(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119574873 A
(43)申请公布日 2025.03.07

(21)申请号 202411182266.2

(22)申请日 2017.05.02

(30)优先权数据
62/330,841 2016.05.02 US
62/339,071 2016.05.19 US
62/376,886 2016.08.18 US

(62)分案原申请数据
201780041157.0 2017.05.02

(71)申请人 ENCODIA公司
地址 美国加利福尼亚州

(72)发明人 马克·朱 凯文·冈德森
迈克尔·菲利普·韦纳

(74)专利代理机构 北京唐颂永信知识产权代理有限公司 11755
专利代理师 刘伟

(51)Int.Cl.
G01N 33/68(2006.01)
C12Q 1/6869(2018.01)
C40B 20/04(2006.01)
C40B 40/04(2006.01)
C40B 70/00(2006.01)

权利要求书3页 说明书116页
序列表（电子公布） 附图70页

(54)发明名称

采用核酸编码的大分子分析

(57)摘要

公开了采用核苷酸编码分析大分子，包括肽、多肽和蛋白质的方法。

CN 119574873 A

1.一种分析肽的方法,其包括以下步骤:

(a)使所述肽或所述肽的组分与多种结合剂接触,每种结合剂能够与所述肽或所述肽的组分结合,其中所述肽或所述肽的组分连接到核酸记录标签,所述核酸记录标签包含所述肽的条形码,并且其中所述多种结合剂的每一种包含核酸编码标签,所述核酸编码标签包含含有关于所述结合剂的识别信息的编码器序列;

(b)在所述结合剂与所述肽或所述肽的组分结合后,通过连接和/或引物延伸在所述结合剂的核酸编码标签与所述肽或所述肽的组分的核酸记录标签之间,传递所述编码器序列和/或所述肽的条形码的序列信息,从而生成包含所述编码器序列和所述肽的条形码的序列信息的延伸核酸构建体;

(c)重复步骤(a)-(b)以对所述肽或所述肽的其它组分进行一个或多个额外的结合周期;以及

(d)分析在步骤(a)-(c)中生成的延伸核酸构建体,其中所述分析包括核酸测序,以及获得在每个结合周期中与所述肽和/或所述肽的组分结合的结合剂的识别信息,从而分析所述肽。

2.根据权利要求1所述的方法,其中分析在步骤(a)-(c)中生成的延伸核酸构建体包括:i)将生成的延伸核酸构建体连接成一个或多个长核酸分子,每个长核酸分子包含两个或两个以上不同的延伸核酸构建体;以及ii)对一个或多个长核酸分子测序,并获得在每个结合周期中与所述肽和/或所述肽的组分结合的结合剂的识别信息。

3.根据权利要求1所述的方法,其中所述结合剂的每个核酸编码标签包含结合周期特异性条形码,所述条形码对于给定结合周期内使用的所有核酸编码标签都是相同的,并且在重复步骤(a)-(b)时,所述条形码与其他结合周期内使用的结合周期特异性条形码相同或不同。

4.根据权利要求1所述的方法,其中在步骤(b)中,将所述肽或其互补体的条形码传递到所述结合剂的核酸编码标签,从而生成包含编码器序列和所述肽或其互补体的条形码的延伸核酸编码标签构建体。

5.根据权利要求1所述的方法,其中在步骤(b)中,将所述编码器序列或其互补体传递到所述肽或所述肽的组分的核酸记录标签,从而生成包含所述肽的条形码和编码器序列或其互补体的延伸核酸记录标签构建体。

6.根据权利要求1所述的方法,其中在步骤(b)中,所述编码器序列和所述肽的条形码分别传递到单独的核酸构建体中,从而生成包含编码器序列和所述肽的条形码的延伸核酸构建体。

7.一种核酸文库,其包含关于多个肽的信息,其中所述核酸文库包含至少1000000个核酸分子,所述核酸文库中的每个核酸分子都是非天然存在的,并包含:

(i)至少三个不同的编码器条形码或其互补体;以及

(ii)独特分子标识符(UMI),和/或用于引物扩增和/或核酸文库测序的通用引发位点;

其中每个编码器条形码提供了与所述多个肽中的一个或多个肽的组分结合的结合剂的标识符标签。

8.一种处理多个多肽的方法,其包括以下步骤:

(a)在多个隔室中提供多个标记多肽,其中(i)多个标记多肽的每一个用核酸标签标

记;(ii)每个隔室包含多个核酸隔室标签,每个核酸隔室标签包含条形码,所述条形码在单个隔室内相同,且与其他隔室的核酸隔室标签的条形码不同;

(b)对于存在于所述隔室中的每个标记多肽,将所述核酸隔室标签或其互补体传递到所述多肽的每个核酸标签,从而生成多个记录标签标记的多肽,其中所述传递是通过引物延伸和/或连接进行的,并且每个记录标签标记的多肽包含一个或多个记录标签,每个记录标签包含条形码或其互补体;以及

(c)将所述多个记录标签标记的多肽片段化,以获得多个记录标签标记的肽,其中每个记录标签标记的肽包含记录标签,所述记录标签包含条形码或其互补体。

9.一种处理多肽的方法,其包括以下步骤:

(a)在隔室中提供标记多肽,其中所述标记多肽具有(i)核酸条形码或独特分子标识符(UMI)标签;和(ii)一个或多个不同于核酸条形码或UMI标签的核酸标签;

(b)将所述条形码或所述UMI标签或其互补体传递到与所述多肽结合的一个或多个核酸标签,其中所述传递是通过引物延伸和/或连接进行的,从而生成记录标签标记的多肽,所述记录标签标记的多肽包含两个或多个以上单独的记录标签,每个记录标签包含条形码或UMI标签或其互补体;以及

(c)将所述记录标签标记的多肽片段化,以获得多个记录标签标记的肽,其中每个记录标签标记的肽包含条形码或UMI标签或其互补体。

10.一种处理多个多肽的方法,其包括以下步骤:

(a)在多个隔室中提供多个标记多肽,其中多个标记多肽的每一个具有(i)核酸条形码或独特分子标识符(UMI)标签;和(ii)一个或多个核酸标签,其中每个隔室平均包含少于一个多肽;

(b)对于每个标记多肽,将所述条形码或所述UMI标签或其互补体传递到一个或多个核酸标签,其中所述传递是通过引物延伸和/或连接进行的,从而生成多个记录标签标记的多肽,其中每个记录标签标记的多肽包含两个或多个以上单独的记录标签,每个记录标签包含条形码或UMI标签或其互补体;以及

(c)将所述多个记录标签标记的多肽片段化,以获得多个记录标签标记的肽,其中每个记录标签标记的肽包含条形码或UMI标签或其互补体。

11.一种分析多肽的方法,其包括以下步骤:

(a)将所述多肽片段化,以提供包含第一片段和第二片段的多肽的两个或多个片段,其中所述多肽的第一片段和第二片段分别连接到核酸记录标签,并且其中所述核酸记录标签包含连接所述多肽的共同条形码序列;

(b)获得所述多肽的每个第一片段和第二片段的至少部分序列信息;

(c)利用所述共同条形码序列,将步骤(b)中获得的所述多肽的第一片段的至少部分序列信息和所述多肽的第二片段的至少部分序列信息组合,从而确定所述多肽的序列。

12.一种分析多个多肽的方法,其包括以下步骤:

(a)将所述多个多肽分隔在多个隔室中,其中每个隔室包含多个核酸隔室标签,每个核酸隔室标签包含条形码序列,所述条形码序列在单个隔室内是相同的,且不同于其他隔室的条形码序列;

(b)将所述多个核酸隔室标签连接到所述每个隔室的多肽,从而在每个隔室内生产多

个标记多肽;

(c)将所述每个隔室中的多个标记多肽进行片段化,以为所述多个标记多肽的标记多肽提供第一片段和第二片段,其中所述标记多肽的第一片段和第二片段分别连接到包含共同条形码序列的核酸隔室标签;

(d)获得所述标记多肽的第一片段和第二片段的至少部分序列信息;

(e)利用所述共同条形码序列,将步骤(d)中获得的所述标记多肽的第一片段的至少部分序列信息和所述标记多肽的第二片段的至少部分序列信息组合,从而确定所述标记多肽的序列。

采用核酸编码的大分子分析

[0001] 本案是申请日为2017年5月2日,发明名称为采用核酸编码的大分子分析,申请号为201780041157.0的案件的分案申请。

[0002] 序列表声明

[0003] 与本申请相关的序列表以文本格式提供以代替纸质副本,并且通过引用并入到本说明书中。包含序列表的文本文件的名称是760229_401WO_SEQUENCE_LISTING.txt。该文本文件为38.7KB,于2017年5月2日创建,并通过EFS-Web以电子方式提交。

技术领域

[0004] 本公开内容一般涉及使用条形码和分子识别事件的核酸编码来分析大分子,包括肽、多肽和蛋白质。

[0005] 相关技术说明编码标签

[0006] 蛋白质在细胞生物学和生理学中发挥着不可或缺的作用,扮演和促进许多不同的生物学功能。由于翻译后修饰(PTM)引入的额外多样性,不同蛋白质分子的集合非常丰富,比转录组复杂得多。另外,细胞内的蛋白质动态地改变(在表达水平和修饰状态中)以响应于环境、生理状态和疾病状态。因此,蛋白质含有大量未经探索的相关信息,特别是与基因组信息有关的信息。总的来说,相对于基因组学分析,蛋白质组学分析方面的创新一直滞后。在基因组学领域,新一代测序(NGS)已经通过在单个仪器运行中分析数十亿个DNA序列改变了这一领域,而在蛋白质分析和肽测序中,通量仍然有限。

[0007] 然而,迫切需要这种蛋白质信息,以便更好地了解健康和疾病中的蛋白质组动态,并帮助实现精准医学。因此,人们对开发“下一代”工具以使这种蛋白质组信息的收集小型化和高度并行化非常感兴趣。

[0008] 由于各种原因,高度平行的大分子表征和蛋白质识别具有挑战性。由于存在一些关键挑战,使用基于亲和力的测定通常是困难的。一个明显的挑战是将一组亲和试剂的读数多路复用到一组同源大分子集合中;另一个挑战是使亲和试剂和脱靶大分子之间的交叉反应最小化;第三个挑战是开发高效的高通量读出平台。该问题的一个例子发生在蛋白质组学中,其中一个目标是鉴定和定量样品中的大多数或所有蛋白质。此外,期望在单分子水平上表征蛋白质上的各种翻译后修饰(PTM)。目前,这是以高通量方式去完成的艰巨任务。

[0009] 蛋白质或肽大分子的分子识别和表征通常使用免疫测定法进行。有许多不同的免疫测定形式,包括ELISA、多路复用ELISA(例如,斑点抗体阵列、液体粒子ELISA阵列)、数字ELISA(例如,Quanterix、Singulex),反相蛋白质阵列(RPPA)等等。这些不同的免疫测定平台都面临着类似的挑战,包括高亲和力和高特异性(或选择性)抗体(结合剂)的开发,在样品和分析物水平上复用的能力有限,灵敏度和动态范围有限,以及交叉反应性和背景信号。结合剂不可知方法,例如通过肽测序(Edman降解或质谱)的直接蛋白质表征提供了有用的替代方法。但是,这些方法都不是非常平行或高通量。

[0010] 基于Edman降解的肽测序首先由Pehr Edman于1950年提出;即通过一系列化学修饰和下游HPLC分析(后来被质谱分析取代)逐步降解肽上的N-末端氨基酸。第一步,在轻度

碱性条件下(NMP/甲醇/H_2O)用异硫氰酸苯酯(PITC)修饰N-末端氨基酸,形成苯基硫代氨基甲酰基(PTC)衍生物。在第二步中,用酸(无水TFA)处理PTC修饰的氨基以产生裂解的环状ATZ(2-苯胺基-5(4)-硫代唑啉酮)修饰的氨基酸,在肽上留下新的N-末端。所述裂解的环状ATZ-氨基酸被转化为PTH-氨基酸衍生物,并通过反相HPLC分析。该过程以迭代方式继续进行,直至从N-末端除去所有或部分数目的包含肽序列的氨基酸并鉴定。通常,Edman降解肽测序是缓慢的并且每天仅有少量肽的有限通量。

[0011] 在过去的10-15年中,使用MALDI、电喷雾质谱(MS)和LC-MS/MS的肽分析在很大程度上取代了Edman降解。尽管MS仪器(Riley et al.,2016,Cell Syst 2:142-143)最近取得了进展,但MS仍然存在一些缺点,包括仪器成本高、对用户的技术水平要求高、定量能力差并且跨越蛋白质组动态范的测量能力有限。例如,由于蛋白质电离效率水平不同,因此样本之间的绝对定量甚至相对定量都具有挑战性。采用质量标签有助于改善相对定量,但需要标记蛋白质组。动态范围是另外一个的难点,其中样品中蛋白质的浓度可以在很大范围内变化(血浆超过10个数量级)。MS通常仅分析高丰度种类,使低丰度蛋白质的表征非常挑战。最后,样品通量通常限于每次运行几千个肽,而对于数据独立分析(DIA),这种通量不足以进行真正的自下而上的高通量蛋白质组分析。此外,对于每个样品记录的数千个复杂MS谱的去卷积存在显着的计算要求。

[0012] 因此,本领域仍需要与大分子测序和/或分析有关的改进技术,应用于蛋白质测序和/或分析,以及用于实现其的产品、方法和试剂盒。需要高度并行化,准确,灵敏和高通量的蛋白质组学技术。本发明提供了具有这些优势及其它益处的方法和设备。

[0013] 参考以下详细描述,本发明的各方面将是明显的。为此,本文阐述了各种参考文献,其更详细地描述了某些背景信息、工艺、化合物和/或组合物,并且每个都通过引用整体并入本文。

发明内容

[0014] 本公开的实施方案总体上涉及高度平行的、高通量数字大分子分析,特别是肽分析的方法。

[0015] 在第一个实施方案中是分析大分子的方法,包括以下步骤:

[0016] (a)提供大分子和连接在固体支持物上的相关记录标签;

[0017] (b)使所述大分子与能够结合所述大分子的第一结合剂接触,其中所述第一结合剂包含具有关于所述第一结合剂的识别信息的第一编码标签;

[0018] (c)将所述第一编码标签的信息传递到所述记录标签,以产生第一次序延伸记录标签;

[0019] (d)使所述大分子与能够结合所述大分子的第二结合剂接触,其中所述第二结合剂包含具有所述第二结合剂的识别信息的第二编码标签;

[0020] (e)将所述第二编码标签的信息传送到第一次序延伸记录标签以产生第二次序延伸记录标签;以及

[0021] (f)分析所述第二次序延伸记录标签。

[0022] 在第二实施方案中是第一实施方案的方法,其中接触步骤(b)和(d)按顺序进行。

[0023] 在第三实施方案中是第一实施方案的方法,其中接触步骤(b)和(d)同时进行。

[0024] 在第四实施例中是第一实施例的方法,还包括在步骤(e)和(f)之间的以下步骤:

[0025] (x)通过采用能够结合所述大分子的第三(或更高次序)结合剂替换所述第二结合剂以重复步骤(d)和(e)一次或多次,其中所述第三(或更高次序)结合剂包含具有第三(或更高次序)结合剂的识别信息的第三(或更高次序的)编码标签;和

[0026] (y)将所述第三(或更高次序)编码标签的信息传送到所述第二(或更高次序)延伸记录标签,以产生第三(或更高次序)延伸记录标签;

[0027] 并且其中在步骤(f)中分析所述第三(或更高次序)延伸记录标签。

[0028] 在第五个实施方案中是分析大分子的方法,包括以下步骤:

[0029] (a)提供大分子,相关的第一记录标签和连接到固体支持物的相关第二记录标签;

[0030] (b)使所述大分子与能够结合所述大分子的第一结合剂接触,其中所述第一结合剂包含具有关于所述第一结合剂的识别信息的第一编码标签;

[0031] (c)将所述第一编码标签的信息传送到所述第一记录标签以产生第一次序延伸记录标签;

[0032] (d)使所述大分子与能够结合所述大分子的第二结合剂接触,其中所述第二结合剂包含具有所述第二结合剂的识别信息的第二编码标签;

[0033] (e)将所述第二编码标签的信息传送到所述第二记录标签以产生第二次序延伸记录标签;并且

[0034] (f)分析所述第一和第二延伸记录标签。

[0035] 在第六实施方案中是第五实施方案的方法,其中接触步骤(b)和(d)按先后顺序进行。

[0036] 在第七实施方案中是第五实施方案的方法,其中接触步骤(b)和(d)同时进行。

[0037] 在第八实施方案中是第五实施方案的方法,其中步骤(a)还包括提供与固体支持物连接的相关的第三(或更高次序段)记录标签。

[0038] 在第九实施例中是第八实施例的方法,还包括在步骤(e)和(f)之间的以下步骤:

[0039] (x)通过采用能够结合所述大分子的第三(或更高次序)结合剂替换所述第二结合剂以重复步骤(d)和(e)一次或多次,其中所述第三(或更高次序)结合剂包含具有第三(或更高次序)结合剂的识别信息的第三(或更高次序的)编码标签;和

[0040] (y)将所述第三(或更高次序)编码标签的信息传送到所述第三(或更高次序)记录标签,以产生第三(或更高次序)延伸记录标签;

[0041] 并且其中在步骤(f)中分析第一,第二和第三(或更高次序)延伸记录标签。

[0042] 在第10个实施方案中,是第五至第九实施方案中任一个的方法,其中所述第一编码标签,第二编码标签和任何更高次序编码标签包含结合的循环特异性间隔序列。

[0043] 在第11个实施方案中,是分析肽的方法,包括步骤:

[0044] (a)提供肽和连接在固体支持物上的相关的记录标签;

[0045] (b)用化学试剂修饰所述肽的N-末端氨基酸(NTAA);

[0046] (c)使所述肽与能够结合所述修饰的NTAA的第一结合试剂接触,其中所述第一结合试剂包含具有所述第一结合试剂的识别信息的第一编码标签;

[0047] (d)将所述第一编码标签的信息传送到所述记录标签以产生一个延伸记录标签;并且

[0048] (e)分析所述延伸记录标签。

[0049] 在第12个实施方案中是第11个实施方案的方法,其中步骤(c)还包括使所述肽与包含第二(或更高次序)编码标签的第二(或更高次序)结合剂接触,所述第二(或更高次序)编码标签具有第二(或更高次序)结合剂的识别信息,其中第二(或更高次序)结合剂能够结合除步骤(b)的修饰的NTAA之外的修饰的NTAA。

[0050] 在第13个实施方案中是第12个实施方案的方法,其中在所述肽与所述第一结合剂接触后,接着使所述肽与第二(或更高级)结合剂接触。

[0051] 在第14个实施方案中是第12个实施方案的方法,其中使所述肽与第二(或更高级)结合剂接触和所述肽与所述第一结合剂的接触同时发生。

[0052] 在第15个实施方案中是第11-14实施方案中任一个的方法,其中所述化学试剂是异硫氰酸酯衍生物,2,4-二硝基苯磺酸(DNBS),4-磺酰基-2-硝基氟苯(SNFB)1-氟-2,4-二硝基苯、丹磺酰氯、7-甲氧基香豆素乙酸、硫代酰化试剂、硫代乙酰化试剂或硫代苄基化试剂。

[0053] 在第16实施方案中是分析肽的方法,包括以下步骤:

[0054] (a)提供肽和连接在固体支持物上的相关的记录标签;

[0055] (b)用化学试剂修饰所述肽的N-末端氨基酸(NTAA)以产生修饰的NTAA;

[0056] (c)使所述肽与能够结合所述修饰的NTAA的第一结合试剂接触,其中所述第一结合试剂包含具有所述第一结合试剂的识别信息的第一编码标签;

[0057] (d)将第一编码标签的信息传递到所述记录标签以产生第一次序延伸记录标签;

[0058] (e)移除修饰的NTAA以暴露新的NTAA;

[0059] (f)采用化学试剂修饰所述新的NTAA以产生新的修饰的NTAA;

[0060] (g)使所述肽与能够结合所述修饰的NTAA的第二结合剂接触,其中所述第二结合剂包含具有所述第二结合剂的识别信息的第二编码标签;

[0061] (h)将所述第二编码标签的信息传送到所述第一次序延伸记录标签以产生第二次序延伸记录标签;以及

[0062] (i)分析所述第二延伸记录标签。

[0063] 在第17实施方案中是分析肽的方法,包括以下步骤:

[0064] (a)提供肽和连接在固体支持物上的相关的记录标签;

[0065] (b)使所述肽与能够结合所述肽的N末端氨基酸(NTAA)的第一结合试剂接触,其中所述第一结合试剂包含具有所述第一结合试剂的识别信息的第一编码标签;

[0066] (c)将所述第一编码标签的信息传送到所述记录标签以产生一个延伸记录标签;并且

[0067] (d)分析所述延伸记录标签。

[0068] 在第18个实施方案中是第17个实施方案的方法,其中步骤(b)还包括使肽与包含第二(或更高次序)编码标签的第二(或更高次序)结合剂接触,所述编码标签具有关于第二(或更高次序)结合剂的识别信息,其中所述第二(或更高次序)结合剂能够结合所述肽的NTAA以外的NTAA。

[0069] 在第19个实施方案中是第18个实施方案的方法,其中在所述肽与所述第一结合剂接触后,接着使所述肽与第二(或更高级)结合剂接触。

[0070] 在第20个实施方案中是第18个实施方案的方法,其中使所述肽与第二(或更高级)结合剂接触和所述肽与所述第一结合剂的接触同时发生

[0071] 在第21实施方案中,是一种用于分析肽的方法,包括以下步骤:

[0072] (a)提供肽和连接在固体支持物上的相关的记录标签;

[0073] (b)使所述肽与能够结合所述肽的N末端氨基酸(NTAA)的第一结合试剂接触,其中所述第一结合试剂包含具有所述第一结合试剂的识别信息的第一编码标签;

[0074] (c)将第一编码标签的信息传递到所述记录标签以产生第一次序延伸记录标签;

[0075] (d)移除所述肽的NTAA以露出新的NTAA;

[0076] (e)使所述肽与能够结合所述新的NTAA的第二结合剂接触,其中所述第二结合剂包含具有所述第二结合剂的识别信息的第二编码标签;

[0077] (h)将所述第二编码标签的信息传送到所述第一次序延伸记录标签以产生第二次序延伸记录标签;以及

[0078] (i)分析所述第二延伸记录标签。

[0079] 在第22实施方案中,是第1至第10实施方案中任一项的方法,其中所述大分子是蛋白质、多肽或肽。

[0080] 在第23实施方案中,是第1至第10实施方案中任一项的方法,其中所述大分子是肽。

[0081] 在第24实施方案中,是第1至第23实施方案中任一项的方法,其中所述肽是通过自生物样品碎片化分离的蛋白而获得。

[0082] 在第25实施方案中,是第1至第10实施方案中任一项的方法,其中所述大分子是脂质、碳水化合物或大环。

[0083] 在第26实施方案中,是第1至第25实施方案中任一项的方法,其中所述记录标签是DNA分子、具有假互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或其组合。

[0084] 在第27实施方案中,是第1至第26实施方案中任一项的方法,其中所述记录标签包含通用引发位点。

[0085] 在第28实施方案中,是第27实施方案中的方法,其中所述通用引发位点包含用于扩增、测序或二者的引物位点。

[0086] 在第29实施方案中,是第1至第28实施方案中任一项的方法,其中所述记录标签包含独特分子标识符(UMI)。

[0087] 在第30实施方案中,是第1至第29实施方案中任一项的方法,其中所述记录标签包含条形码。

[0088] 在第31实施方案中,是第1至第30实施方案中任一项的方法,其中所述记录标签在其3'-末端包含间隔子。

[0089] 在第32实施方案中,是第1至第31实施方案中任一项的方法,其中所述大分子和所述相关的记录标签共价结合在所述固体支持物上。

[0090] 在第33实施方案中,是第1至第32实施方案中任一项的方法,其中所述固体支持物是珠子、多孔珠子、多孔基质、阵列、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、微量滴定孔、ELISA板,旋转干涉测量盘、

硝酸纤维素膜、硝化纤维素基聚合物表面、纳米颗粒或微球。

[0091] 在第34实施方案中,是第33实施方案中的方法,其中所述固体支持物是聚苯乙烯珠子、聚合物珠子、琼脂糖珠子、丙烯酰胺珠子、固体核心珠子、多孔珠子、顺磁珠子、玻璃珠子,或可控孔珠子。

[0092] 在第35实施方案中,是第1至第34实施方案中任一项的方法,其中多个大分子和相关的记录标签结合到所述固体支持物上。

[0093] 在第36实施方案中,是第35实施方案中的方法,其中所述多个大分子在所述固体支持物上以>50nm的平均距离间隔开。

[0094] 在第37实施方案中,是第1至第36实施方案中任一项的方法,其中所述结合剂是多肽或蛋白质。

[0095] 在第38实施方案中,是第37实施方案中的方法,其中所述结合剂是修饰的氨肽酶、修饰的氨酰tRNA合成酶、修饰的anticalin或修饰的ClpS。

[0096] 第39实施方案中,是第1至第38实施方案中任一项的方法,其中所述结合剂能够选择性地结合所述大分子。

[0097] 在第40实施方案中,是第1至第39实施方案中任一项的方法,其中所述记录标签是DNA分子、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或其组合。

[0098] 在第41实施方案中,是第1至第40实施方案中任一项的方法,其中所述记录标签包含编码序列。

[0099] 在第42实施方案中,是第1至第41实施方案中任一项的方法,其中所述记录标签还包含间隔子、结合循环特异性序列、独特分子标识符、通用引发位点、或它们的任何组合。

[0100] 在第43实施方案中,是第1至第42实施方案中任一项的方法,其中所述结合剂和所述编码标签通过接头连接。

[0101] 在第44实施方案中,是第1至第42实施方案中任一项的方法,其中所述结合剂和所述编码标签通过SpyTag/SpyCatcher或SnoopTag/SnoopCatcher肽-蛋白质对连接。

[0102] 在第45实施方案中,是第1至第44实施方案中任一项的方法,其中传送所述编码标签的信息到所述记录标签是通过DNA连接酶介导。

[0103] 在第46实施方案中,是第1至第44实施方案中任一项的方法,其中传送所述编码标签的信息到所述记录标签是通过DNA合成酶介导。

[0104] 在第47实施方案中,是第1至第44实施方案中任一项的方法,其中传送所述编码标签的信息到所述记录标签是通过化学连接介导。

[0105] 在第48实施方案中,是第1至第47实施方案中任一项的方法,其中分析所述延伸记录标签包含核酸测序方法。在第49实施方案中,是第48实施方案中的方法,其中所述核酸测序方法是通过合成测序、连接测序、杂交测序、polony测序、离子半导体测序或焦磷酸测序。

[0106] 第50实施方案中,是第48实施方案中的方法,其中所述核酸测序方法是单分子实时测序、基于纳米孔测序、或使用高级显微镜的DNA直接成像。

[0107] 在第51实施方案中,是第1至第50实施方案中任一项的方法,其中所述延伸记录标签在分析之前扩增。

[0108] 在第52实施方案中,是第1至第51实施方案中任一项的方法,其中包含于所述延伸记录标签中的所述编码标签信息的次序提供关于结合到所述大分子的结合剂的结合次序

的信息。

[0109] 在第53实施方案中,是第1至第52实施方案中任一项的方法,其中所述延伸记录标签中包含的所述编码标签信息的频率提供关于结合到所述大分子的结合剂的结合频率的信息。

[0110] 在第54实施方案中,是第1至第53实施方案中任一项的方法,其中代表多个大分子的多个延伸记录标签被平行分析。

[0111] 在第55实施方案中,是第54实施方案的方法,其中代表多个大分子的多个延伸记录标签在多路复用检测中分析。

[0112] 在第56实施方案中,是第1至第55实施方案中任一项的方法,其中所述多个延伸记录标签在分析之前进行靶标丰度检测。

[0113] 在第57实施方案中,是第1至第56实施方案中任一项的方法,其中所述多个延伸记录标签在分析之前进行减法分析。

[0114] 在第58实施方案中,是第1至第57实施方案中任一项的方法,其中所述多个延伸记录标签在分析之前进行标准化测定以降低高丰度种类。

[0115] 在第59实施方案中,是第1至第58实施方案中任一项的方法,其中所述NTAA采用修饰的氨肽酶、修饰的氨基酸tRNA合成酶、温和型埃德曼降解、埃德曼酶或无水TFA来移除。

[0116] 在第60实施方案中,是第1至第59实施方案中任一项的方法,其中所述至少一个结合剂结合到末端氨基酸残基。

[0117] 在第61实施方案中,是第1至第60实施方案中任一项的方法,其中所述至少一个结合剂结合到翻译后修饰氨基酸。

[0118] 在第62实施方案中,是分析来自于包含多个蛋白复合体、蛋白、或多肽的样本的一个或多个肽,所述方法包含:

[0119] (a)将样品内的多个蛋白质复合物、蛋白质或多肽分区成多个隔室,其中每个隔室包括多个隔室标签,其任选地连接到固体支持物,其中在单个隔室内的所述多个隔室标签是相同的且不同于其他隔室的隔室标签;

[0120] (b)将多种蛋白质复合物、蛋白质和/或多肽片段化成多个肽;

[0121] (c)在足以允许所述多个肽与所述多个隔室内的多个隔室标签退火或连接的条件下,使所述多个肽与所述多个隔室标签接触,从而产生多个隔室标记的肽;

[0122] (d)从所述多个隔室收集隔室标记的肽;并且

[0123] (e)根据第1-21实施方案中以及第26-61实施方案中任一项的方法分析一个或多个隔室标记的肽。

[0124] 在第63实施方案中,是第62实施方案中的方法,所述隔室是微流体液滴。

[0125] 在第64实施方案中,是第62实施方案中的方法,其中所述隔室是微孔。

[0126] 在第65实施方案中,是第62实施方案中的方法,其中所述隔室是表面上分离的区域。

[0127] 在第66实施方案中,是第62至第65实施方案中任一项的方法,其中每个隔室平均包含一个单独细胞。

[0128] 在第67实施方案中,是分析来自于包含多个蛋白复合体、蛋白、或多肽的样本的一个或多个肽,所述方法包含:

[0129] (a)用多个通用DNA标签标记多种蛋白质复合物、蛋白或多肽;

[0130] (b)将样品内的多个蛋白复合物、蛋白或多肽分区成多个隔室,其中每个隔室包括多个隔室标签,其中在单个隔室内的多个隔室标签是相同的且不同于其他隔室的隔室标签;

[0131] (c)在足以允许所述多个肽与所述多个隔室内的所述多个隔室标签退火或连接的条件下,使所述多个蛋白复合物、蛋白、或多肽与所述多个隔室标签接触,从而产生多个隔室标记的蛋白复合物、蛋白、或多肽;

[0132] (d)从所述多个隔室收集隔室标记的蛋白复合物、蛋白、或多肽;

[0133] (e)任选地将隔室标记的蛋白复合物、蛋白或多肽片段化成隔室标记的肽片段化为隔室标签的肽;和

[0134] (f)根据第1-21实施方案中以及第26-61实施方案中任一项的方法分析一个或多个隔室标记的肽。

[0135] 在第68实施方案中,是第62至第67实施方案中任一项的方法,其中所述隔室标签信息通过引物延伸或链接被传送到与肽相关的记录标签。

[0136] 在第69实施方案中,是第62至第68实施方案中任一项的方法,其中所述固体支持物包含珠子。

[0137] 在第70实施方案中,是第69实施方案中的方法,其中所述珠子是聚苯乙烯珠子、聚合物珠子、琼脂糖珠子、丙烯酰胺珠子、固体核心珠子、多孔珠子、顺磁珠子、玻璃珠子或可控孔珠子。

[0138] 在第71实施方案中,是第62至第70实施方案中任一项的方法,其中所述隔室标签包含单链或双链核苷酸分子。

[0139] 在第72实施方案中,是第62至第71实施方案中任一项的方法,其中所述隔室标签包含条形码和任选的UMI。

[0140] 在第73实施方案中,是第72实施方案的方法,其中所述固体支持物是珠子,所述隔室标签包含条形码,进一步其中包含多个连接到其上的隔室标签的珠子通过混合-切分合成形成。

[0141] 在第74实施方案中,是第72实施方案的方法,其中所述固体支持物是珠子,所述隔室标签包含条形码,进一步其中包含多个连接到其上的隔室标签的珠子通过个体合成或固定化形成。

[0142] 在第75实施方案中,是第62至第74实施方案任一项的方法,其中所述间隔标签是记录标签中的成分,所述记录标签任选地还包括间隔、特异分子标签、通用引发位点、或它们的任何组合。

[0143] 在第76实施方案中,是第62至第75实施方案任一项的方法,其中所述间隔标签还包含能够在所述多个蛋白质复合体、蛋白质、或多肽上与内部氨基酸或N-末端氨基酸反应的功能基团。

[0144] 在第77实施方案中,是第76实施方案的方法,所述功能基团是NHS基。

[0145] 在第78实施方案中,是第76实施方案的方法,其中所述功能基团是醛基。

[0146] 第79实施方案是第62至第78实施方案任一项的方法,其中所述多个隔室标签通过以下方法形成:

[0147] 打印、点样、将隔室标签喷墨到隔室中,或它们的组合。

[0148] 在第80实施方案中,是第62至第79实施方案任一项的方法,其中所述隔室标签还包含肽。

[0149] 在第81实施方案中,是第80实施方案的方法,其中所述隔室标签肽包含蛋白连接酶识别序列。

[0150] 第82实施方案是第81实施方案的方法,其中所述蛋白连接酶是butelase I或其同源物。

[0151] 在第83实施方案中,是第62至第82实施方案任一项的方法,其中所述多个多肽用蛋白酶片段化。

[0152] 在第84实施方案中,是第83实施方案的方法,其中所述蛋白酶是金属蛋白酶。

[0153] 在第85实施方案中,是第84实施方案的方法,其中所述金属蛋白酶的活性通过金属阳离子的光子激发释放来调节。

[0154] 在第86实施方案中,是第62至第85实施方案任一项的方法,还包含在将所述多个多肽分区成多个隔室之前,从样品中减去一种或多种高丰度蛋白。

[0155] 在第87实施方案中,是第62至第86实施方案任一项的方法,还包含在将所述多个肽联合到所述隔室标签之前,从所述固体支持物上释放所述隔室标签。

[0156] 在第88实施方案中,是第62实施方案的方法,还包括以下步骤(d),将所述隔室标签肽连接到与记录标签相关的固体支持物上。

[0157] 在第89实施方案中,是第88实施方案的方法,还包括将所述隔室标记的肽上的所述隔室标签的信息传送到相关的记录标签。

[0158] 在第90实施方案中,是第89实施方案的方法,还包括在步骤步骤(e)之前将所述隔室标签从所述隔室标记的肽上移除。

[0159] 在第91实施方案中,是第62至第90实施方案任一项的方法,还包括基于分析的肽的隔室标签序列来确定所分析的肽的来源单细胞的身份。

[0160] 在第92实施方案中,是第62至第90实施方案任一项的方法,还包括基于分析的肽的隔室标签序列来确定所分析的肽来源的蛋白质或蛋白质复合物的身份。

[0161] 在第93实施方案中是分析多个大分子的方法,包括以下步骤:

[0162] (a)提供多个大分子和连接在固体支持物上的相关记录标签;

[0163] (b)使所述多个大分子与能够结合所述多个大分子的多个结合剂接触,其中每个结合剂包含具有所述结合剂的识别信息的编码标签;

[0164] (c)(i)将所述大分子相关记录标签的信息传递到与大分子结合的结合剂的所述编码标签,以产生延伸编码标签;或(ii)将大分子相关记录标签和结合到所述大分子的结合剂的编码标签的信息传递到di-tag构建体;

[0165] (d)收集隔所述延伸编码标签或di-tag构建体;

[0166] (e)任选地重复步骤(b)-(d)一个或多个结合循环;

[0167] (f)分析所述延伸编码标签或di-tag构建体的集合物。

[0168] 在第94实施方案中,是第93实施方案中的方法,其中所述大分子是蛋白。

[0169] 在第95实施方案中,是第93实施方案中的方法,其中所述大分子是肽。

[0170] 在第96实施方案中,是第95实施方案中的方法,其中所述肽是通过碎片化分离自

生物样品的蛋白获得。

[0171] 在第97实施方案中,是第93至第96实施方案中任一项所述的方法,其中所述记录标签是DNA分子、RNA分子、PNA分子、BNA分子、XNA分子、LNA分子、γPNA分子或其组合。

[0172] 在第98实施方案中,是第93至第97实施中任一项的方法,其中所述记录标签包含独特分子标识符(UMI)。

[0173] 在第99实施方案中,是第93-98实施方案中任一项的方法,其中所述记录标签包含隔室标签。

[0174] 在第100实施方案中,是第93至第99实施方案中任一项的方法,其中所述记录标签包含通用引发位点。

[0175] 在第101实施方案中是第93-100实施方案任一项所述的方法,其中所述记录标签在其3'-末端包含间隔子。

[0176] 在第102实施方案中,是第93至第101实施方案中任一项的方法,其中所述记录标签的3'-末端是封闭的,以防止所述记录标签通过合成酶延伸,并且所述大分子相关记录标签的信息和结合到所述大分子的结合剂的编码标签的信息被传送到di-tag构建体。

[0177] 在第103实施方案中,是第93至第102实施方案中任一项的方法,其中所述编码标签包含编码序列。

[0178] 在第104实施方案中,是第93至第103实施方案中任一项的方法,其中所述编码标签包含UMI。

[0179] 在第105实施方案中,是第93至第104实施方案中任一项的方法,其中所述编码标签包含通用引发位点。

[0180] 在第106实施方案中,是第93-105实施方案任一项所述的方法,所述编码标签在其3'-末端包含间隔子。

[0181] 在第107实施方案中,是第93至第106实施方案中任一项的方法,其中所述编码标签包含结合循环特异性序列。

[0182] 在第108实施方案中,是第93至第107实施方案中任一项的方法,其中所述结合剂和所述编码标签通过接头连接。

[0183] 在第109实施方案中,是第93至第108实施方案中任一项的方法,其中传送所述记录标签的信息到所述编码标签是通过引物延伸实现的。

[0184] 在第110实施方案中,是第93至第108实施方案中任一项的方法,其中所述传送记录标签的信息到所述编码标签是通过连接实现的。

[0185] 在第111实施方案中,是第93至第108实施方案中任一项的方法,其中所述di-tag构建体是通过间隙填充、引物延伸或它们二者产生的。

[0186] 在第112实施方案中是实施方案93-97、107、108和111中任一项的方法,其中所述di-tag分子包含衍生自所述记录标签的通用引发位点、衍生自所述记录标签的隔室标签、衍生自所述记录标签的独特分子标识符、衍生自所述记录标签的可选间隔子、衍生自所述编码标签的编码序列、衍生自编码标签的独特分子标识符、衍生自编码标签的可选间隔子和衍生自编码标签的通用引发位点。

[0187] 在第113实施方案中,是第93-112实施方案中任一项的方法,其中所述大分子和相关的记录标签共价结合到所述固体支持物上。

[0188] 在第114实施方案中,是第113实施方案的方法,其中所述固体支持物是珠子、多孔珠子、多孔基质、阵列、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶芯片、流通芯片、包含信号转导电子器件的生物芯片、微量滴定孔、ELISA板、旋转干涉测量盘、硝酸纤维素膜、硝化纤维素基聚合物表面、纳米颗粒或微球。

[0189] 第115实施方案是依据第114实施方案的方法,其中所述固体支持物是聚苯乙烯珠子、聚合物珠子、琼脂糖珠子、丙烯酰胺珠子、固体核心珠子、多孔珠子、顺磁珠子、玻璃珠子,或可控孔珠子。

[0190] 第116实施方案是依据第93至第115实施方案中任一项的方法,其中所述结合剂是多肽或蛋白质。

[0191] 第117实施方案是依据第116实施方案的方法,其中所述结合剂是修饰的氨肽酶、修饰的氨酰tRNA合成酶、修饰的anticalin或者抗体或抗体的结合片段。

[0192] 第118实施方案是第95至第117实施方案中任一项的方法,其中所述结合剂结合单个氨基酸残基、二肽、三肽或肽的翻译后修饰。

[0193] 第119实施方案是依据第118实施方案的方法,其中所述结合剂与N-末端氨基酸残基、C-末端氨基酸残基或内部氨基酸残基结合。

[0194] 第120实施方案是依据第118实施方案的方法,其中所述结合剂结合N-末端肽、C-末端肽或内部肽。

[0195] 第121实施方案是依据第119实施方案的方法,其中所述结合剂与N-末端氨基酸残基结合,并且在每个结合循环后切割N-末端氨基酸残基。

[0196] 第122实施方案是依据第119实施方案的方法,其中所述结合剂结合C-末端氨基酸残基,并且在每个结合循环后切割C-末端氨基酸残基。

[0197] 实施方案123是实施方案121所述的方法,其中所述的N-末端氨基酸残基通过Edman降解法切割。

[0198] 实施方案124是实施方案93所述的方法,其中所述的结合剂是氨基酸或翻译后修饰的位点特异性共价标记。

[0199] 实施方案125是第93至第124实施方案中任一项的方法,其中所述在步骤(b)之后,将包含大分子和相关结合剂的复合物从固体支持物上解离并分配成液滴或微流体液滴的乳液。

[0200] 实施方案126是实施方案125所述的方法,其中所述的每个微流体液滴,平均而言,都包含一种复合物,包括大分子和结合剂。

[0201] 实施方案127是实施方案125或126所述的方法,其中所述的记录标签在产生延伸编码标签或di-tag构建体之前扩增。

[0202] 实施方案128是第125至第127实施方案中任一项的方法,其中所述的乳液融合PCR用于将记录标签信息传递到编码标签或创建一组di-tag构建体。

[0203] 实施方案129是第93至第128实施方案中任一项的方法,其中所述的延伸编码标签或di-tag构建体的集合是在分析之前扩增的。

[0204] 实施方案130是第93至第129实施方案中任一项的方法,其中所述的分析延伸编码标签或二标签构建体的集合包括核酸测序方法。

[0205] 实施方案131是第130实施方案所述的方法,其中所述核酸测序方法是通过合成测

序、连接测序、杂交测序、polony测序、离子半导体测序或焦磷酸测序。

[0206] 实施方案132是实施方案130所述的方法,其中所述核酸测序方法是单分子实时测序、基于纳米孔测序、或使用高级显微镜的DNA直接成像。

[0207] 实施方案133是实施方案130所述的方法,其中所述的大分子的部分组成,是通过使用独特的隔室标签和可选UMI分析多个延伸的编码标签或di-tag标签构建体来确定的。

[0208] 实施方案134是第1至第133实施方案中任一项的方法,其中分析步骤用测序方法进行,所述测序方法具有>5%、>10%、>15%、>20%、>25%或>30%的每碱基错误率。

[0209] 实施方案135是第1至第134实施方案中任一项的方法,其中所述的编码标签、记录标签或两者的识别组件包含纠错码。

[0210] 实施方案136是实施方案135所述的方法,其中所述识别组件选自编码器序列、条形码、UMI、隔室标签、循环特定序列或其任何组合。

[0211] 实施方案137是实施方案135或136所述的方法,其中所述的纠错码,选自汉明码、李距离码、非对称李距离码、里德-所罗门码和Levenshtein-Tenengolts码。

[0212] 实施方案138是第1至第134实施方案中任一项的方法,其中所述的编码标签、记录标签或两者的识别组件能够产生独特的电流或离子通量或光学特征,其中所述的分析步骤包括检测独特的电流或离子通量或光学特征,以识别识别组件。

[0213] 实施方案139是实施方案138所述的方法,其中所述识别组件选自编码器序列、条形码、UMI、隔室标签、循环特定序列或其任何组合。

[0214] 实施方案140是分析多个大分子的方法,包括以下步骤:

[0215] (a)提供多个大分子和连接在固体支持物上的相关记录标签;

[0216] (b)使所述多个大分子与能够结合同源大分子的多个结合剂接触,其中每个结合剂包含具有所述结合剂的识别信息的编码标签;

[0217] (c)将第一结合剂的第一编码标签的信息传递到与第一大分子相关的第一记录标签,以产生第一次序延伸记录标签,其中所述第一结合剂与所述第一大分子结合;

[0218] (d)使所述多个大分子与能够结合同源大分子的多个结合剂接触;

[0219] (e)将第二结合剂的第二编码标签的信息传递到第一次序延伸记录标签,以产生第二次序延伸记录标签,其中所述第二结合剂结合所述第一大分子;

[0220] (f)任选地重复步骤(d)-(e)“n”个结合循环,其中将结合到第一大分子的每个结合剂的每个编码标签的信息传递到前一个结合循环产生的所述延伸记录标签,以产生代表所述第一个大分子的第n次序延伸记录标签;

[0221] (g)分析所述第n次序延伸记录标签。

[0222] 实施方案141是实施方案140所述的方法,其中生成并分析表示多个大分子的多个第n次序延伸记录标签。

[0223] 实施方案142是第140或第141实施方案所述的方法,其中所述大分子为蛋白。

[0224] 实施方案143是第142实施方案所述的方法,其中所述大分子为肽。

[0225] 实施方案144是第143实施方案所述的方法,其中所述肽是通过碎片化分离生物样品的蛋白而获得。

[0226] 实施方案145是第140至第144实施方案中任一项的方法,其中所述多个大分子包含来自多个合并样品的大分子。

[0227] 实施方案146是第140至第145实施方案中任一项的方法,其中所述记录标签是DNA分子、RNA分子、PNA分子、BNA分子、XNA分子、LNA分子、γPNA分子或其组合。

[0228] 实施方案147是第140至第146实施方案中任一项的方法,其中记录标签包含独特分子标识符(UMI)。

[0229] 实施方案148是第140至第147实施方案所述的方法,其中所述记录标签包含隔室标签。

[0230] 实施方案149是第140至第148实施方案中任一项的方法,其中所述的记录标签包含通用引发位点。

[0231] 实施方案150是第140至第149实施方案任一项所述的方法,其中所述记录标签在其3'-末端包含间隔子。

[0232] 实施方案151是第140至第150实施方案中任一项的方法,其中所述编码标签包含编码器序列。

[0233] 实施案例152是第140至第151实施方案中任一项的方法,其中所述编码标签包含UMI。

[0234] 实施方案153是第140至第152实施方案中任一项的方法,其中所述的编码标签包含通用引发位点。

[0235] 实施方案154是第140至第153实施方案中任一项所述的方法,其中所述编码标签在其3'-末端包含间隔子。

[0236] 实施方案155是第140至第154实施方案中任一项的方法,其中所述编码标签包含结合循环特异性序列。

[0237] 实施方案156是第140至第155实施方案中任一项的方法,其中所述的编码标签包含特异分子标签。

[0238] 实施方案157是第140至第156实施方案中任一项的方法,其中所述的结合剂和所述编码标签通过接头连接。

[0239] 实施方案158是第140至第157实施方案中任一项的方法,其中所述记录标签到编码标签的信息传递通过引物延伸介导。

[0240] 实施方案159第140至第158实施方案中任一项的方法,其中所述记录标签到编码标签的信息传递通过连接介导。

[0241] 实施方案160是第140至第159实施方案中任一项的方法,其中所述多个大分子和相关的记录标签或者两者都结合到所述固体支持物上。

[0242] 实施方案161是第140至第160实施方案中任一项的方法,所述固体支持物是珠子、多孔珠子、多孔基质、阵列、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、微量滴定孔、ELISA板、旋转干涉测量盘、硝酸纤维素膜、硝化纤维素基聚合物表面、纳米颗粒或微球。

[0243] 实施方案162是实施方案161所述的方法,其中所述固体支持物是聚苯乙烯珠子、聚合物珠子、琼脂糖珠子、丙烯酰胺珠子、固体核心珠子、多孔珠子、顺磁珠子、玻璃珠子或可控孔珠子。

[0244] 实施方案163是第140至第162实施方案中任一项的方法,其中所述结合剂是多肽或蛋白质。

[0245] 实施方案164是实施方案163所述的方法,其中所述结合剂是修饰的氨肽酶、修饰的氨酰tRNA合成酶、修饰的anticalin或者抗体或抗体的结合片段。

[0246] 实施方案165是第142至第164实施方案中任一项的方法,其中所述结合剂结合单个氨基酸残基、二肽、三肽或肽的翻译后修饰。

[0247] 实施方案166是实施方案165所述的方法,其中所述结合剂与N-末端氨基酸残基、C-末端氨基酸残基或内部氨基酸残基结合。

[0248] 实施方案167是实施方案165所述的方法,其中所述结合剂结合N-末端肽、C-末端肽或内部肽。

[0249] 实施方案168是第142至第164实施方案中任一项的方法,其中所述结合剂结合修饰的N-末端氨基酸残基的化学标签、修饰的C-末端氨基酸残基,或修饰的内部氨基酸残基。

[0250] 实施方案169是第166或168实施方案的方法,其中所述结合剂与N-末端氨基酸残基或修饰的N-末端氨基酸残基的所述化学标签结合,并且所述N-末端氨基酸残基在每个结合循环后被切割。

[0251] 实施方案170是第166或168实施方案的方法,其中所述结合剂与C-末端氨基酸残基或修饰的C-末端氨基酸残基的所述化学标签结合,并且所述C-末端氨基酸残基在每个结合循环后被切割。

[0252] 实施方案171是实施方案169所述的方法,其中所述的N-末端氨基酸残基通过Edman降解法、埃德曼酶、修饰的氨基肽酶、或修饰的酰基肽水解酶切割。

[0253] 实施方案172是实施方案163所述的方法,其中所述的结合剂是氨基酸位点的特异性共价标记或翻译后修饰的位点特异性共价标记。

[0254] 实施方案173是第140-172实施方案中任一项的方法,其中所述多个第n次序延伸记录标签在分析之前扩增。

[0255] 实施方案174是第140至第173实施方案中任一项的方法,其中分析所述第n次序延伸记录标签包含核酸测序方法。

[0256] 实施方案175是实施方案174所述的方法,其中代表多个大分子的多个第n次序延伸记录标签被平行分析。

[0257] 实施方案176是第174或175实施方案所述的方法,其中所述核酸测序方法是通过合成测序、连接测序、杂交测序、polony测序、离子半导体测序或焦磷酸测序。

[0258] 实施方案177是根据实施方案174或175所述的方法,其中核酸测序方法是单分子实时测序、基于纳米孔的测序或采用高级显微镜的DNA的直接成像。

附图说明

[0259] 通过参考附图的示例方式描述本发明的非限制性实施例,其是示意性的并且不旨在按比例绘制。出于说明的目的,并非每个部件都标记在每个图中,也未示出本发明的每个实施例中对于本领域普通技术人员来说理解本发明不是必需的每个部件。

[0260] 附图1A-B:附图1A示出了在所述附图中显示的功能元件的图例。

[0261] 图1B示出了将蛋白质编码转导成DNA编码的一般概述,其中多个蛋白质或多肽被片段化成多个肽,然后将其转化成延伸记录标签文库,以代表多个肽。所述延伸记录标签构成代表所述肽序列的DNA编码文库。所述文库可以适当地修改以在任何下一代测序(NGS)平

台上进行测序。

[0262] 附图2A-2D示出了根据本公开的方法的蛋白质大分子分析的示例,其采用包含编码标签的结合剂(例如抗体、anticalins、N-recognins蛋白(例如,ATP依赖性Clp蛋白酶衔接蛋白(ClpS)),适配体等以及其变体/同源物)、所述编码标签与固定的蛋白质互相作用,所述固定的蛋白质与单个或多个记录标签共定位或共标记。所述记录标签由通用引发位点、条形码(例如,分区条形码、隔室条形码、分级物条形码)、可选的独特分子标识符(UMI)序列,以及用于编码标签的信息传递的间隔序列(Sp)构成。所述间隔标签(Sp)在所有结合循环之间可以是恒定的、结合试剂特异性的、或结合循环数特异性的。所述编码标签由为结合剂提供识别信息的编码序列,可选的UMI,和与记录标签上的互补间隔序列杂交的间隔序列构成,便于将编码标签信息传递给所述记录标签(例如,引物延伸,在本文中也称为合成酶延伸)。图2A示出了通过所述同源结合剂与蛋白质环状结合产生延伸记录标签的过程,以及相应信息从所述结合剂的编码标签到蛋白质记录标签的传送。在一系列按顺序的结合和编码标签信息传送步骤之后,产生最终的延伸记录标签,其包含结合剂编码标签信息,所述结合剂编码标签信息包括编码器序列,其来自于为所述结合剂(例如抗体2(Ab2),抗体3(Ab3),...抗体"n"(Abn))提供识别信息的"n"结合周期、来自于所述记录标签的条形码/可选UMI序列、来自于所述结合剂编码标签的可选UMI序列,以及在所述文库构建体的每一末端的侧翼通用引物序列,以便于通过数字下一代测序进行扩增和分析。图2B示出了用DNA条形码记录标签标记蛋白的示意性例子。在上部图中,N-羟基琥珀酰亚胺(NHS)是胺反应性偶联剂,并且二苯并环辛基(DBCO)是用于"点击"偶联到固体基质的表面的张紧性炔烃。在该方案中,所述记录标签通过NHS基团与所述蛋白的赖氨酸(K)残基(和任选的N-末端氨基酸)的ε胺偶联。在底部图中,用异双功能连接子NHS-炔标记赖氨酸(K)残基的ε胺以产生炔"点击"基团。然后叠氮化物标记的DNA记录标签可以通过标准点击化学将容易地连接到这些反应性炔基上。此外,所述DNA记录标签也可以设计有正交的甲基四嗪(mTet)基团,用于通过逆iEDDA反应下游偶联到TCO衍生的测序底物。图2C示出了采用记录标签的蛋白分析方法的的例子。在顶部附图中,蛋白质大分子通过捕获剂和任选地的通过交联固定在固体支持物上。所述蛋白或捕获剂可以采用记录标签标记。在底部附图中,记录标签相关的蛋白直接固定在固体支持物上。图2D示出了使用同源结合物的DNA编码和所述产生的延伸记录标签的测序的简单蛋白免疫侧定的整体工作流程实例。所述蛋白可以通过记录标签进行条形码化(即,索引)并在循环结合分析之前汇集,大大增加样品通量并节省结合试剂。该方法实际上是用于进行反相蛋白质测定(RPPA)的数字化、更简单且更可扩展的方法。

[0263] 图3A-D通过构建代表肽序列的DNA延伸记录标签进行基于降解的肽测序分析的流程。这是通过在固体支撑物上使用N-末端氨基酸(NTAA)结合的循环过程、编码标签信息传递到与所述肽连接的记录标签、NTAA切割,以及以循环方式重复该过程的Edman降解样方法来实现。提供了对来自肽的N-末端降解的延伸记录标签的示例性构建的概述:(A)标记肽的N-末端氨基酸(例如,采用用苯基硫代氨基甲酰基(PTC)、二硝基苯基(DNP)、磺酰基硝基苯基(SNP)、乙酰基或胍基基团);(B)显示了结合剂与结合到所述标记的NTAA的相关编码标签;(C)显示了结合到固体支持物(例如,珠子)并与记录标签(例如,通过三官能接头)相关的肽,其中在所述结合剂结合到所述肽的NTAA时,所述编码标签的信息被传递到所述记录标签(例如,通过引物延伸)以产生延伸的记录标签;(D)通过化学或酶促方法切割标记的

NTAA以暴露新的NTAA。如箭头所指,所述循环重复“n”次以产生最终的延伸记录标签。所述最终的延伸记录标签可选地通过通用引发位点设置在侧翼以便于下游扩增以及DNA测序。所述正向通用引发位点(例如,Illumina的P5-S1序列)可以是原始记录标签设计的一部分,并且所述反向通用引发位点(例如,Illumina的P7-S2’序列)可以在所述记录标签的延伸的最后一步添加。所述最后一步可以用一个结合试剂独立地完成。

[0264] 图4A-B示出了根据本文公开的方法的示例性蛋白测序流程。图4A示出了具有以浅灰色虚线勾画出的替代模式的示例性工作流程,其中特定实施例以箭头链接的框示出。所述流程的每个步骤的替代模型在箭头下的框中示出。图4B示出了进行循环结合和编码标签信息传送步骤以提高信息传送效率的选项。每个分子可以采用多个记录标签。此外,对于给定的结合事件,编码标签信息传递到所述记录标签可以多次进行,或可选地,采用表面扩增步骤以产生所述延伸记录标签文库的拷贝,等等。

[0265] 图5A-B示出了使用引物延伸以将结合剂的编码标签的识别信息传递到与大分子(例如肽)相关的记录标签以产生延伸记录标签的延伸记录标签的示例性构建的概述。包含具有关于结合剂的识别信息的独特编码序列的编码标签可选地通过共同的间隔序列(Sp’)侧接在每个末端。图5A示出了包含结合到一个连接到珠子的记录标签标记的肽的NTAA的编码标签的NTAA结合剂。所述记录标签通过互补间隔序列(Sp)与所述编码标签退火,并且通过使用间隔区(Sp)作为引物位点,引物延伸反应介导编码标签的信息向所述记录标签的传递。所述编码标签显示,是在远离结合剂的一端具有单链间隔区(Sp’)序列的双链体。该配置使所述编码标签与记录标签中的内部位点的杂交最小化,并有利于所述记录标签的末端间隔区(Sp)序列与所述编码标签的单链间隔突出(Sp’)的杂交。此外,所述延伸记录标签可以与寡核苷酸(与编码器、间隔序列互补)预退火以阻止所述编码标签与内部记录标签序列元件的杂交。图5B显示在“n”个结合循环后产生的最终延伸记录标签(“***”表示在延伸记录标签中未显示的干预结合循环)和所述编码标签信息的传递以及在3’-端添加通用引发位点。

[0266] 图6示出了通过酶连接将编码标签信息传递到延伸记录标签。显示了两种不同的大分子及其各自的记录标签,其中记录标签延伸平行进行。可以通过设计双链编码标签促进连接,使得所述间隔序列(Sp)具有与所述记录标签上的互补间隔序列(Sp’)退火的“粘性末端”突出。双链编码标签的互补链传递信息到所述记录标签。当连接用于延伸所述记录标签,延伸的方向可以是如图所示的5’到3’,或可选的3’到5’。

[0267] 图7示出了通过化学连接将编码标签信息传递到记录标签,以将记录标签或延伸记录标签的3’核苷酸连接到所述编码标签(或其互补体)的5’核苷酸而不插入间隔序列到所述延伸记录标签的“无间隔”方法。所述延伸记录标签和编码标签的方向也可以转换,以便于将所述记录标签的5’端连接到所述编码标签(或互补体)的3’端。在显示的示例中,所述记录标签上的互补“辅助”寡核苷酸序列(“记录辅助”)和所述编码标签之间的杂交用于稳定所述复合物,以促使所述记录标签与编码标签互补链的特异性化学连接。所述产生的延伸记录标签缺乏间隔序列。还示出了可以使用DNA、PNA或类似的核酸聚合物的化学连接(例如,使用叠氮和炔基团(显示为三线符))的“点击化学”版本。

[0268] 图8A-B示出了在N-末端氨基酸降解前,将肽的翻译后修饰(PTM)写入延伸记录标签的示例性方法。附图8A:包含具有关于所述结合剂(例如,包含具有磷酸酪氨酸抗体的识

别信息的编码标签的磷酸酪氨酸抗体)的识别信息的编码标签的结合剂能够结合所述肽。如果磷酸酪氨酸存在于所述记录标签标记的肽中,如图所示,当所述磷酸酪氨酸抗体与磷酸酪氨酸结合后,所述编码标签和记录标签通过互补间隔序列退火并且所述编码标签信息被传递到所述记录标签以产生延伸记录标签。图8B:延伸记录标签可以包含用于初始氨基酸序列(例如“aa_1”,“aa_2”,“aa_3”,…,“aa_N”)的编码标签信息和所述肽的翻译后修饰(例如“PTM_1”,“PTM_2”)的编码标签信息。

[0269] 图9A-B示出了结合剂结合大分子并将附着于结合剂的编码标签的信息传递到多个记录标签中的单个记录标签从而产生共同代表所述大分子的多个延伸记录标签的多个循环过程,所述多个记录标签共定位于单个大分子的一个位点,所述单个大分子附着于固体支持物(例如珠子)。在这个附图中,仅出于举例的目的,所述大分子是肽,并且每个循环涉及将结合剂结合至N-末端氨基酸(NTAA)、通过将编码标签信息传递至记录标签来记录结合事件,然后去除NTAA以曝露新的NTAA。图9A示出了与所述大分子共定位于固体支撑物上的多个记录标签(包含通用的正向引物序列和UMI)。单个的记录标签具有与结合剂的编码标签内的共同间隔序列互补的共同间隔序列(Sp),其可用于引发延伸反应以将编码标签信息传递至记录标签。图9B示出了用于每个连续结合循环的循环特异性NTAA结合剂的不同库,每个库具有循环特异性间隔序列。

[0270] 图10A-C示出了包含将附着于结合剂的编码标签的信息传递到多个记录标签中的单个记录标签从而产生共同代表所述大分子的多个延伸记录标签的多个循环的示例模型,所述多个记录标签共定位于单个大分子的一个位点,所述单个大分子附着于固体支持物(例如珠子)。在这个附图中,仅出于举例的目的,所述大分子是肽,并且每个循环过程涉及结合至N-末端氨基酸(NTAA)、记录结合事件,然后去除NTAA以曝露新的NTAA。

[0271] 图10A示出了与所述大分子共定位于固体支撑物上的多个记录标签(包含通用的正向引物序列和UMI),优选每珠子一个单分子。单个记录标签在它们的3'-末端具有不同的间隔序列,其具有不同的“循环特异性”序列(例如C_1,C_2,C_3,…C_n)。优选地,每个珠子上的所述记录标签分享相同的UMI序列。在第一个结合循环(循环1)中,使多个NTAA结合剂所述大分子接触。循环1中的结合剂具有共同的5'-间隔序列(C'1),其与所述记录标签的所述循环1C_1间隔序列互补。循环1中的结合剂也具有共同的3'-间隔序列(C'$_2$),其与所述循环2间隔子C_2互补。在结合循环1期间,第一个NTAA结合剂结合到所述大分子的自由N-末端,并且第一编码标签的信息通过从杂交的C_1序列到所述互补的C'1间隔序列的引物延伸传递到同源记录标签。除去NTAA以暴露新的NTAA之后,结合循环2使多个NTAA结合剂与所述单分子接触,所述NTAA结合剂具有与所述循环1的3'-间隔序列相同的循环2的5'-间隔序列(C'2)和共同的循环3的3'-间隔序列(C'3)。第2NTAA结合剂结合到所述大分子的N-末端,并且第2编码标签的信息通过互补的C2和C'2间隔序列的引物延伸传递到同源记录标签。这些循环重复到“n”个结合循环,其中最后的延伸记录标签用通用反向引发序列加盖,产生与单个大分子共定位的多个延伸记录标签,其中每个延伸记录标签具有来自一个结合循环的编码标签信息。因为在每个连续结合循环中使用的每组结合剂在所述编码标签中具有循环特异性间隔序列,所以结合循环信息可以与产生的延伸记录标签中的结合剂信息相关联。图10B示出了用于每个连续结合循环的循环特异性NTAA结合剂的不同库,每个库具有循环特异性间隔序列。图10C示出了共定位在所述大分子的位点上的延伸记录标签的集合如何使用循环特

异性间隔序列按顺序基于PCR组装的所述延伸记录标签,从而提供有序的所述大分子的序列。在优选模型中,每个延伸记录标签的多个拷贝通过级联之前的扩增产生。

[0272] 图11A-B示出了信息从记录标签传递到编码标签或di-tag构建体。记录结合信息的两种方法在(A)和(B)中示出。结合剂可以是此处所述的任何类型的结合剂;仅出于说明的目的显示抗磷酸酪氨酸结合剂。对于延伸的编码标签或di-tag构建体,不是将所述编码标签中的结合信息传递到所述记录标签,而是将信息从所述记录标签传递到所述编码标签以生成延伸的编码标签(A),或者将信息从所述记录标签和编码标签传递到第三个di-tag-forming构建体(B)所述di-tag和延伸编码标签包含所述记录标签(包含条形码、可选的UMI序列和可选的隔室标签(CT)序列(未示出))和所述编码标签的信息。所述di-tag和延伸编码标签可从所述记录标签上被洗脱、收集,并可选地被扩增并在下一代测序仪上被读出。

[0273] 图12A-D示出了结合事件的PNA组合条形码/UMI记录标签和di-tag检测的设计。在图12A中,示出了通过四个基本PNA单词序列(A,A'-B,B'-C和C')的化学连接的组合PNA条形码/UMI的构建。杂交DNA臂包括其中以产生用于PNA条形码/UMI的组合装配的无间隔物组合模板。采用化学连接将退火的PNA"单词"缝合在一起。图12B示出了将所述记录标签的PNA信息传递到DNA中间体的方法。所述DNA中间体能够将信息传递给所述编码标签。即,将互补DNA单词序列与PNA退火并进行化学连接(如果发现使用PNA模板的连接酶,则可选酶促连接)。在图12C中,所述DNA中间体设计来通过间隔序列,SP与所述编码标签互作。链置换引物延伸步骤取代连接的DNA并将所述记录标签信息从所述DNA中间体传递到所述编码标签以产生延伸的编码标签。终止子核苷酸可以并入所述DNA中间体的末端以防止编码标签信息通过引物延伸传递到所述DNA中间体。图12D:或者,可以将信息从编码标签传递到所述DNA中间体以产生di-tag构建体。终止子核苷酸可以并入所述DNA中间体的末端以防止记录标签信息从所述DNA中间体传递到所述编码标签。

[0274] 图13A-E示了隔室条形码珠子上的蛋白质组分区,随后通过乳化融合PCR进行di-tag组装,以产生代表肽序列组成的元件库。随后可以通过N-末端测序或者通过与编码标签相关的氨基酸特异性化学标记或结合剂的附着(共价或非共价)来表征所述肽的氨基酸含量。所述编码标签包含通用引发位点、以及用于所述氨基酸识别的条形码、隔室标签、和氨基酸UMI。在信息传递之后,通过记录标签UMI将所述ditags映射回所述原始分子。在图13A中,所述蛋白质组被条形码化的珠子拆分为液滴。具有相关记录标签(包括隔室条形码信息)的肽附着在所述珠子的表面上。所述液体乳液打破后释放具有分区肽的条形码化珠子。在图13B中,在所述肽上的特异性氨基酸位点采用与共价连接到位点特异性标签基团的DNA编码标签化学标记。所述DNA编码标签包含氨基酸条形码信息和任选的氨基酸UMI。图13C:标记的肽-记录标签复合体从所述珠子上释放。图13D:所述标记的肽-记录标签复合物乳化成纳米或微乳液,使得每个隔室平均有少于一个肽-记录标签复合物。图13E:乳液融合PCR将记录标签信息(例如,隔室条形码)传递到所有与氨基酸残基连接的所述DNA编码标签上。

[0275] 图14示出了延伸编码标签从乳化肽记录标签-编码标签复合体产生。来自图13C的所述肽复合体与PCR试剂共乳化成为平均每个液滴具有单个肽复合体的液滴。使用三引物融合PCR方法扩增与肽相关的所述记录标签,将所述扩增的记录标签融合到多个结合剂编码标签或共价标记的氨基酸的编码标签上,通过引物延伸延伸所述编码标签从而转移肽UMI和所述记录标签的隔室标签信息到所述编码标签,并放大得到的延伸编码标签。每个液

滴有多个延伸编码标签种类,每个氨基酸编码序列-UMI编码标签存在不同延伸编码标签种类。以这种方式,在所述肽中的氨基酸身份和数量都能被测定。所述U1通用引物和Sp引物被设计为比所述U2tr通用引物具有更高的熔化Tm。这使得能够进行两步PCR,其中在较高的退火温度下执行前几个循环以放大所述记录标签,然后进行到较低的Tm使得所述记录标签和编码标签在PCR期间互相引导以产生延伸编码标签,并且所述U1和U2tr通用引物用于引发所得的延伸编码标签产物的扩增。在一些实施方案中,来自所述U2tr引物的合成酶延伸可以采用光标记3'封闭基团来防止(Young et al.,2008,Chem.Commun(Camb)4:462-464)。在第一轮PCR扩增所述记录标签,以及第二轮融合PCR步骤(其中所述编码标签Sp_{tr}在扩增的所述记录标签的Sp'序列上引发所述编码标签的延伸)之后,所述$U2_{tr}$的3'封闭基团被移除,并且启动更高温度的PCR以采用U1和$U2_{tr}$引物扩增所述延伸编码标签。

[0276] 图15图示了蛋白质组区和条形码的使用有助于增强蛋白的可映射性和定相。在肽测序中,蛋白通常被消化成肽。在该过程中,源自亲本蛋白分子的各个肽之间的关系的以及它们与亲本蛋白分子的关系的信息被丢失。为了重建该信息,将各个肽序列映射回可能是它们的来源的蛋白序列的集合。采用短和/或部分肽序列,并且随着所述集合的大小和复杂性(例如,蛋白质组序列复杂性)增加,在这样的组中发现特异性匹配的任务变得更加困难。将蛋白质组分区为条形码化(例如,隔室标记的)隔室或分区,随后将所述蛋白质消化成肽,并所述隔室标签连接到所述肽降低了肽序列需要映射上去的所述"蛋白质"空间,在复杂蛋白质样品的情况下极大地简化了所述任务。在消化成肽之前采用独特分子标识符(UMI)标记蛋白质有助于肽映射回所述原始蛋白质分子并允许在源自相同蛋白质分子的翻译后修饰(PTM)变体之间的相信息的注释和单个蛋白质(proteoforms)形式的标识。图15A示出蛋白质组分区的例子,其包含用含有分区条形码的记录标签标记蛋白质并随后将其片段化成记录标签标记的肽。图15B:对于部分肽序列信息或甚至仅仅是组成信息,该映射是高度简并的。然而,部分肽序列或组成信息与来自相同蛋白质的多个肽的信息配对,允许原始蛋白质分子独特标识。

[0277] 图16图示了隔室标记的珠子序列设计的示例性模式。所述隔室标签包含X_{5-20}的条形码以识别单个隔室和N_{5-10}的独特分子标识符(UMI)以识别所述隔室标签连接的肽,其中X和N代表简并核酸碱基或核碱基词。隔室标签可以是单链(上部描绘)或双链(下部描绘)。任选地,隔室标签可以是嵌合分子,其包含具有用于连接目标肽(左图)的蛋白质连接酶(例如,butelase I)的识别序列的肽序列。或者,化学基团可以包含于用于配对到目标肽(例如叠氮化物,如右图所示)的所述隔室标签中。

[0278] 图17A-B图示:(A)代表多个肽的多个延伸记录标签;(B)通过标准杂交捕获技术进行靶肽富集的示例性方法。例如,杂交捕获富集可以使用一种或多种生物素化的"诱饵"寡核苷酸,其与代表一种或多种感兴趣的肽("靶肽")的延伸记录标签杂交,所述一种或多种感兴趣的肽来源于代表肽文库的延伸记录标签库。所述诱饵寡核苷酸:靶延伸记录标签杂交对在杂交后通过所述生物素标签从溶液中下沉,以产生代表肽或目标肽的延伸记录标签的富集分解物。所述延伸记录标签的分离("下拉")可以,例如,使用链霉亲和素包被的磁珠完成。所述生物素基团与所述磁珠上的链霉亲和素结合,并且通过使用磁体固定所述磁珠完成分离同时除去或更换溶液。与代表不需要的或过度丰裕的肽的延伸记录标签竞争性杂交的非生物素化的竞争物富集寡核苷酸可任选地包括在杂交捕获测定的杂交步骤中,以调

节富集的靶肽的量。所述非生物素化的竞争寡核苷酸竞争杂交到所述靶肽,但是由于不存在生物素基团,在捕获步骤期间不捕获杂交双链体。因此,可以通过在大的动态范围内调节竞争寡核苷酸与生物素化的“诱饵”寡核苷酸的比例来调节所述富集的延伸记录标签分级物。该步骤对于解决所述样品中蛋白质丰度的动态范围问题非常重要。

[0279] 图18A-B示出了单细胞和大量蛋白质组分区进单个液滴的示例性方法,每个液滴包含珠子,所述珠子具有附着于其上的多个隔室标签以将肽与其起源蛋白质复合物或源自单个细胞的蛋白质相关联。所述隔室标签包含条形码。液滴形成后液滴成分的操作:(A)单细胞被分区进单个液滴中,接着细胞裂解以释放细胞蛋白质组,以及蛋白水解以将所述细胞蛋白质组消化成肽,并在充分蛋白水解后失活所述蛋白酶;(B)大量蛋白质组被分区进入多个液滴中,其中单个液滴包含蛋白复合物,然后进行蛋白质水解以将所述蛋白复合物消化成肽,并在充分蛋白质水解后灭活所述蛋白酶。在光笼状二价阳离子的光释放(photo-release of photo-caged divalent cations)后,可用热不稳定的金属蛋白酶将包封的蛋白质消化成肽以激活所述蛋白酶。在充分蛋白水解后可以热灭活所述蛋白酶,或者可以螯合所述二价阳离子。液滴含有杂交的或可释放的隔室标签,其包含能够连接至肽的N-或C-末端氨基酸的核酸条形码(分离自记录标签)。

[0280] 图19A-B示出了单细胞和大量蛋白质组分区进单个液滴的示例性方法,每个液滴包含珠子,其具有多个双功能记录标签和连接于记录标签上的隔室标签,以将肽与其原始蛋白质或蛋白质复合物相关联,或将蛋白质与起源相关联。所述隔室标签包含条形码。液滴形成后液滴成分的操作:(A)单细胞被分区成单个液滴中,接着细胞裂解以释放细胞蛋白质组,以及蛋白水解以将所述细胞蛋白质组消化成肽,并在充分蛋白水解后失活所述蛋白酶;(B)大量蛋白质组被分区进入多个液滴中,其中单个液滴包含蛋白复合物,然后进行蛋白质水解以将所述蛋白复合物消化成肽,并在充分蛋白质水解后失活所述蛋白酶。在光笼状二价阳离子(例如,Zn^{2+})的光释放后,可用热不稳定的金属蛋白酶将包封的蛋白质消化成肽。在充分蛋白水解后可以热灭活所述蛋白酶,或者可以螯合所述二价阳离子。液滴含有杂交或可释放的隔室标签,其包含能够连接到肽的N-或C-末端氨基酸的核酸条形码(分离自记录标签)。

[0281] 图20A-L示出了附着于肽的隔室条形码化记录标签的产生。隔室条形码技术(例如,微流体液滴中的条形码化珠子等)可用于将隔室特异性条形码传递到封装在特定隔室内的分子内容物。(A)在一个具体实施方案中,使所述蛋白分子变性,并且使所述赖氨酸残基(K)的ε-胺基与活化的通用DNA标签分子(包含通用引发序列(U1))化学缀合,在5’端用NHS基团显示)。在将通用DNA标签与所述多肽缀合后,除去过量的通用DNA标签。(B)使所述通用DNA标记多肽与结合在珠子上的核酸分子杂交,其中所述结合在单个珠子上的核酸分子包含独特的隔室标签(条形码)序列的群组。通过将所述样品分离进入不同的物理隔室,例如液滴(由虚线椭圆表示)发生区室化。或者,可以通过将标记的多肽固定在所述珠子表面上来直接完成区室化,例如,通过将所述多肽上的通用DNA标签与所述珠子上的所述隔室DNA标签退火,而不需要额外的物理分离。单个多肽分子仅与单个珠子相互作用(例如,单个多肽不跨越多个珠子)。然而,多种多肽可以与同一个珠子相互作用。除了所述隔室条形码序列(BC)之外,与珠子结合的核酸分子可以由共同的Sp(间隔子)序列,独特分子标识符(UMI)以及与所述多肽DNA标签,U1’互补的序列组成。(C)在所述通用DNA标记的多肽与结合

珠子的隔室标签退火后,通过切割所述附属接头从所述珠子释放所述隔室标签。(D)使用源自所述珠子的所述隔室标签核酸分子作为模板,通过基于聚合酶的引物延伸延伸所述退火的U1 DNA标签引物。如(C)中所述引物延伸步骤可以在所述隔室标签从所述珠子释放之后进行,或者任选地,在所述隔室标签仍附着在所述珠子(未示出)时进行。这有效地将所述珠子上的所述隔室标签上的所述条形码序列写到所述多肽上的所述U1 DNA标签序列上。这个新序列构成了记录标签。引物延伸后,蛋白酶,例如Lys-C(切割赖氨酸残基的C-末端),Glu-C(切割谷氨酸残基的C-末端并且在较低程度上切割谷氨酸残基的C-末端),或随机的蛋白酶如蛋白酶K,用于将所述多肽切割成肽片段。(E)每个肽片段用延伸的DNA标签序列标记,该延伸的DNA标签序列在其C-末端赖氨酸上构成记录标签,用于本公开中下游的肽测序。(F)所述记录标记的肽通过张紧性炔标记,DBCO与叠氮化珠子偶联。所述叠氮化珠子可选地还含有与所述记录标签互补的捕获序列,以促进DBCO-叠氮化物固定的效率。应该注意的是,所述肽从原始珠子中移除并重新固定到新的固体支持物(例如珠子)让肽之间形成最佳分子间间隔,有助于本文公开的肽测序方法。除了使用点击化学缀合DNA标签与炔烃预标记多肽(如图2B中所述),图20G-L说明了与图20A-F中所示类似的概念。所述叠氮化物和mTet化学是正交的,允许点击缀合到DNA标签并点击iEDDA缀合(mTet和TCO)到所述测序底物。

[0282] 图21示出了使用用于单细胞的流式聚焦T形结构(T-junction)和采用珠子的隔室标记(例如,条形码)区室化的示例性方法。通过两个液流,可以在液滴形成时容易地引发细胞裂解和蛋白酶活化(Zn^{2+}混合)。

[0283] 图22A-B示出了示例性标记细节。(A)通过采用Butelase I的肽连接将隔室标签(DNA-肽嵌合体)连接到所述肽上。(B)在开始肽测序之前,将标签标签信息转移到相关的记录标签上。任选地,内肽酶AspN,其选择性地切割N-末端到天冬氨酸残基的肽键,可用于在信息转移至所述记录标签后切割所述隔室标签。

[0284] 图23A-C用于基于空间蛋白质组学的组织切片分析的基于阵列的条形码。(A)将一组空间编码的DNA条形码(由BCij表示的特征条形码)阵列与组织切片(FFPE或冷冻)组合。在一个实施方案中,所述组织切片是固定的并且透化的。在优选的实施方案中,所述阵列特征尺寸小于所述细胞大小(对于人细胞为~10μm)。(B)用试剂处理阵列式组织切片以逆转交联(例如,通过采用柠康酸酐的抗原修复方案(Namimatsu,Ghazizadeh等人,2005),然后其中的所述蛋白质用位点反应性DNA标记,从而用DNA记录标签有效地标记所有蛋白质分子(例如,赖氨酸标记,在抗原修复后释放)。标记和洗涤后,阵列结合的DNA条形码序列被切割并扩散到所述阵列式组织切片中并与其中的附着在所述蛋白质上的DNA记录标签杂交。(C)现在对所述阵列式组织进行聚合酶延伸,以将所述杂交的条形码的信息转移到标记所述蛋白质的DNA记录标签上。在转移所述条形码信息后,从所述载玻片上刮下所述阵列式组织,任选地用蛋白酶消化,并将蛋白质或肽提取到溶液中。

[0285] 图24A-B示出两种不同的示例性DNA靶大分子(AB和CD),其固定在珠子上并通过与编码标签连接的结合剂进行测定。该模型系统用于演示编码标签从结合试剂转移到近端报告标签的单分子行为。在优选实施例中,所述编码标签通过引物延伸整合到延伸记录标签中。图24A示出了AB大分子与A特异性结合剂(“A”,是与AB大分子的“A”组分互补的寡核苷酸序列)的相互作用以及相关编码标签的信息通过引物延伸向记录标签的转移,以及B特异性结合剂(“B”,是与AB大分子的“B”组分互补的寡核苷酸序列)和相关编码标签的信息通过引

物延伸向记录标签的转移。编码标签A和B具有不同的序列,并且为了便于在该图示中辨识,它们也具有不同的长度。所述不同的长度有助于通过凝胶电泳分析编码标签转移,而不需要通过下一代测序进行分析。所述A'和B'结合剂的结合被示为单个结合循环的替代可能性。如果增加第二个循环,则所述延伸记录标签将进一步延伸。取决于在第一和第二循环中添加A'或B'结合剂中的哪一种,所述延伸记录标签可以包含AA,AB,BA和BB形式的编码标签信息。因此,所述延伸记录标签包含结合事件次序以及结合子的身份的信息。类似地,图24B演示了CD大分子与C特异性结合剂("C",是与CD大分子的"C"组分互补的寡核苷酸序列)的相互作用以及相关编码标签的信息通过引物延伸转移到记录标签,和D特异性结合剂("D",是与CD大分子的"D"组分互补的寡核苷酸序列)和通过引物延伸将相关编码标签的信息转移到记录标签。编码标签C和D具有不同的序列,并且为了便于识别,在该图示中也具有不同的长度。所述不同长度有助于通过凝胶电泳分析编码标签转移,而不需要通过下一代测序进行分析。C'和D'结合剂的结合被示为单个结合循环的替代可能性。如果添加第二个循环,则所述延伸记录标签将进一步延伸。根据在第一和第二循环中添加C'或D'结合剂中的哪一种,所述延伸记录标签可以包含CC,CD,DC和DD形式的编码标签信息。编码标签可以可选地包含UMI。在编码标签中包含UMI可以记录结合事件的附加信息;并可以在单个结合剂水平上区分结合事件。如果单个结合剂可以参与一个以上的结合事件(例如,其结合亲和力使得它可以足够频繁地脱离和重新结合以参与一个以上的事件),则可能很有用。它也可用于纠错。例如,在某些情况下编码标签可能在同一结合周期中将信息传递到所述记录标签两次或更多次。采用UMI可以暴露那些可能都与单个结合事件相关联的重复的信息传递事件。

[0286] 图25示出了示例性DNA靶大分子(AB)并固定在珠子上,并通过与编码标签连接的结合剂进行测定。A特异性结合剂("A",与AB大分子的A组分互补的寡核苷酸)与AB大分子相互作用,并通过连接将相关编码标签的信息转移至记录标签。B特异性结合剂("B",与AB大分子的B组分互补的寡核苷酸)与AB大分子相互作用,并通过连接将相关编码标签的信息转移至记录标签。编码标签A和B具有不同的序列,并且为了便于识别,在该图示中也具有不同的长度。所述不同长度有助于通过凝胶电泳分析编码标签转移,而不需要通过下一代测序进行分析。

[0287] 图26A-B示出了通过引物延伸的用于结合/编码标签转移的示例性DNA-肽大分子。图26A示出了固定在珠子上的示例性寡核苷酸-肽靶大分子("A"寡核苷酸-CMyc肽)。cMyc特异性结合剂(例如抗体)与所述大分子的所述cMyc肽部分相互作用,并且相关编码标签的信息被转移至记录标签。可以通过凝胶电泳分析cMyc编码标签的信息向记录标签的转移。图26B示出了固定在珠子上的示例性寡核苷酸-肽靶大分子("C"寡核苷酸-血凝素(HA)肽)。HA特异性结合剂(例如,抗体)与所述大分子的所述HA肽部分相互作用,并且相关编码标签的信息被转移至记录标签。可以通过凝胶电泳分析编码标签的信息向记录标签的转移。cMyc抗体-编码标签和HA抗体-编码标签的结合被演示作为单个结合循环的替代可能性。如果执行第二个结合循环,则延伸记录标签将进一步被延伸。依赖于在第一和第二结合周期中添加cMyc抗体-编码标签或HA抗体-编码标签中的是哪一个,所述延伸记录标签可以包含cMyc-HA,HA-cMyc,cMyc-cMyc和HA-HA形式的编码标签信息。尽管未示出,但也可以引入另外的结合剂以能够检测所述大分子的所述A和C寡核苷酸组分。因此,可以通过将信息传递到记录标签并读出所述延伸记录标签来分析包含不同类型骨架的杂合大分子,所述延伸记

录标签含有关于结合事件以及结合剂的身份的次序信息。

[0288] 图27A-D.纠错条形码的生成。(A)采用所述推荐参数[create.dnabarcodes(n＝15,dist＝10)]从所述R软件包'DNABarcodes'(https://bioconductor.riken.jp/packages/3.3/bioc/manuals/DNAB arcodes/man/DNABarcodes.pdf)衍生的77个条形码中选择65个纠错条形码(SEQ ID NO:1-65)的子集。该算法生成15-mer"汉明"条形码,其可以将替换错误纠正到四次替换的距离,并检测到九次替换的错误。通过过滤除那些未显示出各种纳米孔电流水平(用于基于纳米孔的测序)或与该组的其他成员过于相关的条形码而创建了所述65个条形码的子集。(B)15-mer条形码通过所述纳米孔的预测纳米孔电流水平的示意图。通过将每个15-mer条形码字分成11个重叠的5-mer字的复合组来计算预测电流,在条形码一次一个碱基通过所述纳米孔时使用5-mer R9纳米孔电流水平查找表(template_median68pA.5mers.model(https://github.com/jts/nanopolish/tree/master/etc/r9-models)来预测对应的电流水平。从(B)可以理解,这组65个条形码为其每个成员展示出独特的电流特征。(C)作为用于纳米孔测序法的模式延伸记录标签的PCR产物的产生显示采用重叠的DTR组和DTR引物。然后连接PCR扩增子以形成级联的延伸记录标签模型。(D)示例性的"延伸记录标签"模型纳米孔测序读数(读数长度734个碱基)的产生如图27C所示。所述MinIon R9.4 Read的质量得分为7.2(读取质量差)。然而,即使读数质量差(Qscore＝7.2),也可以使用lalign轻松识别条形码序列。15-mer间隔元件加下划线。条形码可以正向或反向排列,用BC或BC'表示。

[0289] 图28A-D.采用记录标签的蛋白质的分析物特异性标记。(A)靶向天然构象的目的蛋白质分析物的结合剂包含与DNA记录标签上的互补分析物特异性条形码(BCA')杂交的分析物特异性条形码(BCA)。或者,所述DNA记录标签可以通过可切割的接头连接在所述结合剂上,并且所述DNA记录标签直接被"点击"到所述蛋白质上,然后从所述结合剂上被切除(通过所述可切割的接头)。所述DNA记录标签包含反应性偶联基团(例如点击化学试剂(如叠氮化物,mTet等),用于偶联所述目标蛋白质和其他功能组分(例如,通用引物序列(Sp)),样品条形码(BCS),分析物特异性条形码(BC_A)和间隔序列(Sp))。样品条形码(BCS)也可用于标记和区分不同样品的蛋白质。所述DNA记录标签也可以包含正交偶联基团(例如mTet),用于随后偶联到基质表面。对于所述记录标签与目标蛋白质的点击化学偶联,所述蛋白质用点击化学偶联基团预标记,该化学偶联基团与所述DNA记录标签(例如,蛋白质上的炔基团与DNA记录标签上的叠氮基团同源)上的点击化学偶联基团同源。用于用点击化学偶联的偶联基团标记DNA记录标签的试剂的实例包括用于赖氨酸标记的炔-NHS试剂、用于光亲和标记的炔-二苯甲酮试剂等。(B)在所述结合剂结合近端靶蛋白后,所述记录标签上的反应性偶联基团(例如叠氮化物)与所述近端蛋白上的同源点击化学偶联基团(显示为三线符号)共价连接。(C)在用所述记录标签标记所述目标蛋白质分析物后,通过采用尿嘧啶特异性切除试剂(例如USER™)消化尿嘧啶(U)除去所述连接的结合剂。(D)采用合适的生物共轭化学反应,例如点击化学(炔-叠氮化物结合对,甲基四嗪(mTET)-反式-环辛烯(TCO)结合对等)将所述DNA记录标签标记的靶蛋白分析物固定到基质表面。在某些实施方案中,使用结合剂库和记录标签库在包含许多不同靶蛋白分析物的单个管中进行整个靶蛋白-记录标签标记测定。在用包含样品条形码(BCS)的记录标签对样品中的蛋白质分析物靶向标记后,可以在(D)中的固定步骤之前汇总多个蛋白质分析物样品。因此,在某些实施方案中,数百个

样品中多达数千种蛋白质分析物可以在单管下一代蛋白质测定(NGPA)中标记和固定,极大地节省昂贵的亲和试剂(例如抗体)。

[0290] 图29A-E.DNA记录标签与多肽的缀合。(A)变性多肽用双功能点击化学试剂,例如炔-NHS酯(乙炔-PEG-NHS酯)试剂或炔二苯甲酮标记,以产生炔标记的(三线符号)多肽。炔烃也可以是张紧炔烃,例如环辛炔,包括二苯并环辛基(DBCO)等。(B)显示了化学偶联到所述炔烃标记的多肽的DNA记录标签设计的实例。所述记录标签包括通用引物序列(P1)、条形码(BC)和间隔序列(Sp)。所述记录标签用mTet基团标记以偶联到基质表面和用叠氮基团标记用于与标记多肽的炔基团偶联。(C)变性的炔烃标记蛋白或多肽通过炔烃和叠氮基团用记录标签标记。任选地,所述记录标签标记的多肽可以进一步用隔室条形码标记,例如通过与连接到隔室珠子和引物延伸(也称为聚合酶延伸)的互补序列退火,或如图20H-J所示。(D)蛋白酶消化所述记录标签标记的多肽产生一群记录标签标记的肽。在一些实施方案中,一些肽不会被任何记录标签标记。在其他实施方案中,一些肽可以附着一个或多个记录标签。(E)使用TCO基团官能化的基质表面和与肽连接的记录标签的mTet基团之间的反转电子Diels-Alder(iEDDA)点击化学反应,将记录标签标记的肽固定在基质表面上。在某些实施例中,可以在所示的不同阶段之间采用清理步骤。使用正交点击化学(例如,叠氮化物-炔烃和mTet-TCO)允许用记录标签对多肽进行点击化学标记,以及用点击化学固定所述记录标签标记的肽到基质表面上(参见,McKay et al.,2014,Chem.Biol.21:1075-1101,其全部内容通过引用并入本文)。

[0291] 图30A-E.在多肽的DNA标签初始标记后将样品条形码写入记录标签。(A)用双功能点击化学试剂如炔-NHS试剂或炔二苯甲酮标记变性的多肽,以产生炔标记的多肽。(B)在炔(或替代点击化学试剂)标记所述多肽之后,包含通用引物序列(P1)并用叠氮基团和mTet基团标记的DNA标签通过叠氮-炔相互作用与所述多肽偶联。应理解,可以采用其它点击化学相互作用。(C)包含样品条形码信息(BCS')和其它记录标签功能组分(例如,通用引物序列(P1'),间隔序列(Sp'))的记录标签DNA构建体通过互补的通用引发序列(P1-P1')与所述DNA标签标记的多肽退火。通过聚合酶延伸将记录标签信息转移到所述DNA标签。(D)蛋白酶消化所述记录标签标记的多肽产生一群记录标签标记的肽。(E)使用TCO基团官能化的基质表面和与肽连接的记录标签的mTet基团之间的反转电子Diels-Alder(iEDDA)点击化学反应,将记录标签标记的肽固定在基质表面上。在某些实施例中,可以在所示的不同阶段之间采用清理步骤。使用正交点击化学(例如,叠氮化物-炔烃和mTet-TCO)允许用记录标签对多肽进行点击化学标记,以及用点击化学固定所述记录标签标记的肽到基质表面上(参见,McKay et al.,2014,Chem.Biol.21:1075-1101,其全部内容通过引用并入本文)。

[0292] 图31A-E.用于条形码多肽的珠子区室化。(A)使用标准生物共轭或光亲和标记技术的异双功能点击化学试剂在溶液中标记多肽。可能的标记位点包括赖氨酸残基的ε-胺(例如,所示的采用NHS-炔烃)或所述肽的碳主链(例如,采用二苯甲酮-炔烃)。(B)包含通用引物序列(P1)的叠氮化物标记的DNA标签与所述标记的多肽的所述炔基团偶联。(C)通过互补DNA序列(P1和P1')将所述DNA标签标记的多肽与DNA记录标签标记的珠子退火。所述珠子上的所述DNA记录标签包含间隔序列(Sp'),区室条形码序列(BC_p'),可选的独特分子标识符(UMI)和通用序列(P1')。所述DNA记录标签信息通过聚合酶延伸(或者,可以采用连接)转移到所述多肽上的所述DNA标签上。在信息转移后,所得多肽包含多个记录标签,其包含隔

室条形码在内的若干功能元件。(D)蛋白酶消化所述记录标签标记的多肽产生一群记录标签标记的肽。所述记录标签标记的肽与所述珠子解离,并且(E)重新固定到测序基质上(例如,如图所示,使用mTet和TCO基团之间的iEDDA点击化学)。

[0293] 图32A-H.下一代蛋白质测定(NGPA)的工作流程示例。蛋白质样品用由几个功能单元,例如通用引用序列(P1)、条形码序列(BC)、可选的UMI序列和间隔序列(Sp)(能够与结合试剂编码标签促使信息传递),组成的DNA记录标签标记。(A)将所述标记的蛋白被固定(被动地或共价地)到基质(例如珠子、多孔珠子或多孔基质)上。(B)所述基质用蛋白封闭,并任选地加入与所述间隔序列互补的竞争寡核苷酸(Sp')以使分析物记录标签序列的非特异性相互作用最小化。(C)将分析物特异性抗体(连有相关编码标签)与基质结合蛋白一起孵育。所述编码标签可包含尿嘧啶碱基,用于随后的尿嘧啶特异性切割。(D)抗体结合后,洗去过量的竞争性寡核苷酸(Sp')(如果加入的话)。所述编码标签通过互补间隔序列瞬时退火到所述记录标签,并且所述编码标签信息在引物延伸反应中转移到所述记录标签以产生延伸记录标签。如果所述固定的蛋白质是变性的,则所述结合的抗体和退火的编码标签可以在碱性条件下,例如用0.1N NaOH,洗涤除去。如果固定的蛋白质处于天然构象,则可能需要较温和的条件来除去结合的抗体和编码标签。在图E-H中概述了较温和的抗体去除条件的实例。(E)在信息从所述编码标签转移到所述记录标签之后,使用尿嘧啶特异性切除试剂(例如USERTM)酶混合物在其尿嘧啶位点切割(切割)所述编码标签。(F)使用高盐、低/高pH洗涤液从蛋白中去除所述结合的抗体。保留在抗体上的所述截短的DNA编码标签很短并且也快速洗脱。较长的DNA编码标签片段可以与所述记录标签保持或不保持退火。(G)如步骤(B)-(D)中,第二结合循环开始并且第二引物延伸步骤通过引物延伸将所述编码标签信息从所述第二抗体转移到所述延伸的记录标签。(H)两个结合循环的结果是来自于附着在所述记录标签的所述第一抗体的结合信息和所述第二抗体的结合信息的级联。

[0294] 图33A-D:使用多种结合剂和酶促介导的顺序信息传递的单步下一代蛋白测定(NGPA)。采用同时结合两种同源结合剂(例如抗体)固定的蛋白质分子的NGPA测定法。在多个同源抗体结合事件后,使用组合的引物延伸和DNA缺口步骤将来自结合抗体的所述编码标签的信息转移至所述记录标签。所述编码标签中的插入符号(^)代表双链DNA切口核酸内切酶位点。(A)在所示的实施例中,与蛋白质的表位1(Epi#1)结合的所述抗体的所述编码标签在互补间隔序列杂交后的引物延伸步骤中将编码标签信息(例如,编码器序列)转移至所述记录标签。(B)一旦所述延伸记录标签和编码标签之间形成所述双链DNA双体,缺口内切酶例如Nt.BsmAI,其在37℃时有活性,仅切割双链DNA基质上的单链DNA,用于切割所述编码标签。在所述缺割步骤之后,由所述截短的编码标签结合剂和延伸的记录标签形成的双体是热力学不稳定的且解离的。所述较长的编码标签片段可以与所述记录标签保持或不保持退火。(C)这让来自与蛋白质的表位#2(Epi#2)结合的所述抗体的所述编码标签通过互补间隔序列与所述延伸记录标签退火,并且所述延伸记录标签通过经引物延伸从Epi#2抗体的所述编码标签到所述延伸记录标签的信息传递来进一步延伸。(D)再次,在延伸记录标签和Epi#2抗体的编码标签之间形成双链DNA双体后,所述编码标签被切口核酸内切酶如Nb.BssSI切割。

[0295] 在某些实施方案中,优选在引物延伸期间使用非链置换聚合酶(也称为聚合酶延伸)。非链置换聚合酶防止所述切割的编码标签短柱延伸,其通过一个以上单碱基与所述记

录标签保持退火。(A)-(D)的过程可以自己重复,直到近端结合的结合剂的所有所述编码标签被所述杂交、信息转移到所述延伸记录标签,和切割步骤“消耗”。所述编码标签可以包括与对给定分析物(例如,同源蛋白)特异的所有结合剂(例如,抗体)相同的编码序列,可以包含表位特异性编码序列,或者可以包含独特分子标识符(UMI)以区分不同的分子事件。

[0296] 图34A-C:记录标签的受控密度-使用滴定基板表面上的反应性基团来固定肽。(A)可以通过控制基质表面上的功能性偶联基团的密度来滴定所述基质的表面上的肽密度。这可以通过活性偶联分子和“虚拟”偶联分子的适当比例对所述基质表面进行衍生来实现。在所示的实例中,NHS-PEG-TCO试剂(活性偶联分子)与NHS-mPEG(虚拟分子)以确定的比例组合和TCO来衍生胺表面。功能化的PEG具有300至超过40,000的各种分子量。(B)使用琥珀酰亚胺基4-(N-马来酰亚胺甲基)环己烷-1(SMCC)双功能交联剂将双功能5’胺DNA记录标签(mTet是其它功能基团)偶联到肽的N-末端Cys残基。所述记录标签上的内部mTet-dT基团是使用mTetrazine-Azide从叠氮化物-dT基团产生的。(C)采用用mTet和TCO的iTDA点击化学反应,将标记肽的所述记录标签固定在(A)的所述活化基质表面上。所述mTet-TCO iEDDA偶联反应非常快速,高效且稳定(mTet-TCO比Tet-TCO更稳定)。

[0297] 图35A-C.下一代蛋白质测序(NGPS)结合周期特异性编码标签。(A)采用循环特异性N-末端氨基酸(NTAA)结合剂编码标签的NGPS测定的设计。NTAA结合剂(例如,对N-末端DNP标记酪氨酸特异的抗体)与肽的DNP标记的NTAA结合,所述肽与包含通用引物序列(P1)、条形码(BC)和间隔序列(SP)的记录标签相关。当所述结合剂结合所述肽的同源NTAA时,与所述NTAA结合剂相关的所述编码标签接近所述记录标签并通过互补间隔序列与所述记录标签退火。编码标签信息通过引物延伸转移到所述记录标签。为了跟踪编码标签代表哪一个结合循环,所述编码标签可以包括循环特异性条形码。在某些实施方案中,结合到分析物的结合剂的编码标签具有独立于循环数的相同的编码器条形码,其与独特的结合循环特异性条形码组合。在其它实施方案中,用于分析物的结合剂的编码标签包含用于组合的分析物结合循环信息的独特编码器条形码。在任一种方法中,共同的间隔序列可用于每个结合循环中的结合剂的编码标签。(B)在该例子中,来自每个结合循环的结合剂具有短的结合循环特异性条形码以识别所述结合循环,其与识别所述结合剂的所述编码器条形码一起提供独特的组合条形码,其识别特别的结合剂-结合循环组合。(C)在所述结合循环完成后,使用加帽循环步骤可将所述延伸记录标签转化为可扩增的文库,其中例如,包含连接到通用引物序列P2的通用引物序列P1和间隔序列Sp’的帽最初通过互补的P1和P1’序列与延伸记录标签退火,使所述帽子接近所述延伸记录标签。所述延伸记录标签中的互补Sp和Sp’序列和帽退火及引物延伸将所述第二通用引物序列(P2)添加到所述延伸记录标签。

[0298] 图36A-F.用于演示从编码标签到记录标签的信息传递的基于DNA的模型系统。通过寡核苷酸模型系统演示了示例性结合和分子内写入。编码标签中的靶向剂A’和B’设计来与记录标签中的靶结合区A和B杂交。通过汇集两个相同浓度的重新编码标签saRT_Abc_v2(A靶标)和saRT_Bbc_V2(B靶标)来制备记录标签(RT)混合物。记录标签在其5’末端生物素化,并含有独特靶结合区、通用正向引物序列、独特的DNA条形码,和8碱基共同间隔序列(Sp)。所述编码标签包含独特的编码器条形码碱基,其侧翼为8个碱基共同间隔序列(Sp’),其中一个通过聚乙二醇接头共价连接到A或B靶试剂。(A)将生物素化的记录标签寡核苷酸(saRT_Abc_v2和saRT_Bbc_V2)和生物素化的Dummy-T10寡核苷酸一起固定在链霉亲和素标

记的珠子。所述记录标签设计有A或B捕获序列(分别被同源结合剂-A'和B'识别)和相应的条形码(rtA_BC和rtB_BC)以鉴定所述结合靶标。该模型系统中的所有条形码选自所述65个15-mer条形码组(SEQ ID NO:1-65)。在某些情况下,15-mer条形码被组合以构成更长的条形码,以便于凝胶分析。特别是,rtA_BC=BC_1+BC_2;rtB_BC=BC_3。还合成了与所述记录标签的所述A和B序列同源的结合剂的两个编码标签,即CT_A'-bc(编码器条形码=BC_5)和CT_B'-bc(编码器条形码=BC_5+BC_6)。对所述编码标签序列的一部分(留下单链Sp'序列)的互补封闭寡核苷酸(DupCT_A'BC和DupCT_AB'BC)任选地在编码标签在退火至所述珠子固定的记录标签之前预先退火至所述编码标签。链置换聚合酶在聚合酶延伸期间去除所述阻断寡核苷酸。条形码图例(插图)表明对所述记录标签和编码标签中的所述功能条形码的15-mer条形码分配。(B)所述记录标签条形码设计和编码标签编码器条形码设计提供了对记录标签和编码标签之间的"分子内"对"分子间"相互作用的简单凝胶分析。在这种设计中,不期望的"分子间"相互作用(A记录标签与B'编码标签;以及B记录标签与A'编码标签)产生的凝胶产物比期望的"分子内"(A记录标签与A'编码标签;B记录标签与B'编码标签)反应产物更长或更短15个碱基。所述引物延伸步骤将A'和B'编码标签条形码(ctA'_BC,ctB'_BC)改变为反向补体条形码(ctA_BC和ctB_BC)。(C)引物延伸测定演示了从编码标签到记录标签的信息转移,以及用于PCR分析的通过退火的EndCap寡核苷酸上的引物延伸添加衔接头序列。(D)通过使用Dummy-T20寡核苷酸滴定记录标签的表面密度来优化"分子内"信息传递。将生物素化的记录标签寡核苷酸与生物素化的Dummy-T20寡核苷酸以1:0、1:10的各种比例混合,一直到1:10000。在降低的记录标签密度(1:10^3和1:10^4)下,"分子内"相互作用相对于"分子间"相互作用占主导地位。(F)作为所述DNA模型系统的简单扩展,虽然图示了含有Nano-Tag15肽-链霉亲和素结合对的简单蛋白质结合系统(KD~4nM)(Perbandt等人,2007,Proteins 67:1147-1153),但是可以使用任何数量的肽结合剂模型系统。Nano-Tag15肽序列是(fM)DVEAWLGARVPLVET(SEQ ID NO:131)(fM=甲酰基-Met)。Nano-Tag15肽还包含短的、柔性的接头肽(GGGGS)和用于偶联所述DNA记录标签的半胱氨酸残基。其它实例肽标签-同源结合剂对包括:钙调蛋白结合肽(CBP)-钙调蛋白(KD~2pM)(Mukherjee等,2015,J.Mol.Biol.427:2707-2725)、淀粉样蛋白-β(Aβ16-27)肽-US7/Lcn2 anticalin(0.2nM)(Rauth等,2016,Biochem.J.473:1563-1578)、PA标签/NZ-1抗体(KD~400pM)、FLAG-M2Ab(28nM)、HA-4B2 Ab(1.6nM)和Myc-9E10 Ab(2.2nM)(Fujii等,2014,Protein Expr.Purif.95:240-247)。(E)作为通过引物延伸对分子内信息从所述结合剂编码标签向所述记录标签转移的测试,结合互补DNA序列"A"的寡核苷酸"结合剂"可用于测试和开发。该杂交事件基本上强于fM亲和力。链霉亲和素可以用作Nano-tag15肽表位的测试结合剂。虽然所述肽标签-结合剂互作是高度亲和的,但是很容易采用酸性和/或高盐洗涤破坏(Perbandt等,同上)。

[0299] 图37A-B.使用纳米或微乳化PCR将信息从UMI标记的N或C末端转移至DNA标签标记多肽体。(A)多肽在其N-或C-末端用包含独特分子标识符(UMI)的核酸分子标记。所述UMI可以侧接用于引发后续PCR的序列。然后所述多肽用单独的DNA标签在其内部位点进行"身体标记",所述单独的DNA标签包含与所述UMI侧翼的引发序列互补的序列。(B)将得到的标记多肽乳化并进行乳化PCR(ePCR)(或者,可以进行乳化体外转录-RT-PCR(IVT-RT-PCR)反应或其它合适的扩增反应)以扩增N-或C-端UMI。形成微乳液或纳米乳液,使得平均液滴直径

为50-1000nm,并且平均每个液滴少于一个多肽。PCR前后的液滴内容快照分别显示在左侧面板和右侧面板中。所述UMI扩增子通过互补引发序列与所述内部多肽体DNA标签杂交,并且所述UMI信息通过引物延伸从所述扩增子转移至所述内部多肽体DNA标签。

[0300] 图38.单细胞蛋白质组学。在含有聚合物形成亚单元(例如丙烯酰胺)的液滴中包封并裂解将细胞。聚合所述聚合物形成亚单元(例如,聚丙烯酰胺),且将蛋白质交联到所述聚合物矩阵。破坏所述乳液液滴并释放含有附着于所述可渗透聚合物矩阵的单细胞蛋白裂解物的聚合凝胶粒。通过在裂解和包封缓冲液中包含变性剂如尿素,将所述蛋白质以其天然构象或变性状态交联到所述聚合物矩阵。使用本领域已知的或本文公开的许多方法包括采用条形码珠粒的乳化或组合索引,将包含隔室条形码和其它记录标签组分(例如,通用引发序列(P1),间隔序列(Sp),任选的独特分子标识符(UMI))的记录标签附着到所述蛋白质上。含有所述单细胞蛋白质的所述聚合凝胶珠粒也可在加入记录标签后进行蛋白酶消化,以产生适合肽测序的记录标签标记的肽。在某些实施方案中,所述聚合物矩阵可以设计成溶解在适合的添加剂中,例如二硫化物交联聚合物,其在暴露于还原剂如三(2-羧乙基)膦(TCEP)或二硫苏糖醇(DTT)时破裂。

[0301] 图39A-E.使用双功能N-末端氨基酸(NTAA)修饰剂和嵌合裂解试剂增强氨基酸切割反应。(A)和(B)连接到固相基质上的肽用双功能NTAA修饰剂修饰,例如生物素-苯基异硫氰酸酯(PITC)。(C)使用链霉亲和素-Edmanase嵌合蛋白将低亲和力的Edmanase(>μM Kd)募集到生物素-PITC标记的NTAA中。(D)由于生物素-链霉亲和素相互作用导致局部有效浓度增加,因此所述Edmanase切割的效率大大提高。(E)所述切割的生物素-PITC标记的NTAA和相关的链霉亲和素-Edmanase嵌合蛋白在切割后扩散开。也可以采用其它许多生物共轭招募策略。叠氮化物修饰的PITC是可商购的(4-叠氮基苯基异硫氰酸酯,Sigma),可以将叠氮化物-PITC简单转化为PITC的其它生物缀合物,例如生物素-PITC通过点击化学与炔烃-生物素反应。

[0302] 图40A-I:从蛋白质裂解物产生C-末端记录标签-标记的肽(可以包封在凝胶珠中)。(A)使变性多肽与酸酐反应以标记赖氨酸残基。在一个实施方案中,使用炔(mTet)-取代的柠康酐+丙酸酐的混合物与mTet标记所述赖氨酸。(显示为条纹状矩形)(B)结果是炔(mTet)标记的多肽,其中一部分赖氨酸被具有丙酸基团(在所述多肽链上显示为正方形)封闭。所述炔烃(mTet)基团可用于基于DNA标记的点击化学。(C)用于炔烃或mTet基团的DNA标签(显示为实心矩形)分别通过使用叠氮化物或反式环辛烯(TCO)标签连接。(D)如图31中所示,使用引物延伸步骤将条形码和功能元件如间隔区(Sp)序列和通用引发序列附加到DNA标签上以产生记录标签标记的多肽。所述条形码可以是样品条形码、分区条形码、隔室条形码、空间位置条形码等,或其任何组合。(E)将得到的所述记录标签标记的多肽用蛋白酶或化学方法片段化成记录标签标记的肽。(F)为了图示,显示了用两个记录标签标记的肽片段。(G)将包含与所述记录标签中的所述通用引物序列互补的通用引物序列的DNA标签连接到肽的C末端。所述C末端DNA标签还包含用于将所述肽缀合至表面的基团。(H)所述C末端DNA标签中的互补通用引物序列和随机选择的记录标签退火。分子内引物延伸反应用于将信息从所述记录标签转移至所述C-末端DNA标签。(I)通过马来酸酐将所述肽上的所述内部记录标签偶联到赖氨酸残基上,该偶联在酸性pH下是可逆的。所述内部记录标签在酸性pH下从肽的赖氨酸残基上切割下来,留下所述C端记录标签。新暴露的赖氨酸残基可任选地用

不可水解的酸酐如丙酸酐封闭。

[0303] 图41.NGPS测定的优选实施方案的工作流程。

[0304] 图42A-D.NGPS测序分析的示例性步骤。记录标签标记的表面结合肽上的N-末端氨基酸(NTAA)乙酰化或酰胺化步骤可以在NTAA结合剂结合之前或之后发生,其取决于NTAA结合剂是否已经被工程化以结合乙酰化NTAA或本地NTAAs。在第一种情况下,(A)所述肽最初在所述NTAA处通过采用乙酸酐的化学手段或采用N-末端乙酰转移酶(NAT)的酶促乙酰化。(B)所述NTAA被NTAA结合剂例如工程化的anticalin、氨酰基tRNA合成酶(aaRS)、ClpS等识别。DNA编码标签附着于所述结合剂并包含识别所述特定的NTAA接合剂的条形码编码序列。(C)在通过所述NTAA结合剂结合所述乙酰化NTAA后,所述DNA编码标签通过互补序列与所述记录标签瞬时退火,并且所述编码标签信息通过聚合酶延伸转移至所述记录标签。在另一个替代实施方案中,所述记录标签信息通过聚合酶延伸转移到所述编码标签。(D)通过工程化的酰基肽水解酶(APH)从所述肽上切割所述乙酰化的NTAA,所述工程化的酰基肽水解酶催化末端乙酰化氨基酸从乙酰化肽的水解。在所述乙酰化的NTAA裂解后,所述循环从所述新暴露的NTAA的乙酰化开始重复它自己。虽然N-末端乙酰化用作为NTAA修饰/切割的示例性模式,但是其他N-末端基团,例如脒基基团,可以被切割化学中的伴随变化取代。如果使用胍基化,可以使用0.5-2%NaOH溶液在温和条件下裂解所述鸟苷酸化的NTAA(参见Hamada,2016,其全部内容通过引用并入)。APH是丝氨酸肽酶,其能够催化从封闭的肽中去除Nα-乙酰化氨基酸,并且它属于所述脯氨酰寡肽酶(POP)家族(氏族SC,家族S9)。它是真核细胞、细菌和古细胞中N末端乙酰化蛋白的关键调节因子。

[0305] 图43A-B:示例性记录标签-编码标签设计特征。(A)示例性记录标签相关蛋白(或肽)和与相关编码标签绑定的结合剂(例如,anticalin)的结构。将胸苷(T)碱基插入所述编码标签上的间隔区(Sp')和条形码(BC')序列之间,以在引物延伸反应中容纳随机的非模板化3'末端的腺苷(A)加成。(B)DNA编码标签通过SpyCatcher-SpyTag蛋白-肽相互作用与结合剂(例如,anticalin)连接。

[0306] 图44A-E.使用裂解剂与记录标签(A)和(B)的杂交增强NTAA裂解反应。附着于固相基质(例如珠子)的记录标签标记的肽在所述NTAA(Mod)上被修饰或标记,例如,用PITC、DNP、SNP、乙酰基修饰剂、胍基化等。(C)A裂解酶(例如,酰基肽水解酶(APH)、氨基肽酶(AP)、Edmanase等)附着于DNA标签,该DNA标签包含与记录标签上的通用引物序列互补的通用引物序列。通过所述裂解酶的DNA标签和所述记录标签上的互补通用引物序列的杂交,将所述裂解酶募集到所述修饰的NTAA中。(D)该杂交步骤极大地提高了所述裂解酶对所述NTAA的有效亲和力。(E)所述切割的NTAA扩散开,并且可以通过剥离所述杂交的DNA标签除去相关的切割酶。

[0307] 图45.使用肽连接酶+蛋白酶+二氨基肽酶的循环降解肽测序。Butelase I将所述TEV-Butelase I肽底物(TENLYFQNHV,SEQ ID NO:132)连接至待测肽的所述NTAA。Butelase需要肽底物C末端的NHV基序。连接后,使用烟草蚀纹病毒(TEV)蛋白酶在谷氨酰胺(Q)残基后切割所述嵌合的肽底物,留下具有连接到待测肽的N-末端的天冬酰胺(N)残基的嵌合肽。二氨基肽酶(DAP)或二肽基肽酶,其从所述N-末端切割两个氨基酸残基,通过两个氨基酸有效去除待测肽上的天冬酰胺残基(N)和原始NTAA以缩短N端添加的待测肽。使用本文提供的结合剂读取所述新暴露的NTAA,然后对测序的"n"个氨基酸重复整个循环"n"次。链霉亲和

素-DAP金属酶嵌合蛋白的应用以及将生物素基团束缚在N-末端天冬酰胺残基上可以控制DAP合成能力。

详细说明

[0309] 本文未具体定义的术语应由本领域技术人员根据本公开和上下文给出含义。然而,如说明书中所使用的,除非有相反的说明,否则这些术语具有指示的含义。

[0310] I.介绍

[0311] 本公开内容部分地提供了高度平行的、高通量数字表征和定量大分子的方法,其直接应用于蛋白和肽的表征和测序(参见图1B、图2A)。本文所述的方法使用包含编码标签的结合剂,所述编码标签具有核酸分子或可测序聚合物形式的识别信息,其中所述结合剂与感兴趣的大分子相互作用。进行多个连续的结合循环,每个循环包含将固定在固体支持物上的多个大分子,优选代表汇集样品,暴露于多种结合剂。在每个结合循环期间,通过将来自结合剂编码标签的信息传递至与大分子共定位的记录标签,来记录结合大分子的每种结合剂的身份,也可选记录结合循环数。在一个替代实施方案中,来自包含相关大分子的识别信息的记录标签的信息可以传递到所结合的结合剂的编码标签(例如,形成延伸编码标签)或第三“di-tag”构建体。多个循环的结合事件在与大分子共定位的记录标签上构建历史结合信息,从而产生包含多个编码标签的延伸记录标签,所述多个编码标签以共线性顺序代表给定大分子的临时结合历史。另外,可以采用循环特定编码标签来跟踪来自每个循环的信息,这样如果一个循环由于某种原因被跳过,所述延伸记录标签可以继续在后续循环中收集信息,并识别信息缺失的循环。

[0312] 或者,可以将信息从包含相关大分子的识别信息的记录标签传递到形成延伸编码标签或第三个di-tag构建体的编码标签,以替代将信息从所述编码标签写入或传递到所述记录标签。可以在每个结合循环后收集形成的延伸编码标签或di-tags,便于随后的序列分析。在包含条形码(例如,分区标签、隔室标签、样本标签、分级物标签、UMI或其任何组合)的所述记录标签上的所述识别信息可用于将所述延伸编码标签或di-tag序列读数映射回原始大分子。以这种方式,产生了代表大分子结合历史的核酸编码文库。可以扩增该核酸编码文库,并使用高通量的下一代数字测序方法进行分析,使每次运行能够分析数百万至数十亿分子。创建核酸的结合信息编码库在另一方面也有用,因为它能够通过利用杂交的基于DNA的技术进行富集、扣减和标准化。这些基于DNA的方法易于快速扩展和定制,并且比直接操作其它类型的大分子文库(如蛋白质库)的方法更具成本效益,因此,可以在测序之前通过一种或多种技术处理核酸编码的结合信息文库,以富集和/或扣减和/或标准化所述序列的代表。这使得可以更有效、更快速、更经济高效地从非常大的文库提取最需要的信息,这些文库中的个体成员最初可能会在丰度上呈现多个数量级的变化。重要的是,这些用于操作文库代表的基于核酸的技术与更常规的方法正交,并且可以与它们组合使用。例如,可以使用基于蛋白质的方法扣减常见的高丰度蛋白,例如白蛋白,可以去除大部分但不是所有不需要的蛋白质。随后,还可以扣减延伸记录标签文库的白蛋白特异性成员,从而实现更完整的整体扣减。

[0313] 一个方面,本公开内容提供了使用Edman样降解方法进行肽测序的高度并行化的方法,允许从大量DNA记录标签标记的肽(例如,数百万至数十亿)进行测序。这些记录标签

标记的肽衍生自蛋白样品的蛋白水解消化或有限水解,并且所述记录标签标记的肽以适当的分子间间隔随机固定在测序基质(例如,多孔珠)上。具有小化学基团的肽的N-末端氨基酸(NTAA)残基的修饰,例如苯基硫代氨基甲酰基(PTC),二硝基苯酚(DNP),磺酰基硝基苯酚(SNP),丹酰基,7-甲氧基香豆素,乙酰基或胍基,催化或募集NTAA裂解反应允许循环控制所述Edman样降解过程。所述修饰的化学基团还可以为同源NTAA结合剂提供增强的结合亲和力。通过结合包含编码标签的同源NTAA结合剂,并将编码标签信息(例如,提供结合剂的识别信息的编码器序列)从所述编码标签传递到所述记录标签(例如,引物延伸或连接),来鉴定每种固定化肽的修饰的NTAA。随后,通过化学方法或酶促方法去除所述修饰的NTAA。在某些实施方案中,酶(例如,Edmanase)被工程化以催化去除所述修饰的NTAA。在其他实施方案中,可以改造天然存在的外肽酶,例如氨肽酶或酰基肽水解酶,从而仅在存在合适的化学修饰条件下切割末端氨基酸。

[0314] II.定义

[0315] 在以下描述中,阐述了某些具体细节以便提供对各种实施例的透彻理解。然而,本领域技术人员将理解,可以在没有这些细节的情况下制备和使用本发明化合物。在其他情况下,未详细示出或描述公知的结构以避免不必要地模糊对实施方案的描述。除非上下文另有要求,否则在整个说明书和随后的权利要求中,词语“包括”及其变体,例如“包括”和“包含”,应以开放的,包含性的意义解释,即“包括但不仅限于”。另外,术语“包括”(和诸如“包含”或“包含”或“具有”或“包括”的相关术语)并不旨在排除在其他某些实施方式中,例如,任何组合物的实施方式。本文所述的物质,组合物,方法或过程等可以由所描述的特征“组成”或“基本上由.....组成”。本文提供的标题仅为了方便,并不解释所要求保护的实施方案的范围或含义。

[0316] 贯穿本说明书对“一个实施方案”或“实施方案”的引用意味着结合该实施方案描述的特定特征,结构或特性包括在至少一个实施方案中。因此,贯穿本说明书在各个地方出现的短语“在一个实施方案中”或“在实施方案中”不一定都指的是同一实施方案。此外,特定特征,结构或特性可以在一个或多个实施方案中以任何合适的方式组合。

[0317] 正如这里所使用的,单数形式“一个(a)”、“一个(an)”及“所述”包含提及物的复数形式,除非另有说明。因此,例如,提及“肽”,包括一种或多种肽或肽的混合物。此外,除非特别说明或从上下文中显而易见,否则如本文所用,术语“或”应理解为包括在内并涵盖“或”和“和”两者。

[0318] 如本文所用,术语“大分子”包括由较小亚单元组成的大的分子。大分子的实例包括但不限于肽、多肽、蛋白、核酸、碳水化合物、脂质、大环化合物。大分子还包括嵌合大分子,其由两种或多种类型的大分子共价连接在一起(例如,肽连接到核酸)的组合组成。大分子还可包括“大分子组装体”,其由两种或更多种大分子的非共价复合物组成。大分子组装体可以由相同类型的大分子(例如蛋白-蛋白)或两种不同类型的大分子(例如蛋白-DNA)组成。

[0319] 如本文所用,术语“肽”包括肽、多肽和蛋白,并且是指包含通过肽键连接的两个或更多个氨基酸的链的分子。一般而言,具有超过20-30个氨基酸的肽通常被称为多肽,而具有超过50个氨基酸的肽通常被称为蛋白。所述肽的氨基酸最典型地是L-氨基酸,但也可以是D-氨基酸、修饰的氨基酸、氨基酸类似物、氨基酸模拟物、或其任何组合。肽可以是天然发

生的、合成产生的、或重组表达的。肽还可以包含修饰所述氨基酸链的其他基团,例如,通过翻译后修饰添加的官能团。

[0320] 如本文所用,术语“氨基酸”是指包含胺基、羧酸基、和对每个氨基酸特异的侧链的有机化合物,其用作肽的单体亚单元。氨基酸包括20种标准的,天然存在的或典型的氨基酸以及非标准氨基酸。标准的天然氨基酸包括丙氨酸(A或Ala),半胱氨酸(C或Cys),天冬氨酸(D或Asp),谷氨酸(E或Glu),苯丙氨酸(F或Phe),甘氨酸(G或Gly),组氨酸(H或His),异亮氨酸(I或Ile),赖氨酸(K或Lys),亮氨酸(L或Leu),蛋氨酸(M或Met),天冬酰胺(N或Asn),脯氨酸(P或Pro)),谷氨酰胺(Q或Gln),精氨酸(R或Arg),丝氨酸(S或Ser),苏氨酸(T或Thr),缬氨酸(V或Val),色氨酸(W或Trp)和酪氨酸(Y或Tyr))。氨基酸可以是L-氨基酸或D-氨基酸。非标准氨基酸可以是修饰氨基酸、氨基酸类似物、氨基酸模拟物、非标准蛋白原氨基酸或天然存在或化学合成的非蛋白原氨基酸。非标准氨基酸的实例,包括但不限于,硒代半胱氨酸、吡咯赖氨酸和N-甲酰基甲硫氨酸、β-氨基酸、同源氨基酸、脯氨酸和丙酮酸衍生物、3-取代丙氨酸衍生物、甘氨酸衍生物、环-取代的苯丙氨酸和酪氨酸衍生物、线性核心氨基酸、N-甲基氨基酸。

[0321] 如本文所用,术语“翻译后修饰”是指在肽被核糖体翻译完成后在肽上发生的修饰。翻译后修饰可以是共价修饰或酶修饰。翻译后修饰的实例,包括但不限于,酰化、乙酰化、烷基化(包括甲基化)、生物素化、丁酰化、氨基甲酰化、羰基化、脱酰胺、deiminiation、二萘胺形成、二硫桥形成、消除(eliminylation)、黄素附着、甲酰化、γ-羧化、谷氨酰化、甘氨酰化、糖基化、glypiation、血红素C附着、羟基化、hypusine形成、碘化、异戊二烯化、脂化(lipidation)、脂质化(lipoylation)、丙酰化、甲基化、肉豆蔻酰化、氧化、棕榈酰化、聚乙二醇化、磷酸酯化、磷酸化、异戊烯化、丙酰化、视黄基希夫碱基形成(retinylidene Schiff base formation)、S-谷胱甘肽化、S-亚硝基化,S-亚磺酰化、硒化、琥珀酰化、硫化、泛素化和C-末端酰胺化。翻译后修饰包括肽的氨基末端和/或羧基末端的修饰。所述末端氨基的修饰包括但不限于脱氨基,N-低级烷基、N-二低级烷基(N-di-lower alkyl)、和N-酰基修饰。所述末端羧基的修饰包括但不限于酰胺、低级烷基酰胺、二烷基酰胺和低级烷基酯修饰(例如,其中低级烷基是C_1-C_4烷基)。翻译后修饰还包括,例如但不限于上述的落在氨基和羧基末端之间的氨基酸的修饰。术语翻译后修饰还可以包括包含一种或多种可检测标记的肽修饰。

[0322] 如本文所用,术语“结合剂”是指与大分子或大分子的组分或特征结合、关联、联合、识别或与其组合的核酸分子、肽、多肽、蛋白、碳水化合物或小分子。结合剂可以与大分子或大分子的组分或特征形成共价关联或非共价关联。结合剂也可以是嵌合结合剂,由两种或多种类型的分子组成,例如核酸分子-肽嵌合结合剂或碳水化合物-肽嵌合结合剂。结合剂可以是天然发生的,合成产生的或重组表达的分子。结合剂可以结合大分子的单个单体或亚单元(例如,肽的单个氨基酸)或结合大分子的多个连接的亚单元(例如,二肽、三肽、长肽的更高级肽、多肽或蛋白分子)。结合剂可以结合线性分子或具有三维结构(也称为构象)的分子。例如,抗体结合剂可以与线性肽、多肽或蛋白结合,或与构象肽、多肽或蛋白结合。结合剂可以结合N-末端肽,C-末端肽、或肽、多肽或蛋白质分子的的干预肽。结合剂可以结合肽分子的N-末端氨基酸,C-末端氨基酸或干预氨基酸。相比于未修饰或未标记的氨基酸,结合剂可以优先结合化学修饰的或标记的氨基酸。例如,结合剂可以优选地与用乙酰基

团、脒基基团、丹酰基基团、PTC基团、DNP基团、SNP部分等修饰的氨基酸结合而不是与不具有这些基团的氨基酸结合。结合剂可以结合肽分子的翻译后修饰。结合剂可以表现出与大分子的组分或特征的选择性结合(例如,结合剂可以选择性地结合20种可能的天然氨基酸残基中的一种并且与其它19种天然氨基酸残基以具有非常低的亲和力结合或者根本不结合)。结合剂可以表现出较低的选择性结合,其中结合剂能够结合大分子的多种组分或特征(例如,结合剂可以以相似的亲和力与两种或更多种不同氨基酸残基结合)。结合剂包含编码标签,其通过接头与所述结合剂连接。

[0323] 如本文所用,术语“接头”是指用于连接两个分子的核苷酸、核苷酸类似物、氨基酸、肽、多肽或非核苷酸化学基团中的一种或多种。接头可用于结合结合剂和码标签、结合记录标签和大分子(例如肽)、结合大分子和固体支持物等、结合记录标签和固体支持物。在某些实施方案中,接头通过酶促反应或化学反应(例如,点击化学)连接两个分子。

[0324] 如本文所用,术语“蛋白质组学”是指定量分析指细胞、组织和体液内的所述蛋白质组,以及分析所述蛋白质组在所述细胞内和组织内的相应空间分布。此外,蛋白质组学研究包括所述蛋白质组的所述动态状态,作为生物学的以及定义的生物或化学刺激的功能在时间内的持续变化。

[0325] 如本文所用,术语“非同源结合剂”是指其在特定结合循环反应中,不能以低亲和力结合大分子特征、组分、或被测亚单元,“同源结合剂”,而“同源结合剂”,其以高亲和力结合相应的大分子特征、组分、或亚单元。例如,如果肽分子的酪氨酸残基在结合反应中被测,则非同源结合剂是指以低亲和力结合或根本不结合所述酪氨酸残基的那些结合剂,使得在适合来自于同源结合剂的编码标签信息转移到所述记录标签的条件下,所述非同源结合剂不能将编码标签信息转移到所述记录标签。或者,如果在结合反应中肽分子的酪氨酸残基被测,则非同源结合剂是以低亲和力结合或根本不结合所述酪氨酸残基的那些试剂,使得在适合于涉及延伸编码标签而不是延伸记录标签的实施方案的条件下,记录标签信息不能有效地转移至所述编码标签。

[0326] 具有自由氨基的肽链一端的末端氨基酸在本文中称为“N-末端氨基酸”(NTAA)。具有自由羧基的链另一端的末端氨基酸在本文中称为“C-末端氨基酸”(CTAA)。构成肽的氨基酸可以按顺序号,其中所述肽的长度为“n”个氨基酸。如本文所用,NTAA被认为是第n个氨基酸(在本文中也称为“nNTAA”)。使用这种命名法,下一个氨基酸是第n-1个氨基酸,然后是第n-2个氨基酸,依此类推从N末端到C末端的肽长度。在某些实施方案中,NTAA、CTAA或两者可以用所述化学基团修饰或标记。

[0327] 如本文所用,术语“条形码”是指约2至约30个碱基的核酸分子(例如,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或30碱基),其为大分子(例如,蛋白质,多肽,肽)、结合剂、来自结合循环的结合剂组、样品大分子、样品组、隔室内(例如,液滴、珠子或隔离的位置)的大分子、隔室内的大分子集、大分子分级物、大分子的分级物集、空间区域或空间区域集、大分子库或结合剂库提供独特标识符标签或起源信息。条形码可以是人工序列或天然发生的序列。在某些实施例中,条形码群体内的每个条形码是不同的。在其他实施方案中,条形码群体中的一部分条形码是不同的,例如,条形码群体中至少约10%,15%,20%,25%,30%,35%,40%,45%,50%,55%60%,65%,70%,75%,80%,85%,90%,95%,97%或99%的条形码是不同的。条形码群可以随机生成

或非随机生成。在某些实施例中,条形码群是纠错条形码。条形码可用于计算地解卷积多路复用测序数据并识别源自单个大分子、样品、文库等的序列读数。条形码还可以用于对大分子集合进行去卷积,所述大分子集合已经分布到小隔室中以增强映射。例如,不是将肽映射回所述蛋白质组,而是将所述肽映射回其起源蛋白分子或蛋白复合物。

[0328] “样品条形码”,也称为“样品标签”,识别大分子来自哪个样品。

[0329] “空间条形码”,其是大分子衍生的2-D或3-D组织切片的区域。空间条形码可用于组织切片上的分子病理学。空间条形码允许来自组织切片的多个样品或文库的多路复用测序。

[0330] 如本文所用,术语“编码标签”是指约2个碱基至约100个碱基,包括任何包括2和100且在其间的整数的核酸分子,其包含它的相关结合剂的识别信息。“编码标签”也可以由“可序列聚合物”制成(参见,例如,Niu等人,2013,Nat.Chem.5:282-292;Roy等人,2015,Nat.Commun.6:7237;Lutz,2015,Macromolecules 48:4759-4767;其中每一篇均通过引用整体并入)。编码标签包含编码器序列,其可选地在一侧侧接一个间隔子或在每侧侧接间隔子。编码标签还可以包含可选的UMI和/或可选的结合循环特异性条形码。编码标签可以是单链或双链的。双链编码标签可包含平末端、突出末端或两者。编码标签可以指直接连接到结合剂的编码标签、与直接连接到结合剂的编码标签杂交的互补序列(例如,用于双链编码标签)、或延伸记录标签中的编码标签信息。在某些实施方案中,编码标签可以进一步包含结合循环特异性间隔子或条形码,独特分子标识符、通用引发位点或其任何组合。

[0331] 如本文所用,术语“编码器序列”或“编码器条形码”是指约2个碱基至约30个碱基的(例如,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或30个碱基)长度的核酸分子,提供其相关的结合剂的识别信息。所述编码器序列可以唯一地标识其关联的结合试剂。在某些实施方案中,编码器序列提供其相关结合剂以及使用该结合剂的结合循环的识别信息。在其它实施例中,编码器序列与编码标签内的单独的结合循环特异性条形码组合。或者,所述编码器序列可以将其关联的结合试剂识别为属于包含两个或更多个不同结合剂的组的成员。在一些实施方案中,这种水平的识别足以用于分析目的。例如,在涉及结合氨基酸的结合剂的一些实施方案中,知道肽在特定位置包含两种可能的氨基酸之一可能就足够了,而不用明确地鉴定该位置的氨基酸残基。在另一个实例中,共同的编码序列用于多克隆抗体,其包含识别蛋白质靶标的一个以上表位的抗体的混合物,并且具有不同的特异性。在其它实施方案中,在编码器序列识别一组可能的结合剂的情况下,可以使用次序解码方法来产生每个结合剂的独特识别。这是通过在重复的结合循环中改变给定结合剂的编码序列来实现的(参见Gunderson等,2004,Genome Res.14:870-7)。当与来自其他循环的编码信息组合时,来自每个结合循环的所述部分识别编码标签信息为所述结合试剂产生的独特标识符,例如,编码标签而不是单独的编码标签(或编码器序列)的特定组合提供为所述结合试剂产生所述独特标识信息。优选地,结合剂库内的所述编码序列具有相同或相似数量的碱基。

[0332] 如本文所用,术语“结合循环特异性标签”,“结合循环特异性条形码”或“结合循环特异性序列”是指用于鉴定在特定结合循环内使用的结合剂文库的独特序列。结合循环特异性标签可包含约2个碱基至约8个碱基(例如,2,3,4,5,6,7,或8个碱基)的长度。结合循环特异性标签可以并入结合剂的编码标签内,作为间隔序列的一部分、编码器序列的一部分、

UMI的一部分、或作为所述编码标签内的单独组件。

[0333] 如本文所用,术语“间隔子”(Sp)是指长度约1碱基至约20个碱基(例如,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,或20碱基)的核酸分子,存在于记录标签或编码标签的末端。在某些实施方案中,间隔序列在一端或两端侧接编码标签的编码序列。在结合剂与大分子结合后,在相关的编码标签上和记录标签上的互补间隔序列之间的分别退火,让结合信息通过引物延伸反应或连接转移到所述记录标签、编码标签或di-互补间隔序列构建体。Sp'是指与Sp互补的间隔序列。优选地,结合剂库内的间隔序列具有相同数量的碱基。可以在结合剂库中使用共同的(共有的或相同的)间隔区。间隔序列可具有“循环特异性”序列,以跟踪特定结合循环中使用的结合剂。所述间隔序列(Sp)在所有结合循环中可以是恒定的,对特定类别的大分子是特异性的,或者是结合循环数特异性的。大分子类别特异性间隔子允许将来自于完成的结合/延伸循环中的延伸记录标签中存在的同源结合剂的编码标签信息退火到另一个结合剂的所编码标签,该另一结合剂在随后的结合循环中通过所述类别特异性间隔子识别相同类别的大分子。只有正确的同源对的所述顺序结合才能产生相互作用的间隔元件和有效的引物延伸。间隔子序列可包含足够数量的碱基以与记录标签中的互补间隔区序列退火以引发引物延伸(也称为聚合酶延伸)反应,或提供用于连接反应的“夹板”,或介导“粘性末端”连接反应。间隔序列可包含比编码标签内的所述编码器序列更少数量的碱基。

[0334] 如本文所用,术语“记录标签”是指核酸分子或可测序聚合物分子(参见,例如,Niu等人,2013,Nat.Chem.5:282-292;Roy等人,2015,Nat。公报6:7237;Lutz,2015,Macromolecules 48:4759-4767;其中每一篇均通过引用整体并入本文),其包含相关大分子的识别信息。在某些实施方案中,在结合剂结合大分子后,来自与结合剂连接的编码标签的信息能转移到与所述大分子相关的所述记录标签,同时所述结合剂结合到所述大分子上。在其他实施方案中,在结合剂结合大分子后,来自与大分子相关的记录标签的信息可以转移到与所述结合剂连接的所述编码标签,同时结合剂结合到大分子上。重新编码标签可以直接与大分子连接、通过多功能接头与大分子连接、或者借助于其在固体支持物上的相近(或共定位)而与大分子结合。记录标签可以通过其5'端或3'端或内部位点连接,只要该链接与用于将编码标签信息转移到所述记录标签的方法兼容,反之亦然。记录标签还可以包括其它功能组件,例如通用引发位点、独特分子标识符、条形码(例如,样品条形码、分级物条形码、空间条形码、隔室标签等)、与编码标签的间隔序列互补的间隔序列、或其任何组合。在使用聚合酶延伸将编码标签信息转移到所述记录标签的实施方案中,记录标签的所述间隔序列优选位于所述记录标签的3'末端。

[0335] 如本文所用,术语“引物延伸”,也称为“聚合酶延伸”,是指由核酸聚合酶(例如,DNA聚合酶)催化的反应,其中与互补链退火的核酸分子(例如,寡核苷酸引物、间隔序列)使用所述互补链作为模板通过聚合酶延伸。

[0336] 如本文所用,术语“独特的分子标识符”或“UMI”是指约3至约40个碱基的核酸分子如本文所用,术语“独特的分子标识符”或“UMI”是指约3至约40个碱基(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,或40碱基)长度的核酸分子,为每个与UMI连接的大分子(例如肽)或结合剂提供独特标识符标签。大分子UMI可用于从多个延伸记录标签计算去卷积测序数据,以识别源

自单个大分子的延伸记录标签。结合剂UMI可用于鉴定结合特定大分子的每种个体结合剂。例如,UMI可用于鉴定对特定肽分子发生的单个氨基酸特异性的结合剂的个体结合事件的数量。应当理解,当UMI和条形码都在结合剂或大分子的背景下被引用时,所述条形码指的个体结合剂或大分子的除UMI之外的识别信息(例如,样品条形码、隔室条形码、结合循环条形码)。

[0337] 如本文所用,术语“通用引发位点”或“通用引物”或“通用引发序列”是指核酸分子,其可用于文库扩增和/或测序反应。通用引发位点可以包括,但不限于用于PCR扩增的引发位点(引物序列)、与流通槽表面上的互补寡核苷酸退火的流通槽适配序列(其促使在一些下一代测序平台中进行桥扩增)、测序引发位点、或其组合。通用引发位点可用于其他类型的扩增,包括通常与下一代数字测序结合使用的扩增。例如,延伸记录标签分子可以环化,并且通用引发位点用于滚环扩增以形成可以用作测序模板的DNA纳米球(Drmanac等,2009,Science 327:78-81)。或者,记录标签分子可以通过来自通用引发位点的聚合酶延伸直接环化和测序(Korlach等,2008,Proc.Natl.Acad.Sci.105:1176-1181)。当在“通用引发位点”或“通用引物”的上下文中使用时,术语“正向”也可称为“5’”或“有义”。当在“通用引发位点”或“通用引物”的上下文中使用时,术语“反向”也可称为“3’”或“反义”。

[0338] 如本文所用,术语“延伸记录标签”是指记录标签,至少一种结合剂的编码标签(或其互补序列)的信息在所述结合剂与大分子结合后已转移到所述记录标签。所述编码标签的信息可以直接(例如,连接)或间接(例如引物延伸)转移到所述记录标签。编码标签的信息可以酶促或化学地转移到所述记录标签。延伸记录标签可以包含编码标签的信息可以酶促或化学地转移到记录标签。延伸记录标签可以包含1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,45,50,55,60,65,70,75,80,85,90,95,100,125,150,175,200或或更多个编码标签的结合剂信息。所述延伸记录标签的碱基序列可以反映由其编码标签鉴定的结合剂的结合时间和顺序。可以反映由所述编码标签鉴定的结合剂的部分结合顺序,或者可以不反映任何由所述编码标签鉴定的结合剂的结合顺序。在某些实施例中,所述延伸记录标签中存在的所述编码标签信息至少25%,30%,35%,40%,45%,50%,55%,60%,65%,70%,75%。,80%,85%,90%,91%,92%,93%,94%,95%,96%,97%,98%,99%或100%同一性代表所分析的大分子序列。在延伸记录标签不以100%同一性代表分析的大分子序列的某些实施方案中,错误可能是由于结合剂的脱靶结合,或由于“错过的”结合循环(例如,因为结合剂在结合循环期间不能与大分子结合,因为引物延伸反应失败),或两者兼而有之。

[0339] 如本文所用,术语“延伸编码标签”是指在结合编码标签的结合剂与大分子结合后,至少一个记录标签(或其互补序列)的信息被转移到其的编码标签,所述记录标签与所述大分子相关。记录标签的信息可以直接(例如,连接)或间接(例如引物延伸)转移到所述编码标签。记录标签的信息可以酶促或化学方式转移。在某些实施例中,延伸编码标签包含记录标签的信息,反映一个结合事件。如本文所用,术语“di-tag”或“di-tag构建体”是指核酸分子,至少一个记录标签(或其互补序列)和至少一个编码标签(或其互补序列)的信息在与所述编码标签连接的结合剂结合后转移到该核酸分子,所述记录标签与所述大分子相关(见,图11B)。记录标签和编码标签的信息可以间接地(例如,引物延伸)转移到所述di-tag。记录标签的信息可以酶促或化学方式转移。在某些实施方案中,di-tag包含记录标签的

UMI、记录标签的隔室标签、记录标签的通用引发位点、编码标签的UMI、编码标签的编码序列、结合循环特异性条形码、编码标签的通用引发位点、或其任何组合。

[0340] 如本文所用,术语“固体支持物”、“固体表面”、或“固体基质”或“基质”是指任何固体材料,包括多孔和非多孔材料,大分子(例如,肽)可以是通过本领域已知的手段,包括共价和非共价相互作用,或其任何组合,直接或间接结合上去。固体支持物可以是二维的(例如平面表面)或三维的(例如凝胶矩阵或珠粒)。固体支持物可以是任何支持表面,包括但不限于,珠子、微珠、阵列、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶芯片、流控芯片、流通槽、包括信号转导电子器件的生物芯片、通道、微量滴定孔、ELISA板、旋转干涉测量盘、硝酸纤维素膜、基于硝酸纤维素的聚合物表面、聚合物矩阵、纳米颗粒、或微球。用于固体载体的材料包括但不限于丙烯酰胺、琼脂糖、纤维素、硝化纤维素、玻璃、金、石英、聚苯乙烯、聚乙烯乙酸乙烯酯、聚丙烯、聚甲基丙烯酸酯、聚乙烯、聚环氧乙烷、聚硅酸盐、聚碳酸酯、特氟隆、碳氟化合物、尼龙、硅橡胶、聚酸酐、聚乙醇酸、聚乳酸、聚原酸酯、官能化硅烷、聚丙基延胡索酸酯、胶原、糖胺聚糖、聚氨基酸、葡聚糖、或其任何组合。固体支持物还包括薄膜、膜、瓶、盘、纤维、编织纤维、成形聚合物,例如管、颗粒、珠子、微球、微粒、或其任何组合。例如,当固体表面是珠子时,所述珠子可以包括但不限于陶瓷珠子、聚苯乙烯珠子、聚合物珠子,甲基苯乙烯珠子、琼脂糖珠子、丙烯酰胺珠子、实心核心珠子、多孔珠、顺磁珠,玻璃珠或可控孔珠。珠子可以是球形或不规则形状。珠粒的尺寸可以为纳米,例如100纳米、毫米,例如1毫米。在某些实施方案中,珠子的尺寸范围为约0.2微米至约200微米,或约0.5微米至约5微米。在一些实施方案中,珠子可以直径约为1,1.5,2,2.5,2.8,3,3.5,4,4.5,5,5.5,6,6.5,7,7.5,8,8.5,9,9.5,10,10.5,15,或20μm。在某些实施方案中,“珠粒”固体支持物可以指个体珠子或多个珠子。

[0341] 如本文所用,术语“核酸分子”或“多核苷酸”是指含有通过3’-5’磷酸二酯键连接的脱氧核糖核苷酸或核糖核苷酸的单链或双链多核苷酸,以及多核苷酸类似物。核酸分子包括,但不限于DNA,RNA和cDNA。多核苷酸类似物可以具有除天然多核苷酸中发现的标准磷酸二酯键以外的骨架,并且任选地具有核糖或脱氧核糖以外的修饰的糖基。多核苷酸类似物含有能够通过Watson-Crick碱基配对与标准多核苷酸碱基形成氢键的碱基,其中所述类似物骨架以允许寡核苷酸类似物分子与标准多核苷酸中的碱基之间的序列特异性方式的氢键的方式呈现所述碱基。多核苷酸类似物的实例包括,但不限于异种核酸(XNA)、桥接核酸(BNA)、乙二醇核酸(GNA)、肽核酸(PNAs)、γPNA、吗啉代多核苷酸、锁核酸(LNAs)、苏糖核酸(TNA),2’-O-甲基多核苷酸、2’-O-烷基核糖基取代的多核苷酸、硫代磷酸酯多核苷酸、和硼磷酸盐多核苷酸。多核苷酸类似物可具有嘌呤或嘧啶类似物,包括例如7-脱氮嘌呤类似物、8-卤嘌呤类似物、5-卤代嘧啶类似物、或可与任何碱基配对的通用碱类似物,包括次黄嘌呤、硝基唑、异喹诺酮类似物、唑类羧酰胺和芳族三唑类似物、或具有额外功能的碱基类似物,例如用于亲和结合的生物素基团。

[0342] 如本文所用,“核酸测序”指确定核酸分子或核酸分子样品中核苷酸的顺序。

[0343] 如本文所用,“下一代测序”是指高通量测序方法,其允许并行测序数百万至数十亿的分子。下一代测序方法的实例包括通过合成法测序,连接测序、杂交测序,Polony测序、离子半导体测序和焦磷酸测序。通过将引物连接到固体基质和与核酸分子的互补序列,核酸分子可以通过所述引物与固体基质杂交,然后通过使用聚合酶在固体基质上的离散区域

中产生多个拷贝来扩增(这些分组有时被称为聚合酶克隆polymerase colonies或polonies)。因此,在测序过程中,特定位置的核苷酸可以被多次测序(例如,数百次或数千次)-这种覆盖深度被称为“深度测序”。高通量核酸测序技术的实例包括Illumina、BGI、Qiagen、Thermo-Fisher和Roche提供的平台,包括如如Service(Science 311:1544-1546,2006)所综述的平行珠子阵列、合成测序、连接测序、毛细管电泳、电子微芯片、“生物芯片”、“微阵列、平行微芯片和单分子阵列格式。

[0344] 如本文所用,“单分子测序”或“第三代测序”是指新一代测序方法,其中来自单分子测序仪器的读数通过对单个DNA分子的测序产生。与依赖于扩增以平行克隆许多DNA分子以便以分阶段方法进行测序的下一代测序方法不同,单分子测序审查单个DNA分子并且不需要扩增或同步。单分子测序包括需要在每次碱基并入入(“洗涤和扫描”循环)后暂停测序反应的方法和不需要在读取步骤之间停止的方法。单分子测序方法的实例包括单分子实时测序(Pacific Biosciences)、基于纳米孔的测序(Oxford Nanopore)、双链中断纳米孔测序,以及用高级显微镜的直接成像DNA。

[0345] 如本文所用,“分析”所述大分子意指大分子的全部或部分组分进行量化、表征、区分或其它们的组合。例如,分析肽、多肽或蛋白质包括测定所述肽的全部或部分氨基酸序列(连续或非连续)。分析大分子还包括对所述大分子的组分的部分鉴定。例如,部分鉴定所述大分子蛋白质序列中的氨基酸可以将蛋白质中的氨基酸鉴定为属于可能的氨基酸子集。分析通常从分析第n个NTAA开始,然后进行到所述肽的下一个氨基酸(即,n-1,n-2,n-3等等)。这通过切割所述第n NTAA,从而将肽的第n-1氨基酸转化为N-末端氨基酸(本文称为“n-1NTAA”)来实现。分析所述肽还可以包括确定肽上翻译后修饰的存在和频率,其可以包括或不包括关于肽上翻译后修饰的先后顺序的信息。分析肽还可以包括确定肽中表位的存在和频率,其可以包括或不包括关于在所肽内的所述表位的顺序或位置的信息。分析所述肽可以包括组合不同类型的分析,例如获得表位信息、氨基酸序列信息、翻译后修饰信息、或其任何组合。

[0346] 如本文所用,术语“隔室”是指将大分子子集与大分子样品分离或分隔的物理区域或体积。例如,隔室可以将单个细胞与其他细胞分离,或者将样品蛋白质组的一部分与样品的蛋白质组的其余部分分开。隔室可以是含水隔室(例如,微流体液滴)、固体隔室(例如,板上的Picotiter孔或微量滴定孔、管、小瓶、凝胶珠),或表面上的分离区域。隔室可包含一个或多个可固定大分子的珠子。

[0347] 如本文所用,术语“隔室标签”或“隔室条形码”是指包含在一个或多个隔室内(例如,微流体液滴)的组分(例如,单个细胞的蛋白质组)的识别信息的约4个碱基至约100个碱基(包括4个碱基,100个碱基和其间的任何整数)的单链或双链核酸分子。隔室条形码识别样品中的大分子子集,例如蛋白样品的子集,其已经从多个(例如,数百万到数十亿)隔室分离进入相同的物理隔室或隔室组。因此,即使在将组分汇集在一起之后,隔室标签也可用于区分来自具有相同隔室标签的一个或多个隔室的成分与具有不同隔室标签的另一隔室中的成分。通过用独特的隔室标签标记每个区室内或两个或更多个区室的组内的所述蛋白和/或肽,可以鉴定衍生自相同蛋白、蛋白复合物或单个区室或区室组内的细胞的肽。隔室标签包含条形码和任选的通用引物,所述条形码任选地在一侧或两侧侧接间隔序列。所述间隔序列可以与记录标签的所述间隔序列互补,使得能够将隔室标签信息转移到所述记录

标签。隔室标签还可以包含通用引发位点、独特分子标识符(用于提供与其连接的肽的识别信息)或两者,特别是对于隔室标签中包含用于下述肽分析方法的记录标签的实施方案。隔室标签可包含用于偶联肽的功能基团(例如,醛、NHS、mTet、炔等)。或者,隔室标签可包含肽,所述肽包含对蛋白连接酶的识别序列,以允许将所述隔室标签连接至目标肽。隔室可以包括单个隔室标签、为可选的UMI序列预留的多个相同的隔室标签、或者两个或更多个不同的隔室标签。在某些实施例中,每个隔室包含独特隔室标签(一对一映射)。在其它实施例中,来自较大隔室群的多个隔室包含相同的隔室标签(多对一映射)。隔室标签可以连接到隔室(例如,珠子)内的固体支持物上或者连接到隔室自身的表面(例如,Picotiter孔的表面)。或者,隔室标签在隔室内的溶液中可以是游离。

[0348] 如本文所用,术语"分区"是指将独特条形码随机分配给来自样品内的大分子群的大分子亚群。在某些实施方案中,可以通过将大分子分配到隔室中来实现分区。隔板可以由单个隔室内的大分子或来自一组隔室中的多个隔室的大分子组成。

[0349] 如本文所用,"分区标签"或"分区条形码"是指包含约4个碱基至约100个碱基(包括4个碱基,100个碱基和其间的任何整数)的单链或双链核酸分子,其包含分区的识别信息。在某些实施方案中,用于大分子的分区标签是指由将大分子分隔成用相同条形码标记的隔室产生的相同隔室标签。

[0350] 如本文所用,术语"分级物"是指样品中的大分子(例如,蛋白)的子集,该样品已经采用物理或化学分离方法从其余样品或细胞器挑选出来,例如按大小、疏水性、等电点、亲和力等分级。分离方法包括HPLC分离、凝胶分离、亲和分离、细胞分馏、细胞器分馏、组织分馏等。流体流动性、磁性、电流、质量、密度及其类似物理性质也可用于分离。

[0351] 如本文所用,术语"分级物条形码"是指包含约4个碱基至约100个碱基(包括4个碱基,100个碱基,和其间的任何整数)的单链或双链核酸分子,其包含对分级物中所述大分子的识别信息。

[0352] III.分析大分子的方法

[0353] 此处描述的所述方法提供了用于大分子分析的高度平行化的方法。高度多路复用化的大分析结合检测被转换成通过下一代测序方法读取的核酸分子文库。在此提供的所述方法对于蛋白和肽测序尤其有用。

[0354] 在优选的实施方案中,蛋白质样品在单分子水平上用至少一种核酸记录标签标记,所述核酸记录标签包括条形码(例如,样品条形码、隔室条形码)和任选的独特分子标识符。所述蛋白质样品经历蛋白水解消化以产生一群记录标签标记的肽(例如,数百万至数十亿)。这些记录标签标记的肽随机地汇集和固定在固体支持物上。所述汇集的、固定的记录标签标记的肽经历多个连续的结合循环,每个结合循环包括暴露于用包含识别相关结合剂的编码器序列的编码标签标记的多种结合剂(例如,所有20种天然存在的氨基酸的结合剂)。在每个结合循环期间,通过将结合剂的编码标签信息传递到所述记录标签(或将记录标签信息传递到编码标签或将记录标签信息和编码标签信息传递到单独的di-tag构建体)来捕获关于结合剂与所述肽的结合的信息。完成结合循环后,生成代表所测定的肽的结合历史的延伸记录标签(或延伸编码标签或di-tag构建体)的文库,其可使用非常高通量的下一代数字测序方法进行分析。在所述记录标签中使用核酸条形码允许大量肽测序数据的解卷积,例如,以鉴定肽序列源自哪个样品、细胞、蛋白质组的子集或蛋白质。

[0355] 在一个方面,提供分析大分子的方法,包括以下步骤:(a)提供大分子和与固体支持物连接的相关或共定位的记录标签;(b)使所述大分子与能够结合所述大分子的第一结合剂接触,其中所述第一结合剂包含具有关于所述第一结合剂的识别信息的第一编码标签;(c)将所述第一编码标签的信息传送到所述记录标签,以产生第一次序延伸记录标签;(d)使所述大分子与能够结合所述大分子的第二结合剂接触,其中所述第二结合剂包含第二编码标签,其具有关于所述第二结合剂的识别信息;(e)将所述第二编码标签的信息传送到所述第一次序延伸记录标签,以产生第二次序延伸记录标签;以及(f)分析所述第二次序延伸标签(参见例如图2A-D)

[0356] 在一些实施方案中,所述接触步骤(b)和(d)按顺序进行,例如,所述第一次结合剂和所述第二结合剂在分开的结合循环反应中与所述大分子接触。在另一些实施方案中,所述接触步骤(b)和(d)同时进行,例如,在单个结合循环反应中包含所述第一次结合剂、所述第二结合剂,以及可选地额外的结合剂。在优选实施方案中,所述接触步骤(b)和(d)包含是所述大分子与多个结合剂接触。

[0357] 在一些实施方案中,所述的方法还包括在步骤(e)和(f)之间的以下步骤:(x)通过用能够结合所述大分子的第三(或更高级)结合剂替换所述第二结合剂重复步骤(d)和(e)一次或多次,其中所述第三(或更高次序)结合剂包含第三(或更高次序)编码标签,其具有关于第三(或更高次序)结合剂的识别信息;(y)将所述第三(或更高次序)编码标签的信息传送到所述第二(或更高次序)延伸记录标签,以产生第三(或更高次序)延伸记录标签;以及(z)分析所述第三(或更高次序)延伸记录标签。

[0358] 所述第三(或更高级)结合剂可以在与所述第一结合剂和所述第二结合剂分开的结合循环反应中与所述大分子接触。或者,所述第三(或更高级)结合剂可以和所述第一结合剂,所述第二结合剂一起在单个结合循环反应中与所述大分子接触。

[0359] 在第二方面,提供的分析大分子的方法包括以下步骤:(a)提供大分子,相关的第一记录标签和与固体支持物连接的相关第二记录标签;(b)使所述大分子与能够结合所述大分子的第一结合剂接触,其中所述第一结合剂包含具有关于所述第一结合剂的识别信息的第一编码标签;(c)将所述第一编码标签的信息传送到所述记录标签以产生第一次序延伸记录标签;(d)使所述大分子与能够结合所述大分子的第二结合剂接触,其中所述第二结合剂包含第二编码标签,其具有关于所述第二结合剂的识别信息;(e)将所述第二编码标签的信息传递到所述第二记录标签以产生第二次序延伸记录标签;以及(f)分析所述第一和第二延伸记录标签。

[0360] 在一些实施方案中,接触步骤(b)和(d)按顺序进行,例如,所述第一次结合剂和所述第二结合剂在分开的结合循环反应中与所述大分子接触。在另一些实施方案中,所述接触步骤(b)和(d)同时进行,例如,在单个结合循环反应中包含所述第一次结合剂、所述第二结合剂,以及可选地额外的结合剂。

[0361] 在一些实施方案中,步骤(a)还包括提供与所述固体支持物连接的相关的第三(或更高次序)记录标签。在进一步实施方案中,所述的方法还包括,在步骤(e)和(f)之间的以下步骤:(x)通过用能够结合所述大分子的第三(或更高级)结合剂替换所述第二结合剂重复步骤(d)和(e)一次或多次,其中所述第三(或更高级)结合剂包含第三(或更高次序)编码标签,其具有关于第三(或更高次序)结合剂的识别信息;(y)将所述第三(或更高次序)编码

标签的信息传送到所述第二(或更高次序)延伸记录标签,以产生第三(或更高次序)延伸记录标签;以及(z)分析所述第三(或更高次序)延伸记录标签。

[0362] 所述第三(或更高级)结合剂可以与所述第一结合剂和所述第二结合剂在分开的结合循环反应中与所述大分子接触。或者,所述第三(或更高级)结合剂可以和所述第一结合剂,所述第二结合剂一起在单个结合循环反应中与所述大分子接触。

[0363] 在一些实施方案中,所述第一编码标签、第二编码标签和任何更高次序编码标签包含结合循环特异性间隔序列。

[0364] 在第三方面,提供的分析大分子的方法包括以下步骤:(a)提供肽,和与固体支持物连接的相关记录标签;(b)用化学基团修饰所述肽的N-末端氨基酸(NTAA)以产生修饰的NTAA;(c)使所述肽与能够结合修饰的NTAA的第一结合剂接触,其中所述第一结合剂包含具有所述第一结合剂的识别信息的第一编码标签;(d)将所述第一编码标签的信息传送到所述记录标签以产生延伸记录标签;(e)分析所述延伸记录标签(参考例如图3)。

[0365] 在一些实施方案在,其中步骤(c)还包含使所述肽与包含第二(或更高次序)编码标签的第二(或更高次序)结合剂接触,所述第二(或更高次序)编码标签具有所述第二(或更高次序)结合剂的识别信息,其中所述第二(或更高次序)结合剂能够结合步骤(b)的修饰的NTAA之外的修饰的NTAA。在进一步的实施方案中,在所述肽与所述第一结合剂接触后,按次序使肽与所述第二(或更高级)结合剂接触,例如,第一结合剂和第二(或更高级)结合剂接触与所述肽在分开的结合循环反应中接触。在其它实施方案中,使所述肽与所述第二(或更高级)结合剂接触和所述肽与所述第一结合剂接触同时发生,例如,在单个结合循环反应中包含第一结合剂和第二(或更高级)结合剂。

[0366] 在某些实施方案中,所述化学基团通过化学反应或酶促反应添加到NTAA。

[0367] 在一些实施方案中,用于修饰所述NTAA的化学基团是苯基硫代氨基甲酰基(PTC)、二硝基苯基(DNP)、磺酰基硝基苯基(SNP)、丹酰基;7-甲氧基香豆素;硫代酰基;硫代乙酰基;乙酰基;胍基;或硫代苄基。

[0368] 通过化学试剂可向所述NTAA添加化学基团。在某些实施方案中,用于修饰具有PTC基团的NTAA的化学试剂是异硫氰酸苯酯或其衍生物;用于修饰具有DNP基团的NTAA的化学试剂是2,4-二硝基苯磺酸(DNBS)或芳基卤化物如1-氟-2,4-二硝基苯(DNFB);用于修饰具有磺酰氧基硝基苯(SNP)基团的NTAA的化学试剂是4-磺酰基-2-硝基氟苯(SNFB);用于修饰具有丹磺酰基的NTAA的化学试剂是磺酰氯,如丹磺酰氯;用于修饰具有7-甲氧基香豆素基团的NTAA的化学试剂是7-甲氧基香豆素乙酸(MCA);用于修饰具有硫代酰基的NTAA的化学试剂是硫代酰化试剂;用于修饰具有硫代乙酰基的NTAA的化学试剂是硫代乙酰化试剂;用于修饰具有乙酰基NTAA的化学试剂是乙酰化试剂(如乙酸酐);用于修饰具有胍基(脒基)NTAA的化学试剂是胍基化试剂,或用于修饰具有硫代苄基的NTAA的化学试剂是硫代苄基化试剂。

[0369] 本公开提供的第四方面,是提供了分析肽的方法,包括以下步骤:(a)提供与固体支持物连接的肽和相关的记录标签;(b)用化学基团修饰所肽的N-末端氨基酸(NTAA)以产生修饰的NTAA;(c)使所述肽与能够结合所述修饰的NTAA的第一结合剂接触,其中所述第一结合剂包含具有关于第一结合剂的识别信息的第一编码标签;(d)将所述第一编码标签的信息传送到所述记录标签,以产生第一延伸记录标签;(e)移除修饰的NTAA以暴露新的

NTAA;(f)用化学基团修饰所述肽的新NTAA以产生新修饰的NTAA;(g)使所述肽与能够结合所述新修饰的NTAA的第二结合剂接触,其中所述第二结合剂包含第二编码标签,其具有关于第二结合剂的识别信息;h)将所述第二编码标签的信息传送到所述第一延伸记录标签,以产生第二延伸记录标签;(i)分析第二延伸记录标签。

[0370] 在一些实施方案中,所述接触步骤(c)和(g)按次序进行,例如,所述第一次结合剂和所述第二结合剂在分开的结合循环反应中与所述肽接触。

[0371] 在一些实施方案中,所述的方法还包括在步骤(h)和(i)之间的以下步骤:(x)通过用能够结合所述修饰的NTAA的第三(或更高级)结合剂替换所述第二结合剂重复步骤(e)、(f)和(g)一次或多次,其中所述第三(或更高级)结合剂包含第三(或更高次序)编码标签,其具有关于第三(或更高次序)结合剂的识别信息;(y)将所述第三(或更高次序)编码标签的信息传送到所述第二(或更高次序)延伸记录标签,以产生第三(或更高次序)延伸记录标签;以及(z)分析所述第三(或更高次序)延伸记录标签。

[0372] 在某些实施方案中,所述化学基团通过化学反应或酶促反应添加到所述NTAA。

[0373] 在一些实施方案中,用于修饰所述NTAA的化学基团是苯基硫代氨基甲酰基(PTC)、二硝基苯基(DNP)、磺酰基硝基苯基(SNP)、丹酰基;7-甲氧基香豆素;硫代酰基;硫代乙酰基;乙酰基;胍基;或硫代苄基。

[0374] 通过化学试剂可向所述NTAA添加化学基团。在某些实施方案中,用于修饰具有PTC基团的NTAA的化学试剂是异硫氰酸苯酯或其衍生物;用于修饰具有DNP基团的NTAA的化学试剂是2,4-二硝基苯磺酸(DNBS)或芳基卤化物如1-氟-2,4-二硝基苯(DNFB);用于修饰具有磺酰氧基硝基苯(SNP)基团的NTAA的化学试剂是4-磺酰基-2-硝基氟苯(SNFB);用于修饰具有丹磺酰基的NTAA的化学试剂是磺酰氯,如丹磺酰氯;用于修饰具有7-甲氧基香豆素基团的NTAA的化学试剂是7-甲氧基香豆素乙酸(MCA);用于修饰具有硫代酰基的NTAA的化学试剂是硫代酰化试剂;用于修饰具有硫代乙酰基的NTAA的化学试剂是硫代乙酰化试剂;用于修饰具有乙酰基NTAA的化学试剂是乙酰化试剂(如乙酸酐);用于修饰具有胍基(脒基)NTAA的化学试剂是胍基化试剂,或用于修饰具有硫代苄基的NTAA的化学试剂是硫代苄基化试剂。

[0375] 在第五方面,提供的分析肽的方法包括以下步骤:(a)提供肽和连接在固体支持物上的相关的记录标签;(b)使所述肽与能够结合所述肽的N-末端氨基酸(NTAA)的第一结合试剂接触,其中所述第一结合试剂包含具有所述第一结合试剂的识别信息的第一编码标签;(c)将所述第一编码标签的信息传送到所述记录标签以产生延伸记录标签;以及(d)分析所述延伸记录标签。

[0376] 在一些实施方案在,其中步骤(b)还包含使所述肽与包含第二(或更高次序)编码标签的第二(或更高次序)结合剂接触,所述第二(或更高次序)编码标签具有所述第二(或更高次序)结合剂的识别信息,其中所述第二(或更高次序)结合剂能够结合所述肽的所述NTAA之外的NTAA。在进一步的实施方案中,在所述肽与所述第一结合剂接触后,按次序使肽与所述第二(或更高级)结合剂接触,例如,第一结合剂和第二(或更高级)结合剂接触与所述肽在分开的结合循环反应中接触。在其它实施方案中,使所述肽与所述第二(或更高级)结合剂接触和所述肽与所述第一结合剂接触同时发生,例如,在单个结合循环反应中包含第一结合剂和第二(或更高级)结合剂。

[0377] 在第六方面,提供的分析肽的方法包括以下步骤:(a)提供肽和连接在固体支持物上的相关的记录标签;(b)使所述肽与能够结合所述肽的所述N末端氨基酸(NTAA)的第一结合剂接触,其中所述第一结合剂包含具有关于第一结合剂的识别信息的第一编码标签;c)将所述第一编码标签的信息传送到所述记录标签,以产生第一延伸记录标签;(d)移除修饰的NTAA以暴露新的NTAA;(e)使所述肽与能够结合所述新修饰的NTAA的第二结合剂接触,其中所述第二结合剂包含第二编码标签,其具有关于第二结合剂的识别信息;(f)将所述第二编码标签的信息传送到所述第一延伸记录标签,以产生第二延伸记录标签;以及(g)分析第二延伸记录标签。

[0378] 在一些实施方案中,所述的方法还包括在步骤(f)和(g)之间的以下步骤:(x)通过用能够结合所述修饰的大分子的第三(或更高级)结合剂替换所述第二结合剂重复步骤(d)、(e)和(f)一次或多次,其中所述第三(或更高级)结合剂包含第三(或更高次序)编码标签,其具有关于第三(或更高次序)结合剂的识别信息;以及(y)将所述第三(或更高次序)编码标签的信息传送到所述第二(或更高次序)延伸记录标签以产生第三(或更高次序)延伸记录标签;以及在步骤(g)中分析所述第三(或更高次序)延伸记录标签。

[0379] 在一些实施方案中,所述接触步骤(b)和(e)按次序进行,例如,所述第一次结合剂和所述第二结合剂在分开的结合循环反应中与所述肽接触。

[0380] 在提供的任何实施方案中,所述方法包含平行分析多个大分子。在优选实施方案中,所述方法包含平行分析多个肽。

[0381] 在本文提供的任何实施方案中,大分子(或肽)与结合剂接触的步骤包括使大分子(或肽)与多种结合剂接触。

[0382] 在提供的任何实施方案中,所述大分子可以是蛋白,多肽或肽。在进一步实施方案中,其中所述肽是通过片段化分离自生物样品的蛋白获得。

[0383] 在提供的任何实施方案中,所述大分子可以是或包含碳水化合物、脂质、核酸或大环。

[0384] 提供的任一实施方案中,其中所述记录标签是DNA分子、具有修饰碱基的DNA分子、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子(Dragulescu-Andrasi et al., 2006,J.Am.Chem.Soc.128:10258-10267),GNA分子或其任意组合。

[0385] 在此提供的任一实施方案中,其中所述的记录标签包含通用引发位点。在进一步实施方案中,其中所述通用引发位点包含用于扩增、测序或二者结合的引物位点。

[0386] 在此提供的任一实施方案中,所述记录标签包含独特分子标识符、隔室标签、分区条形码,样品条形码、分级物条形码、间隔序列或它们的任意组合。

[0387] 在此提供的任一实施方案中,所述记录标签包含独特分子标识符(UMI)、编码器序列、结合循环特异性序列、间隔序列、或它们的任意组合。

[0388] 在此提供的任一实施方案中,所述编码标签中的所述结合循环特异性序列可以是结合循环特异性间隔序列。

[0389] 在某些实施方案中,结合循环特异性序列被编码为与所述编码器序列分开的条形码。

[0390] 在其它实施方案中,所述编码器序列和结合循环特异性序列在单个条形码中列出,该条形码对于所述结合剂和每个结合循环是唯一的。

[0391] 在一些实施方案中,所述间隔序列包含共同的结合循环序列,其在来自多个结合循环的结合剂之间共享。在其它实施方案中,所述间隔序列包含独特的结合循环序列,其在来自相同结合循环的结合剂之间共享。

[0392] 在此提供的任一实施方案中,其中所述的记录标签可以包含条形码。

[0393] 在此提供的任一实施方案中,所述大分子和相关的记录标签可以共价结合到所述固体支持物上。

[0394] 在此提供的任一实施方案中,其中所述固体支持物是珠子、多孔珠子、多孔基质、阵列、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、微量滴定孔、ELISA板、旋转干涉测量盘、硝酸纤维素膜、硝化纤维素基聚合物表面、纳米颗粒或微球。

[0395] 在此提供的任一实施方案中,其中所述固体支持物是聚苯乙烯珠子、聚合物珠子、琼脂糖珠子、丙烯酰胺珠子、固体核心珠子、多孔珠子、顺磁珠子、玻璃珠子或可控孔珠子。

[0396] 在此提供的任一实施方案中,多个大分子和相关的记录标签可以结合到所述固体支持物上。在进一步实施方案中所述多个大分子在所述固体支持物上以≥50nm,≥100nm,或≥200nm的平均距离间隔开。

[0397] 在此提供的任一实施方案中,所述结合剂可以是多肽或蛋白。在进一步实施方案中,其中所述结合剂是修饰的或变体的氨肽酶、修饰的或变体的氨酰tRNA合成酶、修饰的或变体的anticalin,或者修饰的或变体的ClpS。

[0398] 在此提供的任一实施方案中,所述结合剂可以选择性地结合所述大分子。

[0399] 在此提供的任一实施方案中,所述记编码标签可以是DNA分子、具有修饰碱基的DNA分子、RNA分子、BNA分子、XNA分子、LNA分子、GNA分子、PNA分子、γPNA分子或其组合。

[0400] 在此提供的任一实施方案中,所述结合剂和所述编码标签可以通过接头连接。

[0401] 在此提供的任一实施方案中,其中所述的结合剂和所述编码标签通过SpyTag/SpyCatcher或SnoopTag/SnoopCatcher肽-蛋白质对连接(Zakeri,etal.,2012,Proc Natl Acad Sci U S A 109(12):E690-697;Veggiani et al.,2016,Proc.Natl.Acad.Sci.USA 113:1202-1207,每一篇都通过整体引用并入本文)。

[0402] 在此提供的任一实施方案中,其中所述记录标签到编码标签的信息传递通过DNA连接酶介导。或者,其中所述记录标签到编码标签的信息传递通过DNA合成酶或化学连接介导。

[0403] 在此提供的任一实施方案中,分析所述延伸记录标签可以包含核酸测序。在进一步实施方案中,核酸测序方法是合成测序、连接测序、杂交测序、polony测序、离子半导体测序或焦磷酸测序。在其它实施方案中,核酸测序方法是单分子实时测序、基于纳米孔测序、或使用高级显微镜的DNA直接成像。

[0404] 在此提供的任一实施方案中,所述延伸记录标签在分析前被扩增。

[0405] 在此提供的任一实施方案中,其中包含于所述延伸记录标签中的所述编码标签信息的顺序可以通过所述结合剂到大分子提供关于结合顺序的信息以及被所述结合剂检测到的分析物的所述序列信息。

[0406] 在此提供的任一实施方案中,其中包含于所述延伸记录标签中的特定编码标签信息的频率(例如编码器序列)可以通过特定结合剂到所述大分子提供关于结合频率的信息

以及被所述结合剂检测到的所述大分子中的所述分析物的频率的信息。

[0407] 在此公开的任何实施方案中,多个大分子(例如,蛋白质)样品可以汇集,其中每个样品中的大分子群体用包含样品特异性条形码的记录标签标记。这样汇集的大分子样品可以在单个反应管中进行结合循环。

[0408] 在此公开的任何实施方案中,其中代表多个大分子的多个延伸记录标签可以平行分析。

[0409] 在此公开的任何实施方案中,其中代表多个大分子的多个延伸记录标签可以在多路复用检测中分析。

[0410] 在此提供的任一实施方案中,所述多个延伸记录标签在分析之前进行靶标丰度检测。

[0411] 在此提供的任一实施方案中,所述多个延伸记录标签在分析之前进行扣减检测。

[0412] 在此提供的任一实施方案中,所述多个延伸记录标签在分析之前进行标准化检测以降低高丰度种类。

[0413] 在此提供的任一实施方案中,所述NTAA采用修饰的氨肽酶、修饰的氨基酸tRNA合成酶、温和型埃德曼降解、埃德曼酶或无水TFA来移除。

[0414] 在此提供的任一实施方案中,其中至少一种结合剂结合到末端氨基酸残基。在一些实施方案中,所述末端氨基酸残基是N-末端氨基酸或C-末端氨基酸。

[0415] 在此提供的任一实施方案中,其中至少一种结合剂结合到翻译后修饰的氨基酸。

[0416] 前述实施方案的特征在以下部分提供进一步细节。

[0417] IV大分子

[0418] 在一个方面,提供分析大分子的方法,包括以下步骤:大分子是由较小的亚基组成的大分子。在某些实施方案中,大分子是蛋白质、蛋白质复合物、多肽、肽、核酸分子、碳水化合物、脂质、大环或嵌合大分子。

[0419] 根据本文公开的方法分析的大分子(例如,蛋白质,多肽,肽)可以从合适的来源或样品获得,包括但不限于:生物样品,如细胞(原代细胞和培养细胞系),细胞裂解物或提取物,细胞器或囊泡,包括外泌体,组织和组织提取物;活检;粪便;体液(如血液,全血,血清,血浆,尿液,淋巴液,胆汁,脑脊液,间质液,水溶液或玻璃体液,初乳,痰液,羊水,唾液,肛门和阴道分泌物,汗液和精液,渗出物,渗出物(例如,从脓肿或任何其他感染或炎症部位获得的液体)或从关节(正常关节或受类风湿性关节炎,骨关节炎,痛风或化脓性关节炎等疾病影响的关节)获得的液体任何生物体,含有哺乳动物的样品,包括含有微生物组的样品,是优选的和人源样品,特别优选地包括含微生物组的样品;环境样品(如空气,农业,水和土壤样品);微生物样品包括来自微生物生物膜和/或群落的样品,以及微生物孢子;研究样品,包括细胞外液、来自于细胞培养物的细胞外上清液、细菌中的包涵体、包括线粒体区室的细胞区室和细胞周质。

[0420] 在某些实施方案中,大分子是蛋白质、蛋白质复合物、多肽或肽。将肽、多肽或蛋白质的氨基酸序列信息和翻译后修饰转导到核酸编码的文库中,该文库可以通过下一代测序方法进行分析。肽可包含L-氨基酸、D-氨基酸或两者。肽、多肽、蛋白质或蛋白质复合物可包含标准的天然存在的氨基酸、修饰的氨基酸(例如,翻译后修饰)、氨基酸类似物、氨基酸模拟物或其任何组合。在一些实施方案中,肽、多肽或蛋白质是天然存在的、合成产生的或重

组表达的。在任何上述肽实施方案中,肽、多肽、蛋白质或蛋白质复合物可进一步包含翻译后修饰。

[0421] 标准的天然氨基酸包括:丙氨酸(A或Ala)、半胱氨酸(C或Cys)、天冬氨酸(D或Asp)、谷氨酸(E或Glu)、苯丙氨酸(F或Phe)、甘氨酸(G或Gly)、组氨酸(H或His)、异亮氨酸(I或Ile)、赖氨酸(K或Lys)、亮氨酸(L或Leu)、蛋氨酸(M或Met)、天冬酰胺(N或Asn)、脯氨酸(P或Pro)、谷氨酰胺(Q或Gln)、精氨酸(R或Arg)、丝氨酸(S或Ser)、苏氨酸(T或Thr)、缬氨酸(V或Val)、色氨酸(W或Trp)和酪氨酸(Y或Tyr)。非标准氨基酸包括硒代半胱氨酸、吡咯赖氨酸和N-甲酰基甲硫氨酸、β-氨基酸、同源氨基酸、脯氨酸和丙酮酸衍生物、3-取代丙氨酸衍生物、甘氨酸衍生物、环取代苯丙氨酸和酪氨酸衍生物、线性核心氨基酸和N-甲基氨基酸。

[0422] 肽、多肽或蛋白质的翻译后修饰(PTM)可以是共价修饰或酶促修饰。翻译后修饰的实例包括,但不限于酰化、乙酰化、烷基化(包括甲基化)、生物素化、丁酰化、氨基甲酰化、羰基化、脱酰胺、脱酰胺、二萘胺形成、二硫桥形成、消除、黄素附着、甲酰化、γ-羧化、谷氨酰化、甘氨酰化、糖基化(例如,N-连接、O-连接、C-连接、磷酸糖基化)、glypiation、血红素C连接、羟基化、hypusine形成、碘化、异戊二烯化、脂化、脂质化、丙酰化、甲基化、肉豆蔻酰化、氧化、棕榈酰化、聚乙二醇化、磷酸酯化、磷酸化、异戊烯化、丙酰化、亚视黄基席夫碱形成、S-谷胱甘肽化、S-亚硝基化、S-亚磺酰化、硒化、琥珀酰化、硫化、泛素化和C-末端酰胺化。翻译后修饰包括肽、多肽或蛋白质的氨基末端和/或羧基末端的修饰。末端氨基的修饰包括但不限于脱氨基、N-低级烷基、N-二低级烷基和N-酰基修饰。末端羧基的修饰包括但不限于酰胺、低级烷基酰胺、二烷基酰胺和低级烷基酯改性(例如,其中低级烷基是C1-C4烷基)。翻译后修饰还包括落在肽、多肽或蛋白质的氨基和羧基末端之间的氨基酸的修饰,例如但不限于上述修饰。翻译后修饰可以调节细胞内蛋白质的“生物学”,例如其活性、结构、稳定性或定位。磷酸化是最常见的翻译后修饰,并且在蛋白质的调节中起重要作用,特别是在细胞信号传导中(Prabakaran et al.,2012,Wiley Interdiscip Rev Syst Biol Med 4:565-583)。向蛋白质中添加糖,例如糖基化,已经显示出促进蛋白质折叠、改善稳定性和改变调节功能。脂质与蛋白质的附着使得能够靶向细胞膜。翻译后修饰还可以包括肽、多肽或蛋白质修饰以包括一种或多种可检测标记。

[0423] 在某些实施方案中,肽、多肽或蛋白质可以是片段化的。例如,可以通过从样品(例如生物样品)中分离蛋白质来获得片段化的肽。肽、多肽或蛋白质可以通过本领域已知的任何方法片段化,包括通过蛋白酶或内肽酶的片段化。在一些实施方案中,通过使用特异性蛋白酶或内肽酶靶向肽,多肽或蛋白质的片段化。特定的蛋白酶或内肽酶在特定的共有序列(例如,对ENLYFQ\S共有序列特异的TEV蛋白酶)结合并切割。在其他实施方案中,肽、多肽或蛋白质的片段化是非靶向的或随机的,通过使用非特异性蛋白酶或内肽酶。非特异性蛋白酶可以在特定氨基酸残基而不是共有序列处结合和切割(例如,蛋白酶K是非特异性丝氨酸蛋白酶)。蛋白酶和内肽酶是本领域熟知的,并且可用于将蛋白质或多肽切割成较小肽片段的实例包括蛋白酶K、胰蛋白酶、胰凝乳蛋白酶、胃蛋白酶、嗜热菌蛋白酶、凝血酶、因子Xa、弗林蛋白酶、内肽酶、木瓜蛋白酶、胃蛋白酶、枯草杆菌蛋白酶、弹性蛋白酶、肠激酶、Genenase™I、内蛋白酶LysC、内切蛋白酶AspN、内蛋白酶GluC等。(Granvogl et al.,2007,Anal Bioanal Chem 389:991-1002)。在某些实施方案中,肽、多肽或蛋白质被蛋白酶K或任选的不耐热形式的蛋白酶K片段化,以使得能够快速失活。蛋白酶K在变性试剂(例如尿素和

SDS)中非常稳定,能够消化完全变性的蛋白质。可以在附着DNA标签或DNA记录标签之前或之后进行蛋白质和多肽片段化成肽。

[0424] 化学试剂也可用于将蛋白消化成肽片段。化学试剂可以在特定的氨基酸残基处裂解(例如,溴化氰水解甲硫氨酸残基的C-末端的肽键)。用于将多肽或蛋白质片段化成较小肽的化学试剂包括溴化氰(CNBr)、羟胺、肼、甲酸、BNPS-粪臭素[2-(2-硝基苯基亚磺酰基)-3-甲基吲哚]、碘代苯甲酸、NTCB+Ni(2)-硝基-5-硫氰基苯甲酸)等。

[0425] 在某些实施方案中,在酶促或化学切割后,所得肽片段具有大致相同的所需长度,例如,约10个氨基酸至约70个氨基酸,约10个氨基酸至约60个氨基酸,约10个氨基酸。约50个氨基酸,约10至约40个氨基酸,约10至约30个氨基酸,约20个氨基酸至约70个氨基酸,约20个氨基酸至约60个氨基酸,约20个氨基酸约50个氨基酸,约20至约40个氨基酸,约20至约30个氨基酸,约30个氨基酸至约70个氨基酸,约30个氨基酸至约60个氨基酸,约30个氨基酸约50个氨基酸,或约30个氨基酸至约40个氨基酸。通过用包含含有蛋白酶或内肽酶切割位点的肽序列的短测试FRET(荧光共振能量转移)肽掺加蛋白质或多肽样品,可以优选实时监测切割反应。在完整的FRET肽中,荧光基团和猝灭剂基团连接到含有切割位点的肽序列的任一端,并且猝灭剂和荧光团之间的荧光共振能量转移导致低荧光。在用蛋白酶或内肽酶切割测试肽后,分离猝灭剂和荧光团,使荧光大大增加。当达到某种荧光强度时,可以停止裂解反应,从而实现可重现的解理终点。

[0426] 大分子样品(例如,肽、多肽或蛋白质)可以在附着到固体支持物之前经历蛋白质分级方法,其中蛋白质或肽通过一种或多种性质分离,例如细胞定位、分子量、疏水性或等电位点或蛋白质富集方法。或者/另外,蛋白质富集方法可用于选择特定蛋白质或肽(请参见,例如Whiteaker et al.,2007,Anal.Biochem.362:44-54,通过引用整体并入)或选择特定的翻译后修改(请参考,例如Huang et al.,2014.J.Chromatogr.A 1372:1-17,通过引用整体并入)。或者,可以富集或选择特定类别或类别的蛋白质如免疫球蛋白或免疫球蛋白(Ig)同种型如IgG,用于分析。在免疫球蛋白分子的情况下,涉及亲和结合的高变序列的序列和丰度或频率的分析是特别令人感兴趣,特别是因为它们的变化随着疾病进展而变化或与健康、免疫和/或疾病表型相关。使用标准免疫亲和方法也可以从样品中减去丰度过高蛋白。丰度蛋白的消减可用于血浆样品,其中超过80%的蛋白成分是白蛋白和免疫球蛋白。有几种商业产品可用于过丰蛋白的血浆样品的削减,例如PROTIA和PROT20(Sigma-Aldrich)。

[0427] 在某些实施方案中,所述大分子由蛋白或多肽组成。在一个实施方案中,所述蛋白或多肽通过标准胺偶联化学方法用DNA记录标签标记(参见例如图2B,2C,28,29,31,40)。所述-氨基-氨基(例如赖氨酸残基)和N-末端氨基特别易于用胺反应性偶联剂标记,取决于所述反应的pH(Mendoza and Vachet 2009)。在特定实施方案中(参见,例如,图2B和图29),所述记录标签包含反应性基团(例如,用于缀合至固体表面、多功能接头、或大分子)、接头、通用引物序列、条形码(例如,隔室标签、分区条形码、样品条形码、分级物条形码或其任何组合)、可选的UMI、和间隔(Sp)序列,用于促进向/从编码标签的信息转移。在另一个实施方案中,可以首先用通用DNA标签标记所述蛋白,并且所述条形码-Sp序列(代表样品、区室、载玻片上的物理位置等)随后通过酶促或化学偶联步骤附着于所述蛋白。(参见,例如图20,30,31,40)。通用DNA标签包含用于标记蛋白质或多肽大分子的短核苷酸序列,其并且可用作条形码的附着点(例如,隔室标签、记录标签等)。例如,记录标签可在其末端包含与所述通用

DNA标签互补的序列。在某些实施方案中,通用DNA标签是通用引物序列。在所述标记蛋白上的所述通用DNA标签与记录标签中的互补序列(例如,与珠子结合)杂交后,所述退火的通用DNA标签可以通过引物延伸扩增,将所述记录标签信息转移到所述DNA标记的蛋白。在一个具体实施方案中,在蛋白酶消化所述蛋白质成肽之前,用通用DNA标签标记。然后可以将来自消化物的所述标记肽上的所述通用DNA标签转换成信息丰富和有效的记录标签。

[0428] 在某些实施方案中,蛋白质大分子可以通过亲和捕获试剂(并且任选地共价交联)固定到固体支持物上,其中所述记录标签直接与所述亲和捕获试剂结合,或者,所述蛋白可以直接固定到具有记录标签的所述固体支持物上(参见例如图2C)。

[0429] V.固体支持物

[0430] 本公开的大分子连接到固体支持物的表面(也称为“基质表面”)。所述固体载体可以是任何多孔或无孔载体表面,包括但不限于珠子、微珠、阵列、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶芯片、流通槽、流控芯片、包括信号转导电子器件的生物芯片、微量滴定孔、ELISA板,旋转干涉测量盘、硝酸纤维素膜、基于硝酸纤维素的聚合物表面、纳米颗粒或微球。用于固体载体的材料包括但不限于丙烯酰胺、琼脂糖、纤维素、硝化纤维素、玻璃、金、石英、聚苯乙烯、聚乙烯乙酸乙烯酯、聚丙烯、聚甲基丙烯酸酯、聚乙烯、聚环氧乙烷、聚硅酸盐、聚碳酸酯、特氟隆、碳氟化合物、尼龙、硅橡胶、聚酸酐、聚乙醇酸、聚乳酸、聚原酸酯、官能化硅烷、聚丙基延胡索酸酯、胶原、糖胺聚糖、聚氨基酸、葡聚糖、或其任何组合。固体支持物还包括薄膜、膜、瓶、盘、纤维、编织纤维、成形聚合物,例如管、颗粒、珠子、微球、微粒、或其任何组合。例如,当固体表面是珠子时,所述珠子可以包括但不限于陶瓷珠子、聚苯乙烯珠子、聚合物珠子,甲基苯乙烯珠子、琼脂糖珠子、丙烯酰胺珠子、实心珠子、多孔珠、顺磁珠,玻璃珠或可控孔珠。

[0431] 在某些实施方案中,固体支持物是流通槽。流通槽配置可以在不同的下一代测序平台之间变化。例如,所述Illumina流通槽是类似于显微镜载玻片的平面光学透明表面,其包含在其表面结合的寡核苷酸锚草坪。模板DNA包含与末端连接的衔接子,所述衔接子与流通槽表面上的寡核苷酸互补。适配的单链DNA与所述流通槽结合,并在测序前通过“桥”式样固相PCR扩增。所述454流通池(54Life Sciences)支持“picotiter”板,光纤载玻片具有~160万75Picotiter孔。剪切的模板DNA的每单个分子被捕获到分离的珠子上,并且每个珠子在油乳液内的水性PCR反应混合物的专用液滴中区室化。通过PCR在所述珠子表面上克隆扩增模板,然后将载有所述模板的珠子分配到Picotiter滴定板的孔中用于测序反应,理想的是每孔含有一个或更少的珠子。来自Applied Biosystems的SOLiD(支持寡核苷酸连接和检测)仪器,如454系统,通过乳液PCR扩增模板分子。在剔除不含扩增模板的珠子的步骤之后,将结合在珠子的存放在所述流通槽。流通槽也可以是简单的过滤器玻璃料,例如TWIST™ DNA合成柱(Glen Research)。

[0432] 在某些实施方案中,固体支持物是珠子,其可以指单个珠子或多个珠子。在一些实施方案中,所述珠子与选来用于下游分析(例如,SOLiD或454)的下一代测序平台相互兼容。在一些实施方案中,固体支持物是琼脂糖珠子、顺磁珠子、聚苯乙烯珠子、聚合物珠子、丙烯酰胺珠子、实心珠子、多孔珠子、顺磁珠子、玻璃珠子或可控孔珠子。在进一步的实施方案中,珠子可以包被结合官能团(例如,胺基团、用于结合生物素标记的大分子的链霉亲和素、抗体)以促进与大分子的结合。

[0433] 蛋白、多肽或肽可以通过本领域已知的任何方式,包括共价和非共价相互作用,或其任何组合(见,例如Chan et al.,2007,PLoS One 2:e1164;Cazalis et al.,Bioconj.Chem.15:1005-1009;Soellner et al.,2003,J.Am.Chem.Soc.125:11790-11791;Sun et al.,2006,Bioconjug.Chem.17-52-57;Decreau et al.,2007,J.Org.Chem.72:2794-2802;Camarero et al.,2004,J.Am.Chem.Soc.126:14730-14731;Girish et al.,2005,Bioorg.Med.Chem.Lett.15:2447-2451;Kalia et al.,2007,Bioconjug.Chem.18:1064-1069;Watzke et al.,2006,Angew Chem.Int.Ed.Engl.45:1408-1412;Parthasarathy et al.,2007,Bioconjugate Chem.18:469-476;和Bioconjugate Techniques,G.T.Hermanson,Academic Press(2013),并且每个都通过引用整体并入本文)直接或间接地连接到所述固体支持物。例如,所述肽可以通过连接反应与所述固体支持物连接。或者,固体支持物可包括试剂或涂层以促进所述肽直接或间接连接到所述固体支持物上。为此目的可以使用任何合适的分子或材料,包括蛋白质、核酸、碳水化合物和小分子。例如,在一个实施方案中,所述试剂是亲和分子。在另一个实例中,所述试剂是叠氮基团,可以与另一分子中的炔基反应以促进所述固体载体和其它分子之间的联合或结合。

[0434] 蛋白质,多肽或肽可以通过称为“点击化学”的方法连接到固体支持物上。为此目的,任何快速且基本上不可逆的反应可用于将蛋白、多肽或肽连接到所述固体支持物上。示例性反应包括叠氮化物和炔烃的铜催化反应以形成三唑(Huisgen 1,3-偶极环加成),应变促进的叠氮化物炔烃环加成(SPAAC),二烯和亲二烯体(Diels-Alder)的反应,应变-促进的炔烃-硝酮环加成反应,应变烯烃与叠氮化物、四嗪或四唑的反应,烯烃和叠氮化物[3+2]环加成反应,烯烃和四嗪反转电子要求的Diels-Alder(IEDDA)反应(如m-四嗪(mTet)和反式环辛烯(TCO))、烯烃和四唑光反应,叠氮化物和膦的Staudinger连接,以及各种置换反应:例如通过对亲电子原子的亲核攻击的离去基团置换(Horisawa 2014,Knall,Hollauf et al.2014)。示例性置换反应包括胺与:活化酯;N-羟基琥珀酰亚胺酯;异氰酸酯;异硫氰酸酯等的反应。

[0435] 在一些实施方案中,所述大分子和固体支持物通过能够通过两个互补反应基团的反应形成的官能团连接,例如一种前述“点击”反应的产物、官能团。在各种实施方案中,官能团可以通过醛、肟、腙、酰肼、炔烃、胺、叠氮化物、酰基叠氮化物、酰基卤化物、腈、硝酮、巯基、二硫化物、磺酰卤、异硫氰酸酯、酰亚胺酯、活化酯(例如N-羟基琥珀酰亚胺酯、戊炔酸STP酯)、酮、-不饱和羰基、烯烃、马来酰亚胺、-卤代酰亚胺、环氧化物、氮丙啶、四嗪、四唑、膦、生物素或环硫官能团与互补反应性基团的反应形成。示例性反应是胺(例如伯胺)与N-羟基琥珀酰亚胺酯或异硫氰酸酯的反应。

[0436] 在其他实施方案中,所述官能团包含烯烃、酯、酰胺、硫酯、二硫化物、碳环、杂环或杂芳基。在进一步的实施方案中,所述官能团包含烯烃、酯、酰胺、硫酯、二硫化物、碳环、杂环或杂芳基。在其它实施方案中,所述官能团包含酰胺或硫脲。在一些更具体的实施方案中,官能团是三唑基官能团、酰胺、或硫脲官能团。

[0437] 在一个优选实施方案中,iEDDA点击化学用于将大分子(例如,蛋白、多肽、肽)固定到固体支持物上,因为它是快速的并且在低输入浓度下高产率的。在另一个优选的实施方案中,在iEDDA点击化学反应用m-四嗪而不是四嗪,因为m-四嗪具有更好的键稳定性。

[0438] 在一个优选的实施方案中,用TCO官能化基质表面,并通过连接的间-四嗪部分将

记录标签标记的蛋白质,多肽,肽固定在TCO涂覆的基质表面上(图34)。

[0439] 蛋白、多肽或肽可通过其C-末端,N-末端或内部氨基酸例如,通过胺,羧基或巯基固定到固体支持物的表面。用于偶联胺基的标准活化载体包括CNBr活化的、NHS活化的、醛活化的、吖内酯活化的、和CDI活化的载体。用于羧基偶联的标准活化载体包括与胺载体偶联的碳二亚胺活化的羧基基团。半胱氨酸偶联可以使用马来酰亚胺、异乙酰基和吡啶基二硫化物活化的载体。肽羧基末端固定的另一种方式使用脱水胰蛋白酶,胰蛋白酶的催化惰性衍生物,其在C-末端结合含有赖氨酸或精氨酸残基的肽而不切割它们。

[0440] 在某些实施方案中,蛋白质,多肽或肽通过固体表面连接的接头与所述蛋白、多肽或肽的赖氨酸基团的共价连接固定到固体支持物上。

[0441] 记录标签可以在固定到所述固体支持物之前或之后附着到蛋白、多肽或肽上。例如,蛋白、多肽或肽可首先用记录标签标记,然后通过包含两个用于偶联的功能基团的记录标签固定在固体表面上(参见图28)。所述记录标签的一个功能基团与蛋白质偶联,而另一个功能基团将所述记录标签标记的蛋白固定在固体支持物上。

[0442] 或者,在用记录标签标记所述蛋白、多肽或肽之前,将蛋白、多肽或肽固定到固体支持物上。例如,蛋白可以首先用反应基团例如点击化学基团派生化。然后所述活化的蛋白分子可以连接到合适的固体支持物上,然后通过采用互补点击化学基团标记上记录标签。作为例子,用炔和mTet基团衍生的蛋白可以固定在用叠氮化物和TCO衍生的珠子上,并附着在用叠氮化物和TCO标记的记录标签上。

[0443] 应当理解,本文提供的用于将大分子(例如,蛋白、多肽、或肽)连接至所述固体支持物的方法也可用于将记录标签附着至所述固体支持物或将记录标签附着至大分子(例如,蛋白、多肽、或肽)。

[0444] 在某些实施方案中,所述固体支持物的表面被钝化(封闭)以最小化对结合剂的非特异性吸收。“钝化的”表面是指已经用外层材料处理以最小化结合剂的非特异性结合的表面。钝化表面的方法包括来自荧光单分子分析文献的标准方法,包括用聚合物如聚乙二醇(PEG)钝化表面(Pan et al.,2015,Phys.Biol.12:045006)、聚硅氧烷(例如,Pluronic F-127)、星形聚合物(例如,星形PEG)(Groll et al.,2010,Methods Enzymol.472:1-18),疏水性二氯二甲基硅烷(DDS)+自组装吐温-20(Hua et al.,2014,Nat.Methods11:1233-1236)和类金刚石碳(DLC),DLC+PEG(Stavis et al.,2011,Proc.Natl.Acad.Sci.USA 108:983-988)。除了共价表面修饰之外,还可以使用许多钝化剂,包括表面活性剂如Tween-20、溶液状聚硅氧烷(Pluronic系列)、聚乙烯醇(PVA)、和BSA和酪蛋白等蛋白质。或者,当将所述蛋白、多肽、或肽固定到所述固体基质上时,可以通过掺入竞争分子或“虚拟”反应性分子,在所述固体基质的表面上或体积内滴定蛋白、多肽或肽的密度(见图36A)。

[0445] 在多个大分子固定在相同固体支持物上的某些实施方案中,所述大分子可以适当间隔以减少或防止交叉结合或分子间事件的发生,例如,当结合剂结合第一大分子,且其编码标签信息被传送到与相邻大分子相连的记录标签,而不是与第一大分子相连的记录标签。为了控制固体支持物上的大分子(例如,蛋白质、多肽或肽间距)间距,可以在所述基质表面上滴定功能性偶联基团(例如TCO)的密度(参见图34)。在一些实施方案中,多个大分子在所述固体载体的表面上或体积(例如,多孔载体)内间隔开约50nm至约500nm、或约50nm至约400nm,或约50nm至约300nm,或约50nm至约200nm,或约50nm至约100nm的距离。在一些实

施方案中,多个大分子在固体支持物的表面上间隔开,平均距离为至少50nm,至少60nm,至少70nm,至少80nm,至少90nm,至少100nm,至少150nm,至少200nm,至少250nm,至少300nm,至少350nm,至少400nm,至少450nm,或至少500nm。在一些实施方案中,多个大分子在固体支持物的所述表面上间隔开,平均距离为至少50nm。在一些实施方案中,大分子在所述固体支持物的表面上或体积内间隔开,根据经验,使得分子间与分子内事件的相对频率<1:10;<1:100;<1:1000;或<1:10,000。合适的间隔频率可以使用功能测定凭经验确定(参见实施例23),并且可以通过稀释和/或通过掺杂"虚拟"间隔分子来完成,所述间隔分子竞争所述基质表面上的附着位点。

[0446] 例如,如图34所示,PEG-5000(MW~5000)用于阻断所述基质表面(例如珠子表面)上的肽之间的间隙空间。此外,所述肽与功能基团偶联,所述功能基团也与PEG-5000分子连接。在一个优选的实施方案中,这通过将NHS-PEG-5000-TCO+NHS-PEG-5000-甲基的混合物偶联到胺衍生的珠子上来实现(参见图34)。滴定两种PEG(TCO比甲基)之间的化学计量比,以在所述基质表面上产生适当密度的官能偶联基团(TCO基团);所述甲基-PEG对偶联是惰性的。可以通过测量表面上TCO基团的密度来计算TCO基团之间的有效间距。在某些实施方案中,所述固体表面上的偶联基团(例如,TCO)之间的间距为至少50nm,至少100nm,至少250nm或至少500nm。在所述珠子的PEG 5000-TO/甲基衍生化之后,表面上过量的NH_2基团用反应性酸酐(例如乙酸或琥珀酸酐)猝灭。

[0447] VI.记录标签

[0448] 至少一个记录标签直接或间接结合或共定位到所述大分子并连接到所述固体支持物上(见,例如图5)。记录标签可包含DNA、RNA、PNA、γPNA、GNA、BNA、XNA、TNA、多核苷酸类似物、或其组合。记录标签可以是单链的、或部分或完全双链的。记录标签可以具有平末端或粘性末端。在某些实施方案中,在结合剂与大分子结合后,将所述结合剂的编码标签的识别信息转移至所述记录标签以产生延伸记录标签。所述延伸记录标签的进一步延伸可以在随后的结合循环中进行。

[0449] 记录标签可以直接或间接(例如,通过接头)通过本领域已知的任何方式,包括共价和非共价相互作用、或其任何组合,连接到所述固体支持物连接。例如,所述记录标签可以通过连接反应与所述固体支持物连接。或者,所述固体支持物可包括试剂或涂层以便于直接或间接地将所述记录标签连接到所述固体支持物上。将核酸分子固定在固体支持物(例如珠子)上的策略已在美国专利5,900,481中描述。Steinberg et al.(2004,Biopolymers 73:597-605);Lund et al.,1988(Nucleic Acids Res.16:10861-10880);和Steinberg et al.(2004,Biopolymers 73:597-605),其各自通过引用整体并入本文。

[0450] 在某些实施方案中,大分子(例如肽)和相关记录标签的所述共定位是通过将大分子和记录标签缀合至直接连接在所述固体支持物表面的双功能接头来实现的Steinberg et al.(2004,Biopolymers 73:597-605)。在进一步的实施方案中,三官能基团用于衍生所述固体支持物(例如珠子),并且所得的双官能基团与大分子和记录标签二者偶联。

[0451] 方法和试剂(例如点击化学试剂和光亲和标记试剂)例如描述的用于附着大分子和固体支持物的那些,也可用于记录标签的附着。

[0452] 在一个具体实施方案中,单个记录标签附着到大分子(例如肽)上,优选通过解封的N-或C-末端氨基酸的附着。在另一个实施方案中,多个记录标签附着于所述大分子(例

如,蛋白、多肽或肽),优选附着于赖氨酸残基或肽骨架。在一些实施方案中,用多个记录标签标记的大分子(例如,蛋白或多肽)被片段化或消化成较小的肽,每个肽平均标记有一个记录标签。

[0453] 在某些实施方案中,记录标签包含任选的独特分子标识符(UMI),其与所述UMI相关的每个大分子(例如,蛋白质、多肽、肽)提供独特的标识符标签。UMI可为约3至约40个碱基,约3至约30个碱基,约3至约20个碱基,或约3至约10个碱基,或约3至约8个碱基。在一些实施方案中,UMI为约3个碱基,4个碱基,5个碱基,6个碱基,7个碱基,8个碱基,9个碱基,10个碱基,11个碱基,12个碱基,13个碱基,14个碱基,15个碱基,16个碱基,17个碱基,18个碱基,19个碱基,20个碱基,25个碱基,30个碱基,35个碱基或40个碱基的长度。UMI可用于从多个延伸记录标签解卷积测序数据以鉴定来自各个大分子的序列读数。在一些实施方案中,在大分子库内,每个大分子与单个记录标签相关联,每个记录标签包含独特的UMI。在其他实施例中,记录标签的多个副本与单个大分子相关联,所述记录标签的每个副本包含相同的UMI。在一些实施方案中,UMI具有与所述结合剂编码标签内的间隔区或编码序列不同的碱基序列,以便于在序列分析期间区分这些组分。

[0454] 在某些实施例中,记录标签包括条形码,例如而不是UMI(如果存在的话)。条形码是约3至约30个碱基、约3至约25个碱基、约3至约20个碱基、约3至约10个碱基、约3至约10个碱基、约3至约8个碱基长度的核酸分子。在一些实施方案中,条形码是约3个碱基、4个碱基、5个碱基、6个碱基、7个碱基、8个碱基、9个碱基、10个碱基、11个碱基、12个碱基、13个碱基、14个碱基、15个碱基、20个碱基、25个碱基、或30个碱基长度。在一个实施方案中,条形码允许多个样品或文库的多路复用测序。条形码可用于派生大分子(例如肽)的分区、分级物、区室、样品、空间位置或文库。条形码可用于解卷积多路复用序列数据并鉴定来自单个样品或文库的序列读数。例如,条形码珠子可用于涉及样品乳化和分区的方法,例如用于分区所述蛋白质组的目的。

[0455] 条形码可以代表隔室标签,其中隔室,例如液滴、微孔、固体支撑物上的物理区域等,被分配唯一的条形码。隔室与特定条形码的关联可以以任何方式实现,例如通过将单个条形码珠子封装在隔室中,例如通过将条形码液滴直接合并或加入隔室,通过直接印刷或注入条形码试剂到隔室等。隔室内的所述条形码试剂用于将隔室特异性条形码加到所述隔室内的所述大分子或其片段上。应用于蛋白质分区到隔室中,所述条形码可用于将分析的肽映射回所述隔室中的其原始蛋白质分子。这能够极大地方便蛋白质鉴定。隔室条形码也可用于鉴定蛋白质复合物。

[0456] 在其他实施例中,可以为表示隔室群的子集的多个隔室分配代表该子集的唯一条形码。

[0457] 或者,条形码可以是样品识别条形码。样品条形码可用于单个反应容器中的一组样品的多路复用分析或固定到单个固体基质或固体基质的集合(例如,平面载玻片、包含在单个管或容器中的珠子群等)。来自许多不同样品的大分子可以用具有样品特异性条形码的记录标签进行标记,然后在固定到固体支持物、循环结合、和记录标签分析之前将所有所样品汇集在一起。或者,所述样品可以保持分隔开直到产生DNA编码文库后,并且在所述DNA编码文库的PCR扩增过程中附着样品条形码,然后在测序之前混合在一起。当测定不同丰度类别的分析物(例如蛋白质)时,该方法可能是有用的。例如,所述样品可以分开和条形码

化,并且一部分使用降低丰度分析物的结合剂处理,另一部分用较高丰度的分析物的结合剂处理。在一个具体实施方案中,该方法有助于将特定蛋白质分析物测定的动态范围调节到所述蛋白质分析物的标准表达水平的“最佳点”内。

[0458] 在某些实施方案中,来自多个不同样品的肽、多肽或蛋白用具有样品特异性条形码的记录标签标记。所述复合样品条形化的肽、多肽或蛋白可以在循环结合反应之前混合在一起。通过这种方式,有效地创建了数字反相蛋白质阵列(RPPA)的高度复用替代品(Guo,Liu et al.2012,Assadi,Lamerz et al.2013,Akbani,Becker et al.2014,Creighton和Huang 2015)。数字RPPA样分析的创建在翻译研究、生物标记验证、药物发现、临床、和精准医学中有许多应用。

[0459] 在某些实施方案中,记录标签包含通用引发位点,例如,正向或5’通用引发位点。通用引发位点是可用于引发文库扩增反应和/或用于测序的核酸序列。通用引发位点可包括,但不限于用于PCR扩增的引发位点、与流通槽表面上的互补寡核苷酸退火的流通槽衔接子序列(例如,Illumina下一代测序)、测序引发位点或它们的组合。通用引发位点可以是约10个碱基至约60个碱基。在一些实施方案中,通用引发位点包含Illumina P5引物(5’-AATGATACGGCGACCACCGA-3’-SEQ ID NO:133)或Illumina P7引物(5’-CAAGCAGAAGACGGCATACGAGAT-3’-SEQ ID NO:134)。

[0460] 在某些实施方案中,记录标签在其末端例如3’末端包含间隔子。如本文所用,在记录标签的背景下间隔序列包括与其同源结合剂相关的间隔序列相同的间隔序列,或包括与其同源结合剂相关的间隔序列互补的间隔序列。所述记录标签上的末端例如3’间隔子允许同源结合剂的识别信息在第一个结合循环期间从其编码标签转移到所述记录标签(例如,通过互补间隔序列的退火进行引物延伸或粘性末端连接)。

[0461] 在一个实施方案中,所述间隔序列的长度为约1-20个碱基、长度为约2-12个碱基、或长度为5-10个碱基。所述间隔子的长度可取决于诸如用于将编码标签信息转移到所述记录标签的所述引物延伸反应的温度和反应条件等因素。

[0462] 在优选实施例中,所述记录中的所述间隔序列被设计成与所述记录标签中的其它区域具有最小的互补性;同样,所述编码标签中的所述间隔序列应该与所述编码标签中的其它区域具有最小的互补性。换句话说,所述记录标签和编码标签的所述间隔序列应该与所述记录标签或编码标签中成分例如独特分子标识符、条形码(例如,区室、分区、样品、空间位置)、通用引物序列、编码序列、循环特异性序列等等存在于具有最小的序列互补性。

[0463] 如针对所述结合剂间隔子的描述,在一些实施方案中,与大分子文库相关联的记录标签共享共同的间隔序列。在其他实施方案中,与大分子文库相关的所述记录标签具有与其同源结合剂的结合循环特异性间隔序列互补的结合循环特异性间隔序列,其在使用非级联延伸记录标签时是有用的(参见图10)。

[0464] 延伸记录标签的集合可以在事后级联(参见例如图10)。在所述结合循环完成后,所述珠子固体支持物,每个平均包含一个或少于一个大分子的珠子,每个具有延伸记录标签集合的大分子(延伸记录标签集合共定位在所述大分子位点)被放置在乳液中。形成所述乳液使得每个液滴平均至多被1个珠子占据。在乳液中进行可选的组装PCR反应以扩增与所述大分子共定位在所述珠子上的延伸记录标签,并通过在分隔的延伸记录标签上的不同循环特异性序列之间引发来以共线顺序组装它们(Xiong,Peng et al.2008)。然后破坏所述

乳液并对组装的延伸记录标签进行测序。

[0465] 在另一个实施方案中,所述DNA记录标签包含通用引发序列(U1)、一个或多个条形码序列(BC)、和特异于第一结合循环的间隔序列(Sp1)。在所述第一结合循环中,结合剂使用包含Sp1互补间隔区、编码器条形码、和可选的循环条形码、以及第二间隔元件(Sp2)的DNA编码标签。使用至少两种不同间隔元件的实用性是所述第一结合循环选择潜在的几种DNA记录标签中的一种,并且单个DNA记录标签被延伸从而在所述延伸的DNA记录标签的末端产生新的Sp2间隔元件。在第二次和随后的结合循环中,结合剂仅含有Sp2'间隔子而不是Sp1'。以这种方式,在后续循环中仅来自第一循环的所述单个延伸记录标签被延伸。在另一个实施方案中,第二和随后的循环可以使用结合剂特异性间隔子。

[0466] 在一些实施方案中,记录标签从5'至3'方向包含:通用正向(或5')引发序列、UMI、和间隔序列。在一些实施例中,记录标签从5'到3'方向包括:通用正向(或5')引发序列、可选UMI、条形码(例如,样本条形码、分区条形码、隔室条形码、空间条形码或其任何组合)、和间隔序列。在一些其他实施例中,记录标签从5'到3'方向包括:通用正向(或5')引发序列、条形码(例如,样本条形码、分区条形码、隔室条形码、空间条形码或其任何组合)、可选的UMI、和间隔序列。

[0467] 可采用组合方法从修饰的DNA和PNA产生UMI。在一个示例中,UMI可以通过"化学连接"一组短词序列(4-15mer)构建,短词序列被设计为彼此正交(Spiropulos和Heemstra 2012)。DNA模板用于引导所述"词"聚合物的化学连接。所述DNA模板由杂交臂构成,其能够简单地通过在溶液中将子组分混合在一起来组装组合模板结构(参见图12C)。在某些实施方案中,在该设计中没有"间隔物"序列。所述词空间的大小可以从10个词到10,000个词或更多词。在某些实施方案中,所选的这些词使得它们彼此不同以不交叉杂交,但具有相对均一的杂交条件。在一个实施例中,词的长度将在10个碱基的量级上,在子集中具有大约1000个词(这仅是总10-mer~4^{10}词空间的约0.1%=1百万个词)。这些词的集合(子集中含1000)可以级联在一起以生成复杂度=1000^n幂的最终组合UMI。对于连接在一起的4个词,其创造10^{12}个不同元素的UMI多样性。这些UMI序列将在单分子水平上附加到所述大分子(肽、蛋白等)上。在一个实施方案中,UMI的所述多样性超过UMI所附着的大分子的分子数量。通过这种方式,所述UMI可以唯一地识别感兴趣的大分子。组合词UMI的使用有助于在高错误率测序器上(例如纳米孔测序器、纳米间隙通道测序,等)读数,因为不要求单碱基分辨率来读取多碱基长度的词。组合词方法还可用于生成记录标签或编码标签的其它身份信息组件,比如隔室标签、分区条形码、空间条形码、样本条形码、编码器序列、循环特异性序列、和条形码。纳米孔测序和具有容错词(代码)的DNA编码信息有关的方法是本领域已知的(参阅,例如Kiah et al.,2015,Codes for DNA sequence profiles.IEEE International Symposium on Information Theory(ISIT);Gabrys et al.,2015,Asymmetric Lee distance codes for DNA-based storage.IEEE Symposium on Information Theory (ISIT);Laure et al.,2016,Coding in 2D:Using Intentional Dispersity to Enhance the Information Capacity of Sequence-Coded Polymer Barcodes.Angew.Chem.Int.Ed.doi:10.1002/anie.201605279;Yazdi et al.,2015,IEEE Transactions on Molecular,Biological and Multi-Scale Communications 1:230-248;以及Yazdi et al.,2015,Sci Rep 5:14138,其每一篇通过引用整体并入本文)。因此,

在某些实施方案中,本文描述的任何实施方案中的延伸记录标签、延伸编码标签或di-tag构建体包含作为错误纠正码的识别组件(例如,UMI、编码序列、条形码、隔室标签、循环特异性序列,等)。在一些实施例中,所述错误纠正码选自:汉明码、李距离码、非对称李距离码、里德-所罗门码、和Levenshtein-Tenengolts码。对于纳米孔测序,所述电流或离子分布和不对称碱基调用错误是所使用的纳米孔和生物化学类型固有的,并且该信息可用于使用上述错误校正方法设计更稳健的DNA代码。作为采用稳健DNA纳米孔测序条形码的替代方案,可以直接使用条形码序列的电流或离子分布特征(美国专利号7,060,507,其通过引用整体并入),完全避免DNA碱基调用,并如Laszlo等人所述通过映射回到预测的电流/分布特征立即识别所述条形码序列(2014,Nat.Biotechnol.32:829-833,通过引用整体并入)。在这篇文章中,Laszlo等人描述生物纳米孔,MspA,在传递不同字串通过所述纳米孔时产生的电流特征,以及通过将产生的电流特征映射回计算机以预测来自通用序列的可能电流特征来映射和识别DNA链的能力(2014,Nat.Biotechnol.32:829-833)。类似的概念可以应用于DNA代码和通过基于纳米间隙通道电流DNA测序产生的电信号(Ohshiro et al.,2012,Sci Rep 2:501)。

[0468] 因此,在某些实施方案中,编码标签、记录标签或两者的识别组件能够产生独特的电流或离子分布或光学特征,其中本文提供的任何方法的分析步骤包含检测所述独特电流或离子分布或光学特征以识别所述识别组件。在一些实施方案中,所述识别组件选自编码器序列、条形码,UMI、隔室标签、循环特异性序列或其任何组合。

[0469] 在某些实施方案中,样品中的全部或基本量的大分子(例如,蛋白、多肽或肽)(例如,至少50%,55%,60%,65%,70%,75%,80%,85%,90%,95%,96%,97%,98%,99%或100%)被记录标签标记。所述大分子的标记可以在将所述大分子固定到固体支持物之前或之后进行。

[0470] 在其它实施方案中,样品内的大分子(例如,蛋白、多肽或肽)的子集用记录标签标记。在一个具体实施方案中,来自样品的大分子子集用记录标签进行靶向(分析物特异性)标记。蛋白质的靶向记录标签可以使用靶蛋白特异性结合剂(例如,抗体,适配体等)实现,所述靶蛋白特异性结合剂连接短靶特异性DNA捕获探针,例如,分析物特异性条形码,所述捕获探针退火到记录标签中的互补靶特异性诱饵序列,例如分析物特异性条形码(参见图28A)。所述记录标签包含用于所述靶蛋白上存在的同源反应性基团的反应性基团(例如,点击化学标记、光亲和标记)。例如,记录标签可包含用于与炔衍生蛋白相互作用的叠氮基团,或记录标签可包含用于与天然蛋白质相互作用的二苯甲酮,等(参见图28A-B)。在通过所述靶蛋白特异性结合剂结合所述靶蛋白后,所述记录标签和靶蛋白通过其相应的反应性基团偶联(参见图28B-C)。在用记录标签标记所述靶蛋白后,所述靶蛋白特异性结合剂可以通过消化连接在所述靶蛋白特异性结合剂的所述DNA捕获探针去除。例如,所述DNA捕获探针可以设计成含有尿嘧啶碱基,然后将其用尿嘧啶特异性切除试剂(例如USER™)进行靶向消化,且所述靶蛋白特异性结合剂可以从所述靶蛋白中解离。

[0471] 在一个实例中,可以用DNA捕获探针(例如,图28中的分析物条形码BC_A)标记对靶蛋白集特异的抗体,所述DNA捕获探针与设计有互补诱饵序列(例如,图28中的分析物条形码$BC_{A'}$)的记录标签杂交。蛋白的样品特异性标记可以通过使用DNA捕获探针标记的抗体与包含样品特异性条形码的记录标签上的互补诱饵序列杂交来实现。

[0472] 在另一个实例中,靶蛋白特异性适配体用于样品内蛋白子集的靶向记录标签标记。靶特异性适配体与DNA捕获探针连接,所述DNA捕获探针与记录标签中的互补诱饵序列退火。所述记录标签包含反应性化学探针或光反应性化学探针(例如二苯甲酮(BP)),用于与具有相应反应性基团的所述靶蛋白偶联。所述适配体与其靶蛋白分子结合,将所述记录标签带靠近所述靶蛋白,使得所述记录标签偶联到所述靶蛋白。

[0473] 使用附着于小分子蛋白质亲和配体的光反应性化学探针的光亲和(PA)蛋白的标记先前已经描述(Park,Koh et al.2016)。典型的光反应性化学探针包括基于二苯甲酮(benzophenone)(反应性双自由基,365nm)、苯基二氮丙啶(phenyldiazirine)(反应性碳,365nm)、和苯基叠氮化物(反应性氮烯自由基,260nm)的探针,如前所述在照射波长下活化(Smith和Collins 2015)。在一个优选的实施方案中,使用Li等人公开的方法用包含样品条形码的记录标签标记蛋白样品中的靶蛋白质,其中二苯甲酮标记的记录标签中的诱饵序列与附着于同源结合剂(例如,核酸适配体(参见图28)的DNA捕获探针杂交(Li,Liu et al,2013)。对于光亲和标记的蛋白靶标,使用DNA/RNA适配体作为靶蛋白特异性结合剂优于抗体,因为所述光亲和基团可以自我标记所述抗体而不是所述靶蛋白。相比之下,光亲和标记核酸的效率低于标记蛋白质的效率,这使得适配体成为DNA引导的化学或光标记的更好的载体。类似于光亲和标记,也可以采用与Rosen等人描述的方式(Rosen,Kodal et al.2014,Kodal,Rosen et al.2016)类似的方式,在所述适配体结合位点附近使用DNA引导的化学标记反应性赖氨酸(或其它基团)。

[0474] 在前述实施方案中,除杂交之外,其他类型的连接可用于连接所述靶特异性结合剂和所述记录标签(参见图28A)。例如,如图28B所示,一旦所述捕获的靶蛋白(或其它大分子)共价连接到所述记录标签,就能够使用设计为切割并释放所述结合剂的接头共价连接所述两个基团。合适的接头可以连接到所述记录标签的多个位置,例如3'端,或连接到附着于所述记录标签的5'端的接头中。

[0475] VII.结合剂和编码标签

[0476] 本文描述的方法使用能够结合所述大分子的结合剂。结合剂可以是能够结合大分子的组分或特征的任何分子(例如,肽、多肽、蛋白质、核酸、碳水化合物、小分子、以及类似物)。结合剂可以是天然存在的、合成产生的、或重组表达的分子。结合剂可结合大分子的单个单体或亚基(例如,肽的单个氨基酸)或结合大分子的多个连接的亚基(例如,较长肽分子的二肽、三肽或更高次序肽)。

[0477] 在某些实施方案中,结合剂可以设计来共价结合。共价结合可以设计为与正确基团结合时是有条件的或偏爱的。例如,NTAA及其同源NTAA特异性结合剂可各自用反应基团修饰,使得一旦所述NTAA特异性结合剂与所述同源NTAA结合,就进行偶联反应以在其二者之间产生共价连接。所述结合剂与缺乏同源反应基团的其他位置的非特异性结合不会导致共价连接。结合剂与其靶标之间的共价结合允许更严格的洗涤用于去除非特异性结合的结合剂,从而增加所述测定的特异性。

[0478] 在某些实施方案中,结合剂可以是选择性结合剂。如本文所用,选择性结合是指所述结合剂相对于结合到不同配体(例如,氨基酸或氨基酸类)优先结合特定配体(例如,氨基酸或氨基酸类)的能力。选择性通常指一种配体被具有结合剂的复合物中的另一配体置换的反应的平衡常数。通常,这种选择性与所述配体的空间几何形状和/或所述配体与结合剂

结合的方式和程度有关,例如通过氢键或范德华力(非共价互作)或通过可逆或不可逆的与结合剂的共价连接。还应该理解,选择性可以是相对的,并且与绝对相反,并且其可以被包括配体浓度在内的不同的因素影响。因此,在一个实例中,结合剂选择性地结合所述二十种标准氨基酸中的一种。在非选择性结合的实例中,结合剂可以结合所述二十种标准氨基酸中的两种或更多种。

[0479] 在本文公开的方法的实践中,结合剂选择性结合大分子的特征或组分的能力仅需足以让它的编码标签的信息到与大分子相关的所述记录标签的转移,所述记录标签信息到编码标签的转移,或编码标签信息和记录标签信息到di-tag分子的转移。因此,选择性仅需要相对于大分子所暴露的其它结合剂。还应该理解,结合剂的选择性不必绝对地对特定氨基酸,而是可以对一类氨基酸有选择性,例如具有非极性或非极性侧链的氨基酸、或具有带电荷(正或负)侧链的氨基酸,或具有芳香族侧链的,或一些特定类别或大小的侧链的氨基酸,和类似的氨基酸。

[0480] 在一个具体实施方案中,所述结合剂对目标大分子具有高亲和力和高选择性。特别地,具有低解离速率的高结合亲和力对于编码标签和记录标签之间的信息传递是有效的。在某些实施方案中,结合剂具有<10nM、<5nM、<1nM、<0.5nM、或<0.1nM的Kd。在一个具体实施方案中,将所述结合剂以>10X、>100X或>1000X的浓度加入到所述大分子中,以使其结合至完成。Chang等人描述了抗体与单个蛋白质分子的结合动力学的详细讨论(Chang, Rissin et al.2012)。

[0481] 为了增加结合剂对肽的小N-末端氨基酸(NTAA)的亲和力,可以用“免疫原性”半抗原,例如二硝基苯酚(DNP)修饰所述NTAA。可以使用Sanger试剂,二硝基氟苯(DNFB)以循环测序方法实施,该试剂将DNP基团连接到所述NTAA的胺基上。商购的抗DNP抗体在低nM范围内(~8nM,LO-DNP-2)具有亲和力(Bilgicer,Thomas等人2009);因此,有理由认为可以将高亲和力NTAA结合剂制成许多用DNP修饰(通过DNFB)的NTAA,同时对特定NTAA实现良好的结合选择性。在另一个实例中,可以通过4-磺酰基-2-硝基氟苯(SNFB)用磺酰基硝基苯酚(SNP)修饰NTAA。用替代的NTAA改性剂例如乙酰基或脒基(胍基)也可以实现类似的亲和力增强。

[0482] 在某些实施方案中,结合剂可以结合NTAA、CTAA、干预氨基酸、二肽(两个氨基酸的序列)、三肽(三个氨基酸的序列)、或肽分子的更高次序肽。在一些实施方案中,结合剂文库中的每种结合剂选择性结合特定氨基酸,例如二十种标准天然存在的氨基酸之一。标准的天然氨基酸包括丙氨酸(A或Ala)、半胱氨酸(C或Cys)、天冬氨酸(D或Asp)、谷氨酸(E或Glu)、苯丙氨酸(F或Phe)、甘氨酸(G或Gly)、组氨酸(H或His)、异亮氨酸(I或Ile)、赖氨酸(K或Lys)、亮氨酸(L或Leu)、蛋氨酸(M或Met)、天冬酰胺(N或Asn)、脯氨酸(P或Pro)、谷氨酰胺(Q或Gln)、精氨酸(R或Arg)、丝氨酸(S或Ser)、苏氨酸(T或Thr)、缬氨酸(V或Val)、色氨酸(W或Trp)、和酪氨酸(Y或Tyr)。

[0483] 在某些实施方案中,结合剂可以结合氨基酸的翻译后修饰。在一些实施方案中,肽包含一个或多个相同的或不同的翻译后修饰。所述NTAA、CTAA、干预氨基酸、或其组合可以是翻译后修饰的。氨基酸的翻译后修饰包括酰化、乙酰化、烷基化(包括甲基化)、生物素化、丁酰化、氨基甲酰化、羰基化、脱酰胺、脱酰胺、二萘胺形成,二硫键形成、消除、黄素附着、甲酰化、γ-羧化、谷氨酰化、甘氨酰化、糖基化、glypiation、血红素C附着、羟基化、羟丁赖氨酸

(hypusine)形成、碘化、异戊二烯化、脂化、脂酰化、丙二酰化、甲基化、肉豆蔻酰化、氧化、棕榈酰化,聚乙二醇化、磷酸酯化、磷酸化、异戊烯化、丙酰化、retinylidene Schiff碱形成、S-谷胱甘肽化、S-亚硝基化、S-亚磺酰化、硒化、琥珀酰化、硫化、泛素化(ubiquitination)和C-末端酰胺化(也可参阅,Seo和Lee,2004,J.Biochem.Mol.Biol.37:35-44)。

[0484] 在某些实施方案中,凝集素用作结合剂,用于检测蛋白、多肽或肽的糖基化状态。凝集素是碳水化合物结合蛋白,其可以选择性地识别游离碳水化合物或糖蛋白的聚糖表位。识别各种糖基化状态(例如,核心-岩藻糖、唾液酸、N-乙酰基-D-乳糖胺、甘露糖、N-乙酰基-葡糖胺)的凝集素名单包括:A,AAA,AAL,ABA,ACA,ACG,ACL,AOL,ASA,BanLec,BC2L-A,BC2LCN,BPA,BPL,Calsepa,CGL2,CNL,Con,ConA,DBA,Discoidin,DSA,ECA,EEL,F17AG,Gal1,Gal1-S,Gal2,Gal3,Gal3C-S,Gal7-S,Gal9,GNA,GRFT,GS-1,GS-II,GSL-1,GSL-II,HHL,HIHA,HPA,I,II,Jacalin,LBA,LCA,LEA,LEL,扁豆,Lotus,LSL-N,LTL,MAA,MAH,MAL_I,Malectin,MOA,MPA,MPL,NPA,Orysata,PA-IIL,PA-IL,PALa,PHA-E,PHA-L,PHA-P,PHAE,PHAL,PNA,PPL,PSA,PSL1a,PTL,PTL-1,PWM,RCA120,RS-Fuc,SAMB,SBA,SJA,SNA,SNA-1,SNA-II,SSA,STL,TJA-I,TJA-II,TxLCI,UDA,UEA-1,UEA-II,VFA,VVA,WFA,WGA(see,Zhang et al.,2016,MABS 8:524-535)。

[0485] 在某些实施方案中,结合剂可以与修饰的或标记的NTAA结合。修饰或标记的NTAA可以是用PITC、1-氟-2,4-二硝基苯(Sanger试剂,DNFB)、丹磺酰氯(DNS-Cl,或1-二甲基氨基萘-5-磺酰氯)、4-磺酰基-2-硝基氟苯(SNFB)、乙酰化试剂、胍基化试剂、硫代酰化试剂、硫代乙酰化试剂、或硫代苄基化试剂之一标记的NTAA。

[0486] 在某些实施方案中,结合剂可以是适配体(例如,肽适体、DNA适体、或RNA适体)、抗体、anticalin、ATP依赖性Clp蛋白酶衔接蛋白(ClpS)、抗体结合片段、抗体模拟物、肽、肽模拟物、蛋白质、或多核苷酸(例如,DNA、RN、肽核酸(PNA)、γPNA、桥接核酸(BNA)、异种核酸(XNA)、甘油核酸(GNA)、或苏糖核酸(TNA)、或其变体)。

[0487] 如本文所用,术语抗体和抗体以广义使用,不仅包括完整的抗体分子,例如但不限于免疫球蛋白A、免疫球蛋白G、免疫球蛋白、免疫球蛋白E、和免疫球蛋白M,还包括特异性结合至少一个表位的抗体分子的免疫活性组分。抗体可以是天然存在的、合成产生的或重组表达的。抗体可以是融合蛋白。抗体可以是抗体模拟物。抗体的例子包括但不限于Fab片段、Fab'片段、$F(ab')_2$片段、单链抗体片段(scFv)、小抗体、双抗体、交联抗体片段、Affibody™、纳米抗体、单结构域抗体、DVD-Ig分子,α体、affimers、affitins、cyclotides,分子、和其类似物。使用抗体工程或蛋白质工程技术衍生的免疫反应产物也明确地在术语抗体的含义内。抗体和/或蛋白质工程的详细描述,包括相关的方案,可以在J.Maynard和G.Georgiou,2000,Ann.Rev.Biomed.Eng.2:339-76;Antibody Engineering,R.Kontermann和S.Dubel,eds.,Springer Lab Manual,Springer Verlag(2001);U.S.Patent No.5,831,012;和S.Paul,Antibody Engineering Protocols,Humana Press(1995),等其它地方找到,。

[0488] 与抗体一样,可以使用已知方法产生特异性识别肽的核酸和肽适体。适体虽然通常以高度亲和性高度特异、构象依赖方地结合靶分子,但可以选择具有较低结合亲和力的适体。已显示适体基于非常小的结构差异,例如甲基或羟基的存在或不存在,来区分靶标,并且某些适体可区分D-和L-对映体。已经获得了结合小分子靶标,包括药物、金属离子和有机染料、肽、生物素和蛋白质的适体,包括但不限于链霉亲和素、VEGF和病毒蛋白。已显示适

体在生物素化、荧光素标记后以及当附着于玻璃表面和微球时保留功能活性(参阅,Jayasena,1999,Clin Chem 45:1628-50;Kusser2000,J.Biotechnol.74:27-39;Colas,2000,Curr Opin Chem Biol 4:54-9)。特异性结合精氨酸和AMP的适体已经被描述(参阅,Patel和Suri,2000,J.Biotech.74:39-60)。Gold等人已经公开了与特定氨基酸结合的寡核苷酸适体(1995,Ann.Rev.Biochem.64:763-97)。结合氨基酸的RNA适体也已经被描述(Ames和Breaker,2011,RNA Biol.8;82-89;Mannironi et al.,2000,RNA 6:520-27;Famulok,1994,J.Am.Chem.Soc.116:1698-1706)。

[0489] 可以通过基因工程修饰天然存在的或合成产生的蛋白质来制备结合剂,以在氨基酸序列中引入一个或多个突变以产生结合大分子的特定组分或特征的工程化蛋白质(例如,NTAA、CTAA、或翻译后修饰的氨基酸或肽)。例如,可以修饰外肽酶(例如,氨肽酶、羧肽酶)、外切蛋白酶、突变的外切蛋白酶、突变的anticalin、突变的ClpS、抗体或tRNA合成酶,以产生选择性结合特定NTAA的结合剂。在另一个实例中,可以修饰羧肽酶以产生选择性结合特定CTAA的结合剂。结合剂也可以通过设计、或修饰、以及应用,以特异性结合修饰的NTAA或修饰的CTAA,例如具有翻译后修饰的CTAA(例如,磷酸化的NTAA或磷酸化的CTAA)、或经标签(例如,PTC、1-氟-2,4-二硝基苯(使用Sanger试剂,DNFB)、丹磺酰氯((使用DNS-Cl,或1-二甲基氨基萘-5-磺酰氯),或使用硫代酰化试剂、硫代乙酰化试剂、乙酰化试剂、酰胺化(胍基化)试剂或硫代苄基化试剂)修饰的CTAA。蛋白质定向进化的策略是本领域已知的(例如:reviewed by Yuan et al.,2005,Microbiol.Mol.Biol.Rev.69:373-392),包括噬菌体展示、核糖体展示、mRNA展示、CIS显示、CAD显示、乳化、细胞表面展示、,酵母表面展示、细菌表面展示等。

[0490] 在一些实施方案中,可以使用选择性结合修饰的NTAA的结合剂。例如,所述NTAA可以与异硫氰酸苯酯(PITC)反应形成苯基硫代氨基甲酰基-NTAA衍生物。以这种方式,所述结合剂可以被制成选择性地结合苯基硫代氨基甲酰基基团的苯团和所述NTAA的α-碳R基团二者。如下所述,以这种方式使用PITC允许随后通过Edman降解裂解所述NTAA。在另一个实施方案中,所述NTAA可以与Sanger试剂(DNFB)反应,以产生DNP标记的NTAA(参见图3)。任选地,DNFB与离子液体,例如1-[乙基-3-甲基咪唑双[(三氟甲基)磺酰基]酰亚胺([emim][Tf2N])一起使用,其中DNFB是高度可溶的。以这种方式,所述结合剂可以改造以选择性地结合所述NTAA上的所述DNP和所R基团的组合。DNP基团的添加为所述结合剂与所述NTAA的相互作用提供了更大的“手柄”,并且应该产生更高的亲和作用。在另一个实施方案中,结合剂可以是氨肽酶,其已被工程化以识别所述DNP标记的NTAA,从而提供所述肽的氨肽酶降解的循环对照。一旦所述DNP标记的NTAA被切割,就进行另一轮DNFB衍生化以结合并切割新暴露的NTAA。在优选的实施方案中,所述氨肽酶是单体金属蛋白酶,例如由锌活化的氨肽酶(Calcagno和Klein 2016)。在另一个实例中,结合剂可以选择性地结合用磺酰基硝基苯酚(SNP)修饰的NTAA,例如通过使用4-磺酰基-2-硝基氟苯(SNFB)。在另一个实施方案中,结合剂可以选择性地结合乙酰化或酰胺化的NTAA。

[0491] 可用于修饰所述NTAA的其它试剂包括异硫氰酸三氟乙酯、异硫氰酸烯丙酯、和异硫氰酸二甲氨基偶氮苯。

[0492] 为了修饰的NTAA具有高亲和力、修饰的NTAA具有高度特异性,或两者兼而有之,可以工程化改造结合剂。在一些实施方案中,可以通过使用噬菌体展示定向进化有表现好的

亲和支架来开发结合剂。

[0493] 结合并切割单个或小组标记(生物素化的)NTAA的工程化氨肽酶突变体已经被描述过(参见PCT公开号WO2010/065322,其通过引用整体并入)。氨肽酶是从蛋白或肽的N-末端切割氨基酸的酶。天然氨肽酶具有非常有限的特异性,并且以渐进方式一个接一个地切割N-末端氨基酸(Kishor et al.,2015,Anal.Biochem.488:6-8)。然而,残基特异性氨肽酶已经被鉴定出(Eriquez et al.,J.Clin.Microbiol.1980,12:667-71;Wilce et al.,1998,Proc.Natl.Acad.Sci.USA 95:3472-3477;Liao et al.,2004,Prot.Sci.13:1802-10)。可以将氨肽酶工程化以特异性结合20种不同的NTAA,其代表特定基团(例如,PTC,DNP,SNP等)标记的标准氨基酸。通过使用仅在标记存在时具有活性(例如结合活性或催化活性)的工程化氨肽酶来控制肽的N-末端的逐步降解。在另一个例子中,Havranak等人(美国专利公开2014/0273004)描述了工程化氨酰基tRNA合成酶(aaRS)作为特异性NTAA结合物。所述aaRS的氨基酸结合口袋能够天然地结合同源氨基酸,但通常表现出弱的结合亲和力和特异性。而且,这些天然氨基酸结合剂不识别N-末端标记。aaRS支架的定向进化可用于产生更高亲和力、更高特异性的结合剂,其在N-末端标签的环境下识别所述N-末端氨基酸。

[0494] 在另一个实例中,高选择性工程ClpS也已在文献中描述。Emili等描述了通过菌体展示的大肠杆菌ClpS蛋白的定向进化,产生具有对天冬氨酸、精氨酸、色氨酸和亮氨酸残基的NTAA选择性结合的能力的四种不同的变体(美国专利9,566,335,其全部内容通过引用并入)。

[0495] 在一个具体实施方案中,将anticalin工程化以对标记的NTAA(例如DNP,SNP、乙酰化等)具有高亲和力和高特异性。某些品种的anticalin支架凭借其β桶结构的优点具有合适于结合单个氨基酸的形状。N末端氨基酸(有或没有修饰)可能适合并在这个“β桶”桶中被识别。具有工程化新结合活性的高亲和力anticalin已经被描述(Skerra,2008,FEBS J.275:2677-2683综述)。例如,已经设计了具有高亲和力的anticalin结合(低nM)到荧光素和地高辛(digoxygenin)(Gebauer和Skerra 2012)。Banta等人也综述了用于新结合功能的替代支架的工程。(2013,Annu.Rev.Biomed.Eng.15:93-113)。

[0496] 通过使用单价结合剂的二价或更高次序多聚体,可以使给定单价结合剂的功能亲和力(亲合力)增加至少一个数量级(Vauquelin和Charlton 2013)。亲和力是指多种同时非共价结合相互作用的累积强度。个体结合相互作用可以容易地解离。然而,当同时存在多个结合相互作用时,单个结合相互作用的瞬时解离不允许结合蛋白扩散并且结合相互作用可能恢复。增加结合剂亲合力的另一种方法是在与所述结合剂连接的所述编码标签和与所述大分子相关的所述记录标签中包括互补序列。

[0497] 在一些实施方案中,可以采用选择性结合修饰的C-末端氨基酸(CTAA)的结合剂。羧肽酶是切割含有游离羧基的末端氨基酸的蛋白酶。许多羧肽酶表现出氨基酸偏好,例如,羧肽酶B优先在碱性氨基酸如精氨酸和赖氨酸上切割。可以修饰羧肽酶以产生选择性结合特定氨基酸的结合剂。在一些实施方案中,可以改造羧肽酶以选择性结合修饰部分以及CTAA的-碳R基团。因此,工程化的羧肽酶可以特异性识别代表C-末端标记背景中的标准氨基酸的20种不同的CTAA。通过使用仅在标记存在下具有活性(例如结合活性或催化活性)的工程化羧肽酶来控制从肽的C-末端逐步降解。在一个实例中,CTAA可以被对硝基苯胺或7-氨基-4-甲基香豆酰基修饰。

[0498] 可以设计来用于产生本文所述方法的粘合剂的其他潜在支架包括:anticalin,氨基酸tRNA合成酶(aaRS),ClpS,Affilin®,Adnectin™,T细胞受体,锌指蛋白,硫氧还蛋白,GST A1-1,DARPin,affimer,affitin,alphabody,avimer,Kunitz结构域肽,单体,单结构域抗体,EETI-II,HPSTI,胞内抗体,脂质运载蛋白,PHD指,V(NAR)LDTI,evibody,Ig(NAR),knottin,maxibody,neocarzinostatin,pVIII,tendamistat,VLR,蛋白A支架,MTI-II,ecotin,GCN4,Im9,kunitz结构域,微体,PBP,反式体,tetranectin,WW结构域,CBM4-2,DX-88,GFP,iMab,Ldl受体结构域A,Min-23,PDZ结构域,禽胰腺多肽,charybdotoxin/10Fn3,结构域抗体(Dab),a2p8锚蛋白重复,昆虫防御A肽,设计的AR蛋白,C型凝集素结构域,葡萄球菌核酸酶,Src同源结构域3(SH3),或Src同源结构域2(SH2)。

[0499] 结合剂可以设计以承受更高的温度和温和变性条件(例如,存在尿素、硫氰酸胍、离子溶液等)。采用变性剂有助于减少表面结合肽中的二级结构,例如-螺旋结构,发夹,-链,和其他此类结构,其可能干扰结合剂与线性肽表位的结合。在一个实施方案中,离子液体如乙酸1-乙基-3-甲基咪唑乙酸盐([EMIM]+[ACE]用于在结合循环期间减少肽二级结构(Lesch,Heuer et al.2015)。

[0500] 所描述的任何结合剂还包含含有关于所述结合剂的识别信息的编码标签。编码标签是约3碱基至约100个碱基的核酸分子,其为其相关结合剂提供独特的识别信息。编码标签可包含约3至约90个碱基,约3至约80个碱基,约3至约70个碱基,约3至约60个碱基,约3个碱基至约50个碱基,约3个碱基至约40个碱基,约3个碱基至约30个碱基,约3个碱基至约20个碱基,约3个碱基至约10个碱基,或约3个碱基至约8个碱基。在一些实施方案中,编码标签是约3个碱基,4个碱基,5个碱基,6个碱基,7个碱基,8个碱基,9个碱基,10个碱基,11个碱基,12个碱基,13个碱基,14个碱基,15个碱基,16个碱基,17个碱基,18个碱基,19个碱基,20个碱基,25个碱基,30个碱基,35个碱基,40个碱基,55个碱基,60个碱基,65个碱基,70个碱基,75个碱基,80个碱基,85个碱基,90个碱基,95个碱基,或100个碱基。编码标签可以由DNA、RNA、多核苷酸类似物、或其组合组成。多核苷酸类似物包括PNA、γPNA、BNA、GNA、TNA、LNA,吗啉代多核苷酸、2'-O-甲基多核苷酸、烷基核糖基取代的多核苷酸、硫代磷酸酯多核苷酸和7-脱氮嘌呤类似物。

[0501] 编码标签包含编码器序列,其提供关于所述相关结合剂的识别信息。编码序列为约3个碱基至约30个碱基,约3个碱基至约20个碱基,约3个碱基至约10个碱基,或约3个碱基至约8个碱基。在一些实施方案中,编码序列为约3个碱基,4个碱基,5个碱基,6个碱基,7个碱基,8个碱基,9个碱基,10个碱基,11个碱基,12个碱基,13个碱基,14个碱基,15个碱基,20个碱基,25个碱基或30个碱基的长度。所述编码器序列的长度决定了可以生成的独特编码器序列的数量。较短的编码序列产生较少数量的独特编码序列,这在使用少量结合剂时可能是有用的。当分析一群大分子时,可能需要更长的编码器序列。例如,5个碱基的编码序列具有5'-NNNNN-3'的公式(SEQ ID NO:135),其中N可以是任何天然存在的核苷酸或类似物。使用四种天然存在的核苷酸A,T,C和G,具有5个碱基长度的独特编码序列的总数是1,024。在一些实施例中,可以通过排除,例如其中所有碱基相同、至少三个连续碱基相同或两种情况的,编码器序列来减少独特编码序列的总数。在特定实施例中,一组≥50个独特编码器序列用于结合剂库。

[0502] 在一些实施例中,识别编码标签或记录标签的组件,例如编码器序列、条形码、

UMI、隔室标签、分区条形码、样本条形码、空间区域条形码、循环特定序列或其任何组合,受限于汉明距离、Lee距离、不对称Lee距离、Reed-Solomon、Levenshtein-Tenengolts、或类似的纠错方法。汉明距离是指两个相等长度的字符串之间不同的位置的个数。它测量将一个字符串更改为另一个字符串时所需的最小替换次数。汉明距离可以用于通过选择合理距离的编码器序列来纠正错误。因此,在编码器序列是5碱基的示例中,编码器序列的数量减少到256个独特编码器序列(汉明距离1→4^4个编码器序列=256个编码器序列)。在另一个实施例中,所述编码器序列、条形码、UMI、隔室标签、循环特定序列、或其任何组合被设计为易于通过循环解码过程读出(Gunderson,2004,Genome Res.14:870-7)。在另一个实施方案中,所述编码器序列、条形码、UMI、隔室标签、分区条形码、空间条形码、样品条形码,循环特异性序列或其任何组合被设计为通过低精度纳米孔测序读出,因为不需要单个碱基分辨率,而是需要读取复合碱基(长度约5-20个碱基)的词。可以在本公开的方法中使用的15-mer,错误校正汉明条形码的子集如SEQ ID NO:1-65所示,并且它们相应的反向互补序列如SEQ ID NO:66-130所示。

[0503] 在一些实施方案中,结合剂库内的每种独特结合剂具有独特编码序列。例如,20种独特的编码序列可用于结合20种标准氨基酸的20种结合剂的文库。额外的编码标签序列可用于识别修饰的氨基酸(例如,翻译后修饰的氨基酸)。在另一个实例中,30种独特编码序列可用于结合所述20种标准氨基酸和10种翻译后修饰氨基酸(例如磷酸化氨基酸、乙酰化氨基酸、甲基化氨基酸)的30种结合剂的文库。在其它实施例中,两个或更多个不同的结合剂可以共享相同的编码器序列。例如,各自结合不同的标准氨基酸的两种结合剂可以共享相同的编码序列。

[0504] 在某些实施方案中,编码标签还在一端或两端包含间隔序列。间隔序列为约1个碱基至约20个碱基,约1个碱基至约10个碱基,约5个碱基至约9个碱基,或约4个碱基至约8个碱基。在一些实施方案中,间隔子的长度约1个碱基,2个碱基,3个碱基,4个碱基,5个碱基,6个碱基,7个碱基,8个碱基,9个碱基,10个碱基,11个碱基,12个碱基,13个碱基,14个碱基,15个碱基、或20个碱基。在一些实施方案中,编码标签内的间隔子比编码序列短,例如,至少1个碱基,2个碱基,3个碱基,4个碱基,5个碱基,6个碱基,7个碱基,8个碱基,9个碱基,10个碱基,11个碱基,12个碱基,13个碱基,14个碱基,15个碱基,20个碱基或25个碱基。在其它实施例中,编码标签内的间隔子与所述编码器序列的长度相同。在某些实施方案中,间隔子是结合剂特异性的,使得来自先前结合循环的间隔子仅与来自当前结合循环中的适合结合剂的间隔子相互作用。一个例子是含有间隔序列的同源抗体对,仅当两种抗体顺序结合所述大分子时,允许信息转移。间隔序列可用作引物延伸反应的引物退火位点,或连接反应中的夹板或粘端。编码标签上的5'间隔区(参见图5A,"*Sp'")可任选地包含与记录标签上的3'间隔区的伪互补碱基以增加Tm(Lehoud et al.,2008,Nucleic Acids Res.36:3409-3419)。

[0505] 在一些实施方案中,结合剂集合内的所述编码标签共享测定中使用的共同间隔序列(例如,在多路复用结合循环方法中使用的整个结合剂文库在其编码标签中具有共同的间隔子)。在另一个实施方案中,所述编码标签包含结合循环标签,识别特定的结合循环。在其它实施方案中,所述结合剂文库内的编码标签具有结合循环特异性间隔序列。在一些实施方案中,编码标签包含一个结合循环特异性间隔序列。例如,用于第一结合循环中的结合

剂的编码标签包含“循环1”特异性间隔序列,用于第二结合循环的结合剂的编码标签包含“循环2”特异性间隔序列,等等达到“n”个结合循环。在进一步的实施方案中,用于第一结合循环的结合剂的编码标签包含“循环1”特异性间隔序列和“循环2”特异性间隔序列,用于第二结合循环的结合剂的编码标签包括“循环2”特异性间隔序列和“循环3“特异性间隔序列,等等至“n”结合循环。该实施方案可用于在结合循环完成后的非级联延伸记录标签的后续PCR组装(参见图10)。在一些实施方案中,间隔序列包含足够数量的碱基以与记录标签或延伸记录标签中的互补间隔序列退火以引发引物延伸反应或粘性末端连接反应。

[0506] 当一群记录标签与大分子相关联时,循环特异性间隔序列也可用于将编码标签的信息级联到单个记录标签上。第一个结合循环将信息从所述编码标签转移到随机选择的记录标签,随后的结合循环可以通过循环依赖的间隔序列仅引发所述延伸记录标签。更具体地,在第一结合循环中使用的结合剂的编码标签包括“循环1”特异性间隔序列和“循环2”特异性间隔序列,在第二结合循环中使用的结合剂的编码标签包括“循环2”特异性间隔序列和“循环3”特异性间隔序列,等等至“n”结合循环。来自第一结合循环的结合剂的编码标签能够通过互补循环1特异性间隔序列与记录标签退火。在将所述编码标签信息传送到所述记录标签时,所述循环2特异性间隔所述1结束时位于所述延伸记录标签的3’末端。来自第二结合循环的结合剂的编码标签能够通过互补的循环2特异性间隔序列与所述延伸记录标签退火。在将编码标签信息转移到延伸记录标签时,循环3特异性间隔序列在结合周期2结束时位于所述延伸记录标签的3’末端,依此类推走完“n”个结合循环。该实施方案规定,多个结合循环间的某个特定结合循环中的结合信息的转移将仅发生在已经经历先前结合循环的(延伸的)记录标签上。然而,有时结合剂将不能与同源大分子结合。在每个结合循环后包含结合循环特异性间隔子的寡核苷酸作为“追踪”步骤可用于保持所述结合循环同步,即使结合循环事件失败也是如此。例如,如果同源结合剂在结合循环1期间不能结合大分子,则在结合循环1之后用包含循环1特异性间隔子、循环2特异性间隔子和“无效”编码序列的寡核苷酸增加追踪步骤。所述“无效”编码器序列可以缺编码器序列,或者优选地,是肯定地识别“无效”结合循环的特异性条形码。所述“无效”寡核苷酸能够通过所述循环1特异性间隔子与所述记录标签退火,并且所述循环2特异性间隔子被转移至所述记录标签。因此,尽管结合循环1事件失败,来自结合循环2的结合剂仍然能够通过所述循环2特异性间隔子与所述延伸记录标签退火。所述“无效”寡核苷酸将结合循环1标记为延伸记录标签内的失败结合事件。

[0507] 在优选实施例中,结合循环特异性编码器序列用于编码标签。结合循环特定的编码器序列可以通过使用完全独特的分析物结合循环编码器条形码(例如,NTAA)或组合使用连接到循环特异性条形码(见图35)的分析物(例如,NTAA)编码器序列来实现。使用组合方法的优点是需要设计的总条形码更少。对于用于10个循环的一组20种分析物结合剂,仅需要设计20个分析物编码器序列条形码和10个结合循环特异性条形码。相反,如果所述结合循环直接嵌入所述结合剂编码器序列中,则可能需要设计总共200个独立的编码器条形码。将结合循环信息直接嵌入所述编码器序列中的优点在于,当在纳米孔读出上采用纠错条形码时,所述编码标签的总长度可以最小化。使用容错条形码允许使用更容易出错的测序平台和方法进行高度准确的条形码识别,但具有其他优点,例如快速分析、更低成本、和/或更便携的仪器。一个这样的例子是基于纳米孔的测序读数。

[0508] 在一些实施方案中,在所述结合剂附近的第二(3')间隔序列内编码标签包含可切割或可缺刻的DNA链(参见图32)。例如,所述3'间隔子可具有一个或多个尿嘧啶碱基,其可被尿嘧啶特异性切除试剂(USER)切刻。USER在所述尿嘧啶的位置产生单核苷酸间隙。在另一个实例中,所述3'间隔子可包含对切刻核酸内切酶的识别序列,其仅水解双链体的一条链。优选地,用于切割或切刻所述3'间隔序列的酶仅作用于一条DNA链(所述编码标签的3'间隔子),使得属于所述(延伸的)记录标签的双链体内的另一条链保持完整。这些实施方案在对天然构象蛋白质的分析中特别有用,因为它允许在引物延伸发生后从所(延伸的)记录标签中非变性去除所述结合剂,并在可用于后续结合循环的所述延伸记录标签上留下单链DNA间隔序列。

[0509] 所述编码标签也可以设计成包含回文序列。将回文序列包含在编码标签中允许新生的、生长的、延伸的记录标签在编码标签信息转移时折叠自己。所述延伸记录标签折叠成更紧凑的结构,有效地减少了不希望的分子间结合和引物延伸事件。

[0510] 在一些实施方案中,编码标签包含分析物特异性间隔子,其能够仅在用先前识别相同分析物的结合剂延伸的记录标签上引发延伸。可以使用包括分析物特异性间隔子和编码器序列的编码标签从一系列结合事件建立延伸记录标签。在一个实施方案中,第一结合事件使用具有编码标签的结合剂,所述编码标签包含用于下一个结合循环的通用3'间隔引物序列和5'末端的分析物特异性间隔序列;随后的结合循环接着使用具有编码的分析物特异性3'间隔序列的结合剂。该设计导致仅从正确的一系列同源结合事件创建可扩增的文库元件。脱靶和交叉反应的结合相互作用将导致不可扩增的延伸记录标签。在一个实例中,在两个结合循环中使用特定大分子分析物的一对同源结合剂以鉴定所述分析物。第一种所述同源结合剂含有编码标签,该编码标签包含用于在所述记录标签的通用间隔序列上引发延伸的通用间隔子3'序列,以及在5'末端的编码的分析物特异性间隔子,他们将用在下一结合循环中。对于匹配的同源结合剂对,所述第二结合剂的3'分析物特异性间隔子与所述第一结合剂的5'分析物特异性间隔子匹配。以这种方式,只有所述结合剂的同源对的正确结合才会产生可扩增的延伸记录标签。交叉反应性结合剂将不能在所述记录标签上引发延伸,并且不产生可扩增的延伸记录标签产物。该方法极大地增强了本文所公开的方法的特异性。相同的原理可以应用于三联体结合剂组,其中使用3个循环的结合。在第一个结合循环中,所述记录标签上的通用3'Sp序列与结合剂编码标签上的通用间隔子相互作用。引物延伸将编码标签信息,包括分析物特异性5'间隔子,转移至所述记录标签。随后的结合循环采用结合剂编码标签上的分析物特异性间隔子。

[0511] 在某些实施方案中,编码标签可以进一步包含连接在编码标签的所述结合剂的独特分子标识符。所述结合剂的UMI可用于采用延伸编码标签或用于测序读数的di-tag分子实施方案中,其与编码器序列组合提供关于所述结合剂的身份和大分子的独特结合事件的数量的信息。

[0512] 在另一实施例中,编码标签包括随机化序列(一组N',其中N=来自A,C,G,T的随机选项,或来自一组单词的随机选项)。经过一系列"n"结合循环并将编码标签信息转移到所述(延伸)记录标签后,最终的延伸记录标签产品将由一系列这些随机序列组成,这些序列共同形成用于所述最终延伸记录标签的"复合"独特分子标识符(UMI)。例如,如果每个编码标签包含(NN)序列(4*4=16个可能的序列),则在10个测序循环后,形成10个分布式2-mer

的组合组,产生用于延伸记录标签产品的总共16^{10}～10^{12}个多样性可能的复合UMI序列。鉴于肽测序实验使用～10^{9}个分子,这种多样性足以为测序实验创建一组有效的UMI。通过简单地在所述编码标签内使用较长的随机区域(NNN,NNNN等),可以实现更多多样性。

[0513] 编码标签可包括在3'间隔序列的3'末端并入的终止子核苷酸。在结合剂与大分子结合并且其相应的编码标签和记录标签通过互补的间隔序列退火后,引物延伸可以将信息从所述编码标签转移到所述记录标签,或者将信息从所述记录标签转移到所述编码标签。在编码标签的3'末端添加终止子核苷酸可防止记录标签信息传递所述编码标签。应当理解,对于本文所述涉及生成延伸编码标签的实施方案,优选在所述记录标签的3'末端包括终止子核苷酸,以防止编码标签信息转移到所述记录标签。

[0514] 编码标签可以是单链分子,双链分子或部分双链。编码标签可包括平末端、悬挂末端或一样一个。在一些实施方案中,编码标签是部分双链的,其防止所述编码标签与延长的延伸记录标签中的内部编码器和间隔序列退火。

[0515] 编码标签通过本领域已知的任何方式,包括共价和非共价相互作用,直接或间接地与结合剂连接。在一些实施方案中,编码标签可以酶促或化学方式与结合剂连接。在一些实施方案中,编码标签可以通过连接与结合剂结合。在其他实施方案中,编码标签通过亲和结合对(例如,生物素和链霉亲和素)与结合剂连接。

[0516] 在一些实施方案中,通过SpyCatcher-SpyTag相互作用将结合剂连接至编码标签(参见图43B)。所述SpyTag肽通过自发的异肽连接与SpyCatcher蛋白形成不可逆的共价键,从而提供遗传编码方式以产生抵抗力和苛刻条件的肽相互作用(Zakeri et al.,2012,Proc.Natl.Acad.Sci.109:E690-697;Li etal.,2014,J.Mol.Biol.426:309-317)。结合剂可以表达为包含所述SpyCatcher蛋白的融合蛋白。在一些实施方案中,所述SpyCatcher蛋白附加在所述结合剂的N-末端或C-末端。可以使用标准缀合化学法(Bioconjugate Techniques,G.T.Hermanson,Academic Press(2013))将所述SpyTag肽偶联至所述编码标签。

[0517] 在其他实施方案中,通过SnoopTag-SnoopCatcher肽-蛋白质相互作用将结合剂连接至编码标签。所述SnoopTag肽与SnoopCatcher蛋白形成异肽键(Veggiani et al.,Proc.Natl.Acad.Sci.USA,2016,113:1202-1207)。结合剂可以表达为包含SnoopCatcher蛋白的融合蛋白。在一些实施方案中,所述SnoopCatcher蛋白附加在所述结合剂的N-末端或C-末端。可以使用标准缀合化学将所述SnoopTag肽偶联至所述编码标签。

[0518] 在其它实施方案中,结合剂通过所述HaloTag®蛋白质融合标签及其化学配体结合到编码标签。HaloTag是改性的卤代烷脱卤素酶,其被设计来与合成配体(HaloTag配体)共价结合(Los et al.,2008,ACS Chem.Biol.3:373-382)。所述合成配体包含附着到各种有用分子的氯代烷烃接头。在HaloTag和高度特异性的所述氯代烷烃接头之间形成共价键,是在生理条件下快速发生,并且基本上是不可逆的。

[0519] 在某些实施方案中,还使大分子与非同源结合剂接触。如本文所用,非同源结合剂是指对不同的大分子特征或组分,而不是对所考虑的特定大分子具有选择性的结合剂。例如,如果所述n NTAA是苯丙氨酸,并且所述肽与分别对苯丙氨酸、酪氨酸和天冬酰胺具有选择性的三种结合剂接触,则对苯丙氨酸具有选择性的结合剂将是能够选择性结合第n NTAA(即,苯丙氨酸)的第一结合剂,而另外两种结合剂是所述肽的非同源结合剂(因为它们对苯

丙氨酸以外的NTAA具有选择性)。然而,酪氨酸和天冬酰胺结合剂可以是样品中其它肽的同源结合剂。然后如果所述n NTAA(苯丙氨酸)从所述肽上裂解,从而将肽的n-1氨基酸转化为n-1NTAA(例如酪氨酸),然后使所述肽与相同的三种结合剂接触,对酪氨酸具有选择性的结合剂是能够选择性结合所述n-1NTAA(即酪氨酸)的第二种结合剂,而另外两种结合剂是非同源结合剂(因为它们对酪氨酸以外的NTAAs具有选择性)。

[0520] 因此,应该理解,试剂是结合剂还是非同源结合剂取决于目前用于结合的特定大分子特征或组分的性质。并且,如果在多路复用反应中分析多个大分子,则一种大分子的结合剂可以是另一种大分子的非同源结合剂,反之亦然。因此,应该理解,以下关于结合剂的描述适用于本文所述的任何类型的结合剂(即,同源和非同源结合剂)。

[0521] VIII.编码标签信息到记录标签的循环传递

[0522] 在本文所述的方法中,在结合剂与大分子结合时,其连接的编码标签的识别信息被传递到与所述大分子相关的记录标签,从而产生“延伸记录标签”。所述延伸记录标签可以包含来自结合剂的编码标签的信息,该编码标签代表执行的每个结合循环。然而,延伸记录标签也可能“错过”结合循环,例如,因为结合剂不能与大分子结合,因为所述编码标签缺失、损坏或有缺陷,因为引物延伸反应失败。即使发生结合事件,信息从编码标签到记录标签的传递也可能不完整或不是100%准确,例如,因为编码标签损坏或有缺陷,因为在所述引物延伸反应中引入了错误。因此,延伸记录标签可以代表其相关的大分子上发生的100%,或高达95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%的结合事件。此外,所述延伸记录标签中存在的所述编码标签信息可以与相应的编码标签至少30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%一致性。

[0523] 在某些实施方案中,延伸记录标签可以包含来自表示多个连续结合事件的多个编码标签的信息。在这些实施方案中,单个级联的延伸记录标签可以代表单个大分子(参见图2A)。如本文所提到的,将编码标签信息传递到记录标签,还包括传递到延伸记录标签,如在涉及多个连续结合事件的方法中所发生的那样。

[0524] 在某些实施方案中,结合事件信息以循环方式从编码标签传递到记录标签(参见图2A和2C)。通过要求至少两个不同的编码标签(识别两个或更多独立结合事件)映射到相同类别的结合剂(与特定蛋白同源),可以在测序后根据信息滤除交叉反应性结合事件。所述记录标签中可以包含可选的样品或隔室条形码以及可选的UMI序列。所述编码标签还可以包含可选的UMI序列以及所述编码器和间隔子序列。延伸记录标签中也可以包含通用引物序列(U1和U2),用于扩增和NGS测序(参见图2A)。

[0525] 可以使用各种方法将与特定结合剂相关的编码标签信息传递到记录标签。在某些实施方案中,编码标签的信息通过引物延伸传递到记录标签(Chan,McGregor et al.2015)。记录标签或延伸记录标签的3'-末端上的间隔子序列,与编码标签的3'末端上的互补间隔子序列退火,并且聚合酶(例如,链置换聚合酶)使用所述退火的编码标签作为模板延伸所述记录标签序列(参见图5-7)。在一些实施方案中,可以将与编码标签编码器序列和5'间隔子互补的寡核苷酸预先退火至所述编码标签,以防止所述编码标签与延伸记录标签中存在的内部编码器和间隔子序列杂交。所述编码标签上的3'末端间隔子保持单链,优选与所述记录标签上的3'末端间隔子结合。在其他实施方案中,新生的记录标签可以用单

链结合蛋白包被,以防止所述编码标签退火到内部位点。或者,新生的记录标签也可以用RecA(或相关的同源物,例如uvsX)包被,以促进所述3'末端侵入到完全双链编码标签(Bell et al.,2012,Nature 491:274-278)。这种配置防止所述双链编码标签与内部记录标签元件相互作用,然而易受所述延伸记录标签被RecA包裹的3'尾部的链侵入(Bell,et al.,2015,Elife 4:e08646)。单链结合蛋白的存在可促进链置换反应。

[0526] 在一个优选的实施方案中,用于引物延伸的DNA聚合酶具有链置换活性并且具有有限的或没有3'-5核酸外切酶活性。这些聚合酶的许多实例中,有几个包含Klenow exo-(DNA Pol 1的Klenow片段)、T4 DNA聚合酶exo-、T7 DNA聚合酶exo(Sequenase 2.0)、Pfu exo-、Vent exo-、Deep Vent exo-、Bst DNA聚合酶大片段exo-、Bca Pol、9°N Pol和Phi29 Pol exo-。在优选的实施方案中,所述DNA聚合酶在室温至45℃有活性。在另一个实施方案中,采用嗜热聚合酶的"热启动"形式,使得聚合酶被活化并在大约40℃-50℃之间使用。示例性的热启动聚合酶是Bst 2.0Warm Start DNA聚合酶(New England Biolabs)。

[0527] 用于链置换复制的添加剂包括细菌、病毒或真核来源的许多单链DNA结合蛋白(SSB蛋白)中的任何一种:例如,大肠杆菌的SSB蛋白、噬菌体T4基因32产物、噬菌体T7基因2.5蛋白、噬菌体Pf3 SSB、复制蛋白ARPA32和RPA14亚基(Wold,1997);其他DNA结合蛋白,如腺病毒DNA结合蛋白、单纯疱疹蛋白ICP8、BMRF1聚合酶辅助亚基、疱疹病毒UL29SSB样蛋白;已知参与DNA复制的许多复合蛋白中的任何一种,如噬菌体T7解旋酶/引物、噬菌体T4基因41解旋酶、大肠杆菌Rep解旋酶、大肠杆菌recBCD解旋酶、recA、大肠杆菌和真核拓扑异构酶.(Champoux,2001)。

[0528] 通过在引物延伸反应中包含单链结合蛋白(T4基因32,大肠杆菌SSB等),DMSO(1-10%),甲酰胺(1-10%),BSA(10-100μg/ml),TMAC1(1-5mM),硫酸铵(10-50mM),甜菜碱(1-3M),甘油(5-40%),或乙二醇(5-40%),错误引发或自引发事件(例如当重新编码标签的末端间隔序列引发延伸自我延伸)可以最小化。

[0529] 大多数A型聚合酶缺乏3'核酸外切酶活性(内源性或工程化去除),例如Klenow exo-、T7 DNA聚合酶exo-(Sequenase 2.0)和Taq聚合酶催化非模板化添加核苷酸、优选腺苷碱基(在较小程度上,G碱基,取决于序列背景)至双链扩增产物的3'平端。对于Taq聚合酶,3'嘧啶(C>T)使非模板化的腺苷添加最小化,而3'嘌呤核苷酸(G>A)有利于非模板化的腺苷添加。在使用Taq聚合酶进行引物延伸的实施方案中,将胸苷碱基设置在编码标签中位于远离结合剂的间隔序列和相邻条形码序列(例如,编码序列或循环特异性序列)之间,以协调容纳在所述记录标签的间隔序列的3'末端上的偶发性非模板化腺苷酸(图43A)。以这种方式,延伸的记录标签(具有或不具有非模板化的腺苷碱基)可以与编码标签退火并进行引物延伸。

[0530] 或者,通过使用突变聚合酶(嗜温或嗜热)可以减少非模板碱基的添加,其中非模板化的末端转移酶活性已经被一个或多个点突变大大降低,尤其是在O-螺旋区域(参见美国专利7,501,237)(Yang,Astatke等2002)。Pfu exo-,其具有3'核酸外切酶缺陷并具有链置换能力,也不具有非模板化的末端转移酶活性。

[0531] 在另一个实施方案中,最佳聚合酶延伸缓冲液包含40-120mM缓冲剂,例如Tris-乙酸盐、Tris-HCl、HEPES等,pH为6-9。

[0532] 通过在记录/延伸记录标签中包括伪互补碱基,可以最小化通过延伸记录标签的

末端间隔序列与延伸记录标签的内部区域的自退火引发的自引发/误引发事件(Lahoud,Timoshchuk等.2008),(Hoshika,Chen等.2010)。由于存在化学修饰,假互补碱基显示出显著降低的彼此形成双链体的杂交亲和力。然而,许多假互补修饰碱基可以与天然DNA或RNA序列形成强碱基对。在某些实施方案中,编码标签间隔子序列包含多个A和T碱基,可以使用亚磷酰胺寡核苷酸合成将市售的假互补碱基2-氨基腺嘌呤和2-硫代胸腺嘧啶掺入记录标签中。通过向反应中添加假互补核苷酸,可以在引物延伸期间将额外的假互补碱基掺入延伸的记录标签中(Gamper,Arar等.2006)。

[0533] 为了使溶液中所述编码标签标记的结合剂与固定的蛋白的所述记录标签的非特异性相互作用最小化,将与记录标签间隔子序列互补的竞争性(也称为阻断)寡核苷酸添加至结合反应以最小化非特异性相互作用(图32A-D)。封闭寡核苷酸相对较短。在引物延伸之前从结合反应中洗去过量的竞争性寡核苷酸,这有效地使退火的竞争性寡核苷酸从所述记录标签解离,特别是当暴露于稍高的温度(例如,30-50℃)时。封闭寡核苷酸可在其3'末端包含终止子核苷酸以防止引物延伸。

[0534] 在某些实施方案中,所述记录标签上的间隔子序列与编码标签上的互补间隔子序列的退火,在引物延伸反应条件下是亚稳定的(即,退火Tm类似于反应温度)。这允许编码标签的间隔子序列置换退火至记录标签的间隔子序列的任何阻断寡核苷酸。

[0535] 与特定结合剂相关的编码标签信息也可以通过连接传递到记录标签(参见如图6和7)。连接可以是平末端连接或粘末端连接。连接可以是酶促连接反应。连接酶的实例包括但不限于T4 DNA连接酶、T7 DNA连接酶、T3DNA连接酶、Taq DNA连接酶、大肠杆菌DNA连接酶、9°N DNA连接酶、Electroligase®。或者,连接可以是化学连接反应(参见图7)。在图示中,通过使用"记录助手"序列与所述编码标签上的臂的杂交来实现无间隔连接。使用标准化学连接或"点击化学"化学连接退火的补体序列(Gunderson,Huang et al.1998,Peng,Li et al.2010,El-Sagheer,Cheong et al.2011,El-Sagheer,Sanzone et al.2011,Sharma,Kent et al.2012,Roloff and Seitz 2013,Litovchick,Clark et al.2014,Roloff,Ficht et al.2014)。

[0536] 在另一个实施方案中,PNA的传递可以使用公开的技术通过化学连接完成。PNA的结构是具有5'N-末端胺基和非反应性3'-C-末端酰胺。PNA的化学连接需要将末端修饰为具有化学活性。这通常通过用半胱氨酰基部分衍生5'N-末端和用硫酯部分衍生3'-C-末端来完成。这种修饰的PNA容易使用标准的天然化学连接条件偶联(Roloff et al.,2013,Bioorgan.Med.Chem.21:3458-3464)。

[0537] 在一些实施方案中,可以使用拓扑异构酶传递编码标签信息。拓扑异构酶可用于将记录标签上的带顶部的3'磷酸连接到编码标签的5'末端或其互补物上(Shuman et al.,1994,J.Biol.Chem.269:32678-32684)。

[0538] 如本文所述,结合剂可以与翻译后修饰的氨基酸结合。因此,在涉及肽大分子的某些实施方案中,延伸的记录标签包含与氨基酸序列和翻译后修饰有关的编码标签信息。在一些实施方案中,在检测和切割末端氨基酸(例如,NTAA或CTAA)之前,完成对内部翻译后修饰(例如,磷酸化、糖基化、琥珀酰化、泛素化、S-亚硝基化、甲基化、N-乙酰化、脂化等)的氨基酸的检测。在一个实例中,使肽与用于PTM修饰的结合剂接触,并且如上所述将相关的编码标签信息传递至记录标签(参见图8A)。一旦完成了与氨基酸修饰相关的编码标签信息的

检测和传递,就可以在使用N-末端或C-末端降解方法检测和传递所述初始氨基酸序列的编码标签信息之前去除PTM修饰基团。因此,产生的的延伸记录标签指示肽序列中存在翻译后修饰的存在以及初始氨基酸序列信息,虽然不是连续顺序(参见图8B)。

[0539] 在一些实施方案中,内部翻译后修饰的氨基酸的检测可以与初始氨基酸序列的检测同时发生。在一个实例中,NTAA(或CTAA)与翻译后修饰的氨基酸特异性结合剂接触,单独或作为结合剂文库的一部分(例如,由20种标准氨基酸的结合剂组成的文库和选定的翻译后修饰氨基酸)。随后是末端氨基酸切割和与结合剂(或结合剂文库)的接触的连续循环。因此,得到的延伸记录标签指示在初始氨基酸序列中的翻译后修饰的存在和顺序。

[0540] 在某些实施方案中,每个大分子可以采用一组记录标签,以提高编码标签信息传递的整体稳健性和效率(参见如图9)。使用与给定大分子相关的一组记录标签而不是单个记录标签,提高了文库构建效率,因为编码标签与记录标签的潜在偶联产量更高,文库总产量也更高。单个级联延伸记录标签的产量直接取决于级联的逐步产量,而使用能够接受编码标签信息的多个记录标签不会遭受别联的指数损失。

[0541] 在图9和10中示出了这种实施方案的一个实例。在图9A和10A中,多个记录标签与固体支持物上的单个大分子(通过空间共定位到或单个大分子限制到在单个珠子)相关联。结合剂以循环方式暴露于固体支持物,并且它们相应的编码标签在每个循环中将信息传递给共定位的多个记录标签中的一个记录标签。在图9A所示的例子中,结合周期信息被编码到编码标签上存在的间隔子中。对于每个结合循环,用指定的循环特异性间隔子序列标记该组结合剂(图9A和9B)。例如,在NTAA结合剂的情况下,相同氨基酸残基的结合剂用不同的编码标签标记,或在间隔子序列中包含循环特异性信息以表示结合剂特性和循环数。

[0542] 如图9A所示,在第一个结合循环(循环1)中,使多个NTAA结合剂与大分子接触。循环1中使用的结合剂,具有与记录标签的间隔子序列互补的共同间隔子序列。循环1中使用的结合剂还具有包含循环1特异性序列的3'-间隔子序列。在结合循环1期间,第一个NTAA结合剂结合大分子的游离末端,第一个编码标签中的共同间隔子序列的互补序列跟记录标签退火,第一个编码标签的信息通过来自共同间隔序列的引物延伸被传递到同源记录标签。在除去NTAA以暴露新的NTAA后,结合循环2接触多个NTAA结合剂(具有与记录标签的间隔子序列互补的共同间隔子序列)。循环2中使用的结合剂还具有包含循环2特异性序列的3'-间隔子序列。第二个NTAA结合剂与大分子的NTAA结合,第二个编码标签的信息通过引物延伸传递至记录标签。这些循环重复直至"n"个结合循环,产生多个与单个大分子共定位的延伸记录标签,其中每个延伸记录标签具有来自一个结合循环的编码标签信息。因为在每个连续结合循环中使用的每组结合剂在所述编码标签中具有循环特异性间隔子序列,所以结合循环信息可以与产生的延伸记录标签中的结合剂信息相关联。

[0543] 在一个替代实施方案中,多个记录标签与固体支持物(例如,珠子)上的单个大分子相关联,如图9A中所示,但是在这种情况下,在特定结合周期中使用的结合剂具有编码标签,其侧接一个用于当前的结合循环的循环特异性间隔子和一个用于下一结合循环的循环特异性间隔子。这种设计的原因是支持最终组装PCR步骤(图10C)以将延伸记录标签群转换成单个共线延伸记录标签。在测序之前,可以对单个共线延伸记录标签的文库进行富集、扣减和/或标准化方法。在第一个结合循环(循环1)中,在第一个结合剂结合后,包含循环1特异性间隔子(C'1)的编码标签的信息被传递到包含互补的循环1特异性间隔子(C1)的记录

标签上。在第二结合循环(循环2)中,在第二结合剂结合后,包含循环2特异性间隔子(C'2)的编码标签的信息被传递至末端包含互补的循环2特异性间隔子(C2)的不同记录标签。该过程持续到第n个结合循环。在一些实施方案中,延伸记录标签中的第n个编码标签用通用反向引发序列加盖,例如,通用反向引发序列可以作为第n个编码标签设计的一部分并入,或者将所述通用反向引发序列添加到第n个结合循环后的后续反应,比如使用加尾引物的扩增反应中。在一些实施方案中,在每个结合循环中,将大分子暴露于与编码标签连接的结合剂集合,所述编码标签包含关于其相应结合剂的识别信息和结合周期信息(图9和图10)。在一个具体实施方案中,在完成第n个结合循环后,将涂覆有延伸记录标签的珠子基质置于油乳液中,使得平均存在少于或约等于1个珠子/液滴。然后使用组装PCR从珠子扩增所述延伸记录标签,并通过在单独的延伸记录标签内通过循环特异性间隔子序列引发来共线组装多个单独的记录标签(图10C)(Xiong et al.,2008,FEMS Microbiol.Rev.32:522-540)。或者,可以在每个结合循环期间或之后将循环特异性间隔子分别添加到延伸记录标签,而不使用具有结合剂编码标签的循环特异性间隔子。使用一组延伸记录标签,即一组延伸记录标签一起代表一个单独的大分子相比于单个级联延伸记录标签代表一个单独的大分子的一个优点是更高浓度的记录标签可以提高所述编码标签信息的传递效率。此外,结合循环可以重复几次以确保同源结合事件完成。此外,延伸记录标签的表面扩增可能会提供信息传递的冗余(参见图4B)。如果编码标签信息不是一直被传递,则在大多数情况下,仍然可以使用不完整的编码标签信息集合来识别具有非常高信息含量的大分子,例如蛋白。即使是短肽也可以包含大量可能的蛋白序列。例如,10-mer肽具有20^{10}种可能的序列。因此,包含缺失和/或模糊的部分或不完整序列仍然能够唯一地映射。

[0544] 在一些实施方案中,其中蛋白以其天然构象被检测,用含有编码标签的结合剂进行环状结合测定,所述编码标签在临近所述结合剂的间隔元件内包含可切割或可切刻的DNA链(图32)。例如,靠近结合剂的间隔子可具有一个或多个尿嘧啶碱基,其可被尿嘧啶特异性切除试剂(USER)切刻。在另一个实例中,靠近结合剂的间隔子可包含切割核酸内切酶的识别序列,其仅水解双链体的一条链。该设计允许从延伸记录标签中非变性地去除结合剂,并产生游离的单链DNA间隔元件,用于随后的免疫测定循环。在一个优选的实施方案中,将尿嘧啶碱基掺入编码标签中,以允许在引物延伸步骤后,酶促用户(USER)除去结合剂(图32E-F)。在用户切除尿嘧啶后,可以在各种温和条件下,包括高盐(4M NaCl,25%甲酰胺)和温和加热以破坏蛋白质结合剂相互作用。在记录标签上保持退火的另一个截短的编码标签DNA短柱(图32F),在略微升高的温度下容易解离。

[0545] 在接近结合剂的间隔元件内,由可切割或可切刻的DNA链组成的编码标签,也允许单一同源测定,用于从多个结合的结合剂传递编码标签信息(参见图33)。在一个优选的实施方案中,接近结合剂的编码标签包含切口核酸内切酶序列基序,其在dsDNA的背景下以确定的序列基序被切口核酸内切酶识别和切口。在结合多种结合剂后,使用组合的聚合酶延伸(缺乏链置换活性)+切口核酸内切酶试剂混合物,来产生编码标签向近端记录标签或延伸记录标签的重复传递。在每个传递步骤后,得到的延伸记录标签-编码标签双链体被所述切口核酸内切酶切口,释放附着于所述结合剂的截短间隔子并暴露所述延伸记录标签3'间隔序列,其能够与其它近端结合的结合剂的编码标签退火(图33B-D)。设计编码标签间隔序列中切口基序的位置以产生亚稳态杂交体,其可以容易地与未切割的编码标签间隔序列交

换。以这种方式,如果两种或更多种结合剂同时结合相同的蛋白分子,则通过将来自多个结合的结合剂的编码标签的级联将信息结合到记录标签发送,在单个反应混合物中发生,而没有任何循环试剂交换(图33C-D)。该实施方案特别适用于下一代蛋白测定(NGPA),特别是采用多克隆抗体(或混合单克隆抗体群)到蛋白上的多价表位。

[0546] 对于涉及变性蛋白质、多肽和肽的分析的实施方案,可以在引物延伸后通过使用高度变性条件(例如0.1-0.2N NaOH、6M尿素、2.4M异硫氰酸胍、95%甲酰胺等)除去结合的结合剂和退火的编码标签。

[0547] IX.记录标签信息到编码标签或Di-Tag构建体的循环转移

[0548] 在另一方面,不是在结合剂与大分子结合后将信息从所述编码标签写入所述记录标签,而是可以从包含任选的UMI序列(例如鉴定特定肽或蛋白质分子)和至少一个条形码(例如,隔室标签、分区条形码,样本条形码、空间位置条形码等)的记录标签将信息转移到所述编码标签,从而生成延伸编码标签(参见图11A)。在某些实施方案中,在每个结合循环后,并可选地在Edman降解化学步骤之前,收集所述结合剂和相关的延伸编码标签。在某些实施方案中,所述编码标签包含结合周期特异性标签。在完成所有结合循环后,例如在循环Edman降解中检测NTAA,可以扩增和测序延伸编码标签的完整集合,并且根据UMI(肽身份)、编码序列(NTAA结合剂)、隔室标签(蛋白组的单个细胞或子集)、结合循环特异性序列(循环数)或其任何组合之间的关联确定肽的信息。具有相同隔室标签/UMI序列的文库元件映射回相同的细胞、蛋白组子集、分子等,并可以重建所述肽序列。如果所述Edman降解过程中所述记录标签承受太多损坏,该实施方案可能有用。

[0549] 本发明提供了分析多种大分子的方法,包括:(a)提供多个大分子和与固体支持物连接的相关记录标签;(b)使所述多个大分子与多个能够结合所述多个大分子的结合剂接触,其中每种结合剂包含具有关于所述结合剂的识别信息的编码标签;(c)(i)将与所述大分子相关的记录标签的信息传递到与所述大分子结合的所述结合剂的所述编码标签,以产生延伸编码标签(见图11A);或(ii)将与大分子结合的结合剂的大分子相关的记录标签和编码标签的信息传递给di-tag构建体(参见图11B);(d)收集所述延伸编码标签或di-tag构建体;(e)任选地重复步骤(b)-(d)进行一个或多个结合循环;(f)分析延伸编码标签或di-tag构建体的集合。

[0550] 在某些实施方案中,从所述记录标签到所述编码标签的信息传递可以使用引物延伸步骤完成,其中任选地封闭所述记录标签的3'末端以防止所述记录标签的引物延伸(参见如图11A)。在每个结合事件和信息传递完成之后,可以收集得到的延伸编码标签和相关结合剂。在图11B所示的示例中,所述记录标签包括通用引发位点(U2')、条形码(例如,隔室标签"CT")、可选的UMI序列和共同间隔序列(Sp1)。在某些实施方案中,所述条形码是代表单个隔室的隔室标签,UMI可用于将序列读数映射回被查询的特定蛋白或肽分子。如图11B中的所述实例所示,所述编码标签由共同间隔序列(Sp2')、结合剂编码器序列和通用引发位点(U3)组成。在引入所述编码标签标记的结合剂之前,与所述记录标签的U2'通用引发位点互补并包含通用引发序列U1和循环特异性标签的寡核苷酸(U2),与所述记录标签U2'退火。另外,衔接子序列Sp1'-Sp2与所述记录标签Sp1退火。该衔接子序列还能够与所述编码标签上的Sp2'序列相互作用,使所述记录标签和所述编码标签彼此接近。在结合事件之前或之后进行间隙填充延伸连接测定。如果在结合循环之前进行间隙填充,则使用结合后循

环引物延伸步骤来完成di-tag的构建。在跨多个结合循环收集di-tag后,对di-tag的集合进行测序,并通过UMI序列映射回原始肽分子。应理解,为了最大化功效,所述UMI序列的多样性必须超过由UMI标记的单个分子数量的多样性。

[0551] 在某些实施方案中,所述大分子是蛋白或肽。可以通过从生物样品中片段化蛋白来获得所述肽。

[0552] 所述记录标签可以是DNA分子、RNA分子、PNA分子、BNA分子、XNA分子、LNA分子、γPNA分子或其组合。所述记录标签包含识别与其相关的大分子(例如肽)的UMI。在某些实施方案中,所述记录标签还包含隔室标签。所述记录标签还可以包含通用引发位点,其可以用于下游扩增。在某些实施方案中,所述记录标签在其3'末端包含间隔子。间隔子可以与所述编码标签中的间隔子互补。可以封闭所述记录标签的3'-末端(例如,光不稳定的3'封闭基团)以防止聚合酶延伸所述记录标签,促进所述大分子相关记录标签的信息传递到所述编码标签或所述大分子相关记录标签和编码标签的信息传递到di-tag构建体。

[0553] 所述编码标签包含识别所述编码剂所连接的结合剂的编码器序列。在某些实施方案中,所述编码标签还包含与所述编码标签连接的每种结合剂的独特分子标识符(UMI)。所述编码标签还可以包含通用引发位点,其可以用于下游扩增。所述编码标签可在其3'末端包含间隔子。所述间隔子可以与所述记录标签中的间隔子互补,并可用于引发引物延伸反应以将所述记录标签信息转移到所述编码标签。所述编码标签还可以包含结合循环特异性序列,用于识别延伸编码标签或di-tag起源的结合循环。

[0554] 可以通过引物延伸或连接实现将所述记录标签的信息传递到所述编码标签。可以通过间隙填充反应、引物延伸反应或两者,产生所述记录标签和编码标签的信息至di-tag构建体的传递。

[0555] Di-tag分子包含与延伸记录标签类似的功能组分。Di-tag分子可能包括衍生自所述记录标签的通用引发位点、衍生自所述记录标签的条形码(例如,隔室标签)、衍生自所述记录标签的可选独特分子标识符(UMI)、衍生自所述记录标签的可选间隔子、衍生自所述编码标签的编码序列、衍生自所述编码标签的可选独特分子标识符、结合循环特异性序列、衍生自所述编码标签的可选间隔子以及衍生自所述编码标签的通用引发位点。

[0556] 在某些实施例中,可以使用条形码编码字的组合级联来生成所述记录标签。组合编码词的使用提供了一种方法,通过该方法,退火和化学连接可用于将信息从PNA记录标签传递至编码标签或di-tag构建体(参见如图12A-D)。在分析本文公开的肽的方法涉及通过Edman降解切割末端氨基酸的某些实施方案中,可能需要使用对抗Edman降解的苛刻条件有抗性的记录标签,例如PNA。Edman降解方案中的一个苛刻步骤是无水TFA处理以切割N-末端氨基酸。该步骤通常会破坏DNA。与DNA相比,PNA对酸水解具有高度抗性。PNA的挑战是信息传递的酶促方法变得更加困难,即通过化学连接传递信息是优选模式。在图11B中,使用酶促间隙填充延伸连接步骤编写记录标签和编码标签信息,但是目前这对于PNA模板是不可行的,除非研发了使用PNA的聚合酶。由于需要化学连接,产物不易扩增,所以将条形码和UMI从PNA记录标签写入编码标签是有问题的。化学连接方法已在文献中广泛描述(Gunderson et al.1998,Genome Res.8:1142-1153;Peng et al.,2010,Eur.J.Org.Chem.4194-4197;El-Sagheer et al.,2011,Org.Biomol.Chem.9:232-235;El-Sagheer et al.,2011,Proc.Natl.Acad.Sci.USA 108:11338-11343;Litovchick et al.,

2014,Artif.DNA PNA XNA 5:e27896;Roloff et al.,2014,Methods Mol.Biol.1050:131-141)。

[0557] 为了创建组合PNA条形码和UMI序列,可以组合连接来自n-mer文库的一组PNA词。如果每个PNA字来自1,000个词的空间,则四个组合序列产生$1,000^4=10^{12}$个代码的编码空间。以这种方式,从一组起始的4,000个不同DNA模板序列中,可以产生超过10^{12}个PNA代码(图12A)。通过调整级联字的数量或调整基本字的数量,可以生成更小或更大的编码空间。因此,使用DNA词组装杂交和化学连接,可以使用杂交的DNA序列完成信息到PNA记录标签的传递(参见图12B)。在所述PNA模板上组装DNA单词和化学连接所述DNA字后,所得到的中间体可用于将信息传递到所述编码标签/从编码标签传递信息(参见图12C和图12D)。

[0558] 在某些实施方案中,大分子和相关的记录标签与固体支持物共价连接。固体支持物可以是珠子、阵列、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶片芯片、流通芯片、包括信号转导电子器件的生物芯片、微量滴定孔、ELISA板、旋转干涉测量盘、硝酸纤维素膜、硝化纤维素基聚合物表面、纳米颗粒或微球。固体支持物可以是聚苯乙烯珠子、聚合物珠子、琼脂糖珠子、丙烯酰胺珠子、固体核心珠子、多孔珠子、顺磁珠子、玻璃珠子或可控孔珠子。

[0559] 在某些实施方案中,结合剂是蛋白质或多肽。在一些实施方案中,结合剂是修饰的或变体的氨肽酶、修饰的或变体的氨酰基tRNA合成酶、修饰的或变体的anticalin、修饰的或变体的ClpS或者修饰的或变体的抗体或其结合片段。在某些实施方案中,结合剂结合肽的单个氨基酸残基、二肽、三肽或翻译后修饰。在一些实施方案中,结合剂结合N-末端氨基酸残基,C-末端氨基酸残基或内部氨基酸残基。在一些实施方案中,结合剂结合N-末端肽、C-末端肽或内部肽。在一些实施方案中,结合剂是肽的翻译后修饰的氨基酸的位点特异性共价标记。

[0560] 在某些实施方案中,在步骤(b)中使多个大分子与多种结合剂接触后,将包含大分子和相关结合剂的复合物从固体支持物上解离,并分配成液滴或微流体液滴的乳液。在一些实施方案中,每个微流体液滴包含至多一个包含大分子和结合剂的复合物。

[0561] 在某些实施方案中,在产生延伸编码标签或di-tag构建体之前扩增记录标签。在包含大分子和相关结合剂的复合物被分成液滴或微流体液滴使得每个液滴至多有一个复合物的实施方案中,记录标签的扩增提供额外的记录标签作为将信息传递到编码标签或di-tag构建体的模板(见图13和图14)。乳液融合PCR可用于将记录标签信息传递至编码标签或产生di-tag构建体群。

[0562] 可在分析之前,扩增产生的延伸编码标签或di-tag构建体的集合。对延伸编码标签或di-tag构建体的收集的分析,可能包含核酸测序方法。通过合成测序、通过连接测序、通过杂交测序、通过polony测序、离子半导体测序或焦磷酸测序进行测序核酸测序方法可以是单分子实时测序、基于纳米孔的测序或使用高级显微镜的DNA直接成像。

[0563] Edman降解和化学标记N-末端胺的方法如PITC、Sanger剂(DNFB)、SNFB、乙酰化试剂、脒化(胍基)试剂等,也可以修饰标准核酸或上的内部氨基酸和环外胺或PNA碱基,如腺嘌呤、鸟嘌呤和胞嘧啶。在某些实施方案中,在测序之前,用酸酐、guandination剂或类似的封闭试剂封闭肽的赖氨酸残基的-胺。尽管DNA碱基的环外胺对肽的主要N-末端胺的反应性低得多,但控制胺反应性试剂对N-末端胺的反应性降低了对DNA碱基上的内部氨基酸和环

外胺的非靶活性,这对于测序分析很重要。修饰反应的选择性可以通过调节反应条件来调节,例如pH、溶剂(水溶液与有机物、非质子、非极性、极性非质子、离子液体等)、碱和催化剂、共溶剂、温度和时间。此外,环核胺在DNA碱基上的反应性受DNA是否为ssDNA或dsDNA形式的调节。为了最小化修饰,在NTAA化学修饰之前,记录标签可以与互补DNA探针P1'、{Sample BCs}'、{Sp-BC}'等杂交。在另一个实施方案中,也可以使用具有受保护的环外胺的核酸(Ohkubo,Kasuya et al.2008)。在另一个实施方案中,"反应性较低"的胺标记化合物,例如SNFB,会减轻DNA上的内部氨基酸和外环胺的脱靶标记(Carty and Hirs 1968)。由于所述对磺酰基从对硝基吸走电子,导致SNFB的氟取代活性低于DNFB,所以SNFB的反应性低于DNFB。

[0564] 通过仔细选择化学和反应条件(浓度、温度、时间、pH、溶剂类型等),可以滴定偶联条件和偶联试剂,以优化NTAA-胺修饰并最大限度地减少脱靶氨基酸修饰或DNA修饰。例如,已知DNFB在非质子溶剂如乙腈中比在水中更容易与仲胺反应。环外胺的温和修饰仍可允许互补探针与序列杂交,但可能破坏基于聚合酶的引物延伸。还可能保护外酯胺的同时仍然允许氢键结合。这在最近的出版物中有所描述,其中受保护的碱基仍然能够与感兴趣的靶标杂交(Ohkubo,Kasuya et al.2008)。在一个实施方案中,工程化聚合酶用于在所述延伸记录标签期间将核苷酸和受保护碱基并入到DNA编码标签模板上。在另一个实施方案中,工程化聚合酶用于在编码标签的延伸期间将PNA模板(w/或w/o受保护碱基)上的核苷酸并入PNA记录标签模板。在另一个实施方案中,通过将外源寡核苷酸与PNA记录标签退火,可以将信息从记录标签传递到编码标签。通过选择序列空间不同的UMI可以促进杂交的特异性,例如基于n-mer词组装的设计(Gerry,Witowski et al.1999)。

[0565] 虽然可以使用类似Edman的N-末端肽降解测序来确定肽的线性氨基酸序列,但可以使用替代实施方案利用延伸记录标签、延伸编码标签和di-tag方法对肽进行部分组成分析。结合剂或化学标记可用于标记肽上的N-末端和内部氨基酸或氨基酸修饰。化学试剂可以以位点特异性方式共价修饰氨基酸(例如标记)(Sletten and Bertozzi 2009,Basle,Joubert et al.2010)(Spicer and Davis 2014)。编码标签可以附着到靶向单个氨基酸的化学标记试剂上,以促进编码和随后位点特异性标记的氨基酸的识别(参见图13)。

[0566] 肽组成分析不要求循环降解所述肽,因此避免了将含有标签的DNA暴露于苛刻的Edman化学的问题。在循环结合模式中,还可以使用延伸编码标签或di-tag来提供组成信息(氨基酸或二肽/三肽信息)、PTM信息和初始氨基酸序列。在一个实施例中,可以使用本文描述的延伸编码标签或di-tag方法来读出该组成信息。如果与UMI和隔室标签信息结合,延伸编码标签或di-tag的集合提供关于肽及其起源隔室蛋白或起源蛋白的组成信息。延伸编码标签或di-tag的集合映射回相同的隔室标签(并且表面上来源于蛋白分子),是利用部分组成信息来映射肽的有力工具。隔室标记肽的集合不是映射回整个蛋白组,而是映射回蛋白分子的有限子集,大大增加了映射的独特性。

[0567] 本文使用的结合剂可识别单个氨基酸、二肽、三肽或甚至更长的肽序列基序。Tessler(2011,Digital Protein Analysis:Technologies for Protein Diagnostics and Proteomics through Single Molecule Detection.Ph.D.,Washington University in St.Louis)证明,可以带电二肽表位子集生成相对选择性的二肽抗体(Tessler 2011)。将定向进化应用于替代蛋白支架(例如,aaRS、anticalin、ClpS等),适体可用于延伸一组二

肽/三肽结合剂。来自二肽/三肽组成分析的信息与映射回单个蛋白分子的信息,可足以唯一地识别和定量每个蛋白分子。最多,总共有400种可能的二肽组合。然而,最频繁和最具抗原性(带电荷、亲水性、疏水性)二肽的子集应该足以产生结合剂。该数字可以构成一组40-100种不同的结合剂。对于一组40种不同的结合剂,平均10-mer肽有约80%被至少一种结合剂结合的机会。将该信息与衍生自相同蛋白分子的所有肽组合,可以允许识别蛋白分子。关于肽及其起源蛋白的所有这些信息可以组合起来,提供更准确和精确的蛋白序列表征。

[0568] 最近提出了一种使用部分肽序列信息的数字蛋白表征测定法(Swaminathan et al.,2015,PLoS Comput.Biol.11:e1004080)(Yao,Docter et al.2015)。即,该方法采用氨基酸的荧光标记,其易于使用标准化学标记,例如半胱氨酸、赖氨酸、精氨酸、酪氨酸、天冬氨酸/谷氨酸(Basle,Joubert etal.2010)。部分肽序列信息的挑战是映射回蛋白组的映射是一对多的关联,无法标识独特蛋白。这种一对多的映射问题可以通过将整个蛋白组空间减少到肽被映射回的蛋白分子的有限子集来解决。实质上,单个部分肽序列可以映射回100或1000个不同的蛋白序列,但是如果已知一组几个肽(例如,来自单个蛋白分子消化物的10个肽)全部映射回隔室内蛋白分子子集中的单个蛋白分子,则更容易推断出蛋白质分子的身份。例如,来自相同分子的所有肽的肽蛋白质组图谱的交叉点,极大地限制了可能的蛋白身份集(参见图15)。

[0569] 特别地,通过创造性地使用隔室标签和UMI,显著增强了部分肽序列或组合物的可映射性。即,所述蛋白组最初被分隔到条形码化的隔室,其中所述隔室条形码也附着于UMI序列。所述隔室条形码是所述隔室独有的序列,所述UMI是所述隔室内每个条形码分子独有的序列(参见图16)。在一个实施方案中,使用类似于PCT公开WO2016/061517(通过引用全文并入)中公开的方法完成该分区,即通过杂交到附着于所述珠子的DNA隔室条形码杂交,使DNA标签标记的多肽与珠子表面直接相互作用(见图31)。引物延伸步骤将信息从珠子连接的隔室条形码传递到多肽上的DNA标签(图20)。在另一个实施方案中,通过将含有UMI的条形码珠子和蛋白分子共包封成乳液液滴来完成这种分区。此外,所述液滴可选地含有将蛋白消化成肽的蛋白酶。许多蛋白酶可用于消化所述报告标记的多肽(Switzar,Giera et al.2013)。酶促连接酶(例如butelase I)与蛋白酶的共包封可能需要对所述酶进行修饰,例如聚乙二醇化,以使其对蛋白酶消化具有抗性(Frokjaer and Otzen 2005,Kang,Wang et al.2010)。消化后,所述肽会连接到所述条形码-UMI标签上。在优选实施方案中,所述条形码-UMI标签保留在珠子上以促进下游生物化学操作(参见图13)。

[0570] 在条形码-UMI与所述肽连接后,破坏乳液并收获珠子。所述条形码化的肽可以通过它们的初始氨基酸序列或它们的氨基酸组成来表征。所述肽的两种类型的信息都可用于将其映射回所述蛋白组的子集。通常,序列信息映射回所述蛋白组的更小子集而不是组成信息。尽管如此,通过将来自多个肽(序列或组合物)的信息与相同的隔室条形码组合,可以唯一地识别所述蛋白或所述肽的来源蛋白。以这种方式,可以表征和定量所述整个蛋白组。所述肽的初始序列信息可以通过进行肽测序反应来获得,其中延伸记录标签产生代表肽序列的DNA编码文库(DEL)。在优选实施例中,记录标签包含隔室条形码和UMI序列。该信息与从编码标签传递的初级或PTM氨基酸信息一起使用,以产生最终的映射肽信息。

[0571] 肽序列信息的替代方案是产生与隔室条形码和UMI连接的肽氨基酸或二肽/三肽组成信息。这是通过对具有UMI条形码肽的珠子进行氨基酸标记步骤来实现的,其中每个肽

上的选出氨基酸(内部)用包含氨基酸编码信息的DNA标签和另一个氨基酸UMI进行位点特异性标记(AA UMI)(见图13)。最易于化学标记的氨基酸(AAs)是赖氨酸、精氨酸、半胱氨酸、酪氨酸、色氨酸和天冬氨酸/谷氨酸,但为其他AA开发标记方案也是可行的(Mendoza and Vachet,2009)。给定的肽可包含几种相同类型的AA。可以通过所属附着的AA UMI标记区分相同类型的多个氨基酸的存在。每个标记分子在所述DNA标签内都有不同的UMI,使得能够计数氨基酸。化学标记的替代方案是用结合剂"标记"所述AA。例如,可以使用用包含AA和AA UMI的编码信息的编码标签标记的酪氨酸特异性抗体标记所述肽的所有酪氨酸。使用这种方法的警告是大体积抗体遇到的空间位阻,理想情况下较小的scFv,anticalins或ClpS变体将用于此目的。

[0572] 在一个实施方案中,在标记AA之后,通过区室化所述肽复合物使得每个液滴含有单个肽并进行乳液融合PCR以构建一组表征所述区室化的肽的氨基酸组成的延伸编码标签或di-tags,信息在记录标签和与肽上结合或共价偶联的结合剂相关的多个编码标签之间转移。对所述di-tags进行测序后,可以将具有相同条形码的肽的信息映射回单个蛋白质分子。

[0573] 在一个具体实施方案中,所述标记的肽复合物与所述珠子解离(参见图13),分区成小迷你隔室(例如,微乳液),使得给定的隔室内平均仅有一种标记的/结合的结合剂肽复合物。在一个具体实施方案中,通过产生微乳液液滴来实现这种区室化(Shim,Ranasinghe等人,2013,Shembekar,Chaipan等人,2016)。除所述肽复合物外,PCR试剂与三种引物(U1,Sp和U2tr)也一起共同包封在所述液滴中。在液滴形成后,在较高的退火温度下进行几个循环的乳液PCR(约5-10个循环),例如仅使U1和Sp退火并扩增所述记录标签产物(参见图13)。在最初的5-10个PCR循环之后,降低所述退火温度,使得U2tr和所述氨基酸编码标签上的Sptr参与扩增,并进行另外~10轮。所述三引物乳液PCR有效地将所述肽UMI条形码与所有所述AA编码标签组合,产生代表所述肽及其氨基酸组成的di-tag库。也可以进行所述三引物PCR和所述标签级联的其他方式。另一个实施方案是使用通过光解封闭活化的3'封闭的U2引物,或添加油溶性还原剂以启动不稳定封闭的3'核苷酸的3'解封闭。后乳液PCR,可以使用通用引物进行另一轮PCR以格式化用于NGS测序的所述库元件。

[0574] 以这种方式,所述库元件的不同序列组件用于计数和分类目的。对于给定的肽(由隔室条形码-UMI组合标识),存在许多库元件,每个库元件具有识别AA编码标签和AA UMI(参见图13)。所述AA代编码和相关的UMI用于计算给定肽中给定氨基酸类型的出现。因此,所肽(可能是GluC,LysC或Endo AsnN消化物)的特征在于其氨基酸组成(例如,2个Cys,1个Lys,1个Arg,2个Tyr等),而与空间次序无关。尽管如此,这提供了足够的特征以将肽映射到所述蛋白质组的子集,并且当与衍生自相同蛋白质分子的其他肽组合使用时,唯一地识别和定量所述蛋白质。

[0575] X.末端氨基酸(TAA)标记方法

[0576] 在某些实施方案中,在本文所述的方法中,使肽与结合剂接触之前修饰或标记所述肽的末端氨基酸(例如,NTAA或CTAA)。

[0577] 在一些实施方案中,所述NTAA与异硫氰酸苯酯(PITC)反应以产生苯基硫代氨基甲酰基(PTC)-NTAA衍生物。Edman降解通常使用苯基异硫氰酸酯(PITC)来标记所述N-末端。PITC具有两种非常适合于本公开的方法的性质:(1)PITC高效地标记所述N-末端胺基;(2)

所得PTC衍生的NTAA在酸处理时经历自身异构化,导致剩余肽中的氨基酸裂解。

[0578] 可用于标记所述NTAA的其它试剂包括:4-磺基苯基异硫氰酸酯、3-吡啶基异硫氰酸酯(PYITC)、2-哌啶子基乙基异硫氰酸酯(PEITC)、3-(4-吗啉代)丙基异硫氰酸酯(MPITC)、3-(二乙氨基)异硫氰酸丙酯(DEPTIC)(Wang etal.,2009,Anal Chem 81:1893-1900)、(1-氟-2,4-二硝基苯(Sanger试剂,DNFB)、丹磺酰氯(DNS-Cl或1-二甲基氨基萘-5-磺酰氯)、4-磺酰基-2-硝基氟苯(SNFB)、乙酰化试剂、脒化(胍基)试剂,2-羧基-4,6-二硝基氯苯、7-甲氧基香豆素乙酸,硫代酰化试剂、硫代乙酰化试剂和硫代苄基化试剂。如果所述NTAA被封闭而无法标记,则有许多方法可以解封所述末端,例如用酰基肽水解酶(APH)去除N-乙酰基封闭(Farries,Harris et al.,1991,Eur.J.Biochem.196:679-685)。解封肽N-末端的方法是本领域已知的(参阅例如see,e.g.,Krishna et al.,1991,Anal.Biochem.199:45-50;Leone et al.,2011,Curr.Protoc.Protein Sci.,Chapter 11:Unit11.7;Fowler et al.,2001,Curr.Protoc.Protein Sci.,Chapter 11:Unit 11.7,其中每篇都通过引用整体并入本文)。

[0579] 丹磺酰氯与肽的游离胺基反应,得到所述NTAA的丹酰基衍生物。DNFB和SNFB与肽的-胺基反应分别产生DNP-NTAA和SNP-NTAA。另外,DNFB和SNFB也都与赖氨酸残基的-胺反应。DNFB还与酪氨酸和组氨酸氨基酸残基反应。SNFB比DNFB对胺基具有更好的的选择性,并且优选用于NTAA修饰(Carty和Hirs 1968)。在某些实施方案中,在多肽蛋白酶消化成肽之前,赖氨酸-胺用有机酸酐预封闭。

[0580] 另一种有用的NTAA修饰剂是乙酰基,因为存在已知的酶以去除乙酰化的NTAA,即酰基肽水解酶(APH),其切割所述N-末端乙酰化氨基酸,有效地以单个氨基酸缩短所述肽{Chang,2015#373;Friedmann,2013#374}。所述NTAA可用乙酸酐进行化学乙酰化或用N-末端乙酰转移酶(NAT)酶促乙酰化{Chang,2015#373;Friedmann,2013#374}。另一种有用的NTAA改性剂是脒基(胍基)部分,因为已证实的酰胺化NTAA的裂解化学在文献中是已知的,即N-末端酰胺化肽与0.5-2％NaOH的温和温育导致所述N-末端氨基酸的裂解{Hamada,2016#383}。这有效地提供了温和的Edman样化学N-末端降解肽测序过程。此外,某些脒化(胍基)试剂和下游NaOH裂解与DNA编码十分兼容。

[0581] NTAA上DNP/SNP、乙酰基或脒基(胍基)基团的存在可以提供与工程化结合剂相互作用的更好操作。存在许多具有低nM亲和力的商业DNP抗体。标记NTAA的其它方法包括用trypigase(Liebscher et al.,2014,Angew Chem Int Ed Engl 53:3024-3028)和氨基酰基转移酶(Wagner,et al.,2011,J Am Chem Soc 133:15139-15147)标记。

[0582] 在离子液体存在下,异硫氰酸酯已显示出对伯胺的增强的反应性。离子液体是有机化学反应中的优异溶剂(并且用作催化剂),并且可以增强异硫氰酸酯与胺反应以形成硫脲。一个例子是使用离子液体1-丁基-3-甲基咪唑四氟硼酸盐[Bmim][BF4]通过异硫氰酸苯酯(PITC)快速有效地标记芳香族和脂肪胺(Le,Chen et al.2005)。Edman降解涉及所述异硫氰酸酯(如PITC)与肽的氨基N-末端的反应。因此,在一个实施方案中,离子液体通过提供更温和的标记和降解条件用于改善所述Edman降解过程的效率。例如,在25℃下,在离子液体[Bmim][BF4]中使用5％(vol./vol.)PITC 10分钟。比在标准Edman PITC衍生化条件下标记更有效,标准条件是在55℃,在含有吡啶、乙醇和ddH_2O(1:1:1vol./vol./vol.)的溶液中使用5％(vol./vol.)PITCC 60分钟(Wang,Fang et al.2009)。在优选的实施方案中,在片

段化成肽之前,内部赖氨酸、酪氨酸、组氨酸和半胱氨酸氨基酸在所述多肽内被封闭。这样,在所述肽测序反应期间,只有所述NTAA的所述肽-胺基易于修饰。当使用DNFB(Sanger'试剂)和丹磺酰氯时,这尤其重要。

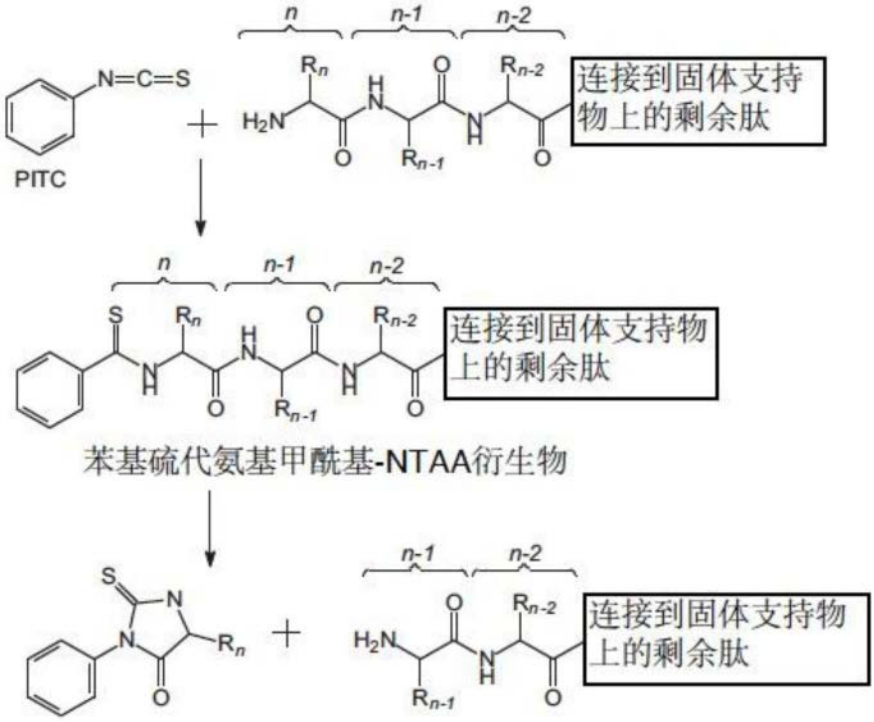
[0583] 在某些实施方案中,所述NTAA在NTAA标记步骤(特别是蛋白的原始N-末端)之前已被封闭。如果是这样,有许多方法可以解除N末端的封闭,例如用酰基肽水解酶(APH)去除N-乙酰基封闭(Farries,Harris et al.1991)。许多其它解封所述肽N-末端的方法是本领域已知的(参阅,例如,Krishna et al.,1991,Anal.Biochem.199:45-50;Leone et al.,2011,Curr.Protoc.Protein Sci.,Chapter 11:Unit11.7;Fowler et al.,2001,Curr.Protoc.Protein Sci.,Chapter 11:Unit 11.7,其各自通过引用整体并入本文)。

[0584] 所述CTAA可以用许多不同的羧基反应试剂进行修饰。在另一个实例中,用混合酸酐和异硫氰酸酯修饰所述CTAA以产生硫代乙内酰脲((Liu and Liang 2001)and U.S.Patent No.5,049,507)。所述硫代乙内酰脲修饰的肽可以在升高的温度下在碱中裂解以暴露倒数第二个CTAA,有效地产生基于C末端的肽降解测序方法(Liu和Liang 2001)。可以对CTAA进行的其他修饰包括添加对硝基苯胺基团和添加7-氨基-4-甲基香豆酰基。

[0585] XI.末端氨基酸裂解方法

[0586] 在涉及分析肽的某些实施方案中,在通过结合剂的末端氨基酸(N-末端或C-末端)结合和编码标签信息转移至记录标签、将记录标签信息转移至编码标签、标签信息和编码标签信息转移到di-tag构建体中之后,从所述肽中除去或切割所述末端氨基酸以暴露新的末端氨基酸。在一些实施方案中,所述末端氨基酸是NTAA。在其它实施方案中,所述末端氨基酸是CTAA。

[0587] 末端氨基酸的切割可以通过许多已知技术完成,包括化学切割和酶促切割。化学切割的一个例子是Edman降解。在Edman降解肽期间,所述nNTAA在温和的碱性条件下与异硫氰酸苯酯(PITC)反应,形成所述苯基硫代氨基甲酰基-NTAA衍生物。接下来,在酸性条件下,所述苯基硫代氨基甲酰基-NTAA衍生物被切割,产生游离的噻唑啉酮衍生物,从而将肽的所述n-1氨基酸转化为N-末端氨基酸(n-1NTAA)。此过程中的步骤如下所示:

[0588] 

[0589] 如上所述,典型的Edman降解需要在长的温育时间内施加苛刻的高温化学条件(例如,无水TFA)。这些条件通常与编码大分子的核酸不兼容。

[0590] 为了将化学Edman降解转化为核酸编码友好的方法,苛刻的化学步骤被温和的化学降解或有效的酶促步骤取代。在一个实施方案中,化学Edman降解可以使用比原始描述的更温和的条件。Edman降解的几种较温和的裂解条件已在文献中描述,包括用乙腈中的三乙胺乙酸酯代替无水TFA(参阅,例如:Barrett,1985,Tetrahedron Lett.26:4375-4378,其全部内容通过引用并入)。所述NTAA的切割也可以使用硫酰化降解来完成,与Edman降解相比,其使用较温和的切割条件(见美国专利4,863,870)。

[0591] 在另一个实施方案中,无水TFA的切割可以用“Edmanase”代替,其是一种工程酶,其通过硫脲硫原子对在温和条件下易断键的羰基的亲核攻击来催化去除所述PITC衍生的N-末端氨基酸(参阅,U.S.Patent Publication US2014/0273004,通过引用整体并入本文)。Edmanase是通过修饰cruzain,来自Trypanosoma cruzi的半胱氨酸蛋白酶(Borgo,2014),来制备。C25G突变除去催化半胱氨酸残基,同时选择三个突变(G65S、A138C、L160Y)以与Edman试剂(PITC)的苯基基团产生空间配合。

[0592] NTAA的酶促切割也可以通过氨肽酶完成。氨肽酶以单体和多聚体酶天然存在,并且可以是金属或ATP依赖性的。天然氨肽酶具有非常有限的特异性,并且以逐步方式一个接一个地切割N-末端氨基酸。对于本文所述的方法,氨肽酶可以被工程改造为具有对仅当被N末端标签修饰时的所述NTAA的特异性结合或催化活性。例如,氨基肽酶可以被工程改造,以使其仅切割例如DNP/SNP、PTC、丹磺酰氯、乙酰基、脒基等基团修饰的N-末端氨基酸。以这种方式,氨肽酶一次仅从N末端切割单个氨基酸,并允许控制所述降解循环。在一些实施方案中,所述修饰的氨肽酶对氨基酸残基身份是非选择性的,同时对所述N-末端标签具有选择性。在其他实施方案中,所述修饰的氨肽酶对氨基酸残基身份和N-末端标签二者都具有选择性。Borgo和Havranek说明了修饰酶促NTAA降解的特异性的模型的实例,其中通过结构-

功能辅助设计,将蛋氨酸氨肽酶转化为亮氨酸氨肽酶(Borgo和Havranek 2014)。可以采用类似的方法用改良的NTAA,例如DNP/SNP修饰的NTAA,其中氨基肽酶被工程化(使用基于结构-功能的设计和定向进化)仅切割存在DNP/SNP基团的N-末端氨基酸。结合并切割标记的(生物素化的)NTAA的单个的或小的基团的工程化氨肽酶突变体已经有人描述(参见PCT公开号WO2010/06532)。

[0593] 在某些实施方案中,紧密的单体金属酶促氨肽酶被工程化以识别和切割DNP标记的NTAA。所述单体金属氨肽酶的使用有两个关键优点:1)紧密的单体蛋白更容易使用噬菌体展示显示和筛选;2)金属氨基肽酶具有独特的优势,即其活性可以通过添加或除去适当的金属阳离子随意打开/关闭。示例性的氨肽酶包括M28家族的氨肽酶,例如Streptomyces sp.KK506(SKAP)(Yoo,Ahn et al.2010)、Streptomyces griseus(SGAP)、Vibrio proteolyticus(VPAP),(Spungin and Blumberg 1989,Ben-Meir,Spungin et al.1993)。这些酶在室温和pH 8.0下稳定,健壮、且有活性,因此与喜欢温和条件的肽分析相容。

[0594] 在另一个实施方案中,通过将所述氨肽酶工程化为仅在N-末端氨基酸标记存在有活性来实现环状切割。此外,所述氨肽酶可以被工程化为非特异性的,使得它不会选择性地识别一种特定氨基酸而非另一种特定氨基酸,而是仅识别所述标记的N-末端。在一个优选的实施方案中,金属肽酶单体氨肽酶(例如Vibro亮氨酸氨肽酶)(Hernandez-Moreno,Villasenor et al.2014)被工程化为仅裂解修饰的NTAA(例如PTC、DNP、SNP、乙酰化、酰化等)。

[0595] 在另一个实施方案中,通过使用工程化的酰基肽水解酶(APH)切割乙酰化的NTAA来实现环状切割。APH是丝氨酸肽酶,其能够从封闭的肽中催化去除Nα-乙酰化氨基酸,并且是真核细胞、细菌和古细胞中N末端乙酰化蛋白的关键调节剂。在某些实施方案中,APH是二聚体并且仅具有外肽酶活性(Gogliettino,Balestrieri et al.2012,Gogliettino,Riccio et al.2014)。所述工程化APH可具有比内源或野生型APH更高的亲和力和更低的选择性。

[0596] 在另一个实施方案中,所述NTAA的脒化(胍基化)采用来促使温和切割所述标记的NTAA(Hamada,2016,其通过引用整体并入)。本领域已知许多脒化(胍基化)试剂,包括:S-甲基异硫脲、3,5-二甲基吡唑-1-甲脒、S-乙基硫脲溴化物、S-乙基硫脲氯化物、O-甲基异脲、O-甲基二异脲硫酸盐、O-甲基异脲硫酸氢盐、2-甲基-1-硝基异脲、氨基亚氨基甲磺酸、氰胺、氰基胍、双氰胺、3,5-二甲基-1-脒基吡唑硝酸盐、和3,5-二甲基吡唑、N,N'-双(邻-氯-Cbz)-S-甲基异硫脲、和N,N'-双(邻-溴-Cbz)-S-甲基异硫脲(Katritzky,2005,通过引用整体并入)。

[0597] NTAA标记、结合和降解工作流程的示例如下(参见图41和42):来自蛋白水解消化的大量记录标签标记肽(例如,5000万至10亿)被以合适的分子内间距随机固定在单个分子测序基质上(例如,多孔珠)。以循环方式,每个肽的N-末端氨基酸(NTAA)用小化学基团(例如,DNP、SNP、乙酰基)修饰,以提供对NTAA降解过程的循环控制,并通过同源结合剂增强结合亲和力。每个固定化肽的修饰的N-末端氨基酸(例如,DNP-NTAA、SNP-NTAA、乙酰基-NTAA)被所述同源NTAA结合剂结合,并且与结合的NTAA结合剂相关的编码标签的信息被转移到与所述固定的肽相关的所述记录标签。在NTAA识别,结合和将编码标签信息转移至记录标签后,通过暴露于工程化氨肽酶(例如,用于DNP-NTAA或SNP-NTAA)或工程化APH(例如,用于乙

酰-NTAA)来去除标记的NTAA。仅在标记存在下才能进行NTAA切割。其他NTAA标记(例如,PITC)也可以采用来与适当工程化的氨肽酶一起使用。在一个具体实施方案中,单一工程化氨肽酶或APH普遍切割具有N-末端氨基酸标记的所有可能的NTAA(包括翻译后修饰变体)。在另一个具体的实施方案中,使用两种、三种、四种或更多种工程化的氨肽酶或APH来切割标记的NTAA的库。

[0598] 对DNP或SNP标记的NTAA具有活性的氨肽酶可以被选择:通过使用结合对apo-酶(在没有金属辅因子的情况下无活性)的紧密结合选择的筛选,然后进行功能性催化选择步骤,如Ponsard等人描述的工程化用于苄青霉素的金属-β-内酰胺酶方法(Ponsard,Galleni et al.2001,Fernandez-Gacio,Uguen et al.2003)。这种两步法选择涉及使用通过添加Zn^{2+}离子活化的金属-AP。在紧密结合选择固定的肽基质后,引入Zn^{2+},并且催化激活能够水解DNP或SNP标记的NTAA的噬菌体以释放结合的噬菌体到上清液中。进行重复选择循环为DNP或SNP标记的NTAA切割富集活性AP。

[0599] 在本文提供的任何实施方案中,NTAA裂解试剂向所述NTAA的募集可以通过嵌合裂解酶和嵌合NTAA修饰剂增强,其中所述嵌合裂解酶和嵌合NTAA修饰剂各自包含能够相互紧密结合的基团(例如,生物素-链霉亲和素)(参见图39)。例如,可以用生物素-PITC修饰NTAA,并且通过链霉亲和素-生物素相互作用将嵌合切割酶(链霉亲和素-Edmanase)募集到所述修饰的NTAA,从而提高所述切割酶的亲和力和效率。所述修饰的NTAA被切割并随着相关的切割酶一起从所述肽中扩散出来。在嵌合Edmanase的例子中,该方法有效地将亲和力K_D从μM增强至亚皮摩尔。通过使用与所述记录标签相互作用的裂解剂上的DNA标签进行束缚,也可以实现类似的切割增强(参见图44)。

[0600] 作为NTAA切割的替代方案,二肽基氨基肽酶(DAP)可用于从肽上切割最后两个N-末端氨基酸。在某些实施方案中,可以切割单个NTAA(参见图45):图45描绘了N-末端降解的方法,其中butelaseI肽底物的N-末端连接将TEV内肽酶底物附着至所述肽的N末端。附着后,TEV内肽酶从所述测试肽(经历测序的肽)切割新连接的肽,留下与所述NTAA连接的单个天冬酰胺(N)。与DAP孵育,其从N末端切割两个氨基酸,导致所述原始NTAA的净去除。整个过程可以在N端降解过程中循环。

[0601] 对于涉及CTAA结合剂的实施方案,从肽切割CTAA的方法也是本领域已知的。例如,美国专利6,046,053公开了一种使肽或蛋白质与烷基酸酐反应以将羧基末端转化为恶唑酮,通过与酸和醇或与酯反应释放所述C-末端氨基酸的方法。CTAA的酶促切割也可以通过羧肽酶完成。几种羧肽酶表现出氨基酸偏好,例如,羧肽酶B优先在碱性氨基酸如精氨酸和赖氨酸上切割。如上所述,羧肽酶也可以与氨肽酶相同的方式进行修饰,以改造成特异性结合具有C-末端标记的CTAA的工程羧肽酶。以这种方式,羧肽酶一次仅从所述C末端切割单个氨基酸,并允许控制所述降解循环。在一些实施方案中,所述修饰的羧肽酶对氨基酸残基身份是非选择性的,同时对C-末端标记具有选择性。在其他实施方案中,修饰的羧肽酶对氨基酸残基身份和C末端标记都具有选择性。

[0602] XII.延伸记录标签,延伸编码标签或Di-Tag的处理和分析

[0603] 可以使用多种核酸测序方法处理和分析延伸记录标签,延伸编码标签和代表感兴趣的大分子的di-tag文库。测序方法的例子包括,但不限于,链终止测序(Sanger测序);下一代测序方法,如合成测序、连接测序、杂交测序、polony测序、离子半导体测序、和焦磷酸

测序;和第三代测序方法,如单分子实时测序、基于纳米孔的测序、双重中断测序、以及使用高级显微镜的DNA直接成像。

[0604] 可以以各种方式扩增延伸记录标签、延伸编码标签或di-tags的库。延伸记录标签、延伸编码标签或di-tags的库可以进行指数扩增,例如通过PCR或乳液PCR。已知乳液PCR产生更均匀的扩增(Hori,Fukano et al.2007)。或者,延伸记录标签、延伸编码标签或di-tags的库可以进行线性扩增,例如,通过使用T7 RNA聚合酶的模板DNA的体外转录。可以使用与其中包含的通用正向引发位点和通用反向引发位点相容的引物扩增所述延伸记录标签、延伸编码标签或di-tags的库。所述延伸记录标签、延伸编码标签或di-tags的库也可以使用有尾引物进行扩增,以将序列添加到所述延伸记录标签、延伸编码标签或di-tags的5'端、3'端或两端。可以添加到所述延伸记录标签、延伸编码标签或di-tags末端的序列包括与用于制作测序平台兼容的所述延伸记录标签、延伸编码标签或di-tags的:允许在单个测序运行中复用的多个文库的库特异性索引序列、衔接子序列、读取引物序列或任何其他序列。制备下一代测序的文库扩增的实例如下:使用从~1mg珠子(~10ng)洗脱的延伸记录标签库,200μMdNTP,正向和反向扩增引物每个1μM,0.5μl(1U)Phusion热启动酶(New England Biolabs),并进行以下循环条件:98℃30秒,然后98℃10秒20个循环,60℃持续30秒,72℃持续30秒,然后是72℃持续7分钟,然后保持在4℃。

[0605] 在某些实施方案中,在扩增之前,期间或之后,所述延伸记录标签、延伸编码标签或di-tags库可以进行靶富集。靶富集可用于在测序之前从所述延伸记录标签、延伸编码标签或di-tags库中选择性地捕获或扩增代表感兴趣大分子的延伸记录标签。由于高成本和难以产生针对靶蛋白的高度特异性结合剂,蛋白序列的靶标富集是具有挑战性的。众所周知,抗体是非特异性的,并且难以跨越数千种蛋白质上规模化生产。本公开的方法通过将所述蛋白代码转换成核酸代码来规避了该问题,该核酸代码然后可以利用可用于DNA文库的多种靶向DNA富集策略。通过富集它们相应的延伸记录标签,可以在样品中富集感兴趣的肽。靶向富集的方法是本领域已知的,并且包括杂交捕获测定、基于PCR的测定,例如TruSeq定制Amplicon(Illumina),挂锁探针(也称为分子倒置探针),及其类似方法(see,Mamanova et al.,2010,Nature Methods 7:111-118;Bodi et al.,J.Biomol.Tech.2013,24:73-86;Ballester et al.,2016,Expert Review of Molecular Diagnostics357-372;Mertes et al.,2011,Brief Funct.Genomics 10:374-386;Nilsson etal.,1994,Science 265:2085-8;每个都通过引用整体并入本文。

[0606] 在一个实施方案中,通过基于杂交捕获的测定来富集延伸记录标签、延伸编码标签或di-tags库(参见例如图17A和17B)。在基于杂交捕获的测定中,所述延伸记录标签、延伸编码标签或di-tags库与靶标特异性寡核苷酸或用亲和标签(例如生物素)标记的"诱饵寡核苷酸"杂交。通过亲和配体(例如,链霉亲和素包被的珠子)和它们的亲和标签,杂交到所述靶特异性寡核苷酸的延伸记录标签、延伸编码标签或di-tags被"拉下",并且和背景(非特异性)延伸记录标签被洗掉(参见例如图17)。然后获得用于阳性富集(例如,从珠子洗脱)富集的延伸记录标签、延伸编码标签或di-tags。

[0607] 对于通过基于阵列的"原位"寡核苷酸合成和随后的寡核苷酸库扩增合成的诱饵寡核苷酸,可以通过在给定的寡核苷酸阵列内使用几组通用引物将竞争诱饵工程化到池中。对于每种类型的通用引物,生物素化引物与非生物素化引物的比例控制富集率。几种引

物类型的使用使得能够将几种富集比设计到最终的寡核苷酸诱饵池中。

[0608] 诱饵寡核苷酸可以设计成与代表感兴趣的大分子的延伸记录标签、延伸编码标签或di-tags互补。诱饵寡核苷酸与延伸记录标签、延伸编码标签或di-tags上的间隔序列的互补程度可以是0％至100％,以及其间的任何整数。通过一些富集实验可以很容易地优化该参数。在一些实施方案中,所述间隔子相对于所述编码序列的长度在编码标签设计中被最小化,或者间隔子被设计为不可与所述诱饵序列杂交。一种方法是使用在辅助因子存在下形成二级结构的间隔子。这种二级结构的一个例子是G-quadruplex,它是由两个或多个彼此堆叠的鸟嘌呤四联体形成的结构(Bochman,Paeschke et al.2012)。鸟嘌呤四联体是由四个鸟嘌呤碱基形成的方形平面结构,其通过Hoogsteen氢键结合。所述G-quadruplex结构在阳离子存在下稳定,例如K+离子vs.Li离子。

[0609] 为了使所用的诱饵寡核苷酸的数量最小化,可以生物信息学地鉴定来自每种蛋白质的一组相对独特的肽,并且将仅与代表目标肽的相应延伸记录标签库互补的那些诱饵寡核苷酸用于杂交捕获检测。也可以使用相同或不同的诱饵组进行连续回合或富集。

[0610] 为了在代表其片段(例如肽)的延伸记录标签、延伸编码标签或di-tags库中富集全长大分子(例如,蛋白质或多肽),可以设计“平铺的”诱饵寡核苷酸以覆盖代表所述蛋白的完整核酸。

[0611] 在另一个实施方案中,引物延伸和基于连接介导的扩增富集(AmpliSeq、PCR、TruSeq TSCA等)可用于选择和调节(module)代表大分子子集的文库元件的富集的分级物。竞争oligos也可用于调节引物延伸、连接或扩增的程度。在最简单的实施方案中,这可以通过混合包含通用引物尾的靶特异性引物和缺乏5’通用引物尾的竞争引物来实现。在初始引物延伸后,仅可扩增具有5’通用引物序列的引物。具有和不具有通用引物序列的引物的比例控制扩增的靶标的分数。在其它实施方案中,用包含的除延伸引物以外的杂交来调节经历引物延伸,连接或扩增的文库元件的分数。

[0612] 靶富集方法也可以以用在负选择模式中,以在测序之前从文库中选择性地去除延伸记录标签、延伸编码标签或di-tag。因此,在上述使用生物素化诱饵寡核苷酸和链霉亲和素包被的珠子的实施例中,保留上清液用于测序,而不分析结合在珠子上的所述诱饵寡核苷酸:延伸记录标签、延伸编码标签或di-tag二标签杂交体。可以去除的不需要的延伸记录标签、延伸编码标签或di-tag的例子是那些代表丰度大分子种类例如蛋白质,白蛋白,免疫球蛋白等的延伸记录标签、延伸编码标签或di-tag。

[0613] 与所述靶标杂交但缺乏生物素基团的竞争寡核苷酸诱饵也可用于杂交捕获步骤以调节富集的任何特定位点的分数。竞争寡核苷酸诱饵与标准生物素化诱饵竞争性杂交所述靶标,有效调节富集期间拉下的靶标分数(图17)。10个数量级动态范围的蛋白质表达可以通过使用这种竞争抑制方法举行几个数量级的压缩,特别是对于过丰种类,例如白蛋白。因此,针对给定位点的捕获的文库元件相对于标准杂交捕获的分数可以从100％调节至0％富集度。

[0614] 另外,文库标准化技术可用于从延伸记录标签、延伸编码标签或di-tag文库中去除过多的种类。这种方法最适用于源自位点特异性蛋白酶例如胰蛋白酶,LysC,GluC等消化产生的肽的定义长度文库。在一个实例中,标准化可以通过使双链文库变性并允许所述文库元件重新退火来完成。由于双分子杂交动力学的二阶速率常数,丰度文库元素比欠丰度

元素更快地重新退火(Bochman,Paeschke et al.2012)。可以使用本领域已知的方法将ssDNA文库元件与丰度dsDNA文库元件分离,例如羟基磷灰石柱上的色谱法(VanderNoot,et al.,2012,Biotechniques 53:373-380)或用来自Kamchatka蟹的双体特异性核酸酶(DSN)处理文库(Shagin et al.,2002,Genome Res.12:1935-42),所述双体特异性核酸酶(DSN)破坏所述dsDNA文库元件。

[0615] 在连接到固体支持物和/或得到的延伸记录标签文库之前,大分子的分级分离、富集和扣减方法的任何组合可以节省测序读数并改善低丰度种类的测量。

[0616] 在一些实施方案中,通过连接或末端互补PCR连接延伸记录标签、延伸编码标签或di-tags的文库以产生分别包含多种不同延伸记录标签、延伸编码标签或di-tags的长DNA分子(Du et al.,2003,BioTechniques 35:66-72;Muecke et al.,2008,Structure 16:837-841;U.S.Patent No.5,834,252,每个通过引用整体并入本文)。该实施方案优选用于纳米孔测序,其中通过所述纳米孔测序装置分析长链DNA。

[0617] 在一些实施方案中,对延伸记录标签、延伸编码标签或di-tags进行直接的单分子分析(参阅,例如Harris et al.,2008,Science 320:106-109)。可以在固体支持物上,例如流通槽或适于加载到流通槽表面(可选地,微孔图形化)上珠子,直接分析所述延伸记录标签、延伸编码标签或di-tags,其中所述流通槽或珠子能够与单分子测序仪或单分子解码仪器整合。对于单分子解码,可以使用几轮合并的荧光标记的解码寡核苷酸的杂交(Gunderson et al.,2004,Genome Res.14:970-7)来确定所述延伸记录标签中的编码标签的身份和顺序。为了解构所述编码标签的结合顺序,可以用如上描述的循环特异性编码标签标记所述结合剂(Gunderson et al.,2004,Genome Res.14:970-7)。循环特异性编码标签既可用于代表单个大分子的单个,级联延伸记录标签,也可用于代表单个大分子的延伸记录标签的集合。

[0618] 在对所述延伸报告标签,延伸编码标签或di-tag文库进行测序后,所得到的序列可以通过它们的UMI折叠,然后相关联到相应的大分子(例如,肽、蛋白、蛋白复合物)并且与细胞中总体大分子类型(例如肽、多肽、蛋白大分子的蛋白质组)对准。得到的序列也可以通过它们的隔室标签折叠并与其相应的隔室蛋白质组相关联,在特定的实施方案中,隔室蛋白质组仅包含单个或非常有限数量的蛋白分子。蛋白质鉴定和定量都可以从这种数字肽信息中轻松获得。

[0619] 在一些实施方案中,可以针对特定测序分析平台优化所述编码标签序列。在特定实施方案中,测序平台是纳米孔测序。在一些实施方案中,测序平台具有>5%,>10%,>15%,>20%,>25%或>30%的每碱基错误率。例如,如果要使用纳米孔测序仪器分析所述延伸记录标签,则所述条形码序列(例如,编码器序列)能设计成在通过纳米孔时是最佳电学上可辨别的。根据本文所述方法的肽测序可能非常适合纳米孔测序,鉴于纳米孔测序的所述单碱基准确度仍然相当低(75%-85%),但所述“编码序列”的判定应该更准确(>99%)。此外,一种称为双重中断纳米孔测序(DI)的技术可以与纳米孔链测序一起使用,而不需要分子马达,大大简化了所述系统设计(Derrington,Butler et al.2010)。通过DI纳米孔测序读出延伸记录标签需要级联的延伸记录标签库中的间隔元件与互补寡核苷酸退火。本文使用的寡核苷酸可包含LNA,或其它经修饰的核酸或类似物,以增加所得双链体的有效Tm。当用这些双链体间隔区修饰的单链延伸记录标记通过所述孔时,所述双链区将在收缩区瞬

时停滞,使得能够读出与所述双链区相邻的约3个碱基的电流。在DI纳米孔测序的特定实施方案中,所述编码器序列以这样的方式设计:与所述间隔元件相邻的三个碱基产生最大程度上可电辨别纳米孔信号(Derrington et al.,2010,Proc.Natl.Acad.Sci.USA 107:16060-5)。作为无马达DI排序的替代方案,所述间隔元件可以设计成采用二级结构,例如G-quartet,当它通过纳米孔时,它将瞬时停顿所述延伸记录标签、延伸编码标签或di-tag,从而能够读出所述相邻的编码器序列(Shim,Tan et al.2009,Zhang,Zhang et al.2016)。在经过所述停顿之后,下一个间隔子将再次产生瞬态停顿,从而能够读出下一个编码器序列,依此类推。

[0620] 本文公开的方法可用于同时(复用)多个大分子(例如肽)的分析,包括检测、定量和/或测序。如本文所用的复用是指在相同测定中分析多个大分子。所述多个大分子可以源自相同的样品或不同的样品。所述多个大分子可以衍生自相同的受试者或不同的受试者。分析的所述多个大分子可以是不同的大分子(例如肽),或源自不同样品的相同大分子(例如肽)。多个大分子包括2个或更多个大分子,5个或更多个大分子,10个或更多个大分子,50个或更多个大分子,100个或更多个大分子,500个或更多个大分子,1000个或更多个大分子,5,000个或更多个大分子,10,000个或更多个大分子,50,000个或更多个大分子,100,000个或更多个大分子,500,000个或更多个大分子,或1,000,000个或更多个大分子。

[0621] 样品复用可以通过前期标记大分子样品的记录标签的条形码化来实现。每个条形码代表不同的样品,并且可以在循环结合测定或序列分析之前合并样品。以这种方式,许多条形码标记的样品可以在单个管中同时处理。该方法是对在反相蛋白质阵列(RPPA)上进行的免疫测定的显着改进(Akbani,Becker et al.2014,Creighton and Huang 2015,Nishizuka和Mills 2016)。以这种方式,本公开内容基本上对具有简单工作流程的RPPA测定提供了高度数字化样品和分析物复用替代物。

[0622] XIII.经由NTAA识别、记录标签延伸、和NTAA切割的循环轮的大分子表征

[0623] 在某些实施方案中,本公开内容提供的用于分析大分子的方法包括多个结合循环,其中大分子与多种结合剂接触,并且结合剂的连续结合将历史结合信息以基于核酸的编码标签的形式转移到至少一个与所述大分子关联的记录标签。以这种方式,以核酸形式产生包含关于多个结合事件的信息的历史记录。

[0624] 在涉及使用基于N-末端降解的方法分析肽大分子的方法的实施方案中(参见图3,图4,图41和图42),在第一结合剂与n个氨基酸的肽的n NTAA接触和结合后,转移所述第一结合剂的编码标签信息到与肽相关的记录标签,从而产生第一次序延伸记录标签,所述nNTAA如本文所述被切割。所述nNTAA的切割所述将肽的n-1氨基酸转化为N-末端氨基酸,其在本文中称为n-1NTAA。如本文所述,nNTAA可任选地用基团(例如,PTC,DNP,SNP,乙酰基,脒基等)标记,其特别适用于与被设计来结合标记形式的NTAA的裂解酶结合。如果所述nNTAA被标记,则所述n-1NTAA然后用相同的基团标记。使第二结合剂与所述肽接触并结合所述n-1NTAA,并将第二结合剂的编码标签信息转移至第一次序延伸记录标签从而产生第二次序延伸记录标签(例如,用于产生代表所述肽的级联的第n次序延伸记录标签),或至不同的记录标签(例如,用于产生多个延伸记录标签,其统一代表所述肽)。所述n-1NTAA的切割将所述肽的n-2氨基酸转化为N-末端氨基酸,其在本文中称为n-2NTAA。另外的结合、转移、切割和任选的NTAA标记可如上所述发生直至n个氨基酸以产生共同代表所述肽的第n次序延伸

记录标签或n个单独的延伸记录标签。如本文所用,当用于指代结合剂,编码标签或延伸记录标签时,n“次序”是指使用所述结合剂及其相关的编码标签的n个结合循环,或指创建所述延伸记录标签的n个结合循环。

[0625] 在一些实施方案中,所述第一结合剂和所述第二结合剂与大分子的接触,与任选的任何其他结合剂(例如,第三结合剂、第四结合剂,第五结合剂等)同时进行。例如,可以将第一结合剂和第二结合剂,以及任选的任何其它序数的结合剂合并在一起,例如形成结合剂文库。在另一个实例中,将所述第一结合剂和第二结合剂,以及任选的任何其它序数的结合剂同时加入到大分子中,而不是合并在一起。在一个实施方案中,结合剂文库包含至少20种选择性结合20种标准天然氨基酸的结合剂。

[0626] 在其它实施方案中,所述第一结合剂和第二结合剂,以及任选的任何其它序数的结合剂各自在分开的结合循环中与大分子接触,按顺序添加。在某些实施方案中,优选同时使用多种结合剂,因为平行方法节省时间并且所述结合剂处于竞争中,这减少了非同源结合剂对被同源结合剂结合的位点的非特异性结合。

[0627] 通过本文描述的方法产生的最终延伸记录标签的长度取决于多个因素,包括编码标签的长度(例如,编码器序列和间隔区)、记录标签的长度(例如,独特分子标识符、间隔区、通用引发位点、条形码)、操作的结合循环数、以及来自每个结合循环的编码标签是否被转移到相同的延伸记录标签或多个延伸记录标签。在代表肽并且由Edman降解样的裂解方法产生的级联延伸记录标签的实例中,如果所述编码标签具有两侧各有5个碱基的间隔子的5碱基编码序列,则在最终延伸记录标签上的所述编码标签信息,其代表肽的结合剂历史,是10个碱基×Edman降解循环数。对于20个循环的运行,所述延伸记录至少为200个碱基(不包括初始记录标签序列)。该长度与标准的下一代测序仪器兼容。

[0628] 在最终结合循环和所述最终结合剂的编码标签信息转移到所述延伸记录标签之后,可以通过连接、引物延伸或本领域已知的其他方法添加通用反向引发位点来对所述记录标签加盖。在一些实施例中,所述记录标签中的通用正向引发位点与附加到所述最终延伸记录标签的通用反向启发位点兼容。在一些实施方案中,通用反向引发位点是Illumina P7引物(5’-CAAGCAGAAGACGGCATACGAGAT-3’-SEQ ID NO:134)或Illumina P5引物(5’-AATGATACGGCGACCACCGA-3’-SEQ ID NO133)。可以附加有义或反义P7,这取决于所述记录标签的链义。延伸记录标签文库可以直接从固体支持物(例如珠子)切割或扩增,并用于传统的下一代测序分析和方案中。

[0629] 在一些实施方案中,在单链延伸记录标签文库中进行引物延伸反应以复制其互补链。

[0630] 所述NGPS肽测序分析包括循环进程中的几个化学和酶促步骤。NGPS测序是单分子的事实赋予该过程几个关键优势。单分子测定的第一个关键优势是对各种循环化学/酶促步骤中的低效率的稳健性。这通过使用存在于所述编码标签序列中的循环特异性条形码来实现。

[0631] 使用循环特异性编码标签,我们跟踪每个循环的信息。由于这是单分子测序方法,因此在测序过程中每个结合/转移循环的70%的效率就足以产生可映射的序列信息。例如,在我们的序列平台上可以将十碱基肽序列“CPVQLWVDST”(SEQ ID NO:169)读作“CPXQXWXDXT”(SEQ ID NO:170)(其中X=任何氨基酸;通过循环数跟踪推断出氨基酸的存

在)。这种部分氨基酸序列读数足以使用BLASTP将其独特地映射回人p53蛋白。因此,我们的流程不必都非常完美。此外,当循环特异性条形码与我们的分区概念相结合时,蛋白质的绝对鉴定可以通过10个位置中仅识别出几个氨基酸来实现,因为我们知道什么肽组映射到原始蛋白质分子(通过隔室条形码)。

[0632] XIV.通过分级、区室化、和有限的结合性树脂的蛋白标准化。

[0633] 蛋白质组学分析的关键挑战之一是解决样品中蛋白丰度的大动态范围。血浆中(甚至是"前20"耗尽血浆)的蛋白超过10个数量级的动态范围。在某些实施方案中,在分析之前从样品中扣减某些蛋白种类(例如,高丰度蛋白)。这可以通过例如使用商业上可获得的蛋白消耗试剂例如Sigma的PROT20免疫耗尽试剂盒来实现,其消耗前20种血浆蛋白。此外,可以大大降低所述动态范围,甚至降低到可控的3-4个数量级的方法是有用。在某些实施方案中,蛋白质样品动态范围可以通过使用标准分级分离方法包括电泳和液相色谱法(Zhou,Ning等人,2012),来分级所述蛋白样品来调节,或者将所述分级物分区到加载有限蛋白结合能力的珠子/树脂(例如羟基化二氧化硅颗粒)(McCormick 1989)的隔室中并洗脱结合蛋白来调节。洗去每个隔室中的过量蛋白质。

[0634] 电泳方法的实例包括毛细管电泳(CE)、毛细管等电聚焦(CIEF)、毛细管等速电泳(CITP)、自由流电泳、凝胶洗脱的液体分级物包埋电泳(GELFrEE)。液相色谱蛋白质分离方法的实例包括反相(RP)、离子交换(IE)、尺寸排除(SE),亲水相互作用等。隔室分区的实例包括乳液、液滴、微孔、平面基板上的物理分离区域等。示例性蛋白结合珠子/树脂包括用酚基团或羟基衍生的二氧化硅纳米粒子(例如,来自Agilent Technologies的StrataClean Resin,来自LabTech的RapidClean,等)。通过限制所述珠子/树脂的结合能力,在给定分级物中洗脱的高丰度蛋白将仅部分地与所述珠子结合,并且过量蛋白被除去。

[0635] XV.分区单个细胞的蛋白质组或分子二次取样

[0636] 在另一方面,本公开内容提供了使用条形码和分区技术对样品中的蛋白进行大规模平行分析的方法。目前的蛋白质分析方法涉及将蛋白质大分子片段化成适合肽测序的较短肽分子。因此,使用此类方法获得的信息受限于片段化步骤的限制并且排除,例如蛋白质的长范围连续性信息,包括翻译后修饰、每个样品中发生的蛋白质-蛋白质互作、样品中存在的蛋白群的组成、或所述蛋白大分子的起源,例如来自特定细胞或细胞群。蛋白质分子内翻译后修饰的长范围信息(例如,蛋白质表征)提供了更完整的生物学图像,以及关于什么肽属于什么蛋白分子的长范围信息提供更稳健的肽序列到底层蛋白质序列的映射(见图15A)。当肽测序技术仅提供不完整的氨基酸序列信息,例如仅来自5种氨基酸类型的信息时,这尤其重要。通过使用本文公开的分区方法,结合来自多种源自相同蛋白质分子的肽的信息,蛋白分子(例如蛋白质组)的身份可以更准确地评估。隔室标签与来自相同隔室的蛋白质和肽的结合促进了分子和细胞信息的重建。在一般的蛋白质组分析中,细胞被裂解并且蛋白质被消化成短肽,破坏了关于哪种蛋白质来自哪种细胞或细胞类型,以及哪种肽来自哪种蛋白质或蛋白质复合物的整体信息。这种整体信息对于理解细胞和组织内的生物学和生物化学非常重要。

[0637] 分区是指将独特条形码随机分配给样品中大分子群体的大分子亚群。可以通过将大分子分配进隔室中来实现分区。分区可以由单个隔室内的大分子或来自一组隔室的多个隔室内的大分子组成。

[0638] 通过独特的隔室标签鉴定已经分隔到多个(例如,数百万到数十亿)隔室的相同物理隔室或隔室组中的大分子子集或蛋白质样品子集。因此,即使在将成分合并在一起之后,隔室标签也可用于将来自具有相同隔室标签的一个或多个隔室的成分从具有不同隔室标签的另一隔室(或隔室组)中的成分中区分出来。

[0639] 本公开内容提供了通过将复杂蛋白质组样品(例如,多种蛋白质复合物、蛋白质或多肽)或复合细胞样品分区成多个隔室来增强蛋白质分析的方法,其中每个隔室包含多个隔室标签,在单个隔室内的隔室标签相同(除了可选的UMI序列)并且与其他隔室的隔室标签不同(参见图18-20)。所述隔室任选地包含连接所述多个隔室标签的固体支撑物(例如珠子)。将多种蛋白质复合物、蛋白或多肽片段化成多个肽,然后在足以允许所述多个肽与所述多个隔室标签在所述多个隔室内退火或连接的条件下与所述多个隔室标签接触,从而产生多个隔室标记的肽。或者,所述多个蛋白复合物、蛋白或多肽在足以允许所述多个蛋白复合物、蛋白或多肽与所述多个隔室标签在多个隔室内退火或连接的条件下连接到多个隔室标签,从而产生多个隔室标记的蛋白复合物、蛋白、多肽。然后从所述多个隔室收集隔室标记的蛋白复合物、蛋白或多肽,并任选地片段化成多个隔室标记的肽。根据本文描述的任何方法分析一种或多种隔室标记的肽。

[0640] 在某些实施方案中,通过引物延伸(图5)或连接(图6)将隔室标签信息转移至与大分子(例如肽)相联的记录标签。

[0641] 在一些实施方案中,所述隔室标签在所述隔室内的溶液中是游离的。在其他实施方案中,所述隔室标签直接连接到隔室的表面(例如,微量滴定板或picotiter板的孔底部)或珠子或隔室内的珠子。

[0642] 隔室可以是含水隔室(例如,微流体液滴)或固体隔室。固体隔室包括,例如,纳米颗粒、微球、阵列上的微量滴定孔或picotiter孔或分离区域、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶芯片、流通槽、流通芯片、包括信号转导电子器件的生物芯片、ELISA板,旋转干涉测量盘、硝酸纤维素膜或基于硝酸纤维素的聚合物表面。在某些实施方案中,每个隔室平均含有单个细胞。

[0643] 固体支持物可以是任何支持表面,包括但不限于珠子、微珠、阵列、玻璃表面、硅表面、塑料表面、过滤器、膜、尼龙、硅晶芯片、流通槽、流通芯片、包括信号转导电子器件的生物芯片、ELISA板,旋转干涉测量盘、硝酸纤维素膜或基于硝酸纤维素的聚合物表面,纳米颗粒或微球。用于固体载体的材料包括但不限于丙烯酰胺、琼脂糖、纤维素、硝化纤维素、玻璃、金、石英、聚苯乙烯、聚乙烯乙酸乙烯酯、聚丙烯、聚甲基丙烯酸酯、聚乙烯、聚环氧乙烷、聚硅酸盐、聚碳酸酯、特氟隆、碳氟化合物、尼龙、硅橡胶、聚酸酐、聚乙醇酸、聚乳酸、聚原酸酯、官能化硅烷、聚丙基延胡索酸酯、胶原、糖胺聚糖、聚氨基酸或其任何组合。在某些实施方案中,固体支持物是珠子,例如,聚苯乙烯珠子、聚合物珠子、琼脂糖珠子、丙烯酰胺珠子、固体核心珠子、多孔珠子、顺磁珠子、玻璃珠子或可控孔珠子。

[0644] 在Shembekar等人(Shembekar,Chaipan等人,2016)中综述了将样品分区到具有隔室标记珠子的隔室的各种方法。在一个实例中,蛋白质组通过乳液分配成液滴,从而能够记录被本文公开的方法对蛋白质分子和蛋白质复合物的全局信息举行记录(参见例如图18和图19)。在某些实施方案中,所述蛋白质组与隔室标记的珠子、可活化的蛋白酶(直接或间接通过热、光等),以及被工程化(例如,修饰的赖氨酸、聚乙二醇化)为具有蛋白酶抗性的肽连

接酶一起被分配到隔室(例如,液滴)中。在某些实施方案中,可以用变性剂处理蛋白质组以评估蛋白或多肽的肽组分。如果需要关于蛋白质天然状态的信息,可以将相互作用的蛋白质复合物分配到隔室,用于随后分析由其衍生的肽。

[0645] 隔室标签包括条形码,其任选地在一侧或两侧侧接间隔子或通用引物序列。所述引物序列可以与记录标签的3'序列互补,从而能够通过引物延伸反应将隔室标签信息转移到所述记录标签(参见图22A-B)。所述条形码可以由附着于固体支持物或隔室的单链核酸分子或其与固体支持物或隔室杂交的互补序列,或这两种链组成(参见例如图16)。隔室标签可包含功能基团,例如附着于所述间隔子上,用于偶联肽。在一个实例中,功能基团(例如,醛)是指能够与所述多个肽上的N-末端氨基酸残基反应的基团。在另一个实例中,功能基团能够与所述多个肽上的内部氨基酸残基(例如,赖氨酸或用"点击"反应性基团标记的赖氨酸)反应。在另一个实施方案中,功能基团可以仅仅地是能够与DNA标签标记的蛋白杂交的互补DNA序列。或者,隔室标签可以是嵌合分子,还包含含有蛋白质连接酶(例如,butelase I或其同源物)的识别序列的肽以允许将所述隔室标签连接至目标肽(参见图22A)。隔室标签可以是较大核酸分子内的组分,其任选地还包含用于提供与其连接的肽的识别信息的独特分子标识符、间隔序列,通用引发位点或其任何组合。该UMI序列通常与隔室内的隔室标签群不同。在某些实施方案中,隔室标签是记录标签内的组分,使得用于提供个体隔室信息的相同标签也用于记录与其连接的肽的单个肽信息。

[0646] 在某些实施方案中,隔室标签可以通过印刷、点样、喷墨到所述隔室中形成隔室标签。在某些实施方案中,如Klein et al.,2015,Cell 161:1187-1201;Macosko et al.,2015,Cell 161:1202-1214;和Fan et al.,2015,Science347:1258367等描述的通过分开-合并寡核苷酸连接或合成形成多个隔室标记的珠子,其中每个珠子有一种条形码类型。隔室标记的珠子也可以通过单独合成或固定形成。在某些实施方案中,所述隔室标记的珠子还包含双功能记录标签,其中一部分包含含有记录标签的隔室标签,另一部分包含功能基团,其上可以偶联所述消化的肽(图19和图20)。

[0647] 在某些实施方案中,所述多个隔室内的多种蛋白或多肽用蛋白酶片段化成多个肽。蛋白酶可以是金属蛋白酶。在某些实施方案中,所述金属蛋白酶的活性通过金属阳离子的光活化释放来调节。可以使用的内肽酶的实例包括:胰蛋白酶、胰凝乳蛋白酶、弹性蛋白酶、嗜热菌蛋白酶、胃蛋白酶、梭菌蛋白酶、谷氨酰内肽酶(GluC)、内肽酶ArgC、肽酰基-asp金属内肽酶(AspN)、内肽酶LysC和内肽酶LysN。它们的活化模式取决于缓冲液和二价阳离子的要求。任选地,在将蛋白质或多肽充分消化成肽片段后,所述蛋白酶被灭活(例如,加热、氟代油或硅油可溶性抑制剂,例如二价阳离子螯合剂)。

[0648] 在具有隔室标签的肽条形码的某些实施方案中,用DNA标签标记蛋白分子(任选地,变性多肽),通过DNA标签与蛋白质的赖氨酸基团的ε-胺基的缀合或间接地通过点击化学附着于用反应性点击基团例如炔预标记的蛋白/多肽(参见图2B和图20A)。然后将所述DNA标签标记的多肽分配进入包含隔室标签(例如,与液滴中包含的珠子结合的DNA条形码)的隔室(参见图20B),其中隔室标签包含识别每个隔室的条形码。在一个实施方案中,单个蛋白质/多肽分子与与珠子相关的单一种类的DNA条形码共同包封(参见图20B)。在另一个实施方案中,所述隔室可以构成珠子的表面,其附着有与PCT公号WO2016/061517(通过引用整体并入)中所描述的类似的隔室(珠子)标签,除了被应用于蛋白质而不是DNA之外。所述

隔室标签可包括条形码(BC)序列、通用引发位点(U1')、UMI序列、和间隔序列(Sp)。在一个实施方案中,伴随着分区或在其后,将所述隔室标签从所述珠子上切割下并与附着于所述多肽的所述DNA标签杂交,例如分别通过DNA标签和隔室标签上的互补U1和U1'序列。为了在珠子上分区,所述DNA标签标记的蛋白质可以直接与所述珠子表面上的所述隔室标签杂交(参见图20C)。在该杂交步骤之后,从隔室中提取具有杂交DNA标签的多肽(例如,乳液"破裂",或从珠子切割的隔室标签),并且使用基于聚合酶的引物延伸步骤将条形码和UMI信息写入所述多肽上的所述DNA标签以产生隔室条形码记录标签(参见图20D)。LysC蛋白酶消化可用于将多肽切割成在其C-末端赖氨酸处用具有含通用引发序列、隔室标签和UMI的记录标签标记的肽(参见图20E)。在一个实施方案中,所述LysC蛋白酶被工程化以耐受DNA标记的赖氨酸残基。将得到的记录标签标记肽以适当的密度固定在固体基质(例如珠子)上,以使记录标记的肽之间的分子间相互作用最小化(参见图20E和20F)。

[0649] 所述肽附着到所述隔室标签(或反之亦然)可以直接与固定的隔室标签或其互补序列(如果是双链的)连接。或者,所述隔室标签可以从所述固体支持物或所述隔室的表面分离,且所述肽和溶液相隔室标签在隔室内连接。在一个实施方案中,所述隔室标签上(例如,在寡核苷酸的末端)的功能基团是醛,其通过席夫碱(Schiff base)直接与所述肽的胺N-末端偶联(参见图16)。在另一个实施方案中,所述隔室标签构建为核酸-肽嵌合分子,其包含用于蛋白连接酶的肽基序(n-X...XXCGSHV-c)。使用肽连接酶,例如butelase I或其同源物,将所述核酸-肽隔室标签构建体与消化的肽缀合。Butelase I,和其它天冬酰胺酰内肽酶(AEP)同源物可用于将寡核苷酸-肽隔室标签构建体的C末端连接至所述消化肽的N末端(Nguyen,Wang et al.2014,Nguyen,Cao et al.2015)。这种反应快速而有效。得到的隔室标记肽随后可以固定至固体支持物,用于如本文所述的核酸肽分析。

[0650] 在某些实施方案中,连接至固体支持物或隔室表面的隔室标签在与具有多个片段化肽的隔室标签连接之前释放(参见图18)。在一些实施方案中,在从多个隔室收集所述隔室标记的肽之后,所述隔室标记的肽连接到与记录标签相关联固体支持物。然后隔室标签信息可以从所述隔室标记的肽上的所述隔室标签转移到所述相关的记录标签上(例如,通过从所述记录标签和隔室标签内的互补间隔序列引发的引物延伸反应)。在一些实施方案中,然后根据本文所述的方法在肽分析之前从隔室标记的肽中移除所述隔室标签。在进一步的实施方案中,最初用于消化所述多种蛋白质的所述序列特异性蛋白酶(例如,Endo AspN)也用于在所述隔室标签信息转移至所述相关记录标签后从所述肽的N末端移除所述隔室标签(见图22B)。

[0651] 基于隔室分区的方法包括通过使用T形接头和流动聚焦的微流体装置形成液滴,使用搅拌产生乳液或通过具有小孔的膜(例如,轨迹蚀刻膜)挤出等(参见图21)。区室化的挑战在于处理隔室的内部。在某些实施方案中,由于交换流体组分是具有挑战性的,因此可能难以在隔室内进行一系列不同的生化步骤。如前所述,通过将试剂添加到所述乳液的氟油中,可以改变液滴内部的有限特征,例如pH、螯合剂、还原剂等。然而,在水相和有机相中都具有溶解性的化合物的数量是有限的。一种方法是将隔室中的反应限制为基本上所述条形码至所述目标分子的转移。

[0652] 在用含隔室标签(条形码)的记录标签标记所述蛋白质/肽后,所述将蛋白质/肽以合适的密度固定在固体支持物上,以有利于从结合的同源结合剂的编码标签到相应的记录

标签/附着在所述肽或蛋白质分子上的记录标签的分子内转移。通过控制所述固体支持物表面上分子的分子间间距来最小化分子间信息传递。

[0653] 在某些实施方案中,所述隔室标签对一组隔室中的每个隔室不需要是唯一的。隔室群中的隔室子集(两个,三个,四个或更多个)可以共享相同的隔室标签。例如,每个隔室可以包含一群珠子表面,其用于从样品中捕获大分子亚群(每个珠子捕获许多分子)。此外,所述珠子包括隔室条形码,其能够附着到捕获的大分子上。每个珠子仅具有单个隔室条形码序列,但是该隔室条形码可以在隔室中的其他珠子上复制(许多珠子映射到相同的条形码)。在物理隔室和隔室条形码之间可以存在(尽管不是必需的)多对一映射,此外,隔室内的大分子之间可以存在(尽管不是必需的)多对一映射。分区条形码被定义为将独特条形码分配给来自样品内的大分子群的大分子的子样品集。该分区条形码可以包含相同的隔室条形码,所述隔室条形码由在由相同条形码标记的隔室内的分区大分子产生。物理隔室的使用有效地对原始样本进行子采样以提供分区条形码的分配。例如,提供了一套标有10,000个不同隔室条形码的珠子。此外,假设在给定的测定中,在测定中使用100万个珠子的群体。平均而言,每个区室条形码有100个珠子(泊松分布)。进一步假设所述珠子捕获了1000万个大分子的聚群。平均而言,每个珠子有10个大分子,每个隔室条形码有100个隔室,每个分区条形码(包含100个不同的物理隔室的100个隔室条形码)实际上有1000个大分子。

[0654] 在另一个实施方案中,通过在N或C末端或两端都用可扩增的DNA UMI标签(例如记录标签)标记多肽(化学或酶促)来完成单分子分区和多肽的分区条形码编码(参见图37)。DNA标签通过非特异性光标记或特异性化学附着到反应性氨基酸如赖氨酸从而附着到所述多肽体(内部氨基酸)如图2B所示。来自附着于所述肽末端的记录标签的信息通过酶促乳液PCR(Williams,Peisajovich et al.2006,Schutze,Rubelt et al.2011)或乳液体外转录/逆转录(IVT/RT)步骤转移至所述DNA标签。在优选的实施方案中,使用纳米乳液,使得每个尺寸为50nm-1000nm的乳液液滴平均少于单个多肽(Nishikawa,Sunami et al.2012,Gupta,Eral et al.2016)。另外,PCR的所有组分都包括在水性乳液混合物中,包括引物、dNTP、Mg^{2+}、聚合酶和PCR缓冲液。如果使用IVT/RT,那么记录标签设计有T7/SP6 RNA聚合酶启动子序列以产生与附着于所述多肽体的所述DNA标签杂交的转录物(Ryckelynck,Baudrey et al.2015)。逆转录酶(RT)将来自杂交的RNA分子的信息复制到DNA标签。以这种方式,乳液PCR或IVT/RT可用于有效地将信息从末端记录标签转移到附着于多肽体的多个DNA标签(Ryckelynck,Baudrey et al.2015)。

[0655] 通过珠子中的凝胶化来封装细胞内容物是单细胞分析的有用方法(Tamminen和Virta 2015,Spencer,Tamminen et al.2016)。条形码单细胞液滴使得来自单个细胞的所有组分都能够用相同的标识符标记(Klein,Mazutis et al.2015,Gunderson,Steemers et al.2016,Zilionis,Nainys et al.2017)。隔室条形码可以通过多种方式实现,包括通过液滴连接(Raindance)将独特条形码直接结合到每个液滴中、通过将条形码珠子引入液滴(10X Genomics)、或通过液滴组分的组合条形码后封装、凝胶化使用和分开-合并组合条形码,如Gunderson等人描述。(Gunderson,Steemers et al.2016)和PCT Publication WO2016/130704,其全部内容通过引用并入本文。如Adey等人所述,类似的组合标记方案也可以应用于细胞核。(Vitak,Torkenczy et al.2017)。

[0656] 上述液滴条形码方法已用于DNA分析,但没用于蛋白质分析。使上述液滴条形码平

台适应蛋白质需要几个创新步骤。第一个是条形码主要由DNA序列组成,需要将这种DNA序列信息赋予所述蛋白质分析物。在DNA分析物的情况下,将DNA信息转移到DNA分析物上相对简单。相反,将DNA信息转移到蛋白质上更具挑战性,特别是当所述蛋白被变性并被消化成肽用于下游分析时。这要求每个肽都用隔室条形码标记。挑战在于,一旦将所述细胞包封成液滴,就难变性所述蛋白质,蛋白酶消化所述得到多肽,并同时用DNA条形码标记所述肽。聚合物形成液滴中的细胞封装及其聚合(凝胶化)成多孔珠粒,可以将其带入水性缓冲液中,这提供了执行多个不同反应步骤的载体,不像液滴中的细胞(Tamminen和Virta 2015,Spencer,Tamminen et al.2016)(Gunderson,Steemers et al.2016)。优选地,所述包封的蛋白质与所述凝胶基质交联以防止它们随后从凝胶珠粒扩散。这种凝胶珠粒形式允许所述凝胶中的包埋蛋白质经化学或酶促变性、用DNA标签标记、蛋白酶消化、并进行许多其他干预。图38描绘了凝胶基质中单个细胞的示例性包封和裂解。

[0657] XVI.组织和单细胞空间蛋白质组学

[0658] 条形码的另一个用途是在空间分布的DNA条形码序列阵列表面上的组织的空间分割。如果用含有条形码的DNA记录标签标记组织蛋白,该条形码反映了安装在阵列表面上的所述细胞组织内的所述蛋白质的空间位置,那么所述组织切片内所述蛋白质分析物的所述空间分布可以在序列分析后重建,就像是如Stahl等人所述,用于空间转录组学研究。(2016,Science353(6294):78-82)和Crosetto et al.(Corsetto,Bienko et al.,2015)。空间条形码的附着可以通过从所述阵列释放阵列结合的条形码并将它们分散到所述组织切片中来实现,或者,所述组织切片中的所述蛋白质可以用DNA记录标签标记,然后用蛋白酶消化所述蛋白质以释放标记的肽,其可以扩散并与所述阵列上的空间条形码杂交。然后可以将所述条形码信息转移(酶促或化学)到与所述肽连接的所述记录标签上。

[0659] 组织内的所述蛋白质的空间条形码可以通过将固定/透化的组织切片,用DNA记录标签化学标记,放置在空间编码的DNA阵列上来实现,其中所述阵列上的每个特征具有空间可识别条形码(参见图23)。为了将阵列条形码附着到所述DNA标签,可以用蛋白酶消化所述组织切片,释放DNA标签标记的肽,其可以扩散并与所述组织切片邻近的近端阵列特征杂交。可以使用化学/酶促连接或聚合酶延伸将所述阵列条形码信息转移至所述DNA标签。或者,不是所述允许标记的肽扩散到所述阵列表面,而是可以切割所述阵列上的条形码序列并使其扩散到所述组织切片上的近端区域并与其中的DNA标签标记的蛋白质杂交。再一次,所述条形码信息可以通过化学/酶促连接或聚合酶延伸来转移。在第二种情况下,蛋白酶消化可以在转移条形码信息后进行。任一种方法的结果都是记录标签标记的蛋白质或肽的集合,其中所述记录标签包括含蛋白质/肽在起源组织内的位置的2-D空间信息的条形码。此外,可以表征翻译后修饰的空间分布。该方法提供了灵敏且高度多路复用的原位数字免疫组织化学分析,并且应该形成导致更准确的诊断和预后的现代分子病理学的基础。

[0660] 在另一个实施方案中,可以在细胞内使用空间条形码来鉴定细胞器和细胞隔室内的蛋白成分/PTM(Christoforou等,2016,Nat.Commun.7:8992,通过引用整体并入)。许多方法可用于提供细胞内空间条形码,其可附着于近侧蛋白质。在一个实施方案中,细胞或组织可以是分级到组成细胞器的亚细胞,以及条形码化的不同的蛋白细胞器分级物。其他空间细胞标记方法在Marx,2015,Nat Methods 12:815-819的综述中描述,其全部内容通过引用并入本文,类似的方法可以在此使用。

[0661] 提供以下实例是为了说明而非限制。

[0662] 示例

[0663] 示例1:用蛋白酶K消化蛋白质样品

[0664] 通过用蛋白酶如胰蛋白酶、蛋白酶K等消化从蛋白样品制备肽文库。胰蛋白酶优选切割如C-末端赖氨酸和精氨酸等带正电荷的氨基酸,而蛋白酶K在所述蛋白非选择性切割。因此,蛋白酶K消化需要使用优选的酶-多肽比例进行仔细滴定以提供足够的蛋白水解以产生短肽(~30个氨基酸),但不过度消化所述样品。通常,需要对给定的蛋白酶K批次进行功能活性的滴定。在该实施例中,用蛋白酶K消化蛋白质样品,在1X PBS/1mM EDTA/0.5mM $CaCl_2$/0.5%SDS(pH 8.0)中以1:10-1:100(w/w)酶:蛋白质比例37℃下消化1小时。温育后,加入PMSF至终浓度为5mM以抑制进一步消化。

[0665] 蛋白酶K的比活性可以通过将"化学底物"苯甲酰精氨酸-p-nitroanilide与蛋白酶K一起温育并测量在~410nm吸收的黄色p-nitroanilide产物的发展来测量。酶活性以单位测量,其中一个单位等于每分钟产生的1μM p-nitroanilide,并且以酶活性/毫克总蛋白质为单位测量比活性。然后通过将所述酶活性除以溶液中蛋白的总量来计算比活性。

[0666] 示例2:使用SP3在珠子上的蛋白酶消化和标记的样品制备

[0667] 如Hughes等人所述,使用SP3样品制备方案提取蛋白质并使其变性,(2014,Mol Syst Biol 10:757)。提取后,所述蛋白质混合物(和珠子)于50mM硼酸盐缓冲液补充有0.02%SDS的(pH 8.0)w/1mM EDTA中,37℃溶解1小时。在蛋白质溶解后,通过添加DTT至终浓度为5mM来降低二硫键,并在50℃下孵育样品10分钟。通过加入碘乙酰胺至终浓度为10mM使半胱氨酸烷基化,并在黑暗中室温下温育20分钟。将反应物在50mM硼酸盐缓冲液中稀释2倍,并最终蛋白酶:蛋白质比例为1:50(w/w)加入Glu-C或Lys-C。在37℃ o/n(~16hrs.)温育所述样品以完成消化。在如Hughes等人描述的样品消化后(同上),通过加入100%乙腈至终浓度为95%乙腈使所述肽与所述珠子结合,并在8分钟内用乙腈洗涤孵化。洗涤后,通过5分钟的移液管混合步骤将肽从10μl的2% DMSO中的珠上洗脱下来。

[0668] 示例3:所述记录标签与所述肽的偶联

[0669] DNA记录标签以几种方式与肽偶联(参见Aslam等,1998,Bioconjugation:Protein Coupling Techniques for the Biomedical Sciences,Macmillan Reference LTD;Hermanson GT,1996,Bioconjugate Techniques,Academic Press Inc.,1996)。在一种方法中,寡核苷酸记录标签构建有采用碳二亚胺化学与所述肽的C-末端偶联的5'胺、和内部张紧性炔烃,使用点击化学与叠氮化物珠偶联的DBCO-dT(Glen Research,VA)。使用大摩尔过量的记录标签将所述记录标签与溶液中的肽偶联以驱动所述碳二亚胺偶联完成,并限制肽-肽偶联。或者,所述寡核苷酸用5'张紧炔(DBCO-dT)构建,并与叠氮化物衍生的肽偶联(通过叠氮化物-PEG-胺和碳二亚胺偶联至肽的C-末端),并与醛-反应性HyNic肼珠偶联。为此目的,可以用内部醛甲酰基吲哚(Trilink)基团容易地标记所述记录标签寡核苷酸。或者,所述记录标签不是偶联到C-末端胺,而是偶联到内部赖氨酸残基上(优选在Lys-C消化后,或者可选地在Glu-C消化后)。在一种方法中,这可以通过用NHS-叠氮化物(或NHS-PEG-叠氮化物)基团活化所述赖氨酸胺然后偶联到5'胺标记的记录标签来实现。在另一种方法中,5'胺标记的记录标签可以与过量的NHS同源双功能交联剂如DSS反应,以产生5'NHS活化的记录标签。该5'NHS活化的记录标签可以直接与肽的赖氨酸残基的-氨基偶联。

[0670] 示例4:肽上氨基酸的位点特异性标记

[0671] 五个不同蛋白质或多肽上的氨基酸的例子,是能用活化的DNA标签(使用异双功能氨基酸位点特异性试剂激活)直接修饰的氨基酸,或能通过采用具有点击基团的位点特异性标记氨基酸的点击化学异双功能试剂间接修饰的氨基酸,所述点击基团后来用于附着于所述DNA标签上的同源点击基团(Lundblad 2014)。典型的蛋白质输入包含在含有0.1%RapiGest™ SF表面活性剂和5mM TCEP的50μl适当水性缓冲液中的1μg蛋白质。RapiGest™ SD可用作酸降解表面活性剂,用于将蛋白质变性为多肽以改善标记或消化。可以使用以下氨基酸标记策略:使用马来酰亚胺化学的半胱氨酸---200μMSulfo-SMCC活化的DNA标签用于在100mM MES缓冲液(pH 6.5)+1%TX-100中对半胱氨酸举行1小时的位点特异性标记;使用NHS化学的赖氨酸---200μM DSS或BS^3活化的DNA标签用于在硼酸盐缓冲液(50mM,pH 8.5)+1%TX-100中室温下对溶液相蛋白上或珠子结合肽上的赖氨酸进行位点特异性标记;酪氨酸用4-苯基-3H-1,2,4-三唑啉-3,5(4H)-二酮(PTAD)或重氮化学修饰用于重氮化学,DNA标签用EDC和4-羧基苯重氮四氟硼酸盐活化(Aikon International,中国)。通过将所述蛋白质或珠子结合肽与200μM重氮衍生的DNA标签在硼酸盐缓冲液(50mM,pH8.5)+1%TX-100中,冰上孵育1小时来产生与酪氨酸连接的重氮(Nguyen,Cao et al.2015)。使用EDC化学修饰天冬氨酸/谷氨酸-将胺标记的DNA标签与所述珠子结合肽和pH6.5MES中的100mM EDC/50mM咪唑一起在室温下温育1小时(Basle et al.,2010,Chem.Biol.17:213-227)。标记后,使用来自C4树脂ZipTips(Millipore)的蛋白结合洗脱除去过量活化的DNA标签。将洗脱的蛋白加入50μl 1X PBS缓冲液中。

[0672] 示例5:将张紧炔烃记录标签标记的肽固定到叠氮活化的珠子上

[0673] 叠氮化物衍生的Dynabeads®M-270珠粒通过使市售的胺Dynabeads®M-270与叠氮化物PEG NHS酯异双功能接头(JenKem Technology,TX)反应生产。此外,叠氮化物的表面密度可以通过混入适当比例甲氧基或羟基PEG NHS酯来滴定。对于给定的肽样品,1-2mg叠氮化物衍生的Dynabeads®M-270珠子(~1.3x10^8珠子)在100μl硼酸盐缓冲液(50mM硼酸钠,pH 8.5)中稀释,加入1ng记录标签-肽,23-37℃孵育1小时。用200μl硼酸盐缓冲液洗涤3次。

[0674] 示例6:制备醛基吲哚活性HYNIC珠子

[0675] 胺珠子的HyNic衍生化产生醛基吲哚反应性珠子。将等分的20mg Dynabeads®M-270胺珠(2.8μm)珠悬浮于200μl硼酸盐缓冲液中。在短暂超声处理后,加入1-2mg Sulfo-S-HyNic(琥珀酰亚胺基6-肼基烟酸酯丙酮腙,SANH)(Catalog#S-1002,Solulink,San Diego)并在室温下震荡反应混合物1小时。然后将所述珠子用硼酸盐缓冲液洗涤2次,用柠檬酸盐缓冲液(200mM柠檬酸钠)洗涤1次。将所述珠子以终浓度为10mg/ml悬浮在柠檬酸盐缓冲液中。

[0676] 示例7:将记录标签醛基吲哚标记的肽固定到活化珠子上

[0677] 将等分的1-2mg HyNic活化的Dynabeads®M-270珠(~1.3×10^8个珠)稀释于补充有50mM苯胺的100μl柠檬酸盐缓冲液中,加入~1ng记录标签肽缀合物并在37℃孵育1小时。所述珠子用200μl柠檬酸盐缓冲液洗涤3次,并重悬于100μl硼酸盐缓冲液中。

[0678] 示例8:寡核苷酸模型系统-通过编码标签的识别信息到循环方式中的记录标签的

转移记录结合剂历史

[0679] 对于核酸编码标签和记录标签,可以使用标准核酸酶学的连接或引物延伸将信息从结合的结合剂上的所述编码标签转移到近侧记录标签。这可以用简单的模型系统来证明,该系统由5'部分代表结合剂靶标,3'部分代表所述记录标签的寡核苷酸组成。可以使用点击化学通过dT-炔烃修饰(DBCO-dT,Glen Research)将所述寡核苷酸固定在内部位点。在图24A所示的示例中,所述固定的寡核苷酸(AB靶标)含有两个靶标结合区域,标记的A和B,同源寡核苷酸"结合剂"可与之结合,所述A寡核苷酸和B寡核苷酸。所述A寡核苷酸和B寡核苷酸与编码标签(序列和长度不同)连接,所述编码标签通过共同的间隔区(Sp)与所述记录标签相互作用以引发引物延伸(或连接)。所述Sp的长度应保持短(例如,6-9个碱基)以最小化结合剂结合期间的非特异性相互作用。在该特定示例中,所述编码标签的长度被设计为通过凝胶分析易于从来自"B"寡核苷酸结合事件(20碱基编码序列)中区分出"A"寡核苷酸结合事件(10碱基编码序列)。

[0680] 通过对PAGE凝胶的简单分析,可以测量A或B编码标签转移的效率,并可以轻松优化实验参数。除所述AB靶标序列外,还可以使用类似的寡核苷酸CD靶序列(参见图24B),除了C和D是与A和B不相互作用的不同杂交序列。此外,C和D含有不同序列和长度的编码标签,分别包含30碱基DNA编码和40碱基DNA编码。第二靶序列CD的目的是评估AB和CD靶分子之间的交叉相互作用。假设特异性杂交,所述CD靶标的所述延伸记录标签不应包含A或B编码标签信息,除非在连接到与所述AB靶标结合的寡核苷酸的A或B编码标签之间发生分子间交叉。同样,所述AB目标的所述延伸记录标签不应包含C或D编码标签信息。在所述AB和CD靶标物理邻近(即<50nm)的情况下,可能存在串扰。因此,重要的是适当地将表面上的靶标大分子间隔开。

[0681] 该寡核苷酸模型系统能够完全表征结合剂历史的记录能力。图25显示了通过连接而不是引物延伸的信息转移。在对凝胶进行初始优化后,进行各种结合和测定方案并通过测序评估。独特分子标识符(UMI)序列用于计数目的,并且能够鉴定源自单个大分子的读数,并提供原始样品中总体大分子复杂性的测量。示例性的历史结合方案包括:A-B-C-B-A,A-B-A-A-B-A,A-B-C-D-A-C等。得到的最终产物应读为:UMI-Sp-A-Sp-B-Sp-B-Sp-A-Sp+UMI-Sp-C-Sp;UMI-SP-A-SP-B-SP-A-SP-A-SP-B-SP-A;UMI-A-Sp-B-Sp-A+UMI-Sp-C-Sp-D-Sp-C-Sp。该分析的结果允许进一步优化。

[0682] 示例9:寡核苷酸-肽模型系统-通过传递编码标签的识别信息到循环中的记录标签的结合剂历史记录

[0683] 在验证所述寡核苷酸模型系统后,通过将肽表位标签与示例性靶寡核苷酸序列的5'末端缀合,从所述寡核苷酸系统构建肽模型系统(图26A和26B)。示例性肽表位标签包括:FLAG(DYKDDDDK)(SEQ ID NO:171),V5(GKPIPNPLLGLDST)(SEQ ID NO:172),c-Myc(EQKLISEEDL)(SEQ ID NO:173),HA(YPYDVPDYA)(SEQ ID NO:174),V5(GKPIPNPLLGLDST)(SEQ ID NO:175),StrepTag II(NWSHPQFEK)(SEQ ID NO:176),etc等。可选的Cys-Ser-Gly接头可以用来将所述肽表位标签偶联到所述寡核苷酸。用A_寡核苷酸-cMyc肽构建体替换实施例7的AB寡核苷酸模板,用C_寡核苷酸-HA肽构建体替换实施例7的CD寡核苷酸模板(参见图26)。所述A_oligonucleotide-cMyc肽构建体还含有CSG接头和N-末端磷酸酪氨酸。同样,所述同源肽结合剂、cMyc抗体和HA抗体分别用所述B寡核苷酸编码标签和D寡核苷酸编

码标签标记。所述磷酸酪氨酸特异性抗体用单独的“E”编码标签标记。以这种方式,所述肽模型系统与所述寡核苷酸系统平行,并且寡核苷酸结合和抗体结合都在该模型系统中测试。

[0684] 固定的DNA肽构建体抗体染色使用抗c-myc抗体(2G8D5,小鼠单克隆,GenScript)、抗HA抗体(5E11D8,小鼠单克隆,GenScript)、strep-tag II抗体(5A9F9,小鼠单克隆,GenScript)、或抗FLAG抗体(5AE85,小鼠单克隆,GenScript)在1X PBST(PBS+0.1%吐温20)中使用0.1-1μg/ml进行。孵育通常在室温下进行30分钟。还进行1X PBST中1%PVP的标准预封闭,和染色后洗涤。通过用高盐(1M NaCl)和低pH(甘氨酸,pH 2.5)或高pH(三乙胺,pH 11.5)洗涤有效地实现抗体去染色。

[0685] 如Williams等人描述,所述靶寡核苷酸含有用于连接叠氮化珠子的内部炔标记,并且其5’末端含有氨基用于SMCC介导的与所述肽的C-末端半胱氨酸的连接。(2010,Curr Protoc Nucleic Acid Chem.Chapter 4:Unit 4.41)。或者,标准碳二亚胺偶联用于所述寡核苷酸和肽的缀合反应(Lu et al.,2010,Bioconjug.Chem.21:187-202)。在这种情况下,过量的寡核苷酸用于驱动碳二亚胺反应并使肽-肽偶联最小化。缀合后,通过从PAGE凝胶上切除并洗脱来纯化最终产物。

[0686] 示例10:通过DNA/PNA编码标签的连接进行编码标签转移以记录标记

[0687] 通过连接将编码标签直接或间接转移到所述记录标签以产生延伸记录标签。在一个实施方案中,将编码标签的退火互补物连接到所述记录标签(图25)。该编码标签互补物可以是核酸(DNA或RNA),肽核酸(PNA),或能够与生长的记录标签连接的一些其他编码分子。在使用标准ATP依赖性和NADH依赖性连接酶的DNA和RNA的情况下,连接可以是酶促的,或者对于DNA/RNA尤其是所述肽核酸PNA,连接可以是化学介导的。

[0688] 对于DNA的酶促连接,所述退火的编码标签要求5’磷酸连接到所述记录标签的3’羟基。示例性酶促连接条件如下(Gunderson,Huang et al.1998):50mM Tris-HCl(pH 7.8),10mM MgCl2,10mM DTT,1mM ATP,50μg/ml BSA,100mM NaCl,0.1% TX-100和2.0U/μl T4 DNA连接酶(New England Biolabs)。大肠杆菌DNA连接酶反应包括40mM Tris-HCl(pH 8.0),10mM $MgCl_2$,5mM DTT,0.5mM NADH,50μg/ml BSA,0.1% TX-100,和0.025U/μl E.coli DNA连接酶(Amersham)。Taq DNA连接反应包括20mM Tris-HCl(pH 7.6),25mM乙酸钾,10mM乙酸镁,10mM DTT,1mM NADH,50μg/ml BSA,0.1%Triton X-100,10%PEG,100mM NaCl和1.0U/μl Taq DNA连接酶(New England Biolabs)。T4和大肠杆菌DNA连接酶反应在室温下进行1小时,Taq DNA连接酶反应在40℃下进行1小时。

[0689] DNA/PNA的模板化学连接的几种方法可用于DNA/PNA编码标签转移。这些包括标准化学连接和点击化学方法。用于模板DNA连接的示例性化学连接条件如下(Gunderson,Huang et al.1998):模板3’磷酸报告标签与5’磷酸编码标签的连接在包含由50mM 2-[N-吗啉代]乙磺酸(MES)(pH 6.0含KOH)、10mM $MgCl_2$、0.001%SDS、新制备的200mM EDC、50mM咪唑(pH6.0含HCl)或50mM HOBt(pH 6.0,含HCl)和3.0-4.0M TMACl(Sigma)组成的反应中,在室温下进行1小时。

[0690] 用于PNA的模板依赖性连接的示例性条件包括NH 2-PNA-CHO聚合物(例如,编码标签补体和延伸的记录标签)的连接,并且由Brudno等人描述(Brudno,Birnbaum et al.2010)。PNA具有5’胺当量和3’醛当量,其中化学连接将两个部分偶联以产生Schiff碱,

随后用氰基硼氢化钠还原。该偶联的典型反应条件是:100mM TAPS(pH 8.5),80mM NaCl和80mM氰基硼氢化钠,室温下60分钟。使用含有5'氨基末端1,2-氨基硫醇(aminothiol)修饰和3'C-末端硫酯修饰的官能化PNA进行天然化学连接的示例性条件由Roloff等人描述。(2014,Methods Mol.Biol.1050:131-141)。其他N-和C-末端PNA基团也可用于连接。另一个例子涉及使用点击化学的PNA的化学连接。使用Peng等人的方法(2010,European J.Org.Chem.2010:4194-4197),PNA可以用5'叠氮化物和3'炔衍生化,并使用点击化学连接。"点击"化学连接的示例性反应条件是:1-2mg含有模板化PNA-PNA的珠子,在100μl含有10mM磷酸钾缓冲液、100mM KCl、5mM THPTA(三羟丙基三唑基胺)、0.5mM $CuSO_4$和2.5mM抗坏血酸钠的反应混合物中。所述化学连接反应在室温下温育1小时。Sakurai等描述了PNA连接的其他示例性方法(Sakurai,Snyder et al.2005)。

[0691] 示例11:PNA翻译成DNA

[0692] 使用退火到所述PNA模板上的DNA寡核苷酸的点击化学介导的聚合将PNA翻译成DNA。所述DNA寡核苷酸含有反应性5'叠氮化物和3'炔烃以产生能够被DNA聚合酶复制的核苷酸间三唑键(El-Sagheer et al.,2011,Proc.Natl.Acad.Sci.USA 108:11338-11343).将一组完整的与PNA中所有可能的编码标签互补的DNA寡核苷酸(10nM,1X杂交缓冲液:10mM硼酸盐(pH8.5),0.2M NaCl)与固相结合的PNA分子孵育(23-50℃)30分钟。退火后,将所述固相结合的PNA-DNA构建体用抗坏血酸钠缓冲液(10mM抗坏血酸钠,200mM NaCl)洗涤1次。所述"点击化学"反应条件如下:将珠子上的PNA-DNA在新鲜的抗坏血酸钠缓冲液中孵育,并与10mM THPTA+2mM $CuSO_4$混合物以1:1结合,并在室温下孵育1小时。然后用杂交缓冲液洗涤所述珠子1次,用PCR缓冲液洗涤2次。化学连接后,在El-Sagheer等人描述的条件下PCR扩增所得到的连接的DNA产物(2011,Proc.Natl.Acad.Sci.USA 108:11338-11343)。

[0693] 示例12:与核酸记录和编码标记相容的温和N-末端EDMAN降解

[0694] N末端Edman降解和DNA编码之间的相容性允许这种方法用于肽测序。N-末端Edman降解法的标准条件(采用无水TFA)破坏DNA。然而,这种影响可以通过开发更温和的切割条件和开发具有更耐酸的修饰的DNA来缓解。N末端Edman降解的较温和条件可以使用苯基硫代氨基甲酰基(PTC)-肽的切割优化和在所述切割条件下测量的DNA/PNA编码文库的稳定性的组合来开发。此外,通过使用碱基修饰,例如在低pH下减少脱嘌呤的7-脱氮嘌呤,以及减少脱嘌呤的5'甲基修饰胞嘧啶,来稳定天然DNA以对抗酸水解(Schneider和Chait,1995,Nucleic Acids Res.23:1570-1575)。鉴于胸腺嘧啶是最稳定的酸片段化基础,富含T的编码标签也可能是有用的。如Barrett等人所述,所述温和N末端Edman降解的条件用在乙腈中采用三乙胺乙酸盐60℃ 10分钟的温和碱基切割代替无水TFA切割。(1985,Tetrahedron Lett.26:4375-4378,通过引用整体并入)。这些温和的条件与大多数类型的DNA报告和编码标签兼容。作为替代方案,PNA用在编码标签中,因为它们完全是酸稳定的(Ray and Norden,2000,FASEB J.14:1041-1060)。

[0695] 使用以下测定证明了使用DNA编码标签/记录标签编码NTAA结合物的身份以及进行温和的N-末端Edman降解反应的相容性。抗磷酸酪氨酸和抗cMyc抗体均用于读出模型肽。C-Myc和N-末端磷酸酪氨酸检测、编码标签写入、和使用单个Edman降解步骤的N-末端磷酸酪氨酸去除。在该步骤之后,再次用抗磷酸酪氨酸和抗cMyc抗体染色所述肽。通过qPCR评估所述记录标签对N-末端降解的稳定性。通过测序、qPCR或凝胶电泳所分析,最终记录标签序

列中缺乏E-寡核苷酸编码标签信息来指示所述磷酸酪氨酸的有效去除

[0696] 示例13:制备隔室标记的珠子。

[0697] 为了制备隔室标记的珠子,使用分开-合并合成方法,采用亚磷酰胺合成或通过分开-合并连接将条形码掺入到固定在珠子上的寡核苷酸中。隔室标签可进一步包含独特分子标识符(UMI)以唯一地标记所述隔室标签连接的每个肽或蛋白分子。示例性的隔室标签序列如下:5'-NH2-GCGCAATCAG-XXXXXXXXXXXX-NNNNN-TGCAAGGAT-3'(SEQ ID NO:177)。所述XXXXXXXXXXXX(SEQ ID NO:178)条形码序列是每个珠子的固定的核碱基序列群,其通过在珠子上的分开-合并合成产生,其中所述固定的序列在珠子与珠子之间不同。所述NNNNN(SEQ ID NO:179)序列在珠子内随机化作为随后与其连接的肽分子的独特分子标识符(UMI)。如Macosko等人所述,可以使用分开-合并方法在珠子上合成所述条形码序列。(2015,Cell161:1202-1214,通过引用整体并入)。所述UMI序列可以通过使用简并碱基混合物(所有四种亚磷酰胺碱的混合物存在于每个偶联步骤中)合成寡核苷酸来产生。所述5'-NH_2用琥珀酰亚胺基4-(N-马来酰亚胺甲基)环己烷-1-羧酸酯(SMCC)活化,并且含有蛋白酶I肽底物(具有从N-末端到C-末端"CGGSSGSNHV"(SEQ ID NO:180)序列)的半胱氨酸使用Williams等人描述的改进方案与SMCC活化的隔室标记珠子偶联。(2010,Curr Protoc Nucleic Acid Chem.Chapter 4:Unit 4.41)。即,将200μl磁珠(10mg/ml)置于1.5ml Eppendorf管中。将1ml偶联缓冲液(100mM KH_2PO_4缓冲液,pH 7.2,含5mM EDTA,0.01%Tween 20,pH 7.4)加入管中并短暂涡旋。将新鲜制备的40μl Sulfo-SMCC(50mg/ml,在DMSO中,ThermoFisher)加入到所述磁珠中并混合。反应物在室温下在旋转混合器上温育1小时。温育后,在磁铁上将所述珠子与上清液分离,并用500μl偶联缓冲液洗涤3次。所述珠子重悬于400μl偶联缓冲液中。向磁珠中加入1mL CGGSSGSNHV(SEQ ID NO:180)肽(在TCEP还原(5mM)和冰冷丙酮沉淀后在偶联缓冲液中1mg/mL)。所述反应物于室温下在旋转混合器上温育2小时。所述反应物用偶联缓冲液洗涤1次。将400μl猝灭缓冲液(100mM KH_2PO_4缓冲液,pH 7.2和10mg/mL巯基琥珀酸,pH 7.4)加入到所述反应混合物中并旋转混合器上孵育2小时。所述反应混合物用偶联缓冲液洗涤3次。将得到的珠子重新悬浮在储存缓冲液(10mM KH_2PO_4缓冲液,pH 7.2含0.02%NaN_3,0.01%吐温20,pH 7.4)中并在4℃下储存。

[0698] 示例14:包封的珠子和蛋白质的生成

[0699] 隔室标记的珠子和蛋白与锌金属内肽酶结合,如内切蛋白酶AspN(Endo AspN),可选的光笼状Zn螯合剂(如ZincCleav I),和工程化耐热的butelase I同源物(Bandara,Kennedy et al.2009,Bandara,Walsh et al.2011,Cao,Nguyen et al.2015)。将来自实施例12的隔室标记的珠子与蛋白质混合并通过T-结构微流控或流体聚焦装置乳化(参见图21)。在双水流配置中,一个流体中的蛋白质和Zn^{2+}可以与来自另一个流体的金属内肽酶组合以在液滴形成时立即启动消化。在一种流动配置中,所有试剂都预混合并乳化在一起。这需要使用任选的光笼蔽的Zn螯合剂(例如,ZincCleav I),通过暴露于UV光下在液滴形成后引发蛋白质消化。所述调节浓度和流动条件,使得平均每个液滴少于一个珠子。在优化的实验中,可以制备10^8个femto-液滴,其中占10%的液滴含有珠子(Shim et al.,2013,ACS Nano 7:5955-5964)。在一种流动方法中,在形成液滴后,通过将乳液暴露于UV-365nm光以释放光笼蔽的Zn 2+,激活Endo AspN蛋白酶来活化蛋白酶。将乳液在37℃温育1小时以消化所述蛋白成肽。消化后,通过加热所述乳液至80℃15分钟使所述Endo AspN失活。在双流体

方案中,在将所述两种流体合并成液滴的过程中引入Zn^{2+}。在这种情况下,所述Endo AspN可以通过使用光活化的Zn^{2+}笼状分子灭活,其中所述螯合剂在暴露于紫外光下时激活,或者通过向油相中添加两亲性Zn^{2+}螯合剂,例如2-烷基丙二酸,或EDTA-MO来激活。两亲性EDTA分子的实例包括:EDTA-MO,EDTA-BO,EDTA-BP,DPTA-MO,DPTA-BO,DPTA-BP等(Ojha,Singh等人2010,Moghaddam,de Campo等人2012)。其它方式也可用于控制液滴内部的反应,包括通过向乳液油中加入两性酸或碱来改变液滴的pH。例如,可以使用水/油可溶性乙酸降低液滴pH。由于乙酸分子的两亲性质,向氟代乳液中加入乙酸导致所述液滴室内的pH降低(Mashaghi和van Oijen,2015,Sci Rep5:11837)。同样地,加入碱、丙胺,使所述液滴内部碱化。类似的方法可用于其他类型的两亲分子,例如油/水可溶性氧化还原试剂、还原剂、螯合剂和催化剂。

[0700] 在将区室化蛋白消化成肽后,使用butelase I或化学连接(例如,醛-氨基等)将所述肽连接到所述珠子上的所述隔室标签(寡核苷酸肽条形码嵌合体)(参见图16和图22A)。在任选的方法中,使用寡核苷酸-硫代缩肽“化学底物”使得所述butelase I连接不可逆(Nguyen,Cao et al.2015)。在连接后,所述乳液“破裂”,并且具有固定的隔室标记的肽构建体的珠子被大量收集,或者所述隔室标记的肽从珠子上切下,并且大量收集。如果所述固定了隔室标记的肽的珠子包含记录标签,则这些珠子可以直接用于本文所述的基于核酸编码的肽分析方法。相反,如果所述隔室标记的肽从所述珠子基质上切下,则隔室标记的肽随后通过与隔室标记的肽的C-末端缀合而记录标签关联,并固定在固体支持物上用于随后的采用如本文所述的编码标记的结合剂和测序分析的结合循环。记录标签与隔室标记肽的结合可以使用三官能连接分子完成。在将带有相关记录标签的隔室标记肽固定到固体支持物上进行循环测序分析后,使用引物延伸或连接将所述隔室信息转移到所述相关的记录标签上(参见图22B)。在将隔室标签信息转移到所述记录标签后,可以使用原始肽消化中使用的相同酶将所述隔室标签从所述肽上切割下来(参见图22B)。这恢复了所述肽的原始N-末端,从而能够实现如本文所述的N-末端降解肽测序方法。

[0701] 示例15:通过三种引物融合乳液PCR用氨基酸特异性编码标签共价修饰肽的相关记录标签的DI-TAG生成

[0702] 具有由隔室标签和分子UMI组成的记录标签的肽用编码标签位点特异性化学标记进行化学修饰。所述编码标签还含有UMI,以便能够计算修饰肽内给定类型的氨基酸的数量。使用来自Tyson和Armor的改进方案(Tyson and Armour 2012),在100μl的总水性体积中制备乳液PCR,其含有1xPHUSION™ GC反应缓冲液(Thermo Fisher Scientific),每种dNTPs 200μM(New England Biolabs),1μM引物U1,1μM引物U2tr,25nM引物Sp,14单位PHUSION™高保真DNA聚合酶(Thermo Fisher Scientific)。每5至10秒向2ml cryo瓶中的轻质矿物油中的200μl油相(4.5%vol./vol)Span 80,0.4%vol./vol.吐温80,0.05%Triton X-100(Sigma)中加入10μl水相。同时以1000rpm搅拌总共5分钟,如Turner和Hurles先前所述(2009,Nat.Protoc.4:1771-1783)。所得乳液的平均液滴尺寸为约5微米。也可以采用其他乳液生成方法,例如使用T形接头和流动聚焦(Brouzes,Medkova et al.2009)。产生乳液后,将100μl水/油混合物转移到0.5ml PCR管中,并在下列条件下进行第一轮扩增:98℃30秒;40个循环,98℃10秒,70℃30秒和72℃30秒;然后在72℃下延伸5分钟。第二轮扩增反应在以下条件下进行:98℃30秒;40个循环,98℃10秒,55℃30秒和72℃30秒;然后在4℃下保

持。在PCR的最后一个循环后,通过将200μl已烷(Sigma)直接加入PCR管中,涡旋20秒,并以13,000g离心3分钟,尽快破坏乳液。

[0703] 示例16:对延伸记录标签,延伸编码标签或DI-TAG构建体进行测序

[0704] 记录标签或编码标签的所述间隔区(Sp)或通用引发位点可以设计为在序列的主体中仅使用三个碱基(例如,A、C和T)和在序列的5'末端使用第四个碱基(例如,G)。对于合成测序(SBS),能够使用标准黑(未标记和未终止)核苷酸(dATP,dGTP和dTTP)和单个ffC染料标记的可逆终止子(例如,完全功能的胞嘧啶三磷酸)的混合物可以使黑碱基快速掺入到所述间隔序列上。通过这种方式,只有相关的编码器序列、独特分子标识符,隔室标签,所述延伸报告标签、延伸编码标签、或di-tag的结合循环序列被SBS测序,并且所述非相关的间隔子或通用引发序列被"跳过"。可以改变所述序列的5'末端的所述间隔子的碱基和第四碱基的身份,上述身份仅出于说明的目的提供。

[0705] 示例17:蛋白裂解物的制备。

[0706] 本领域已知多种用于制备来自各种样品类型的蛋白质裂解物的方案。这些方案的大多数变形取决于细胞类型以及裂解物中提取的蛋白质是否要在非变性或变性状态下进行分析。对于所述NGPA测定,天然构象或变性蛋白质都可以固定在固体基质上(参见图32)。此外,在固定天然蛋白质后,固定在所述基质表面上的所述蛋白质可以变性。使用变性蛋白质的优点是双重的。首先,许多抗体试剂结合线性表位(例如Western Blot Abs),并且变性蛋白质可以更容易访问线性表位。其次,当使用变性蛋白质时,简化了NGPA测定工作流程,因为所述固定的蛋白质已经变性,所以可以使用碱性(例如,0.1NaOH)剥离条件从所述延伸记录标签上剥离所述退火的编码标签。这与使用包含天然构象的蛋白质的测定法中的所述退火的编码标签的移除形成对比,其需要在结合事件和信息转移后酶促去除所述退火的编码标签。

[0707] 非变性蛋白质裂解缓冲液的示例包括:由50mm HEPES(pH 7.4),150mM NaCl,1%Triton X-100,1.5mM $MgCl^2$,10%甘油组成的RPPA缓冲液;和商业缓冲液,例如M-PER哺乳动物蛋白质提取试剂(Thermo-Fisher)。变性裂解缓冲液包含50mm HEPES(pH8),1%SDS。添加尿素(1M-3M)或盐酸胍(1-8M)也可用于变形所述蛋白样品。除了裂解缓冲液的上述组分外,通常还包括蛋白酶和磷酸酶抑制剂。蛋白酶抑制剂和典型浓度的实例包括aptrotinin(2μg/ml),亮肽素(5-10μg/ml),苯甲脒(15μg/ml),胃蛋白酶抑制剂A(1μg/ml),PMSF(1mM),EDTA(5mM)和EGTA(1mM)。磷酸酶抑制剂的实例包括焦磷酸钠(10mM),氟化钠(5-100mM)和正钒酸钠(1mM)。其他添加剂可包括DNAaseI以从蛋白质样品中去除DNA,以及还原剂如DTT以减少二硫键。

[0708] 从组织培养细胞制备的非变性蛋白质裂解物方案的实例如下:用胰蛋白酶(含0.05%胰蛋白酶-EDTA的PBS溶液)处理贴壁细胞,通过离心(200g,5分钟)收集,并在冰冷PBS条件下洗涤2次。添加冰冷的M-PER哺乳动物提取试剂(1毫升,每100mm培养皿或150cm^2培养瓶约10^7个细胞),其补充有蛋白酶/磷酸酶抑制剂和添加剂(例如,添加不含完全抑制剂(罗氏)和PhosStop(罗氏)的EDTA。将得到的细胞悬浮液在旋转振荡器上于4℃温育20分钟。然后在4℃以~12,000rpm(取决于细胞类型)离心20分钟以分离所述蛋白质上清液。使用BCA测定法定量所述蛋白质,并以1mg/ml重悬于PBS中。所述蛋白质裂解物可立即使用或在液氮中快速冷冻并储存在-80℃。

[0709] 基于Hughs等人的SP3方案,从组织培养细胞制备的变性蛋白质裂解物方案的实例如下:贴壁细胞用胰蛋白酶(PBS中含0.05%胰蛋白酶-EDTA)消化,通过离心收集(200g 5min),并在冰冷的PBS中洗涤2次。加入补充了蛋白酶/磷酸酶抑制剂和添加剂(例如1X完全蛋白酶抑制剂混合物(Roche)的冰冷变性裂解缓冲液(1毫升,每100mm培养皿或150cm^2培养瓶约10^7个细胞)。将得到的细胞悬浮液在95℃下孵育5分钟。放在冰上5分钟。将Benzonase核酸酶(500U/ml)加入所述裂解物中并在37℃下孵育30分钟去除DNA和RNA。

[0710] 通过每100uL裂解物加入5μL200mMDTT并在45℃下孵育30分钟来还原所述蛋白质。通过每100uL裂解物加入10uL 400mM碘乙酰胺并在黑暗中于24℃温育30分钟来完成蛋白质半胱氨酸基团的烷基化。反应通过每100uL裂解物加入10uL 200mM DTT来淬灭。任选通过每100ul裂解物加入2ul酸酐和100μl 1M Na_2CO_3(pH 8.5)来酰化蛋白质。在室温下孵育30分钟,推荐使用戊酸酐、苯甲酸酐和丙酸酐而不是乙酸酐,以使"体内"乙酰化赖氨酸与通过酰化的"原位"封闭赖氨酸基团区别开来(Sidoli,Yuan et al.,2015)。所述反应通过加入5mg Tris(2-氨乙基)胺、聚合物(Sigma)并在室温下孵育30分钟淬灭。通过通过0.45微米醋酸纤维素Spin-X管(Corning)以2000g离心裂解物1分钟除去聚合物树脂。使用所述BCA测定法定量所述蛋白质,并以1mg/ml重悬于PBS中。

[0711] 在另外的实施例中,使用过滤辅助样品制备(FASP)方案产生标记的肽,如Erde等人所述,其中MWCO过滤装置用于蛋白质滞留、烷基化和肽酶消化(Erde,Loo et al.2014,Feist和Hummon 2015).

[0712] 示例18:分区标记的肽的生成

[0713] DNA标签(具有任选的样品条形码和正交附着基团)用于使用标准生物共轭方法标记变性多肽的赖氨酸上的-氨基(Hermanson 2013),或者使用光亲和标记(PAL)方法如二苯甲酮与所述多肽连接(Li,Liu等人,2013)。在用赖氨酸基团上的DNA标签标记多肽或在CH基团(通过PAL)上随机标记多肽并通过用酰基酸酰化来封闭未标记的基团后,将所述DNA标签标记的,酰化的多肽退火到具有附着的包含通用引发序列的DNA寡核苷酸的隔室珠子、隔室条形码、任选的UMI,和与附着于所述多肽的所述DNA标签的一部分互补的引物序列。由于多个DNA杂交标签的协同性,单个多肽分子主要与单个珠子相互作用,使得能够将相同的隔室条形码写入所述多肽分子的所有DNA标签。退火后,所述多肽结合DNA标签在退火的珠子结合DNA序列上引发聚合酶延伸反应。以这种方式,将所隔室条形码和其他功能元件写到与结合的多肽连接的所述DNA标签上。完成该步骤后,所述多肽具有多个记录标签附着,其中所述记录标签具有共同的间隔序列、条形码序列(例如样品、分级物、隔室、空间,等),任选的UMI和其他功能元件。使用标准内切蛋白酶如胰蛋白酶、GluC、蛋白酶K等将该标记的多肽消化成肽片段。注意:如果使用胰蛋白酶消化赖氨酸标记的多肽,则该多肽仅在Arg残基处切割而不是Lys残基(因为Lys残基被标记)。所述蛋白酶消化可以直接在珠子上进行,或者所述标记的多肽从条形码化的珠子上移除后进行。

[0714] 示例19:制备用于模型系统的DNA记录标签-肽缀合物。

[0715] 所述记录标签寡核苷酸用5' NH_2基团和内部mTetrazine基团合成,用于随后偶联珠子(炔烃-dT通过mTet-PEG-N_3异双功能交联剂转化为mTetrazine-dT)。如Williams等人所述,使用NHS/马来酰亚胺异双功能交联剂,例如LC-SMCC(ThermoFisher Scientific),将所述寡核苷酸的5' NH_2基团偶联到肽上的反应性半胱氨酸。(Williams和Chaput 2010)。特

别地,将20nmol的5'NH_2标记的寡核苷酸在硅化管中用乙醇沉淀并重悬于180ul磷酸盐偶联缓冲液(0.1M磷酸钾缓冲液,pH7.2)中。将5mg LC-SMCC重悬于1mL DMF(5mg/ml)中(在-20等分储存)。将等分的20μl LC-SMCC(5mg/ml)加入180μl所述重悬的寡核苷酸中,混合并在室温下温育1小时。混合物用2X乙醇沉淀。将得到的马来酰胺衍生的寡核苷酸重悬于200μl磷酸盐偶联缓冲液中。将含有半胱氨酸残基(纯度>95%,脱盐)的肽以1mg/ml(~0.5mM)重悬于DMSO中。将约50nmol肽(100μl)加入到反应混合物中,并在室温下温育过夜。如William等人所述,所述得到的DNA记录标签-肽缀合物使用天然PAGE纯化(Williams和Chaput 2010)。在硅化管中缀合物以浓度为100uM重新悬浮于磷酸盐偶联缓冲液中。

[0716] 示例20:用于DNA-肽固定的底物的开发。

[0717] 适于点击化学固定的磁珠通过将M-270胺磁性Dynabeads转化为能够偶联至炔烃或甲基四嗪标记的寡肽缀合物的叠氮化物衍生或TCO衍生的珠子来产生(参见例如图29D-E;图30D-E)。即,洗涤10mg M-270珠子并重悬于500μl硼酸盐缓冲液(100mM硼酸钠,pH8.5)中。将TCO-PEG(12-120)-NHS(Nanocs)和甲基-PEG(12-120)-NHS的混合物以1mM重悬于DMSO中,并在室温下与M-270胺珠子一起温育过夜。滴定甲基与TCO PEG的比例以调节所述珠子上的最终TCO表面密度,使得<100TCO基团/um^2(参见例如图31E;图34)。未反应的胺基团在室温下用0.1M乙酸酐和0.1MDIEA在DMF(10mg珠粒用500μl)中的混合物加盖2小时。加盖并在DMF中洗涤3次后,所述珠子重悬于10mg/ml的磷酸盐偶联缓冲液中。

[0718] 示例21:将记录标签标记的肽固定到基质。

[0719] 使用记录标签上的mTet基团和活化珠子或基质表面的TCO基团,通过IEDDA点击化学反应将记录标签标记的肽固定在基质上。即使在低输入反应物浓度下,该反应也是快速高效的。此外,使用甲基四嗪对所述键赋予更大的稳定性(Selvaraj和Fox 2013,Knall,Hollauf et al.2014,Wu和Devaraj2016)。将200ng M-270TCO珠重悬于100μl磷酸盐偶联缓冲液中。将5pmol在所述记录标签上包含mTet基团的DNA记录标签标记的肽加入到珠子中,使最终浓度为~50nM。将反应物在室温下温育1小时。固定后,将所述底物上未反应的TCO基团在室温下用含1mM甲基四嗪酸的磷酸盐偶联缓冲液中猝灭1小时。

[0720] 示例22:N-末端氨基酸(NTAA)修饰

[0721] 化学NTAA乙酰化:

[0722] 肽的NTAA通过使用有机的或水溶液的乙酸酐或NHS-乙酸酯(磺基-NHS-乙酸酯)乙酰化。对于乙酸酐衍生化反应,含10mM乙酸酐的DMF与所述肽一起在室温下温育30分钟(Halpin,Lee等人,2004)。或者,所述肽在使用50mM乙酸酐的100mM 2-(N-吗啉代)乙磺酸盐(MES)缓冲液(pH6.0)和1M NaCl的水溶液中室温下乙酰化30分钟(Tse,Snyder等人2008)。对于NHS-乙酸酯衍生化反应,制备磺基-NHS-乙酸酯(在DMSO中100mM)的储备溶液,并以5-10mM的终浓度加入100mM磷酸钠缓冲液(pH8.0)或100mM硼酸盐缓冲液中(pH 9.4)并在RT孵育10-30分钟(Goodnow 2014)。

[0723] 酶促NTAA乙酰化

[0724] 肽的NTAA通过暴露于N-乙酰转移酶(SsArd1来自Sulfolobus solfataricus)酶促乙酰化,使用以下条件:将肽与在NAT缓冲液NAT缓冲液(20mM Tris-HCl,pH 8.0,100mM NaCl,1mM EDTA,1mM acetyl-CoA)的2μM SsArd1在65摄氏度孵育10分钟(Chang和Hsu 2015)。

[0725] 化学NTAA酰胺化(Guanidination)

[0726] 将肽与在DMF中的10mM N,N-双(叔丁氧基羰基)硫脲,20mM三甲胺和12mM Mukayama试剂(2-氯-1-甲基吡啶碘化物)在室温下孵育30分钟。或者,将肽与在DMF中的10mM 1H-吡唑-1-甲脒盐酸盐,10mM DIEA在室温下孵育30分钟。用标准解封闭方法去除保护基团。或者,将肽与PBS缓冲液(pH 8.0)或100mM硼酸盐缓冲液(pH 8.0)中的10mM S-甲基异硫脲一起在10摄氏度温育30分钟(Tse,Snyder et al.2008)。

[0727] PITC标签:

[0728] 将肽与离子液体[Bmim][BF4]中5%(vol./vol.)PITC在室温下孵育5分钟。对反应时间进行优化用于定量NTAA的PITC标记,同时最小化在延伸DNA记录标签中存在的核苷酸碱基上的环外胺的异位标记。

[0729] DNFB标签

[0730] 2,4-二硝基氟苯(DNFB)制备为5mg/ml原液存储于甲醇中。该溶液避光保存并当天使用。通过在10mM硼酸盐缓冲液(pH8.0)中的0.5-5.0μg/ml DNFB中于37℃温育5-30分钟来标记肽。

[0731] SNFB标签:

[0732] 4-磺酰基-2-硝基-氟苯(SNFB)制备为5mg/ml原液存储于甲醇中,该溶液应避光保存且当天使用。通过在10mM硼酸盐缓冲液(pH8.0)中的0.5-5.0μg/ml DNFB中于37℃温育5-30分钟来标记肽。

[0733] 乙酰化的NTAA肽的切割:

[0734] 通过与在25mM Tris-HCl(pH7.5)中的10μM酰基肽水解酶(APH)酶(来自Sulfolobus solfataricus,SSO2693)在90℃温育10分钟,从所述肽上切下所述乙酰化的NTAA(Gogliettino,Balestrieri et al.2012)。

[0735] 酰胺化NTAA肽的切割:

[0736] 通过在0.1N NaOH中在37℃温育10分钟,将所述酰胺化(胍基化)的NTAA从肽上切下(Hamada 2016)。

[0737] 示例23:用模型系统验证分子内编码标签信息到记录标签的转移

[0738] 用DNA模型系统测试编码标签信息到固定到珠子上的记录标签(参见图36A)的“分子内”转移。使用了两种不同类型的记录标签寡核苷酸。saRT_Abc_v2(SEQ ID NO:141)含有“A”DNA捕获序列(SEQ ID NO:153)(“A”结合剂的模拟表位)和相应的“A”条形码(rtA_BC);saRT_Bbc_V2(SEQ ID NO:142)含有“B”DNA捕获序列(SEQ ID NO:154)(“B”结合剂的模拟表位)和相应的“B”条形码(rtB_BC)。这些条形码是基本的65组15-mer条形码(SEQ ID NO:1-65)和它们的反向互补序列(SEQ ID NO:66-130)的组合。rtA_BC是两个条形码BC_1和BC_2的共线性组合,rtB_BC只是一个条形码BC_3。同样,所述编码标签上的所述条形码(编码器序列)也包括来自65个15-mer条形码的基本组的条形码(SEQ ID NO:1-65)。CT_A'-bc_1PEG(SEQ ID NO:144)和CT_B'-bc(SEQ ID NO:147)编码标签分别由互补捕获序列A'和B'组成,并分别被分配了所述15-mer条形码,BC_5,和BC_5&BC_6。这种所述记录标签和编码标签的设计可以轻松地进行凝胶分析。所需的“分子内”引物延伸产生相似大小的寡核苷酸产物,而不希望的“分子间”延伸产生一个比“分子内”产物大15个碱基的寡聚产物和另一个短15个碱基的寡聚产物(图36B)。

[0739] 评估了记录标签密度对“分子内”与“分子间”信息传递的影响。对于正确的信息传递,应该观察到“分子内”信息传递(“A”编码标签到A记录标签;B’编码标签到B记录标签),而不是“分子间”信息传递(A’编码标签结合到A记录而将信息传输到B记录标签,反之亦然)。为了测试记录标签间隔对珠子表面的影响,将生物素化的记录标签寡核苷酸saRT_Abc_v2(SEQ ID NO:141)和saRT_Bbc_v2(SEQ ID NO:142)以1:1的比例混合,然后以1:0,1:10,1:10^2,1:10^3,和1:10^4的比例滴定所述saDummy-T10寡核苷酸(SEQ ID NO:143)。将总共20pmol的记录标签寡核苷酸与在50μl固定化缓冲液(5mM Tris-Cl(pH 7.5),0.5mM EDTA,1M NaCl)中5ul M270链霉亲和素珠(Thermo)在37℃孵育15分钟。在室温下用100μl固定缓冲液洗涤珠子3次。大多数随后的洗涤步骤使用100μl的体积。通过将珠子重新悬浮在25μl 5X退火缓冲液(50mM Tris-Cl(pH 7.5),10mM $MgCl^2$)中并添加所述编码标签混合物,使编码标签(在后续循环中要求与DupCT序列的双重退火)退火至固定在珠子上的记录标签。通过加热至65℃1分钟将所述编码标签退火至所述记录标签,然后使其缓慢冷却至室温(0.2℃/sec)。或者,编码标签可以在37℃的PBST缓冲液中退火。在室温下用PBST(PBS+0.1%Tween-20)洗涤珠子,并在37℃用PBST洗涤2次,持续5分钟,在室温下用PBST洗涤1次,最后用退火缓冲液洗涤一次。将珠子重悬于19.5μl延伸缓冲液(50mM Tris-Cl(pH 7.5),2mM $MgSO_4$,125μMdNTP,50mM NaCl,1mM二硫苏糖醇,0.1%Tween-20和0.1mg/ml BSA)中,并在37℃温育15分钟。将Klenow exo-DNA聚合酶(NEB,5U/ul)加入到所述珠子中,最终浓度为0.125U/ul,并在37℃下孵育5分钟。引物延伸后,珠子用PBST洗涤2次,在室温下用50μl 0.1NaOH洗涤1次5分钟,用PBST洗涤3次,用PBS洗涤1次。为了添加下游PCR衔接序列R1’,将EndCap2T寡核苷酸(由R1(SEQ ID NO:152)组成)杂交在珠子上并延伸,如对所述编码标签寡核苷酸所操作。添加所述衔接序列后,最终的延伸记录标签寡核苷酸通过在95%甲酰胺/10mM EDTA中于65℃温育5分钟从所述链霉亲和素珠子中洗脱。将大约1/100的洗脱产物在20μl中PCR扩增18个循环,在10%变性PAGE凝胶上分析1μlPCR产物。得到的凝胶证明了通过聚合酶延伸将编码标签信息写入记录标签的原理的证据(图36C),以及在珠子表面记录标签密度稀释时产生相对于“分子间”延伸事件的主要“分子内”延伸事件的能力。

[0740] 在该模型系统中,来自包含相应编码序列和通用反向引物位点的记录标签RT_ABC和RT_BBC的PCR产物的大小是100个碱基对(图36C),而通过错误配对saRT_ABC(SEQ ID NO:141)/CT_B’BC(SEQ ID NO:147)和saRT_BBC(SEQ ID NO:142)/CT_A’BC(SEQ ID NO:144)的所述产物的大小分别是115和85个碱基对。如图36D所示,在珠子上存在以高密度的saRT_ABC(SEQ ID NO:141)和saRT_BBC(SEQ ID NO:142)时观察到三条带。预期所述重新编码标签在结合到它自己的近端编码标签上延伸(分子内事件),或在高密度的邻居重新编码标签上延伸(分子间事件)。然而,通过稀释虚拟寡核苷酸中的所述重新编码标签,由错误配对的产物条带减少,并在1:10000的比例时消失。该结果表明,所述记录标签以低密度在珠子表面上间隔开,导致分子间事件减少。

[0741] 表1.模型系统序列

[0742]

Name	Sequence (5'-3')	SEQ ID NO:
saRT_Abc_v2	/5Biosg/TTTTTGCAAATGGCATTCTGACATCCCGTAGTCCGCGACACTAGATGTCTAGCATGCCGCCGTGTCATGTGGAAACTGAGTG	141
saRT_Bbc_v2	/5Biosg/TTTTTTTTTTGACTGGTTCCAATTGACAAGCCGTAGTCCGCGACACTAGTAAGCCGGTATATCAACTGAGTG	142
saDummy-pT10	/5Biosg/TTTTTTTTTT/3SpC3/	143
CT_A'-bc	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGCACTCAGT/3SpC3/	144

[0743]

CT_A'-bc_1PEG	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/	145
CT_A'bc_5PEG	GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18//iSP18//iSP18//iSP18//iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/	146
CT_B'bc	GCTTGTCAATTGGAACCAGTCTTTT/iSp18/CACTCAGTCCTAACGCGTATACGGGAATCTCGGCAGTTCACTCAGT/3SpC3/	147
EndCap2T	CGATTTGCAAGGATCACTCGTCACTCAGTCCTAACGCGTATACG/3SpC3/	148
Sp	ACTGAGTG	149
Sp'	CACTCAGT	150
P1_f2	CGTAGTCCGCGACACTAG	151
R1	CGATTTGCAAGGATCACTCG	152
dupCT_A'BC	CGTATACGCGTTAGGACTGAGTG/3SpC3/	153
dupCT_B'BC	AACTGCCGAGATTCCCGTATACGCGTTAGGACTGAGTG/3SpC3/	154

[0744] /3SpC3/＝3'C3(三碳)间隔子3'C3

[0745] /5Biosg/＝5'生物素

[0746] /iSP18/＝18-原子hexa-乙二醇间隔子

[0747] 示例24:对延伸记录标签,延伸编码标记或DI-TAG构建体的纳米测序DNA条形码可以设计为耐受高度易错的NGS测序仪,例如基于纳米孔的测序仪,其中当前碱基调用错误率大约为10%或更高。许多错误纠错码系统已经在文献中描述。这些包括Hamming码,Reed-Solomon码,Levenshtein码,Lee码,等。容错条形码基于使用R Bioconductor包的汉明码和Levenshtein码,"ve条码"能够纠正插入、删除、和替换错误,取决于所选择的设计参数(Buschmann和Bystrykh 2013)。图27A中显示了一组65种不同的15-mer汉明条形码(如SEQ ID NO:1-65所示,它们的反向互补序列分别在SEQ ID NO:66-130中)。这些条形码的最小汉明距离为10,并且自校正到四个替换误差和两个插入错误,足以在具有10%错误率的纳米孔测序仪上准确读出。此外,使用预测的纳米孔电流特征从一组77个原始条形码中过筛选得到这些条形码(参见图27B)。它们被过滤以在条形码之间上具有大的电流水平差异,并且与该组中的其他条形码最大程度地不相关。通过这种方式,使用这些条形码的分析的实际

原始纳米孔电流水平图可以直接映射到预测的条形码特征,而无需使用碱基调用算法(Laszlo,Derrington et al.2014)。

[0748] 为了模拟使用纳米孔测序的延伸记录标签、延伸编码标签、或di-tag构建体的分析,使用四个正向引物(DTF1(SEQ ID NO:157),DTF2(SEQ ID NO:158),DTF3(SEQ ID NO:159),DTF4(SEQ ID NO:160))和四种反向引物(DTR9(SEQ ID NO:161),DTR10(SEQ ID NO:162),DTR11(SEQ ID NO:163),DTR12(SEQ ID NO:164))生成了包含15-mer条形码小子集的PCR产物(图27C)。将这组8个引物与侧翼正向引物F1(SEQ ID NO:165)和反向引物R1(SEQ ID NO:166)一起包括在PCR反应中。所述DTF和DTR引物通过互补的15-mer间隔序列(Sp15)(SEQ ID NO:167)退火。4个DTF正向和4个DTR反向引物的组合产生一组16种可能的PCR产物。

[0749] PCR条件:

[0750]

试剂	终浓度
F1(5'磷酸化的)(SEQ ID NO: 165)	1 uM
R1(5'磷酸化的)(SEQ ID NO: 166)	1 uM
DTF1-4 (SEQ ID NOS:157-160); DTR9-12 (SEQ ID NOS:161-164)	0.3 nM ea
VeraSeq 缓冲液 2	1X
dNTPs	200 uM
水	
VeraSeq 2.0 Ultra Pol	2 U/100 ul

[0751] PCR循环

[0752]

98 °C	30 秒
50 °C	2 分钟
98 °C	10 秒
55 °C	15 秒
72 °C	15 秒
重复后 3 个步骤 19 个循环	
72 °C	5 分钟

[0753] PCR后,扩增子通过平末端连接(图27C)级联如下:将20ul PCR产物直接与20ul Quick Ligase Mix(NEB)混合,并在室温下温育过夜。使用Zymo纯化柱纯化得到的长度约0.5-2kb的连接产物,并洗脱到20ul水中。将约7约。纯化的连接产物直接用于MinIon Library Rapid Sequencing Prep kit(SQK-RAD002),并在MinION Mk 1B(R9.4)装置上分析。图27D中显示了质量得分7.2(约80%准确度)的734bp纳米孔读数的实例。尽管测序精确度差,但是序列中的大量条形码易于读取,如通过条形码到MinIon序列读数的基于lalign的对准所示(图27D)。

[0754] 示例25:凝胶珠中包封的单个细胞

[0755] 使用标准技术将单细胞包封成液滴(~50μm)(Tamminen和Virta 2015,Spencer,Tamminen et al.2016)(见图38)。聚丙烯酰胺(丙烯酰胺:双丙烯酰胺(29:1)(30%w/vol.))、二苯甲酮甲基丙烯酰胺(BM),和APS与细胞一起包含在非连续相中以产生在连续油相(扩散成液滴)能够在加入TEMED后聚合的液滴。二苯甲酮交联到聚丙烯酰胺凝胶液滴的阵列中。这允许所述蛋白质随后与所述聚丙烯酰胺阵列的光亲和交联(Hughes,Spelke et al.2014,Kang,Yamauchi et al.2016)。固定在所得单细胞凝胶珠内的所述蛋白质可以使用多种方法进行单细胞条形码编码。在一个实施方案中,使用如前所述的胺反应剂或光活性二苯甲酮DNA标签,将DNA标签化学或光化学连接至单细胞凝胶珠中的固定的蛋白上。单细胞凝胶珠可以通过如前所述的条形码珠的共同包封包封在含有条形码的液滴中,并且所述DNA条形码标签转移到蛋白上,或者单细胞凝胶珠中的蛋白可以通过一系列如Amini,Cusanovich和Gunderson等人描述的合并-分开步骤组合索引(Amini,Pushkarev et al.2014,Cusanovich,Daza et al.2015)(Gunderson,Steemers et al.2016)。在最简单的实施中,单细胞凝胶珠中的所述蛋白首先用“点击化学”基团标记(参见图40),然后使用所述合并-分开方法将组合DNA条形码点击到所述蛋白样品上。

[0756] 参考文献:

[0757] Harlow,Ed和David Lane.Using Antibodies.Cold Spring Harbor,New York:Cold Spring Harbor Laboratory Press,1999.

[0758] Hennessy BT,Lu Y,Gonzalez-Angulo AM et al.A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers.Clinical proteomics.2010;6(4):129-151.

[0759] Davidson,G.R.,S.D.Armstrong and R.J.Beynon(2011)."Positional proteomics at the N-terminus as a means of proteome simplification."Methods Mol Biol 753:229-242.

[0760] Zhang,L.,Luo,S.,和Zhang,B.(2016).The use of lectin microarray for assessing glycosylation of therapeutic proteins.mAbs 8,524-535.

[0761] Akbani,R.,K.F.Becker,N.Carragher,T.Goldstein,L.de Koning,U.Korf,L.Liotta,G.B.Mills,S.S.Nishizuka,M.Pawlak,E.F.Petricoin,3rd,H.B.Pollard,B.Serrels和J.Zhu(2014)."Realizing the promise of reverse phase protein arrays for clinical,translational,and basic research:a workshop report:the RPPA(Reverse Phase Protein Array)society."Mol Cell Proteomics 13(7):1625-1643.

[0762] Amini,S.,D.Pushkarev,L.Christiansen,E.Kostem,T.Royce,C.Turk,N.Pignatelli,A.Adey,J.O.Kitzman,K.Vijayan,M.Ronaghi,J.Shendure,K.L.Gunderson和F.J.Steemers(2014)."Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing."Nat Genet 46(12):1343-1349.

[0763] Assadi,M.,J.Lamerz,T.Jarutat,A.Farfsing,H.Paul,B.Gierke,E.Breitinger,M.F.Templin,L.Essioux,S.Arbogast,M.Venturi,M.Pawlak,H.Langen和T.Schindler

(2013)."Multiple protein analysis of formalin-fixed and paraffin-embedded tissue samples with reverse phase protein arrays."Mol Cell Proteomics 12(9):2615-2622.

[0764] Bailey,J.M.和J.E.Shively(1990)."Carboxy-terminal sequencing:formation and hydrolysis of C-terminal peptidylthiohydantoins."Biochemistry29(12):3145-3156.

[0765] Bandara,H.M.,D.P.Kennedy,E.Akin,C.D.Incarvito和S.C.Burdette(2009)."Photoinduced release of $Zn^{2+}$with ZinCleav-1:a nitrobenzyl-based caged complex."Inorg Chem 48(17):8445-8455.

[0766] Bandara,H.M.,T.P.Walsh和S.C.Burdette(2011)."A Second-generation photocage for Zn2+inspired by TPEN:characterization and insight into the uncaging quantum yields of ZinCleav chelators."Chemistry 17(14):3932-3941.

[0767] Basle,E.,N.Joubert和M.Pucheault(2010)."Protein chemical modification on endogenous amino acids."Chem Biol 17(3):213-227.

[0768] Bilgicer,B.,S.W.Thomas,3rd,B.F.Shaw,G.K.Kaufman,V.M.Krishnamurthy,L.A.Estroff,J.Yang和G.M.Whitesides(2009)."A non-chromatographic method for the purification of a bivalently active monoclonal IgG antibody from biological fluids."J Am Chem Soc 131(26):9361-9367.

[0769] Bochman,M.L.,K.Paeschke和V.A.Zakian(2012)."DNA secondary structures:stability and function of G-quadruplex structures."Nat Rev Genet13(11):770-780.

[0770] Borgo,B.和J.J.Havranek(2014)."Motif-directed redesign of enzyme specificity."Protein Sci 23(3):312-320.

[0771] Brouzes,E.,M.Medkova,N.Savenelli,D.Marran,M.Twardowski,J.B.Hutchison,J.M.Rothberg,D.R.Link,N.Perrimon和M.L.Samuels(2009)."Droplet microfluidic technology for single-cell high-throughput screening."Proc Natl Acad Sci U S A 106(34):14195-14200.

[0772] Brudno,Y.,M.E.Birnbaum,R.E.Kleiner和D.R.Liu(2010)."An in vitro translation,selection and amplification system for peptide nucleic acids."Nat Chem Biol 6(2):148-155.

[0773] Calcagno,S.和C.D.Klein(2016)."N-Terminal methionine processing by the zinc-activated Plasmodium falciparum methionine aminopeptidase 1b."Appl Microbiol Biotechnol.

[0774] Cao,Y.,G.K.Nguyen,J.P.Tam和C.F.Liu(2015)."Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation."Chem Commun(Camb)51(97):17289-17292.

[0775] Carty,R.P.and C.H.Hirs(1968)."Modification of bovine pancreatic ribonuclease A with 4-sulfonyloxy-2-nitrofluorobenzene.Isolation and identification of modified proteins."J Biol Chem 243(20):5244-5253.

[0776] Chang,L.,D.M.Rissin,D.R.Fournier,T.Piech,P.P.Patel,D.H.Wilson和D.C.Duffy(2012)."Single molecule enzyme-linked immunosorbent assays: theoretical considerations."J Immunol Methods 378(1-2):102-115.

[0777] Chang,Y.Y.and C.H.Hsu(2015)."Structural basis for substrate-specific acetylation of Nalpha-acetyltransferase Ard1 from Sulfolobus solfataricus."Sci Rep 5:8673.

[0778] Christoforou,A.,C.M.Mulvey,L.M.Breckels,A.Geladaki,T.Hurrell,P.C.Hayward,T.Naake,L.Gatto,R.Viner,A.Martinez Arias和K.S.Lilley(2016)."A draft map of the mouse pluripotent stem cell spatial proteome."Nat Commun 7:8992.

[0779] Creighton,C.J.和S.Huang(2015)."Reverse phase protein arrays in signaling pathways:a data integration perspective."Drug Des Devel Ther 9:3519-3527.

[0780] Crosetto,N.,M.Bienko和A.van Oudenaarden(2015)."Spatially resolved transcriptomics and beyond."Nat Rev Genet 16(1):57-66.

[0781] Cusanovich,D.A.,R.Daza,A.Adey,H.A.Pliner,L.Christiansen,K.L.Gunderson,F.J.Steemers,C.Trapnell和J.Shendure(2015)."Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing."Science 348(6237):910-914.

[0782] Derrington,I.M.,T.Z.Butler,M.D.Collins,E.Manrao,M.Pavlenok,M.Niederweis和J.H.Gundlach(2010)."Nanopore DNA sequencing with MspA."Proc Natl Acad Sci U S A 107(37):16060-16065.

[0783] El-Sagheer,A.H.,V.V.Cheong和T.Brown(2011)."Rapid chemical ligation of oligonucleotides by the Diels-Alder reaction."Org Biomol Chem 9(1):232-235.

[0784] El-Sagheer,A.H.,A.P.Sanzone,R.Gao,A.Tavassoli和T.Brown(2011)."Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli."Proc Natl Acad Sci U S A 108(28):11338-11343.

[0785] Emili,A.,M.McLaughlin,K.Zagorovsky,J.B.Olsen,W.C.W.Chan和S.S.Sidhu(2017).Protein Sequencing Method and Reagents.USPTO.USA,The Governing Council of University of Toronto.9,566,335B1.

[0786] Erde,J.,R.R.Loo和J.A.Loo(2014)."Enhanced FASP(eFASP)to increase proteome coverage and sample recovery for quantitative proteomic experiments."J Proteome Res 13(4):1885-1895.

[0787] Farries,T.C.,A.Harris,A.D.Auffret和A.Aitken(1991)."Removal of N-acetyl groups from blocked peptides with acylpeptide hydrolase.Stabilization of the enzyme and its application to protein sequencing."Eur J Biochem 196(3):679-685.

[0788] Feist,P.和A.B.Hummon(2015)."Proteomic challenges:sample preparation

techniques for microgram-quantity protein analysis from biological samples." Int J Mol Sci 16(2):3537-3563.

[0789] Friedmann,D.R.和R.Marmorstein(2013)."Structure and mechanism of non-histone protein acetyltransferase enzymes."FEBS J 280(22):5570-5581.

[0790] Frokjaer,S.和D.E.Otzen(2005)."Protein drug stability:a formulation challenge."Nat Rev Drug Discov 4(4):298-306.

[0791] Fujii,Y.,M.Kaneko,M.Neyazaki,T.Nogi,Y.Kato和J.Takagi(2014)."PA tag:a versatile protein tagging system using a super high affinity antibody against a dodecapeptide derived from human podoplanin."Protein Expr Purif 95:240-247.

[0792] Gebauer,M.和A.Skerra(2012)."Anticalins small engineered binding proteins based on the lipocalin scaffold."Methods Enzymol 503:157-188.

[0793] Gerry,N.P.,N.E.Witowski,J.Day,R.P.Hammer,G.Barany和F.Barany(1999)."Universal DNA microarray method for multiplex detection of low abundance point mutations."J Mol Biol 292(2):251-262.

[0794] Gogliettino,M.,M.Balestrieri,E.Cocca,S.Mucerino,M.Rossi,M.Petrillo,E.Mazzella和G.Palmieri(2012)."Identification and characterisation of a novel acylpeptide hydrolase from Sulfolobus solfataricus:structural and functional insights."PLoS One 7(5):e37921.

[0795] Gogliettino,M.,A.Riccio,M.Balestrieri,E.Cocca,A.Facchiano,T.M.D'Arco,C.Tesoro,M.Rossi和G.Palmieri(2014)."A novel class of bifunctional acylpeptide hydrolases--potential role in the antioxidant defense systems of the Antarctic fish Trematomus bernacchii."FEBS J 281(1):401-415.

[0796] Granvogl,B.,M.Ploscher和L.A.Eichacker(2007)."Sample preparation by in-gel digestion for mass spectrometry-based proteomics."Anal Bioanal Chem389(4):991-1002.

[0797] Gunderson,K.L.,X.C.Huang,M.S.Morris,R.J.Lipshutz,D.J.Lockhart and M.S.Chee(1998)."Mutation detection by ligation to complete n-mer DNA arrays."Genome Res 8(11):1142-1153.

[0798] Gunderson,K.L.,F.J.Steemers,J.S.Fisher和R.Rigatti(2016).Methods and Compositions for Analyzing Cellular Components.WIPO,Illumina,Inc.

[0799] Gunderson,K.L.,F.J.Steemers,J.S.Fisher和R.Rigatti(2016).Methods and compositions for analyzing cellular components,Illumina,Inc.

[0800] Guo,H.,W.Liu,Z.Ju,P.Tamboli,E.Jonasch,G.B.Mills,Y.Lu,B.T.Hennessy和D.Tsavachidou(2012)."An efficient procedure for protein extraction from formalin-fixed,paraffin-embedded tissues for reverse phase protein arrays."Proteome Sci 10(1):56.

[0801] Hamada,Y.(2016)."A novel N-terminal degradation reaction of peptides via N-amidination."Bioorg Med Chem Lett 26(7):1690-1695.

[0802] Hermanson,G.(2013).Bioconjugation Techniques,Academic Press.

[0803] Hernandez-Moreno,A.V.,F.Villasenor,E.Medina-Rivero,N.O.Perez,L.F.Flores-Ortiz,G.Saab-Rincon和G.Luna-Barcenas(2014)."Kinetics and conformational stability studies of recombinant leucine aminopeptidase."Int J Biol Macromol 64:306-312.

[0804] Hori,M.,H.Fukano和Y.Suzuki(2007)."Uniform amplification of multiple DNAs by emulsion PCR."Biochem Biophys Res Commun 352(2):323-328.

[0805] Horisawa,K.(2014)."Specific and quantitative labeling of biomolecules using click chemistry."Front Physiol 5:457.

[0806] Hoshika,S.,F.Chen,N.A.Leal和S.A.Benner(2010)."Artificial genetic systems:self-avoiding DNA in PCR and multiplexed PCR."Angew Chem Int Ed Engl 49(32):5554-5557.

[0807] Hughes,A.J.,D.P.Spelke,Z.Xu,C.C.Kang,D.V.Schaffer和A.E.Herr(2014)."Single-cell western blotting."Nat Methods 11(7):749-755.

[0808] Hughes,C.S.,S.Foehr,D.A.Garfield,E.E.Furlong,L.M.Steinmetz和J.Krijgsveld(2014)."Ultrasensitive proteome analysis using paramagnetic bead technology."Mol Syst Biol 10:757.

[0809] Kang,C.C.,K.A.Yamauchi,J.Vlassakis,E.Sinkala,T.A.Duncombe和A.E.Herr(2016)."Single cell-resolution western blotting."Nat Protoc 11(8):1508-1530.

[0810] Kang,T.S.,L.Wang,C.N.Sarkissian,A.Gamez,C.R.Scriver和R.C.Stevens(2010)."Converting an injectable protein therapeutic into an oral form:phenylalanine ammonia lyase for phenylketonuria."Mol Genet Metab 99(1):4-9.

[0811] Katritzky,A.R.和B.V.Rogovoy(2005)."Recent developments in guanylating agents."ARKIVOC iv(Issue in Honor of Prof.Nikolai Zefirov):49-87.

[0812] Klein,A.M.,L.Mazutis,I.Akartuna,N.Tallapragada,A.Veres,V.Li,L.Peshkin,D.A.Weitz和M.W.Kirschner(2015)."Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells."Cell 161(5):1187-1201.

[0813] Knall,A.C.,M.Hollauf和C.Slugovc(2014)."Kinetic studies of inverse electron demand Diels-Alder reactions(iEDDA)of norbornenes and 3,6-dipyridin-2-yl-1,2,4,5-tetrazine."Tetrahedron Lett 55(34):4763-4766.

[0814] Le,Z.G.,Z.C.Chen,Y.Hu和Q.G.Zheng(2005)."Organic Reactions in Ionic Liquids:Ionic Liquid-promoted Efficient Synthesis of Disubstituted and Trisubstituted Thioureas Derivatives."Chinese Chemical Letters 16(2):201-204.

[0815] Lesch,V.,A.Heuer,V.A.Tatsis,C.Holm和J.Smiatek(2015)."Peptides in the presence of aqueous ionic liquids:tunable co-solutes as denaturants or protectants?"Phys Chem Chem Phys 17(39):26049-26053.

[0816] Li,G.,Y.Liu,Y.Liu,L.Chen,S.Wu,Y.Liu和X.Li(2013)."Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry."Angew Chem Int Ed Engl 52(36):9544-9549.

[0817] Litovchick,A.,M.A.Clark和A.D.Keefe(2014)."Universal strategies for

the DNA-encoding of libraries of small molecules using the chemical ligation of oligonucleotide tags."Artif DNA PNA XNA 5(1):e27896.

[0818] Liu,Y.和S.Liang(2001)."Chemical carboxyl-terminal sequence analysis of peptides and proteins using tribenzylsilyl isothiocyanate."J Protein Chem 20(7):535-541.

[0819] Lundblad,R.L.(2014).Chemical reagents for protein modification.Boca Raton,CRC Press,Taylor&Francis Group.

[0820] Mashaghi,S.和A.M.van Oijen(2015)."External control of reactions in microdroplets."Sci Rep 5:11837.

[0821] McCormick,R.M.(1989)."A solid-phase extraction procedure for DNA purification."Anal Biochem 181(1):66-74.

[0822] Mendoza,V.L.和R.W.Vachet(2009)."Probing protein structure by amino acid-specific covalent labeling and mass spectrometry."Mass Spectrom Rev 28(5):785-815.

[0823] Mikami,T.,T.Takao,K.Yanagi和H.Nakazawa(2012)."N(alpha)Selective Acetylation of Peptides."Mass Spectrom(Tokyo)1(2):A0010.

[0824] Moghaddam,M.J.,L.de Campo,N.Kirby和C.J.Drummond(2012)."Chelating DTPA amphiphiles:ion-tunable self-assembly structures and gadolinium complexes."Phys Chem Chem Phys 14(37):12854-12862.

[0825] Mukherjee,S.,M.Ura,R.J.Hoey和A.A.Kossiakoff(2015)."A New Versatile Immobilization Tag Based on the Ultra High Affinity and Reversibility of the Calmodulin-Calmodulin Binding Peptide Interaction."J Mol Biol 427(16):2707-2725.

[0826] Namimatsu,S.,M.Ghazizadeh和Y.Sugisaki(2005)."Reversing the effects of formalin fixation with citraconic anhydride and heat:a universal antigen retrieval method."J Histochem Cytochem 53(1):3-11.

[0827] Nguyen,G.K.,Y.Cao,W.Wang,C.F.Liu和J.P.Tam(2015)."Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1and Thiodepsipeptide."Angew Chem Int Ed Engl 54(52):15694-15698.

[0828] Nguyen,G.K.,S.Wang,Y.Qiu,X.Hemu,Y.Lian和J.P.Tam(2014)."Butelase 1is an Asx-specific ligase enabling peptide macrocyclization and synthesis."Nat Chem Biol 10(9):732-738.

[0829] Nishizuka,S.S.和G.B.Mills(2016)."New era of integrated cancer biomarker discovery using reverse-phase protein arrays."Drug Metab Pharmacokinet 31(1):35-45.

[0830] Ohkubo,A.,R.Kasuya,K.Sakamoto,K.Miyata,H.Taguchi,H.Nagasawa,T.Tsukahara,T.Watanobe,Y.Maki,K.Seio和M.Sekine(2008)."Protected DNA Probes'capable of strong hybridization without removal of base protecting groups."Nucleic Acids Res 36(6):1952-1964.

[0831] Ojha,B.,A.K.Singh,M.D.Adhikari,A.Ramesh和G.Das(2010)."2-Alkylmalonic acid:amphiphilic chelator and a potent inhibitor of metalloenzyme."J Phys Chem B 114(33):10835-10842.

[0832] Peng,X.,H.Li和M.Seidman(2010)."A Template-Mediated Click-Click Reaction:PNA-DNA,PNA-PNA(or Peptide)Ligation,and Single Nucleotide Discrimination."European J Org Chem 2010(22):4194-4197.

[0833] Perbandt,M.,O.Bruns,M.Vallazza,T.Lamla,C.Betzel和V.A.Erdmann(2007)."High resolution structure of streptavidin in complex with a novel high affinity peptide tag mimicking the biotin binding motif."Proteins 67(4):1147-1153.

[0834] Rauth,S.,D.Hinz,M.Borger,M.Uhrig,M.Mayhaus,M.Riemenschneider和A.Skerra(2016)."High-affinity Anticalins with aggregation-blocking activity directed against the Alzheimer beta-amyloid peptide."Biochem J 473(11):1563-1578.

[0835] Ray,A.和B.Norden(2000)."Peptide nucleic acid(PNA):its medical and biotechnical applications and promise for the future."FASEB J 14(9):1041-1060.

[0836] Riley,N.M.,A.S.Hebert和J.J.Coon(2016)."Proteomics Moves into the Fast Lane."Cell Syst 2(3):142-143.

[0837] Roloff,A.,S.Ficht,C.Dose和O.Seitz(2014)."DNA-templated native chemical ligation of functionalized peptide nucleic acids:a versatile tool for single base-specific detection of nucleic acids."Methods Mol Biol 1050:131-141.

[0838] Roloff,A.和O.Seitz(2013)."The role of reactivity in DNA templated native chemical PNA ligation during PCR."Bioorg Med Chem 21(12):3458-3464.

[0839] Sakurai,K.,T.M.Snyder和D.R.Liu(2005)."DNA-templated functional group transformations enable sequence-programmed synthesis using small-molecule reagents."J Am Chem Soc 127(6):1660-1661.

[0840] Schneider,K.和B.T.Chait(1995)."Increased stability of nucleic acids containing 7-deaza-guanosine and 7-deaza-adenosine may enable rapid DNA sequencing by matrix-assisted laser desorption mass spectrometry."Nucleic Acids Res 23(9):1570-1575.

[0841] Selvaraj,R.和J.M.Fox(2013)."trans-Cyclooctene--a stable,voracious dienophile for bioorthogonal labeling."Curr Opin Chem Biol 17(5):753-760.

[0842] Sharma,A.K.,A.D.Kent和J.M.Heemstra(2012)."Enzyme-linked small-moleculedetection using split aptamer ligation."Anal Chem 84(14):6104-6109.

[0843] Shembekar,N.,C.Chaipan,R.Utharala和C.A.Merten(2016)."Droplet-based microfluidics in drug discovery,transcriptomics and high-throughput molecular genetics."Lab Chip 16(8):1314-1331.

[0844] Shenoy,N.R.,J.E.Shively和J.M.Bailey(1993)."Studies in C-terminal sequencing:new reagents for the synthesis of peptidylthiohydantoins."J Protein Chem 12(2):195-205.

[0845] Shim,J.U.,R.T.Ranasinghe,C.A.Smith,S.M.Ibrahim,F.Hollfelder,W.T.Huck,D.Klenerman和C.Abell(2013)."Ultrarapid generation of femtoliter microfluidic droplets for single-molecule-counting immunoassays."ACS Nano 7(7):5955-5964.

[0846] Shim,J.W.,Q.Tan和L.Q.Gu(2009)."Single-moleculedetection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity."Nucleic Acids Res 37(3):972-982.

[0847] Sidoli,S.,Z.F.Yuan,S.Lin,K.Karch,X.Wang,N.Bhanu,A.M.Arnaudo,L.M.Britton,X.J.Cao,M.Gonzales-Cope,Y.Han,S.Liu,R.C.Molden,S.Wein,L.Afjehi-Sadat和B.A.Garcia(2015)."Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis."Proteomics 15(9):1459-1469.

[0848] Sletten,E.M.和C.R.Bertozzi(2009)."Bioorthogonal chemistry:fishing for selectivity in a sea of functionality."Angew Chem Int Ed Engl 48(38):6974-6998.

[0849] Spencer,S.J.,M.V.Tamminen,S.P.Preheim,M.T.Guo,A.W.Briggs,I.L.Brito,A.W.D,L.K.Pitkanen,F.Vigneault,M.P.Juhani Virta和E.J.Alm(2016)."Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers."ISME J 10(2):427-436.

[0850] Spicer,C.D.和B.G.Davis(2014)."Selective chemical protein modification."Nat Commun 5:4740.

[0851] Spiropulos,N.G.和J.M.Heemstra(2012)."Templating effect in DNA proximity ligation enables use of non-bioorthogonal chemistry in biological fluids."Artif DNA PNA XNA 3(3):123-128.

[0852] Switzar,L.,M.Giera和W.M.Niessen(2013)."Protein digestion:an overview of the available techniques and recent developments."J Proteome Res12(3):1067-1077.

[0853] Tamminen,M.V.和M.P.Virta(2015)."Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells."Front Microbiol 6:195.

[0854] Tessler,L.(2011).Digital Protein Analysis:Technologies for Protein Diagnostics and Proteomics through Single-Molecule Detection.Ph.D.,WASHINGTON UNIVERSITY IN ST.LOUIS.

[0855] Tyson,J.和J.A.Armour(2012)."Determination of haplotypes at structurally complex regions using emulsion haplotype fusion PCR."BMC Genomics 13:693.

[0856] Vauquelin,G.和S.J.Charlton(2013)."Exploring avidity:understanding the

potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands."Br J Pharmacol 168(8):1771-1785.

[0857] Veggiani,G.,T.Nakamura,M.D.Brenner,R.V.Gayet,J.Yan,C.V.Robinson和M.Howarth(2016)."Programmable polyproteams built using twin peptide superglues."Proc Natl Acad Sci U S A 113(5):1202-1207.

[0858] Wang,D.,S.Fang和R.M.Wohlhueter(2009)."N-terminal derivatization of peptides with isothiocyanate analogues promoting Edman-type cleavage and enhancing sensitivity in electrospray ionization tandem mass spectrometry analysis."Anal Chem 81(5):1893-1900.

[0859] Williams,B.A.和J.C.Chaput(2010)."Synthesis of peptide-oligonucleotide conjugates using a heterobifunctional crosslinker."Curr Protoc Nucleic Acid Chem Chapter 4:Unit4 41.

[0860] Wu,H.和N.K.Devaraj(2016)."Inverse Electron-Demand Diels-Alder Bioorthogonal Reactions."Top Curr Chem(J)374(1):3.

[0861] Xiong,A.S.,R.H.Peng,J.Zhuang,F.Gao,Y.Li,Z.M.Cheng和Q.H.Yao(2008)."Chemical gene synthesis:strategies,softwares,error corrections,and applications."FEMS Microbiol Rev 32(3):522-540.

[0862] Yao,Y.,M.Docter,J.van Ginkel,D.de Ridder和C.Joo(2015)."Single-molecule protein sequencing through fingerprinting:computational assessment."Phys Biol 12(5):055003.

[0863] Zakeri,B.,J.O.Fierer,E.Celik,E.C.Chittock,U.Schwarz-Linek,V.T.Moy和M.Howarth(2012)."Peptide tag forming a rapid covalent bond to a protein,through engineering a bacterial adhesin."Proc Natl Acad Sci U S A 109(12):E690-697.

[0864] Zhang,L.,K.Zhang,S.Rauf,D.Dong,Y.Liu和J.Li(2016)."Single-Molecule Analysis of Human Telomere Sequence Interactions with G-quadruplex Ligand."Anal Chem 88(8):4533-4540.

[0865] Zhou,H.,Z.Ning,A.E.Starr,M.Abu-Farha和D.Figeys(2012)."Advancements in top-down proteomics."Anal Chem 84(2):720-734.

[0866] Zilionis,R.,J.Nainys,A.Veres,V.Savova,D.Zemmour,A.M.Klein和L.Mazutis(2017)."Single-cell barcoding and sequencing using droplet microfluidics."Nat Protoc 12(1):44-73.

[0867] 可以对根据以上详细描述的实施例进行这些和其他改变。通常,在以下权利要求中,所使用的术语不应被解释为将权利要求限制于说明书和权利要求中公开的特定实施例,而是应该被解释权利要求有权要求包括所有可能的实施例以及全部等同范围,因此,权利要求不限于本公开的。

[0868] 可以组合上述各种实施例以提供进一步的实施例。本说明书中提及和/或在申请数据表中列出的所有美国专利,美国专利申请公开,美国专利申请,外国专利,外国专利申请和非专利出版物,包括美国临时专利申请No.62/330,841美国临时专利申请号62/339,

071和美国临时专利申请号62/376,886通过引用整体并入本文,通过引用整体并入本文。如果需要,可以修改实施例的各方面以采用所述各种专利、申请和出版物的概念来提供其它进一步实施例。

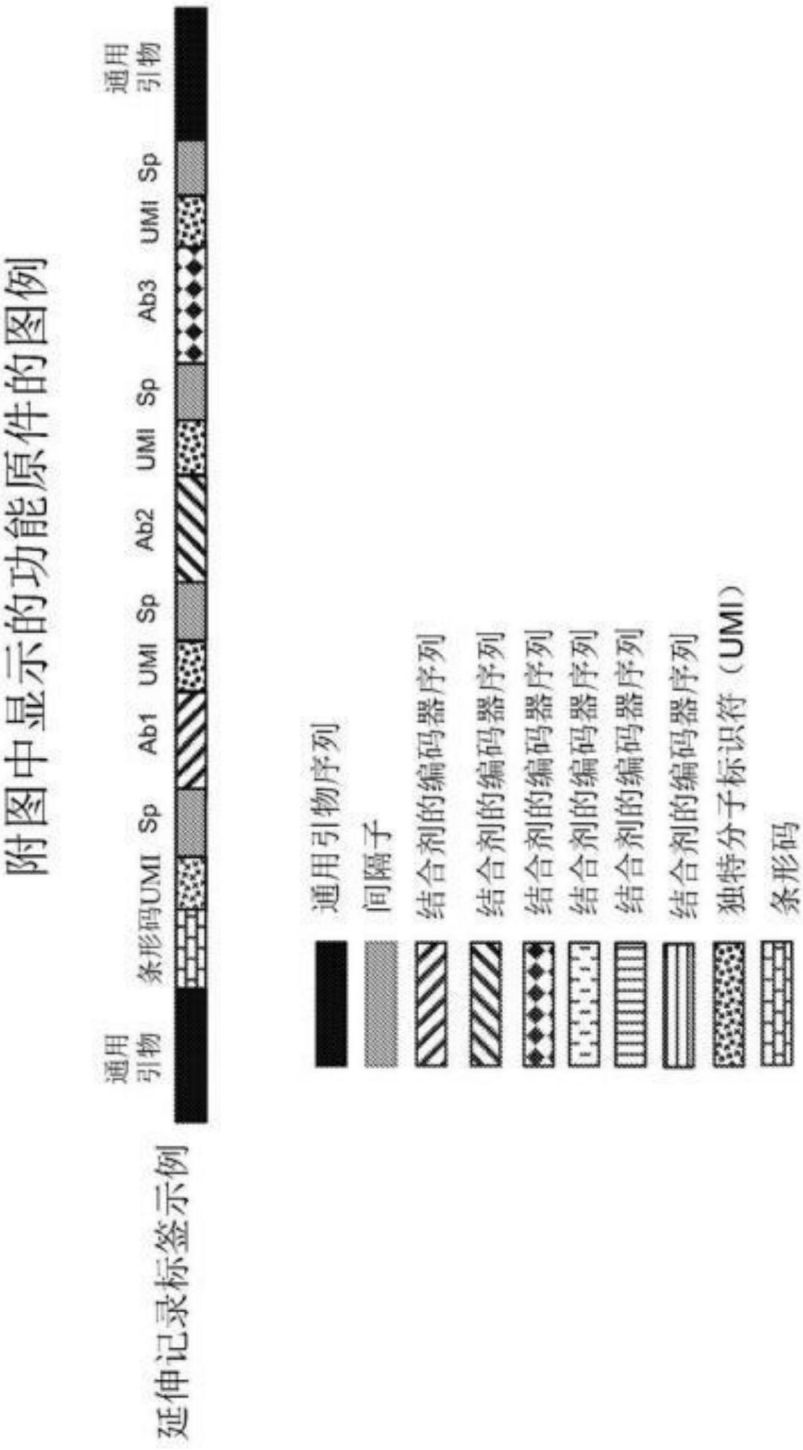

附图中显示的功能原件的图例
延伸记录标签示例
通用引物
条形码UMI
Sp
Ab1
UMI
Sp
Ab2
UMI
Sp
Ab3
UMI
Sp
通用引物
通用引物序列
间隔子
结合剂的编码器序列
结合剂的编码器序列
结合剂的编码器序列
结合剂的编码器序列
结合剂的编码器序列
结合剂的编码器序列
独特分子标识符（UMI）
条形码

图1A

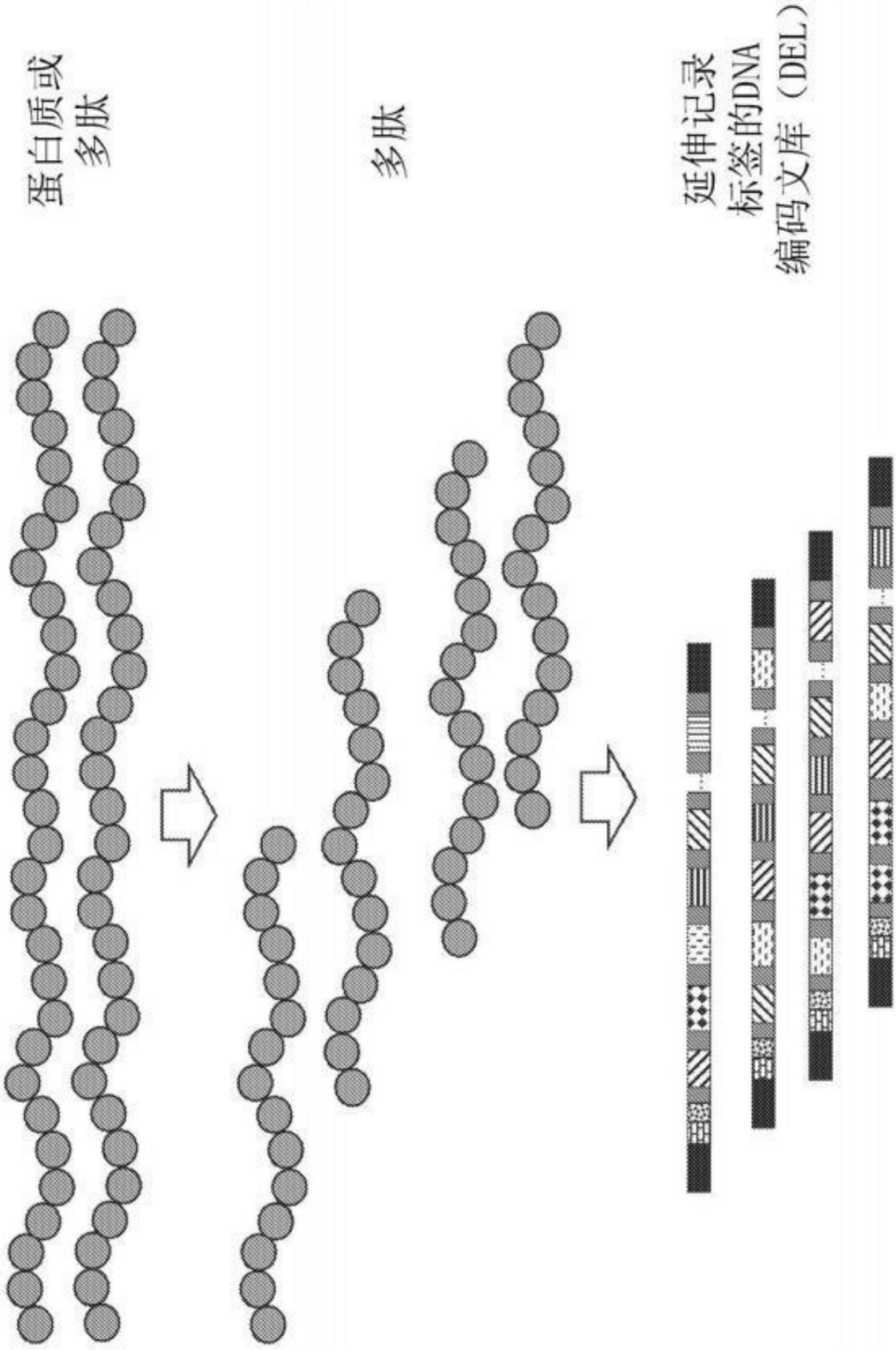


图1B

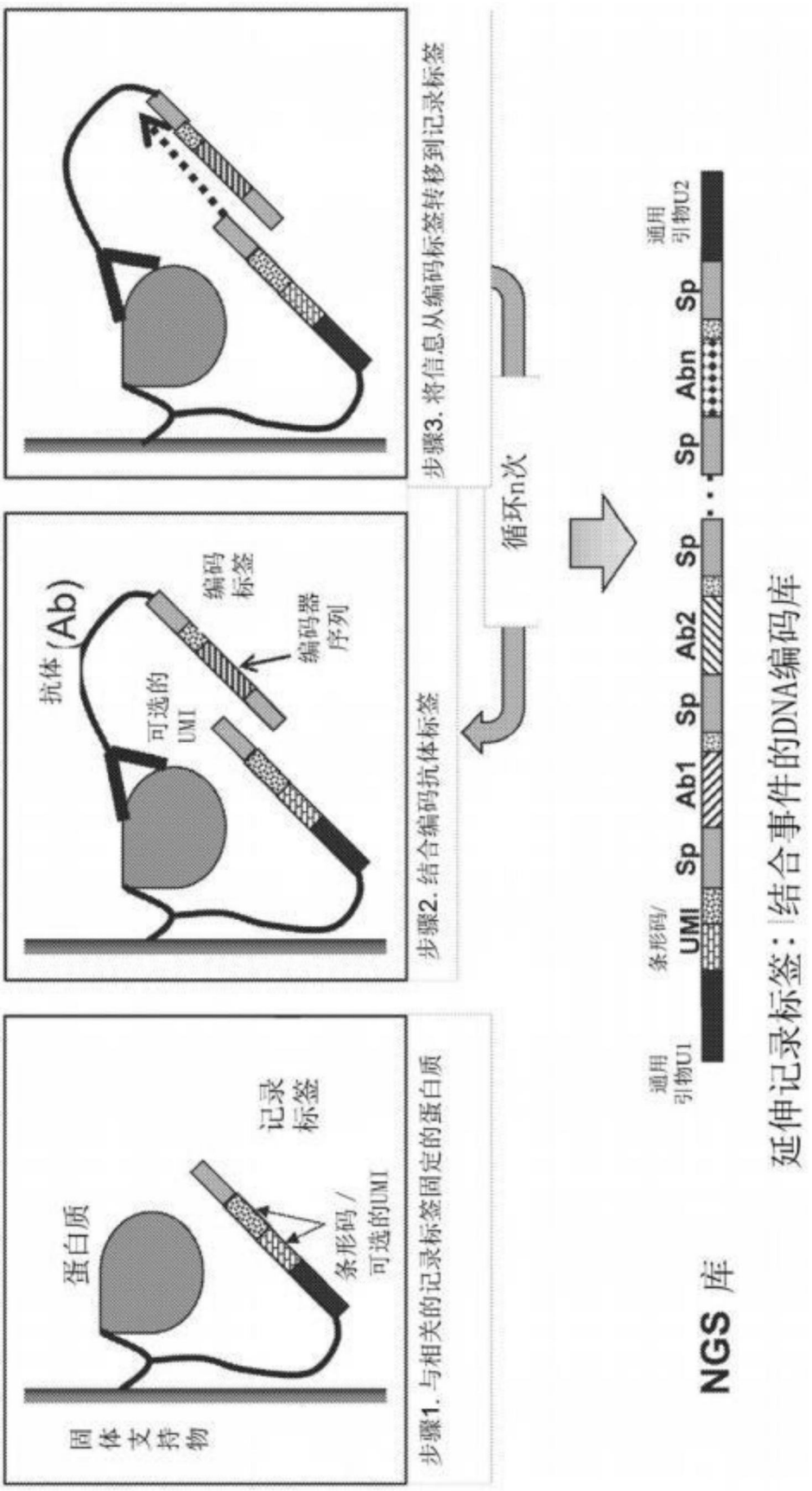

固体支持物
蛋白质
记录标签
条形码 / 可选的UMI
步骤1. 与相关的记录标签固定的蛋白质
抗体 (Ab)
可选的 UMI
编码标签
编码器序列
步骤2. 结合编码抗体标签
步骤3. 将信息从编码标签转移到记录标签
循环n次
通用引物U1
条形码/ UMI
Sp
Ab1
Sp
Ab2
Sp
Sp
Abn
Sp
通用引物U2
NGS 库
延伸记录标签：结合事件的DNA编码库

图2A

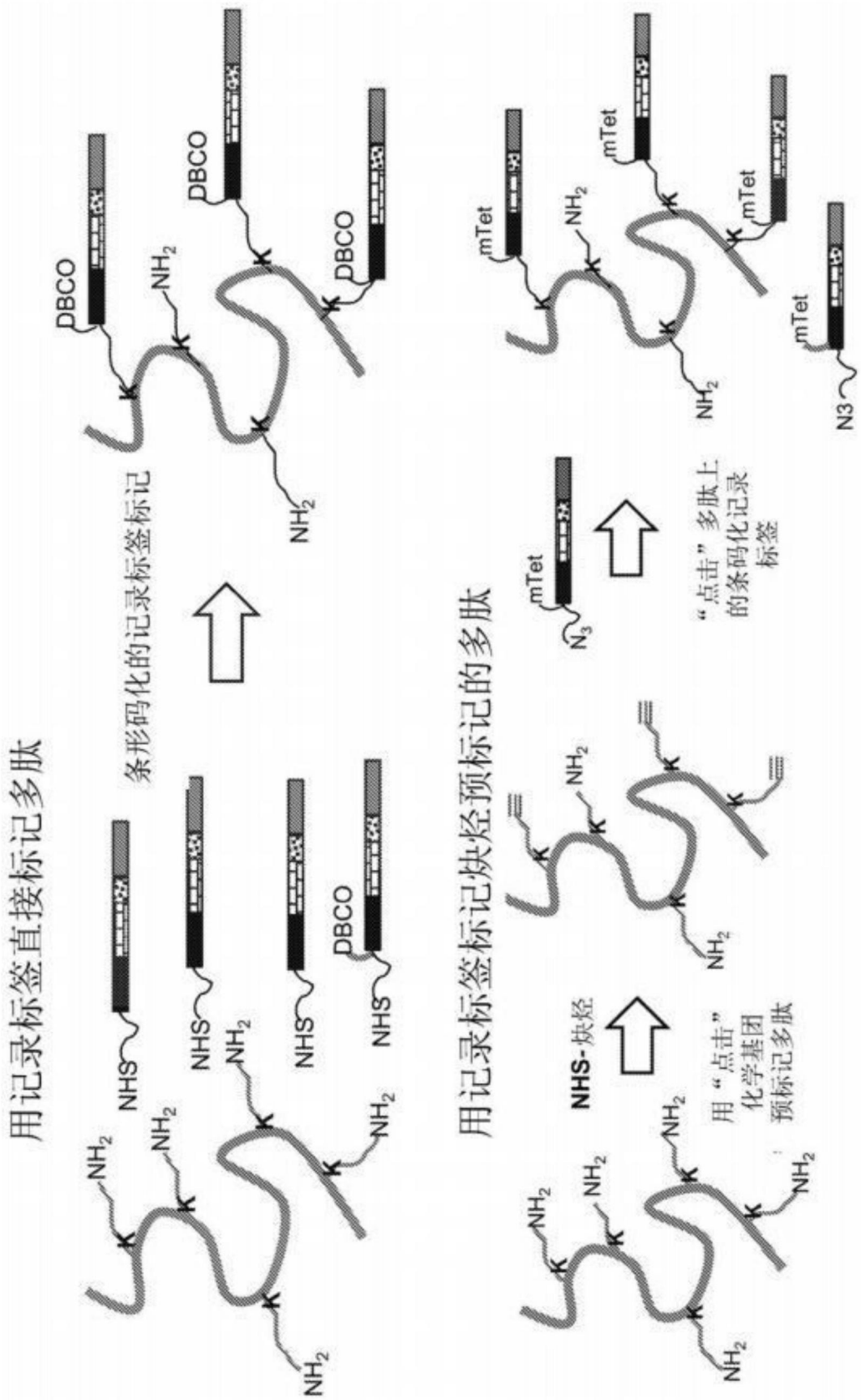

用记录标签直接标记多肽
NHS
DBCO
条形码化的记录标签标记
用记录标签标记炔烃预标记的多肽
NHS-炔烃
用“点击”
化学基团
预标记多肽
mTet
N3
“点击”多肽上
的条码化记录
标签

图2B

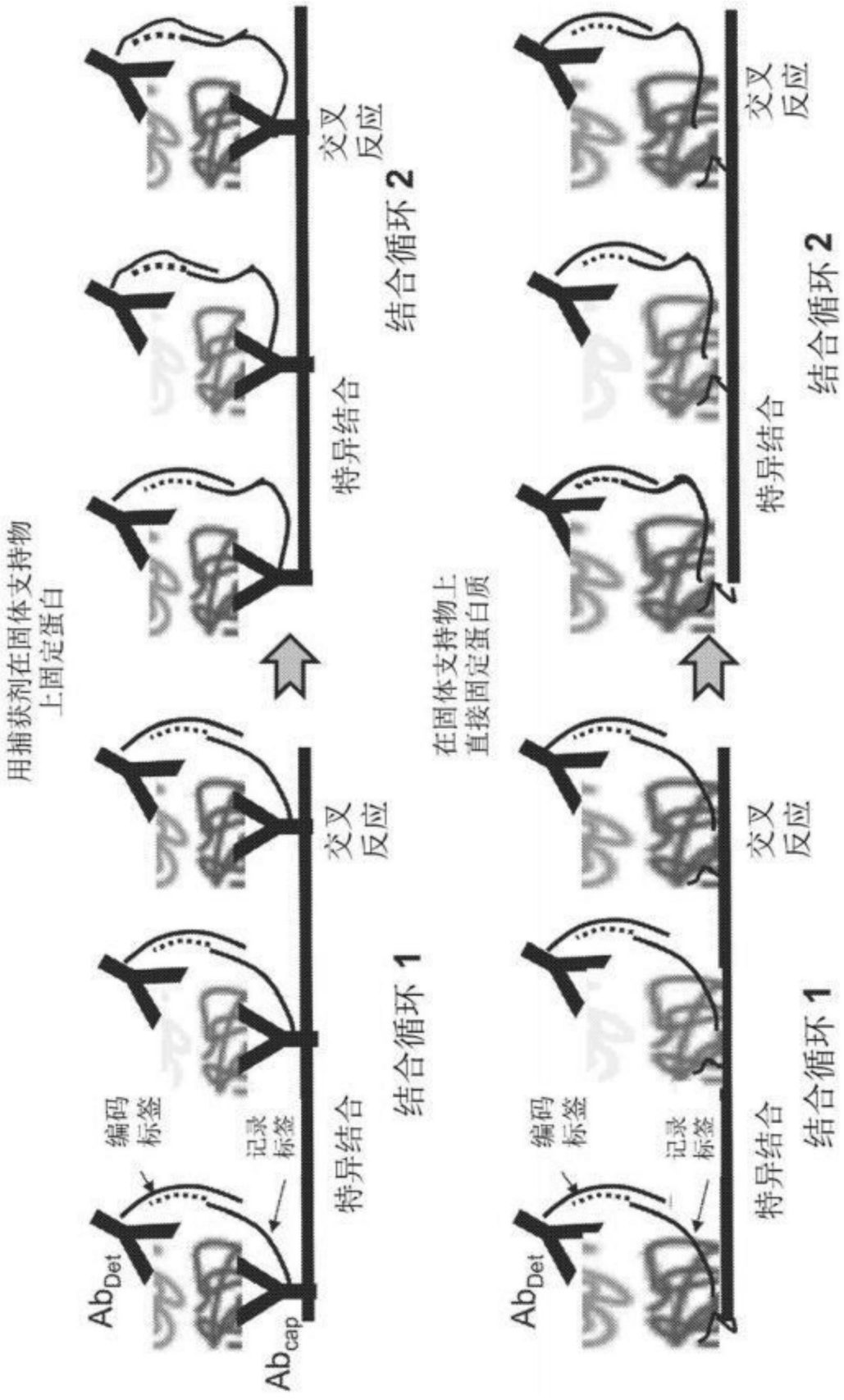

用捕获剂在固体支持物上固定蛋白
在固体支持物上直接固定蛋白质
Abdet
Abcap
编码标签
记录标签
特异结合
交叉反应
结合循环 1
结合循环 2

图2C

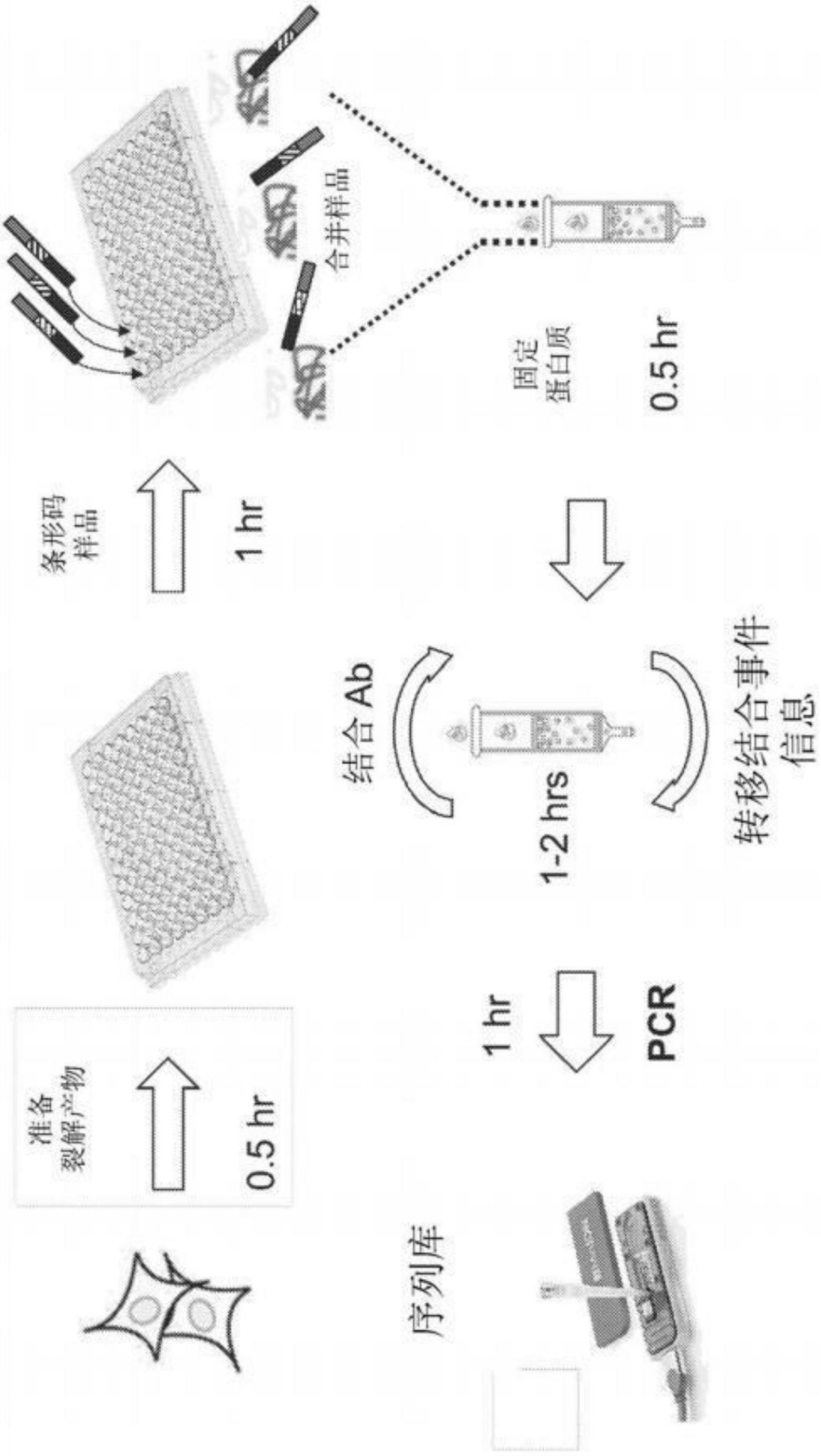

条形码
样品
1 hr
合并样品
固定
蛋白质
0.5 hr
结合 Ab
1-2 hrs
转移结合事件
信息
准备
裂解产物
0.5 hr
1 hr
PCR
序列库

图2D

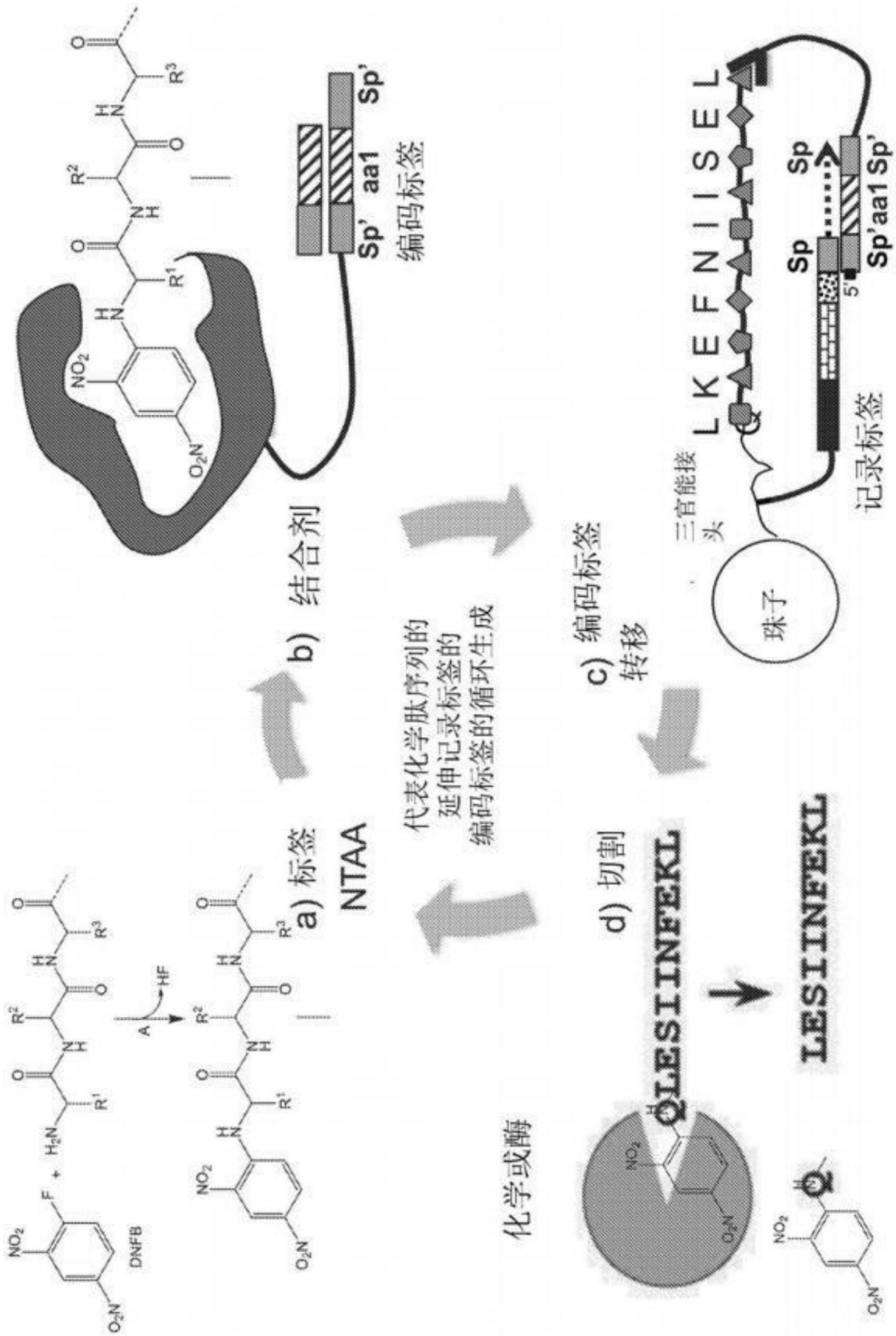

a) 标签
NTAA
b) 结合剂
Sp' aa1 Sp'
编码标签
代表化学肽序列的
延伸记录标签的循环生成
化学或酶
c) 编码标签
转移
三官能接头
珠子
Sp
Sp
Sp'aa1 Sp'
记录标签
d) 切割
QLESIINFEKL
LESIINFEKL
DNFB
HF

图3

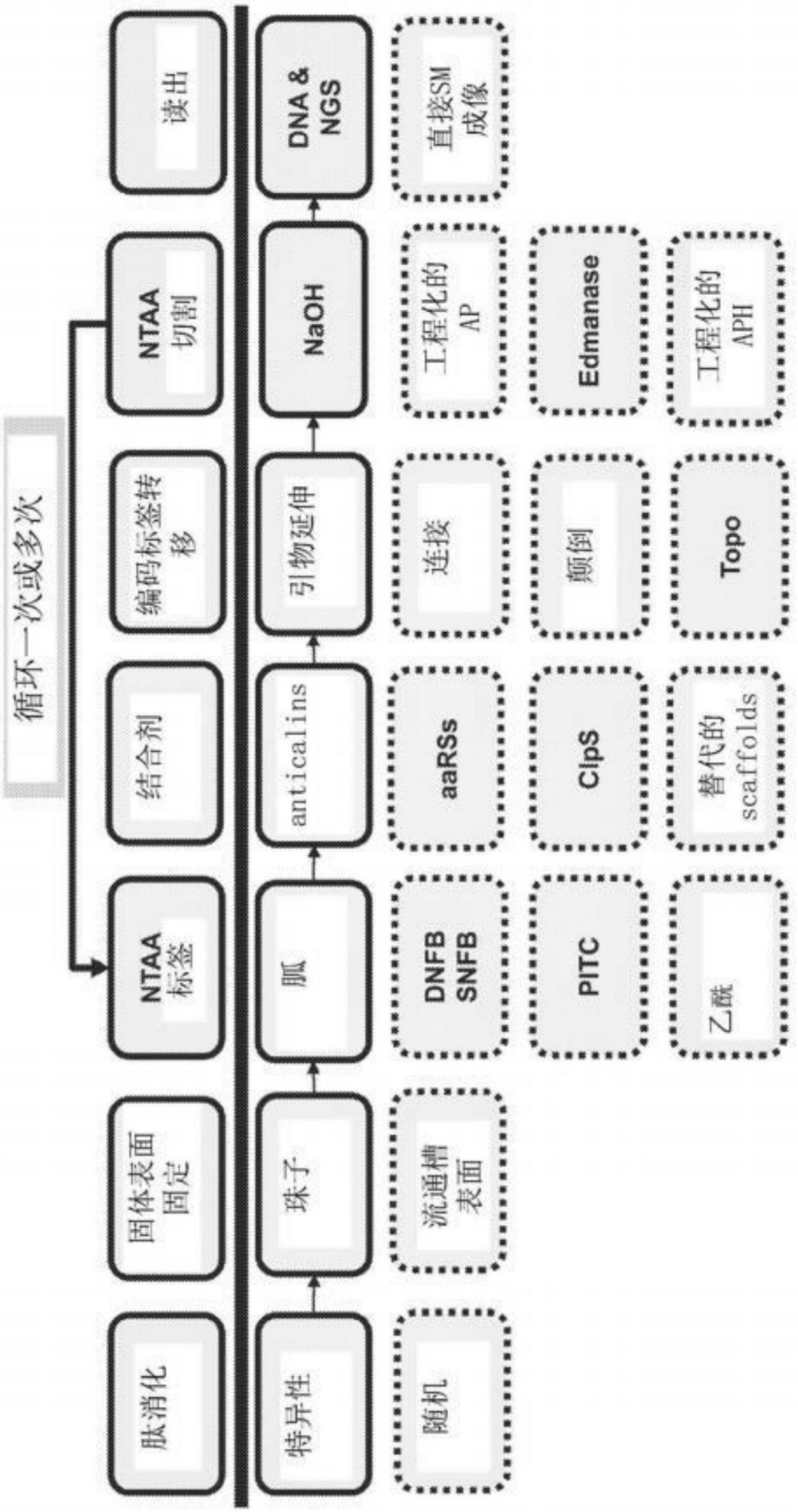

循环一次或多次
肽消化
固体表面固定
NTAA标签
结合剂
编码标签转移
NTAA切割
读出
特异性
珠子
胍
anticalins
引物延伸
NaOH
DNA & NGS
随机
流通槽表面
DNFB SNFB
aaRSs
连接
工程化的AP
直接SM成像
PITC
ClpS
颠倒
Edmanase
乙酰
替代的scaffolds
Topo
工程化的APH

图4A

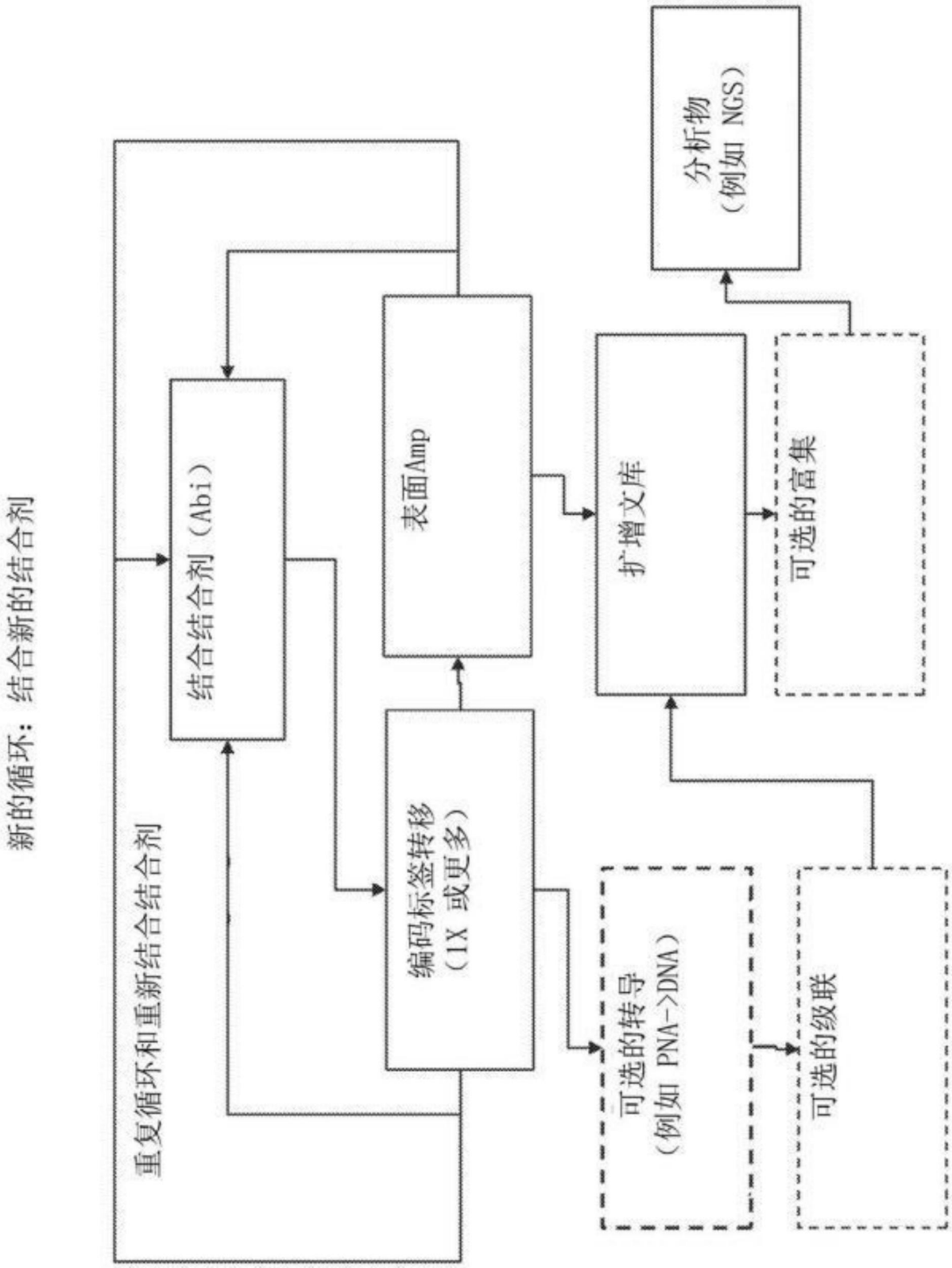

新的循环：结合新的结合剂
重复循环和重新结合结合剂
结合结合剂（Abi）
编码标签转移
（1X 或更多）
表面Amp
扩增文库
分析物
（例如 NGS）
可选的富集
可选的转导
（例如 PNA->DNA）
可选的级联

图4B

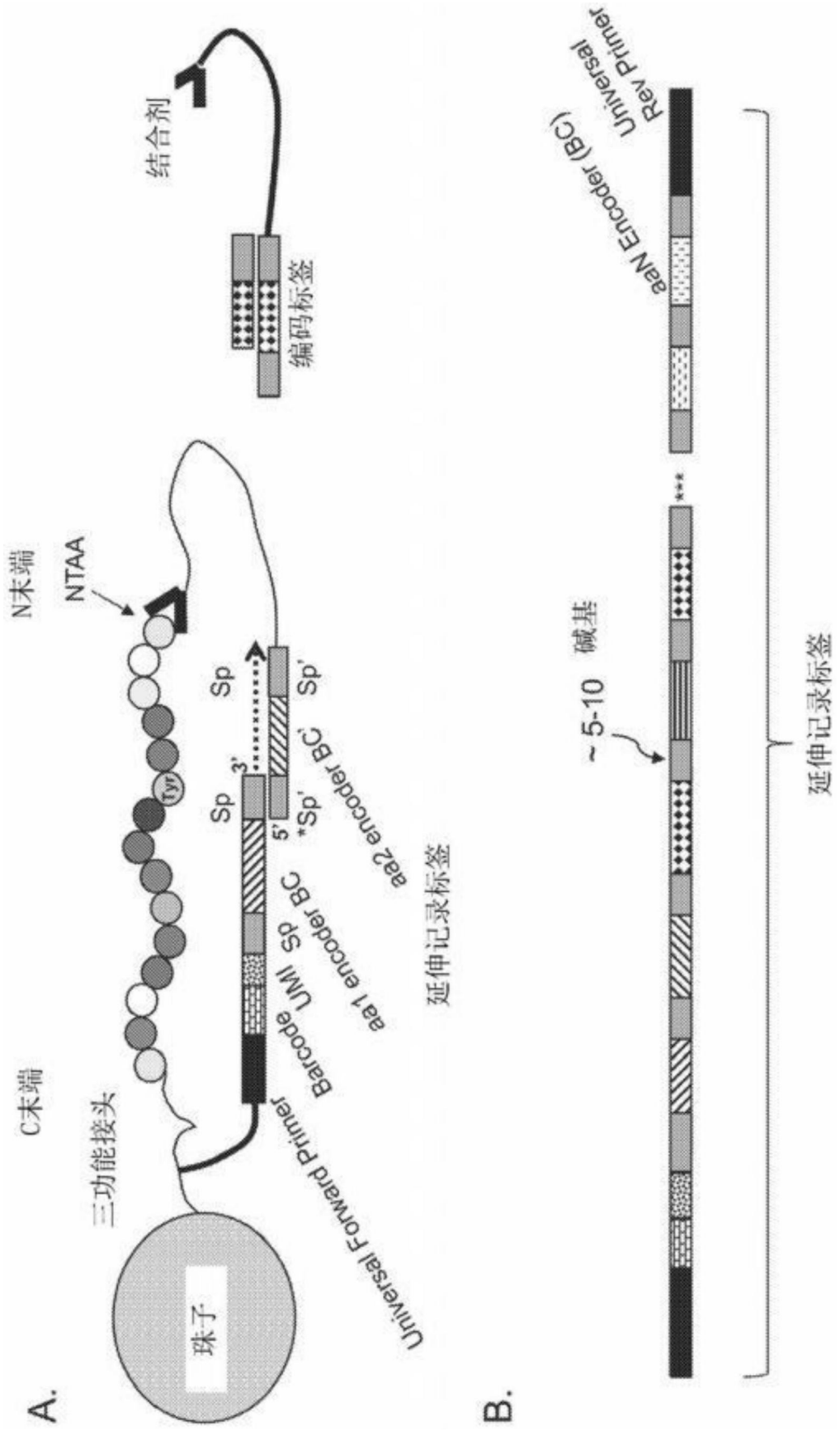

A.
C末端
三功能接头
珠子
Universal Forward Primer
Barcode UMI Sp
aa1 encoder BC
Sp
3'
5'
*Sp'
aa2 encoder BC
Sp'
Sp
tyr
N末端
NTAA
结合剂
编码标签
延伸记录标签
B.
~ 5-10 碱基
aaN Encoder (BC)
Rev Primer Universal
延伸记录标签

图5

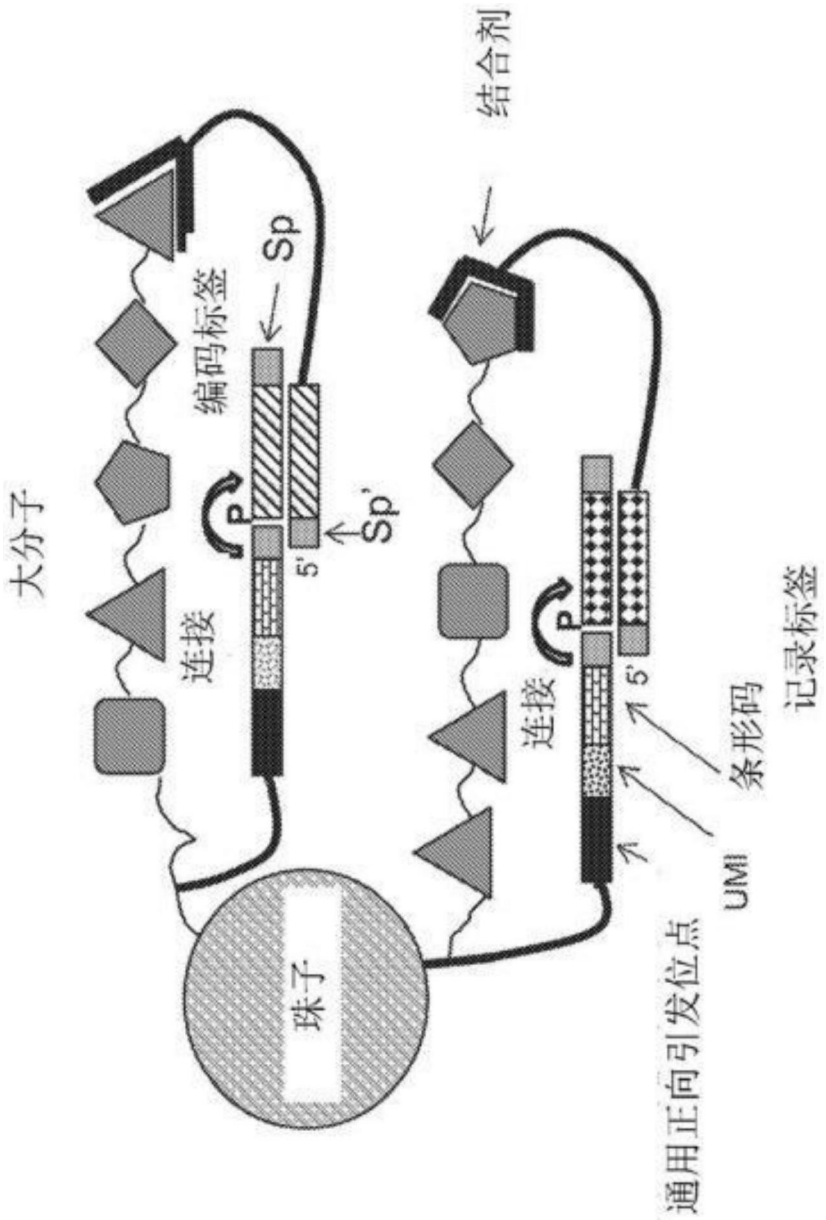


图6

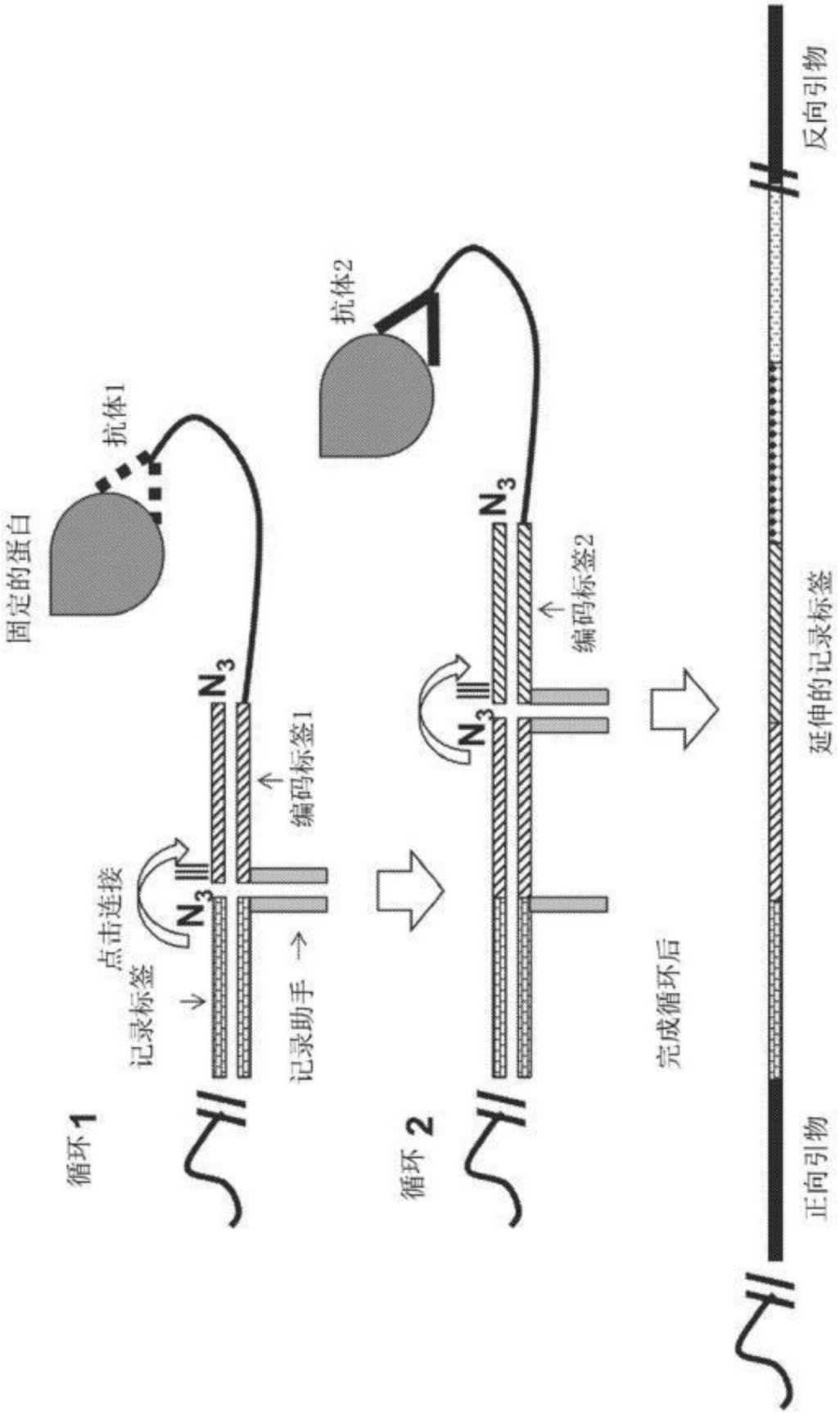

固定的蛋白
抗体1
抗体2
点击连接
记录标签
记录助手
编码标签1
编码标签2
循环 1
循环 2
完成循环后
正向引物
反向引物
延伸的记录标签

图7

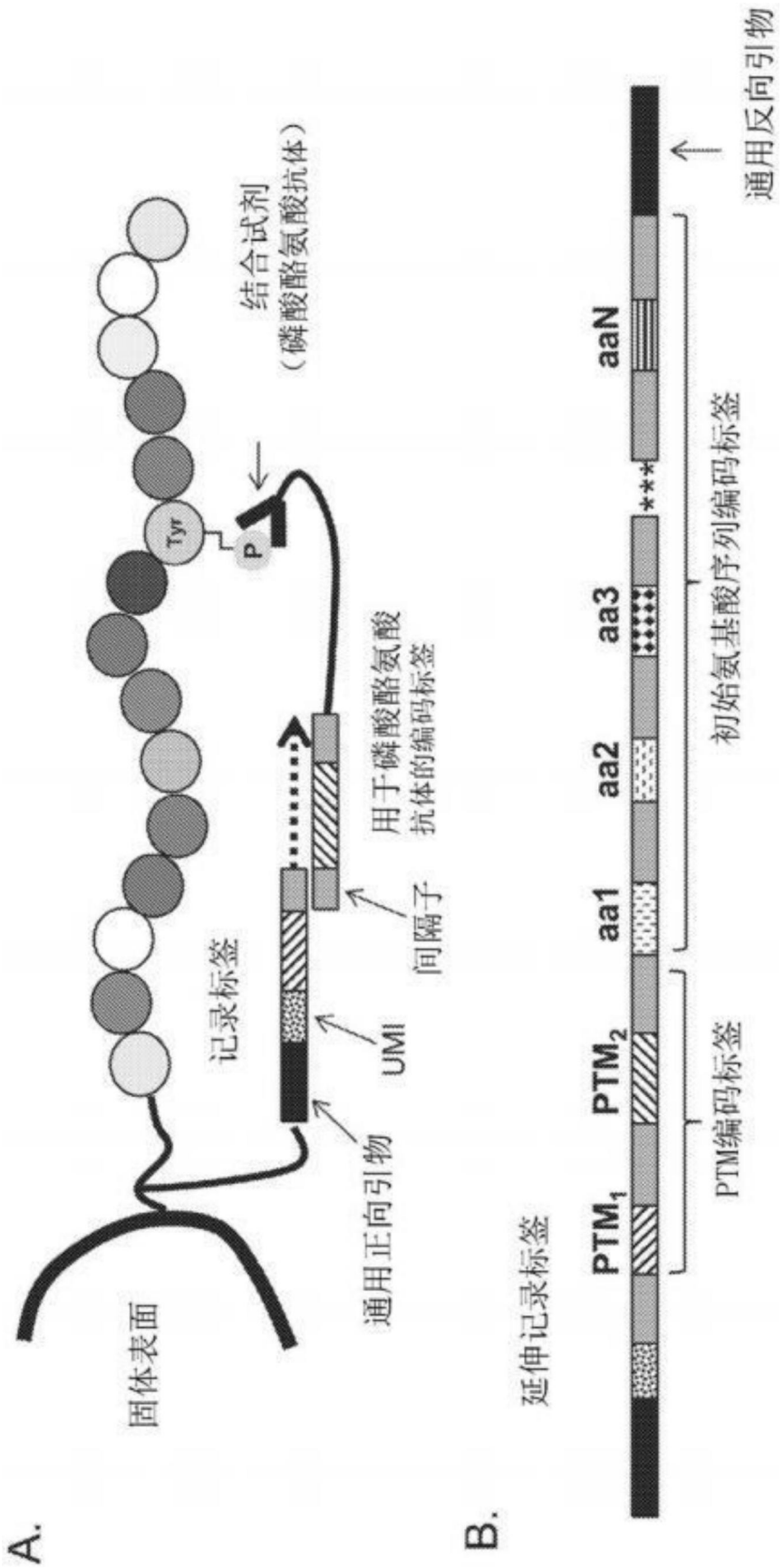

A.
固体表面
记录标签
Tyr
P
结合试剂
（磷酸酪氨酸抗体）
通用正向引物
UMI
间隔子
用于磷酸酪氨酸
抗体的编码标签
B.
延伸记录标签
PTM1
PTM2
aa1
aa2
aa3
***
aaN
PTM编码标签
初始氨基酸序列编码标签
通用反向引物

图8

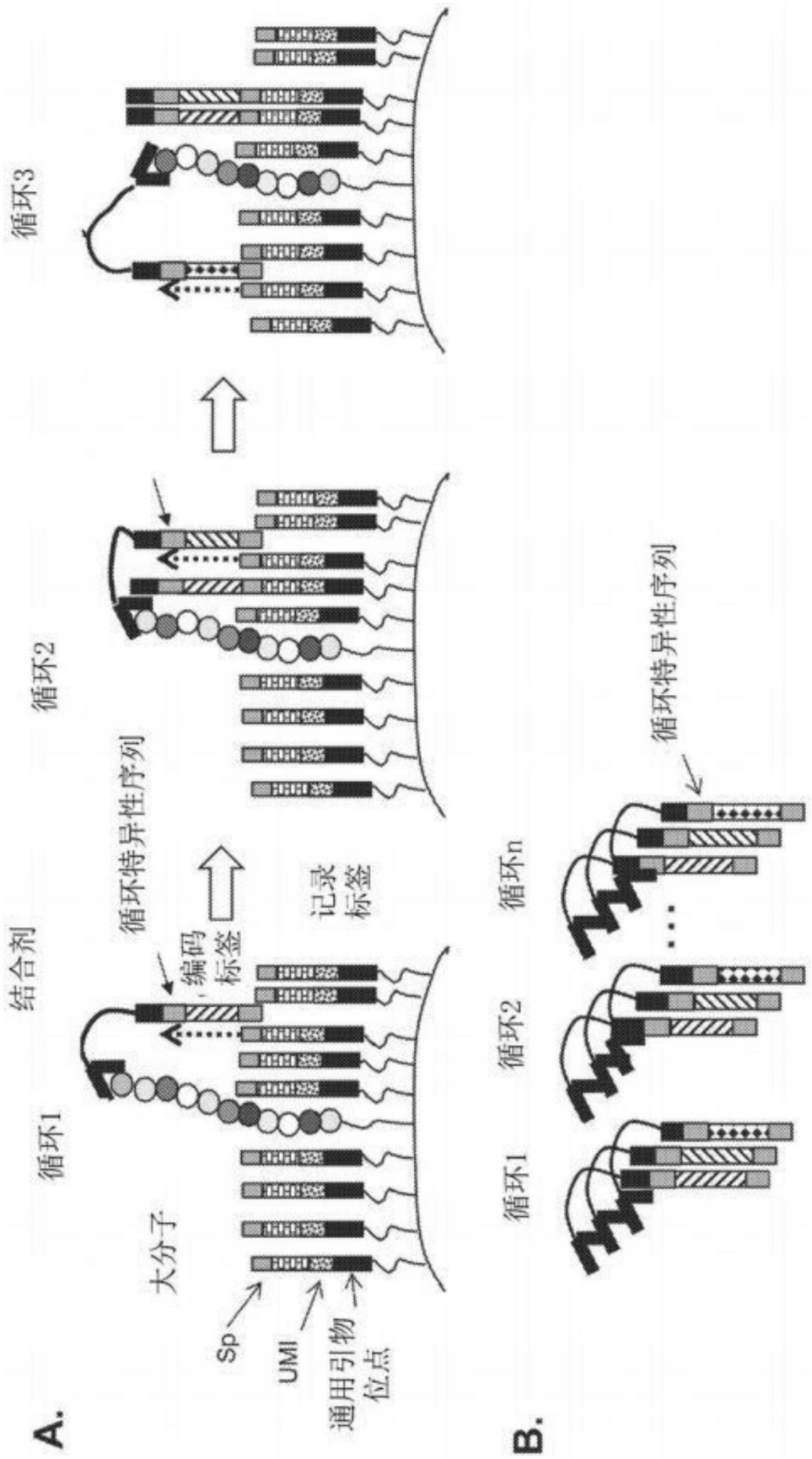

A.
循环1
结合剂
大分子
Sp
UMI
通用引物位点
编码标签
循环特异性序列
记录标签
循环2
循环3
B.
循环1
循环2
循环n
循环特异性序列

图9

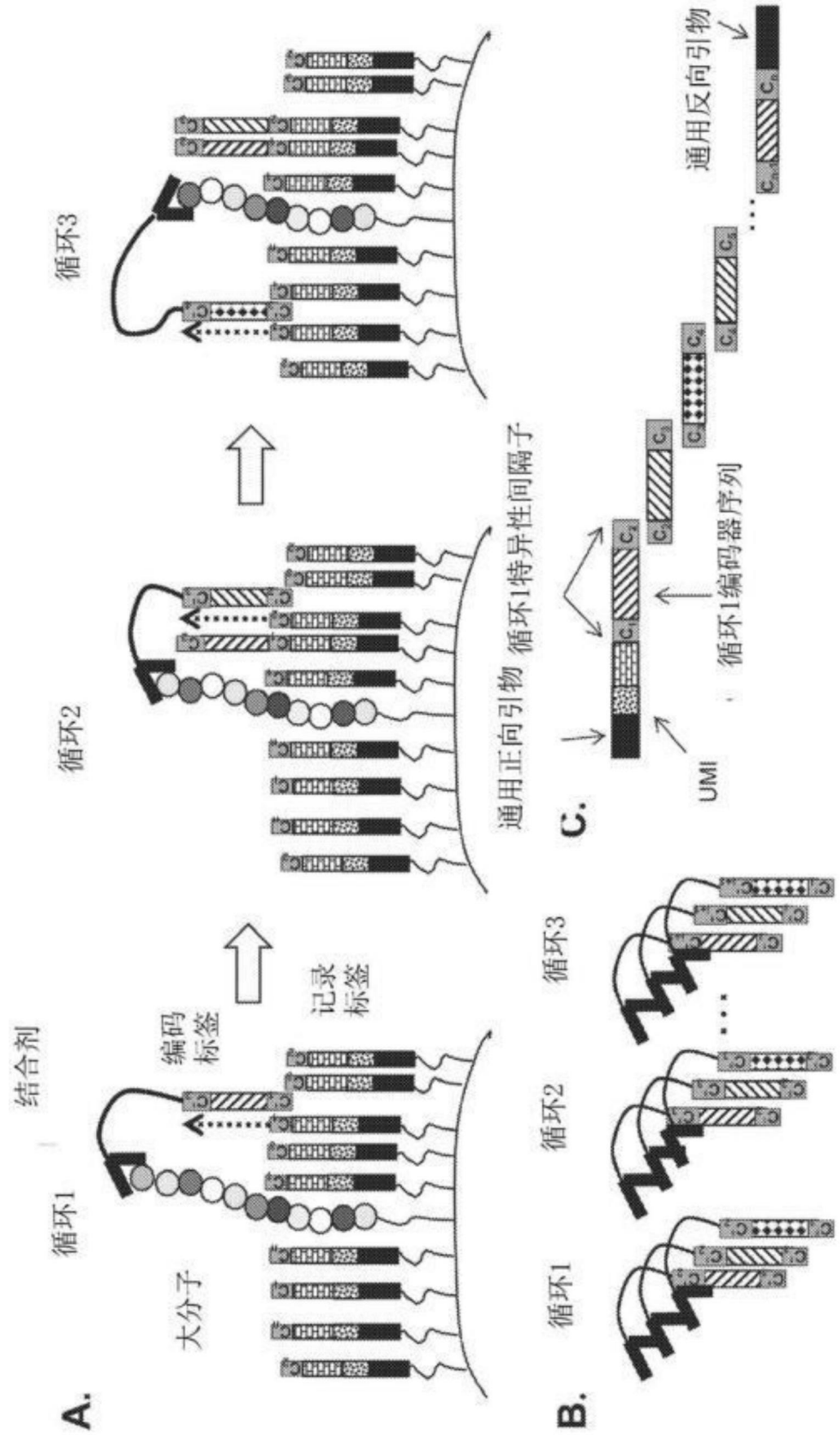

A.
大分子
结合剂
循环1
编码
标签
记录
标签
循环2
循环3
通用正向引物
循环1特异性间隔子
C.
UMI
循环1编码器序列
通用反向引物
B.
循环1
循环2
循环3

图10

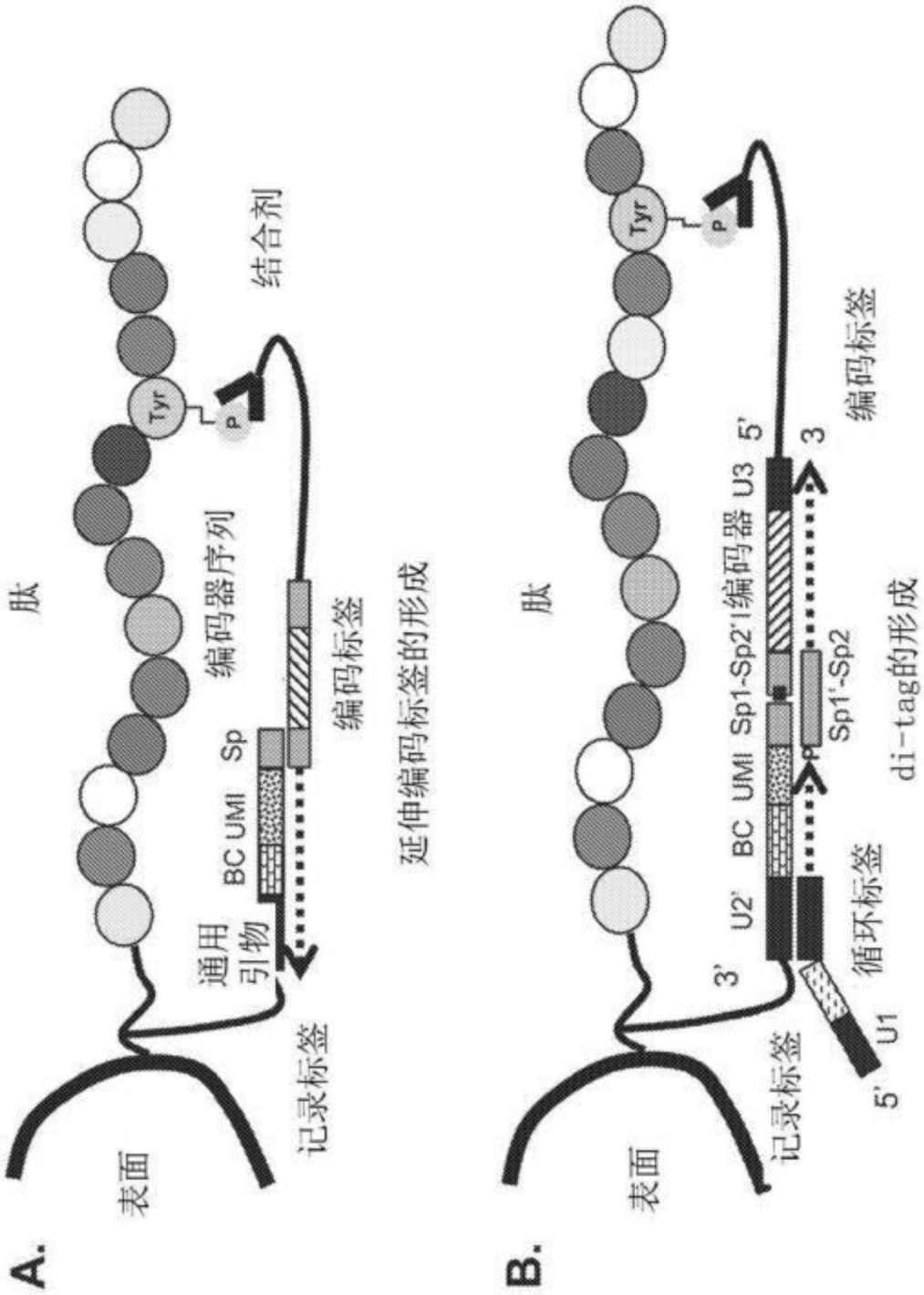

A.
表面
肽
Tyr
P
结合剂
通用引物
BC UMI
Sp
编码器序列
记录标签
编码标签
延伸编码标签的形成
B.
表面
肽
Tyr
P
3'
U2'
BC
UMI
Sp1-Sp2'
编码器
U3
5'
3
编码标签
Sp1'-Sp2
循环标签
U1
5'
di-tag的形成
记录标签

图11

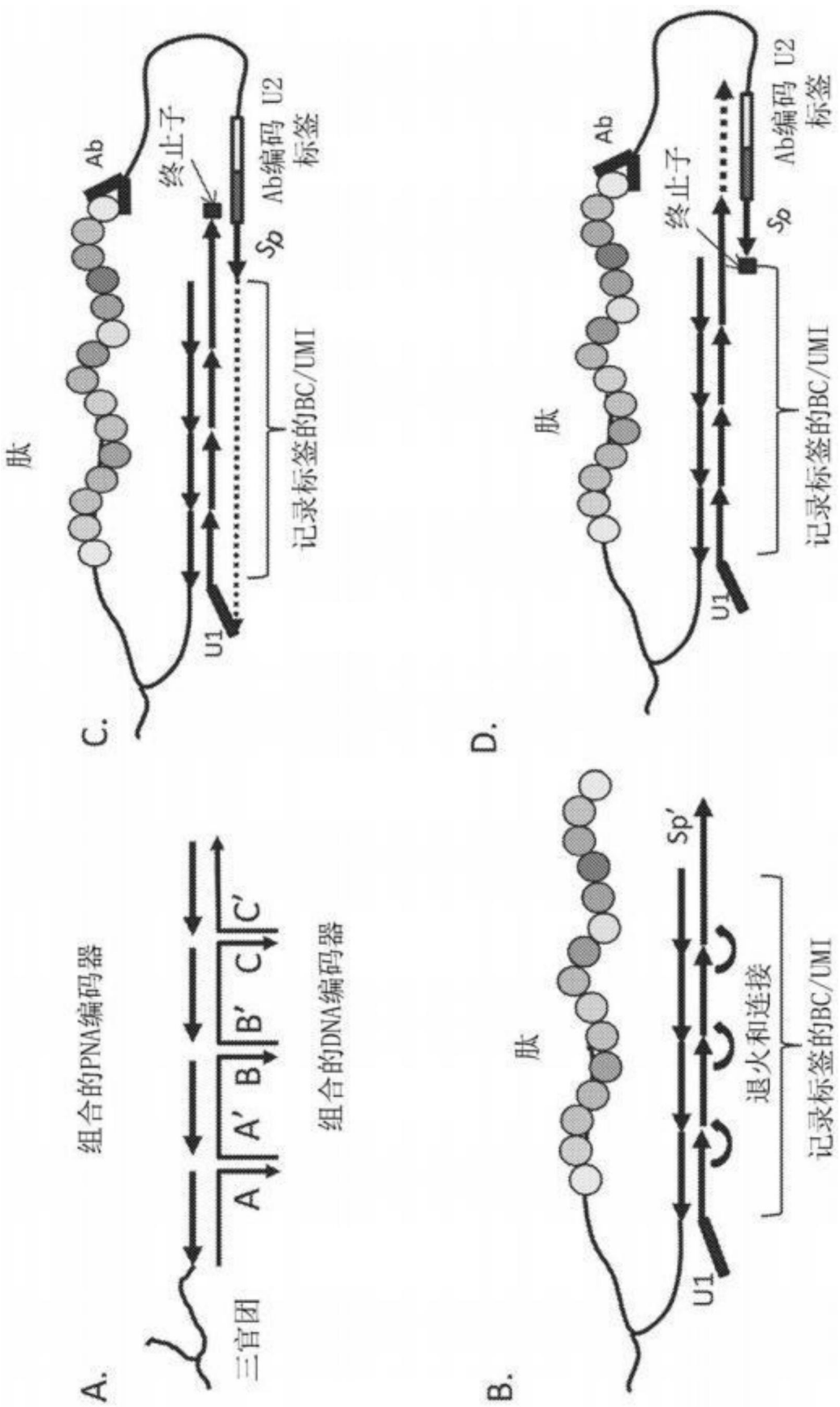

A.
组合的PNA编码器
三官团
A
A′
B
B′
C
C′
组合的DNA编码器
B.
肽
Sp′
退火和连接
U1
记录标签的BC/UMI
C.
肽
Ab
终止子
Sp
Ab编码
U2
标签
U1
记录标签的BC/UMI
D.
肽
Ab
终止子
Sp
Ab编码
U2
标签
U1
记录标签的BC/UMI

图12

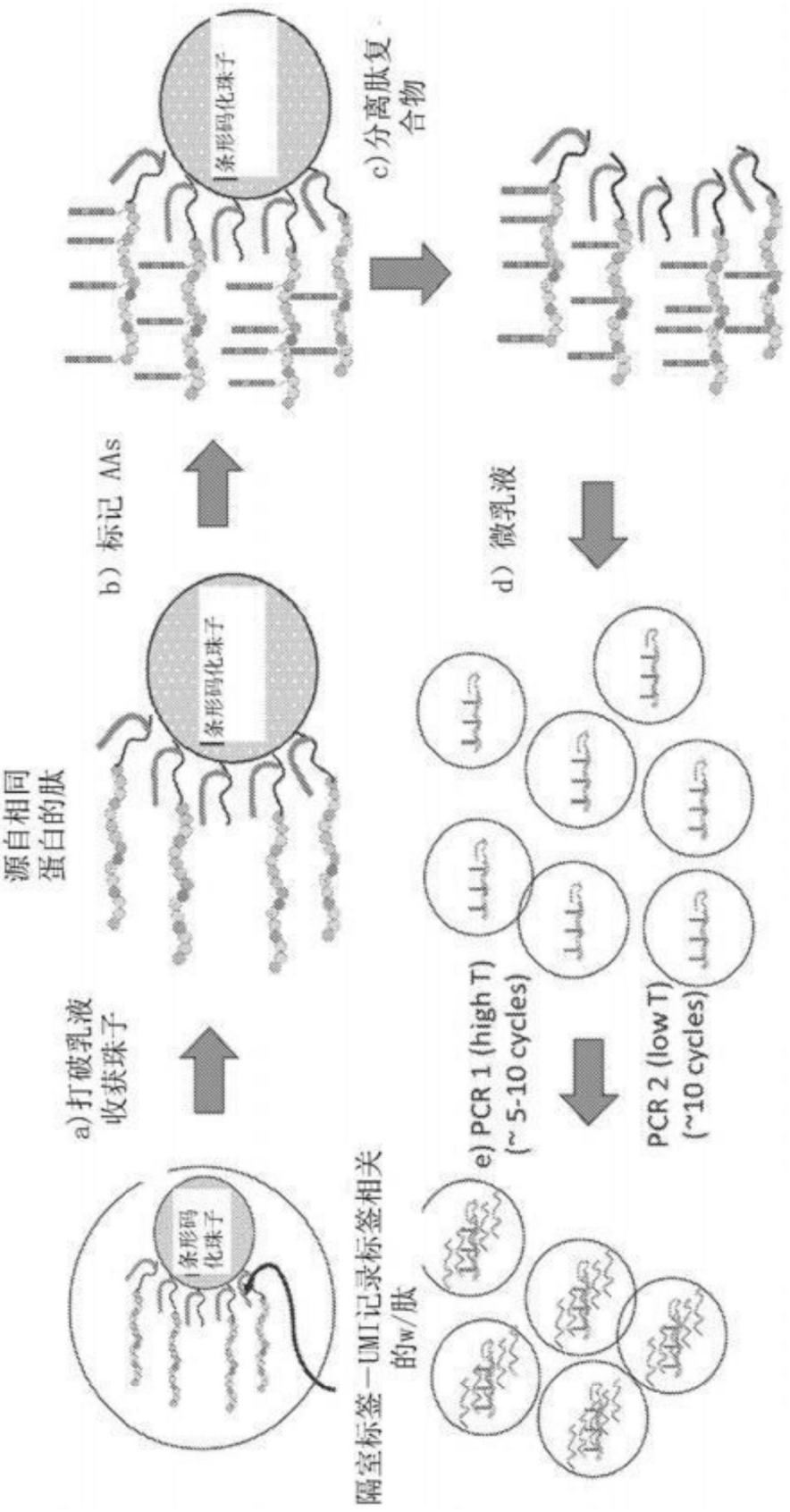

条形码化珠子
a)打破乳液
收获珠子
源自相同
蛋白的肽
b) 标记 AMs
c)分离肽复
合物
d) 微乳液
隔室标签—UMI记录标签相关
的w/肽
e) PCR 1 (high T)
(~ 5-10 cycles)
PCR 2 (low T)
(~10 cycles)

图13

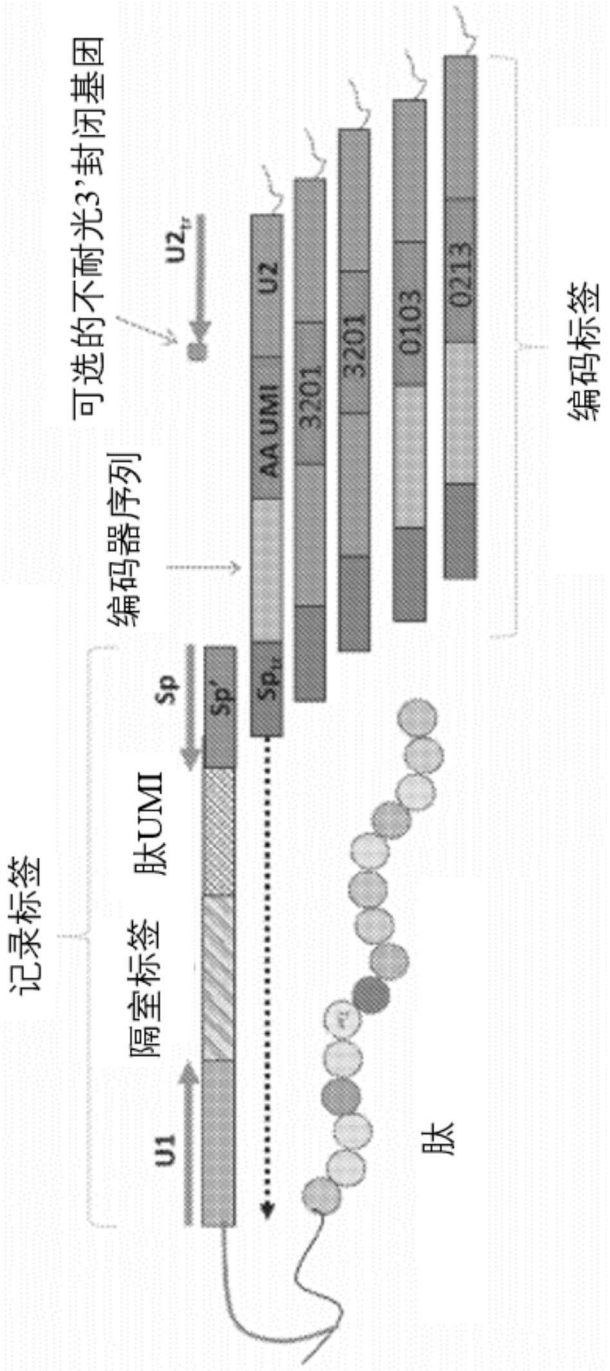

可选的不耐光3'封闭基团
$U2_{tr}$
编码器序列
AA UMI
U2
3201
3201
0103
0213
编码标签
Sp
Sp'
Sp_{tr}
记录标签
隔室标签
肽UMI
U1
肽

图14

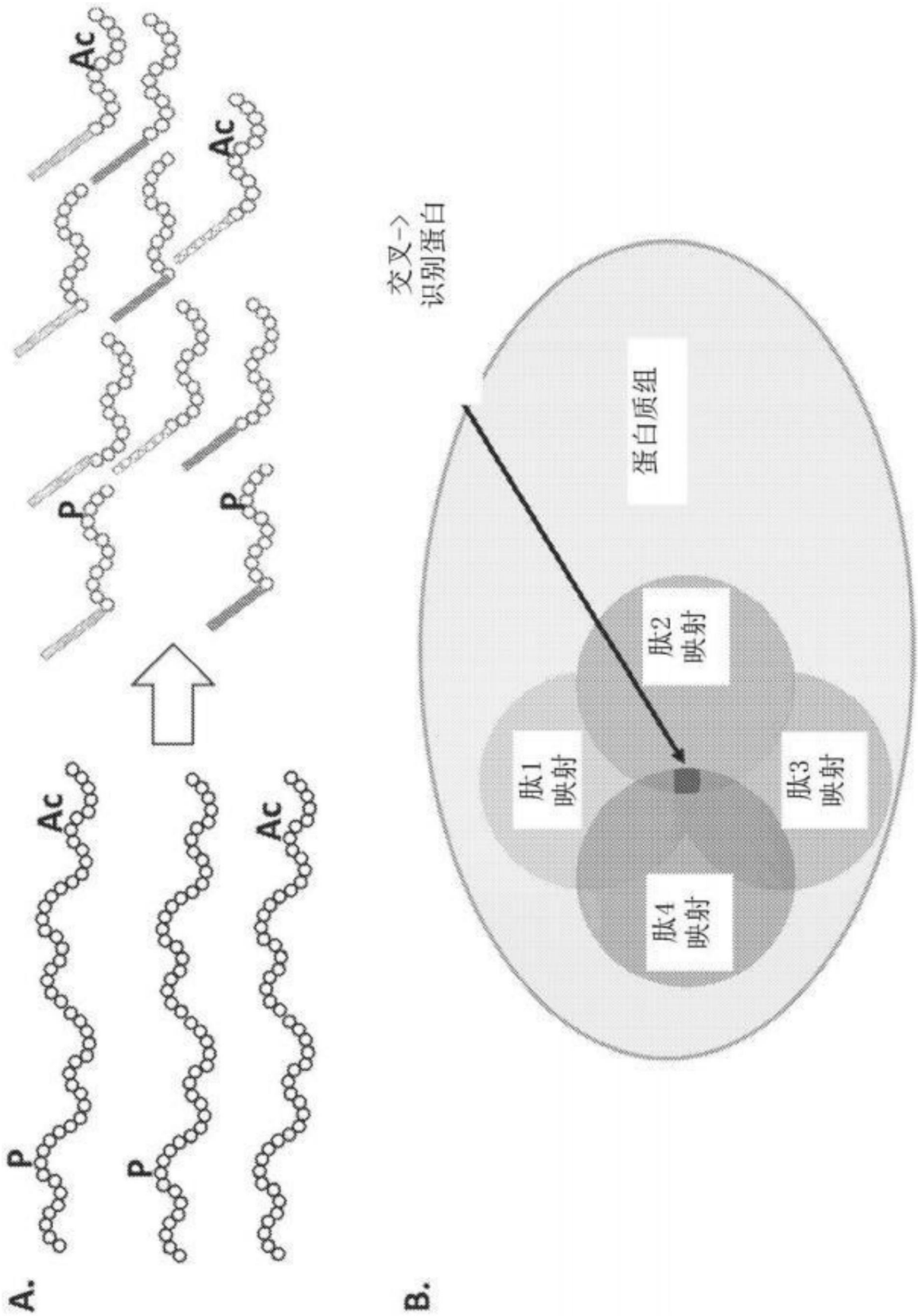


图15

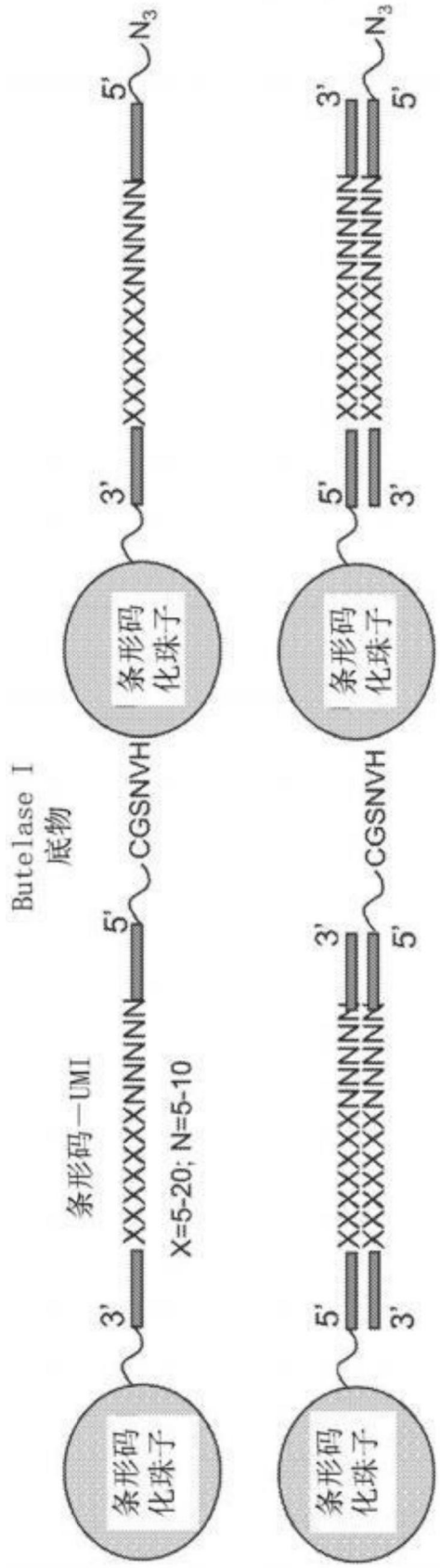

条形码化珠子
条形码化珠子
条形码化珠子
条形码化珠子
5'
3'
N_3
XXXXXXXXNNNNNN
CGSNVH
Butelase I
底物
条形码—UMI
XXXXXXXXNNNNNN
X=5-20; N=5-10

图16

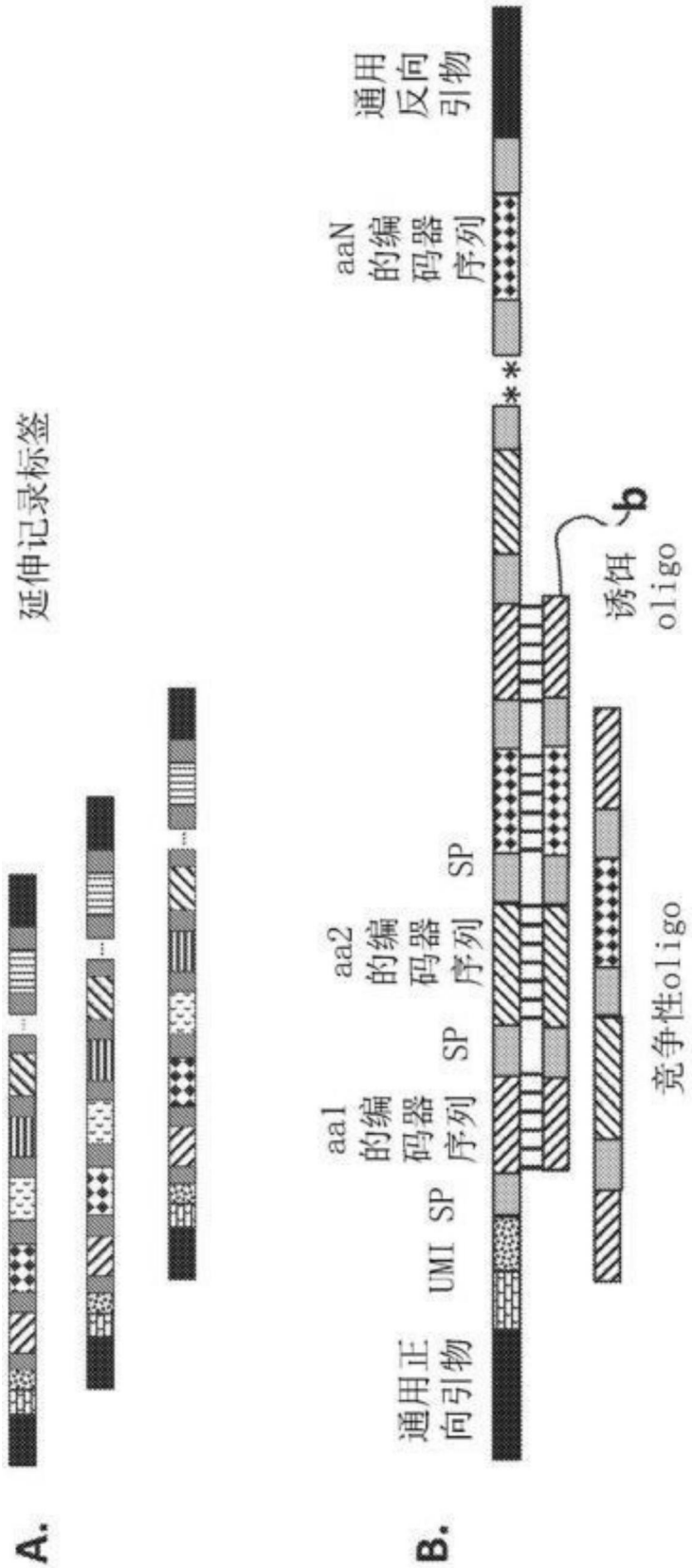

A.
延伸记录标签
B.
通用正向引物
UMI
SP
aa1的编码器序列
SP
aa2的编码器序列
SP
aaN的编码器序列
通用反向引物
诱饵oligo
竞争性oligo

图17

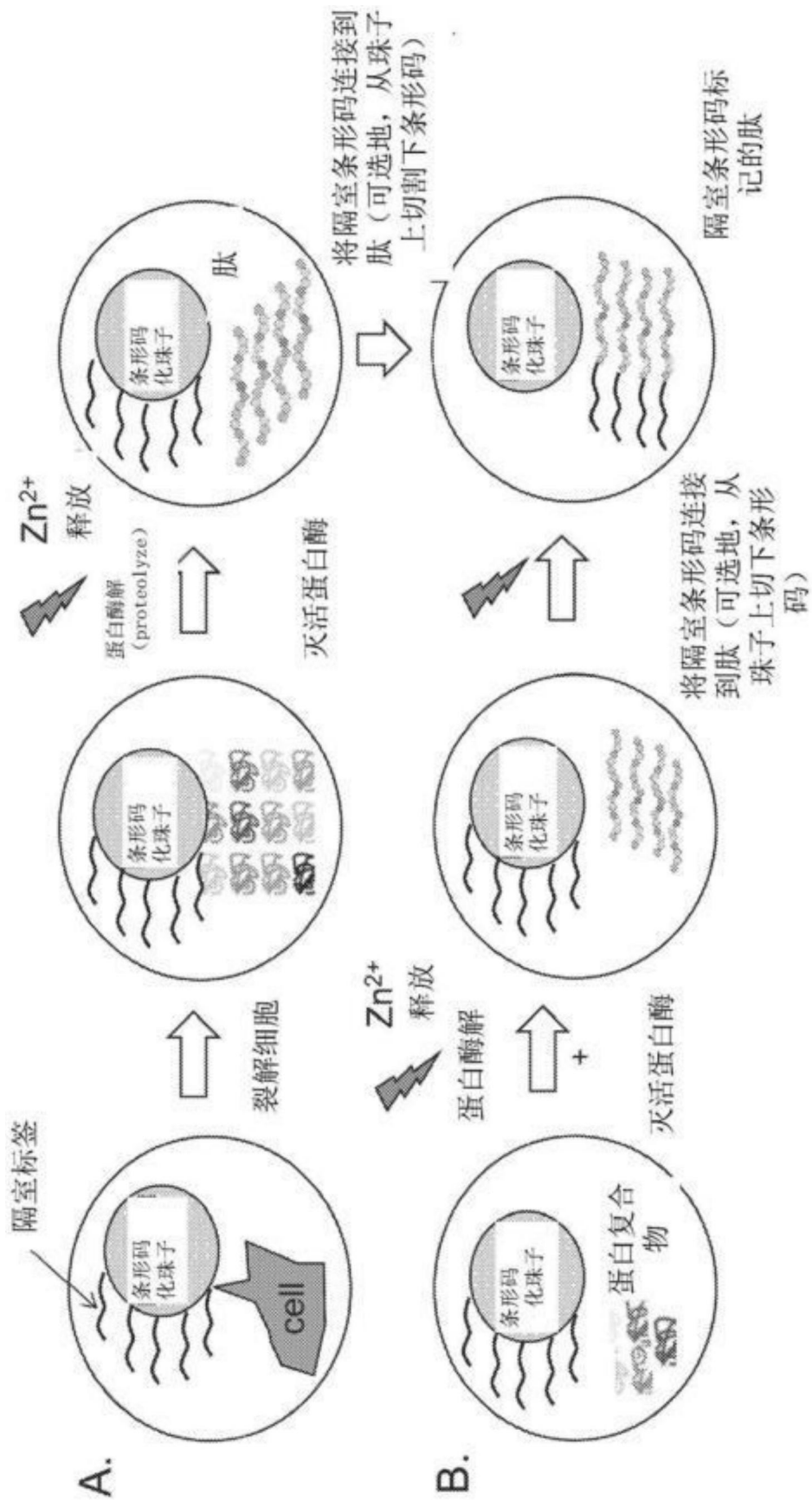

A.
隔室标签
条形码化珠子
cell
裂解细胞
条形码化珠子
蛋白酶解
(proteolyze)
Zn^{2+}
释放
条形码化珠子
肽
灭活蛋白酶
将隔室条形码连接到肽（可选地，从珠子上切割下条形码）
隔室条形码标记的肽
B.
条形码化珠子
蛋白复合物
灭活蛋白酶
+
条形码化珠子
蛋白酶解
Zn^{2+}
释放
条形码化珠子
将隔室条形码连接到肽（可选地，从珠子上切下条形码）

图18

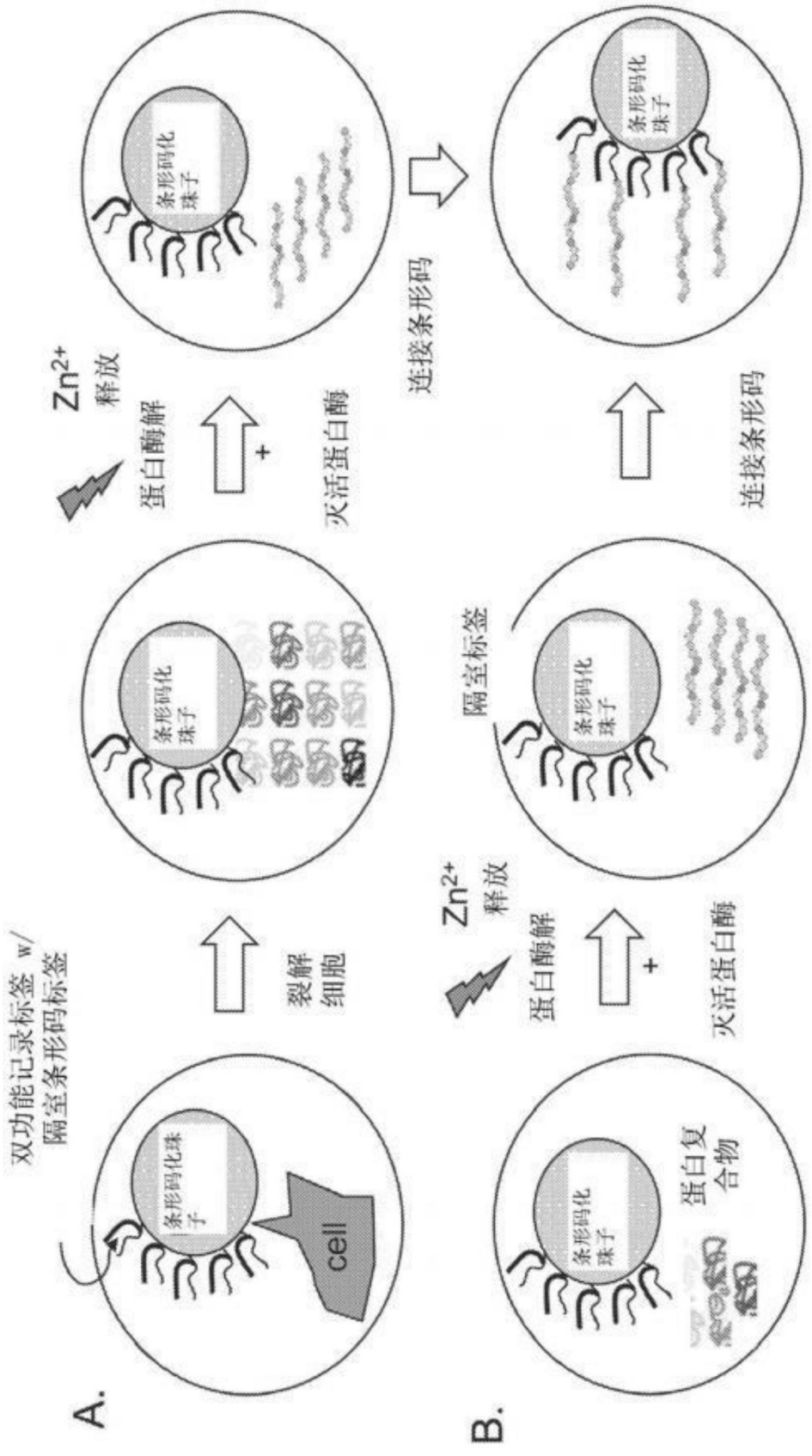

双功能记录标签 w/
隔室条形码标签
条形码化珠子
cell
裂解细胞
条形码化珠子
A.
Zn2+ 释放
蛋白酶解
+
灭活蛋白酶
条形码化珠子
连接条形码
隔室标签
条形码化珠子
Zn2+ 释放
蛋白酶解
+
灭活蛋白酶
条形码化珠子
连接条形码
条形码化珠子
蛋白复合物
B.

图19

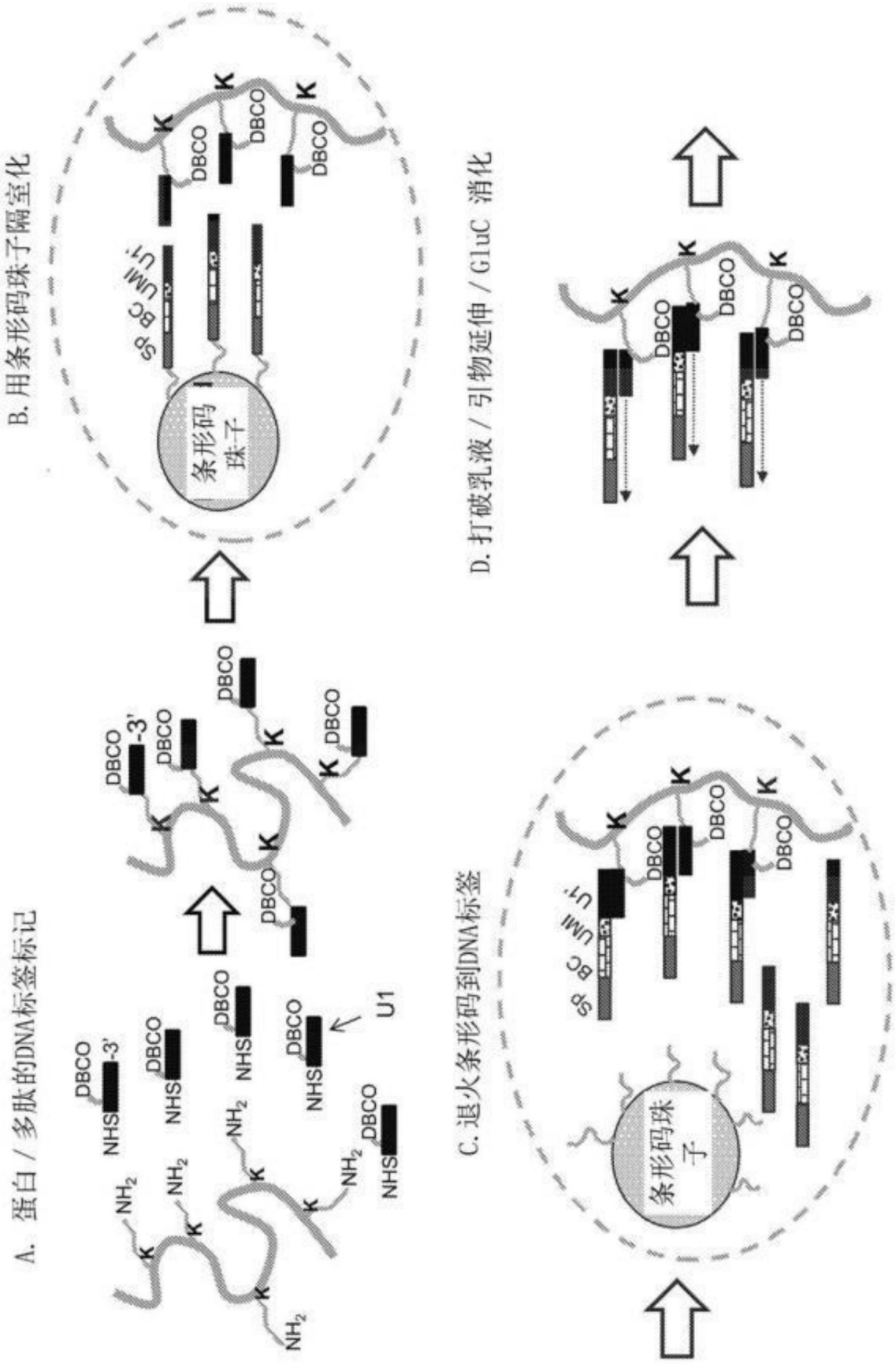

A. 蛋白 / 多肽的DNA标签标记
B. 用条形码珠子隔室化
C. 退火条形码到DNA标签
D. 打破乳液 / 引物延伸 / G1uC 消化
条形码珠子
U1
DBCO
NHS
NH2
K
-3'

图20A-D

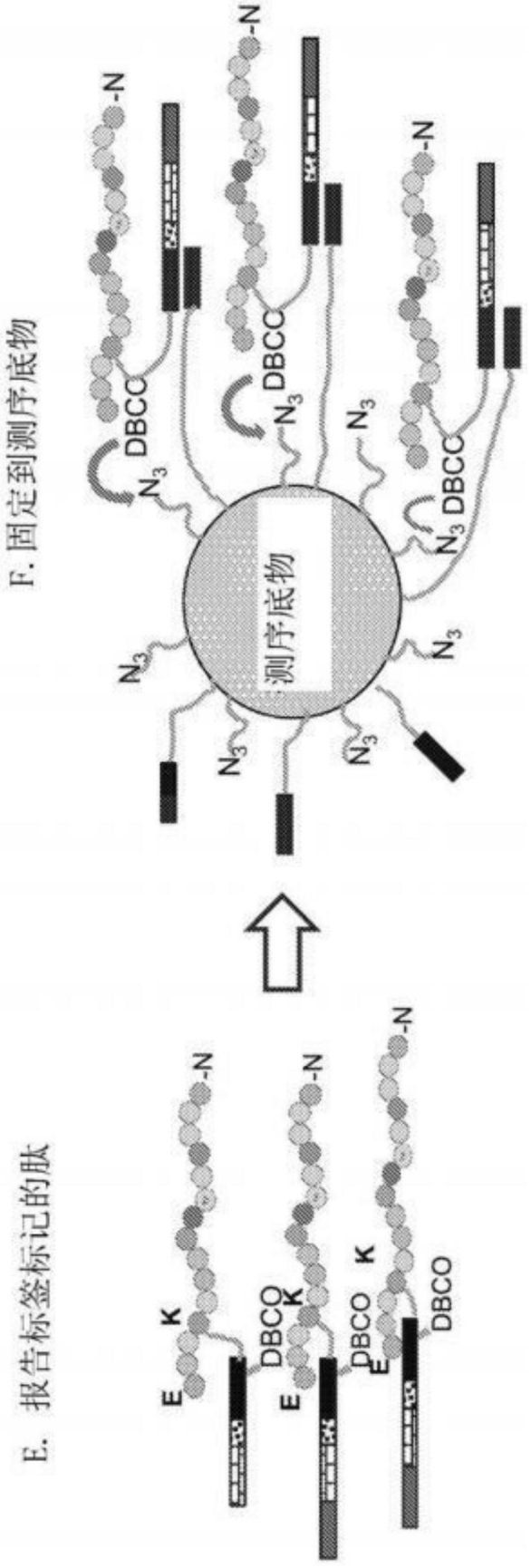
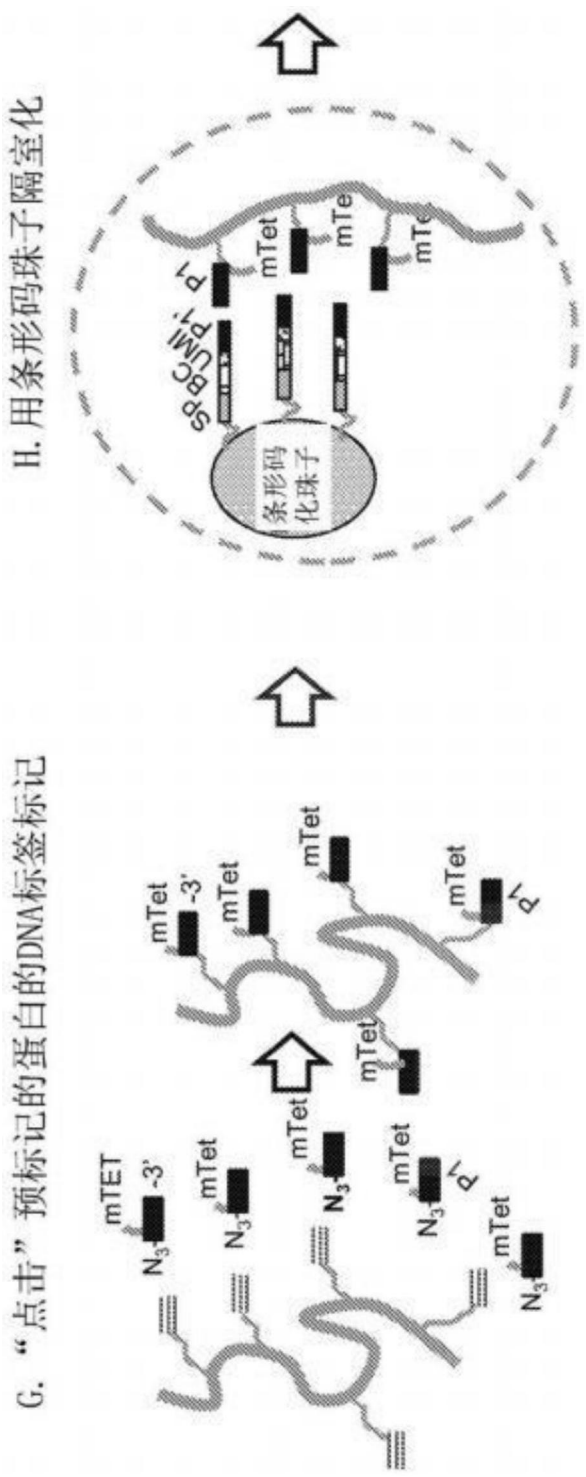


图20E-F


H. 用条形码珠子隔室化
条形码化珠子
mTet
mTet
mTet
G. “点击”预标记的蛋白的DNA标签标记
mTET
-3'
mTet
mTet
mTet

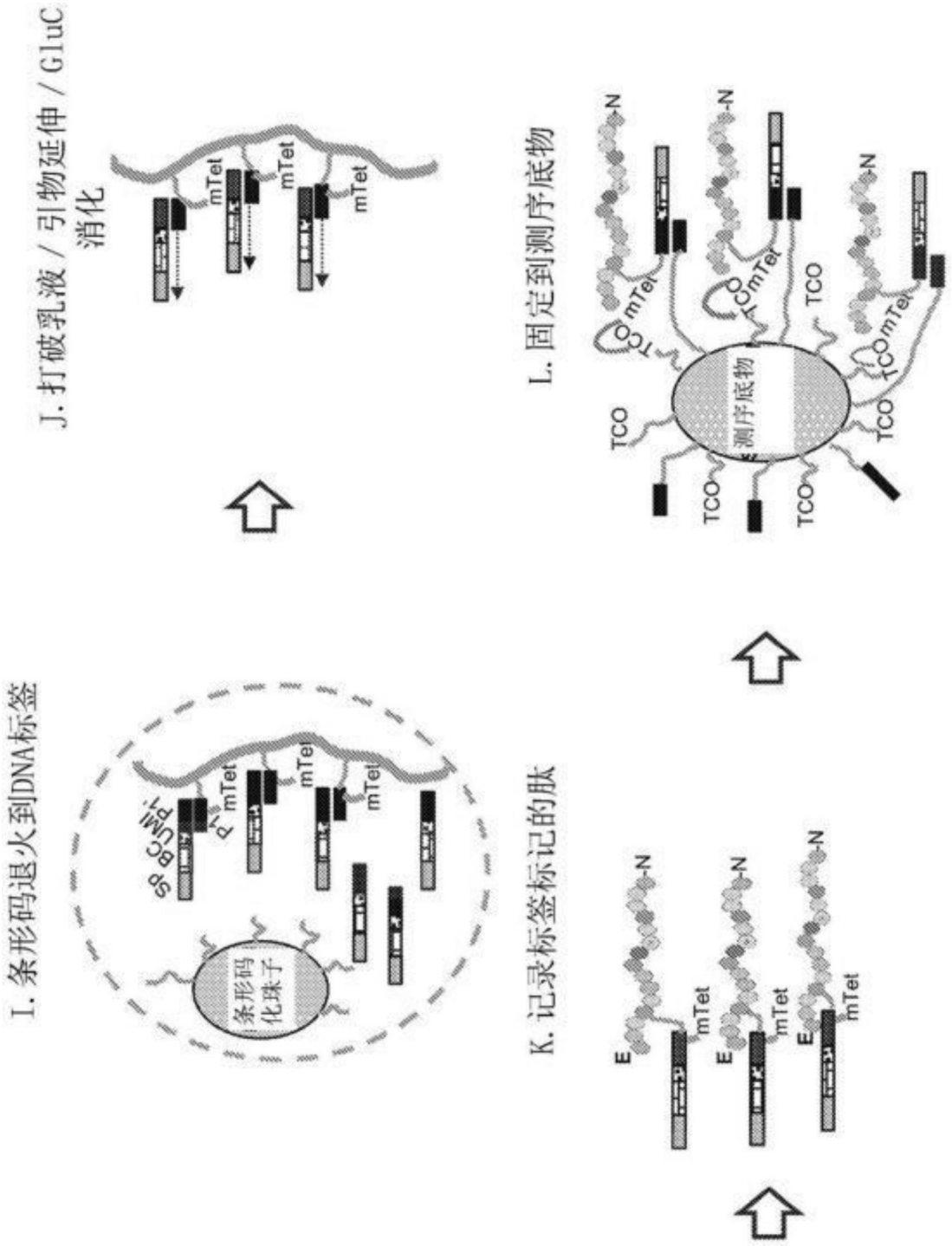

J. 打破乳液 / 引物延伸 / G1uC 消化
mTet
mTet
mTet
L. 固定到测序底物
N
mTet
TCO
测序底物
TCO
TCO
TCO
TCO
I. 条形码退火到DNA标签
mTet
mTet
mTet
条形码化珠子
K. 记录标签标记的肽
E
N
mTet
E
N
mTet
E
N
mTet

图20I-L

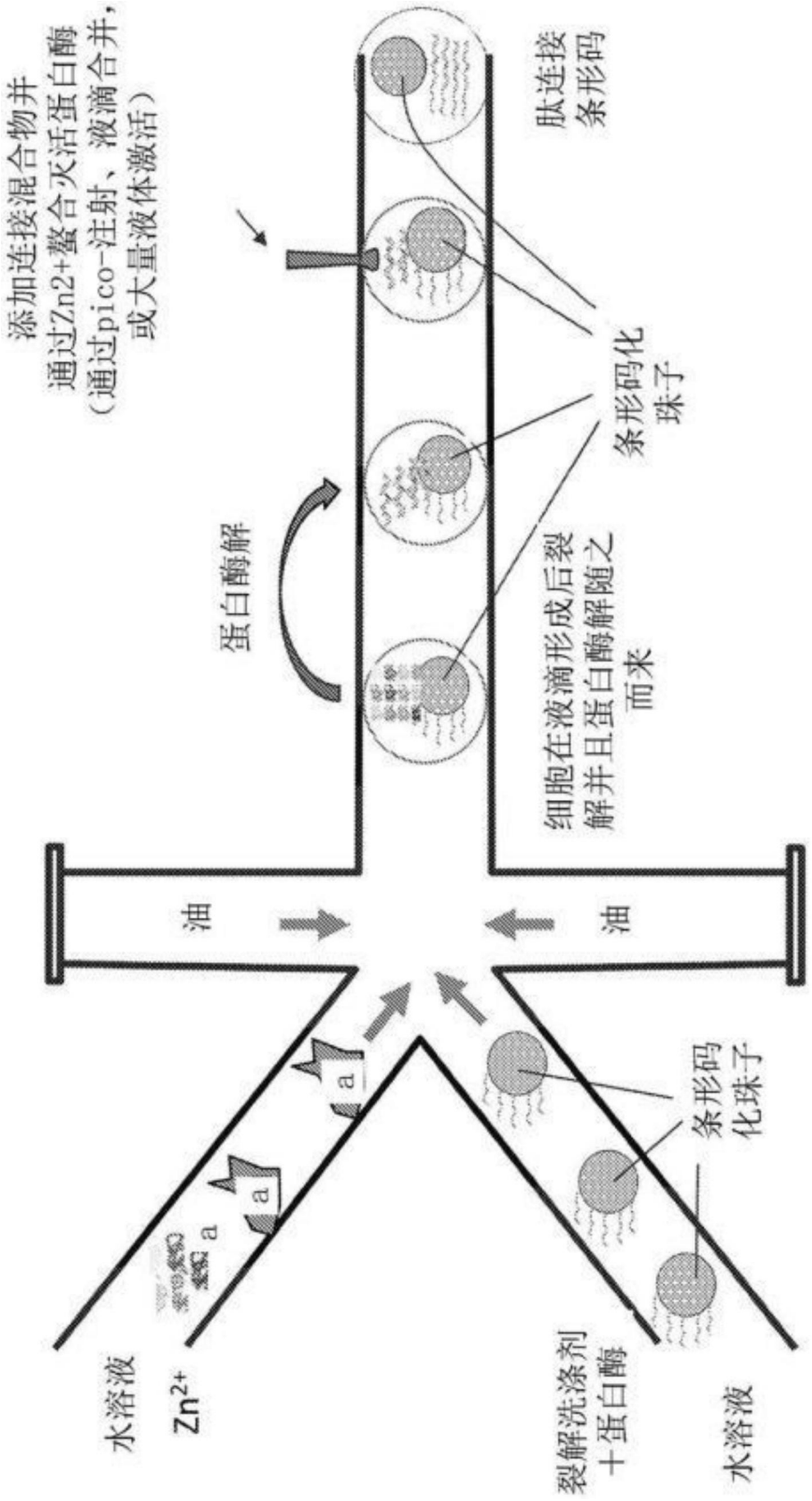

添加连接混合物并
通过Zn2+螯合灭活蛋白酶
（通过pico-注射、液滴合并，
或大量液体激活）
肽连接
条形码
条形码化
珠子
蛋白酶解
细胞在液滴形成后裂
解并且蛋白酶解随之
而来
油
油
a
a
a
水溶液
Zn^{2+}
裂解洗涤剂
+蛋白酶
条形码
化珠子
水溶液

图21

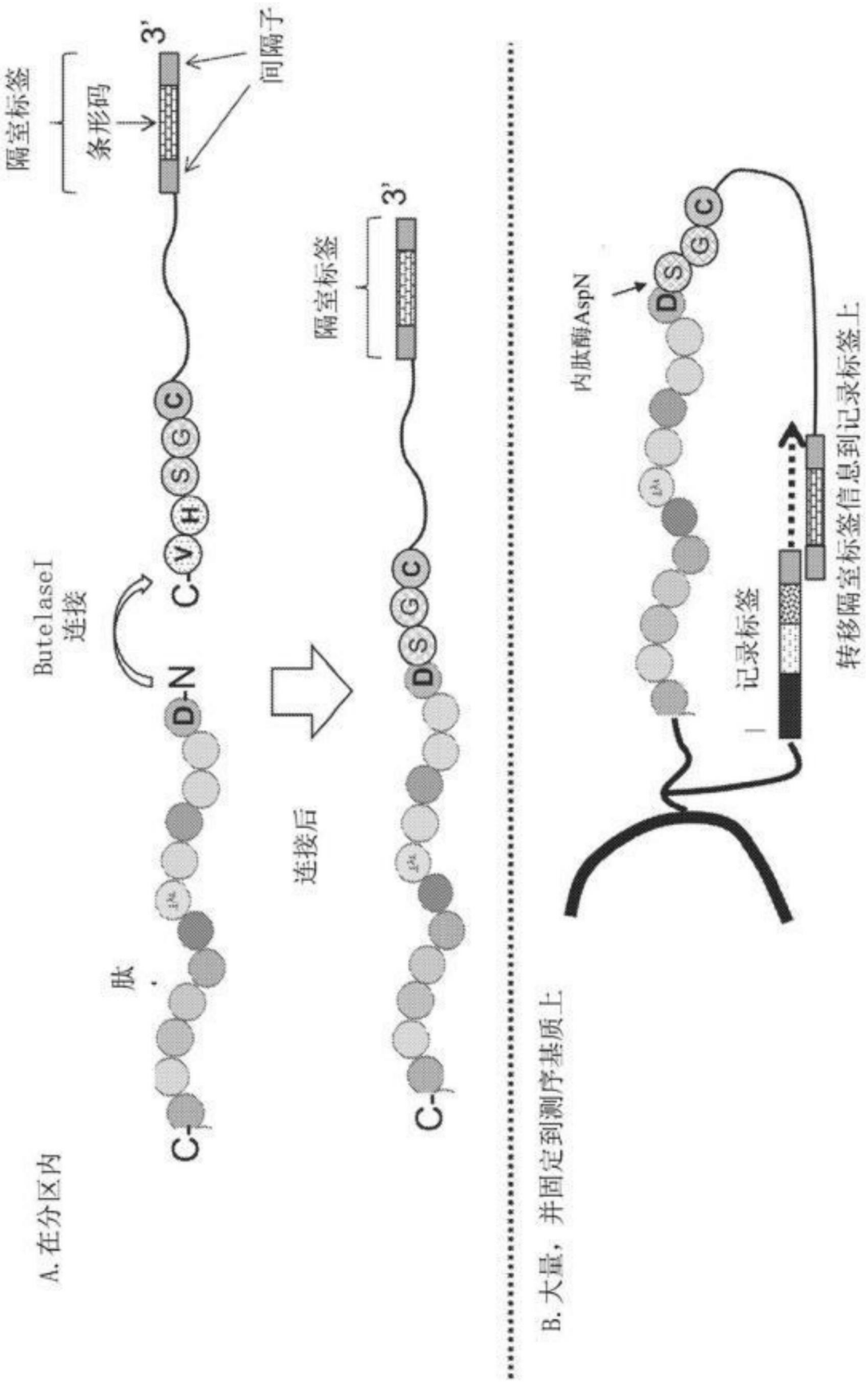

A. 在分区内
肽
C-
D-N
Butelasel
连接
C-V H S G C
隔室标签
条形码
3'
间隔子
连接后
C-
D S G C
隔室标签
3'
B. 大量，并固定到测序基质上
内肽酶AspN
D S G C
记录标签
转移隔室标签信息到记录标签上

图22

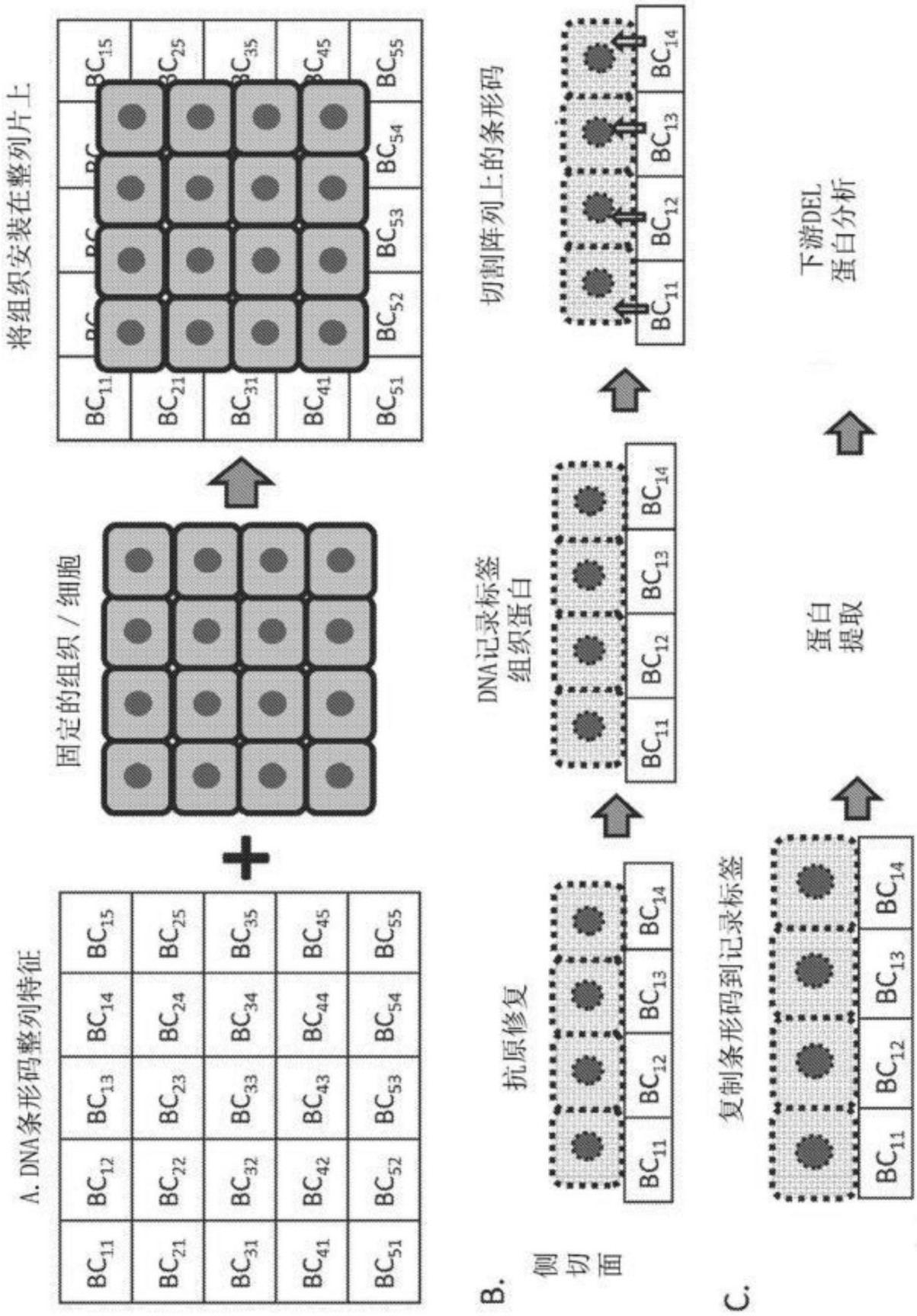

A. DNA条形码整列特征
BC11 BC12 BC13 BC14 BC15
BC21 BC22 BC23 BC24 BC25
BC31 BC32 BC33 BC34 BC35
BC41 BC42 BC43 BC44 BC45
BC51 BC52 BC53 BC54 BC55
固定的组织 / 细胞
将组织安装在整列片上
B.
侧切面
抗原修复
DNA记录标签
组织蛋白
切割阵列上的条形码
蛋白提取
C.
复制条形码到记录标签
下游DEL
蛋白分析

图23

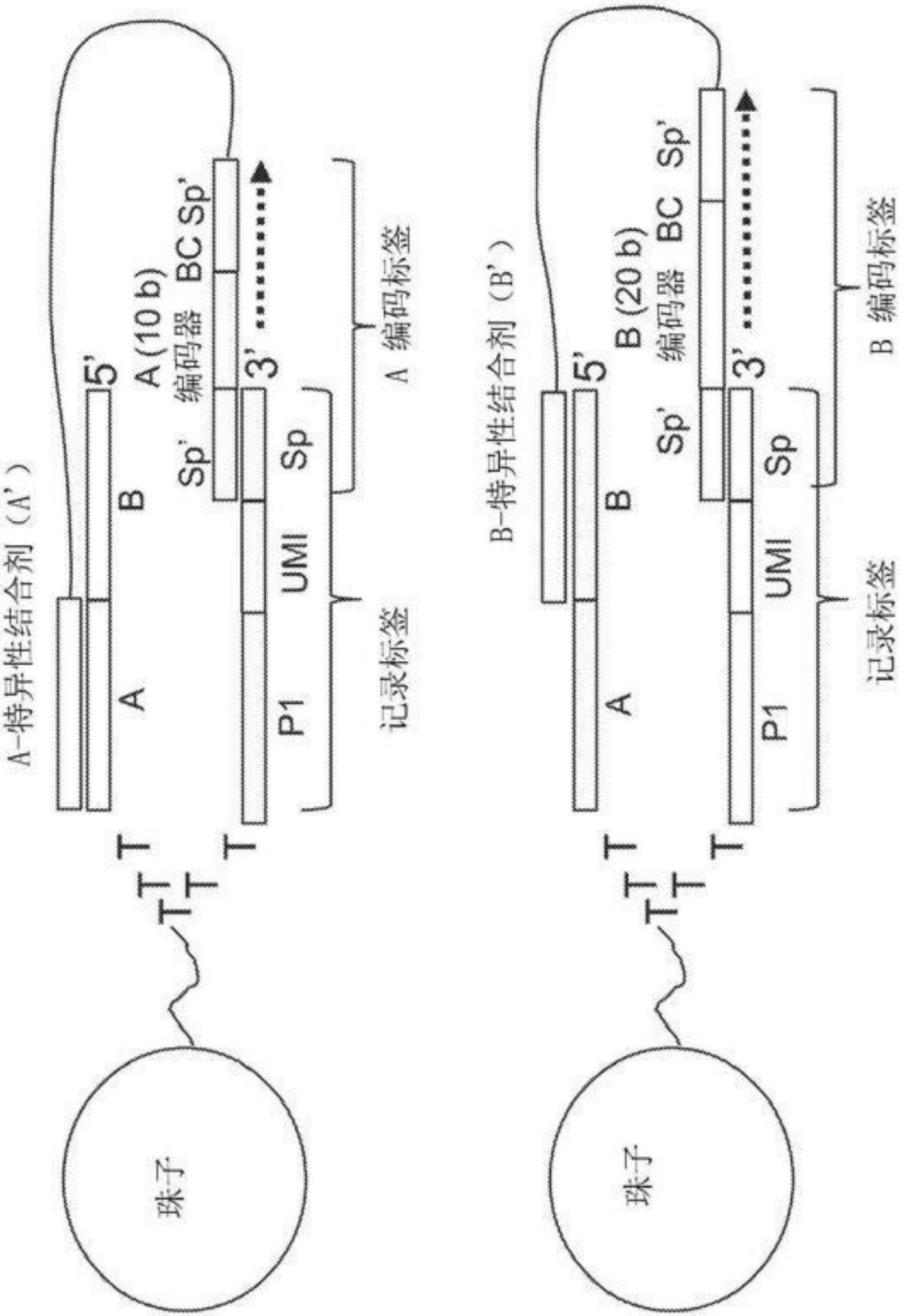

A-特异性结合剂（A’）
5’
A
B
A (10 b)
Sp’ 编码器 BC Sp’
3’
P1
UMI
Sp
A 编码标签
记录标签
珠子
B-特异性结合剂（B’）
5’
A
B
B (20 b)
Sp’ 编码器 BC Sp’
3’
P1
UMI
Sp
B 编码标签
记录标签
珠子

图24A

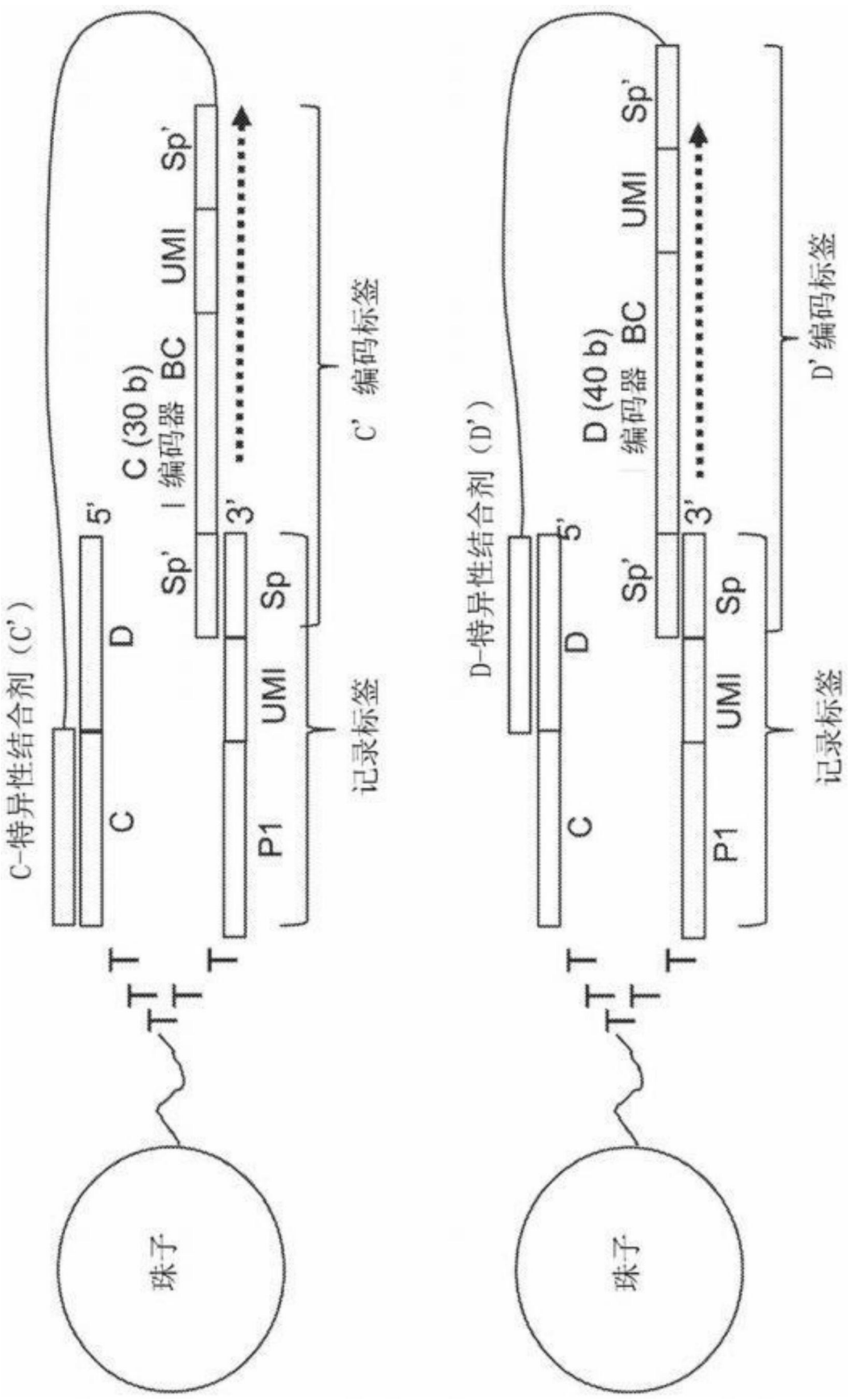

C-特异性结合剂（C’）
珠子
T
T
T
T
T
C
D
5’
C (30 b)
编码器
BC
UMI
Sp’
Sp’
3’
P1
UMI
Sp
记录标签
C’ 编码标签
D-特异性结合剂（D’）
珠子
T
T
T
T
T
C
D
5’
D (40 b)
编码器
BC
UMI
Sp’
Sp’
3’
P1
UMI
Sp
记录标签
D’ 编码标签

图24B

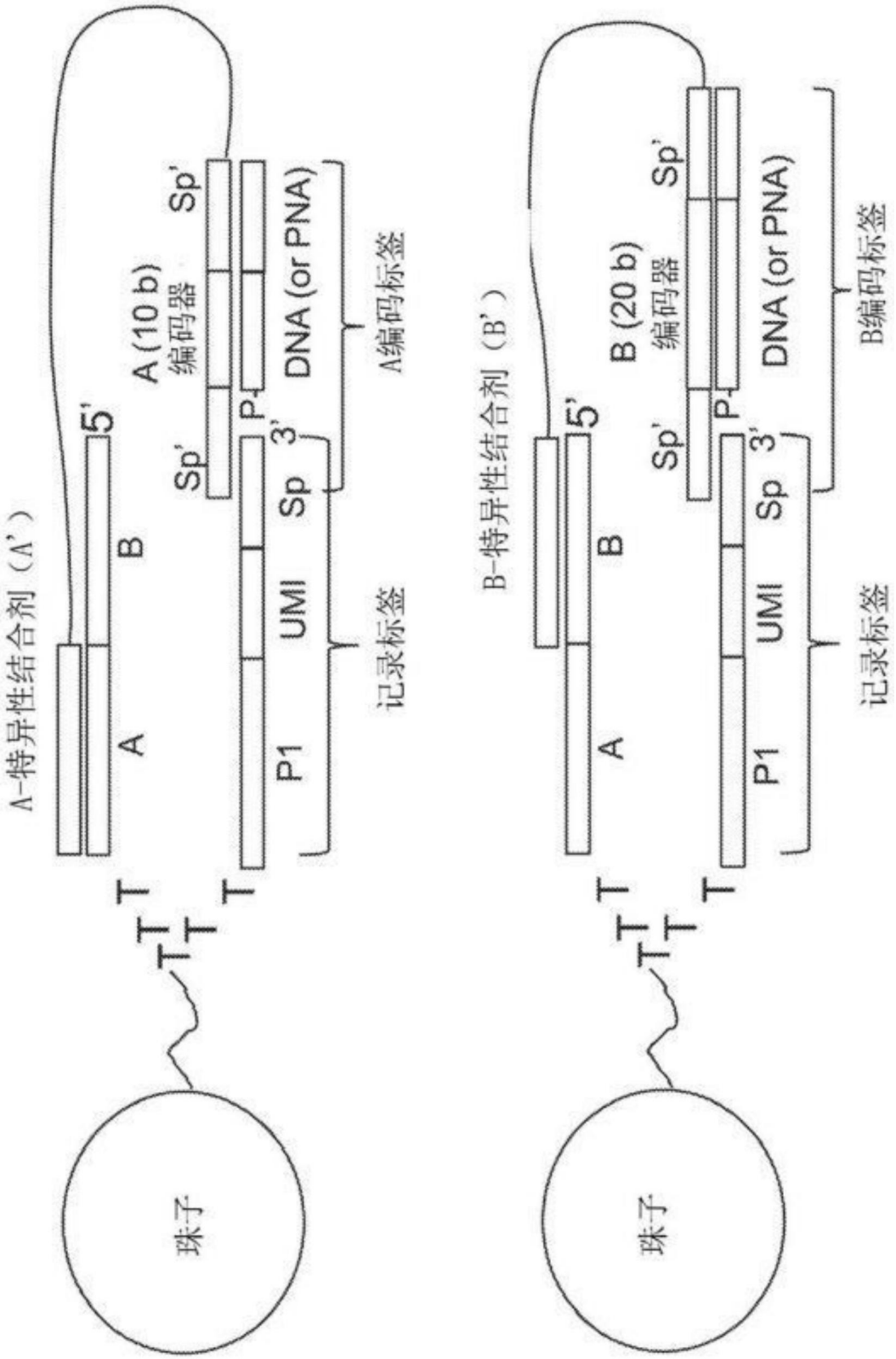

A-特异性结合剂（A’）
5’
A
B
A (10 b)
编码器
Sp’
Sp’
P2
DNA (or PNA)
3’
P1
UMI
Sp
记录标签
A编码标签
T
T
T
T
T
珠子
B-特异性结合剂（B’）
5’
A
B
B (20 b)
编码器
Sp’
Sp’
P2
DNA (or PNA)
3’
P1
UMI
Sp
记录标签
B编码标签
T
T
T
T
T
珠子

图25

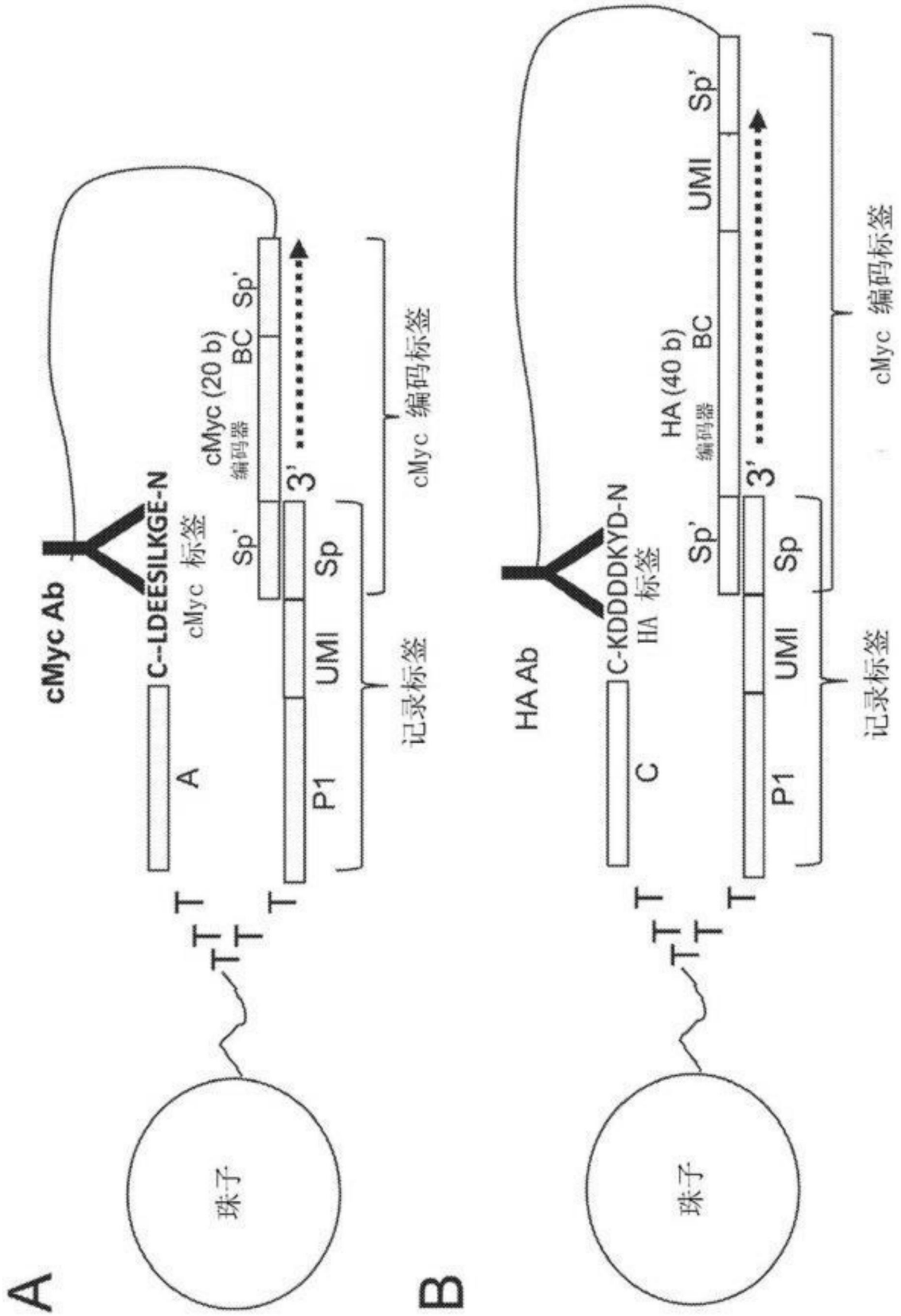
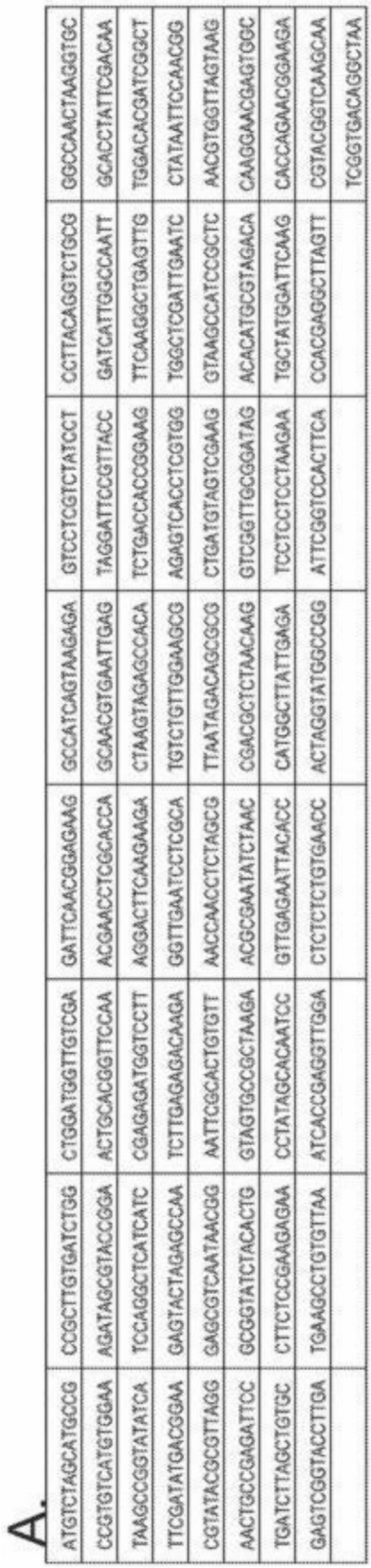

A
cMyc Ab
C--LDEESILKGE-N
cMyc 标签
A
cMyc (20 b)
编码器
BC
Sp'
Sp'
3'
P1
UMI
Sp
记录标签
cMyc 编码标签
珠子
T
T
T
T
T
B
HA Ab
C-KDDDDKYD-N
HA 标签
C
HA (40 b)
编码器
BC
Sp'
Sp'
3'
P1
UMI
Sp
记录标签
cMyc 编码标签
珠子
T
T
T
T
T

图26



A.									
ATGTCTAGCATGCCG	CCGCTTGTGATCTGG	CTGGATGGTTGTCGA	GATTCAACGGAGAAG	GCCATCAGTAAGAGA	GTCCTCGTCTATCCT	CCTTACAGGTCTGCG	GGCCAACTAAGGTGC
CCGTGTCATGTGGAA	AGATAGCGTACCGGA	ACTGCACGGTTCCAA	ACGAACCTCGCACCA	GCAACGTGAATTGAG	TAGGATTCCGTTACC	GATCATTGGCCAATT	GCACCTATTCGACAA
TAAGCCGGTATATCA	TCCAGGCTCATCATC	CGAGAGATGGTCCTT	AGGACTTCAAGAAGA	CTAAGTAGAGCCACA	TCTGACCACCGGAAG	TTCAAGGCTGAGTTG	TGGACACGATCGGCT
TTCGATATGACGGAA	GAGTACTAGAGCCAA	TCTTGAGAGACAAGA	GGTTGAATCCTCGCA	TGTCTGTTGGAAGCG	AGAGTCACCTCGTGG	TGGCTCGATTGAATC	CTATAATTCCAACGG
CGTATACGCGTTAGG	GAGCGTCAATAACGG	AATTCGCACTGTGTT	AACCAACCTCTAGCG	TTAATAGACAGCGCG	CTGATGTAGTCGAAG	GTAAGCCATCCGCTC	AACGTGGTTAGTAAG
AACTGCCGAGATTCC	GCGGTATCTACACTG	GTAGTGCCGCTAAGA	ACGCGAATATCTAAC	CGACGCTCTAACAAG	GTCGGTTGCGGATAG	ACACATGCGTAGACA	CAAGGAACGAGTGGC
TGATCTTAGCTGTGC	CTTCTCCGAAGAGAA	CCTATAGCACAATCC	GTTGAGAATTACACC	CATGGCTTATTGAGA	TCCTCCTCCTAAGAA	TGCTATGGATTCAAG	CACCAGAACGGAAGA
GAGTCGGTACCTTGA	TGAAGCCTGTGTTAA	ATCACCGAGGTTGGA	CTCTCTCTGTGAACC	ACTAGGTATGGCCGG	ATTCGGTCCACTTCA	CCACGAGGCTTAGTT	CGTACGGTCAAGCAA
							TCGGTGACAGGCTAA

图27A

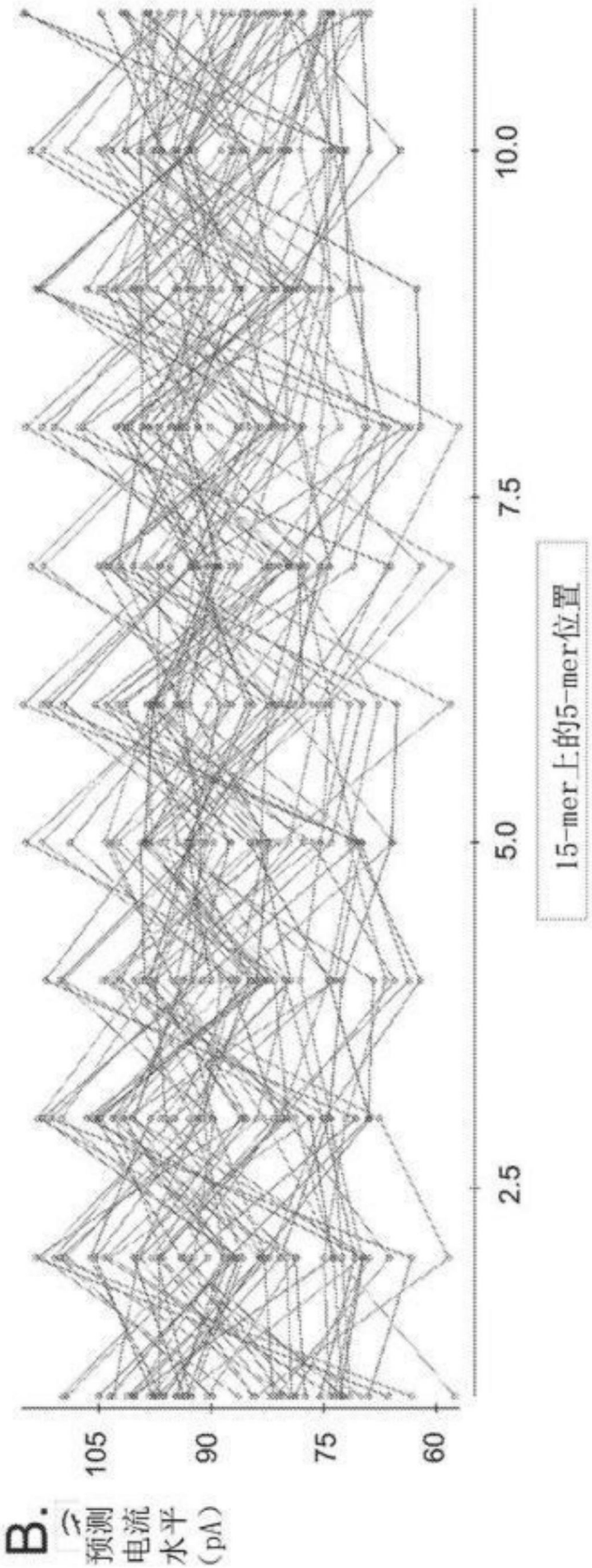


图27B

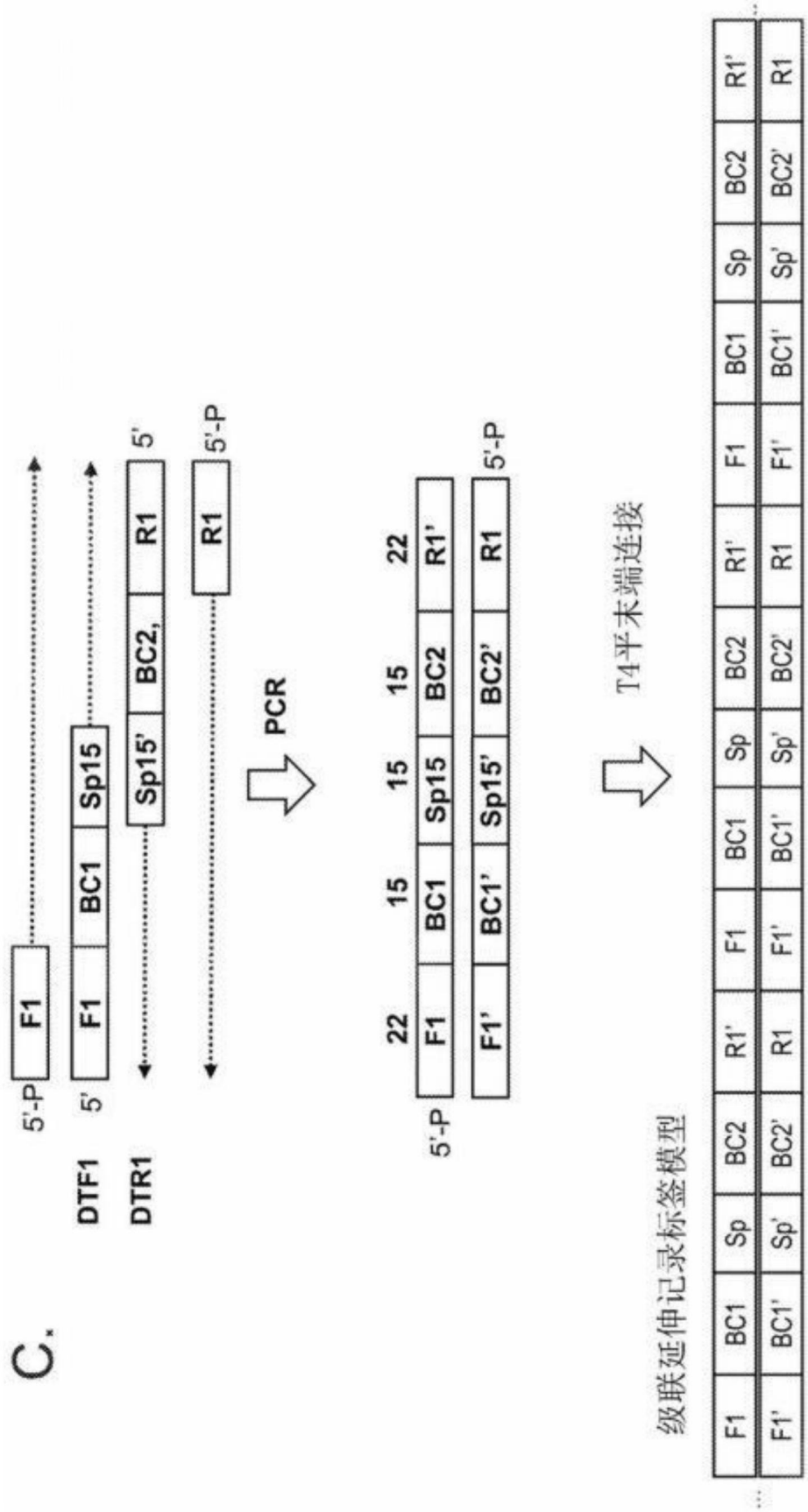

C.
5'-P
F1
DTF1
5'
F1
BC1
Sp15
DTR1
5'
Sp15'
BC2,
R1
5'-P
R1
PCR
5'-P
22
F1
15
BC1
15
Sp15
15
BC2
22
R1'
5'-P
F1'
BC1'
Sp15'
BC2'
R1
T4平末端连接
级联延伸记录标签模型
F1 BC1 Sp BC2 R1' F1 BC1 Sp BC2 R1' F1 BC1 Sp BC2 R1'
F1' BC1' Sp' BC2' R1 F1' BC1' Sp' BC2' R1 F1' BC1' Sp' BC2' R1

图27C

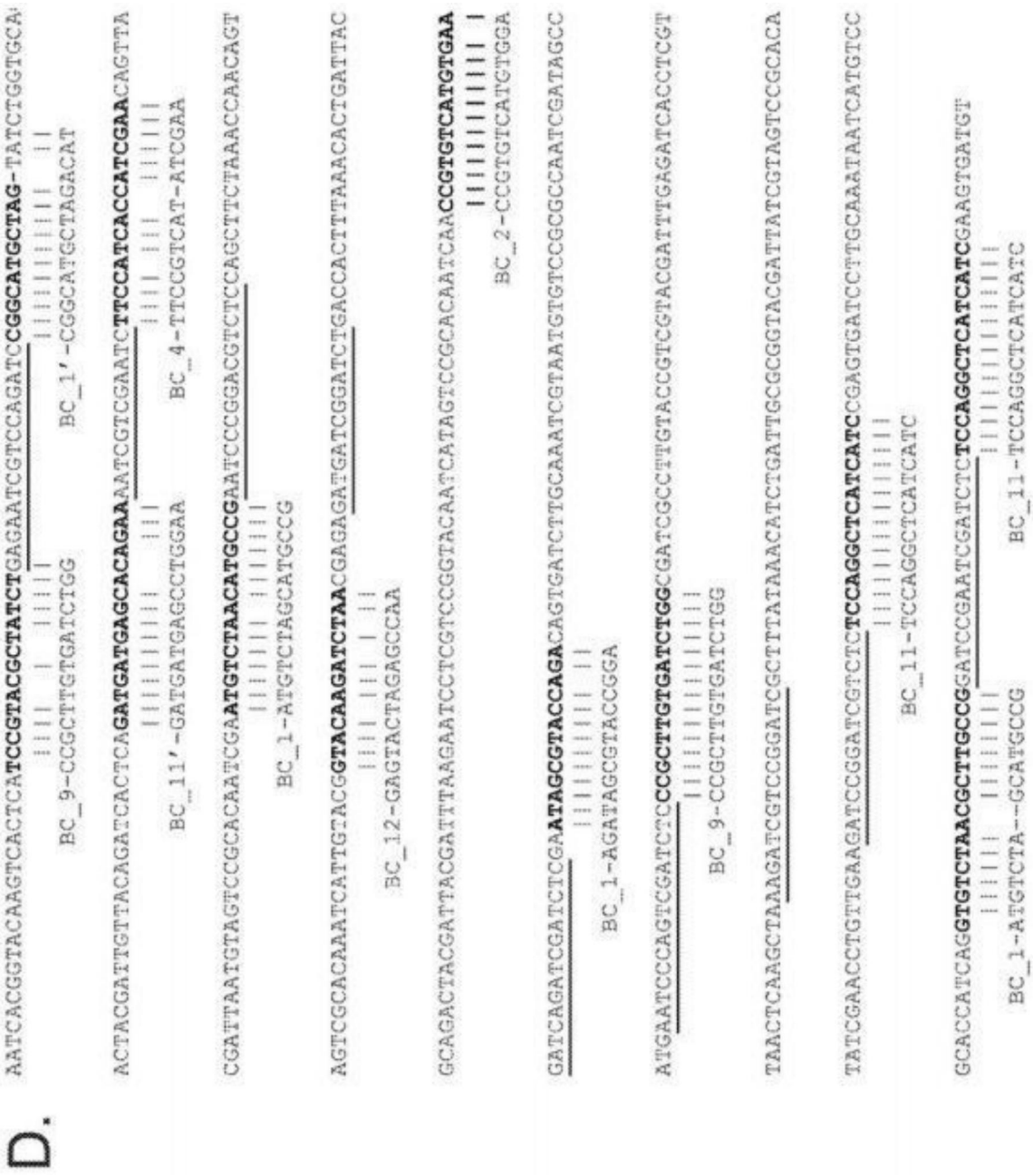

D.
AATCACGGTACAAGTCACTCATCCGTACGCTATCTGAGAATCGTCCAGATCCGGCATGCTAG-TATCTGGTGCA
BC_9-CCGCTTGTGATCTGG
BC_1'-CGGCATGCTAGACAT
ACTACGATTGTTACAGATCACTCAGATGATGAGCACAGAAAATCGTCGAATCTTCCATCACCATCGAACAGTTA
BC_11'-GATGATGAGCCTGGAA
BC_4-TTCCGTCAT-ATCGAA
CGATTAATGTAGTCCGCACAATCGAATGTCTAACATGCCGAATCCCGGACGTCTCCAGCTTCTAAACCAACAGT
BC_1-ATGTCTAGCATGCCG
AGTCGCACAAATCATTGTACGGTACAAGATCTAACGAGAGATGATCGGATCTGACCACTTTAAACACTGATTAC
BC_12-GAGTACTAGAGCCAA
GCAGACTACGATTACGATTTAAGAATCCTCGTCCGGTACAATCATAGTCCGCACAATCAACCGTGTCATGTGAA
BC_2-CCGTGTCATGTGGA
GATCAGATCGATCTCGAATAGCGTACCAGACAGTGATCTTGCAAATCGTAATGTGTCCGCGCCAATCGATAGCC
BC_1-AGATAGCGTACCGGA
ATGAATCCCAGTCGATCTCCCGCTTGTGATCTGGCGATCGCCTTGTACCGTCGTACGATTTGAGATCACCTCGT
BC_9-CCGCTTGTGATCTGG
TAACTCAAGCTAAAGATCGTCCGGATCGCTTTATAAACATCTGATTGCGCGGTACGATTATCGTAGTCCGCACA
TATCGAACCTGTTGAAGATCCGGATCGTCTCTCCAGGCTCATCATCCGAGTGATCCTTGCAAATAATCATGTCC
BC_11-TCCAGGCTCATCATC
GCACCATCAGGTGTCTAACGCTTGCCGGATCCGAATCGATCTCTCCAGGCTCATCATCGAAGTGATGT
BC_1-ATGTCTA--GCATGCCG
BC_11-TCCAGGCTCATCATC

图27D

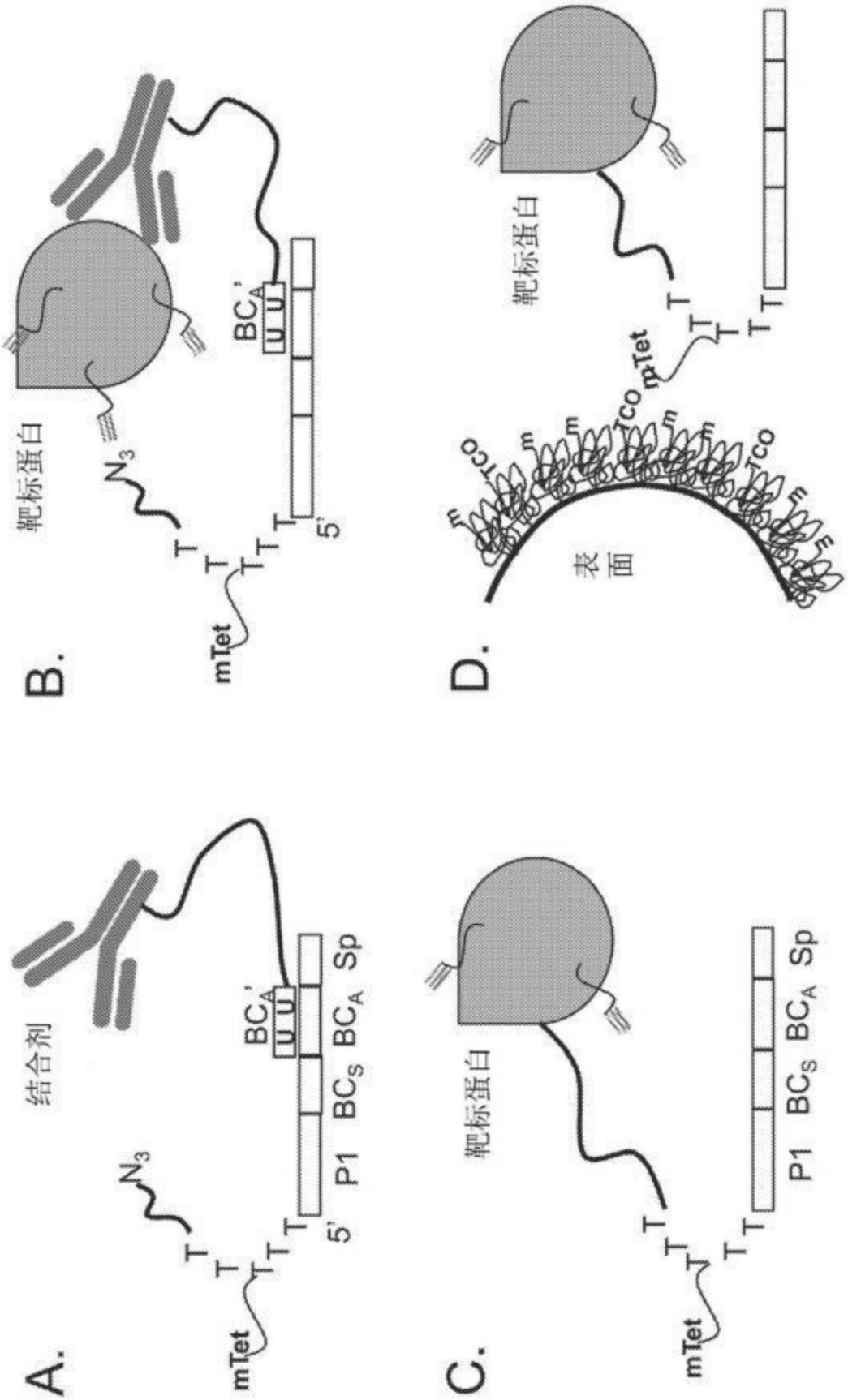

A.
结合剂
N_3
T
T
T
T
T
mTet
BC_A'
U U
5'
P1
BC_S
BC_A
Sp
B.
靶标蛋白
N_3
T
T
T
T
T
mTet
BC_A'
U U
5'
C.
靶标蛋白
T
T
T
T
T
mTet
P1
BC_S
BC_A
Sp
D.
靶标蛋白
T
T
T
T
T
mTet
TCO
m
表面

图28

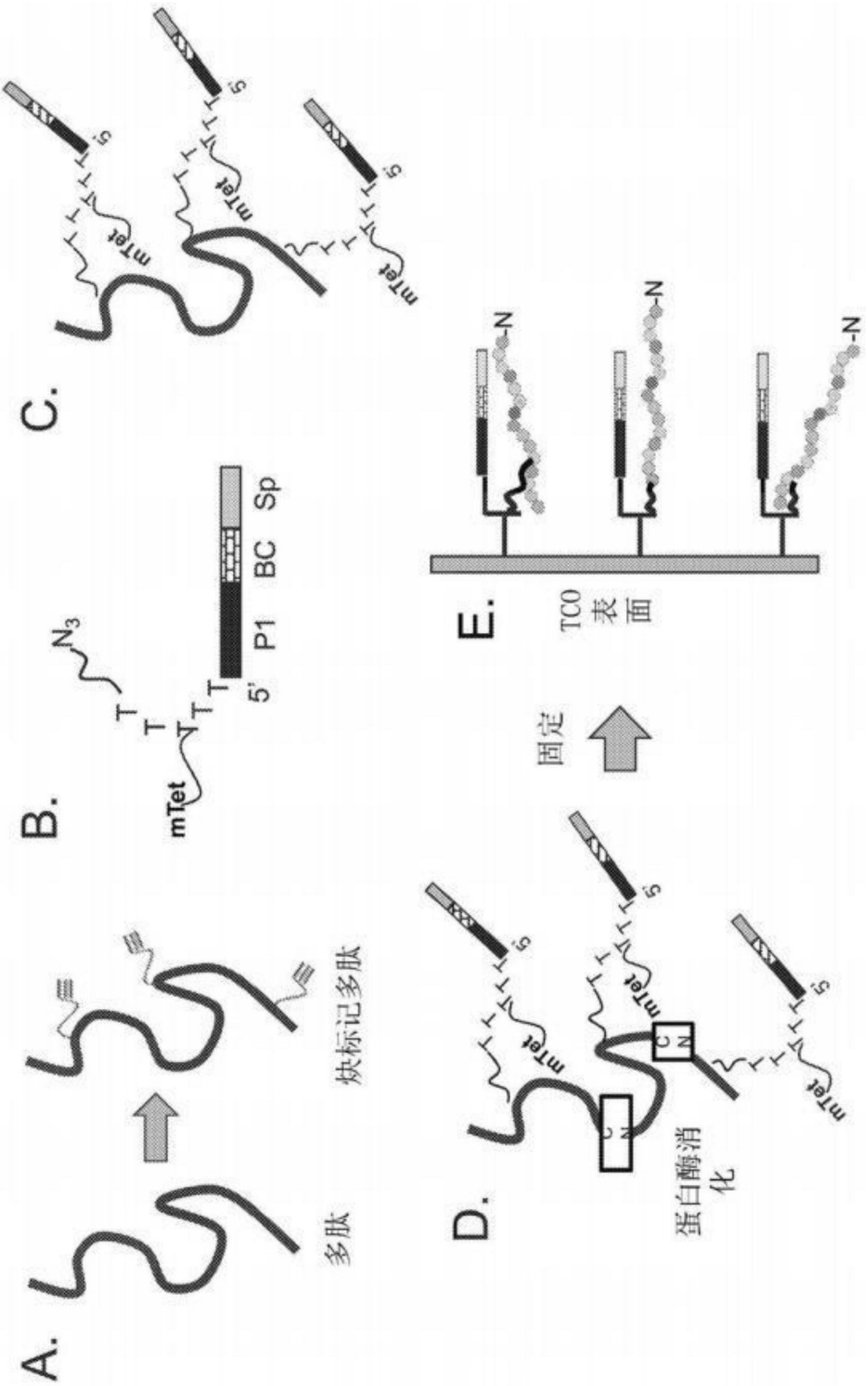


图29

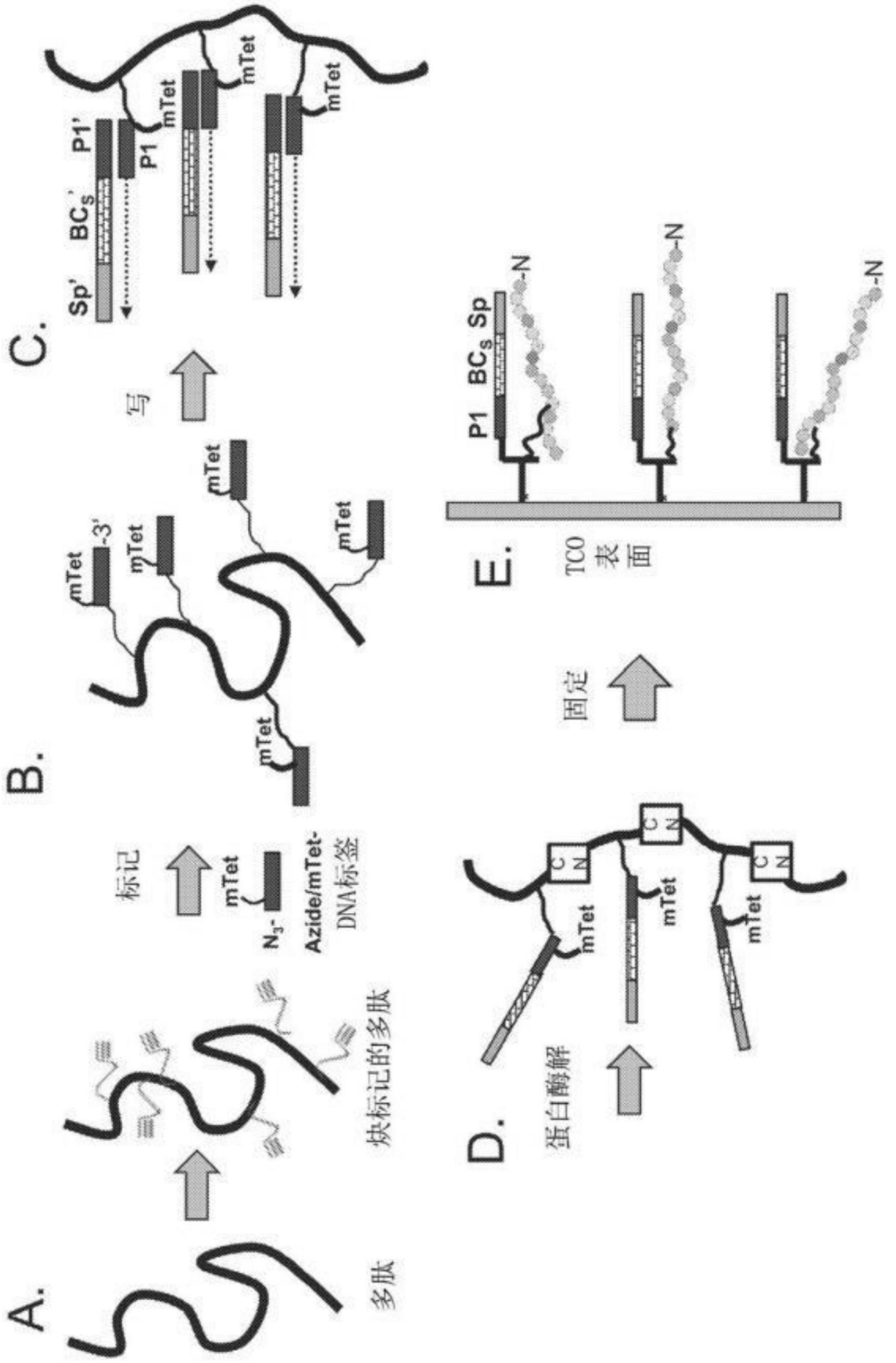

A.
多肽
标记
炔标记的多肽
mTet
N3-
Azide/mTet-
DNA标签
B.
mTet
-3'
mTet
mTet
mTet
mTet
写
C.
Sp'
BCs'
P1'
P1
mTet
mTet
mTet
D.
蛋白酶解
C
N
mTet
mTet
mTet
固定
E.
TCO
表
面
P1
BCs
Sp
-N
-N
-N

图30

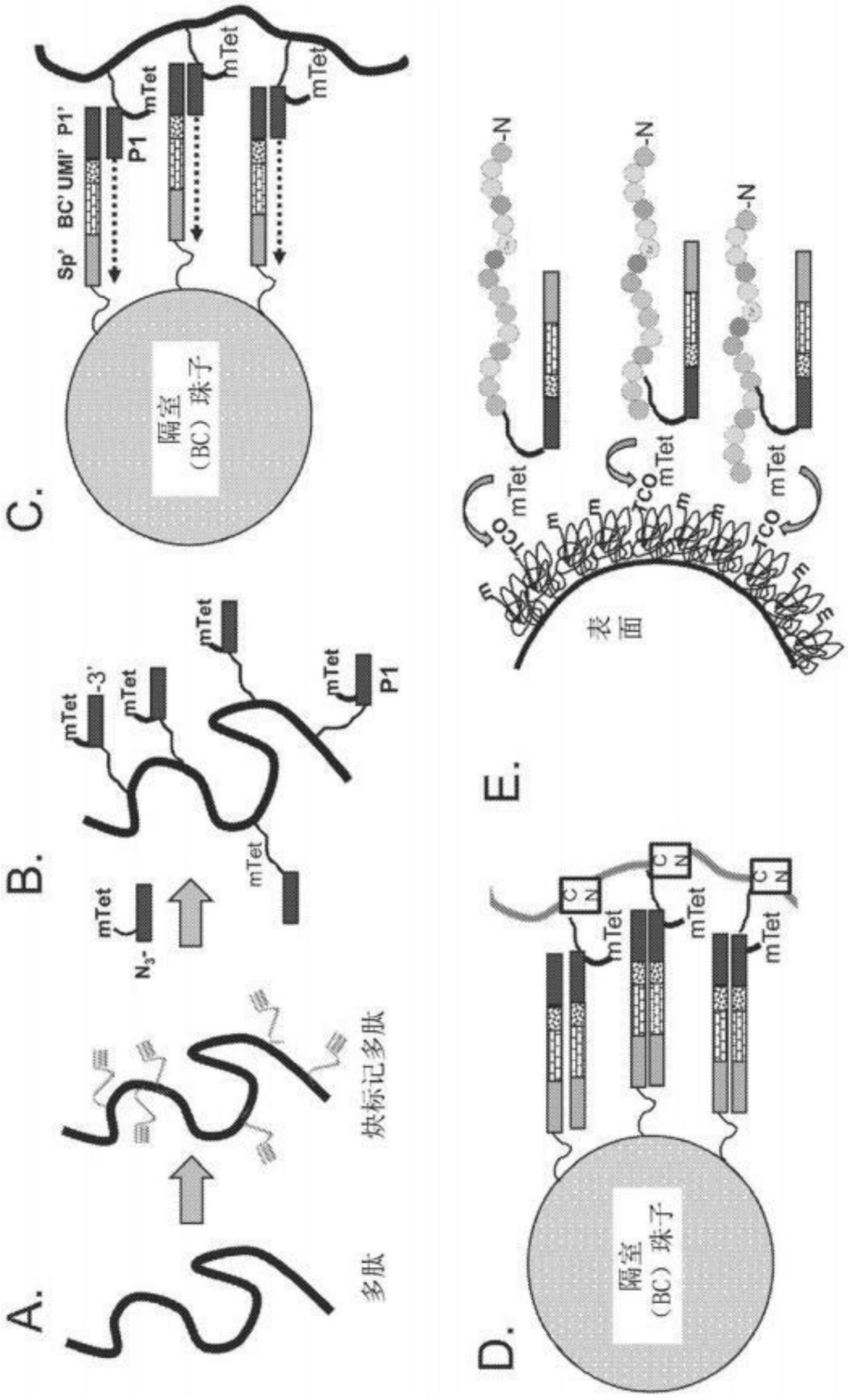

A.
多肽
炔标记多肽
B.
mTet
N3-
mTet
-3'
mTet
mTet
mTet
P1
C.
隔室
（BC）珠子
Sp'
BC'UMI'
P1'
P1
mTet
mTet
mTet
D.
隔室
（BC）珠子
C
N
mTet
E.
表面
TCO
mTet
-N

图31

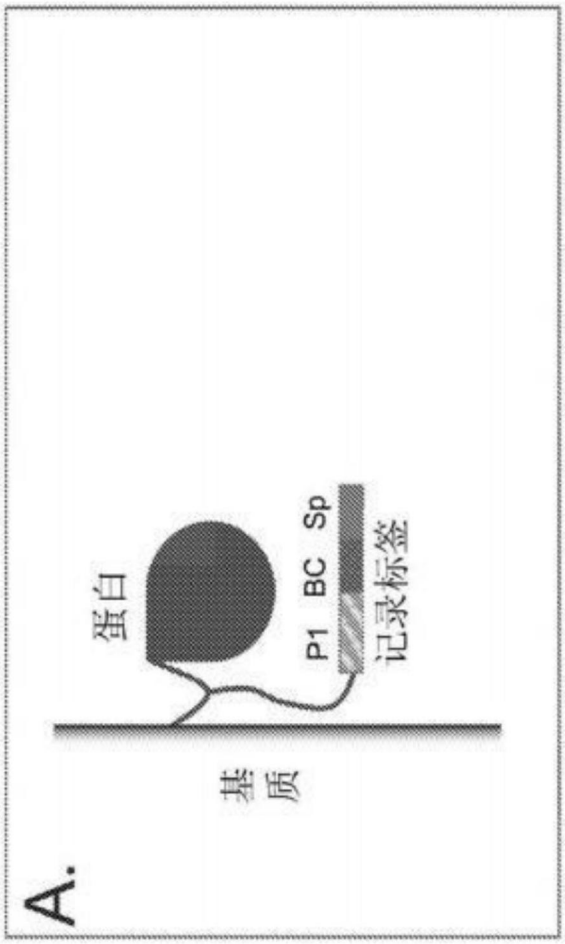


图32A

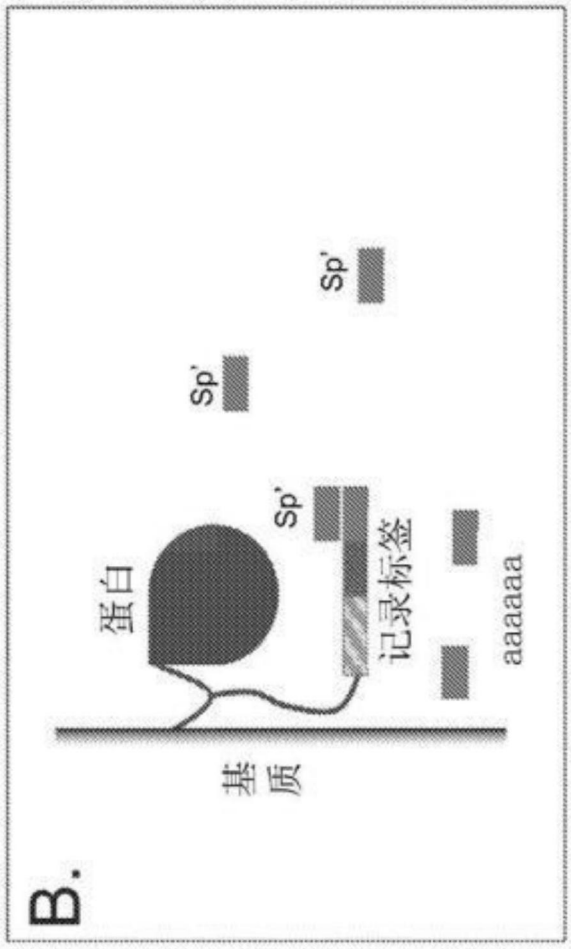


图32B

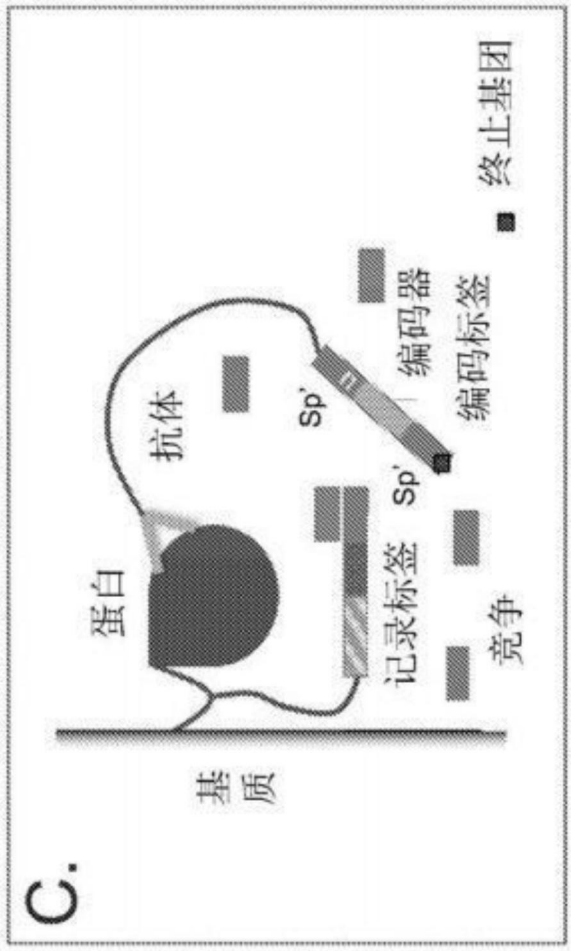


图32C

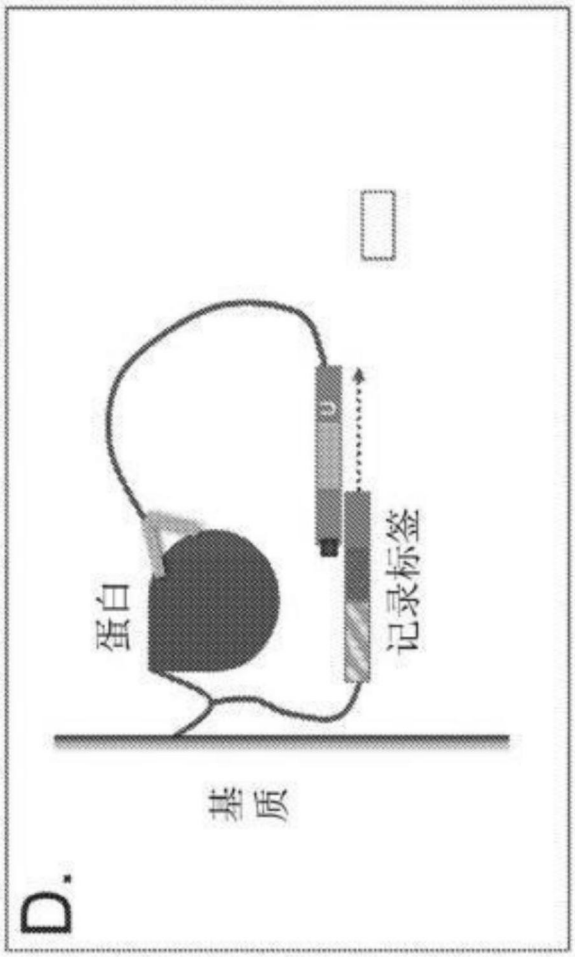


图32D

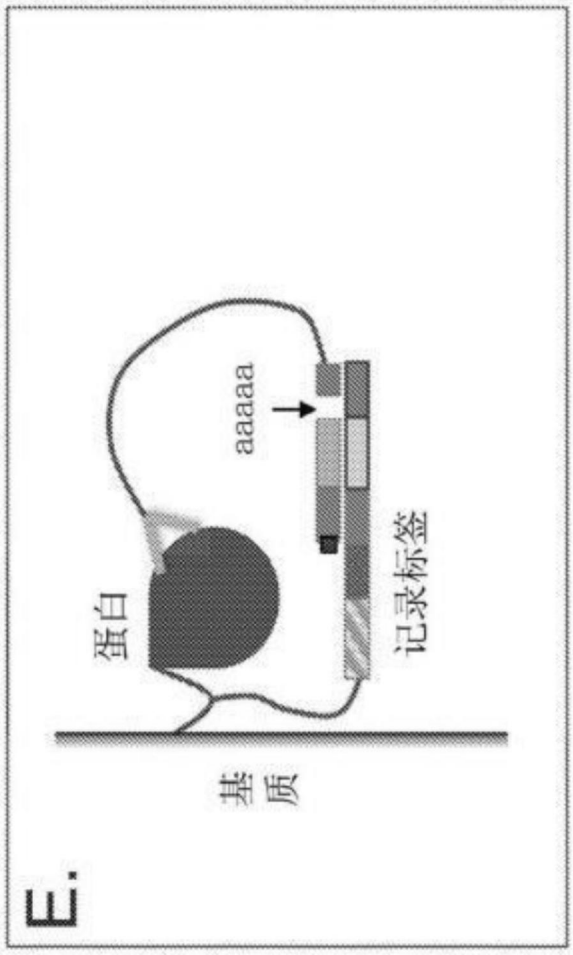


图32E

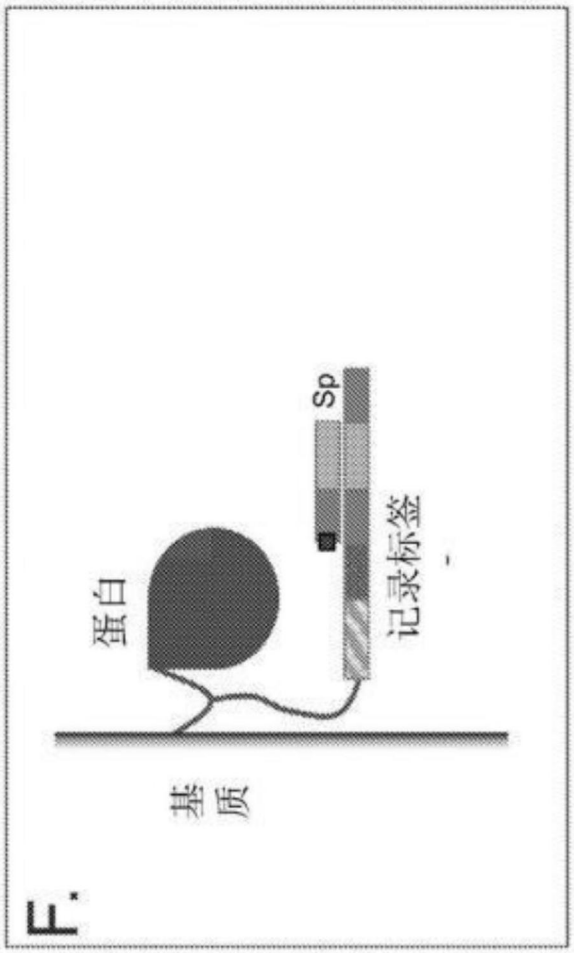


图32F

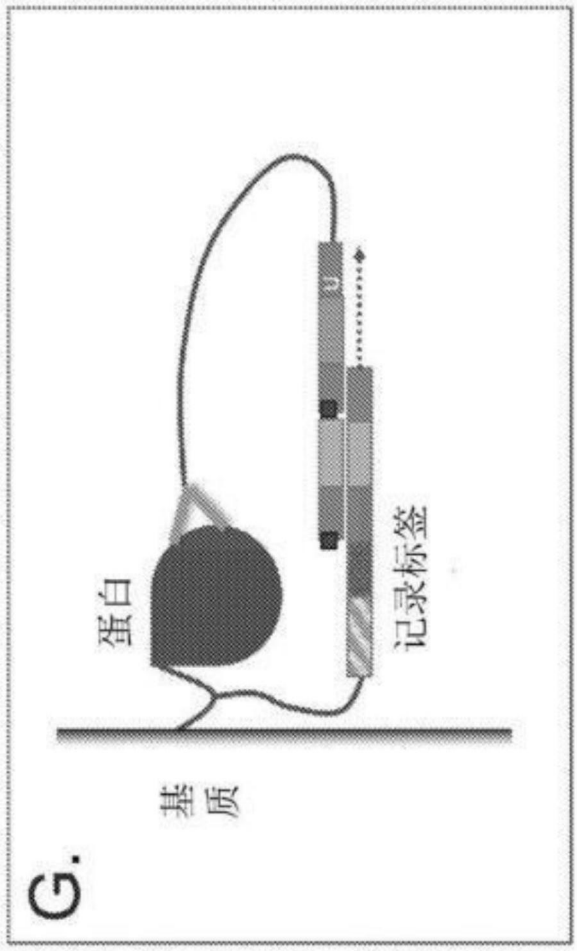


图32G

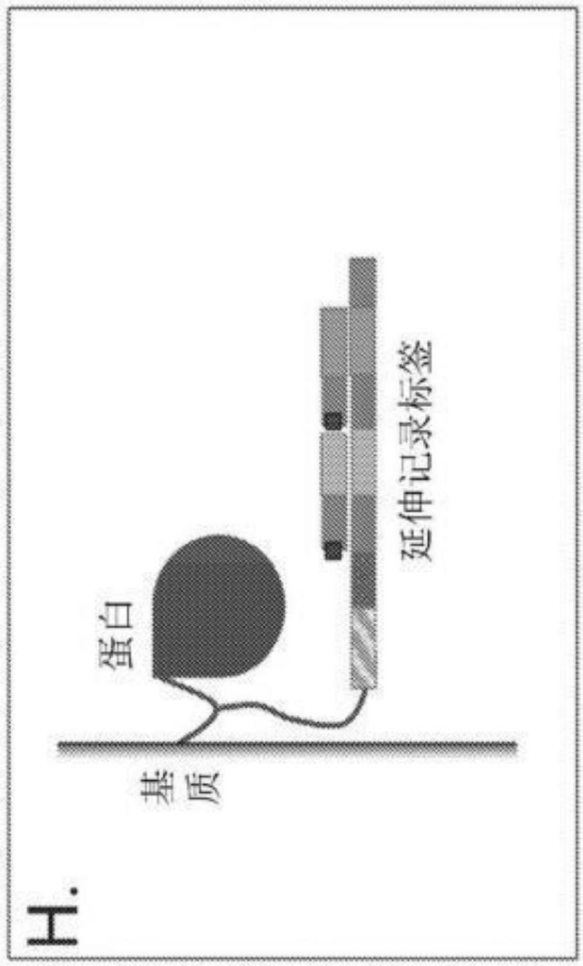


图32H

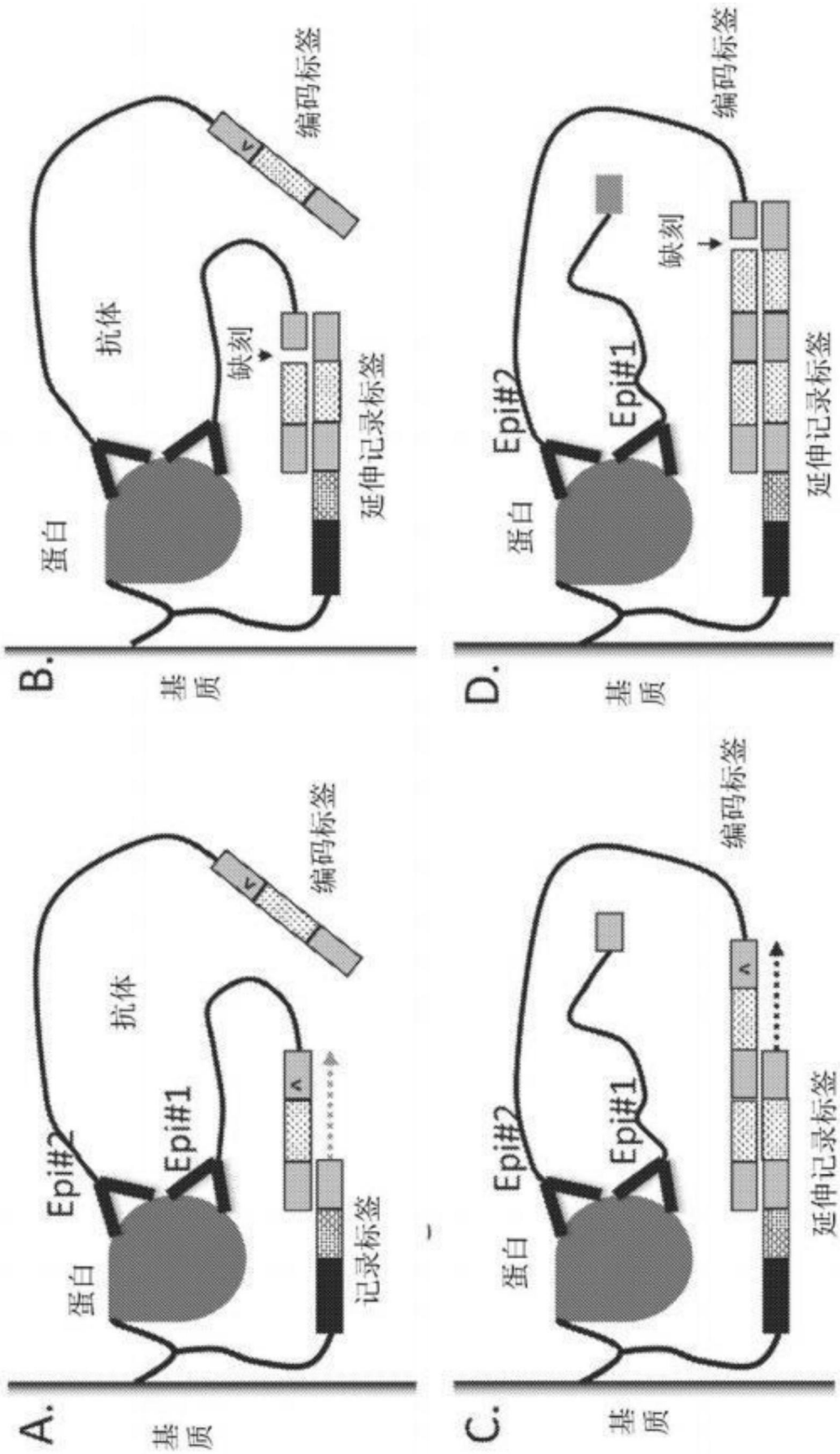

A.
蛋白
Epi#2
Epi#1
抗体
编码标签
记录标签
基质
B.
蛋白
抗体
缺刻
编码标签
延伸记录标签
基质
C.
蛋白
Epi#2
Epi#1
编码标签
延伸记录标签
基质
D.
蛋白
Epi#2
Epi#1
缺刻
编码标签
延伸记录标签
基质

图33

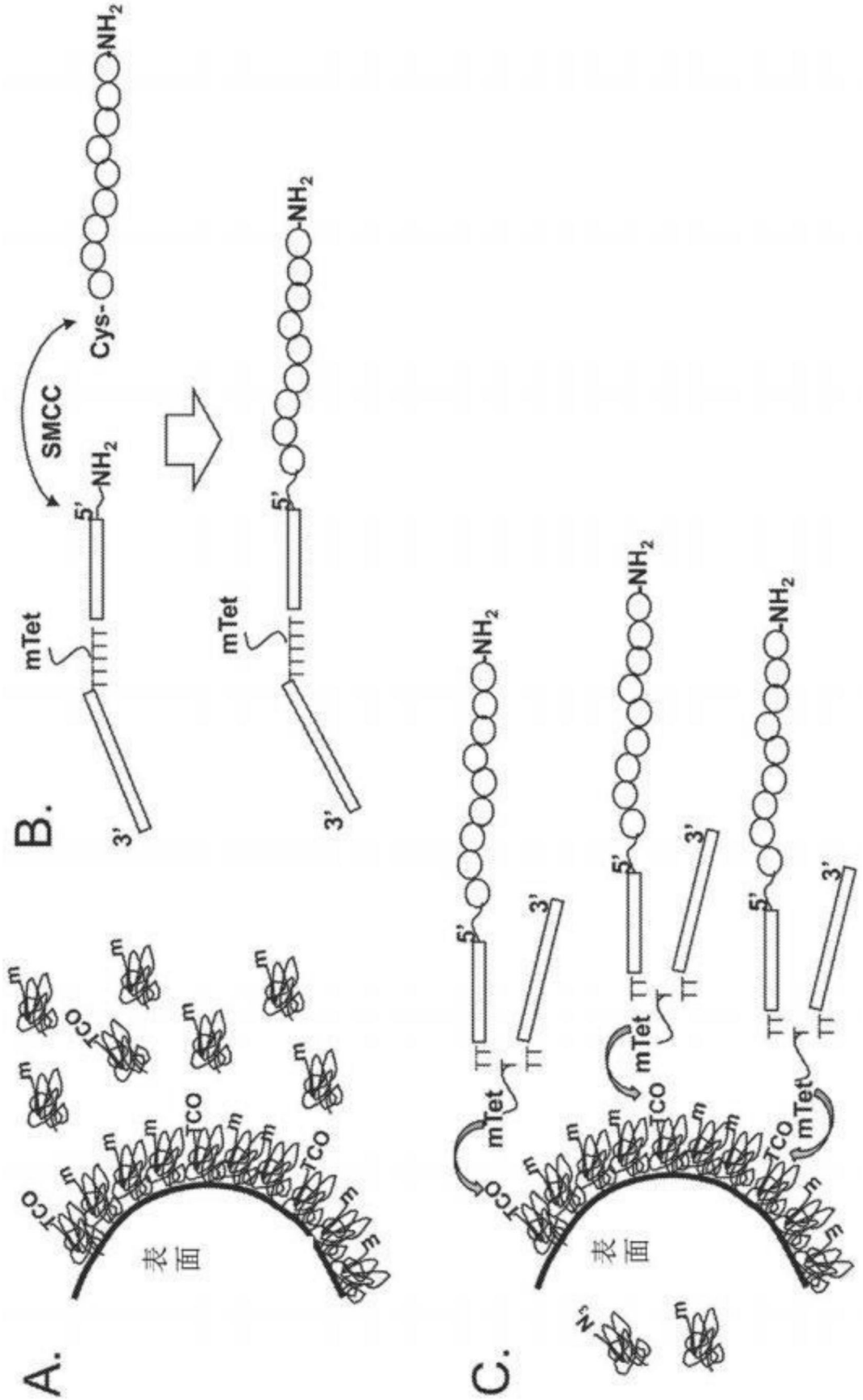

B.
SMCC
Cys-
NH2
5'
NH2
mTet
3'
mTet
5'
NH2
3'
A.
表面
TCO
m
C.
表面
mTet
TCO
5'
3'
NH2

图34

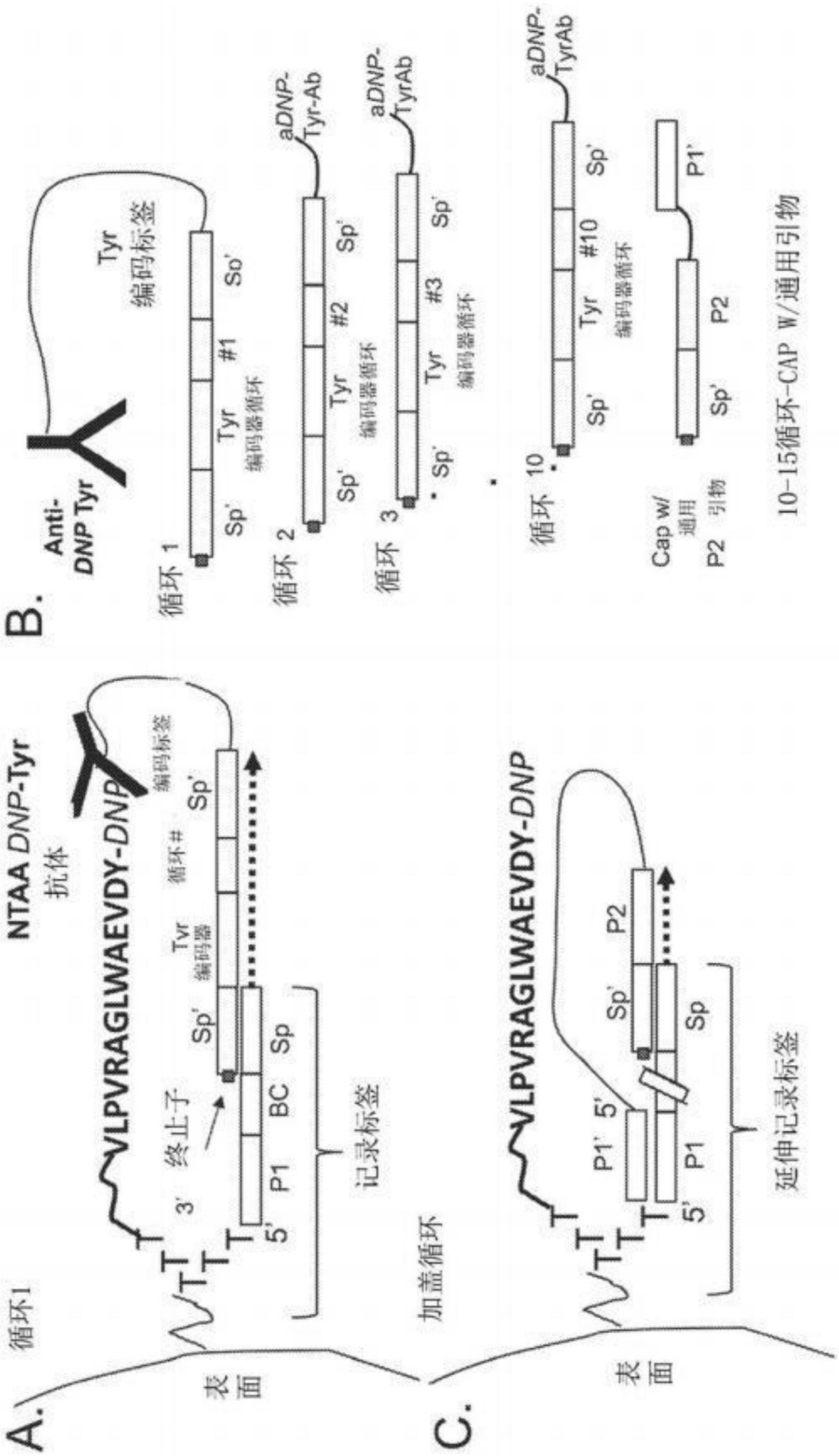

A.
循环1
表面
NTAA DNP-Tyr
抗体
VLPVRAGLWAEVDY-DNP
编码标签
Tyr
编码器
循环#
终止子
3'
Sp'
5'
P1
BC
Sp
记录标签
加盖循环
B.
Anti-
DNP Tyr
Tyr
编码标签
循环 1
Sp'
Tyr
#1
So'
编码器循环
循环 2
aDNP-
Tyr-Ab
Sp'
Tyr
#2
Sp'
编码器循环
循环 3
aDNP-
TyrAb
Sp'
Tyr
#3
Sp'
编码器循环
循环 10
aDNP-
TyrAb
Sp'
Tyr
#10
Sp'
编码器循环
Cap w/
通用
引物
P2
P1'
Sp'
P2
10-15循环-CAP W/通用引物
C.
表面
VLPVRAGLWAEVDY-DNP
P1'
5'
Sp'
P2
5'
P1
Sp
延伸记录标签

图35

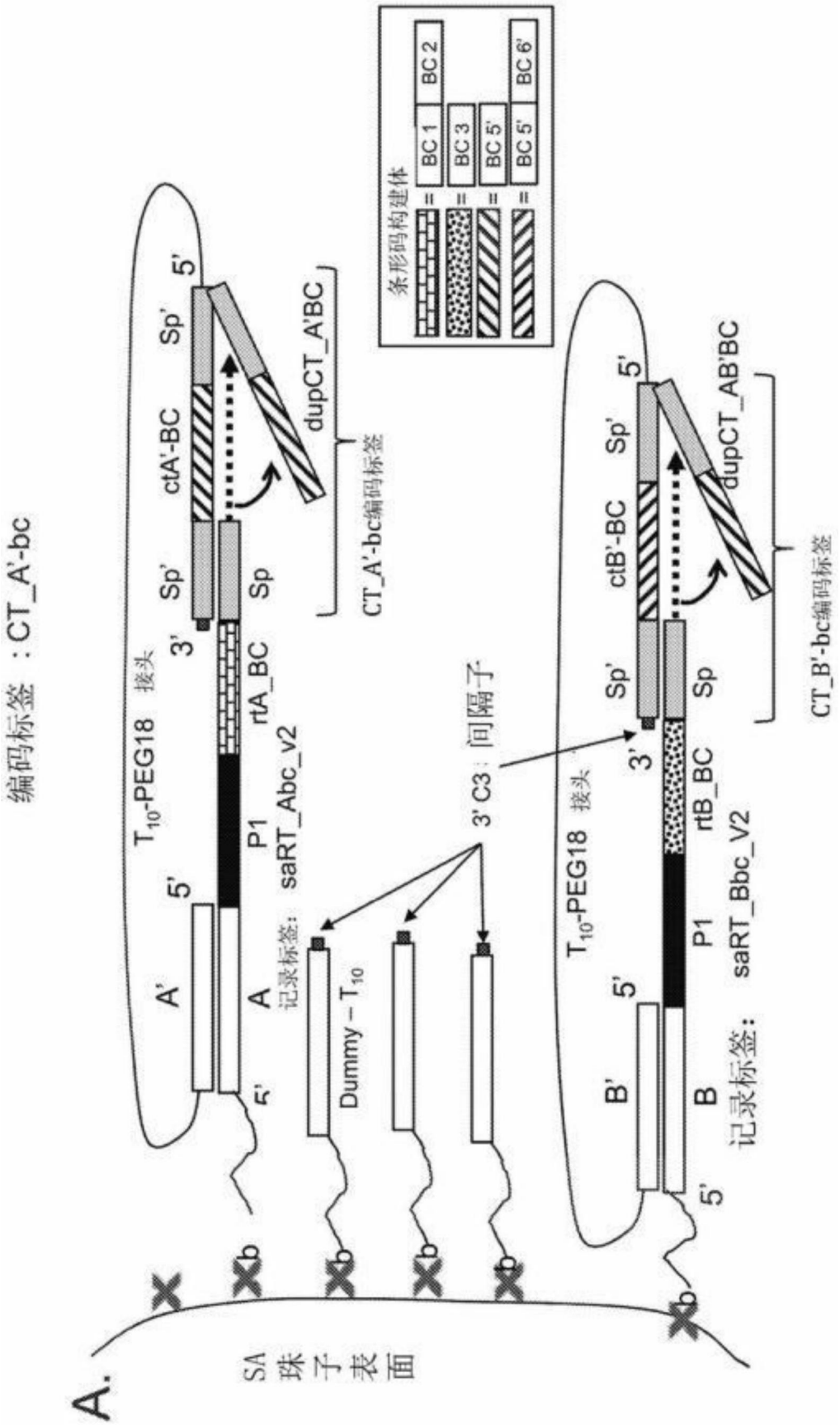

A.
SA 珠子表面
编码标签 ：CT_A'-bc
条形码构建体
BC 1
BC 2
BC 3
BC 5'
BC 5'
BC 6'
T10-PEG18 接头
A'
5'
Sp'
ctA'-BC
Sp'
5'
dupCT_A'BC
CT_A'-bc编码标签
3'
Sp
rtA_BC
P1
saRT_Abc_v2
A
记录标签：
Dummy – T10
3' C3：间隔子
T10-PEG18 接头
B'
5'
Sp'
ctB'-BC
Sp'
5'
dupCT_AB'BC
CT_B'-bc编码标签
3'
Sp
rtB_BC
P1
saRT_Bbc_V2
B
记录标签：
5'

图36A

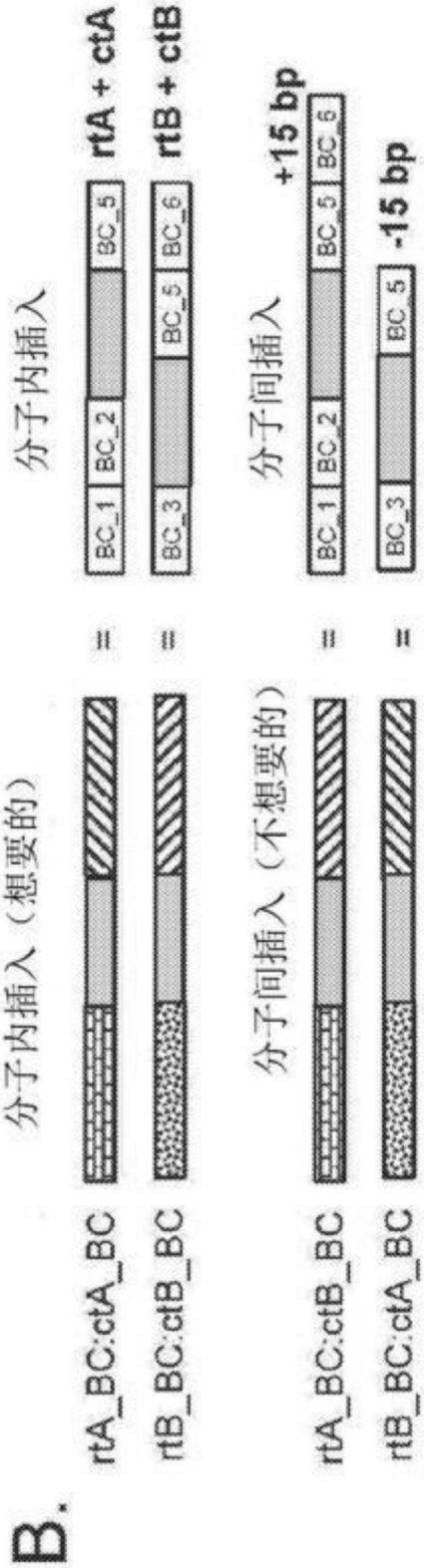

B.
分子内插入（想要的）
rtA_BC:ctA_BC
rtB_BC:ctB_BC
分子内插入
BC_1
BC_2
BC_5
rtA + ctA
BC_3
BC_5
BC_6
rtB + ctB
分子间插入（不想要的）
rtA_BC:ctB_BC
rtB_BC:ctA_BC
分子间插入
+15 bp
-15 bp

图36B

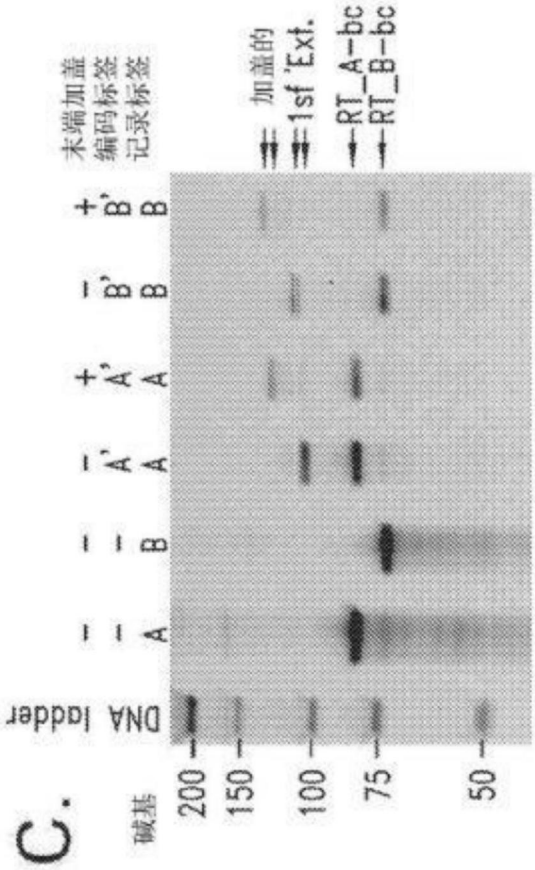

C.
碱基
200
150
100
75
50
DNA ladder
末端加盖
编码标签
记录标签
A
B
A'
A
B'
B
+'
A'
A
−
B'
B
+'
B'
B
加盖的
1sf 'Ext.
RT_A-bc
RT_B-bc

图36C

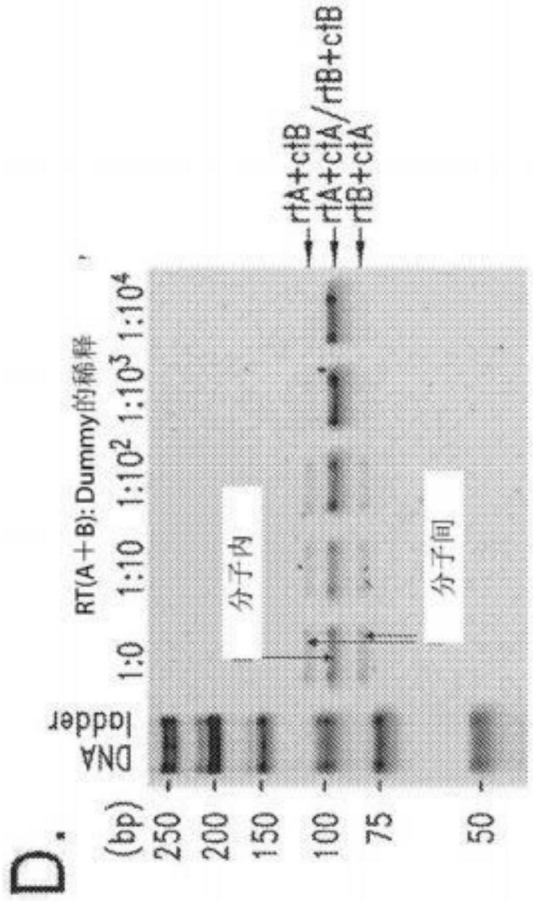

D.
(bp)
250
200
150
100
75
50
DNA ladder
RT(A+B): Dummy的稀释
1:0
1:10
1:10^2
1:10^3
1:10^4
分子内
分子间
rtA+ctB
rtA+ctA/rtB+ctB
rtB+ctA

图36D

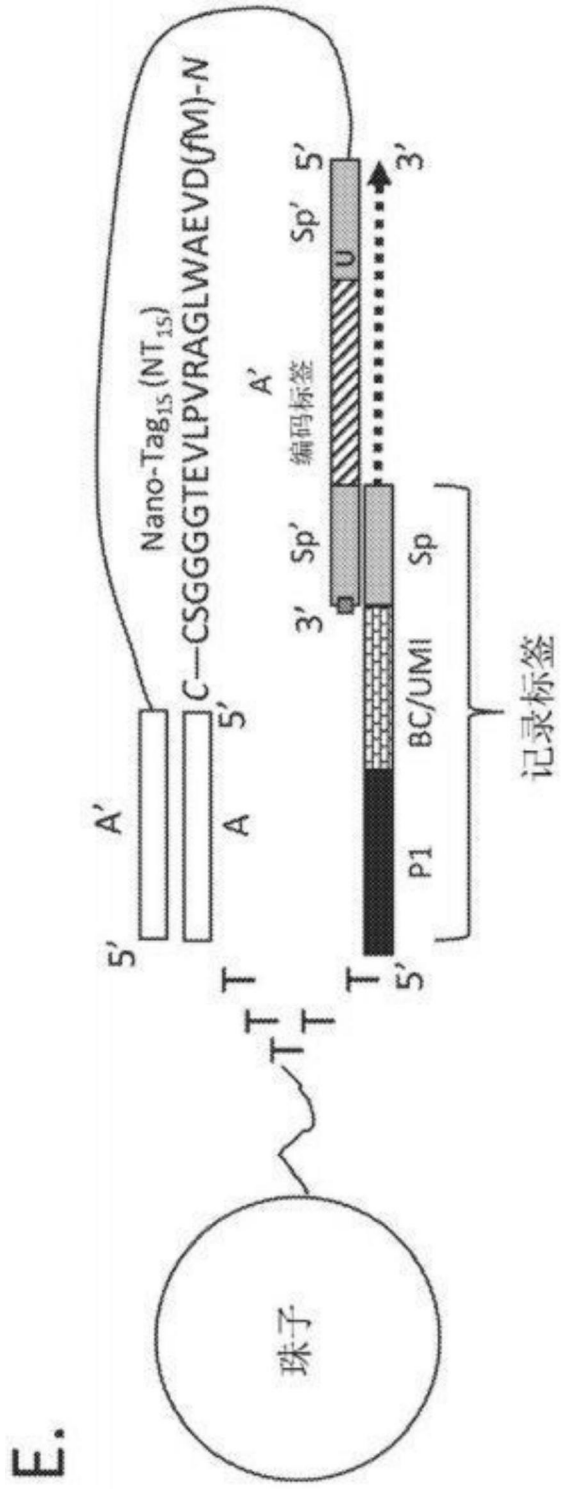

E.
珠子
T
T
T
T
5'
A'
5'
A
C—CSGGGGTEVLPVRAGLWAEVD(fM)-N
Nano-Tag15 (NT15)
3'
Sp'
P1
BC/UMI
Sp
记录标签
A'
编码标签
Sp'
U
5'
3'

图36E

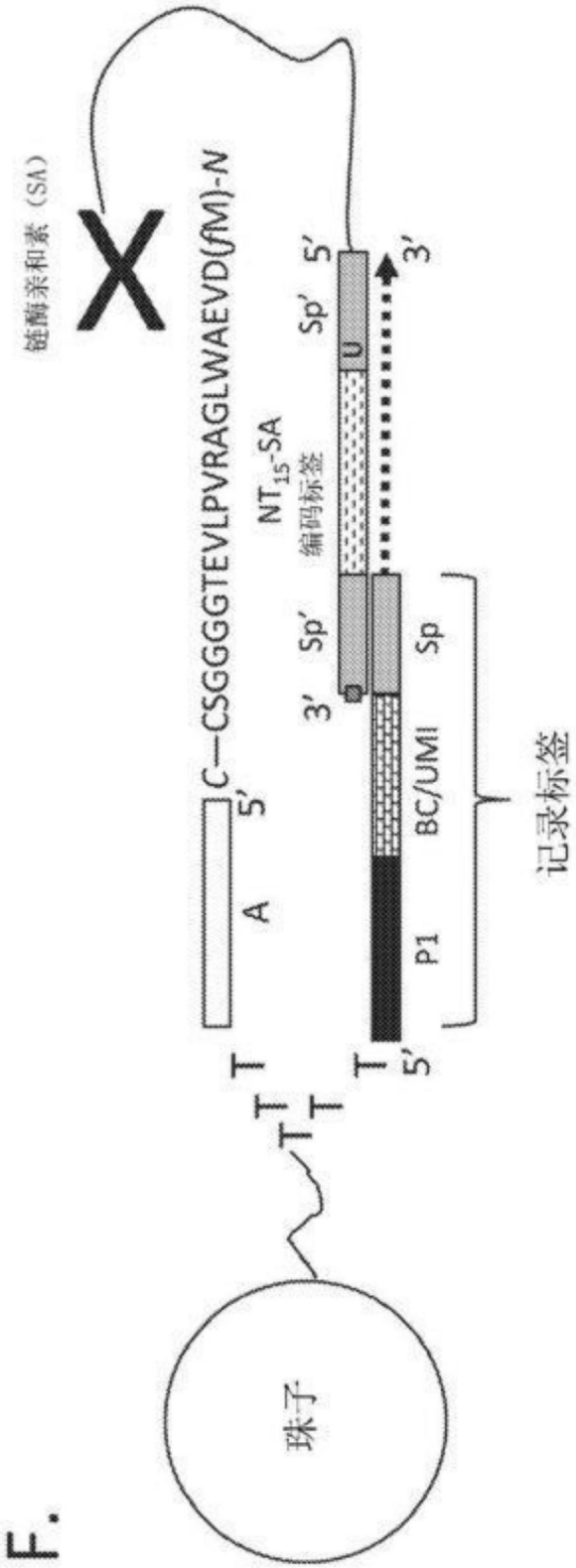

链酶亲和素（SA）
C—CSGGGGTEVLPVRAGLWAEVD(fM)-N
NT15-SA
编码标签
A
5'
3'
Sp'
U
Sp'
5'
3'
BC/UMI
Sp
P1
5'
T
T
T
T
T
记录标签
珠子
F.

图36F

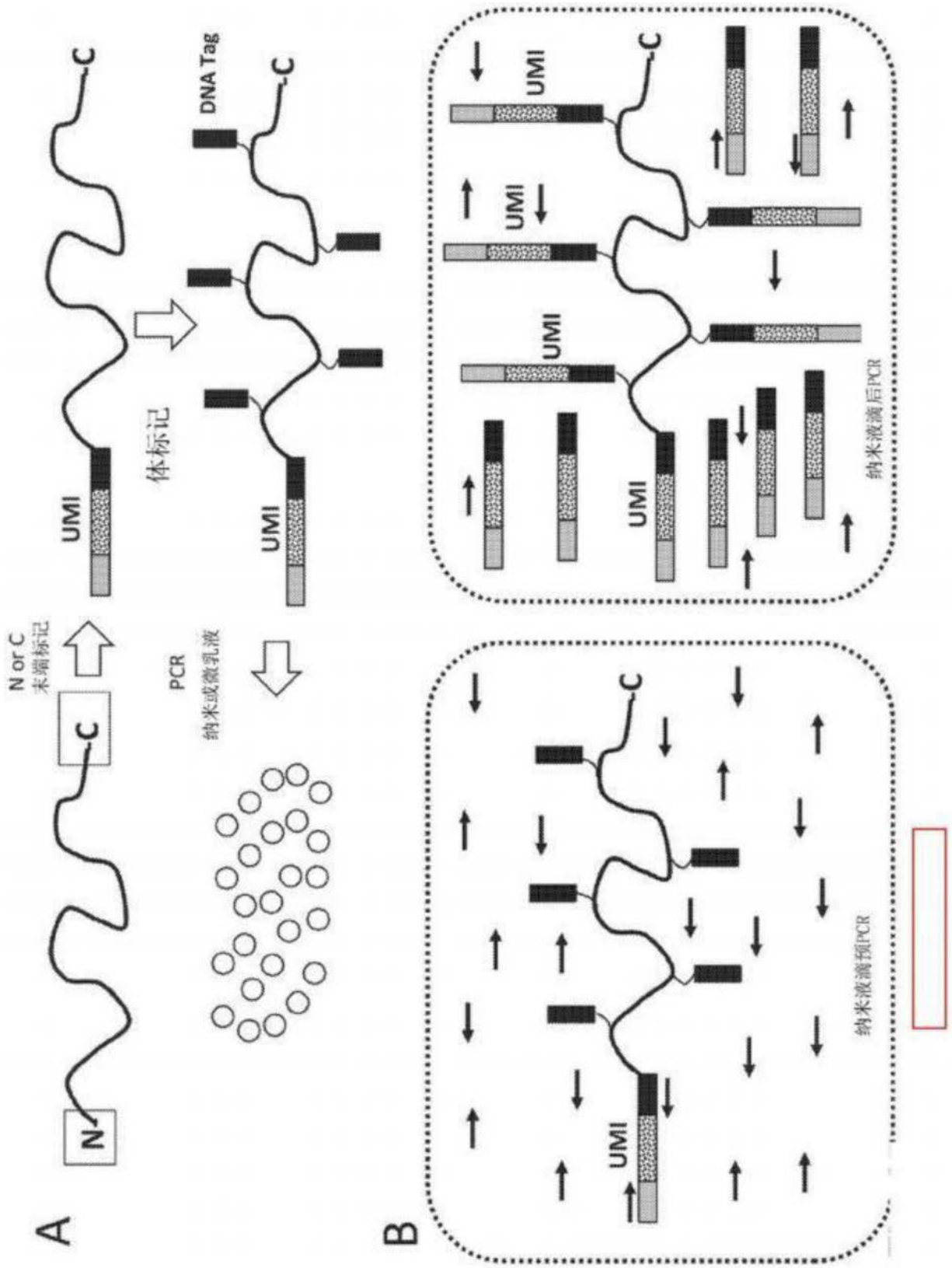


图37

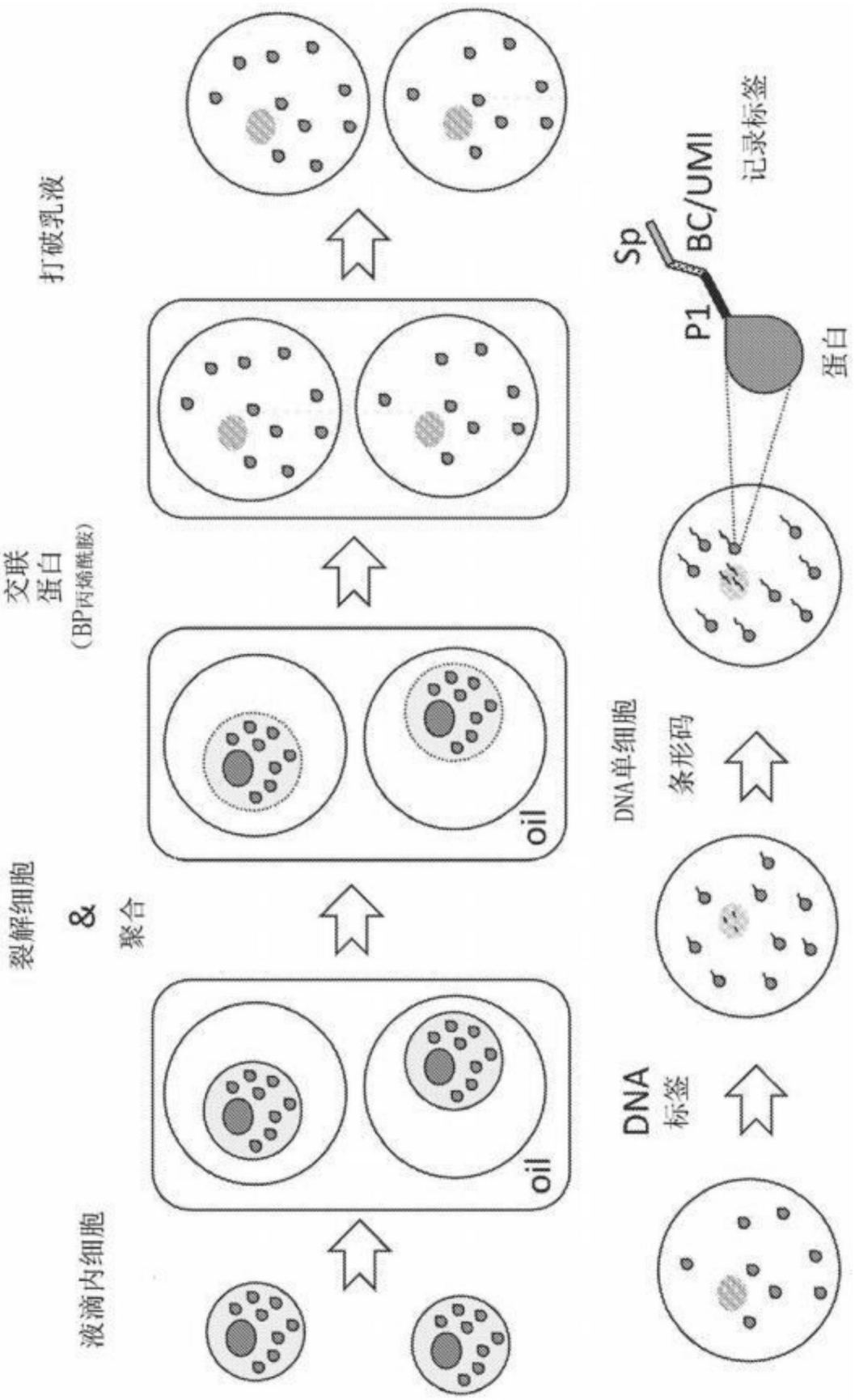

液滴内细胞
裂解细胞
&
聚合
交联
蛋白
(BP丙烯酰胺)
打破乳液
oil
oil
DNA
标签
DNA单细胞
条形码
Sp
P1
BC/UMI
记录标签
蛋白

图38

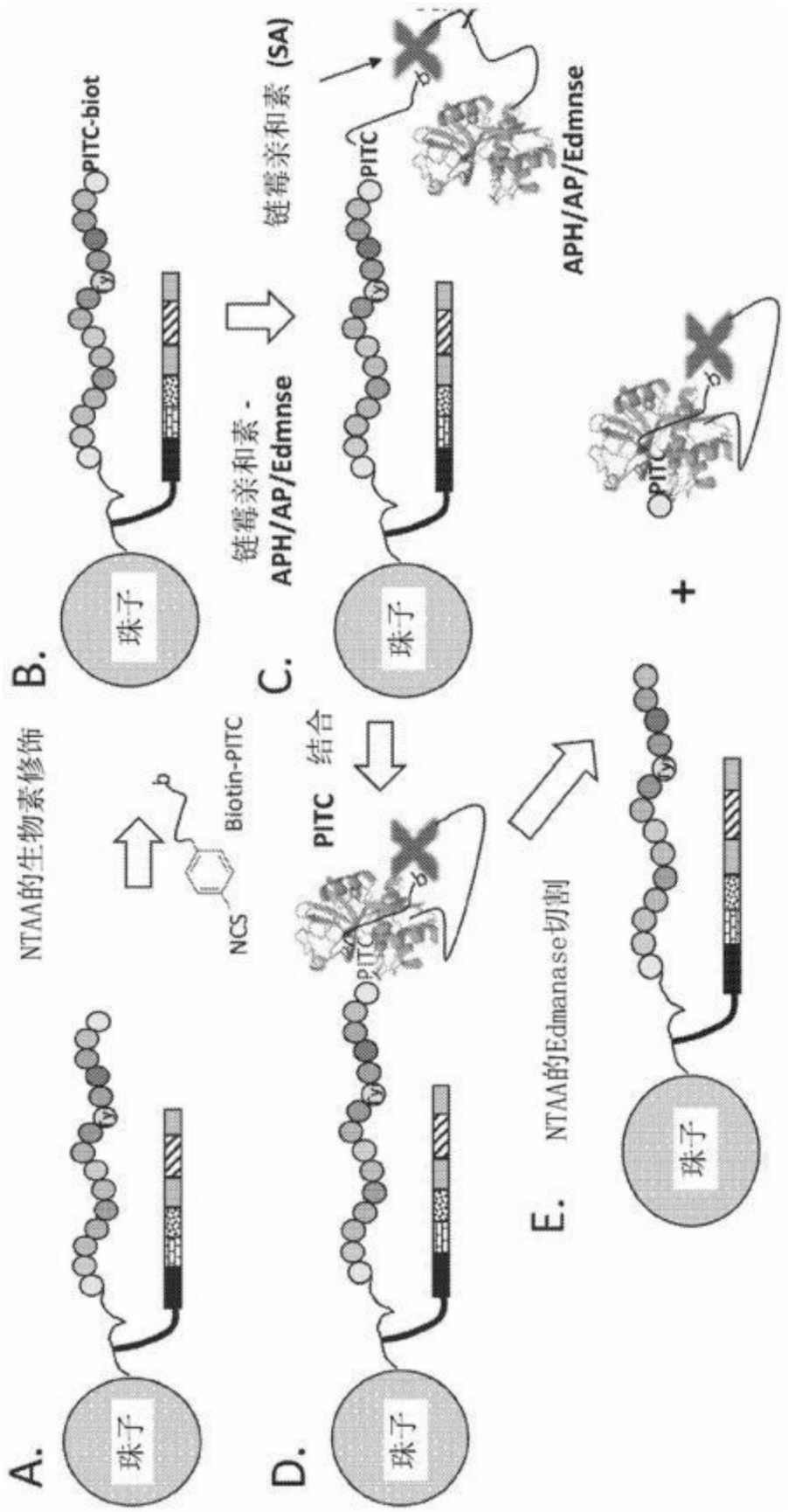

A.
珠子
NTAA的生物素修饰
Biotin-PITC
NCS
B.
PITC-biot
珠子
C.
链霉亲和素 - APH/AP/Edmnse
链霉亲和素 (SA)
APH/AP/Edmnse
PITC
珠子
D.
PITC 结合
PITC
珠子
E.
NTAA的Edmanase切割
PITC
珠子
+

图39

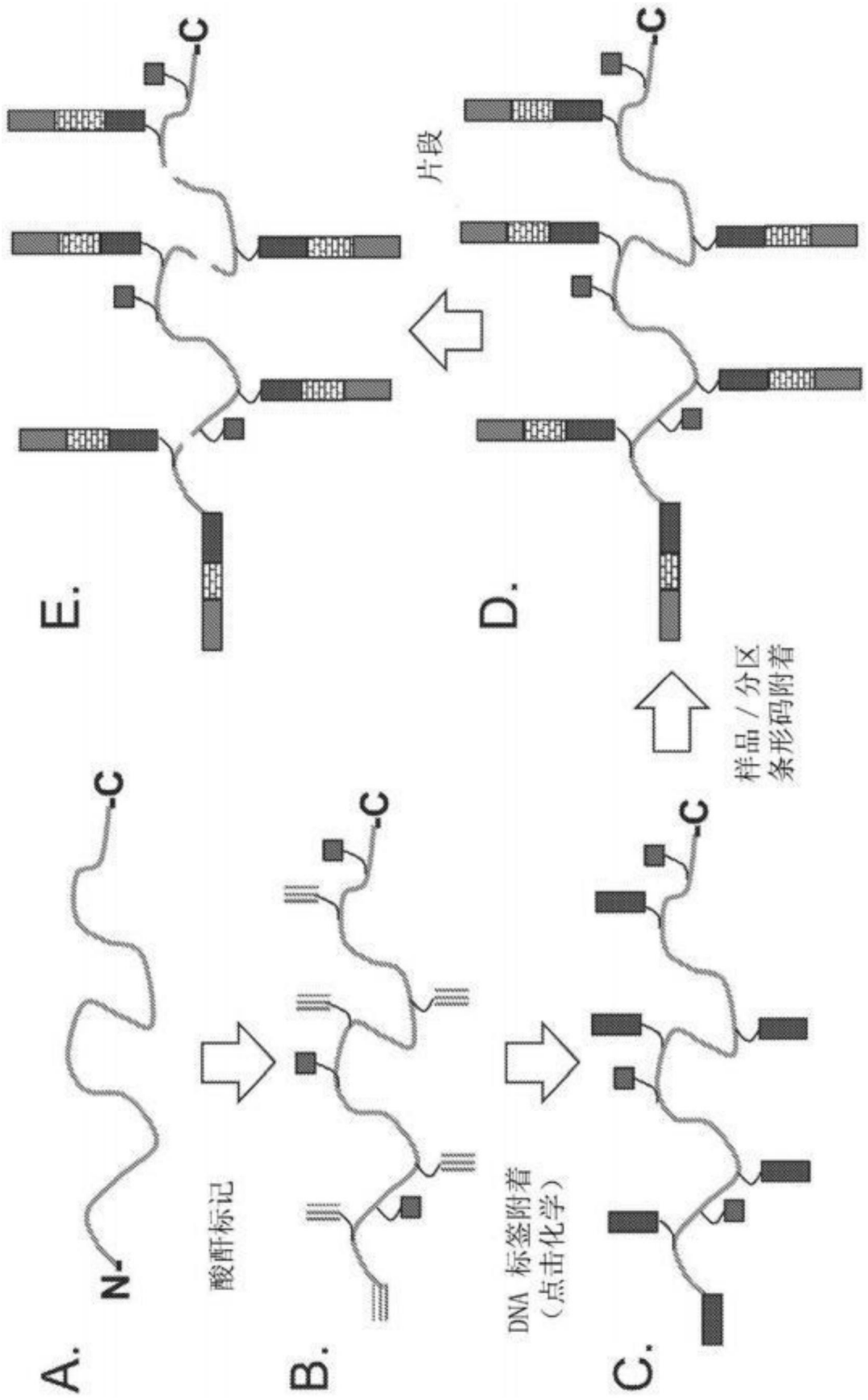


图40A-E

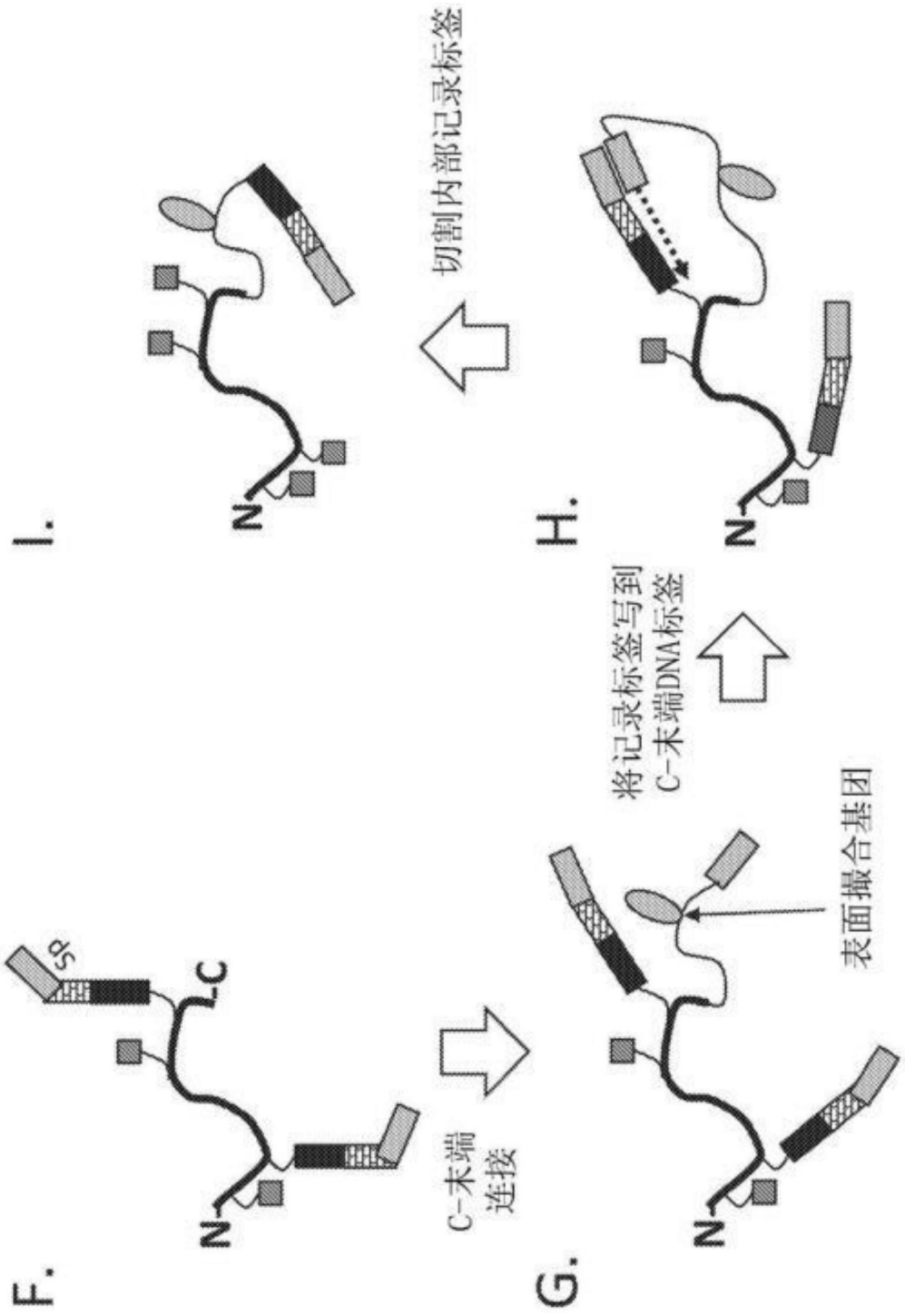


图40F-H

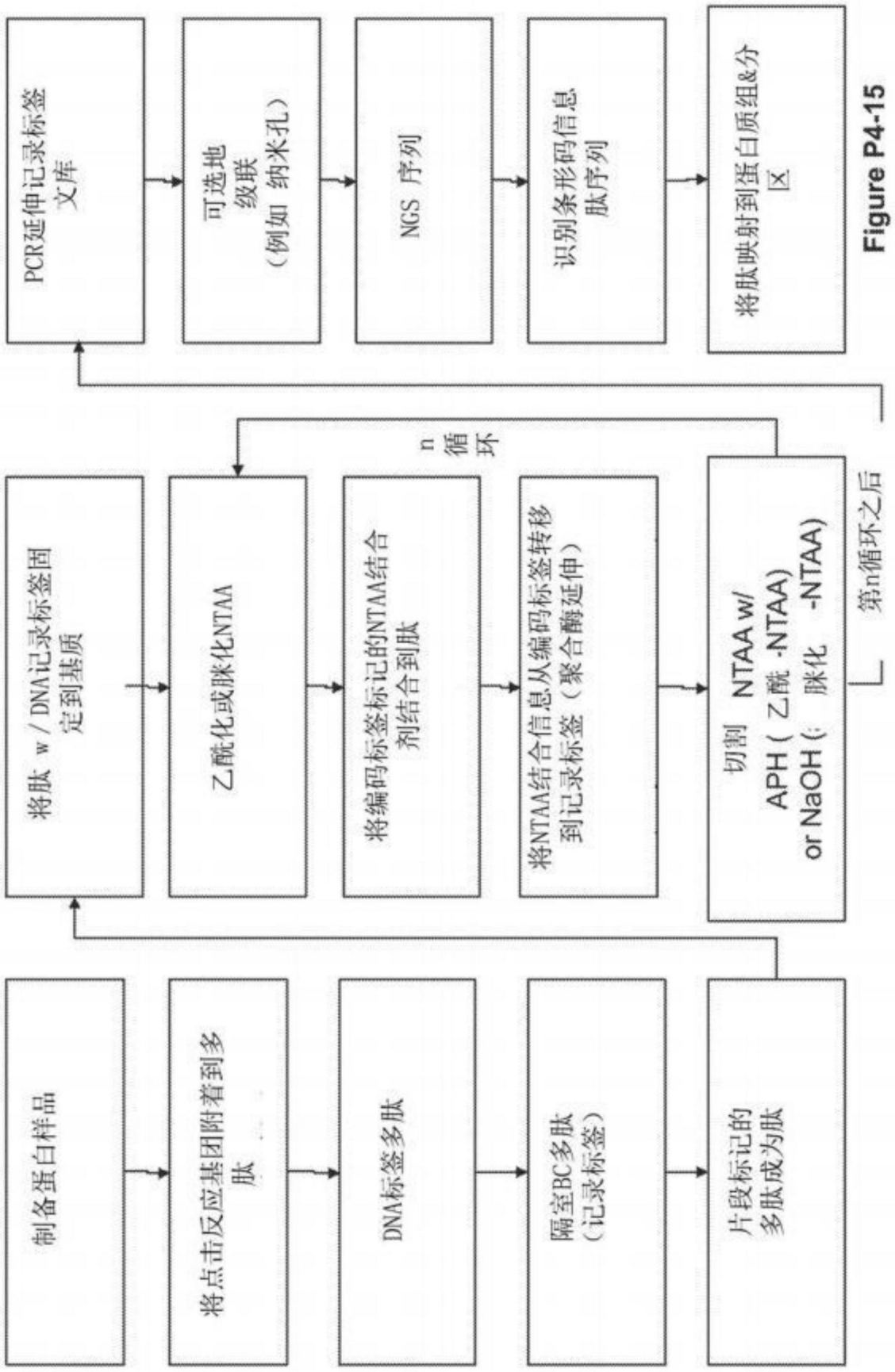

制备蛋白样品
将点击反应基团附着到多肽
DNA标签多肽
隔室BC多肽（记录标签）
片段标记的多肽成为肽
将肽 w / DNA记录标签固定到基质
乙酰化或脒化NTAA
将编码标签标记的NTAA结合剂结合到肽
将NTAA结合信息从编码标签转移到记录标签（聚合酶延伸）
切割 NTAA w/ APH（ 乙酰 -NTAA) or NaOH（ 脒化 -NTAA)
n 循环
第n循环之后
PCR延伸记录标签文库
可选地级联（例如 纳米孔）
NGS 序列
识别条形码信息肽序列
将肽映射到蛋白质组&分区
Figure P4-15

图41

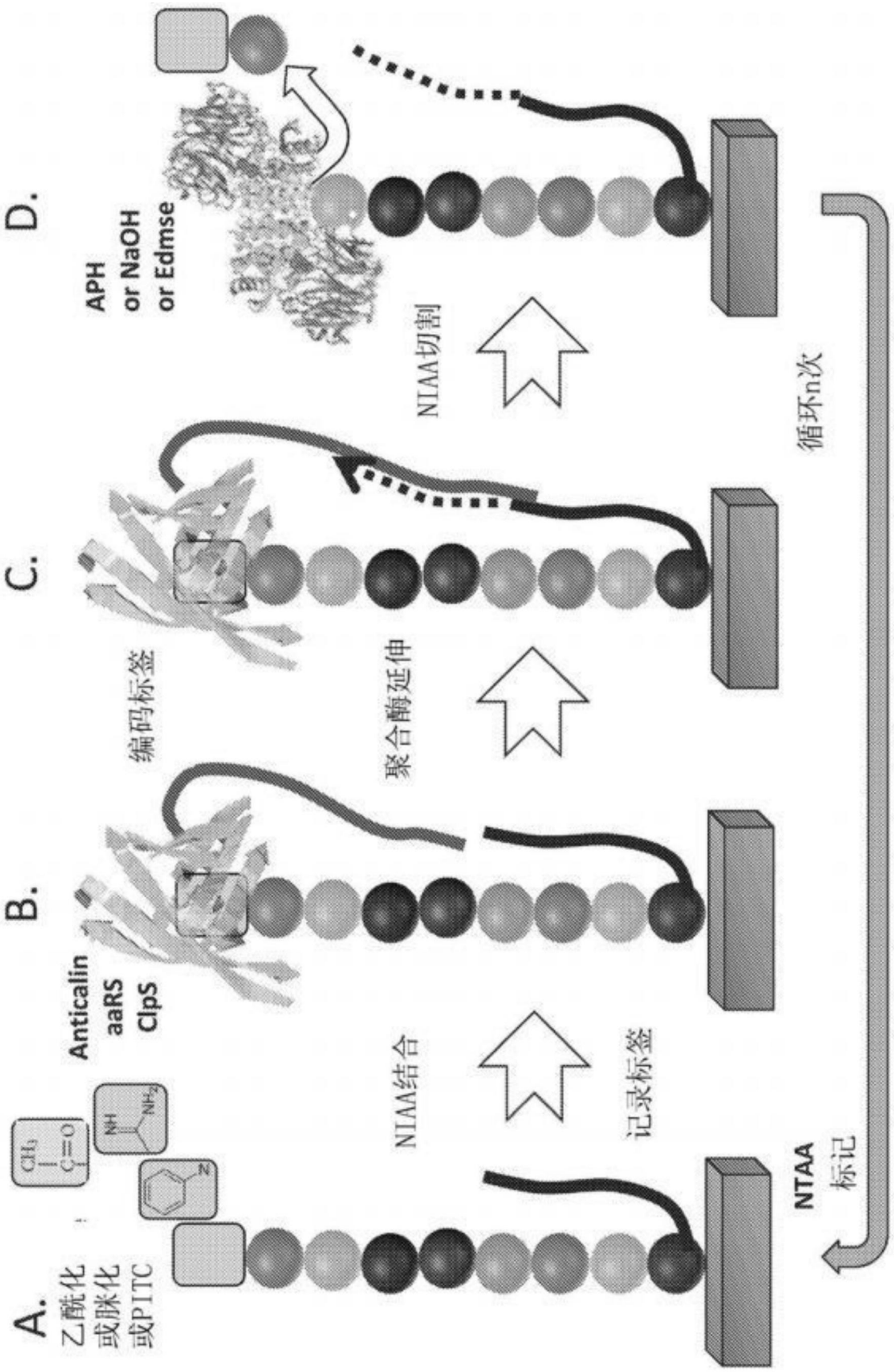

A.
乙酰化
或脒化
或PITC
B.
Anticalin
aaRS
ClpS
NIAA结合
记录标签
C.
编码标签
聚合酶延伸
D.
APH
or NaOH
or Edmse
NIAA切割
循环n次
NTAA
标记

图42

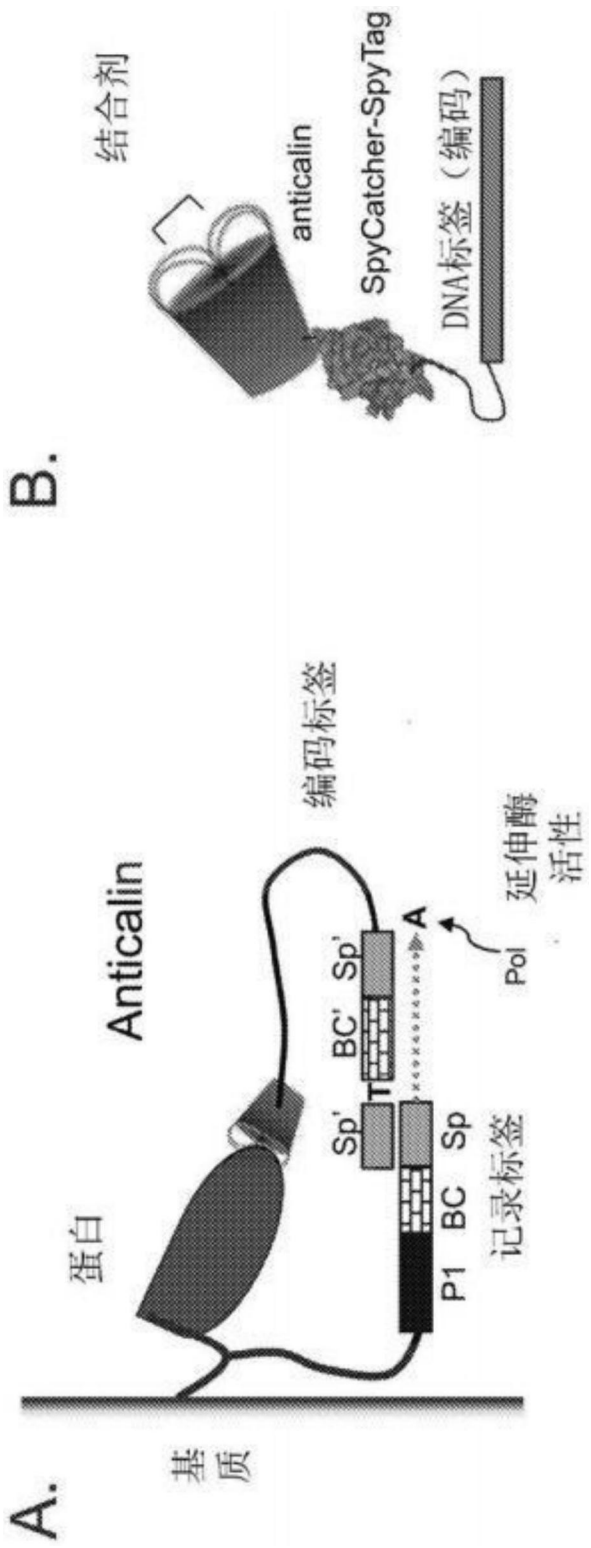

B.
结合剂
anticalin
SpyCatcher-SpyTag
DNA标签（编码）
A.
基质
蛋白
Anticalin
编码标签
Sp'
BC'
Sp'
T
A
Pol
延伸酶活性
P1
BC
Sp
记录标签

图43

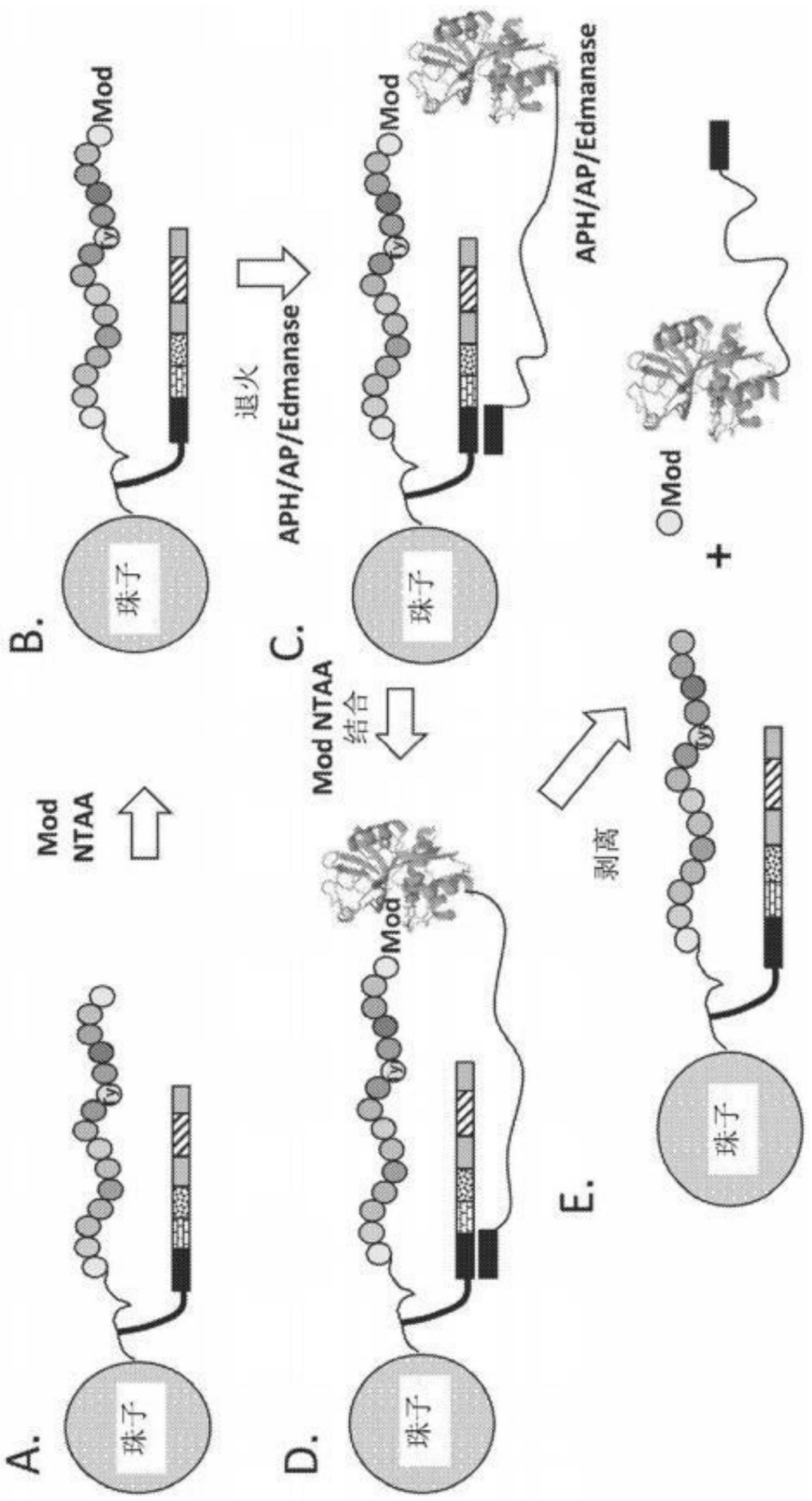

A.
珠子
Mod
NTAA
B.
珠子
Mod
退火
APH/AP/Edmanase
C.
珠子
Mod
APH/AP/Edmanase
Mod NTAA
结合
D.
珠子
Mod
剥离
E.
珠子
Mod
+

图44

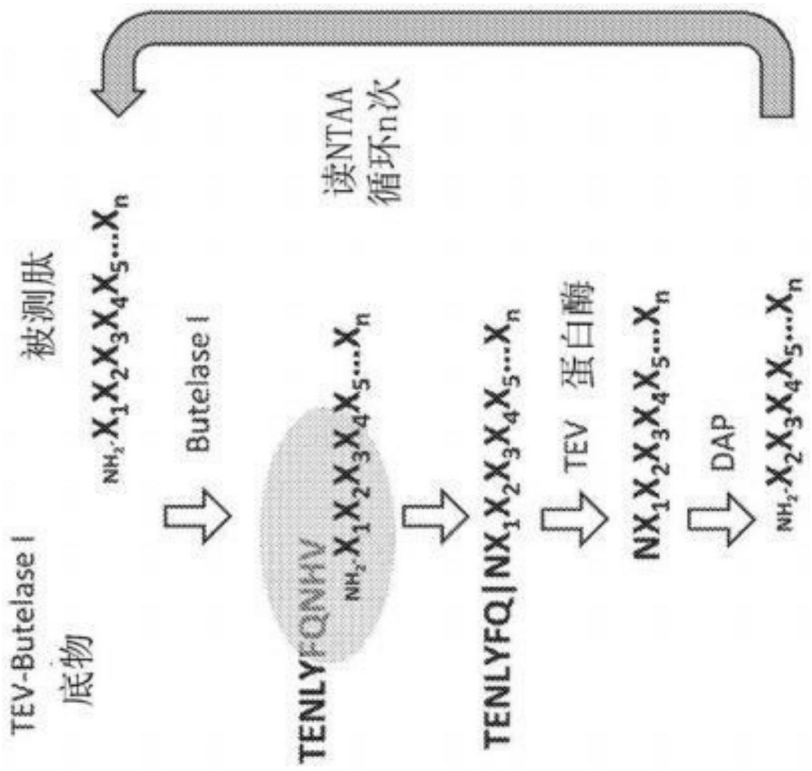

TEV-Butelase 1
底物
被测肽
NH_2-$X_1X_2X_3X_4X_5...X_n$
Butelase 1
TENLYFQNHV
NH_2-$X_1X_2X_3X_4X_5...X_n$
TENLYFQ|$NX_1X_2X_3X_4X_5...X_n$
TEV 蛋白酶
$NX_1X_2X_3X_4X_5...X_n$
DAP
NH_2-$X_2X_3X_4X_5...X_n$
读NTAA
循环n次

图45