在一些情况下,在进一步加工之前将主要来源均质化。在一些情况下,使用缓冲液如RIPA缓冲液裂解细胞。在此阶段也可使用变性缓冲液。样品可被过滤或离心以去除脂质和颗粒物质。样品也可被纯化以去除核酸,或者可用RNA酶和DNA酶处理。样品可含有完整蛋白质、变性蛋白质、蛋白质片段或部分降解的蛋白质。

[0053] 样品可取自患有疾病或病症的受试者。疾病或病症可以是传染病、免疫病症或疾病、癌症、遗传病、退行性疾病、生活方式疾病、损伤、罕见疾病或年龄相关疾病。传染病可能由细菌、病毒、真菌和/或寄生虫引起。癌症的非限制性实例包括膀胱癌、肺癌、脑癌、黑色素瘤、乳腺癌、非霍奇金淋巴瘤、宫颈癌、卵巢癌、结直肠癌、胰腺癌、食道癌、前列腺癌、肾癌、皮肤癌、白血病、甲状腺癌、肝癌和子宫癌。遗传疾病或病症的一些实例包括但不限于囊性纤维化、夏科-马里-图思病(Charcot-Marie-Tooth disease)、亨廷顿病(Huntington's disease)、波伊茨-耶格综合征(Peutz-Jeghers syndrome)、唐氏综合症、类风湿性关节炎

和泰-萨病(Tay-Sachs disease)。生活方式疾病的非限制性实例包括肥胖症、糖尿病、动脉硬化、心脏病、中风、高血压、肝硬化、肾炎、癌症、慢性阻塞性肺病(copd)、听力问题和慢性背痛。损伤的一些实例包括但不限于擦伤、脑损伤、瘀伤、烧伤、脑震荡、充血性心力衰竭、建筑损伤、脱位、连枷胸、骨折、血胸、椎间盘突出、髌骨隆凸挫伤、低体温、撕裂、神经挟捏、气胸、肋骨骨折、坐骨神经痛、脊髓损伤、肌腱韧带筋膜损伤、创伤性脑损伤和鞭伤。可在治疗患有疾病或病症的受试者之前和/或之后取得样品。可在治疗之前和/或之后取得样品。可在治疗或治疗方案期间取得样品。可从受试者取得多个样品以监测治疗随时间的效果。可从已知或疑似患有传染病的受试者取得样品,针对该传染病没有诊断性抗体。

[0054] 样品可取自疑似患有疾病或病症的受试者。样品可取自经历不明原因的症状如疲劳、恶心、体重减轻、酸痛和疼痛、虚弱或记忆丧失的受试者。样品可取自具有明确原因的症状的受试者。样品可取自由于诸如家族史、年龄、环境暴露等因素、生活方式风险因素或存在其他已知风险因素而存在发展疾病或病症风险的受试者。

[0055] 样品可取自胚胎、胎儿或孕妇。在一些实例中,样品可由从母亲血浆中分离的蛋白质组成。在一些实例中,蛋白质从母亲血液中的循环胎儿细胞分离。

[0056] 可对蛋白质进行处理以去除可能干扰表位结合的修饰。例如,对蛋白质进行糖苷酶处理以去除翻译后糖基化。可用还原剂处理蛋白质以减少蛋白质内的二硫键结合。可用磷酸酶处理蛋白质以去除磷酸基团。可以去除的翻译后修饰的其他非限制性实例包括乙酸盐、酰胺基团、甲基基团、脂质、遍在蛋白、豆蔻酰化、棕榈酰化、异戊二烯化或异戊烯化(例如法尼醇和香叶基香叶醇)、法尼基化、香叶基香叶酰化、糖基磷脂酰肌醇化、脂化、黄素部分附接、磷酸泛酰巯基乙胺化和亚视黄基席夫碱形成。还可对样品进行处理以保留翻译后蛋白质修饰。在一些实例中,可将磷酸酶抑制剂添加到样品中。在一些实例中,可添加氧化剂以保护二硫键。

[0057] 接下来,可以使蛋白质完全或部分变性。在一些实施方案中,可以使蛋白质完全变性。可通过施加外部应激如洗涤剂、强酸或碱、浓无机盐、有机溶剂(例如,醇或氯仿)、辐射或加热而使蛋白质变性。可通过添加变性缓冲液而使蛋白质变性。也可将蛋白质沉淀、冻干和悬浮在变性缓冲液中。蛋白质可通过加热而变性。可能优选不太会对蛋白质造成化学修饰的变性方法。

[0058] 在缀合之前或之后,可对样品的蛋白质进行处理以产生更短的多肽。剩余的蛋白质可用酶如蛋白酶K部分消化以生成片段,或可以保持其完整。在进一步的实例中,蛋白质可暴露于蛋白酶如胰蛋白酶。蛋白酶的另外的实例可包括丝氨酸蛋白酶、半胱氨酸蛋白酶、苏氨酸蛋白酶、天冬氨酸蛋白酶、谷氨酸蛋白酶、金属蛋白酶和天冬酰胺肽裂解酶。

[0059] 在一些情况下,去除极大蛋白质(例如肌联蛋白)和小蛋白质可能是有用的,此类蛋白质可通过过滤或其他合适的方法去除。在一些实例中,极大蛋白质可包括超过400kD、450kD、500kD、600kD、650kD、700kD、750kD、800kD或850kD的蛋白质。在一些实例中,极大蛋白质可包括超过约8,000个氨基酸、约8,500个氨基酸、约9,000个氨基酸、约9,500个氨基酸、约10,000个氨基酸、约10,500个氨基酸、约11,000个氨基酸或约15,000个氨基酸的蛋白质。在一些实例中,小蛋白质可包括小于约10kD、9kD、8kD、7kD、6kD、5kD、4kD、3kD、2kD或1kD的蛋白质。在一些实例中,小蛋白质可包括少于约50个氨基酸、45个氨基酸、40个氨基酸、35个氨基酸或约30个氨基酸的蛋白质。可以通过大小排阻色谱法去除极大或小蛋白质。极大

蛋白质可通过大小排阻色谱法分离,用蛋白酶处理以产生中等大小的多肽,并与样品的中等大小的蛋白质重新组合。

[0060] 在一些情况下,蛋白质可按大小排序。在一些情况下,可通过将蛋白质分选到微孔中而将蛋白质排序。在一些情况下,可通过将蛋白质分选到纳米孔中而将蛋白质排序。在一些情况下,可通过使蛋白质通过凝胶如SDS-PAGE凝胶而将蛋白质排序。在一些情况下,可通过其他大小依赖性分级分离方法而将蛋白质排序。在一些情况下,可基于电荷分离蛋白质。在一些情况下,可基于疏水性分离蛋白质。在一些情况下,可基于其他物理特征分离蛋白质。在一些情况下,蛋白质可在变性条件下分离。在一些情况下,蛋白质可在非变性条件下分离。在一些情况下,可将分级分离的蛋白质的不同级分置于基底的不同区域上。在一些情况下,可将分离的蛋白质的不同部分置于基底的不同区域上。在一些情况下,蛋白质样品可在SDS-PAGE凝胶中分离并从SDS-PAGE凝胶转移到基底,使得蛋白质在连续的过程中按大小分选。在一些情况下,蛋白质样品可基于大小分选成三个级分,并且可将三个级分分别施加至基底的第一、第二和第三区域。在一些情况下,可以对本文所述的系统和方法中使用的蛋白质进行分选。在一些情况下,可以不对本文所述的系统和方法中使用的蛋白质进行分选。

[0061] 可以例如用可识别标签标记蛋白质,以允许样品的多路复用。可识别标签的一些非限制性实例包括:荧光团,或核酸条形码化的碱基连接体。所使用的荧光团可包括荧光蛋白,如GFP、YFP、RFP、eGFP、mCherry、tdtomato、FITC、Alexa Fluor 350、Alexa Fluor 405、Alexa Fluor 488、Alexa Fluor 532、Alexa Fluor 546、Alexa Fluor 555、Alexa Fluor 568、Alexa Fluor 594、Alexa Fluor 647、Alexa Fluor 680、Alexa Fluor 750、Pacific Blue、香豆素、BODIPY FL、Pacific Green、Oregon Green、Cy3、Cy5、Pacific Orange、TRITC、Texas Red、R-藻红蛋白、别藻蓝蛋白或其他本领域已知的荧光团。

[0062] 可以对任何数目的蛋白质样品进行多路复用。例如,多路复用的反应可含有来自2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、约20、约25、约30、约35、约40、约45、约50、约55、约60、约65、约70、约75、约80、约85、约90、约95、约100或多于100个起始样品的蛋白质。可识别标签可提供询问每个蛋白质的来源样品的方法,或者可指导来自不同样品的蛋白质隔离到固体支持物上的不同区域。

[0063] 基底

[0064] 在一些实施方案中,随后将蛋白质施加至官能化基底上,从而以化学方式将蛋白质附接至基底。在一些情况下,可通过生物素附接将蛋白质附接至基底。在一些情况下,可通过核酸附接将蛋白质附接至基底。在一些实施方案中,可将蛋白质施加至中间物质,其中随后将该中间物质附接至基底。在一些情况下,蛋白质可与珠子(例如,金珠子)缀合,然后可将珠子捕获在表面(例如,硫醇化表面)上。在一些情况下,一种蛋白质可与每个珠子缀合。在一些情况下,蛋白质可与珠子缀合(例如,每个珠子一个蛋白质),并且可将珠子捕获在表面上(例如在微孔和/或纳米孔中)。

[0065] 基底可以是能够形成固体支撑物的任何基底。如本文所用的基底或固体基底可以指蛋白质可共价或非共价附接的任何固体表面。固体基底的非限制性实例包括颗粒、珠子、载玻片、装置元件的表面、膜、流动池、孔、腔室、宏观流体腔室,其可以是平的或弯曲的,或者可以具有其他形状,并且可以是光滑的或有纹理的。在一些情况下,基底表面可包含微孔。在一些情况下,基底表面可包含纳米孔。在一些情况下,基底表面可包含与一个或多个

纳米孔组合的一个或多个微孔。在一些实施方案中,基底可以由玻璃、碳水化合物如葡聚糖、塑料如聚苯乙烯或聚丙烯、聚丙烯酰胺、胶乳、硅、金属如金,或纤维素组成,并且可被进一步修饰以允许或增强寡核苷酸的共价或非共价附接。例如,基底表面可通过用特定官能团如马来酸或琥珀酸部分修饰进行官能化,或者通过用化学反应性基团如氨基、巯基或丙烯酸基团修饰(例如通过硅烷化)进行衍生化。合适的硅烷试剂包括氨丙基三甲氧基硅烷、氨丙基三乙氧基硅烷和4-氨基丁基三乙氧基硅烷。基底可用N-羟基琥珀酰亚胺(NHS)官能团进行官能化。玻璃表面还可使用例如环氧硅烷、丙烯酸硅烷或丙烯酰胺硅烷,通过如丙烯酸或环氧基等其他反应性基团进行衍生化。寡核苷酸附接的基底和方法优选对于反复的结合、洗涤、成像和洗脱步骤是稳定的。在一些实例中,基底可以是载玻片或流动池。

[0066] 可通过例如光刻、蘸笔纳米刻蚀、纳米压印刻蚀、纳米球刻蚀(nanosphere lithography)、纳米球刻蚀(nanoball lithography)、纳米柱阵列、纳米线刻蚀、扫描探针刻蚀、热化学刻蚀、热扫描探针刻蚀、局部氧化纳米刻蚀、分子自组装、模板刻蚀或电子束刻蚀来创建官能团的有序阵列。有序阵列中的官能团可被定位成使得每个官能团与任何其他官能团相距小于200纳米(nm),或约200nm、约225nm、约250nm、约275nm、约300nm、约325nm、约350nm、约375nm、约400nm、约425nm、约450nm、约475nm、约500nm、约525nm、约550nm、约575nm、约600nm、约625nm、约650nm、约675nm、约700nm、约725nm、约750nm、约775nm、约800nm、约825nm、约850nm、约875nm、约900nm、约925nm、约950nm、约975nm、约1000nm、约1025nm、约1050nm、约1075nm、约1100nm、约1125nm、约1150nm、约1175nm、约1200nm、约1225nm、约1250nm、约1275nm、约1300nm、约1325nm、约1350nm、约1375nm、约1400nm、约1425nm、约1450nm、约1475nm、约1500nm、约1525nm、约1550nm、约1575nm、约1600nm、约1625nm、约1650nm、约1675nm、约1700nm、约1725nm、约1750nm、约1775nm、约1800nm、约1825nm、约1850nm、约1875nm、约1900nm、约1925nm、约1950nm、约1975nm、约2000nm或大于2000nm。可以以一定浓度提供随机间隔的官能团,使得官能团与任何其他官能团平均相距至少约50nm、约100nm、约150nm、约200nm、约250nm、约300nm、约350nm、约400nm、约450nm、约500nm、约550nm、约600nm、约650nm、约700nm、约750nm、约800nm、约850nm、约900nm、约950nm、约1000nm或大于100nm。

[0067] 基底可间接地官能化。例如,可将基底聚乙二醇化,并且可将官能团施加至PEG分子的全部或亚组一部分。另外,如上所述,在一些情况下,珠子(例如,金珠子)可以是缀合的,然后可将珠子捕获在表面(例如,硫醇化表面)上。在一些情况下,一种蛋白质可与每个珠子缀合。在一些情况下,蛋白质可与珠子缀合(例如,每个珠子一个蛋白质),并且可将珠子捕获在表面上(例如在微孔和/或纳米孔中)。

[0068] 可使用适合于微尺度或纳米尺度结构(例如,有序结构如微孔、纳米孔、微柱、单分子阵列、纳米球、纳米柱或纳米线)的技术对基底进行官能化。在一些情况下,基底可具有不同大小的微孔。在一些情况下,微孔可以为1微米(μm),可以为约2 μm 、约3 μm 、约4 μm 、约5 μm 、约6 μm 、约7 μm 、约8 μm 、约9 μm 、约10 μm 、约15 μm 、约20 μm 、约25 μm 、约30 μm 、约35 μm 、约40 μm 、约45 μm 、约50 μm 、约55 μm 、约60 μm 、约65 μm 、约70 μm 、约75 μm 、约80 μm 、约85 μm 、约90 μm 、约95 μm 、约100 μm 、约105 μm 、约110 μm 、约115 μm 、约120 μm 、约125 μm 、约130 μm 、约135 μm 、约140 μm 、约145 μm 、约150 μm 、约155 μm 、约160 μm 、约165 μm 、约170 μm 、约175 μm 、约180 μm 、约185 μm 、约190 μm 、约195 μm 、约200 μm 、约205 μm 、约210 μm 、约215 μm 、约220 μm 、约225 μm 、约230 μm 、约235 μm 、约240 μm

m、约245 μ m、约250 μ m、约255 μ m、约260 μ m、约265 μ m、约270 μ m、约275 μ m、约280 μ m、约285 μ m、约290 μ m、约295 μ m、约300 μ m、约305 μ m、约310 μ m、约315 μ m、约320 μ m、约325 μ m、约330 μ m、约335 μ m、约340 μ m、约345 μ m、约350 μ m、约355 μ m、约360 μ m、约365 μ m、约370 μ m、约375 μ m、约380 μ m、约385 μ m、约390 μ m、约395 μ m、约400 μ m、约405 μ m、约410 μ m、约415 μ m、约420 μ m、约425 μ m、约430 μ m、约435 μ m、约440 μ m、约445 μ m、约450 μ m、约455 μ m、约460 μ m、约465 μ m、约470 μ m、约475 μ m、约480 μ m、约485 μ m、约490 μ m、约495 μ m、约500 μ m或大于500 μ m。在一些情况下,基底可具有大小范围为5 μ m至500 μ m的微孔。在一些情况下,基底可具有大小范围为约5 μ m至约500 μ m的微孔。在一些情况下,基底可具有大小范围为10 μ m至100 μ m的微孔。在一些情况下,基底可具有大小范围为约10 μ m至约100 μ m的微孔。在一些情况下,基底可具有一系列不同大小的微孔,使得不同大小的蛋白质可被分选成不同大小的微孔。在一些情况下,基底中的微孔可按大小分布(例如,较大微孔分布在第一区域中而较小微孔分布在第二区域中)。在一些情况下,基底可具有约十种不同大小的微孔。在一些情况下,基底可具有约20种不同大小、约25种不同大小、约30种不同大小、约35种不同大小、约40种不同大小、约45种不同大小、约50种不同大小、约55种不同大小、约60种不同大小、约65种不同大小、约70种不同大小、约75种不同大小、约80种不同大小、约85种不同大小、约90种不同大小、约95种不同大小、约100种不同大小或超过100种不同大小的微孔。

[0069] 在一些情况下,基底可具有不同大小的纳米孔。在一些情况下,纳米孔可以为约100纳米(nm)、约150nm、约200nm、约250nm、约300nm、约350nm、约400nm、约450nm、约500nm、约550nm、约600nm、约650nm、约700nm、约750nm、约800nm、约850nm、约900nm、约950nm,或950nm至1微米。在一些情况下,基底可具有大小范围为100nm至1微米的纳米孔。在一些情况下,基底可具有大小范围为100nm至500nm的纳米孔。在一些情况下,基底可具有一系列不同大小的纳米孔,使得不同大小的蛋白质可被分选至不同大小的纳米孔中。在一些情况下,基底中的纳米孔可以按大小分布(例如,较大纳米孔分布在第一区域中而较小纳米孔分布在第二区域中)。在一些情况下,基底可具有约十种不同大小的纳米孔。在一些情况下,基底可具有约20种不同大小或超过30种不同大小的纳米孔。

[0070] 在一些情况下,基底可具有一系列不同大小的纳米孔和/或微孔,使得不同大小的蛋白质可被分选至不同大小的纳米孔和/或微孔中。在一些情况下,基底中的纳米孔和/或微孔可以按大小分布(例如,较大微孔分布在第一区域中而较小纳米孔分布在第二区域中)。在一些情况下,基底可具有约十种不同大小的纳米孔和/或微孔。在一些情况下,基底可具有约20种不同大小、约25种不同大小、约30种不同大小、约35种不同大小、约40种不同大小、约45种不同大小、约50种不同大小、约55种不同大小、约60种不同大小、约65种不同大小、约70种不同大小、约75种不同大小、约80种不同大小、约85种不同大小、约90种不同大小、约95种不同大小、约100种不同大小或超过100种不同大小的纳米孔和/或微孔。

[0071] 基底可包含任何材料,包括金属、玻璃、塑料、陶瓷或其组合。在一些优选的实施方案中,固体基底可以是流动池。流动池可以由单层或多层组成。例如,流动池可包含基层(例如,硼硅酸盐玻璃)、覆盖在基层上的通道层(例如,蚀刻的硅)以及覆盖层或顶层。当这些层组装在一起时,可以形成封闭的通道,在任一端具有穿过覆盖层的入口/出口。每层的厚度可以变化,但优选小于约1700 μ m。这些层可由本领域已知的任何合适的材料组成,包括但不限于光敏玻璃、硼硅酸盐玻璃、熔融硅酸盐、PDMS或硅。不同的层可由相同的材料或不同的

材料组成。

[0072] 在一些实施方案中,流动池可包含在流动池底部上的通道的开口。流动池可在可被离散地可视化的位置上包含数百万个附接的靶标缀合位点。在一些实施方案中,与本发明的实施方案一起使用的各种流动池可包含不同数目的通道(例如,1个通道、2个或更多个通道、3个或更多个通道、4个或更多个通道、6个或更多个通道、8个或更多个通道、10个或更多个通道、12个或更多个通道、16个或更多个通道,或超过16个通道)。各种流动池可包含不同深度或宽度的通道,深度或宽度可在单个流动池内的通道之间不同,或在不同流动池的通道之间不同。单个通道的深度和/或宽度也可以变化。例如,在通道内的一个或多个点处,通道可以为小于约50 μm 深、约50 μm 深、小于约100 μm 深、约100 μm 深、约100 μm 深至约500 μm 深、约500 μm 深或大于约500 μm 深。通道可具有任何横截面形状,包括但不限于圆形、半圆形、矩形、梯形、三角形或卵形的横截面。

[0073] 可将蛋白质点样、滴加、移液、流动、洗涤,或以其他方式施加至基底。在基底已经用诸如NHS酯等部分官能化的情况下,不需要对蛋白质进行修饰。在基底已经用替代的部分(例如巯基、胺或连接体核酸)官能化的情况下,可以使用交联试剂(例如辛二酸二琥珀酰亚胺酯、NHS-磺胺)。在基底已经用连接体核酸官能化的情况下,可以用互补核酸标签修饰样品的蛋白质。

[0074] 在一些情况下,蛋白质可与核酸缀合。使用核酸,可以形成核酸纳米球,从而使蛋白质与核酸纳米球连接。当核酸纳米球附接到基底时,与核酸附接的蛋白质通过核酸纳米球附接至基底。DNA纳米球可附接(例如通过吸附或通过缀合)至基底。基底可具有胺官能化表面,核酸纳米球可附接至该表面。

[0075] 在一些情况下,可形成具有功能活性末端(例如马来酰亚胺、NHS酯等)的核酸纳米球。然后,蛋白质可与纳米球缀合,从而使蛋白质与核酸纳米球相连。当核酸纳米球附接至基底时,与核酸附接的蛋白质通过核酸纳米球附接至基底。DNA纳米球可附接(例如通过吸附或通过缀合)至基底。基底可具有胺官能化表面,核酸纳米球可附接至该表面。

[0076] 可光活化的交联剂可用于使样品直接交联至基底上的特定区域。可光活化的交联剂可用于通过将每个样品附接在基底的已知区域中来允许蛋白质样品的多路复用。可光活化的交联剂可例如通过在蛋白质交联之前检测荧光标签来允许已经成功标记的蛋白质的特异性附接。可光活化的交联剂的实例包括但不限于N-5-叠氮基-2-硝基苯甲酰氧基琥珀酰亚胺、磺基琥珀酰亚胺基6-(4'-叠氮基-2'-硝基苯基氨基)己酸酯、琥珀酰亚胺基4,4'-氮杂戊酸酯、磺基琥珀酰亚胺基4,4'-氮杂戊酸酯、琥珀酰亚胺基6-(4,4'-氮杂戊酰胺基)己酸酯、磺基琥珀酰亚胺基6-(4,4'-氮杂戊酰胺基)己酸酯、琥珀酰亚胺基2-((4,4'-氮杂戊酰胺基)乙基)-1,3'-二硫代丙酸酯和磺基琥珀酰亚胺基2-(4,4'-氮杂戊酰胺基)乙基)-1,3'-二硫代丙酸酯。

[0077] 还可通过限制每个样品与基底上的离散区域的结合来使样品多路复用。例如,基底可以组织成小道(lane)。用于多路复用的另一种方法是反复地将样品施加至整个基底,在每次样品施加之后使用非特异性蛋白质结合试剂或染料进行蛋白质检测步骤。在一些情况下,染料的实例可包括荧光蛋白凝胶染色剂如SYPRO[®] Ruby、SYPRO[®] Orange、SYPRO[®] Red、SYPRO[®] Tangerine和Coomassie[™] Fluor Orange。

[0078] 通过在每次添加样品后跟踪所有蛋白质的位置,可以确定基底上的每个位置首先包含蛋白质的阶段,从而确定蛋白质来源于哪个样品。该方法还可在每次施加样品后确定基底的饱和度,并允许蛋白质在基底上结合的最大化。例如,如果在第一次施加样品后仅有30%的官能化位置被蛋白质占据,则可以进行相同样品的第二次施加或不同样品的施加。

[0079] 多肽可通过一个或多个残基附接至基底。在一些实例中,多肽可经由N末端、C末端、两个末端或通过内部残基附接。

[0080] 除了永久性交联剂之外,对于一些应用来说,使用可光切割的连接体可能是合适的,并且这样做使得能够在分析后从基底中选择性地提取蛋白质。在一些情况下,可光切割的交联剂可用于几种不同的多路复用样品。在一些情况下,可光切割的交联剂可用于多路复用反应中的一个或多个样品中。在一些情况下,多路复用反应可包括通过永久性交联剂交联至基底的对照样品和通过可光切割的交联剂交联至基底的实验样品。

[0081] 每个缀合的蛋白质可在空间上彼此分开,使得每个缀合的蛋白质是光学可分辨的。因此,蛋白质可以用独特空间地址单独标记。在一些实施方案中,这可以通过使用低浓度蛋白质和基底上的低密度附接位点进行缀合以使得每个蛋白质分子在空间上彼此分开而来实现。在使用可光活化的交联剂的实例中,可以使用光图案,使得蛋白质固定到预定位置。

[0082] 在一些方法中,已经纯化的大批蛋白质可缀合至基底并使用本文所述的方法处理,以便鉴定纯化的蛋白质。大批蛋白质可包含已收集在一起的纯化蛋白质。在一些实例中,大批蛋白质可在空间上与每个其他的缀合蛋白质或大批蛋白质分开的位置处缀合,使得每个缀合蛋白质或大批蛋白质是光学可分辨的。因此,蛋白质或大批蛋白质可以用独特空间地址单独标记。在一些实施方案中,这可以通过低浓度蛋白质和使用基底上的低密度附接位点进行缀合以使得每个蛋白质分子在空间上彼此分开来实现。在使用可光活化的交联剂的实例中,可以使用光图案,使得一种或多种蛋白质固定到预定位置。

[0083] 在一些实施方案中,每个蛋白质可与独特空间地址相关联。例如,一旦蛋白质在空间上分离的位置处附接至基底,每个蛋白质可以例如通过坐标被指定索引地址。在一些实例中,预先指定的独特空间地址的网格可以预先确定。在一些实施方案中,基底可含有易于识别的固定标记,使得可相对于基底的固定标志确定每个蛋白质的放置。在一些实例中,基底可具有永久标记在表面上的网格线和/或“原点”或其他基准点。在一些实例中,可永久地或半永久地标记基底的表面,以提供用以定位交联蛋白质的参考。图案化本身的形状,如缀合多肽的外部边界也可用作基准点,用于确定每个斑点的独特位置。

[0084] 基底还可含有缀合的蛋白质标准品和对照。缀合的蛋白质标准品和对照可以是缀合至已知位置的已知序列的肽或蛋白质。在一些实例中,缀合的蛋白质标准品和对照可用作测定中的内部对照。蛋白质可以从纯化的蛋白质原料施加至基底,或者可以通过诸如核酸可编程蛋白质阵列(NAPPA)等过程在基底上合成。

[0085] 在一些实例中,基底可包含荧光标准品。这些荧光标准品可用于校准测定之间的荧光信号强度。这些荧光标准品也可用于将荧光信号的强度与区域中存在的荧光团的数目相关联。荧光标准品可包括在测定中使用的一些或所有不同类型的荧光团。

[0086] 亲和试剂

[0087] 一旦基底与来自样品的蛋白质缀合,就可进行多亲和试剂测量。本文所述的测量

过程可采用多种亲和试剂。

[0088] 亲和试剂可以是以可再现的特异性结合蛋白质或肽的任何试剂。例如，亲和试剂可以是抗体、抗体片段、适体或肽。在一些实例中，单克隆抗体可能是优选的。在一些实例中，抗体片段如Fab片段可能是优选的。在一些情况下，亲和试剂可以是市售亲和试剂，如市售抗体。在一些情况下，可通过筛选市售亲和试剂以鉴定具有有用特征的亲和试剂来选择所需的亲和试剂。在一些情况下，可针对亲和试剂结合单一蛋白质的能力对其进行筛选。在一些情况下，可针对亲和试剂结合表位或氨基酸序列的能力对其进行筛选。在一些情况下，可针对亲和试剂组通过差异结合来集体地分辨相似蛋白质（例如具有高度相似序列的蛋白质）的能力对其进行筛选。在一些情况下，亲和试剂可被筛选为具有重叠的结合特征以增加对特定蛋白质的结合特异性。亲和试剂的筛选可以以多种不同方式进行。一个实例是针对NAPPA或表位平铺阵列来筛选亲和试剂。在一些情况下，可以使用并设计为与蛋白质靶标结合的蛋白质特异性亲和试剂（例如市售抗体或适体）。在一些情况下，可以在结合测量之前混合多种蛋白质特异性亲和试剂。例如，对于每个结合测量运转，可以选择蛋白质特异性亲和试剂的新混合物，该新混合物包含从完整组中随机选择的可用亲和试剂的亚组。例如，每种后续混合物可以以相同的随机方式生成，期望许多亲和试剂将存在于多于一种混合物中。在一些情况下，可以使用蛋白质特异性亲和试剂的混合物更快速地生成蛋白质鉴定。在一些情况下，蛋白质特异性亲和试剂的此类混合物可增大亲和试剂在任何单独的运转中结合的未知蛋白质的百分比。亲和试剂的混合物可以由所有可用的亲和试剂的1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%或更多组成。

[0089] 亲和试剂可具有高、中等或低特异性。在一些实例中，亲和试剂可识别几种不同的表位。在一些实例中，亲和试剂可识别存在于两种或更多种不同蛋白质中的表位。在一些实例中，亲和试剂可识别存在于多种不同蛋白质中的表位。在一些情况下，在本公开内容的方法中使用的亲和试剂可以对单个表位具有高度特异性。在一些情况下，在本公开内容的方法中使用的亲和试剂可以对含有翻译后修饰的单个表位具有高度特异性。

[0090] 在一些实施方案中，被导向鉴定靶氨基酸序列的亲和试剂实际上可包含一组不同的组分，这些组分在本文所述的方法中使用彼此无差异或可区分。特别地，可用于鉴定相同靶氨基酸序列的不同组分可使用相同的检测部分来鉴定相同的靶氨基酸序列。例如，无论侧翼序列如何均结合三聚体氨基酸序列(AAA)的亲和试剂可包含在不受侧翼序列的任何影响的情况下结合三聚体AAA序列的单个探针，或者包含一组400个探针，每个探针都与 α AAA β 形式的不同5氨基酸表位结合，其中 α 和 β 可以是任何氨基酸。在第二种情况的一些中，可以组合400个探针，使得每个探针的量相等。在第二种情况的一些中，可以组合400个探针，使得每个探针的量可通过每个探针的特征性结合亲和力加权，使得任何给定的5氨基酸表位被结合的概率相等。

[0091] 可通过本领域已知的任何方法生成新的亲和试剂。开发亲和试剂的方法包括SELEX、噬菌体展示和接种。在一些实例中，可使用基于结构的药物设计方法设计亲和试剂。基于结构的药物设计(或直接药物设计)利用到关于感兴趣表位的三维结构和亲和试剂的结合位点的知识。

[0092] 在一些情况下，亲和试剂可以用核酸条形码标记。在一些实例中，核酸条形码可用于纯化使用后的亲和试剂。在一些实例中，核酸条形码可用于分选亲和试剂以供重复使用。

在一些情况下,亲和试剂可以用荧光团标记,该荧光团可用于分选使用后的亲和试剂。

[0093] 在一些情况下,用核酸条形码标记的多种亲和试剂可以是多路复用的,然后使用互补核酸探针进行检测。可以使用具有不同检测部分的多个互补核酸在单个循环中检测一组多路复用的亲和试剂。在一些情况下,可以使用与检测部分缀合的单个互补核酸在多个循环中检测一组多路复用的亲和试剂。在一些情况下,可以使用各自与不同检测部分缀合的多个互补核酸在多个循环中检测一组多路复用的亲和试剂。在一些情况下,可以使用各自与不同组检测部分缀合的多个互补核酸在多个循环中检测一组多路复用的亲和试剂。

[0094] 在一些情况下,用核酸条形码标记的一种或多种亲和试剂可以与结合的蛋白质交联。一旦一种或多种亲和试剂与蛋白质交联,就可以对条形码进行测序以确定交联亲和试剂的身份。在一些情况下,多种结合蛋白可暴露于一种或多种亲和试剂。在一些情况下,当多种结合的蛋白质与一种或多种亲和试剂交联时,可以对与结合的亲和试剂相关联的条形码进行测序以确定与多种结合的蛋白质中的每一种相关联的交联亲和试剂的身份。

[0095] 亲和试剂家族可包含一种或多种类型的亲和试剂。例如,本公开内容的方法可以使用亲和试剂家族,其包含抗体、抗体片段、Fab片段、适体、肽和蛋白质中的一种或多种。

[0096] 可以修饰亲和试剂。修饰包括但不限于检测部分的附接。检测部分可以直接或间接附接。例如,检测部分可以直接与亲和试剂共价附接,或者可以通过连接体附接,或者可以通过亲和反应附接,如互补核酸标签或生物素链霉亲和素对。能够经受亲和试剂的温和洗涤和洗脱的附接方法可能是优选的。

[0097] 检测部分包括但不限于荧光团、生物发光蛋白、包含恒定区和条形码区的核酸区段,或用于与纳米颗粒如磁性颗粒连接的化学系链(tether)。检测部分可包括具有不同激发或发射模式的几种不同荧光团。

[0098] 检测部分可以从亲和试剂切下。这可以实现从不再感兴趣的亲和试剂上去除检测部分以减少信号污染的步骤。

[0099] 在一些情况下,亲和试剂是未修饰的。例如,如果亲和试剂是抗体,则可以通过原子力显微术检测抗体的存在。亲和试剂可以是未修饰的,并且可以例如通过对一种或多种亲和试剂具有特异性的抗体来检测。例如,如果亲和试剂是小鼠抗体,则可以通过使用抗小鼠第二抗体来检测小鼠抗体。或者,亲和试剂可以是适体,该适体由对适体具有特异性的抗体来检测。可以用如上所述的检测部分修饰第二抗体。在一些情况下,可以通过原子力显微术检测第二抗体的存在。

[0100] 在一些实例中,亲和试剂可包含相同的修饰,例如缀合的绿色荧光蛋白,或者可包含两种或更多种不同类型的修饰。例如,每种亲和试剂可与各自具有不同激发或发射波长的几种不同荧光部分中的一种缀合。这可以允许对亲和试剂的多路复用,因为可以组合和/或区分几种不同的亲和试剂。在一个实例中,第一亲和试剂可与绿色荧光蛋白缀合,第二亲和试剂可与黄色荧光蛋白缀合,并且第三亲和试剂可与红色荧光蛋白缀合,因此三种亲和试剂可以是多路复用的并通过其荧光进行鉴定。在另一个实例中,第一、第四和第七亲和试剂可与绿色荧光蛋白缀合,第二、第五和第八亲和试剂可与黄色荧光蛋白缀合,并且第三、第六和第九亲和试剂可与红色荧光蛋白缀合;在这种情况下,第一、第二和第三亲和试剂可以一起多路复用,而第二、第四和第七,以及第三、第六和第九亲和试剂形成两个进一步的多路复用反应。可以一起多路复用的亲和试剂的数目可取决于用于区分它们的检测部分。

例如,用荧光团标记的亲试剂的多路复用可能受到可用的独特荧光团数目的限制。对于进一步的实例,用核酸标签标记的亲试剂的多路复用可由核酸条形码的长度决定。

[0101] 可以在用于测定之前确定每种亲试剂的特异性。可以在使用已知蛋白质的对照实验中确定亲试剂的结合特异性。可以使用任何合适的实验方法来确定亲试剂的特异性。在一个实例中,基底可以在已知位置处负载已知的蛋白质标准品并用于评估多种亲试剂的特异性。在另一个实例中,基底可包含实验样品以及对照和标准品的小组,使得每种亲试剂的特异性可从与对照和标准品的结合来计算,然后用于鉴定实验样品。在一些情况下,可以包括具有未知特异性的亲试剂以及已知特异性的亲试剂,来自已知特异性的亲试剂的数据可用于鉴定蛋白质,而未知特异性的亲试剂与鉴定的蛋白质的结合模式可用于确定其结合特异性。还可以通过使用其他亲试剂的已知结合数据来重新确认任何单独的亲试剂的特异性,以评估该单独的亲试剂结合的蛋白质。因此,通过亲试剂小组的多次使用,亲试剂的特异性可以随着每次反复而逐渐改善。虽然可以使用对特定蛋白质具有独特特异性的亲试剂,但是本文所述的方法可以不需要它们。另外,方法可以在一定特异性范围内有效。在一些实例中,当亲试剂对任何特定蛋白质都不具特异性但对氨基酸基序(例如三肽AAA)具有特异性时,本文所述的方法可能特别有效。

[0102] 在一些实例中,可以选择一种或多种亲试剂来结合给定长度如2、3、4、5、6、7、8、9、10或大于10个氨基酸的氨基酸基序。在一些实例中,可以选择一种或多种亲试剂来结合具有2个氨基酸到40个氨基酸的一系列不同长度的氨基酸基序。

[0103] 在一些实例中,可以选择具有高、中等或低结合亲和力的亲试剂。在一些情况下,具有低或中等结合亲和力的亲试剂可能是优选的。在一些情况下,亲试剂可具有约 10^{-3}M 、 10^{-4}M 、 10^{-5}M 、 10^{-6}M 、 10^{-7}M 、 10^{-8}M 、 10^{-9}M 、 10^{-10}M 或更低的解离常数。在一些情况下,亲试剂可具有大于约 10^{-10}M 、 10^{-9}M 、 10^{-8}M 、 10^{-7}M 、 10^{-6}M 、 10^{-5}M 、 10^{-4}M 、 10^{-3}M 、 10^{-2}M 或更高的解离常数。

[0104] 可以选择一些亲试剂来结合修饰的氨基酸序列,如磷酸化或遍在蛋白化的氨基酸序列。在一些实例中,可以选择对可以由一种或多种蛋白质包含的表位家族具有广泛特异性的一种或多种亲试剂。在一些实例中,一种或多种亲试剂可以结合两种或更多种不同的蛋白质。在一些实例中,一种或多种亲试剂可与其一种或多种靶标弱结合。例如,亲试剂可与其一种或多种靶标以少于10%、少于10%、少于15%、少于20%、少于25%、少于30%、少于35%或少于35%结合。在一些实例中,一种或多种亲试剂可中等或强烈地与其一种或多种靶标结合。例如,亲试剂可与其一种或多种靶标以超过35%、超过40%、超过45%、超过60%、超过65%、超过70%、超过75%、超过80%、超过85%、超过90%、超过91%、超过92%、超过93%、超过94%、超过95%、超过96%、超过97%、超过98%或超过99%结合。

[0105] 为了补偿弱结合,可将过量的亲试剂施加至基底。亲试剂可以相对于样品蛋白质以约1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1或10:1过量施加。亲试剂可以相对于样品蛋白质中表位的预期出现率以约1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1或10:1过量施加。

[0106] 亲试剂还可包含磁性组分。磁性组分可用于将一些或所有结合的亲试剂操纵到相同的成像平面或z堆叠(stack)中。将一些或所有亲试剂操纵到相同的成像平面中可

以改善成像数据的质量并降低系统中的噪声。

[0107] 结合测量

[0108] 给定一组修饰的亲试剂和缀合的基底,可将亲试剂反复施加至基底。每个测量循环由几个阶段组成。在第一阶段,将亲试剂施加至基底,在基底上它们可吸附到缀合的蛋白质上。

[0109] 接下来,可以轻轻洗涤基底以去除非特异性结合。该洗涤步骤可以在不洗脱与固定化蛋白质结合的亲试剂的条件下进行。可用于该步骤的缓冲液的一些实例包括磷酸盐缓冲盐水、Tris缓冲盐水、具有Tween20的磷酸盐缓冲盐水和具有Tween20的Tris缓冲盐水。

[0110] 吸附后,确定每种修饰的亲试剂的结合地址,如通过测量已直接与亲试剂缀合的荧光团,或测量已缀合至与亲试剂缀合的核酸链的互补核酸的荧光团。检测方法由检测部分的选择决定。可以光学地检测荧光团和生物发光部分,在一些情况下需要第二检测试剂。可以在结合测量之前确定基底上每个固定化蛋白质的独特地址,或者可以通过结合测量生成包含固定化蛋白质的地址列表。

[0111] 接下来,可通过更严格的洗涤来解吸附亲试剂。该洗涤步骤可从固定的基底去除一些或所有亲试剂。在一些情况下,亲试剂可被选择为具有低至中等结合亲和力以促进去除。可以重新捕获使用过的亲试剂以便重复使用或丢弃。在使用具有可切割检测部分的亲试剂的实例中,可以在该阶段切割和去除检测部分。在严格洗涤之后,在一些实例中,可以猝灭任何剩余的荧光,并且施加甚至更严格的洗涤以去除残余的亲试剂。在施加下一个亲试剂之前,可通过对基底进行重新成像来检测残留/污染。还可通过连续监视图像中的再现信号来检测污染。这就结束了一个分析循环。

[0112] 在一些实施方案中,可通过暴露于激活波长下的长时间强光将荧光标记的亲试剂猝灭。荧光标签的猝灭可以代替洗涤步骤来去除亲试剂。在一些实施方案中,可能需要循环 n 个荧光团以区分哪些信号来自先前的 $n-1$ 个循环。

[0113] 对于每种亲试剂或其多路复用继续进行循环。测量阶段的结果是非常大的表格,其列出了每种亲试剂的结合坐标,或者在每个坐标位置处结合的亲试剂,参见例如图10。

[0114] 分析

[0115] 蛋白质鉴定的最后一步可包括软件工具,以从关于哪种亲试剂与坐标结合的信息确定在基底的每个坐标处的每种蛋白质的最可能的身份。该软件可利用关于每种亲试剂的结合特征的信息。例如,如果给定的亲试剂优先与含有三肽表位AAA的蛋白质结合。鉴于关于每种亲试剂的结合特征、样品中蛋白质的数据库和结合坐标列表、结合模式的信息,软件工具为每个坐标分配可能的身份以及对该身份的置信度。在亲试剂与蛋白质之间精确1-1映射的极端情况下,这可以通过简单的查表来完成。然而,在结合更复杂的情况下,这可以通过解决适当的满足问题来执行。在结合特征高度复杂的情况下,可以采用期望最大化方法。

[0116] 软件还可利用其中每种亲试剂均不结合的一些或所有位置的列表,并使用关于表位不存在的信息来确定存在的蛋白质。软件还可利用关于哪些亲试剂与每个地址结合和不结合的信息。因此,软件将使用关于哪些表位存在以及哪些表位不存在的信息。软件可包含数据库。数据库可包含品所获自的物种中的一些或所有已知蛋白质的序列。例如,如果

已知样品是人源的,则可以使用具有一些或所有人蛋白质的序列的数据库。如果样品的物种是未知的,则可以使用一些或所有蛋白质序列的数据库。数据库还可包含一些或所有已知蛋白质变体和突变蛋白质的序列,以及可由DNA移码突变产生的一些或所有可能蛋白质的序列。数据库还可包含可能由过早终止密码子或由降解产生的可能截短蛋白质的序列。

[0117] 该软件可包含一种或多种算法,如机器学习、深度学习、统计学习、监督学习、无监督学习、聚类、期望最大化、最大似然估计、贝叶斯推理、线性回归、逻辑回归、二元分类、多项分类或其他模式识别算法。例如,软件可以执行一种或多种算法来分析以下信息(例如,作为一种或多种算法的输入):(i)每种亲和试剂的结合特征,(ii)样品中蛋白质的数据库,(iii)结合坐标列表和/或(iv)亲和试剂与蛋白质的结合模式,以生成或分配(例如,作为一种或多种算法的输出)(a)每个坐标的可能身份和/或(b)对该身份的置信度(例如,置信水平和/或置信区间)。机器学习算法的实例可包括支持向量机(SVM)、神经网络、卷积神经网络(CNN)、深度神经网络、级联神经网络、k-最近邻(k-NN)分类、随机森林(RF)以及其他类型的分类和回归树(CART)。

[0118] 可通过在基底上执行本公开内容的方法来训练软件,该基底上每个地址处的蛋白质的身份均是预先确定的。例如,可以使用核酸可编程蛋白质阵列或表位平铺阵列作为训练数据集来训练软件。

[0119] 确定样品的特征

[0120] 一旦解码完成,就定义缀合至每个地址的蛋白质的可能身份。因此,可以通过对观察结果进行计数来估计它们在混合物中的丰度。因此,可以编制混合物中存在的每种蛋白质的列表,以及该蛋白质的观察数目。

[0121] 此外,如果使用可光切割连接体或其他形式的特异性可切割连接体将蛋白质附接至基底,则感兴趣的特定蛋白质可从基底释放并收集以供进一步研究。例如,可以鉴定和洗脱特定蛋白质以供进一步研究。本公开内容的方法还可用作从混合物中纯化和/或分离所需蛋白质的方法。在一些情况下,该方法可能能够纯化和/或分离特定同种型或翻译后修饰的蛋白质。在可能的蛋白质和相关序列的完整列表无法获得的样品中,该方法可能能够区分不同蛋白质组中的不同蛋白质,然后可以洗脱这些蛋白质以供进一步研究。例如,对于含有许多未知蛋白质的高度复杂的样品,如肠道微生物组样品,本文所述的方法可用于在质谱法之前将样品分级分离。在一些情况下,一旦可以判定蛋白质的身份,就可以从基底洗脱蛋白质。在鉴定蛋白质时从基底去除蛋白质允许后续轮次的亲和试剂结合继续用于身份尚未被判定的蛋白质,并且可以降低背景噪声和其余轮次的脱靶信号。在一些实例中,可以使用对特定蛋白质具有特异性的一种或多种亲和试剂作为第一轮来鉴定血液样品中的高丰度蛋白质,如血清白蛋白或免疫球蛋白,然后可以在过程中较早地去除这些高丰度蛋白质。在一些情况下,在每轮亲和试剂结合后,或在每第二、第三、第四、第五、第六、第七、第八、第九、第十、第十五、第二十或超过第二十轮的亲和试剂结合后,可以去除基底上的蛋白质的亚组。在每轮蛋白质洗脱后,信噪比可能会增加。

[0122] 在一些情况下,未鉴定的蛋白质可基于其结合模式进行分组或聚类。例如,在一些情况下,样品中存在的蛋白质可能没有在序列数据库中表示。未鉴定的蛋白质可基于其与亲和探针的结合模式聚类成组,目标是每个组含有样品中一组具有相同序列的未知蛋白质。可估计每组的蛋白质量并将其包括在定量分析中,该定量分析包括但不限于健康状态

与疾病状态之间的差异定量、纵向分析或生物标志物发现。在一些情况下,可以从基底选择性地去除未鉴定的基团,以通过质谱法进行鉴定。在其他情况下,可通过执行专门设计用于生成确信鉴定的进一步结合亲和力和测量实验来鉴定未鉴定的组。

[0123] 在一些情况下,在蛋白质或一组蛋白质已经去除后,可将额外的样品添加至基底。例如,血清白蛋白是血清中的高丰度蛋白质,其可占样品中蛋白质的约一半,在第一轮亲和试剂结合后去除血清白蛋白可允许向基底添加另外的血液样品。在一些实施方案中,可优选在将样品固定在基底上之前去除高丰度蛋白质,例如通过免疫沉淀或亲和柱纯化。

[0124] 可使用本公开内容的方法鉴定蛋白质修饰。例如,可通过使用修饰特异性检测试剂的检测与间插的酶处理(例如磷酸酶处理)的反复循环来鉴定翻译后修饰。对不同修饰具有特异性的亲和试剂可用于确定在固定化蛋白质上是否存在此类修饰。该方法还允许在有和没有给定修饰的情况下定量每种蛋白质的实例的数目。

[0125] 可通过将样品蛋白质的结合模式与预测的蛋白质身份之间的匹配不一致性来检测蛋白质中的突变。例如,基底上除了一种亲和试剂的结合之外,与已知蛋白质的亲和试剂结合概况匹配的固定化蛋白质或多肽可具有氨基酸置换。由于亲和试剂可具有重叠表位,尽管具有单个氨基酸置换,但固定化蛋白质可能与预测的亲和结合模式具有数个错配。还可检测到引起过早终止密码子移码的DNA突变。

[0126] 所需的亲和试剂的数目可小于样品中存在的表位的总数。例如,如果选择亲和试剂使得每种亲和试剂识别一种独特的三肽表位,则识别样品中所有可能表位的亲和试剂的总集为 $20 \times 20 \times 20 = 8000$ 。然而,本公开内容的方法可能仅需要这些亲和试剂中的约100、200、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500或6000种。在一些情况下,该方法可能仅需要少于约500、1000、2500、3000、3500、4000、4500、5000、5500或6000种亲和试剂。图13示出了模拟结果,其示出了在给定一组对独特氨基酸3-聚体具有特异性的x种亲和试剂的情况下可鉴定的已知人蛋白质的百分比随每种亲和试剂的结合效率的变化。如图13所示,可以用8000种3-聚体亲和试剂独特地鉴定98%的人蛋白质,并且结合可能性为10%。

[0127] 本公开内容的方法可以是高度准确的。本公开内容的方法可以能够以约90%、91%、92%、93%、94%、95%、96%、97%、98%、98.5%、99%、99.5%、99.9%或大于99.9%的准确度鉴定每种蛋白质。

[0128] 本公开内容的方法可以能够以约90%、91%、92%、93%、94%、95%、96%、97%、98%、98.5%、99%、99.5%、99.9%或大于99.9%的置信度预测每种蛋白质的身份。对于样品中的不同蛋白质,置信度可能不同。例如,具有非常独特序列的蛋白质可以以比其他蛋白质高度相似的蛋白质更高的置信度鉴定。在一些情况下,可以以高置信度将蛋白质鉴定为蛋白质家族的一部分,然而蛋白质的确切身份可能以较低的置信度预测。在一些情况下,极大或极小的蛋白质可以以比更中等大小的蛋白质更低的置信度预测。

[0129] 在一些情况下,可以以高置信度将蛋白质鉴定为蛋白质家族的一部分,然而蛋白质的确切身份可能以较低的置信度预测。例如,含有单个氨基酸变体的蛋白质可能难以以高置信度从蛋白质的规范形式中分辨出来。在这种情况下,规范序列和含有单个氨基酸变体的形式都不具有高置信度,但是可以对作为含有这两种序列的一组蛋白质的一部分的未知蛋白质进行高置信度评估。在蛋白质可具有多个相似序列的相关同种型的情况下可能发

生类似的情况。

[0130] 本公开内容的方法可以能够鉴定给定样品中的一些或所有蛋白质。本公开内容的方法可以能够鉴定样品中约90%、91%、92%、93%、94%、95%、96%、97%、98%、98.5%、99%、99.5%、99.9%或超过99.9%的蛋白质。

[0131] 本公开内容的方法可以能够快速鉴定样品中的蛋白质。本公开内容的方法可以能够每天每个流动池鉴定超过约100、约1000、约5000、约10000、约20,000、约30,000、约40,000、约50,000、约100,000、1,000,000、约10,000,000、约100,000,000、约1,000,000,000、约10,000,000,000、约100,000,000,000、约1,000,000,000,000个蛋白质。本公开内容的方法可以能够每天每个流动池鉴定超过约 10^{10} 、 10^{11} 、 10^{12} 、 10^{13} 、 10^{14} 、 10^{15} 、 10^{16} 、 10^{17} 个或超过约 10^{17} 个蛋白质。本公开内容的方法可以能够每天每个流动池鉴定约 10^{10} - 10^{12} 、 10^{11} - 10^{14} 、 10^{12} - 10^{16} 或 10^{13} - 10^{17} 个蛋白质。本公开内容的方法可以能够每天每个流动池鉴定约10pg、约20pg、约30pg、约40pg、约50pg、约60pg、约70pg、约80pg、约90pg、约100pg、约300pg、约300pg、约400pg、约500pg、约600pg、约700pg、约800pg、约900pg、约1ng、约2ng、约3ng、约4ng、约5ng、约6ng、约7ng、约8ng、约8ng、约10ng、约10ng、约20ng、约30ng、约40ng、约50ng、约60ng、约70ng、约80ng、约90ng、约100ng、约300ng、约300ng、约400ng、约500ng、约600ng、约700ng、约800ng、约900ng、约1μg、约2μg、约3μg、约4μg、约5μg、约6μg、约7μg、约8μg、约8μg、约10μg、约10μg、约20μg、约30μg、约40μg、约50μg、约60μg、约70μg、约80μg、约90μg、约100μg、约300μg、约300μg、约400μg、约500μg、约600μg、约700μg、约800μg、约900μg或大于约1mg蛋白质内的超过95%的蛋白质。

[0132] 本公开内容的方法可用于在实验处理后评估蛋白质组。本公开内容的方法可用于评估治疗干预的效果。

[0133] 本公开内容的方法可用于生物标志物发现。监测患有和不患有疾病的受试者中的蛋白质组表达可鉴定生物标志物。在发展出疾病之前的受试者中或在存在发展出疾病的风险的受试者中监测蛋白质组表达可鉴定预测风险的生物标志物。评估受试者的蛋白质组表达可指示受试者的健康或发展出某些疾病或病症的风险。本公开内容的方法可用于评估疗法,或区分药物/疗法的应答者与非应答者。本公开内容的方法可特别用于个性化医疗。

[0134] 本公开内容的方法可用于诊断疾病。不同的疾病或疾病阶段可与不同的蛋白质表达小组相关。不同的蛋白质表达小组可与每种给定治疗的不同治疗结果相关。受试者的蛋白质组表达数据可用于诊断受试者和/或选择最合适的疗法。

[0135] 本公开内容的方法可用于鉴定样品所来自的个体或物种。例如,本公开内容的方法可用于确定样品是否实际上来自所声称的物种或来源。在具有丰富蛋白质但核酸有限的样品中,本文所述的方法可优于基于PCR的方法。例如,鉴定蜂蜜样品的来源。对于进一步的实例,本公开内容的方法可用于评估食品安全和食品质量控制。

[0136] 本公开内容的方法可用于采用比可能蛋白质的数目更少的亲和试剂鉴定蛋白质分子池中的任何单蛋白质分子。例如,该方法可使用亲和试剂小组,以高于阈值量的确定性鉴定n种可能蛋白质的池中的未鉴定的单蛋白质分子,其中该小组中的亲和试剂的种数为m,并且其中m小于n。未鉴定的蛋白质可以是对应于已知蛋白质和基因序列的已知蛋白质,或者可以是没有已知蛋白质或基因序列的未知蛋白质。在未知蛋白质的情况下,该方法可鉴定未知蛋白质的特征,并由此鉴定该未知蛋白质的存在和数量,但不鉴定氨基酸序列。本

公开内容的方法可用于选择m种亲和试剂的小组,其能够鉴定选自n种可能蛋白质的池的未鉴定蛋白质。本文公开的方法还能够使用m种结合试剂独特地鉴定和定量蛋白质混合物中的n种蛋白质,并且其中通过m种结合试剂的亚组的独特结合概况鉴定每种蛋白质。此外,m可小于n的约二分之一、三分之一、四分之一、五分之一、六分之一、七分之一、十分之一、二十分之一、五十分之一或百分之一。对于进一步的实例,本公开内容可用于选择少于约100、200、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、3500或4000种亲和试剂的小组,使得该小组中的亲和试剂能够独特地鉴定至少约100、500、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、12,000、14,000、16,000、18,000、20,000、25,000、30,000、35,000、40,000、50,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、2,000,000、3,000,000、4,000,000或5,000,000种不同蛋白质中的每一种。

[0137] 本公开内容的方法可以能够鉴定蛋白质组中的大多数蛋白质。本公开内容的方法可以能够鉴定哺乳动物、鸟、鱼、两栖动物、爬行动物、脊椎动物、无脊椎动物、植物、真菌、细菌或古菌的蛋白质组中的大多数蛋白质。本公开内容的方法可以能够鉴定蛋白质组中大于约5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%的蛋白质。

[0138] 实施例

[0139] 实施例1:使用结合独特3-聚体肽的抗体进行蛋白质鉴定

[0140] 进行计算实验以确定所有表位的组在蛋白质组中的百分比覆盖度与可使用本公开内容的方法鉴定的蛋白质组的百分比之间的关系。对于该实验,选择所有3-聚体氨基酸表位的组。没有考虑蛋白质修饰。由于存在20种天然存在的氨基酸,因此所有3-聚体表位的总组是 $20 \times 20 \times 20 = 8000$ 种可能的表位。对于该模拟,将x设定为在实验中筛选的表位数,对于1至8000的每个x值,随机选择x个表位的组,并计算可鉴定的蛋白质组的百分比。图13示出了该模拟的结果。

[0141] 实施例2:使用结合独特3-聚体肽的抗体进行蛋白质鉴定

[0142] 进行进一步的计算实验以确定亲和试剂的数目对可识别性-覆盖度曲线的影响。针对一系列亲和试剂池大小计算数据系列,以显示出对于蛋白质的每一可能的覆盖度,可被鉴定的蛋白质组的百分比(y轴),并且结果显示在表1中。例如,如果这些3-聚体氨基酸表位中的20%被结合(该结合可能足以或不足以鉴定蛋白质),则具有100个氨基酸的蛋白质具有98个3-聚体氨基酸表位“着陆位点”。如图15所示,在亲和试剂池具有250种3-聚体特异性亲和试剂的情况下,如果每种蛋白质的20%的着陆位点被结合,则仅可鉴定约7%的蛋白质组。对于具有8000种亲和试剂的亲和试剂池,则在20%的着陆位点被结合时可以鉴定约98%的蛋白质组。

[0143] 表1:3-聚体解码探针数目对蛋白质组的可识别性-覆盖度曲线的影响

[0144]

	8000	7000	6000	5000	4000	3000	2000	1000	500	250
1.00%	0.1825	0.135	0.0845	0.072	0.042	0.0125	0.0035	0	0	0
2.00%	0.492	0.41	0.3515	0.26	0.156	0.0985	0.037	0.0035	0.0005	0
3.00%	0.677	0.614	0.55	0.455	0.344	0.2175	0.0985	0.015	0.0005	0
4.00%	0.786	0.745	0.676	0.604	0.472	0.334	0.176	0.029	0.003	0
5.00%	0.843	0.811	0.765	0.7005	0.61	0.4765	0.269	0.054	0.0075	0
6.00%	0.9025	0.852	0.809	0.7645	0.6815	0.569	0.3485	0.092	0.012	0.0015
7.00%	0.9005	0.877	0.8435	0.81	0.7285	0.626	0.4345	0.1395	0.022	0.0025
8.00%	0.9275	0.9025	0.8875	0.835	0.782	0.678	0.491	0.192	0.034	0.002
9.00%	0.9415	0.923	0.898	0.8725	0.814	0.728	0.5495	0.221	0.0415	0.0065
10.00%	0.9575	0.941	0.919	0.8835	0.8535	0.751	0.601	0.261	0.0715	0.007
12.00%	0.9635	0.957	0.946	0.913	0.8825	0.81	0.663	0.3445	0.0955	0.0145

[0145]

15.00%	0.978	0.969	0.962	0.9505	0.9185	0.8605	0.7675	0.443	0.1585	0.0295
17.00%	0.981	0.9765	0.9645	0.9575	0.927	0.884	0.8005	0.503	0.1915	0.0435
20.00%	0.9885	0.986	0.9725	0.9635	0.9575	0.9105	0.847	0.584	0.2525	0.0775
25.00%	0.99	0.9865	0.9785	0.9745	0.966	0.9445	0.8915	0.6955	0.357	0.1165
30.00%	0.9865	0.9895	0.985	0.9825	0.973	0.9625	0.9245	0.76	0.4355	0.1665
50.00%	0.9915	0.9915	0.9935	0.9895	0.9855	0.978	0.967	0.89	0.691	0.374

[0146] 实施例3:缀合在基底上的发光蛋白质分子

[0147] 将荧光蛋白样品藻红蛋白在4度的温育室中直接缀合至NHS-酯涂覆的盖玻片,持续4小时。然后采用具有Hamamatsu orca flash 4.0相机的Leica DMi8,使用300ms曝光对荧光蛋白样品进行成像。图16A和16B示出了捕获的所得图像(为清楚起见,将颜色颠倒)。如图16A和16B所示,每个暗点代表指示蛋白质存在的荧光信号区域。图16B是图16A的放大图。图16B中的箭头指示表示蛋白质的信号,其可与背景噪声清楚区分。

[0148] 使第二蛋白质样品绿色荧光蛋白变性,并在4度的温育室中直接缀合至NHS-酯涂覆的盖玻片,持续4小时。初始成像显示没有基线残留荧光,表明绿色荧光蛋白完全变性。然后将蛋白质与附有Alexa-Fluor 647的抗肽抗体一起温育。然后用0.1% Tween-20冲洗抗肽抗体。然后使用TIRF在带有Andor NEO sCMOS相机的Nikon Eclipse Ti上对其进行成像。图17示出了捕获的所得图像(为了清楚起见,将颜色颠倒)。

[0149] 实施例4:蛋白质的鉴定

[0150] 图18中描绘了四种可能的蛋白质即绿色荧光蛋白、RNASE1、LTF和GSTM1的蛋白质组。在该实施例中,来自该蛋白质组的未知蛋白质的单个分子缀合至基底上的位置。通过九种不同亲和试剂的小组依次询问未知蛋白质。九种不同亲和试剂中的每一种识别不同的氨基酸三聚体[AAA、AAC、AAD、AEV、GDG、QSA、LAD、TRK、DGD],并且每一种都用荧光染料进行标记。确定该未知蛋白质被亲和试剂DGD、AEV、LAD、GDG和QSA结合。对该蛋白质组的四种蛋白质的序列的分析表明,只有GFP含有这些三氨基酸基序中的全部五种,这些基序在图18的序列中加下划线。因此,确定未知蛋白质的单个分子是GFP蛋白质。

[0151] 计算机控制系统

[0152] 本公开内容提供了被编程用于实现本公开内容的方法的计算机控制系统。图14示出了计算机系统1401,其被编程或以其他方式配置用于表征和鉴定生物聚合物,如蛋白质。计算机系统1401可调节以下各个方面:评估和分析本公开内容的样品,例如,观察基底的独特空间地址处的信号;基于观察到的信号确定在独特空间地址处与生物聚合物部分连接的

可识别标签的存在；相对于生物聚合物序列数据库评估确定的可识别标签，以确定生物聚合物部分的特征。计算机系统1401可以是用户的电子设备或相对于该电子设备远程定位的计算机系统。电子设备可以是移动电子设备。

[0153] 计算机系统1401包括中央处理单元(CPU,本文中也称为“处理器”和“计算机处理器”)1405,其可以是单核或多核处理器,或者用于平行处理的多个处理器。计算机系统1401还包括存储器或存储器位置1410(例如,随机存取存储器、只读存储器、闪速存储器)、电子存储单元1415(例如,硬盘)、用于与一个或多个其他系统通信的通信接口1420(例如,网络适配器),和外围设备1425,如高速缓冲存储器、其他存储器、数据存储和/或电子显示适配器。存储器1410、存储单元1415、接口1420和外围设备1425通过诸如主板等通信总线(实线)与CPU 1405通信。存储单元1415可以是用于存储数据的数据存储单元(或数据存储库)。计算机系统1401可借助于通信接口1420可操作地耦合到计算机网络(“网络”)1430。网络1430可以是因特网、互联网和/或外联网,或与因特网通信的内联网和/或外联网。在一些情况下,网络1430是电信和/或数据网络。网络1430可包括一个或多个计算机服务器,其可以实现分布式计算,如云计算。在一些情况下借助于计算机系统1401,网络1430可以实现对等网络,这可以使耦合到计算机系统1401的设备能够充当客户端或服务器。

[0154] CPU 1405可执行一系列机器可读指令,该机器可读指令可体现在程序或软件中。该指令可存储在存储器位置,如存储器1410中。该指令可以指向CPU 1405,其可随后编程或以其他方式配置CPU 1405以实现本公开内容的方法。由CPU 1405执行的操作的实例可包括提取、解码、执行和回写。

[0155] CPU 1405可以是电路如集成电路的一部分。系统1401的一个或多个其他组件可包括在电路中。在一些情况下,该电路是专用集成电路(ASIC)。

[0156] 存储单元1415可以存储文件,如驱动程序、文库和保存的程序。存储单元1415可以存储用户数据,例如用户偏好和用户程序。在一些情况下,计算机系统1401可包括一个或多个附加数据存储单元,该附加数据存储单元在计算机系统1401外部,如位于通过内联网或因特网与计算机系统1401通信的远程服务器上。

[0157] 计算机系统1401可通过网络1430与一个或多个远程计算机系统通信。例如,计算机系统1401可与用户的远程计算机系统通信。远程计算机系统的实例包括个人计算机(例如,便携式PC)、平板或平板型PC(例如, **Apple®** iPad、**Samsung®** Galaxy Tab)、电话、智能电话(例如, **Apple®** 电话、支持Android的设备、**Blackberry®**),或个人数字助理。用户可经由网络1430访问计算机系统1401。

[0158] 如本文所述的方法可通过存储在计算机系统1401的电子存储位置上,例如,存储在存储器1410或电子存储单元1415上的机器(例如,计算机处理器)可执行代码来实现。可以以软件的形式提供机器可执行代码或机器可读代码。在使用期间,代码可以由处理器1405执行。在一些情况下,可以从存储单元1415检索代码并将其存储在存储器1410上以备处理器1405获取。在一些情况下,可以排除电子存储单元1415,并且机器可执行指令存储在存储器1410中。

[0159] 代码可被预编译和配置成与具有适于执行该代码的处理器一起使用,或者可以在运行时期中编译。可以以编程语言提供该代码,可选择该编程语言以使得代码能够

以预编译或实时编译的方式执行。

[0160] 本文提供的系统和方法的各方面,如计算机系统501,可在编程中体现。所述技术的各个方面可被认为是“产品”或“制品”,其通常为在一类机器可读介质中携带或体现的机器(或处理器)可执行代码和/或相关数据的形式。机器可执行代码可存储在电子存储单元,如存储器(例如,只读存储器、随机存取存储器、闪速存储器)或硬盘上。“存储”型介质可包括计算机的任何或全部有形存储器、处理器等,或其相关模块,如各种半导体存储器、磁带驱动器、磁盘驱动器等,它们可以在任何时间为软件编程提供非暂时性存储。软件的全部或部分有时可以通过因特网或各种其他电信网络进行通信。这种通信例如可以使得软件能够从一个计算机或处理器加载到另一个计算机或处理器中,例如,从管理服务器或主计算机加载到应用服务器的计算机平台中。因此,可以承载软件元素的另一种类型的介质包括光波、电波和电磁波,诸如跨本地设备之间的物理接口、通过有线和光学陆线网络以及通过各种空中链路而使用的光波、电波和电磁波。携带此类波的物理元件,如有线或无线链路、光学链路等,也可被认为是承载软件的介质。如本文所用,除非限于非暂时性有形“存储”介质,否则诸如计算机或机器“可读介质”等术语是指参与向处理器提供指令以供执行的任何介质。

[0161] 因此,机器可读介质如计算机可执行代码可采取多种形式,包括但不限于有形存储介质、载波介质或物理传输介质。非易失性存储介质包括例如光盘或磁盘,诸如任何计算机中的任何存储设备等,所述存储装置诸如可用于实现附图中所示的数据库等。易失性存储介质包括动态存储器,如这样的计算机平台的主存储器。有形传输介质包括同轴电缆;铜线和光纤,包括构成计算机系统内的总线的导线。载波传输介质可采取电信号或电磁信号或者声波或光波的形式,诸如在射频(RF)和红外(IR)数据通信期间生成的那些。因此,计算机可读介质的常见形式包括例如:软盘、柔性盘、硬盘、磁带、任何其他磁性介质、CD-ROM、DVD或DVD-ROM、任何其他光学介质、穿孔卡片纸带、任何其他具有孔图案的物理存储介质、RAM、ROM、PROM和EPROM、FLASH-EPROM、任何其他存储器芯片或匣盒、传送数据或指令的载波、传送这样的载波的电缆或链路,或者计算机可从中读取编程代码和/或数据的任何其他介质。许多这些形式的计算机可读介质可涉及将一个或多个指令的一个或多个序列载送到处理器以供执行。

[0162] 计算机系统1401可包括电子显示器1435或与电子显示器1435通信,电子显示器1435包含用户界面(UI) 1440。UI的实例包括但不限于图形用户界面(GUI)和基于网络的用户界面。

[0163] 可通过一种或多种算法来实现本公开内容的方法和系统。算法可在由中央处理单元1405执行时通过软件实现。该算法可以例如确定生物聚合物部分如蛋白质部分的特征和/或身份。例如,算法可用于确定候选生物聚合物部分如候选蛋白质部分的最可能的身份。

[0164] 在一些实施方案中,识别存在于许多不同蛋白质中的短表位的适体或肽聚体(peptamer)可被称为数字适体或数字肽聚体。本发明的一个方面提供了数字适体或数字肽聚体的组,其中该组包含至少约15种数字适体或数字肽聚体,其中15种数字适体或数字肽聚体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且其中每种数字适体或数字肽聚体识别包含该数字适体或数字肽聚体所结合的同表位的

多种截然不同的蛋白质。在一些实施方案中,数字适体或数字肽聚体的组包含100种数字适体或数字肽聚体,其结合由3个连续氨基酸组成的表位。在一些实施方案中,数字适体或数字肽聚体的组进一步包含100种数字适体,其结合由4个连续氨基酸组成的表位。在一些实施方案中,数字适体或数字肽聚体的组进一步包含100种数字适体或数字肽聚体,其结合由5个连续氨基酸组成的表位。在一些情况下,数字亲和试剂可以是抗体、适体、肽聚体、肽或Fab片段。

[0165] 在一些实施方案中,数字适体组包含至少约20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000种数字适体。在一些实施方案中,数字适体组包含至少1000种数字适体,其结合由4个连续氨基酸组成的表位。在一些实施方案中,数字适体组进一步包含至少100种数字适体,其结合由5个连续氨基酸组成的表位。数字适体组进一步包含至少100种数字适体,其结合由3个连续氨基酸组成的表位。在一些实施方案中,将数字适体组固定在表面上。在一些实施方案中,该表面是阵列。

[0166] 在另一方面,本发明提供了用于生成包含多种不同蛋白质的样品的蛋白质结合概况的方法,所述方法包括:在允许结合的条件下使所述样品与数字适体组接触,其中该数字适体组包含至少约15种数字适体,其中15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的同表位的多种截然不同的蛋白质;任选地去除未结合的蛋白质;以及检测蛋白质与所述数字适体的结合,从而生成样品的蛋白质结合概况。

[0167] 在一些实施方案中,所述方法进一步包括在允许结合的条件下使样品与该数字适体组接触的步骤(a)之前的用蛋白质切割剂处理样品的步骤。

[0168] 在另一方面,本发明包括两种或更多种不同样品的蛋白质结合概况的文库,所述样品的每一种包含多种蛋白质,所述方法包括:在允许结合的条件下使样品与一组数字适体接触,其中该数字适体组包含至少约15种数字适体,其中15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的同表位的多个截然不同的蛋白质;任选地去除未结合的蛋白质;通过检测蛋白质与数字适体的结合而生成待测样品的蛋白质结合概况,由此生成蛋白质结合概况;以及用至少两种样品重复上述步骤。

[0169] 在一些实施方案中,所述方法进一步包括在允许结合的条件下使样品与该数字适体组接触的步骤之前的用蛋白质切割剂处理样品的步骤。

[0170] 在另一方面,本发明包括用于表征测试样品的方法,其包括:在允许结合的条件下使测试样品与一组数字适体接触,其中该数字适体组包含至少约15种数字适体,其中15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的同表位的多种截然不同的蛋白质;任选地去除未结合的蛋白质;通过检测蛋白质与数字适体的结合而生成所述测试样品的蛋白质结合概况;以及将测试样品的生成的蛋白质结合概况与参考样品的蛋白质结合概况进行比较,以表征测试样品。

[0171] 在另一方面,本发明包括用于确定测试样品中是否细菌、病毒或细胞的方法,所述方法包括:在允许结合的条件下使测试样品与一组数字适体接触,其中该数字适体组包含至少约15种数字适体,其中15种数字适体中的每一种已被表征为与由3或4或5个连续氨基

酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的同表位的多种截然不同的蛋白质;任选地去除未结合的蛋白质;通过检测蛋白质与数字适体的结合而生成测试样品的蛋白质结合概况,由此生成蛋白质结合概况;以及将测试样品的蛋白质结合概况与参考样品的蛋白质结合概况进行比较,由此通过比较确定测试样品中是否存在细菌、病毒或细胞。

[0172] 在另一个方面,本发明包括用于鉴定样品中的测试蛋白质的方法,所述方法包括:使包含或疑似包含测试蛋白质的样品与包含至少约15种数字适体的一组数字适体接触,其中15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的同表位的多种截然不同的蛋白质;以及通过检测测试蛋白质与该数字适体组的结合来确定测试蛋白质的身份,其中至少约六种数字适体结合测试蛋白质;并且其中结合的存在表明在测试蛋白质中存在至少约六种表位,其中使用该至少约六种表位的身份来鉴定测试蛋白质。

[0173] 虽然本文中已经示出并描述了本发明的优选实施方案,但对于本领域技术人员显而易见的是,这些实施方案仅以实例的方式提供。并非旨在通过本说明书中提供的具体实例来限制本发明。虽然已经参考前述说明书描述了本发明,但是本文实施方案的描述和说明并不意味着以限制性的方式来理解。本领域技术人员在不脱离本发明的情况下现将会想到多种变化、改变和替代。此外,应当理解,本发明的所有方面都不限于本文阐述的具体描绘、配置或相对比例,其取决于多种条件和变量。应当理解,本文中描述的本发明实施方案的各种替代方案可用于实施本发明。因此,预期本发明还应涵盖任何这样的替代、修改、变化或等同物。旨在以下述权利要求限定本发明的范围,并由此涵盖这些权利要求范围内的方法和结构及其等同物。

[0174] 尽管有所附权利要求,但本文所述的公开内容还由以下条款限定:

[0175] 1. 数字适体组,其中所述组包含至少约15种数字适体,其中所述15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且其中每种数字适体识别包含该数字适体所结合的同表位的多种截然不同的蛋白质。

[0176] 2. 根据条款1所述的数字适体组,其中所述组包含100种数字适体,所述数字适体结合由3个连续氨基酸组成的表位。

[0177] 3. 根据条款1所述的数字适体组,其中所述组进一步包含100种数字适体,所述数字适体结合由4个连续氨基酸组成的表位。

[0178] 4. 根据条款3所述的数字适体组,其中所述组进一步包含100种数字适体,所述数字适体结合由5个连续氨基酸组成的表位。

[0179] 5. 根据条款1所述的数字适体组,其中所述组包含至少约20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000种数字适体。

[0180] 6. 根据条款1所述的数字适体组,其中所述组包含至少1000种数字适体,所述数字适体结合由4个连续氨基酸组成的表位。

[0181] 7. 根据条款6所述的数字适体组,其中所述组进一步包含至少100种数字适体,所述数字适体结合由5个连续氨基酸组成的表位。

[0182] 8. 根据条款7所述的数字适体组,其中所述组进一步包含至少100种数字适体,所述数字适体结合由3个连续氨基酸组成的表位。

[0183] 9. 根据条款1-8中任一项所述的数字适体组,其中所述数字适体固定在表面上。

[0184] 10. 根据条款9所述的数字适体组,其中所述表面是阵列。

[0185] 11. 一种用于生成包含多种不同蛋白质的样品的蛋白质结合概况的方法,所述方法包括:

[0186] a) 在允许结合的条件下使所述样品与数字适体组接触,其中所述数字适体组包含至少约15种数字适体,其中所述15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的同一表位的多种截然不同的蛋白质;

[0187] b) 任选地去除未结合的蛋白质;以及

[0188] c) 检测蛋白质与所述数字适体的结合,从而生成所述样品的蛋白质结合概况。

[0189] 12. 根据条款11所述的方法,其中所述方法进一步包括在允许结合的条件下使所述样品与所述数字适体组接触的步骤(a)之前的用蛋白质切割剂处理所述样品的步骤。

[0190] 13. 一种用于生成两种或更多种不同样品的蛋白质结合概况的文库的方法,所述样品中的每一种包含多种蛋白质,所述方法包括:

[0191] a) 在允许结合的条件下使样品与数字适体组接触,其中所述数字适体组包含至少约15种数字适体,其中所述15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的同一表位的多种截然不同的蛋白质;

[0192] b) 任选地去除未结合的蛋白质;

[0193] c) 通过检测蛋白质与所述数字适体的结合而生成待测样品的蛋白质结合概况,由此生成蛋白质结合概况;以及

[0194] d) 用至少两种样品重复步骤(a)到(c)。

[0195] 14. 根据条款13所述的方法,其中所述方法进一步包括在允许结合的条件下使所述样品与所述数字适体组接触的步骤(a)之前的用蛋白质切割剂处理所述样品的步骤。

[0196] 15. 一种蛋白质结合概况的文库,其中使用条款13所述的方法制备所述文库。

[0197] 16. 一种用于表征测试样品的方法,其包括:

[0198] a) 在允许结合的条件下使所述测试样品与数字适体组接触,其中所述数字适体组包含至少约15种数字适体,其中所述15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的同一表位的多种截然不同的蛋白质;

[0199] b) 任选地去除未结合的蛋白质;

[0200] c) 通过检测蛋白质与所述数字适体的结合而生成所述测试样品的蛋白质结合概况;以及

[0201] d) 将所述测试样品的生成的蛋白质结合概况与参考样品的蛋白质结合概况进行比较,以表征所述测试样品。

[0202] 17. 一种用于确定测试样品中是否存在细菌、病毒或细胞的方法,所述方法包括:

[0203] a) 在允许结合的条件下使所述测试样品与数字适体组接触,其中所述数字适体组包含至少约15种数字适体,其中所述15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合,并且每种数字适体识别包含该数字适体所结合的

同一表位的多种截然不同的蛋白质；

[0204] b) 任选地去除未结合的蛋白质；

[0205] c) 通过检测蛋白质与所述数字适体的结合而生成所述测试样品的蛋白质结合概况，由此生成蛋白质结合概况；以及

[0206] d) 将所述测试样品的蛋白质结合概况与参考样品的蛋白质结合概况进行比较，由此通过所述比较确定所述测试样品中是否存在细菌、病毒或细胞。

[0207] 18. 一种用于鉴定样品中的测试蛋白质的方法，所述方法包括：

[0208] a) 使包含或疑似包含所述测试蛋白质的样品与包含至少约15种数字适体的数字适体组接触，其中所述15种数字适体中的每一种已被表征为与由3或4或5个连续氨基酸组成的不同表位特异性结合，并且每种数字适体识别包含该数字适体所结合的同一样位的多种截然不同的蛋白质；以及

[0209] b) 通过检测所述测试蛋白质与所述数字适体组的结合来确定所述测试蛋白质的身份，其中至少约六种数字适体结合所述测试蛋白质；并且其中结合的存在表明在所述测试蛋白质中存在至少约六种表位，其中使用所述至少约六种表位的身份来鉴定所述测试蛋白质。

[0210] 19. 一种确定蛋白质特征的方法，所述方法包括：

[0211] 获得基底，其中一种或多种蛋白质的部分缀合至所述基底，使得每个单独的（在分子水平上）蛋白质部分具有独特的、光学可分辨的空间地址；

[0212] 将含有第一至第n组（按顺序）的一种或多种亲和试剂的流体施加至所述基底，其中所述一种或多种亲和试剂中的每一种对所述一种或多种蛋白质的部分的一个表位（连续或非连续氨基酸序列）具有特异性，并且其中所述第一至第n组的一种或多种亲和试剂中的每种亲和试剂与可识别标签连接；

[0213] 在所述第一组一种或多种亲和试剂和随后直至第n组一种或多种亲和试剂各自施加至所述基底后，执行以下步骤：

[0214] 观察所述可识别标签；

[0215] 鉴定所述基底上具有一个或多个观察到的信号的一个或多个独特空间地址；

[0216] 确定具有鉴定的独特空间地址的所述一种或多种蛋白质的每个部分含有与所述一个或多个观察到的信号相关的一个或多个表位；以及

[0217] 基于所述一个或多个表位确定每个蛋白质部分的特征。

[0218] 20. 一种确定蛋白质特征的方法，所述方法包括：

[0219] 获得基底，其中一种或多种蛋白质的部分缀合至所述基底，使得所述基底具有多个位置，每个位置包含单个蛋白质或蛋白质池，所述蛋白质池中至少60%的蛋白质共享相同的氨基酸序列；

[0220] 将含有第一至第n组（按顺序）的一种或多种亲和试剂的流体施加至所述基底，其中所述一种或多种亲和试剂中的每一种对所述一种或多种蛋白质的部分的一个表位（连续或非连续氨基酸序列）具有特异性，并且其中所述第一至第n组的一种或多种亲和试剂中的每种亲和试剂与可识别标签连接；

[0221] 在所述第一组一种或多种亲和试剂和随后直至第n组一种或多种亲和试剂各自施加至所述基底后，执行以下步骤：

- [0222] 观察所述可识别标签；
- [0223] 鉴定所述基底上具有一个或多个观察到的信号的一个或多个独特空间地址；
- [0224] 确定具有鉴定的独特空间地址的所述一种或多种蛋白质的每个部分含有与所述一个或多个观察到的信号相关的一个或多个表位；以及
- [0225] 基于所述一个或多个表位确定每个蛋白质部分的特征。
- [0226] 21. 根据条款19或20所述的方法，其中所述方法可用于鉴定至少400种不同的蛋白质，其比依赖于来自质谱仪的数据的蛋白质鉴定技术快至少10%。
- [0227] 22. 根据条款21所述的方法，其中所述方法以至少50%的准确度鉴定所述至少400种不同的蛋白质。
- [0228] 23. 根据条款22所述的方法，其中所述方法将特定蛋白质鉴定为特定蛋白质家族的成员，而不依赖于所述方法是否以超过10%的阈值置信度鉴定所述特定蛋白质本身。
- [0229] 24. 根据条款19或20所述的方法，其中基于所述蛋白质的大小在所述基底上分离所述一种或多种蛋白质的部分。
- [0230] 25. 根据条款19或20所述的方法，其中基于所述蛋白质的电荷在所述基底上分离所述一种或多种蛋白质的部分。
- [0231] 26. 根据条款19或20所述的方法，其中所述基底包含微孔。
- [0232] 27. 根据条款19或20所述的方法，其中所述基底包括不同大小的微孔。
- [0233] 28. 根据条款19或20所述的方法，其中所述蛋白质通过生物素附接而附接至所述基底。
- [0234] 29. 根据条款19或20所述的方法，其中所述蛋白质通过核酸附接至所述基底。
- [0235] 30. 根据条款29所述的方法，其中所述蛋白质通过核酸纳米球附接至所述基底。
- [0236] 31. 根据条款19或20所述的方法，其中所述蛋白质通过纳米珠附接至所述基底。
- [0237] 32. 根据条款19或20所述的方法，其中获得其中结合有一种或多种蛋白质的部分的基底包括获得具有官能团的有序阵列的基底，以及施加蛋白质样品，使得每个官能团与来自所述样品的不超过一种蛋白质分子缀合。
- [0238] 33. 根据条款32所述的方法，其中获得具有官能团的有序阵列的基底包括使用选自光刻、蘸笔纳米刻蚀、纳米压印刻蚀、纳米球刻蚀、热扫描探针刻蚀、局部氧化纳米刻蚀、分子自组装、模板刻蚀和电子束刻蚀的方法。
- [0239] 34. 根据条款32所述的方法，其中每个官能团与每个其他官能团相距至少约300nm定位。
- [0240] 35. 根据条款19或20所述的方法，其中所述基底包含不同大小微孔的有序阵列。
- [0241] 36. 根据条款19或20所述的方法，其中获得所述基底包括将蛋白质的第一样品缀合至所述基底，使用蛋白质染料检测具有来自所述第一样品的结合蛋白质的每个位置，使第二样品缀合，以及使用蛋白质染料检测具有来自所述第二样品的结合蛋白质的每个位置。
- [0242] 37. 根据条款19或20所述的方法，其中获得所述基底包括将蛋白质的第一样品缀合至所述基底，使用蛋白质染料检测具有来自所述第一样品的结合蛋白质的每个位置，从结合蛋白质的数目确定所述基底上未被蛋白质结合的官能团比例。
- [0243] 38. 根据条款19或20所述的方法，其中亲和试剂可包含组分的池，所述组分结合具

有不同侧翼序列的相同核心序列,使得至少一种组分对于在不考虑侧翼序列的情况下结合所述核心序列的任何实例具有高于阈值的结合亲和力。

	<110> 点燃生物科学有限公司
	<120> 测定蛋白质的方法
	<130> 51612-701.601
	<140> PCT/US2017/064322
	<141> 2017-12-01
	<150> 62/500,455
	<151> 2017-05-02
	<150> 62/429,063
	<151> 2016-12-01
	<160> 15
	<170> PatentIn version 3.5
[0001]	<210> 1
	<211> 30
	<212> PRT
	<213> 人工序列
	<220>
	<223> 人工序列的描述：合成多肽
	<400> 1
	Met Lys Leu Thr Leu Lys Asn Leu Ser Met Ala Ile Met Met Ser Ile
	1 5 10 15
	Val Met Gly Ser Ser Ala Met Ala Ala Asp Ser Asn Glu Lys
	20 25 30
	<210> 2
	<211> 8
	<212> PRT
	<213> 人工序列
	<220>
	<223> 人工序列的描述：合成肽

[0002]

<400> 2

Gly Tyr Leu Pro Glu His Thr Leu

1 5

<210> 3

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 3

Ala Asp Tyr Leu Glu Gln Asp

1 5

<210> 4

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 4

Thr Tyr Pro Gly Arg

1 5

<210> 5

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 5

Leu His Asp His Tyr Leu Asp

1 5

<210> 6
<211> 7
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 6
Asp Arg Ala Arg Lys Asp Gly
1 5

<210> 7
<211> 5
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

[0003] <400> 7
Thr Phe Glu Glu Glu
1 5

<210> 8
<211> 8
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 8
Asp Glu Ile Lys Ser Leu Lys Phe
1 5

<210> 9
<211> 10
<212> PRT
<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 9

Gln Thr Tyr Pro Gly Arg Phe Pro Met Gly
1 5 10

<210> 10

<211> 18

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 10

His Thr Phe Glu Glu Glu Ile Glu Phe Val Gln Gly Leu Asn His Ser
1 5 10 15

Thr Gly

[0004]

<210> 11

<211> 40

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成多肽

<400> 11

Asn Ile Gly Ile Tyr Pro Glu Ile Lys Ala Pro Trp Phe His Gln Glu
1 5 10 15

Gly Lys Asp Ile Ala Ala Lys Thr Leu Glu Val Leu Lys Lys Tyr Gly
20 25 30

Tyr Thr Gly Lys Asp Asp Lys Val
35 40

<210> 12

<211> 244

<212> PRT

<213> 维多利亚多管水母

<400> 12

Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val
1 5 10 15

Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu
20 25 30

Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
35 40 45

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe
50 55 60

[0005]

Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln
65 70 75 80

His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg
85 90 95

Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val
100 105 110

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile
115 120 125

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn
130 135 140

Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly
145 150 155 160

Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val
 165 170 175

Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro
 180 185 190

Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser
 195 200 205

Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val
 210 215 220

Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys His His
 225 230 235 240

His His His His

[0006]

<210> 13
 <211> 150
 <212> PRT
 <213> 家牛

<400> 13
 Met Ala Leu Lys Ser Leu Val Leu Leu Ser Leu Leu Val Leu Val Leu
 1 5 10 15

Leu Leu Val Arg Val Gln Pro Ser Leu Gly Lys Glu Thr Ala Ala Ala
 20 25 30

Lys Phe Glu Arg Gln His Met Asp Ser Ser Thr Ser Ala Ala Ser Ser
 35 40 45

Ser Asn Tyr Cys Asn Gln Met Met Lys Ser Arg Asn Leu Thr Lys Asp
 50 55 60

Arg Cys Lys Pro Val Asn Thr Phe Val His Glu Ser Leu Ala Asp Val
65 70 75 80

Gln Ala Val Cys Ser Gln Lys Asn Val Ala Cys Lys Asn Gly Gln Thr
85 90 95

Asn Cys Tyr Gln Ser Tyr Ser Thr Met Ser Ile Thr Asp Cys Arg Glu
100 105 110

Thr Gly Ser Ser Lys Tyr Pro Asn Cys Ala Tyr Lys Thr Thr Gln Ala
115 120 125

Asn Lys His Ile Ile Val Ala Cys Glu Gly Asn Pro Tyr Val Pro Val
130 135 140

[0007] His Phe Asp Ala Ser Val
145 150

<210> 14
<211> 710
<212> PRT
<213> 智人

<400> 14
Met Lys Leu Val Phe Leu Val Leu Leu Phe Leu Gly Ala Leu Gly Leu
1 5 10 15

Cys Leu Ala Gly Arg Arg Arg Ser Val Gln Trp Cys Ala Val Ser Gln
20 25 30

Pro Glu Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Lys Val
35 40 45

Arg Gly Pro Pro Val Ser Cys Ile Lys Arg Asp Ser Pro Ile Gln Cys

50

55

60

Ile Gln Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr Leu Asp Gly
65 70 75 80

Gly Phe Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro Val
85 90 95

Ala Ala Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr Tyr
100 105 110

Ala Val Ala Val Val Lys Lys Gly Gly Ser Phe Gln Leu Asn Glu Leu
115 120 125

Gln Gly Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly Trp
130 135 140

[0008]

Asn Val Pro Ile Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly Pro
145 150 155 160

Pro Glu Pro Ile Glu Ala Ala Val Ala Arg Phe Phe Ser Ala Ser Cys
165 170 175

Val Pro Gly Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu Cys
180 185 190

Ala Gly Thr Gly Glu Asn Lys Cys Ala Phe Ser Ser Gln Glu Pro Tyr
195 200 205

Phe Ser Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly Asp
210 215 220

Val Ala Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu Ser Asp Glu
225 230 235 240

Ala Glu Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg Lys
245 250 255

Pro Val Asp Lys Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser His
260 265 270

Ala Val Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp Asn
275 280 285

Leu Leu Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro Lys
290 295 300

Phe Gln Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu Phe Lys
305 310 315 320

[0009] Asp Ser Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser Gly
325 330 335

Leu Tyr Leu Gly Ser Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg Lys
340 345 350

Ser Glu Glu Glu Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys Ala
355 360 365

Val Gly Glu Gln Glu Leu Arg Lys Cys Asn Gln Trp Ser Gly Leu Ser
370 375 380

Glu Gly Ser Val Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys Ile
385 390 395 400

Ala Leu Val Leu Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly
405 410 415

Tyr Val Tyr Thr Ala Gly Lys Cys Gly Leu Val Pro Val Leu Ala Glu
420 425 430

Asn Tyr Lys Ser Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val Asp
435 440 445

Arg Pro Val Glu Gly Tyr Leu Ala Val Ala Val Val Arg Arg Ser Asp
450 455 460

Thr Ser Leu Thr Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His Thr
465 470 475 480

Ala Val Asp Arg Thr Ala Gly Trp Asn Ile Pro Met Gly Leu Leu Phe
485 490 495

[0010]

Asn Gln Thr Gly Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser Cys
500 505 510

Ala Pro Gly Ser Asp Pro Arg Ser Asn Leu Cys Ala Leu Cys Ile Gly
515 520 525

Asp Glu Gln Gly Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg Tyr
530 535 540

Tyr Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly Asp
545 550 555 560

Val Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn
565 570 575

Asn Asn Glu Ala Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu
580 585 590

Leu Cys Leu Asp Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser Cys
595 600 605

His Leu Ala Met Ala Pro Asn His Ala Val Val Ser Arg Met Asp Lys
610 615 620

Val Glu Arg Leu Lys Gln Val Leu Leu His Gln Gln Ala Lys Phe Gly
625 630 635 640

Arg Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu
645 650 655

Thr Lys Asn Leu Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg Leu
660 665 670

His Gly Lys Thr Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val Ala
675 680 685

[0011]

Gly Ile Thr Asn Leu Lys Lys Cys Ser Thr Ser Pro Leu Leu Glu Ala
690 695 700

Cys Glu Phe Leu Arg Lys
705 710

<210> 15

<211> 243

<212> PRT

<213> 日本血吸虫

<400> 15

Ser Thr Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val
1 5 10 15

Gln Pro Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu
20 25 30

His Leu Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe
 35 40 45

Glu Leu Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp
 50 55 60

Val Lys Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys
 65 70 75 80

His Asn Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met
 85 90 95

Leu Glu Gly Ala Val Leu Asp Ile Arg Tyr Cys Val Ser Arg Ile Ala
 100 105 110

[0012]

Tyr Ser Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu
 115 120 125

Pro Glu Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr
 130 135 140

Leu Asn Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala
 145 150 155 160

Leu Asp Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro
 165 170 175

Lys Leu Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp
 180 185 190

Lys Tyr Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp
 195 200 205

Gln	Ala	Thr	Phe	Gly	Gly	Gly	Asp	His	Pro	Pro	Lys	Ser	Asp	Leu	Val
210							215					220			

[0013]	Pro	Arg	Gly	Ser	Pro	Leu	Glu	Val	Leu	Phe	Gln	Gly	Pro	Cys	Gly	Ser
	225					230					235				240	

Tyr Gly Ser

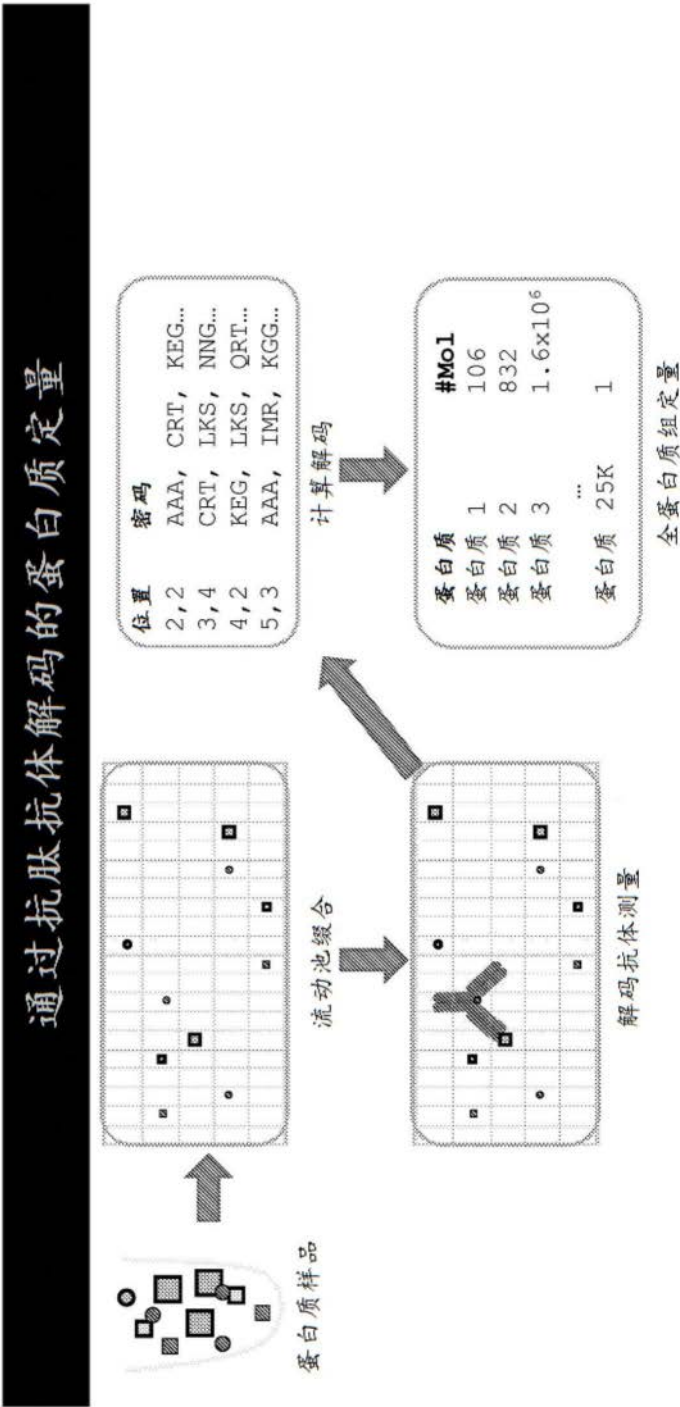


图1

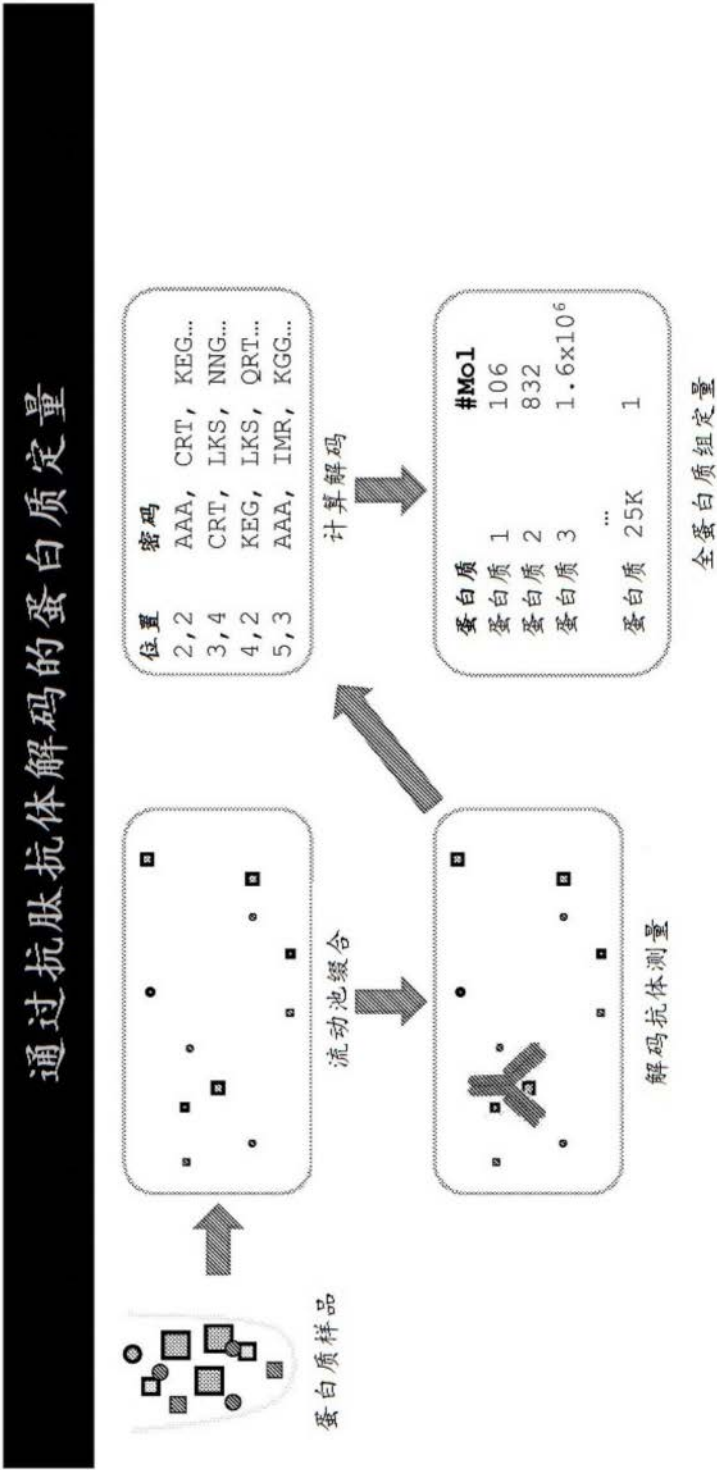


图2

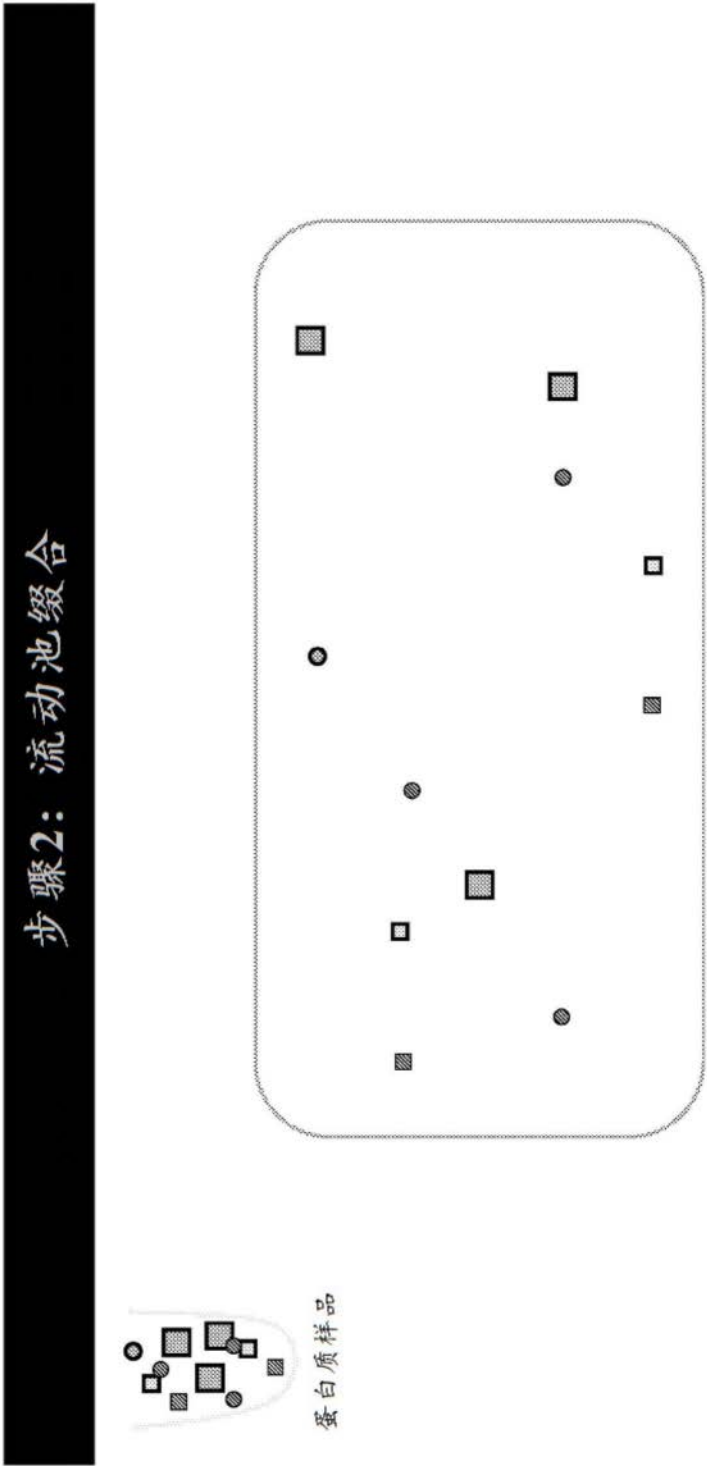


图3

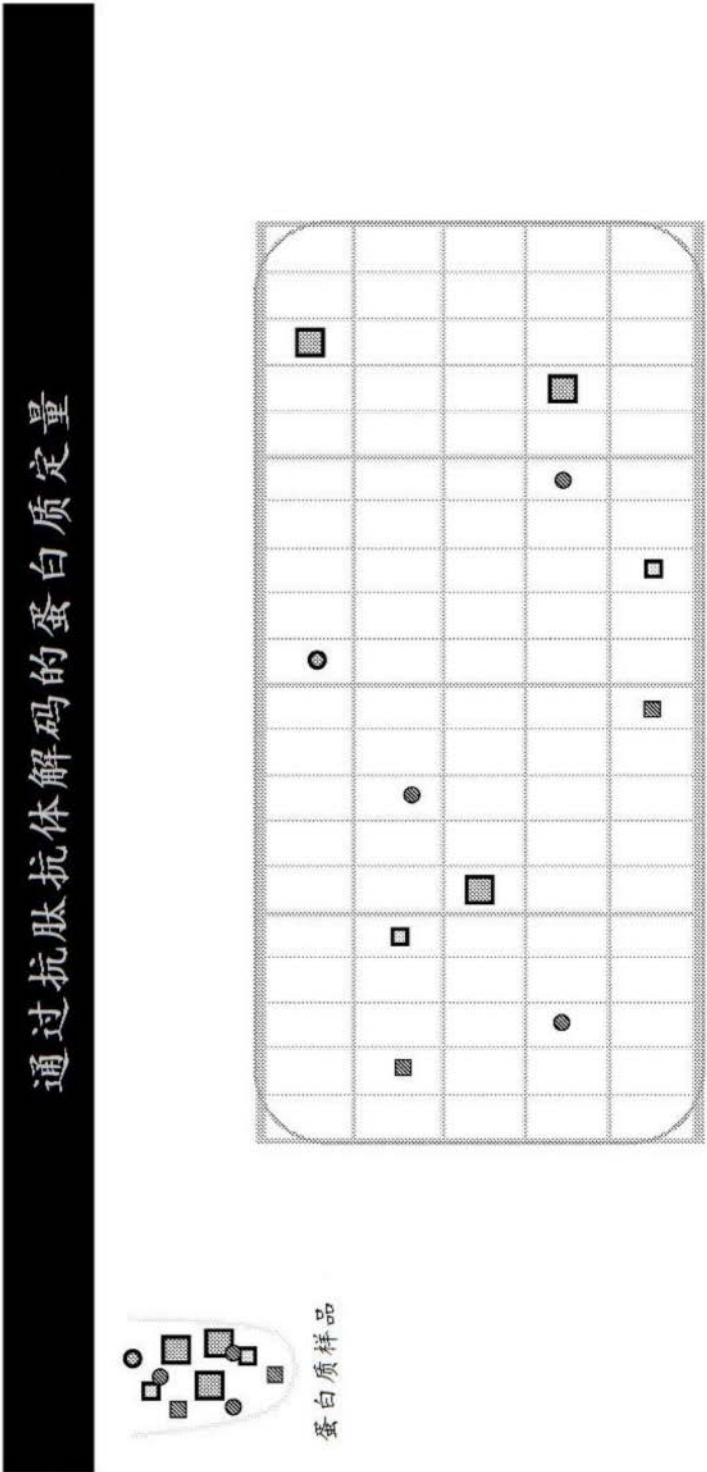


图4



图5

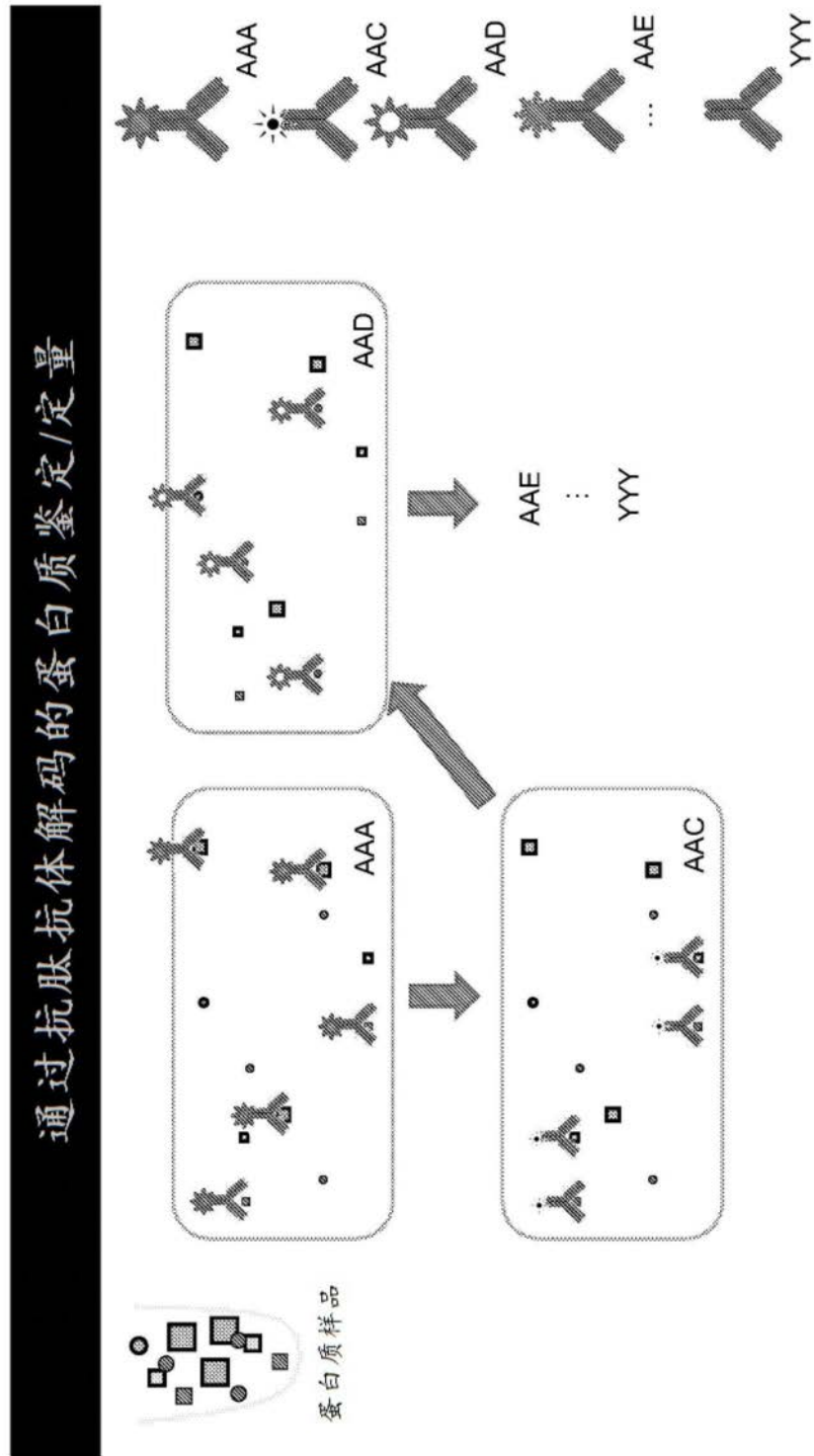


图6

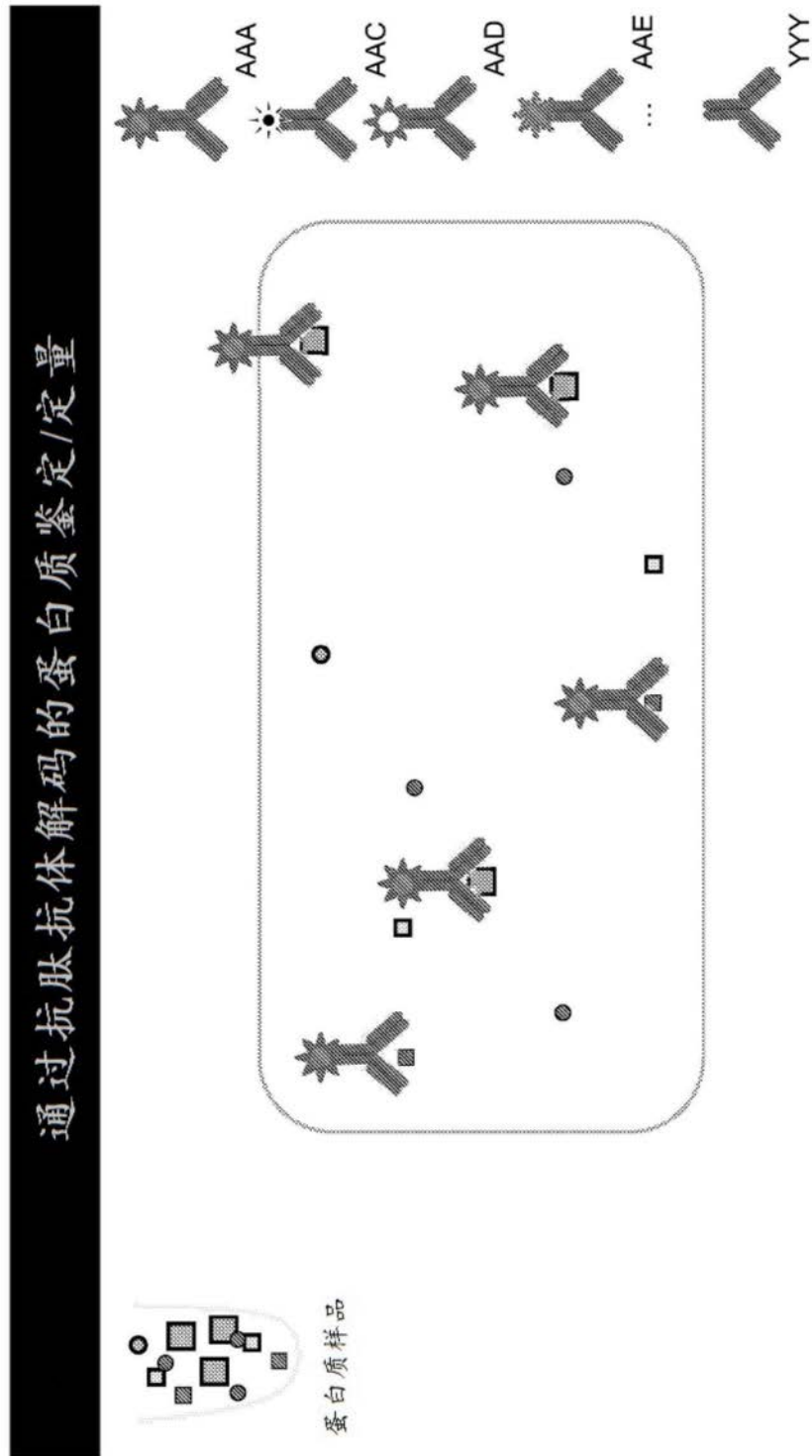


图7

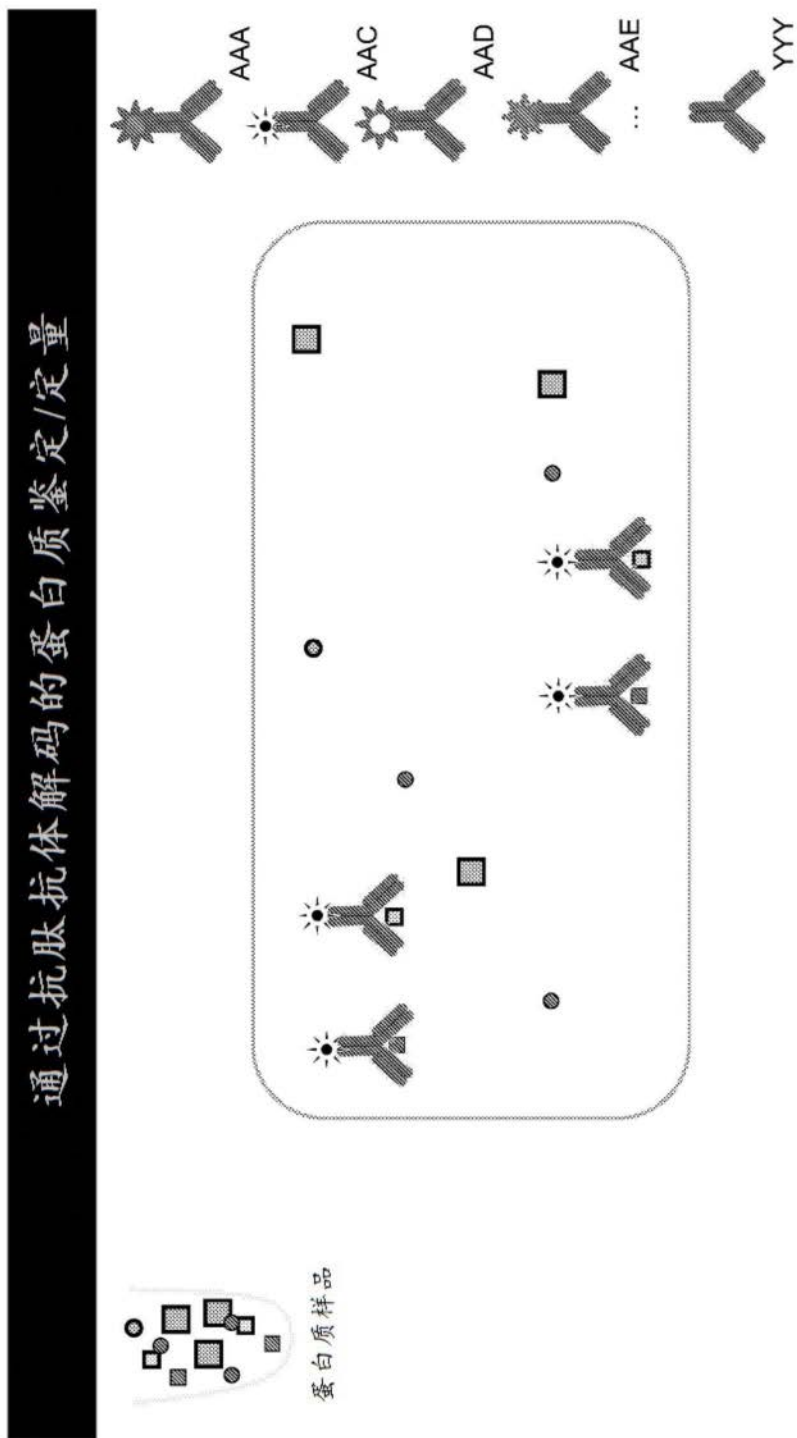


图8

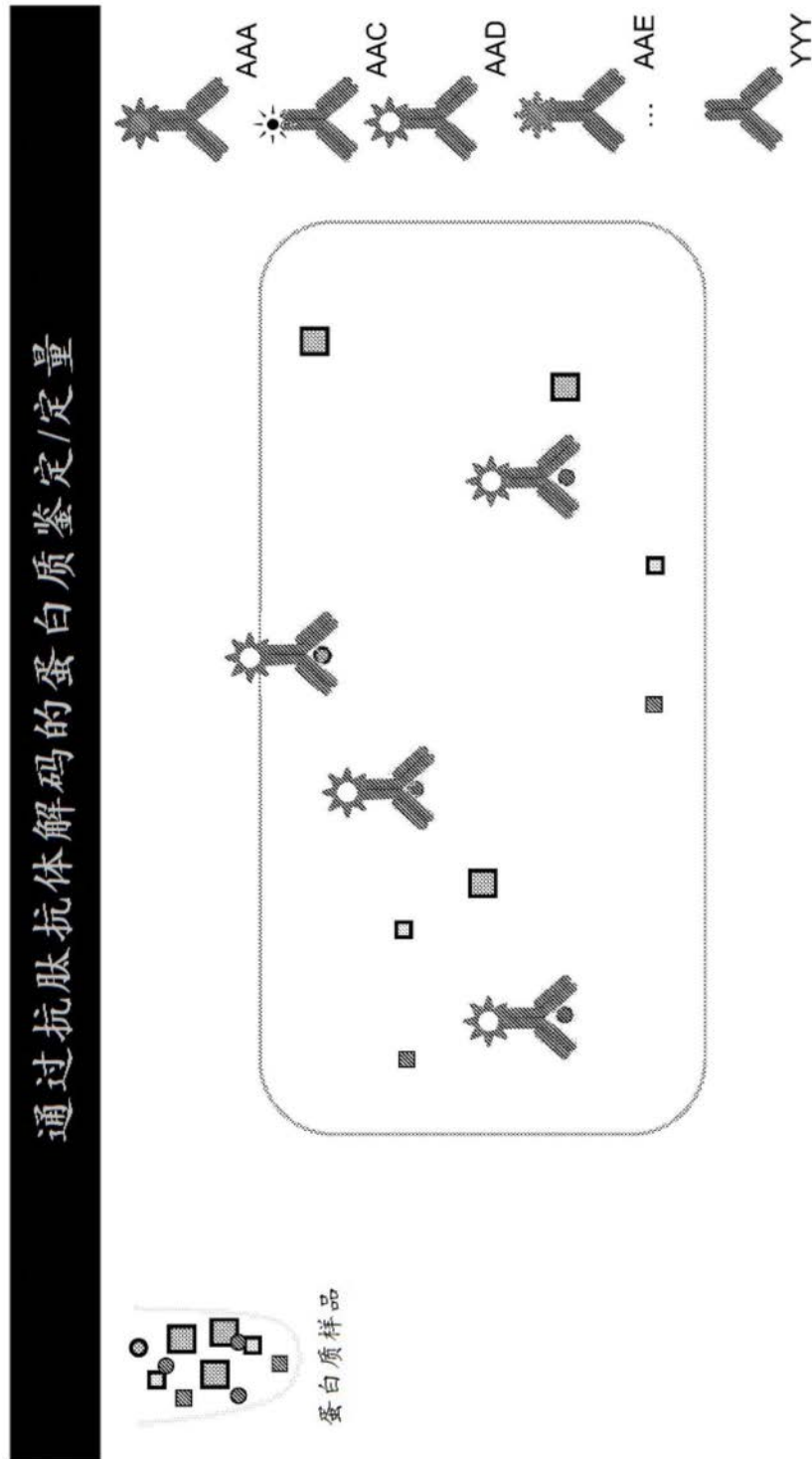


图9

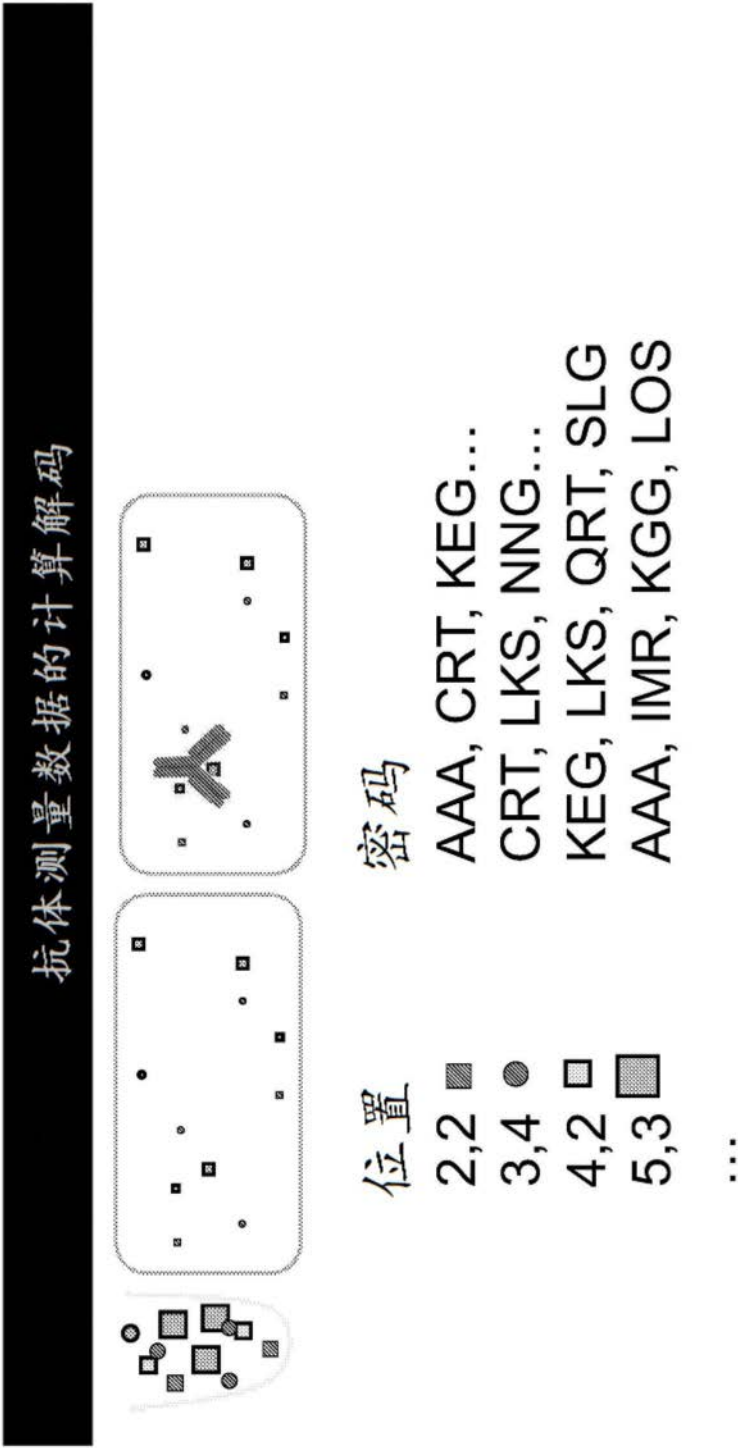


图10

全蛋白质组定量：所有蛋白质的计数			
蛋白质	#Mol	#解码_命中	
蛋白质 1	106	10	
蛋白质 2	832	6	
蛋白质 3	1.6×10^6	30	
...			
蛋白质 25K	1	22	

图11

异常列表：可能的突变体和经修饰蛋白质			
位置	可能的蛋白质	附加密码_命中	
斑点 1, 2	蛋白质 1	TRT, Y', ...	
斑点 12, 36	蛋白质 2	RBM, ...	
...			

图12

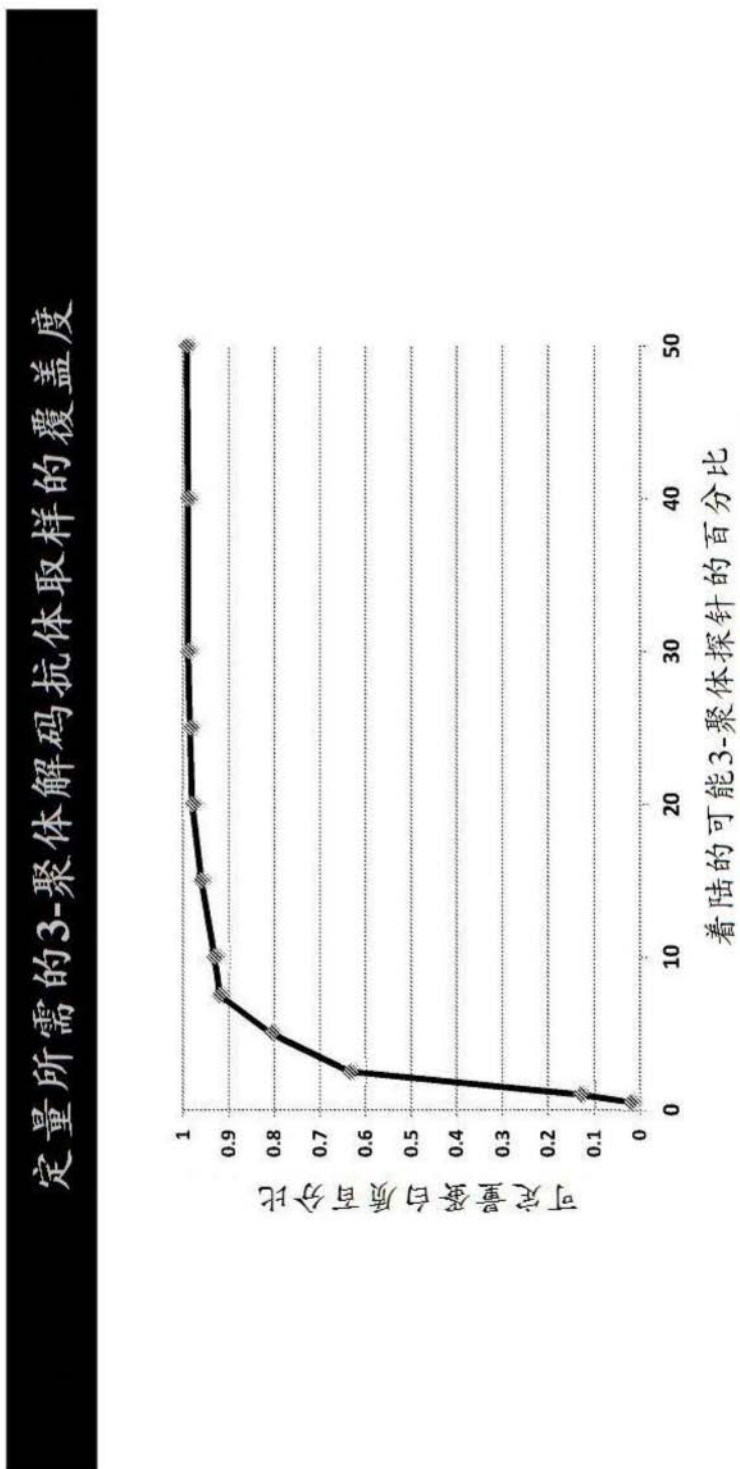


图13

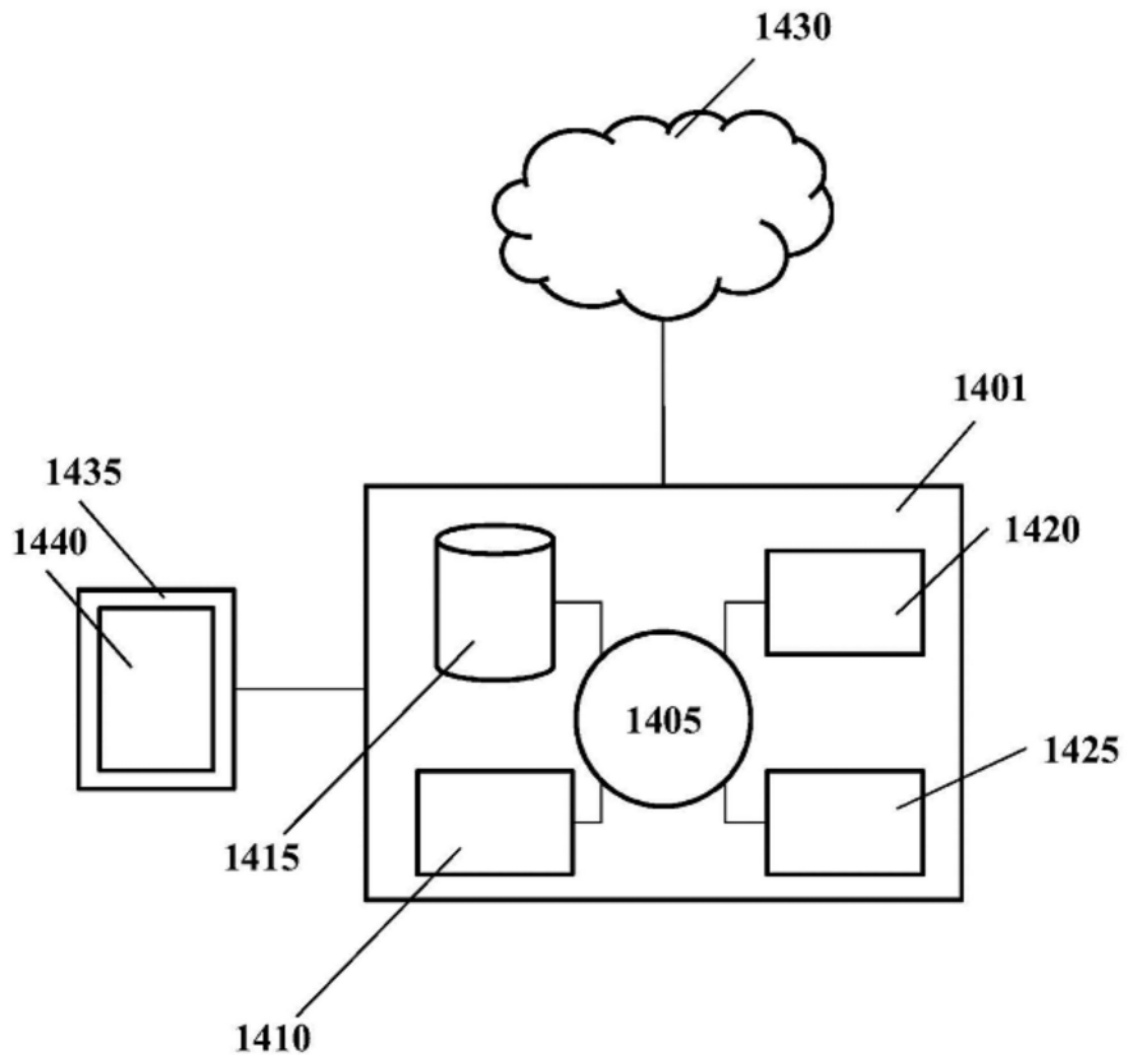


图14

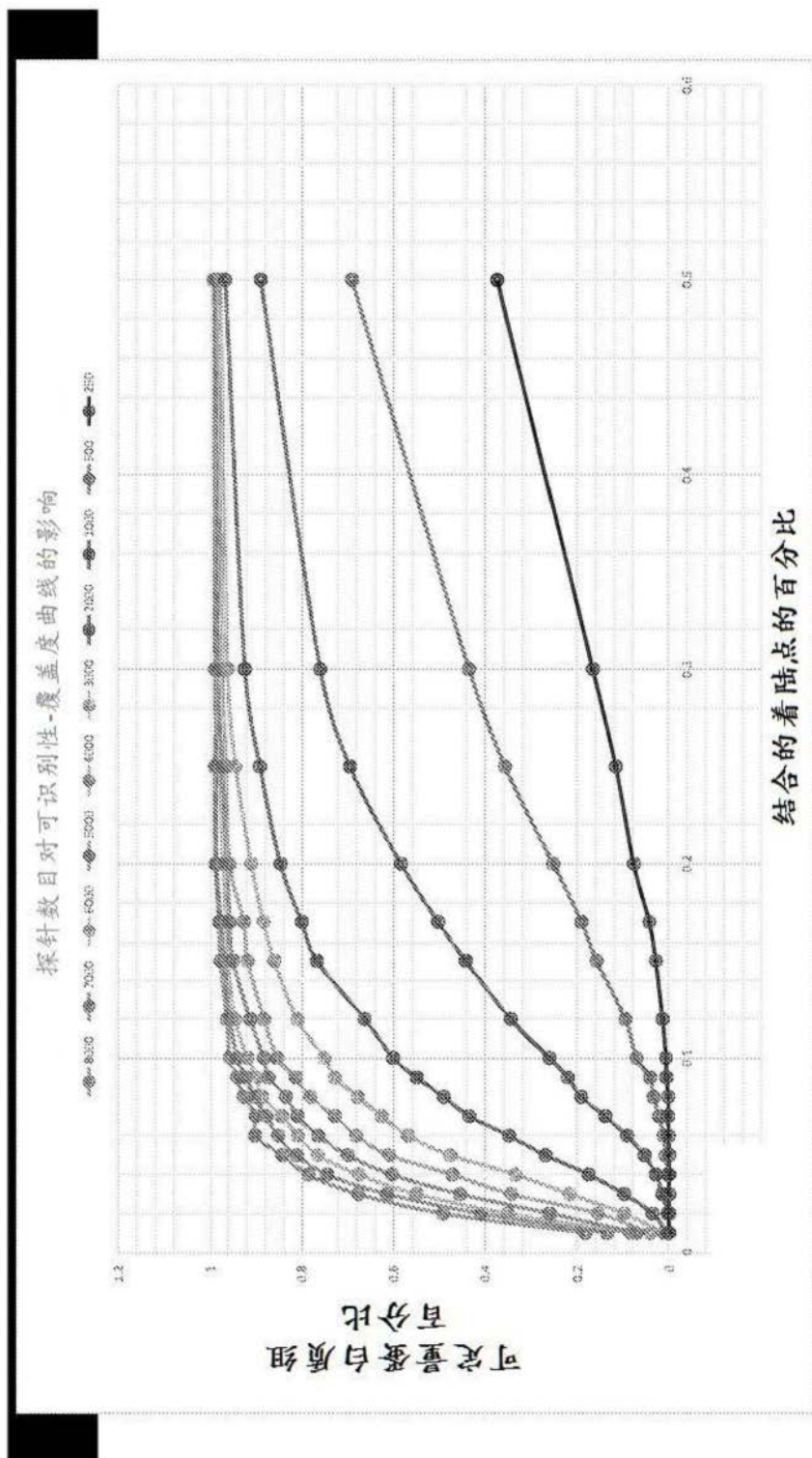


图15

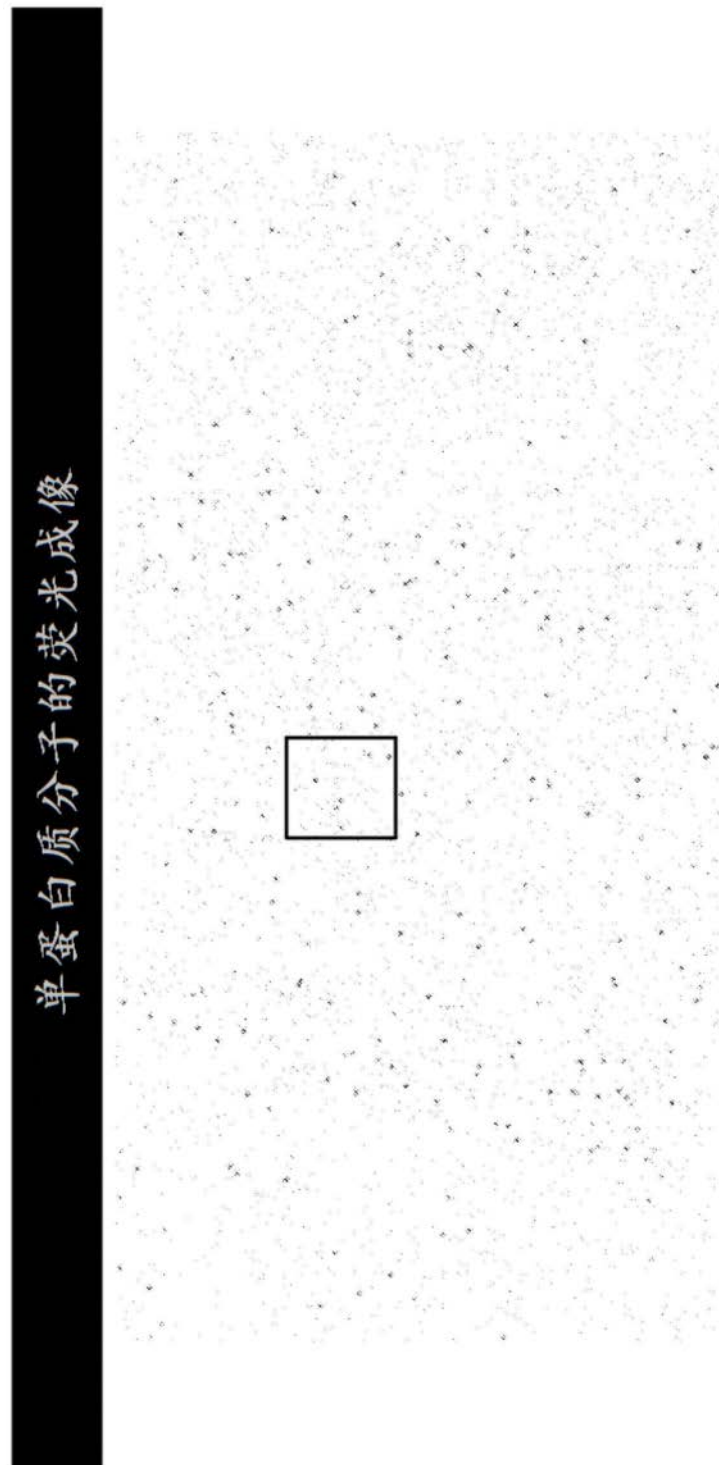


图16A

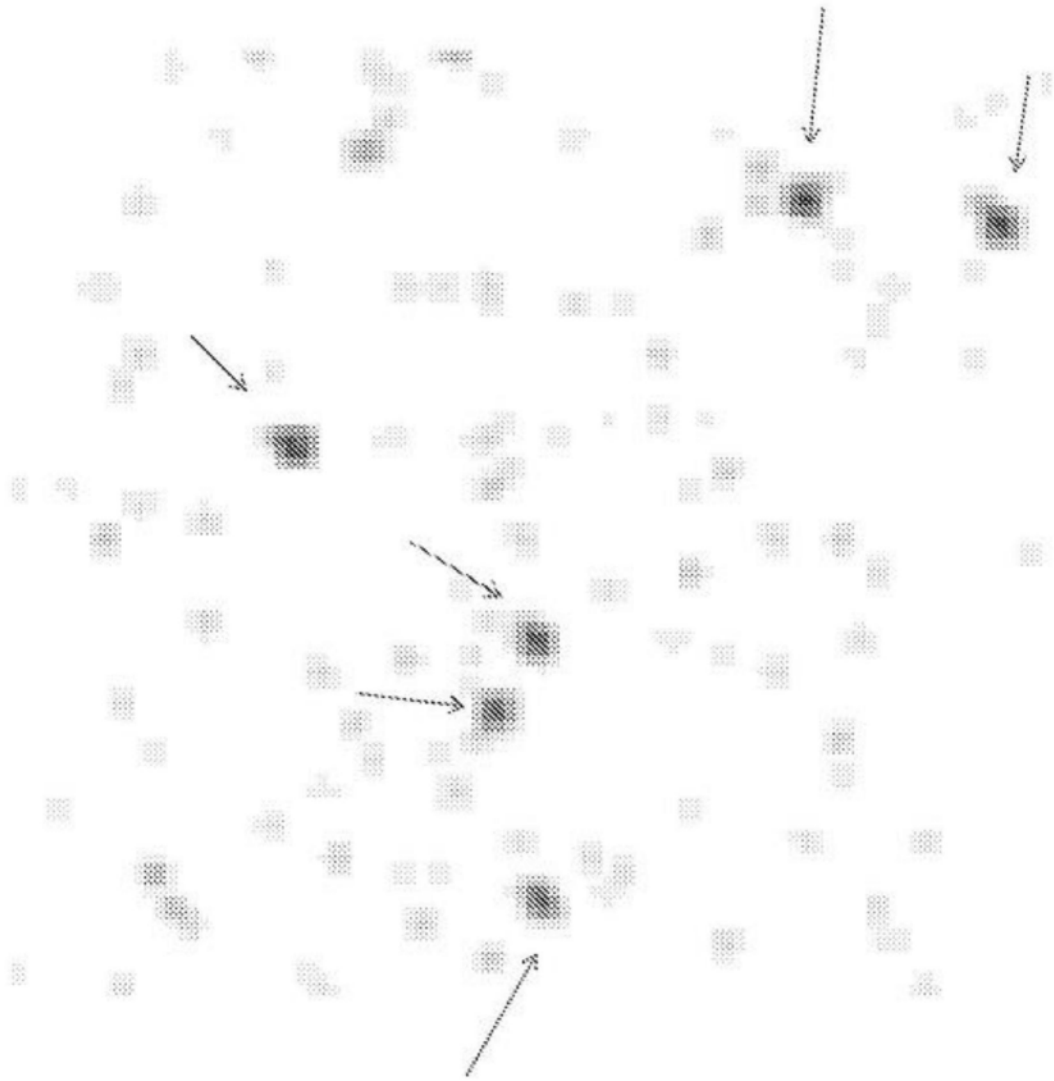


图16B

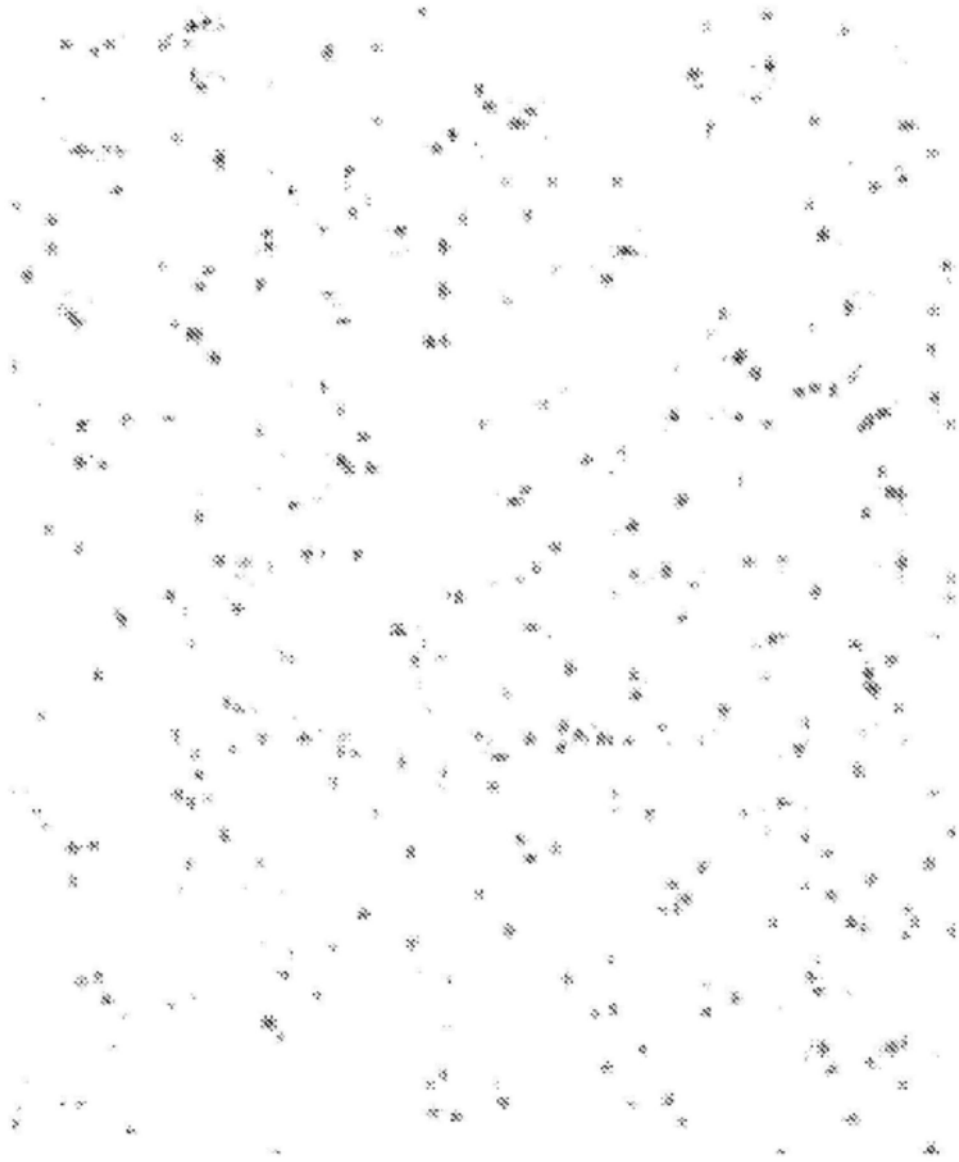


图17

鉴定A蛋白质

```

>sp|P42212|GFP_AEQVI_绿色荧光蛋白 OS=维多利亚多管水母 GN=GFP PE=1 SV=1 6-HIS
MSKGEELFTGVVPIIVEIDGDVNGHKFSVSGEGEGDATYGLTKLTKICTTGKLPVWPFTLVTTTSYGVCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKTEG
DTLVNRITELKGIDFKEDGNIILGHKLEYNYNSHNVITMADKQNGIKYNFKIRHNIEDGSVQLADHYQQNTPIGDGSVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITHG
MDELYKHHHHH

>sp|P61823|RNAS1_BOVIN_胰岛素核糖核酸酶 OS=奶牛 GN=RNASE1 PE=1 SV=1
MALKSLVLLSLVLLVLRVQPSLGKETAAAKFERQHMDSTSAASSSNYCQNMKSRNLTKDRCKPVNTFVHFSLADVQAVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSS
KYPNCAYKTTQANKHIIIVACEGNPVYPVHEDASV

>sp|P02788|TRFL_HUMAN_乳透转蛋白 OS=智人 GN=LTF PE=1 SV=6
MKLVFLVLLFGLAGLRRRSVQWCAVSQPEATKCFQWQRNMKVRGPPVSCIKRDSPIQCIQIAENRADAVTLDGGFIYEAGLAPYKLRPVAAEVYGTQPRTHYAVAV
VKKGSFQNLQGLKSGHTGLRRTAGWNPVIGTLRPFLNWTGPPEPIEAAVAFSSAVPGADKGFPNLCRLCAGTGENKCAFSSQEPYFYSGAFKCLRDGAGDVAFIREST
VFEDLSDEAERDEYELLCPDTRKPVDKFKDCHLARVPSHAVVARSVNGKEDAIWNLRQAEKFGKSPKQLFGSPSGQKDLLFKDSAIGFSRVPPRIDSLGLYLGSGYFTAIQ
NLRKSEFEVAARRARVAVCAVGEQELRKCNQWSGLSEGVTCSSASTEDCIALVLKGEADAMSLDGGVYTAGKGLVPVLAENYKQSSDPDPNCVDRPVEGYLAVAVVRRSD
TSLTWNSVKGKKSCHTAVDRTAGWNIPMGILLFNQTSCKFDEYFSQSCAPGSDPRSNLCALCIGDEQGENKCVPNSNERIYGYTGAFRCLAENAGDVAFVKDVTVLQNTDGNNEA
WAKDLKLADFALLCLDGRKKPVTEARSCHLAMAPNHAVVSRMDKVERLKQVLLHQAKFGRNGSDCPDKFCLFQSETRKLLFNDNTECLARLHGKTTYEKYLGPQYVAGITNLKKC
STSPILLEACEFLRK

>tr|C1L9P1|C1L9P1_SCHJA_谷胱甘肽S-转氨酶 M1 OS=日本血吸虫 GN=GSTM1 PE=2 SV=1 HIS-V5-GSTM1
CAX71422 SJCDD067I637
FTMSPIILGYWKIKGIVQPTRLLILEYLEEKYEHLYERDEGDKWRNKKFELGLEPPLPYYIGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRITAYSKD
FETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDEMLYDALDVVLYNDPMCLOAFPKLVCFKKRIEALPQIDKYLKSSKYIAWPLQGWAQATFGGGDHPPKSDLVPRGSPLEV
LFQGPCGSYGS

```

图18