



(12) 发明专利申请

(10) 申请公布号 CN 115094523 A

(43) 申请公布日 2022. 09. 23

(21) 申请号 202210757923.6

(22) 申请日 2017.03.31

(30) 优先权数据

62/317,353 2016.04.01 US

62/472,504 2017.03.16 US

(62) 分案原申请数据

201780034482.4 2017.03.31

(71) 申请人 库博科学公司

地址 美国亚利桑那州

(72) 发明人 马修·格雷宁 大卫·史密斯

高拉夫·萨伊尼

(74) 专利代理机构 北京安信方达知识产权代理

有限公司 11262

专利代理师 刘晓杰 武晶晶

(51) Int.Cl.

C40B 50/14 (2006.01)

C40B 50/18 (2006.01)

C40B 40/10 (2006.01)

G01N 33/68 (2006.01)

G01N 33/543 (2006.01)

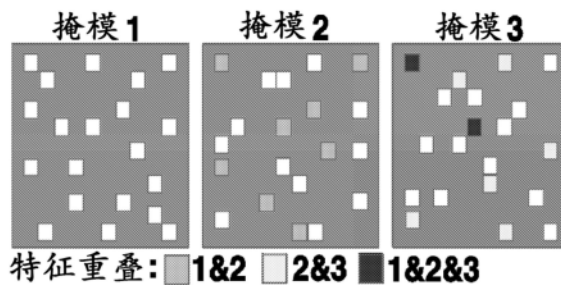
权利要求书3页 说明书73页 附图25页

(54) 发明名称

用于治疗性抗体表征的基于阵列的肽文库

(57) 摘要

本文提供了用于进行原位图案化化学的方法、化学文库和模拟系统。还提供了用于表征抗体-靶标相互作用的包括使用合成化学文库的方法、系统和测定,该方法、系统和测定以基于知识的方式增加探索的蛋白质空间,该方法、系统和测定包括:鉴定抗体的靶蛋白、表征靶蛋白中的抗体结合区、鉴定靶蛋白中的线性表位和结构表位以及确定抗体与靶蛋白结合的倾向。



1. 一种在基底上原位合成化学文库的方法,所述化学文库包含多个分子,所述方法包括:

(a) 接收生物序列和多个合成步骤;

(b) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;

(c) 为每个图案化掩模指定至少一个单体;以及

(d) 将所述单体耦合至所述特征以形成分子;

其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤。

2. 一种原位合成的化学文库,所述化学文库包含多个分子,其中所述合成使用图案化步骤以在基底上构建所述文库,所述合成包括:

(a) 接收生物序列和多个合成步骤;

(b) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;

(c) 为每个图案化掩模指定至少一个单体;以及

(d) 将所述单体耦合至所述特征以形成分子;

其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤。

3. 一种用于模拟在基底上原位合成化学文库的计算系统,所述化学文库包含多个分子,所述计算系统包含:

(a) 处理器和存储器;

(b) 计算机程序,其包括可由所述处理器执行的指令,所述计算机程序包含:

(1) 接收模块,其被配置为接收生物序列和多个合成步骤;

(2) 模拟模块,其被配置为:(i) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;(ii) 为每个图案化掩模指定至少一个单体;以及(iii) 将所述单体耦合至所述特征以形成分子;

其中(i)、(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤。

4. 一种原位合成肽阵列的方法,所述方法包括:

(a) 接收输入氨基酸序列;

(b) 确定多个合成步骤;

(c) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

(d) 为每个图案化掩模指定至少一个单体;以及

(e) 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

5. 一种阵列,其包含所述阵列上的多个原位合成的肽,所述肽由多个图案化掩模产生,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个

顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠。

6. 一种用于表征针对至少一种蛋白质靶标的抗体结合的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(a)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(c) 使用步骤(b)的所述比对得分表征针对所述至少一种蛋白质靶标的所述抗体的结合。

7. 一种用于鉴定靶蛋白中抗体表位的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(a)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(c) 确定步骤(a)的所述单独的肽中的保守氨基酸,以鉴定保守结合肽基序,并将所述单独的基序与所述至少一种靶蛋白比对,以便鉴定所述靶蛋白的至少一种抗体表位。

8. 一种用于表征靶蛋白中的抗体结合区的方法,所述方法包括:

(a) 在存在和不存在多种竞争肽的情况下,使第一肽阵列与所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第一预定阈值内。

(b) 使用选自步骤(a)中至少一种所述单独的肽的输入肽序列、衍生自步骤(a)中所述单独的肽的比对的保守基序或衍生自步骤(a)中所述单独的肽的比对的比对基序来产生第二肽阵列,所述第二肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列;

(c) 使所述第二肽阵列与所述抗体接触以鉴定第二组肽;以及

(d) 在存在多种竞争肽的情况下使所述第二肽阵列与所述抗体接触,并鉴定来自步骤(c)中的第二组单独的肽,所述第二组单独的肽展现出处于步骤(c)中所述结合信号的第二预定阈值内的结合信号;以及

(e) 将所述第二组单独的肽与所述靶蛋白比对,并鉴定所述靶蛋白中与所鉴定的第二组单独的肽比对的区域,从而表征所述靶蛋白中的抗体结合区。

9. 一种用于鉴定抗体的靶蛋白的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使第一肽阵列与一种或多种浓度的所述抗体接触以获得一个或多个输入氨基酸序列,其中所鉴定的输入氨基酸序列展现出在存在所述多种竞争肽的情况下的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下的所述结合信号的第一预定阈值内。

(b) 使用选自步骤(a)中至少一种所述单独的肽的一个或多个输入氨基酸序列、衍生自步骤(a)中所述单独的肽的比对的保守基序或衍生自步骤(a)中所述单独的肽的比对的比对基序来获得一种或多种二级肽阵列,所述一种或多种二级肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

(c) 在存在和不存在所述多种竞争肽的情况下,使每个所述二级肽阵列与所述抗体接触以获得一组肽序列,其中所鉴定的一组肽序列展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第二预定阈值内。

(d) 使所述一组肽序列彼此比对以获得至少一种预测性结合基序;以及

(e) 将所述预测性结合基序作为针对蛋白质数据库的搜索标准进行比对,从而基于所述蛋白质数据库搜索结果得分来鉴定所述抗体的靶蛋白。

10. 一种用于确定抗体与至少一种蛋白质靶标结合倾向的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 使步骤(a)的所述单独的肽与第一蛋白质靶标比对,其中步骤(a)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(c) 重复步骤(a)的所述单独的肽与至少一种另外的蛋白质靶标的比对,其中步骤(a)的所述单独的肽与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(d) 比较步骤(b)和(c)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

## 用于治疗性抗体表征的基于阵列的肽文库

本申请是申请日为2017年03月31日、申请号为201780034482.4、发明名称为“用于治疗性抗体表征的基于阵列的肽文库”的中国专利申请(其对应PCT申请的申请日为2017年03月31日、申请号为PCT/US2017/025546)的分案申请。

### 相关申请的交叉引用

[0001] 本申请要求于2016年4月1日提交的美国临时申请号62/317,353的权益以及2017年3月16日提交的美国临时申请号62/472,504的权益,这两篇美国临时申请均通过引用以其全文并入本文。

### 背景技术

[0002] 癌症是美国第二大常见的死亡原因,其中在美国每天有超过1,600例癌症相关死亡,每年近600,000例癌症相关死亡。2015年诊断出大约165万新发癌症病例,并且由于人口和生活方式的因素,癌症发病率逐渐增加。需要用于检测和治疗癌症的敏感且有效的方法。

### 发明内容

[0003] 随着近来诊断和治疗的改进,癌症死亡人数一直呈下降趋势,并且随着持续监测和后续治疗,癌症正在向慢性疾病发展。虽然癌症患者死亡的比例正在下降,但是由于突破性治疗和包括癌症复发和其他治疗方法的长期慢性病护理的高成本,癌症治疗的经济负担正在迅速增加。癌症治疗费用的快速增长处于不可持续的发展轨迹上,并且按照目前的速度,到2028年,患者的自付费用将是家庭收入中位数的100%。由于成本上升,特别是癌症免疫疗法和抗体治疗的成本,患者需要在治疗和经济稳定性之间做出艰难的选择。

[0004] 免疫疗法和基于抗体的癌症治疗已成为延长患者存活率的两个主要治疗突破。免疫疗法激活并利用患者的免疫系统来杀死癌细胞,而基于抗体的治疗靶向抑制或杀死癌细胞的特定途径。这些方法中的每一种都严重或完全地依赖于高度靶标特异性抗体或生物制剂的发现和开发,并且最近还依赖于多靶标特异性抗体或具有多价结合的生物制剂的发现和开发。即使由免疫疗法和基于抗体的治疗提供的患者存活率显著提高,但仍然存在特定的主要挑战。首先,由于重大的脱靶副作用具有高发生率,因此免疫治疗和基于抗体的治疗仅限于有限的能够具适宜的治疗反应的患者群体。例如,两种最常用的抗体治疗药物Humira和Remicade仅在25%的患者群体中有效。其次,高发现和开发成本是限制免疫疗法和基于抗体的研发项目数目以及市场上竞争者数目的进入壁垒。在很大一部分患者中脱靶效应的高发生率和高研发成本二者导致免疫疗法和基于抗体的治疗具有非常高的价格,这在许多情况下对于患者来说极其昂贵而无法承担。

[0005] 由于研发成本不断上升,目前制药研发的主要威胁之一是生产力下降。缓解这种生产力的下降将需要创新以降低成本、增加正在进行的候选分子的数目以及减少研发周期时间。为了降低与免疫疗法和基于抗体的治疗相关的高研发成本和脱靶风险,需要创新平台来实现从发现过程早期到晚期临床前开发的治疗性先导抗体的全面筛选和表征。此外,新的低成本和高通量抗体表征平台将允许更多候选者进入发现管线,并使更多公司能够进

入免疫治疗发现计划,这将增加创新、竞争和市场潜力。

[0006] 免疫疗法是癌症治疗方面的突破,并且是发展最快的药物市场领域之一。抗体文库筛选和靶/脱靶结合表征是免疫疗法开发中的基本活动。目前,在常规筛选针对治疗靶标的大抗体文库的能力与表征筛选出的治疗性抗体候选物的靶/脱靶结合的有限能力之间存在很大差距。随着候选物的数目远远大于单特异性抗体的多特异性治疗性抗体和生物制剂的出现,这种差距正在扩大。治疗性抗体靶/脱靶结合表征的主要限制是,相对于总的可能的表位(例如,10-聚体肽表位意味着10万亿个可能的序列)而言,可以进行分析的表位相互作用的部分极小。目前包括微阵列、表面等离子体共振(SPR)和干涉测量的抗体表征平台具有10,000-50,000个表位相互作用测量的实际限制。这种有限的治疗性抗体结合谱增加了未检测到的脱靶效应的风险。

[0007] 本文公开了新的平台以显著增加检测到的治疗性抗体相互作用的数目,这可以降低未检测到的脱靶效应的风险。这些技术基于合并的肽合成化学与半导体制造工艺,利用基于掩模的光刻法在8英寸晶片上原位图案化含有超过4000万个肽(潜在表位)的文库。将该晶片切割成13个显微镜载玻片尺寸的芯片以用于下游分析。利用本文所述的这类肽文库芯片,抗体结合谱测定可以按当前抗体表征平台的一小部分成本按比例缩放至每天超过1000万个抗体-靶标相互作用。抗体表位点变异分析证明了肽芯片对抗体表征的适用性。

[0008] 在一个方面,本文公开了在基底上原位合成化学文库的方法,所述化学文库包含多个分子,所述方法包括:(a)接收生物序列和多个合成步骤;(b)确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;(c)为每个图案化掩模指定至少一个单体;以及(d)将所述单体耦合至所述特征以形成分子;其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤。在一些实施方案中,所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。在一些实施方案中,所述输入生物序列包含疾病相关的表位。在一些实施方案中,所述输入生物序列包含肽序列。在一些实施方案中,所述输入生物序列包含表位序列。在一些实施方案中,所述输入生物序列包含随机序列。在一些实施方案中,所述方法包括从所述输入生物序列中导出单体的有序列表。在另外的实施方案中,所述有序列表的大小为所述合成步骤的数目。在一些实施方案中,所述单体的有序列表包含所述输入生物序列。在一些实施方案中,所述单体的有序列表以相反的顺序包含所述输入生物序列。在一些实施方案中,分子为肽或核酸。在一些实施方案中,所述单体的有序列表包含氨基酸序列。在一些实施方案中,所述单体的有序列表包含核苷酸序列。在一些实施方案中,所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。在一些实施方案中,所述多个图案化掩模的数目为所述合成步骤的数目。在一些实施方案中,每个顺序图案化掩模中约20%至约50%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。在一些实施方案中,每个顺序图案化掩模中约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。在一些实施方案中,所述合成步骤基于光刻法。在一些实施方案中,所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。在一些实施方案中,所述文库中至少40%的分子是不同的。在一些实施方案中,所

述文库中至少50%的分子是不同的。在一些实施方案中,所述文库中至少60%的分子是不同的。在一些实施方案中,所述文库中至少70%的分子是不同的。在一些实施方案中,所述文库中至少80%的分子是不同的。在一些实施方案中,所述文库中至少90%的分子是不同的。在一些实施方案中,所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。在一些实施方案中,所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。在一些实施方案中,所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。在一些实施方案中,所述文库包含等于所述生物序列长度的中值单体长度。在一些实施方案中,所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。在一些实施方案中,所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值单体长度。在一些实施方案中,所述基底选自阵列、晶片、载玻片和珠子。在一些实施方案中,所述合成的化学文库包含肽、核苷酸或其组合。在一些实施方案中,所述肽的长度为约5至约25个氨基酸。在一些实施方案中,所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。在一些实施方案中,利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。在一些实施方案中,所述表面间隔物为Cys-Gly-Pro-Gly-Xaan-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaan-(PEG3)-Cys。在一些实施方案中,利用能够用酯键环化的表面间隔物来合成所述化学文库。在一些实施方案中,所述酯键为同双官能的二-NHS酯键。在一些实施方案中,所述表面间隔物为Lys-(PEG3)-Xaan-(PEG3)-赖氨酸。在一些实施方案中,所述基底涂覆有亲水性单层。在一些实施方案中,所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。在一些实施方案中,所述亲水单层是均匀的。

[0009] 在另一方面,本文公开了原位合成的化学文库,其中所述合成使用图案化步骤以在基底上构建所述文库,所述化学文库包含多个分子,所述合成包括:(a)接收生物序列和多个合成步骤;(b)确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;(c)为每个图案化掩模指定至少一个单体;以及(d)将所述单体耦合至所述特征以形成分子;其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤。在一些实施方案中,所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。在一些实施方案中,所述输入生物序列包含疾病相关的表位。在一些实施方案中,所述输入生物序列包含肽序列。在一些实施方案中,所述输入生物序列包含表位序列。在一些实施方案中,所述输入生物序列包含随机序列。在一些实施方案中,所述方法包括从所述输入生物序列中导出单体的有序列表。在另外的实施方案中,所述有序列表的大小为所述合成步骤的数目。在一些实施方案中,所述单体的有序列表包含所述输入生物序列。在一些实施方案中,所述单体的有序列表以相反的顺序包含所述输入生物序列。在一些实施方案中,分子为肽或核酸。在一些实施方案中,所述单体的有序列表包含氨基酸序列。在一些实施方案中,所述单体的有序列表包含核苷酸序列。在一些实施方案

中,所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。在一些实施方案中,所述多个图案化掩模的数目为所述合成步骤的数目。在一些实施方案中,每个顺序图案化掩模中约20%至约50%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。在一些实施方案中,每个顺序图案化掩模中约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。在一些实施方案中,所述合成步骤基于光刻法。在一些实施方案中,所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。在一些实施方案中,所述文库中至少40%的分子是不同的。在一些实施方案中,所述文库中至少50%的分子是不同的。在一些实施方案中,所述文库中至少60%的分子是不同的。在一些实施方案中,所述文库中至少70%的分子是不同的。在一些实施方案中,所述文库中至少80%的分子是不同的。在一些实施方案中,所述文库中至少90%的分子是不同的。在一些实施方案中,所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。在一些实施方案中,所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。在一些实施方案中,所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。在一些实施方案中,所述文库包含等于所述生物序列长度的中值单体长度。在一些实施方案中,所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。在一些实施方案中,所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值单体长度。在一些实施方案中,所述基底选自阵列、晶片、载玻片和珠子。在一些实施方案中,所述合成的化学文库包含肽、核苷酸或其组合。在一些实施方案中,所述肽的长度为约5至约25个氨基酸。在一些实施方案中,所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。在一些实施方案中,利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。在一些实施方案中,所述表面间隔物为Cys-Gly-Pro-Gly-Xaan-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaan-(PEG3)-Cys。在一些实施方案中,利用能够用酯键环化的表面间隔物来合成所述化学文库。在一些实施方案中,所述酯键为同双官能的二-NHS酯键。在一些实施方案中,所述表面间隔物为Lys-(PEG3)-Xaan-(PEG3)-赖氨酸。在一些实施方案中,所述基底涂覆有亲水性单层。在一些实施方案中,所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。在一些实施方案中,所述亲水性单层是均匀的。

[0010] 在另一方面,本文公开了用于模拟在基底上原位合成化学文库的计算系统,所述化学文库包含多个分子,所述计算系统包含:(a)处理器和存储器;(b)计算机程序,其包括可由所述处理器执行的指令,所述计算机程序包含:(1)接收模块,其被配置为接收生物序列和多个合成步骤;(2)模拟模块,所述模拟模块被配置为:(i)确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;(ii)为每个图案化掩模指定至少一个单体;以及(iii)将所述单体耦合至所述特征以形成分子;其中(i)、(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤。在一些实施方案中,所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、

90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。在一些实施方案中,所述输入生物序列包含疾病相关的表位。在一些实施方案中,所述输入生物序列包含肽序列。在一些实施方案中,所述输入生物序列包含表位序列。在一些实施方案中,所述输入生物序列包含随机序列。在一些实施方案中,所述模拟模块包括从所述输入生物序列中导出单体的有序列表。在另外的实施方案中,所述有序列表的大小为所述合成步骤的数目。在一些实施方案中,所述单体的有序列表包含所述输入生物序列。在一些实施方案中,所述单体的有序列表以相反的顺序包含所述输入生物序列。在一些实施方案中,分子为肽或核酸。在一些实施方案中,所述单体的有序列表包含氨基酸序列。在一些实施方案中,所述单体的有序列表包含核苷酸序列。在一些实施方案中,所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。在一些实施方案中,所述多个图案化掩模的数目为所述合成步骤的数目。在一些实施方案中,每个顺序图案化掩模中约20%至约50%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。在一些实施方案中,每个顺序图案化掩模中约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。在一些实施方案中,所述合成步骤基于光刻法。在一些实施方案中,所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。在一些实施方案中,所述文库中至少40%的分子是不同的。在一些实施方案中,所述文库中至少50%的分子是不同的。在一些实施方案中,所述文库中至少60%的分子是不同的。在一些实施方案中,所述文库中至少70%的分子是不同的。在一些实施方案中,所述文库中至少80%的分子是不同的。在一些实施方案中,所述文库中至少90%的分子是不同的。在一些实施方案中,所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。在一些实施方案中,所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。在一些实施方案中,所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。在一些实施方案中,所述文库包含等于所述生物序列长度的中值单体长度。在一些实施方案中,所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。在一些实施方案中,所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值单体长度。在一些实施方案中,所述基底选自阵列、晶片、载玻片和珠子。在一些实施方案中,所述合成的化学文库包含肽、核苷酸或其组合。在一些实施方案中,所述肽的长度为约5至约25个氨基酸。在一些实施方案中,所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。在一些实施方案中,利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。在一些实施方案中,所述表面间隔物为Cys-Gly-Pro-Gly-Xaan-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaan-(PEG3)-Cys。在一些实施方案中,利用能够用酯键环化的表面间隔物来合成所述化学文库。在一些实施方案中,所述酯键为同双官能的二-NHS酯键。在一些实施方案中,所述表面间隔物为Lys-(PEG3)-Xaan-(PEG3)-赖氨酸。在一些实施方案中,所述基底涂覆有亲水性单层。在一些实施方案中,所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。在一些实施方案中,所述亲水性单层是均匀的。

[0011] 还包括用于表征针对至少一种蛋白质靶标的抗体结合的方法和测定,所述方法包括:(a)在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的所述抗体接触以鉴别一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内;(b)将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(a)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及(c)使用步骤(b)的所述比对得分表征针对所述至少一种蛋白质靶标的所述抗体的结合。

[0012] 本文还公开了用于鉴别靶蛋白中的抗体表位的方法和测定,所述方法包括:(a)在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内;(b)将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(a)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及(c)确定步骤(a)的所述单独的肽中的保守氨基酸,以鉴别保守结合肽基序,并将所述单独的基序与所述至少一种靶蛋白比对,以便鉴定所述靶蛋白的至少一种抗体表位。

[0013] 本文公开了用于表征靶蛋白中的抗体结合区的方法和测定,所述方法包括:(a)在存在和不存在多种竞争肽的情况下,使第一肽阵列与所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第一预定阈值内;(b)使用选自步骤(a)中至少一种所述单独的肽的输入肽序列、衍生自步骤(a)中所述单独的肽的比对的保守基序或衍生自步骤(a)中所述单独的肽的比对的比对基序来产生第二肽阵列,所述第二肽阵列通过以下方法合成:i.确定多个合成步骤;ii.确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;iii.为每个图案化掩模指定至少一个单体;以及iv.将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列;(c)使所述第二肽阵列与所述抗体接触以鉴定第二组肽;(d)在存在多种竞争肽的情况下使所述第二肽阵列与所述抗体接触,并鉴定来自步骤(c)的第二组单独的肽,所述第二组单独的肽展现出处于步骤(c)中所述结合信号的第二预定阈值内的结合信号;以及(e)将所述第二组单独的肽与所述靶蛋白比对,并鉴定所述靶蛋白与所鉴定的第二组单独的肽比对的区域,从而表征所述靶蛋白中的抗体结合区。

[0014] 本文还包括用于鉴定抗体的靶蛋白的方法和测定,所述方法包括:(a)在存在和不存在一种或多种浓度的多种竞争肽的情况下,使第一肽阵列与一种或多种浓度的所述抗体接触以获得一个或多个输入氨基酸序列,其中所鉴定的输入氨基酸序列展现出在存在所述多种竞争肽的情况下的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下的所述结合信号的第一预定阈值内;(b)使用选自步骤(a)中至少一种所述单独的肽的一个或多个输入氨基酸序列、衍生自步骤(a)中所述单独的肽的比对的保守基序或衍生自步骤(a)中所述单独的肽的比对的比对基序来获得一种或多种二级肽阵列,所述一种或多种二级

肽阵列通过以下方法合成：(i) 确定多个合成步骤；(ii) 确定多个图案化掩模，其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定，并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠；(iii) 为每个图案化掩模指定至少一个单体；以及(iv) 将所述单体耦合至所述特征，其中(iii)和(iv)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列；(c) 在存在和不存在所述多种竞争肽的情况下，使每个所述二级肽阵列与所述抗体接触以获得一组肽序列，其中所鉴定的一组肽序列展现出在存在所述多种竞争肽的情况下测量的结合信号，所述结合信号处于在不存在所述多种竞争肽时测量的情况下所述结合信号的第二预定阈值内；(d) 使所述一组肽序列彼此比对以获得至少一种预测性结合基序；以及(e) 将所述预测性结合基序作为针对蛋白质数据库的搜索标准进行比对，从而基于所述蛋白质数据库搜索结果得分来鉴定所述抗体的靶蛋白。

[0015] 本文还包括用于确定抗体与至少一种蛋白质靶标结合倾向的方法，所述方法包括：(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下，使肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽，其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号，所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内；(b) 使步骤(a)的所述单独的肽与第一蛋白质靶标比对，其中步骤(a)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分；(c) 重复步骤(a)的所述单独的肽与至少一种另外的蛋白质靶标的比对，其中步骤(a)的所述单独肽与所述另外的蛋白质靶标之间的所述比对被指定为比对得分；以及(d) 比较步骤(b)和(c)的所述比对得分，以获得所述抗体与所述蛋白质靶标结合的相对倾向。

[0016] 本文公开了用于确定抗体与至少一种蛋白质靶标结合倾向的方法和测定，所述方法包括：(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下，使第一肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽，其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号，所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内；(b) 对步骤(a)的所述一种或多种单独的肽进行比对以获得至少一种预测性靶基序；(c) 将所述至少一种预测性靶基序与第一蛋白质靶标比对，其中步骤(c)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分；(d) 重复步骤(b)的至少一种预测性靶基序与至少一种另外的蛋白质靶标的比对，其中步骤(b)的所述至少一种预测性靶基序与所述另外的蛋白质靶标之间的所述比对被指定为比对得分；以及(e) 比较步骤(c)和(d)的所述比对得分，以获得所述抗体与所述蛋白质靶标结合的相对倾向。

[0017] 本文还公开了用于表征针对至少一种蛋白质靶标的抗体结合的试剂盒和系统，所述试剂盒和系统包括：(a) 提供肽阵列；(b) 提供多种竞争肽；(c) 为用户提供说明以在存在和不存在一种或多种浓度的多种竞争肽的情况下，使所述肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽，其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号，所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内；(d) 为所述用户提供说明以将所述单独的肽与所述至少一种蛋白质靶标比对，其中步骤(c)的所述单独的肽与至少一种蛋白

质靶标之间的所述比对被指定为比对得分；以及 (e) 为所述用户提供说明以使用步骤 (d) 的所述比对得分表征针对所述至少一种蛋白质靶标的所述抗体的结合。

[0018] 另外,本文公开了用于鉴定靶蛋白中的抗体表位的试剂盒和系统,所述试剂盒和系统包括:(a) 提供肽阵列;(b) 提供多种竞争肽;(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内;(d) 为所述用户提供说明以将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤 (c) 的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及 (e) 为所述用户提供说明以确定步骤 (c) 的所述单独的肽中的保守氨基酸,以鉴定保守结合肽基序,并将所述单独的基序与所述至少一种靶蛋白比对,以便鉴定所述靶蛋白的至少一个抗体表位。

[0019] 本文还公开了用于鉴定靶蛋白中的抗体表位的试剂盒和系统,所述试剂盒和系统包括:(a) 提供肽阵列;(b) 提供多种竞争肽;(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内;(d) 为所述用户提供说明以将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤 (c) 的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及 (e) 为所述用户提供说明以确定步骤 (c) 的所述单独的肽中的保守氨基酸,以鉴定保守结合肽基序,并将所述单独的基序与所述至少一种靶蛋白比对,以便鉴定所述靶蛋白的至少一种抗体表位。

[0020] 本文进一步公开了用于表征靶蛋白中的抗体结合区的试剂盒和系统,所述试剂盒和系统包括:(a) 提供第一肽阵列;(b) 提供多种竞争肽;(c) 为用户提供说明以在存在和不存在所述多种竞争肽的情况下,使第一肽阵列与抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第一预定阈值内;(d) 为用户提供说明以使用选自步骤 (c) 中至少一种所述单独的肽的输入肽序列、衍生自步骤 (c) 中所述单独的肽的比对的保守基序或衍生自步骤 (c) 中所述单独的肽的比对的比对基序来产生第二肽阵列,所述第二肽阵列通过以下方法合成:(i) 确定多个合成步骤;(ii) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;(iii) 为每个图案化掩模指定至少一个单体;以及 (iv) 将所述单体耦合至所述特征,其中 (ii) 和 (iii) 组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列;(e) 为所述用户提供说明以使所述第二肽阵列与所述抗体接触以鉴定第二组肽;(f) 为所述用户提供说明以在存在所述多种竞争肽的情况下使所述第二肽阵列与所述抗体接触,并鉴定来自步骤 (e) 的第二组单独的肽,所述第二组单独的肽展现出处于在步骤 (e) 中所述结合信号的第二预定阈值内;以及 (g) 为用户提供说明以将所述第二组单独的肽与所述靶蛋白比对,并鉴定所述靶蛋白中与所鉴定的第二组单独的肽

比对的区域,从而表征所述靶蛋白中的抗体结合区。

[0021] 本文还公开了用于确定抗体与至少一种蛋白质靶标结合倾向的试剂盒和系统,所述试剂盒包括:(a)提供肽阵列;(b)提供多种竞争肽;(c)为用户说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内;(d)为所述用户提供说明以将步骤(c)的所述单独的肽与第一蛋白质靶标比对,其中步骤(c)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;(e)为所述用户提供说明以重复步骤(c)的单独的肽与至少一种另外的蛋白质靶标的比对,其中步骤(c)的所述单独肽与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及(f)为所述用户提供说明以比较步骤(c)和(d)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

[0022] 本文公开了用于确定抗体与至少一种蛋白质靶标结合倾向的试剂盒和系统,所述试剂盒和系统包括:(a)提供第一肽阵列;(b)提供多种竞争肽;(c)为用户说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述第一肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内;(d)为所述用户提供说明以对步骤(c)的所述一种或多种单独的肽进行比对以获得至少一种预测性靶基序;(e)为所述用户提供说明以将所述至少一种预测性靶基序与第一蛋白质靶标比对,其中步骤(c)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;(f)为所述用户提供说明以重复步骤(e)的至少一种预测性靶基序与至少一种另外的蛋白质靶标的比对,其中步骤(e)的所述至少一种预测性靶基序与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及(g)为所述用户提供说明以比较步骤(c)和(d)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

[0023] 在本文公开的一些方法、测定、试剂盒和系统中,所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。在一些公开内容中,所述预定阈值为在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少18倍内、至少16倍内、至少14倍内、至少12倍内、至少10倍内、至少9倍内、至少8倍内、至少7倍内、至少6倍内、至少5倍内、至少4倍内、至少3倍内、至少2倍内、至少1倍内、至少0.5倍内或至少0.2倍内。在本文公开的其他方法、测定、试剂盒和系统中,与不存在竞争剂的情况相比,所述预定阈值是在存在竞争肽情况下的结合信号的至少5%的结合信号。在本文公开的其他方法、测定、试剂盒和系统中,所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少80%、至少85%、至少90%、至少95%、至少100%。在一些实施方案中,所述竞争肽包含生物样品。在其他实施方案中,所述生物样品为血清。在另外其他的实施方案中,所述竞争肽衍生自所述靶蛋白。在另外其他的实施方案中,所述竞争肽与所述靶蛋白至少50%相似。在一些实施

方案中,所述竞争肽与所述靶蛋白至少55%相似、至少60%相似、至少65%相似、至少70%相似、至少75%相似、至少80%相似、至少85%相似、至少90%相似、至少95%相似、至少97%相似或至少100%相似。在一些实施方案中,所述竞争肽衍生自所述抗体的已知表位。在一些实施方案中,所述竞争肽与所述抗体的已知表位至少50%相似。在其他实施方案中,所述竞争肽为与所述抗体的已知表位至少55%相似、至少60%相似、至少65%相似、至少70%相似、至少75%相似、至少80%相似、至少85%相似、至少90%相似、至少95%相似、至少97%相似或至少100%相似的竞争肽。在另外其他的实施方案中,所述竞争肽包含生物样品和衍生自本文公开的所述靶蛋白的肽。

[0024] 在一些实施方案中,所述肽阵列包含至少1000种独特的肽。在其他实施方案中,所述肽阵列包含至少10,000种独特的肽。在另外其他的实施方案中,所述肽阵列包含至少100,000种独特的肽。在另外其他的实施方案中,所述肽阵列包含至少1,000,000种独特的肽。在其他实施方案中,所述肽阵列包含至少5000、至少10,000、至少50,000、至少100,000、至少250,000、至少500,000、至少750,000、至少1,000,000、至少2,000,000、至少3,000,000种或更多种独特的肽。在另外其他的实施方案中,所述肽阵列是原位合成的。在另外其他的实施方案中,所述肽阵列通过以下方法合成:a.接收输入氨基酸序列;b.确定多个合成步骤;c.确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;d.为每个图案化掩模指定至少一个单体;以及e.将所述单体耦合至所述特征;其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0025] 在另外其他的实施方案中,所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。在一些实施方案中,在存在和不存在一种或多种浓度的所述竞争肽的情况下确定表观K<sub>d</sub>。在一些实施方案中,使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的所述比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。本文公开的方法、测定、试剂盒和系统还进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自如本文公开的步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。本文公开的方法、测定、试剂盒和系统还进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自如本文公开的步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

#### 援引并入

[0026] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文,其程度如同特别且单独地指出每个单独的出版物、专利或专利申请通过引用而并入。

#### 附图说明

[0027] 本专利或申请文件含有至少一幅彩色附图。具有彩色附图的本专利或专利申请公开的副本将在请求和支付必要费用后由主管局提供。

[0028] 本发明的新颖特征在所附的权利要求中具体阐述。通过参考以下对利用本发明的

原理的说明性实施方案加以阐述的详细描述以及附图,将会获得对本发明的特征和优点的更好的理解,在这些附图中:

[0029] 图1图示了用于构建肽阵列的光刻工艺。

[0030] 图2示出了取自显微镜的示例性图像。

[0031] 图3示出了直接从肽文库阵列上的单个阵列特征获取的质谱。

[0032] 图4示出了p53Ab1单克隆抗体表位(RHSVV)的丙氨酸扫描。

[0033] 图5示出了掩模算法的图形表示。

[0034] 图6图示了掩模的序列。

[0035] 图7示出了有序合成步骤的图形表示。

[0036] 图8示出了来自计算机模拟肽文库合成的肽长度的示例分布。

[0037] 图9示出了使用本文公开的掩模和合成算法生成的模拟文库中的序列长度的示例分布。

[0038] 图10图示了用于从来自表位基序的阵列肽的聚焦文库中获得表位序列的过程,所述表位基序从获自多样化文库的表位基序的阵列肽的聚焦文库中获得。

[0039] 图11图示了通过在多样化文库中由抗HER2 mAb结合的单独的肽(包括显著肽)获得HER2富集的表位基序(A),并使用该基序提供包含由抗HER2 mAb结合的单独的肽(包括显著肽)的阵列肽的聚焦文库(B),从中鉴定HER2的完整表位序列(C)来鉴定靶蛋白中的示例性线性表位的过程。通过ClustalW比对,在HER2比对的单独的肽中最常鉴别的氨基酸(包括显著肽)显示为WebLogo[Crooks GE等人,(2004)Genome Res 14:1188-1190]。沿x轴显示出相应的HER2序列(UniProt ID=#P04626)。垂直示出了任何一个位置的氨基酸,并且由单字母代码的高度描绘了比对的显著文库肽中的出现比例(proportional occurrence)。

[0040] 图12图示了存在于由抗HER2 mAb结合的多样化文库的单独的肽(包括显著肽)中的三聚体比对的示例性评分。

[0041] 图13示出了从多样化文库的肽中鉴定的一组简化的氨基酸与聚焦的肽阵列文库的全组氨基酸的示例性作图。

[0042] 图14A示出了在抗HER2 mAb Thermo MA5-13675(克隆3B5)的剂量-应答测定中从多样化文库中鉴定的单独的肽(包括显著肽)的比对。

[0043] 图14B示出了在抗HER2 mAb(Thermo MA5-13675(克隆3B5))的剂量-应答测定中从聚焦文库中鉴定的单独的肽(包括显著肽)的阵列肽序列(左列)和相应的比对(右列)。

[0044] 图14C示出了在HER2-比对的肽中最常鉴别的氨基酸显示为WebLogo。

[0045] 图14D示出了已知免疫原和预测的表位序列的相应身份。

[0046] 图15A示出了在抗HER2 mAb(Santa Cruz SC-33684(克隆3B5))的剂量-应答测定中从多样化文库中鉴定的单独的肽(包括显著肽)的比对。

[0047] 图15B示出了在抗HER2 mAb(Santa Cruz SC-33684(克隆3B5))的剂量-应答测定中从聚焦文库中鉴定的单独的肽(包括显著肽)的阵列肽序列(左列)和相应的比对(右列)。

[0048] 图15C示出了在HER2-比对的肽中最常鉴别的氨基酸显示为WebLogo。

[0049] 图15D示出了已知免疫原和预测的表位序列的相应身份。

[0050] 图16A示出了在抗HER2 mAb(Cell Signalling 2165(克隆29D8))的剂量-应答测定中从多样化文库中鉴定的单独的肽(包括显著肽)的比对。

[0051] 图16B示出了在HER2比对的肽中最常鉴别的氨基酸显示为在抗HER2 mAb (Cell Signalling 2165 (克隆29D8)) 的剂量-应答测定中从聚焦文库中鉴定的单独的肽 (包括显著肽) 的WebLogo。

[0052] 图16C示出了在HER2-比对的肽中最常鉴别的氨基酸显示为WebLogo。

[0053] 图16D示出了已知免疫原和预测的表位序列的相应身份。

[0054] 图17A图示了从单独的肽 (包括显著肽) 的比对鉴定的HER2的线性组分和结构表位。

[0055] 图17B示出了在由抗HER2 mAb结合的聚焦文库中单独的肽 (包括显著肽) 的比对。

[0056] 图18示出了如图17所示鉴定的HER2结构表位的线性组分的序列以及抗HER2 mAb 曲妥珠单抗Fab (赫赛汀) 晶体结构与线性组分的相互作用。

[0057] 图19示出了在抗HER2 mAb (Cell Signalling 2165 (克隆29D8)) 的剂量-应答测定中从聚焦文库中鉴定的最高10个单独的肽 (包括显著肽) (A) 和中值10个肽 (B) 的BLAST比对结果。

[0058] 图20示出了在抗HER2 mAb (Thermo MA5-13675 (克隆3B5)) 的剂量-应答测定中从聚焦文库中鉴定的最高10个单独的肽 (包括显著肽) (A) 和中值10个肽 (B) 的BLAST比对结果。

[0059] 图21示出了10种单个肽 (包括显著肽) 的BLAST比对结果, (A), 以及在抗HER2 mAb (Santa Cruz SC-33684 (克隆3B5)) 的剂量-应答测定中从聚焦文库中鉴定的最高10个单独的肽 (包括显著肽) (A) 和中值10个肽 (B) 的BLAST比对结果。

[0060] 图22示出了抗HER2 mAb (Cell Signalling (#2165)) 对HER2和EGFR的倾向。

## 具体实施方式

[0061] 免疫疗法是一种利用人体免疫系统寻找和治疗癌症的癌症治疗。免疫疗法开发的一个高度活跃的领域是使靶向细胞表面受体 (如被癌细胞劫持的T细胞抑制受体 (例如抗-CTLA-4、抗-PD-1)) 的抗体或生物制剂工程化的能力 (即检查点疗法)。一些方法基于靶向多种受体 (如将T细胞和癌细胞与单个双特异性分子结合在一起的BiTE抗体结构) 的工程化多特异性抗体。这些多特异性的体系结构引入了额外的挑战 (如需要表征的更多的先导候选物) 以及增加的脱靶结合潜力。虽然抗体已被证明是药物研发 (R&D) 中灵活且治疗相关的平台, 但在全面表征从早期发现到晚期开发的候选抗体的靶和脱靶结合活性的能力存在显著限制。

[0062] 合成的肽文库通常用于抗体结合表征, 但这是昂贵的并且限于序列空间的小样品 (即表位作图/分类)。目前使用相对低通量的方法如表面等离子体共振和干涉测量进行合成的肽文库的抗体表征, 该相对低通量的方法限于测量小于10,000个抗体-肽相互作用 (例如20个抗体与500个肽)。蛋白质和肽微阵列可用于表征大于10,000个抗体-肽相互作用, 但蛋白质和机器人印刷的肽阵列成本过高, 并且原位合成的肽阵列遭受缺乏可扩展性、再现性和生产质量。噬菌体或酵母肽展示文库也用于鉴定抗体-肽相互作用, 但是这些迭代选择方法仅提供关于最高亲和力相互作用和许多中等亲和力的数据, 未检测到临床相关的抗体-靶标相互作用。由于先导候选失败通常是未检测到脱靶结合效应的结果, 因此抗体-靶标表征中的这些限制最终增加了开发成本,

[0063] 本文公开的技术将能够实现治疗性抗体和生物学先导候选物的可靠、高通量、低成本和全面的结合表征。例如,这些技术的益处包括:1) 增加可表征的先导候选物的数目, 2) 提高先导候选物的成功率, 以及3) 降低免疫疗法开发成本。本文公开的技术包括基于原位肽合成的高度可扩展的基于阵列的肽文库平台, 其具有为半导体制造而开发的过程和设备。本文公开的方法和测定还提供了鉴定抗体结合区(包括表位和推定的表位)以及蛋白质靶向抗体的能力, 从而允许阐明可能在例如不良或非靶向相互作用中起作用的可能的脱靶蛋白。

#### 阵列平台

[0064] 本文公开了提供允许增加化学文库合成的多样性和保真度的阵列平台的方法和过程, 阵列平台包含阵列表面上的多个单独特征。每个特征通常包含在阵列表面上原位合成的多个单独分子, 其中分子在特征内是相同的, 但是分子的序列或同一性在特征之间不同。阵列分子包括但不限于核酸(包括DNA、RNA、核苷、核苷酸、结构类似物或其组合)、肽、肽-模拟物及其组合等, 其中阵列分子可包含分子中天然或非天然单体。这样的阵列分子包括大合成肽阵列的合成。在一些实施方案中, 阵列中的分子为模拟表位, 一种模拟表位结构并能够结合表位引发的抗体的分子。在一些实施方案中, 阵列中的分子为互补位或互补位模拟物, 其包含与抗原表位结合的抗体(或T细胞受体)的可变区中的位点。在一些实施方案中, 本发明的阵列为包含随机、半随机或多样化肽序列的肽阵列。在一些实施方案中, 多样化肽序列可衍生自蛋白质组文库, 例如衍生自特定生物体(参见, 例如, 结核分枝杆菌(Mtb)蛋白质组文库(Schubert等人, Cell Host Microbe (2013) 13 (5):602-12)或细胞器(参见, 例如, 线粒体(Mtd)蛋白质组文库(Calvo和Mootha, Annu. Rev. Genomics (2010) 11:25-44)等。

[0065] 在另外其他的实施方案中, 多样化肽序列可衍生自一组所有已知的氨基酸组合, 例如至少100%的所有可能的四聚体、至少90%的所有可能的四聚体、至少85%的所有可能的四聚体、至少80%的所有可能的四聚体、至少75%的所有可能的四聚体、至少70%的所有可能的四聚体、至少65%的所有可能的四聚体、至少60%的所有可能的四聚体、至少55%的所有可能的四聚体、至少50%的所有可能的四聚体、至少45%的所有可能的四聚体、至少40%的所有可能的四聚体、至少35%的所有可能的四聚体、至少30%的所有可能的四聚体或至少25%的所有可能的四聚体。在另外其他的实施方案中, 多样化肽序列可衍生自一组所有可能的五聚体, 例如, 至少100%的所有可能的五聚体、至少95%的所有可能的五聚体、至少90%的所有可能的五聚体、至少85%的所有可能的五聚体、至少80%的所有可能的五聚体、至少75%的所有可能的五聚体、至少70%的所有可能的五聚体、至少65%的所有可能的五聚体、至少60%的所有可能的五聚体、至少55%的所有可能的五聚体、至少50%的所有可能的五聚体、至少45%的所有可能的五聚体、至少40%的所有可能的五聚体、至少35%的所有可能的五聚体、至少30%的所有可能的五聚体或至少25%的所有可能的五聚体。在另外其他的实施方案中, 阵列的多样化肽序列可衍生自一组氨基酸组合, 例如25%-100%的所有可能的六聚体、25%-100%的所有可能的七聚体、25%-100%的所有可能的八聚体、25%-100%的所有可能的九聚体或25%-100%的所有可能的十聚体或其组合。多样化肽序列的表示仅受阵列大小的限制。因此, 较大阵列, 例如, 至少1百万、至少2百万、至少3百万、至少4百万、至少5百万、至少6百万、至少7百万、至少8百万、至少9百万、至少1000万个或更

多的肽可与本文公开的方法、系统和测定一起使用。备选的或另外的,可合成多个基本上不重叠的肽文库/阵列以覆盖解析由生物样品或抗体识别的肽序列或基序所需的序列空间。

[0066] 在一些实施方案中,阵列上的单独肽具有可变和/或不同的长度。在一些实施方案中,肽的长度为约6-20个氨基酸、或长度为约7-18个氨基酸,或长度为约8-15个氨基酸,或长度为约9-14个氨基酸。在其他实施方案中,肽的长度为至少6个氨基酸、至少7个氨基酸、至少8个氨基酸、至少9个氨基酸、至少10个氨基酸、至少11个氨基酸、至少12个氨基酸、至少13个氨基酸、至少14个氨基酸、至少15个氨基酸。在另外其他的实施方案中,肽的长度不超过15个氨基酸、不超过14个氨基酸、不超过13个氨基酸、不超过12个氨基酸、不超过11个氨基酸、不超过10个氨基酸、不超过9个氨基酸或不超过8个氨基酸。在另外其他的实施方案中,阵列上的肽的平均长度为约6个氨基酸、约7个氨基酸、约8个氨基酸、约9个氨基酸、约10个氨基酸、约11个氨基酸、约12个氨基酸、约13个氨基酸、约14个氨基酸或约15个氨基酸。

[0067] 在另外其他的实施方案中,阵列上肽的氨基酸结构单元包含所有天然氨基酸。在其他实施方案中,阵列上肽的氨基酸结构单元由非天然或合成氨基酸组成。在另外其他的实施方案中,仅19个氨基酸用作合成阵列上肽的结构单元。在另外其他的实施方案中,仅18个氨基酸、仅17个氨基酸、仅16个氨基酸、仅15个氨基酸或仅14个氨基酸用作合成阵列上肽的结构单元。在一些实施方案中,在肽合成期间省略半胱氨酸。在其他实施方案中,在肽合成期间省略蛋氨酸。在另外其他的实施方案中,在肽合成期间省略异亮氨酸。在另外其他的实施方案中,在肽合成期间省略苏氨酸。在另外其他的实施方案中,在肽合成期间省略半胱氨酸、甲硫氨酸、异亮氨酸和/或苏氨酸(包括其所有的组合)。

[0068] 在一些实施方案中,本发明的阵列为肽阵列,该肽阵列包含聚焦或有限的一组肽序列,所有肽序列均衍生自输入氨基酸或肽序列、或输入氨基酸或肽基序。一种或多种肽阵列可与本文公开的方法、系统和测定一起使用,该一种或多种肽阵列包括多样化或半随机肽阵列和/或聚焦或有限的一组肽序列。例如,本文公开的方法、系统和测定可利用多样化的一组肽和所选择的聚焦或有限的一组肽。肽阵列可与本文公开的生物样品平行或顺序使用。例如,最初可使用多样化的肽阵列,并且获得单克隆抗体的至少一个基序(基于序列或基于结构的)或序列,例如,具有未知的结合谱。然后可将鉴定的基序或序列用作产生至少一个聚焦或有限的一组肽序列的输入序列以及如本文所述进行的测定。使用本文所述的方法、系统和阵列,多个聚焦或有限的一组肽阵列可用于表征未知单克隆抗体的抗体结合。

[0069] 几乎所有治疗性抗体筛选在选定数目的先导抗体上掺入一定水平的表位作图和表位分类,并且这些数据驱动有关哪些先导抗体向前移动到开发管线中的决定。表位作图研究通常利用肽的系统重叠序列来确定负责抗体-靶标相互作用的氨基酸。表位分类研究绘制几种先导抗体的表位,并随后通过它们对鉴定的表位的结合亲和力/动力学来分类抗体。表位分类研究是鉴定具有不同表位反应性和潜在不同作用模式和脱靶效应的先导抗体的关键决策数据集。通常,使用与已知表位相关的靶向肽序列的合成文库进行表位分类和表位作图表征,由于有限的分析通量和纯化的合成肽文库的高成本,因此其将分析限制为几千个靶向相互作用(例如10个先导抗体与100个肽)。如此少量的抗体-靶标相互作用的表征允许未检测到许多脱靶和/或低亲和力相互作用,这增加了开发管线中晚期候选物的失败率。

[0070] 所有当前表位作图/分类平台的共同弱点是相对于可能的相互作用的总数严格限

制抗体-表位相互作用分析的通量。该分析通量限制迫使抗体发现科学家减少选择进一步开发的先导抗体数目。结果,先导抗体数目减少增加了晚期抗体治疗性候选物失败的风险。这最终增加了那些成功的候选物的成本,并反过来补贴失败候选物的研发成本。与有限的分析通量相关的风险随着多特异性抗体筛选的出现而增加,该多特异性抗体筛选需要选择更多的先导抗体以鉴定具有与靶标疾病相关的特定多特异性和最小脱靶效应的候选物。

[0071] 本文公开的技术包括光刻阵列合成平台,该光刻阵列合成平台合并半导体制造过程和组合化学合成以在硅晶片上产生基于阵列的文库。图1示出了光刻工艺的剖面视图;平台包括用于生长肽合成的基底101。施加掩模102并随后施加UV光103可以控制肽合成。此外,通过依次施加具有UV光曝光的另一掩模,可以建立各种阵列特征。通过利用光刻特征图案化的巨大进步,该阵列合成平台具有高度可扩展性,并且能够在8英寸晶片上产生具有4000万个特征的组合化学文库。使用半导体晶片生产设备在10,000级洁净室中进行光刻阵列合成,以实现高再现性。当将晶片切割成标准显微镜载玻片尺寸时,每个载玻片含有超过300万个不同的化学实体。

[0072] 在一些实施方案中,具有通过本文公开的技术产生的化学文库的阵列用于基于免疫的诊断测定,例如称为免疫特征测定。使用来自与阵列结合的一滴血液的患者抗体库,结合阵列的荧光结合剖面图像提供了足够的信息来对疾病与健康进行分类。图2示出了取自显微镜的示例图像。该图像包含与阵列结合的IgG抗体库的荧光图像。每个方形特征为 $14\mu\text{m}^2$ ,并在显微镜载玻片上以超过300万个不同肽的密度进行图案化。

[0073] 在一些实施方案中,正在开发用于临床应用的免疫特征测定以诊断/监测自身免疫疾病并评估对自身免疫治疗的应答。免疫特征测定的示例性实施方案详细描述于题为“Compound Arrays for Sample Profiling”的美国预授权公开号2012/0190574和题为“Immunosignaturing: A Path to Early Diagnosis and Health Monitoring”的美国预授权公开号2014/0087963”中,对于这样的公开内容,这两篇申请都通过引用并入本文。本文开发的阵列使用包括椭圆偏光法、质谱分析法和荧光的正交分析方法在每个合成阵列内结合分析测量能力。这些测量能够对阵列合成性能进行纵向定性和定量评估。

[0074] 原位合成的肽阵列的主要缺点之一是无法直接测量合成的肽特征的纯度。在一些实施方案中,所述技术包括直接来自硅晶片的合成肽的定性原位质谱分析法。通过在硅表面与合成的肽之间掺入气相可切割的连接器以便在不从阵列特征扩散的情况下完成肽的切割来进行质谱分析法。在肽切割后,通过施加薄的气溶胶基质层直接在硅表面上进行基质辅助激光解吸电离(MALDI)质谱分析法,并随后将MALDI激光聚焦在单独的肽特征上以获取每种合成肽的质谱。

[0075] 图3示出了直接从肽文库阵列上的单个阵列特征获取的质谱。来自使用光刻合成方法产生的肽阵列特征的定性原位MALDI质谱也包括在本文描述的方法和装置中。本领域技术人员已知的其他分析也可用于定量和/或限定本文公开的原位合成过程的保真度。

#### 抗体与肽阵列的结合

[0076] 在各种实施方案中,本文公开的方法、系统和技术提供肽阵列平台,该肽阵列平台用于检测肽阵列上发生的结合事件(包括抗体与肽结合事件)。在一些实施方案中,肽阵列是高密度肽阵列。在一些实施方案中,阵列包含阵列上的特征内的单独的肽,该单独的肽间隔小于0.5nm、小于1nm、小于2nm、小于3nm、小于4nm、小于5nm、小于6nm、小于7nm、小于8nm、

小于9nm、小于10nm、小于11nm、小于12nm、小于13nm、小于14nm或小于15nm。

[0077] 添加生物样品并允许与肽阵列一起温育。生物样品包括血液、干血、血清、血浆、唾液、泪液、泪管液、检查拭子 (check swab)、活检、组织、皮肤、毛发、脑脊液样品、粪便或尿液样品。在一些实施方案中,受试者可以,例如,使用“手指针刺 (fingerstick)”或“手指刺破 (fingerprick)”来抽取少量血液并将其添加到表面 (如滤纸或其他可吸收源),或者添加到小瓶或容器中并任选地干燥。可以浓缩或稀释由受试者提供的生物样品。在另外其他的实施方案中,生物样品是纯化的抗体制剂,其包括单克隆抗体、多克隆抗体、抗体片段、单链抗体、嵌合抗体、人源化抗体、抗体药物缀合物等。在另外其他的实施方案中,生物样品是用于在细胞宿主中繁殖重组抗体的细胞培养物或其他生长培养基。

[0078] 在一些实施方案中,通过本文公开的方法或系统进行分析不需要超过约0.5nl至约50 $\mu$ l的生物样品。在另外其他的实施方案中,分析需要约0.5nl至25 $\mu$ l、约5nl to 10 $\mu$ l、约5nl至5 $\mu$ l、约10nl至5 $\mu$ l、约10nl至2.5 $\mu$ l、约100nl至2.5 $\mu$ l或约100nl至1 $\mu$ l的生物样品。在一些实施方案中,受试者可提供来自例如活检或组织的固体生物样品。在一些实施方案中,通过本文公开的方法或系统进行分析需要约1mg、约5mg、约10mg、约15mg、约20mg、约25mg、约30mg、约35mg、约40mg、约45mg、约50mg、约55mg、约60mg、约65mg、约7mg、约75mg、约80mg、约85mg、约90mg、约95mg或约100mg的生物样品。

[0079] 在一些实施方案中,来自受试者的生物样品的浓度过高,因此在与本发明的阵列接触之前需要稀释。在生物样品与本发明的阵列接触之前,可对该样品进行多次稀释。稀释可以是系列稀释,其可以以对数方式产生几何级数的浓度。例如,10倍系列稀释可以为1M、0.01M、0.001M及其几何级数。稀释可以为,例如,一倍稀释、两倍稀释、三倍稀释、四倍稀释、五倍稀释、六倍稀释、七倍稀释、八倍稀释、九倍稀释、十倍稀释、十六倍稀释、二十五倍稀释、三十二倍稀释、六十四倍稀释和/或一百二十五倍稀释。

#### 结合事件的检测

[0080] 可以以多种形式检测样品组分与肽阵列之间的结合相互作用。在一些形式中,对样品的成分进行标记。标记物可以是放射性同位素或染料等。可通过在获得样品前对患者施用标记物或通过将标记物连接至样品或其选择性成分上来提供标记物。

[0081] 结合相互作用还可使用第二检测试剂如抗体来检测。例如,样品中的抗体与阵列的结合可使用对抗体的同种型 (例如,IgG (包括任何亚型,如IgG1、IgG2、IgG3和IgG4)、IgA、IgM) 具有特异性的第二抗体来检测。第二抗体通常被标记,并且可与被分析样品中特定同种型的所有抗体结合。可使用具有不同的同种型特异性的不同的第二抗体 (例如,来自不同的宿主)。

[0082] 结合相互作用还可使用无标记方法如表面等离子体共振 (SPR) 和质谱分析法来检测。SPR可提供对解离常数和解离率的测量,例如,使用A-100Biocore/GE仪器进行这种类型的分析。

[0083] 结合事件的检测也可以在存在竞争肽的情况下发生。在一些实施方案中,竞争性抑制剂是与本文公开的确定的表位、基序或输入序列相同、相似的肽或与衍生自本文公开的确定的表位、基序或输入序列的肽。在一些实施方案中,竞争性抑制剂肽包含至少2种、至少3种、至少5种、至少10种、至少15种、至少20种、至少25种、至少30种、至少35种、至少40种、至少45种或至少50种不同肽的混合物。在一些实施方案中,竞争肽包含天然和/或非天然氨

基酸。在一些实施方案中,竞争性抑制肽包含与确定的表位、基序或输入序列至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%和/或至少99%相同。在其他实施方案中,竞争性抑制肽包含与确定的表位、基序或输入序列至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%和/或至少99%相似。在一些实施方案中,相似性可以通过序列或通过结构确定。在其他实施方案中,竞争性抑制肽可以包含随机肽或半随机肽的混合物。在另外其他的实施方案中,竞争性肽混合物可以包括添加到本文公开的竞争性抑制肽或代替本文公开的竞争性抑制肽的生物来源,例如血清、血浆或血液。通过将竞争性抑制肽添加到结合反应中,并在不存在和存在竞争性抑制肽的情况下测量结合信号的变化,可以获得特异性测量,该特异性测量传达关于阵列上的肽与生物样品之间相互作用的严格性的信息。特异性可以根据在存在竞争者的情况下测量的亲和力(Kd)和/或具有与生物样品或抗体结合的确定的基序或序列的鉴定的肽的数目来测量,并鉴定为推定的结合位点。

#### 治疗性抗体的开发和表征:抗体表位结合谱

[0084] 在一些实施方案中,检测肽阵列上的抗体结合提出了一些可通过本文公开的技术解决的挑战。该技术可以通过特定的涂层和官能团密度来调整表面性质,该技术已被用于解决使用肽阵列分析抗体结合的两个潜在缺点。首先,通过用适度亲水的单层聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合涂覆硅表面,使肽阵列上的非特异性抗体结合最小化。在一些实施方案中,所述亲水单层是均匀的。其次,使用使肽远离表面移动の間隔物将合成的肽与硅表面连接,使得肽以无阻碍的方向呈递给抗体。此外,与主要是无序结构的线性肽相比,表面间隔物可用于使肽环化,使得所有肽以一致的有序结构呈递给抗体。在一些实施方案中,表面间隔物包括以下序列:循环间隔物例1:半胱氨酸-甘氨酸-脯氨酸-甘氨酸-(可变氨基酸序列(Xaa)<sub>n</sub>)-甘氨酸-脯氨酸-甘氨酸-半胱氨酸,其中两个半胱氨酸残基能够在氧化条件下环化。在其他实施方案中,表面间隔物可包括以下序列:循环间隔物例2:半胱氨酸-(PEG3)-(可变氨基酸序列(Xaa)<sub>n</sub>)-(PEG3)-半胱氨酸,其中两个半胱氨酸残基能够在氧化条件下环化。在另外其他的实施方案中,表面间隔物包括以下序列:循环间隔物例3:赖氨酸-(PEG3)-(可变氨基酸序列(Xaa)<sub>n</sub>)-(PEG3)-赖氨酸,其中两个赖氨酸残基能够与同双官能的二-NHS酯键环化。总之,这些表面开发在与溶液相抗体结合接近或相关的阵列上产生抗体结合谱。

[0085] 图4示出了本文公开的方法的示例性实施方案,其描绘了p53Ab1单克隆抗体表位(RHSVV)的丙氨酸扫描。用丙氨酸单独替换到表位的前三个位置合成丙氨酸扫描文库阵列(RHS)。与本文开发的阵列上的天然表位(401)相比,减少的p53Ab1抗体与丙氨酸替换特征(402、403和404)的结合与公开的p53Ab1表位变体结合相关。

[0086] 在一些实施方案中,本文公开的技术解决了使用阵列检测抗体结合谱的抗体标记方法。抗体的直接荧光标记经常抑制、修饰或消除与已知表位的结合。为了解决这个问题,本文公开的技术包括与夹心ELISA测定相似的“夹心测定”方法,其首先将未标记的第一抗体(被分析的抗体)结合至阵列,随后结合荧光标记的第二抗体,将该第二抗体结合至未标记的第一抗体上的固定表位(例如IgG抗体的Fc区)。在阵列上温育之前验证标记的第二抗

体与第一抗体的结合,以确保标记的第二抗体如预期的那样结合第一抗体。

[0087] 已经用p53结合单克隆抗体(p53Ab1)的已知肽表位(RHSVV)的丙氨酸扫描进行了阵列表面和测定进展产生稳健的抗体结合谱的验证。使用光刻肽阵列合成来合成丙氨酸扫描肽序列组,并且该丙氨酸扫描肽序列组包括:AHSVV、RASVV、RHAVV,其中丙氨酸(A)被替换到表位的前三个位置。如图4所示,使用具有p53AB1抗体和丙氨酸扫描阵列的夹心测定方法,将与每个丙氨酸替换序列的结合与已知表位RHSVV进行比较。从阵列获得的p53Ab1丙氨酸扫描抗体结合谱结果与公开的结果匹配,该公开的结果示出了p53Ab1需要肽表位中的R、H和S以用于高亲和力结合。

#### 掩模算法

[0088] 因为靶表位可以用作算法的输入序列,所以本文公开的新掩模和合成算法与抗体发现和表征特别相关,并且因此可以筛选围绕该表位的化学空间区域,该筛选包括添加、截断、替换和缺失。由于癌症中存在高表位突变率,因此这在筛选癌症靶抗体中特别重要。通过筛选表位周围的序列空间区域,可以检测到大量癌症相关突变。

[0089] 通过在顺序掩模(即掩模 $n$ 与掩模 $n+1$ )之间包括开放特征的设定重叠百分比,可以合成高度多样化的化学文库阵列(例如,肽阵列),这允许对输入序列(例如,靶表位)周围的序列空间进行彻底的作图和分析。

[0090] 在一些实施方案中,使用一组固定的光刻掩模来对由长度为 $n$ 的任何输入序列定义的化学空间区域进行采样,其中 $n$ 为该组中掩模的数目。该算法克服了光图案化合成中的灵活性的主要限制,因为生成具有定义序列的文库通常需要一组新的掩模,这是昂贵且耗时的。

[0091] 一个主要的创新成果是高度可扩展的综合抗体结合表征平台,其具有测量至少500,000个序列变体的结合谱(即表位作图和分类)的能力,该序列变体衍生自直至氨基酸序列的长度为,例如,最多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、31、32、33、34、35、36、37、38、39、40、50、60、70、80、90、100个氨基酸的任何输入肽表位序列。本文公开的平台方法和装置将显著推进治疗性抗体和免疫疗法发现的领域。由该新的平台实现的抗体结合信息的量和细节可以促进发现具有新作用模式和/或尚未实现的最小脱靶副作用的新的基于抗体的治疗。所提出的开发促进了多特异性抗体的开发,该多特异性抗体的开发需要对大量具有更复杂的结合谱的先导候选相对于单特异性先导抗体进行表征。

[0092] 在一些实施方案中,本文公开的技术产生抗体表征平台,该抗体表征平台将抗体-表位谱通量增加至每天数百万次相互作用,相对于当前平台增加至少一个数量级。可以使用本文描述的平台用大量治疗性抗体候选物(数百候选物)进行初始抗体-表位相互作用谱研究(例如表位分类)。

[0093] 在一些实施方案中,设计了规定的光刻掩模和合成算法以生成以任何长度( $k$ )(例如, $k=10$ )的输入序列为中心的序列空间。换言之,单组掩模产生肽变体文库阵列,该肽变体文库阵列衍生自长度为最多10个氨基酸的任何输入序列。在一些实施方案中,因为免疫原性表位肽的长度通常为8-10个氨基酸,所以选择10的输入序列长度。在一些实施方案中,选择另一长度,例如,最多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、31、32、33、34、35、36、37、38、39、40、50、60、70、80、90、100个氨

基酸。

[0094] 在一些实施方案中,通过该掩模和合成算法产生的肽阵列序列空间不限于特定的输入序列,并且该肽阵列序列空间由以下定义:1) 用户指定的输入序列,2) 合成步骤的顺序,3) 掩模 $n$ 和 $(n-1)$ 之间的常见开放特征的级分,其中开放特征是阵列特征,该阵列特征对于光通过特定掩模 $n$ 是开放的,从而导致在该位置添加下一个添加的氨基酸。

[0095] 在一些实施方案中,本文公开的掩蔽和合成算法用于迭代地优化先导序列(例如肽),使得可以筛选以先导序列为中心的序列空间区域的更高活性序列。可以选择新的先导抗体并且迭代地重复该过程以最终鉴定具有所需活性水平的肽序列。在一些实施方案中,所述技术用于亲和试剂(例如肽),使得可以迭代地改善先导肽的亲合力直至达到所需的亲合力或用于酶活性试剂(例如酶抑制剂或激活剂)。

[0096] 在一些实施方案中,所公开的光刻掩模和化学合成算法包括简单且相对较少的化学步骤。该算法可以将掩模和化学合成视为连接的组合问题,使得采样空间由以下定义:1) 输入序列,2) 合成步骤的顺序和3) 掩模之间的顺序特征重叠(即,掩模 $n$ 和 $n-1$ 之间的重叠)。可以在计算机中模拟该算法以使用定义的1) 输入序列、2) 化学步骤的顺序和3) 百分比顺序掩模重叠来计算所生成空间的大小和多样性。

[0097] 本文公开的技术包括利用光刻掩模和合成顺序以对由输入序列(例如肽序列)和化学步骤的顺序定义的化学空间区域进行采样的的算法。该算法确定在用于合成的相邻顺序掩模之间重叠的特征( $p$ )的设定百分比。本文公开的算法可以用数学方法描述。设 $p$ 表示掩模 $n$ 与掩模 $n-1$ 之间重叠的开放特征的百分比; $x$ 表示文库中的特征总数; $\cap$ 表示交集或一组重叠特征(行,列); $|\cdot|$ 表示基数或重叠特征的数目。掩码算法为: $p \times x = |(\text{掩模}n) \cap (\text{掩模}n-1)|$ 。

[0098] 图5示出了掩模算法的图形表示。对于每个掩模 $n$ ,与掩模 $n-1$ 存在重叠百分比(图5中的阴影区域)。根据重叠百分比的比例,在系列中的多个(或所有)顺序掩模之间可存在共享重叠(例如,由双箭头示出)。可以调整该共享重叠以定义采样空间中的序列的多样性和中值长度。

[0099] 图6示出了掩模算法的另一种表示。一组连续的掩码用于选择性地曝光和激活用于合成的阵列特征。每个掩模( $n$ )与掩模( $n-1$ )具有级分特征重叠。使用该算法,文库中最大可能的肽长度等于用于构建文库阵列的掩模和化学步骤的数目,并且文库的中值肽长度取决于掩模 $n$ 与 $(n-1)$ 之间的开放特征重叠的级分。

[0100] 图7示出了具有有序合成步骤的一些实施方案的图形表示。在图7中,氨基酸耦合的顺序基于输入序列。在该实例中,将12-聚体输入序列HVGAAAPVVPQA建立在前12个步骤中全部特征的一小部分上,其中每个步骤对应于掩模编号。该输入序列周围的采样序列空间区域由以下生成:(a) 氨基酸的顺序,(b) 步骤1-12中重叠和非重叠特征的比率,以及(c) 步骤13-25中使用的重叠特征和氨基酸的百分比。存在合成顺序的其他实例,其中输入序列中不存在的氨基酸与输入氨基酸序列交错(参见以下具体实施方案部分)。所公开的算法具有合理地聚焦序列空间或甚至对随机序列空间(即,掩模和合成步骤的顺序是随机的)进行采样的灵活性。

[0101] 在一个示例性实施方案中,设计并产生总共25个光刻掩模以适应包括肽文库阵列合成中的所有20种天然氨基酸的化学步骤。光刻掩模阵列特征将以 $18\mu\text{m}$ 间距进行图案化,

以在25个合成步骤后产生500,000个总阵列特征,其中每个75mm×25mm载玻片具有8个重复阵列,每个载玻片总共有4,000,000个特征(1个载玻片=每个阵列500,000个特征的8个重复阵列=4,000,000个特征/载玻片)。掩模系列中掩模n与(n-1)之间的开放特征重叠的级分将被设置为42%,以实现等于输入序列长度10的中值肽库阵列长度。每个掩模将具有从总共500,000个阵列特征中随机选择的210,000个开放特征,其中掩模n上的42%的开放特征与掩模(n-1)重叠。使用治疗性抗体10-聚体表位输入肽序列(QMWAPQWGP,一种赫赛汀治疗性抗体表位[57])从计算机模拟合成确定该级分重叠,以从25个光刻合成步骤生成458,305个不同序列变体和41,695个重复的序列变体的文库。图8示出了来自使用图6中所示的规定掩模算法的计算机模拟肽文库合成的肽长度的分布,所述掩模算法具有10-聚体输入序列和25个合成步骤,其中中值长度为10。

#### 质谱分析法检测

[0102] 在一些实施方案中,本文公开的技术开发了每个芯片上的一组肽序列的原位质谱分析法检测,其检测每个合成步骤以定量每个步骤的效率和纯度。这些技术建立在初始的MALDI开发基础上,以实现产量和纯度定量。在一些实施方案中,通过掺入气相可切割的安全捕获连接器(SCL)从合成的肽阵列获取原位MALDI质谱,该连接器对于结合测定条件是稳定的并且可以使用氨气从硅表面切割下来而不扩散。该SCL将耦合至胺官能化硅表面,并且肽将由SCL表面连接构建。在8英寸晶片上合成肽阵列后,从晶片切割出13个显微镜载玻片尺寸的芯片,每个芯片具有8个重复的肽阵列,并且保留一个芯片用于MALDI质谱采集。MALDI保留芯片的氨气处理从硅表面切割合成的肽而不扩散。气相切割后,使用微滴气溶胶应用将促进肽解吸/电离的MALDI基质施加至芯片,而不使切割的肽在阵列表面上扩散。最后,通过将MALDI激光与相对于一组比对基准标记物的特异性切割的肽阵列特征比对,从合成的肽阵列原位获得MALDI质谱,以确保激光以用于质谱采集的预期阵列特征为中心。

[0103] 在一些实施方案中,为了使用MALDI质谱分析法定量每个合成步骤的效率,在所产生的掩模(例如,前述25个掩模)上包括一组500 $\mu\text{m}^2$  MALDI合成-分析阵列特征。在8英寸晶片内的所有13个芯片上使对应于25个合成步骤中的每一个的总共25个MALDI分析阵列特征图案化,从而能够对组合合成中的所有步骤进行效率计算。作为第一合成步骤,将共同的C-末端(第一合成位置)氨基酸(例如甘氨酸)耦合至所有MALDI分析阵列特征。在共同氨基酸后,将每个单独的MALDI分析特征与25个阵列合成掩模中的每一个串联地光保护。将用于该合成步骤的相应氨基酸耦合至光保护的MALDI分析特征以产生二聚体序列,该二聚体序列由与共同氨基酸(例如精氨酸-甘氨酸二聚体)耦合的该合成步骤的氨基酸组成。为了使所有肽序列的MALDI电离标准化,将三(2,4,6-三甲氧基苯基)磷(TMPP)信号增强子耦合至所有N末端。在从所有25个特征采集MALDI质谱数据之后,每个合成步骤的效率将计算为所需二聚体与普通单体(例如精氨酸-甘氨酸与甘氨酸)的质谱峰的比率。

#### 结合谱再现性

[0104] 在各种实施方案中,本文公开的技术包括用一组5种工程化抗体定量阵列内和阵列间结合谱再现性,并用肽再合成和表面等离子共振(SPR)确认结合谱。

[0105] 在一些实施方案中,一组5种单克隆抗体和5种单独的阵列用于定量抗体结合谱再现性(即%CV)。通过使用一组确定的抗体,可以严格控制抗体浓度和样品组成,以测量阵列产生的可变性与样品或测定中的可变性。

[0106] 在测试所获得的结合谱再现性的示例性实施方案中,将从文献中鉴定出长度为6-10个氨基酸范围内的五个不相关的肽表位,并用作表位变体文库的5个单独的肽阵列合成的输入序列。使用经工程化以结合所选表位的五种IgG单克隆抗体。五种抗体中的每一种分别与它们各自的变体文库阵列结合。使用荧光标记的抗-IgG Fc第二抗体来标记第一抗体结合,该第二抗体基于夹心测定方案与第一IgG抗体的Fc区结合。将使用一个阵列内的重复肽特征荧光强度来计算阵列内%CV。将使用重复阵列上的相同特征荧光强度来计算阵列间%CV。从五种抗体阵列结合谱(总共25种肽)中的每一种中选择五种表位变体序列用于合成和纯化,随后进行溶液相SPR结合分析。

#### 数字处理设备

[0107] 在一些实施方案中,本文所述的系统、平台、软件、网络和方法包括数字处理设备或其使用。在进一步的实施方案中,数字处理设备包括一个或多个硬件中央处理单元(CPU),即执行设备功能的处理器。在更进一步的实施方案中,数字处理设备进一步包含被配置为进行可执行指令的操作系统。在一些实施方案中,数字处理设备任选地连接计算机网络。在进一步的实施方案中,数字处理设备任选地连接到因特网,使得其访问万维网。在更进一步的实施方案中,数字处理设备任选地连接到云计算基础设施。在其他实施方案中,数字处理设备任选地连接到内联网。在其他实施方案中,数字处理设备任选地连接到数据存储设备。

[0108] 根据本文的描述,作为非限制性实例,合适的数字处理设备包括服务器计算机、台式计算机、膝上型计算机、笔记本计算机、次笔记型计算机、上网本计算机、记事本计算机、机顶计算机、手持式计算机、互联网设备、移动智能电话、平板计算机、个人数字助理、视频游戏控制台和媒介物。本领域技术人员将认识到,许多智能电话适用于本文所述的系统。本领域技术人员还将认识到,具有可选计算机网络连接的选择电视、视频播放器和数字音乐播放器适用于本文所述的系统。合适的平板计算机包括具有本领域技术人员已知的小册子、平板和可转换配置的平板计算机。

[0109] 在一些实施方案中,数字处理设备包括被配置为执行可执行指令的操作系统。操作系统例如包含程序和数据的软件,该软件管理设备的硬件并为应用程序的执行提供服务。本领域技术人员将认识到,作为非限制性实例,合适的服务器操作系统包括FreeBSD、OpenBSD、NetBSD<sup>®</sup>、Linux、Apple<sup>®</sup> Mac OS X Server<sup>®</sup>、Oracle<sup>®</sup> Solaris<sup>®</sup>、Windows Server<sup>®</sup>和Novell<sup>®</sup> NetWare<sup>®</sup>。本领域技术人员将认识到,作为非限制性实例,合适的个人计算机操作系统包括Microsoft<sup>®</sup> Windows<sup>®</sup>、Apple<sup>®</sup> Mac OS X<sup>®</sup>、UNIX<sup>®</sup>和类似UNIX的操作系统如GNU/Linux<sup>®</sup>。在一些实施方案中,操作系统由云计算提供。本领域技术人员还将认识到,作为非限制性实例,合适的移动智能电话操作系统包括Nokia<sup>®</sup> Symbian<sup>®</sup> OS、Apple<sup>®</sup> iOS<sup>®</sup>、Research In Motion<sup>®</sup> BlackBerry OS<sup>®</sup>、Google<sup>®</sup> Android<sup>®</sup>、Microsoft<sup>®</sup> Windows Phone<sup>®</sup> OS、Microsoft<sup>®</sup> Windows Mobile<sup>®</sup> OS、Linux<sup>®</sup>和Palm<sup>®</sup> WebOS<sup>®</sup>。

[0110] 在一些实施方案中,数字处理设备包括存储和/或存储器设备。该存储设备和/或存储器设备是用于临时或永久地存储数据或程序的一个或多个物理设备。在一些实施方案

中,该设备是易失性存储器且需要电力来维护存储的信息。在一些实施方案中,该设备是非易失性存储器,并且在数字处理设备未通电时保留存储的信息。在进一步的实施方案中,非易失性存储器包含闪速存储器。在一些实施方案中,非易失性存储器包含动态随机存取存储器(DRAM)。在一些实施方案中,非易失性存储器包含铁电随机存取存储器(FRAM)。在一些实施方案中,非易失性存储器包含相变随机存取存储器(PRAM)。在其他实施方案中,该设备是存储设备,作为非限制性实例,该存储设备包括CD-ROM、DVD、闪速存储器设备、磁盘驱动器、磁带驱动器、光盘驱动器和基于云计算的存储器。在进一步的实施方案中,存储和/或存储器设备是诸如本文公开的那些设备的组合。

[0111] 在一些实施方案中,数字处理设备包括用于向用户发送视觉信息的显示器。在一些实施方案中,显示器是阴极射线管(CRT)。在一些实施方案中,显示器是液晶显示器(LCD)。在进一步的实施方案中,显示器是薄膜晶体管液晶显示器(TFT-LCD)。在一些实施方案中,显示器是有机发光二极管(OLED)显示器。在各种其他实施方案中,在OLED显示器上是无源矩阵OLED(PMOLED)或有源矩阵OLED(AMOLED)显示器。在一些实施方案中,显示器是等离子显示器。在其他实施方案中,显示器是视频投影仪。在更进一步的实施方案中,显示器是诸如本文公开的那些设备的组合。

[0112] 在一些实施方案中,数字处理设备包括用于从用户接收信息的输入设备。在一些实施方案中,输入设备是键盘。在一些实施方案中,输入设备是指向设备,作为非限制性实例,该指示设备包括鼠标、跟踪球、跟踪板、操纵杆、游戏控制器或触笔。在一些实施方案中,输入设备是触摸屏或多点触摸屏。在其他实施方案中,输入设备是用于捕获语音或其他声音输入的麦克风。在其他实施方案中,输入设备是用于捕获运动或视觉输入的摄像机。在更进一步的实施方案中,输入设备是诸如本文公开的那些设备的组合。

[0113] 在一些实施方案中,数字处理设备包括数码相机。在一些实施方案中,数码相机捕捉数字图像。在一些实施方案中,数码相机是自动对焦相机。在一些实施方案中,数码相机是电荷耦合器件(CCD)相机。在进一步的实施方案中,数码相机是CCD摄像机。在其他实施方案中,数码相机是互补金属氧化物半导体(CMOS)相机。在一些实施方案中,数码相机捕捉静止图像。在其他实施方案中,数码相机捕捉视频图像。在各种实施方案中,合适的数码相机包括100万、200万、300万、400万、500万、600万、700万、800万、900万、1000万、1100万、1200万、1300万、1400万、1500万、1600万、1700万、1800万、1900万、2000万、2100万、2200万、2300万、2400万、2500万、2600万、2700万、2800万、2900万、3000万像素和更高像素的照相机,包括其间的增量。在一些实施方案中,数码相机是标准清晰度相机。在其他实施方案中,数码相机是HD摄像机。在进一步的实施方案中,HD摄像机捕获具有至少约1280x约720像素或至少约1920x约1080像素的图像。在一些实施方案中,数码相机捕获彩色数字图像。在其他实施方案中,数码相机捕获灰度数字图像。在各种实施方案中,数字图像以任何合适的数字图像格式存储。作为非限制性实例,合适的数字图像格式包括联合图像专家组(JPEG)、JPEG 2000、可交换图像文件格式(Exif)、标记图像文件格式(TIFF)、RAW、便携式网络图形(PNG)、图形交换格式(GIF)、Windows<sup>®</sup>位图(BMP)、便携式像素图(PPM)、便携式灰度图(PGM)、便携式位图文件格式(PBM)和WebP。在各种实施方案中,数字图像以任何合适的数字视频格式存储。作为非限制性实例,合适的数字视频格式包括AVI、MPEG、

Apple® QuickTime®、MP4、AVCHD®、Windows Media®、DivX™、Flash Video、Ogg Theora、WebM和RealMedia。

#### 非暂时性计算机可读存储介质

[0114] 在一些实施方案中，本文公开的系统、平台、软件、网络和方法包括用程序编码的一个或多个非暂时性计算机可读存储介质，该程序包括可由任选的联网的数字处理设备的操作系统执行的指令。在进一步的实施方案中，计算机可读存储介质是数字处理设备的有形组件。在更进一步的实施方案中，计算机可读存储介质任选地可从数字处理设备移除。在一些实施方案中，作为非限制性实例，计算机可读存储介质包括CD-ROM、DVD、闪速存储器设备、固态存储器、磁盘驱动器、磁带驱动器、光盘驱动器、云计算系统和服务器等。在一些情况下，所述程序和指令在介质上永久地、基本上永久地、半永久地或非暂时地编码。

#### 计算机程序

[0115] 在一些实施方案中，本文公开的系统、平台、软件、网络和方法包括至少一个计算机程序。计算机程序包括可在数字处理设备的CPU中执行的指令序列，该指令序列被编写以进行指定的任务。鉴于本文提供的公开内容，本领域技术人员将认识到，计算机程序可用各种语言的各种版本来编写。在一些实施方案中，计算机程序包含一个指令序列。在一些实施方案中，计算机程序包含多个指令序列。在一些实施方案中，从一个位置提供计算机程序。在其他实施方案中，从多个位置提供计算机程序。在各个实施方案中，计算机程序包括一个或多个软件模块。在各种实施方案中，计算机程序部分或全部包括一个或多个web应用程序、一个或多个移动应用程序、一个或多个独立应用程序、一个或多个web浏览器插件、扩展、加载项或附加项或其组合。

#### Web应用程序

[0116] 在一些实施方案中，计算机程序包括web应用程序。鉴于本文提供的公开内容，本领域技术人员将认识到，在各种实施方案中，web应用程序利用一个或多个软件框架和一个或多个数据库系统。在一些实施方案中，基于诸如Microsoft® .NET或Ruby on Rails (RoR) 的软件框架创建web应用程序。在一些实施方案中，web应用程序利用一个或多个数据库系统，作为非限制性实例，该数据库系统包括关系数据库系统、非关系数据库系统、面向对象的数据库系统、关联数据库系统和XML数据库系统。在进一步的实施方案中，作为非限制性实例，合适的关系数据库系统包括Microsoft® SQL服务器、MySQL™和Oracle®。本领域技术人员还将认识到，在各种实施方案中，web应用程序以一种或多种语言的一个或多个版本编写。Web应用程序可以用一种或多种标记语言、表示定义语言、客户端脚本语言、服务器端编码语言、数据库查询语言或其组合来编写。在一些实施方案中，web应用程序在某种程度上以诸如超文本标记语言 (HTML)、可扩展超文本标记语言 (XHTML) 或可扩展标记语言 (XML) 的标记语言编写。在一些实施方案中，web应用程序在某种程度上以诸如级联样式表 (CSS) 的表示定义语言编写。在一些实施方案中，web应用程序在某种程度上以诸如异步 Javascript和XML (AJAX)、Flash® 动作脚本、Javascript或Silverlight® 的客户端脚本语言编写。在一些实施方案中，web应用程序在某种程度上以诸如Active Server Pages (ASP)、ColdFusion®、Perl、Java™、JavaServer Pages (JSP)、超文本预处理器 (PHP)、

Python<sup>™</sup>、Ruby、Tcl、Smalltalk、WebDNA<sup>®</sup>或Groovy的服务器端编码语言编写。在一些实施方案中，web应用程序在某种程度上以诸如结构化查询语言 (SQL) 的数据库查询语言编写。在一些实施方案中，web应用程序集成了诸如IBM<sup>®</sup> Lotus Domino<sup>®</sup>的企业服务器产品。在一些实施方案中，用于为艺术家提供允许艺术家上传信息和媒体文件的职业发展网络的网络应用程序包括媒体播放器元件。在各种进一步的实施方案中，媒体播放器元件利用许多合适的多媒体技术中的一种或多种，作为非限制性实例，该多媒体技术包括Adobe<sup>®</sup> Flash<sup>®</sup>、HTML 5、Apple<sup>®</sup> QuickTime<sup>®</sup>、Microsoft<sup>®</sup> Silverlight<sup>®</sup>、Java<sup>™</sup>和Unity<sup>®</sup>。移动应用程序

[0117] 在一些实施方案中，计算机程序包括提供给移动数字处理设备的移动应用程序。在一些实施方案中，移动应用程序在其制造时被提供给移动数字处理设备。在其他实施方案中，经由本文所述的计算机网络将移动应用程序提供给移动数字处理设备。

[0118] 鉴于本文提供的公开内容，使用本领域已知的硬件、语言和开发环境，通过本领域技术人员已知的技术创建移动应用程序。本领域技术人员将认识到，移动应用程序是用几种语言来编写的。作为非限制性实例，合适的编程语言包括C、C++、C#、Objective-C、Java<sup>™</sup>、Javascript、Pascal、Object Pascal、Python<sup>™</sup>、Ruby、VB.NET、WML和具有或不具有CSS的XHTML/HTML或其组合。

[0119] 合适的移动应用程序开发环境可从多个来源获得。作为非限制性实例，商业上可用的开发环境包括AirplaySDK、alcheMo、Appcelerator<sup>®</sup>、Celsius、Bedrock、Flash Lite、.NET Compact Framework、Rhomobile和WorkLight移动平台。其他开发环境可免费获得，作为非限制性实例，该其他开发环境包括Lazarus、MobiFlex、MoSync和Phonegap。此外，移动设备制造商分发软件开发工具包，作为非限制性实例，该软件开发工具包包括iPhone和iPad (iOS) SDK、Android<sup>™</sup> SDK、BlackBerry<sup>®</sup> SDK、BREW SDK、Palm<sup>®</sup> OS SDK、Symbian SDK、webOS SDK和Windows<sup>®</sup> Mobile SDK。

[0120] 本领域技术人员将认识到，多个商业论坛可用于分发移动应用程序，作为非限制性实例，该商业论坛包括Apple<sup>®</sup> App Store、Android<sup>™</sup> Market、BlackBerry<sup>®</sup> App World、用于Palm设备的App Store、App Catalog for webOS、Windows<sup>®</sup> Marketplace for Mobile、用于Nokia<sup>®</sup>设备的Ovi Store、Samsung<sup>®</sup> Apps和Nintendo<sup>®</sup> DSi Shop。

#### 独立应用程序

[0121] 在一些实施方案中，计算机程序包括独立应用程序，该独立应用程序是作为独立计算机进程运行的程序，而不是现有进程的附加项 (例如，不是插件)。本领域技术人员将认识到经常编译独立应用程序。编译器是将用编程语言编写的源代码转换为二进制目标代码如汇编语言或机器代码的计算机程序。作为非限制性实例，合适的编译编程语言包括C、C++、Objective-C、COBOL、Delphi、Eiffel、Java<sup>™</sup>、Lisp、Python<sup>™</sup>、Visual Basic和VB.NET或其组合。通常至少部分地执行编译以创建可执行程序。在一些实施方案中，计算机程序包括一个或多个可执行编译的应用程序。

#### 软件模块

[0122] 在各个实施方案中,本文公开的系统、平台、软件、网络和方法包括软件、服务器和数据库模块。鉴于本文提供的公开内容,使用本领域已知的机器、软件和语言,通过本领域技术人员已知的技术创建软件模块。本文公开的软件模块以多种方式实现。在各种实施方案中,软件模块包含文件、代码段、编程对象、编程结构或其组合。在进一步的各种实施方案中,软件模块包含多个文件、多个代码段、多个编程对象、多个编程结构或其组合。在各种实施方案中,作为非限制性实例,一个或多个软件模块包含web应用程序、移动应用程序和独立应用程序。在一些实施方案中,软件模块在一个计算机程序或应用程序中。在其他实施方案中,软件模块在多于一个计算机程序或应用程序中。在一些实施方案中,软件模块托管在一台机器上。在其他实施方案中,软件模块托管在多于一台机器上。在进一步的实施方案中,软件模块托管在云计算平台上。在一些实施方案中,软件模块托管在一个位置中的一个或多个机器上。在其他实施方案中,软件模块托管在多于一个位置中的一个或多个机器上。

[0123] 试剂盒

[0124] 所公开实施方案的装置和方法可包装成试剂盒。在一些实施方案中,试剂盒包括关于所述装置或方法的使用的书面说明。该书面材料可以是,例如,标签。该书面材料可以提供条件,例如,针对测定或进行测定的步骤。这些说明为用户提供使用装置和/或进行本文公开的方法和测定的最佳指导。

#### 实施方案

[0125] 以下非限制性实施方案提供了本发明的示例性实例,但并不限制本发明的范围。

[0126]

[0127] 实施方案1。在一些实施方案中,本文提供了在基底上原位合成化学文库的方法,所述化学文库包含多个分子,所述方法包括:

(a) 接收生物序列和多个合成步骤;

(b) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;

(c) 为每个图案化掩模指定至少一个单体;以及

(d) 将所述单体耦合至所述特征以形成分子;

(e) 其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤。

[0128] 实施方案2。根据实施方案1所述的方法,其中所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。

[0129] 实施方案3。根据实施方案1所述的方法,其中所述输入生物序列包含疾病相关的表位。

[0130] 实施方案4。根据实施方案1所述的方法,其中所述输入生物序列包含疾病相关的表位。

[0131] 实施方案5。根据实施方案1所述的方法,其中所述输入生物序列包含肽序列。

[0132] 实施方案6。根据实施方案1所述的方法,其中所述输入生物序列包含表位序列。

[0133] 实施方案7。根据实施方案1所述的方法,其中所述输入生物序列包含随机序列。

[0134] 实施方案8。根据实施方案1所述的方法,其进一步包括从所述输入生物序列中导

出单体的有序列表。

[0135] 实施方案9。根据实施方案8所述的方法,其中所述有序列表的大小为所述合成步骤的数目。

[0136] 实施方案10。根据实施方案8所述的方法,其中所述单体的有序列表包含所述输入生物序列。

[0137] 实施方案11。根据实施方案10所述的方法,其中所述单体的有序列表以相反的顺序包含所述输入生物序列。

[0138] 实施方案12。根据实施方案8所述的方法,其中所述分子为肽或核酸。

[0139] 实施方案13。根据实施方案8所述的方法,其中所述单体的有序列表包含氨基酸序列。

[0140] 实施方案14。根据实施方案8所述的方法,其中所述单体的有序列表包含核苷酸序列。

[0141] 实施方案15。根据实施方案1所述的方法,其中所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。

[0142] 实施方案16。根据实施方案1所述的方法,其中所述多个图案化掩模的数目为所述合成步骤的数目。

[0143] 实施方案17。根据实施方案1所述的方法,其中每个顺序图案化掩模中约20%至约50%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

[0144] 实施方案18。根据实施方案1所述的方法,其中每个顺序图案化掩模中约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

[0145] 实施方案19。根据实施方案1所述的方法,其中所述合成步骤基于光刻法。

[0146] 实施方案20。根据实施方案1所述的方法,其中所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。

[0147] 实施方案21。根据实施方案1所述的方法,其中所述文库中至少40%的分子、至少50%的分子、至少60%的分子、至少70%的分子、至少80%的分子或至少90%的分子是不同的。

[0148] 实施方案22。根据实施方案1所述的方法,其中所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

[0149] 实施方案23。根据实施方案1所述的方法,其中所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

[0150] 实施方案24。根据实施方案1所述的方法,其中所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。

[0151] 实施方案25。根据实施方案1所述的方法,其中所述文库包含等于所述生物序列长度的中值单体长度。

[0152] 实施方案26。根据实施方案1所述的方法,其中所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。

[0153] 实施方案27。根据实施方案1所述的方法,其中所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值单体长度。

[0154] 实施方案28。根据实施方案1所述的方法,其中所述基底选自阵列、晶片、载玻片和珠子。

[0155] 实施方案29。根据实施方案1所述的方法,其中所述合成的化学文库包含肽、核苷酸或其组合。

[0156] 实施方案30。根据实施方案29所述的方法,其中所述肽的长度为约5至约25个氨基酸。

[0157] 实施方案31。根据实施方案29所述的方法,其中所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。

[0158] 实施方案32。根据实施方案1所述的方法,其中利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。

[0159] 实施方案33。根据实施方案32所述的方法,其中所述表面间隔物为Cys-Gly-Pro-Gly-Xaa<sub>n</sub>-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-Cys。

[0160] 实施方案34。根据实施方案1所述的方法,其中利用能够用酯键环化的表面间隔物来合成所述化学文库。

[0161] 实施方案35。根据实施方案34所述的方法,其中所述酯键为同双官能的二-NHS酯键。

[0162] 实施方案36。根据实施方案34所述的方法,其中所述表面间隔物为Lys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-赖氨酸。

[0163] 实施方案37。根据实施方案1所述的方法,其中所述基底涂覆有亲水性单层。

[0164] 实施方案38。根据实施方案37所述的方法,其中所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。

[0165]

[0166] 实施方案39。根据实施方案37所述的方法,其中所述亲水性单层是均匀的。

[0167] 实施方案40。在一些实施方案中,本文提供了原位合成的化学文库,所述化学文库包含多个分子,其中所述合成使用图案化步骤以在基底上构建所述文库,所述合成包括:

(a) 接收生物序列和多个合成步骤;

(b) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;

(c) 为每个图案化掩模指定至少一个单体;以及

(d) 将所述单体耦合至所述特征以形成分子;其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤。

[0168] 实施方案41。根据实施方案40所述的化学文库,其中所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。

[0169] 实施方案42。根据实施方案40所述的化学文库,其中所述输入生物序列包括疾病

相关的表位、肽序列、表位序列和/或随机序列。

[0170] 实施方案43。根据实施方案40所述的化学文库,其进一步包括从所述输入生物序列中导出单体的有序列表。

[0171] 实施方案44。根据实施方案43所述的化学文库,其中所述有序列表的大小为所述合成步骤的数目。

[0172] 实施方案45。根据实施方案43所述的化学文库,其中所述单体的有序列表包含所述输入生物序列。

[0173] 实施方案46。根据实施方案43所述的化学文库,其中所述单体的有序列表以相反的顺序包含所述输入生物序列。

[0174] 实施方案47。根据实施方案40所述的化学文库,其中所述分子包含肽或核酸。

[0175] 实施方案48。根据实施方案43所述的化学文库,其中所述单体的有序列表包含氨基酸序列和/或核苷酸序列。

[0176] 实施方案49。根据实施方案40所述的化学文库,其中所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。

[0177] 实施方案50。根据实施方案40所述的化学文库,其中所述多个图案化掩模的数目为所述合成步骤的数目。

[0178] 实施方案51。根据实施方案40所述的化学文库,其中每个顺序图案化掩模中约20%至约50%或约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

[0179] 实施方案52。根据实施方案40所述的化学文库,其中所述合成步骤基于光刻法。

[0180] 实施方案53。根据实施方案40所述的化学文库,其中所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。

[0181] 实施方案54。根据实施方案40所述的化学文库,其中所述文库中至少40%、至少50%、至少60%、至少70%、至少80%或至少90%的分子是不同的。

[0182] 实施方案55。根据实施方案40所述的化学文库,其中所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

[0183] 实施方案56。根据实施方案40所述的化学文库,其中所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

[0184] 实施方案57。根据实施方案40所述的化学文库,其中所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。

[0185] 实施方案58。根据实施方案40所述的化学文库,其中所述文库包含等于所述生物序列长度的中值单体长度。

[0186] 实施方案59。根据实施方案40所述的化学文库,其中所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。

[0187] 实施方案60。根据实施方案40所述的化学文库,其中所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、

170%、180%、190%或200%的中值单体长度。

[0188] 实施方案61。根据实施方案40所述的化学文库,其中所述基底选自阵列、晶片、载玻片和珠子。

[0189] 实施方案62。根据实施方案40所述的化学文库,其中所述合成的化学文库包含肽、核苷酸或其组合。

[0190] 实施方案63。根据实施方案62所述的化学文库,其中所述肽的长度为约5至约25个氨基酸。

[0191] 实施方案64。根据实施方案63所述的化学文库,其中所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。

[0192] 实施方案65。根据实施方案40所述的化学文库,其中利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。

[0193] 实施方案66。根据实施方案65所述的化学文库,其中所述表面间隔物为Cys-Gly-Pro-Gly-Xaa<sub>n</sub>-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-Cys。

[0194] 实施方案67。根据实施方案40所述的化学文库,其中利用能够用酯键环化的表面间隔物来合成所述化学文库。

[0195] 实施方案68。根据实施方案67所述的化学文库,其中所述酯键为同双官能的二-NHS酯键。

[0196] 实施方案69。根据实施方案68所述的化学文库,其中所述表面间隔物为Lys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-赖氨酸。

[0197] 实施方案70。根据实施方案40所述的化学文库,其中所述基底涂覆有亲水性单层。

[0198] 实施方案71。根据实施方案70所述的化学文库,其中所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。

[0199] 实施方案72。根据实施方案70所述的化学文库,其中所述亲水性单层是均匀的。

[0200] 实施方案73。在一些实施方案中,本文提供了用于模拟在基底上原位合成化学文库的计算系统,所述化学文库包含多个分子,所述计算系统包含:

(a) 处理器和存储器;

(b) 计算机程序,其包括可由所述处理器执行的指令,所述计算机程序包含:

(1) 接收模块,其被配置为接收生物序列和多个合成步骤;

(2) 模拟模块,其被配置为:(i) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;(ii) 为每个图案化掩模指定至少一个单体;以及(iii) 将所述单体耦合至所述特征以形成分子;其中(i)、(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤。

[0201] 实施方案74。根据实施方案73所述的系统,其中所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。

[0202] 实施方案75。根据实施方案73所述的系统,其中所述输入生物序列包括疾病相关的表位、肽序列、表位序列和/或随机序列。

[0203] 实施方案76。根据实施方案73所述的系统,其进一步包括从所述输入生物序列中

导出单体的有序列表。

[0204] 实施方案77。根据实施方案76所述的系统,其中所述有序列表的大小为所述合成步骤的数目。

[0205] 实施方案78。根据实施方案76所述的系统,其中所述单体的有序列表包含所述输入生物序列。

[0206] 实施方案79。根据实施方案78所述的系统,其中所述单体的有序列表以相反的顺序包含所述输入生物序列。

[0207] 实施方案80。根据实施方案73所述的系统,其中所述分子包含肽或核酸。

[0208] 实施方案81。根据实施方案73所述的系统,其中所述单体的有序列表包含氨基酸序列和/或核苷酸序列。

[0209] 实施方案82。根据实施方案73所述的系统,其中所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。

[0210] 实施方案83。根据实施方案73所述的系统,其中所述多个图案化掩模的数目为所述合成步骤的数目。

[0211] 实施方案84。根据实施方案73所述的系统,其中每个顺序图案化掩模中约20%至约50%或约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

[0212] 实施方案85。根据实施方案73所述的系统,其中所述合成步骤基于光刻法。

[0213] 实施方案86。根据实施方案73所述的系统,其中所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。

[0214] 实施方案87。根据实施方案73所述的系统,其中所述文库中至少40%、至少50%、至少60%、至少70%、至少80%或至少90%的分子是不同的。

[0215] 实施方案88。根据实施方案73所述的系统,其中所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

[0216] 实施方案89。根据实施方案73所述的系统,其中所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

[0217] 实施方案90。根据实施方案73所述的系统,其中所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。

[0218] 实施方案91。根据实施方案73所述的系统,其中所述文库包含等于所述生物序列长度的中值单体长度。

[0219] 实施方案92。根据实施方案73所述的系统,其中所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。

[0220] 实施方案93。根据实施方案73所述的系统,其中所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值单体长度。

[0221] 实施方案94。根据实施方案73所述的系统,其中所述基底选自阵列、晶片、载玻片

和珠子。

[0222] 实施方案95。根据实施方案73所述的系统,其中所述合成的化学文库包含肽、核苷酸或其组合。

[0223] 实施方案96。根据实施方案95所述的系统,其中所述肽的长度为约5至约25个氨基酸。

[0224] 实施方案97。根据实施方案96所述的系统,其中所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。

[0225] 实施方案98。根据实施方案73所述的系统,其中利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。

[0226] 实施方案99。根据实施方案98所述的系统,其中所述表面间隔物为Cys-Gly-Pro-Gly-Xaa<sub>n</sub>-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-Cys。

[0227] 实施方案100。根据实施方案73所述的系统,其中利用能够用酯键环化的表面间隔物来合成所述化学文库。

[0228] 实施方案101。根据实施方案100所述的系统,其中所述酯键为同双官能的二-NHS酯键。

[0229] 实施方案102。根据实施方案101所述的系统,其中所述表面间隔物为Lys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-赖氨酸。

[0230] 实施方案103。根据实施方案73所述的系统,其中所述基底涂覆有亲水性单层。

[0231] 实施方案104。根据实施方案103所述的系统,其中所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。

[0232] 实施方案105。根据实施方案103所述的系统,其中所述亲水性单层是均匀的。

[0233] 实施方案106。在一些实施方案中,本文提供了原位合成肽阵列的方法,所述方法包括:

(a) 接收输入氨基酸序列;

(b) 确定多个合成步骤;

(c) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

(d) 为每个图案化掩模指定至少一个单体;以及

(e) 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0234] 实施方案107。根据实施方案106所述的方法,其中所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。

[0235] 实施方案108。根据实施方案106所述的方法,其中所述输入生物序列包括疾病相关的表位、肽序列或表位序列。

[0236] 实施方案109。根据实施方案106所述的方法,其进一步包括从所述输入序列中导出单体的有序列表。

[0237] 实施方案110。根据实施方案109所述的方法,其中所述有序列表的大小为所述合

成步骤的数目。

[0238] 实施方案111。根据实施方案109所述的方法,其中所述单体的有序列表包含所述输入序列。

[0239] 实施方案112。根据实施方案111所述的方法,其中所述单体的有序列表以相反的顺序包含所述输入序列。

[0240] 实施方案113。根据实施方案109所述的方法,其中所述单体的有序列表包含氨基酸序列。

[0241] 实施方案114。根据实施方案106所述的方法,其中所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。

[0242] 实施方案115。根据实施方案106所述的方法,其中所述多个图案化掩模的数目为所述合成步骤的数目。

[0243] 实施方案116。根据实施方案106所述的方法,其中每个顺序图案化掩模中约20%至约50%或约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

[0244] 实施方案117。根据实施方案106所述的方法,其中所述合成步骤基于光刻法。

[0245] 实施方案118。根据实施方案106所述的方法,其中所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。

[0246] 实施方案119。根据实施方案106所述的方法,其中所述阵列上至少40%、至少50%、至少60%、至少70%、至少80%或至少90%的肽是不同的。

[0247] 实施方案120。根据实施方案106所述的方法,其中所述阵列上至少50%的肽的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

[0248] 实施方案121。根据实施方案106所述的方法,其中所述阵列上至少50%的肽的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

[0249] 实施方案122。根据实施方案106所述的方法,其中所述阵列上的所述肽包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。

[0250] 实施方案123。根据实施方案106所述的方法,其中所述阵列包含等于所述输入序列长度的中值肽长度。

[0251] 实施方案124。根据实施方案106所述的方法,其中所述阵列包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值肽长度。

[0252] 实施方案125。根据实施方案106所述的方法,其中所述阵列包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值肽长度。

[0253] 实施方案126。根据实施方案106所述的方法,其中所述肽的长度为约5至约25个氨基酸。

[0254] 实施方案127。根据实施方案106所述的方法,其中所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。

[0255] 实施方案128。根据实施方案106所述的方法,其中利用能够在氧化条件下环化的表面间隔物来合成所述肽阵列。

[0256] 实施方案129。根据实施方案128所述的方法,其中所述表面间隔物为Cys-Gly-Pro-Gly-Xaa<sub>n</sub>-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-Cys。

[0257] 实施方案130。根据实施方案106所述的方法,其中利用能够用酯键环化的表面间隔物来合成所述肽阵列。

[0258] 实施方案131。根据实施方案130所述的方法,其中所述酯键为同双官能的二-NHS酯键。

[0259] 实施方案132。根据实施方案130所述的方法,其中所述表面间隔物为Lys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-赖氨酸。

[0260] 实施方案133。根据实施方案106所述的方法,其中所述肽阵列涂覆有亲水性单层。

[0261] 实施方案134。根据实施方案132所述的方法,其中所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。

[0262] 实施方案135。根据实施方案132所述的方法,其中所述亲水性单层是均匀的。

[0263] 实施方案136。在一些实施方案中,本文包括了阵列,其包含所述阵列上的多个原位合成的肽,所述肽由多个图案化掩模产生,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠。

[0264] 实施方案137。在一些实施方案中,本文包括了用于表征针对至少一种蛋白质靶标的抗体结合的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(a)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(c) 使用步骤(b)的所述比对得分表征针对所述至少一种蛋白质靶标的所述抗体的结合。

[0265] 实施方案138。根据实施方案137所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

[0266] 实施方案139。根据实施方案137所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

[0267] 实施方案140。根据实施方案137所述的方法,其中所述竞争肽包含生物样品。

[0268] 实施方案141。根据实施方案137所述的方法,其中所述生物样品为血清。

[0269] 实施方案142。根据实施方案137所述的方法,其中所述竞争肽衍生自所述靶蛋白。

[0270] 实施方案143。根据实施方案142所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

[0271] 实施方案144。根据实施方案137所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

[0272] 实施方案145。根据实施方案144所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

[0273] 实施方案146。根据实施方案137所述的方法,其中所述竞争肽包含生物样品和实施方案142至145中任一项的肽。

[0274] 实施方案147。根据实施方案137所述的方法,其中所述肽阵列包含至少1000、至少10,000、至少100,000或至少1,000,000种独特的肽。

[0275] 实施方案148。根据实施方案137所述的方法,其中所述肽阵列是原位合成的。

[0276] 实施方案149。根据实施方案148所述的方法,其中所述肽阵列通过以下方法合成:

i. 接收输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0277] 实施方案150。根据实施方案137所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

[0278] 实施方案151。根据实施方案137所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

[0279] 实施方案152。根据实施方案137所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的所述比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的趋势。

[0280] 实施方案153。根据实施方案137所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自权利要求169中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0281] 实施方案154。根据实施方案137所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自权利要求169中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0282] 实施方案155。在一些实施方案中,本文公开了用于鉴定靶蛋白中抗体表位的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(a)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(c) 确定步骤(a)的所述单独的肽中的保守氨基酸,以鉴定保守结合肽基序,并将所述单独的基序与所述至少一种靶蛋白比对,以鉴定所述靶蛋白的至少一种抗体表位。

[0283] 实施方案156。根据实施方案155所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

[0284] 实施方案157。根据实施方案155所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,其中所述预定阈值是在存在竞争肽的情况下结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

[0285] 实施方案158。根据实施方案155所述的方法,其中所述竞争肽包含生物样品。

[0286] 实施方案159。根据实施方案155所述的方法,其中所述生物样品为血清。

[0287] 实施方案160。根据实施方案155所述的方法,其中所述竞争肽衍生自所述靶蛋白。

[0288] 实施方案161。根据实施方案160所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

[0289] 实施方案162。根据实施方案155所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

[0290] 实施方案163。根据实施方案162所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

[0291] 实施方案164。根据实施方案155所述的方法,其中所述竞争肽包含生物样品和实施方案160至163中任一项的肽。

[0292] 实施方案165。根据实施方案155所述的方法,其中所述肽阵列包含至少1000、至少10,000、至少100,000或至少1,000,000种独特的肽。

[0293] 实施方案166。根据实施方案155所述的方法,其中所述肽阵列是原位合成的。

[0294] 实施方案167。根据实施方案166所述的方法,其中所述肽阵列通过以下方法合成:

i. 接收输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0295] 实施方案168。根据实施方案155所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

[0296] 实施方案169。根据实施方案155所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观K<sub>d</sub>。

[0297] 实施方案170。根据实施方案155所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的比对得分进行分级,以确定每种抗体与所述蛋白质靶标

结合的倾向。

[0298] 实施方案171。根据实施方案155所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自权利要求190中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0299] 实施方案172。根据实施方案155所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自权利要求190中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0300] 实施方案173。根据实施方案155所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

[0301] 实施方案174。根据实施方案173所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

[0302] 实施方案175。在一些实施方案中,本文公开了用于表征靶蛋白中的抗体结合区的方法,所述方法包括:

(a) 在存在和不存在多种竞争肽的情况下,使第一肽阵列与所述抗体接触以获得一种或多种单独肽,其中所鉴定的一种或多种单独肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第一预定阈值内。

(b) 使用选自步骤(a)中至少一种所述单独的肽的输入肽序列、衍生自步骤(a)中所述单独的肽的比对的保守基序或衍生自步骤(a)中所述单独的肽的比对的比对基序来产生第二肽阵列,所述第二肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧邻的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0303] 实施方案176。根据实施方案175所述的方法,其中所述竞争肽包含生物样品。

[0304] 实施方案177。根据实施方案175所述的方法,其中所述生物样品为血清。

[0305] 实施方案178。根据实施方案175所述的方法,其中所述竞争肽衍生自所述靶蛋白。

[0306] 实施方案179。根据实施方案178所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

[0307] 实施方案180。根据实施方案175所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

[0308] 实施方案181。根据实施方案180所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

[0309] 实施方案182。根据实施方案175所述的方法,其中所述竞争肽包含生物样品和实

施方案178至181中任一项的肽。

[0310] 实施方案183。根据实施方案175所述的方法,其中所述肽阵列包含至少1000、至少10,000、至少100,000或至少1,000,000种独特的肽。

[0311] 实施方案184。根据实施方案175所述的方法,其中所述肽阵列是原位合成的。

[0312] 实施方案185。根据实施方案175所述的方法,其中所述第一肽阵列通过以下方法合成:

i. 接收输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0313] 实施方案186。根据实施方案175所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

[0314] 实施方案187。根据实施方案175所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

[0315] 实施方案188。根据实施方案175所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。

[0316] 实施方案189。根据实施方案175所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自实施方案175中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0317] 实施方案190。根据实施方案175所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自权利要求213中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0318] 实施方案191。根据实施方案175所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

[0319] 实施方案192。根据实施方案191所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

[0320] 实施方案193。根据实施方案175所述的方法,其中所述第一预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

[0321] 实施方案194。根据实施方案175所述的方法,其中所述第二预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

[0322] 实施方案195。根据实施方案175所述的方法,其中所述第一预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

[0323] 实施方案196。根据实施方案175所述的方法,其中所述第二预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

[0324] 实施方案197。根据实施方案175所述的方法,其中所述抗体结合区为所述靶蛋白的线性表位。

[0325] 实施方案198。根据实施方案175所述的方法,其中所述抗体结合区为所述靶区域的结构表位。

[0326] 实施方案199。根据实施方案198所述的方法,其中用选自实施方案213的步骤(a)中至少一种所述单独的肽的另外的肽来重复权利要求175中的步骤(b)至步骤(d)。

[0327] 实施方案200。在一些实施方案中,本文公开了用于鉴定抗体的靶蛋白的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使第一肽阵列与一种或多种浓度的所述抗体接触以获得一个或多个输入氨基酸序列,其中所鉴定的输入氨基酸序列展现出在存在所述多种竞争肽的情况下的结合信号,所述结合信号处于在不存在多种所述竞争肽的情况下的所述结合信号的第一预定阈值内。

(b) 使用选自步骤(a)中至少一种所述单独的肽的一个或多个输入氨基酸序列、衍生自步骤(a)中所述单独的肽的比对的保守基序或衍生自步骤(a)中所述单独的肽的比对的比对基序来获得一种或多种二级肽阵列,所述一种或多种二级肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列;

(c) 在存在和不存在所述多种竞争肽的情况下,使每个所述二级肽阵列与所述抗体接触以获得一组肽序列,其中所鉴定的一组肽序列展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第二预定阈值内。

(d) 使所述一组肽序列彼此比对以获得至少一种预测性结合基序;以及

(e) 将所述预测性结合基序作为针对蛋白质数据库的搜索标准进行比对,从而基于所述蛋白质数据库搜索结果得分来鉴定所述抗体的靶蛋白。

[0328] 实施方案201。根据实施方案200所述的方法,其中所述竞争肽包含生物样品。

[0329] 实施方案202。根据实施方案200所述的方法,其中所述生物样品为血清。

[0330] 实施方案203。根据实施方案200所述的方法,其中所述竞争肽衍生自所述靶蛋白。

[0331] 实施方案205。根据实施方案203所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

[0332] 实施方案206。根据实施方案200所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

[0333] 实施方案207。根据实施方案206所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

[0334] 实施方案208。根据实施方案200所述的方法,其中所述竞争肽包含生物样品和实施方案203至208中任一项的肽。

[0335] 实施方案209。根据实施方案200所述的方法,其中所述肽阵列包含至少1000、至少10,000、至少100,000或至少1,000,000种独特的肽。

[0336] 实施方案210。根据实施方案200所述的方法,其中所述肽阵列是原位合成的。

[0337] 实施方案211。根据实施方案200所述的方法,其中所述第一肽阵列通过以下方法合成:

i. 接受输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模都被指定为所述基底上每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0338] 实施方案212。根据实施方案200所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

[0339] 实施方案213。根据实施方案200所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

[0340] 实施方案214。根据实施方案200所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。

[0341] 实施方案215。根据实施方案200所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自权利要求241中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0342] 实施方案216。根据实施方案200所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自权利要求241中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0343] 实施方案217。根据实施方案200所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

[0344] 实施方案218。根据实施方案217所述的方法,其中所述蛋白质数据库为蛋白质组

数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

[0345] 实施方案219。根据实施方案200所述的方法,其中所述第一预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

[0346] 实施方案220。根据实施方案200所述的方法,其中所述预定阈值是在存在竞争肽的情况下结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

[0347] 实施方案221。在一些实施方案中,本文公开了用于确定抗体与至少一种蛋白质靶标结合倾向的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 使步骤(a)的所述单独的肽与第一蛋白质靶标比对,其中步骤(a)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(c) 重复步骤(a)的所述单独的肽与至少一种另外的蛋白质靶标的比对,其中步骤(a)的所述单独的肽与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(d) 比较步骤(b)和(c)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

[0348] 实施方案222。根据实施方案221所述的方法,其中所述竞争肽包含生物样品。

[0349] 实施方案223。根据实施方案222所述的方法,其中所述生物样品为血清。

[0350] 实施方案224。根据实施方案221所述的方法,其中所述竞争肽衍生自所述靶蛋白。

[0351] 实施方案225。根据实施方案221所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

[0352] 实施方案226。根据实施方案225所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

[0353] 实施方案227。根据实施方案221所述的方法,其中所述竞争肽包含生物样品和实施方案224至226中任一项的肽。

[0354] 实施方案228。根据实施方案221所述的方法,其中所述肽阵列包含至少1000、至少10,000、至少100,000或至少1,000,000种独特的肽。

[0355] 实施方案229。根据实施方案221所述的方法,其中所述肽阵列是原位合成的。

[0356] 实施方案230。根据实施方案221所述的方法,其中所述肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧邻的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(b)和(c)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0357] 实施方案231。根据实施方案221所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

[0358] 实施方案232。根据实施方案221所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

[0359] 实施方案233。根据实施方案221所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的所述比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。

[0360] 实施方案234。根据实施方案221所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自权利要求264中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0361] 实施方案235。根据实施方案221所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自权利要求264中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0362] 实施方案236。根据实施方案221所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

[0363] 实施方案237。根据实施方案236所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

[0364] 实施方案238。根据实施方案221所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

[0365] 实施方案239。根据实施方案221所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

[0366] 实施方案240。在一些实施方案中,本文公开了用于确定抗体与至少一种蛋白质靶标结合倾向的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使第一肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 对步骤(a)的所述一种或多种单独的肽进行比对以获得至少一种预测性靶基序;

(c) 将所述至少一种预测性靶基序与第一蛋白质靶标比对,其中步骤(a)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(d) 重复步骤(b)的至少一种预测性靶基序与至少一种另外的蛋白质靶标的比对,其中步骤(b)的所述至少一种预测性靶基序与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(e) 比较步骤(c)和(d)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合

的相对倾向。

[0367] 实施方案241。根据实施方案240所述的方法,其中所述竞争肽包含生物样品。

[0368] 实施方案242。根据实施方案240所述的方法,其中所述生物样品为血清。

[0369] 实施方案243。根据实施方案240所述的方法,其中所述竞争肽衍生自所述靶蛋白。

[0370] 实施方案244。根据实施方案243所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

[0371] 实施方案245。根据实施方案240所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

[0372] 实施方案246。根据实施方案245所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

[0373] 实施方案247。根据实施方案240所述的方法,其中所述竞争肽包含生物样品和实施方案243至246中任一项的肽。

[0374] 实施方案248。根据实施方案240所述的方法,其中所述肽阵列包含至少1000、至少10,000、至少100,000或至少1,000,000种独特的肽。

[0375] 实施方案249。根据实施方案240所述的方法,其中所述肽阵列是原位合成的。

[0376] 实施方案250。根据实施方案240所述的方法,其中所述肽阵列通过以下方法合成:

i. 接收输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧邻的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

[0377] 实施方案251。根据实施方案240所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

[0378] 实施方案252。根据实施方案240所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观K<sub>d</sub>。

[0379] 实施方案253。根据实施方案240所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对使每种抗体获得的所述比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。

[0380] 实施方案254。根据实施方案240所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自权利要求287中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0381] 实施方案255。根据实施方案240所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自权利要求287中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

[0382] 实施方案256。根据实施方案255所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

[0383] 实施方案257。根据实施方案240所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

[0384] 实施方案258。根据实施方案240所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

[0385] 实施方案259。根据实施方案240所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

[0386] 实施方案260。在一些实施方案中,本文公开了用于表征针对至少一种蛋白质靶标的抗体结合的试剂盒和系统,所述试剂盒和系统包括:

(a) 提供肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(d) 为所述用户提供说明以将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(c)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(e) 为所述用户提供说明以使用步骤(d)的所述比对得分表征针对所述至少一种蛋白质靶标的所述抗体的结合。

[0387] 实施方案261。在一些实施方案中,本文公开了用于鉴定靶蛋白中的抗体表位的试剂盒和系统,所述试剂盒和系统包括:

(a) 提供肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(d) 为所述用户提供说明以将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(c)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(e) 为所述用户提供说明以确定步骤(c)的所述单独的肽中的保守氨基酸,以鉴定保守结合肽基序,并将所述单独的基序与所述至少一种靶蛋白比对,以鉴定所述靶蛋白的至少一种抗体表位。

[0388] 实施方案262。在一些实施方案中,本文公开了用于表征靶蛋白中的抗体结合区的

试剂盒和系统,所述试剂盒和系统包括:

(a) 提供第一肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在所述多种竞争肽的情况下,使第一肽阵列与抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第一预定阈值内。

(d) 为用户提供说明以使用选自步骤(c)中至少一种所述单独的肽的输入肽序列、衍生自步骤(c)中所述单独的肽的比对的保守基序或衍生自步骤(c)中所述单独的肽的比对的比对基序来产生第二肽阵列,所述第二肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧邻的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列;

(e) 为所述用户提供说明以使所述第二肽阵列与所述抗体接触以鉴定第二组肽;以及

(f) 为所述用户提供说明以在存在所述多种竞争肽的情况下使所述第二肽阵列与所述抗体接触,并鉴定来自步骤(e)的第二组单独的肽,所述第二组单独的肽展现出处于在步骤(e)中所述结合信号的第二预定阈值内的结合信号;以及

(g) 为用户提供说明以将所述第二组单独的肽与所述靶蛋白比对,并鉴定所述靶蛋白中与所鉴定的第二组单独的肽比对的区域,从而表征所述靶蛋白中的抗体结合区。

[0389] 实施方案263。在一些实施方案中,本文公开了用于确定抗体与至少一种蛋白质靶标结合倾向的试剂盒和系统,所述试剂盒和系统包括:

(a) 提供肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(d) 为所述用户提供说明以使步骤(c)的所述单独的肽与第一蛋白质靶标比对,其中步骤(c)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(e) 为所述用户提供说明以重复步骤(c)的单独的肽与至少一种另外的蛋白质靶标的比对,其中步骤(c)的所述单独的肽与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(f) 为所述用户提供说明以比较步骤(c)和(d)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

[0390] 实施方案264。在一些实施方案中,本文公开了用于确定抗体与至少一种蛋白质靶标结合倾向的试剂盒和系统,所述试剂盒和系统包括:

(a) 提供第一肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述第一肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(d) 为所述用户提供说明以对步骤(c)的所述一种或多种单独的肽进行比对以获得至少一种预测性靶基序;

(e) 为所述用户提供说明以将所述至少一种预测性靶基序与第一蛋白质靶标比对,其中步骤(c)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(f) 为所述用户提供说明以重复步骤(e)的至少一种预测性靶基序与至少一种另外的蛋白质靶标的比对,其中步骤(e)的所述至少一种预测性靶基序与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(g) 为所述用户提供说明以比较步骤(c)和(d)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

本发明提供了包括但不限于以下实施方式:

1. 一种在基底上原位合成化学文库的方法,所述化学文库包含多个分子,所述方法包括:

(a) 接收生物序列和多个合成步骤;

(b) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;

(c) 为每个图案化掩模指定至少一个单体;以及

(d) 将所述单体耦合至所述特征以形成分子;

其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤。

2. 根据实施方式1所述的方法,其中所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。

3. 根据实施方式1所述的方法,其中所述输入生物序列包含疾病相关的表位。

4. 根据实施方式1所述的方法,其中所述输入生物序列包含肽序列。

5. 根据实施方式1所述的方法,其中所述输入生物序列包含表位序列。

6. 根据实施方式1所述的方法,其中所述输入生物序列包含随机序列。

7. 根据实施方式1所述的方法,其进一步包括从所述输入生物序列中导出单体的有序列表。

8. 根据实施方式7所述的方法,其中所述有序列表的大小为所述合成步骤的数目。

9. 根据实施方式7所述的方法,其中所述单体的有序列表包含所述输入生物序列。

10. 根据实施方式9所述的方法,其中所述单体的有序列表以相反的顺序包含所述输入生物序列。

11. 根据实施方式7所述的方法,其中所述分子为肽或核酸。

12. 根据实施方式7所述的方法,其中所述单体的有序列表包含氨基酸序列。

13. 根据实施方式7所述的方法,其中所述单体的有序列表包含核苷酸序列。

14. 根据实施方式1所述的方法,其中所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。

15. 根据实施方式1所述的方法,其中所述多个图案化掩模的数目为所述合成步骤的数目。

16. 根据实施方式1所述的方法,其中每个顺序图案化掩模中约20%至约50%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

17. 根据实施方式1所述的方法,其中每个顺序图案化掩模中约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

18. 根据实施方式1所述的方法,其中所述合成步骤基于光刻法。

19. 根据实施方式1所述的方法,其中所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。

20. 根据实施方式1所述的方法,其中所述文库中至少40%的分子是不同的。

21. 根据实施方式1所述的方法,其中所述文库中至少50%的分子是不同的。

22. 根据实施方式1所述的方法,其中所述文库中至少60%的分子是不同的。

23. 根据实施方式1所述的方法,其中所述文库中至少70%的分子是不同的。

24. 根据实施方式1所述的方法,其中所述文库中至少80%的分子是不同的。

25. 根据实施方式1所述的方法,其中所述文库中至少90%的分子是不同的。

26. 根据实施方式1所述的方法,其中所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

27. 根据实施方式1所述的方法,其中所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

28. 根据实施方式1所述的方法,其中所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。

29. 根据实施方式1所述的方法,其中所述文库包含等于所述生物序列长度的中值单体长度。

30. 根据实施方式1所述的方法,其中所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。

31. 根据实施方式1所述的方法,其中所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值单体长度。

32. 根据实施方式1所述的方法,其中所述基底选自阵列、晶片、载玻片和珠子。

33. 根据实施方式1所述的方法,其中所述合成的化学文库包含肽、核苷酸或其组合。

34. 根据实施方式33所述的方法,其中所述肽的长度为约5至约25个氨基酸。

35. 根据实施方式34所述的方法,其中所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。

36. 根据实施方式1所述的方法,其中利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。

37. 根据实施方式36所述的方法,其中所述表面间隔物为Cys-Gly-Pro-Gly-Xaa<sub>n</sub>-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-Cys。

38. 根据实施方式1所述的方法,其中利用能够用酯键环化的表面间隔物来合成所述化学文库。

39. 根据实施方式38所述的方法,其中所述酯键为同双官能的二-NHS酯键。

40. 根据实施方式38所述的方法,其中所述表面间隔物为Lys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-赖氨酸。

41. 根据实施方式1所述的方法,其中所述基底涂覆有亲水性单层。

42. 根据实施方式41所述的方法,其中所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。

43. 根据实施方式41所述的方法,其中所述亲水性单层是均匀的。

44. 一种原位合成的化学文库,所述化学文库包含多个分子,其中所述合成使用图案化步骤以在基底上构建所述文库,所述合成包括:

(a) 接收生物序列和多个合成步骤;

(b) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;

(c) 为每个图案化掩模指定至少一个单体;以及

(d) 将所述单体耦合至所述特征以形成分子;

其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤。

45. 根据实施方式44所述的文库,其中所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。

46. 根据实施方式44所述的文库,其中所述输入生物序列包含疾病相关的表位。

47. 根据实施方式44所述的文库,其中所述输入生物序列包含肽序列。

48. 根据实施方式44所述的文库,其中所述输入生物序列包含表位序列。

49. 根据实施方式44所述的文库,其中所述输入生物序列包含随机序列。

50. 根据实施方式44所述的文库,其进一步包括从所述输入生物序列中导出单体的有序列表。

51. 根据实施方式50所述的文库,其中所述有序列表的大小为所述合成步骤的数目。

52. 根据实施方式50所述的文库,其中所述单体的有序列表包含所述输入生物序列。

53. 根据实施方式50所述的文库,其中所述单体的有序列表以相反的顺序包含所述输入生物序列。

54. 根据实施方式44所述的文库,其中所述分子包含肽或核酸。
55. 根据实施方式50所述的文库,其中所述单体的有序列表包含氨基酸序列。
56. 根据实施方式50所述的文库,其中所述单体的有序列表包含核苷酸序列。
57. 根据实施方式44所述的文库,其中所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。
58. 根据实施方式44所述的文库,其中所述多个图案化掩模的数目为所述合成步骤的数目。
59. 根据实施方式44所述的文库,其中每个顺序图案化掩模中约20%至约50%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。
60. 根据实施方式44所述的文库,其中每个顺序图案化掩模中约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。
61. 根据实施方式44所述的文库,其中所述合成步骤基于光刻法。
62. 根据实施方式44所述的文库,其中所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。
63. 根据实施方式44所述的文库,其中所述文库中至少40%的分子是不同的。
64. 根据实施方式44所述的文库,其中所述文库中至少50%的分子是不同的。
65. 根据实施方式44所述的文库,其中所述文库中至少60%的分子是不同的。
66. 根据实施方式44所述的文库,其中所述文库中至少70%的分子是不同的。
67. 根据实施方式44所述的文库,其中所述文库中至少80%的分子是不同的。
68. 根据实施方式44所述的文库,其中所述文库中至少90%的分子是不同的。
69. 根据实施方式44所述的文库,其中所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。
70. 根据实施方式44所述的文库,其中所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。
71. 根据实施方式44所述的文库,其中所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。
72. 根据实施方式44所述的文库,其中所述文库包含等于所述生物序列长度的中值单体长度。
73. 根据实施方式44所述的文库,其中所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。
74. 根据实施方式44所述的文库,其中所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值单体长度。
75. 根据实施方式44所述的文库,其中所述基底选自阵列、晶片、载玻片和珠子。
76. 根据实施方式44所述的文库,其中所述合成的化学文库包含肽、核苷酸或其组合。
77. 根据实施方式76所述的文库,其中所述肽的长度为约5至约25个氨基酸。

78. 根据实施方式77所述的文库,其中所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。

79. 根据实施方式44所述的文库,其中利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。

80. 根据实施方式79所述的文库,其中所述表面间隔物为Cys-Gly-Pro-Gly-Xaa<sub>n</sub>-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-Cys。

81. 根据实施方式44所述的文库,其中利用能够用酯键环化的表面间隔物来合成所述化学文库。

82. 根据实施方式81所述的文库,其中所述酯键为同双官能的二-NHS酯键。

83. 根据实施方式82所述的文库,其中所述表面间隔物为Lys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-赖氨酸。

84. 根据实施方式44所述的文库,其中所述基底涂覆有亲水性单层。

85. 根据实施方式84所述的文库,其中所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。

86. 根据实施方式84所述的文库,其中所述亲水性单层是均匀的。

87. 一种用于模拟在基底上原位合成化学文库的计算系统,所述化学文库包含多个分子,所述计算系统包含:

(a) 处理器和存储器;

(b) 计算机程序,其包括可由所述处理器执行的指令,所述计算机程序包含:

(1) 接收模块,其被配置为接收生物序列和多个合成步骤;

(2) 模拟模块,其被配置为:(i) 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠;(ii) 为每个图案化掩模指定至少一个单体;以及(iii) 将所述单体耦合至所述特征以形成分子;

其中(i)、(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤。

88. 根据实施方式87所述的系统,其中所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。

89. 根据实施方式87所述的系统,其中所述输入生物序列包含疾病相关的表位。

90. 根据实施方式87所述的系统,其中所述输入生物序列包含肽序列。

91. 根据实施方式87所述的系统,其中所述输入生物序列包含表位序列。

92. 根据实施方式87所述的系统,其中所述输入生物序列包含随机序列。

93. 根据实施方式87所述的系统,其进一步包括从所述输入生物序列中导出单体的有序列表。

94. 根据实施方式93所述的系统,其中所述有序列表的大小为所述合成步骤的数目。

95. 根据实施方式93所述的系统,其中所述单体的有序列表包含所述输入生物序列。

96. 根据实施方式95所述的系统,其中所述单体的有序列表以相反的顺序包含所

述输入生物序列。

97. 根据实施方式87所述的系统,其中所述分子包含肽或核酸。

98. 根据实施方式87所述的系统,其中所述单体的有序列表包含氨基酸序列。

99. 根据实施方式87所述的系统,其中所述单体的有序列表包含核苷酸序列。

100. 根据实施方式87所述的系统,其中所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。

101. 根据实施方式87所述的系统,其中所述多个图案化掩模的数目为所述合成步骤的数目。

102. 根据实施方式87所述的系统,其中每个顺序图案化掩模中约20%至约50%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

103. 根据实施方式87所述的系统,其中每个顺序图案化掩模中约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

104. 根据实施方式87所述的系统,其中所述合成步骤基于光刻法。

105. 根据实施方式87所述的系统,其中所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。

106. 根据实施方式87所述的系统,其中所述文库中至少40%的分子是不同的。

107. 根据实施方式87所述的系统,其中所述文库中至少50%的分子是不同的。

108. 根据实施方式87所述的系统,其中所述文库中至少60%的分子是不同的。

109. 根据实施方式87所述的系统,其中所述文库中至少70%的分子是不同的。

110. 根据实施方式87所述的系统,其中所述文库中至少80%的分子是不同的。

111. 根据实施方式87所述的系统,其中所述文库中至少90%的分子是不同的。

112. 根据实施方式87所述的系统,其中所述文库中至少50%的分子的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

113. 根据实施方式87所述的系统,其中所述文库中至少50%的分子的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

114. 根据实施方式87所述的系统,其中所述文库中的所述分子包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。

115. 根据实施方式87所述的系统,其中所述文库包含等于所述生物序列长度的中值单体长度。

116. 根据实施方式87所述的系统,其中所述文库包含长于所述生物序列长度的40%、50%、60%、70%、80%或90%的中值单体长度。

117. 根据实施方式87所述的系统,其中所述文库包含短于所述生物序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值单体长度。

118. 根据实施方式87所述的系统,其中所述基底选自阵列、晶片、载玻片和珠子。

119. 根据实施方式87所述的系统,其中所述合成的化学文库包含肽、核苷酸或其组合。

120. 根据实施方式119所述的系统,其中所述肽的长度为约5至约25个氨基酸。
121. 根据实施方式120所述的系统,其中所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。
122. 根据实施方式87所述的系统,其中利用能够在氧化条件下环化的表面间隔物来合成所述化学文库。
123. 根据实施方式122所述的系统,其中所述表面间隔物为Cys-Gly-Pro-Gly-Xaa<sub>n</sub>-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-Cys。
124. 根据实施方式87所述的系统,其中利用能够用酯键环化的表面间隔物来合成所述化学文库。
125. 根据实施方式124所述的系统,其中所述酯键为同双官能的二-NHS酯键。
126. 根据实施方式125所述的系统,其中所述表面间隔物为Lys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-赖氨酸。
127. 根据实施方式87所述的系统,其中所述基底涂覆有亲水性单层。
128. 根据实施方式127所述的系统,其中所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。
129. 根据实施方式127所述的系统,其中所述亲水性单层是均匀的。
130. 一种原位合成肽阵列的方法,所述方法包括:
- 接收输入氨基酸序列;
  - 确定多个合成步骤;
  - 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;
  - 为每个图案化掩模指定至少一个单体;以及
  - 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。
131. 根据实施方式130所述的方法,其中所述合成步骤的数目大于所述生物序列长度的50%、60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%。
132. 根据实施方式130所述的方法,其中所述输入序列包含疾病相关的表位。
133. 根据实施方式130所述的方法,其中所述输入序列包含肽序列。
134. 根据实施方式130所述的方法,其中所述输入序列包含表位序列。
135. 根据实施方式130所述的方法,其进一步包括从所述输入序列中导出单体的有序列表。
136. 根据实施方式135所述的方法,其中所述有序列表的大小为所述合成步骤的数目。
137. 根据实施方式135所述的方法,其中所述单体的有序列表包含所述输入序列。
138. 根据实施方式137所述的方法,其中所述单体的有序列表以相反的顺序包含所述输入序列。
139. 根据实施方式135所述的方法,其中所述单体的有序列表包含氨基酸序列。

140. 根据实施方式130所述的方法,其中所述多个图案化掩模的数目小于10、15、20、25、30、35、40、45、50、60、70、80、90或100。

141. 根据实施方式130所述的方法,其中所述多个图案化掩模的数目为所述合成步骤的数目。

142. 根据实施方式130所述的方法,其中每个顺序图案化掩模中约20%至约50%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

143. 根据实施方式130所述的方法,其中每个顺序图案化掩模中约30%至约45%的所述激活指定特征与紧接在前的图案化掩模的所述激活指定特征重叠。

144. 根据实施方式130所述的方法,其中所述合成步骤基于光刻法。

145. 根据实施方式130所述的方法,其中所述基底上的特征的直径为约0.5微米至约200微米,并且中心的中心距为约1微米至约300微米。

146. 根据实施方式130所述的方法,其中所述阵列上至少40%的肽是不同的。

147. 根据实施方式130所述的方法,其中所述阵列上至少50%的肽是不同的。

148. 根据实施方式130所述的方法,其中所述阵列上至少60%的肽是不同的。

149. 根据实施方式130所述的方法,其中所述阵列上至少70%的肽是不同的。

150. 根据实施方式130所述的方法,其中所述阵列上至少80%的肽是不同的。

151. 根据实施方式130所述的方法,其中所述阵列上至少90%的肽是不同的。

152. 根据实施方式130所述的方法,其中所述阵列上至少50%的肽的长度为至少3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

153. 根据实施方式130所述的方法,其中所述阵列上至少50%的肽的长度为至多3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体。

154. 根据实施方式30所述的方法,其中所述阵列上的所述肽包含3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90或100个单体的中值长度。

155. 根据实施方式130所述的方法,其中所述阵列包含等于所述输入序列长度的中值肽长度。

156. 根据实施方式130所述的方法,其中所述阵列包含长于所述输入序列长度的40%、50%、60%、70%、80%或90%的中值肽长度。

157. 根据实施方式130所述的方法,其中所述阵列包含短于所述输入序列长度的60%、70%、80%、90%、100%、110%、120%、130%、140%、150%、160%、170%、180%、190%或200%的中值肽长度。

158. 根据实施方式130所述的方法,其中所述肽的长度为约5至约25个氨基酸。

159. 根据实施方式130所述的方法,其中所述氨基酸C、I和M以及任选的Q和E不包括在可用于肽合成的所述氨基酸中。

160. 根据实施方式130所述的方法,其中利用能够在氧化条件下环化的表面间隔物来合成所述肽阵列。

161. 根据实施方式160所述的方法,其中所述表面间隔物为Cys-Gly-Pro-Gly-Xaa<sub>n</sub>-Gly-Pro-Gly-Cys或Cys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-Cys。

162. 根据实施方式130所述的方法,其中利用能够用酯键环化的表面间隔物来合成所述肽阵列。

163. 根据实施方式162所述的方法,其中所述酯键为同双官能的二-NHS酯键。

164. 根据实施方式160所述的方法,其中所述表面间隔物为Lys-(PEG3)-Xaa<sub>n</sub>-(PEG3)-赖氨酸。

165. 根据实施方式130所述的方法,其中所述肽阵列涂覆有亲水性单层。

166. 根据实施方式165所述的方法,其中所述亲水性单层包含聚乙二醇(PEG)、聚乙烯醇、羧甲基葡聚糖及其组合。

167. 根据实施方式165所述的方法,其中所述亲水性单层是均匀的。

168. 一种阵列,其包含所述阵列上的多个原位合成的肽,所述肽由多个图案化掩模产生,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠。

169. 一种用于表征针对至少一种蛋白质靶标的抗体结合的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(a)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(c) 使用步骤(b)的所述比对得分表征针对所述至少一种蛋白质靶标的所述抗体的结合。

170. 根据实施方式169所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

171. 根据实施方式169所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

172. 根据实施方式169所述的方法,其中所述竞争肽包含生物样品。

173. 根据实施方式169所述的方法,其中所述生物样品为血清。

174. 根据实施方式169所述的方法,其中所述竞争肽衍生自所述靶蛋白。

175. 根据实施方式174所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

176. 根据实施方式169所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

177. 根据实施方式176所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

178. 根据实施方式169所述的方法,其中所述竞争肽包含生物样品和实施方式174至177中任一项的肽。

179. 根据实施方式169所述的方法,其中所述肽阵列包含至少1000种独特的肽。

180. 根据实施方式169所述的方法,其中所述肽阵列包含至少10,000种独特的肽。

181. 根据实施方式169所述的方法,其中所述肽阵列包含至少100,000种独特的肽。

182. 根据实施方式169所述的方法,其中所述肽阵列包含至少1,000,000种独特的肽。

183. 根据实施方式169所述的方法,其中所述肽阵列是原位合成的。

184. 根据实施方式183所述的方法,其中所述肽阵列通过以下方法合成:

i. 接收输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

185. 根据实施方式169所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

186. 根据实施方式169所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

187. 根据实施方式169所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的所述比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。

188. 根据实施方式169所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自实施方式169中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

189. 根据实施方式169所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自实施方式169中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

190. 一种用于鉴定靶蛋白中抗体表位的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(a)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(c) 确定步骤(a)的所述单独的肽中的保守氨基酸,以鉴定保守结合肽基序,并将所述单独的基序与所述至少一种靶蛋白比对,以便鉴定所述靶蛋白的至少一种抗体表位。

191. 根据实施方式190所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

192. 根据实施方式190所述的方法,其中所述预定阈值是在存在竞争肽的情况下

的结合信号,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

193.根据实施方式190所述的方法,其中所述竞争肽包含生物样品。

194.根据实施方式190所述的方法,其中所述生物样品为血清。

195.根据实施方式190所述的方法,其中所述竞争肽衍生自所述靶蛋白。

196.根据实施方式195所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

197.根据实施方式190所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

198.根据实施方式197所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

199.根据实施方式190所述的方法,其中所述竞争肽包含生物样品和实施方式193至198中任一项的肽。

200.根据实施方式190所述的方法,其中所述肽阵列包含至少1000种独特的肽。

201.根据实施方式190所述的方法,其中所述肽阵列包含至少10,000种独特的肽。

202.根据实施方式190所述的方法,其中所述肽阵列包含至少100,000种独特的肽。

203.根据实施方式190所述的方法,其中所述肽阵列包含至少1,000,000种独特的肽。

204.根据实施方式190所述的方法,其中所述肽阵列是原位合成的。

205.根据实施方式204所述的方法,其中所述肽阵列通过以下方法合成:

i.接收输入氨基酸序列;

ii.确定多个合成步骤;

iii.确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv.为每个图案化掩模指定至少一个单体;以及

v.将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

206.根据实施方式190所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

207.根据实施方式190所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

208.根据实施方式190所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的趋势。

209.根据实施方式190所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自实施方式190中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

210.根据实施方式190所述的方法,其进一步包括确定每种抗体的度量得分,其中

每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自实施方式190中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

211.根据实施方式190所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

212.根据实施方式211所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

213.一种用于表征靶蛋白中的抗体结合区的方法,所述方法包括:

(a)在存在和不存在多种竞争肽的情况下,使第一肽阵列与所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第一预定阈值内。

(b)使用选自步骤(a)中至少一种所述单独的肽的输入肽序列、衍生自步骤(a)中所述单独的肽的比对的保守基序或衍生自步骤(a)中所述单独的肽的比对的比对基序来产生第二肽阵列,所述第二肽阵列通过以下方法合成:

i.确定多个合成步骤;

ii.确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iii.为每个图案化掩模指定至少一个单体;以及

iv.将所述单体耦合至所述特征,其中(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列;

(c)使所述第二肽阵列与所述抗体接触以鉴定第二组肽;以及

(d)在存在多种竞争肽的情况下使所述第二肽阵列与所述抗体接触,并鉴定来自步骤(c)中的第二组单独的肽,所述第二组单独的肽展现出处于步骤(c)中所述结合信号的第二预定阈值内的结合信号;以及

(e)将所述第二组单独的肽与所述靶蛋白比对,并鉴定所述靶蛋白中与所鉴定的第二组单独的肽比对的区域,从而表征所述靶蛋白中的抗体结合区。

214.根据实施方式213所述的方法,其中所述竞争肽包含生物样品。

215.根据实施方式213所述的方法,其中所述生物样品为血清。

216.根据实施方式213所述的方法,其中所述竞争肽衍生自所述靶蛋白。

217.根据实施方式216所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

218.根据实施方式213所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

219.根据实施方式218所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

220.根据实施方式213所述的方法,其中所述竞争肽包含生物样品和实施方式216至219中任一项的肽。

221.根据实施方式213所述的方法,其中所述肽阵列包含至少1000种独特的肽。

222.根据实施方式213所述的方法,其中所述肽阵列包含至少10,000种独特的肽。

223. 根据实施方式213所述的方法,其中所述肽阵列包含至少100,000种独特的肽。

224. 根据实施方式213所述的方法,其中所述肽阵列包含至少1,000,000种独特的肽。

225. 根据实施方式213所述的方法,其中所述肽阵列是原位合成的。

226. 根据实施方式213所述的方法,其中所述第一肽阵列通过以下方法合成:

i. 接收输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

227. 根据实施方式213所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

228. 根据实施方式213所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

229. 根据实施方式213所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。

230. 根据实施方式213所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自实施方式213中步骤(b)的所述比对得分和单个肽的信号的组合和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

231. 根据实施方式213所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自实施方式213中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

232. 根据实施方式213所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

233. 根据实施方式232所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

234. 根据实施方式213所述的方法,其中所述第一预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽时的所述结合信号的至少20倍内。

235. 根据实施方式213所述的方法,其中所述第二预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

236. 根据实施方式213所述的方法,其中所述第一预定阈值是在存在竞争肽的情

况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

237. 根据实施方式213所述的方法,其中所述第二预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

238. 根据实施方式213所述的方法,其中所述抗体结合区为所述靶蛋白的线性表位。

239. 根据实施方式213所述的方法,其中所述抗体结合区为所述靶区域的结构表位。

240. 根据实施方式239所述的方法,其中用选自实施方式213的步骤(a)中至少一种所述单独的肽的另外的肽来重复实施方式213中的步骤(b)至步骤(d)。

241. 一种用于鉴定抗体的靶蛋白的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使第一肽阵列与一种或多种浓度的所述抗体接触以获得一个或多个输入氨基酸序列,其中所鉴定的输入氨基酸序列展现出在存在所述多种竞争肽的情况下的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下的所述结合信号的第一预定阈值内。

(b) 使用选自步骤(a)中至少一种所述单独的肽的一个或多个输入氨基酸序列、衍生自步骤(a)中所述单独的肽的比对的保守基序或衍生自步骤(a)中所述单独的肽的比对的比对基序来获得一种或多种二级肽阵列,所述一种或多种二级肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

(c) 在存在和不存在所述多种竞争肽的情况下,使每个所述二级肽阵列与所述抗体接触以获得一组肽序列,其中所鉴定的一组肽序列展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第二预定阈值内。

(d) 使所述一组肽序列彼此比对以获得至少一种预测性结合基序;以及

(e) 将所述预测性结合基序作为针对蛋白质数据库的搜索标准进行比对,从而基于所述蛋白质数据库搜索结果得分来鉴定所述抗体的靶蛋白。

242. 根据实施方式241所述的方法,其中所述竞争肽包含生物样品。

243. 根据实施方式241所述的方法,其中所述生物样品为血清。

244. 根据实施方式241所述的方法,其中所述竞争肽衍生自所述靶蛋白。

245. 根据实施方式244所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

246. 根据实施方式241所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

247. 根据实施方式246所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

248. 根据实施方式241所述的方法,其中所述竞争肽包含生物样品和实施方式244至247中任一项的肽。

249. 根据实施方式241所述的方法,其中所述肽阵列包含至少1000种独特的肽。

250. 根据实施方式241所述的方法,其中所述肽阵列包含至少10,000种独特的肽。

251. 根据实施方式241所述的方法,其中所述肽阵列包含至少100,000种独特的肽。

252. 根据实施方式241所述的方法,其中所述肽阵列包含至少1,000,000种独特的肽。

253. 根据实施方式241所述的方法,其中所述肽阵列是原位合成的。

254. 根据实施方式241所述的方法,其中所述第一肽阵列通过以下方法合成:

i. 接收输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

255. 根据实施方式241所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

256. 根据实施方式241所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

257. 根据实施方式241所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的趋势。

258. 根据实施方式241所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自实施方式241中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

259. 根据实施方式241所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自实施方式241中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

260. 根据实施方式241所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

261. 根据实施方式260所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

262. 根据实施方式241所述的方法,其中所述第一预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

263. 根据实施方式241所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

264. 一种用于确定抗体与至少一种蛋白质靶标结合倾向的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 使步骤(a)的所述单独的肽与第一蛋白质靶标比对,其中步骤(a)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(c) 重复步骤(a)的所述单独的肽与至少一种另外的蛋白质靶标的比对,其中步骤(a)的所述单独的肽与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(d) 比较步骤(b)和(c)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

265. 根据实施方式264所述的方法,其中所述竞争肽包含生物样品。

266. 根据实施方式265所述的方法,其中所述生物样品为血清。

267. 根据实施方式264所述的方法,其中所述竞争肽衍生自所述靶蛋白。

268. 根据实施方式264所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

269. 根据实施方式264所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

270. 根据实施方式269所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

271. 根据实施方式264所述的方法,其中所述竞争肽包含生物样品和实施方式267至270中任一项的肽。

272. 根据实施方式264所述的方法,其中所述肽阵列包含至少1000种独特的肽。

273. 根据实施方式264所述的方法,其中所述肽阵列包含至少10,000种独特的肽。

274. 根据实施方式264所述的方法,其中所述肽阵列包含至少100,000种独特的肽。

275. 根据实施方式264所述的方法,其中所述肽阵列包含至少1,000,000种独特的肽。

276. 根据实施方式264所述的方法,其中所述肽阵列是原位合成的。

277. 根据实施方式264所述的方法,其中所述肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧邻的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(b)和(c)组装成一个所述合成步骤并重复

所述合成步骤以形成所述肽阵列。

278. 根据实施方式264所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

279. 根据实施方式264所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

280. 根据实施方式264所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的所述比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。

281. 根据实施方式264所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自实施方式264中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

282. 根据实施方式264所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自实施方式264中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

283. 根据实施方式264所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

284. 根据实施方式283所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

285. 根据实施方式264所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

286. 根据实施方式264所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%。

287. 一种用于确定抗体与至少一种蛋白质靶标结合倾向的方法,所述方法包括:

(a) 在存在和不存在一种或多种浓度的多种竞争肽的情况下,使第一肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(b) 对步骤(a)的所述一种或多种单独的肽进行比对以获得至少一种预测性靶基序;

(c) 将所述至少一种预测性靶基序与第一蛋白质靶标比对,其中步骤(a)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(d) 重复步骤(b)的至少一种预测性靶基序与至少一种另外的蛋白质靶标的比对,其中步骤(b)的所述至少一种预测性靶基序与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(e) 比较步骤(c)和(d)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

288. 根据实施方式287所述的方法,其中所述竞争肽包含生物样品。

289. 根据实施方式288所述的方法,其中所述生物样品为血清。

290. 根据实施方式287所述的方法,其中所述竞争肽衍生自所述靶蛋白。

291. 根据实施方式290所述的方法,其中所述竞争肽与所述靶蛋白至少50%相似。

292. 根据实施方式287所述的方法,其中所述竞争肽衍生自所述抗体的已知表位。

293. 根据实施方式292所述的方法,其中所述竞争肽与所述抗体的已知表位至少50%相似。

294. 根据实施方式287所述的方法,其中所述竞争肽包含生物样品和实施方式290至293中任一项的肽。

295. 根据实施方式287所述的方法,其中所述肽阵列包含至少1000种独特的肽。

296. 根据实施方式287所述的方法,其中所述肽阵列包含至少10,000种独特的肽。

297. 根据实施方式287所述的方法,其中所述肽阵列包含至少100,000种独特的肽。

298. 根据实施方式287所述的方法,其中所述肽阵列包含至少1,000,000种独特的肽。

299. 根据实施方式287所述的方法,其中所述肽阵列是原位合成的。

300. 根据实施方式287所述的方法,其中所述肽阵列通过以下方法合成:

i. 接收输入氨基酸序列;

ii. 确定多个合成步骤;

iii. 确定多个图案化掩模,其中每个图案化掩模都被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧接的图案化掩模的所述激活指定特征重叠;

iv. 为每个图案化掩模指定至少一个单体;以及

v. 将所述单体耦合至所述特征,其中(c)和(d)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列。

301. 根据实施方式287所述的方法,其中所述结合信号被测量为在不存在和存在一种或多种浓度的所述竞争肽的情况下的信号强度。

302. 根据实施方式287所述的方法,其中在存在和不存在一种或多种浓度的所述竞争肽的情况下获得表观Kd。

303. 根据实施方式287所述的方法,其中使至少一种另外的抗体与所述肽阵列接触,并对每种抗体获得的所述比对得分进行分级,以确定每种抗体与所述蛋白质靶标结合的倾向。

304. 根据实施方式287所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一结合谱度量,所述单一结合谱度量衍生自实施方式287中步骤(b)的所述比对得分和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

305. 根据实施方式287所述的方法,其进一步包括确定每种抗体的度量得分,其中每种抗体被指定单一特异性谱度量,所述单一特异性谱度量衍生自实施方式287中步骤(b)的所述比对得分、具有来自步骤(b)的超过一个比对位置的肽的数目和具有来自步骤(b)的超过一个比对位置的步骤(a)的所述单独的肽的信号的组合。

306. 根据实施方式287所述的方法,其进一步包括将所述至少一种抗体表位作为针对蛋白质数据库的搜索标准进行比对。

307. 根据实施方式306所述的方法,其中所述蛋白质数据库为蛋白质组数据库,并且其中鉴定了另外的抗体靶蛋白和/或交叉反应蛋白。

308. 根据实施方式287所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号处于在不存在竞争肽的情况下的所述结合信号的至少20倍内。

309. 根据实施方式287所述的方法,其中所述预定阈值是在存在竞争肽的情况下的结合信号,所述结合信号为与不存在竞争剂的情况下相比的所述结合信号的至少5%的结合信号。

310. 一种用于表征针对至少一种蛋白质靶标的抗体结合的试剂盒,所述试剂盒包括:

(a) 提供肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在一种或多种浓度的所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(d) 为所述用户提供说明以将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(c)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(e) 为所述用户提供说明以使用步骤(d)的所述比对得分表征针对所述至少一种蛋白质靶标的所述抗体的结合。

311. 一种用于鉴定靶蛋白中抗体表位的试剂盒,所述试剂盒包括:

(a) 提供肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的所述抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(d) 为所述用户提供说明以将所述单独的肽与所述至少一种蛋白质靶标比对,其中步骤(c)的所述单独的肽与至少一种蛋白质靶标之间的所述比对被指定为比对得分;以及

(e) 为所述用户提供说明以确定步骤(c)的所述单独的肽中的保守氨基酸,以鉴定保守结合肽基序,并将所述单独的基序与所述至少一种靶蛋白比对,以便鉴定所述靶蛋白的至少一种抗体表位。

312. 一种用于表征靶蛋白中抗体结合区的试剂盒,所述试剂盒包括:

(a) 提供第一肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在所述多种竞争肽的情况下,使第一肽阵列与抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的第一预定阈值内。

(d) 为用户提供说明以使用选自步骤(c)中至少一种所述单独的肽的输入肽序列、衍生自步骤(c)中所述单独的肽的比对的保守基序或衍生自步骤(c)中所述单独的肽的比对的比对基序来产生第二肽阵列,所述第二肽阵列通过以下方法合成:

i. 确定多个合成步骤;

ii. 确定多个图案化掩模,其中每个图案化掩模被指定为所述基底上的每个特征的激活或非激活指定,并且其中每个顺序图案化掩模中约1%至约75%的所述激活指定特征与紧邻的图案化掩模的所述激活指定特征重叠;

iii. 为每个图案化掩模指定至少一个单体;以及

iv. 将所述单体耦合至所述特征,其中(ii)和(iii)组装成一个所述合成步骤并重复所述合成步骤以形成所述肽阵列;

(e) 为所述用户提供说明以使所述第二肽阵列与所述抗体接触以鉴定第二组肽;

(f) 为所述用户提供说明以在存在所述多种竞争肽的情况下使所述第二肽阵列与所述抗体接触,并鉴定来自步骤(e)的第二组单独的肽,所述第二组单独的肽展现出处于在步骤(e)中所述结合信号的第二预定阈值内的结合信号;以及

(g) 为用户提供说明以将所述第二组单独的肽与所述靶蛋白比对,并鉴定所述靶蛋白中与所鉴定的第二组单独的肽比对的区域,从而表征所述靶蛋白中的抗体结合区。

313. 一种用于确定抗体与至少一种蛋白质靶标结合倾向的试剂盒,所述试剂盒包括:

(a) 提供肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(d) 为所述用户提供说明以使步骤(c)的所述单独的肽与第一蛋白质靶标比对,其中步骤(c)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(e) 为所述用户提供说明以重复步骤(c)的单独的肽与至少一种另外的蛋白质靶标的比对,其中步骤(c)的所述单独的肽与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(f) 为所述用户提供说明以比较步骤(c)和(d)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

314. 一种用于确定抗体与至少一种蛋白质靶标结合倾向的试剂盒,所述试剂盒包括:

(a) 提供第一肽阵列;

(b) 提供多种竞争肽;

(c) 为用户提供说明以在存在和不存在一种或多种浓度的所述多种竞争肽的情况下,使所述第一肽阵列与一种或多种浓度的抗体接触以获得一种或多种单独的肽,其中所鉴定的一种或多种单独的肽展现出在存在所述多种竞争肽的情况下测量的结合信号,所述结合信号处于在不存在所述多种竞争肽的情况下测量的所述结合信号的预定阈值内。

(d) 为所述用户提供说明以对步骤(c)的所述一种或多种单独的肽进行比对以获得至少一种预测性靶基序;

(e) 为所述用户提供说明以将所述至少一种预测性靶基序与第一蛋白质靶标比对,其中步骤(c)的所述单独的肽与所述第一蛋白质靶标之间的所述比对被指定为比对得分;

(f) 为所述用户提供说明以重复步骤(e)的至少一种预测性靶基序与至少一种另外的蛋白质靶标的比对,其中步骤(e)的所述至少一种预测靶基序与所述另外的蛋白质靶标之间的所述比对被指定为比对得分;以及

(g) 为所述用户提供说明以比较步骤(c)和(d)的所述比对得分,以获得所述抗体与所述蛋白质靶标结合的相对倾向。

#### 实施例

##### 实施例1:计算机模拟

[0391] 在一些实施方案中,通过计算机模拟方法模拟来掩模算法。在该实施例中,模拟包括以下参数:

- 特征总数:500,000
- 掩模n与掩模n-1之间的重叠百分比:52%
- 输入序列:HVGAAAPVVPQA (疾病相关的表位)
- 合成步骤的数目:21
- 添加的合成顺序:A,R,Q,S,P,W,V,V,P,A,D,A,A,M,G,V,F,H,K,L,Y

o 由用户选择添加的合成顺序以生成与输入序列或多或少密切相关的序列空间(参见上文),然而,输入序列中的所有氨基酸必须包括在添加的合成顺序中。与输入序列中的氨基酸顺序相似的耦合合成顺序将生成与输入序列更密切相关的空间;相反地,与输入序列中的氨基酸的顺序不太相似的耦合合成顺序将生成与输入序列不太密切相关的序列空间。

[0392] 以下描述了模拟的伪代码:

```
librarySize <- 500000
fractionSeqMaskOverlap <- 0.52
inputSequence <- "H V G A A A P V V P Q A "
synthesisOrder <- c("A", "R", "Q", "S", "P", "W", "V", "V", "P", "A", "L",
"A", "A", "M", "G", "V", "F", "H", "K", "L", "Y")

numSteps <- length(synthesisOrder)
numFeaturePerMask <- floor(fractionSeqMaskOverlap*librarySize)

librarySeqs <- rep('', librarySize)
libraryLengths <- rep(0, librarySize)
inputSeqReps <- c()

n_lMask <- rep(FALSE, librarySize)
n_lMask[sample(seq(1:librarySize), numFeaturePerMask)] <- TRUE

for(currStep in synthesisOrder) {

  str(currStep)
  n_lMaskOpen <- which(n_lMask, arr.ind = TRUE)

  for(currIdx in n_lMaskOpen){
    librarySeqs[currIdx] <- paste(currStep, librarySeqs[currIdx])
    libraryLengths[currIdx] <- length(librarySeqs[currIdx])/2
    if(librarySeqs[currIdx] == inputSequence) {
      inputSeqReps <- append(inputSeqReps, currIdx)
    }
  }
}

<- sort(sample(n_lMaskOpen, floor(fractionSeqMaskOverlap*
                                length(n_lMaskOpen))))
diffIndices <- setdiff(1:librarySize, n_lMaskOpen)

nMaskOpen <- sort(sample(diffIndices, numFeaturePerMask-
                        length(overlapIndices)))
nMaskOpenIndices <- sort(append(overlapIndices, nMaskOpen))
nMask <- rep(FALSE, librarySize)
nMask[nMaskOpenIndices] <- TRUE

n_lMask <- nMask
}
```

[0393] 考虑以下输入和参数设置:输入序列为HVGAAAPVVPQA;合成步骤的数目为21;文库大小包括500,000个特征;采样空间中输入序列重复的数目为3;采样空间中所有序列的平均重复数为1.4个重复/序列;采样空间中不同序列的数目为360,064个序列。图9示出了使用本文公开的掩模和合成算法生成的模拟文库中序列长度的分布,其中中值长度为11。

[0394] 此外,下表示出了从使用本文公开的掩模和合成算法的模拟生成的序列空间中选择的一组示例序列。

H M A D V V S Q R A	L H V G A A A P V V P Q A	H V G M D P V W Q A
G A A P P Q A	H V P V V W P Q A	K F V M A A A V S A
L F V G M P V	Y L H F V M D V V Q	M A D A P V W P S Q A
Y K G M A D V W P S Q A	Y L H F Q R A	Y V G A P V W S
H F V G M A A A V P S Q R	K V M A D A V V P Q A	K H A A P W P S
K H F V M A A P V W P A	Y K H V M A D P V Q R	Y L K F M A D V V W P S R A
L K H V G A A A P V V P Q A	L H F V A A P V P S R	L K H V G M D P V V P S
Y L F V G A D P W S R A	L F M A A P Q A	L K H F G A A A V V P Q
F V A D A P V P S Q R A	K H F G V W Q R A	K H V A A D A V W P A
Y L K H V A A A W P S R A	Y L H V M A A A V P A	Y L H V G A A A P V V P Q A
Y H F V G M A A A V A	G M A D A P V P Q R	Y H V G A A P V W A

#### 实施例2:高密度肽阵列上抗体结合谱的表征

[0395] 鉴定抗HER2 mAb与阵列肽的结合

[0396] 设计竞争结合测定以鉴定反映mAb的生物结合的阵列肽。将具有该特征的阵列肽鉴定为单独的肽。可以应用单独的肽的等级,其中显著肽被定义为不具有缺口的多于一个精确匹配,尽管具有不同程度的带有缺口的匹配的肽也是可接受的。进行该测定以鉴定14种商购的针对HER2的治疗或研究单克隆抗体中的显著肽(表1)。该组包括来自产生不同免疫原的不同克隆、产生相同的免疫原的不同克隆以及从不同供应商获得的相同克隆。如下所述,在不存在和存在竞争剂的情况下测量每种mAb的表观Kd(半数最大可饱和结合时的抗体浓度)。

[0397] 表1-抗HER2单克隆抗体组

抗体名称&克隆	宿主	HER2 免疫原序列区	克隆性	部件#	供应商
Neu 抗体(C-12):sc-374382	鼠	983-1017	mono	sc-374382	SCBT
Neu 抗体(C-3):sc-377344	鼠	251-450	mono	sc-377344	SCBT
Neu 抗体(A-2):sc-393712	鼠	1180-1197	mono	sc-393712	SCBT
Neu 抗体(3B5)	鼠	1242-1255	mono	sc-33684	SCBT
单克隆抗 HER2 抗体 4B8	鼠	22-122	mono	WH0002064M6	Sigma
单克隆抗 HER2 抗体 CL0268	鼠	274-400	mono	AMAB90627	Sigma
抗 ERBB2/HER2 抗体 (aa676-1255, 克隆 11A7)	鼠	676-1255	mono	LS-C337488	LSBio
抗 ERBB2/HER2 抗体(aa23-652)	鼠	23-652	mono	LS-C128811	LSBio
HER-2/ErbB2 抗体(6C2)	鼠	750-987	mono	MA5-15702	Thermo
HER-2/ErbB2 抗体(3B5)	鼠	C-末端	mono	MA5-13675	Thermo
HER2/ErbB2(D8F12)XP®兔 mAb	兔	N-末端	mono	#4290	Cell Signaling
HER2/ErbB2(44E7)小鼠 mAb	鼠	C-末端	mono	#2248	Cell Signaling
HER2/ErbB2(29D8)兔 mAb	兔	1242-1255	mono	#2165	Cell Signaling
抗人 ErbB2 治疗性抗体(曲妥珠单抗)	重组人源化	人 ERBB2	mono	TAB-005	Creative Biolabs

[0398] 竞争性结合测定。获得包含多样化肽阵列或聚焦肽文库(在实施例1中描述)的微阵列,并在使用前通过在蒸馏水中温和搅拌浸泡1小时,在PBS中温和搅拌浸泡30分钟以及在初级温育缓冲液(PBST,1%甘露醇)中温和搅拌浸泡1小时进行再水化。将包含微阵列的载玻片装载到ArrayIt微阵列盒(ArrayIt,Sunnyvale,CA)中以使单个微阵列适应微量滴定板印迹。通过在温育缓冲液(PBST,1%甘露醇)中连续稀释原液来制备六种不同浓度的mAb溶液,每种mAb为:3nM、1nM、0.33nM、0.11nM、0.0367nM和0.012nM。

[0399] 在不存在或存在两种不同浓度的血清竞争剂(1/69ND和1/71ND)的情况下或在不存在或存在两种不同浓度的竞争肽混合物(250 $\mu$ M和750 $\mu$ M)的情况下测定mAb结合。竞争肽的混合物由根据以下标准选择的24种肽组成:a)用于提供具有平衡氨基酸组分的混合物,即不富含任何一种氨基酸的肽;b)用于具有GRAVY得分<0,以确保水性测定中的溶解度;以及c)用于具有平衡且连续的等电点(pI)范围,范围从pI=3到pI=10。将阵列与不同mAb溶液在37 $^{\circ}$ C下温育30分钟,并在TeleShake95(INHECO,Martinsried,Germany)上混合以允许抗体-肽结合。温育后,将盒在PBST(PBS-吐温)中以10X室容积洗涤。根据第一抗体的来源,使用与AlexaFluor 647(Thermo-Invitrogen,Carlsbad,CA)耦合的4.0nM山羊抗人IgG(H+L)、山羊抗兔或山羊抗小鼠第二抗体检测结合的mAb。允许第二抗体的结合在37 $^{\circ}$ C下在温育缓冲液(PBST中的3%BSA)中进行1小时,同时在TeleShake95平台混合器上混合。与第二抗体温育后,将载玻片再次用PBST以10X室容积和蒸馏水洗涤,从盒中去除,用异丙醇喷雾并离心干燥。

[0400] 数据采集。使用配有532nm激光和572nm BP 34滤光器(nnopsys,Carbonne,France)的Innopsys 910AL微阵列扫描仪对测定的文库阵列进行成像。Mapix软件应用程序(版本7.2.1)使用自动网格化算法鉴定与每个肽特征相关的图像区域。将每个肽特征的中值像素强度保存为制表符分隔的文本文件,并存储在数据库中进行分析。通过确定每个可寻址肽特征的相对荧光值,以1 $\mu$ M分辨率和1%特征饱和度来获得定量信号测量。对于测定的每种mAb,获得三十结合测量。

[0401] 信号分析。通过定量荧光信号来测量mAb与每个特征的结合。中值特征强度首先相对于阴性对照(仅第二抗体)减去背景,然后进行log<sub>10</sub>转化,随后通过除以log<sub>10</sub>转化的中值进行归一化。

[0402] 阵列肽结合的特异性。本文的特异性是指抗体分化两种不同抗原的程度。(参考文献:Immunology and Infectious Disease,S.A.Frank,2002,Princeton Univ.Press)。通过在不存在和存在竞争剂的情况下获得的结合信号的差异来表征每个阵列肽的结合特异性,并且通过非同源肽竞争剂或血清竞争剂减弱结合的程度提供了mAb特异性的测量。通过在不存在竞争剂的情况下和在存在每种血清和非同源肽竞争剂的情况下的每种阵列肽的表观K<sub>d</sub>值的差异来确定肽结合特异性。

[0403] 结果。数据示出阵列肽在不存在竞争剂的情况下展现出可饱和的mAb剂量-应答结合,并且在存在血清或肽竞争剂的情况下维持该可饱和的结合。随后,相对于在不存在竞争剂的情况下获得的表观K<sub>d</sub>,在存在竞争剂的情况下确定的表观K<sub>d</sub>的降低用于从肽库阵列筛选中选择单独的肽,在这种情况下是显著肽。

[0404] 表2:赫赛汀表观K<sub>d</sub>结果

赫赛汀结合肽序列	750 uM 肽混合竞争剂的表现	
	无竞争剂的表现 Kd	Kd
FGPYKPFGAQ	0.001	0.010
PYKFFP	0.001	0.003
EYKPFWKGAP	0.001	0.010
FGPQYKPFQP	0.001	0.001
FGPYKPIGAQPP	0.002	0.007
FGEQYKPPIWKGAQPP	0.003	0.008
FGPQYKPI	0.003	0.020
QPFWKFP	0.005	0.010
QPFPIWKGAQP	0.010	0.020
FGPQYKPIWKFP	0.020	0.020

[0405] 根据表观Kd的倍数变化对肽进行分级。选择单独的肽(包括显著肽),以便在存在竞争剂的情况下测量时表观Kd有变化,而当在不存在每个竞争剂的情况下测量时Kd降低小于10倍。

[0406] 随后,单独的肽(包括显著肽)用于鉴定线性表位和结构表位,从而鉴定靶表位内的关键氨基酸以确定mAb的结合特异性,并鉴定未知蛋白质靶标。

#### 实施例3:多样化和聚焦的肽阵列

[0407] 多样化文库。制备多样化肽文库以对组合肽文库中表示的高度多样化的序列空间进行采样,并提供包含预测结合表位的富集基序的单独的肽(包括显著肽)。富集的基序作为鉴别用于设计聚焦文库的输入序列的基础。参见图10。

[0408] 在所提供的方法中使用的多样化文库被制备为126,009个肽的初级高度多样化组合文库,其中值长度为9个残基,范围为5至13个氨基酸,并且被设计为包括16个氨基酸(排除甲硫氨酸,M;半胱氨酸,C;异亮氨酸,I;和苏氨酸,T)的99.9%的所有可能的4-聚体和48.3%的所有可能的5-聚体。使用适用于叔丁氧基羰基(BOC)保护基团肽化学的标准半导体光刻工具在200mm氧化硅晶片上合成肽(Legutki JB等人,Nature Communications.2014;5:4785)。简言之,用BOC-甘氨酸涂覆氨基硅烷官能化晶片。接下来,通过旋转涂覆将含有通过UV光激活的光酸产生剂的光致抗蚀剂施加至晶片。通过光掩模将晶片暴露于UV光(365nm)允许使用给定掩模将固定选择晶片上的特征暴露出来。在暴露于UV光之后,加热晶片,从而允许暴露的特征的BOC-去保护。随后洗涤,然后通过施加激活的氨基酸完成循环。对于每个循环,将特定氨基酸添加到位于阵列上特定位置的肽的N-末端。重复这些循环,改变掩模和耦合的氨基酸,以实现组合肽文库。从每个晶片切割出具有标准显微镜载玻片尺寸的十三个矩形区域。将每个完成的晶片切割成13个具有标准显微镜载玻片尺寸(25mm X 75mm)的矩形区域。这些载玻片中的每一个含有8行3列的24个阵列。最后,使用标准混合物去除一些氨基酸侧链上的保护基团。将完成的载玻片储存在干燥的氮气环境中直至需要。进行了许多质量测试,确保在工艺规范内制造阵列,该工艺规范包括对每个步骤使用3 $\sigma$ 统计限值。通过MALDI-MS间歇地对晶片批次进行采样,以鉴定每个氨基酸在正确的步骤中耦合,从而确保构成组合合成的单个步骤是正确的。经由电子自定义关系数据库从头到尾跟踪晶片制造,该数据库用Visual Basic编写并具有带有SQL后端的访问前端。

前端用户界面允许操作员轻松地将生产信息输入数据库。SQL后端提供简单的数据库备份方法,并与其他计算机系统集成以进行数据共享。通常跟踪的数据包括化学品、配方、时间和执行任务的技术人员。在生产晶片之后,检查数据并锁定和储存记录。最后,如下所述,在结合测定中评估每个批次以确认性能。

[0409] 与多样化文库的阵列肽的单克隆结合鉴定了包含3-5聚体基序的单独的肽(包括显著肽),所述单独的肽用于鉴定用于设计聚焦文库的输入序列(图10)。

[0410] 聚焦文库。制备聚焦文库以改变输入序列周围的许多位置,该输入序列包含在多样化文库中鉴定的单独的肽(包括显著肽)的富集基序。使用一系列24个重叠掩模,将所提供的方法中使用的聚焦文库制备为16,920个肽的文库,其产生具有0至17个氨基酸残基的中值长度的合成肽。

[0411] 聚焦文库的肽被设计为各自提供多样化文库的单独的肽(在这种情况下为显著肽)的一个输入序列的变体序列。每个特征的尺寸为44 $\mu\text{m}$  X 44 $\mu\text{m}$ ,设置为50 $\mu\text{m}$  X 50 $\mu\text{m}$ 间距,在特征之间具有6 $\mu\text{m}$ 的间隙空间。如多样化肽文库的合成所述,使用适用于叔丁氧羰基(BOC)保护基团肽化学的标准半导体光刻工具在200mm氧化硅晶片上合成肽(Legutki JB等人,Nature Communications.2014;5:4785)。通过MALDI-MS间歇地对晶片批次进行采样,以鉴定每个氨基酸在正确的步骤中耦合,从而确保构成聚焦合成的单个步骤是正确的。

#### 实施例4:表位的鉴定

[0412] HER2预测表位的鉴定。使用抗HER2 mAb SCBT sc-33684、Thermo MA5-13675、Cell Signaling#2165和Creative Biolabs TAB-005对多样化或聚焦的肽阵列/文库(或二者)进行如实施例2中所述的竞争结合测定,以鉴定单独的肽(包括显著肽)和预测的表位序列。

[0413] 如实施例2中所述,根据合成肽对mAb的相对特异性水平对结合肽进行分级,并且将单独的肽选为表观Kd降低小于10倍。选择单独的肽,特别是显著肽,用于预测测试的每种mAb的HER2表位序列。

[0414] 对每个显著肽的阵列信号进行中值归一化和对数化,并且使用ClustalW和MUSCLE比对将具有>至少2倍以上中值的信号的显著肽与HER2蛋白的重叠6-聚体序列(UNIPROT#P04626)比对。将显著肽的3-聚体的正向和反向序列与任何所有可能的HER2靶标6-聚体进行比对,并确定整个HER2蛋白中每个氨基酸位置的得分。将比对得分计算为每个位置的所有得分的总和,并将其与相应的显著肽的结合信号组合以提供基序得分(图12)。基序得分足以预测靶表位。使用CLUSTALW(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308517/>)和MUSCLE(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/>)软件进行亚基序的线性排列并进行比较。

[0415] 根据基序在显著肽中的富集对基序进行分级。通过确定特定基序的概率/在文库或阵列中随机发现该基序的概率,相对于所有阵列肽(即显著和非显著的文库阵列肽)中基序的发生率计算倍数富集。表3示出了三聚体基序的示例性列表和相应的倍数富集。

[0416] 最后,对显著肽进行比对(CLUSTALW(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308517/>) and MUSCLE(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/>))以确定保守氨基酸的同一性和位置。

[0417] 表3-显著肽中富集的基序

富集的基序	倍数富集	富集P值	富集错误发现率
EVE	5.79	5.6699E-219	8.656E-217
HEV	8.13	5.3367E-206	4.8885E-204
PWE	5.88	3.7323E-144	1.3149E-142
WEV	7.65	1.0867E-119	3.1107E-118
HEVG	8.78	2.74964E-42	8.18782E-40

#### 实施例5:线性表位的鉴定

[0418] 对于测试的每种mAb,如实施例4中所述,在多样化文库中鉴定单独的肽(包括显著肽)。确定相应的富集基序以预测HER2表位和所鉴定的保守氨基酸及其位置。从三种示例性抗HER2抗体的多样化文库中鉴定的最高剂量应答肽序列:MA5-13675(克隆3B5)(Thermo Fisher;Waltham,MA)、sc-33684(克隆3B5)(Santa Cruz BioTechnologies,Dallas,TX)和2165(克隆29D8)(Cell Signalling Technologies,Danvers,MA)分别显示于图14、图15和图16中。

[0419] 将富集的基序与HER2蛋白比对以鉴定包含可以改变以设计聚焦文库的基序的区域。使用简化的一组氨基酸将基序的每个残基定位到蛋白质靶标,以减少样品阵列所需的氨基酸数目,同时充分表示可定位表位的蛋白质组序列(图13)。包含三聚体和四聚体基序的靶蛋白上的区域被显示为在多样化文库中鉴定的单独的肽(特别是显著肽)上高度保守,用作输入序列以获得其变体序列并包含用于设计聚焦文库的保守基序。

[0420] 如实施例2中所述,通过开发聚焦文库算法的过程生成位置变体。这些变体衍生自聚焦文库设计期间定义的输入序列、掩膜顺序和氨基酸顺序。

[0421] 将在每个聚焦文库中鉴定的单独的肽(在这种情况下为显著肽)与HER2靶蛋白比对,根据它们的相对特异性进行评分,并进行比对以鉴定表位的共有序列。

[0422] 图14B、图15B和图16B中示出了从三个聚焦文库鉴定的最高显著肽的比对。mAb MA5-13675(克隆3B5)(Thermo Fisher)(图11B和图14C)的保守氨基酸的位置示出了对多样化文库和聚焦文库的组合筛选的一次迭代,该一次迭代鉴定了包含在免疫原中的线性HER2表位的完整序列(图14D)。

[0423] 类似地,正确鉴定了抗HER单克隆抗体sc-33684(克隆3B5)(Santa Cruz BioTechnologies)和2165(克隆29D8)(Cell Signaling Technologies)的完整线性表位(图15C和图15D,以及图16C和图16D)。

[0424] 在测试的所有抗HER2 mAb中,多样化文库和聚焦文库的组合筛选正确鉴定了HER2的线性表位,该HER2的线性表位对应于用于产生抗HER2 mAb的公开的免疫原序列。

#### 实施例6:结构HER2表位的鉴定

[0425] 为了证明所提供的系统和方法可以鉴定抗HER2 mAb的结构表位,进行曲妥珠单抗Fab单克隆抗体(赫赛汀)与多样化和聚焦文库的结合以鉴定构成由赫赛汀识别的结构表位的三种线性组分。

[0426] 首先,进行如实施例1和实施例2中所述的赫赛汀与多样化文库的结合,以鉴定由

赫赛汀结合的肽中的富集基序,以预测结构表位的线性组分。由曲妥珠单抗Fab(赫赛汀)识别的HER2结构表位的三个单独的线性组分(图17A):FGPEADQ、KDPPFC和IWKFPDEEGACQPC(Chen,H.-S.等人Sci.Rep.5,12411;doi:10.1038/srep12411[2015])。随后使用显著肽中富集的基序来鉴定HER2靶蛋白的3个输入区。基于输入序列设计聚焦文库,该输入序列包含对应于彼此附着的三个结构组分的三个基序。使用赫赛汀筛选聚焦文库,并对鉴定的显著肽进行比对以鉴定保守氨基酸及其相对于公开的结构表位序列的位置。

[0427] 图17B中示出了从聚焦文库中鉴定的最高显著肽的示例性比对。将每个线性组分的位置定位到包含结构表位组分的序列(图17A)。图18示出了曲妥珠单抗的晶体结构(BLASTP Ident得分在<https://www.ncbi.nlm.nih.gov/books/NBK62051/>),及其相对于HER2细胞外部分的定位。HER2的有色部分代表从聚焦文库中的肽序列鉴定的单个线性组分。

[0428] 数据示出鉴定了结构表位的完整序列。这些发现进一步证实了所提供的方法概括HER2 mAb与其HER2靶标之间的生物结合相互作用的能力。

#### 实施例7:从整个蛋白质组鉴定未知抗体靶标

[0429] 使用三种mAb Cell Signalling2165(克隆29D8)、Thermo MA5-13675(克隆3B5)和Santa Cruz SC-33684(克隆3B5)来证明所提供的系统和方法从整个蛋白质组鉴定未知蛋白质靶标的能力。

[0430] 首先从相应的多样化文库中鉴定每种mAb的单独的肽(包括显著肽)和富集基序。对存在这些短基序的蛋白质组的查询通常将会导致许多比对,其中大多数比对是与正在寻找的真正靶标无关的序列。发现了包含3-4聚体基序的序列的后续设计和所得聚焦文库鉴定的9-12聚体序列对人类蛋白质组中精确匹配的筛选。图19A、图20A和图21A示出了在使用从筛选聚焦的相应聚焦文库鉴定的最高10个单独的肽(在这种情况下为显著肽)查询时所鉴定的表位序列的BLAST比对结果。这些图示出了所有最高得分比对都是HER2蛋白,也称为v-erb-b2。相反,具有中值特异性得分的显著肽在与人类蛋白质组比对时未鉴定出相关HER2序列(图19B、图20B、图21C)。

[0431] 这些数据示出了可以高可靠性地鉴定抗体的未知靶蛋白,如BLAST得分所示。

#### 实施例8:确定抗HER2 mAb的特异性

[0432] 可以使用上述多样化和聚焦文库确定单克隆抗体的特异性。

[0433] 首先,可以如实施例2中所述确定肽对多样化文库和聚焦文库二者的结合特异性。已经鉴定了来自聚焦文库的剂量-应答单独肽(包括显著肽),氨基酸的保守程度可用于确定mAb的特异性。在一种情况下,当使用参考抗体或已知与用于鉴定真实表位的mAb无关的一组参照抗体时,可以将鉴定表位共有序列的所有保守氨基酸的比特总和,例如图14D,与相同推定表位序列获得的比特总和进行比较。例如,一组与抗HER2抗体无关的10个mAb可用于作用于与多样化文库和聚焦文库结合的混合物以提供单独的肽(包括显著肽),当比对时,该单独的肽(包括显著肽)将提供可在单独的肽上保守的氨基酸的比特得分。

[0434] 因此,抗体对表位的特异性可以通过推定的表位序列中的氨基酸保守程度来定义。

#### 实施例9:确定抗体对一组不同蛋白质的倾向的方法

[0435] 进行mAb(Cell Signalling(#2165))与HER2和EGFR的结合以证明所提供的多样化

和聚焦的肽阵列文库可用于确定抗体与不同蛋白质靶标结合的倾向。开发了一种算法。

[0436] 从抗HER2 mAb与肽阵列文库的结合确定第一组单独的肽(包括显著肽),并从抗EGFR mAb到相同的肽阵列文库确定第二组单独的肽(包括显著肽)。对于两组中的每一组鉴定富集的单独的肽(包括显著肽)、基序,并将富集的基序与相应的靶蛋白比对。使用3种水平的比对严格性进行每组基序的比对:

高度严格(精确比对)

中度严格(允许小缺口和氨基酸替换),以及

低度严格(允许更宽的缺口和氨基酸替换)。

[0437] 每个比对鉴定了每个靶标中的特定残基。在两种情况下,比对规则越严格,即随着比对严格性的增加,所得到的比对越严格,如用红色标记的残基所示(图22)。每个比对可以通过“红色”残基的数目来评分。

[0438] 图22示出了在相同的比对严格性规则下,相对于与HER2的结合,预测mAb以较小程度结合EGFR,即mAb与HER2结合比与EGFR结合具有更大的倾向。

[0439] 可以使用富集的基序确定抗体与靶蛋白结合的倾向,该富集的基序从多样化文库的单独的肽(在这种情况下为显著肽)中和/或从聚焦文库的单独的肽(在这种情况下为显著肽)中鉴定出。

[0440] 虽然本文已经示出并描述了本发明的优选实施方案,但对于本领域技术人员将显而易见的是,这些实施方案仅以示例的方式提供。本领域技术人员在不脱离本发明的情况下现将会想到多种变化、改变和替代。应当理解,本文中所述的本发明实施方案的各种替代可用于实施本发明。旨在以下权利要求限定本发明的范围,并由此涵盖这些权利要求范围内的方法和结构及其等同物。



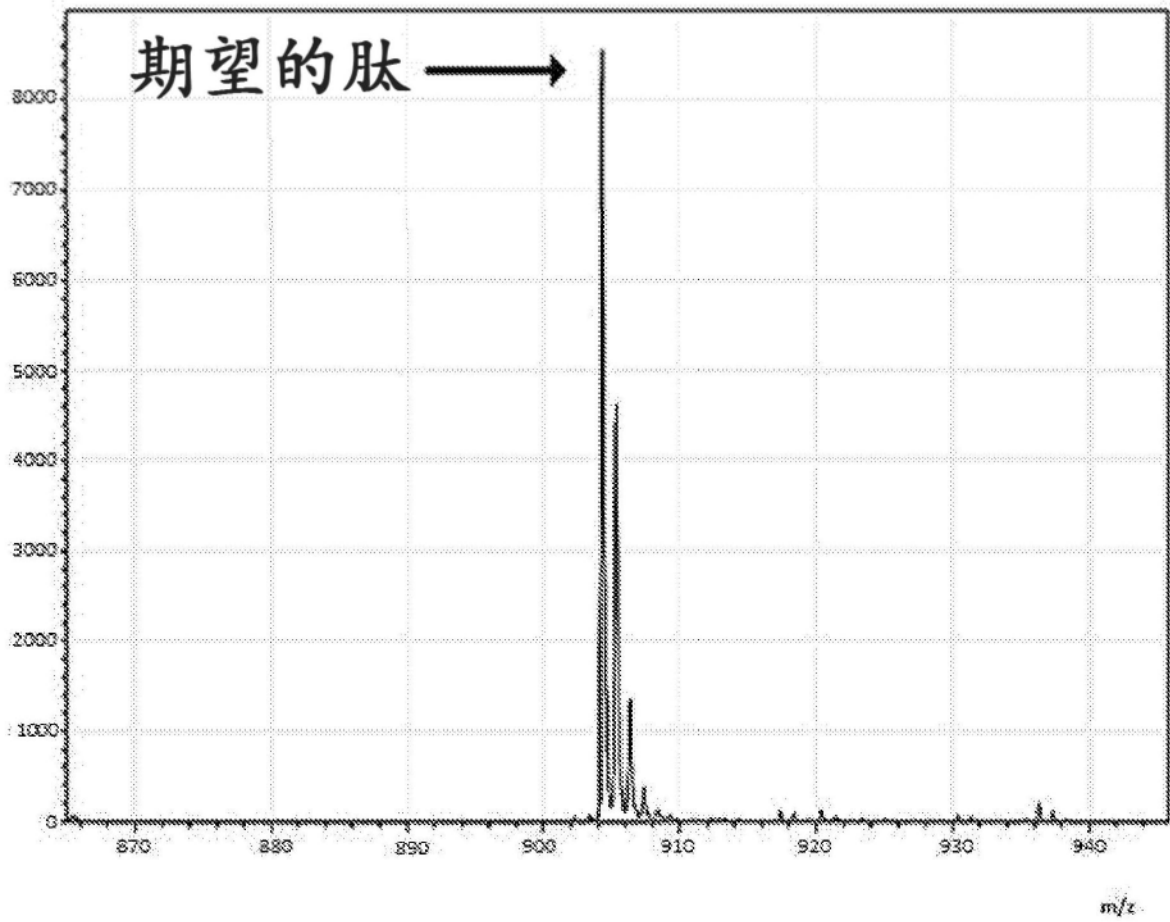


图3

### P53Ab1与表位变体结合

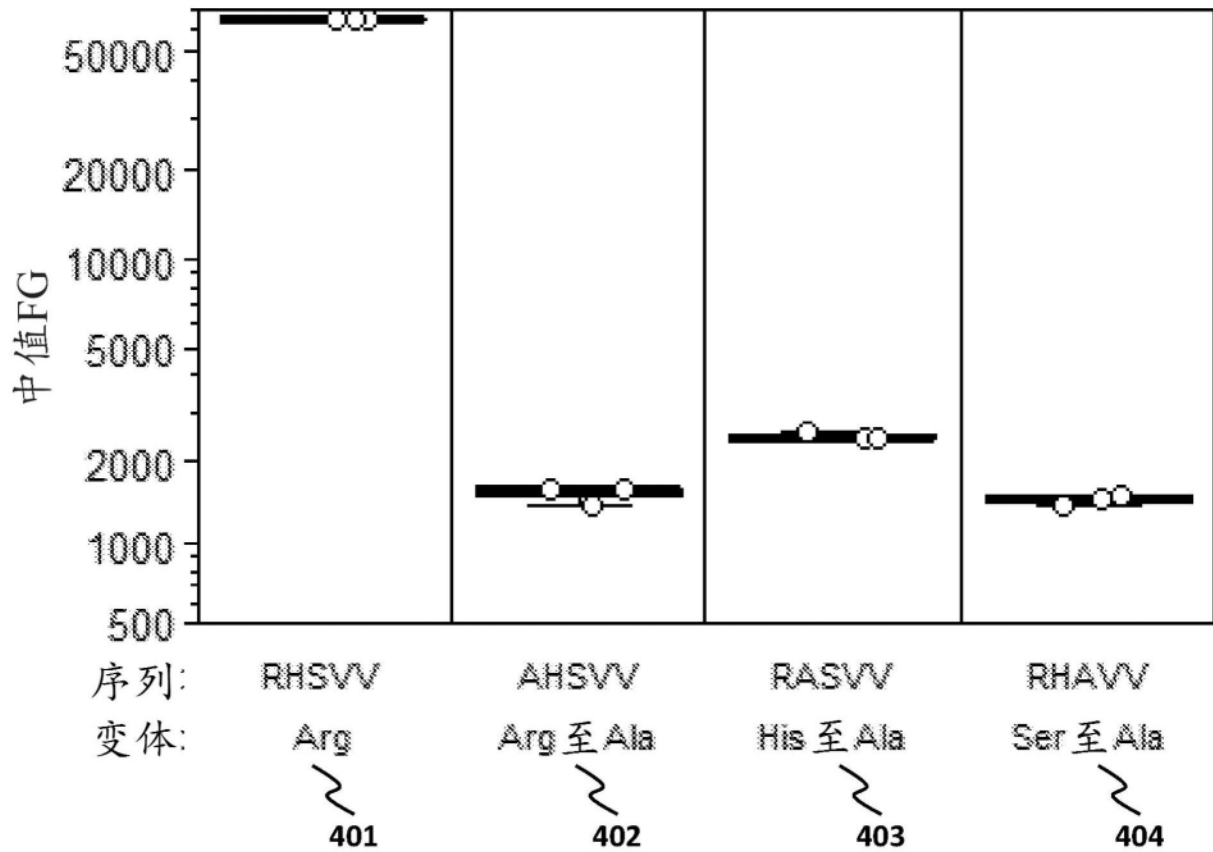


图4

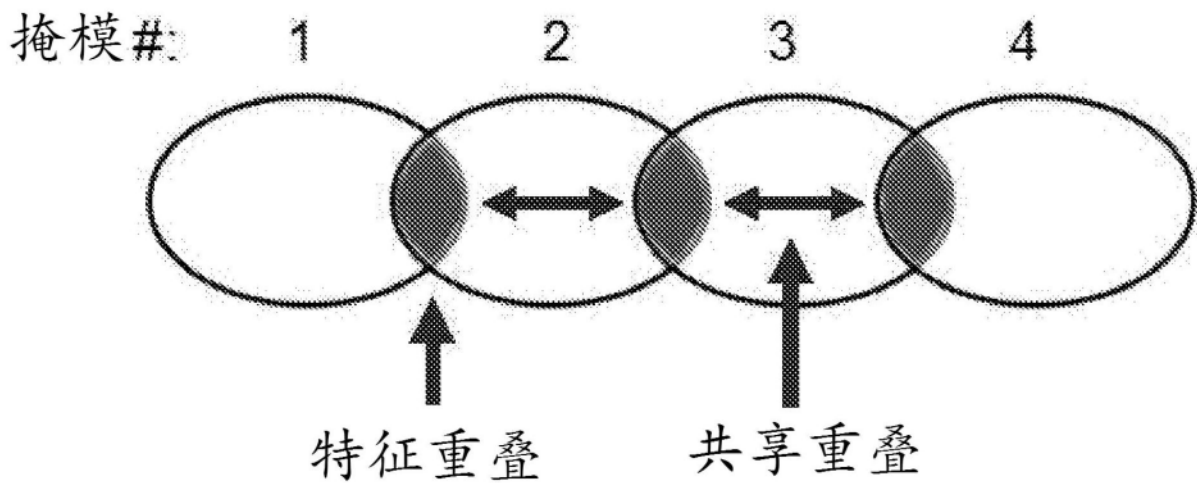


图5

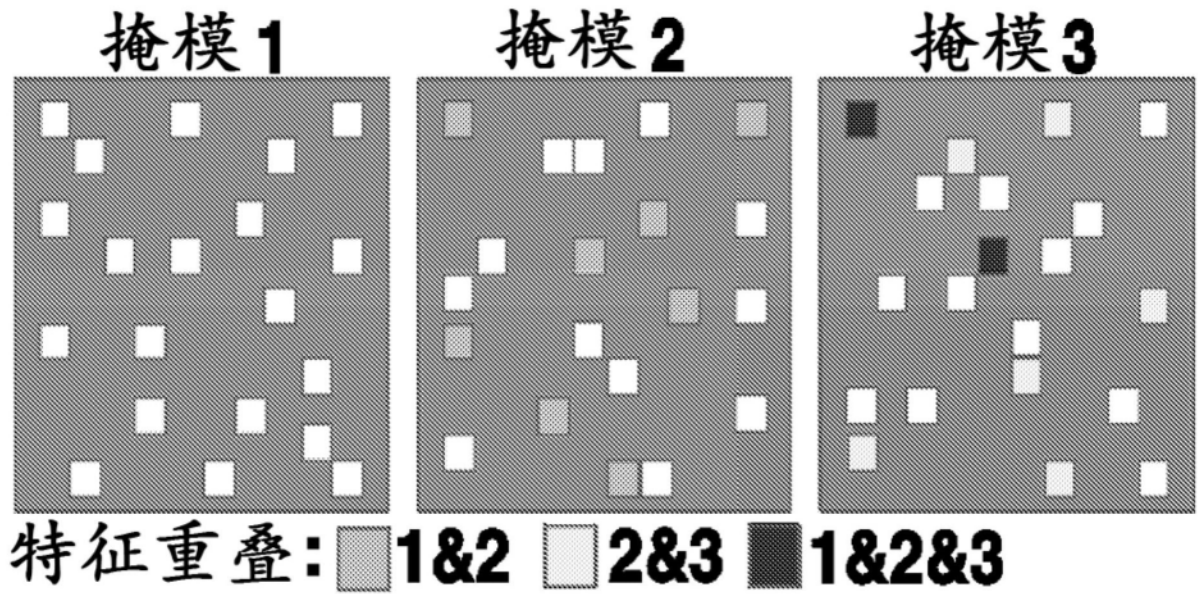


图6

输入序列(N->C): HVGAAAFVVPQA

掩模重叠号: 52

特征数目: 500,000

掩模#: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

氨基酸#: A Q P V V P A A A G V H R S W K N D M F L I E G Y

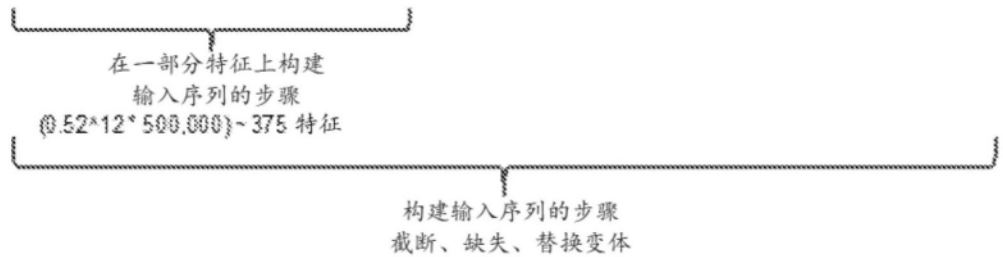


图7

### 序列长度的分布

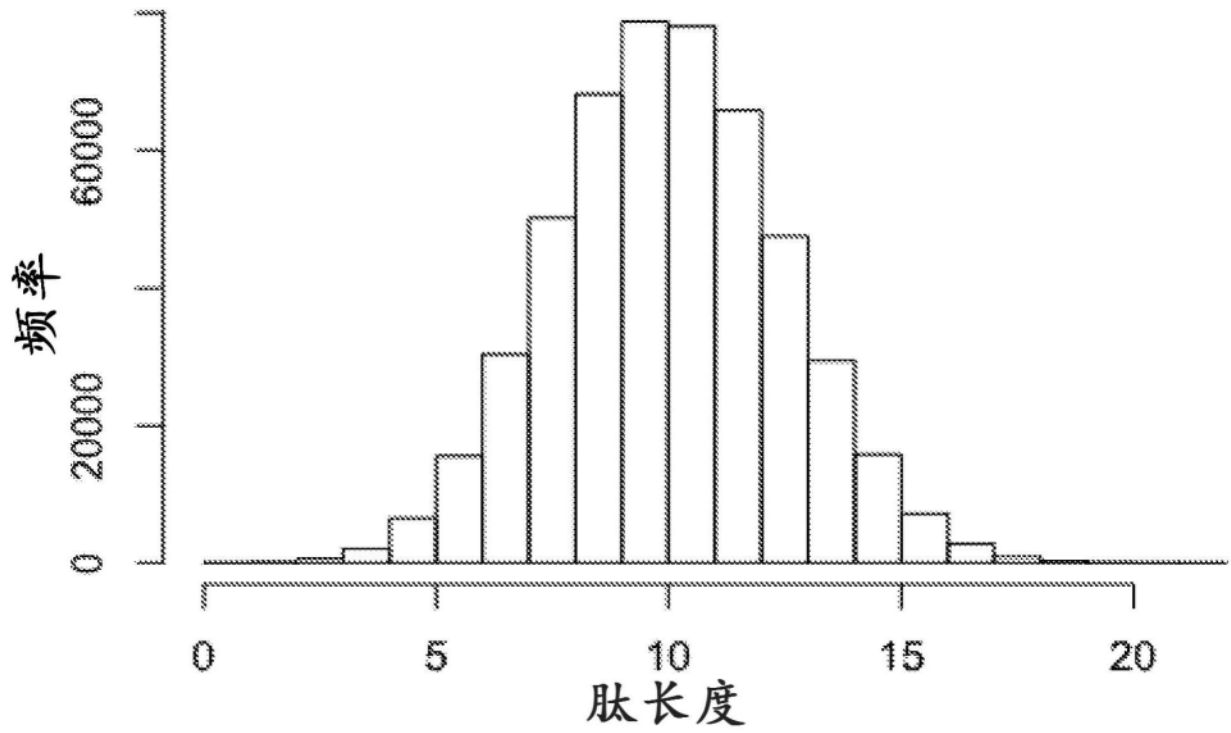


图8

序列长度的分布

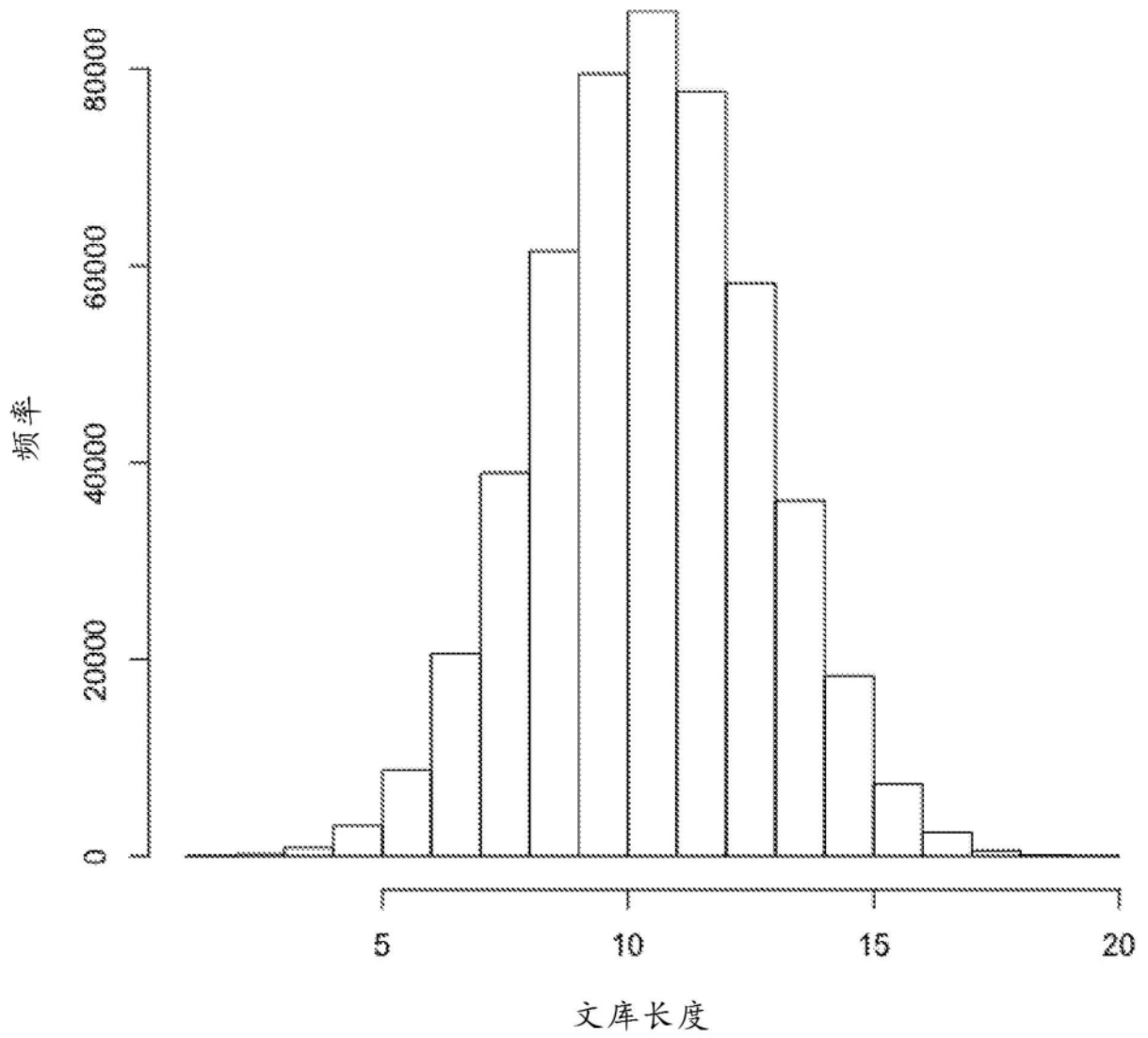


图9

### 多样化-聚焦文库策略使肽空间的采样能够表征高分辨率 和信息内容的结合

我们如何测量跨巨大肽空间的结合谱并放大生物学相关的结合序列？

跨蛋白质序列空间的Ab结合的  
广泛采样

多样化约120 k文库  
捕获3-5聚体结合基序

跨表位体序列空间（位置和氨基酸  
水平）的Ab结合的高分辨率采样

聚焦17 k文库  
捕获全表位相互作用

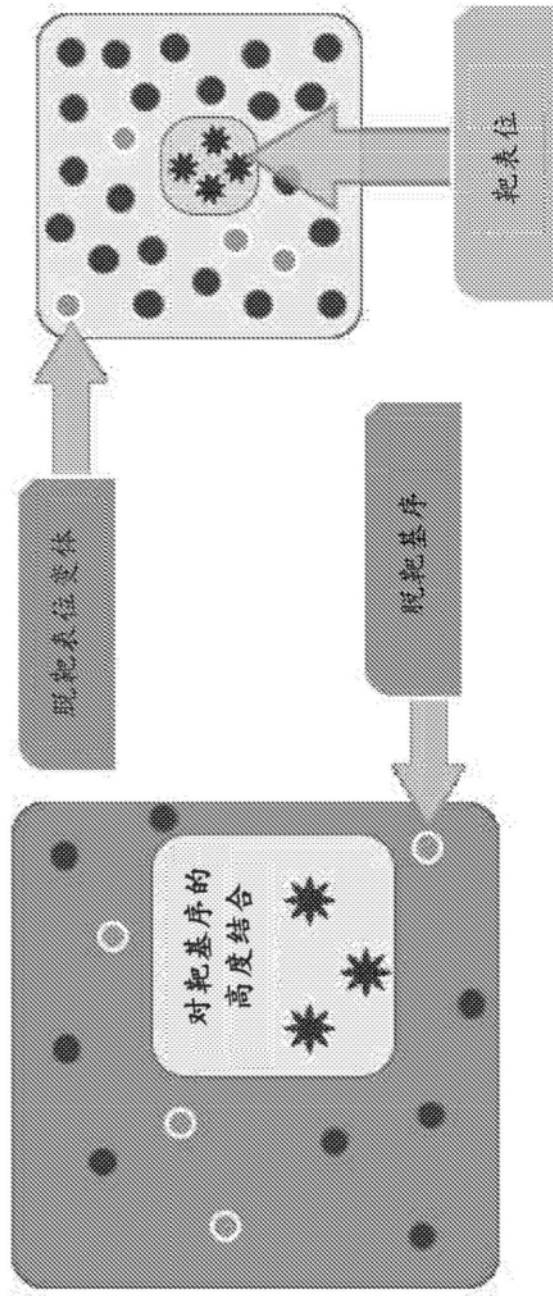


图10



已知的Ab靶序列: ...GAPPSTFKGTPTAENPEYLGLDVPV...  
 重叠6-聚体(Ex): GAPPST

APPSTF  
 PPSTFK  
 .....YLGLDV  
 LGLDVP  
 GLDVVP

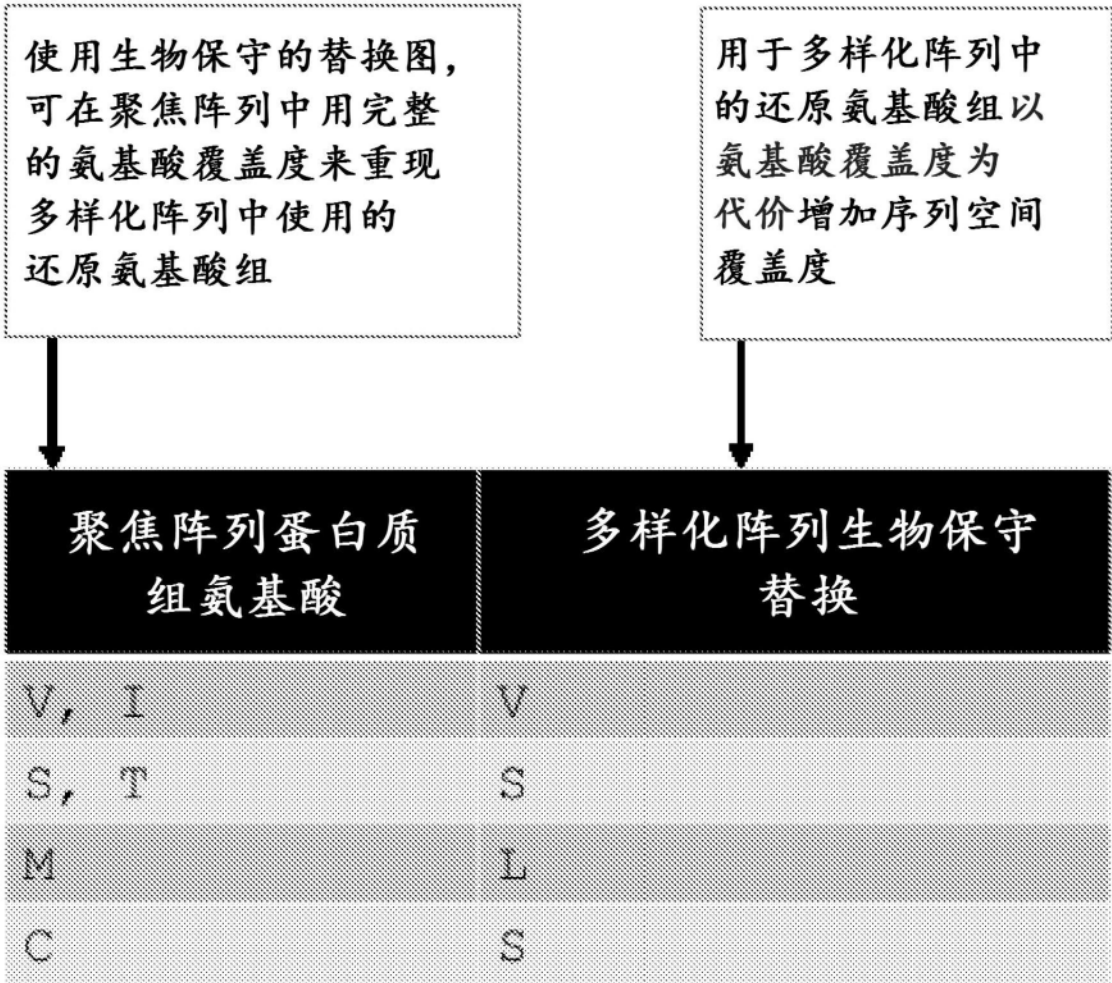
- 将显著肽3-聚体（正向和反向）与靶标6-聚体进行允许空位的比对：

靶标6-聚体: YLGLDV  
 肽3-聚体比对: Y G D 得分 = 1.5 (空位罚分)  
                   H GL 得分 = 1.5  
                   GL V 得分 = 2.5

- 将比对得分与相应的比对的显著肽结合信号组合以产生每个靶序列位置的基序命中评分度量

图12

(生物学上) 保守氨基酸替换用于在具有还原氨基酸组的肽阵列上获得mAb结合和蛋白质组序列覆盖度



全氨基酸组 ← 简化的氨基酸组

保守替换参考文献: Bioinformatics for Genetiobls, M.R. Barnes, I.C. Gray 等人, Wiley, 2003

图13

# 多样化文库最高剂量- 应答肽序列保守

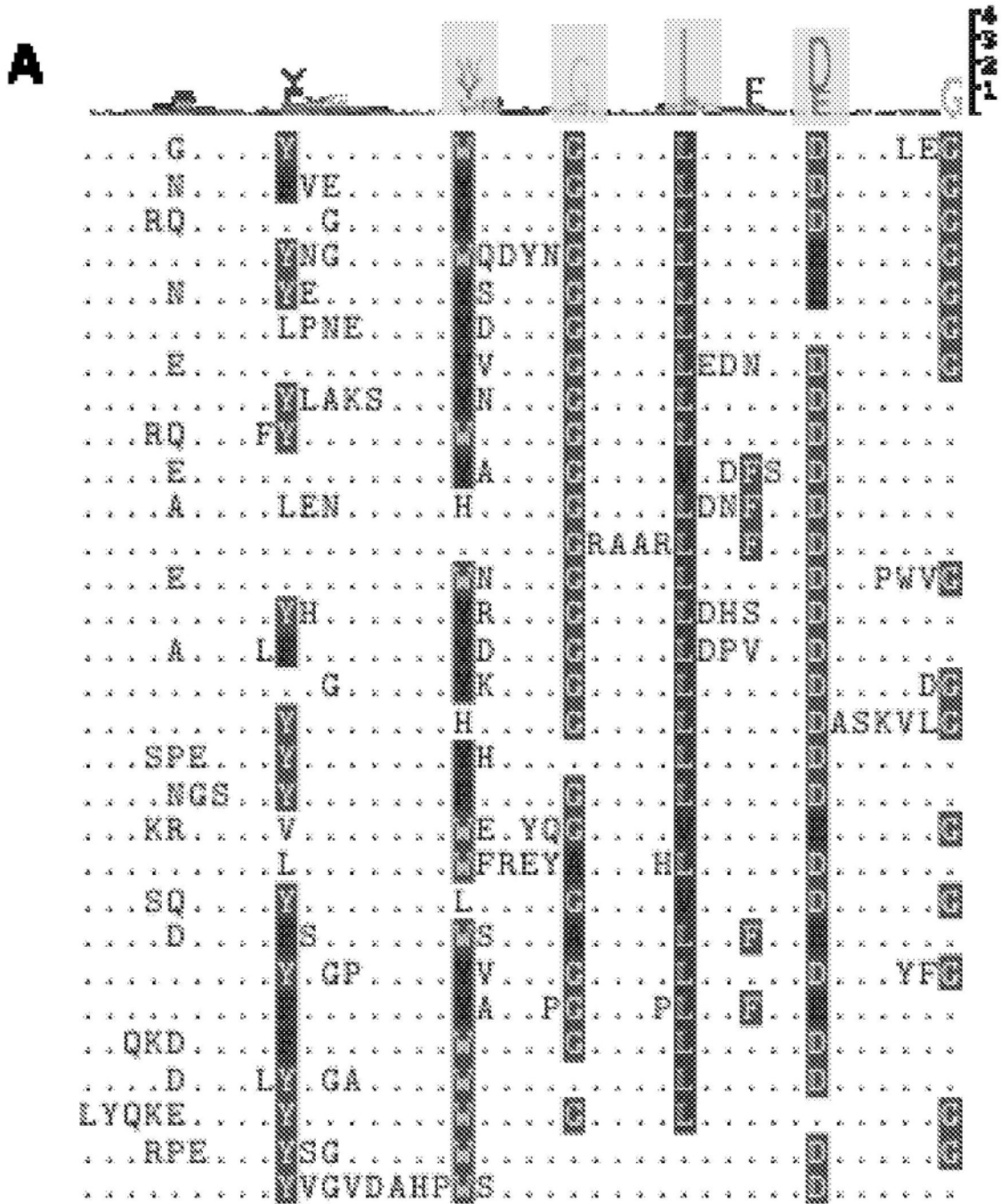


图14A

序列	比对
PTGPEYLGLED	PT-GPE--YLGLED---
APEFVGLDVPV	-A--PE--FVGLDVPV
ENFGLDLPV	-----E--NFGLDLPV
PTAENVGLDVPV	PT--AE--NVGLDVPV
PTENVGLDPV	PT---E--NVGLD-PV
PTGENYLGLEDL	PT--GEN-YLGLED--L
APEYLGLEDVPV	-A--PE--YLGLEDVPV
GAPEFVGLDLPV	-G-APE--FVGLDLPV
PTAENVGLDVPV	PT--AE--NVGLDVPV
GPEYVGLDL	-G--PE--YVGLDL--
PEFLGLDLPV	----PE--FLGLDLPV
ENYLGLED	-----EN-YLGLED---
PTGYLGLED	PT--G---YLGLED---
GAENVGLDVPV	-G--AE--NVGLDVPV
AENYVGLDPV	----AEN-YVGLD-PV
GENVGLDLPV	-G---E--NVGLDLPV
PTAENVGLDLV	PT--AE--NVGLDL-V
GYVGLDPV	-G-----YVGLD-PV
PTENVGLDV	PT---E--NVGLD--V
ENFGLD	-----E--NFGLD---
GENYVGLDV	----GEN-YVGLD--V
PEYVGLVPV	----PE--YVGL-VPV
PTAPEYVGLDPV	PT-APE--YVGLD-PV
PTENPEYLGLEDPV	PTENPE--YLGLED-PV
GAPEYLGLEDLPV	GA--PE--YLGLEDLPV

图14B

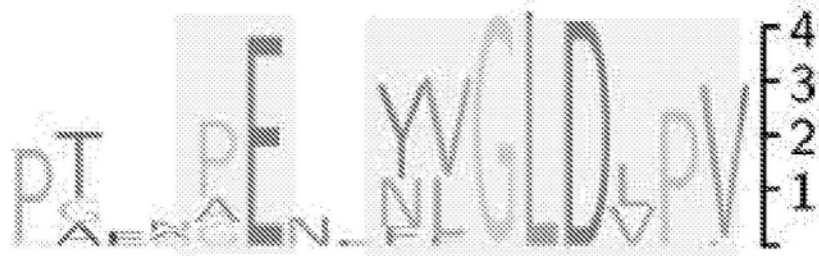


图14C

公开的 免疫原	PTAENPEYLGLDVPV
预测的表位	AENPEYLGLDV
Ab ID	MA5-13675 (克隆3B5)
供应商	Thermo

图14D



序列	比对
PTGPEYLGLD	P-TGP-EYLGLD---
APEFVGLDVPV	---AP-EFVGLDVPV
ENFGLDLFV	-----ENFGLDLFV
PTGENYLGLDL	P-TGE-NYLGLDL--
PEFLGLDLFV	----P-EFLGLDLFV
APEYLGLDVPV	---AP-EYLGLDVPV
AENYVGLDPV	---AE-NYVGLD-PV
PTAENVGLDVPV	--PTA-ENVGLDVPV
ENYLGLD	----E-NYLGLD---
GENYVGLDV	---GE-NYVGLD--V
GYVGLDLFV	---G---YVGLDLFV
PEYVGLDL	----P-EYVGLDL--
GENVGLDLFV	---G--ENVGLDLFV
PTAENVGLDVPV	--PTA-ENVGLDVPV
PTENVGLDPV	----PTENVGLD-PV
YVGLDPV	-----YVGLD-PV
ENYLGLDPV	----E-NYLGLD-PV
APEYLGLD	---AP-EYLGLD---
PTGYLGLD	----PTGYLGLD---
GYVGLDPV	---G---YVGLD-PV
GPEYVGLDL	---GP-EYVGLDL--
AENYVGLDPV	---AE-NYVGLD-PV
GYVGLDPV	---G---YVGLD-PV
APEYVGLDPV	---AP-EYVGLD-PV
APEFLGLDPV	---AP-EFLGLD-PV

图15B



图15C

公开的 免疫原	PTAENPEYLGLDVPV
预测的表位	AENPEYLGLDV
Ab ID	sc-33684 (克隆385)
供应商	Santa Cruz

图15D



序列	比对
PTPEFYGLDVPV	PT--P--E--FYGLDVPV
PTPEYLGLDVPV	PT--P--E--YLGLDVPV
AENFYLGLDVPV	----A--ENFYLGLDVPV
APEYLGLDVPV	-A--P--E--YLGLDVPV
GPEYLGLDVPV	-G--P--E--YLGLDVPV
PTGENFYLGLDVPV	----PTGENFYLGLDVPV
PTYLGLDVPV	----P--T--YLGLDVPV
GAPEYLGLDVPV	-GA-P--E--YLGLDVPV
GAPEFYVGLDVPV	-GA-P--EF-YVGLDVPV
PEFYLDVPV	----P--E--FY-LDVPV
GPEYLGLDVPV	-G--P--E--YLGLDVPV
PEYLGLDVPV	----P--E--YLGLDVPV
PTGENFYLDVPV	----PTGEN-FY-LDVPV
APEFYGLDVPV	-A--P--E--FYGLDVPV
GPEFYLDVPV	-G--P--E--FY-LDVPV
GPEYLGLDVPV	-G--P--E--YLGLDVPV
PTENFYLDVPV	----PT-EN-FY-LDVPV
PTAPEYVGLDVPV	PTA-P--E--YVGLDVPV
GPEYVGLDVPV	-G--P--E--YVGLDVPV
PTAPEYVGLDVPV	PTA-P--E--YVGLDVPV
GFLGLDVPV	-G-----FLGLDVPV
GAPEYVGLDVPV	-GA-P--E--YVGLDVPV
PTENYLGLDV	----PT-EN-YLGLDV--
FYLGLDVPV	-----F--YLGLDVPV
ENYGLDVPV	-----EN-Y-GLDVPV

图16B

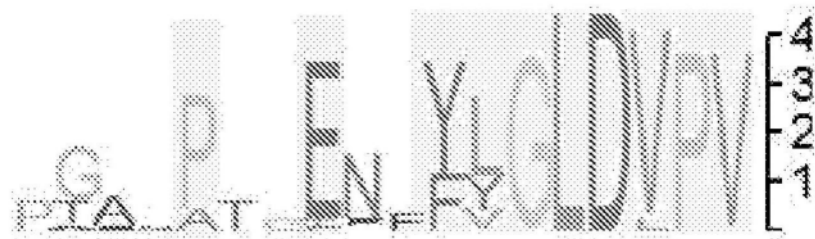


图16C

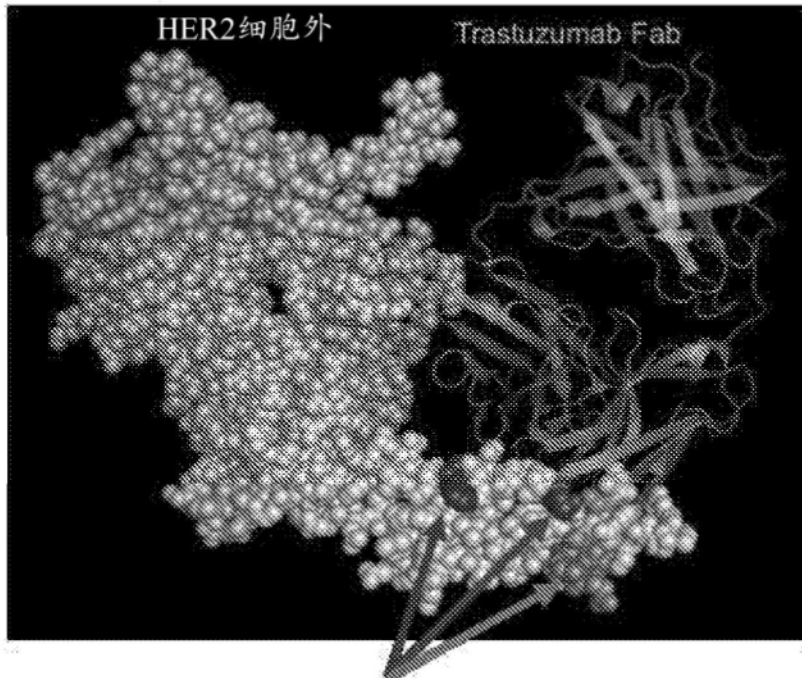
<b>公开的免疫原</b>	<b>PTAENPEYLGLDVPV</b>
<b>预测的表位</b>	<b>YLGLDVPV</b>
<b>Ab ID</b>	<b>2165 (克隆 29D8)</b>
<b>供应商 anta Cruz</b>	<b>Cell Signaling</b>

图16D



聚焦肽阵列将曲妥珠单抗结构表位表示为线性组分

551 CQPQNGSVTC FGPEADQCVA CAHYKDPPFC VARCPGVKVP DLSYMPIWKF 600  
601 PDEEGACOPE PINCTHSCVD LDDKGCPAEQ DIENLYFQGA MAAAARGGPE 650



阵列鉴定出映射到结构的顶部线性组分

图18

聚焦肽阵列将HER2 (a.k.a erbB-2) 鉴定为来自BLAST  
搜索的cell signaling 2165 (克隆29D8) mAb

**A**

**最高10个剂量-应答结合肽的  
BLAST比对搜索**

Database: All non-redundant GenBank CDS  
translations+PDB+SwissProt+PIR+PRF excluding environmental samples  
from WGS projects  
108,789,579 sequences; 39,865,587,614 total letters

Query= 1  
Length=13

Sequences producing significant alignments:	Score (Bits)	E Value
AA018082.1 v-erb-b2 erythroblastic leukemia viral oncogene ho...	44.8	5e-07
AAA75493.1 HER2 receptor [Homo sapiens]	44.8	5e-07
NP_004439.2 receptor tyrosine-protein kinase erbB-2 isoform a...	44.8	5e-07
BAG58195.1 unnamed protein product [Homo sapiens]	44.8	5e-07
NP_001276865.1 receptor tyrosine-protein kinase erbB-2 sofor...	44.8	5e-07
NP_001005862.1 receptor tyrosine-protein kinase erbB-2 sofor...	44.8	5e-07

**B**

**中值10个剂量-应答结合肽  
的BLAST比对搜索**

Database: All non-redundant GenBank CDS  
translations+PDB+SwissProt+PIR+PRF excluding environmental samples  
from WGS projects  
108,789,579 sequences; 39,865,587,614 total letters

Query= 1  
Length=8

Sequences producing significant alignments:	Score (Bits)	E Value
NP_00110441.1 E3 ubiquitin-protein ligase ANFR isoform 8 [Ho...	22.3	10
AA056722.1 autocrine motility factor receptor [Homo sapiens]	22.3	10
NP_001135.3 E3 ubiquitin-protein ligase ANFR isoform c [Homo ...	22.3	10
EA002855.1 hCG2811773, isoform TR8_a [Homo sapiens]	22.3	10
NP_00110440.1 E3 ubiquitin-protein ligase ANFR isoform e [Ho...	22.3	10
AA017043.1 ANFR protein [Homo sapiens]	22.3	10

图19

### 聚焦肽阵列将HER2 (a.k.a erbB-2) 鉴定为来自BLAST 搜索的Thermo MA5-13675 (克隆3B5) mAb靶标

**A**

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects  
 108,789,579 sequences; 39,865,587,614 total letters  
 Query= 1  
 Length=7

Sequences producing significant alignments:	Score (bits)	E Value
AA018082.1 v-erb-b2 erythroblastic leukemia viral oncogene ho...	21.8	20
AAA75493.1 HER2 receptor [Homo sapiens]	21.8	20
NP_004439.2 receptor tyrosine-protein kinase erbB-2 isoform a...	21.8	20
BAG58195.1 unnamed protein product [Homo sapiens]	21.8	20
NP_001276865.1 receptor tyrosine-protein kinase erbB-2 isofo...	21.8	20
NP_001005862.1 receptor tyrosine-protein kinase erbB-2 isofo...	21.8	20

**B**

#### 中值10个剂量-应答结合肽的 BLAST比对搜索

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects  
 108,789,579 sequences; 39,865,587,614 total letters  
 Query= 1  
 Length=7

Sequences producing significant alignments:	Score (bits)	E Value
EAW76813.1 AC02810008, isoform CRA.a [Homo sapiens]	22.7	9.8
NP_001337981.1 rfa GTPase-activating protein 48 [Homo sapiens]	22.7	9.8
Q97C36.3 Sequence: full-length GTPase-activating protein 48; AITN...	22.7	9.8
EXE6654.1 AC02810008, isoform CRA.b [Homo sapiens]	22.7	9.8
Q511859432 Q51180813.4 protein [isolated] - human	22.7	9.8
Q271838918 hypothetical protein 2 - human (fragment)	20.6	57

图20

## 聚焦肽阵列将HER2 (a.k.a erbB-2) 鉴定为来自BLAST搜索的Santa Cruz SC-33684 (克隆3B5) mAb靶标

**A**

```

Database: All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF excluding environmental samples
from WGS projects
      108,789,579 sequences; 39,865,587,614 total letters
Query= 1
Length=7

Sequences producing significant alignments:
                                     Score      E
                                     (Bits)    Value
AA018082.1  v-erb-b2 erythroblastic leukemia viral oncogene ho...  19.7      115
AAA75493.1  HER2 receptor [Homo sapiens]                               19.7      115
NP_004439.2  receptor tyrosine-protein kinase erbB-2 isoform a...    19.7      115
BAG58195.1  unnamed protein product [Homo sapiens]                   19.7      115
NP_001276865.1  receptor tyrosine-protein kinase erbB-2 isofor...    19.7      115
NP_001005862.1  receptor tyrosine-protein kinase erbB-2 isofor...    19.7      115
    
```

**B**

### 中值10个剂量-应答结合肽的 BLAST比对搜索

```

Database: All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF excluding environmental samples
from WGS projects
      108,789,579 sequences; 39,865,587,614 total letters
Query= 3
Length=8

Sequences producing significant alignments:
                                     Score      E
                                     (Bits)    Value
EAW08140.1  regeneration associated serine protease [Homo sapien...  21.8      35
NP_856345.2  inactive serine protease FAM61 isoform a precursor...  21.8      35
BAG55484.1  unnamed protein product [Homo sapiens]                   21.8      35
NP_001003991.1  inactive serine protease FAM61 isoform b precursor...  21.8      35
CA144383.1  hypothetical protein [Homo sapiens]                       21.8      35
CAG45900.1  hypothetical protein [Homo sapiens]                       21.8      35
    
```

图21

