



(12) 发明专利申请

(10) 申请公布号 CN 116200476 A

(43) 申请公布日 2023.06.02

(21) 申请号 202211048976.7

(22) 申请日 2017.03.02

(30) 优先权数据

1603656.8 2016.03.02 GB

1603657.6 2016.03.02 GB

1603658.4 2016.03.02 GB

(62) 分案原申请数据

201780015041.X 2017.03.02

(71) 申请人 牛津纳米孔科技公开有限公司

地址 英国牛津

(72) 发明人 拉科马·贾亚辛格

伊丽莎白·杰恩·华莱士

(74) 专利代理机构 北京同立钧成知识产权代理

有限公司 11205

专利代理师 刘文娜 黄健

(51) Int.Cl.

C12Q 1/6869 (2018.01)

G01N 33/487 (2006.01)

C07K 14/245 (2006.01)

C07K 19/00 (2006.01)

C12N 15/31 (2006.01)

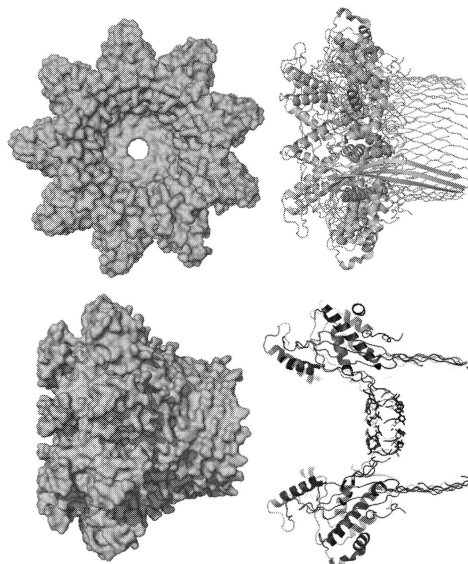
权利要求书2页 说明书127页
序列表(电子公布) 附图47页

(54) 发明名称

靶分析物测定方法、突变CsgG单体及其构筑体、及聚核苷酸和寡聚孔

(57) 摘要

本发明涉及靶分析物测定方法、突变CsgG单体及其构筑体、及聚核苷酸和寡聚孔。本发明涉及CsgG的突变形式。本发明还涉及使用CsgG的分析物检测和表征。



1. 一种用于测定靶分析物是否存在或其一或多个特征的方法,其包括:

(a) 使所述靶分析物与CsgG孔接触,使得所述靶分析物相对于所述孔移动,其中,所述孔包括突变CsgG单体,所述突变CsgG单体包括SEQ ID NO:2中所示序列的变异体,所述变异体包括K94N、K94Q、K94R、K94F、K94Y、K94W、K94L或K94S;和

(b) 在所述分析物相对于所述孔移动时,进行一个或多个测量,且从而测定所述分析物是否存在或其一或多个特征。

2. 根据权利要求1所述的方法,其中所述变异体包括K94Q或K94N。

3. 根据权利要求1所述的方法,其中所述变异体还包括:Q42K或Q42R;E44N或E44Q;L90R或L90K;N91R或N91K;I95R或I95K;A99R或A99K;E101H、E101K、E101N、E101Q或E101T;和/或Q114K。

4. 根据权利要求1所述的方法,其中所述变异体还包括以下的一者或多者:

(a) R93F、R93Y、R93W、R93L、R93I、R93V、R93N、R93Q或R93S;

(b) G103F、G103W、G103S、G103N、G103K或G103R;

(c) Y130W、Y130F、Y130H、Y130Q或Y130N;

(d) K135L、K135V、K135N、K135Q或K135S;

(e) S208V、S208I、S208F、S208W、S208Y、S208L或S208T;

(f) R142Q或R142S;

(g) R97F、R97Y、R97W、R97V、R97I、R97K、R97S、R97Q或R97H;

(h) N102K、N102Q、N102L、N102I、N102V、N102S或N102H;

(i) E44S;

(j) N91K或N91R;

(k) Q114R或Q114K;

(l) E101I、E101L、E101A或E101H;

(m) F48S、F48N、F48Q、F48Y、F48W、F48I、F48V、F48H、F48R或F48K;

(n) T150Y、T150A、T150V、T150L、T150S、T150Q或T150N。

5. 根据前述权利要求任一项所述的方法,其中所述变异体还包括以下的缺失:

-D195、Y196、Q197、R198和L199;

-R192、F193、I194、D195、Y196、Q197、R198、L199和L200;

-Q197、R198、L199和L200;

-I194、D195、Y196、Q197、R198和L199;

-D195、Y196、Q197、R198、L199和L200;

-Y196、Q197、R198、L199、L200和E201;

-Q197、R198、L199、L200和E201;

-Q197、R198、L199;或

-F193、I194、D195、Y196、Q197、R198和L199。

6. 根据前述权利要求任一项所述的方法,其中所述变异体还包括R192D、R192Q、R192F、R192S、R192T、R192N或R192E。

7. 根据前述权利要求任一项所述的方法,其中所述变异体包括R97W、R93W、或R93Y和R97Y;还包括F191T;或还包括V105、A106和I107的缺失。

8. 根据前述权利要求任一项所述的方法, 其中所述变异体包括以下的一者或多者: (i) Q42R或Q42K; (ii) K49R; (iii) N102R、N102F、N102Y或N102W; (iv) D149N、D149Q或D149R; (v) E185N、E185Q或E185R; (vi) D195N、D195Q或D195R; (vii) E201N、E201Q或E201R; (viii) E203N、E203Q或E203R; (ix) E44N、E44Q、E44R或E44K; (x) S54P; (xi) R97N、R97G或R97L; (xii) E101N、E101Q、E101R、E101K、E101F、E101Y或E101W; (xiv) T150I; 和 (xv) R192E或R192N。

9. 根据前述权利要求任一项所述的方法, 其中所述变异体还包括F56Q/N55S、F56N/Y51A、F56Q/Y51A、F56Q/Y51Q、F56Q/Y51T、F56A/Y51A、F56N/N55S/Y51A、F56A/N55S/Y51A或F56Q/N55S/Y51T。

10. 一种突变CsgG单体, 其包括在SEQ ID NO:2中所示序列的变异体, 所述变异体包括K94N/Q/R/F/Y/W/L/S, 且保持形成孔的能力。

11. 根据权利要求10所述的突变CsgG单体, 其中所述变异体为如权利要求2至9中的任一项所定义。

12. 一种构筑体, 其包括两个或更多个共价连接的CsgG单体, 其中所述单体中的至少一个是根据权利要求10或11所述的突变单体, 且其中所述单体是相同的或不同的。

13. 一种聚核苷酸, 其编码根据权利要求10或11所述的突变单体。

14. 一种同源寡聚孔, 其衍生自CsgG, 所述孔包括相同的根据权利要求10或11所述的突变单体或相同的根据权利要求12所述的构筑体。

15. 一种异源寡聚孔, 其衍生自CsgG, 所述孔包括至少一个根据权利要求10或11所述的突变单体或至少一个根据权利要求12所述的构筑体。

靶分析物测定方法、突变CsgG单体及其构筑体、及聚核苷酸和寡聚孔

[0001] 本申请为申请日为2017年03月02日、申请号为201780015041.X、发明名称为“突变孔”的中国专利申请的分案申请。

技术领域

[0002] 本发明涉及CsgG的突变形式。本发明还涉及使用CsgG的分析物检测和表征。

背景技术

[0003] 纳米孔感测是一种依赖于观测分析物分子与受体之间的个别结合或相互作用事件的感测方法。可通过在绝缘膜中放置纳米尺寸单孔和测量在分析物分子存在下通过孔的电压驱动的离子转运来产生纳米孔传感器。通过分析物的独特电流标志来揭露分析物的特性，值得注意的是，电流块的持续时间和程度以及电流电平的变化。此类纳米孔传感器是可商购的，如由牛津纳米孔科技有限公司(Oxford Nanopore Technologies Ltd)销售的MinION™装置，其包含与电子芯片集成的纳米孔阵列。

[0004] 目前仍需要横跨广泛应用范围的快速且廉价的核酸(例如DNA或RNA)测序技术。现有技术缓慢且昂贵，这主要是因为其依赖于扩增技术来产生大体积的核酸且需要大量的用于信号检测的专业荧光化学试剂。纳米孔感测具有通过降低核苷酸和所需的试剂的数量来提供快速且廉价的核酸测序的潜能。

[0005] 使用纳米孔感测进行核酸测序的两个基本组分是(1)控制通过孔的核酸移动和(2)在核酸聚合物通过孔移动时鉴别核苷酸。在过去，为了实现核苷酸鉴别，核酸已通过溶血素突变。这提供了已表明为序列依赖性的电流标志。还表明，大量的核苷酸引起当使用溶血素孔时所观测的电流，使得在所观测的电流与聚核苷酸问题之间具有直接关系。

[0006] 虽然用于核苷酸鉴别的电流范围已通过溶血素孔突变来改进，但如果核苷酸之间的电流差异可以进一步改进，那么测序系统将具有更高个性能。另外，已观测到，当核酸通过孔移动时，一些电流状态显示很高的变化。还表明，一些突变溶血素孔比其它的显示更高的变化。虽然这些状态的变化可含有序列特异性信息，但期望生产具有低变化的孔以简化系统。还期望降低引起所观测电流的核苷酸数量。

发明内容

[0007] 本发明人已出人意料地证实，CsgG和其新颖突变可用于表征分析物，如聚核苷酸。本发明涉及突变CsgG单体。本发明人已出人意料地证实，包含新颖突变单体的孔具有增强的评估分析物特征的能力，如聚核苷酸的序列。突变孔出人意料地提供靶聚核苷酸相对于，如通过孔更加一致移动。突变孔出人意料地显示改进的表征准确度。特定来说，突变孔出人意料地显示电流范围增加和状态变化降低，所述增加的电流范围使得更容易在不同核苷酸之间进行鉴别，所述状态变化降低增加信噪比。另外，突变孔出人意料地更容易捕获核苷酸和聚核苷酸。

[0008] 除非相反地陈述,否则本文所公开的所有氨基酸取代、缺失和/或添加均是参考包含SEQ ID NO:2中所示序列的变异体的突变CsgG单体。

[0009] 提及包含SEQ ID NO:2中所示序列的变异体的突变CsgG单体,涵盖包含如在如下文所公开的其它SEQ ID NO中陈述的序列变异体的突变CsgG单体。可对包含除SEQ ID NO:2中所示以外的序列变异体的CsgG单体,进行等效于本文所公开的参考包含SEQ ID NO:2中所示序列的变异体的突变CsgG单体的那些取代、缺失和/或添加的氨基酸取代、缺失和/或添加。

[0010] 突变单体可被视为经分离单体。

[0011] 在一个实施例中,本发明涉及突变CsgG单体,其中在位置192处的精氨酸(R)已被天冬氨酸(D)、谷氨酰氨(Q)、苯丙氨酸(F)、丝氨酸(S)或苏氨酸(T)取代。本发明人已出人意料地证实,此类单体且特定来说包含R192D取代的单体比在192位置处无取代的单体容易表达许多。

[0012] 在一个实施例中,本发明涉及突变CsgG单体,其中在位置94处的赖氨酸(K)已被天冬氨酸(N)、谷氨酰氨(Q)、精氨酸(R)、苯丙氨酸(F)、酪氨酸(Y)、色氨酸(W)、亮氨酸(L)或丝氨酸(S)取代。本发明人已出人意料地证实,此类单体,且特定来说包含K94Q或K94N取代的单体,比在位置94处包含赖氨酸的其它相同突变的噪声少。

[0013] 因此,本发明提供一种突变CsgG单体,其包含SEQ ID NO:2中所示序列的变异体,其中变异体包含:(a)以下中的一或多个:K94N/Q/R/F/Y/W/L/S、D43S、E44S、F48S/N/Q/Y/W/I/V/H/R/K、Q87N/R/K、N91K/R、K94N/Q/R/F/Y/W/L/S/N、R97F/Y/W/V/I/K/S/Q/H、E101I/L/A/H、N102K/Q/L/I/V/S/H、R110F/G/N、Q114R/K、R142Q/S、T150Y/A/V/L/S/Q/N、R192D/Q/F/S/T和D248S/N/Q/K/R;和/或(b),在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变):I41、R93、A98、Q100、G103、T104、A106、I107、N108、L113、S115、T117、Y130、K135、E170、S208、D233、D238、E244、Q42、E44、L90、N91、I95、A99、E101和Q114。

[0014] 本发明还提供一种突变CsgG单体,其包含SEQ ID NO:2中所示序列的变异体,其中所述变异体包含:(a)以下一或多个位置的缺失:R192、F193、I194、D195、Y196、Q197、R198、L199、L200和E201;和/或(b)以下中的一或多个的缺失:V139/G140/D149/T150/V186/Q187/V204/G205、G137/G138/Q151/Y152/Y184/E185/Y206/T207和A141/R142/G147/A148/A188/G189/G202/E203。

[0015] 突变CsgG单体可包含SEQ ID NO:2中所示序列的变异体,其中变异体包含:(a) K94Q、F191T和R192D/Q/F/S/T/N/E中的一或多个;和/或(b) V105、A106和I107的缺失。

[0016] 突变CsgG单体可以是:

[0017] -突变CsgG单体,其包含SEQ ID NO:2中所示序列的变异体的突变CsgG单体,所述突变CsgG单体包含R97W;

[0018] -包含SEQ ID NO:2中所示序列的变异体的突变CsgG单体,所述突变CsgG单体包含R192D/Q/F/S/T;

[0019] -突变CsgG单体,其包含(a) R192D; (b) R97W/Y和/或R93W/Y,优选地R97W、R93W或R93Y和R97Y; (c) K94Q/N; (d) G103K/R和/或T104K/R; 和/或(e) F191T, V105、A106和I107的缺失和/或F193、I194、D195、Y196、Q197、R198和L199的缺失。

[0020] 突变CsgG单体优选地进一步包含Y51A和F56Q。

[0021] 由本发明提供的特定突变CsgG单体包含SEQ ID NO:2中所示序列的变异体,所述变异体包含以下突变:

[0022] (1) Y51A、F56Q和R192D;

[0023] (2) Y51A、F56Q和R97W。

[0024] (3) Y51A、F56Q、R192D和R97W;

[0025] (4) Y51A、F56Q、R192D和R93W;

[0026] (5) Y51A、F56Q、R192D、R93Y和R97Y;或

[0027] (6) Y51A、F56Q、R192D和R93W。

[0028] (7) 根据(1) - (6)中任一项所述的突变,和:

[0029] (a) V105、A106和I107的缺失。

[0030] (b) K94Q或K94N;

[0031] (c) D195、Y196、Q197、R198和L199的缺失或F193、I194、D195、Y196、Q197、R198和L199的缺失;和/或

[0032] (d) F191T。

[0033] (8) 根据(1) - (6)中任一项所述的突变,和:

[0034] (i) K94Q和V105、A106和I107的缺失;

[0035] (ii) K94N和V105、A106和I107的缺失;

[0036] (iii) F191T和V105、A106和I107的缺失;

[0037] (iv) K94Q和F191T;

[0038] (v) K94N和F191T;

[0039] (vi) K94Q、F191T和V105、A106和I107的缺失;或

[0040] (vii) K94N、F191T和V105、A106和I107的缺失。

[0041] (9) 根据(1) - (8)中任一项所述的突变,和:

[0042] -T104K或T104R;

[0043] -L90R;

[0044] -N91R;

[0045] -I95R;

[0046] -A99R;

[0047] -E101K、E101N、E101Q、E101T或E101H;

[0048] -E44N或E44Q;和/或

[0049] -Q42K。

[0050] 本发明还提供:

[0051] -一种构筑体,其包含两个或更多个共价连接的CsgG单体,其中所述单体中的至少一个是本发明的突变单体;

[0052] -一种聚核苷酸,其编码本发明的突变单体或本发明的构筑体;

[0053] -一种同源寡聚孔,其衍生自CsgG,所述同源寡聚孔包含相同的本发明的突变单体或相同的本发明的构筑体;

[0054] -一种异源寡聚孔,其衍生自CsgG,所述异源寡聚孔包含至少一个本发明的突变单体或至少一个本发明的构筑体;

- [0055] -一种用于测定靶分析物是否存在或其一或多个特征的方法,其包含:
- [0056] (a)使靶分析物与本发明的孔接触,使得靶分析物相对于孔移动;和
- [0057] (b)在所述分析物相对于所述孔移动时,进行一或多个测量,且从而测定所述分析物是否存在或其一或多个特征;
- [0058] -一种形成用于表征靶聚核苷酸的传感器的方法,其包含在本发明的孔与聚核苷酸结合蛋白之间形成复合体且从而形成用于表征靶聚核苷酸的传感器;
- [0059] -一种用于表征靶聚核苷酸的传感器,其包含本发明的孔与聚核苷酸结合蛋白之间的复合体;
- [0060] -一种本发明的孔的用途,其用于测定靶分析物是否存在或其一或多个特征;
- [0061] -一种用于表征靶分析物的试剂盒,其包含(a)本发明的孔和(b)膜组分;
- [0062] -一种用于表征样本中的靶分析物的设备,其包含(a)多个本发明的孔和(b)多个膜;
- [0063] -一种表征靶聚核苷酸的方法,其包含:
- [0064] a)使聚核苷酸与本发明的孔、聚合酶和标记核苷酸接触,使得磷酸标记物种通过聚合酶依序添加到靶聚核苷酸中,其中磷酸物种含有对各核苷酸具有特异性的标记;和
- [0065] b)使用所述孔检测所述磷酸标记物种,且从而表征所述聚核苷酸;和
- [0066] -一种生产本发明的突变单体或本发明的构筑体的方法,其包含在合适宿主细胞中表达本发明的聚核苷酸且从而生产本发明的突变单体或构筑体。

附图说明

- [0067] 图1:示出来自大肠杆菌(E.coli)的CsgG。
- [0068] 图2:示出CsgG的尺寸。
- [0069] 图3:示出在10 Å/ns下的单G易位。对于鸟嘌呤进入到CsgG-Eco的F56环中存在大屏障。*=G进入F56环。A=G停止与56环相互作用。B=G停止与55环相互作用。C=G停止与51环相互作用。
- [0070] 图4:示出在100 Å/ns下的ssDNA易位。通过收缩CsgG-Eco来牵拉DNA需要很大的力。
- [0071] 图5:示出在10 Å/ns下的ssDNA易位。CsgG-F56A-N55S和CsG-F56A-N55S-Y51A突变两个具有较低的ssDNA易位屏障。
- [0072] 图6(a)-(c)到图8(a)-(c):显示与野生型(WT)相比范围增加的突变孔。
- [0073] 图9(a)-(c)和图10(a)-(b):显示与野生型(WT)相比输送量增加的突变孔。
- [0074] 图11和12:显示与野生型(WT)相比插入增加的突变孔。
- [0075] 图13:显示实例2中使用的DNA构筑体X。标记1的区对应于30个SpC3间隔子。标记2的区对应于SEQ ID NO:42。标记3的区对应于四个iSp18间隔子。标记4的区对应于SEQ ID NO:43。标记5的区段对应于四个5-硝基吡啶。标记6的区对应于SEQ ID NO:44。标记7的区对应于SEQ ID NO:45。标记8的区对应于SEQ ID NO:46,其具有在SEQ ID NO:46的3'端处连接的四个iSp18间隔子(标记9的区)。在iSp18间隔子的相对端是3'胆固醇系链(标记10)。标记11的区对应于四个SpC3间隔子。
- [0076] 图14:显示CsgG蛋白的Strep trap(通用电气医疗集团(GE Healthcare))纯化的

实例色谱迹线(x轴标记=洗脱体积(mL)、Y轴标记=吸光度(mAu))。将样本加载到25mM Tris、150mM NaCl、2mM EDTA、0.01%DDM中,且用10mM脱硫生物素洗脱。其中CsgG蛋白洗脱的洗脱峰标记为E1。

[0077] 图15:显示在初始strep纯化之后CsgG蛋白的典型SDS-PAGE观测的实例。在300V下在1×TGS缓冲液中进行4-20%TGX凝胶(伯乐公司(Bio Rad)电泳22分钟。用宝石红染色剂对凝胶进行染色。泳道1-3显示含有如由箭头指示的CsgG蛋白的主要洗脱峰(图14中的标记E1)。泳道4-6对应于含有杂质的主要洗脱峰(图14中的标记E1)尾部的洗脱份。M显示所使用的分子量标记,所述分子量标记是未染色Novex Sharp(单位=kD)。

[0078] 图16:显示CsgG蛋白的尺寸排阻色谱图(SEC)的实例(120mL S200(通用电气医疗集团),x轴标记=洗脱体积(mL)、y轴标记=吸光度(mAu))。在strep纯化和加热蛋白质样本之后,进行SEC。SEC的电泳缓冲液是25mM Tris、150mM NaCl、2mM EDTA、0.01%DDM、0.1%SDS、pH 8.0,且在1毫升/分钟速率下进行色谱柱。标记X的迹线显示在220nm下的吸光度,且标记Y的迹线显示在280nm下的吸光度。收集标记为星形的峰。

[0079] 图17:显示在SEC之后CsgG蛋白的典型SDS-PAGE观测的实例。在300V下在1×TGS缓冲液中进行4-20%TGX凝胶(伯乐公司)电泳22分钟,且用宝石红宝染色剂对凝胶进行染色。泳道1显示在strep纯化和加热之后但在SEC之前的CsgG蛋白样本。泳道2-8显示跨大致48mL-60mL图16(中间峰=55mL)电泳且在图16中用星形标记的峰收集的级分。M显示所使用的分子量标记,所述分子量标记是未染色Novex Sharp(单位=kD)。对应于CsgG-Eco孔的条棒由箭头指示。

[0080] 图18到24:显示与野生型(WT)相比范围增加的突变孔。

[0081] 图25到30:显示与野生型(WT)相比范围增加的突变孔。

[0082] 图31显示在模拟期间于0和20ns获取(1到3回合),在孔(CsgG-Eco-(Y51T/F56Q)-StrepII(C))9(具有突变Y51T/F56Q的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47且在C端处连接,孔突变第20号)上的酶(T4 Dda-(E94C/C109A/C136A/A360C)(具有突变E94C/C109A/C136A/A360C且接着(ΔM1)G1G2的SEQ ID NO:24))的快照。

[0083] 图32显示在模拟期间于30和40ns获取(1到3回合),在孔(CsgG-Eco-(Y51T/F56Q)-StrepII(C))9(具有突变Y51T/F56Q的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47且在C端处连接,孔突变第20号)上的酶(T4 Dda-(E94C/C109A/C136A/A360C)(具有突变E94C/C109A/C136A/A360C且接着(ΔM1)G1G2的SEQ ID NO:24))的快照。

[0084] 图33显示在实例5中描述的模拟期间获取,在孔CsgG-Eco-(Y51A/F56Q/R97W)-StrepII(C))9(具有突变Y51A/F56Q/R97W的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47且在C端处连接,孔突变第26号)上的酶(T4 Dda-(E94C/F98W/C109A/C136A/K194L/A360C)(具有突变E94C/F98W/C109A/C136A/K194L/A360C且接着(ΔM1)G1G2的SEQ ID NO:24))的快照。

[0085] 图34显示电流迹线的两个十秒截屏,其显示在无酶控制的情况下通过MspA突变x=MspA-(De1-L74/G75/D118/L119)D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8(SEQ ID NO:50,其具有突变D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K和氨基酸L74/G75/D118/L119的缺失)的DNA(SEQ ID NO:51)的易位。

[0086] 图35显示电流迹线的两个十秒截屏,其显示在无酶控制的情况下通过CsgG-Eco-

(Y51A/F56Q/R97W/R192D-StrepII(C))9(具有突变Y51A/F56Q/R97W/R192D的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47且在C端处连接)的DNA(SEQ ID NO:51)的易位。

[0087] 图36显示电流迹线的两个十秒截屏,其显示在无酶控制的情况下通过CsgG-Eco-(Y51A/F56Q/R97W/E101S/R192D-StrepII(C))9(具有突变Y51A/F56Q/R97W/E101S/R192D的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47且在C端处连接)的DNA(SEQ ID NO:51)的易位。

[0088] 图37显示CsgG突变孔的两个凝胶过滤色谱图(120ml S200色谱柱)的重叠,所述CsgG突变孔:A)CsgG-Eco-(Y51A/F56Q/R97W)-StrepII(C))9(具有突变Y51A/F56Q/R97W的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47且在C端处连接);和B)CsgG-Eco-(Y51A/F56Q/R97W/R192D)-StrepII(C))9(具有突变Y51A/F56Q/R97W/R192D的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47且在C端处连接)。CsgG-Eco-(Y51A/F56Q/R97W)-StrepII(C))9在A280下的吸光度标记为A,且CsgG-Eco-(Y51A/F56Q/R97W/R192D)-StrepII(C))9在A280下的吸光度标记为B。使两个构筑体均在500ml培养基中生长。使用加载到上色谱柱完全相同的方案和相同的体积来对两个蛋白质进行表达和纯化。电泳缓冲液是25mM Tris、150mM NaCl、2mM EDTA、0.01%DDM、0.1%SDS pH8。CsgG-Eco-(Y51A/F56Q/R97W)-StrepII(C))9孔的级分延迟是归因于在AKTA纯化器10上使用的不同连接配置。吸光度值的差异指示所表达的蛋白质的量,其中吸光度值越高指示所表达的蛋白质的量越高。

[0089] 图38显示CsgG纳米孔的SDS-PAGE分析。泳道A-C含有CsgG-Eco-(Y51A/F56Q/R97W)-StrepII(C))9;泳道D-F含有CsgG-Eco-(Y51A/F56Q/R97W/R192D)-StrepII(C))9;和泳道M含有分子量标记。使用完全相同的方案来表达和纯化两个孔。在4-20%TGX凝胶(伯乐公司,目录号5671093)上在TGS缓冲液中进行电泳22分钟。用宝石红染色剂(生命技术公司(Life Technologies),目录号S1200)对凝胶进行染色。将相同体积的各孔样本加载到凝胶上,以比较在纯化之后所获得的蛋白质的量——泳道A和D含有5μL,泳道B和E含有10μL且泳道C和F含有15μL。

[0090] 图39显示相较于基线孔突变28(CsgG-(WT-Y51A/F56Q/R97W/R192D-StrepII)9)的basecall准确度的八个CsgG突变孔的basecall准确度。除由突变28中的R97W和R192D取代造成的准确度改进以外,D195-L199(突变A)、F193-L199(突变B)或V105-I107(突变D)的缺失,或F191T(突变C)的取代使得准确度进一步改进。还在含有额外K94Q突变(突变E)的突变孔中测试了对于缺失V105-I107的basecall准确度的效果,且仍观测到准确度相较于基线突变28的改进。引入R93W突变(突变F)或两个R93Y和R97W突变(突变H)而非R97W突变(基线突变28),增加basecall准确度。除R93W(突变G)以外的缺失D195-L199引起basecall准确度增强。

[0091] 图40A和图40B显示基线突变28CsgG-(WT-Y51A/F56Q/R97W/R192D-StrepII)9和突变D的模板速度分布(A)和模板准确度分布(B),所述突变D包含额外的V105-I107的缺失。如实例中所描述,制备模板DNA,且通过突变孔。如实例中所描述,测定模板速度和准确度。图40A显示,当使用突变D时,速度分布相较于基线突变更为紧密。图40B显示,突变D相较于基线突变具有更紧密的模板分布。

[0092] 图41显示实例“波浪线”,其显示通过基线突变28CsgG-(WT-Y51A/F56Q/R97W/R192D-StrepII)9呈现的“噪声”孔误差模式。图41的上图显示在“良好”和“噪声”孔状态期

间,通过孔的电流流的差异。图41的下图显示从“良好”状态过渡到“噪声”状态的放大图。

[0093] 图42显示平均在至少5个回合内,与基线突变28具有相同序列但当相较于基线突变28时具有额外K94N突变(突变I)或额外K94Q突变(突变J)的突变孔的噪声孔状态的减少,所述基线突变28含有Y51A/F56Q/R97W/R192D突变。

[0094] 图43示出具有接合到其每一端的衔接子的模板链的结构。衔接子具有与其预结合的T4 Dda解螺旋酶。实例中使用的衔接子的各部分的序列显示于SEQ ID NO:52到55中。

[0095] 图44显示突变CsgG纳米孔的凝血酶结合适体(TBA)事件之间的中值时间,所述突变CsgG纳米孔包含以下取代中的一个:Q42K(突变K)、E44N(突变L)、E44Q(突变M)、L90R(突变N)、N91R(突变O)、I95R(突变P)、A99R(突变Q)、E101H(突变R)、E101K(突变S)、E101N(突变T)、E101Q(突变U)、E101T(突变V)和Q114K(突变W)。中值时间相较于基线孔显著降低,所述基线孔包含突变Y51A/F56Q/K94Q/R97W/R192D-del(V105-I107)(基线突变E),其所有也包括于所测试13个突变中的每一个中。图44显示Q42K、E44N、E44Q、L90R、N91R、I95R、A99R、E101H、E101K、E101N、E101Q、E101T和Q114K取代中的每一个增加模板DNA捕获速率。

[0096] 图45显示对应于SEQ ID NO 2、5、6、7、27、28、29、30、32、36、3、35、31、40、33、34、37、39、38、41和4的21种CsgG同源物的序列比对

[0097] 图46显示与图45相同的相对序列比对,其中所预测的 α 螺旋二级结构区另外加阴影。

[0098] 图47显示与图45相同的相对序列比对,其中所预测的 β 折叠二级结构区另外加阴影。

[0099] 图48显示孔AQ和孔97W的原始电子资料的两个实例。

[0100] 序列表的描述

[0101] SEQ ID NO:1显示来自大肠杆菌(*Escherchia coli*)菌株K-12亚株MC4100的编码野生型CsgG单体的密码子优化聚核苷酸序列。这个单体不含信号序列。

[0102] SEQ ID NO:2显示来自大肠杆菌菌株K-12亚株MC4100的野生型CsgG单体的成熟形式的氨基酸序列。这个单体不含信号序列。对于此使用的缩写为CsgG=CsgG-Eco。

[0103] SEQ ID NO:3显示假定蛋白CK0_02032[差异柠檬酸杆菌属(*Citrobacter koseri*) ATCC BAA-895]的YP_001453594.1:1-248的氨基酸序列,其与SEQ ID NO:2 99%一致。

[0104] SEQ ID NO:4显示卷曲生产组合件/转运组分CsgG、部分[肠道沙门氏菌(*Salmonella enterica*)]的WP_001787128.1:16-238的氨基酸序列,其与SEQ ID NO:2 98%。

[0105] SEQ ID NO:5显示卷曲生产组合件/转运蛋白CsgG[无丙二酸柠檬酸杆菌属(*Citrobacter amalonaticus*)]的KEY44978.1:16-277的氨基酸序列,其与SEQ ID NO:2 98%一致。

[0106] SEQ ID NO:6显示卷曲生产组合件/转运组分[鼠柠檬酸杆菌ICC168(*Citrobacter rodentium* ICC168)]的YP_003364699.1:16-277的氨基酸序列,其与SEQ ID NO:2 97%一致。

[0107] SEQ ID NO:7显示卷曲生产组合件/转运组分CsgG[阿氏肠杆菌LF7a(*Enterobacter asburiae* LF7a)]的YP_004828099.1:16-277的氨基酸序列,其与SEQ ID NO:2 94%一致。

- [0108] SEQ ID NO:8显示编码Phi29 DNA聚合酶的聚核苷酸序列。
- [0109] SEQ ID NO:9显示Phi29 DNA聚合酶的氨基酸序列。
- [0110] SEQ ID NO:10显示衍生自大肠杆菌sbcB基因的密码子优化聚核苷酸序列。其编码大肠杆菌核酸外切酶I (EcoExo I)。
- [0111] SEQ ID NO:11显示大肠杆菌核酸外切酶I (EcoExo I)的氨基酸序列。
- [0112] SEQ ID NO:12显示衍生自大肠杆菌xthA基因的密码子优化聚核苷酸序列。其编码大肠杆菌核酸外切酶III。
- [0113] SEQ ID NO:13显示大肠杆菌核酸外切酶III的氨基酸序列。这个酶执行来自双链DNA(dsDNA)的一条链的5'单磷酸酯核苷在3'-5'方向上的分配消化。链上的酶起始需要大致4个核苷酸的5'悬突。
- [0114] SEQ ID NO:14显示衍生自极端嗜热菌(*T.thermophilus*) recJ基因的密码子优化聚核苷酸序列。其编码极端嗜热菌RecJ酶(TthRecJ-cd)。
- [0115] SEQ ID NO:15显示极端嗜热菌RecJ酶(TthRecJ-cd)的氨基酸序列。这个酶执行来自ssDNA的5'单磷酸酯核苷在5'-3'方向上的进行性消化。链上的酶起始需要至少4个核苷酸。
- [0116] SEQ ID NO:16显示衍生自噬菌体λexo(redX)基因的密码子优化聚核苷酸序列。其编码噬菌体λ核酸外切酶。
- [0117] SEQ ID NO:17显示噬菌体λ核酸外切酶的氨基酸序列。所述序列是组装成三聚体的三个相同亚基中的一个。所述酶执行来自dsDNA的一个链的核苷酸在5'-3'方向上的高度进行性消化(<http://www.neb.com/nebecomm/products/productM0262.asp>)。链上的酶起始优选地需要具有5'磷酸的大致4个核苷酸的5'悬突。
- [0118] SEQ ID NO:18显示Hel308 Mbu的氨基酸序列。
- [0119] SEQ ID NO:19显示Hel308 Csy的氨基酸序列。
- [0120] SEQ ID NO:20显示Hel308 Tga的氨基酸序列。
- [0121] SEQ ID NO:21显示Hel308 Mhu的氨基酸序列。
- [0122] SEQ ID NO:22显示TraI Eco的氨基酸序列。
- [0123] SEQ ID NO:23显示XPD Mbu的氨基酸序列。
- [0124] SEQ ID NO:24显示Dda 1993的氨基酸序列。
- [0125] SEQ ID NO:25显示Trwc Cba的氨基酸序列。
- [0126] SEQ ID NO:26显示转运体[雷金斯堡约克氏菌(*Yokenella regensburgei*)]的WP_006819418.1:19-280的氨基酸序列,其与SEQ ID NO:2 91%一致。
- [0127] SEQ ID NO:27显示卷曲生产组合件/转运蛋白CsgG[灰尘粘菌(*Cronobacter pulveris*)]的WP_024556654.1:16-277的氨基酸序列,其与SEQ ID NO:2 89%一致。
- [0128] SEQ ID NO:28显示卷曲生产组合件/转运蛋白CsgG[水生拉恩菌HX2(*Rahnella aquatilis* HX2)]的YP_005400916.1:16-277的氨基酸序列,其与SEQ ID NO:2 84%一致。
- [0129] SEQ ID NO:29显示CsgG家族卷曲生产组合件/转运组分[抗坏血酸吕克沃尔菌(*Kluyvera ascorbata*) ATCC 33433]的KFC99297.1:20-278的氨基酸序列,其与SEQ ID NO:2 82%一致。
- [0130] SEQ ID NO:30显示CsgG家族卷曲生产组合件/转运组分[蜂房哈夫尼菌(*Hafnia*

alvei)ATCC 13337]的KFC86716.1|:16-274的氨基酸序列,其与SEQ ID NO:2 81%一致。

[0131] SEQ ID NO:31显示涉及形成卷曲聚合物[肠杆菌科细菌菌株FGI 57]的未定性蛋白质的YP_007340845.1|:16-270的氨基酸序列,其与SEQ ID NO:2 76%一致。

[0132] SEQ ID NO:32显示卷曲生产组合件/转运蛋白CsgG[类志贺邻单胞菌(*Plesiomonas shigelloides*)]的WP_010861740.1:17-274的氨基酸序列,其与SEQ ID NO:2 70%一致。

[0133] SEQ ID NO:33显示卷曲生产组合件/转运外膜脂蛋白组分CsgG[费希弧菌(*Vibrio fischeri*)ES114]的YP_205788.1:23-270的氨基酸序列,其与SEQ ID NO:2 60%一致。

[0134] SEQ ID NO:34显示卷曲生产组合件蛋白质CsgG[洛伊氏弧菌(*Aliivibrio logei*)]的WP_017023479.1:23-270的氨基酸序列,其与SEQ ID NO:2 59%一致。

[0135] SEQ ID NO:35显示卷曲生产组合件/转运组分CsgG[发光杆菌属(*Photobacterium* sp.)AK15]的WP_007470398.1:22-275的氨基酸序列,其与SEQ ID NO:2 57%一致。

[0136] SEQ ID NO:36显示卷曲生产组合件蛋白质CsgG[维氏气单胞菌(*Aeromonas veronii*)]的WP_021231638.1:17-277的氨基酸序列,其与SEQ ID NO:2 56%一致。

[0137] SEQ ID NO:37显示卷曲生产组合件/转运蛋白CsgG[希瓦氏菌属(*Shewanella* sp.)ECSMB14101]的WP_033538267.1:27-265的氨基酸序列,其与SEQ ID NO:2 56%一致。

[0138] SEQ ID NO:38显示卷曲生产组合件蛋白质CsgG[恶臭假单胞菌(*Pseudomonas putida*)]的WP_003247972.1:30-262的氨基酸序列,其与SEQ ID NO:2 54%一致。

[0139] SEQ ID NO:39显示卷曲生产组合件/转运组分CsgG[堇色希瓦氏菌(*Shewanella violacea*)DSS12]的YP_003557438.1:1-234的氨基酸序列,其与SEQ ID NO:2 53%一致。

[0140] SEQ ID NO:40显示卷曲生产组合件/转运蛋白CsgG[詹氏海杆菌(*Marinobacterium jannaschii*)]的WP_027859066.1:36-280的氨基酸序列,其与SEQ ID NO:2 53%一致。

[0141] SEQ ID NO:41显示卷曲生产组合件/转运组分CsgG[生鲜奶金黄杆菌属(*Chryseobacterium oranimense*)G311]的CEJ70222.1:29-262的氨基酸序列,其与SEQ ID NO:2 50%一致。

[0142] SEQ ID NO:42显示实例2中使用的聚核苷酸序列。

[0143] SEQ ID NO:43显示实例2中使用的聚核苷酸序列。

[0144] SEQ ID NO:44显示实例2中使用的聚核苷酸序列。

[0145] SEQ ID NO:45显示实例2中使用的聚核苷酸序列。

[0146] SEQ ID NO:46显示实例2中使用的聚核苷酸序列。连接于SEQ ID NO:46的3'端的是六个iSp18间隔子,其在两个胸腺嘧啶和3'胆固醇TEG相对端处连接。

[0147] SEQ ID NO:47显示StrepII(C)的聚核苷酸序列。

[0148] SEQ ID NO:48显示Pro的聚核苷酸序列。

[0149] SEQ ID NO:49显示编码野生型MspA单体的密码子优化聚核苷酸序列。这个突变不含信号序列。

[0150] SEQ ID NO:50显示野生型MspA单体的成熟形式的氨基酸序列。这个突变不含信号序列。

[0151] SEQ ID NO:51显示在实例7和11中使用的凝血酶结合适体的聚核苷酸序列。

- [0152] SEQ ID NO:52显示Y衔接子上链的聚核苷酸序列。
- [0153] SEQ ID NO:53显示Y衔接子阻断链的聚核苷酸序列。
- [0154] SEQ ID NO:54显示Y衔接子胆固醇系链链的聚核苷酸序列。
- [0155] SEQ ID NO:55显示Y衔接子下链的聚核苷酸序列。
- [0156] SEQ ID NO:56显示在实例中使用的3.6kb双链DNA靶序列的聚核苷酸序列。

具体实施方式

[0157] 应理解,所公开产物和方法的不同应用可根据所属领域的特定需要来调适。还应理解,本文所用的术语仅出于描述本发明的特定实施例的目的,并且不打算限制性的。

[0158] 另外除非上下文另外明确规定,否则如本说明书和所附权利要求书中所使用,单数形式“一(a/an)”和“所述(the)”包括多个指示物。因此,举例来说,提及“聚核苷酸”包括两个或更多个聚核苷酸;提及“一聚核苷酸结合蛋白”包括两个或更多个此类蛋白质;提及“一解螺旋酶”包括两个或更多个解螺旋酶;提及“一单体”是指两个或更多个单体;提及“一孔”包括两个或更多个孔和其类似物。

[0159] 本文中(不论上文或下文)所列举的所有公开、专利和专利申请均以全文引用的方式并入本文中。

[0160] 突变CsgG单体

[0161] 本发明的一方面提供突变CsgG单体。突变CsgG单体可用于形成本发明的孔。突变CsgG单体是其序列不同于野生型CsgG单体的序列且保持形成孔的能力的单体。用于确定突变单体形成孔的能力的方法在所属领域中是众所周知的,且在下文更详细地论述。

[0162] 由本发明的一些实施例的CsgG单体(包含修饰R97W)构筑的孔显示,当表征(或测序)靶聚核苷酸时,相较于在97处无修饰的其它相同孔,准确度增加。当本发明的CsgG单体包含修饰R93W或修饰R93Y和R97Y而非R97W时,也看到准确度增加。因此,可由一或多个突变CsgG单体来构筑孔,所述突变CsgG单体包含SEQ ID NO:2的R97或R93处的修饰,使得所述修饰增加氨基酸的疏水性。举例来说,此类修饰可包括被含有疏水性侧链的任何氨基酸,包括例如(但不限于)W和Y,取代的氨基酸取代。

[0163] 本发明的一些实施例的包含R192D/Q/F/S/T的CsgG单体比在位置192处不具有取代的单体更容易表达,这可归因于正电荷的减少。因此,位置192可被降低正电荷的氨基酸取代。本发明的包含R192D/Q/F/S/T的单体还可包含额外修饰,所述修饰改进由单体形成的突变孔与分析物(如聚核苷酸)相互作用和表征的能力。

[0164] 当表征(或测序)靶聚核苷酸时,包含本发明的一些实施例的CsgG单体的孔显示准确度增加,所述CsgG单体包含V105、A106和I107的缺失,F193、I194、D195、Y196、Q197、R198和L199的缺失,或D195、Y196、Q197、R198和L199的缺失和/或F191T。在位置105到107处的氨基酸对应于纳米孔帽中的顺式环;且在位置193到199处的氨基酸对应于孔另一端处的反式环。不希望受理论所束缚,认为,顺式环的缺失改进酶与孔的相互作用,且反式环的移除减少孔反侧上的DNA之间的任何不需要的相互作用。

[0165] 包含本发明的一些实施例的CsgG单体的孔,所述CsgG单体包含K94Q或K94N,显示当表征(或测序)靶聚核苷酸时,相较于在94处无突变的相同孔,噪声孔(即引起信号:噪声比增加的那些孔)的数量减少。发现位置94是在孔的孔室内,且发现是一个关于电流信号噪

声的尤其敏感的位置。

[0166] 包含本发明的一些实施例的CsgG单体的孔,所述CsgG单体包含T104K或T104R、N91R、E101K/N/Q/T/H、E44N/Q、Q114K、A99R、I95R、N91R、L90R、E44Q/N和/或Q42K,均表明当用于表征(或测序)靶聚核苷酸时,相较于在这些位置处无取代的相同孔,捕获靶聚核苷酸的能力改进。

[0167] 使用跨膜孔表征(如测序)聚核苷酸可如在国际申请第PCT/GB2012/052343号(公开为W0 2013/041878)中所公开来进行。由于靶聚核苷酸相对于孔移动或移动通过孔,可通常通过测量流过孔的离子电流来从所产生的独特离子电流标志来表征分析物。在任何特定时间处测量的电流电平通常取决于一组k聚合物(例如核苷酸)单元,其中k是正整数,且典型电流标志可表示为指示特定k聚体的一系列电流电平。聚核苷酸相对于(如通过)孔的移动可视为从一个k聚体移动到另一个或从k聚体移动到k聚体。表征聚核苷酸的分析技术可例如涉及使用HMM、神经网络和例如前向-后向算法或维特比算法(Viterbi algorithm)来测定对应于特定序列的一系列测量的可能性。或者,聚核苷酸可通过测定特征向量和将特征向量与另一特征向量进行比较来表征,所述另一特征向量可为已知的,如在国际申请第PCT/GB2013/050381号(公开为W0 2013/121224)中所公开。然而,用于表征聚核苷酸的分析技术不必限于以上实例。

[0168] 当本发明的单体形成跨膜孔且与聚核苷酸结合蛋白一起使用以表征靶聚核苷酸时,一些经修饰位置与聚核苷酸结合蛋白相互作用。举例来说,当单体形成跨膜孔且与聚核苷酸结合蛋白一起使用以表征靶聚核苷酸时,R97W与聚核苷酸结合蛋白相互作用。修饰根据本发明的CsgG单体通常提供靶聚核苷酸相对于(如通过)包含单体的跨膜孔的更一致移动。在靶聚核苷酸相对于(如通过)孔移动时,修饰通常提供从一个k聚体到另一个或从k聚体到k聚体的更一致移动。修饰通常允许靶聚核苷酸相对于(如通过)跨膜孔更平稳地移动。修饰通常提供靶聚核苷酸相对于(如通过)跨膜孔更规则或更不规则的移动。

[0169] 修饰根据本发明的CsgG单体(例如R97W)通常降低与靶聚核苷酸相对于(如通过)包含单体的孔移动相关的正向滑动的量。在实例中使用的包括Dda解螺旋酶的一些解螺旋酶沿聚核苷酸在5'到3'方向上移动。当聚核苷酸的5'端(远离解螺旋酶移动的末端)由孔捕获时,解螺旋酶按照由所施加的电势造成的场的方向起作用,且将交织的聚核苷酸移动到孔中且移动到反式腔室中。正向滑动涉及DNA相对于孔正向移动(即朝向其3'且远离其5'端移动)至少4个连续核苷酸且通常大于10个连续核苷酸。正向滑动可涉及正向移动100个连续核苷酸或更多,且这可在各链中发生大于一次。

[0170] 根据本发明修饰CsgG单体(例如R97W)降低与靶聚核苷酸相对于(如通过)包含单体的跨膜孔移动相关的噪声。在分析信号时靶聚核苷酸在任何方向上的不需要的移动通常造成电流标志或k聚体电平中的噪声。修饰可通过降低靶聚核苷酸中的与一或多个k聚体(如各k聚体)相关的不需要的移动,来降低这种噪声。修饰可降低靶聚核苷酸中的与一或多个k聚体(如各k聚体)的电流电平或标志相关的噪声。

[0171] 对于聚核苷酸采用的酶马达在完全催化周期中具有多个步骤,其中ATP水解以使聚核苷酸正向移动一个碱基(例如结合ATP.Mg,水解以产生ADP.P.Mg,使聚核苷酸正向移动一个碱基,且释放ADP/P/Mg副产物)。各子步骤过程具有通过所述过程的动力学所测定的特有停留时间分布。如果催化周期的这些子步骤的任一个在读取器中移动聚核苷酸的位置

(例如通过相对于酶移动聚核苷酸,或通过改变孔顶部上的酶的位置),接着这可观测为通过孔的电流变化,只要所述变化持续足够长以通过采集电子设备检测到即可。如果子步骤过程未引起构形改变或聚核苷酸中的移动,或发生地太快而难以观测到,那么在理想系统中,对于聚核苷酸正向移动一个整数碱基,完全催化周期将仅引起电流的一个阶跃变化。

[0172] 对于不含有R97W的孔(例如Pro-CP1-Eco-(WT-Y51A/F56Q-StrepII(C))9),我们观测到以大致取决于ATP.Mg浓度的指数停留分布的很长的停留时间水平(其通过模型预测)。对于孔AQ,我们也观测到主要电平之间的短暂的子步骤电流电平,如在图48中标示。因为子步骤电流电平是短暂的,所以其在两个相隔较宽的电流电平之间的空隙中最容易观测到。子步骤电平对应于聚核苷酸的中等大致0.5个碱基移动,且在這些条件下,具有大致3毫秒的ATP.Mg独立的停留时间。

[0173] 含有R97W的孔(例如Pro-CP1-Eco-(WT-Y51A/F56Q/R97W-StrepII(C))9)显示类似很长的具有ATP.Mg依赖性停留时间的主要电平,但在这些条件下或在采集频率下,不显示独特中间子步骤电流电平的迹象(可能解释为,其未发生,发生地太快而难以观测,或子步骤确实发生且对于观测原则上足够缓慢,但实际上其归因于例如酶与孔相互作用的方式而不能观测到)。

[0174] 原始数据迹线(图48)显示,对于孔Pro-CP1-Eco-(WT-Y51A/F56Q/R97W-StrepII(C))9(孔97W)和Pro-CP1-Eco-(WT-Y51A/F56Q-StrepII(C))9(孔AQ),通过纳米孔的酶控制DNA链易位的相对于时间(x轴,秒)的离子电流(y轴,pA)迹线。各电流电平是保持在纳米孔读取器中的序列改变离子流的结果,且当聚核苷酸改变在纳米孔中的位置时,例如当酶使整个链正向移动一个碱基时,观测到电流逐步变化。在此情况下,DNA链部分地含有重复序列(GGTT)_n。通过将Dda酶负载到合成DNA聚核苷酸上和记录原始数据输出的MinION上运行来获得数据(顺式缓冲液:500mM KCl、25mM HEPES、pH8、0.6mM MgCl₂、0.6mM ATP、140mV、37℃、5kHz采集频率)。孔97W仅显示来自聚核苷酸的整数个逐步动作的主要电流电平,在电平之间无显著数据密度。相比而言,孔AQ具有显著的中间子步骤电平,如在图48中由箭头标示。

[0175] 突变单体优选地具有改进的聚核苷酸读取特性,即显示改进的聚核苷酸捕获和核苷酸鉴别。特定来说,由突变单体构筑的孔优选地比野生型更容易地捕获核苷酸和聚核苷酸。另外,由突变单体构筑的孔优选地显示电流范围增加和状态变化降低,所述电流范围增加使得更容易在不同核苷酸之间进行鉴别,所述状态变化降低增加信噪比。

[0176] 另外,优选地减少在聚核苷酸移动通过由突变构筑的孔时引起电流的核苷酸的数量。这使得其更容易鉴定在聚核苷酸移动通过孔时所观测的电流与聚核苷酸序列之间的直接关系。另外,由突变单体构筑的孔可显示增加的输送量,即更可能与分析物(如聚核苷酸)相互作用。这使得使用孔更容易表征分析物。可更容易地将由突变单体构筑的孔插入到膜中。本发明的突变单体包含SEQ ID NO:2中所示序列的变异体。SEQ ID NO:2是来自大肠杆菌菌株K-12亚株MC4100的野生型CsgG单体。SEQ ID NO:2的变异体是具有不同于SEQ ID NO:2的氨基酸序列且保持其形成孔的能力的多肽。可使用本领域中已知的任何方法来分析变异体形成孔的能力。举例来说,可将变异体连同其它适当亚基一起插入到两亲层中,且可测定其寡聚形成孔的能力。用于将亚基插入到膜(如两亲层)中的方法在本领域中已知。举例来说,可将纯化形式的亚基悬浮于含有三嵌段共聚物膜的溶液中,使得其扩散到膜中且

通过结合于膜和组装成功能性状态来插入。

[0177] 在本文的所有论述中,使用标准的针对氨基酸的一个字母代码。这些如下:丙氨酸(A)、精氨酸(R)、天冬酰胺(N)、天冬氨酸(D)、半胱氨酸(C)、谷氨酸(E)、谷氨酰胺(Q)、甘氨酸(G)、组氨酸(H)、异亮氨酸(I)、亮氨酸(L)、赖氨酸(K)、甲硫氨酸(M)、苯丙氨酸(F)、脯氨酸(P)、丝氨酸(S)、苏氨酸(T)、色氨酸(W)、酪氨酸(Y)和缬氨酸(V)。也使用标准的取代记法,即Q42R意指位置42处的Q被R置换。

[0178] 在本发明的突变单体的一个实施例中,SEQ ID NO:2的变异体包含(a)在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变):I41、R93、A98、Q100、G103、T104、A106、I107、N108、L113、S115、T117、Y130、K135、E170、S208、D233、D238和E244;和/或(b)以下中的一或多个:D43S、E44S、F48S/N/Q/Y/W/I/V/H/R/K、Q87N/R/K、N91K/R、K94R/F/Y/W/L/S/N、R97F/Y/W/V/I/K/S/Q/H、E101I/L/A/H、N102K/Q/L/I/V/S/H、R110F/G/N、Q114R/K、R142Q/S、T150Y/A/V/L/S/Q/N、R192D/Q/F/S/T和D248S/N/Q/K/R。变异体可包含(a);(b);或(a)和(b)。

[0179] 在本发明的一些实施例中,SEQ ID NO:2的变异体包含R97W。

[0180] 在本发明的一些实施例中,SEQ ID NO:2的变异体包含R192D/Q/F/S/T,优选地R192D/Q,更优选地R192D。在(a)中,变异体可包含任何数量的位置和位置的组合处的修饰,如1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18或19个位置。在(a)中,变异体优选地包含以下中的一或多个:I41N、R93F/Y/W/L/I/V/N/Q/S、A98K/R、Q100K/R、G103F/W/S/N/K/R、T104R/K、A106R/K、I107R/K/W/F/Y/L/V、N108R/K、L113K/R、S115R/K、T117R/K、Y130W/F/H/Q/N、K135L/V/N/Q/S、E170S/N/Q/K/R、S208V/I/F/W/Y/L/T、D233S/N/Q/K/R、D238S/N/Q/K/R和E244S/N/Q/K/R。

[0181] 在(a)中,变异体优选地包含提供靶聚核苷酸相对于(如通过)包含单体的跨膜孔的更一致移动的一或多个修饰。特定来说,在(a)中,变异体优选地包含在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变):R93、G103和I107。变异体可包含R93、G103、I107、R93和G103、R93和I107、G103和I107,或R93、G103和I107。变异体优选地包含以下中的一或多个:R93F/Y/W/L/I/V/N/Q/S、G103F/W/S/N/K/R和I107R/K/W/F/Y/L/V。这些可以针对位置R93、G103和I107所示的任何组合存在。

[0182] 在(a)中,变异体优选地包含允许由突变单体构筑的孔优选地更容易捕获核苷酸和聚核苷酸的一或多个修饰。特定来说,在(a)中,变异体优选地包含在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变):I41、T104、A106、N108、L113、S115、T117、E170、D233、D238和E244。变异体可包含任何数量的位置和位置的组合处的修饰,如1、2、3、4、5、6、7、8、9、10或11个位置。变异体优选地包含以下中的一或多个:I41N、T104R/K、A106R/K、N108R/K、L113K/R、S115R/K、T117R/K、E170S/N/Q/K/R、D233S/N/Q/K/R、D238S/N/Q/K/R和E244S/N/Q/K/R。另外或替代地,变异体可包含(c)Q42K/R、E44N/Q、L90R/K、N91R/K、I95R/K、A99R/K、E101H/K/N/Q/T和/或Q114K/R。

[0183] 在(a)中,变异体优选地包含提供更一致移动和增加捕获的一或多个修饰。特定来说,在(a)中,变异体优选地包含在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变)(i)A98、(ii)Q100、(iii)G103和(iv)I107。变异体优选地包含以下中的一或多个:(i)A98R/K、(ii)Q100K/R、(iii)G103K/R和(iv)I107R/K。变异体可包含{i};{ii};

{iii}; {iv}; {i,ii}; {i,iii}; {i,iv}; {ii,iii}; {ii,iv}; {iii,iv}; {i,ii,iii}; {i,ii,iv}; {i,iii,iv}; {ii,iii,iv} 或 {i,ii,iii,iv}。

[0184] 特别优选的是,提供增加捕获分析物(如聚核苷酸)的突变单体包括在位置Q42、E44、E44、L90、N91、I95、A99、E101和Q114中的一或多个处的突变,所述突变移除在突变位置处的负电荷和/或增加正电荷。特定来说,以下突变可包括于本发明的突变单体中,以产生具有改进的捕获分析物的能力的CsgG孔,优选地聚核苷酸:Q42K、E44N、E44Q、L90R、N91R、I95R、A99R、E101H、E101K、E101N、E101Q、E101T和Q114K。包含这些突变中的一个与其它有益突变组合的特定突变单体的实例描述于实例11中。

[0185] 在(a)中,变异体优选地包含提供增加的表征准确度的一或多个修饰。特定来说,在(a)中,变异体优选地包含在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变):Y130、K135和S208,如Y130、K135、S208、Y130和K135,Y130和S208,K135和S208,或Y130、K135和S208。变异体优选地包含以下中的一或多个:Y130W/F/H/Q/N、K135L/V/N/Q/S和R142Q/S。这些取代可以任何数量和组合存在,如针对Y130、K135和S208所陈述。

[0186] 在(b)中,变异体可包含任何数量的取代和取代的组合,如1、2、3、4、5、6、7、8、9、10、11或12个取代。在(b)中,变异体优选地包含提供靶聚核苷酸相对于(如通过)包含单体的跨膜孔更一致移动的一或多个修饰。特定来说,在(b)中,变异体优选地包含以下中的一或多个:(i)Q87N/R/K、(ii)K94R/F/Y/W/L/S/N、(iii)R97F/Y/W/V/I/K/S/Q/H、(iv)N102K/Q/L/I/V/S/H和(v)R110F/G/N。更优选地,变异体包含K94D或K94Q和/或R97W或R97Y。变异体可包含:{i}; {ii}; {iii}; {iv}; {v}; {i,ii}; {i,iii}; {i,iv}; {i,v}; {ii,iii}; {ii,iv}; {ii,v}; {iii,iv}; {iii,v}; {iv,v}; {i,ii,iii}; {i,ii,iv}; {i,ii,v}; {i,iii,iv}; {i,iii,v}; {i,iv,v}; {ii,iii,iv}; {ii,iii,v}; {ii,iv,v}; {iii,iv,v}; {i,ii,iii,iv}; {i,ii,iii,v}; {i,ii,iv,v}; {i,iii,iv,v}; {ii,iii,iv,v} 或 {i,ii,iii,iv,v}。经修饰以提供靶聚核苷酸相对于(如通过)包含单体的跨膜孔更一致移动的其它优选变异体包括(vi)R93W和R93Y。优选变异体可包含R93W和R97W、R93Y和R97W、R93W和R97W,或更优选地R93Y和R97Y。变异体可包含:{vi}; {i,vi}; {ii,vi}; {iii,vi}; {iv,vi}; {v,vi}; {i,ii,vi}; {i,iii,vi}; {i,iv,vi}; {i,v,vi}; {ii,iii,vi}; {ii,iv,vi}; {ii,v,vi}; {iii,iv,vi}; {iii,v,vi}; {iv,v,vi}; {i,ii,iii,vi}; {i,ii,iv,vi}; {i,ii,v,vi}, {i,iii,iv,vi}; {i,iii,v,vi}; {i,iv,v,vi}; {ii,iii,iv,vi}; {ii,iii,v,vi}; {ii,iv,v,vi}; {iii,iv,v,vi}; {i,ii,iii,iv,vi}; {i,ii,iii,v,vi}; {i,ii,iv,v,vi}; {i,iii,iv,v,vi}; {ii,iii,iv,v,vi} 或 {i,ii,iii,iv,v,vi}。

[0187] 在(b)中,变异体优选地包含允许由突变单体构筑的孔优选地更容易捕获核苷酸和聚核苷酸的一或多个修饰。特定来说,在(b)中,变异体优选地包含以下中的一或多个:(i)D43S、(ii)E44S、(iii)N91K/R、(iv)Q114R/K和(v)D248S/N/Q/K/R。变异体可包含:{i}; {ii}; {iii}; {iv}; {v}; {i,ii}; {i,iii}; {i,iv}; {i,v}; {ii,iii}; {ii,iv}; {ii,v}; {iii,iv}; {iii,v}; {iv,v}; {i,ii,iii}; {i,ii,iv}; {i,ii,v}; {i,iii,iv}; {i,iii,v}; {i,iv,v}; {ii,iii,iv}; {ii,iii,v}; {ii,iv,v}; {iii,iv,v}; {i,ii,iii,iv}; {i,ii,iii,v}; {i,ii,iv,v}; {i,iii,iv,v}; {ii,iii,iv,v} 或 {i,ii,iii,iv,v}。

[0188] 在(b)中,变异体优选地包含提供更一致移动和增加捕获的一或多个修饰。特定来说,在(b)中,变异体优选地包含以下中的一或多个:Q87R/K、E101I/L/A/H和N102K,如Q87R/

K;E101I/L/A/H;N102K;Q87R/K和E101I/L/A/H;Q87R/K和N102K;E101I/L/A/H和N102K;或Q87R/K、E101I/L/A/H和N102K。

[0189] 在(b)中,变异体优选地包含提供增加的表征准确度的一或多个修饰。特定来说,在(a)中,变异体优选地包含F48S/N/Q/Y/W/I/V。

[0190] 在(b)中,变异体优选地包含提供表征准确度增加和捕获增加的一或多个修饰。特定来说,在(a)中,变异体优选地包含F48H/R/K。

[0191] 变异体可包含(a)和(b)两个中的提供更一致移动的修饰。变异体可包含(a)和(b)两个中的提供捕获增加的修饰。

[0192] 本发明提供SEQ ID NO:2的变异体,其使用包含变异体的孔提供增加输送量的用于表征分析物(如聚核苷酸)的分析。此类变异体可包含K94处,优选地K94Q或K94N处,更优选地K94Q处的突变。包含K94Q或K94N突变与其它有益突变组合的特定突变单体的实例描述于实例10和11中。

[0193] 本发明提供SEQ ID NO:2的变异体,其使用包含变异体的孔在用于表征分析物(如聚核苷酸)的分析提供增加的表征准确度。此类变异体包括包含以下的变异体:F191处,优选地F191T处的突变;V105-I107的缺失;F193-L199或D195-L199的缺失;和/或R93和/或R97处,优选地R93Y、R97Y处,,或更优选地R97W、R93W或R97Y和R97Y两个处的突变。包含这些突变中的一或多个与其它有益突变组合的特定突变单体的实例描述于实例9中。

[0194] 在本发明的突变单体的另一实施例中,SEQ ID NO:2的变异体包含:(A)一或多个位置R192、F193、I194、D195、Y196、Q197、R198、L199、L200和E201的缺失;和/或(B)以下中的一或多个的缺失:V139/G140/D149/T150/V186/Q187/V204/G205(在本文中称为条带1)、G137/G138/Q151/Y152/Y184/E185/Y206/T207(在本文中称为条带2)和A141/R142/G147/A148/A188/G189/G202/E203(在本文中称为条带3)。

[0195] 在(A)中,变异体可包含任何数量的位置和位置的组合的缺失,如1、2、3、4、5、6、7、8、9或10个位置。在(A)中,变异体优选地包含以下的缺失:

[0196] -D195、Y196、Q197、R198和L199;

[0197] -R192、F193、I194、D195、Y196、Q197、R198、L199和L200;

[0198] -Q197、R198、L199和L200;

[0199] -I194、D195、Y196、Q197、R198和L199;

[0200] -D195、Y196、Q197、R198、L199和L200;

[0201] -Y196、Q197、R198、L199、L200和E201;

[0202] -Q197、R198、L199、L200和E201;

[0203] -Q197、R198、L199;或

[0204] -F193、I194、D195、Y196、Q197、R198和L199。

[0205] 更优选地,变异体包含以下的缺失:D195、Y196、Q197、R198和L199或F193、I194、D195、Y196、Q197、R198和L199。在(B)中,可缺失任何数量的条带和条带1到3的组合,如条带1、条带2、条带3条带1和2、条带1和3、条带2和3,或条带1、2和3。

[0206] 变异体可包含根据(A)、(B),或(A)和(B)的缺失。

[0207] 包含根据以上(A)和/或(B)的一或多个位置的缺失的变异体可进一步包含上文和下文所论述的修饰或取代中的任一个。如果在出现于SEQ ID NO:2中的缺失位置之后的一

或多个位置处进行修饰或取代,那么修饰或取代的一或多个位置的编号必须进行相应的调整。举例来说,如果缺失L199,那么E244变为E243。类似地,如果缺失条带1,那么R192变为R186。

[0208] 在本发明的突变单体的另一实施例中,SEQ ID NO:2的变异体包含(C)一或多个位置V105、A106和I107的缺失。除根据(A)和/或(B)的缺失以外,可进行根据(C)的缺失。

[0209] 上文所描述的缺失通常降低与靶聚核苷酸相对于(如通过)包含单体的跨膜孔移动相关的噪声。因此,可更准确地表征靶聚核苷酸。

[0210] 在于特定位置处的不同氨基酸通过/符号分隔的以上段落中,/符号意指“或”。举例来说,Q87R/K意指Q87R或Q87K。

[0211] 本发明提供SEQ ID NO:2的变异体,其提供增加的分析物(如聚核苷酸)的捕获。此类变异体可包含:T104处,优选地T104R或T104K处的突变;N91处,优选地N91R处的突变;E101处,优选地E101K/N/Q/T/H处的突变;位置E44处,优选地E44N或E44Q处的突变;和/或位置Q42处,优选地Q42K处的突变。

[0212] 在SEQ ID NO:2中不同位置处的突变可以任何可能的方法来进行组合。特定来说,本发明的单体可包含:改进准确度的一或多个突变、降低噪声的一或多个突变、和/或增强捕获分析物的一或多个突变。

[0213] 在本发明的突变单体中,SEQ ID NO:2的变异体优选地包含以下中的一或多个:
(i) 在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变):N40、D43、E44、S54、S57、Q62、R97、E101、E124、E131、R142、T150和R192,如在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变)N40、D43、E44、S54、S57、Q62、E101、E131和T150,或N40、D43、E44、E101和E131;(ii) 在Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处的突变;(iii) Q42R或Q42K;(iv) K49R;(v) N102R、N102F、N102Y或N102W;(vi) D149N、D149Q或D149R;(vii) E185N、E185Q或E185R;(viii) D195N、D195Q或D195R;(ix) E201N、E201Q或E201R;(x) E203N、E203Q或E203R;和(xi) 以下位置中的一或多个的缺失:F48、K49、P50、Y51、P52、A53、S54、N55、F56和S57。变异体可包含(i)到(xi)的任何组合。特定来说,变异体可包含:(其中由空格分隔的圆括号{}中的各变异体表示来自变异体列表的任选变异体,即{i}或{ii}或{iii}或{iv}或{v}等){i}{ii}{iii}{iv}{v}{vi}{vii}{viii}{ix}{x}{xi}{i,ii}{i,iii}{i,iv}{i,v}{i,vi}{i,vii}{i,viii}{i,ix}{i,x}{i,xi}{ii,iii}{ii,iv}{ii,v}{ii,vi}{ii,vii}{ii,viii}{ii,ix}{ii,x}{ii,xi}{iii,iv}{iii,v}{iii,vi}{iii,vii}{iii,viii}{iii,ix}{iii,x}{iii,xi}{iv,v}{iv,vi}{iv,vii}{iv,viii}{iv,ix}{iv,x}{iv,xi}{v,vi}{v,vii}{v,viii}{v,ix}{v,x}{v,xi}{vi,vii}{vi,viii}{vi,ix}{vi,x}{vi,xi}{vii,viii}{vii,ix}{vii,x}{viii,ix}{viii,x}{viii,xi}{ix,x}{ix,xi}{x,xi}{i,ii,iii}{i,ii,iv}{i,ii,v}{i,ii,vi}{i,ii,vii}{i,ii,viii}{i,ii,ix}{i,ii,x}{i,ii,xi}{i,iii,iv}{i,iii,v}{i,iii,vi}{i,iii,vii}{i,iii,viii}{i,iii,ix}{i,iii,x}{i,iii,xi}{i,iv,v}{i,iv,vi}{i,iv,vii}{i,iv,viii}{i,iv,ix}{i,iv,x}{i,iv,xi}{i,v,vi}{i,v,vii}{i,v,viii}{i,v,ix}{i,v,x}{i,v,xi}{i,vi,vii}{i,vi,viii}{i,vi,ix}{i,vi,x}{i,vi,xi}{i,vii,viii}{i,vii,ix}{i,vii,x}{i,vii,xi}{i,viii,ix}{i,viii,x}{i,viii,xi}{i,ix,x}{i,ix,xi}{i,x,xi}{ii,iii,iv}{ii,iii,v}{ii,iii,vi}{ii,iii,vii}{ii,iii,viii}{ii,iii,ix}{ii,iii,x}{ii,iii,xi}{ii,iv,v}{ii,

iv,vi} {ii,iv,vii} {ii,iv,viii} {ii,iv,ix} {ii,iv,x} {ii,iv,xi} {ii,v,vi} {ii,v,vii}
 {ii,v,viii} {ii,v,ix} {ii,v,x} {ii,v,xi} {ii,vi,vii} {ii,vi,viii} {ii,vi,ix} {ii,vi,
 x} {ii,vi,xi} {ii,vii,viii} {ii,vii,ix} {ii,vii,x} {ii,vii,xi} {ii,viii,ix} {ii,
 viii,x} {ii,viii,xi} {ii,ix,x} {ii,ix,xi} {ii,x,xi} {iii,iv,v} {iii,iv,vi} {iii,iv,
 vii} {iii,iv,viii} {iii,iv,ix} {iii,iv,x} {iii,iv,xi} {iii,v,vi} {iii,v,vii} {iii,v,
 viii} {iii,v,ix} {iii,v,x} {iii,v,xi} {iii,vi,vii} {iii,vi,viii} {iii,vi,ix} {iii,
 vi,x} {iii,vi,xi} {iii,vii,viii} {iii,vii,ix} {iii,vii,x} {iii,vii,xi} {iii,viii,
 ix} {iii,viii,x} {i,ii,iii,iv} {i,ii,iii,v} {i,ii,iii,vi} {i,ii,iii,vii} {i,ii,iii,
 viii} {i,ii,iii,ix} {i,ii,iii,x} {i,ii,iii,xi} {i,ii,iv,v} {i,ii,iv,vi} {i,ii,iv,
 vii} {i,ii,iv,viii} {i,ii,iv,ix} {i,ii,iv,x} {i,ii,iv,xi} {i,ii,v,vi} {i,ii,v,vii}
 {i,ii,v,viii} {i,ii,v,ix} {i,ii,v,x} {i,ii,v,xi} {i,ii,vi,vii} {i,ii,vi,viii} {i,
 ii,vi,ix} {i,ii,vi,x} {i,ii,vi,xi} {i,ii,vii,viii} {i,ii,vii,ix} {i,ii,vii,x} {i,
 ii,vii,xi} {i,ii,viii,ix} {i,ii,viii,x} {i,ii,viii,xi} {i,ii,ix,x} {i,ii,ix,xi} {i,
 ii,x,xi} {i,iii,iv,v} {i,iii,iv,vi} {i,iii,iv,vii} {i,iii,iv,viii} {i,iii,iv,ix}
 {i,iii,iv,x} {i,iii,iv,xi} {i,iii,v,vi} {i,iii,v,vii} {i,iii,v,viii} {i,iii,v,ix}
 {i,iii,v,x} {i,iii,v,xi} {i,iii,vi,vii} {i,iii,vi,viii} {i,iii,vi,ix} {i,iii,vi,x}
 {i,iii,vi,xi} {i,iii,vii,viii} {i,iii,vii,ix} {i,iii,vii,x} {i,iii,vii,xi} {i,iii,
 viii,ix} {i,iii,viii,x} {i,iii,viii,xi} {i,iii,ix,x} {i,iii,ix,xi} {i,iii,x,xi} {i,
 iv,v,vi} {i,iv,v,vii} {i,iv,v,viii} {i,iv,v,ix} {i,iv,v,x} {i,iv,v,xi} {i,iv,vi,
 vii} {i,iv,vi,viii} {i,iv,vi,ix} {i,iv,vi,x} {i,iv,vi,xi} {i,iv,vii,viii} {i,iv,
 vii,ix} {i,iv,vii,x} {i,iv,vii,xi} {i,iv,viii,ix} {i,iv,viii,x} {i,iv,viii,xi} {i,
 iv,ix,x} {i,iv,ix,xi} {i,iv,x,xi} {i,v,vi,vii} {i,v,vi,viii} {i,v,vi,ix} {i,v,vi,x}
 {i,v,vi,xi} {i,v,vii,viii} {i,v,vii,ix} {i,v,vii,x} {i,v,vii,xi} {i,v,viii,ix} {i,
 v,viii,x} {i,v,viii,xi} {i,v,ix,x} {i,v,ix,xi} {i,v,x,xi} {i,vi,vii,viii} {i,vi,
 vii,ix} {i,vi,vii,x} {i,vi,vii,xi} {i,vi,viii,ix} {i,vi,viii,x} {i,vi,viii,xi} {i,
 vi,ix,x} {i,vi,ix,xi} {i,vi,x,xi} {i,vii,viii,ix} {i,vii,viii,x} {i,vii,viii,xi}
 {i,vii,ix,x} {i,vii,ix,xi} {i,vii,x,xi} {i,viii,ix,x} {i,viii,ix,xi} {i,viii,x,xi}
 {i,ix,x,xi} {ii,iii,iv,v} {ii,iii,iv,vi} {ii,iii,iv,vii} {ii,iii,iv,viii} {ii,iii,
 iv,ix} {ii,iii,iv,x} {ii,iii,iv,xi} {ii,iii,v,vi} {ii,iii,v,vii} {ii,iii,v,viii}
 {ii,iii,v,ix} {ii,iii,v,x} {ii,iii,v,xi} {ii,iii,vi,vii} {ii,iii,vi,viii} {ii,iii,
 vi,ix} {ii,iii,vi,x} {ii,iii,vi,xi} {ii,iii,vii,viii} {ii,iii,vii,ix} {ii,iii,vii,
 x} {ii,iii,vii,xi} {ii,iii,viii,ix} {ii,iii,viii,x} {ii,iii,viii,xi} {ii,iii,ix,x}
 {ii,iii,ix,xi} {ii,iii,x,xi} {ii,iv,v,vi} {ii,iv,v,vii} {ii,iv,v,viii} {ii,iv,v,
 ix} {ii,iv,v,x} {ii,iv,v,xi} {ii,iv,vi,vii} {ii,iv,vi,viii} {ii,iv,vi,ix} {ii,iv,
 vi,x} {ii,iv,vi,xi} {ii,iv,vii,viii} {ii,iv,vii,ix} {ii,iv,vii,x} {ii,iv,vii,xi}
 {ii,iv,viii,ix} {ii,iv,viii,x} {ii,iv,viii,xi} {ii,iv,ix,x} {ii,iv,ix,xi} {ii,iv,
 x,xi} {ii,v,vi,vii} {ii,v,vi,viii} {ii,v,vi,ix} {ii,v,vi,x} {ii,v,vi,xi} {ii,v,vii,
 viii} {ii,v,vii,ix} {ii,v,vii,x} {ii,v,vii,xi} {ii,v,viii,ix} {ii,v,viii,x} {ii,v,
 viii,xi} {ii,v,ix,x} {ii,v,ix,xi} {ii,v,x,xi} {ii,vi,vii,viii} {ii,vi,vii,ix} {ii,

vi,vii,x} {ii,vi,vii,xi} {ii,vi,viii,ix} {ii,vi,viii,x} {ii,vi,viii,xi} {ii,vi,ix,x} {ii,vi,ix,xi} {ii,vi,x,xi} {ii,vii,viii,ix} {ii,vii,viii,x} {ii,vii,viii,xi} {ii,vii,ix,x} {ii,vii,ix,xi} {ii,vii,x,xi} {ii,viii,ix,x} {ii,viii,ix,xi} {ii,viii,x,xi} {ii,ix,x,xi} {iii,iv,v,vi} {iii,iv,v,vii} {iii,iv,v,viii} {iii,iv,v,ix} {iii,iv,v,x} {iii,iv,v,xi} {iii,iv,vi,vii} {iii,iv,vi,viii} {iii,iv,vi,ix} {iii,iv,vi,x} {iii,iv,vi,xi} {iii,iv,vii,viii} {iii,iv,vii,ix} {iii,iv,vii,x} {iii,iv,vii,xi} {iii,iv,viii,ix} {iii,iv,viii,x} {iii,iv,viii,xi} {iii,iv,ix,x} {iii,iv,ix,xi} {iii,iv,x,xi} {iii,v,vi,vii} {iii,v,vi,viii} {iii,v,vi,ix} {iii,v,vi,x} {iii,v,vi,xi} {iii,v,vii,viii} {iii,v,vii,ix} {iii,v,vii,x} {iii,v,vii,xi} {iii,v,viii,ix} {iii,v,viii,x} {iii,v,viii,xi} {iii,v,ix,x} {iii,v,ix,xi} {iii,v,x,xi} {iii,vi,vii,viii} {iii,vi,vii,ix} {iii,vi,vii,x} {iii,vi,vii,xi} {iii,vi,viii,ix} {iii,vi,viii,x} {iii,vi,viii,xi} {iii,vi,ix,x} {iii,vi,ix,xi} {iii,vi,x,xi} {iii,vii,viii,ix} {iii,vii,viii,x} {iii,vii,viii,xi} {iii,vii,ix,x} {iii,vii,ix,xi} {iii,vii,x,xi} {iii,viii,ix,x} {iii,viii,ix,xi} {iii,viii,x,xi} {iii,ix,x,xi} {iv,v,vi,vii} {iv,v,vi,viii} {iv,v,vi,ix} {iv,v,vi,x} {iv,v,vi,xi} {iv,v,vii,viii} {iv,v,vii,ix} {iv,v,vii,x} {iv,v,vii,xi} {iv,v,viii,ix} {iv,v,viii,x} {iv,v,viii,xi} {iv,v,ix,x} {iv,v,ix,xi} {iv,v,x,xi} {iv,vi,vii,viii} {iv,vi,vii,ix} {iv,vi,vii,x} {iv,vi,vii,xi} {iv,vi,viii,ix} {iv,vi,viii,x} {iv,vi,viii,xi} {iv,vi,ix,x} {iv,vi,ix,xi} {iv,vi,x,xi} {iv,vii,viii,ix} {iv,vii,viii,x} {iv,vii,viii,xi} {iv,vii,ix,x} {iv,vii,ix,xi} {iv,vii,x,xi} {iv,viii,ix,x} {iv,viii,ix,xi} {iv,viii,x,xi} {iv,ix,x,xi} {v,vi,vii,viii} {v,vi,vii,ix} {v,vi,vii,x} {v,vi,vii,xi} {v,vi,viii,ix} {v,vi,viii,x} {v,vi,viii,xi} {v,vi,ix,x} {v,vi,ix,xi} {v,vi,x,xi} {v,vii,viii,ix} {v,vii,viii,x} {v,vii,viii,xi} {v,vii,ix,x} {v,vii,ix,xi} {v,vii,x,xi} {v,viii,ix,x} {v,viii,ix,xi} {v,viii,x,xi} {v,ix,x,xi} {vi,vii,viii,ix} {vi,vii,viii,x} {vi,vii,viii,xi} {vi,vii,ix,x} {vi,vii,ix,xi} {vi,vii,x,xi} {vi,viii,ix,x} {vi,viii,ix,xi} {vi,viii,x,xi} {vi,ix,x,xi} {vii,viii,ix,x} {vii,viii,ix,xi} {vii,viii,x,xi} {vii,ix,x,xi} {viii,ix,x,xi} {i,ii,iii,iv,v} {i,ii,iii,iv,vi} {i,ii,iii,iv,vii} {i,ii,iii,iv,viii} {i,ii,iii,iv,ix} {i,ii,iii,iv,x} {i,ii,iii,iv,xi} {i,ii,iii,v,vi} {i,ii,iii,v,vii} {i,ii,iii,v,viii} {i,ii,iii,v,ix} {i,ii,iii,v,x} {i,ii,iii,v,xi} {i,ii,iii,vi,vii} {i,ii,iii,vi,viii} {i,ii,iii,vi,ix} {i,ii,iii,vi,x} {i,ii,iii,vi,xi} {i,ii,iii,vii,viii} {i,ii,iii,vii,ix} {i,ii,iii,vii,x} {i,ii,iii,vii,xi} {i,ii,iii,viii,ix} {i,ii,iii,viii,x} {i,ii,iii,viii,xi} {i,ii,iii,ix,x} {i,ii,iii,ix,xi} {i,ii,iii,x,xi} {i,ii,iv,v,vi} {i,ii,iv,v,vii} {i,ii,iv,v,viii} {i,ii,iv,v,ix} {i,ii,iv,v,x} {i,ii,iv,v,xi} {i,ii,iv,vi,vii} {i,ii,iv,vi,viii} {i,ii,iv,vi,ix} {i,ii,iv,vi,x} {i,ii,iv,vi,xi} {i,ii,iv,vii,viii} {i,ii,iv,vii,ix} {i,ii,iv,vii,x} {i,ii,iv,vii,xi} {i,ii,iv,viii,ix} {i,ii,iv,viii,x} {i,ii,iv,viii,xi} {i,ii,iv,ix,x} {i,ii,iv,ix,xi} {i,ii,iv,x,xi} {i,ii,v,vi,vii} {i,ii,v,vi,viii} {i,ii,v,vi,ix} {i,ii,v,vi,x} {i,ii,v,vi,xi} {i,ii,v,vii,viii} {i,ii,v,vii,ix} {i,ii,v,vii,x} {i,ii,v,vii,xi} {i,ii,v,viii,ix} {i,ii,v,viii,x} {i,ii,v,viii,xi}

viii,xi} {i,ii,v,ix,x} {i,ii,v,ix,xi} {i,ii,v,x,xi} {i,ii,vi,vii,viii} {i,ii,vi,vii,ix} {i,ii,vi,vii,x} {i,ii,vi,vii,xi} {i,ii,vi,viii,ix} {i,ii,vi,viii,x} {i,ii,vi,viii,xi} {i,ii,vi,ix,x} {i,ii,vi,ix,xi} {i,ii,vi,x,xi} {i,ii,vii,viii,ix} {i,ii,vii,viii,x} {i,ii,vii,viii,xi} {i,ii,vii,ix,x} {i,ii,vii,ix,xi} {i,ii,vii,x,xi} {i,ii,viii,ix,x} {i,ii,viii,ix,xi} {i,ii,viii,x,xi} {i,ii,ix,x,xi} {i,iii,iv,v,vi} {i,iii,iv,v,vii} {i,iii,iv,v,viii} {i,iii,iv,v,ix} {i,iii,iv,v,x} {i,iii,iv,v,xi} {i,iii,iv,vi,vii} {i,iii,iv,vi,viii} {i,iii,iv,vi,ix} {i,iii,iv,vi,x} {i,iii,iv,vi,xi} {i,iii,iv,vii,viii} {i,iii,iv,vii,ix} {i,iii,iv,vii,x} {i,iii,iv,vii,xi} {i,iii,iv,viii,ix} {i,iii,iv,viii,x} {i,iii,iv,viii,xi} {i,iii,iv,ix,x} {i,iii,iv,ix,xi} {i,iii,iv,x,xi} {i,iii,v,vi,vii} {i,iii,v,vi,viii} {i,iii,v,vi,ix} {i,iii,v,vi,x} {i,iii,v,vi,xi} {i,iii,v,vii,viii} {i,iii,v,vii,ix} {i,iii,v,vii,x} {i,iii,v,vii,xi} {i,iii,v,viii,ix} {i,iii,v,viii,x} {i,iii,v,viii,xi} {i,iii,v,ix,x} {i,iii,v,ix,xi} {i,iii,v,x,xi} {i,iii,vi,vii,viii} {i,iii,vi,vii,ix} {i,iii,vi,vii,x} {i,iii,vi,vii,xi} {i,iii,vi,viii,ix} {i,iii,vi,viii,x} {i,iii,vi,viii,xi} {i,iii,vi,ix,x} {i,iii,vi,ix,xi} {i,iii,vi,x,xi} {i,iii,vii,viii,ix} {i,iii,vii,viii,x} {i,iii,vii,viii,xi} {i,iii,vii,ix,x} {i,iii,vii,ix,xi} {i,iii,vii,x,xi} {i,iii,viii,ix,x} {i,iii,viii,ix,xi} {i,iii,viii,x,xi} {i,iii,ix,x,xi} {i,iv,v,vi,vii} {i,iv,v,vi,viii} {i,iv,v,vi,ix} {i,iv,v,vi,x} {i,iv,v,vi,xi} {i,iv,v,vii,viii} {i,iv,v,vii,ix} {i,iv,v,vii,x} {i,iv,v,vii,xi} {i,iv,v,viii,ix} {i,iv,v,viii,x} {i,iv,v,viii,xi} {i,iv,v,ix,x} {i,iv,v,ix,xi} {i,iv,v,x,xi} {i,iv,vi,vii,viii} {i,iv,vi,vii,ix} {i,iv,vi,vii,x} {i,iv,vi,vii,xi} {i,iv,vi,viii,ix} {i,iv,vi,viii,x} {i,iv,vi,viii,xi} {i,iv,vi,ix,x} {i,iv,vi,ix,xi} {i,iv,vi,x,xi} {i,iv,vii,viii,ix} {i,iv,vii,viii,x} {i,iv,vii,viii,xi} {i,iv,vii,ix,x} {i,iv,vii,ix,xi} {i,iv,vii,x,xi} {i,iv,viii,ix,x} {i,iv,viii,ix,xi} {i,iv,viii,x,xi} {i,iv,ix,x,xi} {i,v,vi,vii,viii} {i,v,vi,vii,ix} {i,v,vi,vii,x} {i,v,vi,vii,xi} {i,v,vi,viii,ix} {i,v,vi,viii,x} {i,v,vi,viii,xi} {i,v,vi,ix,x} {i,v,vi,ix,xi} {i,v,vi,x,xi} {i,v,vii,viii,ix} {i,v,vii,viii,x} {i,v,vii,viii,xi} {i,v,vii,ix,x} {i,v,vii,ix,xi} {i,v,vii,x,xi} {i,v,viii,ix,x} {i,v,viii,ix,xi} {i,v,viii,x,xi} {i,v,ix,x,xi} {i,vi,vii,viii,ix} {i,vi,vii,viii,x} {i,vi,vii,viii,xi} {i,vi,vii,ix,x} {i,vi,vii,ix,xi} {i,vi,vii,x,xi} {i,vi,viii,ix,x} {i,vi,viii,ix,xi} {i,vi,viii,x,xi} {i,vi,ix,x,xi} {i,vii,viii,ix,x} {i,vii,viii,ix,xi} {i,vii,viii,x,xi} {i,vii,ix,x,xi} {i,viii,ix,x,xi} {ii,iii,iv,v,vi} {ii,iii,iv,v,vii} {ii,iii,iv,v,viii} {ii,iii,iv,v,ix} {ii,iii,iv,v,x} {ii,iii,iv,v,xi} {ii,iii,iv,vi,vii} {ii,iii,iv,vi,viii} {ii,iii,iv,vi,ix} {ii,iii,iv,vi,x} {ii,iii,iv,vi,xi} {ii,iii,iv,vii,viii} {ii,iii,iv,vii,ix} {ii,iii,iv,vii,x} {ii,iii,iv,vii,xi} {ii,iii,iv,viii,ix} {ii,iii,iv,viii,x} {ii,iii,iv,viii,xi} {ii,iii,iv,ix,x} {ii,iii,iv,ix,xi} {ii,iii,iv,x,xi} {ii,iii,v,vi,vii} {ii,iii,v,vi,viii} {ii,iii,v,vi,ix} {ii,iii,v,vi,x} {ii,iii,v,vi,xi} {ii,iii,v,vii,viii} {ii,iii,v,vii,ix} {ii,iii,v,vii,x} {ii,iii,v,vii,xi} {ii,iii,v,viii,ix} {ii,iii,v,viii,x} {ii,iii,v,viii,xi} {ii,iii,v,ix,x} {ii,iii,v,ix,xi}

{ii,iii,v,ix,xi}{ii,iii,v,x,xi}{ii,iii,vi,vii,viii}{ii,iii,vi,vii,ix}{ii,iii,vi,vii,x}{ii,iii,vi,vii,xi}{ii,iii,vi,viii,ix}{ii,iii,vi,viii,x}{ii,iii,vi,viii,xi}{ii,iii,vi,ix,x}{ii,iii,vi,ix,xi}{ii,iii,vi,x,xi}{ii,iii,vii,viii,ix}{ii,iii,vii,viii,x}{ii,iii,vii,viii,xi}{ii,iii,vii,ix,x}{ii,iii,vii,ix,xi}{ii,iii,vii,x,xi}{ii,iii,viii,ix,x}{ii,iii,viii,ix,xi}{ii,iii,viii,x,xi}{ii,iii,ix,x,xi}{ii,iv,v,vi,vii}{ii,iv,v,vi,viii}{ii,iv,v,vi,ix}{ii,iv,v,vi,x}{ii,iv,v,vi,xi}{ii,iv,v,vii,viii}{ii,iv,v,vii,ix}{ii,iv,v,vii,x}{ii,iv,v,vii,xi}{ii,iv,v,viii,ix}{ii,iv,v,viii,x}{ii,iv,v,viii,xi}{ii,iv,v,ix,x}{ii,iv,v,ix,xi}{ii,iv,v,x,xi}{ii,iv,vi,vii,viii}{ii,iv,vi,vii,ix}{ii,iv,vi,vii,x}{ii,iv,vi,vii,xi}{ii,iv,vi,viii,ix}{ii,iv,vi,viii,xi}{ii,iv,vi,ix,x}{ii,iv,vi,ix,xi}{ii,iv,vi,x,xi}{ii,iv,vii,viii,ix}{ii,iv,vii,viii,xi}{ii,iv,vii,ix,x}{ii,iv,vii,ix,xi}{ii,iv,vii,x,xi}{ii,iv,viii,ix,x}{ii,iv,viii,ix,xi}{ii,iv,ix,x,xi}{ii,v,vi,vii,viii}{ii,v,vi,vii,ix}{ii,v,vi,vii,x}{ii,v,vi,vii,xi}{ii,v,vi,viii,ix}{ii,v,vi,viii,x}{ii,v,vi,viii,xi}{ii,v,vi,ix,x}{ii,v,vi,ix,xi}{ii,v,vi,x,xi}{ii,v,vii,viii,ix}{ii,v,vii,viii,xi}{ii,v,vii,ix,x}{ii,v,vii,ix,xi}{ii,v,vii,x,xi}{ii,v,viii,ix,x}{ii,v,viii,ix,xi}{ii,v,viii,x,xi}{ii,v,ix,x,xi}{ii,vi,vii,viii,ix}{ii,vi,vii,viii,x}{ii,vi,vii,viii,xi}{ii,vi,vii,ix,x}{ii,vi,vii,ix,xi}{ii,vi,vii,x,xi}{ii,vi,viii,ix,x}{ii,vi,viii,ix,xi}{ii,vi,viii,ix,xi}{ii,vi,viii,x,xi}{ii,vi,ix,x,xi}{ii,vii,viii,ix,x}{ii,vii,viii,ix,xi}{ii,vii,viii,ix,xi}{ii,vii,viii,x,xi}{ii,vii,ix,x,xi}{ii,viii,ix,x,xi}{iii,iv,v,vi,vii}{iii,iv,v,vi,viii}{iii,iv,v,vi,ix}{iii,iv,v,vi,x}{iii,iv,v,vi,xi}{iii,iv,v,vii,viii}{iii,iv,v,vii,ix}{iii,iv,v,vii,x}{iii,iv,v,vii,xi}{iii,iv,v,viii,ix}{iii,iv,v,viii,x}{iii,iv,v,viii,xi}{iii,iv,v,ix,x}{iii,iv,v,ix,xi}{iii,iv,v,x,xi}{iii,iv,vi,vii,viii}{iii,iv,vi,vii,ix}{iii,iv,vi,vii,x}{iii,iv,vi,vii,xi}{iii,iv,vi,viii,ix}{iii,iv,vi,viii,x}{iii,iv,vi,viii,xi}{iii,iv,vi,ix,x}{iii,iv,vi,ix,xi}{iii,iv,vi,x,xi}{iii,iv,vii,viii,ix}{iii,iv,vii,viii,x}{iii,iv,vii,viii,xi}{iii,iv,vii,ix,x}{iii,iv,vii,ix,xi}{iii,iv,vii,x,xi}{iii,iv,viii,ix,x}{iii,iv,viii,ix,xi}{iii,iv,viii,x,xi}{iii,iv,ix,x,xi}{iii,v,vi,vii,viii}{iii,v,vi,vii,ix}{iii,v,vi,vii,x}{iii,v,vi,vii,xi}{iii,v,vi,viii,ix}{iii,v,vi,viii,x}{iii,v,vi,viii,xi}{iii,v,vi,ix,x}{iii,v,vi,ix,xi}{iii,v,vi,x,xi}{iii,v,vii,viii,ix}{iii,v,vii,viii,xi}{iii,v,vii,ix,x}{iii,v,vii,ix,xi}{iii,v,vii,x,xi}{iii,v,viii,ix,x}{iii,v,viii,ix,xi}{iii,v,viii,x,xi}{iii,v,ix,x,xi}{iii,vi,vii,viii,ix}{iii,vi,vii,viii,x}{iii,vi,vii,viii,xi}{iii,vi,vii,ix,x}{iii,vi,vii,ix,xi}{iii,vi,vii,x,xi}{iii,vi,viii,ix,x}{iii,vi,viii,ix,xi}{iii,vi,viii,ix,xi}{iii,vi,ix,x,xi}{iii,vii,viii,ix,x}{iii,vii,viii,ix,xi}{iii,vii,viii,ix,xi}{iii,vii,viii,x,xi}{iii,vii,ix,x,xi}{iii,viii,ix,x,xi}{iv,v,vi,vii,viii}{iv,v,vi,vii,ix}{iv,v,vi,vii,x}{iv,v,vi,vii,xi}{iv,v,vi,viii,ix}{iv,v,vi,viii,x}{iv,v,vi,viii,xi}{iv,v,vi,ix,x}{iv,v,vi,ix,xi}{iv,v,vi,x,xi}{iv,v,vii,viii,ix}{iv,v,

vii,viii,x} {iv,v,vii,viii,xi} {iv,v,vii,ix,x} {iv,v,vii,ix,xi} {iv,v,vii,x,xi}
 {iv,v,viii,ix,x} {iv,v,viii,ix,xi} {iv,v,viii,x,xi} {iv,v,ix,x,xi} {iv,vi,vii,
 viii,ix} {iv,vi,vii,viii,x} {iv,vi,vii,viii,xi} {iv,vi,vii,ix,x} {iv,vi,vii,ix,
 xi} {iv,vi,vii,x,xi} {iv,vi,viii,ix,x} {iv,vi,viii,ix,xi} {iv,vi,viii,x,xi} {iv,
 vi,ix,x,xi} {iv,vii,viii,ix,x} {iv,vii,viii,ix,xi} {iv,vii,viii,x,xi} {iv,vii,ix,
 x,xi} {iv,viii,ix,x,xi} {v,vi,vii,viii,ix} {v,vi,vii,viii,x} {v,vi,vii,viii,xi}
 {v,vi,vii,ix,x} {v,vi,vii,ix,xi} {v,vi,vii,x,xi} {v,vi,viii,ix,x} {v,vi,viii,ix,
 xi} {v,vi,viii,x,xi} {v,vi,ix,x,xi} {v,vii,viii,ix,x} {v,vii,viii,ix,xi} {v,vii,
 viii,x,xi} {v,vii,ix,x,xi} {v,viii,ix,x,xi} {vi,vii,viii,ix,x} {vi,vii,viii,ix,
 xi} {vi,vii,viii,x,xi} {vi,vii,ix,x,xi} {vi,viii,ix,x,xi} {vii,viii,ix,x,xi} {i,
 ii,iii,iv,v,vi} {i,ii,iii,iv,v,vii} {i,ii,iii,iv,v,viii} {i,ii,iii,iv,v,ix} {i,
 ii,iii,iv,v,x} {i,ii,iii,iv,v,xi} {i,ii,iii,iv,vi,vii} {i,ii,iii,iv,vi,viii} {i,
 ii,iii,iv,vi,ix} {i,ii,iii,iv,vi,x} {i,ii,iii,iv,vi,xi} {i,ii,iii,iv,vii,viii}
 {i,ii,iii,iv,vii,ix} {i,ii,iii,iv,vii,x} {i,ii,iii,iv,vii,xi} {i,ii,iii,iv,viii,
 ix} {i,ii,iii,iv,viii,x} {i,ii,iii,iv,viii,xi} {i,ii,iii,iv,ix,x} {i,ii,iii,iv,
 ix,xi} {i,ii,iii,iv,x,xi} {i,ii,iii,v,vi,vii} {i,ii,iii,v,vi,viii} {i,ii,iii,v,
 vi,ix} {i,ii,iii,v,vi,x} {i,ii,iii,v,vi,xi} {i,ii,iii,v,vii,viii} {i,ii,iii,v,
 vii,ix} {i,ii,iii,v,vii,x} {i,ii,iii,v,vii,xi} {i,ii,iii,v,viii,ix} {i,ii,iii,v,
 viii,x} {i,ii,iii,v,viii,xi} {i,ii,iii,v,ix,x} {i,ii,iii,v,ix,xi} {i,ii,iii,v,x,
 xi} {i,ii,iii,vi,vii,viii} {i,ii,iii,vi,vii,ix} {i,ii,iii,vi,vii,x} {i,ii,iii,vi,
 vii,xi} {i,ii,iii,vi,viii,ix} {i,ii,iii,vi,viii,x} {i,ii,iii,vi,viii,xi} {i,ii,
 iii,vi,ix,x} {i,ii,iii,vi,ix,xi} {i,ii,iii,vi,x,xi} {i,ii,iii,vii,viii,ix} {i,ii,
 iii,vii,viii,x} {i,ii,iii,vii,viii,xi} {i,ii,iii,vii,ix,x} {i,ii,iii,vii,ix,xi}
 {i,ii,iii,vii,x,xi} {i,ii,iii,viii,ix,x} {i,ii,iii,viii,ix,xi} {i,ii,iii,viii,x,
 xi} {i,ii,iii,ix,x,xi} {i,ii,iv,v,vi,vii} {i,ii,iv,v,vi,viii} {i,ii,iv,v,vi,ix}
 {i,ii,iv,v,vi,x} {i,ii,iv,v,vi,xi} {i,ii,iv,v,vii,viii} {i,ii,iv,v,vii,ix} {i,ii,
 iv,v,vii,x} {i,ii,iv,v,vii,xi} {i,ii,iv,v,viii,ix} {i,ii,iv,v,viii,x} {i,ii,iv,v,
 viii,xi} {i,ii,iv,v,ix,x} {i,ii,iv,v,ix,xi} {i,ii,iv,v,x,xi} {i,ii,iv,vi,vii,
 viii} {i,ii,iv,vi,vii,ix} {i,ii,iv,vi,vii,x} {i,ii,iv,vi,vii,xi} {i,ii,iv,vi,
 viii,ix} {i,ii,iv,vi,viii,x} {i,ii,iv,vi,viii,xi} {i,ii,iv,vi,ix,x} {i,ii,iv,vi,
 ix,xi} {i,ii,iv,vi,x,xi} {i,ii,iv,vii,viii,ix} {i,ii,iv,vii,viii,x} {i,ii,iv,vii,
 viii,xi} {i,ii,iv,vii,ix,x} {i,ii,iv,vii,ix,xi} {i,ii,iv,vii,x,xi} {i,ii,iv,viii,
 ix,x} {i,ii,iv,viii,ix,xi} {i,ii,iv,viii,x,xi} {i,ii,iv,ix,x,xi} {i,ii,v,vi,vii,
 viii} {i,ii,v,vi,vii,ix} {i,ii,v,vi,vii,x} {i,ii,v,vi,vii,xi} {i,ii,v,vi,viii,ix}
 {i,ii,v,vi,viii,x} {i,ii,v,vi,viii,xi} {i,ii,v,vi,ix,x} {i,ii,v,vi,ix,xi} {i,ii,
 v,vi,x,xi} {i,ii,v,vii,viii,ix} {i,ii,v,vii,viii,x} {i,ii,v,vii,viii,xi} {i,ii,v,
 vii,ix,x} {i,ii,v,vii,ix,xi} {i,ii,v,vii,x,xi} {i,ii,v,viii,ix,x} {i,ii,v,viii,
 ix,xi} {i,ii,v,viii,x,xi} {i,ii,v,ix,x,xi} {i,ii,vi,vii,viii,ix} {i,ii,vi,vii,
 viii,x} {i,ii,vi,vii,viii,xi} {i,ii,vi,vii,ix,x} {i,ii,vi,vii,ix,xi} {i,ii,vi,

vii,x,xi} {i,ii,vi,viii,ix,x} {i,ii,vi,viii,ix,xi} {i,ii,vi,viii,x,xi} {i,ii,vi,ix,x,xi} {i,ii,vii,viii,ix,x} {i,ii,vii,viii,ix,xi} {i,ii,vii,viii,x,xi} {i,ii,vii,ix,x,xi} {i,ii,viii,ix,x,xi} {i,iii,iv,v,vi,vii} {i,iii,iv,v,vi,viii} {i,iii,iv,v,vi,ix} {i,iii,iv,v,vi,x} {i,iii,iv,v,vi,xi} {i,iii,iv,v,vii,viii} {i,iii,iv,v,vii,ix} {i,iii,iv,v,vii,x} {i,iii,iv,v,vii,xi} {i,iii,iv,v,viii,ix} {i,iii,iv,v,viii,x} {i,iii,iv,v,viii,xi} {i,iii,iv,v,ix,x} {i,iii,iv,v,ix,xi} {i,iii,iv,v,x,xi} {i,iii,iv,vi,vii,viii} {i,iii,iv,vi,vii,ix} {i,iii,iv,vi,vii,x} {i,iii,iv,vi,vii,xi} {i,iii,iv,vi,viii,ix} {i,iii,iv,vi,viii,x} {i,iii,iv,vi,viii,xi} {i,iii,iv,vi,ix,x} {i,iii,iv,vi,ix,xi} {i,iii,iv,vi,x,xi} {i,iii,iv,vii,viii,ix} {i,iii,iv,vii,viii,x} {i,iii,iv,vii,viii,xi} {i,iii,iv,vii,ix,x} {i,iii,iv,vii,ix,xi} {i,iii,iv,vii,x,xi} {i,iii,iv,viii,ix,x} {i,iii,iv,viii,ix,xi} {i,iii,iv,viii,x,xi} {i,iii,iv,ix,x,xi} {i,iii,v,vi,vii,viii} {i,iii,v,vi,vii,ix} {i,iii,v,vi,vii,x} {i,iii,v,vi,vii,xi} {i,iii,v,vi,viii,ix} {i,iii,v,vi,viii,x} {i,iii,v,vi,viii,xi} {i,iii,v,vi,ix,x} {i,iii,v,vi,ix,xi} {i,iii,v,vi,x,xi} {i,iii,v,vii,viii,ix} {i,iii,v,vii,viii,x} {i,iii,v,vii,viii,xi} {i,iii,v,vii,ix,x} {i,iii,v,vii,ix,xi} {i,iii,v,vii,x,xi} {i,iii,v,viii,ix,x} {i,iii,v,viii,ix,xi} {i,iii,v,viii,x,xi} {i,iii,v,ix,x,xi} {i,iii,vi,vii,viii,ix} {i,iii,vi,vii,viii,x} {i,iii,vi,vii,viii,xi} {i,iii,vi,vii,ix,x} {i,iii,vi,vii,ix,xi} {i,iii,vi,vii,x,xi} {i,iii,vi,viii,ix,x} {i,iii,vi,viii,ix,xi} {i,iii,vi,viii,x,xi} {i,iii,vi,ix,x,xi} {i,iii,vii,viii,ix,x} {i,iii,vii,viii,ix,xi} {i,iii,vii,viii,x,xi} {i,iii,vii,ix,x,xi} {i,iii,viii,ix,x,xi} {i,iv,v,vi,vii,viii} {i,iv,v,vi,vii,ix} {i,iv,v,vi,vii,x} {i,iv,v,vi,vii,xi} {i,iv,v,vi,viii,ix} {i,iv,v,vi,viii,x} {i,iv,v,vi,viii,xi} {i,iv,v,vi,ix,x} {i,iv,v,vi,ix,xi} {i,iv,v,vi,x,xi} {i,iv,v,vii,viii,ix} {i,iv,v,vii,viii,x} {i,iv,v,vii,viii,xi} {i,iv,v,vii,ix,x} {i,iv,v,vii,ix,xi} {i,iv,v,vii,x,xi} {i,iv,v,viii,ix,x} {i,iv,v,viii,ix,xi} {i,iv,v,viii,x,xi} {i,iv,v,ix,x,xi} {i,iv,vi,vii,viii,ix} {i,iv,vi,vii,viii,x} {i,iv,vi,vii,viii,xi} {i,iv,vi,vii,ix,x} {i,iv,vi,vii,ix,xi} {i,iv,vi,vii,x,xi} {i,iv,vi,viii,ix,x} {i,iv,vi,viii,ix,xi} {i,iv,vi,viii,x,xi} {i,iv,vi,ix,x,xi} {i,iv,vi,vii,viii,ix} {i,iv,vi,vii,viii,xi} {i,iv,vi,vii,ix,x} {i,iv,vi,vii,ix,xi} {i,iv,vi,vii,x,xi} {i,iv,vi,viii,ix,x} {i,iv,vi,viii,ix,xi} {i,iv,vi,viii,x,xi} {i,v,vi,vii,viii,ix} {i,v,vi,vii,viii,x} {i,v,vi,vii,viii,xi} {i,v,vi,vii,ix,x} {i,v,vi,vii,ix,xi} {i,v,vi,vii,x,xi} {i,v,vi,viii,ix,x} {i,v,vi,viii,ix,xi} {i,v,vi,viii,x,xi} {i,v,vi,ix,x,xi} {i,v,vii,viii,ix,x} {i,v,vii,viii,ix,xi} {i,v,vii,viii,x,xi} {i,v,vii,ix,x,xi} {i,v,viii,ix,x,xi} {i,vi,vii,viii,ix,x} {i,vi,vii,viii,ix,xi} {i,vi,vii,viii,x,xi} {i,vii,viii,ix,x,xi} {ii,iii,iv,v,vi,vii} {ii,iii,iv,v,vi,viii} {ii,iii,iv,v,vi,ix} {ii,iii,iv,v,vi,x} {ii,iii,iv,v,vi,xi} {ii,iii,iv,v,vii,viii} {ii,iii,iv,v,vii,ix} {ii,iii,iv,v,vii,x} {ii,iii,iv,v,vii,xi} {ii,iii,iv,v,viii,ix} {ii,iii,iv,v,viii,x} {ii,iii,iv,v,viii,xi} {ii,iii,iv,v,ix,x} {ii,iii,iv,v,ix,xi} {ii,iii,iv,v,x,xi} {ii,iii,iv,vi,vii,viii} {ii,iii,iv,vi,vii,ix} {ii,iii,iv,vi,vii,x} {ii,iii,iv,vi,vii,xi} {ii,iii,iv,vi,viii,ix} {ii,iii,iv,vi,viii,x} {ii,iii,iv,vi,viii,xi} {ii,iii,iv,vi,ix,x} {ii,iii,iv,vi,ix,xi} {ii,iii,iv,vi,x,xi} {ii,iii,iv,vii,viii} {ii,iii,iv,vii,ix} {ii,iii,iv,vii,x} {ii,iii,iv,vii,xi} {ii,iii,iv,vi,vii,xi}

{ii,iii,iv,vi,viii,ix} {ii,iii,iv,vi,viii,x} {ii,iii,iv,vi,viii,xi} {ii,iii,iv,
 vi,ix,x} {ii,iii,iv,vi,ix,xi} {ii,iii,iv,vi,x,xi} {ii,iii,iv,vii,viii,ix} {ii,
 iii,iv,vii,viii,x} {ii,iii,iv,vii,viii,xi} {ii,iii,iv,vii,ix,x} {ii,iii,iv,vii,
 ix,xi} {ii,iii,iv,vii,x,xi} {ii,iii,iv,viii,ix,x} {ii,iii,iv,viii,ix,xi} {ii,iii,
 iv,viii,x,xi} {ii,iii,iv,ix,x,xi} {ii,iii,v,vi,vii,viii} {ii,iii,v,vi,vii,ix}
 {ii,iii,v,vi,vii,x} {ii,iii,v,vi,vii,xi} {ii,iii,v,vi,viii,ix} {ii,iii,v,vi,
 viii,x} {ii,iii,v,vi,viii,xi} {ii,iii,v,vi,ix,x} {ii,iii,v,vi,ix,xi} {ii,iii,v,
 vi,x,xi} {ii,iii,v,vii,viii,ix} {ii,iii,v,vii,viii,x} {ii,iii,v,vii,viii,xi} {ii,
 iii,v,vii,ix,x} {ii,iii,v,vii,ix,xi} {ii,iii,v,vii,x,xi} {ii,iii,v,viii,ix,x}
 {ii,iii,v,viii,ix,xi} {ii,iii,v,viii,x,xi} {ii,iii,v,ix,x,xi} {ii,iii,vi,vii,
 viii,ix} {ii,iii,vi,vii,viii,x} {ii,iii,vi,vii,viii,xi} {ii,iii,vi,vii,ix,x} {ii,
 iii,vi,vii,ix,xi} {ii,iii,vi,vii,x,xi} {ii,iii,vi,viii,ix,x} {ii,iii,vi,viii,ix,
 xi} {ii,iii,vi,viii,x,xi} {ii,iii,vi,ix,x,xi} {ii,iii,vii,viii,ix,x} {ii,iii,vii,
 viii,ix,xi} {ii,iii,vii,viii,x,xi} {ii,iii,vii,ix,x,xi} {ii,iii,viii,ix,x,xi}
 {ii,iv,v,vi,vii,viii} {ii,iv,v,vi,vii,ix} {ii,iv,v,vi,vii,x} {ii,iv,v,vi,vii,xi}
 {ii,iv,v,vi,viii,ix} {ii,iv,v,vi,viii,x} {ii,iv,v,vi,viii,xi} {ii,iv,v,vi,ix,x}
 {ii,iv,v,vi,ix,xi} {ii,iv,v,vi,x,xi} {ii,iv,v,vii,viii,ix} {ii,iv,v,vii,viii,x}
 {ii,iv,v,vii,viii,xi} {ii,iv,v,vii,ix,x} {ii,iv,v,vii,ix,xi} {ii,iv,v,vii,x,xi}
 {ii,iv,v,viii,ix,x} {ii,iv,v,viii,ix,xi} {ii,iv,v,viii,x,xi} {ii,iv,v,ix,x,xi}
 {ii,iv,vi,vii,viii,ix} {ii,iv,vi,vii,viii,x} {ii,iv,vi,vii,viii,xi} {ii,iv,vi,
 vii,ix,x} {ii,iv,vi,vii,ix,xi} {ii,iv,vi,vii,x,xi} {ii,iv,vi,viii,ix,x} {ii,iv,
 vi,viii,ix,xi} {ii,iv,vi,viii,x,xi} {ii,iv,vi,ix,x,xi} {ii,iv,vii,viii,ix,x} {ii,
 iv,vii,viii,ix,xi} {ii,iv,vii,viii,x,xi} {ii,iv,vii,ix,x,xi} {ii,iv,viii,ix,x,
 xi} {ii,v,vi,vii,viii,ix} {ii,v,vi,vii,viii,x} {ii,v,vi,vii,viii,xi} {ii,v,vi,
 vii,ix,x} {ii,v,vi,vii,ix,xi} {ii,v,vi,vii,x,xi} {ii,v,vi,viii,ix,x} {ii,v,vi,
 viii,ix,xi} {ii,v,vi,viii,x,xi} {ii,v,vi,ix,x,xi} {ii,v,vii,viii,ix,x} {ii,v,vii,
 viii,ix,xi} {ii,v,vii,viii,x,xi} {ii,v,vii,ix,x,xi} {ii,v,viii,ix,x,xi} {ii,vi,
 vii,viii,ix,x} {ii,vi,vii,viii,ix,xi} {ii,vi,vii,viii,x,xi} {ii,vi,vii,ix,x,xi}
 {ii,vi,viii,ix,x,xi} {ii,vii,viii,ix,x,xi} {iii,iv,v,vi,vii,viii} {iii,iv,v,vi,
 vii,ix} {iii,iv,v,vi,vii,x} {iii,iv,v,vi,vii,xi} {iii,iv,v,vi,viii,ix} {iii,iv,v,
 vi,viii,x} {iii,iv,v,vi,viii,xi} {iii,iv,v,vi,ix,x} {iii,iv,v,vi,ix,xi} {iii,iv,
 v,vi,x,xi} {iii,iv,v,vii,viii,ix} {iii,iv,v,vii,viii,x} {iii,iv,v,vii,viii,xi}
 {iii,iv,v,vii,ix,x} {iii,iv,v,vii,ix,xi} {iii,iv,v,vii,x,xi} {iii,iv,v,viii,ix,
 x} {iii,iv,v,viii,ix,xi} {iii,iv,v,viii,x,xi} {iii,iv,v,ix,x,xi} {iii,iv,vi,vii,
 viii,ix} {iii,iv,vi,vii,viii,x} {iii,iv,vi,vii,viii,xi} {iii,iv,vi,vii,ix,x}
 {iii,iv,vi,vii,ix,xi} {iii,iv,vi,vii,x,xi} {iii,iv,vi,viii,ix,x} {iii,iv,vi,
 viii,ix,xi} {iii,iv,vi,viii,x,xi} {iii,iv,vi,ix,x,xi} {iii,iv,vii,viii,ix,x}
 {iii,iv,vii,viii,ix,xi} {iii,iv,vii,viii,x,xi} {iii,iv,vii,ix,x,xi} {iii,iv,
 viii,ix,x,xi} {iii,v,vi,vii,viii,ix} {iii,v,vi,vii,viii,x} {iii,v,vi,vii,viii,

xi} {iii,v,vi,vii,ix,x} {iii,v,vi,vii,ix,xi} {iii,v,vi,vii,x,xi} {iii,v,vi,viii,
 ix,x} {iii,v,vi,viii,ix,xi} {iii,v,vi,viii,x,xi} {iii,v,vi,ix,x,xi} {iii,v,vii,
 viii,ix,x} {iii,v,vii,viii,ix,xi} {iii,v,vii,viii,x,xi} {iii,v,vii,ix,x,xi} {iii,
 v,viii,ix,x,xi} {iii,vi,vii,viii,ix,x} {iii,vi,vii,viii,ix,xi} {iii,vi,vii,viii,
 x,xi} {iii,vi,vii,ix,x,xi} {iii,vi,viii,ix,x,xi} {iii,vii,viii,ix,x,xi} {iv,v,vi,
 vii,viii,ix} {iv,v,vi,vii,viii,x} {iv,v,vi,vii,viii,xi} {iv,v,vi,vii,ix,x} {iv,v,
 vi,vii,ix,xi} {iv,v,vi,vii,x,xi} {iv,v,vi,viii,ix,x} {iv,v,vi,viii,ix,xi} {iv,v,
 vi,viii,x,xi} {iv,v,vi,ix,x,xi} {iv,v,vii,viii,ix,x} {iv,v,vii,viii,ix,xi} {iv,v,
 vii,viii,x,xi} {iv,v,vii,ix,x,xi} {iv,v,viii,ix,x,xi} {iv,vi,vii,viii,ix,x} {iv,
 vi,vii,viii,ix,xi} {iv,vi,vii,viii,x,xi} {iv,vi,vii,ix,x,xi} {iv,vi,viii,ix,x,
 xi} {iv,vii,viii,ix,x,xi} {v,vi,vii,viii,ix,x} {v,vi,vii,viii,ix,xi} {v,vi,vii,
 viiii,x,xi} {v,vi,vii,ix,x,xi} {v,vi,viii,ix,x,xi} {v,vii,viii,ix,x,xi} {vi,vii,
 viiii,ix,x,xi} {i,ii,iii,iv,v,vi,vii} {i,ii,iii,iv,v,vi,viii} {i,ii,iii,iv,v,vi,
 ix} {i,ii,iii,iv,v,vi,x} {i,ii,iii,iv,v,vi,xi} {i,ii,iii,iv,v,vii,viii} {i,ii,
 iii,iv,v,vii,ix} {i,ii,iii,iv,v,vii,x} {i,ii,iii,iv,v,vii,xi} {i,ii,iii,iv,v,
 viiii,ix} {i,ii,iii,iv,v,viii,x} {i,ii,iii,iv,v,viii,xi} {i,ii,iii,iv,v,ix,x} {i,
 ii,iii,iv,v,ix,xi} {i,ii,iii,iv,v,x,xi} {i,ii,iii,iv,vi,vii,viii} {i,ii,iii,iv,
 vi,vii,ix} {i,ii,iii,iv,vi,vii,x} {i,ii,iii,iv,vi,vii,xi} {i,ii,iii,iv,vi,viii,
 ix} {i,ii,iii,iv,vi,viii,x} {i,ii,iii,iv,vi,viii,xi} {i,ii,iii,iv,vi,ix,x} {i,ii,
 iii,iv,vi,ix,xi} {i,ii,iii,iv,vi,x,xi} {i,ii,iii,iv,vii,viii,ix} {i,ii,iii,iv,
 vii,viii,x} {i,ii,iii,iv,vii,viii,xi} {i,ii,iii,iv,vii,ix,x} {i,ii,iii,iv,vii,
 ix,xi} {i,ii,iii,iv,vii,x,xi} {i,ii,iii,iv,viii,ix,x} {i,ii,iii,iv,viii,ix,xi}
 {i,ii,iii,iv,viii,x,xi} {i,ii,iii,iv,ix,x,xi} {i,ii,iii,v,vi,vii,viii} {i,ii,
 iii,v,vi,vii,ix} {i,ii,iii,v,vi,vii,x} {i,ii,iii,v,vi,vii,xi} {i,ii,iii,v,vi,
 viiii,ix} {i,ii,iii,v,vi,viii,x} {i,ii,iii,v,vi,viii,xi} {i,ii,iii,v,vi,ix,x} {i,
 ii,iii,v,vi,ix,xi} {i,ii,iii,v,vi,x,xi} {i,ii,iii,v,vii,viii,ix} {i,ii,iii,v,
 vii,viii,x} {i,ii,iii,v,vii,viii,xi} {i,ii,iii,v,vii,ix,x} {i,ii,iii,v,vii,ix,
 xi} {i,ii,iii,v,vii,x,xi} {i,ii,iii,v,viii,ix,x} {i,ii,iii,v,viii,ix,xi} {i,ii,
 iii,v,viii,x,xi} {i,ii,iii,v,ix,x,xi} {i,ii,iii,vi,vii,viii,ix} {i,ii,iii,vi,
 vii,viii,x} {i,ii,iii,vi,vii,viii,xi} {i,ii,iii,vi,vii,ix,x} {i,ii,iii,vi,vii,
 ix,xi} {i,ii,iii,vi,vii,x,xi} {i,ii,iii,vi,viii,ix,x} {i,ii,iii,vi,viii,ix,xi}
 {i,ii,iii,vi,viii,x,xi} {i,ii,iii,vi,ix,x,xi} {i,ii,iii,vii,viii,ix,x} {i,ii,
 iii,vii,viii,ix,xi} {i,ii,iii,vii,viii,x,xi} {i,ii,iii,vii,ix,x,xi} {i,ii,iii,
 viiii,ix,x,xi} {i,ii,iv,v,vi,vii,viii} {i,ii,iv,v,vi,vii,ix} {i,ii,iv,v,vi,vii,x}
 {i,ii,iv,v,vi,vii,xi} {i,ii,iv,v,vi,viii,ix} {i,ii,iv,v,vi,viii,x} {i,ii,iv,v,
 vi,viii,xi} {i,ii,iv,v,vi,ix,x} {i,ii,iv,v,vi,ix,xi} {i,ii,iv,v,vi,x,xi} {i,ii,
 iv,v,vii,viii,ix} {i,ii,iv,v,vii,viii,x} {i,ii,iv,v,vii,viii,xi} {i,ii,iv,v,vii,
 ix,x} {i,ii,iv,v,vii,ix,xi} {i,ii,iv,v,vii,x,xi} {i,ii,iv,v,viii,ix,x} {i,ii,iv,
 v,viii,ix,xi} {i,ii,iv,v,viii,x,xi} {i,ii,iv,v,ix,x,xi} {i,ii,iv,vi,vii,viii,ix}

[illegible]

x} {ii,iii,iv,v,vii,ix,xi} {ii,iii,iv,v,vii,x,xi} {ii,iii,iv,v,viii,ix,x} {ii,
iii,iv,v,viii,ix,xi} {ii,iii,iv,v,viii,x,xi} {ii,iii,iv,v,ix,x,xi} {ii,iii,iv,
vi,vii,viii,ix} {ii,iii,iv,vi,vii,viii,x} {ii,iii,iv,vi,vii,viii,xi} {ii,iii,iv,
vi,vii,ix,x} {ii,iii,iv,vi,vii,ix,xi} {ii,iii,iv,vi,vii,x,xi} {ii,iii,iv,vi,
viii,ix,x} {ii,iii,iv,vi,viii,ix,xi} {ii,iii,iv,vi,viii,x,xi} {ii,iii,iv,vi,ix,
x,xi} {ii,iii,iv,vii,viii,ix,x} {ii,iii,iv,vii,viii,ix,xi} {ii,iii,iv,vii,viii,
x,xi} {ii,iii,iv,vii,ix,x,xi} {ii,iii,iv,viii,ix,x,xi} {ii,iii,v,vi,vii,viii,ix}
{ii,iii,v,vi,vii,viii,x} {ii,iii,v,vi,vii,viii,xi} {ii,iii,v,vi,vii,ix,x} {ii,
iii,v,vi,vii,ix,xi} {ii,iii,v,vi,vii,x,xi} {ii,iii,v,vi,viii,ix,x} {ii,iii,v,vi,
viii,ix,xi} {ii,iii,v,vi,viii,x,xi} {ii,iii,v,vi,ix,x,xi} {ii,iii,v,vii,viii,ix,
x} {ii,iii,v,vii,viii,ix,xi} {ii,iii,v,vii,viii,x,xi} {ii,iii,v,vii,ix,x,xi} {ii,
iii,v,viii,ix,x,xi} {ii,iii,vi,vii,viii,ix,x} {ii,iii,vi,vii,viii,ix,xi} {ii,
iii,vi,vii,viii,x,xi} {ii,iii,vi,vii,ix,x,xi} {ii,iii,vi,viii,ix,x,xi} {ii,iii,
vii,viii,ix,x,xi} {ii,iv,v,vi,vii,viii,ix} {ii,iv,v,vi,vii,viii,x} {ii,iv,v,vi,
vii,viii,xi} {ii,iv,v,vi,vii,ix,x} {ii,iv,v,vi,vii,ix,xi} {ii,iv,v,vi,vii,x,xi}
{ii,iv,v,vi,viii,ix,x} {ii,iv,v,vi,viii,ix,xi} {ii,iv,v,vi,viii,x,xi} {ii,iv,v,
vi,ix,x,xi} {ii,iv,v,vii,viii,ix,x} {ii,iv,v,vii,viii,ix,xi} {ii,iv,v,vii,viii,
x,xi} {ii,iv,v,vii,ix,x,xi} {ii,iv,v,viii,ix,x,xi} {ii,iv,vi,vii,viii,ix,x} {ii,
iv,vi,vii,viii,ix,xi} {ii,iv,vi,vii,viii,x,xi} {ii,iv,vi,vii,ix,x,xi} {ii,iv,vi,
viii,ix,x,xi} {ii,iv,vii,viii,ix,x,xi} {ii,v,vi,vii,viii,ix,x} {ii,v,vi,vii,
viii,ix,xi} {ii,v,vi,vii,viii,x,xi} {ii,v,vi,vii,ix,x,xi} {ii,v,vi,viii,ix,x,xi}
{ii,v,vii,viii,ix,x,xi} {ii,vi,vii,viii,ix,x,xi} {iii,iv,v,vi,vii,viii,ix} {iii,
iv,v,vi,vii,viii,x} {iii,iv,v,vi,vii,viii,xi} {iii,iv,v,vi,vii,ix,x} {iii,iv,v,
vi,vii,ix,xi} {iii,iv,v,vi,vii,x,xi} {iii,iv,v,vi,viii,ix,x} {iii,iv,v,vi,viii,
ix,xi} {iii,iv,v,vi,viii,x,xi} {iii,iv,v,vi,ix,x,xi} {iii,iv,v,vii,viii,ix,x}
{iii,iv,v,vii,viii,ix,xi} {iii,iv,v,vii,viii,x,xi} {iii,iv,v,vii,ix,x,xi} {iii,
iv,v,viii,ix,x,xi} {iii,iv,vi,vii,viii,ix,x} {iii,iv,vi,vii,viii,ix,xi} {iii,iv,
vi,vii,viii,x,xi} {iii,iv,vi,vii,ix,x,xi} {iii,iv,vi,viii,ix,x,xi} {iii,iv,vii,
viii,ix,x,xi} {iii,v,vi,vii,viii,ix,x} {iii,v,vi,vii,viii,ix,xi} {iii,v,vi,vii,
viii,x,xi} {iii,v,vi,vii,ix,x,xi} {iii,v,vi,viii,ix,x,xi} {iii,v,vii,viii,ix,x,
xi} {iii,vi,vii,viii,ix,x,xi} {iv,v,vi,vii,viii,ix,x} {iv,v,vi,vii,viii,ix,xi}
{iv,v,vi,vii,viii,x,xi} {iv,v,vi,vii,ix,x,xi} {iv,v,vi,viii,ix,x,xi} {iv,v,vii,
viii,ix,x,xi} {iv,vi,vii,viii,ix,x,xi} {v,vi,vii,viii,ix,x,xi} {i,ii,iii,iv,v,
vi,vii,viii} {i,ii,iii,iv,v,vi,vii,ix} {i,ii,iii,iv,v,vi,vii,x} {i,ii,iii,iv,v,
vi,vii,xi} {i,ii,iii,iv,v,vi,viii,ix} {i,ii,iii,iv,v,vi,viii,x} {i,ii,iii,iv,v,
vi,viii,xi} {i,ii,iii,iv,v,vi,ix,x} {i,ii,iii,iv,v,vi,ix,xi} {i,ii,iii,iv,v,vi,
x,xi} {i,ii,iii,iv,v,vii,viii,ix} {i,ii,iii,iv,v,vii,viii,x} {i,ii,iii,iv,v,vii,
viii,xi} {i,ii,iii,iv,v,vii,ix,x} {i,ii,iii,iv,v,vii,ix,xi} {i,ii,iii,iv,v,vii,
x,xi} {i,ii,iii,iv,v,viii,ix,x} {i,ii,iii,iv,v,viii,ix,xi} {i,ii,iii,iv,v,viii

[illegible]

$\{iv,v,viii,ix,x,xi\} \{ii,iii,iv,vi,vii,viii,ix,x\} \{ii,iii,iv,vi,vii,viii,ix,xi\}$
 $\{ii,iii,iv,vi,vii,viii,x,xi\} \{ii,iii,iv,vi,vii,ix,x,xi\} \{ii,iii,iv,vi,viii,ix,$
 $x,xi\} \{ii,iii,iv,vii,viii,ix,x,xi\} \{ii,iii,v,vi,vii,viii,ix,x\} \{ii,iii,v,vi,vii,$
 $viii,ix,xi\} \{ii,iii,v,vi,vii,viii,x,xi\} \{ii,iii,v,vi,vii,ix,x,xi\} \{ii,iii,v,vi,$
 $viii,ix,x,xi\} \{ii,iii,v,vii,viii,ix,x,xi\} \{ii,iii,vi,vii,viii,ix,x,xi\} \{ii,iv,v,$
 $vi,vii,viii,ix,x\} \{ii,iv,v,vi,vii,viii,ix,xi\} \{ii,iv,v,vi,vii,viii,x,xi\} \{ii,iv,$
 $v,vi,vii,ix,x,xi\} \{ii,iv,v,vi,viii,ix,x,xi\} \{ii,iv,v,vii,viii,ix,x,xi\} \{ii,iv$
 $vi,vii,viii,ix,x,xi\} \{ii,v,vi,vii,viii,ix,x,xi\} \{iii,iv,v,vi,vii,viii,ix,x\}$
 $\{iii,iv,v,vi,vii,viii,ix,xi\} \{iii,iv,v,vi,vii,viii,x,xi\} \{iii,iv,v,vi,vii,ix,x,$
 $xi\} \{iii,iv,v,vi,viii,ix,x,xi\} \{iii,iv,v,vii,viii,ix,x,xi\} \{iii,iv,vi,vii,viii,$
 $ix,x,xi\} \{iii,v,vi,vii,viii,ix,x,xi\} \{iv,v,vi,vii,viii,ix,x,xi\} \{i,ii,iii,iv,v,$
 $vi,vii,viii,ix\} \{i,ii,iii,iv,v,vi,vii,viii,x\} \{i,ii,iii,iv,v,vi,vii,viii,xi\} \{i,$
 $ii,iii,iv,v,vi,vii,ix,x\} \{i,ii,iii,iv,v,vi,vii,ix,xi\} \{i,ii,iii,iv,v,vi,vii,x,$
 $xi\} \{i,ii,iii,iv,v,vi,viii,ix,x\} \{i,ii,iii,iv,v,vi,viii,ix,xi\} \{i,ii,iii,iv,v,$
 $vi,viii,x,xi\} \{i,ii,iii,iv,v,vi,ix,x,xi\} \{i,ii,iii,iv,v,vii,viii,ix,x\} \{i,ii,$
 $iii,iv,v,vii,viii,ix,xi\} \{i,ii,iii,iv,v,vii,viii,x,xi\} \{i,ii,iii,iv,v,vii,ix,x,$
 $xi\} \{i,ii,iii,iv,v,viii,ix,x,xi\} \{i,ii,iii,iv,vi,vii,viii,ix,x\} \{i,ii,iii,iv,vi,$
 $vii,viii,ix,xi\} \{i,ii,iii,iv,vi,vii,viii,x,xi\} \{i,ii,iii,iv,vi,vii,ix,x,xi\} \{i,$
 $ii,iii,iv,vi,viii,ix,x,xi\} \{i,ii,iii,iv,vii,viii,ix,x,xi\} \{i,ii,iii,v,vi,vii,$
 $viii,ix,x\} \{i,ii,iii,v,vi,vii,viii,ix,xi\} \{i,ii,iii,v,vi,vii,viii,x,xi\} \{i,ii,$
 $iii,v,vi,vii,ix,x,xi\} \{i,ii,iii,v,vi,viii,ix,x,xi\} \{i,ii,iii,v,vii,viii,ix,x,$
 $xi\} \{i,ii,iii,vi,vii,viii,ix,x,xi\} \{i,ii,iv,v,vi,vii,viii,ix,x\} \{i,ii,iv,v,vi,$
 $vii,viii,ix,xi\} \{i,ii,iv,v,vi,vii,viii,x,xi\} \{i,ii,iv,v,vi,vii,ix,x,xi\} \{i,ii,$
 $iv,v,vi,viii,ix,x,xi\} \{i,ii,iv,v,vii,viii,ix,x,xi\} \{i,ii,iv,vi,vii,viii,ix,x,$
 $xi\} \{i,ii,v,vi,vii,viii,ix,x,xi\} \{i,iii,iv,v,vi,vii,viii,ix,x\} \{i,iii,iv,v,vi,$
 $vii,viii,ix,xi\} \{i,iii,iv,v,vi,vii,viii,x,xi\} \{i,iii,iv,v,vi,vii,ix,x,xi\} \{i,$
 $iii,iv,v,vi,viii,ix,x,xi\} \{i,iii,iv,v,vii,viii,ix,x,xi\} \{i,iii,iv,vi,vii,viii,$
 $ix,x,xi\} \{i,iii,v,vi,vii,viii,ix,x,xi\} \{i,iv,v,vi,vii,viii,ix,x,xi\} \{ii,iii,iv,$
 $v,vi,vii,viii,ix,x\} \{ii,iii,iv,v,vi,vii,viii,ix,xi\} \{ii,iii,iv,v,vi,vii,viii,x,$
 $xi\} \{ii,iii,iv,v,vi,vii,ix,x,xi\} \{ii,iii,iv,v,vi,viii,ix,x,xi\} \{ii,iii,iv,v,vii,$
 $viii,ix,x,xi\} \{ii,iii,iv,vi,vii,viii,ix,x,xi\} \{ii,iii,v,vi,vii,viii,ix,x,xi\}$
 $\{ii,iv,v,vi,vii,viii,ix,x,xi\} \{iii,iv,v,vi,vii,viii,ix,x,xi\} \{i,ii,iii,iv,v,vi,$
 $vii,viii,ix,x\} \{i,ii,iii,iv,v,vi,vii,viii,ix,xi\} \{i,ii,iii,iv,v,vi,vii,viii,x,$
 $xi\} \{i,ii,iii,iv,v,vi,vii,ix,x,xi\} \{i,ii,iii,iv,v,vi,viii,ix,x,xi\} \{i,ii,iii,iv,$
 $v,vii,viii,ix,x,xi\} \{i,ii,iii,iv,vi,vii,viii,ix,x,xi\} \{i,ii,iii,v,vi,vii,viii,$
 $ix,x,xi\} \{i,ii,iv,v,vi,vii,viii,ix,x,xi\} \{i,iii,iv,v,vi,vii,viii,ix,x,xi\} \{ii,$
 $iii,iv,v,vi,vii,viii,ix,x,xi\}$ 或 $\{i,ii,iii,iv,v,vi,vii,viii,ix,x,xi\}$

[0214] 如果变异体包含(i)和(iii)到(xi)中的任一个,那么其可进一步包含在Y51、N55和F56中的一或多个处的突变,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/

F56处。

[0215] 在(i)中,变异体可包含在任何数量和以下的组合处突变:N40、D43、E44、S54、S57、Q62、R97、E101、E124、E131、R142、T150和R192。在(i)中,变异体优选地包含在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变):N40、D43、E44、S54、S57、Q62、E101、E131和T150。在(i)中,变异体优选地包含在以下位置处的一或多个突变(即在以下位置中的一或多个处的突变):N40、D43、E44、E101和E131。在(i)中,变异体优选地包含在S54和/或S57处的突变。在(i)中,变异体更优选地包含在以下处的突变:(a)S54和/或S57;和(b)Y51、N55和F56中的一或多个,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。如果在(xi)中缺失S54和/或S57,那么其在(i)中不会为突变的且反之亦然。在(i)中,变异体优选地包含在T150处的突变,如T150I。或者,变异体优选地包含在以下处的突变:(a)T150;和(b)Y51、N55和F56中的一或多个,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。在(i)中,变异体优选地包含在Q62处的突变,如Q62R或Q62K。或者,变异体优选地包含在以下处的突变:(a)Q62;和(b)Y51、N55和F56中的一或多个,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。变异体可包含在D43、E44、Q62处或其任何组合的突变,如D43、E44、Q62、D43/E44、D43/Q62、E44/Q62或D43/E44/Q62。或者,变异体优选地包含在以下处的突变:(a)D43、E44、Q62、D43/E44、D43/Q62、E44/Q62或D43/E44/Q62;和(b)Y51、N55和F56中的一或多个,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。

[0216] 在(ii)和由/符号分隔的不同位置的本申请的其它地方中,/符号意指“和”,从而Y51/N55是Y51和N55。在(ii)中,变异体优选地包含在Y51/N55处的突变。已提出,CsgG中的收缩是由通过残基Y51、N55和F56的侧链形成的三个堆叠式同心环构成(Goyal等人,2014,《自然(Nature)》,516,250-253)。因此,(ii)中的这些残基的突变可降低在聚核苷酸移动通过孔时引起电流的核苷酸的数量,且从而使得其更容易鉴定所观测的电流(在聚核苷酸移动通过孔时)与聚核苷酸之间的直接关系。F56可以下文参考适用于本发明的方法中的变异体和孔而论述任一方式来突变。

[0217] 在(v)中,变异体可包含N102R、N102F、N102Y或N102W。变异体优选地包含:(a)N102R、N102F、N102Y或N102W;和(b)在Y51、N55和F56中的一或多个处的突变,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。

[0218] 在(xi)中,可缺失任何数量和以下的组合:K49、P50、Y51、P52、A53、S54、N55、F56和S57。优选地,可缺失K49、P50、Y51、P52、A53、S54、N55和S57中的一或多个。如果在(xi)中缺失Y51、N55和F56中的任一个,那么其在(ii)中不会为突变的且反之亦然。

[0219] 在(i)中,变异体优选地包含以下取代中的一或多个:N40R、N40K、D43N、D43Q、D43R、D43K、E44N、E44Q、E44R、E44K、S54P、S57P、Q62R、Q62K、R97N、R97G、R97L、E101N、E101Q、E101R、E101K、E101F、E101Y、E101W、E124N、E124Q、E124R、E124K、E124F、E124Y、E124W、E131D、R142E、R142N、T150I、R192E和R192N,如以下中的一或多个:N40R、N40K、D43N、D43Q、D43R、D43K、E44N、E44Q、E44R、E44K、S54P、S57P、Q62R、Q62K、E101N、E101Q、E101R、E101K、E101F、E101Y、E101W、E131D和T150I,或以下中的一或多个:N40R、N40K、D43N、D43Q、D43R、D43K、E44N、E44Q、E44R、E44K、E101N、E101Q、E101R、E101K、E101F、E101Y、E101W和E131D。变异体可包含任何数量的取代和这些取代的组合。在(i)中,变异体优选地包含S54P和/或

S57P。在(i)中,变异体优选地包含:(a) S54P和/或S57P;和(b) 在Y51、N55和F56中的一或多个处的突变,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。在Y51、N55和F56中的一或多个处的突变可以是下文论述的那些中的任一个。在(i)中,变异体优选地包含F56A/S57P或S54P/F56A。变异体优选地包含T150I。或者,变异体优选地包含在以下处的突变:(a) T150I;和(b) Y51、N55和F56中的一或多个,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。

[0220] 在(i)中,变异体优选地包含Q62R或Q62K。或者,变异体优选地包含:(a) Q62R或Q62K;和(b) 在Y51、N55和F56中的一或多个处的突变,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。变异体可包含D43N、E44N、Q62R或Q62K或其任何组合,如D43N、E44N、Q62R、Q62K、D43N/E44N、D43N/Q62R、D43N/Q62K、E44N/Q62R、E44N/Q62K、D43N/E44N/Q62R或D43N/E44N/Q62K。或者,变异体优选地包含:(a) D43N、E44N、Q62R、Q62K、D43N/E44N、D43N/Q62R、D43N/Q62K、E44N/Q62R、E44N/Q62K、D43N/E44N/Q62R或D43N/E44N/Q62K;和(b) 在Y51、N55和F56中的一或多个处的突变,如在Y51、N55、F56、Y51/N55、Y51/F56、N55/F56或Y51/N55/F56处。

[0221] 在(i)中,变异体优选地包含D43N。

[0222] 在(i)中,变异体优选地包含E101R、E101S、E101F或E101N。

[0223] 在(i)中,变异体优选地包含E124N、E124Q、E124R、E124K、E124F、E124Y、E124W或E124D,如E124N。

[0224] 在(i)中,变异体优选地包含R142E和R142N。

[0225] 在(i)中,变异体优选地包含R97N、R97G或R97L。

[0226] 在(i)中,变异体优选地包含R192E和R192N。

[0227] 在(ii)中,变异体优选地包含:F56N/N55Q、F56N/N55R、F56N/N55K、F56N/N55S、F56N/N55G、F56N/N55A、F56N/N55T、F56Q/N55Q、F56Q/N55R、F56Q/N55K、F56Q/N55S、F56Q/N55G、F56Q/N55A、F56Q/N55T、F56R/N55Q、F56R/N55R、F56R/N55K、F56R/N55S、F56R/N55G、F56R/N55A、F56R/N55T、F56S/N55Q、F56S/N55R、F56S/N55K、F56S/N55S、F56S/N55G、F56S/N55A、F56S/N55T、F56G/N55Q、F56G/N55R、F56G/N55K、F56G/N55S、F56G/N55G、F56G/N55A、F56G/N55T、F56A/N55Q、F56A/N55R、F56A/N55K、F56A/N55S、F56A/N55G、F56A/N55A、F56A/N55T、F56K/N55Q、F56K/N55R、F56K/N55K、F56K/N55S、F56K/N55G、F56K/N55A、F56K/N55T、F56N/Y51L、F56N/Y51V、F56N/Y51A、F56N/Y51N、F56N/Y51Q、F56N/Y51S、F56N/Y51G、F56Q/Y51L、F56Q/Y51V、F56Q/Y51A、F56Q/Y51N、F56Q/Y51Q、F56Q/Y51S、F56Q/Y51G、F56R/Y51L、F56R/Y51V、F56R/Y51A、F56R/Y51N、F56R/Y51Q、F56R/Y51S、F56R/Y51G、F56S/Y51L、F56S/Y51V、F56S/Y51A、F56S/Y51N、F56S/Y51Q、F56S/Y51S、F56S/Y51G、F56G/Y51L、F56G/Y51V、F56G/Y51A、F56G/Y51N、F56G/Y51Q、F56G/Y51S、F56G/Y51G、F56A/Y51L、F56A/Y51V、F56A/Y51A、F56A/Y51N、F56A/Y51Q、F56A/Y51S、F56A/Y51G、F56K/Y51L、F56K/Y51V、F56K/Y51A、F56K/Y51N、F56K/Y51Q、F56K/Y51S、F56K/Y51G、N55Q/Y51L、N55Q/Y51V、N55Q/Y51A、N55Q/Y51N、N55Q/Y51Q、N55Q/Y51S、N55Q/Y51G、N55R/Y51L、N55R/Y51V、N55R/Y51A、N55R/Y51N、N55R/Y51Q、N55R/Y51S、N55R/Y51G、N55K/Y51L、N55K/Y51V、N55K/Y51A、N55K/Y51N、N55K/Y51Q、N55K/Y51S、N55K/Y51G、N55S/Y51L、N55S/Y51V、N55S/Y51A、N55S/Y51N、N55S/Y51Q、N55S/Y51S、N55S/Y51G、N55G/Y51L、N55G/Y51V、N55G/Y51A、N55G/Y51N、N55G/Y51Q、N55G/

Y51S、N55G/Y51G、N55A/Y51L、N55A/Y51V、N55A/Y51A、N55A/Y51N、N55A/Y51Q、N55A/Y51S、
N55A/Y51G、N55T/Y51L、N55T/Y51V、N55T/Y51A、N55T/Y51N、N55T/Y51Q、N55T/Y51S、N55T/
Y51G、F56N/N55Q/Y51L、F56N/N55Q/Y51V、F56N/N55Q/Y51A、F56N/N55Q/Y51N、F56N/N55Q/
Y51Q、F56N/N55Q/Y51S、F56N/N55Q/Y51G、F56N/N55R/Y51L、F56N/N55R/Y51V、F56N/N55R/
Y51A、F56N/N55R/Y51N、F56N/N55R/Y51Q、F56N/N55R/Y51S、F56N/N55R/Y51G、F56N/N55K/
Y51L、F56N/N55K/Y51V、F56N/N55K/Y51A、F56N/N55K/Y51N、F56N/N55K/Y51Q、F56N/N55K/
Y51S、F56N/N55K/Y51G、F56N/N55S/Y51L、F56N/N55S/Y51V、F56N/N55S/Y51A、F56N/N55S/
Y51N、F56N/N55S/Y51Q、F56N/N55S/Y51S、F56N/N55S/Y51G、F56N/N55G/Y51L、F56N/N55G/
Y51V、F56N/N55G/Y51A、F56N/N55G/Y51N、F56N/N55G/Y51Q、F56N/N55G/Y51S、F56N/N55G/
Y51G、F56N/N55A/Y51L、F56N/N55A/Y51V、F56N/N55A/Y51A、F56N/N55A/Y51N、F56N/N55A/
Y51Q、F56N/N55A/Y51S、F56N/N55A/Y51G、F56N/N55T/Y51L、F56N/N55T/Y51V、F56N/N55T/
Y51A、F56N/N55T/Y51N、F56N/N55T/Y51Q、F56N/N55T/Y51 S、F56N/N55T/Y51G、F56Q/N55Q/
Y51L、F56Q/N55Q/Y51V、F56Q/N55Q/Y51A、F56Q/N55Q/Y51N、F56Q/N55Q/Y51Q、F56Q/N55Q/
Y51S、F56Q/N55Q/Y51G、F56Q/N55R/Y51L、F56Q/N55R/Y51V、F56Q/N55R/Y51A、F56Q/N55R/
Y51N、F56Q/N55R/Y51Q、F56Q/N55R/Y51S、F56Q/N55R/Y51G、F56Q/N55K/Y51L、F56Q/N55K/
Y51V、F56Q/N55K/Y51A、F56Q/N55K/Y51N、F56Q/N55K/Y51Q、F56Q/N55K/Y51S、F56Q/N55K/
Y51G、F56Q/N55S/Y51L、F56Q/N55S/Y51V、F56Q/N55S/Y51A、F56Q/N55S/Y51N、F56Q/N55S/
Y51Q、F56Q/N55S/Y51S、F56Q/N55S/Y51G、F56Q/N55G/Y51L、F56Q/N55G/Y51V、F56Q/N55G/
Y51A、F56Q/N55G/Y51N、F56Q/N55G/Y51Q、F56Q/N55G/Y51S、F56Q/N55G/Y51G、F56Q/N55A/
Y51L、F56Q/N55A/Y51V、F56Q/N55A/Y51A、F56Q/N55A/Y51N、F56Q/N55A/Y51Q、F56Q/N55A/
Y51S、F56Q/N55A/Y51G、F56Q/N55T/Y51L、F56Q/N55T/Y51V、F56Q/N55T/Y51A、F56Q/N55T/
Y51N、F56Q/N55T/Y51Q、F56Q/N55T/Y51 S、F56Q/N55T/Y51G、F56R/N55Q/Y51L、F56R/N55Q/
Y51V、F56R/N55Q/Y51A、F56R/N55Q/Y51N、F56R/N55Q/Y51Q、F56R/N55Q/Y51S、F56R/N55Q/
Y51G、F56R/N55R/Y51L、F56R/N55R/Y51V、F56R/N55R/Y51A、F56R/N55R/Y51N、F56R/N55R/
Y51Q、F56R/N55R/Y51S、F56R/N55R/Y51G、F56R/N55K/Y51L、F56R/N55K/Y51V、F56R/N55K/
Y51A、F56R/N55K/Y51N、F56R/N55K/Y51Q、F56R/N55K/Y51S、F56R/N55K/Y51G、F56R/N55S/
Y51L、F56R/N55S/Y51V、F56R/N55S/Y51A、F56R/N55S/Y51N、F56R/N55S/Y51Q、F56R/N55S/
Y51S、F56R/N55S/Y51G、F56R/N55G/Y51L、F56R/N55G/Y51V、F56R/N55G/Y51A、F56R/N55G/
Y51N、F56R/N55G/Y51Q、F56R/N55G/Y51 S、F56R/N55G/Y51G、F56R/N55A/Y51L、F56R/N55A/
Y51V、F56R/N55A/Y51A、F56R/N55A/Y51N、F56R/N55A/Y51Q、F56R/N55A/Y51S、F56R/N55A/
Y51G、F56R/N55T/Y51L、F56R/N55T/Y51V、F56R/N55T/Y51A、F56R/N55T/Y51N、F56R/N55T/
Y51Q、F56R/N55T/Y51S、F56R/N55T/Y51G、F56S/N55Q/Y51L、F56S/N55Q/Y51V、F56S/N55Q/
Y51A、F56S/N55Q/Y51N、F56S/N55Q/Y51Q、F56S/N55Q/Y51S、F56S/N55Q/Y51G、F56S/N55R/
Y51L、F56S/N55R/Y51V、F56S/N55R/Y51A、F56S/N55R/Y51N、F56S/N55R/Y51Q、F56S/N55R/
Y51S、F56S/N55R/Y51G、F56S/N55K/Y51L、F56S/N55K/Y51V、F56S/N55K/Y51A、F56S/N55K/
Y51N、F56S/N55K/Y51Q、F56S/N55K/Y51S、F56S/N55K/Y51G、F56S/N55S/Y51L、F56S/N55S/
Y51V、F56S/N55S/Y51A、F56S/N55S/Y51N、F56S/N55S/Y51Q、F56S/N55S/Y51S、F56S/N55S/
Y51G、F56S/N55G/Y51L、F56S/N55G/Y51V、F56S/N55G/Y51A、F56S/N55G/Y51N、F56S/N55G/
Y51Q、F56S/N55G/Y51S、F56S/N55G/Y51G、F56S/N55A/Y51L、F56S/N55A/Y51V、F56S/N55A/

Y51A、F56S/N55A/Y51N、F56S/N55A/Y51Q、F56S/N55A/Y51 S、F56S/N55A/Y51G、F56S/N55T/Y51L、F56S/N55T/Y51V、F56S/N55T/Y51A、F56S/N55T/Y51N、F56S/N55T/Y51Q、F56S/N55T/Y51S、F56S/N55T/Y51G、F56G/N55Q/Y51L、F56G/N55Q/Y51V、F56G/N55Q/Y51A、F56G/N55Q/Y51N、F56G/N55Q/Y51Q、F56G/N55Q/Y51S、F56G/N55Q/Y51G、F56G/N55R/Y51L、F56G/N55R/Y51V、F56G/N55R/Y51A、F56G/N55R/Y51N、F56G/N55R/Y51Q、F56G/N55R/Y51S、F56G/N55R/Y51G、F56G/N55K/Y51L、F56G/N55K/Y51V、F56G/N55K/Y51A、F56G/N55K/Y51N、F56G/N55K/Y51Q、F56G/N55K/Y51S、F56G/N55K/Y51G、F56G/N55S/Y51L、F56G/N55S/Y51V、F56G/N55S/Y51A、F56G/N55S/Y51N、F56G/N55S/Y51Q、F56G/N55S/Y51S、F56G/N55S/Y51G、F56G/N55G/Y51L、F56G/N55G/Y51V、F56G/N55G/Y51A、F56G/N55G/Y51N、F56G/N55G/Y51Q、F56G/N55G/Y51S、F56G/N55G/Y51G、F56G/N55A/Y51L、F56G/N55A/Y51V、F56G/N55A/Y51A、F56G/N55A/Y51N、F56G/N55A/Y51Q、F56G/N55A/Y51S、F56G/N55A/Y51G、F56G/N55T/Y51L、F56G/N55T/Y51V、F56G/N55T/Y51A、F56G/N55T/Y51N、F56G/N55T/Y51Q、F56G/N55T/Y51 S、F56G/N55T/Y51G、F56A/N55Q/Y51L、F56A/N55Q/Y51V、F56A/N55Q/Y51A、F56A/N55Q/Y51N、F56A/N55Q/Y51Q、F56A/N55Q/Y51S、F56A/N55Q/Y51G、F56A/N55R/Y51L、F56A/N55R/Y51V、F56A/N55R/Y51A、F56A/N55R/Y51N、F56A/N55R/Y51Q、F56A/N55R/Y51S、F56A/N55R/Y51G、F56A/N55K/Y51L、F56A/N55K/Y51V、F56A/N55K/Y51A、F56A/N55K/Y51N、F56A/N55K/Y51Q、F56A/N55K/Y51S、F56A/N55K/Y51G、F56A/N55S/Y51L、F56A/N55S/Y51V、F56A/N55S/Y51A、F56A/N55S/Y51N、F56A/N55S/Y51Q、F56A/N55S/Y51S、F56A/N55S/Y51G、F56A/N55G/Y51L、F56A/N55G/Y51V、F56A/N55G/Y51A、F56A/N55G/Y51N、F56A/N55G/Y51Q、F56A/N55G/Y51S、F56A/N55G/Y51G、F56A/N55A/Y51L、F56A/N55A/Y51V、F56A/N55A/Y51A、F56A/N55A/Y51N、F56A/N55A/Y51Q、F56A/N55A/Y51S、F56A/N55A/Y51G、F56A/N55T/Y51L、F56A/N55T/Y51V、F56A/N55T/Y51A、F56A/N55T/Y51N、F56A/N55T/Y51Q、F56A/N55T/Y51 S、F56A/N55T/Y51G、F56K/N55Q/Y51L、F56K/N55Q/Y51V、F56K/N55Q/Y51A、F56K/N55Q/Y51N、F56K/N55Q/Y51Q、F56K/N55Q/Y51S、F56K/N55Q/Y51G、F56K/N55R/Y51L、F56K/N55R/Y51V、F56K/N55R/Y51A、F56K/N55R/Y51N、F56K/N55R/Y51Q、F56K/N55R/Y51S、F56K/N55R/Y51G、F56K/N55K/Y51L、F56K/N55K/Y51V、F56K/N55K/Y51A、F56K/N55K/Y51N、F56K/N55K/Y51Q、F56K/N55K/Y51S、F56K/N55K/Y51G、F56K/N55S/Y51L、F56K/N55S/Y51V、F56K/N55S/Y51A、F56K/N55S/Y51N、F56K/N55S/Y51Q、F56K/N55S/Y51S、F56K/N55S/Y51G、F56K/N55G/Y51L、F56K/N55G/Y51V、F56K/N55G/Y51A、F56K/N55G/Y51N、F56K/N55G/Y51Q、F56K/N55G/Y51S、F56K/N55G/Y51G、F56K/N55A/Y51L、F56K/N55A/Y51V、F56K/N55A/Y51A、F56K/N55A/Y51N、F56K/N55A/Y51Q、F56K/N55A/Y51S、F56K/N55A/Y51G、F56K/N55T/Y51L、F56K/N55T/Y51V、F56K/N55T/Y51A、F56K/N55T/Y51N、F56K/N55T/Y51Q、F56K/N55T/Y51S、F56K/N55T/Y51G、F56E/N55R、F56E/N55K、F56D/N55R、F56D/N55K、F56R/N55E、F56R/N55D、F56K/N55E或F56K/N55D。

[0228] 在(ii)中,变异体优选地包含:Y51R/F56Q、Y51N/F56N、Y51M/F56Q、Y51L/F56Q、Y51I/F56Q、Y51V/F56Q、Y51A/F56Q、Y51P/F56Q、Y51G/F56Q、Y51C/F56Q、Y51Q/F56Q、Y51N/F56Q、Y51S/F56Q、Y51E/F56Q、Y51D/F56Q、Y51K/F56Q或Y51H/F56Q。

[0229] 在(ii)中,变异体优选地包含Y51T/F56Q、Y51Q/F56Q或Y51A/F56Q。

[0230] 在(ii)中,变异体优选地包含:Y51T/F56F、Y51T/F56M、Y51T/F56L、Y51T/F56I、Y51T/F56V、Y51T/F56A、Y51T/F56P、Y51T/F56G、Y51T/F56C、Y51T/F56Q、Y51T/F56N、Y51T/

F56T、Y51T/F56S、Y51T/F56E、Y51T/F56D、Y51T/F56K、Y51T/F56H或Y51T/F56R。

[0231] 在(ii)中,变异体优选地包含Y51T/N55Q、Y51T/N55S或Y51T/N55A。

[0232] 在(ii)中,变异体优选地包含:Y51A/F56F、Y51A/F56L、Y51A/F56I、Y51A/F56V、Y51A/F56A、Y51A/F56P、Y51A/F56G、Y51A/F56C、Y51A/F56Q、Y51A/F56N、Y51A/F56T、Y51A/F56S、Y51A/F56E、Y51A/F56D、Y51A/F56K、Y51A/F56H或Y51A/F56R。

[0233] 在(ii)中,变异体优选地包含:Y51C/F56A、Y51E/F56A、Y51D/F56A、Y51K/F56A、Y51H/F56A、Y51Q/F56A、Y51N/F56A、Y51S/F56A、Y51P/F56A或Y51V/F56A。

[0234] 在(xi)中,变异体优选地包含Y51/P52、Y51/P52/A53、P50到P52、P50到A53、K49到Y51、K49到A53的缺失,和K49到S54用单一脯氨酸(P)的置换;和Y51到A53、Y51到S54、N55/F56、N55到S57、N55/F56用单一P的置换;和N55/F56用单一P的置换;和N55/F56用单一甘氨酸(G)的置换;和N55/F56用单一丙氨酸(A)的置换;和用单一P的置换,和Y51N、N55/F56;和用单一P的置换,和Y51Q、N55/F56;和用单一P的置换,Y51S、N55/F56;和用单一G的置换,和Y51N、N55/F56;和用单一G的置换,和Y51Q、N55/F56;和用单一G的置换,和Y51S、N55/F56;和用单一A的置换,和Y51N、N55/F56;和用单一A/Y51Q或N55/F56的置换;和用单一A的置换,和Y51S。

[0235] 变异体更优选地包含:D195N/E203N、D195Q/E203N、D195N/E203Q、D195Q/E203Q、E201N/E203N、E201Q/E203N、E201N/E203Q、E201Q/E203Q、E185N/E203Q、E185Q/E203Q、E185N/E203N、E185Q/E203N、D195N/E201N/E203N、D195Q/E201N/E203N、D195N/E201Q/E203N、D195N/E201N/E203Q、D195Q/E201Q/E203N、D195Q/E201N/E203Q、D195N/E201Q/E203Q、D195Q/E201Q/E203Q、D149N/E201N、D149Q/E201N、D149N/E201Q、D149Q/E201Q、D149N/E201N/D195N、D149Q/E201N/D195N、D149N/E201Q/D195N、D149N/E201N/D195Q、D149Q/E201Q/D195N、D149Q/E201N/D195Q、D149N/E201Q/D195Q、D149Q/E201Q/D195Q、D149N/E203N、D149Q/E203N、D149N/E203Q、D149Q/E203Q、D149N/E185N/E201N、D149Q/E185N/E201N、D149N/E185Q/E201N、D149N/E185N/E201Q、D149Q/E185Q/E201N、D149Q/E185N/E201Q、D149N/E185Q/E201Q、D149Q/E185Q/E201Q、D149N/E185N/E203N、D149Q/E185N/E203N、D149N/E185Q/E203N、D149N/E185N/E203Q、D149Q/E185Q/E203Q、D149N/E185N/E201N/E203N、D149Q/E185N/E201N/E203N、D149N/E185Q/E201N/E203N、D149N/E185N/E201Q/E203N、D149Q/E185N/E201N/E203Q、D149Q/E185Q/E201N/E203N、D149Q/E185N/E201Q/E203N、D149Q/E185N/E201N/E203Q、D149N/E185Q/E201Q/E203N、D149N/E185Q/E201N/E203Q、D149Q/E185Q/E201N/E203Q、D149Q/E185Q/E201Q/E203N、D149N/E185N/D195N/E201N/E203N、D149Q/E185N/D195N/E201N/E203N、D149N/E185Q/D195N/E201N/E203N、D149Q/E185Q/D195N/E201N/E203N、D149N/E185N/D195Q/E201N/E203N、D149Q/E185N/D195Q/E201N/E203N、D149N/E185N/D195N/E201Q/E203N、D149Q/E185N/D195N/E201Q/E203N、D149Q/E185N/D195N/E201N/E203Q、D149Q/E185Q/D195N/E201N/E203N、D149Q/E185N/D195Q/E201N/E203N、D149N/E185Q/D195Q/E201N/E203N、D149N/E185Q/D195N/E201Q/E203N、D149N/E185Q/D195N/E201Q/E203Q、D149Q/E185Q/D195Q/E201N/

E203N、D149Q/E185Q/D195N/E201Q/E203N、D149Q/E185Q/D195N/E201N/E203Q、D149Q/E185N/D195Q/E201Q/E203N、D149Q/E185N/D195Q/E201N/E203Q、D149Q/E185N/D195N/E201Q/E203Q、D149N/E185Q/D195Q/E201Q/E203N、D149N/E185Q/D195Q/E201N/E203Q、D149N/E185Q/D195N/E201Q/E203Q、D149N/E185N/D195Q/E201Q/E203Q、D149Q/E185Q/D195Q/E201Q/E203N、D149Q/E185Q/D195Q/E201N/E203Q、D149Q/E185Q/D195N/E201Q/E203Q、D149Q/E185N/D195Q/E201Q/E203Q、D149N/E185Q/D195Q/E201Q/E203Q、D149Q/E185Q/D195N/E201Q/E203Q、D149Q/E185N/D195Q/E201Q/E203Q、D149N/E185R/E201N/E203N、D149Q/E185R/E201N/E203N、D149N/E185R/E201Q/E203N、D149N/E185R/E201N/E203Q、D149Q/E185R/E201Q/E203N、D149Q/E185R/E201N/E203Q、D149N/E185R/E201Q/E203Q、D149Q/E185R/E201Q/E203Q、D149R/E185N/E201N/E203N、D149R/E185Q/E201N/E203N、D149R/E185N/E201Q/E203N、D149R/E185N/E201N/E203Q、D149R/E185Q/E201Q/E203N、D149R/E185Q/E201N/E203Q、D149R/E185N/E201Q/E203Q、D149R/E185Q/E201Q/E203Q、D149R/E185N/D195N/E201N/E203N、D149R/E185Q/D195N/E201N/E203N、D149R/E185N/D195Q/E201N/E203N、D149R/E185N/D195N/E201Q/E203N、D149R/E185Q/D195N/E201N/E203Q、D149R/E185Q/D195Q/E201N/E203N、D149R/E185Q/D195N/E201N/E203Q、D149R/E185N/D195Q/E201Q/E203N、D149R/E185N/D195Q/E201N/E203Q、D149R/E185N/D195N/E201Q/E203Q、D149R/E185Q/D195Q/E201Q/E203N、D149R/E185Q/D195Q/E201N/E203Q、D149R/E185Q/D195N/E201N/E203Q、D149R/E185N/D195Q/E201Q/E203Q、D149R/E185Q/D195N/E201Q/E203Q、D149R/E185N/D195Q/E201Q/E203Q、D149R/E185Q/D195Q/E201Q/E203Q、D149N/E185R/D195N/E201N/E203N、D149Q/E185R/D195N/E201N/E203N、D149N/E185R/D195Q/E201N/E203N、D149N/E185R/D195N/E201Q/E203N、D149N/E185R/D195N/E201N/E203Q、D149Q/E185R/D195Q/E201N/E203N、D149Q/E185R/D195Q/E201N/E203Q、D149Q/E185R/D195N/E201Q/E203Q、D149Q/E185R/D195Q/E201Q/E203Q、D149Q/E185R/D195Q/E201Q/E203Q、D149N/E185R/D195N/E201R/E203N、D149Q/E185R/D195N/E201R/E203N、D149N/E185R/D195Q/E201R/E203N、D149N/E185R/D195N/E201R/E203Q、D149Q/E185R/D195Q/E201R/E203Q、D149Q/E185R/D195Q/E201R/E203Q、D149Q/E185R/D195Q/E201R/E203Q、D149Q/E185R/D195Q/E201R/E203Q、D149Q/E185R/D195Q/E201R/E203Q、E131D/K49R、E101N/N102F、E101N/N102Y、E101N/N102W、E101F/N102F、E101F/N102Y、E101F/N102W、E101Y/N102F、E101Y/N102Y、E101Y/N102W、E101W/N102F、E101W/N102Y、E101W/N102W、E101N/N102R、E101F/N102R、E101Y/N102R或E101W/N102F。

[0236] 本发明形成孔的优选变异体,其中较少核苷酸在聚核苷酸移动通过孔时引起电流,包含Y51A/F56A、Y51A/F56N、Y51I/F56A、Y51L/F56A、Y51T/F56A、Y51I/F56N、Y51L/F56N或Y51T/F56N,或更优选地Y51I/F56A、Y51L/F56A或Y51T/F56A。如上文所论述,这使得其更容易鉴定所观测的电流(在聚核苷酸移动通过孔时)与聚核苷酸之间的直接关系。

[0237] 形成显示范围增加的孔的优选变异体包含在以下位置处的突变:

[0238] Y51、F56、D149、E185、E201和E203;

[0239] N55和F56;

- [0240] Y51和F56;
- [0241] Y51、N55和F56;或
- [0242] F56和N102。
- [0243] 形成显示范围增加的孔的优选变异体包含:
- [0244] Y51N、F56A、D149N、E185R、E201N和E203N;
- [0245] N55S和F56Q;
- [0246] Y51A和F56A;
- [0247] Y51A和F56N;
- [0248] Y51I和F56A;
- [0249] Y51L和F56A;
- [0250] Y51T和F56A;
- [0251] Y51I和F56N;
- [0252] Y51L和F56N;
- [0253] Y51T和F56N;
- [0254] Y51T和F56Q;
- [0255] Y51A、N55S和F56A;
- [0256] Y51A、N55S和F56N;
- [0257] Y51T、N55S和F56Q;或
- [0258] F56Q和N102R。
- [0259] 形成孔的优选变异体,其中较少核苷酸在聚核苷酸移动通过孔时引起电流,包含在以下位置处的突变:
- [0260] N55和F56,如N55X和F56Q,其中X是任何氨基酸;或
- [0261] Y51和F56,如Y51X和F56Q,其中X是任何氨基酸。
- [0262] 特别优选的是,变异体包含Y51A和F56Q。
- [0263] 形成显示输送量增加的孔的优选变异体包含在以下位置处的突变:
- [0264] D149、E185和E203;
- [0265] D149、E185、E201和E203;或
- [0266] D149、E185、D195、E201和E203。
- [0267] 形成显示输送量增加的孔的优选变异体包含:
- [0268] D149N、E185N和E203N;
- [0269] D149N、E185N、E201N和E203N;
- [0270] D149N、E185R、D195N、E201N和E203N;或
- [0271] D149N、E185R、D195N、E201R和E203N。
- [0272] 形成其中捕获聚核苷酸增加的孔的优选变异体包含以下突变:
- [0273] D43N/Y51T/F56Q;
- [0274] E44N/Y51T/F56Q;
- [0275] D43N/E44N/Y51T/F56Q;
- [0276] Y51T/F56Q/Q62R;
- [0277] D43N/Y51T/F56Q/Q62R;

- [0278] E44N/Y51T/F56Q/Q62R;或
- [0279] D43N/E44N/Y51T/F56Q/Q62R。
- [0280] 优选变异体包含以下突变:
- [0281] D149R/E185R/E201R/E203R或Y51T/F56Q/D149R/E185R/E201R/E203R;
- [0282] D149N/E185N/E201N/E203N或Y51T/F56Q/D149N/E185N/E201N/E203N;
- [0283] E201R/E203R或Y51T/F56Q/E201R/E203R
- [0284] E201N/E203R或Y51T/F56Q/E201N/E203R;
- [0285] E203R或Y51T/F56Q/E203R;
- [0286] E203N或Y51T/F56Q/E203N;
- [0287] E201R或Y51T/F56Q/E201R;
- [0288] E201N或Y51T/F56Q/E201N;
- [0289] E185R或Y51T/F56Q/E185R;
- [0290] E185N或Y51T/F56Q/E185N;
- [0291] D149R或Y51T/F56Q/D149R;
- [0292] D149N或Y51T/F56Q/D149N;
- [0293] R142E或Y51T/F56Q/R142E;
- [0294] R142N或Y51T/F56Q/R142N;
- [0295] R192E或Y51T/F56Q/R192E;或
- [0296] R192N或Y51T/F56Q/R192N。
- [0297] 优选变异体包含以下突变:
- [0298] Y51A/F56Q/E101N/N102R;
- [0299] Y51A/F56Q/R97N/N102G;
- [0300] Y51A/F56Q/R97N/N102R;
- [0301] Y51A/F56Q/R97N;
- [0302] Y51A/F56Q/R97G;
- [0303] Y51A/F56Q/R97L;
- [0304] Y51A/F56Q/N102R;
- [0305] Y51A/F56Q/N102F;
- [0306] Y51A/F56Q/N102G;
- [0307] Y51A/F56Q/E101R;
- [0308] Y51A/F56Q/E101F;
- [0309] Y51A/F56Q/E101N;或
- [0310] Y51A/F56Q/E101G
- [0311] 优选地进一步包含在T150处的突变。形成显示插入增加的孔的优选变异体包含T150I。在T150处的突变(如T150I)可与上文所论述突变或其组合中的任一个进行组合。
- [0312] SEQ ID NO:2的优选变异体包含(a) R97W和(b) 在Y51和/或F56处的突变。SEQ ID NO:2的优选变异体包含(a) R97W和(b) Y51R/H/K/D/E/S/T/N/Q/C/G/P/A/V/I/L/M和/或F56 R/H/K/D/E/S/T/N/Q/C/G/P/A/V/I/L/M。SEQ ID NO:2的优选变异体包含(a) R97W和(b) Y51L/V/A/N/Q/S/G和/或F56A/Q/N。SEQ ID NO:2的优选变异体包含(a) R97W和(b) Y51A和/

或F56Q。SEQ ID NO:2的优选变异体包含R97W、Y51A和F56Q。

[0313] 在本发明的突变单体中,SEQ ID NO:2的变异体优选地包含在R192处的突变。变异体优选地包含R192D/Q/F/S/T/N/E、R192D/Q/F/S/T或R192D/Q。SEQ ID NO:2的优选变异体包含(a) R97W, (b) 在Y51和/或F56处的突变和(c) 在R192处的突变,如R192D/Q/F/S/T/N/E、R192D/Q/F/S/T或R192D/Q。SEQ ID NO:2的优选变异体包含(a) R97W, (b) Y51R/H/K/D/E/S/T/N/Q/C/G/P/A/V/I/L/M和/或F56 R/H/K/D/E/S/T/N/Q/C/G/P/A/V/I/L/M, 和(c) 在R192处的突变,如R192D/Q/F/S/T/N/E、R192D/Q/F/S/T或R192D/Q。SEQ ID NO:2的优选变异体包含(a) R97W, (b) Y51L/V/A/N/Q/S/G和/或F56A/Q/N, 和(c) 在R192处的突变,如R192D/Q/F/S/T/N/E、R192D/Q/F/S/T或R192D/Q。SEQ ID NO:2的优选变异体包含(a) R97W, (b) Y51A和/或F56Q, 和(c) 在R192处的突变,如R192 D/Q/F/S/T/N/E、R192D/Q/F/S/T或R192D/Q。SEQ ID NO:2的优选变异体包含R97W、Y51A、F56Q和R192D/Q/F/S/T或R192D/Q。SEQ ID NO:2的优选变异体包含R97W、Y51A、F56Q和R192D。SEQ ID NO:2的优选变异体包含R97W、Y51A、F56Q和R192Q。在于特定位置处的不同氨基酸通过/符号分隔的以上段落中,/符号意指“或”。举例来说,R192D/Q意指R192D或R192Q。

[0314] 上文所描述的任一个SEQ ID NO:2的优选变异体可进一步。在本发明的突变单体中,SEQ ID NO:2的变异体优选地包含在R93处的突变。SEQ ID NO:2的优选变异体包含(a) R93W和(b) 在Y51和/或F56处的突变,优选地Y51A和F56Q。D或R192N,V105、A106和I107的缺失。

[0315] 以上SEQ ID NO:2的优选变异体中的任一个可包含K94N/Q突变。以上SEQ ID NO:2的优选变异体中的任一个可包含F191T突变。本发明还提供一种突变CsgG单体,其包含SEQ ID NO:2中所示序列的变异体,所述变异体包含在实例中所公开的变异体中存在的突变组合。

[0316] 用于引入或取代天然存在的氨基酸的方法在所属领域中是众所周知的。举例来说,可通过在编码突变单体的聚核苷酸中的相关位置处用精氨酸的密码子(CGT) 置换甲硫氨酸的密码子(ATG),而用精氨酸(R) 将甲硫氨酸(M) 取代。接着,可如下文所述来表达聚核苷酸。

[0317] 用于引入或取代非天然存在的氨基酸的方法在所属领域中也是众所周知的。举例来说,可通过在用于表达突变单体的IVTT系统中包括合成氨基酰基-tRNA来引入非天然存在的氨基酸。替代地,可通过在特定氨基酸营养缺陷型大肠杆菌中在那些特定氨基酸的合成(即非天然存在的) 类似物存在下表达突变单体,来引入所述非天然存在的氨基酸。如果突变单体是使用部分肽合成来生产,那么其还可通过裸接合来生产。

[0318] 变异体

[0319] 除上文所论述的特定突变以外,变异体可包括其它突变。在SEQ ID NO:2的氨基酸序列的整个长度内,基于氨基酸同源性,变异体将优选地与所述序列至少50%同源。更优选地,基于氨基酸同源性,变异体可在整个序列上与SEQ ID NO:2的氨基酸序列至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,并且更优选地至少95%、97%或99%同源。在100个或更多的延伸段内,例如在125、150、175或200个或更多的邻接氨基酸内,可存在至少80%,例如至少85%、90%或95%氨基酸同源性(“硬同源性”)。

[0320] 所属领域中的标准方法可用于测定同源性。举例来说,UWGCG程序包提供BESTFIT

程序,其可用于计算同源性,例如用其默认设置(德弗罗等人(1984)《核酸研究(Nucleic Acids Research)》12,第387-395页)。PILEUP和BLAST算法可用于计算同源性对序列进行排序(如鉴定等同残基或对应序列(通常用其默认设置)),例如如Altschul S.F. (1993)《分子进化杂志(J Mol Evol)》36:290-300;Altschul,S.F等人(1990)《分子进化杂志》215:403-10中所描述。用于执行BLAST分析的软件可通过美国国家生物技术信息中心(National Center for Biotechnology Information)(<http://www.ncbi.nlm.nih.gov/>)公开获得。

[0321] SEQ ID NO:2是来自大肠杆菌菌株K-12亚株MC4100的野生型CsgG单体。SEQ ID NO:2的变异体可包含存在于另一CsgG同源物中的取代中的任一个。优选CsgG同源物在SEQ ID NO:3到7和26到41中显示。与SEQ ID NO:2相比,变异体可包含存在于SEQ ID NO:3到7和26到41中的取代中的一或多个的组合。举例来说,可在SEQ ID NO:2中的任何一或多个位置处进行突变,所述一或多个位置在SEQ ID NO:2与SEQ ID NO:3到7和SEQ ID NO:26到41中的任一个之间是不同的。此类突变可为SEQ ID NO:2中的氨基酸被SEQ ID NO:3到7和SEQ ID NO:26到41中的任一个的对应位置的氨基酸取代。替代地,在这些位置中的任一个处的突变可为用任何氨基酸的取代,或可为缺失或插入突变,如缺失或插入1到10个氨基酸,如缺失或插入2到8或3到6个氨基酸。除本文所公开的突变以外,在SEQ ID NO:2与所有SEQ ID NO:3到7和SEQ ID NO:26到41之间是保守的氨基酸优选地存在于本发明的变异体中。然而,可在任何一或多个位置处进行保守突变,这些一或多个位置在SEQ ID NO:2与所有SEQ ID NO:3到7和SEQ ID NO:26到41之间是保守的。

[0322] 本发明提供一种孔形成CsgG突变单体,其包含本文所描述的氨基酸中的任何一个或多个,所述氨基酸在CsgG单体结构中对应于SEQ ID NO:2中的特定位置的位置处被取代成SEQ ID NO:2的特定位置。可通过所属领域中的标准技术来测定对应位置。举例来说,上文提到的PILEUP和BLAST算法可用于比对CsgG单体的序列与SEQ ID NO:2,且因此鉴定对应残基。

[0323] 特定来说,本发明提供一种孔形成CsgG突变单体,其包含以下中的任何一个或多个:

- [0324] -在对应于SEQ ID NO:2中的R97的位置处的W;
- [0325] -在对应于SEQ ID NO:2中的R93的位置处的W;
- [0326] -在对应于SEQ ID NO:2中的R97的位置处的Y;
- [0327] -在对应于SEQ ID NO:2中的R93的位置处的Y;
- [0328] -在对应于SEQ ID NO:2中的R93和R97的各位置处的Y;
- [0329] -在对应于SEQ ID NO:2中的R192的位置处的D;
- [0330] -在对应于SEQ ID NO:2中的V105-I107的位置处的残基的缺失;
- [0331] -在对应于SEQ ID NO:2中的F193到L199的一或多个位置处的残基的缺失;
- [0332] -在对应于SEQ ID NO:2中的F195到L199的位置处的残基的缺失;
- [0333] -在对应于SEQ ID NO:2中的F193到L199的位置处的残基的缺失;
- [0334] -在对应于SEQ ID NO:2中的F191的位置处的T;
- [0335] -在对应于SEQ ID NO:2中的K49的位置处的Q;
- [0336] -在对应于SEQ ID NO:2中的K49的位置处的N;
- [0337] -在对应于SEQ ID NO:2中的K42的位置处的Q;

- [0338] -在对应于SEQ ID NO:2中的E44的位置处的Q;
- [0339] -在对应于SEQ ID NO:2中的E44的位置处的N;
- [0340] -在对应于SEQ ID NO:2中的L90的位置处的R;
- [0341] -在对应于SEQ ID NO:2中的L91的位置处的R;
- [0342] -在对应于SEQ ID NO:2中的I95的位置处的R;
- [0343] -在对应于SEQ ID NO:2中的A99的位置处的R;
- [0344] -在对应于SEQ ID NO:2中的E101的位置处的H;
- [0345] -在对应于SEQ ID NO:2中的E101的位置处的K;
- [0346] -在对应于SEQ ID NO:2中的E101的位置处的N;
- [0347] -在对应于SEQ ID NO:2中的E101的位置处的Q;
- [0348] -在对应于SEQ ID NO:2中的E101的位置处的T;
- [0349] -在对应于SEQ ID NO:2中的Q114的位置处的K。

[0350] 本发明的CsgG孔形成单体优选地进一步包含在对应于SEQ ID NO:2中的Y51的位置处的A和/或在对应于SEQ ID NO:2中的F56的位置处的Q。

[0351] 孔形成突变单体通常保持形成与野生型CsgG单体相同的3D结构的能力,如与具有SEQ ID NO:2的序列的CsgG单体相同的3D结构。CsgG的3D结构在本领域中已知且公开于例如Cao等人(2014)《美国国家科学院院刊(PNAS)》E5439-E5444中。可在野生型CsgG序列中进行任何数量的除本文所描述突变以外的突变,其限制条件为,CsgG突变单体保持通过本发明的突变赋予其的改进的特性。

[0352] 通常,CsgG单体将保持形成包含三个 α -螺旋和五个 β -折叠的结构的能力。特定来说,本发明人已显示,可至少在CsgG的以下区中进行不影响CsgG单体形成跨膜孔的能力的突变:在第一 α -螺旋(其开始于SEQ ID NO:2中的S63)的N端、在第二 α -螺旋(从SEQ ID NO:2的G85到A99)中、在第二 α -螺旋与第一 β -折叠之间的环(从SEQ ID NO:2的Q100到N120)中、在第四和第五 β -折叠(分别地,SEQ ID NO:2的S173到R192和R198到T107)中和在第四与第五 β -折叠之间的环(SEQ ID NO:2的F193到Q197)中,所述跨膜孔能够使多肽移位。因此,可以设想,可在任何CsgG单体中的这些区中的任一个中进行不影响单体形成可使聚核苷酸移位的孔的能力的其它突变。还预期,可在其它区中进行不影响单体形成可使聚核苷酸移位的孔的能力的突变,所述其它区如:在 α -螺旋中的任一个(SEQ ID NO:2的S63到R76、G85到A99或V211到L236)中;或在 β -折叠中的任一个(SEQ ID NO:2的I121到N133、K135到R142、I146到R162、S173到R192或R198到T107)中。还预期,可在以下中的任一个中进行不影响单体形成可使聚核苷酸移位的孔的能力的一或多个氨基酸的缺失:连接 α -螺旋与 β -折叠的环区,和/或在CsgG单体的N端和/或C端区中。

[0353] 可对SEQ ID NO:2的氨基酸序列进行除上文所论述的那些以外的氨基酸取代,例如至多1、2、3、4、5、10、20或30个取代。保守取代用具有类似化学结构、类似化学特性或类似侧链体积的其它氨基酸来置换氨基酸。所引入的氨基酸可具有与其置换的氨基酸类似的极性、亲水性、疏水性、碱性、酸性、电中性或电荷。替代地,保守取代可引入代替预先存在的芳族或脂族氨基酸的另一芳族或脂族氨基酸。保守氨基酸改变在所属领域中是众所周知的,且可根据如在下文表2中限定的20中主要氨基酸的特性来选择。当氨基酸具有类似极性时,还可参考表3中的氨基酸侧链的亲水性值来决定这个选择。

[0354] 表2-氨基酸的化学特性

[0355]	Ala	脂族、疏水性的、中性	Met	疏水性的、中性
	Cys	极性、疏水性的、中性	Asn	极性、亲水性的、中性
	Asp	极性、亲水性的、带电 (-)	Pro	疏水性的、中性
	Glu	极性、亲水性的、带电 (-)	Gln	极性、亲水性的、中性
	Phe	芳族、疏水性的、中性	Arg	极性、亲水性的、带电 (+)
[0356]	Gly	脂族、中性	Ser	极性、亲水性的、中性
	His	芳族、极性、亲水性的、带电 (+)	Thr	极性、亲水性的、中性
	Ile	脂族、疏水性的、中性	Val	脂族、疏水性的、中性
	Lys	极性、亲水性的、带电 (+)	Trp	芳族、疏水性的、中性
	Leu	脂族、疏水性的、中性	Tyr	芳族、极性、疏水性的

[0357] 表3-亲水性值

[0358]	侧链	亲水性
	Ile	4.5
	Val	4.2
	Leu	3.8
	Phe	2.8
	Cys	2.5
	Met	1.9
	Ala	1.8
	Gly	-0.4
	Thr	-0.7
	Ser	-0.8
	Trp	-0.9
	Tyr	-1.3
	Pro	-1.6
	His	-3.2
	Glu	-3.5
	Gln	-3.5
	Asp	-3.5
	Asn	-3.5
	Lys	-3.9
	Arg	-4.5

[0359] 可从上文所描述的多肽另外缺失SEQ ID NO:2的氨基酸序列的一或多个氨基酸残基。可缺失至多1、2、3、4、5、10、20或30个或更多个残基。

[0360] 变异体可包括SEQ ID NO:2的片段。此类片段保持孔形成活性。片段的长度可为至少50、至少100、至少150、至少200或至少250个氨基酸。此类片段可用于产生孔。片段优选地包含SEQ ID NO:2的跨膜域,即K135-Q153和S183-S208。

[0361] 替代地或另外,可向上文所描述的多肽中添加一或多个氨基酸。可在SEQ ID NO:2

或其变异体或片段的氨基酸序列的氨基端或羧基端处提供延长物。延长物的长度可以非常短,例如1到10个氨基酸。替代地,延长物可以较长,例如至多50或100个氨基酸。可将载体蛋白与根据本发明的氨基酸序列融合。其它融合蛋白在下文更详细地论述。

[0362] 如上文所论述,变异体是具有不同于SEQ ID NO:2的氨基酸序列且保持其形成孔的能力的多肽。变异体通常含有SEQ ID NO:2的引起孔形成的区。CsgG的孔形成能力通过各亚基中的 β -折叠提供,所述CsgG含有 β -折叠桶。SEQ ID NO:2的变异体通常包含SEQ ID NO:2中形成 β -折叠的区,即K135-Q153和S183-S208。可对SEQ ID NO:2的形成 β -折叠的区进行一或多个修饰,只要所得变异体保持其形成孔的能力即可。SEQ ID NO:2的变异体优选地在其-螺旋和/或环区内包括一或多个修饰,如取代、添加或缺失。

[0363] 衍生自CsgG的单体可经修饰以辅助其鉴定或纯化,例如通过添加抗生蛋白链菌素标签或通过添加信号序列以促进其从其中单体并不天然地含有此类序列的细胞分泌。其它合适的标签在下文更详细地论述。单体可用显露标记来进行标记。显露标记可以是使单体被检测到的任何合适标记。合适标记在下文描述。

[0364] 衍生自CsgG的单体还可使用D-氨基酸来产生。举例来说,衍生自CsgG的单体可包含L-氨基酸与D-氨基酸的混合物。这在用于产生此类蛋白质或肽的所属领域中是常规的。

[0365] 衍生自CsgG的单体含有一或多个特异性修饰以便于核苷酸鉴别。衍生自CsgG的单体还可含有其它非特异性修饰,只要其不干扰孔形成即可。本领域中已知多种非特异性侧链修饰,且可对衍生自CsgG的单体的侧链进行所述多种非特异性侧链修饰。此类修饰包括例如,通过与醛反应,随后用 NaBH_4 、用甲基乙酰亚氨酸脒化或用乙酸酐酰化来对氨基酸还原烷化。

[0366] 可使用本领域中已知的标准方法来产生衍生自CsgG的单体。可以合成方式或通过重组手段来制备衍生自CsgG的单体。举例来说,可通过活体外翻译和转录 (IVTT) 来合成单体。用于产生孔和单体的合适方法在国际申请第PCT/GB09/001690号 (公开为W02010/004273)、第PCT/GB09/001679号 (公开为W0 2010/004265) 或第PCT/GB10/000133号 (公开为W0 2010/086603) 中论述。论述用于将孔插入到膜中的方法。

[0367] 在一些实施例中,突变单体进行化学修饰。突变单体可以任何方式和在任何位点处进行化学修饰。突变单体优选地通过将一分子与一或多个半胱氨酸连接 (半胱氨酸连接)、将一分子与一或多个赖氨酸连接、将一分子与一或多个非天然氨基酸连接、表位的酶修饰或末端的修饰来进行化学修饰。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。突变单体可通过连接任何分子来进行化学修饰。举例来说,突变单体可通过连接染料或荧光团来进行化学修饰。

[0368] 在一些实施例中,突变单体用有助于包含单体的孔与靶核苷酸或靶聚核苷酸序列之间的相互作用的分子衔接子来进行化学修饰。衔接子的存在改进孔和核苷酸或聚核苷酸序列的宿主-客体化学,且从而改进由突变单体形成的孔的测序能力。宿主-客体化学的原理在所属领域中是众所周知的。衔接子对于孔的物理或化学特性具有效果,其改进其与核苷酸或聚核苷酸序列的相互作用。衔接子可改变孔的折叠桶或通道的电荷,或特异性地与核苷酸或聚核苷酸序列相互作用或结合于所述核苷酸或聚核苷酸序列,从而促进其与孔相互作用。

[0369] 分子衔接子优选地是环状分子、环糊精、能够杂交的物种、DNA结合剂或嵌入剂、肽

或肽类似物、合成聚合物、芳族平面分子、能够结合氢的带正电小分子或小分子。

[0370] 衔接子可为环状的。环状衔接子优选地具有与孔相同的对称性。衔接子优选地具有八重或九重对称性,这是由于CsgG通常围绕中心轴具有八或九个亚基。这一点在下文更详细论述。

[0371] 衔接子通常通过宿主-客体化学而与核苷酸或聚核苷酸序列相互作用。衔接子通常能够与核苷酸或聚核苷酸序列相互作用。衔接子包含能够与核苷酸或聚核苷酸序列相互作用的一或多个化学基团。所述一或多个化学基团优选地通过非共价相互作用而与核苷酸或聚核苷酸序列相互作用,如疏水性相互作用、氢结合、范德华力(Van der Waal's forces)、 π -阳离子相互作用和/或静电力。能够与核苷酸或聚核苷酸序列相互作用的一或多个化学基团优选地带正电。能够与核苷酸或聚核苷酸序列相互作用的一或多个化学基团优选地包含氨基。氨基可连接于伯、仲或叔碳原子。衔接子甚至更优选地包含氨基环,如6、7或8个氨基的环。衔接子最优选地包含八氨基环。质子化氨基环可与核苷酸或聚核苷酸序列中的带负电磷酸基相互作用。

[0372] 孔内的衔接子的正确定位可通过衔接子与包含突变单体的孔之间的宿主-客体化学来进行。衔接子优选地包含能够与孔中的一或多个氨基酸相互作用的一或多个化学基团。衔接子更优选地包含通过非共价相互作用而能够与孔中的一或多个氨基酸相互作用的一或多个化学基团,所述非共价相互作用如疏水性相互作用、氢结合、范德华力、 π -阳离子相互作用和/或静电力。能够与孔中的一或多个氨基酸相互作用的化学基团通常是羟基或胺。羟基可连接于伯、仲或叔碳原子。羟基可与孔中的不带电氨基酸形成氢键。可使用有助于孔与核苷酸或聚核苷酸序列之间的相互作用的任何衔接子。

[0373] 合适衔接子包括(但不限于)环糊精、环肽和葫芦脲。衔接子优选地是环糊精或其衍生物。环糊精或其衍生物可以是在Eliseev, A.V.和Schneider, H-J. (1994)《美国化学学会杂志(J. Am. Chem. Soc.)》116, 6081-6088中所公开那些化合物中的任一个。衔接子更优选地是七-6-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-胍基)-环糊精($\text{gu}_7\text{-}\beta\text{CD}$)。 $\text{gu}_7\text{-}\beta\text{CD}$ 中的胍基具有比 $\text{am}_7\text{-}\beta\text{CD}$ 中的伯胺高许多的pKa,且因此其带更多的正电。这种 $\text{gu}_7\text{-}\beta\text{CD}$ 衔接子可用于增加孔中的核苷酸的停留时间,增加所测量的剩余电流的准确度、以及增加在高温或低数据获取速率下的碱基检测速率。

[0374] 如果如在下文更详细地论述来使用3-(2-吡啶二硫代)丙酸丁二酰亚胺酯(SPDP)交联剂,那么衔接子优选地是七(6-脱氧-6-胺基)-6-N-单(2-吡啶基)二硫代丙酰基- β -环糊精($\text{am}_6\text{amPDP}_1\text{-}\beta\text{CD}$)。

[0375] 更多合适的衔接子包括 γ -环糊精,其包含9个糖单元且因此具有九重对称性)。 γ -环糊精可含有连接子分子,或可进行修饰以包含所有或更多个在上文所论述的 β -环糊精实例中使用的经修饰糖单元。

[0376] 分子衔接子优选地共价连接于突变单体。可使用本领域中已知的任何方法,将衔接子共价连接于孔。通常通过化学连接来连接衔接子。如果通过半胱氨酸连接来连接分子衔接子,那么通过取代优选地将一或多个半胱氨酸引入到突变中,例如折叠桶中。可通过将分子衔接子连接到突变单体中的一或多个半胱氨酸来对突变单体进行化学修饰。所述一或多个半胱氨酸可为天然存在的,即在SEQ ID NO:2中的位置1和/或215处。替代地,可通过将分子衔接子连接到在其它位置处所引入的一或多个半胱氨酸来对突变单体进行化学修饰。

在位置215处的半胱氨酸可例如通过取代来移除,以确保分子衔接子并不连接到所述位置,而不是在位置1处的半胱氨酸或在另一位置处引入的半胱氨酸。

[0377] 可通过修饰相邻残基来增强半胱氨酸残基的反应性。举例来说,侧接精氨酸、组氨酸或赖氨酸残基的碱性基团将改变半胱氨酸巯基的pKa为更具反应性的S⁻基的pKa。半胱氨酸残基的反应性可通过巯基保护基(如dTNB)来保护。在连接连接子之前,这些可与突变单体的一或多个半胱氨酸残基反应。可将所述分子直接连接于突变单体。使用连接子,如化学交联剂或肽连接子,所述分子优选地连接于突变单体。

[0378] 合适化学交联剂在所属领域中是众所周知的。优选交联剂包括3-(吡啶-2-基二磺酰基)丙酸2,5-二氧化吡咯烷-1-基酯、4-(吡啶-2-基二磺酰基)丁酸2,5-二氧化吡咯烷-1-基酯和8-(吡啶-2-基二磺酰基)辛酸2,5-二氧化吡咯烷-1-基酯。最优选交联剂是3-(2-吡啶二硫代)丙酸丁二酰亚胺酯(SPDP)。通常,在分子/交联剂复合体共价连接于突变单体之前,分子共价连接于双功能交联剂,但也有可能在双功能交联剂/单体复合体连接于分子之前,双功能交联剂与共价连接于单体。

[0379] 连接子优选地对二硫苏糖醇(DTT)具有抗性。合适连接子包括(但不限于)基于碘乙酰胺和基于马来酰亚胺的连接子。

[0380] 在其它实施例中,单体可连接于聚核苷酸结合蛋白。这形成可在本发明的测序方法中使用的模块测序系统。下文论述聚核苷酸结合蛋白。

[0381] 聚核苷酸结合蛋白优选地共价连接于突变单体。可使用本领域中已知的任何方法,将蛋白质共价连接于单体。可对单体和蛋白质进行化学融合或基因融合。如果从单一聚核苷酸序列表达完全构筑体,那么单体和蛋白质是基因融合的。将单体基因融合于聚核苷酸结合蛋白论述于国际申请第PCT/GB09/001679号(公开为W0 2010/004265)中。

[0382] 如果通过半胱氨酸连接来连接聚核苷酸结合蛋白,那么通过取代优选地将一或多个半胱氨酸引入到突变中。将一或多个半胱氨酸优选地引入到在同源物中具有低保守性、指示可容许突变或插入的环区中。因此,其适合于连接聚核苷酸结合蛋白。在此类实施例中,可移除在位置251处的天然存在的半胱氨酸。可通过如上文所描述的修饰来增强半胱氨酸残基的反应性。

[0383] 可将聚核苷酸结合蛋白直接连接于突变单体或通过一或多个连接子连接于突变单体。可使用在国际申请第PCT/GB10/000132号(公开为W0 2010/086602)中描述的杂交连接子,来将分子连接于突变单体。替代地,可使用肽连接子。肽连接子是氨基酸序列。肽连接子的长度、柔性和亲水性通常被设计为使得其不干扰单体和分子的功能。优选柔性肽连接子是2到20,如4、6、8、10或16个丝氨酸和/或甘氨酸的延伸段。更优选的柔性连接子包括(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。优选刚性连接子是2到30,如4、6、8、16或24个脯氨酸的延伸段。更优选的刚性连接子包括(P)₁₂,其中P是脯氨酸。

[0384] 突变单体可用分子衔接子和聚核苷酸结合蛋白来进行化学修饰。

[0385] 可将分子(用所述分子来对单体进行化学修饰)直接连接于单体或通过如以下国际申请中所公开的连接子来连接:第PCT/GB09/001690号(公开为W0 2010/004273)、第PCT/GB09/001679号(公开为W0 2010/004265)或第PCT/GB10/000133号(公开为W02010/086603)。

[0386] 可例如通过添加组氨酸残基(his标签)、天冬氨酸残基(asp标签)、抗生蛋白链菌

素标签、flag标签、SUMO标签、GST标签或MBP标签,或通过添加促进其从其中多肽并不天然地含有此类序列的细胞分泌的信号序列,来对本文所描述的任一种蛋白质(如本发明的突变单体和孔进行修饰)以辅助其鉴定或纯化。引入基因标签的替代性方案是将标签化学反应到蛋白质上的原生或经工程改造的位置上。此的实例将为将凝胶移动试剂与蛋白质外上经工程改造的半胱氨酸反应。这已证实为一种用于分离溶血素异源寡聚物的方法(《生物化学(Chem Biol)》.1997年7月;4(7):497-505)。

[0387] 本文所描述的任一个蛋白质,如本发明的突变单体和孔,可用显露标记来标记。显露标记可以是使蛋白质被检测到的任何合适标记。合适标记包括(但不限于)荧光分子;放射性同位素,例如¹²⁵I、³⁵S;酶;抗体;抗原;聚核苷酸;和配体,如生物素。

[0388] 本文所描述的任一个蛋白质,如本发明的单体或孔,可以合成方式或通过重组手段来制备。举例来说,可通过活体外翻译和转录(IVTT)来合成蛋白质。蛋白质的氨基酸序列可进行修饰以包括非天然存在的氨基酸或增加蛋白质的稳定性。当通过合成手段来生产蛋白质时,可在生产期间引入此类氨基酸。还可根据合成或重组生产来改变蛋白质。

[0389] 还可使用D-氨基酸来生产蛋白质。举例来说,蛋白质可包含L-氨基酸与D-氨基酸的混合物。这在用于产生此类蛋白质或肽的所属领域中是常规的。

[0390] 蛋白质还可含有其它非特异性修饰,只要其不干扰蛋白质的功能即可。在本领域中已知多种非特异性侧链修饰,且可对蛋白质的侧链进行所述多种非特异性侧链修饰。此类修饰包括例如,通过与醛反应,随后用NaBH₄、用甲基乙酰亚氨酸脒化或用乙酸酐酰化来对氨基酸还原烷化。

[0391] 可使用本领域中已知的标准方法来生产本文所描述的任一个蛋白质,包括本发明的单体和孔。可使用所属领域中的标准方法来推导编码蛋白质的聚核苷酸序列且进行复制。可使用所属领域中的标准技术在细菌宿主细胞中表达编码蛋白质的聚核苷酸序列。可通过从重组表达载体原位表达多肽来在细胞中产生蛋白质。表达载体任选地携带诱导性启动子以控制多肽的表达。这些方法描述于Sambrook, J. 和Russell, D. (2001).《分子克隆实验指南(Molecular Cloning: A Laboratory Manual)》第3版. 纽约冷泉港的冷泉港实验室出版社(Cold Spring Harbor Laboratory Press)中。

[0392] 可在从蛋白质产生生物体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,大规模生产蛋白质。典型的蛋白质液相色谱系统包括FPLC、AKTA系统、Bio-Cad系统、Bio-Rad生物系统和Gilson HPLC系统。

[0393] 构筑体

[0394] 本发明还提供一种构筑体,其包含两个或更多个共价连接的CsgG单体,其中所述单体中的至少一个是本发明的突变单体。本发明的构筑体保持其形成孔的能力。可如上文所论述来测定此。本发明的一或多个构筑体可用于形成用于表征(如测序)聚核苷酸的孔。构筑体可包含至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9或至少10个单体。构筑体优选地包含两个单体。两个或更多个单体可相同或不同。

[0395] 构筑体中的至少一个单体是本发明的突变单体。构筑体中的2个或更多个、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、或10个或更多个单体可为本发明的突变单体。构筑体中的所有单体优选地是本发明的突变单体。突变单体可相同或不同。在一优选实施例中,构筑体包含两个本发明的突变单

体。

[0396] 构筑体中的本发明的突变单体的长度优选地大致相同,或相同。构筑体中的本发明的突变的折叠桶的长度优选地大致相同,或相同。长度可以氨基酸数量和/或长度单位的形式测量。

[0397] 构筑体可包含不是本发明的突变单体的一或多个单体。不是本发明的突变单体的CsgG突变单体包括:单体,其包含SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41;或以下的比较变异体:SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41,其中以上所论述的氨基酸/位置无一个突变。构筑体中的至少一个单体可包含:SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41,或以下中所示序列的比较变异体:SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41。基于氨基酸同源性,SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41的比较变异体,在其整个序列内,与SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41至少50%同源。更优选地,基于氨基酸同源性,比较变异体可在整个序列上与SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41的氨基酸序列,至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,并且更优选地至少95%、97%或99%同源。

[0398] 优选地将构筑体中的单体进行基因融合。如果从单一聚核苷酸序列表达完全构筑体,那么单体是基因融合的。可以任何方式来组合单体中的编码序列,以形成编码构筑体的单一聚核苷酸序列。

[0399] 可以任何构形来将单体基因融合。可通过单体的末端氨基酸来融合单体。举例来说,可将一个单体所述氨基端与另一单体的羧基端融合。构筑体中的第二和随后单体(在氨基到羧基方向上)可在其氨基端端包含甲硫氨酸(其中的每一个与前一个单体的羧基端融合)。举例来说,如果M是单体(不具有氨基端甲硫氨酸),且mM是具有氨基端甲硫氨酸的单体,那么构筑体可包含序列M-mM、M-mM-mM或M-mM-mM-mM。这些蛋氨酸的存在通常由在于编码整个构筑体的聚核苷酸内的编码第二或随后单体的聚核苷酸的5'端处的起始密码子(即ATG)的表达造成。构筑体中的第一单体(在氨基到羧基方向上)还可包含甲硫氨酸(例如mM-mM、mM-mM-mM或mM-mM-mM-mM)。

[0400] 可将两个或更多个单体直接基因融合在一起。优选地使用连接子来将单体进行基因融合。连接子可设计成限制单体的移动性。优选连接子是氨基酸序列(即肽连接子)。可使用上文所论述的任一个肽连接子。

[0401] 在另一优选实施例中,将单体化学融合。如果例如通过化学交联剂将两部分化学连接,那么两个单体是化学融合的。可使用上文所论述的任一个化学交联剂。可将连接子连接于引入到本发明的突变单体中一或多个半胱氨酸残基。替代地,可将连接子连接于构筑体中的一个单体的末端。

[0402] 如果构筑体含有不同单体,那么可通过保持连接子的浓度大量过度于单体来预防单体自身的交联。替代地,可在使用两个连接子的情况中使用“锁和钥”排列。各连接子仅一端可反应在一起,以形成较长连接子,且连接子的另一端各自与不同单体反应。此类连接子描述于国际申请第PCT/GB10/000132号(公开为W0 2010/086602)中。

[0403] 聚核苷酸

[0404] 本发明还提供聚核苷酸序列,其编码本发明的突变单体。突变单体可以是上文所论述的那些突变单体中的任一个。基于核苷酸一致性,聚核苷酸序列优选地包含在整个序列上与SEQ ID NO:1的序列至少50%、60%、70%、80%、90%或95%同源的序列。在300个或更多个,例如375、450、525或600或更多个连续核苷酸的延伸段上,可存在至少80%,例如至少85%、90%或95%核苷酸一致性(“硬同源性”)。可如上文所描述来计算同源性。基于基因密码的简并,聚核苷酸序列可包含不同于SEQ ID NO:1的序列。

[0405] 本发明还提供聚核苷酸序列,其编码本发明的基因融合构筑体中的任一个。聚核苷酸优选地包含SEQ ID NO:1中所示序列的两个或更多个变异体。基于核苷酸一致性,聚核苷酸序列优选地包含在整个序列上与SEQ ID NO:1具有至少50%、60%、70%、80%、90%或95%同源性的两个或更多个序列。在600个或更多个,例如750、900、1050或1200或更多个连续核苷酸的延伸段上,可存在至少80%,例如至少85%、90%或95%核苷酸一致性(“硬同源性”)。可如上文所描述来计算同源性。

[0406] 可使用所属领域中的标准方法来推导聚核苷酸序列且进行复制。可从生产孔的生物体(如大肠杆菌)提取编码野生型CsgG的染色体DNA。可使用涉及特异性引物的PCR来扩增编码孔亚基的基因。可接着对所扩增的序列进行定点诱变。定点诱变的合适方法在本领域中已知,且包括例如组合链反应。编码本发明的构筑体的聚核苷酸可使用熟知技术来制备,如在Sambrook, J. 和Russell, D. (2001).《分子克隆实验指南》第3版,纽约冷泉港的冷泉港实验室出版社中所描述的那些技术。

[0407] 可接着将所得聚核苷酸序列并入到重组可复制的载体(如克隆载体)中。载体可用于在相容的宿主细胞中复制聚核苷酸。因此,可通过将聚核苷酸引入到可复制的载体中,将所述载体引入到相容的宿主细胞中,和在产生载体复制的条件下使宿主细胞生长,来制备聚核苷酸序列。可从宿主细胞回收载体。用于克隆聚核苷酸的合适宿主细胞在本领域中已知,且在下文更详细地描述。

[0408] 可将聚核苷酸序列克隆到合适表达载体中。在表达载体中,聚核苷酸序列通常可操作地连接于控制序列,所述控制序列能够通过宿主细胞实现编码序列的表达。此类表现载体可用于表达孔亚基。

[0409] 术语“可操作地连接”是指所描述的组分处于允许其以其预期方式起作用的关系的并接。“可操作地连接”于编码序列的控制序列是以在与控制序列相容的条件下实现编码序列的表达的方式接合。可将相同或不同聚核苷酸序列的多个拷贝引入到载体中。

[0410] 可接着将表达载体引入到合适宿主细胞中。因此,可通过将聚核苷酸序列插入到表达载体中、将载体引入到相容的细菌宿主细胞中和在产生聚核苷酸序列的表达的条件下使宿主细胞生长,来生产本发明的突变单体或构筑体。可将以重组方式表达的单体或构筑体自组装到宿主细胞膜中的孔中。替代地,以此方式生产的重组孔可从宿主细胞移除且插入到另一膜中。当生产包含至少两个不同单体或构筑体的孔时,不同单体或构筑体可分开地在如上文所描述的不同宿主细胞中表达,从宿主细胞移除且组装到独立膜(如兔细胞膜或合成膜)中的孔中。

[0411] 载体可例如为质粒、病毒或噬菌体载体,其具备复制起点、任选地用于表达所述聚核苷酸序列的启动子和任选地启动子的调控子。载体可含有一或多个可选标记基因,例如

四环素抗性基因。可选择启动子和其它表达调控信号,以与被设计成表达载体的宿主细胞相容。通常使用T7、trc、lac、ara或 λ 启动子。

[0412] 宿主细胞通常以高水平表达单体或构筑体。将选择用聚核苷酸序列转化的宿主细胞,以与用于转化细胞的表达载体相容。宿主细胞通常是细菌并优选地大肠杆菌。具有 λ DE3溶源,例如C41 (DE3)、BL21 (DE3)、JM109 (DE3)、B834 (DE3)、TUNER、Origami和Origami B,的任何细胞可表达包含T7启动子的载体。除上文所列的条件以外,列举于以下中的任一个方法可用于表达CsgG蛋白:Cao等人,2014,《美国国家科学院院刊》,《九聚体细菌淀粉状蛋白分泌通道的结构(Structure of the nonameric bacterial amyloid secretion channel)》,doi-1411942111;和Goyal等人,2014,《自然》,516,250-253《对细菌淀粉状蛋白分泌通道CsgG的结构性和机制见解(structural and mechanistic insights into the bacterial amyloid secretion channel CsgG)》。

[0413] 本发明还包含一种生产本发明的突变单体或本发明的构筑体的方法。方法包含在合适宿主细胞中表达本发明的聚核苷酸。聚核苷酸优选地是载体的一部分,且优选地可操作地连接于启动子。

[0414] 孔

[0415] 本发明还提供各种孔。本发明的孔对于表征(如测序)聚核苷酸序列是理想的,这是因为其可以较高度度的敏感性在不同核苷酸之间进行鉴别。孔可出人意料地区别DNA和RNA中的四种核苷酸。本发明的孔可甚至区别甲基化和未甲基化的核苷酸。本发明的孔的基础分辨率出人意料地高。孔显示几乎完全分离所有四种DNA核苷酸。基于孔中的停留时间和流过孔的电流,孔进一步在脱氧胞苷单磷酸酯(dCMP)与甲基-dCMP之间进行鉴别。

[0416] 本发明的孔还可在一系列条件下在不同核苷酸之间进行鉴别。特定来说,孔将在有利于表征(如测序)核酸的条件下在核苷酸之间进行鉴别。本发明的孔可在不同核苷酸之间进行鉴别的程度可通过改变所施加的电势、盐浓度、缓冲液、温度和添加剂的存在(如脲、甜菜碱和DTT)来控制。这允许对孔的功能进行微调,尤其当测序时。这一点在下文更详细论述。本发明的孔还可用于根据与一或多个单体的相互作用而不是根据核苷酸紧接核苷酸,来鉴定聚核苷酸聚合物。

[0417] 本发明的孔可为经分离,大体上经分离、纯化或大体上纯化的。如果本发明的孔完全不含任何其它组分(如脂质或其它孔),那么其是经分离或纯化的。如果孔是与将不干扰其既定用途的载体或稀释剂混合的,那么其是大体上经分离的。举例来说,如果孔是以包含小于10%、小于5%、小于2%或小于1%的其它组分(如三嵌段共聚物、脂质或其它孔)的形式存在,那么其大体上是经分离或大体上经纯化的。替代地,本发明的孔可存在于膜中。下文论述合适膜。

[0418] 本发明的孔可呈个别或单一孔存在。替代地,本发明的孔可以同源孔群体或两种或更多种孔的异源群体存在。

[0419] 同源寡聚孔

[0420] 本发明还提供一种同源寡聚孔,其衍生自CsgG,所述同源寡聚孔包含相同的本发明的突变单体。同源寡聚孔可包含本发明的突变中的任一个。本发明的同源寡聚孔对于表征(如测序)聚核苷酸是理想的。本发明的同源寡聚孔可具有上文所论述的任一个优势。

[0421] 同源寡聚孔可含有任何数量的突变单体。孔通常包含至少7、至少8、至少9或至少

10个相同突变单体,如7、8、9或10个突变单体。孔优选地包含八个或九个相同突变单体。一或多个,如2、3、4、5、6、7、8、9或10个突变单体优选地如上文所论述进行化学修饰。

[0422] 用于制备孔的方法在下文更详细地论述。

[0423] 异源寡聚孔

[0424] 本发明还提供一种异源寡聚孔,其衍生自CsgG,所述同源寡聚孔包含至少一个本发明的突变单体。本发明的异源寡聚孔对于表征(如测序)聚核苷酸是理想的。异源寡聚孔可使用本领域中已知的方法来制备(例如《蛋白质科学(Protein Sci.)》2002年7月;11(7):1813-24)。

[0425] 异源寡聚孔含有足够的单体来形成孔。单体可具有任何类型。孔通常包含至少7、至少8、至少9或至少10个单体,如7、8、9或10个单体。孔优选地包含八个或九个单体。

[0426] 在一优选实施例中,所有单体(如10、9、8或7个单体)均是本发明的突变单体,且其中的至少一个不同于另外的突变单体。在一更优选实施例中,孔包含八个或九个本发明的突变单体,且其中的至少一个不同于另外的突变单体。其所有可彼此不同。

[0427] 孔中的本发明的突变单体的长度优选地大致相同,或相同。孔中的本发明的突变的折叠桶的长度优选地大致相同,或相同。长度可以氨基酸数量和/或长度单位的形式测量。

[0428] 在另一优选实施例中,突变单体中的至少一个不是本发明的突变单体。在这个实施例中,残余单体优选地是本发明的突变单体。因此,孔可包含9、8、7、6、5、4、3、2或1个本发明的突变单体。任何数量的孔中的单体可以不是本发明的突变单体。孔优选地包含七或八个本发明的突变单体,且一单体不是本发明的单体。本发明的突变单体可相同或不同。

[0429] 构筑体中的本发明的突变单体的长度优选地大致相同,或相同。构筑体中的本发明的突变的折叠桶的长度优选地大致相同,或相同。长度可以氨基酸数量和/或长度单位的形式测量。

[0430] 孔可包含不是本发明的突变单体的一或多个单体。不是本发明的突变单体的CsgG单体包括:单体,其包含SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41;或以下的比较变异体:SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41,其中上文关于本发明所论述的氨基酸/位置无一个突变/取代。基于氨基酸同源性,SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41的比较变异体通常在其整个序列上与SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41至少50%同源。更优选地,基于氨基酸同源性,比较变异体可在整个序列上与SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41的氨基酸序列,至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,并且更优选地至少95%、97%或99%同源。

[0431] 在所有上文所论述的实施例中,一或多个,如2、3、4、5、6、7、8、9或10个突变单体优选地如上文所论述进行化学修饰。

[0432] 用于制备孔的方法在下文更详细地论述。

[0433] 含构筑体的孔

[0434] 本发明还提供一种孔,其包含至少一个本发明的构筑体。本发明的构筑体包含两个或更多个共价连接的衍生自CsgG的单体,其中所述单体中的至少一个是本发明的突变单

体。换句话说,构筑体必须含有超过一个单体。孔含有足够的构筑体和(视需要)单体,以形成孔。举例来说,八聚体孔可包含:(a)四个构筑体,其各自包含两个构筑体;(b)两个构筑体,其各自包含四个单体,或(b)一个构筑体,其包含两个单体和六个不形成构筑体的一部分的单体。举例来说,九聚体孔可包含:(a)四个构筑体,其各自包含两个构筑体和一个并不形成构筑体的一部分的单体;(b)两个构筑体,其各自包含四个单体和一并不形成构筑体的一部分的单体,或(b)一个构筑体,其包含两个单体和七个不形成构筑体的一部分的单体。构筑体和单体的其它组合可由所属领域的技术人员来设想。

[0435] 孔中的至少两个单体是呈本发明的构筑体形式。构筑体和因此孔包含至少一个本发明的突变单体。孔通常总共包含至少7、至少8、至少9或至少10个单体,如7、8、9或10个单体(其中的至少两个必须呈构筑体形式)。孔优选地包含八个或九单体(其中的至少两个必须呈构筑体形式)。

[0436] 含有构筑体的孔可为同源寡聚物(即包括相同的构筑体)或为异源寡聚物(即其中至少一个构筑体不同于另外的构筑体)。

[0437] 孔通常含有:(a)一个构筑体,其包含两个单体;和(b)5、6、7或8个单体。构筑体可以是上文所论述的那些突变构筑体中的任一个。单体可以是上文所论述的那些单体中的任一个,包括:本发明的突变单体;单体,其包含SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41;和突变单体,其包含如上文所论述的SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41的比较变异体。

[0438] 另一典型孔包含超过一个本发明的构筑体,如两个、三个或四个本发明的构筑体。视需要,此类孔进一步包含足够的额外单体或构筑体,来形成孔。额外单体可以是上文所论述的那些单体中的任一个,包括:本发明的突变单体;单体,其包含SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41;和突变单体,其包含如上文所论述的SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41的比较变异体。额外构筑体可以是上文所论述的那些构筑体中的任一个,或可为包含两个或更多个共价连接的CsgG单体的构筑体,所述单体各自包含:单体,其包含SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41;或以下的比较变异体:如上文所论述的SEQ ID NO:2、3、4、5、6、7、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40或41。

[0439] 本发明的其它孔仅包含含有2个单体的构筑体,例如孔可包含含有2个单体的4、5、6、7或8个构筑体。至少一个构筑体是本发明的构筑体,即至少一个构筑体中的至少一个单体,并优选地至少一个构筑体中的各单体是本发明的突变单体。包含2个单体的所有构筑体均可为本发明的构筑体。

[0440] 根据本发明的特异性孔包含四个本发明的构筑体,其各自包含两个单体,其中各构筑体中的至少一个单体,并优选地各构筑体中的各单体是本发明的突变单体。构筑体可寡聚成孔,所述孔具有使得仅各构筑体的一个单体造成孔的通道。通常,构筑体的其它单体将在孔的通道外部上。举例来说,本发明的孔可包含7、8、9或10个构筑体,所述构筑体包含2个单体,其中通道包含7、8、9或10个单体。

[0441] 可将突变引入到如上文所描述的构筑体中。突变可为交替的,即突变对于两单体构筑体内的各单体是不同的,且构筑体组装为同源寡聚物,从而产生交替的修饰。换句话

说,将包含MutA和MutB的单体融合和组装,形成A-B:A-B:A-B:A-B孔。替代地,突变可为相邻的,即将相同突变引入到构筑体中的两个单体中,且这接着与不同突变单体或构筑体寡聚。换句话说,将包含MutA的单体融合,随后与含MutB的单体寡聚反应,形成A-A:B:B:B:B:B:B。

[0442] 含构筑体的孔中的本发明的单体中的一或多个可如上文所述进行化学修饰。

[0443] 分析物表征

[0444] 本发明提供一种测定靶分析物是否存在或其一或多个特征的方法。方法涉及:使靶分析物与本发明的孔接触,使得靶分析物相对于(如通过)孔移动,和在分析物相对于孔移动时进行一或多个测量,且从而测定分析物是否存在或其一或多个特征。靶分析物还可被称为模板分析物或所关注的分析物。

[0445] 步骤(a)和(b)优选地在孔两端施加电势的情况下进行。如下文更详细地论述,所施加的电势通常使得在孔与聚核苷酸结合蛋白之间形成复合体。所施加的电势可为电压电势。替代地,所施加的电势可为化学势。此的实例为在两亲层两端使用盐梯度。盐梯度公开于Holden等人《美国化学学会杂志》2007年7月11日;129(27):8650-5中。

[0446] 方法是用于测定靶分析物是否存在或其一或多个特征。方法可用于测定分析物是否存在或其一或多个特征。方法可涉及测定分析物是否存在或其一或多个特征。方法可包含测定任何数量的分析物是否存在或其一或多个特征,如2、5、10、15、20、30、40、50、100种或更多种分析物。可测定一或多种分析物的任何数量的特征,如1、2、3、4、5、10个或更多个特征。

[0447] 靶分析物优选地是金属离子、无机盐、聚合物、氨基酸、肽、多肽、蛋白质、核苷酸、寡核苷酸、聚核苷酸、染料、漂白剂、药物、诊断剂、娱乐性药物、爆炸性或环境污染物。方法可涉及测定两种或更多种相同类型的分析物是否存在或其一或多个特征,如两种或更多种蛋白质、两种或更多种核苷酸或两种或更多种药物。替代地,方法可涉及测定两种或更多种不同类型的分析物是否存在或其一或多个特征,如一或多种蛋白质、一或多种核苷酸和一或多种药物。

[0448] 可从细胞分泌靶分析物。替代地,靶分析物可为存在于细胞内部使得在可进行本发明之前必须从细胞提取分析物的分析物。

[0449] 分析物优选地是氨基酸、肽、多肽和/或蛋白质。氨基酸、肽、多肽或蛋白质可为天然存在的或非天然存在的。多肽或蛋白质可包括在其合成或修饰氨基酸内。对于氨基酸的多种不同类型修饰在本领域中已知。合适氨基酸和其修饰如上文。出于本发明的目的,应理解,可通过所属领域中任何可用的方法来修饰靶分析物。

[0450] 蛋白质可酶、抗体、激素、生长因子或生长调控蛋白,如细胞因子。细胞因子可选自:白介素,优选地IFN-1、IL-1、IL-2、IL-4、IL-5、IL-6、IL-10、IL-12和IL-13;干扰素,优选地IL- γ ;和其它细胞因子,如TNF- α 。蛋白质可为细菌蛋白质、真菌蛋白质、病毒蛋白质或寄生虫来源的蛋白质。

[0451] 靶分析物优选地是核苷酸、寡核苷酸或聚核苷酸。下文论述核苷酸和聚核苷酸。寡核苷酸是短核苷酸聚合物,其通常具有50个或更少个核苷酸,如40个或更少个、30个或更少个、20个或更少个、10个或更少个或5个或更少个核苷酸。寡核苷酸可包含下文论述的任一个核苷酸,包括无碱基和修饰核苷酸。

[0452] 靶分析物,如靶聚核苷酸,可存在于下文论述中的任一个合适样本中。

[0453] 孔通常存在于如下文所论述的膜中。可使用下文论述的方法,来将靶分析物偶联或递送到膜。

[0454] 下文论述中的任一个测量可用于测定靶分析物是否存在或其一或多个特征。方法优选地包含:使靶分析物与本发明的孔接触,使得靶分析物相对于(如移动通过)孔移动,和在分析物相对于孔移动时测量穿过孔的电流,且从而测定分析物是否存在或其一或多个特征。

[0455] 如果电流以对分析物具有特异性的方式流过孔(即如果检测到与分析物相关的流过孔的独特电流),那么靶分析物存在。如果电流并不以对核苷酸具有特异性的方式流过孔,那么分析物不存在。可在分析物存在下进行对照实验,以测定如果影响流过孔的电流的方式。

[0456] 本发明可基于分析物对于穿过孔的电流具有不同效果,而用于区分具有类似结构的分析物。个别分析物可根据当其与孔相互作用时的其电流幅度来在单分子水平下进行鉴定。本发明还可用于测定样本中是否存在特定分析物。本发明还可用于测量样本中的特定分析物的浓度。使用除CsgG以外的孔来进行分析物表征在本领域中已知。

[0457] 聚核苷酸表征

[0458] 本发明提供一种表征靶聚核苷酸,如测序聚核苷酸的方法。对于使用纳米孔表征或测序聚核苷酸存在两种主要策略,即链表征/测序和核酸外切酶表征/测序。本发明的方法可涉及任一个方法。

[0459] 在链测序中,通过所施加的电势或抵抗所施加的电势,DNA通过纳米孔易位。可在孔的顺侧上使用渐进地或逐渐地对双链DNA起作用的核酸外切酶,以在所施加的电势下馈送残余单链,或在逆转电势下在反侧上。同样,使双链DNA解旋的解螺旋酶也可以类似的方式使用。还可使用聚合酶。需要链抵抗施加的电势而易位的测序应用也有可能,但DNA必须首先在逆转或无电势下由酶“捕获”。随着电势接着切换回后续结合,链应以顺式到反式的方式穿过孔,且通过电流保持为延长的构形。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,所述分子马达抵抗施加的电势将最近易位的单链以逐步受控方式(反式到顺式)牵拉回反式通过孔。

[0460] 在一个实施例中,表征靶聚核苷酸的方法涉及使靶序列与本发明的孔和解螺旋酶接触。方法中可以使用任何解螺旋酶。下文论述合适解螺旋酶。解螺旋酶可以相对于孔的两种模式来起作用。首先,优选地使用解螺旋酶进行方法,使得其在由所施加电压造成的场的情况下控制靶序列移动通过孔。在这种模式中,DNA的5'端首先在孔中被捕获,且酶控制DNA移动到孔中,使得靶序列在场的情况下通过孔,直到其最终易位通过到双层的反侧为止。替代地,优选地进行方法,使得解螺旋酶在抵抗由所施加电压造成的场的情况下控制靶序列移动通过孔。在这种模式中,DNA的3'端首先在孔中被捕获,且酶控制DNA移动到孔中,使得靶序列抵抗所施加场的情况下牵拉出孔,直到其最终推回到双层的顺侧为止。

[0461] 在核酸外切酶测序中,核酸外切酶从靶聚核苷酸的一端释放个别核苷酸,且如下文所论述来鉴定这些个别核苷酸。在另一实施例中,表征靶聚核苷酸的方法涉及使靶序列与孔和核酸外切酶接触。可在方法中使用下文论述中的任一个核酸外切酶。酶可共价连接于孔,如下文所论述。

[0462] 核酸外切酶是通常扣在聚核苷酸的一端且从那个末端一次一个核苷酸来消化序

列的酶。核酸外切酶可在5'到3'方向或3'到5'方向上消化聚核苷酸。通常通过选择所使用的酶和/或使用本领域中已知的方法,来决定与核酸外切酶结合的聚核苷酸的末端。在聚核苷酸的任一末端处的羟基或帽结构可通常用于防止或促进核酸外切酶结合于聚核苷酸的特定末端。

[0463] 方法涉及使聚核苷酸与核酸外切酶接触,使得如上文所论述,以允许表征或鉴定一定比例的核苷酸的速率从聚核苷酸的末端消化核苷酸。进行此的方法在所属领域中是众所周知的。举例来说,Edman降解用于从多肽的末端连续地消化单一氨基酸,使得可使用高效液相色谱(HPLC)来鉴定其。在本发明中可使用同源方法。

[0464] 核酸外切酶起作用的速率通常比野生型核酸外切酶的最优速率慢。在本发明的方法中的核酸外切酶的活性的合适速率包括以下的消化速率:0.5到1000个核苷酸每秒、0.6到500个核苷酸每秒、0.7到200个核苷酸每秒、0.8到100个核苷酸每秒、0.9到50个核苷酸每秒或1到20或10个核苷酸每秒。速率优选地是1、10、100、500或1000个核苷酸每秒。核酸外切酶活性的合适速率可以不同方式来实现。举例来说,可根据本发明使用具有降低的最优活性速率的变异体核酸外切酶。

[0465] 在链表征实施例中,方法包含:使聚核苷酸与本发明的孔接触,使得聚核苷酸相对于(如通过)孔移动,和在聚核苷酸相对于孔移动时进行一或多个测量,其中测量指示聚核苷酸的一或多个特征,且从而表征靶聚核苷酸。

[0466] 在核酸外切酶表征实施例中,方法包含:使聚核苷酸与本发明的孔和核酸外切酶接触,使得核酸外切酶从靶聚核苷酸的一端消化个别核苷酸,和个别核苷酸相对于(如通过)孔移动,和在个别核苷酸相对于孔移动时进行一或多个测量,其中测量指示个别核苷酸的一或多个特征,且从而表征靶聚核苷酸。

[0467] 个别核苷酸是单核苷酸。个别核苷酸是不通过核苷酸键结合于另一核苷酸或聚核苷酸的核苷酸。核苷酸键涉及结合于另一核苷酸的糖基团的核苷酸的一个磷酸基。个别核苷酸通常是通过核苷酸键结合于另一聚核苷酸的核苷酸,所述聚核苷酸具有至少5、至少10、至少20、至少50、至少100、至少200、至少500、至少1000或至少5000个核苷酸。举例来说,已从靶聚核苷酸序列(如DNA或RNA链)消化个别核苷酸。核苷酸可为下文论述的那些中的任一个。

[0468] 个别核苷酸可以任何方式且在任何位点处与孔相互作用。核苷酸优选地通过如上文所论述的衔接子或结合衔接子而可逆地结合于孔。在核苷酸穿过跨膜的孔时,核苷酸最优选地通过衔接子或结合衔接子而可逆地结合于孔。在核苷酸穿过跨膜的孔时,核苷酸还可通过衔接子或结合衔接子而可逆地结合于孔的折叠桶或通道。

[0469] 在个别核苷酸与孔之间的相互作用期间,核苷酸通常以对所述核苷酸具有特异性的方式影响流过孔的电流。举例来说,特定核苷酸将减少在特定平均时段流过孔的电流,且减少到特定程度。换句话说,对于特定核苷酸,流过孔的电流是独特的。可进行对照实验,以测定特定核苷酸对于流过孔的电流的影响。接着,可将在测试样本上进行本发明的方法的结果与来自此类对照实验的结果进行比较,以鉴定样本中的特定核苷酸或测定在样品中是否存在特定核苷酸。在所述频率下以指示特定核苷酸的方式影响流过孔的电流的频率,可用于测定样本中的所述核苷酸的浓度。还可计算样本内不同核苷酸的比率。举例来说,可计算dCMP与甲基-dCMP的比率。

[0470] 方法涉及测量靶聚核苷酸的一或多个特征。靶聚核苷酸还可被称为模板聚核苷酸或所关注的聚核苷酸。

[0471] 这个实施例也使用本发明的孔。可使用上文关于靶分析物所论述的孔和实施例中的任一个。

[0472] 聚核苷酸

[0473] 聚核苷酸(如核酸)是包含两个或更多个核苷酸的大分子。聚核苷酸或核酸可包含任何核苷酸的任何组合。核苷酸可以是天然存在的或人工的。聚核苷酸中的一或多个核苷酸可以是氧化或甲基化的。聚核苷酸中的一或多个核苷酸可以是受损的。举例来说,聚核苷酸可以包含嘧啶二聚体。此类二聚体通常与紫外光的损坏相关,且是皮肤黑色素瘤的主要病因。聚核苷酸中的一或多个核苷酸可例如用标记或标签修饰。合适标记在下文描述。聚核苷酸可包含一或多个间隔子。

[0474] 核苷酸通常含有核碱基、糖和至少一个磷酸基。核碱基和糖形成核苷。

[0475] 核碱基通常是杂环的。核碱基包括(但不限于)嘌呤和嘧啶,且更详细地说腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)、尿嘧啶(U)和胞嘧啶(C)。

[0476] 糖通常是戊糖。核苷酸糖包括(但不限于)核糖和脱氧核糖。糖优选地是脱氧核糖。

[0477] 聚核苷酸优选地包含以下核苷:脱氧腺苷(dA)、脱氧尿苷(dU)和/或胸苷(dT)、脱氧鸟苷(dG)和脱氧胞苷(dC)。

[0478] 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。核苷酸可包含超过三个磷酸,如4或5个磷酸。磷酸可连接在核苷酸的5'或3'侧上。核苷酸包括(但不限于):单磷酸腺苷(AMP)、单磷酸鸟苷(GMP)、单磷酸胸苷(TMP)、单磷酸尿苷(UMP)、单磷酸5-甲基胞嘧啶、单磷酸5-羟基甲基胞嘧啶、单磷酸胞嘧啶核苷(CMP)、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、单磷酸脱氧腺苷(dAMP)、单磷酸脱氧鸟苷(dGMP)、单磷酸脱氧胸苷(dTMP)、单磷酸脱氧尿苷(dUMP)、单磷酸脱氧胞苷(dCMP)和单磷酸脱氧甲基胞苷。核苷酸优选地是选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP、dCMP和dUMP。

[0479] 核苷酸可以无碱基(即缺乏核碱基)。核苷酸还可缺乏核碱基和糖(即是C3间隔子)。

[0480] 聚核苷酸中的核苷酸可以任何方式彼此连接。如在核酸中,核苷酸通常通过其糖和磷酸基连接。如在嘧啶二聚体中,核苷酸可通过其核碱基连接。

[0481] 聚核苷酸可以是单链或双链。聚核苷酸优选地是单链。在实例中,单链聚核苷酸表征是被称作1D。聚核苷酸的至少部分可以是双链。

[0482] 聚核苷酸可以是核酸,例如脱氧核糖核酸(DNA)或核糖核酸(RNA)。聚核苷酸可以包含与一条DNA链杂交的一条RNA链。聚核苷酸可以是所属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N-(2-氨基乙基)-甘氨酸重复单元构成。GNA主链由通过磷酸二酯键连接的乙二醇重复单元构成。TNA主链由通过磷酸二酯键连接在一起的重复苏糖构成。LNA是由如上文所论述的具有将核糖部分中的2'氧和4'碳连接的额外桥的核糖核苷酸形成。桥接核酸(BNA)是经修饰的RNA核苷酸。其还可被称为受限或不可接近的RNA。BNA单体可含有五元、六元或甚至七元桥接结构,其具有“固定”的C3'-内糖褶皱。在核糖的2', 4'-位置处合成地并入桥,产生2', 4'-BNA单体。

[0483] 聚核苷酸最优选地是核糖核酸(RNA)或脱氧核糖核酸(DNA)。

[0484] 聚核苷酸可以是任何长度的。举例来说,聚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。聚核苷酸的长度可以是1000个或更多个核苷酸或核苷酸对、5000个或更多个核苷酸或核苷酸对、或100000个或更多个核苷酸或核苷酸对。

[0485] 可研究任何数量的聚核苷酸。举例来说,本发明的方法可涉及表征2、3、4、5、6、7、8、9、10、20、30、50、100或更多个聚核苷酸。如果表征两个或更多个聚核苷酸,那么其可以是不同的聚核苷酸或两个相同聚核苷酸的例子。

[0486] 聚核苷酸可以是天然存在的或人工的。举例来说,方法可用于验证所制造寡核苷酸的序列。通常在活体外进行所述方法。

[0487] 样本

[0488] 聚核苷酸通常存在于任何合适样本中。通常在已知含有或疑似含有聚核苷酸的样本进行上本发明。替代地,可在确认聚核苷酸的一致性的样本上进行本发明,已知或预期所述聚核苷酸在样本中存在。

[0489] 样本可以是生物样本。可使用从任何生物体或微生物获得或提取的样本在活体外进行本发明。生物体或微生物通常是古细菌、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。可在从任何病毒获得或提取的样本上在活体外进行本发明。样本优选地是流体样本。样本通常包含患者的体液。样本可以是尿液、淋巴液、唾液、粘液或羊膜液,但优选地是血液、血浆或血清。

[0490] 通常,样本是人类来源的;但替代地其可以是来自另一哺乳动物,如来自商业养殖动物,如马、牛、绵羊、鱼、鸡或猪;或替代地可以是宠物,如猫或狗。替代地,样本可以是植物来源的,如从经济作物获得的样本,如谷物、豆科植物、果实或植物,例如小麦、大麦、燕麦、芥花、玉米、大豆、稻谷、大黄、香蕉、苹果、蕃茄、马铃薯、葡萄、烟草、菜豆、小扁豆、甘蔗、可可、棉花。

[0491] 样本可以是非生物样本。非生物样本优选地是流体样本。非生物样本的实例包括手术液、水(如饮用水、海水或河水)和实验室测试用试剂。

[0492] 通常在于本发明中使用之前处理样本,例如通过离心或通过穿过滤出不需要的分子或细胞(如红细胞)的膜。可紧接着在获取后进行测量。还可通常在分析之前存储样本,优选地低于-70℃。

[0493] 表征

[0494] 方法可涉及测量聚核苷酸的两个、三个、四个或五个或更多个特征。一或多个特征优选地是选自:(i)聚核苷酸的长度;(ii)聚核苷酸的一致性;(iii)聚核苷酸的序列;(iv)聚核苷酸的二级结构;和(v)聚核苷酸是否被修饰。可根据本发明来测量(i)到(v)的任何组合,如:{i}、{ii}、{iii}、{iv}、{v}、{i,ii}、{i,iii}、{i,iv}、{i,v}、{ii,iii}、{ii,iv}、{ii,v}、{iii,iv}、{iii,v}、{iv,v}、{i,ii,iii}、{i,ii,iv}、{i,ii,v}、{i,iii,iv}、{i,iii,v}、{i,iv,v}、{ii,iii,iv}、{ii,iii,v}、{ii,iv,v}、{iii,iv,v}、{i,ii,iii,iv}、{i,ii,iii,v}、{i,ii,iv,v}、{i,iii,iv,v}、{ii,iii,iv,v}或{i,ii,iii,iv,v}。可对于与第二聚核苷酸进行比较的第一聚核苷酸,测量(i)到(v)的不同组合,包括上文所列的那些组合中的任一个。

[0495] 对于(i),可例如通过测定聚核苷酸与孔之间的相互作用的数量或聚核苷酸与孔之间的相互作用的持续时间,来测量聚核苷酸的长度。

[0496] 对于(ii),可以多种方式来测量聚核苷酸的一致性。可结合测量聚核苷酸的序列或不测量聚核苷酸的序列,来测量聚核苷酸的一致性。前者是直接了当的;测序聚核苷酸且从而进行鉴定。可以若干方式来进行后者。举例来说,可测量聚核苷酸中的特定基序的存在(而不测量聚核苷酸的残余序列)。替代地,方法中的特定电子和/或光学信号的测量可鉴定来自特定来源的聚核苷酸。

[0497] 对于(iii),可如先前所描述来测定聚核苷酸的序列。合适测序方法,尤其使用电测量的那些测序方法,描述于以下中:Stoddart D等人,《美国国家科学院院刊》,12;106(19):7702-7,Lieberman KR等人,《美国化学学会杂志》2010;132(50):17961-72,和国际申请W0 2000/28312。

[0498] 对于(iv),可以多种方式来测量二级结构。举例来说,如果方法涉及电测量,那么可使用停留时间的变化或流过孔的电流的变化来测量二级结构。这允许区分单链和双链聚核苷酸的区。

[0499] 对于(v),可测量是否存在任何修饰。方法优选地包含,用一或多个蛋白质或用一或多个标记、标签或间隔子,测定聚核苷酸是否通过甲基化、通过氧化、通过损坏来修饰。特异性修饰将引起与可使用以下所描述的方法测量的孔的特异性相互作用。举例来说,可基于在孔与各核苷酸相互作用期间流过孔的电流,将甲基胞嘧啶与胞嘧啶区别开。

[0500] 使靶聚核苷酸与本发明的孔接触。孔通常存在于膜中。下文论述合适膜。可使用适合于研究膜/孔系统的任何设备来进行方法,其中孔存在于膜中。可使用适合于跨膜孔感测的任何设备来进行方法。举例来说,设备包含含有水溶液的腔室和将腔室分隔为两个部分的屏障。屏障通常具有开孔,在开孔中形成含有孔的膜。替代地,屏障形成其中存在孔的膜。

[0501] 可使用描述于国际申请第PCT/GB08/000562号(W0 2008/102120)中的设备来进行方法。

[0502] 可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。可能电测量包括:电流测量、阻抗测量、隧道测量(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W0 2005/124888)。可将光学测量与电测量组合(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,如流过孔的离子电流的测量。

[0503] 可使用标准单通道记录装置来进行电测量,如Stoddart D等人,《美国国家科学院院刊》,12;106(19):7702-7,Lieberman KR等人,《美国化学学会杂志》2010;132(50):17961-72,和国际申请W0 2000/28312中所描述。替代地,可使用多通道系统来进行电测量,例如如国际申请W0 2009/077734和国际申请W0 2011/067559中所描述。

[0504] 优选地通过在膜两端施加的电势来进行所述方法。所施加的电势可为电压电势。替代地,所施加的电势可为化学势。此的实例是在膜(如两亲层)两端使用盐梯度。盐梯度公开于Holden等人《美国化学学会杂志》2007年7月11日;129(27):8650-5中。在一些例子中,在聚核苷酸相对于孔移动时穿过孔的电流用于评估或测定聚核苷酸的序列。这是链测序。

[0505] 方法可涉及在聚核苷酸相对于孔移动时测量穿过孔的电流。因此,在方法中使用的设备还可包含能够施加电势和测量膜和孔两端的电信号的电路。可使用贴片钳或电压钳

来进行方法。方法优选地涉及使用电压钳。

[0506] 本发明的方法可涉及在聚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合适条件是所属领域中已知的,且公开于实例中。通常通过在膜和孔两端施加的电压来进行所述方法。所用电压通常是+5V到-5V,如+4V到-4V、+3V到-3V,或+2V到-2V。所用电压通常是-600mV到+600mV或-400mV到+400mV。所用电压优选地在具有下限和上限的范围内,下限选自:-400mV、-300mV、-200mV、-150mV、-100mV、-50mV、-20mV和0mV,上限独立地选自:+10mV、+20mV、+50mV、+100mV、+150mV、+200mV、+300mV和+400mV。所用电压更优选地在100mV到240mV的范围内,并且最优选在120mV到220mV的范围内。有可能通过使用增大的所施加电势来增大通过孔对不同核苷酸之间的鉴别。

[0507] 通常在存在任何电荷载流子的情况下进行方法,电荷载流子如金属盐,例如碱金属盐、卤盐,例如氯化物盐,例如碱金属氯化物盐。电荷载体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。在下文论述的示范性设备中,盐在腔室中的水溶液中存在。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。优选KCl、NaCl,以及亚铁氰化钾和铁氰化钾的混合物。电荷载流子可以是跨膜不对称的。举例来说,电荷载流子的类型和/或浓度可在膜的各侧上不同。

[0508] 盐浓度可以是饱和的。盐浓度可以是3M或更低,且通常0.1到2.5M、0.3到1.9M、0.5到1.8M、0.7到1.7M、0.9到1.6M或1M到1.4M。盐浓度优选地是150mM到1M。优选地使用至少0.3M的盐浓度来进行方法,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且允许在正常电流波动的背景下鉴定指示核苷酸存在的电流。

[0509] 通常在缓冲液存在下进行方法。在下文论述的示范性设备中,缓冲液在腔室中的水溶液中存在。在本发明的方法中可使用任何缓冲液。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。通常在以下的pH下来进行方法:4.0到12.0、4.5到10.0、5.0到9.0、5.5到8.8、6.0到8.7或7.0到8.8或7.5到8.5。所用pH优选地是约7.5。

[0510] 可在以下温度下来进行方法:0℃到100℃、15℃到95℃、16℃到90℃、17℃到85℃、18℃到80℃、19℃到70℃或20℃到60℃。通常在室温下进行方法。任选地在支持酶功能的温度(如约37℃)下进行方法。

[0511] 聚核苷酸结合蛋白

[0512] 链表征方法优选地包含使聚核苷酸与聚核苷酸结合蛋白接触,使得蛋白质控制聚核苷酸相对于(如通过)孔的移动。

[0513] 更优选地,方法包含:(a)使聚核苷酸与本发明的孔和聚核苷酸结合蛋白接触,使得蛋白质控制聚核苷酸相对于(如通过)孔的移动;和(b)在聚核苷酸相对于孔移动时进行一或多个测量,其中测量指示聚核苷酸的一或多个特征,且从而表征聚核苷酸。

[0514] 更优选地,方法包含:(a)使聚核苷酸与本发明的孔和聚核苷酸结合蛋白接触,使得蛋白质控制聚核苷酸相对于(如通过)孔的移动;和(b)在聚核苷酸相对于孔移动时测量通过孔的电流,其中电流指示聚核苷酸的一或多个特征,且从而表征聚核苷酸。

[0515] 聚核苷酸结合蛋白可以是能够结合于聚核苷酸且控制其移动通过孔的任何蛋白质。在所属领域中测定蛋白质是否结合于聚核苷酸是直接了当的。蛋白质通常与聚核苷酸

相互作用,且修改其的至少一个特性。蛋白质可通过裂解聚核苷酸形成个别核苷酸或较短核苷酸链(如二或三核苷酸)来修改聚核苷酸。蛋白质可通过将聚核苷酸定位或移动到特异性位置(即控制其移动)来修改聚核苷酸。

[0516] 聚核苷酸结合蛋白优选地衍生自聚核苷酸操作酶。聚核苷酸操作酶是能够与聚核苷酸相互作用且修改其的至少一个特性的多肽。酶可通过裂解聚核苷酸形成个别核苷酸或较短核苷酸链(如二或三核苷酸)来修改聚核苷酸。酶可通过将聚核苷酸定位或移动到特异性位置来修改聚核苷酸。聚核苷酸操作酶并不需要显示酶促活性,只要其能够结合聚核苷酸且控制其移动通过孔即可。举例来说,酶可进行修饰以移除其酶促活性,或可在防止其充当酶的条件下使用。在下文更详细地论述此类条件。

[0517] 聚核苷酸操作酶优选地衍生自核分解酶。酶的构筑体中所用的聚核苷酸操作酶更优选地是衍生自以下酶分类(EC)组中的任一个的成员: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。酶可以是在国际申请第PCT/GB10/000133号(公开为W0 2010/086603)中所公开任一种酶。

[0518] 优选酶是聚合酶、核酸外切酶、解螺旋酶和拓扑异构酶,如回旋酶。合适酶包括(但不限于)来自大肠杆菌的核酸外切酶I(SEQ ID NO:11)、来自大肠杆菌的核酸外切酶III(SEQ ID NO:13)、来自极端嗜热菌的RecJ(SEQ ID NO:15)和噬菌体 λ 核酸外切酶(SEQ ID NO:17)、TatD核酸外切酶和其变异体。包含SEQ ID NO:15中所示序列的三个亚基或其变异体相互作用,形成三聚体核酸外切酶。这些核酸外切酶还可用于本发明的核酸外切酶方法中。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®)或其变异体。酶优选地是Phi29DNA聚合酶(SEQ ID NO:9)或其变异体。拓扑异构酶优选地是酶分类(EC)组5.99.1.2和5.99.1.3中的任一个的成员。

[0519] 酶最优选地是衍生自解螺旋酶,如Hel308 Mbu(SEQ ID NO:18)、Hel308 Csy(SEQ ID NO:19)、Hel308 Tga(SEQ ID NO:20)、Hel308 Mhu(SEQ ID NO:21)、TraI Eco(SEQ ID NO:22)、XPD Mbu(SEQ ID NO:23)或其变异体。可在本发明中使用任何解螺旋酶。解螺旋酶可以是或衍生自Hel308解螺旋酶、RecD解螺旋酶(如TraI解螺旋酶或TrwC解螺旋酶)、XPD解螺旋酶或Dda解螺旋酶。解螺旋酶可以是在中以下国际申请中所公开的任何解螺旋酶、修饰解螺旋酶或解螺旋酶构筑体:第PCT/GB2012/052579号(公开为W0 2013/057495)、第PCT/GB2012/053274号(公开为W0 2013/098562)、第PCT/GB2012/053273号(公开为W02013098561)、第PCT/GB2013/051925号(公开为W0 2014/013260)、第PCT/GB2013/051924号(公开为W0 2014/013259)、第PCT/GB2013/051928号(公开为W0 2014/013262)和第PCT/GB2014/052736号。

[0520] 解螺旋酶优选地包含SEQ ID NO:25中所示的序列(Trwc Cba)或其变异体、SEQ ID NO:18中所示的序列(Hel308 Mbu)或其变异体,或SEQ ID NO:24中所示的序列(Dda)或其变异体。变异体可以下文关于跨膜孔所论述中的任一个方式来不同于原生序列。SEQ ID NO:24的优选变异体包含:(a)E94C和A360C;或(b)E94C、A360C、C109A和C136A且接着任选地(Δ M1)G1G2(即缺失M1且接着添加G1和G2)。

[0521] 根据本发明可以使用任何数量的解螺旋酶。举例来说,可使用1、2、3、4、5、6、7、8、9、10个或更多个解螺旋酶。在一些实施例中,可使用不同数量的解螺旋酶。

[0522] 本发明的方法优选地包含使聚核苷酸与两个或更多个解螺旋酶接触。两个或更多

个解螺旋酶通常是相同的解螺旋酶。两个或更多个解螺旋酶可以是不同的解螺旋酶。

[0523] 两个或更多个解螺旋酶可以是上文提到的解螺旋酶的任何组合。两个或更多个解螺旋酶可以是两个或更多个Dda解螺旋酶。两个或更多个解螺旋酶可以是一或多个Dda解螺旋酶和一或多个TrwC解螺旋酶。两个或更多个解螺旋酶可以是相同解螺旋酶的不同变异体。

[0524] 两个或更多个解螺旋酶优选地彼此连接。两个或更多个解螺旋酶更优选地彼此共价连接。解螺旋酶可以按任何次序和使用任何方法来连接。用于本发明中的优选解螺旋酶构筑体描述于以下国际申请中：第PCT/GB2013/051925号（公开为WO 2014/013260）、第PCT/GB2013/051924号（公开为WO 2014/013259）、第PCT/GB2013/051928号（公开为WO 2014/013262）和第PCT/GB2014/052736号。

[0525] SEQ ID NO:9、11、13、15、17、18、19、20、21、22、23、24或25的变异体是氨基酸序列不同于SEQ ID NO:9、11、13、15、17、18、19、20、21、22、23、24或25且其保持聚核苷酸结合能力的酶。这可使用所属领域中已知的任何方法来测量。举例来说，可使变异体与聚核苷酸接触，且可测量其结合于聚核苷酸和沿所述聚核苷酸移动的能力。变异体可包括便于聚核苷酸结合和/或便于其在高盐浓度和/或室温下活性的修饰。变异体可进行修饰，使得其结合聚核苷酸（即保持聚核苷酸结合能力）但不充当解螺旋酶（即当具备所有便于移动的必需组分（例如ATP和 Mg^{2+} ）时不沿聚核苷酸移动）。此类修饰是所属领域中已知的。举例来说，解螺旋酶中的 Mg^{2+} 结合域的修饰通常产生不充当解螺旋酶的变异体。这些变异体类型可充当分子制动器（参见下文）。

[0526] 基于氨基酸同源性，在SEQ ID NO:9、11、13、15、17、18、19、20、21、22、23、24或25的氨基酸序列的整个长度上，变异体将优选地与所述序列至少50%同源。更优选地，基于氨基酸同源性，变异体多肽可在整个序列上与SEQ ID NO:9、11、13、15、17、18、19、20、21、22、23、24或25的氨基酸序列，至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%，并且更优选地至少95%、97%或99%同源。在200个或更多个，例如230、250、270、280、300、400、500、600、700、800、900或1000个或更多个邻接氨基酸的延伸段上，可存在至少80%，例如至少85%、90%或95%氨基酸同源性（“硬同源性”）。如上文所描述来测定同源性。变异体可以上文关于以上SEQ ID NO:2所论述中的任一个方式来不同于野生型序列。酶可共价连接于孔。任何方法可用于将酶共价连接于孔。

[0527] 优选分子制动器是TrwC Cba-Q594A（具有突变Q594A的SEQ ID NO:25）。变异体并不充当解螺旋酶（即当具备所有便于移动的必需组分（例如ATP和 Mg^{2+} ）时，结合聚核苷酸但并不沿其移动）。

[0528] 在链测序中，通过所施加的电势或抵抗所施加的电势，聚核苷酸通过纳米孔易位。可在孔的顺侧上使用渐进地或逐渐地对双链聚核苷酸起作用的核酸外切酶，以在所施加的电势下馈送残余单链，或在逆转电势下在反侧上。同样，使双链DNA解旋的解螺旋酶也可以类似的方式使用。还可使用聚合酶。需要链抵抗施加的电势而易位的测序应用也有可能，但DNA必须首先在逆转或无电势下由酶“捕获”。随着电势接着切换回后续结合，链应以顺式到反式的方式穿过孔，且通过电流保持为延长的构形。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，所述分子马达抵抗施加的电势将最近易位的单链以逐步受控方式（反式到顺式）牵拉回反式通过孔。

[0529] 方法中可以使用任何解螺旋酶。解螺旋酶可以相对于孔的两种模式来起作用。首先,优选地使用解螺旋酶来进行方法,使得在由所施加电压造成的场的情况下,解螺旋酶使聚核苷酸通过孔移动。在这种模式中,聚核苷酸的5'端首先在孔中被捕获,且解螺旋酶使聚核苷酸移动到孔中,使得其在场的情况下通过孔,直到其最终易位通过到膜的反侧为止。替代地,优选地进行方法,使得抵抗由所施加电压造成的场的情况下,解螺旋酶使聚核苷酸移动通过孔。在这种模式中,聚核苷酸的3'端首先在孔中被捕获,且解螺旋酶使聚核苷酸移动到孔中,使得其抵抗所施加场的情况下牵拉出孔,直到其最终推回到膜的顺侧为止。

[0530] 还可在相反的方向上进行方法。聚核苷酸的3'端可首先在孔中被捕获,且解螺旋酶可使聚核苷酸移动到孔中,使得其在场的情况下通过孔,直到其最终易位通过到膜的反侧为止。

[0531] 当解螺旋酶不具备便于移动的必需组分,或进行修饰以阻止或防止其移动时,当其通过所施加的场牵拉到孔中时,其可结合于聚核苷酸,且充当减缓聚核苷酸移动的制动器。在非活动模式中,是3'还是5'向下捕获聚核苷酸是不重要的,是所施加的场,通过充当制动器的酶,朝向反式侧而将聚核苷酸牵拉到孔中。当在非活动模式中时,通过解螺旋酶对聚核苷酸的移动控制可以多种方式(包括棘轮、滑动和制动)描述。还可以这种方式来使用缺乏解螺旋酶活性的解螺旋酶变异性。

[0532] 可按任何次序使聚核苷酸与聚核苷酸结合蛋白和孔接触。优选的是,当使聚核苷酸与聚核苷酸结合蛋白(如解螺旋酶)和孔接触时,聚核苷酸首先与蛋白质形成复合体。当在孔两端施加电压时,接着聚核苷酸/蛋白质复合体与孔形成复合体,且控制聚核苷酸通过孔的移动。

[0533] 通常在游离核苷酸或游离核苷酸类似物和有助于聚核苷酸结合蛋白作用的酶辅因子存在下,进行方法中使用聚核苷酸结合蛋白的任何步骤。游离核苷酸可以是上文所论述的任一个个别核苷酸中的一或多个。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)和三磷酸脱氧胞苷(dCTP)。游离核苷酸优选地是选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP或dCMP。游离核苷酸优选地是三磷酸腺苷(ATP)。酶辅因子是一种允许构筑体起作用的因子。酶辅因子优选地是二价金属阳离子。二价金属阳离子优选地是 Mg^{2+} 、 Mn^{2+} 、 Ca^{2+} 或 Co^{2+} 。酶辅因子最优选地是 Mg^{2+} 。

[0534] 解螺旋酶和分子制动器

[0535] 在一优选实施例中,方法包含:

[0536] (a) 提供聚核苷酸与一或多个解螺旋酶以及连接于聚核苷酸的一或多个分子制动器;

[0537] (b) 使聚核苷酸与本发明的孔接触,且在孔两端施加电势,使得一或多个解螺旋酶

和一或多个分子制动器带到一起,且均控制聚核苷酸相对于(如通过)孔的移动;

[0538] (c) 在所述聚核苷酸相对于所述孔移动时,进行一或多个测量,其中测量指示聚核苷酸的一或多个特征,且从而表征聚核苷酸。

[0539] 这个方法类型在国际申请PCT/GB2014/052737中详细论述

[0540] 一或多个解螺旋酶可以是上文所论述的那些解螺旋酶中的任一个。一或多个分子制动器可以是结合于聚核苷酸且减缓聚核苷酸通过孔的移動的任何化合物或分子。一或多个分子制动器优选地包含结合于聚核苷酸的一或多个化合物。一或多个化合物优选地是一或多个巨环化合物。合适巨环化合物包括(但不限于)环糊精、杯芳烃、环肽、冠醚、葫芦脲、柱状芳烃、其衍生物或其组合。环糊精或其衍生物可以是在Eliseev, A. V. 和Schneider, H. J. (1994)《美国化学学会杂志》116, 6081-6088中所公开那些化合物中的任一个。药剂更优选地是七-6-氨基- β -环糊精(am₇- β CD)、6-单脱氧-6-单氨基- β -环糊精(am₁- β CD)或七-(6-脱氧-6-胍基)-环糊精(gu₇- β CD)。

[0541] 一或多个分子制动器优选地是一或多个单链结合蛋白(SSB)。一或多个分子制动器更优选地是:单链结合蛋白(SSB),其包含并不具有净负电荷的羧基端(C端)区;或(ii)经修饰SSB,其包含在其C端区中减少C端区的净负电荷的一或多个修饰。一或多个分子制动器最优选地是国际申请第PCT/GB2013/051924号(公开为W0 2014/013259)中所公开的SSB中的一个。

[0542] 一或多个分子制动器优选地是一或多个聚核苷酸结合蛋白。聚核苷酸结合蛋白可以是能够结合于聚核苷酸且控制其移动通过孔的任何蛋白质。在所属领域中测定蛋白质是否结合于聚核苷酸是直接了当的。蛋白质通常与聚核苷酸相互作用,且修改其的至少一个特性。蛋白质可通过裂解聚核苷酸形成个别核苷酸或较短核苷酸链(如二或三核苷酸)来修改聚核苷酸。所述部分可通过将聚核苷酸定位或移动到特异性位置(即控制其移动)来修改聚核苷酸。

[0543] 聚核苷酸结合蛋白优选地衍生自聚核苷酸操作酶。一或多个分子制动器可衍生自上文所论述的聚核苷酸操作酶中的任一个。Phi29聚合酶(SEQ ID NO:8)的充当分子制动器的经修饰型式公开于美国专利第5,576,204号中。一或多个分子制动器优选地是衍生自解螺旋酶。

[0544] 可使用任何数量的衍生自解螺旋酶的分子制动器。举例来说,1、2、3、4、5、6、7、8、9、10个或更多个解螺旋酶可用作分子制动器。如果两个或更多个解螺旋酶用作分子制动器,那么所述两个或更多个解螺旋酶通常是相同的解螺旋酶。两个或更多个解螺旋酶可以是不同的解螺旋酶。

[0545] 两个或更多个解螺旋酶可以是上文提到的解螺旋酶的任何组合。两个或更多个解螺旋酶可以是两个或更多个Dda解螺旋酶。两个或更多个解螺旋酶可以是一或多个Dda解螺旋酶和一或多个TrwC解螺旋酶。两个或更多个解螺旋酶可以是相同解螺旋酶的不同变异体。

[0546] 两个或更多个解螺旋酶优选地彼此连接。两个或更多个解螺旋酶更优选地彼此共价连接。解螺旋酶可以按任何次序和使用任何方法来连接。衍生自解螺旋酶的一或多个分子制动器优选地进行修饰,以降低聚核苷酸结合域中的开口的大小,通过所述开口,在至少一个构形状态中,聚核苷酸可与解螺旋酶解结合。这公开于W0 2014/013260中。

[0547] 用于本发明中的优选解螺旋酶构筑体描述于以下国际申请中：第PCT/GB2013/051925号(公开为W0 2014/013260)、第PCT/GB2013/051924号(公开为W0 2014/013259)、第PCT/GB2013/051928号(公开为W0 2014/013262)和第PCT/GB2014/052736号。

[0548] 如果以主动模式使用一或多个解螺旋酶(即当一或多个解螺旋酶具备便于移动的所有必需组分(例如ATP和 Mg^{2+})时),那么一或多个分子制动器优选地是:(a)以非活动模式使用(即在不在于便于移动的必需组分下使用,或不能主动移动), (b)以主动模式使用,其中一或多个分子制动器在与一或多个解螺旋酶相反的方向上移动,或(c)以主动模式使用,其中一或多个分子制动器在与一或多个解螺旋酶相同的方向上移动,且比一或多个解螺旋酶更缓慢。

[0549] 如果以非活动模式使用一或多个解螺旋酶(即当一或多个解螺旋酶不具备便于移动的所有必需组分(例如ATP和 Mg^{2+})时,或不能主动移动),那么一或多个分子制动器优选地是:(a)以非活动模式使用(即在不在于便于移动的必需组分下使用,或不能主动移动),或(b)以主动模式使用,其中一或多个分子制动器在与聚核苷酸相同的方向上沿聚核苷酸移动通过孔。

[0550] 一或多个解螺旋酶和一或多个分子制动器可在任何位置处连接于聚核苷酸,使得其带到一起且均控制聚核苷酸通过孔的移动。一或多个解螺旋酶和一或多个分子制动器相隔至少一个核苷酸,如相隔至少5、至少10、至少50、至少100、至少500、至少1000、至少5000、至少10,000、至少50,000个核苷酸或更多。如果方法涉及表征在一端具备Y衔接子且在另一端具备发夹环衔接子的双链聚核苷酸,那么一或多个解螺旋酶优选地连接于Y衔接子,且一或多个分子制动器优选地连接于发夹环衔接子。在这个实施例中,一或多个分子制动器优选地是进行修饰使得其结合聚核苷酸但不充当解螺旋酶的一或多个解螺旋酶。连接于Y衔接子的一或多个解螺旋酶优选地在间隔子处停止,如下文更详细地论述。连接于发夹环衔接子的一或多个分子制动器优选地不在间隔子处停止。当一或多个解螺旋酶达到发夹环时,一或多个解螺旋酶和一或多个分子制动器优选地带到一起。在Y衔接子连接于聚核苷酸之前或在Y衔接子连接于聚核苷酸之后,一或多个解螺旋酶可连接于Y衔接子。在发夹环衔接子连接于聚核苷酸之前或在发夹环衔接子连接于聚核苷酸之后,一或多个分子制动器可连接于发夹环衔接子。

[0551] 一或多个解螺旋酶和一或多个分子制动器优选地不彼此连接。一或多个解螺旋酶和一或多个分子制动器更优选地不彼此共价连接。一或多个解螺旋酶和一或多个分子制动器优选地不连接,如以下国际申请中所描述:第PCT/GB2013/051925号(公开为W02014/013260)、第PCT/GB2013/051924号(公开为W0 2014/013259)、第PCT/GB2013/051928号(公开为W0 2014/013262)和第PCT/GB2014/052736号。

[0552] 间隔子

[0553] 一或多个解螺旋酶可在一个或多个间隔子处停止,如在国际申请第PCT/GB2014/050175号中所论述。在本发明中可使用在国际申请中所公开的一或多个解螺旋酶和一个或多个间隔子的任何构形。

[0554] 当聚核苷酸的一部分进入孔且沿由施加的电势造成的场移动通过孔时,随着聚核苷酸移动通过孔,一或多个解螺旋酶通过孔而移过间隔子。这是因为聚核苷酸(包括一个或多个间隔子)移动通过孔且一或多个解螺旋酶残留在孔上。

[0555] 一或多个间隔子优选地是聚核苷酸的一部分,例如其中断聚核苷酸序列。一或多个间隔子优选地不是与聚核苷酸杂交的一或多个阻断分子(如减速带)的一部分。

[0556] 在聚核苷酸中可存在任何数量的间隔子,如1、2、3、4、5、6、7、8、9、10个或更多个间隔子。在聚核苷酸中优选地存在两个、四个或六个间隔子。在聚核苷酸的不同区中可存在一或多个间隔子,如Y衔接子和/或发夹环衔接子中的一或多个间隔子。

[0557] 一或多个间隔子各自提供能量势垒,一或多个解螺旋酶甚至在主动模式中不能克服所述能量势垒。一或多个间隔子可通过降低解螺旋酶的牵引(例如通过从聚核苷酸中的核苷酸移除碱基)或物理地阻断一或多个解螺旋酶的移动(例如使用大型化学基团)来使一或多个解螺旋酶停止。

[0558] 一或多个间隔子可包含使一或多个解螺旋酶停止的任何分子或所述分子的组合。一或多个间隔子可包含防止一或多个解螺旋酶沿聚核苷酸移动的任何分子或所述分子的组合。测定一或多个解螺旋酶是否在不存在跨膜孔和所施加的电势下在一或多个间隔子处停止是直接了当的。举例来说,可通过PAGE来测量解螺旋酶移动通过间隔子和使DNA互补链移位的能力。

[0559] 一或多个间隔子通常包含线性分子,如聚合物。一或多个间隔子通常具有与聚核苷酸不同的结构。举例来说,如果聚核苷酸是DNA,那么一或多个间隔子通常不是DNA。特定来说,如果聚核苷酸是脱氧核糖核酸(DNA)或核糖核酸(RNA),那么一或多个间隔子优选地包含肽核酸(PNA)、甘油核酸(GNA)、苏糖核酸(TNA)、锁核酸(LNA)或与核苷酸侧链的合成聚合物。一或多个间隔子可包含在从聚核苷酸的相反的方向上的一或多个核苷酸。举例来说,当聚核苷酸是在5'到3'方向上时,一或多个间隔子在3'到5'方向上可包含一或多个核苷酸。核苷酸可以是上文所论述的那些核苷酸中的任一个。

[0560] 一或多个间隔子优选地包含:一或多个硝基吡啶,如一或多个5-硝基吡啶;一或多个肌苷;一或多个吡啶;一或多个2-氨基嘌呤;一或多个2-6-二氨基嘌呤;一或多个5-溴-脱氧尿苷;一或多个反向胸苷(反向dT);一或多个反向双脱氧胸苷(ddT);一或多个双脱氧胞苷(ddC);一或多个5-甲基胞苷;一或多个5-羟甲基胞苷;一或多个2'-O-甲基RNA碱基;一或多个异脱氧胞苷(Iso-dC);一或多个异脱氧鸟苷(Iso-dG);一或多个iSpC3基团(即缺乏糖和碱基的核苷酸);一或多个可光分解(PC)基团;一或多个基团;一或多个间隔子9(iSp9)基团;一或多个间隔子18(iSp18)基团;聚合物或一或多个巯基连接。一或多个间隔子可包含这些基团的任何组合。许多这些基团可购自IDT®(Integrated DNA Technologies®)。

[0561] 一或多个间隔子可含有任意数量的这些基团。举例来说,对于2-氨基嘌呤、2-6-二氨基嘌呤、5-溴-脱氧尿苷、反向dT、ddT、ddC、5-甲基胞苷、5-羟甲基胞苷、2'-O-甲基RNA碱基、Iso-dC、Iso-dG、iSpC3基团、PC基团、己二醇基团和巯基连接,一或多个间隔子优选地包含2、3、4、5、6、7、8、9、10、11、12个或更多个。一或多个间隔子优选地包含2、3、4、5、6、7、8个或更多个iSp9基团。一或多个间隔子优选地包含2、3、4、5或6个或更多个iSp18基团。最优选间隔子是四个iSp18基团。

[0562] 聚合物优选地是多肽或聚乙二醇(PEG)。多肽优选地包含2、3、4、5、6、7、8、9、10、11、12个或更多个氨基酸。PEG优选地包含2、3、4、5、6、7、8、9、10、11、12个或更多个单体单元。

[0563] 一或多个间隔子优选地包含一或多个无碱基核苷酸(即缺乏核碱基的核苷酸),如

2、3、4、5、6、7、8、9、10、11、12个或更多个无碱基核苷酸。在无碱基核苷酸中,核碱基可被-H(idSp)或-OH置换。可通过从一或多个相邻核苷酸移除核碱基,将无碱基间隔子插入到聚核苷酸中。举例来说,聚核苷酸可进行修饰,以包括3-甲基腺嘌呤、7-甲基鸟嘌呤、1,N6-亚乙烯基腺嘌呤肌苷或次黄嘌呤,且可使用人类烷基腺嘌呤DNA糖基化酶(hAAG)来从这些核苷酸移除核碱基。替代地,聚核苷酸可进行修饰,以包括尿嘧啶且用尿嘧啶-DNA糖基化酶(UDG)移除核碱基。在一个实施例中,一或多个间隔子不包含任何无碱基核苷酸。

[0564] 一或多个解螺旋酶可通过(即之前)或在各线性分子间隔子上停止。如果使用线性分子间隔子,那么聚核苷酸优选地具备与一或多个解螺旋酶待移动的各间隔子的末端相邻的聚核苷酸的双链区。双链区通常有助于使一或多个解螺旋酶在相邻间隔子上停止。如果在约100mM或更低的盐浓度下进行方法,那么双链区的存在是特别优选的。各双链区的长度通常是至少10个,如至少12个核苷酸。如果在本发明中所用聚核苷酸是单链,那么可通过将较短聚核苷酸与同间隔子相邻的区杂交来形成双链区。较短聚核苷酸通常由与聚核苷酸相同的核苷酸来形成,但可由不同核苷酸来形成。举例来说,较短聚核苷酸可由LNA来形成。

[0565] 如果使用线性分子间隔子,那么聚核苷酸优选地在与一或多个解螺旋酶待移动的各间隔子的末端相反的各间隔子的末端具备阻断分子。这可有助于确保一或多个解螺旋酶仍在各间隔子上停止。在其于溶液中扩散开的情况下,其还可有助于将一或多个解螺旋酶保持在聚核苷酸上。阻断分子可以是下文论述、物理地引起一或多个解螺旋酶停止的任何化学基团。阻断分子可以是聚核苷酸的双链区。

[0566] 一或多个间隔子优选地包含物理地引起一或多个解螺旋酶停止的一或多个化学基团。一或多个化学基团优选地是一或多个侧链(pendant)化学基团。一或多个化学基团可连接于聚核苷酸中的一或多个核碱基。一或多个化学基团可连接于聚核苷酸主链。可存在任何数量的这些化学基团,如2、3、4、5、6、7、8、9、10、11、12个或更多。合适基团包括(但不限于)荧光团、抗生蛋白链菌素和/或生物素、胆固醇、亚甲基蓝、二硝基苯酚(DNP)、地高辛和/或抗地高辛和二苯基环辛炔基团。

[0567] 聚核苷酸中的不同间隔子可包含不同停止分子。举例来说,一个间隔子可包含上文所论述的一个线性分子,且另一间隔子可包含物理地引起一或多个解螺旋酶停止的一或多个化学基团。间隔子可包含上文所论述中的任一个线性分子和物理地引起一或多个解螺旋酶停止的一或多个化学基团(如一或多个无碱基基团和荧光团)。

[0568] 可取决于聚核苷酸的类型和进行本发明的方法的条件,来设计合适间隔子。大部分解螺旋酶结合DNA且沿DNA移动,且因此可使用不是DNA的任何东西来使其停止。合适分子在上文论述。

[0569] 优选地在存在游离核苷酸和/或存在解螺旋酶辅因子下来进行本发明的方法。这一点在下文更详细论述。在不存在跨膜孔和所施加的电势下,一或多个间隔子优选地能够在存在游离核苷酸和/或存在解螺旋酶辅因子下,使一或多个解螺旋酶停止。

[0570] 如果如下文所论述在存在游离核苷酸和解螺旋酶辅因子下进行本发明的方法(使得更多解螺旋酶中的一个是在主动模式中),那么一或多个较长间隔子通常用于确保在使一或多个解螺旋酶与跨膜孔接触和施加电势之前,所述一或多个解螺旋酶在聚核苷酸上停止。可在不存在游离核苷酸和解螺旋酶辅因子下使用一或多个较短间隔子(使得一或多个解螺旋酶是在非活动模式中)。

[0571] 盐浓度还影响一或多个间隔子使一或多个解螺旋酶停止的能力。在不存在跨膜孔施加的电势下,一或多个间隔子优选地能够在约100mM或更低的盐浓度下使一或多个解螺旋酶停止。在本发明的方法中所用的盐浓度越高,通常使用的一或多个间隔子越短,且反之亦然。

[0572] 优选的特征组合显示于以下表4中。

[0573]	聚核苷酸	间隔子组合物*	间隔子长度(即*的数量)	盐[]	游离核苷酸?	解螺旋酶辅因子?
	DNA	iSpC3	4	1 M	是	是
	DNA	iSp18	4	100-1000 mM	是	是
	DNA	iSp18	6	<100-1000 mM	是	是
	DNA	iSp18	2	1 M	是	是
	DNA	iSpC3	12	<100-1000 mM	是	是
	DNA	iSpC3	20	<100-1000 mM	是	是
	DNA	iSp9	6	100-1000 mM	是	是
	DNA	idSp	4	1 M	是	是

[0574] 表4

[0575] 方法可涉及移动两个或更多个解螺旋酶通过间隔子。在此类情况下,通常增加间隔子的长度,以防止在不存在孔和施加的电势下后置解螺旋酶牵拉前导解螺旋酶通过间隔子。如果方法涉及移动两个或更多个解螺旋酶通过一或多个间隔子,那么可增加上文所论述的间隔子长度至少1.5倍,如2倍、2.5倍或3倍。举例来说,如果方法涉及移动两个或更多个解螺旋酶通过一或多个间隔子,那么可增加以上表4第三列中的间隔子长度1.5倍、2倍、2.5倍或3倍。

[0576] 膜

[0577] 本发明的孔可存在于膜中。在本发明的方法中,通常使聚核苷酸与膜中的本发明的孔接触。根据本发明可以使用任何膜。合适膜在所属领域中是众所周知的。膜优选地是两亲层。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特性两种。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各单体子单元提供的特性。然而,嵌段共聚物可具有由个别子单元形成的聚合物不拥有的独特特性。嵌段共聚物可进行工程改造,使得单体子单元中的一个在水性介质中是疏水性的(即亲脂性),而其它子单元是亲水性的。在此情况下,嵌段共聚物可拥有两亲特性,且可形成模拟生物膜的结构。嵌段共聚物可以是二嵌段的(其由两个单体子单元组成),但也可由超过两个的单体子单元来构筑,形成表现为两亲物的更复杂的排列。共聚物可以是三嵌段、四嵌段或五嵌段共聚物。膜优选地是三嵌段共聚物膜。

[0578] 古细菌双极四醚脂质是进行构筑使得脂质形成单层膜的天然存在的脂质。这些脂质一般发现于存活于苛刻生物环境中的极端条件的生物、嗜热菌、嗜盐菌和嗜酸菌中。认为其稳定性是源于最终双层的融合性质。通过产生具有一般基序亲水性-疏水性-亲水性的三

嵌段聚合物来构筑模拟这些生物实体的嵌段共聚物材料,是直接了当的。这种材料可形成表现类似于脂质双层且涵盖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在生物脂质膜上保持若干优势。因为合成三嵌段共聚物,所以可小心地控制准确的构筑,以提供形成膜和与孔和其它蛋白质相互作用所需的正确链长度和特性。

[0579] 还可由不分类为脂质亚材料的子单元来构筑嵌段共聚物,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小节还可拥有很低的蛋白质结合特性,此允许产生当暴露于原始生物样本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。

[0580] 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。

[0581] 膜最优选地是国际申请第PCT/GB2013/052766号或第PCT/GB2013/052767号中所公开的膜中的一个。

[0582] 两亲分子可进行化学修饰或官能化,以便于偶联聚核苷酸。

[0583] 两亲层可以是单层或双层。两亲层通常是平坦的。两亲层可以是弯曲的。两亲层可以是支撑式的。

[0584] 两亲膜通常天然地是流动的,基本上以大致 $10^{-8} \text{ cm s}^{-1}$ 的脂质扩散速率充当二维液体。这意味着孔和偶联的聚核苷酸可通常在两亲膜内移动。

[0585] 膜可以是脂质双层。脂质双层是细胞膜的模型,且用作一系列实验研究的极佳平台。举例来说,脂质双层可用于通过单通道记录对膜蛋白的活体外研究。替代地,脂质双层可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。合适脂质双层包括(但不限于)平坦脂质双层、支撑式双层或脂质体。脂质双层优选地是平坦脂质双层。合适脂质双层公开于以下中:国际申请第PCT/GB08/000563号(公开为W02008/102121)、国际申请第PCT/GB08/004127号(公开为W0 2009/077734)和国际申请第PCT/GB2006/001057号(公开为W0 2006/100484)。

[0586] 用于形成脂质双层的方法在所属领域中是已知的。脂质双层通常通过蒙塔尔和米勒的方法(《美国国家科学院院报(Proc.Natl.Acad.Sci.USA.)》,1972;69:3561-3566)来形成,其中脂质单层携载于通过开孔两侧的水溶液/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,且接着使在开孔两侧上的水溶液的表面上蒸发一滴溶剂,来将脂质添加到水性电解质溶液的表面。一旦有机溶剂已蒸发,那么开孔两侧上的溶液/空气界面来回物理地移动通过开孔,直到形成双层为止。可跨越膜中的开孔或跨越凹槽中的开口形成平坦脂质双层。

[0587] 蒙塔尔和米勒的方法是常用的,这是因为是节约成本的,且是形成适合于蛋白孔插入的良好品质脂质双层的相对直接了当的方法。双层形成的其它常见方法包括脂质体双层的尖端浸没、双层涂刷和贴片夹持。

[0588] 尖端浸没双层形成需要使开孔表面(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,通过将溶解于有机溶剂中的一滴脂质在溶液表面处蒸发来首先在溶液/空气界面处产生脂质单层。接着,通过朗缪尔-沙佛(Langmuir-Schaefer)过程形成双层,且需要机械自动以使开孔相对于溶液表面移动。

[0589] 对于涂刷的双层,将溶解于有机溶剂中的一滴脂质直接应用于开孔,所述开孔浸没在水性测试溶液中。使用笔刷或等效物,使脂质溶液稀薄地扩散在开孔内。溶剂的薄化使得形成脂质双层。然而,从双层完全移除溶剂是非常困难的,且因此通过这种方法形成的双层较不稳定且更倾向于在电化学测量期间具有噪声。

[0590] 贴片夹持是在生物细胞膜研究中常用的。通过抽汲将细胞膜夹持到移液管的末端,且膜贴片变为连接在开孔内。所述方法适用于通过夹持接着爆裂以离开密封在移液管的开孔内的脂质双层的脂质体来产生脂质双层。所述方法需要稳定的、巨大的且单层脂质体和具有玻璃表面的材料中制造小开孔。

[0591] 可通过超声处理、挤压或Mozafari方法来形成脂质体(Colas等人(2007)《Micron》38:841-847)。

[0592] 在一优选实施例中,如国际申请第PCT/GB08/004127号(公开为W0 2009/077734)中所描述形成脂质双层。在此方法中有利的是,由干燥脂质形成脂质双层。在一最优选实施例中,跨越开口形成脂质双层,如W02009/077734(PCT/GB08/004127)中所描述。

[0593] 由脂质的两个相对层形成脂质双层。两个脂质层被布置成使得其疏水尾部基团面朝彼此,形成疏水性的内部。脂质的亲水性头基朝外面向双层各侧上的水性环境。双层可存在于多种脂质阶段中,所述阶段包括(但不限于)液体无序阶段(液体片层)、液体有序阶段、固体有序阶段(片层凝胶阶段、交错结合的凝胶阶段)和平坦双层晶体(片层亚凝胶阶段、片层结晶阶段)。

[0594] 可使用形成脂质双层的任何脂质组合物。选择脂质组合物,使得脂质双层具有所需的特性,如表面电荷、支持膜蛋白的能力、充填密度或所形成的机械特性。脂质组合物可包含一或多种不同脂质。举例来说,脂质组合物可含有至多100种脂质。脂质组合物优选地含有1到10种脂质。脂质组合物可包含天然存在的脂质和/或人工脂质。

[0595] 脂质通常包含头基、界面部分和可相同或不同的两个疏水尾部基团。合适头基包括(但不限于):中性头基,如二酰基甘油酯(DG)和脑酰胺(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);和带正电头基,如三甲基铵丙烷(TAP)。合适界面部分包括(但不限于)天然存在的界面部分,如基于甘油或基于脑酰胺的部分。合适疏水尾部基团包括(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);和分支链烃链,如植烷酰基。链的长度和不饱和烃链中的双键的位置与数量可变化。链的长度和分支链烃链中的分支(如甲基)的位置和数量可变化。疏水尾部基团可作为醚或酯连接于界面部分。脂质可以是分枝菌酸。

[0596] 脂质还可以进行化学修饰。脂质的头基或尾部基团可以进行化学修饰。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];官能化的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];和针对结合修饰的脂质,如1,2-二油酰基-sn-甘油-3-磷酸乙醇胺-N-(丁二酰基)和1,2-二棕榈酰基-sn-甘油-3-磷酸乙醇胺-N-(生物素基)。尾部基团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(10,12-三辅二炔基)-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;甾化脂质,如1,2-二棕榈酰基-D62-sn-甘油-3-磷酸胆碱;和醚连接的脂质,如1,2-二-O-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,以便于偶联聚核苷酸。

[0597] 两亲层,例如脂质组合物,通常包含将影响层的特性的一或多个添加剂。合适添加剂包括(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;固醇,如胆固醇、麦角固醇、羊毛固醇、谷固醇和豆固醇;溶血磷脂,如1-酰基-2-羟基-sn-甘油-3-磷酸胆碱;和脑酰胺。

[0598] 在另一优选实施例中,膜包含固态层。固态层可由两个有机和无机材料形成,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_2 ;有机和无机聚合物,如聚酰胺;塑料,如Teflon®,或弹性体,如二组分加成固化的聚硅氧橡胶;和玻璃。固态层可由石墨烯形成。合适石墨烯层公开于国际申请第PCT/US2008/010637号(公开为W0 2009/035647)中。如果膜包含固态层,那么孔通常存在于两亲膜或层中,所述两亲膜或层含于固态层内,例如在固态层内的孔洞、孔、空隙、通道、沟槽或缝隙内。所属领域的技术人员可制备合适的固态/两亲性杂合系统。合适系统公开于W02009/020682和W0 2012/005857中。可使用上文所论述的两亲膜或层中的任一个。

[0599] 通常使用以下来进行方法:(i)人工两亲层,其包含孔,(ii)分离包含孔的天然存在的脂质双层,或(iii)其中插入有孔的细胞。通常使用人工两亲层,如人工三嵌段共聚物层,来进行方法。层可包含其它跨膜和/或膜内蛋白质以及除孔以外的其它分子。下文论述合适设备和条件。通常在活体外进行本发明的方法。

[0600] 偶联

[0601] 聚核苷酸优选地与包含孔的膜偶联。方法可包含将聚核苷酸与包含孔的膜偶联。使用一或多个锚,聚核苷酸优选地与膜偶联。使用任何已知的方法,聚核苷酸可与膜偶联。

[0602] 各锚包含与聚核苷酸偶联(或结合)的基团和与膜偶联(或结合)的基团。各锚可与聚核苷酸和/或膜共价偶联(或结合)。如果使用Y衔接子和/或发夹环衔接子,那么使用衔接子,聚核苷酸优选地与膜偶联。

[0603] 使用任何数量的锚,如2、3、4个或更多个锚,聚核苷酸可与膜偶联。举例来说,使用两个锚,聚核苷酸可与膜偶联,所述两个锚中的每一个分别地与聚核苷酸和膜两个偶联(或结合)。

[0604] 一或多个锚可包含一或多个解螺旋酶和/或一或多个分子制动器。

[0605] 如果膜是两亲层,如共聚物膜或脂质双层,那么一或多个锚优选地包含存在于膜中的多肽锚和/或存在于膜中的疏水性锚。疏水性锚优选地是脂质、脂肪酸、固醇、碳纳米管、多肽、蛋白质或氨基酸,例如胆固醇、棕榈酸或生育酚。在优选实施例中,一或多个锚不是孔。

[0606] 膜的组分,如两亲分子、共聚物或脂质,可进行化学修饰或官能化,以形成一或多个锚。膜的组分的合适化学修饰和合适官能化方式的实例在下文更详细地论述。可对任何比例的膜组分进行官能化,例如至少0.01%、至少0.1%、至少1%、至少10%、至少25%、至少50%或100%。

[0607] 聚核苷酸可直接与膜偶联。用于将聚核苷酸与膜偶联的一或多个锚优选地包含连接子。一或多个锚可包含一或多个连接子,如2、3、4个或更多个。可使用一个连接子来将超

过一个聚核苷酸与膜偶联,如2、3、4个或更多个。

[0608] 优选连接子包括(但不限于)聚合物,如聚核苷酸、聚乙二醇(PEG)、多糖和多肽。这些连接子可以是线性、分支链或环状的。举例来说,连接子可以是环状聚核苷酸。聚核苷酸可与环状聚核苷酸连接子上的互补序列杂交。

[0609] 一或多个锚或一或多个连接子可包含可进行剪切分解的组分,如限制位点或光不稳定性基团。

[0610] 官能化连接子和其可偶联分子的方式在所属领域中是已知的。举例来说,用马来酰亚胺基团官能化的连接子将与蛋白质中的半胱氨酸残基反应且连接。在本发明的上下文中,蛋白质可存在于膜中,或可用于与聚核苷酸偶联(或结合)。这一点在下文更详细论述。

[0611] 可使用“锁和钥”排列来避免聚核苷酸的交联。各连接子的仅一端可一起反应形成较长连接子,且连接子的另一端各自分别与聚核苷酸或膜反应。此类连接子描述于国际申请第PCT/GB10/000132号(公开为W0 2010/086602)中。

[0612] 在下文论述的测序实施例中,使用连接子是优选的。如果在当与孔相互作用时聚核苷酸并不解耦(即在步骤(b)或(e)中不解耦)的意义上,聚核苷酸永久性直接与膜偶联,那么在归因于膜与孔之间的距离而测序操作不能持续到聚核苷酸的末端时,将丢失一些序列数据。如果使用连接子,那么聚核苷酸可进行处理完全。

[0613] 偶联可以是永久性或稳定的。换句话说,偶联可以使得当聚核苷酸与孔相互作用时其仍与膜偶联。

[0614] 偶联可以是暂时的。换句话说,偶联可以使得当聚核苷酸与孔相互作用时其可与膜解耦。

[0615] 对于某些应用,如适体检测,暂时性质的偶联是优选的。如果永久性或稳定连接子直接连接于聚核苷酸的5'或3'端,且连接子比膜与跨膜孔的通道之间的距离短,那么在测序操作不能持续到聚核苷酸的末端时,一些序列数据将丢失。如果偶联是暂时的,接着当偶联的末端随机变为不含膜时,那么聚核苷酸可以进行处理完全。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。使用胆固醇或脂肪酰基链,聚核苷酸可以与两亲层或三嵌段共聚物膜暂时偶联。可使用长度为6到30个碳原子的任何脂肪酰基链,如十六烷酸。

[0616] 在优选实施例中,聚核苷酸(如核酸)与两亲层(如三嵌段共聚物膜或脂质双层)偶联。先前已用各种不同网络共享的策略进行核酸与合成脂质双层的偶联。这些概述于下表5中。

[0617] 表5

	锚包含	偶联类型	参考文献
[0618]	巯基	稳定的	Yoshina-Ishii, C.和 S. G. Boxer (2003). “《支撑式脂质双层上的行动拴系囊泡的阵列 (Arrays of mobile tethered vesicles on supported lipid bilayers)》”《美国化学学会杂志》 125 (13): 3696-7。
	生物素	稳定的	Nikolov, V., R.Lipowsky 等人(2007). “《具有锚定 DNA 分子的巨大囊泡的行为 (Behavior of giant vesicles with anchored DNA molecules)》”. 《生物物理杂志 (Biophys J)》 92 (12): 4356-68
	胆固醇	暂时的	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)》”. 《美国化学学会杂志》 126 (33): 10224-5
	表面活性剂 (例如脂质、棕榈酸等)	稳定的	van Lengerich, B., R. J. Rawle 等人 “《脂质囊泡与液体支撑的双层的共价连接允许观测 DNA 介导的囊泡相互作用 (Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions)》”. 《朗缪尔》 26 (11): 8666-72

[0619] 合成聚核苷酸和/或连接子可在合成反应中使用修饰胺基磷酸酯来官能化,所述修饰胺基磷酸酯容易地与直接添加合适锚定基团相容,所述锚定基团如,胆固醇、生育酚、棕榈酸、巯基、脂质和生物素基团。这些不同的连接化学,产生一套用于与聚核苷酸连接的选项。各不同修饰基团以稍微不同的方式偶联聚核苷酸,且偶联未必总是永久性的,因此对于聚核苷酸与膜,得到不同的停留时间。暂时偶联的优势在上文论述。

[0620] 聚核苷酸与连接子或与官能化膜的偶联还可可通过多种其它手段来实现,其限制条件为可向聚核苷酸中添加互补反应性基团或锚定基团。向聚核苷酸的任一末端添加反应性基团,先前已有报导。可使用T4聚核苷酸激酶和ATP γ S向ssDNA或dsDNA的5'中添加巯基(Grant, G.P.和P.Z. Qin(2007). “《一种用于连接在核酸的5'端处的氮氧化物自旋标记的便捷方法(A facile method for attaching nitroxide spin labels at the 5' terminus of nucleic acids)》”.《核酸研究》**35**(10): e77)。可使用T4聚核苷酸激酶和 γ -[2-叠氮基乙基]-ATP或 γ -[6-叠氮基己基]-ATP,向ssDNA或dsDNA的5'-磷酸中添加叠氮基团。使用巯基或点击化学,含有巯基、碘乙酰胺OPSS或马来酰亚胺基团(与巯基具有反应性)或DIBO(二苯并环氧磷)或炔基(与叠氮化物具有反应性)中的任一个的系链,可共价连接于聚核苷酸。可使用末端转移酶将修饰寡核苷酸并入到ssDNA的3'中,来添加更多样的化学基团选择(如生物素、巯基和荧光团)(Kumar, A., P. Tchen等人(1988). “《通过末端脱氧核苷酸转移酶对合成寡核苷酸探针进行非放射性标记(Nonradioactive labeling of synthetic oligonucleotide probes with terminal deoxynucleotidyl transferase)》”.《分析生物化学》**169**(2): 376-82)。抗生蛋白链菌素/生物素和/或抗生蛋白链菌素/脱硫生物素偶联可用于任何其它聚核苷酸。以下实例描述聚核苷酸可如何使用抗生蛋白链菌素/生物素和

抗生蛋白链菌素/脱硫生物素来与膜偶联。也有可能,可使用末端转移酶(例如胆固醇或棕榈酸)向具有合适修饰的核苷酸的聚核苷酸中直接添加锚。

[0621] 一或多个锚优选地通过杂交将聚核苷酸与膜偶联。一或多个锚中的杂交允许以如上文所论述的暂时方式偶联。杂交可存在于一或多个锚的任何部分中,如在一或多个锚与聚核苷酸之间,在一或多个锚内或在一或多个锚与膜之间。举例来说,连接子可包含两个或更多个杂交在一起的聚核苷酸,如3、4或5个聚核苷酸。一或多个锚可与聚核苷酸杂交。一或多个锚可直接与聚核苷酸或直接与Y衔接子和/或连接于聚核苷酸的前导序列或直接与连接于聚核苷酸的发夹环衔接子杂交(如下文所论述)。替代地,一或多个锚可与一或多个(如2或3个)中间聚核苷酸(或“夹板”)杂交,所述中间聚核苷酸与聚核苷酸、与连接于聚核苷酸的Y衔接子和/或前导序列或与连接于聚核苷酸的发夹环衔接子杂交(如下文所论述)。

[0622] 一或多个锚可包含单链或双链聚核苷酸。锚的一个部分可与单链或双链聚核苷酸接合。已报导使用T4 RNA接合酶I对ssDNA的短碎片的接合(Troutt, A.B., M.G. McHeyzer-Williams等人(1992). “《接合锚定的PCR: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》89(20):9823-5)。替代地,单链或双链聚核苷酸可与双链聚核苷酸接合,且接着两个链通过热或化学变性分开。对于双链聚核苷酸,有可能,向双螺旋的一个或两个端中添加一片单链聚核苷酸,或向一个或两个端中添加双链聚核苷酸。对于向双链聚核苷酸中添加单链聚核苷酸,这可使用T4 RNA接合酶I来实现,如与单链聚核苷酸的其它区的接合。对于向双链聚核苷酸中添加双链聚核苷酸,那么分别通过聚核苷酸上的互补3'dA/dT尾和所添加聚核苷酸(入对于许多样本制备型应用常规地进行,以防止多联体或二聚体形成),或使用通过限制性消化聚核苷酸和相容衔接子的接合而产生的“粘稠端”,接合可以是“平末端”的。接着,如果单链聚核苷酸用于在5'端、3'端或两端(当双链聚核苷酸用于接合时)处的接合或修饰,那么当双螺旋熔融时,各单链将具有5'或3'修饰。

[0623] 如果聚核苷酸是合成链,那么可在化学合成聚核苷酸期间并入一或多个锚。举例来说,可使用具有连接于其的反应性基团的引物来合成聚核苷酸。

[0624] 腺苷酸化聚核苷酸是接合反应中的中间物,其中单磷酸腺苷连接于聚核苷酸的5'-磷酸。用于产生这种中间物的各种试剂盒是可获得的,如来自NEB的5'DNA腺苷酰化试剂盒。通过在反应中对于修饰核苷酸三磷酸进行取代ATP,那么可以向聚核苷酸的5'中添加反应性基团(如巯基、胺、生物素、叠氮化物等)。也有可能,可使用5'DNA腺苷酰化试剂盒向具有合适修饰的核苷酸的聚核苷酸中直接添加锚。

[0625] 用于扩增基因组DNA片段的常见技术是使用聚合酶链式反应(PCR)。此处,使用两个合成寡核苷酸引物,可产生大量相同DNA片段的拷贝,其中对于各拷贝,双螺旋中的各链的5'将是合成聚核苷酸。可通过使用聚合酶向单或双链DNA的3'端中添加单个或多个核苷酸。可使用的聚合酶的实例包括(但不限于)末端转移酶、Klenow和大肠杆菌Poly(A)聚合酶)。通过在反应中对于修饰核苷酸三磷酸取代ATP,那么可将锚(如胆固醇、巯基、胺、叠氮化物、生物素或脂质)并入到双链聚核苷酸中。因此,所扩增聚核苷酸的各拷贝将含有锚。

[0626] 理想地,聚核苷酸与膜偶联而无须使聚核苷酸官能化。这可通过将一或多个锚(如聚核苷酸结合蛋白或化学基团)与膜偶联和使一或多个锚与聚核苷酸相互作用或通过使膜

官能化来实现。一或多个锚可通过本文所描述的任一个方法来与膜偶联。特定来说,一或多个锚可包含一或多个连接子,如马来酰亚胺官能化连接子。

[0627] 在这个实施例中,聚核苷酸通常是RNA、DNA、PNA、TNA或LNA,且可以是双或单链。这个实施例尤其适合于基因组DNA聚核苷酸。

[0628] 一或多个锚可包含,与单或双链聚核苷酸、聚核苷酸内的特异性核苷酸序列或聚核苷酸内的修饰核苷酸的图案或存在于聚核苷酸上的任何其它配体偶联、结合或与相互作用的任何基团。

[0629] 用于锚中的合适结合蛋白包括(但不限于):大肠杆菌单链结合蛋白、P5单链结合蛋白、T4 gp32单链结合蛋白、TOPO V dsDNA结合区、人类组蛋白、大肠杆菌HU DNA结合蛋白,和其它古细菌、原核或真核单链或双链聚核苷酸(或核酸)结合蛋白,包括下文所列的那些。

[0630] 特异性核苷酸序列可以由转录因子、核糖体、核酸内切酶、拓扑异构酶或复制起始因子识别的序列。修饰核苷酸的图案可以是甲基化或损坏图案。

[0631] 一或多个锚可包含与聚核苷酸偶联、结合、插入或相互作用的任何基团。基团可通过静电、氢结合或范德华力相互作用插入聚核苷酸中或与其相互作用。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)、通用碱基或通用核苷酸(其可与任何聚核苷酸杂交)和钬络合物(其可与甲基化碱基反应)。因此,聚核苷酸可使用连接于膜的一或多个通用核苷酸而与膜偶联。各通用核苷酸可使用一或多个连接子而与膜偶联。通用核苷酸优选地包含以下核碱基中的一个:次黄嘌呤、4-硝基吡啶、5-硝基吡啶、6-硝基吡啶、甲酰基吡啶、3-硝基吡咯、硝基咪唑、4-硝基吡唑、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)。通用核苷酸更优选地包含以下核苷中的一个: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、2-0'-甲基肌苷、4-硝基吡啶2'-脱氧核苷、4-硝基吡啶核苷、5-硝基吡啶2'-脱氧核苷、5-硝基吡啶核苷、6-硝基吡啶2'-脱氧核苷、6-硝基吡啶核苷、3-硝基吡咯2'-脱氧核苷、3-硝基吡咯核苷、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-硝基吡唑2'-脱氧核苷、4-硝基吡唑核苷、4-硝基苯并咪唑2'-脱氧核苷、4-硝基苯并咪唑核苷、5-硝基咪唑2'-脱氧核苷、5-硝基咪唑核苷、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'-脱氧烟云杯伞素、2'-脱氧异鸟苷、K-2'-脱氧核苷、P-2'-脱氧核苷和吡咯烷。通用核苷酸更优选地包含2'-脱氧肌苷。通用核苷酸更优选地是IMP或dIMP。通用核苷酸最优选地是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)。

[0632] 一或多个锚可通过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其它核碱基旋转180°)与聚核苷酸偶联(或结合)。举例来说,一或多个锚可包含与聚核苷酸形成Hoogsteen氢键或反向Hoogsteen氢键的一或多个核苷酸、一或多个寡核苷酸或一或多个聚核苷酸。这些氢键类型允许第三聚核苷酸链卷绕双链螺旋于周围,且形成三链螺旋。通过与双链双螺旋形成三链螺旋,一或多个锚可与双链聚核苷酸偶联(或结合)。

[0633] 在此实施例中,至少1%、至少10%、至少25%、至少50%或100%的膜组分可进行官能化。

[0634] 当一或多个锚包含蛋白质时,所述一或多个锚可以能够直接锚定到膜中而无需进一步官能化,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)跨膜蛋白质、膜内蛋白质和膜蛋白。替代地,可表达具有与膜相容的基因融合疏水区的蛋白质。此类疏水性蛋白质区在所属领域中是已知的。

[0635] 一或多个锚优选地在与膜接触之前与聚核苷酸混合,但一或多个锚可与膜接触且随后与聚核苷酸接触。

[0636] 在另一方面中,聚核苷酸可使用上文所描述的方法来官能化,使得其可由特异性结合基团识别。具体地说,聚核苷酸可用配体进行官能化,所述配体如,生物素(用于结合于抗生蛋白链菌素)、直链淀粉(用于结合于麦芽糖结合蛋白或融合蛋白)、Ni-NTA(用于结合于聚组氨酸或聚组氨酸标签的蛋白质)或肽(如抗原)。

[0637] 根据一优选实施例,一或多个锚可用于当聚核苷酸连接于前导序列时将聚核苷酸与膜偶联,所述前导序列优选地螺旋到孔中。前导序列在下文更详细地论述。较佳地,聚核苷酸连接(如接合)于优选地螺旋到孔中的前导序列。此类前导序列可包含均聚聚核苷酸或无碱基区。前导序列通常设计成直接与一或多个锚杂交,或通过一或多个中间聚核苷酸(或夹板)与所述一或多个锚杂交。在此类情况下,一或多个锚通常包含与前导序列中的序列或一或多个中间聚核苷酸(或夹板)中的序列互补的聚核苷酸序列。在此类情况下,一或多个夹板通常包含与前导序列中的序列互补的聚核苷酸序列。

[0638] 化学连接中使用的分子的实例是EDC(1-乙基-3-[3-二甲基胺基丙基]碳化二亚胺盐酸盐)。还可使用可商购的试剂盒(Thermo Pierce,产品号22980),来向聚核苷酸的5'中添加反应性基团。合适方法包括(但不限于)使用组氨酸残基和Ni-NTA的暂时亲和力连接,以及通过反应性半胱氨酸、赖氨酸或非天然氨基酸的更稳固共价连接。

[0639] 双链聚核苷酸

[0640] 聚核苷酸可以是双链。如果聚核苷酸是双链,那么在接触步骤之前,方法优选地进一步包含使桥接部分(如发夹环)与聚核苷酸的一端接合。可接着在使聚核苷酸与根据本发明孔接触时或之前,将聚核苷酸的两条链分开。在聚核苷酸通过孔移动受聚核苷酸结合蛋白(如解螺旋酶或分子制动器)控制时,可将两条链分开。

[0641] 以此方式连接和询问双链上的两条链增加表征的效率和准确度。

[0642] 桥接部分能够连接靶聚核苷酸的两条链。桥接部分通常共价连接靶聚核苷酸的两条链。桥接部分可以是能够连接靶聚核苷酸的两条链的任何东西,其限制条件为桥接部分不干扰单链聚核苷酸通过跨膜孔的移动。

[0643] 桥接部分可通过在所属领域中已知的任何合适手段连接于靶聚核苷酸。桥接部分可分开地合成,且化学连接或酶促接合于靶聚核苷酸。替代地,桥接部分可以是在靶聚核苷酸加工中产生。

[0644] 桥接部分连接于在靶聚核苷酸的一端处或附近的靶聚核苷酸。桥接部分优选地连接于在靶聚核苷酸的10个核苷酸末端内的靶聚核苷酸

[0645] 合适桥接部分包括(但不限于)聚合物连接子、化学连接子、聚核苷酸或多肽。优选地,桥接部分包含DNA、RNA、修饰DNA(如无碱基DNA)、RNA、PNA、LNA或PEG。桥接部分更优选地是DNA或RNA。

[0646] 桥接部分最优选地是发夹环或发夹环衔接子。可使用所属领域中已知的方法来设

计合适发夹衔接子。发夹环可以是任何长度。发夹环的长度通常是110个或更少个核苷酸,如100个或更少个核苷酸、90个或更少个核苷酸、80个或更少个核苷酸、70个或更少个核苷酸、60个或更少个核苷酸、50个或更少个核苷酸、40个或更少个核苷酸、30个或更少个核苷酸、20个或更少个核苷酸,或10个或更少个核苷酸。发夹环的长度优选地是约1到110、2到100、5到80或6到50个核苷酸。如果环涉及不同的衔接子的选择能力,那么较长长度的发夹环,如50到110个核苷酸,是优选的。类似地,如果环不涉及如下文所论述的可选择结合,那么较短长度的发夹环,如1到5个核苷酸,是优选的。

[0647] 发夹衔接子可接合于第一和/或第二聚核苷酸的任一端,即5'或3'端。使可用所属领域中已知的方法,将发夹衔接子接合于第一和/或第二聚核苷酸。可使用接合酶,如T4 DNA接合酶、大肠杆菌DNA接合酶、Taq DNA接合酶、Tma DNA接合酶和9°N DNA接合酶,来接合发夹衔接子。

[0648] 可使用所属领域中已知的任何方法,来将聚核苷酸的两条链分开。举例来说,可通过聚核苷酸结合蛋白或使用有利于解杂交的条件,来将其分开(有利于解杂交的条件的实例包括(但不限于)高温、高pH和添加可破坏氢结合或碱基配对的药剂,如甲酰胺和脲)。

[0649] 发夹衔接子优选地包含可选择结合部分。这允许第一和/或第二聚核苷酸纯化或分离。可选择结合部分是可基于其结合特性而选择的部分。因此,可选择结合部分优选地是特异性结合于表面的部分。如果可选择结合部分以比在本发明中使用的任何其它部分高许多的程度结合于表面,那么其特异性结合于表面。在优选实施例中,部分结合于无本发明中使用的其它部分与其结合的表面。

[0650] 合适选择性结合部分在所属领域中是已知的。优选选择性结合部分包括(但不限于):生物素;聚核苷酸序列;抗体;抗体片段,如Fab和ScSv;抗原;聚核苷酸结合蛋白;聚组氨酸尾部和GST标签。最优选选择性结合部分是生物素和可选择聚核苷酸序列。生物素特异性结合于涂布有抗生物素蛋白的表面。可选择聚核苷酸序列特异性地结合(即杂交)于涂布有同源序列的表面。替代地,可选择聚核苷酸序列特异性结合于涂布有聚核苷酸结合蛋白的表面。

[0651] 发夹衔接子和/或可选择结合部分可包含可进行剪切、切割、裂解或水解的区。此类区可设计成允许从其在纯化或分离后结合于其的表面移除第一和/或第二聚核苷酸。合适区在所属领域中是已知的。合适区包括(但不限于)RNA区、包含脱硫生物素和抗生蛋白链霉菌素的区、二硫键和可光裂解区。

[0652] 双链靶聚核苷酸优选地包含在桥接部分(如发夹环或发夹环衔接子)相对端处的前导序列。前导序列在下文更详细地论述。

[0653] 进行转角测序

[0654] 在一优选实施例中,靶双链聚核苷酸在一端具备桥接部分,如发夹环或发夹环衔接子,且方法包含使聚核苷酸与孔接触,使得聚核苷酸的两条链移动通过孔,且在聚核苷酸的两条链相对于孔移动时进行一或多个测量,其中测量指示聚核苷酸的链的一或多个特征,且从而表征靶双链聚核苷酸。在另一优选实施例中,靶双链聚核苷酸在一端具备桥接部分,如发夹环或发夹环衔接子,且方法包含使聚核苷酸与孔和核酸外切酶接触,使得消化聚核苷酸的两条链,形成个别核苷酸。上文所论述的任一个实施例同样应用于这个实施例。

[0655] 前导序列

[0656] 在链表征/测序方法中的接触步骤之前,方法优选地包含将聚核苷酸连接于优选地螺旋到孔中的前导序列。前导序列有助于本发明的方法。前导序列设计成优选地螺旋到孔中,且从而有助于聚核苷酸通过孔的移动。前导序列还可用于将聚核苷酸连接于如上文所论述的一或多个锚。

[0657] 前导序列通常包含聚合物。聚合物优选地是带负电的。聚合物优选地是:聚核苷酸,如DNA或RNA;修饰聚核苷酸(如无碱基DNA);PNA;LNA;聚乙二醇(PEG)或多肽。前导优选地包含聚核苷酸,并且更优选地包含单链聚核苷酸。前导序列可包含上文所论述的任一个聚核苷酸。单链前导序列最优选地包含DNA的单链,如聚dT区段。前导序列优选地包含一或多个间隔子。

[0658] 前导序列可以是任何长度,但其长度通常是10到150个核苷酸,如20到150个核苷酸。前导的长度通常取决于方法中使用的跨膜孔。

[0659] 前导序列优选地是如下文所定义的Y衔接子的一部分。

[0660] 双重偶联

[0661] 本发明的方法可涉及双链聚核苷酸的双重偶联。在一优选实施例中,本发明的方法包含:

[0662] (a) 提供在一端处具有Y衔接子和在另一端处具有桥接部分衔接子(如发夹环衔接子)的双链聚核苷酸,其中Y衔接子包含用于将聚核苷酸与膜偶联的一或多个第一锚,其中桥接部分衔接子包含用于将聚核苷酸与膜偶联的一或多个第二锚,且其中桥接部分衔接子与膜偶联的强度大于Y衔接子与膜偶联的强度;

[0663] (b) 使步骤(a)中提供的聚核苷酸与本发明的孔接触,使得聚核苷酸相对于(如通过)孔移动;和

[0664] (c) 在聚核苷酸相对于孔移动时,进行一或多个测量,其中测量指示聚核苷酸的一或多个特征,且从而表征靶聚核苷酸。

[0665] 这种方法类型在英国申请第1406147.7号中详细论述。

[0666] 双链聚核苷酸在一端处具备Y衔接子和在另一端处具备桥接部分衔接子。Y衔接子和/或桥接部分衔接子通常是聚核苷酸衔接子。其可由上文所论述的任一个聚核苷酸形成。

[0667] Y衔接子通常包含(a)双链区和(b)单链区或在另一端处不互补的区。如果Y衔接子包含单链区,那么其可被描述为具有悬突。Y衔接子中非互补区的存在给予衔接子其Y形状,这是由于不同于双链部分,两条链通常不彼此杂交。Y衔接子包含一或多个第一锚。锚在上文更详细地论述。

[0668] Y衔接子优选地包含优选地螺旋到孔中的前导序列。这一点在上文论述。

[0669] 桥接部分衔接子优选地包含如上文所论述的可选择结合部分。桥接部分衔接子和/或可选择结合部分可包含可如上文所论述进行剪切、切割、裂解或水解的区。

[0670] 如果一或多个解螺旋酶和一或多个分子制动器如上文所论述来使用,那么Y衔接子优选地包含一或多个解螺旋酶,且桥接部分衔接子优选地包含一或多个分子制动器。

[0671] 使可用所属领域中已知的方法,将Y衔接子和/或桥接部分衔接子接合于聚核苷酸。可使用接合酶,如T4 DNA接合酶、大肠杆菌DNA接合酶、Taq DNA接合酶、Tma DNA接合酶和9°N DNA接合酶,来接合衔接子中的一个或两个。替代地,可使用下文论述的本发明的方法,向聚核苷酸中添加衔接子。

[0672] 在一优选实施例中,方法的步骤a)包含修饰双链聚核苷酸,使得其包含在一端处的Y衔接子且在另一端处的桥接部分衔接子。可使用任何修饰方式。方法优选地包含修饰根据本发明的双链聚核苷酸。这一点在下文更详细论述。可以任何方式对修饰和表征方法进行组合。

[0673] 桥接部分衔接子与膜偶联(或结合)的强度大于Y衔接子与膜偶联(或结合)的强度。这可以任何方式来测量。用于测量偶联(或结合)的强度的合适方法公开于英国申请第1406147.7号的实例中。

[0674] 桥接部分衔接子偶联(或结合)的强度优选地是Y衔接子偶联(或结合)的强度的至少1.5倍,如锚衔接子偶联(或结合)的强度的至少两倍、至少三倍、至少四倍、至少五倍或至少十倍。桥接部分衔接子对于膜的亲和力常数(Kd)优选地是Y衔接子的亲和力常数的至少1.5倍,如Y衔接子的偶联的强度的至少两倍、至少三倍、至少四倍、至少五倍或至少十倍。

[0675] 存在若干种使桥接部分衔接子比Y衔接子更强烈地与膜偶联(或结合)的方式。举例来说,桥接部分衔接子可包含比Y衔接子更多的锚。举例来说,桥接部分衔接子可包含2、3个或更多个第二锚,而Y衔接子可包含一个第一锚。

[0676] 一或多个第二锚与膜的偶联(或结合)的强度可大于一或多个第一锚与膜的偶联(或结合)的强度。一或多个第二锚与桥接部分衔接子的偶联(或结合)的强度可大于一或多个第一锚与Y衔接子的偶联(或结合)的强度。一或多个第一锚和一或多个第二锚可通过杂交而连接于其相应衔接子,且一或多个第二锚中的杂交的强度比一或多个第一锚中的杂交的强度高。在本发明中还可使用此等实施例的任何组合。可使用所属领域中的已知技术来测量偶联(或结合)的强度。

[0677] 一或多个第二锚优选地包含以比一或多个第一锚中的一个或多个基团与膜偶联(或结合)更高的强度与膜偶联(或结合)的一个或多个基团。在优选实施例中,桥接部分衔接子/一或多个第二锚使用胆固醇与膜偶联(或结合),且Y衔接子/一或多个第一锚使用棕榈酸与膜偶联(或结合)。胆固醇比棕榈酸更强烈地结合于三嵌段共聚物膜和脂质膜。在一替代实施例中,桥接部分衔接子/一或多个第二锚使用单酰基物种(如棕榈酸)与膜偶联(或结合),且Y衔接子/一或多个第一锚使用二酰基物种(如二棕榈酰基磷脂酰胆碱)与膜偶联(或结合)。

[0678] 添加发夹环和前导序列

[0679] 在提供之前,可使双链聚核苷酸与MuA转座酶和双链MuA底物群体接触,其中群体中的一部分底物是包含前导序列的Y衔接子,且其中群体中的一部分底物是发夹环衔接子。转座酶将双链聚核苷酸分析物分段,且将MuA底物与片段的一个或两个端接合。这产生包含在一端处的前导序列和在另一端处的发夹环的多个修饰双链聚核苷酸。可接着使用本发明的方法来研究修饰双链聚核苷酸。

[0680] 群体中的各底物优选地包含至少一个通用核苷酸的悬突,使得转座酶将模板聚核苷酸分段,且将底物与双链片段的一或两端接合,且从而产生多个片段/底物构筑体,且其中方法进一步包含使悬突与构筑体中的片段接合,且从而生产多个修饰双链聚核苷酸。合适通用核苷酸在上文论述。悬突的长度优选地是五个核苷酸。

[0681] 替代地,群体中的各底物优选地包含:(i)至少一个悬突;和(ii)在与包含不存在于模板聚核苷酸中的核苷的至少一个悬突相同的链中的至少一个核苷酸,使得转座酶将模

板聚核苷酸分段,且使底物与双链片段的一或两端接合,且从而产生多个片段/底物构筑体,且其中方法进一步包含(a)通过选择性地移除至少一个核苷酸来从构筑体移除悬突,且从而生产包含单链间隙的多个双链构筑体,和(b)修复构筑体中的单链间隙,且从而生产多个修饰双链聚核苷酸。聚核苷酸通常包含以下核苷:脱氧腺苷(dA)、脱氧尿苷(dU)和/或胸苷(dT)、脱氧鸟苷(dG)和脱氧胞苷(dC)。不存在于聚核苷酸中的核苷优选地是无碱基核苷、腺苷(A)、尿苷(U)、5-甲基尿苷(m^5U)、胞苷(C)或鸟苷(G),或包含脲、5,6-二羟基胸苷、胸苷乙二醇、5-羟基-5-甲基乙内酰脲、尿嘧啶乙二醇、6-羟基-5,6-二羟基胸苷、甲基酒石酰脲、7,8-二氢-8-氧代鸟嘌呤(8-氧代鸟嘌呤)、8-氧代腺嘌呤、fapy-鸟嘌呤、甲基-fapy-鸟嘌呤、fapy-腺嘌呤、黄霉素B1-fapy-鸟嘌呤、5-羟基-胞嘧啶、5-羟基-尿嘧啶、3-甲基腺嘌呤、7-甲基鸟嘌呤、1,N6-亚乙烯基腺嘌呤、次黄嘌呤、5-羟基尿嘧啶、5-羟甲基尿嘧啶、5-甲酰基尿嘧啶或顺-syn-环丁烷嘧啶二聚体。至少一个核苷酸优选地是来自悬突的10个核苷酸或更少。至少一个核苷酸是悬突中的第一个核苷酸。悬突中的所有核苷酸优选地包含不存在于模板聚核苷酸中的核苷。

[0682] 这些基于MuA的方法公开于国际申请第PCT/GB2014/052505号中。其还在英国申请第1406147.7号中详细论述。

[0683] 在一或多个解螺旋酶与双链聚核苷酸和MuA转座酶接触之前,所述一或多个解螺旋酶可连接于MuA底物Y衔接子。替代地,在一或多个解螺旋酶与双链聚核苷酸和MuA转座酶接触之前,所述一或多个解螺旋酶可连接于MuA底物Y衔接子。

[0684] 在一或多个分子制动器与双链聚核苷酸和MuA转座酶接触之前,所述一或多个分子制动器可连接于MuA底物发夹环衔接子。替代地,在一或多个分子制动器与双链聚核苷酸和MuA转座酶接触之前,所述一或多个分子制动器可连接于MuA底物发夹环衔接子。

[0685] 解偶联

[0686] 本发明的方法可涉及表征多个靶聚核苷酸和解偶联至少第一靶聚核苷酸。

[0687] 在一优选实施例中,本发明涉及表征两个或更多个靶聚核苷酸。方法包含:

[0688] (a)在第一样本中提供第一聚核苷酸;

[0689] (b)在第二样本中提供第二聚核苷酸;

[0690] (c)使用一或多个锚,将第一样本中的第一聚核苷酸与膜偶联;

[0691] (d)使第一聚核苷酸与本发明的孔接触,使得聚核苷酸相对于(如通过)孔移动;

[0692] (e)在第一聚核苷酸相对于所述孔移动时,进行一或多个测量,其中测量指示第一聚核苷酸的一或多个特征,且从而表征第一聚核苷酸;

[0693] (f)将第一聚核苷酸从膜解偶联;

[0694] (g)使用一或多个锚,将第二样本中的第二聚核苷酸与膜偶联;

[0695] (h)使第二聚核苷酸与本发明的孔接触,使得第二聚核苷酸相对于(如通过)孔移动;和

[0696] (i)在第二聚核苷酸相对于孔移动时,进行一或多个测量,其中测量指示第二聚核苷酸的一或多个特征,且从而表征第二靶聚核苷酸。

[0697] 这种方法类型在英国申请第1406155.0号中详细论述。

[0698] 可以在步骤(g)(即在使第二聚核苷酸与膜偶联之前)之前,进行步骤(f)(即将第一聚核苷酸解偶联)。可以在步骤(f)之前进行步骤(g)。如果在第一聚核苷酸解偶联之前,

第二聚核苷酸与膜偶联,那么步骤(f)优选地包含选择性地将第一聚核苷酸从膜解偶联(即从膜解偶联第一聚核苷酸但不解偶联第二聚核苷酸)。所属领域的技术人员可设计其中实现选择性解偶联的系统。可以在同时进行步骤(f)和(g)。这一点在下文更详细论述。

[0699] 在步骤(f)中,至少10%的第一聚核苷酸优选地从膜解偶联。举例来说,至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少95%的第一聚核苷酸可以从膜解偶联。优选地,所有的第一聚核苷酸均从膜解偶联。可使用孔来测定从膜解偶联的第一聚核苷酸的量。这公开于实例中。

[0700] 第一聚核苷酸和第二聚核苷酸可彼此不同。替代地,第一和第二聚核苷酸可以是不同的聚核苷酸。在此类情况下,不需要在添加第二聚核苷酸之前移除至少一部分的第一样本。这一点在下文更详细论述。如果方法涉及研究三个或更多个聚核苷酸,那么其所有均可以彼此不同,或其中一些可以彼此不同。

[0701] 第一聚核苷酸和第二聚核苷酸可以是相同聚核苷酸的两个例子。第一聚核苷酸可以与第二聚核苷酸相同。这允许校对。如果方法涉及研究三个或更多个聚核苷酸,那么其所有均可以所有相同聚核苷酸的三个或更多个例子,或其中一些可以是相同聚核苷酸的独立例子。

[0702] 第一样本和第二样本可以彼此不同。举例来说,第一样本可来源于人类,且第二样本可来源于病毒。如果第一和第二样本彼此不同,那么其可含有或疑似含有相同的第一和第二聚核苷酸。如果方法涉及研究三个或更多个样本,那么其所有均可以彼此不同,或其中一些可以彼此不同。

[0703] 第一样本和第二样本优选地是相同样本的两个例子。第一样本优选地与第二样本相同。这允许校对。如果方法涉及研究三个或更多个样本,那么其所有均可以是相同样本的三个或更多个例子,或其中一些可以是相同样本的独立例子。

[0704] 可研究任何数量的聚核苷酸。举例来说,本发明的方法可涉及表征3、4、5、6、7、8、9、10、20、30、50、100或更多个聚核苷酸。如果使用本发明的方法研究三个或更多个聚核苷酸,那么第二聚核苷酸也从膜解偶联,且对于第三聚核苷酸添加必需的步骤数。对于四个或更多个聚核苷酸来说,也是如此。

[0705] 本发明的方法涉及从膜解偶联第一聚核苷酸。如果研究三个或更多个聚核苷酸,那么本发明的方法可涉及从膜解偶联第二聚核苷酸。

[0706] 可使用任何已知方法,从膜解偶联第一聚核苷酸。优选地在步骤(f)中使用跨膜孔,不从膜解偶联第一聚核苷酸。优选地使用电压或施加的电势,不从膜解偶联第一聚核苷酸。

[0707] 步骤(f)优选地包含通过从膜移除一或多个锚来从膜解偶联第一聚核苷酸。如果移除锚,那么使用其它(或独立)锚将第二聚核苷酸与膜偶联。用于偶联第二聚核苷酸的锚可以是与偶联第一聚核苷酸相同的锚类型或不同的锚类型。

[0708] 步骤(f)更优选地包含使一或多个锚与对于所述一或多个锚具有比膜对于锚具有的亲和力更高的亲和力的药剂接触。用于竞争性结合或免疫放射分析以测定分子的特异性结合能力的各种方案在所属领域中是众所周知的(参见例如Maddox等人,《实验医学杂志(J. Exp. Med.)》158, 1211-1226, 1993)。药剂将锚从膜移除,且从而解偶联第一聚核苷酸。药剂优选地是糖。可使用以比一或多个锚对于膜具有的亲和力更高的亲和力结合于一或多个

锚的任何糖。糖可以是如下文所论述的环糊精或其衍生物。

[0709] 如果一或多个锚包含疏水性锚(如胆固醇),那么药剂优选地是环糊精一其衍生物或脂质。环糊精或其衍生物可以是在Eliseev,A.V.和Schneider,H-J. (1994)《美国化学学会杂志》116,6081-6088中所公开那些化合物中的任一个。药剂更优选地是七-6-氨基- β -环糊精(am₇- β CD)、6-单脱氧-6-单氨基- β -环糊精(am₁- β CD)或七-(6-脱氧-6-胍基)-环糊精(gu₇- β CD)。可使用本文所公开的任一个脂质。

[0710] 如果锚包含抗生蛋白链菌素、生物素或脱硫生物素,那么药剂优选地是生物素、脱硫生物素或抗生蛋白链菌素。生物素和脱硫生物素两个均以比抗生蛋白链菌素结合于膜更高的亲和力结合于抗生蛋白链菌素,且反之亦然。生物素对于抗生蛋白链菌素比脱硫生物素具有更强的亲和力。因此,可使用生物素或抗生蛋白链菌素从膜移除包含抗生蛋白链菌素的锚,且反之亦然。

[0711] 如果锚包含蛋白质,那么药剂优选地是特异性结合于蛋白质的抗体或其片段。如果抗体以优先或高亲和力结合于蛋白质,但并不结合或仅以很低的亲和力结合于其它或不同蛋白质,那么抗体特异性结合于蛋白质。如果抗体以 1×10^{-6} M或更少,更优选地 1×10^{-7} M或更少、 5×10^{-8} M或更少,更优选地 1×10^{-8} M或更少或更优选地 5×10^{-9} M或更少的Kd结合,那么所述抗体以优先或高亲和性结合。如果抗体以 1×10^{-6} M或更多,更优选地 1×10^{-5} M或更多,更优选地 1×10^{-4} M或更多,更优选地 1×10^{-3} M或更多,甚至更优选地 1×10^{-2} M或更多的Kd结合,那么所述抗体以很低的亲和力结合。任何方法可用于检测结合或特异性结合。定量测量抗体与蛋白质的结合的方法在所属领域中是众所周知的。抗体可以是单克隆抗体或多克隆抗体。合适的抗体片段包括(但不限于)Fv、F(ab')和F(ab')₂片段以及单链抗体。此外,抗体或其片段可以是嵌合抗体或其片段、CDR-接枝抗体或其片段或人源化抗体或其片段。

[0712] 步骤(f)优选地包含使一或多个锚与药剂接触,所述药剂降低所述一或多个锚与膜偶联的能力。举例来说,药剂可干扰一或多个锚的结构和/或疏水性,且从而降低其与膜偶联的能力。如果锚包含胆固醇,那么药剂优选地是胆固醇脱氢酶。如果锚包含脂质,那么药剂优选地是磷脂酶。如果锚包含蛋白质,那么药剂优选地是蛋白酶或脲。合适锚与药剂的其它组合对于所属领域的技术人员将是清楚的。

[0713] 步骤(f)优选地包含通过从一或多个锚分离第一聚核苷酸来从膜解偶联第一聚核苷酸。这可以任何手段来进行。举例来说,可以在包含连接子的锚中剪切连接子。这个实施例特别适用于涉及通过杂交连接的锚。此类锚在上文论述。

[0714] 步骤(f)更优选地包含通过使第一聚核苷酸和一或多个锚与药剂接触来从膜解偶联第一聚核苷酸,所述药剂与第一聚核苷酸竞争与一或多个锚结合。用于测定和测量竞争性结合的方法在所属领域中是已知的。药剂优选地是与第一聚核苷酸竞争与一或多个锚杂交的聚核苷酸。举例来说,如果使用涉及杂交的一或多个锚来使第一聚核苷酸与膜偶联,那么可通过使一或多个锚与也与杂交位点杂交的聚核苷酸接触来解偶联聚核苷酸。通常以高于第一聚核苷酸和一或多个锚的浓度的浓度,来添加聚核苷酸药剂。替代地,聚核苷酸药剂可比第一聚核苷酸更强烈地与一或多个锚杂交。

[0715] 步骤(f)更优选地包含:(i)使第一聚核苷酸和一或多个锚与脲、三(2-羧乙基)膦(TCEP)、二硫苏糖醇(DTT)、抗生蛋白链菌素或生物素、UV光、酶或结合剂接触;(ii)将第一聚核苷酸和一或多个锚加热;或(iii)改变pH。脲、三(2-羧乙基)膦(TCEP)或二硫苏糖醇

(DTT)能够破坏锚且从膜分离第一聚核苷酸。如果锚包含抗生蛋白链菌素-生物素连接,那么抗生蛋白链菌素药剂将竞争与生物素结合。如果锚包含抗生蛋白链菌素-脱硫生物素连接,那么生物素药剂将竞争与抗生蛋白链菌素结合。UV光可用于分解光不稳定性基团。酶和结合剂可用于剪切、分解或拆开锚。优选酶包括(但不限于)核酸外切酶、核酸内切酶或解螺旋酶。优选结合剂包括(但不限于)酶、抗体或其片段或单链结合蛋白(SSB)。可使用下文论述的酶或上文所论述的抗体中的任一个。加热和pH可用于破坏杂交和其它连接。

[0716] 如果通过从一或多个锚分离第一聚核苷酸来从膜解偶联第一聚核苷酸,那么一或多个锚将残留在膜中。步骤(g)优选地包含使用从第一聚核苷酸分离的一或多个锚,来使第二聚核苷酸与膜偶联。举例来说,第二聚核苷酸还可具备与残留于膜中的一或多个锚杂交的一或多个聚核苷酸。替代地,步骤(g)优选地包含使用来自从第一聚核苷酸分离的一者(即一或多个其它锚)的一或多个独立锚,来使第二聚核苷酸与膜偶联。所述一或多个独立锚可以是与用于使第一聚核苷酸与膜偶联的锚相同的锚类型,或可以是不同的锚类型。步骤(g)优选地包含使用来自从第一聚核苷酸分离的一或多个锚的一或多个不同锚,来使第二聚核苷酸与膜偶联。

[0717] 在一优选实施例中,步骤(f)和(g)包含通过使膜与第二聚核苷酸接触,来从膜解偶联第一聚核苷酸,使得第二聚核苷酸与第一聚核苷酸竞争与一或多个锚结合,且置换第一聚核苷酸。举例来说,如果使用涉及杂交的一或多个锚来使第一聚核苷酸与膜偶联,那么可通过使锚与连接于也与一或多个锚中的杂交位点杂交的聚核苷酸的第二聚核苷酸接触,来解偶联第一聚核苷酸。通常以高于第一聚核苷酸和一或多个锚的浓度的浓度,来添加第二聚核苷酸。替代地,第二聚核苷酸可比第一聚核苷酸更强烈地与一或多个锚杂交。

[0718] 移除或洗涤

[0719] 尽管在步骤(f)中从膜解偶联第一聚核苷酸,但不必移除或洗掉。如果第二聚核苷酸可容易地区别于第一聚核苷酸,那么不需要移除第一聚核苷酸。

[0720] 在步骤(f)与(g)之间,方法优选地进一步包含从膜移除至少一些第一样本。可移除至少10%的第一样本,如可移除至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%或至少90%的第一样本。

[0721] 方法更优选地进一步包含从膜移除所有的第一样本。这可以任何方式来进行。举例来说,可在已解偶联第一聚核苷酸之后,用缓冲液来洗涤膜。合适缓冲液在下文论述。

[0722] 经修饰聚核苷酸

[0723] 在表征之前,可通过在使用靶聚核苷酸作为模板,聚合酶形成经修饰聚核苷酸条件下,使聚核苷酸与聚合酶和游离核苷酸群体接触,来对靶聚核苷酸进行修饰,其中当形成经修饰聚核苷酸时,聚合酶用不同核苷酸物种置换靶聚核苷酸中的一或多个核苷酸物种。接着,经修饰聚核苷酸可具备连接于聚核苷酸的一或多个解螺旋酶和连接于聚核苷酸的一或多个分子制动器。这种修饰类型描述于英国申请第1403096.9号中。可使用上文所论述的任一个聚合酶。聚合酶优选地是Klenow或9o North。

[0724] 在使用模板聚核苷酸作为模板,聚合酶形成经修饰聚核苷酸条件下,使模板聚核苷酸与聚合酶接触。此类条件是本领域中已知的。举例来说,通常使聚核苷酸与含聚合酶的可商购的聚合酶缓冲液接触,所述缓冲液如来自New England Biolabs®的缓冲液。对于Klenow的温度优选地是20到37℃,或对于9o North的是60到75℃。引物或3'发夹通常用作

聚合酶延伸的成核点。

[0725] 使用跨膜孔表征(如测序)聚核苷酸通常涉及分析由k个核苷酸构成的聚合物单元,其中k是正整数(即‘k-聚体’)。这一点在国际申请第PCT/GB2012/052343号(公开为WO 2013/041878)号中论述。尽管期望在不同聚体的电流测量之间具有清楚的分隔,但这些测量中的一些重叠是常见的。尤其在k聚体中用高数量的聚合物单元,即高的k值,可能变得难以解析由不同k聚体产生的测量,损害关于聚核苷酸的推导信息,例如聚核苷酸的潜在序列的评估。

[0726] 通过用经修饰聚核苷酸中的不同核苷酸物种来置换靶聚核苷酸中的一或多个核苷酸物种,经修饰聚核苷酸含有不同于靶聚核苷酸中的那些k聚体的k聚体。经修饰聚核苷酸中的不同k聚体能够从靶聚核苷酸中的k聚体产生不同的电流测量,且因此经修饰聚核苷酸提供来自靶聚核苷酸的不同信息。来自经修饰聚核苷酸的额外信息可使得其更容易表征靶聚核苷酸。在一些例子中,经修饰聚核苷酸自身可以更容易进行表征。举例来说,经修饰聚核苷酸可以设计成包括在其电流测量之间具有增加的分隔或清楚的分隔的k聚体,或具有噪声减少的k聚体。

[0727] 当形成经修饰聚核苷酸时,聚合酶优选地用不同核苷酸物种置换靶聚核苷酸中的两个或更多个核苷酸物种。聚合酶可用独特的核苷酸物种置换靶聚核苷酸中的两个或更多个核苷酸物种中的每一个。聚合酶可用相同的核苷酸物种置换靶聚核苷酸中的两个或更多个核苷酸物种中的每一个。

[0728] 如果靶聚核苷酸是DNA,那么经修饰聚核苷酸中的不同核苷酸物种通常包含不同于腺嘌呤、鸟嘌呤、胸腺嘧啶、胞嘧啶或甲基胞嘧啶的核碱基,和/或包含不同于脱氧腺苷、脱氧鸟苷、胸苷、脱氧胞苷或脱氧甲基胞苷的核苷。如果靶聚核苷酸是RNA,那么经修饰聚核苷酸中的不同核苷酸物种通常包含不同于腺嘌呤、鸟嘌呤、尿嘧啶、胞嘧啶或甲基胞嘧啶的核碱基,和/或包含不同于腺苷、鸟苷、尿苷、胞苷或甲基胞苷的核苷。不同核苷酸物种可以是上文所论述的任一个通用核苷酸。

[0729] 聚合酶可用不同核苷酸物种来置换一或多个核苷酸物种,所述不同核苷酸物种包含一或多个核苷酸物种缺乏的化学基团或原子。化学基团可以是丙炔基、硫基、氧代基、甲基、羟甲基、甲酰基、羧基、羰基、苯甲基、炔丙基或炔丙胺基。

[0730] 聚合酶可用不同核苷酸物种来置换一或多个核苷酸物种,所述不同核苷酸物种缺乏一或多个核苷酸物种中存在的化学基团或原子。聚合酶可用具有电负性改变的不同核苷酸物种来置换核苷酸物种中的一或多个。具有电负性改变的不同核苷酸物种优选地包含卤素原子。

[0731] 方法优选地进一步包含从经修饰聚核苷酸中的一或多个不同核苷酸物种选择性地移除核碱基。

[0732] 分析物递送

[0733] 靶分析物优选地连接于将分析物递送向膜的微粒。这种递送类型公开于英国申请第1418469.1号中。可使用任何类型的微粒和连接方法。

[0734] 其它表征方法

[0735] 在另一实施例中,利用聚合酶通过检测向靶聚核苷酸中添加的标记物种,来表征聚核苷酸,且接着释放。聚合酶使用聚核苷酸作为模板。各标记的物种对各核苷酸具有特异

性。使聚核苷酸与本发明的孔(如本发明的孔)、聚合酶和标记核苷酸接触,使得磷酸标记的物种通过聚合酶依序添加到靶聚核苷酸中,其中磷酸物种含有对各核苷酸具有特异性的标记。可在从核苷酸释放标记物种之前(即在靶聚核苷酸向中添加其时),或在从核苷酸释放其之后,使用孔来检测标记物种。

[0736] 聚合酶可以是上文所论述的那些聚合酶中的任一个。使用孔检测磷酸标记物种,且从而表征聚核苷酸。这种方法类型公开于欧洲申请第13187149.3(公开为EP 2682460)号中。上文所论述的任一个实施例同样应用于这种方法。

[0737] 标记物种的实例包括(但不限于)聚合物、聚乙二醇、糖、环糊精、荧光团、药物、代谢物、肽。此类标签的非限制性实例可在Kumar等人《科学研究(Sci Rep.)》2012;2:684.电子版2012年9月21的研究中找到。

[0738] 形成传感器的方法

[0739] 本发明还提供一种形成用于表征靶聚核苷酸的传感器的方法。方法包含在本发明的孔与聚核苷酸结合蛋白(如解螺旋酶或核酸外切酶)之间形成复合体。可通过在存在靶聚核苷酸下使孔和蛋白质接触且接着在孔两端施加电势,来形成复合体。所施加的电势可以是如上文所描述的化学势或电压电势。替代地,可通过将孔共价连接于蛋白质来形成复合体。用于共价连接的方法在所属领域中是已知的,且例如公开于国际申请第PCT/GB09/001679号(公开为W0 2010/004265)和第PCT/GB10/000133号(公开为W02010/086603)中。复合体是用于表征靶聚核苷酸的传感器。方法优选地包含在本发明的孔与解螺旋酶之间形成复合体。上文所论述的任一个实施例同样应用于这种方法。

[0740] 本发明还提供一种用于表征靶聚核苷酸的传感器。传感器包含在本发明的孔与聚核苷酸结合蛋白之间的复合体。上文所论述的任一个实施例同样应用于本发明的传感器。

[0741] 试剂盒

[0742] 本发明还提供一种用于表征靶聚核苷酸的试剂盒。试剂盒包含本发明的孔和膜的组分。膜优选地由组分形成。孔优选地存在于膜中。试剂盒可包含上文所公开的任一个膜(如两亲层或三嵌段共聚物膜)的组分。

[0743] 试剂盒可进一步包含聚核苷酸结合蛋白。可使用上文所论述的任一个聚核苷酸结合蛋白。

[0744] 试剂盒可进一步包含用于使聚核苷酸与膜偶联的一或多个锚。

[0745] 试剂盒优选地是用于表征双链聚核苷酸,并优选地包含Y衔接子和发夹环衔接子。Y衔接子优选地具有所连接的一或多个解螺旋酶,且发夹环衔接子优选地具有所连接的一或多个分子制动器。Y衔接子优选地包含用于使聚核苷酸与膜偶联的一或多个第一锚,发夹环衔接子优选地包含用于使聚核苷酸与膜偶联的一或多个第二锚,且发夹环衔接子与膜偶联的强度优选地大于Y衔接子与膜偶联的强度。

[0746] 本发明的试剂盒可另外包含使得能够进行上文提到的任一个实施例的一或多个其它试剂或仪器。此类试剂或仪器包括以下中的一或多个:合适缓冲液(水性溶液)、从个体获得样本的装置(如包含针的容器或仪器)、用于扩增和/或表达聚核苷酸的装置,或电压或贴片钳设备。试剂可以干态形式存在于试剂盒中,使得流体样本再悬浮试剂。试剂盒还可任选地包含使得能够用本发明的方法使用试剂盒的说明书或关于何种生物体可使用所述方法的详情。

[0747] 设备

[0748] 本发明还提供一种用于表征靶分析物(如靶聚核苷酸)的设备。设备包含多个本发明的孔和多个膜。多个孔优选地存在于多个膜中。孔和膜的数量优选地是相等的。优选地,各膜中存在单一孔。

[0749] 设备优选地进一步包含用于进行本发明的方法的说明书。设备可以是用于分析物分析的任何常规设备,如阵列或芯片。上文参考本发明的方法所论述的任一个实施例同样适用于本发明的设备。设备可进一步包含存在于本发明的试剂盒中的任一个特征。

[0750] 设备优选地安设成进行本发明的方法。

[0751] 设备优选地包含:

[0752] 传感器装置,其能够支撑多个孔和膜,所述传感器装置能够操作以使用孔和膜来执行分析物表征;和

[0753] 至少一个通口,其用于递送执行表征用材料。

[0754] 替代地,设备优选地包含:

[0755] 传感器装置,其能够支撑多个孔和膜,所述传感器装置能够操作以使用孔和膜来执行分析物表征;和

[0756] 至少一个储槽,其用于盛放执行表征用材料。

[0757] 设备更优选地包含:

[0758] 传感器装置,其能够支撑多个孔和膜,所述传感器装置能够操作以使用孔和膜来执行分析物表征;

[0759] 至少一个储槽,其用于盛放执行表征用材料;

[0760] 流控系统,其配置成可控地从至少一个储槽供应材料到传感器装置;和

[0761] 一个或多个容器,其用于容纳相应样本,所述流控系统配置成选择性地从一个或多个容器供应样本到传感器装置。

[0762] 设备可以是在国际申请第PCT/GB08/004127号(公开为W0 2009/077734)、第PCT/GB10/000789号(公开为W0 2010/122293)、国际申请第PCT/GB10/002206号(公开为W0 2011/067559)或国际申请第PCT/US99/25679号(公开为W0 00/28312)中所描述的任一种设备。

[0763] 以下实例说明本发明。

[0764] 实例1

[0765] 本实例描述研究在CsgG内的DNA行为所进行的模拟。

[0766] 材料和方法

[0767] 进行所操控的分子动态模拟,以研究CsgG-Eco和各种突变对于DNA易位的高能屏障的量值。使用GROMACS程序包4.0.5版,利用GROMOS 53a6力场和SPC水模型来进行模拟。CsgG-Eco(SEQ ID NO:2)的结构取自蛋白质数据库登录号4UV3。为了制备CsgG-Eco突变的模型,使用PyMOL对野生型蛋白质结构进行突变。所研究的突变是CsgG-Eco-(F56A))(具有突变F56A的SEQ ID NO:2)、CsgG-Eco-(F56A-N55S)(具有突变F56A/N55S的SEQ ID NO:2)和CsgG-Eco-(F56A-N55S-Y51A)(具有突变F56A/N55S/Y51A的SEQ ID NO:2)。

[0768] 接着,将DNA放置到孔中。安设两个不同系统:

[0769] i.将单一鸟嘌呤核苷酸放置到孔中,恰好在收缩区上(大致残基56环上方的5-10

埃)

[0770] ii.沿孔轴放置单链DNA(ssDNA),其中5'端朝向孔的 β -折叠桶侧。在这种安设中,ssDNA预先螺旋到孔的整个长度中。

[0771] 接着,将模拟盒溶剂化,且接着使用最陡下降算法使能量降至最低。

[0772] 使用到300K的Berendsen恒温器和Berendsen恒压器,在NPT集中模拟各系统。在整个模拟中,对于孔的主链应用限制。

[0773] 为了牵拉DNA通过孔,在单一鸟嘌呤模拟中,对于磷原子施加拉力。在ssDNA模拟中,对于链的5'端处的磷原子施加拉力。通过在上文提到的DNA磷原子与以恒定速度平行于孔轴行进的假想点之间连接弹簧,以恒定速度来施加拉力。应注意,弹簧并不具有任何形状也不确实使其进行任何流体动力学拖拽。弹簧常数等于 $5 \text{ kJmol}^{-1} \text{ \AA}^{-2}$ 。

[0774] 结果

[0775] 单一G易位

[0776] 如图3中所示,拉力相对于时间的图展现,对于核苷酸进入野生型CsgG-Eco孔中的苯丙氨酸残基F56的环,存在很大的屏障。对于针对所研究的CsgG-Eco突变观测到的鸟嘌呤易位,无明显屏障。

[0777] ssDNA易位

[0778] 对于ssDNA易位,每孔进行两个模拟,其中各回合具有不同的施加的牵拉速度($100 \text{ \AA/ns}$ 和 $10 \text{ \AA/ns}$)。如图4中所展示,其示出较快牵拉速度模拟,CsgG野生型孔需要最大的拉力来实现ssDNA易位。如图5所示,其示出较慢牵拉速度模拟,CsgG-Eco(野生型,SEQ ID NO:2)和CsgG-Eco-(F56A)孔两个均需要最大的施加力来实现ssDNA易位。在ssDNA易位通过CsgG和MspA基线孔所需的拉力之间进行比较,表明CsgG孔的突变需要允许类似的ssDNA易位的水平。

[0779] 实例2

[0780] 本实例描述若干CsgG突变的表征。

[0781] 材料和方法

[0782] 在设置实验之前,将DNA构筑体X(最终浓度 0.1 nM ,参见图13对于构筑体X的草图表示和描述)与T4 Dda-E94C/C109A/C136A/A360C(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:24,添加到纳米孔系统的最终浓度是 10 nM ,其是以于缓冲液(151.5 mM KCl , 25 mM 磷酸钾, 5% 甘油, $\text{pH} 7.0$, 1 mM EDTA)中的形式提供),在室温下预先培育五分钟。在五分钟后,向预混合物中添加TMAD($100 \mu\text{M}$),且将混合物另外培育5分钟。最终,向预混合物中添加 MgCl_2 (添加到纳米孔系统的最终浓度是 1.5 mM)、ATP(添加到纳米孔系统的最终浓度是 1.5 mM)、KCl(添加到纳米孔系统的最终浓度是 500 mM)和磷酸钾缓冲液(添加到纳米孔系统的最终浓度是 25 mM)。

[0783] 从插入在含嵌段共聚物的缓冲液(25 mM 磷酸钾缓冲液, 150 mM 亚铁(II) 氰化钾, 150 mM 铁(III) 氰化钾, $\text{pH} 8.0$)中的各种单一CsgG纳米孔获得电测量。在达到插入在嵌段共聚物中的单一孔之后,接着缓冲液(2 mL , 25 mM 磷酸钾缓冲液, 150 mM 亚铁(II) 氰化钾, 150 mM 铁(III) 氰化钾, $\text{pH} 8.0$)流过系统,移除任何过量的CsgG纳米孔。接着,使 $150 \mu\text{L}$ 的 500 mM KCl 、 25 mM 磷酸钾、 1.5 mM MgCl_2 、 1.5 mM ATP 、 $\text{pH} 8.0$ 流过系统。在10分钟之后,使 $150 \mu\text{L}$ 的 500 mM KCl 、 25 mM 磷酸钾、 1.5 mM MgCl_2 、 1.5 mM ATP 、 $\text{pH} 8.0$ 流过系统,且接着酶(T4 Dda-

E94C/C109A/C136A/A360C, 10nM最终浓度)、DNA构筑体X (0.1nM, 最终浓度)、燃料 (MgCl₂, 1.5mM最终浓度; ATP, 1.5mM最终浓度) 的预混合物(总共150μL) 流动到单一纳米孔实验系统中。实验是在-120mV下进行, 且进行解螺旋酶控制的DNA移动监测。

[0784] 结果

[0785] 显示范围增加的孔(图6(a) - (c) 到图8(a) - (c) 和图18到30)

[0786] CsgG-Eco- (StrepII (C)) (SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 具有~10pA的范围 (参见图6(a)), 而以下的CsgG-Eco孔突变呈现增加的电流范围:

[0787] 1-CsgG-Eco- (Y51N-F56A-D149N-E185R-E201N-E203N-StrepII (C)) 9 (具有突变Y51N/F56A/D149N/E185R/E201N/E203N的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~30pA的范围 (参见图6(b))。

[0788] 2-CsgG-Eco- (N55A-StrepII (C)) 9 (具有突变N55A的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~35pA的范围 (参见图6(c))。

[0789] 3-CsgG-Eco- (N55S-StrepII (C)) 9 (具有突变N55S的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~40pA的范围 (参见图7(a))。

[0790] 4-CsgG-Eco- (Y51N-StrepII (C)) 9 (具有突变Y51N的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~40pA的范围 (参见图7(b))。

[0791] 5-CsgG-Eco- (Y51A-F56A-StrepII (C)) 9 (具有突变Y51A/F56A的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~30pA的范围 (参见图7(c))。

[0792] 6-CsgG-Eco- (Y51A-F56N-StrepII (C)) 9 (具有突变Y51A/F56N的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~20pA的范围 (参见图8(a))。

[0793] 7-CsgG-Eco- (Y51A-N55S-F56A-StrepII (C)) 9 (具有突变Y51A/N55S/F56A的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~30pA的范围 (参见图8(b))。

[0794] 8-CsgG-Eco- (Y51A-N55S-F56N-StrepII (C)) 9 (具有突变Y51A/N55S/F56N的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~30pA的范围 (参见图8(c))。

[0795] 13-CsgG-Eco- (F56H-StrepII (C)) 9 (具有突变F56H的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~35pA的范围 (参见图18)。

[0796] 14-CsgG-Eco- (F56Q-StrepII (C)) 9 (具有突变F56Q的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~40pA的范围 (参见图19)。

[0797] 15-CsgG-Eco- (F56T-StrepII (C)) 9 (具有突变F56T的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~35pA的范围 (参见图20)。

[0798] 16-CsgG-Eco- (S54P/F56A-StrepII (C)) 9 (具有突变S54P/F56A的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~35pA的范围 (参见图21)。

[0799] 17-CsgG-Eco- (Y51T/F56A-StrepII (C)) 9 (具有突变Y51T/F56A的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~30pA的范围 (参见图22)。

[0800] 18-CsgG-Eco- (F56P-StrepII (C)) 9 (具有突变F56P的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 呈现~30pA的范围 (参见图23)。

[0801] 19-CsgG-Eco- (F56A-StrepII (C)) 9 (具有突变F56A的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)呈现~40pA的范围(参见图24)。

[0802] 20-CsgG-Eco- (Y51T/F56Q-StrepII (C)) 9 (具有突变Y51T/F56Q的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)呈现~30pA的范围(参见图25)。

[0803] 21-CsgG-Eco- (N55S/F56Q-StrepII (C)) 9 (具有突变N55S/F56Q的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)呈现~35pA的范围(参见图26)。

[0804] 22-CsgG-Eco- (Y51T/N55S/F56Q-StrepII (C)) 9 (具有突变Y51T/N55S/F56Q的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)呈现~35pA的范围(参见图27)。

[0805] 23-CsgG-Eco- (F56Q/N102R-StrepII (C)) 9 (具有突变F56Q/N102R的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)呈现~30pA的范围(参见图28)。

[0806] 24-CsgG-Eco- (Y51Q/F56Q-StrepII (C)) 9 (具有突变Y51Q/F56Q的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)呈现~40pA的范围(参见图29)。

[0807] 25-CsgG-Eco- (Y51A/F56Q-StrepII (C)) 9 (具有突变Y51A/F56Q的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)呈现~35pA的范围(参见图30)。

[0808] 显示输送量增加的孔(图9(a) - (c)和10(a) - (b))

[0809] 如可从图9(a) - (c)和10(a) - (b)看出,以下突变孔(以下9-12)呈现在4小时中多个解螺旋酶控制的DNA移动(在图9(a) - (c)和10(a) - (b)中标记为X)每通道,而CsgG-Eco- (StrepII (C)) (SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)(a)通常呈现在4小时中仅1或2个解螺旋酶控制的DNA移动(在图9(a)中标记为X)每通道,且相反呈现延长的块区(在图9(a)中标记为Y)。

[0810] 9-CsgG-Eco- (D149N-E185N-E203N-StrepII (C)) 9 (具有突变D149N/E185N/E203N的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)(图9(b))

[0811] 10-CsgG-Eco- (D149N-E185N-E201N-E203N-StrepII (C)) 9 (具有突变D149N/E185N/E201N/E203N的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)(图9(c))

[0812] 11-CsgG-Eco- (D149N-E185R-D195N-E201N-E203N) -StrepII (C)) 9 (具有突变D149N/E185R/D195N/E201N/E203N的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)(图10(a))

[0813] 12-CsgG-Eco- (D149N-E185R-D195N-E201R-E203N) -StrepII (C)) 9 (具有突变D149N/E185R/D195N/E201R/E203N的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)(图10(b))

[0814] 显示插入增加的孔(图11和12)

[0815] 如可通过比较图11和12看出,图12中显示的突变孔CsgG-Eco- (T150I-StrepII (C)) 9 (具有突变T150I的SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接),存在于膜中,所述膜的孔数量相较于CsgG-Eco- (StrepII (C)) (SEQ ID NO:2,其中StrepII (C)是SEQ ID NO:47,且在C端处连接)孔(图11中显示)增加(~4-5倍)。图11和12中的箭头说明在4小时实验中插入到嵌段共聚物中的CsgG-Eco纳米孔的数量(图11中,130-140;且在图12中1-11,各自对应于独立纳米孔实验)。多余CsgG-Eco- (StrepII (C)) (SEQ ID NO:2,其

中StrepII(C)是SEQ ID NO:47,且在C端处连接),三个实验显示插入一个纳米孔,而对于突变孔(CsgG-Eco-(T150I-StrepII(C))9),各实验显示插入至少一个纳米孔,且若干实验显示多个孔插入。

[0816] 实例3

[0817] 本实例描述研究用于纯化CsgG孔的大肠杆菌纯化方法。

[0818] 材料和方法

[0819] 在美国金斯瑞公司(GenScript USA Inc.)中合成编码多肽Pro-CsgG-Eco-(StrepII(C))(SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接,且其中Pro是SEQ ID NO:48,且在N端处连接)的DNA,且克隆到含有胺苄青霉素抗性基因的pT7载体中。通过异丙基 β -D-1-硫代半乳糖吡喃糖苷(IPTG)来诱导pT7载体的蛋白表达。将DNA溶液的浓度调整到400ng/ μ L。DNA(1 μ L)用于转化Lemo21(DE3)感受态大肠杆菌细胞(50 μ L,NEB,目录号C2528H)。在转化之前,从Lemo21(DE3)细胞(德国的基因桥公司(Gene Bridges GmbH))敲除CsgG基因。接着,将细胞向外涂铺在含有胺苄青霉素(0.1mg/mL)的LB琼脂上,且在37°C下培育大致16小时。

[0820] 使细菌菌落在含有胺苄青霉素的LB板上生长,并入CsgG质粒。一个此类菌落用于接种含有羧苄青霉素(0.1mg/mL)的LB培养基的起子培养物(100mL)。使起子培养物在37°C、搅动下生长,直到OD600达到1.0-1.2为止。起子培养物用于接种含有羧苄青霉素(0.1mg/mL)和鼠李糖(500 μ M)的新鲜500mL LB培养基,直到O.D.600为0.1。使培养物在37°C、搅动下生长,直到OD600达到0.6为止。接着,调整培养物的温度到18°C,且通过添加IPTG(0.2mM最终浓度)来起始诱导。在18°C、搅动下,进行诱导大致18小时。

[0821] 在诱导之后,通过在6,000g下离心30分钟,来集结培养物。将集结粒再悬浮于50mM Tris、300mM NaCl、含有蛋白酶抑制剂(默克密理博公司539138)、benzonase核酸酶(西格玛公司E1014)和1Xbugbuster(默克密理博公司70921)pH8.0(大致10mL缓冲液每克集结粒)中。将悬浮液充分混合,直到其完全均匀为止,接着将样本转移到4°C下的辊式混合器大致5小时。通过在20,000g下离心45分钟来使裂解物集结,且通过0.22 μ m PES针筒过滤器来过滤上清液。向上获取含有CsgG的上清液(称为样本1),以用于通过柱色谱纯化。

[0822] 样本1应用于5mL Strep Trap柱(通用电气医疗集团)。用25mM Tris、150mM NaCl、2mM EDTA、0.01%DDM pH8来洗涤柱,直到维持10柱体积的稳定基线为止。接着,在返回到150mM缓冲液之前,用25mM Tris、2M NaCl、2mM EDTA、0.01%DDM pH8来洗涤柱。用10mM脱硫生物素进行洗脱。CsgG蛋白的Strep trap(通用电气医疗集团)纯化的色谱迹线的实例显示于图14中。洗脱峰标记为E1。图15显示在初始strep纯化之后CsgG-Eco蛋白的典型SDS-PAGE观测的实例。泳道1-3显示含有如由箭头指示的CsgG蛋白的主要洗脱峰(图14中的标记E1)。泳道4-6对应于含有杂质的主要洗脱峰(图14中的标记E1)尾部的洗脱份。

[0823] 合并洗脱峰值,且加热到65°C维持15分钟以移除热不稳定的受污染的蛋白质。在20,000g下对加热的溶液进行离心10分钟,且丢弃集结粒。在120mL葡聚糖凝胶S200柱(通用电气医疗集团)上在25mM Tris、150mM NaCl、2mM EDTA、0.01%DDM、0.1%SDS pH8中对上清液进行凝胶过滤。在220nm下进行监测,这是归因于蛋白质的低色氨酸组成。在大致55mL体积下洗脱样本(图16显示55mL样本峰(用星形标记)的尺寸排阻柱迹线)。在4-20%TGX上进行洗脱峰(参见图17,伯乐公司),以确认所关注的CsgG-Eco-(StrepII(C))(SEQ ID NO:2,

其中StrepII (C) 是SEQ ID NO:47,且在C端处连接)的孔的存在。合并经鉴定级分,且通过50kD Amicon离心柱浓缩。

[0824] 实例4

[0825] 本实例描述为研究CsgG-Eco- (Y51T/F56Q) -StrepII (C)) 9(具有突变Y51T/F56Q的SEQ ID NO:2,其中StrepII (C) 是SEQ ID NO:47,且在C端处连接,孔突变第20号) 与T4Dda- (E94C/C109A/C136A/A360C) (具有突变E94C/C109A/C136A/A360C且接着(Δ M1)G1G2的SEQ ID NO:24) 之间的相互作用进行的模拟。

[0826] 模拟方法

[0827] 使用GROMACS程序包4.0.5版,利用GROMOS 53a6力场和SPC水模型来进行模拟。

[0828] CsgG-Eco- (Y51T/F56Q) -StrepII (C)) 9(具有突变Y51T/F56Q的SEQ ID NO:2,其中StrepII (C) 是SEQ ID NO:47,且在C端处连接,孔突变第20号) 模型是基于在蛋白质数据库登录号4UV3和4Q79中发现的CsgG的晶体结构。使用PyMOL制备相关突变。接着,使用最陡下降算法,使所得孔模型能量降至最低。T4 Dda- (E94C/C109A/C136A/A360C)) (具有突变E94C/C109A/C136A/A360C且接着(Δ M1)G1G2的SEQ ID NO:24) 模型是基于蛋白质数据库登录号3UPU中发现的Dda1993结构。同样,使用PyMOL制备相关突变,且使用最陡下降算法使模型能量降至最低。

[0829] 接着,将T4 Dda- (E94C/C109A/C136A/A360C) (具有突变E94C/C109A/C136A/A360C且接着(Δ M1)G1G2的SEQ ID NO:24) 模型放置在CsgG-Eco- (Y51T/F56Q) -StrepII (C)) 9(具有突变Y51T/F56Q的SEQ ID NO:2,其中StrepII (C) 是SEQ ID NO:47,且在C端处连接,孔突变第20号) 上。用不同的初始酶构形进行三个模拟(回合1到3(0ns),参见图31):

[0830] 在所有酶构形中,定位酶,使得DNA的5'端指向孔,且酶在整个模拟中不受限制。孔主链受限制,且使模拟盒溶剂化。使用到300K的Berendsen恒温器和Berendsen恒压器,在NPT集中模拟系统40ns。

[0831] 使用GROMACS分析软件以及本地写入码两个来分析酶与孔之间的接触。下表显示对于孔和酶氨基酸两个所观测到的接触的数量。表6-8显示孔上的氨基酸接触点,所述氨基酸接触点与酶上的氨基酸接触点相互作用。在三分之二的模拟中,酶在孔上倾斜(参见回合2和3(20、30和40ns),图31和32)。回合1显示,酶尚未倾斜,且因此在表6中显示具有高的相互作用的点可进行优化,以增加孔帽上的酶稳定性。

[0832]

孔	酶	接触数
ASN 102	ASP 198	8200
ASN 102	TYR 438	8130
GLN 100	ASP 212	7369
GLU 101	TRP 195	5979
ARG 97	TYR 350	4873
GLU 101	LEU 215	4851
ASN 102	TRP 195	3988
ARG 97	TYR 415	3798
GLU 101	TYR 350	3759
LEU 113	ASP 212	3718
ASN 102	LYS 358	3124
ARG 97	GLY 211	2765
GLU 101	CYS 412	2715
ARG 97	GLY 193	2708
ASN 102	ILE 196	2342
GLU 101	TYR 415	2268
GLU 101	ARG 216	2158
ARG 110	THR 213	2094
ARG 110	ASP 212	2066
GLY 103	ARG 216	1456
GLU 101	TYR 318	1333
ASN 102	GLU 347	1316
GLU 101	LYS 194	1310
ARG 97	PRO 411	1203
GLU 101	LYS 358	1161
ASN 102	ARG 216	1132

[0833]

ARG 97	TRP 195	888
LYS 94	TYR 415	793
ASN 102	PRO 315	696
ASN 102	LYS 247	541
GLU 101	ALA 214	449
ASN 102	ASP 346	440
ARG 97	ALA 214	366
ARG 97	LYS 194	336
GLU 101	ASP 212	302
ARG 97	VAL 439	267
ARG 110	THR 210	263
ARG 97	THR 210	259
ARG 97	GLN 422	257
GLU 101	TYR 409	228
ALA 98	TRP 195	207
GLU 101	LYS 247	201
ASN 102	GLU 317	179
ARG 110	ARG 216	147
ARG 97	ASP 212	108
ASN 102	VAL 314	87
GLU 101	THR 213	72
ASN 102	LYS 255	70
VAL 105	ARG 216	69
ASN 102	LEU 215	59
ASN 102	THR 210	55
ILE 111	ASP 212	48
ARG 97	HIS 414	48
THR 104	ARG 216	36
ASN 102	TYR 197	32
GLN 100	THR 213	30
ASN 102	GLU 361	28
ARG 97	VAL 418	28
ALA 98	TYR 415	27
GLU 101	LEU 354	17
GLU 101	TYR 197	16
ASN 102	GLY 316	16

[0834]

ARG 97	GLU 361	16
ARG 97	GLU 347	14
ILE 107	ARG 216	12
ASN 102	GLY 208	12
ARG 97	TYR 409	11
ARG 97	LYS 247	11
GLU 101	LYS 364	8
ARG 97	PHE 209	7
LYS 94	GLU 419	6
GLU 101	PRO 411	5
GLU 101	GLU 317	5
ASN 102	ILE 251	5
ARG 97	LEU 354	5
LYS 94	VAL 418	3
ASN 102	ARG 321	3
ARG 97	LYS 243	3
LYS 94	CYS 412	2
LEU 113	THR 210	2
GLY 103	GLU 317	2
GLU 101	LYS 351	2
ASN 102	TYR 318	2
ASN 102	MET 219	2
ASN 102	LYS 194	2
ARG 97	VAL 314	2
ARG 97	LYS 364	2
THR 104	PRO 315	1
GLY 103	THR 213	1
GLU 101	PRO 315	1

[0835] 表6=回合1,酶和孔接触相互作用

[0836]

孔	酶	接触数
GLU 101	THR 210	14155
SER 115	ASP 202	9477
ARG 97	THR 210	9064
ASN 102	VAL 200	5323
THR 104	ASP 202	4476

[0837]

ASN 102	ASN 221	3422
GLU 101	PHE 437	3171
ARG 97	ASP 217	2698
GLU 101	ARG 216	2198
ARG 97	GLY 208	1730
GLU 101	LYS 199	1710
SER 115	SER 224	1440
ASN 102	LYS 199	1351
ASN 102	ASP 212	1298
ASN 102	ARG 405	1219
GLU 101	ARG 207	1180
ASN 102	SER 224	1150
ASN 102	LYS 255	1114
ARG 97	ASP 198	946
GLU 101	PHE 209	931
ARG 97	THR 213	791
ARG 97	ARG 216	599
ASN 102	THR 210	589
GLN 114	ASP 202	530
ASN 102	ASP 202	492
ARG 97	ASP 212	490
GLY 103	ARG 405	474
THR 104	SER 224	451
GLU 101	LYS 255	429
ASN 102	ASP 198	405
ASN 102	PHE 209	400
ASN 102	ARG 178	316
ARG 110	GLU 258	309
ASN 102	ASN 180	257
GLN 100	PHE 223	256
GLU 101	TYR 197	220
GLN 114	SER 228	212
LEU 113	PHE 223	210
ASN 102	ILE 225	204
GLN 114	LYS 227	194
GLU 101	GLY 211	189

[0838]

GLU 101	ASP 212	174
LEU 113	SER 224	159
LEU 113	GLY 203	145
ARG 97	VAL 220	134
GLU 101	THR 213	133
THR 104	SER 228	125
ARG 97	TYR 197	123
LYS 94	ASP 212	118
ASN 102	ARG 216	110
ASN 102	ASN 235	108
ASN 102	GLY 211	104
GLU 101	ARG 405	79
GLN 114	SER 224	69
ASN 102	VAL 220	63
LEU 113	LYS 227	49
ASN 102	VAL 201	42
ARG 97	PHE 209	42
GLU 101	ASN 180	40
ARG 97	TYR 438	38
ARG 97	ARG 207	32
ASN 102	PHE 407	28
SER 115	ASN 221	23
ARG 110	HIS 204	22
GLU 101	PHE 223	21
ARG 97	ASP 189	19
ARG 110	PHE 223	16
THR 104	ILE 225	13
GLY 103	ASN 180	11
ARG 97	LYS 194	11
GLU 101	PHE 407	10
ARG 97	MET 219	9
THR 104	ASN 235	8
ARG 110	ARG 405	8
ARG 97	TRP 195	7
ILE 111	PHE 223	6
GLU 101	GLY 208	6

[0839]	LEU 113	ASP 202	5
	GLU 101	ARG 178	5
	ASN 102	THR 213	5
	ALA 98	ARG 216	5
	ASN 102	ASP 217	4
	ARG 97	LYS 199	4
	THR 104	LEU 229	3
	THR 104	ARG 405	3
	GLU 101	VAL 201	3
	GLU 101	MET 219	3
	ARG 110	ASP 202	3
	ARG 110	ARG 207	2
	THR 104	VAL 201	1
	GLY 103	SER 224	1
	GLY 103	LYS 255	1
	GLY 103	GLU 258	1
	GLY 103	ASN 235	1
	GLU 101	ASP 198	1
	ASN 102	PHE 437	1
	ARG 97	PHE 437	1
	ARG 110	LYS 227	1

[0840] 表7=回合2,酶和孔接触相互作用

[0841]	孔	酶	接触数
	ARG 97	THR 174	15557
	GLN 100	ASP 5	10353
	GLU 101	LYS 177	9238
	ARG 97	SER 179	6630
	LEU 116	ASP 202	6545
	GLU 101	TYR 434	6524
	SER 115	ASP 202	5693
	GLU 101	HIS 204	5457
	ARG 97	GLN 10	5106
	ARG 93	ASP 202	4646
	ARG 93	GLU 8	4446
	SER 115	LYS 11	4342

[0842]

LEU 113	ASP 5	3871
ASN 102	SER 224	3605
GLU 101	ASN 12	3344
GLU 101	GLN 10	3327
ARG 97	GLU 175	3096
GLU 101	SER 224	3028
LEU 116	GLU 8	2936
LYS 94	ASP 185	2708
ARG 97	ASN 180	2700
GLU 101	PHE 3	2500
THR 104	LYS 11	2352
SER 115	GLU 8	2323
ARG 93	ASN 180	1912
ASN 102	LYS 177	1838
LYS 94	ASP 198	1828
ARG 110	ASP 5	1714
ALA 98	GLY 203	1701
ASN 102	ASN 12	1695
GLU 101	TYR 169	1691
ARG 97	THR 7	1593
ARG 110	ASP 4	1404
ARG 97	ASP 212	1381
ASN 102	HIS 204	1226
ASN 102	ASN 15	1173
ARG 97	VAL 176	1096
ALA 98	HIS 204	998
ARG 97	ASP 202	875
ASN 102	TYR 434	850
ALA 98	ASN 12	716
GLU 101	THR 213	702
GLU 101	ARG 178	642
GLU 101	ASN 221	600
ASN 102	LYS 11	588
ARG 97	ASP 217	585
ARG 97	ARG 207	537
GLU 101	ARG 207	525

[0843]

ARG 97	PHE 437	511
GLU 101	ARG 216	510
ASN 102	LYS 19	482
ARG 97	HIS 204	473
LEU 113	LYS 11	409
ARG 97	THR 213	358
ARG 93	ASP 212	354
ARG 97	TYR 169	316
ARG 97	GLY 203	308
ARG 97	ASP 435	300
GLN 87	LYS 199	249
THR 104	ASN 15	221
ARG 97	ALA 181	220
ASN 102	LYS 227	198
LYS 94	ARG 178	184
ASN 102	GLU 8	183
LEU 113	LEU 6	182
ARG 93	SER 179	179
LEU 90	ASN 180	172
LEU 90	ASP 202	144
ARG 97	ILE 225	138
GLU 101	ASN 15	135
GLU 101	LYS 19	113
LYS 94	ASN 180	109
LYS 94	GLU 175	105
ARG 93	THR 7	81
LYS 94	ARG 207	77
GLN 100	PHE 3	72
ASN 102	ARG 216	66
ARG 97	LYS 177	62
GLU 101	THR 210	59
ARG 97	ARG 178	56
LYS 94	ASP 212	55
ARG 97	GLU 172	53
GLU 101	VAL 176	51
ALA 98	ARG 207	49

[0844]

ARG 110	PHE 3	48
ALA 98	ASP 202	47
ARG 97	VAL 200	40
ALA 98	VAL 201	36
LYS 94	THR 210	35
ILE 111	ASP 5	32
ARG 97	ARG 405	27
LEU 90	VAL 200	26
ARG 97	THR 210	26
GLY 103	PHE 3	25
GLU 101	PHE 209	25
ARG 97	ARG 216	22
ASN 102	VAL 220	21
LYS 94	GLY 211	19
ARG 97	PHE 209	17
GLU 101	LYS 227	15
GLN 114	LYS 11	15
GLY 103	LYS 19	13
ARG 97	PHE 3	13
GLU 101	THR 2	12
GLU 101	ILE 225	12
ARG 97	ILE 184	12
ALA 98	GLU 8	12
ALA 98	ARG 178	12
ASN 102	ILE 225	11
LYS 94	LYS 199	10
GLU 101	ARG 433	8
ARG 97	ASN 221	8
LYS 94	VAL 200	7
ASN 102	ASP 202	7
ASN 102	ASN 221	7
ARG 97	LEU 173	7
SER 115	HIS 204	6
ASN 102	GLY 203	6
GLU 101	CYS 171	5
ARG 97	ASN 12	5

[0845]	ASN 102	PHE 223	4
	ASN 102	LYS 166	4
	ARG 97	GLY 211	4
	ARG 97	GLN 170	4
	GLU 101	ARG 405	3
	ASN 102	PHE 3	3
	GLU 101	GLU 175	2
	ARG 97	VAL 220	2
	ARG 93	GLY 203	2
	LYS 94	THR 174	1
	LEU 90	LYS 199	1
	LEU 116	ASN 180	1
	LEU 113	ASP 212	1
	LEU 113	ASP 202	1
	GLY 103	ASN 15	1
	GLU 101	THR 7	1
	GLU 101	PHE 437	1
	GLN 114	ASP 202	1
	ASN 102	ARG 405	1
	ARG 97	TYR 434	1
	ARG 97	PRO 182	1
	ARG 97	GLY 9	1
	ARG 97	GLU 8	1
	ALA 99	ASP 202	1

[0846] 表8=回合3,酶和孔接触相互作用

[0847] 实例5

[0848] 本实例描述为了研究以下之间的相互作用而进行的模拟:a) CsgG-Eco- (Y51A/F56Q) -StrepII(C))9(具有突变Y51A/F56Q的SEQ ID N0:2,其中StrepII(C)是SEQ ID N0:47,且在C端处连接,孔突变第25号)与T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C) (具有突变E94C/F98W/C109A/C136A/K194L/A360C且接着(Δ M1)G1G2的SEQ ID N0:24)之间,和b) CsgG-Eco- (Y51A/F56Q/R97W) -StrepII(C))9(具有突变Y51A/F56Q/R97W的SEQ ID N0:2,其中StrepII(C)是SEQ ID N0:47,且在C端处连接,孔突变第26号)与T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C) (具有突变E94C/F98W/C109A/C136A/K194L/A360C且接着(Δ M1)G1G2的SEQ ID N0:24)之间。

[0849] 模拟方法

[0850] 如实例4中所描述进行模拟。

[0851] CsgG-Eco- (Y51A/F56Q) -StrepII (C)) 9 (具有突变Y51A/F56Q的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接, 孔突变第25号) 和CsgG-Eco- (Y51A/F56Q/R97W) -StrepII (C)) 9 (具有突变Y51A/F56Q/R97W的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接, 孔突变第26号) 模型, 是基于在蛋白质数据库登录号4UV3和4Q79中发现的CsgG的晶体结构。使用PyMOL制备相关突变。接着, 使用最陡下降算法, 使所得孔模型能量降至最低。

[0852] T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C) (具有突变E94C/F98W/C109A/C136A/K194L/A360C且接着(Δ M1)G1G2的SEQ ID NO:24) 模型, 是基于在蛋白质数据库登录号3UPU中发现的Dda1993结构。同样, 使用PyMOL制备相关突变, 且使用最陡下降算法使模型能量降至最低。

[0853] 接着, 将T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C) (具有突变E94C/F98W/C109A/C136A/K194L/A360C且接着(Δ M1)G1G2的SEQ ID NO:24) 模型放置在突变孔25和26上。

[0854] 在模拟中, 定位酶, 使得DNA的5'端指向孔, 且酶在整个模拟中不受限制。对于所研究的突变孔中的每一个, 进行两个模拟: 首先, 孔主链受限制, 且将模拟盒溶剂化; 其次, 除帽区以外孔主链受限制, 且将模拟盒溶剂化。使用到300K的Berendsen恒温器和Berendsen恒压器, 在NPT集中模拟系统40ns。

[0855] 使用GROMACS分析软件以及本地写入码两个来分析酶与孔之间的接触。下表显示对于孔和酶氨基酸两个所观测到的接触的数量 (对于突变25和26, 其具有T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C))。表9 (孔主链受限制) 和10 (孔主链受限制, 其中帽区不受限制) 显示孔突变25上的氨基酸接触点和用酶 (T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C)) 制备的接触的接触数量。表11 (孔主链受限制) 和12 (孔主链受限制, 其中帽区不受限制) 显示孔突变25 (CsgG-Eco- (Y51A/F56Q) -StrepII (C)) 9 (具有突变Y51A/F56Q的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 上的氨基酸接触点, 所述氨基酸接触点与酶 (T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C)) 上的氨基酸接触点相互作用。表13 (孔主链受限制) 和14 (孔主链受限制, 其中帽区不受限制) 显示孔突变26上的氨基酸接触点和用酶 (T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C)) 制备的接触的接触数量。表15 (孔主链受限制) 和16 (孔主链受限制, 其中帽区不受限制) 显示孔突变26 (CsgG-Eco- (Y51A/F56Q/R97W) -StrepII (C)) 9 (具有突变Y51A/F56Q/R97W的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 上的氨基酸接触点, 所述氨基酸接触点与酶 (T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C)) 上的氨基酸接触点相互作用。图33显示孔突变26和T4 Dda- (E94C/F98W/C109A/C136A/K194L/A360C) 的初始快照。

[0856]	氨基酸	位置	接触的数量
	GLU	101	26081
	ARG	97	12985

[0857]

GLN	100	6485
ASN	102	4941
LEU	113	3923
LYS	94	3159
ARG	110	2348
GLN	114	617
ILE	111	195
ALA	98	74
ILE	107	67
THR	104	24
PRO	112	21
SER	115	17
GLY	103	17
GLN	87	13
LEU	90	6
ASN	108	3

[0858] 表9

[0859]

氨基酸	位置	接触的数量
ARG	97	17462
GLU	101	12262
ASN	102	9008
LYS	94	5982
GLN	100	3471
GLY	103	3366
LEU	113	1511
GLN	87	734
THR	104	500
ARG	110	420
ASN	91	273
GLN	114	178
ALA	98	94
ARG	93	44
ILE	111	21
PRO	112	13
LEU	90	13

[0860]

SER	115	4
VAL	105	2
ALA	99	2
ILE	95	1

[0861] 表10

[0862]

孔		酶		接触的数量
氨基酸	位置	氨基酸	位置	
GLU	101	TYR	318	7116
GLU	101	THR	210	6306
GLN	100	ASN	365	4693
GLU	101	TRP	195	4608
LEU	113	ASN	367	3910
LYS	94	ASP	212	2369
ASN	102	LYS	371	2088
GLU	101	LYS	364	1903
ARG	97	SER	224	1883
ARG	97	TYR	438	1764
ARG	110	LYS	368	1727
GLU	101	LYS	358	1443
GLN	100	ASN	367	1390
GLU	101	GLN	422	1356
ARG	97	TRP	195	1318
ARG	97	PHE	209	1095
ARG	97	TYR	318	965
ARG	97	TYR	350	895
ASN	102	LYS	358	833
ARG	97	GLU	361	818
ARG	97	GLY	208	706
GLU	101	GLY	316	697
GLU	101	GLU	317	678
ARG	97	GLU	317	647
GLN	114	ASN	367	617
GLU	101	LYS	371	506
ASN	102	TRP	366	466
ARG	97	THR	210	463

[0863]

ASN	102	ARG	207	457
LYS	94	ASP	217	423
GLU	101	ARG	207	409
ASN	102	GLU	419	405
GLN	100	LYS	368	402
GLU	101	ARG	321	354
ARG	110	GLY	369	332
ARG	97	GLN	422	324
ASN	102	GLY	313	313
ARG	97	ASN	221	306
ARG	97	VAL	439	301
ARG	110	ASN	365	274
SER	257	GLU	258	267
GLU	101	LYS	310	217
ILE	111	ASN	367	195
ARG	97	ASP	198	154
GLU	101	SER	224	154
ARG	97	VAL	418	145
ARG	97	ALA	157	137
GLU	101	GLY	208	136
ARG	97	GLU	419	135
LYS	94	TYR	415	111
ARG	97	VAL	220	105
ASN	102	PRO	315	104
ASN	102	GLU	93	103
ARG	97	LYS	364	97
LYS	94	TYR	350	94
ARG	97	GLU	154	92
ARG	97	ALA	416	82
GLU	101	PHE	223	76
ILE	107	LYS	368	67
ARG	97	GLU	347	66
ARG	97	TYR	415	60
ASN	102	TYR	197	57
LYS	94	GLY	211	49
LYS	94	THR	213	47

[0864]

GLU	101	ASN	365	46
ARG	97	LEU	354	41
LYS	94	ARG	216	41
ALA	98	TRP	195	37
ASN	102	TRP	195	34
ALA	98	LYS	358	32
ASN	102	LEU	194	27
THR	104	TRP	366	24
ARG	97	GLY	211	23
PRO	112	ASN	367	21
GLU	101	PHE	209	21
ASN	102	LYS	364	20
LYS	94	ALA	416	19
SER	115	TRP	366	17
GLY	103	LEU	194	17
ARG	97	ASP	212	16
GLU	101	TYR	415	15
ARG	110	GLY	370	15
ARG	97	LYS	351	14
ASN	102	LYS	310	14
ASN	102	GLU	154	14
ARG	97	CYS	360	13
ARG	97	THR	156	12
GLU	101	LEU	194	11
GLN	87	ASP	212	9
ARG	97	PHE	308	8
GLU	101	LEU	354	8
LEU	113	TRP	366	7
ARG	97	ARG	321	7
GLU	101	ALA	157	7
LEU	113	ASN	365	6
GLU	101	GLY	193	6
ARG	97	ASP	217	5
ALA	98	THR	210	5
GLN	87	ARG	216	4
ARG	97	LEU	319	4

[0865]

ARG	97	ASN	155	4
LYS	94	VAL	220	4
LEU	90	TYR	415	3
LEU	90	ASP	212	3
ARG	97	PHE	223	3
ARG	97	GLY	193	3
ARG	97	ALA	421	3
ASN	108	LYS	368	3
ASN	102	GLY	316	3
GLU	101	VAL	418	3
ARG	97	ILE	159	2
GLU	101	PRO	315	2
SER	257	ASP	260	1
ARG	97	GLY	357	1
ASN	102	THR	210	1
ASN	102	ILE	196	1
ASN	102	GLU	317	1
GLU	101	LYS	227	1
GLU	101	GLU	361	1
GLU	101	ASN	221	1
LYS	94	VAL	418	1
LYS	94	GLU	154	1

[0866] 表11

[0867]

孔		酶		接触的数量
氨基酸	位置	氨基酸	位置	
ARG	97	ASN	365	3386
GLU	101	GLY	316	2967
GLY	103	ASP	189	2773
LYS	94	ASP	212	2524
GLU	101	GLU	317	2401
ASN	102	TYR	197	2377
GLN	100	ASN	365	2241
GLU	101	LYS	358	2110
ARG	97	GLU	317	2031
GLU	101	THR	210	2025

[0868]

ASN	102	ASP	189	1985
LYS	94	THR	213	1689
LEU	113	ASN	367	1509
ARG	97	GLU	361	1421
ARG	97	TRP	195	1361
ARG	97	TYR	438	1315
ARG	97	VAL	439	1145
LYS	94	ASP	217	957
GLU	101	TYR	318	852
GLU	101	ASN	365	797
SER	257	ASP	260	784
ARG	97	GLY	193	744
ARG	97	GLU	419	738
ARG	97	PHE	209	731
ASN	102	ARG	321	719
ASN	102	TRP	195	708
ASN	102	LYS	371	686
ARG	97	GLN	422	637
GLN	87	ASP	212	603
ARG	97	TYR	318	603
ARG	97	ASN	367	575
ARG	97	ASP	198	543
GLN	100	ASN	367	517
LYS	94	TYR	350	470
ARG	97	ALA	157	446
ARG	97	GLY	208	444
SER	257	GLU	258	425
GLU	101	LYS	371	407
GLU	101	LYS	364	400
ASN	102	GLU	93	389
GLN	100	LYS	368	383
ARG	110	LYS	368	357
ARG	97	THR	210	303
ASN	91	ARG	216	273
THR	104	ASN	367	214
GLY	103	TYR	197	204

[0869]

GLN	114	ASN	367	178
GLU	101	ARG	207	173
ASN	102	TRP	366	171
ASN	102	GLY	369	153
ASN	102	LEU	194	123
ASN	102	GLY	313	120
ARG	97	THR	362	100
THR	104	TRP	366	99
GLY	103	THR	164	97
LYS	94	TRP	195	94
GLN	100	GLY	193	84
GLY	103	LYS	177	81
ASN	102	LYS	166	80
ARG	97	ILE	159	78
LYS	94	ALA	416	76
GLN	87	ARG	216	75
ASN	102	SER	306	73
ASN	102	ARG	207	73
ALA	98	TRP	195	68
ARG	97	TYR	158	65
ASN	102	GLU	347	65
ASN	102	GLN	170	65
SER	257	LYS	261	62
ARG	97	ASN	192	62
GLN	100	GLY	369	61
THR	104	LYS	199	60
ARG	97	LEU	194	57
ASN	102	PHE	163	54
ASN	102	GLU	348	53
GLY	103	GLY	313	53
GLN	100	GLU	419	53
ASN	102	LYS	351	52
GLN	87	THR	213	51
LYS	94	GLY	211	49
GLN	100	GLN	423	49
THR	104	LYS	166	48

[0870]

GLU	101	ARG	321	45
ARG	97	PHE	308	44
ARG	93	GLU	419	44
ASN	102	GLY	370	41
ARG	110	GLY	369	41
SER	257	THR	259	40
ARG	97	GLY	211	40
ARG	97	THR	156	39
ASN	102	THR	164	39
LYS	94	ARG	216	39
ARG	97	ARG	321	38
GLY	103	LYS	166	35
ASN	102	GLY	208	33
GLU	101	GLY	370	30
LYS	94	GLU	361	28
ASN	102	SER	224	27
GLU	101	ILE	159	25
THR	104	PHE	308	23
THR	104	ASN	365	21
ASN	102	PRO	315	21
LYS	94	GLU	154	21
ILE	111	ASN	367	21
ARG	110	ILE	159	20
GLY	103	ARG	321	20
GLN	100	LEU	194	20
ARG	97	GLU	347	18
ARG	97	ARG	122	16
GLY	103	TRP	366	16
GLY	103	GLY	369	15
PRO	112	ASN	367	13
ALA	98	ARG	216	13
ARG	97	SER	345	12
ASN	102	LYS	364	12
GLN	100	GLY	370	12
GLU	101	PRO	315	11
GLN	100	ILE	159	11

[0871]

GLN	100	GLU	317	11
LEU	90	VAL	418	10
ASN	102	VAL	220	10
ASN	102	LYS	255	10
ALA	98	LYS	358	10
THR	104	ASN	180	9
GLY	103	LEU	194	9
THR	104	LYS	368	8
LYS	94	TYR	415	8
GLN	100	THR	164	8
ARG	97	VAL	418	7
ARG	97	VAL	220	7
ARG	97	TRP	366	7
ASN	102	LEU	319	7
GLY	103	LYS	255	7
LYS	94	ILE	413	7
GLN	100	PHE	163	7
THR	104	LYS	255	6
THR	104	LYS	177	6
LYS	94	GLU	347	6
GLN	100	ASN	192	6
GLN	87	ALA	416	5
ARG	97	LYS	368	5
ARG	97	ASP	217	5
ASN	102	ARG	216	5
LYS	94	VAL	439	5
ASN	102	VAL	200	4
ASN	102	ASN	365	4
GLU	101	LYS	310	4
GLU	101	GLY	313	4
GLY	103	PHE	223	4
LYS	94	ASN	155	4
GLN	100	LEU	420	4
ARG	97	SER	224	3
ARG	97	ASP	417	3
ASN	102	THR	362	3

[0872]

SER	115	GLU	419	3
GLU	101	LEU	194	3
GLU	101	GLY	369	3
GLY	103	ASN	192	3
LYS	94	ASP	417	3
VAL	105	LYS	166	2
LEU	113	ASN	365	2
LEU	90	ASP	212	2
ARG	97	TYR	92	2
ARG	97	GLU	154	2
ARG	97	ARG	207	2
THR	104	TYR	197	2
THR	104	ASP	185	2
ASN	102	PHE	209	2
ASN	102	ILE	159	2
GLY	103	PHE	308	2
GLY	103	LEU	319	2
GLY	103	ILE	159	2
ALA	98	THR	210	2
ALA	99	ILE	159	2
GLN	100	THR	210	2
SER	257	LYS	254	1
ILE	95	ARG	216	1
LEU	90	THR	213	1
ARG	97	TYR	304	1
ARG	97	ASP	212	1
THR	104	GLY	369	1
THR	104	ASN	192	1
ASN	102	VAL	314	1
ASN	102	THR	210	1
ASN	102	PHE	223	1
ASN	102	LYS	199	1
SER	115	ASP	212	1
ARG	110	GLU	317	1
ARG	110	ASN	365	1
GLY	103	PRO	315	1

[0873]

GLY	103	ASN	367	1
ALA	98	ALA	157	1
LYS	94	VAL	418	1

[0874] 表12

[0875]	氨基酸	位置	接触的数量
	GLU	101	18418
	TRP	97	12195
	ASN	102	5232
	GLY	103	212
	ALA	98	113
	SER	115	89
	THR	104	57
	LYS	94	52
	LEU	113	11
	GLN	100	5
	ARG	110	5
	GLN	114	4
	ARG	93	1

[0876] 表13

[0877]	氨基酸	位置	接触的数量
	TRP	97	16770
	ASN	102	11609
	GLU	101	4947
	GLY	103	2211
	THR	104	2187
	GLN	100	1589
	LYS	94	686
	ALA	98	289
	SER	115	274
	ARG	110	251
	ARG	93	44
	ILE	95	14
[0878]	LEU	113	5
	ASN	91	5
	LEU	116	4
	VAL	105	1
[0879]	LEU	90	1

[0879] 表14

[0880]

孔		酶		接触的数量
氨基酸	位置	氨基酸	位置	
GLU	101	TRP	195	5230
TRP	97	GLY	211	4360
GLU	101	THR	210	3265
GLU	101	LYS	358	3046
GLU	101	GLN	422	2476
ASN	102	ASP	212	1980
TRP	97	ARG	216	1707
TRP	97	ASN	365	1445
TRP	97	GLU	361	944
GLU	101	LYS	368	937
TRP	97	GLU	419	909
GLU	101	ALA	157	906
ASN	102	LYS	368	842
GLU	101	TYR	318	764
TRP	97	THR	210	720
ASN	102	TRP	366	626
GLU	101	ARG	216	518
TRP	97	ASN	155	487
GLU	101	LYS	364	482
ASN	102	GLU	361	409
ASN	102	THR	210	292
GLU	101	ASN	365	284
GLU	101	GLY	211	261
TRP	97	TRP	366	253
ASN	102	TRP	195	243
ASN	102	ASN	365	239
TRP	97	GLN	422	230
TRP	97	VAL	418	205

[0881]

ASN	102	LYS	358	204
TRP	97	THR	156	195
GLY	103	GLY	193	152
TRP	97	GLU	154	148
ASN	102	THR	213	129
TRP	97	ASP	212	128
ALA	98	ASP	212	112
TRP	97	THR	213	112
TRP	97	ALA	157	101
SER	115	ASP	212	76
ASN	102	LEU	194	75
GLU	101	PRO	315	61
ASN	102	VAL	314	59
THR	104	TRP	366	56
TRP	97	TYR	350	48
GLU	101	GLY	193	46
TRP	97	LYS	364	45
ASN	102	PRO	315	45
GLU	101	THR	156	38
ASN	102	GLU	317	37
GLU	101	LEU	194	36
TRP	97	PHE	209	32
GLY	103	LEU	194	31
TRP	97	TYR	318	30
LYS	94	GLU	154	30
TRP	97	GLY	193	28
TRP	97	ALA	214	27
GLU	101	GLU	317	25
ASN	102	GLU	154	22
TRP	97	LEU	354	18
GLY	103	TRP	366	14
ALA	115	TRP	366	13
TRP	97	ASP	417	12
LEU	113	TRP	366	11
GLU	101	VAL	418	11
ASN	102	GLY	211	10

[0882]

GLU	101	ARG	207	10
GLY	103	LYS	368	9
GLU	101	VAL	314	8
LYS	94	VAL	418	8
ASN	102	LYS	166	7
GLU	101	LYS	166	7
GLY	103	ASN	365	6
TRP	97	LYS	358	5
ASN	102	LYS	364	5
ARG	110	ASN	365	5
LYS	94	TYR	350	5
GLN	114	TRP	366	4
LYS	94	ASN	155	4
GLN	100	ASN	365	4
ASN	102	GLY	313	3
GLU	101	ASP	212	3
LYS	94	ALA	416	3
ASN	102	GLY	316	2
ASN	102	ALA	157	2
GLU	101	LYS	126	2
ALA	98	GLY	211	1
TRP	97	VAL	439	1
TRP	97	VAL	314	1
TRP	97	TRP	195	1
TRP	97	GLU	347	1
TRP	97	GLU	317	1
TRP	97	ARG	122	1
THR	104	LYS	368	1
ARG	93	TRP	366	1
ASN	102	THR	362	1
GLU	101	CYS	360	1
GLU	101	ARG	321	1
LYS	94	HIS	414	1
LYS	94	ASP	417	1
GLN	100	LYS	368	1

[0883] 表15

[0884]

孔		酶		接触的数量
氨基酸	位置	氨基酸	位置	
TRP	97	GLU	361	3862
TRP	97	GLU	317	3269
TRP	97	GLU	154	2166
ASN	102	GLY	313	2086
ASN	102	THR	210	2009
GLU	101	TRP	195	1896
ASN	102	ASN	365	1656
THR	104	GLU	419	1510
TRP	97	ASN	365	1488
TRP	97	ARG	216	1435
GLN	100	ALA	157	1204
GLY	103	TRP	195	1191
GLU	101	ASN	365	1106
GLU	101	LYS	368	1100
ASN	102	TRP	195	822
SER	257	ASP	260	779
ASN	102	GLU	361	738
TRP	97	TRP	195	719
TRP	97	ASN	155	591
ASN	102	ARG	207	535
ASN	102	TRP	366	524
GLY	103	LYS	364	514
GLU	101	THR	210	402
TRP	97	THR	210	377
ASN	102	GLY	208	371
TRP	97	TYR	318	365
TRP	97	GLN	422	333
TRP	97	PRO	315	316
TRP	97	VAL	220	309
THR	104	LYS	368	307
TRP	97	THR	156	304
ASN	102	ALA	157	274

[0885]

SER	115	ASP	212	274
TRP	97	GLY	316	266
GLU	101	GLN	422	264
GLY	103	TYR	197	214
ASN	102	ARG	148	203
ASN	102	LYS	368	199
ASN	102	ASP	198	199
ASN	102	LYS	255	192
ASN	102	PHE	308	180
TRP	97	GLU	419	174
LYS	94	GLY	211	174
GLN	100	ASN	155	168
ASN	102	PHE	223	153
TRP	97	LYS	358	152
TRP	97	ASP	198	151
ARG	110	GLY	193	148
GLN	100	GLU	419	144
ASN	102	LYS	364	139
ALA	98	ALA	157	135
ASN	102	ILE	159	122
ALA	98	TRP	195	113
THR	104	TYR	197	113
TRP	97	THR	362	103
GLY	103	GLY	369	103
LYS	94	THR	210	103
LYS	94	ASP	212	89
ASN	102	GLU	154	87
ASN	102	ILE	196	84
LYS	94	TYR	415	84
THR	104	GLU	154	77
TRP	97	GLY	193	75
ASN	102	ARG	321	71
TRP	97	VAL	418	69
GLU	101	TRP	366	69
THR	104	ASP	198	67
THR	104	ALA	157	64

[0886]

GLY	103	GLU	154	60
LYS	94	GLY	153	60
LYS	94	GLU	154	59
TRP	97	TRP	366	58
GLU	101	ARG	207	56
ARG	110	LEU	194	52
ASN	102	GLY	369	46
LYS	94	GLU	361	46
GLY	103	LYS	368	42
ARG	93	ASP	212	40
ASN	102	GLU	317	39
GLN	100	THR	156	34
TRP	97	VAL	314	32
ALA	98	ASP	212	31
GLY	103	ARG	207	31
LYS	94	PHE	209	29
TRP	97	GLY	211	28
ASN	102	ARG	312	28
ARG	110	GLY	313	25
TRP	97	ARG	321	22
ASN	102	VAL	220	22
TRP	97	ASP	417	18
THR	104	GLY	369	18
ASN	102	THR	164	17
LYS	94	ARG	216	17
TRP	97	GLY	153	16
ASN	102	TYR	158	16
GLY	103	LEU	194	16
GLU	101	ILE	159	15
GLN	100	PRO	315	15
ILE	95	ASP	212	14
GLY	103	TRP	366	14
TRP	97	LEU	194	13
GLU	101	ARG	321	13
GLN	100	LYS	368	13
ASN	102	GLY	153	11

[0887]

ARG	110	ASN	192	11
GLN	100	TYR	158	10
ASN	102	LYS	145	9
GLU	101	ARG	216	9
ARG	110	PRO	315	8
ALA	98	THR	210	7
ASN	102	GLY	211	7
ASN	102	ARG	216	7
ARG	110	LYS	368	7
LYS	94	HIS	414	7
LYS	94	ALA	214	7
SER	257	GLU	258	6
TRP	97	LYS	255	6
THR	104	ASP	212	6
LYS	94	ASP	417	6
ASN	91	ASP	212	5
TRP	97	THR	213	5
TRP	97	LYS	364	5
THR	104	TRP	195	5
THR	104	PHE	308	5
GLY	103	LYS	255	5
GLY	103	GLU	419	5
GLY	103	ARG	321	5
LEU	116	ASP	212	4
THR	104	TYR	438	4
THR	104	THR	210	4
THR	104	ILE	159	4
ARG	93	ARG	216	4
GLU	101	ALA	157	4
GLY	103	ILE	159	4
GLY	103	ASP	198	4
LYS	94	ASN	155	4
LEU	113	ARG	216	3
ASN	102	LEU	319	3
GLU	101	VAL	314	3
GLU	101	GLY	193	3

[0888]

LEU	113	TRP	366	2
ASN	102	VAL	314	2
ASN	102	TYR	197	2
ASN	102	PHE	209	2
ASN	102	LEU	194	2
GLU	101	THR	362	2
GLU	101	GLY	313	2
GLY	103	ILE	196	2
ALA	98	PRO	315	1
ALA	98	GLY	211	1
ALA	98	ASN	155	1
VAL	105	LYS	368	1
LEU	90	ASP	212	1
TRP	97	PHE	223	1
TRP	97	PHE	209	1
TRP	97	ASP	217	1
TRP	97	ASP	212	1
THR	104	VAL	200	1
THR	104	GLY	370	1
THR	104	ARG	207	1
ASN	102	PRO	152	1
ASN	102	LYS	310	1
ASN	102	LYS	227	1
ASN	102	GLU	419	1
ASN	102	ARG	122	1
ASN	102	ALA	311	1
GLU	101	VAL	418	1
GLU	101	PHE	308	1
GLU	101	LEU	194	1
GLY	103	PHE	308	1
LYS	94	VAL	439	1
GLN	100	ARG	122	1

[0889] 表16

[0890] 实例6

[0891] 本实例描述显示表征准确度改进的若干CsgG突变的表征。

[0892] 材料和方法

[0893] 在此实例中使用的材料和方法是与上文对于实例2所描述的那些相同。用于控制移动的酶是酶1=T4 Dda-E94C/C109A/C136A/A360C(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:24或酶2=T4 Dda-E94C/F98W/C109A/C136A/K194L/A360C(具有突变E94C/F98W/C109A/C136A/K194L/A360C的SEQ ID NO:24)。

[0894] 使用如在国际申请PCT/GB2012/052343(公开为W0/2013/041878)中公开的方法,来计算1D准确度表征测量。

[0895] 结果

[0896] 对于MspA突变 $x = \text{MspA} - ((\text{Del} - \text{L74/G75/D118/L119})\text{D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K})8$ (SEQ ID NO:50,其具有突变D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K,和氨基酸L74/G75/D118/L119的缺失)(其中DNA易位受T4 Dda-E94C/F98W/C109A/C136A/K194L/A360C控制)的1D basecall表征准确度是68.7%。所测试的所有突变(参见下表17)显示相较于MspA突变X,1D basecall表征准确度改进。

[0897] 27-CsgG-Eco-(Y51A/F56Q/R97W/R192Q-StrepII(C))9(具有突变Y51A/F56Q/R97W/R192Q的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0898] 28-CsgG-Eco-(Y51A/F56Q/R97W/R192D-StrepII(C))9(具有突变Y51A/F56Q/R97W/R192D的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0899] 29-CsgG-Eco-(Y51A/F56Q/K135L/T150I/S208V-StrepII(C))9(具有突变Y51A/F56Q/K135L/T150I/S208V的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0900] 30-CsgG-Eco-(Y51A/F56Q/T150I/S208V-StrepII(C))9(具有突变Y51A/F56Q/T150I/S208V的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0901] 31-CsgG-Eco-(Y51A/F56Q/S208V-StrepII(C))9(具有突变Y51A/F56Q/S208V的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0902] 32-CsgG-Eco-(Y51A/F56Q/T150I-StrepII(C))9(具有突变Y51A/F56Q/T150I的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0903] 33-CsgG-Eco-(Y51A/F56Q/K135V/T150Y-StrepII(C))9(具有突变Y51A/F56Q/K135V/T150Y的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0904] 34-CsgG-Eco-(Y51A/F56Q/K135L-StrepII(C))9(具有突变Y51A/F56Q/K135L的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0905] 35-CsgG-Eco-(Y51A/F56Q/R97F/R192D-StrepII(C))9(具有突变Y51A/F56Q/R97F/R192D的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0906] 36-CsgG-Eco-(Y51A/F56Q/K135L/T150I-StrepII(C))9(具有突变Y51A/F56Q/K135L/T150I的SEQ ID NO:2,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0907] 37-CsgG-Eco-((Del-D195/Y196/Q197/R198/L199)-Y51A/F56Q-StrepII(C))9(SEQ ID NO:2,其具有突变Y51A/F56Q和氨基酸D195/Y196/Q197/R198/L199的缺失,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0908] 38-CsgG-Eco-((Del-R192/F193/I194/D195/Y196/Q197/R198/L199/L200)-Y51A/F56Q-StrepII(C))9(SEQ ID NO:2,其具有突变Y51A/F56Q和氨基酸R192/F193/I194/D195/Y196/Q197/R198/L199/L200的缺失,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0909] 39-CsgG-Eco-((Del-Q197/R198/L199/L200)-Y51A/F56Q-StrepII(C))9(SEQ ID NO:2,其具有突变Y51A/F56Q和氨基酸Q197/R198/L199/L200的缺失,其中StrepII(C)是SEQ ID NO:47,且在C端处连接)

[0910] 40-CsgG-Eco-((Del-I194/D195/Y196/Q197/R198/L199)-Y51A/F56Q-StrepII

(C)) 9 (SEQ ID NO:2, 其具有突变Y51A/F56Q和氨基酸I194/D195/Y196/Q197/R198/L199的缺失, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0911] 41-CsgG-Eco- ((Del-V139/G140/D149/T150/V186/Q187/V204/G205) -Y51A/F56Q-StrepII (C)) 9 (SEQ ID NO:2, 其具有突变Y51A/F56Q和氨基酸V139/G140/D149/T150/V186/Q187/V204/G205的缺失, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0912] 42-CsgG-Eco- ((Del-D195/Y196/Q197/R198/L199/L200) -Y51A/F56Q-StrepII (C)) 9 (SEQ ID NO:2, 其具有突变Y51A/F56Q和氨基酸D195/Y196/Q197/R198/L199/L200的缺失, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0913] 43-CsgG-Eco- ((Del-Y196/Q197/R198/L199/L200/E201) -Y51A/F56Q-StrepII (C)) 9 (SEQ ID NO:2, 其具有突变Y51A/F56Q和氨基酸Y196/Q197/R198/L199/L200/E201的缺失, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0914] 44-CsgG-Eco- ((Del-Q197/R198/L199) -Y51A/F56Q-StrepII (C)) 9 (SEQ ID NO:2, 其具有突变Y51A/F56Q和氨基酸Q197/R198/L199的缺失, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0915] 45-CsgG-Eco- ((Del-F193/I194/D195/Y196/Q197/R198/L199) -Y51A/F56Q-StrepII (C)) 9 (SEQ ID NO:2, 其具有突变Y51A/F56Q和氨基酸F193/I194/D195/Y196/Q197/R198/L199的缺失, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0916] 46-CsgG-Eco- (Y51A/F56Q/R192T-StrepII (C)) 9 (具有突变Y51A/F56Q/R192T的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0917] 47-CsgG-Eco- (Y51A/F56Q/N102S-StrepII (C)) 9 (具有突变Y51A/F56Q/N102S的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0918] 48-CsgG-Eco- (Y51A/F56Q/Q42R-StrepII (C)) 9 (具有突变Y51A/F56Q/Q42R的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0919] 49-CsgG-Eco- (Y51A/F56Q/R192S-StrepII (C)) 9 (具有突变Y51A/F56Q/R192S的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0920] 50-CsgG-Eco- (Y51A/F56Q/G103N-StrepII (C)) 9 (具有突变Y51A/F56Q/G103N的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0921] 51-CsgG-Eco- (Y51A/F56Q/R97N/N102R-StrepII (C)) 9 (具有突变Y51A/F56Q/R97N/N102R的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0922] 52-CsgG-Eco- (Y51A/F56Q/R97L-StrepII (C)) 9 (具有突变Y51A/F56Q/R97L的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0923] 53-CsgG-Eco- (Y51A/F56Q/R192D-StrepII (C)) 9 (具有突变Y51A/F56Q/R192D的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0924] 54-CsgG-Eco- (Y51A/F56Q/R97N/N102G-StrepII (C)) 9 (具有突变Y51A/F56Q/R97N/N102G的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0925] 55-CsgG-Eco- (Y51A/F56Q/F48S-StrepII (C)) 9 (具有突变Y51A/F56Q/F48S的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0926] 56-CsgG-Eco- (Y51A/F56Q/G103S-StrepII (C)) 9 (具有突变Y51A/F56Q/G103S的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)

[0927] 57-CsgG-Eco- (Y51A/F56Q/E101L-StrepII (C)) 9 (具有突变Y51A/F56Q/E101L的SEQ ID N0:2,其中StrepII (C)是SEQ ID N0:47,且在C端处连接)

[0928] 58-CsgG-Eco- (Y51A/F56Q/R192Q-StrepII (C)) 9 (具有突变Y51A/F56Q/R192Q的SEQ ID N0:2,其中StrepII (C)是SEQ ID N0:47,且在C端处连接)

[0929] 59-CsgG-Eco- (Y51A/F56Q/K135N/R142N/R192N-StrepII (C)) 9 (具有突变Y51A/F56Q/K135N/R142N/R192N的SEQ ID N0:2,其中StrepII (C)是SEQ ID N0:47,且在C端处连接)

[0930] 60-CsgG-Eco- (Y51A/F56Q/R97N-StrepII (C)) 9 (具有突变Y51A/F56Q/R97N的SEQ ID N0:2,其中StrepII (C)是SEQ ID N0:47,且在C端处连接)

[0931] 61-CsgG-Eco- (Y51A/F56Q/R192N-StrepII (C)) 9 (具有突变Y51A/F56Q/R192N的SEQ ID N0:2,其中StrepII (C)是SEQ ID N0:47,且在C端处连接)

[0932] 62-CsgG-Eco- (Y51A/F56Q/Y130W-StrepII (C)) 9 (具有突变Y51A/F56Q/Y130W的SEQ ID N0:2,其中StrepII (C)是SEQ ID N0:47,且在C端处连接)

[0933] 63-CsgG-Eco- (Y51A/F56Q/E101G-StrepII (C)) 9 (具有突变Y51A/F56Q/E101G的SEQ ID N0:2,其中StrepII (C)是SEQ ID N0:47,且在C端处连接)

[0934]

CsgC 突变编号	酶	1D Basecall 表征准确度 (%)
25	2	72.6
26	2	79.2
27	2	81.3
28	2	80.2
29	2	72.2
30	2	74.7
31	2	74.5
32	2	73.3
33	2	73.8
34	2	74.3
35	2	75.5
36	2	73.5
37	2	73.4
38	2	75.3

[0935]

39	2	72.0
40	2	74.5
41	2	74.8
42	2	71.0
43	2	70.9
44	2	71.9
45	2	73.0
46	1	75.0
47	1	72.1
48	1	75.0
49	1	74.8
50	1	75.2
51	1	73.1
52	1	75.4
53	1	77.1
54	1	75.9
55	1	73.7
56	1	73.1
57	1	73.2
58	1	76.8
59	1	72.7
60	1	76.1
61	1	75.0
62	1	74.7
63	1	73.1
25	1	74.8
26	1	77.4

[0936] 表17

[0937] 实例7

[0938] 本实例比较多种不同突变纳米孔的DNA捕获。

[0939] 材料和方法

[0940] 从插入在含嵌段共聚物的缓冲液 (25mM磷酸钾缓冲液, 150mM亚铁 (II) 氰化钾, 150mM铁 (III) 氰化钾, pH 8.0) 中的各种单一CsgG或MspA纳米孔获得电测量。在达到插入在嵌段共聚物中的单一孔之后, 接着缓冲液 (2mL, 25mM磷酸钾缓冲液, 150mM亚铁 (II) 氰化钾, 150mM铁 (III) 氰化钾, pH 8.0) 流过系统, 移除任何过量的纳米孔。接着, 使150μL的500mM KCl、25mM磷酸钾、1.5mM MgCl₂、1.5mM ATP、pH8.0流过系统。在10分钟之后, 接着将150μL的DNA (SEQ ID NO:51, 200nM) 流动到单一纳米孔实验系统中。实验是在-120mV下进行, 且进行解螺旋酶控制的DNA移动监测。

[0941] 结果

[0942] CsgG突变CsgG-Eco- (Y51A/F56Q/R97W/E101S/R192D-StrepII (C)) 9 (具有突变Y51A/F56Q/R97W/E101S/R192D的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端

处连接) (参见图36) 显示, 比MspA- (De1-L74/G75/D118/L119) D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K) 8 (参见图34), 更高的捕获速率 (例如更容易地捕获DNA聚核苷酸)。电流迹线中的各尖峰对应于在不受酶控制的情况下DNA聚核苷酸 (SEQ ID NO:51) 通过纳米孔的易位。对于CsgG纳米孔-CsgG-Eco- (Y51A/F56Q/R97W/E101S/R192D-StrepII (C)) 9, 图34中的10秒电流迹线显示比10秒电流迹线更少的DNA易位。

[0943] 当相较于无E101S突变的CsgG突变时, 位置E101到E101S的突变引起捕获率的增加。图35显示, CsgG突变CsgG-Eco- (Y51A/F56Q/R97W/R192D-StrepII (C)) 9 (具有突变Y51A/F56Q/R97W/R192D的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 的10秒电流迹线呈现比CsgG-Eco- (Y51A/F56Q/R97W/E101S/R192D-StrepII (C)) 9更少的易位。CsgG-Eco- (Y51A/F56Q/R97W/R192D-StrepII (C)) 9的易位的平均数是7.25/秒 (n=12), 且CsgG-Eco- (Y51A/F56Q/R97W/E101S/R192D-StrepII (C)) 9的易位的平均数是18/秒 (n=14)。

[0944] 实例8

[0945] 本实例比较两个不同CsgG突变孔的表达水平。

[0946] 材料和方法

[0947] 在此实例中用于制备纳米孔 (A=CsgG-Eco- (Y51A/F56Q/R97W) -StrepII (C)) 9 (具有突变Y51A/F56Q/R97W的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接); 和B=CsgG-Eco- (Y51A/F56Q/R97W/R192D) -StrepII (C)) 9 (具有突变Y51A/F56Q/R97W/R192D的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接)) 的材料和方法与上文对于实例3所描述的那些相同。

[0948] 结果

[0949] 两个纳米孔A=CsgG-Eco- (Y51A/F56Q/R97W) -StrepII (C)) 9 (具有突变Y51A/F56Q/R97W的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接) 和B=CsgG-Eco- (Y51A/F56Q/R97W/R192D) -StrepII (C)) 9 (具有突变Y51A/F56Q/R97W/R192D的SEQ ID NO:2, 其中StrepII (C) 是SEQ ID NO:47, 且在C端处连接), 使用精确相同的方案来表达和纯化, 且使用凝胶过滤色谱图 (120mL S200柱, 参见图37) 和SDS-PAGE分析 (参见图38) 来分析相同体积的各纳米孔。B (470.3mAu) 的吸光度值比A (11.4mAu) 高许多, 此指, CsgG-Eco- (Y51A/F56Q/R97W/R192D) -StrepII (C)) 9以比CsgG-Eco- (Y51A/F56Q/R97W) -StrepII (C)) 9更许多的水平表达。图38中的条带的强度还指示两个孔的表达水平。条带A-C (含有CsgG-Eco- (Y51A/F56Q/R97W) -StrepII (C)) 9) 比条带D-E (含有CsgG-Eco- (Y51A/F56Q/R97W/R192D) -StrepII (C)) 9) 的强力低。分析两个的两个方法指示, 添加R192D突变极大地增加CsgG突变所观测到的表达。

[0950] 实例9

[0951] 本实例描述显示表征准确度改进的若干CsgG突变的表征。

[0952] 材料和方法

[0953] Csg孔

[0954] 测试以下8个CsgG突变孔。上文实例X中描述的突变28用作基线孔。在SEQ ID NO:2中制备突变, 且纯化标签StrepII具有SEQ ID NO:47中所示的序列。

[0955] 基线孔 (突变28): CsgG- (WT-Y51A/F56Q/R97W/R192D-StrepII) 9

[0956] 突变A:CsgG- (WT-Y51A/F56Q/R97W/R192D-StrepII)9-del (D195-L199)

[0957] 突变B:CsgG- (WT-Y51A/F56Q/R97W/R192D-StrepII)9-del (F193-L199)

[0958] 突变C:CsgG- (WT-Y51A/F56Q/R97W/R192D-StrepII)9-F191T

[0959] 突变D:CsgG- (WT-Y51A/F56Q/R97W/R192D-del (V105-I107) -StrepII)9

[0960] 突变E:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107)

[0961] 突变F:CsgG- (WT-Y51A/F56Q/R192D-StrepII)9-R93W

[0962] 突变G:CsgG- (WT-Y51A/F56Q/R192D-StrepII)9-R93W-del (D195-L199)

[0963] 突变H:CsgG- (WT-Y51A/F56Q/R192D-StrepII)9-R93Y/R97Y

[0964] 纳米孔制备

[0965] 为了制备含有嵌段共聚物膜的多个槽孔的纳米孔阵列芯片,所述多个孔各自具有所插入的单一CsgG突变纳米孔,使用以下方法。将在大肠杆菌中表达的CsgG突变纯化,且存储于在pH8下含有25mM Tris、150mM NaCl、2mM EDTA、0.01%DDM、0.1%SDS、0.1%Brij 58的缓冲液中。使用在pH 8.0下包含25mM磷酸钾、150mM亚铁(II)氰化钾、150mM铁(III)氰化钾的缓冲液,以1:1000000将这些突变CsgG孔进行稀释,且添加到芯片中,在各槽孔中获得单一孔。在孔插入之后,用在pH 8.0下包含25mM磷酸钾、150mM亚铁(II)氰化钾、150mM铁(III)氰化钾的1mL缓冲液来洗涤阵列芯片,以移除过量的孔。在几分钟之后,用500mL的测序混合液冲洗各芯片两次,所述测序混合液含有470mM KCl、25mM HEPES、11mM ATP和10mM MgCl₂。

[0966] DNA样本制备

[0967] 使用以下方法制备DNA样本,以用于测序。将1μg DNA分析物与40nM的含有同衔接子和平头TA接合酶(获自<https://store.nanoporetech.com/>)预结合的T4 Dda解螺旋酶的衔接子混合物一起培育10分钟。衔接子的结构显示于图43中,且衔接子中所含的序列陈述于SEQ ID NO:52到55中。接着,使用Spri纯化,纯化接合反应混合物,以移除未接合的游离衔接子。在含有在pH10下的40mM CAPS、40mM KCl和400nM胆固醇系链的25μL洗脱缓冲液中,洗脱最终接合的混合物。对于各芯片,将12μL DNA-衔接子接合的混合物与测序混合液(最终体积,150μL)混合,且添加到芯片,以用于测序。接着,在160mV下进行实验6小时。

[0968] 使用如在W02013/041878中所公开的方法,来计算1D准确度表征测量。

[0969] 测量模板速度

[0970] 通过以下方法来测量模板速度。将各波浪线的basecall与参考序列进行比对。将跨越比对(从比对起始位置减去比对终点位置)的碱基的数量,除对应于比对结束的事件与对应于比对开始的事件之间的时间。

[0971] 结果

[0972] Basecall准确度

[0973] 如图39中所示,发现所有8个CsgG突变孔,相较于实例6中描述的基线孔突变28(CsgG- (WT-Y51A/F56Q/R97W/R192D-StrepII)9),均具有改进的basecall准确度。如表17中所示,突变28显示在实例6中所测试的所有CsgG突变的最高basecall准确度(80.2%),其中实例5中描述的突变25(CsgG- (WT-Y51A/F56Q-StrepII)9)是基线突变。因此,除由突变28中的R97W和R192D取代引起的准确度的改进以外,D195-L199、F193-L199或V105-I107的缺失,或F191T的取代,引起另外的准确度的改进。

[0974] D195-L199、F193-L199或V105-I107的缺失,或F191T的取代,也各自将预期为在存在CsgG序列中的其它突变下或在不存在R97W、R192D、Y51A和/或F56Q突变下准确度得到改进。举例来说,突变E,其含有除del(V105-I107)突变以外的突变K94Q,相较于突变28具有改进的准确度,就如含有del(V105-I107)的突变G一样,所述突变G不含有R97W取代但相反含有R93W取代。

[0975] 实例26中的表17显示,突变26,其含有R97W取代,具有几乎与突变28(80.2%)一样良好的basecall准确度(79.2%)。这表明,R97W突变是basecall准确度增加的基础。在此实例中,对不含有R97W突变(以及不含有R192D突变)但相反含有R93W取代(突变F)或R93Y取代和R97Y取代两个(突变H)的两个突变进行测试,且发现具有比突变28更高的basecall准确度。这显示,R93W和R93Y/R97Y取代可用于改进CsgG纳米孔的basecall准确度。

[0976] 模板速度和模板准确度

[0977] 如图40A中所示,突变D(CsgG-(WT-Y51A/F56Q/R97W/R192D-del(V105-I107)-StrepII)9),相较于基线突变28CsgG-(WT-Y51A/F56Q/R97W/R192D-StrepII)9,具有收紧的速度群体分布。

[0978] 图40B显示,突变D(CsgG-(WT-Y51A/F56Q/R97W/R192D-del(V105-I107)-StrepII)9),相较于基线突变28CsgG-(WT-Y51A/F56Q/R97W/R192D-StrepII)9,具有收紧的模板准确度分布。

[0979] 有利的是,具有收紧的速度和准确度分布,这是由于数据中存在较小的偏差。另外,数据的中值准确度将通过降低所产生的较低准确度数据的量而增加。因此,从CsgG纳米孔的V105到I107缺失氨基酸,可用于产生具有用于表征聚核苷酸的改进的特性的CsgG纳米孔。

[0980] 实例10

[0981] 本实例描述显示噪声孔信号减少的CsgG突变的表征。

[0982] 材料和方法

[0983] CsgG孔

[0984] 测试以下CsgG突变孔。上文实例X中描述的突变28用作基线孔。在SEQ ID NO:2中制备突变,且纯化标签StrepII具有SEQ ID NO:47中所示的序列。

[0985] 基线孔(突变28):CsgG-(WT-Y51A/F56Q/R97W/R192D-StrepII)9

[0986] 突变I:CsgG-(WT-Y51A/F56Q/R97W/R192D-StrepII)9-K94N。

[0987] 突变J:CsgG-(WT-Y51A/F56Q/R97W/R192D-StrepII)9-K94Q。

[0988] 纳米孔制备

[0989] 如实例9中所描述,制备含有嵌段共聚物膜的多个槽孔的芯片,所述多个槽孔各自具有插入的单一CsgG突变纳米孔。

[0990] DNA样本制备

[0991] 使用实例9中描述的方法,制备DNA样本,以用于测序。

[0992] 测定在噪声孔状态中的时间花费

[0993] 如下计算噪声孔状态中的时间花费的百分比。将在各通道内的事件检测的信号分成不重叠的短窗口。对于每个窗口,计算电流电平的平均值和电流电平的分散性。接着,将获得的值传递给分类器,所述分类器返回标示窗口是否含有噪声信号的标记。分类器经过

训练,以通过提供噪声信号具有预标记的数据来检测噪声信号。

[0994] 结果

[0995] 噪声孔状态

[0996] 图41显示实例“波浪线”,其显示通过基线突变28CsgG- (WT-Y51A/F56Q/R97W/R192D-StrepII) 9呈现的“噪声”孔误差模式。图41的上图显示在“良好”和“噪声”孔状态期间,通过孔的电流流的差异。图41的下图显示从“良好”状态过渡到“噪声”状态的放大图。

[0997] 图42显示平均在至少5个回合内,当相较于基线突变28时,突变孔I和J的噪声孔状态的减少。

[0998] 突变J和突变I两个的噪声孔状态的时间花费的百分比相较于基线显著地降低。突变I和J与用作基线的突变28恰好相差一个残基。突变I和突变J两个均含有K94的取代。突变I含有K94N突变,且突变II含有K94Q突变。因此,CsgG纳米孔中的K94的取代,尤其用N或Q取代,可用于产生具有用于表征聚核苷酸的改进的特性的CsgG纳米孔。

[0999] 实例11

[1000] 本实例描述显示捕获活性增加的若干CsgG突变的表征。

[1001] 材料和方法

[1002] CsgG孔

[1003] 测试以下CsgG突变孔。上文实例9中描述的突变E用作基线孔。在SEQ ID NO:2中制备突变,且纯化标签StrepII具有SEQ ID NO:47中所示的序列。

[1004] 基线(突变E):CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) .

[1005] 突变K:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -Q42K

[1006] 突变L:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -E44N

[1007] 突变M:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -E44Q

[1008] 突变N:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -L90R

[1009] 突变O:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -N91R

[1010] 突变P:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -I95R

[1011] 突变Q:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -A99R

[1012] 突变R:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -E101H

[1013] 突变S:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -E101K

[1014] 突变T:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -E101N

[1015] 突变U:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -E101Q

[1016] 突变V:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -E101T

[1017] 突变W:CsgG- (WT-Y51A/F56Q/K94Q/R97W/R192D-del (V105-I107) -Q114K

[1018] 纳米孔制备

[1019] 为了制备含有嵌段共聚物膜的多个槽孔的纳米孔阵列芯片,所述多个槽孔各自具有所插入的单一CsgG突变纳米孔,使用以下方法。将在大肠杆菌中表达的CsgG突变纯化,且存储于在pH8下含有25mM Tris、150mM NaCl、2mM EDTA、0.01%DDM、0.1%SDS、0.1%Brij 58的缓冲液中。使用在pH 8.0下包含25mM磷酸钾、150mM亚铁(II)氰化钾、150mM铁(III)氰化钾的缓冲液,以1:1000000将这些突变CsgG孔进行稀释,且添加到芯片中,在各槽孔中获得单一孔。在孔插入之后,用在pH 8.0下包含25mM磷酸钾、150mM亚铁(II)氰化钾、150mM铁

(III) 氰化钾的1mL缓冲液来洗涤芯片,以移除过量的孔。将1mL在pH 8.0下含有含240nM TBA分析物的25mM磷酸钾、150mM亚铁(II) 氰化钾、150mM铁(III) 氰化钾的溶液冲洗到芯片中。

[1020] 测定捕获能力

[1021] 接着,突变孔捕获DNA分析物的能力通过其捕获凝血酶结合适体(TBA) (SEQ ID NO:51) 的能力来评定。在180mV下进行实验。为了测量通过孔的TBA捕获,计算TBA事件之间的中值时间。

[1022] 结果

[1023] 如图44中所示,13个突变的TBA事件之间的中值时间相较于基线显著地降低,从而指示,所有13个突变均显示模板DNA的捕获速率增加。

[1024] 13突变中的每一个相较于基线孔具有单一氨基酸取代。特定取代是:Q42K、E44N、E44Q、L90R、N91R、I95R、A99R、E101H、E101K、E101N、E101Q、E101T和Q114K。所有这些突变均包括用不带电氨基酸或带正电氨基酸取代带负电氨基酸,或带正电氨基酸取代不带电氨基酸。因此,可推断,在位置Q42、E44、E44、L90、N91、I95、A99、E101和Q114中的一或多个处的氨基酸,用在这些位置处移除负电荷和/或增加正电荷的氨基酸取代,引起聚核苷酸的捕获增加。

[1025] 各种CsgG同源物的序列比对

[1026] 图45显示二十一个CsgG同源物之间的序列比对,如上文详述。对以下序列进行多序列比对:SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:32、SEQ ID NO:33、SEQ ID NO:34、SEQ ID NO:35、SEQ ID NO:36、SEQ ID NO:37、SEQ ID NO:38、SEQ ID NO:39、SEQ ID NO:40和SEQ ID NO:41。

[1027] Praline软件(一种集成同源性延长和二级结构信息的多序列比对工具箱)用于执行比对,<http://www.ibi.vu.nl/programs/pralinewww/>,也参见Simossis VA1,Heringa J.;《核酸研究》2005年7月1日;33(万维网服务器问题):W289-94。使用BLOSUM62残基交换矩阵,来对比对进行记分。关于这种方法的详情,参见例如Henikoff S, Henikoff JG;《美国国家科学院学报(Proc.Natl.Acad.Sci.USA)》第89卷,第10915-10919页,1992年11月。分别地使用12的空隙开口和1的延长罚分。使用PSIPRED的二级结构预测用于引导比对。关于这种方法的详情,参见例如Jones D.T.;《分子生物学杂志》1999年9月17日;292(2):195-202。用于比对序列的以上方法是示范性的,且可使用所属领域中已知的其它序列比对方法。

[1028] 参考图45的序列比对,序列比对的各区段、在各位置处的保守性通过直方图和得分指示。刻度上的数字0-9指示保守性增加,具有产生与保守氨基类似特性的突变的柱标记有加号(‘+’),且星形符号(‘*’)指示在那个位置处的100%序列一致性。可从序列比对的保守值看出,许多残基显示极高或甚至完美的序列一致性,从而指示这些21个同源物是紧密相关的。

[1029] 图46显示与图45相同的相对序列比对,其中所预测的 α 螺旋二级结构区另外加灰色阴影。图47显示与图45相同的相对序列比对,其中所预测的 β 折叠二级结构区另外加灰色阴影。图46和47显示,这些同源物的所预测的 α 螺旋和 β 折叠的区(CsgG纳米孔的重要二级结构)是高度保守的。

[1030] 多序列比对强烈表明所述序列是相关的;不仅是存在很高程度的沿比对的保守性,而且还比对所预测的二级结构元件。

[1031] 图45、46和47中的序列比对可用作显示彼此比对的相对位置的参考。因此,可鉴定相对于SEQ ID NO 2鉴定的氨基酸残基和其它CsgG同源物中的对应氨基酸残基。为了易于鉴定,残基R97和R192已用星号定位。可从表看出,例如SEQ ID NO:2的R192对应于SEQ ID NO:32的残基R191和SEQ ID NO:37的残基K177。

[1032] 如参考图45到47将容易地理解,CsgG单体是高度保守的。此外,根据关于SEQ ID NO:2的突变的知识,有可能测定除SEQ ID NO:2以外的CsgG单体的突变的等效位置。

[1033] 因此,提到包含如SEQ ID NO:2中所示序列的变异体和如在权利要求书和在说明书中其它地方中陈述的其特异性氨基酸突变的突变CsgG单体,还涵盖包含如以下中所示序列的变异体和其对应氨基酸突变的突变CsgG单体:SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:32、SEQ ID NO:33、SEQ ID NO:34、SEQ ID NO:35、SEQ ID NO:36、SEQ ID NO:37、SEQ ID NO:38、SEQ ID NO:39、SEQ ID NO:40和SEQ ID NO:41。同样,提到涉及使用关于包含如SEQ ID NO:2中所示序列的变异体和如在权利要求书中和在说明书中其它地方中陈述的其特异性氨基酸突变的突变CsgG单体的孔的构筑体、孔或方法,还涵盖关于包含根据以上所公开的SEQ ID NO的序列的变异体和其对应氨基酸突变的突变CsgG单体的构筑体、孔或方法。还应理解,本发明延伸到未在说明书中明确鉴定的显示高度保守区的其它变异CsgG单体。

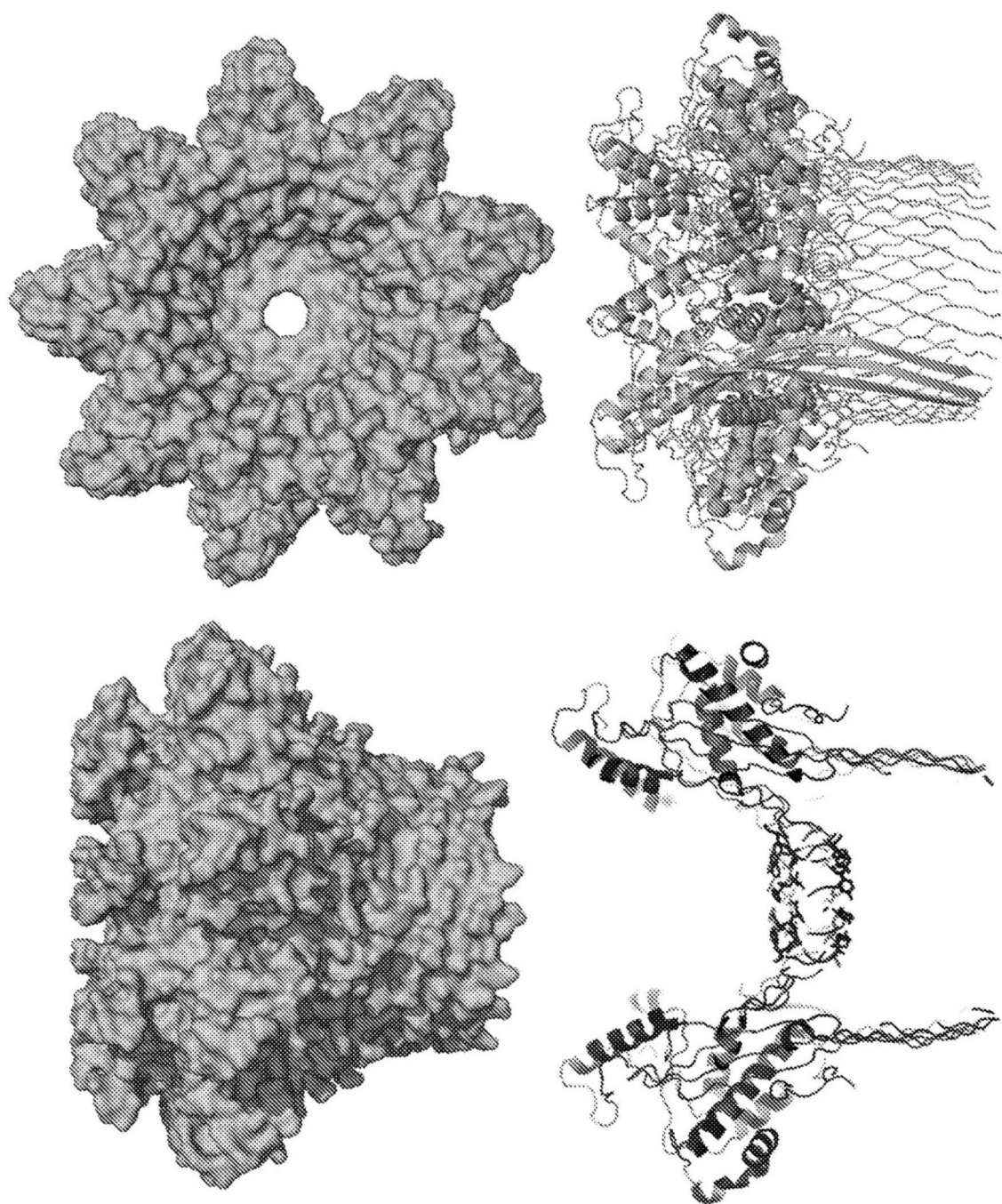


图1

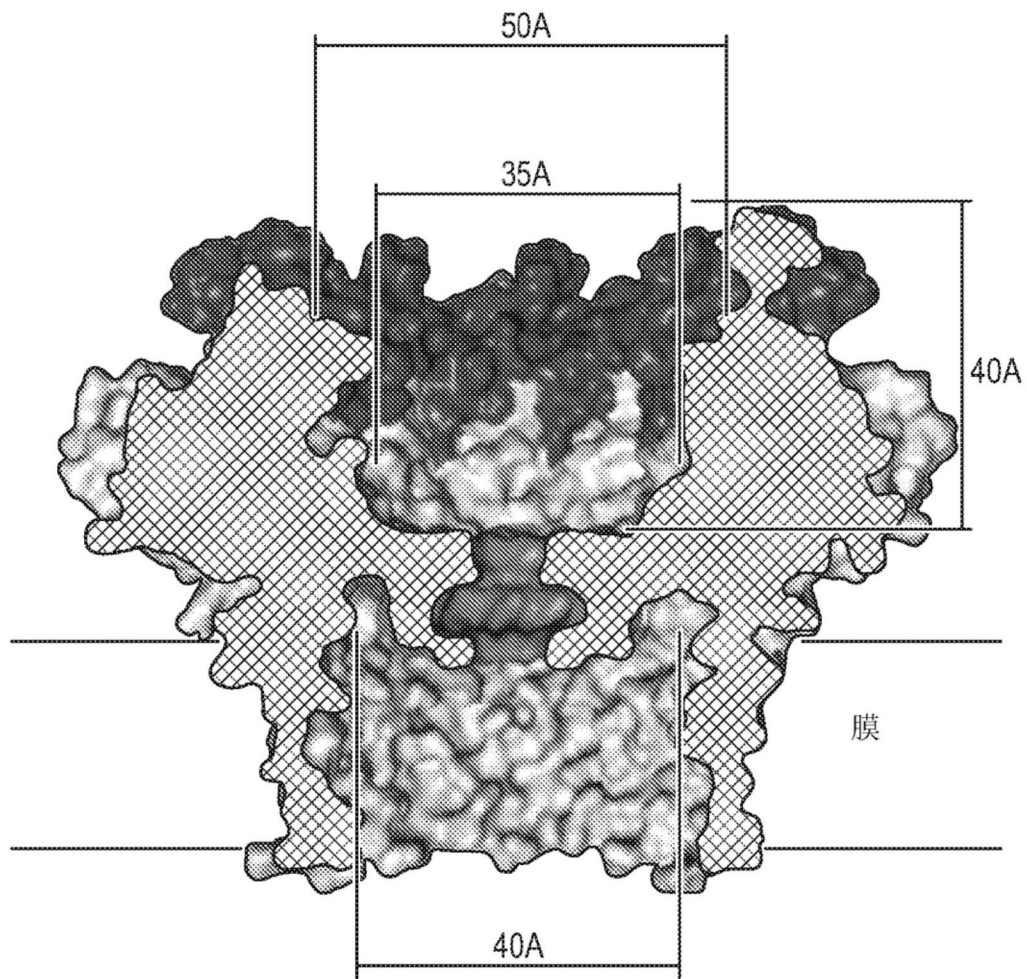


图2

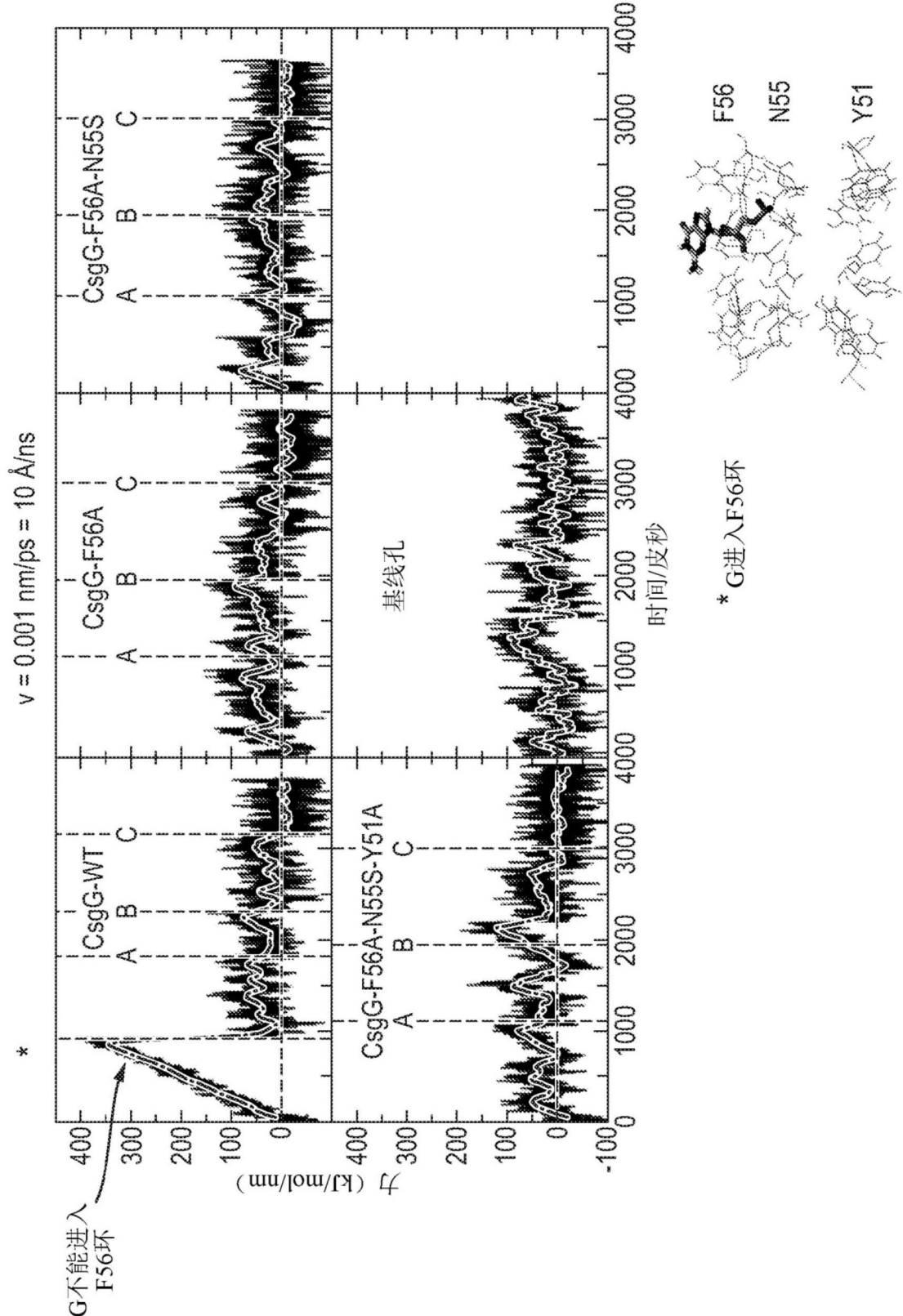


图3

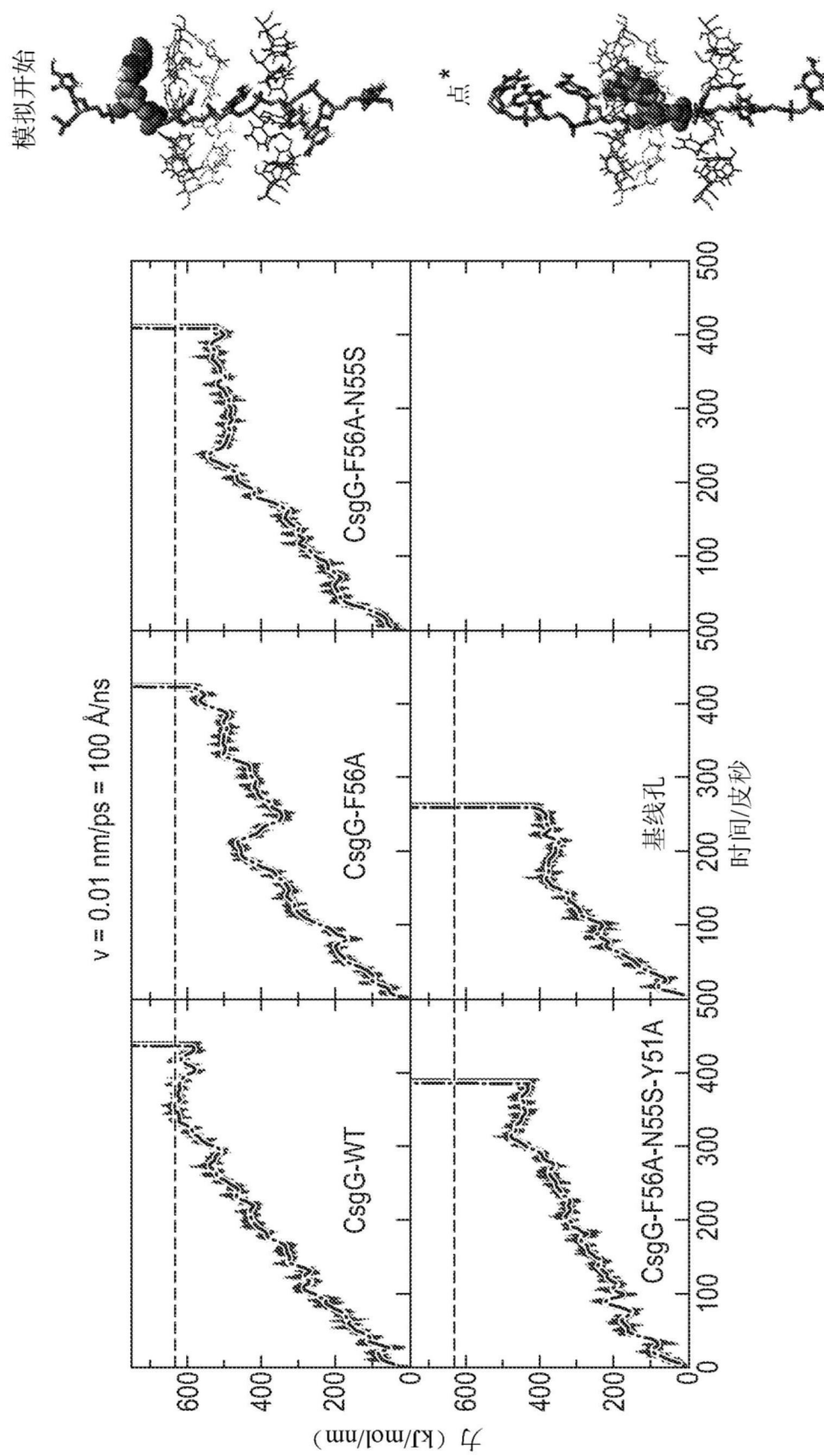


图4

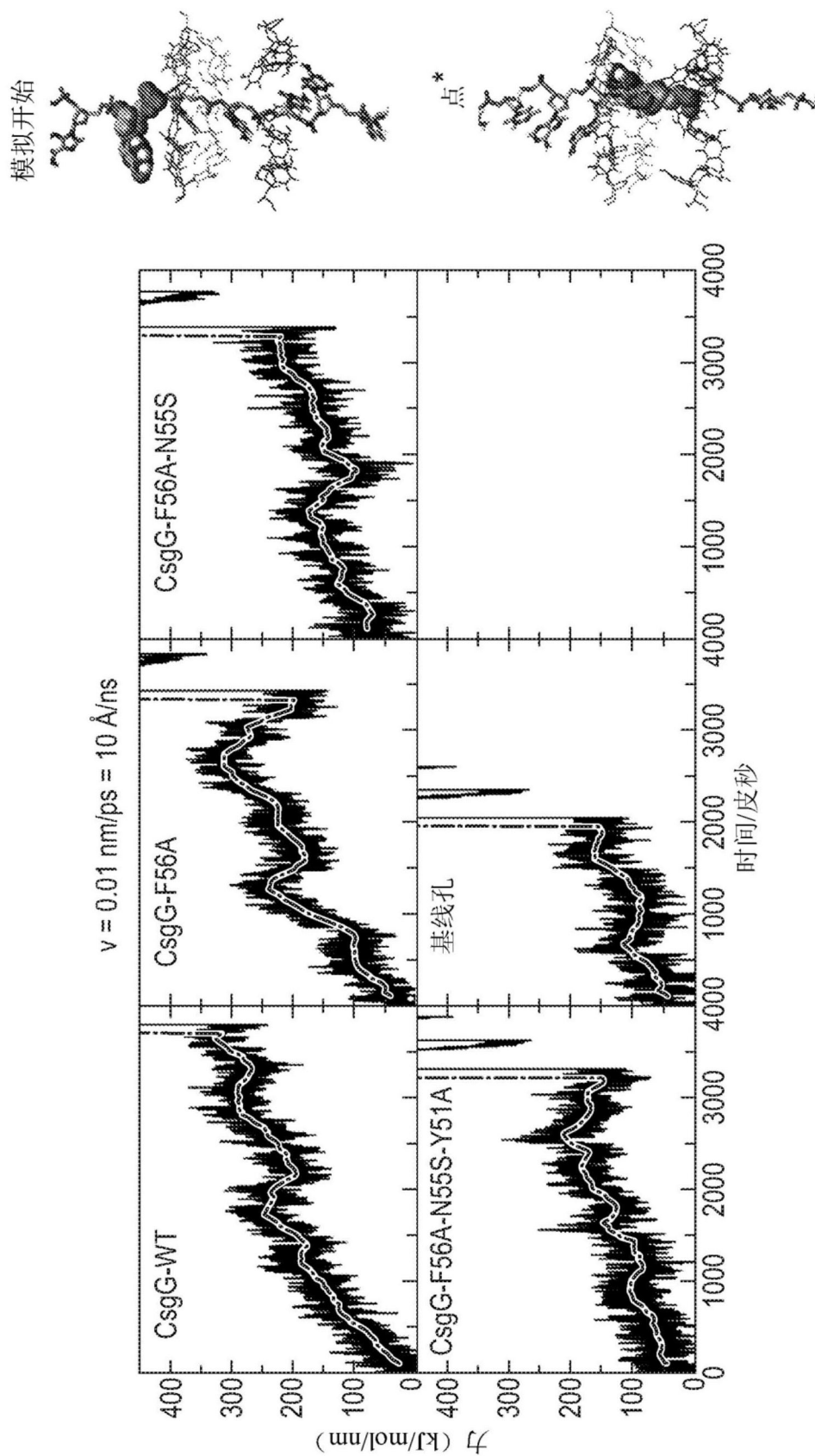


图5

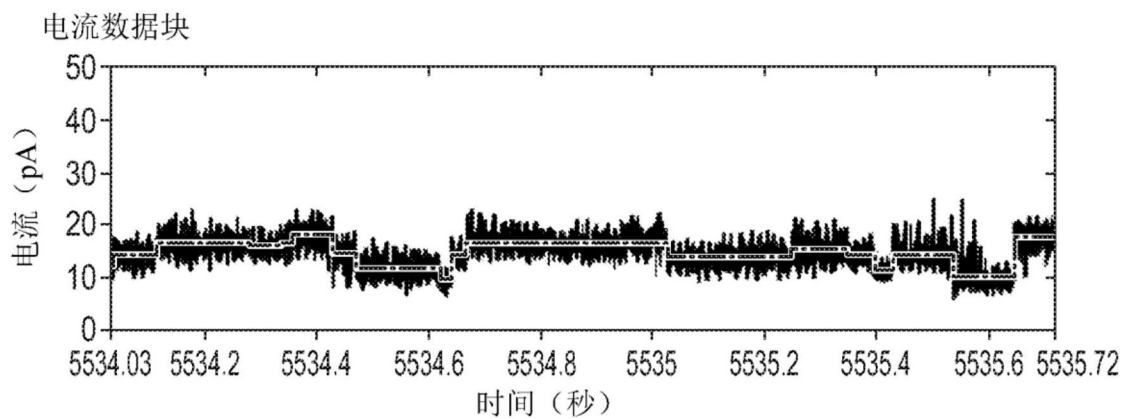


图6(a)

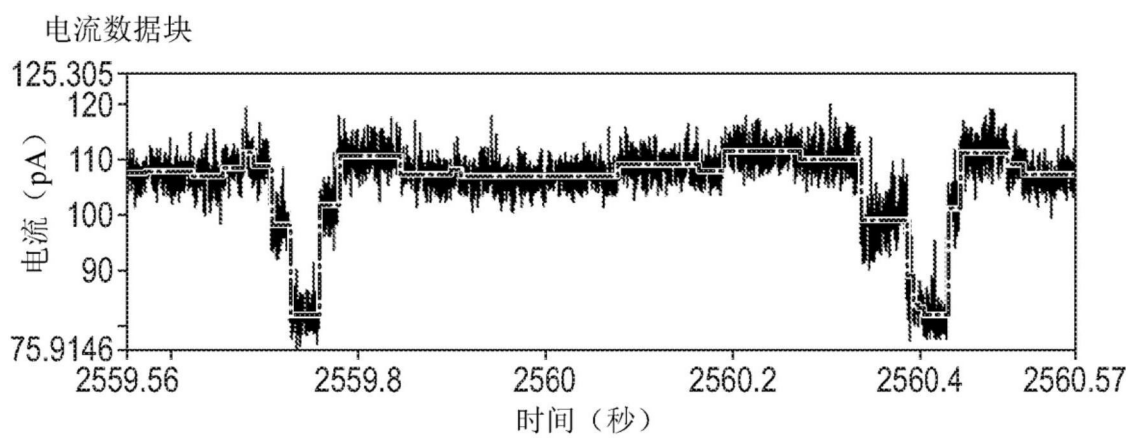


图6(b)

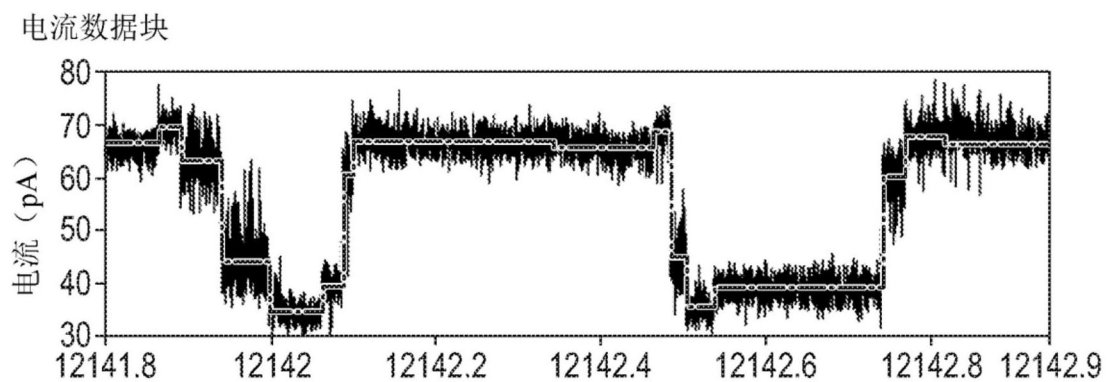


图6(c)

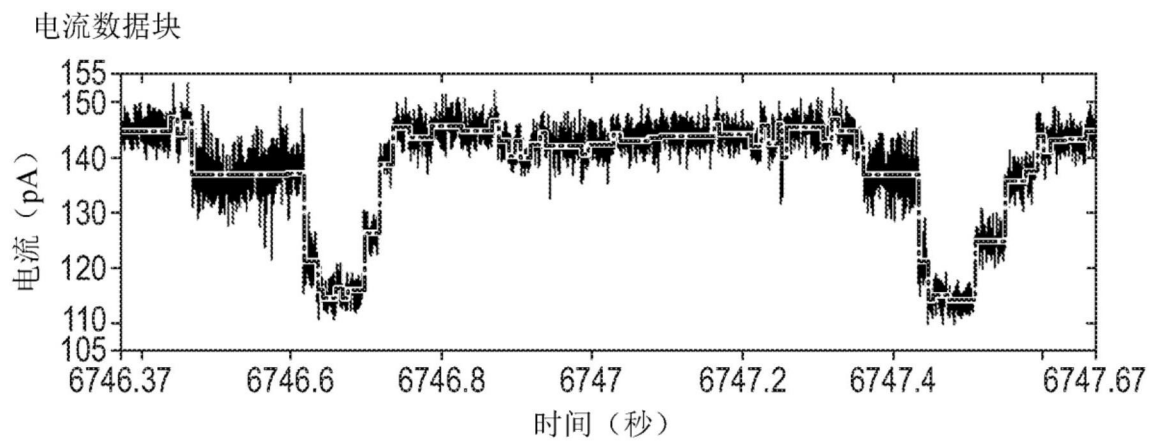


图7(a)

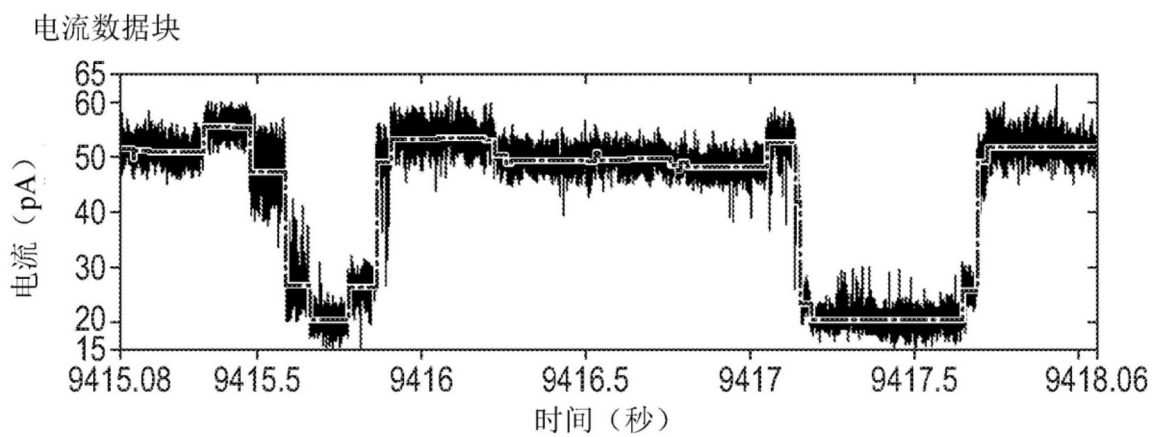


图7(b)

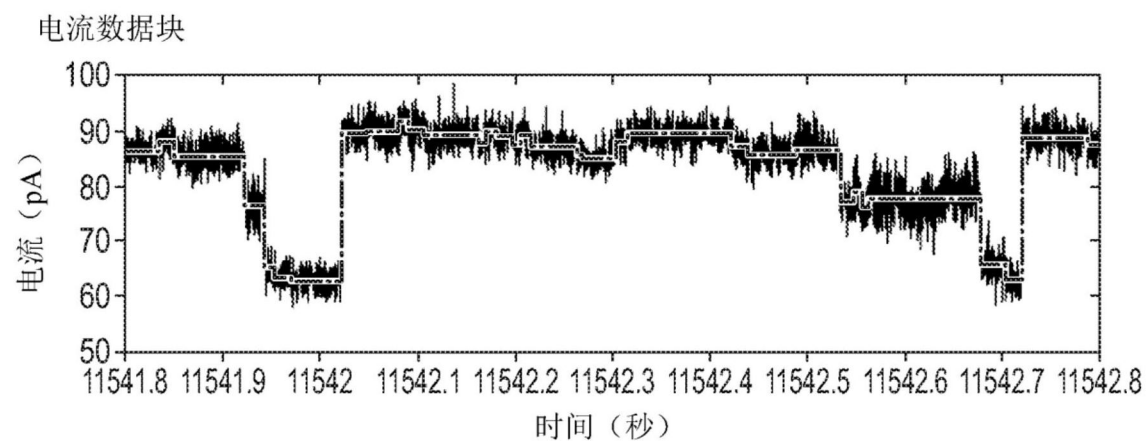


图7(c)

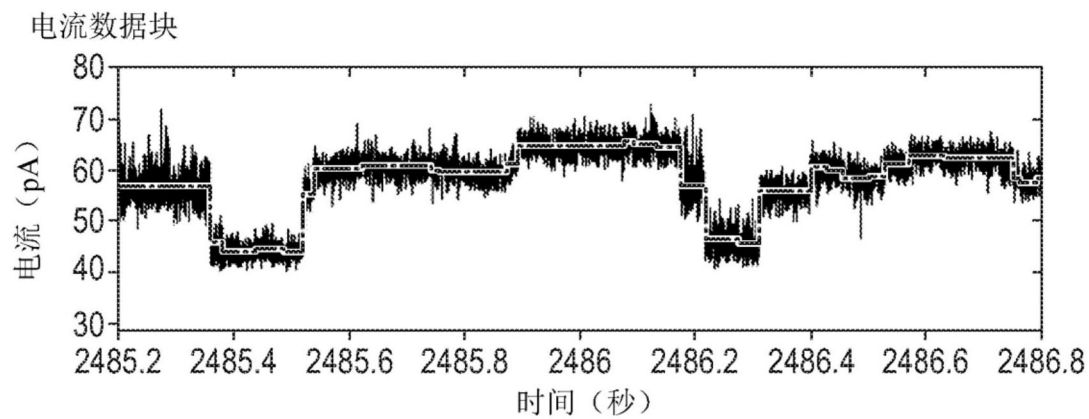


图8(a)

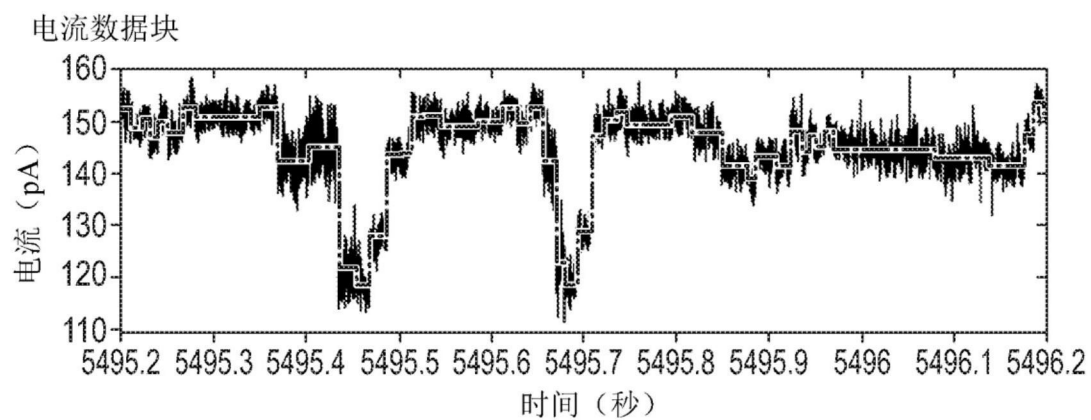


图8(b)

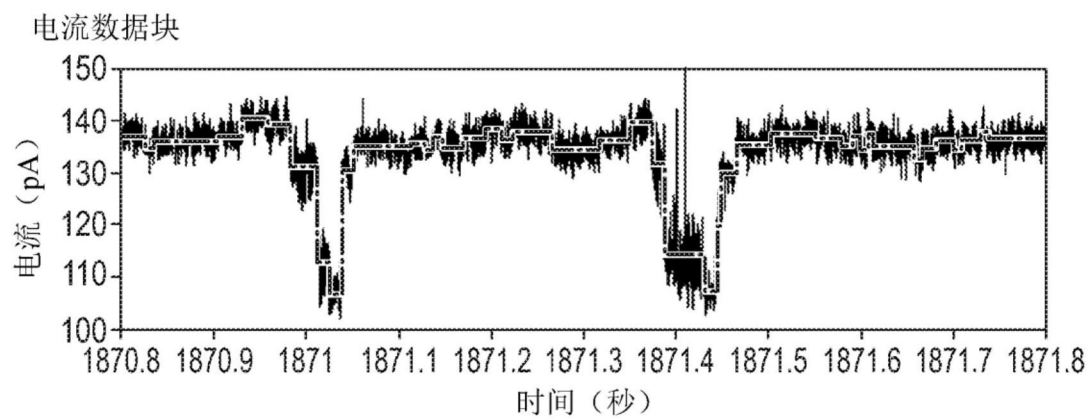


图8(c)

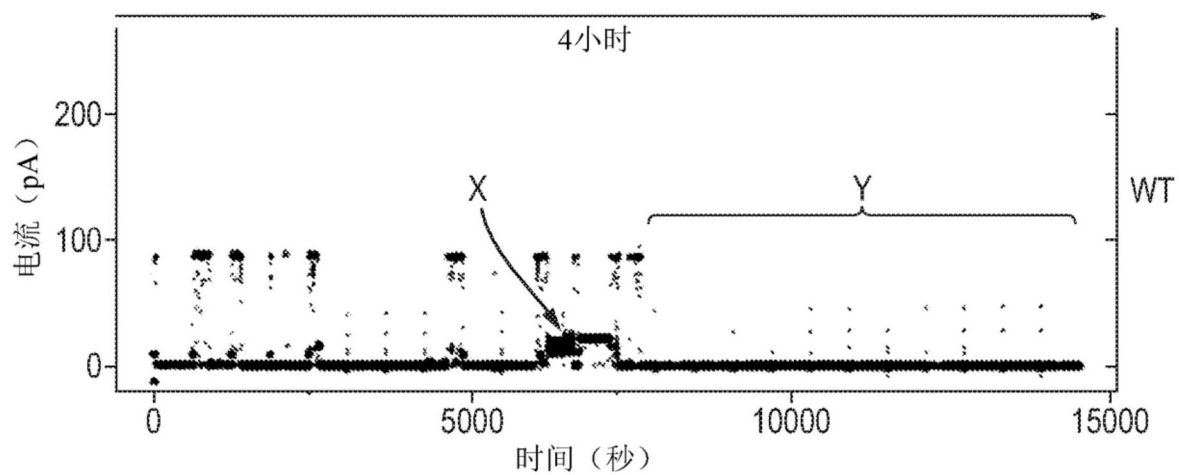


图9(a)

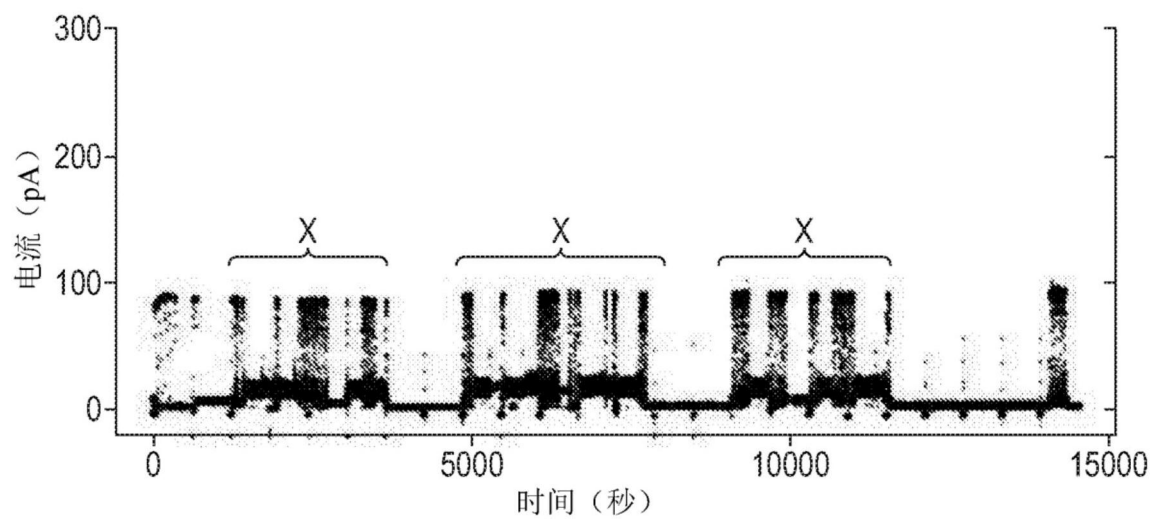


图9(b)

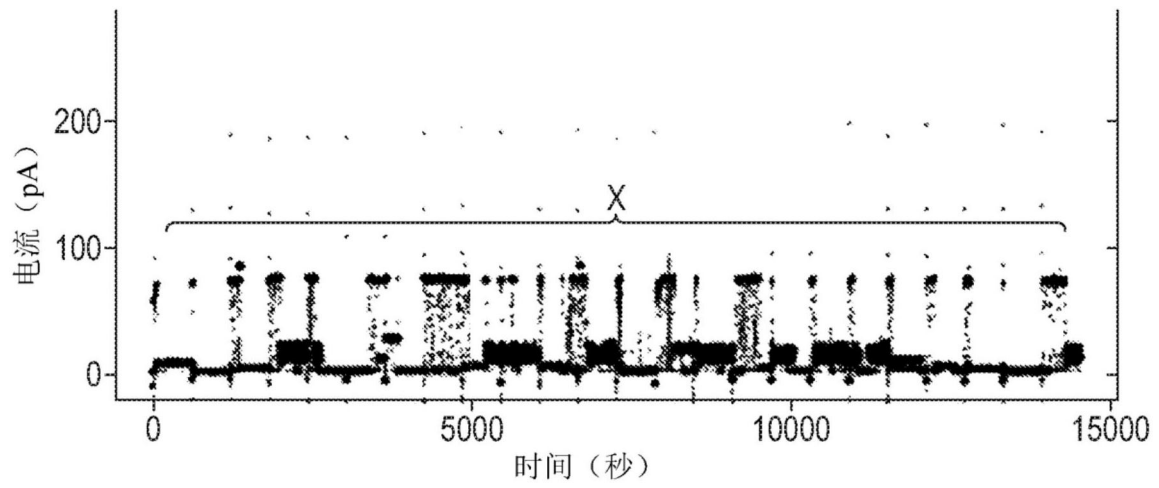


图9(c)

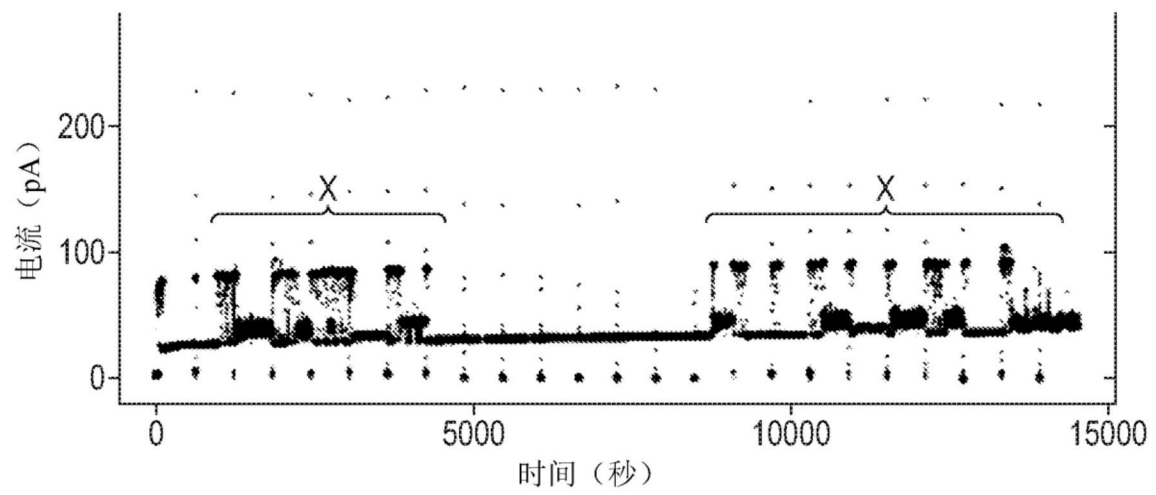


图10(a)

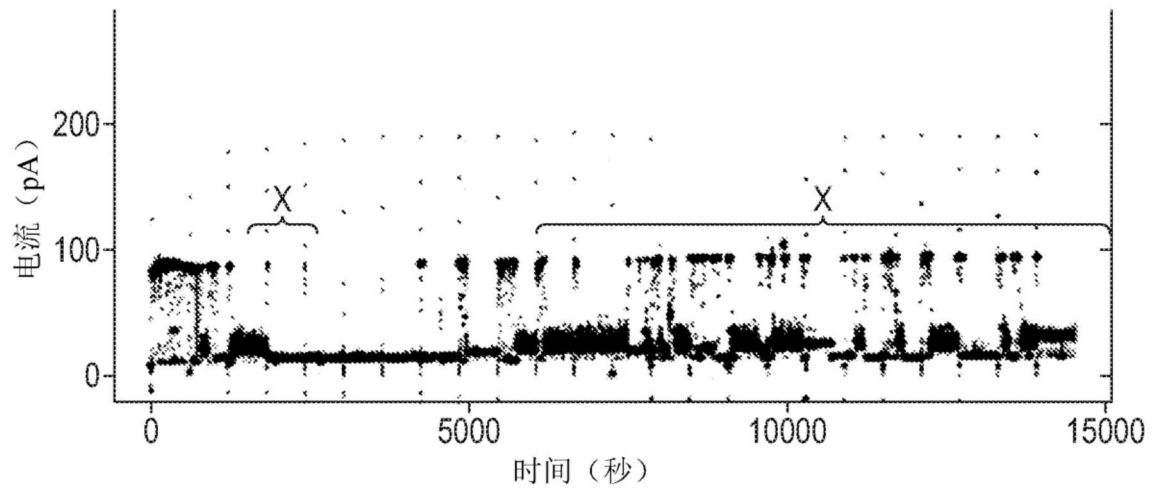


图10(b)

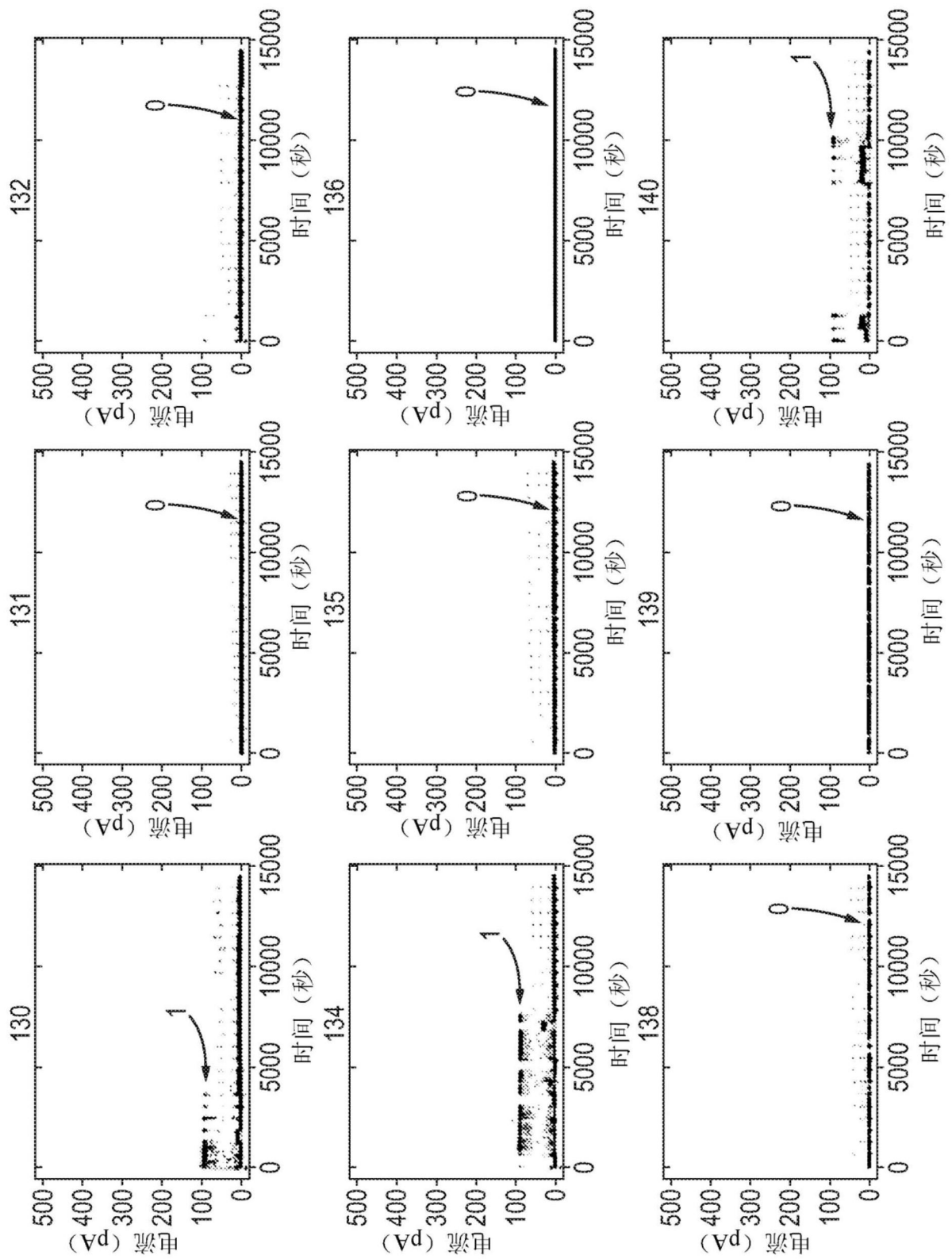


图11

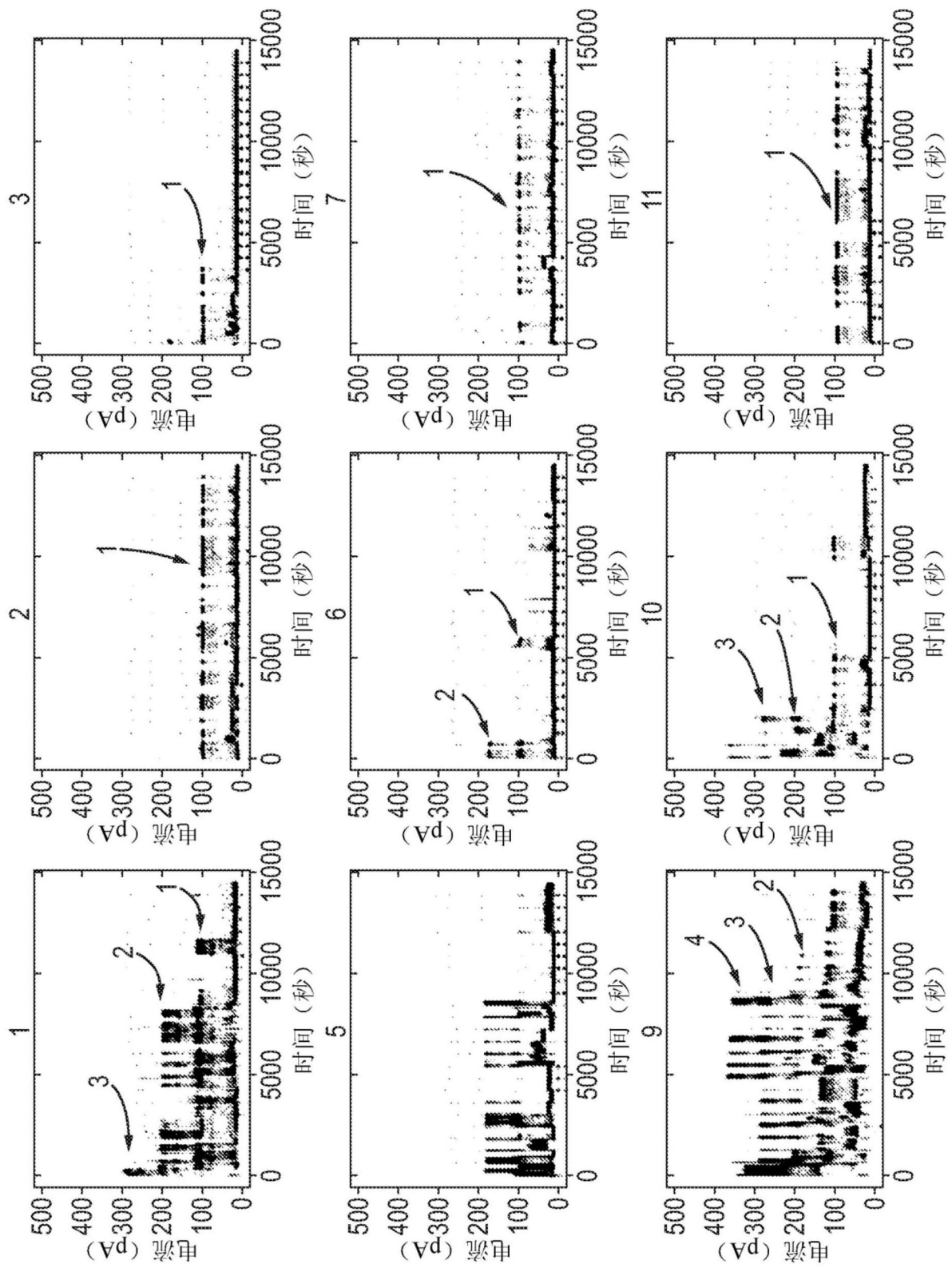


图12

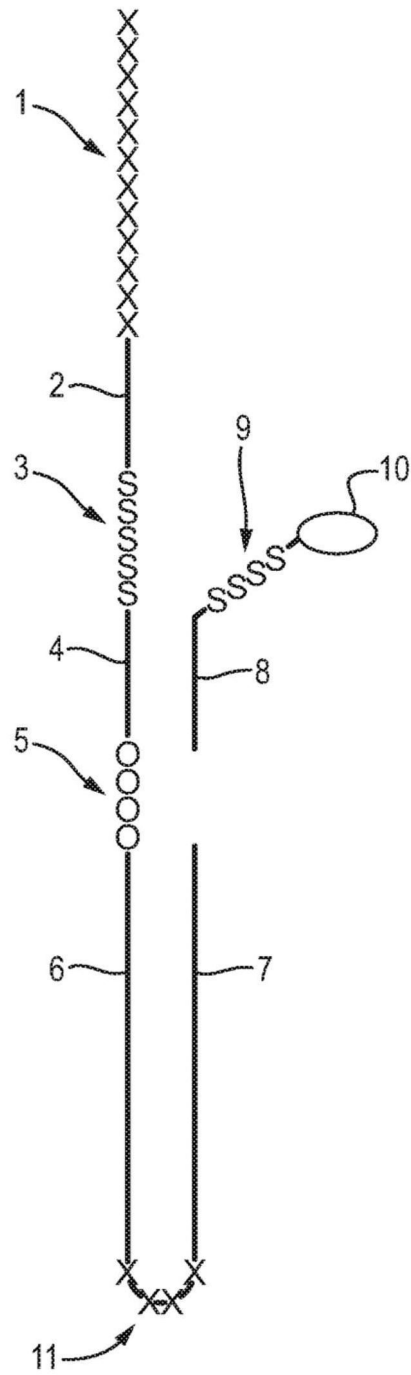


图13

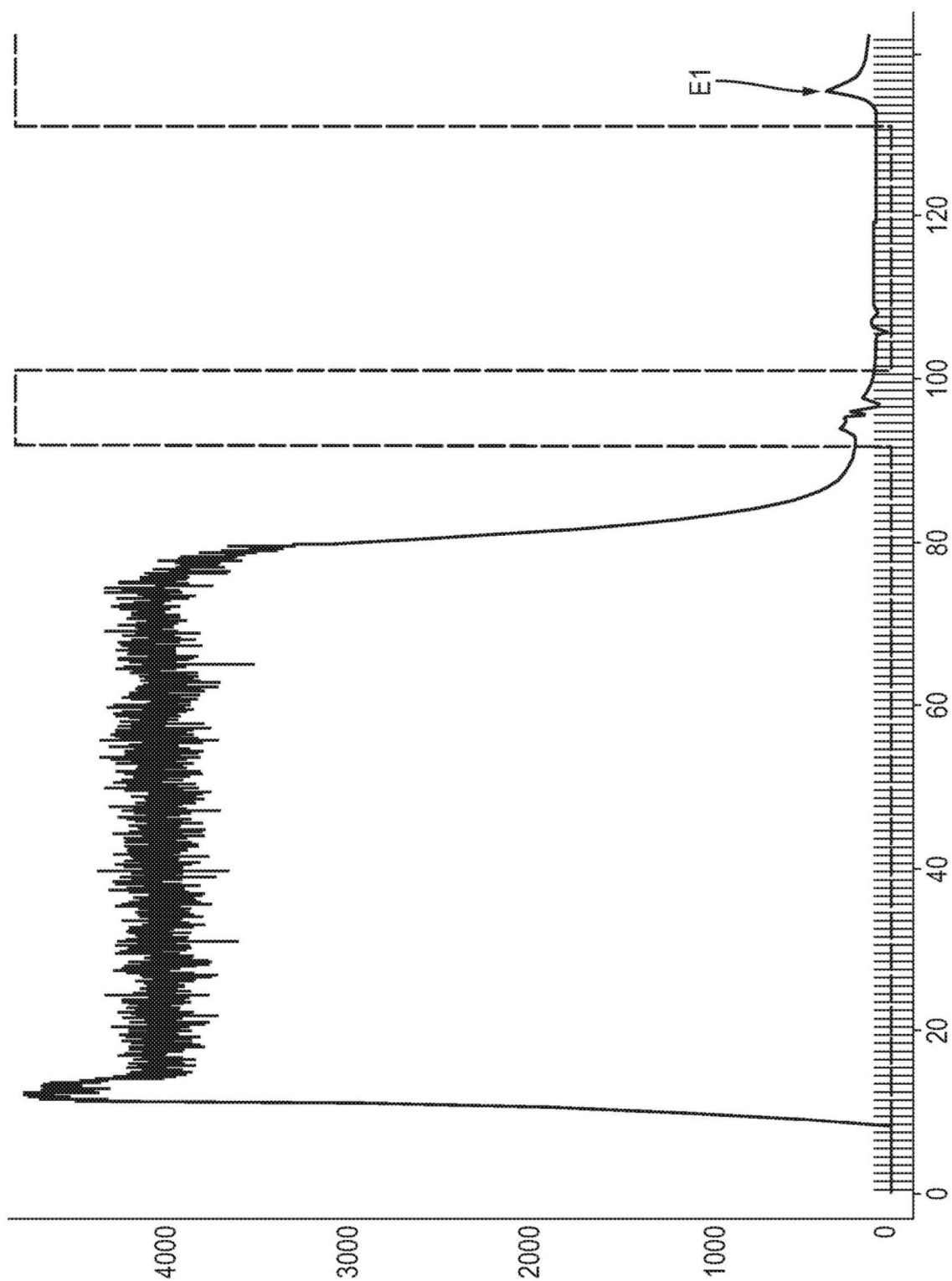


图14

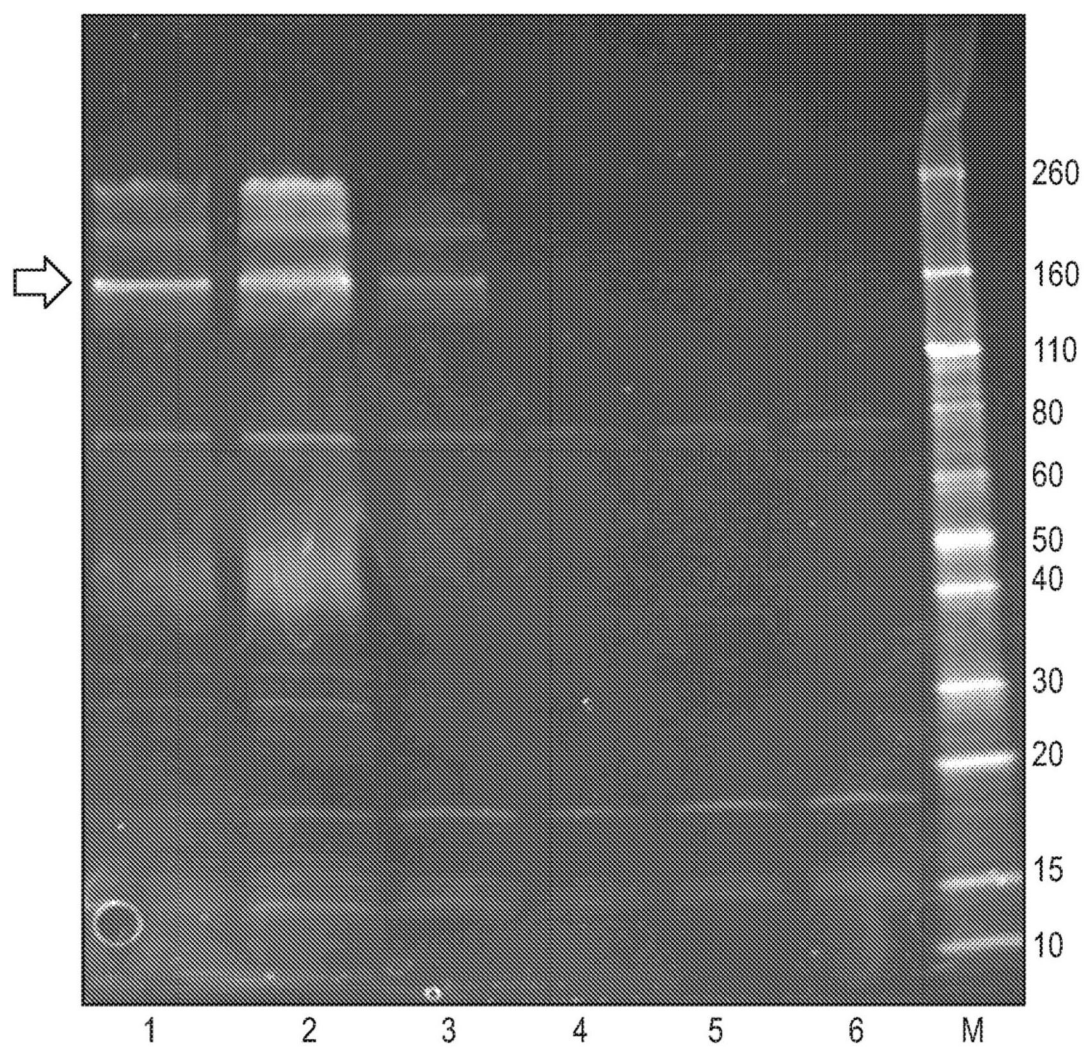


图15

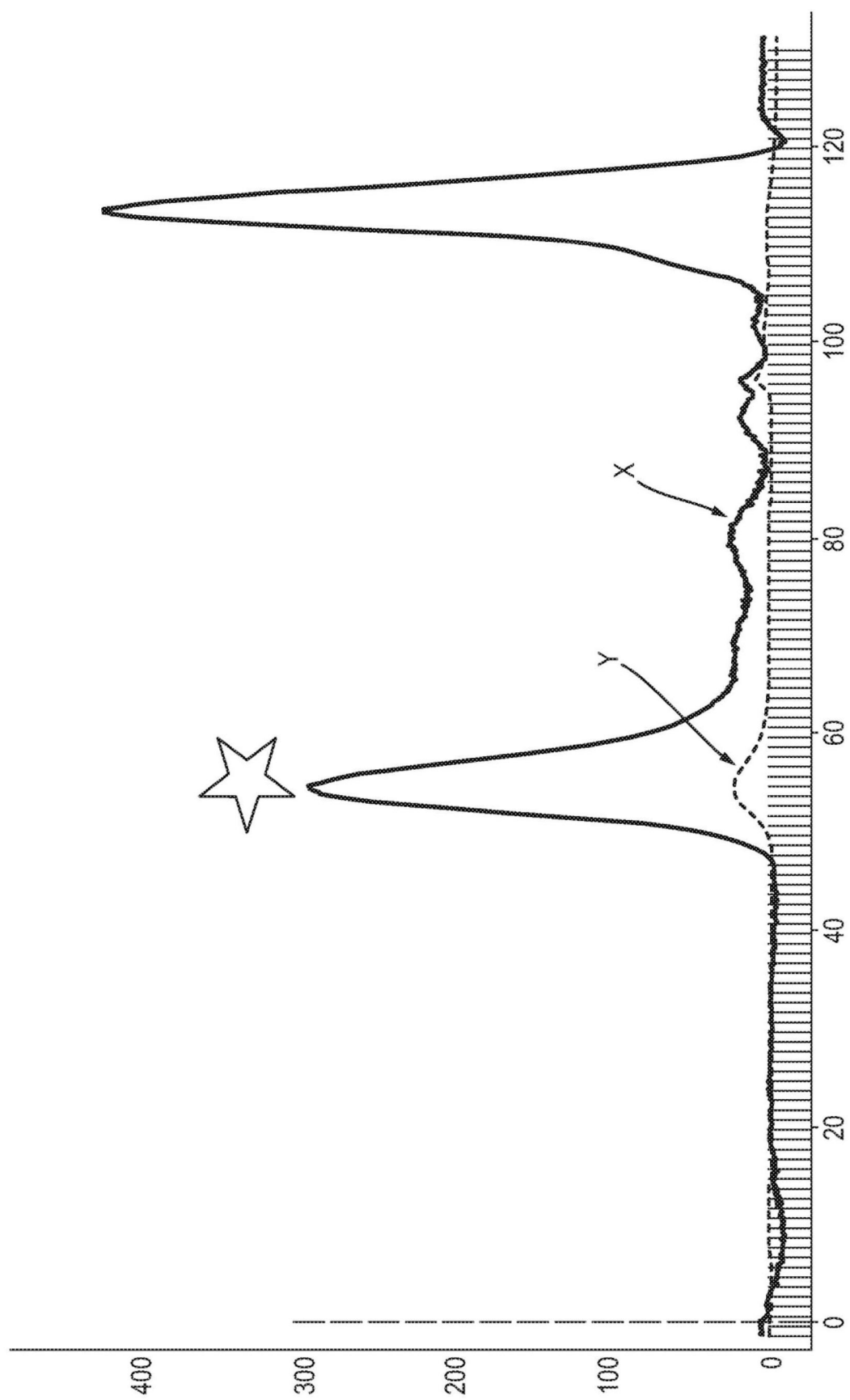


图16

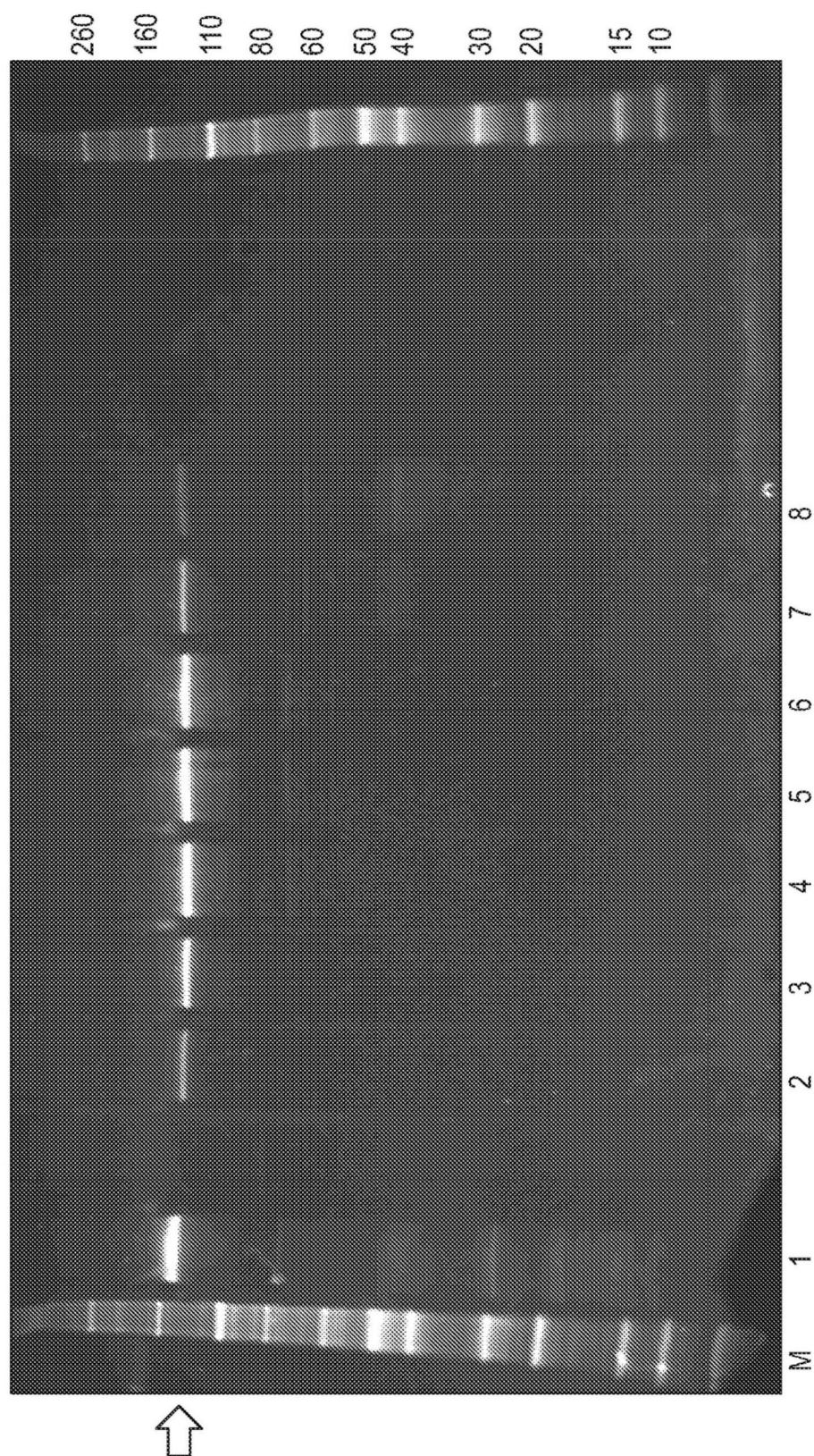


图17

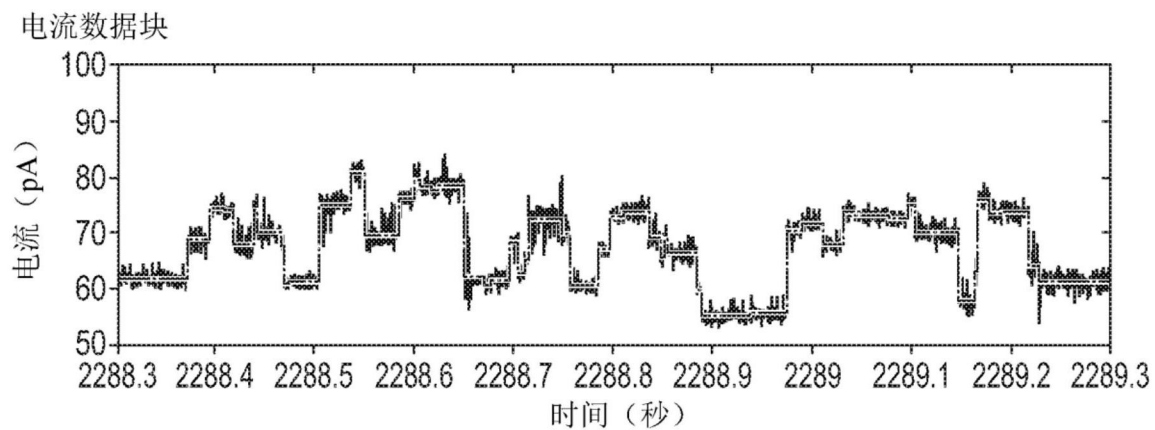


图18

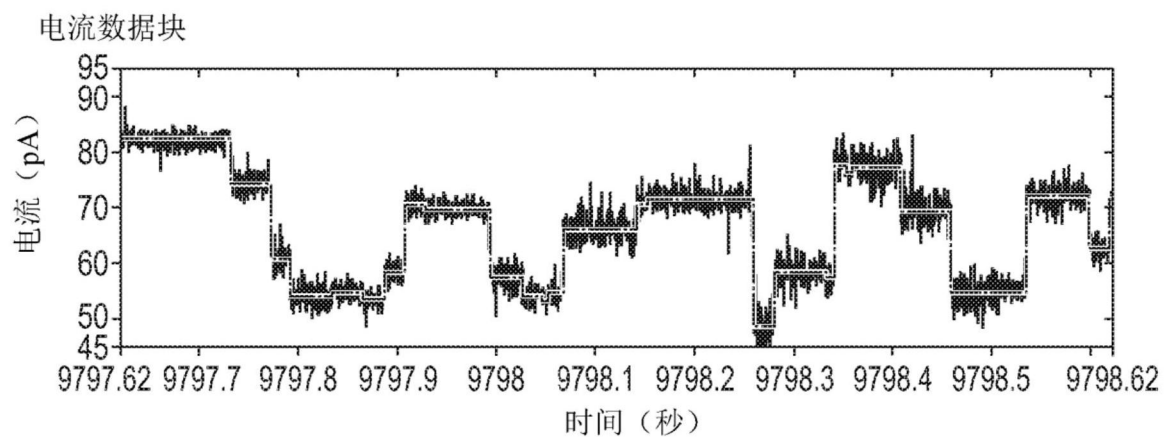


图19

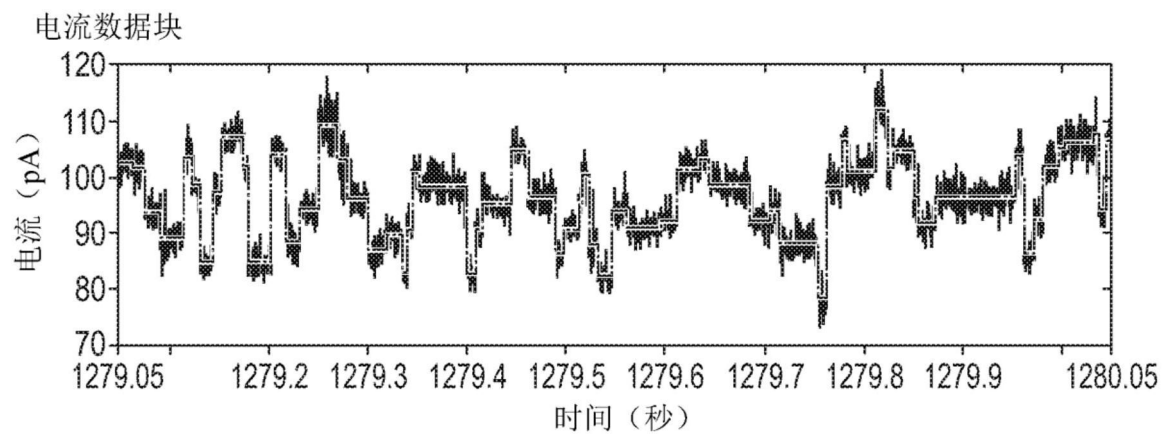


图20

电流数据块

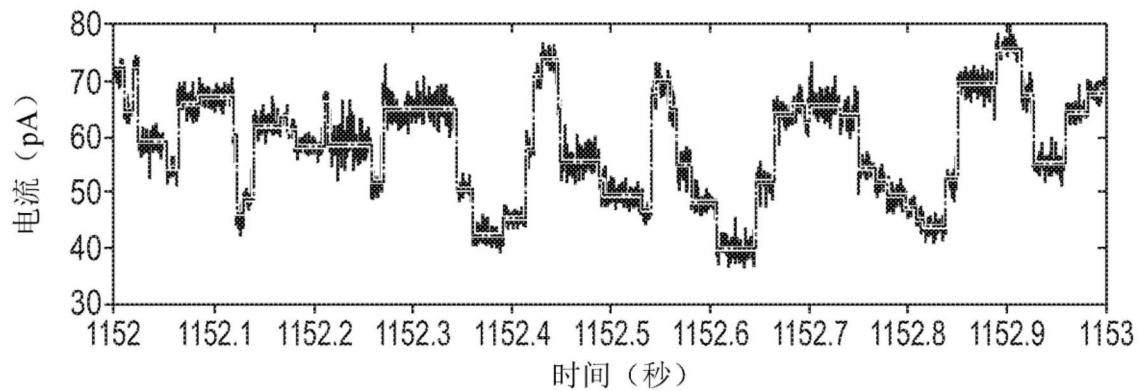


图21

电流数据块

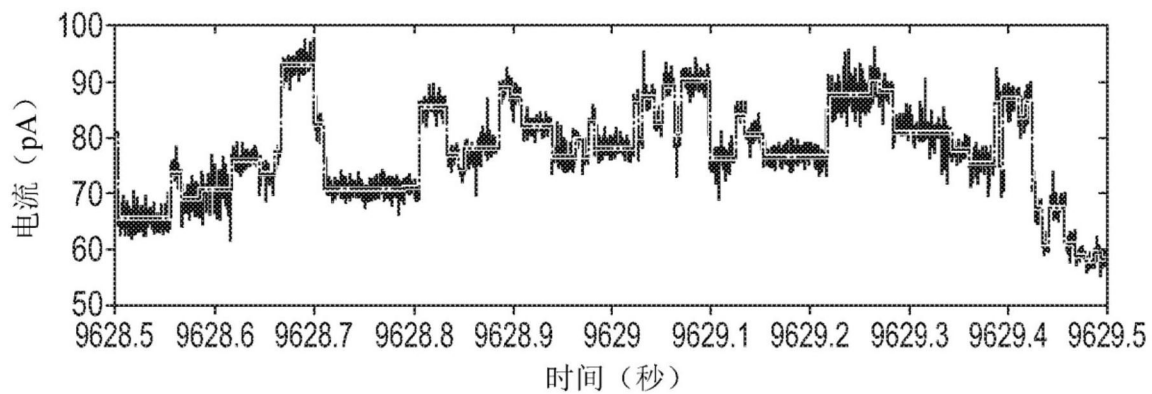


图22

电流数据块

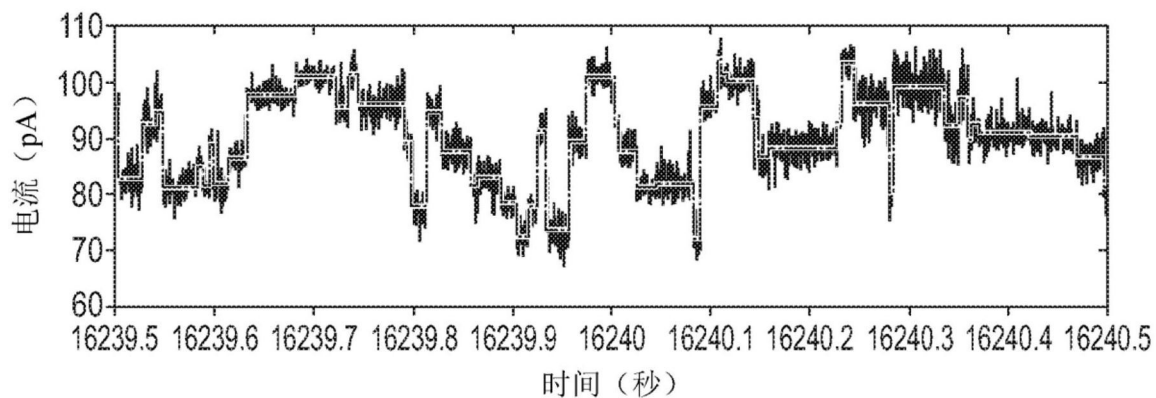


图23

电流数据块

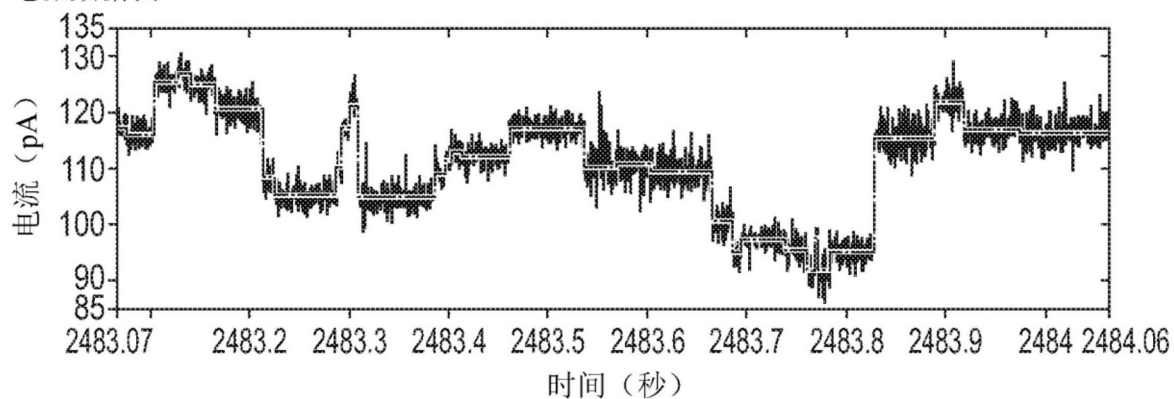


图24

电流数据块

CsgG-Eco-(Y51T/F56Q-StreptII(C))9

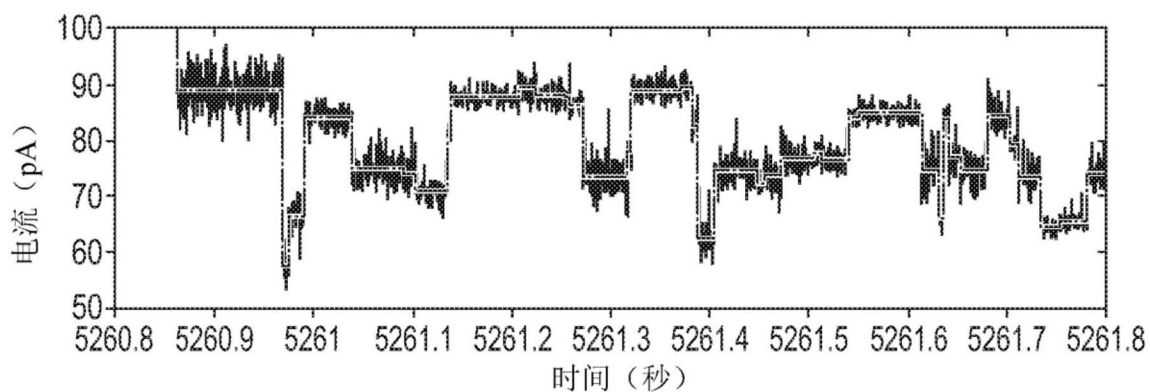


图25

电流数据块

CsgG-Eco-(N55S/F56Q-StreptII(C))9

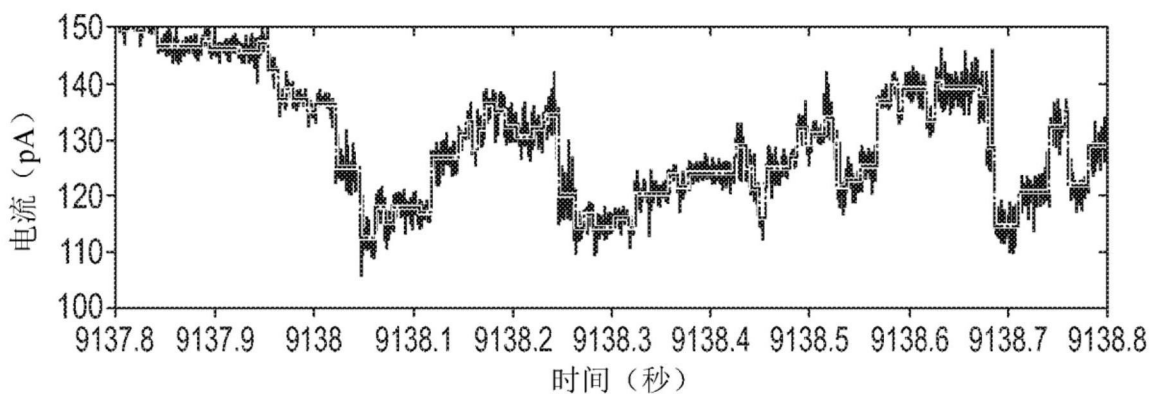


图26

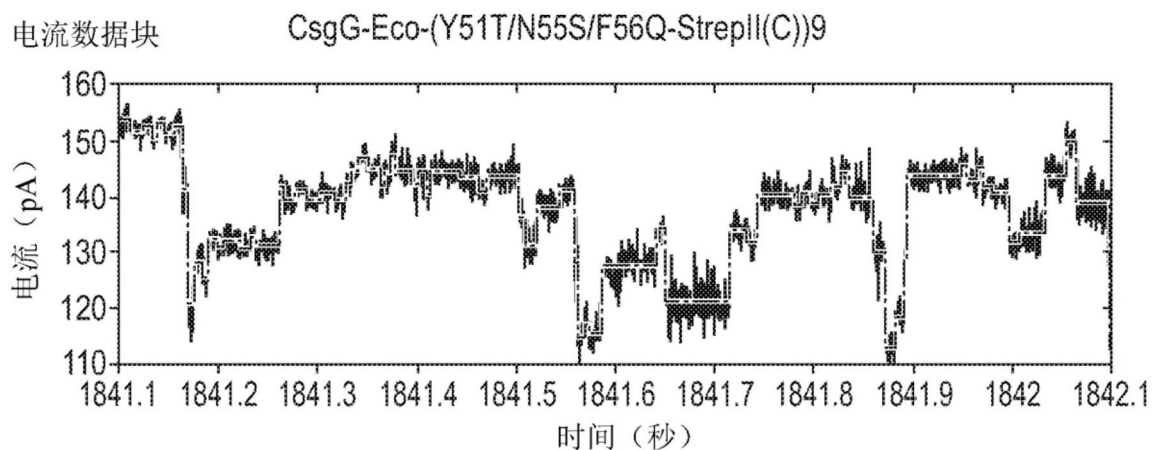


图27

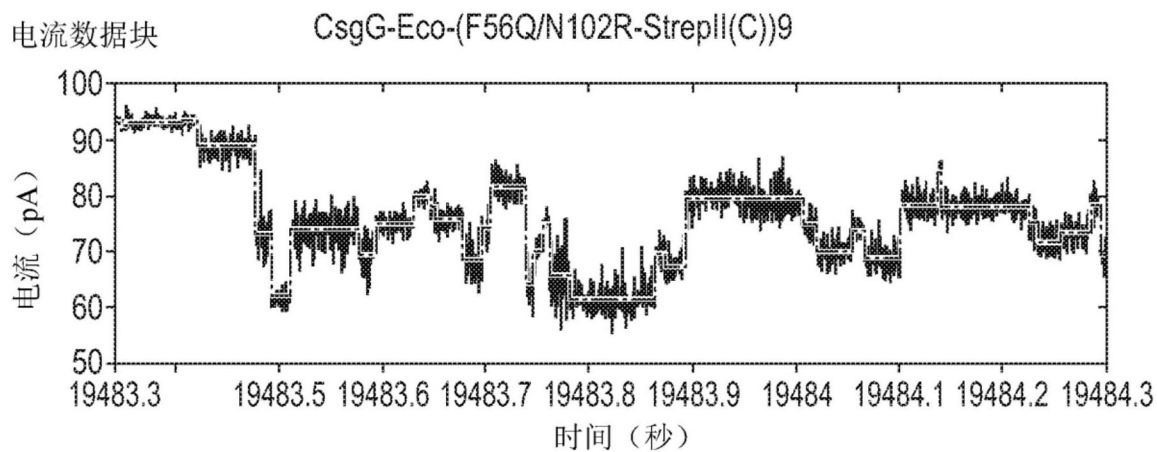


图28

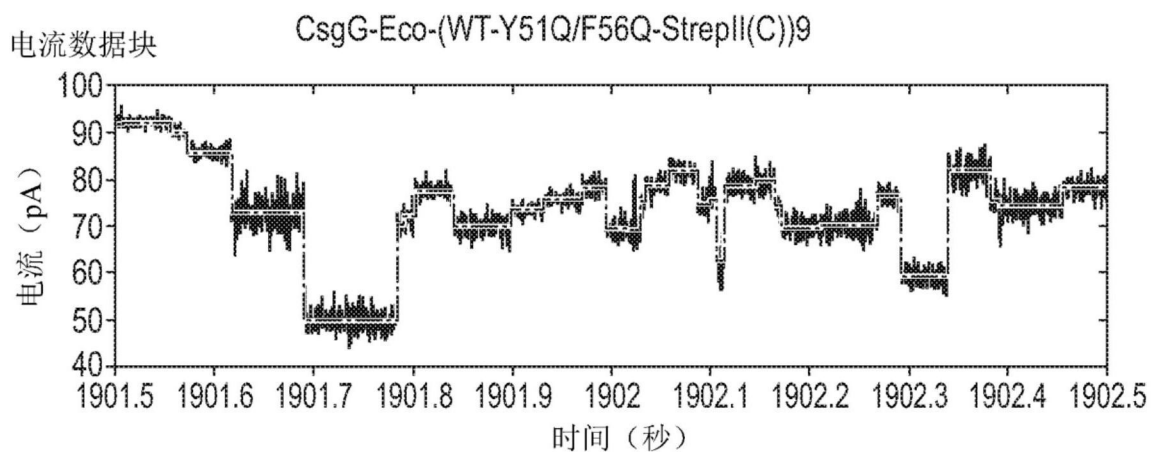


图29

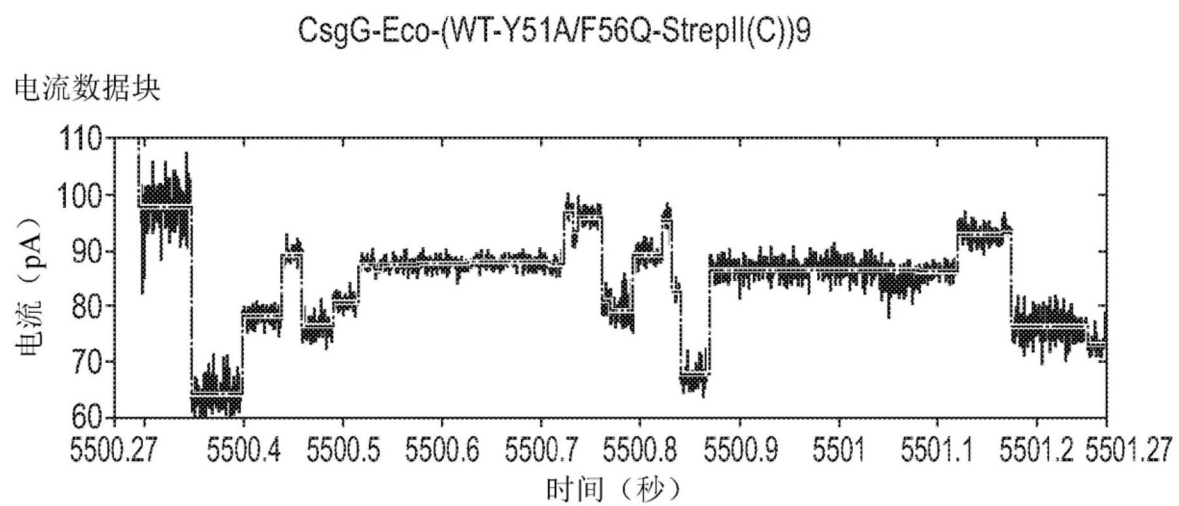


图30

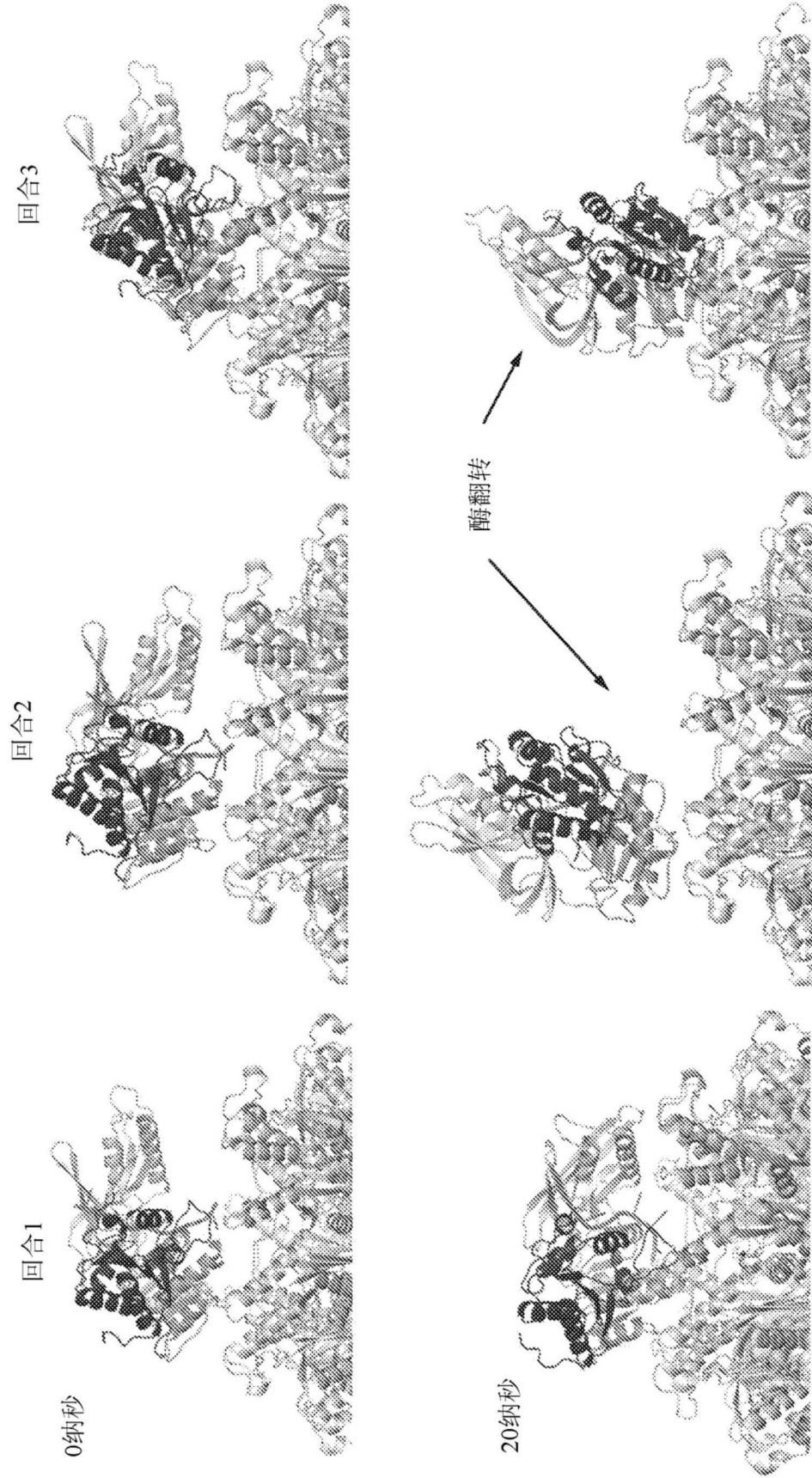


图31

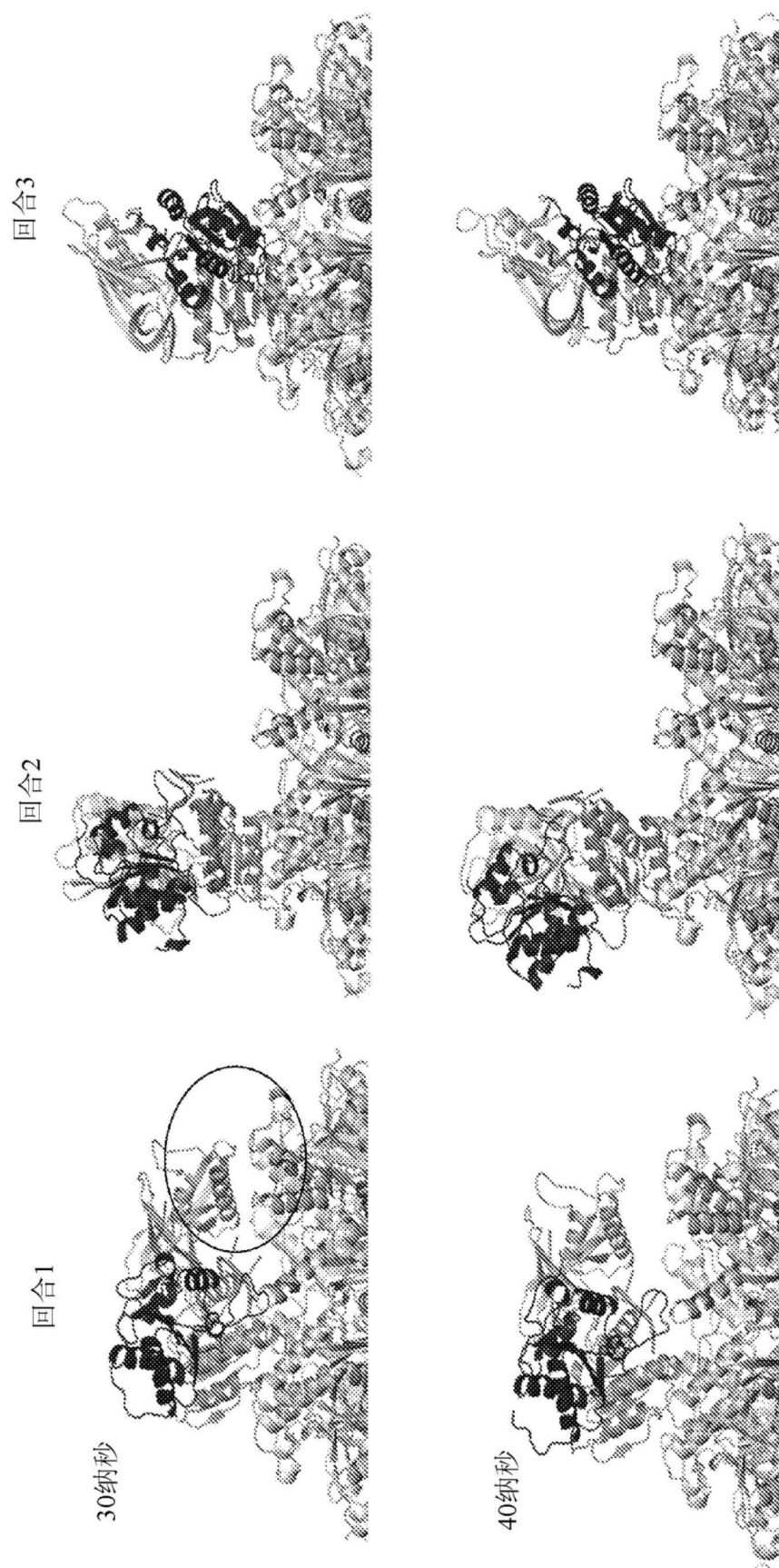


图32

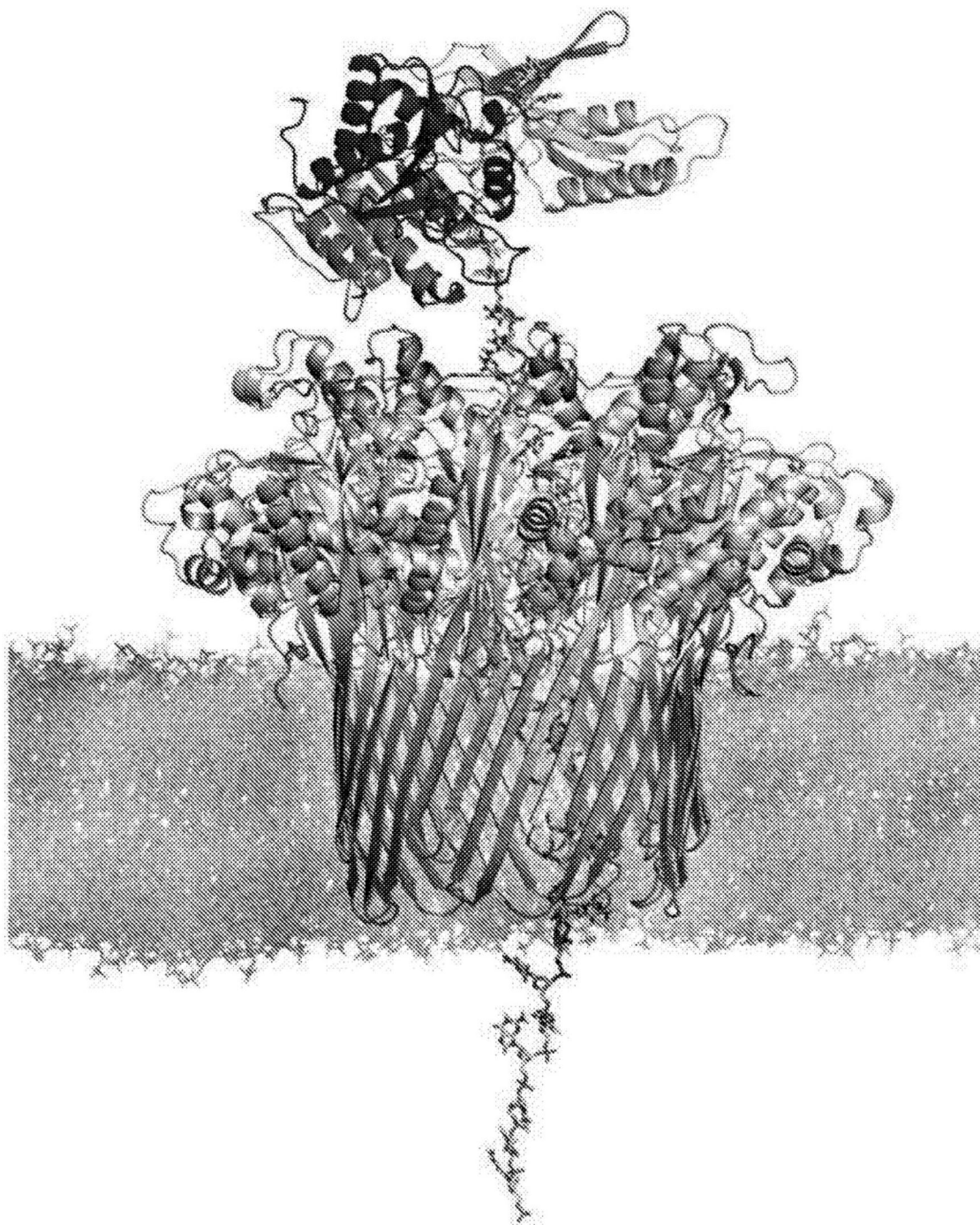


图33

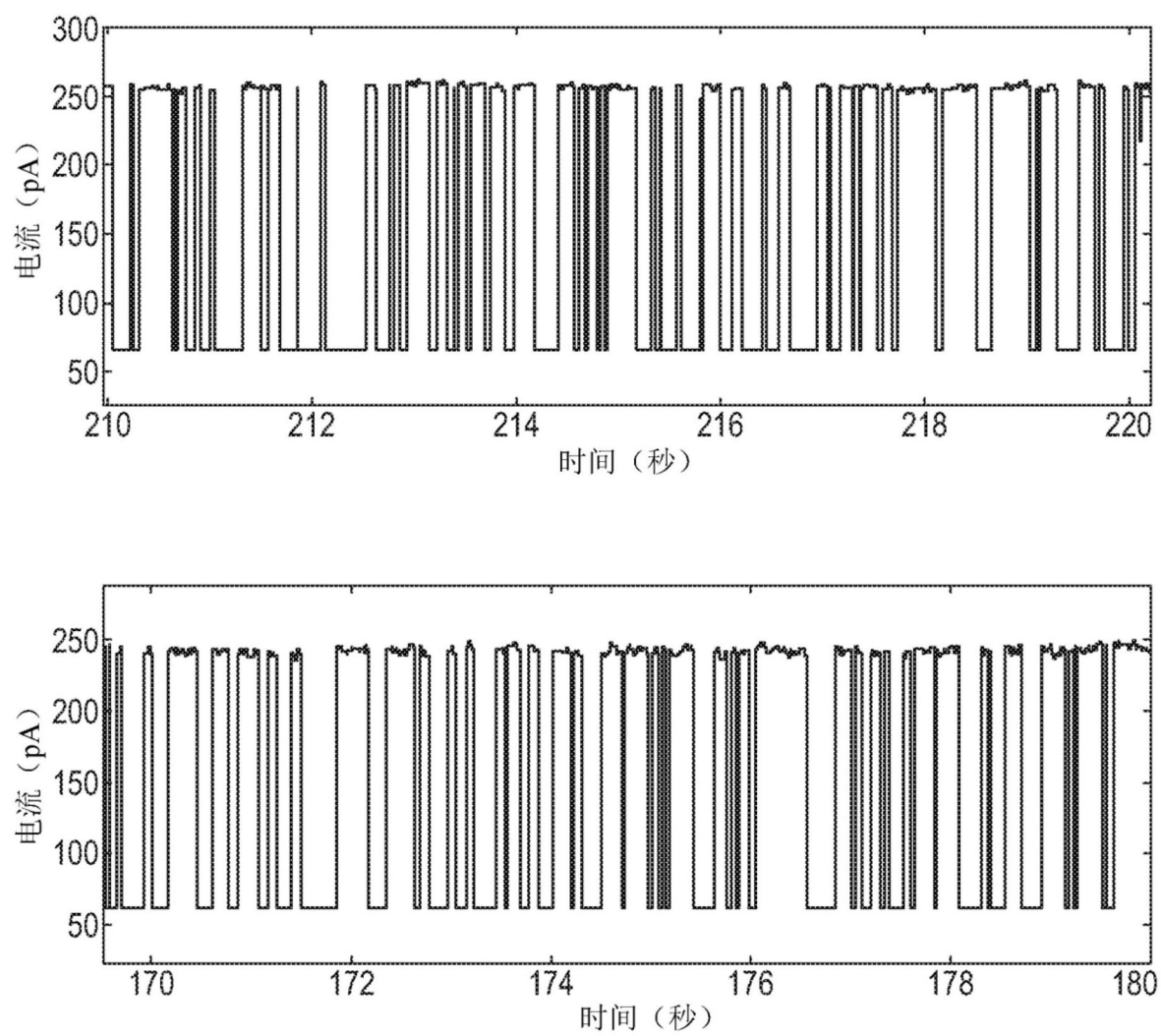


图34

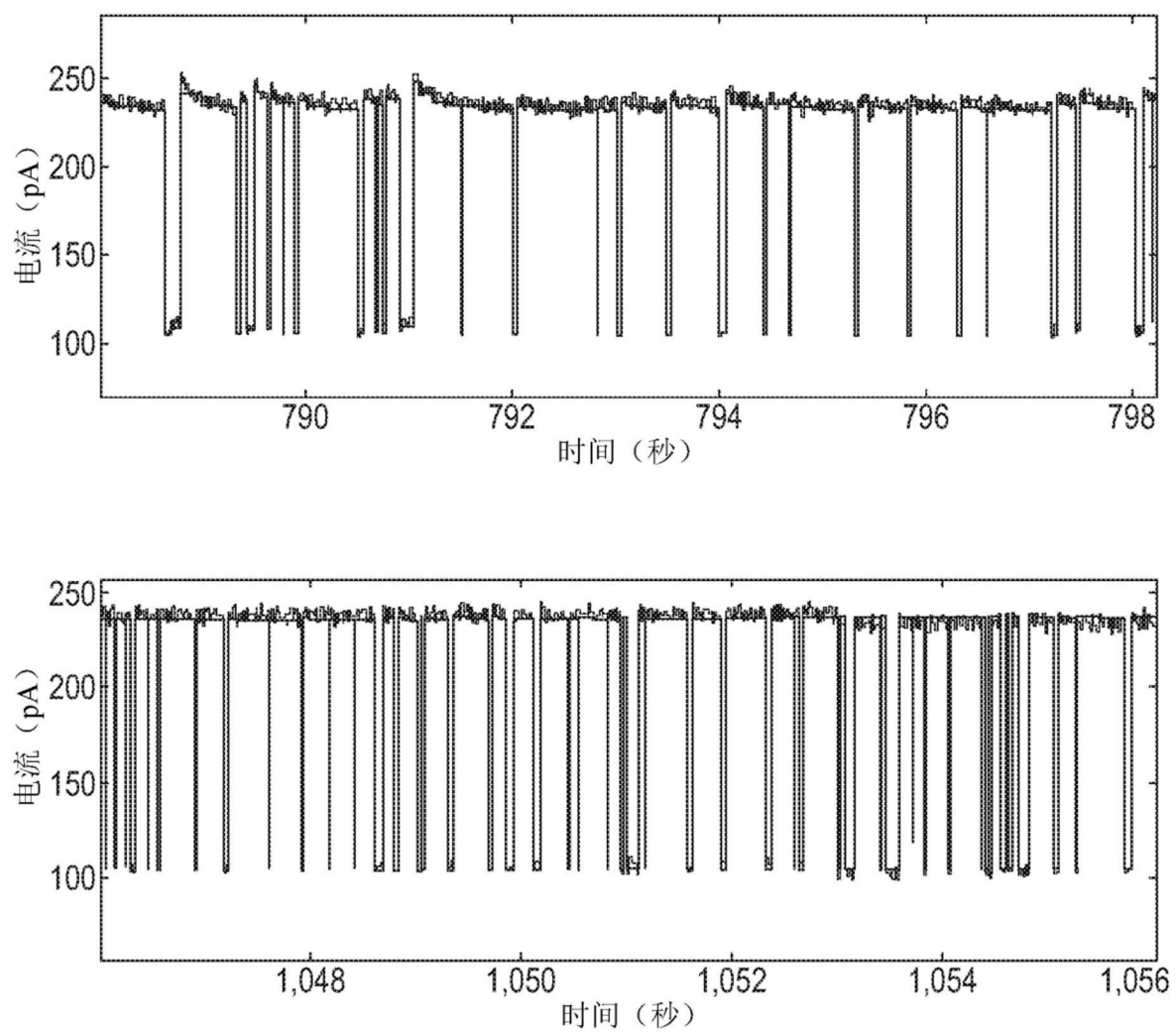


图35

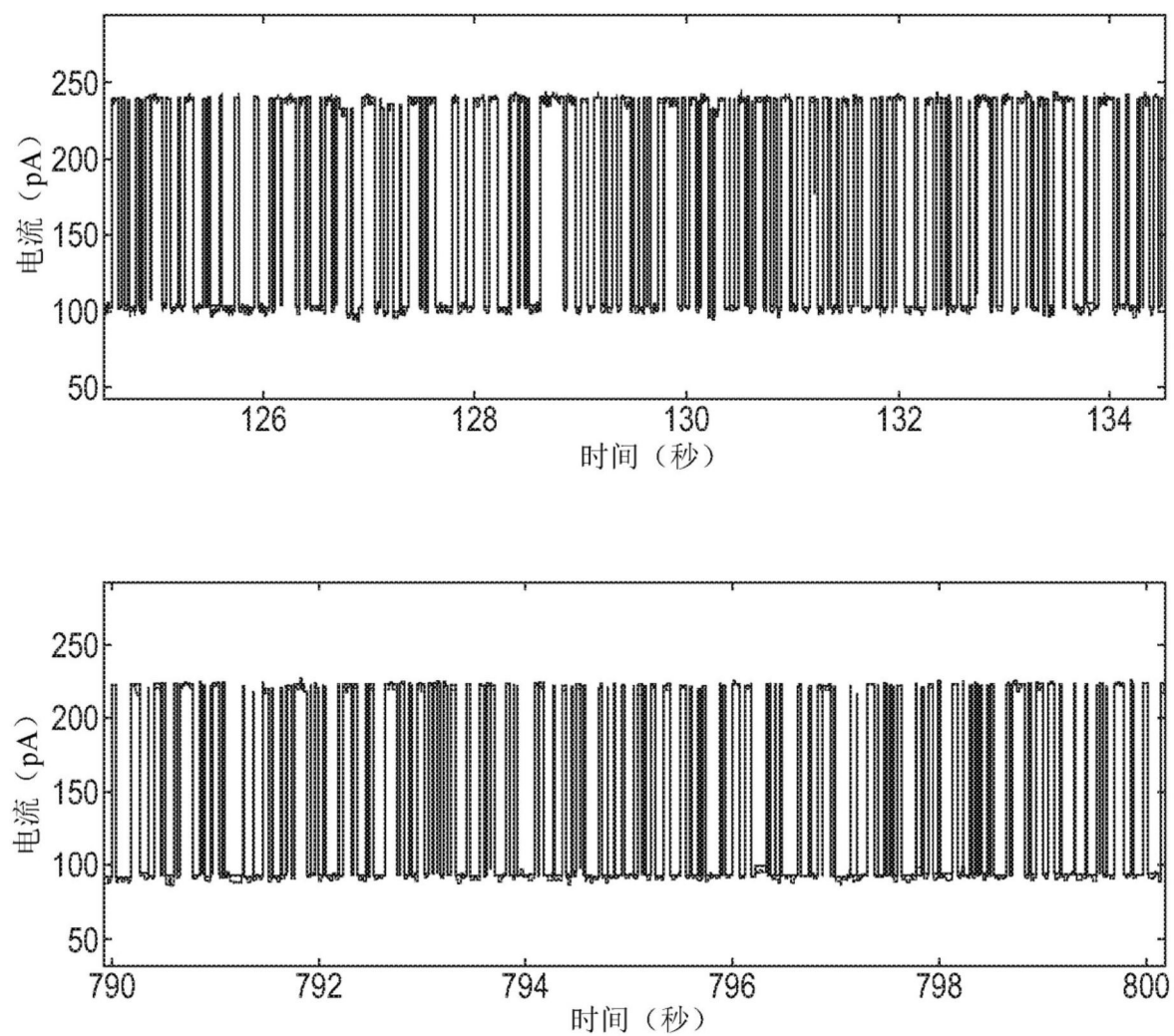


图36

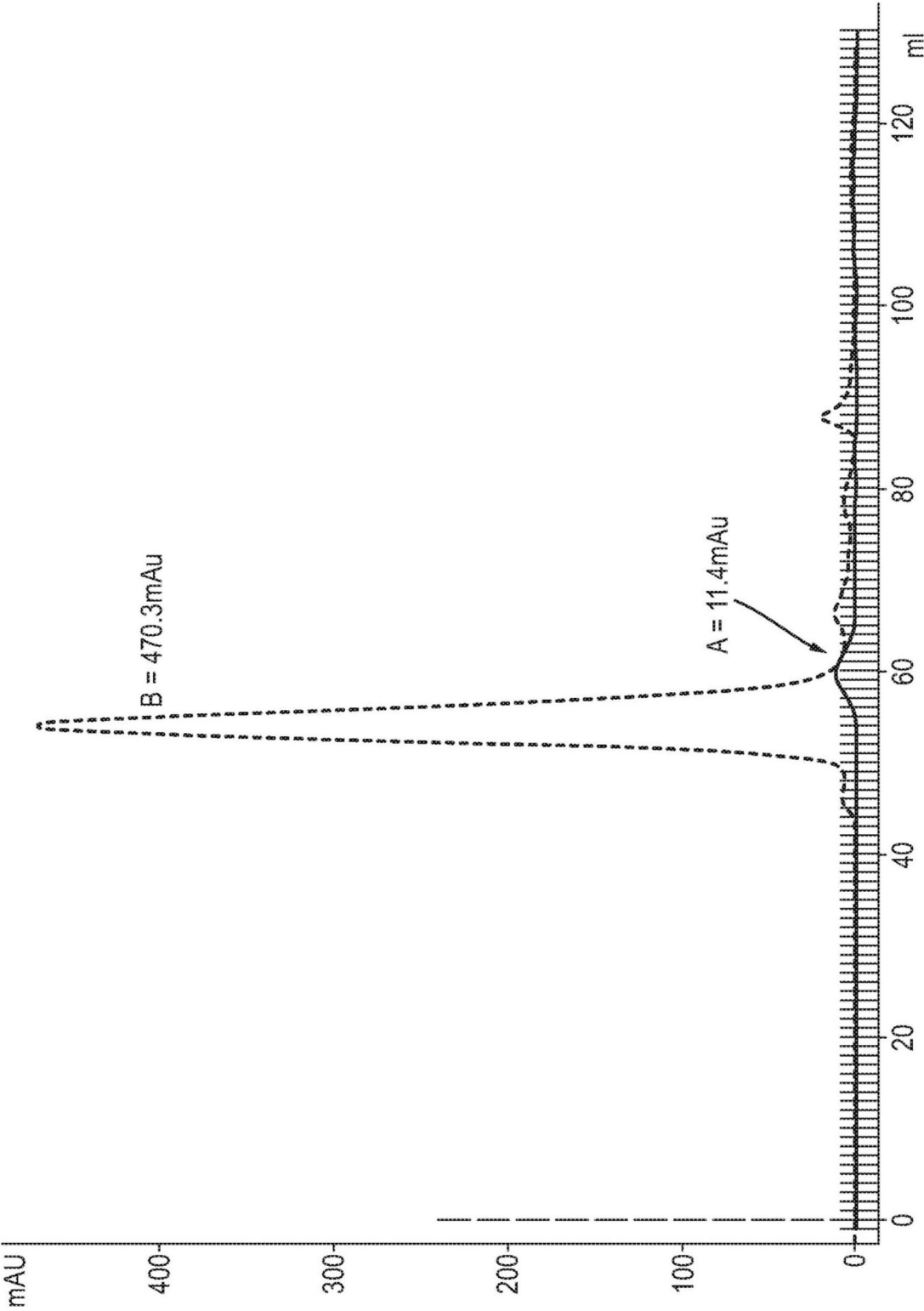


图37

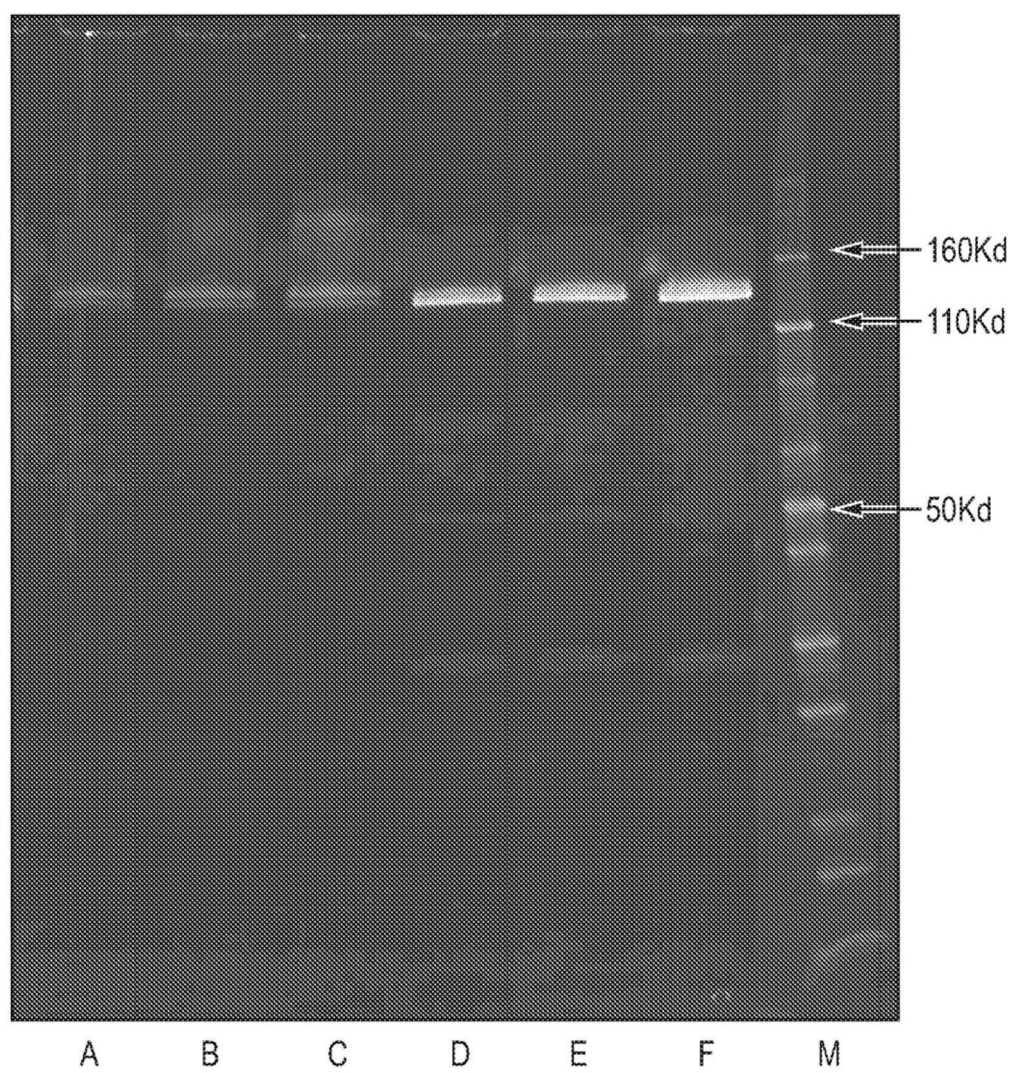


图38

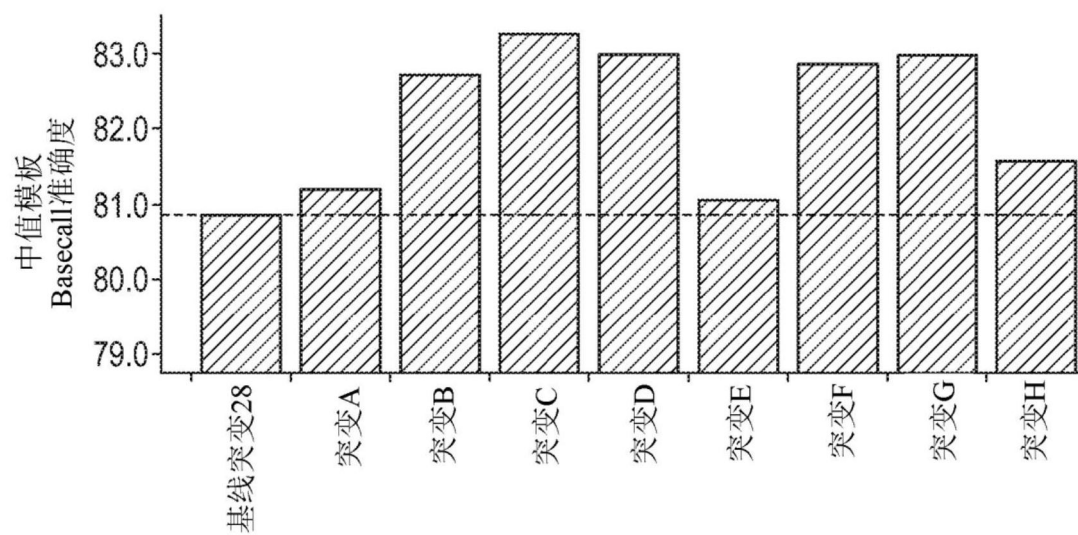


图39

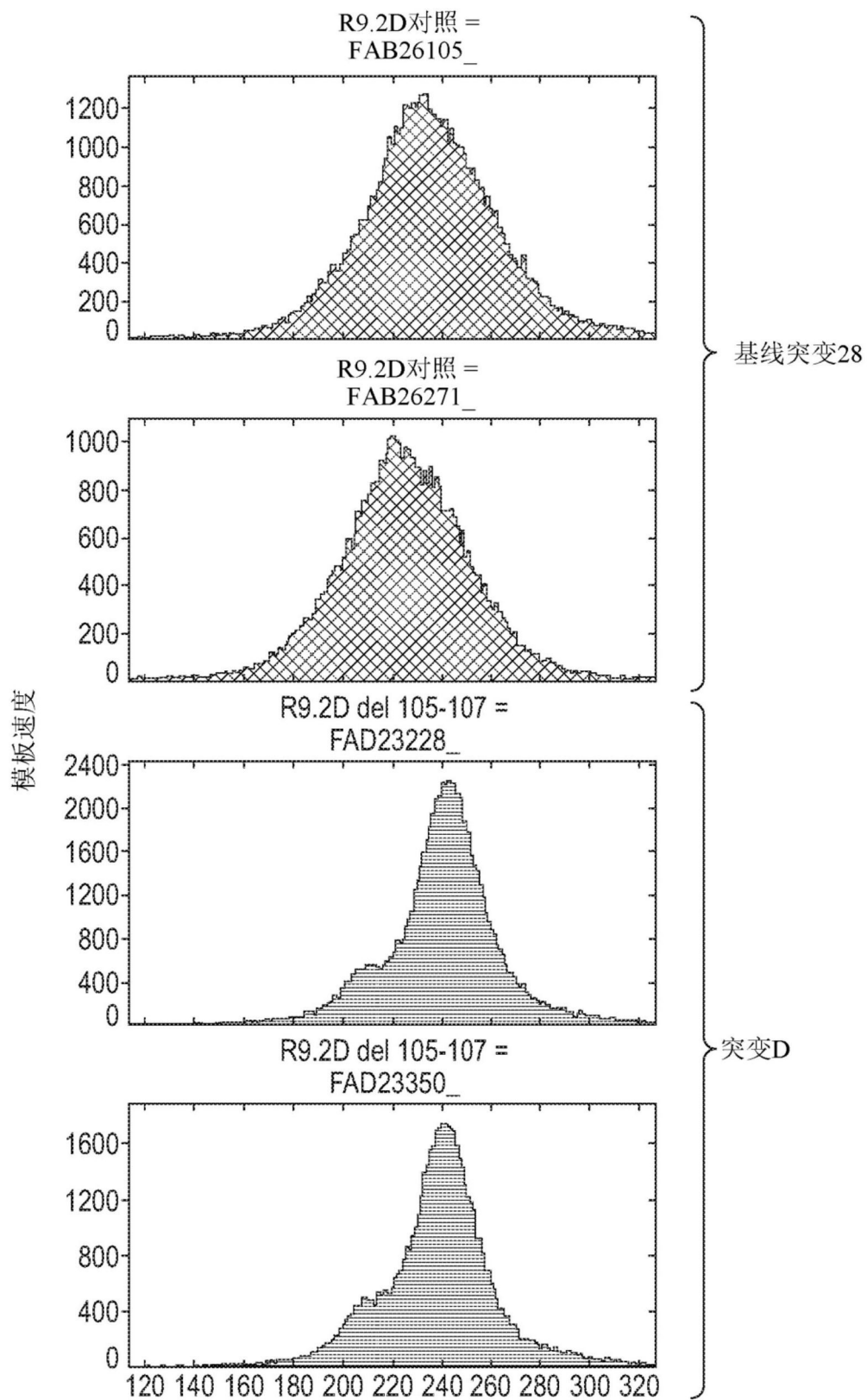


图40A

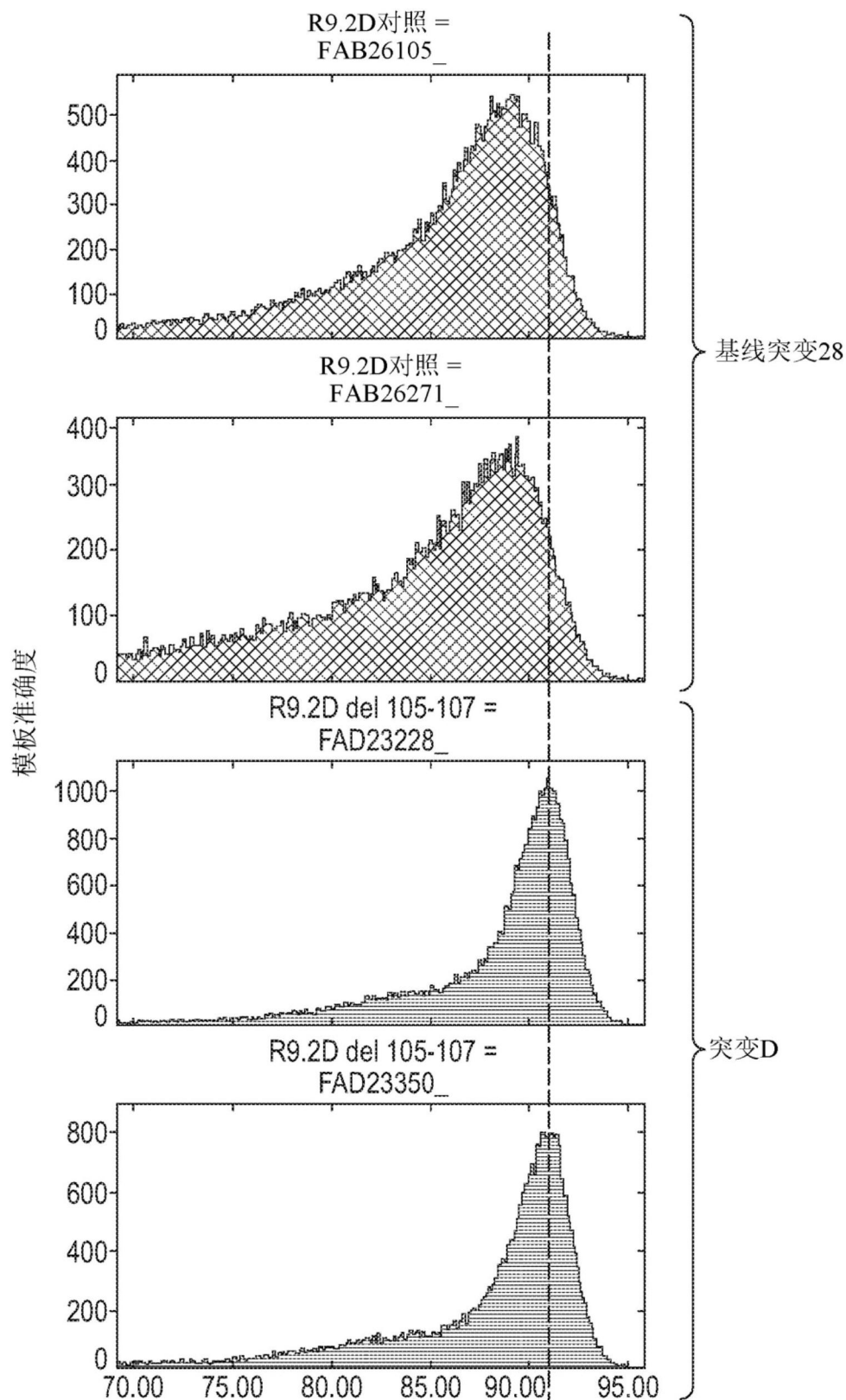


图40B

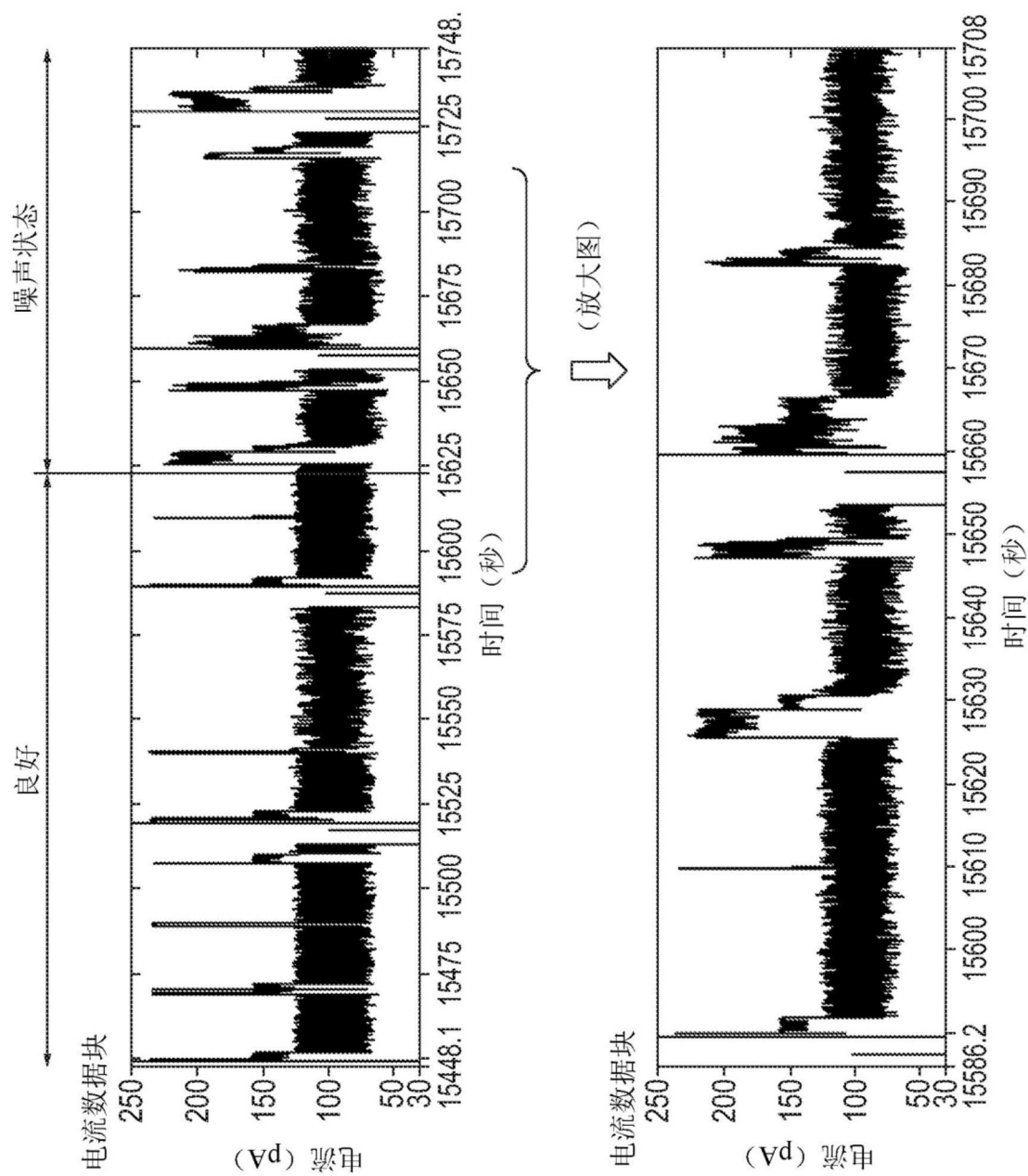


图41

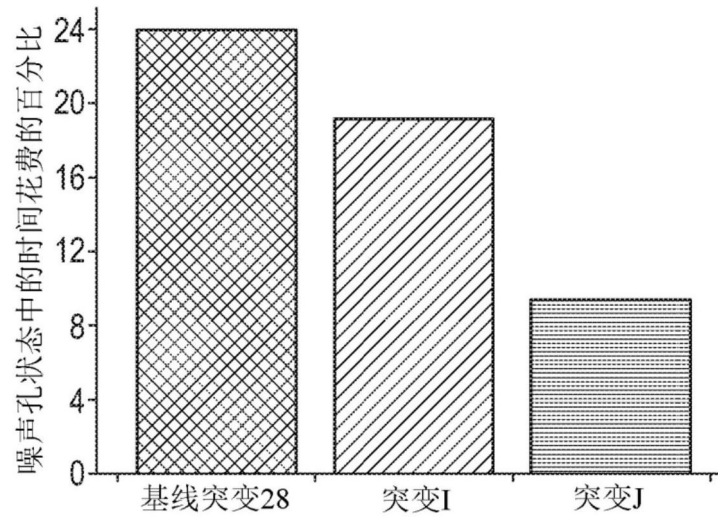


图42

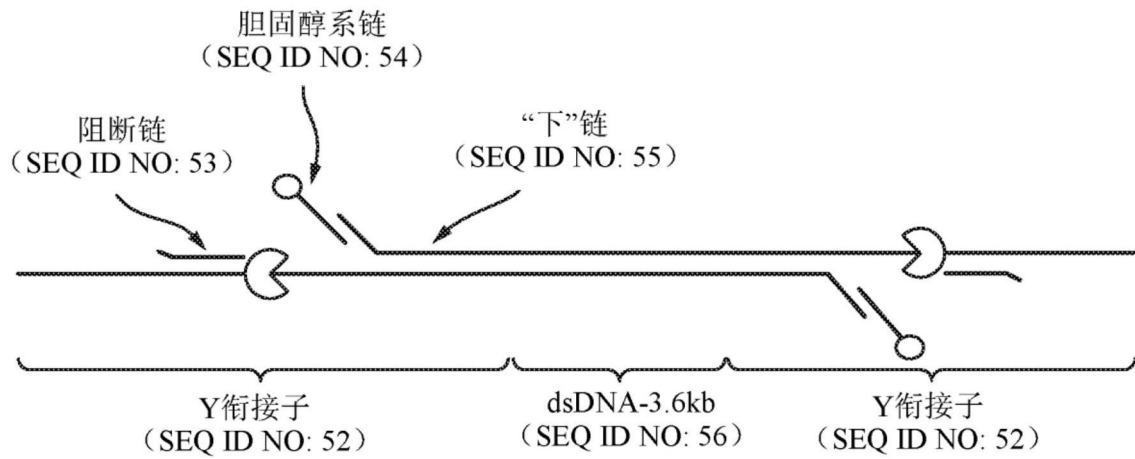


图43

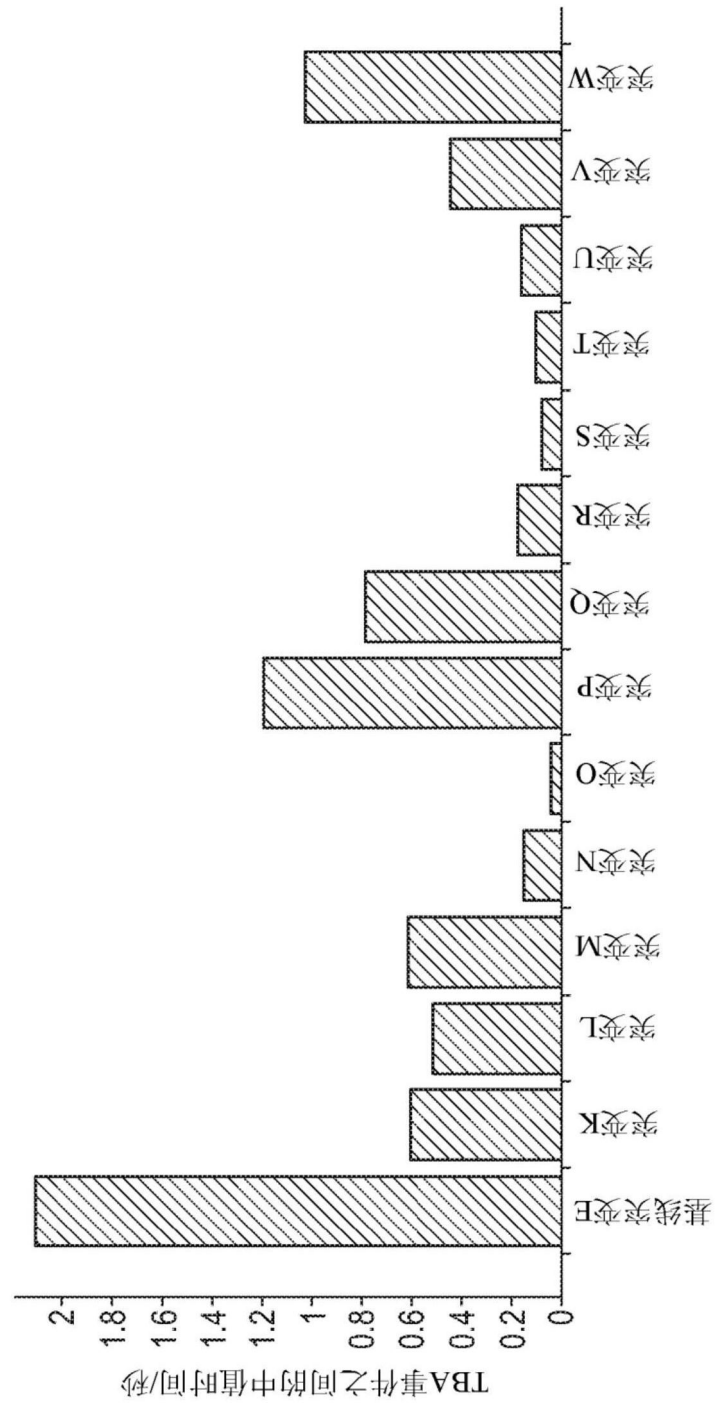


图44

SEQ_ID_NO_2	1 CLT--APPKEAARPTLMPRAQSYKDLTHLPAPTGKIFVS
SEQ_ID_NO_5	1 CLT--APPKEAAKPTLMPRAQSYKDLTHLP IPTGKIFVS
SEQ_ID_NO_6	1 CLT--TPPKEAAKPTLMPRAQSYKDLTHLPVPTGKIFVS
SEQ_ID_NO_7	1 CLT--APPKEAAKPTLMPRAQSYRDLTHLPAPTGKIFVS
SEQ_ID_NO_27	1 CLT--APPKEAAKPTLMPRAQSYRDLTNLPDPKPKGLFVS
SEQ_ID_NO_28	1 CLT--AAPKEAARPTLLPRAPSYTDLTHLPSPQGRIFVS
SEQ_ID_NO_29	1 CIT--SPPKQAAKPTLLPRSQSYQDLTHLPEPQGRIFVS
SEQ_ID_NO_30	1 CLT--APPKQAAKPTLMPRAQSYQDLTHLPEPAGKLFVS
SEQ_ID_NO_32	1 -IT--EVPKEAAKPTLMPRASTYKDLVALPKPNGKIIVS
SEQ_ID_NO_36	1 CATHIGSPVADEKATLMPRSVSYKELISLPKPKGKIVAA
SEQ_ID_NO_3	1 -----MPRAQSYKDLTHLPMPPTGKIFVS
SEQ_ID_NO_35	1 -----PETSKEPTLMARGTAYQDLVSLPLPKGKVYVS
SEQ_ID_NO_31	1 CIT--TPPQEA AKPTLLPRDATYKDLVSLPQPRGKIYVA
SEQ_ID_NO_40	1 -----LTRRMSTYQDLIDMPAPRGKIVTA
SEQ_ID_NO_33	1 -----PETSESPTLMORGANYIDLISLPKPKQKIFVS
SEQ_ID_NO_34	1 -----PDASESPTLMORGATYLDLISLPKPKQKIYVS
SEQ_ID_NO_37	1 -----ASSSLMPKGESYYDLINLPAPOGVMLAA
SEQ_ID_NO_39	1 -----MPKSDTYDDLIGLPHPOGSMMLAA
SEQ_ID_NO_38	1 -----QDSETPTLTPRASTYDDLINMPRPKGRMLAV
SEQ_ID_NO_41	1 -----PSDPERSTMGELTPSTAELRNLPNPKIIVIG
SEQ_ID_NO_4	1 CLT--APPKQAAKPTLMPRAQSYKDLTHLPAPTGKIFVS

保守性



SEQ_ID_NO_2	130 YESNVKSGGVGARYFGIGADTOYQLDQIAVNLRVVNVST
SEQ_ID_NO_5	130 YESNVKSGGVGARYFGIGADTOYQLDQIAVNLRVVNVST
SEQ_ID_NO_6	130 YESNVKSGGAGARYFGIGADTOYQLDQIAVNLRVVNVST
SEQ_ID_NO_7	130 YESNVKSGGVGARYFGIGADTOYQLDQIAVNLRVVNVST
SEQ_ID_NO_27	130 YESNVKSGGVGARYFGIGGDTOYQLDQIAVNLRVVNVST
SEQ_ID_NO_28	130 YESNVKSGGVGARYFGIGASTOYQLDQIAVNLRVVDVNT
SEQ_ID_NO_29	130 YESNVKSGGAGARYFGIGASTOYQLDQIAVNLRVVNVST
SEQ_ID_NO_30	130 YESNVKSGGAGARFFGIGASTOYQLDQIAVNLRVVDVNT
SEQ_ID_NO_32	129 YDSDIKTGGAGARYFGIGADGKYRVDQVAVNLRVVDVRT
SEQ_ID_NO_36	132 YDSNVRTGGAGAKYFGIGASGEYRVDQVTVNLRVVDIRS
SEQ_ID_NO_3	116 YESNVKSGGVGARYFGIGADTOYQLDQIAVNLRVVNVST
SEQ_ID_NO_35	125 YDTNIQTGGAGARYLGVGASGQYRTDOVTVNIRAVDVRT
SEQ_ID_NO_31	130 YASNVTGGFGARYFGIGGSTOYQLDQVAVNLRIVNVHT
SEQ_ID_NO_40	117 YDSNIHTGGAGARYFGIGASEKYRVDEVTVNLRVIDIRT
SEQ_ID_NO_33	125 YDSNIKTGGAGARYLGIGGSGQYRADQVTVNIRAVDVRS
SEQ_ID_NO_34	125 YDSNIKTGGAGARYLGIGGSGQYRADQVTVNIRAVDVRS
SEQ_ID_NO_37	115 YDTNVRTGGAGARYLGIGAATQFRVDTVTVNLRVVDIRT
SEQ_ID_NO_39	110 YDTNIKTGGAGARYLGIGVNSKFRVDTVTVNLRVVDIRT
SEQ_ID_NO_38	124 YDTNVRSGGEGARYLGIDISREYRVDDQVTVNLRVVDVRT
SEQ_ID_NO_41	127 YDSNTMTGGFGARYFGIGASTQYRQDRITIIYLRAVSTLN
SEQ_ID_NO_4	130 YESNVKSGGVGARYFGIGADTOYQLDQIAVNLRVVNVST

保守性



图45

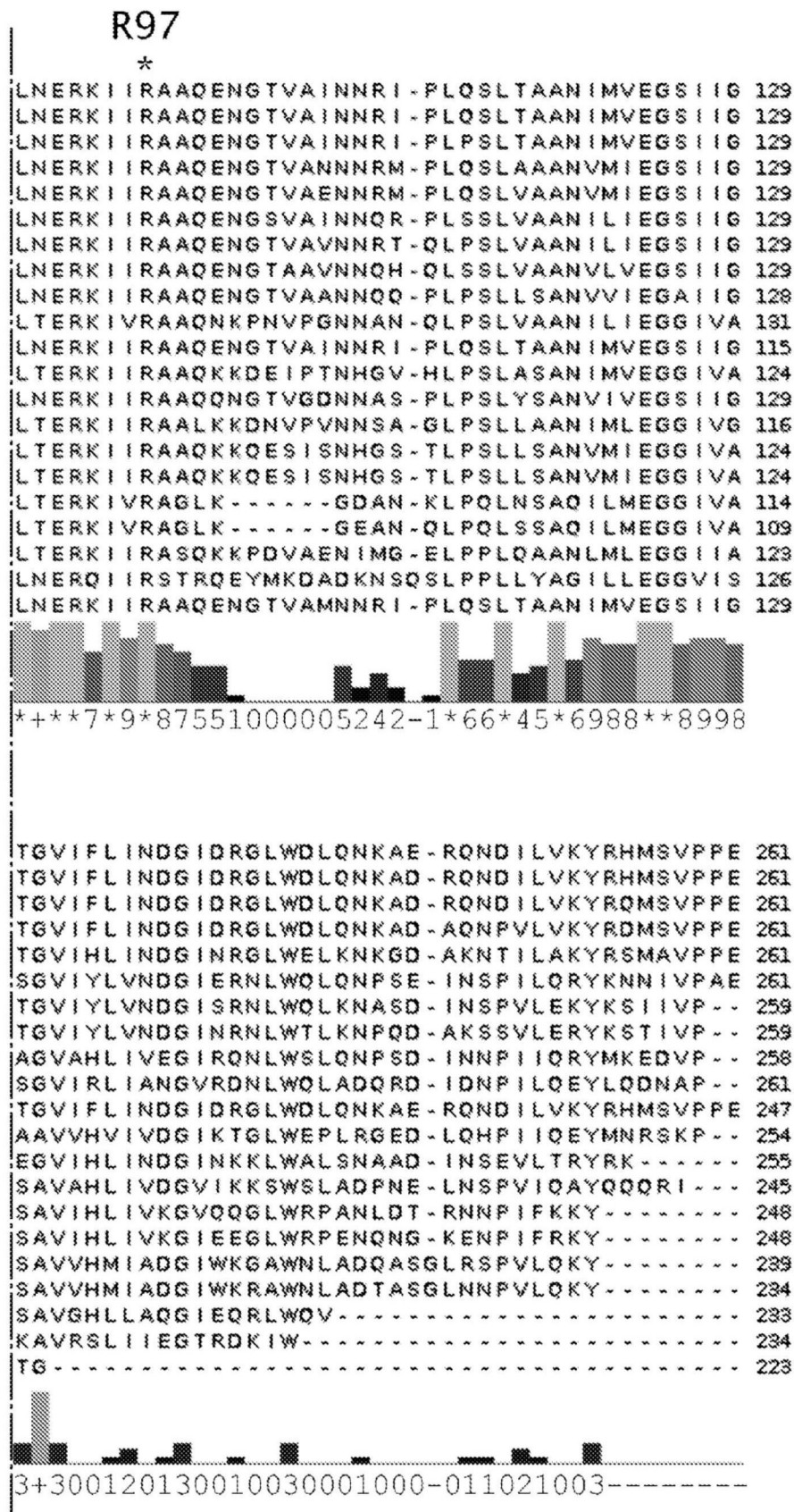


图45(接续II)

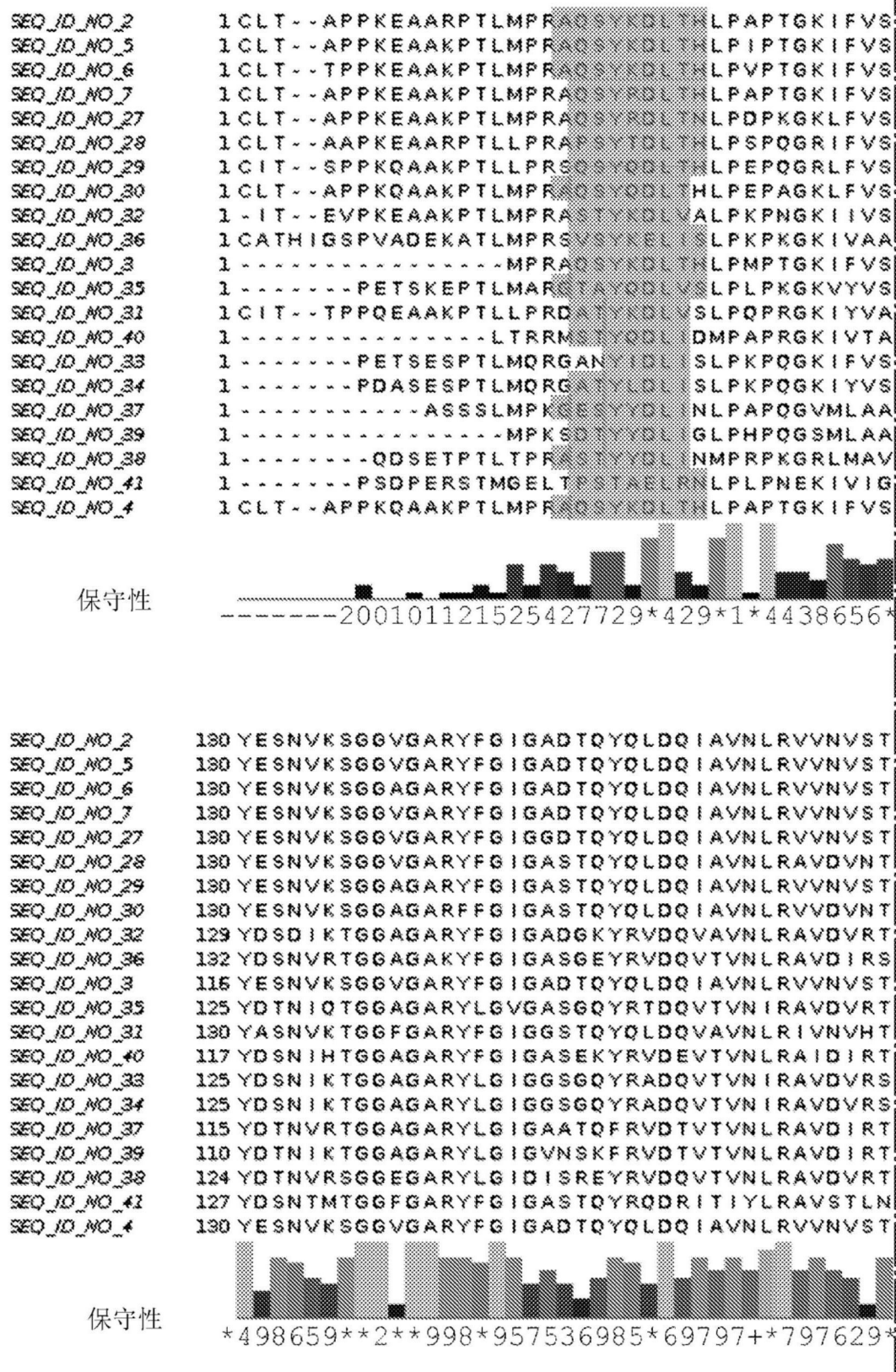


图46

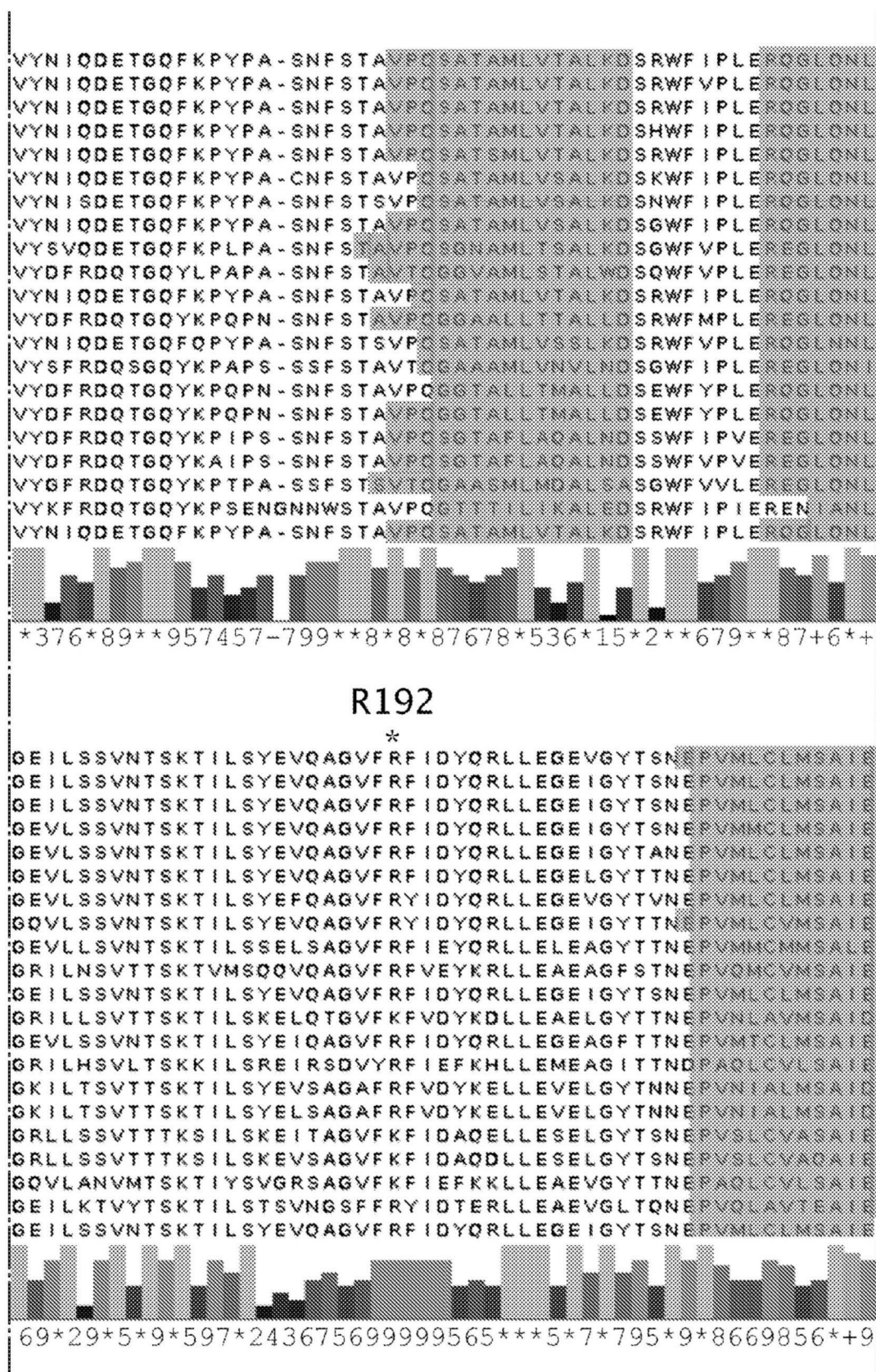


图46(接续I)

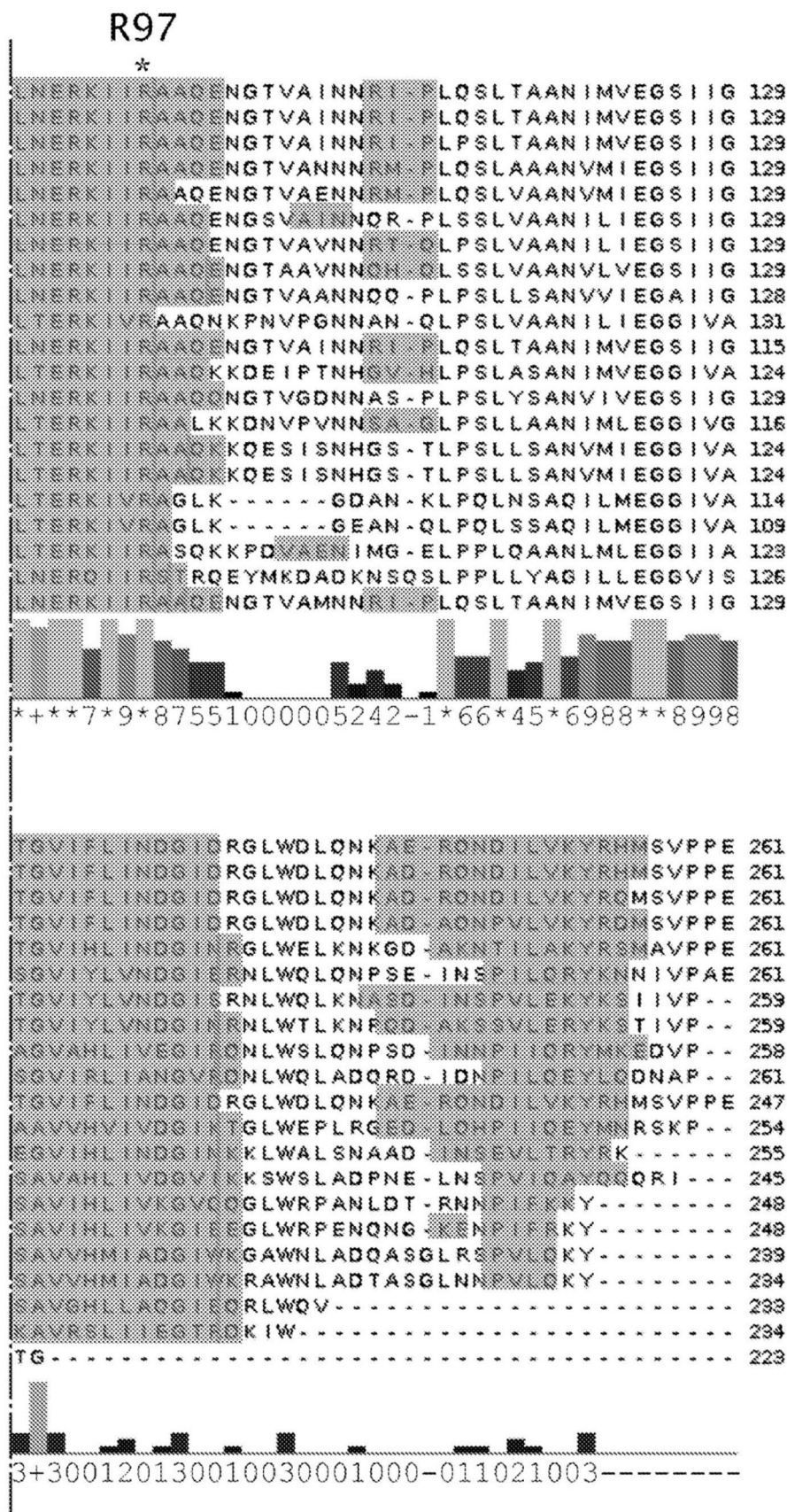


图46(接续II)

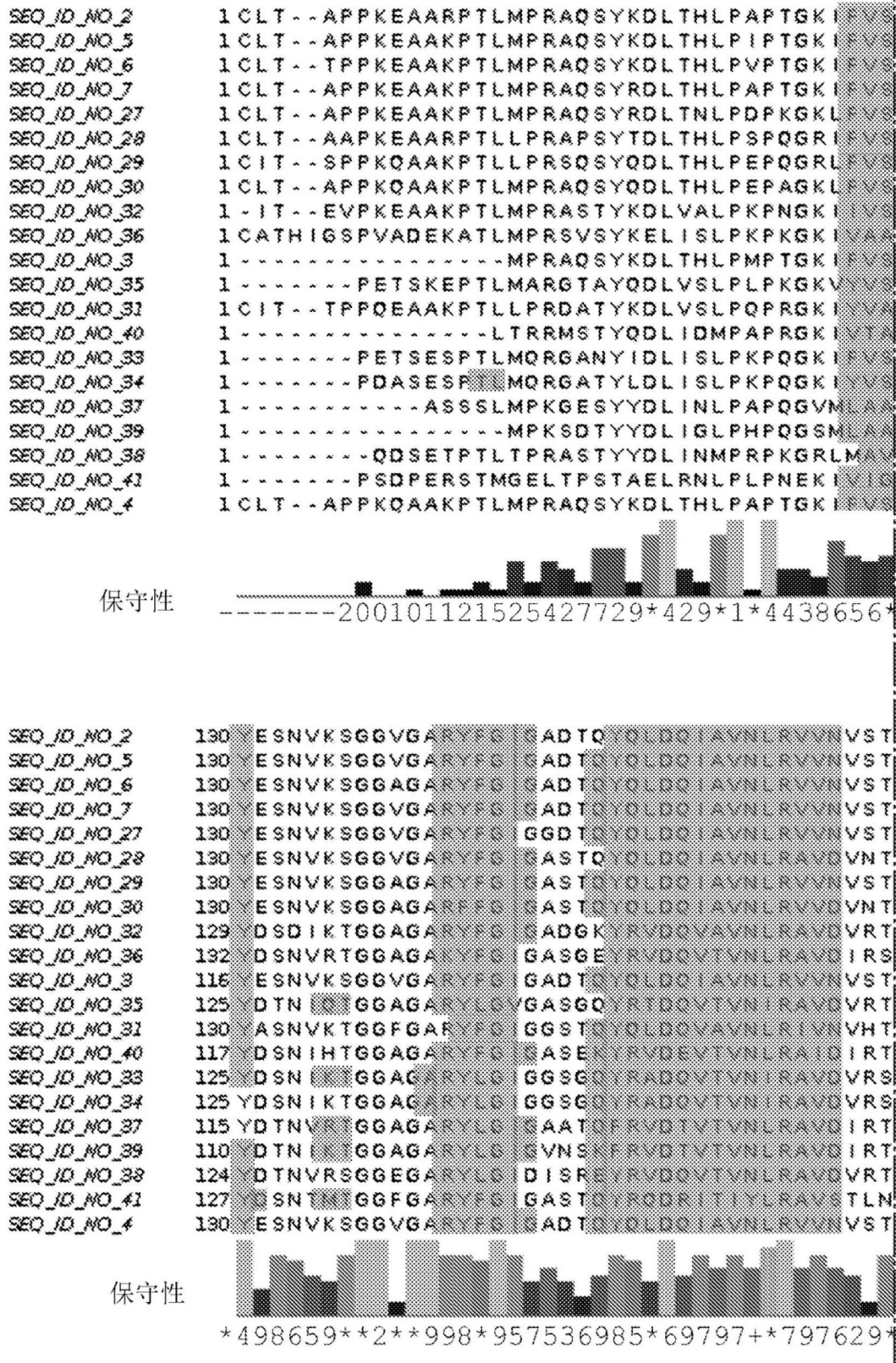


图47

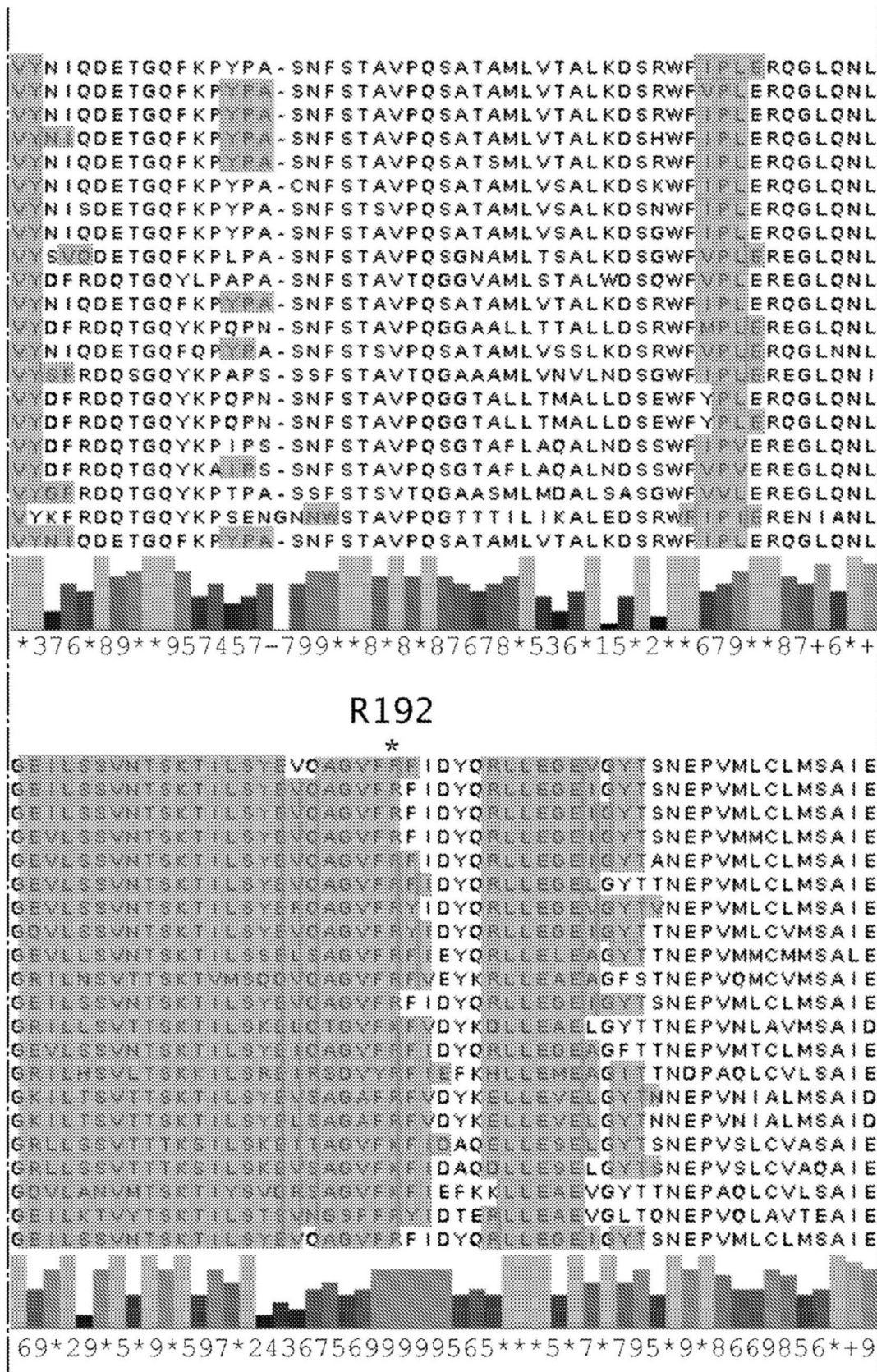


图47(接续I)

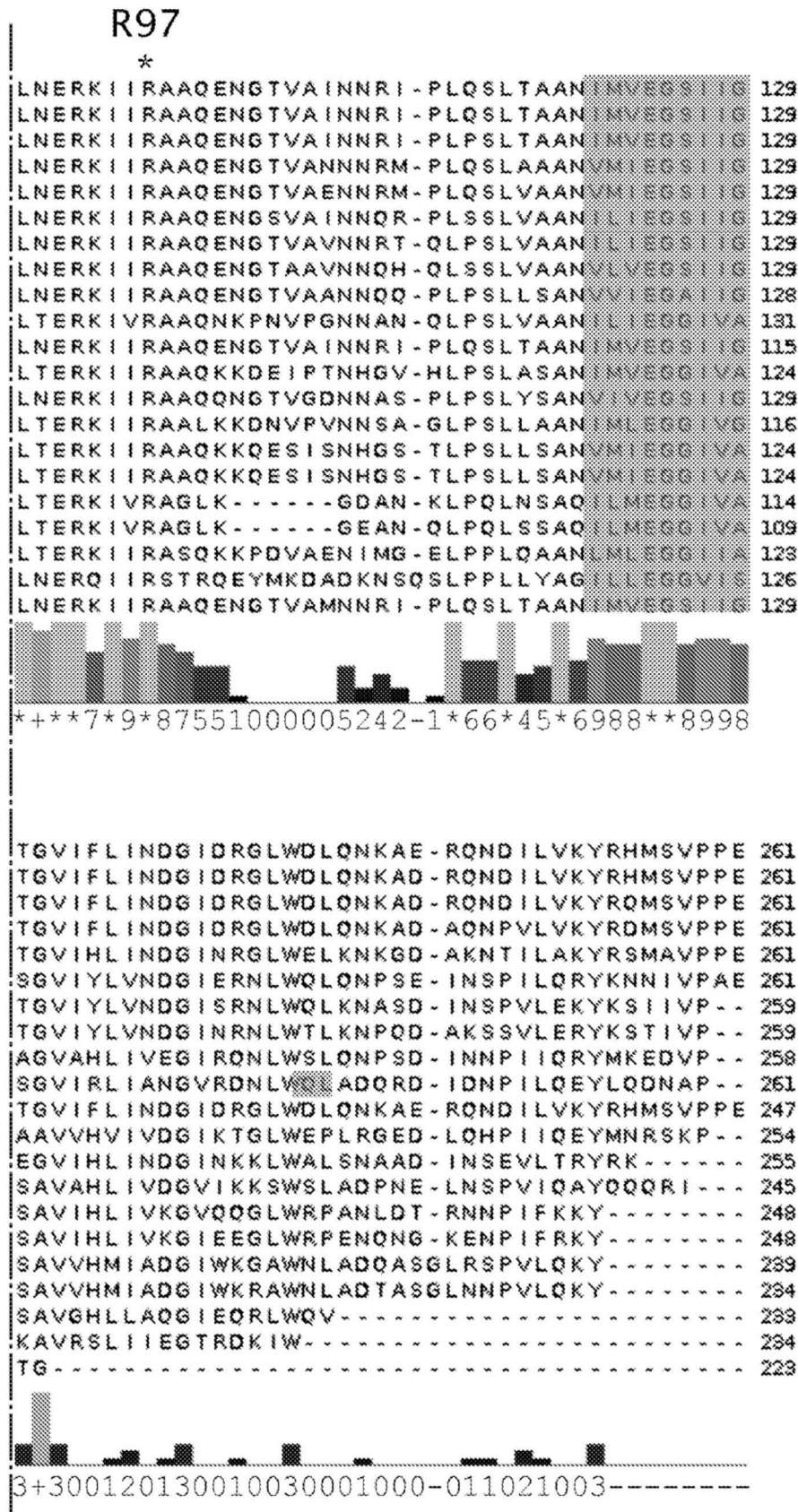


图47(接续II)

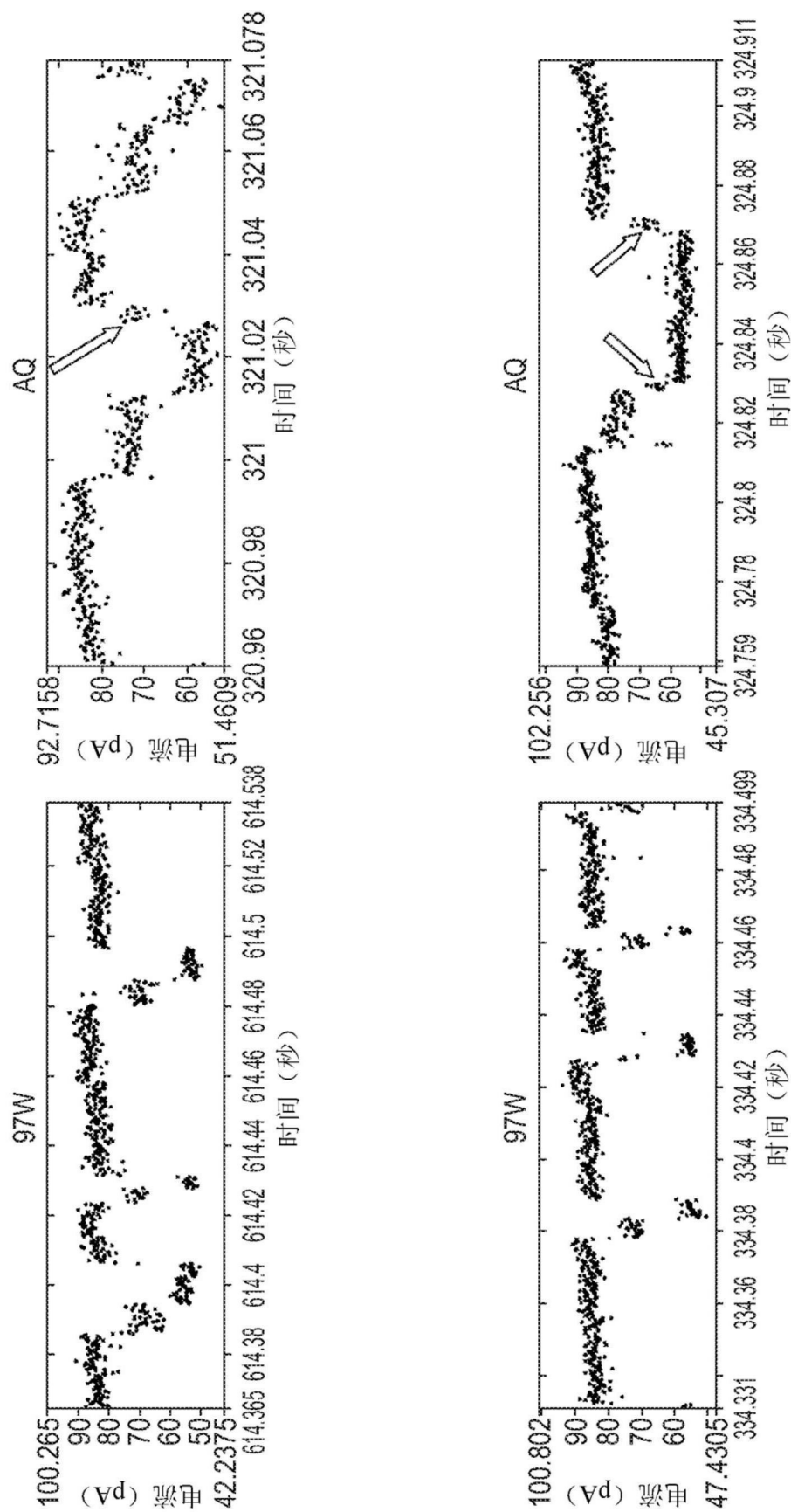


图48