(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119716077 A
(43)申请公布日 2025.03.28

(21)申请号 202411678798.5

(22)申请日 2018.10.31

(30)优先权数据
62/579,844 2017.10.31 US
62/582,312 2017.11.06 US
62/583,448 2017.11.08 US

(62)分案原申请数据
201880084553.6 2018.10.31

(71)申请人 ENCODIA公司
地址 美国加利福尼亚州

(72)发明人 马克·S·朱 约翰·M·贝尔
村中宣仁 凯文·冈德森
迈克尔·飞利浦·威纳 石磊
罗伯特·C·詹姆斯
卢卡·蒙弗雷戈拉

(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 1/6804*(2018.01)
*C12Q 1/6869*(2018.01)
*C40B 40/10*(2006.01)
*C40B 70/00*(2006.01)
*C40B 30/04*(2006.01)

权利要求书2页 说明书257页
序列表（电子公布） 附图95页

(54)发明名称

# 采用核酸编码和/或标签进行分析的试剂盒

(57)摘要

公开了试剂盒以及采用核酸编码来分析包括肽、多肽和蛋白质的大分子的方法。样品分析试剂盒采用分子相互作用和/或反应的核酸编码和/或核酸记录，例如识别事件(例如，抗原与抗体之间，修饰的末端氨基酸残基之间，或小分子或肽治疗剂与靶标之间等)。在试剂盒中可以包括另外的条形码化试剂，例如用于循环特异性条形码化的试剂(例如“计时”)、隔室条形码化试剂、组合条形码化试剂、空间条形码化试剂或其任何组合。样品可以包含大分子，其包括肽、多肽和蛋白质，并且该记录可以产生分子相互作用和/或反应信息，和/或多肽序列信息。试剂盒可用于高通量、多路复用和/或自动化分析，并适用于蛋白质组或其子集的分析。

CN 119716077 A

1.一种用于分析样品中多种多肽的方法,包含:

(a)将多种多肽中的每种多肽连接到核酸标签上以生成核酸-多肽嵌合体;

(b)将所述核酸-多肽嵌合体中的核酸标签与连接到固体支持物上的捕获核酸杂交;

(c)将所述核酸-多肽嵌合体共价固定到所述固体支持物上;以及

(d)将共价连接的核酸-多肽嵌合体与多种结合剂接触,其中所述多种结合剂中的每种结合剂与所述多种多肽中的一种或多种多肽的组分结合,并且所述多种结合剂中的每种结合剂包含:(i)核酸编码标签,所述核酸编码标签具有关于所述结合剂的识别信息、或(ii)可检测标签;从而分析所述共价连接的核酸-多肽嵌合体中的一种或多种多肽。

2.根据权利要求1所述的方法,其中所述多种多肽中的每种多肽通过N-末端氨基酸(NTAA)残基或C-末端氨基酸(NTAA)残基连接到所述核酸标签。

3.根据权利要求1所述的方法,其中所述多种多肽中的每种多肽通过赖氨酸残基连接到所述核酸标签。

4.根据权利要求1所述的方法,其中所述多种结合剂包含三种不同的结合剂,每种结合剂对所述多种多肽中的一种或多种多肽的不同组分具有选择性。

5.根据权利要求4所述的方法,其中至少三种不同的结合剂中的每种结合剂被配置为与所述多种多肽中的一种或多种多肽的NTAA残基或CTAA残基结合。

6.根据权利要求1所述的方法,其中所述多种结合剂中的每种结合剂包含可检测标签,所述可检测标签包含荧光基团、颜色编码的纳米颗粒、量子点或其任何组合。

7.根据权利要求1所述的方法,其中所述多种结合剂中的每种结合剂包含核酸编码标签,所述核酸编码标签具有关于所述结合剂的识别信息。

8.根据权利要求1所述的方法,其中所述多种多肽中的每种多肽与所述核酸标签的内部核苷酸化学缀合。

9.根据权利要求1所述的方法,其中所述捕获核酸包含发夹区域。

10.根据权利要求1所述的方法,其中所述多种多肽中的至少一些多肽各自连接到多个核酸标签,并且每种多肽被片段化或消化成较小的多肽,每个较小的多肽平均连接到一个核酸标签。

11.根据权利要求1所述的方法,其中所述多种多肽是蛋白质组或其子集。

12.根据权利要求1所述的方法,其中所述核酸标签包含条形码或独特分子标识符(UMI),其中所述条形码包含分析物条形码、隔室条形码、分区条形码、样品条形码、分级物条形码或其任何组合。

13.根据权利要求1所述的方法,其中所述固体支持物是珠粒或流通槽。

14.根据权利要求1所述的方法,其中所述核酸-多肽嵌合体随机地连接到所述固体支持物。

15.根据权利要求1所述的方法,其中所述多种多肽包含至少1,000,000种多肽。

16.一种用于分析来自不同样品的一种或多种多肽的试剂盒,包含:

(a)一种或多种核酸记录标签,其被配置为与所述一种或多种多肽直接或间接相关、或与通过裂解所述一种或多种多肽获得的一种或多种组分直接或间接相关,其中所述一种或多种核酸记录标签中的每种核酸记录标签包含独特分子标识符(UMI),并且还包含样品条形码,所述样品条形码被配置用于区分来自不同样品的多肽;以及

(b)一种或多种结合剂,每种结合剂包含:

(i)能够与所述一种或多种多肽或其一种或多种组分结合的结合部分;以及

(ii)连接所述结合部分的核酸编码标签,所述核酸编码标签包含条形码序列,所述条形码序列包含关于所述结合部分的识别信息;

其中所述一种或多种核酸记录标签和所述一种或多种核酸编码标签被配置为,允许在所述每种结合剂与所述一种或多种多肽或其组分结合时,通过连接或引物延伸转移所述条形码序列和/或所述UMI的序列信息,从而生成一种或多种延伸核酸构建体,每种核酸构建体包含条形码序列或其互补体、样品条形码或其互补体、以及UMI或其互补体。

17.一种用于分析多肽的试剂盒,包含:

(a)功能化试剂,其能够修饰所述多肽的N-末端氨基酸(NTAA)以生成所述多肽的功能化NTAA;

(b)一种或多种核酸记录标签,其被配置为与所述多肽直接或间接相关、或与通过裂解所述多肽的功能化NTAA获得的所述多肽的组分直接或间接相关,其中所述一种或多种核酸记录标签中的每种核酸记录标签包含独特分子标识符(UMI);以及

(c)一种或多种结合剂,每种结合剂包含:

(i)能够与所述多肽的功能化NTAA或与通过裂解所述多肽的功能化NTAA获得的所述多肽的组分结合的结合部分;以及

(ii)连接所述结合部分的核酸编码标签,所述核酸编码标签包含条形码序列,所述条形码序列包含关于所述结合部分的识别信息;

其中所述一种或多种核酸记录标签和所述一种或多种核酸编码标签被配置为,允许在所述每种结合剂与所述多肽的功能化NTAA或与通过裂解所述多肽的功能化NTAA获得的所述多肽的组分结合时,通过连接或引物延伸转移所述条形码序列和/或所述UMI的序列信息,从而生成一种或多种延伸核酸构建体,每种核酸构建体包含条形码序列或其互补体、以及UMI或其互补体。

18.根据权利要求17所述的试剂盒,其中所述试剂盒还包含消除试剂,用于移除所述多肽的功能化NTAA。

# 采用核酸编码和/或标签进行分析的试剂盒

[0001] 本案是申请日为2018年10月31日,发明名称为采用核酸编码和/或标签进行分析的试剂盒,申请号为201880084553.6的案件的分案申请。

[0002] 相关申请的交叉引用

[0003] 本申请要求于2017年10月31日提交的标题为“用于使用核酸编码和/或标签进行分析的试剂盒”的美国临时申请序列第62/579,844号和于2017年11月6日提交的标题为“用于使用核酸编码和/或标签进行分析的试剂盒”的美国临时申请序列第62/582,312号,以及2017年11月8日提交的标题为“用于使用核酸编码和/或标签进行分析的试剂盒”的美国临时申请序列第62/583,448号的优先权,这些申请各自的全部内容出于所有目的通过引用并入本文。本申请涉及2016年5月2日提交的标题为“使用核酸编码的大分子分析”的美国临时专利申请第62/330,841号;2016年5月19日提交的标题为“使用核酸编码的大分子分析”的美国临时专利申请第62/339,071号;2016年8月18日提交的标题为“使用核酸编码的大分子分析”的美国临时专利申请第62/376,886号;于2017年5月2日提交的标题为“使用核酸编码的大分子分析”的国际专利申请第PCT/US2017/030702号;2017年10月31日提交的具有代理人案件No.776533000500的美国临时专利申请第62/579,844号,标题为“使用核酸编码和/或标签进行分析的试剂盒”;2017年10月31日提交的具有代理人案号No.776533000600美国临时专利申请第62/579,870号,标题为“用于多肽分析的方法和组合物”;2017年10月31日提交的具有代理人案件No.776533000700的美国临时专利申请第62/579,840号,标题为“使用核酸编码和/或标签的方法和试剂盒”;2017年11月6日提交的具有代理人案件No.776533000501的美国临时专利申请第62/582,312号,标题为“使用核酸编码和/或标签进行分析的试剂盒”;以及于2017年11月7日提交的具有代理人案件No.776533000701美国临时专利申请No.62/582,916号,标题为“使用核酸编码和/或标签的方法和试剂盒”,所述申请的公开内容出于所有目的通过引用并入本文。

[0004] 在ASCII文本文件上提交序列表

[0005] 以下关于ASCII文本文件的提交内容通过引用整体并入本文:序列表(文件名:4614-2000540_20181031_SeqList.txt,记录日期:2018年10月31日,大小:52K字节)的计算机可读形式(CRF)。

## 技术领域

[0006] 本公开总体上涉及使用分子相互作用和/或反应(例如,识别事件)的核酸编码和/或核酸记录的样品分析试剂盒。在一些实施例中,试剂盒可用于高通量、多路复用和/或自动化分析,并适用于蛋白质组或其子集的分析。

## 背景技术

[0007] 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。

[0008] 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。本公开解决了这些和其他需要。

## 发明内容

[0009] 本发明内容不旨在用于限制所要求保护的主题的范围。根据包括附图和所附权利要求中公开的那些方面的详细描述,所要求保护的主题的其它特征、细节、用途和优点将变得容易理解。

[0010] 在本发明内容的任何实施例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话)或其任何部分(例如,通用引物、间隔件、UMI、记录标签条形码、编码序列、结合循环特异性条形码,等等)可以包括或替换为可测序聚合物,例如非核酸可测序聚合物。

[0011] 在一个方面,本文公开了试剂盒,其包含:(a)记录标签,其被配置为与分析物直接或间接相关;(b)(i)编码标签,其包含关于能够结合所述分析物的结合基团的识别信息,并且被配置为与所述结合基团直接或间接相关以形成结合剂,和/或(ii)标签,其中所述记录标签和所述编码标签被配置为允许在所述结合剂和所述分析物之间结合时在它们之间转移信息;和任选地,(c)结合基团。在一个实施例中,记录标签和/或分析物被配置为直接或间接固定到支持物上。在另一个实施例中,记录标签被配置为固定到支持物上,从而固定与记录标签相关的分析物。在另一个实施例中,分析物被配置为固定到支持物上,从而固定与分析物相关的记录标签。在又一个实施例中,记录标签和分析物中的每一个被配置为固定到支持物上。在另一个实施例中,记录标签和分析物被设计成当两者都固定到支持物上时共同定位。在一些实施例中,用于在记录标签与结合至分析物的结合剂的编码标签之间进行信息转移的(i)分析物与(ii)记录标签之间的距离小于约$10^{-6}$nm,约$10^{-6}$n,约$10^{-5}$nm,约$10^{-4}$nm,约0.001nm,约0.01nm,约0.1nm,约0.5nm,约1nm,约2nm,约5nm,或大于约5nm,或在上述范围内的任何值。

[0012] 在任何前述实施例中,试剂盒可进一步包含固定接头,其被配置为:(i)直接或间接固定到支持物上,和(ii)与记录标签和/或分析物直接或间接相关。在一个实施例中,固定接头被配置为与记录标签和分析物相关。

[0013] 在任何前述实施例中,固定连接物可以配置为直接固定到支持物上,从而固定与固定接头相关的记录标签和/或分析物。

[0014] 在任何前述实施例中,试剂盒可以进一步包含支持物。

[0015] 在任何前述实施例中,试剂盒可以进一步包含一种或多种试剂,用于在结合剂和分析物之间结合时在编码标签和记录标签之间转移信息。在一个实施例中,一种或多种试剂被配置为将信息从编码标签转移到记录标签,从而产生延伸记录标签。在另一个实施例中,一种或多种试剂被配置为将信息从记录标签转移到编码标签,从而产生延伸编码标签。

在又一个实施例中,一种或多种试剂被配置为产生包含来自编码标签的信息和来自记录标签的信息的双标签构建体。

[0016] 在任何前述实施例中,试剂盒可以包含至少两个记录标签。在任何前述实施例中,试剂盒可以包含至少两个编码标签,每个编码标签包含关于其相关结合基团的识别信息。在具体的实施例中,每一分析物具有可用于结合到分析物的结合剂的多个记录标签(例如,至少约2个,约5个,约10个,约20个,约50个,约100个,约200个,约500个,约1000个,约2000个,约5000个或更多)。在具体的实施例中,试剂盒包含多个记录标签,例如,至少约2个,约5个,约10个,约20个,约50个,约100个,约200个,约500个,约1000个,约2000个,约5000个,或更多。

[0017] 在任何前述实施例中,试剂盒可以包含结合剂中至少两种。在一个实施例中,试剂盒包含:(i)一种或多种试剂,其在第一结合剂和分析物之间结合时用于将信息从第一结合剂的第一编码标签转移到记录标签,以产生第一次序延伸记录标签,和/或(ii)一种或多种试剂,其在第二结合剂和分析物之间结合时用于将信息从第二结合剂的编码标签转移至第一延伸记录标签,以产生第二次序延伸记录标签,其中(i)的一种或多种试剂和(ii)的一种或多种试剂可以相同或不同。在具体的实施例中,每个分析物具有多个结合剂和/或编码标签,例如至少约2个,约5个,约10个,约20个,约50个,约100个,约200个,约500个,约1000个,约2000个,约5000个或更多个可用于分析物的记录标签,多个结合剂可以依次或并行加入。在具体的实施例中,试剂盒包含多个结合剂和/或编码标签,例如,至少约2个,约5个,约10个,约20个,约50个,约100个,约200个,约500个,约1000个,约2000个,约5000个或更多。

[0018] 在任何前述实施例中,试剂盒可以进一步包含(iii)一种或多种试剂,其用于在第三(或更高次序)结合剂与分析物之间结合时,将信息从第三(或更高次序)结合剂的第三(或更高次序)编码标签转移至第二次序延伸记录标签,以产生第三(或更高次序)次序延伸记录标签。在一个实施例中,试剂盒包含:(i)一种或多种试剂,其在第一结合剂和分析物之间结合时用于将信息从第一结合剂的第一编码标签转移到第一记录标签,以产生第一延伸记录标签,(ii)一种或多种试剂,其在第二结合剂和分析物之间结合时用于将信息从第二结合剂的第二编码标签转移到第二记录标签,以产生第二延伸记录标签,和/或(iii)一种或多种试剂,其在第三(或更高次序)结合剂和分析物之间结合时用于将信息从第三(或更高次序)结合剂的第三(或更高次序)编码标签转移到第三(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,其中(ii)、(ii)和/或(iii)的一种或多种试剂可以相同或不同。

[0019] 在任何前述实施例中,试剂盒可以进一步包含(iii)一种或多种试剂,其在第三(或更高次序)结合剂与分析物之间结合时用于将信息从第三(或更高次序)结合剂的第三(或更高次序)编码标签转移到第三(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。

[0020] 在任何前述实施例中,第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为与分析物直接或间接相关。

[0021] 在任何前述实施例中,第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为固定在支持物上。

[0022] 在任何前述实施例中,第一记录标签、第二记录标签和/或第三(或更高次序)记录

标签可以被配置为与分析物共定位,例如,以在第一、第二或第三(或更高次序)结合剂与分析物之间结合时允许分别在第一、第二或第三(或更高次序)编码标签与第一、第二或第三(或更高次序)记录标签之间转移信息。

[0023] 在任何前述实施例中,第一编码标签、第二编码标签和/或第三(或更高次序)编码标签中的每一个可以包含结合循环特异性条形码,例如结合循环特异性间隔区序列$C_n$和/或编码标签特异性间隔区序列$C_n$,其中n是整数,并且$C_n$表示第n结合剂和多肽之间的结合。或者,可以外源添加结合循环标签$C_n$,例如,结合循环标签$C_n$对于编码标签可以是外源的。

[0024] 在任何前述实施例中,分析物可以包含多肽。在一个实施例中,试剂盒的结合基团能够结合至多肽的一个或多个N-末端、内部或C-末端氨基酸,或能够结合至通过官能化试剂修饰的一个或多个N-末端、内部或C-末端氨基酸。

[0025] 在任何前述实施例中,试剂盒可以进一步包含一种或多种官能化试剂。

[0026] 在任何前述实施例中,试剂盒可以进一步包含用于移除(例如,通过化学裂解或酶裂解)多肽的一个或多个N-末端、内部或C-末端氨基酸,或移除官能化的N-末端、内部或C-末端氨基酸的消除试剂,任选地,其中消除试剂包含羧肽酶或氨肽酶或其变体,突变体或修饰的蛋白质;水解酶或其变体,突变体或其修饰的蛋白质;温和型埃德曼降解试剂;Edmanase酶;无水TFA,碱;或其任何组合。

[0027] 在任何前述实施例中,一个或多个N-末端、内部或C-末端氨基酸可包含:(i)N-末端氨基酸(NTAA);(ii)N-末端二肽序列;(iii)N-末端三肽序列;(iv)内部氨基酸;(v)内部二肽序列;(vi)内部三肽序列;(vii)C-末端氨基酸;(viii)C-末端二肽序列;或(ix)C-末端三肽序列,或其任何组合,任选地其中(i)-(ix)中的任何一个或多个氨基酸残基被修饰或官能化。

[0028] 在另一方面,本文公开了试剂盒,其包含:至少(a)第一结合剂,其包含(i)能够结合待分析多肽的N-末端氨基酸(NTAA)或官能化NTAA的第一结合基团,和(ii)包含关于第一结合基团的识别信息的第一编码标签,任选地(b)记录标签,其被配置为与多肽直接或间接相关,并且进一步任选地(c)官能化试剂,其能够修饰多肽的第一NTAA,以产生第一官能化NTAA,其中记录标签和第一结合剂被配置为允许在第一结合剂和多肽之间结合时在第一编码标签和记录标签之间转移信息。在一个实施例中,试剂盒还包含一种或多种试剂,用于将信息从第一编码标签转移到记录标签,从而产生第一次序延伸记录标签。

[0029] 在任何前述实施例中,官能化试剂可以包含化学剂、酶和/或生物剂,如异硫氰酸酯衍生物,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,或硫苄基化试剂。

[0030] 在前述实施例中的任一个中,试剂盒可进一步包含消除试剂,用于移除(例如,通过化学裂解或酶裂解)第一官能化NTAA,以暴露紧邻的氨基酸残基作为第二NTAA。在一个实施例中,第二NTAA能够通过相同或不同的官能化试剂官能化以产生第二官能化NTAA,其可与第一官能化NTAA相同或不同。在另一个实施例中,试剂盒还包含:(d)第二(或更高次序)结合剂,其包含(i)能够结合到第二官能化NTAA的第二(或更高次序)结合基团,和(ii)第二(或更高次序)编码标签,其包含关于第二(或更高次序)结合基团的识别信息,其中第一编码标签和第二(或更高次序)编码标签可以相同或不同。在又一实施例中,第一官能化NTAA和第二官能化NTAA彼此独立地选自由以下组成的群组:官能化N-末端丙氨酸(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),酪氨酸(Y或Tyr),其任何组合。

[0031] 在任何前述实施例中,试剂盒可以进一步包含一种或多种试剂,用于将信息从第二(或更高次序)编码标签转移到第一次序延伸记录标签,从而产生第二(或更高次序)延伸记录标签。

[0032] 在另一方面,本文公开了试剂盒,其包含:至少(a)一种或多种结合剂,每种结合剂包含(i)能够结合待分析多肽的N-末端氨基酸(NTAA)或官能化NTAA的结合基团,和(ii)包含关于结合基团的识别信息的编码标签,和/或(b)一个或多个记录标签,其被配置为与多肽直接或间接相关,其中一个或多个记录标签和一个或多个结合剂被配置为允许在每个结合剂与多肽之间结合时在编码标签和记录标签之间转移信息,以及任选地(c)官能化试剂,其能够修饰多肽的第一NTAA以产生第一官能化NTAA。在一个实施例中,试剂盒进一步包含消除试剂,用于移除(例如,通过化学裂解或酶裂解)第一官能化NTAA,以暴露紧邻的氨基酸残基(如第二NTAA)。在另一实施例中,第二NTAA能够通过相同或不同的官能化试剂官能化以产生第二官能化NTAA,其可与第一官能化NTAA相同或不同。在又一实施例中,第一官能化NTAA和第二官能化NTAA彼此独立地选自由以下组成的群组:官能化N-末端丙氨酸(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),缬氨酸(V或Val),色氨酸(W或Trp),酪氨酸(Y或Tyr),其任何组合。

[0033] 在任何前述实施例中,试剂盒可以包含:(i)一种或多种试剂,其用于在第一结合剂和多肽之间结合时,将信息从第一结合剂的第一编码标签转移到第一记录标签,以产生第一延伸记录标签,和/或(ii)一种或多种试剂,其用于在第二结合剂和多肽之间结合时,将信息从第二结合剂的第二编码标签转移到第二记录标签结合,以产生第二延伸记录标签,其中(i)的一种或多种试剂和(ii)的一种或多种试剂可以相同或不同。在一个方面,试剂盒还包含:(iii)一种或多种试剂,其用于在第三(或更高次序)结合剂和多肽之间结合时,将信息从第三(或更高次序)结合剂的第三(或更高次序)编码标签转移到第三(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。

[0034] 在任何前述实施例中,第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为与多肽直接或间接相关。

[0035] 在任何前述实施例中,第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为固定在支持物上。

[0036] 在任何前述实施例中,第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为与多肽共定位,例如,以在第一、第二或第三(或更高次序)结合剂和多肽之间结合时允许分别在第一、第二或第三(或更高次序)编码标签与第一、第二或第三(或更高次序)记录标签之间转移信息。

[0037] 在任何前述实施例中,支持物上的第一记录标签、第二记录标签和/或第三(或更

高次序)记录标签之间的距离可以等于或大于约10nm,等于或大于约15nm,等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,或等于或大于约500nm,同时每个记录标签及其相应的分析物被配置为当两者被固定到支持物上时共定位,或每个记录标签及其其相应的分析物之间的距离小于约$10^{-6}$nm,约$10^{-6}$nm,约$10^{-5}$nm,约$10^{-4}$nm,约0.001nm,约0.01nm,约0.1nm,约0.5nm,约1nm,约2nm,约5nm,或大于约5nm,或在上述范围之间的任何值。

[0038] 在任何前述实施例中,第一编码标签、第二编码标签和/或第三(或更高次序)编码标签中的每一个可以包含结合循环特异性条形码,例如结合循环特异性间隔区序列$C_n$和/或编码标签特异性间隔区序列$C_n$,其中n是整数,并且$C_n$表示第n结合剂和多肽之间的结合。或者,可以外源添加结合循环标签$C_n$,例如,结合循环标签$C_n$对于编码标签可以是外源的。

[0039] 在任何前述实施例中,分析物或多肽可包含蛋白质或多肽链或其片段,脂质,碳水化合物或大环,或其组合或复合物。

[0040] 在任何前述实施例中,分析物或多肽可以包含大分子或其复合物,例如蛋白质复合物或其亚基。

[0041] 在任何前述实施例中,记录标签可以包含核酸、寡核苷酸、修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、具有一个或多个保护碱基的DNA或RNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子,或吗啉代,或其组合。

[0042] 在任何前述实施例中,记录标签可以包括通用引发位点。

[0043] 在任何前述实施例中,记录标签可包含用于扩增、测序或两者的引发位点,例如通用引发位点包含用于扩增、测序或两者的引发位点。

[0044] 在任何前述实施例中,记录标签和/或编码标签可以包含独特的分子标识符(UMI)。

[0045] 在任何前述实施例中,记录标签和/或编码标签可以包含条形码和/或核酸酶位点,例如切刻内切核酸酶位点(例如,dsDNA切刻内切核酸酶位点)。

[0046] 在任何前述实施例中,记录标签和/或编码标签在其3'-末端和/或在其5'-末端包含间隔区,例如,记录标签在其3'-末端包含间隔区。

[0047] 在任何前述实施例中,记录标签和/或编码标签可以包含一个或多个核酸酶位点,例如核酸内切酶位点,归巢核酸内切酶位点,限制酶消化位点,切刻核酸内切酶位点或其组合。在一些实施例中,可以在编码标签中提供核酸酶位点,例如,在间隔区序列内或在间隔区序列和编码器序列之间。在一些实施例中,可以在记录标签中提供核酸酶位点,例如,在通用引物序列和支持物之间(例如,用于在支持物上裂解记录标签)。

[0048] 在任何前述实施例中,试剂盒可以包含固体支持物,例如刚性固体支持物,柔性固体支持物或软固体支持物,并且包括多孔支持物或非多孔支持物。

[0049] 在任何前述实施例中,试剂盒可以包括支持物,支持物包含珠粒、多孔珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

在一个实施例中,支持物包含聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[0050] 在任何前述实施例中,试剂盒可以在序列反应、平行反应,或序列和平行反应的组合中包含支持物和/或可以用于分析多种分析物(例如,多肽)。在一个实施例中,分析物在支持物上以等于或大于约10nm,等于或大于约15nm,等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,或等于或大于约500nm的平均距离间隔。

[0051] 在任何前述实施例中,结合基团可以包含多肽或其片段,蛋白质或多肽链或其片段,或蛋白质复合物或其亚基,例如抗体或其抗原结合片段。

[0052] 在任何前述实施例中,结合基团可以包含羧肽酶或氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰蛋白;抗生物素蛋白或其变体,突变体或修饰蛋白;ClpS或其变体,突变体或修饰的蛋白质;UBR盒蛋白质或其变体,突变体或修饰蛋白;结合氨基酸的经修饰的小分子,即万古霉素或其变体,突变体或经修饰的分子;或其任何组合,或其中在每个结合剂中,结合基团包含小分子,编码标签包含识别小分子的多核苷酸,由此多个结合剂形成编码的小分子库,例如DNA编码的小分子库。

[0053] 在任何前述实施例中,结合基团可以选择性地和/或特异性地结合至分析物或多肽。

[0054] 在任何前述实施例中,编码标签可以包含核酸、寡核苷酸、修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、具有一个或多个保护碱基的DNA或RNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代,或其组合。

[0055] 在任何前述实施例中,编码标签可以包含条形码序列,例如编码区序列,如识别结合基团的编码序列。

[0056] 在任何前述实施例中,编码标签可包含间隔区、结合循环特异性序列、独特分子标识符(UMI),通用引发位点或其任何组合。在一个实施例中,在每个结合循环后将结合循环特异性序列添加到记录标签中。

[0057] 在任何前述实施例中,结合基团和编码标签可以通过接头或结合对连接。

[0058] 在任何前述实施例中,结合基团和编码标签可以通过SpyTag/SpyCatcher,SpyTag-KTag/SpyLigase(其中待连接的两个部分具有SpyTag/KTag对,并且SpyLigase将SpyTag连接至KTag,从而连接这两个部分),SnoopTag/SnoopCatcher肽-蛋白质对,HaloTag/HaloTag配体对,或分选酶,如LPXTG标签/分选酶(例如,Sortase A5,ActiveMotif,San Diego,或如US 9,267,127 B2中所公开的,其通过引用并入本文),或其任何组合连接。

[0059] 在任何前述实施例中,试剂盒可以进一步包含用于在模板或非模板反应中在编码标签和记录标签之间转移信息的试剂,任选地其中试剂是(i)化学连接试剂或生物连接试剂,例如,连接酶,例如用于连接单链核酸或双链核酸的DNA连接酶或RNA连接酶,或(ii)用于单链核酸或双链核酸引物延伸的试剂,任选地,其中试剂盒还包含连接试剂,连接试剂包含至少两种连接酶或其变体(例如,至少两种DNA连接酶,或至少两种RNA连接酶,或至少一种DNA连接酶和至少一种RNA连接酶),其中至少两种连接酶或其变体包含腺苷酸化连接酶

和组成型非腺苷酸化连接酶,或任选地其中试剂盒还包含连接试剂,连接试剂包含DNA或RNA连接酶和DNA/RNA脱腺苷酸化酶。

[0060] 在任何前述实施例中,试剂盒可以进一步包含聚合酶,例如DNA聚合酶或RNA聚合酶或逆转录酶,用于在编码标签和记录标签之间转移信息。

[0061] 在任何前述实施例中,试剂盒可以进一步包含一种或多种用于核酸序列分析的试剂。在一个实施例中,核酸序列分析包括通过合成测序、通过连接测序、通过杂交测序、polony测序、离子半导体测序、焦磷酸测序、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,或其任何组合。

[0062] 在任何前述实施例中,试剂盒可以进一步包含一种或多种用于核酸扩增的试剂,例如用于扩增一个或多个延伸记录标签,任选地其中核酸扩增包含指数扩增反应(例如,聚合酶链反应(PCR),例如乳液PCR以减少或消除模板转换)和/或线性扩增反应(例如,通过体外转录的等温扩增,或等温嵌合引物引发的核酸扩增(ICAN))。参见,例如,Uemori等人(2007),“ICAN的分子机制——一种新的基因扩增方法的研究(Investigation of the molecular mechanism of ICAN,a novel gene amplification method)”,J Biochem142(2):283-292;Mukai等(2007),“采用5’-DNA-RNA-3’嵌合引物、RNaseH和strand置换DNA聚合酶三种元素的高效等温DNA扩增系统(Highly efficient isothermal DNA amplification system using three elements of 5’-DNA-RNA-3’chimeric primers, RNaseH and strand-displacing DNA polymerase)”,J Biochem142(2):273-281;Ma等人(2013),“用于下一代测序的等温扩增法(Isothermal amplification method for next-generation sequencing)”,Proc Natl Acad Sci U S A.110(35):14320-14323。110(35):14320-14323,其全部通过引用并入本文用于所有目的。

[0063] 在任何前述实施例中,试剂盒可以包含用于将编码标签信息转移至记录标签以形成延伸记录标签的一种或多种试剂,其中延伸记录标签上的编码标签信息的次序和/或频率指示结合剂结合至分析物或多肽的次序和/或频率。

[0064] 在任何前述实施例中,试剂盒可以进一步包含一种或多种用于靶标富集的试剂,例如,一种或多种延伸记录标签的富集。

[0065] 在任何前述实施例中,试剂盒可以进一步包含一种或多种用于扣除,例如扣除一个或多个延伸记录标签的试剂。

[0066] 在任何前述实施例中,试剂盒可以进一步包含一种或多种用于归一化的试剂,例如,用以减少高丰度种类,如一种或多种分析物或多肽。

[0067] 在任何前述实施例中,试剂盒的至少一种结合剂可以结合到末端氨基酸残基、末端二氨基酸残基或末端三氨基酸残基。

[0068] 在任何前述实施例中,试剂盒的至少一种结合剂可以结合到翻译后修饰的氨基酸。

[0069] 在任何前述实施例中,试剂盒可以进一步包含用于将样品中的多种分析物或多肽分区到多个隔室中的一种或多种试剂或装置,其中每个隔室包含多个隔室标签,多个隔室标签任选地连接到支持物(例如,固体支持物),其中多个隔室标签在单个隔室内是相同的并且不同于其它隔室的隔室标签。在一个实施例中,试剂盒还包含用于将多种分析物或多肽(例如,多个蛋白质复合物、蛋白质和/或多肽)片段化成多种多肽片段的一种或多种试剂

或工具。

[0070] 在任何前述实施例中,试剂盒可以进一步包含一种或多种用于将多个多肽片段与多个隔室中的每一个内的隔室标签退火或连接的试剂或工具,从而产生多个隔室标记的多肽片段。

[0071] 在任何前述实施例中,多个隔室可以包括微流体液滴、微孔,或表面上的分离区域,或其任何组合。

[0072] 在任何前述实施例中,多个隔室中的每一个平均可包含单个单元。

[0073] 在任何前述实施例中,试剂盒可以进一步包含用于标记样品中的多种分析物或多肽的一个或多个通用DNA标签。

[0074] 在任何前述实施例中,试剂盒可以进一步包含一种或多种试剂,用于用一种或多种通用DNA标签标记样品中的多种分析物或多肽。

[0075] 在任何前述实施例中,试剂盒可进一步包含用于引物延伸或连接的一种或多种试剂。

[0076] 在任何前述实施例中,支持物可以包含聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[0077] 在任何前述实施例中,隔室标签可包含单链或双链核酸分子。

[0078] 在任何前述实施例中,隔室标签可包含条形码和任选的UMI。在任何前述实施例中,支持物可以是珠粒,并且隔室标签可以包含条形码。

[0079] 在任何前述实施例中,支持物可以包含珠粒,并且包含连接到其上的多个隔室标签的珠粒可以通过均分与合并合成,单独合成,或固定化,或其任何组合形成。

[0080] 在任何前述实施例中,试剂盒可进一步包含用于均分与合并合成、单独合成或固定化或其任何组合的一种或多种试剂。

[0081] 在任何前述实施例中,隔室标签可以是记录标签内的组分,其中记录标签任选地还可以包含间隔区、条形码序列、独特分子标识符、通用引发位点或其任何组合。

[0082] 在任何前述实施例中,隔室标签可进一步包含能够与多个分析物或多肽(例如,蛋白质复合物、蛋白质或多肽)上的内部氨基酸,肽骨架或N-末端氨基酸反应的官能部分。在一个实施例中,官能部分可以包含醛、叠氮化物/炔、马来酰亚胺/硫醇、环氧/亲核试剂、逆电子需求Diels-Alder(iEDDA)基团、点击试剂或其任何组合。

[0083] 在任何前述实施例中,隔室标签可以进一步包含肽,例如蛋白连接酶识别序列,并且任选地蛋白连接酶可以是butelase I或其同源物。

[0084] 在任何前述实施例中,试剂盒可以进一步包含化学或生物试剂,例如酶,例如蛋白酶(例如,多金属蛋白酶),用于使多种分析物或多肽片段化。

[0085] 在任何前述实施例中,试剂盒可以进一步包含用于从支持物释放隔室标签的一种或多种试剂。

[0086] 在任何前述实施例中,试剂盒可以进一步包含用于形成延伸编码标签或双标签构建体的一种或多种试剂。在一个实施例中,记录标签的3'末端被封闭以防止记录标签被聚合酶延伸。在任何前述实施例中,编码标签可以包含编码器序列、UMI、通用引发位点、在其3'末端的间隔区、结合循环特异性序列,或其任何组合。

[0087] 在任何前述实施例中,双标签构建体可通过间隙填充、引物延伸或其组合产生。

[0088] 在上述任何实施例中,其中双标签分子包含衍生自记录标签的通用引发位点、衍生自记录标签的隔室标签,衍生自记录标签的隔室标签、衍生自记录标签的独特分子标识符、衍生自记录标签的可选间隔区、衍生自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区、和衍生自编码标签的通用引发位点。

[0089] 在任何前述实施例中,结合剂可以是多肽或蛋白质。

[0090] 在任何前述实施例中,结合剂可以包含氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰蛋白;抗生物素蛋白或其变体,突变体或修饰蛋白;ClpS或其变体,突变体或修饰的蛋白质;或结合氨基酸的经修饰的小分子,即万古霉素或其变体,突变体或经修饰的分子(例如结合D-丙氨酰基-D-丙氨酸的万古霉素);或抗体或其结合片段;或其任何组合。

[0091] 在任何前述实施例中,结合剂可结合到单个氨基酸残基(如N-末端氨基酸残基、C-末端氨基酸残基、或内部氨基酸残基),二肽(例如,N-末端二肽、C-末端二肽,或内部二肽),三肽(例如,N-末端三肽、C-末端三肽,或内部三肽),或翻译后修饰的分析物或多肽。

[0092] 在任何前述实施例中,结合剂可结合N-末端多肽、C-末端多肽或内部多肽。

[0093] 在任何前述实施例中,编码标签和/或记录标签可以包含一个或多个纠错码、一个或多个编码器序列、一个或多个条形码、一个或多个UMI、一个或多个隔室标签、一个或多个循环特异性序列,或其任何组合。在一些实施例中,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。

[0094] 在任何前述实施例中,编码标签和/或记录标签可以包含循环标签。

[0095] 在任何前述实施例中,试剂盒可以进一步包含独立于编码标签和/或记录标签的循环标签。

[0096] 在任何前述实施例中,试剂盒可进一步包含:(a)用于产生细胞裂解物或蛋白质样品的试剂;(b)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;(c)蛋白酶,例如胰蛋白酶、LysN或LysC;(d)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;(e)用于基于降解的多肽测序的试剂;和/或(f)用于核酸测序的试剂。

[0097] 在任何前述实施例中,试剂盒可以包含:(a)用于产生细胞裂解物或蛋白质样品的试剂;(b)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;(c)蛋白酶,例如胰蛋白酶、LysN或LysC;(d)用于将多肽(例如蛋白质)固定到包含固定的记录标签的支持物上的试剂;(e)用于基于降解的多肽测序的试剂;和/或(f)用于核酸测序的试剂。

[0098] 在任何前述实施例中,试剂盒可以包含:(a)用于产生细胞裂解物或蛋白质样品的试剂;(b)变性试剂;(c)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;(d)通用DNA引物序列;(e)用通用DNA引物序列标记多肽的试剂;(f)用于通过引物对标记的多肽进行退火的条形码化的珠;(g)用于聚合酶延伸的试剂,用于将条形码从珠粒写到标记的多肽;(h)蛋白酶,例如胰蛋白酶、LysN或LysC;(i)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;(j)用于基于降解的多肽测序的试剂;和/或(k)用于核酸测序的试剂。

[0099] 在任何前述实施例中,试剂盒可以包含:(a)交联剂;(b)用于产生细胞裂解物或蛋

白质样品的试剂;(c)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;(d)通用DNA引物序列;(e)用通用DNA引物序列标记多肽的试剂;(f)用于通过引物对标记的多肽进行退火的条形码化的珠粒;(g)用于聚合酶延伸的试剂,用于将条形码从珠粒写到标记的多肽;(h)蛋白酶,例如胰蛋白酶、LysN或LysC;(i)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;(j)用于基于降解的多肽测序的试剂;和/或(k)用于核酸测序的试剂。

[0100] 在试剂盒的任何前述实施例中,可以在溶液中或在支持物例如固体支持物上提供一种或多种组分。

[0101] 试剂盒组分还可以包括以下示例性方法和/或方面中公开和/或使用的任何分子,分子复合物或缀合物,试剂(例如,化学或生物试剂),剂,结构(例如,支持物、表面、颗粒或珠粒),反应中间体,反应产物,结合复合物,或任何其他制品。本试剂盒可用于分析任何合适的分析物,例如,大分子或多肽。在一些实施例中,本发明试剂盒可用于高度平行、高通量数字分析(例如,小分子分析),特别是多肽分析。在一些实施例中,本发明试剂盒可用于以下用于分析分析物的示例性方法中,分析物例如为大分子或多肽。

[0102] 在第一示例中是分析分析物例如大分子或多肽的方法,包含以下步骤:(a)提供分析物和与固体支持物连接的记录标签;(b)使分析物与能够结合分析物的第一结合剂接触,其中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)将第一编码标签的信息转移到记录标签,以产生第一次序延伸记录标签;(d)使分析物与能够结合分析物的第二结合剂接触,其中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;(e)将第二编码标签的信息转移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(f)分析第二次序延伸标签。

[0103] 在第二示例中是根据第一示例所述的方法,其中接触步骤(b)和(d)按顺序进行。

[0104] 在第三示例中是根据第一示例所述的方法,其中接触步骤(b)和(d)同时进行。

[0105] 在第四示例中是根据第一示例所述的方法,还包含在步骤(e)和(f)之间的以下步骤:(x)通过用能够结合修饰的分析物的第三(或更高次序)结合剂替换第二结合剂重复步骤(d)和(e)一次或多次,其中第三(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;以及(y)将第三(或更高次序)编码标签的信息转移到第二(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)中分析第三(或更高次序)延伸记录标签。

[0106] 在第五示例中是分析分析物例如大分子或多肽的方法,包含以下步骤:(a)提供分析物,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)使分析物与能够结合分析物的第一结合剂接触,其中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)将第一编码标签的信息转移到记录标签以产生第一次序延伸记录标签;(d)使分析物与能够结合分析物的第二结合剂接触,其中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)将第二编码标签的信息转移到第二记录标签以产生第二次序延伸记录标签;以及(f)分析第一和第二延伸记录标签。

[0107] 在第六示例中是根据第五示例所述的方法,其中接触步骤(b)和(d)按顺序进行。

[0108] 在第七示例中是根据第五示例所述的方法,其中接触步骤(b)和(d)同时进行。

[0109] 在第八示例中是根据第五示例所述的方法,其中步骤(a)还包含提供与固体支持

物连接的相关的第三(或更高次序)记录标签。

[0110] 在第九示例中是根据第八示例所述的方法,还包含在步骤(e)和(f)之间的以下步骤:(x)通过用能够结合修饰的分析物的第三(或更高次序)结合剂替换第二结合剂重复步骤(d)和(e)一次或多次,其中第三(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;以及(y)将第三(或更高次序)编码标签的信息转移到第三(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;以及在步骤(f)中分析第一、第二和第三(或更高次序)延伸记录标签。

[0111] 在第10示例中是根据第五至第九示例中任一个所述的方法,其中第一编码标签、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区序列。

[0112] 在第11示例中是分析肽的方法,包含步骤:(a)提供肽,和与固体支持物连接的相关记录标签;(b)用化学剂修饰肽的N-末端氨基酸(NTAA);(c)使肽与能够结合修饰的NTAA的第一结合剂接触,其中第一结合剂包含具有第一结合剂的识别信息的第一编码标签;(d)将第一编码标签的信息转移到记录标签以产生延伸记录标签;(e)分析延伸记录标签。

[0113] 在第12示例中是根据第11示例所述的方法,其中步骤(c)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,第二(或更高次序)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。

[0114] 在第19示例中是根据第18示例所述的方法,其中在肽与第一结合剂接触后,接着使肽与第二(或更高次序)结合剂接触。

[0115] 在第14示例中是根据第12示例所述的方法,其中使肽与第二(或更高次序)结合剂接触跟肽与第一结合剂的接触同时发生

[0116] 在第15示例中是根据第11-14示例中任一个所述的方法,其中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。

[0117] 在第21示例中是一种分析肽的方法,包含以下步骤:(a)提供与固体支持物连接的肽和相关的记录标签;(b)用化学剂修饰所肽的N-末端氨基酸(NTAA)以产生修饰的NTAA;(c)使肽与能够结合所述修饰的NTAA的第一结合剂接触,其中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转移到所述记录标签,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)用化学剂修饰肽的新NTAA以产生新修饰的NTAA;(g)使肽与能够结合新修饰的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)将第二编码标签的信息转移到第一延伸记录标签,以产生第二延伸记录标签;(i)分析第二延伸记录标签。

[0118] 在第21示例中是一种分析肽的方法,包含以下步骤:(a)提供连接在固体支持物上的肽和相关的记录标签;(b)使肽与能够结合肽的N-末端氨基酸(NTAA)的第一结合剂接触,其中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)将第一编码标签的信息转移到记录标签以产生延伸记录标签;以及(d)分析延伸记录标签。

[0119] 在第18示例中是根据第17示例所述的方法,其中步骤(b)还包含使肽与包含第二(或更高次序)编码标签的第二(或更高次序)结合剂接触,编码标签具有关于第二(或更高次序)结合剂的识别信息,其中第二(或更高次序)结合剂能够结合肽的NTAA以外的NTAA。

[0120] 在第19示例中是根据第18示例所述的方法,其中在肽与第一结合剂接触后,接着使肽与第二(或更高次序)结合剂接触。

[0121] 在第20示例中是根据第18示例所述的方法,其中使肽与第二(或更高次序)结合剂接触跟肽与第一结合剂的接触同时发生

[0122] 在第21示例中是一种分析肽的方法,包含以下步骤:(a)提供连接在固体支持物上的肽和相关的记录标签;(b)使肽与能够结合肽的N末端氨基酸(NTAA)的第一结合剂接触,其中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)将第一编码标签的信息转移到记录标签,以产生第一延伸记录标签;(d)移除NTAA以暴露肽的新的NTAA;(e)使肽与能够结合新NTAA的第二结合剂接触,其中第二结合剂包含具有关于第二结合剂的识别信息的第二编码标签;(f)将第二编码标签的信息转移到第一延伸记录标签,以产生第二延伸记录标签;以及(g)分析第二延伸记录标签。

[0123] 在第22示例中是根据第一至第10示例中任一个所述的方法,其中分析物是蛋白质、多肽或肽。

[0124] 在第23示例中是根据第一至第10示例中任一个所述的方法,其中分析物是肽。

[0125] 在第24示例中是根据第11至第23示例中任一个所述的方法,其中肽通过片段化分离自生物样品的蛋白质而获得。

[0126] 在第25示例中是根据第一至第10示例中任一个所述的方法,其中分析物是脂质、碳水化合物或大环。

[0127] 在第26示例中是根据第一至第25示例中任一个所述的方法,其中记录标签是DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或其组合。

[0128] 在第27示例中是根据第一至第26示例中任一个所述的方法,其中记录标签包含通用引发位点。

[0129] 在第28示例中是根据第27示例所述的方法,其中通用引发位点包含用于扩增、测序或二者的引发位点。

[0130] 在第29示例中是根据第一至第28示例中任一个所述的方法,其中记录标签包含独特分子标识符(UMI)。

[0131] 在第30示例中是根据第一至第29示例中任一个所述的方法,其中记录标签包含条形码。

[0132] 在第31示例中是根据第一至第30示例中任一个所述的方法,其中记录标签在其3'-末端包含间隔区。

[0133] 在第32示例中是根据第一至第31示例中任一个所述的方法,其中分析物和相关的记录标签共价连接到固体支持物上。

[0134] 在第33示例中是根据第一至第32示例中任一个所述的方法,其中固体支持物是珠粒、多孔珠粒、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。

[0135] 在第34示例中是根据第33示例所述的方法,其中固体支持物是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒,或可

控孔珠粒。

[0136] 在第35示例中是根据第一至第34示例中任一个所述的方法,其中将多个分析物(例如,相同分析物或不同分析物的分子)和相关的记录标签连接到固体支持物上。

[0137] 在第36实例中是根据第35实例的方法,其中多个分析物(例如,相同分析物或不同分析物的分子)在固体支持物上以>50nm的平均距离间隔。

[0138] 在第37示例中是根据第一至第36示例中任一个所述的方法,其中结合剂是多肽或蛋白质。

[0139] 在第38示例中是根据第37示例所述的方法,其中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。

[0140] 在第39示例中是根据第一至第38示例中任一个所述的方法,其中结合剂能够选择性地结合分析物。

[0141] 在第40示例中是根据第一至第39示例中任一个所述的方法,其中记录标签是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。

[0142] 在第41示例中是根据第一至40示例中任一个所述的方法,其中编码标签包含编码器序列。

[0143] 在第42示例中是根据第一至第41示例中任一个所述的方法,其中记录标签还包含间隔区、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。

[0144] 在第43示例中是根据第一至42示例中任一个所述的方法,其中结合剂和编码标签通过接头连接。

[0145] 在第44示例中是根据第一至42示例中任一个所述的方法,其中结合剂和编码标签通过SpyTag/SpyCatcher,SpyTag-KTag/SpyLigase(其中待连接的两个部分具有SpyTag/KTag对,并且SpyLigase将SpyTag连接至KTag,从而连接这两个部分)、分选酶,或SnoopTag/SnoopCatcher肽-蛋白对连接。

[0146] 在第45示例中是根据第1至第44示例中任一个所述的方法,其中编码标签到记录标签的信息转移通过DNA连接酶介导。

[0147] 在第46示例中是根据第一至第44示例中任一个所述的方法,其中编码标签到记录标签的信息转移通过DNA聚合酶介导。

[0148] 在第47示例中是根据第一至第44示例中任一个所述的方法,其中编码标签到记录标签的转移信息通过化学连接介导。

[0149] 在第48示例中是根据第一至第47示例中任一个所述的方法,其中分析延伸记录标签包含核酸测序方法。

[0150] 在第49示例中是根据第48示例所述的方法,其中核酸测序方法是通过合成测序、通过连接测序、通过杂交测序、polony测序、离子半导体测序或焦磷酸测序。

[0151] 第50示例中是根据第48示例所述的方法,其中核酸测序方法是单分子实时测序、基于纳米孔测序、或使用高级显微镜的DNA直接成像。

[0152] 在第51示例中是根据第一至第50示例中任一个所述的方法,其中延伸记录标签在分析之前扩增。

[0153] 在第52示例中是根据第一至51示例中任一个所述的方法,其中包含在延伸记录标签上的编码标签信息的次序提供关于结合到分析物的结合剂的结合次序的信息。

[0154] 在第53示例中是根据第一至52示例中任一个所述的方法,其中包含在延伸记录标签上的编码标签信息的频率提供关于结合到分析物的结合剂的结合频率的信息。

[0155] 在第54示例中是根据第1至53示例中任一个所述的方法,其中平行分析代表多个分析物(例如,相同分析物或不同分析物的分子)的多个延伸记录标签。

[0156] 在第55示例中是根据第54示例所述的方法,其中在多路复用测定中分析代表多种分析物(例如,相同分析物或不同分析物的分子)的多个延伸记录标签。

[0157] 在第56示例中是根据第一至55示例中任一个所述的方法,其中多个延伸记录标签在分析之前进行靶标富集测定。

[0158] 在第57示例中是根据第一至56示例中任一个所述的方法,其中多个延伸记录标签在分析之前进行扣除测定。

[0159] 在第58示例中是根据第一至第57示例中任一个所述的方法,其中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。

[0160] 在第59示例中是根据第一至58示例中任一个所述的方法,其中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、Edmanase酶或无水TFA来移除。

[0161] 在第60示例中是根据第一至59示例中任一个所述的方法,其中至少一个结合剂结合到末端氨基酸残基。

[0162] 在第61示例中是根据第一至60示例中任一个所述的方法,其中至少一个结合剂结合到翻译后修饰氨基酸。

[0163] 在第67示例中是一种用于分析来自包含多个蛋白质复合物、蛋白质或多肽的样品的一个或多个肽的方法,所述方法包含:(a)将样品内的多个蛋白质复合物、蛋白质或多肽分配到多个隔室中,其中每个隔室包含任选地连接到固体支持物的多个隔室标签,其中多个隔室标签在单个隔室中相同并且不同于其它隔室的隔室标签;(b)将多个蛋白质复合物、蛋白质和/或多肽片段化成多种肽;(c)在足以允许多个肽与多个隔室中的多个隔室标签退火或连接的条件下,使多个肽与多个隔室标签接触,从而产生多个隔室标记的肽;(d)从多个隔室收集隔室标记的肽;和(e)根据第1至21示例和第26至61示例中任一个所述的方法分析一个或多个隔室标记的肽。

[0164] 在第63示例中是根据第62示例所述的方法,其中隔室是微流体液滴。

[0165] 在第64示例中是根据第62示例所述的方法,其中隔室是微孔。

[0166] 在第65示例中是根据第62示例所述的方法,其中隔室是表面上分离的区域。

[0167] 在第66示例中是根据第62至65示例中任一个所述的方法,其中每个隔室平均包含一个单独细胞。

[0168] 在第67示例中是一种用于分析来自于包含多个蛋白质复合物、蛋白质、或多肽的样品的一个或多个肽的方法,所述方法包含:(a)用多个通用DNA标签标记多个蛋白质复合物、蛋白质或多肽;(b)将样品内的多个标记的蛋白质复合物、蛋白质或多肽分配到多个隔室中,其中每个隔室包含多个隔室标签,其中多个隔室标签在单个隔室中相同且不同于其它隔室的隔室标签;(c)在足以允许多个蛋白质复合物、蛋白质或多肽与多个隔室中的多个隔室标签退火或连接的条件下,使多个蛋白质复合物、蛋白质或多肽与多个隔室标签接触,从而产生多个隔室标记的蛋白质复合物、蛋白质或多肽;(d)从多个隔室收集隔室标记的蛋白质复合物、蛋白质或多肽;(e)任选地将隔室标记的蛋白质复合物、蛋白质或多肽片段化

成隔室标记的肽;和(f)根据第1至21示例和第26至61实施例中任一个所述的方法分析一个或多个隔室标记的肽。

[0169] 在第68示例中是根据第62至67示例中任一个所述的方法,其中隔室标签信息通过引物延伸或链接转移到与肽相关的记录标签。

[0170] 在第69示例中是根据第62至68示例中任一个所述的方法,其中固体支持物包含珠粒。

[0171] 在第70示例中是根据第69示例所述的方法,其中珠粒是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[0172] 在第71示例中是根据第62至70示例中任一个所述的方法,其中隔室标签包含单链或双链核酸分子。

[0173] 在第72示例中是根据第62至第71示例中任一个所述的方法,其中隔室标签包含条形码和任选的UMI。

[0174] 在第73示例中是根据第72示例所述的方法,其中固体支持物是珠粒,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粒通过均分与合并合成形成。

[0175] 在第74示例中是根据第72示例所述的方法,其中固体支持物是珠粒,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粒通过个体合成或固定化形成。

[0176] 在第75示例中是根据第62至74示例中任一个所述的方法,其中隔室标签是记录标签中的成分,记录标签任选地还包含间隔区、独特分子标识符、通用引发位点或它们的任何组合。

[0177] 在第76示例中是根据第62至75示例中任一个所述的方法,其中隔室标签还包含能够在多个蛋白质复合物、蛋白质或多肽上与内部氨基酸或N-末端氨基酸反应的官能部分。

[0178] 在第77示例中是根据第76示例所述的方法,官能部分是NHS基。

[0179] 在第78示例中是根据第76示例所述的方法,其中官能部分是醛基。

[0180] 在第79示例是根据第62至78示例中任一个所述的方法,其中多个隔室标签通过以下方法形成:打印、点样、将隔室标签喷墨打印到隔室中,或它们的组合。

[0181] 在第80示例中是根据第62至79示例中任一个所述的方法,其中隔室标签还包含肽。

[0182] 在第81示例中是根据第80示例中任一个所述的方法,其中隔室标签肽包含蛋白连接酶识别序列。

[0183] 在第82示例中是根据第81示例所述的方法,其中蛋白连接酶是butelase I或其同源物。

[0184] 在第83示例中是根据第62至82示例中任一个所述的方法,其中多个多肽用蛋白酶片段化。

[0185] 在第84示例中是根据第83示例所述的方法,其中蛋白酶是金属蛋白酶。

[0186] 在第85示例中是根据第84示例所述的方法,其中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。

[0187] 在第86示例中是根据第62至85示例中任一个所述的方法,还包含在将多个多肽分

区成多个隔室之前,从样品中扣除一种或多种丰度蛋白。

[0188] 在第87示例中是根据第62至第86示例中任一个所述的方法,还包含在将多个肽连接到隔室标签之前,从固体支持物上释放隔室标签。

[0189] 在第88示例中是根据第62示例所述的方法,还包含以下步骤(d),将隔室标签肽连接到与记录标签相关的固体支持物上。

[0190] 在第89示例中是根据第88示例所述的方法,还包含将隔室标记的肽上的隔室标签的信息转移到相关的记录标签。

[0191] 在第90示例中是根据第89示例所述的方法,还包含在步骤(e)之前将隔室标签从隔室标记的肽上移除。

[0192] 在第91示例中是根据第62至90示例中任一个所述的方法,还包含基于分析的肽的隔室标签序列来确定分析的肽的来源单细胞的身份。

[0193] 在第92示例中是根据第62至90示例中任一个所述的方法,还包含基于分析的肽的隔室标签序列来确定分析的肽的来源蛋白质或蛋白质复合物的身份。

[0194] 在第93示例中是一种用于分析多种分析物(例如,相同分析物或不同分析物的分子)的方法,所述方法包含以下步骤:(a)提供多个分析物和连接到固体支持物上的相关记录标签;(b)使多种分析物与能够结合多种分析物的多种结合剂接触,其中每种结合剂包含编码标签,编码标签具有关于结合剂的识别信息;(c)(i)将与分析物相关的记录标签的信息转移至与分析物结合的结合剂的编码标签,以产生延伸编码标签;或(ii)将与分析物相关的记录标签和与分析物结合的结合剂的编码标签的信息转移至双标签构建体;(d)收集延伸编码标签或双标签构建体;(e)任选地将步骤(b)-(d)重复一个或多个结合循环;(f)分析延伸编码标签或双标签构建体的集合。

[0195] 在第94示例中是根据第93示例所述的方法,其中分析物是蛋白质。

[0196] 在第95示例中是根据第93示例所述的方法,其中分析物是肽。

[0197] 在第96示例中是根据第95示例所述的方法,其中肽通过片段化分离自生物样品的蛋白质而获得。

[0198] 在第97示例中是根据第93至96示例中任一个所述的方法,其中记录标签是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。

[0199] 在第98示例中是根据第93至第97实施中任一个所述的方法,其中记录标签包含独特分子标识符(UMI)。

[0200] 在第99示例中是根据第93至98示例中任一个所述的方法,其中记录标签包含隔室标签。

[0201] 在第100示例中是根据第93至第99示例中任一个所述的方法,其中记录标签包含通用引发位点。

[0202] 在第101示例中是根据第93-100示例任一个所述的方法,其中记录标签在其3'-末端包含间隔区。

[0203] 在第102示例中是根据第93至101示例中任一个所述的方法,其中记录标签的3'-末端是封闭的,以防止记录标签通过合成酶延伸,并且分析物相关记录标签的信息和结合到分析物的结合剂的编码标签的信息被转移到双标签构建体。

[0204] 在第103示例中是根据第93至102示例中任一个所述的方法,其中编码标签包含编

码序列。

[0205] 在第104示例中是根据第93至103示例中任一个所述的方法,其中编码标签包含UMI。

[0206] 在第105示例中是根据第93至104示例中任一个所述的方法,其中编码标签包含通用引发位点。

[0207] 在第106示例中是根据第93至105示例任一个所述的方法,编码标签在其3'-末端包含间隔区。

[0208] 在第107示例中是根据第93至106示例中任一个所述的方法,其中编码标签包含结合循环特异性序列。

[0209] 在第108示例中是根据第93至107示例中任一个所述的方法,其中结合剂和编码标签通过接头连接。

[0210] 在第109示例中是根据第93至108示例中任一个所述的方法,其中记录标签到编码标签的信息转移通过引物延伸实现。

[0211] 在第110示例中是根据第93至108示例中任一个所述的方法,其中记录标签到编码标签的信息转移通过连接实现。

[0212] 在第111示例中是根据第93至108示例中任一个所述的方法,其中双标签构建体通过间隙填充、引物延伸或它们二者产生的。

[0213] 在第112示例中是根据示例93-97、107、108和111中任一个所述的方法,其中双标签分子包含衍生自记录标签的通用引发位点、衍生自记录标签的隔室标签、衍生自记录标签的独特分子标识符、衍生自记录标签的可选间隔区、衍生自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区和衍生自编码标签的通用引发位点。

[0214] 在第113示例中是根据第93至112示例中任一个所述的方法,其中分析物和相关的记录标签共价结合到固体支持物上。

[0215] 在第114示例中是根据第113示例的方法,其中固体支持物是珠粒、多孔珠粒、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。

[0216] 在第115示例中是根据第114示例所述的方法,其中固体支持物是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒,或可控孔珠粒。

[0217] 在第116示例中是根据第93至第115示例中任一个所述的方法,其中结合剂是多肽或蛋白质。

[0218] 在第117示例中是根据第116示例所述的方法,其中结合剂是修饰的氨肽酶、修饰的氨酰tRNA合成酶、修饰的anticalin或者抗体或其结合片段。

[0219] 在第118示例中是根据第95至117示例中任一个所述的方法,其中结合剂结合单个氨基酸残基、二肽、三肽或肽的翻译后修饰。

[0220] 在第119示例中是根据第118示例所述的方法,其中结合剂与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。

[0221] 在第120示例中是根据第118示例所述的方法,其中结合剂结合N-末端肽、C-末端肽或内部肽。

[0222] 在第121示例中是根据第119示例所述的方法,其中结合剂与N-末端氨基酸残基结合,并且在每个结合循环后裂解N-末端氨基酸残基。

[0223] 在第122示例中是根据第119示例所述的方法,其中结合剂结合C-末端氨基酸残基,并且在每个结合循环后裂解C-末端氨基酸残基。

[0224] 示例123.根据示例121所述的方法,其中N末端氨基酸残基通过埃德曼降解法裂解。

[0225] 示例124.根据示例93所述的方法,其中结合剂是氨基酸或翻译后修饰的位点特异性共价标签。

[0226] 示例125.根据示例93-124中任一个所述的方法,其中在步骤(b)之后,将包含分析物和相关结合剂的复合物从固体支持物上解离并分散成液滴或微流体液滴乳液。

[0227] 示例126.根据示例125所述的方法,其中每个微流体液滴,平均而言,都包含一个含分析物和结合剂的复合物。

[0228] 示例127.根据示例125或126所述的方法,其中记录标签在产生延伸编码标签或双标签构建体之前扩增。

[0229] 示例128.根据示例第125至127示例中任一个所述的方法,其中乳液融合PCR用于将记录标签信息转移到编码标签或创建一组双标签构建体。

[0230] 示例129.根据示例93至128中任一个所述的方法,其中延伸编码标签或双标签构建体的集合是在分析之前扩增的。

[0231] 示例130.根据示例93至129中任一个所述的方法,其中分析延伸编码标签或双标签构建体的集合包含核酸测序方法。

[0232] 示例131.根据示例130所述的方法,其中核酸测序方法是通过合成测序、通过连接测序、通过杂交测序、polony测序、离子半导体测序或焦磷酸测序。

[0233] 示例132.根据示例130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。

[0234] 示例133.根据示例130所述的方法,其中分析物的部分组成是通过采用特异性隔室标签或可选地采用UMI分析多个延伸编码标签或双标签标签构建体来确定的。

[0235] 示例134.根据示例1至133中任一个所述的方法,其中分析步骤用具有每碱基错误率>5%、>10%、>15%、>20%、>25%或>30%的测序方法进行。

[0236] 示例135.根据示例1至134中任一个所述的方法,其中编码标签、记录标签或两者的识别元件包括纠错码。

[0237] 示例136.根据示例135所述的方法,其中识别元件选自编码器序列、条形码、UMI、隔室标签、循环特异性序列或其任何组合。

[0238] 示例137.根据示例135或136所述的方法,其中纠错码选自汉明码、李距离码、非对称李距离码、里德-所罗门码和Levenshtein-Tenengolts码。

[0239] 示例138.根据示例1至134中任一个所述的方法,其中编码标签、记录标签或两者的识别元件能够产生独特的电流或离子通量或光学特征,其中分析步骤包含检测独特的电流或离子通量或光学特征以识别识别元件。

[0240] 示例139.根据示例138所述的方法,其中识别元件选自编码器序列、条形码、UMI、隔室标签、循环特异性序列或其任何组合。

[0241] 示例140.一种用于分析多种分析物(例如,相同分析物或不同分析物的分子)的方法,所述方法包含以下步骤:(a)提供多个分析物和连接到固体支持物上的相关记录标签;(b)使多种分析物与能够结合同源分析物的多种结合剂接触,其中每种结合剂包含编码标签,编码标签具有关于结合剂的识别信息;(c)将第一结合剂的第一编码标签的信息转移至与第一分析物相关联的第一记录标签,以产生第一次序延伸记录标签,其中第一结合剂结合至第一分析物;(d)使多种分析物与能够结合同源分析物的多种结合剂接触;(e)将第二结合剂的第二编码标签的信息转移至第一次序延伸记录标签,以产生第二次序延伸记录标签,其中第二结合剂结合至第一分析物;(f)任选地将步骤(d)-(e)重复"n"个结合循环,其中将与第一分析物结合的各结合剂的各编码标签的信息转移至前一结合循环产生的延伸记录标签,以产生代表第一分析物的第n次序延伸记录标签;(g)分析第n次序延伸记录标签。

[0242] 示例141.根据示例140所述的方法,其中生成并分析表示多种分析物的多个第n次序延伸记录标签。

[0243] 示例142.根据示例140或141所述的方法,其中分析物是蛋白质。

[0244] 示例143.根据示例142所述的方法,其中分析物是肽。

[0245] 示例144.根据示例143所述的方法,其中肽通过片段化生物样品的蛋白质而获得。

[0246] 示例145.根据示例140至144中任一个所述的方法,其中多种分析物包含来自多个合并的样品的分析物(例如,大分子,诸如多肽、蛋白质、蛋白质复合物)。

[0247] 示例146.根据示例140至145中任一个所述的方法,其中记录标签是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。

[0248] 示例147.根据示例140至146中任一个所述的方法,其中记录标签包含独特分子标识符(UMI)。

[0249] 示例148.根据示例140至147中任一个所述的方法,其中记录标签包含隔室标签。

[0250] 示例149.根据示例140至148中任一个所述的方法,其中记录标签包含通用引发位点。

[0251] 示例150.根据示例140至149中任一个所述的方法,其中记录标签在其3'-末端包含间隔区。

[0252] 示例151.根据示例140至150中任一个所述的方法,其中编码标签包含编码器序列。

[0253] 示例152.根据示例140至151中任一个所述的方法,其中编码标签包含UMI。

[0254] 示例153.根据示例140至152中任一个所述的方法,其中编码标签包含通用引发位点。

[0255] 示例154.根据示例140至153中任一个所述的方法,其中编码标签在其3'-端包含间隔区。

[0256] 示例155.根据示例140至154中任一个所述的方法,其中编码标签包含结合循环特异性序列。

[0257] 示例156.根据示例140至155中任一个所述的方法,其中编码标签包含独特分子标

识符。

[0258] 示例157.根据示例140至156中任一个所述的方法,其中结合剂和编码标签通过接头连接。

[0259] 示例158.根据示例140至157中任一个所述的方法,其中记录标签到编码标签的信息转移通过引物延伸介导。

[0260] 示例159.根据示例140至158中任一个所述的方法,其中记录标签到编码标签的信息转移通过连接介导。

[0261] 示例160.根据示例140至159中任一个所述的方法,其中多种分析物、相关的记录标签,或者两者都共价结合到固体支持物上。

[0262] 示例161.根据示例140至160中任一个所述的方法,固体支持物是珠粒、多孔珠粒、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。

[0263] 示例162.根据示例161所述的方法,其中固体支持物是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[0264] 示例163.根据示例140至162中任一个所述的方法,其中结合剂是多肽或蛋白质。

[0265] 示例164.根据示例163所述的方法,其中结合剂是修饰的氨肽酶、修饰的氨酰tRNA合成酶、修饰的anticalin,或者抗体或其结合片段。

[0266] 示例165.根据示例142至164中任一个所述的方法,其中结合剂结合单个氨基酸残基、二肽、三肽或肽的翻译后修饰。

[0267] 示例166.根据示例165所述的方法,其中结合剂结合N-末端氨基酸残基、C-末端氨基酸残基或内部氨基酸残基。

[0268] 示例167.根据示例165所述的方法,其中结合剂结合N-末端肽、C-末端肽或内部肽。

[0269] 示例168.根据示例142至164中任一项所述的方法,其中结合剂结合修饰的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。

[0270] 示例169.根据示例166或168所述的方法,其中结合剂结合N-末端氨基酸残基或修饰的N-末端氨基酸残基的化学标签,并且N-末端氨基酸残基在每个结合循环后被裂解。

[0271] 示例170.根据示例166或168所述的方法,其中结合剂结合C-末端氨基酸残基或修饰的C-末端氨基酸残基的化学标签,并且C-末端氨基酸残基在每个结合循环后被裂解。

[0272] 示例171.根据示例169所述的方法,其中N-末端氨基酸残基通过埃德曼降解法、埃德曼酶、修饰的氨基肽酶、或修饰的酰基肽水解酶裂解。

[0273] 示例172.根据示例163所述的方法,其中结合剂是氨基酸或翻译后修饰的位点特异性共价标签。

[0274] 示例173.根据示例140至172中任一个所述的方法,其中多个第n次序延伸记录标签在分析之前扩增。

[0275] 示例174.根据示例140至173中任一个所述的方法,其中分析第n次序延伸记录标签包含核酸测序方法。

[0276] 示例175.根据示例174所述的方法,其中代表多种分析物的多个第n次序延伸记录标签被平行分析。

[0277] 示例176.根据示例174或175所述的方法,其中核酸测序方法是通过合成测序、通过连接测序、通过杂交测序、polony测序、离子半导体测序或焦磷酸测序。

[0278] 示例177.根据示例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。

[0279] 示例1'.一种方法,包含:(a)使一组蛋白质与试剂库接触,其中每种蛋白质与记录标签直接或间接相关,其中每种试剂包含(i)小分子,肽或肽模拟物,肽类(例如,类肽,β-肽,或D-肽肽模拟物),多糖,或适体(例如,核酸适体,例如DNA适体或肽适体),和(ii)编码标签,其包含关于小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽拟肽),多糖或适体的识别信息,其中每种蛋白质和/或其相关记录标签,或每个试剂,直接或间接固定在支持物上;(b)允许在(i)与小分子,肽或肽模拟物,肽类(例如,类肽,β-肽,或D-肽肽模拟物),多糖或适体和/或反应的每种蛋白质相关的记录标签,以及(ii)一个或多个试剂的编码标签之间转移信息,以生成延伸记录标签和/或延伸编码标签;以及(c)分析延伸记录标签和/或延伸编码标签。

[0280] 示例2'.根据示例1'所述的方法,其中每种蛋白质与支持物上的其他蛋白质间隔开的平均距离等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[0281] 示例3'.根据示例1'或2'所述的方法,其中每种蛋白质及其相关记录标签与支持物上的其他蛋白质及其相关记录标签间隔开的平均距离等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[0282] 示例4'.根据示例1'至3'中任一个所述的方法,其中将一种或多种蛋白质和/或其相关记录标签共价固定到支持物上(例如,通过接头),或非共价固定到支持物上(例如,通过结合对)。

[0283] 示例5'.根据示例1'至4'中任一个所述的方法,其中蛋白质和/或其相关联的记录标签的一个子集共价地固定支持物上,而该蛋白质和/或其相关联的记录标签的另一个子集非共价地固定到支持物上。

[0284] 示例6'.根据示例1'至5'中任一个所述的方法,其中将一个或多个记录标签固定到支持物上,从而固定相关的蛋白质。

[0285] 示例7'.根据示例1'至6'中任一个所述的方法,其中将一种或多种蛋白质固定到支持物上,从而固定相关的记录标签。

[0286] 示例8’.根据示例1至7中任一个所述的方法,其中至少一种蛋白质与其相关记录标签共定位,同时各自独立地固定到支持物上。

[0287] 示例9’.根据示例1’至8’中任一个所述的方法,其中至少一种蛋白质和/或其相关记录标签直接或间接地与固定接头相关,并且固定接头直接或间接地固定到支持物上,由此将至少一种蛋白质和/或其相关记录标签固定到支持物上。

[0288] 示例10’.根据示例1’至9’中任一个所述的方法,其中固定的记录标签的密度等于或大于固定的蛋白质的密度。

[0289] 示例11’.根据示例10’所述的方法,其中固定的记录标签的密度为固定的蛋白质的密度至少约2倍,至少约3倍,至少约4倍,至少约5倍,至少约6倍,至少约7倍,至少约8倍,至少约9倍,至少约10倍,至少约20倍,至少约50倍,至少约100倍或更多。

[0290] 示例12’.根据示例1’所述的方法,其中每种试剂与固定在支持物上的其他试剂间隔开的平均距离等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[0291] 示例13’.根据示例12’所述的方法,其中将一种或多种试剂共价固定到支持物上(例如,通过接头),或者非共价固定到支持物上(例如,通过结合对)。

[0292] 示例14’.根据示例12’或13’的方法,其中试剂的一个子集共价固定至支持物,而试剂的另一个子集非共价固定至支持物。

[0293] 示例15’.根据示例12’至14’中任一个所述的方法,其中对于一种或多种试剂,将小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体固定至支持物,从而固定编码标签。

[0294] 示例16’.根据示例12’至15’中任一个所述的方法,其中对于一种或多种试剂,将编码标签固定到支持物上,从而固定小分子,肽或肽模拟物,肽类(例如,类肽、β-肽,或D-肽肽模拟物),多糖,或适体。

[0295] 示例17’.根据示例1’至16’中任一个所述的方法,其中将信息从至少一个编码标签转移到至少一个记录标签,由此产生至少一个延伸记录标签。

[0296] 示例18’.根据示例1’至17’中任一个所述的方法,其中将信息从至少一个记录标签转移到至少一个编码标签,由此产生至少一个延伸编码标签。

[0297] 示例19’.根据示例1’至18’中任一个所述的方法,其中产生至少一个双标签构建体,双标签构建体包含来自编码标签的信息和来自记录标签的信息。

[0298] 示例20’.根据示例1’至19’中任一个所述的方法,其中至少一种蛋白质与两种或更多种试剂的小分子,肽或肽模拟物,肽类(例如,类肽、β-肽或D-肽肽模拟物),多糖或适体结合和/或反应。

[0299] 示例21’.根据示例20’所述的方法,其中延伸记录标签或延伸编码标签包含关于两种或更多种试剂的小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体的识别信息。

[0300] 示例22’.根据示例1’至21’中任一个所述的方法,其中蛋白质中的至少一种与两个或更多个记录标签相关,其中两个或更多个记录标签可以相同或不同。

[0301] 示例23’.根据示例1’至22’中任一个所述的方法,其中试剂中的至少一种包含两个或更多个编码标签,其中两个或更多个编码标签可以相同或不同。

[0302] 示例24’.根据示例1’至23’中任一个所述的方法,其中信息的转移通过连接(例如,酶或化学连接,夹板连接,粘端连接,单链(ss)连接例如ssDNA连接,或其任何组合),聚合酶介导的反应(例如,单链核酸或双链核酸的引物延伸),或其任何组合完成。

[0303] 示例25’.根据示例24’所述的方法,其中连接和/或聚合酶介导的反应相对于蛋白质与小分子,肽或肽模拟物,肽类(例如,类肽、β-肽,或D-肽肽模拟物),多糖,或适体之间的结合占用时间或反应时间具有更快的动力学,任选地,其中用于连接和/或聚合酶介导的反应的试剂存在于与蛋白质与小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物,多糖或适体)之间的结合或反应相同的反应体积中,并且进一步任选地其中通过使用伴随的结合/编码步骤,和/或通过使用降低的编码或信息写入步骤的温度来减慢结合剂的解离速率来实现信息转移。

[0304] 示例26’.根据示例1’至25’中任一个所述的方法,其中每种蛋白质通过单独的附着与其记录标签相关,和/或其中每个小分子,肽或肽模拟物,肽类(例如,类肽、β-肽,或D-肽肽模拟物),多糖或适体通过单独的附着与其编码标签相关。

[0305] 示例27’.根据示例26’所述的方法,其中通过核糖体或mRNA/cDNA展示发生附着,其中记录标签和/或编码标签序列信息包含在mRNA序列中。

[0306] 示例28’.根据示例27’所述的方法,其中记录标签和/或编码标签在mRNA序列的3’端包含通用引物序列、条形码和/或间隔区序列。

[0307] 示例29’.根据示例28’的方法,其中记录标签和/或编码标签在3’端进一步包含限制酶消化位点。

[0308] 示例30’.根据示例1’至29’中任一个所述的方法,其中该组蛋白质是蛋白质组或其子集,任选地其中该组蛋白质是使用基因组或其子集的体外转录随后体外翻译产生的,或使用转录组或其子集的体外翻译产生的。

[0309] 示例31’.根据示例30’所述的方法,其中蛋白质组的子集包括激酶组;分泌体;受体组(例如,GPCRome);免疫蛋白质组;营养蛋白质组;由翻译后修饰(例如,磷酸化,泛素化,甲基化,乙酰化,糖基化,氧化,脂质化和/或亚硝基化)定义的蛋白质组子集,例如磷酸蛋白质组(例如,磷酸酪氨酸-蛋白质组,酪氨酸-激酶组和酪氨酸-磷酸酶组),糖蛋白质组等;与组织或器官、发育阶段或生理或病理状况相关的蛋白质组子集;与细胞过程如细胞周期,分化(或去分化),细胞死亡,衰老,细胞迁移,转化或转移相关的蛋白质组子集;或其任何组合。

[0310] 示例32’.根据示例1’至31’中任一个所述的方法,其中该组蛋白质来自哺乳动物,例如人,非人动物,鱼,无脊椎动物,节肢动物,昆虫或植物,例如,酵母,细菌,例如,大肠杆菌,病毒,例如,HIV或HCV,或其组合。

[0311] 示例33’.根据示例1’至32’中任一个所述的方法,其中该组蛋白质包含蛋白质复合物或其亚基。

[0312] 示例34’.根据示例1’至33’中任一个所述的方法,其中记录标签包含核酸、寡核苷

酸,修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代,或它们的组合。

[0313] 示例35'.根据示例1'至34'中任一个所述的方法,其中记录标签包含通用引发位点。

[0314] 示例36'.根据示例1'至35'中任一个所述的方法,其中记录标签包含用于扩增、测序或两者的引发位点,例如,通用引发位点包含用于扩增、测序或两者的引发位点。

[0315] 示例37'.根据示例1'至36'中任一个所述的方法,其中记录标签包含独特分子标识符(UMI)。

[0316] 示例38'.根据示例1'至37'中任一个所述的方法,其中记录标签包含条形码。

[0317] 示例39'.根据示例1-38任一个所述的方法,其中记录标签在其3'-末端包含间隔区。

[0318] 示例40'.根据示例1-39中任一个所述的方法,其中支持物是固体支持物,例如刚性固体支持物、柔性固体支持物或软固体支持物,并且包括多孔支持物或非多孔支持物。

[0319] 示例41'.根据示例1'至40'中任一个所述的方法,其中支持物包含珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[0320] 示例42'.根据示例41'所述的方法,其中支持物包含聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[0321] 示例43'.根据示例1'至42'中任一个所述的方法,其用于平行分析该组蛋白质和小分子库,和/或肽或肽模拟物,和/或肽类(例如,类肽,β-肽,或D-肽肽模拟物),和/或多糖,和/或适体之间的相互作用,以便产生小分子-蛋白质结合基质,和/或肽/肽模拟物-蛋白质结合基质,和/或肽类-蛋白质结合基质(例如,类肽-蛋白质结合基质,β-肽-蛋白质结合基质,或D-肽模拟-蛋白质结合基质),和/或多糖-蛋白质结合基质,和/或适体-蛋白结合基质。

[0322] 示例44'.根据示例43'所述的方法,其中基质尺寸为约$10^2$,约$10^3$,约$10^4$,约$10^5$,约$10^6$,约$10^7$,约$10^8$,约$10^9$,约$10^{10}$,约$10^{11}$,约$10^{12}$,约$10^{13}$,约$10^{14}$或更大,例如约$2\times10^{13}$。

[0323] 示例45'.根据示例1'至44'中任一个所述的方法,其中编码标签包含核酸、寡核苷酸、修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代,或它们的组合。

[0324] 示例46'.根据示例1'至45'中任一个所述的方法,其中编码标签包含识别小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体的编码器序列。

[0325] 示例47'.根据示例1'至46'中任一个所述的方法,其中编码标签包含间隔区、独特分子标识符(UMI)、通用引发位点,或其任何组合。

[0326] 示例48'.根据示例1'至47'中任一个所述的方法,其中小分子,肽或肽模拟物,肽类(例如,类肽,β-肽,或D-肽肽模拟物),多糖,或适体和编码标签通过接头或结合对连接。

[0327] 示例49'.根据示例1'至48'中任一个所述的方法,其中小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体和编码标签通过SpyTag-KTag/SpyLigase(其中待连接的两个部分具有SpyTag-KTag/SpyLigase对,并且SpyLigase将SpyTag连接至KTag,因此将两个部分连接),SpyTag/SpyCatcher,SnoopTag/SnoopCatcher肽-蛋白质对,分选酶,或HaloTag/HaloTag配体对,或其任何组合连接。

[0328] 示例50'.一种用于分析多肽的方法,包含:(a)使(i)多肽的一组片段与(ii)结合剂库接触,其中每个片段与记录标签直接或间接相关,其中每个结合剂包含结合基团和编码标签,编码标签包含关于结合基团的识别信息,其中结合基团能够结合结合至片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合至通过官能化试剂修饰的一个或多个N-末端,内部或C-末端氨基酸,并且其中每个片段和/或其相关记录标签,或每个结合剂直接或间接固定至支持物;(b)当结合基团与片段的一个或多个N-末端、内部或C-末端氨基酸结合时,允许在(i)与每个片段相关的记录标签和(b)编码标签之间转移信息,以产生延伸记录标签和/或延伸编码标签;以及(c)分析延伸记录标签和/或延伸编码标签。

[0329] 示例51'.根据示例50'的方法,其中一个或多个N-末端、内部或C-末端氨基酸包含:(i)N-末端氨基酸(NTAA);(ii)N-末端二肽序列;(iii)N-末端三肽序列;(iv)内部氨基酸;(v)内部二肽序列;(vi)内部三肽序列;(vii)C-末端氨基酸;(viii)C-末端二肽序列;或(ix)C-末端三肽序列,或其任何组合,任选地其中(i)-(ix)中的任何一个或多个氨基酸残基被修饰或官能化。

[0330] 示例52'.根据示例51'所述的方法,其中一个或多个N-末端、内部或C-末端氨基酸在每个残基处独立地选自由以下组成的群组: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),酪氨酸(Y或Tyr),其任何组合。

[0331] 示例53'.根据示例50'至52'中任一个所述的方法,其中结合基团包含多肽或其片段,蛋白质或多肽链或其片段,或蛋白质复合物或其亚基,如抗体或其抗原结合片段。

[0332] 示例54'.根据示例50'至53'中任一个所述的方法,其中结合基团包含抗体或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰的蛋白质;抗生物素蛋白或其变体,突变体或修饰的蛋白质;ClpS或其变体,突变体或修饰的蛋白质;UBR盒蛋白质或其变体,突变体或修饰的蛋白质;或结合氨基酸的修饰的小分子,即万古霉素或其变体,突变体或修饰的分子;或其任何组合。

[0333] 示例55'.根据示例50'至54'中任一个所述的方法,其中结合基团能够选择性地和/或特异性地结合至官能化的N-末端氨基酸(NTAA)、N-末端二肽序列或N-末端三肽序列或其任何组合。

[0334] 示例56'.一种用于分析多种多肽的方法,包含:(a)用多个通用标签标记多个多肽的每个分子;(b)在适于多个通用标签与多个隔室标签退火或连接的条件下,使多个多肽与多个隔室标签接触,从而将多个多肽分区成多个隔室(例如,珠粒表面,微流体液滴,微孔,或表面上的分离区域,或其任何组合),其中多个隔室标签在每个隔室内是相同的并且不同于其他隔室的隔室标签;(c)将每个隔室中的多肽片段化,从而产生一组多肽片段,每个多

肽片段与包含至少一个通用多核苷酸标签和至少一个隔室标签的记录标签相关;(d)将该组多肽片段直接或间接固定到支持物上;(e)使固定的多肽片段组与结合剂库接触,其中每种结合剂包含结合基团和编码标签,编码标签包含关于结合基团的识别信息,其中结合基团能够结合至所述片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合至通过官能化试剂修饰的一个或多个N-末端、内部,或C-(c)端氨基酸;(f)当结合基团与片段的一个或多个N-末端、内部或C-末端氨基酸结合时,允许在(i)与每个片段相关的记录标签和(ii)编码标签之间转移信息,以产生延伸记录标签和/或延伸编码标签;(g)解析延伸记录标签和/或延伸编码标签。

[0335] 示例57'.根据示例56'所述的方法,其中具有相同隔室标签的多种多肽属于相同蛋白质。

[0336] 示例58'.根据示例56'所述的方法,其中具有相同隔室标签的多种多肽属于不同的蛋白质,例如,两种,三种,四种,五种,六种,七种,八种,九种,十种或更多种蛋白质。

[0337] 示例59'.根据示例56'至58'中任一个所述的方法,其中多个隔室标签被固定到多个基板上,每个基板限定隔室。

[0338] 示例60'.根据示例59'所述的方法,其中多个基板选自由以下组成的群组:珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[0339] 示例61'.根据示例59'或60'所述的方法,其中多个基板中的每一个包含条形码化颗粒,如条形码化珠粒,例如,聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[0340] 示例62'.根据示例59'-61'中任一个所述的方法,其中支持物选自由以下组成的群组:珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[0341] 示例63'.根据示例62'的方法,其中支持物包含测序珠粒,例如,聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[0342] 示例64'.根据示例56'-63'中任一个所述的方法,其中每个片段及其相关记录标签与支持物上的其他片段及其相关记录标签间隔开的平均距离等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[0343] 示例65’.一种用于分析多种多肽的方法,包含:(a)将多个多肽固定至多个基板,其中每个基板包含多个记录标签,每个记录标签包含隔室标签,任选地其中每个隔室是珠粒、微流体液滴、微孔,或表面上的间隔区域,或其任何组合;(b)将固定在每个基板上的多肽片段化(例如,通过蛋白酶消化)从而产生一组固定在基板上的多肽片段;(c)使固定的多肽片段组与结合剂库接触,其中每种结合剂包含结合基团和编码标签,编码标签包含关于结合基团的识别信息,其中结合基团能够结合至片段的一个或多个N末端、内部或C-末端氨基酸,或能够结合至通过官能化试剂修饰的一个或多个N-末端、内部,或C-末端氨基酸;(d)当结合基团与每个片段的一个或多个N-末端、内部或C-末端氨基酸结合时,允许在(i)记录标签和(ii)编码标签之间转移信息,以产生延伸记录标签和/或延伸编码标签;以及(e)分析延伸记录标签和/或扩展编码标签。

[0344] 示例66’.根据示例65’所述的方法,其中具有相同隔室标签的多种多肽属于相同蛋白质。

[0345] 示例67’.根据示例65’所述的方法,其中具有相同隔室标签的多种多肽属于不同的蛋白质,例如,两种,三种,四种,五种,六种,七种,八种,九种,十种或更多种蛋白质。

[0346] 示例68’.根据示例65’至67’中任一个所述的方法,其中每个基板限定隔室。

[0347] 示例69’.根据示例65’至68’中任一个所述的方法,其中多个基板选自由以下组成的群组:珠粒、多孔珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒(Fe304)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[0348] 示例70’.根据示例65’-69’中任一个所述的方法,其中多个基板中的每一个包含条形码化颗粒,如条形码化珠粒,例如,聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[0349] 示例71’.根据示例50’-70’中任一个所述的方法,其中官能化试剂包含化学剂、酶和/或生物剂,如异硫氰酸酯衍生物,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。

[0350] 示例72’.根据示例50’-71’中任一个所述的方法,其中记录标签包含核酸、寡核苷酸,修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代,或它们的组合。

[0351] 示例73’.根据示例50’至72’中任一个所述的方法,其中记录标签包含通用引发位点;用于扩增、测序或两者的引发位点;任选地,独特分子标识符(UMI);条形码;任选地,在其3’末端的间隔区;或其组合。

[0352] 示例74’.根据示例50’至73’中任一个所述的方法,其用于测定多肽或多种多肽的序列。

[0353] 示例75’.根据示例50’-74’中任一个所述的方法,其中编码标签包含核酸、寡核苷酸、修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分

子、PNA分子、γPNA分子或吗啉代,或它们的组合。

[0354] 示例76'.根据示例50至75中任一个所述的方法,其中编码标签包含编码器序列,任选的间隔区,任选的独特分子标识符(UMI),通用引发位点,或其任何组合。

[0355] 示例77'.根据示例50'-76'中任一个所述的方法,其中结合基团和编码标签通过接头或结合对连接。

[0356] 示例78'.根据示例50-77中任一个所述的方法,其中结合基团和编码标签通过SpyTag-KTag/SpyLigase(其中待连接的两个部分具有SpyTag/KTag对,并且SpyLigase将SpyTag连接至KTag,由此将两个部分连接),SnoopTag/SnoopCatcher肽-蛋白质对,分选酶,或HaloTag/HaloTag配体对,或其任何组合。

[0357] 示例79'.根据1'至78'中任一个所述的方法,其中编码标签和/或记录标签包含一个或多个纠错码,一个或多个编码器序列,一个或多个条形码,一个或多个UMI,一个或多个隔室标签,或其任何组合。

[0358] 示例80'.根据示例79'所述的方法,其中纠错码选自汉明码、李距离码、非对称李距离码、里德-所罗门码和Levenshtein-Tenengolts码。

[0359] 示例81'.根据示例1'-80'中任一个所述的方法,其中分析延伸记录标签和/或延伸编码标签包含核酸序列分析。

[0360] 示例82'.根据示例81'所述的方法,其中核酸序列分析包含核酸测序方法,例如通过合成测序,通过连接测序,通过杂交测序,polony测序,离子半导体测序或焦磷酸测序,或其任何组合。

[0361] 示例83'.根据示例82'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。

[0362] 示例84'.根据示例1'-83'中任一个所述的方法,进一步包含一个或多个洗涤步骤。

[0363] 示例85'.根据示例1'-84'中任一个所述的方法,其中延伸记录标签和/或延伸编码标签在分析之前被扩增。

[0364] 示例86'.根据示例1'-85'中任一个所述的方法,其中延伸记录标签和/或延伸编码标签在分析之前进行进行靶标富集测定。

[0365] 示例87'.根据示例1'-86'中任一个所述的方法,其中延伸记录标签和/或延伸编码标签在分析之前进行扣除测定。

[0366] 在一个方面,本文公开了试剂盒,其包含:(a)试剂库,其中每个试剂包含(i)小分子,肽或肽模拟物,肽类(例如,类肽,β-肽,或D-肽拟肽),多糖,和/或适体,和(ii)编码标签,其包含关于小分子,肽或肽模拟物,肽类(例如,类肽,β-肽,或D-肽模拟肽),多糖,或适体的识别信息;以及任选地(b)一组蛋白质,其中每种蛋白质直接或间接地与记录标签相关,其中每种蛋白质和/或其相关联的记录标签或每种试剂直接或间接地固定到支持物上,并且其中该组蛋白质、记录标签和试剂库被配置为允许在(i)与一种或多种试剂的小分子,肽或肽模拟物,肽类(例如,类肽,β-肽,或D-肽肽模拟物),多糖,或适体结合和/或反应的每种蛋白质相关的记录标签以及(ii)一个或多个试剂的编码标签之间转移信息,以生成延伸记录标签和/或延伸编码标签。

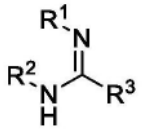
[0367] 在一个方面,本文公开了用于分析多肽的试剂盒,其包含:(a)结合剂库,其中每个

结合剂包含结合基团和编码标签,编码标签包含关于结合基团的识别信息,其中结合基团能够结合片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合通过官能化试剂修饰的一个或多个N-末端、内部,或C-末端氨基酸;以及任选地(b)多肽的一组片段,其中每个片段直接或间接地与记录标签相关,或(b')用于片段化多肽的工具,例如蛋白酶,其中每个片段和/或其相关记录标签,或每个结合剂直接或间接地固定到支持物上,并且其中多肽的该组片段、记录标签和结合剂库被配置为允许在结合基团和片段的一个或多个N-末端、内部或C-末端氨基酸之间结合时在(i)与每个片段相关的记录标签和(ii)编码标签之间转移信息,以产生延伸记录标签和/或延伸编码标签。

[0368] 在一个方面,本文公开了用于分析多种多肽的试剂盒,包含:(a)结合剂库,其中每个结合剂包含结合基团和编码标签,编码标签包含关于结合基团的识别信息,其中结合基团能够结合片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合通过官能化试剂修饰的一个或多个N末端、内部,或C端氨基酸;和(b)多个基板,任选地具有固定于其上的多个多肽,其中每个基板包含多个记录标签,每个记录标签包含隔室标签,任选地其中每个隔室是珠粒、微流体液滴、微孔,或表面上的间隔区域,或其任何组合,其中固定在每个基板上的多肽被配置成片段化(例如,通过蛋白酶裂解)以产生一组固定在基板上的多肽片段,其中多个多肽、记录标签和结合剂库被配置为允许在结合基团与每个片段的一个或多个N-末端、内部或C-末端氨基酸之间结合时在(i)记录标签和(ii)编码标签之间转移信息,以生成延伸记录标签和/或延伸编码标签。

[0369] 以下方面提供了本公开的补充说明。

[0370] 方面1.一种分析多肽的方法,包含以下步骤:(a)提供任选地直接或间接与记录标签结合的多肽;(b)用化学试剂官能化多肽的N末端氨基酸(NTAA),其中化学试剂包含选自由以下组成的群组的化合物:(i)式(I)的化合物

[0371] 
(I)

[0372] 或其盐或缀合物,

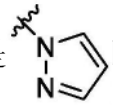
[0373] 其中,

[0374] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

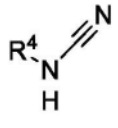
[0375] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[0376] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基未被取代或被取代;

[0377] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[0378] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[0379] (ii)式(II)化合物:

[0380] 
(II)

[0381] 或其盐或缀合物,

[0382] 其中,

[0383] $R^4$是H,$C_{1-6}$烷基,环烷基,-C(O)$R^g$或-C(O)O$R^g$;和

[0384] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[0385] (iii)式(III)化合物:

[0386] $R^5$-N=C=S (iii)

[0387] 或其盐或缀合物,

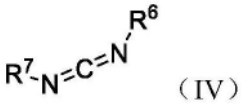
[0388] 其中,

[0389] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[0390] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自由选自卤素,-$NR^hR^i$,-S(O)$_2R^j$或杂环基组成的群组的一个或多个基团取代;

[0391] $R^h$,$R^i$和$R^j$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代。

[0392] (iv)式(IV)化合物:

[0393] 

(IV)

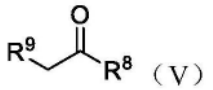
[0394] 或其盐或缀合物,

[0395] 其中,

[0396] $R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基,-$CO_2C_{1-4}$烷基,-$OR^k$,芳基或环烷基,其中$C_{1-6}$烷基,-$CO_2C_{1-4}$烷基,-$OR^k$,芳基和环烷基各自未被取代或被取代;和

[0397] $R^k$为H,$C_{1-6}$烷基或杂环基,其中$C_{1-6}$烷基和杂环基各自未被取代或被取代;

[0398] (v)式(V)的化合物:

[0399] 

(V)

[0400] 或其盐或缀合物,

[0401] 其中,

[0402] $R^8$是卤素或-$OR^m$;

[0403] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[0404] $R^9$是氢,卤素或$C_{1-6}$卤代烷基;

[0405] (vi)式(VI)的金属复合物:

[0406] $ML_n$(VI)

[0407] 或其盐或缀合物,

[0408] 其中,

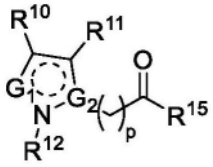
[0409] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[0410] L是选自由以下组成的群组的配体:-OH,-$OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[0411] n为1-8的整数,包括1和8;

[0412] 其中每个L可以相同或不同;和

[0413] (vii)式(VII)化合物:

[0414] 

(VII)

[0415] 或其盐或缀合物,

[0416] 其中,

[0417] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[0418] $G^2$是N或CH;

[0419] p是0或1;

[0420] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

[0421] $R^{15}$是H或OH,

[0422] (c)使多肽与包含能够结合官能化NTAA的第一结合部分的第一结合剂接触;

[0423] (c1)具有关于第一结合剂的识别信息的第一编码标签,或

[0424] (c2)第一可检测标签;

[0425] (d)(d1)将第一编码标签的信息转移到记录标签,以产生延伸记录标签并分析延伸记录标签,或者

[0426] (d2)检测第一可检测标签;

[0427] 其中步骤(b)在步骤(c)之前,步骤(c)之后和步骤(d)之前进行,或者在步骤(d)之后进行。

[0428] 在一些示例中,测序方法采用“埃德曼类”N-末端氨基酸降解方法。类埃德曼降解由两个关键步骤组成:1)肽的NTAA上的-胺的官能化,和2)官能化NTAA的消除。标准的爱德曼官能化化学以及本文所述的爱德曼类官能化化学表现出较差的官能化和N-末端脯氨酸残基的消除。因此,N-末端脯氨酸的存在可导致循环测序反应的“停顿”。因此,在本文所述方法的一些实施例中,通过将靶标多肽暴露于仅特异性裂解N-末端脯氨酸的脯氨酸氨肽酶(脯氨酸亚氨基肽酶),在每个埃德曼类降解循环开始时除去任何N-末端脯氨酸是有益的。因此,在一些实施例中,本文所述的每种方法和测定可任选地包括使被分析的多肽与脯氨酸氨肽酶接触的额外步骤。同样地,用于实施这些方法的试剂盒可以任选地包括至少一种脯氨酸氨肽酶。

[0429] 文献中已知几种脯氨酸氨肽酶(PAP)可用于此目的。在优选的实施例中,小单体PAP(约25-35kDa)用于去除NTAA脯氨酸。用于本文所述的方法和试剂盒的合适的单体PAP包括来自凝结芽孢杆菌,L.delbrueckii,淋病奈瑟菌,脑膜脓毒黄杆菌,粘质沙雷氏菌,以及植物乳杆菌(MEROPS S33.001)的家族成员(Nakajima,Ito等人,2006)(Kitazono,Yoshimoto等人,1992)。合适的多聚PAP也是已知的,并且包括来自D hansenii的酶(Bolumar,Sanz等人,2003年)。可以使用天然的或工程化的PAP。通过本文的方法和测定产生的没有脯氨酸残基的肽序列的有效映射可以通过将肽读回作图至“无脯氨酸”蛋白质组

来完成。在生物信息学水平上,这基本上转化为由19个氨基酸残基而不是20个氨基酸残基组成的蛋白质。

[0430] 或者,为了保留脯氨酸信息,可以在除去脯氨酸之前和之后采用两个结合步骤以能够检测脯氨酸残基,但这是以每个测序循环的额外结合/编码循环的额外成本来实现的。此外,将埃德曼类化学与R-基团特异性氨基肽酶组合的概念可用于去除任何NTF/NTE抗性氨基酸;然而,在优选的实施例中,只有单一的无效氨基残基,通常是脯氨酸,被氨肽酶除去。除去多个残基导致除去的序列的组合爆炸(即除去P和W导致除去具有运行P、运行W和运行P和W的序列)

[0431] 方面2.根据方面1所述的方法,其中步骤(a)包含提供多肽以及结合到支持物(例如,固体支持物)的相关记录标签。

[0432] 方面3.根据方面1所述的方法,其中步骤(a)包含提供在溶液中与相关记录标签连接的多肽。

[0433] 方面4.根据方面1所述的方法,其中步骤(a)包含提供与记录标签间接相关的多肽。

[0434] 方面5.根据方面1所述的方法,其中在步骤(a)中多肽不与记录标签相关。

[0435] 方面6.根据方面1-5中任一个所述的方法,其中步骤(b)在步骤(c)之前进行。

[0436] 方面7.根据方面1-5中任一个所述的方法,其中步骤(b)在步骤(c)之后且在步骤(d)之前进行。

[0437] 方面8.根据方面1-5中任一个所述的方法,其中步骤(b)在步骤(c)和步骤(d)之后进行。

[0438] 方面9.根据方面1-5中任一个所述的方法,其中步骤(a),(b),(c1),和(d1)按顺序发生。

[0439] 方面10.根据方面1-5中任一个所述的方法,其中步骤(a),(b),(c),和(d1)按顺序发生。

[0440] 方面11.根据方面1-5中任一个所述的方法,其中步骤(a),(c1),(d1),和(b)按顺序发生。

[0441] 方面12.根据方面1-5中任一个所述的方法,其中步骤(a),(b),(c2),和(d2)按顺序发生。

[0442] 方面13.根据方面1-5中任一个所述的方法,其中步骤(a),(c2),(b),和(d2)按顺序发生。

[0443] 方面14.根据方面1-5中任一个所述的方法,其中步骤(a),(c2),(d2),和(b)按顺序发生。

[0444] 方面15.根据方面1-14中任一个所述的方法,其中步骤(c)还包含使多肽与第二(或更高次序)结合剂接触,第二(或更高次序)结合剂包含能够结合不同于步骤(b)的官能化NTAA的官能化NTAA的第二(或更高次序)结合部分,以及具有关于第二(或更高次序)结合剂的识别信息的编码标签。

[0445] 方面16.根据方面15所述的方法,其中在多肽与第一结合剂接触后,接着使多肽与第二(或更高次序)结合剂接触。

[0446] 方面17.根据方面15所述的方法,其中多肽与第二(或更高次序)结合剂接触跟多

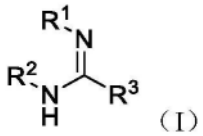
肽与第一结合剂接触同时发生。

[0447] 方面18.一种分析多肽的方法,包含以下步骤:

[0448] (a)提供任选地直接或间接与记录标签相关的多肽;

[0449] (b)用化学试剂官能化多肽的N-末端氨基酸(NTAA)以产生官能化NTAA,其中化学试剂包含选自由以下组成的群组的化合物:

[0450] (i)式(I)化合物:

[0451] 
(I)

[0452] 或其盐或缀合物,

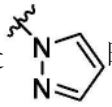
[0453] 其中,

[0454] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-$S(O)_2R^c$;

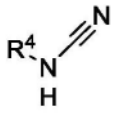
[0455] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[0456] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基是未取代的或取代的;

[0457] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[0458] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[0459] (ii)式(II)化合物:

[0460] 
(II)

[0461] 或其盐或缀合物,

[0462] 其中,

[0463] $R^4$是H,$C_{1-6}$烷基,环烷基,-C(O)$R^g$或-C(O)O$R^g$;和

[0464] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[0465] (iii)式(III)化合物:

[0466] R5-N=C=S (III)

[0467] 或其盐或缀合物,

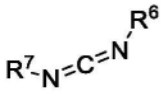
[0468] 其中,

[0469] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[0470] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自由选自卤素,-N$R^h$$R^i$,-$S(O)_2R^j$或杂环基组成的群组的一个或多个基团取代;

[0471] $R^h$,$R^i$和$R^j$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代。

[0472] (iv)式(IV)化合物:

[0473] 
（IV）

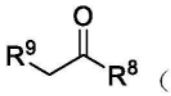
[0474] 或其盐或缀合物，

[0475] 其中，

[0476] $R^6$和$R^7$各自独立地为H，$C_{1-6}$烷基，$-CO_2C_{1-4}$烷基，$-OR^k$，芳基或环烷基，其中$C_{1-6}$烷基，$-CO_2C_{1-4}$烷基，$-OR^k$，芳基和环烷基各自未被取代或被取代；和

[0477] $R^k$为H，$C_{1-6}$烷基或杂环基，其中$C_{1-6}$烷基和杂环基各自未被取代或被取代；和

[0478] (v)式(V)的化合物：

[0479] 
（V）

[0480] 或其盐或缀合物，

[0481] 其中，

[0482] $R^8$是卤素或$-OR^m$；

[0483] $R^m$为H，$C_{1-6}$烷基或杂环基；和

[0484] $R^9$是氢，卤素或$C_{1-6}$卤代烷基；和

[0485] (vi)式(VI)的金属复合物：

[0486] $ML_n$(VI)

[0487] 或其盐或缀合物，

[0488] 其中，

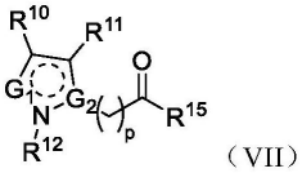
[0489] M是选自由以下组成的群组的金属：Co，Cu，Pd，Pt，Zn和Ni；

[0490] L是选自由以下组成的群组的配体：-OH，$-OH_2$，2，2’-二吡啶(bpy)，1，5-二硫代环辛烷(dtco)，1，2-双(二苯基膦)乙烷(dppe)，乙二胺(en)和三亚乙基四胺(trien)；和

[0491] n为1-8的整数，包括1和8；

[0492] 其中每个L可以相同或不同；和

[0493] (vii)式(VII)化合物：

[0494] 
（VII）

[0495] 或其盐或缀合物，

[0496] 其中，

[0497] $G^1$是N，$NR^{13}$或$CR^{13}R^{14}$；

[0498] $G^2$是N或CH；

[0499] p是0或1；

[0500] $R^{10}$，$R^{11}$，$R^{12}$，$R^{13}$和$R^{14}$各自独立地选自H，$C_{1-6}$烷基，$C_{1-6}$卤代烷基，$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺，其中$C_{1-6}$烷基，$C_{1-6}$卤代烷基，$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代，并且$R_{10}$和$R_{11}$可以任选地一起形成环；和

[0501] $R^{15}$是H或OH，

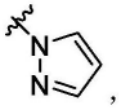
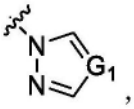
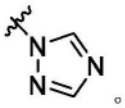
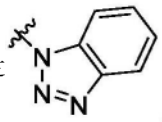
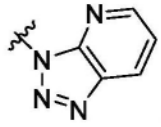
[0502] (c)使多肽与第一结合剂接触,第一结合剂包含能够结合官能化NTAA的第一结合部分和(c1)具有关于第一结合剂的识别信息的第一编码标签,或(c2)第一可检测标签;

[0503] (d)(d1)将第一编码标签的信息转移到记录标签,以产生第一延伸记录标签并分析延伸记录标签,或者

[0504] (d2)检测第一可检测标签,以及

[0505] (e)消除官能化NTAA以暴露新的NTAA;

[0506] 其中步骤(b)在步骤(c)之前,步骤(c)之后和步骤(d)之前进行,或者在步骤(d)之后进行。

[0507] 在用于本文公开的任何方法和试剂盒的式(I)的化合物的一些实施例中,$R^3$是单环杂芳基。在式(I)的一些实施例中,$R^3$是5或6元单环杂芳基。在式(I)的一些实施例中,$R^3$是含有一个或多个N的5或6元单环杂芳基。优选地,$R^3$选自吡唑,咪唑,三唑和四唑,并且经由吡唑,咪唑,三唑或四唑环的氮原子连接到式(I)的脒,并且$R^3$任选地被选自卤素,$C_{1-3}$烷基,$C_{1-3}$卤代烷基和硝基的基团取代。在一些实施例中,$R^3$为其中$G_1$为N,CH或CX,其中X为卤素,$C_{1-3}$烷基,$C_{1-3}$卤代烷基或硝基。在一些实施例中,$R^3$为  或,其中X为Me,F,Cl,$CF_3$或$NO_2$。在一些实施例中,$R^3$为  ,其中$G_1$为N或CH。在一些实施例中,$R^3$是  。在一些实施例中,$R^3$是双环杂芳基。在一些实施例中,R3为9或10元双环杂芳基。在某些实施例中,$R^3$是  或  。

[0508] 方面19.根据方面18所述的方法,其中步骤(a)包含提供多肽以及结合到支持物(例如,固体支持物)的相关记录标签。

[0509] 方面20.根据方面18所述的方法,其中步骤(a)包含提供在溶液中与相关记录标签连接的多肽。

[0510] 方面21.根据方面18所述的方法,其中步骤(a)包含提供与记录标签间接相关的多肽。

[0511] 方面22.根据方面18所述的方法,其中在步骤(a)中多肽不与记录标签相关。

[0512] 方面23.根据方面18-22中任一个所述的方法,其中步骤(b)在步骤(c)之前进行。

[0513] 方面24.根据方面18-22中任一个所述的方法,其中步骤(b)在步骤(c)之后且在步骤(d)之前进行。

[0514] 方面25.根据方面18-22中任一个所述的方法,其中步骤(b)在步骤(c)和步骤(d)之后进行。

[0515] 方面26.根据方面18-22中任一个所述的方法,其中步骤(a),(b),(c1)和(c2)按顺序发生。

[0516] 方面27.根据方面18-22中任一个所述的方法,其中步骤(a),(c1),(b),(d1)和(d1)按顺序发生。

[0517] 方面28.根据方面18-22中任一个所述的方法,其中步骤(a),(c1),(d1)和(b)按顺

序发生。

[0518] 方面29.根据方面18-22中任一个所述的方法,其中步骤(a),(b),(c2)和(d2)按顺序发生。

[0519] 方面30.根据方面18-22中任一个所述的方法,其中步骤(a),(c2),(b)和(d2)按顺序发生。

[0520] 方面31.根据方面18-22中任一个所述的方法,其中步骤(a),(c2),(d2)和(b)按顺序发生。

[0521] 方面32.根据方面18-31中任一个所述的方法,进一步包含以下步骤:

[0522] (f)用化学试剂官能化多肽的新NTAA以产生新官能化NTAA;

[0523] (g)使多肽与第二(或更高次序)结合剂接触,第二(或更高次序)结合剂包含能够结合新官能化NTAA的第二(或更高次序)结合基团和(g1)具有关于第二(或更高次序)结合剂的识别信息的第二编码标签,或(g2)第二可检测标签;

[0524] (h)(h1)将第二编码标签的信息转移到第一延伸记录标签,以生成第二延伸记录标签,并且分析第二延伸记录标签,或者

[0525] (h2)检测第二可检测标签,以及

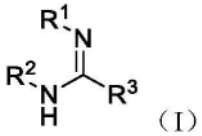
[0526] (i)移除修饰的NTAA以暴露新的NTAA;

[0527] 其中步骤(f)在步骤(g)之前,步骤(g)之后和步骤(h)之前进行,或者在步骤(h)之后进行。

[0528] 在方面1-32的任一个中,所述方法任选地进一步包括通常在所述方法步骤之前,在适于裂解存在的任何N-末端脯氨酸的条件下使多肽与脯氨酸氨肽酶接触的步骤。

[0529] 方面33.根据方面32所述的方法,其中化学试剂包含选自由以下组成的群组的化合物:

[0530] (i)式(I)化合物:

[0531]  (I)

[0532] 或其盐或缀合物,

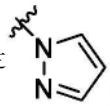
[0533] 其中,

[0534] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

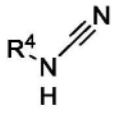
[0535] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[0536] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基是未取代的或取代的;

[0537] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[0538] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[0539] (ii)式(II)化合物:

[0540] 
（II）

[0541] 或其盐或缀合物，

[0542] 其中，

[0543] $R^4$是H，$C_{1-6}$烷基，环烷基，-C(O)$R^g$或-C(O)O$R^g$；和

[0544] $R^g$为H，$C_{1-6}$烷基，$C_{2-6}$烯基，$C_{1-6}$卤代烷基或芳基烷基，其中$C_{1-6}$烷基，$C_{2-6}$烯基，$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代；

[0545] (iii)式(III)化合物：

[0546] $R^5$-N＝C＝S (III)

[0547] 或其盐或缀合物，

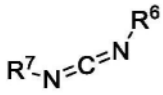
[0548] 其中，

[0549] $R^5$是$C_{1-6}$烷基，$C_{2-6}$烯基，环烷基，杂环烷基，芳基或杂芳基；

[0550] 其中$C_{1-6}$烷基，$C_{2-6}$烯基，环烷基，杂环烷基，芳基或杂芳基各自未被取代或由选自由选自卤素，-$NR^hR^i$，-S(O)$_2R^j$或杂环基组成的群组的一个或多个基团取代；

[0551] $R^h$，$R^i$和$R^j$各自独立地为H，$C_{1-6}$烷基，$C_{1-6}$卤代烷基，芳基烷基，芳基或杂芳基，其中$C_{1-6}$烷基，$C_{1-6}$卤代烷基，芳基烷基，芳基和杂芳基各自未被取代或被取代。

[0552] (iv)式(IV)化合物：

[0553] 
（IV）

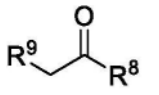
[0554] 或其盐或缀合物，

[0555] 其中，

[0556] $R^6$和$R^7$各自独立地为H，$C_{1-6}$烷基，-$CO_2C_{1-4}$烷基，-$OR^k$，芳基或环烷基，其中$C_{1-6}$烷基，-$CO_2C_{1-4}$烷基，-$OR^k$，芳基和环烷基各自未被取代或被取代；和

[0557] $R^k$为H，$C_{1-6}$烷基或杂环基，其中$C_{1-6}$烷基和杂环基各自未被取代或被取代；和

[0558] (v)式(V)的化合物：

[0559] 
（V）

[0560] 或其盐或缀合物，

[0561] 其中，

[0562] $R^8$是卤素或-$OR^m$；

[0563] $R^m$为H，$C_{1-6}$烷基或杂环基；和

[0564] $R^9$是氢，卤素或$C_{1-6}$卤代烷基；和

[0565] (vi)式(VI)的金属复合物：

[0566] $ML_n$(VI)

[0567] 或其盐或缀合物，

[0568] 其中，

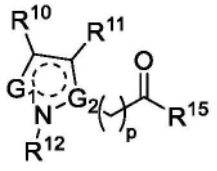
[0569] M是选自由以下组成的群组的金属：Co，Cu，Pd，Pt，Zn和Ni；

[0570] L是选自由以下组成的群组的配体:-OH,-$OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[0571] n为1-8的整数,包括1和8;

[0572] 其中每个L可以相同或不同;和

[0573] (vii)式(VII)化合物:

[0574] 

(VII)

[0575] 或其盐或缀合物,

[0576] 其中,

[0577] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[0578] $G^2$是N或CH;

[0579] p是0或1;

[0580] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

[0581] $R^{15}$为H或OH。

[0582] 方面34.根据方面32或方面33所述的方法,其中步骤(f)在步骤(g)之前进行。

[0583] 方面35.根据方面32或方面33所述的方法,其中步骤(f)在步骤(g)之后并且在步骤(h)之前进行。

[0584] 方面36.根据方面32或方面33所述的方法,其中步骤(f)在步骤(g)和步骤(h)之后进行。

[0585] 方面37.根据方面32或方面33所述的方法,其中步骤(f),(g1)和(h1)按顺序发生。

[0586] 方面38.根据方面32或方面33所述的方法,其中步骤(g1),(f)和(h1)按顺序发生。

[0587] 方面39.根据方面32或方面33所述的方法,其中步骤(g1),(h1)和(f)按顺序发生。

[0588] 方面40.根据方面32或方面33所述的方法,其中步骤(f),(g2)和(h2)按顺序发生。

[0589] 方面41.根据方面32或方面33所述的方法,其中步骤(g2),(f)和(h2)按顺序发生。

[0590] 方面42.根据方面32或方面33所述的方法,其中步骤(g2),(h2)和(f)按顺序发生。

[0591] 方面43.根据方面32-42所述的方法,其中在肽与第一结合剂接触后,接着使肽与第二(或更高次序)结合剂接触。

[0592] 方面44.根据方面32-42所述的方法,其中多肽与第二(或更高次序)结合剂接触跟多肽与第一结合剂接触同时发生。

[0593] 方面45.根据权利要求1-44任一所述的方法,其中多肽通过片段化来自生物样品的蛋白质而获得。

[0594] 方面46.根据方面1-45中任一个所述的方法,其中记录标签包含核酸、寡核苷酸、修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、具有受保护碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代DNA,或其组合。

[0595] 方面47.根据方面46所述的方法,其中DNA分子是骨架修饰的、糖修饰的或核碱基

修饰的。

[0596] 方面48.根据方面46所述的方法,其中DNA分子具有核碱基保护基,如Alloc,亲电子保护基,如thiranes,乙酰基保护基,硝基苄基保护基,磺酸酯保护基,或传统的碱不稳定保护基,包括Ultramild试剂。

[0597] 方面49.根据方面1-48中任一个所述的方法,其中记录标签包含通用引发位点。

[0598] 方面50.根据方面49所述的方法,其中通用引发位点包含用于扩增、测序或二者的引物位点。

[0599] 方面51.根据方面1-50中任一个所述的方法,其中记录标签包含独特分子标识符(UMI)。

[0600] 方面52.根据方面1-51中任一个所述的方法,其中记录标签包含条形码。

[0601] 方面53.根据方面1-52中任一个所述的方法,其中记录标签在其3'-末端包含间隔区。

[0602] 方面54.根据方面1-53中任一个所述的方法,其中多肽和相关的记录标签共价结合到固体支持物上。

[0603] 方面55.根据方面1-54任一所述的方法,支持物是珠粒、多孔珠粒、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。

[0604] 方面56.根据方面55所述的方法,其中支持物包含金、银、半导体或量子点。

[0605] 方面57.根据方面55所述的方法,其中纳米粒子包含金、银或量子点。

[0606] 方面58.根据方面55所述的方法,其中支持物是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[0607] 方面59.根据方面1-58中任一个所述的方法,其中多肽和相关的记录标签连接到固体支持物上。

[0608] 方面60.根据方面59所述的方法,其中多种多肽在支持物上间隔开,其中多肽之间的平均距离为约≥20nm。

[0609] 方面61.根据方面1-60中任一个所述的方法,其中结合剂的结合部分包含肽或蛋白质。

[0610] 方面62.根据方面1-61中任一个所述的方法,其中结合剂的结合部分包含氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰的蛋白质;anticalin或其变体,突变体或修饰的蛋白质;Clps(例如ClpS2)或其变体,突变体或修饰的蛋白质;UBR盒蛋白质或其变体,突变体或修饰的蛋白质;或结合氨基酸的修饰的小分子,即万古霉素或其变体,突变体或修饰的分子;或抗体或其结合片段;或其任何组合。

[0611] 方面63.根据方面1-62中任一个所述的方法,其中结合剂结合到单个氨基酸残基(例如,N-末端氨基酸残基,C-末端氨基酸残基,或内部氨基酸残基),二肽(例如,N-末端二肽,C-末端二肽,或内部二肽),三肽(例如,N-末端三肽,C-末端三肽,或内部三肽),或翻译后修饰的多肽。

[0612] 方面64.根据方面1-62中任一个所述的方法,其中结合剂结合NTAA官能化的单个氨基酸残基,NTAA官能化的二肽,NTAA官能化的三肽或NTAA官能化的多肽。

[0613] 方面65.根据方面1-64中任一个所述的方法,其中结合剂的结合部分能够选择性地结合多肽。

[0614] 方面66.根据方面1-65中任一个所述的方法,其中编码标签是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。

[0615] 方面67.根据方面1-66中任一个所述的方法,其中编码标签包含编码器或条形码序列。

[0616] 方面68.根据方面1-67中任一个所述的方法,其中编码标签还包含间隔区、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。

[0617] 方面69.根据方面1-68中任一个所述的方法,其中结合部分和编码标签通过接头连接。

[0618] 方面70.根据方面1-69所述的方法,其中结合部分和编码标签通过SpyTag/SpyCatcher肽-蛋白质对,SnoopTag/SnoopCatcher肽-蛋白质对或HaloTag/HaloTag配体对连接。

[0619] 方面71.根据方面1-70中任一个所述的方法,其中记录标签到编码标签的信息转移通过DNA连接酶或RNA连接酶介导。

[0620] 方面72.根据方面1-70中任一个所述的方法,其中编码标签到记录标签的信息转移通过DNA聚合酶、RNA聚合酶或逆转录酶介导。

[0621] 方面73.根据方面1-70中任一个所述的方法,其中记录标签到编码标签的信息转移通过化学连接介导。

[0622] 方面74.根据方面73所述的方法,其中使用单链DNA进行化学连接。

[0623] 75.根据方面74所述的方法,其中使用双链DNA进行化学连接。

[0624] 76.根据方面1-72中任一个所述的方法,分析延伸记录标签包含核酸测序方法。

[0625] 77.根据方面76所述的方法,其中核酸测序方法是通过合成测序、通过连接测序、通过杂交测序、polony测序、离子半导体测序或焦磷酸测序。

[0626] 78.根据方面76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。

[0627] 79.根据方面1-78中任一个所述的方法,其中延伸记录标签在分析前扩增。

[0628] 80.根据方面1-79中任一个所述的方法,还包含添加循环标记的步骤。

[0629] 81.根据方面80所述的方法,其中循环标记提供关于结合剂与多肽结合的次序的信息。

[0630] 82.根据方面80或方面81所述的方法,其中将循环标记添加到编码标签。

[0631] 83.根据方面80或方面81所述的方法,其中将循环标记添加到记录标签。

[0632] 84.根据方面80或方面81所述的方法,其中将循环标记添加到结合剂。

[0633] 85.根据方面80或方面81所述的方法,其中独立于编码标签、记录标签和结合剂添加循环标记。

[0634] 86.根据方面1-85中任一个所述的方法,其中包含于延伸记录标签中的编码标签信息的次序提供关于结合到大分子的结合剂的结合次序的信息。

[0635] 87.根据方面1-86中任一个所述的方法,其中包含于延伸记录标签中的编码标签信息的频率提供关于结合到多肽的结合剂的结合频率的信息。

[0636] 88.根据方面1-87中任一个所述的方法,其中平行分析代表多个多肽的多个延伸记录标签。

[0637] 89.根据方面88所述的方法,其中在多路复用测定中分析代表多个多肽的多个延伸记录标签。

[0638] 90.根据方面88或89中任一个所述的方法,其中多个延伸记录标签在分析之前进行靶标丰度检测。

[0639] 91.根据方面88-90中任一个所述的方法,其中多个延伸记录标签在分析之前进行扣除测定。

[0640] 92.根据方面88-91中任一个所述的方法,其中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。

[0641] 93.根据方面1-92中任一个所述的方法,其中通过化学裂解或酶裂解从多肽消除NTAA。

[0642] 94.根据方面93所述的方法,其中NTAA被羧肽酶或氨肽酶或其变体,突变体或修饰的蛋白质;水解酶或其变体,突变体或修饰的蛋白质;温和型埃德曼降解;Edmanase酶;TFA,碱;或其任何组合消除。

[0643] 95.根据方面94所述的方法,其中温和型埃德曼降解使用二氯或单氯酸。

[0644] 96.根据方面94所述的方法,其中温和型埃德曼降解使用TFA、TCA或DCA。

[0645] 97.根据方面94所述的方法,其中温和型埃德曼降解使用三乙基乙酸铵。

[0646] 98.根据方面94所述的方法,其中碱是氢氧化物、烷基化胺、环胺、碳酸盐缓冲剂或金属盐。

[0647] 99.根据方面98所述的方法,其中氢氧化物是氢氧化钠。

[0648] 100.根据方面98所述的方法,其中烷基化胺选自甲胺,乙胺,丙胺,二甲胺,二乙胺,二丙胺,三甲胺,三乙胺,三丙胺,环己胺,苄胺,苯胺,二苯胺,N,N-二异丙基乙胺(DIPEA)和二异丙基胺基锂(LDA)。

[0649] 101.根据方面98所述的方法,其中环胺选自吡啶,嘧啶,咪唑,吡咯,吲哚,哌啶,吡咯烷,1,8-二氮杂双环[5.4.0]十一-7-烯(DBU)和1,5-二氮杂双环[4.3.0]壬-5-烯(DBN)。

[0650] 102.根据方面98所述的方法,其中碳酸盐缓冲液包含碳酸钠,碳酸钾,碳酸钙,碳酸氢钠,碳酸氢钾或碳酸氢钙。

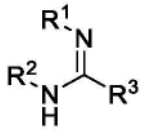
[0651] 103.根据方面98所述的方法,其中金属盐包含银。

[0652] 104.根据方面103所述的方法,其中金属盐是$AgClO_4$。

[0653] 105.根据方面1-104中任一个所述的方法,其中至少一种结合剂结合至末端氨基酸残基、末端二氨基酸残基或末端三氨基酸残基。

[0654] 106.根据方面1-105中任一个所述的方法,其中至少一种结合剂结合到翻译后修饰的氨基酸。

[0655] 107.根据方面1-106中任一个所述的方法,其中化学试剂包含选自由式(IV)的化合物组成的群组的化合物:

[0656] 

(I)

[0657] 或其盐或缀合物，

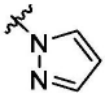
[0658] 其中，

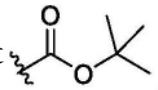
[0659] $R^1$和$R^2$各自独立地为H，$C_{1-6}$烷基，环烷基，$-C(O)R^a$，$-C(O)OR^b$或$-S(O)_2R^c$；

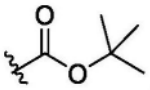
[0660] $R^a$，$R^b$和$R^c$各自独立地为H，$C_{1-6}$烷基，$C_{1-6}$卤代烷基，芳基烷基，芳基或杂芳基，其中$C_{1-6}$烷基，$C_{1-6}$卤代烷基，芳基烷基，芳基和杂芳基各自未被取代或被取代；

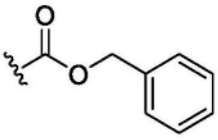
[0661] $R^3$是杂芳基，$-NR^dC(O)OR^e$或$-SR^f$，其中杂芳基是未取代的或取代的；

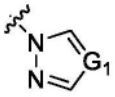
[0662] $R^d$，$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基；和

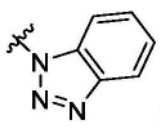
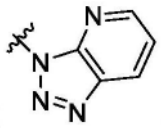
[0663] 任选地其中当$R_3$是  时，$R^1$和$R^2$不都是H。

[0664] 108.根据方面107所述的方法，其中$R^1$是  。

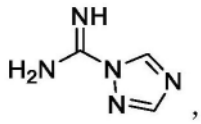
[0665] 109.根据方面107或108的方法，其中$R^2$是  。

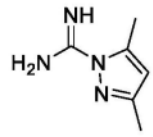
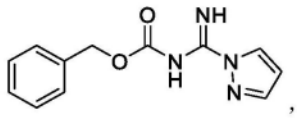
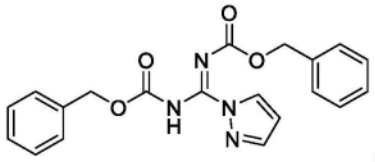
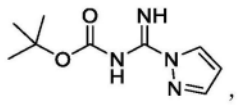
[0666] 110.根据方面107-109中任一个所述的方法，其中$R^1$或$R^2$为  。

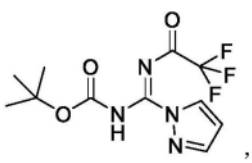
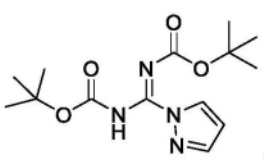
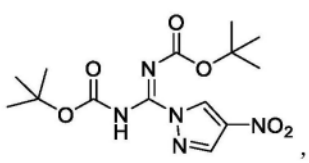
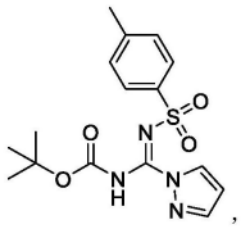
[0667] 111.根据方面107-110中任一个所述的方法，其中$R^3$是  ，其中$G_1$是N或CH。

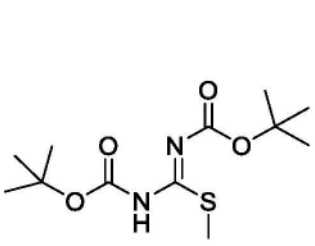
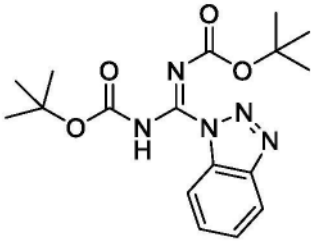
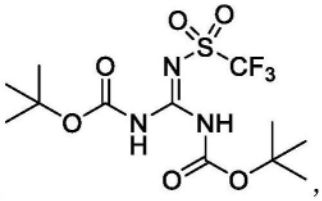
[0668] 112.根据权利要求107-110中任一个所述的方法，其中$R^3$是  或  。

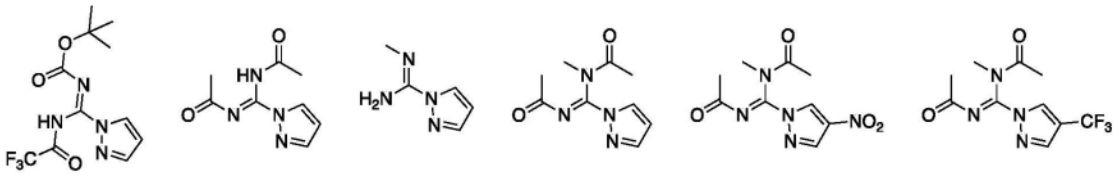
[0669] 113.根据方面107所述的方法，其中式(I)的化合物选自由以下组成的群组：

 ，

 ，  ，  ，  ，

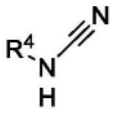
 ，  ，  ，  ，


， 
，以及 
，并且任选地进一步包括选自以下的化合物：


(N-Boc，N'-三氟乙酰基-吡唑甲脒,N,N'-双乙酰基-吡唑甲脒,N-甲基-吡唑甲脒,N,N'-双乙酰基-N-甲基-吡唑甲脒,N,N'-双乙酰基-N-甲基-吡唑甲脒,N,N'-双乙酰基-N-甲基-4-硝基-吡唑甲脒,和N,N'-双乙酰基-N-甲基-4-三氟甲基-吡唑甲脒,或其盐或缀合物。

[0670] 114.根据方面107-113中任一个所述的方法,其中化学试剂另外包含Mukaiyama试剂(2-氯-1-甲基吡啶碘化物)。

[0671] 115.根据方面1-106中任一个所述的方法,其中化学试剂包含选自由式(II)的化合物组成的群组的化合物：

[0672] 
(II)

[0673] 或其盐或缀合物,

[0674] 其中,

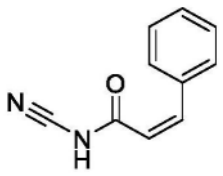
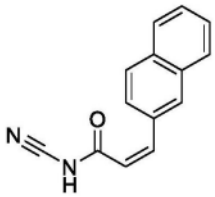
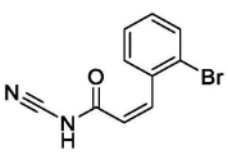
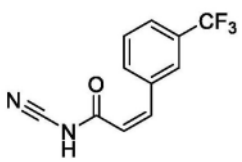
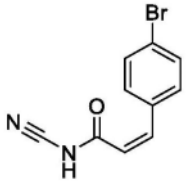
[0675] $R^4$是H,$C_{1-6}$烷基,环烷基,-C(O)$R^g$或-C(O)O$R^g$;和

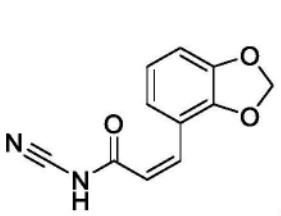
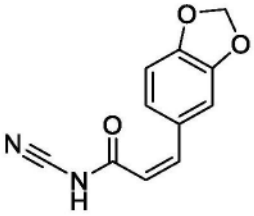
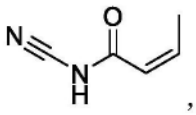
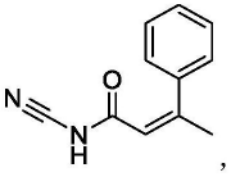
[0676] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[0677] 116.根据方面115所述的方法,其中$R^4$是羧基苄基。

[0678] 117.根据方面115所述的方法,其中$R^4$是-C(O)$R^g$,并且$R^g$是$C_{2-6}$烯基,其任选地被芳基、杂芳基或杂环烷基取代。

[0679] 118.根据方面115所述的方法,其中化合物选自由以下组成的群组：


， 
， 
， 
， 
，

[0680]  ，  ，  ，以及  ，

或其盐或缀合物。

[0681] 119.根据方面115-118中任一个所述的方法，其中化学试剂另外包含TMS-Cl，$Sc(OTf)_2$，$Zn(OTf)_2$或含镧系元素的试剂。

[0682] 120.根据方面1-106中任一个所述的方法，其中化学试剂包含自由式(III)的化合物组成的群组的化合物：

[0683] $R^5$-N＝C＝S (III)

[0684] 或其盐或缀合物，

[0685] 其中，

[0686] $R^5$是$C_{1-6}$烷基，$C_{2-6}$烯基，环烷基，杂环烷基，芳基或杂芳基；

[0687] 其中$C_{1-6}$烷基，$C_{2-6}$烯基，环烷基，杂环烷基，芳基或杂芳基各自未被取代或由选自由选自卤素，$-NR^hR^i$，$-S(O)_2R^j$或杂环基组成的群组的一个或多个基团取代；

[0688] $R^h$，$R^i$和$R^j$各自独立地为H，$C_{1-6}$烷基，$C_{1-6}$卤代烷基，芳基烷基，芳基或杂芳基，其中$C_{1-6}$烷基，$C_{1-6}$卤代烷基，芳基烷基，芳基和杂芳基各自未被取代或被取代；

[0689] 121.根据方面120所述的方法，其中$R^5$是取代的苯基。

[0690] 122.根据方面120或方面121所述的方法，其中$R^5$是苯基，其被一个或多个选自由卤素、$-NR^hR^i$、$-S(O)_2R^j$和杂环基组成的群组的基团取代。

[0691] 123.根据方面120所述的方法，其中式(III)的化合物是三甲基硅烷基异硫氰酸酯(TMSITC)或五氟苯基异硫氰酸酯(PFPITC)。

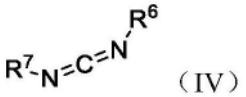
[0692] 124.根据方面120-123中任一个所述的方法，其中化学试剂另外包含碳二亚胺化合物。

[0693] 125.根据方面120-124中任一个所述的方法，其中使用三氟乙酸或盐酸消除NTAA。

[0694] 126.根据方面120-124中任一个所述的方法，其中使用温和型埃德曼降解消除NTAA。

[0695] 127.根据方面120-124中任一个所述的方法，其中使用Edmanase或工程化水解酶、氨肽酶或羧肽酶消除NTAA。

[0696] 128.根据方面1-106中任一个所述的方法，其中化学试剂包含选自由式(IV)的化合物组成的群组的的化合物：

[0697]  （IV）

[0698] 或其盐或缀合物，

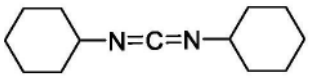
[0699] 其中，

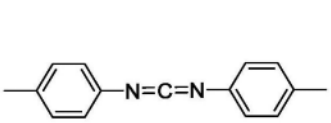
[0700] $R^6$和$R^7$各自独立地为H，$C_{1-6}$烷基，$-CO_2C_{1-4}$烷基，$-OR^k$，芳基或环烷基，其中$C_{1-6}$烷基，$-CO_2C_{1-4}$烷基，$-OR^k$，芳基和环烷基各自未被取代或被取代；和

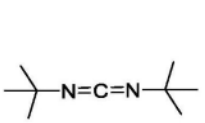
[0701] $R^k$为H，$C_{1-6}$烷基或杂环基，其中$C_{1-6}$烷基和杂环基各自未被取代或被取代。

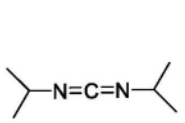
[0702] 129.根据方面128所述的方法,其中$R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基或环烷基。

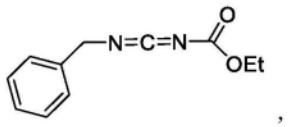
[0703] 130.根据方面128所述的方法,其中化合物选自由以下组成的群组:

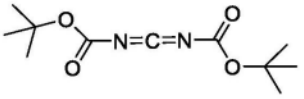

,

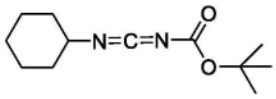

,

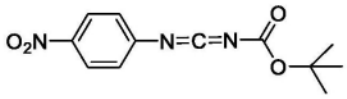

,

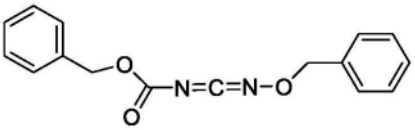

,

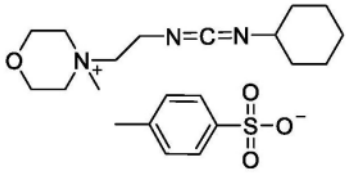

,

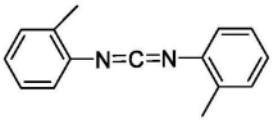

,

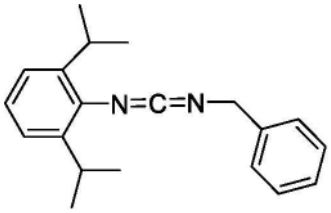

,


,


,


,


以及


或其盐或缀合物。

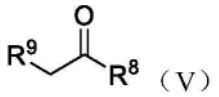
[0704] 131.根据方面128-130中任一个所述的方法,其中式(IV)的化合物通过相应硫脲的脱硫制备。

[0705] 132.根据方面128-131中任一个所述的方法,其中化学试剂另外包含Mukaiyama试剂(2-氯-1-甲基吡啶碘化物)。

[0706] 133.根据方面128-131中任一个所述的方法,其中化学试剂另外包含路易斯酸。

[0707] 134.根据方面133所述的方法,其中路易斯酸选自N-((芳基)亚氨基-乙酰萘酮)$ZnCl_2$,$Zn(OTf)_2$,$ZnCl_2$,$PdCl_2$,CuCl和$CuCl_2$。

[0708] 135.根据方面1-106中任一个所述的方法,其中化学试剂包含选自由式(V)的化合物组成的群组的的化合物:

[0709] 
(V)

[0710] 或其盐或缀合物,

[0711] 其中,

[0712] $R^8$是卤素或$-OR^m$;

[0713] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[0714] $R^9$是氢、卤素或$C_{1-6}$卤代烷基。

[0715] 136.根据方面135所述的方法,其中$R^8$是氯。

[0716] 137.根据方面135或方面136的方法,其中$R^9$是氢或溴代。

[0717] 138.根据方面135-137中任一个所述的方法,其中化学试剂另外包含肽偶联试剂。

[0718] 139.根据方面138所述的方法,其中肽偶联试剂是碳二亚胺化合物。

[0719] 140.根据方面139所述的方法,其中碳二亚胺化合物是二异丙基碳二亚胺(DIC)或1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)。

[0720] 141.根据方面135-140中任一个所述的方法,其中使用酰基肽水解酶(APH)消除NTAA。

[0721] 142.根据方面1-106中任一个所述的方法,其中化学试剂包含式(VI)的金属复合物:

[0722] $ML_n$(VI)

[0723] 或其盐或缀合物,

[0724] 其中,

[0725] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[0726] L是选自由以下组成的群组的配体:-OH,$-OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[0727] n为1-8的整数,包括1和8;

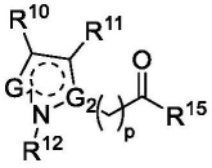
[0728] 其中每个L可以相同或不同。

[0729] 143.根据方面142所述的方法,其中M是Co。

[0730] 144.根据方面142或方面143所述的方法,其中化学试剂包含顺式-β-羟基(三亚乙基四胺)水合钴(III)复合物。

[0731] 145.根据方面142所述的方法,其中化学试剂包含β-$[Co(trien)(OH)(OH_2)]^{2+}$。

[0732] 146.根据方面1-106中任一个所述的方法,其中化学试剂包含选自由式(VII)的化合物组成的群组的化合物:

[0733] 

(VII)

[0734] 或其盐或缀合物,

[0735] 其中,

[0736] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[0737] $G^2$是N或CH;

[0738] p是0或1;

[0739] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

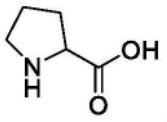
[0740] $R^{15}$为H或OH。

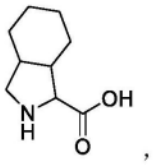
[0741] 147.根据方面146所述的方法,其中$G^1$是$CH_2$且$G^2$是CH,或$G^1$是NH且$G^2$是N。

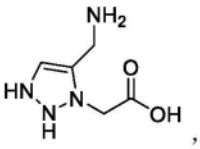
[0742] 148.根据方面146或147所述的方法,其中$R^{12}$是H。

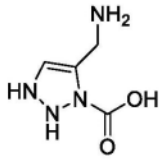
[0743] 149.根据方面146-148中任一个所述的方法,其中$R^{10}$和$R^{11}$各自为H。

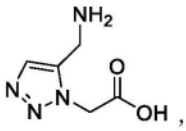
[0744] 150.根据方面146所述的方法,其中式(VII)的化合物选自由以下组成的群组:

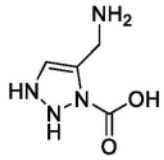

，

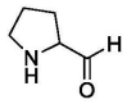

，

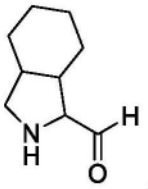

，

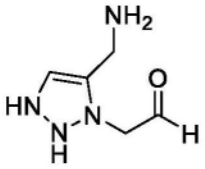

，

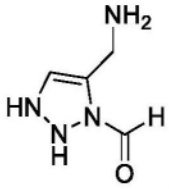

，

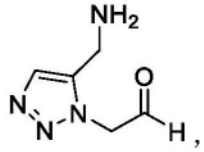

，

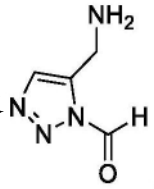

，


，


，


，


，以及


，或其盐或缀合物。

[0745] 151.根据方面107-150中任一个所述的方法，其中使用碱消除NTAA。

[0746] 152.根据方面151所述的方法，其中碱是氢氧化物、烷基化胺、环胺、碳酸盐缓冲剂或金属盐。

[0747] 153.根据方面152所述的方法，其中氢氧化物是氢氧化钠。

[0748] 154.根据方面152所述的方法，其中烷基化胺选自甲胺，乙胺，丙胺，二甲胺，二乙胺，二丙胺，三甲胺，三乙胺，三丙胺，环己胺，苄胺，苯胺，二苯胺，N,N-二异丙基乙胺(DIPEA)，和二异丙基胺基锂(LDA)。

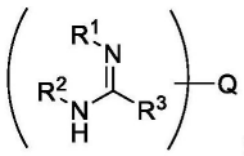
[0749] 155.根据方面152所述的方法，其中环胺选自吡啶，嘧啶，咪唑，吡咯，吲哚，哌啶，吡咯烷，1,8-二氮杂双环[5.4.0]十一-7-烯(DBU)和1,5-二氮杂双环[4.3.0]壬-5-烯(DBN)。

[0750] 156.根据方面152所述的方法，其中碳酸盐缓冲液包含碳酸钠，碳酸钾，碳酸钙，碳酸氢钠，碳酸氢钾或碳酸氢钙。

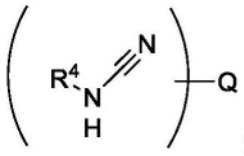
[0751] 157.根据方面152所述的方法，其中金属盐包含银。

[0752] 158.根据方面152所述的方法，其中金属盐是$AgClO_4$。

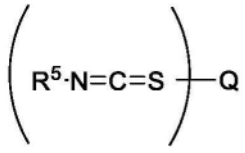
[0753] 159.根据方面1-158中任一个所述的方法，其中化学试剂包含选自由以下组成的群组的缀合物：

[0754] 
式（I）-Q，

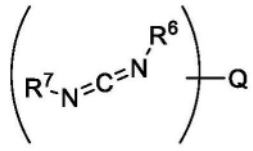
[0755] 其中$R^1$，$R^2$和$R^3$如方面1中对于式(I)所定义，并且Q是配体；

[0756] 
式（II）-Q，

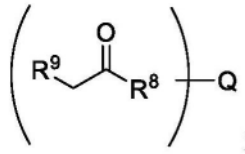
[0757] 其中$R^4$如对方面1中的式(II)所定义，并且Q是配体；

[0758] 
式（III）-Q，

[0759] 其中$R^5$如对方面1中的式(III)所定义，并且Q是配体；

[0760] 
，式（IV）-Q，

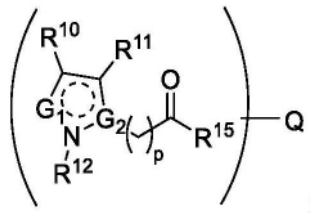
[0761] 其中$R^6$和$R^7$如方面1中对于式(IV)所定义，并且Q是配体；

[0762] 
式（V）-Q，

[0763] 其中$R^8$和$R^9$如方面1中对于式(V)所定义，并且Q是配体；

[0764] $(ML_n)$-Q式(VI)-Q，

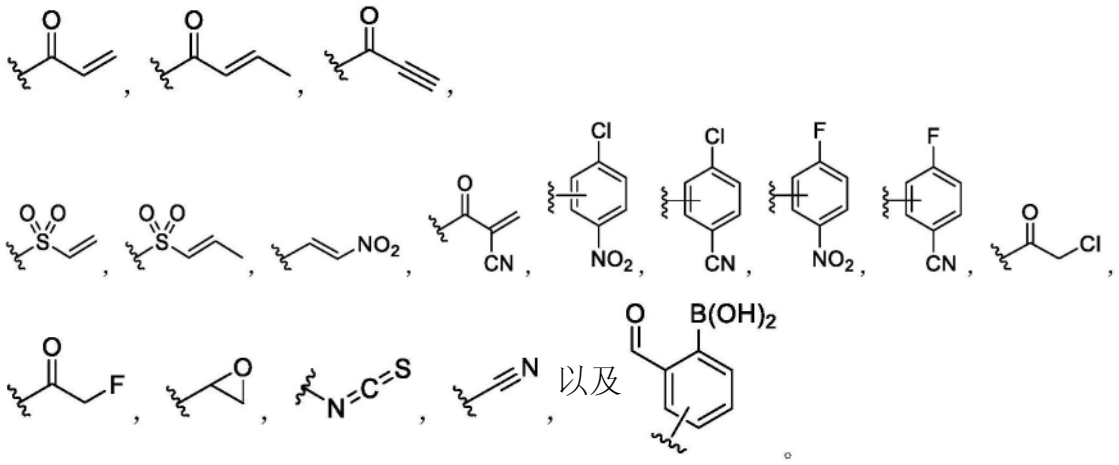
[0765] 其中M、L和N如方面1中对于式(V)所定义，并且Q是配体；

[0766] 
式（VII）-Q，

[0767] 其中$R^{10}$，$R^{11}$，$R^{12}$，$R^{15}$，$G^1$，$G^2$和p如方面1中对于式(VII)所定义，并且Q是配体。

[0768] 160.根据方面159所述的方法，其中Q选自-$C_{1-6}$烷基，-$C_{2-6}$烯基，-$C_{2-6}$炔基，芳基，杂芳基，杂环基，-N＝C＝S，-CN，-C(O)$R^n$，-C(O)O$R^o$，--S$R^p$或-S(O)$_2R^q$；其中-$C_{1-6}$烷基，-$C_{2-6}$烯基，-$C_{2-6}$炔基，芳基，杂芳基和杂环基各自未被取代或被取代，并且$R^n$，$R^o$，$R^p$和$R^q$各自独立地选自-$C_{1-6}$烷基，-$C_{1-6}$卤代烷基，-$C_{2-6}$烯基，-$C_{2-6}$炔基，芳基，杂芳基和杂环基。

[0769] 161.根据方面159所述的方法，其中Q选自由以下组成的群组：


以及

。

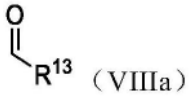
[0770] 162.根据方面159所述的方法，其中Q是荧光团。

[0771] 163.根据方面1-162中任一个所述的方法，其中方面1或方面5所述的步骤(b)包含

NTAA的双官能化。

[0772] 164.根据方面163所述的方法,其中方面1或方面5所述的步骤(b)包含用(1)第一化学试剂和(2)第二化学试剂来双官能化NTAA,第一化学试剂包含选自由以下组成的群组的化合物:式(I),(II),(III),(IV),(V),(VI)和(VII)的化合物,或其盐或缀合物。

[0773] 165.根据方面164所述的方法,其中第二化学试剂包含式(VIIIa)或(VIIIb)的化合物:

[0774]  (VIIIa)

[0775] 或其盐或缀合物,

[0776] 其中,

[0777] $R^{13}$是H,$C_{1-6}$烷基,芳基,杂芳基,环烷基或杂环烷基,其中$C_{1-6}$烷基,芳基,杂芳基,环烷基和杂环烷基各自未被取代或被取代;或

[0778] $R^{13}$-X (VIIIb)

[0779] 其中,

[0780] $R^{13}$为$C_{1-6}$烷基,芳基,杂芳基,环烷基或杂环烷基,其各自未被取代或被取代;和

[0781] X是卤素。

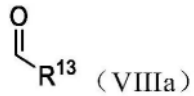
[0782] 166.根据方面164或165所述的方法,其中方面1或方面5所述的步骤(b)包含在用第一化学试剂官能化之前用第二化学试剂官能化NTAA。

[0783] 167.根据方面164或165所述的方法,其中方面1或方面5所述的步骤(b)包含在用第二化学试剂官能化之前用第一化学试剂官能化NTAA。

[0784] 168.根据方面32-167中任一个所述的方法,其中方面6所述的步骤(f)包含NTAA的双官能化。

[0785] 169.根据方面168所述的方法,其中方面6所述的步骤(f)包含用(1)第一化学试剂和(2)第二化学试剂来双官能化NTAA,第一化学试剂包含选自由以下组成的群组的化合物:式(I),(II),(III),(IV),(V),(VI),(VII)和(VIII)的化合物。

[0786] 170.根据方面169所述的方法,其中第二化学试剂包含式(VIIIa)或(VIIIb)的化合物:

[0787]  (VIIIa)

[0788] 或其盐或缀合物,

[0789] 其中,

[0790] $R^{13}$是H,$C_{1-6}$烷基,芳基,杂芳基,环烷基或杂环烷基,其中$C_{1-6}$烷基,芳基,杂芳基,环烷基和杂环烷基各自是未取代的或取代的;或

[0791] $R^{13}$-X (VIIIb)

[0792] 其中,

[0793] $R^{13}$为$C_{1-6}$烷基,芳基,杂芳基,环烷基或杂环烷基,其各自未被取代或被取代;和

[0794] X是卤素。

[0795] 171.根据方面169或170所述的方法,其中方面1或方面5所述的步骤(b)包含在用

第一化学试剂官能化之前用第二化学试剂官能化NTAA。

[0796] 172.根据方面169或170所述的方法,其中方面1或方面5所述的步骤(b)包含在用第二化学试剂官能化之前用第一化学试剂官能化NTAA。

[0797] 173.根据方面165-167和170-172任一个所述的方法,其中式(VIIIa)化合物是甲醛。

[0798] 174.根据方面165-167和170-172中任一个所述的方法,其中式(VIIIb)化合物是碘甲烷。

[0799] 175.一种分析来自于包含多个蛋白质复合物、蛋白质或多肽的样品的一个或多个多肽的方法,所述方法包含:

[0800] (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,其中每个隔室包含多个隔室标签,其任选地连接到支持物(例如,固体支持物),其中在单个隔室内的多个隔室标签是相同的,并且不同于其他隔室的隔室标签;

[0801] (b)将多个蛋白质复合物、蛋白质和/或多肽片段化成多个多肽;

[0802] (c)在足以允许多个多肽与多个隔室内的多个隔室标签退火或连接的条件下,使多个多肽与多个隔室标签接触,从而产生多个隔室标记的多肽;

[0803] (d)从多个隔室收集隔室标记的多肽;并且

[0804] (e)根据方面-174中任一个所述的方法分析一个或多个隔室标记的多肽。

[0805] 176.根据方面175所述的方法,隔室是微流体液滴。

[0806] 177.根据方面175所述的方法,隔室是微孔。

[0807] 178.根据方面175所述的方法,隔室是平面上的分离区域。

[0808] 179.根据方面175-178中任一个所述的方法,其中每一个隔室平均包含单独的细胞。

[0809] 180.一种分析来自于包含多个蛋白质复合物、蛋白质或多肽的样品的一个或多个多肽的方法,所述方法包含:

[0810] (a)用多个通用DNA标签标记多个蛋白质复合物、蛋白质或多肽;

[0811] (b)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,其中在单个隔室内的多个隔室标签是相同的,并且不同于其他隔室的隔室标签;

[0812] (c)在足以允许多个肽与多个隔室内的多个隔室标签退火或连接的条件下,使多个蛋白质复合物、蛋白质或多肽与多个隔室标签接触,从而产生多个隔室标记的蛋白质复合物、蛋白质或多肽;

[0813] (d)从多个隔室收集隔室标记的蛋白质复合物、蛋白质或多肽;

[0814] (e)任选地将隔室标记的蛋白质复合物、蛋白质或多肽片段化成隔室标记的多肽;和

[0815] (f)根据方面1-174中任一个所述的方法分析一个或多个隔室标记的多肽。

[0816] 181.根据方面175-180中任一个所述的方法,其中隔室标签信息通过引物延伸或连接转移到与多肽相关的记录标签。

[0817] 182.根据方面1-181中任一个所述的方法,所述固体支持物是珠粒、多孔珠粒、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。

[0818] 183.根据方面175-181中任一个所述的方法,其中固体支持物是珠粒。

[0819] 184.根据方面183所述的方法,其中珠粒是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[0820] 185.根据方面175-184中任一个所述的方法,其中隔室标签包含单链或双链核酸分子。

[0821] 186.根据方面175-185中任一个所述的方法,其中隔室标签包含条形码和可选的UMI。

[0822] 187.根据方面186所述的方法,其中固体支持物是珠粒,并且隔室标签包含条形码,进一步其中包含多个隔室标签连接到其上的珠粒通过均分与合并合成来形成。

[0823] 188.根据方面186所述的方法,其中固体支持物是珠粒,并且隔室标签包含条形码,进一步其中包含多个隔室标签连接到其上的珠粒通过单个合成或固定化形成。

[0824] 189.根据方面175-188中任一个所述的方法,其中隔室标签是记录标签中的成分,记录标签任选地还包含间隔区、条形码序列、独特分子标识符、通用引发位点或它们的任何组合。

[0825] 190.根据权利要求175-189中任一个所述的方法,其中隔室标签还包含能够与多个蛋白质复合物、蛋白质或多肽的内部氨基酸、肽骨架或N-末端氨基酸反应的官能部分。

[0826] 191.根据方面190所述的方法,其中官能部分是醛,叠氮化物/炔,或马来酰亚胺/硫醇,或环氧化物/亲核试剂,或逆电子需求Diels-Alder(iEDDA)基团。

[0827] 192.根据方面190所述的方法,其中官能部分是醛基。

[0828] 193.根据方面175-192中任一个所述的方法,其中多个隔室标签通过以下步骤形成:打印、点样、将隔室标签喷墨打印到隔室中,或它们的组合。

[0829] 194.根据方面175-193中任一个所述的方法,其中隔室标签还包含多肽。

[0830] 195.根据方面194所述的方法,其中隔室标签多肽还包含蛋白连接酶识别序列。

[0831] 196.根据方面195所述的方法,蛋白连接酶是butelase I或其同源物。

[0832] 197.根据方面175-196中任一个所述的方法,其中多个多肽采用蛋白酶片段化。

[0833] 198.根据方面197所述的方法,蛋白酶是金属蛋白酶。

[0834] 199.根据方面198所述的方法,其中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。

[0835] 200.根据方面175-199中任一个所述的方法,还包含在将多个多肽分区成多个隔室之前,从样品中扣除一种或多种丰度蛋白。

[0836] 201.根据方面175-200中任一个所述的方法,还包含在将多个多肽连接到隔室标签之前,从固体支持物上释放隔室标签。

[0837] 202.根据方面175所述的方法,还包含以下步骤(d),将隔室标记的多肽连接到与记录标签相关的固体支持物上。

[0838] 203.根据方面202所述的方法,还包含将隔室标记的多肽上的隔室标签的信息转移到相关的记录标签。

[0839] 204.根据方面203所述的方法,还包含在步骤(e)之前,将隔室标签从隔室标记的

多肽上移除。

[0840] 205.根据方面175-204中任一个所述的方法,还包含基于分析的多肽的隔室标签序列来确定分析的多肽的来源单细胞的身份。

[0841] 206.根据方面175-204中任一个所述的方法,还包含基于分析的多肽的隔室标签序列来确定分析的多肽的来源蛋白质或蛋白质复合物的身份。

[0842] 207.根据方面32-206中任一个所述的方法,其中用多个氨基酸重复步骤(f)、(g)、(h),以及(i)。

[0843] 208.根据方面207所述的方法,其中用两种或更多种氨基酸重复步骤(f)、(g)、(h),以及(i)。

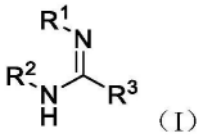
[0844] 209.根据方面207所述的方法,其中用多达约100个氨基酸重复步骤(f)、(g),以及(h)。

[0845] 210.一种用于分析多肽的试剂盒,包含:

[0846] (a)用于提供任选地直接或间接与记录标签相关的多肽的试剂;

[0847] (b)用于官能化多肽的N-末端氨基酸(NTAA)的试剂,其中试剂包含选自由以下组成的群组的化合物:

[0848] (i)式(I)化合物:

[0849] 
(I)

[0850] 或其盐或缀合物,

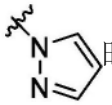
[0851] 其中,

[0852] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

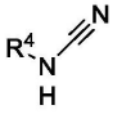
[0853] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[0854] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基是未取代的或取代的;

[0855] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[0856] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[0857] (ii)式(II)化合物:

[0858] 
(II)

[0859] 或其盐或缀合物,

[0860] 其中,

[0861] $R^4$是H,$C_{1-6}$烷基,环烷基,-C(O)$R^g$或-C(O)O$R^g$;和

[0862] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[0863] (iii)式(III)化合物:

[0864] $R^5$-N=C=S (III)

[0865] 或其盐或缀合物，

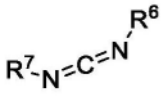
[0866] 其中，

[0867] $R^5$是$C_{1-6}$烷基，$C_{2-6}$烯基，环烷基，杂环烷基，芳基或杂芳基；

[0868] 其中$C_{1-6}$烷基，$C_{2-6}$烯基，环烷基，杂环烷基，芳基或杂芳基各自未被取代或由选自由选自卤素，$-NR^hR^i$，$-S(O)_2R^j$或杂环基组成的群组的一个或多个基团取代；

[0869] $R^h$，$R^i$和$R^j$各自独立地为H，$C_{1-6}$烷基，$C_{1-6}$卤代烷基，芳基烷基，芳基或杂芳基，其中$C_{1-6}$烷基，$C_{1-6}$卤代烷基，芳基烷基，芳基和杂芳基各自未被取代或被取代。

[0870] (iv)式(IV)化合物：

[0871] 
（IV）

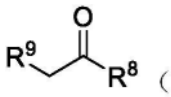
[0872] 或其盐或缀合物，

[0873] 其中，

[0874] $R^6$和$R^7$各自独立地为H，$C_{1-6}$烷基，$-CO_2C_{1-4}$烷基，$-OR^k$，芳基或环烷基，其中$C_{1-6}$烷基，$-CO_2C_{1-4}$烷基，$-OR^k$，芳基和环烷基各自未被取代或被取代；和

[0875] $R^k$为H，$C_{1-6}$烷基或杂环基，其中$C_{1-6}$烷基和杂环基各自未被取代或被取代；

[0876] (v)式(V)的化合物：

[0877] 
（V）

[0878] 或其盐或缀合物，

[0879] 其中，

[0880] $R^8$是卤素或$-OR^m$；

[0881] $R^m$为H，$C_{1-6}$烷基或杂环基；和

[0882] $R^9$是氢，卤素或$C_{1-6}$卤代烷基；

[0883] (vi)式(VI)的金属复合物：

[0884] $ML_n$(VI)

[0885] 或其盐或缀合物，

[0886] 其中，

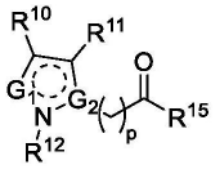
[0887] M是选自由以下组成的群组的金属：Co，Cu，Pd，Pt，Zn和Ni；

[0888] L是选自由以下组成的群组的配体：-OH，$-OH_2$，2，2’-二吡啶(bpy)，1，5-二硫代环辛烷(dtco)，1，2-双(二苯基膦)乙烷(dppe)，乙二胺(en)和三亚乙基四胺(trien)；和

[0889] n为1-8的整数，包括1和8；

[0890] 其中每个L可以相同或不同；和

[0891] (vii)式(VII)化合物：

[0892] 
（VII）

[0893] 或其盐或缀合物,

[0894] 其中,

[0895] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[0896] $G^2$是N或CH;

[0897] p是0或1;

[0898] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

[0899] $R^{15}$为H或OH;

[0900] (c)第一结合剂,其包含能够结合官能化NTAA的第一结合部分和(c1)具有关于第一结合剂的识别信息的第一编码标签,或(c2)第一可检测标签;和

[0901] (d)试剂,其将第一编码标签的信息转移到记录标签,以产生延伸记录标签;并且

[0902] (e)用于分析延伸记录标签的试剂或用于检测第一可检测标签的试剂。

[0903] 211.根据方面210所述的试剂盒,其中(a)的试剂被配置为提供多肽和连接至支持物(例如,固体支持物)的相关记录标签。

[0904] 212.根据方面210所述的试剂盒,其中(a)的试剂被配置为提供直接与溶液中的记录标签相关的多肽。

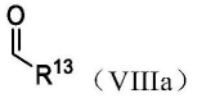
[0905] 213.根据方面210所述的试剂盒,其中(a)的试剂被配置为提供与记录标签间接相关的多肽。

[0906] 214.根据方面210所述的试剂盒,其中(a)的试剂被配置为提供不与记录标签相关的多肽。

[0907] 215.根据方面210-214中任一个所述的试剂盒,其中试剂盒包含用于官能化多肽的NTAA的两种或更多种不同试剂。

[0908] 216.根据方面215所述的试剂盒,其中试剂盒包含第一试剂和第二试剂,第一试剂包含选自由以下组成的群组:如方面125中所述的式(I),(II),(III),(IV),(V),(VI)和(VII)的化合物,或其盐或缀合物。

[0909] 217.根据方面216所述的试剂盒,其中第二试剂包含式(VIIIa)或(VIIIb)的化合物:

[0910]  (VIIIa)

[0911] 或其盐或缀合物,

[0912] 其中,

[0913] $R^{13}$是H,$C_{1-6}$烷基,芳基,杂芳基,环烷基或杂环烷基,其中$C_{1-6}$烷基,芳基,杂芳基,环烷基和杂环烷基各自是未取代的或取代的;或

[0914] $R^{13}$-XX (VIIIb)

[0915] 其中,

[0916] $R^{13}$为$C_{1-6}$烷基,芳基,杂芳基,环烷基或杂环烷基,其各自未被取代或被取代;和

[0917] X是卤素。

[0918] 218.根据方面217所述的试剂盒,其中式(VIIIa)的化合物是甲醛。

[0919] 219.根据方面217所述的试剂盒,其中式(VIIIb)的化合物是碘甲烷。

[0920] 220.根据方面210-219中任一个所述的试剂盒,其中试剂盒包含两种或更多种不同的结合剂。

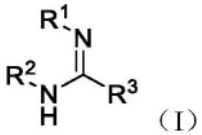
[0921] 221.根据方面210-220中任一个所述的试剂盒,其进一步包含试剂,其用于消除官能化NTAA以暴露新的NTAA。

[0922] 222.根据方面211-221中任一个所述的试剂盒,其中试剂盒包含用于消除官能化NTAA的两种或更多种不同试剂。

[0923] 223.根据方面221或方面222所述的试剂盒,其中用于消除官能化NTAA的试剂包含化学裂解试剂或酶裂解试剂。

[0924] 224.根据方面221或方面222的试剂盒,其中用于消除官能化NTAA的试剂包含羧肽酶或氨肽酶或其变体,突变体或修饰的蛋白质;水解酶或其变体,突变体或修饰的蛋白质;温和型埃德曼降解试剂;Edmanase酶;TFA;碱;或其任何组合。

[0925] 225.根据方面210-224中任一个所述的试剂盒,其中化学试剂包含选自由式(I)的化合物组成的群组的化合物:

[0926]  (I)

[0927] 或其盐或缀合物,

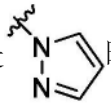
[0928] 其中,

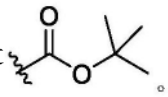
[0929] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

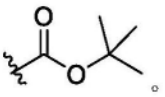
[0930] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

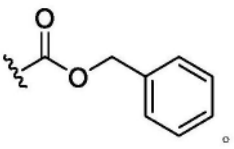
[0931] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基是未取代的或取代的;

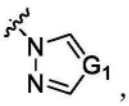
[0932] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[0933] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

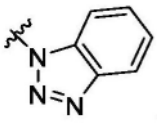
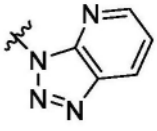
[0934] 226.根据方面225所述的试剂盒,其中$R^1$是  。

[0935] 227.根据方面225或方面226所述的试剂盒,其中R2是  。

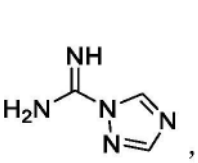
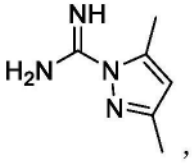
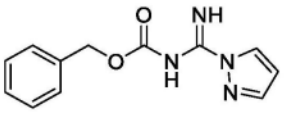
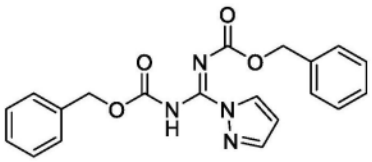
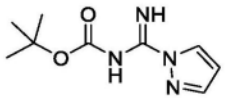
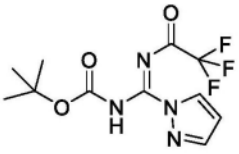
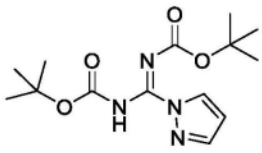
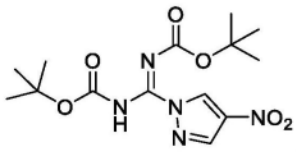
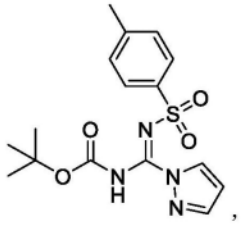
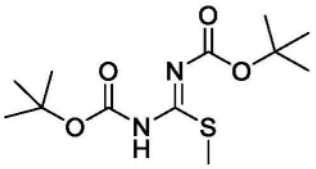
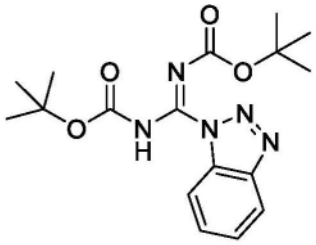
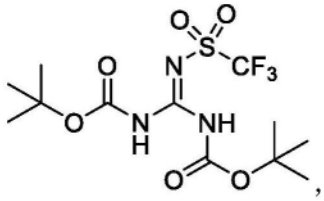
[0936] 228.根据方面225-227中任一个所述的试剂盒,其中$R^1$或$R^2$是  。

[0937] 229.根据方面225-228中任一个所述的试剂盒,其中$R^3$是  ,其中$G_1$是N或CH。

[0938] 230.根据方面225-228中任一个所述的试剂盒,其中$R^3$是

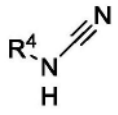

或

。

[0939] 231.根据方面225所述的试剂盒，其中式(I)的化合物选自


，

，

，

，

，

，

，

，

，

，

，以及

或其盐或缀合物。

[0940] 232.根据方面225-231中任一个所述的试剂盒，其中化学试剂另外包含Mukaiyama试剂(2-氯-1-甲基吡啶碘化物)。

[0941] 233.根据方面210-224中任一个所述的试剂盒，其中化学试剂包含选自由式(II)的化合物组成的群组的化合物：

[0942] 
(II)

[0943] 或其盐或缀合物，

[0944] 其中，

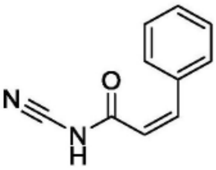
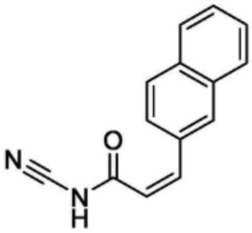
[0945] $R^4$是H，$C_{1-6}$烷基，环烷基，-C(O)$R^g$或-C(O)O$R^g$；和

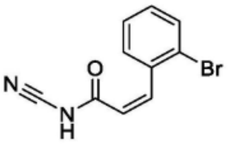
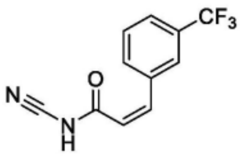
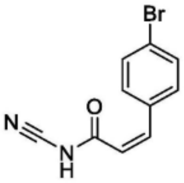
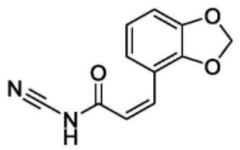
[0946] $R^g$为H，$C_{1-6}$烷基，$C_{2-6}$烯基，$C_{1-6}$卤代烷基或芳基烷基，其中$C_{1-6}$烷基，$C_{2-6}$烯基，$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代；

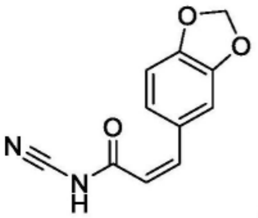
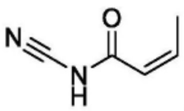
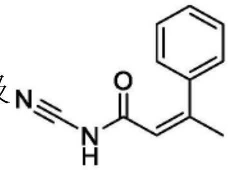
[0947] 234.根据方面142所述的试剂盒，其中$R^4$是羧基苄基。

[0948] 235.根据方面142所述的试剂盒，其中$R^4$是-C(O)$R^g$，并且$R^g$是$C_{2-6}$烯基，其任选地被芳基、杂芳基或杂环烷基取代。

[0949] 236.根据方面142所述的试剂盒，其中化合物选自

 ,  ,

 ,  ,  ,  ,

 ,  , 以及  , 或其盐或缀合物。

[0950] 237.根据方面142-145中任一个所述的试剂盒,其中化学试剂另外包含TMS-Cl,Sc(OTf)$_2$,Zn(OTf)$_2$或含镧系元素的试剂。

[0951] 238.根据方面210-224中任一个所述的试剂盒,其中化学试剂包含选自由式(III)的化合物组成的群组的化合物:

[0952] $R^5$-N＝C＝S (III)

[0953] 或其盐或缀合物,

[0954] 其中,

[0955] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[0956] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自由选自卤素,-$NR^hR^i$,-S(O)$_2$Rj或杂环基组成的群组的一个或多个基团取代;

[0957] $R^h$,$R^i$和Rj各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[0958] 239.根据方面238所述的试剂盒,其中$R^5$是取代的苯基。

[0959] 240.根据方面238或方面239的试剂盒,其中$R^5$是苯基,其被选自由卤素,-$NR^hR^i$,-S(O)$_2$Rj和杂环基组成的群组的一个或多个基团取代。

[0960] 241.根据方面238所述的试剂盒,其中式(III)的化合物是三甲基硅烷基异硫氰酸酯(TMSITC)或五氟苯基异硫氰酸酯(PFPITC)。

[0961] 242.根据方面238-240中任一个所述的试剂盒,其中化学试剂另外包含碳二亚胺化合物。

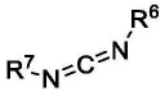
[0962] 243.根据方面238-242中任一个所述的试剂盒,其中用于消除官能化NTAA的试剂包含三氟乙酸或盐酸。

[0963] 244.根据方面238-243中任一个所述的试剂盒,其中用于消除官能化NTAA的试剂包含温和型埃德曼降解试剂。

[0964] 245.根据方面238-243中任一个所述的试剂盒,其中用于消除官能化NTAA的试剂

包含Edmanase或工程改造的水解酶,氨肽酶或羧肽酶。

[0965] 246.根据方面210-224中任一个所述的试剂盒,其中化学试剂包含选自由式(IV)的化合物组成的群组的的化合物:

[0966] 
(IV)

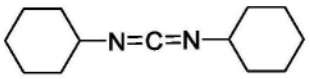
[0967] 或其盐或缀合物,

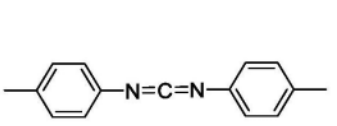
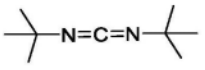
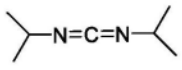
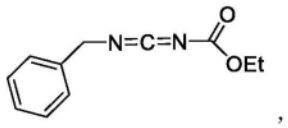
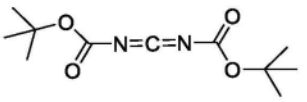
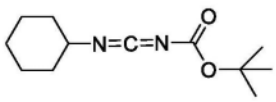
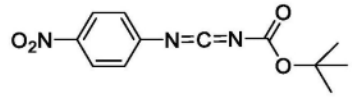
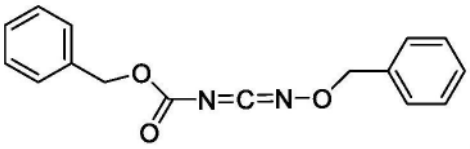
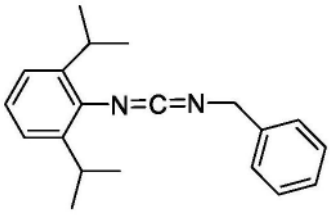
[0968] 其中,

[0969] $R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基,$-CO_2C_{1-4}$烷基,$-OR^k$,芳基或环烷基,其中$C_{1-6}$烷基,$-CO_2C_{1-4}$烷基,$-OR^k$,芳基和环烷基各自未被取代或被取代;和

[0970] $R^k$为H,$C_{1-6}$烷基或杂环基,其中$C_{1-6}$烷基和杂环基各自未被取代或被取代。

[0971] 247.根据方面246所述的试剂盒,其中$R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基或环烷基。

[0972] 248.根据方面246所述的试剂盒,其中化合物选自  ,

 ,
 ,
 ,
 ,
 ,
 ,
 ,
 以及  或其盐或缀合物。

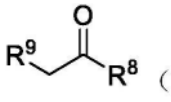
[0973] 249.根据方面246-248中任一个所述的方法,其中式(IV)的化合物通过相应硫脲的脱硫制备。

[0974] 250.根据方面246-249中任一个所述的试剂盒,其中化学试剂另外包含Mukaiyama试剂(2-氯-1-甲基吡啶碘化物)。

[0975] 251.根据方面246-249中任一个所述的试剂盒,其中化学试剂另外包含路易斯酸。

[0976] 252.根据方面251所述的试剂盒,其中路易斯酸选自N-((芳基)亚氨基-乙酰萘酮)$ZnCl_2$,$Zn(OTf)_2$,$ZnCl_2$,$PdCl_2$,CuCl 2和$CuCl_2$。

[0977] 253.根据方面210-224中任一个所述的试剂盒,其中化学试剂包含选自由式(V)的化合物组成的群组的的化合物:

[0978] 
(V)

[0979] 或其盐或缀合物,

[0980] 其中,

[0981] $R^8$是卤素或$-OR^m$;

[0982] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[0983] $R^9$是氢、卤素或$C_{1-6}$卤代烷基。

[0984] 254.根据方面253所述的试剂盒,其中$R^8$是氯。

[0985] 255.根据方面253或方面254所述的试剂盒,其中$R^9$是氢或溴。

[0986] 256.根据方面253-255中任一个所述的试剂盒,其中化学试剂另外包含肽偶联剂。

[0987] 257.根据方面256所述的试剂盒,其中肽偶联试剂是碳二亚胺化合物。

[0988] 258.根据方面257所述的试剂盒,其中碳二亚胺化合物是二异丙基碳二亚胺(DIC)或1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)。

[0989] 259.根据方面253-258中任一个所述的试剂盒,其中用于消除官能化NTAA的试剂包含酰基肽水解酶(APH)。

[0990] 260.根据方面210-224中任一个所述的方法,其中化学试剂包含式(VI)的金属络合物:

[0991] $ML_n$(VI)

[0992] 或其盐或缀合物,

[0993] 其中,

[0994] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[0995] L是选自由以下组成的群组的配体:-OH,-$OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[0996] n为1-8的整数,包括1和8;

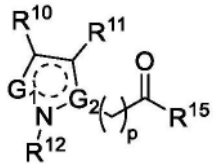
[0997] 其中每个L可以相同或不同。

[0998] 261.根据方面260所述的试剂盒,其中M是Co。

[0999] 262.根据方面260或261所述的试剂盒,其中化学试剂包含顺式-β-羟基(三亚乙基四胺)水合钴(III)复合物。

[1000] 263.根据方面260或261所述的试剂盒,其中所述化学试剂包含β-[Co(trien)(OH)($OH_2$)]$^{2+}$。

[1001] 264.根据方面210-224中任一个所述的试剂盒,其中化学试剂包含选自由式(VII)的化合物组成的群组的的化合物:

[1002] 

(VII)

[1003] 或其盐或缀合物,

[1004] 其中,

[1005] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[1006] $G^2$是N或CH;

[1007] p是0或1;

[1008] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

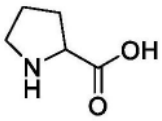
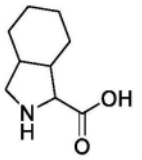
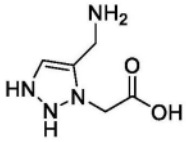
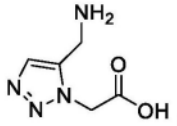
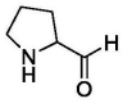
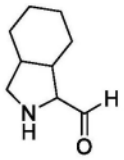
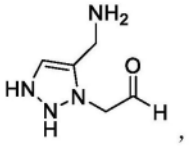
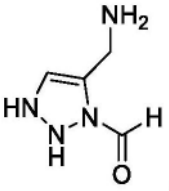
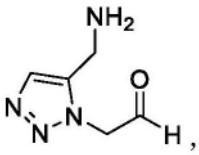
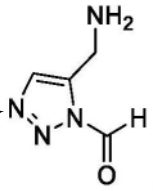
[1009] $R^{15}$为H或OH。

[1010] 265.根据方面264所述的试剂盒,其中$G^1$是$CH_2$且$G^2$是CH,或$G^1$是NH且$G^2$是N。

[1011] 266.根据方面264或265所述的试剂盒,其中$R^{12}$是H。

[1012] 267.根据方面264-266中任一个所述的试剂盒,其中$R^{10}$和$R^{11}$各自为H。

[1013] 268.根据方面264所述的试剂盒,其中式(VII)的化合物选自由以下组成的群组:


,

,

,

,

,

,

,

,

,以及

或其盐或缀合物。

[1014] 269.根据方面225-268中任一个所述的试剂盒,其中用于消除官能化NTAA的试剂包含碱。

[1015] 270.根据方面269所述的试剂盒,其中碱是氢氧化物、烷基化胺、环胺、碳酸盐缓冲剂或金属盐。

[1016] 271.根据方面270所述的试剂盒,其中氢氧化物是氢氧化钠

[1017] 272.根据方面270所述的试剂盒,其中烷基化胺选自甲胺,乙胺,丙胺,二甲胺,二乙胺,二丙胺,三甲胺,三乙胺,三丙胺,环己胺,苄胺,苯胺,二苯胺,N,N-二异丙基乙胺(DIPEA),和二异丙基胺基锂(LDA)。

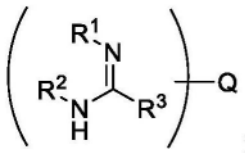
[1018] 273.根据方面270所述的试剂盒,其中环胺选自吡啶,嘧啶,咪唑,吡咯,吲哚,哌啶,吡咯烷,1,8-二氮杂双环[5.4.0]十一-7-烯(DBU)和1,5-二氮杂双环[4.3.0]壬-5-烯(DBN)。

[1019] 274.根据方面270所述的试剂盒,其中碳酸盐缓冲液包含碳酸钠,碳酸钾,碳酸钙,碳酸氢钠,碳酸氢钾或碳酸氢钙。

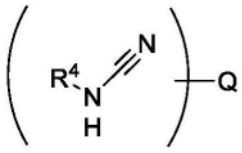
[1020] 275.根据方面270所述的试剂盒,其中金属盐包含银。

[1021] 276.根据方面270所述的试剂盒,其中金属盐是$AgClO_4$。

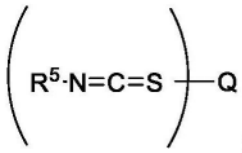
[1022] 277.根据方面210-276中任一个所述的试剂盒,其中化学试剂包含选自由以下组成的群组的缀合物:

[1023] 
式(I)-Q,

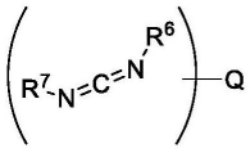
[1024] 其中$R^1$,$R^2$和$R^3$如方面1中对于式(I)所定义,并且Q是配体;

[1025] 
式（II）-Q，

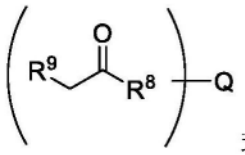
[1026] 其中$R^4$如对方面1中的式(II)所定义，并且Q是配体；

[1027] 
式（III）-Q，

[1028] 其中$R^5$如对方面1中的式(III)所定义，并且Q是配体；

[1029] 
，式（IV）-Q，

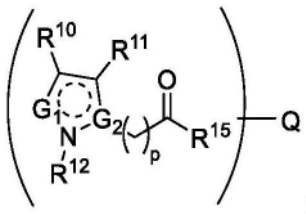
[1030] 其中$R^6$和$R^7$如方面1中对于式(IV)所定义，并且Q是配体；

[1031] 
式（V）-Q，

[1032] 其中$R^8$和$R^9$如方面1中对于式(V)所定义，并且Q是配体；

[1033] $(ML_n)$-Q式(VI)-Q，

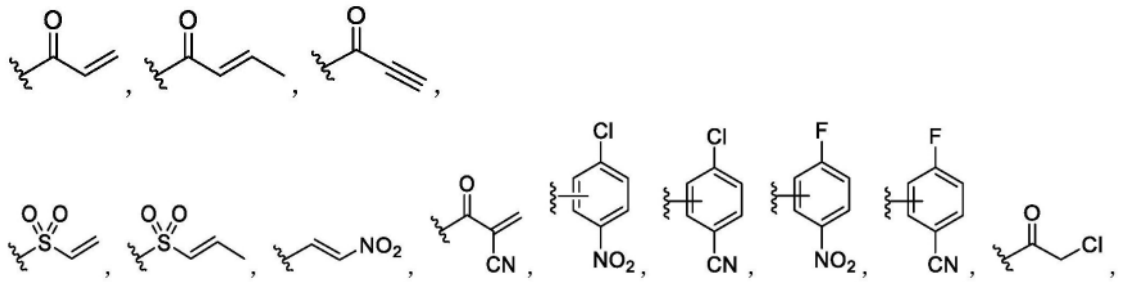
[1034] 其中M、L和N如方面1中对于式(V)所定义，并且Q是配体；

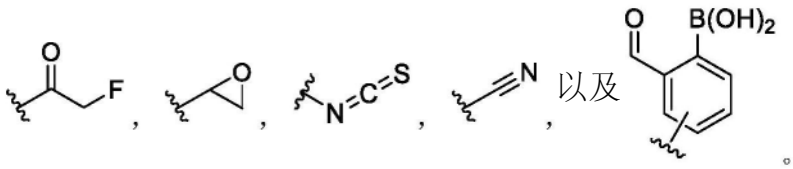
[1035] 
式（VII）-Q，

[1036] 其中$R^{10}$，$R^{11}$，$R^{12}$，$R^{15}$，$G^1$，$G^2$和p如方面1中对于式(VII)所定义，并且Q是配体。

[1037] 278.根据方面277所述的试剂盒，其中Q选自-$C_{1-6}$烷基，-$C_{2-6}$烯基，-$C_{2-6}$炔基，芳基，杂芳基，杂环基，-N＝C＝S，-CN，-C(O)$R^n$，-C(O)O$R^o$，-S$R^p$或-S(O)$_2R^q$；其中-$C_{1-6}$烷基，-$C_{2-6}$烯基，-$C_{2-6}$炔基，芳基，杂芳基和杂环基各自未被取代或被取代，并且$R^n$，$R^o$，$R^p$和$R^q$各自独立地选自-$C_{1-6}$烷基，-$C_{1-6}$卤代烷基，-$C_{2-6}$烯基，-$C_{2-6}$炔基，芳基，杂芳基和杂环基。

[1038] 279.根据方面277所述的试剂盒，其中Q选自由以下组成的群组：






[1039] 280.根据方面277所述的试剂盒,其中Q是荧光团。

[1040] 281.根据方面210-280中任一个所述的试剂盒,其中结合剂结合至末端氨基酸残基、末端二氨基酸残基或末端三氨基酸残基。

[1041] 282.根据方面210-280中任一个所述的方法,其中结合剂结合到翻译后修饰的氨基酸。

[1042] 283.根据方面210-282中任一个所述的试剂盒,其中记录标签包含核酸、寡核苷酸、修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、具有保护碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代DNA,或其组合。

[1043] 284.根据方面283所述的试剂盒,其中DNA分子是骨架修饰的、糖修饰的或核碱基修饰的。

[1044] 285.根据方面283所述的试剂盒,其中DNA分子具有核碱基保护基,如Alloc,亲电子保护基,如thiranes,乙酰基保护基,硝基苄基保护基,磺酸酯保护基,或传统的碱不稳定保护基,包括Ultramild试剂。

[1045] 286.根据方面210-285中任一个所述的试剂盒,其中记录标签包含通用引发位点。

[1046] 287.根据方面286所述的试剂盒,其中通用引发位点包含用于扩增、测序或二者的引物位点。

[1047] 288.根据方面210-287中任一个所述的试剂盒,其中记录标签包含独特分子标识符(UMI)。

[1048] 289.根据方面210-288中任一个所述的试剂盒,其中记录标签包含条形码。

[1049] 290.根据方面210-288中任一个所述的试剂盒,其中记录标签在其3'-末端包含间隔区。

[1050] 291.根据方面210-290中任一个所述的试剂盒,其中用于提供多肽和与支持物连接的相关记录标签的试剂提供多肽和支持物上的相关记录标签的共价连接。

[1051] 292.根据方面210-291中任一个所述的试剂盒,固体支持物是珠粒、多孔珠粒、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。

[1052] 293.根据方面292所述的试剂盒,其中载体包含金、银、半导体或量子点。

[1053] 294.根据方面292所述的试剂盒,其中纳米粒子包含金、银或量子点。

[1054] 295.根据方面292所述的试剂盒,其中珠粒是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[1055] 296.根据方面210-295中任一个所述的试剂盒,其中用于提供多肽和与支持物连接的相关记录标签的试剂提供多种多肽和连接到支持物的相关记录标签。

[1056] 297.根据方面296所述的试剂盒,其中多种多肽在支持物上间隔开,其中多肽之间的平均距离为约≥20nm。

[1057] 298.根据方面210-297中任一个所述的试剂盒,其中结合剂是多肽或蛋白质。

[1058] 299.根据方面210-298中任一个所述的试剂盒,其中结合剂包含氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰蛋白;抗生物素蛋白或其变体,突变体或修饰蛋白;ClpS或其变体,突变体或修饰的蛋白质;或结合氨基酸的修饰的小分子,即万古霉素或其变体,突变体或经修饰的分子;或抗体或其结合片段;或其任何组合。

[1059] 300.根据方面210-299中任一个所述的试剂盒,其中结合剂结合到单个氨基酸残基(例如,N-末端氨基酸残基,C-末端氨基酸残基,或内部氨基酸残基),二肽(例如,N-末端二肽,C-末端二肽,或内部二肽),三肽(例如,N-末端三肽,C-末端三肽,或内部三肽),或分析物或多肽的翻译后修饰。

[1060] 301.根据方面210-300中任一个所述的试剂盒,其中结合剂结合NTAA官能化的单个氨基酸残基,NTAA官能化的二肽,NTAA官能化的三肽或NTAA官能化的多肽。。

[1061] 302.根据方面210-301中任一个所述的试剂盒,结合剂能够选择性地结合多肽。

[1062] 303.根据方面210-302中任一个所述的试剂盒,其中编码标签是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。

[1063] 304.根据方面210-303中任一个所述的试剂盒,其中编码标签包含编码器或条形码序列。

[1064] 305.根据方面210-304中任一个所述的试剂盒,其中编码标签还包含间隔区、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。

[1065] 306.根据方面210-305中任一个所述的试剂盒,其中结合剂中的结合部分和编码标签通过接头连接。

[1066] 307.根据方面210-305中任一个所述的试剂盒,其中结合部分和编码标签通过SpyTag/SpyCatcher肽-蛋白质对,SnoopTag/SnoopCatcher肽-蛋白质对或HaloTag/HaloTag配体对连接。

[1067] 308.根据方面210-307中任一个所述的试剂盒,其中记录标签到编码标签的信息转移通过DNA连接酶或RNA连接酶介导。

[1068] 309.根据方面210-307中任一个所述的试剂盒,其中编码标签到记录标签的信息转移通过DNA聚合酶、RNA聚合酶或逆转录酶介导。

[1069] 310.根据方面210-307中任一个所述的试剂盒,其中用于编码标签到记录标签的信息转移的试剂包含化学连接试剂。

[1070] 311.根据方面310所述的试剂盒,其中化学连接试剂用于单链DNA。

[1071] 312.根据方面310所述的试剂盒,其中化学连接试剂用于双链DNA。

[1072] 313.根据方面210-312中任一个所述的试剂盒,其还包含连接试剂,连接试剂由两种DNA或RNA连接酶变体、腺苷酸化变体和组成型非腺苷酸化变体组成。

[1073] 314.根据方面210-312中任一个所述的试剂盒,其进一步包含由DNA或RNA连接酶和DNA/RNA脱腺苷酸酶组成的连接试剂。

[1074] 315.根据方面210-314中任一个所述的试剂盒,其中试剂盒另外包含用于核酸测序方法的试剂。

[1075] 316.根据方面315所述的试剂盒,其中核酸测序方法是通过合成测序、通过连接测

序、通过杂交测序、polony测序、离子半导体测序或焦磷酸测序。

[1076] 317.根据方面315所述的试剂盒,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。

[1077] 318.根据方面210-317中任一个所述的试剂盒,其中试剂盒另外包含用于扩增延伸记录标签的试剂。

[1078] 319.根据方面210-318中任一个所述的试剂盒,还包括添加循环标记的试剂。

[1079] 320.根据方面319所述的试剂盒,其中循环标记提供关于结合剂与多肽结合的次序的信息。

[1080] 321.根据方面319或方面320所述的试剂盒,其中将循环标记添加到编码标签。

[1081] 322.根据方面319或方面320所述的试剂盒,其中将循环标记添加到记录标签。

[1082] 323.根据方面319或方面320所述的试剂盒,其中将循环标记添加到结合剂。

[1083] 324.根据方面319或方面320所述的试剂盒,其中独立于编码标签、记录标签和结合剂添加循环标记。

[1084] 325.根据方面210-324中任一个所述的试剂盒,其中包含于延伸记录标签中的编码标签信息的次序提供关于结合到多肽的结合剂的结合次序的信息。

[1085] 326.根据方面210-325中任一个所述的试剂盒,其中包含于延伸记录标签中的编码标签信息的频率提供关于结合到多肽的结合剂的结合频率的信息。

[1086] 327.根据方面210-326中任一个所述的试剂盒,其被配置用于分析来自包含多个蛋白质复合物、蛋白质或多肽的样品的一种或多种多肽。

[1087] 328.根据方面327所述的试剂盒,进一步包含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的工具,其中每个隔室包含多个隔室标签,其任选地连接到支持物(例如,固体支持物),其中在单个隔室内的多个隔室标签是相同的,并且不同于其他隔室的隔室标签。

[1088] 329.根据方面327或328所述的试剂盒,其进一步包含用于将多个蛋白质复合物、蛋白质和/或多肽片段化成多种多肽的试剂。

[1089] 330.根据方面328所述的试剂盒,隔室是微流体液滴。

[1090] 331.根据方面328所述的试剂盒,隔室是微孔。

[1091] 332.根据方面328所述的试剂盒,隔室是平面上的分离区域。

[1092] 333.根据方面328-332中任一个所述的试剂盒,其中每一个隔室平均包含单独的细胞。

[1093] 334.根据方面327-333中任一个所述的试剂盒,其还包含用于用多个通用DNA标签标记多个蛋白质复合物、蛋白质或多肽的试剂。

[1094] 335.根据方面328-334中任一个所述的试剂盒,其中用于将隔室标签信息转移至与多肽相关的记录标签的试剂包含引物延伸或连接试剂。

[1095] 336.根据方面328-335中任一个所述的试剂盒,固体支持物是珠粒、多孔珠粒、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。

[1096] 337.根据方面328-335中任一个所述的试剂盒,其中固体支持物包含珠粒。

[1097] 338.根据方面337所述的试剂盒,其中珠粒是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[1098] 339.根据方面328-338中任一个所述的试剂盒,其中隔室标签包含单链或双链核酸分子。

[1099] 340.根据方面328-339中任一个所述的试剂盒,其中隔室标签包含条形码和可选的UMI。

[1100] 341.根据方面340所述的试剂盒,其中固体支持物是珠粒,并且隔室标签包含条形码,进一步其中包含多个隔室标签连接到其上的珠粒通过均分与合并(split and pool)合成来形成。

[1101] 342.根据方面340所述的试剂盒,其中固体支持物是珠粒,并且隔室标签包含条形码,进一步其中包含多个隔室标签连接到其上的珠粒通过单个合成或固定化形成。

[1102] 343.根据方面328-342中任一个所述的试剂盒,其中隔室标签是记录标签中的成分,记录标签任选地还包含间隔区、条形码序列、独特分子标识符、通用引发位点或它们的任何组合。

[1103] 344.根据方面328-343中任一个所述的试剂盒,其中隔室标签还包含能够与多个蛋白质复合物、蛋白质或多肽的内部氨基酸、肽骨架或N-末端氨基酸反应的官能部分。

[1104] 345.根据方面344所述的试剂盒,其中官能部分是醛,叠氮化物/炔,或马来酰亚胺/硫醇,或环氧化物/亲核试剂,或逆电子需求Diels-Alder(iEDDA)基团。

[1105] 346.根据方面344所述的试剂盒,其中官能部分团是醛基。

[1106] 347.根据方面328-346中任一个所述的试剂盒,多个隔室标签通过以下步骤形成:打印、点样、将隔室隔室标签喷墨打印到隔室中,或它们的组合。

[1107] 348.根据方面328-347中任一个所述的试剂盒,其中隔室标签还包含多肽。

[1108] 349.根据方面348所述的方法,其中隔室标签多肽包含蛋白连接酶识别序列。

[1109] 350.根据方面349所述的试剂盒,其中蛋白连接酶是butelase I或其同源物。

[1110] 351.根据方面328-350中任一个所述的试剂盒,其中用于片段化多种多肽的试剂包含蛋白酶。

[1111] 352.根据方面351所述的试剂盒,蛋白酶是金属蛋白酶。

[1112] 353.根据方面352所述的试剂盒,进一步包含用于调节金属蛋白酶的活性的试剂,例如,用于金属蛋白酶的金属阳离子的光子激发释放的试剂。

[1113] 354.根据方面328-353中任一个所述的试剂盒,还包含在将多个多肽分区成多个隔室之前用于从样品中扣除一种或多种丰度蛋白的试剂。

[1114] 355.根据方面328-354中任一个所述的试剂盒,还包含在将多个肽连接到隔室标签之前,从固体支持物上释放隔室标签。

[1115] 356.根据方面328所述的试剂盒,还包含用于将隔室标记的多肽连接到与记录标签相关的固体支持物上的试剂。

[1116] 357.一种用于筛选多肽官能化试剂、氨基酸消除试剂和/或反应条件的方法,该方法包含以下步骤:

[1117] (a)将多核苷酸与多肽官能化试剂和/或氨基酸消除试剂在反应条件下接触;和

[1118] (b)评估步骤(a)对多核苷酸的影响,任选地识别对多核苷酸没有影响或具有最小

影响的多肽官能化试剂、氨基酸消除试剂和/或反应条件。

[1119] 358.根据方面357所述的方法,其中多核苷酸包含至少约4个核苷酸。

[1120] 359.根据方面357所述的方法,其中多核苷酸包含至多约10kb的核苷酸。

[1121] 360.根据权利要求357-359中任一个所述的方法,其中多核苷酸是DNA多核苷酸。

[1122] 361.根据方面357-360中任一个所述的方法,其中多核苷酸是基因组DNA或所述方法在基因组DNA存在下进行。

[1123] 362.根据方面357-360中任一个所述的方法,其中多核苷酸是分离的多核苷酸。

[1124] 363.根据方面357-360中任一个所述的方法,其中多核苷酸是多肽结合剂的一部分。

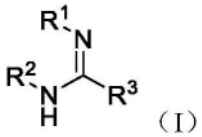
[1125] 364.根据方面357-363中任一个所述的方法,其中在不存在多肽的反应条件下使多核苷酸与多肽官能化试剂和/或氨基酸消除试剂接触。

[1126] 365.根据方面357-363中任一个所述的方法,其中在多肽存在的反应条件下使多核苷酸与多肽官能化试剂和/或氨基酸消除试剂接触。

[1127] 366.根据方面365所述的方法,其中多核苷酸是多肽结合剂的一部分。

[1128] 367.根据方面357-366中任一个所述的方法,其中多肽官能化试剂包含选自由以下组成的群组的化合物:

[1129] (i)式(I)化合物:

[1130] 
(I)

[1131] 或其盐或缀合物,

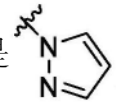
[1132] 其中,

[1133] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

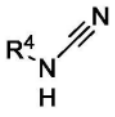
[1134] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[1135] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基是未取代的或取代的;

[1136] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[1137] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[1138] (ii)式(II)化合物:

[1139] 
(II)

[1140] 或其盐或缀合物,

[1141] 其中,

[1142] $R^4$是H,$C_{1-6}$烷基,环烷基,-C(O)$R^g$或-C(O)O$R^g$;和

[1143] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[1144] (iii)式(III)化合物:

[1145] $R^5$-N=C=S(III)

[1146] 或其盐或缀合物,

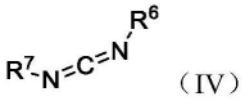
[1147] 其中,

[1148] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[1149] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自由选自卤素,-$NR^hR^i$,-$S(O)_2R^j$或杂环基组成的群组的一个或多个基团取代;

[1150] $R^h$,$R^i$和$R^j$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代。

[1151] (iv)式(IV)化合物:

[1152] 

(IV)

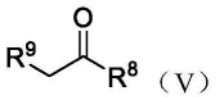
[1153] 或其盐或缀合物,

[1154] 其中,

[1155] $R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基,-$CO_2C_{1-4}$烷基,-$OR^k$,芳基或环烷基,其中$C_{1-6}$烷基,-$CO_2C_{1-4}$烷基,-$OR^k$,芳基和环烷基各自未被取代或被取代;和

[1156] $R^k$为H,$C_{1-6}$烷基或杂环基,其中$C_{1-6}$烷基和杂环基各自未被取代或被取代;

[1157] (v)式(V)的化合物:

[1158] 

(V)

[1159] 或其盐或缀合物,

[1160] 其中,

[1161] $R^8$是卤素或-$OR^m$;

[1162] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[1163] $R^9$是氢,卤素或$C_{1-6}$卤代烷基;

[1164] (vi)式(VI)的金属复合物:

[1165] $ML_n$(VI)

[1166] 或其盐或缀合物,

[1167] 其中,

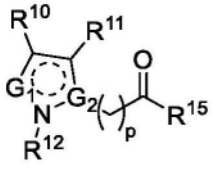
[1168] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[1169] L是选自由以下组成的群组的配体:-OH,-$OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[1170] n为1-8的整数,包括1和8;

[1171] 其中每个L可以相同或不同;和

[1172] (vii)式(VII)化合物:

[1173] 
（VII）

[1174] 或其盐或缀合物，

[1175] 其中，

[1176] $G^1$是N，$NR^{13}$或$CR^{13}R^{14}$；

[1177] $G^2$是N或CH；

[1178] p是0或1；

[1179] $R^{10}$，$R^{11}$，$R^{12}$，$R^{13}$和$R^{14}$各自独立地选自H，$C_{1-6}$烷基，$C_{1-6}$卤代烷基，$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺，其中$C_{1-6}$烷基，$C_{1-6}$卤代烷基，$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代，并且$R_{10}$和$R_{11}$可以任选地一起形成环；和

[1180] $R^{15}$为H或OH。

[1181] 368.根据方面357-367中任一个所述的方法，其中氨基酸消除试剂是化学裂解试剂或酶裂解试剂。

[1182] 369.根据方面368所述的方法，其中氨基酸消除试剂是是羧肽酶或氨肽酶或其变体，突变体或修饰的蛋白质；水解酶或其变体，突变体或修饰的蛋白质；温和型埃德曼降解试剂；Edmanase酶；TFA；碱；或其任何组合。

[1183] 370.根据方面357-369中任一个所述的方法，其中反应条件包含反应时间，反应温度，反应pH，溶剂类型，共溶剂，催化剂和离子液体，和电化学势。

[1184] 371.根据方面357-370中任一个所述的方法，其中步骤(a)在溶液中进行。

[1185] 372.根据方面357-370中任一个所述的方法，其中步骤(a)在固相中进行。

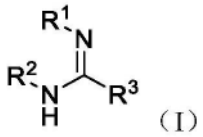
[1186] 373.根据方面357-372中任一个所述的方法，其中通过评定多肽官能化试剂、氨基酸消除试剂和/或反应条件对多核苷酸的修饰的存在、不存在或量来评估步骤(a)对多核苷酸的影响。

[1187] 374.根据方面357-373中任一个所述的方法，其中与在反应条件下不与多肽官能化试剂和/或氨基酸消除试剂接触的相应多核苷酸相比，少于50％的多核苷酸的修饰识别对多核苷酸没有影响或影响最小的多肽官能化试剂、氨基酸消除试剂和/或反应条件。

[1188] 375.一种用于筛选多肽官能化试剂、氨基酸消除试剂和/或反应条件的试剂盒，包含：(a)多核苷酸；(b)多肽官能化试剂和/或氨基酸消除试剂；(c)用于评定用于多肽官能化或消除的多肽官能化试剂、氨基酸消除试剂和/或反应条件对多核苷酸的影响的工具。

[1189] 376.根据方面375所述的试剂盒，其中多肽官能化试剂包含选自由以下组成的群组的化合物：

[1190] (i)式(I)化合物：

[1191] 
（I）

[1192] 或其盐或缀合物，

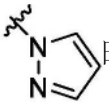
[1193] 其中，

[1194] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

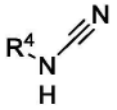
[1195] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[1196] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基未被取代或被取代;

[1197] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[1198] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[1199] (ii)式(II)化合物:

[1200] 

(II)

[1201] 或其盐或缀合物,

[1202] 其中,

[1203] $R^4$是H,$C_{1-6}$烷基,环烷基,-C(O)$R^g$或-C(O)O$R^g$;和

[1204] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[1205] (iii)式(III)化合物:

[1206] $R^5$-N=C=S (III)

[1207] 或其盐或缀合物,

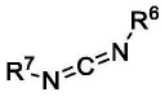
[1208] 其中,

[1209] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[1210] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自由选自卤素,-N$R^h$$R^i$,-S(O)$_2R^j$或杂环基组成的群组的一个或多个基团取代;

[1211] $R^h$,$R^i$和$R^j$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代。

[1212] (iv)式(IV)化合物:

[1213] 

(IV)

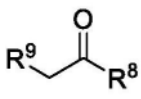
[1214] 或其盐或缀合物,

[1215] 其中,

[1216] $R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基,-$CO_2C_{1-4}$烷基,-O$R^k$,芳基或环烷基,其中$C_{1-6}$烷基,-$CO_2C_{1-4}$烷基,-O$R^k$,芳基和环烷基各自未被取代或被取代;和

[1217] $R^k$为H,$C_{1-6}$烷基或杂环基,其中$C_{1-6}$烷基和杂环基各自未被取代或被取代;

[1218] (v)式(V)的化合物:

[1219] 

(V)

[1220] 或其盐或缀合物,

[1221] 其中,

[1222] $R^8$是卤素或-$OR^m$;

[1223] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[1224] $R^9$是氢,卤素或$C_{1-6}$卤代烷基;

[1225] (vi)式(VI)的金属复合物:

[1226] $ML_n$(VI)

[1227] 或其盐或缀合物,

[1228] 其中,

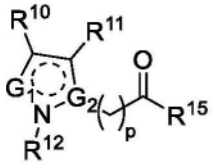
[1229] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[1230] L是选自由以下组成的群组的配体:-OH,-$OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[1231] n为1-8的整数,包括1和8;

[1232] 其中每个L可以相同或不同;和

[1233] (vii)式(VII)化合物:

[1234] 

(VII)

[1235] 或其盐或缀合物,

[1236] 其中,

[1237] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[1238] $G^2$是N或CH;

[1239] p是0或1;

[1240] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

[1241] $R^{15}$为H或OH。

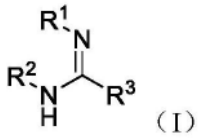
[1242] 377.根据方面375或方面376所述的试剂盒,其包含多肽官能化试剂和氨基酸消除试剂。

[1243] 378.一种对多肽进行测序的方法,该方法包含:

[1244] (a)将多肽附着于支持物或基板,或在溶液中提供多肽;

[1245] (b)用化学试剂官能化多肽的N-末端氨基酸(NTAA),其中化学试剂包含选自由以下组成的群组的化合物

[1246] (i)式(I)化合物:

[1247] 

(I)

[1248] 或其盐或缀合物,

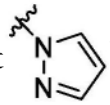
[1249] 其中,

[1250] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-$S(O)_2R^c$;

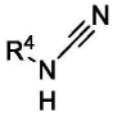
[1251] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[1252] $R^3$是杂芳基,$-NR^dC(O)OR^e$或$-SR^f$,其中杂芳基是未取代的或取代的;

[1253] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[1254] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[1255] (ii)式(II)化合物:

[1256] 

(II)

[1257] 或其盐或缀合物,

[1258] 其中,

[1259] $R^4$是H,$C_{1-6}$烷基,环烷基,$-C(O)R^g$或$-C(O)OR^g$;和

[1260] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[1261] (iii)式(III)化合物:

[1262] $R^5-N=C=S$ (iii)

[1263] 或其盐或缀合物,

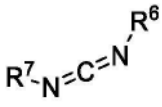
[1264] 其中,

[1265] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[1266] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自由选自卤素,$-NR^hR^i$,$-S(O)_2R^j$或杂环基组成的群组的一个或多个基团取代;

[1267] $R^h$,$R^i$和$R^j$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代。

[1268] (iv)式(IV)化合物:

[1269] 

(IV)

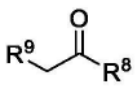
[1270] 或其盐或缀合物,

[1271] 其中,

[1272] $R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基,$-CO_2C_{1-4}$烷基,$-OR^k$,芳基或环烷基,其中$C_{1-6}$烷基,$-CO_2C_{1-4}$烷基,$-OR^k$,芳基和环烷基各自未被取代或被取代;和

[1273] $R^k$为H,$C_{1-6}$烷基或杂环基,其中$C_{1-6}$烷基和杂环基各自未被取代或被取代;

[1274] (v)式(V)的化合物:

[1275] 

(V)

[1276] 或其盐或缀合物,

[1277] 其中,

[1278] $R^8$是卤素或-$OR^m$;

[1279] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[1280] $R^9$是氢,卤素或$C_{1-6}$卤代烷基;

[1281] (vi)式(VI)的金属复合物:

[1282] $ML_n$(VI)

[1283] 或其盐或缀合物,

[1284] 其中,

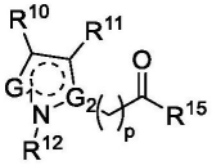
[1285] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[1286] L是选自由以下组成的群组的配体:-OH,-$OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[1287] n为1-8的整数,包括1和8;

[1288] 其中每个L可以相同或不同;和

[1289] (vii)式(VII)化合物:

[1290] 

(VII)

[1291] 或其盐或缀合物,

[1292] 其中,

[1293] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[1294] $G^2$是N或CH;

[1295] p是0或1;

[1296] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

[1297] $R^{15}$为H或OH;

[1298] (c)使多肽与多种结合剂接触,结合剂各自包含能够结合官能化NTAA的结合部分以及可检测标签;

[1299] (d)检测与多肽结合的结合剂的可检测标签,从而识别多肽的N-末端氨基酸;

[1300] (e)消除官能化NTAA以暴露新的NTAA;和

[1301] (f)重复步骤(b)-(d)以确定多肽的至少一部分的序列。

[1302] 379.根据方面378所述的方法,其中步骤(b)在步骤(c)之前进行。

[1303] 380.根据方面378所述的方法,其中步骤(b)在步骤(c)之后且在步骤(d)之前进行。

[1304] 381.根据方面378所述的方法,其中步骤(b)在步骤(c)和步骤(d)之后进行。

[1305] 382.根据方面378-381中任一个所述的方法,其中多肽通过片段化来自生物样品的蛋白质而获得。

[1306] 383.根据方面378-382中任一个所述的方法,其中支持物是珠粒、多孔珠粒、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。

[1307] 384.根据方面378-383中任一个所述的方法,其中通过化学裂解或酶裂解从多肽消除NTAA。

[1308] 385.根据方面378-384所述的方法,其中NTAA被羧肽酶或氨肽酶或其变体,突变体或修饰的蛋白质;水解酶或其变体,突变体或修饰的蛋白质;温和型埃德曼降解;Edmanase酶;TFA,碱;或其任何组合消除。

[1309] 386.根据方面378-385中任一个所述的方法,其中多肽共价附着在支持物或基板上。

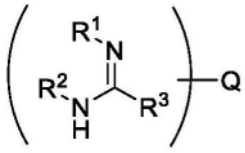
[1310] 387.根据方面378-386中任一个所述的方法,其中支持物或基板是光学透明的。

[1311] 388.根据方面378-387中任一个所述的方法,其中支持物或基板包含多个空间分辨的附着点,并且步骤a)包含将多肽附着至空间分辨的附着点。

[1312] 389.根据方面378-388中任一个所述的方法,其中结合剂的结合部分包含肽或蛋白质。

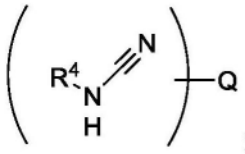
[1313] 390.根据方面378-389中任一个所述的方法,其中结合剂的结合部分包含氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰的蛋白质;anticalin或其变体,突变体或修饰的蛋白质;Clps(例如ClpS2)或其变体,突变体或修饰的蛋白质;UBR盒蛋白质或其变体,突变体或修饰的蛋白质;或结合氨基酸的修饰的小分子,即万古霉素或其变体,突变体或修饰的分子;或抗体或其结合片段;或其任何组合。

[1314] 391.根据方面378-390中任一个所述的方法,其中化学试剂包含选自由以下组成的群组的缀合物:

[1315] 

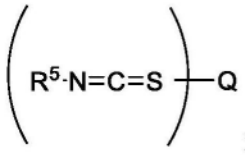
式(I)-Q,

[1316] 其中$R^1$,$R^2$和$R^3$如方面1中对于式(I)所定义,并且Q是配体;

[1317] 

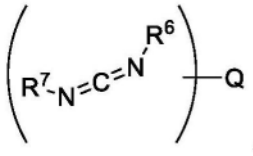
式(II)-Q,

[1318] 其中$R^4$如对方面1中的式(II)所定义,并且Q是配体;

[1319] 

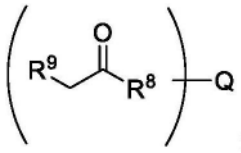
式(III)-Q,

[1320] 其中$R^5$如对方面1中的式(III)所定义,并且Q是配体;

[1321] 

,式(IV)-Q,

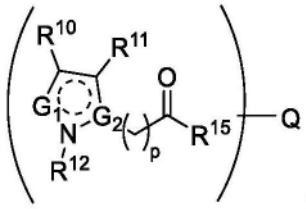
[1322] 其中$R^6$和$R^7$如方面1中对于式(IV)所定义,并且Q是配体;

[1323] 
式（V）-Q,

[1324] 其中$R^8$和$R^9$如方面1中对于式(V)所定义,并且Q是配体;

[1325] $(ML_n)$-Q式(VI)-Q,

[1326] 其中M、L和N如方面1中对于式(V)所定义,并且Q是配体;

[1327] 
式（VII）-Q,

[1328] 其中$R^{10}$,$R^{11}$,$R^{12}$,$R^{15}$,$G^1$,$G^2$和p如方面1中对于式(VII)所定义,并且Q是配体。

[1329] 392.根据方面378-391中任一个所述的方法,其中步骤(b)包含用第二化学试剂官能化NTAA,第二化学试剂选自式(VIIIa)和(VIIIb):

[1330] 
（VIIIa）

[1331] 或其盐或缀合物,

[1332] 其中,

[1333] $R^{13}$是H,$C_{1-6}$烷基,芳基,杂芳基,环烷基或杂环烷基,其中$C_{1-6}$烷基,芳基,杂芳基,环烷基和杂环烷基各自未被取代或被取代;或

[1334] $R^{13}$-X (VIIIb)

[1335] 其中,

[1336] $R^{13}$为$C_{1-6}$烷基,芳基,杂芳基,环烷基或杂环烷基,其各自未被取代或被取代;和

[1337] X是卤素。

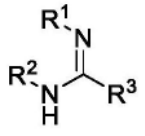
[1338] 393.根据方面378-392中任一个所述的方法,其中多肽是部分或完全消化的蛋白质。

[1339] 394.一种对样品中的多个多肽分子进行测序的方法,该方法包含:

[1340] (a)将样品中的多肽分子附着于支持物或基板上的多个空间分辨的附着点;

[1341] (b)用化学试剂官能化多肽分子的N-末端氨基酸(NTAA),其中化学试剂包含选自由以下组成的群组的化合物:

[1342] (i)式(I)化合物:

[1343] 
（I）

[1344] 或其盐或缀合物,

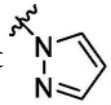
[1345] 其中,

[1346] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

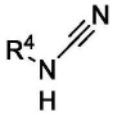
[1347] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[1348] $R^3$是杂芳基,$-NR^dC(O)OR^e$或$-SR^f$,其中杂芳基未被取代或被取代;

[1349] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[1350] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[1351] (ii)式(II)化合物:

[1352] 
(II)

[1353] 或其盐或缀合物,

[1354] 其中,

[1355] $R^4$是H,$C_{1-6}$烷基,环烷基,$-C(O)R^g$或$-C(O)OR^g$;和

[1356] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[1357] (iii)式(III)化合物:

[1358] $R^5-N=C=S$ (iii)

[1359] 或其盐或缀合物,

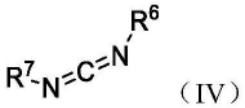
[1360] 其中,

[1361] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[1362] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自由选自卤素,$-NR^hR^i$,$-S(O)_2R^j$或杂环基组成的群组的一个或多个基团取代;

[1363] $R^h$,$R^i$和$R^j$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代。

[1364] (iv)式(IV)化合物:

[1365] 
(IV)

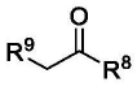
[1366] 或其盐或缀合物,

[1367] 其中,

[1368] $R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基,$-CO_2C_{1-4}$烷基,$-OR^k$,芳基或环烷基,其中$C_{1-6}$烷基,$-CO_2C_{1-4}$烷基,$-OR^k$,芳基和环烷基各自未被取代或被取代;和

[1369] $R^k$为H,$C_{1-6}$烷基或杂环基,其中$C_{1-6}$烷基和杂环基各自未被取代或被取代;

[1370] (v)式(V)的化合物:

[1371] 
(V)

[1372] 或其盐或缀合物,

[1373] 其中,

[1374] $R^8$是卤素或-$OR^m$;

[1375] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[1376] $R^9$是氢,卤素或$C_{1-6}$卤代烷基;

[1377] (vi)式(VI)的金属复合物:

[1378] $ML_n$(VI)

[1379] 或其盐或缀合物,

[1380] 其中,

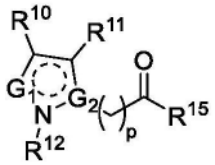
[1381] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[1382] L是选自由以下组成的群组的配体:-OH,-$OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[1383] n为1-8的整数,包括1和8;

[1384] 其中每个L可以相同或不同;和

[1385] (vii)式(VII)化合物:

[1386] 

(VII)

[1387] 或其盐或缀合物,

[1388] 其中,

[1389] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[1390] $G^2$是N或CH;

[1391] p是0或1;

[1392] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

[1393] $R^{15}$为H或OH;

[1394] (c)使多肽与多种结合剂接触,结合剂各自包含能够结合官能化NTAA的结合部分和可检测标签;

[1395] (d)对于在空间上分辨的并附着于支持物或基板的多个多肽分子,光学检测与每个多肽结合的探针的荧光标记;

[1396] (e)消除每个多肽的官能化NTAA;和

[1397] (f)重复步骤b)至d)以确定多个多肽分子的一个或多个中在空间上分辨的并附着于支持物或基板的至少一部分的序列。

[1398] 395.根据方面394所述的方法,其中步骤(b)在步骤(c)之前进行。

[1399] 396.根据方面394所述的方法,其中步骤(b)在步骤(c)之后且在步骤(d)之前进行。

[1400] 397.根据方面394所述的方法,其中步骤(b)在步骤(c)和步骤(d)之后进行。

[1401] 398.根据方面394-397中任一个所述的方法,其中样品包含生物流体、细胞提取物或组织提取物。

[1402] 399.根据方面394-398中任一个所述的方法,其还包含将步骤e)中确定的至少一种多肽分子的序列与参考蛋白质序列数据库进行比较。

[1403] 400.根据方面394-399中任一个所述的方法,其还包含比较步骤e)中确定的每种多肽的序列,将相似多肽序列分组并计数每种相似多肽序列的实例数。

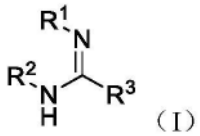
[1404] 401.根据方面394-400中任一个所述的方法,其中荧光标记是荧光基团、颜色编码的纳米颗粒或量子点。

[1405] 402.一种用于测序多肽的试剂盒,包含:

[1406] (a)用于将多肽附着至支持物或基板的试剂,或用于在溶液中提供多肽的试剂;

[1407] (b)用于官能化多肽的N-末端氨基酸(NTAA)的试剂,其中试剂包含选自由以下组成的群组的化合物:

[1408] (i)式(I)化合物:

[1409] 

(I)

[1410] 或其盐或缀合物,

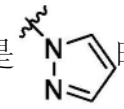
[1411] 其中,

[1412] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

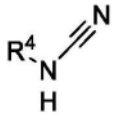
[1413] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[1414] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基是未取代的或取代的;

[1415] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[1416] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[1417] (ii)式(II)化合物:

[1418] 

(II)

[1419] 或其盐或缀合物,

[1420] 其中,

[1421] $R^4$是H,$C_{1-6}$烷基,环烷基,-C(O)$R^g$或-C(O)O$R^g$;和

[1422] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[1423] (iii)式(III)化合物:

[1424] $R^5$-N=C=S (iii)

[1425] 或其盐或缀合物,

[1426] 其中,

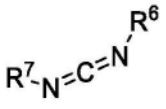
[1427] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[1428] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自

由选自卤素,-$NR^hR^i$,-$S(O)_2R^j$或杂环基组成的群组的一个或多个基团取代;

[1429] $R^h$,$R^i$和$R^j$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代。

[1430] (iv)式(IV)化合物:

[1431] 
（IV）

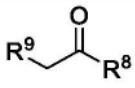
[1432] 或其盐或缀合物,

[1433] 其中,

[1434] $R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基,-$CO_2C_{1-4}$烷基,-$OR^k$,芳基或环烷基,其中$C_{1-6}$烷基,-$CO_2C_{1-4}$烷基,-$OR^k$,芳基和环烷基各自未被取代或被取代;和

[1435] $R^k$为H,$C_{1-6}$烷基或杂环基,其中$C_{1-6}$烷基和杂环基各自未被取代或被取代;

[1436] (v)式(V)的化合物:

[1437] 
（V）

[1438] 或其盐或缀合物,

[1439] 其中,

[1440] $R^8$是卤素或-$OR^m$;

[1441] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[1442] $R^9$是氢,卤素或$C_{1-6}$卤代烷基;

[1443] (vi)式(VI)的金属复合物:

[1444] $ML_n$(VI)

[1445] 或其盐或缀合物,

[1446] 其中,

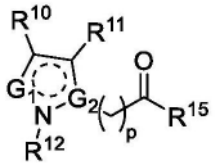
[1447] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[1448] L是选自由以下组成的群组的配体:-OH,-$OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[1449] n为1-8的整数,包括1和8;

[1450] 其中每个L可以相同或不同;和

[1451] (vii)式(VII)化合物:

[1452] 
（VII）

[1453] 或其盐或缀合物,

[1454] 其中,

[1455] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[1456] $G^2$是N或CH;

[1457] p是0或1;

[1458] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

[1459] $R^{15}$为H或OH;和

[1460] (c)包含能够结合官能化NTAA的结合部分和可检测标签的结合剂。

[1461] 403.根据方面402所述的试剂盒,其中试剂盒另外包含用于消除官能化NTAA以暴露新的NTAA的试剂。

[1462] 404.根据方面402或方面403所述的试剂盒,其中肽通过片段化来自生物样品的蛋白而获得。

[1463] 405.根据方面402-404中任一个所述的试剂盒,支持物或基板是珠粒、多孔珠粒、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。

[1464] 406.根据方面403-405所述的试剂盒,其中用于消除官能化NTAA的试剂是羧肽酶或氨肽酶或其变体,突变体或修饰的蛋白质;水解酶或其变体,突变体或修饰的蛋白质;温和型埃德曼降解试剂;Edmanase酶;TFA;碱;或其任何组合。

[1465] 407.根据方面402-406中任一个所述的试剂盒,其中多肽共价附着于支持物或基板。

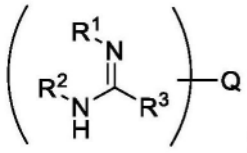
[1466] 408.根据方面402-407中任一个所述的试剂盒,其中支持物或基板是光学透明的。

[1467] 409.根据方面402-408中任一个所述的试剂盒,其中支持物或基板包含多个空间分辨的附着点,并且步骤a)包含将多肽附着至空间分辨的附着点。

[1468] 410.根据方面402-409中任一个所述的试剂盒,其中结合剂的结合部分包含肽或蛋白质。

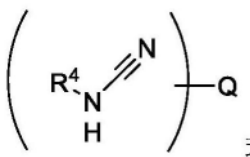
[1469] 411.根据方面402-410中任一个所述的试剂盒,其中结合剂的结合部分包含氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰的蛋白质;anticalin或其变体,突变体或修饰的蛋白质;Clps(例如ClpS2)或其变体,突变体或修饰的蛋白质;UBR盒蛋白质或其变体,突变体或修饰的蛋白质;或结合氨基酸的修饰的小分子,即万古霉素或其变体,突变体或修饰的分子;或抗体或其结合片段;或其任何组合。

[1470] 412.根据方面402-411中任一个所述的试剂盒,其中化学试剂包含选自由以下组成的群组的缀合物:

[1471] 

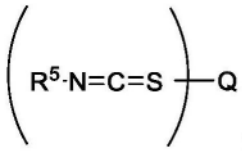
式(I)-Q,

[1472] 其中$R^1$,$R^2$和$R^3$如方面1中对于式(I)所定义,并且Q是配体;

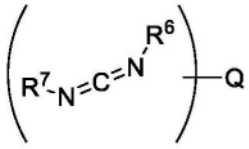
[1473] 

式(II)-Q,

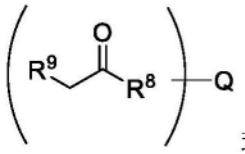
[1474] 其中$R^4$如对方面1中的式(II)所定义,并且Q是配体;

[1475] 
式（III）-Q，

[1476] 其中$R^5$如对方面1中的式(III)所定义，并且Q是配体；

[1477] 
，式（IV）-Q，

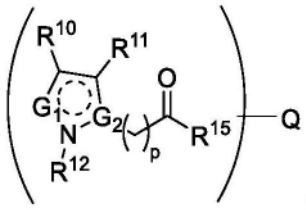
[1478] 其中$R^6$和$R^7$如方面1中对于式(IV)所定义，并且Q是配体；

[1479] 
式（V）-Q，

[1480] 其中$R^8$和$R^9$如方面1中对于式(V)所定义，并且Q是配体；

[1481] $(ML_n)$-Q式(VI)-Q，

[1482] 其中M、L和N如方面1中对于式(V)所定义，并且Q是配体；

[1483] 
式（VII）-Q，

[1484] 其中$R^{10}$，$R^{11}$，$R^{12}$，$R^{15}$，$G^1$，$G^2$和p如方面1中对于式(VII)所定义，并且Q是配体。

[1485] 方面413.根据方面402-412中任一个所述的试剂盒，试剂盒包括选自式(VIIIa)和(VIIIb)的第二化学试剂：

[1486] 
（VIIIa）

[1487] 或其盐或缀合物，

[1488] 其中，

[1489] $R^{13}$是H，$C_{1-6}$烷基，芳基，杂芳基，环烷基或杂环烷基，其中$C_{1-6}$烷基，芳基，杂芳基，环烷基和杂环烷基各自未被取代或被取代；或

[1490] $R^{13}$-XX (VIIIb)

[1491] 其中，

[1492] $R^{13}$为$C_{1-6}$烷基，芳基，杂芳基，环烷基或杂环烷基，其各自未被取代或被取代；和

[1493] X是卤素。

[1494] 414.根据方面402-413中任一个所述的方法，其中多肽是部分或完全消化的蛋白质。

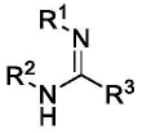
[1495] 415.一种用于测序多种多肽的试剂盒，包含：

[1496] (a)用于将样品中的多肽分子附着于支持物或基板上的多个空间分辨的附着点的

试剂;

[1497] (b)用于官能化多肽分子的N-末端氨基酸(NTAA)的试剂,其中试剂包含选自由以下组成的群组的化合物:

[1498] (i)式(I)化合物:

[1499] 
(I)

[1500] 或其盐或缀合物,

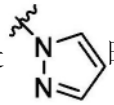
[1501] 其中,

[1502] $R^1$和$R^2$各自独立地为H,$C_{1-6}$烷基,环烷基,-C(O)$R^a$,-C(O)O$R^b$或-S(O)$_2R^c$;

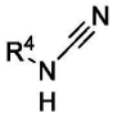
[1503] $R^a$,$R^b$和$R^c$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代;

[1504] $R^3$是杂芳基,-N$R^d$C(O)O$R^e$或-S$R^f$,其中杂芳基未被取代或被取代;

[1505] $R^d$,$R^e$和$R^f$各自独立地为H或$C_{1-6}$烷基;和

[1506] 任选地其中当$R^3$是  时,$R^1$和$R^2$不都是H;

[1507] (ii)式(II)化合物:

[1508] 
(II)

[1509] 或其盐或缀合物,

[1510] 其中,

[1511] $R^4$是H,$C_{1-6}$烷基,环烷基,-C(O)$R^g$或-C(O)O$R^g$;和

[1512] $R^g$为H,$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基或芳基烷基,其中$C_{1-6}$烷基,$C_{2-6}$烯基,$C_{1-6}$卤代烷基和芳基烷基各自未被取代或被取代;

[1513] (iii)式(III)化合物:

[1514] $R^5$-N=C=S (iii)

[1515] 或其盐或缀合物,

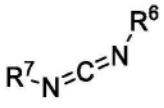
[1516] 其中,

[1517] $R^5$是$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基;

[1518] 其中$C_{1-6}$烷基,$C_{2-6}$烯基,环烷基,杂环烷基,芳基或杂芳基各自未被取代或由选自由选自卤素,-N$R^h$$R^i$,-S(O)$_2R^j$或杂环基组成的群组的一个或多个基团取代;

[1519] $R^h$,$R^i$和$R^j$各自独立地为H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基或杂芳基,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,芳基烷基,芳基和杂芳基各自未被取代或被取代。

[1520] (iv)式(IV)化合物:

[1521] 
(IV)

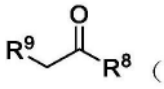
[1522] 或其盐或缀合物,

[1523] 其中,

[1524] $R^6$和$R^7$各自独立地为H,$C_{1-6}$烷基,$-CO_2C_{1-4}$烷基,$-OR^k$,芳基或环烷基,其中$C_{1-6}$烷基,$-CO_2C_{1-4}$烷基,$-OR^k$,芳基和环烷基各自未被取代或被取代;和

[1525] $R^k$为H,$C_{1-6}$烷基或杂环基,其中$C_{1-6}$烷基和杂环基各自未被取代或被取代;

[1526] (v)式(V)的化合物:

[1527] 
(V)

[1528] 或其盐或缀合物,

[1529] 其中,

[1530] $R^8$是卤素或$-OR^m$;

[1531] $R^m$为H,$C_{1-6}$烷基或杂环基;和

[1532] $R^9$是氢,卤素或$C_{1-6}$卤代烷基;

[1533] (vi)式(VI)的金属复合物:

[1534] $ML_n$(VI)

[1535] 或其盐或缀合物,

[1536] 其中,

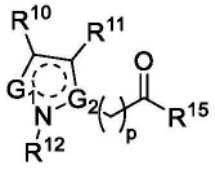
[1537] M是选自由以下组成的群组的金属:Co,Cu,Pd,Pt,Zn和Ni;

[1538] L是选自由以下组成的群组的配体:-OH,$-OH_2$,2,2'-二吡啶(bpy),1,5-二硫代环辛烷(dtco),1,2-双(二苯基膦)乙烷(dppe),乙二胺(en)和三亚乙基四胺(trien);和

[1539] n为1-8的整数,包括1和8;

[1540] 其中每个L可以相同或不同;和

[1541] (vii)式(VII)化合物:

[1542] 
(VII)

[1543] 或其盐或缀合物,

[1544] 其中,

[1545] $G^1$是N,$NR^{13}$或$CR^{13}R^{14}$;

[1546] $G^2$是N或CH;

[1547] p是0或1;

[1548] $R^{10}$,$R^{11}$,$R^{12}$,$R^{13}$和$R^{14}$各自独立地选自H,$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺,其中$C_{1-6}$烷基,$C_{1-6}$卤代烷基,$C_{1-6}$烷基胺和$C_{1-6}$烷基羟基胺各自未被取代或被取代,并且$R_{10}$和$R_{11}$可以任选地一起形成环;和

[1549] $R^{15}$为H或OH;和

[1550] (c)包含能够结合官能化NTAA的结合部分和可检测标签的结合剂。

[1551] 416.根据方面415所述的试剂盒,其中试剂盒另外包含用于消除官能化NTAA以暴露新的NTAA的试剂。

[1552] 方面417.根据方面415或416所述的试剂盒,其中样品包含生物流体、细胞提取物或组织提取物。

[1553] 方面418.根据方面415之417中任一个所述的试剂盒,其中荧光标记是荧光基团、颜色编码的纳米颗粒或量子点。

[1554] 在任何前述实施例中,记录标签和/或编码标签可以是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[1555] 在任何前述实施例中,记录标签和/或编码标签可以是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。

[1556] 在任何前述实施例中,记录标签和/或编码标签可包含可测序聚合物(例如多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[1557] 在任何前述实施例中,编码标签可以是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元,而记录标签是可测序聚合物。在一些实施例中,可以将编码标签(小分子、亚基或单体等)添加到记录标签中,以形成可测序聚合物,例如串(例如,延伸记录标签)上的珠粒(例如,各种编码标签)。

[1558] 在任何前述实施例中,延伸记录标签和/或延伸编码标签可以是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。

[1559] 在一个方面,本文公开了一种方法,其包含:(a)使分析物与能够结合至分析物的第一结合剂接触,其中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部分,分析物与连接至支持物的记录部分相关,并且分析物和/或记录部分固定至支持物;(b)将第一编码部分的信息转移到记录聚合物,以产生第一次序延伸记录部分;(c)使分析物与能够结合至分析物的第二结合剂接触,其中第二结合剂包含具有关于第二结合剂的识别信息的第二编码部分;(d)将第二编码部分的信息转移到第一次序延伸记录部分,以产生第二次序延伸记录部分;以及(e)分析第二次序延伸记录部分。在一个实施例中,所述方法还包含在步骤(d)和(e)之间的以下步骤:(x)通过用能够结合修饰的分析物的第三(或更高次序)结合剂替换第二结合剂重复步骤(c)和(d)一次或多次,其中第三(或更高次序)结合剂包含第三(或更高次序)编码部分,其具有关于第三(或更高次序)结合剂的识别信息;以及(y)将第三(或更高次序)编码部分的信息转移到第二(或更高次序)延伸记录部分,以产生第三(或更高次序)延伸记录部分;以及在步骤(e)中分析第三(或更高次序)延伸记录部分。

[1560] 在任何前述实施例中,记录部分可以是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[1561] 在任何前述实施例中,记录部分可以是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。在一些实施例中,第一、第二、第三和/或更高次序的编码部分包含可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子和/或亚基,单体,残基或结构单元。

[1562] 在任何前述实施例中,第一、第二、第三和/或更高次序的编码部分可以是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[1563] 在任何前述实施例中,第一次序延伸记录部分、第二次序延伸记录部分、第三次序延伸记录部分和/或更高次序延伸记录部分可以是可测序聚合物,例如,多核苷酸或非核酸可测序聚合物。

[1564] 在另一个方面,本文公开了试剂盒,其包含:(a)使多肽与能够结合至多肽的一个

或多个N-末端或C-末端氨基酸的第一结合剂接触,其中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部分,多肽与连接至支持物的记录部分相关,并且多肽和/或记录部分固定至支持物;(b)将第一编码部分的信息转移到记录聚合物,以产生第一次序延伸记录部分;(c)使分析物与能够结合至多肽的不同的一个或多个N-末端、内部或C-末端氨基酸的第二结合剂接触,其中第二结合剂包含具有关于第二结合剂的识别信息的第二编码部分;(d)将第二编码部分的信息转移到第一次序延伸记录部分,以产生第二次序延伸记录部分;以及(e)分析第二次序延伸记录部分。在一个实施例中,多肽的一个或多个N-末端、内部或C-末端氨基酸和/或不同的一个或多个N-末端、内部或C-末端氨基酸是修饰的氨基酸,例如修饰的N-末端氨基酸(NTAA)。

[1565] 在一个方面,本文公开了试剂盒,其包含:(a)使多肽与能够结合多肽的一个或多个N-末端或C-末端氨基酸的第一结合剂接触,其中第一结合剂包含具有关于第一结合剂的鉴定信息的第一编码部分,多肽与连接至支持物的记录部分相关,并且多肽和/或记录部分固定至支持物;(b)将第一编码部分的信息转移到记录部分,以产生第一次序扩展记录部分;(c)消除多肽的一个或多个N-末端或C-末端氨基酸,以暴露多肽的不同的一个或多个N-末端或C-末端氨基酸;(d)使分析物与能够结合多肽的不同的一个或多个N-末端或C-末端氨基酸的第二结合剂接触,其中第二结合剂包含具有关于第二结合剂的识别信息的第二编码部分;(e)将第二编码部分的信息转移到第一次序延伸记录部分,以产生第二次序延伸记录部分;以及(f)分析第二次序延伸记录部分。在一个实施例中,多肽的一个或多个N-末端或C-末端氨基酸和/或不同的一个或多个N-末端或C-末端氨基酸是修饰的氨基酸,例如修饰的N-末端氨基酸(NTAA)。

[1566] 在任何前述实施例中,步骤(a)的记录部分可以是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[1567] 在任何前述实施例中,步骤(a)的记录部分可以是可测序聚合物,例如,多核苷酸或非核酸可测序聚合物。在一些实施例中,第一或第二次序编码部分包含可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子和/或亚基,单体,残基或结构单元。

[1568] 在任何前述实施例中,第一或第二次序编码标签可以是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[1569] 在任何前述实施例中,第一次序延伸记录部分或第二次序延伸记录部分可以是可测序聚合物,例如,多核苷酸或非核酸可测序聚合物。

[1570] 在任何前述实施例中,试剂盒可以进一步包含一种或多种封阻基团。在一个实施例中,当可测序聚合物通过纳米孔时,一个或多个封阻基团能够产生特征电流封阻信号。

[1571] 在任何前述实施例中,可测序聚合物可包含一个或多个能够与结合剂结合的结合基团,结合剂在通过纳米孔时使可测序聚合物停顿。

[1572] 在任何前述实施例中,可测序聚合物可包含一种或多种当通过纳米孔时能够使可测序聚合物停顿的结构。在一个实施例中,一种或多种结构包含二级结构,例如本征亚稳定二级结构。

[1573] 在任何前述实施例中,可测序聚合物可包含使用亚磷酰胺化学构建的聚合物,或类肽聚合物,或其组合。

[1574] 在另一个方面,本文公开的是一种试剂盒,其包含上述任何实施例或其组合中任

一个所述的方法中公开的和/或使用的任何分子,分子复合物或缀合物,试剂(例如,化学或生物试剂),剂,结构(例如,支持物、表面、颗粒或珠粒),反应中间体,反应产物,结合复合物或任何其它制品。

[1575] 可以以单独提供,也可以以任何合适的组合形式提供在示例性试剂盒和方法中公开和/或使用的上述试剂盒组分,任何分子,分子复合物或缀合物,试剂(例如,化学或生物试剂),剂,结构(例如,支持物、表面、颗粒或珠粒),反应中间体,反应产物,结合复合物,或任何其他制品中的任一个,以形成试剂盒。试剂盒可以任选地包含使用说明,例如用于高度平行,高通量数字分析(例如大分子分析),特别是多肽分析。

## 附图说明

[1576] 通过参考附图的示例方式描述本公开的非限制性实施例,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,也未示出本公开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。

[1577] 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。因此,在一个实施例中,本文提供了记录标签或延伸记录标签,其包含一个或多个通用引物序列(或一对或多对通用引物序列,例如,在记录标签的5'端的一对中的一个通用引物和在记录标签或延伸记录标签的3'端的一对中的另一个通用引物),能够在多个记录标签或延伸记录标签中识别记录标签或延伸记录标签的一个或多个条形码序列,一个或多个UMI序列,一个或多个间隔区序列,和/或一个或多个编码器序列(也称为编码序列,例如编码标签的编码序列)。在某些实施例中,延伸记录标签包含(i)一个通用引物序列、一个条形码序列、一个UMI序列和一个间隔区(均来自未延伸记录标签),(ii)一个或多个串联排列的"盒",每个盒包含结合剂的编码序列、UMI序列和间隔区,并且每个盒包含来自编码标签的序列信息,和(iii)另一个通用引物序列,其可以由编码剂的编码标签在第n结合循环中提供,其中n是表示结合循环数的整数,在该结合循环后需要读取测定。在一个实施例中,在将通用引物序列引入到延伸记录标签中之后,可以继续结合循环,可以进一步延伸延伸记录标签,并且可以引入一个或多个另外的通用引物序列。在这种情况下,可以使用通用引物序列的任何组合进行延伸记录标签的扩增和/或测序。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1578] 图1B示出了将蛋白质编码转导或转化成核酸(例如,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库,以代表多个肽。延伸记录标签构成代表肽序列的DNA编码库(DEL)。所述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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1579] 图2A-D说明了根据本文所公开的方法,使用多循环结合剂(例如,抗体,anticalins,N-recognins蛋白质(例如,ClpSs,或UBR盒蛋白质)等及其变体/同源物,以及

适体等)分析多肽(例如,蛋白质)的示例,其包含与固定化蛋白质相互作用的编码标签,固定化蛋白质用单个或多个记录标签共定位或共标记。在该示例中,记录标签包含通用引发位点、条形码(例如,分区条形码、隔室条形码、和/或分级物条形码)、可选的独特分子标识符(UMI)序列,以及可选的用于编码标签和记录标签(或延伸记录标签)之间的信息转移的间隔区序列(Sp)。隔室标签(Sp)在所有结合循环之间可以是恒定的、结合剂特异性的、或结合循环数特异性的(例如,用于"计时"结合循环)。在此实例中,编码标签包含提供用于结合剂(或一类结合剂,例如,全部特异性结合至末端氨基酸(诸如如图3所示的修饰的末端Q)的一类结合剂)的识别信息的编码器序列,任选的UMI和与记录标签上的互补间隔区序列杂交的间隔区序列,有助于将编码标签信息转移到记录标签(例如,通过引物延伸,在本文中也称为聚合酶延伸)。连接也可用于转移序列信息,在这种情况下,可使用间隔区序列,但不是必需的。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1580] 图2A示出了通过同源结合剂与分析物(例如,蛋白质或蛋白质复合物)结合产生延伸记录标签的过程,以及相应信息从结合剂的编码标签到分析物记录标签的转移。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,结合剂编码标签信息包括编码器序列,其来自于为结合剂(例如,抗体1(Ab1),抗体3(Ab3),...抗体"n"(Abn))提供识别信息的"n"结合周期、来自于记录标签的条形码/可选UMI序列、来自于结合剂编码标签的可选UMI序列,以及在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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1581] 图2B示出了用DNA条形码化的记录标签标记蛋白质的示意性示例。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"偶联到固体基板的表面的张紧性炔烃。在该方案中,记录标签通过NHS基团与蛋白质的赖氨酸(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,DNA记录标签也可以设计有正交的甲基四嗪(mTet)基团,用于通过逆电子需求Diels-Alder(iEDDA)反应下游偶联到反式环辛烯(TCO)衍生的测序基板。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1582] 图2C示出了采用记录标签的蛋白质分析方法的示例。在顶部附图中,分析物,例如蛋白质大分子,通过捕获剂和任选地的通过交联固定在固体支持物上。蛋白质或捕获剂可以与记录标签共定位或采用记录标签标记。在底部附图中,记录标签相关的蛋白质直接固定在固体支持物上。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、

隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1583] 图2D示出了使用同源结合物的DNA编码和产生的延伸记录标签的测序的简单蛋白质免疫侧定的整体工作流程实例。蛋白质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。该方法是用于进行反相蛋白质测定(RPPA)的有效的数字的、更简单的和更可缩放的方法,允许以定量的方式同时测量大量生物样品中的蛋白质水平(例如表达水平)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1584] 图3A-D示出了通过构建代表多肽序列的延伸记录标签(例如,DNA序列)进行基于降解的多肽序列测定的流程。这通过使用循环过程的埃德曼降解样方法(例如,N-末端氨基酸(NTAA)结合)实现,将编码标签信息转移至附着于多肽的记录标签,末端氨基酸裂解(例如NTAA裂解)和以循环方式重复所述方法,例如全部在固体支持物上。提供了从肽的N-末端降解构建延伸记录标签的示例性结构:(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,珠粒)并与记录标签(例如,通过三官能接头)相关的多肽,其中在结合剂结合到多肽的NTAA时,编码标签的信息被转移到记录标签(例如,通过引物延伸,包括单链连接或双链连接或平端连接或粘性端连接)以产生延伸的记录标签;(D)通过化学或生物学(例如,酶促)方式裂解标记的NTAA以暴露新的NTAA。如箭头所指,循环重复“n”次以产生最终的延伸记录标签。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。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,并且反向通用引发位点(例如,Illumina的P7-S2’序列)可以在记录标签的延伸的最后一步添加。最后一步可以独立于结合剂完成。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1585] 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。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延伸记录标签库的拷贝,等等。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1586] 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转移到与分析物,

比如大分子(例如多肽)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区序列(Sp')侧接在每个末端。图5A示出了一种NTAA结合剂,该结合剂包含与多肽的NTAA结合的编码标签,该多肽被标记有记录标签并且连接至珠粒。记录标签通过互补间隔区序列(Sp退火至Sp')与编码标签退火,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记录标签的转移。编码标签示为在远离结合剂的末端具有单链间隔区(Sp')序列的双链体。该配置使编码标签与记录标签中的内部位点的杂交最小化,并有利于记录标签的末端间隔区(Sp)序列与编码标签的单链间隔区突出(Sp')的杂交。此外,延伸记录标签可以与一个或多个寡核苷酸(与编码器、间隔区序列互补)预退火以阻止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)和编码标签信息的转移以及在3'-端添加通用引发位点。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1587] 图6示出了通过酶连接将编码标签信息转移到延伸记录标签。显示了两种不同的分析物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,使得间隔区序列(Sp')具有与记录标签上的互补间隔区序列(Sp)退火的在一个链上的"粘性末端"突出。"粘性末端"的长度可以为0-8个碱基,例如约2-4个碱基。双链编码标签的互补链在与记录标签连接后,将信息转移给记录标签。互补链可包含另一间隔区序列,其可与连接前记录标签的Sp相同或不同。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'。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1588] 图7示出了通过化学连接将编码标签信息转移到记录标签,以将记录标签或延伸记录标签的3'核苷酸连接到编码标签(或其互补体)的5'核苷酸而不插入间隔区序列到延伸记录标签的"无间隔"方法。延伸记录标签和编码标签的方向也可以转换,以便于将记录标签的5'端连接到编码标签(或互补体)的3'端。在显示的示例中,记录标签上的互补"辅助"寡核苷酸序列("记录辅助")和编码标签之间的杂交用于稳定复合物,以促使记录标签与编码标签互补链的特异性化学连接。产生的延伸记录标签缺乏间隔区序列。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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1589] 图8A-B示出了在N-末端氨基酸降解前,将多肽的翻译后修饰(PTM)写入延伸记录标签的示例性方法。图8A:包含具有关于结合剂(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多肽。如果磷

酸酪氨酸存在于记录标签标记的多肽中,如图所示,当磷酸酪氨酸抗体与磷酸酪氨酸结合后,编码标签和记录标签通过互补间隔区序列退火并且编码标签信息被转移到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肽的一级氨基酸序列(例如,“$aa_1$”,“$aa_2$”,“$aa_3$”…,“$aa_N$”)和翻译后修饰(例如,$PTM_1$,“$PTM_2$”)的编码标签信息。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1590] 图9A-B示出了结合剂与分析物(例如,多肽等大分子)结合并将附着于结合剂的编码标签的信息转移到多个记录标签中的单个记录标签(例如共定位在附着于固体支持物(例如,珠粒)位点处的单个分析物上)的多个循环的过程,从而产生多个延伸记录标签,其共同表示分析物信息(例如,样品中存在或不存在、水平或量,与结合剂库的结合情况,活性或反应性,氨基酸序列,翻译后修饰,样品来源,或其任何组合)。在这个附图中,仅出于举例的目的,分析物是多肽,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转移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示出了在固体支持物上的多个记录标签(例如,包含通用正向引发序列和UMI),其可用于与分析物结合的结合剂。单个的记录标签具有与结合剂的编码标签内的共同间隔区序列互补的共同间隔区序列(Sp),其可用于引发延伸反应以将编码标签信息转移至记录标签。例如,多个记录标签可以与分析物共定位在支持物上,并且一些记录标签可以比其它记录标签更靠近分析物。在一个方面,可以相对于支持物上的分析物密度来控制记录标签的密度,使得按照统计,每种分析物将具有可用于结合至分析物的结合剂的多个记录标签(例如,至少约2个,约5个,约10个,约20个,约50个,约100个,约200个,约500个,约1000个,约2000个,约5000个,或更多)。该模式可特别用于分析样品中的低丰度蛋白质或多肽。尽管图9A显示不同的记录标签在循环1-3的每一个中被延伸(例如,在每个结合/反应循环中的结合剂中的循环特异性条形码或单独添加的循环特异性条形码可用于结合/反应“计时”),设想延伸记录标签可在任何一个或多个后续结合循环中进一步扩展,并且所得到的延伸记录标签池可以是仅扩展一次、两次、三次或更多次的记录标签的混合。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1591] 图9B示出了用于每个连续结合循环的循环特异性NTAA结合剂的不同池,每个池具有循环特异性间隔区序列。或者,循环特异性序列可在与结合剂分开的试剂中提供。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1592] 图10A-C示出了包含将附着于结合剂的编码标签的信息转移到多个记录标签中的单个记录标签从而产生共同代表分析物的多个延伸记录标签的多个循环的示例模型,多个记录标签共定位于单个分析物的位点,单个分析物附着于固体支持物(例如,珠粒)。在这个附图中,仅出于举例的目的,分析物是多肽,并且每个循环过程涉及结合至NTAA、记录结合事件,然后移除NTAA以曝露新的NTAA。图10A示出了与分析物在固体支持物上的多个记录标

签(包含通用的正向引发序列和UMI),例如,每珠粒一个单分子。单个记录标签在它们的3'-末端具有不同的间隔区序列,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,每个珠粒上的记录标签分享相同的隔室条形码和/或UMI序列。在第一结合循环(循环1)中,使多个NTAA结合剂分析物接触。循环1中的结合剂具有共同的5'-间隔区序列($C'_1$),其与记录标签的循环1$C_1$间隔区序列互补。循环1中的结合剂也具有共同的3'-间隔区序列($C'_2$),其与循环2间隔区$C_2$互补。在结合循环1期间,第一NTAA结合剂结合到分析物的自由N-末端,并且第一编码标签的信息通过从杂交的$C_1$序列到互补的$C'_1$间隔区序列的引物延伸转移到同源记录标签。除去NTAA以暴露新的NTAA之后,结合循环2使多个NTAA结合剂与分析物接触,NTAA结合剂具有与循环1的3'-间隔区序列相同的循环2的5'-间隔区序列($C'_2$)以及共同的循环3的3'-间隔区序列($C'_3$)。第二NTAA结合剂结合到分析物的N-末端,并且第二编码标签的信息通过互补的$C_2$和$C'_2$间隔区序列的引物延伸转移到同源记录标签。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,产生与单个分析物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编码标签中具有循环特异性间隔区序列,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。图10B示出了用于每个连续结合循环的循环特异性NTAA结合剂的不同池,每个池具有循环特异性间隔区序列。图10C示出了共定位在分析物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区序列按顺序基于PCR组装的所述延伸记录标签,从而提供有序的分析物(例如,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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1593] 图11A-B示出了信息从记录标签转移到编码标签或双标签构建体。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。对于延伸编码标签或双标签构造,不是将结合信息从编码标签转移到记录标签,而是将信息从记录标签转移到编码标签以产生延伸编码标签(图11A),或者将信息从记录标签和编码标签转移到第三双标签形成构建体(图11B)。

[1594] 双标签和延伸编码标签包含记录标签(包含条形码,可选的UMI序列,可选的隔室标签(CT)序列(未在图中示出))和编码标签的信息。双标签和延伸编码标签可从记录标签上被洗脱、收集,并可选地被扩增并在下一代测序仪上被读出。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1595] 图12A-D示出了结合事件的PNA组合条形码/UMI记录标签和双标签检测的设计。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区组合模板。采用化学连接将退火的PNA"单词"缝合在一起。图12B示出了将记录标签的PNA信息转移到DNA中间体的方法。DNA中间体能够将信息转移给编码标签。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,

DNA中间体设计来通过间隔区序列SP与编码标签互作。链置换引物延伸步骤取代连接的DNA并将记录标签信息从DNA中间体转移到编码标签以产生延伸编码标签。终止子核苷酸可以并入DNA中间体的末端以防止编码标签信息通过引物延伸转移到DNA中间体。图12D:或者,可以将信息从编码标签转移到DNA中间体以产生双标签构建体。终止子核苷酸可以并入DNA中间体的末端以防止记录标签信息从DNA中间体转移到编码标签。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1596] 图13A-E示了隔室条形码珠粒上的蛋白质组分区,随后通过乳化融合PCR进行双标签组装,以产生代表多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)来表征多肽的氨基酸含量。编码标签包含通用引发位点、以及用于氨基酸识别的编码器序列,隔室标签,和氨基酸UMI。在信息转移之后,通过记录标签UMI将双标签映射回原始分子。在图13A中,蛋白质组被条形码化的珠粒拆分为液滴。具有相关记录标签(包括隔室条形码信息)的肽附着在珠粒的表面上。液体乳液打破后释放具有分区肽的条形码化珠粒。在图13B中,在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记基团的DNA编码标签进行化学标记。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-记录标签复合物从珠粒上释放。图13D: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)转移到所有与氨基酸残基连接的DNA编码标签上。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1597] 图14示出了延伸编码标签从乳化肽记录标签-编码标签复合物产生。来自图13C的肽复合物与PCR试剂共乳化成为平均每个液滴具有单个肽复合物的液滴。使用三引物融合PCR方法扩增与肽相关的记录标签,将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,通过引物延伸延伸编码标签从而转移肽UMI和记录标签的隔室标签信息到编码标签,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,在肽中的氨基酸身份和数量都能被确定。U1通用引物和Sp引物设计来比$U2_{tr}$通用引物具有更高的熔化$T_m$。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记录标签,然后进行到较低的Tm,使得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U1和$U2_{tr}$通用引物用于引发所得的延伸编码标签产物的扩增。在一些实施例中,来自$U2_{tr}$引物的合成酶延伸可以采用光标记3'封闭基团来防止(Young等人,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(其中编码标签Sptr在扩增的记录标签的Sp'序列上引发编码标签的延伸)之后,$U2_{tr}$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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环

特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1598] 图15示出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,蛋白质通常被消化成肽。在该过程中,源自亲本蛋白分子的各个多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质序列的集合。采用短和/或部分肽序列,并且随着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,随后将蛋白质消化成肽,并且隔室标签连接到肽降低了肽序列需要映射上去的"蛋白质"空间,在复杂蛋白质样品的情况下极大地简化了任务。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形式的标识。图15A示出蛋白质组分区的示例,其包含用含有隔室或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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1599] 图16示出了隔室标记的珠粒序列设计的示例性模式。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)的隔室标签中。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1600] 图17A-B示出了:(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库的延伸记录标签库。诱饵寡核苷酸:靶延伸记录标签杂交对在杂交后通过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。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性珠粒完成。生物素基团与磁性珠粒上的链霉亲和素结合,并且通过使用磁体固定磁性珠粒完成分离同时移除或更换溶液。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。非生物素化的竞争寡核苷酸竞争杂交到靶肽,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富集的延伸记录标签分级物。该步骤对于解决样品中蛋白质丰度的动态范围问题非常重要。在该图中所示的任何示例中,编码标签和/或记录标签

(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1601] 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,每个液滴包含珠粒,珠粒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。隔室标签包含条形码。液滴形成后的液滴成分的操纵:(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,以及蛋白水解以将细胞蛋白质组消化成肽,并在充分蛋白水解后使蛋白酶失活;(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蛋白复合物消化成肽,并在充分蛋白质水解后灭活蛋白酶。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蛋白酶,或者可以螯合二价阳离子。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)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1602] 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,每个液滴包含珠粒,珠粒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,或将蛋白质与起源单细胞相关联。在液滴形成后操纵液滴成分:(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,以及蛋白水解以将细胞蛋白质组消化成肽,并在充分蛋白水解后使蛋白酶失活;(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蛋白复合物消化成肽,并在充分蛋白质水解后灭活蛋白酶。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蛋白酶,或者可以螯合二价阳离子。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)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1603] 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,在微流体液滴中的条形码化的珠粒,等等)可以用于将隔室特异性条形码转移至封装在特定隔室内的分子内容物。(A)在一个具体的实施例中,蛋白质分子是变性的,并且赖氨酸残基(K)的ε-胺基与活化的通用DNA标签分子化学缀合(包含通用引发序列(U1),在5'端显示有NHS基团,但也可以使用任何其他生物缀合基团)。在通用DNA标签与多肽缀合后,移除过量的通用DNA标签。(B)使通用DNA标记多肽与结合在珠粒上的核酸分子杂交,其中结合在单个珠粒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)发生隔室化。或者,可以通过将标记的多肽固定在珠粒表面上来直接完成隔室化,例如,通过将多肽上的通用DNA标签与珠粒上的隔室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粒相互作用(例如,单个多肽不跨越多个珠粒)。然而,多种多肽可以与同一个珠粒相互作用。除了隔室条形码序列(BC)之外,结合至珠粒的核酸分子可以由共同的Sp(间隔区)序列,独特的分子标识符(UMI标签)和与多肽DNA标签U1'互补的序列组成。(C)在将通用DNA标记的多肽退火至结合至珠粒的隔室标签之后,经由连接接头的裂解从珠粒释放隔室标签。(D)以来源于珠粒的隔室标签核酸分子为模板,通过基于聚合酶的引物延伸对退火后的U1 DNA标签引物进行延伸。如(C)中所述引物延伸步骤可以在隔室标签从珠粒释放之后进行,或者任选地,在隔室标签仍附着珠粒(未示出)时进行。这有效地将珠粒上的隔室标签上的条形码序列写到多肽上的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(裂解赖氨酸残基的C-末端侧),Glu-C(裂解谷氨酸残基的C-末端侧并且在较低程度上裂解谷氨酸残基的C-末端侧),或随机的蛋白酶如蛋白酶K,用于将多肽裂解成肽片段。(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)记录标记的肽通过张紧性炔标记DBCO与叠氮化珠粒偶联。叠氮化珠粒可选地还含有与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,肽从原始珠粒中移除并重新固定到新的固体支撑物(例如,珠粒)让肽之间形成最佳分子间间距,有助于实现本文公开的肽测序方法。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1604] 除了使用点击化学缀合DNA标签与炔烃预标记的多肽(如图20A-F中所述),图20G-L说明了与图20-F中所示类似的概念。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)到测序基板。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1605] 图21示出了使用用于单细胞的流式聚焦T形结构(T-junction)和采用珠粒的隔室标记(例如,条形码)隔室化的示例性方法。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)。

[1606] 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选择性地裂解N-末端到天冬氨酸残基的肽键,可用于在信息转移至记录标签后裂解隔室标签。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1607] 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$BC_{ij}$表示的特征条形码)阵列与组织切片(FFPE或冷冻)组合。在一个实施例中,组织切片是固定的并且透化的。在优选的实施例中,阵列特征尺寸小于细胞大小(对于人细胞为约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条形码序列被裂解并扩散到阵列式组织切片中并与其中的附着在蛋白质上的DNA记录标签杂交。(C)现在对阵列式组织进行聚合酶延伸,以将杂交的条形码的信息转移到标记蛋白质的DNA记录标签上。在转移条形码信息后,从载玻片上刮下阵列式组织,任选地用蛋白酶消化,并将蛋白质或肽提取到溶液中。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1608] 图24A-B示出两种不同的示例性DNA靶标分析物(AB和CD),其固定在珠粒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,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。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。如果增加第二循环,则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,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,延伸记录标签包含结合事件次序以及结合剂的身份的信息。类似地,图24B示出了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。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。C’和D’结合剂的结合被示为单个结合循环的替代可能性。如果增加第二循环,则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,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转移到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转移事件。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1609] 图25示出了示例性DNA靶大分子(AB)并固定在珠粒上,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。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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1610] 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-肽分析物(例如,大分子)。图26A示出了固定在珠粒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)与大分子的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示出了固定在珠粒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,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大分子的A和C寡核苷酸组分。因此,可以通过将信息转移到记录标签并读出延伸记录标签来分析包含不同类型骨架的杂合大分子,延伸记录标签含有关于结合事件的次序信息以及结合剂的身份。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1611] 图27A-D示出了纠错条形码的生成的示例。(A)采用所述推荐参数[create.dnabarcodes(n＝15,dist＝10)]从R软件包“DNABarcodes”(https://bioconductor.riken.jp/packages/3.3/bioc/manuals/DNABarcodes/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65个条形码的子集。(B)15-mer条形码通过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纳米孔时使用5-mer R9纳米孔电流水平查找表(template_median68pA.5mers.model,查阅https://github.com/jts/nanopolish/tree/master/etc/r9-models)来预测对应的电流水平。如可以从(B)中理解的,该组65个条形码对于其成员中的每一个展现出独特的电流特征。(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。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间隔区元件加下划线。条形码可以正向或反向排列,用BC或BC'表示。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1612] 图28A-D示出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($BC_A'$)杂交的分析物特异性条形码($BC_A$)。或者,DNA记录标签可以通过可裂解的接头连接在所述结合剂上,并且DNA记录标签直接被"点击"到所述蛋白质上,然后从结合剂上被切除(通过可裂解的接头)。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($BC_S$),分析物特异性条形码($BC_A$)和间隔区序列(Sp))。样品条形码($BC_S$)也可用于标记和区分不同样品的蛋白质。DNA记录标签也可以包含正交偶联基团(例如mTet),用于随后偶联到基板表面。对于记录标签与目标蛋白质的点击化学偶联,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质后,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)在用记录标签标记靶蛋白质分析物后,通过采用尿嘧啶特异性切除试剂(例如USERTM)消化尿嘧啶(U)除去连接的结合剂。因此,本发明要求保护的试剂盒还可以包含切刻试剂,例如切刻核酸内切酶。(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DNA记录标签标记的靶蛋白质分析物固定到基板表面。在某些实施例中,使用结合剂库和记录标签库在包含许多不同靶蛋白质分析物的单个管中进行整个靶蛋白质-记录标签标记测定。在用包含样品条形码($BC_S$)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,在某些实施例中,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)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1613] 图29A-E示出了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。记录标签包括通用引物序列(P1)、条形码(BC)和间隔区序列(Sp)。记录标签用mTet基团标记以偶联到基板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质或肽通过炔烃和叠氮基团用记录标签标记。任选地,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

到隔室珠粒和引物延伸(也称为聚合酶延伸)的互补序列退火,或如图20H-J所示。(D)蛋白酶消化记录标签标记的多肽产生记录标签标记的肽群。在一些实施例中,一些肽不会被任何记录标签标记。在其他实施例中,一些肽可以附着一个或多个记录标签。(E)使用TCO基团官能化的基板表面和与肽连接的记录标签的mTet基团之间的逆电子需求Diels-Alder(iEDDA)点击化学反应,将记录标签标记的肽固定在基板表面上。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板表面上(参见,McKay等人,2014,Chem.Biol.21:1075-1101,其全部内容通过引用并入本文)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1614] 图30A-E示出了在多肽的初始DNA标签标记后将样品条形码写入记录标签的示例性过程。(A)用双官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-炔相互作用与多肽偶联。应理解,可以采用其他点击化学相互作用。(C)包含样品条形码信息($BC_S$')和其它记录标签功能组分(例如,通用引物序列(P1'),间隔区序列(Sp'))的记录标签DNA构建体通过互补的通用引发序列(P1-P1')与DNA标签标记的多肽退火。记录标签信息通过聚合延伸转移到DNA标签。(D)蛋白酶消化记录标签标记的多肽产生记录标签标记的肽群。(E)使用TCO基团官能化的表面和与肽连接的记录标签的mTet基团之间的逆电子需求Diels-Alder(iEDDA)点击化学反应,将记录标签标记的肽固定在基板表面上。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多肽到基板表面上(参见,McKay等人,2014,Chem.Biol.21:1075-1101,其全部内容通过引用并入本文)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1615] 图31A-E示出用于条形码多肽的珠粒隔室化。(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标签与标记的多肽的炔基团偶联。(C)通过互补DNA序列(P1和P1')将DNA标签标记的多肽与DNA记录标签标记的珠粒退火。珠粒上的DNA记录标签包含间隔区序列(Sp'),隔室条形码序列($BC_P$'),可选的独特分子标识符(UMI)和通用序列(P1')。DNA记录标签信息通过聚合酶延伸(或者,可以采用连接)转移到多肽上的DNA标签上。在信息转移后,所得多肽包含多个记录标签,其包含包括隔室条形码在内的若干功能元件。(D)蛋白酶消化记录标签标记的多肽产生记录标签标记的肽群。记录标签标记的肽与珠粒解离,并且(E)重新固定到测序基板上(例如,如图所示,使用mTet和TCO基团之间的iEDDA点击化学)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签

或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1616] 图32A-H示出下一代蛋白质测定(NGPA)的工作流程的示例。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序列和间隔区序列(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)将标记的蛋白被固定(被动地或共价地)到基板(例如珠粒、多孔珠粒或多孔基质)上。(B)基板用蛋白封闭,并任选地加入与间隔区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)与基板结合蛋白一起孵育。编码标签可以包含尿嘧啶碱基,用于随后的尿嘧啶特异性裂解。(D)抗体结合后,洗去过量的竞争寡核苷酸(Sp'),若加入的话。编码标签通过互补间隔区序列瞬时退火到所述记录标签,并且编码标签信息在引物延伸反应中转移到记录标签以产生延伸记录标签。如果固定的蛋白质是变性的,则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)在信息从编码标签转移到记录标签之后,使用尿嘧啶特异性切除试剂(例如USER™)酶混合物在其尿嘧啶位点切割(裂解)编码标签。(F)使用高盐、低/高pH洗涤液从蛋白中去除结合的抗体。保留在抗体上的截短的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编码标签信息从第二抗体转移到延伸的记录标签。(H)两个结合循环的结果是来自于附着在记录标签的第一抗体的结合信息和第二抗体的结合信息的级联。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1617] 图33A-D示出使用多种结合剂和酶促介导的顺序信息转移的单步下一代蛋白测定(NGPA)的示例。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编码标签的信息转移至记录标签。编码标签中的插入符(^)代表双链DNA切刻核酸内切酶位点。在图33A中,与蛋白质的表位1(Epi#1)结合的抗体的编码标签在互补间隔区序列杂交后的引物延伸步骤中将编码标签信息(例如,编码区序列)转移至记录标签。在图33B中,一旦延伸记录标签和编码标签之间形成双链DNA双体,缺口内切酶,例如Nt.BsmAI,其在37℃时有活性,仅裂解双链DNA底物上的单链DNA,用于裂解编码标签。在缺割步骤之后,由截短的编码标签结合剂和延伸记录标签形成的双体是热力学不稳定的且解离的。较长的编码标签片段可以与记录标签保持或不保持退火。在图33C中,这让来自与蛋白质的表位#2(Epi#2)结合的抗体的编码标签通过互补间隔区序列与延伸记录标签退火,并且延伸记录标签通过经引物延伸从Epi#2抗体的编码标签到延伸记录标签的信息传递来进一步延伸。在图33D中,再次,在Epi#2抗体的延伸记录标签和编码标签之间形成双链DNA双链体后,编码标签被切刻核酸内切酶,如Nb.bsssi切刻。在某些实施例中,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裂解的编码标签短柱延伸,其通过一个以

上单碱基与记录标签保持退火。图A-D的过程可以自己重复,直到近端结合的结合剂的所有编码标签被杂交、信息转移到延伸记录标签,以及切刻步骤“消耗”。编码标签可以包含与对给定分析物(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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1618] 图34A-C示出使用滴定基板表面上的反应性部分的记录标签-肽固定的控制密度的示例。在图34A中,可以通过控制基板表面上的功能性偶联部分的密度来滴定基板表面上的肽密度。这可以通过活性偶联分子和“虚拟”偶联分子的适当比例对基板表面进行衍生来实现。在所示的示例中,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。官能化的PEG有不同的分子量,从300到40000多。在图34B中,双功能5’胺DNA记录标签(mTet是其它功能部分)使用琥珀酰亚胺基4-(N-马来酰亚胺甲基)环已烷-1(SMCC)双功能交联剂与肽的N-末端Cys残基偶联。使用mTetrazine-Azide从叠氮化物-dT基团创建记录标签上的内部mTet-dT基团。在图34C中,使用与mTet和TCO的iEDDA点击化学反应将记录标签标记的肽固定到图34A的活化的基板表面上。mTet-TCO iEDDA偶联反应非常快速,高效且稳定(mTet-TCO比Tet-TCO更稳定)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1619] 图35A-C示出了下一代蛋白测序(NGPS)结合循环特异性编码标签的示例。(A)设计具有循环特异性N-末端氨基酸(NTAA)结合剂编码标签的NGPS测定法(本文也称为ProteoCode)。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,肽与包含通用引物序列(P1)、条形码(BC)和间隔区序列(SP)的记录标签相关。当结合剂结合肽的同源NTAA时,与NTAA结合剂相关的编码标签接近记录标签并通过互补间隔区序列与记录标签退火。编码标签信息通过引物延伸转移到记录标签。为了跟踪编码标签代表哪一个结合循环,编码标签可以包括循环特异性条形码。在某些实施例中,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。在其它实施例中,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,共同的间隔区序列可用于在每个结合循环中结合剂的编码标签。(B)在该示例中,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,其与识别结合剂的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)在结合循环完成后,使用加帽循环步骤可将延伸记录标签转化为可扩增的库,其中例如,包含连接到通用引物序列P2的通用引物序列P1和间隔区序列Sp’的帽最初通过互补的P1和P1’序列与延伸记录标签退火,使帽接近延伸记录标签。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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可

测序聚合物,例如非核酸可测序聚合物。

[1620] 图36A-F示出了用于演示从编码标签到记录标签的信息转移的基于DNA的模型系统的示例。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碱基共同间隔区序列(Sp)。编码标签包含独特的编码器条形码碱基,其侧翼为8个碱基共同间隔区序列(Sp'),其中一个通过聚乙二醇接头共价连接到A或B靶试剂。在图36A中,将生物素化的记录标签寡核苷酸(saRT_Abc_v2和saRT_Bbc_V2)与生物素化的Dummy-T10寡核苷酸一起固定到链霉亲和素珠粒上。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)以识别结合靶标。在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。还合成了与记录标签的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珠粒固定的记录标签之前预先退火至编码标签。链置换聚合酶在聚合酶延伸期间去除封闭寡核苷酸。条形码图例(插图)表明对记录标签和编码标签中的功能条形码的15-mer条形码分配。在图36B中,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。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)。在图36C中,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子序列。图36D显示通过使用Dummy-T20寡核苷酸滴定记录标签的表面密度优化"分子内"信息转移。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。在图36E中,作为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($K_D$约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-$Tag_{15}$肽序列是(fM)DVEAWLGARVPLVET(SEQ ID NO:131)(fM=甲酰基-Met)。Nano-$Tag_{15}$肽还包含短的、柔性的接头肽(GGGGS)和用于偶联DNA记录标签的半胱氨酸残基。其它示例性肽标签-同源结合剂对包括:钙调蛋白结合肽(CBP)-钙调蛋白($K_D$约2pM)(Mukherjee等人,2015,J.Mol.Biol.427:2707-2725)、淀粉样蛋白-β(Aβ16-27)肽-US7/Lcn2 anticalin(0.2nM)(Rauth等人,2016,Biochem.J.473:1563-1578)、PA标签/NZ-1抗体($K_D$约400pM)、FLAG-M2 Ab(28nM)、HA-4B2 Ab(1.6nM)和Myc-9E10 Ab(2.2nM)(Fujii等人,2014,Protein Expr.Purif.95:240-247)。图36F作为通过引物延伸对分子内信息从结合剂编码标签向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-$tag_{15}$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)。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1621] 图37A-B示出了使用纳米或微乳液PCR将信息从UMI标记的N或C-末端转移至标记多肽体的DNA标签的示例。在图37A中,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。UMI可以侧接用于引发后续PCR的序列。然后多肽用单独的DNA标签在其内部位点进行“身体标记”,单独的DNA标签包含与UMI侧翼的引发序列互补的序列。在图37B中,将得到的标记多肽乳化并进行乳液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-末端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。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UMI信息通过引物延伸从扩增子转移至内部多肽体DNA标签。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1622] 图38示出了单细胞蛋白质组学的示例。在含有聚合物形成亚基(例如丙烯酰胺)的液滴中包封并裂解将细胞。聚合聚合物形成亚基(例如,聚丙烯酰胺),且将蛋白质交联到聚合物矩阵。破坏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),间隔区序列(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。含有单细胞蛋白质的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。在某些实施例中,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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1623] 图39A-E示出了使用双功能N-末端氨基酸(NTAA)修饰剂和嵌合裂解试剂增强氨基酸裂解反应的实例。(A)和(B)连接到固相底物上的肽用双功能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Edmanase裂解的效率大大提高。(E)裂解的生物素-PITC标记的NTAA和相关的链霉亲和素-Edmanase嵌合蛋白在裂解后扩散开。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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1624] 图40A-I示出了从蛋白质裂解物产生C-末端记录标签-标记的肽(可以包封在凝胶珠粒中)的示例。

[1625] (A)使变性多肽与酸酐反应以标记赖氨酸残基。

[1626] 在一个实施例中,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(在多肽链上显示为正方形)封闭。

[1627] 炔(mTet)部分可用于基于点击化学的DNA标记。(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。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)将包含与记录标签中的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。C-末端DNA标签还包含用于将所述肽缀合至表面的基团。(H)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。

[1628] 分子内引物延伸反应用于将信息从记录标签转移至C-末端DNA标签。(I)通过马来酸酐将肽上的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。内部记录标签在酸性pH下从肽的赖氨酸残基上裂解,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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1629] 图41示出了NGPS测定的实施例的示例性工作流程。

[1630] 图42A-D示出了NGPS测序测定的示例性步骤。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NTAA。在第一种情况下,(A)肽最初在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)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编码标签附着于结合剂并包含识别特定的NTAA接合剂的条形码编码序列。(C)在通过NTAA结合剂结合乙酰化NTAA后,DNA编码标签通过互补序列与记录标签瞬时退火,并且编码标签信息通过聚合酶延伸转移至记录标签。在另一个实施例中,通过聚合酶延伸将记录标签信息转移到编码标签。(D)通过工程化的酰基肽水解酶(APH)从肽上裂解乙酰化的NTAA,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。在乙酰化的NTAA裂解后,循环从新暴露的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/裂解的示例性模式,但是其他N-末端基团,例如脒基基团,可以被裂解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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1631] 图43A-B示出了示例性记录标签-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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1632] 图44A-E示出了使用裂解剂与记录标签(A)和(B)的杂交增强NTAA裂解反应的示例。在图44A-B中,附着于固相底物(例如珠粒)的记录标签标记的肽在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。在图44C中,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。通过裂解酶的DNA标签和记录标签上的互补通用引物序列的杂交,将裂解酶募集到修饰的NTAA中。在图44D中,杂交步骤大大提高了裂解酶对NTAA的有效亲和力。在图44E中,裂解的NTAA扩散开,并且可以通过剥离杂交的DNA标签除去相关的裂解酶。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1633] 图45示出了使用肽连接酶+蛋白酶+二氨基肽酶的示例性循环降解肽测序。Butelase I将TEV-Butelase I肽底物(TENLYFQNHV,SEQ ID NO:132)连接至待测肽的NTAA。Butelase需要肽底物的C-末端处的NHV基序。连接后,使用烟草蚀纹病毒(TEV)蛋白酶在谷氨酰胺(Q)残基后裂解嵌合的肽底物,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-末端裂解两个氨基酸残基,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。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1634] 图46A-C示出了通过单链DNA编码标签与单链DNA记录标签连接的示例性"无间隔"

编码标签转移。通过将编码标签连接到记录标签来直接转移单链DNA编码标签,以产生延伸记录标签。(A)经由单链DNA连接的基于DNA的模型系统的概述设计与编码标签缀合的靶向剂B'序列,用于检测记录标签中的B DNA靶。ssDNA记录标签saRT_Bbca_ssLig是5'磷酸化和3'生物素化的,并且由6碱基DNA条形码BCa、通用正向引物序列和靶DNA B序列组成。编码标签CT_B'bcb_ssLig含有通用反向引物序列、尿嘧啶碱基和独特的6碱基编码器条形码BCb。编码标签与B'DNA序列通过聚乙二醇接头共价连接。附着于编码标签的B'序列与附着于记录标签的B序列的杂交使记录标签的5'磷酸基和编码标签的3'羟基在固体表面上紧密接近,导致通过用连接酶如CircLigase II的单链DNA连接进行信息转移。(B)凝胶分析确认单链DNA连接。单链DNA连接测定显示结合信息从编码标签转移到记录标签。47个碱基记录标签与49个碱基编码标签的连接产物大小为96个碱基。如果在同源saRT_Bbca_ssLig记录标签的存在下观察到连接产物带,而在非同源saRT_Abcb_ssLig记录标签的存在下没有观察到产物带,则证明了特异性。(C)编码标签的多周期信息转移。用USER Enzyme处理第一循环连接产物以产生游离的5'磷酸化末端,用于在第二循环信息转移中使用。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1635] 图47A-B示出了通过将双链DNA编码标签连接到双链DNA记录标签的示例性编码标签转移。用基于DNA的模型系统演示了编码标签通过双链DNA连接的多重信息转移。(A)经由双链DNA连接的基于DNA的模型系统的概述制备与编码标签缀合的靶向剂A'序列,用于检测记录标签中的靶结合剂A。记录标签和编码标签均由具有4个碱基突出的两条链组成。当编码标签中的靶向剂A'与固定在固体表面上的记录标签中的靶向剂A杂交时,两个标签的邻近突出端杂交,导致通过连接酶,例如T4 DNA连接酶,经由双链DNA连接进行信息转移。(B)凝胶分析确认双链DNA连接。双链DNA连接测定证明A/A'结合信息从编码标签转移到记录标签。76和54个碱基记录标签与双链编码标签的连接产物大小分别为116和111个碱基。第一循环连接产物用USER Enzyme(NEB)消化,并用于第二循环测定。第二循环连接产物带在约150个碱基处观察到。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1636] 图48A-E示出了示例性的基于肽和基于DNA的模型系统,用于演示从编码标签到具有多个循环的记录标签的信息转移。序列肽和DNA模型系统演示了多重信息转移。(A)基于肽的模型系统中第一循环的概述。制备编码标签缀合的靶向剂抗PA抗体,用于在第一循环信息转移中检测记录标签中的PA-肽标签。此外,还使用Nanotag肽或淀粉样蛋白β(Aβ)肽产生肽记录标签复合物阴性对照。记录标签amRT_Abc分别通过5'端的胺基和内部的炔基共价连接到肽和固相支持物上,amRT_Abc含有A序列靶试剂、聚dT、通用正向引物序列、独特的DNA条形码BC1和BC2,以及8个碱基共同间隔区序列(Sp)。编码标签amCT_bc5在5'端和3'端分别与抗体和C3接头共价连接,amCT_bc5含有侧翼为8个碱基共同间隔区序列(Sp)的独特编码条形码BC5'。当抗PA抗体与PA-标签肽-记录标签(RT)复合物结合时,通过聚合酶延伸完成从编码标签到记录标签的信息转移。(B)在基于DNA的模型测定中的第二循环的概述。

制备与编码标签连接的靶向剂A'序列,用于检测记录标签中的A序列靶试剂。编码标签CT_A'_bc13含有8个碱基的共同间隔区序列(Sp')、独特的编码条形码BC13'、通用反向引物序列。当A'序列与A序列杂交时,通过聚合酶延伸完成从编码标签到记录标签的信息转移。(C)用于PCR分析的记录标签扩增。使用P1_F2和Sp/BC2引物组,通过18个循环PCR来扩增固定的记录标签。在约56bp处观察到记录标签密度依赖性PCR产物。(D)PCR分析确认第一循环的延伸测定。使用P1_F2和Sp/BC5引物组,通过21个循环的PCR来扩增第一循环的延伸记录标签。通过对复合物进行不同密度滴定,在约80bp处观察到PA-肽RT复合物的第一循环的延伸产物的PCR产物的强带。对于纳米和A肽复合物,在最高复合物密度下也观察到小的背景带,表面上归因于非特异性结合。(E)PCR分析确认第二循环的延伸测定。通过21个循环PCR,使用P1_F2和P2_R1引物组来扩增第二延伸记录标签。所有固定肽的珠粒均在117个碱基对处观察到相对较强的PCR产物带,仅对应于原始记录标签(BC1+BC2+BC13)上的第二循环的延伸产物。仅当在测定中使用PA-标签固定珠粒时,在93个碱基对处观察到对应于第一循环的延伸记录标签(BC1+BC2+BC5+BC13)上的第二循环的延伸产物带。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1637] 图49A-B使用p53蛋白测序作为示例来示出蛋白质形式的重要性和测序读段(例如,使用单分子方法获得的那些)的稳健的可作图性。图49A的左图显示完整的蛋白质形式可以被消化成片段,每个片段可以包含一个或多个甲基化氨基酸、一个或多个磷酸化氨基酸,或没有翻译后修饰。翻译后修饰信息可与测序读段一起分析。右图显示了沿着蛋白质的各种翻译后修饰。图49B显示使用分区的映射读段,例如,读段"CPXQXWXDXT"(SEQ ID NO:170,其中X=任何氨基酸)在搜寻整个人蛋白质组后唯一映射回p53(在CPVQLWVDST序列,SEQ ID NO:169)。测序读段不必是长的,例如,约10-15个氨基酸序列可以给出足够的信息来识别蛋白质组内的蛋白质。测序读段可以重叠,并且重叠序列处的序列信息的冗余可以用于推断和/或验证整个多肽序列。

[1638] 图50A-C示出了使用mRNA显示用DNA记录标签来标记蛋白质或肽。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1639] 图51A-E示出了通过N-末端二肽与分区条形码标记的肽结合的单循环蛋白识别。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1640] 图52A-E示出了通过N-末端二肽结合剂与肽固定的分区条形码化珠粒的单循环蛋白质识别。在该图中所示的任何示例中,编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话),或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码器序列、结合循环特异性条形码)可包含或替换为可测序聚合物,例如非核酸可测序聚合物。

[1641] 图53A-B示出了跨越不同种类的细菌的ClpS同源物/变体,以及用于本公开中的示

例性ClpS蛋白质,例如

[1642] 来自登记No.4YJM的ClpS2,根癌土壤杆菌:

[1643] MSDSPVDLKPKPKVKPKLERPKLYKVMLLNDDYTPREFVTVVLKAVFRMSEDTGRRVMMTAHRFG SAVVVVCERDIAETKAKEATDLGKEAGFPLMFTTEPEE(SEQ ID NO:198);

[1644] 来自登录No.2W9R的ClpS,大肠杆菌:

[1645] MGKTNDWLDFDQLAEEKVRDALKPPSMYKVILVNDDYTPMEFVIDVLQKFFSYDVERATQLMLAV HYQGKAICGVFTAEVAETKVAMVNKYARENEHPLLCTLEKAGA(SEQ ID NO:199);

[1646] 和来自保藏No.3DNJ的ClpS,C.crescentus:

[1647] TQKPSLYRVLILNDDYTPMEFVVYVLERFFNKSREDATRIMLHVHQNGVGVCGVYTYEVAETKVAQ VIDSARRHQHPLQCTMEKD(SEQ ID NO:200)。图53A显示了来自聚类至99%同一性的612个不同细菌种类的ClpS氨基酸序列的分级聚类的树状图。图53B是图53A中三个物种的ClpSs之间的氨基酸序列同一性的表。农杆菌ClpS2与大肠杆菌的序列同一性小于35%。并且与新月柄杆菌ClpS的序列同一性小于40%。

[1648] 图54示出了通过单链DNA/聚合物嵌合体编码标签与单链DNA/聚合物嵌合体记录标签连接的聚合物编码标记转移的示例。通过DNA间隔区将编码标签连接到记录标签直接转移单链聚合物编码标签以产生延伸记录标签。(A)经由单链DNA间隔区连接的NTAA检测系统的概述与编码标签缀合的N-末端氨基酸(NTAA)结合剂被设计用于检测记录标签中肽的NTAA。单链DNA/聚合物嵌合体记录标签是5'磷酸化的,并且由用于纳米孔测序的DNA间隔区,聚合物条形码(例如,受保护的DNA、PNA、类肽等),通用聚合物接头(例如,胆固醇、生物素、系链DNA序列等)组成。间隔区长度可以是起始酶连接的1-6个碱基。DNA/聚合物嵌合体编码标签含有通用聚合物衔接子,聚合物条形码(例如,受保护的DNA、PNA、类肽等)和DNA间隔区。编码标签通过聚乙二醇接头与NTAA结合剂共价结合。附着在编码标签上的NTAA结合剂与附着在记录标签上的肽的NTAA的结合使记录标签的5'磷酸基和编码标签的3'羟基在固体表面上紧密接近,导致通过用CircLigase II进行单链DNA连接的信息转移。(B)经由双链DNA间隔区连接的NTAA检测系统的概述制备与DNA/聚合物嵌合体编码标签缀合的NTAA结合剂用于检测记录标签中肽的NTAA。单链DNA/聚合物嵌合体记录标签是5'磷酸化的,并且由用于纳米孔测序的DNA间隔区,聚合物条形码(例如,受保护的DNA、PNA、类肽等),通用聚合物接头(例如,胆固醇、生物素、系链DNA序列等)组成。间隔区长度可以是起始酶连接的5-20个碱基。DNA/聚合物嵌合体编码标签含有通用聚合物衔接头,聚合物条形码(例如,受保护的DNA、PNA、类肽等)和DNA间隔区。编码标签通过聚乙二醇接头与NTAA结合剂共价结合。附着在编码标签上的NTAA结合剂与附着在记录标签上的肽的NTAA的结合使记录标签的5'磷酸基和编码标签的3'羟基在固体表面上紧密接近,导致通过分裂DNA杂交然后用T4 DNA连接酶进行双链DNA连接的信息转移。

[1649] 图55示出了单分子电子测序的示例性方法。(A)离子电流封阻;(B)隧道电流;(C)具有NP/纳米带的面内跨导;(D)具有纳米带的面内跨导。电流封阻基团可以封阻离子电流,或可以介导隧道电流,或可以调节跨导,这取决于电子读出的方式(Heerema等人,2016,Nat Nanotechnol,11(2):127-136).

[1650] 图56示出了用于纳米孔/纳米间隙分析的编码聚合物序列的示例性设计。(A)编码的聚合物由包含代码的组合电流封阻基团(由阴影多边形表示)组成。当电流封阻基团通过

纳米孔时,它们使用多种计算和机器学习方法产生可以与聚合物编码相关联的特征电流封阻信号。“生物素”环构成可以通过结合剂结合的结合基团,结合剂使纳米孔中的DNA“停顿”,允许更准确地读出下游聚合物封闭基团封阻基团。在没有这种转移停顿的特征的情况下,聚合物可能过快地通过孔。在所示示例中,生物素可以通过链霉亲和素(SA)、单体链霉亲和素(mSA2)、抗生物素蛋白等结合。应该选择结合剂和结合条件,使得结合是亚稳定的,并且结合剂将在结合的毫秒内解离,或者可替换地,可以随着跨膜电压的增加而除去。(B)作为外源结合剂“停顿”通过孔的转运的替代,聚合物的内在亚稳定二级结构可用于使聚合物通过孔的转运停顿。同样,应选择二级结构的能量,使得停顿时间在毫秒级。(C)如果纳米孔读出电子产品足够快并且噪声足够低,则可以经由移位通过孔直接读出聚合物。

[1651] 图57示出了使用纳米孔技术的聚合物读出的示例。(A)基本纳米孔设置的插图,其中插入脂质双层膜中的α溶血素(αHly)孔相对于顺式膜电压保持在+100mV反式。这种膜电压极性将带负电的聚合物从顺式侧转移到反式侧。对于带正电的聚合物,膜电压可以反向的。(B)单个封阻基团的样本封阻电流特征。(C)不同纳米孔类型的比较结构:生物纳米孔MspA和aHly、合成纳米孔石墨烯和$SiN_2$。可以用所有孔类型测量封阻电流,条件是封阻基团(阴影多边形)的维度应该在孔径的数量级上。(Laszlo等人,2014,Nature Biotech.,32,829-833;Ying等人,2014,The Analyst,139,3826-3835。)

[1652] 图58示出了通过使具有结合特征的转运停顿来控制聚合物转运通过纳米孔的示例性方法。(A)在这个示例中,可测序聚合物被设计成具有散布有聚合物“捕捉物”(例如,结合至生物素(b)的单体链霉亲和素(mSA2))的当前封阻基团(IB基团),这些捕捉物是结合元件或固有的二级结构元件,它们瞬时使聚合物在孔中的转运停顿以允许充足的时间进行准确的电流封阻(IB)测量。(B)结合剂mSA2紧靠孔,最终在膜电位(即,180mV/cm)下解离。理想地,亚稳态结合在毫秒时间标度上解离。

[1653] 图59A-C示出了三个循环的结合/编码和NGS读出。(A)示出了三个连续的编码步骤。显示肽与记录标签(DNA RT)缀合并固定在磁性珠粒上。该肽含有PA表位。用三个不同循环的特异性条形码(BC4、BC5、BC13)标记的抗-PA抗体进行三个连续的测定循环,每个条形码用于不同的循环(分别为第1、第2和第3循环)。每个循环导致记录标签被相应的循环条形码延伸。(B)显示每个循环后延伸记录标签的PCR产物的分析的琼脂糖凝胶。道1=尺寸标准,道2-4分别对应于第1、第2和第3循环的产物。凝胶清楚地显示高产率编码。PCR使用KOD聚合酶进行24个扩增循环。(C)除凝胶分析外,通过NGS测序分析最终的“三循环”产物(凝胶中的第3道)。初级产物(超过78%的产物)对应于凝胶上的最大带,由所有三种条形码的库元件组成。观察到单和双条形码产物比例较低(2个BC在20.7%,1个BC在1.1%)。

[1654] 图60A-B示出了封闭寡核苷酸在引物延伸编码测定中的应用。在基于引物延伸的编码测定中测试了几种类型的封闭寡核苷酸。(A)示出了用于封闭编码测定的组分的DNA“结合剂”模型系统和三种类型的封闭寡核苷酸的设计:

[1655] RT阻断剂、CT阻断剂和Sp'阻断剂。每个周期的编码标签条形码被设计为周期特异性的并且在周期之间正交(非交互)。(B)编码和PCR扩增3个循环后的编码结果的凝胶分析。在所示的示例中,在每个周期上与输入编码标签同源的CT阻断剂(用*标记的道)提供最高的编码效率,如箭头表示的上带所示。

[1656] 图61A-B示出了减少ProteoCode库扩增过程中的模板转换。(A)筛选嗜热聚合酶在

PCR过程中模板转换的倾向。显示了含有共享的条形码(BC3、BC4、BC5和BC13)和间隔区(Sp)序列的不同长度的两种合成模拟模板(TS Ctrl1和TS Ctrl4)。在通过PCR共扩增后,除了两个原始模板带之外,通过PAGE的带的存在指示模板转换。(B)通过PAGE分析PCR产物。模板转换是PCR条件和聚合酶类型的函数。Taq聚合酶(道2-5;道1=尺寸标记)展示了比Deep Vent exo-(道6-9)低得多的模板转换,特别是在60℃退火时。

[1657] 图62示出了单循环的ProteoCode测定。在测定中使用两种不同的珠粒类型(一种具有氨基A-末端肽(AFA-肽),另一种具有氨基F-末端肽(FA-肽))。在由一个循环的NTF/NTE化学和两个循环的编码组成的全循环ProteoCode分析中使用AFA-肽和FA-肽珠粒,编码由使用工程化苯丙氨酸(F)结合剂的化学前和化学后编码步骤组成。在化学编码前测定与化学编码后测定中使用不同的F-结合剂条形码。在编码的两个循环(预编码和后编码)之后,对延伸记录标签进行PCR和NGS库制备,随后在Illumina MiSeq仪器上进行NGS测序读出。示出了包含"前"和"后"条形码化编码标签的AFA和FA记录标签的归一化读数数量。对于与F-结合剂孵育的FA肽,与化学循环后标签相比观察到更高的"化学前"编码标签读取计数,表明成功移除F残基。相反,对于与F-结合剂孵育的AFA肽,观察到与化学循环后标签相比化学循环前标签的读取计数较低,表明成功移除A残基。

[1658] 图63A-B示出了根据肽嵌合体表面密度的分子间相互作用。对以不同表面嵌合体密度固定在相同珠粒上的4-plex肽-记录标签嵌合体(AA-肽,AF-肽,FA-肽,和没有肽)的F-结合剂测定的相互作用测试。(A)通过PA抗体结合/编码来评估跨四种嵌合体类型的嵌合体固定化的均匀性。每个嵌合体(通过测序或记录标签测定)约占总嵌合体的20-30%。(B)跨4种嵌合体类型的F-结合剂结合和编码的单循环的编码效率。在1:10,000和1:100,000的嵌合体表面稀释度下,FA-肽的分子内编码效率远大于脱靶肽嵌合体(AF、AA和无肽)的分子间编码效率。

[1659] 图64A-B示出了使用ssDNA连接酶用碱基保护的DNA编码的DNA。(A)用于ssDNA连接测定的碱基保护记录标签(RT)和编码标签(CT)的设计。所使用的碱基保护基团是用于寡核苷酸合成的标准亚磷酰胺保护基团:Bz-dA、ibu-dG、Ac-dC。这些保护基在合成后留在原处。用CircLigase II将RT寡核苷酸的5'磷酸化末端与CT寡核苷酸的3'OH连接。(B)ssDNA连接酶连接反应组分和产物的PAGE凝胶分析。道1=RT全长寡核苷酸;道2=在用USER Enzyme处理以从由5'U组成的RT全长产生磷酸化的RT序列之后;道3=将CT寡核苷酸退火至固定在珠粒上的RT寡核苷酸之后;道4=在CT的3'末端通过B-B'相互作用连接到RT的5'末端后,使用CircLigase II。"碱基保护的"DNA的连接是相对有效的,如通过使用CircLigase II形成更高分子量的连接产物所表明的。

[1660] 图65A-E示出了使用与DNA珠粒的杂交和连接来固定DNA标记的肽。(A)设计DNA发夹珠粒,其中DNA发夹上的5'突出用于捕获DNA-肽嵌合体。(B)DNA-肽嵌合体与珠粒上的发夹杂交。(C)DNA连接酶用于将肽上的磷酸化DNA标签的5'端共价连接至发夹。设计肽上的DNA标签使得肽与内部核苷酸化学缀合。肽上的DNA标签构成具有游离3'端的记录标签,来自编码标签的信息可以转移到该记录标签。(D)设计qPCR引物来评估DNA标记的肽在珠粒上的固定化效率。(E)使用三个不同引物组进行qPCR定量以评估发夹(突出)和DNA标记的肽固定化(嵌合体和连接产物)。较低的Cq值表示扩增产物的较高丰度。

## 具体实施方式

[1661] 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。

[1662] 在本申请中提及的所有出版物,包括专利文献、科学论文和数据库,出于所有目的通过引用以其整体并入,其程度如同每个单独的出版物通过引用单独并入一样。如果本文阐述的定义与通过引用并入本文的专利、申请、公开的申请和其他出版物中阐述的定义相反或不一致,则本文阐述的定义优于通过引用并入本文的定义。

[1663] 本文所用的部分标题仅用于组织目的,不应被解释为限制所述主题。

[1664] I.简介和概述

[1665] 由于各种原因,高度平行的大分子表征和蛋白质(例如蛋白质)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,其中一个目标是识别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。

[1666] 蛋白质或多肽分析物的分子识别和表征通常使用免疫测定进行。有许多不同的免疫测定形式,包括ELISA、多重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,在样品水平和分析物水平上多路复用的能力有限,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(埃德曼降解或质谱)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。

[1667] 基于埃德曼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(无水三氟乙酸,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。裂解的环状ATZ-氨基酸被转化为乙内酰苯硫脲(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识别。通常,本领域的埃德曼降解肽测序方法是缓慢的并且每天仅有少量肽的有限通量。

[1668] 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埃德曼降解。尽管MS仪器(Rile等人,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整个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,因此样品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。

[1669] 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。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。

[1670] 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,例如,代表合并样品,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转移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。在一个替代实施例中,来自包含相关大分子的识别信息的记录标签的信息可以转移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"双标签"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,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。

[1671] 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双标签构建体的编码标签,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双标签,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)的记录标签上的识别信息可用于将延伸编码标签或双标签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库。可以扩增该核酸编码库,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、扣除和标准化。这些基于DNA的方法容易且快速地可缩放且可定制,且比可用于直接操纵其它类型的大分子库(例如,蛋白质库)的那些方法更有成本效益。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库以富集和/或扣除和/或归一化序列的表达。这使得可以更有效、更快速、更经济高效地从非常大的库提取最需要的信息,这些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除常见的高丰度蛋白,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,还可以扣除延伸记录标签库的白蛋白特异性成员,从而实现更完整的整体扣除。

[1672] 在一个方面,本公开内容提供了使用埃德曼类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板(例如,多孔珠粒)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埃德曼类降解过程。修饰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)从编码标签转移到记录标签(例如,引物延伸或连接),来识别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修饰的NTAA。在某些实施例中,酶(例如,Edmanase)被工程化以催化去除修饰的NTAA。在其他实施例中,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裂解末端氨基酸。任选地,所述方法包括在每个NTAA移除步骤之前和/或之后使多肽与脯氨酸氨肽酶接触的步骤,因为所述步骤否则可能不裂解末端脯氨酸。

[1673] II.定义

[1674] 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例的描述。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“包含(comprising)”(和诸如“包含(comprise)”或“包含(comprises)”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“由.....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例的范围或含义。

[1675] 贯穿本说明书对“一个实施例”或“实施例”的引用意味着结合该实施例描述的特定特征,结构或特性包括在至少一个实施例中。因此,贯穿本说明书在各个地方出现的短语“在一个实施例中”或“在实施例中”不一定都指的是同一实施例。此外,特定特征、结构或特性可以在一个或多个实施例中以任何合适的方式组合。

[1676] 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,提及的“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。

[1677] 如本文所用,术语“大分子”涵盖由较小亚基组成的大的分子。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。

[1678] 如本文所用,术语“多肽”包括肽和蛋白质,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。在一些实施例中,多肽包含2至50个氨基酸,例如具有超过20-30个氨基酸。在一些实施例中,肽不包含二级、三级或更高级结构。在一些实施例中,蛋白质包含30个或更多个氨基酸,例如具有超过50个氨基酸。在一些实施例中,除了一级结构之外,

蛋白质还包含二级、三级或更高级结构。多肽的氨基酸最典型地是L-氨基酸,但也可以是D-氨基酸、非天然氨基酸、修饰的氨基酸、氨基酸类似物、氨基酸模拟物,或其任何组合。多肽可以是天然发生的、合成产生的、或重组表达的。多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。该聚合物可以是直链或支链的,它可以包含修饰的氨基酸,并且它可以被非氨基酸中断。该术语还涵盖已经天然或通过干预修饰的氨基酸聚合物;例如,二硫键形成,糖基化,脂质化,乙酰化,磷酸化或任何其他操作或修饰,例如与标记组分缀合。

[1679] 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,其用作肽的单体亚基。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。

[1680] 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。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$C_1$-$C_4$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签的肽修饰。

[1681] 如本文所用,术语"结合剂"是指与分析物,例如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、蛋白质、碳水化合物或小分子。结合剂可以与分析物,例如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,二肽、三肽、长肽的更高级肽、多肽或蛋白质分子)。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、多肽或蛋白质结合,或与构象肽、多肽或蛋白质结合。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,结合剂可以例如结合化学修饰的或标记的氨基酸。例如,结合剂可以例如与用乙酰基团、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。结合剂可以结合多肽分子的翻译后修饰。结合剂可以表现出与分析物,例如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分析物,例如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,其可以通过接头与结合剂连接。

[1682] 在一些实施例中,术语"剂"和"试剂"可互换使用。例如,虽然结合剂被称为试剂,但是其编码标签可以具有反应性部分,例如,间隔区序列或反应性端,以与记录标签上的间隔区序列或反应性端反应,从而转移信息。反应可包括连接和/或退火,然后引物延伸。

[1683] 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。在某些实施例中,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。

[1684] 如本文所用,术语"蛋白质组学"是指蛋白质组的分析(例如,定量分析),例如,细胞内的蛋白质组,组织内的蛋白质组和体液内的蛋白质组,以及细胞内和组织内的蛋白质组的相应空间分布。此外,蛋白质组学研究包括蛋白质组的动态状态,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。

[1685] 如本文所用,术语"蛋白质组"可包括由任何生物体的靶标(例如,基因组、细胞、组织或生物体)在某个时间表达的整组蛋白质,多肽或肽(包括其缀合物或复合物)。在一个方面,它是在给定的条件下、在给定的时间、在给定类型的细胞或生物体中表达的蛋白质。蛋白质组学是对蛋白质组的研究。例如,"细胞蛋白质组"可以包括在特定环境条件组(例如,暴露于激素刺激)下在特定细胞类型中发现的蛋白质的集合。生物体的完整蛋白质组可包括来自所有各种细胞蛋白质组的完整组的蛋白质。蛋白质组也可包括某些亚细胞生物系统中蛋白质的收集。例如,病毒中的所有蛋白质可称为病毒蛋白质组。如本文所用,术语"蛋白质组"包括蛋白质组的子集,包括但不限于激酶组;分泌体;受体组(例如,GPCRome);免疫蛋白酶体;营养蛋白质组;由翻译后修饰(例如,磷酸化,泛素化,甲基化,乙酰化,糖基化,氧化,脂质化和/或亚硝基化)定义的蛋白质组子集,例如磷酸蛋白质组(例如,磷酸酪氨酸-蛋白质组,酪氨酸-激酶组和酪氨酸-磷酸化),糖蛋白质组等;与组织或器官、发育阶段或生理或病理状况相关的蛋白质组子集;与细胞过程如细胞周期,分化(或去分化),细胞死亡,衰老,细胞迁移,转化或转移相关的蛋白质组子集;或其任何组合。

[1686] 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、或被测亚基,"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、或亚基。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例的条件下,记录标签信息不能有效地转移至编码标签。

[1687] 如本文所用,术语“脯氨酸氨肽酶”是指能够从多肽特异性裂解N末端脯氨酸的酶。具有这种活性的酶是本领域熟知的,并且也可称为脯氨酸亚氨基肽酶或PAP。已知的单体PAP包括来自凝结芽孢杆菌、L.delbrueckii、淋病奈瑟菌、脑膜脓毒黄杆菌、粘质沙雷氏菌、嗜酸热原体、植物乳杆菌的家族成员(MEROPSS 33.001)(Nakajima,Ito等人,2006)(Kitazono,Yoshimoto等人,1992)。已知的多体PAP包括D.hansenii(Bolumar,Sanz等人,2003)。可以使用天然的或工程化的PAP。

[1688] 如本文所用,术语“特异性结合”是指结合剂(例如抗体、ClpS蛋白质或anticalin)的特异性,使得其优先结合靶标(例如,多肽抗原)。当涉及结合配偶体(例如,蛋白质、核酸、抗体或其他亲和捕获剂等)时,“特异性结合”可以包括具有高亲和力和/或互补性的两个或更多个结合配偶体的结合反应,以确保在指定的测定条件下的选择性杂交。通常,特异性结合将是背景信号标准偏差的至少三倍。因此,在指定条件下,结合配偶体结合其特定靶标分子,并且不以显著量结合样品中存在的其它分子。在存在其它潜在干扰物质的情况下,由特定靶标的结合剂或抗体识别是这种结合的一个特征。例如,对靶标具有特异性或特异性结合靶标的结合剂、抗体或抗体片段以比结合其他非靶标物质更高的亲和力结合靶标。例如,对靶标具有特异性或特异性结合靶标的结合剂、抗体或抗体片段避免与显著百分比的非靶标物质(例如,测试样品中存在的非靶标物质)结合。在一些实施例中,本公开内容的结合剂、抗体或抗体片段避免结合大于约90%的非靶标物质,尽管更高的百分比被清楚地考虑并且是优选的。例如,本公开的结合剂、抗体或抗体片段避免结合约91%,约92%,约93%,约94%,约95%,约96%,约97%,约98%,约99%和约99%或更多的非靶标物质。在其它实施例中,本公开的结合剂、抗体或抗体片段避免大于约10%,20%,30%,40%,50%,60%或70%,或大于约75%,或大于约80%,或大于约85%的非靶标物质结合。

[1689] 如本文所用,多核苷酸,蛋白质或多肽变体、突变体、同源物或修饰型包括与参考多核苷酸、蛋白质或多肽具有核酸或氨基酸序列同一性的蛋白质或多肽,例如,约10%的序列同一性,约15%的序列同一性,约20%的序列同一性,约25%的序列同一性,约30%的序列同一性,约35%的序列同一性,约40%的序列同一性,约45%的序列同一性,约50%的序列同一性,约55%的序列同一性,约60%的序列同一性,约65%的序列同一性,约70%的序列同一性,约75%的序列同一性,约80%的序列同一性,约85%的序列同一性,约90%,91%,92%,93%,94%,95%,96%,97%,98%,或99%的序列同一性,或约100%序列同一性。

[1690] 关于参考多核苷酸序列的“百分比(%)核酸序列同一性”定义为候选序列中与参考多核苷酸序列中的核酸残基相同的核酸残基的百分比,在序列比对和引入缺口(如果需要)以实现最大百分比序列同一性之后,并且不考虑任何保守取代作为序列同一性的一部分。相对于参考多肽序列的“百分比(%)氨基酸序列同一性”定义为候选序列中与参考多肽序列中的氨基酸残基相同的氨基酸残基的百分比,在序列比对和引入缺口(如果需要)以实

现最大百分比序列同一性之后,并且不考虑任何保守取代作为序列同一性的一部分。用于确定氨基酸序列同一性百分比和/或核酸序列百分比的比对可以以本领域技术人员已知的各种方式实现,例如,使用可公开获得的计算机软件,如BLAST,BLAST-2,ALIGN或Megalign(DNASTAR)软件。本领域技术人员可以确定用于比对序列的适当参数,包括在所比较的全长序列上实现最大比对所需的任何算法。

[1691] 例如,本文公开的ClpS蛋白质或多肽包括与参考ClpS蛋白质或多肽,例如根癌土壤杆菌ClpS2(例如,SEQ ID NO:198),大肠杆菌ClpS(例如,SEQ ID NO:199),和/或新月柄杆菌ClpS(例如,SEQ ID NO:200)共有序列同一性的变体、突变体、同源物或修饰型,具有约10%的序列同一性,约15%的序列同一性,约20%的序列同一性,约25%的序列同一性,约30%的序列同一性,约35%的序列同一性,约40%的序列同一性,约45%的序列同一性,约50%的序列同一性,约55%的序列同一性,约60%的序列同一性,约65%的序列同一性,约70%的序列同一性,约75%的序列同一性,约80%的序列同一性,约85%的序列同一性,约90%,91%,92%,93%,94%,95%,96%,97%,98%或99%的序列同一性,或约100%的序列同一性。

[1692] 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编号,其中肽的长度为"n"个氨基酸。如本文所用,NTAA被认为是第n氨基酸(在本文中也称为"第n NTAA")。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。在某些实施例中,NTAA、CTAA或两者可以用化学基团修饰或标记。

[1693] 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、珠粒或隔离的位置)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,条形码群内的每个条形码是不同的。在其他实施例中,条形码群中的一部分条形码是不同的,例如,条形码群中至少约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、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,大分子集合已经分布到小隔室中以增强映射。例如,不是将肽映射回蛋白质组,而是将肽映射回其起源蛋白分子或蛋白复合物。

[1694] "样品条形码",也称为"样品标签",识别大分子来自哪个样品。

[1695] "空间条形码"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库的多重测序。

[1696] 如本文所用,术语"编码标签"是指任何合适长度的多核苷酸,例如约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"可测序聚合物"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)。编码标签可以包含编码器序列,其可选地在一侧侧接一个间隔区或在每侧侧接间隔区。编码标签还可以由可选的UMI和/或可选的结合循环特异性条形码组成。编码标签可以是单链或双链的。双链编码标签可包含平端、突出端或两者。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。在某些实施例中,编码标签可以进一步包含结合循环特异性间隔区或条形码、独特分子标识符、通用引发位点或其任何组合。

[1697] 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。编码器序列可以唯一地识别其相关的结合试剂。在某些实施例中,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,编码器序列可以将其相关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。在一些实施例中,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例中,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,而不用明确地识别该位置的氨基酸残基。在另一个实例中,共同的编码器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。在其它实施例中,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部分识别编码标签信息为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)的特定组合提供为结合试剂产生独特标识信息。例如,结合剂库内的编码序列具有相同或相似数量的碱基。

[1698] 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"是指用于识别在特定结合循环内使用的结合剂库的独特序列。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,作为间隔区序列的一部分、编码器序列的一部分、UMI的一部分、或作为编码标签内的单独组分。

[1699] 如本文所用,术语"间隔区"(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个碱基)的核酸分子,存在于记录标签或编码标签的末端。当间隔区是0碱基时,测定有时称为"无间隔区"形式,例如,如图7和图46所示。在某些实施例中,间隔区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区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记录标签、编码标签或双标签构建体。Sp'指与Sp互补的间隔区序列。例如,结合剂库内的间隔区序列具有相同数目的碱基。可以在结合剂库中使用共同的(共有的或相同的)间隔区。间隔区序列可具有"循环特异性"序列,以跟踪特定结合循环中使用的结合剂。间隔区序列(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区允许将来

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区识别相同类别的大分子。只有正确的同源对的顺序结合才能产生相互作用的间隔区元件和有效的引物延伸。间隔区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。间隔区序列可包含比编码标签内的编码器序列更少数量的碱基。

[1700] 如本文所用,术语"记录标签"是指某部分,例如,化学偶联部分,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 et al.,2015,Nat.Commun.6:7237;Lutz,2015,Macromolecules 48:4759-4767;每一个都是通过引用而整体并入),可以将编码标签的识别信息转移到编码标签,或者可以将关于与记录标签相关联的大分子的识别信息(例如,参见,UMI信息)从编码标签转移到编码标签。识别信息可以包含表征分子的任何信息,例如关于样品,级分,分区,空间位置,相互作用的相邻分子(例如,循环数)等的信息。另外,UMI信息的存在也可以被分类为识别信息。在某些实施例中,在结合剂结合多肽后,来自与结合剂连接的编码标签的信息能转移到与多肽相关的记录标签,同时结合剂结合到多肽上。在其他实施例中,在结合剂结合多肽后,来自与多肽相关的记录标签的信息可以传递到与结合剂连接的编码标签,同时结合剂结合到多肽上。重新编码标签可以直接与多肽连接、通过多功能接头与多肽连接、或者借助于其在固体支持物上的相近(或共定位)而与多肽结合。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分,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)、与编码标签的间隔区序列互补的间隔区序列、或其任何组合。在使用聚合酶延伸将编码标签信息转移到记录标签的实施例中,记录标签的间隔区序列优选位于记录标签的3'末端。

[1701] 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、间隔区序列)使用所述互补链作为模板通过聚合酶延伸。

[1702] 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可用于识别结合特定大分子的每种个体结合剂。例如,UMI可用于识别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)。

[1703] 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等人,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等人,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"。

[1704] 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)的信息在结合剂与分析物,例如大分子结合后已转移到记录标签。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。延伸记录标签的碱基序列可以反映由其编码标签识别的结合剂的结合时间和顺序。可以反映由编码标签识别的结合剂的部分结合顺序,或者可以不反映任何由编码标签识别的结合剂的结合顺序。在某些实施例中,延伸记录标签中存在的编码标签信息至少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例中,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。

[1705] 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)的信息转移到其的编码标签,记录标签与大分子相关。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"双标签"或"双标签构建体"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)的信息在与编码标签连接的结合剂结合后转移到核酸分子,记录标签与大分子相关(参见,例如,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)转移到双标签。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,双标签包含记录标签的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。

[1706] 如本文所用,术语"固体支持物"、"固体表面"、或"固体基板"或"基板"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,珠粒、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

体支持物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、珠粒、微球、微粒、或其任何组合。例如,当固体表面是珠粒时,珠粒可以包括但不限于陶瓷珠粒、聚苯乙烯珠粒、聚合物珠粒、甲基苯乙烯珠粒、琼脂糖珠粒、丙烯酰胺珠粒、实心核心珠粒、多孔珠粒、顺磁珠粒,玻璃珠粒或可控孔珠粒。珠粒可以是球形或不规则形状。珠粒的尺寸可以为纳米(例如,100nm)至毫米(例如,1mm)。在某些实施例中,珠粒的尺寸范围为约0.2微米至约200微米,或约0.5微米至约5微米。在一些实施例中,珠粒可以是约1μm,1.5μm,2μm,2.5μm,2.8μm,3μm,3.5μm,4μm,4.5μm,5μm,5.5μm,6μm,6.5μm,7μm,7.5μm,8,8.5μm,9μm,9.5μm,10μm,10.5μm,15,或20μm的直径。在某些实施例中,“珠粒”固体支持物可以指单个珠粒或多个珠粒。如本文所用,术语“基板”包括机械支持物,材料可设置在机械支持物上以提供功能性,无论是机械,生物,光学,化学还是其它功能性。基板可以是未图案化的或图案化的,分区的或未分区的。基板上的分子可以按特征布置或者可以均匀地布置在基板表面上。

[1707] 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,其包括次黄嘌呤、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。

[1708] 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。

[1709] 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板和与核酸分子的互补序列,核酸分子可以通过引物与固体基板杂交,然后通过使用聚合酶在固体基板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)。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Service(Science 311:1544-1546,2006)所综述的平行珠粒阵列、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。

[1710] 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。

[1711] 如本文所用,“分析”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大分子的组分的部分识别。例如,部分识别大分子蛋白质序列中的氨基酸可以将蛋白质中的氨基酸识别为属于可能的氨基酸子集。分析通常从分析第n NTAA开始,然后进行到肽的下一个氨基酸(即,n-1、n-2、n-3等等)。这通过裂解所述第n NTAA,从而将肽的第n-1个氨基酸转化为N-末端氨基酸(本文称为“第(n-1)个NTAA”)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肽内的表位的顺序或位置的信息。分析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。

[1712] 如本文所用,术语“隔室”是指从大分子样品中分离或隔离大分子子集的物理区域或体积。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子集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粒。

[1713] 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隔室内或两个或更多个隔室的组内的蛋白质和/或肽,可以识别衍生自相同蛋白质、蛋白质复合物或单个隔室或隔室组内的细胞的肽。隔室标签包含条形码和任选的通用引物,条形码任选地在一侧或两侧侧接间隔区序列。间隔区序列可以与记录标签的间隔区序列互补,使得能够将隔室标签信息转移到记录标签。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例。隔室标签可包含用于偶联肽的官能部分(例如,醛、NHS、mTet、炔等)。或者,隔室标签可包含肽,肽包含对蛋白连接酶的识别序列,以允许将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,珠粒)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。

[1714] 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。在某些实施例中,可以通过将大分子分配到隔室中来实现分区。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。

[1715] 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。在某些实施例中,用于大分子的分区标签是指由将大分子分区成用相同条形码标记的隔室产生的相同隔室标签。

[1716] 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。

[1717] 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,其包含对分级物中大分子的识别信息。

[1718] 如本文所用,术语“多路复用”或“多路复用测定”在本文中可包括测定或其它分析方法,其中可通过使用多于一个捕获探针缀合物同时测定多个靶标(例如,多个核酸序列)的存在和/或量,其各自具有至少一种不同的检测特性,例如,荧光特性(例如,激发波长,发射波长,发射强度,FWHM(半峰高最大值全宽度)或荧光寿命)或独特的核酸或蛋白质序列特性。

[1719] III.用于分析包括大分子的分析物的试剂盒

[1720] A.制品和试剂盒的概述

[1721] 本文描述的试剂盒和试剂盒组分提供了用于分析物(例如,大分子)分析的高度平行化方法。高度多路复用的分析物大分析结合测定被转换成通过下一代测序方法读取的核酸分子库。本文提供的试剂盒和试剂盒组分对于蛋白质或肽测序特别有用。

[1722] 在优选的实施例中,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,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。蛋白质样品经历蛋白水解消化以产生记录标签标记的肽群(例如,数百万至数十亿)。这些记录标签标记的肽随机地合并和固定在固体支持物上。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转移到所述记录标签(或将记录标签信息转移到编码标签或将记录标签信息和编码标签信息转移到单独的双标签构建体)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(或延伸编码标签或双标签构建体)的库,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,以识别肽序列源自哪个样品、细胞、蛋白质组的子集或蛋白质。

[1723] 在一个方面,这里提供了在方法中使用的分析大分子的试剂盒和试剂盒组分,该方法包含:(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)使大分子

与能够结合大分子的第一结合剂接触,其中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)将第一编码标签的信息转移到记录标签,以产生第一次序延伸记录标签;(d)使大分子与能够结合大分子的第二结合剂接触,其中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)将第二编码标签的信息转移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(f)分析第二次序延伸标签(参见例如图2A-D)。

[1724] 在某些实施例中,接触步骤(b)和(d)按顺序进行,例如,第一结合剂和第二结合剂在分开的结合循环反应中与大分子接触。在另一些实施例中,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第一结合剂、第二结合剂,以及可选地额外的结合剂。在优选实施例中,接触步骤(b)和(d)包含使大分子与多个结合剂接触。提供试剂盒组分以执行这些步骤。

[1725] 在某些实施例中,所述方法还包含在步骤(e)和(f)之间的以下步骤:(x)通过用能够结合大分子的第三(或更高次序)结合剂替换第二结合剂重复步骤(d)和(e)一次或多次,其中第三(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)将第三(或更高次序)编码标签的信息转移到第二(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)分析第三(或更高次序)延伸记录标签。提供试剂盒组分以执行这些步骤。

[1726] 第三(或更高次序)结合剂可以在与第一结合剂和第二结合剂分开的结合循环反应中与大分子接触。在一个实施例中,在第n结合循环中使第n结合剂与分析物(例如大分子)接触,并且将信息从第n编码标签(第n结合剂的)转移到在第(n-1)结合循环中形成的延伸记录标签,以便形成进一步的延伸记录标签(第n延伸记录标签),其中n是整数: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,或约50,约100,约150,约200或更大。类似地,第(n+1)结合剂在第(n+1)结合循环中与分析物接触,等等。

[1727] 或者,第三(或更高次序)结合剂可以在具有第一结合剂和第二结合剂的单个结合循环反应中与大分子接触。在这种情况下,可以使用结合循环特异性序列,如结合循环特异性编码标签。例如,编码标签可以包含结合周期特异性间隔区序列,使得仅在信息从第n编码标签转移到第(n-1)延伸记录标签以形成第n延伸记录标签之后,那么第(n+1)结合剂(其可能已经或可能尚未与分析物结合)将能够将第(n+1)结合标签的信息转移到第n延伸记录标签。

[1728] 在第二方面,这里提供了在方法中使用的试剂盒和试剂盒组分,该方法包含以下步骤:(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)使大分子与能够结合大分子的第一结合剂接触,其中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)将第一编码标签的信息转移到记录标签以产生第一次序延伸记录标签;(d)使大分子与能够结合大分子的第二结合剂接触,其中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;(e)将第二编码标签的信息转移到第二记录标签以产生第二次序延伸记录标签;以及(f)分析第一和第二延伸记录标签。

[1729] 在某些实施例中,接触步骤(b)和(d)按顺序进行,例如,第一次结合剂和第二结合剂在分开的结合循环反应中与所述大分子接触。在另一些实施例中,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第一结合剂、第二结合剂,以及可选地额外的结合剂。提供试剂盒组分以执行这些步骤。

[1730] 在一些实施例中,步骤(a)还包含提供与固体支持物连接的相关的第三(或更高次序)记录标签。在进一步实施例中,所述方法还包含在步骤(e)和(f)之间的以下步骤:(x)通过用能够结合大分子的第三(或更高次序)结合剂替换第二结合剂重复步骤(d)和(e)一次或多次,其中第三(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)将第三(或更高次序)编码标签的信息转移到第三(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)分析第一、第二和第三(或更高次序)延伸记录标签。提供试剂盒组分以执行这些步骤。

[1731] 第三(或更高次序)结合剂可以在与第一结合剂和第二结合剂分开的结合循环反应中与大分子接触。或者,第三(或更高次序)结合剂可以在具有第一结合剂、第二结合剂的单个结合循环反应中与大分子接触。

[1732] 在试剂盒的某些实施例中,第一编码标签、第二编码标签和任何更高次序编码标签各自具有结合循环特异性序列。

[1733] 在第三方面,这里提供了在方法中使用的试剂盒和试剂盒组分,该方法包含以下步骤:(a)提供肽以及与固体支持物连接的相关记录标签;(b)用化学基团修饰肽的N-末端氨基酸(NTAA)以产生修饰的NTAA;(c)使肽与能够结合修饰的NTAA的第一结合剂接触,其中第一结合剂包含具有第一结合剂的识别信息的第一编码标签;(d)将第一编码标签的信息转移到记录标签以产生延伸记录标签;(e)分析延伸记录标签(参考例如图3)。

[1734] 在某些实施例在,其中步骤(c)还包含使肽与包含第二(或更高次序)编码标签的第二(或更高次序)结合剂接触,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)结合剂能够结合步骤(b)的修饰的NTAA之外的修饰的NTAA。在进一步的实施例中,在肽与第一结合剂接触后,按次序使肽与所述第二(或更高次序)结合剂接触,例如,第一结合剂和第二(或更高次序)结合剂接触与肽在分开的结合循环反应中接触。在其它实施例中,使肽与第二(或更高次序)结合剂接触和肽与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(或更高次序)结合剂。提供试剂盒组分以执行这些步骤。

[1735] 在某些实施例中,化学基团通过化学反应或酶促反应添加到NTAA。

[1736] 在一些实施例中,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。

[1737] 通过化学剂可向NTAA添加化学基团。在某些实施例中,用于修饰具有PTC基团的NTAA的化学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的化学剂是硫代酰化试剂;用于修饰具有硫代乙酰基的NTAA的化学剂是硫代乙酰化试剂;用于修饰具有乙酰基NTAA的化学剂是乙酰化试剂(如乙酸酐);用于修饰具有胍基(脒基)NTAA的化学剂是胍基化试剂,

或用于修饰具有硫代苄基的NTAA的化学剂是硫代苄基化试剂。任何这些化学基团和化学剂可以作为本文公开的试剂盒的组分提供。

[1738] 本公开提供的第四方面提供了在方法中使用的试剂盒和试剂盒组分,包含以下步骤:(a)提供与固体支持物连接的肽和相关的记录标签;(b)用化学基团修饰肽的N-末端氨基酸(NTAA)以产生修饰的NTAA;(c)使肽与能够结合修饰的NTAA的第一结合剂接触,其中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)将第一编码标签的信息转移到记录标签,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)用化学基团修饰肽的新NTAA以产生新修饰的NTAA;(g)使肽与能够结合新修饰的NTAA的第二结合剂接触,其中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)将第二编码标签的信息转移到第一延伸记录标签,以产生第二延伸记录标签;(i)分析第二延伸记录标签。

[1739] 在一些实施例中,接触步骤(b)和(d)按顺序进行,例如,第一次结合剂和第二结合剂在分开的结合循环反应中与肽接触。提供试剂盒组分以执行这些步骤。

[1740] 在某些实施例中,所述方法还包含在步骤(h)和(i)之间的以下步骤:(x)通过用能够结合修饰的NTAA的第三(或更高次序)结合剂替换第二结合剂重复步骤(e)、(f)和(g)一次或多次,其中第三(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)将第三(或更高次序)编码标签的信息转移到所述第二(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)分析第三(或更高次序)延伸记录标签。提供试剂盒组分以执行这些步骤。

[1741] 在某些实施例中,化学基团通过化学反应或酶促反应添加到NTAA。

[1742] 在一些实施例中,用于修饰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-甲氧基香豆素基;硫代酰基;硫代乙酰基;乙酰基;胍基;或硫代苄基。

[1743] 通过化学剂可向NTAA添加化学基团。在某些实施例中,用于修饰具有PTC基团的NTAA的化学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的化学剂是硫代酰化试剂;用于修饰具有硫代乙酰基的NTAA的化学剂是硫代乙酰化试剂;用于修饰具有乙酰基NTAA的化学剂是乙酰化试剂(如乙酸酐);用于修饰具有胍基(脒基)NTAA的化学剂是胍基化试剂,或用于修饰具有硫代苄基的NTAA的化学剂是硫代苄基化试剂。任何这些化学基团和化学剂可以作为本文公开的试剂盒的组分提供。

[1744] 在第三方面,这里提供了在方法中使用的分析多肽的试剂盒和试剂盒组分,该方法包含以下步骤:(a)提供肽和连接在固体支持物上的相关的记录标签;(b)使肽与能够结合肽的N-末端氨基酸(NTAA)的第一结合剂接触,其中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)将第一编码标签的信息转移到记录标签以产生延伸记录标签;以及(d)分析延伸记录标签。

[1745] 在一些实施例中,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,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)结合剂能够结合肽的NTAA之外的NTAA。在进一步的实施例中,在肽与第一结合剂接触后,按次序使肽与第二(或更高次序)结合剂接触,例如,第一结合剂和第二(或更高次序)结合剂接触与肽在分开的结合循环反应中接触。在其它实施例中,使肽与第二(或更高次序)结合剂接触和肽与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(或更高次序)结合剂。

[1746] 在第六方面,这里提供了在方法中使用的试剂盒和试剂盒组分,所述方法包含以下步骤:(a)提供肽和连接在固体支持物上的相关的记录标签;(b)使肽与能够结合肽的N末端-氨基酸(NTAA)的第一结合剂接触,其中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)将第一编码标签的信息转移到所述记录标签,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)使肽与能够结合新修饰的NTAA的第二结合剂接触,其中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)将第二编码标签的信息转移到第一延伸记录标签,以产生第二延伸记录标签;以及(g)分析第二延伸记录标签。

[1747] 在某些实施例中,所述方法还包含在步骤(f)和(g)之间的以下步骤:(x)通过用能够结合修饰的大分子的第三(或更高次序)结合剂替换第二结合剂重复步骤(d)、(e)和(f)一次或多次,其中第三(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;以及(y)将第三(或更高次序)编码标签的信息转移到第二(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)中分析第三(或更高次序)延伸记录标签。提供试剂盒组分以执行这些步骤。

[1748] 在一些实施例中,所述接触步骤(b)和(e)按顺序进行,例如,第一次结合剂和第二结合剂在分开的结合循环反应中与肽接触。

[1749] 在提供的任何实施例中,所述方法包含平行分析多个大分子。在优选实施例中,所述方法包含平行分析多个肽。

[1750] 在本文提供的任何实施例中,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;

[1751] 在提供的任何实施例中,大分子可以是蛋白质、多肽或肽。再进一步实施例中,其中肽是通过片段化分离自生物样品的蛋白获得。

[1752] 在提供的任何实施例中,大分子可以是蛋白质、多肽或肽。

[1753] 提供的任一实施例中,其中记录标签是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等人,2006,J.Am.Chem.Soc.128:10258-10267)、GNA分子或其任意组合。

[1754] 在此提供的任一实施例中,其中记录标签包含通用引发位点。在进一步实施例中,其中通用引发位点包含用于扩增、测序或二者结合的引发位点。

[1755] 在此提供的任一实施例中,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、间隔区序列或它们的任意组合。

[1756] 在此提供的任一实施例中,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、间隔区序列、或它们的任意组合。

[1757] 在此提供的任一实施例中,编码标签中的结合循环特异性序列可以是结合循环特

异性间隔区序列。

[1758] 在某些实施例中,结合循环特异性序列被编码为与编码器序列分开的条形码。在其它实施例中,编码器序列和结合循环特异性序列在单个条形码中列出,条形码对于结合剂和每个结合循环是唯一的。

[1759] 在一些实施例中,间隔区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。在其它实施例中,间隔区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。

[1760] 在此提供的任一实施例中,其中记录标签可以包含条形码。

[1761] 在此提供的任一实施例中,大分子和相关的记录标签可以共价结合到固体支持物上。

[1762] 在此提供的任一实施例中,其中固体支持物是珠粒、多孔珠粒、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。

[1763] 在此提供的任一实施例中,其中所述固体支持物是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[1764] 在此提供的任一实施例中,多个大分子和相关的记录标签可以结合到固体支持物上。在进一步实施例中多个分析物(比如,大分子)在固体支持物上以>50nm,>100nm,或>200nm的平均距离间隔开。

[1765] 在此提供的任一实施例中,结合剂可以是多肽或蛋白质。在进一步实施例中,其中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。

[1766] 在此提供的任一实施例中,结合剂可以选择性地结合大分子。

[1767] 在此提供的任一实施例中,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。

[1768] 在此提供的任一实施例中,结合剂和编码标签可以通过接头连接。

[1769] 在本文提供的任何实施例中,结合剂和编码标签可以通过SpyTag/SpyCatcher,SpyTag-KTag/SpyLigase(其中待连接的两个部分具有SpyTag/KTag对,并且SpyLigase将SpyTag连接至KTag,从而连接两个部分),分选酶,和/或SnoopTag/SnoopCatcher肽-蛋白质对(Zakeri等人,2012,Proc Natl Acad Sci USA 109(12):E690-697;Veggiani等人,2016,Proc.Natl.Acad.Sci.USA 113:1202-1207,通过引用将其全文并入本文。

[1770] SpyLigase是将两个肽标签彼此连接以形成不可逆的肽-肽相互作用的蛋白质结构域。Fierer等人的文献(2014年10月)中,“SpyLigase肽-肽连接聚合亲合体,以增强磁性癌细胞捕获”,PNAS111(13):E1176-E1181,其通过引用整体并入本文。

[1771] 分选酶也可用于连接结合剂和编码标签。例如,分选酶可以选择性地将LPXTG标签(其中X代表任何氨基酸)连接到聚G部分,例如GGG。可以使用例如来自圣地亚哥ActiveMotif或如US 9,267,127 B2中所公开的分选酶A5。在另一个实例中,一种分选酶包含与金黄色葡萄球菌(S.aureus)的氨基酸序列至少约50%,至少约60%,至少约70%,至少约75%,至少约80%,至少约85%,至少约90%,至少约95%,至少约99%,或100%相同的氨

基酸序列。可以使用分选酶A。

[1772] 在本文提供的任何实施例中,编码标签的到记录标签的信息转移是由连接酶介导的,例如DNA连接酶,例如,单链DNA(ssDNA)连接酶。在一个方面,阻断待连接的两个单链多核苷酸之一的5'端以防止在该5'端连接。在任何前述实施例中,ssDNA连接酶可以是栖热菌属噬菌体RNA连接酶,例如噬菌体TS2126 RNA连接酶(例如CircLigase™和CircLigase II™)或其变体、同源物、突变体或修饰型,或古细菌RNA连接酶,例如嗜热自养甲烷杆菌RNA连接酶1或其变体、同源物、突变体或修饰型。在其它方面,ssDNA连接酶是RNA连接酶,例如T4 RNA连接酶,例如T4 RNA连接酶2,T4 RNA连接酶2(截短型),T4 RNA连接酶2(截短型KQ),或T4 RNA连接酶2(截短型K227Q)。在一个方面,可以通过使用拥挤剂如PEG 4000来优化连接反应。当要连接的DNA末端在互补DNA序列上彼此相邻退火时,也可以使用T4 DNA连接酶和Ampligase®DNA连接酶。

[1773] 或者,其中记录标签到编码标签的信息转移通过DNA合成酶或化学连接介导。在一些实施例中,化学连接包含迈克尔加成,例如硫醇-马来酰亚胺偶联。在一些实施例中,化学连接包含铜催化的叠氮化物-炔环加成。在一些实施例中,化学连接包含酰胺键形成。在一些实施例中,酰胺键形成在NHS酯和胺或羧酸和胺之间在肽偶联剂的存在下进行。在一些实施例中,化学连接包含经由胺和醛形成亚胺,然后使用氰基硼氢化钠还原。在一些实施例中,化学连接包含硫醇烯偶联。在一些实施例中,化学连接包含天然的化学连接(例如,硫代(1)酯加半胱氨酸)。在一些实施例中,用于化学连接的试剂包含应变的炔烃(例如,环辛炔)和叠氮化物。在一些实施例中,化学连接包含无痕和非无痕施陶丁格连接。

[1774] 在此提供的任一实施例中,分析延伸记录标签可以包含核酸测序。在进一步实施例中,核酸测序方法是通过合成测序、通过连接测序通过、杂交测序、polony测序、离子半导体测序或焦磷酸测序。在其它实施例中,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。

[1775] 在此提供的任一实施例中,延伸记录标签在分析前被扩增。

[1776] 在此提供的任一实施例中,其中包含于延伸记录标签中的编码标签信息的顺序可以通过结合剂到大分子提供关于结合顺序的信息以及被结合剂检测到的分析物的序列信息。

[1777] 在此提供的任一实施例中,其中包含于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大分子中的分析物的频率的信息。

[1778] 在此公开的任何实施例中,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。

[1779] 在此公开的任何实施例中,其中代表多个大分子的多个延伸记录标签可以平行分析。

[1780] 在此公开的任何实施例中,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测定中分析。

[1781] 在此提供的任一实施例中,多个延伸记录标签在分析之前进行靶标丰度检测。

[1782] 在此提供的任一实施例中,多个延伸记录标签在分析之前进行扣除测定。

[1783] 在此提供的任一实施例中,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。

[1784] 在此提供的任一实施例中,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。

[1785] 在此提供的任一实施例中,其中至少一种结合剂结合到末端氨基酸残基。在一些实施例中,末端氨基酸残基是N-末端氨基酸或C-末端氨基酸。

[1786] 在此提供的任一实施例中,其中至少一种结合剂结合到翻译后修饰的氨基酸。

[1787] 在一些实施例中,本文提供了小分子-蛋白质结合基质。在一些实施例中,这可以是相当大的基质,即使当$10^9$个多样性的小分子库与整个固定的人蛋白质组($2\times10^4$)一起孵育时也是如此。基质尺寸为$10^9\times2\times10^4$,具有$2\times10^{13}$个不同的测量值。在一些实施例中,即使在洗涤步骤之后,也可以用上述方法容易地记录高亲和力结合相互作用(<低nM),但是中等至低亲和力相互作用将可能丧失。在一些实施例中,为了记录低亲和力相互作用(-μM),需要发生结合和同时信息转移的均相反应。在一些实施例中,这要求聚合或连接步骤需要具有相对于结合占用时间的快速动力学。聚合时间标度是在秒的数量级上,并且与在μM浓度下的结合占用时间(约10-100秒/秒)相容。连接时间尺度也可以是秒的数量级(例如,对于快速化学连接,参见,例如,Abe等人,"用于扩增RNA和DNA信号的快速DNA化学连接",Bioconjug Chem 19(1):327-333)。在一些实施例中,反应混合中可以包括连接酶或聚合酶以及一种或多种结合剂,特别是用于分析或检测低亲和力相互作用。或者,可以使用通过与记录标签内的位点相互作用的DNA标签的杂交将连接酶或聚合酶共定位到试剂位点。通过进行结合和"写"(即,信息转移)步骤的组合,更容易"记录"低亲和力相互作用。

[1788] 在一个实施例中,DNA标签附着于蛋白质靶标或肽或小分子配体(或肽类配体,例如,类肽配体,β-肽配体或D-肽肽类配体,或多糖配体)可以通过将条形码化的DNA标签单独附着至单独产生的蛋白质靶标或单独产生的小分子/肽靶标来实现。在另一个实施例中,DNA标签与蛋白质的附着通过核糖体或mRNA/cDNA展示发生,其中DNA条形码包含在mRNA序列内。在mRNA/cDNA展示的过程中,使用体外翻译反应将mRNA翻译成蛋白质。mRNA被配置为具有经由接头连接到mRNA的3'端上的嘌呤霉素分子(参见,例如,Leu等人,(2000),"用于体外蛋白质选择的RNA-蛋白质融合体的优化合成",酶学方法,学术出版社,318:268-293)。嘌呤霉素,一种酪氨酰-tRNA的类似物,系在mRNA转录物的3'端并在核糖体接近3'端时掺入到生长的多肽链中,终止翻译,并有效地产生将RNA/DNA转录物和相关的条形码连接至其相应的翻译蛋白的mRNA-蛋白融合体(例如US2012/0258871 A1、US2017/0107566 A1中所公开的,或Kozlov等人,(2012),PLoS One 7(6):e37441,其全部通过引用并入本文)。

[1789] 在一个实施例中,将用于下游信息转移的条形码序列和相关DNA记录标签功能元件如通用引物位点和间隔区序列工程化到mRNA转录物的3'端中。在一个实施例中,还只包括DNA条形码和DNA记录标签功能元件的5'的独特限制性位点。在体外翻译和产生mRNA-蛋白融合后,mRNA被反向翻译成cDNA。通过退火外源寡核苷酸和用限制酶孵育,cDNA被裂解为条形码的5'。这留下了DNA标签用作后续测定的记录标签(参见,例如,图50)。

[1790] 在另一方面,本文公开了具有单个循环的下一代蛋白质测定。在一些实施例中,该测定使用氨基酸结合剂,例如NTAA结合剂(例如N末端的单氨基酸、二氨基酸或三氨基酸)。

[1791] 通过自底向上质谱法(MS)进行的典型蛋白质识别使用蛋白酶消化步骤,例如胰蛋

白酶消化,以从每种蛋白质产生一组肽(约20-40个肽)。可在LC-MS/MS上分析单一肽种类并提供整个蛋白质种类的识别。通过引用并入本文的WO/2015/042506也公开了在单分子水平上的蛋白质识别方法。WO/2015/042506的试剂和方法步骤可与本公开组合。

[1792] 在单分子水平上,用于蛋白质识别的替代方法是将蛋白质分子分区成数百万至数十亿个分区(例如,分区包含隔室条形码),并且用分区条形码对样品中的所有分子进行条形码化。在一个方面,对于给定的隔室(例如,单独的珠粒表面),该隔室内的所有分子用相同的条形码标记。在另一方面,分区包含共享相同条形码的所有隔室。例如,如果在$10^8$个珠粒上有$10^6$个条形码,则一些珠粒将具有相同的条形码。在该示例中,每个单独的珠粒表面可以是隔室,并且共享相同条形码的隔室形成基于条形码的差异可与其他隔室区分开的隔室。在优选的实施例中,首先将蛋白质变性和烷基化,然后用遍及多肽的“点击”化学处理进行标记。在连接“点击”处理之后,使用点击化学将通用DNA引物附着到“点击”处理。DNA引物将形成DNA记录标签的起始部分。从过量的DNA引物中纯化多肽。在用DNA引物标记多肽后,将多肽暴露于具有隔室条形码和与多肽上的通用引物互补的通用引物的DNA条形码化的珠粒群(参见,例如,图51)。单独的多肽分子经由通用引物序列的杂交与特定的珠粒相互作用,并且基本上在珠粒上“增添”。一个以上的分子可以与任何给定的珠粒结合,但理想地,分子的数量保持最小。一旦多肽被“增添”到珠粒上,使用引物延伸或连接反应将珠粒的条形码信息转移到附着到多肽上的许多通用引物上。以这种方式,多肽现在是“隔室”条形码化的,形成蛋白质组分区的基础。多肽分子上延伸的DNA标签构成测定中后续步骤的记录标签。接下来,使用蛋白酶消化步骤(例如胰蛋白酶消化)将(可以首先去除的)珠粒上的多肽片段化。消化后,从珠粒上洗脱记录标签标记的肽。将标记的肽以单分子密度固定到基板上(例如,肽间距离使得固定的肽之间的平均距离在很大程度上将信息转移限制为仅分子内结合事件。间距的示例性距离可等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[1793] 在一个实施例中,肽来源于胰蛋白酶消化。对于人蛋白质组,大约有500000种不同的胰蛋白酶肽,每种蛋白质约20-30个胰蛋白酶肽。为了定量和表征样品中的蛋白质分子,将蛋白质转化为用分区条形码化记录标签标记的一组肽。来自与来源于给定分子的胰蛋白酶肽结合的所有信息可以通过条形码分区来映射回分子(或分子的小子集)。在一个优选的实施例中,将N-末端二肽和/或三肽(也称为N-末端二氨基/三氨基酸)结合剂与固定化肽一起孵育(使得N-末端是游离的),结合剂标记有包含识别结合剂的条形码信息的DNA编码标签。在结合时或结合后,通过聚合酶延伸或连接将信息从编码标签转移到记录标签(或反之亦然)。该过程可以重复多个循环以减少来自交叉反应性结合事件的错误。一组完整的N-末端二肽结合物将构成一组400个不同的结合剂,每个结合剂用不同的识别条形码标记。单个结合循环应该足以使用分区概念识别整个人蛋白质组。例如,即使仅检测到来自给定蛋白质分子的一组20个肽中的5个胰蛋白酶肽,N末端双氨基酸信息结合胰蛋白酶消化信息也足以唯一地识别蛋白质。

[1794] 作为首先将多肽固定到分区条形码化的珠粒并再固定到分析基板的替代方案,可以通过共价或非共价相互作用将变性的肽直接固定到DNA条形码化的珠粒上,其中整个珠粒用相同的条形码化的记录标签群覆盖(参见,例如,图52)。多肽在珠粒上被消化成片段,如胰蛋白酶消化物。在蛋白酶消化后,N-末端二肽结合剂与条形码化珠粒上的多肽的暴露N-末端结合。信息转移发生在二肽结合剂上的编码标签和附着于珠粒的任何记录标签之间。与先前的"距离间隔的"肽模型不同(参见,例如,图51所示),肽复合物不必如先前实施例中那样间隔开,只要仅进行单个结合循环即可。在结合和信息转移后,可以使用线性扩增或PCR将得到的延伸记录标签直接从珠粒表面扩增至溶液中。

[1795] 在一些实施例中,上述公开的测定不需要在编码标签和记录标签之间循环转移信息。因此,这些测定可以在单个结合循环中进行。

[1796] 前述实施例的特征在以下部分提供进一步细节。

[1797] B.大分子

[1798] 在一个方面,本公开涉及大分子的分析。大分子是由较小的亚基组成的大分子。在某些实施例中,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。

[1799] 根据本文公开的试剂盒和方法分析的大分子(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、包括线粒体隔室的细胞隔室和细胞周质。

[1800] 在某些实施例中,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库中,该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。在一些实施例中,肽、多肽或蛋白质是天然存在的、合成产生的或重组表达的。在任何上述肽实施例中,肽、多肽、蛋白质或蛋白质复合物可进一步包含翻译后修饰。

[1801] 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。

[1802] 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等人,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签。

[1803] 在某些实施例中,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。在一些实施例中,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)结合并裂解。在其他实施例中,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裂解(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裂解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等人,2007,Anal Bioanal Chem 389:991-1002)。在某些实施例中,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。

[1804] 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。

[1805] 在某些实施例中,在酶促或化学裂解后,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个氨基酸至约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个氨基酸至约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个氨基酸至约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裂解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,可以例如实时监测裂解反应。在完整的FRET肽中,荧光基团和猝灭剂基团连接到含有裂解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裂解测试肽后,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。

[1806] 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(请参考例如Whiteaker等人,2007,Anal.Biochem.362:44-54,通过引用以其整体并入)或选择特定的翻译后修饰(参见,例如,Huang等人,2014.J.Chromatogr.A 1372:1-17,通过引用以其整体并入本文)。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扣除丰度过高蛋白。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)。

[1807] 在某些实施例中,大分子由蛋白质或多肽组成。在一个实施例中,通过标准胺偶联化学,用DNA记录标签标记蛋白质或多肽(参见,例如,图1B,图2C,图28,图29,图31,图40)。根据反应的pH,ε-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(Mendoza和Vachet 2009)。在一个特定的实施例(参见,例如,图2B和图29),记录标签由反应部分(例如,对固体表面接合,多功能接头,或高分子),接头,通用引发序列,条形码(如隔室标签,分区条形码,样品条形码,连接器条形码,或任何组合),可选的UMI和间隔区(Sp)序列,促进到/从编码标记的信息转移。在另一个实施例中,可以首先用通用DNA标签标记所述蛋白,并且条形码-Sp序列(代表样品、隔室、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蛋白质。(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通用DNA标签互补的序列。在某些实施例中,通用DNA标签是通用引物序列。在标记蛋白上的通用DNA标签与记录标签中的互补序列(例如,与珠粒结合)杂交后,退火的通用DNA标签可以通过引物延伸扩增,将记录标签信息转移到DNA标记的蛋白。在具体实施例中,在由蛋白酶消化成肽之前,蛋白质用通用DNA标签标记。然后可以将来自消化物的标记肽上的通用DNA标签转换成信息丰富和有效的记录标签。

[1808] 在某些实施例中,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,其中记录标签直接与亲和捕获试剂结合,或者,蛋白质可以直接固定到具有记录标签的固体支持物上(参见,例如,图2C)。

[1809] C.支持物,例如固体支持物

[1810] 本公开的大分子连接到固体支持物的表面(也称为“基板表面”)。固体支持物可以是任何多孔或无孔载体表面,包括但不限于珠粒、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。用于固体支持物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。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、珠粒、微球、微粒、或其任何组合。例如,当固体表面是珠粒时,珠粒可以包括但不限于陶瓷珠粒、聚苯乙烯珠粒、聚合物珠粒,甲基苯乙烯珠粒、琼脂糖珠粒、丙烯酰胺珠粒、实心珠粒、多孔珠粒、顺磁珠粒,玻璃珠或可控孔珠粒。

[1811] 在某些实施例中,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,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包含与端连接的衔接子,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。454流通槽(54Life Sciences)支持“picotiter”板,光纤载玻片具有约160万75-Picotiter孔。剪切的模板DNA的每单个分子被捕获到分离的珠粒上,并且每个珠粒在油乳液内的水性PCR反应混合物的专用液滴中隔室化。通过PCR在珠粒表面上克隆扩增模板,然后将载有模板的珠粒分配到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粒。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粒的步骤之后,将结合在珠粒的存放在流通槽。流通槽也可以是简单的过滤器玻璃料,例如TWISTTM DNA合成柱(Glen Research)。

[1812] 在某些实施例中,固体支持物是珠粒,其可以指单个珠粒或多个珠粒。在一些实施例中,珠粒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。在一些实施例中,固体支持物是琼脂糖珠粒、顺磁珠粒、聚苯乙烯珠粒、聚合物珠粒、丙烯酰胺珠粒、实心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。在进一步的实施例中,珠粒可以包被结合官能(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。

[1813] 蛋白质、多肽或肽可以通过本领域已知的任何方法,包括共价和非共价相互作用,或它们的任何组合,直接或间接地连接到固体支持物上(参见,例如,Chan等人,2007,PLoS One 2:e1164;Cazalis等人,Bioconj.Chem.15:1005-1009;Soellner等人,2003,J.Am.Chem.Soc.125:11790-11791;Sun等人,2006,Bioconjug.Chem.17-52-57;Deceau等人,2007,J.Org.Chem.72:2794-2802;Camarero等人,2004,J.Am.Chem.Soc.126:14730-14731;Girish等人,2005,Bioorg.Med.Chem.Lett.15:2447-2451;Kalia等人,2007,Bioconjug.Chem.18:1064-1069;Watzke等人,2006,Angew Chem.Int.Ed.Engl.45:1408-1412;Parthasarathy等人,2007,Bioconjugate Chem.18:469-476;以及Bioconjugate Techniques,G.T.Hermanson,Academic Press(2013),各自通过引用整体并入本文)。例如,肽可以通过连接反应与固体支持物连接。或者,固体支持物可包括试剂或涂层以促进肽直

接或间接连接到固体支持物上。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,在一个实施例中,试剂是亲和分子。在另一个实例中,试剂是叠氮基团,可以与另一分子中的炔基反应以促进固体支持物和其它分子之间的联合或结合。

[1814] 可以使用称为“点击化学”的方法将蛋白质、多肽或肽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质、多肽或肽连接到固体支持物上。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等人,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。

[1815] 在一些实施例中,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。在各种实施例中,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。

[1816] 在其他实施例中,官能团包含烯烃、酯、酰胺、硫酯、二硫化物、碳环、杂环或杂芳基。在进一步的实施例中,官能团包含烯烃、酯、酰胺、硫酯、二硫化物、碳环、杂环或杂芳基。在其它实施例中,官能团包含酰胺或硫脲。在一些更具体的实施例中,官能团是三唑基官能团、酰胺、或硫脲官能团。

[1817] 在一个优选实施例中,iEDDA点击化学用于将大分子(例如,蛋白质、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例中,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。

[1818] 在一个优选的实施例中,用TCO官能化基板表面,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涂覆的基板表面上(图34)。

[1819] 蛋白质、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、叠氮内酯活化的、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裂解它们。

[1820] 在某些实施例中,蛋白质、多肽或肽通过固体表面连接的接头与蛋白质、多肽或肽的赖氨酸基团的共价连接固定到固体支持物上。

[1821] 记录标签可以在固定到固体支持物之前或之后附着到蛋白质、多肽或肽上。例如,

蛋白质、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官能部分的记录标签固定在固体表面上(参见图28)。记录标签的一个官能部分与蛋白质偶联,而另一个官能部分将记录标签标记的蛋白固定在固体支持物上。

[1822] 或者,在用记录标签标记蛋白质、多肽或肽之前,将蛋白质、多肽或肽固定到固体支持物上。例如,蛋白质可以首先用反应基团例如点击化学基团派生化。然后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基团衍生的蛋白质可以固定在用叠氮化物和TCO衍生的珠粒上,并附着在用叠氮化物和TCO标记的记录标签上。

[1823] 应当理解,本文提供的用于将大分子(例如,蛋白质、多肽、或肽)连接至固体支持物的方法也可用于将记录标签附着至固体支持物或将记录标签附着至大分子(例如,蛋白质、多肽、或肽)。

[1824] 在某些实施例中,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等人,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等人,2010,Methods Enzymol.472:1-18),疏水性二氯二甲基硅烷(DDS)+自组装吐温20(Hua等人,2014,Nat.Methods 11:1233-1236)和类金刚石碳(DLC),DLC+PEG(Stavis等人,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,包括表面活性剂如吐温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,当将蛋白质、多肽、或肽固定到固体基板上时,可以通过掺入竞争分子或“虚拟”反应性分子,在固体基板的表面上或体积内滴定蛋白质、多肽或肽的密度(见图36A)。

[1825] 在多个大分子固定在相同固体支持物上的某些实施例中,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,且其编码标签信息转移到与相邻大分子相关的记录标签,而不是与第一大分子相关的记录标签。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,可以在基板表面上滴定功能性偶联基团(例如,TCO)的密度(参见,例如,图34)。在一些实施例中,多个大分子在固体支持物的表面上或体积(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。在一些实施例中,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。在一些实施例中,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。在一些实施例中,大分子在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<1:10000。合适的间隔频率可以使用功能测定凭经验确定(参见示例23),并且可以通过稀释和/或通过掺杂“虚拟”间隔分子来完成,间隔分子竞争所述基板表面上的附着位点。

[1826] 例如,如图34所示,PEG-5000(MW约5000)用于阻断基板表面(例如珠粒表面)上的肽之间的间隙空间。此外,肽与官能部分偶联,官能部分也与PEG-5000分子连接。在一个优选的实施例中,这通过将NHS-PEG-5000-TCO+NHS-PEG-5000-甲基的混合物偶联到胺衍生的

珠粒上来实现(参见,例如,图34)。滴定两种PEG(TCO和甲基)之间的化学计量比,以在基板表面上产生适当密度的官能偶联基团(TCO基团);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。在某些实施例中,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。在珠粒的PEG5000-TCO/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)抑制剂猝灭。

[1827] 在特定实施例中,分析物分子和/或记录标签以一定的密度固定在基板或支持物上,使得(i)与第一分析物结合的编码剂(特别是该结合的编码剂中的编码标签)和(ii)第二分析物和/或其记录标签之间的相互作用减少、最小化或完全消除。因此,可以减少、最小化或消除由“分子间”接合产生的假阳性测定信号。

[1828] 在某些实施例中,针对每种类型的分析物确定分析物分子和/或基板上的记录标签的密度。例如,变性的多肽链越长,为了防止“分子间”相互作用,密度应越低。在某些方面,增加分析物分子和/或记录标签之间的间距(即,降低密度)增加了本公开的测定的信号背景比。

[1829] 在一些实施例中,分析物分子和/或记录标签以如下的平均密度沉积或固定在基板上的:约0.0001个分子/$\mu m^2$,0.001个分子/$\mu m^2$,0.01个分子/$\mu m^2$,0.1个分子/$\mu m^2$,1个分子/$\mu m^2$,约2个分子/$\mu m^2$,约3个分子/$\mu m^2$,约4个分子/$\mu m^2$,约5个分子/$\mu m^2$,约6个分子/$\mu m^2$,约7个分子/$\mu m^2$,约8个分子/$\mu m^2$,约9个分子/$\mu m^2$,或约10个分子/$\mu m^2$。在其它实施例中,分析物分子和/或记录标签以如下的平均密度固定和/或沉积在基板上:约15,约20,约25,约30,约35,约40,约45,约50,约55,约60,约65,约70,约75,约80,约85,约90,约95,约100,约105,约110,约115,约120,约125,约130,约135,约140,约145,约150,约155,约160,约165,约170,约175,约180,约185,约190,约195,约200,或约200个分子/$\mu m^2$。在其它实施例中,分析物分子和/或记录标签以如下的平均密度沉积或固定:约1个分子/$cm^2$,约10个分子/$cm^2$,约50个分子/$cm^2$,约100个分子/$cm^2$,约150个分子/$cm^2$,约200个分子/$cm^2$,约250个分子/$cm^2$,约300个分子/$cm^2$,约350个分子/$cm^2$,400个分子/$cm^2$,约450个分子/$cm^2$,约500个分子/$cm^2$,约550个分子/$cm^2$,约600个分子/$cm^2$,约650个分子/$cm^2$,约700个分子/$cm^2$,约750个分子/$cm^2$,约800个分子/$cm^2$,约850个分子/$cm^2$,约900个分子/$cm^2$,约950个分子/$cm^2$,或约1000个分子/$cm^2$。在其它实施例中,分析物分子和/或记录标签以在如下之间的平均密度沉积或固定在基板上:在约$1\times10^3$个分子/$mm^2$与约$0.5\times10^4$个分子/$mm^2$之间,在约$0.5\times10^4$个分子/$mm^2$与约$1\times10^4$个分子/$mm^2$之间,在约$1\times10^4$个分子/$mm^2$与约$0.5\times10^5$个分子/$mm^2$之间,在约$0.5\times10^5$与约$1\times10^5$个分子/$mm^2$之间,在约$1\times10^5$与约$0.5\times10^6$个分子/$mm^2$之间,或在约$0.5\times10^6$与约$1\times10^6$个分子/$mm^2$之间。在其他实施例中,沉积或固定在基板上的分析物分子和/或记录标签的平均密度可以例如是在约1个分子/$cm^2$与约5个分子/$cm^2$之间,在约5个分子/$cm^2$与约10个分子/$cm^2$之间,在约10个分子/$cm^2$与约50个分子/$cm^2$之间,在约50个分子/$cm^2$与约100个分子/$cm^2$之间,在约100与约$0.5\times10^3$个分子/$cm^2$之间,在约$0.5\times10^3$与约$1\times10^3$个分子/$cm^2$之间,在$1\times10^3$与约$0.5\times10^4$个分子/$cm^2$之间,在约$0.5\times10^4$与约$1\times10^4$个分子/$cm^2$之间,在约$1\times10^4$与约$0.5\times10^5$个分子/$cm^2$之间,在约$0.5\times10^5$与约$1\times10^5$个分子/$cm^2$之间,在约$1\times10^5$与约$0.5\times10^6$个分子/$cm^2$之间,或在约$0.5\times10^6$与约$1\times10^6$个分子/$cm^2$之间。

[1830] 在某些实施例中,控制溶液中结合剂的浓度以减少测定的背景和/或假阳性结果。

[1831] 在一些实施例中,结合剂的浓度为约0.0001nM,约0.001nM,约0.01nM,约0.1nM,约1nM,约2nM,约5nM,约10nM,约20nM,约50nM,约100nM,约200nM,约500nM或约1000nM。在其他实施例中,用于测定的可溶性缀合物浓度在约为0.0001nM与约0.001nM之间,约0.001nM与0.01nM之间,约0.01nM与0.1nM之间,约0.1nM与约1nM之间,约1nM与约2nM之间,在约2nM与约5nM之间,在约5nM与约10nM之间,在约10nM与约20nM之间,在约20nM与约50nM之间,在约50nM与约100nM之间,在约100nM与约200nM之间,在约200nM与约500nM之间,在约500nM与约1000nM之间,或大于约1000nM。

[1832] 在一些实施例中,可溶性结合剂分子与固定的分析物分子和/或记录标签之间的比率为约0.00001:1,约0.0001:1,约0.001:1,约0.01:1,约0.1:1,约1:1,约2:1,约5:1,约10:1,约15:1,约20:1,约25:1,约30:1,约35:1,约40:1,约45:1,约50:1,约55:1,约60:1,约65:1,约70:1,约75:1,约80:1,约85:1,约90:1,约95:1,约100:1,约$10^4$:1,约$10^5$:1,约$10^6$:1或更高,或上述比率之间的任何比率。可溶性结合剂分子与固定的分析物分子和/或记录标签之间的较高比率可用于驱动结合和/或编码标签/记录标签信息转移完成。这对于检测和/或分析样品中的低丰度蛋白质分析物特别有用。

[1833] D.记录标签

[1834] 记录标签例,至少一个记录标签直接或间接结合或共定位到大分子并连接到固体支持物上(参见,例如,图5)。记录标签可包含DNA,RNA,PNA,γPNA,GNA,BNA,XNA,TNA,受保护的DNA或RNA碱基,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端或突出端。在某些实施例中,在结合剂与大分子结合后,将结合剂的编码标签的识别信息转移至记录标签以产生延伸记录标签。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。

[1835] 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,连接到固体支持物连接。例如,记录标签可以通过连接反应与固体支持物连接。或者,固体支持物可包括试剂或涂层以便于直接或间接地将记录标签连接到固体支持物上。将核酸分子固定在固体支持物(例如珠粒)上的策略已在美国专利5,900,481中描述(Steinberg等人(2004,Biopolymers73:597-605);Lund等人,1988(Nucleic Acids Res.16:10861-10880);以及Steinberg等人(2004,Biopolymers73:597-605),其各自通过引用整体并入本文。

[1836] 在某些实施例中,大分子(例如肽)和相关记录标签的共定位是通过将大分子和记录标签缀合至直接连接在固体支持物表面的双功能接头来实现的(2004,Biopolymers 73:597-605)。在进一步的实施例中,三官能部分用于衍生固体支持物(例如珠粒),并且所得的双官能部分与大分子和记录标签二者偶联。

[1837] 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。

[1838] 在一个具体实施例中,单个记录标签附着到大分子(例如,肽)上,例如通过解封的N-或C-末端氨基酸的附着。在另一个实施例中,多个记录标签附着于大分子(例如,蛋白质、多肽或肽),例如附着于赖氨酸残基或肽骨架。在一些实施例中,用多个记录标签标记的大分子(例如,蛋白质或肽)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。

[1839] 在某些实施例中,记录标签包含任选的独特分子标识符(UMI),其与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。在一些实施例中,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识别来自各个大分子的序列读数。在一些实施例中,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,记录标签的每个副本包含相同的UMI。在一些实施例中,UMI具有与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。

[1840] 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。在一些实施例中,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。在一个实施例中,条形码允许多个样品或库的多重测序。条形码可用于派生大分子(例如肽)的分区、分级物、隔室、样品、空间位置或库。条形码可用于解卷积多重序列数据并识别来自单个样品或库的序列读数。例如,条形码珠粒可用于涉及样品乳化和分区的方法,例如用于蛋白质组分区的目的。

[1841] 条形码可以代表隔室标签,其中隔室,例如液滴、微孔、固体支持物上的物理区域等,被分配唯一的条形码。隔室与特定条形码的相关可以以任何方式实现,例如通过将单个条形码珠粒封装在隔室中,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。隔室内的条形码试剂用于将隔室特异性条形码加到隔室内的大分子或其片段上。应用于蛋白质分区到隔室中,条形码可用于将分析的肽映射回隔室中的其原始蛋白质分子。这能够极大地方便蛋白质识别。隔室条形码也可用于识别蛋白质复合物。

[1842] 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。

[1843] 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板或固体基板的集合(例如,平面载玻片、包含在单个管或容器中的珠粒群等)。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样品汇集在一起。或者,样品可以保持分隔开直到产生DNA编码库后,并且在DNA编码库的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,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。在一个具体实施例中,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蛋白质分析物的标准表达水平的“最佳点”内。

[1844] 在某些实施例中,来自多个不同样品的肽、多肽或蛋白质用具有样品特异性条形码的记录标签标记。复合样品条形化的肽、多肽或蛋白质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等人,2012,Assadi,Lamerz等人,2013,Akbani,Becker等人,2014,Creighton以及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。

[1845] 在某些实施例中,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。在一些实施例中,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)。

[1846] 在某些实施例中,记录标签在其末端例如3'端包含间隔区。如本文所用,在记录标签的背景下间隔区序列包括与其同源结合剂相关的间隔区序列相同的间隔区序列,或包括与其同源结合剂相关的间隔区序列互补的间隔区序列。记录标签上的末端例如3'间隔区允许同源结合剂的识别信息在第一结合循环期间从其编码标签转移到记录标签(例如,通过互补间隔区序列的退火进行引物延伸或粘性端连接)。

[1847] 在一个实施例中,间隔区序列的长度为约1-20个碱基、长度为约2-12个碱基、或长度为5-10个碱基。间隔区的长度可取决于诸如用于将编码标签信息转移到记录标签的引物延伸反应的温度和反应条件等因素。

[1848] 在优选实施例中,记录中的间隔区序列被设计成与记录标签中的其它区域具有最小的互补性;同样,编码标签中的间隔区序列应该与编码标签中的其它区域具有最小的互补性。换句话说,记录标签和编码标签的间隔区序列应该与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,隔室、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。

[1849] 如针对结合剂间隔区的描述,在一些实施例中,与大分子库相关联的记录标签共享共同的间隔区序列。在其他实施例中,与大分子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区序列互补的结合循环特异性间隔区序列,其在使用非级联延伸记录标签时是有用的(参见,例如,图10)。

[1850] 延伸记录标签的集合可以在事后级联(参见,例如,图10)。在结合循环完成后,珠粒固体支持物,每个平均包含一个或少于一个大分子的珠粒,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大分子位点)被放置在乳液中。形成乳液使得每个液滴平均至多被1个珠粒占据。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珠粒上的延伸记录标签,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等人,2008)。然后破坏乳液并对组装的延伸记录标签进行测序。

[1851] 在另一个实施例中,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区序列(Sp1)。在第一结合循环中,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、以及第二间隔区元件(Sp2)的DNA编码标签。使用至少两种不同间隔区元件的实用性是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延伸的DNA记录标签的末端产生新的Sp2间隔区元件。在第二次和随后的结合循环中,结合剂仅含有Sp2'间隔区而不是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。在另一个实施例中,第二和随后的循环可以使用结合剂特异性间隔区。

[1852] 在一些实施例中,记录标签从5'至3'方向包含:通用正向(或5')引发序列、UMI、和间隔区序列。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,样品条形码、分区条形码、隔室条形码、空间条形码或其任何组合)、和间隔区序列。在一些其他实施例中,记录标签从5'到3'方向包含:通用正向(或5')引发序列、条形码(例如,样品条形码、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、和间隔区序列。

[1853] 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"词"聚合物的化学连接。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)。在某些实施例中,在该设计中没有"间隔区"序列。词空间的大小可以从10个词到10000个词或更多词。在某些实施例中,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个单词(这仅是总10-mer词空间约$4^{10}$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大分子(肽、蛋白等)上。在一个实施例中,UMI的多样性超过UMI所附着的大分子的分子数量。通过这种方式,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分,比如隔室标签、分区条形码、空间条形码、样品条形码、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(参加,例如,Kiah等人,2015,DNA序列谱的编码(Codes for DNA sequence profiles).IEEE International Symposium on Information Theory(ISIT);Gabrys等人,2015,基于DNA存储的非对称Lee距离码(Asymmetric Lee distance codes for DNA-based storage).IEEE Symposium on Information Theory (ISIT);Laure等人,2016,Coding in 2D:通过有目的的分散性增强序列编码聚合物条形码的信息容量(Using Intentional Dispersity to Enhance the Information Capacity of Sequence-Coded Polymer Barcodes).Angew.Chem.Int.Ed.doi:10.1002/anie.201605279;Yazdi等人,2015,IEEE Transactions on Molecular,Biological and Multi-Scale Communications 1:230-248;and Yazdi等人,2015,Sci Rep 5:14138,其各自以全文引用的方式并入)。因此,在某些实施例中,本文描述的任何实施例中的延伸记录标签、延伸编码标签或双标签构建体包含作为错误纠正码的识别组分(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,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,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(美国专利第7,060,507号,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)。

[1854] 因此,在某些实施例中,编码标签、记录标签或两者的识别组分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分。在一些实施例中,识别组分选自编码器序列、条形码,UMI、隔室标签、循环特异性序列或其任何组合。

[1855] 在某些实施例中,样品中的全部或基本量的大分子(例如,蛋白质、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。大分子的标记可以在将大分子固定到固体支持物之前或之后进行。

[1856] 在其它实施例中,样品内的大分子(例如,蛋白质、多肽或肽)的子集用记录标签标记。在一个具体实施例中,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质特异性结合剂(例如,抗体,适体等)实现,靶蛋白质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,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,例如,图28A)。记录标签包含用于靶蛋白质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,例如,图28A-B)。在通过靶蛋白质特异性结合剂结合靶蛋白质后,记录标签和靶蛋白质通过其相应的反应性基团偶联(参见,例如,图28B-C)。在用记录标签标记靶蛋白质后,靶蛋白质特异性结合剂可以通过消化连接在靶蛋白质特异性结合剂的DNA捕获探针去除。例如,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,且靶蛋白质特异性结合剂可以从靶蛋白质中解离。

[1857] 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$BC_A$)标记对靶蛋白质集特异的抗体,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。蛋白质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。

[1858] 在另一个实例中,靶蛋白质特异性适体用于样品内蛋白子集的靶向记录标签标记。靶特异性适体与DNA捕获探针连接,DNA捕获探针与记录标签中的互补诱饵序列退火。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靶蛋白质偶联。适体与其靶蛋白分子结合,将记录标签带靠近靶蛋白质,使得记录标签偶联到所述靶蛋白质。

[1859] 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等人,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 and Collins 2015)。在一个优选的实施例中,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,例如,图28)的DNA捕获探针杂交(Li,Liu等人,2013)。对于光亲和标记的蛋白质靶标,使用DNA/RNA适体作为靶蛋白质特异性结合剂优于抗体,因为光亲和基团可以自我标记抗体而不是靶蛋白质。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等人,2014,Kodal,Rosen等人,2016)类似的方式,在适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)。

[1860] 在前述实施例中,除杂交之外,其他类型的连接可用于连接靶特异性结合剂和记录标签(参见,例如,图28A)。例如,如图28B所示,一旦捕获的靶蛋白质(或其它大分子)共价连接到记录标签,就能够使用设计为裂解并释放结合剂的接头共价连接两个基团。合适的接头可以连接到记录标签的多个位置,例如3’端,或连接到附着于记录标签的5’端的接头中。

[1861] E.结合剂和编码标签

[1862] 在一个方面,本文所述的试剂盒包含能够结合至分析物(例如,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))。

[1863] 在某些实施例中,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NTAA特异性结合剂与同源NTAA结合,就进行偶联反应以在其二者之间产生共价连接。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,从而增加测定的特异性。

[1864] 在某些实施例中,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。在非选择性结合的示例中,结合剂可以结合二十种标准氨基酸中的两种或更多种。

[1865] 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,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双标签分子的转移。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。

[1866] 在一个具体实施例中,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转移是有效的。在某些实施例中,结合剂具有的$K_d$为或小于约50nM,或小于约10nM,或小于约5nM,或小于约1nM,或小于约0.5nM,或小于约0.1nM。在一个具体实施例中,将结合剂以>10×、>100×或>1000×$K_d$的浓度加入到大分子中,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等人.2012)。

[1867] 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NTAA的胺基上。商购的抗DNP抗体在低nM范围内(约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。

[1868] 在某些实施例中,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。在一些实施例中,结合剂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)。

[1869] 在某些实施例中,结合剂可以结合氨基酸的翻译后修饰。在一些实施例中,肽包含一个或多个相同的或不同的翻译后修饰。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(另外参见Seo以及Lee,2004,J.Biochem.Mol.Biol.37:35-44)。

[1870] 在某些实施例中,凝集素用作结合剂,用于检测蛋白质、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(参见Zhang等人,2016,MABS 8:524-535)。

[1871] 在某些实施例中,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。

[1872] 在某些实施例中,结合剂可以是适体(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)。

[1873] 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以及G.Georgiou,2000,Ann.Rev.Biomed.Eng.2:339-76;Antibody Engineering,R.Kontermann以及S.Dubel,eds.,Springer Lab Manual,Springer Verlag(2001);U.S.专利第5,831,012号;以及S.Paul,Antibody Engineering Protocols,Humana Press(1995),等其它地方找到,。

[1874] 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标分子,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(参见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(参见Patel以及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以及Breaker,2011,RNA Biol.8;82-89;Mannironi等人,2000,RNA 6:520-27;Famulok,1994,J.Am.Chem.Soc.116:1698-1706)。

[1875] 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,Yuan等人评论,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。

[1876] 在一些实施例中,可以使用选择性结合修饰的NTAA的结合剂。例如,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,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允许随后通过埃德曼降解裂解NTAA。在另一个实施例中,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,结合剂可以改造以选择性地结合NTAA上的DNP和R基团的组合。DNP基团的添加为结合剂与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。在另一个实施例中,结合剂可以是氨肽酶,其已被工程化以识别DNP标记的NTAA,从而提供肽的氨肽酶降解的循环对照。一旦DNP标记的NTAA被裂解,就进行另一轮DNFB衍生化以结合并裂解新暴露的NTAA。在优选的实施例中,氨肽酶是单体金属蛋白酶,例如由锌活化的氨肽酶(Calcagno以及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)。在另一个实施例中,结合剂可以选择性地结合乙酰化或酰胺化的NTAA。

[1877] 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。

[1878] 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。在一些实施例中,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。对于中度至较低亲和力的结合剂,可通过使用伴随的结合/编码步骤实现有效的信息传递,或可选地,可降低编码或信息写入步骤的温度以减慢结合剂的解离速率。在一些实施例中,伴随的结合/编码步骤可以与编码或信息写入步骤的降低的温度相结合。在一些实施例中,在结合基团与分析物的初始结合之后,结合剂的结合基团可以共价地(直接地或间接地,例如通过交联)结合到分析物,以便减慢结合剂的解离速率。这可用于中度至较低亲和力结合剂。

[1879] 结合并裂解单个或小组标记(生物素化的)NTAA的工程化氨肽酶突变体已经被描述过(参见,例如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-末端裂解氨基酸的酶。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裂解N-末端氨基酸(Kishor等人,2015,Anal.Biochem.488:6-8)。然而,残基特异性氨肽酶已经被鉴定出(Eriquez等人,J.Clin.Microbiol.1980,12:667-71;Wilce等人,1998,Proc.Natl.Acad.Sci.USA 95:3472-3477;Liao等人,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。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。

[1880] 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)。在一个实施例中,结合剂的结合基团包含参与天然N-末端蛋白识别和结合的接合蛋白的进化保守的ClpS家族的成员或其变体。在以下文献中描述了细菌中接合蛋白的ClpS家族:Schuenemann等人,(2009),“通过ClpAP衔接蛋白ClpS在大肠杆菌中识别N-端规则基板的结构基础(Structural basis of N-end rule substrate recognition in Escherichia coli by the ClpAP adaptor protein ClpS)”,“EMBO Reports 10(5),以及Roman-Hernandez等人,(2009),“通过N-末端规则接合蛋白ClpS的基板选择的分子基础(Molecular basis of substrate selection by the N-end rule adaptor protein ClpS)”,NAS106(22):8888-93。还参见Guo等人,(2002),JBC 277(48):46753-62,和Wang等人的“N-端规则识别的分子基础(The molecular basis of N-end rule recognition)”,Molecular Cell 32:406-414在一些实施例中,对应于Schueneman等人识别的ClpS疏水性结合袋的氨基酸残基经修饰以产生具有所需选择性的结合基团。ClpS家族成员包括但不限于ClpS1 RPA1203,ClpS2 bll5154,ClpS2 RPA3148,ClpS BAV2091,ClpS BP2756,ClpS ECP_0896,ClpS ECSE_0939,ClpS SG1101,ClpS SCO2916,SCE19A.16c,ClpS Sputw3181_1781,ClpS VC_1143,ClpS WS0336,ClpS XCC1966,ClpS Tcr_1111,ClpS TDE_2123,ClpS ZMO1725,ClpS1 bll2636,ClpS BCAN_A1188,ClpS BamMC406_2437,ClpS Bamb_2567,ClpS BMA10247_2157,ClpS BMASAVP1_A0575,ClpS BURPS1106A_0963,ClpS Bcep1808_2597,ClpS CCNA_02552,及其片段,变体,突变体,同源物或修饰型。

[1881] 在一个实施例中,结合剂的结合基团包含ClpS2蛋白质或肽或其片段,如在Stein等人,(2016),“具有更严格的特异性的N-Degron衔接子的结构基础(Structural Basis of an N-Degron Adaptor with More Stringent Specificity)”,Structure 24(2):232-242中公开的。在一些实施例中,本文公开的ClpS蛋白质或多肽包括与参考ClpS蛋白质或多肽(例如,根癌土壤杆菌ClpS2(例如,SEQ ID NO:198),大肠杆菌ClpS(例如,SEQ ID NO:199),和/或新月柄杆菌ClpS(例如,SEQ ID NO:200))共有序列同一性的变体、突变体、同源物或

修饰型。在一个方面,结合剂的结合基团包含具有约10％的序列同一性,约15％的序列同一性,约20％的序列同一性,约25％的序列同一性,约30％的序列同一性,约35％的序列同一性,约40％的序列同一性,约45％的序列同一性,约50％的序列同一性,约55％的序列同一性,约60％的序列同一性,约65％的序列同一性,约70％的序列同一性,约75％的序列同一性,约80％的序列同一性,约85％的序列同一性,约90％,91％,92％,93％,94％,95％,96％,97％,98％,或99％的序列同一性,或约100％的序列同一性,具有SEQ ID NO:198,具有SEQ ID NO:199,和/或具有SEQ ID NO:200的多肽顺序。

[1882] 在一个实施例中,结合基团包含UBR盒识别序列家族的成员或UBR盒识别序列家族的变体。UBR识别盒描述于Tasaki等人,(2009),JBC 284(3):1884-95。例如,结合基团可以包含UBR1、UBR2或其突变体、变体或同源物。

[1883] 在某些实施例中,除了结合基团之外,结合剂还包含一种或多种可检测标签,例如荧光标签。在一些实施例中,结合剂不包含多核苷酸,如编码标签。任选地,结合剂包含合成的或天然的抗体。在一些实施例中,结合剂包含适体。在一个实施例中,结合剂包含多肽,例如衔接蛋白的ClpS家族的修饰成员,例如大肠杆菌ClpS结合多肽的变体,以及可检测标签。在一个实施例中,可检测标签是光学可检测的。在一些实施例中,可检测标签包含荧光基团、颜色编码的纳米颗粒、量子点或其任何组合。在一个实施例中,标签包含聚苯乙烯染料,聚苯乙烯染料包含核心染料分子例如FluoSphere™,尼罗红,荧光素,若丹明,衍生的若丹明染料,例如TAMRA,磷光剂,polymethadine染料,荧光亚磷酰胺,TEXAS RED,绿色荧光蛋白,吖啶,花青,花青5染料,花青3染料,5'-(2'-氨基乙基)-氨基萘-1-磺酸(EDANS),BODIPY,120ALEXA或前述任一项的衍生物或修饰。在一个实施例中,可检测的标签耐受光漂白,同时在独特且容易检测的波长下产生大量信号(例如,光子),其具有高信噪比。

[1884] 在一个具体实施例中,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)。

[1885] 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结合剂连接的编码标签和与大分子相关的记录标签中包括互补序列。

[1886] 在一些实施例中,可以采用选择性结合修饰的C-末端氨基酸(CTAA)的结合剂。羧肽酶是裂解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裂解。可以修饰羧肽酶以产生选择性结合特定氨基酸的结合剂。在一些实施例中,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。

[1887] 可以设计来用于产生本文所述方法的结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)。

[1888] 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,例如例-螺旋结构,β-发夹,β-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。在一个实施例中,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等人,2015)。

[1889] 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。在一些实施例中,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、、PNA、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。

[1890] 编码标签包含编码器序列,其提供关于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。在一些实施例中,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。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。当分析大分子群时,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1024。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。

[1891] 在一些实施例中,识别编码标签或记录标签的组分,例如编码器序列、条形码、UMI、隔室标签、分区条形码、样品条形码、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→44个编码器序列=256个编码器序列)。在另一个实施例中,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)。在另一个实施例中,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 NOs: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。

[1892] 在一些实施例中,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。

[1893] 在某些实施例中,编码标签还在一端或两端包含间隔区序列。间隔区序列为约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。在一些实施例中,间隔区的长度约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。在一些实施例中,编码标签内的间隔区比编码序列短,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,编码标签内的间隔区与编码器序列的长度相同。在某些实施例中,间隔区是结合剂特异性的,使得来自先前结合循环的间隔区仅与来自当前结合循环中的适合结合剂的间隔区相互作用。一个示例是含有间隔区序列的同源抗体对,仅当两种抗体顺序结合大分子时,允许转移信息。间隔区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,例如,图5A,"*Sp'")可任选地包含记录标签上的3'间隔区的伪互补碱基以增加Tm(Lehoudden等人,2008,2008,Nucleic Acids Res.36:3409-3419)。

[1894] 在一些实施例中,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区序列(例如,在多重结合循环方法中使用的整个结合剂库在其编码标签中具有共同的间隔区)。在另一个实施例中,编码标签包含结合循环标签,识别特定的结合循环。在其它实施例

中,结合剂库内的编码标签具有结合循环特异性间隔区序列。在一些实施例中,编码标签包含一个结合循环特异性间隔区序列。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”特异性间隔区序列,用于第二结合循环的结合剂的编码标签包含“循环2”特异性间隔区序列,等等达到“n”个结合循环。在进一步的实施例中,用于第一结合循环的结合剂的编码标签包含“循环1”特异性间隔区序列和“循环2”特异性间隔区序列,用于第二结合循环的结合剂的编码标签包括“循环2”特异性间隔区序列和“循环3“特异性间隔区序列,等等至“n”结合循环。该实施例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)。在一些实施例中,间隔区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区序列退火以引发引物延伸反应或粘性末端连接反应。

[1895] 当记录标签群与大分子相关联时,循环特异性间隔区序列也可用于将编码标签的信息级联到单个记录标签上。第一结合循环将信息从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区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”特异性间隔区序列和“循环2”特异性间隔区序列,在第二结合循环中使用的结合剂的编码标签包括“循环2”特异性间隔区序列和“循环3”特异性间隔区序列,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特异性间隔区序列与记录标签退火。在将编码标签信息转移到记录标签时,循环2特异性间隔区序列在结合循环1结束时位于延伸记录标签的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特异性间隔区序列与延伸记录标签退火。在将编码标签信息转移到延伸记录标签时,循环3特异性间隔区序列在结合周期2结束时位于延伸记录标签的3’末端,依此类推走完“n”个结合循环。该实施例规定,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区的寡核苷酸作为“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特异性间隔区、循环2特异性间隔区和“无效”编码序列的寡核苷酸增加追踪步骤。“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。“无效”寡核苷酸能够通过循环1特异性间隔区与记录标签退火,并且循环2特异性间隔区被转移至记录标签。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特异性间隔区与延伸记录标签退火。“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。

[1896] 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,如果结合循环直接嵌入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,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。

[1897] 在一些实施例中,在所述结合剂附近的第二(3')间隔区序列内编码标签包含可裂解或可缺刻的DNA链(参见,例如,图32)。例如,3'间隔区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在尿嘧啶的位置产生单核苷酸间隙。在另一个实例中,3'间隔区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,用于裂解或切刻3'间隔区序列的酶仅作用于一条DNA链(编码标签的3'间隔区),使得属于(延伸的)记录标签的双链体内的另一条链保持完整。这些实施例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(延伸的)记录标签中非变性去除结合剂,并在可用于后续结合循环的延伸记录标签上留下单链DNA间隔区序列。

[1898] 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、延伸记录标签在编码标签信息转移时折叠自己。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。

[1899] 在一些实施例中,编码标签包含分析物特异性间隔区,其能够仅在用先前识别相同分析物的结合剂延伸记录标签上引发延伸。可以使用包括分析物特异性间隔区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。在一个实施例中,第一结合事件使用具有编码标签的结合剂,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区序列;随后的结合循环接着使用具有编码的分析物特异性3'间隔区序列的结合剂。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区序列上引发延伸的通用间隔区3'序列,以及在5'末端的编码的分析物特异性间隔区,他们将用在下一结合循环中。对于匹配的同源结合剂对,第二结合剂的3'分析物特异性间隔区与第一结合剂的5'分析物特异性间隔区匹配。以这种方式,只有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。在第一结合循环中,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区相互作用。引物延伸将编码标签信息,包括分析物特异性5'间隔区,转移至记录标签。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区。

[1900] 在某些实施例中,编码标签可以进一步包含连接在编码标签的结合剂的独特分子标识符。结合剂的UMI可用于采用延伸编码标签或用于测序读数的双标签分子实施例中,其与编码器序列组合提供关于结合剂的身份和大分子的独特结合事件的数量的信息。

[1901] 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(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

验使用约$10^9$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。

[1902] 编码标签可包括在3'间隔区序列的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区序列退火后,引物延伸可以将信息从编码标签转移到记录标签,或者将信息从记录标签转移到编码标签。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转移编码标签。应当理解,对于本文涉及生成延伸编码标签的实施例,优选在记录标签的3'末端包括终止子核苷酸,以防止编码标签信息转移到记录标签。

[1903] 编码标签可以是单链分子、双链分子或部分双链。编码标签可包括平端、悬挂端或一样一个。在一些实施例中,编码标签是部分双链的,其防止编码标签与延长的延伸记录标签中的内部编码器和间隔区序列退火。

[1904] 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。在一些实施例中,编码标签可以酶促或化学方式与结合剂连接。在一些实施例中,编码标签可以通过连接与结合剂结合。在其他实施例中,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。

[1905] 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,例如,图43B)。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等人,2012,Proc.Natl.Acad.Sci.109:E690-697;Li等人,2014,J.Mol.Biol.426:309-317)。结合剂可以表达为包含SpyCatcher蛋白的融合蛋白。在一些实施例中,SpyCatcher蛋白或其变体(例如,在Keeble等人的"对膜动力学进行分析的利用SpyTag/SpyCatcher的演变的加速酰胺化(Evolving accelerated amidation by SpyTag/SpyCatcher to analyze membrane dynamics),"Angew.Chem.Int.Ed.10.1002/anie.201707623中公开的SpyTag002或SpyCatcher002变体)连接于结合剂的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SpyTag肽偶联至编码标签。

[1906] 在其他实施例中,通过SnoopTag-SnoopCatcher肽-蛋白质相互作用将结合剂连接至编码标签。SnoopTag肽与SnoopCatcher蛋白形成异肽键(Veggiani等人,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,SnoopCatcher蛋白附加在结合剂的N-末端或C-末端。可以使用标准缀合化学将SnoopTag肽偶联至编码标签。

[1907] 在其它实施例中,结合剂通过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等人,2008,ACS Chem.Biol.3:373-382)。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。

[1908] 在某些实施例中,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n NTAA是苯丙氨酸,并且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。然后如果将n NTAA(苯丙氨酸)从肽上裂解,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NTAA具有选择性)。

[1909] 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,如果在多重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)。

[1910] F.用于将编码标签信息循环转移到记录标签的试剂盒和试剂盒组分

[1911] 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转移到与大分子相关的记录标签,从而产生“延伸记录标签”。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转移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,延伸记录标签中存在的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。

[1912] 在某些实施例中,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。在这些实施例中,单个级联的延伸记录标签可以代表单个大分子(参见,例如,图2A)。如本文所提到的,将编码标签信息转移到记录标签,还包括转移到延伸记录标签,如在涉及多个连续结合事件的方法中所发生的那样。

[1913] 在某些实施例中,结合事件信息以循环方式从编码标签转移到记录标签(参见,例如,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。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。编码标签还可以包含可选的UMI序列以及编码器和间隔区序列。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,例如,图2A)。

[1914] 可以使用各种方法将与特定结合剂相关的编码标签信息转移到记录标签。在某些实施例中,编码标签的信息通过引物延伸传递到记录标签(Chan,McGregor等人,2015)。记录标签或延伸记录标签的3’-末端上的间隔区序列,与编码标签的3’末端上的互补间隔区序列退火,并且聚合酶(例如,链置换聚合酶)使用退火的编码标签作为模板延伸记录标签序列(参见,例如,图5-7)。在一些实施例中,可以将与编码标签编码器序列和5’间隔区互补的寡核苷酸预先退火至编码标签,以防止编码标签与延伸记录标签中存在的内部编码器和间隔区序列杂交。编码标签上的3’末端间隔区保持单链,优选与记录标签上的3’末端间隔

区结合。在其他实施例中,新生的记录标签可以用单链结合蛋白包被,以防止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3'末端侵入到完全双链编码标签(Bell等人,2012,Nature 491:274-278)。这种配置防止双链编码标签与内部记录标签元件相互作用,然而易受延伸记录标签被RecA包被的3'尾部的链侵入(Bell等人,2015,Elife 4:e08646)。单链结合蛋白的存在可促进链置换反应。

[1915] 在一个优选的实施例中,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)、T4DNA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-。在一个优选的实施例中,DNA聚合酶在室温和高达45℃下是有活性的。在另一个实施例中,使用嗜热聚合酶的"热启动"形式,使得聚合酶被活化并在约40℃-50℃下使用。示例性的热启动聚合酶是Bst 2.0热启动DNA聚合酶(新英格兰生物公司)。

[1916] 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Pf3SSB、复制蛋白A RPA32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UL29 SSB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)。

[1917] 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%),可以最小化错误引发或自引发事件,例如当重新编码标签的末端间隔区序列引发延伸自我延伸。

[1918] 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例中,将胸苷碱基设置在编码标签中位于远离结合剂的间隔区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,以协调容纳在记录标签的间隔区序列的3'末端上的偶发性非模板化腺苷酸(参见,例如,图43A)。以这种方式,延伸记录标签(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。

[1919] 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,例如,美国专利7,501,237)(Yang,Astatke等人,2002)。Pfu exo-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。

[1920] 在另一个实施例中,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。

[1921] 通过在记录/延伸记录标签中包括伪互补碱基,可以最小化通过扩展记录标签的

末端间隔区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等人,2008),(Hoshika,Chen等人,2010)。由于存在化学修饰,伪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,许多伪互补修饰碱基可以与天然DNA或RNA序列形成强碱基对。在某些实施例中,编码标签间隔区序列包含多个A和T碱基,可以使用亚磷酰胺寡核苷酸合成将市售的伪互补碱基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等人,2006)。

[1922] 为了使溶液中编码标签标记的结合剂与固定的蛋白的记录标签的非特异性相互作用最小化,将与记录标签间隔区序列互补的竞争(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(参见,例如,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寡核苷酸,这有效地使退火的竞争寡核苷酸从记录标签解离,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。

[1923] 在某些实施例中,记录标签上的间隔区序列与编码标签上的互补间隔区序列的退火,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区序列置换退火至记录标签的间隔区序列的任何封闭寡核苷酸。

[1924] 与特定结合剂相关的编码标签信息也可以通过连接转移到记录标签(参见,例如,图6和7)。连接可以是平端连接或粘端连接。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、T3 DNA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,例如,图7)。在图示中,通过使用"记录助手"序列与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等人,1998,Peng,Li等人,2010,El-Sagheer,Cheong等人,2011,El-Sagheer,Sanzone等人,2011,Sharma,Kent等人,2012,Roloff以及Seitz 2013,Litovchick,Clark等人,2014,Roloff,Ficht等人,2014)。在一个方面,"无间隔"连接也可以使用ssDNA连接酶完成,CircLigase I或II(例如,来自Lucigen),例如,如图46所示。"无间隔"连接的使用有利于大大减少库元件的全长和/或减少或最小化库扩增期间的模板转换。

[1925] 在一个实施例中,试剂盒包含单链DNA(ssDNA)连接酶。ssDNA连接酶能够在没有互补序列的情况下连接ssDNA的末端。例如,CircLigase™ssDNA连接酶和CircLigase™II ssDNA连接酶都是热稳定的连接酶,它们通常用于催化具有5-磷酸和3-羟基的ssDNA模板的分子内连接(即,环化)。与连接互补DNA序列上彼此相邻退火的DNA末端的T4 DNA连接酶和Ampligase® DNA连接酶相反,ssDNA连接酶在不存在互补序列的情况下连接ssDNA的末端。因此,该酶可用于从线性ssDNA制备环状ssDNA分子。环状ssDNA分子可用作滚环复制或滚环转录的基板。除了其对ssDNA的活性之外,CircLigase还具有连接单链核酸的活性,该单链核酸具有3-羟基核糖核苷酸和5-磷酸化核糖核苷酸或脱氧核糖核苷酸。

[1926] 在本公开中可以使用CircLigase™ssDNA连接酶或CircLigase™II ssDNA连接酶。这两种酶的不同之处在于,CircLigase I比CircLigase II的腺苷酰化少得多,并且为了最好的活性需要ATP。在ATP存在下,CircLigase I使ssDNA再循环。CircLigase II几乎100%被腺苷酸化,因此不必向反应缓冲液中加入ATP。CircLigase II作为化学计量反应起作用,

其中酶结合在酶活性位点中被腺苷酸化的寡聚物的5'-端,然后连接寡核苷酸并终止。由于该反应不含有ATP,CircLigase II以1:1的酶:寡核苷酸构型起作用。在具体的实施例中,本文的试剂盒包含栖热菌属噬菌体RNA连接酶,例如噬菌体TS2126 RNA连接酶(例如,CircLigase™和CircLigase II™),或古细菌RNA连接酶,例如嗜热自养甲烷杆菌RNA连接酶1。在其它方面,试剂盒可以包含RNA连接酶,例如T4 RNA连接酶,例如T4 RNA连接酶2,T4 RNA连接酶2(截短型),T4 RNA连接酶2(截短型KQ),或T4 RNA连接酶2(截短型K227Q)。

[1927] 在另一个实施例中,PNA的转移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等人,2013,Bioorgan.Med.Chem.21:3458-3464)。

[1928] 在一些实施例中,可以使用拓扑异构酶转移编码标签信息。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等人,1994,J.Biol.Chem.269:32678-32684)。

[1929] 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例中,延伸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。在一些实施例中,在检测和裂解末端氨基酸(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转移至记录标签(参见,例如,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转移,就可以在使用N-末端或C-末端降解方法检测和转移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,例如,图8B)。

[1930] 在一些实施例中,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,单独或作为结合剂库的一部分(例如,由20种标准氨基酸的结合剂组成的库和选定的翻译后修饰氨基酸)。随后是末端氨基酸裂解和与结合剂(或结合剂库)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。

[1931] 在某些实施例中,每个大分子可以采用一组记录标签,以提高编码标签信息转移的整体稳健性和效率(参见,例如,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,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。

[1932] 在图9和10中示出了这种实施例的一个示例。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粒)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转移给共定位的多个记录标签中的一个记录标签。在图9A所示的示例中,结合周期信息被编码到编码标签上存在的间隔区中。对于每个结合循环,用指定的循环特异性间隔区序列标记该组结合剂(参见,例如,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,或在间隔区序列中包含循环特异性信息以表示结合剂特性和循环数。

[1933] 如图9A所示,在第一结合循环(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区序列互补的共同间隔区序列。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'-间隔区序列。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区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区序列的引物延伸被转移到同源记录标签。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区序列互补的共同间隔区序列)。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'-间隔区序列。第二NTAA结合剂与大分子的NTAA结合,第二编码标签的信息通过引物延伸转移至记录标签。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编码标签中具有循环特异性间隔区序列,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。

[1934] 在一个替代实施例中,多个记录标签与固体支持物(例如,珠粒)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区和一个用于下一结合循环的循环特异性间隔区(参见,例如,图10A和图10B)。这种设计的原因是支持最终组装PCR步骤(参见,例如,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库进行富集、扣除和/或标准化方法。在第一结合循环(循环1)中,在第一结合剂结合后,包含循环1特异性间隔区(C'1)的编码标签的信息被转移到包含互补的循环1特异性间隔区(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特异性间隔区(C'2)的编码标签的信息被转移至末端包含互补的循环2特异性间隔区(C2)的不同记录标签。该过程持续到第n结合循环。在一些实施例中,延伸记录标签中的第n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编码标签设计的一部分并入,或者将通用反向引发序列添加到第n结合循环后的后续反应,比如使用加尾引物的扩增反应中。在一些实施例中,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,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(参见,例如,图9和图10)。在一个具体实施例中,在完成第n结合循环后,将涂覆有延伸记录标签的珠粒基板置于油乳液中,使得平均存在少于或约等于1个珠粒/液滴。然后使用组装PCR从珠粒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区序列引发来共线组装多个单独的记录标签(参见,例如,图10C)(Xiong等人,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区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区。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转移效率。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转移的冗余(参见,例如,图4B)。如果编码标签信息不是一直被转移,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,例如蛋白质。即使是短肽也可以包含大量可能的蛋白质序列。例如,10-mer肽具有2010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。

[1935] 在一些实施例中,其中蛋白质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,编码标签在临近结合剂的间隔区元件内包含可裂解或可切刻的DNA链(图32)。例如,靠近结合剂的间隔区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,靠近结合剂的间隔区可包含切刻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间隔区元件,用于随后的免疫测定循环。在一个优选的实施例中,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(参见,例如,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(参见,例如,图32F),在略微升高的温度下容易解离。

[1936] 在接近结合剂的间隔区元件内,由可裂解或可切刻的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转移编码标签信息(参见,例如,图33)。在一个优选的实施例中,接近结合剂的编码标签包含切刻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刻核酸内切酶识别和切刻。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)+切刻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转移。在每个转移步骤后,得到的延伸记录标签-编码标签双链体被切刻核酸内切酶切刻,释放附着于结合剂的截短间隔区并暴露所述延伸记录标签3'间隔区序列,其能够与其它近端结合的结合剂的编码标签退火(参见,例如,图33B-D)。设计编码标签间隔区序列中切刻基序的位置以产生亚稳态杂交体,其可以容易地与未裂解的编码标签间隔区序列交换。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(参见,例如,图33C-D)。该实施例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。

[1937] 对于涉及变性蛋白质、多肽和肽的分析的实施例,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。

[1938] G.用于记录标签信息到编码标签或双标签构建体的循环转移的试剂盒和试剂盒组分

[1939] 在另一方面,不是在结合剂与大分子结合后将信息从编码标签写入记录标签,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,样品条形码、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,例如,图11A)。在某些实施例中,在每个结合循环后,并可选地在埃德曼降解化学步骤之前,收集结合剂和相关的延伸编码标签。在某些实施例中,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,例如在循环埃德曼降解中检测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(蛋白质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序列的库元件映射回相同的细胞、蛋白质组子集、分子等,并可以重建肽序列。如果埃德曼降解过程中记录标签承受太多损坏,该实施例可能有用。

[1940] 本文提供了用于分析多个大分子的方法,其包含:(a)提供连接到固体支持物上的多个大分子和相关记录标签;(b)使多种大分子与能够结合多种大分子的多种结合剂接触,其中每种结合剂包含编码标签,编码标签具有关于结合剂的识别信息;(c)(i)将与大分子相关的记录标签的信息转移至与大分子结合的结合剂的编码标签,以产生延伸编码标签(参见,例如,图11A);或(ii)将与大分子相关的记录标签和与大分子结合的结合剂的编码标签的信息转移至双标签构建体;(d)收集延伸编码标签或双标签构建体;(e)任选地将步骤(b)-(d)重复一个或多个结合循环;(f)分析延伸编码标签或双标签构建体的集合。

[1941] 在某些实施例中,从记录标签到编码标签的信息转移可以使用引物延伸步骤完成,其中任选地封闭记录标签的3'末端以防止记录标签的引物延伸(参见,例如,图11A)。在每个结合事件和信息转移完成之后,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,记录标签由通用引发位点(U2')、条形码(例如,隔室标签"CT")、可选的UMI序列和共同间隔区序列(Sp1)组成。在某些实施例中,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,编码标签由共同间隔区序列(Sp2')、结合剂编码器序列和通用引发位点(U3)组成。在引入编码标签标记的结合剂之前,与记录标签的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),与记录标签U2'退火。另外,衔接子序列Sp1'-Sp2与记录标签Sp1退火。衔接子序列还能够与编码标签上的Sp2'序列相互作用,使记录标签和编码标签彼此接近。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双标签的构建。在跨多个结合循环收集双标签后,对双标签的集合进行测序,并通过UMI序列映射回原始肽分子。应理解,为了最大化功效,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。

[1942] 在某些实施例中,大分子是蛋白质或肽。可以通过从生物样品中片段化蛋白质来获得肽。

[1943] 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。记录标签包含识别与其相关的大分子(例如肽)的UMI。在某些实施例中,记录标签还包含隔室标签。记录标签还可以包含通用引发位点,其可以用于下游扩增。在某些实施例中,记录标签在其3'末端包含间隔区。间隔区可以与编码标签中的间隔区互补。可以封闭记录标签的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,促进大分子相关记录标签的信息转移到编码标签或大分子相关记录标签和编码标签的信息转移到双标签构建体。

[1944] 编码标签包含识别编码剂所连接的结合剂的编码器序列。在某些实施例中,编码标签还包含与编码标签连接的每种结合剂的独特分子标识符(UMI)。编码标签还可以包含通用引发位点,其可以用于下游扩增。编码标签可在其3'-末端包含间隔区。间隔区可以与记录标签中的间隔区互补,并可用于引发引物延伸反应以将记录标签信息转移到编码标签。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双标签起源的结合循环。

[1945] 可以通过引物延伸或连接实现将记录标签的信息转移到编码标签。可以通过间隙填充反应、引物延伸反应或两者,产生记录标签和编码标签的信息至双标签构建体的转移。

[1946] 双标签分子包含与延伸记录标签类似的功能组分。双标签分子可能包括衍生自记录标签的通用引发位点、衍生自记录标签的条形码(例如,隔室标签)、衍生自记录标签的可选独特分子标识符(UMI)、衍生自记录标签的可选间隔区、衍生自编码标签的编码序列、衍生自编码标签的可选独特分子标识符、结合循环特异性序列、衍生自编码标签的可选间隔区以及衍生自编码标签的通用引发位点。

[1947] 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记录标签转移至编码标签或双标签构建体(参见,例如,图12A-D)。在分析本文公开的肽的方法涉及通过埃德曼降解裂解末端氨基酸的某些实施方案中,可能需要使用对抗埃德曼降解的苛刻条件有抗性的记录标签,例如PNA。埃德曼降解方案中的一个苛刻步骤是无水TFA处理以裂解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的挑战是信息转移的酶促方法变得更加困难,即通过化学连接转移信息是优选模式。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等人,1998,Genome Res.8:1142-1153;Peng et al.,2010,Eur.J.Org.Chem.4194-4197;El-Sagheer等人,2011,Org.Biomol.Chem.9:232-235;El-Sagheer等人,2011,Proc.Natl.Acad.Sci.USA 108:11338-11343;Litovchick等人,2014,Artif.DNA PNA XNA 5:e27896;Roloff等人,2014,Methods Mol.Biol.1050:131-141)。

[1948] 为了创建组合PNA条形码和UMI序列,可以组合连接来自n-mer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$1000^4=10^{12}$个代码的编码空间。以这种方式,从一组起始的4000个不同DNA模板序列中,可以产生超过12个PNA代码(参见,例如,图12A)。通过调整级联词的数量或调整基本词的数量,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记录标签的转移(参见,例如,图12B)。在PNA模板上组装DNA单词和化学连接DNA词后,所得到的中间体可用于将信息转移到编码标签/从编码标签转移信息(参见,例如,图12C和图12D)。

[1949] 在某些实施例中,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粒、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粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[1950] 在某些实施例中,结合剂是蛋白质或多肽。在一些实施例中,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。在某些实施例中,结合剂结合肽的单个氨基酸残基、二肽、三肽或翻译后修饰。在一些实施例中,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。在一些实施例中,结合剂结合N-末端肽、C-末端肽或内

部肽。在一些实施例中,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。

[1951] 在某些实施例中,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。在一些实施例中,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。

[1952] 在某些实施例中,在产生延伸编码标签或双标签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例中,记录标签的扩增提供额外的记录标签作为将信息转移到编码标签或双标签构建体的模板(参见,例如,图13和图14)。乳液融合PCR可用于将记录标签信息转移至编码标签或产生双标签构建体群。

[1953] 可在分析之前,扩增产生的延伸编码标签或双标签构建体的集合。对延伸编码标签或双标签构建体的收集的分析,可能包含核酸测序方法。核酸测序方法是通过合成测序、通过连接测序通过、杂交测序、polony测序、离子半导体测序或焦磷酸测序。。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采用高级显微镜的DNA的直接成像。

[1954] 埃德曼降解和化学标记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。在某些实施例中,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。在另一个实施例中,也可以使用具有受保护的环外胺的核酸(Ohkubo,Kasuya等人,2008)。在另一个实施例中,"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以及Hirs,1968)。由于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。

[1955] 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 N-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等人,2008)。在一个实施例中,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。在另一个实施例中,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。在另一个实施例中,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转移到编码标签。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等人,1999)。

[1956] 虽然可以使用类似埃德曼的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例利用延伸记录标签、延伸编码标签和双标签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和Bertozzi 2009,Basle,Joubert等人,2010)(Spicer和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,例如,图13)。

[1957] 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暴露于苛刻的埃德曼化学的问题。在循环结合模式中,还可以使用延伸编码标签或双标签来提供组成信息(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双标签方法来读出该组成信息。如果与UMI和隔室标签信息结合,延伸编码标签或双标签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。延伸编码标签或双标签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质组,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。

[1958] 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,数字蛋白质分析:通过单分子检测实现蛋白质诊断和蛋白质组学技术(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质序列表征。

[1959] 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等人,2015,PLoS Comput.Biol.11:e1004080)(Yao,Docter等人,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等人,2010)。部分肽序列信息的挑战是映射回蛋白质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质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个不同的蛋白质序列,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)。

[1960] 具体地,通过创新地使用隔室标签和UMI显著增强了部分肽序列或组合物的可作图性。即蛋白质组被初步分区成条形码化的隔室,隔室条形码也附着在UMI序列上。隔室条形码是隔室独有的序列,所述UMI是隔室内每个条形码分子独有的序列(参见,例如,图16)。在一个实施例中,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珠粒的DNA隔室条形码杂交,使DNA标签标记的多肽与珠粒

表面直接相互作用(参见,例如,图31)。引物延伸步骤将信息从珠粒连接的隔室条形码转移到多肽上的DNA标签(参见,例如,图20)。在另一个实施例中,通过将含有UMI的条形码珠粒和蛋白分子共包封成乳液液滴来完成这种分区。此外,液滴可选地含有将蛋白质消化成肽的蛋白酶。许多蛋白酶可用于消化报告分子标记的多肽(Switzar,Giera等人,2013)。酶促连接酶(例如butelase I)与蛋白酶的共包封可能需要对酶进行修饰,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等人,2010)。消化后,肽会连接到条形码-UMI标签上。在优选实施例中,条形码-UMI标签保留在珠粒上以促进下游生物化学操作(参见,例如,图13)。

[1961] 在条形码-UMI与肽连接后,破坏乳液并收获珠粒。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。肽的两种类型的信息都可用于将其映射回蛋白质组的子集。通常,序列信息映射回蛋白质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,可以唯一地识别蛋白质或肽的来源蛋白质。以这种方式,可以表征和定量整个蛋白质组。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转移的初级或PTM氨基酸信息一起使用,以产生最终的映射肽信息。

[1962] 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条形码肽的珠粒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)(参见,例如,图13)。最易于化学标记的氨基酸(AA)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以及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标签区分相同类型的多个氨基酸的存在。每个标记分子在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”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。

[1963] 在一个实施例中,在标记AA之后,通过隔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隔室化的肽的氨基酸组成的延伸编码标签或双标签,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。对双标签进行测序后,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。

[1964] 在一个具体实施例中,标记的肽复合物与珠粒解离(参见,例如,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。在一个具体实施例中,通过产生微乳液液滴来实现这种隔室化(Shim,Ranasinghe等人,2013,Shembekar,Chaipan等人,2016)。除肽复合物外,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,例如,图13)。在最初的5-10个PCR循环之后,降低退火温度,使得U2tr和所述氨基酸编码标签上的Sptr参与扩增,并进行另外约10轮。三引物乳液PCR有效地将肽UMI条形码与所有AA编码标签组合,产生

代表肽及其氨基酸组成的双标签库。也可以进行三引物PCR和标签级联的其他方式。另一个实施例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测序的库元件。

[1965] 以这种方式,库元件的不同序列组分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)。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,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,唯一地识别和定量蛋白质。

[1966] H.用于末端氨基酸(TAA)标记的试剂盒和试剂盒组分

[1967] 在某些实施例中,在本文所述的方法中,使肽与结合剂接触之前修饰或标记肽的末端氨基酸(例如,NTAA或CTAA)。本文描述了用于这种修饰和/或标记的试剂盒和试剂盒组分。

[1968] 在一些实施例中,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。埃德曼降解通常使用苯基异硫氰酸酯(PITC)来标记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,导致氨基酸从剩余肽中裂解。

[1969] 可用于标记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等人,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等人,1991,Eur.J.Biochem.196:679-685)。解封肽N-末端的方法是本领域已知的(参见,例如,Krishna et al.,1991,Anal.Biochem.199:45-50;Leone等人,2011,Curr.Protoc.Protein Sci.,Chapter 11:Unit11.7;Fowler等人,2001,Curr.Protoc.Protein Sci.,Chapter 11:Unit11.7,其中每篇都通过引用整体并入本文)。

[1970] 丹磺酰氯与肽的游离胺基反应,得到NTAA的丹酰基衍生物。DNFB和SNFB与肽的与ε-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也-胺反应。DNFB还与酪氨酸和组氨酸氨基酸残基反应。SNFB比DNFB对胺基具有更好的选择性,并且优选用于NTAA修饰(Carty和Hirs,1968)。在某些实施例中,在多肽蛋白酶消化成肽之前,赖氨酸赖-胺用有机酸酐预封闭。

[1971] 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),其裂解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}。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的温和孵育导致N-末端氨基酸的裂解{Hamada,2016#383}。这有效地提供了温和的埃德曼类化学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。

[1972] 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等人,2014,Angew Chem Int Ed Engl 53:3024-3028)和氨基酰基转移酶(Wagner等人,2011,J Am Chem Soc 133:15139-15147)标记。

[1973] 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。一个示例是使用离子液体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等人,2005)。埃德曼降解涉及异硫氰酸酯(如PITC)与肽的氨基N-末端的反应。因此,在一个实施例中,离子液体通过提供更温和的标记和降解条件用于改善埃德曼降解过程的效率。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ddH2O(1:1:1vol./vol./vol.)的溶液中使用5%(vol./vol.)PITCC 60分钟(Wang,Fang等人,2009)。在优选的实施例中,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多肽内被封闭。这样,在肽测序反应期间,只有NTAA的肽α-胺基易于修饰。当使用DNFB(Sanger'试剂)和丹磺酰氯时,这尤其重要。

[1974] 在某些实施例中,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等人,1991)。解封肽的N-末端的多种方法是本领域已知的(参见,例如,Krishna等人,1991,Anal.Biochem.199:45-50;Leone等人,2011,Curr.Protoc.Protein Sci.,Chapter 11:Unit11.7;Fowler等人,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)。

[1975] CTAA可以用许多不同的羧基反应试剂进行修饰。在另一个示例中,用混合酸酐和异硫氰酸酯修饰CTAA以产生硫代乙内酰脲((Liu以及Liang 2001)以及美国专利第5,049,507号)。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。

[1976] I.用于末端氨基酸裂解的试剂盒和试剂盒组分

[1977] 在涉及分析肽的某些实施例中,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标签构建体中之后,从肽中除去或裂解末端氨基酸以暴露新的末端氨基酸。在一些实施例中,末端氨基酸是NTAA。在其它实施例中,末端氨基酸是CTAA。

[1978] 末端氨基酸的裂解可以通过许多已知技术完成,包括化学裂解和酶促裂解。化学裂解的一个示例是埃德曼降解。在埃德曼降解肽期间,n NTAA在温和的碱性条件下与异硫氰酸苯酯(PITC)反应,形成苯基硫代氨基甲酰基-NTAA衍生物。接下来,在酸性条件下,所述

苯基硫代氨基甲酰基-NTAA衍生物被裂解,产生游离的噻唑啉酮衍生物,从而将肽的第(n-1)氨基酸转化为N-末端氨基酸(第(n-1)NTAA)。此过程中的步骤如下所示:

[1979] 如上所述,典型的埃德曼降解需要在长的孵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。

[1980] 为了将化学埃德曼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。在一个实施例中,化学埃德曼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(参见,例如,Barrett,1985,Tetrahedron Lett.26:4375-4378,其全部内容通过引用并入)。NTAA的裂解也可以使用硫酰化降解来完成,与埃德曼降解相比,其使用较温和的裂解条件(见美国专利4,863,870)。

[1981] 在另一个实施例中,无水TFA的裂解可以用"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(参见,美国专利公开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),进行制备。C25G突变除去催化半胱氨酸残基,同时选择三个突变(G65S、A138C、L160Y)以与埃德曼试剂(PITC)的苯基基团产生空间配合。

[1982] NTAA的酶促裂解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裂解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,以使其仅裂解例如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末端裂解单个氨基酸,并允许控制降解循环。在一些实施例中,修饰的氨肽酶对氨基酸残基身份是非选择性的,同时对N-末端标签具有选择性。在其他实施例中,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)仅裂解存在DNP/SNP基团的N-末端氨基酸。结合并裂解标记的(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公开第WO2010/065322号)。

[1983] 在某些实施例中,紧密的单体金属酶促氨肽酶被工程化以识别和裂解DNP标记的NTAA。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等人,2010)、Streptomyces griseus(SGAP)、Vibrio proteolyticus(VPAP),(Spungin and Blumberg 1989,Ben-Meir,Spungin等人,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。

[1984] 在另一个实施例中,通过将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裂解。此外,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,而是仅识别标记的N-末端。在一个优选的实施例中,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)。

[1985] 在另一个实施例中,通过使用工程化的酰基肽水解酶(APH)裂解乙酰化的NTAA来实现环状裂解。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等人,2014)。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。

[1986] 在另一个实施例中,NTAA的脒化(胍基化)采用来促使温和裂解所述标记的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)。

[1987] 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板上(例如,多孔珠粒)。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固定的肽相关的记录标签。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裂解。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。在一个具体实施例中,单一工程化氨肽酶或APH普遍裂解具有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例中,使用两种、三种、四种或更多种工程化的氨肽酶或APH来裂解标记的NTAA的库。

[1988] 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等人,2001,Fernandez-Gacio,Uguen等人,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底物后,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裂解富集活性AP。

[1989] 在本文提供的任何实施例中,NTAA裂解试剂向NTAA的募集可以通过嵌合裂解酶和嵌合NTAA修饰剂增强,其中嵌合裂解酶和嵌合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,例如,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裂解酶(链霉亲和素-Edmanase)募集到修饰的NTAA,从而提高裂解酶的亲和力和效率。修饰的NTAA被裂解并随着相关的裂解酶一起从肽

中扩散出来。在嵌合Edmanase的例子中,该方法有效地将亲和力KD从μM增强至亚皮摩尔。通过使用与记录标签相互作用的裂解剂上的DNA标签进行束缚,也可以实现类似的裂解增强(参见图44)。

[1990] 作为NTAA裂解的替代方案,二肽基氨基肽酶(DAP)可用于从肽上裂解最后两个N-末端氨基酸。在某些实施例中,可以裂解单个NTAA(参见,例如,图45)。图45描绘了N-末端降解的方法,其中butelase I肽底物的N-末端连接将TEV内肽酶底物附着至肽的N末端。附着后,TEV内肽酶从测试肽(经历测序的肽)裂解新连接的肽,留下与NTAA连接的单个天冬酰胺(N)。与DAP孵育,其从N末端裂解两个氨基酸,导致原始NTAA的净去除。整个过程可以在N端降解过程中循环。

[1991] 对于涉及CTAA结合剂的实施例,从肽裂解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的酶促裂解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裂解。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C-末端裂解单个氨基酸,并允许控制降解循环。在一些实施例中,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。在其他实施例中,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。

[1992] J.用于延伸记录标签、延伸编码标签或双标签的处理和分析的试剂盒和试剂盒组分

[1993] 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双标签库。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,如通过合成测序、通过连接测序、通过杂交测序、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。

[1994] 可以以各种方式扩增延伸记录标签、延伸编码标签或双标签的库。延伸记录标签、延伸编码标签或双标签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等人,2007)。在一些实施例中,对于具有重复条形码或间隔区序列的序列,PCR模板转换的风险高,并且可以使用乳液PCR或线性扩增来降低该风险。例如,可以使用US 9,593,375 B2中公开的乳液PCR,该文献在此引入作为参考用于所有目的。

[1995] 或者,延伸记录标签、延伸编码标签或双标签的库可以进行线性扩增,例如,通过使用T7 RNA聚合酶和逆转录(RT)聚合酶的模板DNA的体外转录来复制回cDNA。在一个方面,线性扩增还减少或消除模板转换。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、延伸编码标签或双标签的库。延伸记录标签、延伸编码标签或双标签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、延伸编码标签或双标签的5'端、3'端或两端。可以添加到所述延伸记录标签、延伸编码标签或双标签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、延伸编码标签或双标签的:允许在单个测序运行中复用的多个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。制备下一代测序的库扩增的实例如下:使用从约1mg珠粒(约

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℃持续30秒,然后98℃持续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℃。

[1996] 在某些实施例中,在扩增之前,期间或之后,延伸记录标签、延伸编码标签或双标签库可以进行靶标富集。靶标富集可用于在测序之前从延伸记录标签、延伸编码标签或双标签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质的高度特异性结合剂,蛋白质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(参见,Mamanova等人,2010,Nature Methods 7:111-118;Bodi et al.,J.Biomol.Tech.2013,24:73-86;Ballester等人,2016,Expert Review of Molecular Diagnostics 357-372;Mertes等人,2011,Brief Funct.Genomics 10:374-386;Nilsson等人,1994,Science 265:2085-8;每个都通过引用整体并入本文)。

[1997] 在一个实施例中,通过基于杂交捕获的测定来富集延伸记录标签、延伸编码标签或双标签库(参见,例如,图17A和17B)。在基于杂交捕获的测定中,延伸记录标签、延伸编码标签或双标签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粒)和它们的亲和标签,杂交到靶特异性寡核苷酸的延伸记录标签、延伸编码标签或双标签被“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,从珠粒洗脱)富集的延伸记录标签、延伸编码标签或双标签。

[1998] 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。

[1999] 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、延伸编码标签或双标签互补。诱饵寡核苷酸与延伸记录标签、延伸编码标签或双标签上的间隔区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。在一些实施例中,间隔区相对于编码序列的长度在编码标签设计中被最小化,或者间隔区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区。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等人,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。G-quadruplex结构在阳离子存在下稳定,例如,$K^+$离子与$Li^+$离子。

[2000] 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。

[2001] 为了在代表其片段(例如肽)的延伸记录标签、延伸编码标签或双标签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蛋白质的完整核酸。

[2002] 在另一个实施例中,引物延伸和基于连接介导的扩增富集(AmpliSeq、PCR、TruSeq TSCA等)可用于选择和调节代表大分子子集的库元件的富集的分级物。竞争寡核苷酸也可用于调节引物延伸、连接或扩增的程度。在最简单的实施例中,这可以通过混合包含通用引物尾的靶标特异性引物和缺乏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。在其它实施例中,用包含的除延伸引物以外的杂交来调节经历引物延伸、连接或扩增的库元件的分数。

[2003] 靶标富集方法也可以以用在负选择模式中,以在测序之前从库中选择性地去除延伸记录标签、延伸编码标签或双标签。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粒的实施例中,保留上清液用于测序,而不分析结合在珠粒上的诱饵寡核苷酸:延伸记录标签、延伸编码标签或双标签二标签杂交体。可以去除的不需要的延伸记录标签、延伸编码标签或双标签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、延伸编码标签或双标签。

[2004] 与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杂交所述靶标,有效调节富集期间拉下的靶标分数(参见,例如,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。

[2005] 另外,库标准化技术可用于从延伸记录标签、延伸编码标签或双标签库中移除过多的种类。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库。在一个示例中,标准化可以通过使双链库变性并允许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,丰度库元素比欠丰度元素更快地重新退火(Bochman,Paeschke等人,2012)。可以使用本领域已知的方法将ssDNA库元件与丰度dsDNA库元件分离,例如羟基磷灰石柱上的色谱法(VanderNoot等人,2012,Biotechniques 53:373-380)或用来自Kamchatka蟹的双体特异性核酸酶(DSN)处理库(Shagin等人,2002,Genome Res.12:1935-42),双体特异性核酸酶(DSN)破坏dsDNA库元件。

[2006] 在连接到固体支持物和/或得到的延伸记录标签库之前,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。

[2007] 在一些实施例中,通过连接或末端互补PCR连接延伸记录标签、延伸编码标签或双标签的库以产生分别包含多种不同延伸记录标签、延伸编码标签或双标签的长DNA分子(Du等人,2003,BioTechniques35:66-72;Muecke等人,2008,Structure 16:837-841;U.S.Patent No.5,834,252,每个通过引用整体并入本文)。该实施例例如用于纳米孔测序,其中通过纳米孔测序装置分析长链DNA。

[2008] 在一些实施方案中,对延伸记录标签、延伸编码标签或双标签进行直接的单分子

分析(参见,例如,Harris等人,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)上珠粒,直接分析延伸记录标签、延伸编码标签或双标签,其中流通槽或珠粒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等人,2004,Genome Res.14:970-7)来确定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等人,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。

[2009] 在对延伸报告标签、延伸编码标签或双标签库进行测序后,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、蛋白质、蛋白质复合物)并且与细胞中总体大分子类型(例如肽、多肽、蛋白质大分子的蛋白质组)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,在特定的实施例中,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。

[2010] 在一些实施例中,可以针对特定测序分析平台优化编码标签序列。在特定实施例中,测序平台是纳米孔测序。在一些实施例中,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延伸记录标签,则条形码序列(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,鉴于纳米孔测序的单碱基准确度仍然相当低(75%-85%),但"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等人,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区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签通过孔时,双链区将在收缩区瞬时停顿,使得能够读出与双链区相邻的约3个碱基的电流。在DI纳米孔测序的特定实施方案中,编码器序列以这样的方式设计:与间隔区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等人,2010,Proc.Natl.Acad.Sci.USA 107:16060-5)。作为无马达DI排序的替代方案,间隔区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,从而能够读出相邻的编码器序列(Shim,Tan等人,2009,Zhang,Zhang等人,2016)。在经过所述停顿之后,下一个间隔区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。

[2011] 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。多个大分子可以源自相同的样品或不同的样品。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。

[2012] 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等人,2014,Creighton以及Huang,2015,Nishizuka以及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。

[2013] K.经由NTAA识别、记录标签延伸、和NTAA裂解的循环轮的大分子表征的试剂盒和试剂盒组分

[2014] 在某些实施例中,本公开内容提供的使用本试剂盒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大分子相关的记录标签。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。

[2015] 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,例如,图3,图4,图41和图42),在第一结合剂与n个氨基酸的肽的第n NTAA接触和结合后,转移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,n NTAA如本文所述被裂解。第n NTAA的裂解将肽的第n-1氨基酸转化为N-末端氨基酸,其在本文中称为第(n-1)NTAA。如本文所述,第n NTAA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。如果第n NTAA被标记,则第(n-1)NTAA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第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,其统一代表肽)。第(n-1)NTAA的裂解将肽的第(n-2)氨基酸转化为N-末端氨基酸,其在本文中称为第(n-2)NTAA。另外的结合、转移、裂解和任选的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,第n“次序”是指使用结合剂及其相关的编码标签的第n结合循环,或指创建延伸记录标签的n个结合循环。

[2016] 在一些实施例中,第一结合剂和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,例如形成结合剂库。在另一个示例中,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同时加入到大分子中,而不是合并在一起。在一个实施例中,结合剂库包含至少20种选择性结合20种标准天然氨基酸的结合剂。

[2017] 在其它实施例中,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。在某些实施例中,优选同时使用多种结合剂,因为平行方法节省时间并且结合剂处于竞争中,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。

[2018] 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。在代表肽并且由埃德曼降解样的裂解方法产生的级联延伸记录标签的实例中,如果编码标签具有两侧各有5个碱基的间隔区的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×埃德曼降解循环数量。对于20个循环的运行,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。

[2019] 在最终结合循环和最终结合剂的编码标签信息转移到延伸记录标签之后,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记录标签加盖。在一些实施例中,记录标签中的通用正向引发位点与附加到最终延伸记录标签的通用反向启发位点兼容。在一些实施例中,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 NO:133)。可以附加有义或反义P7,这取决于所述记录标签的链义。延伸记录标签库可以直接从固体支持物(例如珠粒)裂解或扩增,并用于传统的下一代测序分析和方案中。

[2020] 在一些实施例中,在单链延伸记录标签库中进行引物延伸反应以复制其互补链。

[2021] NGPS肽测序测定包含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。这通过使用存在于编码标签序列中的循环特异性条形码来实现。

[2022] 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)。

[2023] 通过分级、隔室化、和有限的结合性树脂的蛋白标准化。

[2024] 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。在某些实施开中,在分析之前从样品中扣除某些蛋白种类(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。在某些实施例中,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),来分级蛋白质样品来调节,或者将分级物分区到加载有限蛋白质结合能力的珠粒/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。洗去每个隔室化分级物中的过量蛋白质。

[2025] 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。示例性蛋白结合珠粒/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)。通过限制珠粒/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珠粒结合,并且过量蛋白质被除去。

[2026] L.用于单细胞蛋白质组分区或分子亚取样的试剂盒和试剂盒组分

[2027] 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质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、样品中存在的蛋白质群的组成、或蛋白质大分子的起源,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,蛋白质形式表征)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(参见,例如,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(例如蛋白质形式)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。

[2028] 分区是指将独特条形码随机分配给样品中大分子群的大分子亚群。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。

[2029] 通过独特的隔室标签识别已经分隔到多个(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。

[2030] 本公开内容提供了通过将复杂蛋白质组样品(例如,多个蛋白质复合物、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,例如,图18-20)。隔室任选地包含连接多个隔室标签的固体支持物(例如,珠粒)。将多个蛋白质复合物、蛋白质或多肽片段化成多个肽,然后在足以允许多个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与多个隔室标签接触,从而产生多个隔室标记的肽。或者,多个蛋白质复合物、蛋白质或多肽在足以允许多个蛋白质复合物、蛋白质或多肽与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。然后从多个隔室收集隔室标记的蛋白质复合物、蛋白质或多肽,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。

[2031] 在某些实施例中,通过引物延伸(参见例如图5和图48)或连接(参见,例如,图6、图

46和图47)将室标签信息转移到与分析物,例如大分子(例如多肽)相关的记录标签。

[2032] 在一些实施例中,隔室标签在隔室内的溶液中是游离的。在其他实施例中,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)或珠粒或隔室内的珠粒。

[2033] 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。在某些实施例中,每个隔室平均含有单个细胞。

[2034] 固体支持物可以是任何支持表面,包括但不限于珠粒、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、基于硝酸纤维素的聚合物表面、纳米颗粒或微球。用于固体支持物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。在某些实施例中,固体支持物是珠粒,例如,聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[2035] 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的珠粒的隔室的各种方法。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)。在某些实施例中,蛋白质组与隔室标记的珠粒、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。在某些实施例中,可以用变性剂处理蛋白质组以评估蛋白质或肽的肽组分。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。

[2036] 隔室标签包括条形码,其任选地在一侧或两侧侧接间隔区或通用引物序列。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,例如,图22A-B)。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)。隔室标签可包含官能部分,例如附着于间隔区上,用于偶联肽。在一个实例中,官能部分(例如,醛)是指能够与多个肽上的N-末端氨基酸残基反应的基团。在另一个实例中,官能部分能够与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。在另一个实施例中,官能部分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,还包含含有蛋白连接酶(例如,butelase I或其同源物)的识别序列的肽以允许将隔室标签连接至目标肽(参见,例如,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、间隔区序列,通用引发位点或其任何组合。UMI序列通常与隔室内的隔室标签群不同。在某些实施例中,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。

[2037] 在某些实施例中,隔室标签可以通过印刷、点样、喷墨打印到隔室中形成隔室标签。在某些实施例中,如Klein等人,2015,Cell 161:1187-1201;Macosko等人,2015,Cell 161:1202-1214;以及Fan等人,2015,Science 347:1258367等描述的通过均分与合并寡核苷酸连接或合成形成多个隔室标记的珠粒,其中每个珠粒有一种条形码类型。隔室标记的珠粒也可以通过单独合成或固定形成。在某些实施例中,隔室标记的珠粒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,另一部分包含官能部分,其上可以偶联消化的肽(参见,例如,图19和图20)。

[2038] 在某些实施例中,多个隔室内的多种蛋白质或肽用蛋白酶片段化成多个肽。蛋白酶可以是金属蛋白酶。在某些实施例中,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示例包括: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,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)。

[2039] 在具有隔室标签的肽条形码的某些实施例中,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质/多肽(参见,例如,图2B和图20A)。然后将DNA标签标记的多肽分配进入包含隔室标签(例如,与液滴中包含的珠粒结合的DNA条形码)的隔室(参见,例如,图20B),其中隔室标签包含识别每个隔室的条形码。在一个实施例中,单个蛋白质/多肽分子与与珠粒相关的单一种类的DNA条形码共同包封(参见,例如,图20B)。在另一个实施例中,隔室可以构成珠粒的表面,其附着有与PCT开第WO2016/061517号(通过引用整体并入)中所描述的类似的隔室(珠粒)标签,除了被应用于蛋白质而不是DNA之外。隔室标签可包括条形码(BC)序列、通用引发位点(U1')、UMI序列、和间隔区序列(Sp)。在一个实施例中,伴随着分区或在其后,将隔室标签从珠粒上裂解下并与附着于多肽的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。为了在珠粒上分区,DNA标签标记的蛋白质可以直接与珠粒表面上的隔室标签杂交(参见,例如,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",或从珠粒裂解的隔室标签),并且使用基于聚合酶的引物延伸步骤将条形码和UMI信息写入多肽上的DNA标签以产生隔室条形码记录标签(参见,例如,图20D)。LysC蛋白酶消化可用于将多肽裂解成在其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,例如,图20E)。在一个实施例中,LysC蛋白酶被工程化以耐受DNA标记的赖氨酸残基。将所得到的记录标签标记的肽以适当的密度固定到固体基板(例如,珠粒)上,以使记录标签标记的肽之间的分子间相互作用最小化(参见,例如,图20E和20F)。

[2040] 肽附着到隔室标签(或反之亦然),可以直接附着到固定的隔室标签,或附着到其互补序列(如果是双链的)。或者,隔室标签可以从固体支持物或隔室的表面分离,且肽和溶液相隔室标签在隔室内连接。在一个实施例中,隔室标签上(例如,在寡核苷酸的末端)的官能部分是醛,其通过席夫碱(Schiff base)直接与所述肽的胺N-末端偶联(参见,例如,图16)。在另一个实施例中,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末端连接至消化肽的N末端(Nguyen,Wang等人,2014,Nguyen,Cao等人,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。

[2041] 在某些实施例中,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,例如,图18)。在一些实施例中,在从多个隔室收集隔室标记的肽之后,隔室标记的肽连接到与记录标签相关联固体支持物。然后可以将隔室标签信息从隔室标记的肽上的室标签转移到相关的记录标签(例如,经由从记录标签和隔室标签内的互补间隔区序列引发的引物延伸反应)。在一些实施例中,然后在根据本文所述的方法进行肽分析之前从隔室标记的肽中除去隔室标签。在进一步的实施例中,最初用于消化所述多种蛋白质的序列特异性蛋白酶(例如,Endo AspN)也用于在隔室标签信息转移至相关记录标签后从肽的N末端移除隔室标签(参见,例如,图22B)。

[2042] 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,例如,图21)。隔室化的挑战在于处理隔室的内部。在某些实施例中,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条形码至目标分子的转移。

[2043] 在用含隔室标签(条形码)的记录标签标记蛋白质/肽后,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/附着在肽或蛋白质分子上的记录标签的分子内转移。通过控制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转移。

[2044] 在某些实施例中,隔室标签不必对于隔室群中的每个隔室是唯一的。隔室群中的隔室子集(两个,三个,四个或更多个)可共享相同的隔室标签。例如,每个隔室可以包含珠粒表面群,其用于从样品中捕获大分子亚群(每个珠粒捕获许多分子)。此外,珠粒包括隔室条形码,其能够附着到捕获的大分子上。每个珠粒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粒上复制(许多珠粒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。分区条形码可以包含相同的隔室条形码,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品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10000个不同隔室条形码的珠粒。此外,假设在给定的测定中,在测定中使用100万个珠粒的群。平均而言,每个隔室条形码有100个珠粒(泊松分布)。进一步假设珠粒捕获了1000万个大分子的聚群。平均而言,每个珠粒有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。

[2045] 在另一个实施例中,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,例如,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多肽体(内部氨基酸)如图2B所示。来自附着于肽末端的记录标签的信息通过酶促乳液PCR(Williams,Peisajovich等人,2006,Schutze,Rubelt et al.2011)或乳液体外转录/逆转录(IVT/RT)步骤转移至DNA标签。在优选的实施例中,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等人,2012,Gupta,Eral等人,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/SP6RNA聚合酶启动子序列以产生与附着于多肽体的DNA标签杂交的转录物(Ryckelynck,Baudrey等人,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。

[2046] 通过珠粒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以及Virta 2015,Spencer,Tamminen等人,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等人,2015,Gunderson,Steemers等人,2016,Zilionis,Nainys等人,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、通过将条形码珠粒引入液滴(10X Genomics)、或通过液滴组分的组合条形码后封装、凝胶化使用和均分与合并组合条形码,如Gunderson等人描述。(Gunderson,Steemers等人,2016)和PCT公开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等人,2017)。

[2047] 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,特别是当蛋白质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,就难变性蛋白质,蛋白酶消化得到的多肽,并同时用DNA条形码标记肽。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以及Virta 2015,Spencer,Tamminen等人,2016)(Gunderson,Steemers等人,2016)。优选地,包封的蛋白质与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。

[2048] M.用于组织和单细胞空间蛋白质组学的试剂盒和试剂盒组分

[2049] 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细胞组织内的蛋白质的空间位置,那么组织切片内蛋白质分析物的空间分布可以在序列分析后重建,就像是如Stahl等人所述的用于空间转录组学的研究(2016,Science 353(6294):78-82)以及rosetto等人,(Corsetto,Bienko等人,2015)。空间条形码的附着可以通过从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,组织切

片中的所述蛋白质可以用DNA记录标签标记,然后用蛋白酶消化蛋白质以释放标记的肽,其可以扩散并与阵列上的空间条形码杂交。然后可以将条形码信息转移(酶促或化学)到与肽连接的记录标签上。

[2050] 组织内的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,其中阵列上的每个特征具有空间可识别条形码(参见,例如,图23)。为了将阵列条形码附着到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阵列条形码信息转移至DNA标签。或者,不是允许标记的肽扩散到阵列表面,而是可以裂解阵列上的条形码序列并使其扩散到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,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,其中记录标签包含含有蛋白质/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重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。

[2051] 在另一个实施例中,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质成分/PTM(Christoforou等人,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。在一个实施例中,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。

[2052] N.基于纳米孔的测序

[2053] 在任何前述实施例中,分析编码标签和/或记录标签(和/或双标签、隔室标签或分区标签,如果适用的话)或其任何部分(例如通用引物、间隔区、UMI、记录标签条形码、编码序列、结合循环特异性条形码)的序列信息,可以使用单分子测序方法,例如基于纳米孔的测序技术来完成。在一个方面,单分子测序方法是直接单分子测序方法。WO 2017125565 A1公开了示例性的基于纳米孔的测序的某些方面,其内容通过引用以其整体并入。

[2054] DNA和RNA的纳米孔测序可以通过DNA和RNA的链测序和/或外测序来实现。链测序包含当多核苷酸模板的核苷酸穿过纳米孔时直接确定样品多核苷酸链的核苷酸碱基的方法。或者,多核苷酸链的链测序通过测定并入与样品模板链的核苷酸互补的生长链中的核苷酸来间接确定测定模板的序列。

[2055] 在一些实施例中,可以通过检测当核苷酸通过聚合酶并入与酶-聚合物复合物中与聚合酶相关的模板的链互补的链中时从核苷酸碱基释放的标记的核苷酸的标签,对DNA(例如,单链DNA)进行测序。使用标记核苷酸的通过合成的基于单分子纳米孔的测序(Nano-SBS)技术描述于例如WO2014/074027中,其通过引用以其整体并入。因此,在一些实施例中,可以附着到插入的纳米孔的酶-多核苷酸复合物可以是DNA聚合酶-DNA复合物。在一些实施例中,DNA聚合酶-DNA复合物可以附着于野生型或变体单体纳米孔。在一些实施例中,DNA聚合酶-DNA复合物可以附着于野生型、变体或修饰的变体均低聚纳米孔。在一些实施例中,DNA聚合酶-DNA复合物可以附着于野生型、变体或修饰的变体异源低聚纳米孔。在一些实施

例中,DNA聚合酶-DNA复合物可以附着于野生型、变体或修饰的变体αHL纳米孔。在其它实施例中,DNA聚合酶-DNA复合物可以附着于野生型OmpG纳米孔或其变体。

[2056] 在其它实施例中,酶-多核苷酸复合物可以是RNA聚合酶-RNA复合物。RNA聚合酶-RNA复合物可以附着于野生型或变体寡聚或单体纳米孔。在一些实施例中,RNA聚合酶-RNA复合物附着于野生型或变体OmpG纳米孔。在其它实施例中,RNA聚合酶-RNA复合物附着于野生型或变体αHL纳米孔。在其它实施例中,酶-多核苷酸复合物可以是逆转录酶-RNA复合物。逆转录酶-RNA复合物可以附着于野生型或变体寡聚或单体纳米孔。在一些实施例中,逆转录酶-RNA复合物附着于野生型或变体OmpG纳米孔。在其它实施例中,逆转录酶-RNA复合物附着于野生型或变体αHL纳米孔。在一些实施例中,单独的核酸可以通过在其由持续核酸外切酶释放时识别核苷5'-单磷酸酯来测序(Astier等人,2006,J Am Chem Soc128:1705-1710)。因此,在一些实施例中,可以附着到插入的纳米孔的酶-多核苷酸复合物可以是核酸外切酶-多核苷酸复合物。在一些实施例中,核酸外切酶-多核苷酸复合物可以附着于野生型或变体单体纳米孔。在一些实施例中,核酸外切酶-多核苷酸复合物可以附着于野生型或变体同源寡聚纳米孔。在一些实施例中,核酸外切酶-多核苷酸复合物可以附着于野生型或变体异源寡聚纳米孔。在一些实施例中,核酸外切酶-多核苷酸复合物可以附着于野生型αHL纳米孔或其变体。在其它实施例中,核酸外切酶-多核苷酸复合物可以附着于野生型OmpG纳米孔或其变体。

[2057] 在一些实施例中,非核酸聚合物也可以移动通过纳米孔并被测序。例如,蛋白质和多肽可以移动通过纳米孔,并且使用纳米孔的蛋白质或多核苷酸的测序可以通过控制蛋白质的解折叠和易位通过纳米孔来进行。受控的解折叠和随后的易位可以通过与待测序的蛋白质偶联的解折叠酶的作用实现(参见,例如,Nivala等人,2013,Nature Biotechnol 31:247-250)。在一些实施例中,附着在膜中的纳米孔上的酶-聚合物复合物可以是酶-多肽复合物,例如,解折叠酶-蛋白质复合物。在一些实施例中,解折叠酶-蛋白质复合物可以附着至野生型或变体单体纳米孔。在一些实施例中,解折叠酶-蛋白质复合物可以附着至野生型或变体同源寡聚纳米孔。在一些实施例中,解折叠酶-蛋白质复合物可以附着于野生型或变体异源寡聚纳米孔。在一些实施例中,解折叠酶-蛋白质复合物可以附着于野生型αHL纳米孔或其变体。在其他实施例中,解折叠酶-蛋白质复合物可以附着于野生型OmpG纳米孔或其变体。

[2058] 在一些实施例中,其它非核酸聚合物也可以例如通过移动通过纳米孔来测序。例如,WO 1996013606A1公开了糖材料(例如,包括硫酸乙酰肝素(HS)和肝素的多糖)的外切测序,并且US 8,846,363 B2公开了酶(例如,来自肝素黄杆菌(Flavobacterium heparinum)的硫酸酯酶),其可以应用(例如,串联)于多糖(例如,肝素衍生的寡糖)的外切测序。出于所有目的,将这两个专利文献的全部内容通过引用并入本文。

[2059] 酶-聚合物复合物和酶-纳米孔复合物的酶可以包括多核苷酸和多肽加工酶,例如DNA和RNA聚合酶,逆转录酶,核酸外切酶,解折叠酶,连接酶和硫酸酯酶。酶-聚合物复合物和酶-纳米孔复合物中的酶可以是野生型酶,或者可以是野生型酶的变体形式。变体酶可以被工程化以具有相对于亲本酶改变的特性。在一些实施例中,改变的酶-聚合物复合物的酶是聚合酶。聚合酶的改变的特性包括酶活性、保真度、持续能力、伸长率、稳定性或溶解度的改变。聚合酶可以被突变以降低聚合酶将核苷酸并入核酸链(例如,核酸生长链)的速率。通

过纳米孔蛋白质的位点特异性诱变和将DNA加工酶(例如,DNA聚合酶)并入纳米孔中的组合,可以实现降低的速度(和提高的灵敏度)。

[2060] 在一些情况下,核苷酸并入核酸链的速率可以通过将核苷酸和/或模板链官能化(例如通过模板核酸链的甲基化)以提供空间位阻而降低。在一些情况下,通过并入甲基化核苷酸降低速率。

[2061] 酶-聚合物复合物和酶-纳米孔复合物的酶可以被修饰以包含一个或多个附着组分和/或附着位点,附着组分和/或附着位点用于将酶-聚合物复合物连接至插入膜中的纳米孔。类似地,酶-纳米孔复合物的纳米孔和与酶-聚合物复合物连接的纳米孔也可以被修饰为包含一个或多个连接组分和/或连接位点以将纳米孔连接至酶-聚合物复合物。

[2062] 纳米孔测序复合体的纳米孔包括但不限于生物纳米孔、固态纳米孔和杂化生物-固态纳米孔。在一些实施例中,纳米孔测序复合物包括固态纳米孔,如石墨烯,(氧化铟锡)ITO,具有碳纳米管的孔线,或其他无机固态层,或前述任一种的组合。纳米孔测序复合物的生物纳米孔包括来自大肠杆菌、沙门氏菌、志贺氏菌和假单胞菌的OmpG,以及来自金黄色葡萄球菌的α溶血素,来自耻垢分枝杆菌(M.smegmatis sp.)的MspA。纳米孔可以是野生型纳米孔、变体纳米孔或修饰的变体纳米孔。变体纳米孔可以被工程化以具有相对于亲本酶改变的特性。参见,例如,2015年10月28日提交的题为"具有改变的特征的α-溶血素变体"的美国专利申请第14/924,861号。在一些实施例中,特性相对于野生型酶而改变。在一些实施例中,纳米孔测序复合体的变体纳米孔被工程化以降低衍生它的亲本纳米孔的离子电流噪声。具有改变的特性的变体纳米孔的一个示例是在收缩位点处具有一个或多个突变的OmpG纳米孔(WO2017050722 A1,标题为"OmpG变体",于2015年9月22日提交,其通过引用以其整体并入本文),其相对于亲本OmpG的离子噪声水平降低离子噪声水平。降低的离子电流噪声提供了这些OmpG纳米孔变体在多核苷酸和蛋白质的单分子感测中的用途。在其它实施例中,变体OmpG多肽可进一步突变以结合分子衔接子,分子衔接子在驻留于孔中的同时减缓分析物(例如,核苷酸碱基)通过孔的移动,并因此提高分析物识别的准确性。修饰的变体纳米孔大多是或通常是多聚体纳米孔,其亚基已经被工程化以影响亚基间相互作用(US 2017-0088890A1,标题为"Alpha-溶血素变体",其全文通过引用并入本文)。可以利用改变的亚基相互作用来指定单体寡聚化以形成脂质双层中的多聚体纳米孔的序列和次序。该技术提供了对形成纳米孔的亚基的化学计量的控制。多聚体纳米孔的示例是αHL纳米孔,其亚基可以被修饰以确定在寡聚化期间亚基的相互作用的序列。

[2063] 附着方式

[2064] 酶-聚合物复合物可以以任何合适的方式附着到纳米孔。可以使用SpyTag/SpyCatcher肽系统实现将酶-聚合物复合物附着到纳米孔上(Zakeri等人2012,PNAS109:E690-E697)天然化学连接(Thapa等人,2014,Molecules 19:14461-14483),分选酶系统(Wu以及Guo,2012,J Carbohydr Chem 31:48-66;Heck等人,2013,Appl Microbiol Biotechnol 97:461-475),转谷氨酰胺酶系统(Dennler等人,2014,Bioconjug Chem 25:569-578),甲酰甘氨酸连接(Rashidian等人,2013,Bioconjug Chem 24:1277-1294),或本领域已知的其它化学连接技术。

[2065] 在一些情况下,使用Solulink™化学将酶连接至纳米孔。参见例如WO 2017/125565 A1。Solulink™可以是(6-肼基-烟酸,芳族肼)与4FB(4-甲酰基苯甲酸,芳族醛)之间的反应。

在一些实例中,使用Click化学(可从例如LifeTechnologies获得;还参见例如WO 2014/190322 A2)。在一些情况下,将锌指突变引入溶血素分子中,然后使用分子(例如,DNA中间分子)将酶连接至溶血素上的锌指位点。

[2066] O.非核酸聚合物

[2067] 在任何前述实施例中(例如,上文部分A-N中描述的那些),编码标签和/或记录标签或其任何部分(例如,通用引物、间隔区、UMI、记录标签、条形码序列、结合循环特异性条形码等)可包含或被可测序聚合物(例如,非核酸可测序聚合物)替代。在一些方面,编码标签和/或记录标签,或其任何部分(例如,通用引物、间隔区、UMI、记录标签条形码、编码序列、结合循环特异性条形码等)例如通过基于纳米孔的测序给出清楚可分辨的信号。可以使用多种聚合物,并且任选地使用点击化学,可以使用与埃德曼型降解相容的标签读出靶蛋白质和/或多肽的序列信息。在一些实施例中,本文公开的方法包含扩增步骤。在其他实施例中,在没有扩增步骤的情况下,可以通过纳米孔多次处理单个分子以给出多个读取,以便给出可分辨的信号。

[2068] 在任何前述实施例中,记录标签包含待读取的聚合物,其中编码标签内的编码器包含或者是一个或多个小分子部分,可测序的非核酸聚合物的一个或多个亚基,可测序的非核酸聚合物的一个或多个单体。在该示例中,编码器部分类似于串上的珠粒(例如,延伸记录标签)。

[2069] 在一些实施例中,可测序聚合物是线性的。在其它实施例中,可测序聚合物包含一个或多个直链,一个或多个支链,一个或多个环或其任何组合。

[2070] 在一些实施例中,可测序聚合物是未交联的。在其它实施例中,可测序聚合物是交联的,例如弱交联的(例如,交联的聚合物包含约0.1%至约1%的交联剂)。

[2071] 在一些实施例中,可测序聚合物是中性的,例如,具有整体中性电荷。在具体的实施例中,可测序聚合物包括丙烯酰胺,丙烯酸酯,苯乙烯,尼龙,聚酰亚胺,NiPAM聚合物,PEG,聚内酰胺,聚内酯,环氧化物,聚硅酮,聚酯,聚碳二亚胺,聚氨酯,聚吡咯烷酮,聚异硫氰酸酯,或它们的任何组合。

[2072] 在其它实施例中,可测序聚合物是带电的。在具体的实施例中,可测序聚合物是带负电荷的,例如,可测序聚合物可包含聚磺酸酯,和/或带负电荷的丙烯酰胺(例如,与丙烯酸共聚的丙烯酰胺)。在其它实施例中,可测序聚合物是带正电荷的,例如,可测序聚合物可包含聚胺,聚苯胺,聚亚胺和/或具有可被取代或未被取代的胺官能团的丙烯酰胺。在其它实施例中,可测序聚合物可包含离聚物。

[2073] 在其它实施例中,可测序聚合物包括聚吗啉代和/或两性离子聚合物。

[2074] 在一些实施例中,可测序聚合物包含疏水性聚合物和/或亲水性聚合物。在一些方面,可测序聚合物包含疏水性聚合物,例如苯乙烯,聚酰亚胺(kapton),含氟聚合物或异戊二烯或其任何组合。在其它方面,可测序聚合物包含亲水性聚合物,例如聚乙二醇,聚赖氨酸,聚亚铵,聚亚胺,丙烯酰胺或丙烯酸酯,或它们的任何组合。

[2075] 在一些实施例中,可测序聚合物包含复分解产物,例如,通过开环或线性复分解形成的聚合物产物。

[2076] 在一些实施例中,可测序聚合物包含共聚物,嵌段共聚物,二嵌段共聚物,或三嵌段共聚物,或其任何组合。在一些实施例中,可测序聚合物通过聚合物产物的连续聚合或连

接形成。

[2077] 在其它实施例中,可测序聚合物包含生物聚合物。在具体的实施例中,可测序聚合物包含多糖,多肽,肽,或聚酰胺,或其任何组合。

[2078] 在一些实施例中,可测序聚合物包含天然存在的生物聚合物,生物聚合物模拟物,或合成的生物聚合物,或它们的杂合体。

[2079] 在一些实施例中,可测序聚合物包含生物聚合物模拟物,例如类肽。在一些实施例中,可测序聚合物包含星形聚合物或树状聚合物。

[2080] 在一些实施例中,可测序聚合物包括刺激触发的可降解聚合物。刺激的示例包括但不限于化学刺激,生物化学刺激,酶刺激,光,时间和温度。

[2081] 在一些实施例中,可测序聚合物包含笼状或受保护的聚合物。

[2082] 在一些实施例中,可测序聚合物包含可逆共价键合的聚合物。在一些实施例中,可测序聚合物包含二硫化物,硫(1)酯,模板衍生的硝基氧聚合物,或聚亚胺,或其任何组合。

[2083] 在一些实施例中,可测序聚合物包含具有低多分散指数(PDI)或高PDI的聚合物。多分散性指数(PDI),或不均匀性指数,或简单的分散性指数(Đ)是给定聚合物样品中分子质量分布的量度。典型的分散度基于聚合机理而变化,并可受到各种反应条件的影响。在合成聚合物中,其可因反应物比率、聚合完成的接近程度等而大幅变化。对于典型的加成聚合,Đ可以为约5至20。对于典型的逐步聚合,Đ最可能的值为约2或以下。活性聚合是加成聚合的特殊情况,其值非常接近1。这也是在生物聚合物中的情况,其中分散度可以非常接近或等于1,表明仅存在一个长度的聚合物。

[2084] 在一些实施例中,可测序聚合物包含具有低PDI的聚合物,例如活性聚合的产物。在其它实施例中,可测序聚合物包含具有较高PDI的聚合物,例如ATRP的产物或其它统计驱动分布。

[2085] 在一些实施例中,可测序聚合物包含能够充当发色团或导电材料的共轭聚合物。

[2086] 在其它实施例中,可测序聚合物包含结构化或非结构化(例如,折叠或非折叠)聚合物,例如折叠体。在一些实施例中,编码聚合物相对较短,例如,由2个亚基组成。在其它实施例中,聚合物是大的,例如由总计约100kDa的亚基组成。

[2087] 许多聚合物骨架适于包含可测序的聚合物。Ouahabi等人描述的聚磷酸盐骨架可用于纳米孔测序(Al Ouahabi,Charles等人,2015)。磷酸盐基团赋予骨架负电荷,使得能够通过孔进行电控制的易位。不同的R基团可用于提供不同的电流阻断标记。类肽聚合物提供了另一种有用的聚合物。类肽由甘氨酸骨架组成,其中侧链R基团连接到类肽骨架的氮上,而不是肽中的α-碳。类肽的固相合成提供了合成效率、高产率和R基团侧链多样性(可获得数百个不同的R基团),同时减少了合成时间和成本(Knight,Zhou等人,2015)。类肽骨架是中性的,但可以沿着聚合物的长度使用带电荷的R基团以促进电控制的易位。示例性带电R基团包括N-(2-氨基乙基)甘氨酸(Nae)和N-(2-羧乙基)甘氨酸(Nce)。对于类肽的综述,参见例如,Seto等人,2011,"类肽:"合成、表征和纳米结构(Peptoids:Synthesis,Characterization,and Nanostructures)",Comprehensive Biomaterials,vol.2,pp.53-76,其通过引用并入本文。

[2088] 用于当前阻断的R基团可以选自官能部分的长列表(US20160075734,以引用方式并入本文中)。官能部分包括取代的和/或官能化的基团,例如烷氧基,卤素,巯基,叠氮基,

氰基,甲酰基,羧基,羟基,硝基,酰基,芳氧基,烷硫基,氨基,烷基氨基,芳基烷基氨基,取代的氨基,酰氨基,酰氧基,酯,硫酯,羧基硫酯,醚,酰胺,脒基,硫酸酯,磺基,磺酰基,磺酰基,磺酸,磺酰胺,尿素,烷氧基酰氨基,氨基酰氧基,胍基,醛,酮,亚胺,腈,磷酸酯,硫醇,环氧化物,过氧化物,硫氰酸盐,脒,肟,腈,重氮基,杂环基等,这些术语包括这些基团的组合。如本文单独或作为另一基团的一部分使用的"杂环基团"或"杂环基"是指脂族(例如,完全或部分饱和的杂环基)或芳族(例如,杂芳基)单环或二环环系。单环环系示例为含有1、2、3或4个独立地选自氧、氮和硫的杂原子的任何5或6元环。5元环具有0-2个双键,并且6元环具有0-3个双键。单环环系的代表性示例包括但不限于氮杂环丁烷,吖庚因,氮杂环丙烷,二氮杂卓,1,3-二氧戊环,二恶烷,二噻烷,呋喃,咪唑,咪唑啉,咪唑烷,异噻唑,异噻唑啉,异噻唑烷,异恶唑,异恶唑啉,异恶唑烷,吗啉,恶二唑,恶二唑啉,恶二唑烷,恶唑,恶唑啉,恶唑烷,哌嗪,哌啶,吡喃,吡嗪,吡唑,吡唑啉,吡唑烷,吡啶,嘧啶,哒嗪,吡咯,吡咯啉,吡咯烷,四氢呋喃,四氢噻吩,四嗪,四唑,噻二唑,噻二唑啉,噻二唑烷,噻唑,噻唑啉,噻唑烷,噻吩,硫代吗啉,硫代吗啉砜,噻喃,三嗪,三唑,三噻烷等。二环环系的示例为与如本文定义的芳基、如本文定义的环烷基或如本文定义的另一单环环系融合的任何上述单环环系。双环环系的代表性示例包括但不限于例如苯并咪唑,苯并噻唑,苯并噻二唑,苯并噻吩,苯并恶二唑,苯并恶唑,苯并呋喃,苯并吡喃,苯并噻喃,苯并二恶英(benzodioxine),1,3-苯并二氧杂环戊烯,噌啉,吲唑,吲哚,二氢吲哚,吲嗪,萘啶,异苯并呋喃,异苯并噻吩,异吲哚,异二氢吲哚,异喹啉,酞嗪,嘌呤,吡喃并吡啶,喹啉,喹嗪,喹喔啉,喹唑啉,四氢异喹啉,四氢喹啉,噻喃并吡啶等。这些环包括其季铵化衍生物,并且可以任选地被选自以下的基团取代:卤素,烷基,卤代烷基,烯基,炔基,环烷基,环烷基烷基,芳基,芳基烷基,杂环基,杂环烷基,羟基,烷氧基,烯氧基,炔氧基,卤代烷氧基,环烷氧基,环烷基烷氧基,芳氧基,芳基烷氧基,杂环氧基,杂环烷基烷氧基,巯基,烷基-S(O)m,卤代烷基-S(O)m,烯基-S(O)m,炔基-S(O)m,环烷基-S(O)m,环烷基烷基-S(O)m,芳基-S(O)m,芳基烷基-S(O)m,杂环基-S(O)m,杂环烷基-S(O)m,氨基,烷基氨基,烯基氨基,炔基氨基,卤代烷基氨基,环烷基氨基,环烷基烷基氨基,芳基氨基,芳基烷基氨基,杂环氨基,杂环烷基氨基,双取代的氨基,酰基氨基,酰氧基,酯,酰胺,磺酰胺,尿素,烷氧基酰基氨基,氨基酰氧基,硝基或氰基,其中m=0,1,2或3。

[2089] 在任何前述实施例中,记录标签和/或编码标签可以是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[2090] 在任何前述实施例中,记录标签和/或编码标签可以是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。

[2091] 在任何前述实施例中,记录标签和/或编码标签可包含可测序聚合物(例如多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[2092] 在任何前述实施例中,编码标签可以是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元,而记录标签是可测序聚合物。在一些实施例中,可以将编码标签(小分子、亚基或单体等)添加到记录标签中,以形成可测序聚合物,例如串(例如,延伸记录标签)上的珠粒(例如,各种编码标签)。

[2093] 在任何前述实施例中,延伸记录标签和/或延伸编码标签可以是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。

[2094] IV.示例性实施例

[2095] 以下实施例提供本公开的附加说明。

[2096] 实施例1.一种试剂盒,其包含:(a)记录标签,其被配置为与分析物直接或间接相关;(b)(i)编码标签,其包含关于能够结合分析物的结合基团的识别信息,并且被配置为与结合基团直接或间接相关以形成结合剂,和/或(ii)标签,其中记录标签和编码标签被配置为允许在结合剂和分析物之间结合时在它们之间转移信息;和任选地,(c)结合基团。

[2097] 实施例2.根据实施例1所述的试剂盒,其中记录标签和/或分析物被配置为直接或间接固定到支持物上。

[2098] 实施例3.根据实施例2所述的试剂盒,其中记录标签被配置为固定到支持物上,从而固定与记录标签相关的分析物。

[2099] 实施例4.根据实施例2所述的试剂盒,其中分析物被配置为固定到支持物上,从而固定与分析物相关的记录标签。

[2100] 实施例5.根据实施例2所述的试剂盒,其中记录标签和分析物中的每一个被配置为固定到支持物上。

[2101] 实施例6.根据实施例5所述的试剂盒,其中记录标签和分析物被设计成当两者都固定到支持物上时共同定位。

[2102] 实施例7.根据实施例1-6中任一个所述的试剂盒,其还包含固定接头,固定接头被配置为:(i)直接或间接固定到支持物上,和(ii)直接或间接与记录标签和/或分析物相关。

[2103] 实施例8.根据实施例7所述的试剂盒,其中固定接头被配置为与记录标签和分析物相关。

[2104] 实施例9.根据实施例7或8所述的试剂盒,其中固定接头被配置为直接固定至支持物,从而固定与固定接头相关的记录标签和/或分析物。

[2105] 实施例10.根据实施例2-9中任一个所述的试剂盒,进一步包含支持物。

[2106] 实施例11.根据实施例1-10中任一个所述的试剂盒,其进一步包含一种或多种试剂,用于在结合剂和分析物之间结合时在编码标签与记录标签之间转移信息。

[2107] 实施例12.根据实施例11所述的试剂盒,其中一种或多种试剂被配置为将信息从编码标签转移到记录标签,从而产生延伸记录标签。

[2108] 实施例13.根据实施例11所述的试剂盒,其中一种或多种试剂被配置为将信息从记录标签转移到编码标签,从而产生延伸编码标签。

[2109] 实施例14.根据实施例11所述的试剂盒,其中一种或多种试剂被配置为产生包含来自编码标签的信息和来自记录标签的信息的双标签构建体。

[2110] 实施例15.根据实施例1-14中任一个所述的试剂盒,其包含记录标签中的至少两个。

[2111] 实施例16.根据实施例1-15中任一个所述的试剂盒,其包含编码标签中的至少两个,每个编码标签包含关于其相关结合基团的识别信息。

[2112] 实施例17.根据实施例1-16中任一个所述的试剂盒,其包含结合剂中的至少两种。

[2113] 实施例18.根据实施例17所述的试剂盒,其包含:(i)一种或多种试剂,其在第一结合剂和分析物之间结合时用于将信息从第一结合剂的第一编码标签转移到记录标签,以产生第一次序延伸记录标签,和/或(ii)一种或多种试剂,其在第二结合剂和分析物之间结合时用于将信息从第二结合剂的编码标签转移至第一延伸记录标签,以产生第二次序延伸记

录标签,其中(i)的一种或多种试剂和(ii)的一种或多种试剂可以相同或不同。

[2114] 实施例19.根据实施例18所述的试剂盒,其进一步包含:(iii)一种或多种试剂,其用于在第三(或更高次序)结合剂与分析物之间结合时,将信息从第三(或更高次序)结合剂的第三(或更高次序)编码标签转移至第二次序延伸记录标签,以产生第三(或更高次序)次序延伸记录标签。

[2115] 实施例20.根据实施例17所述的试剂盒,其包含:(i)一种或多种试剂,其用于在第一结合剂和分析物之间结合时,将信息从第一结合剂的第一编码标签转移到第一记录标签,以产生第一延伸记录标签,和/或(ii)一种或多种试剂,其用于在第二结合剂和多肽之间结合时,将信息从第二结合剂的第二编码标签转移到第二记录标签结合,以产生第二延伸记录标签,其中(i)的一种或多种试剂和(ii)的一种或多种试剂可以相同或不同。

[2116] 实施例21.根据实施例20所述的试剂盒,其进一步包含:(iii)一种或多种试剂,其用于在第三(或更高次序)结合剂和分析物之间结合时,将信息从第三(或更高次序)结合剂的第三(或更高次序)编码标签转移到第三(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。

[2117] 实施例22.根据实施例20或21所述的试剂盒,其中第一记录标签、第二记录标签和/或第三(或更高次序)记录标签被配置为与分析物直接或间接相关。

[2118] 实施例23.根据实施例20-22中任一个所述的试剂盒,其中第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为固定在支持物上。

[2119] 实施例24.根据实施例20-23中任一个所述的试剂盒,其中第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为与分析物共定位,例如,以在第一、第二或第三(或更高次序)结合剂和分析物之间结合时允许分别在第一、第二或第三(或更高次序)编码标签与第一、第二或第三(或更高次序)记录标签之间转移信息。

[2120] 实施例25.根据实施例20-24中任一个所述的试剂盒,其中第一编码标签、第二编码标签和/或第三(或更高次序)编码标签中的每一个包含结合循环特异性条形码,如结合循环特异性间隔区序列$C_n$和/或编码标签特异性间隔区序列$C_n$,其中n是整数,并且$C_n$表示第n结合剂与多肽之间的结合;或者其中结合循环标签$C_n$是外源添加的,例如,结合循环标签$C_n$对于编码标签可以是外源的。

[2121] 实施例26.根据实施例1-25所述的试剂盒,其中编码标签包含多肽。

[2122] 实施例27.根据实施例26所述的试剂盒,结合基团能够结合多肽的一个或多个N-末端或C-末端氨基酸,或能够结合通过官能化试剂修饰的一个或多个N-末端或C-末端氨基酸。

[2123] 实施例28.根据实施例26或27所述的试剂盒,其进一步包含官能化试剂。

[2124] 实施例29.根据实施例26-28中任一个所述的试剂盒,其进一步包含包含用于移除(例如,通过化学裂解或酶裂解)多肽的一个或多个N-末端、内部或C-末端氨基酸,或移除官能化的N-末端、内部或C-末端氨基酸的消除试剂,任选地,其中消除试剂包含羧肽酶或氨肽酶或其变体,突变体或修饰的蛋白质;水解酶或其变体,突变体或其修饰的蛋白质;温和型埃德曼降解试剂;Edmanase酶;无水TFA,碱;或其任何组合。

[2125] 实施例30.根据实施例26-29中任一个所述的试剂盒,其中一个或多个N-末端、内部或C-末端氨基酸包含:(i)N-末端氨基酸(NTAA);(ii)N-末端二肽序列;(iii)N-末端三肽

序列;(iv)内部氨基酸;(v)内部二肽序列;(vi)内部三肽序列;(vii)C-末端氨基酸;(viii)C-末端二肽序列;或(ix)C-末端三肽序列,或其任何组合,任选地其中(i)-(ix)中的任何一个或多个氨基酸残基被修饰或官能化。

[2126] 实施例31.一种试剂盒,包含:至少(a)第一结合剂,其包含(i)能够结合待分析多肽的N-末端氨基酸(NTAA)或官能化NTAA的第一结合基团,和(ii)包含关于第一结合基团的识别信息的第一编码标签,任选地(b)记录标签,其被配置为与多肽直接或间接相关,并且进一步任选地(c)官能化试剂,其能够修饰多肽的第一NTAA,以产生第一官能化NTAA,其中记录标签和第一结合剂被配置为允许在第一结合剂和多肽之间结合时在第一编码标签和记录标签之间转移信息。

[2127] 实施例32.根据实施例31所述的试剂盒,其还包含一种或多种试剂,用于将信息从第一编码标签转移到记录标签,从而产生第一次序延伸记录标签。

[2128] 实施例33.根据实施例31或32所述的试剂盒,其中官能化试剂包含化学剂、酶和/或生物剂,如异硫氰酸酯衍生物,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。

[2129] 实施例34.根据实施例31-33所述的试剂盒,进一步包含消除试剂,用于移除(例如,通过化学裂解或酶裂解)第一官能化NTAA,以暴露紧邻的氨基酸残基(如第二NTAA)。

[2130] 实施例35.根据实施例34所述的试剂盒,其中第二NTAA能够通过相同或不同的官能化试剂官能化以产生第二官能化NTAA,其可与第一官能化NTAA相同或不同。

[2131] 实施例36.根据实施例35所述的试剂盒,其进一步包含:(d)第二(或更高次序)结合剂,其包含(i)能够结合到第二官能化NTAA的第二(或更高次序)结合基团,和(ii)第二(或更高次序)编码标签,其包含关于第二(或更高次序)结合基团的识别信息,其中第一编码标签和第二(或更高次序)编码标签可以相同或不同。

[2132] 实施例37.根据实施例36所述的试剂盒,其中第一官能化NTAA和第二官能化NTAA彼此独立地选自由以下组成的群组:官能化N-末端丙氨酸(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),缬氨酸(V或Val),色氨酸(W或Trp),酪氨酸(Y或Tyr),其任何组合。

[2133] 实施例38.根据实施例36或37所述的试剂盒,其进一步包含一种或多种试剂,用于将信息从第二(或更高次序)编码标签转移到第一次序延伸记录标签,从而产生第二(或更高次序)延伸记录标签。

[2134] 实施例39.一种试剂盒,其包含:至少(a)一种或多种结合剂,每种结合剂包含(i)能够结合待分析多肽的N-末端氨基酸(NTAA)或官能化NTAA的结合基团,和(ii)包含关于结合基团的识别信息的编码标签,和/或(b)一个或多个记录标签,其被配置为与多肽直接或间接相关,其中一个或多个记录标签和一个或多个结合剂被配置为允许在每个结合剂与多肽之间结合时在编码标签和记录标签之间转移信息,以及任选地(c)官能化试剂,其能够修饰多肽的第一NTAA以产生第一官能化NTAA。

[2135] 实施例40.根据实施例39所述的试剂盒,其进一步包含消除试剂,用于移除(例如,

通过化学裂解或酶裂解)第一官能化NTAA,以暴露紧邻的氨基酸残基(如第二NTAA)。

[2136] 实施例41.根据实施例40所述的试剂盒,其中第二NTAA能够通过相同或不同的官能化试剂官能化以产生第二官能化NTAA,其可与第一官能化NTAA相同或不同。

[2137] 实施例42.根据实施例41所述的试剂盒,其中第一官能化NTAA和第二官能化NTAA彼此独立地选自由以下组成的群组:官能化N-末端丙氨酸(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),缬氨酸(V或Val),色氨酸(W或Trp),酪氨酸(Y或Tyr),其任何组合。

[2138] 实施例43.根据实施例39-42中任一个所述的试剂盒,其包含:(i)一种或多种试剂,其用于在第一结合剂和多肽之间结合时,将信息从第一结合剂的第一编码标签转移到第一记录标签,以产生第一延伸记录标签,和/或(ii)一种或多种试剂,其用于在第二结合剂和多肽之间结合时,将信息从第二结合剂的第二编码标签转移到第二记录标签结合,以产生第二延伸记录标签,其中(i)的一种或多种试剂和(ii)的一种或多种试剂可以相同或不同。

[2139] 实施例44.根据实施例43所述的试剂盒,其进一步包含:(iii)一种或多种试剂,其用于在第三(或更高次序)结合剂和多肽之间结合时,将信息从第三(或更高次序)结合剂的第三(或更高次序)编码标签转移到第三(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。

[2140] 实施例45.根据实施例43或44所述的试剂盒,其中第一记录标签、第二记录标签和/或第三(或更高次序)记录标签被配置为与分析物直接或间接相关。

[2141] 实施例46.根据实施例43-45中任一个所述的试剂盒,其中第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为固定在支持物上。

[2142] 实施例47.根据实施例43-46中任一个所述的试剂盒,其中第一记录标签、第二记录标签和/或第三(或更高次序)记录标签可以被配置为与多肽共定位,例如,以在第一、第二或第三(或更高次序)结合剂和多肽之间结合时允许分别在第一、第二或第三(或更高次序)编码标签与第一、第二或第三(或更高次序)记录标签之间转移信息。

[2143] 实施例48.根据实施例43-47中任一个所述的试剂盒,其中第一编码标签、第二编码标签和/或第三(或更高次序)编码标签中的每一个包含结合循环特异性条形码,例如结合循环特异性间隔区序列$C_n$和/或编码标签特异性间隔区序列$C_n$,其中n是整数,并且$C_n$表示第n结合剂和多肽之间的结合。或者,可以外源添加结合循环标签$C_n$,例如,结合循环标签$C_n$对于编码标签可以是外源的。

[2144] 实施例49.根据实施例1-48中任一个所述的试剂盒,其中分析物或多肽包含蛋白质或多肽链或其片段,脂质,碳水化合物或大环。

[2145] 实施例50.根据实施例1-49中任一个所述的试剂盒,其中分析物或多肽包含大分子或其复合物,如蛋白质复合物或其亚基。

[2146] 实施例51.根据实施例1-50中任一个所述的试剂盒,其中记录标签包含核酸、寡核苷酸、修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、具有受保护碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代,或其组合。

[2147] 实施例52.根据实施例1-51中任一个所述的试剂盒,其中记录标签包含通用引发位点。

[2148] 实施例53.根据实施例1-52中任一个所述的试剂盒,其中记录标签包含用于扩增、测序或两者的引发位点,例如,通用引发位点包含用于扩增、测序或两者的引发位点。

[2149] 实施例54.根据实施例1-53中任一个所述的试剂盒,其中记录标签包含独特分子标识符(UMI)。

[2150] 实施例55.根据权利要求1-54任一个所述的方法,其中记录标签包含条形码。

[2151] 实施例56.根据实施例1-55中任一个所述的试剂盒,其中记录标签在其3'-末端包含间隔区。

[2152] 实施例57.根据实施例1-56中任一个所述的试剂盒,其包含固体支持物,例如刚性固体支持物、柔性固体支持物或软固体支持物,并且包括多孔支持物或非多孔支持物。

[2153] 实施例58.根据实施例1-57中任一个所述的试剂盒,其中支持物包含珠粒、多孔珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[2154] 实施例59.根据实施例58所述的方法,其中支持物包含聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[2155] 实施例60.根据实施例1-59中任一个所述的试剂盒,其包含支持物并且用于在顺序反应、平行反应或顺序和平行反应的组合中分析多种分析物或多肽。

[2156] 实施例61.根据实施例60所述的试剂盒,其中分析物在支持物上以等于或大于约10nm,等于或大于约15nm,等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,或等于或大于约500nm的平均距离间隔。

[2157] 实施例62.根据实施例1-61中任一个所述的试剂盒,其中结合基团包含多肽或其片段,蛋白质或多肽链或其片段,或蛋白质复合物或其亚基,如抗体或其抗原结合片段。

[2158] 实施例63.根据实施例1-62中任一个所述的试剂盒,其中结合基团可以包含羧肽酶或氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰蛋白;抗生物素蛋白或其变体,突变体或修饰蛋白;ClpS或其变体,突变体或修饰的蛋白质;UBR盒蛋白质或其变体,突变体或修饰蛋白;结合氨基酸的经修饰的小分子,即万古霉素或其变体,突变体或经修饰的分子;或其任何组合,或其中在每个结合剂中,结合基团包含小分子,编码标签包含识别小分子的多核苷酸,由此多个结合剂形成编码的小分子库,例如DNA编码的小分子库。

[2159] 实施例64.根据实施例1-63任一所述的试剂盒,结合剂能够选择性和/或结合分析物或多肽。

[2160] 实施例65.根据实施例1-64中任一个所述的试剂盒,其中记录标签包含核酸,寡核

苷酸,修饰的寡核苷酸,DNA分子,具有伪互补碱基的DNA,具有受保护碱基的DNA,RNA分子,BNA分子,XNA分子,LNA分子,PNA分子,γPNA分子或吗啉代DNA,或其组合。

[2161] 实施例66.根据实施例1-65中任一个所述的试剂盒,其中编码聚合物包含条形码序列,例如编码器序列,例如识别结合基团的编码器序列。

[2162] 实施例67.根据实施例1-66中任一个所述的试剂盒,其中编码标签包含间隔区,结合循环特异性序列,独特分子识别符(UMI标签),通用引发位点或其任何组合,任选地其中在每个结合循环后将结合循环特异性序列添加至记录标签。

[2163] 实施例68.根据实施例1-67中任一个所述的试剂盒,其中结合基团和编码标签通过接头或结合对连接。

[2164] 实施例69.根据实施例1-68中任一个所述的试剂盒,其中结合基团和编码标签通过SpyTag/SpyCatcher,SpyTag-KTag/SpyLigase(其中有待连接的两个部分具有SpyTag/KTag对,并且SpyLigase将SpyTag连接至KTag,由此将这两个部分连接),分选酶,SnoopTag/SnoopCatcher肽-蛋白质对,或HaloTag/HaloTag配体对,或其任何组合。

[2165] 实施例70.根据实施例1-69中任一个所述的试剂盒,其进一步包含用于在模板或非模板反应中在编码标签和记录标签之间转移信息的试剂,任选地其中试剂是(i)化学连接试剂或生物连接试剂,例如,连接酶,例如用于连接单链核酸或双链核酸的DNA连接酶或RNA连接酶,或(ii)用于单链核酸或双链核酸引物延伸的试剂,任选地,其中试剂盒还包含连接试剂,连接试剂包含至少两种连接酶或其变体(例如,至少两种DNA连接酶,或至少两种RNA连接酶,或至少一种DNA连接酶和至少一种RNA连接酶),其中至少两种连接酶或其变体包含腺苷酸化连接酶和组成型非腺苷酸化连接酶,或任选地其中试剂盒还包含连接试剂,连接试剂包含DNA或RNA连接酶和DNA/RNA脱腺苷酸化酶。

[2166] 实施例71.根据实施例1-70中任一个所述的试剂盒,其进一步包含聚合酶,例如DNA聚合酶或RNA聚合酶或逆转录酶,用于在编码标签和记录标签之间转移信息。

[2167] 实施例72.根据实施例1-71中任一个所述的试剂盒,其进一步包含一种或多种用于核酸序列分析的试剂。

[2168] 实施例73.根据实施例72所述的试剂盒,核酸序列分析包含通过合成测序、通过连接测序、通过杂交测序、polony测序、离子半导体测序、焦磷酸测序、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,或其任何组合。

[2169] 实施例74.根据实施例1-73中任一个所述的试剂盒,其进一步包含一种或多种用于核酸扩增的试剂,例如用于扩增一个或多个延伸记录标签,任选地其中核酸扩增包含指数扩增反应(例如,聚合酶链反应(PCR),例如乳液PCR以减少或消除模板转换)和/或线性扩增反应(例如,通过体外转录的等温扩增,或等温嵌合引物引发的核酸扩增(ICAN))。

[2170] 实施例75.根据实施例1-74中任一个所述的试剂盒,其包含用于将编码标签信息转移至记录标签以形成延伸记录标签的一种或多种试剂,其中延伸记录标签上的编码标签信息的次序和/或频率指示结合剂结合至分析物或多肽的次序和/或频率。

[2171] 实施例76.根据实施例1-75中任一个所述的试剂盒,其还包含一种或多种用于靶标富集,例如富集一种或多种延伸记录标签的试剂。

[2172] 实施例77.根据实施例1-76中任一个所述的试剂盒,还包含一种或多种用于扣除,例如扣除一个或多个延伸记录标签的试剂。

[2173] 实施例78.根据实施例1-77中任一个所述的试剂盒,进一步包含一种或多种用于归一化的试剂,例如,用以减少高丰度种类,如一种或多种分析物或多肽。

[2174] 实施例79.根据实施例1-78中任一个所述的试剂盒,其中至少一种结合剂结合至末端氨基酸残基、末端二氨基酸残基或末端三氨基酸残基。

[2175] 实施例80.根据实施例1-79中任一个所述的试剂盒,其中至少一种结合剂结合到翻译后修饰的氨基酸。

[2176] 实施例81.根据实施例1-80中任一个所述的试剂盒,其进一步包含用于将样品中的多种分析物或多肽分区到多个隔室中的一种或多种试剂或装置,其中每个隔室包含多个隔室标签,多个隔室标签任选地连接到支持物(例如,固体支持物),其中多个隔室标签在单个隔室内是相同的并且不同于其它隔室的隔室标签。

[2177] 实施例82.根据实施例81所述的试剂盒,其进一步包含用于将多种分析物或多肽(例如,多个蛋白质复合物、蛋白质和/或多肽)片段化成多种多肽片段的一种或多种试剂或工具。

[2178] 实施例83.根据实施例81或82所述的试剂盒,其进一步包含一种或多种用于将多个多肽片段与多个隔室中的每一个内的隔室标签退火或连接的试剂或工具,从而产生多个隔室标记的多肽片段。

[2179] 实施例84.根据实施例81-83中任一个所述的试剂盒,其中多个隔室包含微流体液滴,微孔,或表面上的分离区域,或其任何组合。

[2180] 实施例85.根据实施例81-84中任一个所述的试剂盒,其中每一个隔室平均包含单独的细胞。

[2181] 实施例86.根据实施例81-85中任一个所述的试剂盒,进一步包含用于标记样品中的多种分析物或多肽的一个或多个通用DNA标签。

[2182] 实施例87.根据实施例81-86中任一个所述的试剂盒,其进一步包含一种或多种试剂,用于用一种或多种通用DNA标签标记样品中的多种分析物或多肽。

[2183] 实施例88.根据实施例81-87中任一个所述的试剂盒,其进一步包含一种或多种用于引物延伸或连接的试剂。

[2184] 实施例89.根据实施例81-88中任一个所述的试剂盒,其中珠粒是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒。

[2185] 实施例90.根据实施例81-89中任一个所述的试剂盒,其中隔室标签包含单链或双链核酸分子。

[2186] 实施例91.根据实施例81-90中任一个所述的试剂盒,其中隔室标签包含条形码和可选的UMI。

[2187] 实施例92.根据实施例81-91中任一个所述的试剂盒,其中支持物是珠粒并且隔室标签包含条形码。

[2188] 实施例93.根据实施例81-92中任一个所述的试剂盒,其述支持物包含珠粒,并且其中包含连接到其上的多个隔室标签的珠粒通过均分与合并合成、单独合成或固定化形成。

[2189] 实施例94.根据实施例81-93中任一个所述的试剂盒,其进一步包含一种或多种用

于均分与合并合成、个别合成或固定化的试剂。

[2190] 实施例95.根据实施例81-94中任一个所述的试剂盒,其中隔室标签是记录标签中的成分,记录标签任选地还包含间隔区、条形码序列、独特分子标识符、通用引发位点或它们的任何组合。

[2191] 实施例96.根据实施例81-95中任一个所述的试剂盒,其中隔室标签还包含能够与多个分析物或多肽(例如,蛋白质复合物、蛋白质或多肽)上的内部氨基酸,肽骨架或N-末端氨基酸反应的官能部分。

[2192] 实施例97.根据实施例96所述的试剂盒,官能部分可以包含醛、叠氮化物/炔、马来酰亚胺/硫醇、环氧/亲核试剂、逆电子需求Diels-Alder(iEDDA)基团、点击试剂或其任何组合。

[2193] 实施例98.根据实施例81-97中任一个所述的试剂盒,其中隔室标签进一步包含肽,诸如蛋白连接酶识别序列,任选地其中蛋白连接酶是butelase I或其同源物。

[2194] 实施例99.根据实施例81-98中任一个所述的试剂盒,其进一步包含化学或生物试剂,例如酶,例如蛋白酶(例如,多金属蛋白酶),用于使多种分析物或多肽片段化。

[2195] 实施例100.根据实施例81-99中任一个所述的试剂盒,进一步包含用于从支持物释放隔室标签的一种或多种试剂。

[2196] 实施例101.根据实施例1-100中任一个所述的试剂盒,其进一步包含一种或多种用于形成延伸编码聚合物或双标签构建体的试剂。

[2197] 实施例102.根据实施例101所述的试剂盒,其中记录标签的3’末端被封闭以防止记录标签通过聚合酶延伸。

[2198] 实施例103.根据实施例101或102所述的试剂盒,其中编码标签包含编码器序列、UMI、通用引发位点、在其3’末端的间隔区、结合循环特异性序列,或其任何组合。

[2199] 实施例104.根据实施例101-103中任一个所述的试剂盒,其中双标签构建体通过间隙填充、引物延伸或它们二者产生。

[2200] 实施例105.根据实施例101-104中任一个所述的试剂盒,其中双标签分子包含衍生自记录标签的通用引发位点、衍生自记录标签的隔室标签,衍生自记录标签的隔室标签、衍生自记录标签的独特分子标识符、衍生自记录标签的可选间隔区、衍生自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区、和衍生自编码标签的通用引发位点。

[2201] 实施例106.根据实施例101-105中任一个所述的试剂盒,其中结合剂是多肽或蛋白质。

[2202] 实施例107.根据实施例101-106中任一个所述的试剂盒,其中结合剂包含氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰蛋白;抗生物素蛋白或其变体,突变体或修饰蛋白;ClpS或其变体,突变体或修饰的蛋白质;或结合氨基酸的修饰的小分子,即万古霉素或其变体,突变体或经修饰的分子;或抗体或其结合片段;或其任何组合。

[2203] 实施例108.根据实施例101-107中任一个所述的试剂盒,其中结合剂结合到单个氨基酸残基(例如,N-末端氨基酸残基,C-末端氨基酸残基,或内部氨基酸残基),二肽(例如,N-末端二肽,C-末端二肽,或内部二肽),三肽(例如,N-末端三肽,C-末端三肽,或内部三

肽),或分析物或多肽的翻译后修饰。

[2204] 实施例109.根据实施例101-107中任一个所述的试剂盒,其中结合剂结合N-末端多肽、C-末端多肽或内部多肽。

[2205] 实施例110.根据实施例1-109中任一个所述的试剂盒,其中编码标签和/或记录标签包含一个或多个纠错码、一个或多个编码器序列、一个或多个条形码、一个或多个UMI、一个或多个隔室标签、一个或多个循环特异性序列,或其任何组合。

[2206] 实施例111.根据实施例110所述的试剂盒,其中纠错码选自汉明码、李距离码、非对称李距离码、里德-所罗门码和Levenshtein-Tenengolts码。

[2207] 实施例112.根据实施例1-111中任一个所述的试剂盒,其中编码标签和/或记录标签包含循环标签。

[2208] 实施例113.根据实施例1-112中任一个所述的试剂盒,其进一步包含独立于编码标签和/或记录标签的循环标签。

[2209] 实施例114.根据实施例1-113中任一个所述的试剂盒,其包含:(a)用于产生细胞裂解物或蛋白质样品的试剂;(b)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;(c)蛋白酶,例如胰蛋白酶、LysN或LysC;(d)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;(e)用于基于降解的多肽测序的试剂;和/或(f)用于核酸测序的试剂。

[2210] 实施例115.根据实施例1-113中任一个所述的试剂盒,其包含:(a)用于产生细胞裂解物或蛋白质样品的试剂;(b)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;(c)蛋白酶,例如胰蛋白酶、LysN或LysC;(d)用于将多肽(例如蛋白质)固定到包含固定的记录标签的支持物上的试剂;(e)用于基于降解的多肽测序的试剂;和/或(f)用于核酸测序的试剂。

[2211] 实施例116.根据实施例1-113中任一个所述的试剂盒,其包含:(a)用于产生细胞裂解物或蛋白质样品的试剂;(b)变性试剂;(c)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;(d)通用DNA引物序列;(e)用通用DNA引物序列标记多肽的试剂;(f)用于通过引物对标记的多肽进行退火的条形码化的珠;(g)用于聚合酶延伸的试剂,用于将条形码从珠粒写到标记的多肽;(h)蛋白酶,例如胰蛋白酶、LysN或LysC;(i)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;(j)用于基于降解的多肽测序的试剂;和/或(k)用于核酸测序的试剂。

[2212] 实施例117.根据实施例1-113中任一个所述的试剂盒,其包含:(a)交联剂;(b)用于产生细胞裂解物或蛋白质样品的试剂;(c)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;(d)通用DNA引物序列;(e)用通用DNA引物序列标记多肽的试剂;(f)用于通过引物对标记的多肽进行退火的条形码化的珠粒;(g)用于聚合酶延伸的试剂,用于将条形码从珠粒写到标记的多肽;(h)蛋白酶,例如胰蛋白酶、LysN或LysC;(i)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;(j)用于基于降解的多肽测序的试剂;和/或(k)用于核酸测序的试剂。

[2213] 实施例118.根据实施例1-117中任一个所述的试剂盒,其中一种或多种组分在溶液中或在支持物例如固体支持物上提供。

[2214] V.示例

[2215] 包括以下示例仅用于说明的目的,并不意图限制本发明的范围。

[2216] 示例1:用蛋白酶K消化蛋白质样品

[2217] 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库。胰蛋白酶例如裂解如C-末端赖氨酸和精氨酸等带正电荷的氨基酸,而蛋白酶K在蛋白质上非选择性裂解。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(约30个氨基酸),但不过度消化样品。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 CaCl2/0.5%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。孵育后,加入PMSF至终浓度为5mM以抑制进一步消化。

[2218] 蛋白酶K的比活性可以通过将"化学底物"苯甲酰精氨酸-对硝基苯胺与蛋白酶K一起孵育并测量在约410nm吸收的黄色对硝基苯胺产物的发展来测量。酶活性以单位测量,其中一个单位等于每分钟产生的1μM对硝基苯胺,并且以酶活性/毫克总蛋白质为单位测量比活性。然后通过将酶活性除以溶液中蛋白质的总量来计算比活性。

[2219] 示例2:在珠粒蛋白酶消化和标记上使用SP3的样品制备

[2220] 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,蛋白质混合物(和珠粒)在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来减少二硫键,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,并在黑暗中室温下孵育20分钟。将反应物在50mM硼酸盐缓冲液中稀释2倍,并以最终蛋白酶:蛋白质比例为1:50(w/w)加入Glu-C或Lys-C。在37℃o/n(约16小时)孵育样品以完成消化。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%乙腈使肽与珠粒结合,并在8分钟内用乙腈洗涤孵育。洗涤后,通过5分钟的移液管混合步骤将肽从10μl的2% DMSO中的珠粒上洗脱下来。

[2221] 示例3:记录标签与肽的偶联

[2222] DNA记录标签以几种方式与肽偶联(参见Aslam等人,1998,生物偶联:生物医学的蛋白质偶联技术(Bioconjugation:Protein Coupling Techniques for the Biomedical Sciences),Macmillan Reference LTD;Hermanson GT,1996,生物共轭体技术(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肽的C-末端偶联的5'胺、以及内部张紧性炔烃,使用点击化学与叠氮化物珠粒偶联的DBCO-dT(Glen Research,VA)。使用大摩尔过量的记录标签将记录标签与溶液中的肽偶联以驱动碳二亚胺偶联完成,并限制肽-肽偶联。或者,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)基团容易地标记记录标签寡核苷酸。或者,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(例如,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。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。

[2223] 示例4:肽上氨基酸的位点特异性标记

[2224] 氨基酸可以直接或间接地用DNA标签进行位点选择性修饰。对于直接标记,可以使用位点选择性化学来活化DNA标签,或者对于间接标记,可以使用异双官能化学来将特异性氨基酸反应性部分转化为通用点击化学,之后可以将DNA标签附着到通用点击化学(Lundblad 2014)。在一个实例中,说明了五个不同氨基酸的位点选择性标记的示例,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,点击基团后来用于附着于DNA标签上的同源点击基团(Lundblad 2014)。

[2225] 典型的蛋白质输入包含溶有1μg蛋白质的50μl适当水性缓冲液(含有0.1%RapiGest™ SF表面活性剂和5mM TCEP)。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---200μM Sulfo-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BS3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粒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)。通过溶于硼酸盐缓冲液(50mM,pH 8.5)+1%TX-100的200μM重氮衍生的DNA标签来冰上孵育蛋白质或珠粒结合肽1小时来产生与酪氨酸连接的重氮(Nguyen,Cao等人,2015)。使用EDC化学修饰天冬氨酸/谷氨酸,使用溶于pH 6.5MES的珠粒结合肽和100mM EDC/50mM咪唑将胺标记的DNA标签在室温下孵育1小时(Basle等人,2010,Chem.Biol.17:213-227)。标记后,使用来自C4树脂ZipTips(Millipore)的蛋白质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。

[2226] 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粒上

[2227] 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珠粒(约1.3x$10^8$珠粒)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。

[2228] 示例6:制备醛基吲哚活性HyNic珠粒

[2229] 胺珠粒的HyNic衍生化产生醛基吲哚反应性珠粒。将等分的20mgDynabeads®M-270胺珠(2.8μm)珠粒悬浮于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。然后将珠粒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。将珠粒以终浓度为10mg/ml悬浮在柠檬酸盐缓冲液中。

[2230] 示例7:将记录标签醛基吲哚标记的肽固定到活化珠粒上

[2231] 将等分的1-2mg HyNic活化的Dynabeads®M-270珠(约1.3×$10^8$个珠粒)稀释于补充有50mM苯胺的100μl柠檬酸盐缓冲液中,加入约1ng记录标签肽缀合物并在37℃孵育1小时。珠粒用200μl柠檬酸盐缓冲液洗涤3次,并重悬于100μl硼酸盐缓冲液中。

[2232] 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来记录结合剂历史

[2233] 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)将寡核苷酸固定在内部位点。在图24A所示的示例中,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,A寡核苷酸和B寡核苷酸。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,编码标签通过共同的间隔区(Sp)与记录标签相互作用以引发引物延伸(或连接)。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,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)。

[2234] 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靶标分子之间的交叉相互作用。假设特异性杂交,CD靶标的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,AB目标的延伸记录标签不应包含C或D编码标签信息。在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大分子间隔开。

[2235] 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。到的最终产物应分别读为: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。

[2236] 示例9:寡核苷酸-肽模型系统-通过转移编码标签的识别信息到循环中的记录标签的来记录结合剂历史

[2237] 在验证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,从寡核苷酸系统构建肽模型系统(参见,例如,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)等。可选的Cys-Ser-Gly接头可以用来将肽表位标签偶联到寡核苷酸。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,例如,图26)。A_寡核苷酸-cMyc肽构建体还含有CSG接头和N-末端磷酸酪氨酸。同样,同源肽结合剂、cMyc抗体和HA抗体分别用B寡核苷酸编码标签和D寡核苷酸编码标签标记。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,肽模型系统与寡核苷酸系统平行,并且寡

核苷酸结合和抗体结合都在该模型系统中测试。

[2238] 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。通常在室温下孵育30分钟。还进行在1次PBST中使用1%PVP的标准预封闭以及染色后洗涤。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。

[2239] 如Williams等人描述,靶寡核苷酸含有用于附着叠氮化珠粒的内部炔标签,并且其5'末端含有氨基用于SMCC介导的与肽的C-末端半胱氨酸的附着(2010,Curr Protoc Nucleic Acid Chem.Chapter 4:Unit4.41)。或者,标准碳二亚胺偶联用于寡核苷酸和肽的缀合反应(Lu et等人,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。

[2240] 示例10:通过DNA/PNA编码标签连接补充记录标签来进行编码标签转移

[2241] 通过连接将编码标签直接或间接转移到记录标签以产生延伸记录标签。在一个实施例中,将编码标签的退火互补物连接到记录标签(参见,例如,图25)。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的情下,连接可以是酶促的,或者对于DNA/RNA尤其是肽核酸PNA,连接可以是化学介导的。

[2242] 对于DNA的酶促连接,退火的编码标签要求5'磷酸连接到记录标签的3'羟基。示例性酶促连接条件如下(Gunderson,Huang等人,1998):标准T4 DNA连接反应包括:50mM Tris-HCl(pH 7.8),10mM $MgCl_2$,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大肠杆菌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。

[2243] 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等人,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(pH 6.0,含HCl)或50mM HOBt(pH6.0,含HCl)和3.0-4.0M TMACl(Sigma)组成的反应中,在室温下进行1小时。

[2244] 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记录标签)的连接,并且由Brudno等人描述(Brudno,Birnbaum等人,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的珠粒,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。化学连接反应在室温下孵育1小时。Sakurai等描述了PNA连接的其他示例性方法(Sakurai,Snyder等人,2005)。

[2245] 示例11:PNA翻译成DNA

[2246] 使用退火到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。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等人,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(pH 8.5),0.2M NaCl)与固相结合的PNA分子孵育(23-50℃)30分钟。退火后,将固相结合的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。“点击化学”反应条件如下:将珠粒上的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珠粒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)。

[2247] 示例12:与核酸记录和编码标记相容的温和型N-末端埃德曼降解

[2248] N末端埃德曼降解和DNA编码之间的相容性允许这种方法用于肽测序。N-末端埃德曼降解法的标准条件(采用无水TFA)破坏DNA。然而,这种影响可以通过开发更温和的裂解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)-肽的裂解优化和在裂解条件下测量的DNA/PNA编码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 and 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,温和型N末端埃德曼降解的条件用在乙腈中采用三乙胺乙酸盐60℃10分钟的温和碱基裂解代替无水TFA裂解(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)。除了上述方法之外,DNA条形码可以通过筛选既能经受埃德曼标记又能经受裂解的经验来选择。

[2249] 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型N-末端埃德曼降解反应的相容性。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、和使用单个埃德曼降解步骤的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移除

[2250] 示例13:制备隔室标记的珠粒。

[2251] 为了制备隔室标记的珠粒,使用均分与合并合成方法,采用亚磷酰胺合成或通过均分与合并连接将条形码掺入到固定在珠粒上的寡核苷酸中。隔室标签可进一步包含独特分子标识符(UMI)以唯一地标记隔室标签连接的每个肽或蛋白质分子。示例性的隔室标签序列如下:5'-$NH_2$-GCGCAATCAG-XXXXXXXXXXXX-NNNNN-TGCAAGGAT-3'(SEQ ID NO:177)。XXXXXXXXXXXX(SEQ ID NO:178)条形码序列是每个珠粒的固定的核碱基序列群,其通过在珠粒上的均分与合并合成产生,其中固定的序列在珠粒与珠粒之间不同。NNNNN(SEQ ID NO:179)序列在珠粒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,可以使用均分与合并方法在珠粒上合成条形码序列。(2015,Cell 161:1202-1214,通过引用整体并入)。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。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活化的隔室标记珠粒偶联。(2010,Curr Protoc Nucleic Acid Chem.Chapter 4:Unit 4.41)。即,将200μl磁性珠粒(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%吐温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,Thermo Fisher)加入到磁性珠粒中并混合。反应物在室温下在旋转混合器上孵育1小时。孵育后,在磁铁上将珠粒与上清液分离,并用500μl偶联缓冲液洗涤3次。将这些珠粒再悬浮在400μl偶联缓冲液中。向磁性珠粒中加入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)。反应物于室温下在旋转混合器上孵育2小时。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,pH7.4)加入到反应混合物中并旋转混合器上孵育2小时。反应混合物用偶联缓冲液洗涤3次。将得到的珠粒重悬于储存缓冲液(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。

[2252] 示例14:包封的珠粒和蛋白质的生成

[2253] 隔室标记的珠粒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等人,2009,Bandara,Walsh等人,2011,Cao,Nguyen等人,2015)。将来自实施例12的隔室标记的珠粒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,例如,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。调节浓度和流动条件,使得平均每个液滴少于一个珠粒。在优化的实验中,可以制备$10^8$个femto-液滴,其中占10%的液滴含有珠粒(Shim等人,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℃孵育1小时以消化蛋白质成肽。消化后,通过加热乳液至80℃持续15分钟使Endo AspN失活。在双流体方案中,在将两种流体合并成液滴的过程中引入$Zn^{2+}$。在这种情况下,Endo AspN可以通过使用光活化的$Zn^{2+}$

笼状分子灭活,其中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。

[2254] 在将隔室化蛋白质消化成肽后,使用butelase I或化学连接(例如,醛-氨基等)将肽连接到珠粒上的隔室标签(寡核苷酸肽条形码嵌合体)(参见,例如,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butelase I连接不可逆(Nguyen,Cao等人,2015)。在连接后,乳液"破裂",并且具有固定的隔室标记的肽构建体的珠粒被大量收集,或者隔室标记的肽从珠粒上切下,并且大量收集。如果固定了隔室标记的肽的珠粒包含记录标签,则这些珠粒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,如果隔室标记的肽从珠粒基板上切下,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进行循环测序分析后,使用引物延伸或连接将隔室信息转移到所述相关的记录标签上(参见,例如,图22B)。在将隔室标签信息转移到记录标签后,可以使用原始肽消化中使用的相同酶将隔室标签从肽上裂解下来(参见,例如,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。

[2255] 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双标签生成

[2256] 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。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1x PHUSIONPHUSION™ 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PHUSIONPHUSION™高保真DNA聚合酶(Thermo Fisher Scientific)。每5至10秒向2ml冷沉淀瓶中溶解在轻质矿物油中的200μl油相(4.5%vol./vol.,Span 80,0.4%vol./vol.,吐温80,以及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等人,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℃持续30秒;40个循环,98℃持续10秒,70℃持续30秒和72℃持续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℃持续30秒;40个循环,98℃持续10秒,55℃持续30秒和72℃持续30秒;然后保持在4℃。在PCR的最后一个循环后,通过将200μl己烷(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。

[2257] 示例16:对延伸记录标签、延伸编码标签或双标签构建体进行测序

[2258] 记录标签或编码标签的间隔区(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区序列上。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,延伸报告标签、延伸编码标签、或双标签的结合循环序列被SBS测序,并且非相关的间隔区或通用引发序列被"跳过"。可以改变所述序列的5'末端的间隔区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。

[2259] 示例17:蛋白裂解物的制备。

[2260] 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NGPA测定,天然构象或变性蛋白质都可以固定在固体基板上(参见图32)。此外,在固定天然蛋白质后,固定在基板表面上的蛋白质可以变性。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,0.1N NaOH)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移除退火的编码标签。

[2261] 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(pH 8),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。

[2262] 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(1mL,每100mm培养皿或150$cm^2$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℃孵育20分钟。然后在4℃以约12,000rpm(取决于细胞类型)离心20分钟以分离蛋白质上清液。使用BCA测定法定量所述蛋白质,并以1mg/ml重悬于PBS中。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℃。

[2263] 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示例如下: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,持续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150$cm^2$

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)加入裂解物中并在37℃下孵育30分钟去除DNA和RNA。

[2264] 通过每100uL裂解物加入5μL的200mM DTT并在45℃下孵育30分钟来还原蛋白质。通过每100uL裂解物加入10uL 400mM碘乙酰胺并在黑暗中于24℃孵育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的200mM DTT来淬灭。任选通过每100ul裂解物加入2μl酸酐和100μl 1M $Na_2CO_3$(pH 8.5)来酰化蛋白质。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等人,2015)。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BCA测定法定量蛋白质,并以1mg/ml重悬于PBS中。

[2265] 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等人,2014,Feist and Hummon 2015).

[2266] 示例18:分区标记的肽的生成

[2267] 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用-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寡核苷酸的隔室珠粒、隔室条形码、任选的UMI,和与附着于多肽的DNA标签的一部分互补的引物序列(参见,例如,图31)。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粒相互作用,使得能够将相同的隔室条形码写入多肽分子的所有DNA标签。退火后,多肽结合DNA标签在退火的珠粒结合DNA序列上引发聚合酶延伸反应。以这种方式,将隔室条形码和其他功能元件写到与结合的多肽附着的DNA标签上。完成该步骤后,多肽具有多个记录标签附着,其中记录标签具有共同的间隔区序列、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残基处裂解而不是Lys残基(因为Lys残基被标记)。蛋白酶消化可以直接在珠粒上进行,或者标记的多肽从条形码化的珠粒上移除后进行。

[2268] 示例19:制备用于模型系统的DNA记录标签-肽缀合物。

[2269] 记录标签寡核苷酸用5' $NH_2$基团和内部mTetrazine基团合成,用于随后偶联珠粒(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),将寡核苷酸的5' $NH_2$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,pH 7.2)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的重悬的寡核苷酸中,混合并在室温下孵育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(约0.5mM)重悬于DMSO中。将约50nmol肽(100μl)加入到反应混合物中,并在室温下孵育过夜。如William等人所述,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重悬于磷酸盐偶联缓冲液中。

[2270] 示例20:用于DNA-肽固定的基板的开发。

[2271] 在一些实验中,适于点击化学固定的磁性珠粒通过将M-270胺磁性Dynabead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衍生的珠粒来产生(参见,例如,图29D-E;图30D-E)。即,洗涤10mg M-270珠粒并重悬于500μl硼酸盐缓冲液(100mM硼酸钠,pH 8.5)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胺珠粒一起孵育过夜。滴定甲基与TCO PEG的比例以调节珠粒上的最终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溶于DMF的0.1M乙酸酐和0.1M DIEA(10mg珠粒用500μl)的混合物加盖2小时。加盖并在DMF中洗涤3次后,珠粒重悬于10mg/ml的磷酸盐偶联缓冲液中。

[2272] 在一个实验中,在室温下,在0.1M Hepes,pH 7.5,5mM甲基四嗪-PEG4-叠氮化物(BroadPharm),2.5mM $CuSO_4$,5mM Tris(3-羟丙基三唑基甲基)胺,8mM抗坏血酸钠)中孵育总共10nmol的包括5'氨基和内部炔基修饰的记录标签。孵育16小时后,用$LiClO_4$和83％丙酮沉淀记录标签,然后溶解在50μl的150mM NaCl/NaPhos(pH 7.4),1mM EDTA中。通过分光光度计(Thermo Fisher)测量A260来估计记录标签浓度,并调节至100μM。

[2273] TCO(反式-环辛烯)珠粒由Dynabead M-270胺(Thermo Fisher)制备。将总计0.45mg的Dynabead M-270胺用150μl的150mM NaCl/NaPhos(pH 7.4)洗涤两次,并重悬于150μl的150mM NaCl/NaPhos(pH 7.4)中。另一方面,通过对mPEG-SCM,MW 550(Creative PEGWorks)滴定TCO-PEG1 2-NHS酯(BroadPharm)/DMF来制备TCO-PEG12-NHS/mPEG-SCM混合物的DMF溶液,比率为$1:10^2$,$1:10^3$和$1:10^4$。将总共10μl的单独的TCO-PEG12-NHS/mPEG-SCM混合物加入到150ul预洗涤的珠粒悬浮液中,并且在室温下孵育30分钟以用不同分布的TCO涂覆珠粒。用150μL的150mM NaCl/NaPhos(pH 7.4)洗涤TCO包被的珠两次,并重悬于150ul的150mM NaCl/NaPhos(pH 7.4)中。通过在室温下用6.7μM NHS-乙酸酯(TCI)孵育15分钟来封闭珠粒上的残余胺基。将得到的珠粒用150μL的150mM NaCl/NaPhos(pH 7.4)洗涤两次。

[2274] 将总计0.45mg的单独的TCO-改性的珠粒重悬于10μl的100uM甲基四嗪改性的记录标签溶液中,在室温下孵育30分钟,并且用150μl的150mM的NaCl/NaPhos(pH 7.4)洗涤两次。将记录标签固定的珠粒重悬于75μl的150mM NaCl/NaPhos(pH 7.4),1mM EDTA中。

[2275] 含半胱氨酸的肽通过SM(PEG)8(Thermo Fisher)附着于珠粒上记录标签的5'胺基。将总计5μl的100mM SM(PEG)/DMF溶液添加至150μl的珠粒悬浮液中,在室温下孵育30分钟。将得到的珠粒用150μl的150mM NaCl/NaPhos(pH 7.4)洗涤两次,并重悬于20μl的150mM NaCl/NaPhos(pH 7.4),1mM EDTA中。将总计6μl的8mM肽/DMF溶液添加至20μL的珠粒悬浮液中,并且在室温下孵育2小时并且然后在4℃下孵育16小时。将这些珠粒用150μl的150mM的NaCl/NaPhos(pH 7.4)洗涤两次,重悬于150μl的10μg/ml L-半胱氨酸(Sigma-Aldrich)溶液中,并且在室温下孵育2小时以去除残留的马来酰亚胺基团。用150μl的150mM的NaCl/NaPhos(pH 7.4)洗涤珠粒两次,并且重悬于75μl的150mM的NaCl/NaPhos(pH 7.4),1mM EDTA中。

[2276] 示例21:将记录标签标记的肽固定到基板。

[2277] 出于分析的目的,使用记录标签上的mTet基团和活化珠粒或基板表面的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板上。即使在低输入反应物浓度下,该反应也是快速高效的。此外,使用甲基四嗪对键赋予更大的稳定性(Selvaraj以及Fox 2013,Knall,Hollauf等人,2014,Wu以及Devaraj 2016)。将约200μg与约500μg之间的M-270TCO珠粒重悬于100μl磷酸盐偶联缓冲液中。将5pmol在记录标签上包含mTet基团的DNA记录标签标记的肽加入到珠粒中,使最终浓度为约50nM。将反应物在室温下孵育1小时。固定后,将基板上未反应的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。

[2278] 肽珠粒的大量制备。在室温下,在0.1M Hepes,pH 7.5,5mM甲基四嗪-PEG4-叠氮化物(BroadPharm),2.5mM $CuSO_4$,5mM Tris(3-羟丙基三唑基甲基)胺,8mM抗坏血酸钠)中孵育总共10nmol的包括5'氨基和内部炔基修饰的记录标签。孵育16小时后,用$LiClO_4$和83%丙酮沉淀记录标签,然后溶解在50μl的150mM NaCl/NaPhos(pH 7.4),1mM EDTA中。通过分光光度计(Thermo Fisher)测量A260来估计记录标签浓度,并调节至100μM。TCO(反式-环辛烯)珠粒由Dynabead M-270胺(Thermo Fisher)制备。将总计0.45mg的Dynabead M-270胺用150μl的150mM NaCl/NaPhos(pH 7.4)洗涤两次,并重悬于150μl的150mM NaCl/NaPhos(pH 7.4)中。另一方面,通过对mPEG-SCM,MW 550(Creative PEGWorks)滴定TCO-PEG1 2-NHS酯(BroadPharm)/DMF来制备TCO-PEG12-NHS/mPEG-SCM混合物的DMF溶液,比率为$1:10^2$,$1:10^3$和$1:10^4$。将总共10μl的单独的TCO-PEG12-NHS/mPEG-SCM混合物加入到150μl预洗涤的珠粒悬浮液中,并且在室温下孵育30分钟以用不同分布的TCO涂覆珠粒。用150μL的150mM NaCl/NaPhos(pH 7.4)洗涤TCO包被的珠粒两次,并重悬于150ul的150mM NaCl/NaPhos(pH 7.4)中。通过在室温下用6.7μM NHS-乙酸酯(TCI)孵育15分钟来封闭珠粒上的残余胺基。将得到的珠粒用150μL的150mM NaCl/NaPhos(pH 7.4)洗涤两次。将总计0.45mg的单独的TCO-改性的珠粒重悬于10μl的100μM甲基四嗪改性的记录标签溶液中,在室温下孵育30分钟,并且用150μl的150mM NaCl/NaPhos(pH 7.4)洗涤两次。将记录标签固定的珠粒重悬于75μl的150mM NaCl/NaPhos(pH 7.4),1mM EDTA中。含半胱氨酸的肽通过SM(PEG)8(Thermo Fisher)附着于珠粒上记录标签的5'胺基。将总计5μl的100mM SM(PEG)/DMF溶液添加至150μl的珠粒悬浮液中,在室温下孵育30分钟。将得到的珠粒用150μl的150mM NaCl/NaPhos(pH 7.4)洗涤两次,并重悬于20μl的150mM NaCl/NaPhos(pH 7.4),1mM EDTA中。将总计6μl的8mM肽/DMF溶液添加至20μL的珠粒悬浮液中,并且在室温下孵育2小时并且然后在4℃下孵育16小时。将珠粒用150μl的150mM的NaCl/NaPhos(pH 7.4)洗涤两次,重悬于150μl的10μg/ml L-半胱氨酸(Sigma-Aldrich)溶液中,并且在室温下孵育2小时以去除残留的马来酰亚胺基团。将珠粒用150μl的150mM NaCl/NaPhos(pH 7.4)洗涤两次,并重悬于75μl的150mM NaCl/NaPhos(pH 7.4),1mM EDTA中。

[2279] 示例22:N-末端氨基酸(NTAA)修饰

[2280] 化学NTAA乙酰化:

[2281] 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与肽一起在室温下温育30分钟(Halpin,Lee等人,2004)。或者,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)并在室温孵育10-30分钟(Goodnow 2014)。

[2282] 酶促NTAA乙酰化:

[2283] 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:使用溶于NAT缓冲液(20mM Tris-HCl,pH 8.0,100mM NaCl,1mM EDTA,1mM乙酰辅酶A)的2μM SsArd1将肽在65℃孵育10分钟(Chang和Hsu 2015)。

[2284] 化学NTAA脒化(胍基化):

[2285] 使用溶于DMF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)将肽在室温下孵育30分钟。或者,使用溶于DMF的10mM 1H-吡唑-1-甲脒盐酸盐,10mM DIEA将肽在室温下孵育30分钟。用标准解封闭方法去除保护基团。或者,使用溶于PBS缓冲液(pH 8.0)或100mM硼酸盐缓冲液(pH 8.0)的10mM S-甲基异硫脲将肽在10℃孵育30分钟(Tse,Snyder等人,2008)。

[2286] PITC标签:

[2287] 使用离子液体[Bmim][BF4]中5%(vol./vol.)PITC将肽在室温下孵育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。

[2288] DNFB标签

[2289] 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。通过将溶于10mM硼酸盐缓冲液(pH 8.0)的0.5-5.0μg/ml DNFB于37℃孵育5-30分钟来标记肽。

[2290] SNFB标签:

[2291] 4-磺酰基-2-硝基-氟苯(SNFB)在甲醇中制备为5mg/ml储备液。该溶液应避光保存并当天使用。通过将溶于10mM硼酸盐缓冲液(pH 8.0)中的0.5-5.0μg/ml DNFB于37℃孵育5-30分钟来标记肽。

[2292] 乙酰化的NTAA肽的裂解:

[2293] 通过将溶于25mM Tris-HCl(pH 7.5)的10μM酰基肽水解酶(APH)酶(来自Sulfolobus solfataricus,SSO2693)在90℃孵育10分钟,从肽上切下乙酰化的NTAA(Gogliettino,Balestrieri等人,2012)。

[2294] 酰胺化(胍基化)NTAA肽的裂解:

[2295] 通过在约0.1N至约0.5N的NaOH中在37℃孵育10分钟,将酰胺化(胍基化)的NTAA从肽上切下(Hamada 2016)。

[2296] 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

[2297] 用DNA模型系统测试编码标签信息到固定到珠粒上的记录标签(参见,例如,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 NOs:1-65)和它们的反向互补序列(SEQ ID NOs: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,编码标签上的条形码(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 NOs: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15-mer条形码,BC_5,和BC_5&BC_6。这种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(参见,例如,图36B))。

[2298] 评估了记录标签密度对"分子内"与"分子间"信息转移的影响。对于正确的信息转移,应该观察到"分子内"信息转移("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"信息转移(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粒表面的影响,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链霉亲和素珠粒(Thermo)在37℃孵育15分钟。在室温下用100μl固定缓冲液洗涤珠粒3次。大多数随后的洗涤步骤使用100μl的体积。通过将珠粒重悬于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)退火至固定在珠粒上的记录标签。通过加热至65℃持续1分钟将编码标签退火至记录标签,然后使其缓慢冷却至室温(0.2℃/秒)。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%吐温20)洗涤珠粒,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。将珠粒重悬于19.5μl $MgSO_4$延伸缓冲液(50mM Tris-Cl(pH 7.5),2mM $MgSO_4$,125μM dNTP,50mM NaCl,1mM二硫苏糖醇(DTT),0.1%吐温20和0.1mg/ml BSA)中,并在37℃孵育15分钟。将Klenow exo-DNA聚合酶(NEB,5U/ul)加入到珠粒中,最终浓度为0.125U/ul,并在37℃下孵育5分钟。引物延伸后,珠粒用PBST洗涤2次,在室温下用50μl0.1NaOH洗涤1次持续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)杂交在珠粒上并延伸,如对编码标签寡核苷酸所操作。添加衔接序列后,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℃孵育5分钟从链霉亲和素珠粒中洗脱。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1μl PCR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(参见,例如,图36C),以及在珠粒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。

[2299] 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(参见,例如,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)的产物的大小分别是115和85个碱基对。如图36D所示,在珠粒上存在以高密度的saRT_ABC(SEQ ID NO:141)和saRT_BBC(SEQ ID NO:142)时观察到三条带。预期重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,记录标签以低密度在珠粒表面上间隔开,导致分子间事件减少。

[2300] 表1.模型系统序列

[2301]

| 名称 | 序列(5'-3') | SEQ ID NO: |
|---|---|---|
| saRT_Abc_v2 | /5Biosg/TTTTTGCAAATGGCATTCTGACATCCCGTAGTCCGCGACACTAGATGTCTAGCATGCCGCCGTGTCATGTGGAAACTGAG | 141 |
| saRT_Bbc_v2 | /5Biosg/TTTTTTTTTTGACTGGTTCCAATTGACAAGCCGTAGTCCGCGACACTAGTAAGCCGGTATATCAACTGAGTG | 142 |
| saDummy-pT10 | /5Biosg/TTTTTTTTTT/3SpC3/ | 143 |
| CT_A'-bc | GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGCACTCAGT/3SpC3/ | 144 |
| CT_A'-bc_1PEG | GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/ | 145 |
| CT_A'bc_5PEG | GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18//iSP18//iSP18//iSP18//iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3 | 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/ | 155 |
| dupCT_B'BC | AACTGCCGAGATTCCCGTATACGCGTTAGGACTGAGTG/3SpC3/ | 156 |

[2302] /3SpC3/=3' C3(三碳)间隔区

[2303] /5Biosg/=5'生物素

[2304] /iSP18/=18-原子已-乙二醇间隔区

[2305] 示例24:对延伸记录标签、延伸编码标记或双标签构建体的纳米测序

[2306] 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,etc等。容错条形码基于使用R Bioconductor包的汉明码和Levenshtein码,"DNA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,例如,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等人,2014)。

[2307] 为了模拟使用纳米孔测序的延伸记录标签、延伸编码标签、或双标签构建体的分析,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(参见,例如,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。DTF和DTR引物通过互补的15-mer间隔区序列(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。

[2308] PCR条件:

| 试剂 | 终浓度 |
|---|---|
| F1(5'磷酸化的)(SEQ ID NO:165) | 1uM |
| R1(5'磷酸化的)(SEQ ID NO:166) | 1uM |
| 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 | 200uM |
| 水 | |
| VeraSeq 2.0 Ultra Pol | 2U/100ul |

[2309]

| PCR 循环: | |
|---|---|
| 98℃ | 30 秒 |
| 50℃ | 2 分钟 |
| 98℃ | 10 秒 |
| 55℃ | 15 秒 |
| 72℃ | 15 秒 |
| 重复最后 3 步 19 个循环 | |
| 72℃ | 5 分钟 |

[2310] PCR后,扩增子通过平端连接(图27C)级联如下:将20ul PCR产物直接与20ul Quick Ligase Mix(NEB)混合,并在室温下孵育过夜。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7μl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(参见,例如,图27D,显示SEQ ID NO:168,延伸记录标签构建体的序列)。

[2311] 示例1475:凝胶珠粒中包封的单个细胞

[2312] 使用标准技术将单细胞包封成液滴(约50μm)(Tamminen以及Virta 2015,Spencer,Tamminen等人,2016)(参见,例如,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等人,2014,Kang,Yamauchi等人,2016)。固定在所得单细胞凝胶珠内的蛋白质可以使用多种方法进行单细胞条形码编码。在一个实施例中,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质上。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DNA条形码标签转移到蛋白质上,或者单细胞凝胶珠中的蛋白质可以通过一系列如Amini,Cusanovich和Gunderson等人描述的汇集-分开步骤组合索引(Amini,Pushkarev等人,2014,Cusanovich,Daza等人,2015)(Gunderson,Steemers等人,2016)。在最简单的实施中,单细胞凝胶珠中的蛋白质首先用“点击化学”基团标记(参见,例如,图40),然后使用汇集-分开方法将组合DNA条形码点击到蛋白质样品上。

[2313] 示例26:用基于DNA的模型系统演示单链DNA连接的信息转移

[2314] 用DNA模型系统测试编码标签信息到固定到珠粒上的记录标签(参见,例如,图46A)的转移。使用两种不同类型的记录标签寡核苷酸:saRT_Bbca_ssLig(SEQ ID NO:181)ssDNA构建体,其是5’磷酸化和3’生物素化的,并且含有独特的6碱基DNA条形码,BCa,通用正向引物序列和靶结合“B”序列;saRT_Abca_ssLig(SEQ ID NO:182)ssDNA构建体,其是5’磷酸化的和3’生物素化的,并含有独特的6碱基DNA条形码,BCa,通用正向引物序列和靶结合“A”序列。编码标签寡核苷酸CT_B’bcb_ssLig(SEQ ID NO:183)含有B’序列。这种记录标签和编码标记以及相关结合元件的设计使得凝胶分析变得容易。通过CircLigase II(Lucigen)产生期望的单链DNA连接产物,其中通过使编码标签上的B’序列与固定在固体表面上的记录标签上的B序列退火,使记录标签的5’磷酸基和编码标签的3’羟基紧密接近。

[2315] 通过凝胶分析评价经由B编码标签和B记录标签之间的特异性相互作用的信息转移。通过用生物素化的记录标签寡核苷酸,SaRT_Bbc_ssLig或SaRT_Abc_ssLig以1:10的比率滴定mPEG-生物素MW 550(Creative PEGWorks)来调节表面上的记录标签的密度。将总共2pmol记录标签寡核苷酸与5μl M270链霉亲和素珠粒(Thermo)在50μl固定化缓冲液(5mM Tris-Cl,pH 7.5,0.5mM EDTA,1M NaCl)中37℃下孵育15分钟,用150μl固定化缓冲液洗涤一次,并且用150μl PBST+40％甲酰胺洗涤一次。对于模型测定,将总计40pmol CT_B’bcb_ssLig与5μl记录标签固定的珠粒在50μl PBST中37℃下孵育15分钟。在室温下用150μl PBST+40％甲酰胺洗涤珠粒3次。将这些珠粒重悬于10ul CircLigase II反应混合物(0.033M Tris-Acetate,pH 7.5,0.066M乙酸钾,0.5mM DTT,2mM $MnCl_2$,0.5M甜菜碱,和4U/uL CircLigase II ssDNA连接酶)中并且在45℃下孵育2小时。连接反应后,珠粒用固定化缓冲液+40％甲酰胺洗涤一次,用PBST+40％甲酰胺洗涤一次。通过在10μl 95％甲酰胺/10mM EDTA中65℃下孵育5分钟,将最终延伸记录标签寡核苷酸从链霉亲和素珠粒中洗脱,并且将2.5μl洗脱液加载到15％ PAGE-尿素凝胶中。

[2316] 在该模型系统中,来自47个碱基记录标签的连接产物的大小为96个碱基(参见,例如,图46B)。在saRT_Bbca_ssLig存在下观察到连接产物带,而在saRT_Abcb_ssLig存在下没有观察到产物带。这一结果表明,特异性B/B’序列结合事件是通过编码标签向记录标签的信息转移而编码的。此外,第一循环连接产物用USER Enzyme处理,用于第二信息转移。这些事件是通过凝胶分析观察到的(参见,例如,图46C)。

[2317] 表2.基于肽和基于DNA的模型系统序列

[2318]

| 名称 | 序列（5'-3'） | SEQ ID NO: |
|---|---|---|
| saRT_Bbca_ssLig | /5Phos/TGACATCTAGTGTCGCGGACTACGTGCTTGTCAATTGGAACCAGTCT/3Bio/ | 181 |
| saRT_Abca_ssLig | /5Phos/TGACATGTGAAATTGTTATCCGCTCATGGATGTCAGAATGCCATTTGCT/3Bio/ | 182 |
| CT_B'bcb_ssLig | GACTGGTTCCAATTGACAAGC/iSP18//iSP18//iSP18/CGATTTGCAAGGATCACTCGUTTTAGGT | 183 |

[2319] /5Phos/＝5'-生物素

[2320] /3Bio/＝3'-生物素

[2321] /iSP18/＝18-原子已-乙二醇间隔区

[2322] 示例27:用基于DNA的模型系统演示双链DNA连接的信息转移

[2323] DNA模型系统用于测试将编码标签到固定在珠粒上的记录标签上的信息转移(参见图47A)。记录标签寡核苷酸由两条链组成。saRT_Abc_dsLig(SEQ ID NO:184)为5'生物素化DNA,其含有靶结合剂A序列、通用正向引物序列、两个独特的15个碱基的DNA条形码BC1和BC2,以及4个碱基的突出;Blk_RT_Abc_dsLig(SEQ ID NO:185)是5'磷酸化的和3'C3间隔区修饰的DNA,其含有两个独特的15个碱基的DNA条形码BC2'和BC1,通用正向引物序列。双链编码标签寡核苷酸由两条链组成。一条链,CT_A'bc5_dsLig(SEQ ID NO:186)其包含dU、独特的BC5以及经由聚乙二醇接头与靶向剂A'序列连接的突出。编码标签的另一条链是Dup_CT_A'bc5(SEQ ID NO:187),其含有5'磷酸、dU和独特的条形码BCM5'。这种记录标签和编码标签的设计可以轻松地进行凝胶分析。当两个标签的5'磷酸基和3'羟基彼此接近时,通过编码标签中的靶向剂A'与固定在固体表面上的记录标签中的靶结合剂A的杂交,用T4 DNA连接酶(NEB)连接所需的双链DNA连接产物。

[2324] 评价了靶结合剂A和靶向剂A'之间通过特异性相互作用的信息传递。为了将记录标签在珠粒表面上隔开,将生物素化的记录标签寡核苷酸,总计2pmol与Blk_RT_Abc_dsLig杂交的saRT_Abc_dsLig,以1:10的比率针对mPEG-SCM,MW550(Creative PEGWorks)进行滴定,并且与5μlM 270链霉亲和素珠粒(Thermo)在50μl固定化缓冲液(5mM Tris-Cl,pH 7.5,0.5mM EDTA,1M NaCl)中37℃下持续15分钟。将记录标签固定的珠粒用150μl固定化缓冲液洗涤1次,并用150μl固定化缓冲液+40％甲酰胺洗涤1次。对于第一循环测定,将总共40pmol双链编码标签CT_A'bc5_dsLig:Dup_CT_A'bc5与5μl记录标签固定的珠粒在50μl PBST中37℃下孵育15分钟。在室温下用150μl PBST+40％甲酰胺洗涤珠粒3次。将珠粒重悬于10μl T4 DNA连接酶反应混合物(50mM Tris-HCl,pH 7.5,10mM $MgCl_2$,1mM DTT,1mM ATP,7.5％PAG8000,0.1μg/μl BSA,和20U/μl T4 DNA连接酶)中并在室温下孵育60分钟。连接反应后,珠粒用固定化缓冲液+40％甲酰胺洗涤一次,用PBST+40％甲酰胺洗涤一次。将珠粒用USER Enzyme(NEB)处理以去除双链编码标签,并用于使用CT_A'bc13-R_dsLig:Dup_CT_A'bc13-R_dsLig(分别是SEQ ID NO:188以及SEQ ID NO:189)的第二循环连接测定。在每次处理之后,通过在65℃下在10μl95％甲酰胺/10mM EDTA中孵育5分钟从链霉亲和素珠粒上洗脱双链记录标签,并且将2.5μl洗脱液加载到15％ PAGE-尿素凝胶上。

[2325] 在该模型系统中,具有双链编码标签的76个碱基和54个碱基记录标签的连接产物的大小分别为116个和111个碱基(参见,例如,图4)。第一循环连接产物用USER Enzyme(NEB)完全消化,并用于第二循环测定。第二循环连接产物带在约150个碱基处观察到。这些

结果表明在第一循环和第二循环双链连接测定中编码特异性ASeq/A’seq结合事件。

[2326] 表3.基于肽和基于DNA的模型系统序列

[2327]

| 名称 | 序列（5’-3’） | SEQ ID NO: |
|---|---|---|
| saRT_Abc_dsLig | /5Biosg/TTTTTGCAAATGGCATTCTGACATCCCGTAGTCCGCGACACTAGATGTCTAGCATGCCGCCGTGTCATGTGGAAGA | 184 |
| Blk_RT_Abc_dsLig | /5Phos/CTCTTCTTCCACATGACACGGCGGCATGCTAGACATCTAGTGTCGCGGACTACG/3SpC3/ | 185 |
| CT_A’bc5_dsLig | GGATGUCAGAAUGCCATTTGCTTTTTTTTTT/iSP18/CGGTCTCUCTCTTCCCTAACGCGTATACGGA | 186 |

[2328]

| | | |
|---|---|---|
| Dup_CT_A’bc5_dsLig | /5Phos/AGAGTCCGTATACGCGTTAGGGAUGAGAGAGACCG/3SpC3/ | 187 |
| CT_A’bc13-R_dsLig | GGATGUCAGAAUGCCATTTGCTTTTTTTTTT/iSP18/CGGTCTCUCGATTTGCAAGGATCACTCGCCGTTATTGACGCTCGA | 188 |
| Dup_CT_A’bc13-R_dsLig | /5Phos/AGAGTCGAGCGTCAATAACGGCGAGTGATCCTTGCAAATCGAGAGACCG/3SpC3/ | 189 |

[2329] /5Biosg/＝5’生物素

[2330] /3SpC3/＝3’C3(三碳)间隔区

[2331] /5Phos/＝5’-生物素

[2332] /iSP18/＝18-原子已-乙二醇间隔区

[2333] 示例28:基于肽和DNA的模型系统的序列信息转移循环演示

[2334] 肽模型系统用于测试编码标签信息向固定在珠粒上的记录标签复合物的第一次循环转移(参见,例如,图48A)。将PA肽序列(SEQ ID NO:195)附着到固定在珠粒上的记录标签寡核苷酸amRT_Abc(SEQ ID NO:190)上。amRT_Abc序列含有“A”DNA捕获序列(用于“A”结合剂的模拟表位)和相应的“A”条形码(rtA-BC)。rtA_BC序列是两个条形码BC_1和BC_2(SEQ ID NOs:1-65)的共线组合。对于结合剂,将抗-PA抗体附着于编码标签寡核苷酸amCT_bc5(SEQ ID NO:191)上,该编码标签寡核苷酸amCT_bc5由15-mer条形码BC5(SEQ ID NOs:66-130)组成。此外,使用DNA模型系统来测试编码标签信息到记录标签的第二循环标签(参见,例如,图48B)。CT_A’_bc13由互补的捕获序列A’组成,并指定为15-mer条形码,BC5(SEQ ID NOs:66-130)。这种设计使得在用特异性引物组进行PCR扩增后能够容易地进行凝胶分析。

[2335] 内部炔修饰的记录标签寡核苷酸,amRT_Abc(SEQ ID NO:190)用甲基四嗪-PEG4-叠氮化物(BroadPharm)进行修饰。为了控制珠粒表面上记录标签的密度,由M-270胺Dynabeads(Thermo Fisher)制备具有不同密度的功能偶联位点(反式环辛烯,TCO)的珠粒,M-270胺Dynabeads(Thermo Fisher)通过相对于mPEG-SCM,MW550(Creative PEGworks)以$1:10^2$,$1:10^3$和$1:10^4$的比率滴定TCO-PEG12-NHS酯(Broadpharm)衍生化。将甲基四嗪修饰的amRT-Abc记录标签附着到反式-环辛烯(TCO)衍生的珠粒上。将含有Cys的肽通过SM(PEG)8(Thermo Fisher)附着到amRT_Abc的5’胺基上。使用蛋白质-寡核苷酸缀合试剂盒(Solulink)完成抗PA抗体(Wako Chemicals)与amCT_bc5编码标签的缀合。简单地说,将amCT_bc5的5’胺基用S-4FB修饰,然后用0.5mL的Zeba柱脱盐。抗PA抗体经S-HyNic修饰后,用0.5mL的Zeba柱脱盐。最后,将4FB修饰的amCT_bc5与HyNic修饰的抗PA抗体混合,制备抗体-编码标签缀合物,然后使用Bio-Gel P100(Bio-Rad)进行尺寸排阻。

[2336] 对于第一循环结合测定,将5μl的肽-记录标签(RT)固定的珠粒与SuperBlock T20(TBS)封闭缓冲液(Thermo Fisher)在室温下孵育15分钟以封闭珠粒。将总计2pmol的抗体编码标签缀合物与5μl的肽记录标签固定的珠粒在50μl PBST中在37℃下孵育30分钟。在室温下用1000μl PBST+30％甲酰胺洗涤珠粒3次。将这些珠粒重悬于50μl延伸反应主混合物(50mM Tris-Cl(pH 7.5,2mM $MgSO_4$,125μM dNTPs,50mM NaCl,1mM二硫苏糖醇(DTT),0.1％吐温20,0.1mg/ml BSA,和0.05U/μL Kenow exo-DNA聚合酶)中并且在37℃下孵育5分钟。引物延伸后,用固定化缓冲液(5mM Tris-Cl(pH 7.5),0.5mM EDTA,1M NaCl,30％甲酰胺),用50μl的0.1N氢氧化钠在室温下5分钟洗涤一次,用PBST+30％甲酰胺洗涤一次,用PBS洗涤一次。对于第二结合循环测定,CT_A'_bc13用于结合记录标签内的其同源A序列,并且使得第一循环延伸记录标签的延伸能够在第二循环编码标签序列上延伸。在延伸之后,将最终延伸记录标签寡核苷酸在20μl PCR混合物中用特异性引物进行PCR扩增,并且在10％ PAGE凝胶上分析1μl PCR产物。所得凝胶演示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48C-E)。

[2337] 在图48A所示的模型系统中,来自使用引物组P1_F2和Sp/BC2的记录标签amRT_Abc的PCR产物的大小为56个碱基对。如图48C所示,观察到amRT-Abc密度依赖性带强度。对于使用抗-PA抗体-amCT_bc5缀合物的第一循环结合测定,当在测定中使用同源的PA-标签固定的珠粒时,观察到80个碱基对处的强带PCR产物,而当使用非同源的淀粉样β(A 16-27)或纳米标签固定的珠粒时,观察到最小的PCR产物产率(参见,例如,图48D)。对于使用附着于CT_A'_bc13编码标签的A'DNA标签的第二结合测定(参见,例如,图48B),所有三种风味的肽记录标签缀合物在退火的CT_A'_bc13序列上延伸。如图48E所示,对于所有固定肽的珠粒,在117个碱基对处观察到相对强的PCR产物带,这仅仅对应于原始记录标签(BC1+BC2+BC13)上的第二循环延伸。仅当在测定中使用PA-标签固定的珠粒时,在93个碱基对处观察到对应于第一延伸记录标签(BC1+BC2+BC5+BC13)上的第二延伸的带。这些结果表明特异性肽/抗体和Aseq/A'seq结合事件分别在第一循环和第二循环测定中编码。

[2338] 表4.基于肽和基于DNA的模型系统序列

[2339]

| 名称 | 序列（5’-3’） | SEQ ID NO: |
|---|---|---|
| amRT_Abc | /5AmMC6/GCAAATGGCATTCTGACATCCTT/i5OctdU/TTCGUAGUCCGCGACACTAGATGTCTAGCATGCCGCCGTGTCATGTGGAAACTGAGTG | 190 |
| amCT_bc5 | /5AmMC6//iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/ | 191 |
| CT_A’_bc13 | GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CGATTTGCAAGGATCACTCGCCGTTATTGACGCTCTCACTCAGT/3SpC3/ | 192 |
| Sp | ACTGAGTG | 149 |
| Sp’ | CACTCAGT | 150 |
| P1_f2 | CGTAGTCCGCGACACTAG | 151 |
| R1 | CGATTTGCAAGGATCACTCG | 152 |
| Sp/BC2 | CACTCAGTTTCCACATGACACGGC | 193 |
| Sp/BC5 | CACTCAGTCCTAACGCGTATA | 194 |
| PA 肽 | GVAMPGAEDDVVGGGGSC | 195 |
| 纳米标签肽 | 甲酰基-MDVEAWLGARVPLVETGSGSGSC | 196 |
| Aβ 肽 | HQKLVFFAEDVGSGSGSC | 197 |

[2340] /3SpC3/＝3’C3(三碳)间隔区

[2341] /i5OctdU/＝5’-辛二炔基dU

[2342] /iSP18/＝18-原子己-乙二醇间隔区

[2343] 示例29:使用mRNA展示用DNA记录标签来标记蛋白质或肽

[2344] 通过PCR将单独的条形码安装在每个DNA编码蛋白质的3’端,并汇集条形码化的DNA。使用扩增的T7 Flash(Lucigen)转录扩增的DNA库。使用RNeasy Mini Kit(Qiagen)清理转录反应并通过NanoDrop3000(Fisher Scientific)定量。用T4 DNA连接酶(NEB)将DNA衔接子连接到mRNA的3’端。使用10％TBE-尿素变性凝胶纯化连接的mRNA分子。利用PURExpress试剂盒(NEB)在体外翻译mRNA-嘌呤霉素分子。在体外翻译期间,停顿的核糖体允许嘌呤霉素残基进入核糖体A-位点并附着于蛋白质的C-末端,产生蛋白质-mRNA融合体。通过与硅珠粒连接的互补寡核苷酸捕获蛋白质-mRNA融合体。mRNA部分使用ProtoScript II逆转录酶(NEB)转化为cDNA。用RNase H(NEB)和RNase cocktail(Thermo Fisher)处理蛋白质-cDNA/RNA库,产生蛋白质-cDNA,然后通过切滤器纯化。加入与cDNA中II型限制性位点互补的序列,形成双链,与限制性酶孵育,在cDNA的3’端产生间隔区序列(Sp)。池的一部分用于测序以表征起始蛋白质-cDNA池中的蛋白质表达。

[2345] 示例30:基于核糖体展示的蛋白质条形码

[2346] 对于相对小尺寸的蛋白质库(例如,在该工作中,<200),可以通过用条形码化的引物进行单独的PCR反应将条形码化序列引入DNA模板。汇集条形码化的线性DNA模板,并使用

HiScribe T7试剂盒(NEB)进行体外转录。将转录的mRNA用无DNA的试剂盒(Ambion)处理,用RNeasy Mini试剂盒(Qiagen)纯化并通过Nanodrop 1000(Thermo Scientific)定量。为了产生mRNA-cDNA杂合体,通过在缓冲液(50mM Tris-HCl,pH 8.3,75mM KCl和5mM $MgCl_2$)中,0.10μM mRNA,1μM 5'-丙烯酸和脱硫生物素修饰的引物,dNTP各0.5mM,10U/μLSuperscript III,2U/μL RNaseOUT(Invitrogen)和5mM二硫苏糖醇(DTT)在50℃下孵育30分钟,通过异丙醇沉淀富集得到的mRNA-cDNA杂合体,合成cDNA。通过异丙醇沉淀富集得到的mRNA-cDNA杂合体,并用链霉亲和素包被的磁性珠粒(Dynabeads M-270链霉亲和素,Life Technologies)纯化。将PURExpressΔ核糖体试剂盒(NEB)应用在大肠杆菌核糖体上的展示蛋白质。将与0.40μM mRNA-cDNA杂合体和0.30μM核糖体的250μL IVT反应物在37℃孵育30分钟,通过加入250μL冰冷的缓冲液HKM(50mM HEPES,pH 7.0,250mM KOAc,25mM Mg(OAc)2,0.25U/mL RNasin(Promega),0.5mg/mL氯霉素,5mM 2-巯基乙醇和0.1%(v/v)吐温20)淬灭,并离心(14000g,4℃)10分钟以除去不溶性组分。将总是保存在冰上或冷藏室中的PRMC复合物进行两步Flag标记和脱硫生物素标签亲和纯化以富集全长和条形码化的靶蛋白质。因此,使用抗Flag M2(Sigma-Aldrich)和链霉亲和素磁性珠粒相继纯化蛋白质,链霉亲和素磁性珠粒用补充有100μg/mL酵母tRNA和10mg/mL BSA的缓冲液HKM封闭。用含有100μg/ml Flag肽或5mM生物素的缓冲液HKM洗脱结合的蛋白质,并通过实时PCR定量它们的条形码化DNA。

[2347] 示例31:用NTAA检测系统进行单链DNA间隔区连接的信息转移

[2348] 对于DNA/聚合物嵌合体编码标签和记录标签,可通过使用标准核酸酶学的连接将信息从结合的结合剂上的编码标签转移至邻近的记录标签(图54)。NTAA检测系统可用于测试将聚合物编码标签信息转移至固定在珠粒上的记录标签(参见图1A)。使用DNA/聚合物嵌合体记录标签:单链DNA/聚合物嵌合体构建体,其是5'磷酸化的内部炔修饰的和3'氨基修饰的并且含有DNA间隔区(例如1-6个碱基),聚合物条形码(例如受保护的DNA,PNA,类肽,非DNA聚合物),通用聚合物接头(例如,胆固醇,生物素,系链DNA序列等等),用于USER Enzyme(NEB)消化的dU。DNA/Polymer嵌合体编码标签是5'氨基修饰的,含有通用聚合物衔接子,聚合物条形码(例如保护的DNA,PNA,类肽等)和DNA间隔区(例如1-6个碱基)。这种记录标签和编码标记以及相关结合元件的设计使得凝胶分析变得容易。通过CircLigase II(Lucigen)产生期望的单链DNA连接产物,其中记录标签的5'磷酸基和编码标签的3'羟基经由编码标签上的NTAA与固定在固体表面上的记录标签上的肽的NTAA的结合紧密接近。

[2349] 通过凝胶分析可以评估通过NTAA结合剂编码标签和肽记录标签的NTAA之间的特异性相互作用的信息传递。表面上记录标签的密度可以如下调整。内部炔修饰的DNA/Polymer嵌合体记录标签用甲基四嗪-PEG4-叠氮化物(BroadPharm)修饰。为了控制珠粒表面上记录标签的密度,由M-270胺Dynabeads(Thermo Fisher)制备具有不同密度的功能偶联位点(反式-环辛烯,TCO)的珠粒,M-270胺Dynabeads(Thermo Fisher)通过相对于mPEG-SCM,MW550(Creative PEGWorks)以$1:10^2$,$1:10^3$和$1:10^4$的比率滴定TCO-PEG12-NHS酯(Broadpharm)衍生化。将甲基四嗪修饰的记录标签附着于反式-环辛烯(TCO)衍生珠粒。含有Cys的肽通过SM(PEG)8(Thermo Fisher)附着于珠粒上记录标签的3'胺基。

[2350] 将总共2pmol记录标签与5μl M270链霉亲和素珠粒(Thermo)在50μl固定化缓冲液(5mM Tris-Cl,pH 7.5,0.5mM EDTA,1M NaCl)中于37℃下孵育15分钟,用150μl固定化缓冲

液洗涤一次,并用150μl PBST+30％甲酰胺洗涤一次。对于模型系统,将总计30pmol附着有NTAA结合剂的编码标签与5ul记录标签固定的珠粒在50ul PBST中在37℃下孵育15分钟。在室温下用150μl PBST+30％甲酰胺洗涤珠粒3次。将这些珠粒重悬于10ul CircLigase II反应混合物(0.033M Tris-Acetate,pH 7.5,0.066M乙酸钾,0.5mM DTT,2mM MnCl2,0.5M甜菜碱,和4U/uL CircLigase II ssDNA连接酶)中并且在45℃下孵育2小时。连接反应后,珠粒用固定化缓冲液+30％甲酰胺洗涤一次,用PBST+30％甲酰胺洗涤一次。通过在37℃下5μl的USER Enzyme处理5分钟将最终延伸记录标签从珠粒上洗脱下来,并将2.5μl洗脱液加载到15％ PAGE-尿素凝胶上。连接的记录标签可用于纳米孔测序。

[2351] 示例32:用NTAA检测系统进行双链DNA间隔区连接的信息转移

[2352] 对于DNA/聚合物嵌合体编码标签和记录标签,可通过使用标准核酸酶学的连接将信息从结合的结合剂上的编码标签转移至邻近的记录标签。NTAA检测系统可用于测试将聚合物编码标签信息转移至固定在珠粒上的记录标签(参见图54B)。使用DNA/聚合物嵌合体记录标签:单链DNA/聚合物嵌合体构建体,其是5’磷酸化的内部炔修饰的和3’氨基修饰的并且含有DNA间隔区(例如5-20个碱基),聚合物条形码(例如受保护的DNA,PNA,类肽等),通用聚合物接头(例如,胆固醇,生物素,系链DNA序列,等等),用于USER Enzyme(NEB)消化的dU。DNA/Polymer嵌合体编码标签是5’氨基修饰的,并且含有通用聚合物衔接子,聚合物条形码(例如保护的DNA,PNA,类肽等)和DNA间隔区(例如5-20个碱基)。

[2353] 这种记录标签和编码标签以及相关结合元件的设计使得凝胶分析变得容易。通过夹板DNA杂交和随后的T4 DNA连接酶(NEB)产生期望的单链DNA连接产物,其中记录标签的5’磷酸基和编码标签的3’羟基经由编码标签上的NTAA与固定在固体表面上的记录标签上的肽的NTAA的结合紧密接近。

[2354] 通过凝胶分析可以评估通过NTAA结合剂编码标签和肽记录标签的NTAA之间的特异性相互作用的信息传递。表面上记录标签的密度可以如下调整。内部炔修饰的DNA/Polymer嵌合体记录标签用甲基四嗪-PEG4-叠氮化物(BroadPharm)修饰。为了控制珠粒表面上记录标签的密度,由M-270胺Dynabeads(Thermo Fisher)制备具有不同密度的功能偶联位点(反式-环辛烯,TCO)的珠粒,M-270胺Dynabeads(Thermo Fisher)通过相对于mPEG-SCM,MW550(Creative PEGWorks)以1:102,1:103和1:104的比率滴定TCO-PEG12-NHS酯(Broadpharm)衍生化。将甲基四嗪修饰的记录标签附着于反式-环辛烯(TCO)衍生珠粒。含有Cys的肽通过SM(PEG)8(Thermo Fisher)附着于珠粒上记录标签的3’胺基。

[2355] 将总共2pmol记录标签与5μl M270链霉亲和素珠粒(Thermo)在50μl固定化缓冲液(5mM Tris-Cl,pH 7.5,0.5mM EDTA,1M NaCl)中于37℃下孵育15分钟,用150μl固定化缓冲液洗涤一次,并用150μl PBST+30％甲酰胺洗涤一次。对于模型系统,将总计30pmol附着有NTAA结合剂的编码标签与5ul记录标签固定的珠粒在50ul PBST中在37℃下孵育15分钟。在室温下用150μl PBST+30％甲酰胺洗涤珠粒3次。将珠粒重悬于10μl T4 DNA连接酶反应混合物(50mM Tris-HCl,pH 7.5,10mM MgCl2,1mM DTT,1mM ATP,7.5％ PAG8000,0.1μg/μl BSA,30pmol夹板DNA和20U/μl T4 DNA连接酶)中并在室温下孵育60分钟。连接反应后,用固定化缓冲液+30％甲酰胺洗涤珠粒一次,用0.1N NaOH在室温下洗涤珠粒一次持续5分钟,用PBST+30％甲酰胺洗涤珠粒一次。通过在37℃下5μl的USER Enzyme处理5分钟将最终延伸记录标签从珠粒上洗脱下来,并将2.5μl洗脱液加载到15％ PAGE-尿素凝胶上。连接的记录标

签可用于纳米孔测序。

[2356] 示例33:编码聚合物通过化学连接转移到记录聚合物

[2357] 通过连接将编码聚合物直接或间接地从结合剂-编码聚合物缀合物转移到记录聚合物上,产生延伸的记录聚合物。在一个实施中,该方法使用化学连接方法进行。所使用的编码聚合物可以是线性的或支化的,带电荷的或不带电荷的,疏水的或亲水的。所使用的连接化学可以经由迈克尔加成(硫醇-马来酰亚胺),铜催化的叠氮化物-炔烃环加成,环应变的叠氮化物-炔烃环加成,经由活化酯和胺形成的酰胺键,或本文详述的其他连接化学来进行。聚合物连接步骤可以在每个循环中使用相同的连接反应,在每个其它循环中使用交替的连接反应(以防止环状副产物),或在每个循环中使用不同的连接反应进行。

[2358] 在一个示例中,存在珠粒的溶液,每个珠粒具有待分析的多肽和锚定到其上的炔封端的PEG链(约4000Da,PDI 2.1)记录聚合物。用NTAA特异性结合剂的混合物(80μL的结合剂混合物,磷酸盐缓冲液,pH=8.0)处理珠粒,NTAA特异性结合剂各自经由二异丙基甲硅烷基醚键连接至特异性编码聚合物链的骨架。结合剂可以是anticalins、ClpS或本文指出的其它结合剂衍生物。在一个这样的示例中,结合剂连接的编码聚合物之一可以是用庞大基团(如金刚烷)和带电基团(如羧酸)官能化的线性丙烯酰胺,总分子量为约1000Da(50%丙烯酰胺,20%金刚烷基丙烯酰胺,30%丙烯酸进料比,PDI=2.2)。线性丙烯酰胺编码聚合物在每个末端用叠氮化物封端。在通过结合剂识别NTAA后,使叠氮化物封端的编码聚合物与存在于珠粒上的炔封端的记录聚合物紧密接近。将硫酸铜(II)(10μL的10mM的储备溶液)和抗坏血酸(10μL的500mM的储备溶液)加入到混合物中,编码聚合物通过叠氮化物-炔烃环加成与记录聚合物连接。然后使用氟试剂如四正丁基氟化铵(TBAF,终浓度为1mM)将结合剂从聚合物骨架上裂解下来,得到现在用叠氮化物封端的现在的二嵌段记录聚合物(PEG-丙烯酰胺)缀合物。

[2359] 使用埃德曼降解条件裂解所分析的多肽的NTAA。首先用PITC(10μL的DMF的5M储备溶液在40℃下添加到90μL珠粒溶液中,磷酸盐缓冲液pH=8.5,2h)处理NTAA。或者,首先用PITC(10μL的DMF的5M储备溶液,在40℃下添加到90μL珠粒溶液中,ACN/水(约2:1v/v)处理NTAA。ACN指乙腈。将珠粒洗涤。然后用纯TFA处理PITC封端的NTAA以除去NTAA并产生新的NTAA。

[2360] 在洗涤珠粒之后,再次用一系列结合剂-编码聚合物缀合物(80μL结合剂混合物,磷酸盐缓冲液,pH=8.0)处理新NTAA。在该循环中,聚合物侧链在两端用炔封端。在一种这样的情况下,编码聚合物是聚丙烯酸酯(约5kDa,PDI=1.8)。在通过结合剂识别NTAA后,使炔封端的编码聚丙烯酸酯聚合物与存在于珠粒上的叠氮化物封端的记录聚合物紧密接近。将硫酸铜(II)(10μL的10mM的储备溶液)和抗坏血酸(10μL的500mM的储备溶液)加入到混合物中,编码聚合物通过叠氮化物-炔烃环加成与记录聚合物连接。用缓冲液充分洗涤珠粒。然后使用氟试剂,例如四正丁基氟化铵(TBAF,THF中的终浓度为1mM)将结合剂从聚合物骨架上裂解,得到以炔封端的现在的三嵌段记录聚合物(PEG-丙烯酰胺-丙烯酸酯)。

[2361] 示例34:构建用于纳米孔测序的可测序聚合物

[2362] 记录标签和编码标签可以由一种可测序聚合物和/或编码子单元组成,并且最终的延伸记录标签由一种可测序聚合物组成。对于纳米电子测序例如纳米孔,该可测序聚合物需要具有编码性质,其中编码单元在通过纳米孔或通过纳米间隙或纳米带装置时调节电

学特征(Heerema和Dekker,2016)(参见图56)。编码单元可以由一串不同的基本电流封阻(或调制)部分构成,以创建组合电流封阻特征(参见图56)。组合电流封阻部分可以散布有“停顿”区域,该“停顿”区域使聚合物通过纳米孔的转运的停顿(参见图56A和56B)。或者,如果电子足够快,则聚合物可直接在纳米孔上读出,可能具有封阻部分的多次重复以提供更好的信号(图56C)。

[2363] 示例35:编码的聚合物的纳米孔测序

[2364] 由包埋在脂双层中的M2-NNN MsPa组成的纳米孔用于测量转运聚合物的封阻电流(Laszlo,Derrington等人,2014)(参见图57)。总之,通过将1,2-二植烷酰基-sn-丙三醇-3-磷酸胆碱脂质(Avanti极性脂质)跨越含有我们的操作缓冲液的两个约60μL室的水平约20μm直径孔涂抹而产生脂质双层。Axopatch 200B集成膜片钳增强器(Axon Instruments)在双层上施加180mV电压(反式侧正(trans side positive)),并测量通过孔的离子电流。用4极贝塞尔滤波器在10kHz下对模拟信号进行低通滤波,并在低通滤波频率的5倍下对其进行数字化。在顺式和反式侧都使用盐溶液(1M KCl溶于10mM HEPES/KOH(pH 8.0)),并且在室温(23±1℃)下进行实验。将带有漂白剂的新鲜银丝制成的氯化银电极插入顺式和反式样品池中,用于离子电流测量。将生物蛋白质纳米孔M2-NNN MspA加入研磨的顺式隔室至约2.5ng/ml的终浓度。一旦单个孔插入,如电导的特征增加所示,用无MspA的缓冲液替换缓冲液以防止额外的孔形成。将附有胆固醇衔接子的聚合物(w/生物素部分)加入顺式隔室至1nM的终浓度。将单体链霉亲和素(10μM,Kd约2-3nM)以100nM加入到顺式侧(Demante,Drake等人,2013)。胆固醇衔接子通过将胆固醇分配到膜中而将聚合物捕获在膜表面。带负电荷的聚合物被捕获在孔中并发生易位。聚合物易位通过纳米孔,直到第一生物素-mSA2复合物到达孔入口(参见图58)。在180mV的电压下,聚合物在孔中停顿,直到由于施加的力的能量使得scSA解离,使得平均停顿时间在毫秒值的量级上。

[2365] 示例36:多循环结合/编码测定

[2366] 肽模型系统用于测试编码的三个循环,其中将来自编码标签的信息转移到固定在珠粒上的记录标签上(图59)。将PA肽序列(SEQ ID NO:195)附着到固定在珠粒上的记录标签寡核苷酸amRT_Abc(SEQ ID NO:214)上。amRT_Bbc序列含有BC_3(SEQ ID NOs:66-130)。对于结合剂,将抗-PA抗体(Wako Chemical)附着至编码标签寡核苷酸amCT_bc4、amCT_bc5和amCT_bc13r1(SEQ ID NOs:209-211),其分别包含15-mer条形码BC4、BC5和BC13R1(SEQ ID NOs:66-130)。该设计便于用特异性引物组进行PCR扩增后的凝胶分析。

[2367] 磁性测定珠粒生成的两步法。

[2368] DNA记录标签-肽嵌合体用两步固定法以受控密度固定在磁性珠粒上,该两步固定法类似于图34中描述的一步测定珠粒法,除了偶联记录标签后固定的肽而不是偶联嵌合体作为单个单元。即,通过用mPEG-SCM,MW550(Creative PEGWorks)以1:103的比率滴定TCO-PEG12-NHS酯(BroadPharm),M-270胺Dynabeads珠粒(Thermo Fisher)用反式环辛烯(TCO)基团稀疏地表面官能化。过量的胺基用NHS-乙酸酯封端。在两步嵌合体固定方法中,通过首先将内部炔寡核苷酸修饰物转化为mTet,然后直接偶联到TCO珠粒上,然后将肽偶联到寡聚记录标签的5’胺上,将由内部炔修饰物和5’胺修饰物组成的双功能DNA记录标签固定到TCO珠粒上。具体地,内部炔修饰的记录标签寡核苷酸用mTet-PEG4-叠氮化物(BroadPharm)修饰,然后附着于TCO衍生的珠粒。为了将肽偶联至珠粒上的5’胺记录标签,将含半胱氨酸的

肽经由NHS-PEG8-Mal(Thermo Fisher)附着至固定化记录标签的5’胺基。

[2369] 编码标记的PA抗体的产生

[2370] 使用蛋白质-寡核苷酸缀合试剂盒(Solulink)完成抗PA抗体与编码标签(amCT_bc4,或amCT_bc5,或amCT_bc13r1寡核苷酸)的缀合。简单地说,编码标签的5’胺基用S-4FB修饰,然后用0.5mL Zeba柱脱盐。抗PA抗体经S-HyNic修饰后,用0.5mL的Zeba柱脱盐。最后将4FB修饰的编码标签与HyNic修饰的抗PA抗体混合,制备抗体-编码标签偶联物,然后用0.5mL Zeba柱脱盐。

[2371] 基于抗体的结合/编码测定:

[2372] PA肽-记录标签amRT_Bbc(SEQ ID NO:214)如上所述制备珠粒。对于第一循环结合测定,通过在室温(室温)下与SuperBlock T20(TBS)封闭缓冲液(Thermo Fisher)孵育15分钟,最初封闭一百万个珠粒。将40nM的编码标签(BC4)标记的PA抗体与测定珠粒在50ul的PBST(1xPBS,0.1%吐温20)中37℃下孵育30分钟。在室温下用1ml含有10%甲酰胺的PBST洗涤珠粒一次。将珠粒重悬于50μl延伸反应主混合物(50mM Tris-Cl(pH 7.5),2mM $MgSO_4$,125μM dNTPs,50mM NaCl,1mM二硫苏糖醇,0.1%吐温20,0.1mg/ml BSA和0.05U/μL Klenow exo-DNA聚合酶)中并在37℃下孵育5分钟。引物延伸后,将珠粒用150μl PBST/10%甲酰胺洗涤(室温下)一次,用150μl 0.1N氢氧化钠洗涤5分钟一次,用150μl PBST/10%甲酰胺洗涤一次,用50μl PBST洗涤一次。对于第二和第三结合循环测定(参见图59),在结合/编码测定中分别使用编码标签(BC5)未标记的PA抗体和编码标签(BC13)未标记的PA抗体。在第三编码循环的延伸后,用条形码特异性引物在20μl PCR混合物中PCR扩增最终延伸的记录标签,并在10% PAGE凝胶上分析1μl PCR产物。得到的凝胶证明了通过聚合酶延伸将编码标签信息的三个循环结合和写到记录标签的原理的证据(参见图5B)。除了凝胶分析外,用引物扩增经第三结合测定后产生的延伸记录标签,以产生NGS测序库。将所得库在Illumina MiSeq仪器上测序,结果示于图59C中。超过78%的延伸记录标签含有3个条形码,表明相对有效的逐步编码。

[2373] 示例37:采用测定封闭寡核苷酸的增强编码

[2374] (也参见图3和图36A)

[2375] 在三循环编码测定中使用不同的DNA阻断剂评估结合和编码的寡核苷酸模型系统,以评估对编码效率的影响(图60)。记录标签寡核苷酸,amRT_Bbc(SEQ ID NO:214)如前所述固定在珠粒上。该记录标签序列含有“B”DNA捕获序列(“B”结合剂的模拟表位)和BC_3条形码(SEQ ID NOs:1-65)。在该模型系统中,结合剂编码标签使用具有附加到DNA编码标签的B互补序列(B’)的寡核苷酸。测试了三种类型的DNA阻断剂(图6A)。CT阻断剂与编码标签5’Sp’区和条形码序列杂交。3’Sp’序列对于退火是自由的。在记录标签上,Sp’阻断剂和RT阻断剂分别与Sp间隔区序列和记录标签上的Sp条形码杂交。编码后,延伸记录标签进行PCR扩增以及PAGE分析。如图60B所示,在CT阻断剂存在下,而不是在单独Sp’或RT阻断剂存在下,在来自3个循环的PCR产物中观察到第三循环编码的带。数据表明CT阻断剂有助于将编码标签信息有效转移到记录标签。

[2376] 示例38:使延伸记录标签扩增期间的模板切换最小化。

[2377] 使用两个延伸记录标签寡核苷酸模型模板TS_Ctrl1(SEQ ID NO:212)和TS_Ctrl4(SEQ ID NO:213)(图61A)。这些模型模板被设计成共享共同的Sp和条形码序列以模拟延伸

记录标签模板结构。将图61A所示的模型模板以1:1的比例混合,并在不同温度下用不同的聚合酶通过PCR进行扩增。通过使用10% PAGE凝胶分析PCR产物并通过NGS读出(未显示)来评估模板转换速率。

[2378] 详细地,使用如图61所示的Fwd和Rev引物在10μl PCR反应中扩增上述两个模板的10fmol的1:1混合物,并通过10% PAGE凝胶进行分析,并且还通过NGS使用MiSeq DNA测序仪(Illumina)进行分析。在50-68℃退火温度窗口测定以下聚合酶的模板转换速率:Q5热启动(New England Biolabs)、Taq(New England Biolabs)、Deep Vent(exo-)(New England Biolabs)、KOD Xtreme热启动(EMD Millipore)和Herculase II融合DNA聚合酶(Agilent)。如图61所示,观察到对应于来自TS_Ctrl1和TS_Ctrl4的PCR产物的两个主要条带(125bp和173bp)。在许多聚合酶中观察到模板转换副产物带,包括Deep Vent exo-如凝胶上的较低分子量带所示(参见图6B),但是用Taq DNA聚合酶(参见图6B)和KOD Xtreme热启动DNA聚合酶(数据未显示),并且这些聚合酶随后通过NGS更详细地分析,显示5.6%和2.4%的映射读取部分,其分别显示用Taq和KOD Xtreme热启动DNA聚合酶的模板转换。

[2379] 示例39:全循环ProteoCode肽测序测定的演示。

[2380] 图62描述了单周期ProteoCodeNGPS测定的演示。通过我们的N-末端官能化和N-末端消除(NTF/NTE)试剂移除肽的N-末端氨基酸,这通过F-结合剂结合测定的两个循环来检测(参见图62)。具有内部PA肽序列(SEQ ID NOs:201,202,和203)的N-末端FA、AF以及A肽分别附着到使用示例1中描述的两步法固定在珠粒上的记录标签寡核苷酸amRT_Cbc(SEQ ID NO:215)。

[2381] F-结合剂与编码标签amCT_bc4和amCT_bc5(SEQ ID NOs:209,210)使用如前所述的SpyTag/SpyCatcher蛋白质偶联方法完成。简单地说,融合至SpyCatcher的F-结合剂表示为大肠杆菌,通过Ni-NTA柱纯化,并在PBS中透析。将含有Cys的肽SpyTag(SEQ ID NO:198)经由SM(PEG)24(Thermo Fisher)附接至amCT_bc4和amCT_bc5的5'胺。最后将SpyTag修饰的编码标签与SpyCatcher融合的F-结合剂混合,制备F-结合剂编码标签缀合物。

[2382] 进行两个循环的F结合和编码,前NTF/NTE化学和后NTF/NTE化学。如示例1中所述进行结合和编码测定。第一循环F-结合剂结合/编码测定后,用NTF/NTE试剂处理测定珠粒以除去NTAA。对于NTF处理,将测定珠粒与150ul的溶液(将15mM N-Boc-N'-TFA-吡唑-1-甲脒(Sigma)溶于pH 8.5的0.5M三乙基乙酸铵/乙腈(1:1),0.05% F-127)在50℃下孵育1小时。用200μl的0.5μM三乙基乙酸铵/乙腈(1:1v/v,0.05% F-127)洗涤珠粒。在40℃下,用150μl含有0.05%吐温20的0.5M NaOH孵育测定珠粒1小时,进行NTE处理。用1ml含10%甲酰胺的PBST洗涤珠粒,并用F-结合剂-编码标签进行第二循环F-结合剂结合测定。对于化学前与化学后结合/编码循环,将F-结合剂与不同的循环特异性条形码编码标签缀合。通过NGS分析该测定的延伸记录标签。NGS结果显示,F-结合剂在第一循环检测到FA肽,而AF肽的检测程度最低,相反,F-结合剂在NTF/NTE处理后第二循环检测到AF肽的效率较高,而FA肽的检测较差。总之,在化学之前和之后在AF肽记录标签上检测的F-结合剂编码增加了大约4倍,而在化学之前和之后在FA肽上编码的F-结合剂编码减少了大约4倍,这有效证明了使用DNA编码的原理性单循环肽测序的证据。

[2383] 示例40:串扰测量

[2384] 图36示出了使用DNA模型系统嵌合体密度对珠粒和分子间串扰的影响。利用mTet

衍生的M270-Dynal珠粒上滴定肽-RT嵌合体,建立了4-plex肽-记录标签(RT)串扰模型体系。为了控制肽-RT嵌合体在珠粒表面上的密度,从活化的M-270Carboxyl Dynabeads(Thermo Fisher)制备具有功能性偶联位点的珠粒(甲基四嗪,mTet),活化的M-270Carboxyl Dynabeads(Thermo Fisher)通过以1:100,1:1000,1:10000和1:100000的比率针对m-PEG4-胺(BroadPharm)滴定NH2-PEG4-甲基四嗪(BroadPharm)而衍生化。通过TCO/mTet偶联,将4-plex、无肽、FA、AA和AF肽附着的记录标签以1:100,1:1000,1:10000和1:100000的比率固定在珠粒上。

[2385] 首先通过偶联到TCO-PEG12-NHS酯上“激活”RT寡核苷酸上的5’胺产生一组肽-RT嵌合体(点击化学工具)。在TCO活化后,设计具有内部炔基的RT寡核苷酸与含有叠氮化物的FA、AA和AF肽偶联。将具有N-末端FA、AA和AF氨基酸序列和内部PA表位(SEQ ID NO:195)的肽分别单独附着到记录标签寡核苷酸amRT_Cs2,amRT_Cs4和amRT_Cs5(SEQ ID NOs:217-219)。第四记录标签amRT_Cs1(SEQ ID NOs:216)作为无肽对照。将F-binder结合剂缀合至编码标签寡核苷酸amCT_s7(SEQ ID NO:220)由8-mer条形码BC_s7(SEQ ID NO:221)组成。使用iEDDA TCO-mTet化学将四种嵌合体(FA肽、AA肽、AF肽和无肽)组合并固定至mTet珠粒。如实施例4所述制备F-结合剂编码标签构建体。

[2386] 通过在4-plex测定珠粒上的一个循环F-结合/编码测定来检测记录标签之间的串扰测定信号(参见图63)。在示例1中描述了测定条件,除了使用200nM F-binder结合剂。通过与通用PA抗体结合/编码的第一循环测定4种不同嵌合体的绝对负载,其中所有3种肽类型均含有PA抗原序列。图63A显示所有四种嵌合体以大致相等的量装载在珠粒上。通过在N-末端F-肽记录标签上的F-结合剂编码鉴定特异性编码,如通过延伸记录标签中条形码的NGS读出所确定的。通过在N-末端非-F(AA肽,AF肽,和无肽)肽记录标签上编码的F-结合剂来识别串扰。如图63B所示,所有嵌合体记录标签在1:100和1:1000活性位点珠粒上接收F-结合条形码,而只有FA记录标签在1:10000和1:100000珠粒上接收F-结合剂条形码。结果表明,在高记录标签密度的珠粒(1:100和1:1000)上进行分子间和分子内编码,在较低密度的珠粒(1:10000和1:100000)上进行分子内编码。总之,该模型系统证明了分子内的单分子结合和在稀少填充的珠粒上的编码。注意,本示例的密度滴定与示例例36的DNA模型系统不直接相关,因为使用了不同的珠粒化学。

[2387] 示例41:用碱基保护的DNA编码

[2388] 我们NTF/NTE埃德曼类化学的挑战是,除了修饰肽的N-末端胺之外,高反应性NTF试剂还可以修饰DNA核碱基上的环外和其他胺基。DNA修饰可能损害测定中DNA的编码功能。使用核碱基保护的DNA可以大大降低DNA的修饰。该方法允许使用更广泛的一组化学反应和反应条件,促进NTAA基团的快速改性。

[2389] 使用核碱基保护的DNA实施上述方法的一个挑战是大多数核碱基保护基降低DNA形成杂交体的能力,即,引物延伸和许多连接编码测定的有用先决条件。

[2390] 令人惊奇的是,在DNA编码过程中使用受保护的ssDNA能够通过使用ssDNA连接酶CircLigase将完全受保护的寡核苷酸连接在一起。我们测试了CircLigase将完全保护的寡核苷酸与天然“T”寡核苷酸间隔区序列在连接接合处的任一侧(3’或5’)连接的能力。我们发现,在5’侧没有T间隔区的情况下,使用紧邻连接接合处的受保护的G碱基,可以完成有效的连接。在3’侧使用天然T间隔区。

[2391] 寡核苷酸模型系统用于测试编码标签信息向固定在珠粒上的记录标签复合物的第一循环转移(图64)。使用链霉亲和素官能化珠粒上的强生物素-链霉亲和素相互作用将PRT_OT_B'_bio固定在珠粒上。PRT_OT_B'_bio序列包含"B'"DNA捕获序列("B"结合剂的模拟表位)和相应的由受保护的dA、dG和dC组成的"B'"条形码。对于结合剂,互补的"B"结合序列CT_B_thio(SEQ ID NO:206)通过异双功能接头耦联到由受保护的dA、dG和dC组成的编码标签寡核苷酸PCT_1T(SEQ ID NO:207)。利用CircLigaseII通过单链连接编码特异性B-seq/B'seq结合事件,并通过凝胶分析证实。

[2392] 含硫醇的"B'"序列RT_thio_B'_bio(SEQ ID NO:204)经由SM(PEG)8(Thermo Fisher)附着至PRT_OT(dU)的5'胺基(SEQ ID NO:205),并用15%的变性PAGE对所得到的标签PRT_OT_B'bio进行纯化。将记录标签固定在M-270链霉亲和素Dynabeads(Thermo Fisher)上,链霉亲和素Dynabeads(Thermo Fisher)通过针对mPEG-生物素MW1000(Creative PEGWorks)以1:10的比率滴定而衍生化。用USER Enzyme(New England Biolabs)处理珠粒以在记录标签-5'磷酸G的5'端产生磷酸化位点。经由SM(PEG)8(Thermo Fisher)连接"B'"结合序列的缀合物B_thio与PCT_1T,并通过15%变性PAGE来纯化编码g标签。

[2393] 对于第一循环,结合测定与先前描述的类似,除了通过溶于50ul PBST的5'磷酸化记录标签固定的珠粒将800nM编码标签缀合物在37℃孵育30分钟。用50μl含有40%甲酰胺的PBST洗涤珠粒两次,用50μl的33mM Tris-HCl缓冲液(pH7.5)在室温下洗涤一次。将珠粒重悬于50ul CircLigase II连接反应主混合物(33mM Tris-乙酸盐,pH 7.5,66mM乙酸钾,0.5mM DTT,2.5mM $MnCl_2$,1M甜菜碱和0.25U CircLigase II ssDNA连接酶(Illumina)),并且在37℃下孵育30分钟。连接反应后,用固定化缓冲液(5mM Tris-Cl(pH 7.50),0.5mM EDTA,1M NaCl,40%甲酰胺)洗涤珠粒一次,用含有40%甲酰胺的PBST洗涤一次。通过在65℃在95%甲酰胺/10mM EDTA中孵育5分钟,将记录标签从珠粒上洗脱下来。通过PAGE凝胶分析洗脱的缀合物。得到的凝胶,在图64的道4中条带有位移,证明了使用单链DNA连接将编码标签信息写入包含受保护的dA、dC和dG碱基的记录标签的原理的证据(参见图64)。

[2394] 示例42:使用与DNA珠粒的杂交和连接来固定DNA标记的肽。

[2395] 肽-记录标签DNA嵌合体在珠粒上的有效固定是ProteoCode测定的重要组成部分,并且可以通过化学方法完成,但也可以通过杂交捕获和酶促或化学连接完成。我们发现基于杂交的捕获比化学偶联更有效约10000倍。在此我们证明了将DNA标记的肽杂交和酶连接固定化在DNA捕获珠粒上的用途。

[2396] 将DNA-肽嵌合体杂交并连接到化学固定在磁性珠粒上的发夹捕获DNA上(参见图65)。使用基于反式-环辛烯(TCO)和甲基四嗪(mTet)的点击化学将捕获DNA缀合至珠粒。TCO修饰的短发夹DNA(16个碱基对茎,24个碱基5'突出)与mTet包被的磁性珠粒反应。在五次SSC,0.02% SDS中将磷酸化的DNA-肽嵌合体(1nM)退火至发夹DNA珠粒,并且在37℃孵育30分钟。将珠粒用PBST洗涤一次,并重悬于含有和不含T4 DNA连接酶的1x快速连接溶液(NEB)中。在25℃孵育30分钟后,用PBST洗涤珠粒一次并重悬于20μl PBST中。使用特异性引物组通过qPCR对总捕获DNA、DNA标记肽和连接的DNA标记肽进行定量(图65D)。如图65E所示,在qPCR中仅在连接酶的存在下用珠粒/DNA肽杂合体获得低Ct值,表明DNA标记的肽被有效连接并固定到珠粒上。

[2397] 表5.基于肽和基于DNA的模型系统序列

[2398]

| 名称 | 序列（5’-3’） | SEQ ID NO: |
|---|---|---|
| amRT_Abc | /5AmMC6/GCAAATGGCATTCTGACATCCTT/i5OctdU/TTCGUAGUCCGCGACACTAGATGTCTAGCATGCCGCCGTGTCATGTGGAAACTGAGTG | 190 |
| amCT_bc5 | /5AmMC6//iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/ | 191 |
| CT_A’_bc13 | GGATGTCAGAATGCCATTTGCTTTTTTTTTT/iSP18/CGATTTGCAAGGATCACTCGCCGTTATTGACGCTCTCACTCAGT/3SpC3/ | 192 |
| Sp | ACTGAGTG | 149 |
| Sp’ | CACTCAGT | 150 |
| P1_f2 | CGTAGTCCGCGACACTAG | 151 |
| R1 | CGATTTGCAAGGATCACTCG | 152 |
| Sp/BC2 | CACTCAGTTTCCACATGACACGGC | 193 |
| Sp/BC5 | CACTCAGTCCTAACGCGTATA | 194 |
| PA 肽 | GVAMPGAEDDVVGGGGSC | 195 |
| 纳米标签肽 | 甲酰基-MDVEAWLGARVPLVETGSGSGSC | 196 |
| Aβ 肽 | HQKLVFFAEDVGSGSGSC | 197 |
| Spytag | CGSGSGSAHIVMVDAYKPTK | 198 |
| FA-PA 肽 | FAGVAMPGAEDDVVGSGSK（叠氮化物） | 201 |
| A-PA 肽 | AGVAMPGAEDDVVGSGSK（叠氮化物） | 202 |
| AFA-PA 肽 | AFAGVAMPGAEDDVVGSGSK（叠氮化物） | 203 |
| RT_thio_B’_bio | /5ThioMC6-D/GCTTGTCAATTGGAACCAGTCT/3Bio/ | 204 |
| PRT_0T（dU） | （5’OH）TTTTTTT（dU）G*AC*A*TC*TA*G*TG*TC*G*C*G*G*AC*TA*C*G*（氨基己基-dT）-3ProtP | 205 |
| CT_B_thio | GCTTGTCAATTGGAACCAGTC/3ThioMC3-D/ | 206 |
| PCT_1T | 5ProtP-（氨基己基-dT）C*G*A*TTTG*C*A*A*G*G*A*TC*AC*TC*G*A*G*G*T（3’OH） | 207 |
| amRT_Bb | /5ThioMC6-D/GCTTGTCAATTGGAACCAGTCT/3Bio/ | 208 |

[2399]

[2400]

| | | |
|---|---|---|
| amCT_bc4 | /5AmMC6//iSP18/CACTCAGTTTCCGTCATATCGAATCACTCAGT/3SpC3/ | 209 |
| amCT_bc5 | /5AmMC6//iSP18/CACTCAGTCCTAACGCGTATACGTCACTCAGT/3SpC3/ | 210 |
| amCT_bc13r1 | /5AmMC6//iSP18/CGATTTGCAAGGATCACTCGCCGTTATTGACGCTCTCACTCAGT/3SpC3/ | 211 |
| TS_Ctrl1 | CGTAGTCCGCGACACTAGAATAAGCCGGTATATCAACTGAGTGATTCGATATGACGGAAACTGAGTGACGTATACGCGTTAGGACCTGAGTGAGAGCGTCAATAACGGCGAGTGATCCTTGCAAATCGCCCTATAGTGAGTCGTATTAATTCGC | 212 |
| TS_Ctrl4 | CGTAGTCCGCGACACTAGCGTATACGCGTTAGGACTGAGTGAGAGCGTCAATAACGGACTGAGTGAGAGTCGGTACCTTGAACTGAGTGAAACTGCCGAGATTCCACTGAGTGATTCGATATGACGGAAACTGAGTGATAAGCCGGTATATCACGAGTGATCCTTGCAAATCGCCCTATAGTGAGTCGTATTAATTCGC | 213 |
| amRT_Bbc | /5AmMC6/GACTGGTTCCAATTGACAAGCTT/i5OctdU/TTTTTTUCGTAGTCCGCGACACTAGTAAGCCGGTATATCAACTGAGTG | 214 |
| amRT_Cbc | /5AmMC6/TTT/i5OctdU/TTUCGTAGTCCGCGACACTAGTAAGCCGGTATATCAACTGAGTG | 215 |
| amRT_Cs1 | /5AmMC6/TTT/i5OctdU/TTUCGTAGTCCGCGACACTAGNNNNNNNNNNTTAAGTCGACTGAGTG | 216 |
| amRT_Cs2 | /hanAmMC6/TTT/ihanOctdU/TTUCGTAGTCCGCACACTAGNNNNNNGTTAATGGACTGAGTG | 217 |
| amRT_Cs4 | /5AmMC6/TTT/i5OctdU/TTUCGTAGTCCGCGACACTAGNNNNNNNNNNCAGTACCGACTGAGTG | 218 |
| amRT_Cs5 | /5AmMC6/TTT/i5OctdU/TTUCGTAGTCCGCGACACTAGNNNNNNNNNNGTTGGTTAACTGAGTG | 219 |
| amCT_s7 | /5AmMC6//iSP18/CACTCAGTCAGACTATTCACTCAGT/3SpC3/ | 220 |
| BC_s7 | ATAGTCTG | 221 |

[2401] /3SpC3/＝3’C3(三碳)间隔区

[2402] /i5OctdU/＝5’-辛二炔基dU

[2403] /iSP18/＝18-原子已-乙二醇间隔区

[2404] 5ProtP＝5’氰乙基保护的磷酸盐

[2405] 3ProtP＝3’氰乙基保护的磷酸盐

[2406] *＝受保护的碱基(Bz-dA,ibu-dG,Ac-dC)

[2407] VI.进一步的示例性实施例

[2408] 以下实施例提供本公开的额外说明。

[2409] 1A.一种试剂盒,包含:

[2410] (a)记录聚合物,其被配置为与分析物直接或间接相关;

[2411] (b)(i)编码聚合物,其包含关于能够与所述分析物结合的结合基团的识别信息,并且被配置为与所述结合基团直接或间接相关以形成结合剂,和/或(ii)标签,

[2412] 其中所述记录聚合物和所述编码聚合物被配置为允许在所述结合剂和所述分析物之间结合时在它们之间转移信息;和

[2413] 任选地(c)所述结合基团。

[2414] 2A.根据实施例1A所述的试剂盒,其中记录聚合物和/或分析物被配置为直接或间接固定到支持物上。

[2415] 3A.根据实施例A所述的试剂盒,其中记录聚合物被配置为固定到支持物上,从而固定与记录聚合物相关的分析物。

[2416] 4A.根据实施例2A所述的试剂盒,其中分析物被配置为固定至支持物,从而固定与分析物相关的记录聚合物。

[2417] 5A.根据实施例2A所述的试剂盒,其中记录聚合物和分析物中的每一个被配置为固定到支持物上。

[2418] 6A.根据实施例5A所述的试剂盒,其中记录聚合物和分析物被配置为当两者都固定到支持物上时共定位。

[2419] 7A.根据实施例1A-6A中任一个所述的试剂盒,其还包含固定接头,固定接头被配置为:

[2420] (i)直接或间接固定到支持物上,和

[2421] (ii)直接或间接地与记录聚合物和/或分析物相关。

[2422] 8A.根据实施例7A所述的试剂盒,其中固定接头被配置为与记录聚合物和分析物相关。

[2423] 9A.根据实施例7A或8A所述的试剂盒,其中固定连接物被配置为直接固定至支持物,从而固定与固定接头相关的记录聚合物和/或分析物。

[2424] 10A.根据实施例2A-9A中任一个所述的试剂盒,其进一步包含支持物。

[2425] 11A.根据实施例1A-10A中任一个所述的试剂盒,进一步包含一种或多种试剂,用于在结合剂与分析物之间结合时在编码聚合物与记录聚合物之间转移信息。

[2426] 12A.根据实施例11A所述的试剂盒,其中一种或多种试剂被配置为将信息从编码聚合物转移至记录聚合物,从而产生延伸记录聚合物。

[2427] 13A.根据实施例11A所述的试剂盒,其中一种或多种试剂被配置为将信息从记录聚合物转移至编码聚合物,从而产生延伸编码聚合物。

[2428] 14A.根据实施例11A所述的试剂盒,其中一种或多种试剂被配置为产生包含来自编码聚合物的信息和来自记录聚合物的信息的双标签构建体。

[2429] 15A.根据实施例1A-14A中任一个所述的试剂盒,其包含记录聚合物中的至少两种。

[2430] 16A.根据实施例1A-15A中任一个所述的试剂盒,其包含编码聚合物中的至少两种,每种编码聚合物包含关于其相关结合基团的识别信息。

[2431] 17A.根据实施例1A-16A中任一个所述的试剂盒,其包含结合剂中的至少两种。

[2432] 18A.根据实施例17A所述的试剂盒,其包含:

[2433] (i)一种或多种试剂,其用于在第一结合剂和分析物之间结合时,将信息从第一结

合剂的第一编码聚合物转移到记录聚合物,以产生第一次序延伸记录聚合物,和/或

[2434] (ii)一种或多种试剂,用于在第二结合剂和分析物之间结合时,将信息从第二结合剂的第二编码聚合物转移到第一次序延伸记录聚合物,以产生第二次序延伸记录聚合物,

[2435] 其中(i)的一种或多种试剂和(ii)的一种或多种试剂可以相同或不同。

[2436] 19A.根据实施例18A所述的试剂盒,其进一步包含:

[2437] (iii)一种或多种试剂,用于在第三(或更高次序)结合剂与分析物之间结合时,将信息从第三(或更高次序)结合剂的第三(或更高次序)编码标签转移至第二次序延伸记录标签,以产生第三(或更高次序)次序延伸记录标签。

[2438] 20A.根据实施例17A所述的试剂盒,其包含:

[2439] (i)一种或多种试剂,用于在第一结合剂和分析物之间结合时,将信息从第一结合剂的第一编码聚合物转移到第一记录聚合物,以产生第一延伸记录聚合物,和/或

[2440] (ii)一种或多种试剂,用于在第二结合剂和分析物之间结合时,将信息从第二结合剂的第二编码聚合物转移到第二记录聚合物,以产生第二延伸记录聚合物,

[2441] 其中(i)的一种或多种试剂和(ii)的一种或多种试剂可以相同或不同。

[2442] 21A.根据实施例20A所述的试剂盒,其进一步包含:

[2443] (iii)一种或多种试剂,其用于在第三(或更高次序)结合剂和分析物之间结合时,将信息从第三(或更高次序)结合剂的第三(或更高次序)编码聚合物转移到第三(或更高次序)记录聚合物,以产生第三(或更高次序)延伸记录聚合物。

[2444] 22A.根据实施例20A或21A所述的试剂盒,其中第一记录聚合物、第二记录聚合物和/或第三(或更高次序)记录聚合物被配置为与分析物直接或间接相关。

[2445] 23A.根据实施例20A-22A中任一个所述的试剂盒,其中第一记录聚合物、第二记录聚合物和/或第三(或更高次序)记录聚合物被配置为固定在支持物上。

[2446] 24A.根据实施例20A-23A中任一个所述的试剂盒,其中第一记录聚合物、第二记录聚合物和/或第三(或更高次序)记录聚合物可以被配置为与分析物共定位,例如,以在第一、第二或第三(或更高次序)结合剂和分析物之间结合时允许分别在第一、第二或第三(或更高次序)编码聚合物与第一、第二或第三(或更高次序)记录聚合物之间转移信息。

[2447] 25A.根据实施例20A-24A中任一个所述的试剂盒,其中第一编码聚合物、第二编码聚合物和/或第三(或更高次序)编码聚合物中的每一个包含结合环特异性条形码,例如结合环特异性间隔区聚合物序列$C_n$和/或编码聚合物特异性间隔区聚合物序列$C_n$,其中n是整数,并且$C_n$表示第n种结合剂与多肽之间的结合;或者其中结合循环标签$C_n$是外源添加的,例如,结合循环标签$C_n$对于编码聚合物可以是外源的。

[2448] 26A.根据实施例1A-25A中任一个所述的试剂盒,其中分析物包含多肽。

[2449] 27A.根据实施例26A所述的试剂盒,结合基团能够结合多肽的一个或多个N-末端或C-末端氨基酸,或能够结合通过官能化试剂修饰的一个或多个N-末端或C-末端氨基酸。

[2450] 28A.根据实施例26A或27A所述的试剂盒,其进一步包含官能化试剂。

[2451] 29A.根据实施例26-28中任一个所述的试剂盒,其进一步包含包含用于移除(例如,通过化学裂解或酶裂解)多肽的一个或多个N-末端、内部或C-末端氨基酸,或移除官能化的N-末端、内部或C-末端氨基酸的消除试剂,任选地,其中消除试剂包含羧肽酶或氨肽酶

或其变体,突变体或修饰的蛋白质;水解酶或其变体,突变体或其修饰的蛋白质;温和型埃德曼降解试剂;Edmanase酶;无水TFA,碱;或其任何组合。

[2452] 30A.根据实施例26A-29A中任一个所述的试剂盒,其中一个或多个N-末端、内部或C-末端氨基酸包含:

[2453] (i)N-末端氨基酸(NTAA);

[2454] (ii)N-末端二肽序列;

[2455] (iii)N-末端三肽序列;

[2456] (iv)内部氨基酸;

[2457] (v)内部二肽序列;

[2458] (vi)内部三肽序列;

[2459] (vii)C-末端氨基酸(CTAA);

[2460] (viii)C-末端二肽序列;或

[2461] (ix)C-末端三肽序列,

[2462] 或其任何组合,

[2463] 任选地其中(i)-(ix)中的任何一个或多个氨基酸残基被修饰或官能化。

[2464] 31A.一种试剂盒,包含:

[2465] 至少(a)第一结合剂,其包含(i)能够结合待分析多肽的N末端氨基酸(NTAA)或官能化NTAA的第一结合基团,和(ii)包含关于所述第一结合基团的识别信息的第一编码聚合物,

[2466] 任选地(b)记录聚合物,其被配置为与所述多肽直接或间接相关,以及

[2467] 进一步任选地(c)官能化试剂,其能够修饰所述多肽的第一NTAA以产生第一官能化NTAA,

[2468] 其中所述记录聚合物和所述第一结合剂被配置为允许在所述第一结合剂和所述多肽之间结合时在所述第一编码聚合物和所述记录聚合物之间转移信息。

[2469] 32A.根据实施例31A所述的试剂盒,其还包含一种或多种试剂,用于将信息从第一编码聚合物转移到记录聚合物,从而产生第一次序延伸记录聚合物。

[2470] 33A.根据实施例31A或32A所述的试剂盒,其中官能化试剂包含化学剂、酶和/或生物剂,如异硫氰酸酯衍生物,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。

[2471] 34A.根据实施例31A-33A所述的试剂盒,进一步包含消除试剂,用于移除(例如,通过化学裂解或酶裂解)第一官能化NTAA,以暴露紧邻的氨基酸残基(如第二NTAA)。

[2472] 35A.根据实施例34A所述的试剂盒,其中第二NTAA能够通过相同或不同的官能化试剂官能化以产生第二官能化NTAA,其可与第一官能化NTAA相同或不同。

[2473] 36A.根据实施例35A所述的试剂盒,其进一步包含:

[2474] (d)第二(或更高次序)结合剂,其包含(i)能够结合到第二官能化NTAA的第二(或更高次序)结合基团,和(ii)第二(或更高次序)编码聚合物,其包含关于第二(或更高次序)结合基团的识别信息,

[2475] 其中第一编码聚合物和第二(或更高次序)编码聚合物可以相同或不同。

[2476] 37A.根据实施例36A所述的试剂盒,其中第一官能化NTAA和第二官能化NTAA彼此独立地选自由以下组成的群组:官能化N-末端丙氨酸(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),缬氨酸(V或Val),色氨酸(W或Trp),酪氨酸(Y或Tyr),其任何组合。

[2477] 38A.根据实施例36A或37A所述的试剂盒,其进一步包含一种或多种试剂,用于将信息从第二(或更高次序)编码聚合物转移到第一次序延伸记录聚合物,从而产生第二(或更高次序)延伸记录聚合物。

[2478] 39A.一种试剂盒,包含:

[2479] 至少(a)一种或多种结合剂,其各自包含(i)能够结合待分析多肽的N末端氨基酸(NTAA)或官能化NTAA的结合基团,和(ii)包含关于所述结合基团的识别信息的编码聚合物,和/或

[2480] (b)一种或多种记录聚合物,其被配置为与所述多肽直接或间接相关,

[2481] 其中所述一种或多种记录聚合物和所述一种或多种结合剂被配置为允许在每种结合剂和所述多肽之间结合时在所述编码聚合物和所述记录聚合物之间转移信息,并且

[2482] 任选地(c)官能化试剂,其能够修饰所述多肽的第一NTAA以产生第一官能化NTAA。

[2483] 40A.根据实施例39A所述的试剂盒,其进一步包含消除试剂,用于移除(例如,通过化学裂解或酶裂解)第一官能化NTAA,以暴露紧邻的氨基酸残基(如第二NTAA)。

[2484] 41A.根据实施例40A所述的试剂盒,其中第二NTAA能够通过相同或不同的官能化试剂官能化以产生第二官能化NTAA,其可与第一官能化NTAA相同或不同。

[2485] 42A.根据实施例41A所述的试剂盒,其中第一官能化NTAA和第二官能化NTAA彼此独立地选自由以下组成的群组:官能化N-末端丙氨酸(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),缬氨酸(V或Val),色氨酸(W或Trp),酪氨酸(Y或Tyr),其任何组合。

[2486] 43A.根据实施例39A-42A中任一个所述的试剂盒,其包含:

[2487] (i)和/或一种或多种试剂,用于在第一结合剂和多肽之间结合时,将信息从第一结合剂的第一编码聚合物转移到第一记录聚合物,以产生第一延伸记录聚合物,和/或

[2488] (ii)一种或多种试剂,用于在第二结合剂和多肽之间结合时,将信息从第二结合剂的第二编码聚合物转移到第二记录聚合物,以产生第二延伸记录聚合物,

[2489] 其中(i)的一种或多种试剂和(ii)的一种或多种试剂可以相同或不同。

[2490] 44A.根据实施例43A所述的试剂盒,其进一步包含:

[2491] (iii)一种或多种试剂,用于在第三(或更高次序)结合剂和多肽之间结合时,将信息从第三(或更高次序)结合剂的第三(或更高次序)编码聚合物转移到第三(或更高次序)记录聚合物,以产生第三(或更高次序)延伸记录聚合物。

[2492] 45A.根据实施例43A或44A所述的试剂盒,其中第一记录聚合物、第二记录聚合物和/或第三(或更高次序)记录聚合物被配置为与多肽直接或间接相关。

[2493] 46A.根据实施例43A-45A中任一个所述的试剂盒,其中第一记录聚合物、第二记录聚合物和/或第三(或更高次序)记录聚合物被配置为固定在支持物上。

[2494] 47A.根据实施例43A-46A中任一个所述的试剂盒,其中第一记录聚合物、第二记录聚合物和/或第三(或更高次序)记录聚合物可以被配置为与多肽共定位,例如,以在第一、第二或第三(或更高次序)结合剂和多肽之间结合时允许分别在第一、第二或第三(或更高次序)编码聚合物与第一、第二或第三(或更高次序)记录聚合物之间转移信息。

[2495] 48A.根据实施例43A-47A中任一个所述的试剂盒,其中第一编码聚合物、第二编码聚合物,和/或第三(或更高次序)编码聚合物各自包含结合环特异性条形码,例如结合循环特异性间隔区序列$C_n$和/或编码聚合物特异性间隔区序列$C_n$,或外源性结合环标签$C_n$,其中n是整数,并且$C_n$表示第n结合剂和多肽之间的结合。

[2496] 49A.根据实施例1A-48A中任一个所述的试剂盒,其中分析物或多肽包含蛋白质或多肽链或其片段,脂质,碳水化合物,或大环。

[2497] 50A.根据实施例1A-49A中任一个所述的试剂盒,其中分析物或多肽包含大分子或其复合物,例如蛋白质复合物或其亚基。

[2498] 51A.根据实施例1A-50A中任一个所述的试剂盒,其中记录聚合物包含核酸,寡核苷酸,修饰的寡核苷酸,DNA分子,具有伪互补碱基的DNA,具有一个或多个保护的碱基的DNA或RNA,RNA分子,BNA分子,XNA分子,LNA分子,PNA分子,γPNA分子或吗啉代,非核酸可测序聚合物,例如多糖,多肽,肽或聚酰胺,或其组合。

[2499] 52A.根据实施例1A-51A中任一个所述的试剂盒,其中记录聚合物包含通用引发位点。

[2500] 53A.根据实施例1A-52A中任一个所述的试剂盒,其中记录聚合物包含用于扩增、测序或两者的引发位点,例如,通用引发位点包含用于扩增、测序或两者的引发位点。

[2501] 54A.根据实施例1A-53A中任一个所述的试剂盒,其中记录聚合物和/或编码聚合物包含独特分子标识符(UMI)。

[2502] 55A.根据实施例1A-54A中任一个所述的试剂盒,其中记录聚合物和/或编码聚合物包含条形码和/或核酸酶位点,例如切刻内切核酸酶位点(例如,dsDNA切刻内切核酸酶位点)。

[2503] 56A.根据实施例1A-55A中任一个所述的试剂盒,其中记录聚合物和/或编码聚合物在其3'-末端和/或在其5'-末端包含间隔区聚合物,例如记录聚合物在其3'-末端包含间隔区聚合物。

[2504] 57A.根据实施例1A-56A中任一个所述的试剂盒,其包含固体支持物,例如刚性固体支持物、柔性固体支持物或软固体支持物,并且包括多孔支持物或非多孔支持物。

[2505] 58A.根据实施例1A-57A中任一个所述的试剂盒,其中支持物包含珠粒、多孔珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[2506] 59A.根据实施例58A所述的试剂盒,其中支持物包含聚苯乙烯珠粒、聚合物珠粒、

琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒，或其任何组合。

[2507] 60A.根据实施例1A-59A中任一个所述的试剂盒，其包含支持物并且用于在顺序反应、平行反应或顺序和平行反应的组合中分析多种分析物或多肽。

[2508] 61A.根据实施例60A所述的试剂盒，其中分析物或多肽在支持物上以等于或大于约10nm，等于或大于约15nm，等于或大于约20nm，等于或大于约50nm，等于或大于约100nm，等于或大于约150nm，等于或大于约200nm，等于或大于约250nm，等于或大于约300nm，等于或大于约350nm，等于或大于约400nm，等于或大于约450nm，或等于或大于约500nm的平均距离间隔。

[2509] 62A.根据实施例1A-61A中任一个所述的试剂盒，其中结合基团包含多肽或其片段，蛋白质或多肽链或其片段，或蛋白质复合物或其亚基，如抗体或其抗原结合片段。

[2510] 63A.根据实施例1A-62A中任一个所述的试剂盒，其中结合基团包含抗体或其变体，突变体或修饰的蛋白质；氨酰tRNA合成酶或其变体，突变体或修饰的蛋白质；anticalin或其变体，突变体或修饰的蛋白质；ClpS或其变体，突变体或修饰的蛋白质；UBR盒蛋白质或其变体，突变体或修饰的蛋白质；或结合氨基酸的修饰的小分子，即万古霉素或其变体，突变体或修饰的分子；或其任何组合。

[2511] 其中在每个结合剂中，结合基团包含小分子，编码聚合物包含识别小分子的多核苷酸，由此多个结合剂形成编码的小分子库，例如DNA编码的小分子库。

[2512] 64A.根据实施例1A-63A中任一个所述的试剂盒，其中结合基团能够选择性地和/或特异性地结合至分析物或多肽。

[2513] 65A.根据实施例1A-64A中任一个所述的试剂盒，其中编码聚合物包含核酸，寡核苷酸，修饰的寡核苷酸，DNA分子，具有伪互补碱基的DNA，具有一个或多个保护的碱基的DNA或RNA，RNA分子，BNA分子，XNA分子，LNA分子，PNA分子，γPNA分子或吗啉代，或非核酸可测序聚合物，例如，多糖，多肽，肽或聚酰胺，其组合。

[2514] 66A.根据实施例1A-65A中任一个所述的试剂盒，其中编码聚合物包含条形码序列，例如编码器序列，例如识别结合基团的编码序列。

[2515] 67A.根据实施例1A-57A中任一个所述的试剂盒，其中编码聚合物包含间隔区聚合物、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。

[2516] 任选地，其中在每个结合循环之后向记录聚合物添加结合循环特异性序列。

[2517] 68A.根据实施例1A-67A中任一个所述的试剂盒，其中结合基团和编码聚合物通过接头或结合对连接。

[2518] 69A.根据实施例1A-68A中任一个所述的试剂盒，其中结合基团和编码标签通过SpyTag/SpyCatcher，SpyTag-KTag/SpyLigase(其中有待连接的两个部分具有SpyTag/KTag对，并且SpyLigase将SpyTag连接至KTag，由此将这两个部分连接)，分选酶，SnoopTag/SnoopCatcher肽-蛋白质对，或HaloTag/HaloTag配体对，或其任何组合。

[2519] 70A.根据实施例1A-69A中任一个所述的试剂盒，其还包含用于在模板或非模板反应中在编码聚合物和记录聚合物之间转移信息的试剂，任选地其中试剂是(i)化学连接试剂或生物连接试剂，例如连接酶，例如用于连接单链核酸或双链核酸的DNA连接酶或RNA连接酶，或(ii)用于单链核酸或双链核酸的引物延伸的试剂，

[2520] 任选地,其中试剂盒还包含连接试剂,连接试剂包含至少两种连接酶或其变体(例如,至少两种DNA连接酶,或至少两种RNA连接酶,或至少一种DNA连接酶和至少一种RNA连接酶),其中至少两种连接酶或其变体包含腺苷酸化连接酶和组成型非腺苷酸化连接酶,或

[2521] 任选地,其中试剂盒还包含连接试剂,连接试剂包含DNA或RNA连接酶和DNA/RNA脱腺苷酸酶。

[2522] 71A.根据实施例1A-70A中任一个所述的试剂盒,其进一步包含聚合酶,例如DNA聚合酶或RNA聚合酶或逆转录酶,[添加连接酶]用于在编码聚合物和记录聚合物之间转移信息。

[2523] 72A.根据实施例1A-71A中任一个所述的试剂盒,其进一步包含一种或多种用于核酸序列分析的试剂。

[2524] 73A.根据实施例72A所述的试剂盒,核酸序列分析包含通过合成测序、通过连接测序、通过杂交测序、polony测序、离子半导体测序、焦磷酸测序、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,或其任何组合。

[2525] 74A.根据实施例1A-73A中任一个所述的试剂盒,进一步包含一种或多种用于核酸扩增的试剂,例如用于扩增一个或多个延伸记录标签,任选地其中核酸扩增包含指数扩增反应(例如,聚合酶链反应(PCR),例如乳液PCR以减少或消除模板转换)和/或线性扩增反应(例如,通过体外转录的等温扩增,或等温嵌合引物引发的核酸扩增(ICAN))。

[2526] 75A.根据实施例1A-74A中任一个所述的试剂盒,其包含用于将编码标签信息转移至记录标签以形成延伸记录标签的一种或多种试剂,其中延伸记录标签上的编码标签信息的次序和/或频率指示结合剂结合至分析物或多肽的次序和/或频率。

[2527] 76A.根据实施例1A-75A中任一个所述的试剂盒,其还包含一种或多种用于靶标富集,例如富集一种或多种延伸记录聚合物的试剂。

[2528] 77A.根据实施例1A-76A中任一个所述的试剂盒,还包含一种或多种用于扣除,例如扣除一个或多个延伸记录聚合物的试剂。

[2529] 78A.根据实施例1A-77A中任一个所述的试剂盒,进一步包含一种或多种用于归一化的试剂,例如,用以减少高丰度种类,如一种或多种分析物或多肽。

[2530] 79A.根据实施例1A-78A中任一个所述的试剂盒,其中至少一种结合剂结合至末端氨基酸残基、末端二氨基酸残基或末端三氨基酸残基。

[2531] 80A.根据实施例1A-79A中任一个所述的试剂盒,其中至少一种结合剂结合到翻译后修饰的氨基酸。

[2532] 81A.根据实施例1A-80A中任一个所述的试剂盒,进一步包含用于将样品中的多种分析物或多肽分区到多个隔室中的一种或多种试剂或工具,其中每个隔室包含任选地连接到支持物(例如,固体支持物)上的多个隔室标签(例如,非核酸可测序聚合物形式),其中多个隔室标签在单个隔室中是相同的并且不同于其它隔室的隔室标签。

[2533] 82A.根据实施例81A所述的试剂盒,其还包含用于将多种分析物或多肽(例如,多个蛋白质复合物、蛋白质和/或多肽)片段化成多种多肽片段的一种或多种试剂或工具。

[2534] 83A.根据实施例81A或82A所述的试剂盒,其进一步包含一种或多种用于将多个多肽片段与多个隔室中的每一个内的隔室标签退火或连接的试剂或工具,从而产生多个隔室标记的多肽片段。

[2535] 84A.根据实施例81A-83A中任一个所述的试剂盒,其中多个隔室包含微流体液滴,微孔,或表面上的分离区域,或其任何组合。

[2536] 85A.根据实施例81A-84A中任一个所述的试剂盒,其中每一个隔室平均包含单独的细胞。

[2537] 86A.根据实施例81A-85A中任一个所述的试剂盒,进一步包含用于标记样品中的多种分析物或多肽的一个或多个通用DNA标签。

[2538] 87A.根据实施例81A-86A中任一个所述的试剂盒,其进一步包含一种或多种试剂,用于用一种或多种通用DNA标签标记样品中的多种分析物或多肽。

[2539] 88A.根据实施例81A-87A中任一个所述的试剂盒,其还包含一种或多种用于引物延伸或连接的试剂。

[2540] 89A.根据实施例81A-88A中任一个所述的试剂盒,其中珠粒是聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任意组合。

[2541] 90A.根据实施例81A-89A中任一个所述的试剂盒,其中隔室标签包含单链或双链核酸分子。

[2542] 91A.根据实施例81A-90A中任一个所述的试剂盒,其中隔室标签包含条形码和可选的UMI。

[2543] 92A.根据实施例81A-90A中任一个所述的试剂盒,其中支持物是珠粒并且并且隔室标签包含条形码。

[2544] 93A.根据实施例81A-92A中任一个所述的试剂盒,其述支持物包含珠粒,并且其中包含连接到其上的多个隔室标签的珠粒通过均分与合并合成、单独合成或固定化形成。

[2545] 94A.根据实施例81A-93A中任一个所述的试剂盒,其进一步包含一种或多种用于分开-汇集合成、个别合成或固定化的试剂。

[2546] 95A.根据实施例81A-94A中任一个所述的试剂盒,其中隔室标签是记录聚合物中的成分,其中记录聚合物任选地还包含间隔区聚合物、条形码序列、独特分子标识符、通用引发位点或它们的任何组合。

[2547] 96A.根据实施例81A-95A中任一个所述的试剂盒,其中隔室标签还包含能够与多个分析物或多肽(例如,蛋白质复合物、蛋白质或多肽)上的内部氨基酸,肽骨架或N-末端氨基酸反应的官能部分。

[2548] 97A.根据实施例96A所述的试剂盒,官能部分包含醛、叠氮化物/炔、马来酰亚胺/硫醇、环氧/亲核试剂、逆电子需求Diels-Alder(iEDDA)基团、点击试剂或其任何组合。

[2549] 98A.根据实施例81A-97A中任一个所述的试剂盒,其中隔室标签进一步包含肽,诸如蛋白连接酶识别序列,任选地其中蛋白连接酶是butelase I或其同源物。

[2550] 99A.根据实施例81A-98A中任一个所述的试剂盒,其进一步包含化学或生物试剂,例如酶,例如蛋白酶(例如,多金属蛋白酶),用于使多种分析物或多肽片段化。

[2551] 100A.根据实施例81A-99A中任一个所述的试剂盒,进一步包含用于从支持物释放隔室标签的一种或多种试剂。

[2552] 101A.根据实施例1A-100A中任一个所述的试剂盒,其进一步包含一种或多种用于形成延伸编码聚合物或双标签构建体的试剂。

[2553] 102A.根据实施例101A所述的试剂盒,其中记录聚合物的3'末端被封闭以防止记录聚合物通过聚合酶延伸。

[2554] 103A.根据实施例101A或102A所述的试剂盒,其中编码聚合物包含编码序列、UMI、通用引发位点、在其3'末端的间隔区聚合物、结合循环特异性序列,或其任何组合。

[2555] 104A.根据实施例101A-103A中任一个所述的试剂盒,其中双标签构建体通过间隙填充、引物延伸或它们二者产生。

[2556] 105A.根据实施例101A-104A中任一个所述的试剂盒,其中双标签分子包含衍生自记录聚合物的通用引发位点、衍生自记录聚合物的隔室标签、衍生自记录聚合物的独特分子标识符、衍生自记录聚合物的可选间隔区聚合物、衍生自编码聚合物的编码器序列、衍生自编码聚合物的独特分子标识符、衍生自编码聚合物的可选间隔区聚合物、和衍生自编码聚合物的通用引发位点。

[2557] 106A.根据实施例101A-105A中任一个所述的试剂盒,其中结合剂是多肽或蛋白质。

[2558] 107A.根据实施例101A-106A中任一个所述的试剂盒,其中结合剂的结合基团包含氨肽酶或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰的蛋白质;anticalin或其变体,突变体或修饰的蛋白质;Clps(例如ClpS2)或其变体,突变体或修饰的蛋白质;UBR盒蛋白质或其变体,突变体或修饰的蛋白质;或结合氨基酸的修饰的小分子,即万古霉素或其变体,突变体或修饰的分子;或抗体或其结合片段;或其任何组合。

[2559] 108A.根据实施例101A-107A中任一个所述的试剂盒,其中结合剂结合到单个氨基酸残基(例如,N-末端氨基酸残基,C-末端氨基酸残基,或内部氨基酸残基),二肽(例如,N-末端二肽,C-末端二肽,或内部二肽),三肽(例如,N-末端三肽,C-末端三肽,或内部三肽),或分析物或多肽的翻译后修饰。

[2560] 109A.根据实施例101A-107A中任一个所述的试剂盒,其中结合剂结合N-末端多肽、C-末端多肽或内部多肽。

[2561] 110A.根据实施例1A-109A中任一个所述的试剂盒,其中编码聚合物和/或记录聚合物包含一种或多种纠错码(例如,以非核酸序列可测序聚合物的形式),一个或多个编码器序列(例如,以非核酸序列可测序聚合物的形式),一种或多种条形码(例如,以非核酸可测序聚合物的形式),一个或多个UMI(例如,以非核酸可测序聚合物的形式),一个或多个隔室标签(例如,以非核酸可测序聚合物的形式),一个或多个循环特异性序列(例如,以非核酸可测序聚合物的形式)或其任何组合。

[2562] 111A.根据实施例110A所述的试剂盒,其中纠错码选自汉明码、李距离码、非对称李距离码、里德-所罗门码和Levenshtein-Tenengolts码。

[2563] 112A.根据实施例1A-111A中任一个所述的试剂盒,其中编码聚合物和/或记录聚合物包含循环标签。

[2564] 113A.根据实施例1A-112A中任一个所述的试剂盒,其进一步包含独立于编码聚合物和/或记录聚合物的循环标签。

[2565] 114A.根据实施例1A-113A中任一个所述的试剂盒,其包含:

[2566] (a)用于产生细胞裂解物或蛋白质样品的试剂;

[2567] (b)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;

[2568] (c)蛋白酶,例如胰蛋白酶,LysN或LysC;

[2569] (d)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;

[2570] (e)用于基于降解的多肽测序的试剂;和/或

[2571] (f)用于核酸测序的试剂。

[2572] 115A.根据实施例1A-113A中任一个所述的试剂盒,其包含:

[2573] (a)用于产生细胞裂解物或蛋白质样品的试剂;

[2574] (b)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;

[2575] (c)蛋白酶,例如胰蛋白酶、LysN或LysC;

[2576] (d)用于将多肽(例如蛋白质)固定到包含固定的记录聚合物的支持物上的试剂;

[2577] (e)用于基于降解的多肽测序的试剂;和/或

[2578] (f)用于核酸测序的试剂。

[2579] 116A.根据实施例1A-113A中任一个所述的试剂盒,其包含:

[2580] (a)用于产生细胞裂解物或蛋白质样品的试剂;

[2581] (b)变性试剂;

[2582] (c)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;

[2583] (d)通用DNA引物序列;

[2584] (e)用通用DNA引物序列标记多肽的试剂;

[2585] (e)用于通过引物对标记的多肽进行退火的条形码化的珠粒;

[2586] (g)用于聚合酶延伸的试剂,用于将条形码从珠粒写到标记的多肽;

[2587] (h)蛋白酶,例如胰蛋白酶、LysN或LysC;

[2588] (i)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;

[2589] (j)用于基于降解的多肽测序的试剂;和/或

[2590] (k)用于核酸测序的试剂。

[2591] 117A.根据实施例1A-113A中任一个所述的试剂盒,其包含:

[2592] (a)交联剂;

[2593] (b)用于产生细胞裂解物或蛋白质样品的试剂;

[2594] (c)用于封闭氨基酸侧链的试剂,例如通过半胱氨酸或封闭赖氨酸的烷基化;

[2595] (d)通用DNA引物序列;

[2596] (e)用通用DNA引物序列标记多肽的试剂;

[2597] (e)用于通过引物对标记的多肽进行退火的条形码化的珠粒;

[2598] (g)用于聚合酶延伸的试剂,用于将条形码从珠粒写到标记的多肽;

[2599] (h)蛋白酶,例如胰蛋白酶、LysN或LysC;

[2600] (i)用于将核酸标记的多肽(例如DNA标记的蛋白质)固定到支持物上的试剂;

[2601] (j)用于基于降解的多肽测序的试剂;和/或

[2602] (k)用于核酸测序的试剂。

[2603] 118A.根据实施例1A-117A中任一个所述的试剂盒,其中一种或多种组分在溶液中或在支持物例如固体支持物上提供。

[2604] 1B.一种方法,其包含:

[2605] (a)使一组蛋白质与试剂库接触,其中每种蛋白质直接或间接地与记录聚合物相

关,其中每种试剂包含(i)小分子,肽或肽模拟物,肽类(例如,类肽、β-肽或D-肽肽模拟物),多糖,或适体(例如,核酸适体,如DNA适体或肽适体),和(ii)编码聚合物,其包含关于小分子,肽或肽模拟物,肽类(例如,类肽、β-肽或D-肽肽模拟物,多糖或适体的识别信息,

[2606] 其中将每种蛋白质和/或其相关记录聚合物或每种试剂直接或间接固定到支持物上;

[2607] (b)允许在(i)与一种或多种试剂的所述小分子,肽或肽模拟物,肽类(例如,类肽、β-肽或D-肽肽模拟物),多糖,或适体(例如,核酸适体,如DNA适体或肽适体)结合和/或反应的每种蛋白质相关的记录聚合物,和(ii)一种或多种试剂的编码聚合物之间转移信息,以产生延伸记录聚合物和/或延伸编码聚合物;和

[2608] (c)分析所述延伸记录聚合物和/或所述延伸编码聚合物。

[2609] 2B.根据实施例1B所述的方法,其中每种蛋白质与固定在支持物上的其他蛋白质间隔开的平均距离等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[2610] 3B.根据实施例1B或2B所述的方法,其中每种蛋白质及其相关记录聚合物与支持物上的其他蛋白质及其相关记录聚合物间隔开的平均距离等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[2611] 4B.根据实施例1B-3B中任一个所述的方法,其中将一种或多种蛋白质和/或其相关记录聚合物共价固定到支持物上(例如,通过接头),或非共价固定到支持物上(例如,通过结合对)。

[2612] 5B.根据实施例1B-4B中任一个所述的方法,其中蛋白质和/或其相关联的记录聚合物的一个子集共价地固定支持物上,而蛋白质和/或其相关联的记录聚合物的另一个子集非共价地固定到支持物上。

[2613] 6B.根据实施例1B-5B中任一个所述的方法,其中将一个或多个记录聚合物固定到支持物上,从而固定相关的蛋白质。

[2614] 7B.根据实施例1B-6B中任一个所述的方法,其中将一种或多种蛋白质固定到支持物上,从而固定相关的记录聚合物。

[2615] 8B.根据实施例1B-7B中任一个所述的方法,其中至少一种蛋白质与其相关记录聚合物共定位,同时各自独立地固定到支持物上。

[2616] 9B.根据实施例1B-8B中任一个所述的方法,其中至少一种蛋白质和/或其相关记录聚合物直接或间接地与固定接头相关,并且固定接头直接或间接地固定到支持物上,由此将至少一种蛋白质和/或其相关记录聚合物固定到支持物上。

[2617] 10B.根据实施例1B-9B中任一个所述的方法,其中固定的记录聚合物的密度等于或大于固定的蛋白质的密度。

[2618] 11B.根据实施例10B所述的方法,其中固定的记录聚合物的密度为固定的蛋白质的密度至少约2倍,至少约3倍,至少约4倍,至少约5倍,至少约6倍,至少约7倍,至少约8倍,至少约9倍,至少约10倍,至少约20倍,至少约50倍,至少约100倍或更多。

[2619] 12B.根据实施例1B所述的方法,其中每种试剂与固定在支持物上的其他试剂间隔开的平均距离等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[2620] 13B.根据实施例12B所述的方法,其中将一种或多种试剂共价固定到支持物上(例如,通过接头),或者非共价固定到支持物上(例如,通过结合对)。

[2621] 14B.根据实施例12B或13B所述的方法,其中试剂的一个子集共价固定至支持物,而试剂的另一个子集非共价固定至支持物。

[2622] 15B.根据实施例12B-14B中任一个所述的方法,其中对于一种或多种试剂,将小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体固定至支持物,从而固定编码聚合物。

[2623] 16B.根据实施例12B-15B中任一个所述的方法,其中对于一种或多种试剂,将编码聚合物固定到支持物上,从而固定小分子,肽或肽模拟物,肽类(例如,类肽、β-肽,或D-肽肽模拟物),多糖,或适体。

[2624] 17B.根据实施例1B-16B中任一个所述的方法,其中将信息从至少一个编码聚合物转移到至少一个记录聚合物,由此产生至少一个延伸记录聚合物。

[2625] 18B.根据实施例1B-17B中任一个所述的方法,其中将信息从至少一个记录聚合物转移到至少一个编码标签,由此产生至少一个延伸编码标签。

[2626] 19B.根据实施例1B-18B中任一个所述的方法,其中产生至少一个双标签构建体,双标签构建体包含来自编码聚合物的信息和来自记录聚合物的信息。

[2627] 20B.根据实施例1B-19B中任一个所述的方法,其中至少一种蛋白质与两种或更多种试剂的小分子,肽或肽模拟物,肽类(例如,类肽、β-肽或D-肽肽模拟物),多糖或适体结合和/或反应。

[2628] 21B.根据实施例20B所述的方法,其中延伸记录聚合物或延伸编码聚合物包含关于两种或更多种试剂的小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体的识别信息。

[2629] 22B.根据实施例1B-21B中任一个所述的方法,,其中蛋白质中的至少一种与两个或更多个记录聚合物相关,其中两个或更多个记录聚合物可以相同或不同。

[2630] 23B.根据实施例1B-22B中任一个所述的方法,其中试剂中的至少一种包含两个或更多个编码聚合物,其中两个或更多个编码聚合物可以相同或不同。

[2631] 24B.根据实施例1B-23B中任一个所述的方法,其中信息的转移通过连接(例如,酶

或化学连接,夹板连接,粘端连接,单链(ss)连接例如ssDNA连接,或其任何组合),聚合酶介导的反应(例如,单链核酸或双链核酸的引物延伸),或其任何组合完成。

[2632] 25B.根据实施例24B所述的方法,其中连接和/或聚合酶介导的反应相对于蛋白质与小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体之间的结合占用时间或反应时间具有更快的动力学,

[2633] 任选地其中用于连接和/或聚合酶介导的反应的试剂存在于与蛋白质和小分子,肽或肽模拟物,肽模拟物(例如肽,拟肽,β-肽或D-肽肽模拟物),多糖或适体之间的结合或反应相同的反应体积中。

[2634] 26B.根据实施例1B-25B中任一个所述的方法,其中每种蛋白质通过单独的附着与其记录聚合物相关,和/或其中每个小分子,肽或肽模拟物,肽类(例如,类肽、β-肽,或D-肽肽模拟物),多糖或适体通过单独的附着与其编码聚合物相关。

[2635] 27B.根据实施例26B所述的方法,其中通过核糖体或mRNA/cDNA展示发生附着,其中记录聚合物和/或编码聚合物序列信息包含在mRNA序列中。

[2636] 28B.根据实施例27B所述的方法,其中记录聚合物和/或编码聚合物在mRNA序列的3'端包含通用引物序列、条形码和/或间隔区序列。

[2637] 29B.根据实施例28B所述的方法,其中记录聚合物和/或编码聚合物在3'端进一步包含限制酶消化位点。

[2638] 30B.根据实施例1B-29B中任一个所述的方法,其中该组蛋白质为蛋白质组或其子集,

[2639] 任选地其中该组蛋白质是使用基因组或其子集的体外转录随后体外翻译产生的,或使用转录组或其子集的体外翻译产生的。

[2640] 31B.根据实施例30B所述的方法,其中蛋白质组的子集包含激酶组;分泌体;受体组(例如,GPCRome);免疫蛋白酶体;营养蛋白质组;由翻译后修饰(例如,磷酸化,泛素化,甲基化,乙酰化,糖基化,氧化,脂质化和/或亚硝基化)定义的蛋白质组子集,例如磷酸蛋白质组(例如,磷酸酪氨酸-蛋白质组,酪氨酸-激酶组和酪氨酸-磷酸化),糖蛋白质组等;与组织或器官、发育阶段或生理或病理状况相关的蛋白质组子集;与细胞过程如细胞周期,分化(或去分化),细胞死亡,衰老,细胞迁移,转化或转移相关的蛋白质组子集;或其任何组合。

[2641] 32B.根据实施例1B-31B中任一个所述的方法,其中该组蛋白质来自哺乳动物,例如人,非人动物,鱼,无脊椎动物,节肢动物,昆虫或植物,例如,酵母,细菌,例如,大肠杆菌,病毒,例如,HIV或HCV,或其组合。

[2642] 33B.根据实施例1B-32B中任一个所述的方法,其中该组蛋白质包含蛋白质复合物或其亚基。

[2643] 34B.根据实施例1B-29B中任一个所述的方法,其中记录聚合物包含核酸、寡核苷酸,修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子、吗啉代、非核酸可测序聚合物,如多糖、多肽、肽或聚酰胺,或其组合。

[2644] 35B.根据实施例1B-34B中任一个所述的方法,其中记录聚合物包含通用引发位点。

[2645] 36B.根据实施例1B-35B中任一个所述的方法,其中记录聚合物包含用于扩增、测

序或两者的引发位点,例如,通用引发位点包含用于扩增、测序或两者的引发位点。

[2646] 37B.根据实施例1B-36B中任一个所述的方法,其中记录聚合物包含独特分子标识符(UMI)。

[2647] 38B.根据实施例1B-37B中任一个所述的方法,其中记录聚合物包含条形码。

[2648] 39B.根据实施例1B-38B中任一个所述的方法,其中记录聚合物在其3'-末端包含间隔区聚合物。

[2649] 40B.根据实施例1B-39B中任一个所述的方法,其中支持物是固体支持物,例如刚性固体支持物、柔性固体支持物或软固体支持物,并且包括多孔支持物或非多孔支持物。

[2650] 41B.根据实施例1B-40B中任一个所述的方法,其中支持物包含珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[2651] 42B.根据实施例41B所述的方法,其中支持物包含聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任意组合。

[2652] 43B.根据实施例1B-42B中任一个所述的方法,其用于平行分析该组蛋白质和小分子库,和/或肽或肽模拟物,和/或肽类(例如,类肽,β-肽,或D-肽肽模拟物),和/或多糖,和/或适体之间的相互作用,以便产生小分子-蛋白质结合基质,和/或肽/肽模拟物-蛋白质结合基质,和/或肽类-蛋白质结合基质(例如,类肽-蛋白质结合基质,β-肽-蛋白质结合基质,或D-肽模拟-蛋白质结合基质),和/或多糖-蛋白质结合基质,和/或适体-蛋白结合基质。

[2653] 44B.根据实施例43B所述的方法,其中基质尺寸为约$10^2$,约$10^3$,约$10^4$,约$10^5$,约$10^6$,约$10^7$,约$10^8$,约$10^9$,约$10^{10}$,约$10^{11}$,约$10^{12}$,约$10^{13}$,约$10^{14}$或更大,例如约$2\times10^{13}$。

[2654] 45B.根据实施例1B-44B中任一个所述的方法,其中所述记录聚合物包含核酸、寡核苷酸,修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代、非核酸可测序聚合物,如多糖、多肽、肽或聚酰胺,或其组合。

[2655] 46B.根据实施例1B-45B中任一个所述的方法,其中编码聚合物包含识别小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体的编码器序列。

[2656] 47B.根据实施例1B-46B中任一个所述的方法,其中编码聚合物包含间隔区聚合物、独特分子标识符(UMI)、通用引发位点,或其任何组合。

[2657] 48B.根据实施例1B-47B中任一个所述的方法,其中小分子,肽或肽模拟物,肽类(例如,类肽,β-肽,或D-肽肽模拟物),多糖,或适体和编码聚合物通过接头或结合对连接。

[2658] 49B.根据实施例1B-48B中任一个所述的方法,其中小分子,肽或肽模拟物,肽类(例如,类肽,β-肽或D-肽肽模拟物),多糖或适体和编码聚合物通过SpyTag-KTag/SpyLigase(其中待连接的两个部分具有SpyTag-KTag/SpyLigase对,并且SpyLigase将SpyTag连接至KTag,因此将两个部分连接),SpyTag/SpyCatcher,SnoopTag/SnoopCatcher肽-蛋白质对,分选酶,或HaloTag/HaloTag配体对,或其任何组合连接。

[2659] 50B.一种用于分析多肽的方法,其包含:

[2660] (a)使(i)多肽的片段组与(ii)结合剂库接触,其中每个片段与记录聚合物直接或间接相关,其中每个结合剂包含结合基团和编码聚合物,所述编码聚合物包含关于所述结合基团的识别信息,

[2661] 其中结合基团能够结合片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合通过官能化试剂修饰的一个或多个N-末端、内部或C-末端氨基酸,并且

[2662] 其中每个片段和/或其相关记录聚合物,或每个结合剂直接或间接固定到支持物上;

[2663] (b)当结合基团与片段的一个或多个N-末端、内部或C-末端氨基酸之间结合时,允许在(i)与每个片段相关的记录聚合物和(ii)编码聚合物之间转移信息,以产生延伸记录聚合物和/或延伸编码聚合物;和

[2664] (c)分析延伸记录聚合物和/或延伸编码聚合物。

[2665] 51B.根据实施例1B-48B中任一个所述的方法,其中所述一个或多个N-末端、内部或C-末端氨基酸包含:

[2666] (i)N-末端氨基酸(NTAA);

[2667] (ii)N-末端二肽序列;

[2668] (iii)N-末端三肽序列;

[2669] (iv)内部氨基酸;

[2670] (v)内部二肽序列;

[2671] (vi)内部三肽序列;

[2672] (vii)C-末端氨基酸(CTAA);

[2673] (viii)C-末端二肽序列;或

[2674] (ix)C-末端三肽序列,

[2675] 或其任何组合,

[2676] 任选地其中(i)-(ix)中的任何一个或多个氨基酸残基被修饰或官能化。

[2677] 52B.根据实施例1B-48B中任一个所述的方法,其中一个或多个N-末端、内部或C-末端氨基酸在每个残基处独立地选自由以下组成的群组: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),酪氨酸(Y或Tyr),其任何组合。

[2678] 53B.根据实施例50B-52B中任一个所述的方法,其中结合基团包含多肽或其片段,蛋白质或多肽链或其片段,或蛋白质复合物或其亚基,如抗体或其抗原结合片段。

[2679] 54B.根据实施例50B-53B中任一个所述的方法,其中结合基团包含anticalin或其变体,突变体或修饰的蛋白质;氨酰tRNA合成酶或其变体,突变体或修饰的蛋白质;anticalin或其变体,突变体或修饰的蛋白质;ClpS(例如,ClpS2)或其变体,突变体或修饰的蛋白质;UBR盒蛋白质或其变体,突变体或修饰的蛋白质;或结合氨基酸的修饰的小分子,即万古霉素或其变体,突变体或修饰的分子;或其任何组合。

[2680] 55B.根据实施例50B-53B中任一个所述的方法,其中结合基团能够选择性地和/或

特异性地结合至官能化的N-末端氨基酸(NTAA)、N-末端二肽序列或N-末端三肽序列或其任何组合。

[2681] 56B.一种用于分析多种多肽的方法,包含:

[2682] (a)用多个通用标签标记多个多肽的每个分子;

[2683] (b)在适于使所述多个通用标签与所述多个隔室标签退火或连接的条件下,使所述多个多肽与非核酸可测序聚合物形式的多个隔室标签接触,从而将所述多个多肽分区进入多个隔室(例如,珠粒表面、微流体液滴、微孔、或表面上的间隔区域,或其任何组合),其中在每个隔室内的所述多个隔室标签是相同的,并且不同于其他隔室的隔室标签;

[2684] (c)使每个隔室中的多肽片段化,从而产生多肽片段组,每个多肽片段与包含至少一个通用多核苷酸标签和至少一个隔室标签的记录聚合物相关;

[2685] (d)将多肽片段组直接或间接固定到支持物上;

[2686] (e)使固定的多肽片段组与结合剂库接触,其中每个结合剂包含结合基团和编码聚合物,编码聚合物包含关于结合基团的识别信息,

[2687] 其中,结合基团能够结合片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合通过官能化试剂修饰的一个或多个N-末端、内部或C-末端氨基酸;

[2688] (f)当结合基团与片段的一个或多个N-末端、内部或C-末端氨基酸之间结合时,允许在(i)与每个片段相关的记录聚合物和(ii)编码聚合物之间转移信息,以产生延伸记录聚合物和/或延伸编码聚合物;和

[2689] (g)分析延伸记录聚合物和/或延伸编码聚合物。

[2690] 57B.根据实施例56B所述的方法,其中具有相同隔室标签的多种多肽属于相同蛋白质。

[2691] 58B.根据实施例56B所述的方法,其中具有相同隔室标签的多种多肽属于不同的蛋白质,例如,两种,三种,四种,五种,六种,七种,八种,九种,十种或更多种蛋白质。

[2692] 59B.根据实施例56B-58B所述的方法,其中多个隔室标签被固定到多个基板上,每个基板限定隔室。

[2693] 60B.根据实施例59B所述的方法,其中多个基板选自由以下组成的群组:珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[2694] 61B.根据实施例59B或60B所述的方法,其中多个基板中的每一个包含条形码化颗粒,如条形码化珠粒,例如,聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[2695] 62B.根据实施例59B-61B中任一个所述的方法,其中支持物选自由以下组成的群组:珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒(Fe304)、金纳米颗粒和/或银纳米颗

粒)、量子点、纳米壳、纳米微球,或其任何组合。

[2696] 63B.根据实施例59B-61B中任一个所述的方法,支持物包含测序珠粒,例如,聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[2697] 64B.根据实施例56B-63B中任一个所述的方法,其中每个片段及其相关记录聚合物与支持物上的其他片段及其相关记录聚合物间隔开的平均距离等于或大于约20nm,等于或大于约50nm,等于或大于约100nm,等于或大于约150nm,等于或大于约200nm,等于或大于约250nm,等于或大于约300nm,等于或大于约350nm,等于或大于约400nm,等于或大于约450nm,等于或大于约500nm,等于或大于约550nm,等于或大于约600nm,等于或大于约650nm,等于或大于约700nm,等于或大于约750nm,等于或大于约800nm,等于或大于约850nm,等于或大于约900nm,等于或大于约950nm,或等于或大于约1μm。

[2698] 65B.一种用于分析多种多肽的方法,其包含:

[2699] (a)将多个多肽固定至多个基板,其中每个基板包含多个记录聚合物,每个记录聚合物包含隔室标签,任选地其中每个隔室是珠粒、微流体液滴、微孔、或表面上的间隔区域,或其任何组合;

[2700] (b)对固定在每个基板上的多肽进行片段化(例如,通过蛋白酶消化),从而产生固定在所述基板上的多肽片段组;

[2701] (c)使固定的多肽片段组与结合剂库接触,其中每个结合剂包含结合基团和编码聚合物,编码聚合物包含关于结合基团的识别信息,

[2702] 其中,结合基团能够结合片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合通过官能化试剂修饰的一个或多个N-末端、内部或C-末端氨基酸;

[2703] (D)当结合基团和每个片段的一个或多个N-末端、内部或C-末端氨基酸之间结合时,允许在(i)记录聚合物和(ii)编码聚合物之间转移信息,以产生延伸记录聚合物和/或延伸编码聚合物;和

[2704] (e)分析延伸记录聚合物和/或延伸编码聚合物。

[2705] 66B.根据实施例65B所述的方法,其中具有相同隔室标签的多种多肽属于相同蛋白质。

[2706] 67B.根据实施例65B所述的方法,其中具有相同隔室标签的多种多肽属于不同的蛋白质,例如,两种,三种,四种,五种,六种,七种,八种,九种,十种或更多种蛋白质。

[2707] 68B.根据实施例65B-67B中任一个所述的方法,其中每个基板限定隔室。

[2708] 69B.根据实施例65B-68B中任一个所述的方法,其中多个基板选自由以下组成的群组:珠粒、多孔珠粒、多孔基质、阵列、表面、玻璃表面、硅表面、塑料表面、载玻片、过滤器、尼龙、芯片、硅晶芯片、流通芯片、包括信号转导电子器件的生物芯片、孔、微量滴定孔、板、ELISA板、盘、旋转干涉测量盘、膜、硝化纤维素膜、基于硝化纤维素的聚合物表面、纳米颗粒(例如,包含金属,如磁性纳米颗粒($Fe_3O_4$)、金纳米颗粒和/或银纳米颗粒)、量子点、纳米壳、纳米微球,或其任何组合。

[2709] 70B.根据实施例65B-69B中任一个所述的方法,其中多个基板中的每一个包含条形码化颗粒,如条形码化珠粒,例如,聚苯乙烯珠粒、聚合物珠粒、琼脂糖珠粒、丙烯酰胺珠粒、固体核心珠粒、多孔珠粒、磁性珠粒、顺磁珠粒、玻璃珠粒或可控孔珠粒,或其任何组合。

[2710] 71B.根据实施例50B-70B中任一个所述的方法,其中官能化试剂包含化学剂、酶和/或生物剂,如异硫氰酸酯衍生物,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。

[2711] 72B.根据实施例50B-71B中任一个所述的方法,其中记录聚合物包含核酸、寡核苷酸,修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代、非核酸可测序聚合物,如多糖、多肽、肽或聚酰胺,或其组合。

[2712] 73B.根据实施例50B-72B中任一个所述的方法,其中记录聚合物包含通用引发位点;用于扩增、测序或两者的引发位点;任选地,独特分子标识符(UMI);条形码;任选地,在其3'末端的间隔区;或其组合。

[2713] 74B.根据实施例50B-73B中任一个所述的方法,其用于测定多肽或多种多肽的序列。

[2714] 75B.根据实施例50B-74B中任一个所述的方法,其中记录聚合物包含核酸、寡核苷酸,修饰的寡核苷酸、DNA分子、具有伪互补碱基的DNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、γPNA分子或吗啉代、非核酸可测序聚合物,如多糖、多肽、肽或聚酰胺,或其组合。

[2715] 76B.根据实施例50B-75B中任一个所述的方法,其中编码聚合物包含编码器序列,任选的间隔区聚合物,任选的独特分子标识符(UMI),通用引发位点,或其任何组合。

[2716] 77B.根据实施例50B-76B中任一个所述的方法,其中结合基团和编码聚合物通过接头或结合对连接。

[2717] 78B.根据实施例50B-77B中任一个所述的方法,其中结合基团和编码标签通过SpyTag-KTag/SpyLigas(其中待连接的两个部分具有SpyTag/KTag对,并且SpyLigase将SpyTag连接至KTag,因此将两个部分连接),SpyTag/SpyCatcher,SnoopTag/SnoopCatcher肽-蛋白质对,分选酶,或HaloTag/HaloTag配体对,或其任何组合连接。

[2718] 79B.根据实施例1B-78B中任一个所述的方法,其中编码聚合物和/或记录聚合物包含一个或多个纠错码(例如,以非核酸可测序聚合物的形式),一个或多个编码器序列(例如,以非核酸可测序聚合物的形式),一种或多种条形码(例如,以非核酸可测序聚合物的形式),一种或多种UMI(例如,以非核酸可测序聚合物的形式),一种或多种隔室标签(例如,以非核酸可测序聚合物的形式),或其任何组合。

[2719] 80B.根据实施例79B所述的方法,其中纠错码选自汉明码、李距离码、非对称李距离码、里德-所罗门码和Levenshtein-Tenengolts码。

[2720] 81B.根据实施例1B-80B中任一个所述的方法,其中分析延伸记录聚合物和/或延伸编码聚合物包含核酸序列分析。

[2721] 82B.根据实施例81B所述的方法,核酸序列分析包含核酸测序方法,例如通过合成测序,通过连接测序,通过杂交测序,polony测序,离子半导体测序或焦磷酸测序,或其任何组合。

[2722] 83B.根据实施例82B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。

[2723] 84B.根据实施例1B-83B中任一个所述的方法,进一步包含一个或多个洗涤步骤。

[2724] 85B.根据实施例1B-84B中任一个所述的方法,其中延伸记录聚合物和/或延伸编码聚合物在分析之前被扩增。

[2725] 86B.根据实施例1B-85B中任一个所述的方法,其中延伸记录聚合物和/或延伸编码聚合物在分析之前进行进行靶标富集测定。

[2726] 87B.根据实施例1B-86B中任一个所述的方法,其中延伸记录聚合物和/或延伸编码聚合物在分析之前进行扣除测定。

[2727] 88B.一种试剂盒,其包含:

[2728] (a)试剂库,其中每个试剂包含(i)小分子,肽或肽模拟物,肽类(例如,类肽、β-肽或D-肽肽模拟物,多糖和/或适体,和(ii)编码聚合物,其包含关于所述小分子、肽或肽模拟物,肽类(例如,类肽、β-肽,或D-肽肽模拟物),多糖,或适体的识别信息;和

[2729] 任选地(b)一组蛋白质,其中每种蛋白质与记录聚合物直接或间接相关,

[2730] 其中将每种蛋白质和/或其相关记录聚合物或每种试剂直接或间接固定到支持物上,以及

[2731] 其中该组蛋白质、记录聚合物,和试剂库被配置为允许在(i)与一种或多种试剂的小分子、肽或肽模拟物,肽类(例如,类肽、β-肽或D-肽肽模拟物),多糖和/或适体结合和/或反应的每种蛋白质相关的记录聚合物和(ii)一种或多种试剂的编码聚合物之间转移信息,以产生延伸记录聚合物和/或延伸编码聚合物。

[2732] 89B.一种用于分析多肽的试剂盒,其包含:

[2733] (a)结合剂库,其中每种结合剂包含结合基团和编码聚合物,编码聚合物包含关于所述结合基团的识别信息,其中结合基团能够结合至片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合至通过官能化试剂修饰的一个或多个N-末端、内部或C-末端氨基酸;和

[2734] 任选地(b)多肽的片段组,其中每个片段直接或间接地与记录聚合物相关,或(b')用于将多肽例如蛋白酶片段化的工具,

[2735] 其中每个片段和/或其相关记录聚合物,或每个结合剂直接或间接固定到支持物上,和

[2736] 其中多肽片段组、记录聚合物和结合剂库被配置为在结合基团与片段的一个或多个N-末端、内部或C-末端氨基酸之间结合时允许在(i)与每个片段相关的记录聚合物和(ii)编码聚合物之间转移信息,以产生延伸记录聚合物和/或延伸编码聚合物。

[2737] 90B.一种用于分析多种多肽的试剂盒,其包含:(a)结合剂库,其中每个结合剂包含结合基团和编码聚合物,编码聚合物包含关于结合基团的识别信息,其中结合基团能够结合至片段的一个或多个N-末端、内部或C-末端氨基酸,或能够结合至通过官能化试剂修饰的一个或多个N-末端、内部或C-末端氨基酸;和(b)多个基板,其任选地具有固定于其上的多个多肽,其中每个基板包含多个记录聚合物,每个记录聚合物包含隔室标签(例如,以非核酸可测序聚合物的形式),任选地其中每个隔室是珠粒、微流体液滴、微孔,或表面上的间隔区域,或其任何组合,其中固定在每个基板上的多肽被配置为片段化(例如,通过蛋白酶裂解)以产生固定至基板的多肽片段组,其中多个多肽、记录聚合物和结合剂库被配置为在结合基团与每个片段的一个或多个N-末端、内部或C-末端氨基酸之间结合时允许在(i)

记录聚合物和(ii)编码聚合物之间转移信息,以产生延伸记录聚合物和/或延伸编码聚合物。

[2738] 1C.根据前述实施例中任一个所述的方法或试剂盒,其中记录标签是可测序聚合物(例如多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[2739] 2C.根据前述实施例中任一个所述的方法或试剂盒,其中记录标签是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。

[2740] 3C.根据前述实施例中任一个所述的方法或试剂盒,其中编码标签包含可测序聚合物(例如多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[2741] 4C.根据前述实施例中任一个所述的方法或试剂盒,其中编码标签是可测序聚合物(例如多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[2742] 5C.根据实施例4C所述的方法或试剂盒,其中记录标签是可测序聚合物。

[2743] 6C.根据前述实施例中任一个所述的方法或试剂盒,其中延伸记录标签和/或延伸编码标签是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。

[2744] 1D.一种方法,其包含:

[2745] (a)使分析物与能够结合至所述分析物的第一结合剂接触,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部分,所述分析物与连接至支持物的记录部分相关,并且所述分析物和/或所述记录部分固定至支持物;

[2746] (b)将第一编码部分的信息转移到记录聚合物,以产生第一次序延伸记录部分;

[2747] (c)使分析物与能够结合至分析物的第二结合剂接触,其中第二结合剂包含具有关于第二结合剂的识别信息的第二编码部分;

[2748] (d)将第二编码部分的信息转移到第一次序延伸记录部分,以产生第二次序延伸记录部分;以及

[2749] (e)分析第二次序延伸记录部分。

[2750] 2D.根据实施例1D所述的方法,还包含在步骤(d)和(e)之间的以下步骤:

[2751] (x)通过采用能够结合分析物的第三(或更高次序)结合剂替换第二结合剂以重复步骤(c)和(d)一次或多次,其中第三(或更高次序)结合剂包含具有关于第三(或更高次序)结合剂的识别信息的第三(或更高次序的)编码部分;和

[2752] (y)将第三(或更高次序)编码部分的信息转移到第二(或更高次序)延伸记录部分,以产生第三(或更高次序)延伸记录部分;

[2753] 其中在步骤(e)中分析第三(或更高次序)延伸记录部分。

[2754] 3D.根据实施例1D或2D所述的方法,其中步骤(a)的记录部分是可测序聚合物例(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[2755] 4D.根据实施例1D或2D所述的方法,其中步骤(a)的记录部分是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。

[2756] 5D.根据实施例4D所述的方法,其中第一、第二、第三和/或更高次序的编码部分包含可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子和/或亚基,单体,残基或结构单元。

[2757] 6D.根据实施例4D或5D所述的方法,其中第一、第二、第三和/或更高次序编码部分是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或

结构单元。

[2758] 7D.根据实施例1D-6D中任一个所述的方法,其中第一次序延伸记录部分、第二次序延伸记录部分、第三次序延伸记录部分和/或更高次序延伸记录部分是可测序聚合物,例如,多核苷酸或非核酸可测序聚合物。

[2759] 8D.一种方法,其包含:

[2760] (a)使多肽与能够结合至所述多肽的一个或多个N-末端、内部或C-末端氨基酸的第一结合剂接触,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部分,所述多肽与连接至支持物的记录部分相关,并且所述多肽和/或所述记录部分固定至支持物;

[2761] (b)将所述第一编码部分的信息转移到所述记录聚合物,以产生第一次序延伸记录部分;

[2762] (c)使分析物与能够结合至多肽的不同的一个或多个N-末端、内部或C-末端氨基酸的第二结合剂接触,其中第二结合剂包含具有关于第二结合剂的识别信息的第二编码部分;

[2763] (d)将第二编码部分的信息转移到第一次序延伸记录部分,以产生第二次序延伸记录部分;以及

[2764] (e)分析第二次序延伸记录部分。

[2765] 9D.根据实施例8D所述的方法,其中多肽的一个或多个N-末端、内部或C-末端氨基酸和/或不同的一个或多个N-末端、内部或C-末端氨基酸是修饰的氨基酸,例如修饰的N-末端氨基酸(NTAA)。

[2766] 10D.一种方法,其包含:

[2767] (a)使多肽与能够结合至所述多肽的一个或多个N-末端或C-末端氨基酸的第一结合剂接触,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部分,所述多肽与连接至支持物的记录部分相关,并且所述多肽和/或所述记录部分固定至支持物;

[2768] (b)将第一编码部分的信息转移到记录聚合物,以产生第一次序延伸记录部分;

[2769] (c)消除所述多肽的一个或多个N-末端或C-末端氨基酸,以暴露所述多肽的不同的一个或多个N-末端或C-末端氨基酸;

[2770] (d)使分析物与能够结合至多肽的不同的一个或多个N-末端或C-末端氨基酸的第二结合剂接触,其中第二结合剂包含具有关于第二结合剂的识别信息的第二编码部分;

[2771] (e)将第二编码部分的信息转移到第一次序延伸记录部分以产生第二次序延伸记录部分;以及

[2772] (f)分析第二次序延伸记录部分。

[2773] 11D.根据实施例10D所述的方法,其中多肽的一个或多个N-末端或C-末端氨基酸和/或不同的一个或多个N-末端或C-末端氨基酸是修饰的氨基酸,诸如修饰的N-末端氨基酸(NTAA)。

[2774] 12D.根据实施例8D-11D中任一个所述的方法,其中步骤(a)的记录部分是可测序聚合物例(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[2775] 13D.根据实施例8D-11D中任一个所述的方法,步骤(a)的记录部分可以是可测序聚合物,例如,多核苷酸或非核酸可测序聚合物。

[2776] 14D.根据实施例13D所述的方法,其中第一或第二次序的编码部分包含可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子和/或亚基,单体,残基或结构单元。

[2777] 15D.根据实施例13D或14D所述的方法,其中第一或第二次序编码部分是可测序聚合物(例如,多核苷酸或非核酸可测序聚合物)的小分子,或亚基,单体,残基或结构单元。

[2778] 16D.根据实施例8D-15D中任一个所述的方法,其中第一次序延伸记录部分或第二次序延伸记录部分是可测序聚合物,例如多核苷酸或非核酸可测序聚合物。

[2779] 17D.根据实施例3D-16D中任一个所述的方法,其中可测序聚合物包含一个或多个封阻基团。

[2780] 18D.根据实施例17D所述的方法,其中当可测序聚合物通过纳米孔时,一个或多个封阻基团能够产生特征电流阻断信号。

[2781] 19D.根据实施例3D-18D中任一个所述的方法,其中可测序聚合物包含一个或多个能够与结合剂结合的结合基团,结合剂在通过纳米孔时使可测序聚合物停顿。

[2782] 20D.根据实施例3D-19D中任一个所述的方法,其中可测序聚合物包含一种或多种当通过纳米孔时能够使可测序聚合物停顿的结构。

[2783] 21D.根据实施例20D所述的方法,其中一种或多种结构包含二级结构,例如本征亚稳定二级结构。

[2784] 22D.根据实施例3D-21D中任一个所述的方法,其中可测序聚合物包含使用亚磷酰胺化学构建的聚合物,或类肽聚合物,或其组合。

[2785] 23D.一种试剂盒,其包含根据实施例1D-23D所述的方法中公开的和/或使用的任何分子,分子复合物或缀合物,试剂(例如,化学或生物试剂),剂,结构(例如,支持物、表面、颗粒或珠粒),反应中间体,反应产物,结合复合物或任何其它制品。

[2786] 本公开不旨在将范围限于所提供的具体公开的实施例,例如,用于说明本发明的各个方面的那些实施例。根据本文的描述和教导,对所述组合物和方法的各种修改将变得容易理解。在不脱离本公开的真实范围和精神的情况下可以实践这些变化,并且这些变化旨在落入本公开的范围内。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,权利要求不受本公开的限制。

[2787] 参考文献:

[2788] Harlow,Ed,and David Lane.Using Antibodies.Cold Spring Harbor,New York:Cold Spring Harbor Laboratory Press,1999.

[2789] 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.

[2790] 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.

[2791] Zhang,L.,Luo,S.,and Zhang,B.(2016).The use of lectin microarray for assessing glycosylation of therapeutic proteins.mAbs 8,524-535.

[2792] 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 and 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.

[2793] 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 and F.J.Steemers(2014)."Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing."Nat Genet 46(12):1343-1349.

[2794] 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 and 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.

[2795] Bailey,J.M.and J.E.Shively(1990)."Carboxy-terminal sequencing:formation and hydrolysis of C-terminal peptidylthiohydantoins."Biochemistry 29(12):3145-3156.

[2796] Bandara,H.M.,D.P.Kennedy,E.Akin,C.D.Incarvito and S.C.Burdette(2009)."Photoinduced release of $Zn^{2+}$with ZinCleav-1:a nitrobenzyl-based caged complex."Inorg Chem 48(17):8445-8455.

[2797] Bandara,H.M.,T.P.Walsh and S.C.Burdette(2011)."A Second-generation photocage for $Zn^{2+}$inspired by TPEN:characterization and insight into the uncaging quantum yields of ZinCleav chelators."Chemistry 17(14):3932-3941.

[2798] Basle,E.,N.Joubert and M.Pucheault(2010)."Protein chemical modification on endogenous amino acids."Chem Biol 17(3):213-227.

[2799] Bilgicer,B.,S.W.Thomas,3rd,B.F.Shaw,G.K.Kaufman,V.M.Krishnamurthy,L.A.Estroff,J.Yang and 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.

[2800] Bochman,M.L.,K.Paeschke and V.A.Zakian(2012)."DNA secondary structures:stability and function of G-quadruplex structures."Nat Rev Genet 13(11):770-780.

[2801] Borgo,B.and J.J.Havranek(2014)."Motif-directed redesign of enzyme specificity."Protein Sci 23(3):312-320.

[2802] Brouzes,E.,M.Medkova,N.Savenelli,D.Marran,M.Twardowski,J.B.Hutchison,J.M.Rothberg,D.R.Link,N.Perrimon and M.L.Samuels(2009)."Droplet microfluidic technology for single-cell high-throughput screening."Proc Natl Acad Sci U S A 106(34):14195-14200.

[2803] Brudno,Y.,M.E.Birnbaum,R.E.Kleiner and D.R.Liu(2010)."An in vitro translation,selection and amplification system for peptide nucleic acids."Nat Chem Biol 6(2):148-155.

[2804] Calcagno,S.and C.D.Klein(2016)."N-Terminal methionine processing by the zinc-activated Plasmodium falciparum methionine aminopeptidase 1b."Appl Microbiol Biotechnol.

[2805] Cao,Y.,G.K.Nguyen,J.P.Tam and C.F.Liu(2015)."Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation."Chem Commun (Camb)51(97):17289-17292.

[2806] 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.

[2807] Chang,L.,D.M.Rissin,D.R.Fournier,T.Piech,P.P.Patel,D.H.Wilson and D.C.Duffy(2012)."Single molecule enzyme-linked immunosorbent assays:theoretical considerations."J Immunol Methods 378(1-2):102-115.

[2808] 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.

[2809] Christoforou,A.,C.M.Mulvey,L.M.Breckels,A.Geladaki,T.Hurrell,P.C.Hayward,T.Naake,L.Gatto,R.Viner,A.Martinez Arias and K.S.Lilley(2016)."A draft map of the mouse pluripotent stem cell spatial proteome."Nat Commun 7:8992.

[2810] Creighton,C.J.and S.Huang(2015)."Reverse phase protein arrays in signaling pathways:a data integration perspective."Drug Des Devel Ther 9:3519-3527.

[2811] Crosetto,N.,M.Bienko and A.van Oudenaarden(2015)."Spatially resolved transcriptomics and beyond."Nat Rev Genet 16(1):57-66.

[2812] Cusanovich,D.A.,R.Daza,A.Adey,H.A.Pliner,L.Christiansen,K.L.Gunderson,F.J.Steemers,C.Trapnell and J.Shendure(2015)."Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing."Science 348(6237):910-914.

[2813] Derrington,I.M.,T.Z.Butler,M.D.Collins,E.Manrao,M.Pavlenok,M.Niederweis and J.H.Gundlach(2010)."Nanopore DNA sequencing with MspA."Proc Natl Acad Sci U S A 107(37):16060-16065.

[2814] El-Sagheer,A.H.,V.V.Cheong and T.Brown(2011)."Rapid chemical ligation

of oligonucleotides by the Diels-Alder reaction."Org Biomol Chem 9(1):232-235.

[2815] El-Sagheer,A.H.,A.P.Sanzone,R.Gao,A.Tavassoli and 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 A108(28):11338-11343.

[2816] Emili,A.,M.McLaughlin,K.Zagorovsky,J.B.Olsen,W.C.W.Chan and S.S.Sidhu(2017).Protein Sequencing Method and Reagents.USPTO.USA,The Governing Council of University of Toronto.9,566,335B1.

[2817] Erde,J.,R.R.Loo and 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.

[2818] Farries,T.C.,A.Harris,A.D.Auffret and 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.

[2819] Feist,P.and 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.

[2820] Friedmann,D.R.and R.Marmorstein(2013)."Structure and mechanism of non-histone protein acetyltransferase enzymes."FEBS J 280(22):5570-5581.

[2821] Frokjaer,S.and D.E.Otzen(2005)."Protein drug stability:a formulation challenge."Nat Rev Drug Discov4(4):298-306.

[2822] Fujii,Y.,M.Kaneko,M.Neyazaki,T.Nogi,Y.Kato and 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.

[2823] Gebauer,M.and A.Skerra(2012)."Anticalins small engineered binding proteins based on the lipocalin scaffold."Methods Enzymol 503:157-188.

[2824] Gerry,N.P.,N.E.Witowski,J.Day,R.P.Hammer,G.Barany and F.Barany(1999)."Universal DNA microarray method for multiplex detection of low abundance point mutations."J Mol Biol 292(2):251-262.

[2825] Gogliettino,M.,M.Balestrieri,E.Cocca,S.Mucerino,M.Rossi,M.Petrillo,E.Mazzella and G.Palmieri(2012)."Identification and characterisation of a novel acylpeptide hydrolase from Sulfolobus solfataricus:structural and functional insights."PLoS One 7(5):e37921.

[2826] Gogliettino,M.,A.Riccio,M.Balestrieri,E.Cocca,A.Facchiano,T.M.D'Arco,C.Tesoro,M.Rossi and 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.

[2827] Granvogl,B.,M.Ploscher and L.A.Eichacker(2007)."Sample preparation by in-gel digestion for mass spectrometry-based proteomics."Anal Bioanal Chem 389(4):991-1002.

[2828] 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.

[2829] Gunderson,K.L.,F.J.Steemers,J.S.Fisher and R.Rigatti(2016).Methods and Compositions for Analyzing Cellular Components.WIPO,Illumina,Inc.

[2830] Gunderson,K.L.,F.J.Steemers,J.S.Fisher and R.Rigatti(2016).Methods and compositions for analyzing cellular components,Illumina,Inc.

[2831] Guo,H.,W.Liu,Z.Ju,P.Tamboli,E.Jonasch,G.B.Mills,Y.Lu,B.T.Hennessy and 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.

[2832] Hamada,Y.(2016)."A novel N-terminal degradation reaction of peptides via N-amidination."Bioorg Med Chem Lett 26(7):1690-1695.

[2833] Hermanson,G.(2013).Bioconjugation Techniques,Academic Press.

[2834] Hernandez-Moreno,A.V.,F.Villasenor,E.Medina-Rivero,N.O.Perez,L.F.Flores-Ortiz,G.Saab-Rincon and G.Luna-Barcenas(2014)."Kinetics and conformational stability studies of recombinant leucine aminopeptidase."Int J Biol Macromol 64:306-312.

[2835] Hori,M.,H.Fukano and Y.Suzuki(2007)."Uniform amplification of multiple DNAs by emulsion PCR."Biochem Biophys Res Commun 352(2):323-328.

[2836] Horisawa,K.(2014)."Specific and quantitative labeling of biomolecules using click chemistry."Front Physiol 5:457.

[2837] Hoshika,S.,F.Chen,N.A.Leal and S.A.Benner(2010)."Artificial genetic systems:self-avoiding DNA in PCR and multiplexed PCR."Angew Chem Int Ed Engl 49(32):5554-5557.

[2838] Hughes,A.J.,D.P.Spelke,Z.Xu,C.C.Kang,D.V.Schaffer and A.E.Herr(2014)."Single-cell western blotting."Nat Methods 11(7):749-755.

[2839] Hughes,C.S.,S.Foehr,D.A.Garfield,E.E.Furlong,L.M.Steinmetz and J.Krijgsveld(2014)."Ultrasensitive proteome analysis using paramagnetic bead technology."Mol Syst Biol 10:757.

[2840] Kang,C.C.,K.A.Yamauchi,J.Vlassakis,E.Sinkala,T.A.Duncombe and A.E.Herr(2016)."Single cell-resolution western blotting."Nat Protoc 11(8):1508-1530.

[2841] Kang,T.S.,L.Wang,C.N.Sarkissian,A.Gamez,C.R.Scriver and 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.

[2842] Katritzky,A.R.and B.V.Rogovoy(2005)."Recent developments in guanylating agents."ARKIVOC iv(Issue in Honor of Prof.Nikolai Zefirov):49-87.

[2843] Klein,A.M.,L.Mazutis,I.Akartuna,N.Tallapragada,A.Veres,V.Li,L.Peshkin,D.A.Weitz and M.W.Kirschner(2015)."Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells."Cell 161(5):1187-1201.

[2844] Knall,A.C.,M.Hollauf and 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.

[2845] Le,Z.G.,Z.C.Chen,Y.Hu and 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.

[2846] Lesch,V.,A.Heuer,V.A.Tatsis,C.Holm and 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.

[2847] Li,G.,Y.Liu,Y.Liu,L.Chen,S.Wu,Y.Liu and X.Li(2013)."Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry."Angew Chem Int Ed Engl 52(36):9544-9549.

[2848] Litovchick,A.,M.A.Clark and 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.

[2849] Liu,Y.and S.Liang(2001)."Chemical carboxyl-terminal sequence analysis of peptides and proteins using tribenzylsilyl isothiocyanate."J Protein Chem 20(7):535-541.

[2850] Lundblad,R.L.(2014).Chemical reagents for protein modification.Boca Raton,CRC Press,Taylor&Francis Group.

[2851] Mashaghi,S.and A.M.van Oijen(2015)."External control of reactions in microdroplets."Sci Rep 5:11837.

[2852] McCormick,R.M.(1989)."A solid-phase extraction procedure for DNA purification."Anal Biochem 181(1):66-74.

[2853] Mendoza,V.L.and R.W.Vachet(2009)."Probing protein structure by amino acid-specific covalent labeling and mass spectrometry."Mass SpectromRev 28(5):785-815.

[2854] Mikami,T.,T.Takao,K.Yanagi and H.Nakazawa(2012)."N(alpha)Selective Acetylation of Peptides."Mass Spectrom(Tokyo)1(2):A0010.

[2855] Moghaddam,M.J.,L.de Campo,N.Kirby and C.J.Drummond(2012)."Chelating DTPA amphiphiles:ion-tunable self-assembly structures and gadolinium complexes."Phys Chem Chem Phys 14(37):12854-12862.

[2856] Mukherjee,S.,M.Ura,R.J.Hoey and 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.

[2857] Namimatsu,S.,M.Ghazizadeh and 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.

[2858] Nguyen,G.K.,Y.Cao,W.Wang,C.F.Liu and J.P.Tam(2015)."Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide."Angew Chem Int Ed Engl 54(52):15694-15698.

[2859] Nguyen,G.K.,S.Wang,Y.Qiu,X.Hemu,Y.Lian and J.P.Tam(2014)."Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis."Nat Chem Biol 10(9):732-738.

[2860] Nishizuka,S.S.and G.B.Mills(2016)."New era of integrated cancer biomarker discovery using reverse-phase protein arrays."Drug Metab Pharmacokinet 31(1):35-45.

[2861] Ohkubo,A.,R.Kasuya,K.Sakamoto,K.Miyata,H.Taguchi,H.Nagasawa,T.Tsukahara,T.Watanobe,Y.Maki,K.Seio and M.Sekine(2008)."'Protected DNA Probes'capable of strong hybridization without removal of base protecting groups."Nucleic Acids Res 36(6):1952-1964.

[2862] Ojha,B.,A.K.Singh,M.D.Adhikari,A.Ramesh and G.Das(2010)."2-Alkylmalonic acid:amphiphilic chelator and a potent inhibitor of metalloenzyme."J Phys Chem B 114(33):10835-10842.

[2863] Peng,X.,H.Li and 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.

[2864] Perbandt,M.,O.Bruns,M.Vallazza,T.Lamla,C.Betzel and 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.

[2865] Rauth,S.,D.Hinz,M.Borger,M.Uhrig,M.Mayhaus,M.Riemenschneider and A.Skerra(2016)."High-affinity Anticalins with aggregation-blocking activity directed against the Alzheimer beta-amyloid peptide."Biochem J 473(11):1563-1578.

[2866] Ray,A.and B.Norden(2000)."Peptide nucleic acid(PNA):its medical and biotechnical applications and promise for the future."FASEB J 14(9):1041-1060.

[2867] Riley,N.M.,A.S.Hebert and J.J.Coon(2016)."Proteomics Moves into the Fast Lane."Cell Syst 2(3):142-143.

[2868] Roloff,A.,S.Ficht,C.Dose and 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.

[2869] Roloff,A.and O.Seitz(2013)."The role of reactivity in DNA templated native chemical PNA ligation during PCR."Bioorg Med Chem 21(12):3458-3464.

[2870] Sakurai,K.,T.M.Snyder and 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.

[2871] Schneider,K.and B.T.Chait(1995)."Increased stability of nucleic acids containing 7-deaza-guanosine and7-deaza-adenosine may enable rapid DNA sequencing by matrix-assisted laser desorption mass spectrometry."Nucleic Acids Res 23(9):1570-1575.

[2872] Selvaraj,R.and J.M.Fox(2013)."trans-Cyclooctene--a stable,voracious dienophile for bioorthogonal labeling."Curr Opin Chem Biol 17(5):753-760.

[2873] Sharma,A.K.,A.D.Kent and J.M.Heemstra(2012)."Enzyme-linked small-molecule detection using split aptamer ligation."Anal Chem 84(14):6104-6109.

[2874] Shembekar,N.,C.Chaipan,R.Utharala and C.A.Merten(2016)."Droplet-based microfluidics in drug discovery,transcriptomics and high-throughput molecular genetics."Lab Chip 16(8):1314-1331.

[2875] Shenoy,N.R.,J.E.Shively and J.M.Bailey(1993)."Studies in C-terminal sequencing:new reagents for the synthesis of peptidylthiohydantoins."J Protein Chem 12(2):195-205.

[2876] Shim,J.U.,R.T.Ranasinghe,C.A.Smith,S.M.Ibrahim,F.Hollfelder,W.T.Huck,D.Klenerman and C.Abell(2013)."Ultrarapid generation of femtoliter microfluidic droplets for single-molecule-counting immunoassays."ACS Nano 7(7):5955-5964.

[2877] Shim,J.W.,Q.Tan and L.Q.Gu(2009)."Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity."Nucleic Acids Res 37(3):972-982.

[2878] 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 and 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.

[2879] Sletten,E.M.and C.R.Bertozzi(2009)."Bioorthogonal chemistry:fishing for selectivity in a sea of functionality."Angew Chem Int Ed Engl 48(38):6974-6998.

[2880] 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 and E.J.Alm(2016)."Massively

parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers."ISME J 10(2):427-436.

[2881] Spicer,C.D.and B.G.Davis(2014)."Selective chemical protein modification."Nat Commun 5:4740.

[2882] Spiropulos,N.G.and 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.

[2883] Switzar,L.,M.Giera and W.M.Niessen(2013)."Protein digestion:an overview of the available techniques and recent developments."J Proteome Res 12(3):1067-1077.

[2884] Tamminen,M.V.and M.P.Virta(2015)."Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells."Front Microbiol 6:195.

[2885] Tessler,L.(2011).Digital Protein Analysis:Technologies for Protein Diagnostics and Proteomics through Single-Molecule Detection.Ph.D.,WASHINGTON UNIVERSITY IN ST.LOUIS.

[2886] Tyson,J.and J.A.Armour(2012)."Determination of haplotypes at structurally complex regions using emulsion haplotype fusion PCR."BMC Genomics 13:693.

[2887] Vauquelin,G.and 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.

[2888] Veggiani,G.,T.Nakamura,M.D.Brenner,R.V.Gayet,J.Yan,C.V.Robinson and M.Howarth(2016)."Programmable polyproteams built using twin peptide superglues."Proc Natl Acad Sci US A 113(5):1202-1207.

[2889] Wang,D.,S.Fang and 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.

[2890] Williams,B.A.and J.C.Chaput(2010)."Synthesis of peptide-oligonucleotide conjugates using a heterobifunctional crosslinker."Curr Protoc Nucleic Acid Chem Chapter 4:Unit4 41.

[2891] Wu,H.and N.K.Devaraj(2016)."Inverse Electron-Demand Diels-Alder Bioorthogonal Reactions."Top Curr Chem (J)374(1):3.

[2892] Xiong,A.S.,R.H.Peng,J.Zhuang,F.Gao,Y.Li,Z.M.Cheng and Q.H.Yao(2008)."Chemical gene synthesis:strategies,softwares,error corrections,and applications."FEMS Microbiol Rev 32(3):522-540.

[2893] Yao,Y.,M.Docter,J.van Ginkel,D.de Ridder and C.Joo(2015)."Single-molecule protein sequencing through fingerprinting:computational assessment."

Phys Biol 12(5):055003.

[2894] Zakeri,B.,J.O.Fierer,E.Celik,E.C.Chittock,U.Schwarz-Linek,V.T.Moy and 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.

[2895] Zhang,L.,K.Zhang,S.Rauf,D.Dong,Y.Liu and J.Li(2016)."Single-Molecule Analysis of Human Telomere Sequence Interactions with G-quadruplex Ligand." Anal Chem 88(8):4533-4540.

[2896] Zhou,H.,Z.Ning,A.E.Starr,M.Abu-Farha and D.Figeys(2012)."Advancements in top-down proteomics."Anal Chem 84(2):720-734.

[2897] Zilionis,R.,J.Nainys,A.Veres,V.Savova,D.Zemmour,A.M.Klein and L.Mazutis(2017)."Single-cell barcoding and sequencing using droplet microfluidics."Nat Protoc 12(1):44-73.

[2898] Vigneron et al.,Proc.Natl.Acad.Sci.1996,93,9682-9686.

[2899] Yong et al.,J.Org.Chem.1997,62,1540-1542.

[2900] Xu et al.,Organometallics.2015,34,1787-1801.

[2901] Kwon et al.,Org.Lett.2014,16,6048-6051.

[2902] Martin et al.,Organometallics.2006,34,1787-1801.

[2903] Bhattacharjree et al.,J.Chem.Sci.2016,128(6):875-881.

[2904] Chi et al.,2015,Chem.Eur.J.2015,21,10369-10378.

[2905] Hamada,Y.,Bioog.Med.Chem.Lett.2016,26,1690-1695.

[2906] Katritzky et al.,Arkivoc.2005,iv,49-87.

[2907] Bachor et al.,Mol.Divers.2013,17,605-611.

[2908] Borgo et al.,Protein Science.2015,24(4),571-579.

[2909] Krishna et al.,Protein Science.1992,1(5),582-589.

[2910] Tian et al.,J.Am.Chem.Soc.,2016,138(43),pp.14234-14237.

[2911] Zhang et al.,Org.Lett.,2001,3(15),2341-2344.

[2912] Musiol et al.,Org.Lett.,2001,3(15),2341-2344.

[2913] Bentley et al.,Biochem.J.1973(135),507-511.

[2914] Bentley et al.,Biochem.J.1976(153),137-138.

[2915] Tam et al.,2007,J.Am.Chem.Soc.2007,129,12670-12671

[2916] Wu et al.,J.Am.Chem.Soc.2016,138(44),14554-14557

[2917] Huo et al.,J.Am.Chem.Soc.2007,139,9819-9822

[2918] Tornqvist et al.,Anal.Biochem.1986,154,255-266

[2919] Rydberg et al.,Chem.Res.Toxicol.,2002,15(4),570-581.

[2920] Proulx et al.,Peptide Science,2016,106(5),726-736.

[2921] Fang et al.,Peptide Science,2010,96(1),97-102.

[2922] Buckingham et al.,J.Am.Chem.Soc.1970,92(19),5571-5579.

[2923] Bader et al.,Arch Occup Environ Healt,1994,65(6),411-414.

[2924] Barrett et al.,Tetrahedron Lett.,1985,26(36),4375-4378.

[2925] Sutton et al,Acc.Chem.Res.1987,20(10),357-364.

[2926] Al Ouahabi,A.,L.Charles and J.F.Lutz(2015)."Synthesis of non-natural sequence-encoded polymers using phosphoramidite chemistry."J Am Chem Soc 137(16):5629-5635.

[2927] Demonte,D.,E.J.Drake,K.H.Lim,A.M.Gulick and S.Park(2013)."Structure-based engineering of streptavidin monomer with a reduced biotin dissociation rate."Proteins 81(9):1621-1633.

[2928] Heerema,S.J.and C.Dekker(2016)."Graphene nanodevices for DNA sequencing."Nat Nanotechnol 11(2):127-136.

[2929] Knight,A.S.,E.Y.Zhou,M.B.Francis and R.N.Zuckermann(2015)."Sequence Programmable Peptoid Polymers for Diverse Materials Applications."Adv Mater 27(38):5665-5691.

[2930] Laszlo,A.H.,I.M.Derrington,B.C.Ross,H.Brinkerhoff,A.Adey,I.C.Nova,J.M.Craig,K.W.Langford,J.M.Samson,R.Daza,K.Doering,J.Shendure and J.H.Gundlach(2014)."Decoding long nanopore sequencing reads of natural DNA."Nat Biotechnol 32(8):829-833.

[2931] 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,权利要求不受本公开的限制。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/或在申请资料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第62/330,841号,美国临时专利申请第62/339,071号和美国临时专利申请第62/376,886号,国际专利申请第PCT/US2017/030702号,美国临时专利申请第62/579,844号、第62/582,312号、第62/583,448号、第62/579,870号、第62/579,840号、第62/582,916号,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。

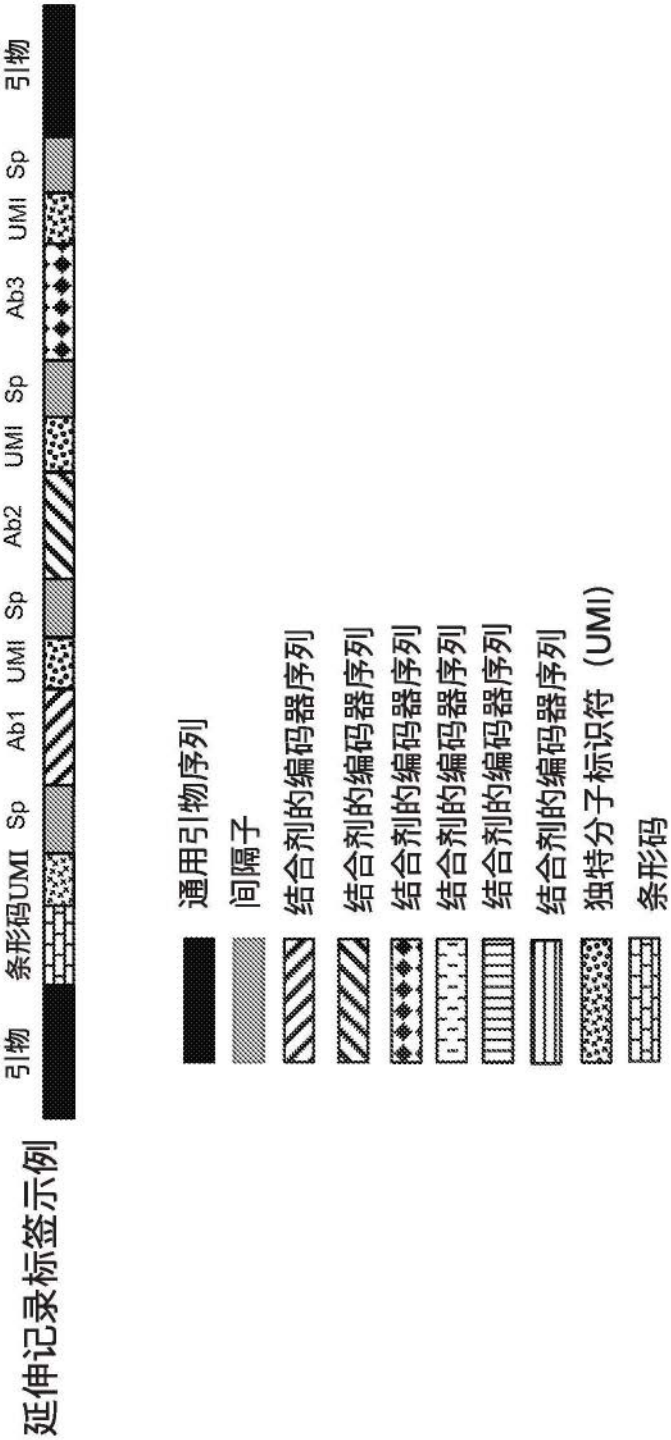

附图中显示的功能元件的图例
延伸记录标签示例
通用引物
条形码UMI
Sp
Ab1
UMI
Sp
Ab2
UMI
Sp
Ab3
UMI
Sp
通用引物
通用引物序列
间隔子
结合剂的编码器序列
结合剂的编码器序列
结合剂的编码器序列
结合剂的编码器序列
结合剂的编码器序列
结合剂的编码器序列
独特分子标识符（UMI）
条形码

图1A

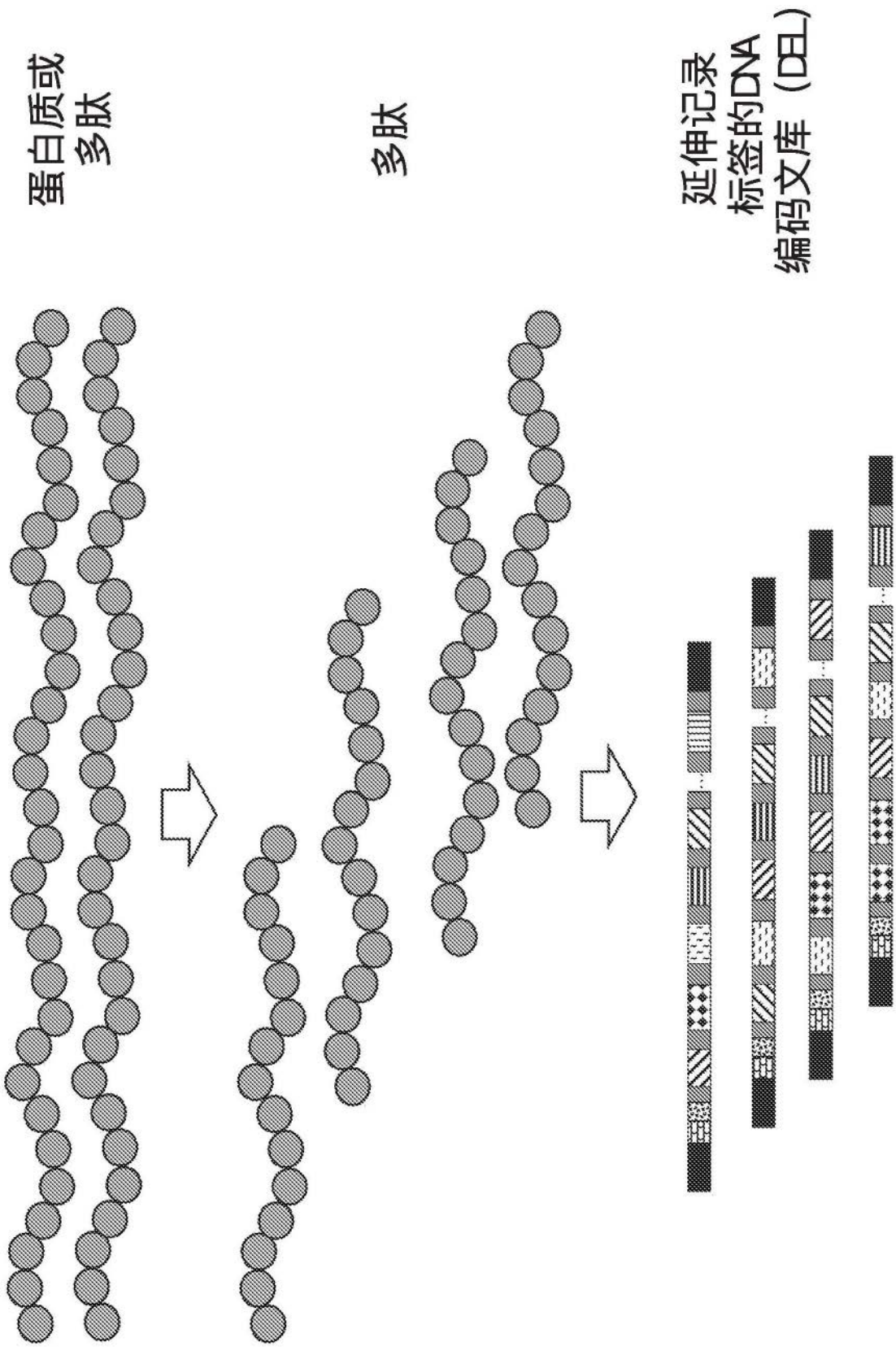



图1B

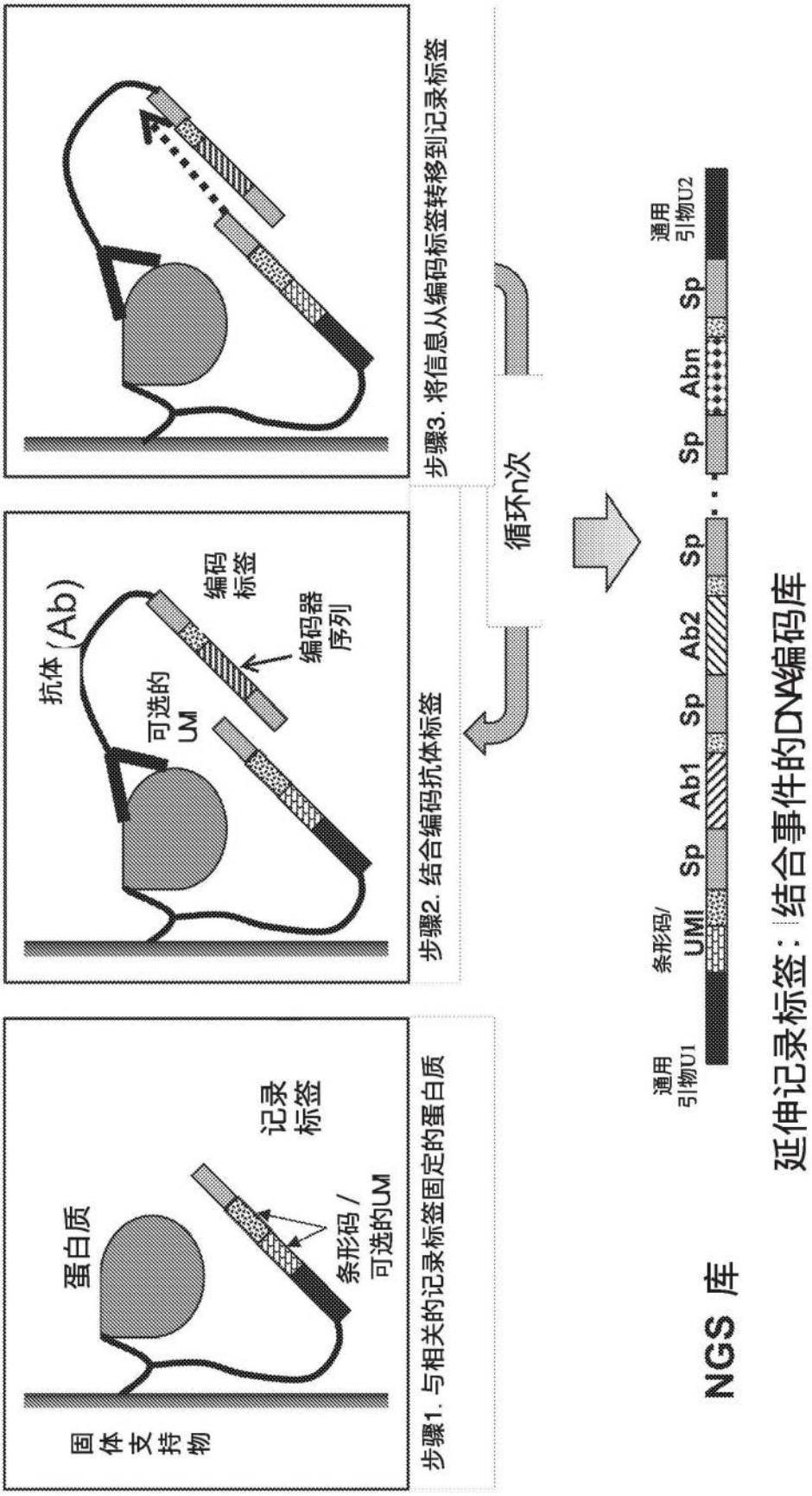

固体支持物
蛋白质
记录标签
条形码 / 可选的UM
步骤1. 与相关的记录标签固定的蛋白质
抗体 (Ab)
可选的UM
编码标签
编码器序列
步骤2. 结合编码抗体标签
步骤3. 将信息从编码标签转移到记录标签
循环n次
通用引物U1
条形码/UMI
Sp
Ab1
Sp
Ab2
Sp
Sp
Abn
Sp
通用引物U2
NGS 库
延伸记录标签：结合事件的DNA编码库

图2A

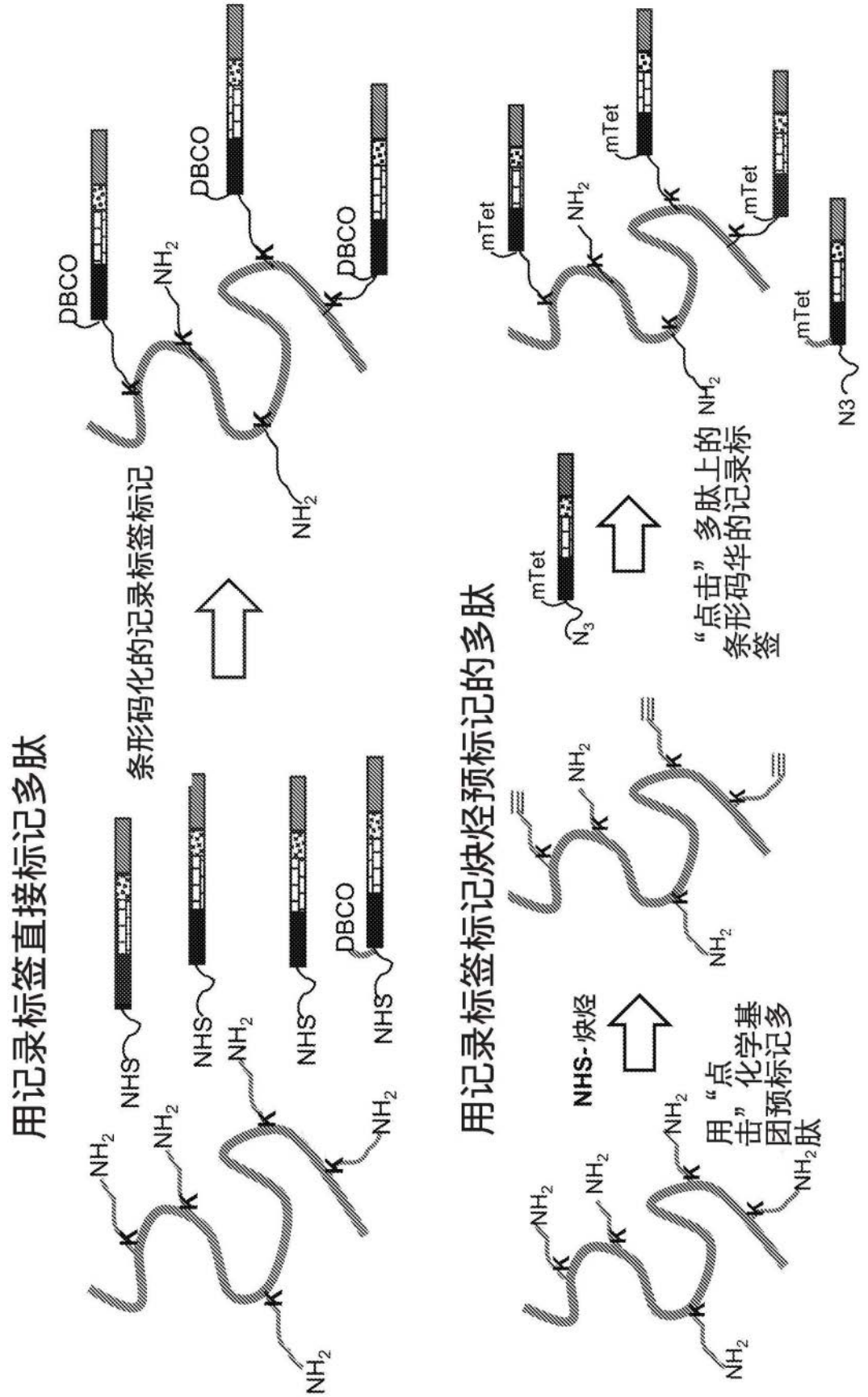

用记录标签直接标记多肽
条形码化的记录标签标记
用记录标签标记炔烃预标记的多肽
NHS-炔烃
用“点击”化学基团预标记多肽
“点击”多肽上的条形码华的记录标签
DBCO
NHS
mTet
N3
NH2

图2B

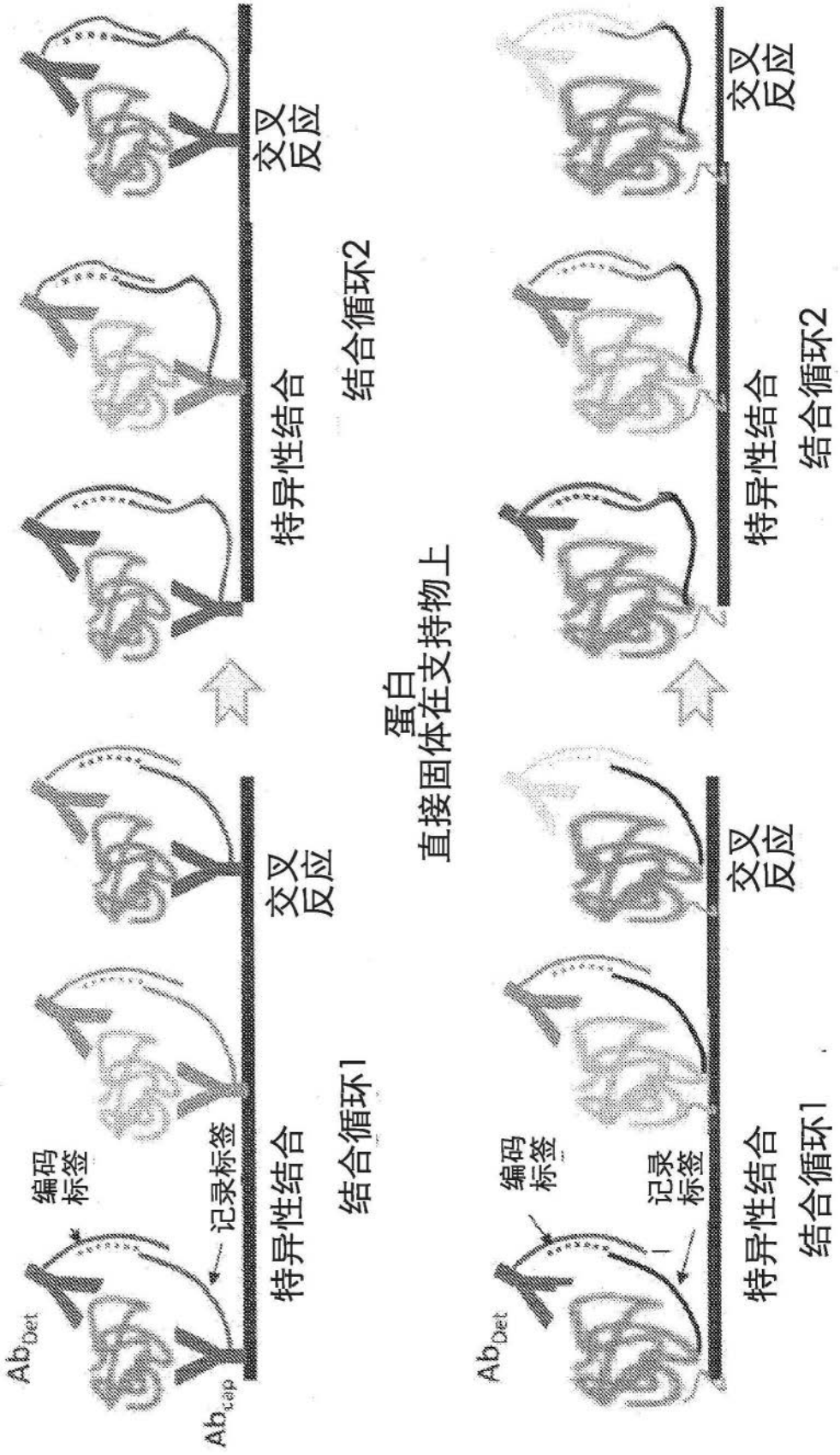

用捕获剂将蛋白
固定在固体支持物上
蛋白
直接固体在支持物上
编码
标签
记录标签
$Ab_{Det}$
$Ab_{cap}$
特异性结合
交叉
反应
结合循环1
结合循环2

图2C

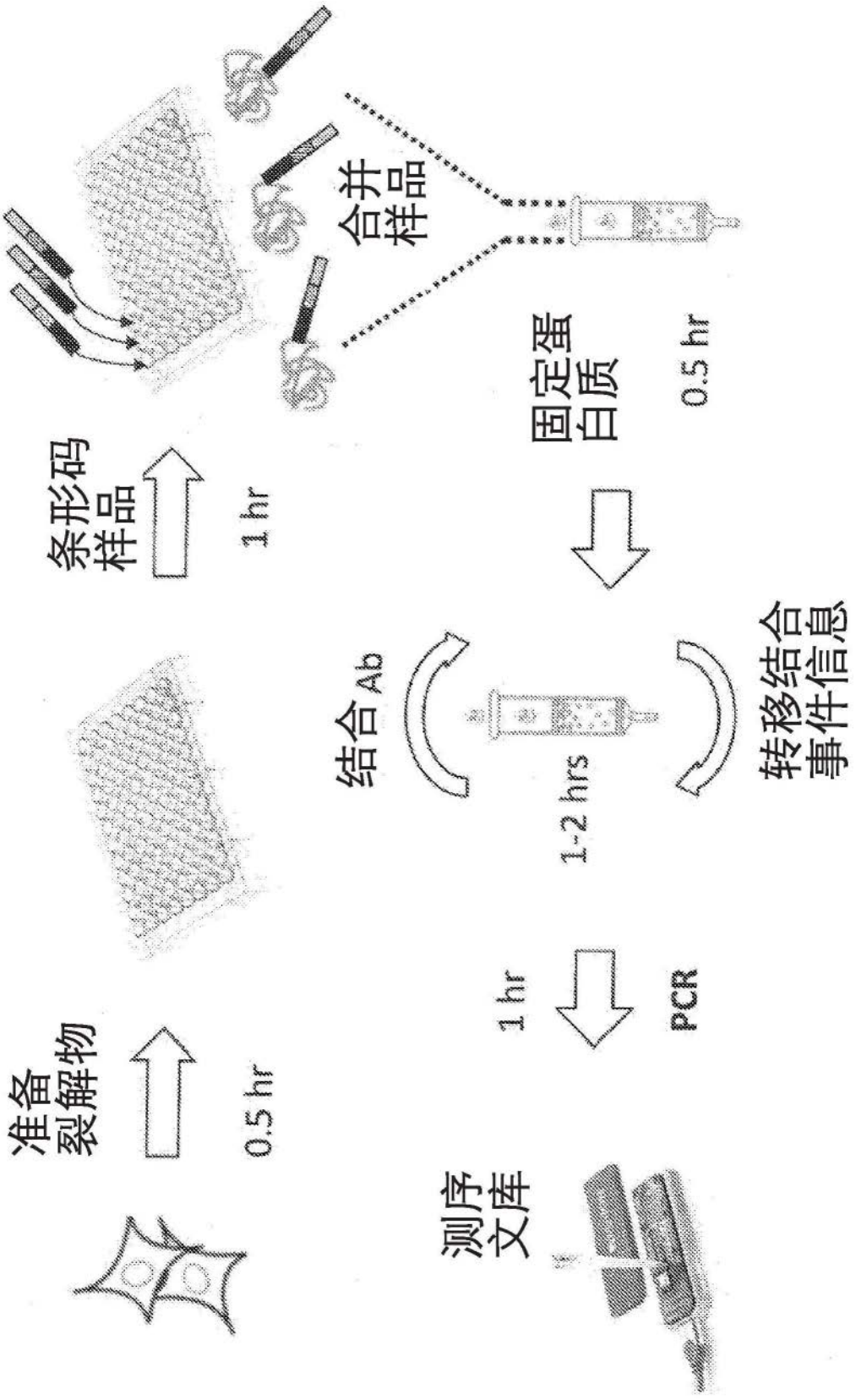

合并样品
固定蛋白质
0.5 hr
条形码样品
1 hr
结合 Ab
1-2 hrs
转移结合事件信息
准备裂解物
0.5 hr
1 hr
PCR
测序文库

图2D

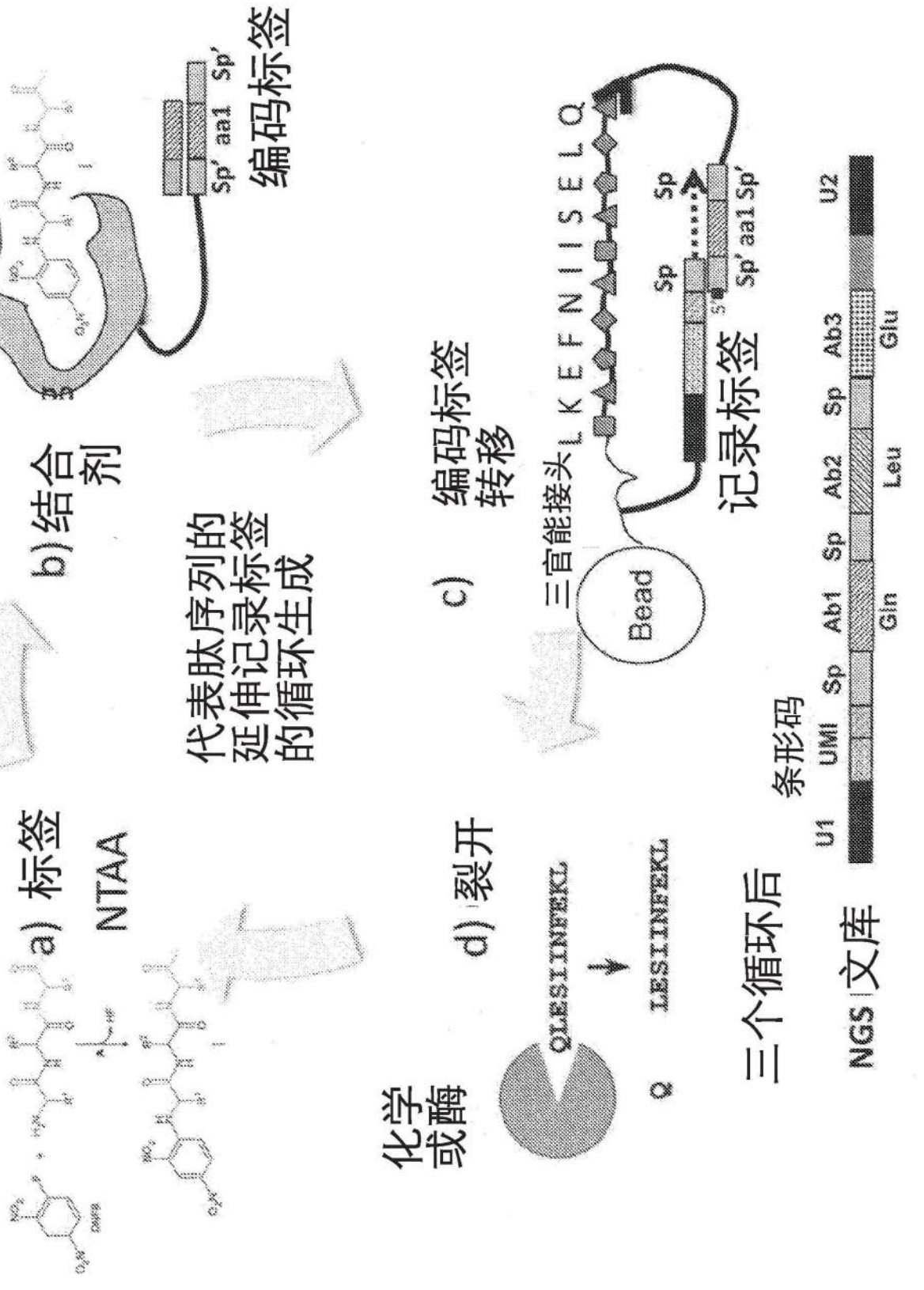

a) 标签
NTAA
b)结合剂
Sp' aa1 Sp'
编码标签
代表肽序列的延伸记录标签的循环生成
c) 编码标签转移
三官能接头
L K E F N I I S E L Q
Bead
Sp
Sp' aa1 Sp'
记录标签
d) 裂开
化学或酶
QLESTINFEKL
Q LESTINFEKL
三个循环后
条形码
U1 UMI Sp Ab1 Sp Ab2 Sp Ab3 U2
Gln Leu Glu
NGS文库

图3

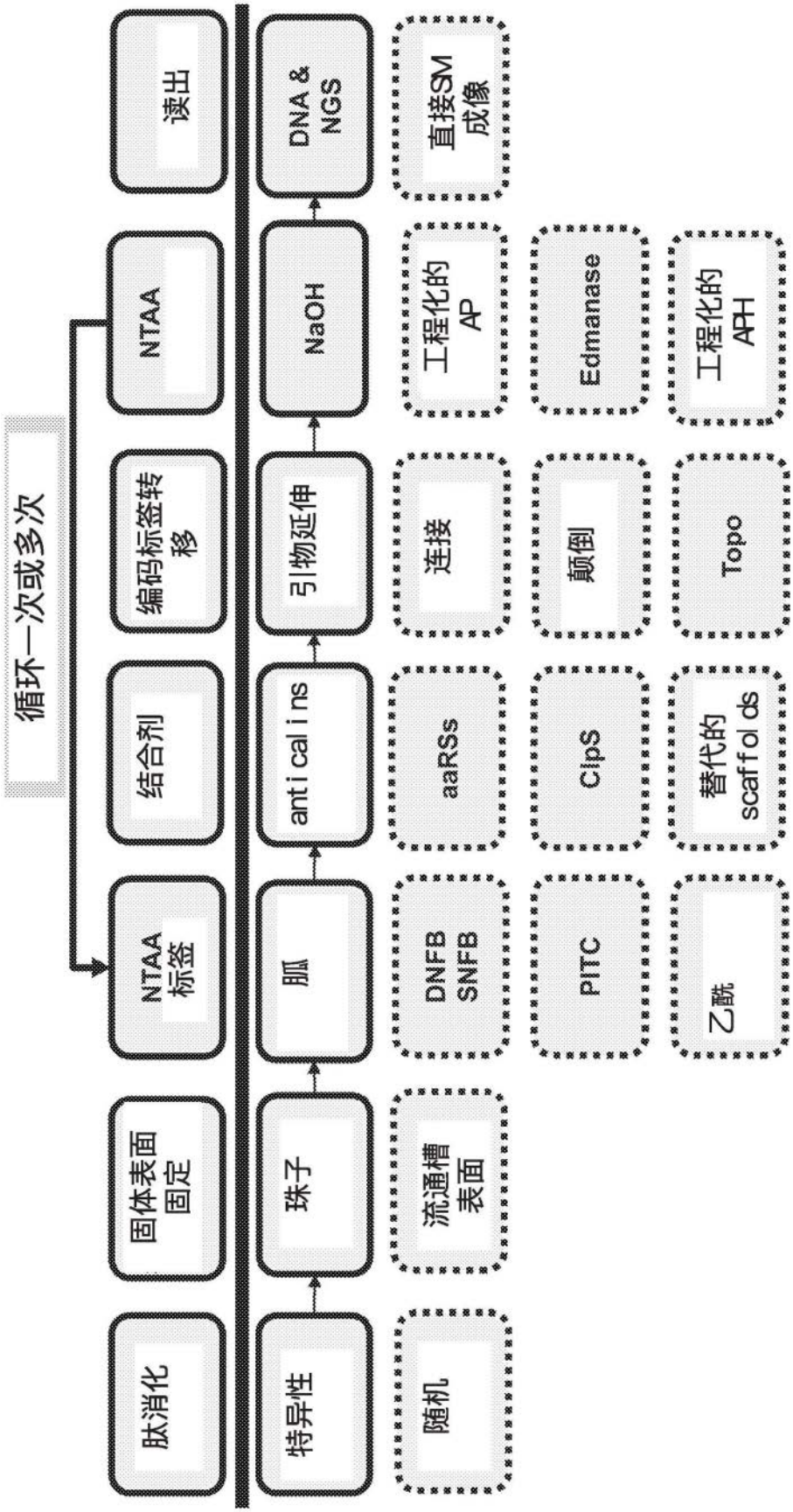

循环一次或多次
肽消化
固体表面固定
NTAA标签
结合剂
编码标签转移
NTAA
读出
特异性
珠子
胍
anticalins
引物延伸
NaOH
DNA & NGS
随机
流通槽表面
DNFB
SNFB
aaRSs
连接
工程化的AP
直接SM成像
PITC
ClpS
颠倒
Edmanase
乙酰
替代的scaffolds
Topo
工程化的APH

图4A

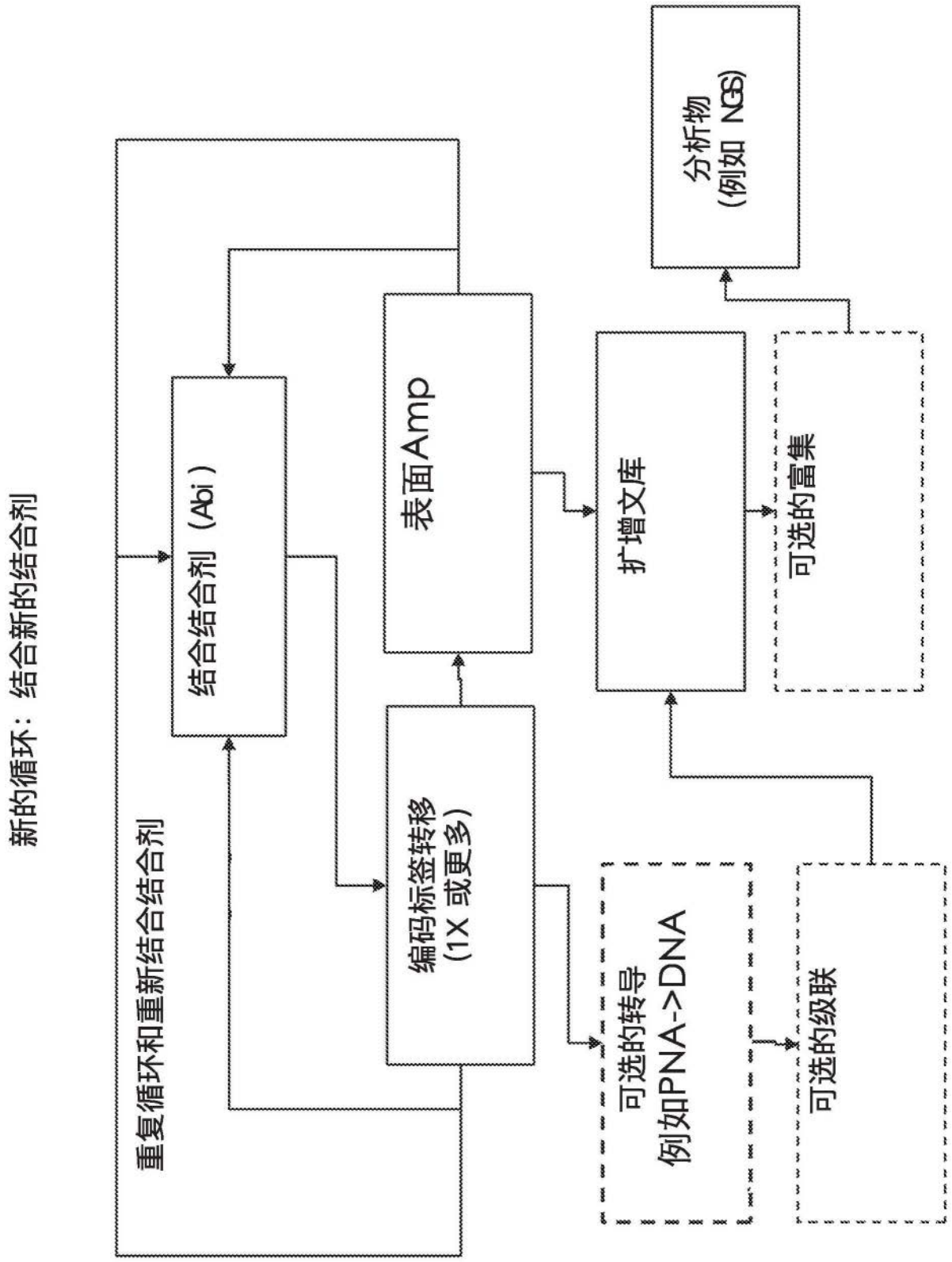

新的循环：结合新的结合剂
重复循环和重新结合结合剂
结合结合剂（Abi）
编码标签转移
（1X 或更多）
表面Amp
扩增文库
可选的富集
分析物
（例如 NGS）
可选的转导
例如PNA->DNA
可选的级联

图4B

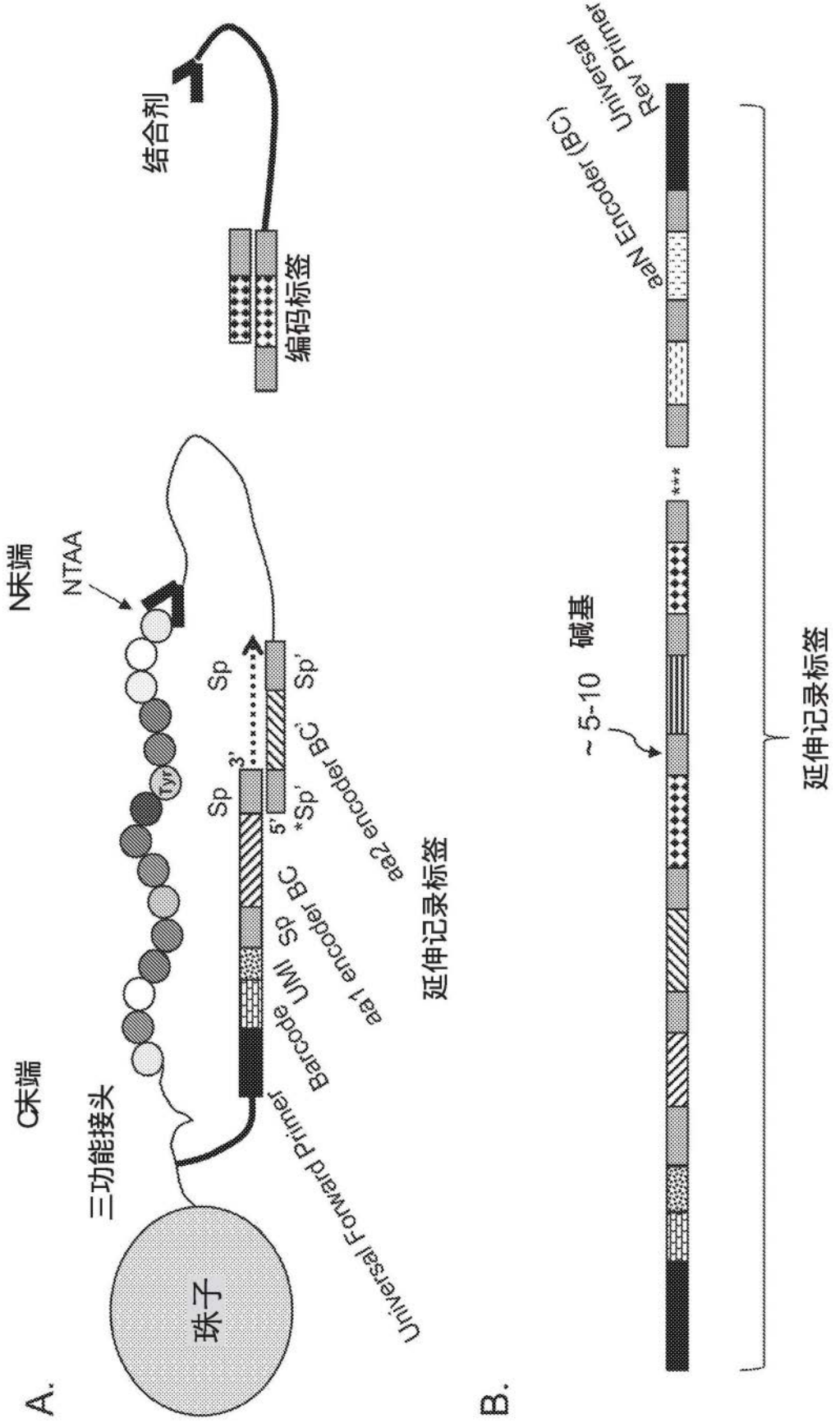

A.
B.
结合剂
编码标签
N末端
NTAA
C末端
三功能接头
珠子
Universal Forward Primer
Barcode UMI Sp
aa1 encoder BC
aa2 encoder BC
Sp
Sp'
延伸记录标签
aaN Encoder (BC)
Universal Rev Primer
~ 5-10 碱基
延伸记录标签

图5

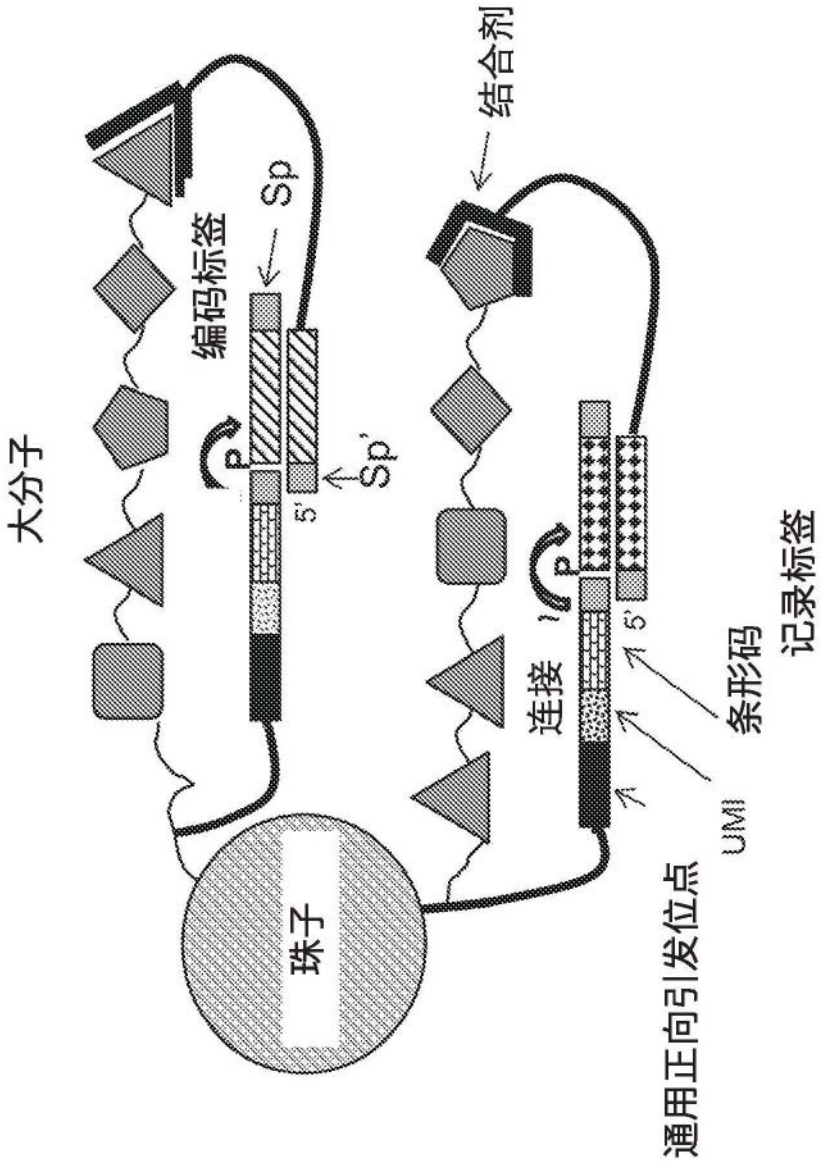



图6

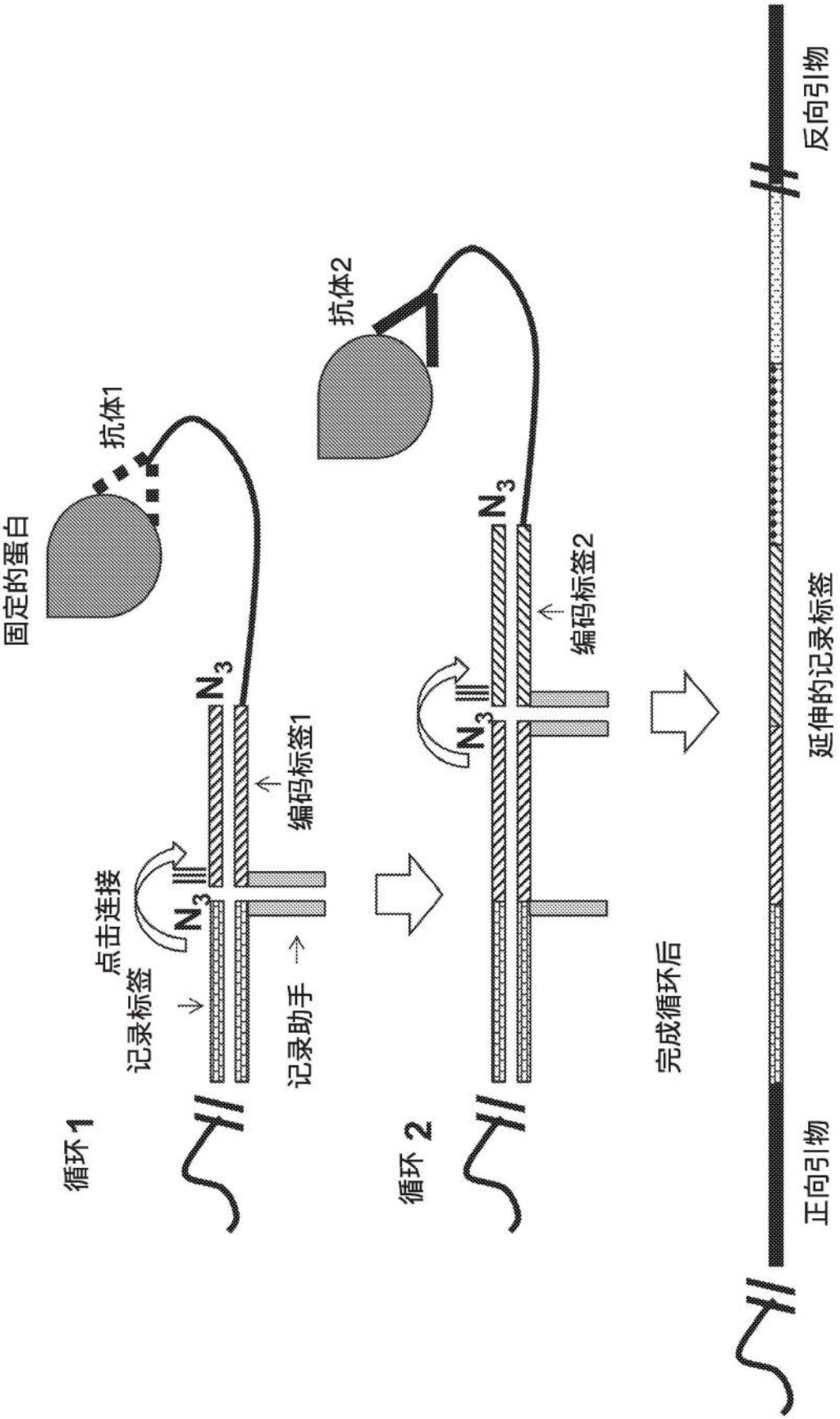

固定的蛋白
抗体1
抗体2
点击连接
记录标签
记录助手
编码标签1
编码标签2
循环 1
循环 2
完成循环后
正向引物
反向引物
延伸的记录标签

图7

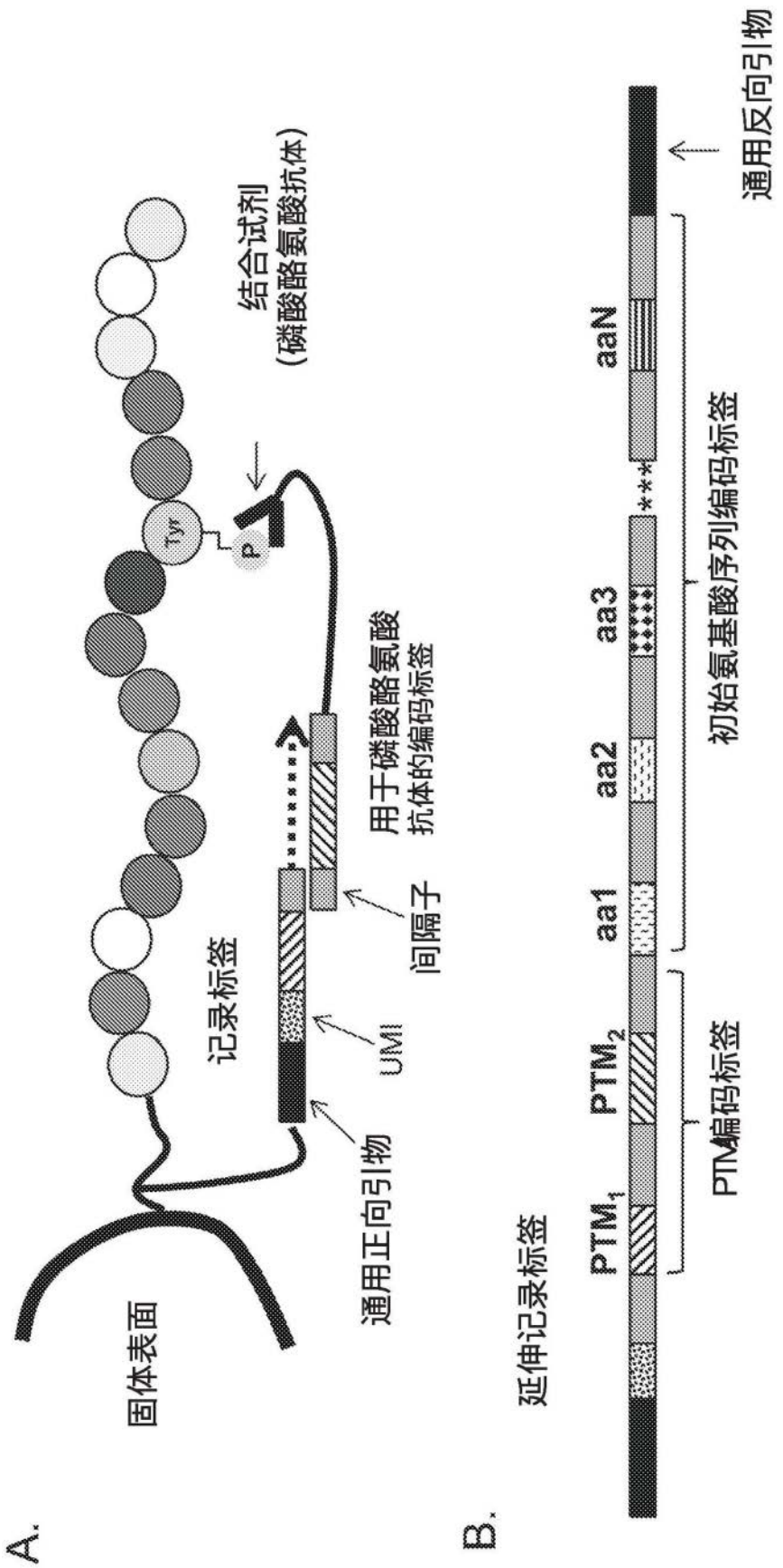

A.
固体表面
记录标签
Tyr
P
结合试剂
(磷酸酪氨酸抗体)
通用正向引物
UMI
间隔子
用于磷酸酪氨酸
抗体的编码标签
B.
延伸记录标签
PTM1
PTM2
aa1
aa2
aa3
aaN
PTM编码标签
初始氨基酸序列编码标签
通用反向引物

图8

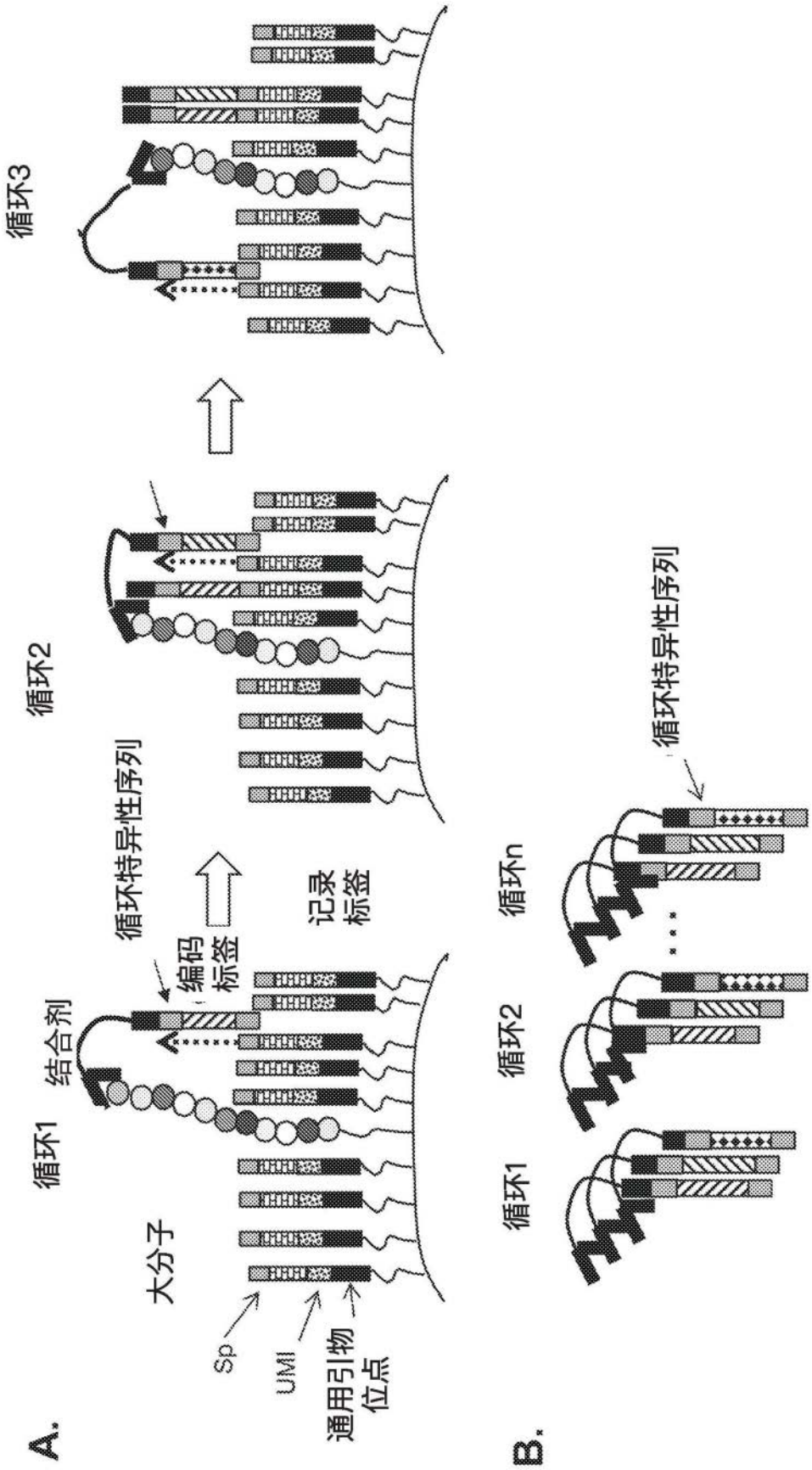

A.
循环1
结合剂
编码标签
循环特异性序列
记录标签
循环2
循环3
大分子
Sp
UMI
通用引物位点
B.
循环1
循环2
循环n
循环特异性序列

图9

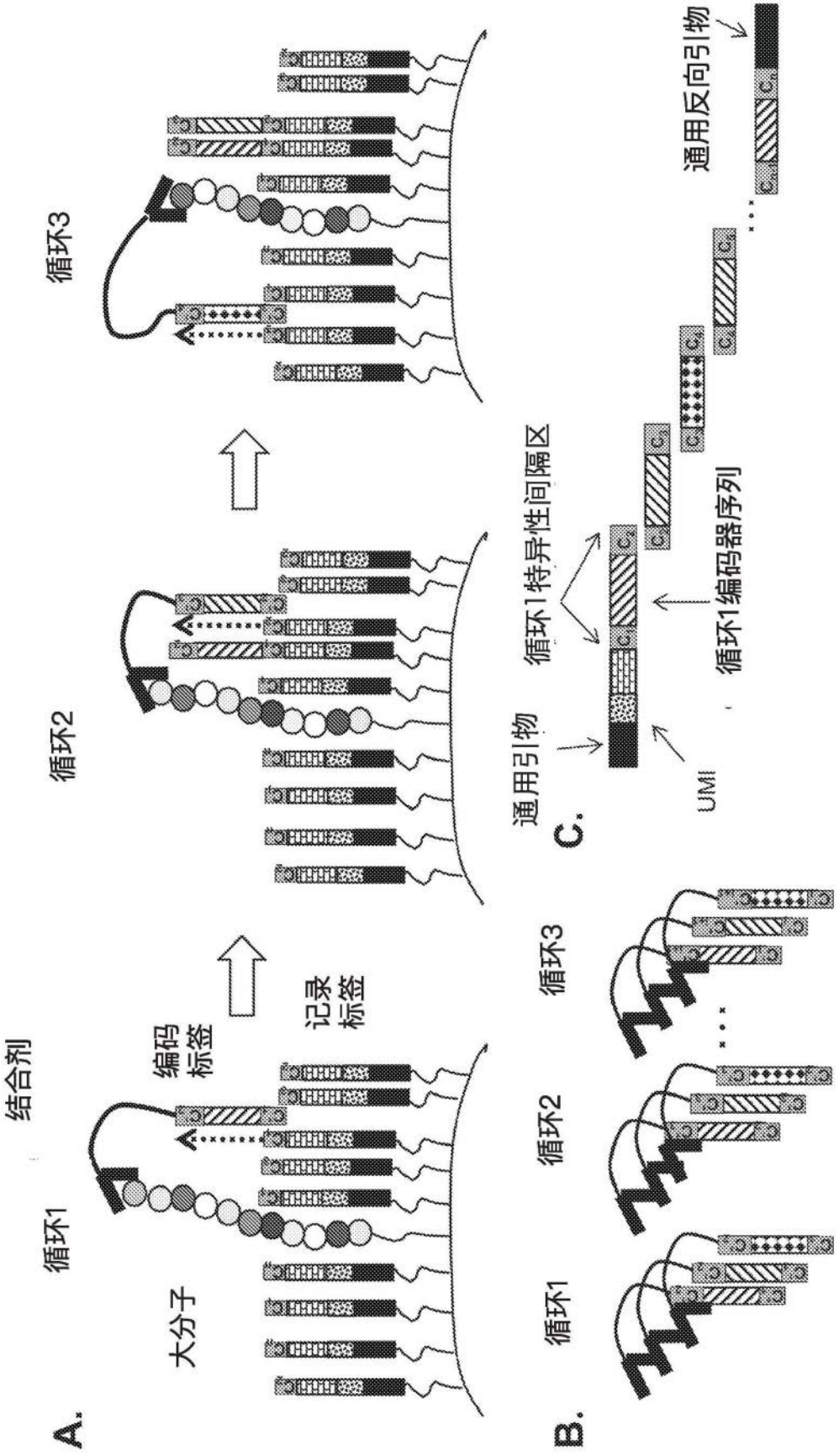

A.
循环1
结合剂
大分子
编码标签
记录标签
循环2
循环3
B.
循环1
循环2
循环3
C.
通用引物
循环1特异性间隔区
UMI
循环1编码器序列
通用反向引物

图10

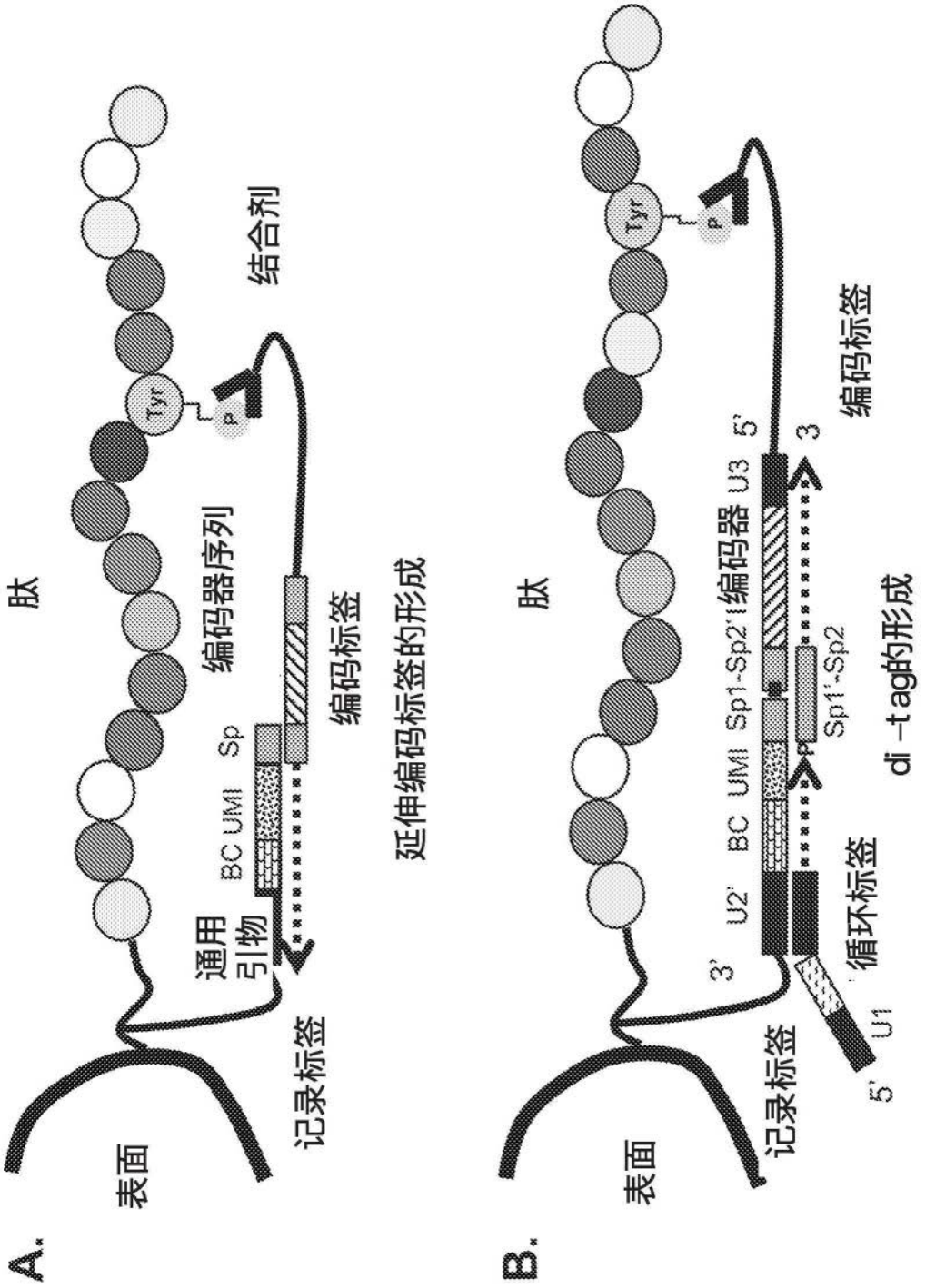

A.
表面
肽
Tyr
P
结合剂
编码器序列
通用
引物
BC UMI
Sp
编码标签
记录标签
延伸编码标签的形成
B.
表面
肽
Tyr
P
3'
U2'
BC
UMI
Sp1-Sp2'
编码器
U3
5'
3
编码标签
Sp1'-Sp2
循环标签
U1
5'
记录标签
di –t ag的形成

图11

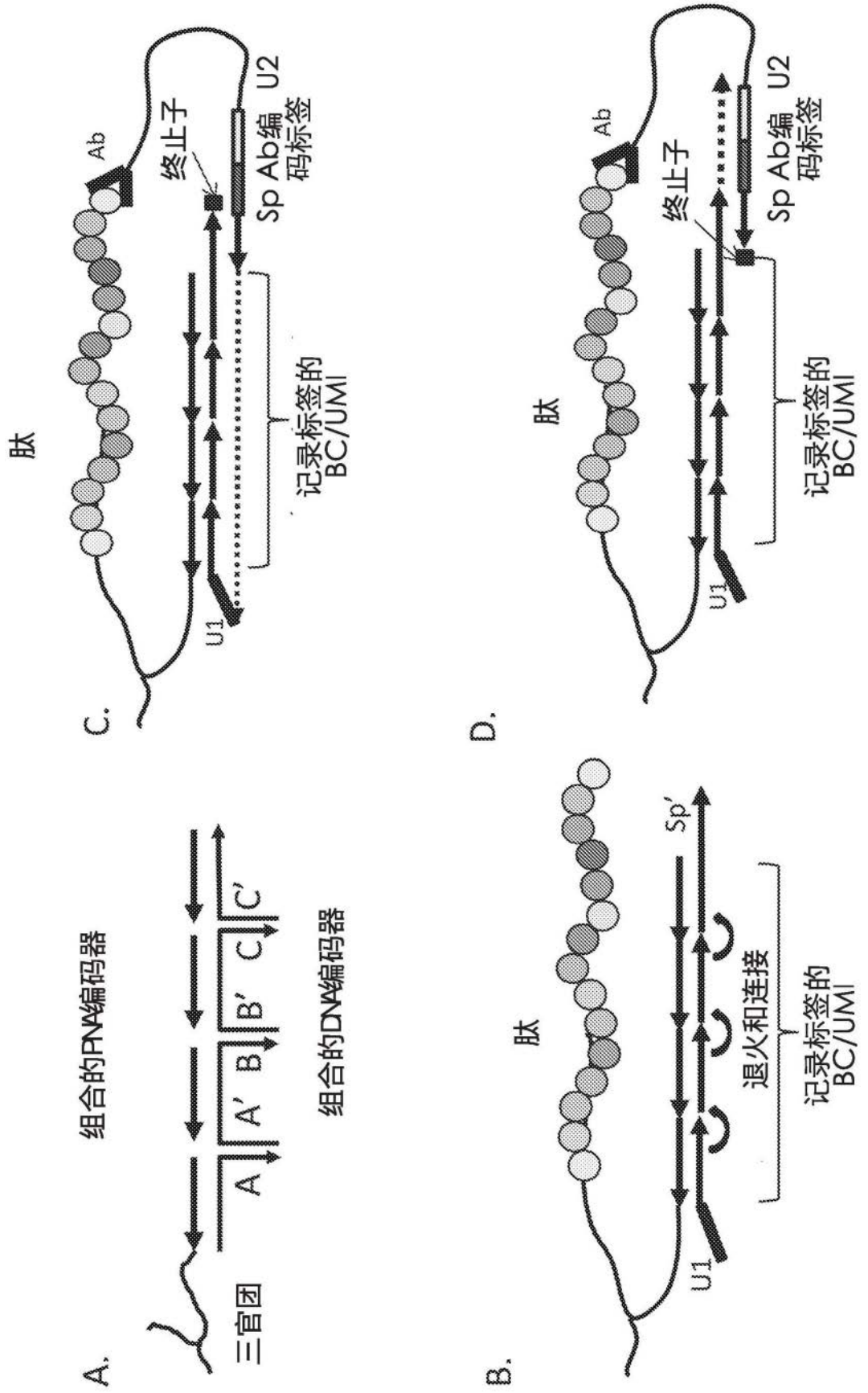

A.
组合的DNA编码器
三官团
A
A'
B
B'
C
C'
组合的DNA编码器
B.
肽
Sp'
U1
退火和连接
记录标签的
BC/UMI
C.
肽
Ab
终止子
Sp
Ab编
码标签
U2
U1
记录标签的
BC/UMI
D.
肽
Ab
终止子
Sp
Ab编
码标签
U2
U1
记录标签的
BC/UMI

图12

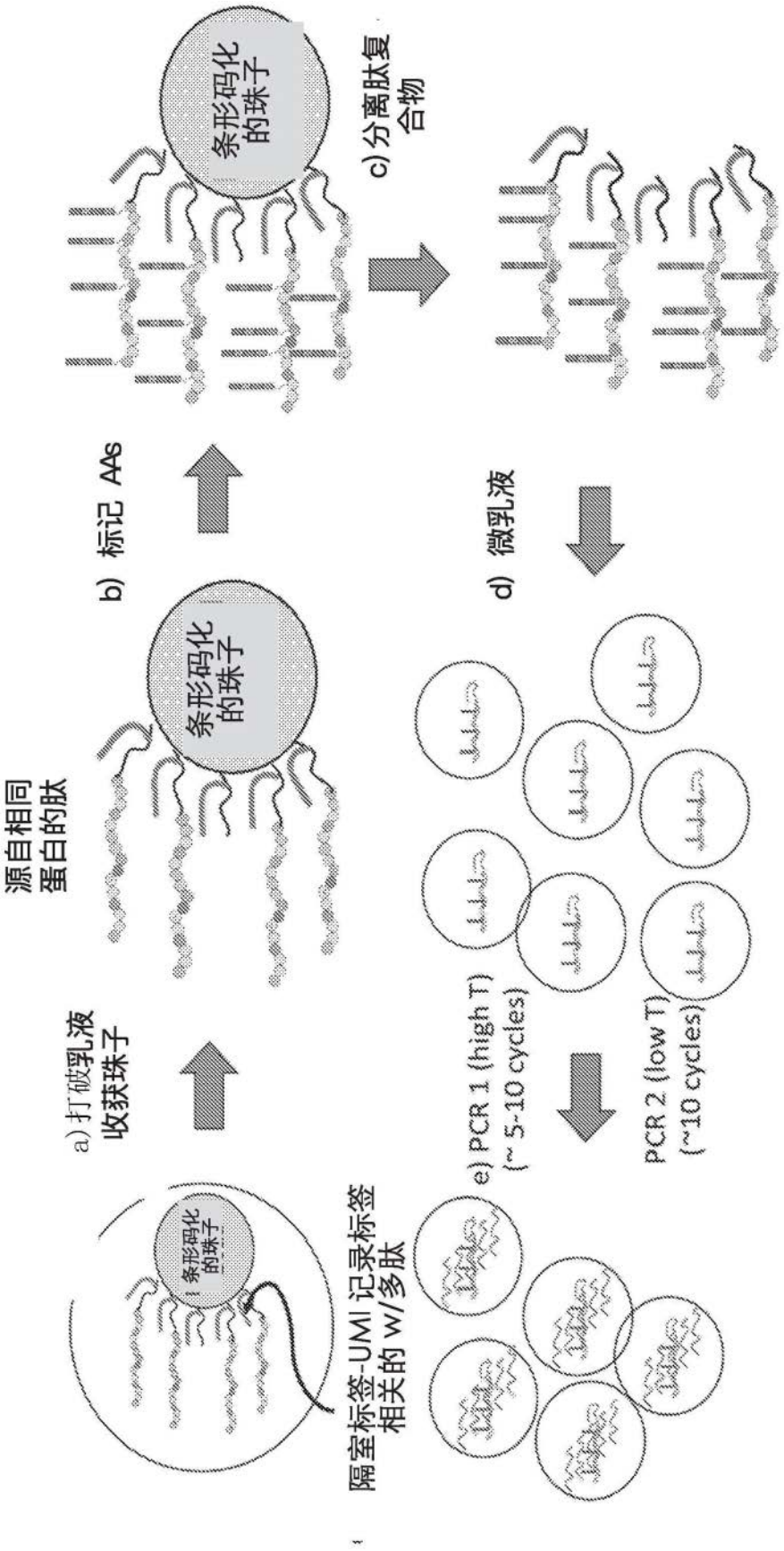

条形码化的珠子
c) 分离肽复合物
b) 标记 AAs
条形码化的珠子
源自相同蛋白的肽
d) 微乳液
a) 打破乳液收获珠子
条形码化的珠子
e) PCR 1 (high T)
(~ 5-10 cycles)
PCR 2 (low T)
(~10 cycles)
隔室标签-UMI 记录标签相关的 w/多肽

图13

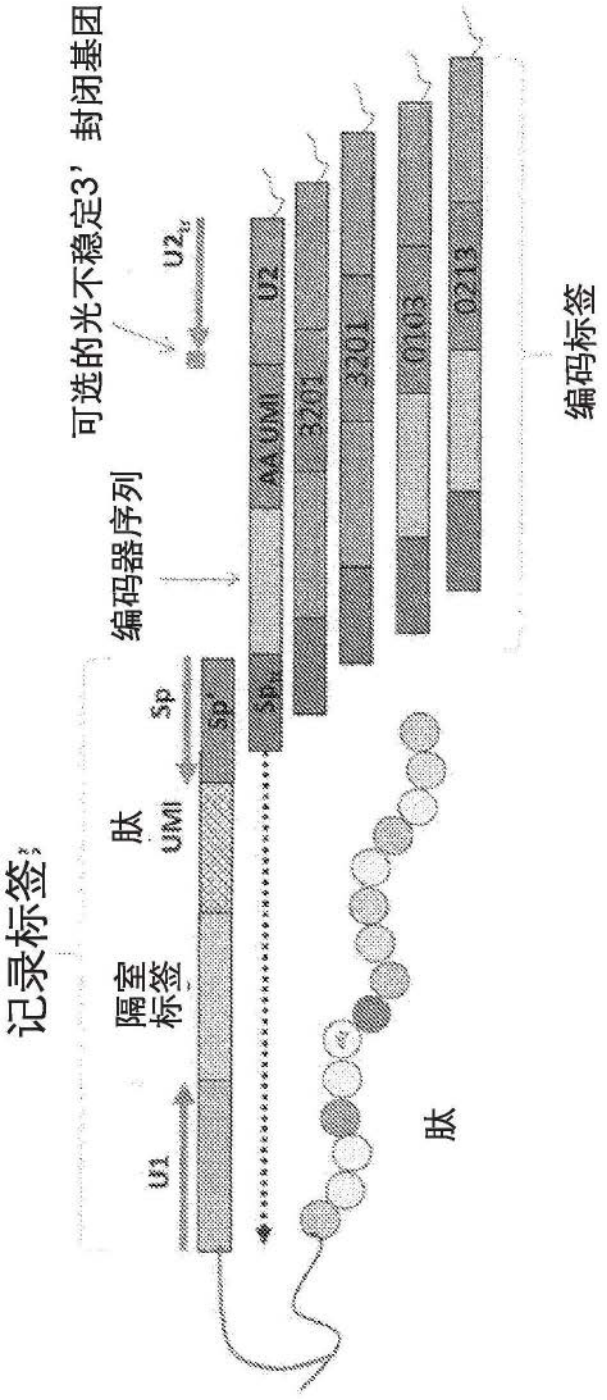

记录标签;
隔室标签
肽
UMI
Sp
Sp'
U1
编码器序列
可选的光不稳定3'封闭基团
U2
AA UMI
3201
0103
0213
编码标签
肽

图14

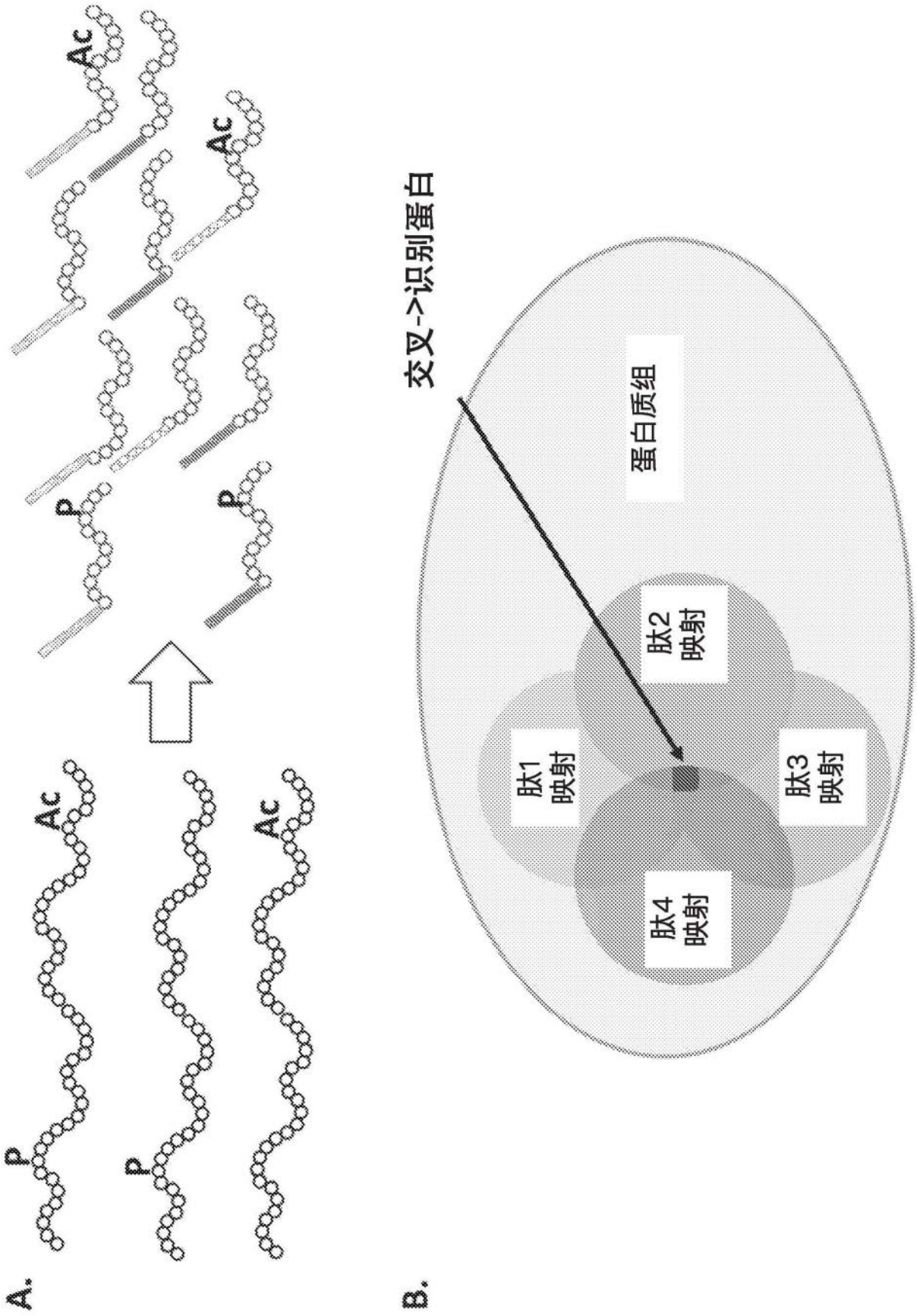



图15

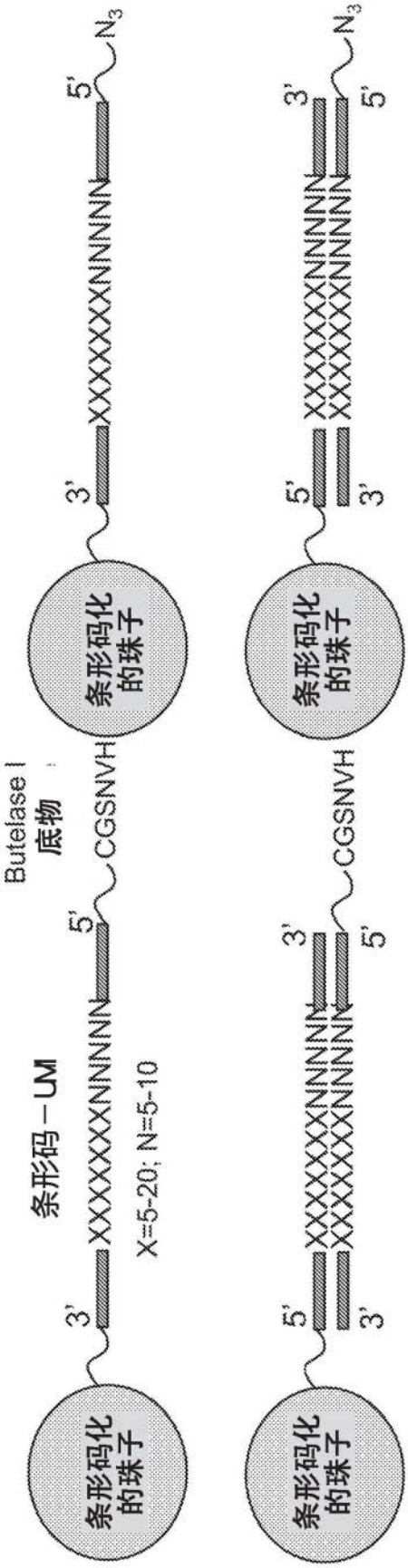

条形码化的珠子
条形码化的珠子
条形码化的珠子
条形码化的珠子
3'
5'
条形码－UM
XXXXXXXNNNNNN
X=5-20; N=5-10
Butelase I
底物
CGSNVH
5'
3'
XXXXXXXNNNNNN
$N_3$

图16

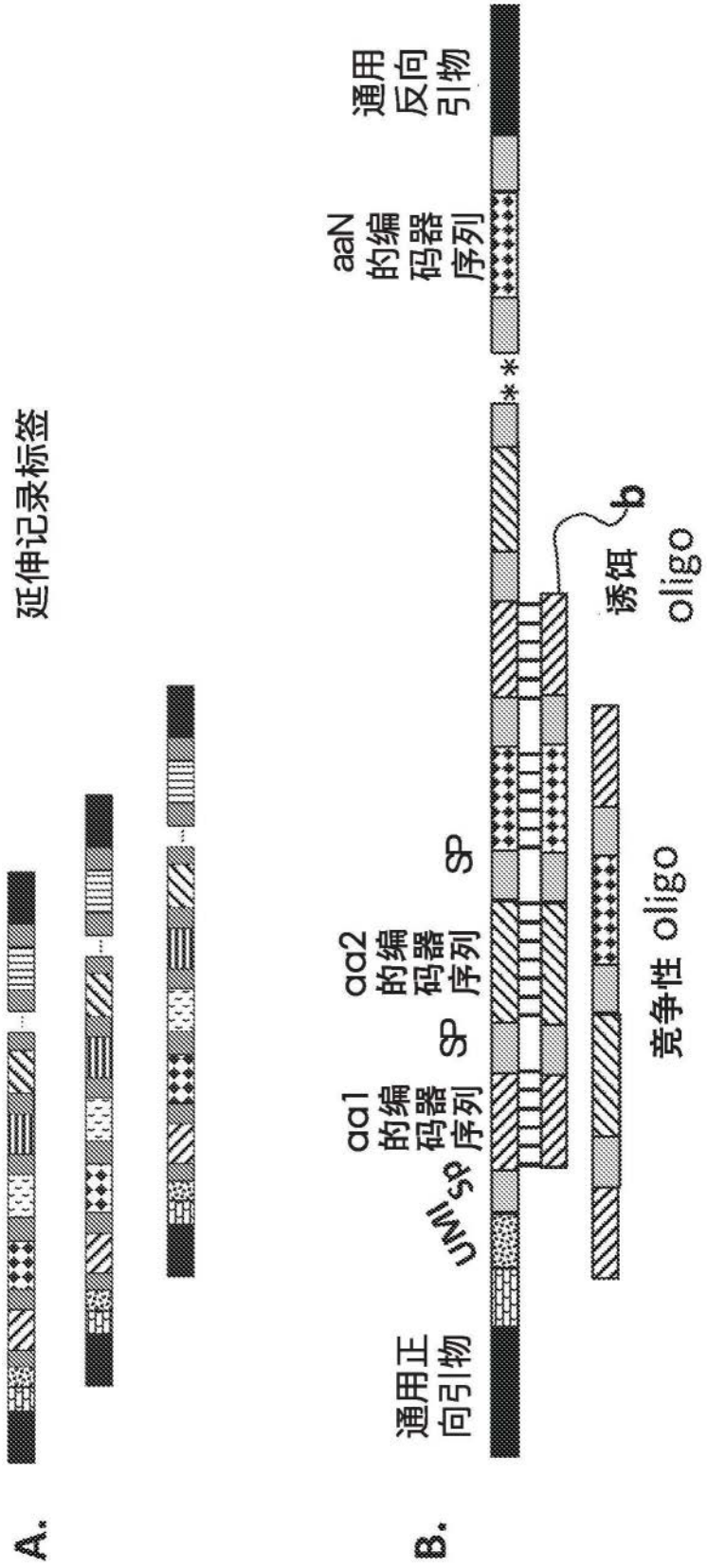

A.
延伸记录标签
B.
通用正向引物
UMI
SP
aa1的编码器序列
SP
aa2的编码器序列
SP
aaN的编码器序列
通用反向引物
竞争性 oligo
诱饵 oligo

图17

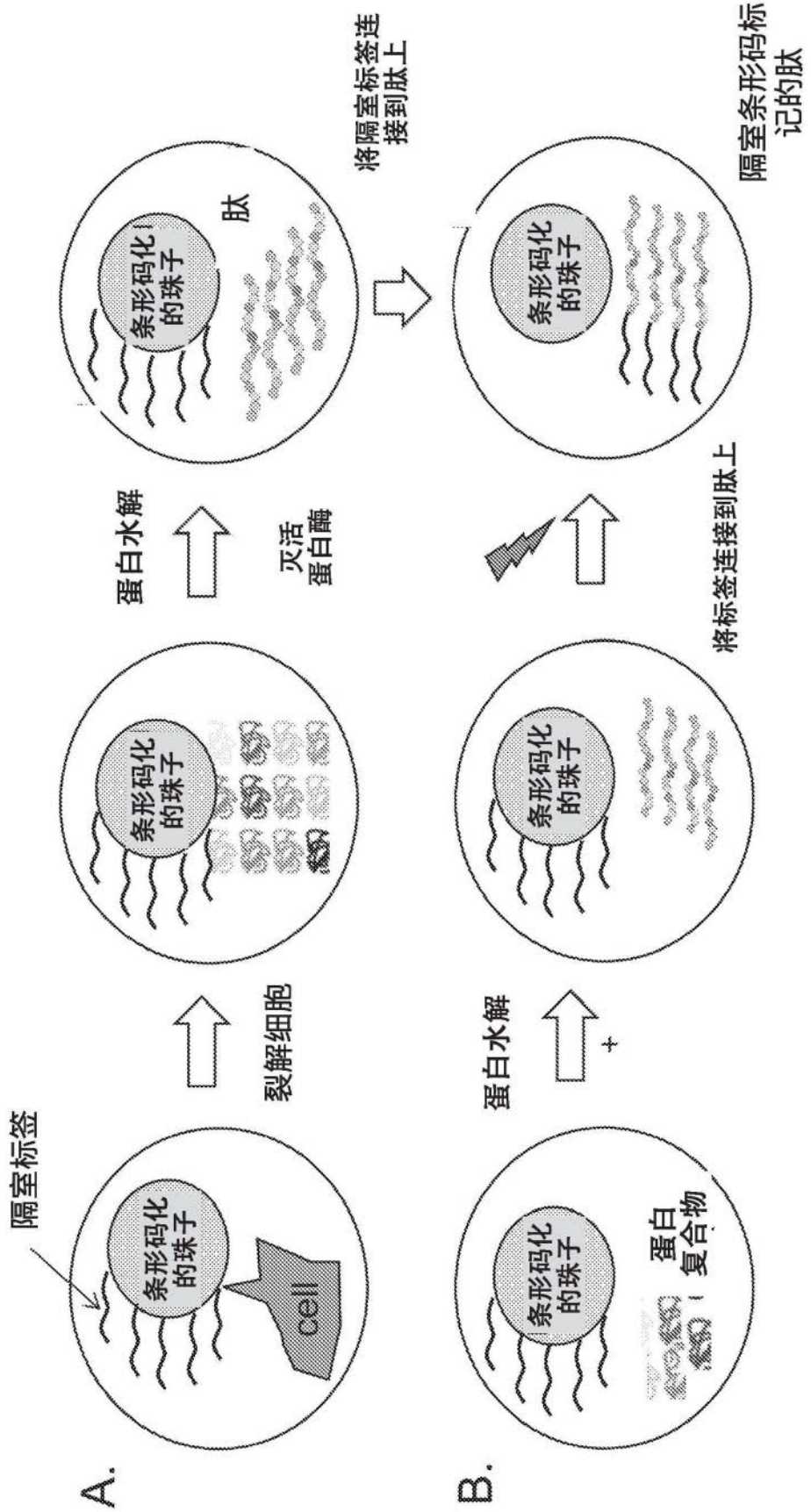

A.
隔室标签
条形码化的珠子
cell
裂解细胞
蛋白水解
灭活蛋白酶
肽
将隔室标签连接到肽上
隔室条形码标记的肽
B.
条形码化的珠子
蛋白复合物
蛋白水解
+
将标签连接到肽上

图18

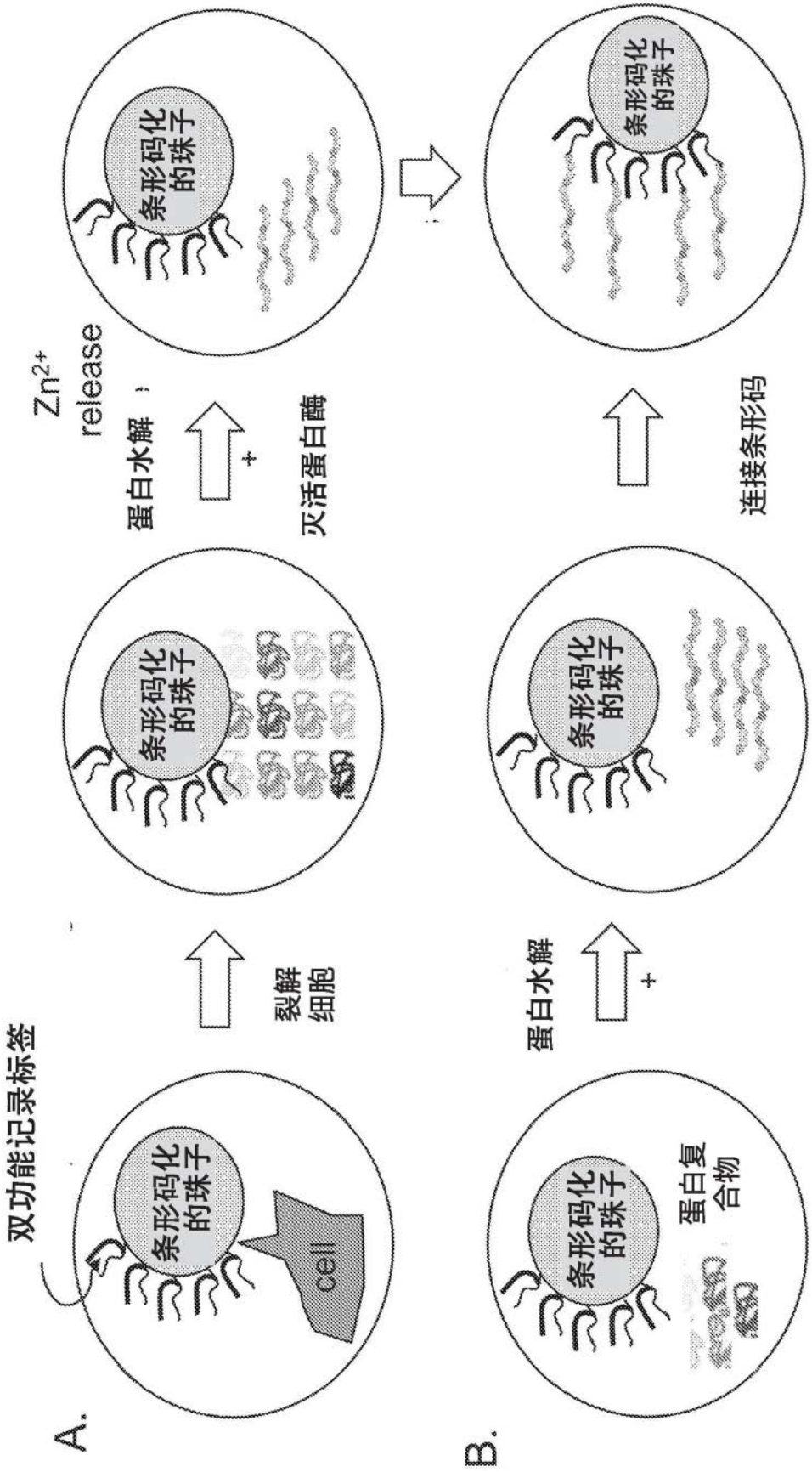

双功能记录标签
条形码化的珠子
cell
裂解细胞
蛋白水解
+
灭活蛋白酶
$Zn^{2+}$ release
连接条形码
蛋白复合物
A.
B.

图19

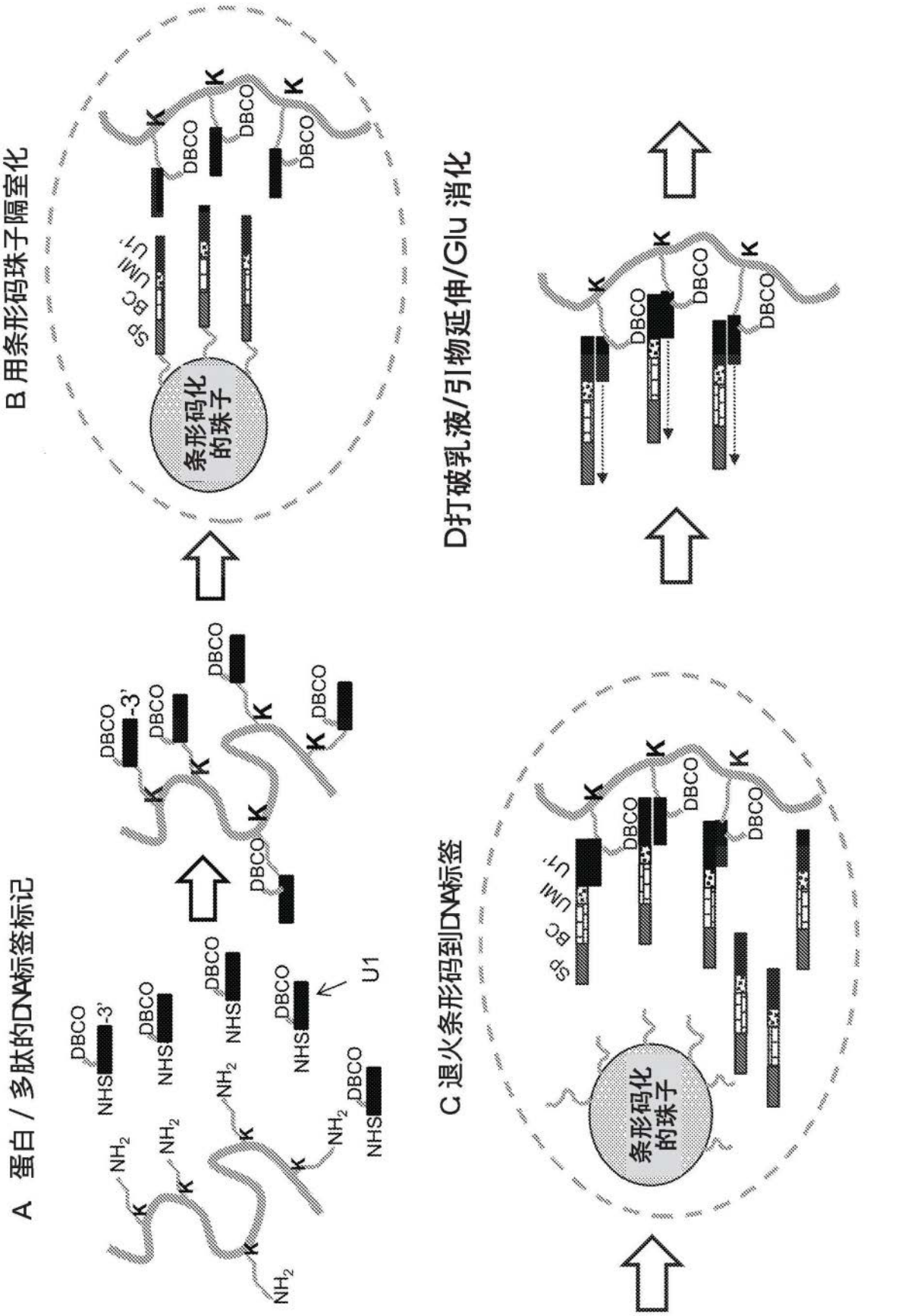

A 蛋白 / 多肽的DNA标签标记
B 用条形码珠子隔室化
C 退火条形码到DNA标签
D打破乳液/引物延伸/Glu 消化
条形码化的珠子
DBCO
NHS
NH2
K
U1
-3'

图20A-D

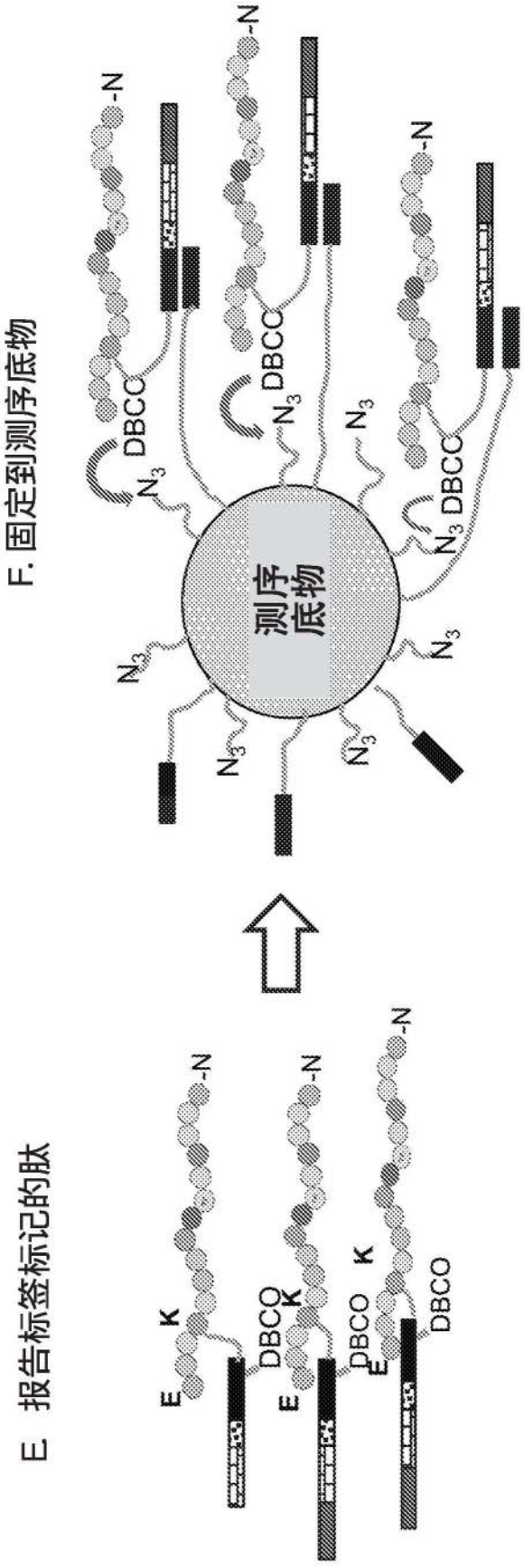



图20E-F

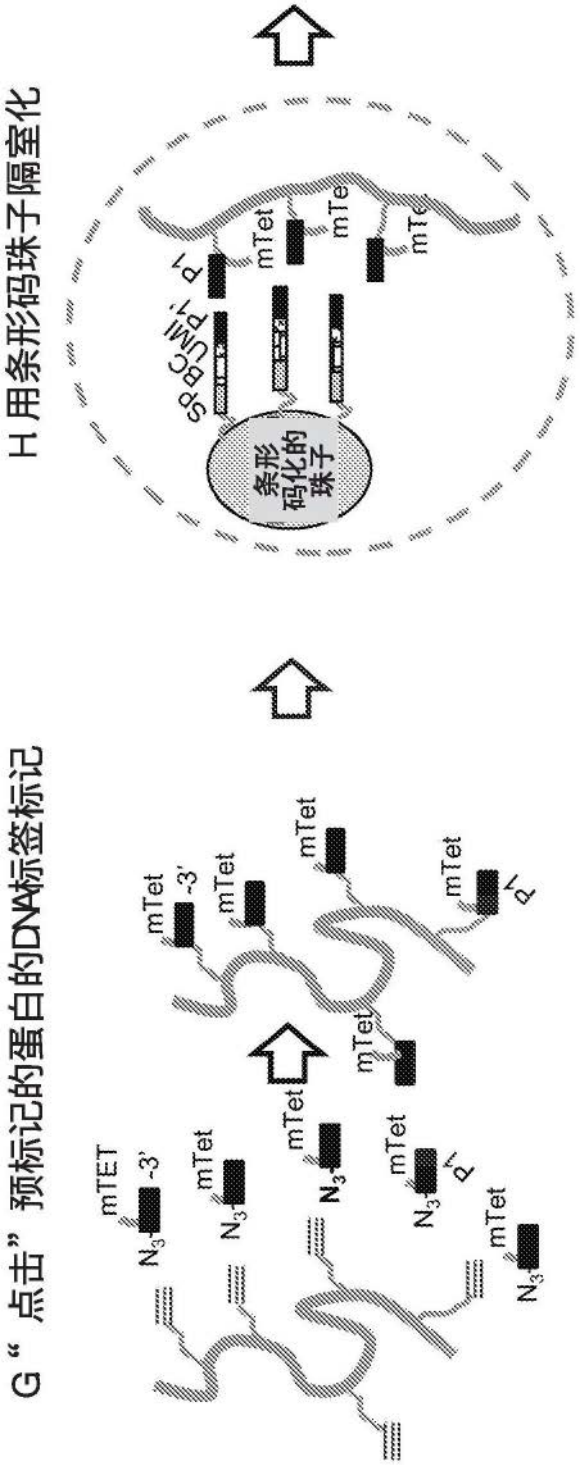



图20G-H

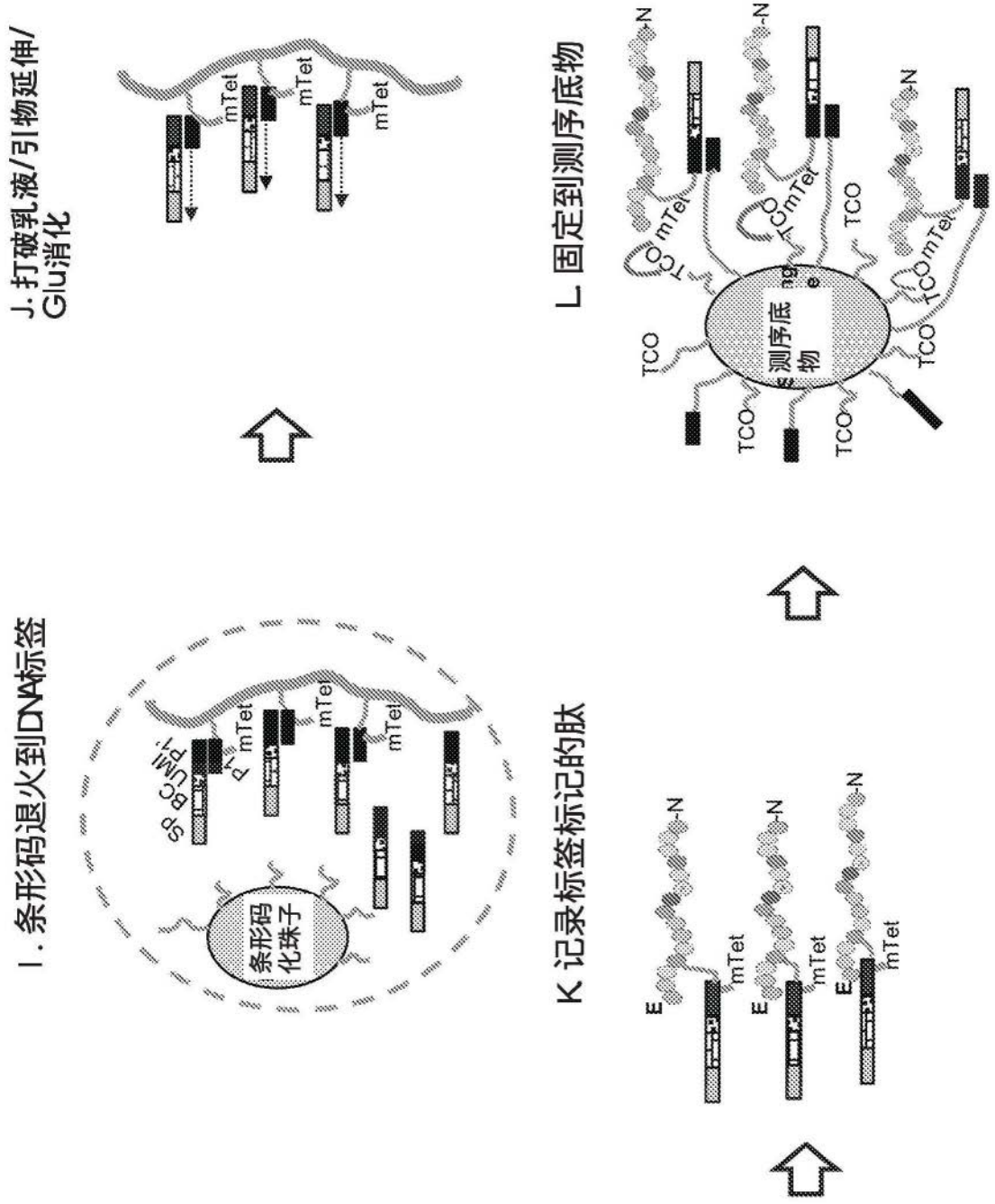

I. 条形码退火到DNA标签
条形码化珠子
mTet
J. 打破乳液/引物延伸/Glu消化
K 记录标签标记的肽
E
N
L 固定到测序底物
测序底物
TCO

图20I-L

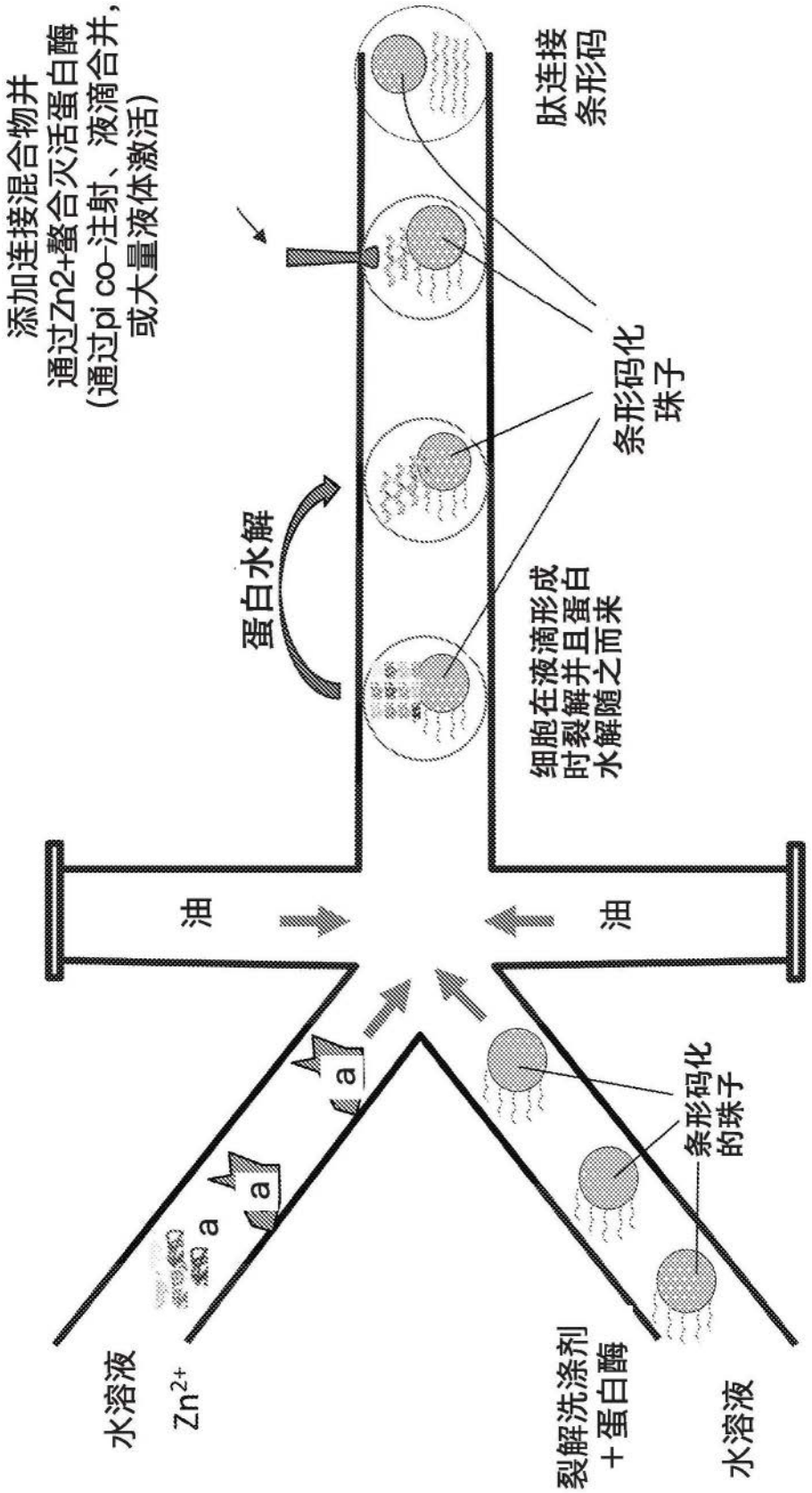

添加连接混合物并
通过Zn2+螯合灭活蛋白酶
(通过pi co-注射、液滴合并,
或大量液体激活)
肽连接
条形码
条形码化
珠子
蛋白水解
细胞在液滴形成
时裂解并且蛋白
水解随之而来
油
油
a
a
a
条形码化
的珠子
水溶液
$Zn^{2+}$
裂解洗涤剂
+蛋白酶
水溶液

图21

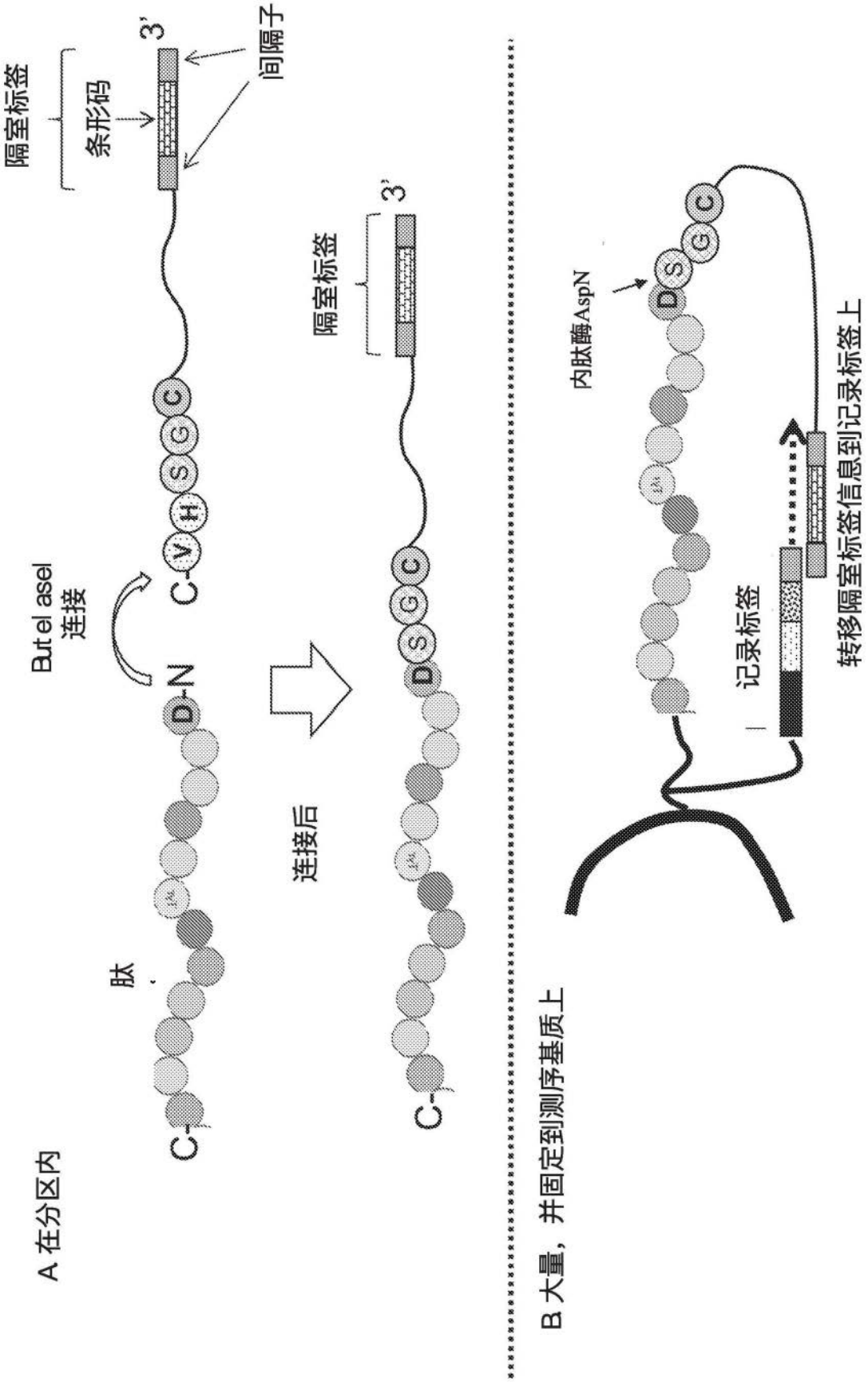

隔室标签
条形码
3'
间隔子
But el asel
连接
肽
C-
D-N
C-V H S G C
连接后
隔室标签
3'
D S G C
C-
A 在分区内
B 大量，并固定到测序基质上
内肽酶AspN
D S G C
记录标签
转移隔室标签信息到记录标签上

图22

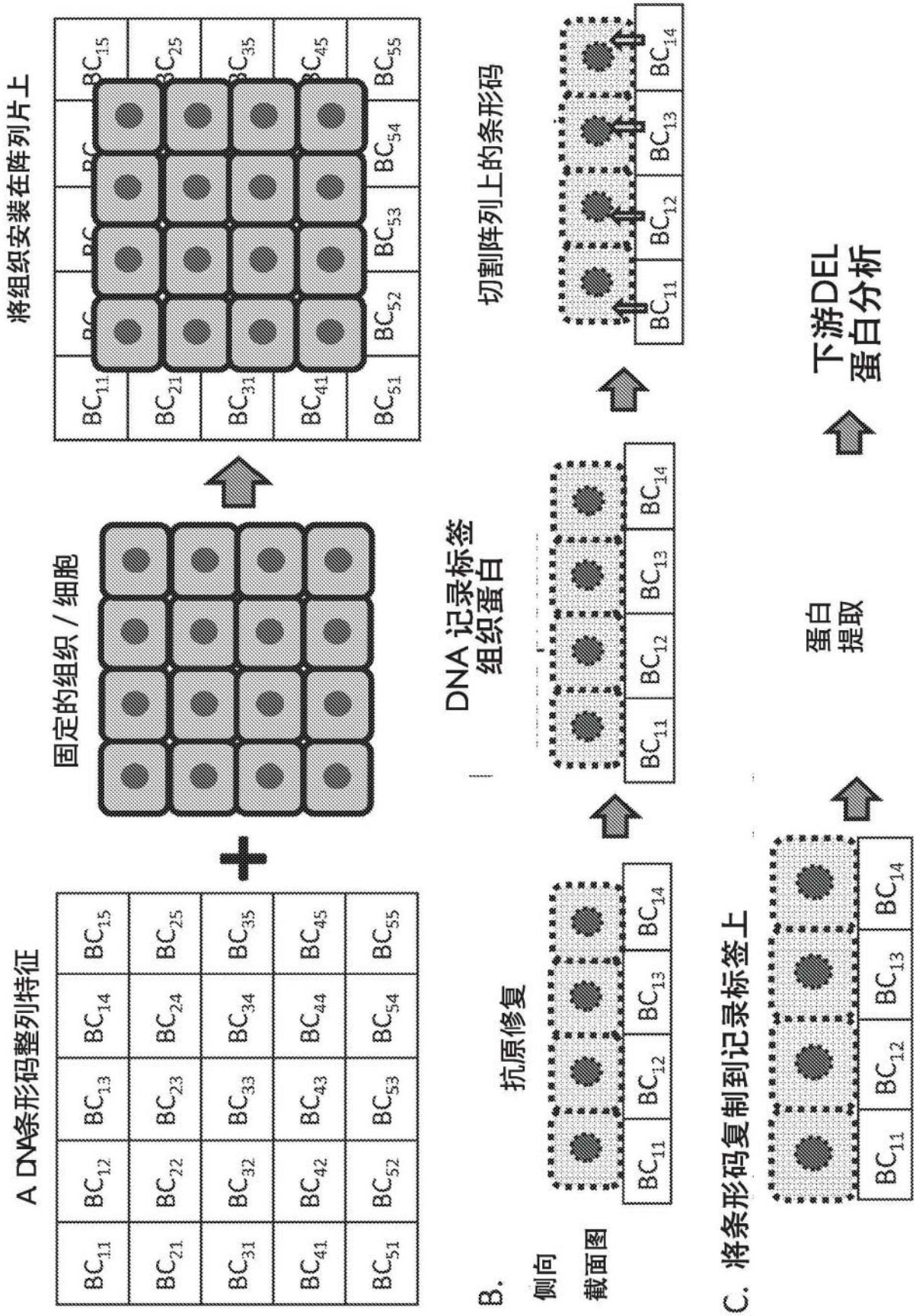

A DNA条形码整列特征
固定的组织 / 细胞
将组织安装在阵列片上
DNA 记录标签
组织蛋白
切割阵列上的条形码
B.
侧向
截面图
抗原修复
蛋白
提取
下游DEL
蛋白分析
C. 将条形码复制到记录标签上

图23

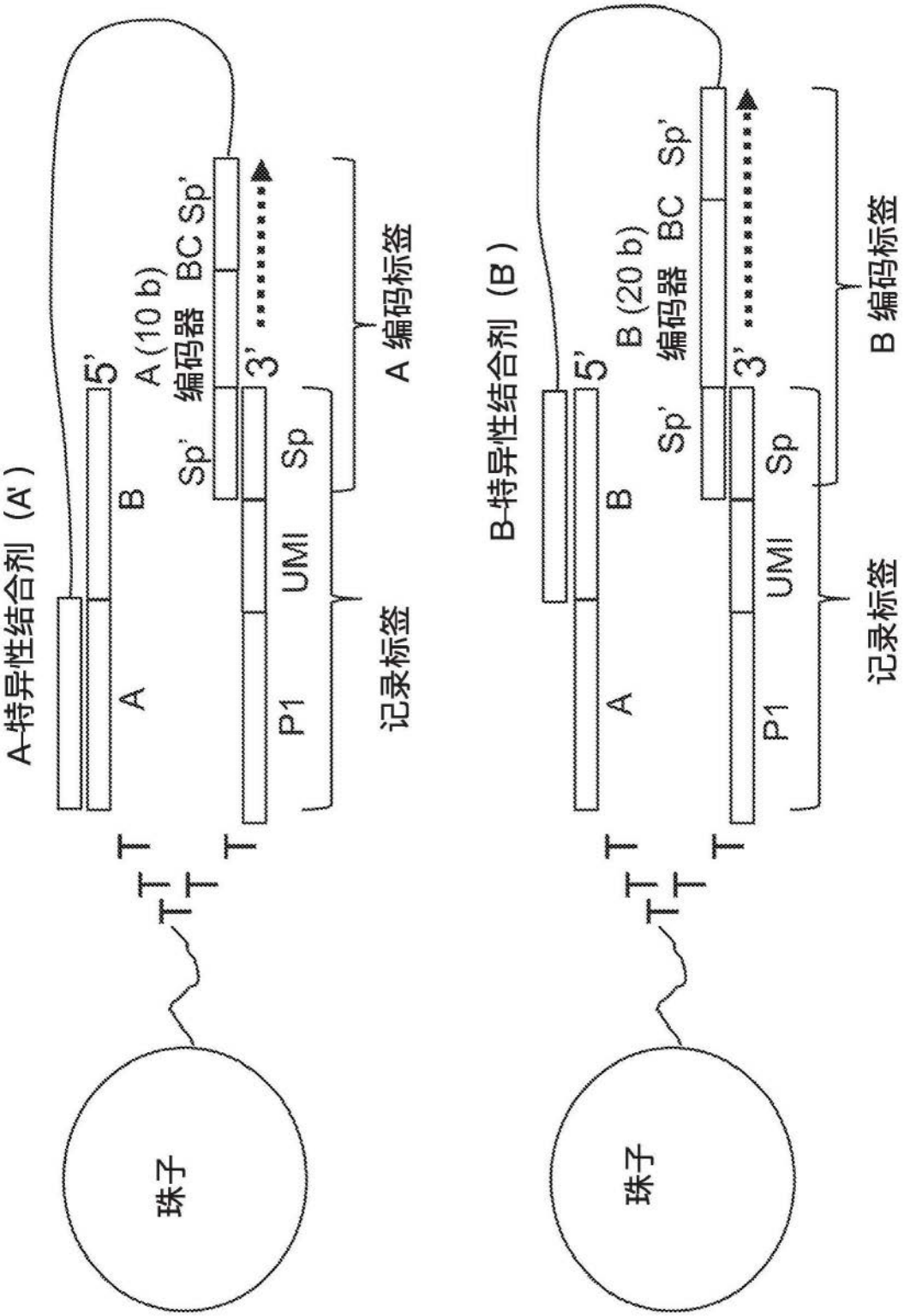

A-特异性结合剂（A'）
5'
A
B
A（10 b）
编码器
BC
Sp'
3'
Sp'
P1
UMI
Sp
A 编码标签
记录标签
T
珠子
B-特异性结合剂（B）
B（20 b）
B 编码标签

图24A

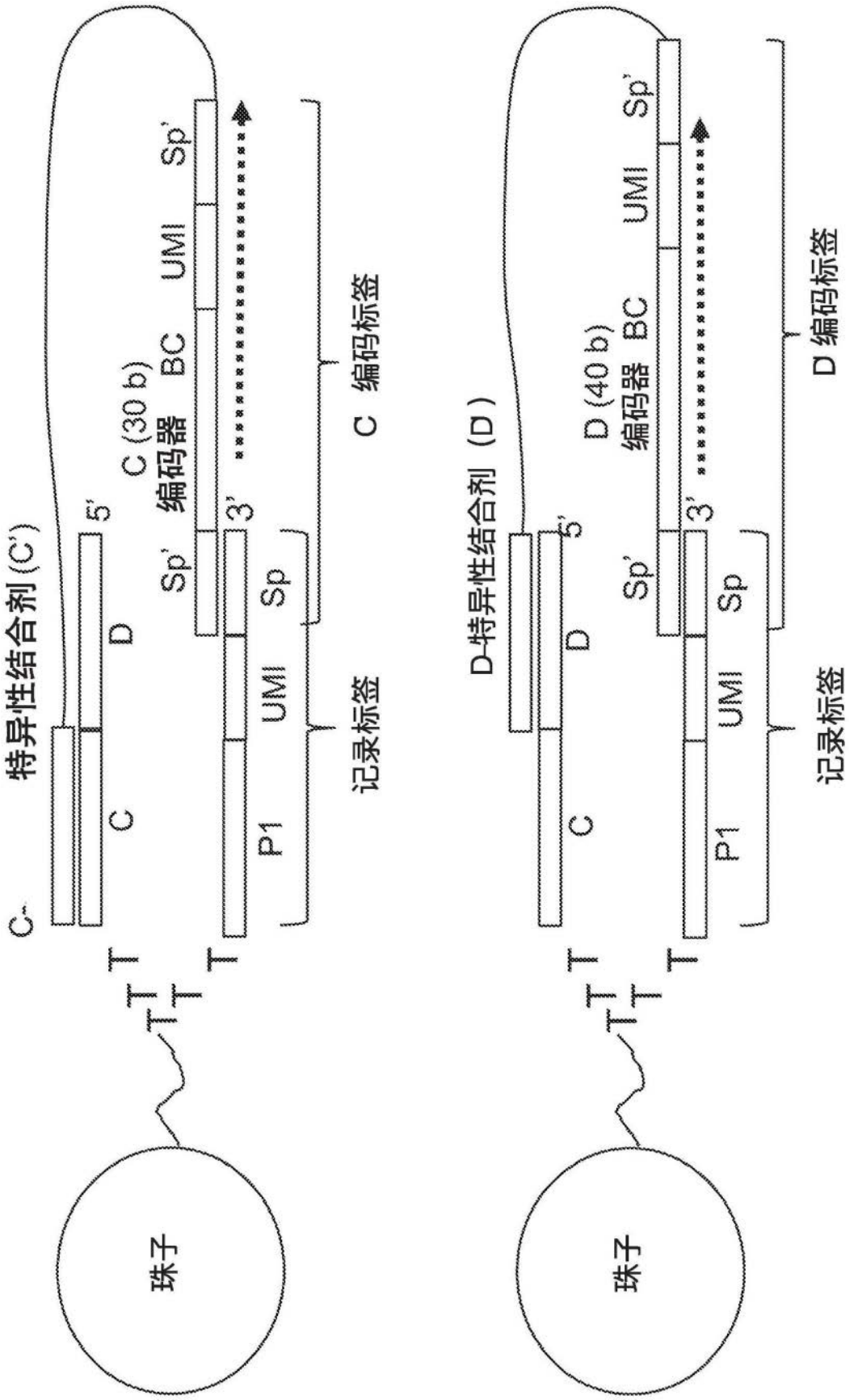

C- 特异性结合剂 (C')
5'
C
D
C (30 b)
Sp' 编码器 BC UMI Sp'
3'
P1 UMI Sp
C 编码标签
记录标签
T
T
T
T
T
珠子
D-特异性结合剂 (D)
5'
C
D
D (40 b)
Sp' 编码器 BC UMI Sp'
3'
P1 UMI Sp
D 编码标签
记录标签
T
T
T
T
T
珠子

图24B

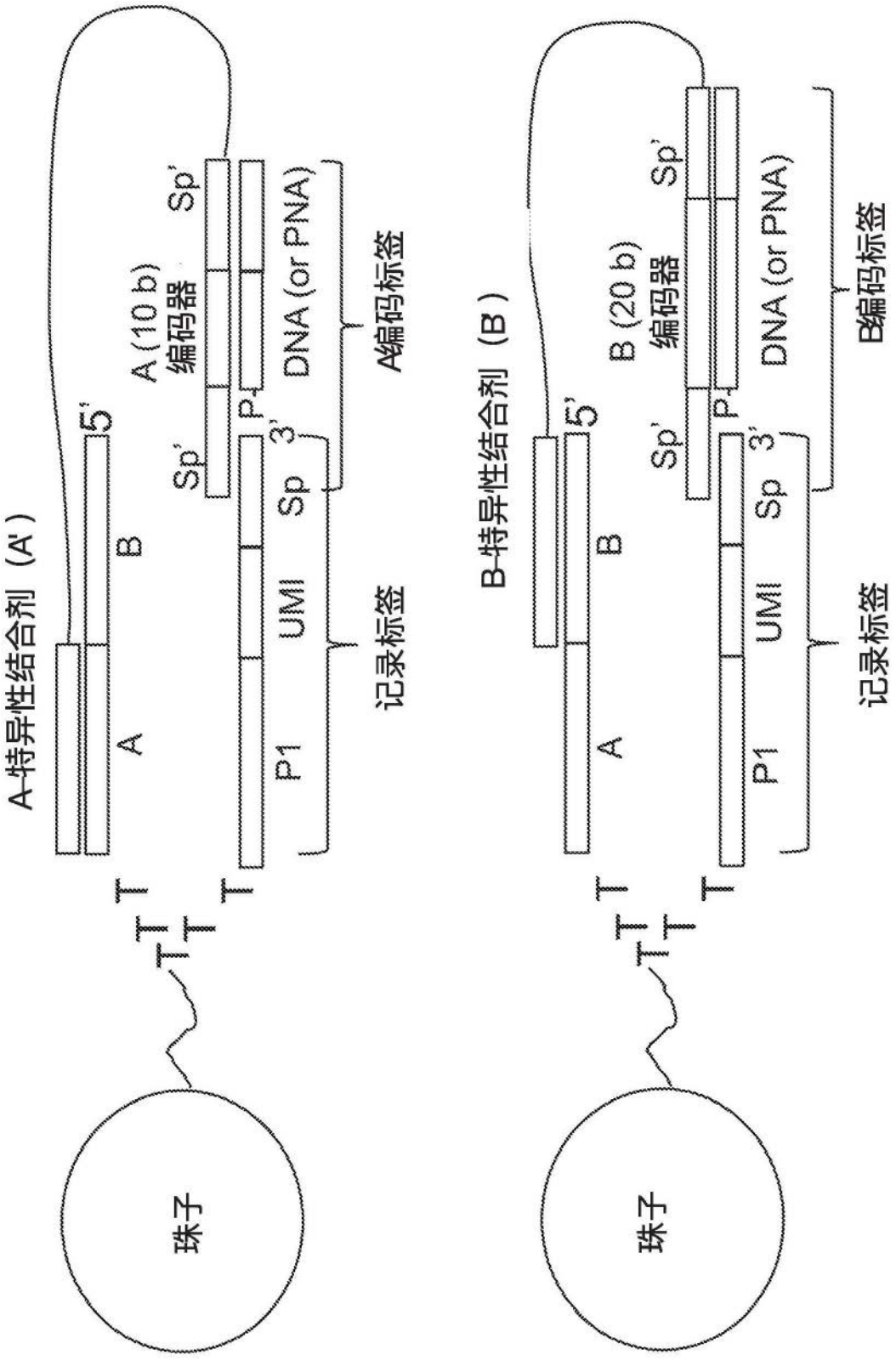

A-特异性结合剂（A'）
5'
A
B
A（10 b）
编码器
Sp'
Sp'
P1
DNA（or PNA）
3'
Sp
UMI
P1
A编码标签
记录标签
T
T
T
T
T
珠子
B-特异性结合剂（B）
5'
A
B
B（20 b）
编码器
Sp'
Sp'
P1
DNA（or PNA）
3'
Sp
UMI
P1
B编码标签
记录标签
T
T
T
T
T
珠子

图25

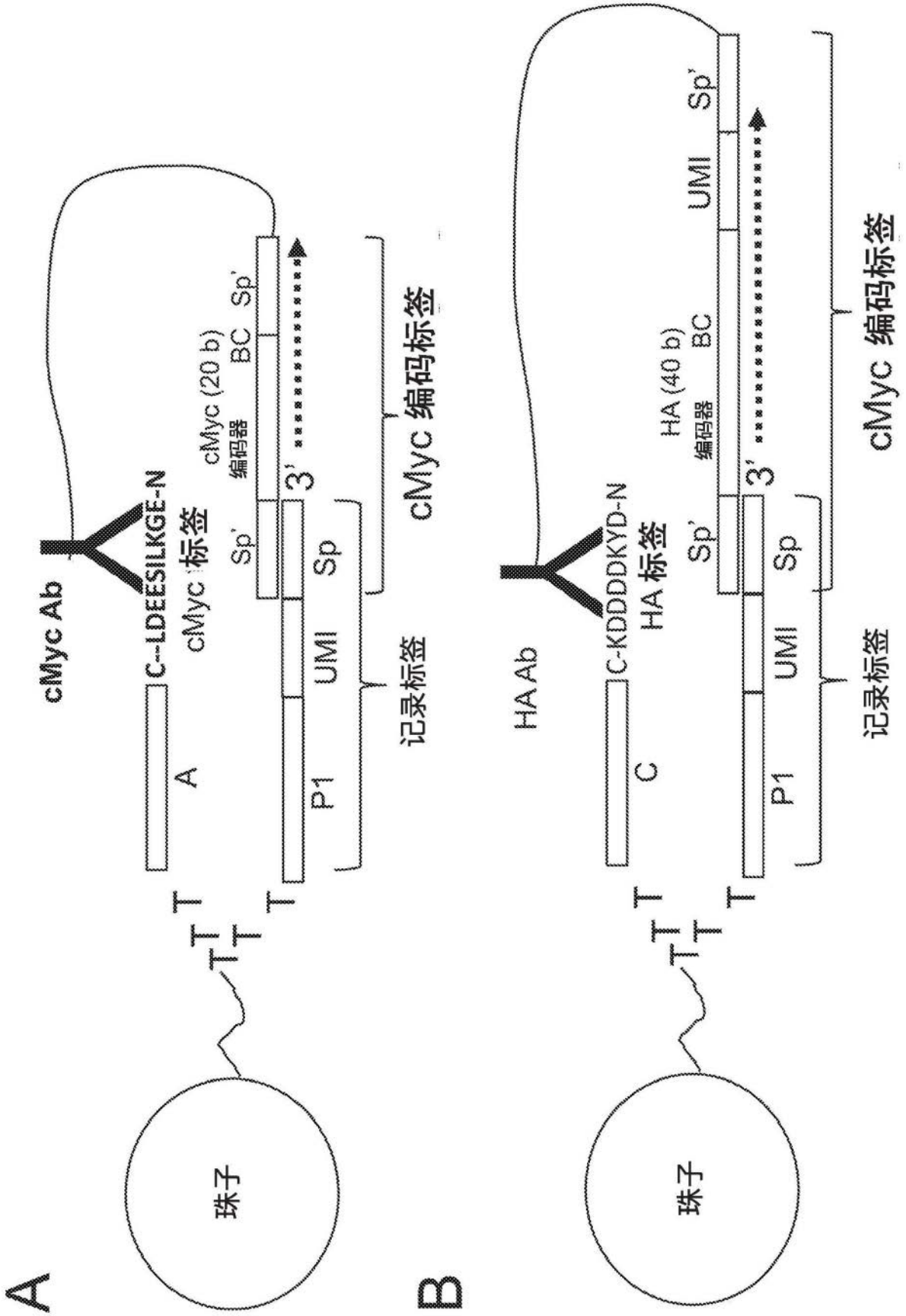

A
cMyc Ab
C--LDEESILKGE-N
cMyc 标签
A
cMyc (20 b)
编码器
BC
Sp'
3'
Sp'
P1
UMI
Sp
T
T
T
T
T
记录标签
cMyc 编码标签
珠子
B
HA Ab
C-KDDDDKYD-N
HA 标签
C
HA (40 b)
编码器
BC
Sp'
UMI
Sp'
3'
Sp'
P1
UMI
Sp
T
T
T
T
T
记录标签
cMyc 编码标签
珠子

图26

A.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ATGTCTAGCATGCCG | CCGCTTGTGATCTGG | CTGGATGGTTGTCGA | GATTCAACGGAGAAG | GCCATCAGTAAGAGA | GTCCTCGTCTATCCT | CCTTACAGGTCTGCG | GGCCAACTAAGGTGC |
| CCGTGTCATGTGGAA | AGATAGCGTACCGGA | ACTGCACGGTTCCAA | ACGAACCTGGCACCA | GCAACGTGAATTGAG | TAGGATTCCGTTACC | GATCATTGGCCAATT | GCACCTATTCGACAA |
| TAAGCCGGTATATCA | TCCAGGCTCATCATC | CGAGAGATGGTCCTT | AGGACTTCAAGAAGA | CTAAGTAGAGCCACA | TCTGACCACCGGAAG | TTCAAGGCTGAGTTG | TGGACACGATCGGCT |
| TTCGATATGACGGAA | GAGTACTAGAGCCAA | TCTTGAGAGACAAGA | GGTTGAATCCTCTGCA | 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 | CCACCGAGGCTTAGTT | CGTACGGTCAAGCAA |
| | | | | | | | TCGGTGACAGGCTAA |

图27A

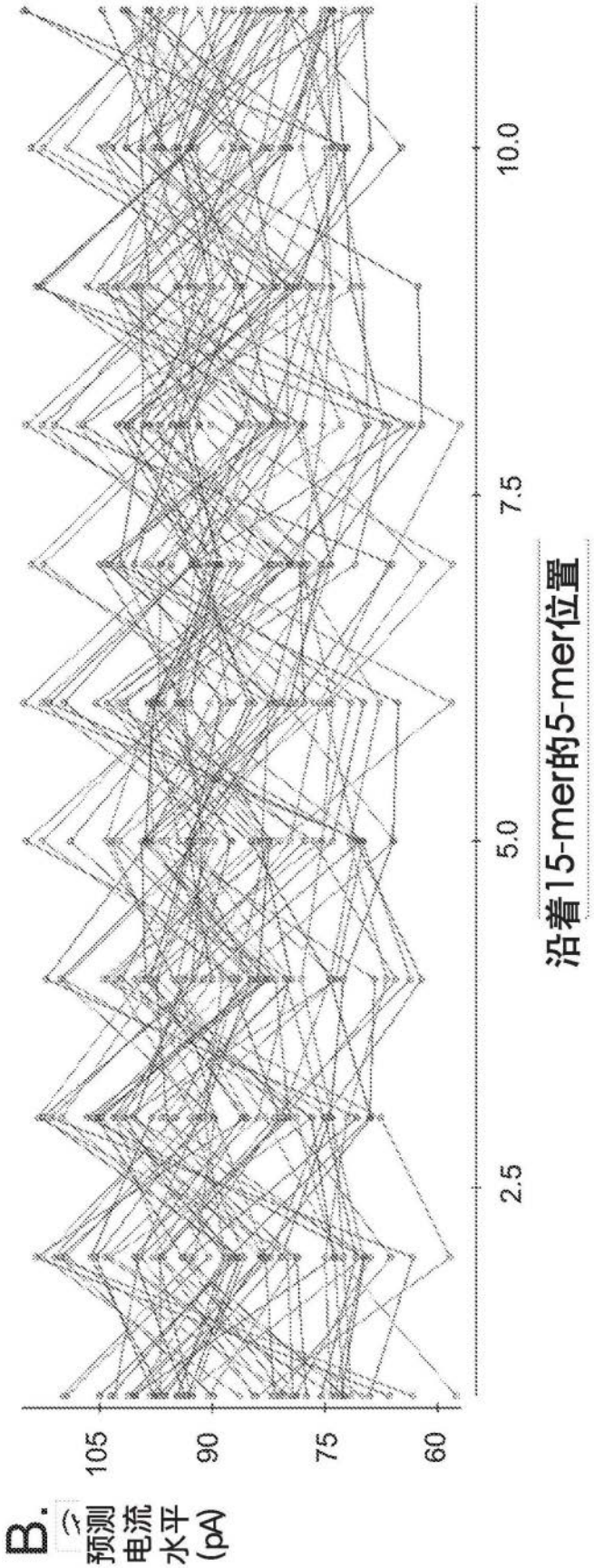



图27B

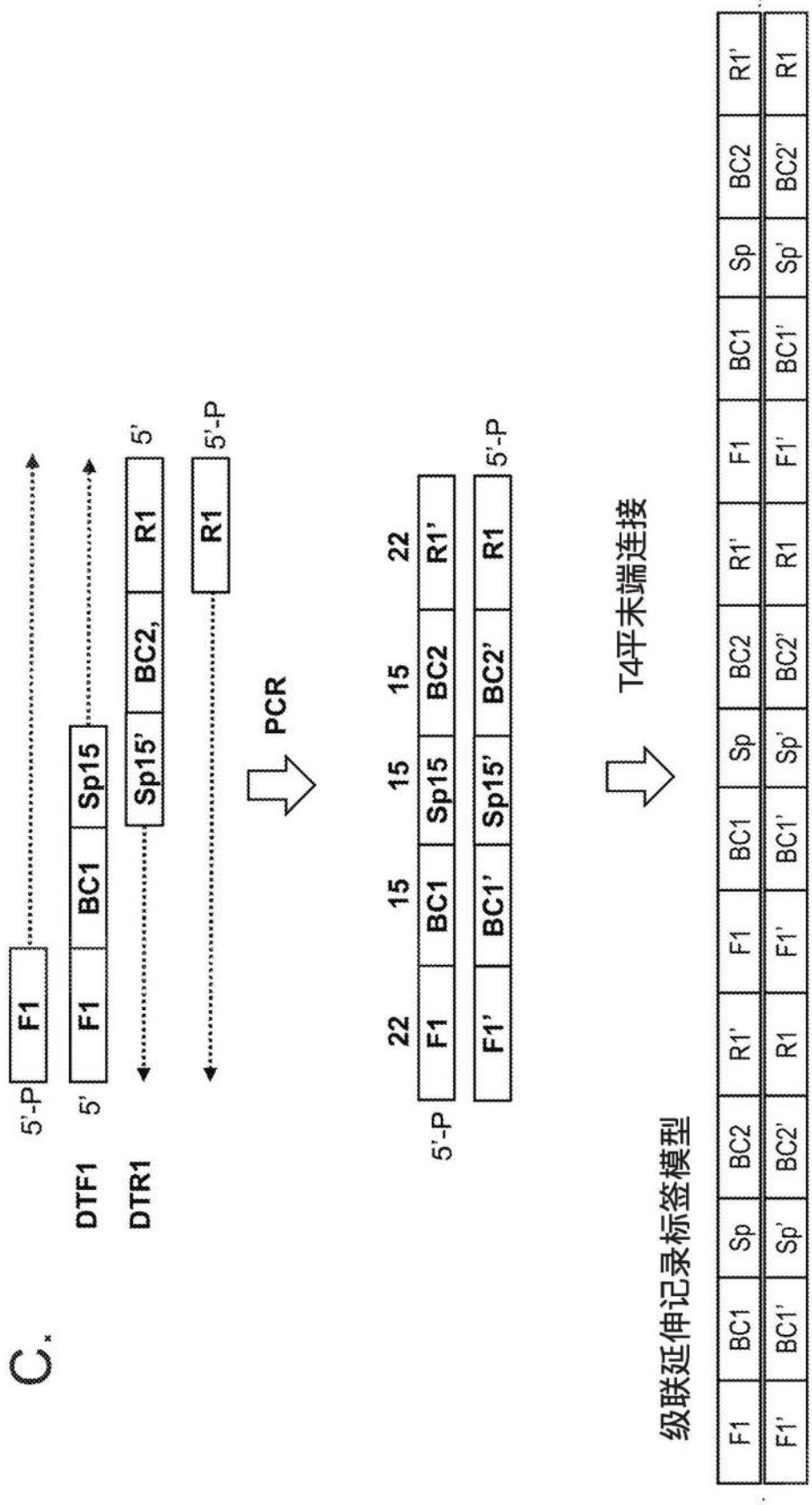

C×
5'-P
F1
DTF1
5'
F1
BC1
Sp15
DTR1
5'
Sp15'
BC2,
R1
5'-P
R1
PCR
22
15
15
15
22
5'-P
F1
BC1
Sp15
BC2
R1'
F1'
BC1'
Sp15'
BC2'
R1
5'-P
T4平末端连接
级联延伸记录标签模型
F1
BC1
Sp
BC2
R1'
F1
BC1
Sp
BC2
R1'
F1
BC1
Sp
BC2
R1'
F1'
BC1'
Sp'
BC2'
R1
F1'
BC1'
Sp'
BC2'
R1
F1'
BC1'
Sp'
BC2'
R1

图27C

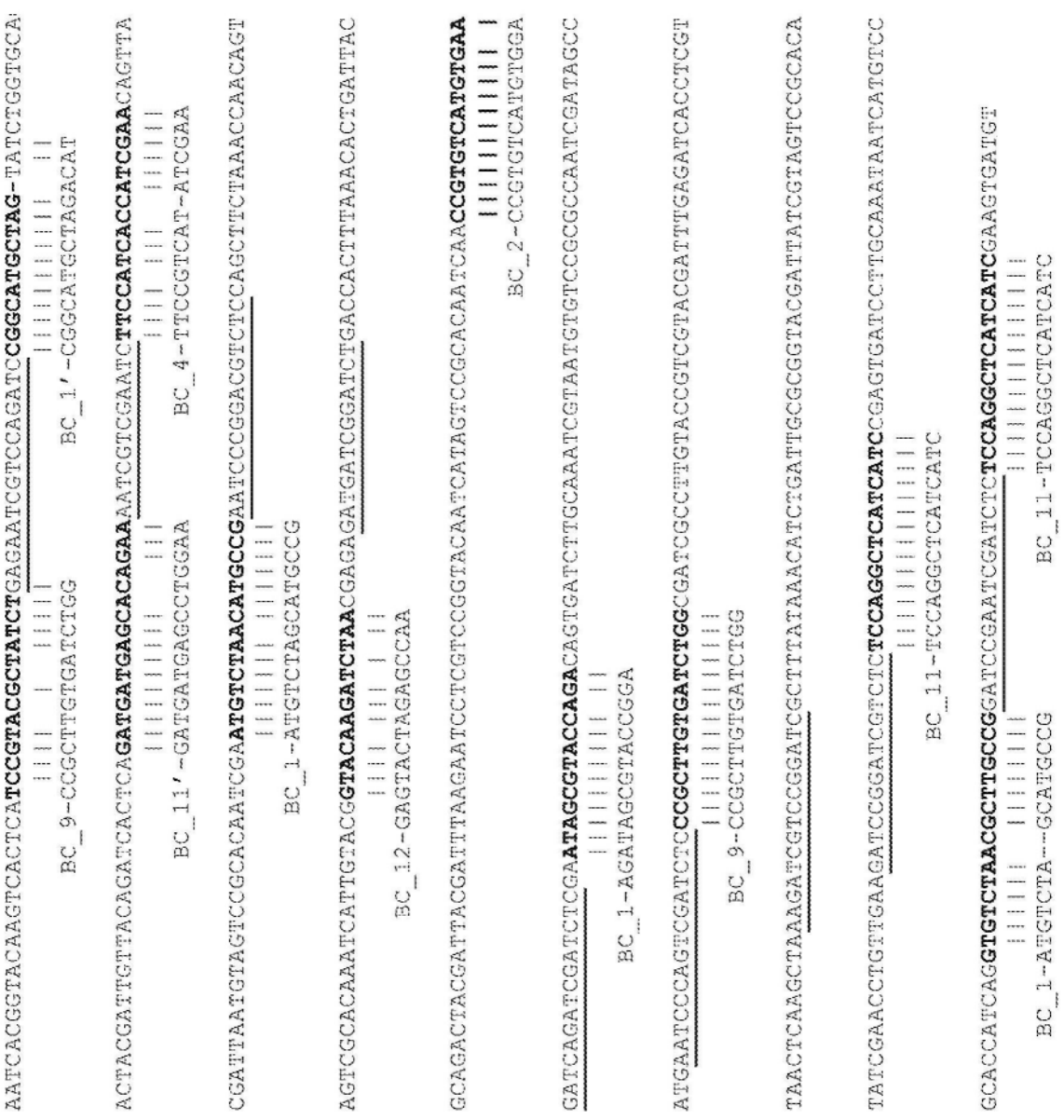

D.
AATCACGGTACAAGTCACTCATCCGTACGCTATCTGAGAATCGTCCAGATCCGGCATGCTAG-TATCTGGTGCA
BC_9-CCGCTTGTGATCTGG
BC_1'-CGGCATGCTAGACAT
ACTACGATTGTTACAGATCACTCAGATGATGAGCACAGAAAATCGTCGAATCTTCCATCACCATCGAACAGTTA
BC_11'-GATGATGAGCCTGGAA
BC_4-TTCCGTCAT-ATCGAA
CGATTAATGTAGTCCGCACAATCGAATGTCTAACATGCCGAATCCCGGACGTCTCCAGCTTCTAAACCAACAGT
BC_1-ATGTCTAGCATGCCG
AGTCGCACAAATCATTGTACGGTACAAGATCTAACGAGAGATGATCGGATCTGACCACTTTAAACACTGATTAC
BC_1.2-GAGTACTAGAGCCAA
GCAGACTACGATTACGATTTAAGAATCCTCGTCCGGTACAATCATAGTCCGCACAATCAACCGTGTCATGTGAA
BC_2-CCGTGTCATGTGGA
GATCAGATCGATCTCGAATAGCGTACCAGACAGTGATCTTGCAAATCGTAATGTGTCCGCGCCAATCGATAGCC
BC_1-AGATAGCGTACCGGA
ATGAATCCCAGTCGATCTCCCGCTTGTGATCTGGCGATCGCCTTGTACCGTCGTACGATTTGAGATCACCTCGT
BC_9-CCGCTTGTGATCTGG
TAACTCAAGCTAAAGATCGTCCGGATCGCTTTATAAACATCTGATTGCGCGGTACGATTATCGTAGTCCGCACA
TATCGAACCTGTTGAAGATCCGGATCGTCTCCTCCAGGCTCATCATCCGAGTGATCCTTGCAAATAATCATGTCC
BC_11-TCCAGGCTCATCATC
GCACCATCAGGTGTCTAACGCTTGCCGGATCCGAATCGGATCTCTCCAGGCTCATCATCGAAGTGATGT
BC_1-ATGTCTA---GCATGCCG
BC_11-TCCAGGCTCATCATC

图27D

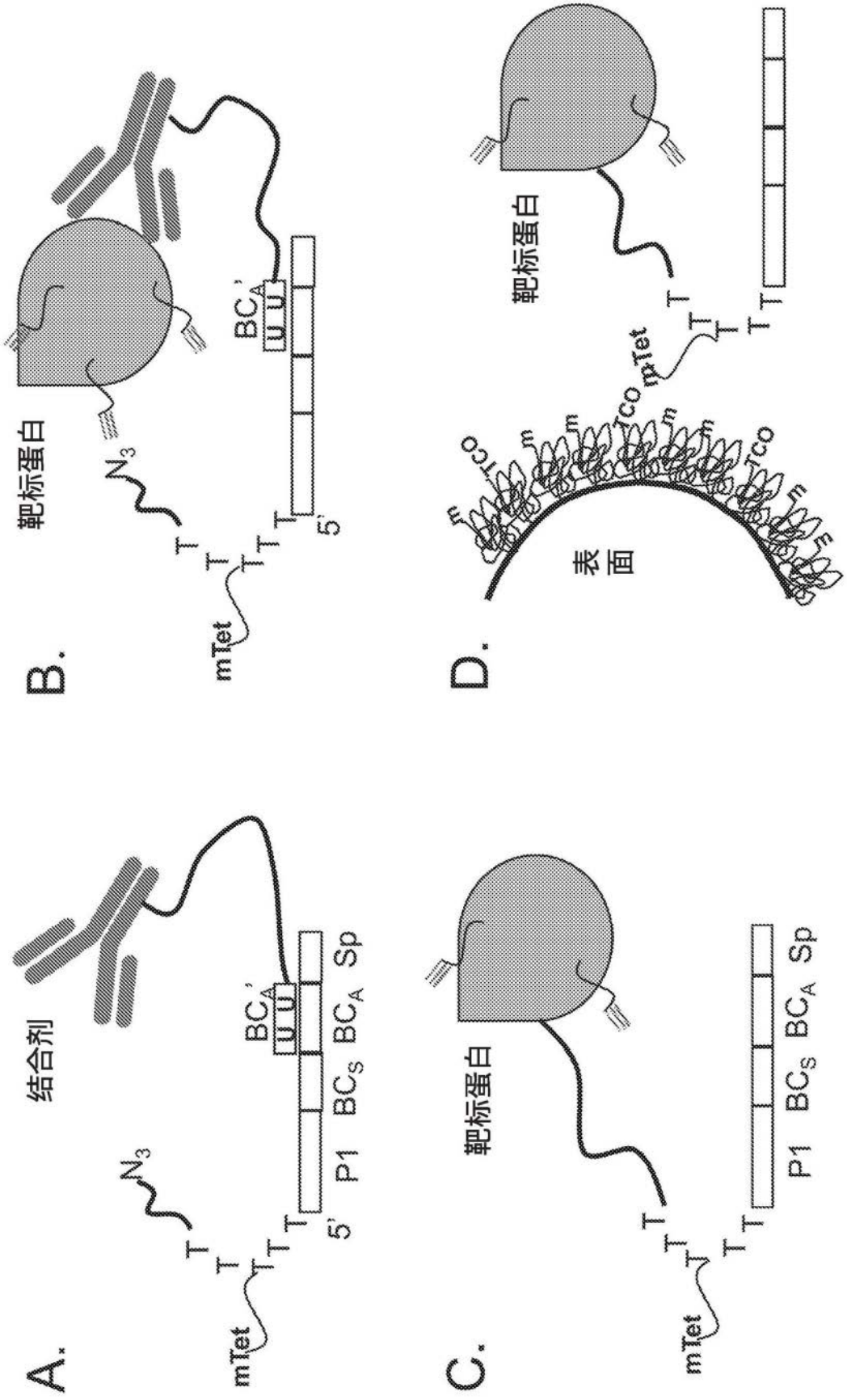

A.
结合剂
$N_3$
T
T
T
T
T
mTet
5'
P1
$BC_S$
$BC_A$
$BC_A'$
U
U
U
B.
靶标蛋白
$N_3$
T
T
T
T
T
mTet
$BC_A'$
U
U
U
5'
C.
靶标蛋白
T
T
T
T
T
mTet
P1
$BC_S$
$BC_A$
Sp
D.
靶标蛋白
T
T
T
T
T
mTet
TCO
m
表面

图28

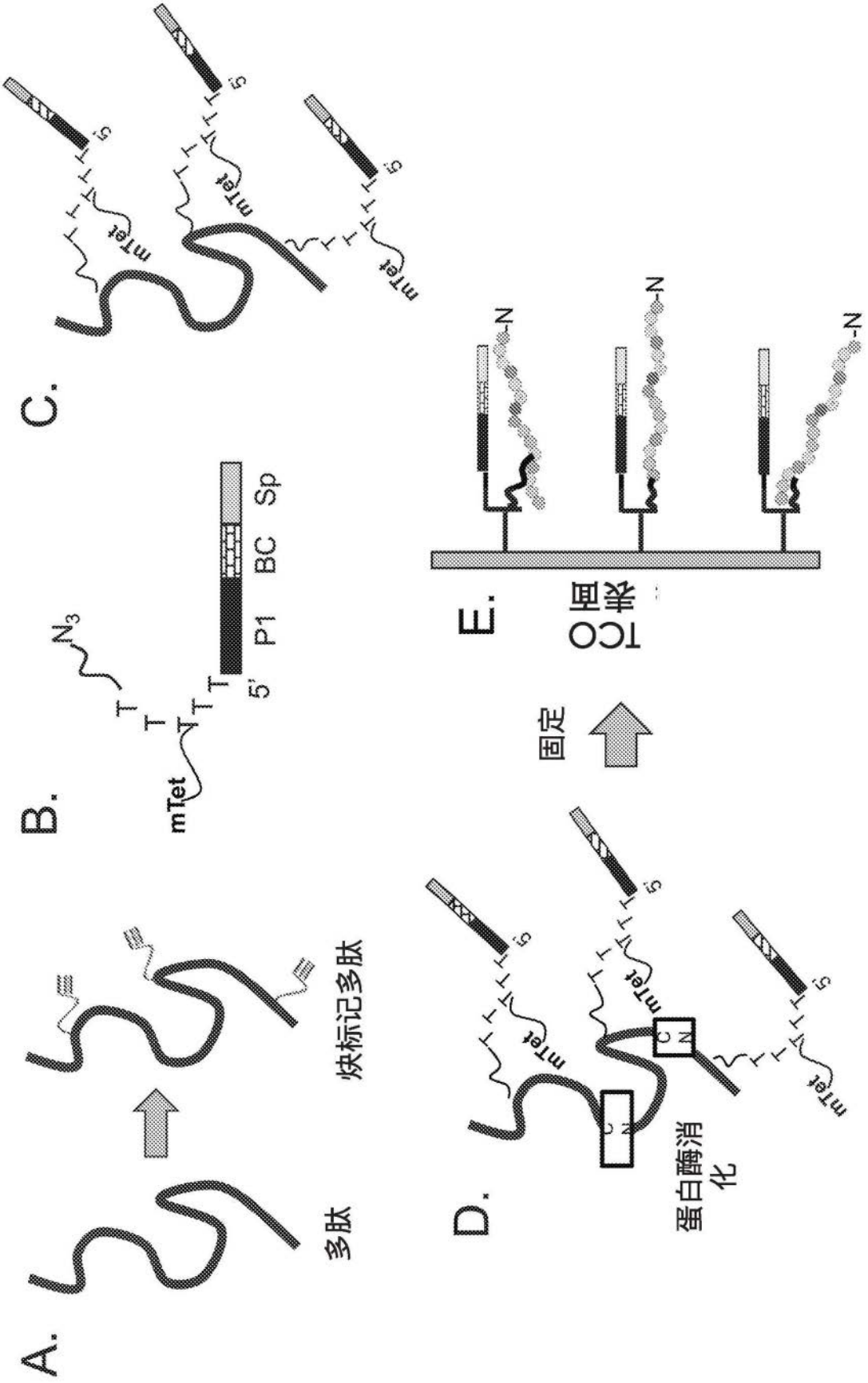



图29

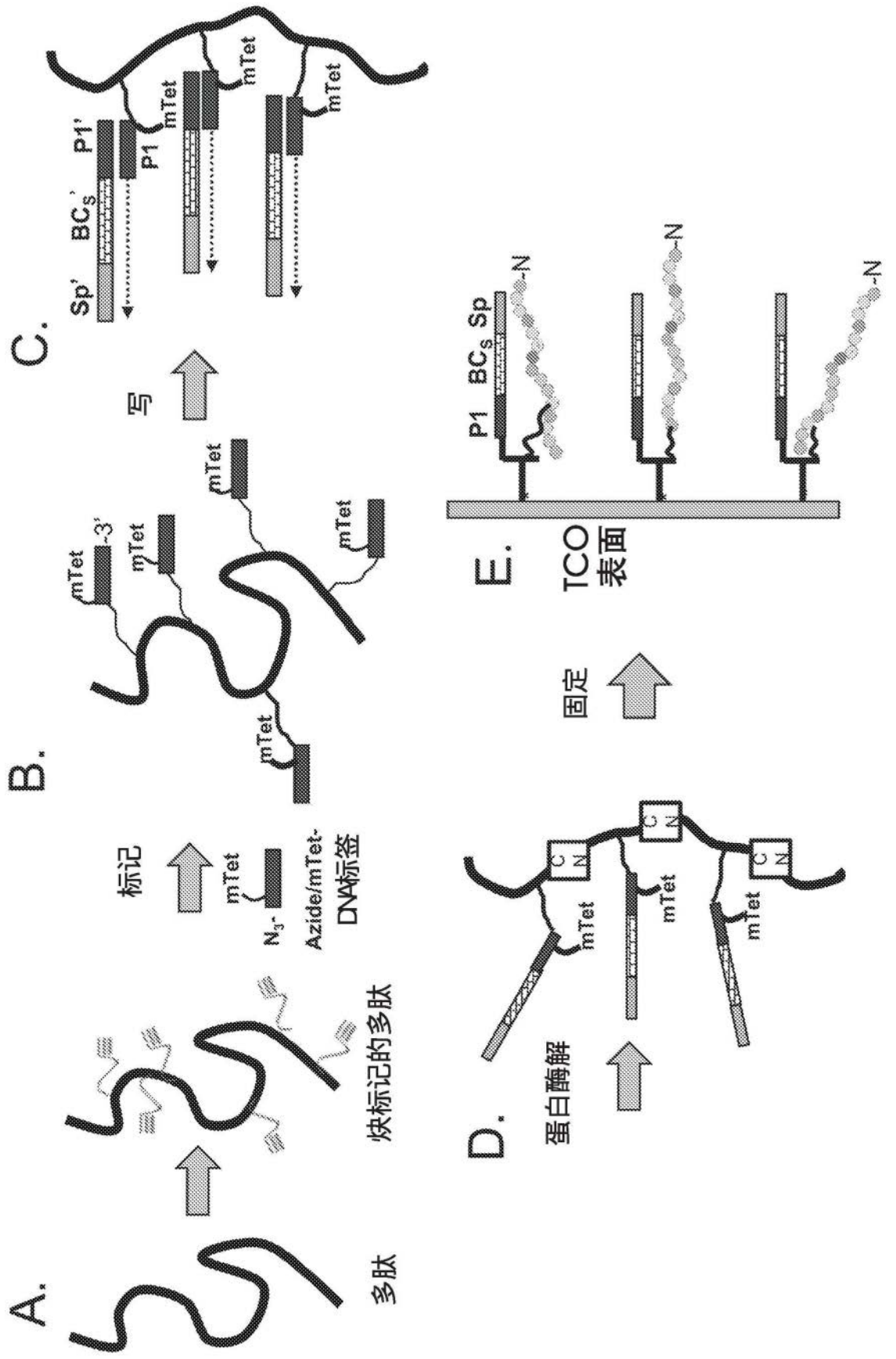

A.
多肽
炔标记的多肽
标记
mTet
N3-
Azide/mTet-DNA标签
B.
mTet
-3'
C.
写
Sp'
BCS'
P1'
P1
E.
TCO表面
固定
P1
BCS
Sp
-N
D.
蛋白酶解
C
N

图30

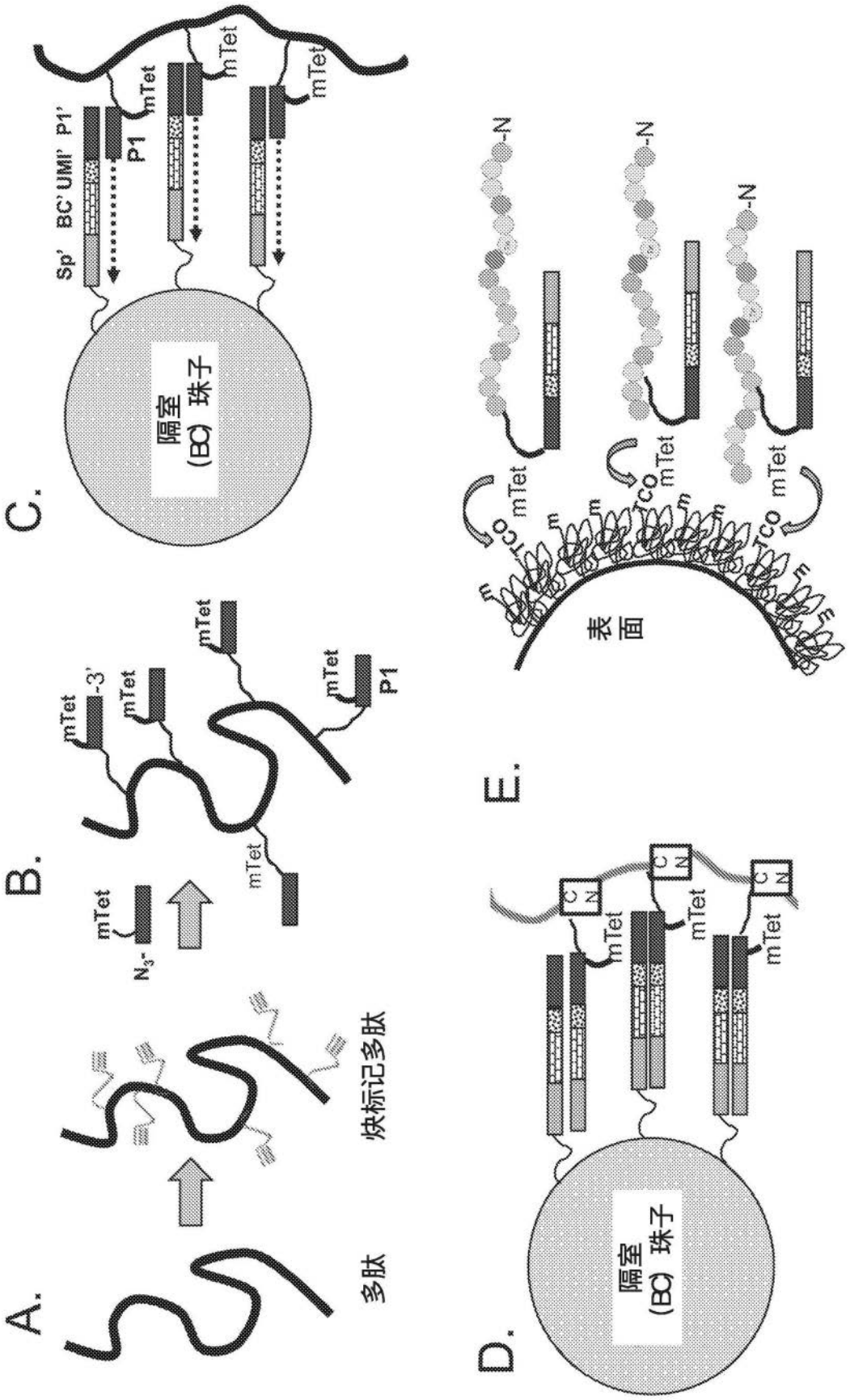

A.
多肽
炔标记多肽
B.
mTet
$N_3$-
mTet
-3'
mTet
mTet
mTet
P1
C.
隔室
(BC) 珠子
Sp'
BC'
UMI'
P1'
P1
mTet
mTet
mTet
D.
隔室
(BC) 珠子
C
N
mTet
E.
表面
-N
mTet
TCO

图31

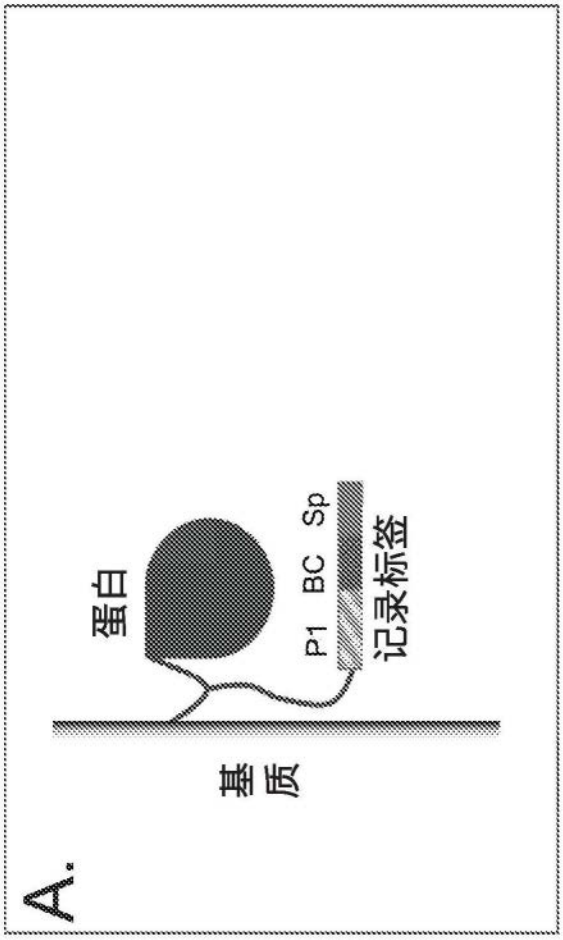



图32A

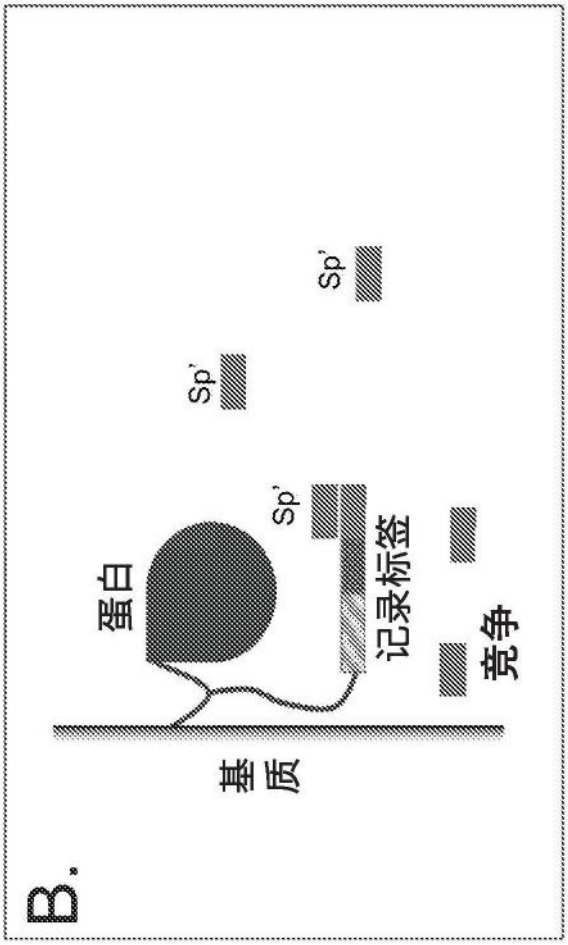



图32B

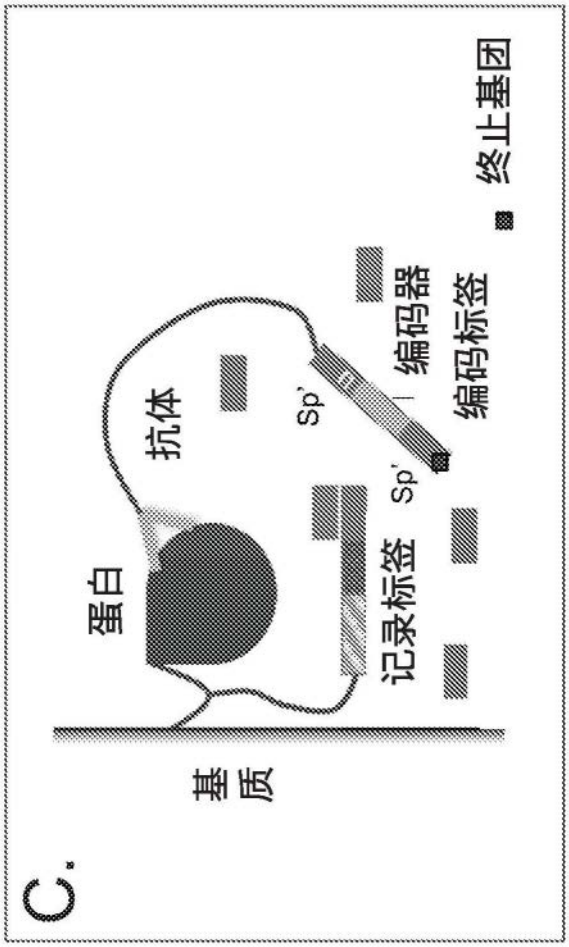



图32C

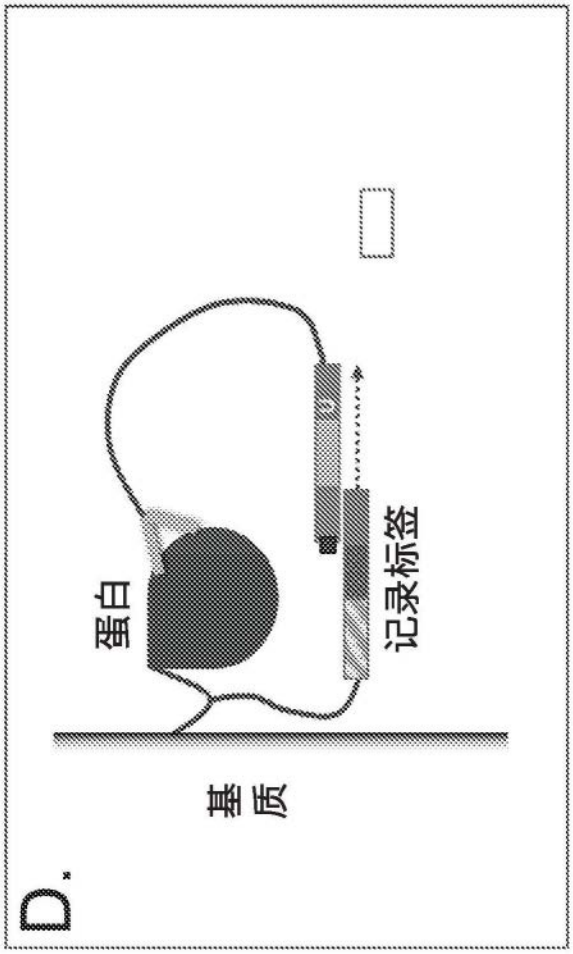



图32D

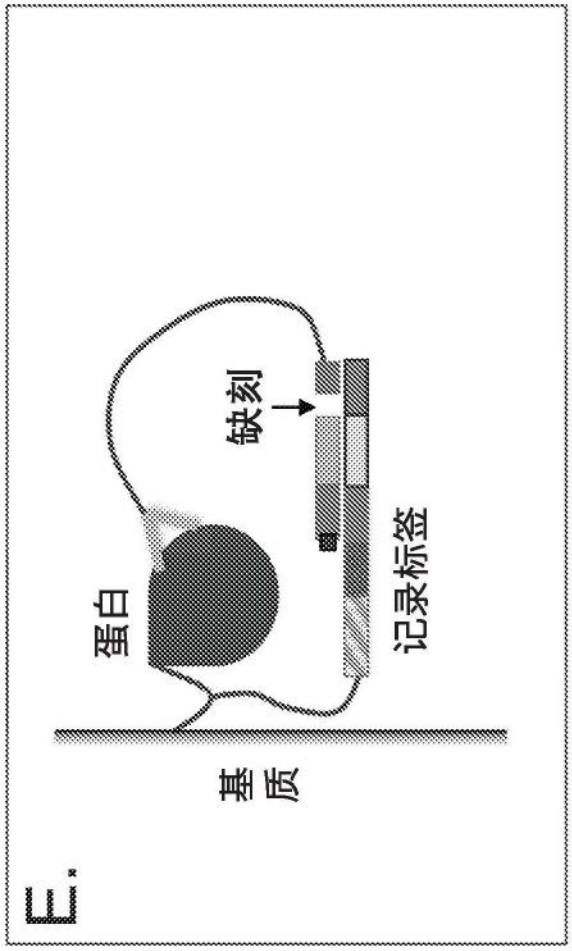



图32E

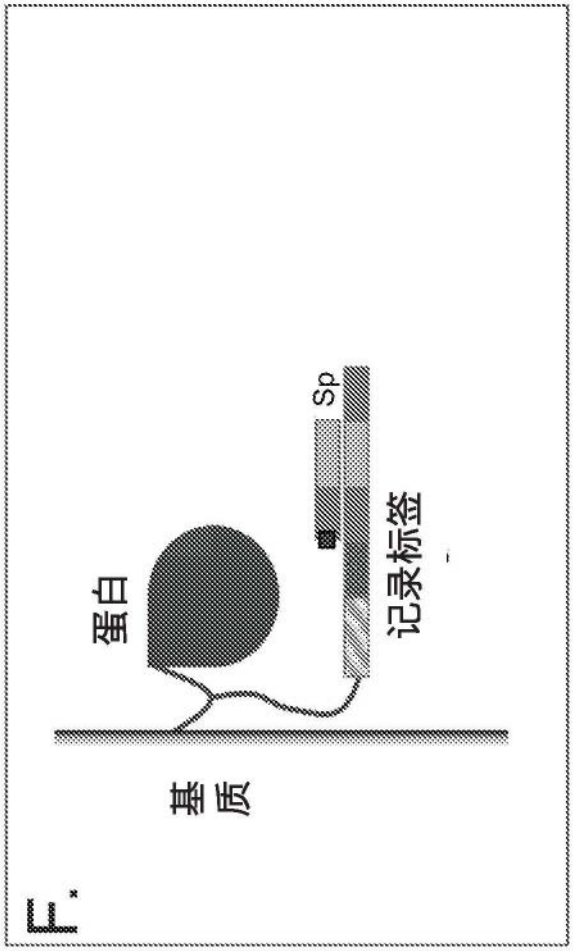



图32F

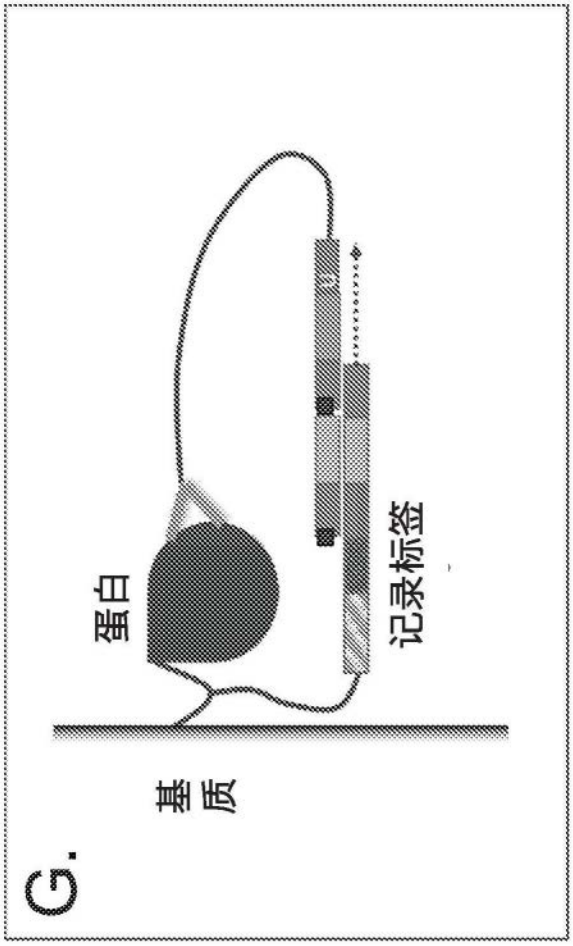



图32G

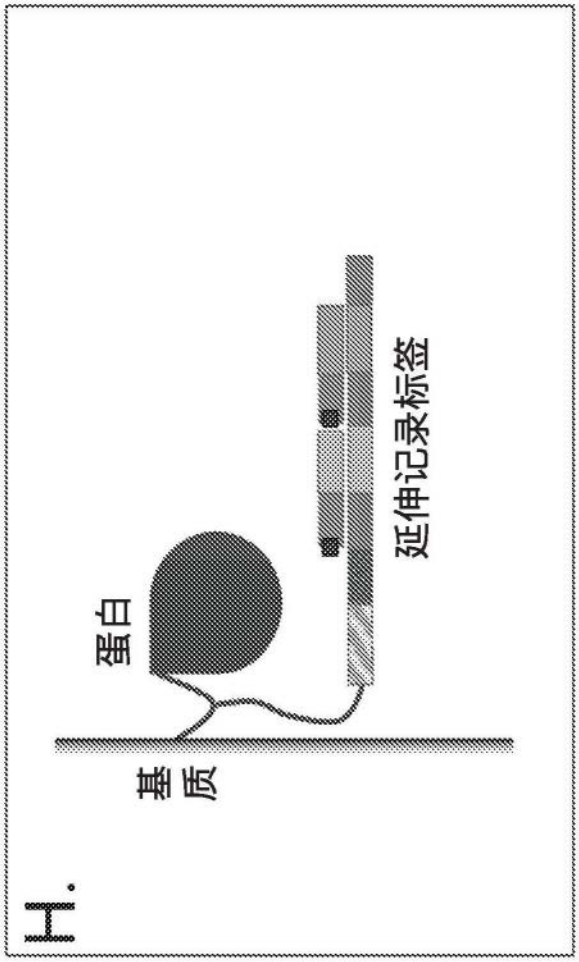



图32H

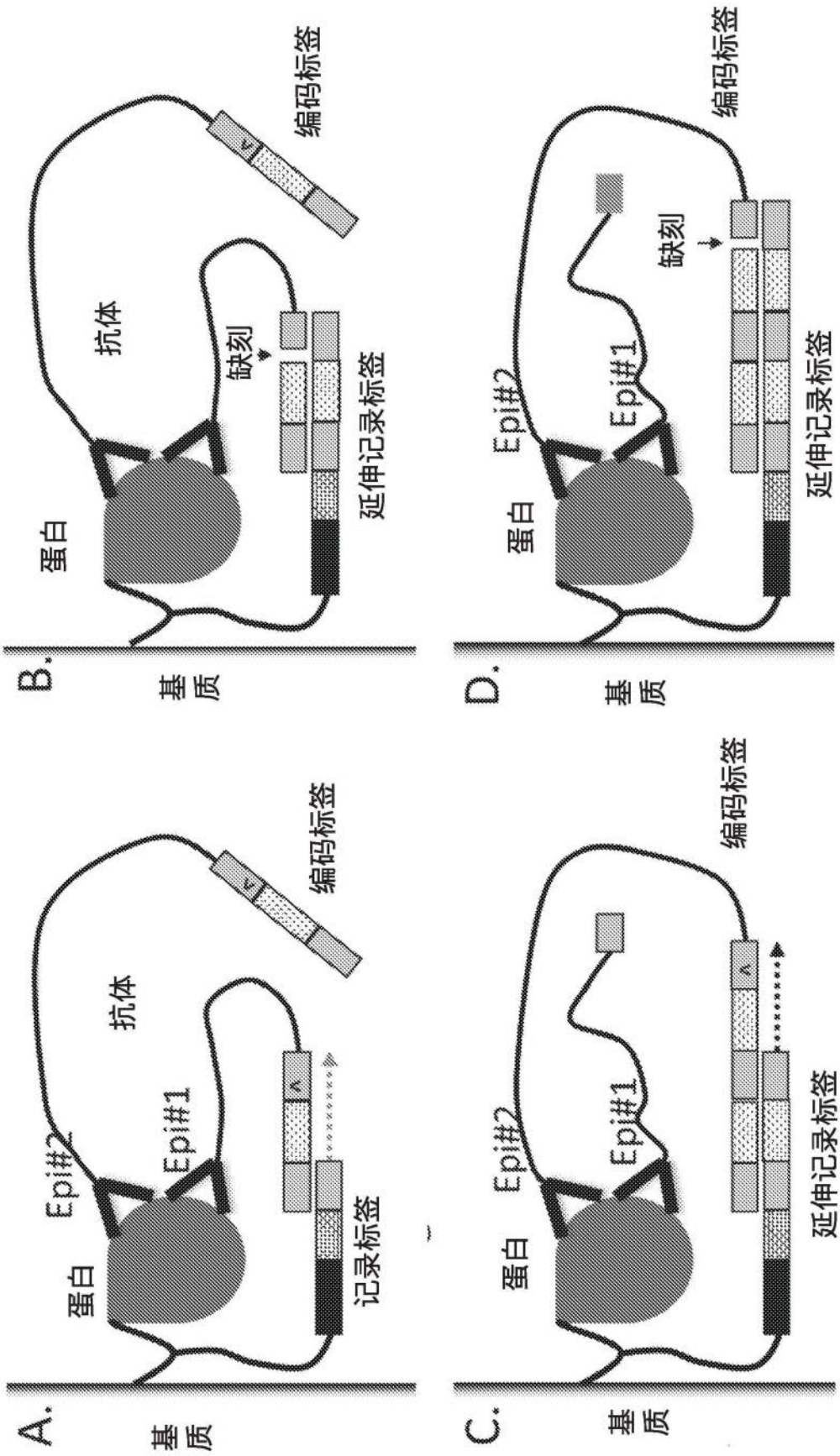

A.
基质
蛋白
Epi#2
抗体
Epi#1
编码标签
记录标签
B.
基质
蛋白
抗体
缺刻
编码标签
延伸记录标签
C.
基质
蛋白
Epi#2
Epi#1
编码标签
延伸记录标签
D.
基质
蛋白
Epi#2
Epi#1
缺刻
编码标签
延伸记录标签

图33

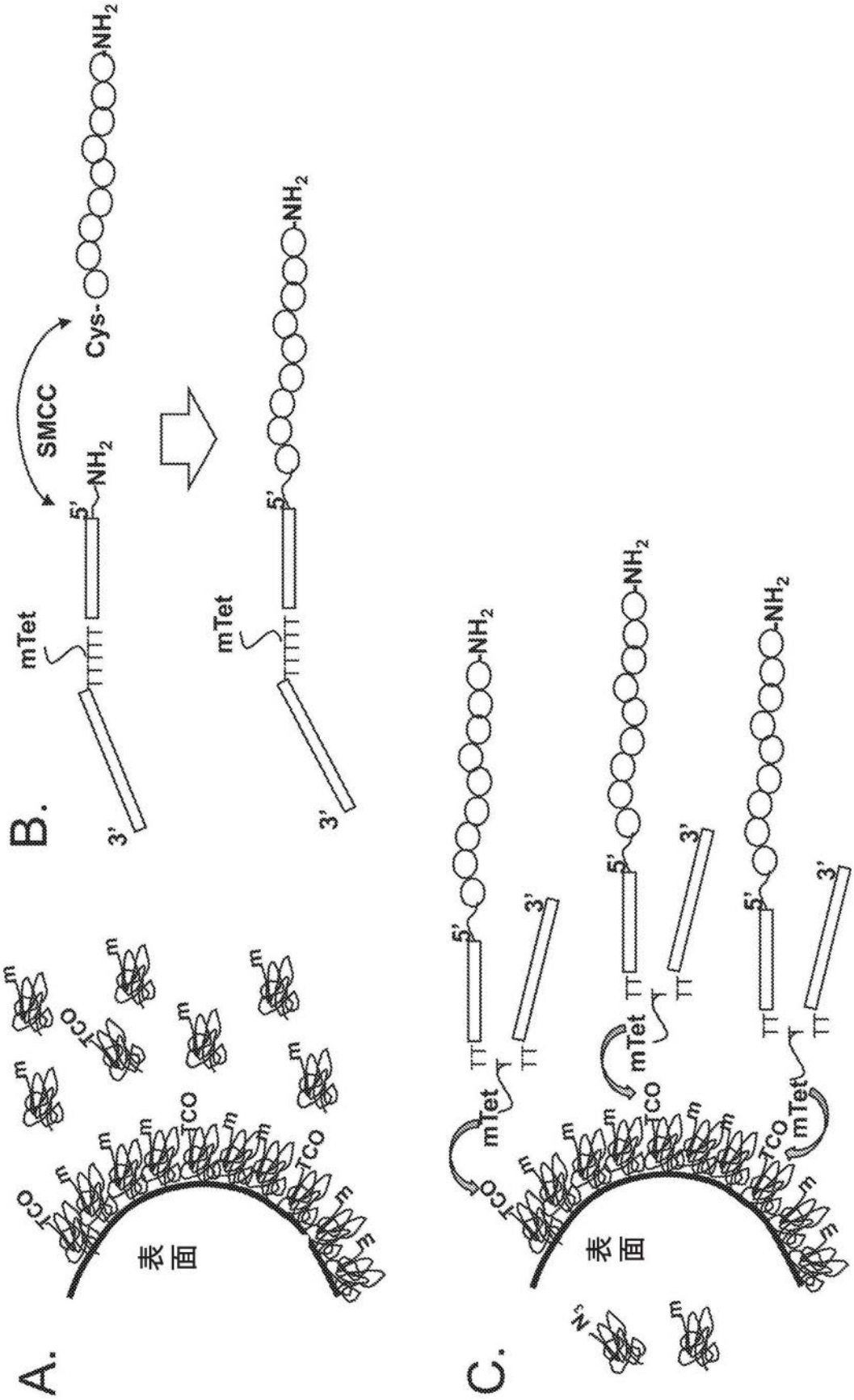



图34

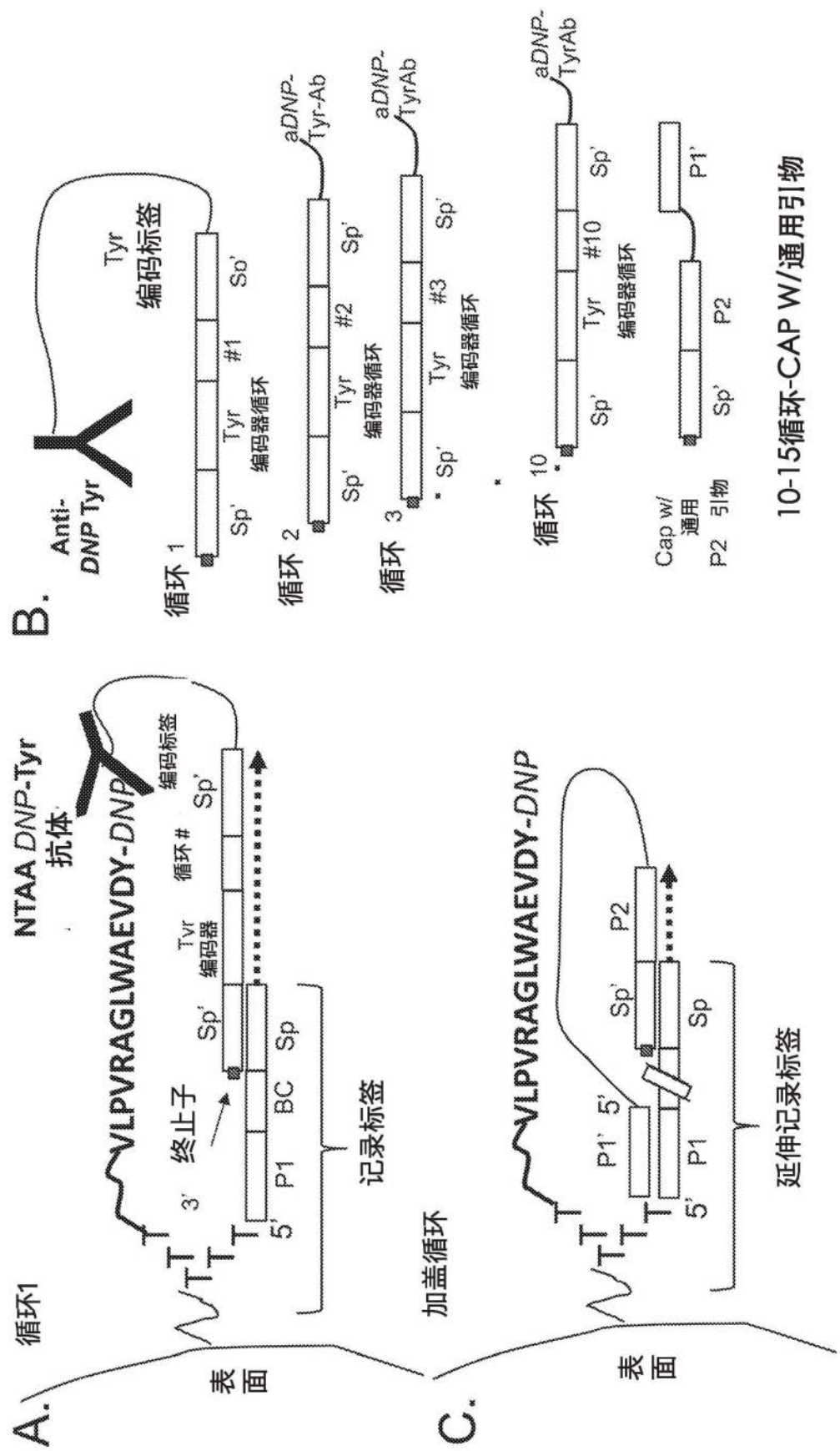

A.
循环1
NTAA DNP-Tyr
抗体
VLPVRAGLWAEVDY-DNP
终止子
3'
Sp'
Tyr
编码器
循环 #
编码标签
Sp'
5'
P1
BC
Sp
记录标签
表面
加盖循环
B.
Anti-
DNP Tyr
Tyr
编码标签
循环 1
Sp'
Tyr
#1
So'
编码器循环
循环 2
aDNP-
Tyr-Ab
Sp'
Tyr
#2
Sp'
编码器循环
循环 3
aDNP-
TyrAb
Sp'
Tyr
#3
Sp'
编码器循环
循环 10
aDNP-
TyrAb
Sp'
Tyr
#10
Sp'
编码器循环
Cap w/
通用
P2 引物
P1'
Sp'
P2
10-15循环-CAP W/通用引物
C.
VLPVRAGLWAEVDY-DNP
P1'
5'
Sp'
P2
5'
P1
Sp
延伸记录标签
表面

图35

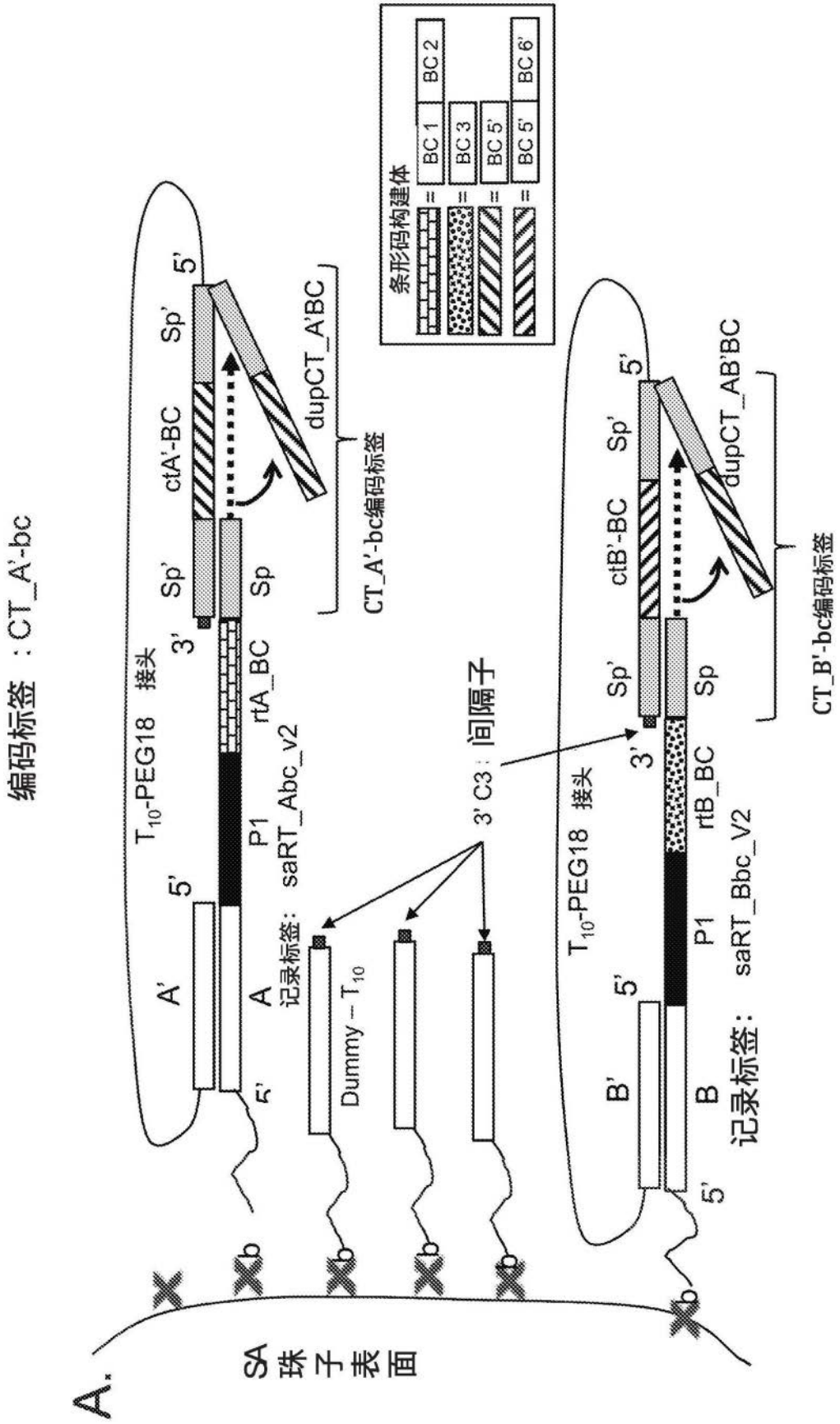

A.
SA 珠子表面
编码标签 ：CT_A'-bc
条形码构建体
BC 1
BC 2
BC 3
BC 5'
BC 5'
BC 6'
A'
5'
T10-PEG18 接头
3'
Sp'
ctA'-BC
Sp'
5'
dupCT_A'BC
CT_A'-bc编码标签
Sp
rtA_BC
P1
saRT_Abc_v2
A
记录标签：
Dummy – T10
5'
3' C3 间隔子
B'
5'
T10-PEG18 接头
3'
Sp'
ctB'-BC
Sp'
5'
dupCT_AB'BC
CT_B'-bc编码标签
Sp
rtB_BC
P1
saRT_Bbc_V2
B
记录标签：
5'

图36A

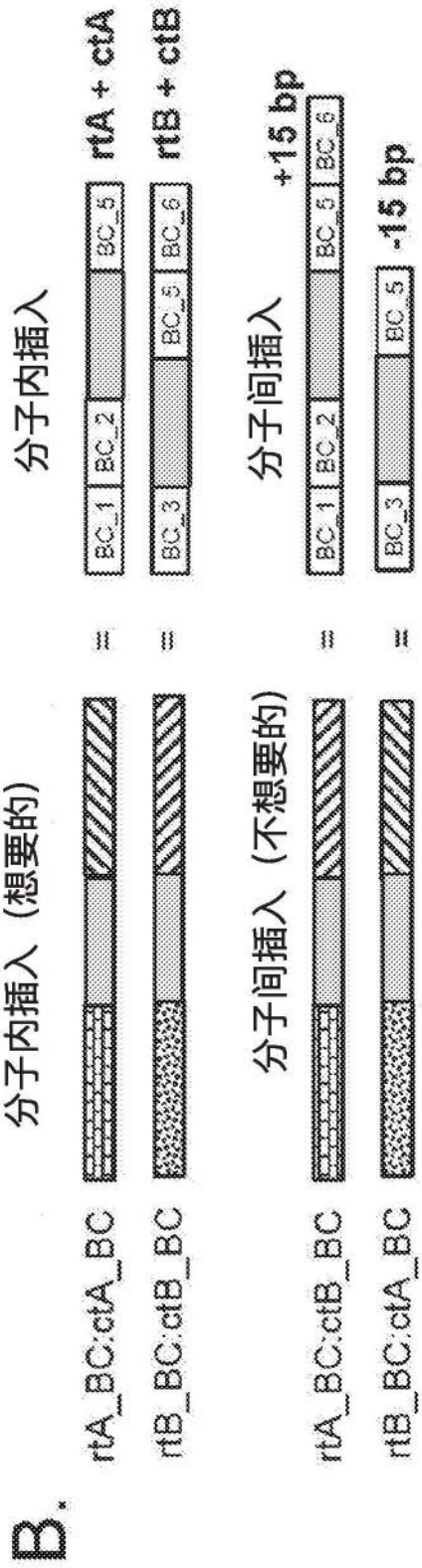

B.
分子内插入（想要的）
rtA_BC:ctA_BC
rtB_BC:ctB_BC
分子内插入
BC_1
BC_2
BC_5
rtA + ctA
BC_3
BC_5
BC_6
rtB + ctB
分子间插入（不想要的）
rtA_BC:ctB_BC
rtB_BC:ctA_BC
分子间插入
+15 bp
BC_1
BC_2
BC_5
BC_6
-15 bp
BC_3
BC_5

图36B

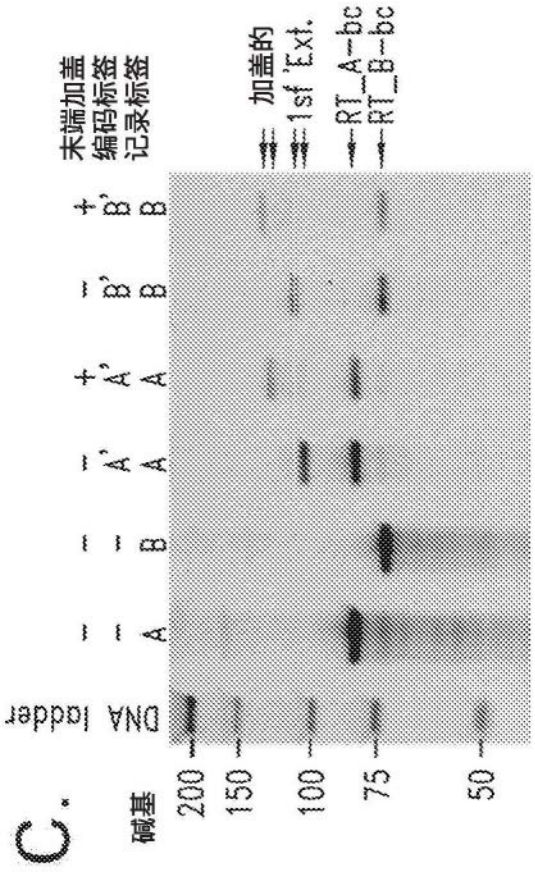

C.
碱基
200
150
100
75
50
DNA ladder
末端加盖
编码标签
记录标签
加盖的
1sf 'Ext.
RT_A-bc
RT_B-bc

图36C

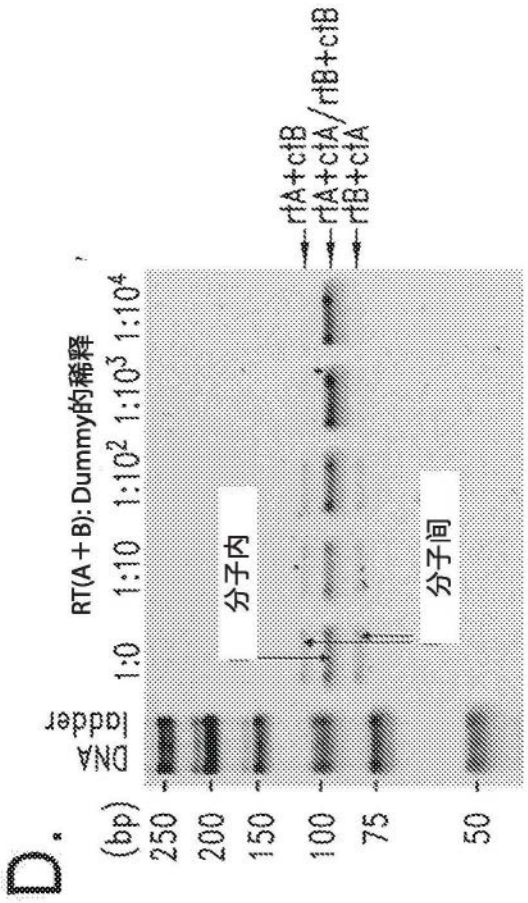

D.
(bp)
250
200
150
100
75
50
DNA ladder
RT(A + B): Dummy的稀释
1:0
1:10
1:10²
1:10³
1:10⁴
分子内
分子间
rtA+ctB
rtA+ctA/rtB+ctB
rtB+ctA

图36D

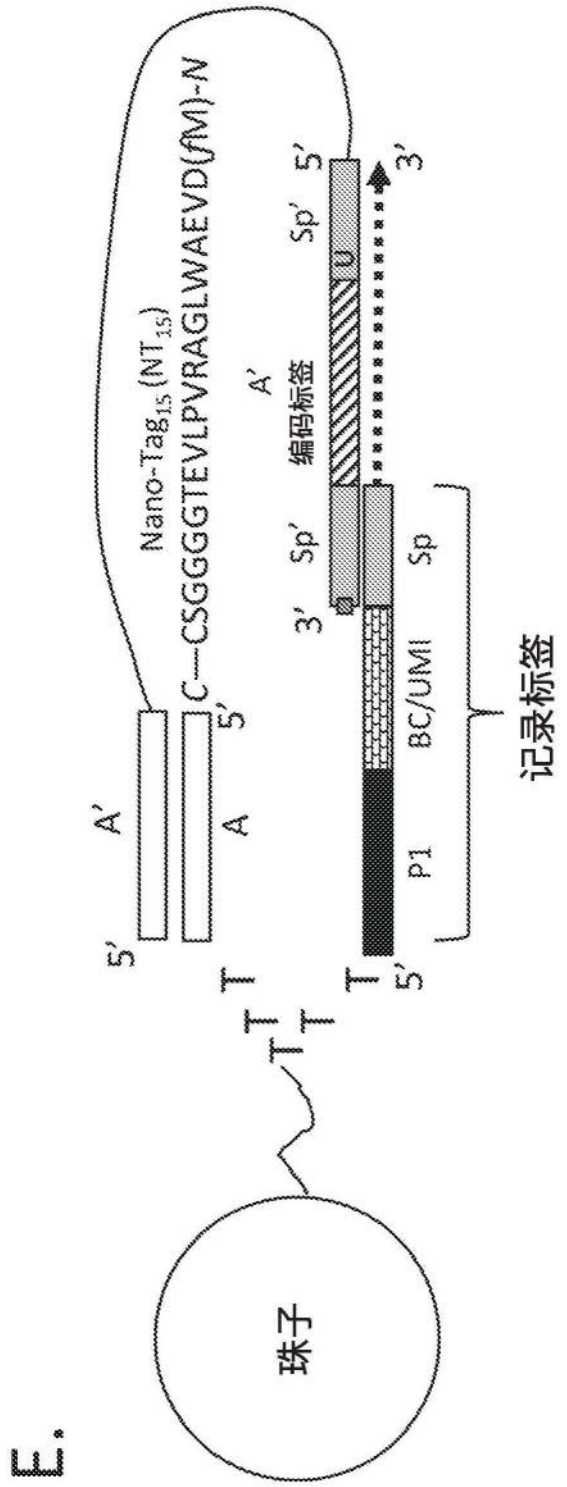
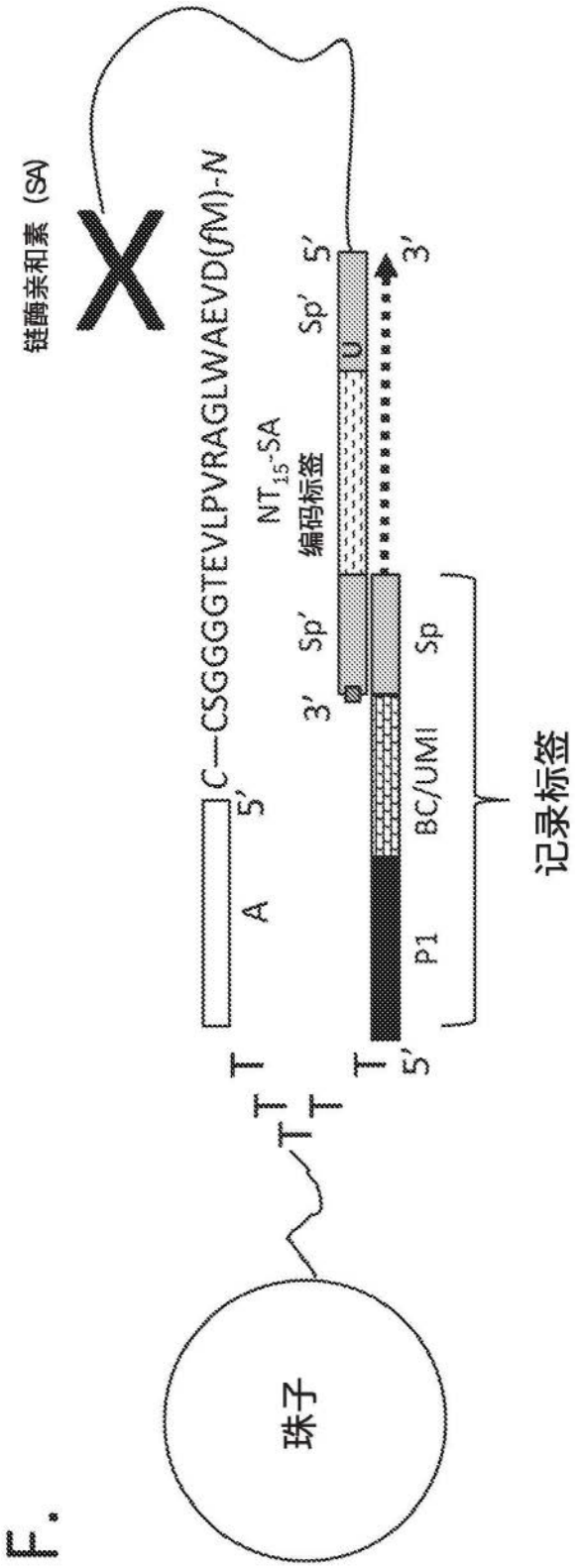

E.
珠子
T
T
T
T
T
5'
A'
5'
A
C—CSGGGGTEVLPVRAGLWAEVD(fM)-N
Nano-Tag15 (NT15)
A'
编码标签
3'
Sp'
U
5'
3'
Sp'
Sp
BC/UMI
P1
记录标签

图36E


链酶亲和素 (SA)
C—CSGGGGTEVLPVRAGLWAEVD(fM)-N
5'
A
T
T
T
T
T
T
5'
P1
BC/UMI
Sp
3'
Sp'
Ø
NT15-SA
编码标签
Sp'
U
5'
3'
记录标签
珠子

F.

图36F

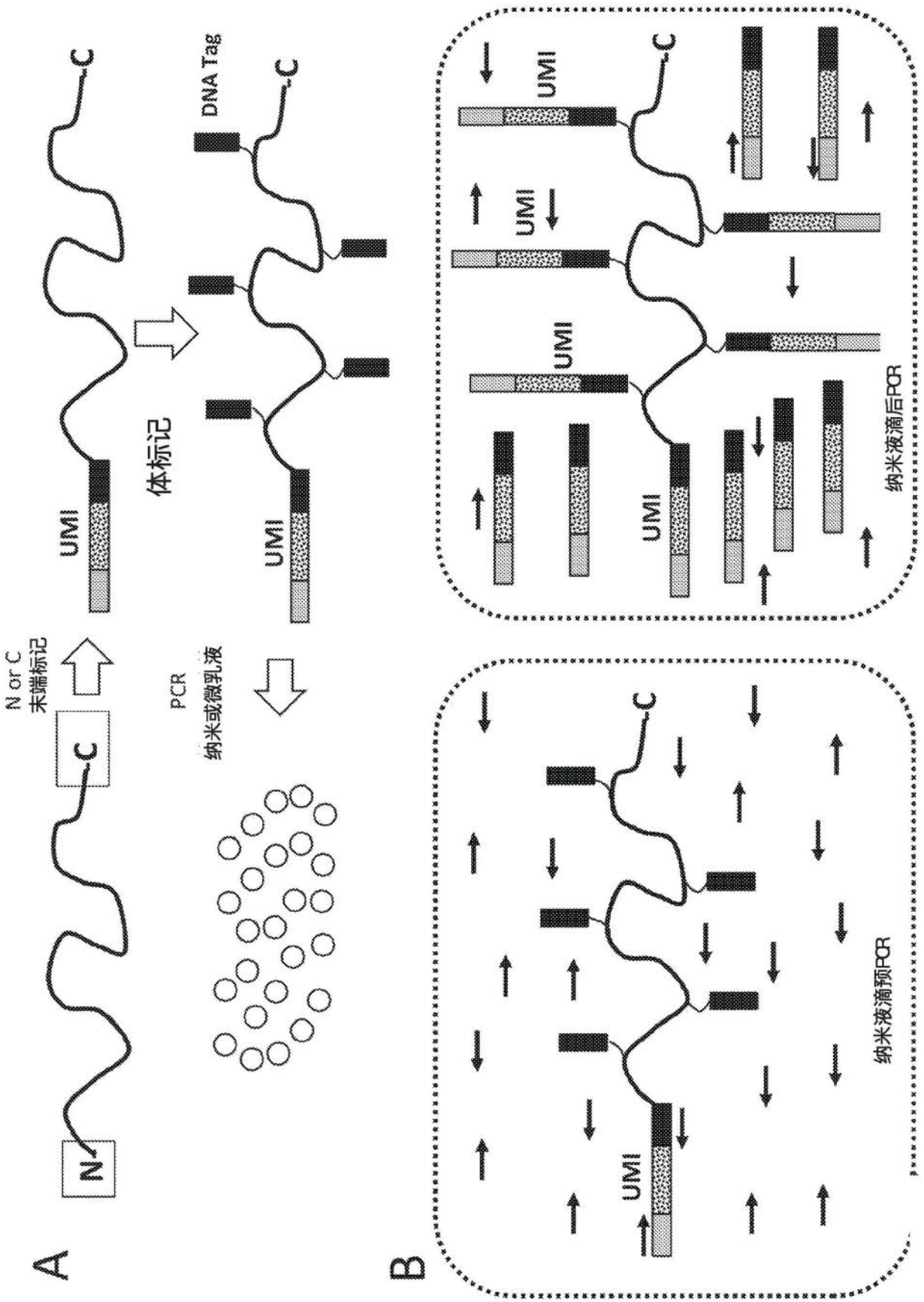



图37

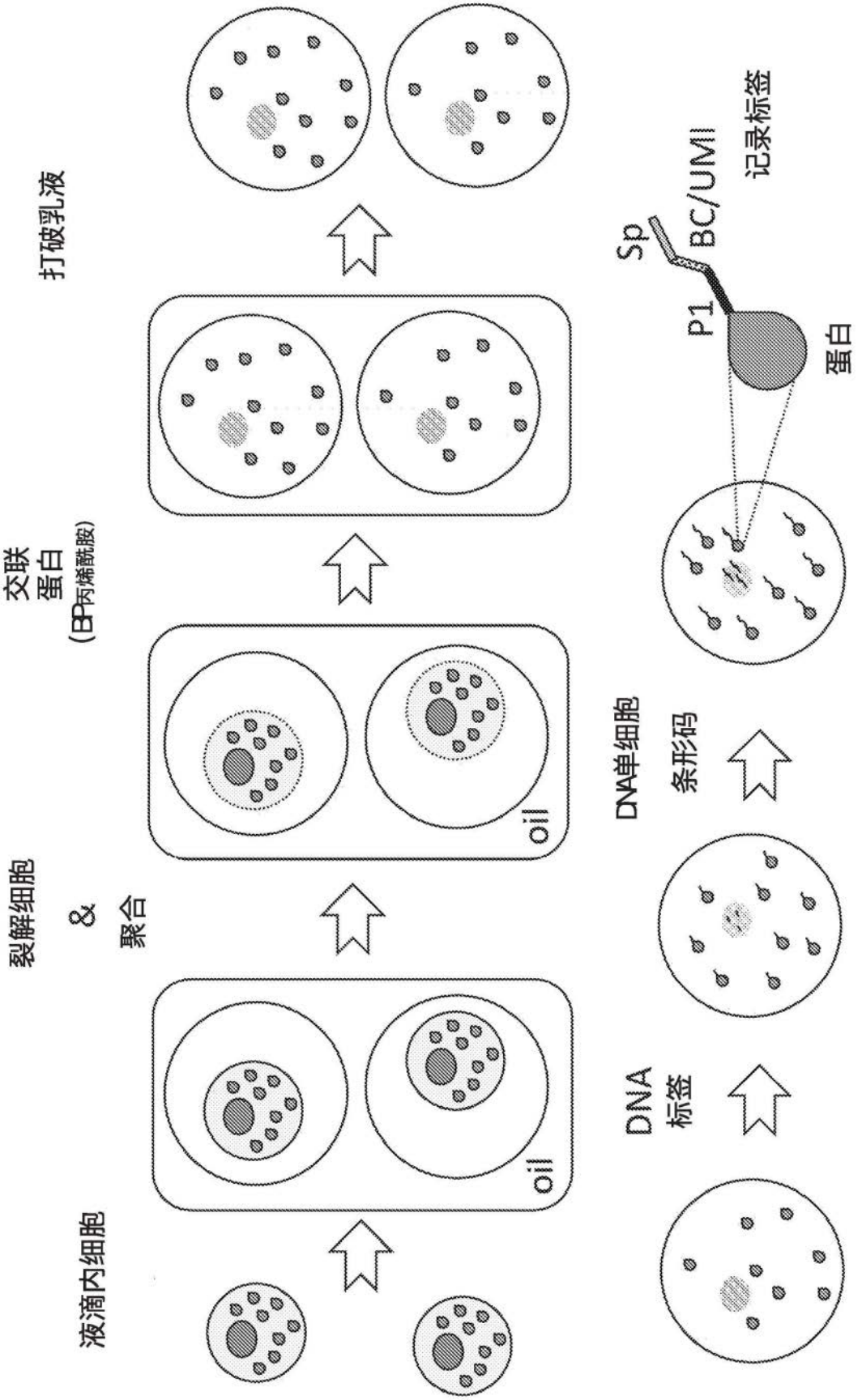

液滴内细胞
裂解细胞
&
聚合
交联
蛋白
(BP丙烯酰胺)
打破乳液
oil
oil
DNA
标签
DNA单细胞
条形码
Sp
P1
BC/UMI
记录标签
蛋白

图38

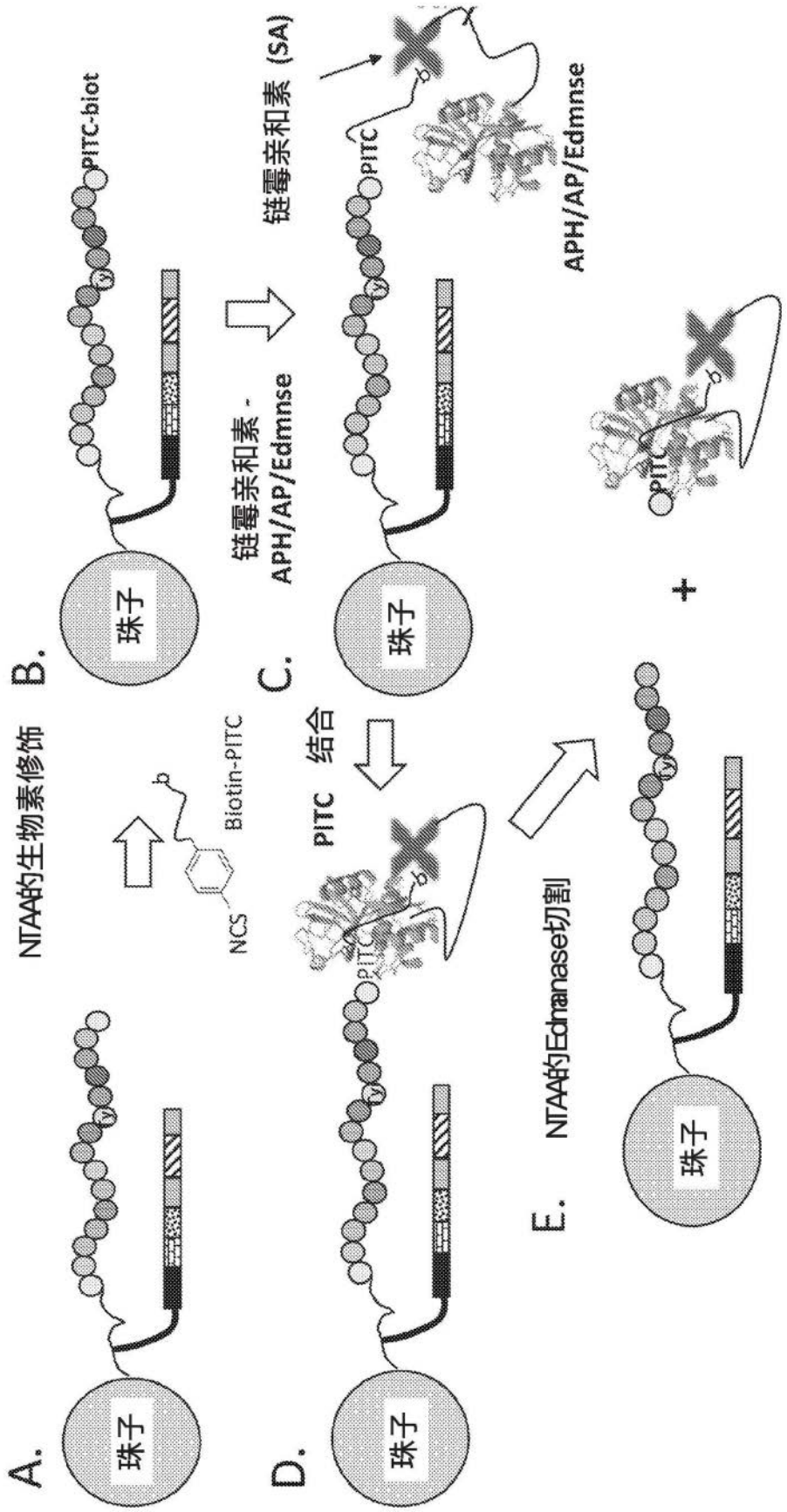

A.
珠子
NTAA的生物素修饰
Biotin-PITC
NCS
B.
PITC-biot
珠子
C.
链霉亲和素 - APH/AP/Edmnse
链霉亲和素 (SA)
PITC
珠子
APH/AP/Edmnse
D.
PITC 结合
PITC
珠子
E.
NTAA的Edmanase切割
珠子
+

图39

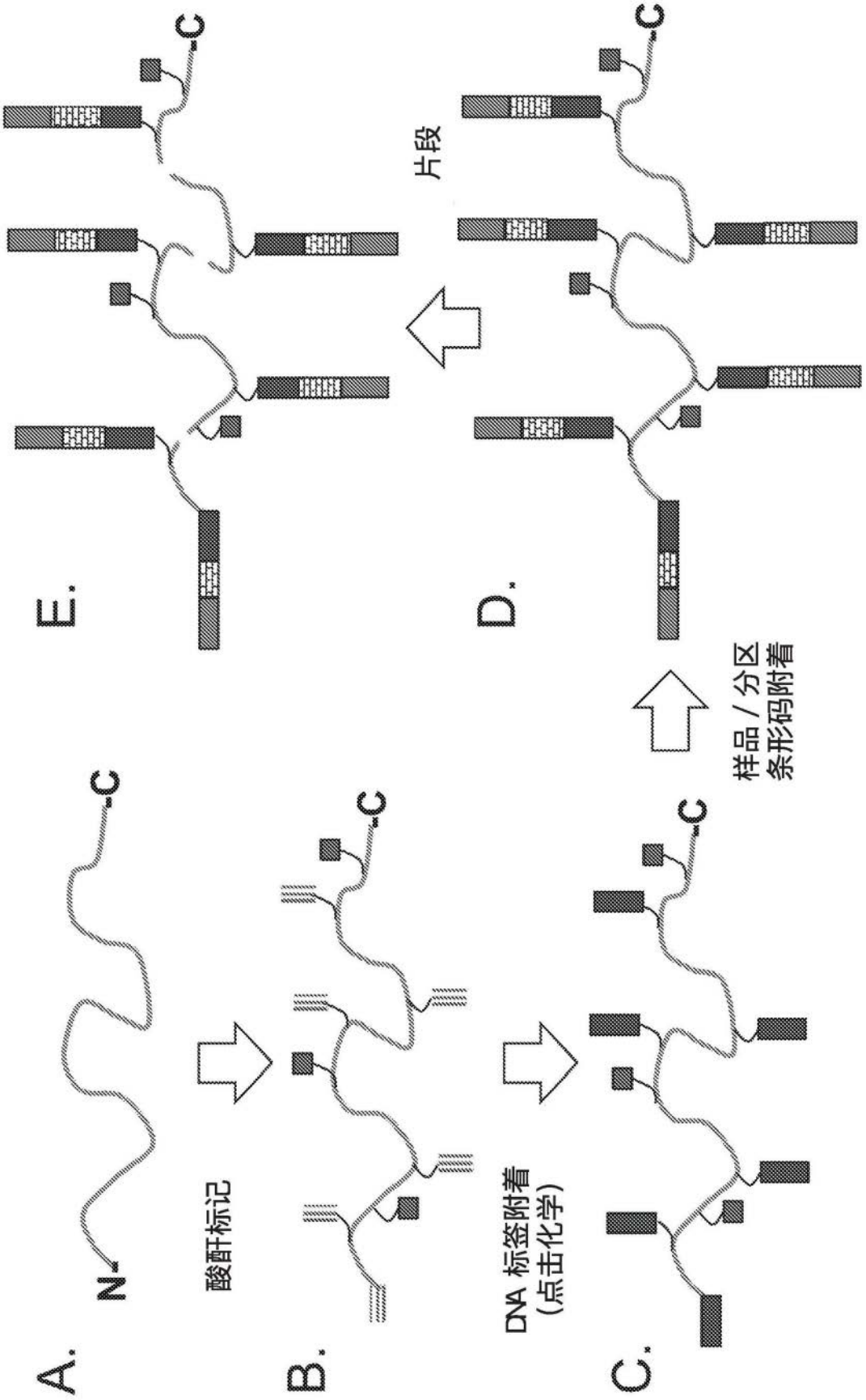



图40A-E

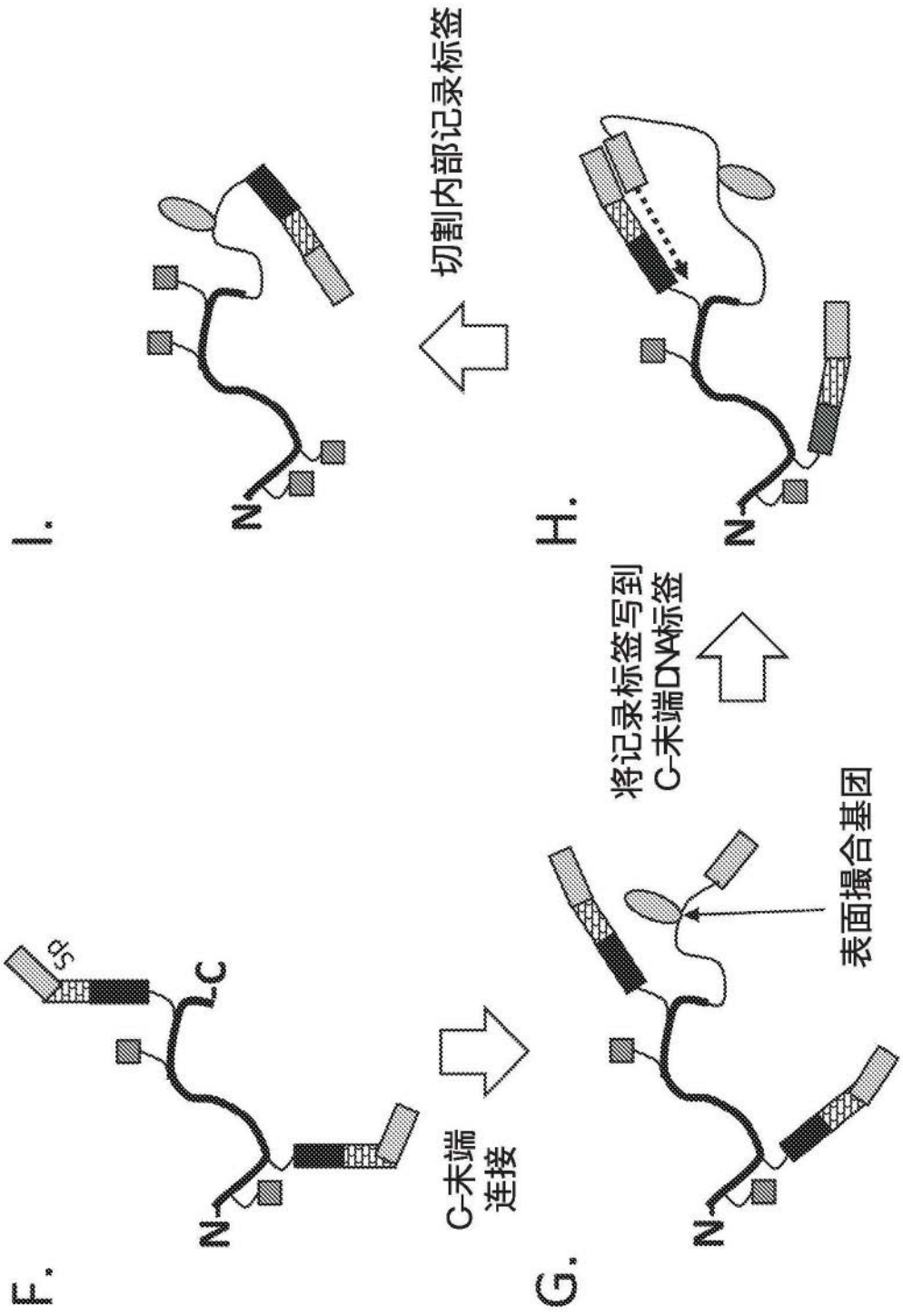



图40F-I

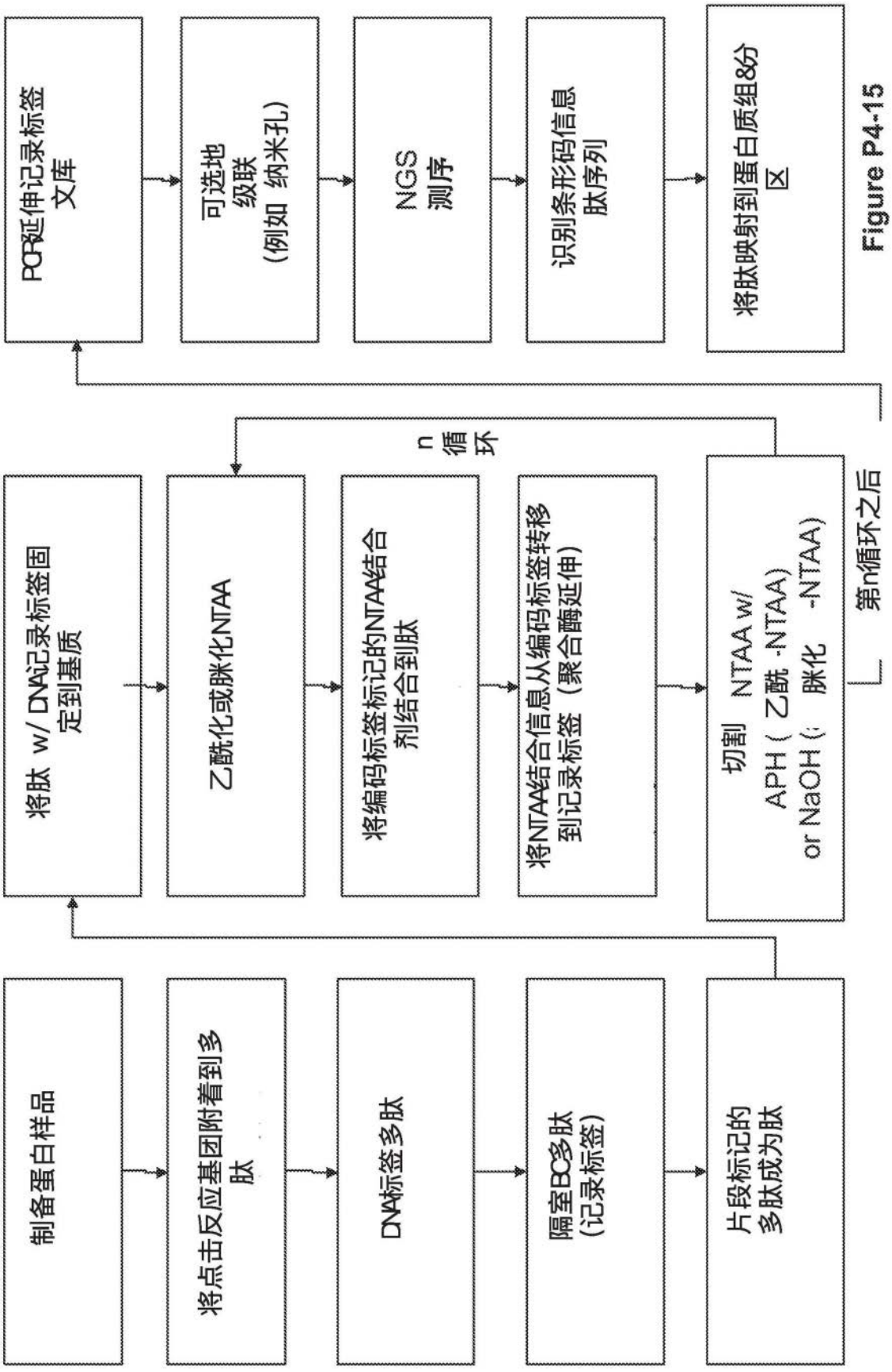

制备蛋白样品
将点击反应基团附着到多肽
DNA标签多肽
隔室BC多肽(记录标签)
片段标记的多肽成为肽
将肽 w/ DNA记录标签固定到基质
乙酰化或脒化NTAA
将编码标签标记的NTAA结合剂结合到肽
将NTAA结合信息从编码标签转移到记录标签（聚合酶延伸）
切割 NTAA w/ APH（乙酰 -NTAA) or NaOH（ 脒化 -NTAA)
第n循环之后
n 循环
PCR延伸记录标签文库
可选地级联(例如 纳米孔)
NGS 测序
识别条形码信息肽序列
将肽映射到蛋白质组&分区
Figure P4-15

图41

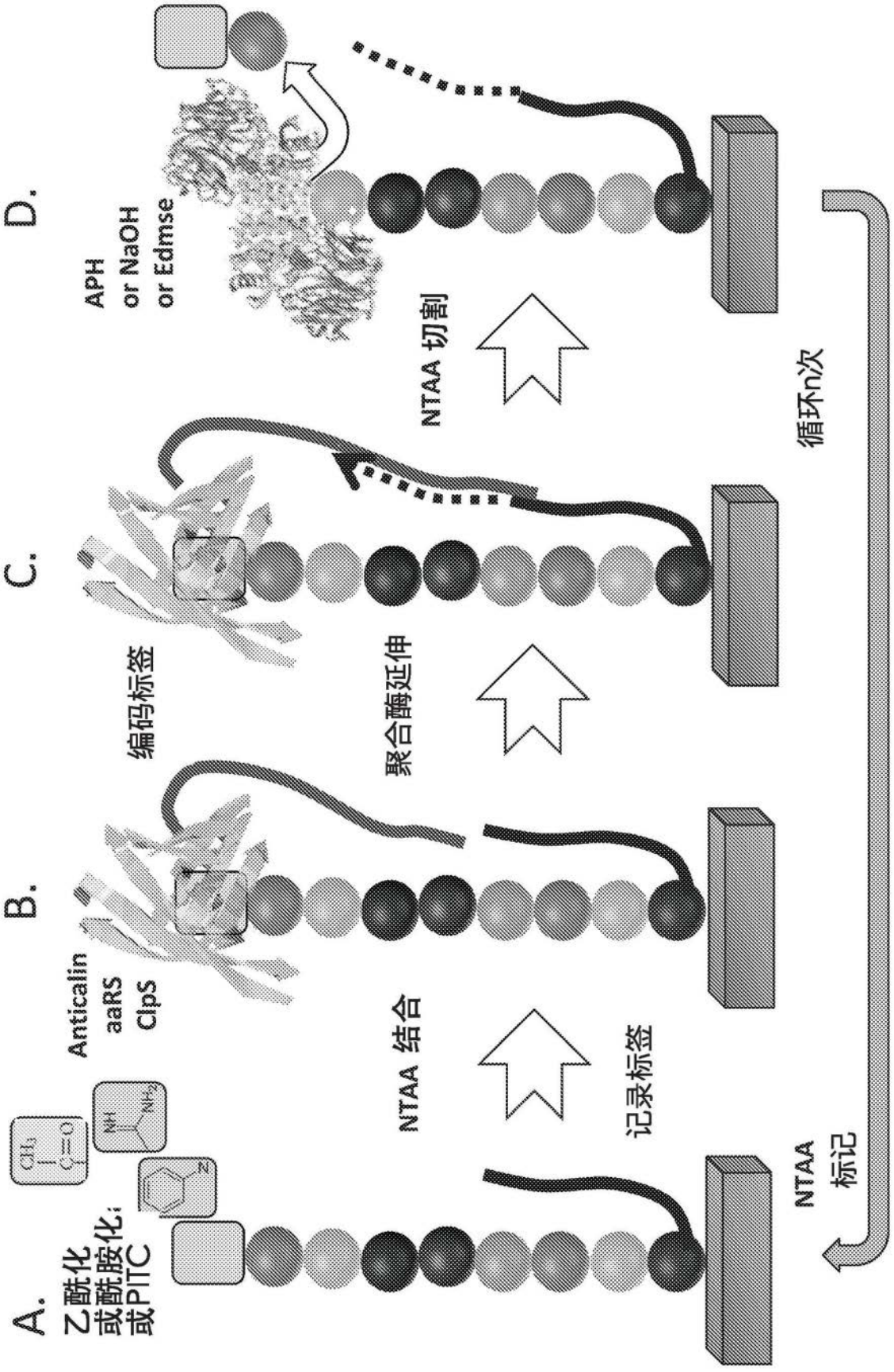

A.
乙酰化
或酰胺化;
或PITC
B.
Anticalin
aaRS
ClpS
NTAA 结合
记录标签
C.
编码标签
聚合酶延伸
D.
APH
or NaOH
or Edmse
NTAA 切割
循环n次
NTAA
标记

图42

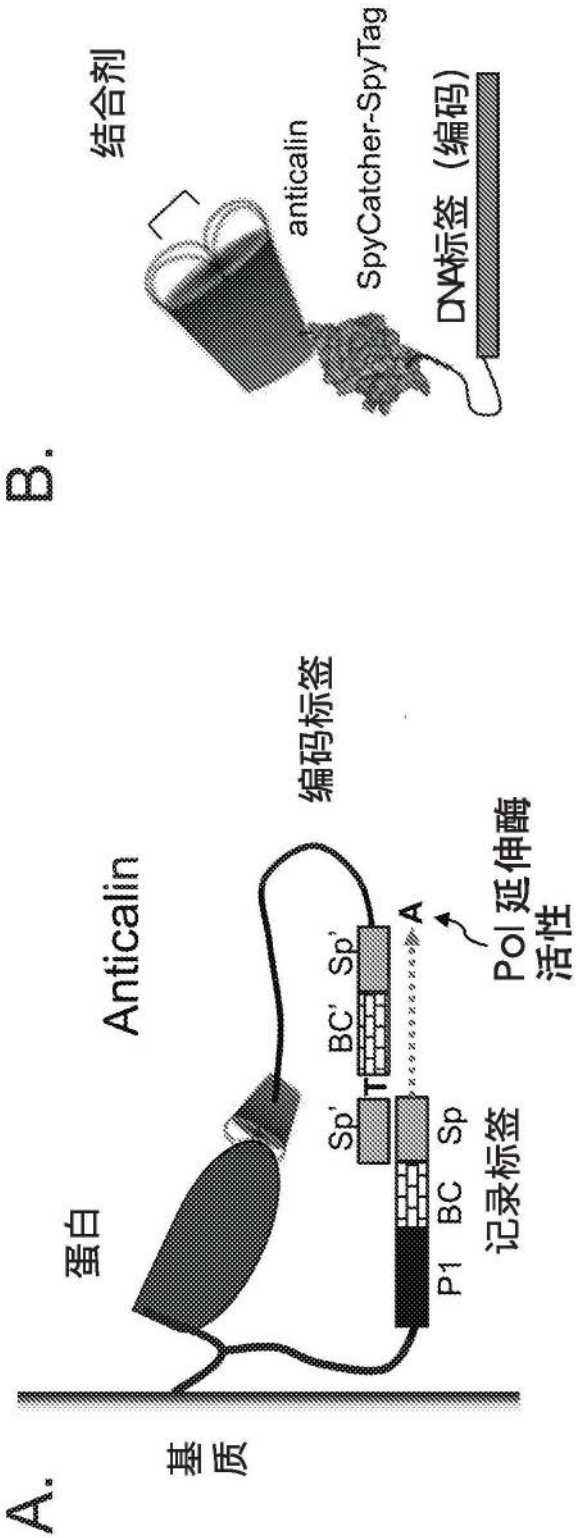

B.
结合剂
anticalin
SpyCatcher-SpyTag
DNA标签（编码）
A.
基质
蛋白
Anticalin
编码标签
Sp'
BC'
T
A
Pol 延伸酶活性
Sp'
Sp
P1
BC
记录标签

图43

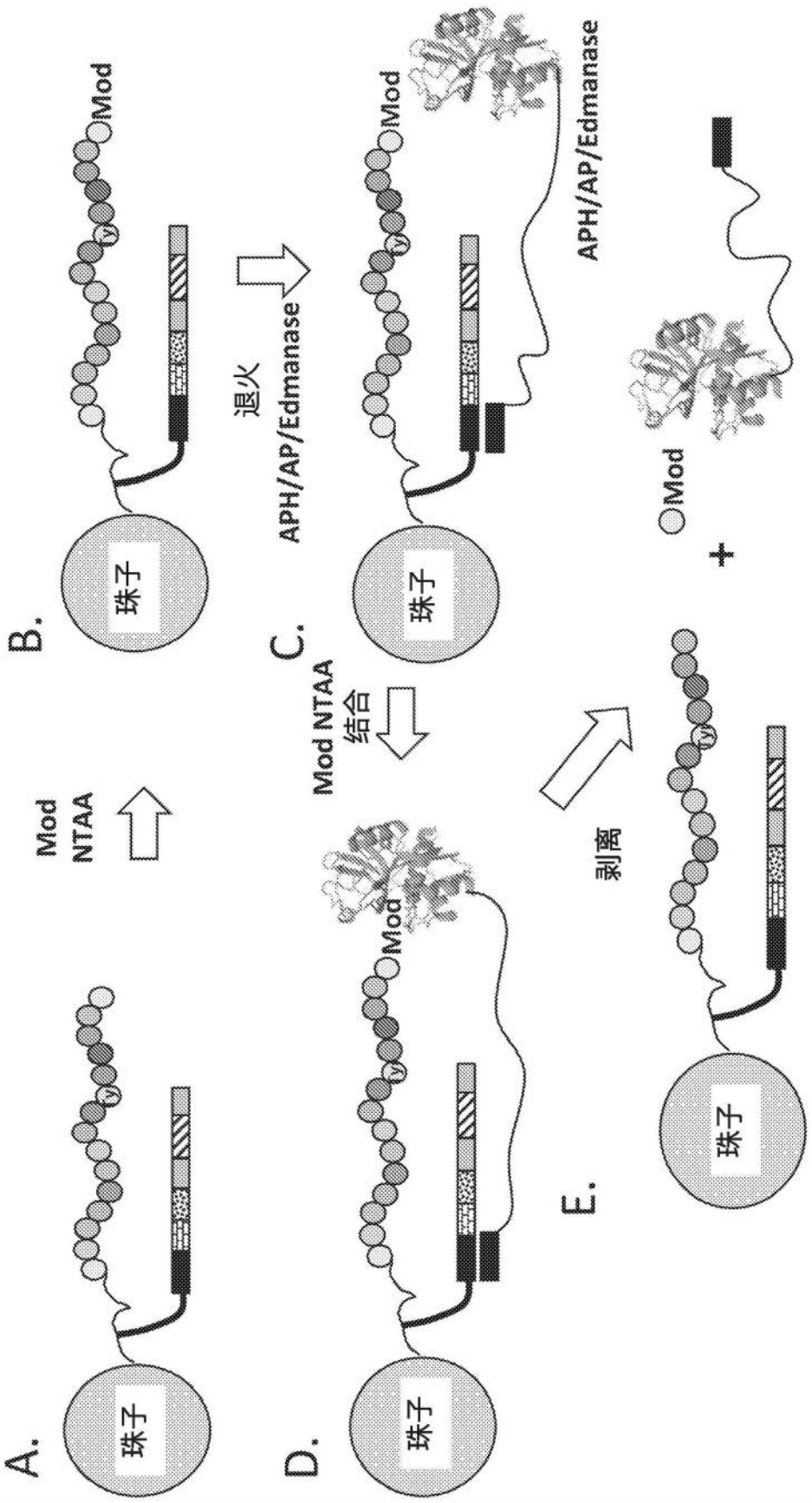

A.
珠子
Mod
NTAA
B.
珠子
Mod
退火
APH/AP/Edmanase
C.
珠子
Mod
APH/AP/Edmanase
Mod NTAA
结合
D.
珠子
Mod
剥离
E.
珠子
Mod
+

图44

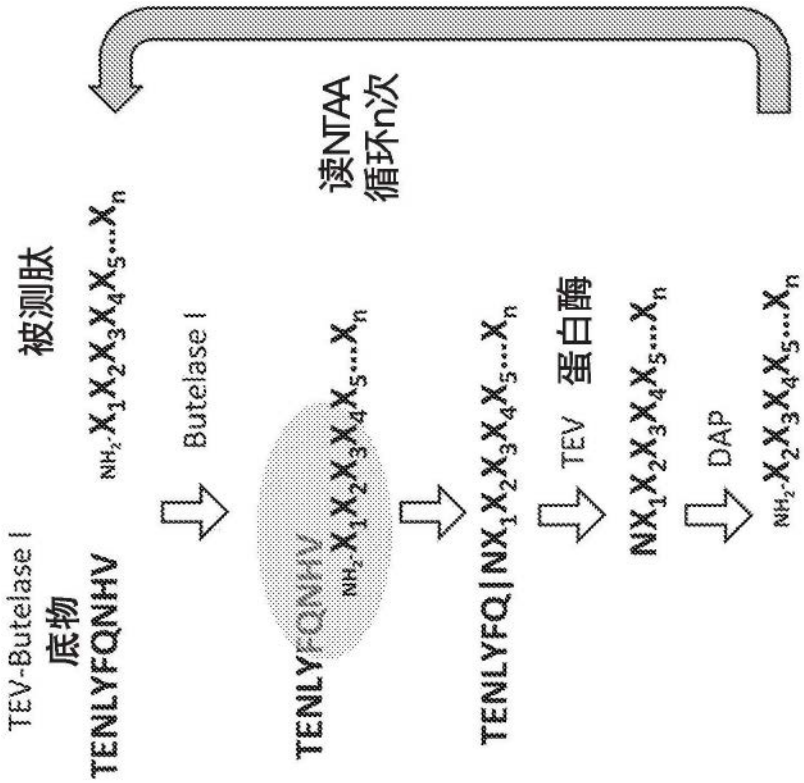

TEV-Butelase 1
底物
TENLYFQNHV
被测肽
$_{NH_2}$-$X_1X_2X_3X_4X_5...X_n$
Butelase 1
TENLYFQNHV
$_{NH_2}$-$X_1X_2X_3X_4X_5...X_n$
TENLYFQ|N$X_1X_2X_3X_4X_5...X_n$
TEV 蛋白酶
N$X_1X_2X_3X_4X_5...X_n$
DAP
$_{NH_2}$-$X_2X_3X_4X_5...X_n$
读NTAA
循环n次

图45

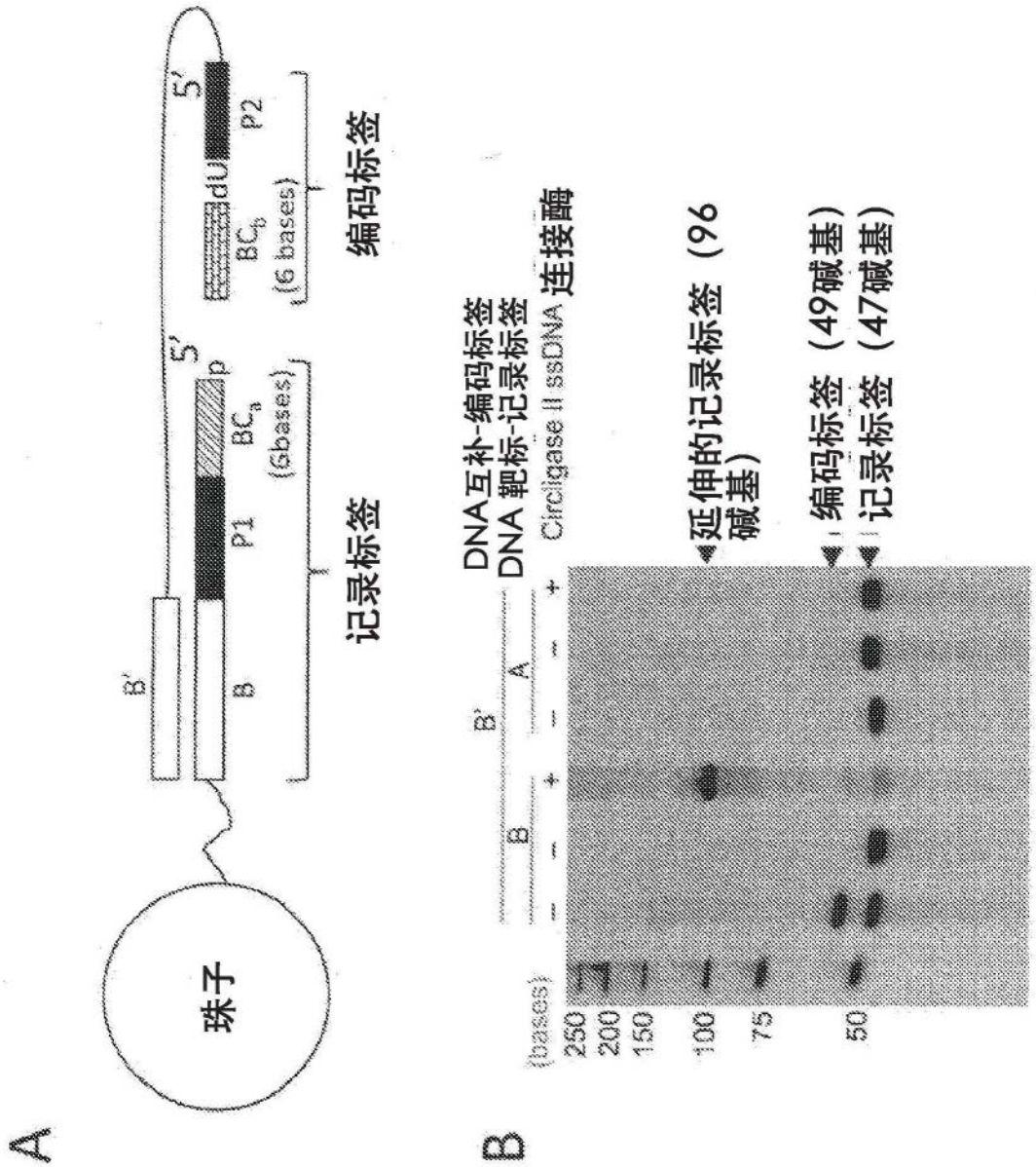

A
珠子
B'
B
P1
BCn
(Gbases)
记录标签
5'
P
5'
dU
BC6
P2
(6 bases)
编码标签
B
DNA互补-编码标签
DNA 靶标-记录标签
Circligase II ssDNA 连接酶
B'
A
B
(bases)
250
200
150
100
75
50
延伸的记录标签（96碱基）
编码标签（49碱基）
记录标签（47碱基）

图46A-B

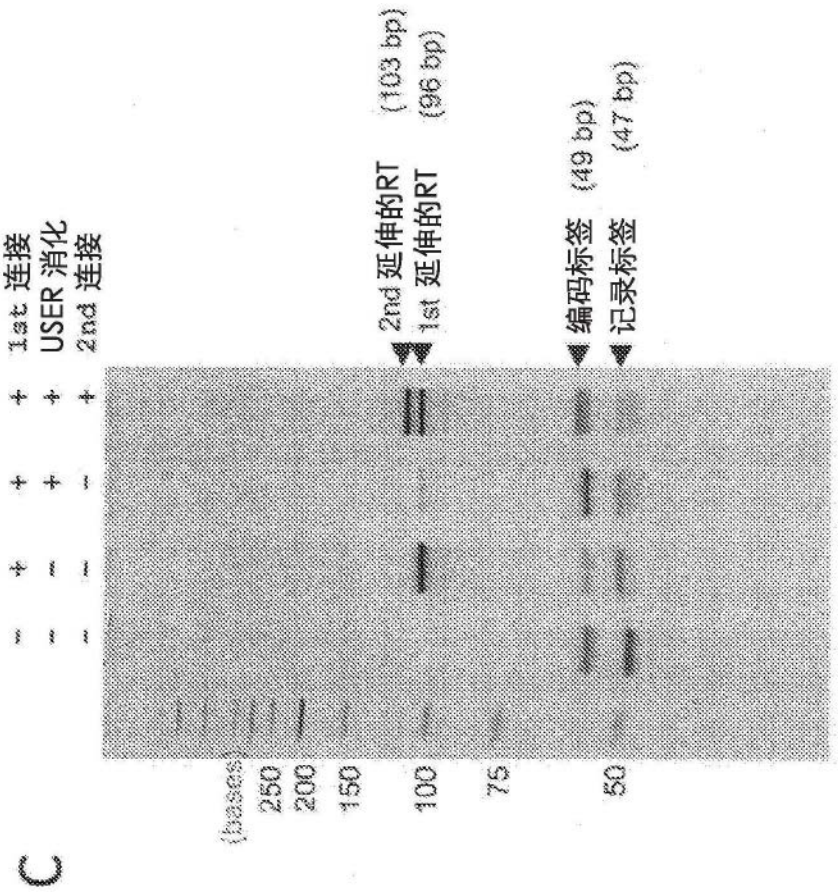

C
1st 连接
USER 消化
2nd 连接
2nd 延伸的RT (103 bp)
1st 延伸的RT (96 bp)
编码标签 (49 bp)
记录标签 (47 bp)
(bases)
250
200
150
100
75
50

图46C

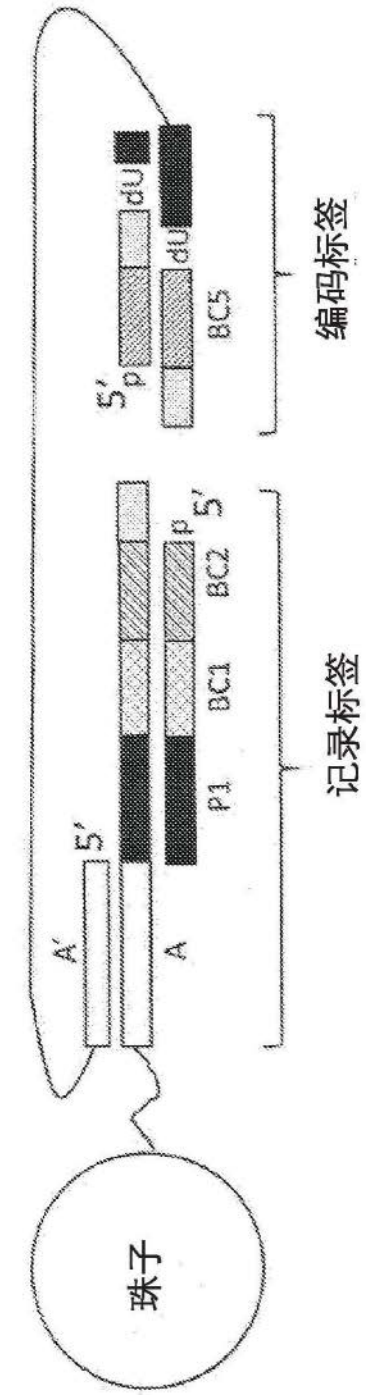



图47A

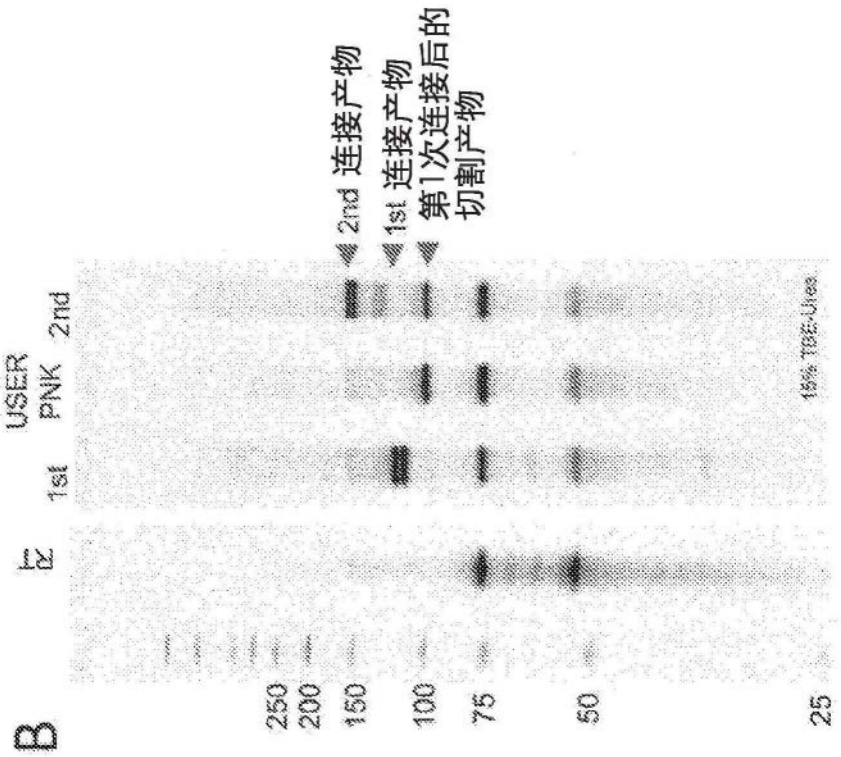

B
kb
250
200
150
100
75
50
25
USER
PNK
1st
2nd
2nd 连接产物
1st 连接产物
第1次连接后的
切割产物
15% TBE-Urea

图47B

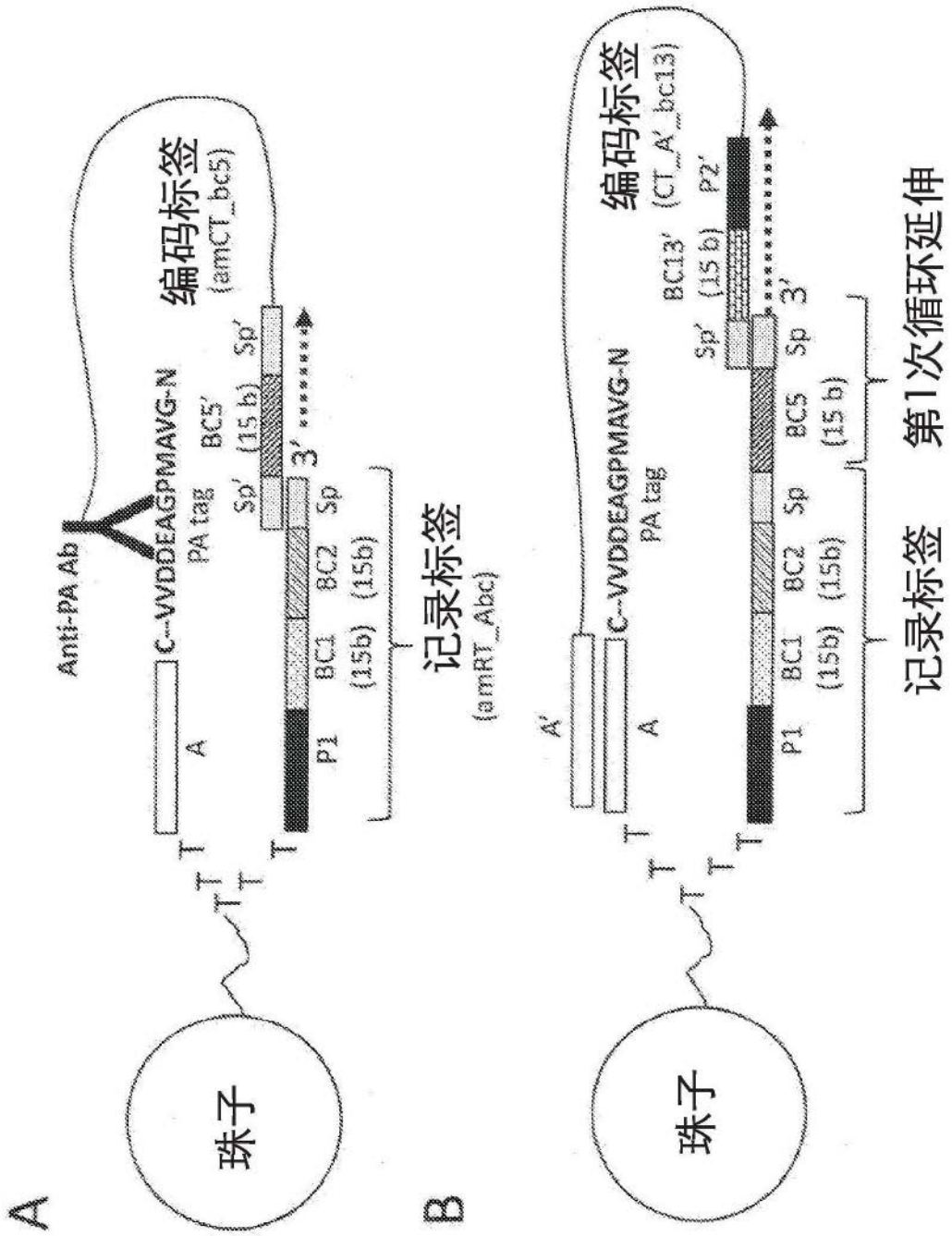

A
珠子
Anti-PA Ab
C--VVDDEAGPMAVG-N
PA tag
编码标签
(amCT_bc5)
A
BC5'
(15 b)
Sp'
Sp'
3'
P1
BC1
(15b)
BC2
(15b)
Sp
记录标签
(amRT_Abc)
B
珠子
A'
C--VVDDEAGPMAVG-N
PA tag
编码标签
(CT_A'_bc13)
A
BC13'
(15 b)
P2'
Sp'
3'
P1
BC1
(15b)
BC2
(15b)
Sp
BC5
(15 b)
Sp
记录标签
第1次循环延伸

图48A-B

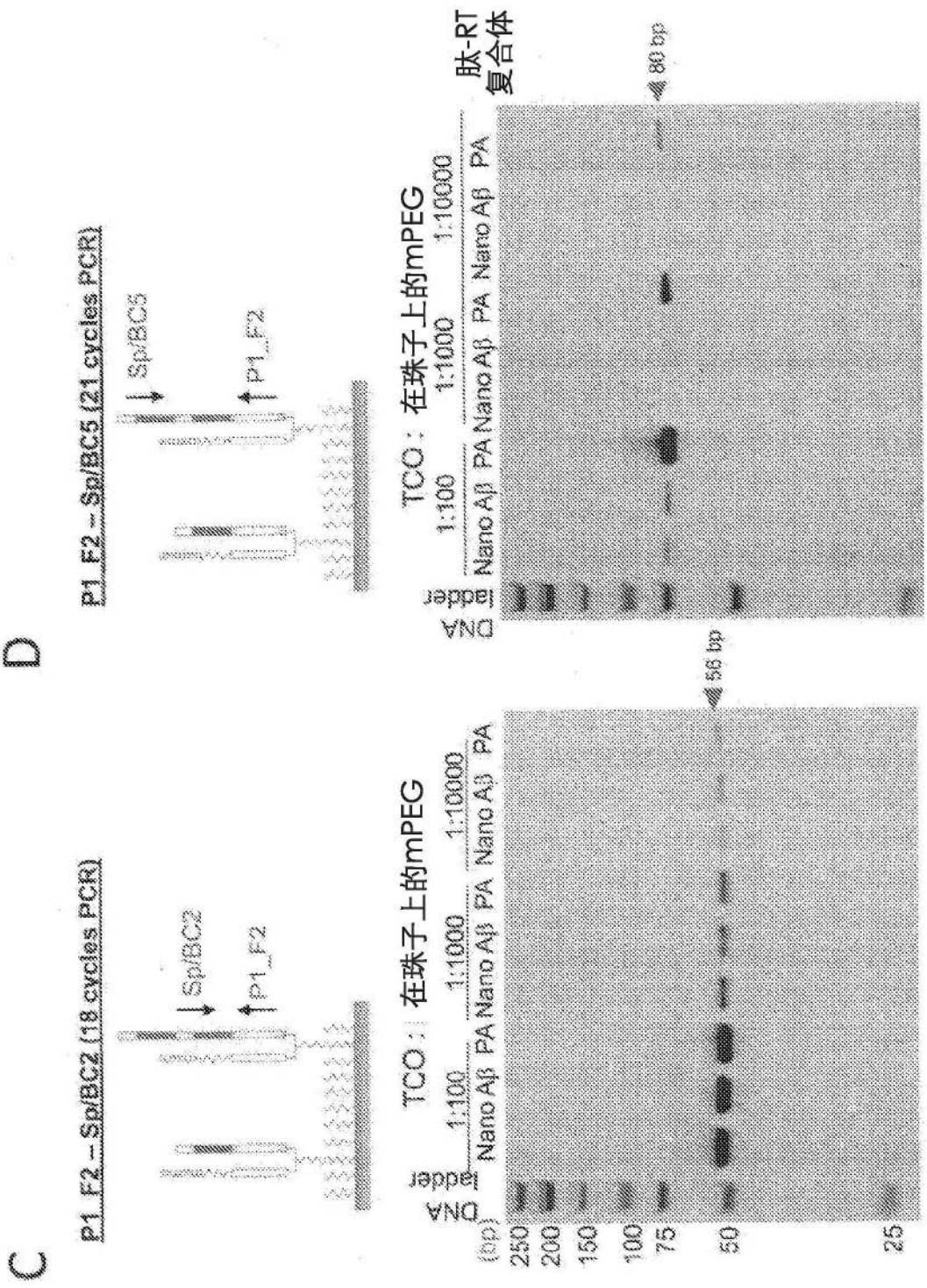

C
P1_F2 – Sp/BC2 (18 cycles PCR)
Sp/BC2
P1_F2
TCO：在珠子上的mPEG
DNA ladder
1:100
1:1000
1:10000
Nano Aβ
PA
Nano Aβ
PA
Nano Aβ
PA
(bp)
250
200
150
100
75
50
25
56 bp
D
P1_F2 – Sp/BC5 (21 cycles PCR)
Sp/BC5
P1_F2
TCO：在珠子上的mPEG
肽-RT复合体
DNA ladder
1:100
1:1000
1:10000
Nano Aβ
PA
Nano Aβ
PA
Nano Aβ
PA
80 bp

图48C-D

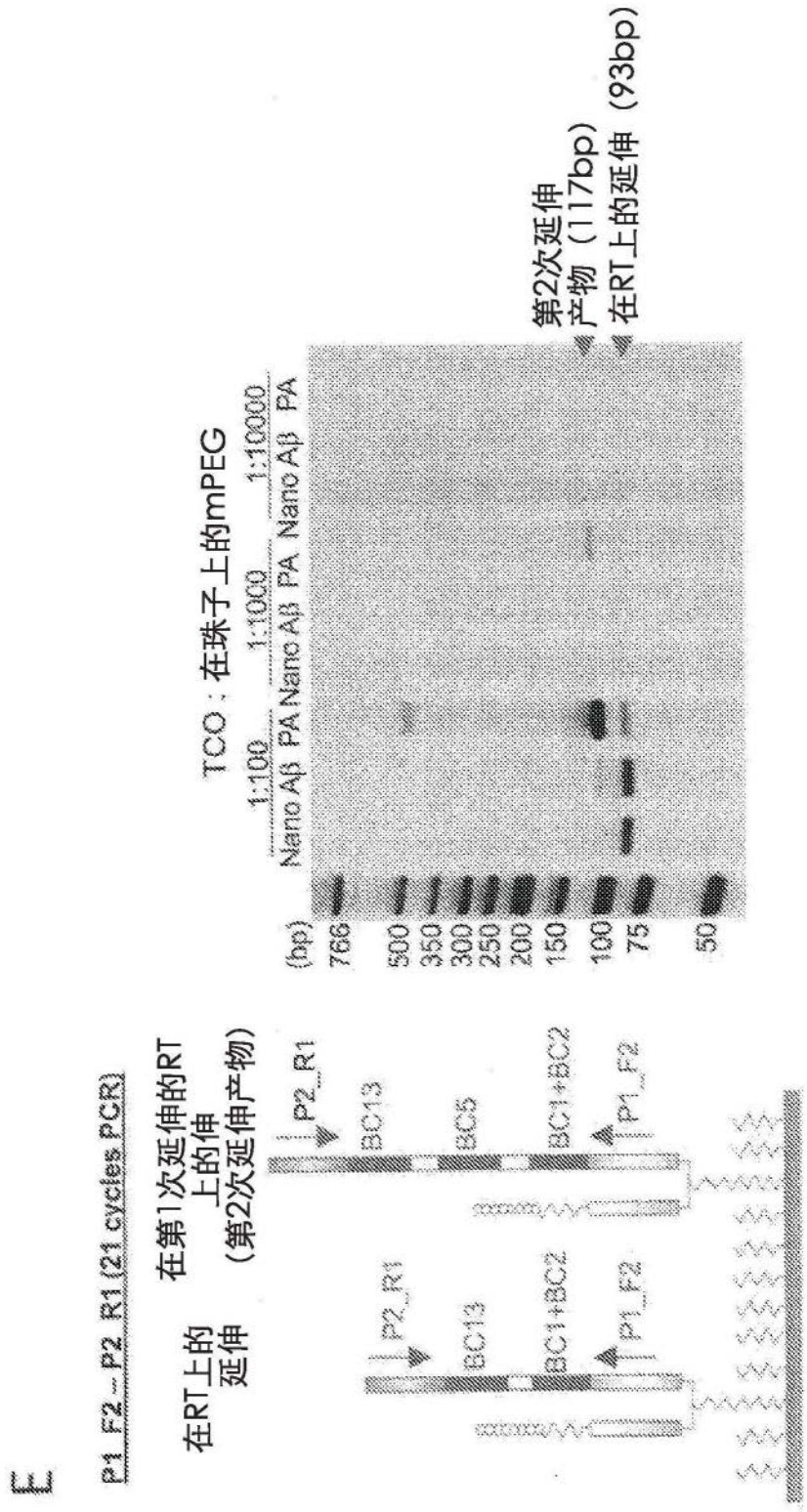

E
P1_F2 – P2_R1 (21 cycles PCR)
在RT上的延伸
在第1次延伸的RT上的伸
(第2次延伸产物)
P2_R1
BC13
BC5
BC1+BC2
P1_F2
TCO：在珠子上的mPEG
1:100
1:1000
1:10000
Nano Aβ PA Nano Aβ PA Nano Aβ PA
(bp)
766
500
350
300
250
200
150
100
75
50
第2次延伸产物（117bp）
在RT上的延伸（93bp）

图48E

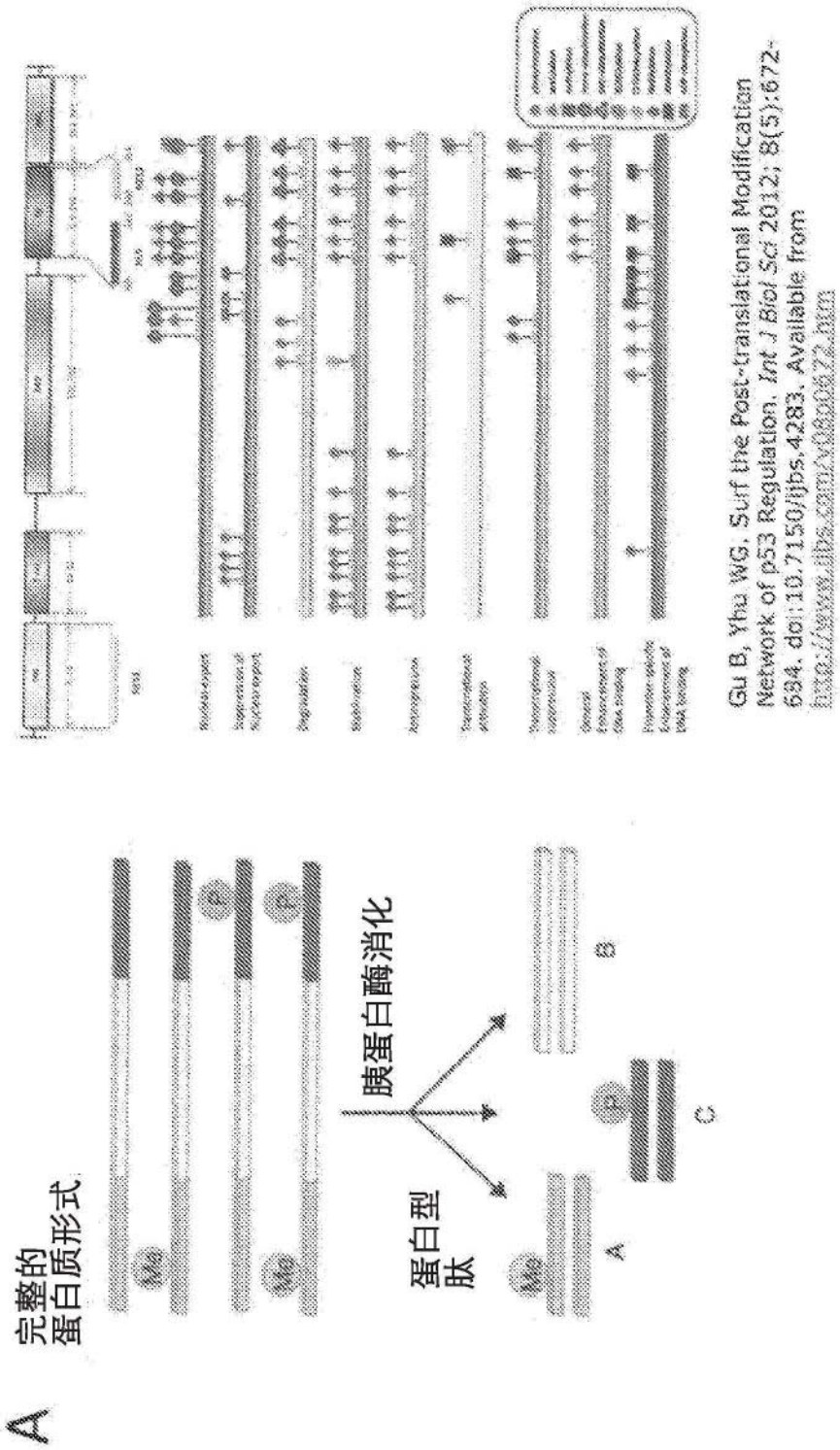

A
完整的
蛋白质形式
胰蛋白酶消化
蛋白型
肽
A
B
C
Gu B, Yhu WG, Surf the Post-translational Modification Network of p53 Regulation, Int J Biol Sci 2012; 8(5):672-684. doi:10.7150/ijbs.4283. Available from http://www.ijbs.com/v08p0672.htm

图49A

B

>sp|P04637|P53_HUMAN
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPDE
APRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAK
SVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHE
RCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNS
SCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELP
PGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPG
GSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD



测序读段“CPXQXWXDXT”
在人类蛋白质组中
唯一地映射回p53

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGP
MEEPXSDPXZEXPLSQ ZMXDLMLSPXDIEQX
SZEPPXSQEXFSD PEXNZLSPLXSQ
QETFXDLWKLXPENNZL DXEQWFXEDPZP
PQXDPSZEXPLSQE WKXLPENXZLSPL MLSXDDIEXWFTL

图49B

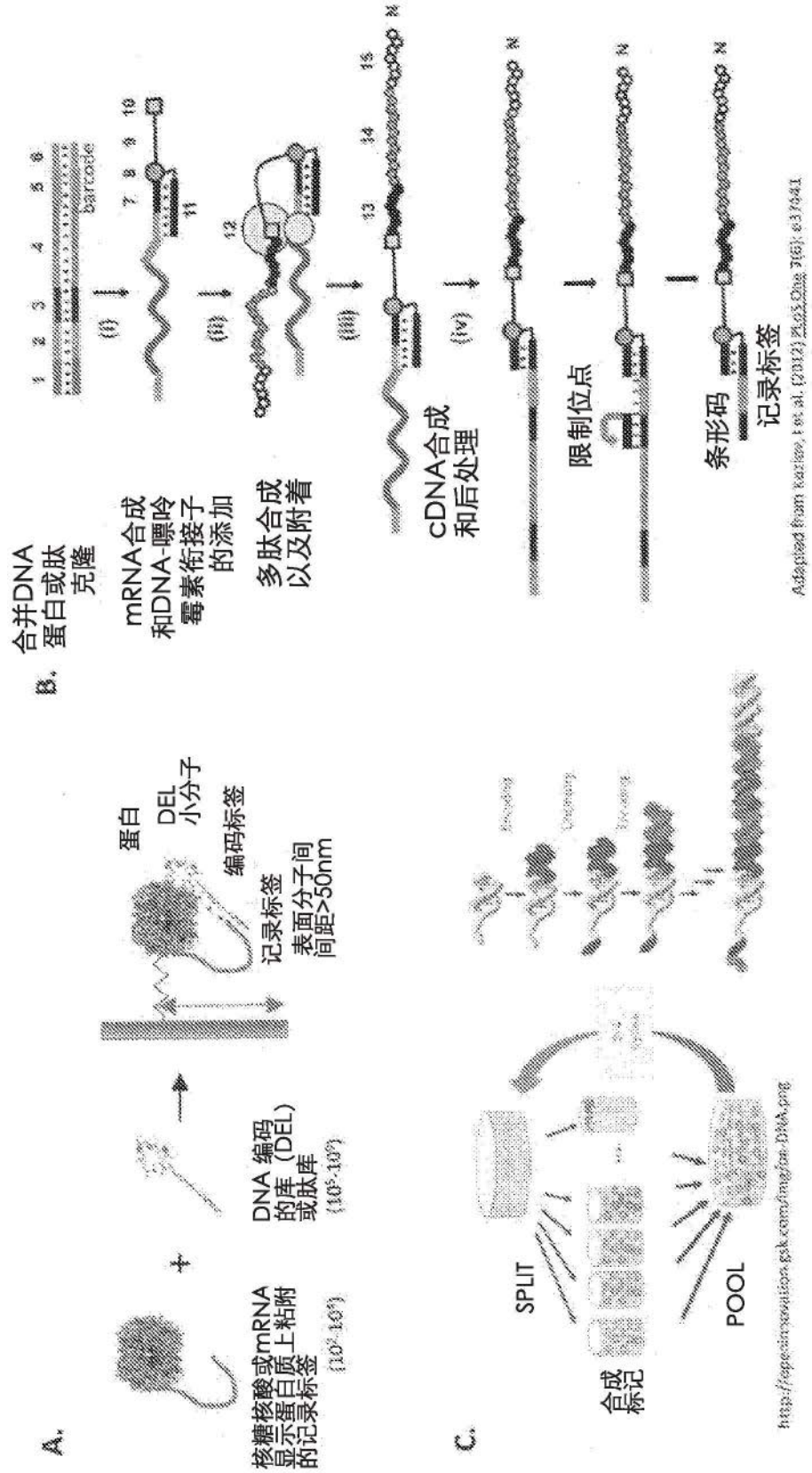

A,
核糖核酸或mRNA显示蛋白质上粘附的记录标签
DNA 编码的库 或肽库
蛋白
DEL 小分子
编码标签
记录标签
表面分子间间距>50nm
B,
合并DNA 蛋白或肽 克隆
mRNA合成 和DNA-嘌呤霉素衔接子 的添加
多肽合成 以及附着
cDNA合成 和后处理
限制位点
条形码
记录标签
C,
SPLIT
POOL
合成标记

图50

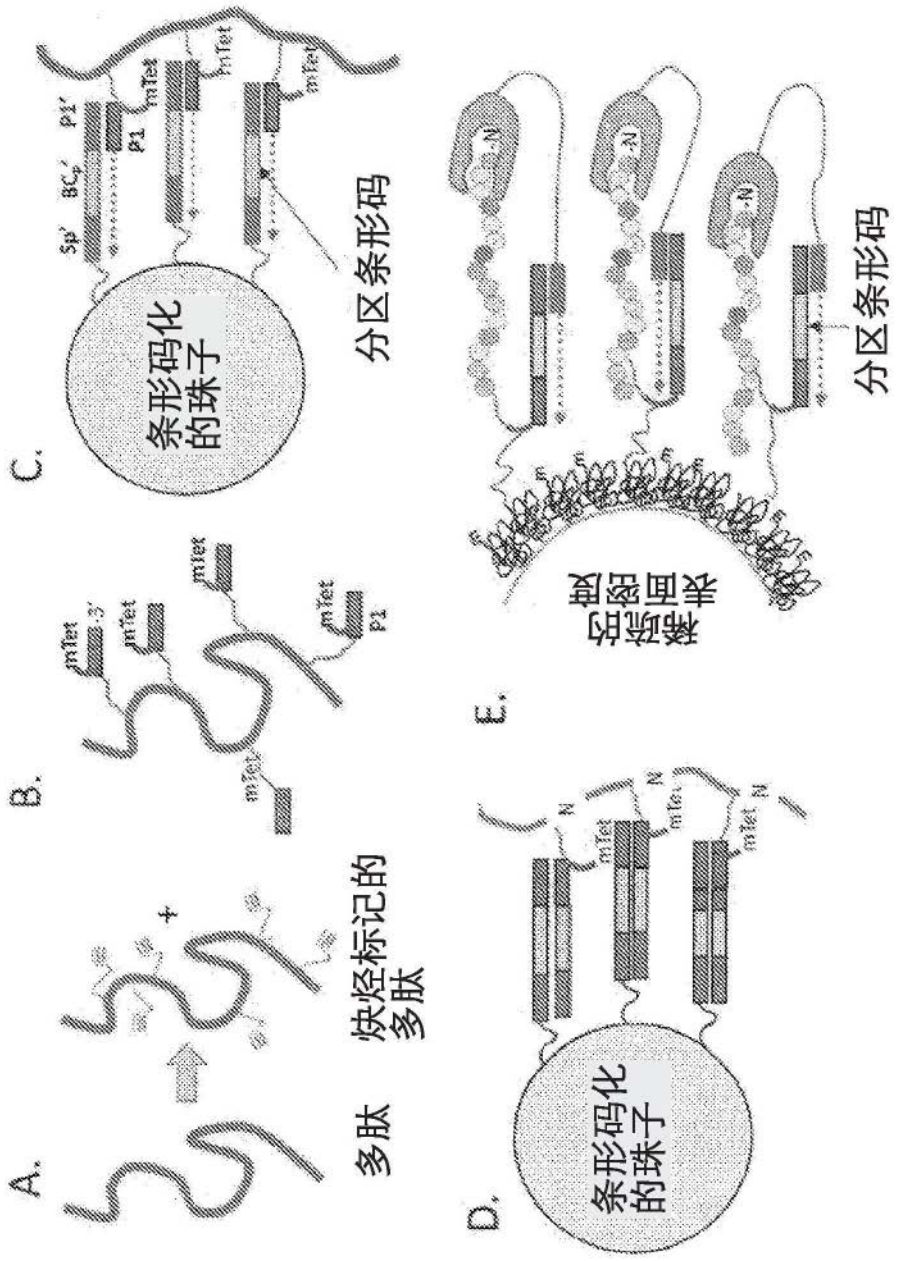



图51

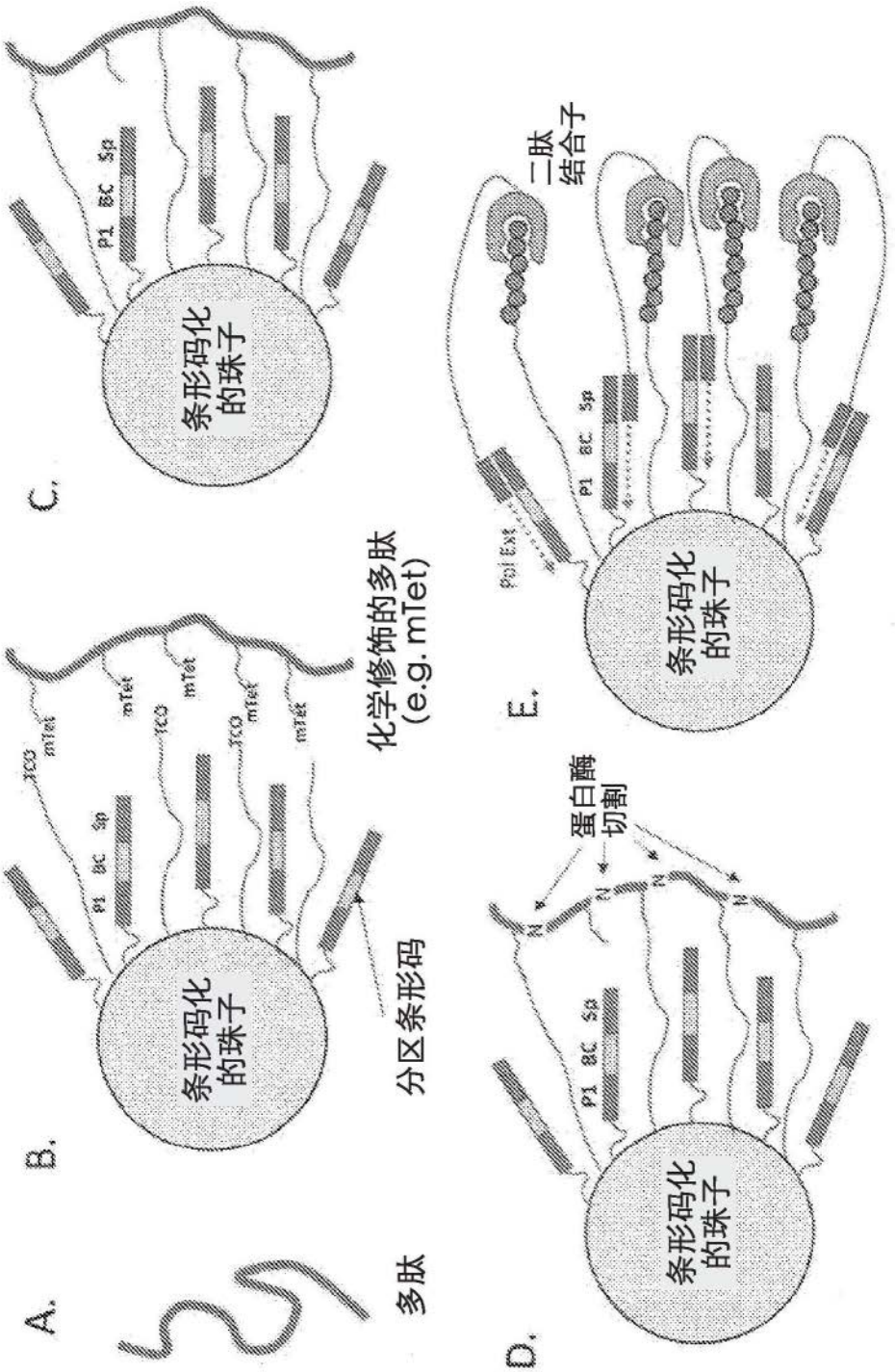
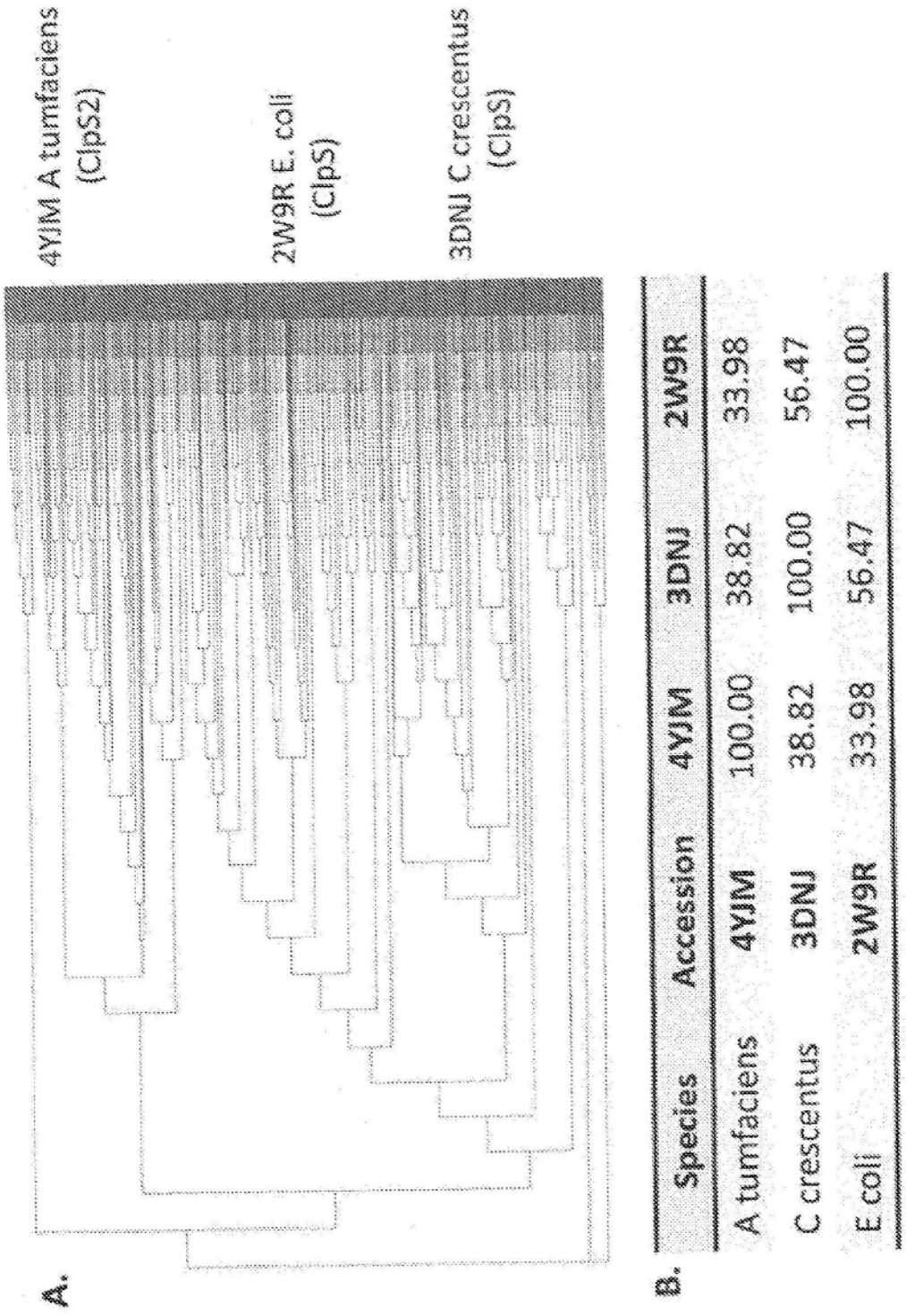

A.
多肽
B.
条形码化的珠子
P1 BC Sp
TCO
mTet
分区条形码
化学修饰的多肽
(e.g. mTet)
C.
条形码化的珠子
P1 BC Sp
D.
蛋白酶
切割
条形码化的珠子
P1 BC Sp
E.
二肽
结合子
Pol EXT
条形码化的珠子
P1 BC Sp

图52




| Species | Accession | 4YJM | 3DNJ | 2W9R |
|---|---|---|---|---|
| A tumfaciens | 4YJM | 100.00 | 38.82 | 33.98 |
| C crescentus | 3DNJ | 38.82 | 100.00 | 56.47 |
| E coli | 2W9R | 33.98 | 56.47 | 100.00 |

图53

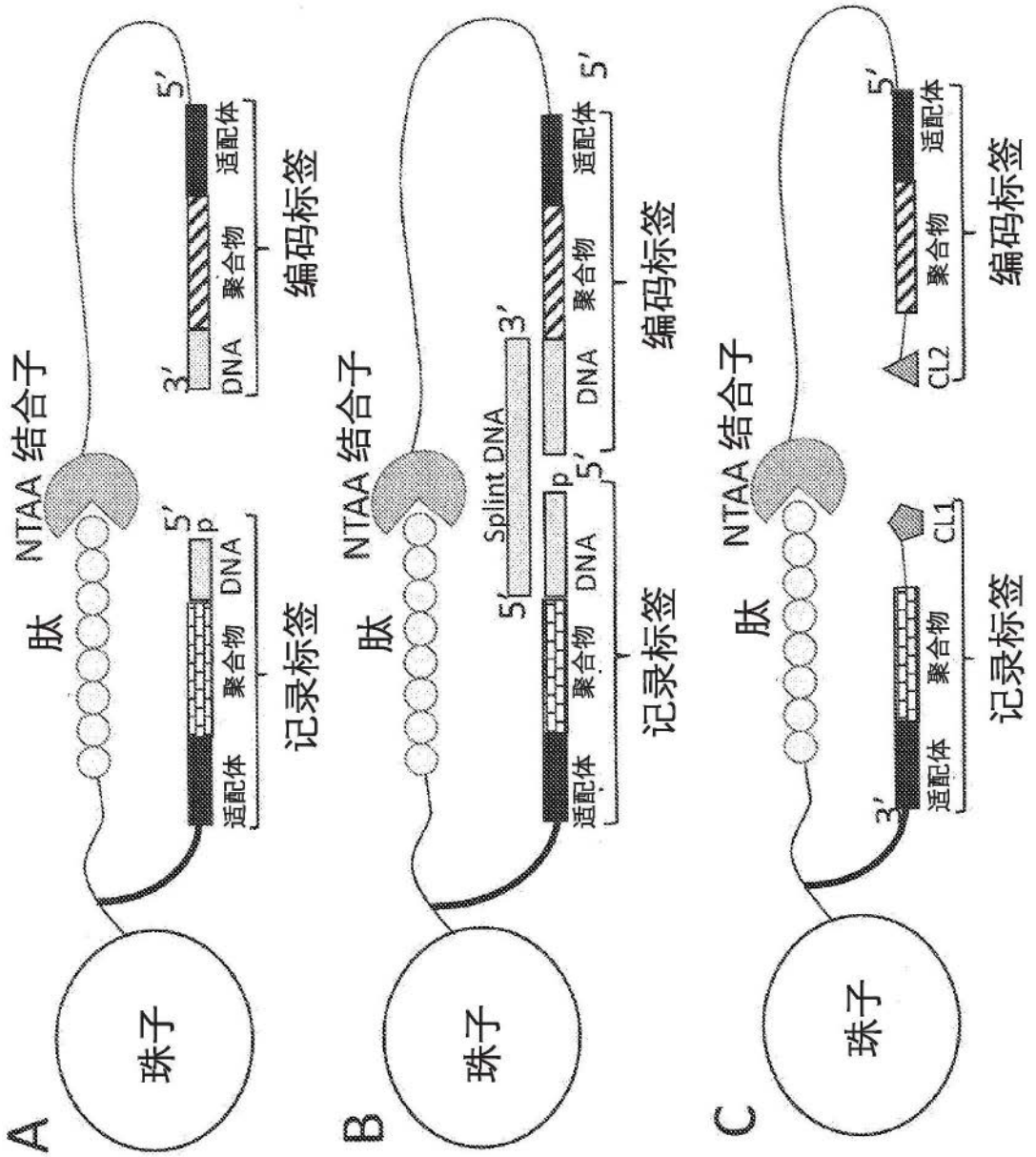



图54

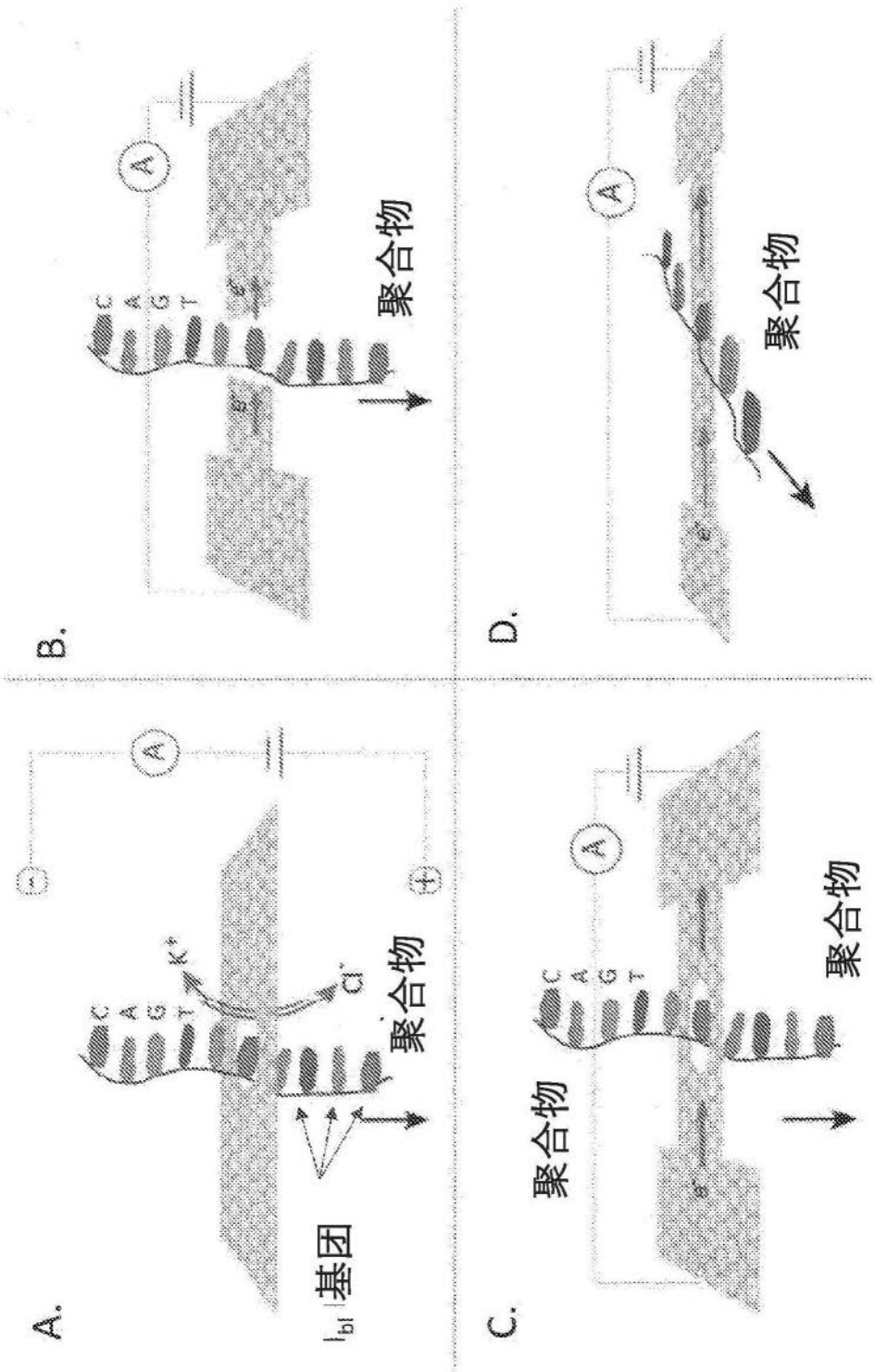



图55

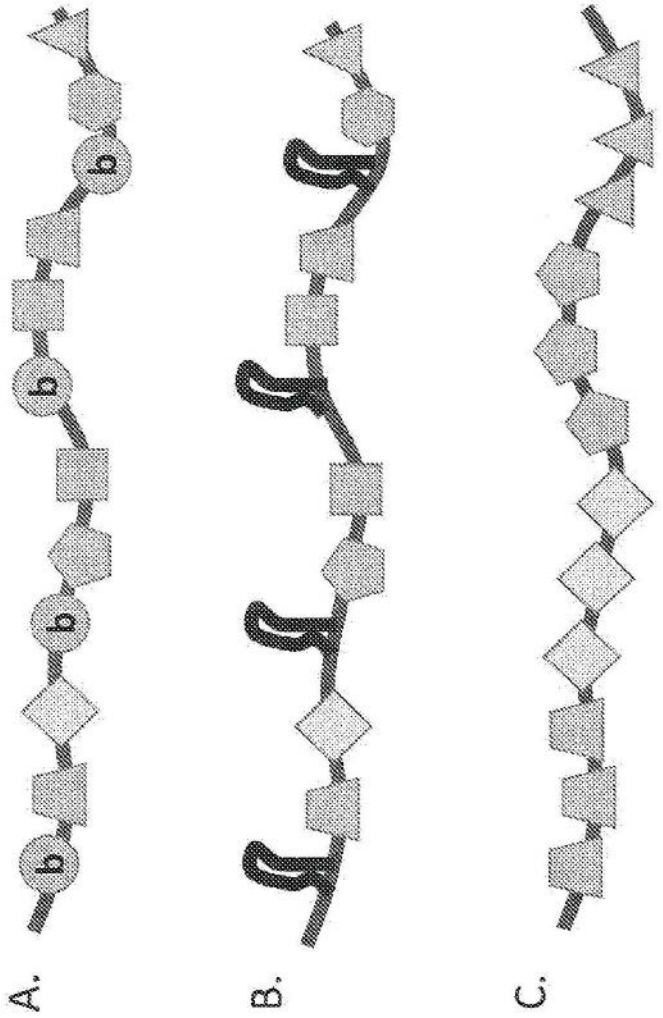



图56

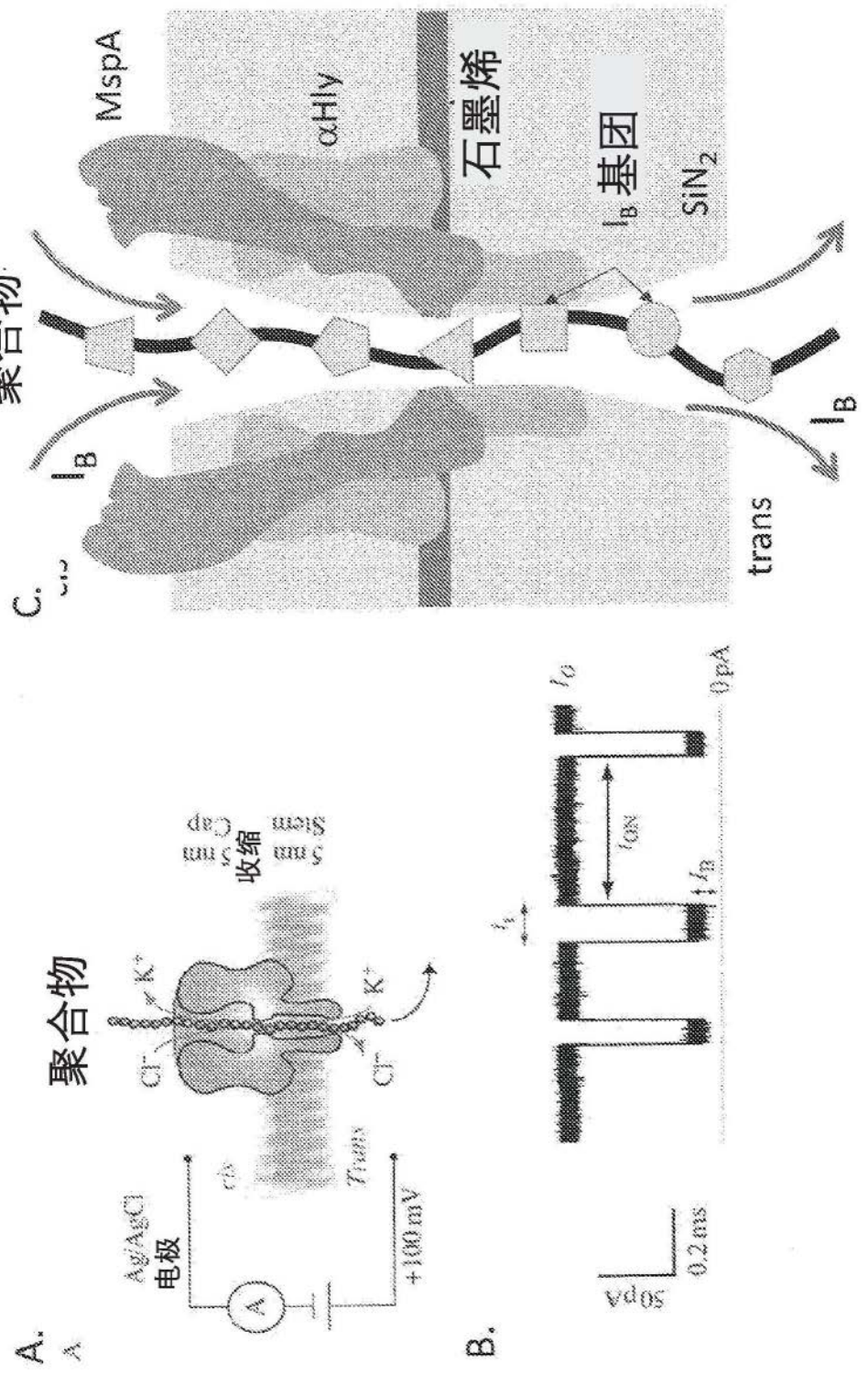

聚合物
MspA
αHly
石墨烯
$I_B$ 基团
$SiN_2$
$I_B$
trans
$I_B$
C.
聚合物
A.
Ag/AgCl
电极
+100 mV
cis
Trans
收缩
5 nm
Stem
Cap
B.
50 pA
0.2 ms
0 pA

图57

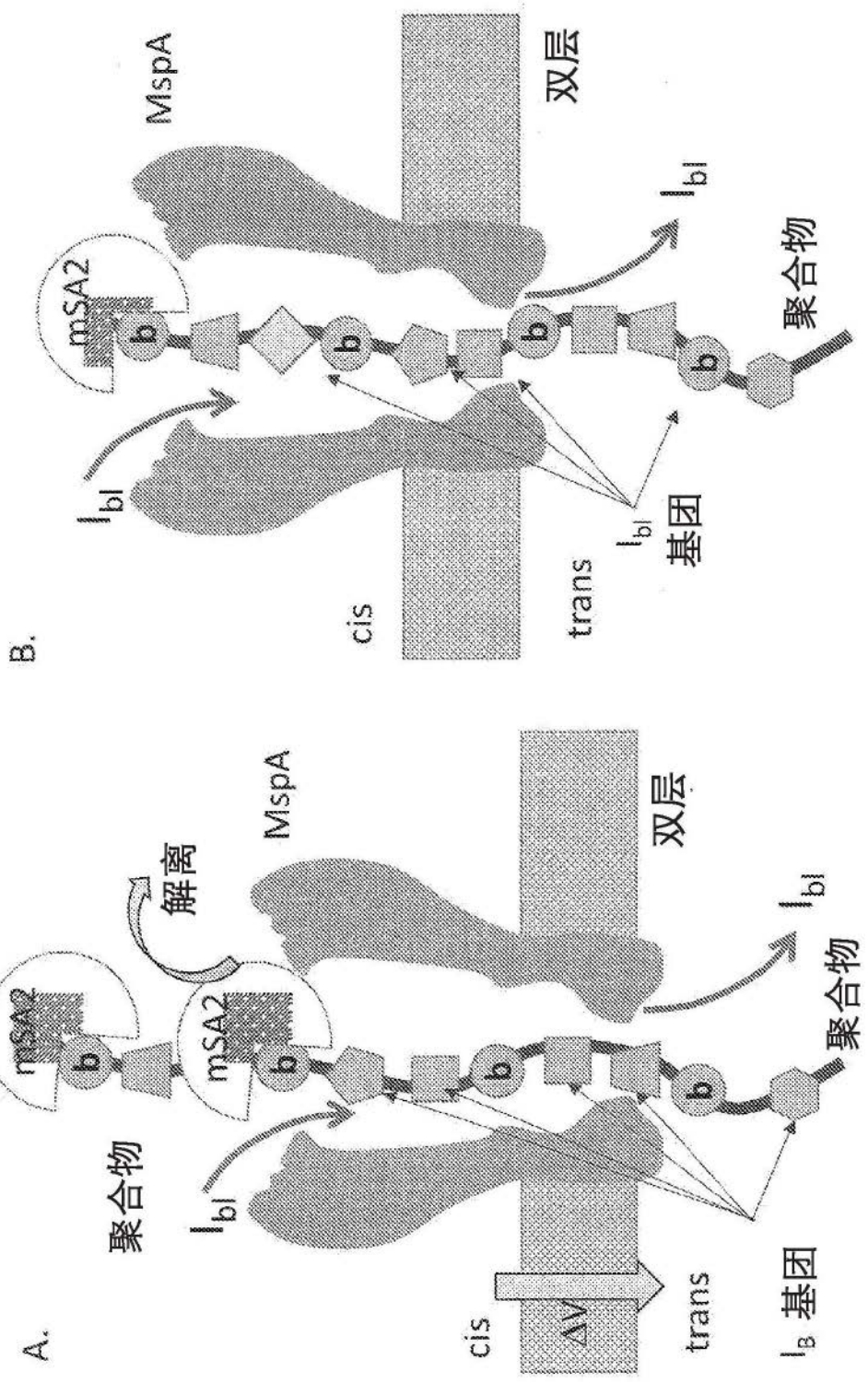



图58

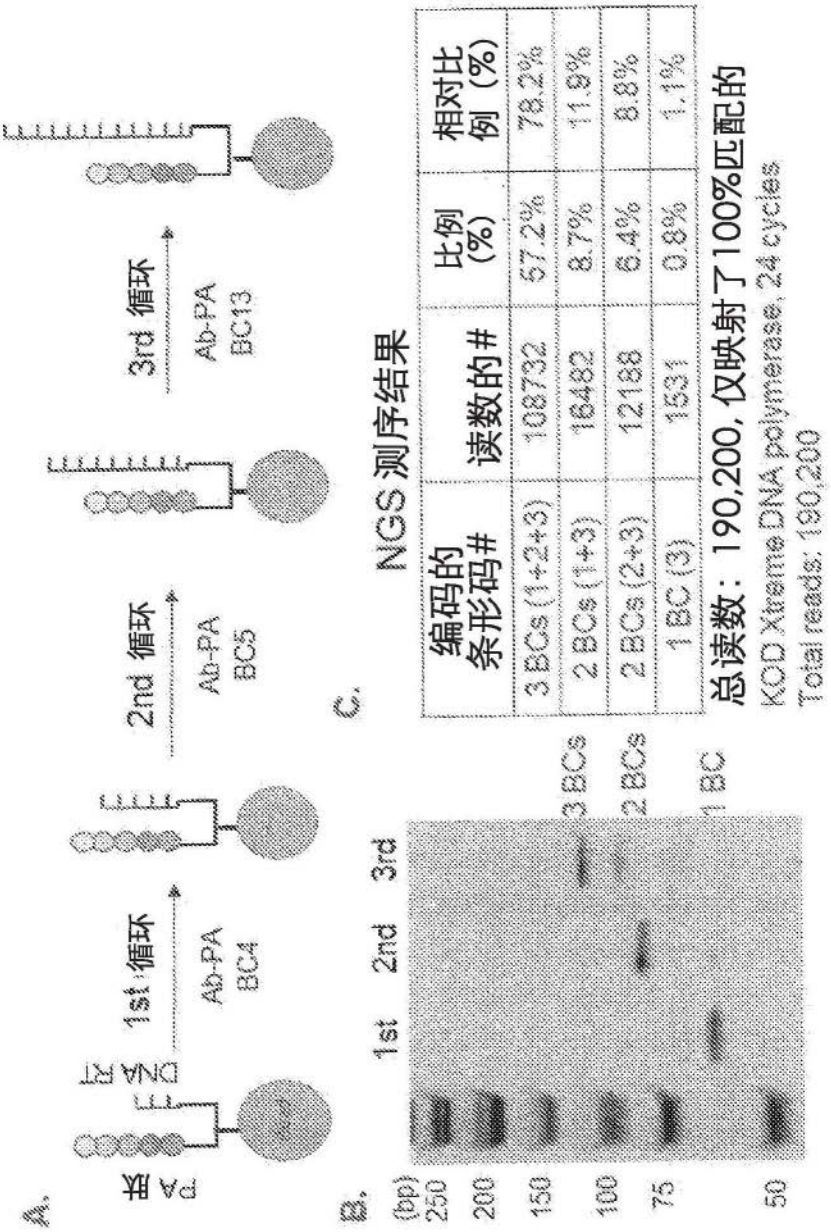

A.
PA
DNA
RT
1st 循环
Ab-PA
BC4
2nd 循环
Ab-PA
BC5
3rd 循环
Ab-PA
BC13
B.
1st
2nd
3rd
3 BCs
2 BCs
1 BC
(bp)
250
200
150
100
75
50
C.
NGS 测序结果
编码的条形码#
读数的#
比例(%)
相对比例(%)
3 BCs (1+2+3)
108732
57.2%
78.2%
2 BCs (1+3)
16482
8.7%
11.9%
2 BCs (2+3)
12188
6.4%
8.8%
1 BC (3)
1531
0.8%
1.1%
总读数：190,200，仅映射了100%匹配的
KOD Xtreme DNA polymerase, 24 cycles
Total reads: 190,200

图59

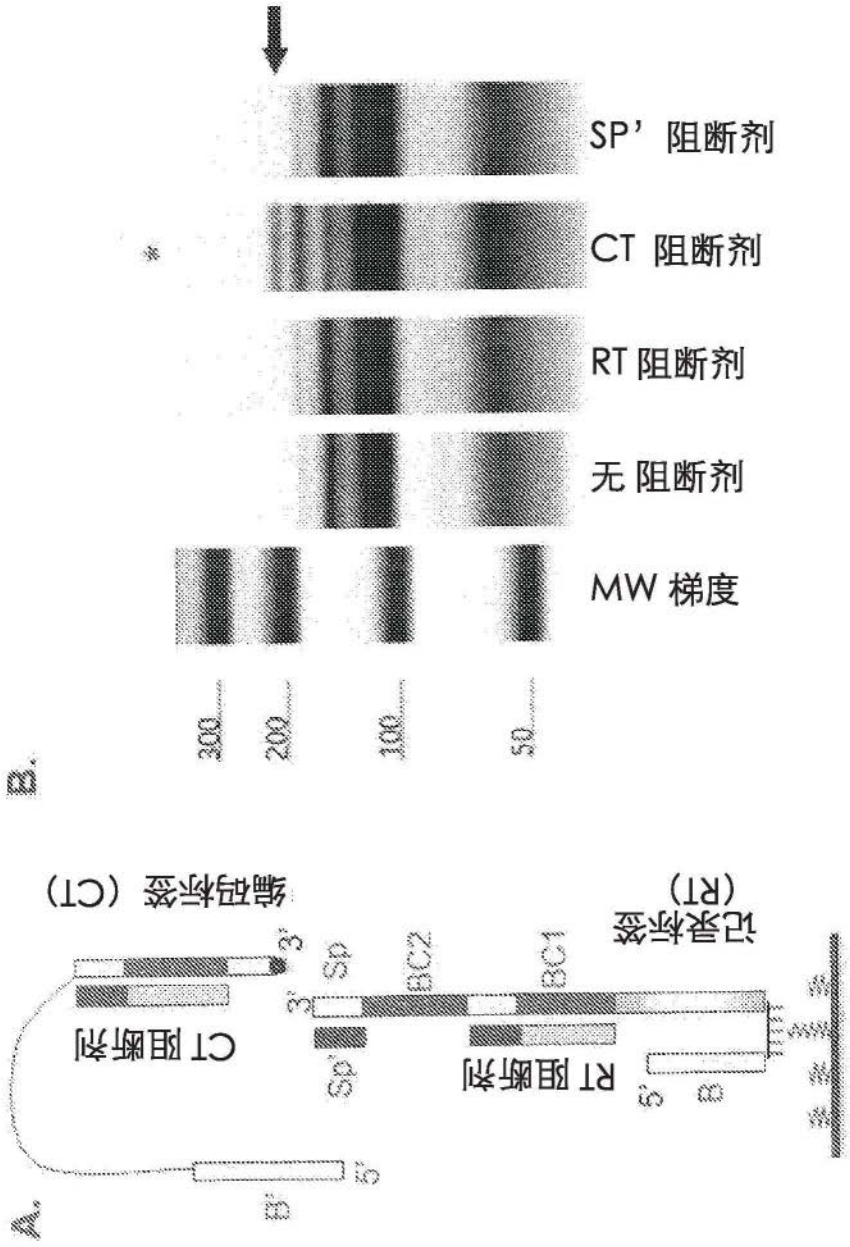

SP’ 阻断剂
CT 阻断剂
RT 阻断剂
无 阻断剂
MW 梯度
300
200
100
50
B.
编码标签 (CT)
记录标签 (RT)
3’
Sp
BC2
BC1
CT 阻断剂
RT 阻断剂
Sp’
A.

图60

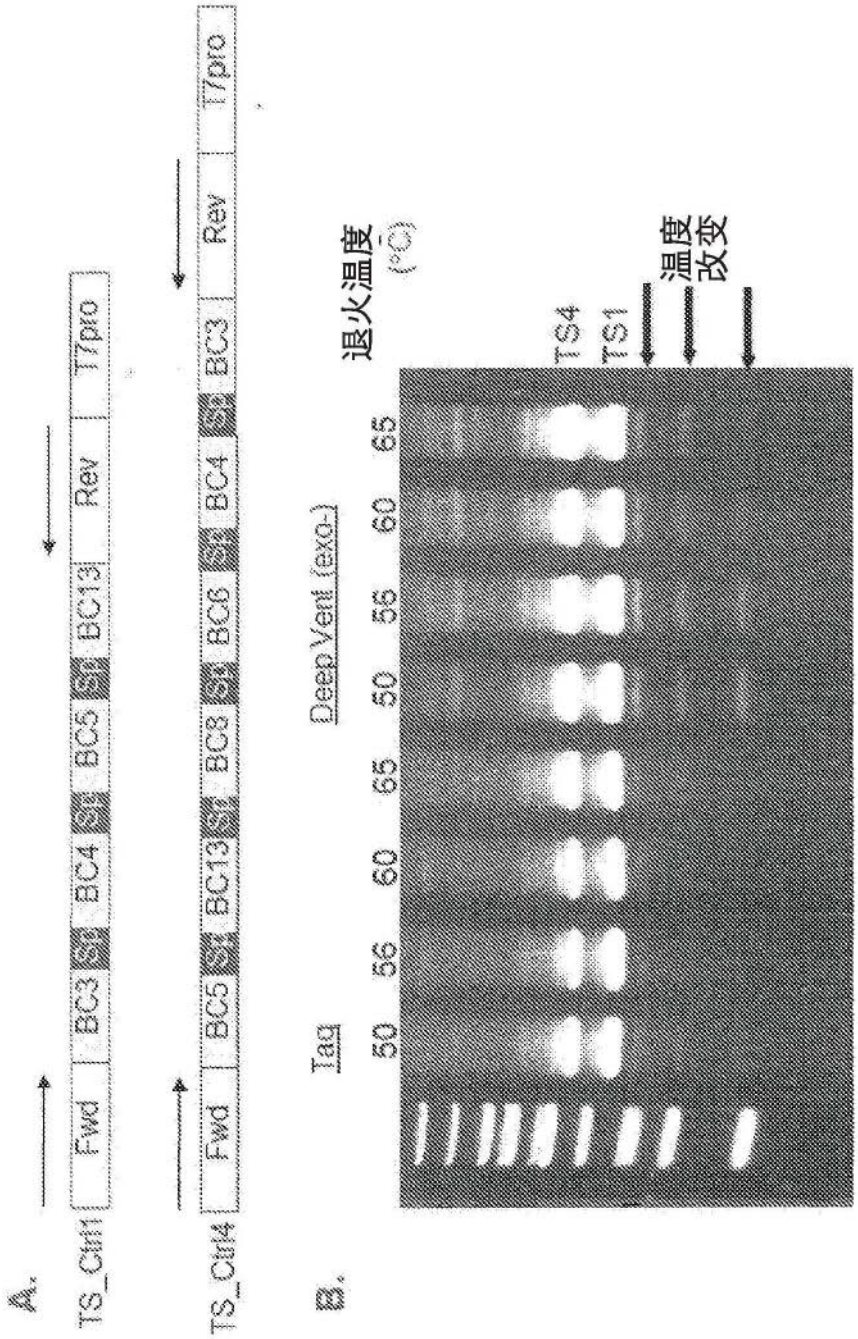

A.
TS_Ctrl1
Fwd
BC3
S1
BC4
S1
BC5
S1
BC13
Rev
T7pro
TS_Ctrl4
Fwd
BC5
S1
BC13
S1
BC8
S1
BC6
S1
BC4
S2
BC3
Rev
T7pro
B.
Taq
Deep Vent (exo-)
退火温度
(°C)
50
56
60
65
50
55
60
65
TS4
TS1
温度
改变

图61

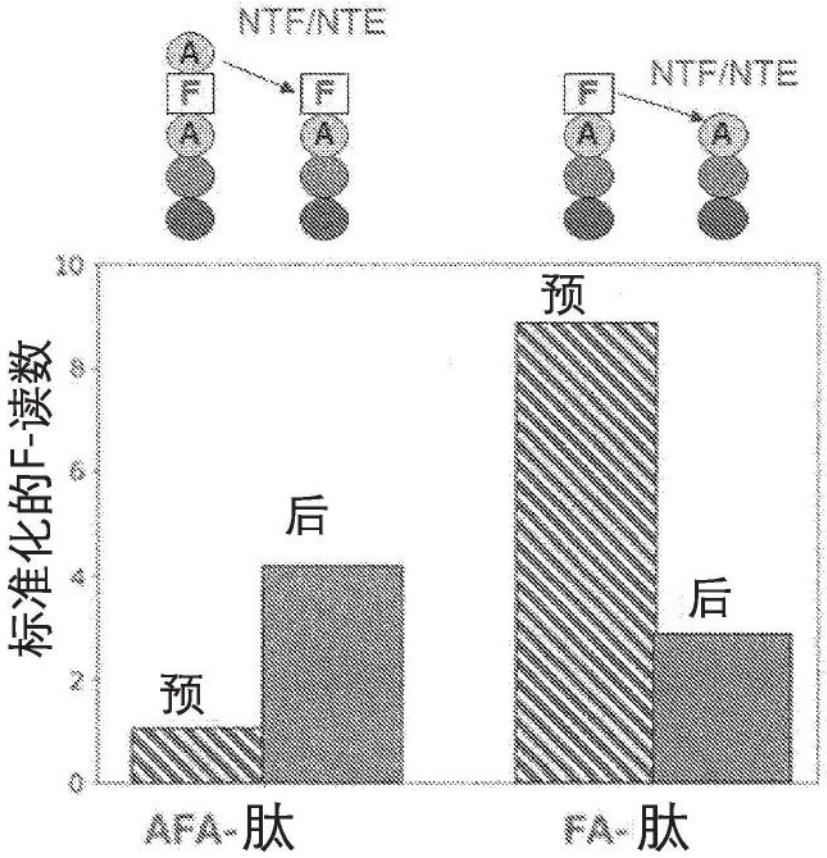



图62

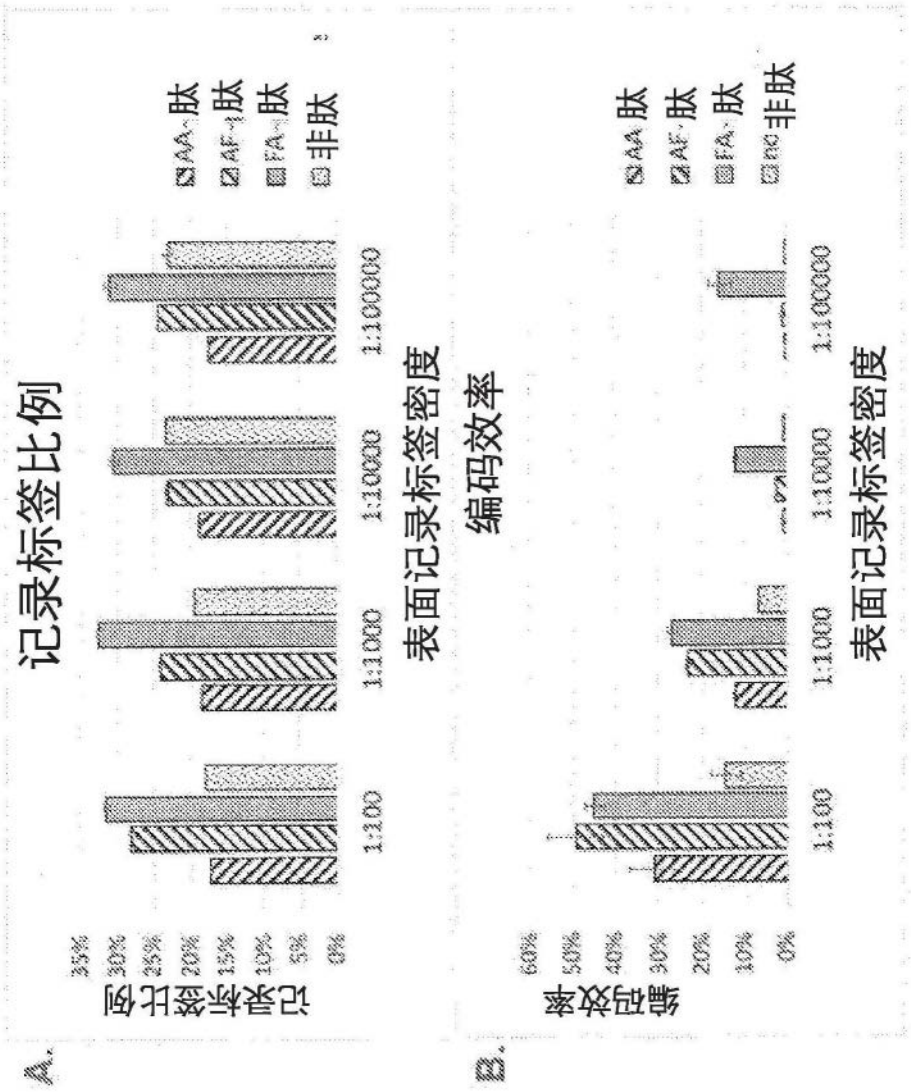

A.
记录标签比例
35%
30%
25%
20%
15%
10%
5%
0%
1:100
1:1000
1:10000
1:100000
表面记录标签密度
AA-肽
AF-肽
FA-肽
非肽
B.
编码效率
60%
50%
40%
30%
20%
10%
0%
1:100
1:1000
1:10000
1:100000
表面记录标签密度
AA-肽
AF-肽
FA-肽
非肽

图63

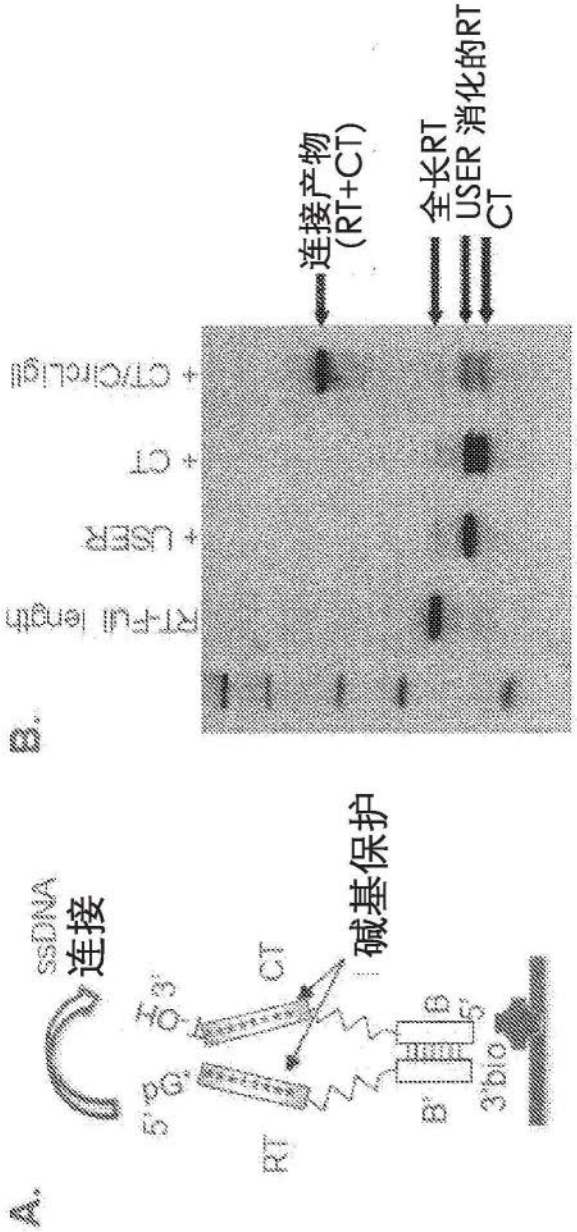

连接产物
(RT+CT)
全长RT
USER 消化的RT
CT
+ CT/CircLigII
+ CT
+ USER
RT-Full length
B.
碱基保护
ssDNA
连接
CT
HO-3'
5' pG'
RT
B
5'
B'
3'bio
A.

图64

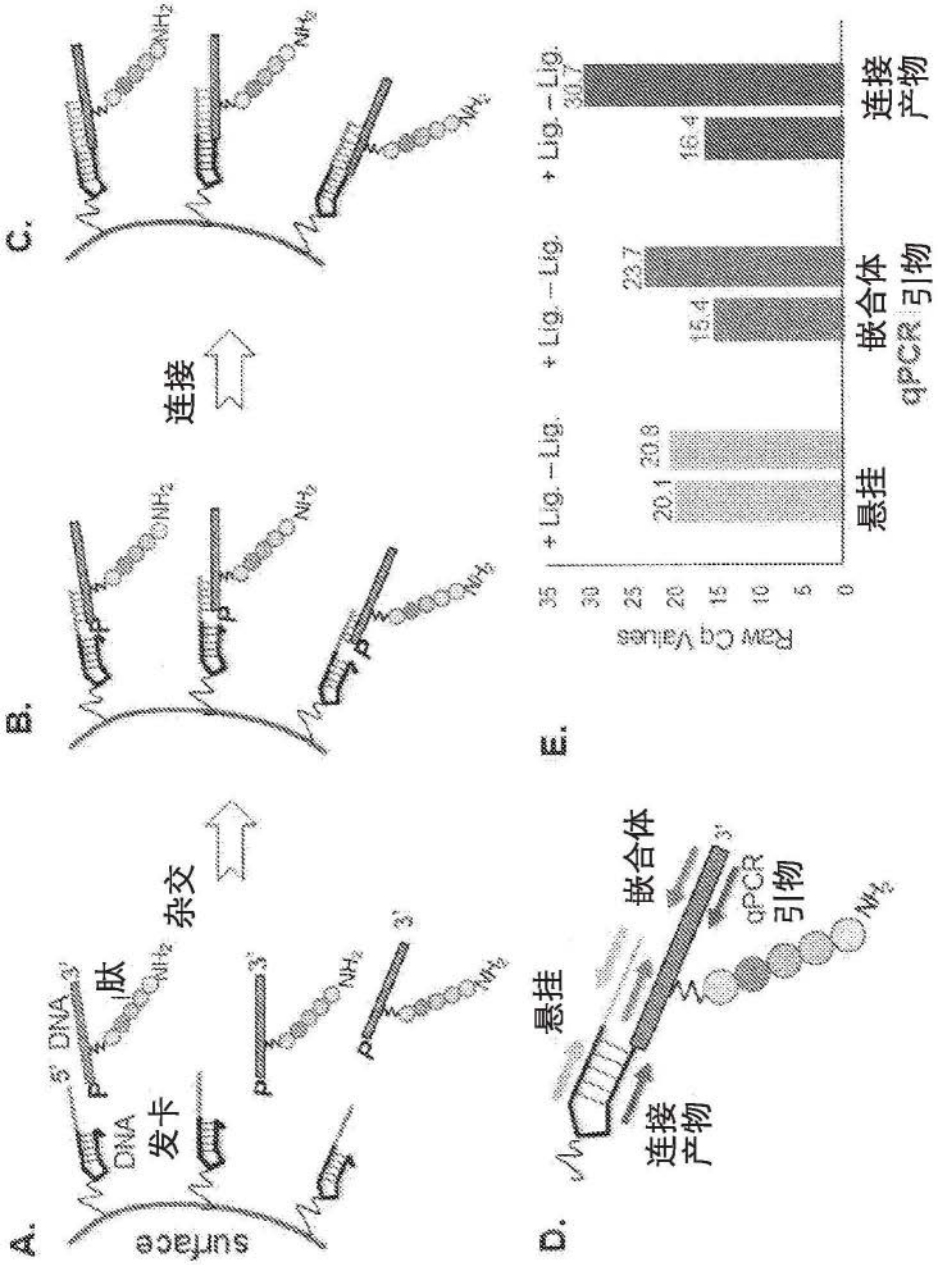

A.
surface
DNA
发卡
5' DNA 3'
肽
NH2
杂交
B.
连接
C.
D.
悬挂
嵌合体
连接
产物
qPCR
引物
E.
Raw Cq Values
35
30
25
20
15
10
5
0
+ Lig. – Lig.
+ Lig. – Lig.
+ Lig. – Lig.
20.1
20.8
23.7
15.4
30.7
16.4
悬挂
嵌合体
qPCR 引物
连接
产物

图65