



(12) 发明专利

(10) 授权公告号 CN 116377044 B

(45) 授权公告日 2025.05.06

(21) 申请号 202310379925.0

(22) 申请日 2017.11.29

(65) 同一申请的已公布的文献号

申请公布号 CN 116377044 A

(43) 申请公布日 2023.07.04

(30) 优先权数据

1620450.5 2016.12.01 GB

62/471,338 2017.03.14 US

(62) 分案原申请数据

201780083105.X 2017.11.29

(73) 专利权人 牛津纳米孔科技公开有限公司

地址 英国牛津

(72) 发明人 詹姆斯·克拉克 詹姆斯·怀特

理查德·马斯喀特

杰西卡·玛丽·梅·克诺特

拉米兹·伊克巴尔·纳撒尼

安德鲁·约翰·赫伦

马克·约翰·布鲁斯

拉科马·尼尚萨·贾亚辛格

多梅尼科·卡普罗蒂

戴维·杰克逊·斯托达特

瑞贝卡·维多利亚·鲍恩

克里斯多夫·詹姆斯·莱特

保罗·理查德·穆迪

(74) 专利代理机构 北京市柳沈律师事务所

11105

专利代理人 陈钘

(51) Int.CI.

C12Q 1/6869 (2018.01)

(56) 对比文件

WO 2015022544 A1, 2015.02.19

审查员 毛颖

权利要求书1页 说明书67页

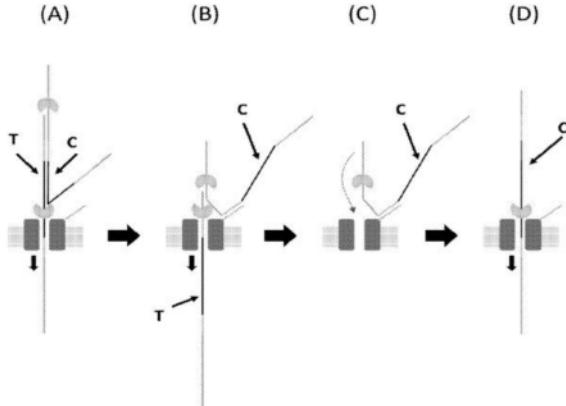
序列表(电子公布) 附图49页

(54) 发明名称

用于使两个非共价结合的互补核酸链依次
易位通过纳米孔的方法和复合物

(57) 摘要

使用纳米孔表征分析物的方法。一个方面的特征在于使用纳米孔表征双链多核苷酸，例如不使用连接所述双链多核苷酸的模板和互补序列的发夹的方法。另一方面特征在于使用标签修饰的纳米孔表征分析物的方法，其具有增加的灵敏度和/或更高的通量。还提供了可以用于所述方法中的组合物和系统，其包括例如用于附接于双链多核苷酸的衔接子和标签修饰的纳米孔。



1. 一种用于使两个非共价结合的互补核酸链依次易位通过纳米孔的方法，所述方法包括：

使一对非共价结合的互补核酸链在促进所述一对非共价结合的互补核酸链的第一成员易位通过所述纳米孔的条件下与所述纳米孔接触，其中衔接子附接于所述互补核酸链的两个末端中的一个或两个上，每个衔接子包含双链体茎和结合有多核苷酸解链酶的从所述双链体茎延伸的第一单链，其中，当所述衔接子附接于所述互补核酸链的两个末端时，一个衔接子的第一单链与所述第一成员邻接且另一个衔接子的第一单链与所述一对的第二成员邻接，

其中在所述第一成员易位通过所述纳米孔期间，所述第一成员拉开双链体茎以暴露所述第二成员上的结合位点，

并且其中所述结合位点与缀合至所述纳米孔内腔外部的外缘处的寡核苷酸标签可逆性结合，并且其中所述第二成员上的所述结合位点是具有与所述标签互补的序列的核酸的一部分。

2. 根据权利要求1所述的方法，其中所述一对互补核酸链包含附接于衔接子的靶核酸，并且其中所述结合位点存在于所述衔接子上。

3. 一种用于权利要求1所述的方法的复合物，其包含：

(i) 具有标签的纳米孔，所述标签是缀合至所述纳米孔内腔外部的外缘处的寡核苷酸标签；

(ii) 通过所述标签与所述纳米孔结合的互补多核苷酸链，其中所述互补多核苷酸链上的结合位点是具有与所述标签互补的序列的核酸的一部分；和

(iii) 与互补多核苷酸链部分杂交的模板多核苷酸链，其中所述模板多核苷酸链易位通过所述纳米孔的内腔，其中衔接子附接于包含所述互补多核苷酸链和所述模板多核苷酸链的双链多核苷酸的两个末端中的一个或两个上，每个衔接子包含双链体茎和结合有多核苷酸解链酶的从所述双链体茎延伸的第一单链，其中，当所述衔接子附接于所述互补核酸链的两个末端时，一个衔接子的第一单链与所述模板多核苷酸链邻接且另一个衔接子的第一单链与所述互补多核苷酸链邻接。

用于使两个非共价结合的互补核酸链依次易位通过纳米孔的方法和复合物

[0001] 本申请是申请日为2017年11月29日,申请号为201780083105.X,发明名称为“使用纳米孔表征分析物的方法和系统”的分案申请。

技术领域

[0002] 本文所提供的涉及使用纳米孔表征分析物的方法。还提供了可以用于所述方法中的组合物和系统,包括例如用于附接于分析物,如双链多核苷酸的衔接子,和标签修饰的纳米孔。在一些实施方案中,本文提供了使用跨膜孔为一种或多种靶多核苷酸测序的方法。

背景技术

[0003] 当前在广泛的应用范围内需要快速且便宜的多核苷酸(例如,DNA或RNA)测序和鉴定技术。

[0004] 跨膜孔(例如,纳米孔)已用于鉴定小分子或折叠的蛋白质并用于监测单分子水平的化学或酶促反应。跨膜孔(例如,纳米孔)作为聚合物和各种小分子的直接、电生物传感器具有巨大的潜力。具体来说,最近正在关注纳米孔作为潜在的DNA测序技术和生物标志物识别。

[0005] 可以在跨纳米孔施加的电位差下测量通过纳米孔的离子流。分析物与纳米孔的相互作用可以引起离子流的特征变化,并且对所产生的信号的测量可以用于表征分析物。例如,测得的信号可以是电流的,并且可以例如用于确定多核苷酸的序列。可以使多核苷酸链易位通过该孔,并且可以从测得的信号得到核苷酸的同一性,如序列。此类测序方法公开于例如W00142782、W02016034591、W02013041878、W02014064443和W02013153359中。

[0006] 已经开发了用于为双链多核苷酸测序的方法,例如涉及通过发夹连接的模板和互补链的易位。链测序通常涉及使用多核苷酸结合蛋白(如解旋酶)来控制多核苷酸通过纳米孔的运动。此类方法公开于例如W02013057495中。纳米孔的尺寸可以使其仅允许单链多核苷酸易位。双链多核苷酸可以通过在易位通过纳米孔之前将链分离以提供单链多核苷酸来确定。多核苷酸结合蛋白如解旋酶可用于同时分离双链多核苷酸并控制所得单链易位通过纳米孔的速率。双链多核苷酸的两条链可以通过桥接部分如发夹环连接,并且制备此类构建体的方法描述于例如W02013057495中。这确保在正向(模板)链易位之后反向(互补)链易位。以这种方式测量两条链是有利的,因为来自两条相连互补链的信息可以组合并用于提供比仅测量模板链可以获得的置信度更高的观察结果。然而,此类发夹连接的多核苷酸的制备会增加样品制备时间并导致有价值的分析物的损失。此外,发夹连接的模板和互补多核苷酸链易位通过纳米孔可以引起纳米孔的另一(反)侧上的链的再杂交。这可以改变易位速率,从而降低测序精确度。此外,由于模板链和互补链的电流-时间数据的差异,使用两种算法进行计算,这使得计算更加复杂和密集。

[0007] 因此,需要改进的表征分析物(例如双链多核苷酸)的方法,其具有提高的精确度和更高的效率/通量。

发明内容

[0008] 本公开总体上涉及使用纳米孔表征分析物的方法和可用于本文所述方法中的组合物，例如衔接子和纳米孔。本公开内容部分基于意外发现，双链多核苷酸的两条链均可以依次易位通过纳米孔以提供序列信息，而不需要通过桥接部分（如发夹环）共价连接两条链。例如，在一些实施方案中，可以向双链多核苷酸的每个末端提供具有双链体茎的衔接子，所述衔接子茎包含与缀合至纳米孔的孔标签互补的捕获序列，其中捕获序列仅在该链解链时显露。因此，当双链多核苷酸的第一链穿过标签修饰的纳米孔时，它拉开衔接子的双链体茎以暴露双链多核苷酸的第二链上的捕获序列，然后被纳米孔的孔标签捕获。此类方法不但使第二链保持靠近纳米孔，而且缩短了读取第一链和第二链之间的时间延迟，从而提高了测序方法的总体精确度和效率。还发现在纳米孔处捕获随后易位通过纳米孔的多种分析物可以增强表征分析物的灵敏度和/或通量。

[0009] 本发明人还发现，当使用多核苷酸结合蛋白分离双链多核苷酸的两条链，同时控制其中一条链通过跨膜孔的运动时，第二链可以留在孔附近并且在第一链易位通过孔之后，第二链可以被孔捕获，并且多核苷酸结合蛋白可以用于控制第二链通过孔的运动。

[0010] 因此，本发明的一个方面提供了为靶多核苷酸测序的方法，其包括：

[0011] (a) 使跨膜孔接触；

[0012] (i) 包含靶多核苷酸和与靶多核苷酸互补的多核苷酸的双链多核苷酸，其中靶多核苷酸和与靶多核苷酸互补的多核苷酸各自包含单链前导序列；和

[0013] (ii) 能够分离双链多核苷酸的链并控制多核苷酸通过跨膜孔的运动的多核苷酸结合蛋白；

[0014] (b) 检测对应于通过所述孔的离子流的信号，以检测易位通过所述孔的多核苷酸；

[0015] (c) 鉴定对应于所述靶多核苷酸的易位的信号和对应于与所述靶多核苷酸互补的多核苷酸的单独易位的连续信号；

[0016] (d) 分析(c)中鉴定的所述信号，

[0017] 从而为所述靶多核苷酸测序。

[0018] 在这方面，双链条形码序列可以附接于靶双链多核苷酸的一端或两端，前导序列可以包含在衔接子中，衔接子可以包含双链区和至少一个单链区，所述衔接子可以包含双链条形码序列，所述衔接子可以包含膜系链或孔系链，附接于靶双链多核苷酸的两个末端的前导序列可以不同，所述双链多核苷酸可以在其每个末端具有不同衔接子和/或多核苷酸结合蛋白，如解旋酶或聚合酶，可以与前导序列结合。当多核苷酸结合蛋白与前导序列结合时，多核苷酸结合蛋白的活性可以停滞，直到多核苷酸与跨膜孔接触。当双链条形码序列附接于靶双链多核苷酸的一端或两端时，可以将唯一的条形码序列附接于样品中的每个双链多核苷酸。在这方面，双链多核苷酸可以附接于微粒和/或可以修饰孔以增强多核苷酸的捕获。例如，吸引或结合多核苷酸或衔接子的一个或多个分子可以与孔连接。此类分子可选自，例如，PNA标签、PEG接头、短寡核苷酸、带正电荷的氨基酸和适体。在这方面，跨膜孔可以是，例如，蛋白质孔，如源自或基于Msp、 α -溶血素 (α -HL)、胞溶素、CsgG、ClyA、Sp1或FraC的孔，或固态孔和/或膜可以是两亲性层或固态层。

[0019] 该方法优于为双链多核苷酸测序的已知方法，其中两条链使用桥接部分如发夹环连接。该方法也优于仅测量模板多核苷酸链的已知方法。具体而言，本发明的方法结合了仅

模板链的方法和发夹环方法的优点,而没有所提到的发夹环方法的缺点。

[0020] 例如,W02013/014451中公开的方法使用多个衔接子,并且样品中仅一些双链多核苷酸将在一端添加Y衔接子,在另一端添加包含桥接部分的衔接子,丢弃样品中的其他多核苷酸。本发明的方法可以使用可以添加到双链多核苷酸的两端的单个前导序列或衔接子来进行。当使用此类单个前导序列/衔接子的系统时,需要丢弃较少的样品(如果有的话)。

[0021] 在该方法中,双链靶多核苷酸的任一端均可被孔捕获。与W02013/014451中公开的方法相比,这提高了灵敏度,其中只有不包含桥接部分的双链多核苷酸的末端可被孔捕获。

[0022] 本发明还提供了衔接子群体,其包含双链条形码序列、单链前导序列和能够分离双链多核苷酸的链并控制多核苷酸通过跨膜孔的运动的多核苷酸结合蛋白,其中所述群体中的每个衔接子中的条形码序列是唯一的。

[0023] 本文提供的另一方面涉及表征多核苷酸的方法。多核苷酸可包含DNA或RNA。所述方法包括:

[0024] (i) 在溶液中组合:

[0025] (a) 包含具有模板链和互补链的双链多核苷酸的构建体,其中所述模板链和所述互补链不共价连接,与

[0026] (b) 纳米孔,其中与所述构建体的一部分结合的一个或多个标签与所述纳米孔缀合,

[0027] 其中所述构建体和所述纳米孔在所述构建体与所述纳米孔结合的条件下组合;

[0028] (ii) 提供条件以便容许所述构建体的模板链进入所述纳米孔,以便容许所述模板链的分离和所述模板链的至少一部分易位通过所述纳米孔;

[0029] (iii) 测量指示所述模板链易位通过所述纳米孔的性质的变化;并且

[0030] (iv) 基于在所述模板链易位通过所述纳米孔时所测量的性质变化来表征所述多核苷酸。

[0031] 在一些实施方案中,所述溶液是离子的并且所测量的性质是通过所述纳米孔的离子电流。

[0032] 在一些实施方案中,衔接子附接于所述双链多核苷酸的两个末端中的一个或两个上,每个衔接子包含双链体茎和从所述双链体茎延伸的第一单链,其中一个衔接子的第一单链与所述模板链邻接且另一个衔接子的第一单链与所述互补链邻接。

[0033] 在一些实施方案中,步骤(ii)包括:在所述膜上施加电位差,以便容许与所述构建体的所述模板链邻接的所述第一单链进入所述纳米孔,将所述纳米孔上的电位差保持足够时间段,以便容许所述模板链分离和所述模板链的至少一部分易位通过所述纳米孔。

[0034] 在一些实施方案中,多核苷酸解链酶与两个衔接子中的一个或每一个预先结合。在所述纳米孔的内腔中提供多核苷酸解链酶。

[0035] 在一些实施方案中,与寡核苷酸具有互补性的衔接子部分可以在与所述第一单链邻接的链上的双链体茎内。可以将所述电位差保持足够时间段以容许所述多核苷酸解链到一定程度,使得其第一单链与所述互补链邻接的所述衔接子部分可用于与标签杂交。

[0036] 在一些实施方案中,一个或多个标签与所述纳米孔的外缘缀合。在一些实施方案中,与衔接子的一部分结合的一个或多个标签与纳米孔缀合。在一些实施方案中,与所述构建体的一部分结合的所述一个或多个标签中的至少一个是与所述构建体部分具有序列互

互补性的核酸。在一些实施方案中,与所述构建体的一部分结合的所述一个或多个标签中的至少一个是与所述衔接子部分具有序列互补性的核酸。核酸可以是不带电荷的,包括例如但不限于PNA或吗啉基。

[0037] 在一些实施方案中,所述条件是纳米孔上的电位差。

[0038] 本领域已知的任何纳米孔均可用于本文所述的方法中。在一些实施方案中,所述纳米孔还起到使所述多核苷酸解链的作用。在一些实施方案中,纳米孔可以是马达蛋白纳米孔,例如phi29马达蛋白纳米孔。在一些实施方案中,纳米孔置于膜内。

[0039] 在一些实施方案中,步骤(iii)包括测量当所述模板链易位通过所述纳米孔时通过所述纳米孔的离子电流的变化。在一些实施方案中,步骤(iv)包括基于在所述模板链易位通过所述纳米孔时所测量的通过所述纳米孔的离子电流的变化来表征所述多核苷酸。

[0040] 在本文所述方法的一些实施方案中,多核苷酸解链酶存在于所述溶液中所述纳米孔的顺式-开口侧。在一些实施方案中,一种或多种多核苷酸解链酶与所述多核苷酸预先结合。

[0041] 在一些实施方案中,对于每个衔接子,多核苷酸解链酶与从所述双链体茎延伸的所述第一单链结合。在一些实施方案中,通过其相应结合的多核苷酸解链酶促进所述模板链的解链。

[0042] 在本文所述方法的一些实施方案中,所述多核苷酸的解链使所述互补链的一部分露出以与标签杂交。

[0043] 在本文所述的各个方面的一些实施方案中,所述方法还可包括保持所述条件足够时间段以在所述模板链易位通过所述纳米孔之后容许所述互补链进入所述纳米孔并且使所述纳米孔易位。例如,所述方法还可包括保持所述电位差足够时间段,以在所述模板链易位通过所述纳米孔之后容许与所述互补链邻接的所述第一单链进入所述纳米孔并且容许所述互补链易位通过所述纳米孔。

[0044] 在本文所述的各个方面的一些实施方案中,所述方法还可包括测量指示所述互补链易位通过纳米孔的性质的变化。所述性质可以是当所述互补链易位通过所述纳米孔时通过所述纳米孔的离子电流。在一些实施方案中,所述方法还可包括进一步基于所测量的指示所述互补链易位通过所述纳米孔的性质的变化来表征所述多核苷酸。在一些实施方案中,可获得指示所测量的指示所述互补链和所述模板链均易位通过所述纳米孔的性质的数据并用于表征所述多核苷酸。可以将所述模板链数据与所述互补链数据进行比较或组合以表征所述多核苷酸。

[0045] 在本文所述方法的一些实施方案中,所述纳米孔可包含第一标签和第二标签。所述第一标签和所述第二标签分别与所述衔接子的所述第一单链邻接所述模板链的一部分和与所述衔接子的所述第一单链邻接所述互补链的一部分结合。

[0046] 在本文所述方法的一些实施方案中,每个衔接子可包含从双链体茎延伸的第二单链。所述一个衔接子的第二个单链与互补链邻接和/或另一个衔接子的第二个单链与模板链邻接。在一些实施方案中,与所述衔接子的一部分结合的一个或多个标签中的至少一个可以是与所述衔接子在所述第二单链内的一部分具有序列互补性的寡核苷酸。在一些实施方案中,与所述衔接子的一部分结合的一个或多个标签中的两个或更多个是与所述衔接子在所述第二单链内的一部分具有序列互补性的寡核苷酸。

[0047] 在本文所述方法的一些实施方案中,所述方法还可包括:基于当所述模板链易位通过所述孔时所述测量性质的变化的测量来确定所述模板链的序列;基于当所述互补链易位通过所述孔时所述测量性质的变化的测量来确定所述互补链的序列;并且比较所述模板链的序列和所述互补链的序列以确立多核苷酸的序列。

[0048] 还提供了用于表征多核苷酸的系统,例如,其可用于本文所述方法的任何方面。所述系统包括:(i)包含具有模板链和互补链的多核苷酸的构建体,和(ii)置于膜中的纳米孔,所述纳米孔包含外缘,所述外缘与至少一种与衔接子的一部分具有序列互补性的核酸缀合。

[0049] 在一些实施方案中,模板链和互补链不共价连接。衔接子可以附接在多核苷酸的两个末端中的每一个上,每个衔接子包含双链体茎和从所述双链体茎延伸的第一单链。一个衔接子的第一单链可与所述模板链邻接且另一个衔接子的第一单链可与所述互补链邻接。对于每个衔接子,多核苷酸解链酶可以与从双链体茎延伸的第一单链结合。在一些实施方案中,所述衔接子的所述部分是在与所述第一单链邻接的链上的双链体茎内。

[0050] 在一些实施方案中,每个衔接子可包含从所述双链体茎延伸的第二单链,其中所述一个衔接子的所述第二单链与所述互补链邻接且所述另一个衔接子的所述第二单链与所述模板链邻接。在一些实施方案中,所述衔接子的所述部分可以在第二单链内。

[0051] 在一些实施方案中,与纳米孔缀合的至少一个核酸(a)与所述衔接子在与所述第一单链邻接的链上的双链体内的一部分具有序列互补性,并且(b)还与所述衔接子在所述第二单链内的一部分具有序列互补性。

[0052] 在一些实施方案中,至少两个核酸与所述纳米孔缀合,其中所述至少两个核酸中的一个与所述衔接子在与所述第一单链邻接的链上的双链体内的一部分具有序列互补性,并且其中所述至少两个核酸中的另一个与所述衔接子在所述第二单链内的一部分具有序列互补性。在一些实施方案中,至少两个核酸可与所述纳米孔的外缘缀合。

[0053] 另一方面涉及一种制备用于表征多核苷酸的系统的方法。所述方法包括:(i)获得包含具有模板链和互补链的多核苷酸的构建体,其中所述模板链和所述互补链不共价连接,并且(ii)在所述构建体暴露于所述纳米孔的外缘的条件下将所述构建体与置于膜中的纳米孔组合,其中至少一个与所述衔接子的一部分具有序列互补性的核酸与所述纳米孔的外缘缀合。

[0054] 在一些实施方案中,衔接子可以附接在所述多核苷酸的两个末端中的每一个上,每个衔接子包含双链体茎和从所述双链体茎延伸的第一单链,其中一个衔接子的第一单链与所述模板链邻接且另一个衔接子的第一单链与所述互补链邻接。对于每个衔接子,多核苷酸解链酶可以与从双链体茎延伸的第一单链结合。

[0055] 包含在本文所述方法的任何方面中形成的两种或更多种组分的复合物也在本公开的范围内。在一些实施方案中,复合物包含:(i)具有标签的纳米孔,(ii)经由标签与纳米孔结合的互补多核苷酸链,和(iii)与互补序列多核苷酸链部分杂交的模板多核苷酸链,其中所述模板多核苷酸链部分置于所述纳米孔的内腔中。在其他实施方案中,复合物可包含(i)具有两个或更多个标签的纳米孔,和(ii)包含模板链和互补链的双链多核苷酸,其中每条链与所述两个或更多个标签中的一个结合。在本文所述复合物的任何实施方案中,标签可位于其内腔外部的外缘处。

[0056] 本文还提供了用于确定分析物的特征的方法和系统。一方面，系统包括：(i) 置于膜中的纳米孔，所述纳米孔包括存在至少一个共同标签的外缘；和(ii) 多种不同的分析物，每种不同的分析物附接于所述至少一个标签的结合结合伴侣上。在一些实施方案中，每种分析物均可以是生物聚合物。此类生物聚合物的实例包括但不限于多核苷酸、多肽、多糖和脂质。

[0057] 在一些实施方案中，每种分析物均为多核苷酸。在每种分析物均为多核苷酸的一些实施方案中，所述结合伴侣可以是附接于所述多核苷酸的衔接子的核苷酸序列，并且至少一个共同标签可以是与所述衔接子的核苷酸序列具有序列互补性的核酸。

[0058] 另一方面，一种使用纳米孔确定分析物特征的方法包括：(i) 提供分析物；(ii) 使一种或多种分析物与纳米孔结合，纳米孔的外缘在纳米孔的内腔外，并且(iii) 在相对于所述纳米孔移动所述分析物的同时获得已经与所述纳米孔结合的所述分析物的测量值，其中所述测量值指示所述分析物的一个或多个特征；并且(iv) 基于步骤(iii)中获得的测量值表征分析物。

[0059] 在一些实施方案中，所述一种或多种分析物可以结合纳米孔内腔外部的纳米孔外缘。在一些实施方案中，一种以上的分析物可以分别与一个以上缀合至所述纳米孔的标签结合。在一些实施方案中，所述一个或多个标签可以与所述纳米孔内腔外部的所述纳米孔外缘缀合。

[0060] 在一些实施方案中，第一分析物可以是多核苷酸。

[0061] 在一些实施方案中，所述方法还可包括：在相对于所述纳米孔移动第二分析物的同时获得已经与所述纳米孔结合的所述第二分析物的测量值，其中获得的所述第二分析物的所述测量值指示所述第二分析物的一个或多个特征，并且基于所获得的第二分析物的测量值表征第二分析物。第二分析物可以是多核苷酸。

[0062] 在一些实施方案中，在所述第一分析物相对于所述纳米孔移动期间第二分析物可以与所述纳米孔结合。第二分析物可以是多核苷酸。

[0063] 本文还提供了一种复合物，其包含具有多个标签的纳米孔，其中部分位于所述纳米孔内腔中的第一分析物和第二分析物与捕获部分中的一个结合。

[0064] 在一些实施方案中，所述多个标签可以在其内腔外部的外缘上。在一些实施方案中，第一分析物和第二分析物是多核苷酸。

[0065] 用于使两个非共价结合的分子依次易位通过纳米孔的方法也在本公开的范围内。所述方法包括：使一对非共价结合的分子在促进该对非共价结合的分子的第一成员易位通过纳米孔的条件下与所述纳米孔接触，其中该对的第二成员上的结合位点在第一成员易位通过纳米孔期间暴露，并且其中结合位点与纳米孔上存在的标签可逆性结合。

[0066] 在一些实施方案中，非共价结合的分子是互补核酸链。在一些实施方案中，该对非共价结合的分子可包含附接于衔接子的靶核酸，并且结合位点可存在于衔接子上。

[0067] 在一些实施方案中，纳米孔上的标签可以是寡核苷酸，第二成员上的结合位点可以是具有与标签互补的序列的核酸的一部分。

[0068] 本文还提供了一种表征多核苷酸的方法，其包括：使一对非共价结合的分子在促进该对非共价结合的分子的第一成员易位通过纳米孔的条件下与所述纳米孔接触，接着依次使该对非共价结合的分子的第二成员易位；测量指示该对的第一和第二成员易位的性

质，并获得指示所测性质的数据；并且基于所获得的第一和第二成员的数据确定特征。

附图说明

[0069] 以下附图构成本说明书的一部分，并且将其包括在内以进一步说明本公开的某些方面，通过参考这些附图中的一个或多个并结合本文给出的具体实施方案的详细描述，可以更好地理解本公开的某些方面。

[0070] 仅出于说明目的，本文所述的图中的链根据其捕获末端标记为“模板”和“互补序列”。穿过纳米孔的第一链标记为模板，在第一链之后的互补链标记为互补序列。在分析从第一链和第二链获得的序列信息后，确定双链多核苷酸的实际模板和互补序列。

[0071] 图1A-1B说明了使用跨膜孔为双链多核苷酸(例如，DNA)构建体测序的现有技术方法，其中模板和互补链通过发夹环连接并且模板链包含5'前导序列。图1A是多核苷酸(例如，DNA)构建体在酶的控制下易位通过纳米孔的示意图。模板进入纳米孔并且相同的酶在发夹周围前进以控制模板后面的互补序列的运动。一旦发夹区易位通过纳米孔，发夹就可在纳米孔的反侧上重新形成。图1B显示了表示从模板易位，从互补序列易位以及将从模板易位和互补序列易位获得的序列信息在算法上组合时获得的序列信息的准确度的峰。

[0072] 图2A-2B说明了根据本文所述的一个实施方案在不使用发夹的情况下对双链多核苷酸(例如，DNA)构建体进行“跟随(follow-on)”测序的方法。模板和互补多核苷酸(例如，DNA)链在每个末端包含衔接子，该衔接子包含前导序列。图2A是双链多核苷酸(例如，DNA)构建体在酶控制下易位通过纳米孔的示意图。双链多核苷酸的模板和互补序列不共价连接，并且每条链具有负载在衔接子上的酶。在模板链穿过纳米孔(并且酶解离)后，通过孔分别捕获互补链并测序。在不存在将模板连接到互补序列的发夹的情况下，在纳米孔的反侧上几乎没有形成二级发夹结构。图2B显示了表示从模板易位，从互补序列易位以及将从模板易位和互补序列易位获得的序列信息在算法上组合时获得的序列信息的准确度的峰。

[0073] 图3A-3B说明了根据本文所述的一个实施方案的载酶衔接子的结构。图3A是载酶衔接子的示意图。标记代表以下内容：(1)间隔区(例如，前导序列)；(2)多核苷酸结合蛋白(例如，多核苷酸解链酶)，如解旋酶(例如，Dda解旋酶)；(3)间隔区；(4)锚，如胆固醇锚。其他实线表示多核苷酸序列。图3B显示了附接于双链多核苷酸(如基因组DNA片段)的每个末端的衔接子，有多核苷酸结合蛋白(例如，多核苷酸解链酶)负载在每个衔接子上。

[0074] 图4是在多核苷酸链易位通过纳米孔期间随时间测量的电流信号的示意图。

[0075] 图5是双链多核苷酸在纳米孔界面处分离和随后单链多核苷酸易位通过纳米孔的图示。

[0076] 图6是在为多核苷酸测序期间图4的电流时间信号的一部分的事件检测的图示。

[0077] 图7是使用递归神经网络(RNN)模型分析信号测量的示意图。

[0078] 图8是如何采用维特比算法(Viterbi algorithm)来确定通过具有最高似然性的可能转变的路径的图示。

[0079] 图9A-9B说明了根据本文所述的一个实施方案的载酶衔接子的结构。图9A是载酶衔接子的示意图。标记代表以下内容：(1)间隔区(例如，前导序列)；(2)多核苷酸结合蛋白(例如，多核苷酸解链酶)，如解旋酶(例如，Dda解旋酶)；(3)间隔区；(4)锚，如胆固醇锚，其是任选的；和(5)置于(1)间隔区(例如，前导序列)的相对端的双链体茎，所述双链体茎包含

与(1)间隔区(例如,前导序列)对准的链上的捕获序列,其中捕获序列与缀合至纳米孔外缘的标签(例如,捕获多核苷酸)互补。其他实线表示多核苷酸序列。图9B显示当衔接子附接于双链多核苷酸的每个末端时的构建体,有多核苷酸结合蛋白(例如,多核苷酸解链酶)负载在每个衔接子上。

[0080] 图10(图A-D)说明了根据本文所述的一个实施方案,使用纳米孔为双链多核苷酸测序的方法的示意图。该方法包括提供(i)每个末端附接于衔接子(例如,如图9A中所示,没有锚(4)并且在衔接子上负载有多核苷酸结合蛋白(例如,多核苷酸解链酶)的双链多核苷酸,和(ii)具有与纳米孔的外缘缀合的捕获多核苷酸的纳米孔。通过使附接于第二链的衔接子的捕获序列与缀合至纳米孔的外缘的标签(例如,捕获多核苷酸)结合,使双链多核苷酸的第二链(互补序列)偶联至纳米孔。

[0081] 图11A-11B显示了使用根据本文所述的一个实施方案的方法获取的链数据的示例部分。链数据显示单个通道的电数据的电流(pA)与时间(秒)。图11A显示了链数据的示例部分,其显示没有阻断电流的链的开孔水平为约200pA。当链被捕获时,电流降低至50-100pA范围,这取决于序列组成。当链完全穿过孔时,电流恢复到200pA的开孔水平。对于成对的模板链和互补链而言,将单独的链标记为T_n(例如,T₁、T₂、……)和C_n(例如,C₁、C₂、……),并且对于其中T后面没有其互补对的链而言,标记为T。标记和数据与表4中的数据相对应。图11A说明一对中的第二互补链通常紧跟在模板之后,在链间具有非常短的时间。图11B显示了电迹线的放大部分,突出显示了跟随的多对之一,标记为模板₁和互补序列₁。

[0082] 图12A-12D显示了使用根据本文所述的一个实施方案的方法获取的链数据的示例部分。链数据显示MinION芯片上单个通道的电数据的电流(pA)与时间(秒)。图12A显示了跟随模板-互补序列对的另一示例电迹线。图12B、图12C和图12D显示了图12A中的迹线的放大图,其中星形标记双链体茎中的sp18间隔区(例如,互补标签部分),其被添加到链中以使得与孔-标签(例如,与纳米孔的外缘缀合的捕获多核苷酸)的偶联成为可能。图12B显示了模板链起始处的sp18。图12C显示了模板末端和互补序列起始处的sp18,并且图12D显示了互补序列末端的sp18。这些标志可用于证明dsDNA底物具有附接于dsDNA两端的酶-衔接子,并测量附接效率。

[0083] 图13显示了当连续链穿过孔(从MinION芯片上的所有通道聚集)时,随后的链(x轴)之间的开孔时间分布的直方图。上图,对照衔接子(如实施例2中所述)显示,在可以与孔标签(例如,捕获核苷酸)偶联的多核苷酸构建体中没有捕获序列的情况下进行测序时,链间的时间分布平均大约为3秒。下图,跟随衔接子2(如实施例4中所述)显示,当链在互补序列中含有可以与孔标签偶联的捕获序列(例如,捕获多核苷酸)时,观察到大约50毫秒的新群体。50ms的短群体来自其模板对后不久快速捕获的互补链。捕获很快,因为互补链通过互补序列与孔标签的结合保持与孔非常靠近,因此不允许其扩散。

[0084] 图14显示了从模板易位,从互补序列易位以及将从模板易位和互补序列易位获得的序列信息在算法上组合时获得的序列信息(随机片段化大肠杆菌)的碱基识别(basecall)的准确度分布的直方图。使用如图10所示的方法获得序列信息。

[0085] 图15显示了具有两种或更多种类型标签的纳米孔的示例性实施方案。例如,可以提供一种标签以增加表征分析物的方法的灵敏度(“灵敏度标签”),同时可以提供另一种标签以增加对双链多核苷酸的模板链之后的互补链测序的可能性(“跟随标签”)。孔标签可以

以多种方式构造。例如，寡聚孔的每个单体可以具有相同类型的标签构型(例如，具有多个结合位点，如Tag-A和Tag-B所示)。Tag-A和Tag-B可以组合形成单个标签，每个单体包含Tag-A/Tag-B组合标签。可替代地，寡聚孔可包含有不同标签附接的混合单体，使得至少一种单体具有与其他单体不同的标签构型。在另一个实例中，Tag-A和Tag-B可以保持为单独的标签，并且每个单体可以包括两种单独的标签。如果灵敏度标签和跟随标签与衔接子中使用的独特序列互补，则可以单独组合，如衔接子原理图设计的下图所示。

[0086] 图16是可如何使用具有两种不同标签类型的纳米孔从溶液中捕获链(以便提高灵敏度)的示意图。附接于双链多核苷酸末端的衔接子包含可用于偶联至第一孔标签的捕获序列(例如，形成Y-衔接子的非互补臂)，同时双链体茎内只有在被拉开时才显露的单独捕获序列允许互补序列结合第二孔标签，因此使互补序列捕获成为可能以进行跟随测序。

[0087] 图17A是可以如何在衔接子的两个位置使用相同捕获序列的示意图，一个显露允许链从溶液中结合纳米孔的孔标签以提高灵敏度，另一个最初未显露并且是在模板拉开通过孔隙时显露，变得可以与孔(仅具有一种类型的标签的孔)上的多个标签中的另一个结合，以实现跟随测序。图17B提供了可用于此目的一些示例序列。顶部构建体显示了示例Y衔接子的一部分。“F0001/F0002”和“F0003/F0004”序列是双链体茎的实例，其可以连接到示例Y衔接子上产生单个衔接子构建体，其可以实现根据本文所述的一个或多个实施方案的方法。纳米孔的相同结合序列位点可以在衔接子的双链体茎内使用不止一次(例如，两次)。

[0088] 图18显示了衔接子设计的示意图，该设计可实现跟随测序和增加灵敏度。孔结合序列(在图18中标记为“吗啉基孔标签的Hyb夹板”)暴露于周围溶液并且最初可用于结合孔标签，因此其提高了灵敏度。当附接时，孔结合位点也与互补链邻接，以致使当模板链穿过孔时，互补链保持与孔结合。该过程在图19中示意性地示出。

[0089] 图19是显示双链多核苷酸的示意图，其中图18的衔接子附接于每个末端。链通过暴露的孔结合位点从溶液中偶联到纳米孔，从而提高随后捕获附近模板链的灵敏度。结合位点也与互补链邻接，以致使当模板已经穿过纳米孔时，互补序列保持与纳米孔结合。互补序列可以在最终捕获和测序之进入许多可能的构象，如下所示，以实现跟随测序。在图19中，分别附接于待检测链的两端的互补区段可以结合在一起形成包含所述链的发夹结构，通过使链更靠近纳米孔来促进测序过程以增加测序效率。当待检测链是长链时，这可能特别有益。

[0090] 图20显示了衔接子设计的示意图，其中双链体茎内的相同序列在衔接子的不同位置重复，如图所示，以实现图21中所示的跟随方法。

[0091] 图21是显示双链多核苷酸的示意图，其中图20的衔接子附接于每个末端。如图20所述，dsDNA链通过侧臂上附接于衔接子的结合位点与孔结合。当模板被捕获在孔中时，侧臂序列被拉开并保持与孔标签结合，如图所示。后来进入模板拉开，具有相同序列的第二位点显露以与侧臂结合(本身仍然与孔标签结合)。以这种方式，孔上的单个标签可以用于提高捕获灵敏度，并且可以重复用于稍后实现底物的互补序列的跟随。最后，孔标签保留了侧臂序列，但是侧臂本身被孔捕获并从孔标签中剥离以释放孔标签用于另一个循环。

[0092] 图22显示当模板-酶接近模板链的末端时，显露的序列如何暴露以与孔标签偶联。跟随过程的效率可以增加，例如，通过包括间隔区(例如，所示序列中的4个sp18间隔区，例如六乙二醇)或使酶短暂停留的类似特征(这允许更多时间进行偶联)，或具有优化几何特

性或灵活性的特征。显露部分中的双重结合位点也提高了与孔标签偶联的机会。

[0093] 图23提供了可以实现图21中公开的方法的示例衔接子/序列。

[0094] 图24提供了可以实现图21中公开的方法，并且在使酶暂停方面更加优化的其他衔接子/序列。

[0095] 图25提供了可以实现图19中公开的方法的示例衔接子/序列。

[0096] 图26提供了构成上图中描述的衔接子的组分的示例序列。

[0097] 图27A显示SYPRO Ruby蛋白凝胶，其显示出经或未经吗啉基孔标签修饰的CsgG的单体和寡聚纳米孔。图27B显示经吡啶基-二硫代吗啉基修饰的纳米孔的示意图。

[0098] 图28显示Cy3荧光凝胶，其显示分析物与吡啶基-二硫代吗啉基修饰的孔的杂交。

[0099] 图29显示SYBR金核酸凝胶染剂，其显示分析物与吡啶基-二硫代吗啉基修饰的孔的杂交。

[0100] 图30显示SYPRO Ruby蛋白凝胶，其显示分析物与吡啶基-二硫代吗啉基修饰的孔的杂交。

[0101] 图31说明了显示纳米孔（例如，CsgG纳米孔）的计算机渲染的图，突出显示了可以添加半胱氨酸以与孔标签缀合的位置。当纳米孔置于膜中时，孔标签可以缀合至纳米孔的外表面，例如，在膜的顺侧或反侧。

[0102] 图32A显示了Y衔接子设计的一个实施方案，其包括两个杂交位点，一个用于孔系链，另一个用于膜或珠粒系链。在该设计中，孔系链紧邻前导序列。图32B显示连接的分析物，例如在一端具有Y衔接子的双链多核苷酸。

[0103] 图33是显示图32A中说明的Y衔接子设计的示例序列的示意图。

[0104] 图34A显示了Y衔接子设计的不同实施方案，其包括两个杂交位点，一个用于孔系链，另一个用于膜或珠粒系链。在该设计中，膜系链紧邻前导序列。图34B显示连接的分析物，例如在一端具有Y衔接子的双链多核苷酸。

[0105] 图35是显示图34A中说明的Y衔接子设计的示例序列的示意图。

[0106] 图36是显示Y衔接子设计的一个替代实施方案的示意图，其包括两个杂交位点，一个用于珠粒系链，另一个用于膜系链。在该设计中，珠粒具有两条不同的系链，一条连接到分析物，另一条连接到孔。

[0107] 图37是显示图36中说明的Y型衔接子设计的示例序列并显示分析物与孔的间接附接的示意图。

[0108] 图38显示了易位通过纳米孔的连续链的示例迹线，所述纳米孔没有可以与链结合以允许跟随测序的孔标签。在这些实例中，链间隔时间范围为2-5秒。

[0109] 图39显示了根据本文所述的一个实施方案，使经修饰的孔易位的连续链的示例迹线。在这些实例中，链间隔时间范围为0.02-3秒。

[0110] 图40显示了以对数标度说明连续链之间的间隔时间的直方图。左图显示具有单一分布的纳米孔（例如，CsgG孔），其中链间隔时间大于1秒。右图显示了易位通过拴系孔的链间隔时间。这显示了两个群体，即快速捕获群体，并且链间隔时间低于0.1秒。

[0111] 图41描绘了显示从输入20ng DNA起的6小时内每个芯片测序的碱基数量的图表。图42显示来自大肠杆菌运行的数据表，其显示具有拴系孔的跟随链的数量增加。图43显示了可用于连接单链和双链核酸的方法。

[0112] 图44显示了表征和串接许多双链靶多核苷酸的方法,其中第一双链靶多核苷酸的互补链募集许多其他双链靶多核苷酸并使其达到孔的局部浓度。这样在孔周围提供比在一般本体溶液中更高的局部浓度,因此双链靶多核苷酸彼此跟随通过开孔,链间隔时间最短。当双链靶多核苷酸的浓度较低时,这尤其有用。使用由与单链结合蛋白偶联的寡核苷酸组成的系链。当对第一双链靶多核苷酸的模板链进行测序时,互补链作为ssDNA释放到溶液中。其他双链靶多核苷酸的单链结合蛋白能够与ssDNA结合。当对互补链进行测序时,互补链的3'被拉回到孔中。ssDNA互补链上的单链结合蛋白在遇到马达蛋白时从互补链中移位,因此沉积在孔周围,增加局部浓度。

具体实施方式

[0113] 虽然跨膜孔(例如,蛋白质纳米孔或固态纳米孔)可用作检测或表征生物聚合物的传感器,但仍存在使用跨膜孔提高检测方法的准确度和/或效率的挑战。例如,通过发夹连接的双链多核苷酸的模板链和互补链易位通过纳米孔存在各种缺点。虽然以这种方式测量两条链是有利的,因为来自两条相连互补链的信息可以组合并用于提供比仅测量模板链可以获得的更高的准确度,这种发夹连接的多核苷酸的制备更复杂且更耗时并且可导致有价值的分析物的损失。此外,发夹连接的模板和互补多核苷酸链易位通过纳米孔可以引起纳米孔的另一(反)侧上的链的再杂交。这可以改变易位速率,从而降低测序精确度。具有发夹结构的链也更难以像线性单链那样快速易位。另外,由于模板链和互补链的电流-时间数据的差异,使用两种算法进行计算,这使得计算更加复杂和密集。

[0114] 对于分析物检测,通常在一个分析物的易位和下一个分析物的易位之间存在时间延迟。这种延迟可以是几秒到几分钟的量级,这可以导致表征更慢,孔开放电流更高(更快地耗尽参比电极),和/或当孔打开时纳米孔被阻塞的可能性增加。因此,需要开发提高使用纳米孔表征分析物的准确度和/或效率或通量的方法和组合物。

[0115] 本公开内容部分基于意外发现,双链多核苷酸的两条链均可以依次易位通过纳米孔以提供序列信息,而不需要通过桥接部分(如发夹环)共价连接两条链。例如,一方面,本发明人发现,当使用多核苷酸结合蛋白(例如,多核苷酸解链酶)分离双链多核苷酸的两条链,同时控制其中一条链通过跨膜孔的运动时,第二链可以留在孔附近并且在第一链易位通过孔之后,第二链可以被孔捕获,并且多核苷酸结合蛋白可以用于控制第二链通过孔的运动。

[0116] 另一方面,本发明人已经发现,可以向双链多核苷酸的每个末端提供具有双链体茎的衔接子,所述衔接子茎包含与缀合至纳米孔的孔标签互补的捕获序列,其中捕获序列仅在该链分离或解链时显露。因此,当双链多核苷酸的第一链穿过标签修饰的纳米孔时,它拉开衔接子的双链体茎以暴露双链多核苷酸的第二链上的捕获序列,然后被纳米孔的孔标签捕获。这种方法使第二链(其通常会扩散开)保持靠近纳米孔,以在为模板测序后进行测序。具体而言,与通常在典型纳米孔测序中观察到的时间的0.1% - 1%相比,本文所述的方法可以使模板易位后互补序列跟随易位的可能性显著增加,达到所述时间的至少约60%。

[0117] 还发现,纳米孔的修饰包含多个分析物的多个结合位点,使得一个或多个分析物可以通过结合位点与纳米孔结合,而通过纳米孔来表征分析物,可以增强表征分析物的灵敏度和/或通量。不希望受理论束缚,在纳米孔的外缘处偶联或捕获分析物可以增强该孔处

分析物的局部浓度。此外,纳米孔附近的至少一种或多种分析物可以容易地一个接一个地进入纳米孔以进行表征,从而减少时间延迟并因此减少每个分析物表征间隔的开孔电流时间。

[0118] 因此,本文的各个方面涉及使用纳米孔表征一种或多种分析物的方法,以及可以用于本文所述方法中的组合物和系统,其包括例如衔接子和纳米孔。一些方面的特征在于使用纳米孔表征双链多核苷酸,例如不使用连接所述双链多核苷酸的模板和互补序列的发夹的方法和组合物。其他方面的特征在于使用标签修饰的纳米孔表征分析物的方法和组合物,其具有增加的灵敏度和/或更高的通量。

[0119] 用于表征分析物(例如,双链多核苷酸)的方法

[0120] 一方面,本公开提供了一种为靶多核苷酸测序的方法,其包括:

[0121] (a)使跨膜孔接触:

[0122] (i)包含靶多核苷酸和与靶多核苷酸互补的多核苷酸的双链多核苷酸,其中靶多核苷酸和与靶多核苷酸互补的多核苷酸各自包含单链前导序列;和

[0123] (ii)能够分离双链多核苷酸的链并控制多核苷酸通过跨膜孔的运动的多核苷酸结合蛋白(例如,多核苷酸解链酶);

[0124] (b)检测对应于通过所述孔的离子流的信号,以检测易位通过所述孔的多核苷酸;

[0125] (c)鉴定对应于所述靶多核苷酸的易位的信号和对应于与所述靶多核苷酸互补的多核苷酸的单独易位的连续信号;

[0126] (d)分析(c)中鉴定的所述信号,

[0127] 从而为所述靶多核苷酸测序。

[0128] 所述方法在步骤(a)之前还可包括将单链前导序列附接于靶多核苷酸和互补多核苷酸的步骤。所述方法在步骤(a)之前还可包括消化靶多核苷酸的一端以在互补链上产生前导序列和/或消化互补多核苷酸的一端以在靶链上产生前导序列的步骤。所述方法还可包括使多核苷酸结合蛋白(例如,多核苷酸解链酶)与前导序列结合。(a)(ii)中的多核苷酸结合蛋白(例如,多核苷酸解链酶)可以与(a)(i)中的前导序列结合。

[0129] 在这个方面,第一多核苷酸结合蛋白(例如,多核苷酸解链酶)分离双链多核苷酸的靶链和互补链,并控制靶多核苷酸或互补多核苷酸通过跨膜孔的运动。第二多核苷酸结合蛋白(例如,多核苷酸解链酶),其可以是与第一多核苷酸结合蛋白(例如,多核苷酸解链酶)相同类型的另一种蛋白,或者可以是不同类型的多核苷酸结合蛋白(例如,多核苷酸解链酶),控制靶多核苷酸(其中互补多核苷酸已经易位通过该孔)或互补多核苷酸(其中靶多核苷酸已经易位通过该孔)通过跨膜孔的运动。第二多核苷酸结合蛋白(例如,多核苷酸解链酶)不需要分离靶多核苷酸和互补多核苷酸,因为双链多核苷酸的两条链的分离已经发生(第一多核苷酸结合蛋白(例如,多核苷酸解链酶)已将两条链分开,同时使其中一条链穿过该孔)。第一多核苷酸结合蛋白(例如,多核苷酸解链酶)通常是加工双链多核苷酸的多核苷酸结合蛋白。第二多核苷酸结合蛋白(例如,多核苷酸解链酶)通常是加工单链多核苷酸的多核苷酸结合蛋白。第一和/或第二多核苷酸结合蛋白(例如,多核苷酸解链酶)可以能够加工双链多核苷酸和单链多核苷酸。

[0130] 本文所述的另一方面的特征在于一种用于使两个非共价结合的分子依次易位通过纳米孔的方法。所述方法包括:使一对非共价结合的分子在促进该对非共价结合的分子

的第一成员易位通过纳米孔的条件下与所述纳米孔接触,其中该对的第二成员上的结合位点在第一成员易位通过纳米孔期间暴露,并且其中暴露的结合位点与纳米孔上存在的标签或系链结合。在第一成员易位通过纳米孔之前,第二成员上的结合位点未暴露(或被遮蔽)。

[0131] 如本文所用,术语“非共价结合的分子”是指包含第一成员和第二成员的分子,其中第一成员和第二成员通过非共价附接的方式彼此缔合并且可以作为单独实体彼此分离。第一成员和第二成员之间的分离和结合过程是可逆的。非共价附接方式的实例包括但不限于互补碱基配对、离子相互作用、疏水性相互作用和/或范德华(Van der Waals)相互作用。

[0132] 在一些实施方案中,非共价结合的分子包含互补多核苷酸链。

[0133] 在一些实施方案中,纳米孔上的标签是寡核苷酸,并且第二成员上的所述结合位点是具有与所述标签互补的序列的核酸的一部分。

[0134] 在一些实施方案中,该对非共价结合的分子包含与衔接子核酸附接的靶核酸(例如,靶双链多核苷酸),并且其中所述结合位点存在于所述衔接子上。

[0135] 仅作为实例,图9B显示了非共价结合的分子,其包含附接于每个末端的互补多核苷酸链(例如,模板链和互补链)和衔接子(例如,衔接子核酸)。如图9A所示,衔接子包含双链体茎(5)和从双链体茎的模板链延伸的第一单链多核苷酸(1)。双链体茎(5)包含与第一单链多核苷酸(1)对准的链上的捕获序列,其中捕获序列与缀合至纳米孔外缘的标签(例如,捕获多核苷酸)互补。在一些实施方案中,第一单链多核苷酸(1)还可包含前导序列。虽然图9A显示了从双链体茎的互补链(例如,互补序列链)延伸的第二单链多核苷酸,但不是必需的。然而,在一些实施方案中,可能需要具有一个或多个第二单链多核苷酸,其包含一条或多条用于固体基质,例如膜或珠粒和/或纳米孔的系链。当第二单链多核苷酸不与第一单链多核苷酸互补时,形成Y-衔接子,例如,如图9A中所示。

[0136] 图9A显示了包含至少一个用于固体基质(例如,膜或珠粒)的锚的示例衔接子,而图15显示了包含至少两个锚的示例衔接子,其中第一锚能够拴系到固体基质,例如,膜或珠粒,并且第二锚能够拴系到纳米孔。纳米孔的第二锚可以构造成与缀合至纳米孔的标签结合。在一些实施方案中,纳米孔的第二锚可以构造成与缀合至纳米孔的标签直接结合。例如,纳米孔的第二锚可以包含与缀合至纳米孔的捕获多核苷酸标签互补的序列。在替代实施方案中,用于纳米孔的第二系链可以构造成与缀合至纳米孔的标签间接结合。例如,图36显示附接于分析物的衔接子可以通过微粒与纳米孔上的标签偶联,这将在下面的“微粒”部分中进一步详细描述。

[0137] 应注意,本文描述的衔接子可以附接于双链多核苷酸的任一端或两端。在一些实施方案中,相同的衔接子附接于双链多核苷酸的两端。在一些实施方案中,不同的衔接子可以附接于双链多核苷酸的末端。例如,通过将两个或更多个不同衔接子群体与双链多核苷酸混合在一起,可以实现不同衔接子与双链多核苷酸末端的附接。通常,形成与不同衔接子附接的双链多核苷酸的混合物,但也存在获得所需异衔接子混合物的方法(例如,通过纯化或通过控制衔接子与双链多核苷酸末端的附接)。

[0138] 在一些实施方案中,双链多核苷酸可具有连接至其3'末端或5'末端的衔接子。

[0139] 可以将平端双链多核苷酸捕获到纳米孔中并拉开。因此,在一些实施方案中,没有衔接子的平端构建体(例如,如本文所述的一种)可用于本文所述方法的任何方面。虽然不是必需的,但在一些实施方案中,期望具有与双链多核苷酸的至少一个末端偶联的前导序

列,例如,以提高纳米孔的捕获效率。

[0140] 在衔接子附接于双链多核苷酸的两端的一些实施方案中,本领域的普通技术人员将容易认识到,当衔接子的第一单链多核苷酸与双链多核苷酸一端的模板链偶联时,另一衔接子的第一个单链多核苷酸与双链多核苷酸相对端的互补链偶联。

[0141] 在一些实施方案中,衔接子可以在靶多核苷酸的每个末端具有预先结合的多核苷酸结合蛋白(例如,多核苷酸解链酶)。在一些实施方案中,所述方法还可包括在溶液中添加多核苷酸结合蛋白(例如,多核苷酸解链酶),使其结合靶多核苷酸每个末端的衔接子。如图10所示,双链多核苷酸的每个末端与如本文所述的衔接子附接,并且多核苷酸结合蛋白(例如,多核苷酸解链酶)负载在衔接子上,其中衔接子包含具有与缀合至纳米孔的标签互补的捕获序列的双链体茎(参见,例如,图9A作为示例衔接子)。从包含双链多核苷酸的构建体延伸的单链多核苷酸(其可任选地包含前导序列)进入纳米孔(图A)。进入纳米孔的第一链标记为模板(T),捕获的第一链的反向互补序列标记为互补序列(C)。当第一模板链在多核苷酸结合蛋白(例如,多核苷酸解链酶,例如解旋酶马达)的控制下穿过孔时,互补序列逐渐被拉开。图B显示,在模板链的末端,互补序列上在双链体茎内的捕获序列通过拉开而暴露,并因此与纳米孔上的标签(例如,捕获多核苷酸)偶联。在图C中,当模板链最终穿过纳米孔并且酶解离时,互补链通过与纳米孔上的标签(例如,捕获多核苷酸)结合而保持与纳米孔偶联。在一定时间后,互补链被其前导序列捕获。在图D中,互补链在负载的第二多核苷酸结合蛋白,例如多核苷酸解链酶如解旋酶马达的控制下穿过纳米孔。当互补链穿过纳米孔时,捕获序列将在某一时刻被从纳米孔的标签(例如,捕获多核苷酸)上拉开,从而释放纳米孔的标签(例如,捕获多核苷酸),使其可用于下一链。在该实施方案中,两种多核苷酸结合蛋白(例如,多核苷酸解链酶)都能够加工双链多核苷酸。在一些实施方案中,在两端负载的多核苷酸结合蛋白(例如,多核苷酸解链酶)可以相同或不同。

[0142] 虽然图10(图B)说明使双链体内的捕获序列暴露(用于结合纳米孔上的标签)发生在模板链易位即将结束时(例如,捕获远端衔接子,从而将其保持在孔的局部)但是更通常地,可以将链设计成使得双链多核苷酸(例如,通过末端或尾部)与纳米孔的结合可以就在拉开过程开始时发生。在这种情况下,捕获序列可以定位在衔接子中(例如,作为Y-衔接子部分的非互补臂),使其在拉开的整个持续时间或甚至在拉开过程之前暴露以与纳米孔结合。

[0143] 在一些实施方案中,没有使多核苷酸结合蛋白(例如,多核苷酸解链酶)结合附接于靶多核苷酸的衔接子,可以将多核苷酸结合蛋白(例如,多核苷酸解链酶)固定在纳米孔的内腔中,使得可以使用单个多核苷酸结合蛋白(例如,多核苷酸解链酶)控制两条链通过纳米孔的运动。

[0144] 在一些实施方案中,当一条链易位通过纳米孔时,链的解链或分离受多核苷酸结合蛋白(例如,多核苷酸解链酶)控制。在一些实施方案中,当一条链易位通过纳米孔时,链的解链或分离可以在不存在多核苷酸结合蛋白(例如,多核苷酸解链酶)的情况下进行。控制多核苷酸的两条链的运动和/或分离的这种无酶方法是本领域已知的。例如,某些纳米孔本身可以提供使多核苷酸如马达蛋白纳米孔解链的力,包括例如phi29马达蛋白纳米孔,例如,如Wendell等人"Translocation of double-stranded DNA through membrane-adapted phi29 motor protein nanopores"Nat Nanotechnol, 4 (2009), 第765-772页所

述,和/或美国专利号8,986,528中所述的纳米孔,其各自的内容是通过引用整体并入本文。

[0145] 如本文所用,术语“易位”是指沿着纳米孔的至少一部分运动。在一些实施方案中,易位从纳米孔的顺侧移动到纳米孔的反侧。

[0146] 靶多核苷酸通常存在于包含靶多核苷酸的多个拷贝的样品中和/或存在于包含多种不同多核苷酸的样品中。在一些实施方案中,本文所述任何方面的方法均可包括确定样品中一个或多个靶多核苷酸的序列。所述方法可包括使孔与两个或更多个双链多核苷酸接触。例如,所述方法可包括使孔与样品接触,其中基本上所有双链多核苷酸在其两条链的每一条上都具有单链前导序列。在一些实施方案中,双链多核苷酸仅通过互补碱基配对彼此偶联。在这些实施方案中,双链多核苷酸可具有四个游离末端,其中游离末端是多核苷酸链的末端。多核苷酸链的末端可为单链,例如单链突出端,或与另一条多核苷酸链碱基配对。在一些实施方案中,测序的双链多核苷酸的两条链不共价附接(例如,无发夹或其他共价附接)。然而,可以将不桥接模板和互补多核苷酸的部分添加到一个或多个游离末端。

[0147] 在本文所述的各个方面的一些实施方案中,所述方法还可包括在与纳米孔接触之前,产生单链前导序列或使其附接于样品中基本上所有双链多核苷酸的两条链上的步骤。添加的前导序列可以有一种或多种多核苷酸结合蛋白(例如,多核苷酸解链酶)与其附接,使得双链多核苷酸群体各自包含在其两条链中的每一条的一端有多核苷酸结合蛋白(例如,多核苷酸解链酶)与其附接的前导序列。

[0148] 包含靶多核苷酸(例如,模板)和与靶多核苷酸互补的多核苷酸(例如,互补序列)的双链多核苷酸可以具有衔接子,所述衔接子包含附接于其每个末端的单链前导序列。在各方面的方法均涉及多核苷酸结合蛋白(例如,多核苷酸解链酶)的一些实施方案中,所述方法可以包括使孔与两种或更多种可以相同或不同的多核苷酸结合蛋白(例如,多核苷酸解链酶)接触。不同的多核苷酸结合蛋白(例如,多核苷酸解链酶)可以与单独的前导序列结合,所述前导序列可以相同或不同。例如,在5'至3'方向起作用的多核苷酸结合蛋白(例如,多核苷酸解链酶)可以与靶多核苷酸的5'末端和/或互补多核苷酸的5'末端的前导序列结合。在3'至5'方向起作用的多核苷酸结合蛋白(例如,多核苷酸解链酶)可以与靶多核苷酸的3'末端和/或互补多核苷酸的3'末端的前导序列结合。

[0149] 本文所述的各个方面的方法中使用的衔接子可以进一步构造为允许分析物(例如,靶多核苷酸或非共价结合的分子)与纳米孔结合,以提高所述表征方法的灵敏度和/或通量。这与暴露衔接子的双链体内的捕获序列以允许非共价结合的分子的第二成员(例如,双链多核苷酸的互补链)结合的实施方案的目的不同,例如,如图10所示,是为了增加模板易位后互补序列易位的可能性,从而提高测序信息的准确度。如图16中所示,衔接子进一步构造为包括用于结合纳米孔的系链,例如,以促进分析物的捕获。因此,在分析物的第一成员(例如,双链多核苷酸的模板链)进入纳米孔之前,分析物与缀合至纳米孔的第一标签结合。当第一成员易位通过孔并拉开双链体茎以暴露第二成员上的捕获序列时,第二个成员与缀合至纳米孔的第二标签结合,使得第二成员保持靠近纳米孔以便在第一成员后进行后续表征。纳米孔上的第一标签和第二标签可以是不同的(例如,如图16所示),或者它们可以是相同的(例如,如图17A所示),这取决于衔接子的设计。

[0150] 因此,本文提供的另一方面涉及使用纳米孔确定分析物特征的方法,所述方法包括:(a)提供纳米孔,所述纳米孔经修饰为包含至少两个或更多个在纳米孔内腔外的标签,

其中标签为至少两种或更多种分析物提供结合位点；(b) 在使得至少一种或多种(例如, 至少一种、至少两种、至少三种、至少四种、至少五种、至少六种、至少七种、至少八种、至少九种或更多种)分析物与纳米孔上的标签结合的条件下使多种分析物与纳米孔接触, 同时来自所述多种的分析物易位通过纳米孔。

[0151] 不同于例如, 通过改变纳米孔内腔内的氨基酸的电荷和/或疏水性修饰以改善纳米孔与靶分析物的相互作用的纳米孔, 本文所述的纳米孔经修饰以提供多个标签, 以在分析物易位通过纳米孔时捕获多个分析物以进行表征。这减少了每次分析物表征间隔的孔开放时间, 并且还增加了分析物的局部浓度, 从而增加了该方法的灵敏度。在一些实施方案中, 如下文“标签或系链修饰的纳米孔(例如, 用于增强分析物捕获如多核苷酸捕获)”部分中描述的标签修饰的纳米孔可用于实现此目的。

[0152] 在一些实施方案中, 可以修饰分析物以与纳米孔上的标签结合。在一些实施方案中, 分析物包含如本文所述的衔接子, 例如包含用于纳米孔的锚的衔接子。

[0153] 纳米孔上的标签与分析物上的结合位点(例如, 附接于分析物的衔接子中存在的结合位点, 其中结合位点可由衔接子的锚或前导序列提供或由衔接子的双链体内的捕获序列提供)之间的相互作用可以是可逆的。例如, 分析物可以例如通过其衔接子与纳米孔上的标签结合, 并且在某一时刻, 例如在通过纳米孔表征分析物期间释放。强的非共价键(例如, 生物素/抗生物素蛋白)仍然是可逆的, 并且可用于本文所述方法的一些实施方案中。例如, 为了确保在模板易位后双链多核苷酸的互补序列易位, 可能需要设计一对孔标签和分析物衔接子以在双链多核苷酸的互补序列(或与互补序列附接的衔接子的一部分)与纳米孔之间提供足够的相互作用, 使得互补序列保持靠近纳米孔(在模板易位期间没有从纳米孔脱离和扩散)但是在其易位通过纳米孔时能够从纳米孔释放。

[0154] 因此, 在一些实施方案中, 本文所述方法中使用的一对孔标签和分析物衔接子可以构造成使得分析物上的结合位点(例如, 附接于分析物的衔接子中存在的结合位点, 其中结合位点可由衔接子的锚或前导序列提供或由衔接子的双链体茎内的捕获序列提供)与纳米孔上的标签的结合强度或亲和力足以将纳米孔和分析物之间的偶联维持一段时间, 直到施加的力处于其上以使结合的分析物从纳米孔释放。在分析物是双链多核苷酸的一些实施方案中, 施加的力可以是互补链末端易位通过纳米孔。

[0155] 在本文所述的各个方面的一些实施方案中, 所述方法还可包括, 在施加跨膜电位时, 响应于分析物(例如, 多核苷酸)穿过纳米孔检测信号。在一些实施方案中, 可以通过提供离子流的渗透失衡来驱动电位差。在一些实施方案中, 可以在位于纳米孔两侧的两个电极之间跨纳米孔施加电位差。信号可以是电测量和/或光学测量。可能的电测量包括: 电流测量、阻抗测量、隧穿测量或电子隧穿测量(Ivanov AP等人,《纳米快报》(Nano Lett.) 2011年1月12日; 11(1): 279-85)和FET测量(国际申请WO 2005/124888), 例如电压FET测量。在一些实施方案中, 信号可以是跨固态纳米孔的电子隧穿或跨固态纳米孔的电压FET测量。光学测量可以与电测量结合(Soni GV等人,《科学仪器综述》(Rev Sci Instrum.) 2010年1月; 81(1): 014301)。测量可以是跨膜电流测量, 例如对流过孔的离子电流的测量。图11A-11B显示了使用本文所述的方法在多核苷酸结合蛋白(例如, 多核苷酸解链酶)的控制下通过纳米孔为双链多核苷酸测序期间, 随时间推移测量的典型电流信号。

[0156] 替代性地, 测量可以是指示通过通道的离子流的荧光测量, 如通过Heron等人,《美

国化学协会期刊(J.Am.Chem.Soc.)》2009,131(5),1652-1653所公开的或使用FET测量跨膜的电压。在一些实施方案中,所述方法还可包括,当施加跨膜电位时,在多核苷酸相互作用和/或移动通过纳米孔时检测流过纳米孔的离子电流。在一些实施方案中,可以使用膜片钳或电压钳来进行所述方法。在一些实施方案中,可以使用电压钳来进行所述方法。电测量可以使用标准信号通道记录设备来进行,如Stoddart D等人,《美国国家科学院院刊》(Proc Natl Acad Sci),12;106(19):7702-7;Lieberman KR等人,《美国化学会志》(J Am Chem Soc.)2010;132(50):17961-72;以及国际申请WO 2000/28312中所述。可替代地,可以使用多通道系统进行电测量,例如如国际申请WO 2009/077734和国际申请WO 2011/067559中所述。

[0157] 可以提供一系列纳米孔以增加通量并因此增加对多核苷酸链的测量,例如国际申请WO2014/064443中公开的,其内容通过引用并入本文。

[0158] 信号测量分析

[0159] 在一些实施方案中,所述方法包括检测对应于通过孔的离子流,指示相互作用,例如多核苷酸易位通过孔的信号。在一些实施方案中,可以通过提供离子流的渗透失衡来驱动电位差。在一些实施方案中,可以在位于孔两侧的两个电极之间在跨膜孔上施加电位差。可替代地,测量可以是指示通过通道的离子流的荧光测量,如通过Heron等人,《美国化学协会期刊(J.Am.Chem.Soc.)》2009,131(5),1652-1653所公开的。可以提供一系列纳米孔以增加通量并因此增加对多核苷酸链的测量,例如WO2014064443中公开的。图4显示了在酶控制下多核苷酸易位通过纳米孔期间随时间测量的典型电流信号。待易位的多核苷酸通过发夹连接时,可在发夹中提供非核苷酸或经修饰的核苷酸以提供指示发夹的信号。电流信号反映了多核苷酸在易位通过纳米孔时的序列。因此,可以确定信号的哪些部分指示模板和互补序列。通常,酶使多核苷酸松脱通过纳米孔,产生特征电流水平。信号随时间的大小取决于纳米孔的性质,并且多于一个核苷酸可以在任何特定时间影响电流。

[0160] 在一些实施方案中,在任何特定时间影响电流的核苷酸的数量可以取决于一组 k 个核苷酸单元,其中 k 是复数整数,下文中称为“ k 聚体”。这可以在概念上被认为是具有比测量的聚合物单元大的“平端读取头”的纳米孔。在这种情况下,要求解的不同 k 聚体的数量增加到 k 的幂。例如,如果存在 n 个可能的聚合物单元,则要求解的不同 k 聚体的数量为 n^k 。用高数量的 k 聚体时,可能变得难以求解由不同 k 聚体产生的测量,不利于估计聚合物单元的潜在序列。

[0161] 通过执行已知分析技术如运行‘t-检验’,可以将电流与时间采样数据点关联成连续组,运行‘t-检验’试图找到信号的局部均值的变化。这些组称为事件。指示特定 k 聚体的事件可以如图6中所示进行确定。事件用一些概括特征表示(关联组内数据点的平均电流和平均电流的标准偏差)。

[0162] 为了确定多核苷酸序列,在一些实施方案中,可以参考模型,其考虑了 k 聚体之间可能的转变数量,并且还考虑了电流水平。此类分析技术公开于WO2013041878中,其特此通过引用并入,其中提到使用概率分析技术如隐马尔可夫模型(HMM)来确定可能的转变总数,并且其中随后通过分析技术如维特比算法来确定最可能的转变。可以采用递归神经网络(RNN)作为HMM的替代,并且在例如描述事件与产生事件的核苷酸数量之间的潜在关系方面,提供比HMM更大的数学表达自由度。此类采用RNN的方法在图7中通过实例进行了说明,

其中由事件推导出包含来自其他相邻事件的信息的特征。这为RNN提供了额外信息，RNN是一种数学模型，其输出取决于先前对数据序列的计算。图8中示意性地和简单地示出了如何采用维特比算法来确定通过具有最高似然性的可能转变的路径的实例。

[0163] k聚体之间最可能的转变可用于确定k聚体序列，从而确定核苷酸链的核苷酸序列。由于所采用的数学方法的性质，核苷酸序列常常以%准确度表示。

[0164] 在仅涉及测量模板链的方法中，上述方法可用于确定模板序列。然而，在测量模板及其反向互补序列的情况下，模板与其互补序列之间的配对关系可以为测序测量提供额外的功效。利用这种关系的特定技术的实例公开于WO2013041878中，由此将模板事件 t_i 对于可能的k聚体状态 s_k 和互补序列事件 c_j 对于可能的k聚体状态的两个矩阵组合形成三维矩阵(2D)模型。2D模型找到两个系列事件的最可能对准，以及解释这些观察结果的k聚体。该模型考虑了模板和互补序列的组合。

[0165] 原则上，2D碱基识别器必须检查 (t_i, c_j, s_k) 的所有三元组以及模板事件、互补序列路径和k具体之间可能的关联数。在实践中，完整的2D计算在长读段长度下在计算上变得不切实际，因此有用的近似值会约束2D模型，如WO201514035中所公开的。

[0166] 可替代地，可以通过比较模板和互补序列事件数据或核苷酸序列并确定模板和互补序列之间的最佳关联来进行核苷酸序列的确定。然而，这种方法不能提供更高的2D模型准确度，因为它未考虑模板和互补链之间组合的最高可能性。公开了比较模板和互补序列碱基识别的共有方法的实例

[0167] 与模板和互补链通过发夹连接的情况不同，互补链可能并非总是按顺序跟随模板链进入纳米孔中。例如，在双链构建体的模板链易位后，存在来自第二双链构建体的模板链可能使纳米孔易位的可能性。此外，在模板(第一)链易位后，存在互补(第二)链可能不被纳米孔的结合位点捕获的可能性。这可能是由于例如纳米孔的一个或多个结合位点已经被一条或多条互补链占据，因此结合位点不可用于互补链。未被纳米孔捕获的任何互补链都将会可能从纳米孔扩散而不被纳米孔捕获。因此，为了利用模板链和互补链的附加功效，首先必须确定信号测量是否对应于模板及其相应的互补序列。

[0168] 在本文所述的各个方面的一些实施方案中，该方法还可包括鉴定对应于所述靶多核苷酸的易位的信号和对应于与所述靶多核苷酸互补的多核苷酸的单独易位的连续信号，并分析这样鉴定的信号。对于这样鉴定的那些信号，可以使用上述数学方法确定靶标的核苷酸序列，其中所述方法利用靶标(模板)和互补序列的信息及其相关优点。

[0169] 为了鉴定信号(例如，连续信号)是否对应于靶标及其互补序列，可以将事件彼此比对以便确定对准程度。根据对准程度，可以确定实际上信号是否对应于靶标及其互补序列。用于产生正相关的对准%可以任意选择，并且可以例如大于95%。可以采用已知的成对比对方法，例如Smith-Waterman或Needleman-Wunsch算法。可以使用的比对方法的合适实例公开于WO2015/140535或WO 2016/059427中。

[0170] 已经观察到，通常，互补链跟随其模板链以连续顺序进入纳米孔或者从纳米孔扩散开。在与该特定互补链无关的另一条链进入纳米孔后，互补链跟随其模板链的可能性要低得多。模板和互补序列对具有独特特征。例如，模板/互补序列对通常倾向于共有相同长度的核苷酸(事件数量)。此外，该对的跟随链可比新链快得多地进入纳米孔，和/或互补序列的测序往往更快，等等。一个或多个这些特征可用于鉴定模板/互补序列对，无需复杂的

计算分析。

[0171] 在一些实施方案中,为了减少计算需求,可以将比对限制为相邻的链测量。一旦确立了链间的模板-互补序列关系,序列的确定就可以利用对模板或互补链的测量或对模板和互补链的测量。例如,可以确定模板链的序列,其中可以认为测序准确度不够高。在这种情况下,该方法可以选择通过考虑模板和互补序列数据来确定序列,以提供高于仅通过确定模板序列而获得的序列准确度。可替代地,可以认为模板链的序列准确度足够好,以致不需要考虑模板和互补序列数据两者。确定是否使用模板序列数据或者是否使用模板和互补序列数据两者的因素可以是例如基础序列是否具有难以精识别的碱基或碱基组,或例如特定碱基是否是单个核苷酸多态性变体。

[0172] 在没有确定序列关系的情况下,可以按照对于仅模板链的测量所进行的相同方式确定该特定链的序列。在确定序列关系的情况下,可以按照对于模板和互补链的测量所进行的相同方式确定该特定链的序列。可以组合该信息以提供整体序列测定。

[0173] 在一些实施方案中,本文所述的方法还包括分析当第一链(例如,靶多核苷酸)易位通过跨膜孔时产生的信号以及当与第一链互补的第二链易位通过相同纳米孔时产生的信号。第一链(例如,靶多核苷酸)及其互补序列(第二链)通过碱基配对连接。因此,一旦第一多核苷酸结合蛋白(例如,多核苷酸解链酶)沿着双链多核苷酸的长度移动,第一链和第二链就不再连接。参见,例如,图10。因此,第二链易位通过纳米孔是与第一链易位通过纳米孔分开的事件,例如,如图11所示,其中在第一链和第二链易位间期观察到约200pA的开孔电流(没有阻断电流的链)。虽然不是必需的,但是期望第二链的易位在第一链易位后尽快,例如立即(例如,小于1秒)发生。参见,例如,图13。

[0174] 该方法包括鉴定对应于所述靶多核苷酸和与所述靶多核苷酸互补的多核苷酸的连续易位的信号的步骤。连续易位包括互补多核苷酸易位通过与靶多核苷酸相同的孔的情况。靶多核苷酸和互补多核苷酸可以以任何顺序易位通过该孔。其他多核苷酸,例如1、2、3、4或5个至约10个多核苷酸可以穿过靶多核苷酸和互补多核苷酸之间的孔。优选地,靶多核苷酸和互补多核苷酸以任何顺序连续地穿过该孔。孔优选在第一靶多核苷酸和互补多核苷酸易位通过该孔与第二靶多核苷酸和互补多核苷酸易位通过该孔之间恢复开放状态。

[0175] 通过将靶多核苷酸和/或互补多核苷酸拴系到膜和/或孔上,可以促进靶多核苷酸和互补多核苷酸连续穿过该孔。促进双链多核苷酸的两条链连续易位的其他方法包括将靶多核苷酸和/或互补多核苷酸附接于微粒和/或修饰孔以增加/增强多核苷酸捕获。

[0176] 可以使用条形码来促进对应于靶多核苷酸易位的信号和对应于与靶多核苷酸互补的多核苷酸的单独易位的连续信号的鉴定。通常,双链条形码包含在双链多核苷酸中或与双链多核苷酸附接。当靶多核苷酸和互补多核苷酸分离时(通过多核苷酸结合蛋白(例如,多核苷酸解链酶)),条形码保留在靶多核苷酸和互补多核苷酸两者中。条形码易位通过跨膜孔将产生该条形码特有的信号。第二次和随后检测条形码易位通过孔可用于确定靶多核苷酸及其互补序列已依次易位通过该孔。从而可以鉴定对应于靶多核苷酸的易位和对应于与所述靶多核苷酸互补的多核苷酸的单独易位的信号。

[0177] 标签或系链修饰的纳米孔(例如,用于增强分析物捕获,如多核苷酸捕获)

[0178] 将用于本文所述方法的纳米孔修饰为包含一个或多个结合位点,用于结合一种或多种分析物。在一些实施方案中,可以将纳米孔修饰为包含一个或多个结合位点,用于结合

附接于分析物的衔接子。例如,在一些实施方案中,纳米孔可以与附接于分析物的衔接子的前导序列结合。在一些实施方案中,纳米孔可以与附接于分析物的衔接子中的单链序列结合。在一些实施方案中,纳米孔可以与附接于分析物的衔接子的双链体内的捕获序列结合,其中捕获序列仅在双链体茎解链时显露。

[0179] 在一些实施方案中,可以将纳米孔修饰为包含一个或多个结合位点,用于结合附接于双链寡核苷酸的第一链或第二链的衔接子,例如,以促进第一链和第二链连续易位通过跨膜孔。

[0180] 在一些实施方案中,将纳米孔修饰为包含一个或多个标签或系链,每个标签或系链均包含分析物的结合位点。

[0181] 在一些实施方案中,将纳米孔修饰为包含两个或更多个标签或系链。例如,可以提供一个标签或系链以增加表征分析物如多核苷酸的方法的灵敏度(“灵敏度标签”),同时可以提供另一标签或系链以增加对多核苷酸的模板链之后的互补链测序的可能性(“跟随标签”)。如图15所示,孔标签可以以多种方式构造。仅举例来说,在一些实施方案中,寡聚孔的每个单体可以具有相同类型的标签构型(例如,具有多个结合位点,如Tag-A和Tag-B所示)。Tag-A和Tag-B可以组合形成单个标签,至少一个或多个单体包含Tag-A/Tag-B组合标签。可替代地,寡聚孔可包含有不同标签附接的混合单体,使得至少一种单体具有与其他单体不同的标签构型。在另一个实例中,Tag-A和Tag-B可以保持为单独的标签,并且至少一个或多个单体可以包括两种单独的标签。如果灵敏度标签和跟随标签与如本文所述的衔接子中使用的独特序列互补,则可以单独组合。

[0182] 图16是可如何使用具有两种不同标签类型的纳米孔从溶液中捕获链(以便提高灵敏度)的示意图。附接于双链多核苷酸末端的衔接子包含可用于偶联至第一孔标签的捕获序列(例如,形成Y-衔接子的非互补臂),同时双链体茎内只有在被拉开时才显露的单独捕获序列允许互补序列结合第二孔标签,因此使互补序列捕获成为可能以进行跟随测序。

[0183] 纳米孔上的标签与分析物上的结合位点(例如,附接于分析物的衔接子中存在的结合位点,其中结合位点可由衔接子的锚或前导序列提供或由衔接子的双链体内的捕获序列提供)之间的相互作用可以是可逆的。例如,分析物可以例如通过其衔接子与纳米孔上的标签结合,并且在某一时刻,例如在通过纳米孔表征分析物期间释放。强的非共价键(例如,生物素/抗生物素蛋白)仍然是可逆的,并且可用于本文所述方法的一些实施方案中。例如,为了确保在模板易位后双链多核苷酸的互补序列易位,可能需要设计一对孔标签和分析物衔接子以在双链多核苷酸的互补序列(或与互补序列附接的衔接子的一部分)与纳米孔之间提供足够的相互作用,使得互补序列保持靠近纳米孔(在模板易位期间没有从纳米孔脱离和扩散)但是在其易位通过纳米孔时能够从纳米孔释放。

[0184] 因此,在本文所述各个方面的一些实施方案中,一对孔标签和分析物衔接子可以构成使得分析物上的结合位点(例如,附接于分析物的衔接子中存在的结合位点,其中结合位点可由衔接子的锚或前导序列提供或由衔接子的双链体内的捕获序列提供)与纳米孔上的标签的结合强度或亲和力足以维持纳米孔和分析物之间的偶联,直到施加的力处于其上以使结合的分析物从纳米孔释放。在分析物是双链多核苷酸的一些实施方案中,施加的力可以是互补链末端易位通过纳米孔。

[0185] 在一些实施方案中,标签或系链不带电。这可以确保标签或系链在电位差的影响

下不被吸入纳米孔中。

[0186] 吸引或结合多核苷酸或衔接子的一个或多个分子可以与孔连接。可以使用与衔接子和/或靶多核苷酸杂交的任何分子。附接于孔的分子可选自PNA标签、PEG接头、短寡核苷酸、带正电荷的氨基酸和适体。具有与它们连接的此类分子的孔是本领域已知的。例如, Howarka等人(2001)Nature Biotech.19:636-639和WO 2010/086620中公开了有短寡核苷酸与其附接的孔,并且在Howarka等人(2000)J.Am.Chem.Soc.122(11):2411-2416中公开了包含附接在孔内腔中的PEG的孔。

[0187] 附接于跨膜孔的短寡核苷酸,该寡核苷酸包含与前导序列中的序列或衔接子中的另一单链序列互补的序列,可用于增强本文所述任何方面的方法中靶多核苷酸和/或互补多核苷酸的捕获。

[0188] 在一些实施方案中,标签或系链可包含或是寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基)。寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基)可具有约10-30个核苷酸的长度或约10-20个核苷酸的长度。示例性寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基)可包含SEQ ID NO:8中所示的序列。在一些实施方案中,用于标签或系链中的寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基)可具有至少一个被修饰用于缀合至其他修饰或固体基质表面(包括,例如珠粒)的末端(例如,3'-或5'-末端)。末端改性剂可以添加可以用于缀合的反应性官能团。可添加的官能团的实例包括但不限于氨基、羧基、硫醇、马来酰亚胺、氨氧基及其任何组合。官能团可以与不同长度的间隔区(例如,C3、C9、C12、间隔区9和18)组合以增加官能团与寡核苷酸序列末端的物理距离。在一些实施方案中,标签或系链可以是寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基),其具有SEQ ID NO:8所示的具有5'-马来酰胺修饰的序列。在一些实施方案中,标签或系链可以是寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基),其具有SEQ ID NO:8所示的具有3'-马来酰胺修饰的序列。在一些实施方案中,标签或系链可以是寡核苷酸(例如,DNA、RNA或PNA),其具有SEQ ID NO:8所示的具有5'-C9-硫醇修饰的序列。在一些实施方案中,标签或系链可以是寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基),其具有SEQ ID NO:8所示的具有3'-C9-硫醇修饰的序列。在一些实施方案中,标签或系链可以是寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基),其具有SEQ ID NO:8所示的具有5'-硫醇修饰的序列。在一些实施方案中,标签或系链可以是寡核苷酸(例如,DNA、RNA、LNA、BNA、PNA或吗啉基),其具有SEQ ID NO:8所示的具有3'-硫醇修饰的序列。

[0189] 在一些实施方案中,标签或系链可包含或是吗啉基寡核苷酸。吗啉基寡核苷酸可具有约10-30个核苷酸的长度或约10-20个核苷酸的长度。示例性吗啉基寡核苷酸可包含SEQ ID NO:8中所示的序列。吗啉基寡核苷酸可以是修饰的或未修饰的。例如,在一些实施方案中,吗啉基寡核苷酸可以在寡核苷酸的3'和/或5'末端被修饰。吗啉基寡核苷酸的3'和/或5'末端上的修饰的实例包括但不限于3'亲和标签和用于化学连接的官能团(包括,例如3'-生物素、3'-伯胺、3'-二硫化物酰胺、3'-吡啶基二硫基及其任何组合);5'末端修饰(包括,例如5'-伯胺和/或5'-dabcyl),用于点击化学的修饰(包括,例如3'-叠氮化物、3'-炔烃、5'-叠氮化物、5'-炔烃)及其任何组合。在一些实施方案中,标签或系链可以是吗啉基寡核苷酸,其具有SEQ ID NO:8所示的具有5'-叠氮化物修饰的序列。在一些实施方案中,标签或系链可以是吗啉基寡核苷酸,其具有SEQ ID NO:8所示的具有3'-叠氮化物修饰的序

列。在一些实施方案中，标签或系链可以是吗啉基寡核苷酸，其具有SEQ ID N0:8所示的具有5'-炔烃修饰的序列。在一些实施方案中，标签或系链可以是吗啉基寡核苷酸，其具有SEQ ID N0:8所示的具有3'-炔烃修饰的序列。在一些实施方案中，标签或系链可以是吗啉基寡核苷酸，其具有SEQ ID N0:8所示的具有3'-吡啶基二硫基修饰的序列。

[0190] 在一些实施方案中，标签或系链还可包含聚合物接头，例如以促进与纳米孔的偶联。示例性聚合物接头包括但不限于聚乙二醇(PEG)。聚合物接头可具有约500Da至约10kDa(包括端值)，或约1kDa至约5kDa(包括端值)的分子量。聚合物接头(例如，PEG)可以用不同的官能团官能化，包括例如但不限于马来酰亚胺、NHS酯、二苯并环辛炔(DBCO)、叠氮化物、生物素、胺、炔烃、醛及其任何组合。在一些实施方案中，标签或系链还可包含具有5'-马来酰亚胺基团和3'-DBCO基团的1kDa PEG。在一些实施方案中，标签或系链还可包含具有5'-马来酰亚胺基团和3'-DBCO基团的2kDa PEG。在一些实施方案中，标签或系链还可包含具有5'-马来酰亚胺基团和3'-DBCO基团的3kDa PEG。在一些实施方案中，标签或系链还可包含具有5'-马来酰亚胺基团和3'-DBCO基团的5kDa PEG。

[0191] 标签或系链的其它实例包括，但不限于His标签、生物素或链霉抗生物素蛋白，结合分析物的抗体，结合分析物的衔接子，分析物结合结构域如DNA结合结构域(包括，例如肽拉链如亮氨酸拉链，单链DNA结合蛋白(SSB))及其任何组合。

[0192] 标签或系链可以使用本领域已知的任何方法附接于纳米孔的外表面，例如，在膜的顺侧上。例如，一个或多个标签或系链可以通过一个或多个半胱氨酸(半胱氨酸连键)、一个或多个伯胺如赖氨酸、一个或多个非天然氨基酸、一个或多个组氨酸(His标签)、一个或多个生物素或链霉抗生物素蛋白、一个或多个基于抗体的标签、表位的一个或多个酶修饰(包括，例如乙酰转移酶)及其任何组合附接到纳米孔。用于进行此类修饰的合适方法在所属领域中是众所周知的。合适的非天然氨基酸包含但不限于4-叠氮基-L-苯丙氨酸(Faz)，及Liu C.C.和Schultz P.G., Annu. Rev. Biochem., 2010, 79, 413-444的图1中编号为1-71的氨基酸中的任一种。

[0193] 在一个或多个标签或系链经由半胱氨酸连键附接于纳米孔的一些实施方案中，可以通过取代将所述一个或多个半胱氨酸引入一个或多个形成纳米孔的单体中。在一些实施方案中，可以通过附接如下各项来对纳米孔进行化学修饰：(i) 马来酰亚胺，包括二溴马来酰亚胺如：4-苯氮霉素(4-phenylazomaleinanol)、1.N-(2-羟乙基)马来酰亚胺、N-环己基马来酰亚胺、1,3-马来酰亚胺基丙酸、1,1-4-氨基苯基-1H-吡咯-2,5,二酮、1,1-4-羟基苯基-1H-吡咯-2,5,二酮、N-乙基马来酰亚胺、N-甲氧基羰基马来酰亚胺、N-叔丁基马来酰亚胺、N-(2-氨基乙基)马来酰亚胺、3-马来酰亚胺基-PROXYL、N-(4-氯苯基)马来酰亚胺、1-[4-(二甲基氨基)-3,5-二硝基苯基]-1H-吡咯-2,5-二酮、N-[4-(2-苯并咪唑基)苯基]马来酰亚胺、N-[4-(2-苯并恶唑基)苯基]马来酰亚胺、N-(1-萘基)马来酰亚胺、N-(2,4-二甲苯基)马来酰亚胺、N-(2,4-二氟苯基)马来酰亚胺、N-(3-氯-对-甲苯基)-马来酰亚胺、1-(2-氨基-乙基)-吡咯-2,5-二酮盐酸盐、1-环戊基-3-甲基-2,5-二氢-1H-吡咯-2,5-二酮、1-(3-氨基丙基)-2,5-二氢-1H-吡咯-2,5-二酮盐酸盐、3-甲基-1-[2-氧代-2-(哌嗪-1-基)乙基]-2,5-二氢-1H-吡咯-2,5-二酮盐酸盐、1-苄基-2,5-二氢-1H-吡咯-2,5-二酮、3-甲基-1-(3,3,3-三氟丙基)-2,5-二氢-1H-吡咯-2,5-二酮、1-[4-(甲基氨基)环己基]-2,5-二氢-1H-吡咯-2,5-二酮三氟乙酸、SMILES O=C1C=CC(=O)N1CC=2C=CN=CC2、SMILES O=

C₁C=CC(=O)N1CN2CCNCC2、1-苄基-3-甲基-2,5-二氢-1H-吡咯-2,5-二酮、1-(2-氟苯基)-3-甲基-2,5-二氢1H-吡咯-2,5-二酮、N-(4-苯氧基苯基)马来酰亚胺、N-(4-硝基苯基)马来酰亚胺；(ii) 碘代乙酰胺,如3-(2-碘乙酰氨基)-PROXYL、N-(环丙基甲基)-2-碘乙酰胺、2-碘-N-(2-苯乙基)乙酰胺、2-碘-N-(2,2,2-三氟乙基)乙酰胺、N-(4-乙酰基苯基)-2-碘代乙酰胺、N-(4-(氨基磺酰基)苯基)-2-碘代乙酰胺、N-(1,3-苯并噻唑-2-基)-2-碘代乙酰胺、N-(2,6-二乙基苯基)-2-碘代乙酰胺、N-(2-苯甲酰基-4-氯苯基)-2-碘代乙酰胺；(iii) 溴代乙酰胺:如N-(4-(乙酰氨基)苯基)-2-溴代乙酰胺、N-(2-乙酰基苯基)-2-溴代乙酰胺、2-溴-N-(2-氟基苯基)乙酰胺、2-溴-N-(3-(三氟甲基)苯基)乙酰胺、N-(2-苯甲酰基苯基)-2-溴代乙酰胺、2-溴-N-(4-氟苯基)-3-甲基丁酰胺、N-苄基2-溴-N-苯基丙酰胺、N-(2-溴-丁酰基)-4-氯-苯磺酰胺、2-溴-N-甲基-N苯基乙酰胺、2-溴-N-苯乙基-乙酰胺、2-金刚烷-1-基-2-溴-N-环己基-乙酰胺、2-溴-N-(2-甲基苯基)丁酰胺、乙酰替对溴苯胺；(iv) 二硫化物,如:ALDRITHIOL-2、ALDRITHIOL-4、异丙基二硫化物、1-(异丁基二硫烷基)-2-甲基丙烷、二苄基二硫化物、4-氨基苯基二硫化物、3-(2-吡啶基二硫代)丙酸酸、3-(2-吡啶基二硫代)丙酸酰肼、3-(2-吡啶基二硫代)丙酸N-琥珀酰亚胺酯、am6amPDP1-βCD;以及(v) 硫醇,如:4-苯基噻唑-2-硫醇、Pulpald、5,6,7,8-四氢-喹唑啉-2-硫醇。

[0194] 在一些实施方案中,标签或系链可以直接附接于纳米孔或通过一个或多个接头附接。可以使用WO 2010/086602中描述的杂交接头将标签或系链连接至纳米孔。可替代地,可以使用肽接头。肽接头是氨基酸序列。肽接头的长度、柔性和亲水性通常被设计为使得其不干扰单体和孔的功能。优选的柔性肽接头是2个到20个,如4个、6个、8个、10个或16个丝氨酸和/或甘氨酸的延伸段。更优选的柔性接头包含(SG)₁、(SG)₂、(SG)₃、(SG)₄、(SG)₅和(SG)₈,其中S是丝氨酸且G是甘氨酸。优选的刚性接头是2个到30个,如4个、6个、8个、16个或24个脯氨酸的延伸段。更优选的刚性接头包含(P)₁₂,其中P是脯氨酸。

[0195] 可以修饰跨膜孔以增强多核苷酸的捕获。例如,可以修饰孔以增加通向孔的入口内和/或孔的筒体内的正电荷。此类修饰是所属领域中已知的。例如,WO 2010/055307公开了α-溶血素中的突变,其增加孔的筒体内的正电荷。

[0196] WO 2012/107778、WO 2013/153359和WO 2016/034591中分别公开了包含增强多核苷酸捕获的突变的经修饰的MspA、胞溶素和CsgG孔。这些出版物中公开的任何经修饰的孔均可以在本文中使用。

[0197] 在一些实施方案中,可以将CsgG纳米孔修饰为包含如本文所述的一个或多个标签或系链。可以在以下位置通过氨基酸修饰将一个或多个标签或系链附接至CsgG纳米孔的一个或多个单体(例如,1、2、3、4、5、6、7、8、9个或更多个):SEQ ID NO:7的T3、K7、R11、Q19、K22、A29、T31、R76、N102、G103、N108、R110、Q114、E170、C215、L216、D238、A243、D248和H255。在一些实施方案中,可以在以下位置通过氨基酸取代将一个或多个标签或系链附接至CsgG纳米孔的一个或多个单体(例如,1、2、3、4、5、6、7、8、9个或更多个):SEQ ID NO:7的T3C、K7C、R11C、Q19C、K22C、A29C、T31C、R76C、E170C、D238C、A243C、D248C、H255C、C215A/T/S/M/G/I/L、L216V。

[0198] 在一些实施方案中,可以进一步修饰CsgG纳米孔以改善分析物通过纳米孔的捕获和/或易位,改善分析物(例如,多核苷酸)识别或区分,改善与多核苷酸解链酶的相互作用,和/或改善信噪比。例如,在一些实施方案中,形成CsgG纳米孔的至少一种单体可包含如WO

2016/034591公开的一种或多种突变。

[0199] 在一些实施方案中,CsgG纳米孔可包含以下氨基酸取代的组合之一(相对于SEQ ID N0:7) : (T3C) ; (K7C) ; (R11C) ; (Q19C) ; (K22C) ; (A29C) ; (T31C) ; (R76C) ; (E170C) ; (D238C) ; (A243C) ; (D248C) ; (H255C) ; (C215A) ; (C215T) ; (C215S) ; (C215M) ; (C215G) ; (C215I) ; (C215L) ; (C215A、L216V) ; (A29C、C215T) ; (T31C、C215T) ; (R76C、C215T) ; (T3C、C215A) ; (K7C、C215A) ; (R11C、C215A) ; (Q19C、C215A) ; (K22C、C215A) ; (A29C、C215A) ; (T31C、C215A) ; (R76C、C215A) ; (E170C、C215A) ; (C215A、D238C) ; (C215A、A243C) ; (C215A、D248C) ; (C215A、H255C) ; (R76C、N91R、C215A) ; (R76C、N91R、C215A) ; (R76C、C215A) ; 和 (R76C、C215T) 。

[0200] 多核苷酸

[0201] 多核苷酸可以是核酸,如脱氧核糖核酸(DNA)或核糖核酸(RNA)。多核苷酸可包含与一个DNA链杂交的一个RNA链。多核苷酸可以是本领域中已知的任何合成核酸,如肽核酸(PNA)、甘油核酸(GNA)、苏糖核酸(TNA)、锁核酸(LNA)或具有核苷酸侧链的其它合成聚合物。PNA主链由通过肽键连接的重复N-(2-氨基乙基)-甘氨酸单元构成。GNA主链由通过磷酸二酯键连接的重复二醇单元构成。TNA主链由通过磷酸二酯键连接在一起的重复苏糖构成。LNA由如上文所论述的具有额外的连接核糖部分中的2'氧和4'碳的桥的核糖核苷酸形成。

[0202] 多核苷酸优选是DNA、RNA或DNA或RNA杂交体,最优选是DNA。靶多核苷酸可以是双链。靶多核苷酸可以包含单链区和具有其它结构的区域,例如发夹环、三链体和/或四链体。DNA/RNA杂交体可以在同一条链上包含DNA和RNA。优选地,DNA/RNA杂交体包含与RNA链杂交的一条DNA链。

[0203] 在一些实施方案中,靶多核苷酸不包含连接模板和互补序列的发夹结构或任何共价连接。在一些实施方案中,靶多核苷酸(例如,模板)和与靶多核苷酸互补的多核苷酸(例如,互补序列)不通过桥接部分如发夹环连接。然而,在一些实施方案中,当单链(例如,模板或互补序列)易位通过纳米孔时,由于衔接子在其两端的相互作用,链本身可以形成发夹结构。参见,例如,图19。这种衔接子设计可有益于表征长多核苷酸,例如,通过保持链的另一端靠近纳米孔。

[0204] 靶多核苷酸可以是任何长度。例如,多核苷酸的长度可以是至少10、至少50、至少100、至少150、至少200、至少250、至少300、至少400或至少500个核苷酸或核苷酸对。靶多核苷酸可以是1000个或更多个核苷酸或核苷酸对,长度为5000个或更多个核苷酸或核苷酸对或长度为100000个或更多个核苷酸或核苷酸对或长度为500,000个或更多个核苷酸或核苷酸对,或长度为1,000,000个或更多个核苷酸或核苷酸对,长度为10,000,000或更多个核苷酸或核苷酸对,或长度为100,000,000或更多个核苷酸或核苷酸对,或长度为200,000,000或更多个核苷酸或核苷酸对,或染色体的整个长度。靶多核苷酸可以是寡核苷酸。寡核苷酸是短核苷酸聚合物,其通常具有50个或更少核苷酸,如40个或更少、30个或更少、20个或更少、10个或更少或5个或更少核苷酸。靶寡核苷酸的长度优选是约15至约30个核苷酸,例如长度是约20至约25个核苷酸。例如,寡核苷酸的长度可以是约15、约16、约17、约18、约19、约20、约21、约22、约23、约24、约25、约26、约27、约28、约29或约30个核苷酸。

[0205] 靶多核苷酸可以是较长的靶标多核苷酸的片段。在该实施方案中,较长的靶多核苷酸通常片段化为多个,例如两个或更多个较短的靶多核苷酸。本发明的方法可用于对那

些较短靶多核苷酸中的一个或多个,例如2、3、4、5个或更多个测序。

[0206] 在一些实施方案中,本文所述各个方面的方法可用于对样品内的多个靶多核苷酸,例如2、3、4或5个至10、15、20个或更多个多核苷酸取样。

[0207] 在一些实施方案中,本文所述各个方面的方法可用于为样品中以双链形式存在的多核苷酸测序。

[0208] 在一些实施方案中,本文所述各个方面的方法可用于通过首先合成单链多核苷酸的互补序列以产生双链多核苷酸来为单链多核苷酸测序。例如,单链多核苷酸可以是RNA,如mRNA,并且可以合成互补cDNA链以产生双链多核苷酸,用于在本发明的方法中进行测序。例如,单链多核苷酸可以是DNA,并且可以合成互补链以产生双链DNA多核苷酸,用于在本发明的方法中进行测序。

[0209] 在一些实施方案中,多核苷酸可以是串接的多核苷酸。串接多核苷酸的方法在PCT/GB2017/051493中有描述。在一个实施方案中,用于将多核苷酸连接在一起的附接方法是点击化学。在该实施方案中,当模板和互补序列不共价连接时,使用纳米孔表征第一双链多核苷酸的模板(捕获的第一链)和互补序列(第一链的反向互补序列)。当模板和互补序列分离时,与连接到第一双链多核苷酸的跟随衔接子中的孔系链互补的序列在互补序列中暴露,并且互补序列结合附接于纳米孔的孔系链。在该实施方案中,串接衔接子也连接到第一双链多核苷酸,使得互补链可以连接到第二双链多核苷酸。

[0210] 制备含有马达蛋白和释放蛋白的串接衔接子复合物。该串接衔接子可以连接到靶多核苷酸的两端。马达蛋白和释放蛋白都可以在连接的衔接子复合物上停滞,直到多核苷酸被孔捕获。一旦捕获了第一多核苷酸,阻断化学物质就被两种蛋白质克服,并且马达蛋白如前所述控制多核苷酸与孔的相互作用。释放蛋白,可以比马达蛋白更快地易位,到达第一多核苷酸的3'以释放与串接衔接子复合物的前导链的5'核酸序列互补的杂交位点。通过显露该杂交位点,第二多核苷酸然后可以与显露的位点杂交,并且第一多核苷酸的3'末端与第二多核苷酸的5'的共价偶联可以发生(图43)。然后重复该过程以进一步连接靶多核苷酸。

[0211] 在一个实施方案中,提供了表征和串接双链靶多核苷酸的方法,其中附接方法是非共价的。在该实施方案中,第一双链靶多核苷酸的互补链募集第二双链靶多核苷酸并使其达到孔的局部浓度。反过来,当对第一互补链测序时,募集的第二双链靶多核苷酸变得从互补链去杂交,而是与孔系链杂交。这使得第一和第二(以及随后的,第三、第四、第五等)双链靶多核苷酸能够彼此跟随通过该孔,链间隔时间最短。当双链靶多核苷酸的浓度较低时,这特别有用,因为第二靶多核苷酸可以在为第一靶多核苷酸测序时募集。

[0212] 在另一个实施方案中,表征和串接双链靶多核苷酸的方法(例如其中附接方法是非共价的),可以使用双组分捕捞系链进行,所述双组分捕捞系链为跟随序列和孔系链提供第二杂交位点,以增加所见事件的比例。

[0213] 在一个实施方案中,表征和串接许多双链靶多核苷酸的方法(例如其中附接方法是非共价的),可用于将多个,例如2至20个,例如4、5、6、8、10、12或15个双链靶多核苷酸带到孔处。第一双链靶多核苷酸的互补链可以募集其他双链靶多核苷酸并使其集中在孔附近。这样在孔周围提供比在一般本体溶液中更高的局部浓度,因此双链靶多核苷酸彼此跟随通过开孔,链间隔时间最短。当双链靶多核苷酸的浓度较低时,这尤其有用。在该实施方

案中,可以使用由与单链结合蛋白偶联的寡核苷酸组成的系链。当对第一双链靶多核苷酸的模板链进行测序时,互补链作为ssDNA释放到溶液中。其他双链靶多核苷酸的单链结合蛋白能够与ssDNA结合。作为跟随过程的一部分,当对互补链进行测序时,互补链的3'被拉回到孔中。ssDNA互补链上的单链结合蛋白在遇到控制互补序列通过孔的运动的马达蛋白时从互补链中移位,因此沉积在孔周围,增加局部浓度。这在图44中进行了描绘。如果靶多核苷酸的序列是已知的,则可以进行这种分析物拖网,但也可以将互补序列添加到孔系链的3',其可以用于平铺互补链的区段。

[0214] 样品

[0215] 分析物(包括,例如蛋白质、肽、分子、多肽、多核苷酸)可以存在于样品中。样品可以是任何合适的样品。样品可以是生物样品。可以在体外对从任何生物体或微生物获得或提取的样品进行本文所述方法的任何实施方案。生物体或微生物通常是太古细菌、原核或真核微生物,且通常属于以下五界中的一个:植物界、动物界、真菌界、原核生物界和原生生物界。在一些实施方案中,可以对从任何病毒中获得或提取的样品在体外执行本文所述各个方面的方法。

[0216] 样品优选是流体样品。样品通常包含体液。体液可以从人或动物获得。人或动物可能患有疾病,疑似患有疾病或有患病风险。样品可以是尿液、淋巴液、唾液、粘液、精液或羊水,但优选是全血、血浆或血清。通常,样品来源于人,但其可替代地可以来自另一种哺乳动物,如来自商业养殖动物,如马、牛、羊或猪,或者可替代地可以是宠物,如猫或狗。

[0217] 可替代地,来源于植物的样品通常从经济作物获得,如谷物、豆科植物、水果或蔬菜,例如小麦、大麦、燕麦、油菜、玉米、大豆、水稻、香蕉、苹果、蕃茄、马铃薯、葡萄、烟草、菜豆、小扁豆、甘蔗、可可、棉花、茶叶或咖啡。

[0218] 样品可以是非生物样品。非生物样品优选是流体样品。非生物样品的实例包括手术液;水,如饮用水、海水或河水;以及实验室测试用试剂。

[0219] 可以在分析前处理样品,例如通过离心或通过膜,所述膜过滤掉不需要的分子或细胞,如红血细胞。样品可以在获取之后立即测量。通常还可以在分析之前,优选在低于-70°C下储存样品。

[0220] 在一些实施方案中,样品可以包含基因组DNA。基因组DNA可以被片段化,或者本文所述任何方法还可包括使基因组DNA片段化。可以通过任何合适的方法使DNA片段化。例如,DNA的片段化方法是本领域中已知的。此类方法可以使用转座酶,例如MuA转座酶。

[0221] 前导序列

[0222] 前导序列通常包含聚合物。聚合物优选是带负电的。聚合物优选是多核苷酸,如DNA或RNA;修饰后的多核苷酸(如无碱基DNA);PNA;LNA;聚乙二醇(PEG)或多肽。前导序列优选包含多核苷酸,并且更优选包含单链多核苷酸。单链前导序列最优选包含DNA的单链,如poly dT区段。前导序列优选包含一个或多个间隔区。

[0223] 前导序列可以是任何长度,但其长度通常是10至150个核苷酸,例如长度是20至150个核苷酸。前导序列的长度通常取决于所述方法中使用的跨膜孔。

[0224] 前导序列优先旋入跨膜孔中,并且从而促进多核苷酸通过孔的移动。前导序列还可以用于将多核苷酸连接到如本文所论述的一个或多个锚。

[0225] 通常,前导序列存在于靶多核苷酸的一端和与靶多核苷酸互补的多核苷酸的一

端。前导序列可以存在于靶多核苷酸的5'末端和靶多核苷酸的互补序列的5'末端。可替代地，前导序列可以存在于靶多核苷酸的3'末端和靶多核苷酸的互补序列的3'末端。前导序列可以存在于靶多核苷酸的5'末端和互补多核苷酸的3'末端，或反之亦然。在后面的这些实施方案中，通常使用两种不同的多核苷酸结合蛋白（例如，多核苷酸解链酶），其中第一多核苷酸结合蛋白（例如，多核苷酸解链酶）沿着多核苷酸以5'至3'方向移动并且第二多核苷酸结合蛋白（例如，多核苷酸解链酶）沿着多核苷酸以3'至5'方向移动。

[0226] 前导序列可以通过任何合适的方法与双链多核苷酸附接。例如，前导序列可以与靶多核苷酸和/或其互补序列连接。可替代地，可以通过消化双链多核苷酸的一条链以在另一条链上产生单链突出端来产生前导序列。

[0227] 多核苷酸结合蛋白（例如，多核苷酸解链酶）可以在与靶多核苷酸或其互补序列附接之前与前导序列结合。多核苷酸结合蛋白（例如，多核苷酸解链酶）可以与双链多核苷酸中存在的前导序列结合。与前导序列结合的多核苷酸结合蛋白（例如，多核苷酸解链酶）的活性可以停滞，直到多核苷酸与跨膜孔接触。停滞多核苷酸结合蛋白（例如，多核苷酸解链酶）的方法是本领域已知的，例如在WO 2014/135838中已知。

[0228] 衔接子

[0229] 前导序列可以存在于衔接子中，其中所述衔接子包含双链区（例如，双链体茎）和至少一个单链区。单链区中的至少一个可以是前导序列。衔接子可以包含至少一个非多核苷酸区域。附接于靶双链多核苷酸的两个末端的衔接子可以相同或不同。优选地，该对中的衔接子是相同的。

[0230] 前导序列优选存在于衔接子的一条链的5'末端（或3'末端）的第一单链区中。第二单链区可以存在于衔接子另一条链的3'末端（或5'末端）。第一和第二单链区不互补。在该实施方案中，衔接子可以称为Y衔接子。

[0231] Y衔接子通常包含(a)双链区（例如，双链体茎）和(b)单链区或在另一端不互补的区域。如果Y衔接子包含单链区，则可以将其描述为具有突出端。Y衔接子中非互补区的存在赋予了衔接子Y形状，因为两条链通常不彼此杂交，不同于双链部分。Y衔接子可以包含一个或多个锚。

[0232] 在一些实施方案中，所述衔接子可包含一个或多个（例如，至少一个、至少两个、至少三个或更多个）用于纳米孔上的一个或多个（例如，至少一个、至少两个、至少三个或更多个）标签的结合位点。在一些实施方案中，用于纳米孔上的标签的结合位点可以是在双链区（例如，双链体茎）内，使得在双链区的两条链分离时结合位点被暴露。参见，例如，图10。另外或可替代地，用于纳米孔上的标签的结合位点可以是在衔接子的单链部分上。仅作为示例，图9A显示了包含至少一个用于固体基质（例如，膜或珠粒）的锚的示例衔接子，而图15显示了包含至少两个锚的示例衔接子，其中第一锚能够偶联到固体基质，例如，膜或珠粒，并且第二锚能够偶联到纳米孔。纳米孔的第二锚可以构造成与缀合至纳米孔的标签结合。

[0233] Y衔接子包含优先旋入孔中的前导序列。

[0234] 使用所属领域中已知的方法，将Y衔接子附接于聚核苷酸。例如，可以使用连接酶，如T4 DNA连接酶、大肠杆菌DNA连接酶、Taq DNA连接酶、Tma DNA连接酶以及9°N DNA连接酶，来连接衔接子中的一个或两个。

[0235] 在一个优选实施方案中，修饰双链多核苷酸，例如样品中的双链多核苷酸，使得它

们在两端包含Y衔接子。可以使用任何修饰方式。该方法可包括通过添加衔接子来修饰双链靶多核苷酸。

[0236] 通过使多核苷酸与MuA转座酶和一群双链MuA底物接触,可以向双链多核苷酸提供衔接子,如Y衔接子或锚。转座酶将双链多核苷酸片段化并且将MuA底物连接到片段的一端或两端。这样就产生了多个修饰后的双链多核苷酸,其包含衔接子或锚。然后可以使用本发明的方法研究修饰后的双链多核苷酸。

[0237] 这些基于MuA的方法公开于WO 2015/022544和WO 2016/059363中。还在WO2015/150786中详细论述了这些方法。

[0238] 所述衔接子还可包含锚,以将包含靶多核苷酸和/或其互补序列的双链多核苷酸拴系到跨膜孔或包含孔的膜上,即衔接子还可包含膜-系链或孔-系链。锚优选附接于不是前导序列的单链区。

[0239] 多核苷酸结合蛋白(例如,多核苷酸解链酶)可以与衔接子中的前导序列结合,或者可以在衔接子附接于双链多核苷酸后添加多核苷酸结合蛋白(例如,多核苷酸解链酶)。与前导序列结合的多核苷酸结合蛋白(例如,多核苷酸解链酶)的活性可以停滞,直到多核苷酸与跨膜孔接触。

[0240] 前导序列或衔接子可以通过任何合适的方法与双链多核苷酸附接。例如,前导序列可以与靶多核苷酸和/或其互补序列连接,或者衔接子可以与双链多核苷酸连接。

[0241] 在一些实施方案中,双链条形码序列可以连接到靶双链多核苷酸的一端或两端。可以在添加前导序列或衔接子之前将条形码序列添加至双链多核苷酸。例如,条形码序列可位于靶双链多核苷酸的末端和衔接子之间。优选地,条形码序列包含在衔接子中。

[0242] 可以将唯一的条形码序列附接,例如链接到样品中的每个双链多核苷酸。条形码序列可用于鉴定对应于靶多核苷酸和与靶多核苷酸互补的多核苷酸连续易位通过孔的信号。

[0243] 在一些实施方案中,本文所述的衔接子可包含一个或多个间隔区以防止预先结合的多核苷酸结合蛋白(例如,多核苷酸解链酶)沿着双链多核苷酸移动并且使双链多核苷酸解链。这些间隔区阻止多核苷酸结合蛋白(例如,多核苷酸解链酶)的进一步移动,直到多核苷酸结合蛋白(例如,多核苷酸解链酶)位于孔处并且在孔上施加电位差。由电位差提供的附加力将多核苷酸结合蛋白(例如,多核苷酸解链酶)推过间隔区并且使其解链并控制多核苷酸通过纳米孔的移动。因此,多核苷酸结合蛋白(例如,多核苷酸解链酶)的移动通常仅在多核苷酸位于纳米孔中时而不是之前发生。用于防止预先结合的多核苷酸结合蛋白(例如,多核苷酸解链酶)沿双链多核苷酸移动并且使双链多核苷酸解链直至多核苷酸处于纳米孔中的间隔区和方法的实例在例如WO2015/110813中有描述,其内容通过引用整体并入本文。

[0244] 条形码

[0245] 多核苷酸条形码是本领域中众所周知的(Kozarewa, I. 等人, (2011),《分子生物学方法》(Methods Mol. Biol.) 733, 第279-298页)。条形码是多核苷酸的特定序列,其以特定且已知的方式影响流过孔的电流。

[0246] 条形码可包含核苷酸序列。核苷酸通常含有核碱基、糖以及至少一个磷酸基。核碱基通常是杂环的。核碱基包括但不限于嘌呤和嘧啶,以及更具体地,腺嘌呤、鸟嘌呤、胸腺嘧啶、尿嘧啶以及胞嘧啶。糖通常是戊糖。核苷酸糖包括但不限于核糖和脱氧核糖。核苷酸通

常是核糖核苷酸或脱氧核糖核苷酸。核苷酸通常含有单磷酸、二磷酸或三磷酸。磷酸可以连接在核苷酸的5'或3'侧上。

[0247] 核苷酸包括但不限于：单磷酸腺苷(AMP)、二磷酸腺苷(ADP)、三磷酸腺苷(ATP)、单磷酸鸟苷(GMP)、二磷酸鸟苷(GDP)、三磷酸鸟苷(GTP)、单磷酸胸苷(TMP)、二磷酸胸苷(TDP)、三磷酸胸苷(TTP)、单磷酸尿苷(UMP)、二磷酸尿苷(UDP)、三磷酸尿苷(UTP)、单磷酸胞苷(CMP)、二磷酸胞苷(CDP)、三磷酸胞苷(CTP)、5-甲基胞苷单磷酸、5-甲基胞苷二磷酸、5-甲基胞苷三磷酸、5-羟甲基胞苷单磷酸、5-羟甲基胞苷二磷酸、5-羟甲基胞苷三磷酸、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、单磷酸脱氧腺苷(dAMP)、二磷酸脱氧腺苷(dADP)、三磷酸脱氧腺苷(dATP)、单磷酸脱氧鸟苷(dGMP)、二磷酸脱氧鸟苷(dGDP)、三磷酸脱氧鸟苷(dGTP)、单磷酸脱氧胸苷(dTMP)、二磷酸脱氧胸苷(dTDP)、三磷酸脱氧胸苷(dTTP)、单磷酸脱氧尿苷(dUMP)、二磷酸脱氧尿苷(dUDP)、三磷酸脱氧尿苷(dUTP)、单磷酸脱氧胞苷(dCMP)、二磷酸脱氧胞苷(dCDP)和三磷酸脱氧胞苷(dCTP)、5-甲基-2'-脱氧胞苷单磷酸、5-甲基-2'-脱氧胞苷二磷酸、5-甲基-2'-脱氧胞苷三磷酸、5-羟甲基-2'-脱氧胞苷单磷酸、5-羟甲基-2'-脱氧胞苷二磷酸以及5-羟甲基-2'-脱氧胞苷三磷酸。衔接子中的核苷酸优选地选自AMP、TMP、GMP、UMP、dAMP、dTDP、dGMP或dCMP。核苷酸可以是无碱基的(即缺乏核碱基)。核苷酸可以含有额外修饰。具体来说，合适的修饰后核苷酸包括但不限于2'氨基嘧啶(如2'-氨基胞苷和2'-氨基尿苷)、2'-羟基嘌呤(如2'-氟嘧啶(如2'-氟胞苷和2'-氟尿苷)、羟基嘧啶(如5'- α -P-硼烷尿苷)、2'-0-甲基核苷酸(如2'-0-甲基腺苷、2'-0-甲基鸟苷、2'-0-甲基胞苷和2'-0-甲基尿苷)、4'-硫代嘧啶(如4'-硫代尿苷和4'-硫代胞苷)，并且核苷酸具有对核碱基的修饰(如5-戊炔基-2'-脱氧尿苷、5-(3-氨丙基)-尿苷以及1,6-二氨基己基-N-5-氨基甲酰基甲基尿苷)。

[0248] 条形码可以包含一种或多种不同的核苷酸种类。例如，T k聚体(即其中中心核苷酸是基于胸腺嘧啶的k聚体，如TTA、GTC、GTG和CTA)通常具有最低的电流状态。可以将修饰版本的T核苷酸引入修饰后的多核苷酸中以进一步降低电流状态，并且从而增加当条形码移动通过孔时所见的总电流范围。

[0249] G k聚体(即中心核苷酸是基于鸟嘌呤的k聚体，例如TGA、GGC、TGT和CGA)往往会受到k聚体中其它核苷酸的强烈影响，并且因此对修饰后的多核苷酸中的G核苷酸进行修饰可以帮助它们具有更独立的当前位置。

[0250] 包括同一核苷酸种类而不是三个不同种类的三个拷贝可以促进表征，因为那样仅需要定位修饰后的多核苷酸中的例如3-核苷酸k聚体。但是，这种修饰确实减少了条形码提供的信息。

[0251] 条形码中可包括一个或多个无碱基核苷酸。使用一个或多个无碱基核苷酸导致特征性电流尖峰。这允许清楚地突出条形码中一个或多个核苷酸种类的位置。

[0252] 条形码中的核苷酸种类可以包含化学原子或基团，例如丙炔基、巯基、氧化基、甲基、羟甲基、甲酰基、羧基、羰基、苯甲基、炔丙基或炔丙胺基。化学基团或原子可以是或可以包含荧光分子、生物素、地高辛(digoxigenin)、二硝基苯酚(DNP)、光不稳定基团、炔烃、DBCO、叠氮化物、游离氨基、氧化还原染料、汞原子或硒原子。

[0253] 条形码可以包含含有卤素原子的核苷酸种类。卤素原子可以连接到不同核苷酸种类，例如核碱基和/或糖上的任何位置。卤素原子优选地是氟(F)、氯(Cl)、溴(Br)或碘(I)。

卤素原子最优先是F或I。

[0254] 锚

[0255] 靶多核苷酸可以使用锚(膜-系链)与膜偶联。可以使用一个或多个锚来将靶多核苷酸与膜偶联。通常，一个或多个锚附接于靶多核苷酸的每条链。锚可以是衔接子的一部分。

[0256] 如果膜是如三嵌段共聚物膜等两亲层，则所述一个或多个锚优选包含可以被插入膜中的多肽锚和/或疏水性锚。疏水性锚优选是脂质、脂肪酸、固醇、碳纳米管、多肽、蛋白质或氨基酸，例如胆固醇、棕榈酸盐或生育酚。在优选实施例中，所述一个或多个锚不是孔。

[0257] 膜组分，如两亲分子、共聚物或脂质，可以经过化学修饰或功能化而形成所述一个或多个锚。下文更详细地论述了合适的化学修饰和使膜组分功能化的合适方式的实例。可以对任何比例的膜组分进行功能化，例如至少0.01%、至少0.1%、至少1%、至少10%、至少25%、至少50%或100%。

[0258] 一个或多个锚优选包含接头。一个或多个锚可以包含一个或多个，例如2、3、4个或更多个接头。

[0259] 优选的接头包括但不限于聚合物，如多核苷酸、聚乙二醇(PEG)、多糖以及多肽。这些接头可以是线性的、支化的或环状的。例如，接头可以是环状多核苷酸。多核苷酸可以与环状多核苷酸接头上的互补序列杂交。

[0260] 一个或多个锚或一个或多个接头可以包含能被切割或分解的组分，如限制位点或光不稳定基团。

[0261] 功能化的接头和其与分子的偶联方式是本领域中已知的。例如，用马来酰亚胺基团功能化的接头将与蛋白质中的半胱氨酸残基反应并且与之连接。

[0262] 可以使用“锁和钥”布置来避免多核苷酸的交联。每个接头只有一端可以一起反应形成更长的接头，并且接头的另一端各自分别与多核苷酸或膜反应。这种接头描述于W0 2010/086602中。

[0263] 在本发明的测序方法中使用接头是优选的。如果多核苷酸是在它不会在与孔相互作用时解偶联的意义上被永久性地直接与膜偶联的，则因为该轮测序由于膜与孔之间的距离而不能持续到多核苷酸的末端，所以一些序列数据将丢失。如果使用接头，则多核苷酸可以得到完全处理。

[0264] 偶联可以是永久的或稳定的。换句话说，偶联可以使多核苷酸在与孔相互作用时保持与膜偶联。

[0265] 偶联可以是暂时的。换句话说，偶联可以使多核苷酸在与孔相互作用时可以与膜解耦。对于多核苷酸测序，暂时性质的偶联是优选的。如果永久性或稳定接头直接连接到多核苷酸的5'或3'端并且接头比膜与跨膜孔的通道之间的距离短，则因为该轮测序不能持续到多核苷酸的末端，所以一些序列数据将丢失。如果偶联是暂时的，则当偶联的末端随机摆脱膜时，多核苷酸就可以得到完全处理。形成永久性/稳定或暂时连接的化学基团在下文更详细地论述。可以使用胆固醇或脂肪酰基链，将靶多核苷酸和/或其互补序列暂时与膜，如两亲层，例如三嵌段共聚物膜或脂质膜偶联。可以使用长度为6至30个碳原子的任何脂肪酰基链，如十六烷酸。

[0266] 在优选实施方案中，锚将靶多核苷酸和/或其互补序列与两亲层，如三嵌段共聚物

膜或脂质双层偶联。先前已用各种不同的系链策略进行了核酸与合成脂质双层的偶联。这些汇总在下文表1中。

[0267] 表1

锚包含	偶联类型	参考文献
[0268]	硫醇;	稳定 Yoshina-Ishii, C. 和 S. G. Boxer (2003). “支撑脂质双层上的可移动拴系囊泡阵列(Arrays of mobile tethered vesicles on supported lipid bilayers).” 《美国化学会志》 (<i>J Am Chem Soc</i>) 125 (13): 3696-7。
	生物素	稳定 Nikolov, V., R. Lipowsky 等人 (2007). “具有锚定 DNA 分子的巨型囊泡的行为(Behavior of giant vesicles with anchored DNA molecules).” 《生物物理杂志》 (<i>Biophys J</i>) 92 (12): 4356-68
	胆固醇	暂时性的 Pfeiffer, I. 和 F. Hook (2004). “《寡核苷酸与脂膜组装件的基于二价胆固醇的偶联(Bivalent cholesterol-based coupling of oligonucleotides to lipid membrane assemblies)》”, 《美国化学协会期刊(J Am Chem Soc)》 126 (33): 10224-5
	表面活性剂(例如脂质、棕榈酸等)	稳定 van Lengerich, B., R. J. Rawle 等人“脂质囊泡与液体支撑双层的共价连接允许观测DNA介导的囊泡相互作用(Cooperative attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions).” 《朗缪尔》 (<i>Langmuir</i>) 26 (11): 8666-72

[0269] 合成多核苷酸和/或接头可以在合成反应中使用修饰后的亚磷酰胺来功能化,所述修饰后的亚磷酰胺容易与直接加入的合适的锚定基团相容,所述锚定基团例如胆固醇、生育酚、棕榈酸盐、硫醇、脂质以及生物素基团。这些不同的连接化学为多核苷酸的连接提供了一套选择。每个不同的修饰基团以稍微不同的方式偶联多核苷酸,并且偶联未必总是永久性的,因此赋予多核苷酸在膜上不同的停留时间。

[0270] 多核苷酸与接头或与功能化膜的偶联还可以通过多种其它手段来实现,条件是可以向多核苷酸中加入互补反应性基团或锚定基团。先前已经报道了向多核苷酸的任一末端加入反应性基团。可以使用T4多核苷酸激酶和ATP γ S向ssDNA或dsDNA的5'加入硫醇基(Grant, G.P. 和 P.Z. Qin (2007) . “一种用于在核酸的5' 端处连接氮氧化物自旋标记的简易方法(A facile method for attaching nitroxide spin labels at the 5' terminus of nucleic acids).”《核酸研究》35(10) :e77)。可以使用T4多核苷酸激酶和 γ - [2-叠氮基乙基]-ATP或 γ - [6-叠氮基己基]-ATP向ssDNA或dsDNA的5' -磷酸中加入叠氮基团。使用硫醇或点击化学,含有硫醇、碘乙酰胺OPSS或马来酰亚胺基团(对硫醇具有反应性)或二苯并环辛炔(DIBO)或炔基(对叠氮化物具有反应性)中的任一个的系链可以共价连接至多核苷酸。可以使用末端转移酶增加更多样的化学基团选择,如生物素、硫醇和荧光团,从而将修饰后的寡核苷酸掺入ssDNA的3' (Kumar, A., P. Tchen 等人 (1988) . “末端脱氧核苷酸转移酶对合

成寡核苷酸探针的非放射性标记(Nonradioactive labeling of synthetic oligonucleotide probes with terminal deoxynucleotidyl transferase).”《分析生物化学》(Anal Biochem) 169 (2) :376-82)。链霉亲和素/生物素和/或链霉亲和素/脱硫生物素偶联可以用于任何其它多核苷酸。也有可能,可以使用具有适当修饰的核苷酸(例如胆固醇或棕榈酸盐)的末端转移酶,直接向多核苷酸加入锚。

[0271] 所述一个或多个锚可以通过杂交将靶多核苷酸和/或其互补序列与膜偶联。杂交可以是在一个或多个锚与靶多核苷酸和/或其互补序列之间,在一个或多个锚内或在一个或多个锚与膜之间。如上文所论述的,在一个或多个锚中的杂交允许以暂时方式进行偶联。举例来说,接头可以包含两个或更多个杂交在一起的多核苷酸,如3、4或5个多核苷酸。所述一个或多个锚可以与靶多核苷酸或与靶多核苷酸互补的多核苷酸杂交。所述一个或多个锚可以直接与附接于靶多核苷酸和/或其互补序列的Y衔接子和/或前导序列杂交。可替代地,所述一个或多个锚可以与一个或多个,如2或3个中间多核苷酸(或“夹板”)杂交,所述中间多核苷酸与多核苷酸,与附接于靶多核苷酸和/或其互补序列的Y衔接子和/或前导序列杂交。

[0272] 所述一个或多个锚可以包含单链或双链多核苷酸。锚的一部分可以连接到单链或双链多核苷酸分析物上。已经报道过使用T4 RNA连接酶I来连接ssDNA的短片(Troutt, A.B., M.G. McHeyzer-Williams等人(1992).“连接锚定PCR:一种具有单侧特异性的简单扩增技术(Ligation-anchored PCR:a simple amplification technique with single-sided specificity).”《美国国家科学院院刊》(Proc Natl Acad Sci U S A) 89 (20) : 9823-5)。或者,可以将单链或双链多核苷酸连接到双链多核苷酸上,并且然后通过热或化学变性分离两条链。对于双链多核苷酸,可以将一片单链多核苷酸加入到双链体的一端或两端,或将双链多核苷酸加入到一端或两端。为了将单链多核苷酸加入到双链多核苷酸中,可以使用T4 RNA连接酶I来实现与单链多核苷酸的其它区域的连接。为了将双链多核苷酸加入到双链多核苷酸中,则连接可以是“平端的”,其中互补的3' dA/dT尾分别在多核苷酸和所加入的多核苷酸上(这是许多样品制备应用的常规做法,用以防止多联体或二聚体形成),或使用通过限制性消化多核苷酸和连接相容的衔接子生成的“粘末端”。接着,当双链体解链时,如果使用单链多核苷酸进行连接,则每条单链将具有5' 或3' 端修饰;或者如果使用双链多核苷酸进行连接,则每条单链将在5' 端、3' 端或两端具有修饰。

[0273] 如果衔接子或靶多核苷酸的互补序列是合成链,则可以在衔接子或互补序列化学合成期间掺入所述一个或多个锚。例如,可以使用有反应性基团附接到其上的引物来合成衔接子或互补序列。

[0274] 腺苷酸化多核苷酸是连接反应中的中间物,其中单磷酸腺苷附接至多核苷酸的5' -磷酸。可获得各种用于生成这种中间物的试剂盒,例如来自NEB的5' DNA腺苷酰化试剂盒。通过用反应中的ATP取代修饰后的三磷酸核苷酸,就可以向多核苷酸的5' 加入反应性基团(如硫醇、胺、生物素、叠氮化物等)。还可以使用具有适当修饰的核苷酸(例如胆固醇或棕榈酸盐)的5' DNA腺苷酰化试剂盒将锚直接加入到多核苷酸中。

[0275] 用于扩增基因组DNA区段的常见技术是使用聚合酶链式反应(PCR)。这里,使用两个合成的寡核苷酸引物,可以生成同一个DNA区段的许多拷贝,其中对于每个拷贝,双链体中每条链的5' 将是合成的多核苷酸。可以通过采用聚合酶将单个或多个核苷酸加入到单链

CN 116377044 B

或双链DNA的3' 端。可以使用的聚合酶的实例包括但不限于末端转移酶、Klenow以及大肠杆菌Poly (A) 聚合酶。通过用反应中的ATP取代修饰后的三磷酸核苷酸，就可以将锚，如胆固醇、硫醇、胺、叠氮化物、生物素或脂质，掺入到双链多核苷酸中。因此，所扩增多核苷酸的每个拷贝都将含有锚。

[0276] 理想地，多核苷酸与膜偶联而不必使多核苷酸功能化。这可以通过将一个或多个锚，如多核苷酸结合蛋白（例如，多核苷酸解链酶）或化学基团，与膜偶联并且使所述一个或多个锚与多核苷酸相互作用或通过使膜功能化来实现。所述一个或多个锚可以通过本文所述的任何方法与膜偶联。具体来说，所述一个或多个锚可以包含一个或多个接头，如马来酰亚胺功能化的接头。在这个实施例中，多核苷酸通常是RNA、DNA、PNA、TNA或LNA，并且可以是双链的或单链的。此实施例特别适合基因组DNA多核苷酸。

[0277] 所述一个或多个锚可以包含与单链或双链多核苷酸、多核苷酸内的特定核苷酸序列或多核苷酸内的修饰后的核苷酸的模式或多核苷酸上存在的任何其它配体偶联、结合或相互作用的任何基团。

[0278] 适合用于锚中的结合蛋白包括但不限于：大肠杆菌单链结合蛋白、P5单链结合蛋白、T4gp32单链结合蛋白、TOPO V dsDNA结合区、人类组蛋白蛋白质、大肠杆菌HU DNA结合蛋白，以及其他古细菌的、原核生物的或真核生物的单链或双链多核苷酸（或核酸）结合蛋白，包括下文所列的那些。

[0279] 特异性核苷酸序列可以是由转录因子、核糖体、核酸内切酶、拓扑异构酶（topoisomerase）或复制起始因子识别的序列。修饰后的核苷酸的模式可以是甲基化模式或损伤模式。

[0280] 所述一个或多个锚可以包含与多核苷酸偶联、结合、插入或相互作用的任何基团。基团可以通过静电、氢键或范德华（Van der Waals）相互作用而插入多核苷酸中或与多核苷酸相互作用。此类基团包括赖氨酸单体、聚赖氨酸（其将与ssDNA或dsDNA相互作用）、溴化乙锭（其插入dsDNA中）、通用碱基或通用核苷酸（其可以与任何多核苷酸杂交）以及锇络合物（其可以与甲基化碱基反应）。因此，多核苷酸可以使用连接至膜的一个或多个通用核苷酸而与膜偶联。每个通用核苷酸可以使用一个或多个接头与膜偶联。通用核苷酸优选包含以下核碱基中的一种：次黄嘌呤、4-硝基胞嘧啶、5-硝基胞嘧啶、6-硝基胞嘧啶、甲酰基胞嘧啶、3-硝基胞嘧啶、硝基咪唑、4-硝基咪唑、4-硝基苯并咪唑、5-硝基咪唑、4-氨基苯并咪唑或苯基（C6芳环）。通用核苷酸更优选包含以下核苷中的一种：2'-脱氧肌苷、肌苷、7-脱氮-2'-脱氧肌苷、7-脱氮-肌苷、2-氮杂-脱氧肌苷、2-氮杂-肌苷、2'-甲基肌苷、4-硝基胞嘧啶2'-脱氧核苷、4-硝基胞嘧啶核苷、5-硝基胞嘧啶2'-脱氧核苷、5-硝基胞嘧啶核苷、6-硝基胞嘧啶2'-脱氧核苷、6-硝基胞嘧啶核苷、3-硝基胞嘧啶2'-脱氧核苷、次黄嘌呤的非环糖类似物、硝基咪唑2'-脱氧核苷、硝基咪唑核苷、4-硝基咪唑2'-脱氧核苷、4-硝基咪唑核苷、4-硝基苯并咪唑2'-脱氧核苷、4-硝基苯并咪唑核苷、5-硝基胞嘧啶2'-脱氧核苷、5-硝基胞嘧啶核苷、4-氨基苯并咪唑2'-脱氧核苷、4-氨基苯并咪唑核苷、5-硝基胞嘧啶2'-脱氧核苷、5-硝基胞嘧啶核苷、4-氨基苯并咪唑2'-脱氧核苷、4-氨基苯并咪唑核苷、苯基C-核苷、苯基C-2'-脱氧核糖基核苷、2'-脱氧烟酰胺核苷、2'-脱氧异鸟苷、K-2'-脱氧核苷、P-2'-脱氧核糖以及胞嘧啶。通用核苷酸更优选包含2'-脱氧肌苷。通用核苷酸更优选是IMP或dIMP。通用核苷酸最优选是dPMP（2'-脱氧-P-核苷单磷酸）或dKMP（N6-甲氧基-2,6-二氨基嘌呤单磷酸）。

[0281] 一个或多个锚可以通过Hoogsteen氢键(其中两个核碱基通过氢键保持在一起)或反向Hoogsteen氢键(其中一个核碱基相对于另一个核碱基旋转180°)与多核苷酸偶联(或结合)。例如,一个或多个锚可以包含与多核苷酸形成Hoogsteen氢键或反向Hoogsteen氢键的一个或多个核苷酸、一个或多个寡核苷酸或一个或多个多核苷酸。这些类型的氢键允许第三多核苷酸链缠绕双链螺旋体并形成三链体。一个或多个锚可以通过与双链双链体形成三链体而与双链多核苷酸偶联(或结合)。

[0282] 在此实施例中,可以使至少1%、至少10%、至少25%、至少50%或100%的膜组分功能化。

[0283] 当一个或多个锚包含蛋白质时,所述一个或多个锚可能能够直接锚定至膜而无需进一步功能化,例如当其已具有与膜相容的外部疏水区时。此类蛋白质的实例包括但不限于跨膜蛋白、膜内蛋白以及膜蛋白。或者,蛋白质可以表达成具有与膜相容的基因融合的疏水区。这种疏水性蛋白质区域是本领域中已知的。

[0284] 一个或多个锚优选在递送到膜之前与多核苷酸混合,但是一个或多个锚可以与膜接触并且随后与多核苷酸接触。

[0285] 另一方面,可以使用上文所述的方法使多核苷酸功能化,使得其可以被特异性结合基团识别。具体来说,多核苷酸可以用配体进行功能化,所述配体例如生物素(用于与链霉亲和素结合)、直链淀粉(用于与麦芽糖结合蛋白或融合蛋白结合)、Ni-NTA(用于与聚组氨酸或聚组氨酸标记的蛋白质结合)或肽(如抗原)。

[0286] 根据一个优选实施例,一个或多个锚可以用于在多核苷酸与前导序列连接时将多核苷酸与膜偶联,所述前导序列优选旋入孔中。优选地,多核苷酸与优先旋入孔中的前导序列附接(attach)(如连接(ligate))。这种前导序列可以包含同聚多核苷酸或无碱基区。前导序列通常设计成直接与一个或多个锚杂交,或通过一个或多个中间多核苷酸(或夹板)与所述一个或多个锚杂交。在这些情况下,一个或多个锚通常包含与前导序列中的序列或一个或多个中间多核苷酸(或夹板)中的序列互补的多核苷酸序列。在这些情况下,一个或多个夹板通常包含与前导序列中的序列互补的多核苷酸序列。

[0287] 上文论述的任何用于使多核苷酸与例如两亲层的膜偶联的方法当然可以应用于其它多核苷酸和膜组合。在一些实施例中,氨基酸、肽、多肽或蛋白质与例如三嵌段共聚物层或脂质双层的两亲层偶联。可以获得各种用于这类多核苷酸的化学连接的方法。化学连接中使用的分子的实例是EDC(1-乙基-3-[3-二甲氨基丙基]碳化二亚胺盐酸盐)。还可以使用市售试剂盒(Thermo Pierce,产品号22980)向多核苷酸的5'中加入反应性基团。合适的方法包括但不限于使用组氨酸残基和Ni-NTA的暂时亲和力附接,以及通过反应性半胱氨酸、赖氨酸或非天然氨基酸的更稳健的共价附接。

[0288] 微粒

[0289] 微粒(通常为珠粒)可用于将分析物(例如,多核苷酸或多肽)递送至跨膜孔。这在WO 2016/059375中有描述,其内容通过引用整体并入本文。可以在本发明的方法中使用任何数量的微粒。例如,本发明的方法可以使用单个微粒或2个、3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、50个、100个、1,000个、5,000个、10,000个、100,000个、500,000个或1,000,000个或更多个微粒。如果使用两个或更多个微粒,则微粒可以是相同的。或者,可以使用不同微粒的混合物。

[0290] 每个微粒可以附接有一个分析物(例如,多核苷酸或多肽)。可替代地,每个微粒可以附接有两个或更多个分析物(例如,多核苷酸或多肽),如3个或更多个、4个或更多个、5个或更多个、6个或更多个、7个或更多个、8个或更多个、9个或更多个、10个或更多个、20个或更多个、30个或更多个、50个或更多个、100个或更多个、500个或更多个、1,000个或更多个、5,000个或更多个、10,000个或更多个、100,000个或更多个、1,000,000个或更多个、或5000,000个或更多个分析物(例如,多核苷酸或多肽)。微粒可以基本上或完全被分析物(例如,多核苷酸或多肽)涂布或覆盖。微粒可以在其基本上全部或全部表面上附接有分析物(例如,多核苷酸或多肽)。微粒可以通过衔接子附接于分析物(例如,多核苷酸或多肽)。衔接子可以是Y衔接子,例如,如图36所示。

[0291] 合适的结合部分的实例包括:蛋白质结合标签(链霉抗生物素蛋白标签、flag标签等)、缀合附接(多核苷酸、聚合物、生物素、肽)以及氨基酸(半胱氨酸、Faz等)。

[0292] 在一些实施方案中,多核苷酸可以附接于两个或更多个微粒。

[0293] 微粒是微观颗粒,其大小通常以微米(μm)为单位测量。微粒也可以被称为微球或微珠。微粒可以是纳米颗粒。纳米颗粒是微观颗粒,其大小通常以纳米(nm)为单位测量。

[0294] 微粒通常具有约 $0.001\mu\text{m}$ 至约 $500\mu\text{m}$ 的粒径。例如,纳米颗粒可以具有约 $0.01\mu\text{m}$ 至约 $200\mu\text{m}$ 或约 $0.1\mu\text{m}$ 至约 $100\mu\text{m}$ 的粒径。更常见的是,微粒具有约 $0.5\mu\text{m}$ 至约 $100\mu\text{m}$,或例如约 $1\mu\text{m}$ 至约 $50\mu\text{m}$ 的粒径。微粒可以具有约 1nm 至约 1000nm ,如约 10nm 至约 500nm 、约 20nm 至约 200nm 或约 30nm 至约 100nm 的粒径。

[0295] 微粒可以是球形或非球形。球形微粒可以被称为微球。非球形颗粒可以是例如板状、针状、不规则状或管状。如本文所使用的术语“粒径”意指颗粒的直径,如果颗粒是球形的,或者如果颗粒是非球形的,则意指基于体积的粒径。基于体积的粒径是与所论述的非球形颗粒具有相同体积的球体的直径。

[0296] 如果在所述方法中使用两个或更多个微粒,则微粒的平均粒径可以是上文所论述的任何大小,如约 $0.5\mu\text{m}$ 至约 $500\mu\text{m}$ 。两个或更多个微粒的群体优选具有 10% 或更小,如 5% 或更小或 2% 或更小的变异系数(标准差与平均值的比率)。

[0297] 可以使用任何方法来确定微粒的大小。合适的方法包括但不限于流式细胞术(参见例如Chandler等人,《血栓形成与止血杂志》(J Thromb Haemost.) 2011年6月;9 (6) : 1216-24)。

[0298] 微粒可以由任何材料形成。微粒优选由陶瓷、玻璃、二氧化硅、聚合物或金属形成。聚合物可以是天然聚合物,如聚羟基烷酸酯、葡聚糖、聚丙交酯、琼脂糖、纤维素、淀粉或壳聚糖,或合成聚合物,如聚氨酯、聚苯乙烯、聚(氯乙烯)、硅烷或甲基丙烯酸酯。合适的微粒是本领域中已知的并且是可商购的。陶瓷和玻璃微球可购自3M®。二氧化硅和聚合物微粒可购自EPRUI Nanoparticles&Microspheres Co.Ltd.。微粒也可以购自Polysciences Inc.、Bangs Laboratories Inc.和Life Technologies。

[0299] 微粒可以是固体。微粒可以是中空的。微粒可以由聚合物纤维形成。

[0300] 微粒可以衍生自用于提取和分离多核苷酸的试剂盒。

[0301] 微粒表面可以与分析物相互作用并附接分析物。所述表面可自然地与分析物如多核苷酸或多肽相互作用,不需要功能化。通常将微粒表面功能化以促进分析物的附接。合适的功能化是本领域已知的。例如,微粒表面可以用以下各者进行功能化:多组氨酸标签(六

组氨酸标签、6xHis-标签、His6标签或His-tag®)、Ni-NTA、链霉亲和素、生物素、寡核苷酸、多核苷酸(如DNA、RNA、PNA、GNA、TNA或LNA)、羧基、季胺基、硫醇基、叠氮基、炔基、DIBO、脂质、FLAG-标签(FLAG八肽、多核苷酸结合蛋白(例如,多核苷酸解链酶)(包括下文所论述的任何一种)、肽、蛋白质、抗体或抗体片段。微粒还可以用下文所论述的任何接头或基团进行功能化。

[0302] 微粒可以用与多核苷酸特异性地结合的分子或基团进行功能化。在这种情况下,将要与微粒附接并被递送到跨膜孔的多核苷酸可以被称为靶多核苷酸。这允许微粒从含有其它多核苷酸的样品中选择或捕获靶多核苷酸。如果分子或基团优先地或以高亲和力与靶多核苷酸结合,但不与其它多核苷酸或不同的多核苷酸结合或仅以低亲和力结合,则所述分子或基团与靶多核苷酸特异性地结合。如果分子或基团以 1×10^{-6} M或更小、更优选地 1×10^{-7} M或更小、 5×10^{-8} M或更小、更优选地 1×10^{-8} M或更小或更优选地 5×10^{-9} M或更小的Kd结合,则所述分子或基团优先地或以高亲和力结合。如果分子或基团以 1×10^{-6} M或更大、更优选地 1×10^{-5} M或更大、更优选地 1×10^{-4} M或更大、更优选地 1×10^{-3} M或更大、甚至更优选地 1×10^{-2} M或更大的Kd结合,则所述分子或基团以低亲和力结合。

[0303] 优选地,所述分子或基团与靶多核苷酸结合的亲和力是其对其它多核苷酸的亲和力的至少10倍,如至少50倍、至少100倍、至少200倍、至少300倍、至少400倍、至少500倍、至少1000倍或至少10,000倍。亲和力可以使用已知的结合分析来测量,如利用荧光和放射性同位素的结合分析。竞争性结合分析也是本领域中已知的。肽或蛋白质与多核苷酸之间的结合强度可以使用纳米孔力光谱来测量,如Hornblower等人,《自然·方法》(Nature Methods.) 4:315-317. (2007) 中所述。

[0304] 微粒可以用与靶多核苷酸或衔接子特异性地杂交或包含与靶多核苷酸或衔接子的一部分或区域互补的部分或区域的寡核苷酸或多核苷酸进行功能化。这允许微粒从含有其它多核苷酸的样品中选择或捕获靶多核苷酸。

[0305] 当寡核苷酸或多核苷酸优先地或以高亲和力与靶多核苷酸杂交,但与其它多核苷酸基本上不杂交、不杂交或仅以低亲和力杂交时,所述寡核苷酸或多核苷酸与靶多核苷酸特异性地杂交。如果寡核苷酸或多核苷酸与靶多核苷酸以解链温度(T_m)比其对其它序列的 T_m 高出至少2°C,如至少3°C、至少4°C、至少5°C、至少6°C、至少7°C、至少8°C、至少9°C或至少10°C杂交,则所述寡核苷酸或多核苷酸会发生特异性杂交。更优选地,寡核苷酸或多核苷酸与靶多核苷酸以 T_m 比其对其它核酸的 T_m 高出至少2°C,如至少3°C、至少4°C、至少5°C、至少6°C、至少7°C、至少8°C、至少9°C、至少10°C、至少20°C、至少30°C或至少40°C杂交。优选地,寡核苷酸或多核苷酸与靶多核苷酸以 T_m 比其对与靶多核苷酸相差一个或多个核苷酸,如1个、2个、3个、4个或5个或更多个核苷酸的序列的 T_m 高出至少2°C,如至少3°C、至少4°C、至少5°C、至少6°C、至少7°C、至少8°C、至少9°C、至少10°C、至少20°C、至少30°C或至少40°C杂交。寡核苷酸或多核苷酸通常与靶多核苷酸以至少90°C,如至少92°C或至少95°C的 T_m 杂交。 T_m 可以使用已知技术通过实验测量,包括使用DNA微阵列,或者可以使用公众可获得的 T_m 计算器计算,如可通过互联网获得的计算器。

[0306] 允许杂交的条件是本领域中众所周知的(例如,Sambrook等人,2001,《分子克隆:实验室手册》(Molecular Cloning:a laboratory manual),第3版,冷泉港实验室出版社(Cold Spring Harbor Laboratory Press);和《当代分子生物学实验手册》(Current

Protocols in Molecular Biology), 第2章, Ausubel等人编, 格林出版与威利交叉科学出版社(Greene Publishing and Wiley-Interscience), 纽约(1995)。杂交可以在低严格度条件下进行, 例如在37°C下在30%至35%甲酰胺、1M NaCl以及1%十二烷基硫酸钠(SDS)的缓冲溶液存在下, 然后在50°C下在1X(0.1650M Na⁺)至2X(0.33M Na⁺)标准柠檬酸钠(SSC)中洗涤20次。杂交可以在中等严格度条件下进行, 例如在37°C下在40%至45%甲酰胺、1M NaCl以及1% SDS的缓冲溶液存在下, 然后在55°C下在0.5X(0.0825M Na⁺)至1X(0.1650M Na⁺)SSC中洗涤。杂交可以在高严格度条件下进行, 例如在37°C下在50%甲酰胺、1M NaCl以及1% SDS的缓冲溶液存在下, 然后在60°C下在0.1X(0.0165M Na⁺)SSC中洗涤。

[0307] 多核苷酸可以包括与靶多核苷酸的部分或区域基本上互补的部分或区域。因此, 与靶多核苷酸中的部分或区域相比, 多核苷酸的区域或部分可以在5个、10个、15个、20个、21个、22个、30个、40个或50个核苷酸的区上具有1个、2个、3个、4个、5个、6个、7个、8个、9个、10个或更多个错配。

[0308] 区域的一部分通常为50个核苷酸或更少, 如40个核苷酸或更少、30个核苷酸或更少、20个核苷酸或更少、10个核苷酸或更少或5个核苷酸或更少。

[0309] 微粒优选是顺磁性或磁性的。微粒优选包含顺磁性或超顺磁性材料或顺磁性或超顺磁性金属, 如铁。可以使用任何合适的磁性微粒。例如, 可以使用可购自例如Clontech、Promega、Invitrogen ThermoFisher Scientific以及NEB的磁珠。在一些实施例中, 微粒包含连接着有机基团的磁性颗粒, 有机基团例如金属螯合基团, 如次氨基三乙酸(NTA)。有机组分可以例如包含选自以下的基团: -C(=O)-O-、-C-O-C-、-C(=O)-、-NH-、-C(=O)-NH、-C(=O)-CH₂-I、-S(=O)₂-以及-S-。有机组分可以包含金属螯合基团, 如次氨基三乙酸(NTA)。通常, 如金、铁、镍或钴等金属也连接在金属螯合基团上。这种磁珠通常用于捕获His标记的蛋白质, 但也适用于本发明。

[0310] 微粒最优选是可购自Life Technologies的His-Tag Dynabead®、来自IBA的Mag Strept珠粒、来自NEB的链霉亲和素磁珠、来自Beckman Coulter的固相可逆固定化(Solid Phase Reversible Immobilization, SPRI)珠粒或Agencourt AMPure XP珠粒或Dynabeads ®MyOne™链霉亲和素C1(ThermoFisher Scientific)。

[0311] 多核苷酸结合蛋白(例如, 多核苷酸解链酶)

[0312] 多核苷酸结合蛋白(例如, 多核苷酸解链酶)可以是能够与多核苷酸结合并且控制其通过孔的移动的任何蛋白质。在本领域中确定蛋白质是否与多核苷酸结合是很简单的。蛋白质通常与多核苷酸相互作用并且修饰其至少一种性质。蛋白质可以通过裂解多核苷酸以形成单独的核苷酸或如二核苷酸或三核苷酸的更短的核苷酸链来修饰多核苷酸。所述部分可通过将多核苷酸定位或移动到特异性位置(即控制其移动)来修饰多核苷酸。

[0313] 多核苷酸结合蛋白(例如, 多核苷酸解链酶)优选衍生自多核苷酸处理酶。多核苷酸处理酶是能够与多核苷酸相互作用并且修饰其至少一种性质的多肽。所述酶可以通过裂解多核苷酸以形成单独的核苷酸或如二核苷酸或三核苷酸的更短的核苷酸链来修饰多核苷酸。所述酶可以通过将多核苷酸定向或移动到特定位置来修饰多核苷酸。多核苷酸处理酶并不需要展现酶活性, 只要其能够结合多核苷酸并且控制其通过孔的移动即可。例如, 酶可以经过修饰以去除其酶活性或可以在防止其起酶作用的条件下使用。下文更详细地论述了这种条件。

[0314] 多核苷酸处理酶优选衍生自溶核酶。在酶的构建体中使用的多核苷酸处理酶更优选衍生自以下酶分类(Enzyme Classification, EC)组中的任何组的成员:3.1.11、3.1.13、3.1.14、3.1.15、3.1.16、3.1.21、3.1.22、3.1.25、3.1.26、3.1.27、3.1.30和3.1.31。所述酶可以是WO 2010/086603中所公开的任何酶。

[0315] 优选的酶是聚合酶、解旋酶、易位酶以及拓扑异构酶,如旋转酶。聚合酶可以是PyroPhage®3173DNA聚合酶(其可购自Lucigen®公司)、SD聚合酶(可购自Bioron®)或其变体。聚合酶优选是Phi29 DNA聚合酶或其变体。拓扑异构酶优选是部分分类(EC)组5.99.1.2和5.99.1.3中的任一个的成员。

[0316] 所述酶最优选衍生自解旋酶。解旋酶可以是或衍生自He1308解旋酶;RecD解旋酶,如TraI解旋酶或TrwC解旋酶;XPD解旋酶或Dda解旋酶。解旋酶可以是或衍生自He1308 Mbu、He1308 Csy He1308 Tga、He1308 Mhu、TraI Eco、XPD Mbu或其变体。

[0317] 解旋酶可以是WO 2013/057495、WO 2013/098562、W02013098561、WO 2014/013260、WO 2014/013259、WO 2014/013262以及WO/2015/055981中所公开的任何解旋酶、修饰后的解旋酶或解旋酶构建体。

[0318] Dda解旋酶优选包含WO/2015/055981和WO 2016/055777中所公开的任何修饰。

[0319] 根据本发明可以使用任何数量的解旋酶。例如,可以使用1、2、3、4、5、6、7、8、9、10个或更多个解旋酶。在一些实施例中,可以使用不同数量的解旋酶。可以使用上述两个或更多个解旋酶的任何组合。两个或更多个解旋酶可以是两个或更多个Dda解旋酶。两个或更多个解旋酶可以是一个或多个Dda解旋酶和一个或多个TrwC解旋酶。两个或更多个解旋酶可以是同一种解旋酶的不同变体。

[0320] 两个或更多个解旋酶优选彼此附接。两个或更多个解旋酶更优选彼此共价附接。解旋酶可以按任何次序并且使用任何方法附接。用于本发明的优选解旋酶构建体描述于WO 2014/013260、WO 2014/013259、WO 2014/013262和W02015/055981中。

[0321] 在一些实施方案中,多核苷酸结合蛋白是多核苷酸解链酶。多核苷酸解链酶是能够使双链多核苷酸解链成单链的酶。在一些实施方案中,多核苷酸解链酶能够使双链DNA解链成单链。在一些实施方案中,多核苷酸解链酶是具有解旋酶活性的酶。多核苷酸解链酶的实例包括例如本文所述的解旋酶。

[0322] 多核苷酸结合能力可以使用本领域中已知的任何方法来测量。例如,可以使蛋白质与多核苷酸接触,并且可以测量蛋白质与多核苷酸结合并沿所述多核苷酸移动的能力。蛋白质可以包括有助于多核苷酸结合和/或有助于其在高盐浓度和/或室温下的活性的修饰。蛋白质可进行修饰,使得其结合多核苷酸(即保持多核苷酸结合能力)但不充当解链酶(即当具备所有便于移动的必需组分(例如ATP和Mg²⁺)时不沿多核苷酸移动)。此类修饰是所属领域中已知的。例如,解旋酶中的Mg²⁺结合结构域的修饰通常产生不起解旋酶作用的变体。这些变体类型可以充当分子刹车。

[0323] 酶可以共价附接至孔。可以使用任何方法将酶共价附接至孔。

[0324] 在链测序中,多核苷酸顺着或逆着外加电位易位通过孔。在双链多核苷酸上逐渐或逐步起作用的核酸外切酶可以在孔的顺侧使用以外加电位下供给剩余的单链,或在反侧使用以在反向电位下供给剩余的单链。同样,还可以以类似的方式使用使双链DNA解链的解旋酶。还可以使用聚合酶。需要逆着外加电位发生链易位的测序应用也是有可能的,但是

DNA必须首先在相反电位或无电位下被酶“捕获”。随着电位随后在结合后转换，链将以顺式到反式的方式通过孔并且通过电流保持呈延长的构型。单链DNA核酸外切酶或单链DNA依赖性聚合酶可充当分子马达，以将最近易位的单链以受控的逐步方式逆着外加电位从反式到顺式从孔中拉回来。

[0325] 可以在本发明中使用任何解旋酶。解旋酶可以两种模式对孔起作用。首先，优选使用解旋酶来进行所述方法，使得在由外加电压造成的场的作用下，所述解旋酶使多核苷酸移动通过孔。在这种模式中，多核苷酸的5' 端首先被捕获在孔中，并且解旋酶使多核苷酸移动到孔中，使其在场的作用下通过孔，直到其最终易位通过，到达膜的反侧为止。或者，优选这样进行所述方法，解旋酶使多核苷酸逆着由外加电压造成的场而移动通过孔。在这种模式中，多核苷酸的3' 端首先被捕获在孔中，并且解旋酶使多核苷酸移动通过孔，使得其逆着外加场被拉出孔，直到其最终被推回到膜的顺侧为止。

[0326] 还可以沿相反的方向进行所述方法。多核苷酸的3' 端可以首先被捕获在孔中，并且解旋酶可以使多核苷酸移动到孔中，使得其在场的作用下通过孔，直到其最终易位通过，到达膜的反侧为止。

[0327] 当解链酶不具备便于移动的必需组分或被修饰成阻止或防止其移动时，所述解链酶可以与多核苷酸结合并且在多核苷酸被外加场拉入孔中时充当刹车以减慢多核苷酸的移动。在非活性模式中，多核苷酸的3' 或5' 是否被捕获不重要，在充当刹车的酶的作用下将多核苷酸朝向反侧拉入孔中的是外加场。当在非活性模式中时，解链酶对多核苷酸的移动的控制可以用多种方式来描述，包括齿合、滑动和制动。还可以用这种方式来使用缺乏解链酶活性的解链酶变体。

[0328] 多核苷酸与多核苷酸结合蛋白(例如，多核苷酸解链酶)和孔可以按任何次序接触。优选的是，当使多核苷酸与如解旋酶的多核苷酸结合蛋白(例如，多核苷酸解链酶)和孔接触时，多核苷酸首先与多核苷酸结合蛋白(例如，多核苷酸解链酶)形成复合物。当跨孔施加电压时，多核苷酸/多核苷酸结合蛋白(例如，多核苷酸解链酶)复合物就会与孔形成复合物并且控制多核苷酸通过孔的移动。

[0329] 使用多核苷酸结合蛋白(例如，多核苷酸解链酶)的方法中的任何步骤通常在游离核苷酸或游离核苷酸类似物和促进多核苷酸结合蛋白(例如，多核苷酸解链酶)的作用的酶辅因子存在下进行。游离核苷酸可以是上文论述的任何单独的核苷酸中的一种或多种。游离核苷酸包括但不限于：单磷酸腺苷(AMP)、二磷酸腺苷(ADP)、三磷酸腺苷(ATP)、单磷酸鸟苷(GMP)、二磷酸鸟苷(GDP)、三磷酸鸟苷(GTP)、单磷酸胸苷(TMP)、二磷酸胸苷(TDP)、三磷酸胸苷(TTP)、单磷酸尿苷(UMP)、二磷酸尿苷(UDP)、三磷酸尿苷(UTP)、单磷酸胞苷(CMP)、二磷酸胞苷(CDP)、三磷酸胞苷(CTP)、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、单磷酸脱氧腺苷(dAMP)、二磷酸脱氧腺苷(dADP)、三磷酸脱氧腺苷(dATP)、单磷酸脱氧鸟苷(dGMP)、二磷酸脱氧鸟苷(dGDP)、三磷酸脱氧鸟苷(dGTP)、单磷酸脱氧胸苷(dTMP)、二磷酸脱氧胸苷(dTDP)、三磷酸脱氧胸苷(dTTP)、单磷酸脱氧尿苷(dUMP)、二磷酸脱氧尿苷(dUDP)、三磷酸脱氧尿苷(dUTP)、单磷酸脱氧胞苷(dCMP)、二磷酸脱氧胞苷(dCDP)以及三磷酸脱氧胞苷(dCTP)。游离核苷酸优选选自AMP、TMP、GMP、CMP、UMP、dAMP、dTDP、dGMP或dCMP。游离核苷酸优选是三磷酸腺苷(ATP)。酶辅因子是允许构建体起作用的因子。酶辅因子优选是二价金属阳离子。二价金属阳离子优选是Mg²⁺、Mn²⁺、Ca²⁺或Co²⁺。酶辅因子最优选是Mg²⁺。

[0330] 分子刹车可以是与多核苷酸结合并且减慢多核苷酸通过孔的移动的任何化合物或分子。分子刹车可以是上文论述的任何一种。分子刹车优选包含与多核苷酸结合的化合物。所述化合物优选是大环。合适的大环包括但不限于环糊精、杯芳烃、环肽、冠醚、葫芦脲、柱芳烃、其衍生物或其组合。环糊精或其衍生物可以是Eliseev, A.V. 和 Schneider, H-J. (1994)《美国化学会志》116, 6081-6088中所公开的任何环糊精或其衍生物。环糊精更优选是七-6-氨基- β -环糊精($\text{am}_7\text{-}\beta\text{CD}$)、6-单脱氧-6-单氨基- β -环糊精($\text{am}_1\text{-}\beta\text{CD}$)或七-(6-脱氧-6-氨基)-环糊精($\text{gu}_7\text{-}\beta\text{CD}$)。

[0331] 膜

[0332] 可根据本文所述的各个方面使用任何膜。合适的膜是本领域熟知的。膜优选是两亲层或固态层。

[0333] 两亲层是由两亲分子形成的层，所述两亲分子如磷脂，其具有亲水性和亲脂性两种特性。两亲分子可以是合成的或天然存在的。非天然存在的两亲物和形成单层的两亲物在所属领域中是已知的，且包括例如嵌段共聚物(Gonzalez-Perez等人，《朗缪尔(Langmuir)》，2009, 25, 10447-10450)。嵌段共聚物是聚合在一起的两个或更多个单体子单元产生单一聚合物链的聚合材料。嵌段共聚物通常具有通过每个单体亚基贡献的性质。然而，嵌段共聚物可以具有由单独的亚单元形成的聚合物所不具备的独特性质。嵌段共聚物可以被工程化为使得单体亚基之一在水性介质中是疏水性的(即亲脂性)，而一个或多个其它亚基是亲水性的。在这种情况下，嵌段共聚物可以具备两亲性质，并且可以形成模拟生物膜的结构。嵌段共聚物可以是二嵌段(由两个单体亚单元组成)，但也可以由多于两个单体亚单元构建以形成表现为两亲物的更复杂布置。共聚物可以是三嵌段、四嵌段或五嵌段共聚物。膜优选是三嵌段共聚物膜。

[0334] 古细菌双极性四醚脂质是天然存在的脂质，其被构建成使得脂质形成单层膜。这些脂质一般发现于在苛刻生物环境中存活的嗜极生物、嗜热生物、嗜盐生物和嗜酸生物中。认为其稳定性是源于最终双层的融合性质。直接了当的做法是，通过产生具有一般基序亲水性-疏水性-亲水性的三嵌段聚合物来构建模拟这些生物实体的嵌段共聚物材料。这种材料可形成表现类似于脂质双层并且涵盖从囊泡到层状膜的一系列阶段表现的单体膜。由这些三嵌段共聚物形成的膜相对于生物脂质膜保持若干优势。因为三嵌段共聚物是合成的，所以可小心地控制准确的构建，以提供形成膜并与孔和其它蛋白质相互作用所需的正确链长度和特性。

[0335] 还可以由不被归类为脂质亚材料的亚单元来构建嵌段共聚物，例如可由硅氧烷或其它非基于烃的单体来制成疏水性聚合物。嵌段共聚物的亲水性亚区段还可以具备低蛋白质结合特性，这允许产生当暴露于原始生物样品时具有高度抗性的膜。这种头基单元还可来源于非经典的脂质头基。

[0336] 与生物脂质膜进行比较，三嵌段共聚物膜还具有增加的机械和环境稳定性，例如高许多的操作温度或pH范围。嵌段共聚物的合成性质提供定制用于广泛范围应用的基于聚合物的膜的平台。

[0337] 膜最优选是W02014/064443或W02014/064444中所公开的膜中的一种。

[0338] 两亲分子可进行化学修饰或官能化，以便于偶联多核苷酸。

[0339] 两亲性层可以是单层或双层。两亲层通常是平面的。两亲层可以是弯曲的。两亲层

可以是支撑式的。两亲层可以是凹入的。两亲层可以从凸起的柱子上悬挂下来,使得两亲层的周边区域(其与柱子连接)高于两亲层区域。这可以允许微粒如上文所描述的沿着膜行进、移动、滑动或滚动。

[0340] 两亲膜通常天然地是流动的,基本上以大致 10^{-8}cm s^{-1} 的脂质扩散速率充当二维液体。这意味着孔和偶联的多核苷酸可通常在两亲膜内移动。

[0341] 膜可以是脂质双层。脂质双层是细胞膜的模型,并且用作一系列实验研究的极佳平台。例如,脂质双层可以用于通过单通道记录对膜蛋白进行体外研究。或者,脂质双层可以用作检测一系列物质的存在的生物传感器。脂质双层可以是任何脂质双层。适当的脂质双层包括但不限于平面脂质双层、支撑双层或脂质体。脂质双层优选是平面脂质双层。合适的脂质双层公开于WO 2008/102121、WO 2009/077734和WO 2006/100484中。

[0342] 用于形成脂质双层的方法是本领域中已知的。脂质双层通常通过Montal和Mueller(《美国国家科学院院刊》,1972;69:3561-3566)的方法形成,其中脂质单层携载于通过开孔的任一侧的水溶液/空气界面上,所述开孔垂直于所述界面。通常通过首先将脂质溶解在有机溶剂中,并且然后使一滴溶剂在开孔的任一侧上的水溶液的表面上蒸发,来将脂质加入到电解质水溶液的表面。一旦有机溶剂已蒸发,则开孔任一侧上的溶液/空气界面物理地上下移动通过开孔,直到形成双层。平面脂质双层可以穿过膜中的开孔或穿过凹槽中的开口形成。

[0343] Montal和Mueller的方法很受欢迎,因为它是一种经济有效且相对简单的形成适合于蛋白孔插入的优质脂质双层的方法。其它常见的双层形成方法包括脂质体双层的尖端浸渍、涂刷双层和膜片钳。

[0344] 尖端浸渍双层形成法需要使开孔表面(例如移液管尖端)接触到携载脂质单层的测试溶液的表面。同样,首先通过将溶解于有机溶剂中的一滴脂质在溶液表面蒸发而在溶液/空气界面处产生脂质单层。然后通过朗缪尔-沙佛(Langmuir-Schaefer)过程形成双层,并且所述双层需要机械自动化以使开孔相对于溶液表面移动。

[0345] 对于涂刷的双层,将溶解于有机溶剂中的一滴脂质直接施加至开孔,所述开孔浸没在水性测试溶液中。使用画刷或等同物将脂质溶液薄薄地铺展在孔上。溶剂的薄化导致脂质双层的形成。然而,难以从双层完全去除溶剂,并且因此通过这种方法形成的双层在电化学测量期间不太稳定并且更容易产生噪声。

[0346] 膜片钳常用于生物细胞膜的研究中。通过抽吸将细胞膜夹在移液管的末端,并将膜片贴在开孔上。所述方法已经适用于通过夹持脂质体来产生脂质双层,然后脂质体破裂以留下密封在移液管开孔上的脂质双层。所述方法需要稳定的、巨大的和单层的脂质体以及在具有玻璃表面的材料中制造小开孔。

[0347] 脂质体可以通过超声处理、挤出或Mozafari方法(Colas等人(2007)《微米》(Micron)38:841-847)来形成。

[0348] 在一个优选的实施例中,脂质双层如WO 2009/077734中所述形成。在此方法中有利的是,脂质双层是由干燥的脂质形成的。在一个最优选的实施例中,脂质双层跨开口形成,如WO 2009/077734中所描述。

[0349] 脂质双层由相对的两层脂质形成。两个脂质层被布置成使得其疏水性尾基面朝彼此以形成疏水性内部。脂质的亲水性头基向外朝向双层各侧上的水环境。双层可以存在于

许多脂质相中,包括但不限于液体无序相(流体层状)、液体有序相、固体有序相(层状凝胶相、交叉凝胶相)以及平面双层晶体(层状亚凝胶相、层状结晶相)。

[0350] 可以使用形成脂质双层的任何脂质组合物。对脂质组合物加以选择,以便形成具有所需特性的脂质双层,所述特性例如表面电荷、支撑膜蛋白的能力、堆积密度或机械特性。脂质组合物可以包含一种或多种不同脂质。例如,脂质组合物可以含有多达100种脂质。脂质组合物优选含有1至10种脂质。脂质组合物可以包含天然存在的脂质和/或人工脂质。

[0351] 脂质通常包含头基、界面部分和可以相同或不同的两个疏水性尾基。合适的头基包括但不限于:中性头基,如二酰基甘油酯(DG)和神经酰胺(CM);两性离子头基,如磷脂酰胆碱(PC)、磷脂酰乙醇胺(PE)和鞘磷脂(SM);带负电头基,如磷脂酰甘油(PG);磷脂酰丝氨酸(PS)、磷脂酰肌醇(PI)、磷脂酸(PA)和心磷脂(CA);以及带正电头基,如三甲基铵-丙烷(TAP)。合适的界面部分包括但不限于天然存在的界面部分,如基于甘油或基于神经酰胺的部分。合适的疏水性尾基包括但不限于:饱和烃链,如月桂酸(正十二烷酸)、肉豆蔻酸(正十四烷酸)、棕榈酸(正十六烷酸)、硬脂酸(正十八烷酸)和花生酸(正二十烷酸);不饱和烃链,如油酸(顺-9-十八烷酸);以及支链烃链,如植烷酰基。不饱和烃链中链的长度和双键的位置与数量可以变化。支链烃链中链的长度和支链,如甲基的位置和数量可以变化。疏水性尾基可以作为醚或酯连接于界面部分。脂质可以是分枝菌酸。

[0352] 脂质还可以进行化学修饰。脂质的头基或尾基可以进行化学修饰。头基已进行化学修饰的合适脂质包括但不限于:PEG修饰的脂质,如1,2-二酰基-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000];功能化的PEG脂质,如1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺-N-[生物素基(聚乙二醇)2000];以及针对缀合修饰的脂质,如1,2-二油酰基-sn-甘油-3-磷酸乙醇胺-N-(琥珀酰基)和1,2-二棕榈酰基-sn-甘油-3-磷酸乙醇胺-N-(生物素基)。尾基已进行化学修饰的合适脂质包括但不限于:可聚合脂质,如1,2-双(10,12-二十三碳二炔基)-sn-甘油-3-磷酸胆碱;氟化脂质,如1-软脂酰基-2-(16-氟软脂酰基)-sn-甘油-3-磷酸胆碱;氘化脂质,如1,2-二棕榈酰基-D62-sn-甘油-3-磷酸胆碱;以及醚连接的脂质,如1,2-二-0-植烷基-sn-甘油-3-磷酸胆碱。脂质可以进行化学修饰或官能化以便于偶联多核苷酸。

[0353] 两亲层,例如脂质组合物,通常包括将影响层的性质的一个或多个添加剂。合适的添加剂包括但不限于:脂肪酸,如棕榈酸、肉豆蔻酸和油酸;脂肪醇,如棕榈醇、肉豆蔻醇和油醇;甾醇,如胆固醇、麦角固醇、羊毛甾醇、谷甾醇和豆甾醇;溶血磷脂,如1-酰基-2-羟基-sn-甘油-3-磷酸胆碱;以及神经酰胺。

[0354] 固态层可以由有机材料和无机材料形成,所述材料包括但不限于:微电子材料;绝缘材料,如 Si_3N_4 、 Al_2O_3 和 SiO ;有机和无机聚合物,如聚酰胺;塑料,如Teflon®,或弹性体,如双组分加成固化的硅橡胶;以及玻璃。固态层可以由石墨烯形成。合适的石墨烯层公开在W0 2009/035647中。Yusko等人,《自然·纳米技术》(Nature Nanotechnology),2011;6:253-260和美国专利申请第2013/0048499号描述了在不使用微粒的情况下将蛋白质递送至固态层中的跨膜孔。本发明的方法可用于改进这些文献中公开的方法中的递送。

[0355] 通常使用以下来进行所述方法:(i)包含孔的人工两亲层;(ii)包含孔的分离的天然存在的脂质双层;或(iii)插入有孔的细胞。通常使用人工两亲层,如人工三嵌段共聚物层来进行所述方法。所述层可以包含其它跨膜蛋白和/或膜内蛋白以及除孔以外的其它分

子。下文论述了合适的设备和条件。通常在体外进行本发明的方法。

[0356] 根据本发明的方法将多核苷酸递送到其上的膜包含在液体中。液体使膜保持“湿润”并防止其变干。液体通常是水溶液。水溶液通常具有与水相同的密度。水溶液的密度通常是约 $1\text{g}/\text{cm}^3$ 。溶液的密度可以根据温度和溶液的确切组成而变化。水溶液的密度通常在约0.97与约 $1.03\text{g}/\text{cm}^3$ 之间。

[0357] 膜通常将两个体积的水溶液隔开。膜抵挡所述体积之间的电流流动。插入膜中的跨膜孔选择性地允许离子通过膜，这可以记录为由两个体积的水溶液中的电极检测到的电信号。包含靶多核苷酸的复合物的存在可调节离子的流动，并通过观察所产生的电信号变化来检测。

[0358] 阵列

[0359] 膜通常是膜阵列的一部分，其中每个膜优选包含跨膜孔。因此，本发明提供了一种使用膜阵列检测靶多核苷酸的方法。

[0360] 膜可以被包含在具有电隔离膜阵列的设备中，每个膜使用其自己的电极单独寻址，使得阵列等同于从测试样品平行测量的许多单独传感器。膜可以相对密集地填充，允许大量的膜用于给定体积的测试样品。在本领域中，例如在WO 2009/077734和WO2012/042226中描述了合适的膜阵列和设备。例如，WO 2009/077734公开了在微孔孔口阵列上形成的多个可单独寻址的脂质双层，每个微孔含有电极和与脂质双层接触的水性介质。

[0361] 所述设备通常以‘即用’状态提供给最终用户，其中膜和跨膜孔是预先插入的。以‘即用’状态提供的典型设备包含两亲膜阵列，每个膜包含跨膜孔并且跨含有液体的孔设置。WO2014/064443公开了这样的设备和其制造方法。将待分析的测试液体施加到两亲膜的上表面。

[0362] 然而，以‘即用’状态提供的设备还需要考虑另外的因素，即传感器不会变干，即液体不会通过两亲膜而从孔中损失，那样的话会导致性能损失或损坏传感器。解决传感器变干问题的一种解决方案是在两亲膜的表面上设置含缓冲液的装置，使得通过膜表面的任何蒸发最小化，并且在膜的任一侧提供的液体可以具有相同的离子强度以减少任何渗透效应。在使用中，可以从两亲膜表面去除缓冲液，并引入待分析的测试液体以接触表面。

[0363] 一些应用可以使用跨膜电特性的测量，例如离子电流。为了提供这样的测量，所述设备可以进一步在每个隔室中包含相应的电极，使其与包含极性介质的体积电接触。可以进行其它类型的测量，例如光学测量，例如荧光测量和FET测量。光学测量和电测量可以同时进行(Heron AJ等人，《美国化学会志》2009;131(5):1652-3)。

[0364] 所述设备可以进一步包含公共电极。所述设备可以进一步包含连接在公共电极与每个隔室中的相应电极之间的电路，布置所述电路是为了进行电测量。这种电测量可以取决于在膜处或穿过膜发生的过程。

[0365] 所述设备可以包含FET阵列来测量纳米孔阵列。

[0366] 跨膜孔

[0367] 跨膜孔是在某种程度上穿过膜的结构。它允许由施加的电位驱动的水合离子流过膜或膜内。跨膜孔通常穿过整个膜，使得水合离子可以从膜的一侧流到膜的另一侧。然而，跨膜孔不是必须得穿过膜。它的一端可以封闭。例如，孔可以是膜中的阱、间隙、通道、沟槽或狭缝，水合离子可以沿着它流动或流入其中。

[0368] 本发明中可以使用任何跨膜孔。孔可以是生物的或人造的。合适的孔包括但不限于蛋白质孔、多核苷酸孔和固态孔。孔可以是DNA折纸孔 (Langecker等人,《科学》(Science),2012;338:932-936)。孔可以是马达蛋白纳米孔,例如,允许双链多核苷酸易位的纳米孔。在一些实施方案中,马达蛋白纳米孔能够使双链多核苷酸解链。示例性马达蛋白纳米孔包括但不限于phi29马达蛋白纳米孔,例如,如Wendell等人“双链DNA易位通过膜适应的phi29马达蛋白纳米孔(Translocation of double-stranded DNA through membrane-adapted phi29 motor protein nanopores)”《自然纳米技术》(Nat Nanotechnol),4(2009),第765-772页所述。在一些实施方案中,如Feng等人“基于纳米孔的第四代DNA测序技术(Nanopore-based fourth-generation DNA sequencing technology)”《基因组学、蛋白质组学和生物信息学》(Genomics, Proteomics & Bioinformatics) (2015) 第13卷,第1期,第4-16页所述或引用的任何纳米孔,均可用于本文所述的各个方面。

[0369] 跨膜孔优选是跨膜蛋白孔。跨膜蛋白孔是多肽或多肽的集合,其允许如多核苷酸的水合离子从膜的一侧流到膜的另一侧。在本发明中,跨膜蛋白孔能够形成孔,所述孔允许由外加电位驱动的水合离子从膜的一侧流到另一侧。跨膜蛋白孔优选地允许多核苷酸从膜,如三嵌段共聚物膜的一侧流到另一侧。跨膜蛋白孔允许如DNA或RNA的多核苷酸移动通过孔。

[0370] 跨膜蛋白孔可以是单体或低聚物。孔优选由几个重复亚单元组成,如至少6个、至少7个、至少8个、至少9个、至少10个、至少11个、至少12个、至少13个、至少14个、至少15个或至少16个亚单元。孔优选是六聚体、七聚体、八聚体或九聚体孔。孔可以是同型低聚物或异型低聚物。

[0371] 跨膜蛋白孔通常包含可以让离子流过的筒体或通道。孔的亚单元通常围绕中心轴线并且为跨膜筒体或通道或跨膜 α -螺旋束或通道提供链。

[0372] 跨膜蛋白孔的筒体或通道通常包含促进与核苷酸、多核苷酸或核酸相互作用的氨基酸。这些氨基酸优选地位于筒体或通道的收缩部附近。跨膜蛋白孔通常包含一个或多个带正电荷的氨基酸,如精氨酸、赖氨酸或组氨酸,或芳香族氨基酸,如酪氨酸或色氨酸。这些氨基酸通常促进孔与核苷酸、多核苷酸或核酸之间的相互作用。

[0373] 根据本发明使用的跨膜蛋白孔可以衍生自 β -筒形孔或 α -螺旋束孔。 β -筒体孔包括由 β -链形成的筒体或通道。合适的 β -筒形孔包括但不限于 β -毒素,如 α -溶血素、炭疽毒素和杀白细胞素,以及细菌的外膜蛋白/孔蛋白,如耻垢分枝杆菌 (*Mycobacterium smegmatis*) 孔蛋白 (Msp),例如MspA、MspB、MspC或MspD、CsgG,外膜孔蛋白F (OmpF)、外膜孔蛋白G (OmpG)、外膜磷脂酶A以及奈瑟氏球菌 (*Neisseria*) 自主转运蛋白脂蛋白 (NalP) 以及其他孔,如胞溶素 (lysenin)。 α -螺旋束孔包括由 α -螺旋形成的筒体或通道。合适的 α -螺旋束孔包括但不限于内膜蛋白和 α 外膜蛋白,如WZA和ClyA毒素。

[0374] 跨膜孔可以衍生自或基于Msp、 α -溶血素 (α -HL)、胞溶素、CsgG、ClyA、Sp1以及溶血蛋白溶血毒素 (fragaceatoxin) C (FraC)。跨膜蛋白孔优选衍生自CsgG,更优选衍生自大肠杆菌菌株K-12亚株MC4100的CsgG。WO 2016/034591中公开了衍生自CsgG的合适的孔。跨膜孔可以衍生自胞溶素。WO 2013/153359中公开了衍生自胞溶素的合适的孔。

[0375] 野生型 α -溶血素孔由7个相同的单体或亚单元形成(即,它是七聚体)。 α -溶血素-

NN的一个单体或亚单元的序列公开在例如WO2016/059375中。

[0376] 跨膜蛋白孔优选衍生自Msp,更优选衍生自MspA。WO 2012/107778中公开了衍生自MspA的合适的孔。

[0377] 可以例如通过加入组氨酸残基(his标签)、天冬氨酸残基(asp标签)、链霉亲和素标签、flag标签、SUMO标签、GST标签或MBP标签,或在多肽天然地不含信号序列的情况下通过加入这样的序列以促进其从细胞分泌,来修饰本文所述的任何蛋白质,如跨膜蛋白孔,从而帮助其鉴定或纯化。引入基因标签的替代方案是将标签化学反应到孔或构建体上的天然的或工程化的位置上。这种操作的实例是使凝胶迁移试剂与孔外部的工程化半胱氨酸反应。这已被证明是一种用于分离溶血素异型低聚物的方法(《生物化学》(Chem Biol).1997年7月;4(7):497-505)。

[0378] 孔可以用显露标记来进行标记。显露标记可以是允许孔被检测到的任何合适的标记。合适的标记包括但不限于荧光分子;放射性同位素,例如¹²⁵I、³⁵S;酶;抗体;抗原;多核苷酸;以及配体,如生物素。

[0379] 本文所描述的任何蛋白质,如跨膜蛋白孔,可以合成方式或通过重组手段制得。例如,可以通过体外翻译和转录(IVTT)来合成孔。孔的氨基酸序列可以经过修饰而包括非天然存在的氨基酸或增加蛋白质的稳定性。当通过合成手段制造蛋白质时,可以在制造期间引入此类氨基酸。还可以在合成或重组制造后改变孔。

[0380] 本文所述的任何蛋白质,如跨膜蛋白孔,可以使用本领域中已知的标准方法制造。编码孔或构建体的多核苷酸序列可以使用本领域中的标准方法衍生和复制。编码孔或构建体的多核苷酸序列可以使用本领域中的标准技术在细菌宿主细胞中表达。可以通过从重组表达载体原位表达多肽而在细胞中产生孔。表达载体任选地携带诱导型启动子以控制多肽的表达。这些方法描述于Sambrook,J.和Russell,D.(2001).《分子克隆:实验室手册》,第3版.纽约冷泉港的冷泉港实验室出版社中。

[0381] 可以在通过任何蛋白质液相色谱系统纯化之后或在重组表达之后,从产蛋白质生物体大规模产生孔。典型的蛋白质液相色谱系统包括FPLC、AKTA系统、Bio-Cad系统、Bio-Rad BioLogic系统以及Gilson HPLC系统。

[0382] 测量离子流

[0383] 可以使用电测量和/或光学测量来监测通过跨膜孔的离子流。

[0384] 电测量可以是电流测量、阻抗测量、隧穿测量或场效应晶体管(field effect transistor,FET)测量。

[0385] 可以检测当多肽易位通过跨膜孔时通过所述孔的离子流的变化作为电流、电阻或光学性质的变化。所测量的效应可以是穿过跨膜孔的电子隧穿。所测量的效应可以是由于多核苷酸与跨膜孔的相互作用引起的电位改变,其中在FET测量中使用局部电位传感器监测效应。

[0386] 可以进行各种不同类型的测量。这包括但不限于:电测量和光学测量。《美国化学会志》2009,131 1652-1653公开了一种涉及荧光测量的合适的光学方法。可能的电测量包括:电流测量、阻抗测量、隧穿测量(Ivanov AP等人,《纳米快报》2011年1月12日;11(1):279-85)以及FET测量(国际申请WO 2005/124888)。光学测量可以与电测量结合(Soni GV等人,《科学仪器综述》(Rev Sci Instrum.)2010年1月;81(1):014301)。测量可以是跨膜电流

测量,例如对流过孔的离子电流的测量。

[0387] 可以使用标准单通道记录设备来进行电测量,如以下文献中所述:

[0388] Stoddart D等人,《美国国家科学院院刊》,12;106(19):7702-7;Lieberman KR等人,《美国化学会志》2010;132(50):17961-72;以及国际申请WO 2000/28312。或者,可以使用多通道系统来进行电测量,所述系统例如,如WO2009/077734和WO 2011/067559中所述。

[0389] 所述方法优选在跨膜施加电位的情况下进行。外加电位可以是电压电位。或者,外加电位可以是化学势。在一些实施方案中,施加的电位可以由渗透失衡驱动。化学势的实例是跨膜,如跨两亲层使用盐梯度。Holden等人,《美国化学会志》2007年7月11日;129(27):8650-5中公开了盐梯度。在一些情况下,使用在多核苷酸相对于孔移动时通过孔的电流来评估或确定多核苷酸的序列。

[0390] 多核苷酸表征

[0391] 在本文所述各个方面的一些实施方案中,该方法可以包括进一步表征靶多核苷酸。在靶多核苷酸与孔接触时,随着多核苷酸相对于孔移动进行一种或多种测量,其指示靶多核苷酸的一个或多个特征。

[0392] 该方法可以包括确定多核苷酸是否被修饰。可以测量任何修饰的存在或不存在。所述方法优选包含:用一个或多个蛋白质或用一个或多个标记、标签或间隔区确定多核苷酸是否通过甲基化、通过氧化、通过损坏来修饰。特定的修饰将引起与孔的特定的相互作用,这可以使用下文所描述的方法测量。例如,可以在孔与每个核苷酸相互作用期间通过孔的离子流的基础上区别甲基胞嘧啶和胞嘧啶。

[0393] 设备

[0394] 可以使用适合于研究膜/孔系统的任何设备来进行所述方法,在所述系统中,孔存在于膜中。可以使用适合于跨膜孔感测的任何设备来进行所述方法。例如,设备包含腔室,所述腔室包含水溶液和将腔室分成两个部分的屏障。屏障通常具有孔口,在其中形成含有孔的膜。或者,屏障形成其中存在孔的膜。

[0395] 可以使用在WO 2008/102120中所述的设备来进行所述方法。

[0396] 可以进行各种不同类型的测量。这包括但不限于:电测量和光学测量。《美国化学会志》2009,131 1652-1653公开了一种合适的光学方法,其涉及荧光测量。可能的电测量包括:电流测量、阻抗测量、隧穿测量(Ivanov AP等人,《纳米快报》(Nano Lett.)2011年1月12日;11(1):279-85)以及FET测量(国际申请WO 2005/124888)。光学测量可以与电测量结合(Soni GV等人,《科学仪器综述》(Rev Sci Instrum.)2010年1月;81(1):014301)。测量可以是跨膜电流测量,例如对流过孔的离子电流的测量。

[0397] 可以使用标准单通道记录设备来进行电测量,如以下文献中所述:Stoddart D等人,《美国国家科学院院刊》,12;106(19):7702-7;Lieberman KR等人,《美国化学会志》2010;132(50):17961-72;以及国际申请WO 2000/28312。可替代地,可以使用多通道系统进行电测量,例如如国际申请WO2009/077734和国际申请WO 2011/067559中所述。

[0398] 所述方法优选在跨膜施加电位的情况下进行。外加电位可以是电压电位。或者,外加电位可以是化学势。化学势的实例是跨膜,如跨两亲层使用盐梯度。Holden等人,《美国化学会志》2007年7月11日;129(27):8650-5中公开了盐梯度。在一些情况下,使用在多核苷酸相对于孔移动时通过孔的电流来评估或确定多核苷酸的序列。

[0399] 所述方法可以涉及测量在多核苷酸相对于孔移动时通过孔的电流。因此，所述设备还可以包含能够跨膜和孔施加电位并测量电信号的电路。可以使用膜片钳或电压钳来进行所述方法。所述方法优选地涉及使用电压钳。

[0400] 本发明的方法可以涉及在多核苷酸相对于孔移动时测量穿过孔的电流。用于测量通过跨膜蛋白孔的离子电流的适当条件在本领域中是已知的并且在实施例中公开。所述方法通常在跨膜和孔施加电压的情况下进行。所使用的电压通常是+5V到-5V，如+4V到-4V、+3V到-3V或+2V到-2V。所使用的电压通常是-600mV到+600mV或-400mV到+400mV。所使用的电压优选地处于具有下限和上限的范围内，所述下限选自-400mV、-300mV、-200mV、-150mV、-100mV、-50mV、-20mV和0mV，并且所述上限独立地选自+10mV、+20mV、+50mV、+100mV、+150mV、+200mV、+300mV和+400mV。所使用的电压更优选地处于100mV到240mV的范围内，并且最优选地处于120mV到220mV的范围内。通过使用增加的外加电位，可以增加孔对不同核苷酸的区分。

[0401] 通常在存在任何电荷载流子的情况下进行所述方法，所述电荷载流子如金属盐，例如碱金属盐；卤盐，例如氯化物盐，如碱金属氯化物盐。电荷载流子可以包括离子液体或有机盐，例如四甲基氯化铵、三甲基苯基氯化铵、苯基三甲基氯化铵或1-乙基-3-甲基氯化咪唑鎓。在上文论述的示例性设备中，盐存在于腔室中的水溶液中。通常使用氯化钾(KCl)、氯化钠(NaCl)、氯化铯(CsCl)，或亚铁氰化钾和铁氰化钾的混合物。KCl、NaCl以及亚铁氰化钾和铁氰化钾的混合物是优选的。电荷载流子可以是跨膜不对称的。例如，电荷载流子的类型和/或浓度可以在膜的各侧上不同。

[0402] 盐浓度可以是饱和的。盐浓度可以是3M或更低，并且通常是0.1至2.5M、0.3至1.9M、0.5至1.8M、0.7至1.7M、0.9至1.6M或1M至1.4M。盐浓度优选是150mM至1M。所述方法优选使用至少0.3M的盐浓度进行，例如至少0.4M、至少0.5M、至少0.6M、至少0.8M、至少1.0M、至少1.5M、至少2.0M、至少2.5M或至少3.0M。高盐浓度提供高信噪比并允许相对于正常电流波动的背景鉴定指示核苷酸的存在的电流。

[0403] 通常在存在缓冲液的情况下进行所述方法。在上文论述的示例性设备中，缓冲液存在于腔室中的水溶液中。在本发明的方法中可以使用任何缓冲液。通常，缓冲液是磷酸盐缓冲液。其它合适的缓冲液是HEPES和Tris-HCl缓冲液。通常在以下pH下进行所述方法：4.0至12.0、4.5至10.0、5.0至9.0、5.5至8.8、6.0至8.7或7.0至8.8或7.5至8.5。所用pH优选是约7.5。

[0404] 可以在以下温度下进行所述方法：0°C至100°C、15°C至95°C、16°C至90°C、17°C至85°C、18°C至80°C、19°C至70°C或20°C至60°C。通常在室温下进行所述方法。任选地在支持酶功能的温度下进行所述方法，如在约37°C下进行。

[0405] 游离核苷酸和辅因子

[0406] 通常在存在游离核苷酸或游离核苷酸类似物和/或促进多核苷酸结合蛋白的作用的酶辅因子的情况下进行所述方法。所述方法也可以在不存在游离核苷酸或游离核苷酸类似物并且不存在酶辅因子的情况下进行。游离核苷酸可以是上文论述的任何单独的核苷酸中的一种或多种。游离核苷酸包括但不限于：单磷酸腺苷(AMP)、二磷酸腺苷(ADP)、三磷酸腺苷(ATP)、单磷酸鸟苷(GMP)、二磷酸鸟苷(GDP)、三磷酸鸟苷(GTP)、单磷酸胸苷(TMP)、二磷酸胸苷(TDP)、三磷酸胸苷(TTP)、单磷酸尿苷(UMP)、二磷酸尿苷(UDP)、三磷酸尿苷

(UTP)、单磷酸胞昔(CMP)、二磷酸胞昔(CDP)、三磷酸胞昔(CTP)、环单磷酸腺苷(cAMP)、环单磷酸鸟昔(cGMP)、单磷酸脱氧腺苷(dAMP)、二磷酸脱氧腺苷(dADP)、三磷酸脱氧腺苷(dATP)、单磷酸脱氧鸟昔(dGMP)、二磷酸脱氧鸟昔(dGDP)、三磷酸脱氧鸟昔(dGTP)、单磷酸脱氧胸昔(dTMP)、二磷酸脱氧胸昔(dTDP)、三磷酸脱氧胸昔(dTTP)、单磷酸脱氧尿昔(dUMP)、二磷酸脱氧尿昔(dUDP)、三磷酸脱氧尿昔(dUTP)、单磷酸脱氧胞昔(dCMP)、二磷酸脱氧胞昔(dCDP)以及三磷酸脱氧胞昔(dCTP)。游离核昔酸优选选自AMP、TMP、GMP、CMP、UMP、dAMP、dTDP、dGMP或dCMP。游离核昔酸优选是三磷酸腺昔(ATP)。酶辅因子是允许多核昔酸结合蛋白(例如,多核昔酸解链酶)起作用的因子。酶辅因子优选是二价金属阳离子。二价金属阳离子优选是Mg²⁺、Mn²⁺、Ca²⁺或Co²⁺。酶辅因子最优选是Mg²⁺。

[0407] 试剂盒

[0408] 本发明还提供了衔接子群体,其包含双链条形码序列、单链前导序列和能够分离双链多核昔酸的链并控制多核昔酸通过跨膜孔的运动的多核昔酸结合蛋白(例如,多核昔酸解链酶),其中所述群体中的每个衔接子中的条形码序列是唯一的。

[0409] 本发明还提供了用于本发明方法的试剂盒。该试剂盒通常包含根据本发明的衔接子群体。该试剂盒可另外包含一种或多种膜锚、多核昔酸结合蛋白(例如,多核昔酸解链酶)(其可以与衔接子预先结合)、连接酶、聚合酶和/或游离核昔酸或辅因子。

[0410] 试剂盒可以包含上文所公开的任何膜的组分,所述膜例如两亲层或三嵌段共聚物膜。试剂盒可以进一步包含跨膜孔。上文参考本发明方法论述的实施例中任一个同样适用于试剂盒。

[0411] 本发明的试剂盒可以另外包括一种或多种其它试剂或仪器,所述试剂或仪器能够实施上述实施方案中的任一个。此类试剂或仪器包括以下各项中的一个或多个:合适的缓冲液(水溶液)、用于从受试者获得样品的装置(如包含针的容器或仪器)、用于扩增和/或表达多核昔酸的装置、如上文所定义的膜或电压钳或膜片钳设备。试剂可以按干燥状态存在于试剂盒中,使得流体样品用于再悬浮试剂。试剂盒还可以任选地包括使试剂盒能够在本发明的方法中使用的说明书。所述试剂盒可以包含磁铁或电磁铁。试剂盒可以任选地包含核昔酸。

[0412] 以下实施例说明了本申请的非限制性方面。

[0413] 实施例

[0414] 实施例1

[0415] 该实施例显示当具有附接于模板链和互补链的单链前导序列的双链基因组DNA与CsgG纳米孔接触时,模板链和互补链在Dda解旋酶的控制下依次但分开地易位通过纳米孔(模板未通过发夹与互补序列连接)。与通过发夹连接在一起时相同模板/互补序列的易位相比,观察到测序准确度增加。

[0416] 材料和方法

[0417] DNA样品制备

[0418] 如下将基因组DNA片段化。将46μl的1μg基因组DNA转移至Covaris g-TUBE。然后将g-TUBE在室温下以根据制造商的方案建议所需的速度旋转1分钟。然后将g-TUBE倒置并再旋转1分钟以收集片段化DNA。将片段化DNA转移到干净的1.5ml Eppendorf DNA LoBind管中。根据制造商的方案,通过使用Agilent Bioanalyzer 12000DNA芯片分析1

μ l样品来评估片段化过程的成功。

[0419] 根据制造商的方案,用NEB的FFPE修复试剂盒以62 μ l体积处理回收的DNA,并使用1x Agencourt AMPure XP珠粒纯化DNA,之后在46 μ l无核酸酶的水中洗脱。

[0420] 然后用NEB的Ultra II End-prep模块处理经FFPE修复的DNA,向片段化DNA的每个末端添加5'磷酸和单个dA-核苷酸。根据制造商的方案,使用45 μ l经FFPE修复的DNA在60 μ l体积中进行反应,之后使用1x Agencourt AMPure XP珠粒纯化并在31 μ l无核酸酶的水中洗脱。使用QuBit荧光计定量1 μ l回收的准备完成的DNA。

[0421] 接下来,将30 μ l准备完成的DNA添加到干净的1.5ml Eppendorf DNA LoBind管中。然后添加含有如图3A所示的其上附接有Dda解旋酶的衔接子的溶液(20 μ l的Oxford Nanopore的SQK-LSK108衔接子混合物,其具有附接到每个衔接子的Dda解旋酶,如图3A所示(连接测序试剂盒1D(R9.4)的组分可从牛津纳米孔技术公司(Oxford Nanopore Technologies)购得的(该体积的衔接子混合物优化用于~350ng,其片段大小大于或等于8kb),然后添加50 μ l NEB的平端/TA连接主混合物,之后通过倒置5次将其混合。然后将反应物在室温下放置10分钟。

[0422] 为了纯化DNA,将40 μ l Ampure XP珠粒添加到来自前一步骤的衔接子连接反应中,并通过倒置5次将其混合。然后将管在旋转混合器(Hula混合器)上在室温下温育5分钟。然后将管置于磁架上,使珠粒沉淀,然后通过移液移除上清液。在添加衔接子珠粒结合缓冲液(140 μ l Oxford Nanopore的SQK-LSK108衔接子珠粒结合缓冲液(可从Oxford Nanopore Technologies商购获得的连接测序试剂盒1D(R9.4))的组分之前,将管从架上移除。然后通过轻弹管使珠粒重新悬浮。在这之后将管放回磁架上,使珠粒沉淀,然后通过移液移除上清液。将该步骤重复第二次。洗涤步骤后,将管从磁架上移除,使沉淀重新悬浮在洗脱缓冲液(25 μ l的Oxford Nanopore的SQK-LSK108洗脱缓冲液(可从Oxford Nanopore Technologies商购获得的连接测序试剂盒1D(R9.4)的组分))中。然后将管在室温下温育10分钟并放回到磁架上以使珠粒沉淀。将洗脱液移至干净的1.5ml Eppendorf DNA LoBind管中。

[0423] 然后制备测序反应混合物以产生图3B中描绘的基因组DNA构建体。通过向12.0 μ l回收文库中添加均来自于Oxford Nanopore的SQK-LSK108试剂盒的37.5 μ l的RBF和25.5 μ l的LLB(可从Oxford Nanopore Technologies商购获得的连接测序试剂盒1D(R9.4)的组分)来制备测序反应混合物。

[0424] 为了产生具有将模板连接到互补序列的发夹的双链基因组DNA(参见图1A中构建体的示意图),按照与上述类似的程序。在类似的程序中,使用Oxford Nanopore SQK-LSK102衔接子混合物和SQK-LSK208 HP衔接子(可从Oxford Nanopore Technologies商购获得的连接测序试剂盒2D(R9.4)的组分),而不是上面提到的衔接子混合物,并且使用的所有其他步骤和组分与上面提到的相同。

[0425] 纳米孔测序

[0426] 使用Oxford Nanopore MinION R9.4流通池获得电测量值。将双链基因组DNA(有或没有将模板连接到互补序列的发夹)添加到纳米孔系统中。进行实验,并且监测解链酶控制的DNA移动。

[0427] 分析

[0428] 事件的比对

[0429] 使用W02016059427中公开的方法进行信号的比对。95%或更高的比对值表示各个事件指示模板及其各自的互补序列

[0430] 比对信号的分析

[0431] 使用如上所述的2D方法进行对比信号的后续分析,以确定核苷酸序列。

[0432] 结果

[0433] 图1包括DNA构建体的示意图,所述DNA构建体包含与Y-衔接子和发夹附接的模板和互补DNA序列,在酶的控制下易位通过纳米孔(图1A),并且图1包括显示获得的测序准确度的图表(图1B)。图2A和2B包括DNA构建体的卡通图示,所述DNA构建体包含均与前导序列附接的模板和互补DNA序列,在酶的控制下易位通过纳米孔(图2A),并且包括显示获得的测序精确度的图表(图2B)。图1B和2B显示了说明使用单独的模板、单独的互补序列和来自模板和互补序列的组合信息获得的测序精确度的峰。通过比较图1B和2B,可以清楚地看出,当模板和互补序列未通过发夹环连接时,单独的互补序列的准确度增加(从~75%增加到>85%),并且当模板和互补序列未连接时,将来自模板和互补序列的信息组合时,准确度也增加。

[0434] 实施例2(对照:经修饰的孔,未向分析物添加结合位点)

[0435] 该实施例描述了当模板和互补序列不共价连接时表征双链多核苷酸的模板(捕获的第一链)和互补序列(第一链的反向互补)的方法。如下所述在数据分析后对所确定的模板和互补序列的鉴定在实施例中称为“跟随配对”。在一些实施方案中,当该配对在彼此1分钟内出现>80%重叠时,则鉴定潜在的跟随配对。在一些实施方案中,当跟随配对立即出现时,并且具有95-100%的重叠,则鉴定潜在的跟随配对。

[0436] 仅模板链(即那些未被分类为属于跟随配对的链)在本文中称为“T”。属于跟随配对的模板链在本文中称为“T_n”,属于跟随配对的互补链在本文中称为“C_n”。“n”可用于标识跟随配对的T_n和C_n成分,例如,T₁和C₁是跟随配对。

[0437] 在该实施例中,使用含有与附接于经修饰的纳米孔的孔标签(例如,捕获多核苷酸)不互补的多核苷酸序列的对照衔接子来说明使用经修饰的纳米孔检测跟随配对的频率。

[0438] 示例性材料和方法

[0439] 对照衔接子与基因组DNA的连接

[0440] 对照衔接子是NB01(天然条形码1),包含条形码上链(SEQ ID NO:1)和条形码下链(SEQ ID NO:2)。

[0441] SEQ ID NO:1

[0442] /5Phos/AAGGTTAACACAAAGACACCGACAACCTTCTTCAGCACCT

[0443] SEQ ID NO:2

[0444] /5Phos/GGTGCTGAAGAAAGTTGTCGGTGTCTTGTTAACCTTAGCAAT

[0445] 使用Oxford Nanopore Technologies测序试剂盒按照制造商指导进行对照衔接子与基因组DNA的连接。将1000ng末端修复的和dA尾的大肠杆菌基因组DNA在室温下以100μl与来自上面的5.5μl的640nM对照衔接子在1x Blunt/TA主混合物(NEB M0367L)中连接20分钟。如下进行样品的SPRI纯化:添加40μL的Agencourt AMPure珠粒(Beckman Coulter),用移液管混合样品,并在室温下温育5分钟。用不含核酸酶的水(AmbionTM)和无水乙醇

(FisherScientific) 制备70%乙醇溶液,并将该溶液储存在冰上。

[0446] 使珠粒在磁力架上沉淀并去除上清液。将沉淀的珠粒用500 μ L来自以上的70%乙醇溶液洗涤,而不破坏沉淀。去除上清液,再次用500 μ L的70%乙醇溶液洗涤沉淀的珠粒。去除70%乙醇溶液,并将沉淀在离心机中短暂脉冲,然后放回磁力架,然后去除最后残留的70%乙醇溶液。

[0447] 通过移液管混合使沉淀重新悬浮在50 μ L无核酸酶的水(AmbionTM)中,并使样品在冰上从珠粒上洗脱10分钟。使珠粒沉淀,并将含有样品的上清液移至新鲜的DNA LoBind管(Eppendorf)中,该样品称为对照-衔接子-基因组-DNA。

[0448] 酶-衔接子复合物与对照-衔接子-基因组DNA的连接

[0449] 将可从Oxford Nanopore Technologies测序试剂盒商购获得的等分试样的BAM(条形码衔接子混合物)在冰上解冻。将20 μ L的BAM在室温下以100 μ L与50 μ L对照-衔接子-基因组-DNA、20 μ L的NEBNext Quick连接反应缓冲液和10 μ L的QuickT4 DNA连接酶(E6056L)连接10分钟。如下进行样品的SPRI纯化2:然后添加40 μ L的Agencourt AMPure珠粒(Beckman Coulter),用移液管混合样品,并在室温下温育5分钟。使珠粒在磁力架上沉淀并去除上清液。用140 μ L衔接子珠粒结合缓冲液洗涤沉淀的珠粒,通过连续两次180°旋转磁架上的Eppendorf管,使珠粒重新悬浮在衔接子珠粒结合缓冲液中。使珠粒在磁力架上沉淀并去除上清液。再次用140 μ L缓冲液洗涤沉淀的珠粒,通过连续两次180°旋转磁架上的Eppendorf管,使珠粒重新悬浮在缓冲液中。使珠粒在磁架上沉淀并去除缓冲液,将沉淀在离心机中短暂脉冲,然后放回到磁力架,然后去除最后残留的缓冲液。

[0450] 通过移液管混合使沉淀重新悬浮在25 μ L无核酸酶的水(AmbionTM)中,并使该文库在冰上从珠粒上洗脱10分钟。

[0451] 标签修饰的纳米孔的制备

[0452] 制备经修饰的CsgG纳米孔以允许孔标签的缀合。例如,修饰CsgG单体(例如,通过氨基酸取代),例如提供半胱氨酸、非天然碱基等,用于孔标签缀合。使用含有编码氨基酸序列SEQ ID NO:7,具有如本文所述的一个或多个氨基酸取代的质粒的PT7载体制备经修饰的CsgG单体。

[0453] SEQ ID NO:7:无信号序列的野生型大肠杆菌CsgG的氨基酸序列(Uniprot登录号P0AEA2)

[0454] CLTAPPKEAARPTLMPRAQSYKDLTHLPAPTGKIFVSVYNIQDETGQFKPYPASNFSATAMLV
TALKDSRWFIPLERQGLQNLLNERKIIRAAQENGTVAINNRIPLQLSLTAANIMVEGSIIGYESNVKSGGVGARYFGI
GADTQYQLDQIAVNLRVVNVSTGEILSSVNTSKTILSYEVQAGVFRFIDYQRLLEGEVGYTSNEPVMLCLMSAIETG
VIFLINDGIDRGLWDLQNKAERQNDILVKYRHMSVPPE

[0455] 将质粒转化到BL21衍生细胞系中,使其突变以置换具有卡那霉素抗性的内源CsgG基因。将细胞接种在含有氨苄青霉素(100 μ g/ml)和卡那霉素(30 μ g/ml)的琼脂板上并在37°C下温育16小时。使用单集落接种100ml含有羧苄青霉素(100 μ g/ml)和卡那霉素(30 μ g/ml)的LB培养基,然后使起始培养物在37°C/250rpm下生长16小时。使用4x 25ml起始培养物接种4x 500ml含有羧苄青霉素(100 μ g/ml)、卡那霉素(30 μ g/ml) 3mM ATP、15mM MgSo₄和0.5mM鼠李糖的LB。使培养物生长直至达到平稳期,然后在37°C/250rpm下再生长2小时。添加葡萄糖达到0.2%,将温度降低到18°C,一旦培养物处于18°C,就可通过添加1% α -乳糖一水合物

诱导蛋白质表达。将培养物在18°C/250rpm下温育16小时。

[0456] 通过离心收获细胞并进行洗涤剂溶解(Bugbuster)。一旦溶解,就将样品进行初始链霉亲和素纯化(5ml HP链霉亲和素阱),将洗脱级分加热到60°C,离心并将上清液进行qIEX纯化(1ml Hi trap Q HP)。合并含有正确蛋白质的级分,浓缩并在24ml Superdex上进行最终精制。

[0457] 如下所述,用吗啉基寡核苷酸(SEQ ID NO:8)修饰上述纳米孔的等分试样:

[0458] (SEQ ID NO:8):由GeneTools提供的吗啉基寡核苷酸

[0459] /5' -GGAACCTCTTGACAA/-3' -毗啶基-二硫代/

[0460] 将1.3μL的1M DTT(二硫苏糖醇)添加到130μL来自以上的纳米孔中,其含有大约9.75μg的纳米孔,并使其在室温下温育1小时。按照制造商的指导,使用0.5mL 7MWCO Zeba脱盐柱(Thermo Fisher Scientific)将该样品缓冲液交换到反应缓冲液(25mM Tris、150mM NaCl、2mM EDTA、0.1% SDS和0.1% Brij58,pH 7)中。按照制造商的指导,使用7MWCO Zeba脱盐柱(Thermo Fisher Scientific)将该样品再次缓冲液交换到反应缓冲液中。通过将由GeneTools提供的300nmol吗啉基寡核苷酸溶解在150μL无核酸酶的水(AmbionTM)中来制备2mM的吗啉基寡核苷酸原液(SEQ ID NO:8)。将其添加到上述经缓冲液交换的样品中达到终浓度为500μM,并使其在室温下温育过夜。这称为经修饰的纳米孔。

[0461] 电测量

[0462] 从插入在磷酸盐缓冲液(例如,含亚铁氰化钾(II)和铁氰化钾(III),pH 8.0)中的嵌段共聚物中的单个经修饰纳米孔获取电测量。在到达插入嵌段共聚物中的单个经修饰纳米孔后,通过用2mL缓冲液冲洗去除任何过量的经修饰纳米孔。

[0463] 使引发缓冲液流过纳米孔系统。为了制备测序混合物,将400nM系链(SEQ ID NO:9),回收的珠粒纯化的文库和文库负载珠粒按照制造商的说明在缓冲液中混合。然后将测序混合物添加到纳米孔系统中。在180mV下进行实验,并且监测解链酶控制的DNA移动。

[0464] SEQ ID NO:9

[0465] /5Chol-TEG/TT/iSp18//iSp18//iSp18//iSp18/TTGACCGCTCGCCTC

[0466] 数据分析

[0467] 当DNA链穿过经修饰的纳米孔时,测量并收集通过纳米孔的电流变化。然后使用碱基识别算法(例如,递归神经网络(RNN)算法)确定链的序列,以获得fastq数据。随后使用本领域已知的序列比对工具将fastq序列数据与参考基因组进行比对。

[0468] 为了鉴定成对的相互互补的链(成对的模板和互补链),计算了链间的重叠分数。重叠分数定义为基因组中两条链共有的碱基的连续区段的长度(以碱基计),通过两条链所跨越(不一定重叠)的基因组部分的长度(以碱基计)标准化。计算每条链的最大重叠分数作为链与在该链的1分钟内(在链之前或之后)通过孔的所有其他链之间的最大重叠。因此,高的最大分数表明,给定的链属于互补对(作为模板或互补序列),而低的重叠分数表明该链不属于互补对。

[0469] 表2显示了使用如本文所述的未修饰或修饰的纳米孔测量的具有不同最大重叠分数值的链的百分比

	最大重叠分数	0.0-0.02	0.02-0.8	0.8-0.9	0.9-0.98	0.98-1
[0470]	如实施例 5 中所述使用未修饰的孔测量的%链	98.4	0.8	0.0	0.0	0.8
	如实施例 2 中所述使用经修饰的孔测量的%链	96.3	1.0	0.2	0.2	2.3
	如实施例 4 中所述具有经修饰的孔和跟随衔接子的%链	33.9	7.1	2.5	3.5	52.9

[0471] 结果

[0472] 解旋酶,例如Dda解旋酶,例如国际PCT公开号W02015/055981(其内容通过引用整体并入本文)中描述的解旋酶,用于控制多核苷酸移动通过经修饰的纳米孔的,例如,如国际PCT公开号W0 2016/034591(其内容通过引用整体并入本文)中所述的经修饰的CsgG纳米孔。表3显示了如本文所述的实施例2-5的数据。它列出了映射链的数量(例如,使用本领域已知的映射方法),其被指定属于如上面数据分析部分中描述的T、T_n和C_n分类。下面表3的第2行显示了在使用来自实施例2的对照衔接子的实例中,通过数据分析指定的T、T_n和C_n链的数目。2.9%的链被归类为跟随配对。

[0473] 在实施例5中,在使用未修饰的纳米孔的情况下,跟随配对的频率较低(表3,第5行),仅为0.6%。这表明即使在分析物不含确切的序列时,在纳米孔上使用标签也会增强跟随事件。不希望受理论束缚,这可能是因为孔-标签可以低效率结合互补序列的暴露ssDNA(相对于确定的位点)

[0474] 实施例3(经修饰的孔和跟随衔接子实施方案1)

[0475] 该实施例描述了当模板和互补序列不共价连接时表征双链多核苷酸的模板(捕获的第一链)和互补序列(第一链的反向互补)的方法。如下所述在数据分析后对所确定的模板和互补序列的鉴定在实施例中称为“跟随配对”。在一些实施方案中,当该配对在彼此1分钟内出现>80%重叠时,则鉴定潜在的跟随配对。在一些实施方案中,当跟随配对立即出现时,并且具有95-100%的重叠,则鉴定潜在的跟随配对。

[0476] 仅模板链(即那些未被分类为属于跟随配对的链)在本文中称为“T”。属于跟随配对的模板链在本文中称为“T_n”,属于跟随配对的互补链在本文中称为“C_n”。“n”可用于标识跟随配对的T_n和C_n成分,例如,T₁和C₁是跟随配对。

[0477] 在该实施例中,根据本文所述的一个实施方案,通过跟随衔接子的连接来实现检测跟随配对的频率增加。如图9-10所示,衔接子在双链体茎内含有捕获多核苷酸序列,使得捕获多核苷酸序列仅在链解链时才显露。捕获多核苷酸序列与附接于经修饰的纳米孔的多核苷酸序列互补。在该实施例中,捕获多核苷酸序列在双链体茎内不含间隔区,例如sp18,并且此类衔接子产生约10%的跟随效率(即,所有链的跟随%达到约10%)。

[0478] 材料和方法

[0479] 跟随衔接子与基因组DNA的连接

[0480] 跟随衔接子的一个实施方案包括条形码上链(SEQ ID N0:3)和条形码下链(SEQ ID N0:4),其分别以10μM和11μM在50mM HEPES pH 8、100mM乙酸钾中以2°C/分钟从95°C到22°C一起退火。杂交的DNA被称为条形码衔接子2。将6.4μL跟随衔接子添加到93.6μL的50mM Tris-HCl pH7.5、20mM氯化钠中以产生跟随衔接子1的640nM稀释液。

- [0481] SEQ ID NO:3
[0482] /5Phos/GCGTCTGCTTGGGTGTTAACCTTTGTCAGAGAGGTTCCAAGTCAGAGAGG TTCCT
[0483] SEQ ID NO:4
[0484] /5Phos/GGAACCTCTCTGACTTGGAACCTCTGACAAAAAGGTAAACACCCAAAGCAGA CGCCAGCAAT
- [0485] 使用Oxford Nanopore Technologies测序试剂盒按照制造商指南进行跟随衔接子与基因组DNA的连接。将1000ng末端修复的和dA尾的大肠杆菌基因组DNA在室温下以100 μ L与来自上面的5.5 μ L的640nM跟随衔接子在1x Blunt/TA主混合物(NEB M0367L)中连接20分钟。然后如实施例2中所述进行SPRI纯化。该样品将被称为跟随衔接子-基因组DNA。
- [0486] 将酶-衔接子复合物与跟随衔接子-基因组DNA的连接
- [0487] 将可从Oxford Nanopore Technologies测序试剂盒商购获得的等分试样的BAM(条形码衔接子混合物)在冰上解冻。将20 μ L的BAM在室温下以100 μ L与50 μ L跟随衔接子-基因组DNA、20 μ L的NEBNext Quick连接反应缓冲液和10 μ L的QuickT4 DNA连接酶(E6056L)连接10分钟。然后如实施例2中所述进行SPRI纯化2。
- [0488] 标签修饰的纳米孔的制备
- [0489] 制备纳米孔的方法以与上述实施例2中所述类似的方式进行,所述纳米孔经修饰w为包括与衔接子的捕获多核苷酸序列互补的多核苷酸序列。
- [0490] 电测量
- [0491] 当链穿过纳米孔时测量并获得电测量的方法以与以上实施例2中所述类似的方式进行。
- [0492] 数据分析
- [0493] 以与以上实施例2中所述类似的方式进行对收集到的电测量值的数据处理和分析。
- [0494] 结果
- [0495] 解旋酶,例如Dda解旋酶,例如国际PCT公开号W02015/055981(其内容通过引用整体并入本文)中描述的解旋酶,用于控制多核苷酸移动通过经修饰的纳米孔的,例如,如国际PCT公开号WO 2016/034591(其内容通过引用整体并入本文)中所述的经修饰的CsgG纳米孔。表3的第3行显示了在使用如本实施例中所述的跟随衔接子的实例中通过数据分析所指定的T、T_n和C_n链的数目。所有链中有7.7%被归类为跟随配对。
- [0496] 实施例4(经修饰的孔和跟随衔接子实施方案2)
- [0497] 该实施例描述了当模板和互补序列不共价连接时表征双链多核苷酸的模板(捕获的第一链)和互补序列(第一链的反向互补)的方法。如下所述在数据分析后对所确定的模板和互补序列的鉴定在实施例中称为“跟随配对”。在一些实施方案中,当该配对在彼此1分钟内出现>80%重叠时,则鉴定潜在的跟随配对。在一些实施方案中,当跟随配对立即出现时,并且具有95-100%的重叠,则鉴定潜在的跟随配对。
- [0498] 仅模板链(即那些未被分类为属于跟随配对的链)在本文中称为“T”。属于跟随配对的模板链在本文中称为“T_n”,属于跟随配对的互补链在本文中称为“C_n”。“n”可用于标识跟随配对的T_n和C_n成分,例如,T₁和C₁是跟随配对。
- [0499] 在该实施例中,根据本文所述的一个实施方案,通过跟随衔接子的连接来实现检

测跟随配对的频率增加。衔接子在双链体茎内含有捕获多核苷酸序列，使得捕获多核苷酸序列仅在链解链时才显露。捕获多核苷酸序列与附接于经修饰的纳米孔的多核苷酸序列互补。该衔接子还含有增强C_n捕获效率的化学物质。例如，在该实施例中，捕获多核苷酸序列在双链体茎内含间隔区，例如sp18，并且此类衔接子产生约60%的跟随效率(即，所有链的跟随%达到~60%)。不希望受理论束缚，这种改善可能是由于多种原因，例如，因为酶在遇到sp18时短暂停留，从而给予更多的偶联时间，从而提高捕获多核苷酸序列与经修饰的纳米孔上的互补序列标签的杂交效率，和/或因为sp18s对于耦合几何形状给予了更大的灵活性，和/或因为sp18s增加的长度是更好的几何形状并且显示更靠近孔标签的序列(例如，C_n的取向被改变，使得捕获效率得到改善，和/或遮挡/阻止T被捕获)。

[0500] 材料和方法

[0501] 跟随衔接子与基因组DNA的连接

[0502] 跟随衔接子的一个实施方案包括条形码上链(SEQ ID NO:5)和条形码下链(SEQ ID NO:6)，其分别以10μM和11μM在50mM HEPES pH 8、100mM乙酸钾中以2°C/分钟从95°C到22°C一起退火。杂交的DNA被称为条形码衔接子2。将6.4μL跟随衔接子添加到93.6μL的50mM Tris-HCl pH7.5、20mM氯化钠中以产生跟随衔接子2的640nM稀释液。

[0503] SEQ ID NO:5

[0504] /5Phos/GCGTCTGCTGGGTGTTAAC*i*Sp18//*i*Sp18//*i*Sp18//*i*Sp18/TTTTGTCAG
AGAGGTTCCAAGTCAGAGAGGTTCC*T*

[0505] SEQ ID NO:6

[0506] /5Phos/GAACCTCTTGACTTGGAACCTCTGACAAAAA/*i*Sp18//*i*Sp18//*i*Sp18//
*i*Sp18/GGT TAAACACCAAGCAGACGCCAGCAAT

[0507] 使用Oxford Nanopore Technologies测序试剂盒按照制造商指南进行跟随衔接子与基因组DNA的连接。将1000ng末端修复的和dA尾的大肠杆菌基因组DNA在室温下以100μL与来自上面的5.5μL的640nM跟随衔接子在1x Blunt/TA主混合物(NEB M0367L)中连接20分钟。然后如实施例2中所述进行SPRI纯化。该样品将被称为跟随衔接子-基因组DNA。

[0508] 将酶-衔接子复合物与跟随衔接子-基因组DNA的连接

[0509] 将可从Oxford Nanopore Technologies测序试剂盒商购获得的等分试样的BAM(条形码衔接子混合物)在冰上解冻。将20μL的BAM在室温下以100μL与50μL跟随衔接子-基因组DNA、20μL的NEBNext Quick连接反应缓冲液和10μL的QuickT4 DNA连接酶(E6056L)连接10分钟。然后如实施例2中所述进行SPRI纯化2。

[0510] 标签修饰的纳米孔的制备

[0511] 制备纳米孔的方法以与上述实施例2中所述类似的方式进行，所述纳米孔经修饰w为包括与衔接子的捕获多核苷酸序列互补的多核苷酸序列。

[0512] 电测量

[0513] 当链穿过纳米孔时测量并获得电测量的方法以与以上实施例2中所述类似的方式进行。

[0514] 数据分析

[0515] 以与以上实施例2中所述类似的方式进行对收集到的电测量值的数据处理和分析。

[0516] 结果

[0517] 解旋酶,例如Dda解旋酶,例如国际PCT公开号W02015/055981(其内容通过引用整体并入本文)中描述的解旋酶,用于控制多核苷酸移动通过经修饰的纳米孔的,例如,如国际PCT公开号WO 2016/034591(其内容通过引用整体并入本文)中所述的经修饰的CsgG纳米孔。表3的第4行显示了在使用如本实施例中所述的跟随衔接子的实例中通过数据分析所指定的T、 T_n 和 C_n 链的数目。45.1%的链被归类为跟随配对。

[0518] 图11A-11B显示了来自该实施例中描述的实验的电流迹线。将T、 T_n 和 C_n 标记为对应下面表4中所示的数据。如数据分析部分中所述,获得下面表4中所示的数据,并且是表3中针对实施例2、3和4汇总的数据的实例部分。

[0519] 图12A-12D显示了跟随配对多核苷酸的模板(T_n)和互补序列(C_n)在它易位通过经修饰的纳米孔时的电流迹线。纳米孔返回到图12A中测序的模板和互补序列之间的开孔(参见图12C),说明模板和互补链不共价连接的事实。

[0520] SEQ ID NO:10和SEQ ID NO:11是由图12A至12D的多核苷酸的Oxford Nanopore Technologies碱基识别RNN算法推导的序列。这些可以高保真度进行比对以证明通过经修饰的纳米为来自随机片段化双链多核苷酸片段的模板和互补序列孔进行测序。

[0521] 在图12A-12D中,用*标记的位置是在SEQ ID NO:5和SEQ ID NO:6中发现的Sp18间隔区,该基序的存在表明如本实施例中描述的跟随衔接子成功连接到多核苷酸上,并且该跟随衔接子的存在大大增强了跟随配对的百分比(如表3所示)。

[0522] 图13以对数标度显示了“链间隔时间”,以秒为单位。标记“链”用于将解旋酶控制的多核苷酸序列通过纳米孔的运动的电信号分类。这使用本领域已知的方法进行分类。链容易因返回开孔而分裂。箭头表示来自实施例4的数据中的显著群体,其使用如该实施例中所述的跟随衔接子。该群体表示跟随配对的比例增加,因为在实施例2和4中, T_n 和 C_n 间隔的时间通常短于T与T或T与 T_n 或 C_n 与T间隔的时间。这说明跟随衔接子也可用于在链间隔时间减少时增加通量。

[0523] 实施例5

[0524] 该实施例描述了当模板和互补序列不共价连接时表征双链多核苷酸的模板(捕获的第一链)和互补序列(第一链的反向互补)的方法。如下所述在数据分析后对所确定的模板和互补序列的鉴定在实施例中称为“跟随配对”。在一些实施方案中,当该配对在彼此1分钟内出现>80%重叠时,则鉴定潜在的跟随配对。在一些实施方案中,当跟随配对立即出现时,并且具有95-100%的重叠,则鉴定潜在的跟随配对。

[0525] 仅模板链(即那些未被分类为属于跟随配对的链)在本文中称为“T”。属于跟随配对的模板链在本文中称为“ T_n ”,属于跟随配对的互补链在本文中称为“ C_n ”。“n”可用于标识跟随配对的 T_n 和 C_n 成分,例如, T_1 和 C_1 是跟随配对。

[0526] 在该实施例中,来自Oxford Nanopore Technologies测序试剂盒的标准组分与Minion和Flowcell(包含未经修饰为包括如本文所述的孔标签的纳米孔阵列的SpotON流通池)一起用于证明在没有如本文所述的经修饰的纳米孔和跟随衔接子的情况下,检测跟随配对的频率。

[0527] 材料和方法

[0528] 对照衔接子与基因组DNA的连接

[0529] 使用Oxford Nanopore Technologies测序试剂盒按照制造商指南进行对照衔接子(例如,如实施例2中所述)与基因组DNA的连接。将1000ng末端修复的和dA尾的大肠杆菌基因组DNA在室温下以100 μ L与来自上面的5.5 μ L的640nM跟随衔接子在1x Blunt/TA主混合物(NEB M0367L)中连接20分钟。然后如实施例2中所述进行SPRI纯化。该样品将被称为跟随衔接子-基因组DNA。

[0530] 将酶-衔接子复合物与跟随衔接子-基因组DNA的连接

[0531] 将可从Oxford Nanopore Technologies测序试剂盒商购获得的等分试样的BAM(条形码衔接子混合物)在冰上解冻。将20 μ L的BAM在室温下以100 μ L与50 μ L对照-衔接子-基因组-DNA、20 μ L的NEBNext Quick连接反应缓冲液和10 μ L的QuickT4 DNA连接酶(E6056L)连接10分钟。然后如实施例2中所述进行SPRI纯化2。

[0532] 电测量

[0533] 当链穿过纳米孔时测量并获得电测量的方法以与以上实施例2中所述类似的方式进行。

[0534] 数据分析

[0535] 以与以上实施例2中所述类似的方式进行对收集到的电测量值的数据处理和分析。

[0536] 结果

[0537] 解旋酶,例如Dda解旋酶,例如国际PCT公开号W02015/055981(其内容通过引用整体并入本文)中描述的解旋酶,用于控制多核苷酸移动通过经修饰的纳米孔的,例如,如国际PCT公开号WO 2016/034591(其内容通过引用整体并入本文)中所述的经修饰的CsgG纳米孔。下面表3的第5行显示了在实例中,通过数据分析指定的T、T_n和C_n链的数目。所有链中0.6%被归类为跟随配对。

[0538] 表3

	T (仅模板)	T _n (模板)	C _n (互补序列)
实例 2 (对照)	29835	448	448
实例 3 (跟随衔接子 1)	13179	546	546
实例 4 (跟随衔接子 2)	23623	9707	9707
实例 5	426384	1301	1301

[0540] 表3显示了对于本文的实施例2-5,确定为仅模板(未跟随其互互补序列)、模板-n(跟随配对的第一链)和互补序列-n(跟随配对的反向互补序列)的链数。

[0541] 表4

fast5	开始 / 秒	结束/秒	持续时间/秒	基因组_起始_位置	基因组_结束_位置	分类	间隔时间/秒	重叠率	重叠_碱基数			
90e5fe72_read_ch241_file335.fast5	15678.130	15710.975	32.846	4554265	4565926	模板 1	0.047	0.954	11124			
90e5fe72_read_ch241_file336.fast5	15711.022	15739.518	28.497	4554266	4565390	互补序列 1						
90e5fe72_read_ch241_file337.fast5	15740.340	15750.546	10.206	2273577	2277115	模板 2		0.110	0.977	3498		
90e5fe72_read_ch241_file338.fast5	15750.656	15760.139	9.483	2273617	2277156	互补序列 2						
90e5fe72_read_ch241_file339.fast5	15760.295	15780.086	19.791	919284	926128	模板 3						
<hr/>												
<hr/>												
.fast5												
90e5fe72_read_ch241_file340.fast5	15780.234	15796.935	16.702	919310	926054	互补序列 3	N/A	N/A	N/A			
90e5fe72_read_ch241_file341.fast5	15797.159	15824.821	27.662	1332188	1341764	仅模板						
90e5fe72_read_ch241_file342.fast5	15826.033	15845.976	19.944	4382073	4386574	仅模板						
90e5fe72_read_ch241_file343.fast5	15846.421	15865.318	18.897	3754934	3761744	模板 4		0.997	6792			
90e5fe72_read_ch241_file344.fast5	15865.351	15881.103	15.752	3754952	3761745	互补序列 4						
90e5fe72_read_ch241_file345.fast5	15883.544	15893.891	10.347	1569939	1573232	仅模板						
90e5fe72_read_ch241_file346.fast5	15901.567	15912.124	10.557	3213966	3216413	模板 5	0.998	2447				
90e5fe72_read_ch241_file347.fast5	15912.176	15921.117	8.941	3213966	3216419	互补序列 5						

[0544] 表4包含来自MinION跟随运行的单个通道(通道241)的图11中所示的链的分析数据。该表包含以下各列:

[0545] Fast5=链的唯一文件名,在MinION运行期间保存为fast5输出

[0546] 开始=链的起始时间(秒)

[0547] 结束=链的结束时间(秒)

[0548] 持续时间=链的持续时间(秒)

[0549] 基因组_起始_位置=在将碱基识别的链与大肠杆菌参考比对后,基因组参考中比对链的区段的起始位置

[0550] 基因组_结束_位置=在将碱基识别的链与大肠杆菌参考比对后,基因组参考中比对链的区段的结束位置

[0551] 间隔时间=对于配对而言,链的间隔时间(秒)

[0552] 重叠率=对于配对而言,比较对准重叠(基因组_起始_位置和基因组_结束_位置之间)的重叠率

[0553] 通过链间非常短的间隔时间来识别配对,并且当与参考比对时它们重叠。

[0554] 实施例6

[0555] 该实施例描述了当将低浓度DNA加入测序装置时增加测序DNA分子的数量的方法。

[0556] 在该实施例中,这通过产生经修饰的DNA-酶衔接子来实现,所述经修饰的DNA-酶衔接子含有与附接于经修饰的纳米孔的多核苷酸序列互补的DNA序列。在该实施例中使用的示例性衔接子在图32A-35中示出。

[0557] 示例性材料和方法

[0558] 衔接子制备(衔接子设计A,例如,如图32A所示)

[0559] 通过以2°C/分钟将温度从95°C调节至22°C,使上链(SEQ ID NO:12)、阻断链(SEQ ID NO:13)和下链(SEQ ID NO:14)分别以10μM、11μM和11μM在50mM Hepes、100mM KOAc, pH8(总体积40μL)中退火。将退火的链与800μL的2.8μM解旋酶(例如,包括如本领域已知的野生型或其突变体的Dda解旋酶)混合,并在室温下温育5分钟。向该溶液添加10μL的8.1mM四甲基偶氮二甲酰胺,并在35°C下温育1小时。添加NaCl使最终浓度为500mM,添加MgCl₂至最终浓度为10mM,添加ATP至最终浓度为1mM,并将溶液在室温下温育30分钟。通过HPLC纯化样品。

[0560] SEQ ID NO:12

[0561] /5SpC3//iSpC3//GGCGTCTGCTGGGTGTTAACCTTTTTTTT/iSp18/AATGTACTTCGTTACGTATTGCT

[0562] SEQ ID NO:13

[0563] /5BNA-G//iBNA-G//iBNA-T//iBNA-T//iBNA-A/AACACCCAAGCAGACGCC
TAAGTCAGAGAGGTTCC

[0564] SEQ ID NO:14

[0565] /5Phos/GCAATACGTAACGAACGAAGT/iBNA-A//iBNA-meC//iBNA-A//iBNA-T//iBNA-T/TTT GAGGCGAGCGGTCAA

[0566] 衔接子制备(衔接子设计B,例如,如图34A所示)

[0567] 通过以2°C/分钟将温度从95°C调节至22°C,使上链(SEQ ID NO:12)、阻断链(SEQ ID NO:15)和下链(SEQ ID NO:16)分别以10μM、11μM和11μM在50mM Hepes、100mM KOAc, pH8(总体积40μL)中退火。将退火的链与800μL的2.8μM解旋酶(例如,包括如本领域已知的野生型或其突变体的Dda解旋酶)混合,并在室温下温育5分钟。向该溶液添加10μL的8.1mM四甲

基偶氮二甲酰胺，并在35°C下温育1小时。添加NaCl至最终浓度为500mM，添加MgCl₂至最终浓度为10mM，添加ATP至最终浓度为1mM，并将溶液在室温下温育30分钟。通过HPLC纯化样品。

[0568] SEQ ID NO:15

[0569] GGTTAACACCCAAGCAGACGCC TTT GAGGCGAGCGGTCAA

[0570] SEQ ID NO:16

[0571] /5Phos/GCAAT ACGTAACGTAAACGAAGT/iBNA-A//iBNA-meC//iBNA-A//iBNA-T//iBNA-T/TTT TAAGTCAGAGAGGTTCC

[0572] 连接制备

[0573] 使用75μL 7k MWCO zeba离心柱，使用2根柱，每柱10μL，并遵循制造商方案，将衔接子设计A缓冲液交换到50mM Tris、20mM NaCl, pH8中。正如使用Qubit®dsDNA测定试剂盒测量的那样，将样品稀释至2.65ng/μL。将以下组分在1.5mL DNA低结合管(由Eppendorf提供)中合并、混合，并在室温下温育10分钟：11.3μL衔接子(在50mM Tris、20mM NaCl中, pH8, 2.65ng/μL)、8.7μL 50mM Tris、20mM NaCl (pH8)、0.9μL 3.6kb dA-尾部DNA (40ng/μL; SEQ ID NO:26))、29.1μL蒸馏水和50μL TA连接酶主混合物(由New England Biolabs(英国)提供)。

[0574] 然后添加40μL的Agencourt AMPure珠粒(Beckman Coulter)，用移液管混合样品，并在室温下温育5分钟。使珠粒在磁力架上沉淀并去除上清液。用140μL衔接子珠粒结合缓冲液洗涤沉淀的珠粒，通过连续两次180°旋转磁架上的Eppendorf管，使珠粒重新悬浮在衔接子珠粒结合缓冲液中。使珠粒在磁力架上沉淀并去除上清液。再次用140μL缓冲液洗涤沉淀的珠粒，通过连续两次180°旋转磁架上的Eppendorf管，使珠粒重新悬浮在缓冲液中。使珠粒在磁架上沉淀并去除缓冲液，将沉淀在离心机中短暂脉冲，然后放回到磁力架，然后去除最后残留的缓冲液。

[0575] 通过移液管混合使沉淀重新悬浮在25μL含系链(SEQ ID NO:9)的洗脱缓冲液中，并使该文库在冰上从珠粒上洗脱10分钟。

[0576] 孔修饰

[0577] 制备经修饰的CsgG纳米孔以允许孔标签的缀合。例如，修饰CsgG单体(例如，通过氨基酸取代)，例如提供半胱氨酸、非天然碱基等，用于孔标签缀合。使用含有编码氨基酸序列SEQ ID NO:7，具有如本文所述的一个或多个氨基酸取代的质粒的PT7载体制备经修饰的CsgG单体。将质粒转化到BL21衍生细胞系中，使其突变以置换具有卡那霉素抗性的内源CsgG基因。将细胞接种在含有氨苄青霉素(100μg/ml)和卡那霉素(30μg/ml)的琼脂板上并在37°C下温育16小时。使用单集落接种100ml含有羧苄青霉素(100μg/ml)和卡那霉素(30μg/ml)的LB培养基，然后使起始培养物在37°C/250rpm下生长16小时。使用4x 25ml起始培养物接种4x 500ml含有羧苄青霉素(100μg/ml)、卡那霉素(30μg/ml) 3mM ATP、15mM MgSo₄、0.5mM鼠李糖的LB。使培养物生长直至达到平稳期，然后在37°C/250rpm下再生长2小时。添加葡萄糖达到0.2%，将温度降低到18°C，一旦培养物处于18°C，就可通过添加1%α-乳糖一水合物诱导蛋白质表达。将培养物在18°C/250rpm下温育16小时。

[0578] 通过离心收获细胞并进行洗涤剂溶解(Bugbuster)。一旦溶解，就将样品进行初始链霉亲和素纯化(5ml HP链霉亲和素阱)，将洗脱级分加热到60°C，离心并将上清液进行

qIEX纯化(1ml Hi trap Q HP)。合并含有正确蛋白质的级分,浓缩并在24ml Superdex上进行最终精制。

[0579] 如下所述,用吗啉基寡核苷酸(SEQ ID N0:8)修饰上述纳米孔的等分试样:将1.3 μ L的1MDTT(二硫苏糖醇)添加到130 μ L来自以上的纳米孔中,其含有大约9.75 μ g的纳米孔,并使其在室温下温育1小时。按照制造商的指导,使用0.5mL 7MWCO Zeba脱盐柱(Thermo Fisher Scientific)将该样品缓冲液交换到反应缓冲液(25mM Tris、150mM NaCl、2mM EDTA、0.1% SDS和0.1% Brij58, pH 7)中。按照制造商的指导,使用7MWCO Zeba脱盐柱(Thermo Fisher Scientific)将该样品再次缓冲液交换到反应缓冲液中。通过将由GeneTools提供的300nmol吗啉基寡核苷酸溶解在150 μ L无核酸酶的水(AmbionTM)中来制备2mM的吗啉基寡核苷酸原液(SEQ ID N0:8)。将其添加到上述经缓冲液交换的样品中达到终浓度为500 μ M,并使其在室温下温育过夜。这称为经修饰的纳米孔。

[0580] 电测量

[0581] 从插入在缓冲液(25mM磷酸盐缓冲液、150mM亚铁氰化钾(II)、150mM铁氰化钾(III), pH 8.0)中的嵌段共聚物中的单一经修饰的纳米孔获取电测量。在到达插入在嵌段共聚物中的单一经修饰的孔之后,2mL缓冲液(25mM磷酸盐缓冲液、150mM亚铁(II)氰化钾、150mM铁(III)氰化钾,pH 8.0)流过系统以去除任何过量的经修饰的纳米孔。

[0582] 使引发缓冲液流过纳米孔系统。为了制备测序混合物,将引发缓冲液、(正如使用Qubit[®]dsDNA HS测定试剂盒测量的那样)和蒸馏水混合在一起。然后将测序混合物添加到纳米孔系统中。

[0583] 结果

[0584] 解旋酶(例如,如本领域已知的解旋酶,包括其野生型或突变体)用于控制所述多核苷酸通过经修饰的纳米孔的移动。图38显示了当多核苷酸易位通过未修饰的纳米孔时的电流迹线,即这是在纳米孔和分析物之间不存在寡核苷酸杂交的系统。在该系统中,链间隔时间均匀分布在1秒左右(参见图40)。

[0585] 图39显示了当多核苷酸易位通过经SEQ ID N0:8修饰的纳米孔时的电流迹线,即这是允许纳米孔和分析物之间杂交的系统。在该系统中,存在两个链间隔时间群:(a)均匀分布在1秒左右,和(b)快速捕获分析物(<0.1秒)(参见图40)。这表明分析物与孔杂交,而孔正在对另一条链进行测序。

[0586] 两个系统之间的这种差异(孔和分析物之间没有杂交相对于孔和分析物之间杂交)汇总在图40中,其显示了两个系统之间的链间隔时间的直方图。这证明在没有杂交的情况下,仅观察到一种捕获类型,并且当分析物可以与孔杂交时,存在另外的捕获类型,其中分析物在前一分析物之后被快速捕获(<0.1秒)。这种减少的链间隔时间增加了测序的链的总数图41。

[0587] 分析物的两端都能够拴系到孔上。

[0588] 实施例7

[0589] 以下是使用吡啶基二硫代吗啉基修饰纳米孔(例如,CsgG纳米孔)以具有附接于纳米孔外表面的孔标签的示例性方案。

[0590] 标签修饰的纳米孔的制备

[0591] 制备经修饰的纳米孔(例如,CsgG纳米孔)以允许孔标签的缀合。例如,修饰CsgG单

体(例如,通过氨基酸取代),例如提供半胱氨酸、非天然碱基等,用于孔标签缀合。使用含有编码氨基酸序列SEQ ID NO:7,具有如本文所述的一个或多个氨基酸取代的质粒的PT7载体制备经修饰的CsgG单体。将质粒转化到BL21衍生细胞系中,使其突变以置换具有卡那霉素抗性的内源CsgG基因。将细胞接种在含有氨苄青霉素(100 μ g/ml)和卡那霉素(30 μ g/ml)的琼脂板上并在37°C下温育16小时。使用单集落接种100ml含有羧苄青霉素(100 μ g/ml)和卡那霉素(30 μ g/ml)的LB培养基,然后使起始培养物在37°C/250rpm下生长16小时。使用4x 25ml起始培养物接种4x 500ml含有羧苄青霉素(100 μ g/ml)、卡那霉素(30 μ g/ml) 3mM ATP、15mM MgSo₄和0.5mM鼠李糖的LB。使培养物生长直至达到平稳期,然后在37°C/250rpm下再生长2小时。添加葡萄糖达到0.2%,将温度降低到18°C,一旦培养物处于18°C,就可通过添加1% α -乳糖一水合物诱导蛋白质表达。将培养物在18°C/250rpm下温育16小时。

[0592] 通过离心收获细胞并进行洗涤剂溶解(Bugbuster)。一旦溶解,就将样品进行初始链霉亲和素纯化(5ml HP链霉亲和素阱),将洗脱级分加热到60°C,离心并将上清液进行qIEX纯化(1ml Hi trap Q HP)。合并含有正确蛋白质的级分,浓缩并在24ml Superdex上进行最终精制。

[0593] 如下所述,用孔标签如吗啉基寡核苷酸(例如,如SEQ ID NO:8所示)修饰上述纳米孔的等分试样。将1.3 μ L的1M DTT(二硫苏糖醇)添加到130 μ L来自以上的纳米孔中,其含有大约9.75 μ g的纳米孔,并使其在室温下温育1小时。按照制造商的指导,使用0.5mL 7MWCO Zeba脱盐柱(Thermo Fisher Scientific)将该样品缓冲液交换到反应缓冲液(25mM Tris、150mM NaCl、2mM EDTA、0.1% SDS和0.1% Brij58,pH 7)中。按照制造商的指导,使用7MWCO Zeba脱盐柱(Thermo Fisher Scientific)将该样品再次缓冲液交换到反应缓冲液中。在无核酸酶的水(AmbionTM)中制备孔标签如吗啉基寡核苷酸(例如,如SEQ ID NO:8所示)的2mM原液。将其添加到上述经缓冲液交换的样品中达到终浓度为500 μ M,并使其在室温下温育过夜,产生经修饰的纳米孔。

[0594] 分析和质量控制

[0595] SDS-PAGE-方案

[0596] 将2uL经修饰和未经修饰的纳米孔添加到8uL反应缓冲液中。将样品在PCR块中加热4分钟达到95°C,以使样品由寡聚体分解成单体。向每份样品中添加10uL的2x Laemmli样品缓冲液(pH 6.8的65.8mM Tris-HCl、26.3% (w/v) 甘油、2.1% SDS、0.01% 溴酚蓝)。将样品在4-20%TGX凝胶上以300mV电泳23分钟并使用SYPRO Ruby蛋白凝胶染色。凝胶的结果示于图27A中。

[0597] 与吡啶基-二硫代吗啉基修饰的孔的杂交-方案

[0598] 10uL经修饰的纳米孔具有相应的荧光杂交序列,其比经修饰的纳米孔过量2倍添加。将样品制成总体积为20uL。将每份样品在室温下放置1小时。添加5uL的5x无染料负载(pH 8.0的50mM Tris-HCl、25% 甘油、5mM EDTA)。样品在4-20% TBE凝胶上以160mV电泳80分钟。使凝胶在Cy3荧光的凝胶扫描仪上运行,使用SYBR Gold核酸凝胶染剂染色,并使用SYPRO Ruby蛋白凝胶染色。凝胶的结果示于图28-30中。

[0599] 实施例7

[0600] 该实施例描述了表征连接多核苷酸的方法,其中用于将多核苷酸连在一起的连接方法是点击化学。在该实施例中,当模板和互补序列不共价连接时,使用纳米孔表征第一双

链多核苷酸的模板(捕获的第一链)和互补序列(第一链的反向互补序列)。当模板和互补序列分离时,与连接到第一双链多核苷酸的跟随衔接子中的孔系链互补的序列在互补序列中暴露,并且互补序列结合附接于纳米孔的孔系链。串接衔接子也连接到第一双链多核苷酸,使得互补链可以串接到第二双链多核苷酸。

[0601] 材料和方法

[0602] 跟随衔接子与基因组DNA的连接

[0603] 跟随衔接子包括条形码上链(SEQ ID NO:17)和条形码下链(SEQ ID NO:18),其分别以10 μ M和11 μ M在50mM HEPES pH 8、100mM乙酸钾中以2°C/分钟从95°C到22°C一起退火。杂交DNA称为跟随衔接子。将6.4 μ L跟随衔接子添加到93.6 μ L的50mM Tris-HCl pH7.5、20mM氯化钠中以产生跟随衔接子的640nM稀释液。

[0604] SEQ ID NO:17:跟随衔接子上链

[0605] /5Phos/TAACGAGGTGTTCTATCTCGCGTCTGCTGGGTGTTAACC/iSp18//iSp18//iSp18//iSp18/TTTTGTCAGAGAGGTTCCAAGTCAGAGAGGTTCC

[0606] SEQ ID NO:18:跟随衔接子下链

[0607] /5Phos/GGAACCTCTCTGACTTGGAACCTCTGACAAAAA/iSp18//iSp18//iSp18//iSp18/GGTAAACACCAAGCAGACGCCAGATAGAACAAACCATCAGATTGTGTTAGTCGCT/iSp18//iSp18//iSp18//iSp18//AGCGACTAACAAACACAATCTGATG/DBCO/

[0608] 使用Oxford Nanopore Technologies测序试剂盒按照制造商指南进行跟随衔接子与基因组DNA的连接。将1000ng末端修复的和dA尾的大肠杆菌基因组DNA在室温下以100 μ L与来自上面的5.5 μ L的640nM跟随衔接子在1x Blunt/TA主混合物(NEB M0367L)中连接20分钟。如实施例2中所述进行样品的SPRI纯化。该样品将被称为跟随衔接子-基因组DNA。

[0609] 串接酶-衔接子复合物与跟随衔接子-基因组DNA的连接

[0610] 前导链(SEQ ID NO:19)、阻断链(SEQ ID NO:20)和下链(SEQ ID NO:21)分别以5.5uM、6uM和6uM在50mM HEPES pH 8、100mM乙酸钾中以2°C/分钟从95°C到22°C一起退火。杂交的DNA被称为串接酶-衔接子复合物。

[0611] SEQ ID NO:19:串接测序衔接子上链

[0612] /叠氮化物/GGTTGTTCTATCTC/iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3/GCGTCTGCTTGGGTGTTAACCTTTTTTT

[0613] /iSp18/AATGTACTTCGTTCAGTTACGT

[0614] SEQ ID NO:20:连接测序衔接子阻断链

[0615] GGTTAACACCCAAGCAGACGCCCTTGAGGCGAGCGGTCAA

[0616] SEQ ID NO:21:串接测序衔接子下链

[0617] /5Phos/TCGTTAACGTAACGAACGAAGT/iBNA-A//iBNA-meC//iBNA-A//iBNA-T//iBNA-T/

[0618] 等分试样的T4 Dda-(E94C/F98W/C109A/C136A/A360C)(SEQ ID NO:24,具有突变E94C/F98W/C109A/C136A/A360C然后(Δ M1)G1G2(其中(Δ M1)G1G2=M1的缺失然后加入G1和G2)在冰上解冻,然后根据制造商的说明,通过0.5ml Zeba柱将50 μ l缓冲液交换到50mM HEPES pH 8,100mM乙酸钾,2mM EDTA中。使用A280纳米值将回收的蛋白质进行定量,并使用

相同的缓冲液调节至 0.25mg ml^{-1} 。

[0619] 将 $27\mu\text{l}$ 缓冲液交换蛋白与 $3\mu\text{l}$ 串接酶-衔接子复合物在DNA低结合eppendorf中混合，并在 35°C 下温育10分钟。然后添加 $0.37\mu\text{l}$ 的 8.1mM TMAD，将样品在 35°C 下温育60分钟。然后加入 $30\mu\text{l}$ 的 50mM HEPES pH 8、 1M NaCl、 2mM MgCl₂、 2mM rATP，并在室温下再放置20分钟。

[0620] 加入 $222\mu\text{l}$ 的Agencourt AMPure珠(Beckman Coulter)并将样品在室温下在旋转器上温育5分钟。使珠粒在磁力架上沉淀并去除上清液。在仍然在磁架上时，用 $500\mu\text{l}$ 的 50mM Tris pH 7.5、 2.5M NaCl、 20% PEG 8,000洗涤珠粒，转 360° 以浸洗架子上的沉淀。去除洗涤缓冲液，并将沉淀在离心机中短暂脉冲，然后返回磁力架以去除最后残留的溶液。将沉淀物在室温下重新悬浮于 $30\mu\text{l}$ 的 50mM Tris pH 7.5、 20mM NaCl中5分钟，然后置于磁架上以回收纯化的衔接子，该衔接子称为预装串接酶-衔接子复合物。

[0621] 将 $20\mu\text{L}$ 的预装串接酶-衔接子复合物在室温下以 $100\mu\text{L}$ 与 $50\mu\text{L}$ 跟随衔接子-基因组DNA、 $20\mu\text{L}$ 的NEBNext Quick连接反应缓冲液和 $10\mu\text{L}$ 的QuickT4 DNA连接酶(E6056L)连接10分钟。如下进行SPRI纯化，然后添加 $40\mu\text{L}$ 的Agencourt AMPure珠粒(Beckman Coulter)，通过移液管混合样品，并在室温下温育5分钟。使珠粒在磁力架上沉淀并去除上清液。用 $140\mu\text{L}$ 衔接子珠粒结合缓冲液洗涤沉淀的珠粒，通过连续两次 180° 旋转磁架上的Eppendorf管，使珠粒重新悬浮在衔接子珠粒结合缓冲液中。使珠粒在磁力架上沉淀并去除上清液。再次用 $140\mu\text{L}$ 缓冲液洗涤沉淀的珠粒，通过连续两次 180° 旋转磁架上的Eppendorf管，使珠粒重新悬浮在缓冲液中。使珠粒在磁架上沉淀并去除缓冲液，将沉淀在离心机中短暂脉冲，然后放回到磁力架，然后去除最后残留的缓冲液。

[0622] 通过移液管混合使沉淀重新悬浮在 $25\mu\text{L}$ 无核酸酶的水(AmbionTM)中，并使该文库在冰上从珠粒上洗脱10分钟。

[0623] 标签修饰的纳米孔的制备

[0624] 制备纳米孔的方法以与上述实施例2中所述类似的方式进行，所述纳米孔经修饰w为包括与衔接子的捕获多核苷酸序列互补的多核苷酸序列。

[0625] 电测量

[0626] 当链穿过纳米孔时测量并获得电测量的方法以与以上实施例2中所述类似的方式进行。

[0627] 数据分析

[0628] 当DNA链穿过经修饰的纳米孔时，测量并收集通过纳米孔的电流变化。然后使用碱基识别算法(例如，递归神经网络(RNN)算法)确定链的序列，以获得fastq数据。随后使用本领域已知的序列比对工具将fastq序列数据与参考基因组进行比对。

[0629] 实施例8

[0630] 该实施例描述了表征和串接双链靶多核苷酸的方法，其中附接方法是非共价的。第一双链靶多核苷酸的互补链募集第二双链靶多核苷酸并使其达到孔的局部浓度。反过来，当对第一互补链测序时，募集的第二双链靶多核苷酸变得从互补链去杂交，而是以与实施例2中进行的类似方式与孔系链杂交。这使得第一和第二(以及随后的，第三、第四、第五等)双链靶多核苷酸能够彼此跟随通过该孔，链间隔时间最短。当双链靶多核苷酸的浓度较低时，这特别有用，因为第二靶多核苷酸可以在为第一靶多核苷酸测序时募集。

[0631] 以与实施例7中所述相同的方式制备分析物,但使用SEQ ID NO:5、6、22、15和23,而不是SEQ ID NO:17-21。所有其他程序、试剂和条件与实施例7中描述的相同。

[0632] SEQ ID NO:5:跟随衔接子上链

[0633] /5Phos/GCGTCTGCTGGGTGTTAAC*/iSp18//iSp18//iSp18//iSp18/*
TTTTTGTCAGAGAGG TTCCAAGTCAGAGAGGTTCT

[0634] SEQ ID NO:6:跟随衔接子下链

[0635] /5Phos/GGAACCTCTGACTTGGAACCTCTGACAAAAA/iSp18//iSp18//iSp18//
iSp18/GGT TAAACACCCAAGCAGGCCAGCAAT

[0636] SEQ ID NO:22:捕捞衔接子上链

[0637] /5SpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//
iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//TTGTCAGAGAGGTTCC/iSpC3//iSpC3//
iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//iSpC3//
iSpC3//iSpC3/GCGTCTGCTGGGTGTTAACCTTTT TTTT/iSp18/AATGTAACCGTTACGT

[0638] SEQ ID NO:15:捕捞衔接子阻断链

[0639] GGTTAACACCCAAGCAGACGCCTTGAGGCGAGCGGTCAA

[0640] SEQ ID NO:23:捕捞衔接子下链

[0641] /5Phos/ACGTAACGTGAACGAACT//iBNA-A//iBNA-mecC//iBNA-A//iBNA-T//iBNA-T//

[0642] 实施例9

[0643] 这是描述表征和串接双链靶多核苷酸的方法的另一个实施例，其中附接方法是非共价的。这种方法与实施例8完全相同，但使用与SEQ ID NO:9不同的系链。双组分捕捞系链为跟随序列和孔系链提供第二杂交位点，以增加所见事件的比例。

[0644] 当形成测序混合物时,SEQ ID NO:9用400nM退火的SEQ ID NO:24和SEQ ID NO:25替换。所有其他程序,试剂和条件与实施例8中描述的相同。

[0645] SEQ ID NO:24:捕捞系链上链

[0646] TTGTCAGAGAGGTTCTTTTTTTTTTTTTTTTTGGTGTTCGTTGG
TGCTGATATTGCTTTTGACCGCTCGCCTC

[0647] SEQ ID NO:25:捕捞系链下链

[0648] GCAATATCAGCACCAACAGAAACAAACCTT/iSp18//iSp18//iSp18//iSp18//iSp18//
iSp18/TT/3Cho1TEG/

[0649] 实施例10

[0650] 该实施例描述了表征和串接许多双链靶多核苷酸的方法,其中附接方法是非共价的。第一双链靶多核苷酸的互补链募集许多其他双链靶多核苷酸并使其集中在孔附近。这样在孔周围提供比在一般本体溶液中更高的局部浓度,因此双链靶多核苷酸彼此跟随通过开孔,链间隔时间最短。当双链靶多核苷酸的浓度较低时,这尤其有用。该实施例像实施例8那样进行。然而,不是使SEQ ID N0:9退火,而是使用由与单链结合蛋白偶联的寡核苷酸组成的系链。

[0651] 当对第一双链靶多核苷酸的模板链进行测序时，互补链作为ssDNA释放到溶液中。其他双链靶多核苷酸的单链结合蛋白能够与ssDNA结合。作为跟随过程的一部分，当对互补链进行测序时，互补链的3'被拉回到孔中。ssDNA互补链上的单链结合蛋白在遇到控制互补

序列通过孔的运动的马达蛋白时从互补链中移位,因此沉积在孔周围,增加局部浓度。

[0652] 其他实施方案

[0653] 本说明书中公开的所有特征可以任意组合。本说明书中公开的每个特征可以用用于相同、等效或类似目的替代特征代替。因此,除非另有明确说明,否则所公开的每个特征仅是一系列等效或类似特征的实例。

[0654] 从以上描述中,本领域技术人员可以容易地确定本公开的基本特征,并且在不脱离其精神和范围的情况下,可以对本公开进行各种改变和修改以使其适应各种用途和条件。因此,其他实施方案也在权利要求内。

[0655] 等效方案

[0656] 尽管本文已经描述和说明了本发明的若干个实施方案,但本领域普通技术人员将容易想到用于执行本文描述的功能和/或获得这些结果和/或这些优点中的一个或多个优点的各种其它装置和/或结构,并且此类变型和/或修改中的每一个被认为是在本文所述的发明实施方案的范围内。更一般来讲,本领域的技术人员将容易认识到,本文中描述的所有参数、尺寸、材料以及构造意味着是示例性的,并且实际参数、尺寸、材料和/或构造将取决于发明传授内容所用于的一种或多种具体应用。本领域的技术人员仅仅使用常规实验将认识到或能够确认本文中描述的本发明的具体实施例的许多等效物。因此,应理解,前述实施例是仅通过实例方式来介绍的,并且在所附权利要求和其等效物的范围内,发明实施方案可以按与具体描述和要求不同的方式来实践。本公开的发明实施方案涉及本文中描述的每个单独的特征、系统、物品、材料和/或方法。此外,两个或更多个这样的特征、系统、物品、材料、试剂盒和/或方法的任何组合,如果这样的特征、系统、物品、材料、试剂盒和/或方法并不相互矛盾,被包含在本公开的发明范围内。

[0657] 如本文定义和使用的所有定义应理解为先于字典定义,通过引用并入的文献中的定义,和/或所定义的术语的普通含义。

[0658] 本文公开的所有参考文献、专利和专利申请均通过引用关于各自所引用的主题而并入本文,在某些情况下可涵盖整个文件。

[0659] 除非明确相反指出,否则本说明书和权利要求书中使用的不定冠词“一”和“一个”应理解为表示“至少一个”。

[0660] 如本文在说明书中使用的,短语“和/或”应当理解为是指这样联合的要素中的“任一个或两个”,即,要素在一些情况下共同存在而在其它情况下分开存在。用“和/或”列出的多个要素应以相同的方式解释,即“一个或多个”如此结合的要素。除了用“和/或”短语具体标识的要素,其它要素可以任选地存在,无论是与具体标识的那些要素相关还是不相关。因此,作为非限制性实例,当连同开放式语言,如“包含”使用时,提到“A和/或B”,在一个实施方案中,可以指仅有A(任选地包括除了B之外的要素);在另一个实施方案中,指仅有B(任选地包括除了A之外的要素);在又另一个实施方案中,指A和B两者(任选地包括其他要素);等等。

[0661] 如本文中在本说明书和权利要求中所使用的,“或”应被理解为具有与如上所定义的“和/或”相同的含义。例如,当将列表中的项目分开时,“或”或“和/或”应被解释为包容性的,即包括多个要素或要素清单中的至少一个要素、而且还包括多于一个要素,以及可选地其它未列出的项。仅仅清楚地指示相反的用语,如“……中的仅一个”或“……中的确切一

个”或者在权利要求中使用时“由……组成”将指包括多个要素或要素清单中的恰好一个要素。一般而言,当之前有排他性术语、比如“任一个”、“……之一”、“……中的仅一个”、或“……中的确切一个”时,本文中使用的用语“或”应当仅被解释为指示排他性替代品(即,“一个或另一个、而不是两个”)。当在权利要求中使用时,“基本上由……组成”应当具有如在专利法领域中所使用的普通含义。

[0662] 如在本说明书和权利要求中所使用的,关于一个或多个要素的清单的短语“至少一个”应被理解为是指选自要素清单中的任一个或多个要素的至少一个要素、但不一定包括要素清单内具体列出的每一个要素的至少一个、并且不排除要素清单中要素的任何组合。这个定义还允许可以可选地存在除了在短语“至少一个”所指的要素清单内具体标识的要素之外的要素,而无论是否与具体标识的那些元素相关还是不相关。因此,作为非限制性实例,“A和B中的至少一个”(或等效地,“A或B中的至少一个”,或等效地“A和/或B中的至少一个”)在一个实施例中可以是指至少一个、任选地包括多于一个A,而不存在B(并且任选地包括除了B的元件);在另一个实施例中,可以是指至少一个、任选地包括多于一个B,而不存在A(并且任选地包括除了A的元件);在又另一个实施例中,可以是指至少一个、任选地包括多于一个A,以及至少一个、任选地包括多于一个B(并且任选地包括其它元件);等。还应该理解,除非明确指出相反,否则在本文要求保护的包括一个以上步骤或操作的任何方法中,该方法的步骤或操作的顺序不一定限于叙述的该方法的步骤或步骤的顺序。

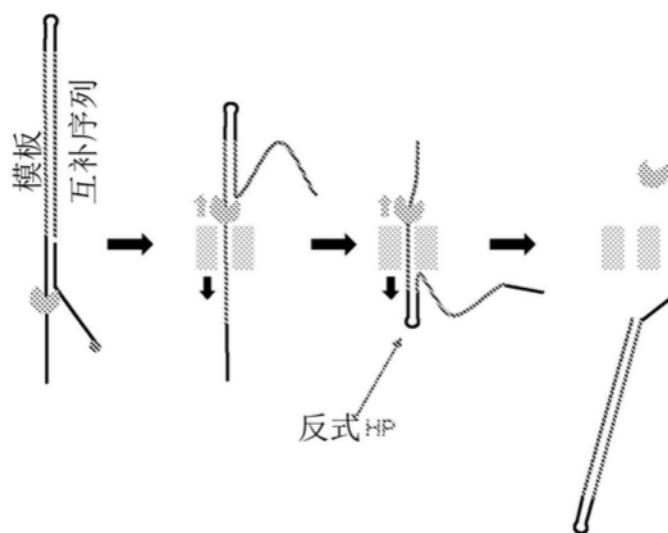


图1A

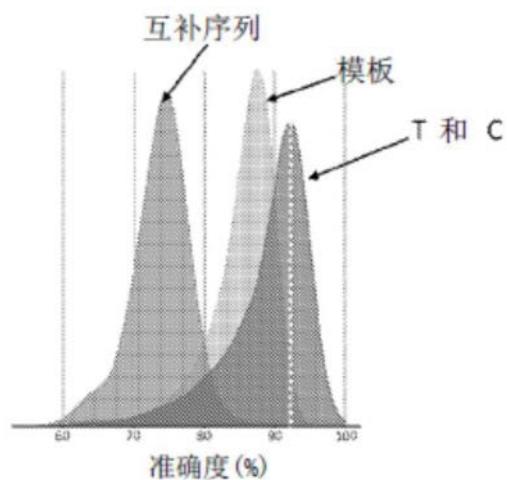


图1B

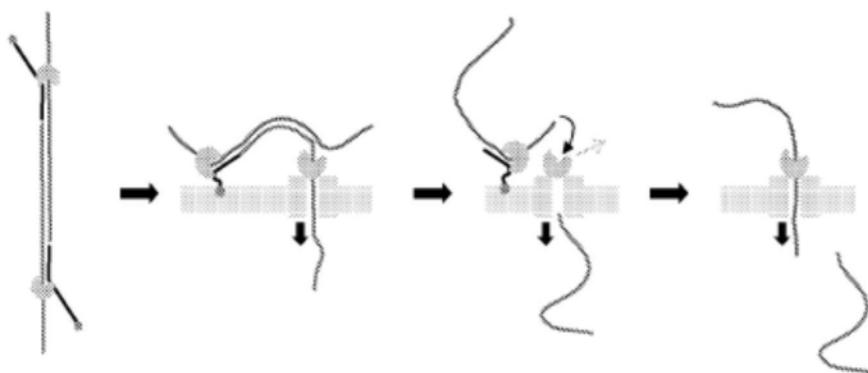


图2A

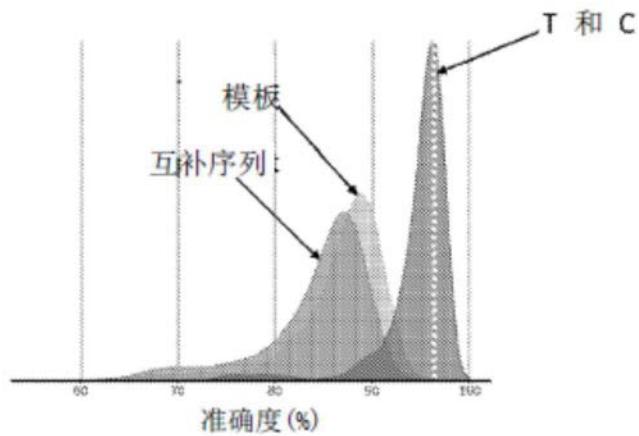


图2B

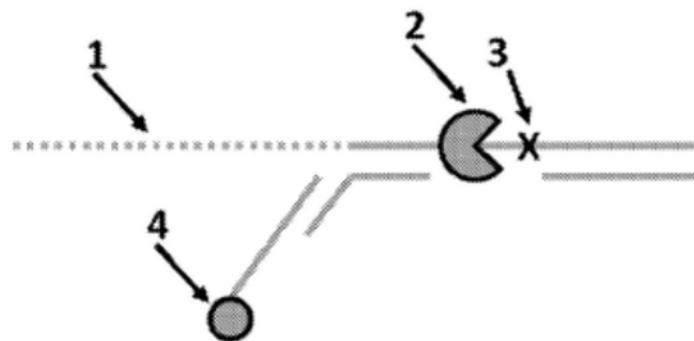


图3A

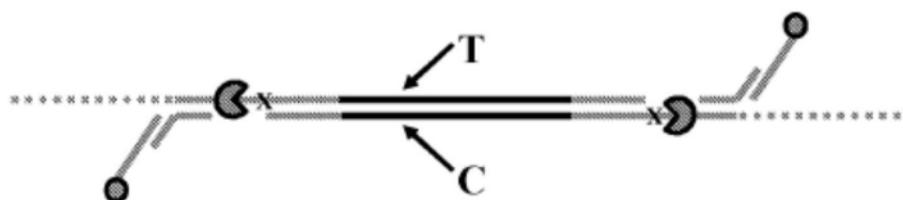


图3B

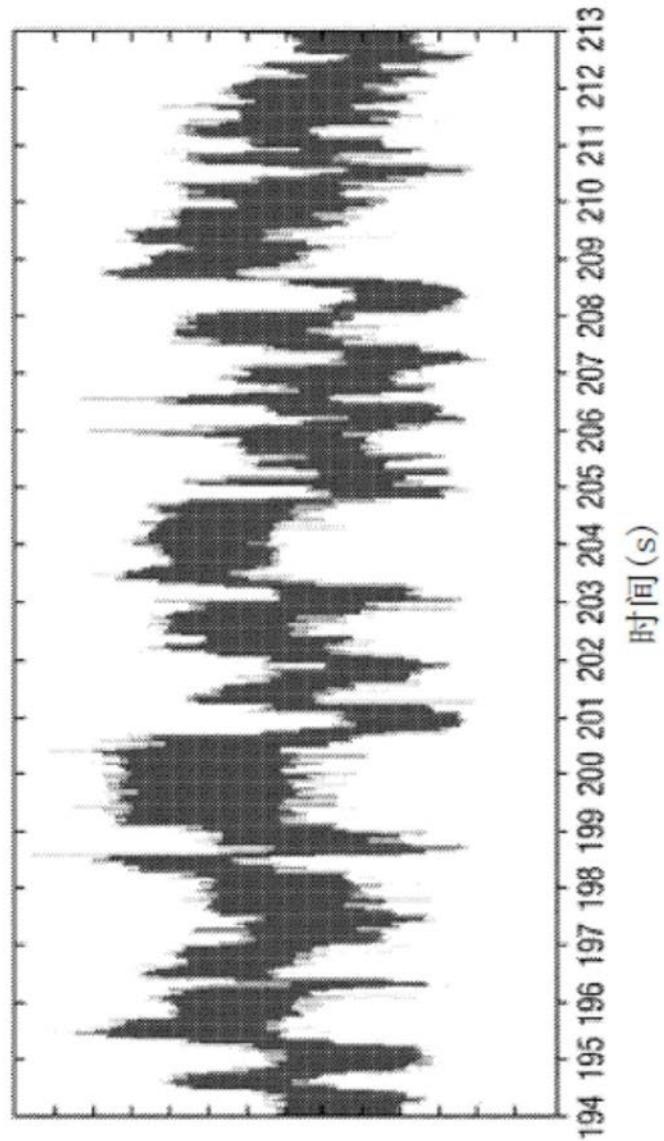


图4

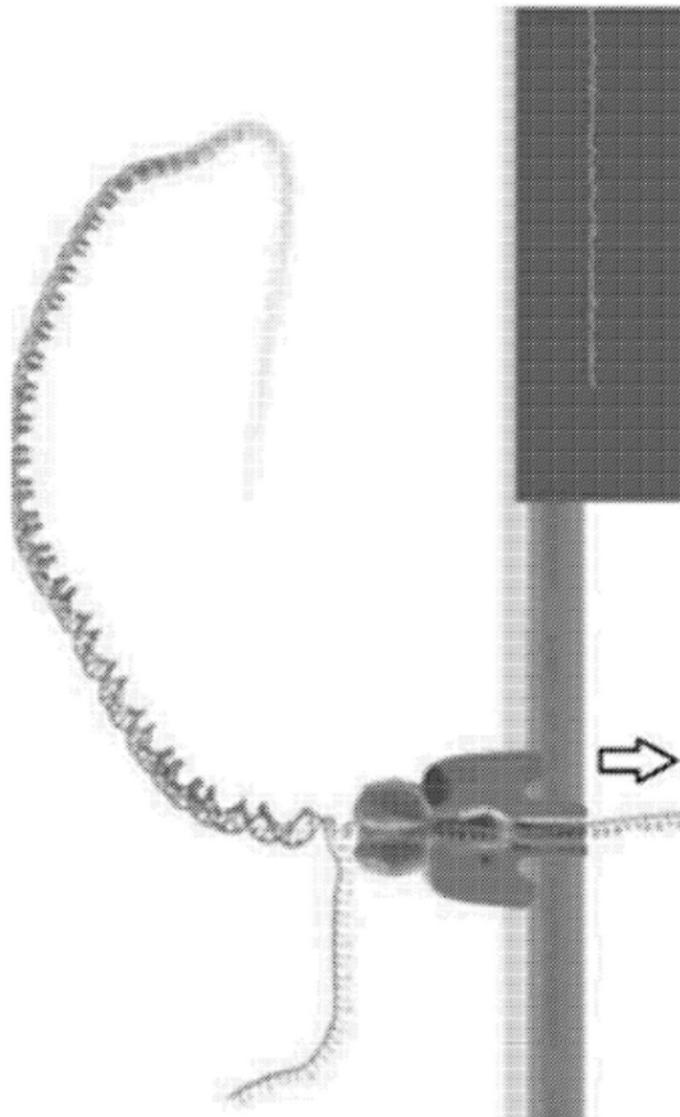


图5

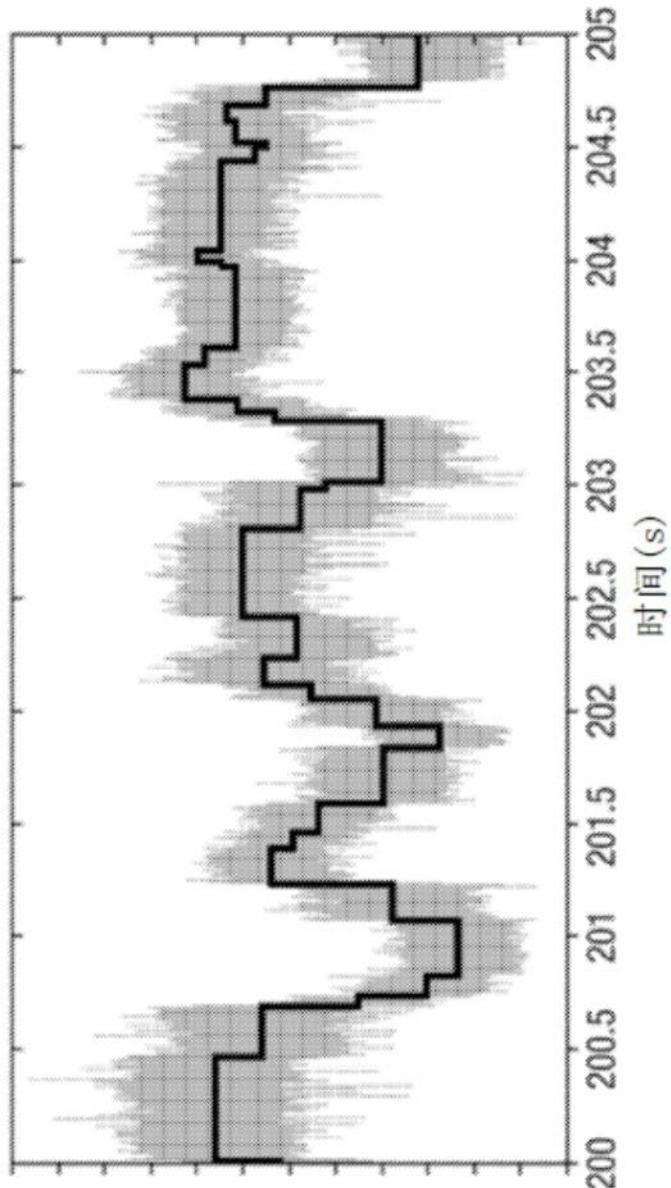


图6

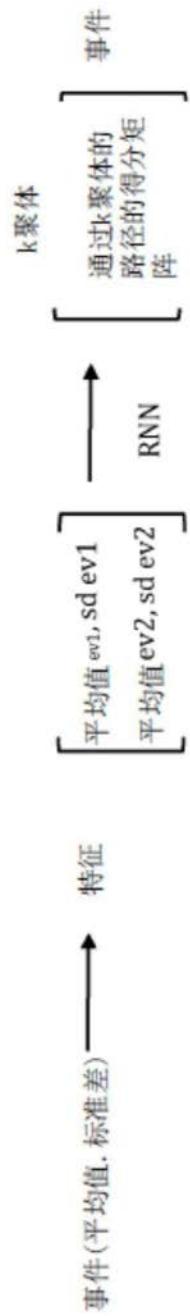


图7

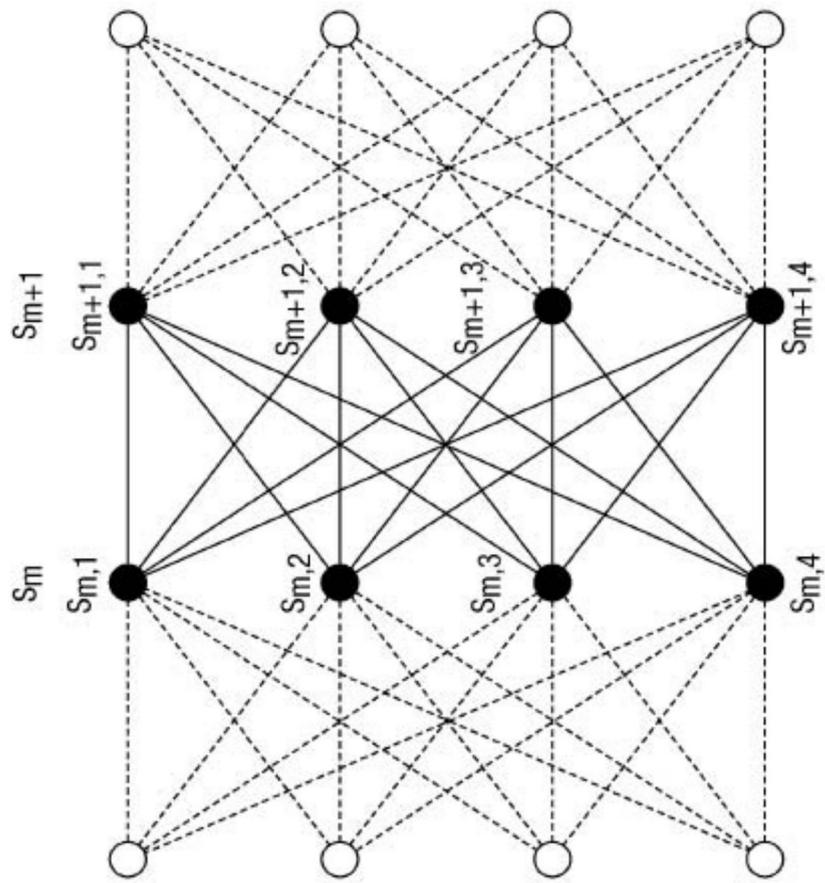


图8

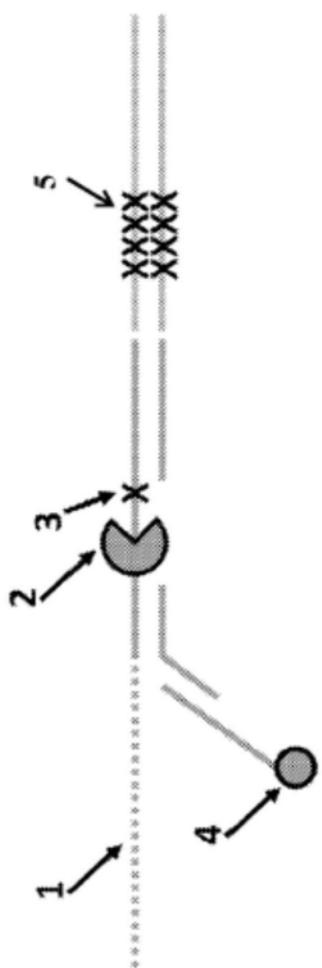


图9A

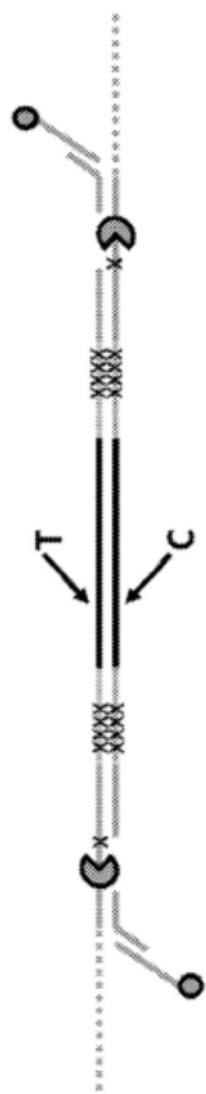


图9B

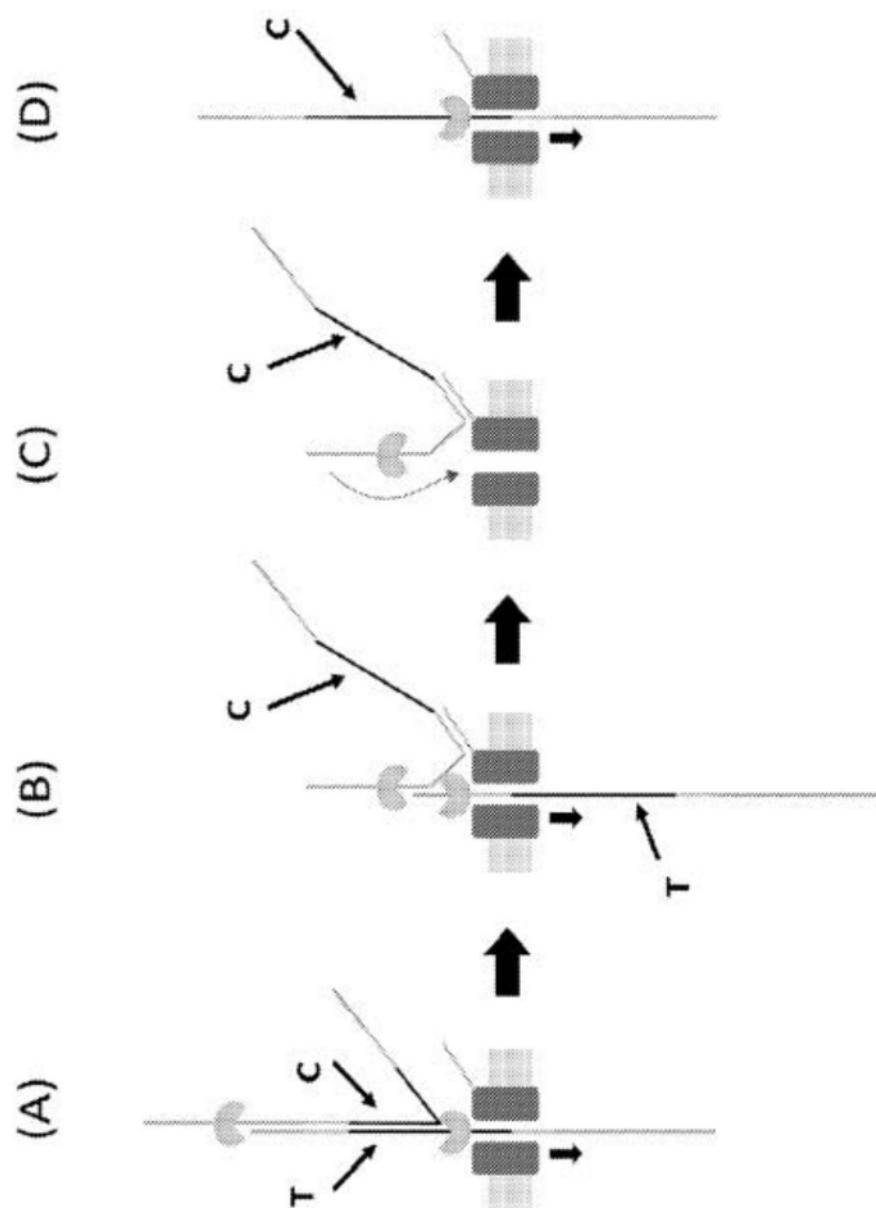


图10

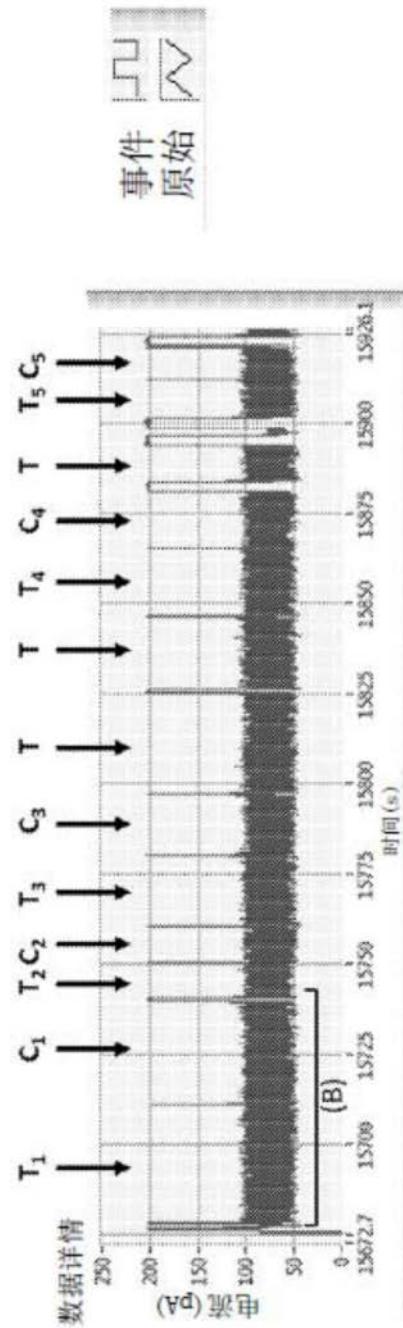


图11A

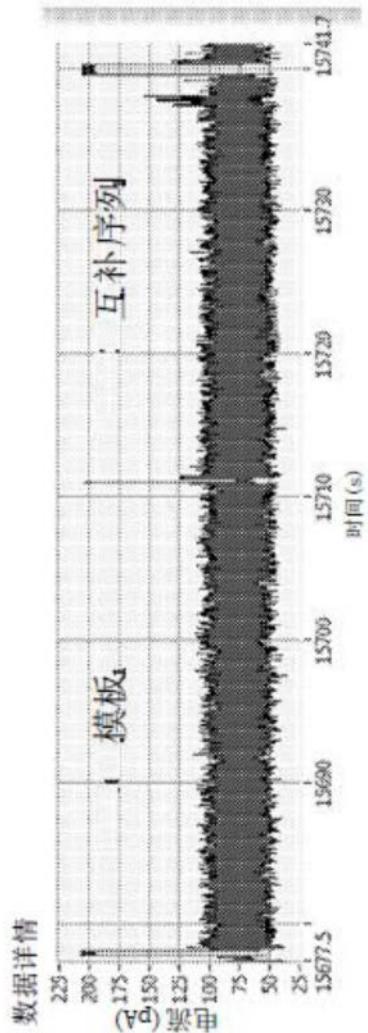


图11B

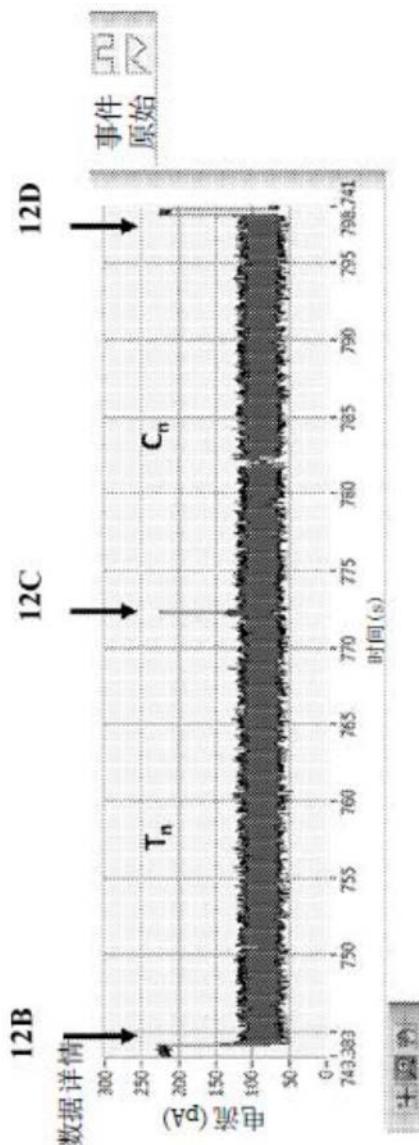


图12A

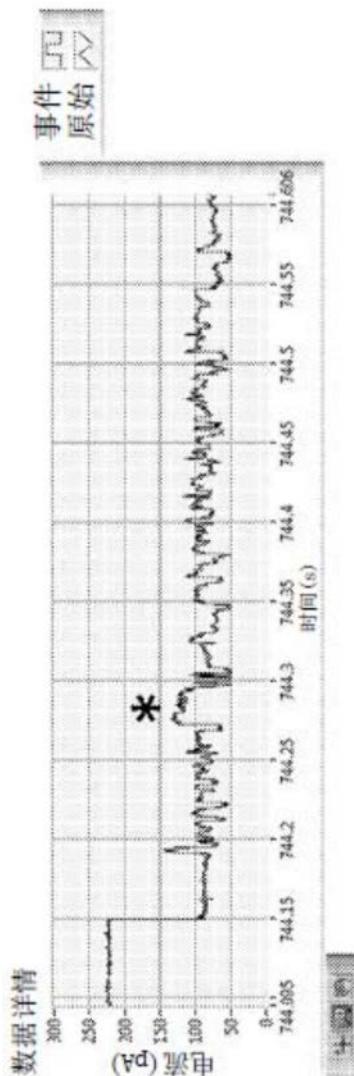


图12B

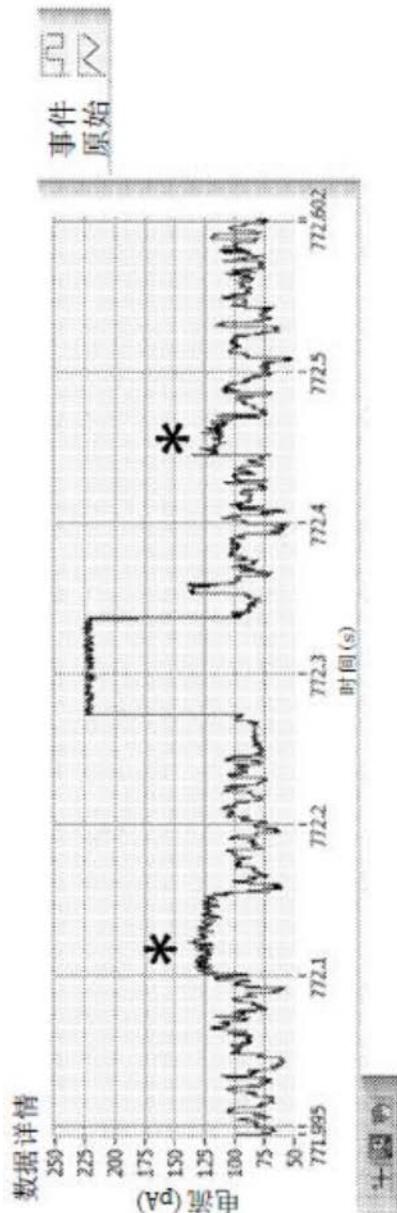


图12C

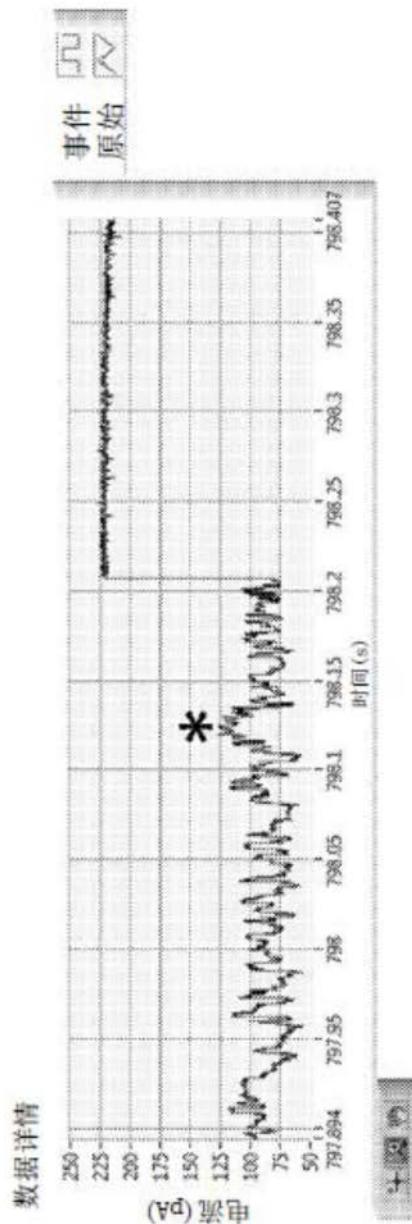


图12D

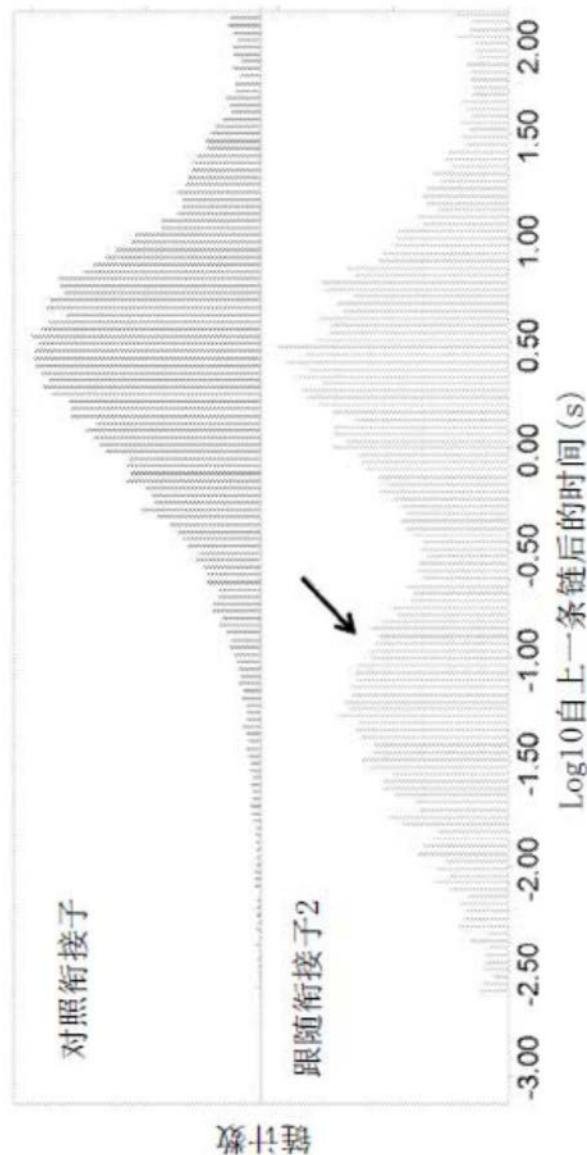


图13

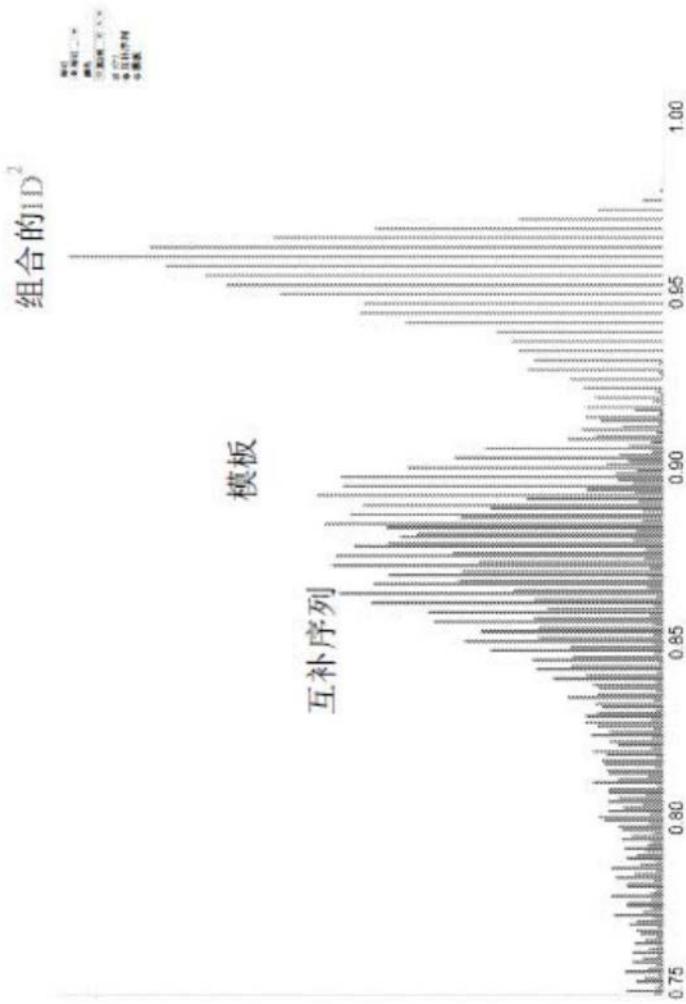


图14

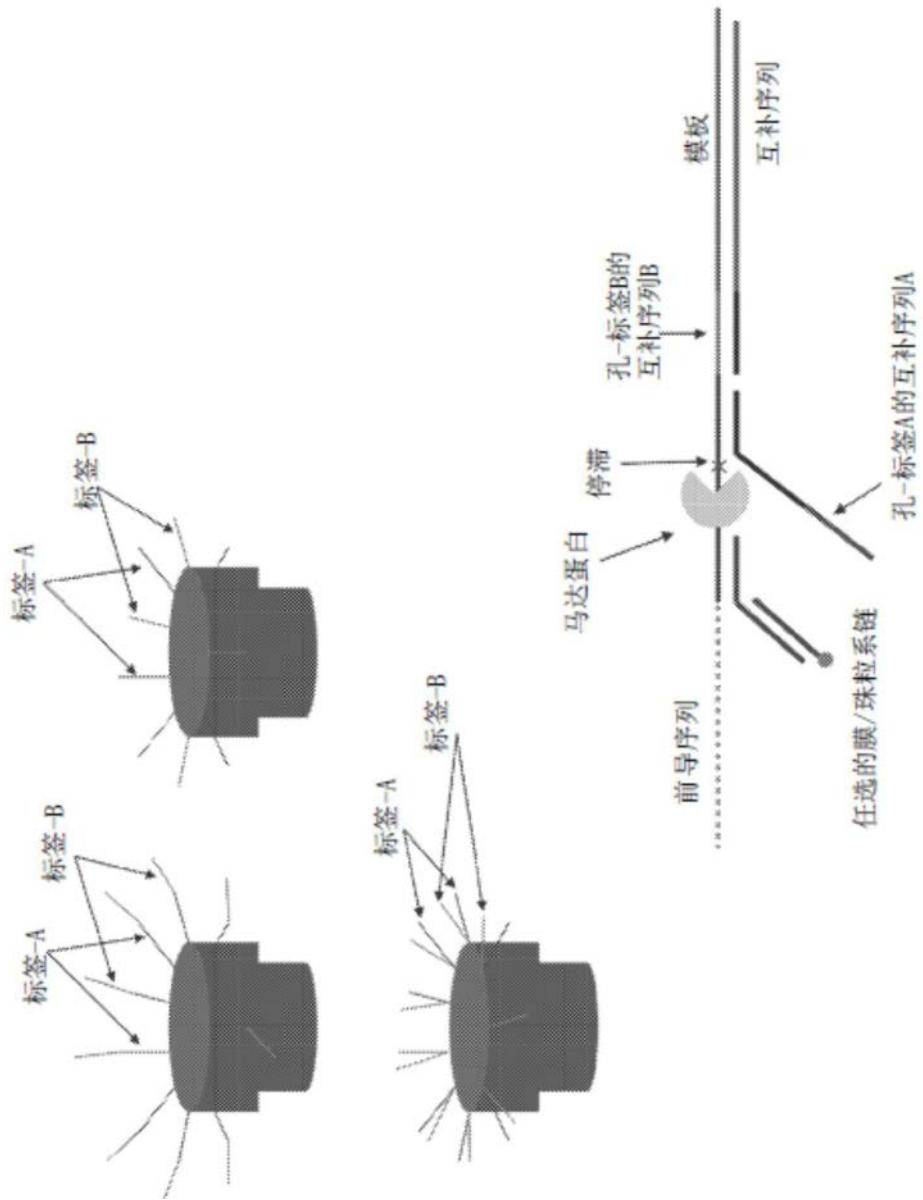


图15

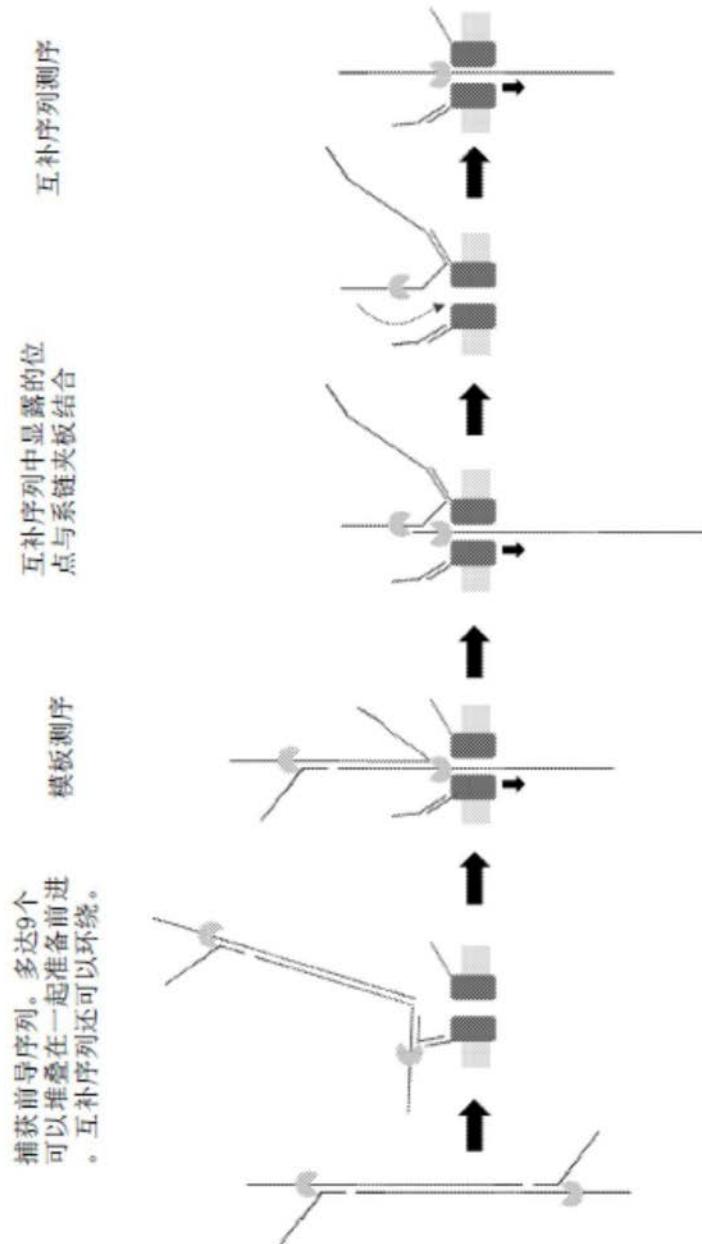


图16

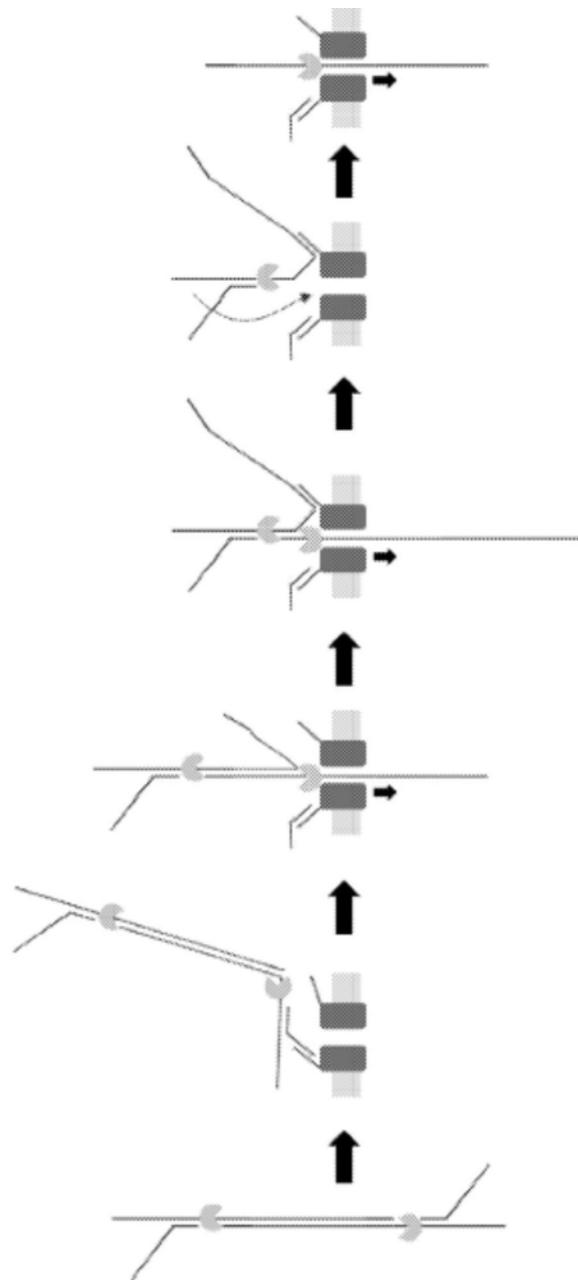


图17A



图17B

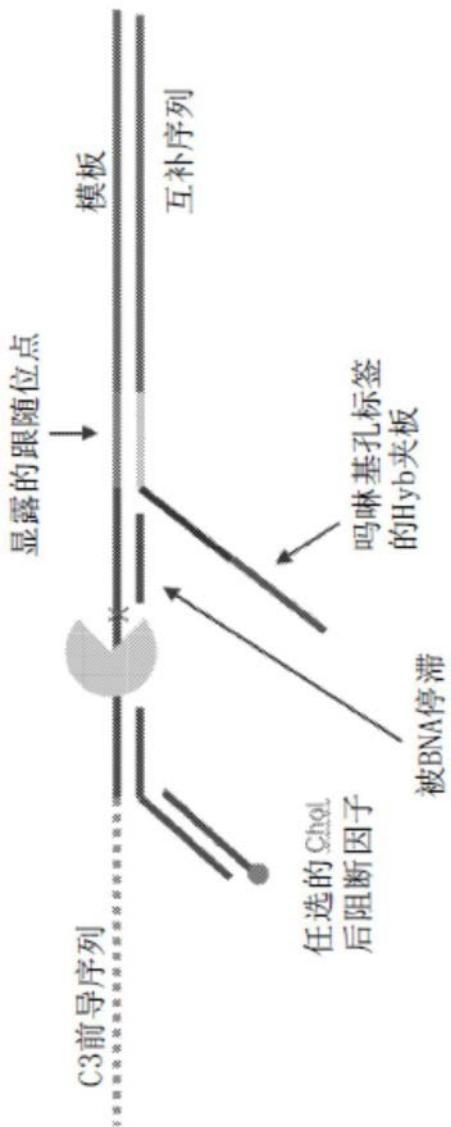


图18

捕获前导序列。多达9个互补序列可以堆叠在一起准备前进。互补序列还可以以环状优化，或者立即与相邻位点杂交形成环。互补序列保持被其尾部结合，或者如果几何形状优化，则立即与相邻位点杂交形成环。互补序列可以立即被捕获，或者如果孔被占据，则可以形成使前导序列靠近孔的分子内再次杂交。

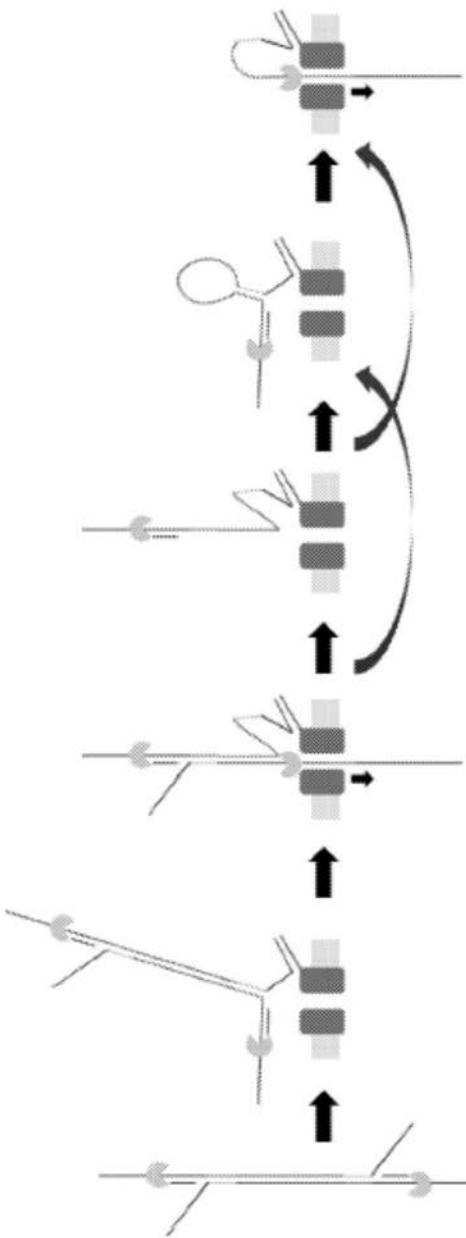


图19

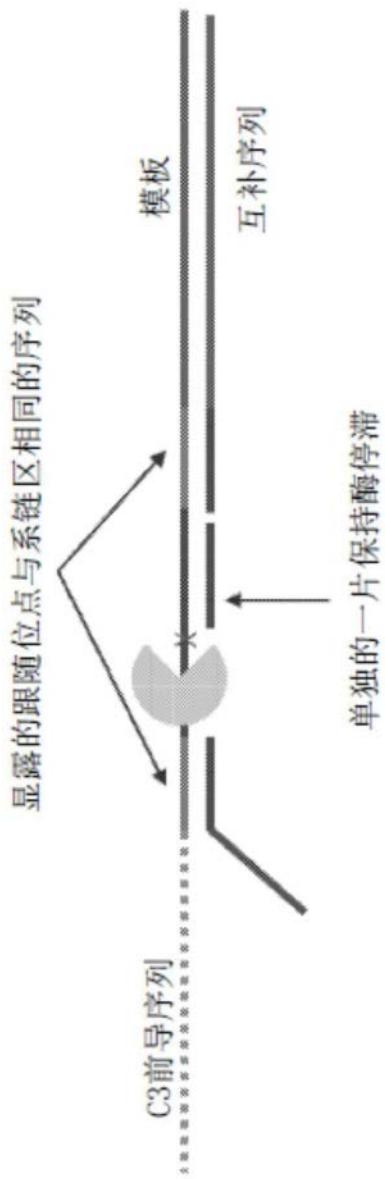


图20

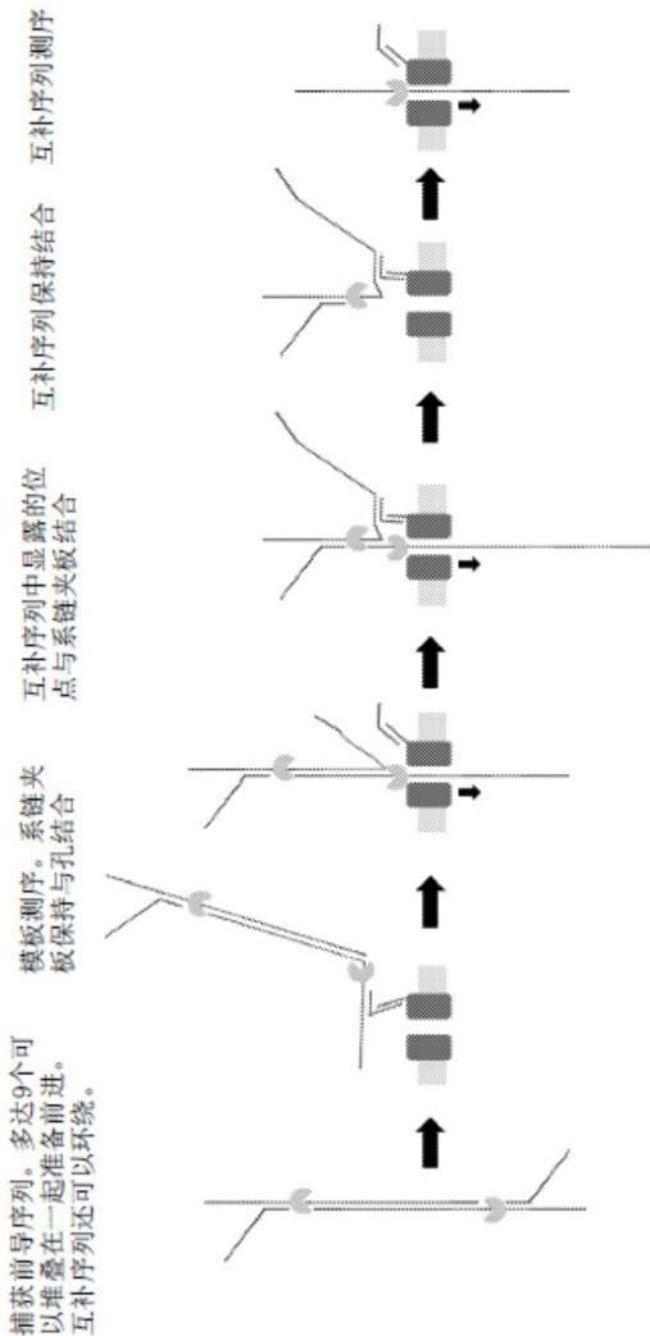


图21

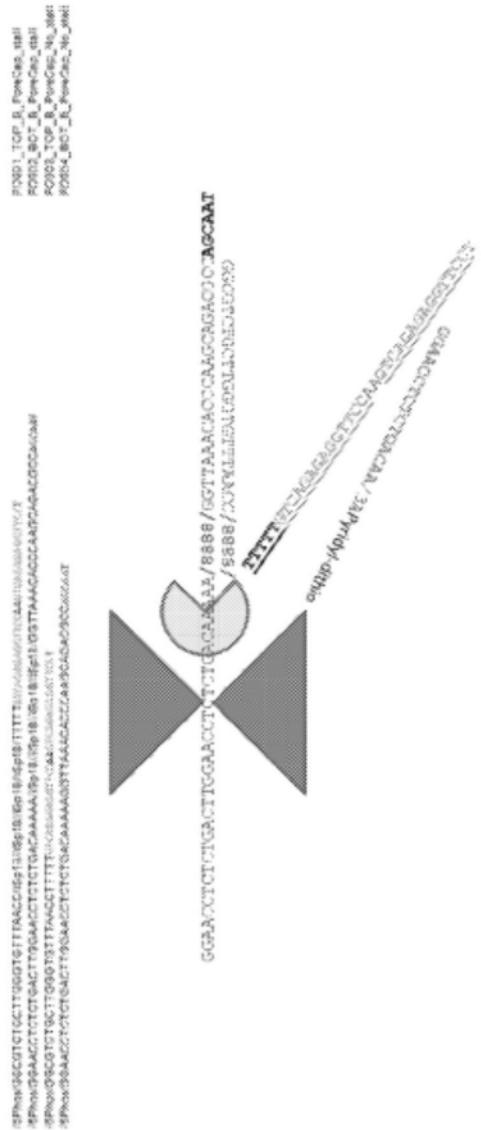


图22

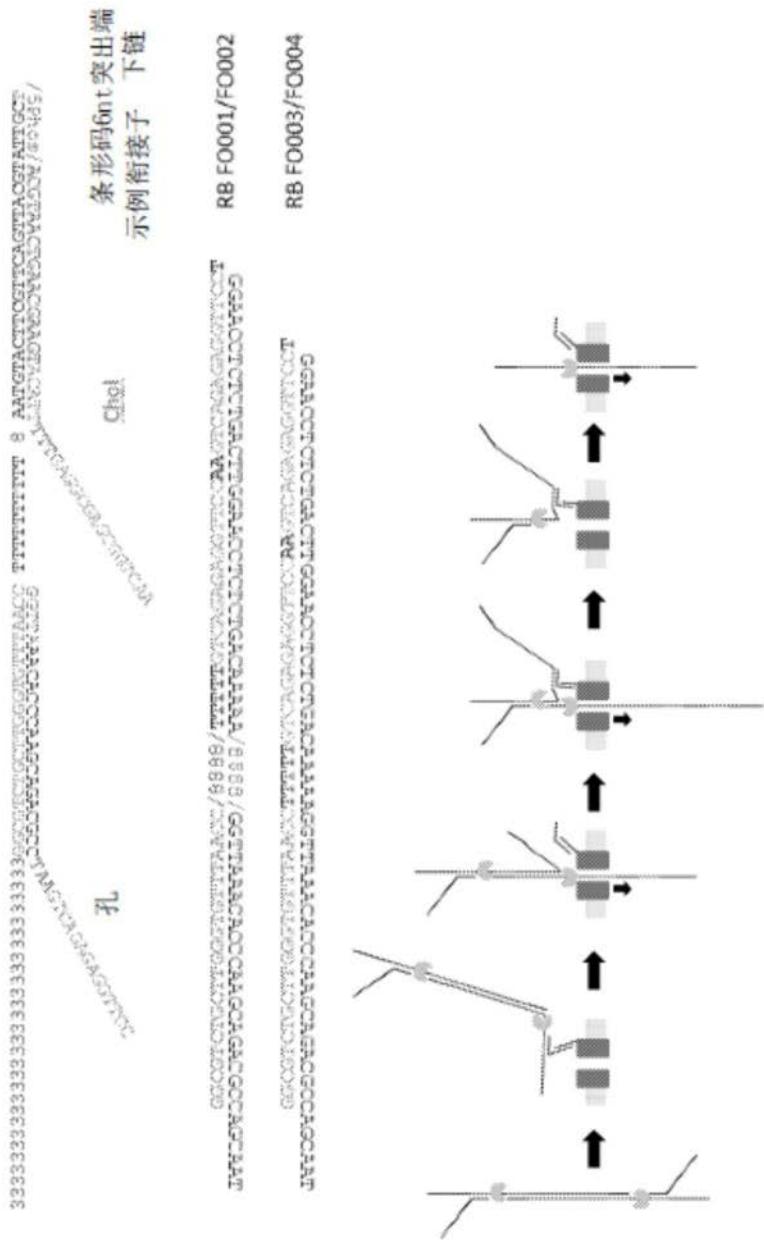


图23

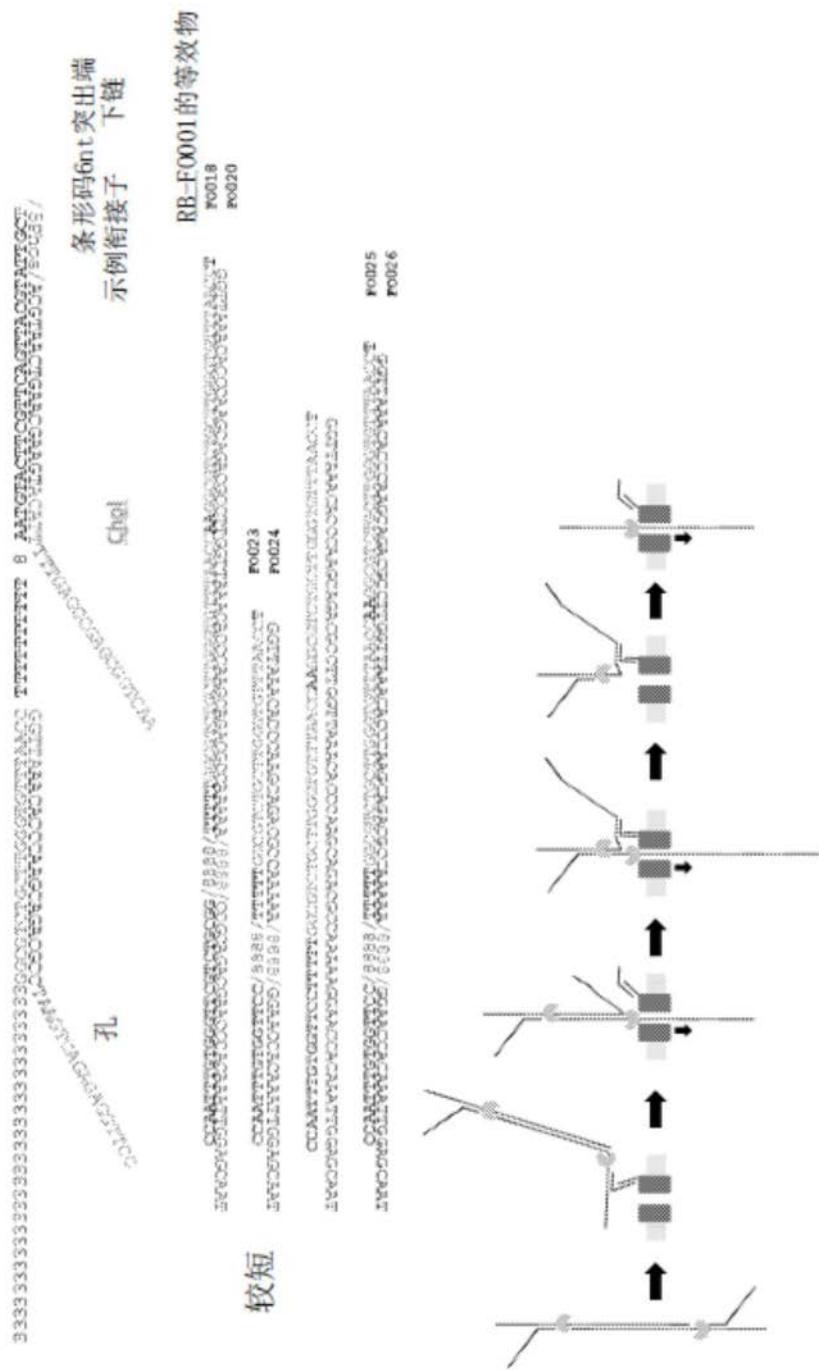


图24

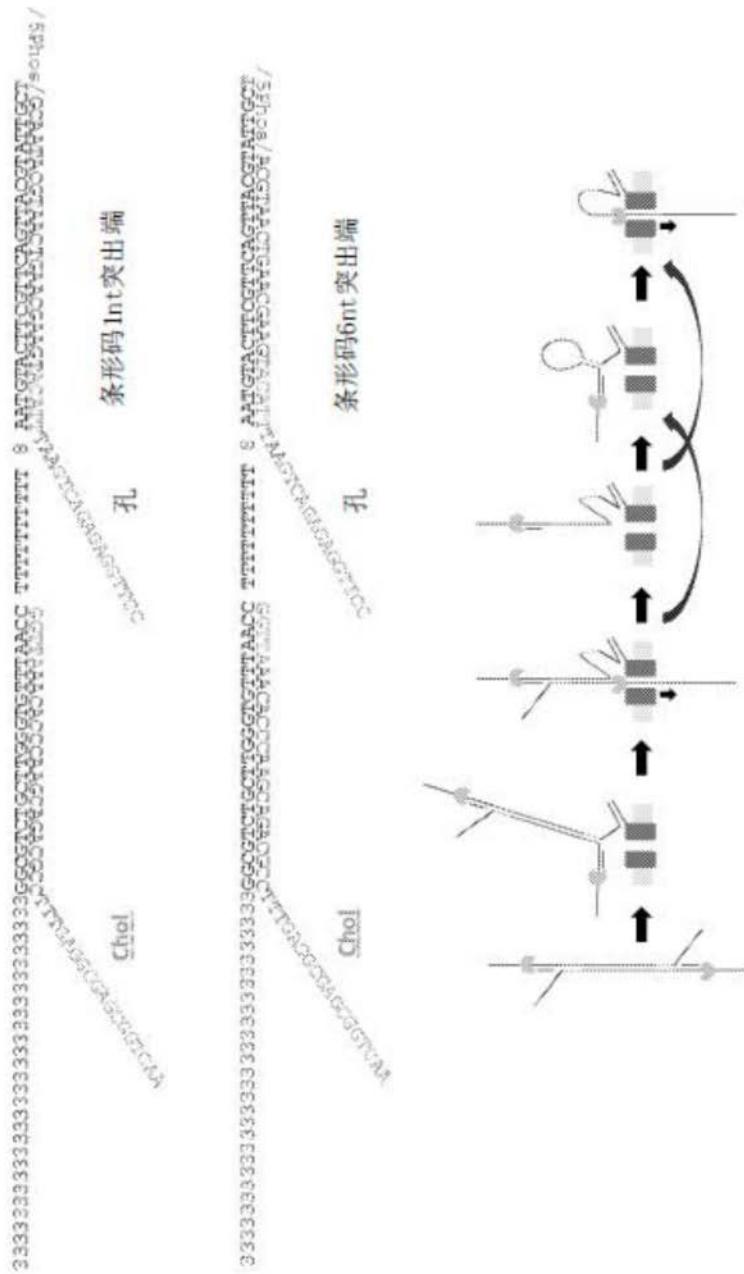


图25

图26

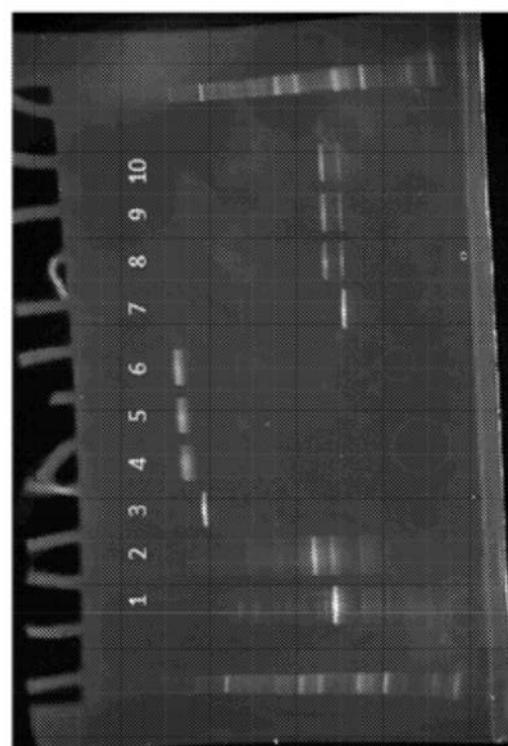
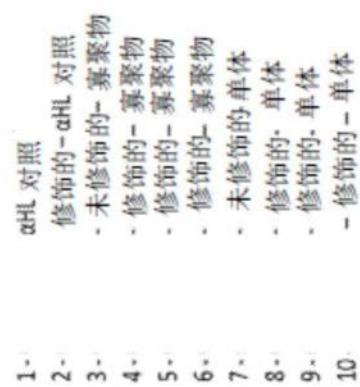


图27A

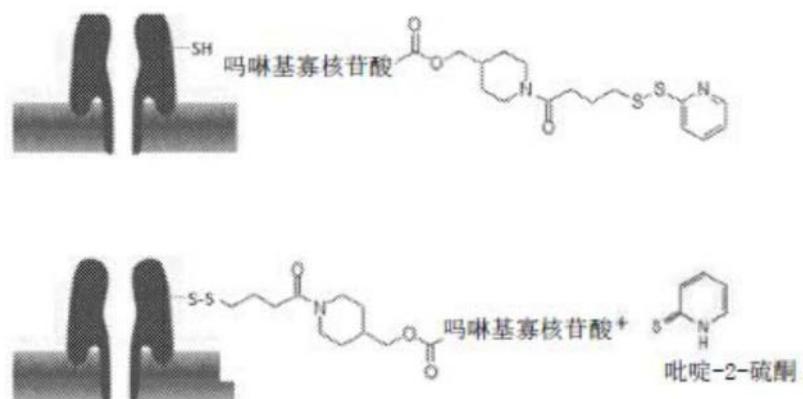


图27B

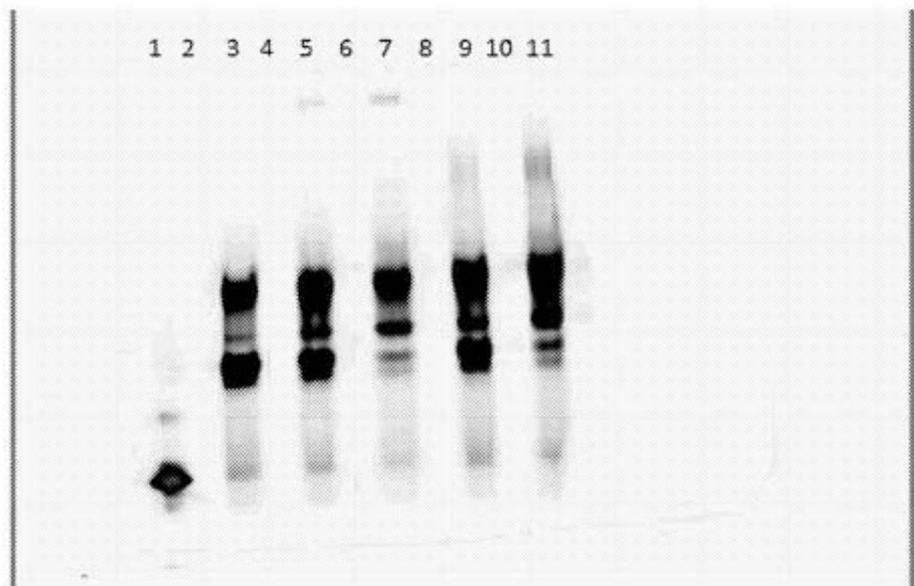


图28

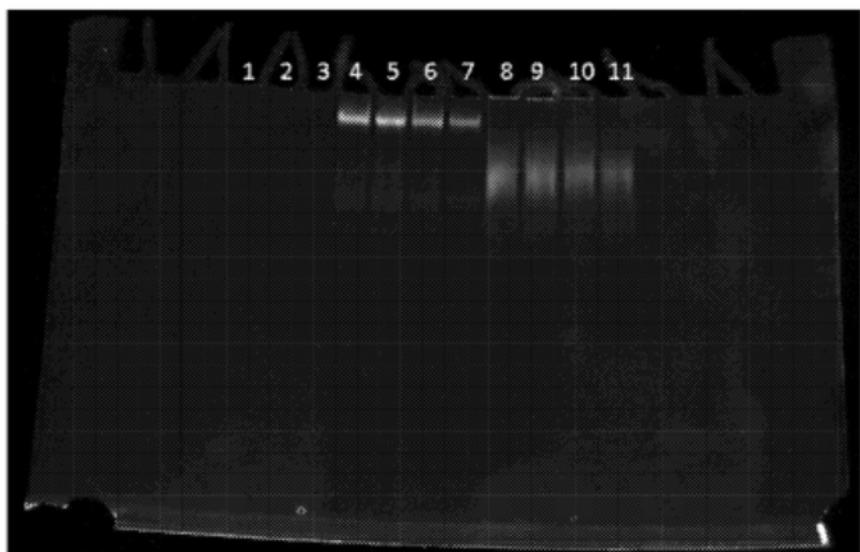


图29

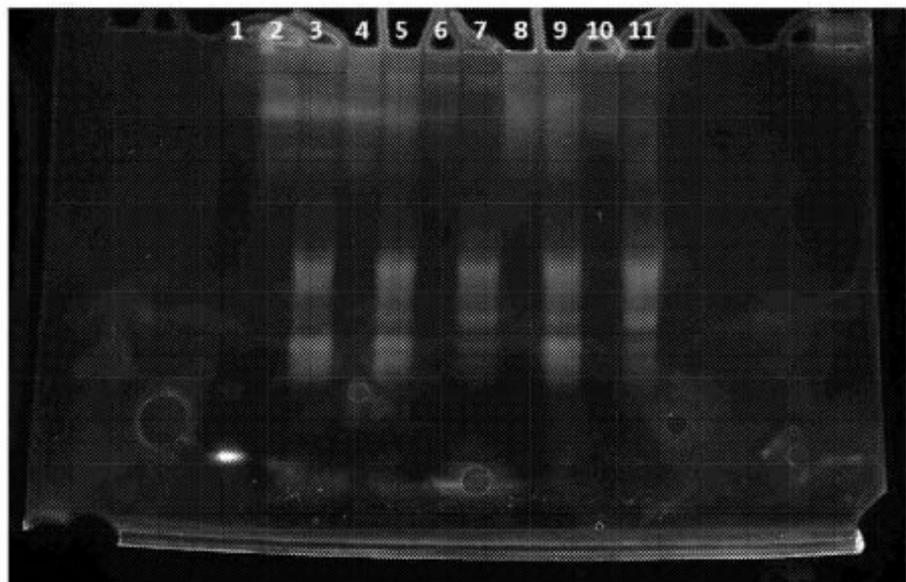


图30

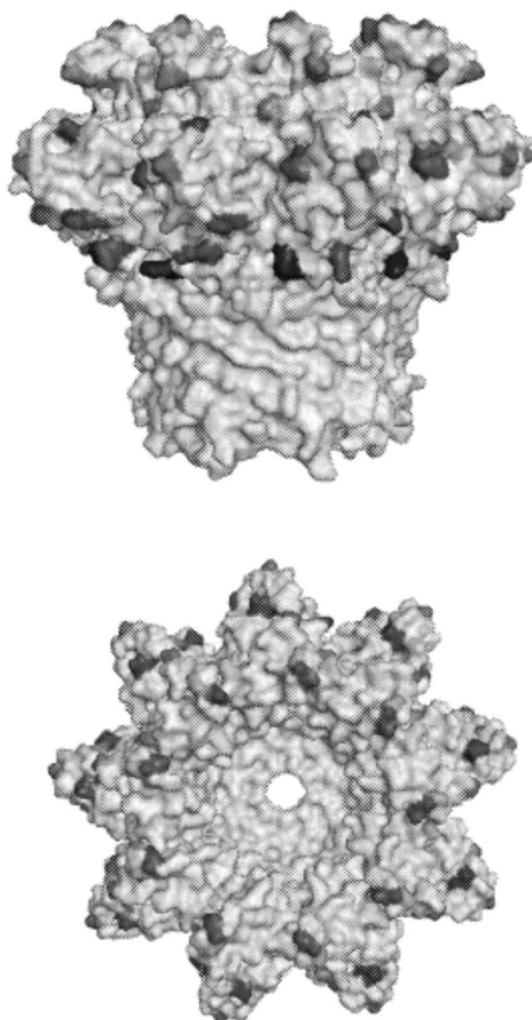


图31



图32A



图32B

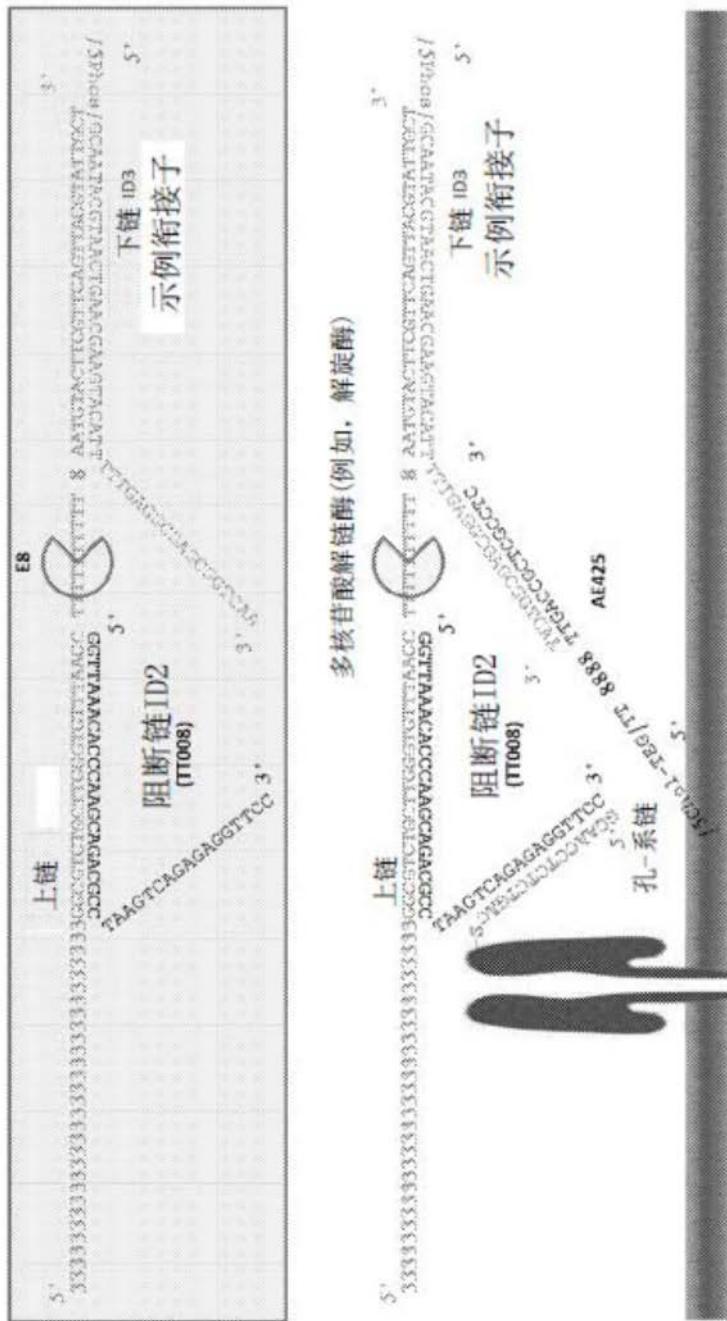


图33



图34A

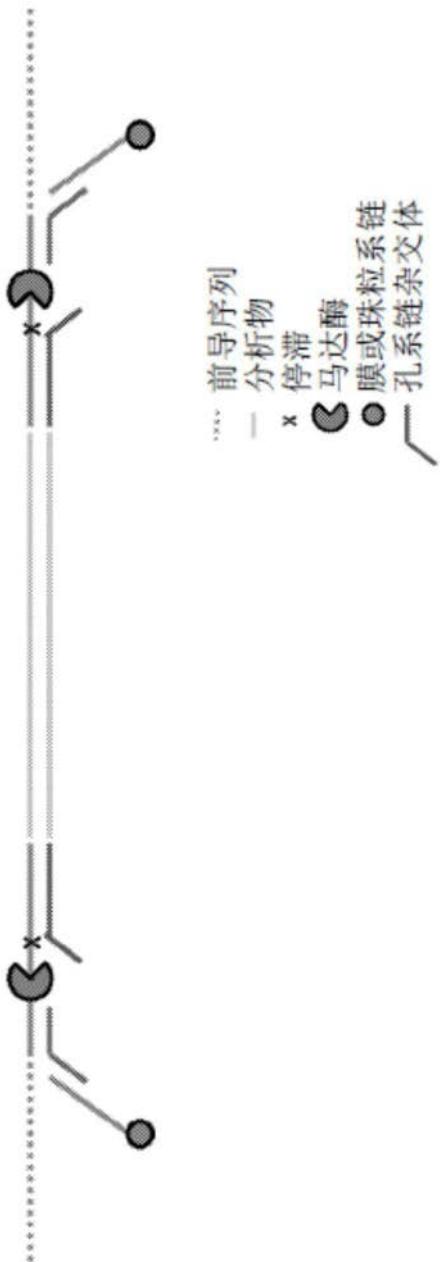


图34B

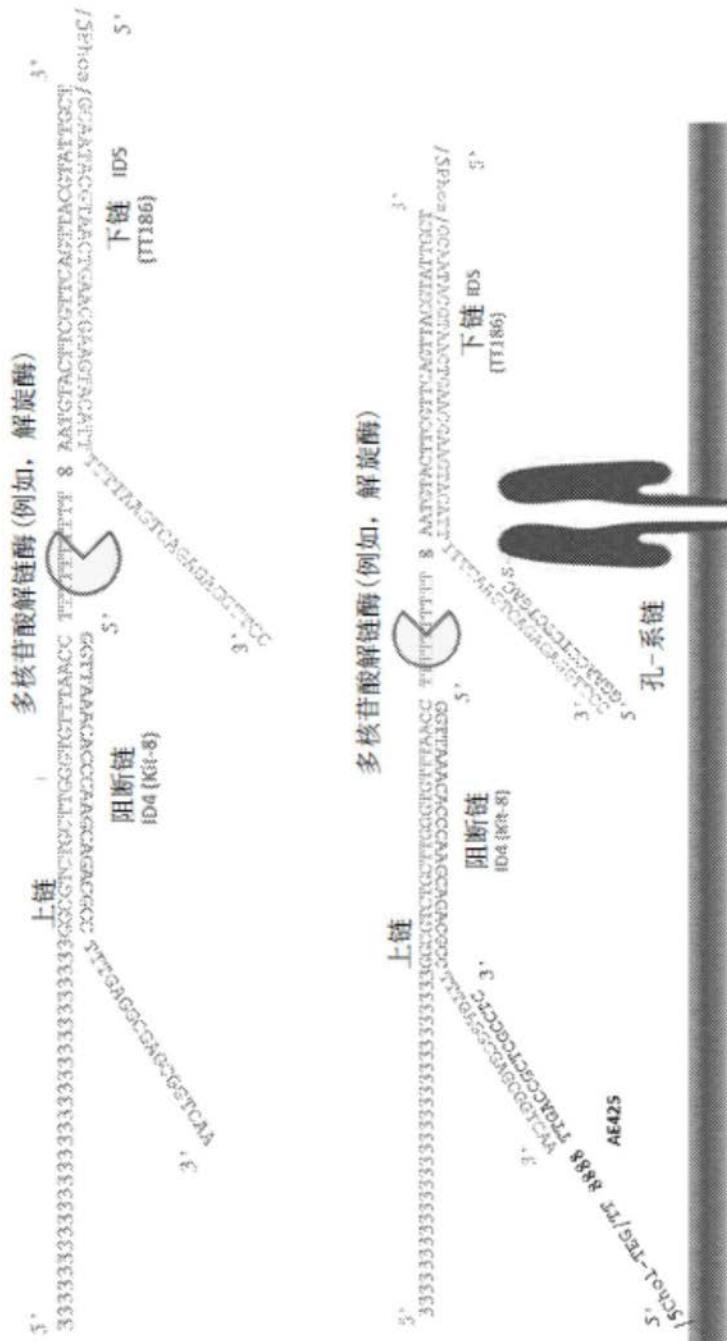


图35

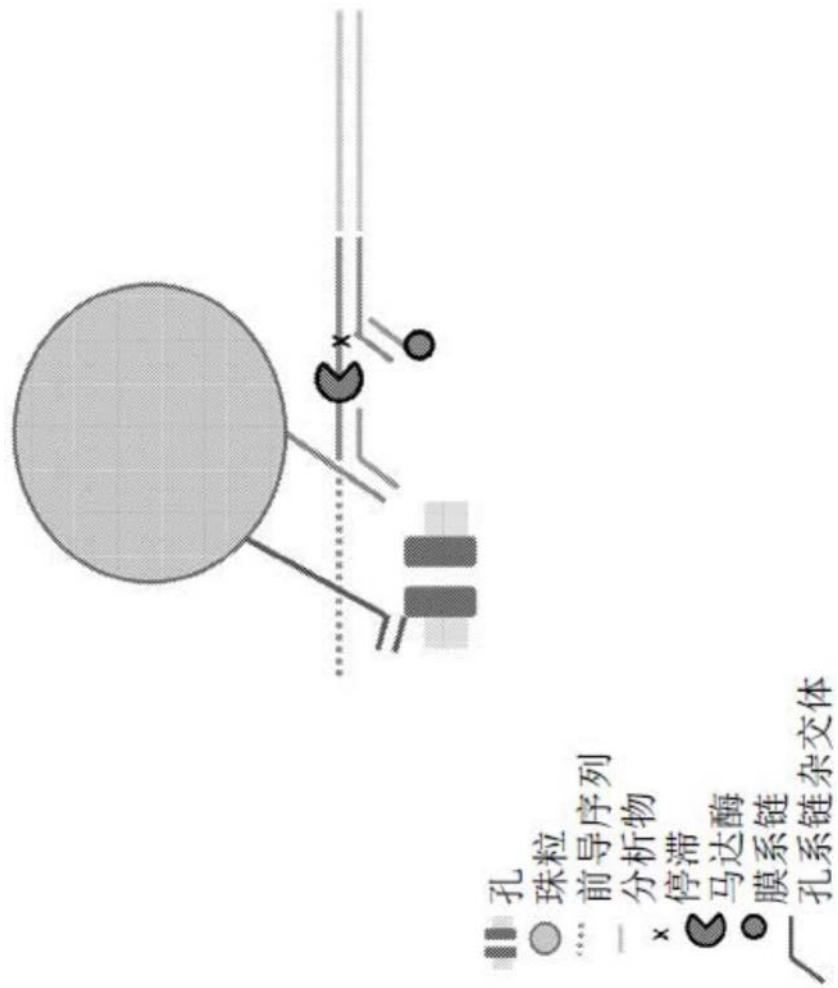


图36

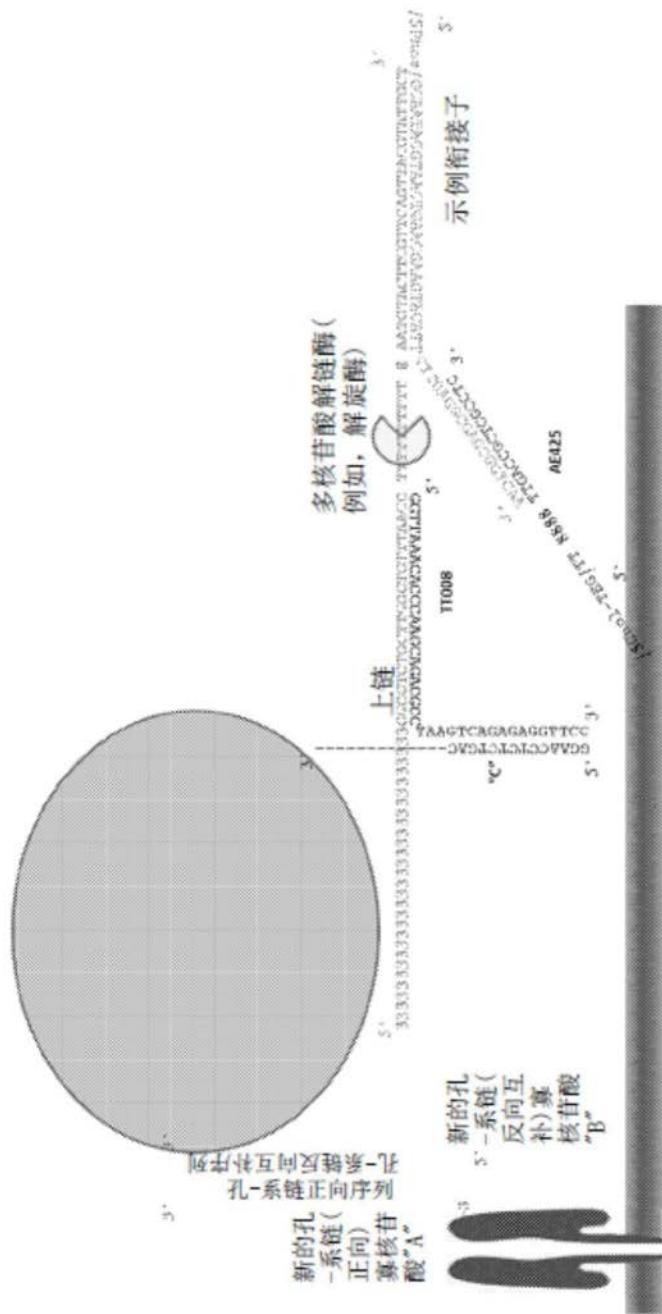


图37

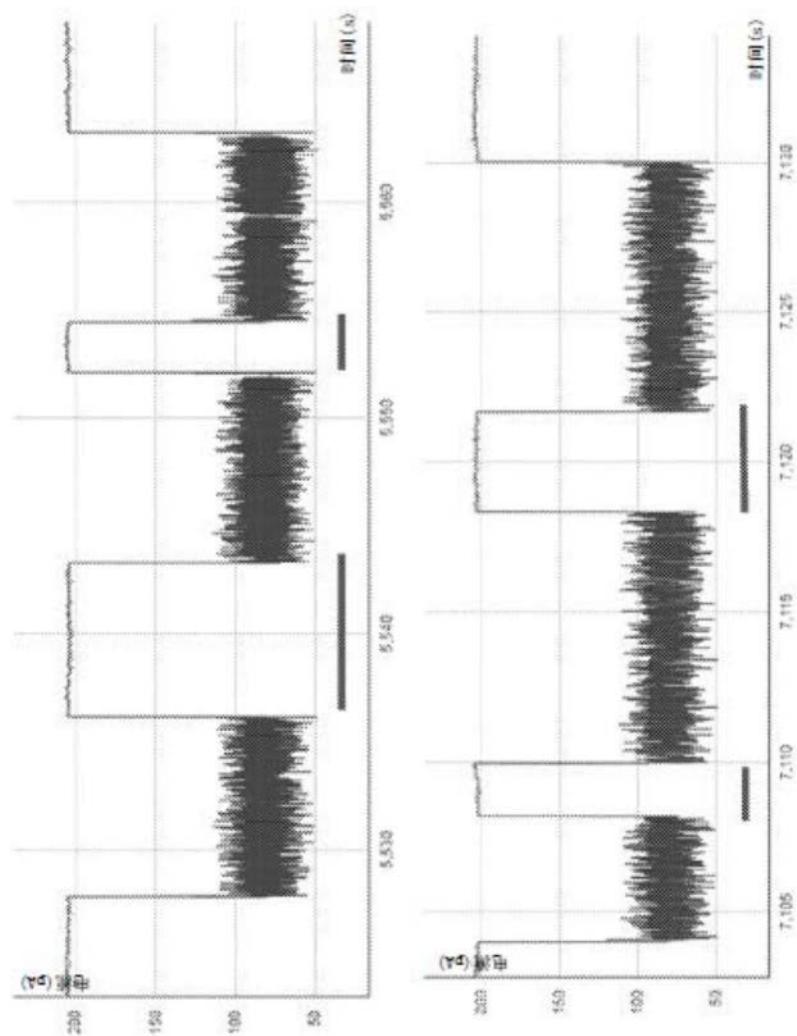


图38

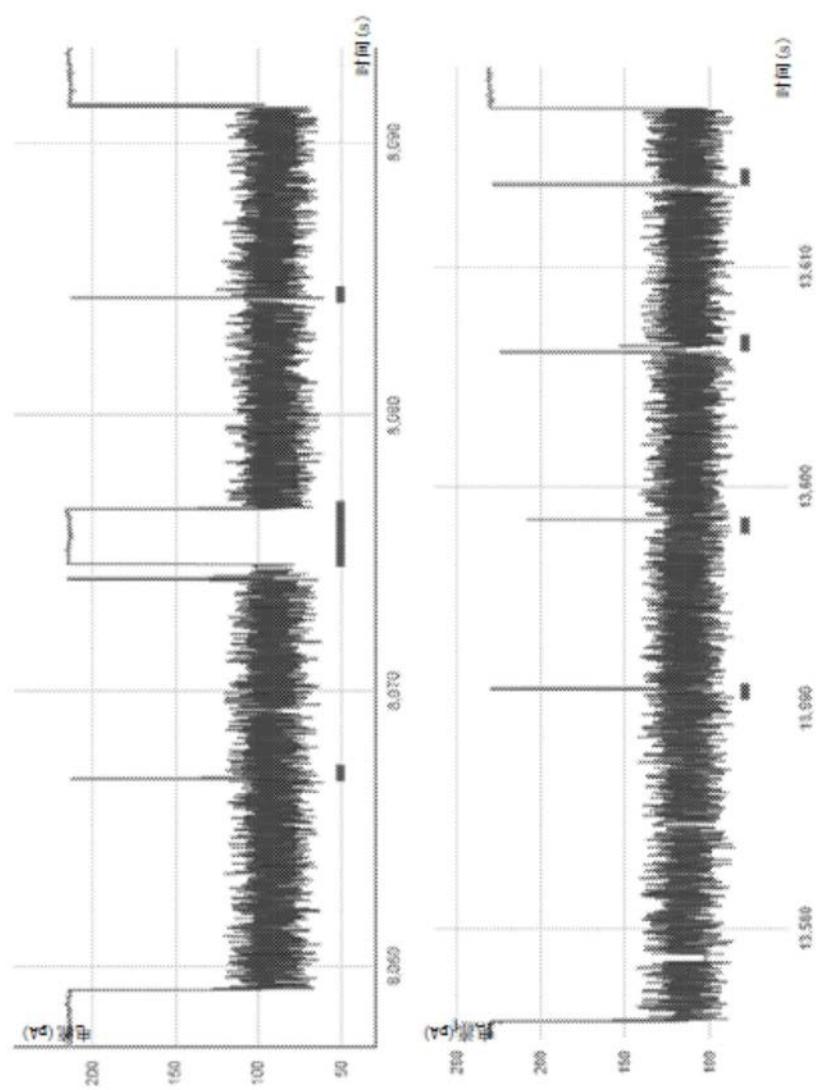


图39

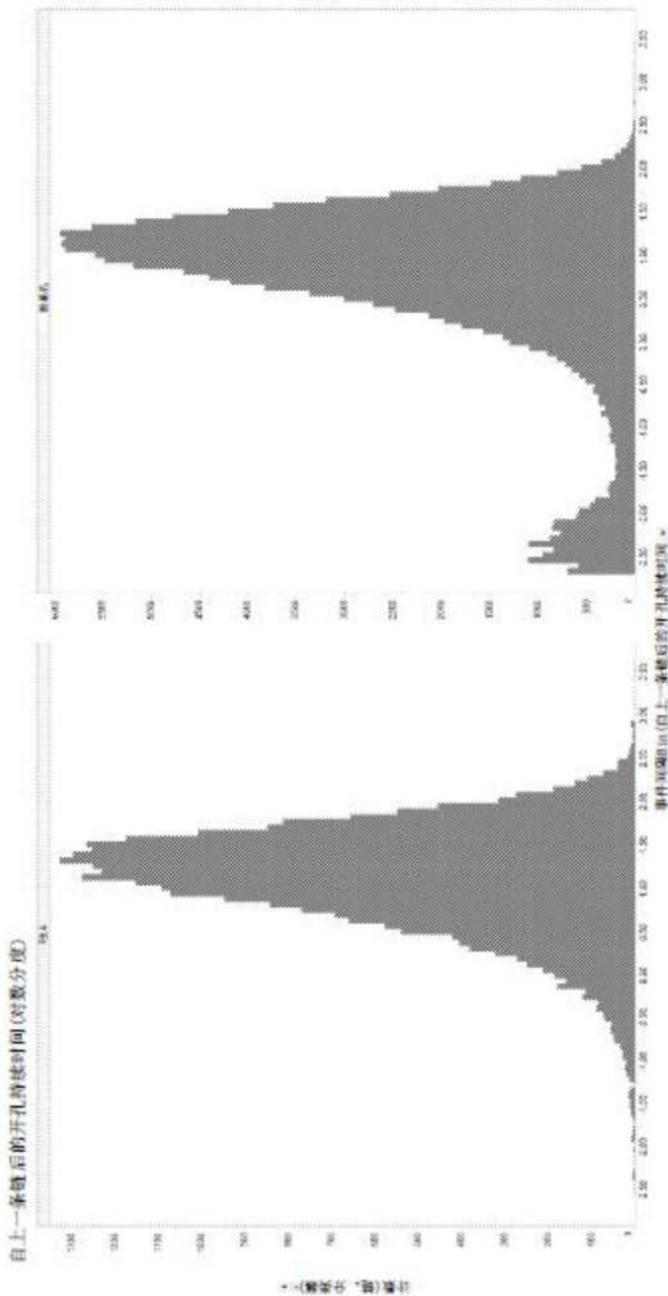


图40

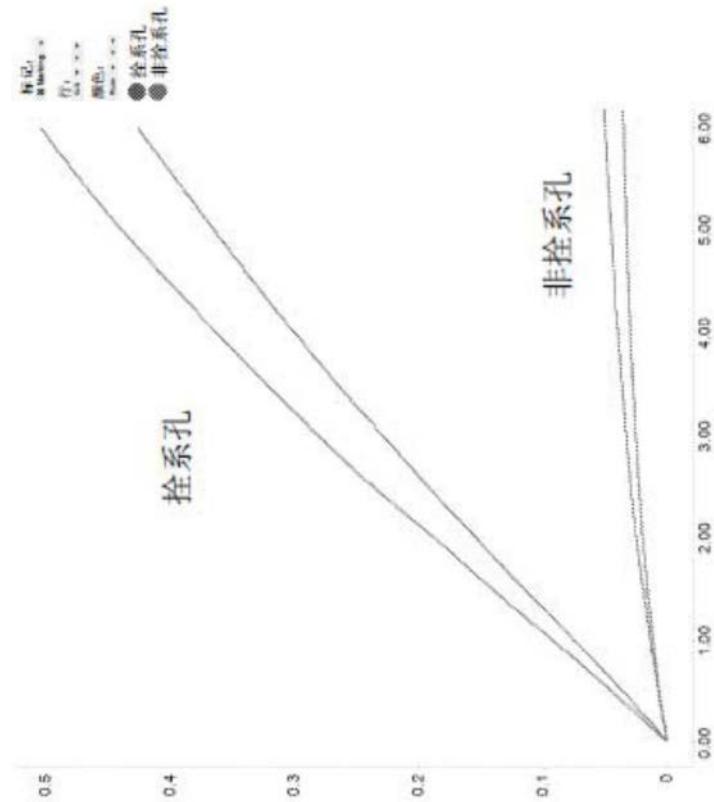


图41

样品	孔	链	测序孔	映射链	跟随链(%)	
					101	195
1	拴系	5299		2525	195	7.7%
2	拴系	5114	99	2482	161	6.5%
3	拴系	228185	409	20517	1187	5.76%
4	拴系	219216	454	22369	1391	6.22%
5	拴系	106680	164	76097	6908	9%
6	拴系	103487	208	74211	6045	8.1%
7	非拴	19719	299	11113	23	0.2%
8	非拴	12566	285	5483	53	0.96%
9	非拴	118773	341	12199	28	0.23%
10	非拴	69880	304	4998	2	0.04%

图42

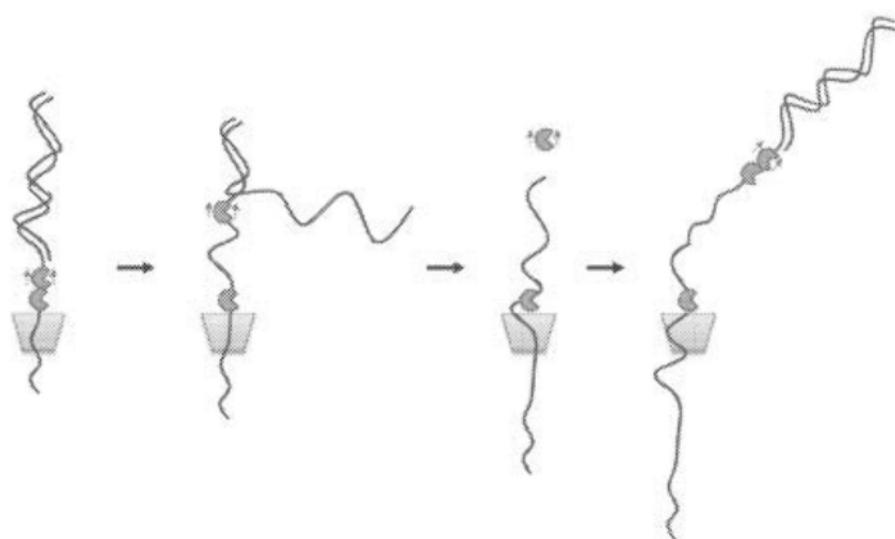


图43

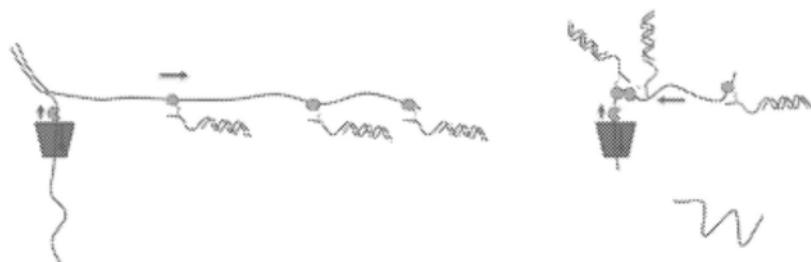


图44