(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113316638 A
(43)申请公布日 2021.08.27

(21)申请号 201980089624.6
(22)申请日 2019.11.20
(30)优先权数据
62/770,136 2018.11.20 US
(85)PCT国际申请进入国家阶段日
2021.07.20
(86)PCT国际申请的申请数据
PCT/US2019/062482 2019.11.20
(87)PCT国际申请的公布数据
WO2020/106889 EN 2020.05.28
(71)申请人 诺迪勒思生物科技公司
地址 美国加利福尼亚州
(72)发明人 乔什·克莱因 贾勒特·艾格特森
特蕾西·陈 索纳尔·托纳皮
帕拉格·马利克 郑珉溶
阿卜杜拉·奥泽
埃利奥特·索雷勒 戴维·斯特恩
(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
代理人 刘晓杰 韦昌金
(51)Int.Cl.
C12N 15/10(2006.01)
C12N 15/115(2006.01)
C40B 30/04(2006.01)
C40B 40/06(2006.01)
G01N 33/53(2006.01)

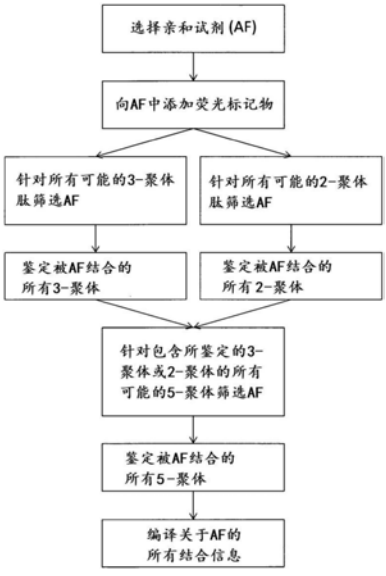
权利要求书4页 说明书64页 附图20页

(54)发明名称

亲和试剂的设计和选择

(57)摘要

本文提供了用于选择结合多种序列环境中的所需肽表位的亲和试剂的方法和系统。该方法依赖于获得肽文库,并使用该肽文库选择亲和试剂,其中每个肽具有序列αXβ,其中X是所需的肽表位,其中α和β各自包含氨基酸。



CN 113316638 A

1.一种鉴定针对表位的亲和试剂的方法,其包括:

获得包含第一多个肽的第一肽文库,每个肽包含式αXβ的序列,其中X是所需表位,并且其中α和β是包含已知氨基酸序列的侧翼结构域;

将所述第一肽文库暴露于多种亲和试剂,从而使所述多种亲和试剂中的至少一种亲和试剂与所述第一肽文库结合;

确定与所述第一多个肽结合的所述至少一种亲和试剂的一个或多个序列,从而形成缩减的亲和试剂池;

获得包含第二多个肽的第二肽文库,每个肽包含式γXδ的序列,其中X是所需表位,并且其中γ和δ是包含已知氨基酸序列的侧翼结构域;

将所述第二肽文库暴露于所述缩减的亲和试剂池,从而使所述缩减的亲和试剂池中的至少一种亲和试剂与所述第二肽文库结合;以及

确定与所述第二肽文库结合的所述缩减的亲和试剂池中的至少一种亲和试剂的一个或多个序列,从而鉴定针对表位X的亲和试剂。

2.根据权利要求1所述的方法,其中α和β包含相同的氨基酸序列。

3.根据权利要求1所述的方法,其中α和β包含不同的氨基酸序列。

4.根据权利要求1所述的方法,其中α和β包含与γ和δ不同的氨基酸序列。

5.根据权利要求1所述的方法,其中α和β包含具有均一的化学多样性的氨基酸序列。

6.根据权利要求5所述的方法,其中所述具有均一的化学多样性的氨基酸序列包括包含两个或更多个具有等同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

7.根据权利要求1所述的方法,其中α和β包含具有不均一的化学多样性的氨基酸序列。

8.根据权利要求7所述的方法,其中所述具有不均一的化学多样性的氨基酸序列包括包含两个或更多个具有不同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

9.根据权利要求1所述的方法,其中γ和δ包含具有均一的化学多样性的氨基酸序列。

10.根据权利要求9所述的方法,其中所述具有均一的化学多样性的氨基酸序列包括包含两个或更多个具有等同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

11.根据权利要求1所述的方法,其中γ和δ包含具有不均一的化学多样性的氨基酸序列。

12.根据权利要求11所述的方法,其中所述具有不均一的化学多样性的氨基酸序列包括包含两个或更多个具有不同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

13.根据权利要求1所述的方法,其进一步包括洗涤与所述第一肽文库结合的所述多种亲和试剂中的所述至少一种亲和试剂。

14.根据权利要求1所述的方法,其进一步包括洗涤与所述第二肽文库结合的所述缩减的亲和试剂池中的所述至少一种亲和试剂。

15.根据权利要求1所述的方法,其中所述第一或第二肽文库内的肽处于天然状态。

16.根据权利要求1所述的方法,其中所述第一或第二肽文库内的肽处于非天然状态。

17.根据权利要求1所述的方法，其中所述非天然状态是变性状态或部分折叠状态。

18.根据权利要求1所述的方法，其中所述至少一种亲和试剂与多于一个表位结合。

19.根据权利要求18所述的方法，其中所述多于一个表位包含简并氨基酸序列。

20.根据权利要求19所述的方法，其中所述简并氨基酸序列相差一个氨基酸。

21.根据权利要求19所述的方法，其中所述简并氨基酸序列相差多于一个氨基酸。

22.根据权利要求19、20或21所述的方法，其中所述差异氨基酸包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

23.根据权利要求19、20或21所述的方法，其中所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

24.根据权利要求1所述的方法，其包括重复一次或多次将所述第一肽文库暴露于所述多种亲和试剂以及确定所述至少一种亲和试剂的一个或多个序列的步骤。

25.根据权利要求1所述的方法，其包括重复一次或多次将所述第二肽文库暴露于所述缩减的亲和试剂池以及确定所述至少一种亲和试剂的一个或多个序列的步骤。

26.根据权利要求1所述的方法，其中所述多种亲和试剂中的亲和试剂包含寡核苷酸、肽、微型蛋白质结合物、抗体、抗体片段或其组合。

27.根据权利要求1所述的方法，其中所述多种亲和试剂中的亲和试剂包含寡核苷酸。

28.一种鉴定针对表位的亲和试剂的方法，其包括：

使第一介质中肽文库内的多个肽与第二介质中亲和试剂池内的多种亲和试剂接触以形成肽-亲和试剂复合物，其中每个肽-亲和试剂复合物包含肽和至少一种结合的亲和试剂，并且其中每个肽包含表位X；

收集所述肽-亲和试剂复合物；

分离未结合的亲和试剂；以及

确定所述肽-亲和试剂复合物的亲和试剂的一个或多个序列，从而鉴定针对表位X的至少一种亲和试剂。

29.根据权利要求28所述的方法，其中所述肽文库内的肽处于天然状态。

30.根据权利要求28所述的方法，其中所述肽文库内的肽处于非天然状态。

31.根据权利要求28所述的方法，其中所述非天然状态是变性状态或部分折叠状态。

32.根据权利要求30或31所述的方法，其进一步包括对所述处于天然状态的多个肽重复所有步骤。

33.根据权利要求32所述的方法，其中针对处于天然状态的表位X的所述至少一种亲和试剂不同于针对处于非天然状态的表位X的所述至少一种亲和试剂。

34.根据权利要求32所述的方法，其中针对处于天然状态的表位X的所述至少一种亲和试剂与针对处于非天然状态的表位X的所述至少一种亲和试剂相同。

35.根据权利要求28所述的方法，其中所述表位X与多于一种亲和试剂结合。

36.根据权利要求28所述的方法，其中所述至少一种亲和试剂与多于一个表位结合。

37.根据权利要求36所述的方法，其中所述多于一个表位包含简并氨基酸序列。

38.根据权利要求37所述的方法，其中所述简并氨基酸序列相差一个氨基酸。

39.根据权利要求37所述的方法，其中所述简并氨基酸序列相差多于一个氨基酸。

40.根据权利要求37、38或39所述的方法，其中所述差异氨基酸包括相似的化学性质，

其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

41.根据权利要求37、38或39所述的方法，其中所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

42.根据权利要求28所述的方法，其中所述多种亲和试剂中的亲和试剂包含寡核苷酸。

43.根据权利要求28所述的方法，其中所述确定与所述多个肽结合的所述至少一种亲和试剂的一个或多个序列包括：

将所述至少一种亲和试剂与所述多个肽分离；以及

对所述至少一种亲和试剂进行测序，从而确定所述至少一种亲和试剂的所述一个或多个序列。

44.一种包含$S_1S_2S_3$结构的亲和试剂，其中S_1和S_3是能够结合互补衔接子的衔接子结构域，并且其中S_2是表位结合域，其中所述表位结合域以小于10^{-6}M的平衡结合常数K_d与所述一个或多个肽表位结合。

45.根据权利要求44所述的亲和试剂，其中所述亲和试剂包含寡核苷酸。

46.根据权利要求44所述的亲和试剂，其中所述衔接子结构域是下一代测序衔接子。

47.根据权利要求44所述的亲和试剂，其中S_1或S_3包含条形码。

48.根据权利要求45所述的亲和试剂，其中S_2包含不超过约90个核苷酸。

49.根据权利要求48所述的亲和试剂，其中S_2包含不超过约60个核苷酸。

50.根据权利要求49所述的亲和试剂，其中S_2包含不超过约30个核苷酸。

51.根据权利要求44所述的亲和试剂，其中所述亲和试剂以小于10^{-6}M的平衡结合常数K_d与多于一个表位结合。

52.根据权利要求51所述的亲和试剂，其中所述亲和试剂以小于10^{-9}M的平衡结合常数K_d与一个或多个表位结合。

53.根据权利要求52所述的亲和试剂，其中所述亲和试剂以小于10^{-12}M的平衡结合常数K_d与一个或多个表位结合。

54.根据权利要求44所述的亲和试剂，其中所述至少一种亲和试剂与多于一个表位结合。

55.根据权利要求54所述的亲和试剂，其中所述多于一个表位包含简并氨基酸序列。

56.根据权利要求55所述的亲和试剂，其中所述简并氨基酸序列相差一个氨基酸。

57.根据权利要求55所述的亲和试剂，其中所述简并氨基酸序列相差多于一个氨基酸。

58.根据权利要求55、56或57所述的亲和试剂，其中所述差异氨基酸包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

59.根据权利要求55、56或57所述的亲和试剂，其中所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

60.根据权利要求44所述的亲和试剂，其中当所述肽处于天然状态时，所述亲和试剂与一个或多个肽表位结合。

61.根据权利要求44所述的亲和试剂，其中当所述肽处于非天然状态时，所述亲和试剂与一个或多个肽表位结合。

62.根据权利要求61所述的亲和试剂，其中所述非天然状态包括变性状态或部分折叠状态。

63.一种亲和试剂池,其包含:

多种亲和试剂;

其中所述多种亲和试剂中的两种或更多种亲和试剂包含结构$S_1S_2S_3$,其中S_1和S_3包含共同的衔接子结构域,并且S_2包含表位结合域;并且

其中所述两种或更多种亲和试剂中的第一亲和试剂包含具有针对肽表位X_1的表征的特异性的第一结构域S_2,并且其中所述两种或更多种亲和试剂中的第二亲和试剂包含具有针对肽表位X_2的表征的特异性的第二结构域S_2,其中所述第一结构域S_2不同于所述第二结构域S_2。

64.一种鉴定针对表位的亲和试剂的方法,其包括:

获得包含多个肽的肽文库,其中所述多个肽中的肽在其序列中包含多于一个表位X;

将所述肽文库暴露于包含多种亲和试剂的亲和试剂池,从而使至少一种亲和试剂与在其序列中包含多于一个表位X的所述肽结合;以及

确定与在其序列中包含多于一个表位X的所述肽结合的所述至少一种亲和试剂的一个或多个序列,从而鉴定针对表位X的亲和试剂。

亲和试剂的设计和选择

交叉引用

[0001] 本申请要求2018年11月20日提交的第62/770,136号美国临时申请的权益,该临时申请通过引用并入本文。

背景技术

[0002] 用于生成结合试剂的选择方法一般被设计为选择对单个表位(epitope)或蛋白质具有高亲和力和特异性的结合试剂。对于某些应用,选择结合多个表位的结合试剂或表征对单个蛋白质或表位没有特异性的结合试剂的结合模式可能是有用的。

发明内容

[0003] 本公开提供了用于选择和表征亲和试剂的方法和系统。在一些实施方案中,本公开提供了选择亲和试剂以与多种序列环境中的表位结合的方法。本文所述的方法和系统也可用来表征亲和试剂的结合模式,以及序列环境对亲和试剂与表位的结合的影响。另外,本文描述的方法和系统可用来表征并选择跨多个表位结合的亲和试剂,如跨具有相同长度的多个表位结合。通过证明在相同长度的多个表位上的混杂性,特定的亲和试剂可用于将序列鉴定为包含该特定亲和试剂所结合的多个表位中的至少一个表位。

[0004] 在一方面,提供了一种鉴定针对表位的亲和试剂的方法,其包括:获得包含第一多个肽的第一肽文库,每个肽包含式αXβ的序列,其中X是所需表位,并且其中α和β是包含已知氨基酸序列的侧翼结构域;将所述第一肽文库暴露于多种亲和试剂,从而使所述多种亲和试剂中的至少一种亲和试剂与所述第一肽文库结合;确定与所述第一多个肽结合的所述至少一种亲和试剂的一个或多个序列,从而形成缩减的亲和试剂池;获得包含第二多个肽的第二肽文库,每个肽包含式γXδ的序列,其中X是所需表位,并且其中γ和δ是包含已知氨基酸序列的侧翼结构域;将所述第二肽文库暴露于所述缩减的亲和试剂池,从而使所述缩减的亲和试剂池中的至少一种亲和试剂与所述第二肽文库结合;以及确定与所述第二肽文库结合的所述缩减的亲和试剂池中的至少一种亲和试剂的一个或多个序列,从而鉴定针对表位X的亲和试剂。

[0005] 在一些情况下,α和β包含相同的氨基酸序列。在一些情况下,α和β包含不同的氨基酸序列。在一些情况下,α和β包含与γ和δ不同的氨基酸序列。在一些情况下,α和β包含具有均一的化学多样性的氨基酸序列。在一些情况下,所述具有均一的化学多样性的氨基酸序列包括包含两个或更多个具有等同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。在一些情况下,α和β包含具有不均一的化学多样性的氨基酸序列。在一些情况下,所述具有不均一的化学多样性的氨基酸序列包括包含两个或更多个具有不同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

[0006] 在一些情况下,γ和δ包含具有均一的化学多样性的氨基酸序列。在一些情况下,所述具有均一的化学多样性的氨基酸序列包括包含两个或更多个具有等同电荷、疏水性、

亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。在一些情况下，γ和δ包含具有不均一的化学多样性的氨基酸序列。在一些情况下，所述具有不均一的化学多样性的氨基酸序列包括包含两个或更多个具有不同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

[0007] 在一些情况下，所述方法进一步包括洗涤与所述第一肽文库结合的所述多种亲和试剂中的所述至少一种亲和试剂。在一些情况下，所述方法进一步包括洗涤与所述第二肽文库结合的所述缩减的亲和试剂池中的所述至少一种亲和试剂。

[0008] 在一些情况下，所述第一或第二肽文库内的肽处于天然状态。在一些情况下，所述第一或第二肽文库内的肽处于非天然状态。在一些情况下，所述非天然状态是变性状态或部分折叠状态。

[0009] 在一些情况下，所述至少一种亲和试剂与多于一个表位结合。在一些情况下，所述多于一个表位包含简并氨基酸序列(degenerate amino acid sequences.)。在一些情况下，所述简并氨基酸序列相差一个氨基酸。根据权利要求20所述的方法，其中所述简并氨基酸序列相差多于一个氨基酸。在一些情况下，所述差异氨基酸包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。在一些情况下，所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

[0010] 在一些情况下，所述方法包括重复一次或多次将所述第一肽文库暴露于所述多种亲和试剂以及确定所述至少一种亲和试剂的一个或多个序列的步骤。在一些情况下，所述方法包括重复一次或多次将所述第二肽文库暴露于所述缩减的亲和试剂池以及确定所述至少一种亲和试剂的一个或多个序列的步骤。

[0011] 在一些情况下，所述多种亲和试剂中的亲和试剂包含寡核苷酸、肽、微型蛋白质结合物、抗体、抗体片段或其组合。在一些情况下，所述多种亲和试剂中的亲和试剂包含寡核苷酸。

[0012] 在另一方面，提供了一种鉴定针对表位的亲和试剂的方法，其包括：使第一介质中肽文库内的多个肽与第二介质中亲和试剂池内的多种亲和试剂接触以形成肽-亲和试剂复合物，其中每个肽-亲和试剂复合物包含肽和至少一种结合的亲和试剂，并且其中每个肽包含表位X；收集所述肽-亲和试剂复合物；分离未结合的亲和试剂；以及确定所述肽-亲和试剂复合物的亲和试剂的一个或多个序列，从而鉴定至少一种针对表位X的亲和试剂。

[0013] 在一些情况下，所述肽文库内的肽处于天然状态。在一些情况下，所述肽文库内的肽处于非天然状态。在一些情况下，所述非天然状态是变性状态或部分折叠状态。在一些情况下，所述方法进一步包括对所述处于天然状态的多个肽重复所有步骤。在一些情况下，所述至少一种针对处于天然状态的表位X的亲和试剂不同于所述至少一种针对处于非天然状态的表位X的亲和试剂。在一些情况下，所述至少一种针对处于天然状态的表位X的亲和试剂与所述至少一种针对处于非天然状态的表位X的亲和试剂相同。在一些情况下，所述表位X与多于一种亲和试剂结合。

[0014] 在一些情况下，所述至少一种亲和试剂与多于一个表位结合。在一些情况下，所述多于一个表位包含简并氨基酸序列。在一些情况下，所述简并氨基酸序列相差一个氨基酸。在一些情况下，所述简并氨基酸序列相差多于一个氨基酸。在一些情况下，所述差异氨基酸

包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。在一些情况下，所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

[0015] 在一些情况下，所述多种亲和试剂中的亲和试剂包含寡核苷酸。

[0016] 在一些情况下，确定与所述多个肽结合的所述至少一种亲和试剂的一个或多个序列包括将所述至少一种亲和试剂与所述多个肽分离，以及对所述至少一种亲和试剂进行测序，从而确定所述至少一种亲和试剂的一个或多个序列。

[0017] 在另一方面，提供了一种包含$S_1S_2S_3$结构的亲和试剂，其中S_1和S_3是能够结合互补衔接子的衔接子结构域，并且其中S_2是表位结合域，其中所述表位结合域以小于10^{-6}M的平衡结合常数K_d与所述一个或多个肽表位结合。

[0018] 在一些情况下，所述亲和试剂包含寡核苷酸。在一些情况下，所述衔接子结构域是下一代测序衔接子。在一些情况下，S_1或S_3包含条形码。在一些情况下，S_2包含不超过约90个核苷酸。在一些情况下，S_2包含不超过约60个核苷酸。在一些情况下，S_2包含不超过约30个核苷酸。

[0019] 在一些情况下，所述亲和试剂以小于10^{-6}M的平衡结合常数K_d与多于一个表位结合。在一些情况下，所述亲和试剂以小于10^{-9}M的平衡结合常数K_d与一个或多个表位结合。在一些情况下，所述亲和试剂以小于10^{-12}M的平衡结合常数K_d与一个或多个表位结合。

[0020] 在一些情况下，所述至少一种亲和试剂与多于一个表位结合。在一些情况下，所述多于一个表位包含简并氨基酸序列。在一些情况下，所述简并氨基酸序列相差一个氨基酸。在一些情况下，所述简并氨基酸序列相差多于一个氨基酸。在一些情况下，所述差异氨基酸包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。在一些情况下，所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

[0021] 在一些情况下，当所述肽处于天然状态时，所述亲和试剂与一个或多个肽表位结合。在一些情况下，当所述肽处于非天然状态时，所述亲和试剂与一个或多个肽表位结合。在一些情况下，所述非天然状态包括变性状态或部分折叠状态。

[0022] 在另一方面，提供了一种亲和试剂池，其包含多种亲和试剂，其中所述多种亲和试剂中的两种或更多种亲和试剂包含结构$S_1S_2S_3$，其中S_1和S_3包含共同的衔接子结构域，并且S_2包含表位结合域，并且其中所述两种或更多种亲和试剂中的第一亲和试剂包含具有表征的针对肽表位X_1的特异性的第一结构域S_2，并且其中所述两种或更多种亲和试剂中的第二亲和试剂包含具有表征的针对肽表位X_2的特异性的第二结构域S_2，其中所述第一结构域S_2不同于所述第二结构域S_2。

[0023] 在另一方面，提供了一种鉴定针对表位的亲和试剂的方法，其包括获得包含多个肽的肽文库，其中所述多个肽中的肽在其序列中包含多于一个表位X，将所述肽文库暴露于包含多种亲和试剂的亲和试剂池，从而使至少一种亲和试剂与在其序列中包含多于一个表位X的肽结合，以及确定与在其序列中包含多于一个表位X的肽结合的所述至少一种亲和试剂的一个或多个序列，从而鉴定针对表位X的亲和试剂。

援引并入

[0024] 本说明书中所提及的所有出版物、专利和专利申请均通过引用并入本文，其程度

犹如具体地且单独地指出每个单独的出版物、专利或专利申请均通过引用而并入。

附图说明

[0025] 本发明的新颖特征在所附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以及附图，将会对本发明的特征和优点获得更好的理解，在这些附图中：

[0026] 图1示出了根据一些实施方案，用于选择亲和试剂的固定的靶标，以及包含该靶标的肽的示例性列表。

[0027] 图2示出了根据一些实施方案的亲和试剂选择过程的流程图。

[0028] 图3A示出了根据一些实施方案，亲和试剂表征过程中的筛选步骤的示例性流程图。

[0029] 图3B示出了根据一些实施方案，用于进行筛选步骤以确定侧翼序列对亲和试剂与表位AAA结合的影响的5-聚体αXβ肽阵列。

[0030] 图4示出了被编程或以其他方式配置为实现本文提供的方法的计算机控制系统。

[0031] 图5示出了嵌入具有不同二级结构的较长序列中的靶标的实例。图5A示出了根据一些实施方案，与嵌入α螺旋肽的不同区域中的靶标(显示为复选框)形成α螺旋的肽的实例。图5B示出了根据一些实施方案，与嵌入β折叠形成肽的不同区域中的靶标(显示为黑色实心框)形成β折叠的肽的实例。

[0032] 图6示出了根据一些实施方案，在适体选择筛选中鉴定的几种适体的靶标结合。

[0033] 图7示出了根据一些实施方案，所鉴定的适体的初步表征。

[0034] 图8示出了根据一些实施方案，所鉴定的适体的初步表征。

[0035] 图9示出了根据一些实施方案，所合成的肽的质谱法验证。

[0036] 图10示出了根据一些实施方案的簇扩增的实例。

[0037] 图11示出了根据一些实施方案，荧光标记的肽靶标与适体簇在流动池上的结合。

[0038] 图12示出了根据一些实施方案，几种适体对肽靶标的结合亲和力。

[0039] 图13A显示了针对包含氨基酸序列LRD重复序列的肽测试的亲和试剂的游离溶液SELEX富集数据。图13B显示了针对包含氨基酸序列TQA重复序列的肽测试的亲和试剂的游离溶液SELEX富集数据。

[0040] 图14A显示了针对包含氨基酸序列YSL重复序列的肽测试的亲和试剂的磁珠SELEX富集数据。图14B显示了针对包含氨基酸序列EVT的肽测试的亲和试剂的磁珠SELEX富集数据。

[0041] 图15A显示了对于针对500个肽筛选的第一亲和试剂，预测的三聚体结合概率相对于真实结合概率的对数图。图15B显示了来自图15A的数据在非对数标度上的图示。

[0042] 图16A显示了对于针对500个肽筛选的第二亲和试剂，预测的三聚体结合概率相对于真实结合概率的对数图。图16B显示了来自图16A的数据在非对数标度上的图示。

[0043] 图17A显示了对于针对500个肽筛选的第三亲和试剂，预测的三聚体结合概率相对于真实结合概率的对数图。图17B显示了来自图17A的数据在非对数标度上的图示。

具体实施方式

[0044] 在整个生命科学领域，人们对能够结合特定蛋白质、代谢物、细胞或细胞界面的亲和试剂的开发具有浓厚的兴趣。最近，亲和试剂选择技术已经扩展到包括非天然核苷酸和氨基酸。这些方法的目的是开发仅与给定表位结合的试剂。

[0045] 结合的排他性被认为是亲和试剂的理想特性。已经做出了大量的努力来确保亲和试剂仅与一种蛋白质结合，而与其他蛋白质的结合最低。例外的情况是针对功能上重要的残基如磷酸酪氨酸的抗体。

[0046] 亲和试剂所面临的一个特殊挑战可能是环境敏感性。例如，亲和试剂可以与核心表位完美地结合，但是取决于侧翼残基，可能偏向于较好地结合或根本不结合。例如，当生成针对肽n-Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu的亲和试剂时，如果n-末端残基是Gly，相比于Ser，该试剂可以结合得更好。

[0047] 尽管通常可能不希望生成对任何一种蛋白质都不具有特异性的亲和试剂，但是存在特定的使用情况，其中亲和试剂对许多蛋白质中可能出现的一种或多种肽具有特异性可能是最佳的。

[0048] 如本文所用的，术语“表位”可指被亲和试剂识别的诸如蛋白质或肽等大分子的一部分。在一些情况下，表位可以是被抗体识别的蛋白质或肽的一部分。在一些情况下，表位可以是被抗体片段识别的蛋白质或肽的一部分。在一些情况下，表位可以是被适体识别的蛋白质或肽的一部分。在一些情况下，表位可以是被肽识别的蛋白质或肽的一部分。

[0049] 如本文所用的，术语“抗原性”可指化学结构(抗原、半抗原、表位或氨基酸序列)与一组具有适应性免疫的特定产物或与一类亲和试剂特异性结合的能力。术语“抗原性”可以与术语“适体原性(aptagenicity)”——化学结构被适体识别的能力——互换使用。术语“抗原性”还可以与术语“亲和试剂原性”互换使用，后者通常是指化学结构被亲和试剂识别的能力。

[0050] 本文所述的方法可用来生成具有所需性质的亲和试剂。例如，本文所述的方法可用来生成适体，使得当适体固定在颗粒上时，肽靶标保持与该适体结合，在缓冲液洗涤后，大约15分钟或更长时间后的信号损失小于10％。在另一个实例中，本文所述的方法可用来生成适体，使得当适体固定在颗粒上时，在洗脱缓冲液的存在下，大约15秒或更短时间后，来自结合的肽靶标的结合信号在缓冲液洗涤后降低至检测不到的水平。在另一个实例中，本文所述的方法可用来生成适体，使得在包含侧翼氨基酸残基身份(identities)的混合物的靶肽的存在下，适体结合性质得以保留。在一些情况下，在固定时保留结合性质的适体可能是理想的。

亲和试剂的选择

[0051] 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指数富集的配体系统进化(SELEX)、噬菌体展示、酵母展示、哺乳动物细胞展示、昆虫细胞展示、核糖体展示、颗粒展示、肽聚体(peptimer)进化、肽聚体设计和接种。示例性亲和试剂可包括寡聚体、适体、肽聚体(peptimers)、抗体、抗体片段、微型蛋白结合物、合成分子和/或其组合。在一些实例中，可使用基于结构的药物设计方法来设计亲和试剂。基于结构的药物设计(或直接药物设计)利用关于目的表位的三维结构和亲和试剂的结合位点的知识。

[0052] 在一些情况下，可以针对与序列环境无关地结合所需表位的能力来选择本公开的

亲和试剂。在一些实施方案中，亲和试剂可以被设计为当蛋白质处于部分或完全变性环境中时结合所需表位。在一些实施方案中，可以针对结合折叠或未折叠情况下的蛋白质中所需表位的能力来选择本公开的亲和试剂。在一些实施方案中，已经变性的蛋白质可以在蛋白质内包含或产生微折叠。在一些实施方案中，被选择用来识别所需表位AAA的亲和试剂可以同样好地或几乎同样好地与所有包含序列AAA的肽结合。在一些情况下，本公开的亲和试剂可以是根据表位的序列或结构环境具有不同亲和力的所需表位。在一些情况下，本公开的亲和试剂可以结合几个不同的表位，而与序列环境无关。在一些情况下，本公开的亲和试剂可以根据序列环境以不同的亲和力结合几个不同的表位。在一些情况下，可以基于表位的化学或物理性质而不是表位的氨基酸序列，将本公开的亲和试剂设计成与蛋白质表位结合。例如，亲和试剂可以(或者可以不)基于局部电荷、局部空间相互作用、疏水性、亲水性、酸性、碱性和其他任何化学或物理性质与表位结合。

[0053] 可以选择本公开的亲和试剂(例如，适体、肽聚体、微型蛋白质结合物等)以具有所需水平的结合混杂性。结合混杂性可指亲和试剂与多于一个特定表位结合的能力。非混杂亲和试剂可以是仅对单个表位具有结合特异性的任何亲和试剂。混杂亲和试剂可以是对多于一个表位具有结合特异性的任何亲和试剂。亲和试剂混杂性可指表位与表位家族结合的能力。在一些情况下，表位家族可能因共同的氨基酸结构而相关。例如，表位家族可具有结构GλG，其中λ可以是任何氨基酸(例如，GRG、GHG、GKG、GDG等)。在一些情况下，表位家族可能因共同的化学性质而相关。例如，表位家族可包含具有疏水侧链的所有3氨基酸表位(例如，AAV、AIM、WLV等)。在一些情况下，混杂亲和试剂可以与随机的一组没有可辨别的序列或结构基序的表位结合。在一些情况下，未表征的混杂亲和试剂可以与已知或表征的一组肽或蛋白质中的未知的一组表位结合。

[0054] 如果亲和试剂与约2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450、500、600、700、800、900、1000个或超过1000个表位结合，则其可以是混杂的。。如果亲和试剂与至少约2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450、500、600、700、800、900、1000个或超过1000个表位结合，则其可以是混杂的。。如果亲和试剂与不超过约1000、900、800、700、600、500、450、400、350、300、250、200、150、100、90、80、70、60、50、45、40、35、30、25、20、15、10、9、8、7、6、5、4、3个或少于3个表位结合，则其可以是混杂的。。

[0055] 亲和试剂的结合混杂性可以以二元或概率方式来表征。结合混杂性的二元表征可包括评估亲和试剂针对所有已知表位(例如，ARA-结合；AHA-结合；AKA-不结合；ADA-结合，等等)的结合或不结合。结合混杂性的概率表征可包括评估检测在特定结合条件或结合条件范围下亲和试剂与已知表位结合的可能性(例如，ARA -75％的结合几率；AHA-81％的结合几率；AKA-0.05％的结合几率；ADA-61％的结合几率，等等)。如果测得亲和试剂具有与表位结合的阈值可能性，则可以认为该亲和试剂结合该表位。如果测得亲和试剂在给定条件下具有约40％、45％、50％、55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％、99.5％、99.9％或高于99.9％的与表位结合的几率，则可以认为该亲和试剂结合该表位。如果测得亲和试剂在给定条件下具有至少约40％、45％、50％、55％、60％、65％、70％、75％、80％、85％、90％、95％、96％、97％、98％、99％、99.5％、99.9％或高于99.9％的与表位结合的几率，则可以认为该亲和试剂结合该表位。如果测得亲和试剂在给

定条件下具有不超过约99.9％、99.5％、99％、98％、97％、96％、95％、90％、85％、80％、75％、70％、65％、60％、55％、50％、45％、40％或低于40％的与表位结合的几率，则可以认为该亲和试剂结合该表位。

[0056] 在一些情况下，这类亲和试剂的鉴定可以通过三步筛选过程来实现：1)针对与包含表位的靶标的结合的初始筛选，2)肽水平的资格筛选，以表征该亲和试剂与表位和/或表位变体的结合；以及3)蛋白质水平的筛选，以确认步骤2中所见的结合表征。在一些情况下，步骤3可以在步骤2之前或部分在步骤2之前进行。在一些情况下，可以省略步骤2，并且步骤3可能足以表征亲和试剂的结合。在一些情况下，可以省略步骤3以确定亲和试剂。

[0057] 在一些情况下，所需表位可以是肽。在一些情况下，可能需要几个不同的表位，在这种情况下，可以选择结合所需表位的亲和试剂。在一些情况下，所需表位或目标表位可包含表位家族。在一些情况下，表位家族可包含具有简并序列的表位。例如，表位家族可包含具有包含EXE的氨基酸序列的表位，其中X是任何氨基酸。还可以基于化学性质表征包含简并序列的表位家族。在一些情况下，表位家族可包含一些或所有具有总体非中性电荷(例如，RRD、DGE)、一些空间位阻氨基酸侧基(例如，WYA、FFL)、极性侧基、疏水性侧基、亲水性侧基等的表位。在一些情况下，具有简并序列的表位家族可包含具有特定化学性质排序的氨基酸序列。例如，表位家族可包括含有连续具有极性侧基氨基酸/带正电荷的侧基氨基酸/非极性侧基氨基酸(例如，TKG)的氨基酸序列的任何表位。在一些情况下，所需表位或目标表位或表位家族可以被称为X。在一些情况下，该表位是不连续的表位。例如，表位可包含肽或蛋白质一级结构中的连续或邻近氨基酸序列。在另一个实例中，表位可包含每第二个氨基酸残基。在另一个实例中，表位可包含在肽或蛋白质二级、三级或四级结构中彼此邻近的几个氨基酸残基，即使这些残基在蛋白质序列或一级结构中不邻近。在一些情况下，该表位是连续的表位。在一些情况下，所需表位X为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或超过20个氨基酸的短氨基酸序列。在一些情况下，所需表位X为至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或超过20个氨基酸的短氨基酸序列。在一些情况下，所需表位X为不超过20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3个或少于3个氨基酸的短氨基酸序列。在一些情况下，X包含几个不同的短氨基酸序列。在一些情况下，X包含一个或多个不连续的短氨基酸序列。例如，X可包含例如C-KYW，其中半胱氨酸残基在该肽的二级或三级结构中的赖氨酸、色氨酸或酪氨酸的近侧。在一些实施方案中，所需表位X是三氨基酸序列$X_1X_2X_3$。可以通过筛选与包含所需表位的靶标结合的亲和试剂来鉴定结合多种序列环境中的该所需表位的亲和试剂。

[0058] 靶标可包括包含所需序列X的肽。在一些情况下，靶标是均为序列X的肽的池。在一些实施方案中，靶标可包含序列αXβ的肽的池，其中X是所需表位，并且侧翼结构域α和β可以是零个、一个、两个、三个、四个或多于四个氨基酸的任何序列。例如，如果所需表位X是AAA，则在靶肽中可以发现的序列的实例可以包括：AAAAA、AAAAC、CAAAA、CAAAC和CAAAD。在一些情况下，α和β可以各自是任何单个氨基酸。在一些情况下，侧翼结构域α和β中的至少一个可以是2、3、4、5、6、7、8、9、10个或超过10个氨基酸。在一些情况下，α和β中的至少一个可以包含连接体或间隔体。该连接体或间隔体可以是本领域已知的任何连接体或间隔体。在一些情况下，该连接体是氨基酸连接体。在一些情况下，该连接体是聚乙二醇(PEG)或PEG聚合物链。该PEG链可以由2、3、4、5、6、7、8、9、10、12、14、16、18、20、22、24、26、28、30、34、36、38、40、

42、44、46、48、50个或超过50个PEG部分组成。在一些情况下，该连接体可以是碳链。在一些情况下，该连接体或间隔体可包含末端标签，如His-标签。所述肽还可包含N末端或C末端修饰，例如加帽。在一些情况下，可以修饰所述肽以除去电荷，例如，末端酰胺化(C-末端)或乙酰化(N-末端)。在一些情况下，所述αXβ肽可包含非天然存在的氨基酸。在一些情况下，所述αXβ肽可以用连接体和官能团修饰。例如，该分子可以具有结构F-L-αXβ，其中F是官能团，而L是连接体。在其他情况下，该分子可以具有结构αXβ-L-F，其中F是官能团，而L是连接体。在一些情况下，α和β可以各自为相同的氨基酸，例如甘氨酸，或者可以各自为一个或多个相同的氨基酸，例如多个甘氨酸残基。在一些实施方案中，可以修饰残基以改变其适体原性。例如，可以通过以下方式改变残基：添加正电荷；添加负电荷；添加疏水基团；修饰以便添加糖；或其他修饰以便加化学多样性。

[0059] 在一些情况下，靶标可包括含有重复系列表位的肽，例如XX、XXX、XXXX等，其中X是N-聚体表位(例如，3-聚体)。例如，对于具有KYW序列的X，肽可包含KYW、KYWKYW、KYWKYWKYW等的序列。在一些情况下，靶标可包含带有侧翼序列α和β的重复系列的表位X，例如αXXβ、αXXXβ、αXXXXβ等。在一些情况下，α和β可以是单个氨基酸。在一些情况下，包含重复表位X的肽可用作从亲和试剂池中初始富集对表位X具有特异性的亲和试剂的靶标。在初始富集对X具有特异性的亲和试剂后，可以测试剩余的亲和试剂池中不包含表位X的重复序列的靶标。包含给定N-聚体的重复序列的肽可以重复该序列约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20次或超过20次。包含给定N-聚体的重复序列的肽可以重复该序列至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20次或超过20次。包含给定N-聚体的重复序列的肽可以重复该序列不超过约20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3次或少于3次。

[0060] 在一些情况下，靶标可包含在肽序列内的几个位置处含有表位X的肽。例如，靶标可包含序列αXβXγXδ，其中α、β、γ和δ各自独立地为间隔体、间隔序列、连接体或连接序列。间隔体、间隔序列、连接体和/或连接序列可以是不同的长度。间隔序列或连接序列可包含一个或多个氨基酸残基。间隔序列或连接序列可以被选择为具有很少的或没有折叠或微结构，和/或具有很小的与表位X相互作用的亲和力。间隔体或连接体可包含PEG链或其他官能团。间隔体或连接体可以被选择为具有很少的或没有折叠或微结构，和/或具有很小的与表位X相互作用的亲和力。含有表位X的肽可以在肽链中约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或超过20个位置处含有表位X。含有表位X的肽可以在肽链中至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或超过20个位置处含有表位X。含有表位X的肽可以在肽链中不超过约20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5、4、3个或少于3个位置处含有表位X。

[0061] 可以使用本领域已知的任何方法来合成肽。存在几种用于肽合成的商业平台，如MultiPep Rsi合成仪(Intavis，德国)。可以使用液相或固相方法合成肽。合成的肽可以使用任何已知的肽分析方法来验证。例如，可以使用以下方法来验证肽：质谱法、基质辅助激光解吸/电离飞行时间质谱法(MALDI-TOF)、基质辅助激光解吸/电离、AMS(加速器质谱法)、气相色谱-MS、液相色谱-MS、电感耦合等离子体质谱法(ICP-MS)、同位素比质谱法(IRMS)、离子迁移谱-MS、串联MS、热电离-质谱法(TIMS)或火花源质谱法(SSMS)。合成肽的浓度也可以通过光谱法来评估。肽合成反应和验证的实例在实施例5中提供。

[0062] 图1示出了根据一些实施方案，用于选择亲和试剂的固定的靶标，以及包含该靶标的肽的示例性列表。在图1的示例中，所需表位X具有氨基酸序列AAA，并且靶标的肽包含序列αAAAβ，其中侧翼结构域α和β各自为单个氨基酸。在该示例中，靶标包含400种不同的肽，代表αAAAβ的每种可能的序列，其中α和β各自为单个氨基酸。图1的示例不应被视为限制可能的实施方案。例如，所描述的一个或多个靶标可以容易地配置有多于或少于三个氨基酸的表位，例如αAAAAβ，或者靶标可包含多于一个氨基酸，例如γαAAAAβδ，其中α、β、γ和δ各自是单个氨基酸。

[0063] 这样，对于任何给定的具有侧翼氨基酸序列α和β的N-聚体表位，组成池的靶序列的总数可以被计算为$20^{(L\alpha+L\beta)}$，其中L_α和L_β分别是整数数字长度的包含α和β的氨基酸。例如，对于3-聚体表位，包含5-聚体池的靶标可包含400种不同的序列(α的20种可能性和β的20种可能性，其中α和β各自为单个氨基酸)。在一些情况下，靶标可包含长于5个氨基酸的肽的池，其中α和β中的每一个或两者可包含两个或更多个氨基酸。在一些情况下，α和β之一可包含零个氨基酸，而α或β中的另一个可包含一个或多个氨基酸。在一些情况下，靶标可以包含序列X的肽，而没有另外的氨基酸。在一些情况下，α、X或β中可能有超过20个氨基酸残基，例如非天然的或修饰的氨基酸。

[0064] 在一些情况下，可以用选自形式αXβ的所有序列的池中的不同靶标进行多轮选择，其中α和β各自是氨基酸序列或单个氨基酸，并且X是包含含有至少一个氨基酸的氨基酸序列的肽表位。在一些情况下，αXβ可包含具有已知程度的化学复杂性或化学多样性的目标表位。化学复杂性或化学多样性可包括均一的或不均一的化学复杂性或化学多样性。化学复杂性或化学多样性可以被定义为具有产生特定化学性质如电荷、空间大小、疏水性、亲水性、亲核性、亲电性、极性、酸性或碱性的部分、官能团、侧链、侧基、间隙(interstitial)基团、间隙元件或其他化学实体。如果α或β的氨基酸序列包含具有相似化学性质的氨基酸，则侧翼结构域α或β可包含均一的化学复杂性或化学多样性。例如，由于氨基酸侧基的相似化学性质，诸如GAG或VIL等氨基酸序列可包含具有均一的化学复杂性或化学多样性的侧翼结构域序列。如果α或β的氨基酸序列包含具有不相似的化学性质的氨基酸，则侧翼结构域α或β可包含不均一的化学复杂性或化学多样性。例如，由于氨基酸侧基的不相似的化学性质，诸如GEW或FRE等氨基酸序列可包含具有不均一的化学复杂性或化学多样性的侧翼结构域序列。在一些情况下，如果侧翼结构域α或β的化学复杂性或化学多样性与X的氨基酸序列的化学复杂性或化学多样性相似，则目标表位αXβ可包含均一的化学复杂性或化学多样性。在一些情况下，化学复杂性或化学多样性可改变亲和试剂结合的可能性。

[0065] 在一些情况下，靶序列X可以嵌入更长的序列中。例如，靶序列X可以嵌入15-聚体中。靶序列X可以嵌入15-聚体内的任何位置，例如在三氨基酸靶序列X的情况下，靶序列X可以在15-聚体的第1、2、3、4、5、6、7、8、9、10、11、12或13位开始。包含嵌入的靶序列的肽可以在溶液中合成，或者可以在芯片如PEPperPRINT芯片或其他肽阵列上合成。在一些实施方案中，包含嵌入的靶序列的肽可以结合或合成到单分子蛋白质阵列上。可以选择更长的序列以形成二级结构，或者缺乏二级结构。这类二级结构的实例包括α螺旋、β折叠、脯氨酸弯曲、转角、环和半胱氨酸桥。在一些情况下，更长的序列可包含非天然存在的氨基酸或其他基团。

[0066] 在一些情况下，所需表位或目标表位X可以装饰在更大的分子或结合基底如蛋白

质或肽上。例如，所需表位或目标表位X可以通过共价或非共价连接体(例如NHS-马来酰亚胺连接体)附接到更大的分子或结合基底上。表位X可以通过连接体与氨基酸侧链的结合而与蛋白质或肽连接。通过将表位添加到蛋白质或肽序列的末端部分，可以将表位X添加到蛋白质或肽上。蛋白质或肽可包含含有表位X的多个拷贝的尾。用所需表位或目标表位X装饰的更大分子或结合基底可以在用表位X装饰之前或之后附加至表面或基底(例如，磁珠、流动池等)上。

[0067] 初始选择步骤可包括针对包含所需表位的靶标筛查亲和试剂文库。亲和试剂文库可包含具有随机序列或具有与已知蛋白质结合适体序列相似的序列的DNA、RNA或肽适体。在一些情况下，亲和试剂文库将包含非规范或非天然核苷酸(例如，尿嘧啶、LNA、BNA、PNA)。在一些情况下，适体文库可以是商业文库。在一些情况下，适体文库可以从研究所、大学、学术中心或研究中心获得。在一些情况下，文库可以包括珠子文库。在一些情况下，适体文库可以从已知序列的文库或从随机序列生成。在一些情况下，可以基于先前针对相关或不相关的所需表位或目标表位或表位家族进行的选择的结果来设计适体文库。在一些情况下，适体文库可以包含具有特定结构的适体，例如茎环文库。在一些情况下，该适体文库可以包含可切换的适体——可在两种构象之间切换的适体。例如，适体可能需要金属离子辅因子来形成第一构象，加入螯合剂如EDTA或EGTA，螯合金属离子并使适体变为不同的构象。可以用来诱导适体切换的其他因素包括光、pH、温度、溶液离子强度、磁场和电流。

[0068] 在一些情况下，适体可包含寡核苷酸。在一些情况下，适体可包含多于一个寡核苷酸。适体可包含长度约为5、10、15、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100个或超过100个核苷酸的寡核苷酸。适体可包含长度至少约为5、10、15、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100个或超过100个核苷酸的寡核苷酸。适体可包含长度不超过约100、99、98、97、96、95、94、93、92、91、90、89、88、87、86、85、84、83、82、81、80、79、78、77、76、75、74、73、72、71、70、69、68、67、66、65、64、63、62、61、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、10、5个或少于5个核苷酸的寡核苷酸。

[0069] 适体可包含各种区域、部分或组分，如结合区域、非结合区域、衔接子、条形码和标记的区域。适体区域、部分或组分可包含长度约为5、10、15、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100个或超过100个核苷酸的寡核苷酸。适体区域、部分或组分可包含长度至少约为5、10、15、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、

74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100个或超过100个核苷酸的寡核苷酸。适体区域、部分或组分可包含长度不超过约100、99、98、97、96、95、94、93、92、91、90、89、88、87、86、85、84、83、82、81、80、79、78、77、76、75、74、73、72、71、70、69、68、67、66、65、64、63、62、61、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、10、5个或少于5个核苷酸的寡核苷酸。

[0070] 在一些情况下，适体文库可包含随机序列。在一些情况下，适体文库可包含合理设计的文库。在一些情况下，适体文库可包含随机序列或合理设计的序列，以及存在于该文库所有适体上的恒定区。在一些情况下，文库的适体可包含序列$S_1S_2S_3$，其中S_1和S_3包含该文库中所有适体共有的恒定序列，并且S_2是对每个适体可能为唯一的可变序列。在一些情况下，适体文库可以被设计为使得所有适体都包含序列$S_1S_2S_3$，其中S_1和S_3包含适用于下一代测序反应的衔接子。例如，S_1和S_3可包含Illumina衔接子，使得文库的适体可以直接在Illumina测序反应中使用，而无需进一步修饰。例如，S_1可包含序列5'-CAAGCAGAAGACGGCATACGAGAT-3'，或者S_3可包含序列5'-GTGTAGATCTCGGTGGTCGCCGTATCATT-3'。在一些情况下，S_3可以位于适体的3'末端，并且S_3可能不包含用于测序的衔接子。在这样的情况下，S_3可以具有最小的长度，以允许添加衔接子，但是一旦添加了衔接子，S_3-衔接子序列就可以包含限制位点，使得可以去除衔接子而留下S_3序列。

[0071] 在一些情况下，S_1和S_3中的至少一个也可以包含条形码。条形码对于选择目标、选择条件、实验日期、来源文库或多孔板的孔可能是唯一的。在一些情况下，条形码对于多孔板的孔可能是唯一的，使得可以在多孔板中进行多个不同的选择实验，并且可以通过对每个孔中的适体的条形码进行测序来评估孔的交叉污染。

[0072] 针对靶标筛查适体文库可以通过本领域已知的任何方法进行。在一方面，可以将靶标固定在固体支持物上，并且可以在允许以低特异性结合适体的条件下添加适体。未结合的适体可以通过一系列严格性逐渐提高的洗涤从靶标中洗出。筛选方法可包括每个富集循环约1、2、3、4、5、6、7、8、9、10个或超过10个洗涤步骤。筛选方法可包括每个富集循环至少约1、2、3、4、5、6、7、8、9、10个或超过10个洗涤步骤。筛选方法可包括每个富集循环不超过约10、9、8、7、6、5、4、3、2次或少于2次洗涤。可以对通过洗涤步骤仍保持与靶标结合的适体进行测序和扩增，以用于进一步的选择轮次，或用于设计具有高序列相似性的其他适体。可以重复数轮靶标结合、洗涤、测序和扩增，或设计新的适体，直到生成具有所需特异性和结合亲和力的适体。还可以利用各自包含适体的多个拷贝的珠子，使用基于珠子的方法筛查适体文库。还可以使用基于阵列的方法筛查适体文库，例如通过将文库的每个适体的多个拷贝点样到阵列上，然后评估靶标所结合的斑点。适体文库也可以使用颗粒展示方法来筛查。在一些实施方案中，适体文库可以使用单分子蛋白质阵列来筛查。

[0073] 可以以任何可想到的配置将亲和试剂暴露于目标池。在第一种情况下，可将目标池固定到基底上，其中亲和试剂可以随后暴露于目标池以测量结合。在第二种情况下，可将亲和试剂池固定到基底上，其中蛋白质靶标可以随后暴露于亲和试剂池以测量结合。在第三种情况下，靶标和亲和试剂可以在游离于溶液中的情况下接触，然后固定到基底上以测量结合。在第三种情况的第一个实施方案中，亲和试剂和肽可以在溶液中接触一段时间，然后固定到靶标上，使得肽与基底连接或结合。蛋白质固定后，亲和试剂可以通过洗脱步骤从

肽上解离，随后被捕获用于测序和/或扩增。在第三种情况的第二个实施方案中，亲和试剂和肽可以在溶液中接触一段时间，然后固定到靶标上，使得亲和试剂与基底连接或结合。亲和试剂固定后，蛋白质可以通过洗脱步骤从亲和试剂上解离。固定的亲和试剂随后可以从基底上释放出来用于测序和/或扩增。使亲和试剂与靶标池接触的方法可涉及一个或多个漂洗、洗涤或分离步骤，以在测量结合之前去除任何未结合的或弱结合的亲和试剂或肽。在一些情况下，在从基底上洗脱亲和试剂之前，可以从基底上释放或洗脱与基底结合的肽。

[0074] 在一些情况下，可以例如通过将亲和试剂的结合拷贝的数目与可用于结合的靶肽的数目进行比较，来测量已鉴定的亲和试剂所结合的靶标池的百分比。在一些实施方案中，亲和试剂可以与约0.00001％、0.0001％、0.001％、0.01％、0.1％、1％、10％、20％、30％、40％、50％、60％、70％、80％或超过80％的包含靶标的肽结合。在一些实施方案中，亲和试剂可以与至少约0.00001％、0.0001％、0.001％、0.01％、0.1％、1％、10％、20％、30％、40％、50％、60％、70％、80％或超过80％的包含靶标的肽结合。在一些实施方案中，亲和试剂可以与不超过约80％、70％、60％、50％、40％、30％、20％、10％、1％、0.1％、0.01％、0.001％、0.0001％、0.00001％、0.000001％或少于0.000001％的包含靶标的肽结合。另外，一旦鉴定并选择了特定的亲和试剂，就可以验证该亲和试剂。在一些实施方案中，可以针对含有表位(亲和试剂被表征为与之结合)的多个序列来验证所选择的亲和试剂。在一些实施方案中，可以通过针对单分子阵列上的多个肽或蛋白质序列评估所选择的亲和试剂来验证所选择的亲和试剂。

[0075] 在一些情况下，可以测量与所需表位或目标表位结合的亲和试剂池的百分比，例如通过将检测到的独特亲和试剂序列的数目与已知的序列池进行比较。在一些实施方案中，表位或表位家族可以与亲和试剂池的约0.000000001％、0.00000001％、0.0000001％、0.000001％、0.00001％、0.0001％、0.001％、0.01％、0.1％、1％、10％、20％、30％、40％、50％、60％、70％、80％或超过80％结合。在一些实施方案中，表位或表位家族可以与亲和试剂池的至少约0.000000001％、0.00000001％、0.0000001％、0.000001％、0.00001％、0.0001％、0.001％、0.01％、0.1％,0.00001％、0.0001％、0.001％、0.01％、0.1％、1％、10％、20％、30％、40％、50％、60％、70％、80％或超过80％结合。在一些实施方案中，表位或表位家族可以与亲和试剂池的不超过约80％、70％、60％、50％、40％、30％、20％、10％、1％、0.1％、0.01％、0.001％、0.0001％、0.00001％、0.000001％或少于0.000001％结合。

[0076] 图2示出了根据一些实施方案的亲和试剂选择过程的流程图。首先，获得靶标并将其固定在固体支持物上。然后使靶标暴露于亲和试剂文库。在这个示例中，亲和试剂是适体。从靶标上洗去未结合的适体，并洗脱剩余的适体。在一些情况下，可以使用PCR扩增洗脱的适体，然后可以去除反义链以再生ssDNA适体。在一些情况下，可以使用生物素标记的反向引物扩增洗脱的适体，从而可以使用链霉亲和素标记的珠子去除反义链。在一些情况下，可以使用磷酸化的反向引物扩增洗脱的适体，从而可以使用λ外切核酸酶去除反义链。在一些情况下，洗脱的适体的等分试样或扩增并再生的适体的等分试样可以在合并测定中使用，以确定总靶标结合。随着朝向优化池的多轮选择进展，洗脱的适体池或扩增并再生的适体池的靶标结合应随着初始选择轮次而增加。在一些情况下，过多轮次的选择可导致结合减少。适体与靶标的结合可以使用本领域已知的任何方法来测量。例如，适体与靶标的结合可以通过ELISA测定、ELONA测定、斑点印迹测定、表面等离子体共振、生物层干涉量度法或

MicroScale热泳来测量。在一些情况下，可以以保留适体的方式对洗脱的适体进行测序，或者可以对洗脱的适体的等分试样进行测序。在该过程的一些实施方案中，可以使用测序结果来预测单个适体或一小组适体是否高度富集，从而指示强结合，并且应选择以供进一步筛选，或者是否有许多不同的适体显示轻度富集，在这种情况下，洗脱的适体可以扩增并重新施加至靶标。可以重复这些步骤，直到产生具有所需结合亲和力的适体。在随后的洗涤步骤中可以增加洗涤步骤的严格性。可以选择洗涤步骤的长度，以获得具有慢的解离速率(off rate)的亲和试剂。可以优化洗脱步骤中使用的溶液的组成以选择从靶标上快速解离的适体，例如，洗脱步骤中使用的溶液可导致pH或游离离子浓度的变化。

[0077] 本公开的方法可以利用一种或多种缓冲液、溶液或混合物来表征亲和试剂对靶标的结合强度和/或结合特异性。缓冲液、溶液或混合物可用于储存、调节肽或蛋白质靶标(例如，变性缓冲液)，用于洗涤未结合的或弱结合的亲和试剂，用于将亲和试剂与靶标结合，用于从基底上洗脱结合的蛋白质，或者用于从肽上洗脱结合的亲和试剂。在一些情况下，缓冲液、溶液或混合物可以是水性的或非水性的。缓冲液、溶液或混合物可包含常见试剂，如Tris、磷酸盐缓冲液或盐水溶液。缓冲液或溶液可包含其他试剂，如盐(例如，NaCl、KCl、$MgCl_2$)、去污剂、离液剂、螯合剂或表面活性剂。示例性缓冲添加剂可包括吐温20、Triton X-100、十二烷基硫酸钠、乙腈、胆酸盐、脱氧胆酸盐、毛地黄皂苷、CHAPS、脲、氯化胍、EDTA和EGTA。可以选择缓冲液或溶液来改变靶标的物理或化学性质。例如，可以选择具有足够酸或碱强度的缓冲液或溶液，以便对氨基酸侧链胺基团(例如，组氨酸、精氨酸、赖氨酸、丝氨酸、苏氨酸、天冬酰胺、谷氨酰胺、色氨酸)进行质子化，或对氨基酸侧链羧酸基团(例如，天冬氨酸、谷氨酸)进行去质子化。缓冲液或溶液可改变靶标的二级或三级结构，而不会使蛋白质变性。例如，缓冲液或溶液可导致构象变化，从而使先前未暴露的表位暴露于缓冲液或溶液。缓冲液或溶液可以氧化或还原蛋白质的部分(例如，二硫键)。

[0078] 在一些情况下，可以将缓冲液选择为用于亲和剂筛选方法的选择剂。可以选择亲和试剂筛选方法中缓冲液的性质，以增加或降低亲和试剂结合的可能性。例如，可以选择具有特定离子强度或存在去污剂或离液剂的缓冲液。可以由于其在采用亲和试剂的下游测定中的应用而将缓冲液选择为选择剂。例如，采用本公开的亲和试剂的蛋白质表征或鉴定方法可能需要结合缓冲液、洗涤缓冲液或检测缓冲液。由于在具有与蛋白质表征测定中使用的那些相似或相同的组成的缓冲液的存在下具有阈值结合水平，因此亲和试剂可以通过上文和下文描述的方法来选择。

[0079] 缓冲液或溶液的特征可在于特定的离子或盐强度。缓冲液或溶液的离子或盐强度可以是至少约10μM、50μM、100μM、250μM、500μM、750μM、1mM、5mM、10mM、25mM、50mM、100mM、250mM、500mM、1M、2M、5M或超过5M。缓冲液或溶液的离子或盐浓度可以不超过约5M、2M、1M、500mM、200mM、100mM、50mM、25mM、10mM、5mM、1mM、750μM、500μM、250μM、100μM、50μM、10μM或低于10μM。缓冲液或溶液的特征可以是具有特定的pH。缓冲液或溶液的pH可以是约1、2、3、4、5、6、7、8、9、10、11、12或13。缓冲液或溶液的pH可以是至少约1、2、3、4、5、6、7、8、9、10、11、12、13或超过13。缓冲液或溶液的pH可以不超过约13、12、11、10、9、8、7、6、5、4、3、2、1或低于1。缓冲液可包含至少约0.01wt％、0.05wt％、0.1wt％、0.5wt％、1wt％、2wt％、5wt％、10wt％或超过10wt％的量的组分，如盐、去污剂、离液剂或表面活性剂。缓冲液可包含不超过约10wt％、5wt％、2wt％、1wt％、0.5wt％、0.1wt％、0.05wt％、0.01wt％或低于0.01wt％

的量的组分，如盐、去污剂、离液剂或表面活性剂。

[0080] 本公开的缓冲液、溶液或混合物可在特定温度下提供或在测定或过程期间加热和/或冷却至特定温度。可以改变缓冲液、溶液或混合物的温度以优化特定过程，例如亲和试剂结合、洗涤或亲和试剂洗脱。缓冲液、溶液或混合物的温度可以是约-80℃、-70℃、-60℃、-50℃、-40℃、-30℃、-20℃、-10℃、0℃、4℃、10℃、15℃、20℃、25℃、30℃、35℃、40℃、45℃、50℃、55℃、60℃、65℃、70℃、75℃、80℃、85℃、90℃、95℃或高于95℃。缓冲液、溶液或混合物的温度可以是至少约-80℃、-70℃、-60℃、-50℃、-40℃、-30℃、-20℃、-10℃、0℃、4℃、10℃、15℃、20℃、25℃、30℃、35℃、40℃、45℃、50℃、55℃、60℃、65℃、70℃、75℃、80℃、85℃、90℃、95℃或高于95℃。缓冲液、溶液或混合物的温度可以不超过约95℃、90℃、85℃、80℃、75℃、70℃、65℃、60℃、55℃、50℃、45℃、40℃、35℃、30℃、25℃、20℃、15℃、10℃、4℃、0℃、-10℃、-20℃、-30℃、-40℃、-50℃、-60℃、-70℃、-80℃或低于-80℃。

[0081] 靶标可包含可测试亲和试剂针对它的结合的肽、蛋白质的蛋白质片段。靶标可以是天然蛋白质或蛋白质片段。靶标可以是合成的肽、蛋白质或蛋白质片段。可以获得靶标，然后结合或固定在基底上。靶标可以直接在基底上合成。

[0082] 在一些情况下，靶标可以固定在珠子上。在一些情况下，靶标可包含不同序列的集合，并且可以固定到珠子上，使得每个珠子包含两个或更多个不同的序列，或者每个珠子包含单个靶序列。在一些情况下，靶标可以固定到珠子上，使得每个珠子包含来自所有可能序列的不同子集的靶序列。例如，第一珠子可包含序列G-X-G的靶标，其中X为希望与其结合的最终序列，并且G为甘氨酸，而第二珠子可包含一个或多个选自包含αXβ的序列的靶标，其中α和β各自为单个氨基酸，并且α和β中的至少一个不是甘氨酸。在另一个实例中，第一珠子可包含序列G-X-G的靶标，其中X为希望与其结合的最终序列，并且G为甘氨酸，而第二珠子可包含一个或多个选自包含αXβ的序列的靶标，其中α和β各自为任何单个氨基酸。在一些情况下，所述珠子可以是磁珠。

[0083] 在一些情况下，可以使用连接体将靶标固定在诸如珠子的固体支持物上。连接体的实例包括短氨基酸序列，如GGG、AGA，或与用于靶序列组合物的左旋氨基酸相反的右旋氨基酸的短序列。在一些情况下，使用的连接体可以是聚乙二醇(PEG)连接体。PEG的结构通常表示为$H-(O-CH_2-CH_2)_n-OH$。在一些情况下，可以选择PEG连接体，以使得PEG连接体足够长，从而使固定的靶标对于待筛选或选择的亲和试剂而言是可及的。在一些情况下，可以选择PEG连接体，以使得PEG连接体足够短，从而避免PEG连接体自身折叠并降低靶标对于待筛选或选择的亲和试剂而言的可及性。在一些情况下，PEG连接体可包含2至32、4至12、4至10或4至8的聚合物链长。在一些情况下，所使用的PEG连接体可以是PEG-2、PEG-3、PEG-4、PEG-5、PEG-6、PEG-7、PEG-8、PEG-9、PEG-10、PEG-11、PEG-12、PEG-13、PEG-14、PEG-15、PEG-16、PEG-17或PEG-18。在一些情况下，PEG连接体可以是PEG-4或PEG-8。

[0084] 在一些情况下，固体支持物上靶标的密度可影响所选亲和试剂的结合。例如，高靶标密度可导致协同结合。在一些情况下，可以优化固体支持物上靶标的密度或支持物上靶标的浓度，以促进具有所需性质的亲和试剂的选择。在一些情况下，固体支持物上靶标的密度或支持物上靶标的浓度可以根据靶标的性质——例如靶标的电荷而变化。在一些情况下，可以将带正电荷的靶标以低于不带电荷或带负电荷的靶标的浓度施加到固体支持物上。在一些情况下，可以将带正电荷的靶标以高于不带电荷或带负电荷的靶标的浓度施加

到固体支持物上。

[0085] 在其他方面，初始选择步骤可包括将文库的亲和试剂固定在固体支持物上并添加标记的靶标。该固体支持物可以是载玻片、珠子、磁珠、流动池内的表面。文库的亲和试剂可以作为单拷贝固定，或者可以以池的形式固定。例如，单个亲和试剂的多个拷贝可以被固定在固体支持物的区域上，而其他亲和试剂的多个拷贝被固定在所述固体支持物的其他区域上。在一些情况下，文库的亲和试剂可以用衔接子修饰，并与寡核苷酸包被的固体支持物杂交。对于基于寡核苷酸的亲和试剂，然后可以利用簇扩增来局部扩增固体支持物上的每个适体。在一些情况下，固体支持物是载玻片。在一些情况下，固体支持物是流动池。在一些情况下，固体支持物是适合于荧光成像的流动池。在一些情况下，固体支持物是磁珠或多个磁珠。多个磁珠中的每个磁珠可以包被在单个不同的亲和试剂的多个拷贝中，使得每个磁珠被彼此不同的亲和试剂包被。

[0086] 在其他方面，亲和试剂的初始选择步骤可包括将靶标注射到能够产生针对该靶标的抗体的宿主动物中。该宿主动物可以是能够产生抗体的任何动物，例如兔、山羊、小鼠、大鼠、马、牛、绵羊、骆驼或驴。可以从宿主动物的血清中回收抗体。来自宿主动物的血清可以原样使用，或者可以从血清中纯化抗体。纯化抗体的方法包括物理化学分级分离、类别特异性亲和力或抗原特异性亲和力。类别特异性亲和力可涉及抗体与对免疫球蛋白具有特异性亲和力的固定化生物配体的结合。抗原特异性亲和力可涉及使用固定化的靶标拉下结合该靶标的抗体。可以对纯化的抗体进行测序以鉴定来自血清的抗体，然后可以在体内或体外合成所鉴定的抗体。

[0087] 在一些情况下，不是从经免疫的宿主动物的血清中提取抗体，而是可以提取脾脏并且可以使脾细胞无限增殖化。使脾细胞无限增殖化的一种方法可以是将细胞与骨髓瘤细胞融合以形成杂交瘤。可以从杂交瘤中分离出单个克隆，每个克隆将产生一种单克隆抗体。可以使用与上述相似的方法针对靶标筛选不同的单克隆抗体。简言之，可以将靶标固定在固体支持物上，并且可以在越来越严格的条件下添加抗体，以确定以所需亲和力结合的抗体。所选择的单克隆抗体的序列可以通过对蛋白质进行测序或对来自产生它的细胞系的编码序列进行测序而得到。

[0088] 在一些情况下，可以使用噬菌体展示来选择抗体、抗体片段、肽或另一种具有可变区的合适的蛋白质支架。可以获得噬菌体文库，其中每个噬菌体在其表面上表达抗体、抗体片段、肽或另一种蛋白质支架，并且编码用于生成该蛋白质或肽的序列。可以将噬菌体文库施加至肽靶标，并且可以洗去表达并编码不与靶标结合的蛋白质或肽的噬菌体。表达并编码结合靶标的抗体的噬菌体可以从靶标中洗脱出来。可以通过用噬菌体感染细胞来扩增洗脱的噬菌体，并且可以用扩增的噬菌体以所需的重复次数重复选择步骤。随后的重复可以利用严格性逐渐增加或洗涤时间逐渐增加的洗涤缓冲液。一旦完成了所需的重复次数，就可以分析选定的噬菌体，例如通过裂解该噬菌体以对编码所表达的蛋白质或肽的DNA进行测序。该序列可用来构建表达所选蛋白质或肽的细胞系。

[0089] 在一些情况下，可以使用细菌、哺乳动物细胞、昆虫细胞或酵母展示来选择抗体或其他亲和试剂。这些方法类似于以上描述的噬菌体展示，但是获得了细菌、哺乳动物细胞、昆虫细胞或酵母文库，其中每个细菌、哺乳动物细胞、昆虫细胞或酵母细胞在其表面上表达蛋白质或肽，并且编码用于生成该蛋白质或肽的序列。选择方法与针对噬菌体展示所描述

的选择方法相同。

[0090] 本公开的亲和试剂选择方法可用于任何可想到的亲和试剂，包括寡聚体、肽聚体、微型蛋白质结合物、抗体、抗体片段、包括合成或非天然组分(例如，非天然核苷酸或氨基酸)的亲和试剂及其组合。

[0091] 在一些实施方案中，可以针对许多不同的靶标平行地选择亲和试剂。例如，对于适体亲和试剂，许多不同的靶标可以固定到阵列上的不同位置，并与SELEX选择方法一起使用。在另一个实例中，许多不同的靶标可以附接至磁珠上并与SELEX选择方法一起使用。在一些实施方案中，可以针对单分子蛋白质阵列上的靶标选择亲和试剂。

[0092] 在一个实施方案中，下一代测序平台(例如，Illumina、Nanopore)可以适合于高通量适体筛选。在一些情况下，Illumina型下一代测序平台可以适合于高通量DNA适体筛选。在一些情况下，可以通过逆转录方法对基于RNA的适体进行筛选和测序。肽适体可以通过任何蛋白质测序方法进行测序。在一些情况下，本公开的亲和试剂可以具有结构$S_1S_2S_3$，其中S_1和S_3包含适合于捕获、检测和/或鉴定亲和试剂的衔接子。例如，如上所述从SELEX选择中洗脱的适体池(其中该池中的所有适体均包含序列$S_1S_2S_3$，其中S_1和S_3包含适合于下一代测序反应的衔接子)可以在流动池中与寡核苷酸杂交。固态扩增可以用来从适体文库的每个起始适体创建适体簇。反义链可以通过多种方法从流动池中去除，例如通过使用从固体支持物上特异性切割反义链的酶。使用在SELEX阶段并入衔接子的适体文库可防止添加衔接子时适体折叠的破坏，并确保流动池中存在的适体与SELEX期间生成的输出池具有相同的长度和序列。这最大限度地增加了流动池中的适体以与它们在SELEX阶段相同的方式折叠的机会，从而允许进行比较。如果适体包含不包含衔接子序列而是在附接至衔接子时形成限制位点的S_3序列，则可以用合适的限制酶处理扩增的链，以重新生成在选择步骤中使用的适体序列。现在可以将流动池的表面覆盖在多个核酸簇中，每个核酸簇包含来自适体文库的单个适体的许多拷贝。可以使用核酸聚合酶、引物和四种不同的荧光标记的可逆终止子(A、T、C和G)在流动池上运行测序反应。这可以在流动池的每个位置处提供适体的序列。一旦已知每个适体簇的身份，就可以将一种或多种可检测地标记的肽靶标添加到流动池中并与适体簇一起孵育。孵育一段时间后，洗去未结合的靶标，并对流动池进行成像，以确定靶标所结合的位置。当进行多个阶段的孵育和洗涤步骤时，可以生成k_a、k_d和K_D值。特别是，在去除未结合的靶标的初始洗涤步骤之后，可以拍摄初始图像，以确定仍然存在哪些结合的靶标。在拍摄初始图像后，可以进行第二洗涤步骤，以去除在经过的时间中已变成未结合的目标，然后可以拍摄第二图像。可以在多个阶段上重复该过程，以便基于它们变成未结合的时间跨多个靶标计算结合性质。在一些情况下，初始洗涤步骤之后可以立即进行第二洗涤步骤。还可以使用包括例如表面等离子体共振或生物层干涉量度法(BLI)技术在内的技术进行结合性质的确定。

[0093] 在一些情况下，可以将具有不同的可辨析标记物的许多不同靶标添加到流动池中，以进行多个平行选择。实施例4中提供了该方法的一个实例。

[0094] 在一些情况下，如上所述使用Illumina型流动池的高通量适体筛选可以减少鉴定具有所需结合性质的适体所需的SELEX选择轮数。在一些情况下，在仅1、2、3、4、5、6或7轮SELEX选择后即可鉴定具有所需结合性质的适体。在一些情况下，不需要SELEX轮次，因为可以使用Illumina型流动池对合理设计的文库进行整体筛选。

[0095] 在一些实施方案中，亲和试剂选择的容易程度可能受表位的序列/结构影响。具有高“免疫原性”、“抗原性”、“适体原性”或“亲和试剂原性”的表位可能更易于针对其设计亲和试剂。例如，包含具有非常不同的化学性质或多样化的化学性质(例如，侧基电荷、侧基空间大小、侧基疏水性、侧基pKa、侧基极性)的氨基酸残基的表位可能更易于对其选择亲和试剂。诸如GGG或AAA的表位可能比诸如KWK或DCY的表位更难以针对其选择亲和试剂。在一些实施方案中，可以修饰表位，例如使用化学生物缀合，以鉴定与该修饰的表位相互作用的亲和试剂。氨基酸K、R、D、E、Y、W和C尤其容易被修饰。在一些实施方案中，当鉴定与所需表位结合的亲和试剂时，用添加了易于使用亲和试剂进行鉴定的组分的化学反应修饰DYW可能是有利的。特别是，可以修饰DYW以添加正电荷或核酸，然后可以使用针对该组分定制的亲和试剂。这样，DYW表位可能更容易凸显出来。此外，当运行平台以鉴定已被修饰的所需表位时，蛋白质样品(如血液样品)可以暴露于化学反应，该化学反应将以预期的和特定的方式修饰DYW表位，然后可以允许针对特定亲和试剂评估该蛋白质，以确定原始蛋白质样品中是否存在与具有正电荷或核酸组分的修饰DYW表位相对应的未修饰的DYW表位。在一些实施方案中，对DYW表位的化学反应可以是可逆的。在一些实施方案中，对DYW表位的化学反应可以是不可逆的。以这种方式，可以运行定制的测定，以鉴定修饰的DYW表位。

[0096] 在一些情况下，可以基于靶标表位的预测的“抗原性”或“亲和试剂原性”将潜在靶标的集合分层为不同的组。例如，在一些情况下，将高抗原性靶标和低抗原性靶标两者组合到单个平行选择步骤中，可导致针对高抗原性靶标的亲和试剂湮灭来自针对低抗原性靶标的亲和试剂的信号。预测的抗原性可以基于靶标中氨基酸残基的化学性质，其中带电荷的和较大的氨基酸预计具有较高的抗原性。含有两个或更多个相同或化学上相似氨基酸残基的靶标的抗原性可能低于含有具有非常不同化学性质的氨基酸残基的靶标的抗原性。在一些情况下，可以根据初始亲和试剂选择筛选的结果，通过实验确定不同靶标的抗原性。

[0097] 在一些情况下，可以通过阴性选择筛选方法放置亲和试剂文库或亲和试剂。阴性选择筛选方法的目的可以是鉴定未显示阈值程度的与不希望的表位、表面、材料或其组合的结合的亲和试剂。在一些情况下，阴性选择筛选可以鉴定不与基底材料(例如，玻璃、聚合物珠子、流体管)结合的亲和试剂。在一些情况下，阴性选择筛选可以鉴定不与肽或蛋白质非特异性结合的亲和试剂。例如，如果发现亲和试剂结合随机文库(例如，包含来自生物样品的天然蛋白质的文库)中所有肽或蛋白质中的至少约10％、20％、30％、40％、50％、60％、70％、80％、90％或更多，则可以不选择该亲和试剂。在一些情况下，亲和试剂可以在其表位特异性被表征之前在阴性选择筛选中进行鉴定。在其他情况下，可以表征亲和试剂的表位特异性，然后通过阴性选择筛选进行验证。

[0098] 阴性选择筛选可以包括以下步骤：a)使介质中的亲和试剂与一种或多种材料或基底接触，b)在接触后从该介质中收集未结合的亲和试剂，c)对未结合的亲和试剂进行测序，d)从所述一种或多种材料或基底收集结合的亲和试剂，e)对结合的亲和试剂进行测序，以及f)比较结合和未结合的亲和试剂的序列，以确定哪些亲和试剂对测试的材料或基底没有亲和性或具有最小的亲和性。从材料或基底收集结合的亲和试剂可包括一次或多次洗脱洗涤或漂洗。

亲和试剂的表征筛选

[0099] 可以对被鉴定为结合靶标的任何亲和试剂进行肽表征或资格筛选。亲和试剂可以

在如上所述的筛选中被鉴定，或者可以通过本领域已知的任何方法获得。在一些实施方案中，第二表征筛选可以用来确定亲和试剂对所需表位相对于其他可能的序列的结合亲和力或特异性。对于第二筛选，可以使用可能的N-聚体序列的文库来确定亲和试剂是否识别其他序列。在一些情况下，该文库包含所有可能的N-聚体序列中的至少80％、85％、90％、95％、96％、97％、98％、99％、99.9％或超过99.9％。在一些情况下，该文库包含所有可能的N-聚体序列。如果认为亲和试剂结合3-聚体，则可以选择其他序列来代表所有可能的3-聚体(20^3＝8000)。在一些实施方案中，被选择用来代表可能的目的表位的序列可以嵌入其他序列如侧翼序列内。在一些实施方案中，被选择用来代表所需表位的序列可以具有侧翼序列或化学连接体，该侧翼序列或化学连接体被添加至所需表位的与固体基底附接的一侧，使得所需表位的位置距固体基底足够远，以便允许亲和试剂接近所需表位。如果认为亲和试剂结合4-聚体，则可以选择其他序列来代表所有可能的4-聚体(20^4＝160,000)。在一个实施方案中，可以将不同N-聚体的池固定在多孔板的不同孔中，或固定在固体支持物的不同区域上。然后可以将亲和试剂与文库一起孵育，并轻轻洗去未结合的亲和试剂。亲和试剂已经与文库结合的区域可以通过任何适合于亲和试剂可视化的方法来检测。在一些实例中，亲和试剂可以通过与蛋白质和/或核酸结合的染料或试剂来可视化，或者亲和试剂可以在添加至文库之前用可检测的部分来标记。在一些情况下，亲和试剂与靶标的结合可以通过荧光共振能量转移(FRET)显微术、表面等离子体共振(SPR)、生物发光共振能量转移(BRET)、NanoBRET或生物层干涉量度法(BLI)来检测。

[0100] 在一些情况下，可以在几种不同的条件下确定亲和试剂的存在与否——例如不同的洗涤严格性，这可以通过例如使用不同的洗涤缓冲液、不同的洗涤时间、不同的洗涤温度或不同的搅动水平来实现。在一些情况下，亲和试剂的结合可以随时间反复测量，以生成结合的时间过程。在一些情况下，在不同条件下和/或在不同时间确定亲和试剂的结合可以使得能够估计亲和试剂针对靶标的K_D、k_a或k_d。亲和试剂或亲和试剂池的结合可以测量约1秒(s)、30s、1分钟(min)、2min、3min、5min、10min、15min、20min、30min、45min、1小时(hr)、2hr、3hr、6hr、12hr或超过12hr的时间段。亲和试剂或亲和试剂池的结合可以测量至少约1s、30s、1min、2min、3min、5min、10min、15min、20min、30min、45min、30小时(hr)、2hr、3hr、6hr、12hr或超过12hr的时间段。亲和试剂或亲和试剂池的结合可以测量不超过约12hr、6hr、3hr、2hr、1hr、45min、30min、20min、5min、10min、5min、3min、2min、1min、30s、1s或短于1s的时间段。亲和试剂或亲和试剂池的结合可以在一段时间内以一定间隔进行测量。时间间隔可以是规律的、模式化的或随机的。亲和试剂或亲和试剂池的结合可以以约1s、5s、10s、15s、30s、1min、5min、10min、15min、30min、1hr、2hr、3hr、6hr或超过6hr的间隔进行测量。亲和试剂或亲和试剂池的结合可以以至少约1s、5s、10s、15s、30s、1min、5min、10min、15min、30min、1hr、2hr、3hr、6hr或超过6hr的间隔进行测量。亲和试剂或亲和试剂池的结合可以以不超过约6hr、3hr、2hr、1hr、30min、15min、10min、5min、1min、30s、15s、10s、5s、1s或短于1s的间隔进行测量。

[0101] 在一些实施方案中，可以使用不同的洗涤和检测条件来鉴定具有期望的性质的亲和试剂，所述性质例如是期望的缔合速率和解离速率，或平衡缔合常数(k_a)和解离常数(k_d)。在一些情况下，期望的亲和试剂可具有慢解离速率。在以下情况下，亲和试剂表位结合反应达到平衡：

[亲和试剂]·[表位]·k_a=[亲和试剂-表位]·k_d。

在一些情况下,优选的亲和试剂的k_d值可低于约1s^{-1}、$10^{-1}s^{-1}$、$10^{-2}s^{-1}$、$10^{-3}s^{-1}$、$10^{-4}s^{-1}$、$10^{-5}s^{-1}$、$10^{-6}s^{-1}$、$10^{-7}s^{-1}$、$10^{-8}s^{-1}$、$10^{-9}s^{-1}$或$10^{-10}s^{-1}$。在一些情况下,解离速率可能比缔合速率更关键,因为可以通过增加添加到反应中的亲和试剂的量或通过延长洗涤前的孵育时间来补偿较差的缔合速率。在一些情况下,选择亲和试剂,使得解离速率足以使亲和试剂能够在一个或多个洗涤步骤后保持结合,直到完成成像步骤后。在一些情况下,亲和试剂可以保持与表位结合至少约1min、2min、3min、4min、5min、6min、7min、8min、9min、10min、11min、12min、13min、14min、15min、16min、17min、18min、19min、20min、25min、30min、35min、40min、45min、50min、55min、60min或超过约60min。

[0102] 在一些情况下,优选的亲和试剂仅与文库中的N-聚体(例如,3-聚体)之一结合。在其他情况下,优选的亲和试剂可以与2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、40、50、60、70、80、90、100种或超过100种不同的N-聚体结合。在一些实施方案中,可以以逐渐增加的严格性重复洗掉亲和试剂的步骤,并且可以在每个洗涤步骤之后检测保持结合的亲和试剂。在一些情况下,优选的亲和试剂与侧翼为任何氨基酸残基的N-聚体结合。在一些情况下,优选的亲和试剂与侧翼为所有可能氨基酸残基的子集的N-聚体(例如,3-聚体)结合。在一些情况下,针对特定目标表位或表位家族选择的亲和剂可以被表征为还结合与目标表位或表位家族不同的表位。例如,可以针对具有序列AXA的目标表位鉴定亲和试剂,其中X可包含任何氨基酸。针对AXA的亲和试剂的后续表征也可以鉴定与表位RSR、FFW和YIC的结合。在一些情况下,如果特定的混杂表位已知,则与其目标表位家族之外的表位结合的混杂亲和试剂仍可能是用于鉴定肽表位的有价值的亲和试剂。

[0103] 在另一种方法中,可以修饰或成熟化针对所需表位或目标表位或表位家族的经鉴定的亲和试剂,以生成新的靶标文库。在一些情况下,亲和试剂的修饰或成熟化可包括在所鉴定的亲和试剂的结合区序列中创建单核苷酸或多核苷酸置换的文库。在其他情况下,亲和试剂的修饰或成熟化可包括置换一个或多个非天然核苷酸或其他化学修饰,以针对相同表位创建修饰的亲和试剂的文库。可以针对所需表位或目标表位或表位家族测试修饰的或成熟化的亲和试剂文库,以鉴定具有不同结合亲和力,例如针对表位目标家族之外的表位的更强结合或不同混杂性的亲和试剂。

[0104] 在另一种方法中,可以将亲和试剂固定在固体支持物上,并且可以将所有可能的n-聚体的池添加到亲和试剂中,可以通过洗涤除去未结合的N-聚体,并且可以分析剩余的N-聚体以发现被亲和试剂结合的序列。重复的筛选步骤可以用不同严格性的洗涤步骤进行。可以使用任何合适的方法,例如质谱法或高效液相色谱质谱法来鉴定结合的N-聚体。在一些情况下,可以使用随机N-聚体池,而不是所有可能的N-聚体的池。在随机N-聚体池用于表征筛选的情况下,结合的和未结合的N-聚体均可以得到鉴定。可以指出在结合的或未结合的N-聚体池中均未代表的序列以供在随后的表征筛选中追踪。在一些情况下,可以在施加至亲和试剂之前对随机N-聚体池进行表征,以确保该池包含至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或超过99%的可能的N-聚体序列。在一些情况下,可以用随机N-聚体进行初步表征筛选,并且可以选择结合少于约30、25、20、15、10或5种N-聚体的亲和试剂以供使用所有可能的N-聚体的池进行进一步表征。

[0105] 可以进行进一步的肽资格筛选，以确定亲和试剂是否特异性结合至所需表位，或者亲和试剂是否结合至表位的亚部分。例如，被认为对3-聚体序列AKD具有特异性的亲和试剂实际上可能对二聚体序列AK具有特异性。因此，对于每种N-聚体亲和试剂，可以针对所有(N-1)-聚体的集合进行筛选步骤。在一些情况下，当已经进行了表征筛选并且被亲和试剂结合的所有N-聚体的集合为已知时，筛选可以仅限于包含在被亲和试剂结合的N-聚体内的(N-1)-聚体。例如，如果被亲和试剂结合的唯一4-聚体是AAKD，则可以针对AAK和AKD而不是所有可能的3-聚体进行筛选。如果亲和试剂确实与(N-1)-聚体结合，则可以针对(N-2)-聚体进行另外的筛选，以进一步限定该亲和试剂的特异性。在一些情况下，可以不选择与(N-1)-聚体结合的亲和试剂进行进一步筛选。在其他情况下，可以选择与不超过一种(N-1)-聚体结合的亲和试剂。在其他情况下，可以保留与超过一种(N-1)-聚体结合的亲和试剂进行进一步筛选。

[0106] 可以进行另外的肽资格筛选，以确定侧翼序列对亲和试剂对其一个或多个表位的亲和力的影响。对于该步骤，可以创建序列αXβ的肽文库(其中X是亲和试剂表位，或多个表位，并且α和β可以是任何一个或多个氨基酸)。在与单个表位结合的亲和试剂的实例中，如果α和β各自为单个氨基酸，假定设置结合的表位的序列，则无论所需表位的长度如何，存在400种可能的αXβ序列。对于亲和试剂结合两个或更多个不同表位，或者α和β可能各自为零个、一个或多于一个氨基酸的情况，可能存在许多可能的αXβ序列。在一些情况下，在该筛选步骤中使用的一组αXβ肽可以是构成用于选择亲和试剂的靶标的相同一组αXβ肽。可以使用标准筛选方法针对这组αXβ序列筛选亲和试剂。查看亲和试剂所结合的αXβ序列的列表以及相对结合亲和力可以允许确定侧翼序列对亲和试剂的结合亲和力的影响。如果观察到强烈的影响，则可以用更长的侧翼序列进行进一步的筛选，例如，如果第一个序列文库由肽αXβ组成，其中α和β各自为单个氨基酸，则可以用由γXδ组成的肽文库进行后续筛选，其中γ和δ各自包含两个氨基酸。

[0107] 图3A示出了根据一些实施方案，亲和试剂表征过程中的筛选步骤的示例性流程图。用荧光标记物标记亲和试剂，以促进亲和试剂的检测。然后将亲和试剂施加至包含所有可能的N-聚体(例如，3-聚体)序列的肽池。包含所有可能的N-聚体序列的肽池可以作为阵列固定在固体支持物上。在一些实施方案中，可以将包含所有可能的N-聚体序列的肽池固定在一个或多个单分子蛋白质阵列上。在一些实施方案中，可以将多个N-聚体序列固定在固体基底上。在一些实施方案中，可以将多个N-聚体序列固定在单分子蛋白质阵列上。在可以平行进行的另一个步骤中，然后将亲和试剂施加到包含所有可能的(N-1)-聚体序列的肽池中。鉴定了被亲和试剂结合的所有N-聚体和(N-1)-聚体肽。获得了(N+2)-聚体肽池，其包含所有可能的(N+2)-聚体肽，所述(N+2)-聚体肽包含所鉴定的结合的N-聚体和(N-1)-聚体的序列。例如，如果已知亲和试剂仅结合表位AAA，则所有可能的(N+2)-聚体肽的池由肽αAAAβ组成，如图3B所示。然后将亲和试剂施加至(N+2)-聚体肽池，并编译来自所有三个筛选步骤的数据，以提供亲和试剂的结合表征。在其他实例中，可以省略步骤或以不同的顺序进行步骤。可以进行蛋白质资格筛选，以确认亲和试剂结合蛋白质而不是肽环境中的一个或多个所需表位。可以将已知身份和序列的蛋白质固定在固体支持物上，然后暴露于亲和试剂。蛋白质可以从纯化的蛋白质储备物施加至固体支持物，或者可以通过诸如核酸可编程蛋白质阵列(NAPPA)等过程在固体支持物上合成。在一些情况下，用于该筛选的蛋白质可以

由包括亲和试剂的一个或多个预测或已知表位的蛋白质组成。在一些情况下，用于该筛选的蛋白质可以由不包括亲和试剂的一个或多个预测或已知表位的蛋白质组成。在一些情况下，用于该筛选的蛋白质可以由包括亲和试剂的一个或多个预测或已知表位的蛋白质和不包括这些表位的蛋白质组成。亲和试剂与不同蛋白质的结合可以通过本领域已知的任何方法来评估。例如，结合的亲和试剂可以通过亲和试剂特异性抗体，通过与蛋白质或核酸结合的染料，或者通过在筛选之前用可检测部分标记亲和试剂来鉴定。一旦确定了结合的和未结合的蛋白质，就可以比较序列，以确定亲和试剂与其一个或多个表位结合的频率如何，以及该结合是否受周围蛋白质序列的影响。也可以将结合结果和蛋白质序列输入机器学习算法中，以验证最可能的结合位点。也可以通过该方法确定脱靶影响。在一些情况下，如果针对表位的亲和试剂结合至少约1％、2％、3％、5％、7％、10％、15％、20％、30％、40％、50％、60％、70％、80％、90％、95％或超过95％的包含该表位的蛋白质，则可以选择该亲和试剂。在一些情况下，如果针对表位的亲和试剂结合少于约50％、40％、30％、20％、10％、5％或少于5％的不包含该表位的蛋白质，则可以选择该亲和试剂。在一些情况下，可以针对相同的表位选择多种亲和试剂。所述多种亲和试剂可以合并在一起，并用作合并的亲和试剂。

[0108] 亲和试剂可以以绝对或概率方式来表征。绝对表征可包括构建一组已知表位、表位家族、蛋白质性质或亲和试剂对其具有亲和力的其他表征参数。亲和试剂的概率表征可包括与表位或表位家族结合的可能性的依赖性或环境非依赖性测量。环境依赖性亲和剂表征可能是，例如，当肽或蛋白质处于其天然状态时，亲和试剂结合表位X的可能性为25％，而当肽或蛋白质变性时，结合表位X的可能性为75％。环境非依赖性表征可以是，例如，亲和试剂结合表位X的可能性为35％。亲和试剂表征可以在许多结合条件下发生，如温度范围、缓冲条件(例如，pH、摩尔浓度)、结合竞争物的存在、肽或蛋白质折叠状态(例如，天然的、非天然的)，以及可能影响亲和试剂结合的其他任何条件。

利用来自与已知蛋白质相互作用的结合数据表征亲和试剂

[0109] 寻找表征特定亲和试剂对构成蛋白质组(例如人类蛋白质组、酵母蛋白质组、大肠杆菌蛋白质组)的蛋白质的结合亲和力的方法是有用的。甚至当查看Uniprot(https://www.uniprot.org/proteomes/UP000005640)上的参考蛋白质组数据库中定义的约70,000个规范蛋白质序列时，也需要花费大量精力来表征一组特定亲和试剂对人类蛋白质组中的整个一组蛋白质的结合亲和力。此外，当考虑蛋白质形式(proteoform)时，可以被鉴定的人类蛋白质组内不同蛋白质形式的数目可以是数十万或数百万。这样，以能够应用于多种未知蛋白质的方式有效地表征亲和试剂的结合亲和力是有利的。在一些实施方案中，通过评估亲和试剂与已知蛋白质的相互作用来产生亲和试剂的结合亲和力。在一些实施方案中，通过评估亲和试剂与具有来源于序列数据库(如来自NCBI或Uniprot的参考蛋白质组)的序列的蛋白质的相互作用来产生亲和试剂的结合亲和力。在一些实施方案中，在本文提供的方法中使用的蛋白质序列可能没有已知的天然来源，如具有随机序列的蛋白质。在一些实施方案中，合成具有非天然序列的蛋白质可用于向模型(如本文所讨论的那些模型)中提供额外的输入。在一些实施方案中，亲和试剂对已知蛋白质内特定靶标的结合亲和力可基于该靶标在已知蛋白质内的存在和基于该靶标在已知蛋白质内的拷贝数来评估。

[0110] 因此，在一些情况下，亲和试剂，如本文公开的亲和试剂，可以通过针对已知蛋白质阵列筛选亲和试剂来表征。通过针对已知蛋白质筛选亲和试剂而生成的结合数据可用来

生成结合亲和力信息，该结合亲和力信息可用来评估在未知蛋白质内是否存在亲和剂所结合的一个表位或多个表位。特别是，每种亲和试剂可以与一个或多个表位结合。此外，每种蛋白质可以包含多个表位，每个表位可以与特定的亲和试剂结合。对于包含与特定亲和试剂结合的多个表位的蛋白质，该蛋白质可以具有特定表位的多个拷贝和/或可以具有不同表位的多个拷贝。这样，在一些情况下，特定蛋白质可以与多个亲和试剂分子结合，所述多个亲和试剂分子中的每一个可能附着在蛋白质内存在的多个表位之一上。因此，有助于表征亲和试剂并预测与蛋白质的结合的信息可适用于鉴定未知蛋白质的方法。在一些实施方案中，所生成的结合数据可以用来鉴定其序列在蛋白质数据库内得到代表的未知蛋白质。在一些实施方案中，该结合数据可用来生成特定的概况，该概况可用来表征未知蛋白质，直到鉴定出与该未知蛋白质相关的序列。特别是，从亲和试剂与已知蛋白质阵列中的蛋白质相互作用生成的结合数据可用来确定亲和试剂对已知蛋白质内的一个或多个表位的结合亲和力。继而，这可以用于评估所筛选的亲和试剂对具有所述一个或多个表位的一个或多个拷贝的蛋白质的结合亲和力。另外或备选地，亲和试剂对一个或多个表位的结合亲和力可以用于评估所筛选的亲和试剂对具有与所表征的一个或多个表位相似或相同的其他表位的蛋白质的结合亲和力。在一些实施方案中，来自针对已知蛋白质阵列中的蛋白质筛选的亲和试剂的相互作用的结合数据可以用来确定已知蛋白质阵列内一种或多种蛋白质中特定类型的表位的亲和力特性，所述表位例如是特定长度的每个可能的表位，完整蛋白质序列的特定子集内的表位，折叠蛋白质结构的特定位置的表位，预测或观察到在折叠蛋白质结构中高度可及的每个表位，或从根据经验的或计算机模拟的结合筛选中鉴定的每个表位。在一些实施方案中，来自针对已知蛋白阵列中蛋白质筛选的亲和试剂的相互作用的结合数据可以用来确定亲和试剂与已知蛋白质阵列内一种或多种蛋白质中的多个表位的结合亲和力。在一些实施方案中，来自针对已知蛋白阵列中蛋白质筛选的亲和试剂的相互作用的结合数据可以用来确定亲和试剂与具有特定表位的多个拷贝的蛋白质的结合亲和力。

[0111] 下面描述了针对已知蛋白质阵列进行亲和试剂筛选的方法。如下所述，该方法可以提供，所测试的亲和试剂结合至少一个N氨基酸表位(例如，三聚体)或N-聚体的组合。该方法的实施方案可以包括一个或多个除三聚体以外的不同长度的表位，如二聚体、4-聚体、5-聚体、6-聚体、7-聚体、更长的表位和不连续表位。在一些实施方案中，方法可以提供，所测试的亲和试剂结合至少一个二聚体或二聚体的组合。在一些实施方案中，方法可以提供，所测试的亲和试剂结合至少一个二聚体或三聚体，或二聚体或三聚体的组合。在一些实施方案中，方法可以提供，所测试的亲和试剂结合两个或更多个N-聚体，所述n-聚体选自二聚体、三聚体、4-聚体、5-聚体、6-聚体、7-聚体、长于7-聚体的N-聚体和不连续表位。在一些实施方案中，方法可以提供，所测试的亲和试剂结合三个或更多个N-聚体，所述n-聚体选自二聚体、三聚体、4-聚体、5-聚体、6-聚体、7-聚体、长于7-聚体的N-聚体和不连续表位。在一些实施方案中，方法可以提供，所测试的亲和试剂结合超过三个N-聚体，所述N-聚体选自二聚体、三聚体、4-聚体、5-聚体、6-聚体、7-聚体、长于7-聚体的N-聚体和不连续表位。

[0112] 使用本文提供的方法，可以计算蛋白质推断。在一些实施方案中，可以基于以下一种或多种考虑因素来计算蛋白质推断，如特定亲和试剂与蛋白质的结合测量结果；与蛋白质可能的候选蛋白质的蛋白质序列有关的信息；以及可以从中推导出具有每种亲和试剂与每种候选蛋白质的预期结合程度的特定置信度的预测的信息，如结合亲和力信息。提供了

本文所述的一些方法，用于产生亲和试剂的统计模型，在提供蛋白质的一级序列的情况下，可以从该模型计算出亲和试剂与该蛋白质结合的概率。

[0113] 在一些实施方案中，可以使用多种特定亲和试剂与多种已知蛋白质中的每一种的一系列结合测量结果来训练统计模型。当没有直接的经验测量可用时，该方法可以用来预测亲和试剂-蛋白质结合。在将方法应用于人类蛋白质组中的蛋白质的情况下，候选人类蛋白质的最小列表可以包含约20,000种蛋白质。此外，允许剪接变体、多态性、部分降解和另外的非人类基因组可以使候选蛋白质迅速扩充至数十万或数百万的数目。

[0114] 在一些情况下，本公开可包括通过针对具有随机序列的一组肽或其中只有子集包含靶向表位的一组肽筛选所选亲和试剂来表征所选亲和试剂的方法。在一些情况下，可以使用插补模型来估算所选亲和试剂与筛选的一组肽或蛋白质中不存在的表位(例如，包含非天然氨基酸的表位)的结合。在一些情况下，可以通过所选亲和试剂针对多种肽或蛋白质的结合测量来训练插补模型。在一些情况下，可以通过其他亲和试剂针对多种蛋白质或肽的结合测量来训练插补模型。在一些情况下，插补模型可以补充由例如氨基酸置换矩阵(例如PAM、BLOSUM)、氨基酸性质(例如电荷、疏水性、极性表面积)、基于生理化学的氨基酸距离(如Grantham距离)或其组合计算的表位相似性度量。

方法

[0115] 在一些实施方案中，获得已知蛋白质序列的阵列。在一些实施方案中，可以通过将纯蛋白质样品点样到芯片上来生成这种已知蛋白质阵列。在一些实施方案中，可以通过将蛋白质直接翻译到芯片上，例如通过使用核酸可编程蛋白质阵列(NAPPA)来生成这种已知蛋白质阵列。可以跨一个或多个特定阵列分析的已知蛋白质的数目可包含少于50个不同的蛋白质序列，可包含50个不同的蛋白质序列，可包含超过50个蛋白质序列，可包含大约100、150、200、250、300、350、400、500、600、700、800、900、1,000、2,000、3,000、5,000、10,000、15,000、20,000个或超过20,000个蛋白质序列。作为一个广义的概念，当分析更多蛋白质时，所开发的模型的准确性可以提高。在一些实施方案中，可以被分析的已知蛋白质的数目可以是约300至约3000个不同蛋白质序列；约500至约2000个不同蛋白质序列；约600至约1500个不同蛋白质序列；约600至约1000个不同蛋白质序列；或约1000至3000个不同蛋白质序列，等等。在一些实施方案中，每种已知蛋白质可以在阵列上的单个位置处以多个拷贝存在。在一些实施方案中，每种蛋白质可以在整个阵列的许多位置处以多个拷贝存在。在一些实施方案中，每种蛋白质可以存在于整个阵列的两个位置、三个位置、四个位置、五个位置、六个位置、七个位置、八个位置、超过八个位置或不同数目的位置的组合。在一些实施方案中，每种蛋白质可以在整个阵列的一个、两个、三个、四个、五个、六个、七个、八个或超过八个位置的每个位置处以多于约1000个拷贝、2000个拷贝、3000个拷贝、5000个拷贝、10,000个拷贝、20,000个拷贝、50,000个拷贝、100,000个拷贝、500,000个拷贝、1,000,000个拷贝、2,000,000个拷贝、5,000,000个拷贝、10,000,000个拷贝、100,000,000个拷贝、10亿个拷贝、100亿个拷贝、1000亿个拷贝或超过1000亿个拷贝存在。此外，在一些实施方案中，还可以通过评估特定位置的少量蛋白质，如每个位置1个拷贝、每个位置2个拷贝、每个位置3个拷贝、每个位置4个拷贝、每个位置5个拷贝、每个位置10个拷贝或每个位置超过10个拷贝来表征亲和试剂。在一些实施方案中，当每个特定位置基于光学检测或其他检测传感器可分辨地不同于另一位置时，位置可以彼此区分开。

[0116] 在一些实施方案中，可以通过使荧光标记的亲和试剂与阵列杂交并测量阵列上每个斑点处观察到的荧光，来评估一种或多种亲和试剂与阵列内的已知蛋白质的结合。在一些实施方案中，每个斑点处每种蛋白质的身份可以是已知的，从而可以使荧光(代表结合)对应于蛋白质身份。

[0117] 在一些实施方案中，N-聚体(例如，三聚体)表位可以用于评估亲和试剂与比所需肽或目标肽更长的任何肽(例如，蛋白质、蛋白质片段、肽、部分降解的蛋白质)内特定表位的结合。在一些实施方案中，可以使用三聚体表位应用本文讨论的方法。在一些实施方案中，蛋白质水平的荧光测量可以用来从单独的表位，例如从每个单独的表位，得出对荧光测量的分数贡献。在一些实施方案中，可以将蛋白质结合模拟为蛋白质序列中可能的N-聚体，例如蛋白质序列中8000种可能的三聚体各自的计数与来自这些N-聚体的分数荧光的线性组合。对于三聚体，这可以表示为：

$$F_{pr} = \sum_{t=1}^{t=8000} c_{t,pr}\beta_t$$

其中：

F_{pr}＝蛋白质pr的测量结果(例如荧光)

$c_{t,pr}$＝蛋白质pr中三聚体t的计数

β_t＝来自亲和试剂与三聚体t的结合的分数荧光

[0118] 在一些实施方案中，三聚体表位的亲和试剂针对多种蛋白质的测量结果可以形成线性方程组：

$$\vec{F} = C\vec{\beta} + \epsilon$$

其中：

$\vec{F}$

是含有对于每种蛋白质观察到的测量结果(例如荧光)的长度N列向量；

C是三聚体计数的N x 8000矩阵，其中每列是每种测量的蛋白质中特定三聚体的计数；

$\vec{\beta}$

是来自试剂与每种可能的三聚体结合的分数荧光的长度8000列向量；

∈是针对背景结合或本底噪声进行校正的标量常数。

[0119] 在一些实施方案中，本文提供的方法可包括非标准氨基酸和/或模拟不同长度的N-聚体。在已知的NAPPA或类似的结合测量的情况下，$\vec{F}$和C是已知的变量，而$\vec{\beta}$和∈的值可以通过线性回归或相关方法得出。特别是，非负最小二乘法和非负最小绝对收缩和选择算子(LASSO)回归可能非常适合于本文所述方法的考虑因素。非负最小二乘法将$\vec{\beta}$的解限制为非负值，而非负LASSO回归进一步施加了稀疏约束。当方程组欠定时，即，当所测得的独特蛋白质的数目少于独特三聚体的数目(在此示例中为8,000)时，LASSO回归特别有效。在一些实施方案中，源自每个表位的分数荧光可以用来估计特定亲和试剂与各个表位的结合特

性，如结合动力学。在一些实施方案中，可以认为相对分数荧光与亲和试剂与每个表位的相对结合亲和力成比例。在一些实施方案中，可以通过除以每个荧光团的荧光计数的数目，然后除以阵列上每个斑点的预期蛋白质分子数，将分数荧光转换为所结合的位点的分数的计算值。在实施例2中提供了对该方法的模拟，其中使用理论亲和试剂的一组结合亲和力来预测亲和试剂与720种人类蛋白质的结合，然后求解预测的结合数据，以确定该亲和试剂对每个可能的三聚体表位的亲和力。

[0120] 从对蛋白质与亲和试剂结合进行模拟的实施方案中收集的数据可以用来帮助训练从蛋白质的一级序列预测蛋白质结合亲和力的其他模型。在一些实施方案中，从对蛋白质与亲和试剂结合进行模拟的实施方案中收集的数据可以用来帮助训练从一级序列的衍生物(例如氨基酸组成、三聚体计数、预测的三维结构)预测蛋白质结合亲和力的其他模型。在一些实施方案中，模型可包括与针对不同长度的表位(例如1聚体、2聚体、4聚体等)或具有不同长度的表位的混合物所讨论的那些参数相似的参数。在一些实施方案中，模型可以使用非线性模型，例如采用将三聚体亲和力与观察到的荧光相关联的指数函数，或者其中将三聚体水平亲和力相乘以生成蛋白质水平亲和力的乘法模型。在一些实施方案中，可以使用神经网络从蛋白质序列或三聚体组成预测蛋白质结合亲和力。另外，可以使用支持向量回归模型从蛋白质序列或三聚体组成预测蛋白质结合亲和力。

[0121] 当只有蛋白质水平结合测量可用时，另一种方法可允许鉴定驱动蛋白质水平结合的N-聚体。这种方法可以在结合分子(例如，肽)具有比表位更长的序列(直到并包括蛋白质)的任何情况下工作(例如，从缀合有15-聚体的PEPperPrint芯片学习三聚体表位)。在一些情况下，亲和试剂显示对单个蛋白质具有亲和性。在没有后续实验的情况下，确定蛋白质的哪些子序列驱动亲和试剂对蛋白质的亲和力在计算上可能是棘手的。然而，如果针对数百种蛋白质获得该亲和试剂的结合测量值，并且亲和试剂与其中许多蛋白质结合，则模式可能会开始出现。在一些情况下，例如，亲和试剂可能有利于其中具有特定N-聚体(例如，三聚体)的蛋白质。本文描述的算法是从蛋白质结合测量的集合确定亲和试剂的三聚体结合概率(如果在蛋白质上看到的话，亲和试剂结合8000个可能的三聚体中的每一个的概率)的方法。更具体地，蛋白质结合测量是探针与蛋白质的结合分数。例如，如果将1,000个拷贝的单一蛋白质沉积在阵列上，并且在杂交后，亲和试剂与其中250个蛋白质结合，则结合分数将为0.25。所描述的方法可以扩展到包括其他特性或序列长度的表位。

[0122] 所述方法可以使用在三聚体结合概率上建立的亲和试剂蛋白质结合的概率模型。使用该模型，在给定蛋白质的一级序列的情况下，可以估计亲和试剂的单个分子与任何蛋白质的单个拷贝结合的概率。

[0123] 亲和试剂与蛋白质的结合概率p被建模为：

$$p = 1 - \prod_{j=1}^{8000} \theta_j^{x_j}$$

其中：

X:蛋白质底物中每个三聚体j的计数

$X = \{x_1, x_2, x_3 \cdots x_{8000}\}$，其中 $x_j \in \mathbb{Z}^*$ θ:结合模型参数。探针不与每个三聚体结合的

概率的向量

$\theta=\{\theta_1,\theta_2,\theta_3,\cdots\theta_{8000}\}$，其中$0\leqslant\theta_j\leqslant1$

亲和试剂不与蛋白质结合的概率等于1-p。

亲和试剂与任何特定三聚体j结合的概率等于$1-\theta_j$。

[0124] 在给出观察到的与单个蛋白质底物的结合结果的情况下，特定亲和试剂结合模型的可能性为：

$$L(\theta|y,X)=\left(\prod_{j=1}^{8000}\theta_j^{x_j}\right)^{1-y}\left(1-\prod_{j=1}^{8000}\theta_j^{x_j}\right)^{y}$$

其中：

y:结合结果

$$y=\begin{cases}\text{如果结合则为 }1\\ \text{如果不结合则为 }0\end{cases}$$

[0125] 对数似然为：

$$log\big(L(\theta|y,X)\big)=(1-y)\sum_{j=1}^{8000}x_j\log\theta_j+y\log\left(1-\prod_{j=1}^{8000}\theta_j^{x_j}\right)$$

[0126] 也可以计算多个结合结果的对数似然。如果针对N个单拷贝蛋白质底物观察到N个结合结果：

Y为N个结合结果的向量；

X为N个基底中的每一个的三聚体计数的N x 8000矩阵，其中任何行X_i为第i个底物的8000个三聚体中每一个的计数。

[0127] 在给出这组观测值的情况下，结合模型的对数似然θ为：

$$log\big(L(\theta|Y,X)\big)=\sum_{i=1}^{N}\left((1-y_i)\sum_{j=1}^{8000}x_{ij}\log\theta_j+y_i log\left(1-\prod_{j=1}^{8000}\theta_j^{x_{ij}}\right)\right)$$

[0128] 可以使用上述似然函数的最大化来估计亲和试剂结合模型的参数θ的$\hat{\theta}$。使用当前公式，同一蛋白质的拷贝的多个结合结果将被表示为向量Y中的单个条目和矩阵X中的行，这在使用软件计算对数似然时可导致大量内存占用。为了减少所需的计算资源，可以使用结合数据的更紧凑表示来执行相同的计算，其中结合结果被折叠为结果集合中每个独特蛋白质底物的结合和非结合事件的数目。通过这种重新公式化，用于计算的矩阵和向量的维度与数据集中独特蛋白质的数目而不是观察到的结合结果的数目相对应。

[0129] 在替代公式中：

$$log\left(L(\theta|U,B,T)\right)=\sum_{i=1}^{M}\left(u_i\sum_{j=1}^{8000}t_{ij}\log\theta_j+b_i log\left(1-\prod_{j=1}^{8000}\theta_j^{t_{ij}}\right)\right)$$

其中：

结合观察值由M个独特蛋白质组成

B为结合计数的向量，U为每个独特蛋白质的非结合计数的向量。两者的长度均为M

T为每个独特蛋白质的三聚体计数的矩阵，有M行和8000列。

[0130] 这种方法通过使用L-BFGS-B算法将函数$f\left(\hat{\theta}\right)=-log\left(L\left(\hat{\theta}\middle|U,B,T\right)\right)$最小化来估计结合模型$\hat{\theta}$，其中每个参数$\hat{\theta}$被约束为大于零且小于1。

[0131] 在给出参数估计值$\hat{\theta}$的情况下，通过提供函数$J=j\left(\hat{\theta}\right)$来直接计算f的分析梯度向量J来改善最大似然估计的速度。分析梯度为：

$$J=\boldsymbol{j}(\hat{\theta})=\left[\frac{\partial f}{\partial\widehat{\theta_1}}\quad\frac{\partial f}{\partial\widehat{\theta_2}}\quad\cdots\quad\frac{\partial f}{\partial\widehat{\theta_{8000}}}\right]$$

$$\frac{\partial\mathbf{f}}{\partial\widehat{\theta_q}}=-\sum_{i=1}^{M}\left(\frac{u_i t_{iq}}{\widehat{\theta_q}}-\frac{b_i t_{iq}\prod_{j=1}^{8000}\hat{\theta}_j{}^{t_{ij}}}{\hat{\theta}_q\left(1-\prod_{j=1}^{8000}\hat{\theta}_j{}^{t_{ij}}\right)}\right)$$

[0132] 正如所证明的，所述方法可以采用一组实验观察值中每个独特蛋白质的结合和非结合事件的单独计数而工作。然而，在只能确定或估计分数结合的情况下，可以用长度为M的向量F替换向量B，该向量F含有与数据集中M个独特蛋白质中的每一个的分数结合，并且向量U被替换为向量D＝1-F。例如，在荧光标记的亲和试剂候选物与核酸可编程蛋白质阵列(NAPPA)杂交，然后成像以测量荧光强度的情况下，F的值可以被估计为给定蛋白质的荧光强度除以在分数结合为1(所有蛋白质均结合)时观察到的荧光强度的估计值。

[0133] 如果特定的三聚体在结合数据集中根本没有观察到，仅在单个独特蛋白质序列中被观察到，或者在其蛋白质序列环境的每一个中总是与另一个特定三聚体一起被观察到，则三聚体的结合概率可能难以学习。对于这样的三聚体，可以估算结合概率，例如通过将三聚体的概率设置为与基于序列或生化特性而相似的另一个三聚体的概率相等。在一些情况下，可以通过将三聚体的概率设置为其他所有未估算的三聚体的平均结合概率来进行估算。

[0134] 上述方法可以用于学习由二聚体、三聚体、4-聚体、5-聚体等构建的结合模型。此外，它还可以应用于包含上述N-聚体的子集或混合物的模型。为此，表示每个三聚体的每个蛋白质中的计数的列(1-8000)可以被替换为表示矩阵M中每个蛋白质中每个N-聚体的计数的列。此外，参数向量θ将是模型中使用的每个N-聚体的结合概率。

亲和试剂表征的改进

[0135] 上述计算机实现的算法可以改进亲和试剂结合特异性的表征并且允许用于预测未表征的亲和试剂的结合亲和力的快速筛选方法。例如，可以使用已知的经表征的亲和试

剂池针对包含数百万肽或蛋白质的蛋白质池或肽文库测量结合。然后，后续的未表征的亲和试剂可以针对相同的蛋白质池或文库测量其结合。上述算法可以实现对未表征的亲和试剂的结合亲和力的快速表征。使用上述方法可以改进对未知结合亲和力的亲和剂的筛选和表征，从而导致将一种或多种亲和试剂加入亲和试剂池中以用于蛋白质表征或鉴定方法。

[0136] 上述计算机实现的算法可以使用单分子测量数据来运行。上述计算机实现的算法可能能够处理完整的蛋白质序列。在表征亲和试剂时，上述计算机实现的算法可能能够处理数亿或数十亿的数据点。

[0137] 上述计算机实现的算法可能能够从更长的序列中提取可靠的N-聚体水平的结合测量值，从而允许许多新的潜在亲和试剂表征方法，如果没有这样的途径，这些方法可能不可行。上述算法可能能够基于不同来源的肽或表位结合数据来评估结合亲和力，例如：肽下拉与质谱法、点样肽阵列、点样蛋白质阵列、噬菌体展示、NAPPA和打印的肽阵列(如PEPperPrint)。上述计算机实现的算法可允许先前未表征的亲和试剂用于蛋白质鉴定，而无需对其结合亲和力进行严格或绝对的测量。

评估二级结构对亲和试剂蛋白质结合的影响

[0138] 亲和试剂与蛋白质的结合可能会受到存在二级结构元素如螺旋、转角、环和折叠的影响。在一些实施方案中，利用肽阵列和蛋白质结构数据库的方法可以用来评估二级结构的存在如何改变亲和试剂与表位的结合。在一些实施方案中，与已知蛋白质结构有关的信息可以用来鉴定形成不同二级结构如α螺旋和β折叠并且还包含目的表位的蛋白质区域。在一些实施方案中，目的表位可以在所鉴定的二级结构内。在一些实施方案中，目的表位可以在所鉴定的二级结构附近。在一些实施方案中，可以合成具有已知二级结构的蛋白质区域，并用其与尚未观察到具有二级结构的不同蛋白质区域相比，来评估亲和试剂的结合。在一些实施方案中，可以合成具有已知二级结构的蛋白质区域，并用其与包含目的表位和与结构化区域相同的氨基酸组成但氨基酸序列不同的杂乱序列相比，来评估亲和试剂的结合。

[0139] 在一些实施方案中，用于测量亲和试剂对不同二级结构的结合亲和力的方法可包括在打印的肽阵列(例如可从PEPperPRINT获得的那些)上使用数百或数千个肽。例如，可以打印包含11,000个肽的阵列，每个肽具有10-30个残基的长度，然后与荧光标记的亲和试剂杂交。可以认为在阵列上每个斑点处测得的荧光量与亲和试剂对位于该斑点处的相应肽的结合亲和力成比例。在一些实施方案中，可以提供测试结构影响的方法。例如，可以打印阵列，该阵列含有“结构”肽——例如，可能预期具有也包含目的表位的特定二级结构的肽，如15-聚体。另外，可以备选地或另外地打印阵列，以包括尚未观察到具有二级结构并且还包括相同的目的表位的“非结构化”肽。“非结构化”肽可以如下构建：修饰特定序列的部分，以修饰与二级结构相关的一个或多个序列部分，从而产生修饰的序列，该修饰的序列被观察到具有比原始肽更少的二级结构，或者被观察到没有任何二级结构。通过测量结合亲和力的不同指标(例如，荧光的量，或亲和试剂与肽之间的另一种附接指标)，可以确定二级结构组分的存在如何可以影响特定亲和试剂针对一个或多个目的表位的特定结合亲和力。

发现结构化肽

[0140] 在一些实施方案中，用于结构化肽的序列可以来源于已知被结构化的内源蛋白质的区域。在一些实施方案中，可以通过最初搜索参考蛋白质组(例如，Uniprot人类参考蛋白

质组)以查找表位在目的物种的带注释蛋白质序列中的所有出现来发现这些肽。一旦确定了表位在目的物种的带注释蛋白质序列中的出现,就可以将包含表位的序列位置与来自蛋白质数据库如Protein Data Bank(PDB)的结构数据进行交叉参考。另外或备选地,可以使用预测二级结构的软件来确定包含表位的序列位置是否与目的结构基序(例如二级结构)相关。可用来预测二级结构的软件程序包括但不限于Rosetta(https://www.rosettacommons.org/software)、I-TASSER(https://zhanglab.ccmb.med.umich.edu/I-TASSER/)和PEP-FOLD(http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/),等等。

[0141] 一旦确定了与结构数据相关的序列位置/序列部分,就可以去除不在与目的结构基序(例如,β折叠、α螺旋、转角)相关的序列部分内的表位。此外,对于在结构元件内的表位的每种情况,可以从表位的任一侧提取表位和结构化的侧翼序列。可以对每个表位进行几种不同的提取,例如,其中表位在肽的开始处的情况,其中表位靠近肽的中间的情况,以及表位在肽的末端的情况。在一些情况下,可以选择肽的长度,例如,肽的长度可以少于10、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40个或超过40个残基。在一些实施方案中,可以选择表位和侧翼序列的可能的片段。例如,在表位长度为3且肽长度为15的情况下,可以根据该表位是否起始于该肽的第1-13位而选择13种不同的序列。在一些情况下,可以选择多个肽长度。此外,对于每个提取的肽序列(表位+结构化侧翼序列),“非结构化”肽,其中与二级结构相关的序列部分被修饰,以改变或去除二级结构特征。在一些实施方案中,可以通过改组一个或多个侧翼序列但将表位保留在相同位置来生成“非结构化”肽。序列的改组可包括在保持组成的同时改变序列中残基的顺序。另外,组成可被视为序列中每个残基的计数。在一些实施方案中,序列的改组可包括在保持组成的同时随机改变序列中残基的顺序。在一些实施方案中,可以通过在保持序列组成的同时非随机地改变序列内残基的顺序来生成“非结构化”肽。在一些实施方案中,可以通过将非天然的或修饰的氨基酸残基掺入序列中来生成“非结构化”肽。在一些实施方案中,可以通过在序列中或对与表位相关的一个或多个侧翼序列的子集组合地应用一种或多种这样的方法来生成“非结构化”肽。在一些实施方案中,可以通过合成肽并使用诸如圆二色谱法的方法来评估“非结构化”肽的二级结构。

评估结构化肽

[0142] 一旦如上所述已鉴定了包含结构化和非结构化表位的肽,就可以打印具有每个结构化和非结构化肽的许多重复物的阵列。然后可以将该阵列与荧光标记的亲和试剂杂交。可以测量在该阵列上每个斑点处测得的荧光的量,并且可以进行比较结构化肽与非结构化肽之间的荧光测量结果的配对t-检验,以确定是否存在例如由二级结构引起的显著的结合增加或减少。如果结构化肽的荧光相对于非结构化肽降低,则这可能表明由于二级结构元素导致的亲和试剂结合的某种程度的破坏。由于蛋白质折叠的挑战,当打印在阵列上时,并非所有“结构化”肽都可能折叠成预测的“天然”构象。对每个表位使用多个不同的结构化肽序列和冗余片段可能有助于解决这一挑战。可以选择与具有相同表位的其他肽相比显示出结合亲和力显著变化的任何肽进行进一步研究。例如,可以通过圆二色谱法(CD光谱法)评估该肽,以确定其是否已经采取了预期的二级结构。在一些情况下,在与亲和试剂杂交之前,可以通过CD光谱法评估所有肽以确定二级结构。

[0143] 用于评估二级结构对亲和试剂结合的影响的另一种方法可以是使用来自无细胞表达或来自细胞裂解物的全长蛋白质。在一些情况下，可以从组织样品中分离出高度结构化的蛋白质，例如朊病毒蛋白质、淀粉样蛋白斑或tau原纤维。例如，可以获得具有已知二级结构的不同重组蛋白的样品，并用其与非结构化蛋白质、杂乱蛋白质或变性蛋白质相比，来评估亲和试剂与结构化蛋白质的结合。如上所述的方法可以用于位于所选结构化蛋白质的结构化区域内的任何表位。所述蛋白质可以与固体支持物结合，以促进对结合的评估，或者可以保持在溶液中，并且可以使用诸如亲和下拉的方法评估结合。

[0144] 在一些情况下，可以进行肽资格筛选和蛋白质资格筛选，以选择具有所需性质的亲和试剂。在其他情况下，可以在进行任一个筛选之前选择亲和试剂，并且可以进行筛选以收集关于亲和试剂的结合数据。

[0145] 可以针对给定N-聚体的所有可能的表位，例如针对所有8000种可能的3-聚体，进行筛选和选择步骤。在一些情况下，可以仅对可能的N-聚体的子集进行筛选步骤。例如，不是针对所有8000种3-聚体筛选亲和试剂，而是可以选择800种3-聚体作为所需表位，而可以不针对其他7200种3-聚体进行筛选。

[0146] 在一些情况下，使用本文所述方法选择的亲和试剂可以与蛋白质组中所有可能的蛋白质的约0.1％、0.5％、1％、5％、10％、15％、20％、25％、30％、35％、40％、45％或50％结合。在一些情况下，使用本文所述方法选择的亲和试剂可以与蛋白质组中所有可能的蛋白质的至少约0.1％、0.5％、1％、5％、10％、15％、20％、25％、30％、35％、40％、45％、50％或超过50％结合。在一些情况下，使用本文所述方法选择的亲和试剂可以与蛋白质组中所有可能的蛋白质的不超过约50％、45％、40％、35％、30％、25％、20％、15％、10％、5％、1％、0.5％、0.1％或不超过0.1％结合。在一些情况下，可以选择具有不同结合性质的多种亲和试剂并将其合并在一起，以使该亲和试剂池与蛋白质组中所有蛋白质的大约一半结合。在一些情况下，可以使用亲和试剂池代替单一亲和试剂。

[0147] 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。在一些实例中，适体可能是优选的。在一些情况下，亲和试剂可以是可商购获得的亲和试剂，如可商购获得的抗体。在一些情况下，可通过筛选可商购获得的亲和试剂以鉴定具有有用特性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共同地辨析相似蛋白质(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，可针对重叠的结合特性对亲和试剂进行筛选，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的方式进行。一个实例是针对NAPPA或表位平铺(tiling)阵列来筛选亲和试剂。在一些情况下，可以使用被设计为与蛋白质靶标结合的蛋白质特异性亲和试剂(例如可商购获得的抗体或适体)。在一些情况下，可以在进行结合测量之前混合多种蛋白质特异性或表位特异性亲和试剂。例如，对于每个结合测量运行，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子集。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行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亲和试剂的1％、5％、10％、20％、30％、40％、50％、60％、70％、80％、90％或超过90％组成。亲和试剂的混合物可以由所有可用亲和试剂的至少1％、5％、10％、20％、30％、40％、50％、60％、70％、80％、90％或超过90％组成。亲和试剂的混合物可以由所有可用亲和试剂的不超过90％、80％、70％、60％、50％、40％、30％、20％、10％、5％、1％或少于1％组成。亲和试剂的混合物可以由2、3、5、10、20、30、50、100、150、200、250、300、400、500、600、700、800、900、1000、1200、1500、2000、3000、4000、5000、7000、8000、10000种或超过10000种亲和试剂组成。亲和试剂的混合物可以由至少2、3、5、10、20、30、50、100、150、200、250、300、400、500、600、700、800、900、1000、1200、1500、2000、3000、4000、5000、7000、8000、10000种或超过10000种亲和试剂组成。亲和试剂的混合物可以由不超过10000、8000、7000、5000、4000、3000、2000、1500、1200、1000、900、800、700、600、500、400、300、250、200、150、100、50、30、20、10、5、3种或少于3种亲和试剂组成。

[0148] 亲和试剂可具有高、中或低特异性。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，在本公开的方法中使用的亲和试剂可以对单个表位是高度特异性的。在一些情况下，在本公开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。

[0149] 在一些实施方案中，用来鉴定目标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时彼此无差异或不可区分。特别是，可用来鉴定相同目标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目标氨基酸序列。例如，无论侧翼序列如何均结合三聚体氨基酸序列(AAA)的亲和试剂可包含结合三聚体AAA序列而不受侧翼序列的任何影响的单个探针，或者包含一组400个探针，每个探针都与αAAAβ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情况下，可以组合所述400个探针，使得每个探针的量相等。在第二种情况的一些情况下，可以组合所述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进行加权，使得任何给定的5氨基酸表位被结合的概率相等。

[0150] 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来设计亲和试剂。基于结构的药物设计(或直接药物设计)利用关于目的表位的三维结构和亲和试剂的结合位点的知识。

[0151] 在一些情况下，亲和试剂可以用核酸条形码进行标记。在一些实例中，核酸条形码可用来纯化使用后的亲和试剂。在一些实例中，核酸条形码可用来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进行标记，该荧光团可用来分选使用后的亲和试剂。

[0152] 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化的(multiplexed)，然后使用互补核酸探针来检测。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化的亲和试剂。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化的亲和试剂。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化的亲和试剂。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化的亲和试剂。

[0153] 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。一旦所述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，以确定所交联的亲和试剂的身份。在一些情况下，多种结合蛋白可以暴露于所述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，以确定与所述多种结合的蛋白质中的每一种相关联的所交联的亲和试剂的身份。

[0154] 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。

[0155] 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，检测部分可以直接共价附接至亲和试剂，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。超过一个检测部分可以附接到每个亲和试剂上。检测部分可以附接到亲和试剂的任何位置，包括末端。检测部分可以附接在不改变亲和试剂的构象或表位的表位结合区的位置。

[0156] 检测部分包括但不限于荧光团、生物发光蛋白质(GFP)、酶蛋白质(例如，HRP)、纳米颗粒(例如，量子点)、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的荧光团。

[0157] 检测部分可以是可从亲和试剂上切下的。这可以允许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。

[0158] 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该小鼠抗体。或者，亲和试剂可以是适体，该适体由对该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。

[0159] 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之一缀合。这可以允许亲和试剂的多路化，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，而第三亲和试剂可与红色荧光蛋白缀合，因此这三种亲和试剂可以是多路化的并通过它们的荧光进行鉴定。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，而第三、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化，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化反应。可以一起多路化的亲和试剂的数目可取决于用来区分它们的检测部分。例如，用荧光团标

记的亲和试剂的多路化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和试剂的多路化可由核酸条形码的长度决定。

[0160] 可以在用于测定之前确定每种亲和试剂的特异性。可以在使用已知蛋白质的对照实验中确定亲和试剂的结合特异性。可以使用任何合适的实验方法来确定亲和试剂的特异性。在一个实例中，固体支持物可以在已知位置处负载已知的蛋白质标准品，并用来评估多种亲和试剂的特异性。在另一个实例中，固体支持物可包含实验样品以及一组对照和标准品，使得每种亲和试剂的特异性可从与所述对照和标准品的结合来计算，然后用来鉴定所述实验样品。在一些情况下，可以包括具有未知特异性的亲和试剂以及已知特异性的亲和试剂，来自已知特异性的亲和试剂的数据可用来鉴定蛋白质，而未知特异性的亲和试剂与所鉴定的蛋白质的结合模式可用来确定其结合特异性。还可以通过使用其他亲和试剂的已知结合数据来重新确认任何单独的亲和试剂的特异性，以评估该单独的亲和试剂结合哪些蛋白质。因此，通过亲和试剂组的多次使用，亲和试剂的特异性可以随着每次迭代而逐渐改善。虽然可以使用对特定蛋白质具有独特特异性的亲和试剂，但是本文所述的方法可能不需要它们。另外，方法可能对一系列特异性有效。在一些实例中，当亲和试剂对任何特定蛋白质都不具有特异性，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。

[0161] 在一些实例中，可以选择一种或多种亲和试剂来结合给定长度如2、3、4、5、6、7、8、9、10或多于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和试剂来结合具有2个氨基酸至40个氨基酸的一系列不同长度的氨基酸基序。

[0162] 在一些实例中，可以选择具有高、中或低结合亲和力的亲和试剂。在一些情况下，具有低或中结合亲和力的亲和试剂可能是优选的。在一些情况下，亲和试剂可具有约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，亲和试剂可具有大于约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。

[0163] 可以选择一些亲和试剂来结合修饰的氨基酸序列，如磷酸化的或遍在蛋白化的氨基酸序列。在一些实例中，可以选择对可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和试剂。在一些实例中，一种或多种亲和试剂可以结合两种或更多种不同的蛋白质。在一些实例中，一种或多种亲和试剂可以与其一种或多种靶标弱结合。例如，亲和试剂可以与其一种或多种靶标以低于10％、低于10％、低于15％、低于20％、低于25％、低于30％、低于35％或低于35％结合。在一些实例中，一种或多种亲和试剂可以与其一种或多种靶标中等或强烈地结合。例如，亲和试剂可以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。

[0164] 为了补偿弱结合，可将过量的亲和试剂施加至固体支持物。亲和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。亲和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。

[0165] 亲和试剂还可包含磁性组分。该磁性组分可用于将一些或所有结合的亲和试剂操纵到同一成像平面或z堆叠(stack)中。将一些或所有亲和试剂操纵到同一成像平面中可以

改善成像数据的质量并降低系统中的噪声。

基底/固体支持物

[0166] 在一些实施方案中，可以将肽施加至官能化的基底，如固体支持物，以促进亲和试剂选择或筛选步骤。图3B显示了具有与固体支持物缀合的阵列肽的固体支持物的实例，其用于亲和试剂表征筛选，以确定侧翼序列对亲和试剂与表位AAA的结合的影响。在一些情况下，可以将靶肽直接施加至固体支持物。在一些情况下，可以在固体支持物上合成或生长靶肽。在一些情况下，可以在诸如阵列或珠子的固体支持物上合成肽。例如，可以制造肽阵列，以在该阵列的每个位置处包含单个肽序列的多个拷贝。定制的肽微阵列也可以商业购买，例如PEPperPRINT。

[0167] 基底可以是能够形成固体支持物的任何基底。如本文所用的基底，如固体支持物，可指肽可以共价或非共价附接至其上的任何固体表面。固体支持物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、微流体腔室、通道、微流体通道或其他任何表面。固体支持物表面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。固体支持物表面可含有微孔。在一些实施方案中，固体支持物可以由玻璃、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、金属或金属氧化物如金或氧化锆，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，固体支持物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(例如通过硅烷化或包含磷酸基团的化合物的结合)进行衍生化。合适的硅烷试剂可包括氨基丙基三甲氧基硅烷、氨基丙基三乙氧基硅烷和4-氨基丁基三乙氧基硅烷。固体支持物可用N-羟基琥珀酰亚胺(NHS)官能团进行官能化。玻璃表面还可以使用例如环氧硅烷、丙烯酸硅烷或丙烯酰胺硅烷，通过诸如丙烯酸或环氧基等其他反应性基团进行衍生化。供寡核苷酸附接的固体支持物和方法优选地对于反复的结合、洗涤、成像和洗脱步骤是稳定的。在一些实例中，固体支持物可以是载玻片或流动池。

[0168] 固体支持物上的官能团的间隔可以是有序的或随机的。官能团的有序阵列可以通过光刻法来创建。有序阵列中的官能团可以定位成使得每个官能团与其他任何官能团相距至少约10nm、约25nm、约50nm、约100nm、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或超过约1000nm。可以以一定的浓度提供随机间隔的官能团，使得官能团与其他任何官能团相距平均至少约10nm、约25nm、约50nm、约100nm、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或超过约1000nm。

[0169] 基底，如固体支持物，可以被间接地官能化。例如，可对固体支持物进行聚乙二醇化，并且可将官能团施加至全部或一部分PEG分子。

[0170] 基底可包含任何材料，包括金属、玻璃、塑料、陶瓷或其组合。基底可包含用于各种目的的附加层或材料，如反射涂层、抗反射涂层、钝化基团或阻断基团。在一些优选的实施方案中，固体支持物可以是流动池。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，硼硅酸盐玻璃层)、覆盖在基层上的通道层(例如，蚀刻的硅层)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1700μm。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。构成基底的材料可以被选择为具有低水平的自发荧光或

[0171] 在一些实施方案中，流动池可在流动池底部上包含通道开口。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，通道可以是小于约50μm深、约50μm深、小于约100μm深、约100μm深、约100μm至约500μm深、约500μm深或超过约500μm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。

[0172] 可将肽点样、滴加、移液、流动、洗涤或以其他方式施加至固体支持物。在固体支持物已经用诸如NHS酯的部分进行官能化的情况下，不需要对肽进行修饰。在固体支持物已经用替代部分(例如巯基、胺或连接体DNA)进行官能化的情况下，可以使用交联试剂(例如辛二酸二琥珀酰亚胺酯、NHS、磺酰胺)。在固体支持物已经用连接体DNA进行官能化的情况下，可以用互补DNA标签修饰靶标的肽。

[0173] 可以使用可光活化的交联剂来引导样品与固体支持物上的特定区域的交联。可以使用可光活化的交联剂通过将每个样品附接在固体支持物的已知区域中来允许肽样品的多路化。可光活化的交联剂可以例如通过在肽交联之前检测荧光标签来允许已经成功标记的肽的特异性附接。可光活化的交联剂的实例包括但不限于N-5-叠氮基-2-硝基苯甲酰基氧基琥珀酰亚胺、6-(4'-叠氮基-2'-硝基苯基氨基)己酸磺基琥珀酰亚胺酯、4,4'-氮杂戊酸琥珀酰亚胺酯、4,4'-氮杂戊酸磺基琥珀酰亚胺酯、6-(4,4'-氮杂戊酰胺基)己酸琥珀酰亚胺酯、6-(4,4'-氮杂戊酰胺基)己酸磺基琥珀酰亚胺酯、2-((4,4'-氮杂戊酰胺基)乙基)-1,3'-二硫代丙酸琥珀酰亚胺酯和2-((4,4'-氮杂戊酰胺基)乙基)-1,3'-二硫代丙酸磺基琥珀酰亚胺酯。

[0174] 肽可通过一个或多个残基附接至基底。在一些实例中，肽可经由N末端、C末端、两个末端或经由内部残基附接。

[0175] 在一些实施方案中，每个肽序列可与独特空间地址相关联。例如，一旦肽在空间上分离的位置处附接至固体支持物，则每个肽序列可以例如通过坐标被分配索引化的地址。在一些实例中，预先分配的独特空间地址的网格可以预先确定。在一些实施方案中，固体支持物可含有易于辨识的固定的标志，使得可以相对于该固体支持物的固定的标志确定每个肽的放置。在一些实例中，固体支持物可具有永久地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固体支持物的表面，以提供用以定位所交联的肽的参考。图案化的形状本身，如缀合肽的外部边界，也可以用作基准点，以供确定每个斑点的独特位置。

实施例

实施例1:选择方法

步骤1:使用SELEX的预富集

[0176] 目的:比较一组预富集方法的以下能力:(1)大大降低输入文库针对肽靶标的多样性,(2)产生表现出与目标适体概况(profile)一致的性质的低复杂性适体池。

[0177] 关键操作:为了缩小针对每个肽靶标的最终适体池,使用包含大约10^{15}个随机文库序列的多样性的输入文库,并使用典型的基于珠子的SELEX进行预富集。待测试的文库设计可包括结构化文库和原始文库。

[0178] 为了促进具有慢解离速率的靶标的预富集,将适体池与肽-珠池之间在结合缓冲液中的SELEX孵育进行过夜,以允许慢解离速率结合物与快解离速率结合物竞争珠子占据。另外,还探索孵育后珠-肽复合物的一系列连续更长的洗涤时间,以偏向于慢解离率结合的富集。将过夜孵育条件与1小时孵育进行比较,从而产生两个SELEX轨迹:短孵育和长孵育。

[0179] 为了促进在特定洗脱物的存在下快速解离的靶标的预富集,在10mM EDTA(或其他待定义的洗脱物)中洗脱结合的适体,并使用洗脱物生成适体池用于下一轮SELEX。通过加热从剩余的珠子上洗脱材料并保存,以允许探索替代洗脱条件的选项。

[0180] 为了确定对于短孵育和长孵育SELEX轨迹,无论侧翼氨基酸残基的组成如何,最能促进结合靶标的预富集的策略,比较四种富集方法,产生总共八条SELEX轨迹:

[0181] 预富集方法1:针对“低严格性”甘氨酸侧翼(“G-flank”,GXG)肽靶标进行预富集SELEX。

[0182] 预富集方法2:在第一组SELEX轮次中针对G-侧翼肽靶标,然后在后续SELEX轮次中针对侧翼为一组特定氨基酸残基(“αXβ”)的“高严格性”肽靶标进行预富集SELEX。

[0183] 预富集方法3:在第一组SELEX轮次中针对αXβ肽靶标,然后在后续SELEX轮次中针对GXG肽靶标进行预富集SELEX。确定从高到低严格性的过渡轮次。

[0184] 预富集方法4:仅针对αXβ目标肽进行预富集SELEX。

[0185] 预富集后,使用粒子显示(Particle Display)测量进度。该数据用来决定将哪些预富集的适体池转入步骤2。qPCR、流式细胞术和其他生成的数据可用来指导指定选择哪些适体池。

[0186] 作为预富集过程的一部分,下一代测序(NGS)可以在所有或部分SELEX轮次上进行。文库制备可包括多路化(multiplexing)。

在整个预富集过程中,每轮SELEX都将保留扩增后材料,以充当用于粒子展示的潜在适体池。

步骤2:粒子展示

[0187] 目的:鉴定表现出目标适体概况的序列的集合。

[0188] 关键操作:在指定的预富集的适体池上进行必要次数的粒子显示分选轮次(通常为三到五轮),以分离表现出目标适体概况的适体池。细化在粒子展示执行过程中的实验设计,以相应地调整分选的严格性。为了促进表现出目标适体概况的适体的分选,每个粒子展示轮次由两个分选步骤组成,为每个输入池产生两个适体候选输出池。在该步骤中,仅使用αXβ目标肽。

[0189] 第一分选步骤旨在促进表现出性质1的适体的富集。将适体池与荧光团1标记的目标肽和生物素标记的阴性对照肽混合。在用荧光链霉亲和素标记之前,让过量的结合的靶标和阴性对照肽从适体珠子上解离15分钟,并且根据高目标肽信号、低阴性对照肽信号和

高适体信号对分选进行门控。扩增该输出池，而保留该池的一部分以用于NGS。

[0190] 第二分选步骤旨在促进表现出性质2的适体的富集。来自第一分选步骤的输出池的剩余部分以与用于第一分选步骤的输入池相同的方式进行准备，不同之处在于省略阴性对照肽，并将该池转移到洗脱缓冲液中并在分选之前孵育15秒。根据低肽信号和高适体信号对该池进行门控。扩增该输出，而保留该池的一部分以用于NGS，该池的剩余部分用于下一轮粒子展示。

步骤3：测序和适体表征

[0191] 目的：鉴定特定序列和序列家族，并确认它们的结合性质符合目标适体概况。

[0192] 关键操作：基于上述结果鉴定待通过NGS进行评价的粒子展示输出池。准备文库用于测序并用来生成NGS输出。

[0193] 合成候选序列(“适体命中物”)，并且/或者将其与珠子缀合，并如下针对目标适体概况进行筛选：

[0194] 为了针对性质1对适体命中物进行评分，将适体珠子与1uM荧光标记的目标肽混合3分钟，洗涤，并稀释至适合于流式细胞术分析的浓度。然后通过在1小时的时间过程中对混合物进行取样来确定解离速率。

[0195] 为了针对性质2对适体命中物进行评分，如上所述制备适体珠子，不同之处在于稀释剂由洗脱缓冲液组成，并且在加入稀释剂后15秒内或在技术上可行的情况下尽快对混合物进行取样。

[0196] 为了针对性质3对适体命中物进行评分，分别使用G-侧翼和X-侧翼肽靶标，将适体候选物针对性质1和性质2进行评分。

实施例2：实现实施例

[0197] 在该实施例中考虑的是亲和试剂与核酸可编程蛋白质阵列(NAPPA)的结合测量，该NAPPA包含720种独特的蛋白质，阵列上包含各3-5个重复物。在一些实施方案中，这些重复物可以分散在整个阵列上。通过将荧光标记的亲和试剂与该阵列杂交并测量在该阵列上的每个斑点处观察到的荧光来评估结合。在该实施例中，每个斑点处每种蛋白质的身份是已知的。因此，荧光可以对应至蛋白质身份。

[0198] 在该实施例中，将蛋白质结合测量(荧光)模拟为蛋白质序列中8000种可能的三聚体各自的计数与对归因于亲和试剂与该三聚体结合的观察到的荧光的分数贡献的线性组合。即：

$$F_{pr} = \sum_{t=1}^{t=8000} c_{t,pr}\beta_t$$

其中：

F_{pr}＝蛋白质pr的荧光

$c_{t,pr}$＝蛋白质pr中三聚体t的计数

β_t＝来自亲和试剂与三聚体t的结合的分数荧光亲和试剂针对多种蛋白质的测量结果形成线性方程组：

$$\vec{F} = C\vec{\beta} + \epsilon$$

其中：

$\vec{F}$是含有对于每种蛋白质观察到的荧光的长度N列向量；

C是三聚体计数的N x 8000矩阵，其中每列是每种测量的蛋白质中特定三聚体的计数；

$\vec{\beta}$是对于每种可能的三聚体的分数荧光的长度8000列向量；

∈是针对背景结合或本底噪声进行校正的标量常数。

[0199] 可将本文所述的模型扩展到包括非标准氨基酸或模拟不同长度的n-聚体。在已知的NAPPA或类似的结合测量的情况下，$\vec{F}$和C是已知的变量，而$\vec{\beta}$和∈的值可以通过线性回归或相关方法得出。特别是，非负最小二乘法和非负最小绝对收缩和选择算子(LASSO)回归可能非常适合于该问题。非负最小二乘法将$\vec{\beta}$的解限制为非负值，而非负LASSO回归进一步施加了稀疏约束。当方程组欠定时，即，当所测得的独特蛋白质的数目少于独特三聚体的数目(在此示例中为8,000)时，LASSO回归可能特别有效。

[0200] 对于具有以下相对亲和力的亲和试剂，模拟与上述NAPPA阵列的结合：

LAS:0.75

GEL:0.05

YIC:0.001

STK:0.001

与所有其他三聚体的结合被分配亲和力0.00001。

[0201] 通过使用所描述的统计模型和施加有高斯噪声的上述结合亲和力(具有预期结合亲和力的5％的标准偏差)计算预期结合，来模拟与NAPPA阵列上的每种蛋白质的结合。

[0202] 从这些模拟的数据，使用LASSO(LASSO稀疏度参数α被设置为1e-3)从蛋白质结合测量结果了解试剂的结合亲和力。学习的亲和力是：

LAS	0.750034
GEL	0.040796
ASW	0.012120
SGD	0.005001
QPH	0.003975
LGC	0.003788
KGR	0.002475
VPS	0.002416
EIK	0.002305
SES	0.002149
RIS	0.001235
SSE	0.001067
STS	0.000953
APP	0.000889
HSD	0.000785
KSQ	0.000735
VQK	0.000637
PPT	0.000538
GCS	0.000417
RPR	0.000106
QPQ	0.000051
ACS	0.000026

与所有其他三聚体的亲和力均为零。

[0203] 该实施例表明，该方法能够建立对亲和试剂结合的合理估计，尽管具有可容忍的噪声量。

实施例3:KW亲和试剂的选择

[0204] 进行亲和试剂选择筛选，以选择与四个不同靶表位——KW、LFQ、IRN和EGE——中的任一个结合的适体。使用基于阵列的表面等离子体共振成像选择从适体文库中选择结合靶标的适体。对与靶标KW结合的适体进行了初步验证。图6显示了如通过表面等离子体共振测量的，5个排名前列的命中物与KW靶标的结合。图7显示了与阴性对照肽相比，最佳抗KW适体的结合。

[0205] 为了确定来自图6的排名最高的抗KW适体的表位特异性，对该适体进行了荧光标记并施加至肽阵列。基于标记的适体与肽阵列的结合，被适体识别的表位被确定为KW和KR，对两者具有相对类似的亲和力。结合数据和初步基序分析在图8中示出。

实施例4:高通量测序荧光配体相互作用概况分析

[0206] 通过获得适体文库，并在该文库的适体的两端掺入衔接子序列，进行了高通量测序荧光配体相互作用概况分析。将具有衔接子的适体的文库施加至包含固定的寡核苷酸引物的流动池，并如图10所示进行簇扩增。簇扩增后，使用可逆染料终止子对簇(每个簇代表来自适体文库的单个适体)进行测序。然后使簇变性以除去测序试剂，洗涤，并使其折叠成其天然适体构象。将荧光标记的肽靶标以四种不同的浓度施加至簇，并对其进行成像以显示与荧光标记的肽靶标结合的簇。图11显示了具有结合的荧光标记的肽靶标和两种不同浓度的流动池的代表性图像。将荧光数据与测序数据组合，以提供在流动池上代表的适体文库中每个适体的亲和力测量结果。图12显示了几种不同的适体的序列和对荧光标记的肽靶标(LFQ)的结合亲和力。

实施例5:肽合成与验证

材料：

[0207] 所有试剂和溶剂均为肽合成级或可获得的最高纯度。氨基酸衍生物获自Aapptech(Louisville,KY USA)，而所有用于肽合成、SPE和RP±HPLC的溶剂均获自Acros Organics,USA。

肽合成：

[0208] 使用标准Fmoc/tBu化学方法，在MultiPep RSi合成仪(Intavis,德国)上合成肽。氨基酸衍生物被N,N,N′,N′-四甲基-O-(1H-苯并三唑-1-基)脲鎓六氟磷酸盐(HBTU)(0.5M)活化。在96容器的微反应模块中使用标准合成方案合成肽。Rink Amide Resin(Intavis,德国)以0.53mmol/g的胺含量加载至每个容器2微摩尔。

添加后用序列合成肽：

树脂制备		
1	备注	合成规模：2 μmol
1	漂洗针	1500 / 2000 ul (DMF)
2	移液	1500μl /1500μl CapMixture (乙酸酐)->排出
3	漂洗针	1500 / 2000 ul (DMF)
4	移液	1500μl /1500μl N-甲基哌啶溶液->排出
5	漂洗针	1000 / 2000 ul (DMF)
6	处理歧管	10000 μl, DMF
7	洗涤树脂	1200 μl, DMF, 4x
8	提取	60 s
循环		1.-> N. (计数= N) (N – 偶联数)
9	脱保护	50 μl, N-甲基哌啶
10	脱保护	50 μl, N-甲基哌啶
11	漂洗针	1000 / 2000 ul
12	洗涤树脂	1800 μl, DMF, 2x
13	洗涤树脂	1200 μl, DMF, 8x

14	提取	30 s
15	偶联	20μl HBTU + 5μl NMM +2μl NMP + 21μl氨基酸衍生物
16	偶联	20μl HBTU + 5μl NMM +2μl NMP + 21μl氨基酸衍生物
17	加帽	50 μl, CapMixture (乙酸酐)
18	漂洗针	1000 / 2000 ul (DMF)
19	洗涤树脂	1200 μl, DMF, 6x
20	提取	60 s
最终动作		
21	脱保护	80 μl, N-甲基哌啶
22	脱保护	80 μl, N-甲基哌啶
23	脱保护	80 μl, N-甲基哌啶
24	漂洗针	500 / 2500 ul (DMF)
25	洗涤树脂	1800 μl, DMF, 2x
26	洗涤树脂	1200 μl, DMF, 8x
27	处理歧管	10000 μl, 乙醇
28	洗涤树脂	1200 μl, 乙醇, 2x
29	提取	300 s
30	漂洗针	1000 / 3000 ul (DMF)

[0209] 除Fmoc-赖氨酸(生物素)(0.3M)外,所有氨基酸衍生物均以DMF中的0.5M溶液使用。

N-甲基哌啶作为DMF中的2M溶液使用,乙酸酐作为DMF中的5%溶液(0.529M)使用。每个循环继续偶联30min,每个结构单元重复2次。氨基酸相对于胺以10倍过量使用,以确保完全偶联。在室温下,用5%水、2.5%三异丙基硅烷在三氟乙酸(TFA)中的溶液裂解肽2h,

在-20℃下用叔丁基甲基醚沉淀，并用冷的叔丁基甲基醚(200ul)洗涤3次。将干燥的肽溶解在1mL的0.1％TFA水溶液(500微升)中，并在-20℃下保存。

固相提取

[0210] 使用Multi 96真空歧管(Waters,USA)，在每孔中含有40mg C18吸附剂的Sep-Pak C18 Multi 96孔板(WAT054955)上纯化粗制肽。将洗脱物收集在2mL收集架(WAT058956)中。在典型方案中，将柱或板用1ml MeOH洗涤，并在加载肽样品之前用0.1％的TFA水溶液调节。通过在不施加真空的情况下使溶液缓慢通过固定相来加载肽，以完成接近定量的肽结合。固相用0.1％TFA水溶液洗涤，以除去盐和其他极性杂质，之后用1mL 70％乙腈水溶液洗脱结合的肽。使用Centrivap将溶剂蒸发过夜，并将沉淀物重新溶解在500微升的去离子水中。使用Tecan Spark读板仪，以水作为空白，并收集200至1000nm范围内的吸收光谱，来测量浓度。使用214nm处的吸收和消光系数来计算肽的浓度。使用紫外线透明的96孔板(ThermoFisher 8404)测量肽的浓度。

质谱法：

[0211] 基质溶液购自Agilent Technologies(目录号G2037A)。通过将1微升的肽储备溶液和1微升的基质共点样到MALDI样品板上并让样品在环境条件下干燥来制备样品。在MALDI-TOF质谱仪AB SCIEX 5800TOF/TOF上记录质谱。代表性质谱在图9中示出。

实施例6:采用磁珠肽文库的亲和试剂SELEX选择

在炔烃磁珠上的肽缀合

[0212] 制备了炔烃磁珠的缀合溶液。制备了0.01％吐温20的第一溶液(5mL水+0.5μL吐温20)。由溶解在1mL水中的25mg $CuSO_4 \cdot 5H_2O$制备100mM $CuSO_4 \cdot 5H_2O$(MW:249.69g/mol)的第二溶液。通过将100mg抗坏血酸钠溶解在1mL水中制备500mM抗坏血酸钠(MW:198.11g/mol)。

[0213] 洗涤储备炔烃磁珠。每个样品使用5μL提供的珠子。通过涡旋>30秒或倾斜并旋转5分钟将珠子重新悬浮在储备小瓶中。从储备小瓶中取出珠子，并使用磁性管架拉下并除去上清液。将1mL 0.01％吐温20添加至珠子，将珠子重新悬浮，然后使用磁性管架拉下，随后去除上清液。在吐温20中的漂洗再重复4次。将洗涤过的珠子重新悬浮在0.01％吐温20中。

[0214] 如下所述将肽与炔烃磁珠缀合。将每份磁珠(200μL珠子)与100nmol制备的肽、20μL 100mM CuSO4·5H2O(炔基数的20倍)、20μL500mM抗坏血酸钠(炔基数的100倍)和足够的0.01％吐温20合并，以使反应体积达到1000μL。将合并的肽-珠混合物在热混合器(Thermomixer)中摇动16小时。混合后，将缀合的珠子拉下。取出并保存上清液以在蛋白质分析仪上测量肽浓度。珠子用0.01％吐温20洗涤两次，然后重新悬浮在0.01％吐温20中。

亲和试剂的SELEX选择

[0215] 在SELEX前，针对肽靶标预富集文库三轮。在预富集轮次(PER)1中，亲和试剂文库在水中用热洗脱。在PE R2和R3之前，将另一个125pmol原始(naïve)文库添加到热洗脱产物中。

部分1:用于SELEX的缓冲液的制备：

[0216] 制备用于SELEX选择的缓冲液。通过将47.66g HEPES(MW 238.30g/mol)、140.256g NaCl(MW 58.44g/mol)、9.5211g MgCl2(MW 95.211g/mol)和7.455g KCl(MW 74.55g/mol)溶解在750mL水中制备1L NVT的20x储备液[200mM HEPES,2.4M NaCl,100mM MgCl2,100mM KCl,2％吐温20pH 7.4]。使用HCl/NaOH将pH调节至7.4。将20mL吐温20添加到溶液中，然后

使最终体积达到1L。由1x NVT与300uM硫酸葡聚糖和0.1mg/ml剪切鲑精DNA制备结合缓冲液。用于SELEX的洗涤缓冲液是1x NVT。

部分2:用于SELEX的文库的制备:

[0217] 在1X NVT中制备中间文库储备液。将文库加热至95℃ 5分钟,然后冷却10分钟。

部分3:肽珠洗涤和文库结合:

[0218] 创建了5个具有200μL 1XNVT的多孔洗涤板。1个多孔珠板每孔加载来自珠肽储备微量离心管的40μL珠子(来自原始储备板的5μL珠子),然后与每孔160μL 1xNVT合并。将梳子(comb)添加到该板上。产生了1个多孔文库板,每孔具有在200μL 1X结合缓冲液中的100pmol加热的原始文库。创建了1个标准板,每孔含有50μL水。

[0219] 从珠板上收集珠子3次,每次收集20秒。通过释放珠子10秒、混合30秒并收集珠子5次,每次收集30秒,在洗涤板1中洗涤珠子。在洗涤板2中重复该洗涤方法。将珠子移动到文库板进行文库结合,方法是通过释放珠子10秒,结合6轮快速混合9分30秒加上慢速混合30秒,然后收集珠子5轮30秒。然后洗涤珠子3次(洗涤板3、4和5),方法是通过释放珠子10秒,通过中度混合4分30秒加上缓慢混合30秒洗涤2次,然后收集珠子2轮30秒。将珠子在水中洗脱,方法是通过将珠子释放14秒,以中速混合30秒并缓慢混合30秒,然后收集珠子2轮21秒。将珠子转移回珠板。收集热洗脱板,并将其内容物转移至PCR板以在下一轮之前储存。

SELEX(6轮)

[0220] 使用原始和预富集的亲和试剂文库对长肽和短肽进行SELEX。前两轮的洗脱分为两组:EDTA和肽竞争洗脱。从第3轮开始,所有样品都用EDTA洗脱。所有轮次都包括第二次热洗脱。

部分1:用于SELEX的文库的制备:

[0221] 在1X NVT中制备中间文库储备液。将文库加热至95℃ 5分钟,然后冷却10分钟。

部分2:肽珠洗涤和文库结合:

[0222] 在MagMax上进行珠子洗涤。生成5个多孔洗涤板,每孔含有200μL 1XNVT。用来自珠肽储备微量离心管的16μL珠子(来自原始储备板的5μL珠子)生成1个多孔珠板,并填充至160μL 1xNVT。将梳子添加至该板。在134μL 1X结合缓冲液中,在50μL加热的原始亲和试剂文库或50μL预富集的亲和试剂文库中以100pmol生成1个多孔文库板。将16μL珠子加入到适当的孔中。在适当的孔中用50μL EDTA生成1个标准板。其余的孔用水填充。在适当的孔中用1XNVT中的50μL肽(500pmol)生成1个标准板。剩余的孔用1xNVT填充。用50μL水生成1个标准板。

[0223] 从珠板上收集珠子3次各20秒。通过释放珠子10秒,中度混合30秒,然后收集珠子5次各30秒,在洗涤板1中洗涤珠子。在洗涤板2中重复洗涤。将珠子移动到文库板进行文库结合,方法是通过结合6次(中度混合9分30秒加上缓慢混合30秒),收集珠子5次各30秒。洗涤珠子3次(洗涤板3、4和5),方法是通过释放珠子10秒,洗涤2次(中度混合4分30秒加上底部混合30秒),然后收集珠子2次各30秒。对于在EDTA中的洗脱,将珠子释放14秒,然后中度混合30秒并缓慢混合30秒,并收集珠子2次21秒。对于肽中的洗脱,将珠子释放14秒,然后中度混合3次4分30秒并缓慢混合30秒,然后收集珠子2次各21秒。对于在水中的洗脱,将珠子释放14秒,然后中度混合30秒并缓慢混合30秒,然后收集珠子2次各21秒。将珠子转移回文库结合板。从所有洗脱板收集液体,并将内容物转移至PCR板进行储存。将EDTA和肽洗脱产物

合并到一个板中用于PCR扩增。

部分3:SELEX输出的扩增

[0224] 将内容物从肽/EDTA板转移到qPCR板。制备两种不同的主混合物(一种用于EDTA洗脱,一种用于肽洗脱)。通过组合2x Kapa HiFi 100x1/2＝50μL、正向引物100x1/500＝0.2μL、反向引物100x1/500＝0.2μL、SyBr绿100x1/100＝1μL、MgCl2 100x4.5/500＝0.9μL、EDTA洗脱＝45μL(对于第1轮)和2.7μ水,以总体积100μL制备用于EDTA洗脱的主混合物。通过组合2x Kapa HiFi 100x1/2＝50μL、正向引物100x1/500＝0.2μL、反向引物100x1/500＝0.2μL、SyBr绿100x1/100＝1μL、EDTA 100x2.5/100＝2.5μL、肽洗脱＝45μL(对于第1轮)和1.1μL水,生成用于肽洗脱的主混合物。使用TissueLyser 3次以混合该板。将板以2000x g离心1分钟。qPCR程序以98℃ 3min、10个循环的98℃ 15s、65℃ 30s和72℃ 30s的热循环运行。

[0225] 创建了用于两种洗脱的一种主混合物,并向每个样品添加100μL。PCR总体积为200μL。主混合物是通过组合2x Kapa HiFi 100x1/2＝50μL、正向引物100x1/500＝0.2μL、反向引物100x1/500＝0.2μL、SyBr绿100x1/100＝1μL和48.6μL水而创建的。记录每个样品达到1000000-1250000FLU所需的PCR循环数。根据阳性对照样品峰达到稳定的位置,选择包括NTC在内的24个样品在生物分析仪上运行,以检查扩增子的大小以及浓度,以确保为后续轮次生成足够的材料。如果检测到的扩增子不足,则用PCR将样品再扩增一轮,以消除气泡形成并产生更多材料。

部分4:文库的重新生成

[0226] 通过组合10mM Tris-HCl、1mM EDTA和2M NaCl,并将溶液的pH调节至7.5,生成2x结合和洗涤缓冲液(B&W缓冲液)。所需珠子的量被计算为300μl PCR反应需要75μl珠子。所需的NaOH量被计算为每75μl珠子需要80μl 20mM NaOH。在1x B&W缓冲液中手动洗涤珠子。创建2个多孔洗涤板,每孔含有200μL 1XB&W。用75μL珠子生成1个多孔珠板,并添加至125μL 1xB&W。将梳子添加至该板。用300μL PCR反应和300μL 2X B&W缓冲液创建1个多孔文库板。在适当的孔中用80μL NaOH创建1个标准板。其余的孔用水填充。1个PCR收集板用20μL 40mM HCL和5μL 20x NVT生成。从珠板上取下梳子。如下所述将珠子与文库板中的亲和试剂文库混合用于文库结合,这是通过结合4次(中度混合4分55秒加上缓慢混合5秒)并收集珠子5次各30秒。将珠子洗涤两次(洗涤板1和2),这是通过释放珠子10秒,洗涤珠子10秒,以及收集珠子5次各30秒。使肽在NaOH中变性,这是通过将肽释放10秒,以中速混合10分钟并缓慢混合5秒,然后收集珠子5次各30秒。在珠板中更换梳子。

[0227] 所有样品都使用Oligreen测定进行定量。将Oligreen标准品在1x NVT中稀释至2ng/μL。在黑色96孔板中对标准品进行2倍稀释(最终体积为100μL)。将2μL ssDNA添加到板中(最终体积为100μL)。将200x Oligreen试剂稀释至1x。将100μL 1x Oligreen试剂添加到板上的每个孔中。在进行下一轮选择之前,在TecanF200上以485(20)nm/535(25)nm读取读数。

实施例7:钴SELEX亲和试剂选择

用于钴琼脂糖珠子的缓冲液的制备

[0228] 为基于琼脂糖珠的SELEX制备缓冲液。通过将47.66g HEPES(MW 238.30g/mol)、140.256g NaCl(MW 58.44g/mol)、9.5211g MgCl2(MW 95.211g/mol)和7.455g KCl(MW 74.55g/mol)溶解在750mL水中制备1L NVT的20x储备液[200mM HEPES,2.4M NaCl,100mM

MgCl2,100mM KCl,2%吐温20pH 7.4],并使用HCl/NaOH将pH调节至7.4。添加20mL吐温20,并使最终体积达到1L。通过将1x NVT与300uM硫酸葡聚糖和0.1mg/ml剪切鲑精DNA混合制备结合缓冲液。

文库的制备

[0229] 在1X NVT中制备中间文库储备液。将文库加热至95℃ 5分钟,然后冷却10分钟。

预富集(3轮)

[0230] 在SELEX轮次1之前,针对肽靶标预富集文库三轮。在预富集轮次(PE R)1中,文库在水中用热洗脱。在PE R2和R3之前,将另一个125pmol原始文库添加到热洗脱中。

储备钴琼脂糖珠的初始洗涤

[0231] 所需的珠子数被计算为在10mg/ml的珠子浓度下每个样品10μL珠子。通过涡旋>30秒或倾斜并旋转5分钟将珠子重新悬浮在小瓶中。从储备小瓶中取出珠子,并加入到1.5mL微量离心管中。将10μL珠子加入到96孔滤板的孔中。在添加之间涡旋1.5mL微量离心管以确保琼脂糖珠适当悬浮。将200μL 1x NVT添加到每个孔中。将板以1000x g旋转2分钟。这重复2次并清空收集板。

肽固定

[0232] 将200μL的1μM肽加入到孔中。将板振摇1小时。然后将板以1000x g离心2分钟,并清空收集板。将200μL 1x NVT添加到每个孔中。将板以1000x g离心2分钟。再重复固定程序2次并清空收集板。

针对珠子的阳性选择

[0233] 在NVT结合缓冲液中制备125pmol亲和试剂文库。将200μL在NVT结合缓冲液中的文库加入到每个孔中。将板振摇1小时。将板以1000x g离心2分钟,并清空收集板。将200μL 1x NVT添加到每个孔中。将板以1000x g离心2分钟,然后振摇10分钟。将板以1000x g离心2分钟,并清空收集板。该程序再重复2次。将收集板更换为干净的收集板。向每孔加入50μL 95℃的水。将板在70℃下振摇3分钟。将板以1000x g离心2分钟。

SELEX(6轮)

[0234] 使用原始和预富集的文库对长肽和短肽进行SELEX。前两轮的洗脱分为两组:无盐和肽竞争洗脱。从第3轮开始,所有样品都用热洗脱。

部分1:用于SELEX的珠子的初始洗涤

[0235] 所需的珠子数被计算为在10mg/ml的珠子浓度下每个样品10μL珠子。通过涡旋>30秒或倾斜并旋转5分钟将珠子重新悬浮在小瓶中。从储备小瓶中取出珠子,并加入到1.5mL微量离心管中。将10μL珠子加入到96孔滤板的孔中。在添加之间涡旋1.5mL微量离心管以确保琼脂糖珠适当悬浮。将200μL 1x NVT添加到每个孔中。将板以1000x g旋转2分钟。这重复2次并清空收集板。

部分2:肽固定

[0236] 将200μL的1μM肽加入到孔中。将板振摇1小时。然后将板以1000x g离心2分钟,并清空收集板。将200μL 1x NVT添加到每个孔中。将板以1000x g离心2分钟。再重复固定程序2次并清空收集板。

部分3:针对珠子的阳性选择

[0237] 将190μL NVT结合缓冲液加入到板的每个孔中。向每孔中加入10μL文库,并将板振

摇1小时。将板以1000x g离心2分钟，并清空收集板。将200μL 1x NVT添加到每个孔中，并将板振摇10分钟。这再重复2次。将50μL的5μM肽加入到肽洗脱孔中。将板振摇14分钟。将50μL洗脱缓冲液加入到孔中并将板振摇1分钟。将收集板更换为干净的收集板并以1000x g离心2分钟。将收集板更换为干净的收集板并向每孔中加入50μL 95℃的水。将板在70℃下振摇3分钟。将板以1000x g离心2分钟。

部分4:SELEX输出的扩增

[0238] 将内容物从肽/无盐板转移到qPCR板。制备两种不同的主混合物(一种用于EDTA洗脱，一种用于肽洗脱)。通过组合Kapa HiFi主混合物(1x最终)、正向引物/R-bio引物(1uM最终)、SYBR绿(1x最终)、EDTA(4mM最终)、40ul SELEX洗脱和水至最终体积，制备用于洗脱的主混合物。使用TissueLyser 3次以混合每个样品。将板以2000x g离心1分钟。每个样品扩增10个循环。将100μL主混合物添加到每个样品中。记录每个样品达到1000000-1250000FLU所需的PCR循环数。根据阳性对照样品峰达到稳定的位置，选择包括NTC在内的12个样品在生物分析仪上运行，以检查扩增子的大小以及浓度，以确保为SELEX的后续轮次生成足够的材料。

部分5:文库的重新生成

[0239] 通过组合10mM Tris-HCl、1mM EDTA和2M NaCl，并将溶液的pH调节至7.5，生成2x结合和洗涤缓冲液(B&W缓冲液)。所需的磁珠量被计算为300μl PCR反应需要75μl珠子。所需的NaOH量被计算为每75μl珠子需要80μl 20mM NaOH。

[0240] 在1mL 1x B&W缓冲液中洗涤磁珠(50μL珠子x 100μL PCR反应的数目)。使用磁性管架将珠子拉下并除去缓冲液。将珠子在相同体积的珠子中洗涤。去除缓冲液，并重复洗涤。将等体积的2x B&W缓冲液添加到珠子中。

[0241] 样品以MagMAX方案运行。位于200μl B&W中的梳子从深孔板中0μl体积处的梳子中取出。在室温下使用梳子将珠子与文库混合20分钟。将100μL PCR反应与100μL珠子在2x B&W缓冲液中合并，在深孔板中总体积为200μl。每孔中度混合10秒，然后中度混合4分50秒，然后缓慢混合10秒。该程序重复2次。

[0242] 收集带有结合的文库的Dynabeads并转移到洗液1。每个孔在深孔板中具有200μl的体积，并中度混合10秒。收集带有结合的文库的Dynabeads并转移到洗液2。每个孔在深孔板中具有200μl的体积，并中度混合10秒。收集带有结合的文库的Dynabeads并将其转移到NaOH板上，以供在20mM NaOH中洗脱。每个孔在深孔板中具有80μl的体积，并中度混合10分钟。收集不含文库的Dynabeads并转移到梳子板上以供释放。将NaOH洗脱内容物转移到PCR板，并加入5μl 20X NV缓冲液和20μl 40mM HCl。使用pH试纸(pH 7.2-7.4)检查pH。选择了一些样品，以在SYBR Gold 2%EX凝胶上检查它们的大小。所有样品都使用Oligreen测定进行定量。将Oligreen标准品在1x NVT中稀释至2ng/μL。在黑色96孔板中对标准品进行2倍稀释(最终体积为100μL)。将2μL ssDNA添加到板中(最终体积为100μL)。将200x Oligreen试剂稀释至1x。将100μL 1x Oligreen试剂添加到板上的每个孔中。在进行下一轮选择之前，在TecanF200上以485(20)nm/535(25)nm读取读数。

实施例8:游离溶液SELEX方案

预富集(3轮)

[0243] 在SELEX轮次1之前，针对肽靶标预富集文库三轮。在预富集轮次(PE R)1中，文库

在水中用热洗脱。在PE R2和R3之前，将另一个125pmol原始文库添加到热洗脱中。

部分1：用于SELEX的缓冲液的制备

[0244] 为基于琼脂糖珠的SELEX制备缓冲液。通过将47.66g HEPES(MW 238.30g/mol)、140.256g NaCl(MW 58.44g/mol)、9.5211g MgCl2(MW 95.211g/mol)和7.455g KCl(MW 74.55g/mol)溶解在750mL水中制备1L NVT的20x储备液[200mM HEPES，2.4M NaCl，100mM MgCl2，100mM KCl，2％吐温20pH 7.4]，并使用HCl/NaOH将pH调节至7.4。添加20mL吐温20，并使最终体积达到1L。通过将1x NVT与300μM硫酸葡聚糖和0.1mg/ml剪切鲑精DNA混合制备结合缓冲液。

部分2：文库的制备

[0245] 在1X NVT中制备中间文库储备液。将文库加热至95℃ 5分钟，然后冷却10分钟。

部分3：文库与溶液中的肽靶标的结合

[0246] 在MagMAX深孔板中的250μL结合缓冲液中制备100pmol肽。根据板的图加入125pmol条形码化的文库，并在室温下振摇该板1小时。在孵育过程中，每孔取10μL琼脂糖链霉亲和素珠(切掉移液管尖头的尖端)。样品在微量离心管中以500xg离心1分钟。取出上清液并记录体积。样品用1ml 1x NVT洗涤两次。将珠子重新悬浮在与原始上清液相同体积的1x NVT中，使得总体积为孔数的10x。向空的96孔微量离心柱的每个孔中加入10μL洗涤过的珠子，并用40μL结合缓冲液洗涤膜和珠子。将板以1000xg离心2分钟。将肽文库混合物与珠子合并，并在振摇下孵育10分钟。将板以1000xg离心2分钟，并在室温下用250μL 1xNVT洗孔10分钟。将板以1000xg离心2分钟，再重复洗涤两次。在洗涤过程中，将管式热混合器设置为95℃，而板式热混合器设置为70℃。将4ml水在管式热混合器中加热，然后每孔加入50μL加热的水。将孔在板式热混合器上于70℃混合孵育3分钟。向滤板装配新的收集板，并离心以收集流过液。将内容物转移到PCR板中以供保存。

SELEX(6轮)

[0247] 使用原始和预富集的文库对长肽和短肽进行SELEX。前两轮的洗脱分为两组：EDTA和肽竞争洗脱。从第3轮开始，所有样品都用EDTA洗脱。所有轮次都包括第二次热洗脱。

部分1：用于SELEX的文库的制备

[0248] 将文库加热至95℃ 5分钟，然后冷却10分钟。

部分2：文库与溶液中的肽靶标的结合

[0249] 在MagMAX深孔板中的250μL结合缓冲液中制备100pmol肽。根据板的图加入125pmol条形码化的文库，并在室温下振摇该板1小时。在孵育过程中，每孔取10μL琼脂糖链霉亲和素珠(切掉移液管尖头的尖端)。样品在微量离心管中以500xg离心1分钟。取出上清液并记录体积。样品用1ml 1x NVT洗涤两次。将珠子重新悬浮在与原始上清液相同体积的1x NVT中，使得总体积为孔数的10x。向空的96孔微量离心柱的每个孔中加入10μL洗涤过的珠子，并用40μL结合缓冲液洗涤膜和珠子。将板以1000xg离心2分钟。将肽文库混合物与珠子合并，并在振摇下孵育10分钟。将板以1000xg离心2分钟，并在室温下用250μL 1xNVT洗孔10分钟。将板以1000xg离心2分钟，再重复洗涤两次。在洗涤过程中，将管式热混合器设置为95℃，而板式热混合器设置为70℃。将4ml水在管式热混合器中加热，然后向每孔中加入50μL加热的水。将孔在板式热混合器上于70℃混合孵育3分钟。向滤板装配新的收集板，并离心以收集流过液。将内容物转移到PCR板中以供保存。部分3：SELEX输出的扩增

[0250] 将内容物从肽/EDTA板转移到qPCR板。制备两种不同的主混合物(一种用于EDTA洗脱,一种用于肽洗脱)。通过组合2x Kapa HiFi 100x1/2＝50μL、正向引物100x1/500＝0.2μL、反向引物100x1/500＝0.2μL、SyBr绿100x1/100＝1μL、MgCl2 100x4.5/500＝0.9μL、EDTA洗脱＝45μL(对于第1轮)和2.7μ水,以总体积100μL制备用于EDTA洗脱的主混合物。通过组合2x Kapa HiFi 100x1/2＝50μL、正向引物100x1/500＝0.2μL、反向引物100x1/500＝0.2μL、SyBr绿100x1/100＝1μL、EDTA 100x2.5/100＝2.5μL、肽洗脱＝45μL(对于第1轮)和1.1μL水,生成用于肽洗脱的主混合物。使用TissueLyser 3次以混合该板。将板以2000x g离心1分钟。qPCR程序以98℃ 3min、10个循环的98℃ 15s、65℃ 30s和72℃ 30s的热循环运行。每个样品扩增10个循环。

[0251] 创建了用于两种洗脱的一种主混合物,并向每个样品添加100μL。PCR总体积为200μL。主混合物是通过组合2x Kapa HiFi 100x1/2＝50μL、正向引物100x1/500＝0.2μL、反向引物100x1/500＝0.2μL、SyBr绿100x1/100＝1μL和48.6μL水而创建的。记录每个样品达到1000000-1250000FLU所需的PCR循环数。根据阳性对照样品峰达到稳定的位置,选择包括NTC在内的24个样品在生物分析仪上运行,以检查扩增子的大小以及浓度,以确保为后续轮次生成足够的材料。如果检测到的扩增子不足,则用PCR将样品再扩增一轮,以消除气泡形成并产生更多材料。

部分4:文库的重新生成

[0252] 通过组合10mM Tris-HCl、1mM EDTA和2M NaCl,并将溶液的pH调节至7.5,生成2x结合和洗涤缓冲液(B&W缓冲液)。所需珠子的量被计算为300μl PCR反应需要75μl珠子。所需的NaOH量被计算为每75μl珠子需要80μl 20mM NaOH。在1x B&W缓冲液中手动洗涤珠子。创建2个多孔洗涤板,每孔含有200μL 1XB&W。用75μL珠子生成1个多孔珠板,并添加至125μL 1xB&W。将梳子添加至该板。用300μL PCR反应和300μL 2X B&W缓冲液创建1个多孔文库板。在适当的孔中用80μL NaOH创建1个标准板。其余的孔用水填充。1个PCR收集板用20μL 40mM HCL和5μL 20x NVT生成。从珠板上取下梳子。如下所述将珠子与文库板中的亲和试剂文库混合用于文库结合,这是通过结合4次(中度混合4分55秒加上缓慢混合5秒)并收集珠子5次各30秒。将珠子洗涤两次(洗涤板1和2),这是通过释放珠子10秒,洗涤珠子10秒,以及收集珠子5次各30秒。使肽在NaOH中变性,这是通过将肽释放10秒,以中速混合10分钟并缓慢混合5秒,然后收集珠子5次各30秒。在珠板中更换梳子。

[0253] 所有样品都使用Oligreen测定进行定量。将Oligreen标准品在1x NVT中稀释至2ng/μL。在黑色96孔板中对标准品进行2倍稀释(最终体积为100μL)。将2μL ssDNA添加到板中(最终体积为100μL)。将200x Oligreen试剂稀释至1x。将100μL 1x Oligreen试剂添加到板上的每个孔中。在进行下一轮选择之前,在TecanF200上以485(20)nm/535(25)nm读取读数。

实施例9:链霉亲和素琼脂糖SELEX方案

预富集(3轮)

[0254] 在SELEX轮次1之前,针对肽靶标预富集文库三轮。在预富集轮次(PE R)1中,文库在水中用热洗脱。在PE R2和R3之前,将另一个125pmol原始文库添加到热洗脱中。

部分1:用于SELEX的缓冲液的制备

[0255] 为基于琼脂糖珠的SELEX制备缓冲液。通过将47.66g HEPES(MW 238.30g/mol)、

140.256g NaCl(MW 58.44g/mol)、9.5211g MgCl2(MW 95.211g/mol)和7.455g KCl(MW 74.55g/mol)溶解在750mL水中制备1L NVT的20x储备液[200mM HEPES,2.4M NaCl,100mM MgCl2,100mM KCl,2%吐温20pH 7.4],并使用HCl/NaOH将pH调节至7.4。添加20mL吐温20,并使最终体积达到1L。通过将1x NVT与300μM硫酸葡聚糖和0.1mg/ml剪切鲑精DNA混合制备结合缓冲液。洗涤缓冲液被制备为1x NVT。

部分2:用于SELEX的文库的制备

[0256] 在1X NVT中制备中间文库储备液。将文库加热至95℃ 5分钟,然后冷却10分钟。

部分3:预富集轮次

储备链霉亲和素琼脂糖珠的初始洗涤

[0257] 所需的链霉亲和素-琼脂糖珠的数目被计算为在10mg/ml的珠子浓度下每个样品使用10μL珠子。通过涡旋>30秒或倾斜并旋转5分钟将珠子重新悬浮在小瓶中。从储备小瓶中取出珠子,并加入到1.5mL微量离心管中。将10μL珠子加入到96孔滤板的孔中。在添加之间涡旋1.5mL微量离心管以确保琼脂糖珠适当悬浮。将200μL 1x NVT添加到每个孔中。将板以1000x g旋转2分钟。这重复2次并清空收集板。

肽固定

[0258] 将200μL的1μM肽加入到孔中。将板振摇1小时。然后将板以1000x g离心2分钟,并清空收集板。将200μL 1x NVT添加到每个孔中。将板以1000x g离心2分钟。再重复固定程序2次并清空收集板。将200μL的1μM生物素-PEG4添加到孔中。将板振摇10分钟。将板以1000x g旋转2分钟,并清空收集板。将200μL 1x NVT添加到每个孔中,并将板以1000x g旋转2分钟。再重复该洗涤2次并清空收集板。

针对珠子的阳性选择

[0259] 在NVT结合缓冲液中制备125pmol亲和试剂文库。将200μL在NVT结合缓冲液中的文库加入到每个孔中。将板振摇10分钟。将板以1000x g离心2分钟,并清空收集板。将200μL 1x NVT添加到每个孔中。将板以1000x g离心2分钟,然后振摇10分钟。将板以1000x g离心2分钟,并清空收集板。该程序再重复2次。将收集板更换为干净的收集板。向每孔加入50μL 95℃的水。将板在70℃下振摇3分钟。将板以1000x g离心2分钟。

SELEX 6轮

[0260] 使用原始和预富集的文库对长肽和短肽进行SELEX。前两轮的洗脱分为两组:EDTA和肽竞争洗脱。从第3轮开始,所有样品都用EDTA洗脱。所有轮次都包括第二次热洗脱。

部分1:用于SELEX的珠子的初始洗涤

[0261] 所需的链霉亲和素-琼脂糖珠的数目被计算为在10mg/ml的珠子浓度下每个样品使用10μL珠子。通过涡旋>30秒或倾斜并旋转5分钟将珠子重新悬浮在小瓶中。从储备小瓶中取出珠子,并加入到1.5mL微量离心管中。将10μL珠子加入到96孔滤板的孔中。在添加之间涡旋1.5mL微量离心管以确保琼脂糖珠适当悬浮。将200μL 1x NVT添加到每个孔中。将板以1000x g旋转2分钟。这重复2次并清空收集板。

部分2:肽固定

[0262] 将200μL的1μM肽加入到孔中。将板振摇1小时。然后将板以1000x g离心2分钟,并清空收集板。将200μL 1x NVT添加到每个孔中。将板以1000x g离心2分钟。再重复固定程序2次并清空收集板。将200μL的1μM生物素-PEG4添加到孔中。将板振摇10分钟。将板以1000x

g旋转2分钟，并清空收集板。将200μL 1x NVT添加到每个孔中，并将板以1000x g旋转2分钟。再重复该洗涤2次，然后清空收集板。

部分3:针对珠子的阳性选择

[0263] 在NVT结合缓冲液中制备125pmol亲和试剂文库。将200μL在NVT结合缓冲液中的文库加入到每个孔中。将板振摇1小时。将板以1000x g离心2分钟，并清空收集板。将200μL 1x NVT添加到每个孔中。将板以1000x g离心2分钟，然后振摇10分钟。将板以1000x g离心2分钟，并清空收集板。该程序再重复2次。将50μL的0.5μM肽加入到肽洗脱孔中，并将板振摇14分钟。将50μL 10mM EDTA加入到EDTA洗脱孔中并将板振摇1分钟。将收集板更换为干净的收集板。向每孔加入50μL 95℃的水。将板在70℃下振摇3分钟。将板以1000x g离心2分钟。

部分4:SELEX输出的扩增

[0264] 将内容物从肽/无盐板转移到qPCR板。制备两种不同的主混合物(一种用于EDTA洗脱，一种用于肽洗脱)。通过组合Kapa HiFi主混合物(1x最终)、正向引物/R-bio引物(1uM最终)、SYBR绿(1x最终)、EDTA(4mM最终)、40ul SELEX洗脱和水至最终体积，制备用于洗脱的主混合物。使用TissueLyser 3次以混合每个样品。将板以2000x g离心1分钟。每个样品扩增10个循环。将100μL主混合物添加到每个样品中。记录每个样品达到1000000-1250000FLU所需的PCR循环数。根据阳性对照样品峰达到稳定的位置，选择包括NTC在内的12个样品在生物分析仪上运行，以检查扩增子的大小以及浓度，以确保为SELEX的后续轮次生成足够的材料。

部分5:文库的重新生成

[0265] 将内容物从肽/无盐板转移到qPCR板。制备两种不同的主混合物(一种用于EDTA洗脱，一种用于肽洗脱)。通过组合Kapa HiFi主混合物(1x最终)、正向引物/R-bio引物(1uM最终)、SYBR绿(1x最终)、EDTA(4mM最终)、40ul SELEX洗脱和水至最终体积，制备用于洗脱的主混合物。使用TissueLyser 3次以混合每个样品。将板以2000x g离心1分钟。每个样品扩增10个循环。将100μL主混合物添加到每个样品中。记录每个样品达到1000000-1250000FLU所需的PCR循环数。根据阳性对照样品峰达到稳定的位置，选择包括NTC在内的12个样品在生物分析仪上运行，以检查扩增子的大小以及浓度，以确保为SELEX的后续轮次生成足够的材料。

[0266] 通过组合10mM Tris-HCl、1mM EDTA和2M NaCl，并将溶液的pH调节至7.5，生成2x结合和洗涤缓冲液(B&W缓冲液)。所需的磁珠量被计算为300μl PCR反应需要75μl珠子。所需的NaOH量被计算为每75μl珠子需要80μl 20mM NaOH。

[0267] 在1mL 1x B&W缓冲液中洗涤磁珠(50μL珠子x 100μL PCR反应的数目)。使用磁性管架将珠子拉下并除去缓冲液。将珠子在相同体积的珠子中洗涤。去除缓冲液，并重复洗涤。将等体积的2x B&W缓冲液添加到珠子中。

[0268] 样品以MagMAX方案运行。位于200μl B&W中的梳子从深孔板中0μl体积处的梳子中取出。在室温下使用梳子将珠子与文库混合20分钟。将100μL PCR反应与100μL珠子在2x B&W缓冲液中合并，在深孔板中总体积为200μl。每孔中度混合10秒，然后中度混合4分50秒，然后缓慢混合10秒。该程序重复2次。

[0269] 收集带有结合的文库的Dynabeads并转移到洗液1。每个孔在深孔板中具有200μl的体积，并中度混合10秒。收集带有结合的文库的Dynabeads并转移到洗液2。每个孔在深孔

板中具有200μl的体积，并中度混合10秒。收集带有结合的文库的Dynabeads并将其转移到NaOH板上，以供在20mM NaOH中洗脱。每个孔在深孔板中具有80μl的体积，并中度混合10分钟。收集不含文库的Dynabeads并转移到梳子板上以供释放。将NaOH洗脱内容物转移到PCR板，并加入5μl 20X NV缓冲液和20μl 40mM HCl。使用pH试纸(pH 7.2-7.4)检查pH。选择了一些样品，以在SYBR Gold 2%EX凝胶上检查它们的大小。所有样品都使用Oligreen测定进行定量。将Oligreen标准品在1x NVT中稀释至2ng/μL。在黑色96孔板中对标准品进行2倍稀释(最终体积为100μL)。将2μL ssDNA添加到板中(最终体积为100μL)。将200x Oligreen试剂稀释至1x。将100μL 1x Oligreen试剂添加到板上的每个孔中。在进行下一轮选择之前，在TecanF200上以485(20)nm/535(25)nm读取读数。

实施例10:池ELONA测定

[0270] 创建至少15μl的100μM亲和试剂池，以及单独的生物素化寡核苷酸。使用2x NVT缓冲液(10mM HEPES,120mM NaCl,5mM MgCl2,5mM KCl,pH 7.4,含0.01%吐温20)将亲和试剂池稀释至50μM。将文库加热至95℃ 5分钟，然后冷却10分钟。

ELONA方案

[0271] ELONA板(Pierce™ Streptavidin Coated High Capacity Plates,透明,384孔)用100μl 1X NV缓冲液+0.1%吐温20洗涤两次。将板与在25μL肽结合缓冲液(含0.1mg/ml sssDNA(剪切鲑精DNA)的1X NV缓冲液)中的1μM生物素化肽一起孵育，并以约500RPM振摇30分钟。链霉亲和素板的每个孔中加入总共100pmol的肽。对于接受生物素而不是肽的孔，加入1.6μM生物素。用100μl 1X NV缓冲液+0.1%吐温20将板洗涤6次。每个孔用100μL SuperG封闭缓冲液+100μM D-生物素封闭，并在室温下以约500RPM摇动30分钟。在封闭孵育期间，稀释并滴定重新折叠的适体。封闭后，将板用100μl 1X NV缓冲液+0.1%吐温20洗涤6次。将折叠的适体在25μL含有100uM D-生物素的适体结合缓冲液(含有0.1mg/ml sssDNA和300μM硫酸葡聚糖的1X NV缓冲液)中加入到孔中，并以约500RPM振摇1小时。

[0272] 在孵育期结束前5分钟制备检测溶液(在适体结合缓冲液中的0.5ug/ml(储备液的1:1000稀释)链霉亲和素HRP缀合物)。将板用100μl 1X NV+0.1%吐温20洗涤6次。将25μL检测溶液添加到样品孔中并以约500RPM振摇30分钟。用100μl 1X NV+0.1%吐温20将板洗涤6次。将25μl室温TMB底物加入到每个孔中并孵育1分30秒。通过向每个孔中加入25μl 1N HCl来终止显色。颜色变为黄色，在450nm处具有最大吸光度。在Tecan Spark上测量每个孔在450nm处的吸光度。

实施例11:游离溶液SELEX亲和试剂富集

[0273] 如实施例8中所述，在溶液中进行生物素化肽和适体文库结合一小时。结合后，肽连同结合的适体一起被固定在链霉亲和素琼脂糖珠上。3次10分钟洗涤后，将结合的适体在EDTA中洗脱，PCR扩增，并再生为单链文库。该过程重复6轮，然后通过下一代测序评估富集。针对具有包含表位LRDLRDLRDLRD或TQATQATQATQA的肽的肽文库测试亲和试剂文库。

[0274] 图13A显示了针对序列LRDLRDLRDLRD测试的亲和试剂的亲和试剂富集数据。观察到结合该表位的序列的可检测富集。图13B显示了针对序列TQATQATQATQA测试的亲和试剂的亲和试剂富集数据。观察到结合该表位的序列的可检测富集，尽管在6轮SELEX后观察到的富集少于针对表位LRDLRDLRDLRD所观察到的富集。

实施例12:使用磁珠的SELEX亲和试剂富集

[0275] 如实施例6中所述，将叠氮化物修饰的肽缀合到磁性炔烃珠上。缀合后，将肽结合的珠与适体文库一起孵育1小时，并在3次10分钟洗涤后，将结合的适体在EDTA中洗脱，PCR扩增，并再生为单链文库。该过程重复6轮，然后通过下一代测序评估富集。针对具有包含表位YSLYSLYSLYSL或EVT的肽的肽文库测试亲和试剂文库。

[0276] 图14A显示了针对序列YSLYSLYSLYSL测试的亲和试剂的亲和试剂富集数据。观察到结合该表位的序列的可检测富集。图14B显示了针对序列EVT测试的亲和试剂的亲和试剂富集数据。观察到结合该表位的序列的可检测富集，尽管在6轮SELEX后观察到的富集少于针对表位YSLYSLYSLYSL所观察到的富集。

实施例13:概率亲和结合模型

[0277] 使用概率亲和结合模型从结合数据集学习三种亲和试剂的三聚体结合概率，该结合数据集包括每种亲和试剂与500种独特的、随机选择的酵母蛋白质的结合。对于每种亲和试剂，对500种蛋白质中的每一种的10,000个拷贝积累了结合测量值(结合或不结合)。每种亲和试剂以0.25的概率结合5种三聚体，以及以更低的概率结合其他三聚体。对于这三种亲和试剂中的每一种，将使用这种方法预测的三聚体结合概率(y轴)与亲和试剂的真实结合概率(x轴)相比较地作图。绘制的虚线表示预测概率与真实概率之间的完美一致性。图15A、16A和17A显示了对数变换的概率，为清楚起见，示出了图中四个忙区(busy region)中的三聚体计数。此外，在500种蛋白质的组中不存在的68种三聚体使用不同的阴影来绘制，尽管它们可能被图中的其他点遮挡。图15B、16B和17B显示了没有对概率进行对数变换的相同数据。

[0278] 虽然本文已经显示并描述了本发明的优选实施方案，但是对于本领域技术人员显而易见的是，这些实施方案仅以示例的方式提供。并非打算用本说明书中提供的具体实例来限制本发明。尽管已经参照上述说明书对本发明进行了描述，但并不意味着对本文实施方案的描述和说明以限制性的意义来解释。在不脱离本发明的情况下，本领域技术人员现将想到许多变化、改变和替换。此外，应当理解，本发明的所有方面均不限于本文所阐述的具体描述、配置或相对比例，其取决于多种条件和变量。应当理解，在实施本发明的过程中可以采用本文所述的本发明实施方案的各种替代方案。因此可以预期，本发明还应涵盖任何这类替代、改变、变化或等同物。旨在以所附权利要求书限定本发明的范围，由此涵盖在这些权利要求范围内的方法和结构及其等同物。

[0279] 尽管有所附权利要求书，但本文阐述的公开内容还由以下条款定义：

1.一种鉴定针对表位的亲和试剂的方法，其包括：

获得包含第一多个肽的第一肽文库，每个肽包含式αXβ的序列，其中X是所需表位，并且其中α和β是包含已知氨基酸序列的侧翼结构域；

将所述第一肽文库暴露于多种亲和试剂，从而使所述多种亲和试剂中的至少一种亲和试剂与所述第一肽文库结合；

确定与所述第一多个肽结合的所述至少一种亲和试剂的一个或多个序列，从而形成缩减的亲和试剂池；

获得包含第二多个肽的第二肽文库，每个肽包含式γXδ的序列，其中X是所需表位，并且其中γ和δ是包含已知氨基酸序列的侧翼结构域；

将所述第二肽文库暴露于所述缩减的亲和试剂池，从而使所述缩减的亲和试剂池

中的至少一种亲和试剂与所述第二肽文库结合;以及

确定与所述第二肽文库结合的所述缩减的亲和试剂池中的至少一种亲和试剂的一个或多个序列,从而鉴定针对表位X的亲和试剂。

2.根据条款1所述的方法,其中α和β包含相同的氨基酸序列。

3.根据条款2所述的方法,其中α和β包含单一氨基酸。

4.根据条款3所述的方法,其中α和β是相同的单一氨基酸。

5.根据条款4所述的方法,其中α和β是丙氨酸或甘氨酸。

6.根据条款1所述的方法,其中α和β包含不同的氨基酸序列。

7.根据条款1所述的方法,其中γ和δ包含相同的氨基酸序列。

8.根据条款7所述的方法,其中γ和δ包含单一氨基酸。

9.根据条款8所述的方法,其中γ和δ是相同的单一氨基酸。

10.根据条款9所述的方法,其中γ和δ是精氨酸、天冬氨酸、谷氨酰胺、脯氨酸或色氨酸。

11.根据条款1所述的方法,其中γ和δ包含不同的氨基酸序列。

12.根据条款1所述的方法,其中α或β包含与γ和δ不同的氨基酸序列。

13.根据条款12所述的方法,其中α和β包含与γ和δ不同的氨基酸序列。

14.根据条款13所述的方法,其中α和β包含相同的第一氨基酸序列并且γ和δ包含相同的第二氨基酸序列。

15.根据条款1所述的方法,其中X包含长度为m的氨基酸序列。

16.根据条款15所述的方法,其中m处于2至7个氨基酸的范围内。

17.根据条款16所述的方法,其中m为3个氨基酸。

18.根据条款1所述的方法,其中X包含在所述多个肽的一级结构中邻近的氨基酸序列。

19.根据条款1所述的方法,其中X包含在所述多个肽的二级、三级或四级结构中邻近的氨基酸序列。

20.根据条款1所述的方法,其中所述多种亲和试剂中的至少0.001%与所述第一肽文库结合。

21.根据条款20所述的方法,其中所述多种亲和试剂中的至少1%与所述第一肽文库结合。

22.根据条款1所述的方法,其中所述缩减的亲和试剂池的不超过0.001%与所述第二肽文库结合。

23.根据条款1所述的方法,其中所述缩减的亲和试剂池的不超过0.000001%与所述第二肽文库结合。

24.根据条款1所述的方法,其中所述多种亲和试剂中的亲和试剂包含寡核苷酸。

25.根据条款24所述的方法,其中所述寡核苷酸包含条形码。

26.根据条款24所述的方法,其中所述寡核苷酸包含用于测序反应的衔接子。

27.根据条款24所述的方法,其中所述寡核苷酸包含可检测标记或标签。

28.根据条款27所述的方法,其中所述可检测标记或标签包括荧光团、生物发光标记、放射性标记或酶标签。

29.根据条款1所述的方法,其中所述确定与所述第一多个肽结合的所述亲和试剂的一个或多个序列包括:

从所述第一肽库中分离亲和试剂;以及

在分离后,对所述亲和试剂进行测序,从而确定所述亲和试剂的一个或多个序列。

30.根据条款1所述的方法,其中所述确定与所述第二肽文库结合的所述缩减的亲和试剂池中的至少一种亲和试剂的一个或多个序列包括:

从所述第二肽文库分离所述至少一种亲和试剂;以及

在分离后,对所述至少一种亲和试剂进行测序,从而确定所述至少一种亲和试剂的一个或多个序列。

31.根据条款29或30所述的方法,其中所述测序包括进行下一代测序反应。

32.根据条款29或30所述的方法,其中所述分离包括从所述肽洗脱所述至少一种亲和试剂。

33.根据条款32所述的方法,其中洗脱所述至少一种亲和试剂包括使用洗脱缓冲液。

34.根据条款32所述的方法,其中所述洗脱缓冲液包含离液剂。

35.根据条款32所述的方法,其中洗脱所述至少一种亲和试剂包括使所述至少一种亲和试剂与水接触,其中水处于至少50℃的温度。

36.根据条款1所述的方法,其中α和β包含具有中性电荷的氨基酸序列。

37.根据条款1所述的方法,其中α或β包含具有非中性电荷的氨基酸序列。

38.根据条款1所述的方法,其中α或β包含不会空间阻断表位X的至少一部分的氨基酸序列。

39.根据条款1所述的方法,其中α和β包含空间阻断表位X的至少一部分的氨基酸序列。

40.根据条款1所述的方法,其中α和β包含具有均一的化学多样性的氨基酸序列。

41.根据条款40所述的方法,其中所述具有均一的化学多样性的氨基酸序列包括包含两个或更多个具有等同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

42.根据条款1所述的方法,其中α和β包含具有不均一的化学多样性的氨基酸序列。

43.根据条款42所述的方法,其中所述具有不均一的化学多样性的氨基酸序列包括包含两个或更多个具有不同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

44.根据条款1所述的方法,其中γ和δ包含具有中性电荷的氨基酸序列。

45.根据条款1所述的方法,其中γ和δ包含具有非中性电荷的氨基酸序列。

46.根据条款1所述的方法,其中γ和δ包含不会空间阻断表位X的至少一部分的氨基酸序列。

47.根据条款1所述的方法,其中γ和δ包含空间阻断表位X的至少一部分的氨基酸序列。

48.根据条款1所述的方法,其中γ和δ包含具有均一的化学多样性的氨基酸序列。

49.根据条款48所述的方法，其中所述具有均一的化学多样性的氨基酸序列包括包含两个或更多个具有等同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

50.根据条款1所述的方法，其中γ和δ包含具有不均一的化学多样性的氨基酸序列。

51.根据条款48所述的方法，其中所述具有不均一的化学多样性的氨基酸序列包括包含两个或更多个具有不同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

52.根据条款1所述的方法，其中所述亲和试剂池进一步包含液体介质。

53.根据条款52所述的方法，其中所述液体介质是水溶液。

54.根据条款53所述的方法，其中所述水溶液是缓冲的溶液。

55.根据条款1所述的方法，其进一步包括洗涤与所述第一肽文库结合的所述多种亲和试剂中的所述至少一种亲和试剂。

56.根据条款1所述的方法，其进一步包括洗涤与所述第二肽文库结合的所述缩减的亲和试剂池中的所述至少一种亲和试剂。

57.根据条款55或56所述的方法，其中所述洗涤包括使用洗涤缓冲液。

58.根据条款57所述的方法，其中所述洗涤缓冲液包含去污剂、表面活性剂或离液剂。

59.根据条款1所述的方法，其中所述第一或第二肽文库内的肽处于天然状态。

60.根据条款1所述的方法，其中所述第一或第二肽文库内的肽处于非天然状态。

61.根据条款1所述的方法，其中所述非天然状态是变性状态或部分折叠状态。

62.根据条款60或61所述的方法，其进一步包括对所述处于天然状态的多个肽重复所有步骤。

63.根据条款62所述的方法，其中所述至少一种针对处于天然状态的表位X的亲和试剂不同于所述至少一种针对处于非天然状态的表位X的亲和试剂。

64.根据条款62所述的方法，其中所述至少针对处于天然状态的表位X的亲和试剂与所述至少一种针对处于非天然状态的表位X的亲和试剂相同。

65.根据条款1所述的方法，其中所述表位X与多于一种亲和试剂结合。

66.根据条款1所述的方法，其中所述至少一种亲和试剂与多于一个表位结合。

67.根据条款66所述的方法，其中所述多于一个表位包含简并氨基酸序列。

68.根据条款67所述的方法，其中所述简并氨基酸序列相差一个氨基酸。

69.根据条款67所述的方法，其中所述简并氨基酸序列相差多于一个氨基酸。

70.根据条款67、68或69所述的方法，其中所述差异氨基酸包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

71.根据条款67、68或69所述的方法，其中所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

72.根据条款1所述的方法，其中所述第一肽文库的所述多个肽或所述第二肽文库的所述多个肽与基底结合。

73.根据条款1所述的方法，其中所述第一肽文库的所述多个肽或所述第二肽文库

的所述多个肽不与基底结合。

74.根据条款72所述的方法,其中所述基底包括流动池。

75.根据条款1所述的方法,其包括重复一次或多次将所述第一肽文库暴露于所述多种亲和试剂以及确定所述至少一种亲和试剂的一个或多个序列的步骤。

76.根据条款1所述的方法,其包括重复一次或多次将所述第二肽文库暴露于所述缩减的亲和试剂池以及确定所述至少一种亲和试剂的一个或多个序列的步骤。

77.根据条款1所述的方法,其中所述多种亲和试剂中的亲和试剂包含寡核苷酸、肽、微型蛋白质结合物、抗体、抗体片段或其组合。

78.一种鉴定针对表位的亲和试剂的方法,其包括:

获得包含多个肽的肽文库,其中所述多个肽中的肽在其序列中包含多于一个表位X;

将所述肽文库暴露于包含多种亲和试剂的亲和试剂池,从而使至少一种亲和试剂与在其序列中包含多于一个表位X的肽结合;以及

确定与在其序列中包含多于一个表位X的肽结合的所述至少一种亲和试剂的一个或多个序列,从而鉴定针对表位X的亲和试剂。

79.根据条款78所述的方法,其中所述多于一个表位X是连续的。

80.根据条款78所述的方法,其中所述多于一个表位X是不连续的。

81.根据条款80所述的方法,其中所述多于一个表位X包含间插氨基酸序列。

82.根据条款81所述的方法,其中所述多个肽中的肽包含由包含相同间插序列的氨基酸序列隔开的表位X的至少三个实例。

83.根据条款81所述的方法,其中所述多个肽中的肽包含由包含不同间插序列的氨基酸序列隔开的表位X的至少三个实例。

84.根据条款78所述的方法,其中所述多个肽中的肽进一步包含表位Y。

85.根据条款84所述的方法,其中所述表位Y与表位X相连。

86.根据条款85所述的方法,其中所述表位Y与每个表位X相连。

87.根据条款84所述的方法,其中所述表位Y与每个表位X不相连。

88.根据条款87所述的方法,其中每个表位Y在所述肽的氨基酸序列中相对于表位X有规律地间隔开。

89.根据条款87所述的方法,其中至少两个表位Y在所述肽的氨基酸序列中相对于表位X随机间隔开。

90.根据条款78所述的方法,其中X包含长度为m的氨基酸序列。

91.根据条款90所述的方法,其中m处于2至7个氨基酸的范围内。

92.根据条款91所述的方法,其中m为3个氨基酸。

93.根据条款79所述的方法,其中所述多于一个连续表位X进一步包含侧翼结构域α和β。

94.根据条款93所述的方法,其中α和β包含相同的氨基酸序列。

95.根据条款94所述的方法,其中α和β包含单一氨基酸。

96.根据条款95所述的方法,其中α和β是相同的单一氨基酸。

97.根据条款96所述的方法,其中α和β是丙氨酸或甘氨酸。

98.根据条款93所述的方法，其中α和β包含不同的氨基酸序列。

99.根据条款78所述的方法，其中所述多种亲和试剂中的亲和试剂包含寡核苷酸。

100.根据条款99所述的方法，其中所述寡核苷酸包含条形码。

101.根据条款99所述的方法，其中所述寡核苷酸包含用于测序反应的衔接子。

102.根据条款99所述的方法，其中所述寡核苷酸包含可检测标记或标签。

103.根据条款102所述的方法，其中所述可检测标记或标签包括荧光团、生物发光标记、放射性标记或酶标签。

104.根据条款78所述的方法，其中所述确定与所述肽文库结合的所述至少一种亲和试剂的一个或多个序列包括：

从所述肽文库分离至少一种亲和试剂；以及

对所述至少一种亲和试剂进行测序，从而确定所述至少一种亲和试剂的一个或多个序列。

105.根据条款104所述的方法，其中所述测序包括进行下一代测序反应。

106.根据条款104所述的方法，其中所述分离包括从所述肽洗脱所述至少一种亲和试剂。

107.根据条款106所述的方法，其中洗脱所述至少一种亲和试剂包括使用洗脱缓冲液。

108.根据条款106所述的方法，其中所述洗脱缓冲液包含离液剂。

109.根据条款106所述的方法，其中洗脱所述至少一种亲和试剂包括使所述至少一种亲和试剂与水接触，其中水处于至少50℃的温度。

110.根据条款93所述的方法，其中α和β包含具有均一的化学多样性的氨基酸序列。

111.根据条款110所述的方法，其中所述具有均一的化学多样性的氨基酸序列包括包含两个或更多个具有等同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

112.根据条款93所述的方法，其中α和β包含具有不均一的化学多样性的氨基酸序列。

113.根据条款112所述的方法，其中所述具有不均一的化学多样性的氨基酸序列包括包含两个或更多个具有不同电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合的氨基酸侧基的氨基酸序列。

114.根据条款78所述的方法，其中所述亲和试剂池进一步包含液体介质。

115.根据条款114所述的方法，其中所述液体介质是水溶液。

116.根据条款115所述的方法，其中所述水溶液是缓冲的溶液。

117.根据条款78所述的方法，其进一步包括洗涤与所述第一肽文库结合的所述多种亲和试剂中的所述至少一种亲和试剂。

118.根据条款78所述的方法，其进一步包括洗涤与所述肽结合的所述亲和试剂池中的所述至少一种亲和试剂。

119.根据条款117或118所述的方法，其中所述洗涤包括使用洗涤缓冲液。

120.根据条款119所述的方法，其中所述洗涤缓冲液包含去污剂、表面活性剂或离

液剂。

121.根据条款78所述的方法,其中所述肽文库内的肽处于天然状态。

122.根据条款78所述的方法,其中所述肽文库内的肽处于非天然状态。

123.根据条款78所述的方法,其中所述非天然状态是变性状态或部分折叠状态。

124.根据条款122或123所述的方法,其进一步包括对所述处于天然状态的多个肽重复所有步骤。

125.根据条款124所述的方法,其中所述至少一种针对处于天然状态的表位X的亲和试剂不同于所述至少一种针对处于非天然状态的表位X的亲和试剂。

126.根据条款124所述的方法,其中所述至少针对处于天然状态的表位X的亲和试剂与所述至少一种针对处于非天然状态的表位X的亲和试剂相同。

127.根据条款78所述的方法,其中所述表位X与多于一种亲和试剂结合。

128.根据条款78所述的方法,其中所述至少一种亲和试剂与多于一个表位结合。

129.根据条款128所述的方法,其中所述多于一个表位包含简并氨基酸序列。

130.根据条款129所述的方法,其中所述简并氨基酸序列相差一个氨基酸。

131.根据条款129所述的方法,其中所述简并氨基酸序列相差多于一个氨基酸。

132.根据条款129、130或131所述的方法,其中所述差异氨基酸包括相似的化学性质,其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

133.根据条款129、130或131所述的方法,其中所述差异氨基酸包括不相似的化学性质,其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

134.根据条款78所述的方法,其中所述肽文库的所述多个肽与基底结合。

135.根据条款78所述的方法,其中所述肽文库的所述多个肽不与基底结合。

136.根据条款134所述的方法,其中所述基底包括流动池。

137.一种鉴定针对表位的亲和试剂的方法,其包括:

使第一介质中肽文库内的多个肽与第二介质中亲和试剂池内的多种亲和试剂接触以形成肽-亲和试剂复合物,其中每个肽-亲和试剂复合物包含肽和至少一种结合的亲和试剂,并且其中每个肽包含表位X;

收集所述肽-亲和试剂复合物;

分离未结合的亲和试剂;以及

确定所述肽-亲和试剂复合物的亲和试剂的一个或多个序列,从而鉴定至少一种针对表位X的亲和试剂。

138.根据条款137所述的方法,其中所述第一介质或第二介质包括水溶液。

139.根据条款138所述的方法,其中所述水溶液包括缓冲的溶液。

140.根据条款138所述的方法,其中所述第一介质和第二介质包括缓冲的水溶液。

141.根据条款137所述的方法,其中所述肽文库内的肽处于天然状态。

142.根据条款137所述的方法,其中所述肽文库内的肽处于非天然状态。

143.根据条款137所述的方法,其中所述非天然状态是变性状态或部分折叠状态。

144.根据条款142或143所述的方法,其进一步包括对所述处于天然状态的多个肽重复所有步骤。

145.根据条款144所述的方法,其中所述至少一种针对处于天然状态的表位X的亲

和试剂不同于所述至少一种针对处于非天然状态的表位X的亲和试剂。

146.根据条款145所述的方法，其中所述至少一种针对处于天然状态的表位X的亲和试剂与所述至少一种针对处于非天然状态的表位X的亲和试剂相同。

147.根据条款137所述的方法，其中所述表位X与多于一种亲和试剂结合。

148.根据条款137所述的方法，其中所述至少一种亲和试剂与多于一个表位结合。

149.根据条款148所述的方法，其中所述多于一个表位包含简并氨基酸序列。

150.根据条款149所述的方法，其中所述简并氨基酸序列相差一个氨基酸。

151.根据条款149所述的方法，其中所述简并氨基酸序列相差多于一个氨基酸。

152.根据条款149、150或151所述的方法，其中所述差异氨基酸包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

153.根据条款149、150或151所述的方法，其中所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

154.根据条款137所述的方法，其中所述收集肽-亲和试剂复合物包括使所述肽-亲和试剂复合物与基底结合。

155.根据条款137所述的方法，其中所述收集肽-亲和试剂复合物包括从所述第一或第二介质分离所述肽-亲和试剂复合物；。

156.根据条款137所述的方法，其中X包含长度为m的氨基酸序列。

157.根据条款156所述的方法，其中m处于2至7个氨基酸的范围内。

158.根据条款157所述的方法，其中m为3个氨基酸。

159.根据条款137所述的方法，其中X包含在所述多个肽的一级结构中邻近的氨基酸序列。

160.根据条款137所述的方法，其中X包含在所述多个肽的二级、三级或四级结构中邻近的氨基酸序列。

161.根据条款137所述的方法，其中所述多种亲和试剂中的亲和试剂包含寡核苷酸。

162.根据条款161所述的方法，其中所述寡核苷酸包含条形码。

163.根据条款161所述的方法，其中所述寡核苷酸包含用于测序反应的衔接子。

164.根据条款161所述的方法，其中所述寡核苷酸包含可检测标记或标签。

165.根据条款161所述的方法，其中所述可检测标记或标签包括荧光团、生物发光标记、放射性标记或酶标签。

166.根据条款137所述的方法，其中所述确定与所述多个肽结合的所述至少一种亲和试剂的一个或多个序列包括：

将所述至少一种亲和试剂与所述多个肽分离；以及

对所述至少一种亲和试剂进行测序，从而确定所述至少一种亲和试剂的一个或多个序列。

167.根据条款166所述的方法，其中所述测序包括进行下一代测序反应。

168.根据条款166所述的方法，其中所述分离包括从所述肽洗脱所述至少一种亲和试剂。

169.根据条款168所述的方法，其中洗脱所述至少一种亲和试剂包括使用洗脱缓

冲液。

170.根据条款169所述的方法，其中所述洗脱缓冲液包含离液剂。

171.根据条款168所述的方法，其中洗脱所述至少一种亲和试剂包括使所述至少一种亲和试剂与水接触，其中水处于至少50℃的温度。

172.根据条款137所述的方法，其进一步包括洗涤与所述多个肽结合的所述多种亲和试剂中的所述至少一种亲和试剂。

173.根据条款172所述的方法，其中所述洗涤包括使用洗涤缓冲液。

174.根据条款173所述的方法，其中所述洗涤缓冲液包含去污剂、表面活性剂或离液剂。

175.一种包含$S_1S_2S_3$结构的亲和试剂，其中S_1和S_3是能够结合互补衔接子的衔接子结构域，并且其中S_2是表位结合域，其中所述表位结合域以小于10^{-6}M的平衡结合常数K_d与所述一个或多个肽表位结合。

176.根据条款175所述的亲和试剂，其中所述亲和试剂包含寡核苷酸。

177.根据条款175所述的亲和试剂，其中所述衔接子结构域是下一代测序衔接子。

178.根据条款175所述的亲和试剂，其中S_1或S_3包含条形码。

179.根据条款175所述的亲和试剂，其中S_1或S_3包含链霉亲和素标签。

180.根据条款176所述的方法，其中所述寡核苷酸包含可检测标记或标签。

181.根据条款180所述的亲和试剂，其中所述可检测标记或标签包括荧光团、生物发光标记、放射性标记或酶标签。

182.根据条款176所述的亲和试剂，其中S_2包含不超过约90个核苷酸。

183.根据条款182所述的亲和试剂，其中S_2包含不超过约80个核苷酸。

184.根据条款183所述的亲和试剂，其中S_2包含不超过约70个核苷酸。

185.根据条款184所述的亲和试剂，其中S_2包含不超过约60个核苷酸。

186.根据条款185所述的亲和试剂，其中S_2包含不超过约50个核苷酸。

187.根据条款186所述的亲和试剂，其中S_2包含不超过约40个核苷酸。

188.根据条款187所述的亲和试剂，其中S_2包含不超过约30个核苷酸。

189.根据条款175所述的亲和试剂，其中所述亲和试剂以小于10^{-6}M的平衡结合常数K_d与多于一个表位结合。

190.根据条款189所述的亲和试剂，其中所述多于一个表位包含简并氨基酸序列。

191.根据条款190所述的亲和试剂，其中所述简并氨基酸序列相差一个氨基酸。

192.根据条款190所述的亲和试剂，其中所述简并氨基酸序列相差多于一个氨基酸。

193.根据条款190、191或192所述的亲和试剂，其中所述差异氨基酸包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

194.根据条款190、191或192所述的亲和试剂，其中所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

195.根据条款189所述的亲和试剂，其中所述亲和试剂以小于10^{-7}M的平衡结合常数K_d与一个或多个表位结合。

196.根据条款195所述的亲和试剂，其中所述亲和试剂以小于10^{-8}M的平衡结合常数K_d与一个或多个表位结合。

197.根据条款196所述的亲和试剂，其中所述亲和试剂以小于10^{-9}M的平衡结合常数K_d与一个或多个表位结合。

198.根据条款197所述的亲和试剂，其中所述亲和试剂以小于10^{-10}M的平衡结合常数K_d与一个或多个表位结合。

199.根据条款198所述的亲和试剂，其中所述亲和试剂以小于10^{-11}M的平衡结合常数K_d与一个或多个表位结合。

200.根据条款199所述的亲和试剂，其中所述亲和试剂以小于10^{-12}M的平衡结合常数K_d与一个或多个表位结合。

201.根据条款175所述的亲和试剂，其中所述亲和试剂与一个或多个表位结合至少1分钟的时间段。

202.根据条款175所述的亲和试剂，其中所述亲和试剂与一个或多个表位结合至少15分钟的时间段。

203.根据条款202所述的亲和试剂，其中所述亲和试剂与一个或多个表位结合至少1小时的时间段。

204.根据条款175所述的亲和试剂，其中当所述肽处于天然状态时，所述亲和试剂与一个或多个肽表位结合。

205.根据条款175所述的亲和试剂，其中当所述肽处于非天然状态时，所述亲和试剂与一个或多个肽表位结合。

206.根据条款205所述的亲和试剂，其中所述非天然状态包括变性状态或部分折叠状态。

207.根据条款175所述的亲和试剂，其中所述亲和试剂进一步包含液体介质。

208.根据条款207所述的亲和试剂，其中所述液体介质包括水性介质。

209.一种亲和试剂池，其包含：

多种亲和试剂；

其中所述多种亲和试剂中的两种或更多种亲和试剂包含结构$S_1S_2S_3$，其中S_1和S_3包含共同的衔接子结构域，并且S_2包含表位结合域；并且

其中所述两种或更多种亲和试剂中的第一亲和试剂包含具有表征的针对肽表位X_1的特异性的第一结构域S_2，并且其中所述两种或更多种亲和试剂中的第二亲和试剂包含具有表征的针对肽表位X_2的特异性的第二结构域S_2，其中所述第一结构域S_2不同于所述第二结构域S_2。

210.根据条款209所述的亲和试剂池，其中X_1包含与X_2不同的氨基酸序列。

211.根据条款209所述的亲和试剂池，其中X_1包含与X_2相同的氨基酸序列。

212.根据条款209所述的亲和试剂池，其进一步包含第三亲和试剂，该第三亲和试剂包含具有表征的针对肽表位X_3的特异性的第三结构域S_2，其中第三结构域S_2不同于第一结构域S_2和第二结构域S_2。

213.根据条款212所述的亲和试剂池，其中X_3包含与X_1或X_2不同的氨基酸序列。

214.根据条款213所述的亲和试剂池，其中X_3包含与X_1和X_2不同的氨基酸序列。

215.根据条款214所述的亲和试剂池，其中X_3包含与X_1和X_2相同的氨基酸序列。

216.根据条款2015所述的亲和试剂池，其中X_3包含与X_1和X_2相同的氨基酸序列。

217.根据条款209所述的亲和试剂池，其中所述两种或更多种亲和试剂包含寡核苷酸。

218.根据条款209所述的亲和试剂池，其中所述共同衔接子结构域是下一代测序衔接子。

219.根据条款209所述的亲和试剂池，其中S_1或S_3包含条形码。

220.根据条款209所述的亲和试剂池，其中S_1或S_3包含链霉亲和素标签。

221.根据条款217所述的亲和试剂池，其中所述寡核苷酸包含可检测标记或标签。

222.根据条款221所述的亲和试剂池，其中所述可检测标记或标签包括荧光团、生物发光标记、放射性标记或酶标签。

223.根据条款217所述的亲和试剂池，其中S_2包含不超过约90个核苷酸。

224.根据条款223所述的亲和试剂池，其中S_2包含不超过约80个核苷酸。

225.根据条款224所述的亲和试剂池，其中S_2包含不超过约70个核苷酸。

226.根据条款225所述的亲和试剂池，其中S_2包含不超过约60个核苷酸。

227.根据条款226所述的亲和试剂池，其中S_2包含不超过约50个核苷酸。

228.根据条款227所述的亲和试剂池，其中S_2包含不超过约40个核苷酸。

229.根据条款228所述的亲和试剂池，其中S_2包含不超过约30个核苷酸。

230.根据条款209所述的亲和试剂池，其中所述第一亲和试剂和所述第二亲和试剂以小于10^{-6}M的平衡结合常数K_d与所述肽表位结合。

231.根据条款230所述的亲和试剂池，其中X_1和X_2包含简并氨基酸序列。

232.根据条款231所述的亲和试剂池，其中所述简并氨基酸序列相差一个氨基酸。

233.根据条款231所述的亲和试剂池，其中所述简并氨基酸序列相差多于一个氨基酸。

234.根据条款231、232或233所述的亲和试剂池，其中所述差异氨基酸包括相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

235.根据条款231、232或233所述的亲和试剂池，其中所述差异氨基酸包括不相似的化学性质，其中所述化学性质选自电荷、疏水性、亲水性、空间大小、极性、分子结构或其组合。

236.根据条款230所述的亲和试剂池，所述第一亲和试剂和所述第二亲和试剂以小于10^{-7}M的平衡结合常数K_d与所述肽表位结合。

237.根据条款236所述的亲和试剂池，所述第一亲和试剂和所述第二亲和试剂以小于10^{-8}M的平衡结合常数K_d与所述肽表位结合。

238.根据条款237所述的亲和试剂池，其中所述第一亲和试剂和所述第二亲和试剂以小于10^{-9}M的平衡结合常数K_d与所述肽表位结合。

239.根据条款238所述的亲和试剂池，其中所述第一亲和试剂和所述第二亲和试剂以小于10^{-10}M的平衡结合常数K_d与所述肽表位结合。

240.根据条款239所述的亲和试剂池，其中所述第一亲和试剂和所述第二亲和试

剂以小于10^{-11}M的平衡结合常数K_d与所述肽表位结合。

241.根据条款240所述的亲和试剂池，其中所述第一亲和试剂和所述第二亲和试剂以小于10^{-12}M的平衡结合常数K_d与所述肽表位结合。

242.根据条款209所述的亲和试剂池，其中所述第一亲和试剂和所述第二亲和试剂与所述肽表位结合至少1分钟的时间段。

243.根据条款209所述的亲和试剂池，其中所述第一亲和试剂和所述第二亲和试剂与所述肽表位结合至少15分钟的时间段。

244.根据条款209所述的亲和试剂池，所述第一亲和试剂和所述第二亲和试剂与所述肽表位结合至少1小时的时间段。

245.根据条款209所述的亲和试剂池，其中当所述肽处于天然状态时，所述第一亲和试剂和所述第二亲和试剂与所述肽表位结合。

246.根据条款209所述的亲和试剂池，其中当所述肽处于非天然状态时，所述第一亲和试剂和所述第二亲和试剂与所述肽表位结合。

247.根据条款246所述的亲和试剂池，其中所述非天然状态包括变性状态或部分折叠状态。

248.根据条款209所述的亲和试剂池，其中所述亲和试剂池进一步包含液体介质。

249.根据条款248所述的亲和试剂池，其中所述液体介质包括水性介质。

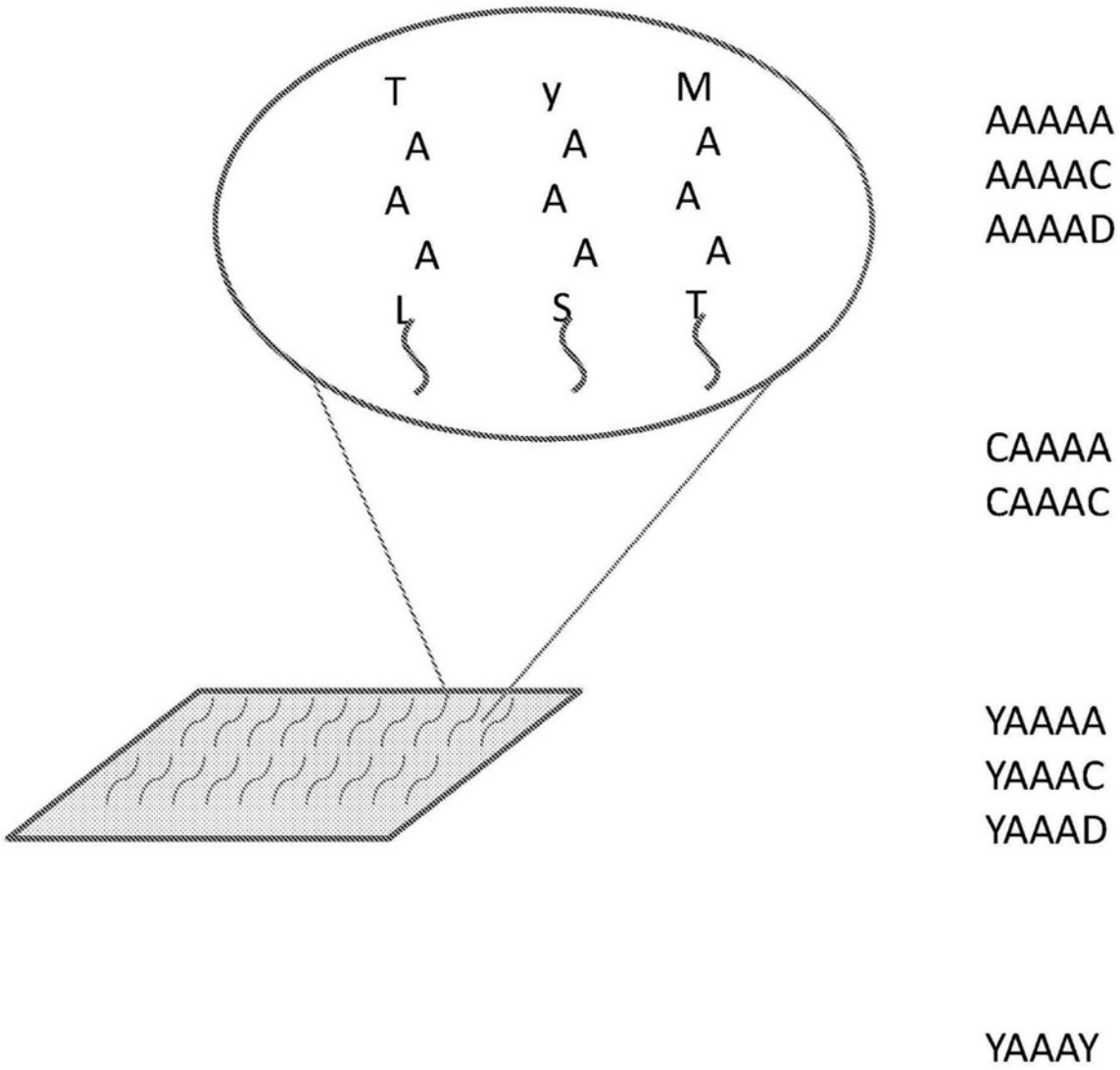

T
A
A
A
L
y
A
A
A
S
M
A
A
A
T
AAAAA
AAAAC
AAAAD
CAAAA
CAAAC
YAAAA
YAAAC
YAAAD
YAAAY

图1

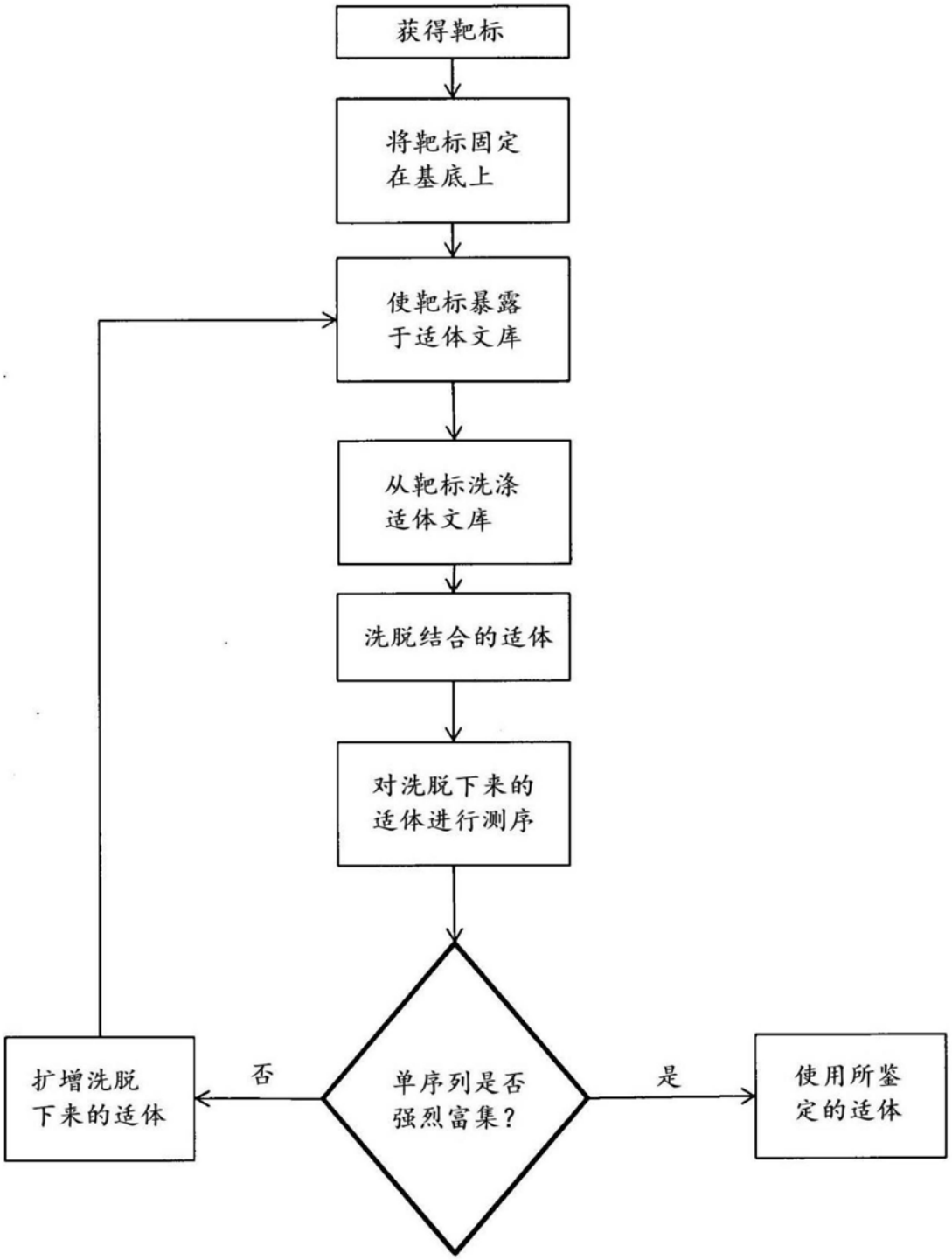

获得靶标
将靶标固定在基底上
使靶标暴露于适体文库
从靶标洗涤适体文库
洗脱结合的适体
对洗脱下来的适体进行测序
单序列是否强烈富集？
否
是
扩增洗脱下来的适体
使用所鉴定的适体

图2

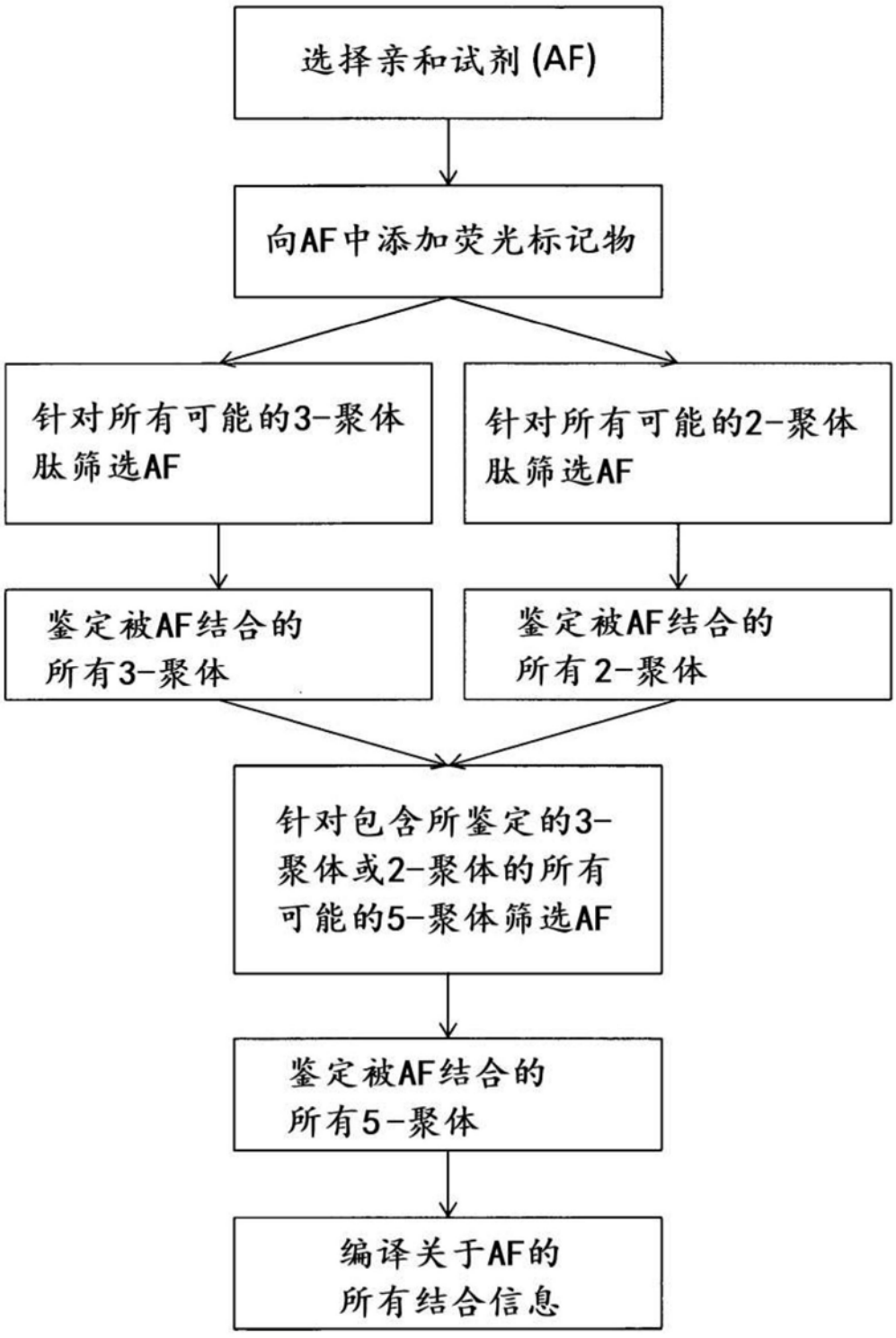

选择亲和试剂(AF)
向AF中添加荧光标记物
针对所有可能的3-聚体肽筛选AF
针对所有可能的2-聚体肽筛选AF
鉴定被AF结合的所有3-聚体
鉴定被AF结合的所有2-聚体
针对包含所鉴定的3-聚体或2-聚体的所有可能的5-聚体筛选AF
鉴定被AF结合的所有5-聚体
编译关于AF的所有结合信息

图3A

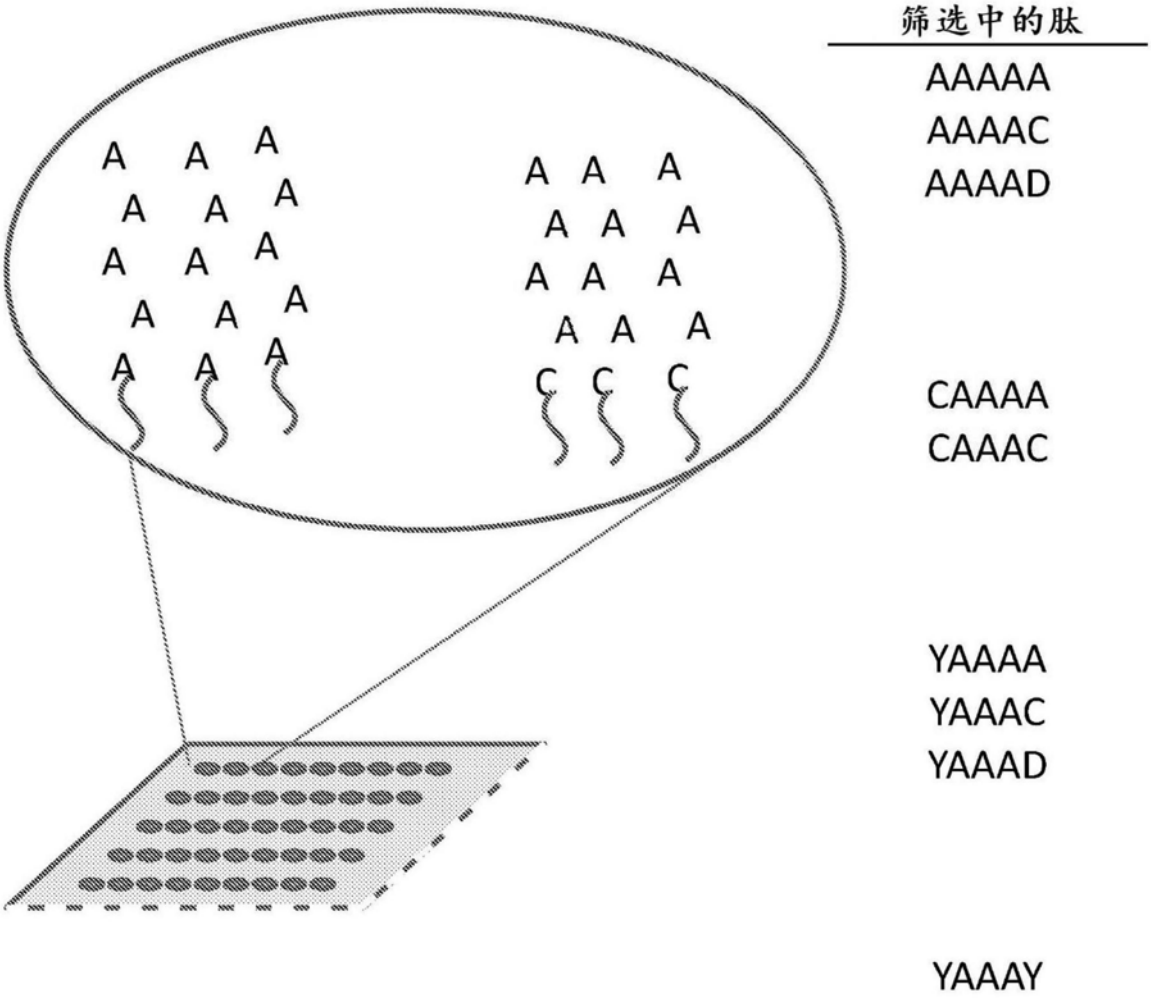

筛选中的肽
AAAAA
AAAAC
AAAAD
CAAAA
CAAAC
YAAAA
YAAAC
YAAAD
YAAAY

图3B

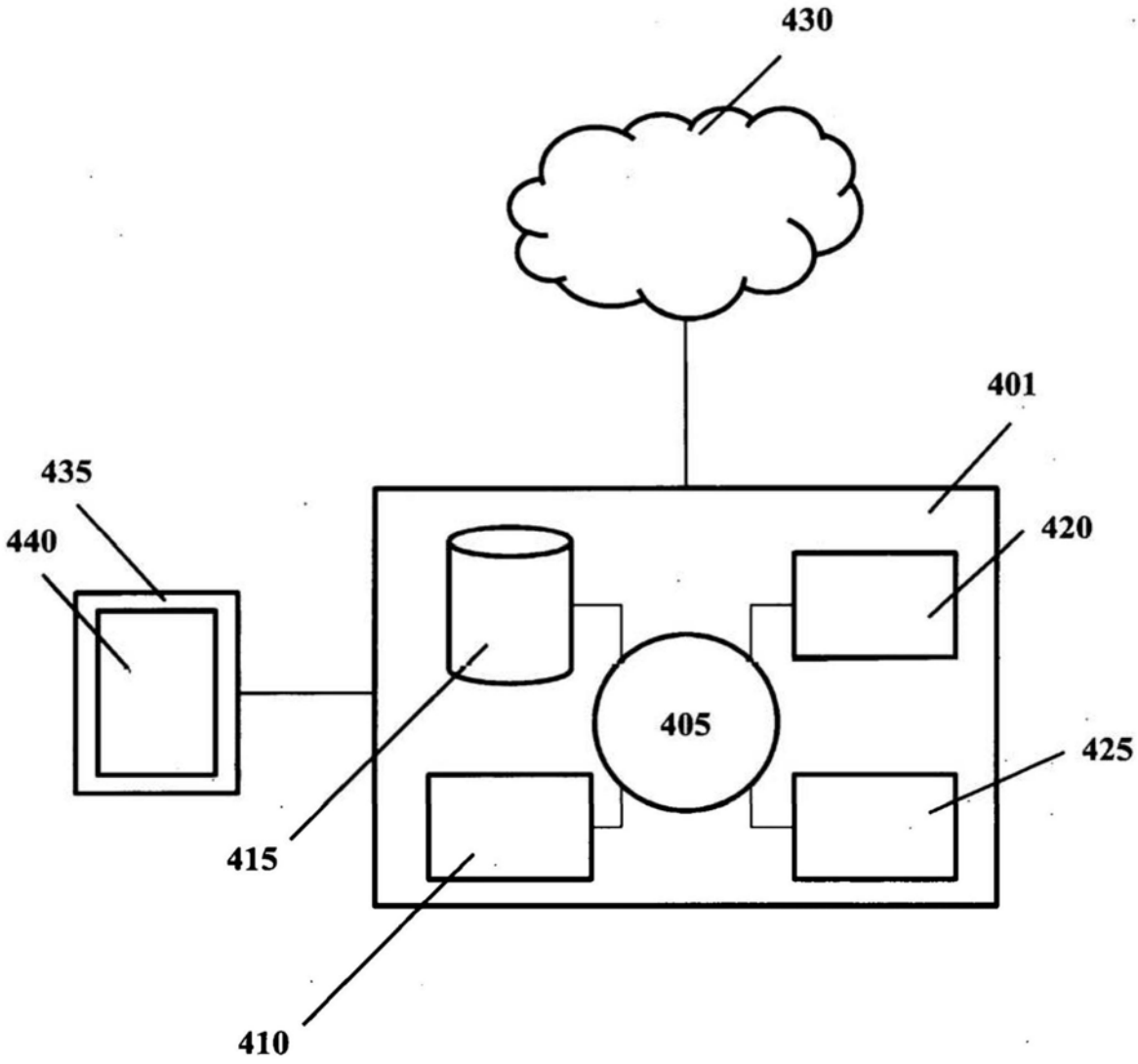


图4

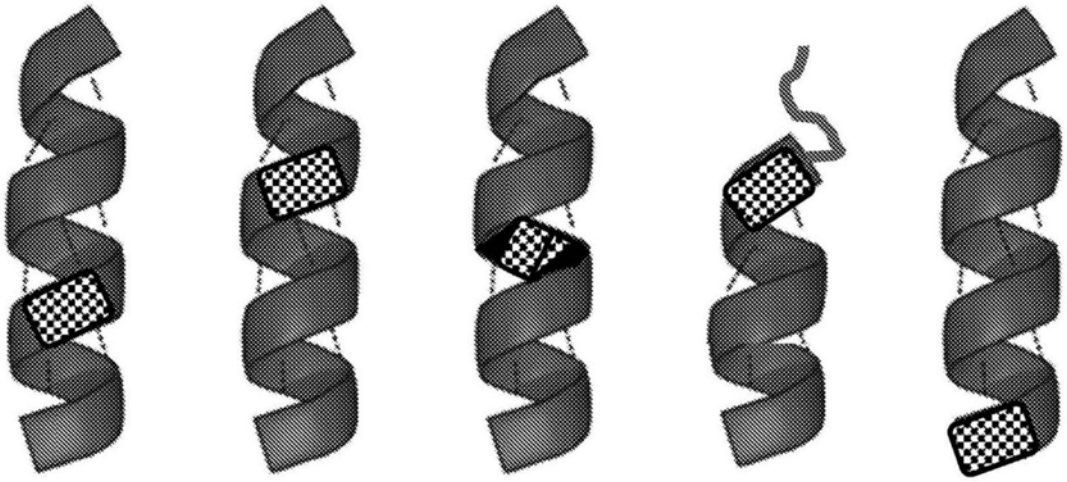


图5A

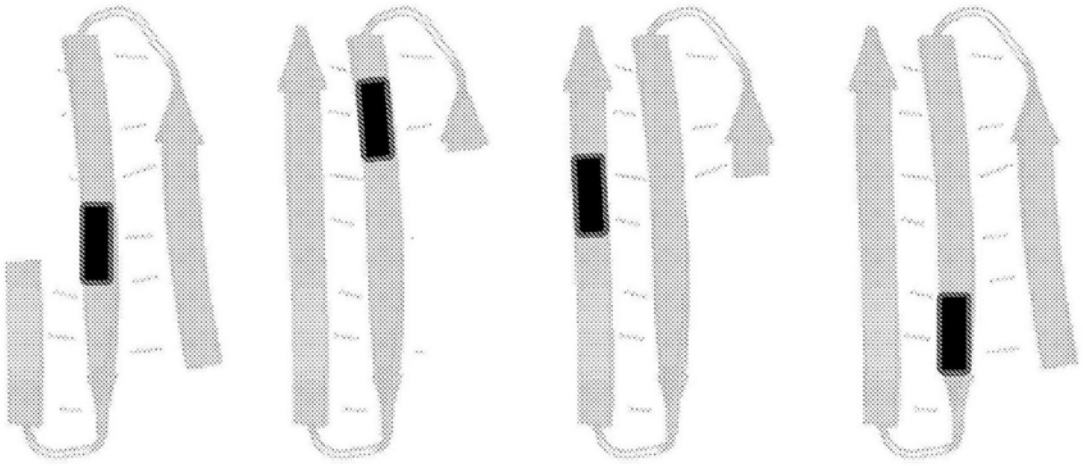


图5B

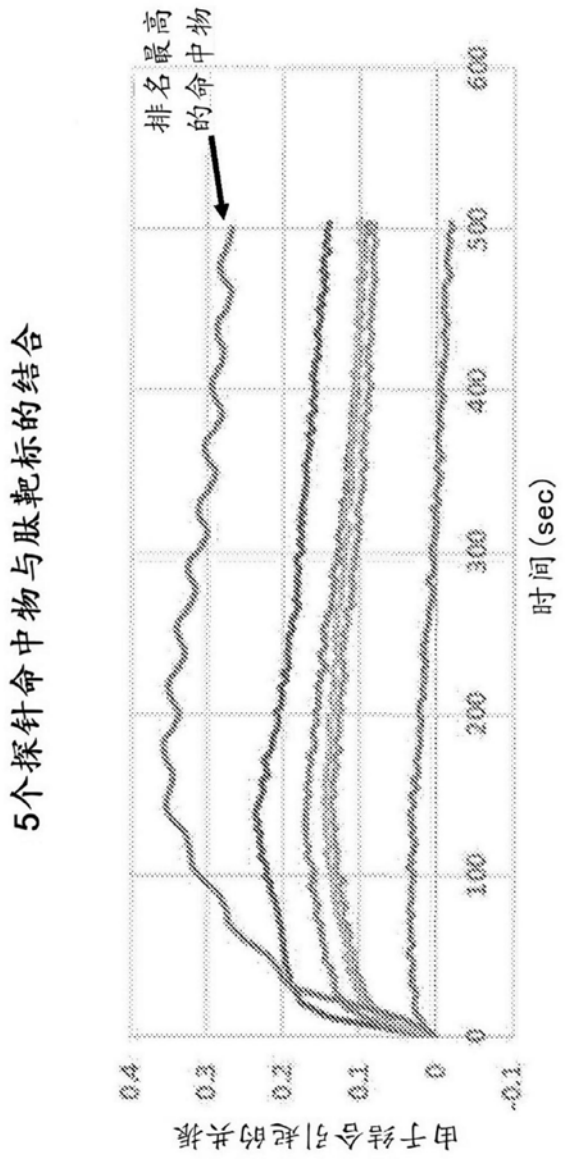

5个探针命中物与肽靶标的结合
排名最高的命中物
由于结合引起的共振
0.4
0.3
0.2
0.1
0
-0.1
0
100
200
300
400
500
600
时间(sec)

图6

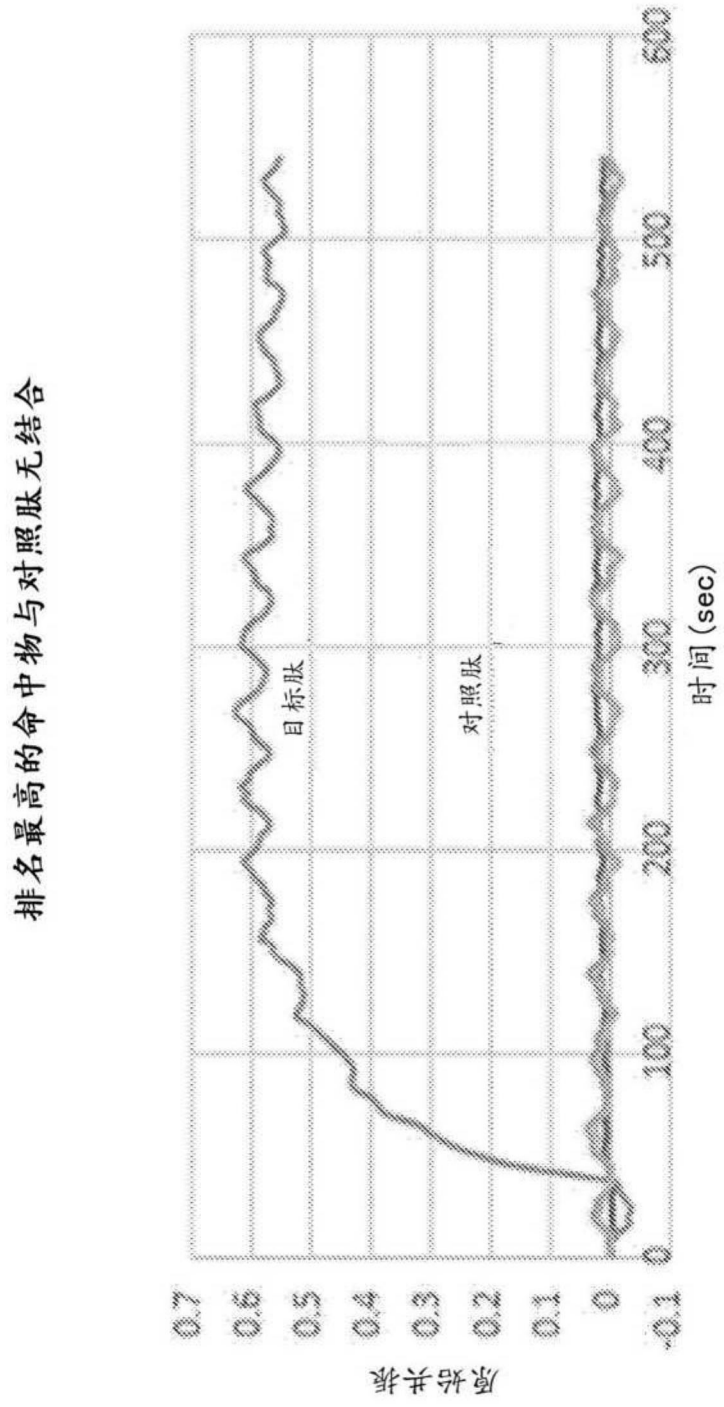

排名最高的命中物与对照肽无结合
目标肽
对照肽
原始共振
时间(sec)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
0
100
200
300
400
500
600

图7

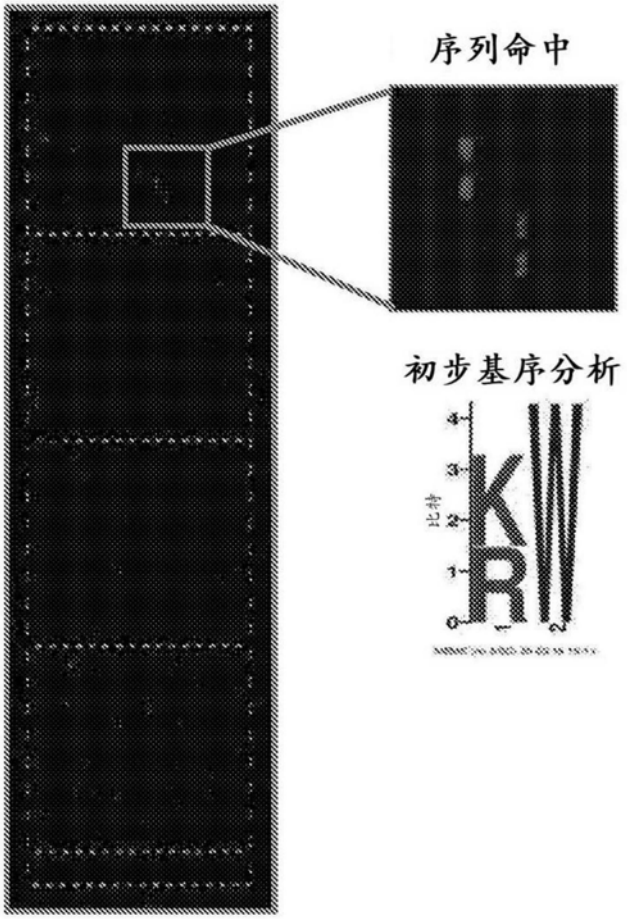


图8

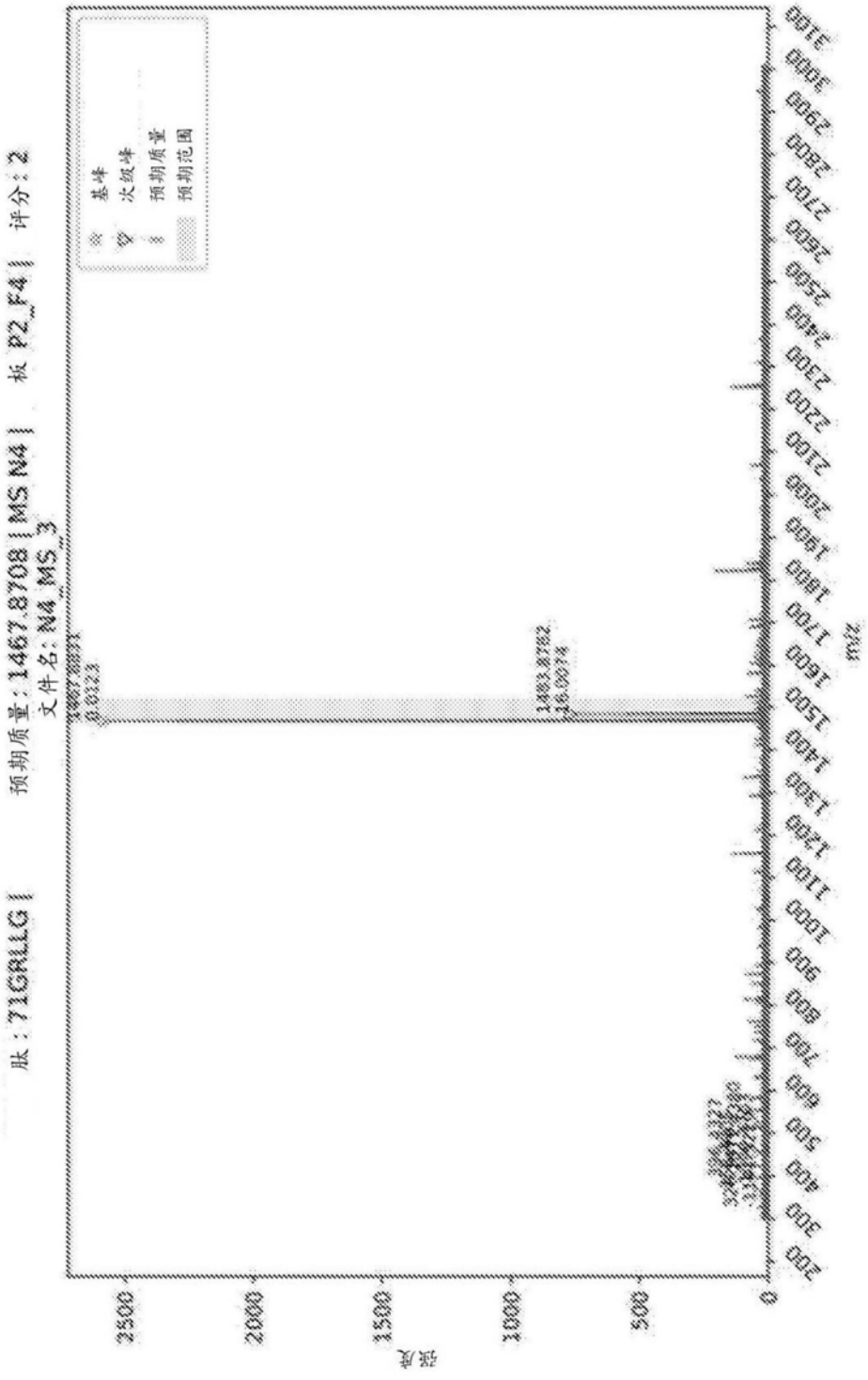

肽：71GRLLG | 预期质量：1467.8708 | MS N4 | 板 P2_F4 | 评分：2
文件名：N4_MS_3
基峰
次级峰
预期质量
预期范围
强度
m/z
2500
2000
1500
1000
500
0

图9

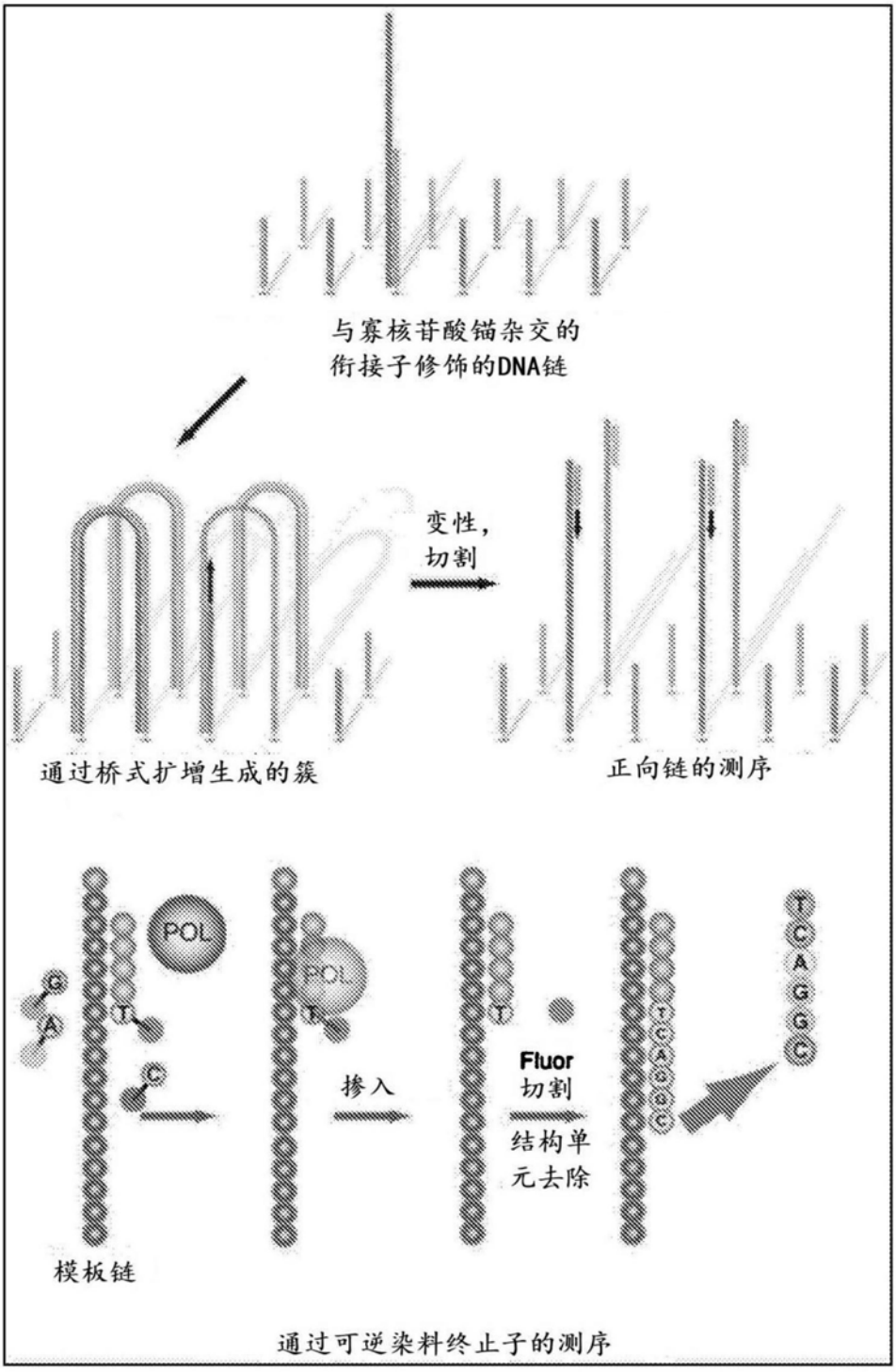

与寡核苷酸锚杂交的
衔接子修饰的DNA链
变性，
切割
通过桥式扩增生成的簇
正向链的测序
POL
G
A
T
C
掺入
Fluor
切割
结构单
元去除
T
C
A
G
G
C
模板链
通过可逆染料终止子的测序

图10

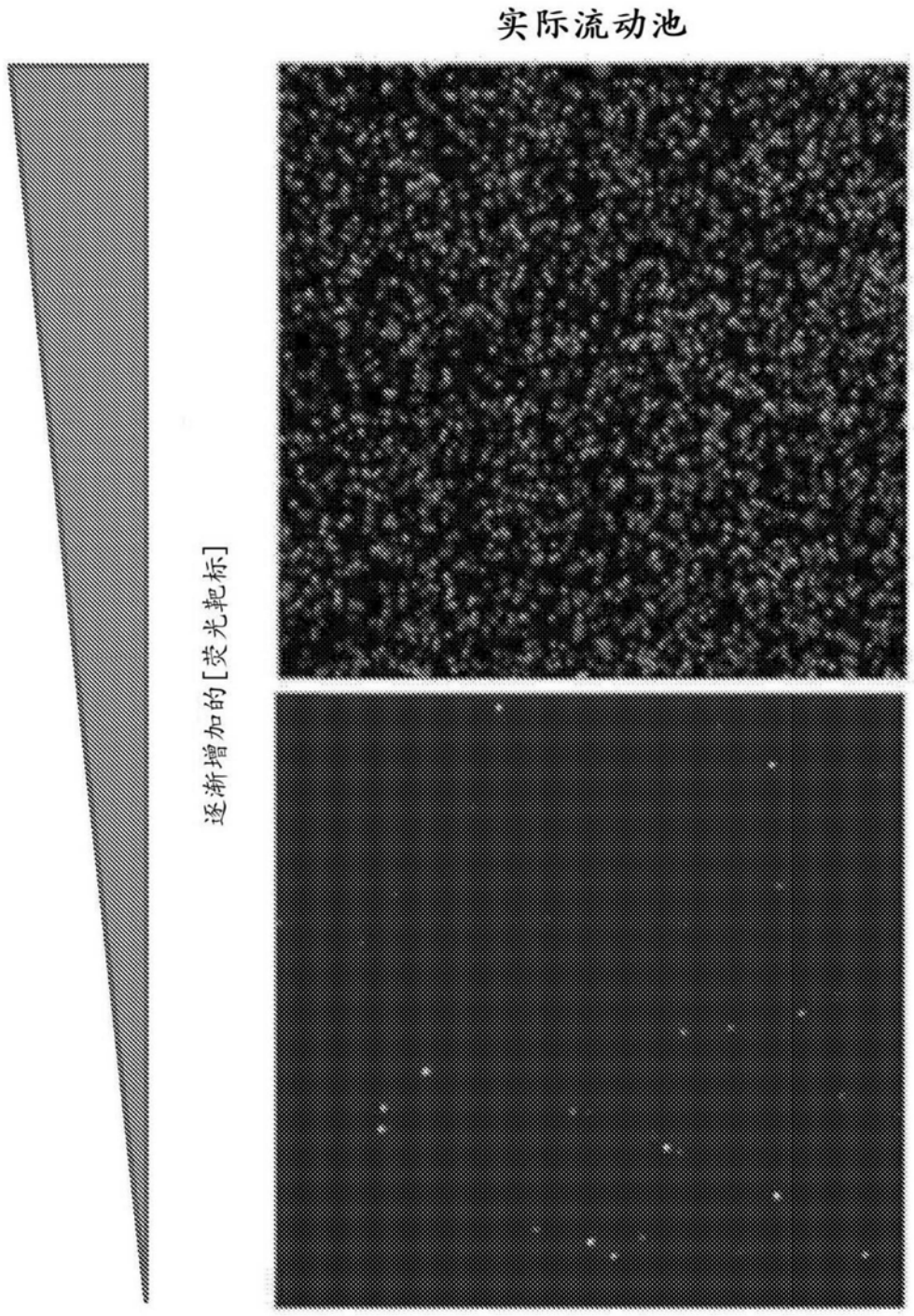


图11

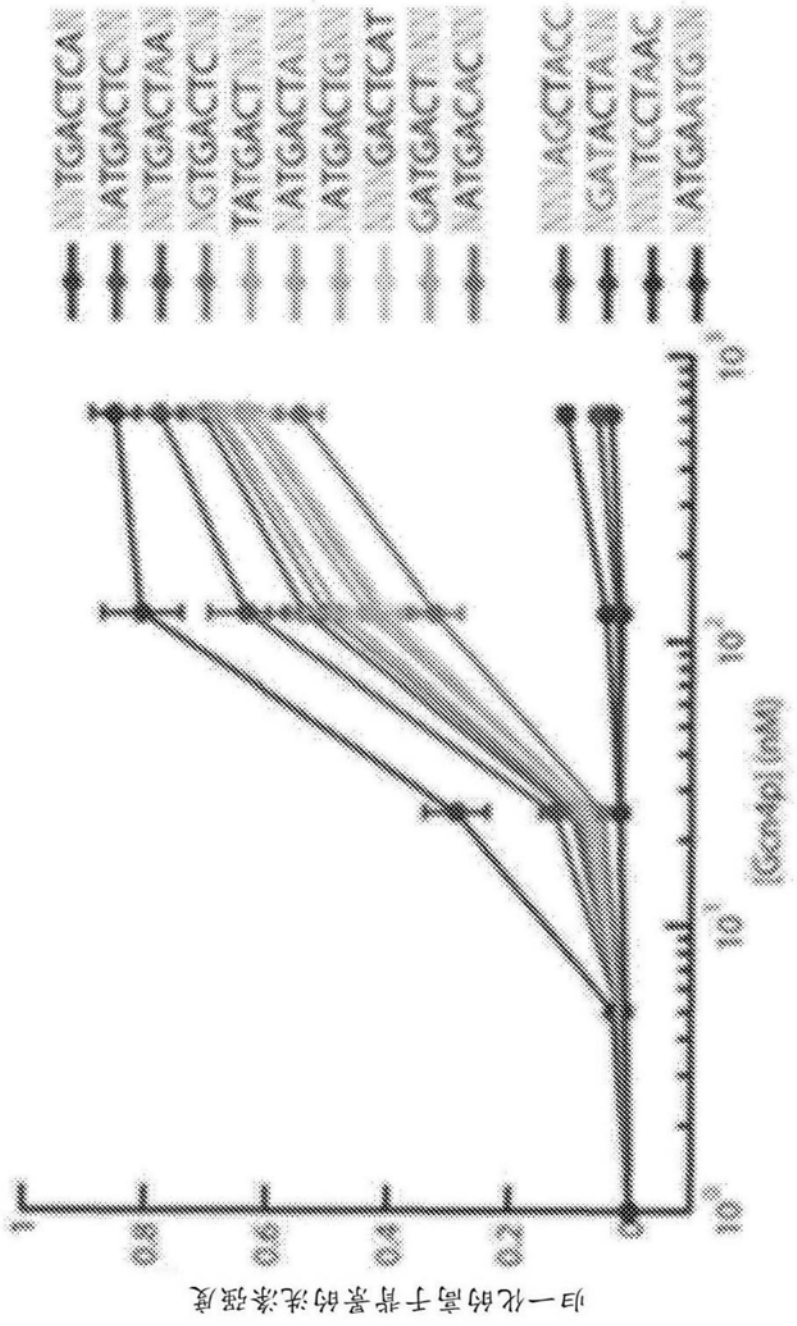

TGACTCA
ATGACTC
TGACTAA
GTGACTC
TATGACT
ATGACTA
ATGACTG
GACTCAT
GATGACT
ATGACAC
AGCTACC
GATACTA
TCCTAAC
ATGAATG
10^-2
10^-3
10^-4
(Conc)(mM)
归一化的高于背景的洗涤强度

图12

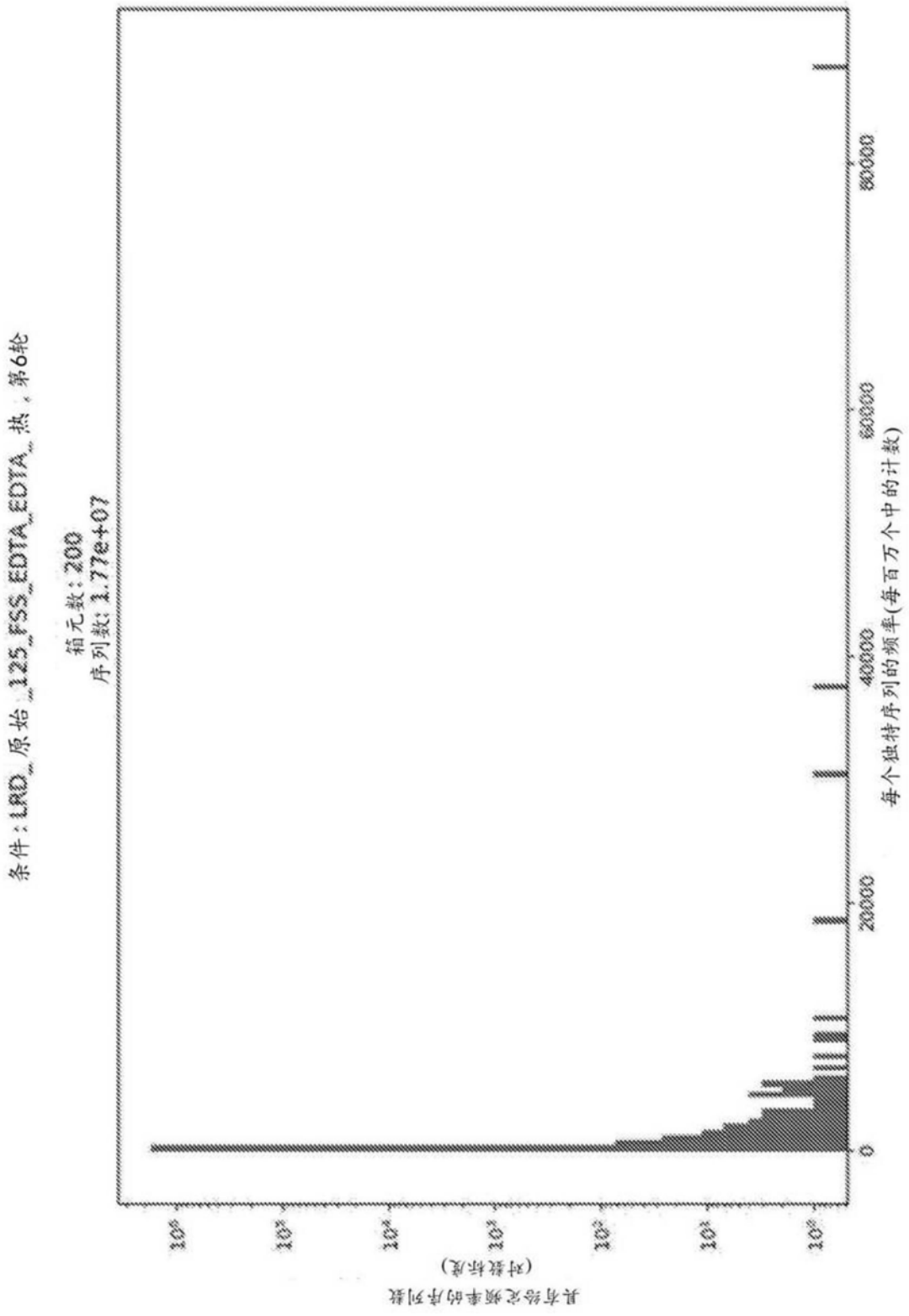

条件：LRD_原始_125_FSS_EDTA_EDTA_热，第6轮
箱元数：200
序列数：1.77e+07
具有给定频率的序列数
(对数标度)
每个独特序列的频率(每百万个中的计数)
0
200000
400000
600000
800000

图13A

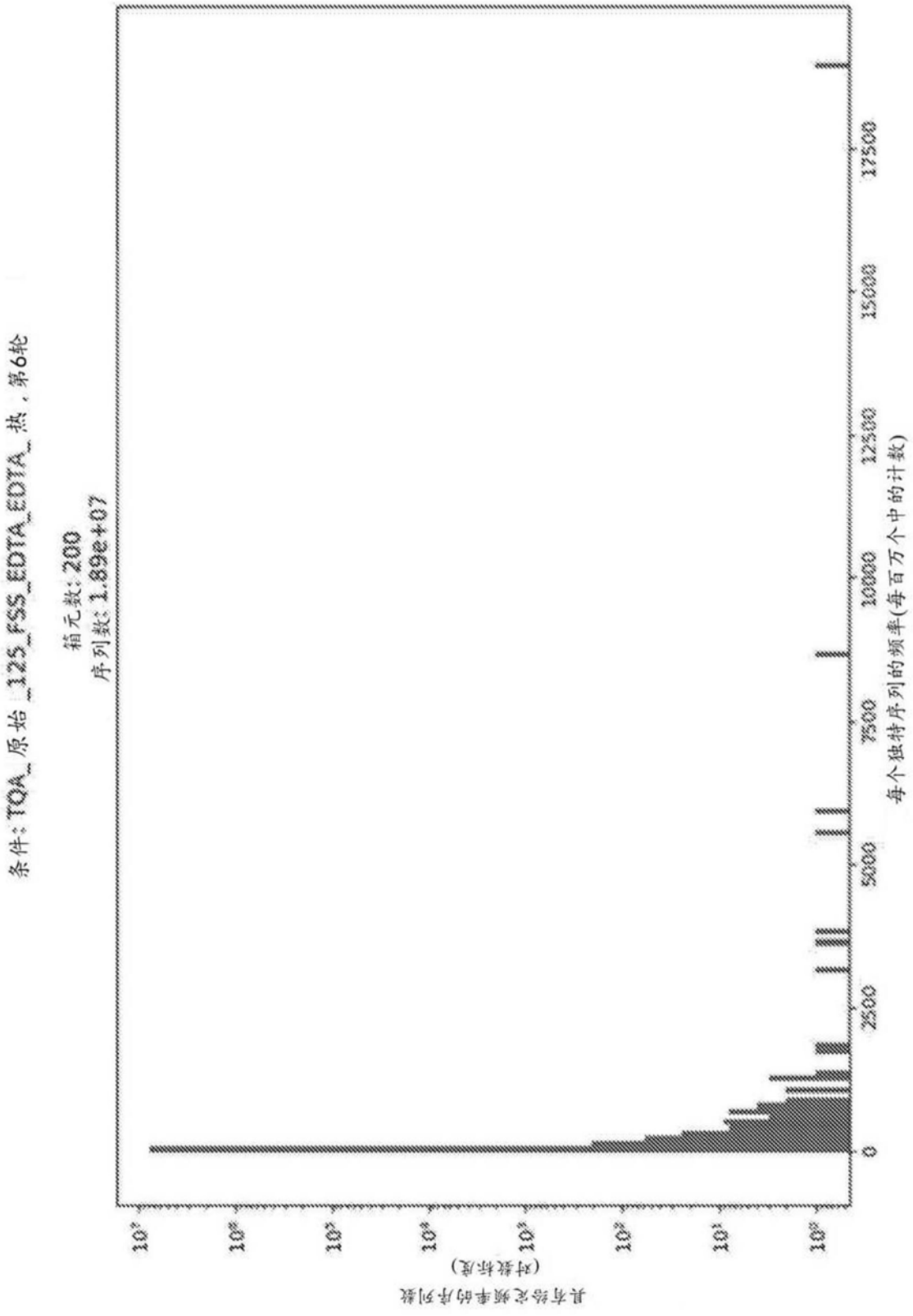

条件: TQA_原始_125_FSS_EDTA_EDTA_热，第6轮
箱元数: 200
序列数: 1.89e+07
具有给定频率的序列数
(对数标度)
每个独特序列的频率(每百万个中的计数)
0
2500
5000
7500
10000
12500
15000
17500

图13B

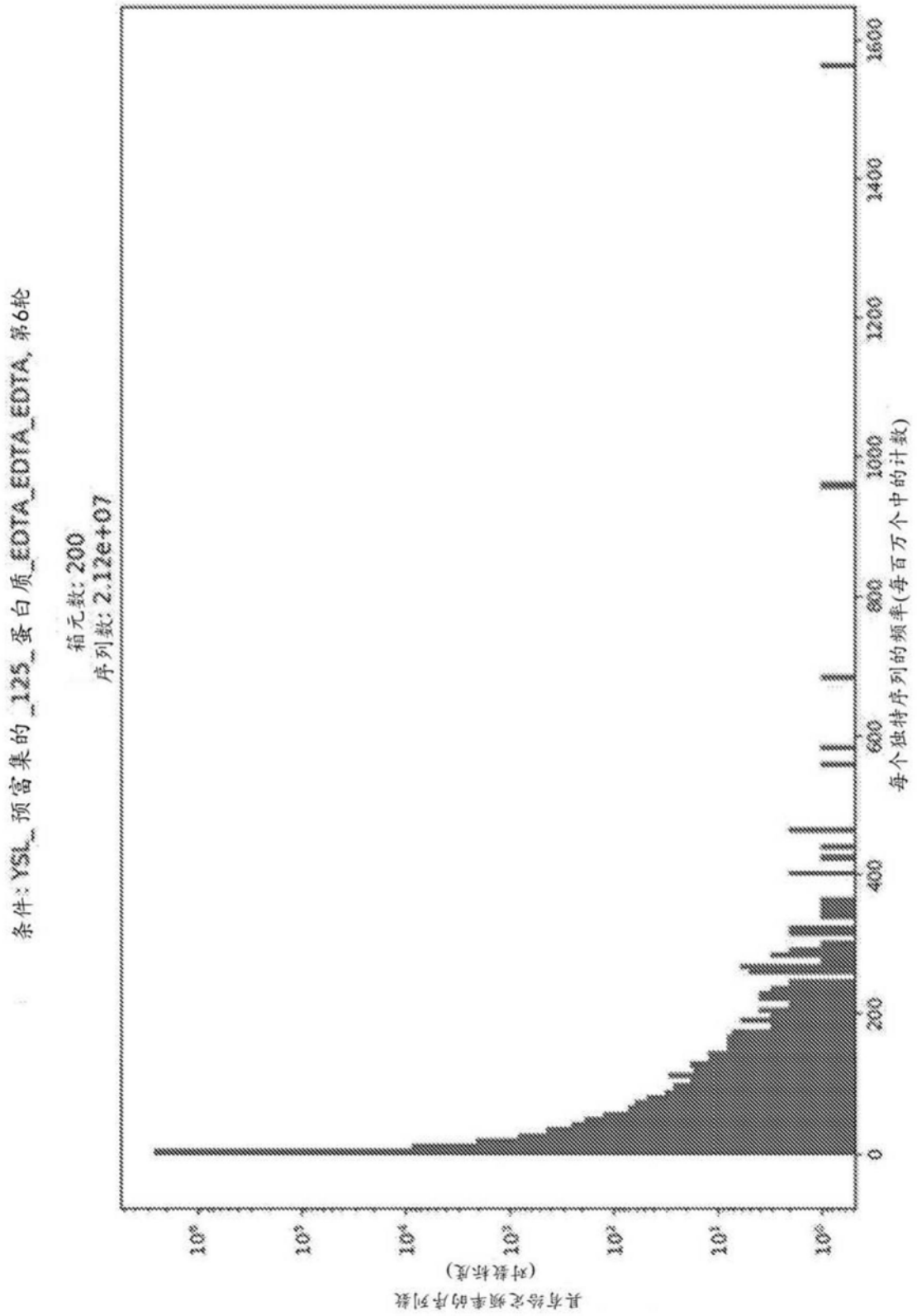

条件: YSL_预富集的_125_蛋白质_EDTA_EDTA_EDTA, 第6轮
箱元数: 200
序列数: 2.12e+07
具有给定频率的序列数
(对数标度)
每个独特序列的频率(每百万个中的计数)

图14A

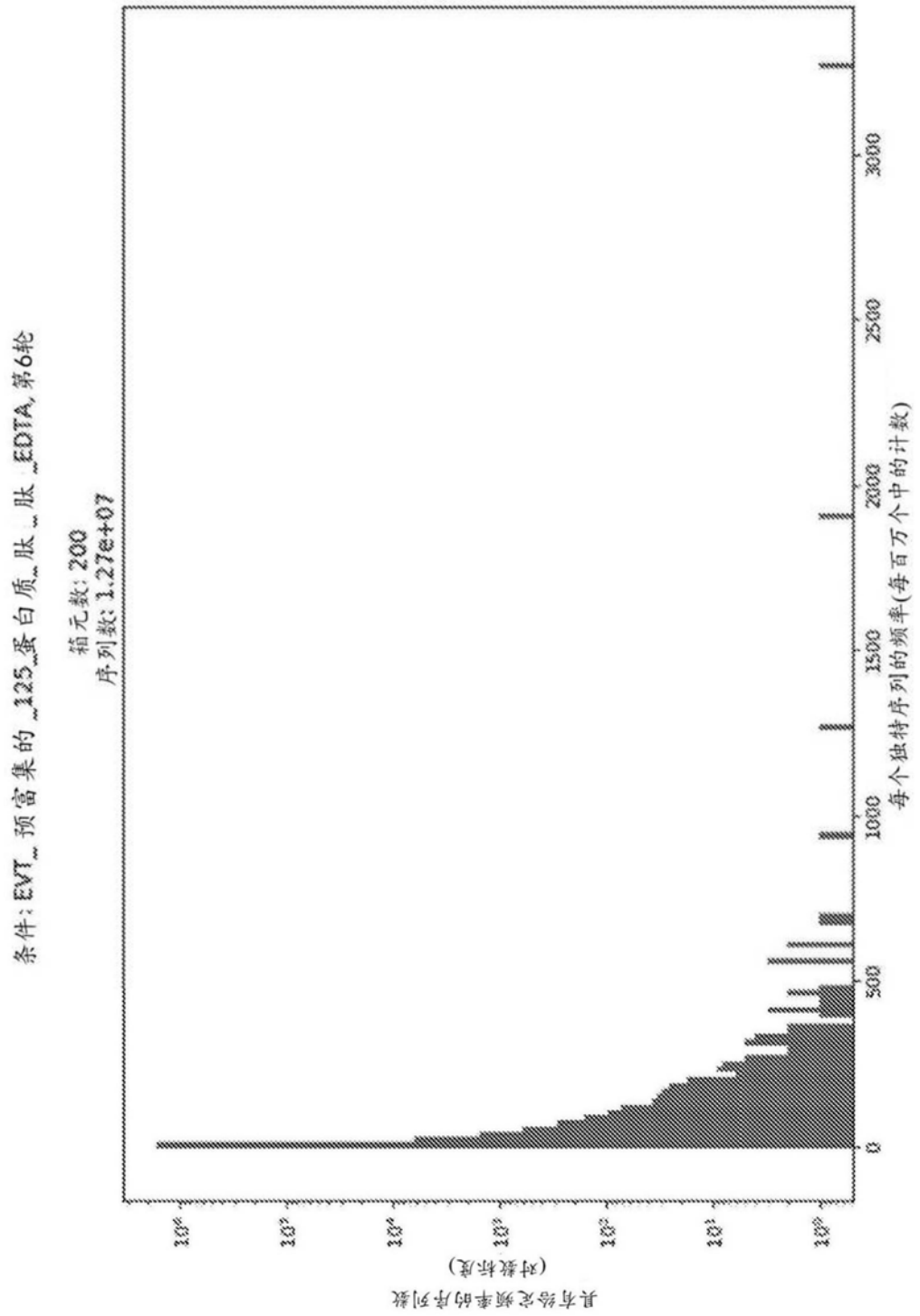

条件:EVT_预富集的_125_蛋白质_肽_肽_EDTA,第6轮
箱元数:200
序列数:1.27e+07
具有给定频率的序列数
(对数标度)
每个独特序列的频率(每百万个中的计数)
0
500
1000
1500
2000
2500
3000

图14B

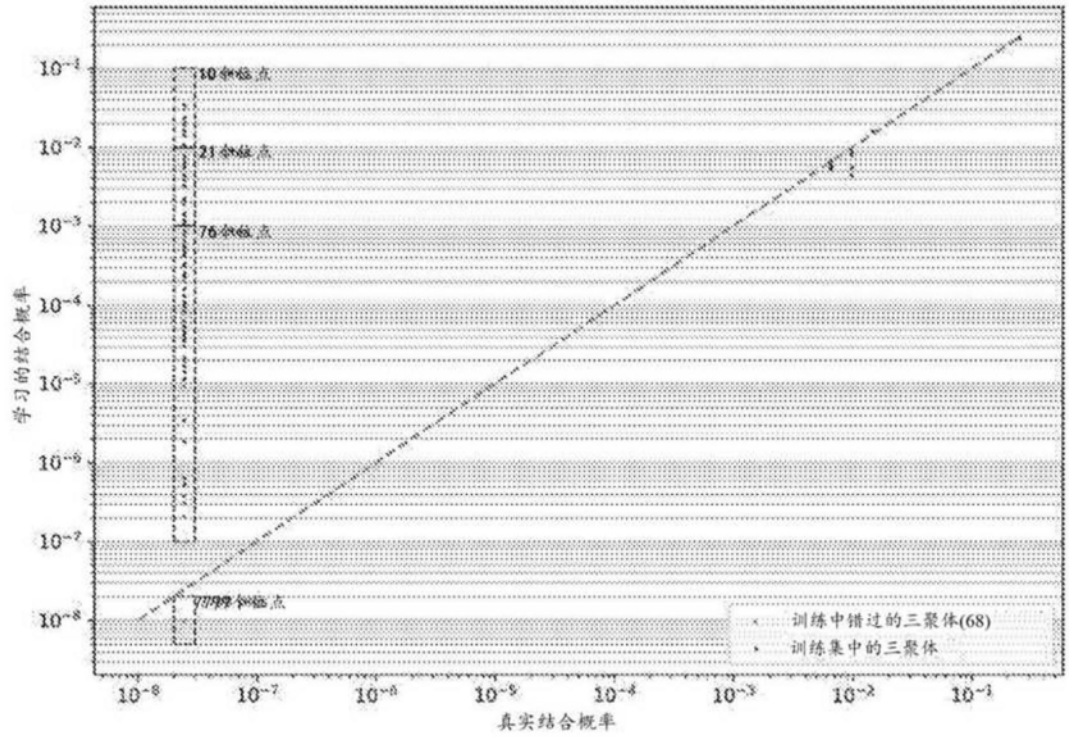

10 个结点
21 个结点
76 个结点
7799 个结点
训练中错过的三聚体(68)
训练集中的三聚体
学习的结合概率
真实结合概率

图15A

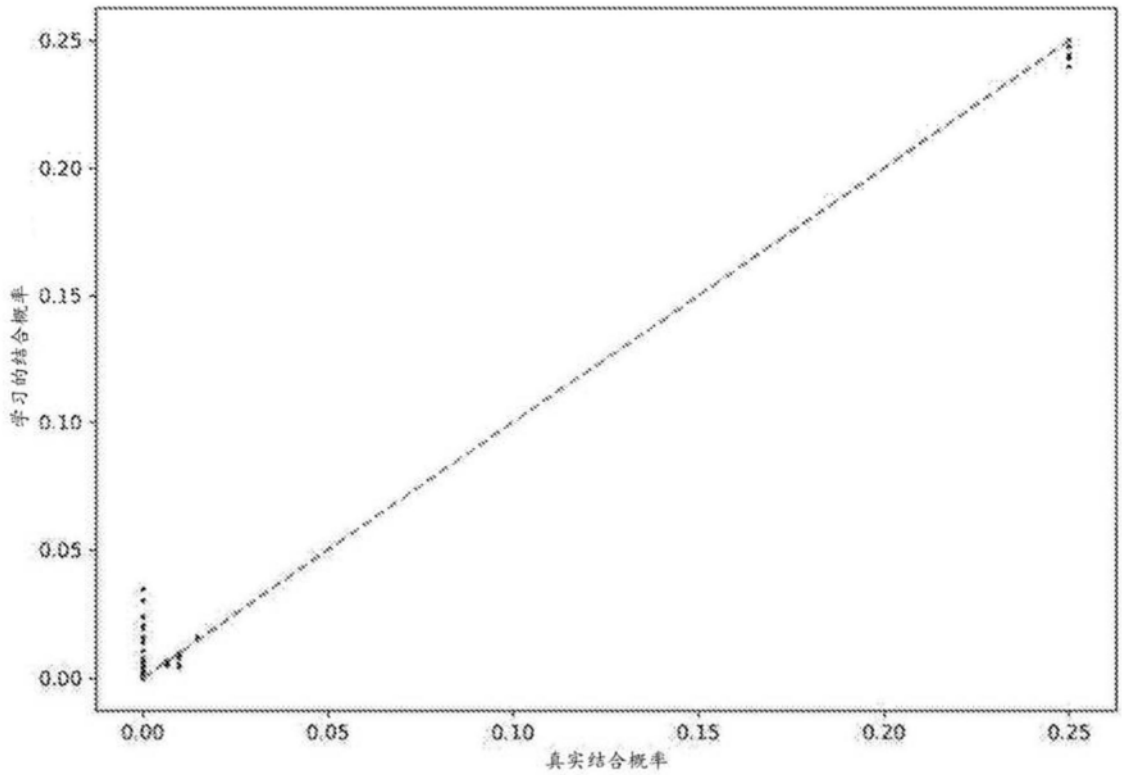

0.25
0.20
0.15
0.10
0.05
0.00
学习的结合概率
真实结合概率

图15B

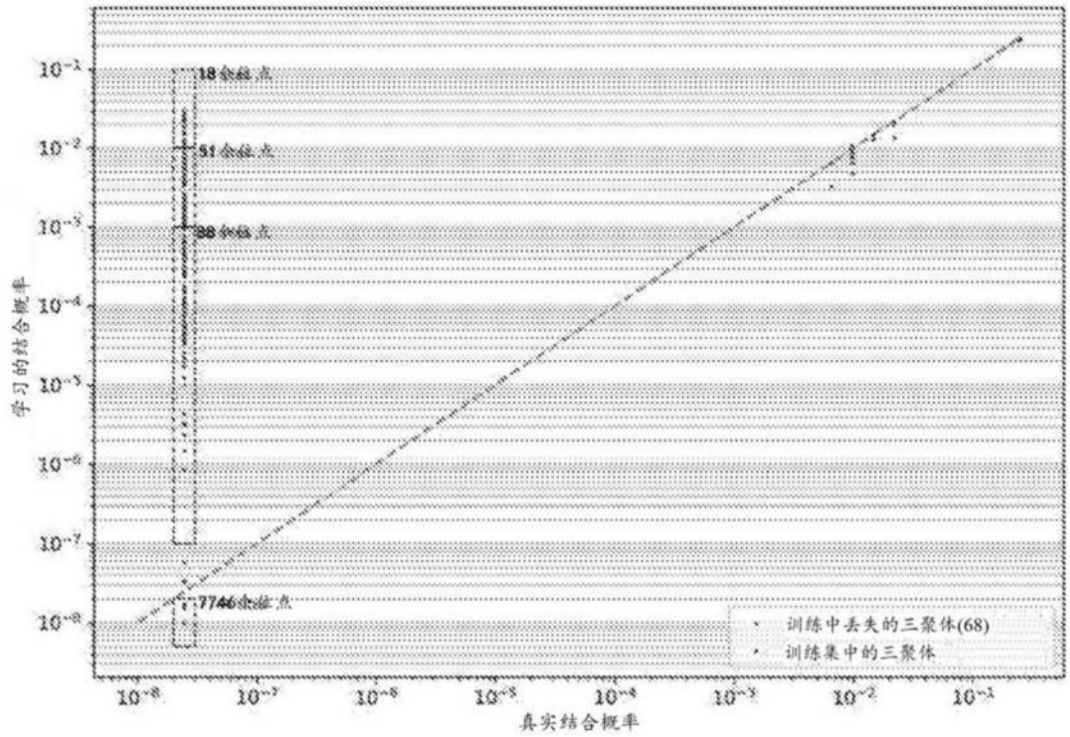

学习的结合概率
真实结合概率
训练中丢失的三聚体(68)
训练集中的三聚体

图16A

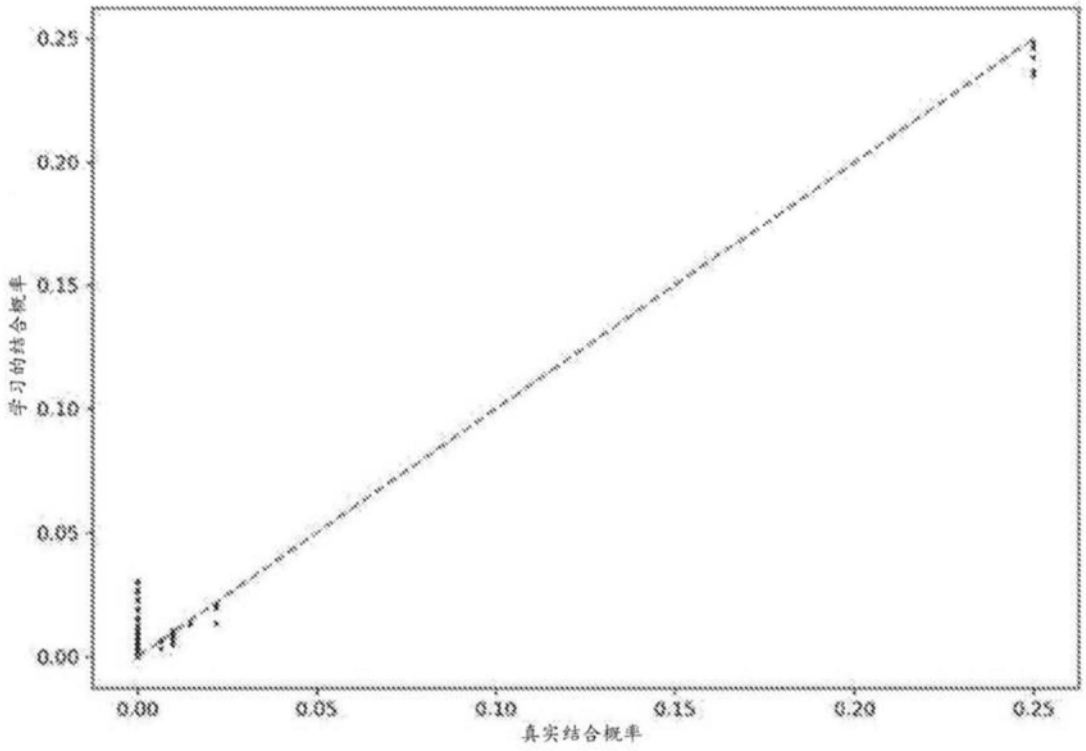

学习的结合概率
真实结合概率
0.00
0.05
0.10
0.15
0.20
0.25

图16B

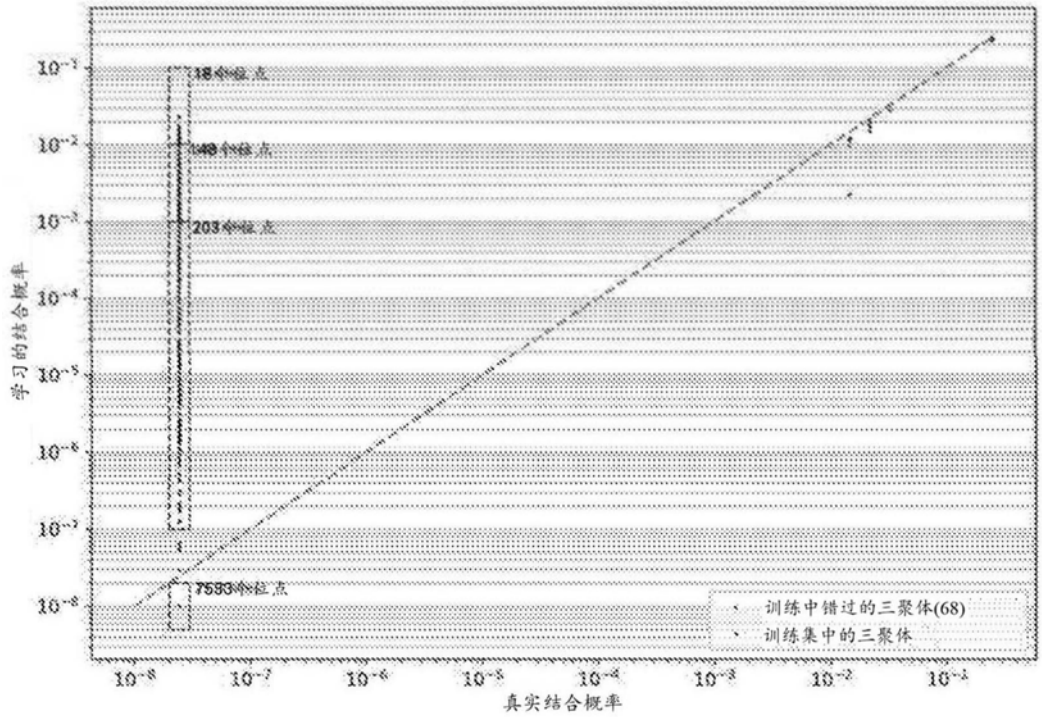

学习的结合概率
真实结合概率
训练中错过的三聚体(68)
训练集中的三聚体

图17A

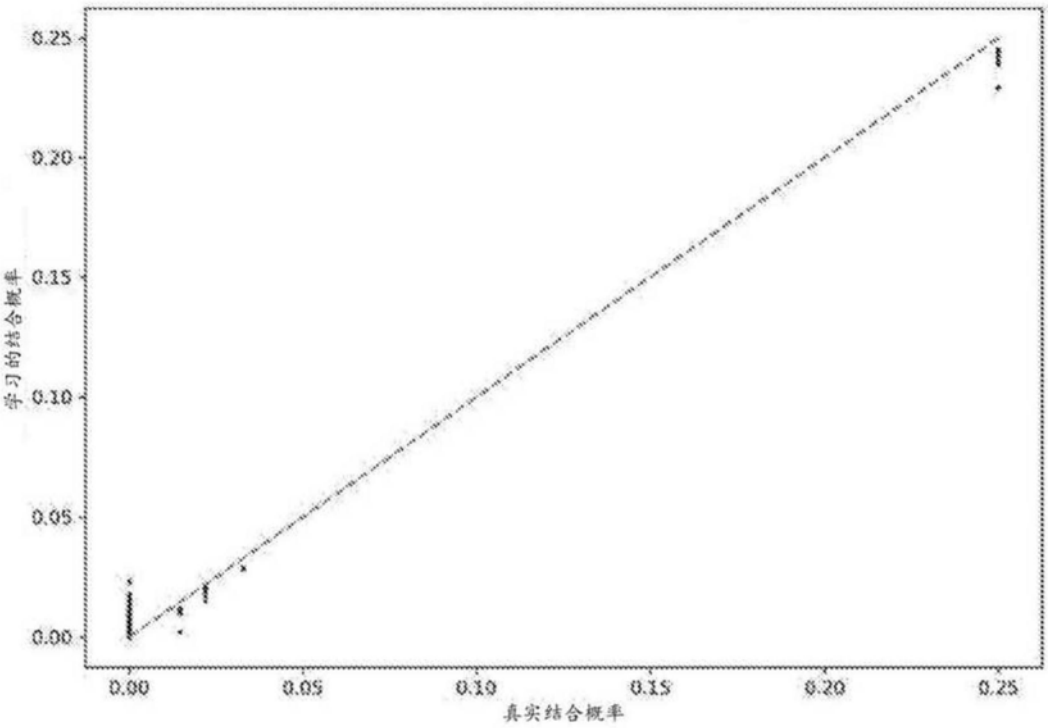

学习的结合概率
真实结合概率
0.00
0.05
0.10
0.15
0.20
0.25

图17B