(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119899249 A
(43)申请公布日 2025.04.29

(21)申请号 202510056986.2

(22)申请日 2015.09.01

(30)优先权数据
1415455.3 2014.09.01 GB
1422079.2 2014.12.11 GB
1506489.2 2015.04.16 GB
1506754.9 2015.04.21 GB
1508287.8 2015.05.14 GB
1511203.0 2015.06.25 GB
1515240.8 2015.08.27 GB

(62)分案原申请数据
201580059505.8 2015.09.01

(71)申请人 弗拉芒区生物技术研究所
地址 比利时根特
申请人 布鲁塞尔自由大学
牛津纳米孔科技公开有限公司

(72)发明人 汉·瑞曼特 拉科马·贾亚辛格
斯蒂芬·霍沃卡
伊丽莎白·杰恩·华莱士
詹姆斯·安东尼·克拉克
理查德·乔治·汉布利
乔纳森·班克兹·皮尤

(74)专利代理机构 北京天昊联合知识产权代理有限公司 11112
专利代理师 麦善勇 张天舒

(51)Int.Cl.
C07K 14/245(2006.01)
C12Q 1/6869(2018.01)
C12Q 1/6874(2018.01)
G01N 27/447(2006.01)

权利要求书6页 说明书132页
序列表（电子公布） 附图47页

(54)发明名称

突变CSGG孔

(57)摘要

本发明涉及位于外模(outer membrane-located)的脂蛋白CsgG的突变形式,特别是在Tyr51、Asn55和Phe56中一个或多个位置处的改性。本发明还涉及使用所述突变CsgG进行的分析物检测和表征。

CN 119899249 A

1.一种用于确定目标分析物的存在、不存在或一个或多个特征的方法,所述目标分析物为寡核苷酸或多核苷酸,所述方法包括:

a.使所述目标分析物与绝缘膜内的CsgG孔或其突变体接触,使得所述目标分析物在所述孔内或所述孔附近结合,或者通过所述孔移动的步骤,其中所述突变体具有与SEQ ID NO:390的氨基酸序列具有至少90%的同一性的氨基酸序列,且维持形成孔的能力;以及

b.在所述分析物在所述孔内或所述孔附近结合、或者通过所述孔移动时,进行一次或多次电学测量,由此确定所述分析物的存在、不存在或一个或多个特征的步骤,

所确定的所述特征为(i)分析物的长度、(ii)分析物的同一性、以及(iii)分析物的序列中的至少一个。

2.根据权利要求1所述的方法,其中,所述目标分析物为目标多核苷酸,步骤a)还包括使所述多核苷酸与多核苷酸结合蛋白接触,使得所述蛋白控制所述多核苷酸通过所述孔的移动,任选地,所述多核苷酸结合蛋白为解旋酶或衍生自解旋酶。

3.根据权利要求1或2所述的方法,其中,所述CsgG突变体包括:

(i)包含SEQ ID NO:390所示序列变体的突变体CsgG单体,其中所述变体包含在Y51,N55和F56中的一个或多个位置处的突变;或者

(ii)九个单体,并且所述单体中的至少一个为SEQ ID NO 390的变体,所述变体包括:(a)以下位置中的一个或多个处的突变:N40,Q42,D43,E44,K49,Y51,S54,N55,F56,S57,Q62,E101,N102,E124,E131,R142,D149,T150,E185,R192,D195,E201和E203;和/或(b)以下位置中的一个或多个的缺失:F48,K49,P50,Y51,P52,A53,S54,N55,F56和S57。

4.根据权利要求3(ii)所述的方法,其中,所述变体包括以下中的一个或多个:(i)以下位置处的一个或多个突变:N40,D43,E44,S54,S57,Q62,R97,E101,E124,E131,R142,T150和R192;(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)以下位置中的一个或多个的缺失:F48,K49,P50,Y51,P52,A53,S54,N55,F56和S57,任选地,

(a)所述变体包括在(i)一个或多个以下取代中:N40R,N40K,D43N,D43Q,D43R,D43K,E44N,E44Q,E44R,E44K,S54P,S57P,Q62R,Q62K,R97N,R97G,R97L,E101 N,E101Q,E101R,E101K,E101F,E101Y,E101W,E124N,E124Q,E124R,E124K,E124F,E124Y,E124W,E131D,R142E,R142N,T150I,R192E和R192N;

(b)所述变体包括在下述(ii)中:(ii)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/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/Y51S,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/Y51S,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/Y51S,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/Y51S,F56S/N55A/Y51G,F56S/N55T/Y51L,F56S/N55T/Y51V,F56S/N55T/Y51A,F56S/N55T/Y51N,F56S/N55T/Y51Q,F56S/N55T/Y51S,F56S/N55TY51G,F56G/N55QY51L,F56G/N55QY51V,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/Y51AF56G/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/Y51S,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/Y51S,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;和/或

(c)所述变体包括在(xi)中:(xi)缺失Y51/P52,缺失Y51/P52/A53,缺失P50至P52,缺失P50至A53,缺失K49至Y51,缺失K49至A53并用单个脯氨酸(P)替代,缺失K49至S54并用单个P替代,缺失Y51至A53,缺失Y51至S54,缺失N55/F56,缺失N55至S57,缺失N55/F56并用单个P替代,缺失N55/F56并用单个甘氨酸(G)替代,缺失N55/F56并用单个丙氨酸(A)替代,缺失N55/F56并用单个P和Y51N替代,缺失N55/F56并用单个P和Y51Q替代,缺失N55/F56并用单个P和Y51S替代、缺失N55/F56并用单个G和Y51N替代,缺失N55/F56并用单个G和Y51Q替代,缺失N55/F56并用单个G和Y51S替代,缺失N55/F56并用单个A和Y51N替代,缺失N55/F56并用单个A/Y51Q替代,或缺失N55/F56并用单个A和Y51 S替代。

5.根据权利要求4所述的方法,其中,所述变体包括: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/E201

N/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/E203N,D149Q/E185N/E203Q,

D149N/E185Q/E203Q,D149Q/E185Q/E203Q,D149N/E185N/E201N/E203N,

D149Q/E185N/E201N/E203N,D149N/E185Q/E201N/E203N,D149N/E185N/E201Q/E203N,

D149N/E185N/E201N/E203Q,D149Q/E185Q/E201N/E203N,D149Q/E185N/E201Q/E203N,

D149Q/E185N/E201N/E203Q,D149N/E185Q/E201Q/E203N,D149N/E185Q/E201N/E203Q,

D149N/E185N/E201Q/E203Q,D149Q/E185Q/E201Q/E203Q,D149Q/E185Q/E201N/E203Q,

D149Q/E185N/E201Q/E203Q,D149N/E185Q/E201Q/E203Q,D149Q/E185Q/E201Q/E203N,

D149N/E185N/D195N/E201N/E203N,D149Q/E185N/D195N/E201N/E203N,

D149N/E185Q/D195N/E201N/E203N,D149N/E185N/D195Q/E201N/E203N,D149N/E185N/D195N/E201Q/E203N,D149N/E185N/D195N/E201N/E203Q,

D149Q/E185Q/D195N/E201N/E203N,D149Q/E185N/D195Q/E201N/E203N,

D149Q/E185N/D195N/E201Q/E203N,D149Q/E185N/D195N/E201N/E203Q,

D149N/E185Q/D195Q/E201N/E203N,D149N/E185Q/D195N/E201Q/E203N,
D149N/E185Q/D195N/E201N/E203Q,D149N/E185N/D195Q/E201Q/E203N,
D149N/E185N/D195Q/E201N/E203Q,D149N/E185N/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/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/E201Q/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/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/D195N/E201Q/E203N,D149Q/E185R/D195N/E201N/E203Q,
D149N/E185R/D195Q/E201Q/E203N,D149N/E185R/D195Q/E201N/E203Q,
D149N/E185R/D195N/E201Q/E203Q,D149Q/E185R/D195Q/E201Q/E203N,
D149Q/E185R/D195Q/E201N/E203Q,D149Q/E185R/D195N/E201Q/E203Q,
D149N/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/E203N,D149Q/E185R/D195N/E201R/E203Q,D149N/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。

6.根据权利要求4所述的方法,其中,所述变体包括在T150处的突变。

7.根据权利要求3(ii)所述的方法,其中,所述变体包括以下取代中的一个或多个:(i)F56N,F56Q,F56R,F56S,F56G,F56A或F56K或F56A,F56P,F56R,F56H,F56S,F56Q,F56I,F56L,F56T或F56G;(ii)N55Q,N55R,N55K,N55S,N55G,N55A或N55T;(iii)Y51L,Y51V,Y51A,Y51N,Y51Q,Y51S或Y51G;(iv)T150I;(v)S54P;和(vi)S57P。

8.一种用于在权利要求1至7中任一项所述的方法中使用的传感器,其用于表征目标多核苷酸,所述传感器包括CsgG孔或其突变体与多核苷酸结合蛋白之间的复合物。

9.根据权利要求8所述的传感器,其中,所述CsgG突变体如权利要求3至7中任一项所定义。

10.一种用于在权利要求1至7中任一项所述的方法中使用的试剂盒,其用于表征目标分析物,所述试剂盒包括:(a)CsgG孔或其突变体,和(b)膜的组分。

11.根据权利要求10所述的试剂盒,其中,所述CsgG突变体如权利要求3至7中任一项所定义。

12.一种用于在权利要求1至7中任一项所述的方法中使用的组合物,包含突变体CsgG单体,所述CsgG单体包含SEQ ID NO 390所示的序列的变体,

(a)所述变体包含以下中的突变:

(i)Y51,这里突变被F,M,L,I,V,A,P,G,C,N,T,S,E,D,K,H或R取代,

(ii)N55,这里突变被F,M,L,I,V,P,G,C,Q,T,E,D,K,H或R取代,

(iii)F56,

(iv)Y51和N55,这里突变不是Y51Q和N55S,

(v)Y51和F56,

(vi)N55和F56,或

(vii)Y51、N55和F56,或者

(b)所述变体为Y51Q和N55S以外的权利要求4或5中所定义的变体。

13.一种用于在权利要求1至7中任一项所述的方法中使用的组合物,包含含有两个或更多个共价连接的CsgG单体的构建体,其中,所述单体中的至少一个是权利要求12中所限定的突变体单体。

14.一种用于在权利要求1至7中任一项所述的方法中使用的组合物,包含含有至少一个CsgG单体的修饰的CsgG孔,其中,所述单体包含在SEQ ID NO 390所示的序列的38至63位之间的修饰。

15.根据权利要求14所述的修饰CsgG组合物,其中,

(a)所述孔包含权利要求12(a)中所限定的单体,或者

(b)所述单体具有SEQ ID NO:4至388的从位置38至63的多肽序列。

16.一种用于在权利要求1至7中任一项所述的方法中使用的组合物,包含权利要求12中所限定的相同的突变体单体或权利要求13中所限定的相同的构建体。

17.一种用于在权利要求1至7中任一项所述的方法中使用的组合物,包含对权利要求12中所限定的突变体CsgG单体进行编码的多核苷酸。

突变CSGG孔

[0001] 本专利申请是申请号为201580059505.8、申请日为2015年9月1日、发明名称为“突变CSGG孔”的专利申请的分案申请。

技术领域

[0002] 本发明涉及一种新型蛋白孔及其用途。特别地,其涉及在核酸测序应用及分子感测中的生物纳米孔。

[0003] 本发明涉及CsgG的突变形式。本发明还涉及采用CsgG的分析物检测和表征。

背景技术

[0004] 蛋白孔是形成膜中通道的跨膜多肽和复合物,离子和某些分子可以通过该通道。所述通道的最小直径典型的处在纳米范围内(10^{-9}米),因此将这些多肽中的某些命名为“纳米孔”。

[0005] 纳米孔具有作为生物传感器的巨大潜力。当结合在膜上的纳米孔被施加电势时,离子从所述通道中流过。该离子流可以作为电流被测量。例如,WO2000/28312和D.Stoddart et al.,Proc.Natl.Acad.Sci.,2010,106,7702-7中描述了使用单通道记录设备的合适的电学检测技术。例如,WO2009/077734中描述了多通道记录技术。

[0006] 通过所述孔,或结合到所述孔中或接近所述孔的分子转录用于阻碍并从而降低通过所述通道的离子流。离子流的降低程度指示孔内或其附近的阻塞的尺寸,该离子流降低程度通过电流的下降进行测量。因此被测量的电流可以用于测量通道阻塞的大小或程度。电流的变化可以用于识别分子或分子的一部分已经在所述孔(分子传感)或接近所述孔或在特定体系中结合,其可以用于基于分子大小测定位于孔内分子的身份(核酸测序)。

[0007] “链测序”方法是已知的采用生物纳米孔对核酸进行测序的方法。将单链多核苷酸通过纳米孔,当它们瞬时穿过纳米孔通道时,每个核苷酸上的碱基通过检测电流的变化进行测定。该方法相对于核酸测序的传统方法节省了大量时间和成本。

[0008] 之前报道的蛋白纳米孔,例如突变体MspA(Manrao et al.,Nature Biotechnology,2012,30(4),349-353)以及α-溶血素纳米孔(Nat.Nanotechnol.,2009,4(4),265-70)已经用于使用“链测序”方法的核酸测序。相似地,对于蛋白传感其他孔例如α-溶血素(J Am Chem Soc,2012,134(5),2781-7)和ClyA(Am.Chem.Soc.Nano.2014,8(12),12826-35)(J.Am.Chem.Soc,2013,135(36),13456-63)也已经被采用。

[0009] 对克服现有技术缺陷的新型纳米孔的需求依然存在,尤其是优化用于分子传感应用和例如核酸测序应用的孔的维度和特征。

[0010] 纳米孔传感是一种依靠观察个体结合或分析物分子和受体间相互作用的传感方法。纳米孔传感器可以通过在绝缘膜中放置纳米维度的单孔并测量在分析物分子存在下通过该孔的电压驱动的离子传输而产生。分析物的识别通过其独特的电流信号、显著的持续时间和电流块的范围以及电流水平的变化显示出来。

[0011] 现在存在对快速且便宜、应用范围广泛的核酸(例如DNA或RNA)测序技术的需求。

现有技术速度慢并且价格昂贵主要是因为其依赖扩增技术产生大量核酸并需要大量的用于信号检测的专用荧光化学物质。纳米孔传感有潜力通过减少核苷酸和所需试剂用量而提供一种快速且便宜的核酸测序方法。

[0012] 采用纳米孔传感进行核酸测序的两个必要组分是(1)控制核酸移动通过该孔;和(2)核酸聚合物移动通过孔时对核苷酸的识别。在过去,为实现核苷酸的识别,使核酸穿过溶血素的突变体。这会提供电流信号,该电流信号显示是由序列决定的。这也表明当使用溶血素孔时,大量的核苷酸有助于观测电流,使得观测电流和多核苷酸之间的直接关系令人深思。

[0013] 虽然通过溶血素孔的突变已经改善了用于识别核苷酸的电流范围,但是如果核苷酸之间的电流差异进一步提高,则测序系统将具有更好的性能。另外,观察到,当核酸移动通过孔时,某些电流状态表现出高变化性。还显示,某些突变溶血素孔比其他的展现出更高的变化性。虽然这些状态的变化可以包含序列的特异性信息,但是希望产生具有低变化性的孔来简化该系统。还期望减少有助于观测到的电流的核苷酸的数量。

发明内容

[0014] 发明人已经鉴定出细菌淀粉样蛋白分泌通道CsgG的结构。所述CsgG通道是一种跨膜寡聚蛋白,其形成最小直径约为0.9nm的通道。CsgG纳米孔的结构使其适合用于蛋白质传感应用中,特别是用于核酸测序。CsgG多肽的修饰变体可以用于进一步增强所述通道对上述特定应用的适用性。

[0015] 因为其结构更适合于DNA测序应用,CsgG孔较现有蛋白质孔例如ClyA或α-溶血素具有优势。CsgG孔具有更有利的纵横比,包括比ClyA具有更短的跨膜通道。CsgG孔与α-溶血素孔相比具有更宽的通道开口。对于某些应用,这可以促进酶的附着,例如对于核酸测序应用。在这些实施方式中,其也可以使位于酶和读取头(定义为孔的最窄部分)之间的核酸链段的长度最小化,从而导致改善的读出信号。在涉及核酸测序的本发明的实施方式中,CsgG孔的通道中狭窄的内部收缩结构(constriction)还促进单链DNA的易位。所述收缩结构由两个环形环组成,这两个环形环通过相邻蛋白质单体中位置51处的酪氨酸残基(Tyr 51)以及分别在位置56和55处的苯丙氨酸残基和天冬酰胺残基(Phe 56和Asn 55)并置而形成。所述收缩结构的尺寸可以被修改。ClyA具有更宽的内部收缩结构,使得目前没有用于测序的双链DNA通过。所述α-溶血素孔不仅具有一个1.3nm宽的内部收缩结构,同时还具有一个2nm宽的具有额外读取头特征的β-桶结构。

[0016] 在第一方面,本发明涉及一种用于分子传感的方法,其包括:

[0017] a)提供一种由绝缘层内的至少一种CsgG单体形成的CsgG生物孔;

[0018] b)在所述绝缘层上施加电势,从而建立通过所述生物孔的电流;

[0019] c)使所述CsgG生物孔与测试底物接触;以及

[0020] d)测量流过所述生物孔的电流。

[0021] 典型地,所述绝缘层为一种膜,例如磷脂双分子层。在一种实施方式中,通过所述孔的电流是由从所述绝缘层的第一侧流到所述绝缘层的第二侧的可溶性离子流携带。

[0022] 在本发明的一种实施方式中,所述分子传感为分析物检测。在特定实施方式中,分析物检测的方法包括在步骤(d)之后的另外的步骤:通过与测试底物不存在时通过所述生

物孔的电流相比,通过所述生物孔的电流减少,确定出测试底物的存在。

[0023] 在本发明一种替代的实施方式中,所述分子传感是核酸测序。典型地,通过所述方法测序的核酸种类为DNA或RNA。在本发明特定的实施方式中,所述CsgG生物孔适合于容纳额外的辅助蛋白。典型地,所述额外的辅助蛋白为选自以下的核酸加工酶:DNA或RNA聚合酶;异构酶;拓扑异构酶;促旋酶;端粒酶;核酸外切酶;和解旋酶。

[0024] 在本发明的实施方式中,所述CsgG生物孔为一种修饰的CsgG孔,其中所述修饰的CsgG孔对所述CsgG孔的至少一个CsgG单体中的单体野生型大肠杆菌CsgG多肽序列至少具有一种修饰。典型地,对形成所述CsgG孔的所有CsgG单体进行相同的修饰。在本发明具体的实施方式中,所述修饰的CsgG单体具有根据SEQ ID NO:4至388的从位置38至63的多肽序列。

[0025] 在第二方面,本发明涉及修饰的CsgG生物孔,包括至少一个CsgG单体,其中所述修饰的CsgG生物孔具有不多于一个的通道收缩结构,所述收缩结构的直径范围为0.5nm至1.5nm。典型地,所述修饰在所述CsgG单体多肽序列的位置38至63之间。合适的,所述修饰处于选自以下的位置:Tyr51;Asn55;和Phe 56。在具体的实施方式中,所述修饰为在Tyr 51位置,或在Asn55和Phe56两个位置。

[0026] 在本发明的实施方式中,对所述CsgG单体的修饰选自以下:自然存在的氨基酸的取代;自然存在的氨基酸的缺失;和自然存在的氨基酸侧链的修饰。合适地,所述修饰降低或除去未修饰氨基酸的空间阻碍。在具体的实施方式中,所述孔的至少一个CsgG单体具有根据SEQ ID NO 4至388的从位置38至63的多肽序列。

[0027] 在第三方面,本发明涉及分离的多肽,所述多肽编码本发明第二方面的修饰的CsgG生物孔的至少一种CsgG单体。

[0028] 在第四方面,本发明涉及分离的核酸,所述核酸编码本发明第三方面的分离的多肽。

[0029] 在第五方面,本发明涉及一种生物传感器,包括:

[0030] a)绝缘层;

[0031] b)在所述绝缘层中的CsgG生物孔;以及

[0032] c)用于测量通过所述生物孔电流的设备。

[0033] 在具体实施方式中,在所述生物传感器中的所述CsgG生物孔为根据本发明第二方面所述的修饰的CsgG生物孔。

[0034] 在第六方面,本发明涉及用于生物传感应用的CsgG生物孔的用途,其中所述生物感应应用为分析物检测或核酸测序。

[0035] 本发明第六方面的一种实施方式中,所述核酸测序为DNA测序或RNA测序。

[0036] 发明人意外地证明,CsgG及其新突变体可用于表征分析物,例如多核苷酸。本发明涉及突变CsgG单体,其中进行一个或多个修饰以提高所述单体与分析物,例如多核苷酸,相互作用的能力。发明人还意外的证明,包含新型突变单体的孔具有增强的与分析物,例如多核苷酸,相互作用的能力,并因此表现出用于评估分析物特性(例如多核苷酸序列)的改进的性能。所述突变孔意外的显示出改善的对核苷酸的鉴别。特别地,突变孔意外的显示出增加的电流范围,这使其更容易区分不同的核苷酸,并减少会增加信噪比的状态变化。另外,多核苷酸移动通过所述孔时促成电流的核苷酸数量降低。这使得确定多核苷酸移动通过孔

时观察到的电流与多核苷酸之间的直接关系更加容易。另外,所述突变孔可以显示出增加的通量,例如,更可能与分析物,例如多核苷酸,相互作用。这使得采用所述孔表征分析物更加容易。所述突变孔可以更容易地插入膜。

[0037] 因此,本发明提供一种包含SEQ ID NO:390所示序列变体的突变型CsgG单体,其中所述变体包含在Y51,N55和F56中的一个或多个位置处的突变。

[0038] 因此,本发明提供一种包含SEQ ID NO:390所示序列变体的突变型CsgG单体,其中所述变体包括以下的一种或多种:(i)在以下位置的一个或多个突变(即在一个或多个以下位置的突变)N40,D43,E44,S54,S57,Q62,R97,E101,E124,E131,R142,T150和R192;(ii)在Y51/N55,Y51/F56,N55/F56或Y51/N55/F56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)一个或多个以下位置的缺失:F48,K49,P50,Y51,P52,A53,S54,N55,F56和S57。

[0039] 本发明还提供:

[0040] -包含两个或多个共价连接的CsgG单体的构建体,其中至少一个单体是本发明所述的突变型单体;

[0041] -编码本发明突变型单体或本发明构建体的多核苷酸;

[0042] -一种衍生自包含本发明的相同突变型单体或本发明的相同构建体的CsgG的同源寡聚孔;

[0043] -一种衍生自包含至少一个本发明突变型单体或至少一个本发明构建体的CsgG的异源寡聚孔;

[0044] -用于确定目标分析物的存在、不存在或一种或多种特性的方法,包括:

[0045] a)将目标分析物与CsgG孔或其突变体接触,以使所述目标分析物相对于所述孔移动;和

[0046] b)所述分析物相对于所述孔移动时获取一个或多个测量值并因此确定所述分析物的存在、不存在或其一种或多种特性;

[0047] -一种形成用于表征目标多核苷酸的传感器的方法,包括形成CsgG孔或其突变体和多核苷酸结合蛋白之间的复合物,并从而形成用于表征目标多核苷酸的传感器;

[0048] -用于表征目标多核苷酸的传感器,包括CsgG孔或其突变体和多核苷酸结合蛋白之间的复合物;

[0049] -CsgG孔或其突变体在确定目标分析物存在、不存在或一个或多个特性中的应用;

[0050] -一种用于表征目标分析物的试剂盒,包括(a)Csg孔或其突变体,以及

[0051] (b)膜的组分;

[0052] -一种用于表征样品中目标分析物的设备,包括(a)多个CsgG孔或其突变体和(b)多个膜;

[0053] -表征目标多核苷酸的方法,包括:

[0054] a)使所述多核苷酸与CsgG孔或其突变体、聚合酶和标记的核苷接触,以使磷酸标记物(phosphate labelled specie)通过聚合酶顺序添加到目标多核苷酸,其中所述磷酸标记物含有对每个核苷酸特异性的标记;以及

[0055] b)采用所述孔检测所述磷酸标记物并从而表征所述多核苷酸;和

[0056] 生产本发明所述突变单体或本发明所述构建体的方法,包括在合适的宿主细胞中表达本发明所述的多核苷酸并从而产生本发明所述的突变单体或构建体。

附图说明

[0057] 图1表示带面表示法(ribbon and surface representation)中CsgG九聚物在其通道构象中的侧面剖视图。

[0058] 图2表示CsgG通道收缩结构(即在纳米孔感测应用环境下的孔读取头)的截面图以及相关的直径测量。

[0059] 图3表示形成孔收缩结构的结构基序,包括三个堆叠的同心侧链层:Tyr 51,Asn 55和Phe 56。

[0060] 图4表示CsgG同源物中的序列同源性,包括CsgG样蛋白的多重序列比对(SEQ ID NO:442至SEQ ID NO:448)。所选序列选自跨CsgG样序列系统发生树的单系分支(monophyletic clade)(未显示),以给出序列多样性的代表性视图。二级结构元素显示为,对于β-链和α-螺旋分别用箭头或棒状物表示,并且所述二级结构元素基于大肠杆菌CsgG晶体结构。重要地,相当于大肠杆菌Tyr 51,Asn55和Phe 56的残基由箭头突出显示。这些残基形成孔的内部收缩结构,即在前述纳米孔传感应用中的所述孔读取头。

[0061] 图5表示在平面磷脂双分子层中重组并在+50mV(n=33)或-50mV(n=13)的电场下测量的CsgG的代表性单通道电流记录(a)和电导直方图(b)。

[0062] 图6表示PPB重组的CsgG的单通道电流记录,该电流记录在+50mV或-50mV,并辅以CsgE浓度增量的条件下进行。水平刻度条位于0pA。

[0063] 图7表示a.C8E4/LDAO溶解的CsgG原始负染色EM照片。箭头表示在图g所示尺寸洗脱曲线中标记的不同颗粒群,其为(I)CsgG九聚体的聚集体,(II)CsgG十八聚体和(III)CsgG九聚体。比例尺(Scale bar),20nm。b,所指示低聚态顶视图和侧视图的类平均示意图,c,顶视图中LDAO溶解的CsgG的旋转自相关函数图,显示出九重对称性,d,$CsgG_{C1S}$的原始负染色EM照片-箭头表示图g中通过尺寸排阻色谱观察到的十六聚体(IV)和八聚体(V)颗粒。e,$CsgG_{C1S}$低聚物侧视图的类平均示意图。没有观察该构建体的俯视图,f,由图g中表示的尺寸排阻色谱观察到的$CsgG_{C1S}$和CsgG颗粒的洗脱体积(EV)、计算的分子质量(MWcalc)、对应于CsgG低聚状态($CsgG_n$)的预期分子质量(MW_{CsgG})以及通过负染色电镜(EM)和X射线晶体观察到的颗粒的对称性的表格。g,$CsgG_{C1S}$(黑色)和C8E4/LDAO-溶解的CsgG(灰色)在Superdex 200 10/300GL(GE Healthcare)上进行的尺寸排阻色谱图,h,i,结晶低聚物顶视图和侧视图的带状示意图,其显示$CsgG_{C1S}$(h)的D_8十六聚体和膜提取的CsgG(i)中的D_9十八聚体。一个原聚体从N端(蓝色)到C端(红色)用彩虹色进行着色。在晶体中捕获的形成尾接尾二聚体的两个C_8八聚体($CsgG_{C1S}$)或C_9九聚体(CsgG)被着色成蓝色和棕褐色。r和θ分别表示半径和原聚体间的转角(interprotomer rotation)。

[0064] 图8表示$CsgG_{C1S}$在2.8Å下的电子密度映射图(map),其采用NCS平均化和密度修改的实验SAD相计算,并在1.5g下轮廓在1.5g绘制等值线。映射图显示了通道构建的区域(CL;标记了单个原聚体),并且被覆盖在最终细化模型上。$CsgG_{C1S}$是一种突变CsgG,其中成熟CsgG序列的N端Cys,即Cys 1,被Ser替换,导致缺少通过大肠杆菌LOL途径的脂修饰。这将获得一种可溶的以前孔(pre-pore)构象存在的同源八聚低聚体(参见图42),所述构象与通过

天然的、酯修饰的CsgG形成的膜靶向同源九聚体孔相反(图43)。

[0065] 图9表示采用以C端到N端方向显示的伸展构象的聚丙氨酸链螺旋穿过所述通道模型化的CsgG收缩结构的顶视图(图9a)和侧视图(图9b)。聚丙氨基酸链的模拟溶剂化(modelled solvation),即图9b中的位置,如图9c所示,其中为清楚起见,C环结构被除去(显示的溶剂分子是距整个聚丙氨酸链10Å内的分子)。

[0066] 图10:表示大肠杆菌中的CsgG。

[0067] 图11:表示CsgG的尺寸。

[0068] 图12:说明在10 Å/ns下单个G的移位。在大肠杆菌CsgG(CsgG-Eco)中,鸟嘌呤进入F56环存在很大的障碍。*=G进入F56环。A=G停止与环56相互作用。B=G停止与环55相互作用。C=G停止与环51相互作用。

[0069] 图13:说明在100 Å/ns下ssDNA的移位。将DNA拉动通过大肠杆局CsgG的收缩结构需要很大的力。

[0070] 图14:说明在10 Å/ns下ssDNA的移位。CsgG-F56A-N55S和CsG-F56A-N55S-Y51A突变体均对ssDNA的移位具有较低的阻碍。

[0071] 图15-17:突变孔与野生型(WT)相比,表现出增加的范围。

[0072] 图18-19:突变孔与野生型(WT)相比,表现出增加的通量。

[0073] 图20-21:突变孔与野生型(WT)相比,表现出增强的插入。

[0074] 图22:表示用于实施例18中的DNA构建体X。标记为1的区域对应于30个SpC3间隔区。标记为2的区域对应于SEQ ID NO:415。标记为3的区域对应于4个iSp18间隔区。标记为4的区域对应于SEQ ID NO:416。标记为5的区段对应于4个5-硝基吲哚。标记为6的区域对应于SEQ ID NO:417。标记为7的区域对应于SEQ ID NO:418。标记为8的区域对应于SEQ ID NO:419,其具有4个iSp18间隔区(标记为9的区域)连接在SEQ ID NO:419的3'端。在iSp18间隔区的另一端是3'胆固醇系链(标记为10)。标记为11的区域对应于4个SpC3间隔区。

[0075] 图23:表示CsgG蛋白的链球菌性陷阱(Strep trap)(GE Healthcare)纯化的示例性色谱踪迹(x轴标记=洗脱体积(mL),Y轴标记=吸光度(mAu))。将样品加样在25mM Tris,150mM NaCl,2mM EDTA,0.01%DDM中,并用10mM脱硫生物素洗脱。将洗脱的CsgG蛋白中的洗脱峰标记为E1。

[0076] 图24:表示在初始strep纯化后,CsgG蛋白的典型SDS-PAGE可视化的示例。4-20%TGX凝胶(Bio Rad)在1X TGS缓冲液中300V下运行22分钟。所述凝胶用Sypro Ruby染色剂染色。泳道1-3表示如箭头所示的含有CsgG蛋白的主要洗脱峰(图23中标记为E1)。泳道4-6对应于含有污染物的主要洗脱峰(图23中标记为E1)的尾部的洗脱馏分。M表示所使用的分子量标记物,其为未染色的Novex Sharp(单位=kD)。

[0077] 图25:表示CsgG蛋白尺寸排阻色谱(SEC)的示例(120mL S200GE healthcare,x轴标记=洗脱体积(mL),Y轴标记=吸光度(mAu))。所述SEC在strep纯化和加热蛋白质样品后进行。SEC的电泳缓冲液为25mM三羟甲基氨基甲烷(Tris),150mM NaCl,2mM EDTA,0.01%DDM,0.1%SDS,pH 8.0,并且色谱柱以1mL/分钟的速率运行。标记为X的踪迹在220nm处有吸光度,标记为Y的踪迹在280nm处有吸光度。收集标记为星号的峰。

[0078] 图26:表示SEC后,CsgG蛋白的典型SDS-PAGE可视化的示例。4-20%TGX凝胶(Bio Rad)在1X TGS缓冲液中300V下运行22分钟,并且所述凝胶用Sypro Ruby染色剂染色。泳道1

表示strep纯化后并加热但在SEC之前的CsgG蛋白样品。泳道2-8表示图25中运行大约48mL-60mL(中间峰＝55mL)的峰上收集的馏分,并在图25中用星号标记。M表示所使用的分子量标记,其为未染色的Novex Sharp(单位＝kD)。对应于大肠杆菌CsgG孔的条带由箭头表示。

[0079] 图27至33:突变孔与野生型(WT)相比,表现出增加的范围。

[0080] 图34-39:突变孔与野生型(WT)相比,表现出增加的范围。

[0081] 图40表示模拟中在0和20ns取得的位于孔(CsgG-Eco-(Y51T/F56Q)-Strepll(C))9(具有突变Y51T/F56Q的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且在C端被连接到孔突变体No.20))顶部的酶(T4Dda-(E94C/C109A/C136A/A360C)(具有突变E94C/C109A/C136A/A360C以及(△M1)G1G2)的SEQ ID NO:412)的快照(运行图1至3)。

[0082] 图41表示模拟中在30和40ns取得的位于孔(CsgG-Eco-(Y51T/F56Q)-Strepll(C)9(具有突变Y51T/F56Q的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且在C端被连接到孔突变体No.20))顶部的酶(T4Dda-(E94C/C109A/C136A/A360C)(具有突变E94C/C109A/C136A/A360C以及(△M1)G1G2)的SEQ ID NO:412)的快照(运行图1至3)。

[0083] 图42表示处于前孔构象的$CsgG_{C1S}$的X射线结构。a,$CsgG_{C1S}$单体的带状图,其从N端到C端着色为蓝色到红色的彩虹。二级结构元素根据ABD样折叠进行标记,额外的N-端和C-端的α-螺旋和连接β1和α1的延伸环分别标记为αN,αC和C环(CL)。b,$CsgG_{C1S}$ C8八聚体的侧视图,其具有由颜色区分的亚单元和一个具有如图a中进行取向和着色的亚单元。

[0084] 图43表示处于其通道构象的CsgG的结构。a,$CsgG_{C1S}$(蓝色)和膜抽提的CsgG(红色)的ATR-FTIR光谱的酰胺I区域(1,700-1,600cm^{-1})。b,TM1和TM2序列(SEQ ID NO:449和SEQ ID NO:450)(面向双分子层的残基,为蓝色)以及与野生型csgG(WT)、空载体或缺乏TM1或TM2的下划线片段的csgG互补的大肠杆菌BW25141ΔcsgG的刚果红结合区。数据采用3组生物学重复试验表示。c,处于前孔构象(浅蓝色;TM1粉色,TM2紫色)和通道构象(褐色;TM1绿色;TM2橙色)的CsgG单体的覆盖图。CL,C环。d,e,带面表示法中的CsgG九聚体的侧视图(d)和截面图(e);螺旋2,核心结构域和TM发夹结构分别用蓝色,浅蓝色和褐色表示。单个原聚体如图42a进行着色。品红色球表示Leu 2的位置。OM,外膜。

[0085] 图44表示CsgG通道收缩结构。a,CsgG通道收缩结构的截面图及其排除溶剂的直径。b,收缩结构由三个堆叠的同心侧链层组成:Tyr 51,Asn 55和Phe 56,其前面是来自周质侧的Phe 48。c,CsgG通道的拓扑结构。d,与csgG(WT)、空载体或带有显示收缩结构突变的csgG互补的大肠杆菌BW25141ΔcsgG的刚果红结合区。数据采用6组生物学重复试验表示。e,f,表示在平面磷脂双分子层中重组并在+50mV(n＝33)或-50mV(n＝13)的电场下测量的CsgG代表性单通道电流记录(e)和电导直方图(f)。

[0086] 图45表示CsgG转运机制的模型。a,CsgE(E),CsgG(G)和补充过量CsgG(E+G)的CsgG的非变性凝胶电泳图(NativePAGE),显示了CsgG-CsgE复合物(E-G*)的形成。数据采用包括4组不同批次蛋白的7组实验表示。b,从非变性凝胶电泳(native PAGE)回收的CsgE(E),CsgG(G)和E-G*复合物的SDS-PAGE。数据采用2组重复试验表示。M,分子量标记。c,CsgG-CsgE颗粒的选取种类的平均图。从左至右:通过冷冻电子显微镜(cryo-EM)看到的顶视图和侧视图,以及负染色侧视图与CsgG九聚体的比较图。d,顶视和斜侧视的CsgE颗粒冷冻电子显微镜(cryo-EM)平均值。旋转自相关显示为九重对称。e,CsgG-CsgE的三维重构图(24A°分辨率,1221个单颗粒)显示了包括CsgG(蓝色)的九聚体颗粒以及另外表示为CsgE九聚体(橙

色)的密度。f,PPB重组CsgG的单通道电流记录,该电流记录在+50mV或-50mV,并补充增加CsgE浓度的条件下进行。水平刻度条位于0pA处。g,CsgG介导的蛋白质分泌的试验模型(Tentative model)。CsgG和CagE被提出用于形成捕获CsgA的分泌复合物(在图54中讨论),在通道上产生熵势。在通道收缩结构中捕获CsgA后,DS-整流的布朗扩散(DS-rectified Brownian diffusion)促使多肽逐渐移位穿过外膜。

[0087] 图46表示大肠杆菌中的Curli生物合成途径。较大的curli亚单元CsgA(浅绿色)作为可溶性单体蛋白从细胞中分泌出来。较小的curli亚单元CsgB(深绿色)与外膜(OM)相连并且作为将CsgA从可溶性蛋白转化为淀粉样蛋白沉淀的成核剂。CsgG(橙色)组装成外膜中的寡聚curli-特异性转运通道。CsgE(紫色)和CsgF(浅蓝色)形成多产CsgA以及CsgB的转运和沉积所需的可溶性辅助蛋白。CsgC形成功能未知的假定的氧化还原酶。所有curli蛋白具有假定的用于跨细胞质(内)膜(IM)转运的Sec信号序列。

[0088] 图47表示通过尺寸排阻色谱和负染色电子显微镜分析的CsgG和$CsgG_{C1S}$在溶液中的低聚态。a,C8E4/LDAO溶解的CsgG原始负染色EM照片。箭头表示如g中所示的尺寸洗脱曲线中标记的不同颗粒群,其为(I)CsgG九聚体的聚集体,(II)CsgG十八聚体和(III)CsgG九聚体。比例尺,20nm。b,所指示的低聚态的顶视图和侧视图的类平均示意图。c,顶视图中LDAO溶解的CsgG的旋转自相关函数图,显示为九重对称性。d,$CsgG_{C1S}$的原始负染色EM照片,箭头表示图g中通过尺寸排阻色谱观察到的十六聚体(IV)和八聚体(V)颗粒。e,$CsgG_{C1S}$低聚物侧视图的类平均示意图。未观察该构建体的顶视图。f,由图g中表示的尺寸排阻色谱观察到的$CsgG_{C1S}$和CsgG颗粒的洗脱体积(EV)、计算的分子质量(MWcalc)、对应CsgG低聚态($CsgG_n$)的预期分子质量(MWCsgG)以及通过负染色EM和X射线晶体衍射观察的颗粒的对称性的表格。g,$CsgG_{C1S}$(黑色)和C8E4/LDAO-溶解的CsgG(灰色)在Superdex 200 10/300GL(GE Healthcare)上进行的尺寸排阻色谱图。h,i,顶视图和侧视图中结晶低聚物的带状示意图,显示了$CsgG_{C1S}$(h)的D8十六聚体和膜提取的CsgG(i)的D9十八聚体。一个原聚体从N端(蓝色)到C端(红色)以彩虹色进行着色。在晶体中捕获的形成尾到尾相接(tail-to-tail)的二聚体的两个C8八聚体($CsgG_{C1S}$)或C9九聚体(CsgG)被染成蓝色和棕褐色。r和h分别表示半径和原聚体间的转角。

[0089] 图48表示CsgG与结构同源物的比较和与CsgG接触的原聚体间的接触。a,b,$CsgG_{C1S}$单体(例如为前孔构象的CsgG)(a)和TolB的核苷酸结合结构域样结构域(b)(PDB 2hqs)的条带图,二者均从N端(蓝色)到C端(红色)着色为彩虹色。普通的二级结构元素以同样的方式进行标记。c,$CsgG_{C1S}$(灰色),从左到右,与油菜黄单胞菌属罕见脂蛋白(Xanthomonas campestris rare lipoprotein)B(PDB 2r76,着色为粉红色)、希瓦氏菌假定脂蛋白(Shewanella oneidensis hypothetical lipoprotein)DUF330(PDB 2iqi,着色为粉红色)和大肠杆菌TolB(PDB 2hqs,其N-端和b-螺旋桨(b-propeller)结构域分别着色为粉红色和黄色)层叠在一起。CsgG特异性结构元素被标记和上色,如左上图所示。d,e,在CsgG结构中发现的两个相邻的原聚体的条带图,该图沿双分子层的平面从低聚物的外部(c)或内部(d)观察。第一个原聚体从N端到C端用彩虹色(深蓝色至红色)表示;第二个原聚体用浅蓝色(核心结构域)、蓝色(螺旋2)和褐色(TM结构域)表示。四个主要的低聚界面是明显的:b6-b39主链在b-桶结构内的相互作用,收缩环(CL),与b1-b3-b4-b5抵接的螺旋1(α1)的侧链堆叠以及螺旋2(α2)的螺旋-螺旋堆叠。还观察到,连接脂质锚定物(品红色球表示Leu 2中Ca位置)

和N端螺旋(αN)的18个残基的N-端环包裹相邻的两个原聚体。脂质锚定物的投影位置预期与+2原聚体的TM1和TM2发夹结构相抵接(为清楚起见未示出)。

[0090] 图49表示CsgG寡聚体中选取的表面残基的Cys可接近性(accessibility)测试。a-c,周质(a),侧面(b)和胞外(c)视图中显示的CsgG九聚体的条带示意图。一个原聚体从N端(蓝色)到C端(红色)着色为彩虹色。半胱氨酸取代基被标记,并且S原子的等效位置显示为球体,根据对大肠杆菌外膜中MAL-PEG(5,000Da)标记的可接近性进行上色。d,在SDS-PAGE上分析的与MAL-PEG反应的样品的蛋白质印迹(Western blot),显示了所引入的半胱氨酸在MAL-PEG上的结合增加了5kDa。可接近的(11和111)、中等可接近的(1)和不可接近的(2)位点在a-e中分别着色为绿色、橙色和红色。对于Arg 97和Arg 110,存在44kDa处的第二物质,对应于引入的和天然的半胱氨酸均被标记的蛋白质的一部分。数据采用4组独立生物学重复性试验表示。e,存在于X射线结构中的D9十八聚体中的二聚界面的侧视图。在二聚体界面或D9颗粒的内腔内引入的半胱氨酸被标记。在与膜结合的CsgG中,这些残基对于MAL-PEG是可接近的,证明D9颗粒是膜提取的CsgG浓缩溶液的产物并且C9复合物形成生理上相关的物质。发现,C端螺旋中的残基(aC;Lys 242,Asp 248和His255)是不可接近至难以接近的,表明aC可与大肠杆菌细胞膜,可能是肽聚糖层,形成额外的接触。

[0091] 图50表示用聚丙氨酸链模型对CsgG收缩结构进行的分子动力学模拟。a,b,以伸展构象螺旋穿过通道的聚丙氨酸链模拟的CsgG收缩结构的顶视图(a)和侧视图(b),此处采用C-端至N-端的方向表示。通过CsgG转运体的底物通道本身不是序列特异性的[参考文献16,23]。为了清楚,将聚丙氨酸链用于模拟一正经过的多肽链的假定的相互作用。所述模拟的区域由9个同心的CsgG C环结构组成,每个C环结构均包含残基47-58。连接所述收缩结构的侧链用棒状示意图表示,其中Phe 51着色为板岩色(slate blue),Asn 55(酰胺夹)着色为青色,Phe 48和Phe 56(Φ夹))分别着色为浅橙色和深橙色。N,O和H原子(只示出了羟基或侧链上的酰胺H原子)分别着色为蓝色、红色和白色。所述聚丙氨酸链中的C,N,O和H原子分别着色为绿色、蓝色、红色和白色。在所述收缩结构中所述聚丙氨酸残基10Å范围内的溶剂分子(水)表示为红点。c,所述聚丙氨酸的模拟溶剂化作用,处于图b中所示位置并且为了清楚除去了C-环结构(示出的溶剂分子为在整个聚丙氨酸10Å范围内的溶剂分子)。在酰胺夹和Φ夹的高度,所述聚丙氨酸链的溶剂化作用减弱为桥接肽骨架和酰胺夹侧链的单一水壳。Tyre 51环中的多数侧链,与其在CsgG(和$CsgG_{C1S}$)X射线结构中观察到的向内的指向中心的位置相比,朝向所述溶剂转动。该模型是使用GROMACS[参考文献53]进行40ns全原子显式溶剂分子动力学模拟的结果,该模拟采用AMBER99SB-ILDN54力场并且位于C环末端的残基(Gin 47和Thr 58)的Cα原子的位置被限制。

[0092] 图51表示CsgG同源物中的序列保守性。a,CsgG九聚体的表面示意图,依据序列相似性着色(对着色的黄色到蓝色进行从低到高的保守性打分),并从周质(最左侧)、侧面(左中)、细胞外环境(右中)观察或作为横截面视图(最右侧)。所述附图显示了,周质前庭的入口、收缩环的前庭侧和TM结构域的腔表面的最高序列保守性映射图的区域。b,CsgG样脂蛋白的多重序列比对。所选序列选自跨系统发育的三个CsgG样序列的单系分支(monophyletic clades)(未显示),以给出序列多样性的代表性视图。二级结构元素的β-链和α-螺旋分别用箭头或棒状物表示,并且所述二级结构元素基于大肠杆菌CsgG晶体结构。c,d,CsgG原聚体的二级结构示意图(c)和CsgG九聚体表面示意图的横截面视图(d),其均为

灰色,并且三个连续的高序列保守性的区域着色为红色(HCR1)、蓝色(HCR2)和黄色(HCR3)。HCR1和HCR2形成收缩环的前庭侧;HCR3对应于螺旋2,位于周质前庭的入口处。在所述收缩结构内部,Phe 56为100%保守,而Asn 55可以被Ser或Thr保守性取代,例如被可作为氢键供体/受体的小极性侧链取代。所述收缩结构出口处的同心侧链环(Tyr51)不是保守的。在所述收缩结构入口处的Phe-环的存在于拓扑结构上类似于炭疽保护性抗原PA63中的Phe 427环(称为Φ夹)),其中显示其催化多肽的捕获和通过过程[参考文献20]。toxB超家族蛋白的MST显示在Phe环高度处具有保守的基序D(D/Q)(F)(S/N)S。这类似于curli样转运体中所见的基序S(Q/N T)(F)ST。尽管所述孔构象中原子级分辨率的PA63结构尚无法得到,但已得到的结构表明,所述Phe-环可以类似地随后由保守性氢键供体/受体(Ser/Asn428),作为随后的移位通道中的同心环(需说明的是,在两个转运体中所述元素的取向都是反转的)。

[0093] 图52表示CsgG和CsgG:CsgE孔的单通道电流分析.a,在负场电位下,CsgG孔显示出两种电导状态。左上图和右上图分别显示了正常(在+50,0和-50mV下测量)的和低电导形式(在0,+50和-50mV下测量)的代表性单通道电流轨迹。在总观察时间(n=22)内,没有观察到两种状态之间的转换,表明所述电导状态具有长寿命(时间尺度为秒到分钟)。左下图显示在+50和-50mV(n=33)下获得的CsgG孔的正常的和低电导形式的电流直方图。CsgG孔在常规和低电导下的I-V曲线用右下图表示。所述数据表示在至少四个独立记录下的平均值和标准方差。所述低电导形式的自然或生理存在是未知的。b,采用辅助因子CsgE滴定的CsgG通道的电生理学性质。该曲线显示了作为CsgE浓度的函数的处于开放状态、中间态和闭合状态的通道的比例。CsgG的开放和关闭状态如图45f所示。将CsgE浓度提高到多于10nM导致CsgG孔闭合。所述效应发生在+50mV(左图)和-50mV(右图)下,排除了CsgE(计算的pI 4.7)电泳到CsgG孔中导致所述孔堵塞的可能性。低概率(,5%)的中间状态大约具有开放通道电导的一半。其可能表示CsgE引起的所述CsgG通道的不完全闭合;或者,其可能表示由来自电解质溶液的残留CsgG单体结合到嵌入膜的孔引起的CsgG二聚体的暂时形成。从单通道电流迹线的全点直方图分析可获得上述三种状态的比例。上述直方图产生的最多三个状态的峰值区域,以及某一给定状态的比例可通过将相应的峰面积除以记录中所有其他状态的总和获得。在负场电位下,两个开放电导状态被识别,类似于对CsgG的观察结果(参见图a)。因为两种开放通道的变化都被较高的CagE浓度所阻碍,所以b中的“开放”迹线结合了两种电导形式。曲线图中的数据表示从三个独立记录得到的平均值和标准方差。c,晶体结构,尺寸排阻色谱和EM显示,洗涤剂提取的CsgG孔在较高蛋白质浓度下形成非天然尾-尾堆叠的二聚体(例如,两个如D9颗粒的九聚体;图47)。这些二聚体也可以在单通道记录中观察到。上图显示在+50,0和-50mV(从左到右)条件下堆叠的CsgG孔的单通道电流迹线。左下图显示在+50和-50mV条件下记录的二聚CsgG孔的电流直方图。分别在+50和-50mV条件下的实验电导值+16.2±1.8和-16.0±3.0pA(n=15)均接近理论计算值23pA。右下图显示堆叠的CsgG孔的I-V曲线。数据表示六个独立记录的平均值和标准方差。d,通过电生理学测试CsgE结合和阻断堆叠的CsgG孔的能力。其显示的为在+50mV(上图)和-50mV(下图)条件下,10或100nM CsgE存在下,堆叠的CsgG孔的单通道电流迹线。所述电流迹线表明,CsgE的原本饱和浓度并不会导致堆叠的CsgG二聚体的孔闭合。这些观察结果与CsgG-CagE接触区至螺旋2和CsgG周质腔入口的映射图很一致,如由EM和定点突变所识别(图45和图52)。

[0094] 图53表示CsgE寡聚体和CsgG-CsgE复合物,a,CsgE的尺寸排阻色谱(Superose 6,

16/600;电泳缓冲液20mM Tris-HCl pH8,100mM NaCl,2.5%甘油)显示两种寡聚状态,1和2,的一种平衡,表观分子质量比为9.16:1。峰1的负染色电子显微镜(Negative-stain EM)观察显示,离散的CsgE颗粒(插图为五个代表性种类平均图像,按增加的倾斜角度排序)与9个CsgE复本的尺寸相当。b,由冷冻电子显微镜(cryo-EM)从顶部视角观察到的CsgE低聚物的所选种类平均图像并且其旋转自相关图显示C9对称性的存在。c,CsgG-CsgE冷冻电子显微镜(cryo-EM)模型的FSC分析。采用对应1221个颗粒的125类,在相关性为0.5的阈值下,通过FSC分析确定了,三维重构实现了24Å的分辨率。d,CsgG-CsgE冷冻电子显微镜密度和X射线结构中观察到的所述CsgG九聚体的叠加图(Overlay)。所述叠加图显示为从侧面视角的半透明密度图(左)或横截面图。e,与野生型csgG(WT),空载体(ΔcsgG)或csgGhelix 2突变体(用单字母编码标记的单个氨基酸替换)互补的大肠杆菌BW25141ΔcsgG的刚果红结合区。数据采用4组生物学重复试验表示。f,胆汁盐毒性对与csgG(WT)互补的大肠杆菌LSR12或携带不同螺旋2突变的csgG(与csgE互补(1)或不互补(2))的影响。从107个细菌开始的10倍连续稀释液被点在MacConkey琼脂平板上。外膜中CsgG孔的表达导致胆汁盐敏感性的增加,其可以通过CsgE的共表达(n=6,三组生物学重复试验,每组重复两次)进行阻断。g,分子表面示意图中CsgG X射线结构的横截面视图。对刚果红结合区或毒素没有影响的CsgG突变体显示为蓝色;对由CsgE介导的胆汁盐敏感性的救助进行干扰的突变体以红色表示。

[0095] 图54表示CsgG分泌复合物的组装和底物征集。所述Curli转运体CsgG与可溶性分泌辅助因子CsgE形成化学计量比为9:9的分泌复合物,其围成一个24,000Å3的腔室,该腔室被认为用于捕获CsgA底物并促进其跨外膜(OM;参见上文和图45)的熵驱动扩散。在理论上而言,对于底物征集和分泌复合物的组装可以设想3种假设途径(a-c)。a,“捕捉-和-盖帽”(“catch-and-cap”)机制需要CsgA与载脂蛋白CsgG移位通道结合(1),导致后者的构象发生变化,为CsgE的结合暴露出高亲和力的结合平台(2)。CsgE的结合导致对底物笼结构的盖帽(capping)。在CsgA分泌时,CsgG将回到其低亲和力的构象,导致CsgE的解离和分泌通道的释放,以进行新的分泌循环。b,在“对接-和-捕获”('dock-and-trap')机制中,CagE首先捕获细胞周质的CsgA(1),导致CagE采用高亲合性复合物,该复合物对接到CsgG移位孔上(2),围住分泌复合物中的CsgA。CsgA可以直接与CsgE寡聚体或CagE单体结合,后者导致随后的寡聚和CsgG的结合。CsgA的分泌导致CagE回到其低亲和力的构象并与分泌通道解离。c,CsgG和CagE形成构成型(constitutive)复合物,其中在CsgA征集和分泌过程中CagE的构象动力学在开放和关闭形式之间循环。目前公布的或现有的数据无法使我们区分这些推定的征集模式或其衍生模式,或着提出上述模式中的一种模式。

[0096] 图55表示$CsgG_{C1S}$和CsgG的数据采集统计图和电子密度映射图。a,$CsgG_{C1S}$和CsgG X射线结构的数据采集统计图。b,在2.8Å下$CsgG_{C1S}$的电子密度映射图,其采用NCS平均和密度修改的实验SAD相计算,并在1.5σ下绘制等值线。所述映射图显示了通道构建的区域(CL;标记了单个原聚体),并且被覆盖在最终细化模型上。c,CsgG TM结构域中的电子密度映射图(沿着倒易矢量a*,b*和c*的分辨率分别为3.6,3.7和3.8Å),其由NCS平均的和密度修饰的分子置换相计算得到(输入模型中不存在TM环);B-因子(B-factor)锐化-20$Å^2$并在1.0σ下绘制等值线。该图显示,单个CsgG原聚体的TM1(Lys 135-Leu 154)和TM2(Leu 182-Asn 209)区域,覆盖在最终的细化模型上。

[0097] 图56表示与携带单链DNA突出端的DNA发夹结构相互作用的CsgG WT蛋白质的单通

道电流迹线(左)和其放大区域(右)。所述迹线显示,响应于在+50mV或-50mV间隔下测量的电位而改变的电流(如箭头所示)。最后的+50mV区段的下行电流的阻滞表明发夹双链同时进入孔腔内,并将单链发夹末端螺旋入孔的内部结构,导致几乎完全的电流阻滞。将电场反转至-50mV导致孔的电泳解阻塞(electrophoretic unblocking)。新的+50mV阶段再次导致DNA发夹结构的进入/螺旋化以及孔的堵塞。在所述+50mV区段上,发夹结构的解旋可以导致所述电流阻滞的停止,如所述电流阻滞的反转所示。所述发夹结构——其具有序列:3'GCGGGGA GCGTATT AGAGTTG GATCGGATGCA GCTGGCTACTGACGTCATGACGTCAGTAGCCAGCATGCAT CCGATC-5'——被加入到所述腔室的顺式侧,终浓度为10nM。

[0098] 图57表示CsgG-ΔPYPA的纯化和通道性质,所述CsgG-ΔPYPA是突变的CsgG孔,其中,在所述收缩残基Y51处PYPA序列(残基50-53)突变为GG。

[0099] 序列表说明

[0100] SEQ ID NO:1表示包含信号序列的野生型大肠杆菌CsgG的氨基酸序列(Uniprot访问号P0AEA2)。

[0101] SEQ ID NO:2表示包含信号序列的野生型大肠杆菌CsgG的多核苷酸序列(基因ID:12932538)。

[0102] SEQ ID NO:3表示SEQ ID NO:2中位置53至77的野生型大肠杆菌CsgG的氨基酸序列。这对应于成熟野生型大肠杆菌CsgG单体的位置38至63的氨基酸序列(即缺少所述信号序列)。

[0103] SEQ ID NO:4至388表示缺失信号序列的CsgG的修饰单体的第38至63位的氨基酸序列。

[0104] SEQ ID NO:389表示编码来自大肠杆菌Str.K-12substr.MC4100的野生型CsgG单体的密码子优化的多核苷酸序列。该单体缺少所述信号序列。

[0105] SEQ ID NO:390表示来自大肠杆菌Str.K-12substr.MC4100的成熟的野生型CsgG单体的氨基酸序列。该单体缺少所述信号序列。用于此处的缩写CsgG=CsgG-Eco.

[0106] SEQ ID NO:391表示假设蛋白CKO_02032的YP_001453594.1:1-248的氨基酸序列[克氏柠檬酸杆菌(Citrobacter koseri)ATCC BAA-895],其与SEQ ID NO:390序列99%相同。

[0107] SEQ ID NO:392表示curli生产组装/运输组件CsgG的WP_001787128.1:16-238的氨基酸序列,部分[沙门氏菌(Salmonella enterica)],其与SEQ ID NO:390序列98%相同。

[0108] SEQ ID NO:393表示curli生产组装/运输蛋白CsgG的KEY44978.1|:16-277的氨基酸序列[无丙二酸柠檬酸杆菌(Citrobacter amalonaticus)],其与SEQ ID NO:390序列98%相同。

[0109] SEQ ID NO:394表示curli生产组装/运输组件的YP_003364699.1:16-277的氨基酸序列[鼠柠檬酸杆菌(Citrobacter rodentium)ICC168],其与SEQ ID NO:390序列97%相同。

[0110] SEQ ID NO:395表示curli生产组装/运输组件CsgG的YP_004828099.1:16-277的氨基酸序列[阿氏肠杆菌(Enterobacter asburiae)LF7a],其与SEQ ID NO:390序列94%相同。

[0111] SEQ ID NO:396表示编码Phi29DNA聚合酶的多核苷酸序列。

[0112] SEQ ID NO:397表示Phi29DNA聚合酶的氨基酸序列。

[0113] SEQ ID NO:398表示源自大肠杆菌sbcB基因的密码子优化的多核苷酸序列。它编码来自大肠杆菌的核酸外切酶I酶(EcoExo I)。

[0114] SEQ ID NO:399表示来自大肠杆菌的核酸外切酶I酶(EcoExo I)的氨基酸序列。

[0115] SEQ ID NO:400表示源自大肠杆菌的xthA基因的密码子优化的多核苷酸序列。它编码来自大肠杆菌的核酸外切酶III酶。

[0116] SEQ ID NO:401表示来自大肠杆菌的核酸外切酶III酶的氨基酸序列。该酶从3'至5'方向对双链DNA(dsDNA)的其中一条链进行5'单磷酸核苷的分布式消解。所述酶在链上的引发需要约4个核苷酸的5'突出结构。

[0117] SEQ ID NO:402表示源自极端嗜热链球菌(T.thermophilus)的recJ基因的密码子优化的多核苷酸序列。它编码来极端嗜热链球菌的RecJ酶(TthRecJ-cd)。

[0118] SEQ ID NO:403表示来自极端嗜热链球菌的RecJ酶的氨基酸序列(TthRecJ-cd)。该酶从5'至3'方向对单链DNA(ssDNA)的5'单磷酸核苷进行进行性消解(processive digestion)。所述酶在链上的引发需要至少4个核苷酸。

[0119] SEQ ID NO:404表示源自细菌噬菌体λ外切(redX)基因的密码子优化的多核苷酸序列。它编码所述细菌噬菌体λ核酸外切酶。

[0120] SEQ ID NO:405表示细菌噬菌体λ核酸外切酶的氨基酸序列。该序列是组装成三聚体的三个相同的亚基之一。该酶从5'至3'方向对双链DNA(dsDNA)的其中一条链进行核苷酸的高度进行性消解(http://www.neb.com/nebecomm/products/productM0262.asp)。所述酶在链上的引发优选需要约4个带有5'磷酸的核苷酸的突出结构。

[0121] SEQ ID NO:406表示Hel308Mbu的氨基酸序列。

[0122] SEQ ID NO:407表示Hel308Csy的氨基酸序列。

[0123] SEQ ID NO:408表示Hel308Tga的氨基酸序列。

[0124] SEQ ID NO:409表示Hel308Mhu的氨基酸序列。

[0125] SEQ ID NO:410表示Tral Eco的氨基酸序列。

[0126] SEQ ID NO:411表示XPD Mbu的氨基酸序列。

[0127] SEQ ID NO:412表示Dda 1993的氨基酸序列。

[0128] SEQ ID NO:413表示Trwc Cba的氨基酸序列。

[0129] SEQ ID NO:414表示转运体的WP_006819418.1:19-280[雷金斯堡约克氏菌(Yokenella regensburgei)]的氨基酸序列,其与SEQ ID NO:390序列91%相同.

[0130] SEQ ID NO:415表示curli生产组装/运输蛋白CsgG的WP_024556654.1:16-277[粉尘克洛诺菌(Cronobacter pulveris)]的氨基酸序列,其与SEQ ID NO:390序列89%相同。

[0131] SEQ ID NO:416表示curli生产组装/运输蛋白CsgG的YP_005400916.1:16-277[水生拉恩氏菌(Rahnella aquatilis)HX2]的氨基酸序列,其与SEQ ID NO:390序列84%相同。

[0132] SEQ ID NO:417表示CsgG家族curli生产组装/运输组件的KFC99297.1:20-278的氨基酸序列[抗坏血酸克吕沃尔菌(Kluyvera ascorbata)ATCC 33433],其与SEQ ID NO:390序列82%相同。

[0133] SEQ ID NO:418表示CsgG家族curli生产组装/运输组件的KFC86716.11:16-274的氨基酸序列[蜂房哈夫尼菌(Hafnia alvei)ATCC 13337],其与SEQ ID NO:390序列81%相

同。

[0134] SEQ ID NO:419表示参与形成curli聚合物的未表征蛋白的YP_007340845.1|:16-270的氨基酸序列[肠杆菌科细菌菌株(Enterobacteriaceae bacterium strain)FGI 57],其与SEQ ID NO:390序列76%相同。

[0135] SEQ ID NO:420表示curli生产组装/运输蛋白CsgG的WP_010861740.1:17-274的氨基酸序列[邻单胞痢疾杆菌(Plesiomonas shigelloides)],其与SEQ ID NO:390序列70%相同。

[0136] SEQ ID NO:421表示curli生产组装/运输外膜脂蛋白组分CsgG的YP_205788.1:23-270的氨基酸序列[费氏弧菌(Vibrio fischeri)ES1 14],其与SEQ ID NO:390序列60%相同。

[0137] SEQ ID NO:422表示curli生产组装蛋白CsgG的WP_017023479.1:23-270的氨基酸序列[变形菌(Aliivibrio logei)],其与SEQ ID NO:390序列59%相同。

[0138] SEQ ID NO:423表示Curli生产组装/运输组分CsgG的WP_007470398.1:22-275的氨基酸序列[发光杆菌属(Photobacterium sp.)AK15],其与SEQ ID NO:390序列57%相同。

[0139] SEQ ID NO:424表示curli生产组装蛋白CsgG的WP_021231638.1:17-277的氨基酸序列[维氏气单胞菌(Aeromonas veronii)],其与SEQ ID NO:390序列56%相同。

[0140] SEQ ID NO:425表示Curli生产组装/运输蛋白CsgG的WP_033538267.1:27-265的氨基酸序列[希瓦氏菌属(Shewanella sp.)ECSMB14101],其与SEQ ID NO:390序列56%相同。

[0141] SEQ ID NO:426表示curli生产组装蛋白CsgG的WP_003247972.1:30-262的氨基酸序列[恶臭假单胞菌(Pseudomonas putida)],其与SEQ ID NO:390序列54%相同。

[0142] SEQ ID NO:427表示curli生产组装/运输组分CsgG的YP_003557438.1:1-234的氨基酸序列[紫色希瓦氏菌(Shewanella violacea)DSS12],其与SEQ ID NO:390序列53%相同。

[0143] SEQ ID NO:428表示curli生产组装/运输蛋白CsgG的WP_027859066.1:36-280的氨基酸序列[詹氏海杆菌(Marinobacterium jannaschii)],其与SEQ ID NO:390序列53%相同。

[0144] SEQ ID NO:429表示Curli生产组装/运输组分CsgG的CEJ70222.1:29-262的氨基酸序列[金黄杆菌(Chryseobacterium oranimense)G311],其与SEQ ID NO:390序列50%相同。

[0145] SEQ ID NO:430表示实施例18中使用的多核苷酸序列。

[0146] SEQ ID NO:431表示实施例18中使用的多核苷酸序列。

[0147] SEQ ID NO:432表示实施例18中使用的多核苷酸序列。

[0148] SEQ ID NO:433表示实施例18中使用的多核苷酸序列。

[0149] SEQ ID NO:434表示实施例18中使用的多核苷酸序列。连接到SEQ ID NO:434的3'端的是6个iSp18间隔区,其另一端连接到两个胸腺嘧啶和3'胆固醇TEG上。

[0150] SEQ ID NO:435表示StepII(C)的氨基酸序列。

[0151] SEQ ID NO:436表示Pro的氨基酸序列。

[0152] SEQ ID NO:437至440表示来自实施例1的引物。

[0153] SEQ ID NO:441表示来自实施例21的发夹结构。

[0154] SEQ ID NO:442至448表示来自图4的序列。

[0155] SEQ ID NO:449至450表示来自图43的序列。

具体实施方式

[0156] 除非特别说明,本发明的实施采用化学,分子生物学,微生物学,重组DNA技术以及化学方法中的常规技术,其在本领域普通技术人员的能力范围之内。这些技术也均在文献中具有解释,例如,M.R.Green,J.Sambrook,2012,分子克隆:实验室手册,第四版,1-3册,Cold Spring Harbor Laboratory出版社,Cold Spring Harbor,NY;Ausubel,F.M.ef al.(1995和定期补充;分子生物学电流协议,第9,13和16章,John Wiley&Sons,New York,N.Y.);B.Roe,J.Crabtree,and A.Kahn,1996,DNA分离和测序:基本技术,John Wiley&Sons;J.M.Polak and James O'D.McGee,1990,原位杂交:原理与实践,牛津大学出版社;M.J.Gait(编辑),1984,寡核苷酸的合成:实用方法,IRL出版社;以及D.M.J.Lilley和J.E.Dahlberg,1992,酶学方法:DNA结构A部分:酶学中DNA合成和物理分析方法,Academic出版社。上述这些文献的全部内容均通过引用纳入本文。

[0157] 在阐述本发明之前,提供许多定义将有助于本发明的理解。本文引用的所有参考文献通过援引整体并入本文。除非特别说明,本文使用的所有技术和科学术语与本发明所属领域的普通技术人员通常的理解具有相同的含义。

[0158] 如本文所使用的,术语"包括"是指必须包括任何所列举的要素,并且也可以任选地包括其他元素。"基本上由...组成"意味着必须包括任何列举的要素,排除对列举的要素的基本和新特性具有重大影响的元素,并且可以可选地包括其它元素。"由...组成"是指不包括所有未列举的元素。由这些术语中的每一个定义的实施例在本发明的范围内。

[0159] 本文所用的术语"核酸"是指单链或双链共价连接的核苷酸序列,其中每个核苷酸上的3'和5'末端通过磷酸二酯键连接。所述多核苷酸可以由脱氧核糖核苷酸碱基或核糖核苷酸碱基组成。核酸可以包括DNA和RNA,并可以在体外合成制备或从自然资源中分离。核酸可以进一步包括修饰的DNA或RNA,例如甲基化的DNA或RNA,或经过翻译后修饰的RNA,例如用7-甲基鸟苷进行的5'-盖帽,3'-端加工,例如裂解和多腺苷化,以及拼接。核酸还可以包括合成核酸(XNA),例如己糖醇核酸(HNA),环己烯核酸(CeNA),苏糖核酸(TNA),甘油核酸(GNA),锁核酸(LNA)和肽核酸(PNA)。核酸(本文也称为"多核苷酸")的大小通常用双链多核苷酸的碱基对(bp)数目表示,或在单链多核苷酸的情况下用核苷酸的数目(nt)表示。1千个bp或nt等于一个千碱基对(kb)。长度小于约40个核苷酸的多核苷酸通常称为"寡核苷酸",并且可以包含用于DNA操作(例如通过聚合酶链式反应(PCR))中的引物。

[0160] 本发明中的术语"氨基酸"使用其最广泛的含义,其含义包括天然存在的Lα-氨基酸或残基。本文使用的天然存在氨基酸通常使用单字母或三字母缩写为:A=Ala;C=Cys;D=Asp;E=Glu;F=Phe;G=Gly;H=His;I=IIe;K=Lys;L=Leu;M=Met;N=Asn;P=Pro;Q=Gln;R=Arg;S=Ser;T=Thr;V=Val;W=Trp;and Y=Tyr(Lehninger,A.L,(1975)生物化学,第2版,71-92页,Worth Publishers,New York)。一般术语"氨基酸"还包括D-氨基酸,复古反转氨基酸(retro-inverso amino acid)以及化学修饰的氨基酸,例如氨基酸类似物,通常不包含在蛋白质中的天然存在的氨基酸,例如正亮氨酸,以及化学合成的具有本领

域已知的氨基酸特征的化合物,如β-氨基酸。例如,苯丙氨酸或脯氨酸的类似物或模拟物——其允许所述肽化合物与天然Phe或Pro具有相同的构象限制——包括在氨基酸的定义中。这种类似物和模拟物在本文中称为相应氨基酸的“功能等价物”。氨基酸的其它实例列举在:Roberts和Vellaccio,肽:分析,合成,生物学,Gross and Meiehofer,eds.,第5卷,341页,Academic Press,Inc.,N.Y.1983,其通过引用并入本文。

[0161] “多肽”是指通过肽键连接的氨基酸残基的聚合物,无论是天然产生的还是通过合成方法体外产生的。长度小于约12个氨基酸残基的多肽通常称为“肽”,长度在约12至约30个氨基酸残基之间的多肽可称为“寡肽”。本文所用的术语“多肽”表示天然存在的多肽,前体形式或前蛋白的产物。多肽也可以经过熟化或翻译后修饰过程,其可以包括但不限于:糖基化,蛋白水解断裂,脂化,信号肽裂解,前肽裂解,磷酸化等。本文所用术语“蛋白质”是指含有一个或多个多肽链的大分子。

[0162] “生物孔”是指限定通道或孔的跨膜蛋白质结构,其允许分子和离子从膜的一侧移位到另一侧。离子类物质穿过孔的移位可以通过施加到所述孔的任一侧的电势差来驱动。纳米孔是一种生物孔,其中分子或离子通过的通道的最小直径是纳米级的(10^{-9}米)。

[0163] 对本发明所有实施方式或方面,多核苷酸可以包含与野生型大肠杆菌CsgG相比具有至少50%,60%,70%,80%,90%,95%或99%完整序列同一性的多核苷酸,如SEQ ID NO:2所示。同样,所述多肽可以包含与野生型大肠杆菌CsgG相比具有至少50%,60%,70%,80%,90%,95%或99%完整序列同一性的多肽,如SEQ ID NO:1所示。多肽可以包含含有PFAM结构域PF03783的多肽,PFAM结构域PF03783为CsgG样蛋白的特征。目前已知的CsgG同源物和CsgG体系结构的列表可以在http://pfam.xfam.org//family/PF03783找到。序列同一性也因而可以是全长多核苷酸或多肽的片段或部分。因此,序列可以与本发明的序列具有仅50%的总体序列同一性,但是特定区域、结构域或亚基可以与本发明的序列共享80%,90%或至多99%的序列同一性。根据本发明,对SEQ ID NO:2的核酸序列的同源性并不仅限于序列同一性。许多核酸序列彼此能够表现出显著的生物学同源性,尽管其具有明显很低的序列同一性。在本发明中,同源核酸序列被认为是在低严格条件下彼此杂合的序列(M.R.Green,J.Sambrook,2012,分子克隆:实验室手册,第四版,1-3册,Cold Spring Harbor Laboratory出版社,Cold Spring Harbor,NY)。

[0164] 术语“载体”用于表示线性或圆形的DNA分子,可以将适当大小的另一核酸(通常为DNA)序列片段整合到其中。上述DNA片段可以包括为由DNA序列片段编码的基因的转录准备的额外片段。所述额外片段可以包括但不限于启动子,转录终止子,增强子,内部核糖体进入位点,非翻译区,多腺苷化信号,可选择标记,复制起点等等。可以使用各种适用于原核(例如,β-内酰胺酶和乳糖启动子系统,碱性磷酸酶,色氨酸(trp)启动子系统,大肠杆菌的lac,tac,T3,T7启动子)宿主和真核(例如,猿猴病毒40早期或晚期启动子,劳斯氏肉瘤病毒(Rous sarcoma virus)长末端重复启动子,巨细胞病毒启动子,腺病毒晚期启动子,EG-1a启动子)宿主的启动子。表达载体通常来自质粒,粘粒,病毒载体和酵母人工染色体;载体通常是含有来自几种来源的DNA序列的重组分子。本发明的具体实施方式提供编码如本文所述的野生型或修饰的CsgG多肽的表达载体。术语“可操作地连接”,当应用于DNA序列时,例如在如上所述的核酸载体中,表示所述序列进行排列使得它们协同地发挥作用以达到其预定目的,即启动子序列允许转录的启始通过相关的编码序列进行到终止序列。

[0165] 生物孔的跨膜蛋白结构本质上可以是单体或寡聚体。通常,所述孔包括围绕中心轴布置的多个多肽亚基,从而形成蛋白质内衬的通道,该通道的延伸基本上垂直于纳米孔残基所在的膜。多肽亚基的数量不受限制。通常,亚基的数量从5至最多30,合适的亚基数量为6至10。或者,在裂解素(perfringolysin)或相关的大膜孔的情况下,亚基的数量未限定。在形成蛋白内衬通道的纳米孔内的蛋白亚基部分通常包括二级结构基序,其可以包括一种或多种跨膜β-桶状结构,和/或α-螺旋部分。

[0166] 应当理解,所公开产品和方法的不同应用可以根据本领域的具体需要进行调整。还应当理解,本文使用的术语仅出于描述本发明具体实施方式的目的,而不是限制性的。

[0167] 另外如本说明书和所附权利要求书所使用的,除非另有内容明确规定,单数形式“一”,“一个”以及“所述”包括复数指代。因此,例如,提及“一个多核苷酸”时包括两个或多个多核苷酸,提及“一个多核苷酸结合蛋白”时包括两个或多个这种蛋白,提及“一个解旋酶”时包括两个或多个解旋酶,提及“一个单体”时是指两种或多种单体,提及“一个孔”时包括两个或多个孔等。

[0168] 本文引用的所有出版物,专利和专利申请,无论是在上文还是下文,它们的全部内容均通过援引包含在本文中。

[0169] 本发明一部分涉及细菌淀粉样蛋白分泌通道(CsgG),其制备方法及其在核酸测序应用和分子传感中的用途。

[0170] CsgG是存在于大肠杆菌外膜中的膜脂蛋白(Uniprot访问号P0AEA2;基因ID:12932538)。在所述外脂质膜中,CsgG形成纳米孔,其包含九个CsgG单体亚基的寡聚复合物。通过II型(脂蛋白)信号序列,CsgG前蛋白移位穿过SEC易位子,随后在成熟CsgG(即带有裂解的II型信号序列的CsgG)的N末端Cys残基上进行三酰化。三酰化或“脂质化”的CsgG被输送到革兰氏阴性宿主的外膜,其中它插入双分子层中作为九聚体孔。非脂质化形式的CsgG,例如,$CsgG_{C1S}$作为处于前孔构象的可溶性蛋白存在于细胞周质中(图42)。

[0171] 野生型CsgG纳米孔的X射线结构(Goyal et al.,自然,2014,516(7530),250-3)显示其具有120Å的宽度和85Å的高度(在下文中,术语纳米孔的“宽度”是指其与膜表面平行的尺寸,而术语纳米孔的“高度”是指其垂直于膜的尺寸)。所述CsgG孔复合物通过36股β-桶状结构穿过膜以提供为40Å内直径的通道(图1)。所述CsgG通道的组装单体各自具有保守的12残基的环结构(C环,“CL”;图2),其共同作用以在通道中形成直径约9.0Å的收缩结构(图1和2)。野生型CsgG纳米孔的收缩结构由三个堆叠的同心环组成,所述环由CsgG寡聚物中存在的每个CsgG单体的氨基酸残基Tyr51,Asn55和Phe56的侧链形成(图3)。这一残基的编号基于在N末端缺失天然15个氨基酸信号序列的成熟蛋白。因此,成熟蛋白对应于SEQ ID NO:1的残基16至277。Tyre56位于SEQ ID NO:1中的66位,Asn在SEQ ID NO:1中的70位,而Phe56位于SEQ ID NO:1的71位。

[0172] 该收缩结构的作用是限制离子和其他分子通过CsgG通道的路径。使用标准电解质条件和+50mV或-50mV的电位,在平面磷脂双分子层中重组的CsgG的单通道电流记录分别导致43.1±4.5pA(n=33)或-45.1±4.0pA(n=13)的稳定电流(图5)。

[0173] 通过添加化学计量的周质因子CsgE(Uniprot访问号P0AE95;实施例10至12)可以有效地阻止电流流过所述CsgG通道。不希望受理论束缚,电流的证据指向一种机制,即CagE与CsgG孔形成复合物,用于封闭通道的一端。可以使用标准单通道记录技术检测到通过

CsgG通道的离子流量的显著减少(实施例12和13,图6)。发明人已经发现,根据本发明的一个实施方案,电流的测量参数(最大电流和监测电流变化的能力)使得纳米孔适合用于核酸测序和分子传感的应用中。

[0174] 因此,本发明的一部分涉及基于流过纳米孔的电流的电学测量值的变化,CsgG纳米孔蛋白复合物在核酸测序中的方法和用途。

[0175] 核酸特别适用于纳米孔测序。DNA和RNA中的天然存在的核酸碱基可以通过其物理尺寸来区分。当核酸分子或单个碱基通过纳米孔的通道时,碱基之间的尺寸差异导致通过通道的离子流的直接相关的降低。离子流中的所述变化可以被记录。用于记录离子流变化的合适的电学测量技术描述于例如WO 2000/28312和D.Stoddart et al.,Proc.Natl.Acad.Sci.,2010,106,pp 7702-7中(单通道记录设备);以及例如WO 2009/077734(多通道记录技术)中。通过适当的校准,离子流的特征的减少可用于识别实时穿过通道的特定核苷酸和相关碱基。

[0176] 跨膜通道中最窄收缩结构的尺寸通常是决定纳米孔用于核酸测序应用的适宜性的关键因素。如果收缩结构太小,待测序的分子将无法通过。然而,为了实现对通过通道的离子流的最大影响,在其最窄点处(即在收缩结构处),通道不应该太大。理想地,任何收缩结构在直径上应尽可能地与通过的碱基的尺寸相近。对于核酸和核酸碱基的测序,合适的收缩结构直径在纳米范围内(10^{-9}米范围)。合适地,直径应在0.5至1.5nm范围内,通常直径在0.7至1.2nm范围内。野生型大肠杆菌CsgG中的收缩结构具有约9Å(0.9nm)的直径。发明人推断,所述CsgG通道中的收缩结构的大小和构象适用于核酸测序。

[0177] 对于与核酸测序相关的应用,所述CsgG纳米孔可以以野生型形式使用,或者可以进一步修饰,例如通过特定氨基酸残基的定向突变,以进一步增强所使用的纳米孔所需要的性质。例如,在本发明的实施方式中,考虑采用突变改变通道内收缩结构的数量,大小,形状,位置或取向。修饰的突变CsgG纳米孔复合物可以通过已知的基因工程技术制备,其导致多肽序列中特异性靶向氨基酸残基的插入,取代和/或缺失。在寡聚CsgG纳米孔的情况下,可以在每个单体多肽亚基,或任何一个单体或所有单体上进行突变。合适地,在本发明的一个实施方式中,对在寡聚蛋白结构内的所有单体多肽进行所述的突变。

[0178] 根据本发明的一个实施方式,提供了一种修饰的突变CsgG纳米孔,其中所述孔内通道收缩结构的数量减少。

[0179] 野生型大肠杆菌CsgG孔包括两个通道收缩结构(参见图1)。这些收缩结构由(i)氨基酸残基Phe56和Asn55,和(ii)氨基酸残基Tyr 51形成,作为更宽结构的一部分,其包括从54位至53位的另外的氨基酸以及C-环基序(图2和图3)。

[0180] 在通常的纳米孔核酸测序中,当感兴趣的核酸序列的各个核苷酸顺序地通过所述纳米孔的通道时,由于所述核苷酸对所述通道的部分阻塞,所述开放通道的离子流被减少。使用上述合适的记录技术测量上述这种离子流的减少。离子流的减少可以校准为已知核苷酸通过所述通道时所测得的离子流的减少,得到用于确定哪个核苷酸通过所述通道的方法,并且从而得到当核酸顺序通过所述通道时,确定通过所述纳米孔的核酸的核苷酸序列的方法。为了精确测定各个核苷酸,通常要求通过所述通道的离子流的减少与通过所述单个收缩结构(或“读取头”)的各个核苷酸的大小直接相关。应当理解,可以通过例如相关的聚合酶的作用,使所述测序是针对"螺旋"穿过(“threaded”through)所述孔的完整核酸聚

合物进行的。或者,序列可以由已经从靠近孔的目标核酸上顺序除去的核苷酸三磷酸碱基的通过进行测定(参见例如WO 2014/187924)。

[0181] 当存在两个或更多个收缩结构并且它们相互间隔开时,每个收缩结构可以同时在核酸链中读取分离的核苷酸或与其相互作用。在这种情况下,通过所述通道的离子流的减少将是所有含有核苷酸的收缩结构的流的合并限制的结果。因此,在一些情况下,双重收缩结构可能导致复合电流信号。在某些情况下,当存在两个这样的读取头时,对一个收缩结构或"读取头"的电流读数可能不能被单独确定。

[0182] 野生型CsgG孔结构可以通过重组基因技术重新设计,以扩大,改变或去除两个收缩结构中的一个,使得在通道内留下单个收缩结构,从而限定单个读取头。CsgG寡聚孔中的收缩结构基序位于野生型大肠杆菌CsgG多肽中38至63位置的氨基酸残基处。该区域的野生型氨基酸序列如SEQ ID NO:3所示。在考虑该区域时,在位置50至53,54至56和58至59的任何氨基酸残基的突变均被视为在本发明的范围内。基于与CsgG同源物的序列相似性(图4),位置38至49,53,57以及61至63的氨基酸残基被认为是高度保守的,因此可能不太适合用于取代或其他修饰。由于Tyr51,Asn55和Phe56侧链在野生型CsgG结构的通道内的关键定位,这些位置的突变可能是有利的,以便修饰或改变读取头的特性。

[0183] 在单体CsgG蛋白指定位置的突变可以导致处于该位置的野生型氨基酸被任意其它天然或非天然氨基酸取代。在本发明的一个实施方式中,希望扩大或除去一个收缩结构;合适地,选择修饰的CsgG蛋白中的氨基酸侧链,使得与其所取代的野生型结构中的氨基酸侧链相比具有较小的空间位阻。指定位置上替换的氨基酸残基可以具有相似的静电性质,或者其也可以具有不同的静电性质。合适地,替换的氨基酸侧链将具有与其所取代的野生型结构中的氨基酸侧链相似的静电电荷,以最小化对二级结构或通道性质的破坏。

[0184] 替换的氨基酸的选择可以基于BLOSUM62矩阵,其提供了基于大量多重序列比对计算氨基酸被另一个取代的可能性的标准方法。BLOSUM62矩阵的实例对技术人员而言可在互联网上免费获得;参见例如国家生物技术信息中心的网站(http:

[0185] //www.ncbi.nlm.nih.gov/Class/Structure/aa/aa_explorer.cgi)。

[0186] 对于野生型CsgG结构中的Tyr51,其在CsgG通道中形成第一收缩结构,提供任意氨基酸的取代。特别地,在本发明的某些实施方式中,Tyr51可被丙氨酸,甘氨酸,缬氨酸,亮氨酸,异亮氨酸,天冬酰胺,谷氨酰胺和苯丙氨酸取代(SEQ ID:39-318)。在本发明的实施方式中,用丙氨酸或甘氨酸对Tyr51的取代是特别合适的(SEQ ID:39-108)。在实施方式中,残基50至53(野生型序列中的PYPA)可以用甘氨酸-甘氨酸(GG)取代(SEQ ID:354-388)。

[0187] 对于Asn55——其有助于所述CsgG通道中的第二收缩结构,提供任意氨基酸的取代。特别地,在本发明的某些实施方式中,Asn 55可以被丙氨酸,甘氨酸,缬氨酸,丝氨酸或苏氨酸取代(SEQ ID:9-33,44-68,79-103,14-14,138,149-173,184-208,219-243,254-278,289-313和324-348)。

[0188] 对于Phe56——其形成所述CsgG通道中的第二收缩结构的一部分,提供任意氨基酸的取代。特别地,在本发明的某些实施方式中,Phe56可以被丙氨酸,甘氨酸,缬氨酸,亮氨酸,异亮氨酸,天冬酰胺和谷氨酰胺取代(SEQ ID:5-13,15-18,20-23,25-28,30-33,40-43,45-48,50-53,55-58,60-63,65-68,75-78,80-83,85-88,90-93,95-98,100-103,110-113,115-118,120-123,125-128,130-133,135-138,145-148,150-153,155-158,160-163,165-

168,170-173,180-183,185-188,190-193,195-198,200-203,205-208,215-218,220-223,225-228,230-233,235-238,240-243,250-253,255-258,260-263,265-268,270-273,275-578,285-288,290-293,295-298,300-303,305-308,310-313,320-323,325-328,330-333,335-338,340-343和345-348)。在本发明的实施方式中,丙氨酸和甘氨酸对Phe56的取代是特别合适的(SEQ ID:5,10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,105,1 10,1 15,120,125,130,135,140,145,150,155,160,165,170,175,180,185,190,195,200,205,210,215,220,225,230,235,240,245,250,255,260,265,270,275,280,285,290,295,300,305,310,315,320,325,330,335,340,345,350,355,360,365,370,375,380,385,6,11,16,21,26,31,36,41,46,51,56,61,66,71,76,81,86,91,96,101,106,111,116,121,126,131,136,141,146,151,156,161,166,171,176,181,186,191,196,201,206,211,216,221,226,231,236,241,246,251,256,261,266,271,276,281,286,291,296,301,306,311,316,321,326,331,336,341 346,351,356,361,366,371,376,381和386)。

[0189] 在给定的突变CsgG蛋白中,Tyr51的取代可以与位置55和56中的任一个同时进行(SEQID:44,49,54,59,64,79,84,89,94,99,1 14,1 19,124,129,134,149,154,159,164,169,184,189,194,199,204,219,224,229,234,239,254,259,264,269,274,289,294,299,304,309,359,364,369,374,379,40,41,42,75,76 77,110,111,112,145,146,147,180,181,182,215,216,217,250,251,252,285,286和287),其中所述通道内至少一个合适尺寸的收缩结构被保持。或者,Tyr51的取代与位置Asn55和Phe56处的取代相互排斥(SEQ ID:39,74,109,144,179,214,249,284,354,10,1 1,12,15,16,17,20,21,22,25,26,27,30,31和32)。

[0190] 或者,野生型CsgG蛋白中Tyr51,Asn55或Phe56中的一个或多个可以缺失(SEQ ID:319-353,34-38,69-73,104-108,139-143,174-178,209-213,244-248,279-283,314-318,384-388,8,13,18,23,28,33,43,48,53,58,63,68,78,83,88,93,98,103,1 13,1 18,123,128,133,138,148,153,158,163,168,173,183,188,193,198,203,208,218,223,228,233,238,243,253,258,263,268,273,278,288,293,298,303,308和313)。为保持通道中至少一个收缩结构,在给定的实施方式中,氨基酸残基Tyr51的缺失与氨基酸残基Asn55和Phe56的缺失相互排斥(SEQID:319-322,324-327,329-332,334-337,339-342,344-347,38,73,108,143,178,213,248,283和318)。位置53和54以及48和49的某些相邻的氨基酸残基也可以缺失。

[0191] 应当理解,本发明提供的实施方式,其中上述修饰可以单独或以任何组合方式进行。

[0192] 除去在Tyr51的收缩结构或在Asn55/Tyr56的收缩结构,导致CsgG通道内的单个收缩结构。不希望被理论束缚,假设Asn55/Tyr56的收缩结构与可能是期望的在Tyr51的收缩结构相比具有更高的构象稳定性。然而,与核苷酸相比,Asn55/Tyr56收缩结构可能太高(沿中心孔的轴测量)。这可能导致DNA链移位中各个碱基对的分辨率差。

[0193] 对于Tyr51收缩结构,则可能是相反的。除去Asn55/Tyr56收缩结构后,寡聚物中剩下的Tyr51残基的环可能与天然结构相比在构象上不太稳定。然而,Tyr51收缩结构较短(当沿着中心孔的轴测量时),并且可能更能够在通道内提供可区分各个碱基的收缩结构。

[0194] 在任一实施方式中,当在核酸测序应用中使用所述孔时,单个狭窄收缩结构的存

在(当沿中心孔轴测量时)可能降低电流读取的复杂性。核酸通过所述孔移位时产生的观测电流的调节将因此仅反映通过单个收缩结构或读数头的分离的核苷酸的通过。

[0195] 一个收缩结构的有效去除也可以增加孔突变体的开放通道电流。增加的开放通道电流将是有利的,因为较高的背景电导导致针对不同核酸碱基对信号的更高分辨率的电流阻滞水平。采用这种方式,对读取头的修饰可以改善生物孔对核酸测序和其他分子传感应用的适用性。

[0196] 作为替代的实施方式,或除了上述序列修饰之外,还提供了,Asn55/Phe56收缩结构可进一步适于调整其高度(沿中心孔轴测量)。Asn55/Phe56收缩结构的这种进一步调整可能伴随着或未伴随着Tyr51或CsgG通道内其他位置的突变。合适地,Asn55/Phe56收缩结构的进一步调整考虑作为扩大或去除由Tyr51残基形成的收缩结构的突变的一部分。

[0197] 在野生型形式中,Asn55/Phe56通道收缩结构由两个彼此垂直相邻而设的氨基酸环组成。因此,所述收缩结构具有大于1nm的长度。1nm长的收缩结构可能不允许分辨对核酸链移位中分离碱基的离子流产生的电信号。通常地,用于核酸测序的已知纳米孔的所述收缩结构一般长度小于1nm。例如,用于DNA测序的MspA纳米孔具有0.6nm的收缩结构高度(沿中心孔轴测量;Manrao et al.,Nature Biotechnology,2012,30(4),349-353)。

[0198] 为降低Asn55/Phe56收缩结构的高度(沿中心孔轴测量),两个残基中的任一个可以被取代或缺失,导致收缩结构顶部或底部的扩大。

[0199] 对于形成CsgG通道中第二收缩结构一部分的Asn55,可以考虑用任意氨基酸取代。特别地,用丙氨酸,甘氨酸,缬氨酸,丝氨酸或苏氨酸取代(SEQ ID:9-33,44-68,79-103,14-14,138,149-173,184-208,219-243,254-278,289-313和324-348)。

[0200] 对于形成CsgG通道中第二收缩结构一部分的Phe56,可以考虑用任意氨基酸取代。特别地,用丙氨酸,甘氨酸,缬氨酸,亮氨酸,异亮氨酸,天冬酰胺和谷氨酰胺取代(SEQ ID:5-13,15-18,20-23,25-28,30-33,40-43,45-48,50-53,55-58,60-63,65-68,75-78,80-83,85-88,90-93,95-98,100-103,110-113,115-118,120-123,125-128,130-133,135-138,145-148,150-153,155-158,160-163,165-168,170-173,180-183,185-188,190-193,195-198,200-203,205-208,215-218,220-223,225-228,230-233,235-238,240-243,250-253,255-258,260-263,265-268,270-273,275-578,285-288,290-293,295-298,300-303,305-308,310-313,320-323,325-328,330-333,335-338,340-343和345-348)。在本发明的实施方式中,丙氨酸和甘氨酸对Phe56的取代是特别合适的(SEQ ID:5,10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,105,1 10,1 15,120,125,130,135,140,145,150,155,160,165,170,175,180,185,190,195,200,205,210,215,220,225,230,235,240,245,250,255,260,265,270,275,280,285,290,295,300,305,310,315,320,325,330,335,340,345,350,355,360,365,370,375,380,385,6,11,16,21,26,31,36,41,46,51,56,61,66,71,76,81,86,91,96,101,106,111,116,121,126,131,136,141,146,151,156,161,166,171,176,181,186,191,196,201,206,211,216,221,226,231,236,241,246,251,256,261,266,271,276,281,286,291,296,301,306,311,316,321,326,331,336,341 346,351,356,361,366,371,376,381和386)。

[0201] 用于调整野生型CsgG孔的收缩结构的最小直径的修饰也是可以想到的。CsgG孔中两个收缩结构的最小直径约为0.9nm(9Å),其直径小于已知有效用于DNA测序的MspA纳米

孔中收缩结构的1.2nm直径(Manrao et al.,Nature Biotechnology,2012,30(4),349-353)。能在最小直径为0.5至1.5nm的修饰的CsgG孔中提供剩余的收缩结构的任意上述突变是合适的。

[0202] 上述列出的任意修饰可以有利地改变收缩结构处的氨基酸的亲水性和电荷分布,以改善移位核酸链的通过和与移位核酸链的非共价相互作用,从而改善电流读出。上述列出的任意突变也可以有利地改变靠近通道收缩结构的亲水性和电荷分布,以优化通过所述收缩结构的电解质离子的流动,并且实现更好的区别移位核酸链中通过的核苷酸。

[0203] CsgG蛋白的进一步修饰被认为可能导致改变通道腔内的表面电荷分布。在本发明的一个实施方式中,进行上述这些修饰可以避免移位核酸不希望地静电吸附到通道壁上。由于核酸带有负电荷,并且CsgG通道内腔含有一些正电荷(图1),假定静电相互作用可能干扰核酸测序时的螺旋或移位。合适地带正电荷的氨基酸残基例如赖氨酸,组氨酸和精氨酸,可以被中性或带负电荷的侧链取代,以进一步提高核酸移位通过孔的效率,从而提高电流读数的清楚度。

[0204] 本发明的实施方式还提供了对野生型大肠杆菌CsgG或突变CsgG的通道内腔进行修饰以促进核酸链(或各个核苷酸)移位或螺旋进入所述孔收缩结构。具有内部收缩结构的CsgG跨膜部分类似于带有以中心孔为特征的盖状结构的桶状结构。通过在最靠近所述桶状结构和盖状结构的孔内腔中增加额外的环,可以促进螺旋进入过程。

[0205] 本发明具体实施方式提供了,所述CsgG孔可以由共价连接的一种或多种单体,二聚体或寡聚物组成。作为非限制性实施例,单体可以以任意构象进行基因融合,例如通过它们的末端氨基酸。在这种情况下,一个单体的氨基末端可以与另一个单体的羧基末端融合。

[0206] 根据本发明的一个实施方式,还提供了,所述CsgG孔可以适应于容纳额外的辅助蛋白,该辅助蛋白可以在分子从所述孔通过时具有有利的性质。对所述孔的适应性可能有助于核酸加工酶的锚定。核酸加工酶可以包括DNA或RNA聚合酶;异构酶;拓扑异构酶;促旋酶;端粒酶;和解旋酶。将这些酶中的一种或多种与纳米孔相关联可以在增强核酸螺旋进入所述孔以及控制核酸链移位通过孔的速度方面具有益处(Manrao et al.,Nature Biotechnology,2012,30(4),349-353)。控制核酸链通过孔的移位速度具有以下优点:在更适合读取且更一致的离子流的电流测量方面提供改进的响应。

[0207] 在本发明的实施方式中,设想对纳米孔膜外区域的修饰可能有助于通过提供一个或多个结合/锚定位点促进合适的核酸加工酶(例如DNA-聚合酶)在通道的内腔中或邻近通道内腔位置的对接(docking)。合适的锚定位点可以包括用于酶静电结合的静电区块(electrostatic patches);允许共价偶联的一个或多个半胱氨酸残基;和/或用于提供空间锚的跨膜通道部分的改变的内部宽度。

[0208] 本发明在下面更详细列出的具体实施方式中提供用于核酸测序的CsgG野生型孔的进一步适应性变化。

[0209] 在本发明的实施方式中,所述CsgG孔的膜外区域(如图1所示的底部)可以被截短或除去,以促进核酸链在所述通道内腔的另一侧上的离去。膜外区域的截短或除去也可以改善来自通道中离子流的电信号的电流分辨率。后者的益处是通过由膜外区域降低对离子流的阻力而带来的。在本发明的CsgG孔中,跨膜通道、内部收缩结构和盖帽区域代表连续的三个阻力区域。去除或降低上述区域之一的贡献会增加开放通道电流,从而提高电流分辨

率。这种改变可以包括缺失α-螺旋2(α2),C-末端α-螺旋(αC)(图1),和/或它们的组合。

[0210] 在本发明的进一步实施方式中,面向在野生型大肠杆菌CsgG孔的外表面上的氨基酸的膜可以被修饰以促进所述孔插入到所述膜中。在某些实施方式中,提供了,单个氨基酸取代可以用合适的更疏水的类似物替代野生型残基。例如,残基Ser136,Gly138,Gly140,Ala148,Ala188或Gly202中的一个或多个可以改变为Ala,Val,Leu或IIe。此外,芳香类残基如酪氨酸或色氨酸可以取代适当的野生型氨基酸,使得它们位于疏水膜和亲水溶剂的界面处。例如,残基Leu 154或Leu182中的一个或多个可被Tyr,Phe或Trp取代。

[0211] 在本发明的实施方式中,可以增加所述蛋白孔的热稳定性。这导致测序装置中纳米孔的保质期的有利增加。在实施方式中,蛋白孔的热稳定性的增加通过修饰β-折叠序列(beta-turn sequences)或改善蛋白表面的静电相互作用实现。在一种实施方式中,跨膜区域中的β-发夹结构可以通过在相邻β-发夹结构中或相邻β-发夹结构之间跨两个相邻β-链形成共价二硫化物进行稳定。这种交叉链半胱氨酸对的实例可以是Val139-Asp203;Gly139-Gly205;Lys135-Thr207;Glu201-Ala141;Gly147-Gly189;Asp149-Gln187;Gln151-Glu185;

[0212] Thr207-Glu185;Gly205-Gln187;Asp203-Gly189;Ala153-Lys135;Gly137-Gln151;

[0213] Val139-Asp149或Ala141-Gly147。

[0214] 在本发明的进一步实施方式中,根据Biotechnol.J.,201 1,6(6),650-659中描述的方法,用于多肽序列的密码子可以被修饰以允许以高水平和低错误率表达CsgG蛋白。所述修饰也可以靶向mRNA的任何二级结构。

[0215] 本发明还提供蛋白序列的改变以改善其对蛋白酶的稳定性。这可以通过去除柔性环区域实现,例如缺失α-螺旋2(α2),C-末端α-螺旋(αC)(图1),和/或其组合。

[0216] 在本发明的实施方式中,改变所述CsgG多肽序列/表达系统以避免蛋白质可能的聚集。

[0217] 本发明还提供多肽序列的改变以提高蛋白质的折叠效率。合适的技术提供在Biotechnol.J.,2011,6(6),650-659中。

[0218] 本发明的一种实施方式进一步提供生物亲和性标签的替换或增加以便于所述CsgG蛋白的纯化。公开的所述CsgG孔的结构包含Strepll标签(Goyal et al.,Nature,2014,516(7530),250-3)。本发明的实施方式包括其它生物亲和性标签,例如用于促进通过金属螯合亲和色谱纯化的组氨酸标签。在本发明的替代实施方式中,所述标签可以包括FLAG标签或表位标签,例如Myc-标签或HA-标签。在本发明的进一步实施方式中,所述纳米孔可以通过采用生物素或其类似物(例如脱硫生物素)的生物素化进行修饰,从而有助于通过与链霉亲和素的相互作用进行纯化。

[0219] 在本发明的实施方式中,可以添加在蛋白质末端的负电荷以增加所述多肽的净电荷并促进蛋白质在聚丙烯酰胺凝胶电泳中的迁移。这可能导致在对这些种类感兴趣的情况下对CsgG异源寡聚体的改进的分离(参见Howorka et al.Proc.Nat.Acad.Sci.,2001,98(23),12996-13001)。异源寡聚体可用于对每孔引入一个单半胱氨酸残基,这可以有利于促进合适的核酸加工酶的连接,例如如上所述的DNA聚合酶。

[0220] 本发明的一部分涉及野生型或修饰的CsgG纳米孔基于流过纳米孔电流的电学测

量值的变化在分子感测应用中的用途。

[0221] 分子在所述CsgG孔的通道中或在所述通道任一开口附近的结合对通过所述孔的开放通道离子流有影响。以与上述核酸测序应用相似的方式,可以使用合适的测量技术通过电流变化来测量开放通道离子流的变化(例如,WO 2000/28312和D.Stoddart et al,Proc.Natl.Acad.Sci.,2010,106,7702-7或WO 2009/077734)。离子流的减少程度(通过电流的减少测量)与孔内或孔附近的阻塞尺寸有关。因此感兴趣的分子,也称为分析物,在孔内或孔附近的结合提供了可检测和可测量的事件,从而形成生物传感器的基础。用于纳米孔感应的合适分子包括核酸;蛋白质;肽类和小分子如药物,毒素或细胞因子。

[0222] 对生物分子存在的检测可以应用于个性化药物开发,医药,诊断,生命科学研究,环境监测以及安全和/或国防工业。

[0223] 在本发明的实施方式中,本申请公开的野生型或修饰的大肠杆菌CsgG纳米孔或其同源物可用作分子传感器。分析物检测过程描述于Howorka et al.Nature Biotechnology(2012)Jun 7;30(6)506-7。待检测的所述分析物分子可以与通道的任意一面结合,或者结合在通道本身的内腔内。结合位置可以通过待感测分子的大小确定。野生型CsgG孔可以作为传感器,或本发明的实施方式,通过重组或化学方法来修饰CsgG孔,以增加结合强度,结合位置或待感测分子的结合特异性。通常的修饰包括加入与待感测分子的结构互补的特异性结合部分。当分析物分子包含核酸时,该结合部分可以包含环糊精或寡核苷酸;对于小分子,这可以是已知的互补结合区域,例如抗体或非抗体分子的抗原结合部分,包括来自T细胞受体(TCR)的单链可变片段(scFv)区域或抗原识别结构域;或对于蛋白质,其可以是目标蛋白质的已知配体。以这种方式,野生型或修饰的大肠杆菌CsgG纳米孔或其同源物可以表现出能够用作分子传感器,用于检测样品中可以包括细胞表面抗原的合适抗原(包括表位)的存在,包括受体,实体肿瘤或血液癌细胞标志物(如淋巴瘤或白血病),病毒抗原,细菌抗原,原生动物抗原,过敏源,过敏相关分子,白蛋白(如人,啮齿动物或牛),荧光分子(包括荧光素),血液组抗原,小分子,药物,酶,酶或酶底物的催化位点,以及酶底物的过渡态类似物。

[0224] 可以使用已知的基因工程和重组DNA技术来实现修饰。任何适应的定位将取决于待感测分子的性质,例如尺寸,三维结构及其生物化学性质。适应性结构的选择可以利用计算机结构设计。一系列定制的CsgG纳米孔被设想出来,每个都特异性适用于其注定的感测应用。蛋白-蛋白相互作用或蛋白质-小分子相互作用的测定和优化可以使用例如BIAcore®等技术进行研究,所述技术使用表面等离子体共振来检测分子相互作用(BIAcore,Inc.,Piscataway,NJ;也参见www.biacore.com)。

[0225] 在一个实施方式中,所述CsgG孔可以是水溶性的八聚体形式,其中N端Cys残基被另一氨基酸替代,以防止蛋白质N端的脂化。在替代实施方式中,蛋白质可以通过除去N-端前导序列在细胞质中表达,以避免通过细菌脂质化途径的加工。

[0226] Goyal等人(Nature,2014;516(7530)250-3)描述了CsgG单体可溶性蛋白,八聚体可溶性蛋白和寡聚脂质化CsgG孔的制备方法,该文献通过援引加入本申请以及实施例1和2中。

[0227] 突变CsgG单体

[0228] 本发明提供突变型CsgG单体。突变CsgG单体可用于形成本发明所述的孔。突变

CsgG单体是其序列从野生型CsgG单体的序列变化而言且保留形成孔的能力的单体。用于确认突变单体形成孔的能力的方法在本领域中是公知的,下面将更详细地讨论。

[0229] 所述突变单体具有提高的多核苷酸读取性质,即显示改进的多核苷酸捕获和核苷酸识别。特别地,由突变单体构建的孔比野生型更容易捕获核苷酸和多核苷酸。此外,由突变单体构建的孔显示增加的电流范围(这使得更容易区分不同的核苷酸)和减少的状态变化(这增加了信噪比)。此外,随着多核苷酸移动通过由突变体构建的孔,贡献于电流的核苷酸数目减少。这使得更容易鉴定当多核苷酸移动通过所述孔时观察到的电流与多核苷酸序列之间的直接关系。此外,由突变单体构建的孔,可以显示增加的通量,即更可能与分析物例如多核苷酸相互作用。这使得使用所述孔表征分析物更容易。由突变单体构建的孔可以更容易地插入膜中。

[0230] 本发明所述突变单体包括SEQ ID NO:390所示序列的变体。SEQ ID NO:390是来自大肠杆菌Str.K-12亚株MC4100的野生型CsgG单体。SEQ ID NO:390的变体是具有与SEQ ID NO:390的序列不同的氨基酸序列并保留其形成孔的能力的多肽。可以使用本领域已知的任何方法测定变体形成孔的能力。例如,可以将变体与其他合适的亚基一起插入两亲性分子层,并且可以确定变体寡聚而形成孔的能力。将亚基插入膜,例如双亲性分子层,中的方法是本领域已知的。例如,亚基可以在含有三嵌段共聚物膜的溶液中以纯化的形式悬浮,使得其扩散到所述膜并通过结合到膜上并组装成功能状态而插入。

[0231] 在本申请的所有讨论中,对氨基酸使用标准的单字母代码。这些代码如下: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),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)和缬氨酸(V)。还使用标准取代符号,即Q42R表示42位置的Q被R替换。

[0232] 在一种实施方式中,本发明的突变单体包括SEQ ID NO:390的突变体,其包含以下中的一个或多个:(i)在以下位置的一个或多个突变(即在一个或多个以下位置的突变):N40,D43,E44,S54,S57,Q62,R97,E101,E124,E131,R142,T150和R192,例如以下位置的一个或多个突变(即在一个或多个以下位置的突变):N40,D43,E44,S54,S57,Q62,E101,E131和T150或N40,D43,E44,E101和E131;(ii)在Y51N55,Y51F56,N55F56或Y51N55F56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)缺失以下一个或多个位置:F48,K49,P50,Y51,P52,A53,S54,N55,F56和S57。所述变体可以包含(i)至(xi)的任意组合。特别地,所述变体可以包括{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}{vii,xi}{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,vii,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}{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,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,x}{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,x}{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,viii,x,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,x}{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,x,xi}{ii,vi,ix,x,xi}{ii,vii,viii,ix,x}{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,x}{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,x,xi}{iii,vi,ix,x,xi}

{iii,vii,viii,ix,x}{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,x{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,vii,viii,ix,x}{i,iv,vii,viii,ix,xi}{i,iv,vii,viii,x,xi}{i,iv,vii,ix,x,xi}{i,iv,viii,ix,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,xr{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,vi,vii,ix,x,xi}{i,vi,viii,ix,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,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,viii,x,xi}{v,vi,vii,ix,x,xi v,vi,viii,ix,x,xi}{v,vii,viii,ix,x,xi}{vi,vii,viii,ix,x,xi}{i,ii,iii,iv,v,vi,vii}{i,ii,iii,iv,v,vi,viii}{i,ii,iii,iv,v,vi,ixi,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,xii,iv,v,viii,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,xii,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,xi,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,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,xii,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,xii,v,vi,viii,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,xii,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,xi,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,xii,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,xii,viii,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,ixiv,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,xiv,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}{i,ii,iv,vi,vii,viii,x}{i,ii,iv,vi,vii,viii,xiiv,vi,vii,ix,x}{i,ii,iv,vi,vii,ix,xi}{i,ii,iv,vi,vii,x,xi}{i,ii,iv,vi,viii,ix,x}{i,ii,iv,vi,viii,ix,xi}{i,ii,iv,vi,viii,x,xiiv,vi,ix,x,xi}{i,ii,iv,vii,viii,ix,x}{i,ii,iv,vii,viii,ix,xi}{i,ii,iv,vii,viii,x,xi}{i,ii,iv,vii,ix,x,xi}{i,ii,iv,viii,ix,x,xi}{i,ii,v,vi,vii,viii,ix}{i,ii,v,vi,vii,viii,x}{i,ii,v,vi,vii,viii,xi}{i,ii,v,vi,vii,ix,x}{i,ii,v,vi,vii,ix,xi}{i,ii,v,vi,vii,x,xiv,vi,viii,ix,x}{i,ii,v,vi,viii,ix,xi}{i,ii,v,vi,viii,x,xi}{i,ii,v,vi,ix,x,xi}{i,ii,v,vii,viii,ix,x}{i,ii,v,vii,viii,ix,xiv,vii,viii,x,xi}{i,ii,v,vii,ix,x,xi}{i,ii,v,viii,ix,x,xi}{i,ii,vi,vii,viii,ix,x}{i,ii,vi,vii,viii,ix,xi}{i,ii,vi,vii,viii,x,xivi,vii,ix,x,xi}{i,ii,vi,viii,ix,x,xi}{i,ii,vii,viii,ix,x,xi}{i,iii,iv,v,vi,vii,viii}{i,iii,iv,v,vi,vii,ix}{i,iii,iv,v,vi,vii,x}{i,iii,iv,v,vi,vii,xi}{i,iii,iv,v,vi,viii,ix}{i,iii,iv,v,vi,viii,x}{i,iii,iv,v,vi,viii,xi}{i,iii,iv,v,vi,ix,x}{i,iii,iv,v,vi,ix,xi,iv,v,vi,x,xi}{i,iii,iv,v,vii,viii,ix}{i,iii,iv,v,vii,viii,x}{i,iii,iv,v,vii,viii,xi}{i,iii,iv,v,vii,ix,x}{i,iii,iv,v,vii,ix,xi,iv,v,vii,x,xi}{i,iii,iv,v,viii,ix,x}{i,iii,iv,v,viii,ix,xi}{i,iii,iv,v,viii,x,xi}{i,iii,iv,v,ix,x,xi}{i,iii,iv,vi,vii,viii,ix,iv,vi,vii,viii,x}{i,iii,iv,vi,vii,viii,xi}{i,iii,iv,vi,vii,ix,x}{i,iii,iv,vi,vii,ix,xi}{i,iii,iv,vi,vii,x,xi}{i,iii,iv,vi,viii,ix,x}{i,iii,iv,vi,viii,ix,xi}{i,iii,iv,vi,viii,x,xi}{i,iii,iv,vi,ix,x,xi}{i,iii,iv,vii,viii,ix,x}{i,iii,iv,vii,viii,ix,xi}{i,iii,iv,vii,viii,x,xi}{i,iii,iv,vii,ix,x,xi}{i,iii,iv,viii,ix,x,xi}{i,iii,v,vi,vii,viii,ix}{i,iii,v,vi,vii,viii,x}{i,iii,v,vi,vii,viii,xi}{i,iii,v,vi,vii,ix,x}{i,iii,v,vi,vii,ix,xi}{i,iii,v,vi,vii,x,xi}{i,iii,v,vi,viii,ix,x}{i,iii,v,vi,viii,ix,xi}{i,iii,v,vi,viii,x,xi}{i,iii,v,vi,ix,x,xi}{i,iii,v,vi,ix,x,xi,v,vii,viii,ix,x}{i,iii,v,vii,viii,ix,xi}{i,iii,v,vii,viii,x,xi}{i,iii,v,vii,ix,x,xi}{i,iii,v,viii,ix,x,xi}{i,iii,vi,vii,viii,ix,x}{i,iii,vi,vii,viii,ix,xi}{i,iii,vi,vii,viii,x,xi}{i,iii,vi,vii,ix,x,xi}{i,iii,vi,viii,ix,x,xi}{i,iii,vii,viii,ix,x,xi}{i,iv,v,vi,vii,viii,ix}{i,iv,v,vi,vii,viii,x}{i,iv,v,vi,vii,viii,xi}{i,iv,v,vi,vii,ix,x}{i,iv,v,vi,vii,ix,xi}{i,iv,v,vi,vii,x,xi}{i,iv,v,vi,viii,ix,x}{i,iv,v,vi,viii,ix,xi}{i,iv,v,vi,viii,x,xi}{i,iv,v,vi,ix,x,xi}{i,iv,v,vii,viii,ix,x}{i,iv,v,vii,viii,ix,xi}{i,iv,v,vii,viii,x,xi}{i,iv,v,vii,ix,x,xi}{i,iv,v,viii,ix,x,xi}{i,iv,vi,vii,viii,ix,x}{i,iv,vi,vii,viii,ix,xi}{i,iv,vi,vii,viii,x,xi}{i,iv,vi,vii,ix,x,xiv,vi,viii,ix,x,xi}{i,iv,vii,viii,ix,x,xi}{i,v,vi,vii,viii,ix,x}{i,v,vi,vii,viii,ix,xi}{i,v,vi,vii,viii,x,xi}{i,v,vi,vii,ix,x,xi}{i,v,vi,viii,ix,x,xi}{i,v,vii,viii,ix,x,xi}{i,vi,vii,viii,ix,x,xi}{ii,iii,iv,v,vi,vii,viii}{ii,iii,iv,v,vi,vii,ix}{ii,iii,iv,v,vi,vii,x}{ii,iii,iv,v,vi,vii,xi}{ii,iii,iv,v,vi,viii,ix}{ii,iii,iv,v,vi,viii,x}{ii,iii,iv,v,vi,viii,xi}{ii,iii,iv,v,vi,ix,x}{ii,iii,iv,v,vi,ix,x}{ii,iii,iv,v,vi,x,xi}{ii,

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

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

[0233] 如果所述变体包括(i)和(iii)至(xi)中的任意一种,其可以进一步包括位于Y51,

N55和F56中的一个或多个位置的突变,例如位于Y51,N55,F56,Y51N55,Y51/F56,N55/F56或Y51/N55/F56。

[0234] 在(i)中,所述变体可以包括N40,D43,E44,S54,S57,Q62,R97,E101,E124,E131,R142,T150和R192中任意数量和组合的突变。在(i)中,所述变优选包括位于以下位置的一个或多个突变(即在以下一个或多个位置进行突变):N40,D43,E44,S54,S57,Q62,E101,E131和T150。在(i)中,所述变优选包括位于以下位置的一个或多个突变(即在一个或多个以下位置的突变):N40,D43,E44,E101和E131。在(i)中,所述变体优选包括位于S54和/或S57的突变。在(i)中,所述变体更优选包括位于(a)S54和/或S57和(b)Y51,N55和F56中的一个或多个,例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56的突变。如果在(xi)中缺失S54和/或S57,则它/它们不能在(i)中突变,反之亦然。在(i)中,所述变体优选包括位于T150,例如T150I的突变。或者所述变体优选包括位于(a)T150和(b)Y51,N55和F56中的一个或多个,例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56的突变。在(i)中,所述变体优选包括位于Q62,例如Q62R或Q62K的突变。或者所述变体优选包括位于(a)Q62和(b)Y51,N55和F56中的一个或多个,例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56的突变。所述变体可以包括位于D43,E44,Q62或其任意组合的突变,例如D43,E44,Q62,D43/E44,D43/Q62,E44/Q62或D43/E44/Q62。或者所述变体优选包括位于(a)D43,E44,Q62,D43/E44,D43/Q62,E44/Q62或D43/E44/Q62和(b)Y51,N55和F56中的一个或多个,例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56的突变。

[0235] 在本申请的(ii)和其他地方中,符号/表示“和”,例如Y51/N55是指Y51和N55。在(ii)中,所述变体优选包括位于Y51/N55的突变。已经提出,CsgG中的所述收缩结构由三个由残基Y51,N55和F56的侧链形成的堆叠的同心环组成(Goyal et al,2014,Nature,516,250-253)。在(ii)中这些残基的突变因而可以减少当多核苷酸移动穿过所述孔时对电流有贡献的核苷酸的数量,从而使得更容易识别观察到的电流(当多核苷酸移动通过所述孔时)与多核苷酸之间的直接关系。Y56可以以下文关于本发明方法中可用变体和孔而论述的任何方式进行突变。

[0236] 在(v)中,所述变体可以包括N102R,N102F,N102Y或N102W。所述变体优选包括(a)N102R,N102F,N102Y或N102W和(b)位于Y51,N55和F56中的一个或多个位置例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56的突变。

[0237] 在(xi)中,任意数量和组合的以下位置:K49,P50,Y51,P52,A53,S54,N55,F56和S57可以缺失。优选地,K49,P50,Y51,P52,A53,S54,N55和S57中的一个或多个可以缺失。如果Y51,N55和F56中的任意一个在(xi)中缺失,则它/它们不能在(ii)中突变,反之亦然。

[0238] 在(i)中,所述变体优选包括以下取代中的一个或多个: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,例如N40R,N40K,D43N,D43Q,D43R,D43K,E44N,E44Q,E44R,E44K,S54P,S57P,Q62R,Q62K,E101N,E101Q,E101R,E101K,E101F,E101Y,E101W,E131D和T150I中的一个或多个,或N40R,N40K,D43N,D43Q,D43R,D43K,E44N,E44Q,E44R,E44K,E101N,E101Q,E101R,E101K,E101F,E101Y,E101W和E131D中的一个或多个。所述变体可以包括任意数量的这些取代和这些取代的任意组合。在(i)中,所述变体优选包括S54P

和/或S57P。在(i)中,所述突变体优选包括(a)S54P和/或S57P和(b)位于Y51,N55和F56中的一个或多个位置的突变,例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56。位于Y51,N55和F56中一个或多个位置的所述突变可以是如下所述突变中的任意一种。在(i)中,所述变体优选包括F56A/S57P或者S54P/F56A。所述变体优选包括T150I。或者所述变体优选包括位于(a)T150I和(b)Y51,N55和F56中的一个或多个,例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56的突变。

[0239] 在(i)中,所述变体优选包括Q62R或Q62K。或者所述变体优选包括(a)Q62R或Q62K和(b)位于Y51,N55和F56中的一个或多个位置的突变,例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56。所述变体可以包括D43N,E44N,Q62R或Q62K或其任意的组合,例如D43N,E44N,Q62R,Q62K,D43N/E44N,D43N/Q62R,D43N/Q62K,E44N/Q62R,E44N/Q62K,D43N/E44N/Q62R或D43N/E44N/Q62K。或者所述变体优选包括(a)D43N,E44N,Q62R,Q62K,D43N/E44N,D43N/Q62R,D43N/Q62K,E44N/Q62R,E44N/Q62K,D43N/E44N/Q62R或D43N/E44N/Q62K和(b)位于Y51,N55和F56中的一个或多个位置例如位于Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56的突变。

[0240] 在(i)中,所述变体优选包括D43N。

[0241] 在(i)中,所述变体优选包括E101R,E101S,E101F或E101N。

[0242] 在(i)中,所述变体优选包括E124N,E124Q,E124R,E124K,E124F,E124Y,E124W或E124D,例如E124N。

[0243] 在(i)中,所述变体优选包括R142E和R142N。

[0244] 在(i)中,所述变体优选包括R97N,R97G或者R97L。

[0245] 在(i)中,所述变体优选包括R192E和R192N。

[0246] 在(ii)中,所述变体优选包括F56N/N55Q,F56N/N55R,F56N/N55K,F56N/N55SF56N/N55G,F56N/N55A,F56N/N55T,F56Q/N55Q,F56Q/N55R,F56Q/N55K,F56Q/N55SF56Q/N55G,F56Q/N55A,F56Q/N55T,F56R/N55Q,F56R/N55R,F56R/N55K,F56R/N55SF56R/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/Y51NF56G/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/T51A,N55A/T51N,N55A/Y51Q,N55AY51S N55A/Y51G,N55T/Y51L,N55T/Y51V,N55T/T51A,N55T/T51N,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/Y51S,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/T51N,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/Y51S,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/Y51S,F56R/N55G/Y51G;F56R/N55A/Y51L,F56R/N55A/Y51V,F56R/N55A/Y51A,F56R/N55A/Y51N,F56R/N55A/Y51Q,F56R/N55A/Y51S,F56R/N55A/Y51GF56R/N55T/Y51L,F56R/N55T/Y51V,F56R/N55T/Y51A,F56R/N55T/Y51N,F56R/N55T/Y51QF56R/N55T/Y51S,F56R/N55T/Y51G,F56S/N55Q/Y51L,F56S/N55Q/Y51V,F56S/N55Q/Y51AF56S/N55Q/Y51N,F56S/N55Q/Y51Q,F56S/N55Q/Y51S,F56S/N55Q/Y51G,F56S/N55R/Y51LF56S/N55R/Y51V,F56S/N55R/Y51A,F56S/N55R/Y51N,F56S/N55R/Y51Q,F56S/N55R/Y51SF56S/N55R/Y51G,F56S/N55K/Y51L,F56S/N55K/Y51V,F56S/N55K/Y51A,F56S/N55K/Y51NF56S/N55K/Y51Q,F56S/N55K/Y51S,F56S/N55K/Y51G,F56S/N55S/Y51L,F56S/N55S/Y51VF56S/N55S/Y51A,F56S/N55S/Y51N,F56S/N55S/Y51Q,F56S/N55S/Y51S,F56S/N55S/Y51GF56S/N55G/Y51L,F56S/N55G/Y51V,F56S/N55G/Y51A,F56S/N55G/Y51N,F56S/N55G/Y51QF56S/N55G/Y51S,F56S/N55G/Y51G,F56S/N55A/Y51L,F56S/N55A/Y51V,F56S/N55A/

Y51AF56S/N55A/Y51N,F56S/N55A/Y51Q,F56S/N55A/Y51S,F56S/N55A/Y51G,F56S/N55T/Y51LF56S/N55T/Y51V,F56S/N55T/Y51A,F56S/N55T/Y51N,F56S/N55T/Y51Q,F56S/N55T/Y51SF56S/N55T/Y51G,F56G/N55Q/Y51L,F56G/N55Q/Y51V,F56G/N55Q/Y51A,F56G/N55Q/Y51NF56G/N55Q/Y51Q,F56G/N55Q/Y51S,F56G/N55Q/Y51G;F56G/N55R/Y51L,F56G/N55R/Y51VF56G/N55R/Y51A,F56G/N55R/Y51N,F56G/N55R/Y51Q,F56G/N55R/Y51S,F56G/N55R/Y51GF56G/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/Y51S,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/Y51 V,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/Y51S,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。

[0247] 在(ii)中,所述变体优选包括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。

[0248] 在(ii)中,所述变体优选包括Y51T/F56Q,Y51Q/F56Q或者Y51A/F56Q。

[0249] 在(ii)中,所述变体优选包括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 or Y51T/F56R。

[0250] 在(ii)中,所述变体优选包括Y51T/N55Q,Y51T/N55S或者Y51T/N55A。

[0251] 在(ii)中,所述变体优选包括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。

[0252] 在(ii)中,所述变体优选包括Y51C/F56A,Y51E/F56A,Y51D/F56A,Y51K/F56A,Y51H/F56A,Y51Q/F56A,Y51N/F56A,Y51S/F56A,Y51P/F56A或Y51V/F56A。

[0253] 在(xi)中,所述变体优选包括Y51/P52,Y51/P52/A53,P50至P52,P50至A53,K49至Y51,K49至A53的缺失和采用单脯氨酸(P),K49至S54进行的替换和采用单P,Y51至A53,Y51至S54,N55/F56,N55至S57,N55/F56进行的替换,和采用单P,N55/F56进行的替换和采用单甘氨酸(G),N55/F56进行的替换和采用单丙氨酸(A),N55/F56进行的替换和采用P和Y51N,N55/F56进行的替换和采用单P和Y51Q,N55/F56进行的替换和采用单P和Y51S,N55/F56进行的替换和采用单甘氨酸(G)和Y51N,N55/F56进行的替换和采用单G和Y51Q,N55/F56进行的替换和采用单G和Y51S,N55/F56进行的替换和采用单丙氨酸(A)和Y51N,N55/F56进行的替换和采用单丙氨酸(A),Y51Q或N55/F56进行的替换和采用单丙氨酸(A)和Y51S进行的替换。

[0254] 所述变体更优选包括D195N/E203N,D195Q/E203N,D195N/E203Q,

[0255] D195Q/E203Q,E201N/E203N,E201Q/E203N,E201N/E203Q,E201Q/E203Q,E185N/E203Q,

[0256] E185Q/E203Q,E185N/E203N,E185Q/E203N,D195N/E201N/E203N,D195Q/E201N/E203N,

[0257] D195N/E201Q/E203N,D195N/E201N/E203Q,D195Q/E201Q/E203N,D195Q/E201N/E203Q,

[0258] D195N/E201Q/E203Q,D195Q/E201Q/E203Q,D149N/E201N,D149Q/E201N,D149N/E201Q,

[0259] D149Q/E201Q,D149N/E201N/D195N,D149Q/E201N/D195N,D149N/E201Q/D195N,

[0260] D149N/E201N/D195Q,D149Q/E201Q/D195N,D149Q/E201N/D195Q,D149N/E201Q/D195Q,

[0261] D149Q/E201Q/D195Q,D149N/E203N,D149Q/E203N,D149N/E203Q,D149Q/E203Q,

[0262] D149N/E185N/E201N,D149Q/E185N/E201N,D149N/E185Q/E201N,D149N/E185N/E201Q,

[0263] D149Q/E185Q/E201N,D149Q/E185N/E201Q,D149N/E185Q/E201Q,D149Q/E185Q/E201Q,

[0264] D149N/E185N/E203N,D149Q/E185N/E203N,D149N/E185Q/E203N,D149N/E185N/E203Q,

[0265] D149Q/E185Q/E203N,D149Q/E185N/E203Q,D149N/E185Q/E203Q,D149Q/E185Q/E203Q,

[0266] D149N/E185N/E201N/E203N,D149Q/E185N/E201N/E203N,D149N/E185Q/E201N/E203N,

[0267] D149N/E185N/E201Q/E203N,D149N/E185N/E201N/E203Q,D149Q/E185Q/E201N/

E203N,

[0268] D149Q/E185N/E201Q/E203N,D149Q/E185N/E201N/E203Q,D149N/E185Q/E201Q/E203N,

[0269] D149N/E185Q/E201N/E203Q,D149N/E185N/E201Q/E203Q,D149Q/E185Q/E201Q/E203Q,

[0270] D149Q/E185Q/E201N/E203Q,D149Q/E185N/E201Q/E203Q,D149N/E185Q/E201Q/E203Q,

[0271] D149Q/E185Q/E201Q/E203N,D149N/E185N/D195N/E201N/E203N,

[0272] D149Q/E185N/D195N/E201N/E203N,D149N/E185Q/D195N/E201N/E203N,

[0273] D149N/E185N/D195Q/E201N/E203N,D149N/E185N/D195N/E201Q/E203N,

[0274] D149N/E185N/D195N/E201N/E203Q,D149Q/E185Q/D195N/E201N/E203N,

[0275] D149Q/E185N/D195Q/E201N/E203N,D149Q/E185N/D195N/E201Q/E203N,

[0276] D149Q/E185N/D195N/E201N/E203Q,D149N/E185Q/D195Q/E201N/E203N,

[0277] D149N/E185Q/D195N/E201Q/E203N,D149N/E185Q/D195N/E201N/E203Q,

[0278] D149N/E185N/D195Q/E201Q/E203N,D149N/E185N/D195Q/E201N/E203Q,

[0279] D149N/E185N/D195N/E201Q/E203Q,D149Q/E185Q/D195Q/E201N/E203N,

[0280] D149Q/E185Q/D195N/E201Q/E203N,D149Q/E185Q/D195N/E201N/E203Q,

[0281] D149Q/E185N/D195Q/E201Q/E203N,D149Q/E185N/D195Q/E201N/E203Q,

[0282] D149Q/E185N/D195N/E201Q/E203Q,D149N/E185Q/D195Q/E201Q/E203N,

[0283] D149N/E185Q/D195Q/E201N/E203Q,D149N/E185Q/D195N/E201Q/E203Q,

[0284] D149N/E185N/D195Q/E201Q/E203Q,D149Q/E185Q/D195Q/E201Q/E203N,

[0285] D149Q/E185Q/D195Q/E201N/E203Q,D149Q/E185Q/D195N/E201Q/E203Q,

[0286] D149Q/E185N/D195Q/E201Q/E203Q,D149N/E185Q/D195Q/E201Q/E203Q,

[0287] D149Q/E185Q/D195Q/E201Q/E203Q,D149N/E185R/E201N/E203N,D149Q/E185R/E201N/E203N,

[0288] D149N/E185R/E201Q/E203N,D149N/E185R/E201N/E203Q,D149Q/E185R/E201Q/E203N,

[0289] D149Q/E185R/E201N/E203Q,D149N/E185R/E201Q/E203Q,D149Q/E185R/E201Q/E203Q,

[0290] D149R/E185N/E201N/E203N,D149R/E185Q/E201N/E203N,D149R/E185N/E201Q/E203N,

[0291] D149R/E185N/E201N/E203Q,D149R/E185Q/E201Q/E203N,D149R/E185Q/E201N/E203Q,

[0292] D149R/E185N/E201Q/E203Q,D149R/E185Q/E201Q/E203Q,D149R/E185N/D195N/E201N/E203N,

[0293] D149R/E185Q/D195N/E201N/E203N,D149R/E185N/D195Q/E201N/E203N,

[0294] D149R/E185N/D195N/E201Q/E203N,D149R/E185Q/D195N/E201N/E203Q,

[0295] D149R/E185Q/D195Q/E201N/E203N,D149R/E185Q/D195N/E201Q/E203N,

[0296] D149R/E185Q/D195N/E201N/E203Q,D149R/E185N/D195Q/E201Q/E203N,

[0297] D149R/E185N/D195Q/E201N/E203Q,D149R/E185N/D195N/E201Q/E203Q,

[0298] D149R/E185Q/D195Q/E201Q/E203N,D149R/E185Q/D195Q/E201N/E203Q,

[0299] D149R/E185Q/D195N/E201Q/E203Q,D149R/E185N/D195Q/E201Q/E203Q,

[0300] D149R/E185Q/D195Q/E201Q/E203Q,D149N/E185R/D195N/E201N/E203N,

[0301] D149Q/E185R/D195N/E201N/E203N,D149N/E185R/D195Q/E201N/E203N,

[0302] D149N/E185R/D195N/E201Q/E203N,D149N/E185R/D195N/E201N/E203Q,

[0303] D149Q/E185R/D195Q/E201N/E203N,D149Q/E185R/D195N/E201Q/E203N,

[0304] D149Q/E185R/D195N/E201N/E203Q,D149N/E185R/D195Q/E201Q/E203N,

[0305] D149N/E185R/D195Q/E201N/E203Q,D149N/E185R/D195N/E201Q/E203Q,

[0306] D149Q/E185R/D195Q/E201Q/E203N,D149Q/E185R/D195Q/E201N/E203Q,

[0307] D149Q/E185R/D195N/E201Q/E203Q,D149N/E185R/D195Q/E201Q/E203Q,

[0308] D149Q/E185R/D195Q/E201Q/E203Q,D149N/E185R/D195N/E201R/E203N,

[0309] D149Q/E185R/D195N/E201R/E203N,D149N/E185R/D195Q/E201R/E203N,

[0310] D149N/E185R/D195N/E201R/E203Q,D149Q/E185R/D195Q/E201R/E203N,

[0311] D149Q/E185R/D195N/E201R/E203Q,D149N/E185R/D195Q/E201R/E203Q,

[0312] D149Q/E185R/D195Q/E201R/E203Q,E131D/K49R,E101N/N102F,E101N/N102Y,E101N/N102W,

[0313] E101F/N102F,E101F/N102Y,E101F/N102W,E101Y/N102F,E101Y/N102Y,E101Y/N102W,

[0314] E101W/N102F,E101W/N102Y,E101W/N102W,E101N/N102R,E101F/N102R,E101Y/N102R或E101W/N102F。

[0315] 本发明形成所述孔的优选变体——其中当多核苷酸移动通过孔时较少的核苷酸有助于电流——包括Y51A/F56A,Y51A/F56N,Y51I/F56A,Y51L/F56A,Y51T/F56A,Y51I/F56N,Y51L/F56N或Y51T/F56N或更优选Y51I/F56A,Y51L/F56A或Y51T/F56A。如上所述,这使得更容易鉴定观察到的电流(当多核苷酸移动通过所述孔时)与多核苷酸之间的直接关系。

[0316] 形成显示增加的范围的孔的优选变体包括以下位置的突变:

[0317] Y51,F56,D149,E185,E201和E203;

[0318] N55和F56;

[0319] Y51和F56;

[0320] Y51,N55和F56;或

[0321] F56和N102。

[0322] 形成显示增加的范围的孔的优选变体包括:

[0323] Y51N,F56A,D149N,E185R,E201N和E203N;

[0324] N55S和F56Q;

[0325] Y51A和F56A;

[0326] Y51A和F56N;

[0327] Y51I和F56A;

[0328] Y51L和F56A;

[0329] Y51T和F56A;

[0330] Y51I和F56N;

[0331] Y51L和F56N;

[0332] Y51T和F56N;

[0333] Y51T和F56Q;

[0334] Y51A,N55S和F56A;

[0335] Y51A,N55S和F56N;

[0336] Y51T,N55S和F56Q;或

[0337] F56Q和N102R。

[0338] 形成所述孔的优选变体——其中当多核苷酸移动通过所述孔时较少的核苷酸贡献于电流——包括在下列位置的突变:

[0339] N55和F56,例如N55X和F56Q,其中X是任意氨基酸;或

[0340] Y51和F56,例如Y51X和F56Q,其中X是任意氨基酸。

[0341] 形成显示增加的通量的孔的优选变体包括以下位置的突变:

[0342] D149,E185和E203;

[0343] D149,E185,E201和E203;或者

[0344] D149,E185,D195,E201和E203。

[0345] 形成显示增加的通量的孔的优选变体包括:

[0346] D149N,E185N和E203N;

[0347] D149N,E185N,E201N和E203N;

[0348] D149N,E185R,D195N,E201N和E203N;或

[0349] D149N,E185R,D195N,E201R和E203N。

[0350] 形成所述孔的优选变体——其中多核苷酸的捕获增加——包括以下突变:

[0351] D43N/Y51T/F56Q;

[0352] E44N/Y51T/F56Q;

[0353] D43N/E44N/Y51T/F56Q;

[0354] Y51T/F56Q/Q62R;

[0355] D43N/Y51T/F56Q/Q62R;

[0356] E44N/Y51T/F56Q/Q62R;或

[0357] D43N/E44N/Y51T/F56Q/Q62R。

[0358] 优选的变体包括以下突变:

[0359] D149R/E185R/E201R/E203R或Y51T/F56Q/D149R/E185R/E201R/E203R;

[0360] D149N/E185N/E201N/E203N或Y51T/F56Q/D149N/E185N/E201N/E203N;

[0361] E201R/E203R或Y51T/F56Q/E201R/E203R

[0362] E201N/E203R或Y51T/F56Q/E201N/E203R;

[0363] E203R或Y51T/F56Q/E203R;

[0364] E203N或Y51T/F56Q/E203N;

[0365] E201R或Y51TF56Q/E201R;

[0366] E201N或Y51T/F56Q/E201N;

[0367] E185R或Y51T/F56Q/E185R;

[0368] E185N或Y51T/F56Q/E185N;

[0369] D149R或Y51T/F56Q/D149R;

[0370] D149N或Y51T/F56Q/D149N;

[0371] R142E或Y51T/F56Q/R142E;

[0372] R142N或Y51T/F56Q/R142N;

[0373] R192E或Y51T/F56Q/R192E;或者

[0374] R192N或Y51T/F56Q/R192N。

[0375] 优选的变体包括以下突变:

[0376] Y51A/F56Q/E101N/N102R;

[0377] Y51A/F56Q/R97N/N102G;

[0378] Y51A/F56Q/R97N/N102R;

[0379] Y51A/F56Q/R97N;

[0380] Y51A/F56Q/R97G;

[0381] Y51A/F56Q/R97L;

[0382] Y51A/F56Q/N102R;

[0383] Y51A/F56Q/N102F;

[0384] Y51A/F56Q/N102G;

[0385] Y51A/F56Q/E101R;

[0386] Y51A/F56Q/E101F;

[0387] Y51A/F56Q/E101N;或

[0388] Y51A/F56Q/E101G。

[0389] 本发明还提供了包含SEQ ID NO:390所示序列的变体的突变型CsgG单体,其中所述变体包括在T150位置的突变。形成表现出增加的插入的孔的优选变体包括T150I。在T150位置的突变,例如T150I,可以与上述任意突变或突变的组合进行结合。

[0390] 本发明还提供包含SEQ ID NO:390所示序列的变体的突变型CsgG单体,所述变体包含实施例中公开的变体中存在的突变的组合。

[0391] 引入或取代天然存在氨基酸的方法是本领域公知的。例如,在编码突变单体的多核苷酸的相关位置,通过用精氨酸的密码子(CGT)替换甲硫氨酸的密码子(ATG),可使甲硫氨酸(M)被精氨酸(R)取代。所述多核苷酸可以随后如下所述进行表达。

[0392] 引入或取代非天然存在氨基酸的方法也是本领域公知的。例如,非天然存在的氨基酸可以通过在用于表达突变单体的IVTT系统中包含合成氨酰基-tRNA来引入。或者,它们可以通过在那些特定氨基酸的合成(即非天然存在的)类似物的存在下通过在对特定氨基酸为营养缺陷型的大肠杆菌中表达所述突变单体来引入。如果使用部分肽合成来产生突变单体,则它们也可以通过裸连接来制备。

[0393] 本发明的其他单体

[0394] 在另一个实施方式中,本发明提供一种突变CsgG单体,其包括SEQ ID NO:390所示的序列的变体,其中所述变体包括在位置Y51,N55和F56中的一个或多个的突变。所述变体可以包含在位置Y51;N55;F56;Y51/N55;Y51/F56;N55/F56;或Y51/N55/F56的突变。所述变体优选包括位置Y51,N55或F56的突变。所述变体可以包含上述讨论的在Y51,N55和F56中一

个或多个位置的任意特异性突变和其任意组合。Y51,N55和F56的一个或多个可以用任意氨基酸取代。Y51可以被F,M,L,I,V,A,P,G,C,Q,N,T,S,E,D,K,H或R取代,例如A,S,T,N或Q。N55可以被F,M,L,I,V,A,P,G,C,Q,,T,S,E,D,K,H或R取代,例如A,S,T,或Q。F56可以被M,L,I,V,A,P,G,C,Q,N,T,S,E,D,K,H或R取代,例如A,S,T,N或Q。

[0395] 所述变体可以进一步包括以下中的一个或多个:(i)在以下位置的一个或多个突变(即在一个或多个以下位置的突变):(i)N40,D43,E44,S54,S57,Q62,R97,E101,E124,E131,R142,T150和R192,(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)缺失以下位置中的一个或多个:F48,K49,P50,Y51,P52,A53,S54,N55,F56和S57。所述变体可以包含上述讨论的(i)和(iii)至(xi)的任意组合。所述变体可以包含上述讨论的对(i)和(iii)至(xi)的任意实施方式。

[0396] 变体

[0397] 除上述具体突变之外,变体可以包括其他突变。在SEQ ID NO:390的氨基酸序列的整个长度上,基于氨基酸同一性,变体优选与所述序列至少50%同源。更优选,基于氨基酸同一性,所述变体与SEQ ID NO:390的氨基酸序列在整个长度上可以至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%以及更优选至少95%,97%或99%同源。在100个以上,例如125,150,175或200以上的连续氨基酸延伸链上,可以具有至少80%,例如至少85%,90%或95%的氨基酸同一性(“严格硬同源性”(hard homology))。

[0398] 现有技术的标准方法可以用于测定同源性。例如UWGCG软件包提供了可用于计算同源性的BESTFIT程序,例如使用其默认设置(Devereux et al.(1984)Nucleic Acids Research 12,p387-395)。PILEUP和BLAST算法可用于计算同源性或序列排序(例如识别等效残基或相应序列(通常以其默认设置)),例如如Altschul S.F.(1993)J Mol Evol 36290-300;Altschul,S.Fet al(1990)J Mol Biol 215403-10中描述的。用于进行BLAST分析的软件可通过国家生物技术信息中心公开获得(http://www.ncbi.nlm.nih.gov)。

[0399] SEQ ID NO:390是来自大肠杆菌Str.K-12substr.MC4100的野生型CsgG单体。序列SEQ ID NO:390的变体可以包括存在于另一个CsgG同源物中的任意取代。优选的CsgG同源物在SEQ ID:NO:391至395和414至429中示出。所述变体可以包括与SEQ ID NO:390相比存在于SEQ ID:NO:391至395和414至429中的一个或多个取代的组合。

[0400] 除上述讨论的氨基酸序列之外,还可以对SEQ ID NO:390的氨基酸序列进行氨基酸取代,例如至多1,2,3,4,5,10,20或30个取代。保守取代是将氨基酸用其他具有相似化学结构、相似化学性质或相似侧链体积的氨基酸替换。引入的所述氨基酸可与其替换的氨基酸相比,具有相似的极性,亲水性,疏水性,碱性,酸性,中性或电荷。或者,所述保守取代可以引入另一芳族的或脂肪族的氨基酸代替之前存在的芳族或脂肪族氨基酸。保守的氨基酸改变是本领域公知的,并且可以根据下表1中定义的20个主要氨基酸的性质进行选择。当氨基酸具有相似的极性时,这也可以参考表2中氨基酸侧链的亲水性水平来确定。

[0401] 表1-氨基酸的化学性质

[0402]

Ala	脂肪族，疏水，中性	Met	疏水，中性
Cys	极性，疏水，中性	Asn	极性，亲水，中性
Asp	极性，亲水，带电荷(-)	Pro	疏水，中性
Glu	极性，亲水，带电荷(-)	Gln	极性，亲水，中性
Phe	芳香族，疏水，中性	Arg	极性，亲水，带电荷（+）
Gly	脂肪族，中性	Ser	极性，亲水，中性
His	芳香族，极性，亲水，带电荷（+）	Thr	极性，亲水，中性
Ile	脂肪族，疏水，中性	Val	脂肪族，疏水，中性
Lys	极性，亲水，带电荷（+）	Trp	芳香族，疏水，中性
Leu	脂肪族，疏水，中性	Tyr	芳香族，极性，疏水

[0403] 表2-亲水性水平

[0404]

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

[0405]

[0406] SEQ ID NO:390的氨基酸序列的一个或多个氨基酸残基可另外从上述多肽中缺失。多达1,2,3,4,5,10,20或30个以上残基可以缺失。

[0407] 变体可以包含SEQ ID NO:390的片段。所述片段保留形成孔的活性。所述片段的长度可以为至少50,至少100,至少150,至少200或至少250个氨基酸。所述片段可以用于产生所述孔。片段优选包括SEQ ID NO:390的跨膜结构域,即K135-Q153和S183-S208。

[0408] 一个或多个氨基酸可以替换地或额外地加入到上述多肽中。可以在SEQ ID NO:390或其多肽变体或其片段的氨基酸序列的氨基末端或羧基末端提供延伸。所述延伸可以是相当短的,例如1至10个氨基酸长度。替代地,所述延伸可以更长,例如达到50或100个氨基酸。载体蛋白可以根据本发明与氨基酸序列融合。下面更详细地讨论其他融合蛋白。

[0409] 如上所述,变体是具有由SEQ ID NO:390的序列变化而来并且保留其形成孔能力

的氨基酸序列的多肽。变体通常包含SEQ ID NO:390中负责孔形成的区域。包含β-桶结构的CsgG的孔形成能力由每个亚基中的β-片体(β-sheets)提供。SEQ ID NO:390的变体通常包括SEQ ID NO:390中形成β-变体的区域,即K135-Q153和S183-S208。只要所得到的变体保持其形成孔的能力,就可以对SEQ ID NO:390中形成β-片体的区域进行一个或多个修饰。SEQ ID NO:390的变体优选包括在其α-螺旋和/或环结构区域内的一个或多个修饰,例如取代,增加或缺失。

[0410] 衍生自CsgG的单体可以进行修饰,以帮助对它们的鉴定或纯化,例如通过添加链霉亲和素标签或通过添加用以促进其从细胞中分泌的信号序列,其中所述单体不天然含有这种序列。下面更详细地讨论其他合适的标签。所述单体可以用显示标签标记。所述显示标签可以是允许所述单体被检测的任意合适的标签。合适的标签如下所述。

[0411] 衍生自CsgG的单体也可以使用D-氨基酸制备。例如,衍生自CsgG的单体可以包括L-氨基酸和D-氨基酸的混合物。这在制备上述蛋白质或肽的技术领域中是常规的。

[0412] 衍生自CsgG的单体包含一个或多个特异性修饰以便于核苷酸的区分。衍生自CsgG的单体也可以包含其他非特异性修饰,只要其不会干扰孔的形成。许多非特异性侧链修饰是本领域已知的,并且可以针对衍生自CsgG的单体的侧链进行。所述修饰包括,例如通过下述进行的氨基酸还原性烷基化:与醛反应然后用$NaBH_4$还原,用亚氨酰乙酸甲酯(methylacetimidate)进行脒基化或用乙酸酐进行酰化来氨基酸。

[0413] 所述衍生自CsgG的单体可以采用本领域已知的方法产生。所述衍生自CsgG的单体可以合成制备或通过重组方法制备。例如,所述单体可以通过体外翻译和转录(IVTT)来合成。生产所述孔和单体的合适方法在国际申请号PCT/GB09/001690(公开为WO 2010/004273),PCT/GB09/001679(公开为WO 2010/004265)或PCT/GB10/000133(公开为WO 2010/086603)中有所讨论。将孔插入膜的方法如下所述。

[0414] 在一些实施方式中,所述突变单体经过化学修饰。所述突变单体可以通过任意方式和在任意位点进行化学修饰。所述突变单体优选通过将分子连接到一个或多个半胱氨酸(半胱氨酸连接),分子连接到一个或多个赖氨酸,将分子连接到一个或多个非天然氨基酸,抗原表位的酶修饰或末端修饰,来进行化学修饰。实施上述修饰的合适方法是本领域公知的。所述突变单体可以通过任意分子的连接进行化学修饰。例如,所述突变单体可以通过连接染料或荧光团进行化学修饰。

[0415] 在一些实施方式中,所述突变单体用分子适配器(adaptor)进行化学修饰,分子适配器促进包含所述单体的孔和目标核苷酸或目标多核苷酸序列之间的相互作用。所述适配器的存在改善所述孔和所述核苷酸或多核苷酸的主-客体化学,因而提高了由所述突变单体形成的孔的测序能力。主-客体化学的原理是本领域公知的。所述适配器对所述孔的物理或化学性质有影响,能改善其与核苷酸或多核苷酸序列的相互作用。所述适配器可以改变孔的桶状结构或通道结构的电荷,或特异性地与所述核苷酸或多核苷酸序列相互作用或结合从而促进其与所述孔的相互作用。

[0416] 所述分子适配器优选为环状分子,环糊精,能够进行杂合的物质,DNA结合剂或螯合剂,肽或肽类似物,合成聚合物,芳族平面分子,带正电荷的小分子或能够进行氢键合的小分子。

[0417] 所述适配器可以为环状的。环状适配器优选具有与所述孔相同的对称性。由于

CsgG通常围绕中心轴具有8个或9个亚基,因此所述适配器优选具有8重或9重的对称性。这将在下面更详细地讨论。

[0418] 所述适配器通常通过主-客体化学与所述核苷酸或多核苷酸序列相互作用。所述适配器通常能够与所述核苷酸或多核苷酸序列相互作用。所述适配器包括一个或多个能够与所述核苷酸或多核苷酸序列相互作用的化学基团。所述一个或多个化学基团优选通过非共价相互作用与所述核苷酸或多核苷酸序列相互作用,例如,疏水性相互作用,氢键,范得华力,π-阳离子相互作用和/或静电力。能够与所述核苷酸或多核苷酸序列相互作用的一个或多个化学基团优选为带正电荷的。能够与所述核苷酸或多核苷酸序列相互作用的一个或多个化学基团更优选包括氨基基团。所述氨基基团可以连接在伯、仲或叔碳原子上。所述适配器甚至更优选包括氨基基团的环,例如6,7或8个氨基基团的环。所述适配器最优选包括8个氨基基团的环。质子化的氨基基团的环可以与在所述核苷酸或多核苷酸序列中带负电荷的磷酸基团相互作用。

[0419] 通过所述适配器与包括所述突变单体的孔之间的主-客体化学可以促进所述适配器在所述孔中的正确定位。所述适配器优选包括一个或多个能够与一个或多个所述孔中的氨基酸相互作用的化学基团。所述适配器更优选包括一个或多个能够通过非共价相互作用与所述孔中一个或多个氨基酸相互作用的化学基团,例如,疏水性相互作用,氢键,范得华力,π-阳离子相互作用和/或静电力。所述能够与所述孔中一个或多个氨基酸相互作用的化学基团通常为羟基或胺。所述羟基基团可以连接在伯、仲或叔碳原子上。所述羟基基团可以与所述孔中未带电荷的氨基酸形成氢键。可以使用任意促进所述孔和所述核苷酸或多核苷酸序列的适配器。

[0420] 合适的适配器包括,但不限于,环糊精、环肽和葫芦脲。所述适配器优选为环糊精或其衍生物。所述环糊精或其衍生物可以为在Eliseev,A.V.,和Schneider,H-J.(1994)J.Am.Chem.Soc.116,6081-6088中公开的任意一种。所述适配器更优选为七-6-氨基-β-环糊精(am_7-βCD),6-单脱氧-6-单氨基-β-环糊精(am_1-βCD)或七-(6-脱氧-6-胍基)-环糊精(gu_7-βCD)。所述gu_7-βCD中的胍基基团比am_7-βCD中的伯胺具有更高的PKa值,因此其更容易带正电荷。所述gu_7-βCD适配器可用于增加所述孔中核苷酸的停留时间,以提高测量的残留电流的准确度,以及增加高温或低数据采集率下的基础检出率。

[0421] 如果如下更详细所述的使用3-(2-吡啶基二硫代)丙酸琥珀酰亚胺酯(SPDP)交联剂,则适配器优选为七(6-脱氧-6-氨基)-6-N-单(2-吡啶基)二硫丙酰基-β-环糊精(am_6amPDP_1-βCD)。

[0422] 更合适的适配器包括包含9个糖单元的γ-环糊精(并因此具有九重对称)。γ-环糊精可以包含连接体分子,或者可以被修饰成包含上述β-环糊精实施例中使用的所有或多个修饰的糖单元。

[0423] 所述分子适配器优选为共价连接到所述突变单体上。所述适配器可以通过任意本领域已知的方法共价连接到所述孔上。所述适配器通常通过化学连接进行连接。如果通过半胱氨酸连接基团连接所述分子适配器,则优选将一个或多个半胱氨酸通过取代引入所述突变体,例如引入所述桶结构中。所述突变单体可以通过将分子适配器连接到所述突变单体中的一个或多个半胱氨酸上进行化学修饰。所述一个或多个半胱氨酸可以是天然存在的,即位于SEQ ID NO:390中1和/或215位置。替代地,所述突变单体可以通过将分子连接到

引入到其他位置的一个或多个半胱氨酸上进行化学修饰。位于位置215的半胱氨酸可以被除去,例如通过取代除去,以保证所述分子适配器不会连接到除位于位置1的半胱氨酸以外的位置或连接到在另一位置引入的半胱氨酸上。

[0424] 半胱氨酸残基的反应活性可以通过修饰其邻近残基进行增强。例如,侧链精氨酸,组氨酸或赖氨酸残基的碱性基团会将半胱氨酸硫醇基团的pKa改变为更具反应性的S^-基团的pKa。所述半胱氨酸残基的反应活性可以被硫醇保护基如dTNB保护。在连接连接体之前,这些保护基团可以与所述突变单体的一个或多个半胱氨酸残基反应。

[0425] 所述分子可以直接连接到所述突变单体上。所述分子优选通过连接体,例如化学交联剂或肽连接体,连接到所述突变单体上。

[0426] 合适的化学交联剂在本领域是公知的。优选的交联剂包括2,5-二氧代吡咯烷-1-基3-(吡啶-2-基硫烷基)丙酸酯,2,5-二氧代吡咯烷-1-基4-(吡啶-2-基硫烷基)丁酸乙酯和2,5-二氧代吡咯烷-1-基8-(吡啶-2-基二硫烷基)辛酸酯。最优选的交联剂为3-(2-吡啶基二硫代)丙酸琥珀酰亚胺酯(SPDP)通常地,在所述分子/交联剂复合物共价连接到所述突变单体之前,所述分子共价连接到所述双功能交联剂上,但也可能是在所述双功能交联剂/单体复合物连接到所述分子之前,将所述双功能交联剂共价连接到所述单体上。

[0427] 所述连接体优选对二硫苏糖醇(DTT)具有抗性。合适的连接体包括,但不限于,碘乙酰胺系和马来酰亚胺系连接体。

[0428] 在另一实施方式中,所述单体可以连接到多核苷酸结合蛋白上。这形成可用于本发明测序方法中的模块化测序系统。以下内容讨论多核苷酸结合蛋白。

[0429] 所述多核苷酸结合蛋白优选共价连接到所述突变单体上。所述蛋白可以采用任意本领域已知方法共价连接到所述单体上。所述单体和蛋白可以进行化学融合或遗传融合。如果整个构建体由单个多核苷酸序列表达,则单体和蛋白质可以进行遗传融合。单体与多核苷酸结合蛋白的遗传融合在国际申请号PCT/GB09/001679(公开为WO 2010/004265)中有所讨论。

[0430] 如果所述多核苷酸结合蛋白通过半胱氨酸连接基团连接,则所述一个或多个半胱氨酸优选通过取代引入到所述突变体中。所述一个或多个半胱氨酸优选引入到环状区域,所述环状区域在同源物中具有低保守性,这表明可以耐受突变或插入。它们因此适合用于连接多核苷酸结合蛋白。在这类实施方式中,251位置的天然存在的半胱氨酸可以被除去。半胱氨酸残基的反应活性可以通过上文描述的修饰进行增强。

[0431] 所述多核苷酸结合蛋白可以被直接或通过一个或多个连接体连接到所述突变单体上。可以使用国际申请号PCT/GB10/000132(公开为WO 2010/086602)中描述的杂合连接体将所述分子连接到所述突变体单体上。替代地,可以使用肽连接体。肽连接体为氨基酸序列。所述肽连接体的长度,柔性和亲水性通常被设计成使其不破坏所述单体和分子的功能。优选的柔性肽连接体具有2至20个,例如4,6,8,10或16个丝氨酸和/或甘氨酸的链长。更优选的柔性连接体包括$(SG)_1$,$(SG)_2$,$(SG)_3$,$(SG)_4$,$(SG)_5$and$(SG)_8$,其中S为丝氨酸而G为甘氨酸。优选的刚性连接体具有2至30个,例如4,6,8,16或24个脯氨酸的链长。更优选的刚性连接体包括$(P)_{12}$,其中P为脯氨酸。

[0432] 所述突变单体可以采用分子适配器和多核苷酸结合蛋白进行化学修饰。

[0433] 所述分子(用其对所述单体进行化学修饰)可以直接连接到所述单体或如在国际

申请PCT/GB09/001690(公布号WO 2010/004273),PCT/GB09/001679(公布号WO 2010/004265)或PCT/GB10/000133(公开号WO 2010/086603)中公开的通过连接体连接。

[0434] 本文所述的任意蛋白质,例如本发明所述的突变体单体和孔,可以被修饰以有助于对它们的鉴定或纯化,例如通过添加组氨酸残基(his标签),天冬氨酸残基(asp标签),链霉亲和素标签,旗号标签(flag tag),SUMO标签,GST标签或MBP标签,或者通过添加促进它们从细胞分泌的信号序列,其中多肽不天然含有这类序列。引入遗传标签的替代方法是将标签化学反应到蛋白上的天然或改造过的位置。这样的一个例子可以是使凝胶移位试剂与在蛋白质外部改造的半胱氨酸反应。该方法已倍证明是一种分离溶血素异源寡聚体的方法(Chem Biol.1997Jul;4(7)497-505)。

[0435] 任意本文所述的蛋白质,例如本发明所述的突变单体和孔,可以用显示标记进行标记。所述显示标记可以是允许所述蛋白质被检测的任意合适标记。合适的标记包括,但不限于,荧光分子,放射性同位素例如^{125}I,^{35}S,酶,抗体,抗原,多核苷酸和配体例如生物素。

[0436] 任意本文所述的蛋白质,例如本发明所述的单体和孔,可以合成制备或通过重组方法制备。例如,所述蛋白质可以通过体外翻译和转录(IVTT)来合成。所述蛋白质的氨基酸序列可以被修饰为,包括非天然存在的氨基酸或增加所述蛋白质的稳定性。当通过合成方法制备所述蛋白质时,可以在制备过程中引入这种氨基酸。在合成或重组制备之后,所述蛋白也可能被改变。

[0437] 蛋白质也可以采用D-氨基酸制备。例如,所述蛋白质可以包括L-氨基酸和D-氨基酸的混合物。这在制备上述蛋白质或肽的技术领域中是常规的。

[0438] 所述蛋白质也可以包含其他非特异性修饰,只要它们不干扰所述蛋白质的功能。许多非特异性侧链修饰是本领域已知的并可以对所述蛋白的侧链进行。所述修饰包括,例如通过下述进行的氨基酸还原性烷基化:与醛反应然后用$NaBH_4$还原,用亚氨酰乙酸甲酯进行脒基化或用乙酸酐进行酰化来氨基酸。

[0439] 本文所述的任意蛋白,包括本发明所述单体和孔,可以采用本领域已知的标准方法制备。编码蛋白的多核苷酸序列可以采用本领域的标准方法进行衍生和复制。编码蛋白的多核苷酸序列采用本领域标准技术在细菌宿主细胞中表达。所述蛋白可以从重组表达载体通过在细胞中原位表达所述多肽进行制备。所述表达载体任选地带有诱导型启动子以控制所述多肽的表达。这些方法在Sambrook,J.and Russell,D.(2001).Molecular Cloning A Laboratory Manual,3rd Edition.Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY中描述。

[0440] 从产生蛋白质的生物体中或重组表达后经任意蛋白质液相色谱系统纯化后,蛋白可以规模生产。典型的蛋白质液相色谱系统包括FPLC,AKTA系统,Bio-Cad系统,Bio-Rad BioLogic系统和Gilson HPLC系统。

[0441] 构建体

[0442] 本发明也提供一种包括两个以上共价连接的CsgG单体的构建体,其中至少一个所述单体是本发明所述的突变单体。本发明所述的构建体保持其形成孔的能力。其可以如上所述进行确定。本发明的一个或多个构建体可以用于形成用于表征的孔,例如多核苷酸测序。所述构建体可以包括至少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。所述构建体优选包括2个单体。所述两个或多个单体可以为相同的或不同的。

[0443] 所述构建体中至少一个单体为本发明所述的突变单体。所述构建体中2个以上,3个以上,4个以上,5个以上,6个以上,7个以上,8个以上,9个以上,10个以上单体可以为本发明所述的突变单体。所述构建体中所有单体优选为本发明中所述突变单体。所述突变单体可以是相同的或不同的。在优选的实施方式中,所述构建体包括两个本发明所述的突变单体。

[0444] 所述构建体中本发明所述突变单体优选为具有大约相同的长度或具有相同的长度。所述构建体中本发明突变单体的桶状结构优选为具有大约相同的长度或具有相同的长度。长度可以用氨基酸数量和/或单元长度进行测量。

[0445] 所述构建体可以包括一个或多个不是本发明突变单体的单体。不为本发明突变单体的CsgG突变单体包括包含SEQ ID NO:

[0446] 390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的单体或SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,414,421,422,423,424,425,426,427,428或429的对照变体,其中上述氨基酸/位置均未进行突变。所述构建体中的至少一个单体可以包括SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429或者SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429中所示序列的对照变体。SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的对照变体与390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428or 429,40或41在其整个序列上基于氨基酸的同一性至少50%同源。更优选,所述对照变体与SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的氨基酸序列在整个长度上基于氨基酸同一性可以至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%以及更优选至少95%,97%或99%同源。

[0447] 所述构建体中的单体优选为遗传融合的。如果整个构建体由单个多核苷酸序列表达,则单体可以进行遗传融合。所述单体的编码序列可以以任意方式重组以形成编码所述构建体的单个多核苷酸序列。

[0448] 所述单体可以以任意构象进行遗传融合。所述单体可以通过其末端氨基酸进行融合。例如,一个所述单体的氨基末端可以与另一单体的羧基末端融合。构建体中第二和其后的单体(从氨基至羧基方向)可以在它们的氨基末端(各自与之前单体的羧基末端融合)包含一个甲硫氨酸。例如,如果M是一个单体(不带有氨基末端甲硫氨酸)且mM是带有氨基末端甲硫氨酸的单体,则所述构建体可以包括序列M-mM,M-mM-mM或M-mM-mM-mM。这些甲硫氨酸的存在通常来自编码整个构建体的多核苷酸中编码第二个或其后的单体的多核苷酸的5'末端的起始密码子(即ATG)的表达。所述构建体中的第一单体(从氨基到羧基方向)也可以包括甲硫氨酸(例如mM-mM,mM-mM-mM或mM-mM-mM-mM)。

[0449] 所述两个或更多个单体可以直接遗传融合在一起。所述单体优选采用连接体遗传融合。所述连接体可以设计为限制所述单体的移动。优选的连接体是氨基酸序列(即肽连接体)。上文所述的任意肽连接体均可以使用。

[0450] 在另一优选的实施方式中,所述单体是化学融合的。如果两个单体是化学连接的,

例如通过化学交联剂连接,则所述两个单体是化学融合的。下文所述的任意化学交联剂均可使用。所述连接体可以连接到一个或多个引入到本发明突变单体中的半胱氨酸残基上。或者,所述连接体可以连接到所述构建体中所述单体之一的末端。

[0451] 如果构建体包含不同单体,可以通过在大量过量的所述单体中保持所述连接体的浓度防止单体彼此间的交联。替代地,可以在使用两个连接体时采用“锁和钥”设置。每个连接体的仅一端可以一起反应以形成更长的连接体,并且所述连接体的另一端各自与不同的单体反应。这种连接体在国际申请号PCT/GB10/000132(公开号:WO 2010/086602)中描述。

[0452] 多核苷酸

[0453] 本发明还提供编码本发明所述突变单体的多核苷酸序列。所述突变单体可以是如上所述的任意一种。所述多核苷酸序列优选包括基于核苷酸同一性与序列SEQ ID NO:389的序列在整个序列上至少50%,60%,70%,80%,90%或95%同源的序列。在300个以上,例如375,450,525或600个以上的连续核苷酸延伸链上,可以具有至少80%例如至少85%,90%或95%的核苷酸同一性(“严格同源性”)。同源性可以如上所述进行计算。所述多核苷酸序列可以包含基于遗传密码的简并性而与SEQ ID NO:389不同的序列。

[0454] 本发明还提供编码任意本发明所述遗传融合构建体的多核苷酸序列。所述多核苷酸优选包括SEQ ID NO:389中所示序列的两个或更多个变体。所述多核苷酸序列优选包括两个或更多个与SEQ ID NO:389序列基于核苷酸同一性在整个序列上具有至少50%,60%,70%,80%,90%或95%同源性的序列。在600个以上,例如750,900,1050或1200个以上的连续核苷酸延伸链上,可以具有至少80%例如至少85%,90%或95%的核苷酸同一性(”严格同源性”)。同源性可以如上所述进行计算。

[0455] 多核苷酸序列可以采用本领域的标准方法进行衍生和复制。编码野生型CsgG的染色体DNA可以从产生孔的生物体如大肠杆菌中提取。编码所述孔亚基的基因可以使用包括特异性引物的PCR进行扩增。所述扩增的序列随后可以进行定点突变。定点突变的合适方法是本领域已知的并且包括,例如,组合链式反应。编码本发明构建的多核苷酸可以采用本领域公知的技术制备,例如在Sambrook,J.and Russell,D.(2001).Molecular Cloning A Laboratory Manual,3rd Edition.Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY中描述的那些。

[0456] 所得到的多核苷酸序列随后可以被整合到重组可复制载体上,例如克隆载体。所述载体可以用于在相容的宿主细胞中复制所述多核苷酸。因此多核苷酸序列可以通过将多核苷酸引入到可复制载体中,将载体引入相容的宿主细胞中,并在引起载体复制的条件下使宿主细胞生长而进行制备。所述载体可以从所述宿主细胞中回收。用于多核苷酸克隆的合适宿主细胞是本领域已知的并且在下文进行更具体的描述。

[0457] 所述多核苷酸序列可以克隆进合适的表达载体中。在表达载体中,所述多核苷酸序列通常可操作地连接到能够提供宿主细胞对编码序列的表达的控制序列。所述表达载体可以用于表达孔亚基。

[0458] 术语"可操作地连接"是指并置连接(juxtaposition),其中所描述的组分处于允许它们以其预期方式发挥作用的关系。控制序列“可操作地连接”到编码序列是指以使得编码序列的表达在与对照序列相容的条件下实现的方式连接。可以将相同或不同的多核苷酸序列的多个复本引入到所述载体中。

[0459] 所述表达载体可以随后被引入到合适的宿主细胞中。因此,本发明的突变体单体或构建体可以通过将多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,并在能引起所述多核苷酸序列的表达的条件下使宿主细胞生长来制备。重组表达的单体或构建体可以自组装到在所述宿主细胞膜中的孔中。替代地,以这种方式制备的重组孔可以从所述宿主细胞中移出并插入到另一膜中。当产生包含至少两种不同单体或构建体的孔时,所述不同的单体或构建体可以如上所述在不同宿主细胞中分别表达,从宿主细胞中移出并组装到分离的膜中的孔中,例如兔细胞膜或合成膜。

[0460] 所述载体可以为例如提供有复制起点的质粒,病毒或噬菌体载体,任选地用于表达所述多核苷酸序列的启动子,和任选地所述启动子的调控子。所述载体可以包含一个或多个可选择的标记基因,例如,四环素抗性基因。可以对启动子和其他表达调节信号进行选择,使其与设计表达载体所针对的宿主细胞相容。通常使用T7,trc,lac,ara或λ_L启动子。

[0461] 所述宿主细胞通常以高水平表达所述单体或构建体。用多核苷酸序列转化的宿主细胞被选择为,与用于转化所述细胞的表达载体相容。所述宿主细胞通常为细菌并优选大肠杆菌。具有λDE3溶源菌(lysogen)的任何细胞,例如C41(DE3),BL21(DE3),JM109(DE3),B834(DE3),TUNER,Origami和Origami B,可以表达包含T7启动子的载体。除上述条件之外,在Cao et al,2014,PNAS,Structure of the nonameric bacterial amyloid secretion channel,doi-1411942111和Goyal et al,2014,Nature,516,250-253structural and mechanistic insights into the bacterial amyloid secretion channel CsgG中引用的任何方法也可以用于表达CsgG蛋白。

[0462] 本发明还包括产生本发明突变单体或本发明构建体的方法。所述方法包括在合适的宿主细胞中表达本发明的多核苷酸。所述多核苷酸优选为载体的一部分并且优选可操作地连接到启动子上。

[0463] 孔

[0464] 本发明还提供多种孔。本发明所述的孔理想的用于表征多核苷酸序列,例如测序多核苷酸序列,因为它们可以高灵敏度区分不同的核苷酸。所述孔可以出人预料的区分DNA和RNA中的四种核苷酸。本发明所述孔甚至可以区分甲基化和未甲基化的核苷酸。本发明所述孔的碱基分辨率出人预料的高。所述孔显示对全部四种DNA核苷酸的几乎完全分离。所述孔基于在所述孔中的停留时间和流过所述孔的电流,进一步区分脱氧胞嘧啶单磷酸(dCMP)和甲基-dCMP。

[0465] 本发明所述孔还可以在一系列条件下区分不同核苷酸。特别地,所述孔在有利于核酸表征如测序的条件下区分核苷酸。通过改变施加的电势,盐浓度,缓冲液,温度和添加剂如脲,甜菜碱和DTT的存在,可以控制本发明所述的孔区分不同核苷酸的程度。这允许孔的功能被精细调控,特别是在测序时。这将在下面更详细地讨论。本发明所述的孔也可以用于通过与一种或多种单体的相互作用而不是在以核苷酸为基础的核苷酸上,来鉴定多核苷酸聚合物。

[0466] 本发明所述孔可以是分离的,基本上分离的,纯化的或基本纯化的。如果完全不含任何其它组分,例如脂质体或其它孔,则本发明所述的孔被分离或纯化。如果孔与不会干扰其预期用途的载体或稀释剂混合,则该孔被基本上被分离。例如,如果孔以包含小于10%,小于5%,小于2%或小于1%的其它组分如三嵌段共聚物,脂质体或其它孔的形式存在,则

所述孔基本上被分离或基本上被纯化。替代地,本发明所述孔可以存在于膜中。合适的膜如下所述。

[0467] 本发明的孔可以以单独的或单一的孔存在。替代地,本发明所述孔可以存在于两个或更多个孔的同源或异源群体中。

[0468] 同源寡聚孔

[0469] 本发明还提供一种衍生自包含相同本发明突变单体的CsgG的同源寡聚孔。所述同源寡聚孔可以包括本发明的任意突变体。本发明所述的同源寡聚孔理想的用于多核苷酸表征,例如测序。本发明所述同源寡聚孔可以具有上述的任何优点。

[0470] 所述同源寡聚孔可以包含任意数量的突变单体。所述孔一般包含至少7,至少8,至少9或至少10个相同的突变单体,例如,7,8,9或10个突变单体。所述孔优选包含8或9个相同的突变单体。所述突变单体的一个或多个,例如2,3,4,5,6,7,8,9或10个,优选如上述讨论的被化学修饰。

[0471] 下面更详细的讨论用于制备所述孔的方法。

[0472] 异源寡聚孔

[0473] 本发明还提供衍生自包含至少一种本发明的突变单体的CsgG的异源寡聚孔。本发明所述的异源寡聚孔理想的用于多核苷酸表征,例如测序。异源寡聚孔可以用本领域已知的方法制备(例如Protein Sci.2002Jul;11(7)1813-24)。

[0474] 所述异源寡聚孔包含足够的单体以形成所述孔。所述单体可以为任何种类。所述孔一般包含至少7,至少8,至少9或至少10个单体,例如,7,8,9或10个单体。所述孔优选包含8或9个单体。

[0475] 在优选的实施方式中,所有所述单体(例如10,9,8或7个单体)为本发明的突变单体,并且其中至少一个单体与其他单体不同。在更优选的实施方式中,所述孔包括八个或九个本发明所述的突变单体,并且其中至少一个单体与其他单体不同。它们也可以全部彼此不同。

[0476] 在所述孔中本发明的突变单体优选为具有大约相同的长度或具有相同的长度。所述孔中本发明所述突变单体的桶状结构优选为具有大约相同的长度或具有相同的长度。长度可以用氨基酸数量和/或长度单元进行测量。

[0477] 在另一优选的实施方式中,至少一个所述突变单体不是本发明的突变单体。在这个实施方式中,剩下的单体优选为本发明所述的突变单体。因此,所述孔可以包括9,8,7,6,5,4,3,2或1个本发明的突变单体。所述孔中任意数量的单体可以不是本发明的突变单体。所述孔优选包括七个或八个本发明的突变单体以及一个不是本发明的单体的单体。本发明所述的突变单体可以是相同或不同的。

[0478] 所述构建体中本发明所述突变单体优选为具有大约相同的长度或具有相同的长度。所述构建体中本发明所述突变单体的桶状结构优选为具有大约相同的长度或具有相同的长度。长度可以用氨基酸数量和/或长度单元进行测量。

[0479] 所述孔可以包括一个或多个不是本发明突变单体的单体。作为不是本发明突变单体的CsgG单体包括包含SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的单体或者SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,414,421,422,423,424,425,426,427,428或429

的对照变体,其中所有如上所述的与本发明相关的氨基酸/位置均没有被突变/取代。SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的对照变体通常为与SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429在其整个序列上基于氨基酸的同一性至少50%同源。更优选,所述对照变体可以为基于氨基酸同一性,与SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的氨基酸序列在整个序列上,至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%以及更优选至少95%,97%或99%同源。

[0480] 在如上所述的所有实施方式中,一个或更多,例如2,3,4,5,6,7,8,9或10个,所述突变单体优选如上所述进行化学修饰。

[0481] 下面更详细的讨论用于制备所述孔的方法。

[0482] 含有构建体的孔

[0483] 本发明还提供一种包括至少一个本发明构建体的孔。本发明的构建体包括两个或更多个共价连接的衍生自CsgG的单体,其中至少一个所述单体为本发明的突变单体。也就是说,构建体必须包含多于一个的单体。所述孔含有足够的构建体和,如果必要,足够的单体以形成所述孔。例如,八聚体孔可以包含(a)四个构建体,每个构建体包含两个单体,(b)两个构建体,每个构建体包含四个单体或(b)一个构建体,其包含两个单体和六个不构成构建体一部分的单体。例如,九聚体孔可以包括(a)各自包含两个单体的四个构建体以及一个不形成构建体一部分的单体,(b)各自包含四个单体的两个构建体以及一个不形成构建体一部分的单体或(b)一个包含两个单体的构建体和七个不构成构建体一部分的单体。本领域技术人员可以设想构建体和单体的其他组合。

[0484] 至少两个在所述孔中的所述单体为本发明构建体的形式。所述构建体,以及因此所述孔,包括至少一个本发明的突变单体。所述孔一般共包含至少7,至少8,至少9或至少10个单体,例如,7,8,9或10个单体(其中的至少两个必须在构建体中)。所述孔优选包括八个或九个单体(其中的至少两个必须在构建体中)。

[0485] 包含构建体的孔可以为同源寡聚体(即包括相同的构建体)或为异源寡聚体(即其中至少一个构建体与其他构建体不同)。

[0486] 孔通常包含(a)包含两个单体的一个构建体和(b)5,6,7或8个单体。所述构建体可以是如上所述的任意一种。所述单体可以为如上所述任意一种,包括本发明的突变单体,包含SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的单体以及包含如上所述SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,414,421,422,423,424,425,426,427,428或429的对照变体的突变单体。

[0487] 另一种典型的孔包括多于一个的本发明构建体,例如两个,三个或四个本发明的构建体。如果必要,所述孔进一步包含足够的额外单体或构建体以形成所述孔。所述额外单体可以为如上所述任意一种,包括本发明的突变单体,包含SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的单体以及包含如上所述SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,414,421,422,423,424,425,426,427,428或429的对照变体的突变单体。所述额外构建体可

以为如上所述任意一种或者可以为包括两个或更多个共价连接的CsgG单体的构建体,所述CsgG单体各自包括包含SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429的单体以及包含如上所述SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,414,421,422,423,424,425,426,427,428或429的对照变体的单体。

[0488] 本发明进一步的孔仅包括包含两个单体的构建体,例如孔可以包括4,5,6,7或8个包含两个单体的构建体。至少一个构建体为本发明的构建体,即,在至少一个构建体中的至少一个单体,并且优选在至少一个构建体中的每个单体,为本发明的突变单体。所有包含两个单体的构建体可以为本发明的构建体。

[0489] 本发明的特定孔包括四个本发明的构建体,每个构建体包含两个单体,其中每个构建体中的至少一个单体,并且优选每个构建体中的每个单体,为本发明的突变单体。所述构建体可以寡聚成具有使得每个构建体中只有一个单体有助于孔通道的结构的孔。通常地,所述构建体中的其他单体将在所述孔的通道的外部。例如,本发明的孔可以包括7,8,9或10个包含两个单体的构建体,其中所述通道包括7,8,9或10个单体。

[0490] 突变可以引入到如上所述的构建体中。所述突变可以是交替的,即所述突变对两个单体构建体中的每个单体均是不同的,并且所述构建体被组装为导致交替修饰的同源寡聚体。也就是说,包含MutA和MutB的单体被融合并且被组装以形成A-B:A-B:A-B:A-B孔。替代地,所述突变可以是相邻的,即将相同的突变引入一个构建体中的两个单体中,然后与不同的突变单体或构建体进行寡聚。也就是说,包含MutA的单体进行融合,然后与含MutB的单体进行寡聚以形成A-A:B:B:B:B:B:B。

[0491] 在含有构建体的孔中一个或多个本发明的单体可以如上所述进行化学修饰。

[0492] 分析物的表征

[0493] 本发明提供一种确定目标分析物的存在、不存在或一种或多种特性的方法。该方法涉及将所述目标分析物与CsgG孔或其突变体(例如本发明的孔)接触,使得所述目标分析物相对于,例如穿过,所述孔移动,并且当所述分析物相对于所述孔移动时获取一个或多个测量值,从而确定所述分析物的存在、不存在或一种或多种特性。所述目标分析物也可以被称为模板分析物或感兴趣的分析物。

[0494] 所述方法包括将所述目标分析物与CsgG孔或其突变体(例如本发明的孔)接触,使得所述目标分析物移动穿过所述孔。所述孔一般包含至少7,至少8,至少9或至少10个单体,例如,7,8,9或10个单体。所述孔优选包含8或9个相同的单体。所述单体中的一个或多个,例如2,3,4,5,6,7,8,9或10个,优选如上述讨论的被化学修饰。

[0495] 所述CsgG孔可以来源于任何生物体。所述CsgG孔可以包含包括SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429中所示序列的单体。所述CsgG孔可以包含至少7,至少8,至少9或至少10个单体,例如,7,8,9或10个单体,每个单体包括SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429中所示序列(即所述孔是包含来自相同生物体的相同单体的同源寡聚物)。所述CsgG孔可以包含单体的任意组合,每个所述单体包括SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429中所示序列。例如,所述孔可以包含7个包括SEQ ID

NO:390中所示序列的单体和2个包括SEQ ID NO:391中所示序列的单体。

[0496] 所述CsgG突变体可以包含任意数量的突变单体,例如至少7,至少8,至少9或至少10个单体,例如,7,8,9或10个单体。所述突变单体可以包括SEQ ID NO:390,391,392,393,394,395,414,415,416,417,418,419,420,421,422,423,424,425,426,427,428或429中所示序列的比较变体。比较变体已经在前文中讨论过。比较变体必须能够形成孔,并可以相比于本发明所述孔具有任意百分比的前文讨论的同源性。

[0497] 所述CsgG突变体优选包含9个单体并且所述单体中至少一个是SEQ ID NO:390中所示序列的变体,所述变体包括:(a)在一个或多个以下位置的突变:N40,Q42,D43,E44,K49,Y51,S54,N55,F56,S57,Q62,E101,N102,E124,E131,R142,D149,T150,E185,R192,D195,E201和E203,例如在一个或多个以下位置的突变:N40,Q42,D43,E44,K49,Y51,S54,N55,F56,S57,Q62,E101,N102,E131,D149,T150,E185,D195,E201和E203,或一个或多个在一个或多个以下位置的突变:N40,Q42,D43,E44,K49,Y51,N55,F56,E101,N102,E131,D149,T150,E185,D195,E201和E203;和/或(b)一个或多个以下位置的缺失:F48,K49,P50,Y51,P52,A53,S54,N55,F56和S57。所述变体可以包含(a);(b);或(a)和(b)。在(a)中,任意数量和组合的以下位置:N40,Q42,D43,E44,K49,Y51,S54,N55,F56,S57,Q62,E101,N102,R124,E131,R142,D149,E185,R192,D195,E201以及E203,可以发生突变。如前文所讨论的,Y51,N55和F56中的一个或多个发生突变可以减少当多核苷酸移动穿过所述孔时对电流有贡献的核苷酸的数量,从而使得更容易识别当多核苷酸移动通过所述孔时观察到的电流与多核苷酸之间的直接关系。

[0498] 在(b)中,任意数量和组合的以下位置:F48,K49,P50,Y51,P52,A53,S54,N55,F56和S57可以缺失。所述变体可以包含前文针对本发明突变单体讨论的任意特定突变/取代或其组合。

[0499] 用于本发明所述方法的优选变体包括一个或多个以下取代:(a)F56N,F56Q,F56R,F56S,F56G,F56A或F56K或F56A,F56P,F56R,F56H,F56S,F56Q,F56I,F56L,F56T或F56G;(b)N55Q,N55R,N55K,N55S,N55G,N55A或N55T;(c)Y51L,Y51V,Y51A,Y51N,Y51Q,Y51S或Y51G;(d)T150I;(e)S54P;以及(f)S57P。所述变体可以包含(a)至(f)的任意数量和组合。

[0500] 用于本发明所述方法的优选变体包含Q62R或Q62K。

[0501] 用于本发明所述方法的优选变体包含在D43,E44,Q62或其任意组合的突变,例如D43,E44,Q62,D43/E44,D43/Q62,E44/Q62或D43/E44/Q62。

[0502] 所述变体可以包含以下位置的突变:

[0503] Y51,F56,D149,E185,E201和E203,例如Y51N,F56A,D149N,E185R,E201N和E203N;

[0504] N55,例如N55A或N55S;

[0505] Y51,例如Y51N或Y51T;

[0506] S54,例如S54P;

[0507] S57,例如S57P;

[0508] F56,例如F56N,F56Q,F56R,F56S,F56G,F56A或F56K或F56A,F56P,F56R,F56H,F56S,F56Q,F56I,F56L,F56T或F56G;

[0509] Y51和F56,例如Y51A和F56A,Y51A和F56N,Y51I和F56A,Y51L和F56A,Y51T和F56A,Y51T和F56Q,Y51I和F56N,Y51L和F56N或Y51T和F56N,优选,Y51I和F56A,Y51L和F56A或Y51T

和F56A,更优选Y51T和F56Q或进一步优选Y51X和F56Q,其中X是任意一种氨基酸;

[0510] N55和F56,例如N55X和F56Q,其中X是任意一种氨基酸;

[0511] Y51,N55和F56,例如Y51A,N55S和F56A,Y51A,N55S和F56N或Y51T,N55S和F56Q;

[0512] S54和F56,例如S54P和F56A或S54P和F56N;

[0513] F56和S57,例如F56A和S57P或F56N和S57P;

[0514] D149,E185和E203,例如D149N,E185N和E203N;

[0515] D149,E185,E201和E203,例如D149N,E185N,E201N和E203N;

[0516] D149,E185,D195,E201和E203,例如D149N,E185R,D195N,E201N和E203N或D149N,E185R,D195N,E201R和E203N;

[0517] F56和N102,例如F56Q和N102R;

[0518] (a)Q62,例如Q62R或Q62K,和(b)Y51,N55和F56的一个或多个,例如在Y51,N55,F56,Y51N55,Y51F56,N55F56或Y51N55F56,例如Y51TF56QQ62R;

[0519] (i)D43,E44,Q62或其任意组合,例如D43,E44,Q62,D43/E44,D43/Q62,E44/Q62或D43/E44/Q62和(ii)Y51,N55和F56的一种或多种,例如在Y51,N55,F56,Y51/N55,Y51/F56,N55/F56或Y51/N55/F56,例如D43N/Y51T/F56Q,E44N/Y51T/F56Q,D43N/E44N/Y51T/F56Q,D43N/Y51T/F56Q/Q62R,E44N/Y51T/F56Q/Q62R或D43N/E44N/Y51T/F56Q/Q62R;或

[0520] T150,例如T150I。

[0521] 用于本发明方法的优选孔包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包括一个或多个以下取代:(a)F56N,F56Q,F56R,F56S,F56G,F56A或F56K或F56A,F56P,F56R,F56H,F56S,F56Q,F56I,F56L,F56T或F56G;(b)N55Q,N55R,N55K,N55S,N55G,N55A或N55T;(c)Y51L,Y51V,Y51A,Y51N,Y51Q,Y51S或Y51G;(d)T150I;(e)S54P;以及(f)S57P。所述变体可以包含(a)至(f)中的任意数量和组合。所述单体优选在这些优选的孔中是相同的。

[0522] 用于本发明所述方法的优选孔包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包括:

[0523] Y51N,F56A,D149N,E185,E201N和E203N;

[0524] N55A;

[0525] N55S;

[0526] Y51N;

[0527] S54P;

[0528] S57P;

[0529] F56N,F56Q,F56R,F56S,F56G,F56A或F56K;

[0530] F56A,F56P,F56R,F56H,F56S,F56Q,F56I,F56L,F56T或F56G;

[0531] Y51A和F56A;

[0532] Y51A和F56N;

[0533] Y51I和F56A;

[0534] Y51L和F56A;

[0535] Y51T和F56A;

[0536] Y51T和F56Q;

[0537] Y51I和F56N;

[0538] Y51L和F56N;

[0539] Y51T和F56N;

[0540] N55S和F56Q;

[0541] Y51A,N55S和F56A;

[0542] Y51A,N55S和F56N;

[0543] Y51T,N55S和F56Q;

[0544] S54P和F56A;

[0545] S54P和F56N;

[0546] F56A和S57P;

[0547] F56N和S57P;

[0548] D149N,E185N和E203N;

[0549] D149N,E185N,E201N和E203N;

[0550] D149N,E185R,D195N,E201N和E203N;

[0551] D149N,E185R,D195N,E201R和E203N;

[0552] T150I;

[0553] F56Q和N102R;

[0554] F56和N102,例如F56Q和N102R;

[0555] Y51T/F56Q/Q62R;

[0556] D43N/Y51T/F56Q;

[0557] E44N/Y51T/F56Q;

[0558] D43N/E44N/Y51T/F56Q;

[0559] D43N/Y51T/F56Q/Q62R;

[0560] E44N/Y51T/F56Q/Q62R;或

[0561] D43N/E44N/Y51T/F56Q/Q62R。

[0562] 用于本发明方法的所述CsgG突变体优选包含9个单体并且所述单体中的至少一个是SEQ ID NO:390中所示序列的变体,所述变体包括在位置Y51,N55和F56中的一个或多个位置的突变。用于本发明所述方法的优选孔包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含在位置Y51,N55和F56中的一个或多个处的突变。所述单体优选在这些优选的孔中是相同的。所述变体可以包含在位置Y51;N55;F56;Y51/N55;Y51/F56;N55/F56;或Y51/N55/F56的突变。所述变体可以包含上述讨论的位置Y51,N55和F56中的一个或多个位置处的任意特异性突变和其任意组合。Y51,N55和F56中的一个或多个可以用任意氨基酸取代。Y51可以被F,M,L,I,V,A,P,G,C,Q,N,T,S,E,D,K,H或R取代,例如A,S,T,N或Q。N55可以被F,M,L,I,V,A,P,G,C,Q,,T,S,E,D,K,H或R取代,例如A,S,T,或Q。F56可以被M,L,I,V,A,P,G,C,Q,N,T,S,E,D,K,H或R取代,例如A,S,T,N或Q。所述变体可以进一步包含以下中的一个或多个:(i)在以下位置的一个或多个突变(即在一个或多个以下位置的突变)(i)N40,D43,E44,S54,S57,Q62,R97,E101,E124,E131,R142,T150和R192;(iii)Q42Ror 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;以及(ix)缺失一个或多个以下位置:F48,K49,P50,Y51,P52,A53,S54,N55,F56和S57。所述变体可以包含上述讨论的(i)和(iii)至(xi)的任意组合。所述变体可以包含上述针对(i)和(iii)至(xi)讨论的的任意实施方式。

[0563] (1)用于本发明所述方法的优选变体或(2)用于本发明所述方法的优选孔,包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含: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。

[0564] (1)用于本发明所述方法的优选突变体或(2)用于本发明所述方法的优选孔,包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含:Y51TF56Q,Y51QF56Q或Y51A/F56Q。

[0565] (1)用于本发明所述方法的优选突变体或(2)用于本发明所述方法的优选孔,包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含: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。

[0566] (1)用于本发明所述方法的优选变体或(2)用于本发明所述方法的优选孔,包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含:Y51T/N55Q,Y51T/N55S或Y51T/N55A。

[0567] (1)用于本发明所述方法的优选变体包含或(2)用于本发明所述方法的优选孔,包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含: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。

[0568] (1)用于本发明所述方法的优选变体或(2)用于本发明所述方法的优选孔,包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含:Y51C/F56A,Y51E/F56A,Y51D/F56A,Y51K/F56A,Y51H/F56A,Y51Q/F56A,Y51N/F56A,Y51S/F56A,Y51P/F56A或Y51V/F56A。

[0569] (1)用于本发明所述方法的优选变体或(2)用于本发明所述方法的优选孔,包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含:

[0570] D149R/E185R/E201R/E203R或Y51T/F56Q/D149R/E185R/E201R/E203R;

[0571] D149N/E185N/E201N/E203N或Y51T/F56Q/D149N/E185N/E201N/E203N;

[0572] E201R/E203R或Y51T/F56Q/E201R/E203R

[0573] E201N/E203R或Y51T/F56Q/E201N/E203R;

[0574] E203R或Y51T/F56Q/E203R;

[0575] E203N或Y51T/F56Q/E203N;

[0576] E201R或Y51T/F56Q/E201R;

[0577] E201N或Y51T/F56Q/E201N;

[0578] E 185R或Y51T/F56Q/E 185R;

[0579] E185N或Y51T/F56Q/E185N;

[0580] D149R或Y51T/F56Q/D149R;

[0581] D149N或Y51TF56Q/D149N;

[0582] R142E或Y51T/F56Q/R142E;

[0583] R142N或Y51T/F56Q/R142N;

[0584] R192E或Y51T/F56Q/R192E;或者

[0585] R192N或Y51T/F56Q/R192N;

[0586] (1)用于本发明所述方法的优选变体或(2)用于本发明所述方法的优选孔,包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含:

[0587] Y51A/F56Q/E101N/N102R;

[0588] Y51A/F56Q/R97N/N 102G;

[0589] Y51A/F56Q/R97N/N102R;

[0590] Y51A/F56Q/R97N;

[0591] Y51A/F56Q/R97G;

[0592] Y51A/F56Q/R97L;

[0593] Y51A/F56Q/N102R;

[0594] Y51A/F56Q/N102F;

[0595] Y51A/F56Q/N102G;

[0596] Y51A/F56Q/E101R;

[0597] Y51A/F56Q/E101F;

[0598] Y51A/F56Q/E101N;或

[0599] Y51A/F56Q/E101G。

[0600] 所述单体优选在这些优选的孔中是相同的。

[0601] 用于本发明所述方法的优选孔包含至少7,至少8,至少9或至少10个单体,例如7,8,9或10个单体,每个所述单体包括SEQ ID NO:390的变体,所述变体包含F56A,F56P,F56R,F56H,F56S,F56Q,F56I,F56L,F56T或F56G。所述单体优选在这些优选的孔中是相同的。

[0602] 所述CsgG突变体可以包含实施例中公开的SEQ ID NO:390的任何变体或者可以是实施例中公开的任何孔。

[0603] 所述CsgG突变体最优选为本发明的孔。

[0604] 步骤(a)和(b)优选在跨孔施加电势的情况下进行。如下更详细讨论的内容,所施加的电势一般导致在所述孔和多核苷酸结合蛋白之间形成复合物。所述施加的电势可以是电压电势。替代地,所施加的电势可以是化学电势。其一个例子是跨越双亲性分子层使用盐梯度。盐梯度被公开在Holden et al.,J Am Chem Soc.2007Jul 11;129(27)8650-5中。

[0605] 所述方法用于确定目标分析物的存在、不存在或一种或多种特性。所述方法可以用于确定至少一种目标分析物的存在、不存在或一种或多种特性。所述方法可以涉及确定两种或更多种目标分析物的存在、不存在或一种或多种特性。所述方法可以包括确定任意数量目标分析物的存在、不存在或一种或多种特性,例如2,5,10,15,20,30,40,50,100或更

多种目标分析物。一种或多种目标分析物的任意数量的特性可以被确定,例如1,2,3,4,5,10或更多种特性。

[0606] 所述目标分析物优选为金属离子,无机盐,聚合物,氨基酸,肽,多肽,蛋白,核苷酸,寡核苷酸,多核苷酸,染料,漂白剂,药物,诊断剂,爆炸物或环境污染物。所述方法可以涉及确定两种或更多种相同类别的目标分析物的存在、不存在或一种或多种特性,例如,两种以上蛋白,两种以上核苷酸或两种以上药物。或者,所述方法可以涉及确定两种或更多种不同类别的目标分析物的存在、不存在或一种或多种特性,例如,一种或多种蛋白,一种或多种核苷酸和一种或多种药物。

[0607] 所述目标分析物可以通过细胞分泌。或者,所述目标分析物可以是存在细胞内的分析物使得在进行本发明前所述分析物必须被从细胞中抽提出来。

[0608] 所述分析物优选为氨基酸,肽,多肽和/或蛋白。所述氨基酸,肽,多肽或蛋白可以是天然存在的或非天然存在的。所述多肽或蛋白可以在其中包含合成或修饰的氨基酸。对氨基酸的大量不同类型的修饰是本领域已知的。合适的氨基酸和对其的修饰如上所述。为了本发明的目的,可以理解的是,目标分析物可以通过任何本领域已知方法进行修饰。

[0609] 所述蛋白可以是酶,抗体,激素,生长因子或生长调节蛋白,例如细胞因子。所述细胞因子可以选自白介素,优选IFN-1,IL-1,IL-2,IL-4,IL-5,IL-6,IL-10,IL-12和IL-13,干扰素,优选IL-γ,以及其他细胞因子,例如TNF-a。所述蛋白可以是细菌蛋白,真菌蛋白,病毒蛋白或寄生虫源蛋白。

[0610] 所述目标分析物优选为核苷酸,寡核苷酸或多核苷酸。以下讨论核苷酸和多核苷酸。寡核苷酸是短链核苷酸聚合物,一般具有50个或更少的核苷酸,例如40个或更少,30个或更少,20个或更少,10个或更少或5个个或更少的核苷酸。所述寡核苷酸可以包括以下讨论的任何核苷酸,包括无碱基和修饰的核苷酸。

[0611] 所述目标分析物,例如目标多核苷酸,可以存在于下面讨论的任何合适的样品中。

[0612] 如下所述,所述孔一般存在于膜中。可以采用如下讨论的方法将目标分析物偶联或传递至所述膜上。

[0613] 如下讨论的任何测量方法可以用于确定目标分析物的存在、不存在或一种或多种特性。所述方法优选包括将所述目标分析物与所述孔接触使得所述分析物相对于(例如移动穿过)所述孔移动,并且测量当所述分析物相对于所述孔移动时穿过所述孔的电流,从而确定所述分析物的存在、不存在或一种种或多种特性。

[0614] 如果所述电流以对分析物特异性的方式流过孔(即如果检测到与所述分析物相关的特异性电流流过所述孔),则目标分析物存在。如果电流没有以对核苷酸特异性的方式流过所述孔,则所述分析物不存在。对照实验可以在所述分析物存在下进行以确定是否影响电流流过所述孔的方式。

[0615] 本发明可以用于区分相似结构的分析物——基于它们对穿过孔的电流的不同影响。各个分析物可以从它们与所述孔相互作用时它们的电流幅值在单个分子水平上进行识别。本发明也可以用于确定样品中是否存在特定分析物。本发明也可以用于测量样品中特定分析物的浓度。使用除CsgG之外的孔进行分析物表征是本领域已知的。

[0616] 多核苷酸的表征

[0617] 本发明提供一种表征目标多核苷酸的方法,例如多核苷酸测序。存在两种使用纳

米孔表征或测序多核苷酸的主要策略,即链表征/测序和外切核酸酶表征/测序。本发明所述方法可以涉及任何一种方法。

[0618] 在链测序中,DNA顺着或逆着所施加的电势位移通过纳米孔。以渐进方式或前进方式(processively)作用在双链DNA上的外切核酸酶可以用在所述孔的顺式侧以在施加电势条件下将剩余单链送入并通过所述孔,或在反向电势条件下在所述孔的反式侧上将剩余单链送入并通过所述孔。相似地,解开双链DNA的解旋酶也可以采用类似的方式使用。也可以使用聚合酶。还可能用于要求逆着施加电势进行链移位的测序应用,但是所述DNA必须在反向电势或无电势下首先被酶“捕获”。然后随着电势在结合后切换回来,所述链将从顺式到反式穿过所述孔并且通过电流保持处于伸展构象。单链DNA外切核酸酶或单链DNA依赖性聚合酶可以作为分子马达,以受控的逐步方式,逆着施加的电势,从反式到顺式,将刚刚移位的单链拉回并通过所述孔。

[0619] 在一种实施方式中,所述表征目标多核苷酸的方法包括使目标序列与孔和解旋酶接触。任何解旋酶均可以用于所述方法。以下讨论合适的解旋酶。解旋酶可以相对于所述孔以两种模式工作。第一,所述方法优选使用解旋酶进行,使其通过由施加电压产生的场控制目标序列通过所述孔的运动。在这种模式中,首先将DNA的5’端捕获到所述孔中,并且所述酶控制DNA移动进入所述孔中,使得所述目标序列在所述电场下通过所述孔直到其最终移位到双分子层的反式侧。或者,所述方法优选这样进行:使得解旋酶控制所述目标序列逆着由施加电压产生的场通过所述孔的运动。在这种模式中,DNA的3’端首先被捕获到所述孔中,并且所述酶控制DNA通过所述孔的运动,使得所述目标序列逆着所施加的场被拉出所述孔直到其最终弹回到双分子层的顺式侧。

[0620] 在核酸外切酶测序中,核酸外切酶从目标多核苷酸的一端释放各单个核苷酸,并且如下所述鉴定这些单个核苷酸。在另一种实施方式中,表征目标多核苷酸的方法包括使目标序列与孔和核酸外切酶接触。如下所述的任意核酸外切酶均可以用于所述方法。所述酶可以如下所述共价连接到所述孔上。

[0621] 核酸外切酶是通常锁定到多核苷酸的一端并从该端一次消化所述序列的一个核苷酸的酶。所述核酸外切酶可以在5’至3’方向或3’至5’方向消化所述多核苷酸。多核苷酸的结合有核酸外切酶的末端通常通过选择所使用的酶和/或采用本领域已知的方法决定。多核苷酸任意一端的羟基或盖帽结构通常可以用于预防或促进核酸外切酶与多核苷酸特定末端的结合。

[0622] 所述方法包括将所述多核苷酸与所述核酸外切酶接触,使得所述核苷酸从所述多核苷酸的末端以允许表征或鉴别如上所述的一部分核苷酸的速率被消化。进行这些步骤的方法是本领域公知的。例如,埃德曼(Edman)降解用于从多肽的末端顺序消化单个氨基酸,使得它们可以使用高效液相色谱(HPLC)进行鉴定。同源性的方法可以用于本发明。

[0623] 所述核酸外切酶作用的速率一般低于野生型核酸外切酶的最优速率。本发明所述方法中的所述核酸外切酶活性的合适速率包括每秒消化0.5到1000个核苷酸,每秒消化0.6到500个核苷酸,每秒消化0.7到200个核苷酸,每秒消化0.8到100个核苷酸,每秒消化0.9到50个核苷酸或每秒消化1到20或10个核苷酸。所述速率优选为每秒1,10,100,500或1000个核苷酸。核酸外切酶活性的合适速率可以通过多种方式实现。例如,根据本发明可以使用具有降低的最佳活性速率的不同的核酸外切酶。

[0624] 在链表征的实施方式中,所述方法包括使所述多核苷酸与CsgG孔或其突变体,例如本发明的孔,接触,使得所述多核苷酸相对于所述孔移动,例如通过所述孔,并且当所述多核苷酸相对于所述孔移动时,获取一个或多个测量值,其中所述测量值指示所述多核聚苷酸的一个或多个特征,从而表征所述目标多核苷酸。

[0625] 在核酸外切酶表征的实施方式中,所述方法包括使所述多核苷酸与CsgG孔或其突变体,例如本发明中的孔,以及核酸外切酶接触,使得所述核酸外切酶从所述目标多核苷酸的一端消化各个核苷酸并且所述各个核苷酸相对于所述孔移动,例如通过所述孔,并且当所述各个核苷酸相对于所述孔移动时获取一个或多个测量值,其中所述测量值指示所述各个核苷酸的一个或多个特征,从而表征所述目标多核苷酸。

[0626] 各个核苷酸是单个的核苷酸。各个核苷酸是不通过核苷酸键与另一个核苷酸或多核苷酸结合的核苷酸。核苷酸键包括与另一核苷酸的糖基结合的核苷酸的磷酸基团之一。各个核苷酸通常不通过核苷酸键与另一个具有至少5个,至少10个,至少20个,至少50个,至少100个,至少200个,至少500个,至少1000个或至少5000个核苷酸的多核苷酸结合。例如,所述各个核苷酸已经从目标多核苷酸序列被消化,例如,DNA或RNA链。所述核苷酸可以是如下所述的任意一种。

[0627] 所述各个核苷酸可以与所述孔以任意方式在任意位点相互作用。所述核苷酸优选通过适配器或与适配器连接而可逆地与所述孔结合,如上文所述。所述核苷酸最优选在它们跨膜穿过所述孔时通过适配器或与适配器连接而可逆地与所述孔结合。所述核苷酸也可以当它们跨膜穿过所述孔时通过适配器或与所述适配器接合而可逆地与所述孔的桶状结构或通道结合。

[0628] 在所述各个核苷酸与所述孔的相互作用中,所述核苷酸通常以对核苷酸特异性的方式影响流过所述孔的电流。例如,一种特定的核苷酸将以特定的平均时间周期和特定的程度减少流过所述孔的电流。也就是说,流过所述孔的电流对特定核苷酸是特有的。可以进行对照实验来确定特定核苷酸对流过所述孔的电流的影响。然后将对测试样品实施本发明方法得到的结果与来自上述对照实验的结果进行比较,以便鉴定样品中的特定核苷酸或确定样品中是否存在特定核苷酸。以指明特定核苷酸的方式影响流过所述孔的电流的频率可用于确定样品中所述核苷酸的浓度。样品中不同核苷酸的比例也可以计算得出。例如,dCMP对甲基-dCMP的比例可以计算得出。

[0629] 所述方法包括测量所述目标多核苷酸的一种或多种特性。所述目标多核苷酸也可以被称为模板多核苷酸或感兴趣的多核苷酸。

[0630] 这个实施方式也可以使用CsgG孔或其突变体,例如本发明的孔。上文针对目标分析物所述的任意孔和实施方式均可以使用。

[0631] 多核苷酸

[0632] 多核苷酸,例如核酸,是包含两个或多个核苷酸的大分子。所述多核苷酸或核酸可以包含任意核苷酸的任意组合。所述核苷酸可以是天然存在的或人工合成的。所述多核苷酸中的一个或多个核苷酸可以被氧化或甲基化。所述多核苷酸中的一个或多个核苷酸可以被损伤。例如,所述多核苷酸可以包含嘧啶二聚体。这种二聚体通常与由紫外线造成的损伤有关并且是皮肤黑色素瘤的主要成因。所述多核苷酸中的一个或多个核苷酸可以被修饰,例如用标记或标签。合适的标记如下所述。所述多核苷酸可以包含一个或多个间隔区。

[0633] 核苷酸通常含有核碱基,糖和或至少一个磷酸基团。所述核碱基和糖形成核苷。

[0634] 所述核碱基通常是杂环的。核碱基包括,但不限于,嘌呤和嘧啶,更具体的是腺嘌呤(A),鸟嘌呤(G),胸腺嘧啶(T),尿嘧啶(U)和胞嘧啶(C)。

[0635] 所述糖通常是戊糖。核苷糖包括,但不限于,核糖和脱氧核糖。所述糖优选是脱氧核糖。

[0636] 所述多核苷酸优选包括如下核苷:脱氧腺苷(dA),脱氧尿苷(dU)和/或脱氧胸苷(dT),脱氧鸟苷(dG)和脱氧胞苷(dC)。

[0637] 所述核苷酸通常是核糖核苷酸或脱氧核糖核苷酸。所述核苷酸通常含有单磷酸,二磷酸或三磷酸。所述核苷酸可以包含多于三个磷酸,例如4或5个磷酸。磷酸可以连接在核苷酸的3'侧或5'侧。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-甲基胞苷,单磷酸5-羟基甲基胞苷,单磷酸胞苷(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。所述核苷酸优选选自AMP,TMP,GMP,CMP,UMP,dAMP,dTMP,dGMP,dCMP和dUMP。

[0638] 核苷酸可以是无碱基的(即缺少核碱基)。核苷酸也可以缺少核碱基和糖(即为C3间隔区)。

[0639] 所述多核苷酸中的核苷酸可以任意方式相互连接。所述核苷酸通常通过其糖基和磷酸基团连接,如在核酸中一样。所述核苷酸可以通过其核碱基连接,如在嘧啶二聚体中一样。

[0640] 所述多核苷酸可以是单链或双链的。所述多核苷酸的至少一部分优选是双链的。

[0641] 所述多核苷酸可以是核酸,例如脱氧核糖核酸(DNA)或核糖核酸(RNA)。所述多核苷酸可以包含一条RNA链,所述RNA链杂合到一条DNA链。所述多核苷酸可以是任意现有技术已知的合成核酸,例如肽核酸(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他合成聚合物。所述PNA骨架是由通过肽键连接的重复的N-(2-氨基乙基)-甘氨酸单元组成。所述GNA骨架是由通过磷酸二酯键连接的重复乙二醇单元组成。所述TNA骨架是由通过磷酸二酯键连接在一起的重复苏糖基组成。LNA由上述核糖核酸形成,具有连接核糖部分中2'氧和4'碳的额外桥连结构。桥连的核酸(BNA)是修饰的RNA核苷酸。它们也可以称为限制的或不可接近的RNA。BNA单体可以含有5元,6元或甚至7元桥连结构并带有"固定的"C3'-内糖折叠结构(C3'-endo sugar puckering)。所述桥连结构被合成引入核糖的2',4'-位,以产生2',4'-BNA单体。

[0642] 所述多核苷酸最优选核糖核酸(RNA)或脱氧核糖核酸(DNA)。

[0643] 所述多核苷酸可以为任意长度。例如,所述多核苷酸的长度可以是至少10,至少50,至少100,至少150,至少200,至少250,至少300,至少400或至少500个核苷酸或核苷酸对。所述多核苷酸的长度可以为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。

[0644] 任意数量的多核苷酸可以被研究。例如本发明方法可以涉及表征2,3,4,5,6,7,8,9,10,20,30,50,100个或更多个多核苷酸。如果两个或更多个多核苷酸被表征,它们可以是不同的多核苷酸或相同多核苷酸的两种情形。

[0645] 所述多核苷酸可以是天然存在的或人工合成的。例如,所述方法可用于验证所制备的寡核苷酸的序列。所述方法通常在体外进行。

[0646] 样品

[0647] 所述多核苷酸通常存在于任意合适的样品中。本发明通常针对已知包含或怀疑含有所述多核苷酸的样品进行。或者,本发明可以针对某一样品进行以确认已知存在或预期存在于所述样品中的多核苷酸的身份。

[0648] 所述样品可以是生物样品。本发明可以针对从任何生物体或微生物中获得或提取的样品在体外实施。所述生物体或微生物通常是古细菌的(archaeal),原核的或真核的,并且通常属于以下五界之一:植物界,动物界,真菌,原核生物和原生生物。本发明针对从任何病毒中获得或提取的样品在体外实施。所述样品优选是流体样品。样品通常包括患者的体液。所述样品可以是尿液,淋巴液,唾液,粘液或羊水,但优选血液,血浆或血清。

[0649] 通常,所述样品来源于人,但替代地可以是来自其他哺乳动物,如自商业上养殖的动物如马,牛,绵羊,鱼,鸡或猪,或者替代地可以是宠物如猫或狗。或者,所述样品可以来源于植物,例如从商业作物获得的样品,如谷类,豆类,水果或蔬菜,例如小麦,大麦,燕麦,芸苔,玉米,大豆,水稻,大黄,香蕉,苹果,番茄,土豆,葡萄,烟草,菜豆,小扁豆,甘蔗,可可,棉花。

[0650] 所述样品可以是非生物样品。所述非生物样品优选为流体样品。非生物样品的实例包括手术液,水如饮用水、海水或河水,以及用于实验室试验的试剂。

[0651] 所述样品通常是在本发明中使用前处理,例如通过离心,或通过膜过滤掉不需要的分子或细胞,例如红细胞。所述样品可在采集后立即测量。样品也可通常在测定前被存储,优选低于-70℃存储。

[0652] 表征

[0653] 所述方法可以涉及测量多核苷酸的两个,三个,四个或五个或更多个特征。所述一个或多个特征优选选自(i)多核苷酸的长度,(ii)多核苷酸的身份,(iii)多核苷酸的序列,(iv)多核苷酸的二级结构,以及(v)多核苷酸是否被修饰。根据本发明可以测量(i)至(v)的任意组合,例如{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},

[0654] {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}。与第二多核苷酸相比,(i)至(v)的不同组合可以测量第一多核苷酸,包括上述所列出的任意组合。

[0655] 对于(i),可以例如通过确定多核苷酸和孔之间相互作用的数量或多核苷酸和孔之间相互作用的持续时间对多核苷酸的长度进行测量。

[0656] 对于(ii),可以以多种方式测量多核苷酸的身份,多核苷酸的身份可以结合多核苷酸序列的测量或不结合多核苷酸序列的测量进行测量。前者较为简单;对多核苷酸进行测序进而进行识别。后者可以通过几种不同方式完成。例如,可以测量多核苷酸中特定基序的存在(无需测量多核苷酸的其余序列)。或者,所述方法中特定的电和/或光信号的测量可以识别出所述多核苷酸来自特定来源。

[0657] 对于(iii),多核苷酸的序列可以如先前所述进行测定。合适的测序方法,特别是

使用电学测量方法的测序方法,描述在Stoddart D et al.,Proc Natl Acad Sci,12;106(19)7702-7,Lieberman KR et al,J Am Chem Soc.2010;132(50)17961-72,以及国际申请WO2000/28312中。

[0658] 对于(iv),二级结构可以采用多种方法测量。例如,如果所述方法涉及电学测量方法,则可以使用停留时间的变化或流过孔的电流的变化来测量所述二级结构。这允许区分单链和双链多核苷酸的区域。

[0659] 对于(v),可以测量是否存在任何修饰。所述方法优选包括,测定多核苷酸是否通过甲基化,氧化,损伤,用一种或多种蛋白或用一种或多种标记,标签或间隔区进行修饰。特定的修饰将导致与所述孔的特异性相互作用,其可以使用下述方法进行测量。例如,甲基胞嘧啶可以基于其与每个核苷酸相互作用期间流过所述孔的电流而与胞嘧啶区分开来。

[0660] 所述目标多核苷酸与CsgG孔或其突变体接触,例如如本发明的孔。所述孔通常存在于膜中。合适的膜如下所述。所述方法可以使用适合于研究膜/孔系统——其中孔存在于膜中——的的任何装置进行。所述方法可以使用适合用于跨膜孔感侧的任何装置进行。例如,所述装置包括包含水性溶液的腔室和将腔室分成两个部分的屏障。所述屏障通常具有孔洞,在孔洞中形成包含孔的膜。或者所述屏障形成膜,所述膜中存在孔。

[0661] 所述方法可以使用描述于国际申请号PCT/GB08/000562(WO 2008/102120)中的装置进行。

[0662] 可以进行各种不同类型的测量。这包括但不限于电学测量和光学测量。可能的电学测量包括电流测量,阻抗测量,隧道测量(tunnelling measurement)(Ivanov AP et al.,Nano Lett.2011Jan 12;11(1):279-85)以及FET测量(国际申请WO 2005/124888)。光学测量可以与电学测量结合(Soni GV et al.,Rev Sci Instrum.2010Jan;81(1)014301)。所述测量可以是跨膜电流测量,例如流过所述孔的离子电流的测量。

[0663] 电学测量可以使用描述在Stoddart D et al.,Proc Natl Acad Sci,12;106(19)7702-7,Lieberman KR et al,J Am Chem Soc.2010;132(50)17961-72和国际申请WO2000/28312中的标准单通道记录设备进行。替代地,电学测量可以使用多通道系统进行,例如如国际申请WO 2009/077734和国际申请WO 2011/067559中描述的。

[0664] 所述方法优选采用跨膜施加的电势进行。所述施加的电势可以是电压电势。替代地,所施加的电势可以是化学电势。其一实例为采用跨膜,例如双亲性分子层,的盐梯度进行。盐梯度被公开在Holden et al.,J Am Chem Soc.2007Jul 11;129(27):8650-5中。在一些情况下,多核苷酸相对所述孔移动时流过所述孔的电流用于估算或确定所述多核苷酸的序列。这就是链测序。

[0665] 所述方法可以包括测量多核苷酸相对所述孔移动时流过所述孔的电流。因此用于所述方法的设备也可以包括能够施加电势并测量穿过膜和孔的电信号的电路。所述方法可以采用膜片钳或电压钳进行。所述方法优选涉及使用电压钳。

[0666] 本发明方法可以包括测量多核苷酸相对所述孔移动时流过所述孔的电流。测量通过跨膜蛋白孔的离子流的合适条件是本领域已知的并且在实施例中公开。所述方法通常通过施加在所述膜和所述孔上的电压进行。所使用的电压通常为从+5V至-5V,例如从从+4V至-4V,从+3V至-3V或从+2V至-2V。所使用的电压通常为从-600mV至+600V或-400mV至+400mV。所使用的电压优选在具有选自-400mV,-300mV,-200mV,-150mV,-100mV,-50mV,-

20mV和0mV的下限和独立地选自+10mV,+20mV,+50mV,+100mV,+150mV,+200mV,+300mV和+400mV的上限的范围内。所使用的电压更优选在100mV至240mV的范围内并且最优选在120mV至220mV的范围内。通过使用增加的施加电势,可以增加孔对不同核苷酸的识别。

[0667] 所述方法通常在存在任何电荷载体的情况下进行,例如金属盐例如碱金属盐,卤化物盐例如氯化物盐,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-甲基咪唑鎓氯化物。在上述示例性装置中,盐存在于所述腔室中的水性溶液中。通常使用氯化钾(KCl),氯化钠(NaCl),氯化铯(CsCl)或亚铁氰化钾和铁氰化钾的混合物。KCl,NaCl和亚铁氰化钾和铁氰化钾的混合物是优选的。电荷载体在所述膜上可以是不对称的。例如,电荷载体的类型和/或浓度可以在所述膜的每一侧上不同。

[0668] 所述盐的浓度可以是饱和的。所述盐的浓度可以为3M或更低,并且通常为0.1至2.5M,0.3至1.9M,0.5至1.8M,0.7至1.7M,0.9至1.6M或1M至1.4M。所述盐的浓度优选为150mM至1M。所述方法优选使用至少0.3M,例如至少0.4M,至少0.5M,至少0.6M,至少0.8M,至少1.0M,至少1.5M,至少2.0M,至少2.5M或至少3.0M的盐浓度进行。高盐浓度提供高的信噪比,并允许通过电流指示在正常电流波动背景下待识别的核苷酸的存在。

[0669] 所述方法通常在存在缓冲液的情况下进行。在上述示例性装置中,所述缓冲液存在于所述腔室中的水性溶液中。任意缓冲液可以用于本发明的方法。通常地,所述缓冲液为磷酸缓冲液。其他合适的缓冲液为HEPES或Tris-HCl缓冲液。所述方法通常在pH为4.0至12.0,4.5至10.0,5.0至9.0,5.5至8.8,6.0至8.7或7.0至8.8或7.5至8.5下进行。使用的pH值优选约7.5。

[0670] 所述方法可以在0℃至100℃,15℃至95℃,16℃至90℃,17℃至85℃,18℃至80℃,19℃至70℃或20℃至60℃温度下进行。所述方法通常在室温下进行。所述方法任选的在支持酶功能的温度下进行,例如约37℃。

[0671] 多核苷酸结合蛋白

[0672] 链表征方法优选包括使多核苷酸与多核苷酸结合蛋白接触,使得所述蛋白控制多核苷酸相对于孔的移动,例如,通过所述孔。

[0673] 更优选地,所述方法包括(a)使多核苷酸与CsgG孔或其突变体(例如,本发明的孔)和多核苷酸结合蛋白接触,使得所述蛋白控制多核苷酸相对于孔的移动,例如,通过孔,和(b)当多核苷酸相对于孔移动时获取一个或多个测量值,其中,所述测量值指示多核苷酸的一个或多个特征,从而表征多核苷酸。

[0674] 更优选地,所述方法包括(a)使多核苷酸与CsgG孔或其突变体(例如,本发明的孔)和多核苷酸结合蛋白接触,使得所述蛋白控制多核苷酸相对于孔的移动,例如,通过孔,和(b)当多核苷酸相对于孔移动时测量通过孔的电流,其中,所述电流指示多核苷酸的一个或多个特征,从而表征多核苷酸。

[0675] 多核苷酸结合蛋白可以是能够结合多核苷酸并控制其移动通过孔的任何蛋白。本领域中判断蛋白是否与多核苷酸结合是比较简单的。蛋白通常与多核苷酸相互作用并改性多核苷酸的至少一种性质。蛋白可以通过裂解多核苷酸以形成各单个核苷酸或核苷酸的短链(例如,二核苷酸或三核苷酸)来对其进行改性。蛋白可以通过使多核苷酸定向或将其移动到特定位置来对其进行改性,即,控制它的移动。

[0676] 多核苷酸结合蛋白优选衍生自多核苷酸处理酶。多核苷酸处理酶是能够与多核苷酸相互作用并改性多核苷酸的至少一种性质的多肽。所述酶可以通过裂解多核苷酸以形成各单个核苷酸或核苷酸的短链(例如,二核苷酸或三核苷酸)来对其进行改性。所述酶可以通过使多核苷酸定向或将其移动到特定位置来对其进行改性。多核苷酸处理酶不需要显示酶活性,只要其能够结合多核苷酸并控制其通过孔的移动即可。例如,可以对所述酶进行改性以去除其酶活性,或者可以在防止其用作酶的条件下进行使用。以下对这种条件进行更详细的讨论。

[0677] 多核苷酸处理酶优选衍生自溶核酶(nucleolytic enzyme)。在酶的构建体中使用的多核苷酸处理酶更优选地衍生自酶分类(EC)基团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中的任一个。所述酶可以是申请号为PCT/GB10/000133(公开为WO2010/086603)的国际申请中公开的那些酶中的任一个。

[0678] 优选地,所述酶为聚合酶、外切核酸酶、解旋酶和拓扑异构酶,例如,促旋酶。合适的酶包括但不限于来自大肠杆菌的核酸外切酶I(SEQ ID NO:399)、来自大肠杆菌的核酸外切酶III酶(SEQ ID NO 401)、来自极端嗜热链球菌的RecJ(SEQ ID NO:403)和细菌噬菌体λ外切核酸酶(SEQ ID NO:405)、TatD外切核酸酶及其变体。三个包括SEQ ID NO:403或其变体所示序列的亚基相互作用形成三聚体外切核酸酶。这些外切核酸酶也可以用于本发明的外切核酸酶方法中。聚合酶可以是PyroPhage®3173DNA聚合酶(其可从Lucigen®公司购得)、SD聚合酶(可从Bioron®购得)或其变体。该酶优选为Phi29DNA聚合酶(SEQ ID NO:397)或其变体。拓扑异构酶优选为酶分类(EC)组5.99.1.2和5.99.1.3中的任一个。

[0679] 所述酶最优选衍生自解旋酶,例如,Hel308Mbu(SEQ ID NO:406)、Hel308Csy(SEQ ID NO:407)、Hel308Tga(SEQ ID NO:408)、Hel308Mhu(SEQ ID NO:409)、Tral Eco(SEQ ID NO:410)、XPD Mbu(SEQ ID NO:411)或其变体。本发明中可以使用任何解旋酶。所述解旋酶可以为或衍生自Hel308解旋酶、RecD解旋酶,例如,Tral解旋酶或TrwC解旋酶、XPD解旋酶或Dda解旋酶。所述解旋酶可以是以下国际申请中公开的解旋酶、改性解旋酶或解旋酶构建体中的任一个:PCT/GB2012/052579(公开为WO2013/057495);PCT/GB2012/053274(公开为WO2013/098562);PCT/GB2012/053273(公开为WO2013098561);PCT/GB2013/051925(公开为WO2013/057495);PCT/GB2012/053274(公开为WO2013/098562);PCT/GB2012/053273(公开为WO2013098561);PCT/GB2013/051925(公开为WO2014/013260);PCT/GB2013/051924(公开为WO2014/013259);PCT/GB2013/051928(公开为WO2014/013262)和PCT/GB2014/052736。

[0680] 解旋酶优选地包括SEQ ID NO:413(Trwc Cba)或其变体所示的序列、SEQ ID NO:406(Hel308Mbu)或其变体所示的序列,或者SEQ ID NO:412(Dda)或其变体所示的序列。变体可能以下述针对跨膜孔所讨论的任何方式不同于天然序列。SEQ ID NO:412的优选变体包括(a)E94C和A360C或(b)E94C,A360C,C109A和C136A,然后,可选地(ΔM1)G1G2(即,缺失M1,然后加入G1和G2)。

[0681] 根据本发明,可以使用任何数量的解旋酶。例如,可以使用1,2,3,4,5,6,7,8,9,10或更多个解旋酶。在一些实施例中,可以使用不同数目的解旋酶。

[0682] 本发明的方法优选包括使多核苷酸与两个或更多个解旋酶接触。所述两个或更多个解旋酶通常是相同的解旋酶。所述两个或更多个解旋酶可以是不同的解旋酶。

[0683] 所述两个或更多个解旋酶可以是上述解旋酶的任意组合。所述两个或更多个解旋酶可以是两个或更多个Dda解旋酶。所述两个或更多个解旋酶可以是一种或多种Dda解旋酶和一种或多种TrwC解旋酶。所述两个或更多个解旋酶可以是相同解旋酶的不同变体。

[0684] 所述两个或更多个解旋酶优选地彼此连接。所述两个或更多个解旋酶更优选地彼此共价连接。解旋酶可以以任何顺序并使用任何方法连接。本发明中采用的优选解旋酶构建体在以下国际申请中进行描述:PCT/GB2013/051925(公开为WO2014/013260);PCT/GB2013/051924(公开为WO2014/013259);PCT/GB2013/051928(公开为WO2014/013262)和PCT/GB2014/052736。

[0685] SEQ ID NO:397,399,401,403,405,406,407,408,409,410,411,412或413的变体为具有氨基酸序列的酶,所述氨基酸序列从SEQ ID NO:397,399,401,403,405,406,407,408,409,410,411,412或413的氨基酸序列变化而来并且保留多核苷酸结合能力。这可以使用本领域已知的任何方法来测量。例如,变体可以与多核苷酸接触,并且可以对其与多核苷酸结合及沿多核苷酸移动的能力进行测量。变体可以包括促进多核苷酸结合和/或促进其在高盐浓度和/或室温下活性的改性。可以对变体进行改性,使得它们结合多核苷酸(即,保留多核苷酸结合能力),但不起到解旋酶的作用(即,当提供有促进移动的所有必需组分时,不沿多核苷酸移动,例如ATP和Mg^{2+})。这种改性是本领域已知的。例如,解旋酶中Mg^{2+}结合结构域的改性通常导致变体不能用作解旋酶。这些类型的变体可以作为分子制动剂(见下文)。

[0686] 在SEQ ID NO:397,339,401,403,405,406,407,408,409,410,411,412或413的氨基酸序列的整个长度上,基于氨基酸同一性,变体优选与该序列至少50%同源。更优选地,基于氨基酸同一性,变体多肽可以与SEQ ID NO:397,399,401,403,405,406,407,408,409,410,411,412或413的氨基酸序列在整个序列上至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,更优选地,至少95%,97%或99%同源。在200或更多个,例如,230,250,270,280,300,400,500,600,700,800,900或1000或更多个连续氨基酸的长度段上,可以存在至少80%,例如,至少85%,90%或95%的氨基酸同一性("严格同源性")。如上所述,确定同源性。变体可以按上述针对SEQ ID NO:390所讨论的任何方式,不同于野生型序列。酶可以与孔共价连接。可以采用任何方法将酶与孔共价连接。

[0687] 优选的分子制动剂为TrwC Cba-Q594A(具有突变Q594A的SEQ ID NO:413)。这种变体不用作解旋酶(即,当提供有促进移动的所有必需组分时,与多核苷酸结合,但不沿其移动,例如,ATP和Mg^{2+})。

[0688] 在链测序时,多核苷酸随着或逆着所施加的电势移位通过孔。可以将对双链多核苷酸逐渐或渐进性起作用的外切核酸酶用在孔的顺式侧上以在施加电势的情况下将剩余的单链送入并通过所述孔,或者在反向电势的情况下,在反式侧上将剩余单链送入并通过所述孔。同样地,解开双链DNA的解旋酶也可以以类似的方式进行使用。也可以使用聚合酶。还可能用于要求逆着施加电势进行链移位的测序应用,但DNA必须在反向电势或不存在电势的情况下首先被酶"捕获"。然后随着电势在结合后切换回来,所述链将从顺式到反式穿过所述孔并且通过电流保持处于伸展构象。单链DNA外切核酸酶或单链DNA依赖性聚合酶可以作为分子马达,以受控的逐步方式,逆着施加的电势,从反式到顺式,将刚刚移位的单链拉回并通过所述孔。

[0689] 所述方法中可以使用任何解旋酶。解旋酶相对于孔可以以两种模式工作。第一,所述方法优选使用解旋酶进行,使得解旋酶通过由施加电压产生的场移动所述多核苷酸通过所述孔。在这种模式中,首先将多核苷酸的5'端捕获到所述孔中,并且所述解旋酶将所述多核苷酸移动到所述孔中,使得其在所述电场下通过所述孔直到其最终移位到膜的反式侧。或者,所述方法优选这样进行:使得解旋酶逆着由施加电压产生的场使多核苷酸移动通过所述孔。在这种模式中,多核苷酸的3'端首先被捕获到所述孔中,并且所述解旋酶使多核苷酸移动通过所述孔,使得其逆着所施加的场被拉出所述孔直到其最终弹回到膜的顺式侧。

[0690] 也可以沿相反方向执行所述方法。可以在孔中首先捕获多核苷酸的3'端,并且解旋酶可以移动多核苷酸到孔中,使得其利用场通过所述孔,直到其最终移位到膜的反式侧。

[0691] 当解旋酶未提供有促进移动的必要组分或解旋酶被改性以阻碍或阻止其移动时,其可以结合多核苷酸并且在其被所施加的场牵引进孔中时用作减慢多核苷酸移动的制动剂。在非主动模式中,多核苷酸的3'或5'端是否被捕获是无关紧要的,是所施加的场通过用作制动剂的酶将多核苷酸朝向反式侧牵引到孔中。当处于非主动模式时,解旋酶对多核苷酸进行的移动控制可以以多种方式进行描述,包括棘轮(ratcheting)、滑动和制动。也可以以这种方式使用缺乏解旋酶活性的解旋酶变体。

[0692] 可以以任何顺序使多核苷酸与多核苷酸结合蛋白和孔接触。优选地,当多核苷酸与多核苷酸结合蛋白(比如:解旋酶)和孔接触时,多核苷酸首先与蛋白形成复合物。当在孔上施加电压时,多核苷酸/蛋白复合物此时与孔形成复合物并控制多核苷酸移动通过孔。

[0693] 采用多核苷酸结合蛋白的方法中的任何步骤通常在游离核苷酸或游离核苷酸类似物以及促进多核苷酸结合蛋白的作用的酶辅助因子存在的情况下进行。游离核苷酸可以是上面讨论的任何一个或多个单个核苷酸。游离核苷酸包括但不限于腺苷一磷酸(AMP)、腺苷二磷酸(ADP)、腺苷三磷酸(ATP)、鸟苷一磷酸(GMP)、鸟苷二磷酸(GDP)、鸟苷三磷酸(GTP)、胸苷单磷酸(TMP)、胸苷二磷酸(TDP)、胸苷三磷酸(TTP)、尿苷单磷酸(UMP)、尿苷二磷酸(UDP)、尿苷三磷酸(UTP)、胞苷单磷酸(CMP)、胞苷二磷酸(CDP)、胞苷三磷酸(CTP)、环腺苷酸(cAMP)、环磷酸鸟苷(cGMP)、脱氧腺苷一磷酸(dAMP)、脱氧腺苷二磷酸(dADP)、脱氧腺苷三磷酸(dATP)、脱氧鸟苷一磷酸(dGMP)、脱氧鸟苷二磷酸(dGDP)、脱氧鸟苷三磷酸(dGTP)、脱氧胸苷单磷酸(dTMP)、脱氧胸苷二磷酸(dTDP)、脱氧胸苷三磷酸(dTTP)、脱氧尿苷单磷酸(dUMP)、脱氧尿苷二磷酸(dUDP)、脱氧尿苷三磷酸(dUTP)、脱氧胞苷单磷酸(dCMP)、脱氧胞苷二磷酸(dCDP)和脱氧胞苷三磷酸(dCTP)。游离核苷酸优选地选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP或dCMP。游离核苷酸优选为腺苷三磷酸(ATP)。酶辅助因子为使得构建体起作用的因子。酶辅助因子优选为二价金属阳离子。二价金属阳离子优选为Mg^{2+}、Mn^{2+}、Ca^{2+}、或Co^{2+}。酶辅助因子最优选为Mg^{2+}。

[0694] 解旋酶和分子制动剂

[0695] 在优选实施例中,所述方法包括:

[0696] (a)向多核苷酸提供一个或多个解旋酶以及与多核苷酸连接的一个或多个分子制动剂;

[0697] (b)使多核苷酸与CsgG孔或其突变体,例如本发明的孔,接触,并在孔上施加电势使得一个或多个解旋酶以及一个或多个分子制动剂聚集在一起,二者均控制多核苷酸相对于孔的移动,例如,通过孔;

[0698] (c)在多核苷酸相对于孔移动时,获取一个或多个测量值,其中,所述测量值指示多核苷酸的一个或多个特征,从而表征多核苷酸。

[0699] 这类方法在申请号为PCT/GB2014/052737的国际申请中进行详细讨论。

[0700] 所述一个或多个解旋酶可以是上述讨论的任何一种。所述一个或多个分子制动剂可以是与多核苷酸结合并减慢多核苷酸通过孔的移动的任何化合物或分子。所述一个或多个分子制动剂优选包括一个或多个与多核苷酸结合的化合物。所述一个或多个化合物优选为一个或多个大环化合物。合适的大环化合物包括但不限于环糊精、杯芳烃、环肽、冠醚、瓜环、柱芳烃、其衍生物或其组合。环糊精或其衍生物可以是Eliseev,A.V.和Schneider,H-J.(1994)J.Am.Chem.Soc 116,6081-6088中公开的任一种。该试剂更优选为七-6-氨基-β-环糊精(am_7-βCD)、6-单脱氧-6-单氨基-β-环糊精(am_1-βCD)或七-(6-脱氧-6-胍基)-环糊精(gu_7-βCD)。

[0701] 所述一个或多个分子制动剂优选为一个或多个单链结合蛋白(SSB)。所述一个或多个分子制动剂更优选地为包括不具有净负电荷的羧基末端(C-末端)区域的单链结合蛋白(SSB)或(ii)在其C-末端区域具有一处或多处修饰的改性SSB,所述修饰减少了C-末端区域的净负电荷。所述一个或多个分子制动剂最优选地为国际申请PCT/GB2013/051924(公开为WO2014/013259)中公开的SSB之一。

[0702] 所述一个或多个分子制动剂优选地为一个或多个多核苷酸结合蛋白。多核苷酸结合蛋白可以是任何能够与多核苷酸结合并控制其通过孔的移动的蛋白。本领域中可以容易地判断蛋白是否与多核苷酸结合。蛋白通常与多核苷酸相互作用并改性多核苷酸的至少一种性质。蛋白可以通过裂解多核苷酸形成各单个核苷酸或核苷酸的短链(例如,二核苷酸或三核苷酸)来对其进行改性。该部分可以通过使多核苷酸定向或将其移动到特定位置,即,控制其的移动,来对其进行改性。

[0703] 多核苷酸结合蛋白优选衍生自多核苷酸处理酶。所述一个或多个分子制动剂可以衍生自上述任何一种多核苷酸处理酶。作为分子制动剂的Phi29聚合酶(SEQ ID NO:396)的改性版本公开在美国专利5,576,204中。所述一个或多个分子制动剂优选衍生自解旋酶。

[0704] 可以使用衍生自解旋酶的任意数量的分子制动剂。例如,可以将1,2,3,4,5,6,7,8,9,10或更多个解旋酶用作分子制动剂。如果两个或更多个解旋酶被用作分子制动剂,则所述两个或更多个解旋酶通常是相同的解旋酶。所述两个或更多个解旋酶可以是不同的解旋酶。

[0705] 所述两个或更多个解旋酶可以是上述解旋酶的任意组合。所述两个或更多个解旋酶可以是两个或更多个Dda解旋酶。所述两个或更多个解旋酶可以是一个或多个Dda解旋酶和一个或多个TrwC解旋酶。所述两个或更多个解旋酶可以是同一解旋酶的不同变体。

[0706] 所述两个或更多个解旋酶优选地彼此连接。所述两个或更多个解旋酶更优选地彼此共价连接。解旋酶可以以任何顺序并使用任何方法进行连接。优选地,对衍生自解旋酶的所述一个或多个分子制动剂进行改性以减小多核苷酸结合结构域中开口的大小,其中,多核苷酸在至少一个构象状态下通过所述多核苷酸结合结构域从解旋酶上解开。这在WO2014/013260中进行了公开。

[0707] 本发明中使用的优选解旋酶构建体在以下国际申请中进行描述:PCT/GB2013/051925(公开为WO2014/013260);PCT/GB2013/051924(公开为WO2014/013259);PCT/

GB2013/051928(公开为WO2014/013262)和PCT/GB2014/052736。

[0708] 如果所述一个或多个解旋酶在主动模式下使用(即,当所述一个或多个解旋酶提供有便于移动的所有必需组分时,例如,ATP和Mg^{2+}),所述一个或多个分子制动剂优选地(a)在非主动模式下使用(即,在不存在促进移动的必要组件或不能主动移动的情况下使用);(b)在主动模式下使用,其中,所述一个或多个分子制动剂沿与所述一个或多个解旋酶相反的方向移动;或(c)在主动模式下使用,其中,所述一个或多个分子制动剂沿与所述一个或多个解旋酶相同的方向移动,并且比所述一个或多个解旋酶移动得更慢。

[0709] 如果所述一个或多个解旋酶在非主动模式下(即,当所述一个或多个解旋酶未提供有促进移动的所有必需组分,例如,ATP和Mg^{2+}时或不能主动移动时)使用,所述一个或多个分子制动剂优选地(a)在非主动模式下(即,用在不存在促进移动的必要组件或不能主动移动的情况下)使用或(b)在主动模式下使用,其中,所述一个或多个分子制动剂以与多核苷酸通过孔的方向相同的方向沿着多核苷酸移动。

[0710] 所述一个或多个解旋酶和一个或多个分子制动剂可以在任何位置与多核苷酸连接,使得它们被聚集在一起,并且二者均控制多核苷酸通过孔的移动。所述一个或多个解旋酶和一个或多个分子制动剂是至少一个分开的核苷酸,例如,至少5个、至少10个、至少50个、至少100个、至少500个、至少1000个、至少5000个、至少10,000个、至少50,000个或更多个分离开的核苷酸。如果该方法涉及表征一端设置有Y适配器且另一端设置有发夹环适配器的双链多核苷酸,则所述一个或多个解旋酶优选地与Y适配器连接,且所述一个或多个分子制动剂优选地与发夹环适配器连接。在这一实施方式中,所述一个或多个分子制动剂优选地为一个或多个进行了改性使得它们与多核苷酸结合但不用作解旋酶的解旋酶。与Y适配器连接的所述一个或多个解旋酶优选地停滞在以下进行更详细讨论的间隔区处。与发夹环适配器连接的所述一个或多个分子制动剂优选地不停滞在间隔区处。当所述一个或多个解旋酶到达发夹环时,所述一个或多个解旋酶和一个或多个分子制动剂优选地被聚集在一起。在Y适配器与多核苷酸连接之前或在Y适配器与多核苷酸连接之后,可以将所述一个或多个解旋酶与Y适配器连接。在发夹环适配器与多核苷酸连接之前或者在发夹环适配器与多核苷酸连接之后,可以将所述一个或多个分子制动剂与发夹环适配器连接。

[0711] 所述一个或多个解旋酶和所述一个或多个分子制动剂优选地不相互连接。所述一个或多个解旋酶和所述一个或多个分子制动剂更优选地彼此不共价连接。如国际申请PCT/GB2013/051925(公开为WO2014/013260);PCT/GB2013/051924(公开为WO2014/013259);PCT/GB2013/051928(公开为WO2014/013262)和PCT/GB2014/052736中所述,所述一个或多个解旋酶和一个或多个分子制动剂优选地不连接。

[0712] 间隔区

[0713] 所述一个或多个解旋酶可以停滞在国际申请PCT/GB2014/050175中所述的一个或多个间隔区处。在本发明中可以使用该国际申请中公开的一个或多个解旋酶和一个或多个间隔区的任何构象。

[0714] 当多核苷酸的一部分进入孔并沿着由所施加电势产生的场移动通过孔时,随着多核苷酸移动通过孔,所述一个或多个解旋酶通过孔移动通过间隔区。这是因为多核苷酸(包括所述一个或多个间隔区)移动通过孔,而所述一个或多个解旋酶留在孔的顶部。

[0715] 所述一个或多个间隔区优选是多核苷酸的一部分,例如,它们打断多核苷酸序列。

所述一个或多个间隔区优选地不是与多核苷酸杂交的一个或多个阻断分子例如减速块(speed bump)的一部分。

[0716] 多核苷酸中可以存在任意数量的间隔区,例如1,2,3,4,5,6,7,8,9,10或更多个间隔区。多核苷酸中优选存在两个、四个或六个间隔区。多核苷酸的不同区域中可以存在一个或多个间隔区,例如,Y适配器和/或发夹环适配器中的一个或多个间隔区。

[0717] 所述一个或多个间隔区各自提供一个能垒,所述一个或多个解旋酶即使在主动模式下也无法克服该能垒。所述一个或多个间隔区可以通过减少解旋酶的牵引力(例如,通过去除多核苷酸中的核苷酸的碱基)或物理性阻断所述一个或多个解旋酶的移动(例如,采用庞大的化学基团)来阻滞所述一个或多个解旋酶。

[0718] 所述一个或多个间隔区可以包括使所述一个或多个解旋酶停滞的任一分子或分子组合。所述一个或多个间隔区可以包括防止所述一个或多个解旋酶沿多核苷酸移动的任一分子或分子组合。在不存在跨膜孔和施加电势的情况下,直接确定所述一个或多个解旋酶是否停滞在一个或多个间隔区处。例如,可以通过PAGE测量解旋酶移动穿过间隔区并替代DNA互补链的能力。

[0719] 所述一个或多个间隔区通常包括线性分子,例如,聚合物。所述一个或多个间隔区通常具有与多核苷酸不同的结构。例如,如果多核苷酸是DNA,则所述一个或多个间隔区通常不是DNA。特别地,如果多核苷酸是脱氧核糖核酸(DNA)或核糖核酸(RNA),则所述一个或多个间隔区优选包括肽核酸(PNA)、甘油核酸(GNA)、苏糖核酸(TNA)、锁核酸(LNA)或具有核苷酸侧链的合成聚合物。所述一个或多个间隔区可以包括与多核苷酸方向相反的一个或多个核苷酸。例如,当多核苷酸处于5'至3'方向时,所述一个或多个间隔区可以在3'至5'方向上包括一个或多个核苷酸。核苷酸可以是以上所讨论的任一种。

[0720] 所述一个或多个间隔区优选包括一个或多个硝基吲哚,例如,一个或多个5-硝基吲哚、一个或多个肌苷、一个或多个吖啶、一个或多个2-氨基嘌呤、一个或多个2-6-二氨基嘌呤、一个或多个5-溴-脱氧尿苷、一个或多个反式胸苷(反式dT)、一个或多个反式二脱氧胸腺嘧啶(ddT)、一个或多个双脱氧胞苷(ddC)、一个或多个5-甲基胞苷、一个或多个5-羟甲基胞苷、一个或多个2'-O-甲基RNA碱基、一个或多个异脱氧胞苷(Iso-dC)、一个或多个异脱氧鸟苷(Iso-dG)、一个或多个iSpC3基团(即,缺乏糖和碱基的核苷酸)、一个或多个光可裂解(PC)基团、一个或多个己二醇基团、一个或多个间隔区9(iSp9)基团、一个或多个间隔区18(iSp18)基团、聚合物或一个或多个硫醇连接。所述一个或多个间隔区可以包括这些基团的任一组合。许多这些基团可以从IDT®(Integrated DNA Technologies®)购得。

[0721] 所述一个或多个间隔区可以包括任一数量的这些基团。例如,对于2-氨基嘌呤、2-6-二氨基嘌呤、5-溴-脱氧尿苷、反式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、己二醇基团和硫醇连接,所述一个或多个间隔区优选包括2,3,4,5,6,7,8,9,10,11,12或更多个。所述一个或多个间隔区优选包括2,3,4,5,6,7,8或更多个iSp9基团。所述一个或多个间隔区优选包括2,3,4,5或6或更多个iSp18基团。最优选的间隔区为四个iSp18基团。

[0722] 聚合物优选为多肽或聚乙二醇(PEG)。所述多肽优选包括2,3,4,5,6,7,8,9,10,11,12或更多个氨基酸。所述PEG优选包括2,3,4,5,6,7,8,9,10,11,12或更多个单体单元。

[0723] 所述一个或多个间隔区优选包括一个或多个无碱基核苷酸(即,缺少核碱基的核

苷酸),例如,2,3,4,5,6,7,8,9,10,11,12或更多个无碱基核苷酸。所述核碱基可以由无碱基核苷酸中的-H(idSp)或-OH替代。通过从一个或多个相邻核苷酸中去除核碱基,可以将无碱基间隔区插入多核苷酸中。例如,可以将多核苷酸改性为包括3-甲基腺嘌呤、7-甲基鸟嘌呤、1,N6-亚乙烯基腺嘌呤肌苷(ethenoadenine inosine)或次黄嘌呤,并且可以使用人类烷基腺嘌呤DNA糖基化酶(hAAG)从这些核苷酸中去除核碱基。替代地,可以将多核苷酸改性成包括尿嘧啶和用尿嘧啶-DNA糖基化酶(UDG)去除的核碱基。在一实施例中,所述一个或多个间隔区不包括任何无碱基核苷酸。

[0724] 所述一个或多个解旋酶可以由每个线性分子间隔区阻滞(即,在间隔区之前)或在每个线性分子间隔区上阻滞。如果使用线性分子间隔区,则多核苷酸优选地设置有与每个间隔区末端相邻的多核苷酸双链区域,所述一个或多个解旋酶移动通过每个间隔区末端。所述双链区域通常有助于在相邻间隔区上阻滞所述一个或多个解旋酶。如果在约100mM或更低的盐浓度下执行该方法,则特别优选存在双链区域。每个双链区域通常是至少10个,例如,至少12个,长度的核苷酸。如果本发明中使用的多核苷酸为单链的,则可以通过将较短的多核苷酸与邻近间隔区的区域进行杂合来形成双链区域。所述较短的多核苷酸通常由与多核苷酸相同的核苷酸形成,但也可以由不同的核苷酸形成。例如,所述较短的多核苷酸可以由LNA形成。

[0725] 如果使用线性分子间隔区,则多核苷酸优选在每个间隔区的末端设置有阻断分子,所述末端与所述一个或多个解旋酶移动通过的末端相对。这可以有助于确保所述一个或多个解旋酶在每个间隔区上保持停滞。还可以有助于在所述一个或多个解旋酶在溶液中扩散的情况下,使它/它们保留在多核苷酸上。所述阻断分子可以是以下讨论的任一化学基团,这些化学基团物理上导致所述一个或多个解旋酶停滞。所述阻断分子可以是多核苷酸的双链区域。

[0726] 所述一个或多个间隔区优选地包括一个或多个物理上导致所述一个或多个解旋酶停滞的化学基团。所述一个或多个化学基团优选为一个或多个侧链化学基团。所述一个或多个化学基团可以与多核苷酸中的一个或多个核碱基连接。所述一个或多个化学基团可以与多核苷酸骨架连接。可以存在任一数目的这些化学基团,例如,2,3,4,5,6,7,8,9,10,11,12或更多个。合适的基团包括但不限于荧光团、链霉亲和素和/或生物素、胆固醇、亚甲蓝、二硝基苯酚(DNP)、地高辛和/或抗地高辛和二苄基环辛炔基。

[0727] 多核苷酸中的不同间隔区可以包括不同的阻滞分子。例如,一个间隔区可以包括上述线性分子之一,另一个间隔区可以包括物理上使所述一个或多个解旋酶停滞的一个或多个化学基团。间隔区可以包括上述线性分子中的任何一种和物理上导致所述一种或多种解旋酶停滞的一个或多个化学基团,例如,一个或多个脱碱基(abasics)和荧光团。

[0728] 可以根据多核苷酸的类型和执行本发明所述方法的条件来设计合适的间隔区。大多数解旋酶结合DNA并沿着DNA移动,因此可能会被任何不是DNA的物质停滞。上文对合适的分子进行了讨论。

[0729] 优选在游离核苷酸和/或解旋酶辅助因子存在的情况下,执行本发明所述方法。这在以下进行更详细的讨论。在不存在跨膜孔和施加电势的情况下,所述一个或多个间隔区优选能够在存在游离核苷酸和/或存在解旋酶辅助因子的情况下阻滞所述一个或多个解旋酶。

[0730] 如果在存在游离核苷酸和解旋酶辅助因子的情况下执行本发明所述方法(使得所述一个或多个解旋酶处于主动模式),如下所述,则通常使用一个或多个较长的间隔区来确保所述一个或多个解旋酶在与跨膜孔相接触且施加电势之前停滞在多核苷酸上。在不存在游离核苷酸和解旋酶辅助因子的情况下,可以使用一个或多个较短的间隔区(使得所述一个或多个解旋酶处于非主动模式)。

[0731] 盐浓度也影响所述一个或多个间隔区阻滞所述一个或多个解旋酶的能力。在不存在跨膜孔和施加电势的情况下,所述一个或多个间隔区优选能够在约100mM或更低的盐浓度下使所述一个或多个解旋酶停滞。本发明所述方法中使用的盐浓度越高,通常使用的所述一个或多个间隔区越短,反之亦然。

[0732] 特征的优选组合在下表3中示出。

[0733]

多核苷酸	间隔区成分*	间隔区长度(即,数量)	盐[ ]	游离核苷酸?	解旋酶辅助因子?
DNA	iSpC3	4	1 M	是	是
DNA	iSp18	4	100-1000 mM	是	是

[0734]

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

[0735] 表3

[0736] 该方法可以涉及使两个或更多个解旋酶移动通过间隔区。在这种情况下,通常增加间隔区的长度,以防止尾部解旋酶在不存在孔和施加电势的情况下推动先导解旋酶经过间隔区。如果该方法涉及使两个或更多个解旋酶移动通过一个或多个间隔区,则上述间隔区长度可以增加至少1.5倍,例如,2倍、2.5倍或3倍。例如,如果该方法涉及使两个或更多个解旋酶移动通过一个或多个间隔区,则上述表3第三列中的间隔区长度可以增加1.5倍、2倍、2.5倍或3倍。

[0737] 膜

[0738] 本发明所述孔可以存在于膜中。在本发明所述方法中,多核苷酸通常在膜中与CsgG孔或其突变体例如本发明所述孔接触。根据本发明,可以使用任何膜。合适的膜是本领域公知的。膜优选为两亲层。两亲层是由两亲分子形成的层,例如,磷脂,其具有亲水性和亲油性。两亲分子可以是合成的或天然存在的。非天然存在的两亲物和形成单层的两亲物是本领域已知的,包括:例如,嵌段共聚物(Gonzalez-Perez等人;Langmuir;2009,25,10447-10450)。嵌段共聚物为聚合物材料,其中,两个或更多个单体子单元聚合在一起形成单一聚合物链。嵌段共聚物通常具有由每个单体子单元贡献的性质。然而,嵌段共聚物可以具有由各个子单元形成的聚合物所不具有的独特性质。可以对嵌段共聚物进行设计,使得其中一个单体子单元为疏水的(即亲油的),而另一个子单元在水性介质中为亲水的。在这种情况

下,嵌段共聚物可以具有两亲性质并且可以形成模拟生物膜的结构。嵌段共聚物可以是二嵌段(由两个单体子单元组成),但也可以由多于两个的单体子单元构成,从而形成表现为两亲物的更加复杂的结构。共聚物可以是三嵌段、四嵌段或五嵌段共聚物。膜优选为三嵌段共聚物膜。

[0739] 古细菌双极性四醚脂质是天然存在的脂质,其构建成使得脂质形成单层膜。通常在存活于恶劣生物环境、嗜热菌、嗜盐菌和嗜酸性细菌的极端微生物中发现这些脂质。其稳定性被认为是源于最终双层的融合性质。通过产生具有亲水-疏水-亲水这一通用基序的三嵌段聚合物来直接构建模拟这些生物实体的嵌段共聚物材料。该材料可以形成单体膜,所述单体膜行为与脂质双层类似,并且包括从囊泡到层式膜发一系列相行为。由这些三嵌段共聚物形成的膜相比生物脂质膜具有多个优点。由于三嵌段共聚物是合成的,因此可以仔细控制精确的结构,以提供形成膜以及与孔和其它蛋白相互作用所需的正确链长度和性质。

[0740] 嵌段共聚物也可以由未分类为脂质子物质的子单元构成;例如,疏水性聚合物可以由硅氧烷或其它非烃系单体制成。嵌段共聚物的亲水性部分也可以具有低的蛋白结合性质,这使得膜暴露于原始生物样本时具有高度抗性。这种头部基团单元还可以衍生自非典型脂质头部基团。

[0741] 与生物脂质膜相比,三嵌段共聚物膜也增加了机械和环境稳定性,例如,更高的操作温度或pH范围。嵌段共聚物的合成性质为定制基于聚合物的能广泛应用的膜提供了平台。

[0742] 最优选地,膜为国际申请PCT/GB2013/052766或PCT/GB2013/052767中公开的膜之一。

[0743] 可以对两亲性分子进行化学改性或功能化以促进多核苷酸的偶联。

[0744] 两亲层可以是单层或双层。两亲层通常是平面的。两亲层可以是弯曲的。可以对两亲层进行支撑。

[0745] 两亲性膜通常是天然可移动的,基本上作为二维流体,其脂质扩散速率约为10^{-8}cm s^{-1}。这意味着孔和偶联的多核苷酸通常可以在两亲膜内移动。

[0746] 膜可以为脂质双层。脂质双层为细胞膜的模型,并作为一系列实验研究的优秀平台。例如,脂质双层通过单通道记录可以用于膜蛋白的体外研究。替代地,脂质双层可用作用以检测一系列物质存在的生物传感器。脂质双层可以是任何脂质双层。合适的脂质双层包括但不限于平面脂质双层、支撑双层或脂质体。脂质双层优选为平面脂质双层。以下国际申请中公开了合适的脂质双层:PCT/GB08/000563(公开为WO2008/102121)、申请号为PCT/GB08/004127(公开为WO2009/077734)和申请号为PCT/GB2006/001057(公开为WO2006/100484)。

[0747] 脂质双层的形成方法是本领域已知的。脂质双层通常通过Montal和Mueller(Proc.Natl.Acad.Sci.USA.),1972;69:3561-3566)的方法形成,其中脂质单层承载在水溶液/空气界面上,经过与该界面垂直的孔的任一侧。通常,通过先将脂质溶解在有机溶剂中,然后在孔的任一侧上使所述溶剂滴在水溶液表面上蒸发,从而将脂质添加到水性电解质溶液的表面。一旦有机溶剂蒸发,孔的任一侧上的溶液/空气界面物理地上下移动通过孔,直到形成双层。可以在膜中的孔上或在凹陷内的开口上形成平面脂质双层。

[0748] Montal和Mueller的方法受大众所喜爱,因为它是一种用于形成适合蛋白孔插入的优质脂质双层的经济且较简单方法。其它形成双层的常见方法包括尖端浸渍、涂覆双层和脂质体双层的贴片钳。

[0749] 尖部浸渍形成双层必需使孔表面(例如,移液管尖部)接触到承载脂质单层的测试溶液的表面。另外,首先通过将溶解在有机溶剂中的脂质的液滴在溶液表面蒸发而在溶液/空气界面处产生脂质单层。然后由Langmuir-Schaefer工艺形成所述双层,并且所述双层需要机械自动化以相对于溶液表面移动所述孔。

[0750] 对于涂覆双层,将溶解在有机溶剂中的脂质的液滴直接施加到所述孔,该孔浸没在水性测试溶液中。使用画笔或等效物将脂质溶液薄薄地涂在孔上。溶剂的稀释(thinning)导致脂质双层的形成。然而,从双层完全去除溶剂是困难的,并因此由该方法形成的双层是较不稳定的且在电化学测量过程中更容易有噪声。

[0751] 膜片钳通常在生物细胞膜的研究中使用。细胞膜通过抽吸被夹到移液管的端部,且膜的小块贴附到孔上。该方法已被改为用于通过夹住脂质体,然后脂质体爆裂而使脂质双层密封移液管的孔,而产生脂质双层。该方法需要稳定的、巨大的和单层的脂质体,以及需要在具有玻璃表面的材料中制造小孔。

[0752] 脂质体可通过超声、挤出或Mozafari法(Colas等人(2007)Micron 38:841-847)来形成。

[0753] 在优选的实施方案中,脂质双层如在国际申请No.PCT/GB08/004127(公开为WO2009/077734)中描述的来形成。本方法中有利的是,脂质双层是由干燥的脂质形成的。在最优选的实施方案中,脂质双层跨越开口而形成,如在WO2009/077734(PCT/GB08/004127)中所述。

[0754] 脂质双层是由脂质的两个相对的层形成的。脂质的两层被排列为使得它们的疏水性尾部基团面向彼此以形成疏水性内部。脂质的亲水性头部基团面向外朝向该双层的每一侧上的含水环境。该双层可存在于许多脂质相中,所述脂质相包括但不限于,液态无序相(流体层状)、液态有序相、固态有序相(层状凝胶相、相互交叉的凝胶相)和平面双层晶体(层状子凝胶相、层状结晶相)。

[0755] 可使用形成脂质双层的任何脂质组合物。脂质组合物被选择为,使得形成的脂双层具有所需的特性,例如表面电荷,具有支撑膜蛋白的能力、填充密度或机械特性。脂质组合物可包含一个或多个不同的脂质。例如,脂质组合物可含有多达100个脂质。脂质组合物优选含有1至10个脂质。脂质组合物可以包括天然存在的脂质和/或人工脂质。

[0756] 脂质通常包括头部基团、界面部分和可以相同或不同的两个疏水尾部基团。合适的头部基团包括,但不限于,中性头部基团,如二酰基甘油酯(DG)和神经酰胺(CM);两性离子头部基团,如磷脂酰胆碱(PC)、磷脂酰乙醇胺(PE)和鞘磷脂(SM);带负电荷的头部基团,如磷脂酰甘油(PG);磷脂酰丝氨酸(PS)、磷脂酰肌醇(PI)、磷酸(PA)和心磷脂(CA);和带正电荷的头部基团,如三甲基铵-丙烷(TAP)。合适的界面部分包括,但不限于,天然存在的界面部分,例如基于甘油或基于神经酰胺的部分。合适的疏水性尾部基团包括,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);和支链烃链,如植烷酰(phytanoyl)。链的长度和不饱和烃链中双键的位置和数量可以改变。链的长度和支

链烃链中支链(如甲基基团)的位置和数目可以改变。疏水性尾部基团可以被连接到作为醚或酯的界面部分。脂质可以是霉菌酸。

[0757] 脂质也可以是被化学修饰的。脂质的头部基团或尾部基团可以是经化学修饰的。头部基团已被化学修饰的合适的脂质包括,但不限于,PEG修饰的脂质,例如1,2-二酰基-sn-甘油基-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-sn-甘油基-3-磷酸乙醇胺-N-[生物素基(聚乙二醇)2000];为共轭而修饰的脂质,如1,2-二油酰-sn-甘油基-3-磷酸乙醇胺-N-(琥珀酰基)和1,2-二棕榈酰基-sn-甘油基-3-磷酸乙醇胺-N-(生物素基)。尾部基团已被化学修饰的合适的脂质包括,但不限于,可聚合的脂质,如1,2-双(10,12-二十三二炔酰)-sn甘油基-3-磷酸胆碱;氟化脂质,例如1-棕榈酰基-2-(16-氟代棕榈酰基)-sn-甘油基-3-磷酸胆碱;氘化脂质,如1,2-二棕榈酰基-D62-sn-甘油基-3-磷酸胆碱;和醚连接的脂质,如1,2-二-O-植烷基-sn-甘油基-3-磷酸胆碱。脂质可以是化学修饰的或官能化的,以有助于多核苷酸的偶联。

[0758] 两亲层例如,脂质组合物,通常包含将影响该层的性能的一个或多个添加剂。合适的添加剂包括,但不限于,脂肪酸,如棕榈酸、肉豆蔻酸和油酸;脂肪醇,如棕榈醇、肉豆蔻醇和油醇;甾醇,如胆固醇、麦角固醇、羊毛甾醇,谷甾醇和豆甾醇;溶血磷脂,如1-酰基-2-羟基-sn-甘油基-3-磷酸胆碱;和神经酰胺。

[0759] 在另一个优选的实施方案中,所述膜包括固态层。固态层可以由有机和无机材料形成,所述有机和无机材料包括但不限于,微电子材料、绝缘材料,如Si_3N_4、Al_2O_3和SiO,有机和无机聚合物,例如聚酰胺、塑料如Teflon®,或弹性体如双组分加成固化硅橡胶,和玻璃。固态层可以由石墨烯形成。合适的石墨烯层在国际申请PCT/US2008/010637(公开为WO2009/035647)中公开。如果膜包括固态层,则孔通常存在于两亲膜中或固态层内包括的层中,例如,固态层内的孔洞、阱、间隙、通道、沟槽或狭缝中。本领域技术人员可以制备合适的固态/两亲性混合系统。WO2009/020682和WO2012/005857中公开了合适的系统。可以使用上述任一两亲膜或层。

[0760] 该方法通常使用(i)包含孔的人工两亲层,(ii)分离的、包含孔的天然存在的脂质双层,或(iii)具有嵌入其中的孔的细胞,而进行。该方法通常使用人工两亲层进行,如人工三嵌段共聚物层。除了孔以外,所述层可包含其它跨膜和/或膜内蛋白以及其他分子。合适的设备和条件在下面讨论。本发明的方法通常在体外实施。

[0761] 偶联

[0762] 多核苷酸优选与包括孔的膜偶联。所述方法可以包括将多核苷酸与包括孔的膜偶联。优选采用一个或多个锚(anchor)将多核苷酸与膜偶联。可以使用任何已知方法将多核苷酸与膜偶联。

[0763] 每个锚包括与多核苷酸偶联(或结合)的基团和与膜偶联(或结合)的基团。每个锚可与多核苷酸和/或膜共价偶联(或结合)。如果使用了Y适配器和/或发夹环适配器,则优选使用该适配器将多核苷酸与膜偶联。

[0764] 可以采用任一数量的锚将多核苷酸与膜偶联,例如,2,3,4或更多个锚。例如,可以采用两个锚,将多核苷酸与膜偶联,其中,每个锚均与多核苷酸和膜偶联(或结合)。

[0765] 所述一个或多个锚可以包括所述一个或多个解旋酶和/或所述一个或多个分子制动剂。

[0766] 如果膜为两亲层,例如,共聚物膜或脂质双层,则所述一个或多个锚优选包括存在于膜中的多肽锚和/或存在于膜中的疏水锚。疏水锚优选为脂质、脂肪酸、固醇、碳纳米管、多肽、蛋白或氨基酸,例如,胆固醇、棕榈酸酯或生育酚。在优选实施例中,所述一个或多个锚不是孔。

[0767] 可以对膜的组分,例如,两亲分子、共聚物或脂质,进行化学改性或官能化以形成所述一个或多个锚。以下对合适的化学改性以及膜组分的合适官能化方式的示例进行更详细的讨论。可以对任一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%。

[0768] 可以直接将多核苷酸与膜偶联。用于将多核苷酸与膜偶联的一个或多个锚优选包括连接体。所述一个或多个锚可以包括一个或多个,例如,2,3,4或更多个连接体。可以使用一个连接体将多于一个,例如,2,3,4或更多个多核苷酸与膜偶联。

[0769] 优选的连接体包括但不限于诸如多核苷酸、聚乙二醇(PEG)、多糖和多肽等聚合物。这些连接体可以是直链、支链或环状的。例如,连接体可以是环状多核苷酸。多核苷酸可以杂合到所述环状多核苷酸连接体上的互补序列。

[0770] 所述一个或多个锚或一个或多个连接体可以包括可以切割分解(cut to broken down)的组分,例如,限制位点或光不稳定基团。

[0771] 官能化连接体及其可以偶联分子的方式是本领域已知的。例如,用马来酰亚胺基团官能化的连接体将与蛋白中的半胱氨酸残基反应并连接。在本发明的上下文中,蛋白可以存在于膜中或可以用于与多核苷酸偶联(或结合)。这将在以下进行更详细的讨论。

[0772] 可以采用“锁和钥匙”结构来避免多核苷酸交联。各连接体仅一端可以一起反应以形成更长的连接体,各连接体的另一端分别与多核苷酸或膜反应。这样的连接体在国际申请PCT/GB10/000132(公开为WO2010/086602)中进行描述。

[0773] 优选在以下所讨论的测序实施例中使用连接体。多核苷酸在与孔相互作用时不偶联(即,在步骤(b)或(e)中不偶联),就这一点而言,如果多核苷酸直接与膜持久偶联,则由于膜和孔之间的距离,当测序运行不能继续到多核苷酸的末端时,某些序列数据将丢失。如果使用连接体,则多核苷酸可以进行到完成为止。

[0774] 偶联可以是持久的或稳定的。换言之,偶联可以为以下这样:使得当与孔相互作用时,多核苷酸仍与膜偶联。

[0775] 偶联可以是瞬态的。换言之,偶联可以为以下这样:使得当与孔相互作用时,多核苷酸可以从膜上解偶联。

[0776] 对于某些应用,如适体(aptamer)检测,优选偶联为瞬态性质的。如果持久或稳定的连接体直接与多核苷酸的5’或3’末端连接,并且连接体比膜与跨膜孔通道之间的距离短,则当测序运行不能继续到多核苷酸的末端时,某些序列数据将丢失。如果偶联是瞬态的,则当偶联端随机摆脱膜时,则可以完全处理多核苷酸。以下对形成持久性/稳定性或瞬态链接的化学基团进行更详细的讨论。多核苷酸可以利用胆固醇或脂肪酰基链与两亲层或三嵌段共聚物膜瞬时偶联。可以使用长度为6至30个碳原子的任一脂肪酰基链,例如,十六烷酸。

[0777] 在优选实施例中,将多核苷酸(例如,核酸)与诸如三嵌段共聚物膜或脂质双层等两亲层偶联。之前已使用各种不同的系链策略来将核酸与合成脂质双层偶联。这些总结在

下表4中。

[0778] 表4

[0779]

锚包括	偶联类型	参考文献
硫醇	稳定	Yoshina-Ishii, C. and S. G. Boxer (2003). "Arrays of mobile tethered vesicles on supported lipid bilayers." J Am Chem Soc **125**(13): 3696-7.
生物素	稳定	Nikolov, V., R. Lipowsky, et al. (2007). " Behavior of giant vesicles with anchored DNA molecules." Biophys J **92**(12): 4356-68
胆固醇	瞬态	Pfeiffer, I. and F. Hook (2004). " Bivalent cholesterol-based coupling of oligonucletides to lipid membrane assemblies." J Am Chem Soc **126**(33): 10224-5
表面活性剂（例如脂质，棕榈酸酯等）	稳定	van Lengerich, B., R. J. Rawle, et al. " Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions." Langmuir **26**(11): 8666-72

[0780] 可以在合成反应中使用经改性的亚磷酰胺对合成多核苷酸和/或连接体进行官能化,这易于直接加入合适的锚定基团,例如,胆固醇、生育酚、棕榈酸酯、硫醇、脂质和生物素基团。这些不同的连接用化学物质提供了一套用于与多核苷酸连接的选择。每个不同的改性基团以稍微不同的方式偶联多核苷酸,并且偶联并不总是永久的,因此为多核苷酸与膜的连接提供了不同的停留时间。上文中讨论了瞬态偶联的优点。

[0781] 如果可以将互补的反应基团或锚定基团加入到多核苷酸中,则多核苷酸与连接体或官能化膜的偶联也可以通过许多其它方式实现。向多核苷酸的任一末端添加反应基团从前已经被报道。可以采用T4多核苷酸激酶和ATPγS向ssDNA或dsDNA的5’末端添加硫醇基团(Grant,G.P.和P.Z.Qin(2007);“A facile method for attaching nitroxide spin labels at the 5’terminus of nucleic acids.”NucleicAcidsRes 35(10):e77)。可以采用T4多核苷酸激酶和γ-[2-叠氮基乙基]-ATP或γ-[6-叠氮基己基]-ATP向ssDNA或dsDNA的5’-磷酸酯添加叠氮化物基团。采用硫醇或点击化学,将系链(tether)与多核苷酸共价连接,所述系链含有硫醇、碘乙酰胺OPSS或马来酰亚胺基团(对硫醇具有反应性)或DIBO(二苯并环辛炔(dibenzocyclooxtyne))或炔基(对叠氮化物具有反应性)。可以采用末端转移酶使化学基团(例如,生物素、硫醇和荧光团)的选择更多样化,从而将改性的寡核苷酸结合到ssDNA的3’端(Kumar,A.,P.Tchen等人(1988);“Nonradioactive labeling of synthetic oligonucleotide probes with terminal deoxynucleotidyl transferase.” AnalBiochem 169(2):376-82)。链霉亲和素/生物素和/或链霉亲和素/脱硫生物素偶联可用于任何其它多核苷酸。以下实施例描述了如何使用链霉亲和素/生物素和链霉亲和素/脱硫生物素将多核苷酸与膜偶联。也可以采用具有适当改性的核苷酸的末端转移酶(例如胆固醇或棕榈酸酯)将锚直接加入到多核苷酸中。

[0782] 所述一个或多个锚优选通过杂合将多核苷酸与膜偶联。所述一个或多个锚中的杂合允许以如上所述的瞬时方式偶联。杂合可以存在于所述一个或多个锚的任何部分中,例

如,在所述一个或多个锚与多核苷酸之间、在所述一个或多个锚内或者在所述一个或多个锚与膜之间。例如,连接体可以包括两个或更多个杂合在一起的多核苷酸,例如,3,4或5个多核苷酸。所述一个或多个锚可以与多核苷酸杂合。所述一个或多个锚可以直接与多核苷酸杂合,或直接与连接到多核苷酸的Y适配器和/或前导序列杂合,或直接与连接到多核苷酸的发夹环适配器杂合(如下所述)。替代地,所述一个或多个锚可以与一个或多个(例如,2或3个)杂合到多核苷酸的中间多核苷酸(或"夹板(splint)")杂合,与连接到多核苷酸的Y适配器和/或前导序列杂合,或与连接到多核苷酸的发夹环适配器杂合(如下所述)。

[0783] 所述一个或多个锚可以包括单链或双链多核苷酸。锚的一部分可以与单链或双链多核苷酸连接。采用T4RNA连接酶I对ssDNA的短片段的连接已被报道(Troutt,A.B.,M.G.McHeyzer-Williams等人(1992);"Ligation-anchored PCR:a simple amplification technique with single-sided specificity."Proc Natl Acad Sci U S A 89(20):9823-5)。替代地,单链或双链多核苷酸可以与双链多核苷酸连接,然后通过热或化学变性分离这两条链。对于双链多核苷酸,可以将一段单链多核苷酸添加到所述双链体的一端或两端,或将双链多核苷酸添加到所述双链体的一端或两端。对于将单链多核苷酸添加到双链多核苷酸,这可以同连接到单链多核苷酸其它区域一样,使用T4RNA连接酶I来实现。对于将双链多核苷酸添加到双链多核苷酸,此时连接可以用分别在多核苷酸和添加的多核苷酸上的互补的3'dA/dT尾部(这是常规针对许多样品制备应用进行的,以防止串多联体(concatemer)或者二聚体形成)或使用由多核苷酸的限制消化以及相容适配器的连接而产生的"粘性末端",进行"平端终止(blunt-ended)"。然后,当双链体熔化时,每条单链将具有5'或3'修饰,如果使用单链多核苷酸进行连接或修饰,则在5'末端或3'末端进行修饰,或如果使用双链多核苷酸进行连接,则在这两端进行修饰。

[0784] 如果多核苷酸是合成链,则在化学合成多核苷酸期间可以掺入所述一个或多个锚。例如,可以采用具有与其连接的反应基团的引物合成多核苷酸。

[0785] 腺苷酸化多核苷酸是连接反应中的中间体,其中,腺苷单磷酸与多核苷酸的5'-磷酸连接。可以使用各种试剂盒来产生该中间体,例如,来自NEB的5'DNA腺苷酸化试剂盒。通过在反应中用ATP替代修饰的核苷酸三磷酸酯,然后可以向多核苷酸的5'添加反应基团(如硫醇、胺、生物素、叠氮化物等)。也可以采用带有适当修饰的核苷酸(例如,胆固醇或棕榈酸酯)的5'DNA腺苷酸化试剂盒将锚直接添加到多核苷酸。

[0786] 扩增基因组DNA片段的常见技术是使用聚合酶链反应(PCR)。这里,采用两个合成的寡核苷酸引物,可以产生同一DNA片段的多个复本,其中,对于每个复本,双链体中每个链的5'将是合成的多核苷酸。可以通过采用聚合酶将单个或多个核苷酸添加到单链或双链DNA的3'末端。可以使用的聚合酶的实例包括但不限于末端转移酶、克列诺酶(Klenow)和大肠杆菌Poly(A)聚合酶。通过在反应中用ATP替代修饰的核苷酸三磷酸酯,就可将诸如胆固醇、硫醇、胺、叠氮化物、生物素或脂质等锚掺入双链多核苷酸中。因此,扩增的多核苷酸的每个复本都含有锚。

[0787] 理想地,多核苷酸与膜偶联而不必使多核苷酸官能化。这可以通过将所述一个或多个锚(例如,多核苷酸结合蛋白或化学基团)与膜偶联并使所述一个或多个锚与所述多核苷酸相互作用或通过使膜官能化来实现。所述一个或多个锚可以通过本文所述的任何方法与膜偶联。特别地,所述一个或多个锚可以包括一个或多个连接体,例如,马来酰亚胺官能

化连接体。

[0788] 在本实施例中,多核苷酸通常是RNA、DNA、PNA、TNA或LNA,并且可以是双链或单链的。本实施例特别适用于基因组DNA多核苷酸。

[0789] 所述一个或多个锚可以包括与单链或双链多核苷酸、多核苷酸内的特定核苷酸序列或多核苷酸内改性核苷酸的模式或存在于多核苷酸上的任何其它配体进行偶联、结合或相互作用的任何基团。

[0790] 合适的用于锚中的结合蛋白包括但不限于:大肠杆菌单链结合蛋白、P5单链结合蛋白、T4gp32单链结合蛋白、TOPO V dsDNA结合区、人组蛋白、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多核苷酸(或核酸)结合蛋白,包括下述所列的那些。

[0791] 特定核苷酸序列可以是转录因子、核糖体、内切核酸酶、拓扑异构酶或复制起始因子所识别的序列。改性核苷酸的模式可以是甲基化或损伤的模式。

[0792] 所述一个或多个锚可以包括与多核苷酸偶联、结合、嵌入或相互作用的任一基团。所述基团可以通过静电、氢键或范德华相互作用而插入多核苷酸或与多核苷酸相互作用。这种基团包括赖氨酸单体、聚赖氨酸(其与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(其可以与任何多核苷酸杂合)和锇复合物(其可以与甲基化碱基反应)。因此,可以采用与膜连接的一个或多个通用核苷酸将多核苷酸与膜偶联。每个通用核苷酸可以采用一个或多个连接体与膜偶联。通用核苷酸优选包括以下核碱基之一:次黄嘌呤、4-硝基吲哚、5-硝基吲哚、6-硝基吲哚、福尔马林、3-硝基吡咯、硝基咪唑、4-硝基吡唑、4-硝基苯并咪唑、5-硝基吲唑、4-氨基苯并咪唑或苯基(C6-芳环)。通用核苷酸更优选地包括以下核苷之一:2'-脱氧肌苷、肌苷、7-脱氮杂-2'-脱氧肌苷、7-脱脱氮肌苷、2-氮杂-脱氧肌苷、2-氮杂-肌苷、2-O'-甲基肌苷、4-硝基吲哚2'-脱氧核糖核苷、4-硝基吲哚核糖核苷、5-硝基吲哚2'-脱氧核糖核苷、5-硝基吲哚核糖核苷、6-硝基吲哚2'-脱氧核糖核苷、6-硝基吲哚核糖核苷、3-硝基吡咯2'-脱氧核糖核苷、3-硝基吡咯核糖核苷、次黄嘌呤的非环状糖类似物、硝基咪唑2'-脱氧核糖核苷、硝基咪唑核糖核苷、4-硝基吡唑2'-脱氧核糖核苷、4-硝基吡唑核糖核苷、4-硝基苯并咪唑2'-脱氧核糖核苷、4-硝基苯并咪唑核糖核苷、5-硝基吲唑2'-脱氧核糖核苷、5-硝基吲唑核糖核苷、4-氨基苯并咪唑2'-脱氧核糖核苷、4-氨基苯并咪唑核糖核苷、苯基C-核糖核苷、苯基C-2'-脱氧核糖基核苷、2'-脱氧水粉蕈素、2'-脱氧异鸟苷、K-2'-脱氧核糖、P-2'脱氧核糖和吡咯烷。通用核苷酸更优选包含2'-脱氧肌苷。通用核苷酸更优选为IMP或dIMP。通用核苷酸最优选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)。

[0793] 所述一个或多个锚可以通过Hoogsteen氢键(其中,两个核碱基通过氢键保持在一起)或反向Hoogsteen氢键(其中,一个核碱基相对于另一个核碱基旋转180°)与多核苷酸偶联(或结合)。例如,所述一个或多个锚可以包括一个或多个核苷酸、一个或多个寡核苷酸或一个或多个与多核苷酸形成Hoogsteen氢键或反向Hoogsteen氢键的多核苷酸。这些类型的氢键允许第三个多核苷酸链缠绕在双链螺旋上,形成三链体。所述一个或多个锚可以通过与双链体形成三链体而与双链多核苷酸偶联(或结合)。

[0794] 在本实施例中,至少1%、至少10%、至少25%、至少50%或100%的膜组分可以被官能化。

[0795] 当所述一个或多个锚包括蛋白时,它们能够无需进行进一步官能化而直接锚定到

膜中,例如,如果它已经具有与膜相容的外部疏水区域。这种蛋白的实例包括但不限于跨膜蛋白、膜内蛋白和膜蛋白。替代地,蛋白可以用与膜相容的基因融合疏水区域表达。这种疏水蛋白区域是本领域已知的。

[0796] 所述一个或多个锚在与膜接触之前优选与多核苷酸混合,但是所述一个或多个锚可以与膜接触并随后与多核苷酸接触。

[0797] 另一方面,可以使用上述方法对多核苷酸进行官能化,使得其可以被特定的结合基团识别。具体地,可以用诸如生物素(用于与链霉亲和素结合)、直链淀粉(用于与麦芽糖结合蛋白或融合蛋白结合)、Ni-NTA(用于与多组氨酸或多组氨酸标记蛋白结合)或肽(例如,抗原)等配体对多核苷酸进行官能化。

[0798] 根据优选实施例,当多核苷酸连接到优先旋入孔的前导序列时,所述一个或多个锚可以用于将多核苷酸与膜偶联。以下将对前导序列进行更详细的讨论。优选地,将多核苷酸连接(例如,接合)到优先旋入孔的前导序列。这种前导序列可以包括均聚多核苷酸或无碱区域。前导序列通常设计为直接地或经由一个或多个中间多核苷酸(或夹板)与所述一个或多个锚杂合。在这种情况下,所述一个或多个锚通常包括与前导序列中的序列或一个或多个中间多核苷酸(或夹板)中的序列互补的多核苷酸序列。在这种情况下,所述一个或多个夹板通常包括与前导序列中的序列互补的多核苷酸序列。

[0799] 化学连接中使用的分子的实例是EDC(1-乙基-3-[3-二甲基氨基丙基]碳二亚胺盐酸盐)。也可以采用市售的试剂盒(Thermo Pierce,货号:22980)向多核苷酸的5'末端添加反应基团。合适的方法包括但不限于使用组氨酸残基和Ni-NTA进行瞬时亲和连接以及通过反应性半胱氨酸、赖氨酸或非天然氨基酸进行更稳定的共价连接。

[0800] 双链多核苷酸

[0801] 多核苷酸可以是双链的。如果多核苷酸是双链的,则所述方法优选还包括:在接触步骤之前将诸如发夹环等桥接部分与多核苷酸的一端连接。然后,根据本发明,多核苷酸的两条链可以在多核苷酸与孔接触之时或接触之前分离开来。当多核苷酸移动通过孔是由诸如解旋酶或分子制动剂等多核苷酸结合蛋白控制时,两条链可以分离开来。

[0802] 以这种方式在双链构建体上连接和询问两条链增加了表征的效率和准确性。

[0803] 桥接部分能够连接目标多核苷酸的两条链。桥接部分通常共价连接目标多核苷酸的两条链。假如桥接部分不干扰单链多核苷酸移动通过跨膜孔,则桥接部分可以是任何能够连接目标多核苷酸的两条链的物质。

[0804] 桥接部分可以通过本领域已知的任一合适方法与目标多核苷酸连接。桥接部分可以单独合成并且化学连接或酶连接到目标多核苷酸。替代地,桥接部分可以在处理目标多核苷酸时产生。

[0805] 桥接部分在目标多核苷酸的一端或靠近目标多核苷酸的一端与目标多核苷酸连接。桥接部分优选在目标多核苷酸末端的10个核苷酸内与目标多核苷酸连接。

[0806] 合适的桥接部分包括但不限于聚合物连接体、化学连接体、多核苷酸或多肽。优选地,桥接部分包括DNA、RNA、改性DNA(例如,脱碱基DNA)、RNA、PNA、LNA或PEG。桥接部分更优选为DNA或RNA。

[0807] 桥接部分最优选是发夹环或发夹环适配器。可以使用本领域已知的方法来设计合适的发夹适配器。发夹环可以是任何长度。发夹环长度通常为110个或更少的核苷酸,例如,

100个或更少的核苷酸、90个或更少的核苷酸、80个或更少的核苷酸、70个或更少的核苷酸、60个或更少的核苷酸、50个或更少的核苷酸、40个或更少的核苷酸、30个或更少的核苷酸、20个或更少的核苷酸或10个或更少的核苷酸。发夹环的长度优选为约1至10、2至100、5至80或6至50个核苷酸。如果所述环涉及适配器的差分选择性，则优选长度较长的发夹环，例如，50至110个核苷酸。类似地，如下所述，如果所述环不涉及如下所述的选择性结合，则优选长度较短的发夹环，例如，1至5个核苷酸。

[0808] 发夹适配器可以与第一和/或第二多核苷酸的任一端，即，5’或3’末端，连接。可以使用本领域已知的任何方法将发夹适配器与第一和/或第二多核苷酸连接。可以使用连接酶连接发夹适配器，例如，T4DNA连接酶、大肠杆菌DNA连接酶、Taq DNA连接酶、Tma DNA连接酶和9°N DNA连接酶。

[0809] 可以使用本领域已知的任何方法将多核苷酸的两条链分离开来。例如，它们可以被多核苷酸结合蛋白或者利用有利于脱杂合的条件(有利于脱杂合的条件的实例包括但不限于高温、高pH以及添加可以破坏氢键合或碱基配对的试剂，如甲酰胺和脲)分离开来。

[0810] 发夹适配器优选包括选择性结合部分。这使得第一和/或第二多核苷酸能够被纯化或分离。选择性结合部分为可以基于其结合性质进行选择的部分。因此，选择性结合部分优选为特异性与表面结合的部分。如果选择性结合部分与表面结合的程度比本发明中使用的任何其它部分更大，则所述选择性结合部分与表面特异性结合。在优选实施例中，所述部分与本发明中使用的任何其它部分所未结合的表面结合。

[0811] 合适的选择性结合部分是本领域已知的。优选的选择性结合部分包括但不限于：生物素、多核苷酸序列、抗体、抗体片段，例如，Fab和ScSv，抗原、多核苷酸结合蛋白、多组氨酸尾部和GST标签。最优选的选择性结合部分为生物素和选择性多核苷酸序列。生物素特特异性地与涂覆有亲和素的表面结合。选择性多核苷酸序列特异性地与涂覆有同源序列的表面结合(即，杂合)。替代地，选择性多核苷酸序列特异性地与涂覆有多核苷酸结合蛋白的表面结合。

[0812] 发夹适配器和/或选择性结合部分可以包括能够切割、切开(nicked)、裂解或水解的区域。这种区域可以设计为，使第一和/或第二多核苷酸在纯化或分离后从其所结合的表面去除。合适的区域是本领域已知的。合适的区域包括但不限于RNA区域、包括脱硫生物素和链霉亲和素的区域、二硫键和光可裂解区域。

[0813] 双链目标多核苷酸优选地在桥接部分的相对端包括前导序列，例如，发夹环或发夹环适配器。前导序列将在以下进行更详细的讨论。

[0814] 拐角测序(Round the corner sequencing)

[0815] 在优选实施例中，目标双链多核苷酸的一端设置有桥接部分，例如，发夹环或发夹环适配器，并且所述方法包括使多核苷酸与孔接触，使得多核苷酸的两条链移动通过孔，以及在多核苷酸的两条链相对于孔移动时获取一个或多个测量值，其中，所述测量值指示多核苷酸的两条链的一个或多个特征，从而表征目标双链多核苷酸。在另一个优选实施例中，目标双链多核苷酸的一端设置有桥接部分，例如，发夹环或发夹环适配器，并且所述方法包括使多核苷酸与孔和外切核酸酶接触，使得多核苷酸的两条链均被消化从而形成单个核苷酸。上述任一实施方式同样适用于本实施方式。

[0816] 前导序列

[0817] 链表征/测序方法中的接触步骤之前,所述方法优选包括使多核苷酸与优先旋入孔的前导序列连接。前导序列有利于本发明方法。前导序列被设计为优先旋入孔中,从而促进多核苷酸通过孔的移动。前导序列也可以用于将多核苷酸与如上所述的一个或多个锚连接。

[0818] 前导序列通常包括聚合物。所述聚合物优选是带负电荷的。所述聚合物优选为多核苷酸(例如DNA或RNA)、改性多核苷酸(例如,脱碱基DNA)、PNA、LNA、聚乙二醇(PEG)或多肽。前导序列优选包括多核苷酸,更优选地包括单链多核苷酸。前导序列可以包括上述任一多核苷酸。单链前导序列最优选包括单链DNA,例如,聚dT区段。前导序列优选地包括所述一个或多个间隔区。

[0819] 前导序列可以是任何长度,但是长度通常为10至150个核苷酸,例如,长度为20至150个核苷酸。前导序列的长度通常取决于所述方法中使用的跨膜孔。

[0820] 前导序列优选为下文限定的Y适配器的一部分。

[0821] 双偶联

[0822] 本发明的方法可以涉及双链多核苷酸的双偶联。在优选实施方式中,本发明的方法包括:

[0823] (a)在双链多核苷酸的一端设置Y适配器,另一端设置诸如发夹环适配器等桥接部分适配器,其中,所述Y适配器包括一个或多个用于将多核苷酸与膜偶联第一锚,所述桥接部分适配器包括一个或多个用于将多核苷酸与膜偶联的第二锚,其中,桥接部分适配器与膜偶联的强度大于Y适配器与膜偶联的强度;

[0824] (b)使步骤(a)中提供的多核苷酸与CsgG孔或其突变体,比如本发明的孔,相接触,使得多核苷酸相对于孔移动,例如,通过孔;以及

[0825] (c)在多核苷酸相对于孔移动时获取一个或多个测量值,其中,所述测量值指示多核苷酸的一个或多个特征,从而表征目标多核苷酸。

[0826] 这种方法在申请号为1406147.7的英国申请中进行了详细讨论。

[0827] 双链多核苷酸的一端设置有Y适配器,另一端设置有桥接部分适配器。所述Y适配器和/或桥接部分适配器通常是多核苷酸适配器。它们可以由上述任一多核苷酸形成。

[0828] Y适配器通常包括(a)双链区域和(b)单链区域或在另一端不互补的区域。如果Y适配器包括单链区域,则可以将其描述为具有突出端(overhang)。在Y适配器中存在非互补区域使得适配器具有Y形状,因为与双链部分不同,这两条链通常彼此不杂合。Y适配器包括所述一个或多个第一锚。上文对锚进行了更详细的讨论。

[0829] Y适配器优选地包括优先旋入孔的前导序列。这在上文中讨论过。

[0830] 如上所述,桥接部分适配器优选包括选择性结合部分。如上所述,桥接部分适配器和/或选择性结合部分可以包括能够切割、切开、裂解或水解的区域。

[0831] 如上所述,如果使用一个或多个解旋酶和一个或多个分子制动剂,则Y适配器优选地包括所述一个或多个解旋酶,并且桥接部分适配器优选包括所述一个或多个分子制动剂。

[0832] 可以使用本领域已知的任何方法将Y适配器和/或桥接部分适配器与多核苷酸连接。可以使用诸如T4DNA连接酶、大肠杆菌DNA连接酶、Taq DNA连接酶、Tma DNA连接酶和9°NDNA连接酶等连接酶连接所述适配器中的一者或两者。替代地,可以使用以下所讨论的本

发明的方法将适配器添加到多核苷酸。

[0833] 在优选实施例中,所述方法的步骤a)包括对双链多核苷酸进行改性,使得其一端包括Y适配器,另一端包括桥接部分适配器。可以采用任何改性方式。所述方法优选包括根据本发明对双链多核苷酸进行改性。这将在以下进行更详细的讨论。改性和表征方法可以以任何方式进行组合。

[0834] 桥接部分适配器与膜的偶联(或结合)强度大于Y适配器与膜的偶联(或结合)强度。这可以以任何方式进行测量。申请号为1406147.7的英国申请的实施例中公开了一种用于测量偶联(或结合)强度的合适方法。

[0835] 桥接部分适配器的偶联(或结合)强度优选为Y适配器偶联(或结合)强度的至少1.5倍,例如,为锚适配器的偶联(或结合)强度的至少两倍、至少三倍、至少四倍、至少五倍或至少十倍。用于膜的桥接部分适配器的亲和常数(Kd)优选为Y适配器的亲和常数的至少1.5倍,例如为Y适配器偶联强度的至少两倍,至少三次,至少四次,至少五次或至少十倍。

[0836] 桥接部分适配器比Y适配器更强地与膜偶联(或结合)的方式有多种。例如,桥接部分适配器可以包括比Y适配器更多的锚。例如,桥接部分适配器可以包括2,3或更多个第二锚,而Y适配器可以包括一个第一锚。

[0837] 所述一个或多个第二锚与膜偶联(或结合)的强度可以大于所述一个或多个第一锚与膜偶联(或结合)的强度。所述一个或多个第二锚与桥接部分适配器偶联(或结合)的强度可以大于所述一个或多个第一锚与Y适配器偶联(或结合)的强度。所述一个或多个第一锚和所述一个或多个第二锚可以通过杂合与其各自的适配器连接,并且所述一个或多个第二锚中的杂合强度比所述一个或多个第一锚中的杂合强度大。这些实施例的任一组合也可用于本发明。可以使用本领域的已知技术来测量偶联(或结合)强度。

[0838] 所述一个或多个第二锚优选地包括一个或多个与膜偶联(或结合)的基团,其强度大于所述一个或多个与膜偶联(或结合)的第一锚中的一个或多个基团。在优选实施例中,采用胆固醇将桥接部分适配器/一个或多个第二锚与膜偶联(或结合),采用棕榈酸酯将Y适配器/一个或多个第一锚与膜偶联(或结合)。胆固醇比棕榈酸酯能更强地与三嵌段共聚物膜和脂质膜结合。在替代实施例中,桥接部分适配器/一个或多个第二锚采用诸如棕榈酸酯等单酰基物质与膜偶联(或结合),而Y适配器/一个或多个第一锚采用诸如二棕榈酰卵磷脂等二酰基物质与膜偶联(或结合)。

[0839] 添加发夹环和前导序列

[0840] 在提供之前,双链多核苷酸可以与MuA转座酶和一组双链MuA底物接触,其中,所述组中的一部分底物为包括前导序列的Y适配器,并且所述组中的一部分底物为发夹环适配器。转座酶将双链多核苷酸分析物分裂成片段,并将MuA底物与片段的一端或两端连接。这产生了多个改性双链多核苷酸,其一端包括前导序列,另一端包括发夹环。然后,可以使用本发明的方法对改性双链多核苷酸进行研究。

[0841] 所述组中的每个底物优选包括通用核苷酸的至少一个突出端,使得转座酶将模板多核苷酸分裂成片段,并将底物与双链片段的一端或两端连接,从而产生多个片段/底物构建体,其中,所述方法还包括将突出端与构建体中的片段连接,从而产生多个经改性的双链多核苷酸。以上对合适的通用核苷酸进行了讨论。突出端的长度优选为5个核苷酸。

[0842] 替代地,组中的每个底物优选包括(i)至少一个突出端和(ii)在与所述至少一个

突出端相同的链中的至少一个核苷酸,其包括不存在于模板多核苷酸中的核苷,使得转座酶将模板多核苷酸分裂成片段并将底物与双链片段的一端或两端连接,从而产生多个片段/底物构建体,所述方法还包括(a)通过选择性地去除所述至少一个核苷酸来从构建体中去除突出端,从而产生多个包括单链间隙的双链构建体;和(b)修复构建体中的单链间隙,从而产生多个经改性的双链多核苷酸。多核苷酸通常包括核苷脱氧腺苷(dA)、脱氧尿苷(dU)和/或胸苷(dT)、脱氧鸟苷(dG)和脱氧胞苷(dC)。不存在于多核苷酸中的核苷优选为脱碱基的腺苷(A)、尿苷(U)、5-甲基尿苷(m^5U)、胞苷(C)或鸟苷(G),或包括脲、5,6-二羟基胸腺嘧啶、胸腺嘧啶、5-羟基-5-甲基乙内酰脲(5-hydroxy-5methylhydanton)、尿嘧啶二醇、6-羟基-5,6-二氢硫胺、甲基丙醇二酰脲(tartronylurea)、7,8-二氢-8-氧鸟嘌呤(8-氧鸟嘌呤)、8-氧代腺嘌呤、fapy-鸟嘌呤、甲基-fapy-鸟嘌呤、fapy-腺嘌呤、黄曲霉毒素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,N 6-亚乙烯基腺嘌呤、次黄嘌呤、5-羟基尿嘧啶、5-羟基甲基尿嘧啶、5-甲酰基尿嘧啶或顺式-环丁烷嘧啶二聚体。所述至少一个核苷酸优选为所述突出端中的10个或更少个核苷酸。所述至少一个核苷酸为突出端中的第一个核苷酸。突出端中的所有核苷酸优选包括不存在于模板多核苷酸中的核苷。

[0843] 这些基于MuA的方法在申请号为PCT/GB2014/052505的国际申请中进行了公开。这些方法也在申请号为1406147.7的英国申请中进行了详细讨论。

[0844] 一个或多个解旋酶可以在其与双链多核苷酸和MuA转座酶接触之前与MuA底物Y适配器连接。替代地,一个或多个解旋酶可以在其与双链多核苷酸和MuA转座酶接触之前与MuA底物Y适配器连接。

[0845] 一个或多个分子制动剂可以在其与双链多核苷酸和MuA转座酶接触之前与MuA底物发夹环适配器连接。替代地,一个或多个分子制动剂可以在其与双链多核苷酸和MuA转座酶接触之前与MuA底物发夹环适配器连接。

[0846] 解偶联

[0847] 本发明所述方法可以涉及表征多个目标多核苷酸和解偶联至少第一目标多核苷酸。

[0848] 在优选实施例中,本发明涉及表征两个或更多个目标多核苷酸。该方法包括:

[0849] (a)在第一样本中提供第一多核苷酸;

[0850] (b)在第二样本中提供第二多核苷酸;

[0851] (c)采用一个或多个锚将所述第一样本中的所述第一多核苷酸与膜偶联;

[0852] (d)使所述第一多核苷酸与CsgG孔或其突变体例如本发明的孔接触,使得多核苷酸相对于孔移动,例如,通过孔;

[0853] (e)当所述第一多核苷酸相对于孔移动时获取一个或多个测量值,其中,所述测量值指示所述第一多核苷酸的一个或多个特征,从而表征所述第一多核苷酸;

[0854] (f)将所述第一多核苷酸从膜中解偶联;

[0855] (g)采用一个或多个锚将所述第二样本中的所述第二多核苷酸与膜偶联;

[0856] (h)使所述第二多核苷酸与CsgG孔或其突变体例如本发明的孔接触,使得所述第二多核苷酸相对于孔移动,例如,通过孔;以及

[0857] (i)当所述第二多核苷酸相对于孔移动时获取一个或多个测量值,其中,所述测量

值指示所述第二多核苷酸的一个或多个特征,从而表征所述第二个多核苷酸。

[0858] 这种方法在申请号为1406155.0的英国申请中进行详细讨论。

[0859] 步骤(f)(即,第一多核苷酸的解偶联)可以在步骤(g)之前(即,在将第二多核苷酸与膜偶联之前)执行。步骤(g)可以在步骤(f)之前执行。如果第二多核苷酸在第一多核苷酸解偶联之前与膜偶联,则步骤(f)优选包括选择性地将第一多核苷酸与膜解偶联(即,将第一多核苷酸而不是第二多核苷酸与膜解偶联)。技术人员可以设计一种实现选择性解偶联的系统。步骤(f)和(g)可以同时执行。这将在以下进行更详细的讨论。

[0860] 在步骤(f)中,优选至少10%的第一多核苷酸从膜中解偶联。例如,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%的第一多核苷酸可以从膜中解偶联。优选地,所有的第一多核苷酸从膜中解偶联。可以使用孔来确定从膜中解偶联的第一多核苷酸的数量。这在各实施例中进行公开。

[0861] 第一多核苷酸和第二多核苷酸可以彼此不同。替代地,第一和第二多核苷酸可以是不同的多核苷酸。在这种情况下,在添加第二多核苷酸之前可以无需去除至少部分所述第一样本。这将在以下进行更详细的讨论。如果该方法涉及研究三个或更多个多核苷酸,则它们可以彼此不同,或者它们中的一些可能彼此不同。

[0862] 第一多核苷酸和第二多核苷酸可以是同一多核苷酸的两个实例。第一多核苷酸可以与第二多核苷酸相同。这使得可以校阅(proof reading)。如果该方法涉及研究三个或更多个多核苷酸,则它们可以全部为同一多核苷酸的三个或更多个实例,或者它们中的一些可以是同一多核苷酸的不同实例。

[0863] 第一样本和第二样本可以彼此不同。例如,第一样本可以来源于人,第二样本可以来源于病毒。如果第一和第二样本彼此不同,则它们可以含有或被怀疑含有相同的第一和第二多核苷酸。如果所述方法涉及研究三个或更多个样本,则它们可以全部彼此不同,或者它们中的一些可以彼此不同。

[0864] 第一样本和第二样本优选为同一样本的两个实例。第一样本优选与第二样本相同。这使得可以校阅。如果所述方法涉及研究三个或更多个样本,则它们可以全部是同一样本的三个或更多个实例,或者它们中的一些可以是同一样本的不同实例。

[0865] 可以研究任一数目的多核苷酸。例如,本发明方法可以涉及表征3,4,5,6,7,8,9,10,20,30,50,100或更多个多核苷酸。如果使用本发明所述方法研究三个或更多个多核苷酸,则第二多核苷酸也从膜中解偶联,并针对第三多核苷酸添加必需的步骤数。对于四个或更多个多核苷酸也是如此。

[0866] 本发明所述方法涉及将第一多核苷酸从膜中解偶联。如果正在研究三个或更多个多核苷酸,则本发明所述方法可以涉及将第二多核苷酸从膜中解偶联。

[0867] 可以采用任何已知的方法将第一多核苷酸从膜中解偶联。优选地,不采用跨膜孔将第一多核苷酸从步骤(f)中的膜解偶联。优选地,不采用电压或施加电势将第一多核苷酸从膜中解偶联。

[0868] 步骤(f)优选包括通过从膜中去除一个或多个锚来使第一多核苷酸从膜中解偶联。如果锚被去除,则采用其它(或单独的)锚将第二多核苷酸与膜偶联。用于偶联第二多核苷酸的锚可以与用于偶联第一多核苷酸的锚类型相同或可以为不同类型的锚。

[0869] 步骤(f)更优选地包括使一个或多个锚与一种试剂接触,所述试剂对于一个或多

个锚的亲和性比锚对于膜的亲和性高。各种用于竞争性结合或免疫放射测定法以确定分子的特定结合能力的方案是本领域公知的(例如,参见Maddox等人,J.Exp.Med.158,1211-1226,1993)。该试剂将锚从膜中去除,从而使第一多核苷酸解偶联。该试剂优选为糖。可以采用任何这样的糖:其与一个或多个锚结合的亲和性要比一个或多个锚对于膜的亲和性高。所述糖可以是如下所述的环糊精或其衍生物。

[0870] 如果一个或多个锚包括疏水锚,例如,胆固醇,则所述试剂优选为环糊精或其衍生物或脂质。环糊精或其衍生物可以是Eliseev,A.V.和Schneider,H-J(1994)J.Am.Chem.Soc.116,6081-6088中公开的任一种。所述试剂更优选为七-6-氨基-β-环糊精(am_7-βCD)、6-单脱氧-6-单氨基-β-环糊精(am_1-βCD)或七-(6-脱氧-6-胍基)-环糊精(gu_7-βCD)。可以使用本文中公开的任何脂质。

[0871] 如果锚包括链霉亲和素、生物素或脱硫生物素,则该试剂优选为生物素、脱硫生物素或链霉亲和素。生物素和脱硫生物素与链霉亲和素结合的亲和性均高于链霉亲和素与膜结合的亲和性,反之亦然。生物素比脱硫生物素具有更强的对链霉亲和素的亲和性。因此,可以使用生物素或链霉亲和素将包括链霉亲和素的锚从膜中去除,反之亦然。

[0872] 如果锚包括蛋白,则所述试剂优选为与蛋白特异性结合的抗体或其片段。如果抗体优先或以高亲和性与蛋白结合而不是仅以低亲和性与其它或不同蛋白结合,则所述抗体与蛋白特异性结合。如果抗体以1×10^{-6}M或更小的,更优选地1×10^{-7}M或更小的,5×10^{-8}M或更小的,更优选地1×10^{-8}M或更小的或更优选地5×10^{-9}M或更小的Kd结合,则其优先或以高亲和性结合。如果抗体以1×10^{-6}M或更多的,更优选地1×10^{-5}M或更多的,更优选地1×10^{-4}M或更多的,更优选地1×10^{-3}M或更多的,进一步优选地1×10^{-2}M或更多的Kd结合,则其是以低亲和性结合。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结合的方法是本领域已知的。抗体可以是单克隆抗体或多克隆抗体。抗体的合适片段包括但不限于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移植抗体或其片段或人源化抗体(humanised antibody)或其片段。

[0873] 步骤(f)优选包括使所述一个或多个锚与一种能降低所述一个或多个锚与膜偶联的能力的试剂接触。例如,所述试剂可能干扰一个或多个锚的结构和/或疏水性,并从而降低它们与膜偶联的能力。如果锚包括胆固醇,则所述试剂优选为胆固醇脱氢酶。如果锚包括脂质,则所述试剂优选为磷脂酶。如果锚包括蛋白,则所述试剂优选为蛋白酶或脲。合适的锚和试剂的其它组合对于本领域技术人员而言是清楚的。

[0874] 步骤(f)更优选地包括通过使第一多核苷酸与一个或多个锚分离来将第一多核苷酸从膜中解偶联。这可以以任何方式完成。例如,连接体可以在包括连接体的锚中被切割。本实施例特别适用于涉及通过杂合而连接的锚。这种锚在上文中进行了讨论。

[0875] 步骤(f)更优选地包括通过使第一多核苷酸及一个或多个锚与试剂接触来将第一多核苷酸从膜中解偶联,其中,所述试剂与第一多核苷酸竞争以便与一个或多个锚结合。竞争性结合的确定和测量方法是本领域已知的。所述试剂优选为与第一多核苷酸竞争以便与所述一个或多个锚杂合的多核苷酸。例如,如果采用一个或多个涉及杂合的锚将第一多核苷酸与膜偶联,则可以通过使所述一个或多个锚与同样杂合到杂合位点的多核苷酸接触来对多核苷酸进行解偶联。通常,以高于第一多核苷酸和一个或多个锚浓度的浓度添加多核苷酸试剂。替代地,与第一多核苷酸相比,多核苷酸试剂可以更强地与一个或多个锚杂合。

[0876] 步骤(f)更优选地包括(i)使第一多核苷酸及一个或多个锚与脲、三(2-羧基乙基)膦(TCEP)、二硫苏糖醇(DTT)、链霉亲和素或生物素、UV光、酶或结合剂接触;(ii)加热第一多核苷酸和一个或多个锚;或(iii)改变pH值。脲、三(2-羧基乙基)膦(TCEP)或二硫苏糖醇(DTT)能够破坏锚并将第一多核苷酸与膜分离。如果锚包括链霉亲和素-生物素连接,则链霉亲和素试剂争取与生物素结合。如果锚包括链霉亲和素-脱硫生物素连接,则生物素试剂争取与链霉亲和素结合。UV光可用于降解光不稳定基团。酶和结合剂可用于切割、分解或解开锚。优选的酶包括但不限于外切核酸酶、内切核酸酶或解旋酶。优选的结合剂包括但不限于酶、抗体或其片段或单链结合蛋白(SSB)。可以使用任何以下讨论的酶或上述讨论的抗体。热量和pH可用于破坏杂合和其它连接。

[0877] 如果通过将第一多核苷酸与一个或多个锚分离来将第一多核苷酸从膜中解偶联,则一个或多个锚将保留在膜中。步骤(g)优选地包括采用与第一多核苷酸分离的一个或多个锚将第二多核苷酸与膜偶联。例如,第二多核苷酸还可以具有一个或多个与保留在膜中的一个或多个锚杂合的多核苷酸。替代地,步骤(g)优选地包括采用与第一多核苷酸(即一个或多个其它锚)分离的锚中的一个或多个单独的锚将第二多核苷酸与膜偶联。所述一个或多个单独的锚可以与将第一多核苷酸与膜偶联的锚是同一类型,或者可以是不同类型的锚。步骤(g)优选地包括采用一个或多个锚将第二多核苷酸与膜偶联,其中所述一个或多个锚不同于与第一多核苷酸分离的一个或多个锚。

[0878] 在优选实施例中,步骤(f)和(g)包括通过使膜与第二多核苷酸接触使得第二多核苷酸与第一多核苷酸竞争以便与一个或多个锚结合并取代第一多核苷酸来将第一多核苷酸从膜中解偶联。例如,如果采用一个或多个涉及杂合的锚来将第一多核苷酸与膜偶联,则可以通过使锚与连接到多核苷酸(其也与一个或多个锚中的杂合位点杂合)的第二多核苷酸接触来对第一多核苷酸进行解偶联。通常以高于第一多核苷酸和一个或多个锚浓度的浓度添加第二多核苷酸。替代地,相比第一多核苷酸,第二多核苷酸可以更强地与一个或多个锚杂合。

[0879] 去除或洗涤

[0880] 虽然在步骤(f)中将第一多核苷酸从膜中解偶联,但其并不一定被去除或洗去。如果可以容易地将第二多核苷酸与第一多核苷酸区分开来,则不需要去除第一多核苷酸。

[0881] 在步骤(f)和(g)之间,所述方法优选地还包括从膜中去除至少一些第一样本。可以去除至少10%的第一样本,例如,可以去除至少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。

[0882] 所述方法更优选地还包括从膜中去除所有第一样本。这可以以任何方式完成。例如,在第一多核苷酸进行解偶联后,可以用缓冲液洗涤膜。以下对合适的缓冲液进行讨论。

[0883] 改性的多核苷酸

[0884] 在表征之前,在使用目标多核苷酸作为模板将聚合酶形成改性多核苷酸的条件下,可以通过使多核苷酸与聚合酶和一组游离核苷酸接触来对目标多核苷酸进行改性,其中,当形成改性多核苷酸时,所述聚合酶用不同核苷酸种类替代目标多核苷酸中的一个或多个核苷酸种类。然后,可以向改性多核苷酸提供与多核苷酸连接的一个或多个解旋酶以及与多核苷酸连接的一个或多个分子制动剂。这种类型的改性在申请号为1403096.9的英国申请中进行了描述。可以采用上述任一聚合酶。聚合酶优选为克列诺酶(Klenow)或9°

North。

[0885] 在使用模板多核苷酸作为模板将聚合酶形成改性多核苷酸的条件下,使模板多核苷酸与聚合酶接触。这些条件是本领域已知的。例如,多核苷酸通常与市售的聚合酶缓冲液中的聚合酶接触,例如,来自New England Biolabs®的缓冲液。对于克列诺酶,温度优选为20至37℃,或对于9°North,温度优选为60至75℃。通常将引物或3'发夹用作聚合酶延伸的成核点。

[0886] 使用跨膜孔进行的多核苷酸表征,如测序,通常包括对由k个核苷酸组成的聚合物单元进行分析,其中,k是正整数(即,'k-聚体(k-mers)')。这在申请号为PCT/GB2012/052343(公开为WO2013/041878)的国际申请中进行了讨论。尽管希望在不同k-聚体的电流测量值之间有明确的分离,但是这些测量值中的一些发生重叠是很常见的。特别是k-聚体中具有高数量的聚合物单元,即,高k值,分辨不同k-聚体产生的测量值变得比较困难,从而不利于得到有关多核苷酸的信息,例如,估计多核苷酸的潜在序列。

[0887] 通过用改性多核苷酸中的不同核苷酸种类替代目标多核苷酸中的一个或多个核苷酸种类,所述改性多核苷酸包含与目标多核苷酸不同的k-聚体。改性多核苷酸中的不同k-聚体能够产生与目标多核苷酸中的k-聚体不同的电流测量值,因此改性多核苷酸提供与目标多核苷酸不同的信息。来自改性多核苷酸的附加信息可以使得更容易表征目标多核苷酸。在一些情况下,改性多核苷酸本身可能更易于表征。例如,改性多核苷酸可以设计为包括电流测量值之间具有增加的分离或清楚的分离的k-聚体,或具有降低的噪声的k-聚体。

[0888] 当形成改性多核苷酸时,聚合酶优选用不同核苷酸种类替代目标多核苷酸中的两个或更多个核苷酸。聚合酶可以用不同的核苷酸种类替代目标多核苷酸中的所述两个或更多个核苷酸种类的每一个。聚合酶可以用同一核苷酸种类替代目标多核苷酸中的所述两个或更多个核苷酸种类的每一个。

[0889] 如果目标多核苷酸为DNA,则改性多核苷酸中的不同核苷酸种类通常包括不同于腺嘌呤、鸟嘌呤、胸腺嘧啶、胞嘧啶或甲基胞嘧啶的核碱基和/或包括不同于脱氧腺苷、脱氧鸟苷、胸苷、脱氧胞苷或脱氧甲基胞苷的核苷。如果目标多核苷酸为RNA,则改性多核苷酸中的不同核苷酸种类通常包括不同于腺嘌呤、鸟嘌呤、尿嘧啶、胞嘧啶或甲基胞嘧啶的核碱基和/或包括不同于腺苷、鸟苷、尿苷、胞苷或甲基胞苷的核苷。不同的核苷酸种类可以是上述任一通用核苷酸。

[0890] 聚合酶可以用不同的核苷酸种类替代一个或多个核苷酸种类,其中,所述不同的核苷酸种类包括所述一个或多个核苷酸种类中不存在的化学基团或原子。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、苄基、炔丙基或炔丙胺基。

[0891] 聚合酶可以用不同核苷酸种类替代一个或多个核苷酸种类,其中,所述不同核苷酸种类缺少所述一个或多个核苷酸种类中存在的化学基团或原子。聚合酶可以用具有改变的电负性的不同核苷酸种类替代一个或多个核苷酸种类。具有改变的电负性的不同核苷酸种类优选地包括卤素原子。

[0892] 所述方法优选地还包括选择性地从改性多核苷酸的一个或多个不同核苷酸种类中去除核碱基。

[0893] 分析物输送

[0894] 目标分析物优选地与将分析物向膜输送的微粒连接。这种类型的输送在申请号为

1418469.1的英国申请中进行了公开。

[0895] 可以使用任何类型的微粒和连接方法。

[0896] 其它表征方法

[0897] 在另一实施例中,通过检测聚合酶添加给目标多核苷酸的标记种类然后进行释放来对多核苷酸进行表征。聚合酶使用多核苷酸作为模板。每个标记种类对每个核苷酸具有特异性。多核苷酸与CsgG孔或其突变体例如本发明所述孔、聚合酶以及标记的核苷酸接触,使得磷酸酯标记物质由聚合酶顺序地加入到多核苷酸,其中,磷酸酯物质包含对每个核苷酸特异性的标记。标记的物质可以在其从核苷酸释放之前(即,当其加入到目标多核苷酸中时)或在其从核苷酸释放之后采用孔进行检测。

[0898] 聚合酶可以是上述任一种。采用孔检测磷酸酯标记物质,从而表征多核苷酸。这种方法在申请号为13187149.3(公开为EP2682460)的欧洲申请中进行了公开。上述任一实施例同样适用于该方法。

[0899] 标记的物质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。这种标签的非限制性实例可以在Kumar等人Sci Rep.2012;2:684电子出版2012年9月21日的著作中发现。

[0900] 传感器的形成方法

[0901] 本发明还提供了形成传感器以用于表征目标多核苷酸的方法。所述方法包括在诸如本发明所述孔等CsgG孔或其突变体与诸如解旋酶或外切核酸酶等多核苷酸结合蛋白之间形成复合物。可以通过在存在目标多核苷酸的情况下使孔和蛋白接触,然后在孔上施加电势来形成复合物。施加的电势可以是如上所述的化学电势或电压电势。替代地,可以通过使孔与蛋白共价连接来形成复合物。用于共价连接的方法是本领域已知的并在以下国际申请中进行了公开:例如,申请号为PCT/GB09/001679(公开为WO2010/004265)和PCT/GB10/000133(公开为WO2010/086603)的国际申请。复合物是一种用于表征目标多核苷酸的传感器。所述方法优选地包括在诸如本发明所述孔等CsgG孔或其突变体和解旋酶之间形成复合物。上述任一实施例同样适用于该方法。

[0902] 本发明还提供了一种用于表征目标多核苷酸的传感器。传感器包括在诸如本发明所述孔等CsgG孔或其突变体与多核苷酸结合蛋白之间的复合物。上述任一实施例同样适用于本发明所述传感器。

[0903] 试剂盒

[0904] 本发明还提供了一种用于表征目标多核苷酸的试剂盒。试剂盒包括CsgG孔或其突变体,例如,本发明所述孔,以及膜的组分。膜优选由各组分形成。孔优选存在于膜中。试剂盒可以包括上述任一膜的组分,例如,两亲层或三嵌段共聚物膜。

[0905] 试剂盒还可以包括多核苷酸结合蛋白。可以使用上述任一多核苷酸结合蛋白。

[0906] 试剂盒还可以包括一个或多个用于将多核苷酸与膜偶联的锚。

[0907] 试剂盒优选用于表征双链多核苷酸,并且优选包括Y适配器和发夹环适配器。Y适配器优选具有一个或多个连接的解旋酶,发夹环适配器优选地具有一个或多个连接的分子制动剂。Y适配器优选包括一个或多个用于将多核苷酸与膜偶联的第一锚,发夹环适配器优选地包括一个或多个用于将多核苷酸与膜偶联的第二锚,并且发夹环适配器与膜的偶联强度优选地大于Y适配器与膜的偶联强度。

[0908] 本发明所述试剂盒可以另外包括一个或多个能够使上述任一实施例得以执行的其它试剂或仪器。这种试剂或仪器包括以下一种或多种:合适的缓冲液(水性溶液)、从受体(例如,血管或包括针头的仪器)获得样本的装置、用于扩增和/或表达多核苷酸的装置或电压或膜片钳装置。试剂可以以干燥状态存在于试剂盒中,使得流体样本使试剂重新悬浮。试剂盒还可以可选地包括能够使试剂盒用于本发明所述方法的说明或有关所述方法可以用于哪种生物体的细节。

[0909] 设备

[0910] 本发明还提供了一种用于表征目标分析物(例如,目标多核苷酸)的设备。所述设备包括多个CsgG孔或其突变体和多个膜。所述多个孔优选存在于所述多个膜中。孔和膜的数量优选相等。优选地,每个膜中存在单个孔。

[0911] 所述设备优选地还包括用于实施本发明所述方法的指令。所述设备可以是任一用于分析物分析的常规设备,例如,阵列或芯片。结合本发明所述方法所讨论的任一实施例同样适用于本发明所述设备。所述设备还可以包括本发明所述试剂盒中存在的任何特征。

[0912] 优选地,对所述设备进行设置以便执行本发明所述方法。

[0913] 所述设备优选地包括:

[0914] 传感器装置,其能够支撑多个孔和膜并且可操作地采用孔和膜进行分析物表征;和

[0915] 至少一个端口,用于传送用以进行表征材料。

[0916] 替代地,所述设备优选地包括:

[0917] 传感器装置,其能够支撑多个孔和膜并可操作地采用孔和膜进行分析物表征;和

[0918] 至少一个储存器,用于容纳用以进行表征的材料。

[0919] 所述设备更优选地包括:

[0920] 传感器装置,其能够支撑膜和多个孔,并且可操作地采用孔和膜进行分析物表征;

[0921] 至少一个储存器,用于容纳用以进行表征的材料;

[0922] 流体系统,用于可控地将材料从所述至少一个储存器向所述传感器装置供应;和

[0923] 一个或多个容器,用于接收相应的样本,所述流体系统用于选择性地将样本从所述一个或多个容器向所述传感器装置供应。

[0924] 所述设备可以是以下国际申请中描述的任一种:申请号为PCT/GB08/004127(公开为WO2009/077734),PCT/GB10/000789(公开为WO2010/122293),申请号为PCT/GB10/002206(公开WO 2011/067559)或申请号为PCT/US99/25679(公开为WO00/28312)的国际申请。

[0925] 下列实施例对本发明进行说明。

[0926] 实施例

[0927] 实施例1:克隆和CsgG菌株

[0928] 已经描述了用于生产外膜局部C-末端StrepII标记的CsgG(pPG1)和周质C-末端StrepII标记的$CsgG_{C1S}$(pPG2)的表达构建体(Goyal,P.et al.,Acta Crystallogr.F.Struct.Biol.Cryst.Commun.2013,69,1349-1353)。对于硒代甲硫氨酸(selenomethionine)标记,StrepII标记的$CsgG_{C1S}$由于产率增加而在细胞质中进行表达。因此,对pPG2进行改变以便采用带有引物5'-TCT TTA AC CGC CCC GCC TAA AG-3'(正向)(SEQ ID NO:437)和5'-CAT TTT TTG CCC TCGTTA TC-3'(反向)(pPG3)(SEQ ID NO:438)的反向PCR去除N末端信号

肽。对于表型测定,通过所描述的方法[Datsenko,K.A.et al.,Proc.Natl Acad.Sci.USA 97,6640-6645(2000)]构建大肠杆菌BW25141(大肠杆菌NVG2)的csgG缺失突变体(用引物5'-AAT AAC TCA ACC GAT TTT TAA GCC CCA GCT TCA TAA GGA AAA TAA TCG TGT AGG CTG GAG CTG CTT C-3'(SEQ ID NO:439)和5'-CGC TTA AAC AGT AAA ATG CCG GAT GAT AAT TCC GGC TTT TTT ATC TGC ATA TGA ATA TCC TCC TTA G-3'(SEQ ID NO:440))。在trc启动子控制下,通过始于pMC2(含有csgG的pTRC99a载体)的定点诱变(QuikChange方案;Stratagene)来构建用于Cys可接近性试验并用于通道收缩结构的表型探测的各种CsgG取代型突变体(Robinson,L.S.et al.,Mol.Microbiol.2006,59,870-881)。

[0929] 实施例2:蛋白的表达和纯化

[0930] 如所述(Robinson,L.S.et al.,Mol.Microbiol.2006,59,870-881)对CsgG和$CsgG_{C1S}$进行表达和纯化。简言之,在大肠杆菌BL21(DE3)中重组产生CsgG,其中,所述大肠杆菌BL21(DE3)采用pPG1转化并使用缓冲液A(50mM Tris-HCl pH8.0,500mM NaCl,1mM EDTA;1mM二硫苏糖醇(DTT))中1%的正十二烷基-β-D-麦芽糖苷(DDM)从分离的外膜中进行提取。将StrepII标记的CsgG装载到5毫升Strep-Tactin琼脂糖柱(Iba GmbH)上,并通过用20个柱体积的补充有0.5%的四甘醇单辛基醚(C8E4;Affymetrix)和4mM月桂基二甲基胺-N-氧化物(LDAO;Affymetrix)的缓冲液A洗涤来进行洗涤剂交换。将蛋白通过添加2.5mM D-脱硫生物素进行洗脱,并浓缩至5mg ml^{-1}用于结晶实验。对于硒代甲硫氨酸标记,在用pPG3转化并在补充有40mg l^{-1}L-硒代甲硫氨酸的M9基本培养基上生长的Met营养缺陷型菌株B834(DE3)中产生$CsgG_{C1S}$。使细胞颗粒(Cell pellet)重悬于补充有完整的蛋白酶抑制剂混合物(Protease Inhibitor Cocktail(Roche))的50mM Tris-HCl pH 8.0,150mM NaCl,1mM EDTA,5mM DTT中,并通过穿过以20×10^{3}lb in^{-2}操作的TS系列细胞破碎剂(Constant Systems Ltd)而破碎。如所述(Robinson,L.S.et al.,Mol.Microbiol.2006,59,870-881),对标记的$CsgG_{C1S}$进行纯化。在整个纯化过程中加入DTT(5mM)以避免硒代甲硫氨酸的氧化。

[0931] 在携带有pNH27的大肠杆菌NEBC2566细胞中产生CsgE(Nenninger,A.A.et al.,Mol.Microbiol.2011,81,468-499)。将25mM Tris-HCl pH 8.0,150mM NaCl,25mM咪唑,5%(v/v)glycerol甘油中的细胞溶解产物装载在HisTrap FF上(GE Healthcare)。用在20mM Tris-HCI pH 8.0,150mM NaCl,5%(v/v)甘油缓冲液中的500mM咪唑采用线性梯度对CsgE-组氨酸(CsgE-his)进行洗脱。对含有CsgE的馏分补充250mM$(NH_4)_2SO_4$,并施加到用20mM Tris-HCl pH 8.0,100mM NaCl,250mM$(NH_4)_2SO_4$、5%(v/v)甘油平衡的5ml HiTrap苯基HP柱(GE Healthcare)上。用20mM Tris-HCl pH 8.0,10mM NaCl,5%(v/v)甘油以线性梯度进行洗脱。将含有馏分的CsgE装载到Superose 6Prep Grade 10/600柱(GE Healthcare)上,所述柱在20mM Tris-HCl pH 8.0,100mM NaCl,5%(v/v)甘油中平衡。

[0932] 实施例3:溶液中寡聚状态评估

[0933] 将各约0.5mg的洗涤剂溶解的CsgG(0.5%C8E4,4mM LDAO)和$CsgG_{C1S}$施用到用25mM Tris-HCl pH 8.0,500mM NaCl,1mM DTT,4mM LDAO和0.5%C8E4(CsgG)平衡或用25mM Tris-HCl pH 8.0,200mM NaCl($CsgG_{C1S}$)平衡的Superdex 200 10/300GL分析凝胶过滤柱(GE Healthcare)上,并以0.7ml min^{-1}运行。用牛甲状腺球蛋白、牛γ-球蛋白、鸡卵清蛋白、马肌球蛋白和维生素B_{12}(Bio-Rad)对柱洗脱体积进行校准(图7)。采用Swamy,M.,et al.,Sci.STKE 2006,pl4[http://dx.doi.org/10.1126/stke.3452006pl4(2006)]中所描述的

程序,将膜提取的CsgG,即20μg洗涤剂溶解的蛋白,也在3-10%蓝色天然PAGE上运行(图7)。NativeMark(Life Technologies)未染色蛋白标准物(7μl)用于分子质量估计。主要发现成熟的CsgG是具有C9对称性的离散型九聚体成孔颗粒以及九聚体孔的尾对尾二聚体(即,具有D9对称性的十八聚体)。出于纳米孔感测应用的目的,蛋白的优选状态为单个九聚体孔。可以通过在体积排阻色谱法和/或插入用于纳米孔感测应用的脂质双层之前加热样本来增加九聚体孔对D9十八聚体孔的数量。

[0934] 实施例4:结晶、数据收集和结构测定

[0935] 通过对包含有100mM pH值为4.2的乙酸钠、8%PEG 4000和100mM pH值为7.0的丙二酸钠的溶液进行坐滴式(sitting-drop)蒸气扩散,将硒代甲硫氨酸标记的$CsgG_{C1S}$浓缩至3.8mg ml^{-1}并进行结晶。在补充有15%甘油的晶体缓冲液中对晶体进行培养,并在液氮中快速冷冻。通过对包含有100mM Tris-HCl pH8.0、8%PEG 4000、100mM NaCl和500mM $MgCl_2$的溶液进行悬滴式(hanging-drop)蒸气扩散,将洗涤剂溶解的CsgG浓缩至5mg ml^{-1}并进行结晶。将晶体在液氮中快速冷冻并通过存在于晶体溶液中的洗涤剂进行冷冻保护。为了优化结晶条件并筛选具有良好衍射质量的晶体,在光束线Proxima-1和Proxima-2a(Soleil,法国)、PX-1(瑞士光源,瑞士)、102、103、104和124(钻石光源,英国)以及ID14eh2、ID23eh1和ID23eh2(ESRF,法国)上对晶体进行分析。分别在光束线104和103收集用于$CsgG_{C1S}$和CsgG结构测定的最终衍射数据(表5)。

数据收集和细化统计		
数据收集	$CsgG_{C1S}$	CsgG
空间群	P1	C2
细胞大小		
a, *b*, *c* (Å)	101.3, 103.6, 141.7	161.9, 372.8, 161.9
α, *β*, *γ* (°)	111.3, 90.5, 118.2	90.0, 92.9, 90.0
分辨率(Å)*	30-2.8 (2.9-2.8)	30-3.6 (3.7-3.6) 30 – 3.6 (a*), –3.7 (b*), –3.8 (c*)
Rmeas *	15.1 (81.8)	16.2 (90.6) †
I/σ*I**	9.82 (2.03)	6.80 (1.89) †
完整度(%)*	98.7 (98.3)	91.57 (27.26) 99.9 (99.1) †
冗余*	11.2 (7.0)	4.4 (4.3)
Wilson B($Å^2$)	46.7	101.0

[0936]

细化		
分辨率(Å)*	30-2.8 (2.9-2.8)	30-3.6 (3.7-3.6) 30 – 3.6 (a*), –3.7 (b*), –3.8 (c*)
反射数量	112419 (11159)	102130 (11094)
R_{work}/R_{free}	0.1881 / 0.2337	0.3024 / 0.3542
原子数量		
蛋白	28853	34165
配体/离子	0	0
水	0	0
B-因子($Å^2$)		
蛋白	57.3	116.7
配体/离子		
水		
R.m.s偏差		
键长 (Å)	0.01	0.03
键角 (°)	1.31	1.87

[0937] CsgGC1S和膜提取的CsgG(均从单个晶体收集)的数据统计;

[0938] *最高分辨率壳在括号中示出。

[0939] †沿互逆方向a*,b*和c*上的各向异性截断的纠正值。

[0940] 表5

[0941] 采用Xia2和XDS封装(Winter,G.,J.Appl.Cryst.,2010,43,186-190;Kabsch,W.,Acta Crystallogr.D Bi01.Crystallogr.2010,66,125-132)处理$CsgG_{C1s}$的衍射数据。$CsgG_{C1s}$的晶体属于空间群P1,晶胞大小为a=101.3 Å, b=103.6 Å, c=141.7 Å,α=111.3°,β=90.5°,γ=118.2°,不对称单元中含有16个蛋白复本。为了结构测定和细化,在0.9795 Å波长处收集的数据基于最高分辨率层中2的//σ/截止在2.8 Å截断。采用从单一异常色散(SAD)实验计算的实验相来求解结构。通过使用ShelxC和ShelxD(Sheldrick,G.M.,Acta Crystallogr.D Biol.Crystallogr.201 0,66,479-485),在不对称单元中定位总共92个硒位点,将其细化并使用Sharp(Bricogne,G.,Acta Crystallogr.D Biol.Crystallogr.2003,59,2023-2030)(相位功率0.79,品质因数(FOM)0.25)用于相位计

算。通过非晶体学对称性(NCS)采用Parrot(Cowtan,K.,Acta Crystallogr.D Biol.Crystallogr.2010,66,470-478(图7;FOM 0.85)将实验相进行密度修正和平均。采用Buccaneer(Cowtan,K.,Acta Crystallogr.D Biol.Crystallogr.2006,62,1002-1011)构建初始模型,并采用Phenix细化(Adams,P.D.et al.Acta Crystallogr.D Biol.Crystallogr.2010,66,213-221)和手工检查以及在Coot(Emsley,P.et al.,Acta.Crystallogr.D Biol.Crystallogr.2010,66,486-501)中(重建)构建的模型通过最大似然细化的迭代循环进行细化。最终结构含有28,853个原子,这些原子在属于16条$CsgG_{C1S}$链的3,700个残基中(图7),其中,molprobity(Davis,I.W.etal.,Nucleic Acids Res.35(Suppl 2),W375-W383)值为1.34;98%的残基位于拉氏图(允许区域为99.7%)的优选区域中。电子密度图显示与可能的溶剂分子对应的不明确密度,因此未建立水分子或离子。在整个细化过程中,维持十六倍的NCS平均,分别在细化早期和晚期采用严格和局部NCS约束。

[0942] 在空间群C2中,采用XDS封装(Winter,G.,J.Appl.Cryst.2010,43,186-190;Kabsch,W.,Acta Crystallogr.D Biol.Crystallogr.2010,66,125-132)从0.9763Å波长处的单晶收集CsgG的衍射数据,并编入索引且进行缩放,其中,所述空间群的晶胞尺寸为a=161.7Å,b=372.3Å,c=161.8Å和β=92.9°,包括非对称单元中的18个CsgG复本和72%的溶剂含量。沿互逆单元方向a*,b*和c*将用于结构测定和细化的衍射数据椭圆形地截取为3.6Å,3.7Å和3.8Å的分辨率极限,并用衍射各向异性服务器(Diffraction Anisotropy Server)进行各向异性地缩放(Strong,M.et al.,Proc.Natl.Acad.Sci.USA 2006,103,8060-8065)。采用$CsgG_{C1S}$单体的分子置换被证明不成功。对自旋函数的分析揭示了不对称单元中的D9对称性(未示出)。在$CsgG_{C1S}$结构基础上,假设从C_8去到C_9寡聚物之后,原聚体间的角度(interprotomer angle)从45°变化到40°,颗粒周围的原聚体间的弧度(interprotomer arc)保持大致相同,从而使半径计算的增加约为4 Å,生成九聚搜索模型。将所计算的九聚体用作搜索模型,用Phaser(McCoy,A.J.et al.,J.Appl.Colyst.2007,40,658-674)找到包含两个复本的分子置换溶液。对密度修正的和NCS平均的电子密度图(Parrot[Cowtan,K.,Acta Crystallogr.D Biol.Crystallogr.2010,66,470-478];图8)的检查使得能够手工建立TM1和TM2并重塑蛋白核心中的相邻残基,以及构建残基2-18,其从$CsgG_{C1S}$模型中缺失并将α1螺旋物连接到N-末端脂质锚。分别使用用于初始和最终细化循环的Buster-TNT(Smart.O.S.et al.,Acta Crystallogr.D Biol.Crystallogr.2012,68,368-380)和Refmac5(Murshudov,G.N.et al.,Acta Crystallogr.D Biol.Crystallogr.,2011,67,355-367)对CsgG模型进行细化。在整个细化过程中应用十八倍的局部NCS限制,并且采用具有由Prosmart(Nicholls,R.A.,Long,F.&Murshudov,G.N.,Acta Crystallogr.D Biol.Crystallogr.2011,68,404-417)产生的σ0.01和氢键约束的凝胶体细化来运行Refmac 5。最终结构包括4,451个残基中的34,165个原子,这些残基属于18个CsgG链(图7),其中molprobity值为2.79;93.0%的残基位于拉式图(允许区域为99.3%)的优选区域。不能识别对应于N-末端脂质锚的不明确电子密度。

[0943] 实施例5:刚果红试验

[0944] 为了分析刚果红结合,在LB(Lysogeny Broth)培养基中对LB中在37℃生长的细菌过夜培养物进行稀释,直至达到0.5的D_{600}。在补充有氨苄青霉素(100mg l^{-1})、刚果红(100mg

l^{-1})和0.1％(w/v)异丙基β-D-硫代半乳糖苷(IPTG)的LB琼脂板上点样5μl样本。在室温(20-22℃)下将平板培养48小时以诱导卷曲菌毛表达。观察菌落形态发展和染色结合情况48h。

[0945] 实施例6:半胱氨酸可接近性试验

[0946] 采用定点突变在pMC2中产生半胱氨酸突变体,并在大肠杆菌LSR12(Chapman,M.R.et al.,Science,2002,295,851-855)中进行表达。将生长过夜的细菌培养物点样到含有1mM IPTG和100mg l^{-1}氨苄青霉素的LB琼脂板上。将平板在室温下培养,并在48小时后刮下细胞,重悬于1ml的PBS中并用D_{600}归一化。通过超声降解法裂解细胞,并在4℃温度下以3,000g离心20秒,以便从细胞裂解液和悬浮膜中去除未破裂的细胞。将上清液中的蛋白在室温下用15mM甲氧基聚乙二醇-马来酰亚胺(MAL-PEG 5kDa)标记1小时。用100mM DTT终止反应,并在4℃温度下在50.4Ti转子中以40,000r.p.m(~100,000g)离心20分钟以便将总的膜制成小球。用1％月桂酰肌氨酸钠洗涤所述小球以溶解细胞质膜并再次离心。将所得的外膜用含有1％DDM的PBS重新悬浮并溶解。使用镍珠的金属亲和性下拉(pulldowns)用于SDS-PAGE和anti-His蛋白质印迹。具有空载pMC2载体的大肠杆菌LSR12细胞用作阴性对照。

[0947] 实施例7:ATR-FTIR光谱法

[0948] 在Equinox 55红外光谱仪(Bruker)上进行ATR-FTIR测量,连续用干燥空气吹扫,配备有液氮制冷型汞镉碲化物探测器和Golden Gate反射附件(Specac)。内反射元件为金刚石晶体(2mm×2mm),光束入射角为45°。将每个纯化的蛋白样本(1μl)在晶体表面铺展并在气态氮气流下干燥以形成膜。以2cm^{-1}分辨率记录的每个光谱为128个累计值的平均值用于提高的信噪比。所有光谱采用水蒸汽贡献减法来处理,通过衍射控象法(apodization)在最终分辨率为4cm^{-1}情况下变平滑,并在Amide I谱带(1,700-1,600cm^{-1})的区域上进行归一化以使其进行比较(Goormaghtigh,E.;Ruysschaert,J.M.,Spectrochim.Acta,1994,50A,2137-2144)。

[0949] 实施例8:阴性染色EM和对称性测定

[0950] 阴性染色EM用于监测CsgG、$CsgG_{C1S}$和CsgE的溶液中(in-solution)寡聚状态。将CsgE、$CsgG_{C1S}$和两亲性结合(amphipol-bound)的CsgG调节成0.05mg ml^{-1}的浓度,并施加到辉光放电的碳涂覆铜网格上(CF-400;Electron Microscopy Sciences)。培养1分钟后,将样本染上印记,然后在2％的乙酸双氧铀中洗涤并染色。在Tecnai T12BioTWIN LaB6显微镜上收集图像,操作电压为120kV,放大倍数为×49,000,散焦为800至2,000nm之间。如对于cryo-EM所述,执行对比传递函数(CTF)、相位翻转和颗粒选择。对于膜提取的(membrane-extracted)CsgG,分别从九聚体和顶视图分析十八聚体颗粒(共1,780个)。对于纯化的CsgE,总共分析了2,452个颗粒。正如以下CsgG-CsgE cryo-EM分析所述,获得三维模型,并通过几轮多参考对准(multi-reference alignment,MRA)、多统计分析(MSA)和锚设置细化进行细化。在所有情况下,归一化和中心化后,采用如cryo-EM部分中所述的IMAGIC-4D对图像进行分类。对于每组颗粒,选择与特征视图对应的最佳分类。采用最佳分类平均值进行CsgG顶视图的对称性测定,每类大约有20张图像。使用IMAGIC计算旋转自相关函数并进行绘制。

[0951] 实施例9:阴性染色EM和对称性测定

[0952] 阴性染色EM用于监测CsgG、$CsgG_{C1S}$和CsgE的溶液中(in-solution)寡聚状态。将CsgE、$CsgG_{C1S}$和两亲性聚合物结合(amphipol-bound)的CsgG调节成0.05mg ml^{-1}的浓度,并

施加到辉光放电的碳涂覆铜网上(CF-400;Electron Microscopy Sciences)。培养1分钟后,将样本染上印迹,然后在2%的乙酸双氧铀中洗涤并染色。在Tecnai T12BioTWIN LaB6显微镜上收集图像,操作电压为120kV,放大倍数为×49,000,散焦为800至2,000nm之间。如对于cryo-EM所述,执行对比传递函数(CTF)、相位翻转和颗粒选择。对于膜提取的CsgG,分别通过九聚体和顶视图分析十八聚体颗粒(共1,780个)。对于纯化的CsgE,总共分析了2,452个颗粒。针对以下CsgG-CsgE cryo-EM分析所述,获得三维模型,并通过几轮多参考对准(multi-reference alignment,MRA)、多统计分析(MSA)和锚设置细化进行细化。在所有情况下,归一化和中心化后,采用如cryo-EM部分中所述的IMAGIC-4D对图像进行分类。对于每组颗粒,选择与特征视图对应的最佳分类。采用最佳分类平均值进行CsgG顶视图的对称性测定,每类大约有20张图像。使用IMAGIC计算旋转自相关函数并进行绘制。

[0953] 实施例10:CsgG-CsgE复合物形成

[0954] 对于CsgG-CsgE复合物形成,通过向300ml洗涤剂去稳定的脂质体(1mg ml^{-1} 1,2-二肉豆蔻酰-sn-甘油-3-磷酸胆碱(DMPC)和0.4%的LDAO)添加120ml CsgG(0.5%C8E4中的24mg ml^{-1}蛋白、4mM LDAO、25mM pH值为8.0的Tris-HCl、500mM NaCl、1mM DTT)并在加入90ml(在100mg ml^{-1}原料(stock)下)的A8-35两亲聚合物(A8-35amphipols)前在冰上培养5分钟将纯化的CsgG中的增溶洗涤剂替换为两亲聚合物A8-35(Anatrace)。在冰上培养另外15分钟后,将样本装载到Superose 6 10/300GL(GE Healthcare)柱上,并在200mM NaCl、2.5%木糖醇、25mM pH为8的Tris-HCl、0.2mM DTT中进行凝胶过滤。将200mM NaCl、2.5%木糖醇、25mM pH值为8的Tris-HCl、0.2mM DTT中的等体积纯化单体CsgE以对于CsgE和CsgG分别为15和5μM的最终蛋白浓度加入到两亲聚合物溶解的CsgG,样本在温度为18℃、为、125V电压下在0.5×TBE缓冲液中4.5%的天然PAGE上运行。对于第二个变性维度,将对应于CsgG-CsgE复合物的条带从在同一凝胶上平行运行的未染色的泳道(lanes)中切除,在Laemmli缓冲液(60mM pH值为6.8的Tris-HCl、2%的SDS、10%甘油、5%的2-巯基乙醇、0.01%溴酚蓝)中煮沸5分钟,并在4-20%的SDS-PAGE上运行。纯化的CsgE和CsgG作为对照样本与复合物一起运行。用考马斯亮蓝(InstantBlue Coomassie)对凝胶进行染色,以便进行目视检查,或用SYPRO橙染色,以便通过对SDS-PAGE上的CsgE和CsgG条带进行荧光检测(Typhoon FLA 9000)来对CsgG-CsgE复合物进行化学计量评估,得到0.97的CsgG/CsgE比。

[0955] 实施例11:CsgG-CsgECryo-EM

[0956] Cryo电子显微镜用于测定C_9CsgG-CsgE复合物的溶液中(in-solution)结构。将如上所述制备的CsgG-CsgE复合物与HisTrap FF(GE Healthcare)结合并洗脱以去除未结合的CsgG,洗脱后立即将其应用于Quantifoil R2/2碳涂覆网格(Quantifoil Micro Tools GmbH),所述Quantifoil R2/2碳涂覆网格已在20mA下辉光放电30秒。使用自动化系统(Leica)将样本冷冻浸入液氮中,并在FEI F20显微镜下观察,其操作电压为200kV,并且在低剂量条件下标称放大倍数为×50,000,散焦范围为1.4-3mm。在Falcon II探测器上记录图像帧。标本级别的像素大小为每像素1.9Å。采用CTFFIND3(Mindell,J.A.&Grigorieff,N.,J.Struct.Biol.2003,142,334-347)评估CTF参数,并在SPIDER(Shaikh,T.R.et al.,Nature Protocols,2008,3,1941-1974)中进行相位翻转。采用BOXER(EMAN2;Tang,G.et al.,J.Struct.Biol.,2007,157,38-46)从CTF校正的显微照片中自动挑选颗粒。丢弃散光超过10%的图像。从75张显微照片中总共选出1,221个颗粒,并且窗口化成128-像素×128-

像素盒子。将图像归一化成同一平均值和标准偏差,并在~200 Å的低分辨率截止下进行高通滤波。使其居中,然后进行第一轮MSA。采用IMAGIC-4D(图像科学软件(Image Science Software))中的参考自由分类获得初始参考集。与C_9圆柱形颗粒的特征侧视图对应的最佳类别被用作MRA的参考。对于CsgG-CsgE复合物,从最佳的125个特征视图(具有良好的对比度和明确定义的特征)中计算出第一个三维模型,所述125个特征视图包括复合物的1,221个颗粒,其角度重构由方向确定(Image Science Software)。三维图通过MRA、MSA和锚设置细化的多轮迭代进行改进。根据0.5标准水平,分辨率通过傅里叶壳相关性(Fourier shell correlation,FSC)估计为24 Å(图5)。在UCSF Chimera(Pettersen,E.F.et al.,J.Comput.Chem.,2004,25,1605-1612)中进行图和数字的可视化。

[0957] 实施例12:CsgG和CsgG+CsgE孔的单通道电流分析

[0958] 在负场电势下,CsgG孔显示出两种电导状态。图5分别显示正常(在+50,0和-50mV下测量)和低电导形式(在0,+50和-50mV下测量)的代表性单通道电流迹线。在总观察时间(n=22)期间,没有观察到两种状态之间的转换,表明电导状态具有长的寿命(秒到分的时间量程)。左下图显示了在+50和-50mV(n=33)情况下获得的CsgG孔的正常和低电导形式的电流直方图。右下图显示了具有规则和低电导率的CsgG孔的I-V曲线。数据表示至少四个独立记录值的平均值和标准偏差。低电导形式的性质或生理存在是未知的。

[0959] 图6显示了用辅助因子CsgE滴定的CsgG通道的电生理学结果。该图将开放、中间和闭合通道的馏分显示为CsgE浓度的函数。CsgG的开放和闭合状态如图6所示(分别为0nM和100nM的CsgE)。将CsgE的浓度增加到10nM以上导致CsgG孔闭合。该效应发生在+50mV(左)和-50mV(右),排除了将CsgE(计算的pI为4.7)电泳到CsgG孔中导致孔堵塞的可能性。不频繁(<5%)的中间状态的电导大约是开放通道的一半。这可能表示CsgE引起CsgG通道不完全闭合;可选地,这可能表示电解质溶液中的残留CsgG单体与膜嵌入孔结合而引起CsgG二聚体的暂时形成。通过对单通道电流迹线进行全点直方图分析获得三种状态的馏分。直方图中产生的多达三种状态的峰值区域、给定状态的馏分通过将相应的峰值区域除以记录中所有其它状态的总和来获得。在负场电势下,可以看到两个开放电导状态,类似于CsgG的观察结果(见a)。由于两个开放通道的变化都被较高的CsgE浓度阻断,图6(0nM)中的“开放”迹线结合了两种电导形式。图中的数据表示三个独立记录值的平均值和标准偏差。

[0960] 晶体结构、尺寸排阻层析和EM显示了洗涤剂提取的CsgG孔在较高的蛋白浓度下形成非天然的尾对尾层叠式(stacked)二聚体(例如,两个九聚体作为D9颗粒,图7)。这些二聚体也可以在单通道记录中观察到。上图显示了+50,0和-50mV(从左到右)情况下的层叠式CsgG孔的单通道电流迹线。左下图显示了+50和-50mV情况下记录的二聚CsgG孔的电流直方图。在+50和-50mV情况下,+16.2±1.8和-16.0±3.0pA(n=15)的实验电导值分别接近理论计算值23pA。右下图显示了层叠式CsgG孔的I-V曲线。数据表示六个独立记录值的平均值和标准偏差。

[0961] 通过电生理学对CsgE结合和阻断层叠式CsgG孔的能力进行测试。示出了在+50mV(上位)和-50mV(下位)情况下,在10或100nM CsgE存在下,层叠式CsgG孔的单通道电流迹线。电流痕迹表明,CsgE的其它饱和浓度不会导致层叠式CsgG二聚体的孔闭合。由EM和定点诱变可以看出,这些观察结果与CsgG-CsgE接触区映射到螺旋物2以及CsgG周质腔体的开口很好地吻合(图5和图6)。

[0962] 实施例13:胆汁盐毒性试验

[0963] 通过在含有胆汁盐的琼脂板上减慢生长来研究外膜渗透性。将大肠杆菌LSR12(Chapman,M.R.et al.,Science,2002,295,851-855)细胞的十倍连续稀释液(5ml)点样在McConkey琼脂板上,其中,所述稀释液同时含有pLR42(Nenninger,A.A.et al.,Mol.Microbiol.,2011,81,486-499)和pMC2(Robinson,L.S.et al.,Mol.Microbiol.2006,59,870-881)(或衍生的螺旋物2突变体),所述琼脂板含有100mg l^{-1}氨苄青霉素、25mg l^{-1}氯霉素、1mM含有或不含0.2%(w/v)L-阿拉伯糖的IPTG。在37℃下培养过夜后,检查菌落生长。

[0964] 实施例14:单通道电流记录

[0965] 采用带有来自Nanion的Orbit 16试剂盒的并行高分辨率电记录进行单通道电流记录。简言之,在16通道多电极空腔阵列(MECA)芯片(lonera)(Del Rio Martinez,J.M.,ACS NanoSmall,2014,5,8080-8088)中的(亚皮升体积的)微腔上形成1,2-二植烷酰基-sn-甘油基-3-磷酸胆碱的水平双层(Avanti Polar Lipids)。双层上方和下方的顺式和反式腔均含有1.0M KCl、25mM pH值为8.0的Tris-HCl。为了将通道插入膜中,将溶解在25mM pH值为8.0的Tris-HCl、500mM NaCl、1mM DTT、0.5%C8E4、5mM LDAO中的CsgG加入到顺式腔中直至最终浓度为90-300nM。为了测试CsgG通道与CsgE的相互作用,将溶解在25mM pH值为8.0的Tris-HCl、150mM NaCl中的后一种蛋白的溶液加入到顺式腔中直至最终浓度为0.1、1、10和100nM。采用Tecella Triton 16通道放大器,在保持电势为+50mV和-50mV(顺式侧接地)情况下,记录跨膜电流,其中,所述16通道放大器的低通滤波频率为3kHz且采样频率为10kHz。使用pClamp套件(分子装置)的Clampfit来分析电流迹线。使用Origin 8.6(Microcal)(Movileanu,L.,Nature Biotechnol.,2000,18,1091-1095)生成图。

[0966] 基于CsgG X射线结构的孔尺寸,将测量电流与计算出的电流进行比较,模拟由三段组成:跨膜段、周质前庭和连接两者的内通道收缩。将前两个段建模为圆锥形,而收缩表示为圆柱体。相应的电阻R_1,R_2和R_3分别计算为

[0967] $R_1=L_1/(\pi D_1 d_1 k)$

[0968] $R_2=L_2/(\pi D_2 d_2 k)$

[0969] $R_3=L_3/(\pi d_1 d_2 k)$

[0970] 其中,L_1、L_2和L_3为各段的轴向长度,分别测量为3.5、4.0和2.0nm,并且D_1、d_1、D_2和d_2为段1和段2的最大和最小直径,分别测量为4.0、0.8、3.4和0.8nm。电导率k的值为10.6Sm^{-1}。通过将R_1、R_2和R_3以及电压V=50mV插入到下式来计算电流:

[0971] $I=V/(R_1+R_2+R_3)$

[0972] 未发现接入电阻显著改变预测电流。

[0973] 可以在存在分析物的情况下,进行如上所述的单通道电流记录,从而使得通道起到生物传感器的作用。

[0974] 实施例15:采用模型聚丙氨酸链对CsgG收缩结构进行分子动力学模拟

[0975] 利用聚丙氨酸链穿过扩展构象的通道对CsgG收缩结构进行建模,如图9所示,在C端至N端方向。穿过CsgG转运体的底物通道本身不是序列特异性的(Nenninger,A.A.et al.,Mol.Microbiol.,2011,81,486-499;Van Gerven,N.et al.,Mol.Microbiol.2014,91,1022-1035 2014)。为了清楚起见,使用聚丙氨酸链来对穿过多肽链的推定相互作用进行建

模。建模区域由9个同心式CsgG C-环组成,每个包括残基47-58。沿收缩结构排列成行的侧链在图中示为棒状,其中标记出Asn 55和Phe 56。收缩结构内的10Å的聚丙氨酸残基(残基标记为+1至+5)内的溶剂分子(水)显示为点。图9c示出了如图9b所示的聚丙氨酸链的建模溶剂化作用,为清楚起见去除了C-环(所示的溶剂分子是10Å的完整聚丙氨酸链内的)。在Asn 55和Phe 56的环的高度处,聚丙氨酸链的溶剂化作用减少到单个水壳,所述水壳将肽骨架和酰胺侧链桥接。与CsgG(和$CsgG_{C1S}$)X射线结构中观察到的侧链的内部、中心指向位置相比,Tyr 51环中的大多数侧链朝向溶剂转动。该模型是采用GROMACS(Pronk,S.et al.,Bioinformatics,2013,29,845-854)和残基的Ca原子对40ns全原子显示溶剂进行分子动力学模拟的结果,其中,所述GROMACS利用AMBER99SB-ILDN(Lindorff-Larsen,K.et al.,Proteins 2010,78,1950-1958)力场,所述残基的Ca原子位于位置受限的C环末端(Gln 47和Thr 58)。

[0976] 实施例16:CsgG纳米孔用于核酸测序的用途

[0977] Phi29DNA聚合酶(DNAP)可以与位于膜内的野生型或突变型CsgG纳米孔一起用作分子马达,以允许寡聚探针DNA链通过所述孔的可控运动。可以跨所述孔施加电压,并且由所述纳米孔任一侧的盐溶液中的离子的运动产生电流。随着探针DNA移动通过所述孔,流经所述孔的离子流相对于所述DNA变化。该信息已显示为序列依赖性的,并允许从电流测量值,例如上述实施例14中所述的测量值,精确地读取探针的序列。

[0978] 实施例17

[0979] 本实施例描述了为研究CsgG中的DNA行为而进行的模拟。

[0980] 材料与方法

[0981] 实施拉伸分子动力学模拟以研究CsgG-Eco和各种突变体对DNA移位的能垒大小。所述模拟使用4.0.5版GROMACS软件包,并采用GROMOS 53a6力场和SPC水模型实施。从蛋白质数据库中获取CsgG-Eco(SEQ ID NO:390)的结构,登录号为4UV3。为了制作CsgG-Eco突变体的模型,其野生型蛋白质结构使用PyMOL进行突变。研究的突变体为CsgG-Eco-

[0982] (F56A)(带有突变F56A的SEQ ID NO:390),CsgG-Eco-(F56A-N55S)(带有突变F56A/N55S的SEQ ID NO:390)以及CsgG-Eco-(F56A-N55S-Y51A)(带有突变F56A/N55S/Y51A的SEQ ID NO:390)。

[0983] 然后DNA被放入所述孔中。建立两种不同的系统:

[0984] i.将单个鸟嘌呤核苷酸放置在所述孔内,正好在收缩结构区域的上方。

[0985] (所述残基56环上方大约5-10埃)

[0986] ii.单链DNA(ssDNA)沿所述孔的轴线放置,并且其5'端朝向所述孔的β-桶结构侧。在上述建立中,所述ssDNA预穿过所述孔的整个长度。

[0987] 然后模拟盒被溶剂化,并使用最速下降算法将能量最小化。

[0988] 每个系统都在NPT系统中模拟,使用Berendsen恒温算法和Berendsen恒压算法至300K。在整个模拟过程中,所述孔的骨架受到限制。

[0989] 为了将DNA拉过所述孔,在单个鸟嘌呤模拟中对其磷原子施加拉力。在所述ssDNA模拟中,所述拉力施加到所述链5'端的磷原子上。通过连接上述DNA磷原子和平行于孔轴线以恒定速率行进的虚拟点之间的弹性元件,以恒定速率施加所述拉力。需注意的是,所述弹性元件没有任何形状,也不会发生任何流体动力学阻力。弹性常数等于$5\ kJmol^{-1}Å^{-2}$。

[0990] 实验结果

[0991] 单个G的移位

[0992] 如图12所示,所述拉力与时间的图表明,核苷酸进入所述野生型CsgG-Eco孔中的苯丙氨酸残基F56的环的能垒较大。对于所研究的CsgG-Eco突变体,未观察到鸟嘌呤移位存在显著能垒。

[0993] ssDNA移位

[0994] 对于ssDNA移位,每个孔进行两次模拟,每次模拟施加不同的拉力速度(100Å/ns和10Å/ns)。如图13所示,其表示更快的拉力速率模拟,所述CsgG野生型孔需要最大的拉力以使ssDNA移位。如图14所示,其表示较慢的拉力速率模拟,CsgG-Eco(野生型,SEQ ID NO:390)和CsgG-Eco-(F56A)孔都需要施加最大的力来实现ssDNA移位。通过比较用于ssDNA移位穿过CsgG孔和MspA基线孔所需的拉力,表明需要CsgG孔的突变以允许相似水平的ssDNA移位。

[0995] 实施例18

[0996] 本实施例描述对多种CsgG突变体的表征。

[0997] 材料与方法

[0998] 在建立实验之前,DNA构建体X(终浓度为0.1nM,参见用于图像显示构建体X的图22和说明书)与T4Dda-E94C/C109A/C136A/A360C(带有突变E94C/C109A/C136A/A360C的SEQ ID NO:412,加入到纳米孔体系的终浓度为10nM,其在缓冲液(151.5mM KCl,25mM磷酸钾,5%甘油,pH 7.0,1mM EDTA)中提供)一起在室温下预培养5分钟。5分钟后,将TMAD(100μM)加入到所述预混合物中,并将所述混合物再培养5分钟。最后,将MgCl2(加入到所述纳米孔体系中的终浓度为1.5mM),ATP(加入到所述纳米孔体系中的终浓度为1.5mM),KCl(加入到所述纳米孔体系中的终浓度为500mM)和磷酸钾缓冲液(加入到所述纳米孔体系中的终浓度为25mM)加入到所述预混合物中。

[0999] 从插入到缓冲液(25mM K磷酸盐缓冲液,150mM亚铁(II)氰化钾,150mM铁(III)氰化钾,pH8.0)的嵌段共聚物中的各种单个CsgG纳米孔获得电学测量结果。在实现将单个孔插入所述嵌段共聚物中后,缓冲液(2mL,25mM K磷酸盐缓冲液,150mM亚铁(II)氰化钾,150mM铁(III)氰化钾,pH 8.0)流过所述系统以除去任何过量CsgG纳米孔。150uL的500mM KCl,25mM K磷酸盐,1.5mM MgCl2,1.5mM ATP,pH 8.0的溶液随后流过所述系统。10分钟后,150uL的500mM KCl,25mM K磷酸钾,1.5mM MgCl2,1.5mM ATP,pH8.0的溶液流过所述系统,并且然后酶(T4Dda-E94C/C109A/C136A/A360C,终浓度为10nM),DNA构建体X(终浓度为0.1nM),养料(MgCl2终浓度1.5mM,ATP终浓度1.5mM)的预混液(共150uL)流入所述单个纳米孔实验系统。所述实验在-120mV和监测解旋酶控制的DNA移动下进行。

[1000] 实验结果

[1001] 孔显示增加的范围(图15至17以及图27至39)

[1002] CsgG-Eco-(Strepll(C))(SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在其C末端)具有~10pA的范围(参见图15(a)),而以下CsgG-Eco孔突变体显示增加的电流范围-

[1003] 1-CsgG-Eco-(Y51N-F56A-D149N-E185R-E201N-E203N-Strepll(C))9(具有突变Y51N/F56A/D149N/E185R/E201N/E203N的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端)显示为~30pA的范围(参见图15(b))。

[1004] 2-CsgG-Eco-(N55A-Strepll(C))9(具有突变N55A的SEQ ID NO:390,其中Strepll(C)具有SEQ ID NO:435并且连接在C末端)显示为~35pA的范围(参见图15(c))。

[1005] 3-CsgG-Eco-(N55S-Strepll(C))9(具有突变N55S的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~40pA的范围(参见图16(a))。

[1006] 4-CsgG-Eco-(Y51N-Strepll(C))9(具有突变Y51N的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~40pA的范围(参见图16(b))。

[1007] 5-CsgG-Eco-(Y51A-F56A-Strepll(C))9(具有突变Y51A/F56A的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~30pA的范围(参见图16(c))。

[1008] 6-CsgG-Eco-(Y51A-F56N-Strepll(C))9(具有突变Y51A/F56N的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~20pA的范围(参见图17(a))。

[1009] 7-CsgG-Eco-(Y51A-N55S-F56A-Strepll(C))9(具有突变Y51A/N55S/F56A的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~30pA的范围(参见图17(b))。

[1010] 8-CsgG-Eco-(Y51A-N55S-F56N-Strepll(C))9(具有突变Y51A/N55S/F56N的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~30pA的范围(参见图17(c))。

[1011] 13-CsgG-Eco-(F56H-Strepll(C))9(具有突变F56H的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~35pA的范围(参见图27)。

[1012] 14-CsgG-Eco-(F56Q-Strepll(C))9(具有突变F56Q的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在其C末端)显示为~40pA的范围(参见图28)。

[1013] 15-CsgG-Eco-(F56T-Strepll(C))9(具有突变F56T的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~35pA的范围(参见图29)。

[1014] 16-CsgG-Eco-(S54P/F56A-Strepll(C))9(具有突变S54P/F56A的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~35pA的范围(参见图30)。

[1015] 17-CsgG-Eco-(Y51T/F56A-Strepll(C))9(具有突变Y51T/F56A的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~30pA的范围(参见图31)。

[1016] 18-CsgG-Eco-(F56P-Strepll(C))9(具有突变F56P的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~30pA的范围(参见图32)。

[1017] 19-CsgG-Eco-(F56A-Strepll(C))9(具有突变F56A的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~40pA的范围(参见图33)。

[1018] 20-CsgG-Eco-(Y51T/F56Q-Strepll(C))9(具有突变Y51T/F56Q的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~30pA的范围(参见图34)。

[1019] 21-CsgG-Eco-(N55S/F56Q-Strepll(C))9(具有突变N55S/F56Q的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~35pA的范围(参见图35)。

[1020] 22-CsgG-Eco-(Y51T/N55S/F56Q-Strepll(C))9(具有突变Y51T/N55S/F56Q的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~35pA的范围(参见图36)。

[1021] 23-CsgG-Eco-(F56Q/N102R-Strepll(C))9(具有突变F56Q/N102R的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端)显示为~30pA的范围(参见图37)。

[1022] 24-CsgG-Eco-(Y51Q/F56Q-Strepll(C))9(具有突变Y51Q/F56Q的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端)显示为~40pA的范围(参见图38)。

[1023] 25-CsgG-Eco-(Y51A/F56Q-Strepll(C))9(具有突变Y51A/F56Q的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端)显示为~35pA的范围(参见图39)。

[1024] 显示增加的通量的孔(图18和19)

[1025] 从图18和19可以看出,以下突变孔(以下9-12)在4小时内显示出每个通道多个解旋酶控制的DNA移动(图18和19中标记为X),而如图18(a)所示的CsgG-Eco-(Strepll(C))(SEQ ID NO:390,其中StepII(C)为SEQ ID NO:435并且连接在C末端),通常在4小时内每个通道仅显示1或2个解旋酶控制的DNA运动(在图18(a)中标记为X),并相反地,显示延长的阻挡区域(在图18(a)中标记为Y)。

[1026] 9-CsgG-Eco-(D149N-E185N-E203N-Strepll(C))9(具有突变D149N/E185N/E203N的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端)(参见图18(b))。

[1027] 10-CsgG-Eco-(D149N-E185N-E201N-E203N-Strepll(C))9(具有突变

[1028] D149N/E185N/E201N/E203N的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端)(参见图18(c))。

[1029] 11-CsgG-Eco-(D149N-E185R-D195N-E201N-E203N)-Strepll(C))9(具有突变D149N/E185R/D195N/E201N/E203N的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端)(参见图19(a))。

[1030] 11-CsgG-Eco-(D149N-E185R-D195N-E201R-E203N)-Strepll(C))9(具有突变D149N/E185R/D195N/E201R/E203N的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端)(参见图19(b))。

[1031] 显示增加的插入的孔(图20和21)

[1032] 通过对比图20和21可以看出,与CsgG-Eco-(Strepll(C))(SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端)孔相比(如图20所示),如图21中所示的突变孔CsgG-Eco-(T150l-Strepll(C))9(具有突变T150l的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端),以增加的孔数(~4-5倍)存在于膜中。图20和21中的箭头表示在4小时的实验中插入到嵌段共聚物中的CsgG-Eco纳米孔的数目(图20中的130-140以及图21中的1-11,各自对应于单独的纳米孔实验)。对于CsgG-Eco-(Strepll(C))(SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端),三次实验显示插入一个纳米孔,而对于突变孔(CsgG-Eco-(T150l-Strepll(C))9),每次实验显示插入至少一个纳米孔并且多次实验显示多个孔的插入。

[1033] 实施例19

[1034] 本实施例描述发展用于纯化CsgG孔的大肠杆菌纯化方法。

[1035] 材料与方法

[1036] 编码多肽Pro-CsgG-Eco-(Strepll(C)(SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并连接在C末端,并且其中Pro为SEQ ID NO:436并连接在N末端)的DNA在GenScript USA Inc.中合成并被克隆到含有氨苄青霉素抗性基因的pT7载体中。所述pT7载体的蛋白质表达通过异丙基β-D-1-硫代吡喃半乳糖苷(lPTG)诱导。所述DNA溶液的浓度调节至400ng/uL。使用DNA(1μl)来转化Lemo21(DE3)感受态(competent)大肠杆菌细胞(50μl,NEB,目录号

C2528H)。在转化前,CsgG基因从Lemo21(DE3)细胞(Gene Bridges GmbH,德国)中敲除。然后所述细胞接种在含有氨苄青霉素(0.1mg/mL)的LB琼脂上,并在37℃下培养约16小时。

[1037] 在含有氨苄青霉素的LB平板上生长的菌落含有CsgG质粒。使用一个这样的菌落来接种含有氨苄青霉素(0.1mg/mL)的LB培养基(100mL)上的起始培养物。所述起始培养物在37℃,搅拌下生长直到OD600达到1.0-1.2。所述起始培养物用于接种到新鲜的500mL的含有氨苄青霉素(0.1mg/mL)和鼠李糖(500μM)的LB培养基上,以达到0.1的O.D.600。所述起始培养物在37℃,搅拌下生长直到OD600达到0.6。然后将所述培养物的温度调至18℃,并通过添加IPTG(终浓度为0.2mM)引发诱导。诱导在18℃搅拌下进行约18小时。

[1038] 诱导后,所述培养物通过在6000g下离心30分钟制成小球。所述小球在含有蛋白酶抑制剂(Merck Millipore 539138),苯甲酸酶核酸酶(Sigma E1014)和1X抽提液(Merck Millipore70921)pH8.0的50mM Tris,300mM NaCl溶液(每克小球约10ml缓冲液)中重悬。所述悬液充分混合直到其完全均匀,然后样品在4℃下转移到滚筒混合器中约5小时。通过在20000g下离心45分钟使细胞裂解物制成小球,上清液通过0.22μM PES注射过滤器进行过滤。含有CsgG的上清液(称为样品1),进一步通过柱层析提纯。

[1039] 样品1上样到5ml Strep Trap柱(GE Healthcare)。所述柱用25mM Tris,150mM NaCl,2mM EDTA,0.01%DDM pH8的溶液洗涤,直到维持10个柱体积的稳定基线。然后所述柱在回到150mM缓冲液之前,用25mM Tris,2M NaCl,2mM EDTA,0.01%DDM pH8的溶液洗涤。洗脱采用10mM脱硫生物素进行。图23中显示CsgG蛋白的Strep trap(GE Healthcare)纯化的色谱痕迹的例子。洗脱峰用E1标记。图24示出在初始Strep纯化后,CsgG-Eco蛋白的典型SDS-PAGE可视化的示例。泳道1-3显示了如箭头所示的含有CsgG蛋白的主要洗脱峰(图23中标记为E1)。泳道4-6对应于含有污染物的主要洗脱峰(图23中标记为E1)的尾部的洗脱馏分。

[1040] 合并所述洗脱峰,并加热至65℃15分钟以除去热不稳定的杂质蛋白。所述加热的溶液在20000g下离心10分钟,并弃去所述小球。上清液在120mL Sephadex S200柱(GE Healthcare)上通过25mM Tris,150mM NaCl,2mM EDTA,0.01%DDM,0.1%SDS pH8溶液进行凝胶过滤。由于蛋白质的色氨酸组分低,监测在220nM下进行。样品在约55mL体积处洗脱出来(图25显示尺寸排阻柱的痕迹并用星号标记在55mL处的样品峰)。洗脱峰在4-20%TGX(参见图26,Bio Rad)上进行电泳,以证实感兴趣的CsgG-Eco-Strepll(C))(SEQ ID NO:390,其中Strepll为SEQ ID NO:435并且连接在C末端)的孔的存在。合并鉴定出的馏分,并用50kD的Amicon自旋柱进行浓缩。

[1041] 实施例20

[1042] 本实施例描述用来研究CsgG-Eco-(Y51T/F56Q-Strepll(C))9(具有突变Y51T/F56Q的SEQ ID NO:390,其中Strepll(C)为SEQ ID NO:435并且连接在C末端孔突变体No.20)与T4Dda-(E94C/C109A/C136A/A360C)(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:412并且然后(ΔM1)G1G2)之间相互作用的模拟。

[1043] 模拟方法

[1044] 所述模拟使用4.0.5版GROMACS软件包,并采用GROMOS 53a6力场和SPC水模型进行。

[1045] CsgG-Eco-(Y51T/F56Q-Strepll(C))9(具有突变Y51T/F56Q的SEQ ID NO:390,其

中Stepll(C)为SEQ ID NO:435并且连接在C末端孔突变体No.20)的模型是基于蛋白质数据库中发现的CsgG的晶体结构,登录号为4UV3和4Q79。相关的突变采用PyMOL制作。然后所得到的孔模型使用最速下降算法将能量最小化。T4Dda-(E94C/C109A/C136A/A360C)(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:412以及然后的(ΔM1)G1G2)的模型是基于蛋白质数据库中发现的Dda1993结构,登录号为3UPU。再次,相关的突变采用PyMOL制作,并使用最速下降算法将能量最小化。

[1046] T4Dda-(E94C/C109A/C136A/A360C)(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:412以及然后的(ΔM1)G1G2)的模型然后放置于CsgG-Eco-(Y51T/F56Q-Strepll(C))9(具有突变Y51T/F56Q的SEQ ID NO:390,其中Stepll(C)为SEQ ID NO:435并且连接在C末端孔突变体No.20)的上方。使用不同的初始酶构象进行三次模拟(试验1至3(0ns),参见图40):

[1047] 在所有酶构象中,酶被取向使得DNA的5'末端指向孔,并且在整个模拟过程中酶是无限定的。孔骨架受到限制并且模拟盒被溶剂化。所述系统在NPT系统中模拟40ns,使用Berendsen恒温算法和Berendsen恒压算法至300K。

[1048] 酶与孔之间的接触使用GROMACS分析软件以及本地书写的代码进行分析。下面的表格显示了观察到的孔和酶氨基酸的接触次数。表6-8显示与酶上的氨基酸接触位点相互作用的孔上的氨基酸接触位点。在三次模拟中的两次模拟中,酶在孔的顶部倾斜(参见试验2和3(20,30和40ns),图40和41)。试验1显示酶没有倾斜,因此可以优化表6中显示具有高相互作用的位点,以增加孔帽上的酶稳定性。

[1049]

孔	酶	# 接触
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

[1050]

孔	酶	# 接触
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
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

[1051]

孔	酶	# 接触
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
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

[1052]

孔	酶	# 接触
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

[1053] 表6＝试验1酶和孔接触相互作用

[1054]

孔	酶	# 接触
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
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

[1055]

孔	酶	# 接触
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
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

[1056]

孔	酶	# 接触
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
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

[1057]

孔	酶	# 接触
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

[1058] 表7＝试验2酶和孔接触相互作用

[1059]

孔	酶	# 接触
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
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

[1060]

孔	酶	# 接触
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
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

[1061]

孔	酶	# 接触
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
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

[1062]

孔	酶	# 接触
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
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

[1063]

孔	酶	# 接触
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

[1064] 表8＝试验3酶和孔接触相互作用

[1065] 实施例21:CsgG纳米孔在通道收缩结构中对捕获核酸的能力

[1066] 纳米孔用于核酸测序的用途要求通过纳米孔捕获和穿过单链DNA。在本实施例中,CsgG WT蛋白的单通道电流迹线伴随着携带单链DNA突出结构的DNA发夹的存在。在图56中表示的所述实施例的迹线显示对应于在+50mV或-50mV间隔(如箭头所示)测量的电势而改变的电流。在最后+50mV段的下行电流的阻滞表明将单链发夹结构末端穿入内部的孔结构,所述内部的孔结构导致几乎完全的电流阻滞。将电场反转至-50mV导致孔的电泳解除阻塞。新的+50mV时段再次导致DNA发夹结构的结合以及孔的堵塞。在所述+50mV段上,发夹结构的展开可以导致所述电流阻滞的停止,如所述电流阻滞的反转所示。具有序列为:3'

[1067] GCGGGGAGCGTATTAGAGTTGGATCGGATGCAGCTGGCTACTGACGTCATGACGTCAGTAGCCAGCATGCATCCGATC-5'(SEQ ID NO:441)的发夹结构在10nM的终浓度下被加入到所述腔室的顺式侧。

[1068] 实施例22本实施例描述具有改变的内部收缩结构的突变CsgG孔的产生。

[1069] 在本实施例中,证明了CsgG-ΔPYPA的稳定性和通道性质,所述CsgG-ΔPYPA为一种突变CsgG孔,其中收缩结构环中的序列PYPA(残基50-53)被GG取代。对于该突变体,位置38至63的收缩结构基序对应于SEQ ID NO:354。这种突变预见到从所述孔的收缩结构中除去Y51以及收缩结构环的缩短,以便降低在所述孔读取头中的复杂性,即所述孔的最窄部分,在感测应用中在此处测量电导率,对结合或穿过所述孔的底物的性质最为敏感。所述PYPA序列的替换保留所述CsgG九聚体的稳定性,并导致具有如图57所示增加的电导率的孔。

[1070] 利用来自Nanion(德国慕尼黑)的Orbit 16试剂盒,使用平行高分辨率电子记录,所述CsgG-ΔPYPA孔突变体通过单通道电流记录进行分析。简而言之,在16-通道多电极

空腔阵列(MECA)芯片(lonera,弗莱堡,德国)51中的(亚皮秒体积的)微腔上形成1,2-二植烷酰基-sn-甘油-3-磷酸胆碱(Avanti Polar Lipids)的水平双分子层。所述双分子层上方和下方的顺式和反式腔室均含有1.0M KCl,25mM Tris-HCl,pH8.0。为了将通道插入所述膜中,将溶解在25mM Tris pH 8.0,500mM NaCl,1mM DTT,0.5%C8E4,5mM LDAO中的CsgG加入到顺式腔室中至终浓度为100nM。使用Tecella Triton 16通道放大器,在3kHz的低通滤波频率和10kHz的采样频率下,在+50mV和-50mV(顺式侧接地)的保持电势记录H跨膜电流。电流迹线使用pClamp组件的Clampfit(Molecular Devices,USA)进行分析。

[1071] 实施例23:

[1072] 本实施例描述对细菌淀粉样蛋白分泌通道CsgG结构和机理的深入观察。

[1073] Curli是功能性淀粉样蛋白纤维,其构成由拟杆菌和变形杆菌形成的防护生物膜中的细胞外基质的主要蛋白质组分(主要是a类和c类)[1-3]。它们提供致病菌株中的适应性优势,并在菌血症期间诱导强烈的促炎症反应[1,4,5]。Curli的形成需要专门的蛋白质分泌机制,所述机制包括外膜脂蛋白CsgG和两种可溶性辅助蛋白,CsgE和CsgF[6,7]。我们在这里报道了大肠杆菌CsgG非脂化,可溶形式以及天然膜提取构象的X射线结构。CsgG形成由9个反密码子-结合-结构域样单元组成的寡聚转运复合体,其产生横贯所述双分子层的β-桶结构并连接到周质中的笼状前庭(vestibule)。其跨膜和周质结构域被由三个堆叠的同心苯丙氨酸,天冬酰胺和酪氨酸环组成的0.9nm通道收缩结构分开,所述环可以引导伸展的多肽底物通过所述分泌孔。特异性因子CagE形成结合和封闭分泌通道的周质面的九聚体接合体,产生一个24000Å3的预收缩结构腔室。我们的结构,功能和电生理学分析意味着CsgG是一种非门态的,非选择性的蛋白分泌通道,预期采用基于扩散的熵驱动的转运机制。

[1074] Curli是细菌表面附属物,其具有淀粉样蛋白纤维的结构和生理特征,该蛋白纤维在人类退行性疾病中最为人所知[7-9]。然而,细菌淀粉样蛋白例如Curli的作用是促进生物膜的形成[4,10]。与作为蛋白错误折叠产物的致病性淀粉样蛋白不同,Curli的形成由在两个专用操纵子,csgBAC(curli特异性基因BAC)和大肠杆菌中的csgDEFG,中编码的蛋白调控(图46)[6,7]。分泌后,CsgB使CsgA亚基成核形成curli纤维[7,11,12]。CsgA和CsgB的分泌和细胞外沉积依赖于两种可溶性辅助因子,分别为CsgE和CsgF,以及位于外膜中的262-残基的脂蛋白CsgG[13-16]。由于在外膜上缺少水解能源或离子梯度,因此CsgG属于一种特殊类型的蛋白质移位体,其必须通过交替供能的运输机制进行操作。在没有结构模型的情况下,CsgG如何以跨生物膜的调节方式促进高度稳定的淀粉样蛋白纤维的分泌和组装的动力学工作方式迄今仍然是未知的。

[1075] 在插入所述外膜之前,脂蛋白通过脂蛋白定位(Lol)途径穿过周质[17]。我们观察到,非脂化CsgG($CsgG_{C1s}$)可以分离作为可溶性周质中间体,类似于在成孔蛋白和毒素中观察到的前孔形态[18]。除少量的离散低聚复合物之外,CsgGc1s主要作为单体被发现(图47)[19]。所述可溶的$CsgG_{C1S}$寡聚体进行结晶,并且其结构确定至2.8Å,揭示出具有八倍二面体对称性(D8)的十六聚体颗粒,该颗粒由两个以尾-尾相互作用连接的环形八聚体复合物(C8)组成。$CsgG_{C1S}$原聚体表现出一个反密码子-结合-结构域(ABD)样的折叠,其在氨基和羧基末端用两个α-螺旋(分别为αN和αC;图42以及图48a-c)延伸。附加的CsgG特异性元件是连接β1和α1的延伸环,连接β3-β4和β5-α3的环上的两个插入结构以及一个延伸的α2螺旋,所述螺旋通过在相邻单体间堆积而参与CSgG寡聚过程(图42b)。在β3-β5片体的背面和延伸的β1-α1环

之间形成原聚体的进一步接触(图48d,e)。

[1076] 在$CsgG_{C1S}$结构中,将α1连接到脂质锚蛋白N-末端的残基1-17是无序的,并且没有明显的跨膜(TM)结构域可以被辨别(图42)。$CsgG_{C1S}$和天然的膜提取CsgG的衰减全反射傅里叶变换红外光谱(ATR-FTIR)显示后者在β折叠区域(1625-1630cm^{-1})具有更高的吸收,而无规则卷曲和α-螺旋区域(分别为1645-1650cm^{-1}和1656cm^{-1};图43a)则伴随吸收减少,表明膜相关的CsgG含有β-桶结构域。在连接β3-β4(残基134-154)和β5-α3(残基184-204)的无序的延伸环中发现延伸用于β-链形成的候选序列;这些序列的敲除导致无法形成Curli(图43b)。洗涤剂提取的CsgG的晶体结构证实两个区域的构象重排为两个相邻的β-发夹结构,延伸通过β3-β4(TM1)和β5-α3(TM2)形成的β-折叠(图43c)。它们在CsgG寡聚体中的并排放置产生复合36-链的β-桶结构(图43d)。而结晶的$CsgG_{C1S}$寡聚物显示D8对称性,CsgG结构显示出D9对称性,并且CsgG原聚体保留了相当的原聚体间的接触,除相对于中轴的5°旋转以及沿径向轴4Å的翻译(图47)。上述观察结果与通过单粒子电子显微镜发现的溶液中(in-solution)的寡聚状态相一致,其仅发现对于膜提取CsgG的C9和D9对称性(图47)。非脂化样品中单体的主要存在形式和与膜结合蛋白的对称性错配表明,在插入膜之前,CsgG以单体的,LOlA-结合的形式靶向外膜,并且所述C8和D8颗粒是高度浓缩的$CsgG_{C1S}$溶液的人工制品。此外,我们发现是C9九聚体而不是D9复合物形成生理学上相关的颗粒,因为在分离的大肠杆菌外膜中,通过观察到的尾-尾二聚化包含在残基中的半胱氨酸取代可利用马来酰亚胺聚乙二醇(PEG,5kDa;图49)进行标记。

[1077] 因此,CsgG形成宽度为120Å且高度为85Å的九聚转运复合物。所述复合物穿过外部膜通过内径为40Å的36-链的β-桶结构(图43e)。通过围绕(wraps over)在相邻原聚体的18-残基连接器,所述N-末端的脂质锚蛋白从所述核心结构中分离出来(图48d)。N-末端Cys上的二酰基甘油和酰胺连接的酰基链未在电子密度图中解析出来,但是基于Leu 2的位置,预计脂质锚蛋白位于所述β-桶结构外壁的侧面。在所述周质侧,所述转运体形成内径为35Å,高度为40Å的溶剂可进入的大空腔,其向所述周质通过螺旋2形成50Å的开口(图43e)。在其最高点,该周质前庭(vestibule)通过连接β1至α1(C-环;图43e和44a,b)的保守性12-残基环从TM通道中分离,其将分泌管道收缩至排除溶剂的直径为9.0 Å(图44a,c)。这些孔的尺寸与一个或两个(例如环状结构)延伸的多肽片段相适合,其中五个残基跨过收缩结构的高度(图50)。收缩结构的管腔衬里层由残基Tyr 51,Asn 55和Phe 56的侧链形成的三个堆叠的同心环组成(图44a,b)。在炭疽PA63毒素中,拓扑等效的同心Phe环(称为Φ夹)排列于移位通道的入口并催化多肽的捕获和通过[20-22]。CsgG样转运体的多重序列比对显示Phe 56的绝对保守性和由Asn 55到Ser或Thr的保守突变(图51)。Phe 56或Asn 55到Ala的突变会导致Curli产量的近似损失(图44d),而Asn 55→Ser取代保留了野生型的分泌水平,上述现象一起揭示出对Φ-夹的堆叠结构的需要,然后是在CsgG收缩结构中的氢键供体/受体(图44b和图51)。

[1078] 在磷脂双分子层平面中重组的CsgG的单通道电流记录使用标准电解质条件和分别为+50mV或-50mV的电势,导致43.1±4.5pA(n=33)或-45.1±4.0pA(n=13)的稳定电流(图44e,f以及图52)。所述观察到的电流与基于简单三-片段孔模型计算的46.6pA的预测值和X射线结构中观察到的中心收缩结构的尺寸非常一致(图44c)。在负电场电势(-26.2±3.6pA(n=13);图52)条件下还可以观察到第二个低电导构象(图52).然而,这种构象是否

存在于生理条件下并不清楚。

[1079] 我们的结构数据和单通道记录表明CsgG形成了非门控的肽扩散通道。在PA_{63},即模型肽扩散通道中,多肽的通过取决于ΔρH驱动的布朗棘轮模型,其对移位通道中的扩散步骤进行校正[20-22]。然而,这种质子梯度并不存在于外膜上,需要替代的驱动力。而在升高的浓度下,CsgG促进周质多肽的非选择性扩散性渗出,在天然条件下对CsgA而言分泌是特异性的,并且需要周质因子CsgE[16,23]。在过量的CagE存在下,纯化的CsgG在天然性PAGE上形成更缓慢迁移物种(图45a)。SDS-PAGE分析显示,该新物种由以1:1化学计量比存在的CsgG-CagE复合物组成(图45b)。通过下拉亲和纯化分离的CsgG-CsgE的冷冻电子显微镜(cryo-EM)可视化发现了对应于CsgG九聚体的九倍对称颗粒,和在周质门庭的入口处的另外的加盖密度,其尺寸和形状类似于同样通过单粒子EM和尺寸排阻色谱观察到的C9CsgE寡聚物(图45c-e以及图53)。观察到的CsgG-CagE接触界面的位置被CsgG螺旋2中的阻塞点突变所证实。与加帽功能一致,单通道记录显示CagE与所述移位体结合导致其离子电导的特异性沉默(图45f以及图52)。通道中CsgE的加帽似乎在CsgE九聚体结合的功能中是一个全有-或-全无的反应。在饱和条件下,CsgE的结合诱导通道的完全阻断,而在约10nM条件下,CsgE结合与解离事件之间的平衡导致间歇性阻断的或完全开放的移位体。在1nM或以下的条件下,瞬时(<1ms)局部阻塞事件可能源于短期与单体CsgE的遭遇。

[1080] 因此,CsgG和CsgE似乎形成一个包含~24000Å^{3}的中心腔的笼状复合物,使其联想到在GroEL伴侣蛋白和GroES共-伴侣蛋白中观察到的底物结合腔和包封盖结构[24]。所述CsgG-CagE的外壳与129-残基CsgA的全部或部分捕获相适应。最近在ABC毒素中观察到移位底物的封闭[25]。展开的多肽的空间限制导致其构象空间的减少,产生在伴侣蛋白情况下显示为有利于多肽折叠的熵电位[24,26]。类似地,我们推测在curli转运中,CsgG前庭中Curli亚基的局部高浓度和构象限制会在移位通道上产生熵自由能梯度(图45g)。在捕获进入收缩结构期间,所述多肽链然后预期通过布朗扩散逐渐向外移动,通过从所述分泌通道附近的CsgE介导的限制和/或底物捕获产生的熵电位进行校正。为了预收缩结构腔内的完全限制,展开的129-残基多肽逃逸到大量溶剂中将相当于释放高达~80kcal mol^{-1}(约340kJ mol^{-1};参考文献27)的熵自由能。底物对接和限制的初始熵消耗可能至少部分地通过在CsgG-CsgE-CsgA复合物组装过程中释放的结合能以及在周质中已经降低的CsgA熵进行补偿。在理论上,可以设想三种可能的CsgA补充到分泌复合物的途径(图54)。

[1081] Curli诱导的生物膜在致病性肠杆菌中形成适合因子和毒力因子[4,5]。其独特的分泌和组装性能也迅速获得对(生物)技术应用的兴趣[23,28,29]。我们对两个关键分泌成分的结构表征和生物化学研究提供了一种在不存在可水解能源,膜电势或离子梯度的情况下,展开的蛋白底物的膜移位的重复机制的试验模型(图45e和图54)。所提出的分泌模型中贡献因子的完全验证和解构将需要移位子的体外重组,以允许在单分子水平下精确地遵循转运动力学。

[1082] 方法

[1083] 克隆和菌株。已在参考文献19中描述了用于制备局部外膜的C-末端Strepll标记的CsgG(pPG1)和周质C-末端Strepll标记的$CsgG_{C1S}$(pPG2)的表达构建体。对于硒代甲硫氨酸标记,Strepll标记的$CsgG_{C1S}$由于增加的产量而在细胞质中进行表达。因此,pPG2改变为采用带有引物5'-TCTTTAACCGCCCCGCCTAA AG-3'(正向)和5'-CATTTTTTGCCCTCGTTATC-3'

(反向)(pPG3)的反向PCR去除N末端信号肽。对于表型测定,通过参考文献30中所描述的方法(带有引物5'-AATAACTCAACC GAT TTT TAA GCC CCA GCT TCA TAA GGA AAA TAA TCG TGT AGG CTG GAG CTG CTT C-3'和5'-CGC TTA AAC AGT AAA ATG CCG GAT GAT AAT TCC GGC TTT TTT ATC TGC ATA TGA ATA TCC TCC TTA G-3')构建大肠杆菌BW25141(大肠杆菌NVG2)的CsgG缺失突变体。在trc启动子控制下,通过始于pMC2,即含有CsgG的pTRC99a载体,的定点诱变(QuikChange方案;Stratagene)来构建各种CsgG替代型突变体,所述CsgG替代型突变体用于Cys可接近性试验并用于通道收缩的表型探测[14]。

[1084] 蛋白表达和纯化。如上所述,对CsgG和$CsgG_{C1S}$进行表达和纯化[19]。简言之,在大肠杆菌BL21(DE3)中重组地产生CsgG,其中,所述大肠杆菌采用pPG1转化,并使用缓冲液A(50mM Tris-HCl pH 8.0;500mM NaCl;1mM EDTA;1mM二硫苏糖醇(DTT))中的1%正十二烷基-b-D-麦芽糖苷(DDM)从分离的外膜中进行提取。将Strepll标记的CsgG装载到5毫升Strep-Tactin琼脂糖凝胶柱(lba GmbH)上,并通过用20个柱体积的补充有0.5%的四甘醇单辛基醚(C8E4;Affymetrix)和4mM月桂基二甲基氧化胺(LDAO:lauryldimethylamine-N-oxide;Affymetrix)的缓冲液A洗涤来进行清洁剂交换。通过添加2.5mM D-脱硫生物素将蛋白洗脱,并浓缩至5mg ml^{-1}用于结晶实验。对于硒代甲硫氨酸标记,在用pPG3转化的Met营养缺陷型菌株B834(DE3)中制备$CsgG_{C1S}$,并在补充有40mg l^{-1}L-硒代甲硫氨酸的M9基本培养基上生长。细胞颗粒重悬于补充有完整的罗氏蛋白酶抑制剂混合物(Protease Inhibitor Cocktail(Roche))的50mM Tris-HCl pH 8.0、150mM NaCl、1mM EDTA、5mM DTT中,并通过穿过以20×10^{3}lb in^{-2}操作的TS系列细胞破碎剂(Constant Systems Ltd)而分裂。如上所述,对标记的$CsgG_{C1S}$进行纯化[19]。在整个纯化过程中添加DTT(5mM)以避免硒代甲硫氨酸氧化。

[1085] 在携带有pNH27的大肠杆菌NEBC2566细胞中产生CsgE(参考文献16)。将25mM pH值为8.0的Tris-HCl、150mM NaCl、25mM咪唑、5%(v/v)甘油中的细胞溶解产物装载在HisTrap FF上(GE Healthcare)。CsgE-his用20mM Tris-HCl pH 8.0,150mM NaCl,5%(v/v)甘油缓冲液中的500mM咪唑进行线性梯度洗脱。含有CsgE的馏分用250mM$(NH_4)_2SO_4$增补,并施加到用20mM Tris-HCl pH8.0、100mM NaCl、250mM$(NH_4)_2SO_4$、5%(v/v)甘油平衡的5ml HiTrap Phenyl HP柱上(GE Healthcare)。向20mM Tris-HCl pH 8.0、10mM NaCl、5%(v/v)甘油施加线性梯度以便进行洗脱。将含有馏分的CsgE装载到Superose 6Prep Grade10/600(GE Healthcare)柱上,所述柱在20mM Tris-HCl pH 8.0、100mM NaCl、5%(v/v)甘油中平衡。

[1086] 溶液中寡聚状态评估。将各约0.5mg的洗涤剂溶解的CsgG(0.5%C8E4,4mM LDAO)和$CsgG_{C1S}$施加到用25mM Tris-HCl pH 8.0、500mM NaCl、1mM DTT、4mM LDAO和0.5%C8E4(CsgG)或25mM Tris-HCl pH 8.0、200mM NaCl($CsgG_{C1S}$)平衡的Superdex 200 10/300GL分析凝胶过滤柱(GE Healthcare)上,并以0.7ml min^{-1}运行。用牛甲状腺球蛋白、牛γ-球蛋白、鸡卵清蛋白、马肌球蛋白和维生素B_{12}(Bio-Rad)对柱洗脱体积进行校准(图47)。采用参考文献31中描述的程序(图47),将膜提取的CsgG,即20mg洗涤剂溶解的蛋白也在3-10%蓝色天然PAGE上运行。NativeMark(Life Technologies)未染色蛋白标准物(7ml)用于分子量估计。

[1087] 结晶、数据收集和结构测定。通过对包含有100mM pH值为4.2的乙酸钠、8%PEG 4000和100mM丙二酸钠pH 7.0的溶液进行坐滴式(sitting-drop)蒸气扩散,将硒代甲硫氨酸标记的$CsgG_{C1S}$浓缩至3.8mg ml^{-1}并进行结晶。在补充有15%甘油的晶体缓冲液中对晶体

进行培养,并在液氮中快速冷冻。通过对包含有100mM Tris-HCl pH8.0、8％PEG 4000、100mM NaCl和500mM $MgCl_2$的溶液进行悬滴式(hanging-drop)蒸气扩散,将洗涤剂溶解的CsgG浓缩至5mg ml^{-1}并进行结晶。将晶体在液氮中快速冷冻并通过存在于晶体溶液中的洗涤剂进行冷冻保护。为了优化结晶条件并筛选具有良好衍射质量的晶体,在光束线Proxima-1和Proxima-2a(Soleil,法国)、PX-1(瑞士光源,瑞士)、102、103、104和124(钻石光源,英国)以及ID14eh2、ID23eh1和ID23eh2(ESRF,法国)上对晶体进行分析。分别在光束线104和103上收集到用于$CsgG_{C1S}$和CsgG结构测定的最终衍射数据(数据收集和细化统计参见图55a)。采用Xia2和XDS[32,33]封装处理$CsgG_{C1S}$的衍射数据。$CsgG_{C1S}$的晶体属于空间群P1,晶胞大小为a=101.3 Å, b=103.6 Å, c=141.7 Å,α=111.3°,β=90.5°, γ =118.2°,不对称单元中含有16个蛋白复本。为了实现结构测定和细化,以0.9795 Å波长收集的数据基于最高分辨率层中2的//σ/截止在2.8 Å截断。采用从单一异常色散(SAD)实验计算的实验相来求解结构。通过使用ShelxC和ShelxD[34],在不对称单元中定位总共92个硒位点,进行细化并采用Sharp[35](相位功率0.79,品质因数(FOM)0.25)用于相位计算。通过非晶体学对称性(NCS)采用Parrot[36](图55;FOM 0.85)对实验相进行密度修正和平均。采用Buccaneer[37]构建初始模型,并采用Phenix细化[38]和手工检查以及在Coot[39]中(重建)构建的模型通过最大似然细化的迭代循环进行细化。最终结构含有28,853个原子,这些原子在属于16条$CsgG_{C1S}$链的3,700个残基中(图47),其中,molprobity[40]值为1.34;98％的残基位于拉氏图(允许区域为99.7％)的优选区域中。电子密度图未显示与可能的溶剂分子对应的不明确密度,因此未建立水分子或离子。在整个细化过程中,维持十六倍的NCS平均,分别在细化早期和晚期采用严格和局部NCS约束。

[1088] 在空间群C2中,采用XDS封装[32,33]从0.9763 Å波长的单晶中收集CsgG的衍射数据,并编入索引且进行缩放,其中,所述空间群的晶胞尺寸为a=161.7 Å, b=372.3 Å, c=161.8 Å和β=92.9°,包括非对称单元中的18个CsgG复本和72％的溶剂含量。沿互逆单元方向a*,b*和c*将用于结构测定和细化的衍射数据椭圆形地截取为3.6Å,3.7Å和3.8Å的分辨率极限,并用衍射各向异性服务器(Diffraction Anisotropy Server)[41]进行各向异性地缩放。采用$CsgG_{C1S}$单体的分子置换被证明不成功。对自旋函数的分析揭示了不对称单元中的D9对称性(未示出)。在$CsgG_{C1S}$结构基础上,假设从C_8去到C_9寡聚物之后,原聚体间的角度(interprotomer angle)从45°变化到40°,颗粒周围的原聚体间的弧度(interprotomer arc)保持大致相同,从而使半径计算的增加约为4 Å,生成九聚搜索模型。将所计算的九聚体用作搜索模型,用Phaser[42]找到包含两个复本的分子置换溶液。对进行了密度修正和NCS平均的电子密度图(Parrot[36];图55)的检查使得能够手工建立TM1和TM2并重塑蛋白核心中的相邻残基,以及构建残基2-18,其从$CsgG_{C1S}$模型中缺失并将α1螺旋物连接到N-末端脂质锚。分别使用用于初始和最终细化循环的Buster-TNT[43]和Refmac5(参考文献44)对CsgG模型进行细化。在整个细化过程中应用十八倍的局部NCS限制,并且采用具有由Prosmart[45]产生的σ0.01和氢键约束的凝胶体细化来运行Refmac5。最终结构包括4,451个残基中的34,165个原子,这些残基属于18个CsgG链(图47),其中molprobity值为2.79;93.0％的残基位于拉式图(允许区域为99.3％)的优选区域。不能识别出对应于N-末端脂质锚的不明确电子密度。

[1089] 刚果红试验。为了分析刚果红结合,在LB(Lysogeny Broth)培养基中对LB中在37

℃生长的细菌过夜培养物进行稀释,直至达到0.5的D_{600}。在补充有氨苄青霉素(100mg l^{-1})、刚果红(100mg l^{-1})和0.1%(w/v)异丙基β-D-硫代半乳糖苷(IPTG)的LB琼脂板上点样5μl样本。在室温(20-22℃)下将平板培养48小时以诱导卷曲菌毛表达。观察菌落形态发展和染色结合情况。

[1090] 半胱氨酸可接近性试验。采用定点突变在pMC2中产生半胱氨酸突变体,并在大肠杆菌LSR12中进行表达(参考文献7)。将生长过夜的细菌培养物点样到含有1mM IPTG和100mg l^{-1}氨苄青霉素的LB琼脂板上。将板在室温下培育,并在48小时后刮下细胞,重悬于1ml的PBS中并用D_{600}归一化。通过超声处理裂解细胞,并在4℃温度下以3,000g离心20秒,以便从细胞裂解液和悬浮膜中去除未破裂的细胞。将上清液中的蛋白在室温下用15mM甲氧基聚乙二醇-马来酰亚胺(MAL-PEG 5kDa)标记1小时。用100mM DTT终止反应,并在4℃温度下在50.4Ti转子中以40,000r.p.m(~100,000g)离心20分钟以将整个膜制成小球。将小球用1%月桂酰肌氨酸钠洗涤以溶解细胞质膜并再次离心。将所得的外膜用含有1%DDM的PBS重新悬浮并溶解。用镍珠进行的金属亲和性下拉(Metal-affinity pulldowns with nickel beads)用于SDS-PAGE和抗-His蛋白质免疫印迹。带有空pMC2载体的大肠杆菌LSR12细胞用作阴性对照。

[1091] ATR-FTIR光谱法。在Equinox 55红外光谱仪(Bruker)上进行ATR-FTIR测量,所述红外光谱仪连续用干燥空气吹扫,并配备有液氮制冷型碲汞镉探测器(mercury cadmium telluride detector)和Golden Gate反射附件(Specac)。内反射元件为金刚石晶体(2mm×2mm),光束入射角为45°。将每个纯化的蛋白样本(1μl)在晶体表面铺展并在气态氮气流下干燥以形成膜。以2cm^{-1}分辨率记录的每个光谱为128个累计值的平均值,用以提高信噪比。所有光谱采用水蒸汽贡献减法(water vapour contribution subtraction)来处理,通过变迹法(apodization)在4cm^{-1}最终分辨率下变平滑,并在酰胺I谱带(1,700-1,600cm^{-1})区域上归一化以对它们进行比较[46]。

[1092] 阴性染色EM和对称性测定。阴性染色EM用于监测CsgG、$CsgG_{C1S}$和CsgE的溶液中(in-solution)寡聚状态。将CsgE、$CsgG_{C1S}$和两亲聚合物结合的CsgG调节成0.05mg ml^{-1}的浓度,并施加到辉光放电的碳涂覆铜网格上(CF-400;Electron Microscopy Sciences)。培育1分钟后,将样本染上印记,然后在2%的乙酸双氧铀中洗涤并染色。在Tecnai T12BioTWIN LaB6显微镜上收集图像,操作电压为120kV,放大倍数为×49,000,散焦为800至2,000nm。如针对cryo-EM所述,执行对比传递函数(CTF)、相位翻转和颗粒选择。对于膜提取的CsgG,分别通过九聚体和顶视图分析十八聚体颗粒(共1,780个)。对于纯化的CsgE,总共分析了2,452个颗粒。按以下针对CsgG-CsgE cryo-EM分析所述,获得三维模型,并通过几轮多参考比对(multi-reference alignment,MRA)、多统计分析(MSA)和锚设置细化(anchor set refinement)来细化三维模型。在所有情况下,归一化和中心化后,如cryo-EM部分中所述采用IMAGIC-4D对图像进行分类。对于每组颗粒,选择与特征视图对应的最佳分类。采用最佳分类平均值进行CsgG顶视图的对称性测定,每类大约有20张图像。使用IMAGIC计算旋转自相关函数并进行绘制。

[1093] CsgG-CsgE复合物形成。对于CsgG-CsgE复合物形成,通过向300μl洗涤剂去稳定的脂质体(1mg ml^{-1} 1,2-二肉豆蔻酰-sn-甘油基-3-磷酸胆碱(DMPC)和0.4%的LDAO)中加入120μl CsgG(0.5%C8E4中的24mg ml^{-1}蛋白、4mM LDAO,25mM Tris-HCl pH 8.0,500mM

NaCl,1mM DTT)并在加入90ml(在100mg ml^{-1}原料下)的A8-35两亲聚合物前在冰上培育5分钟,从而将纯化的CsgG中的增溶洗涤剂交换为两亲聚合物A8-35(Anatrace)。在冰上培育另外15分钟后,将样本装载到Superose 6 10/300GL(GE Healthcare)柱上,并在200mM NaCl,2.5%木糖醇,25mM Tris-HCl pH为8,0.2mM DTT中进行凝胶过滤。,将200mM NaCl、2.5%木糖醇、25mM pH值为8的Tris-HCl、0.2mM DTT中的等体积纯化单体CsgE以对于CsgE和CsgG分别为15和5μM的最终蛋白浓度加入到两亲聚合物溶解的CsgG,样本在温度为18℃、电压为125V情况下在0.5×TBE缓冲液中在4.5%的天然PAGE上运行。对于第二个变性维度,将对应于CsgG-CsgE复合物的条带从在同一凝胶上平行运行的未染色通道(lane)中切除,在Laemmli缓冲液(60mM pH值为6.8的Tris-HCl、2%的SDS、10%甘油、5%的2-巯基乙醇、0.01%溴酚蓝)中煮沸5分钟,并在4-20%的SDS-PAGE上运行。纯化的CsgE和CsgG作为对照样本与复合物一起运行。用考马斯亮蓝(InstantBlue Coomassie)对凝胶进行染色,以便进行目视检查,或用SYPRO橙染色,以便通过对SDS-PAGE上的CsgE和CsgG条带进行荧光检测(Typhoon FLA 9000)来对CsgG-CsgE复合物进行化学计量评估,得到0.97的CsgG/CsgE比。

[1094] CsgG-CsgE Cryo-EM。Cryo电子显微镜用于测定C_9CsgG-CsgE复合物的溶液内(in-solution)结构。将如上所述制备的CsgG-CsgE复合物与HisTrap FF(GE Healthcare)结合并从其上洗脱以去除未结合的CsgG,洗脱后立即将其应用于在20mA下辉光放电30秒的Quantifoil R2/2碳涂覆网格(Quantifoil Micro Tools GmbH)。使用自动化系统(Leica)将样本冷冻浸入液氮中,并在FEI F20显微镜下观察,其中操作电压为200kV,并且在低剂量条件下标称放大倍数为×50,000,散焦范围为1.4-3μm。在Falcon II探测器上记录图像帧。标本级别的像素大小为每像素1.9Å。采用CTFFIND3(参考文献47)评估CTF参数,并在SPIDER[48]中进行相位翻转。采用BOXER(EMAN2;参考文献49)从CTF校正的显微照片中自动选择颗粒。丢弃散光超过10%的图像。从75张显微照片中总共选出1,221个颗粒,并且将这些颗粒窗口化成128像素×128像素的盒子。将图像归一化成相同的平均值和标准偏差,并在~200Å的低分辨率截止值下进行高通滤波。使它们居中,然后进行第一轮MSA。采用IMAGIC-4D(图像科学软件(Image Science Software))中的参考免费分类获得初始参考集。与C_9圆柱形颗粒的特征侧视图对应的最佳类别被用作MRA的参考。对于CsgG-CsgE复合物,从最佳的125个特征视图(具有良好的对比度和明确定义的特征)中计算出第一个三维模型,所述125个特征视图包括复合物的1,221个颗粒,其方向通过角度重构确定(Image Science Software)。三维图通过MRA、MSA和锚设置细化而进行的多轮迭代进行细化。分辨率通过傅里叶壳相关性(Fourier shell correlation,FSC)根据0.5标准水平估计为24 Å(图52)。在UCSF Chimera[50]中进行图和数字的可视化。

[1095] 胆汁盐毒性试验。通过在含有胆汁盐的琼脂板上减慢的生长来研究外膜渗透性。将大肠杆菌LSR12(参考文献7)细胞的十倍连续稀释液(5μl)点样在McConkey琼脂板上,其中,所述稀释液含有pLR42(参考文献16)和pMC2(参考文献14)(或衍生的螺旋物2突变体),所述琼脂板含有100μg l^{-1}氨苄青霉素、25μg l^{-1}氯霉素、1mM含有或不含0.2%(w/v)L-阿拉伯糖的IPTG。在37℃下培育过夜后,检查菌落生长。

[1096] 单通道电流记录。采用带有来自Nanion的Orbit 16试剂盒的并行高分辨率电记录进行单通道电流记录。简言之,在16通道多电极空腔阵列(MECA)芯片(lonera)[51]中的微腔(亚皮升体积(subpicolitre volume)的)上形成1,2-二植烷酰基-sn-甘油基-3-磷酸胆碱

(Avanti Polar Lipids)的水平双层。双层上方和下方的顺式和反式腔均含有1.0M KCl、25mM pH值为8.0的Tris-HCl。为了将通道插入膜中,将溶解在25mM pH值为8.0的Tris-HCl、500mM NaCl、1mM DTT、0.5%C8E4、5mM LDAO中的CsgG加入到顺式腔室中直至最终浓度为90-300nM。为了测试CsgG通道与CsgE的相互作用,将溶解在25mM pH值为8.0的Tris-HCl、150mM NaCl中的CsgE的溶液加入到顺式腔室中直至最终浓度为0.1、1、10和100nM。采用Tecella Triton 16通道放大器,在保持电势为+50mV和-50mV(顺式侧接地)情况下,记录跨膜电流,其中,所述16通道放大器的低通滤波频率为3kHz且采样频率为10kHz。使用pClamp套件(Molecular Devices)的Clampfit分析电流迹线。使用Origin 8.6(Microcal)[52]生成图。

[1097] 基于CsgG X射线结构的孔尺寸,将测量的电流与计算出的电流进行比较,建模为由三段组成:跨膜段、周质前庭和连接两者的内通道收缩结构。将前两个段建模为圆锥形,而收缩结构表示为圆柱体。相应的电阻R_1,R_2和R_3分别计算为

[1098] $R1=L1/(\pi D_1 d_4 \kappa)$

[1099] $R2=L2/(\pi D_2 d_2 \kappa)$

[1100] $R3=L3/(\pi D_1 d_2 \kappa)$

[1101] 其中,L_1、L_2和L_3为各段的轴向长度,经测量分别为3.5、4.0和2.0nm,并且D_1、d_1、D_2和d_2为段1和段2的最大和最小直径,经测量分别为4.0、0.8、3.4和0.8nm。电导率k的值为10.6Sm^{-1}。通过将R_1、R_2和R_3以及电压V=50mV代入到下式来计算电流:

[1102] $I=V/(R1+R2+R3)$

[1103] 未发现接入电阻显著改变预测电流。

[1104] 参考文献:

[1105] 1.Olsen,A.,Jonsson,A.&Normark,S.Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli.Nature 338,652-655(1989).

[1106] 2.Collinson,S.K.et al.Thin,aggregative fimbriae mediate binding of Salmonella enteritidis to fibronectin.J.Bacteriol.175,12-18(1993).

[1107] 3.Dueholm,M.S.,Albertsen,M.,Otzen,D.&Nielsen,P.H.Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure.PLoS ONE 7,e51274(2012).

[1108] 4.Cegelski,L.et al.Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation.Nature Chem.Biol.5,913-919(2009).

[1109] 5.Herwald,H.et al.Activation of the contact-phase system on bacterial surfaces--a clue to serious complications in infectious diseases.Nature Med.4,298-302(1998).

[1110] 6.Hammar,M.,Arnqvist,A.,Bian,Z.,Olsen,A.&Normark,S.Expression of two csg operons is required for production of fibronectin-and Congo red-binding curli polymers in Escherichia coli K-12.Mol.Microbiol.18,661-670(1995).

[1111] 7.Chapman,M.R.et al.Role of Escherichia coli curli operons in directing amyloid fiber formation.Science 295,851-855(2002).

[1112] 8.Wang,X.,Smith,D.R.,Jones,J.W.&Chapman,M.R.In vitro polymerization of a functional Escherichia coli amyloid protein.J.Biol.Chem.282,3713-3719(2007).

[1113] 9.Dueholm,M.S.et al.Fibrillation of the major curli subunit CsgA under a wide range of conditions implies a robust design of aggregation.Biochemistry 50,8281-8290(2011).

[1114] 10.Hung,C.et al.Escherichia coli biofilms have an organized and complex extracellular matrix structure.MBio 4,e00645-13(2013).

[1115] 11.Hammar,M.,Bian,Z.&Normark,S.Nucleator-dependent intercellular assembly of adhesive curli organelles in Escherichia coli.Proc.Natl Acad.Sci.USA 93,6562-6566(1996).

[1116] 12.Bian,Z.&Normark,S.Nucleator function of CsgB for the assembly of adhesive surface organelles in Escherichia coli.EMBO J.16,5827-5836(1997).

[1117] 13.Loferer,H.,Hammar,M.&Normark,S.Availability of the fibre subunit CsgA and the nucleator protein CsgB during assembly of fibronectin-binding curli is limited by the intracellular concentration of the novel lipoprotein CsgG.Mol.Microbiol.26,11-23(1997).

[1118] 14.Robinson,L.S.,Ashman,E.M.,Hultgren,S.J.&Chapman,M.R.Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein.Mol.Microbiol.59,870-881(2006).

[1119] 15.Nenninger,A.A.,Robinson,L.S.&Hultgren,S.J.Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF.Proc.Natl Acad.Sci.USA 106,900-905(2009).

[1120] 16.Nenninger,A.A.et al.CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation.Mol.Microbiol.81,486-499(2011).

[1121] 17.Okuda,S.&Tokuda,H.Lipoprotein sorting in bacteria.Annu.Rev.Microbiol.65,239-259(2011).

[1122] 18.lacovache,I.,Bischofberger,M.&van der Goot,F.G.Structure and assembly of pore-forming proteins.Curr.Opin.Struct.Biol.20,241-246(2010).

[1123] 19.Goyal,P.,Van Gerven,N.,Jonckheere,W.&Remaut,H.Crystallization and preliminary X-ray crystallographic analysis of the curli transporter CsgG.Acta Crystallogr.F Struct.Biol.Cryst.Commun.69,1349-1353(2013).

[1124] 20.Krantz,B.A.et al.A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore.Science 309,777-781(2005).

[1125] 21.Janowiak,B.E.,Fischer,A.&Collier,R.J.Effects of introducing a single charged residue into the phenylalanine clamp of multimeric anthrax protective antigen.J.Biol.Chem.285,8130-8137(2010).

[1126] 22.Feld,G.K.,Brown,M.J.&Krantz,B.A.Ratcheting up protein translocation with anthrax toxin.Protein Sci.21,606-624(2012).

[1127] 23.Van Gerven,N.et al.Secretion and functional display of fusion proteins through the curli biogenesis pathway.Mol.Microbiol.91,1022-1035(2014).

[1128] 24.Brinker,A.et al.Dual function of protein confinement in chaperonin-assisted protein folding.Cell 107,223-233(2001).

[1129] 25.Busby,J.N.,Panjikar,S.,Landsberg,M.J.,Hurst,M.R.&Lott,J.S.The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device.Nature 501,547-550(2013).

[1130] 26.Takagi,F.,Koga,N.&Takada,S.How protein thermodynamics and folding mechanisms are altered by the chaperonin cage:molecular simulations.Proc.Natl Acad.Sci.USA 100,11367-11372(2003).

[1131] 27.Zhou,H.X.Protein folding in confined and crowded environments.Arch.Biochem.Biophys.469,76-82(2008).

[1132] 28.Chen,A.Y.et al.Synthesis and patterning of tunable multiscale materials with engineered cells.Nature Mater.13,515-523(2014).

[1133] 29.Sivanathan,V.&Hochschild,A.A bacterial export system for generating extracellular amyloid aggregates.Nature Protocols 8,1381-1390(2013).

[1134] 30.Datsenko,K.A.&Wanner,B.L.One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.Proc.Natl Acad.Sci.USA97,6640-6645(2000).

[1135] 31.Swamy,M.,Siegers,G.M.,Minguet,S.,Wollscheid,B.&Schamel,W.W.A.Blue native polyacrylamide gel electrophoresis(BN-PAGE)for the identification and analysis of multiprotein complexes.Sci.STKE 2006,pl4,http://dx.doi.org/10.1126/stke.3452006pl4(2006).

[1136] 32.Winter,G.xia2:an expert system for macromolecular crystallography data reduction.J.Appl.Cryst.43,186-190(2010).

[1137] 33.Kabsch,W.Xds.Acta Crystallogr.D Biol.Crystallogr.66,125-132(2010).

[1138] 34.Sheldrick,G.M.Experimental phasing withSHELXC/D/E:combining chain tracing with density modification.Acta Crystallogr.D Biol.Crystallogr.66,479-485(2010).

[1139] 35.Bricogne,G.,Vonrhein,C.,Flensburg,C.,Schiltz,M.&Paciorek,W.Generation,representation and flow of phase information in structure determination:recent developments in and around SHARP 2.0.Acta Crystallogr.D Biol.Crystallogr.59,2023-2030(2003).

[1140] 36.Cowtan,K.Recent developments in classical density modification.Acta Crystallogr.D Biol.Crystallogr.66,470-478(2010).

[1141] 37.Cowtan,K.The Buccaneer software for automated model building.1.Tracing protein chains.Acta Crystallogr.D Biol.Crystallogr.62,

1002-1011(2006).

[1142] 38.Adams,P.D.et al.PHENIX:a comprehensive Python-based system for macromolecular structure solution.Acta Crystallogr.D Biol.Crystallogr.66,213-221(2010).

[1143] 39.Emsley,P.,Lohkamp,B.,Scott,W.G.&Cowtan,K.Features and development of Coot.Acta Crystallogr.D Biol.Crystallogr.66,486-501(2010).

[1144] 40.Davis,I.W.et al.MolProbity:all-atom contacts and structure validation for proteins and nucleic acids.Nucleic Acids Res.35(Suppl 2),W375-W383(2007).

[1145] 41.Strong,M.et al.Toward the structural genomics of complexes:crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis.Proc.Natl Acad.Sci.USA 103,8060-8065(2006).

[1146] 42.McCoy,A.J.et al.Phaser crystallographic software.J.Appl.Cryst.40,658-674(2007).

[1147] 43.Smart,O.S.et al.Exploiting structure similarity in refinement:automated NCS and target-structure restraints in BUSTER.Acta Crystallogr.D Biol.Crystallogr.68,368-380(2012).

[1148] 44.Murshudov,G.N.et al.REFMAC5 for the refinement of macromolecular crystal structures.Acta Crystallogr.D Biol.Crystallogr.67,355-367(2011).

[1149] 45.Nicholls,R.A.,Long,F.&Murshudov,G.N.Low-resolution refinement tools in REFMAC5.Acta Crystallogr.D Biol.Crystallogr.68,404-417(2012).

[1150] 46.Goormaghtigh,E.&Ruysschaert,J.M.Subtraction of atmospheric water contribution in Fourier transform infrared spectroscopy of biological membranes and proteins.Spectrochim.Acta 50A,2137-2144(1994).

[1151] 47.Mindell,J.A.&Grigorieff,N.Accurate determination of local defocus and specimen tilt in electron microscopy.J.Struct.Biol.142,334-347(2003).

[1152] 48.Shaikh,T.R.et al.SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs.Nature Protocols 3,1941-1974(2008).

[1153] 49.Tang,G.et al.EMAN2:an extensible image processing suite for electron microscopy.J.Struct.Biol.157,38-46(2007).

[1154] 50.Pettersen,E.F.et al.UCSF Chimera--a visualization system for exploratory research and analysis.J.Comput.Chem.25,1605-1612(2004).

[1155] 51.Del Rio Martinez,J.M.,Zaitseva,E.,Petersen,S.,Baaken,G.&Behrends,J.C.Automated formation of lipid membrane microarrays for ionic single molecule sensing with protein nanopores.Small http://dx.doi.org/10.1002/smll.201402016(13 Aug ust 2014).

[1156] 52.Movileanu,L.,Howorka,S.,Braha,O.&Bayley,H.Detecting protein analytes that modulate transmembrane movement of a polymer chain within a

single protein pore.Nature Biotechnol.18,1091-1095(2000).

[1157] 53.Pronk,S.et al.GROMACS 4.5:a high-throughput and highly parallel open source molecular simulation toolkit.Bioinformatics 29,845-854(2013).

[1158] 54.Lindorff-Larsen,K.et al.Improved side-chain torsion potentials for the Amber ff99SB protein force field.Proteins 78,1950-1958(2010).

[1159] 55.Capra,J.A.&Singh,M.Predicting functionally important residues from sequence conservation.Bioinformatics 23,1875-1882(2007).

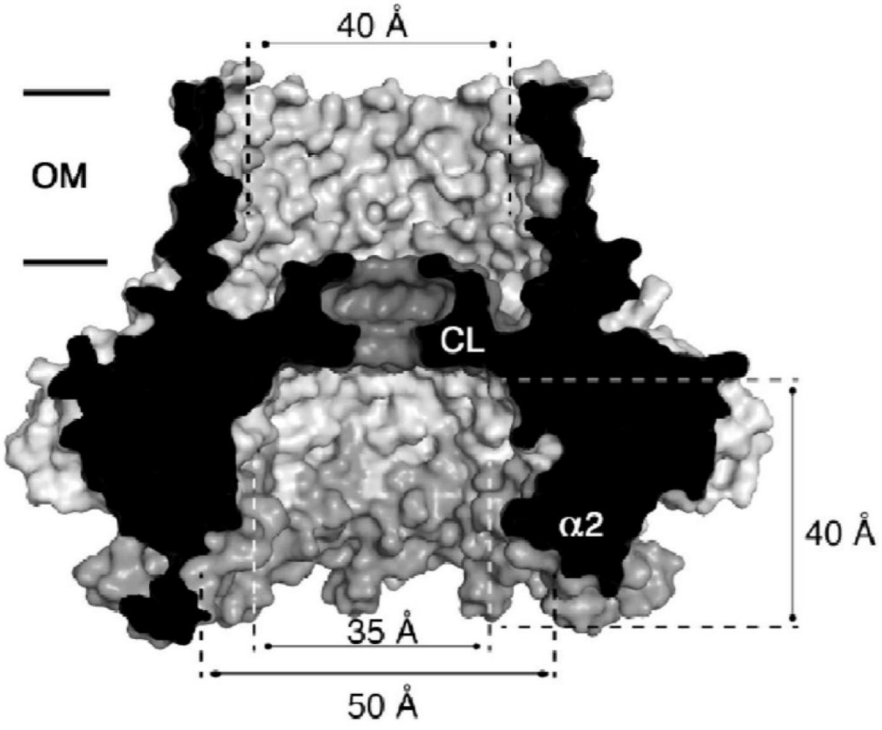


图1

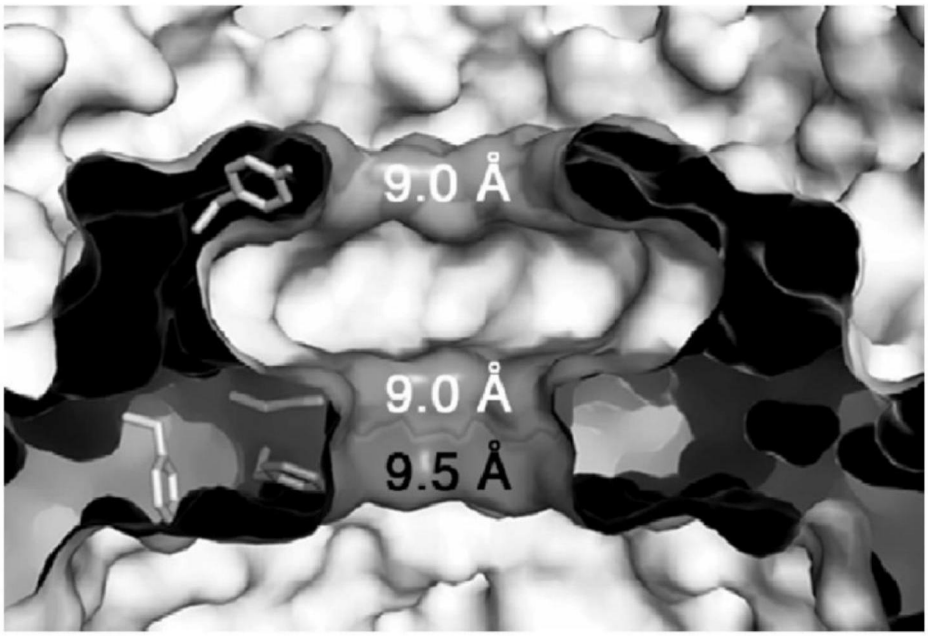


图2

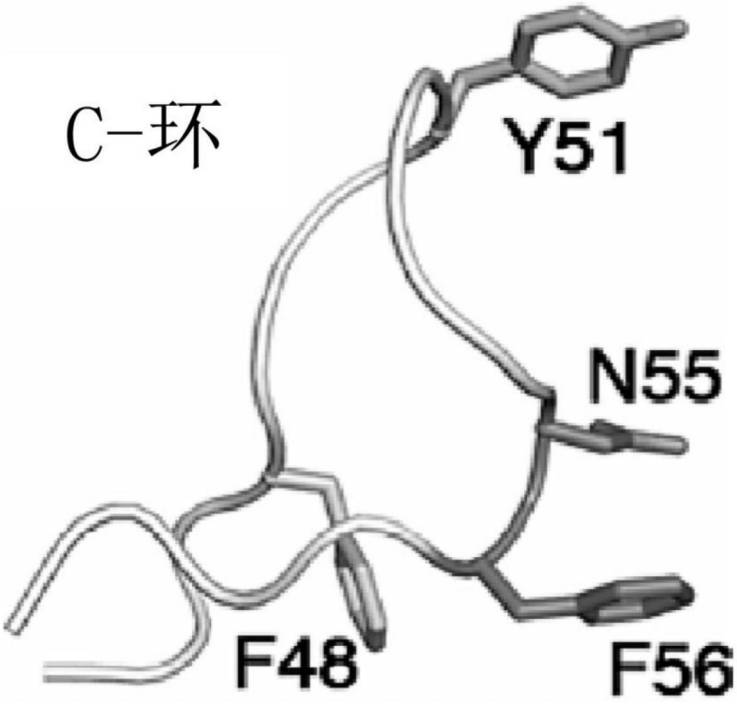


图3

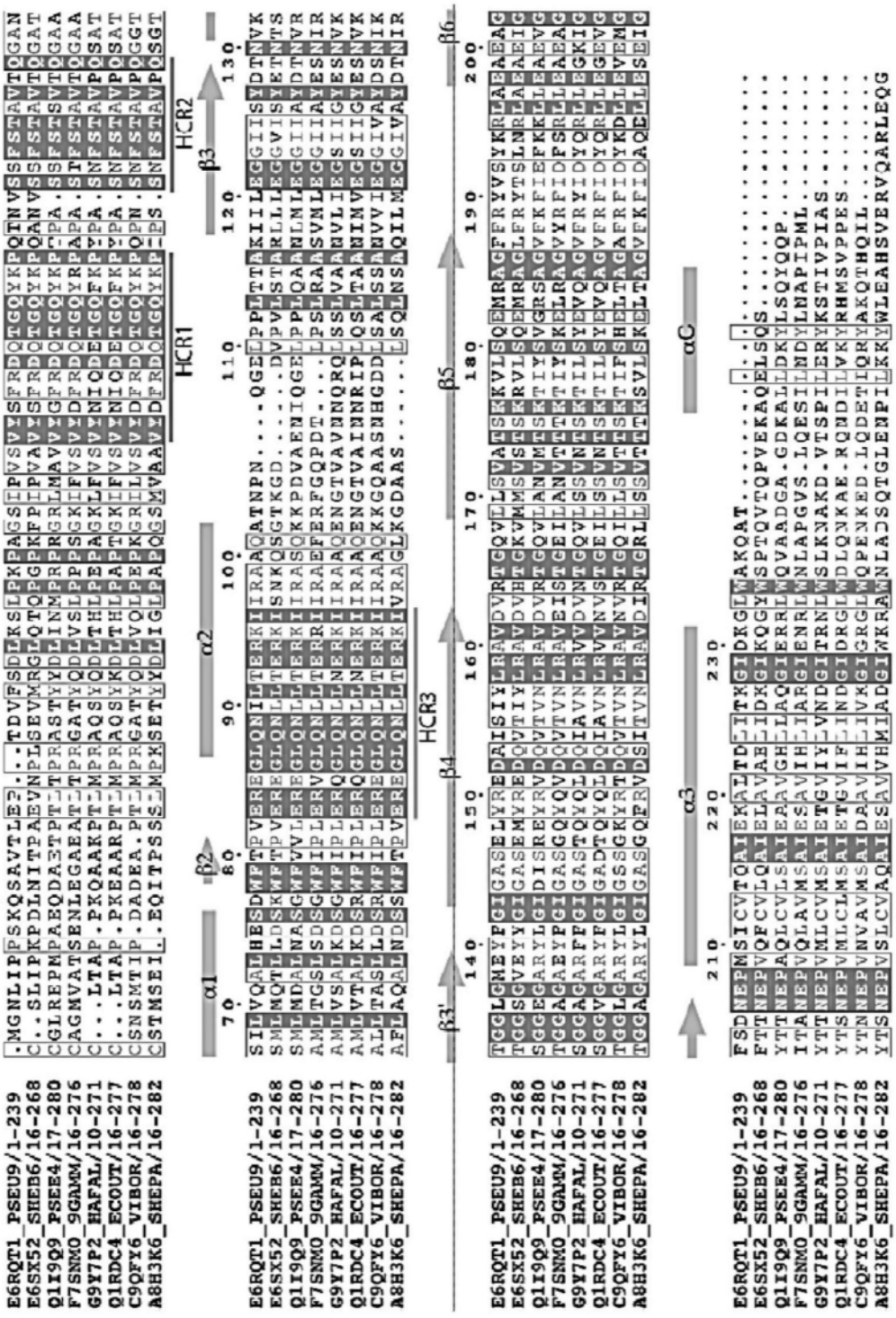

αN
β1
Y51
N55 F56
HCR1
HCR2
β3
α1
β2
α2
HCR3
β3'
β4
β5
β6
α3
αC
E6RQT1_PSEU9/1-239
E6SX52_SHEB6/16-268
Q1I9Q9_PSEE4/17-280
F7SNMO_9GAMM/16-276
G9Y7P2_HAFAL/10-271
Q1RDC4_ECOUT/16-277
C9QFY6_VIBOR/16-278
A8H3K6_SHEPA/16-282

图4

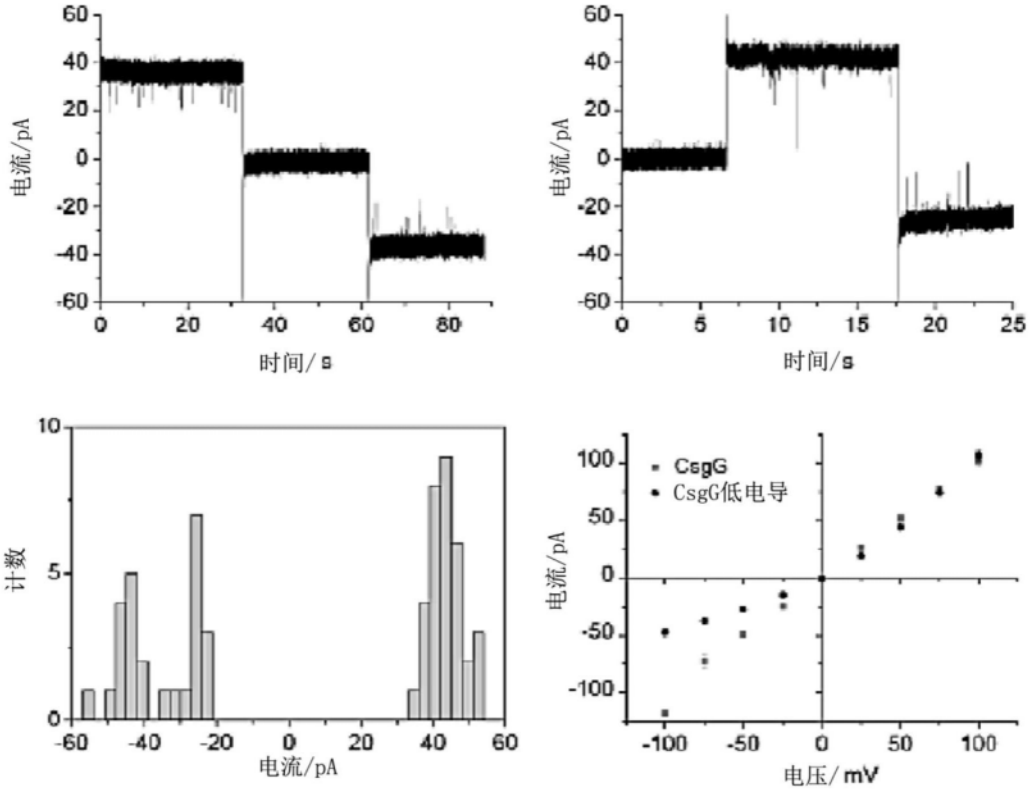

电流/pA
时间/s
计数
电流/pA
CsgG
CsgG低电导
电压/mV

图5

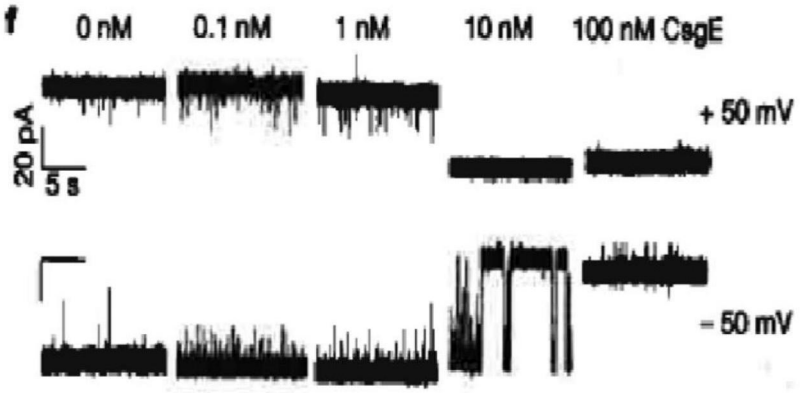

f
0 nM
0.1 nM
1 nM
10 nM
100 nM CsgE
+ 50 mV
20 pA
5 s
- 50 mV

图6

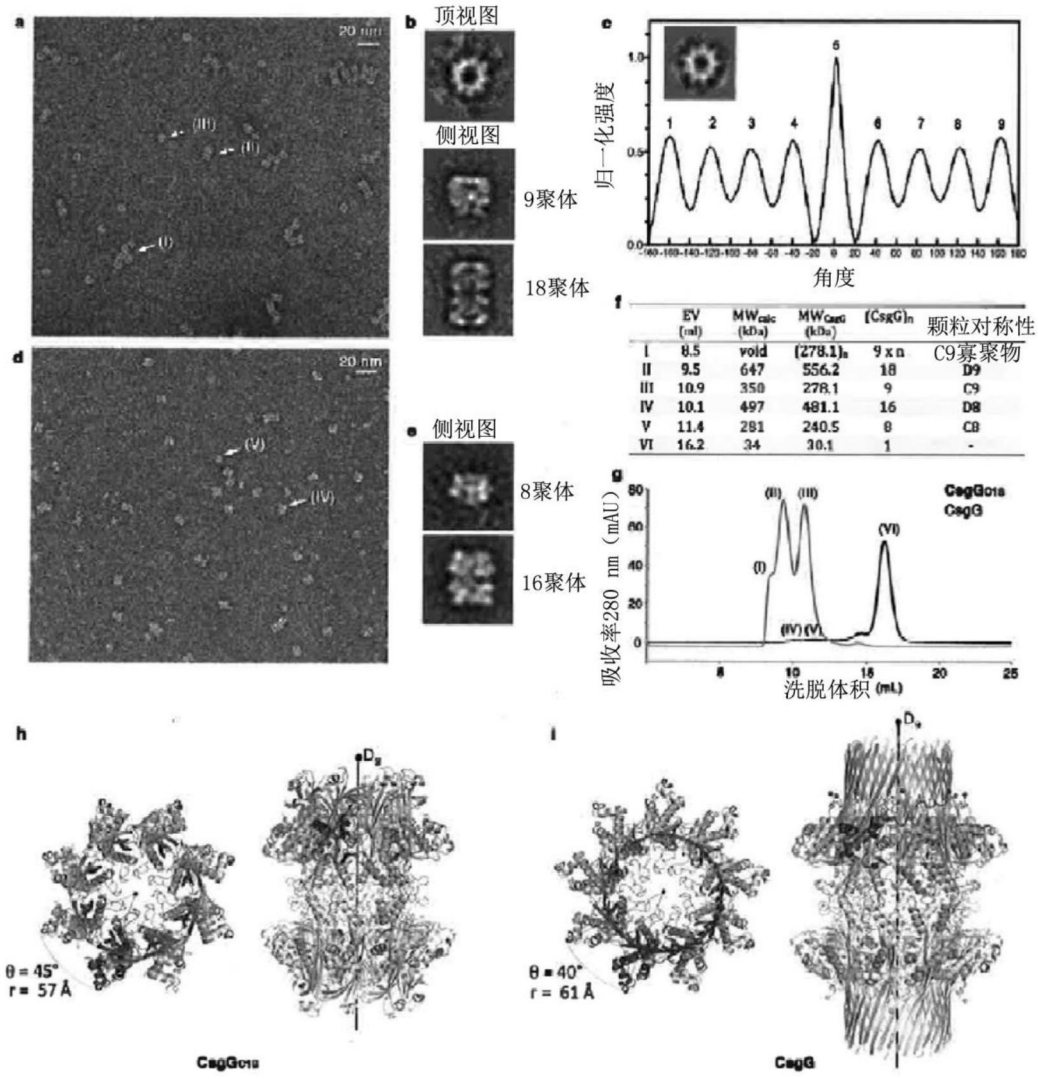

a
20 nm
(III)
(II)
(I)
d
20 nm
(V)
(IV)
b
顶视图
侧视图
9聚体
18聚体
e
侧视图
8聚体
16聚体
c
归一化强度
角度
f
EV (ml)
MWcalc (kDa)
MWCagG (kDa)
(CsgG)n
颗粒对称性
I 8.5 void (278.1)n 9 x n C9寡聚物
II 9.5 647 556.2 18 D9
III 10.9 350 278.1 9 C9
IV 10.1 497 481.1 16 D8
V 11.4 281 240.5 8 C8
VI 16.2 34 30.1 1 -
g
吸收率280 nm (mAU)
洗脱体积 (mL)
CsgG18
CsgG
h
θ = 45°
r = 57 Å
CsgG18
i
θ = 40°
r = 61 Å
CsgG

图7

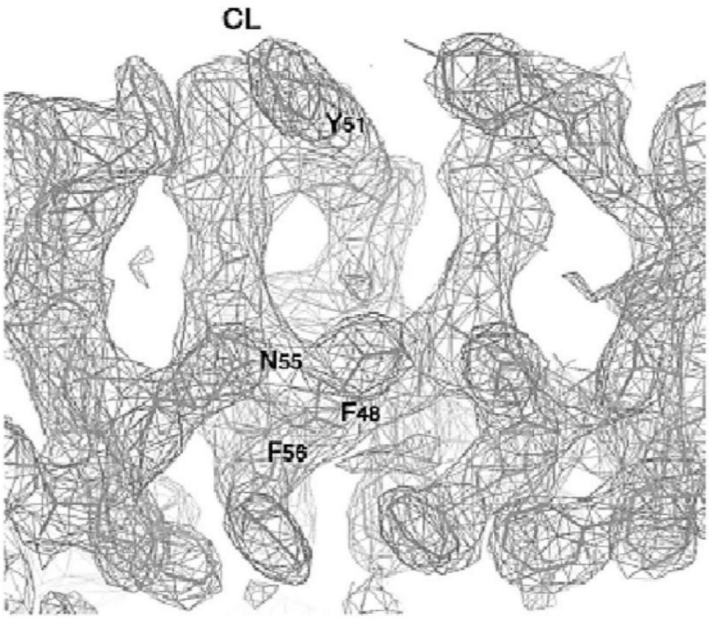


图8

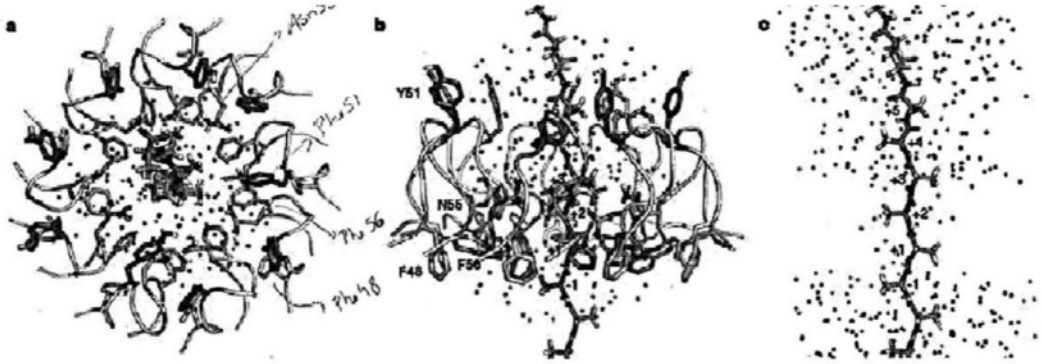


图9

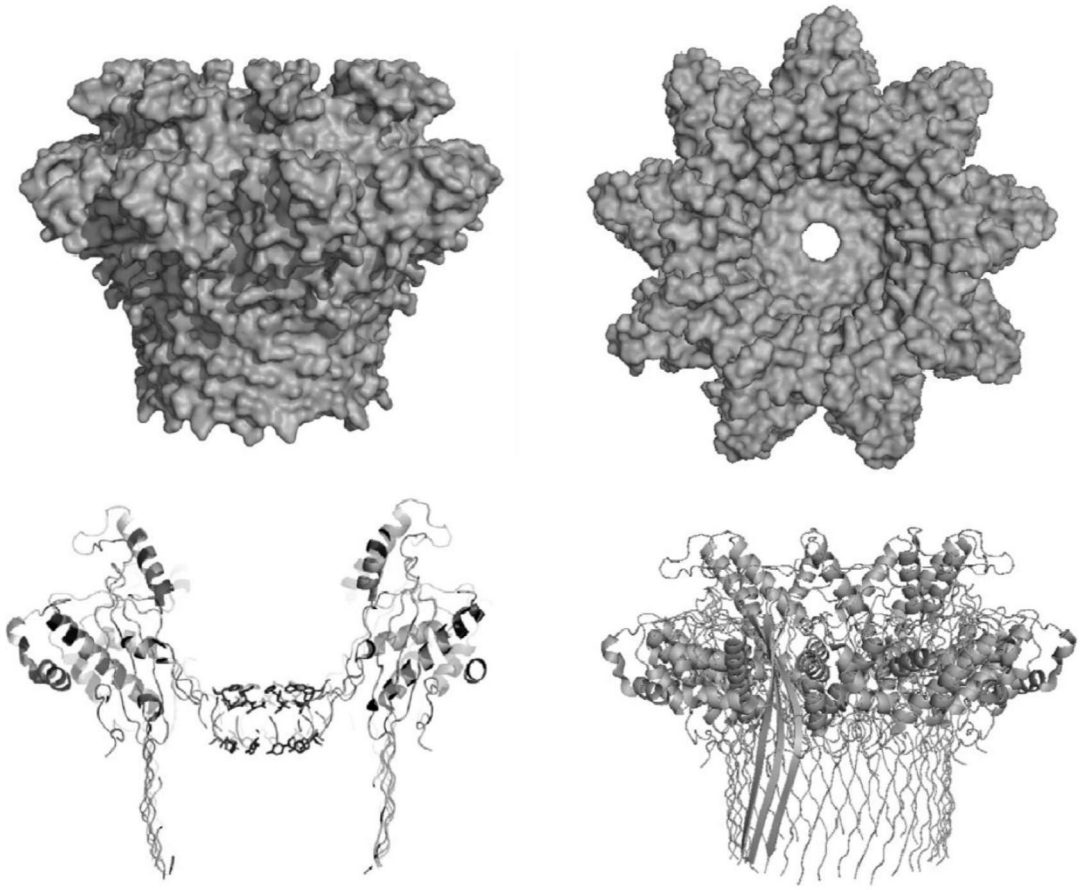


图10

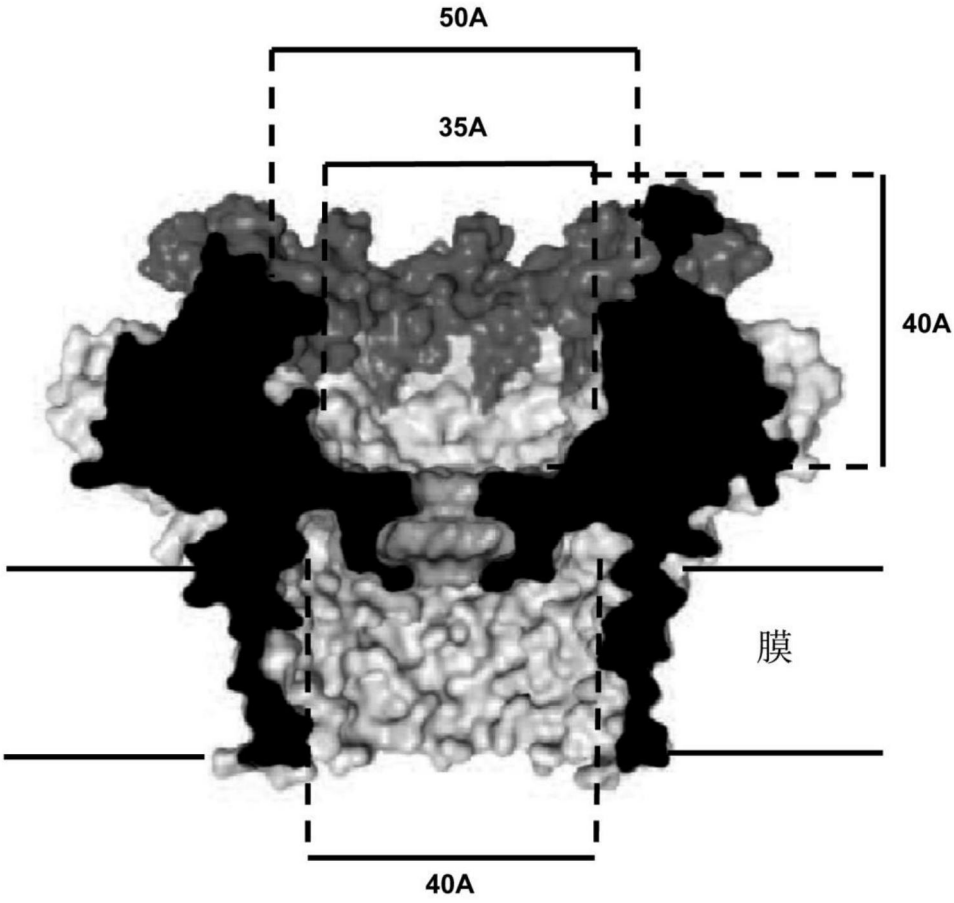


图11

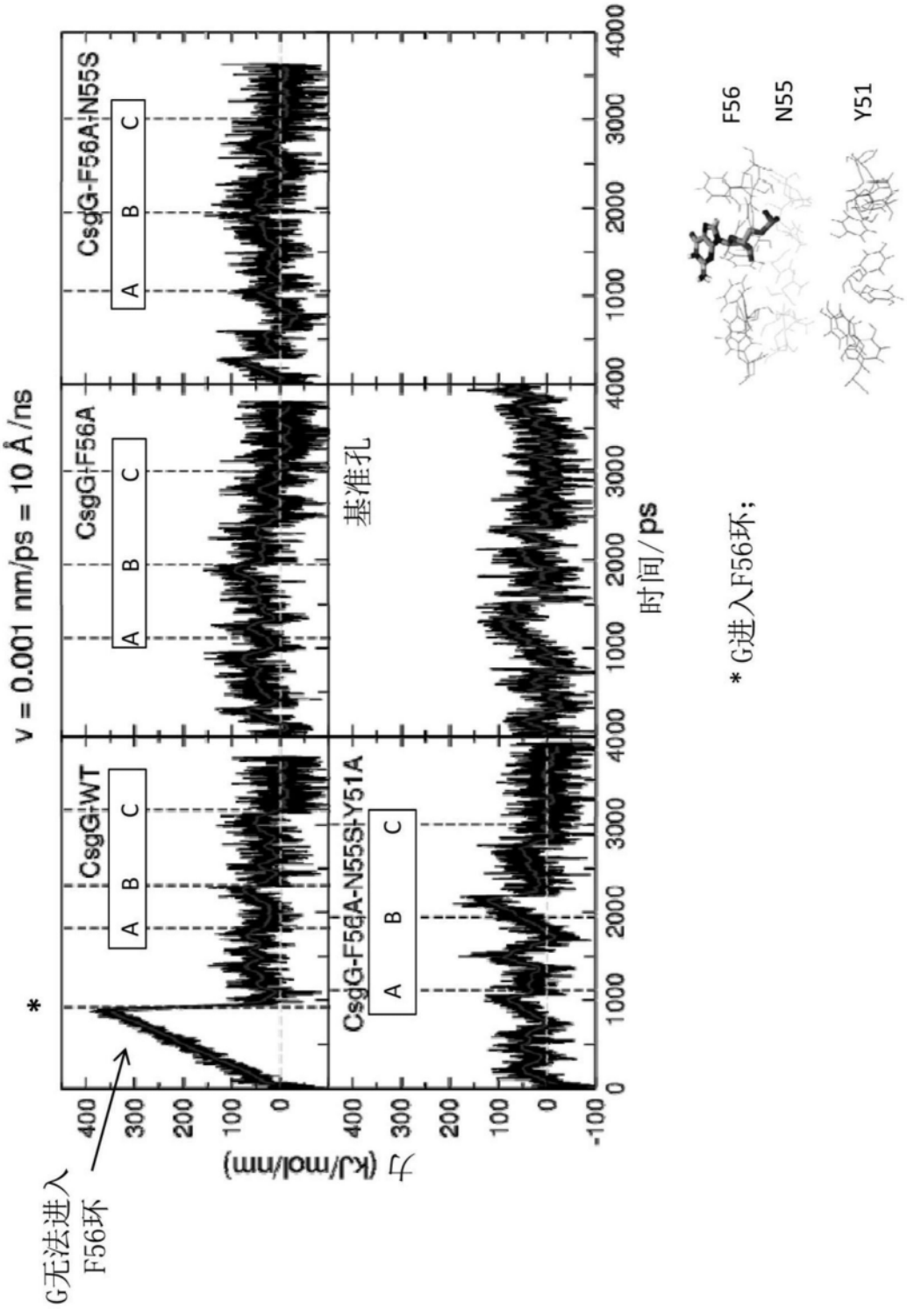

v = 0.001 nm/ps = 10 Å/ns
*
G无法进入
F56环
CsgG-WT
CsgG-F56A-N55S-Y51A
CsgG-F56A
CsgG-F56A-N55S
A B C
力 (kJ/mol/nm)
基准孔
时间/ps
* G进入F56环;
F56
N55
Y51

图12

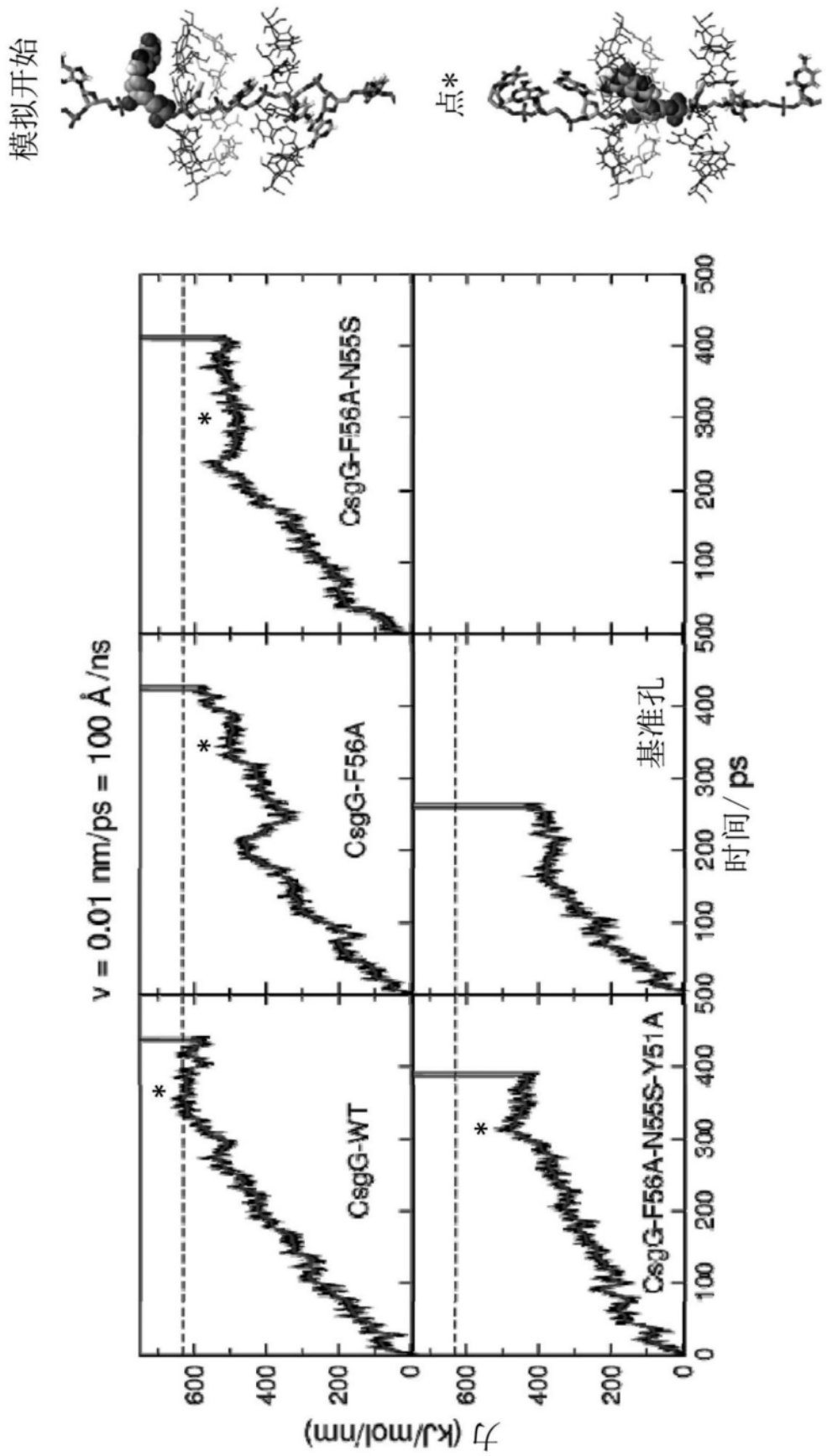

模拟开始
点*
v = 0.01 nm/ps = 100 Å/ns
CsgG-WT
CsgG-F56A
CsgG-F56A-N55S
CsgG-F56A-N55S-Y51A
基准孔
力 (kJ/mol/nm)
时间/ ps
0
200
400
600
0
100
200
300
400
500

图13

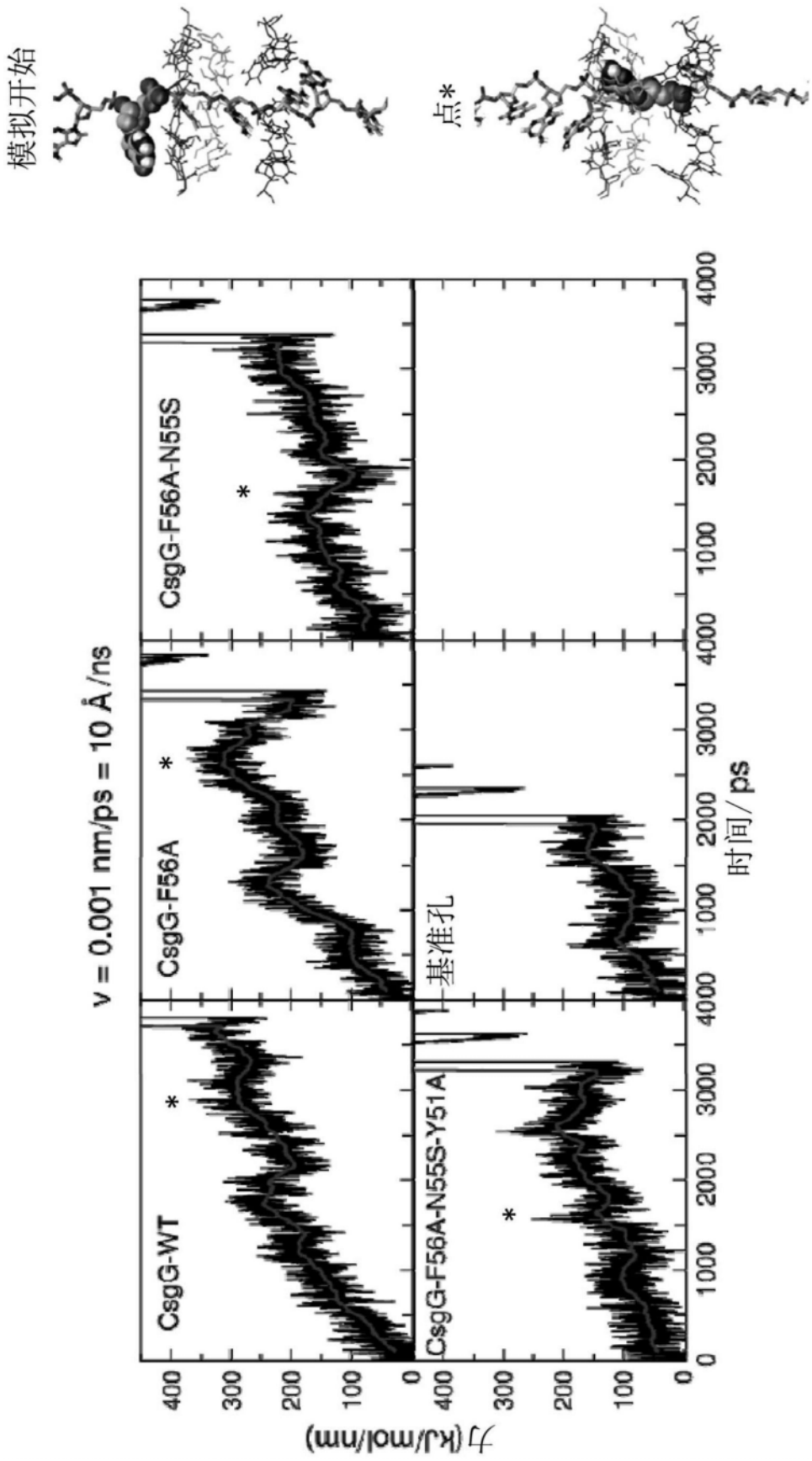

模拟开始
点*
v = 0.001 nm/ps = 10 Å/ns
CsgG-WT
CsgG-F56A
CsgG-F56A-N55S
CsgG-F56A-N55S-Y51A
基准孔
力 (kJ/mol/nm)
时间 / ps
0
100
200
300
400
1000
2000
3000
4000

图14

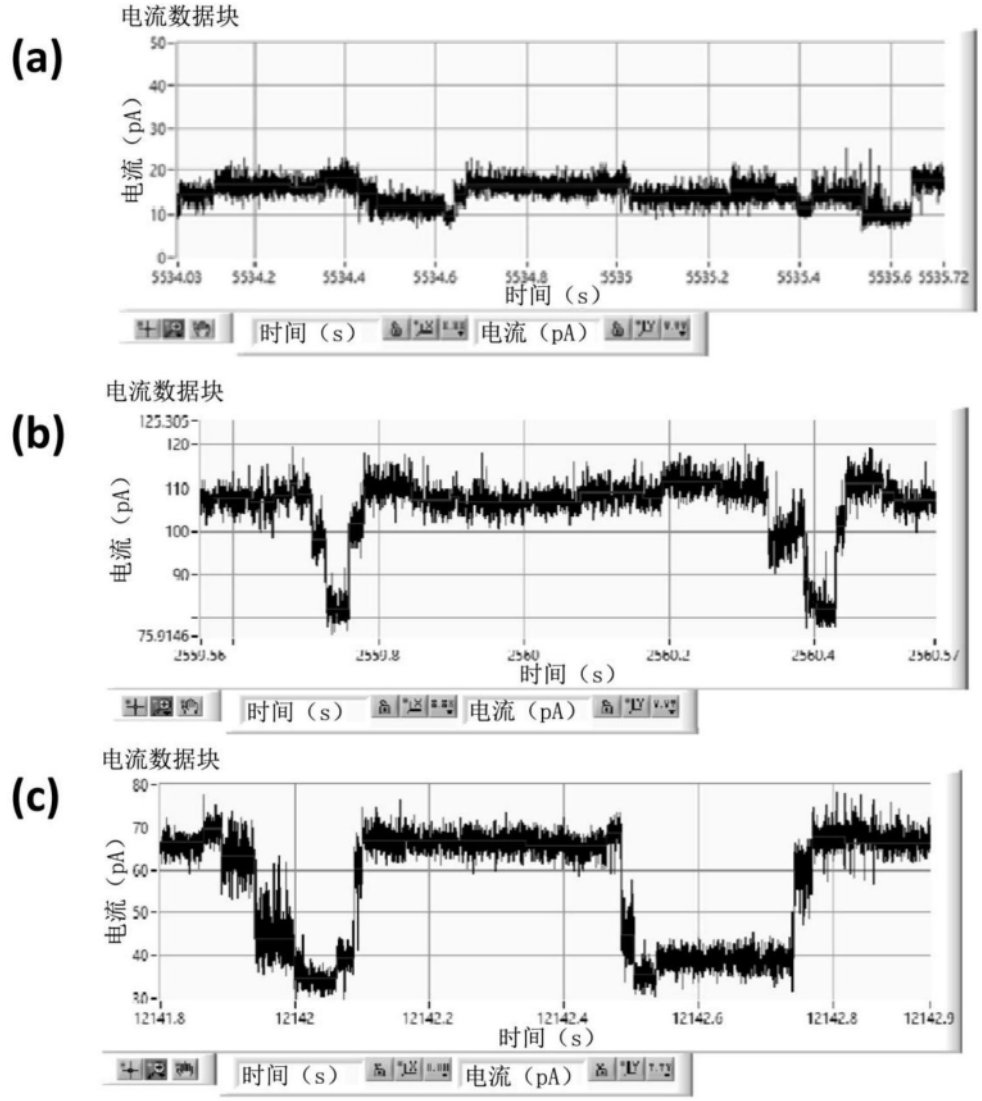

(a)
电流数据块
电流（pA）
时间（s）
5534.03
5534.2
5534.4
5534.6
5534.8
5535
5535.2
5535.4
5535.6
5535.72
时间（s）
电流（pA）
(b)
电流数据块
125.305
120
110
100
90
75.9146
电流（pA）
2559.56
2559.8
2560
2560.2
2560.4
2560.57
时间（s）
时间（s）
电流（pA）
(c)
电流数据块
80
70
60
50
40
30
电流（pA）
12141.8
12142
12142.2
12142.4
12142.6
12142.8
12142.9
时间（s）
时间（s）
电流（pA）

图15

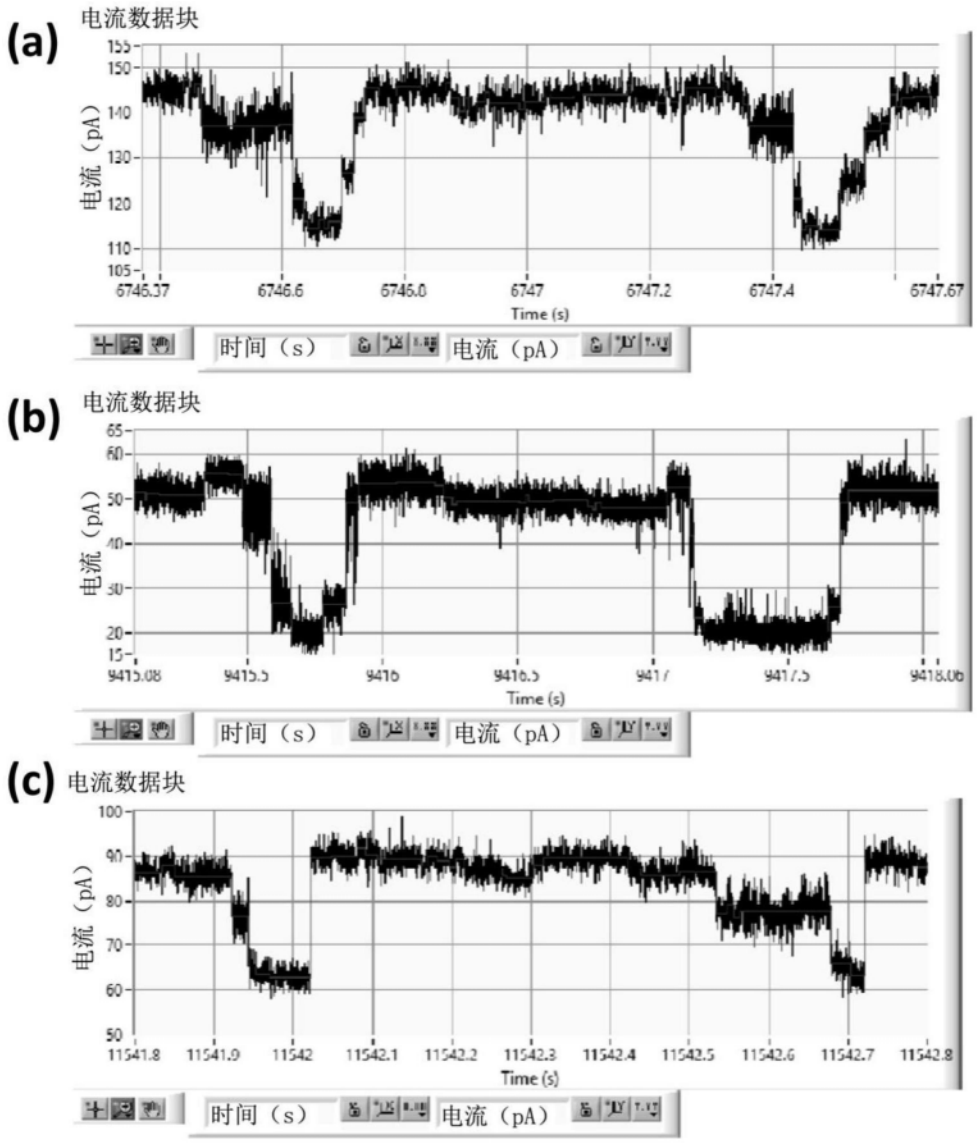

(a)
电流数据块
电流（pA）
Time (s)
时间（s）
电流（pA）
(b)
电流数据块
电流（pA）
Time (s)
时间（s）
电流（pA）
(c)
电流数据块
电流（pA）
Time (s)
时间（s）
电流（pA）

图16

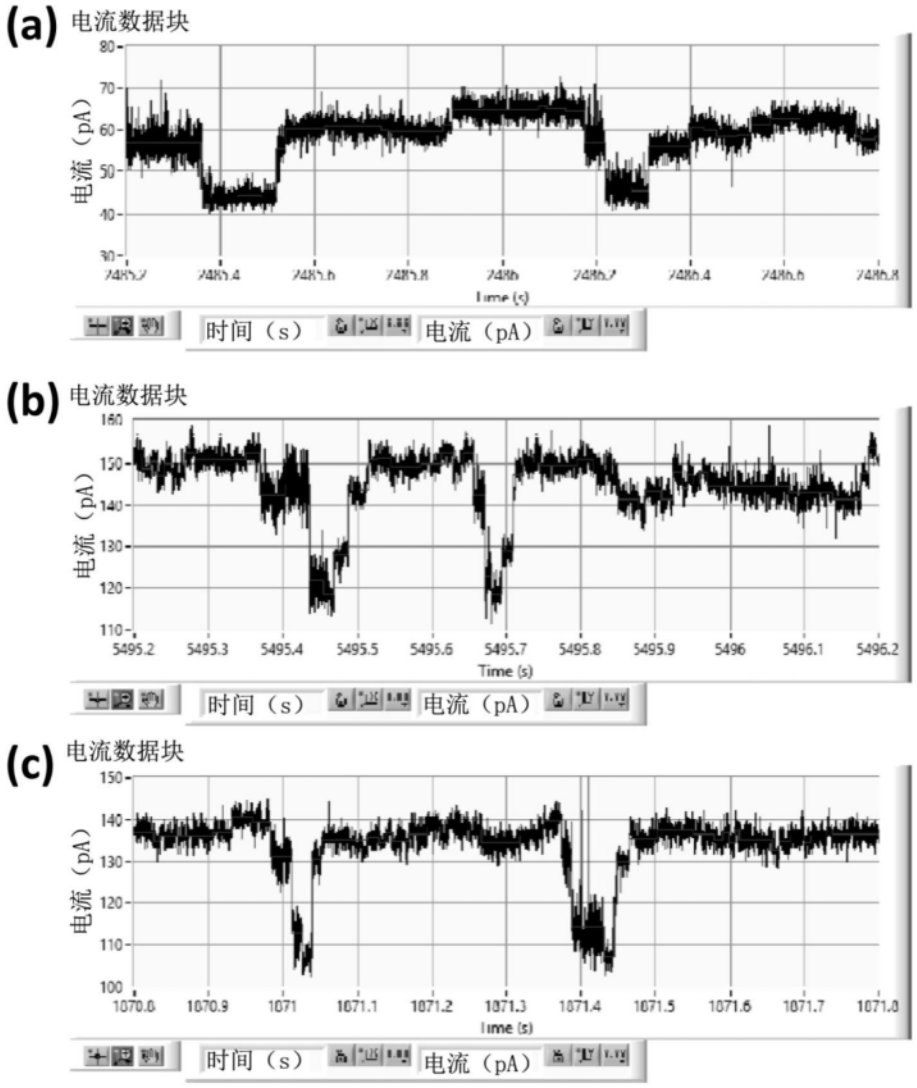

(a)
电流数据块
电流（pA）
80
70
60
50
40
30
Time (s)
时间（s）
电流（pA）
(b)
电流数据块
电流（pA）
160
150
140
130
120
110
5495.2
5495.3
5495.4
5495.5
5495.6
5495.7
5495.8
5495.9
5496
5496.1
5496.2
Time (s)
时间（s）
电流（pA）
(c)
电流数据块
电流（pA）
150
140
130
120
110
100
1070.8
1070.9
1071
1071.1
1071.2
1071.3
1071.4
1071.5
1071.6
1071.7
1071.8
Time (s)
时间（s）
电流（pA）

图17

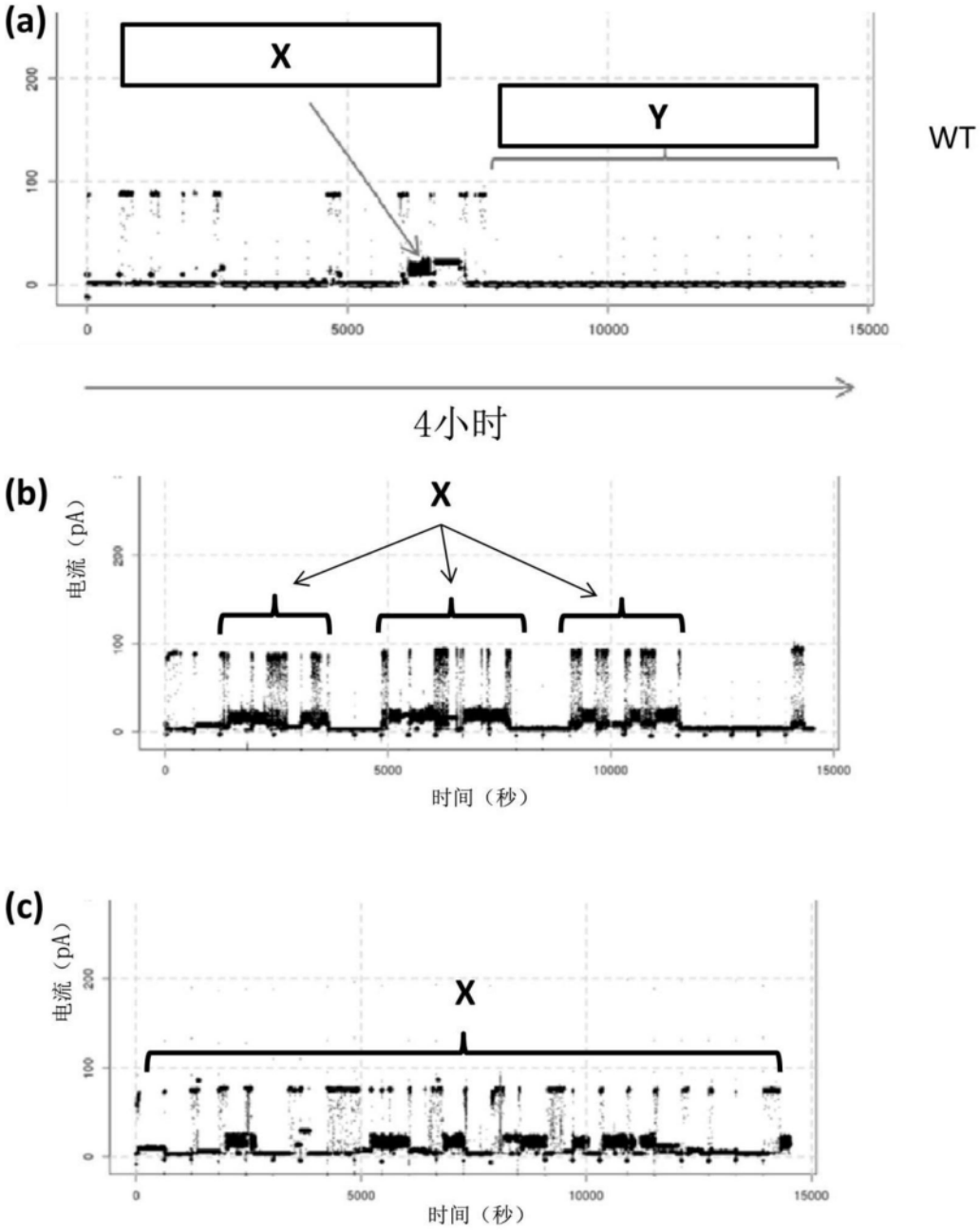

(a)
X
Y
WT
200
100
0
0
5000
10000
15000
4小时
(b)
X
电流（pA）
200
100
0
0
5000
10000
15000
时间（秒）
(c)
X
电流（pA）
200
100
0
0
5000
10000
15000
时间（秒）

图18

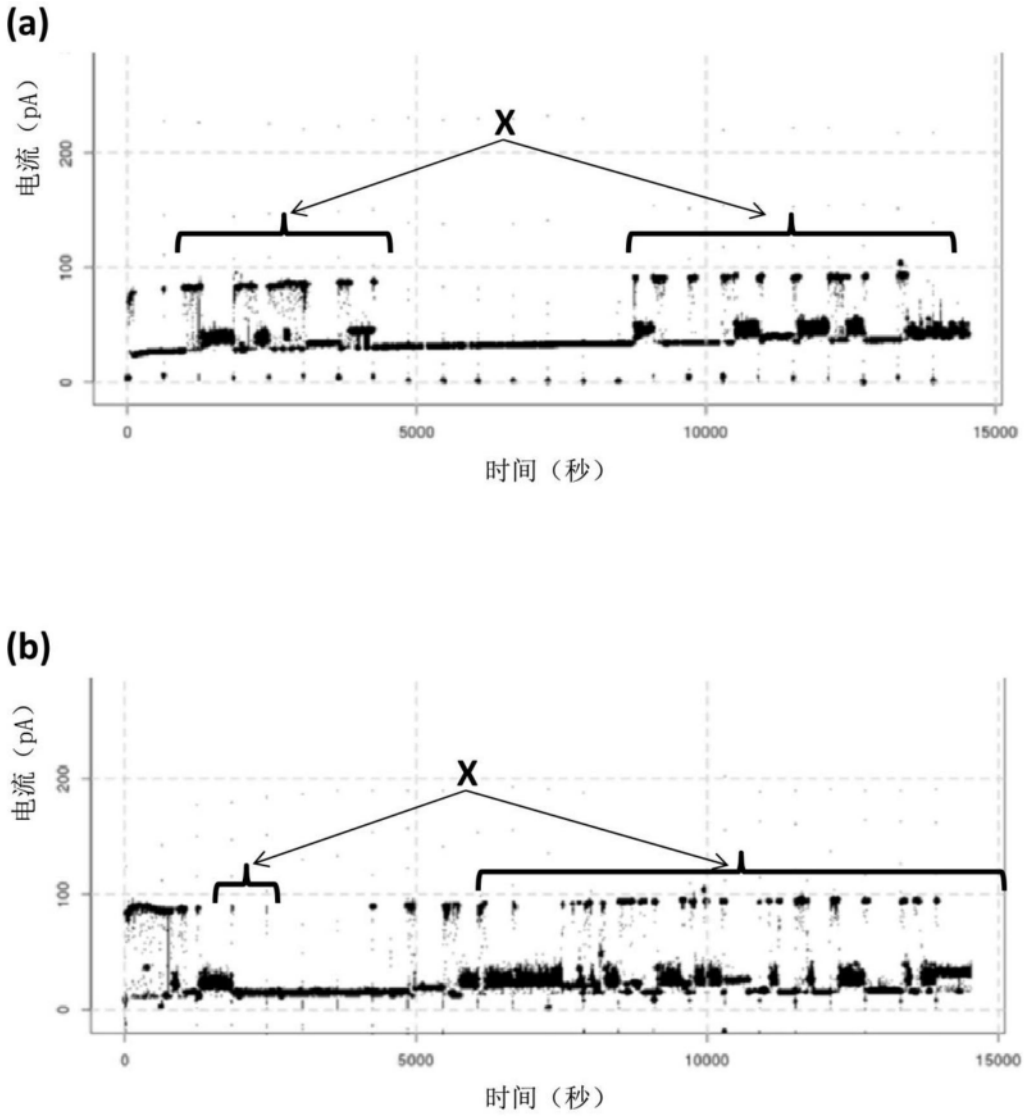

(a)
电流（pA）
X
0
100
200
0
5000
10000
15000
时间（秒）
(b)
电流（pA）
X
0
100
200
0
5000
10000
15000
时间（秒）

图19

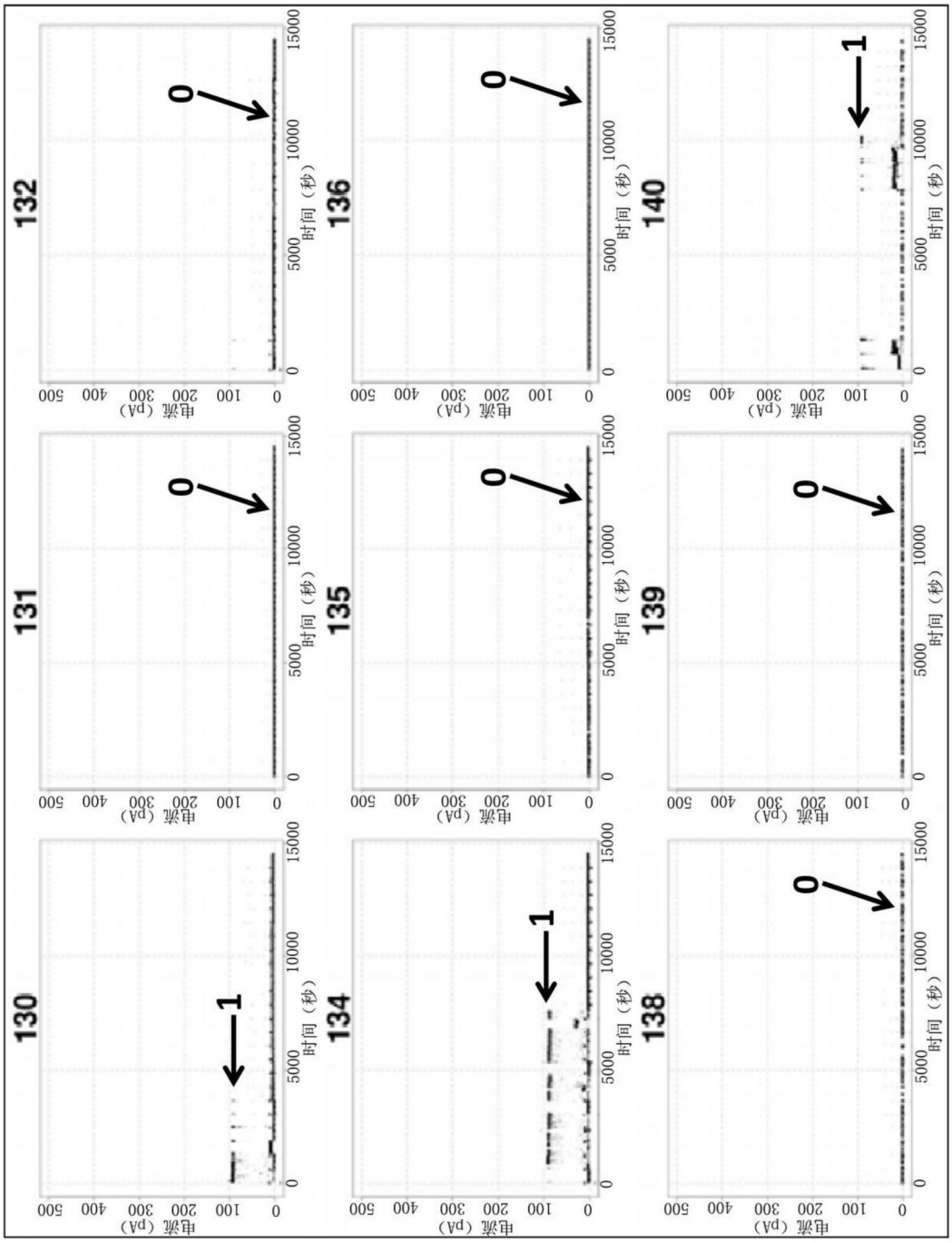


图20

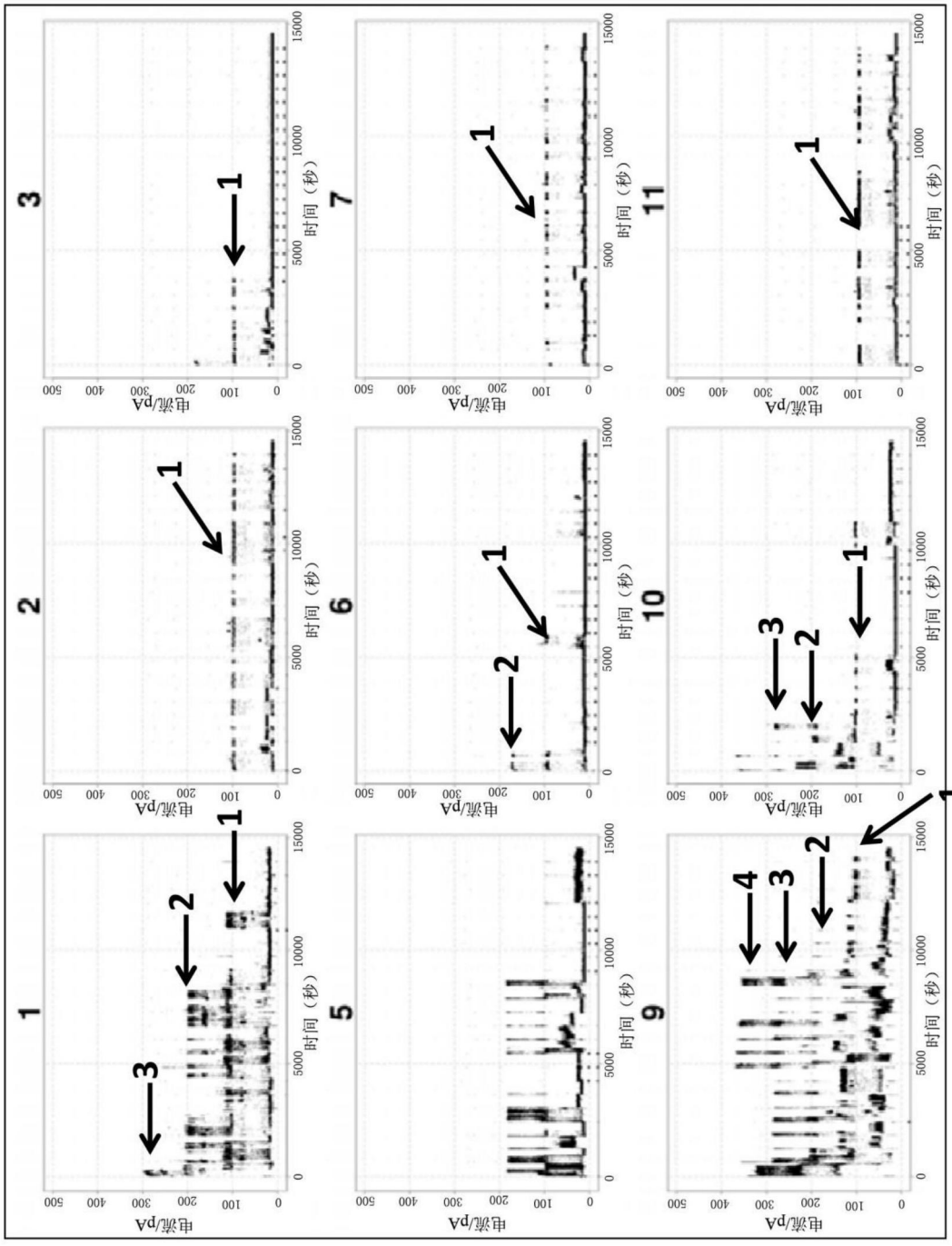


图21

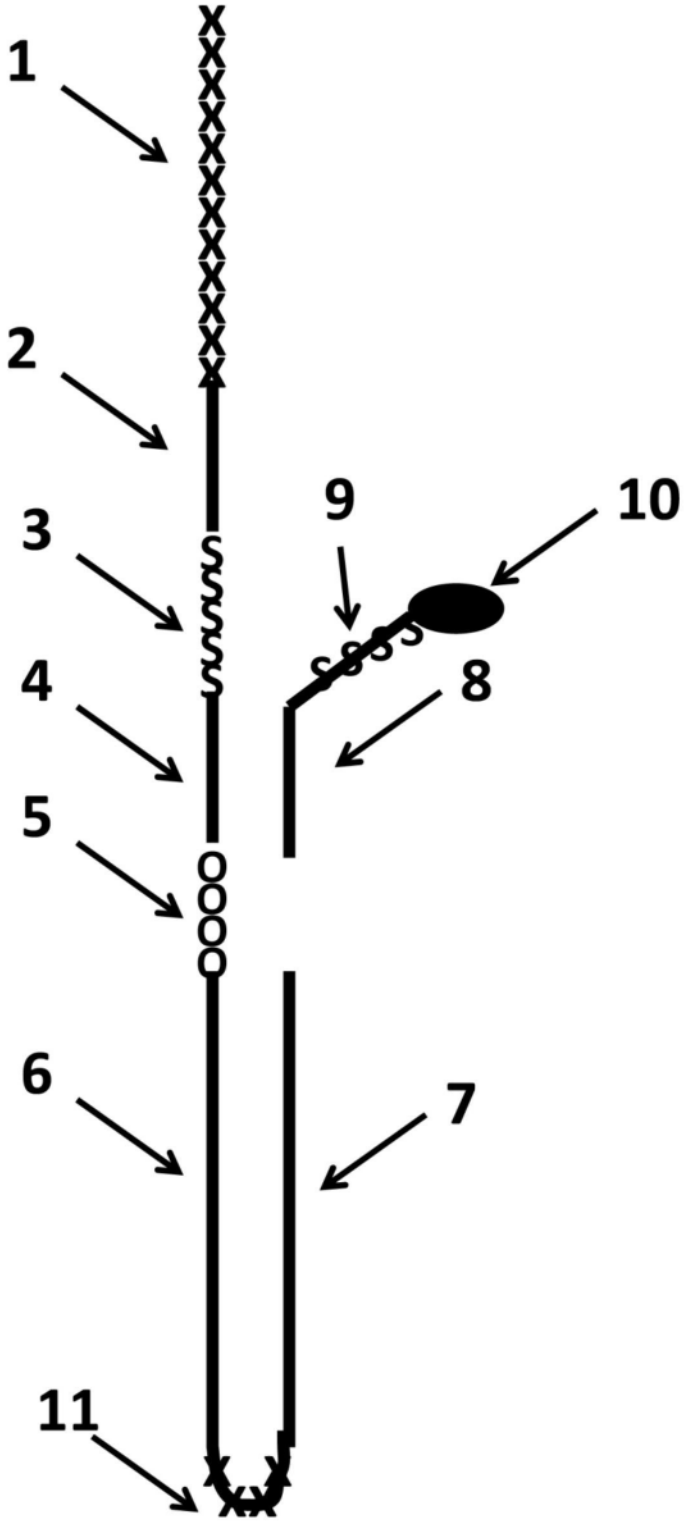


图22

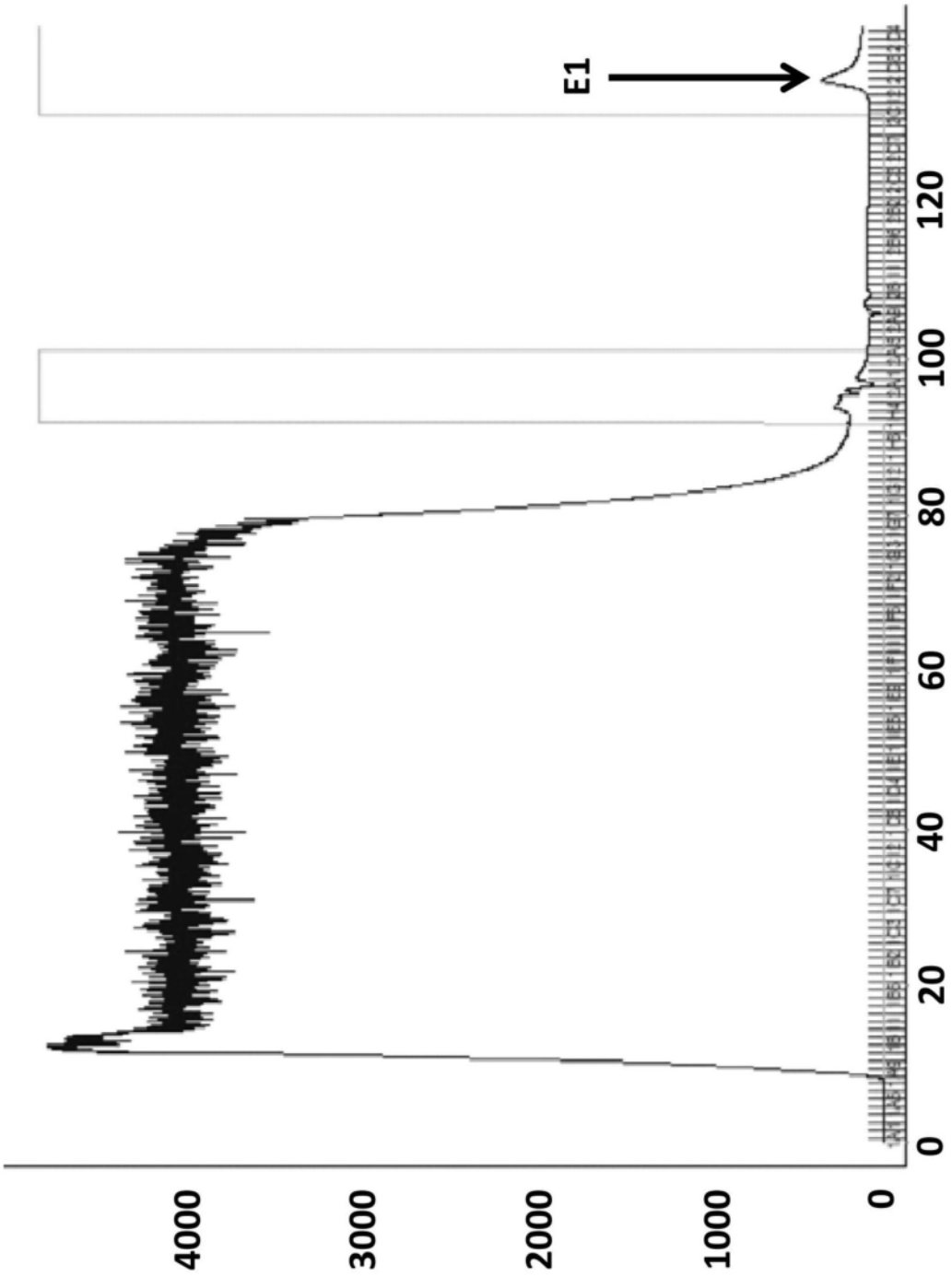


图23

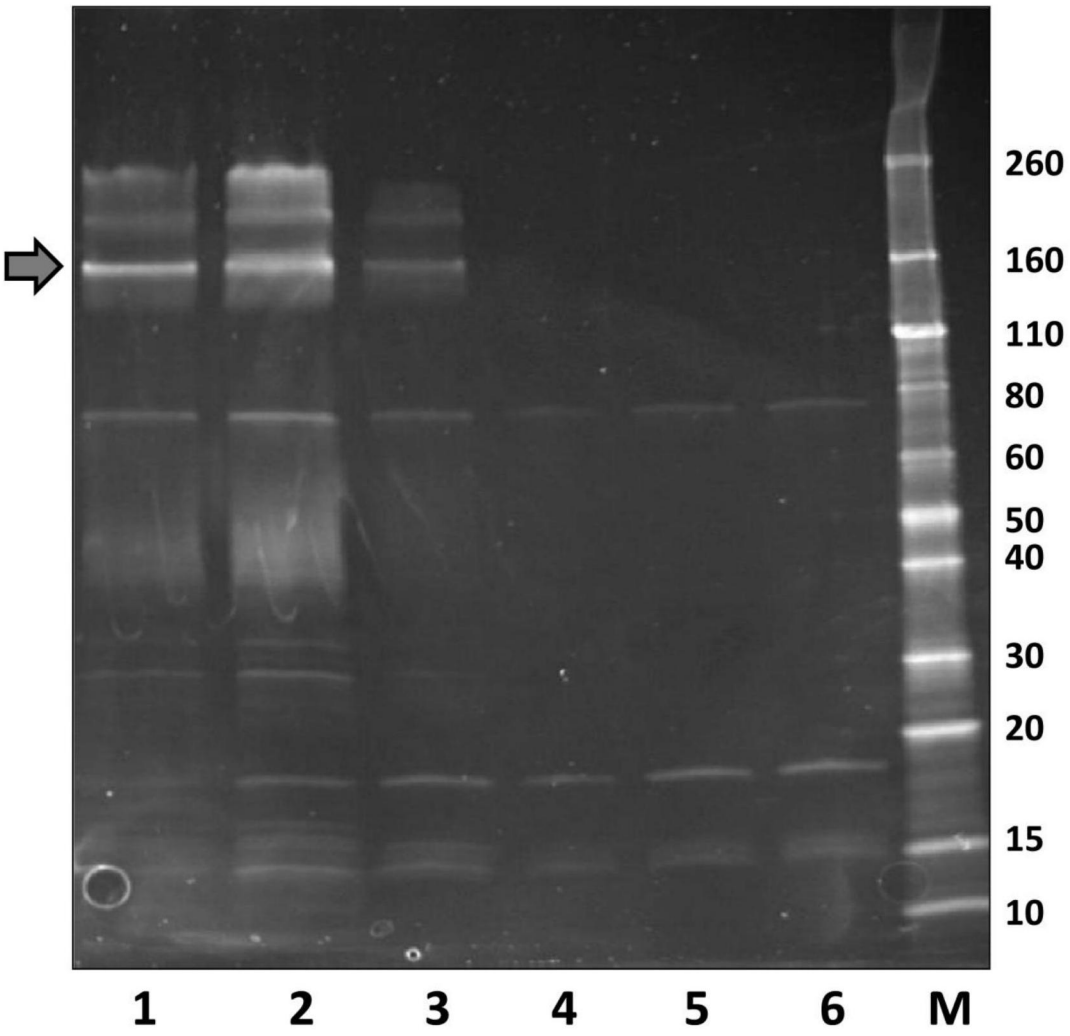


图24

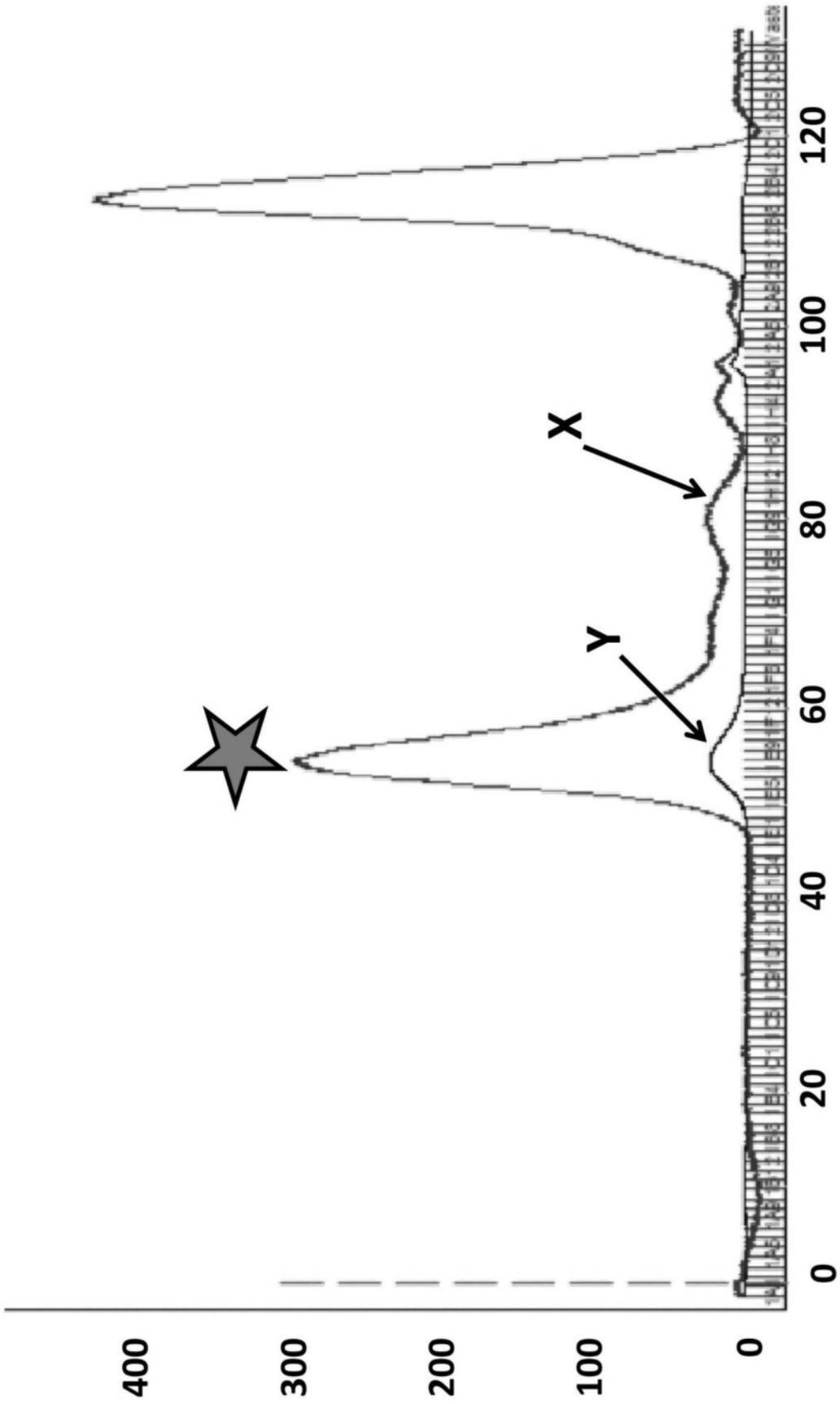


图25

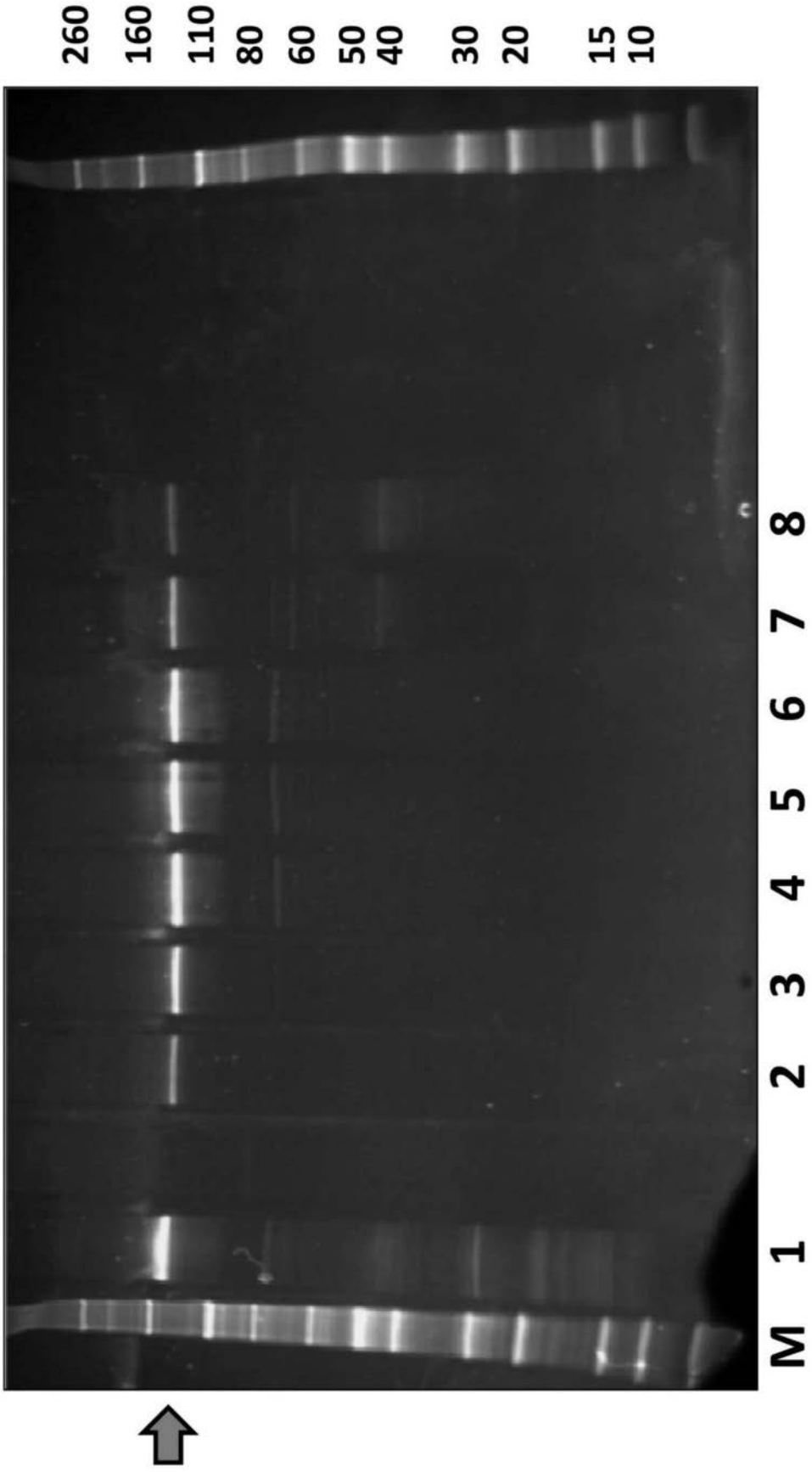


图26

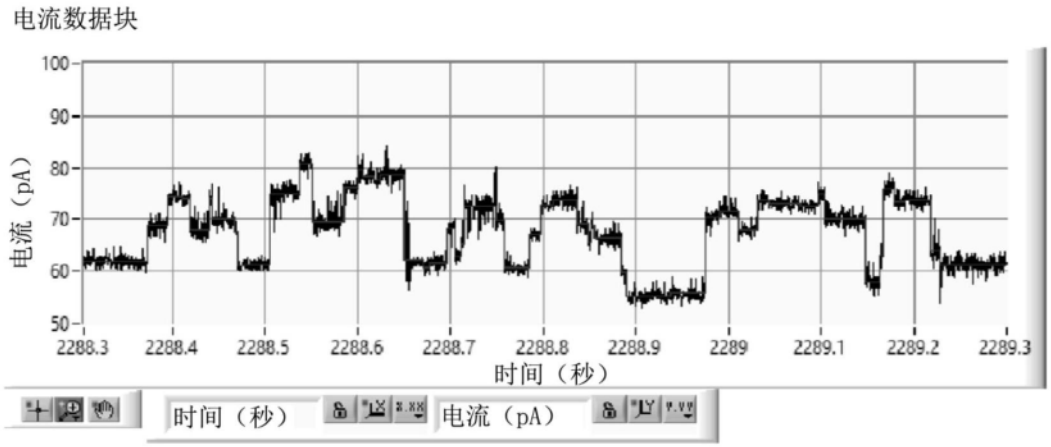

电流数据块
100
90
80
70
60
50
电流（pA）
2288.3
2288.4
2288.5
2288.6
2288.7
2288.8
2288.9
2289
2289.1
2289.2
2289.3
时间（秒）
时间（秒）
电流（pA）

图27

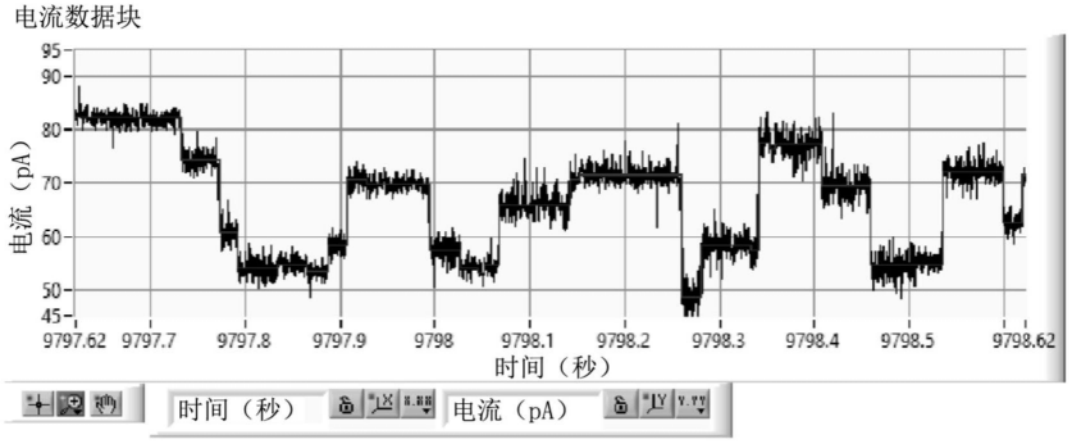

电流数据块
95
90
80
70
60
50
45
电流（pA）
9797.62
9797.7
9797.8
9797.9
9798
9798.1
9798.2
9798.3
9798.4
9798.5
9798.62
时间（秒）
时间（秒）
电流（pA）

图28

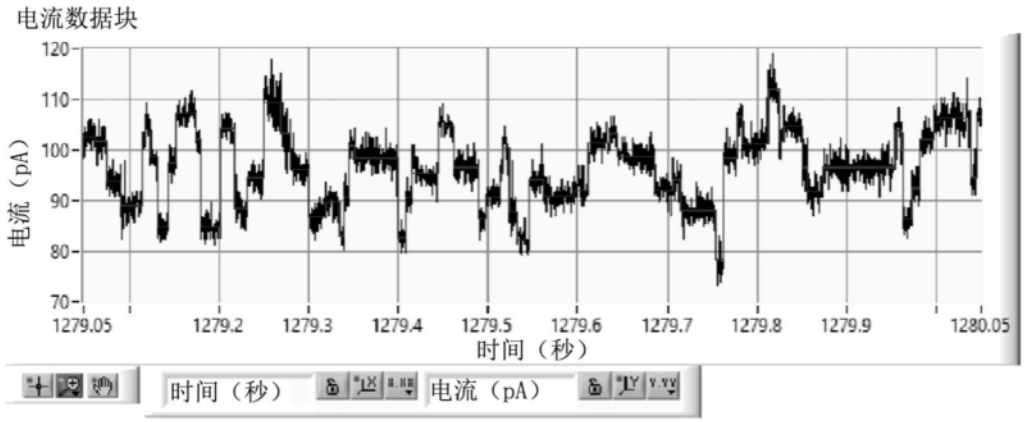

电流数据块
120
110
100
90
80
70
电流（pA）
1279.05
1279.2
1279.3
1279.4
1279.5
1279.6
1279.7
1279.8
1279.9
1280.05
时间（秒）
时间（秒）
电流（pA）

图29

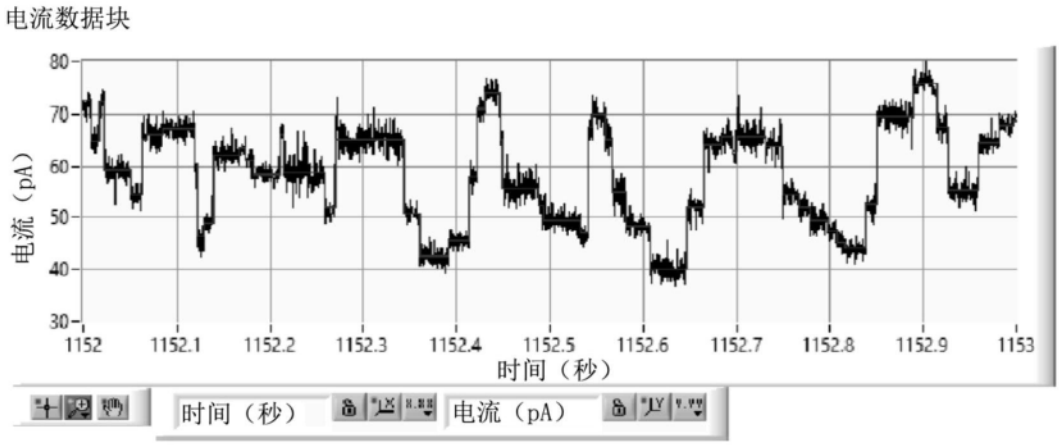

电流数据块
80
70
60
50
40
30
电流（pA）
1152
1152.1
1152.2
1152.3
1152.4
1152.5
1152.6
1152.7
1152.8
1152.9
1153
时间（秒）
时间（秒）
电流（pA）

图30

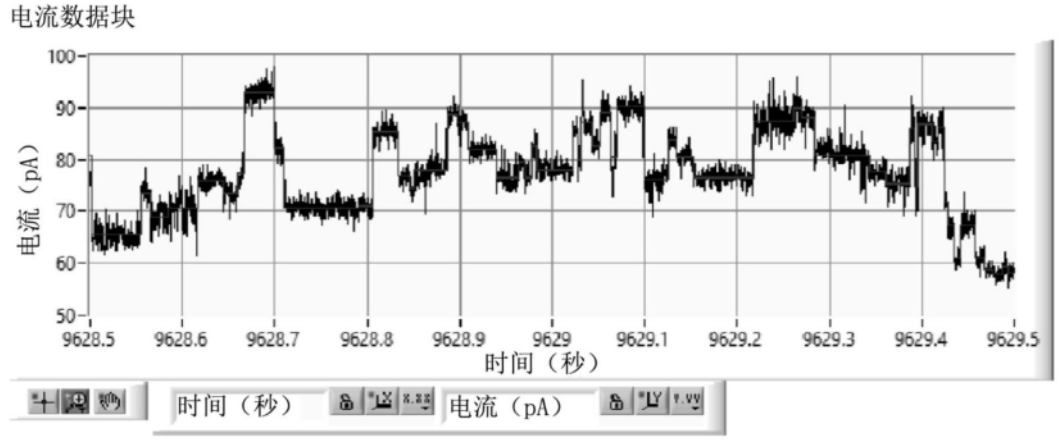

电流数据块
100
90
80
70
60
50
电流（pA）
9628.5
9628.6
9628.7
9628.8
9628.9
9629
9629.1
9629.2
9629.3
9629.4
9629.5
时间（秒）
时间（秒）
电流（pA）

图31

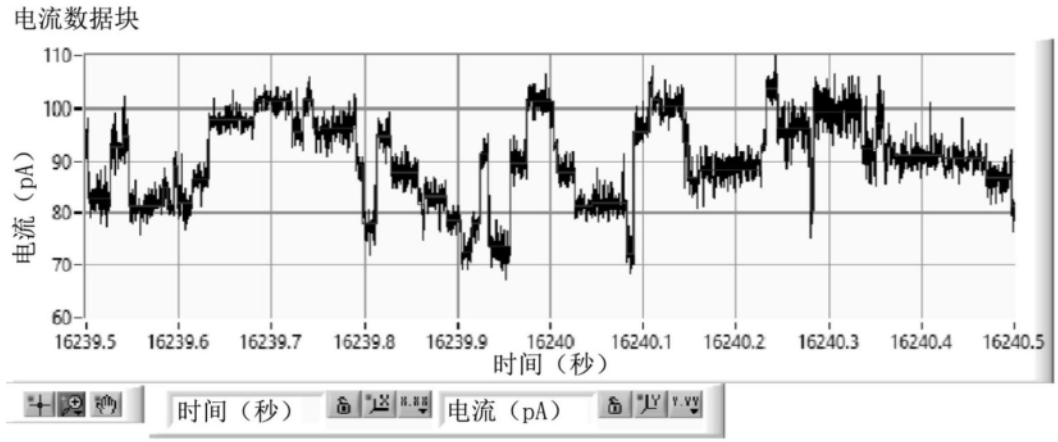

电流数据块
110
100
90
80
70
60
电流（pA）
16239.5
16239.6
16239.7
16239.8
16239.9
16240
16240.1
16240.2
16240.3
16240.4
16240.5
时间（秒）
时间（秒）
电流（pA）

图32

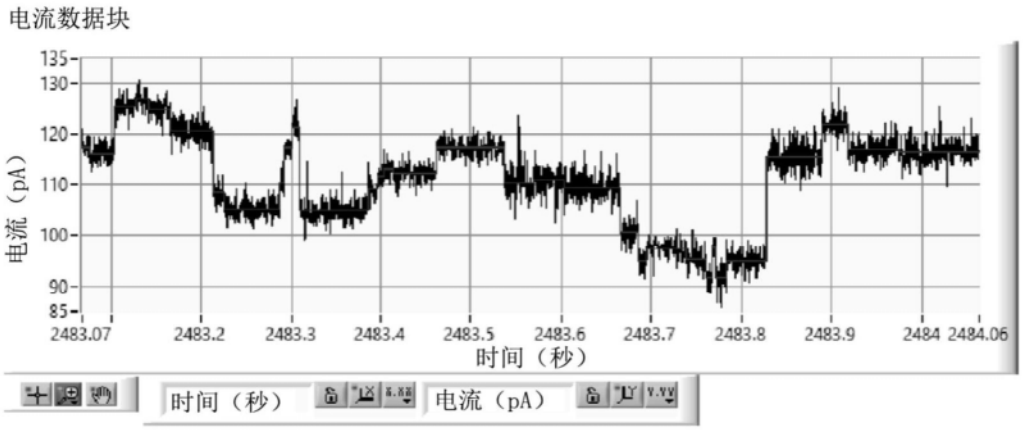


图33

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

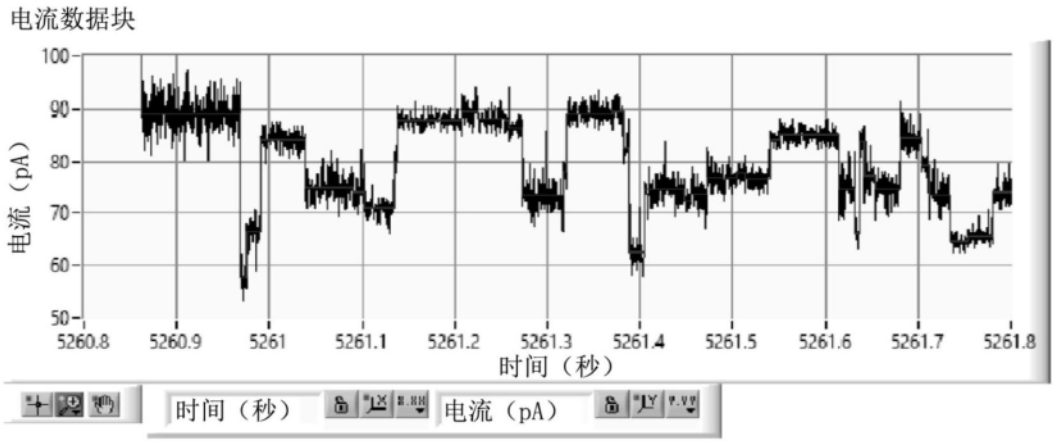


图34

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

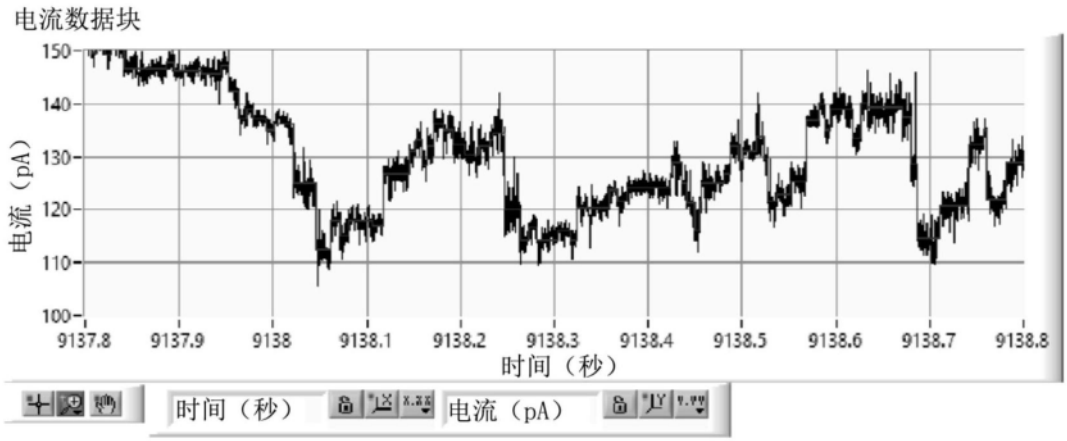


图35

CsgG-Eco-(Y51T/N55S/F56Q-StrepII(C))9

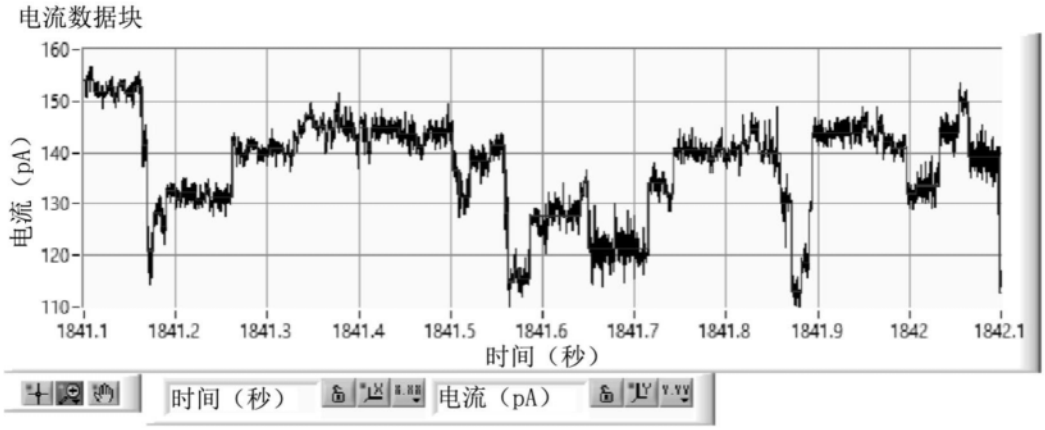


图36

CsgG-Eco-(F56Q/N102R-StrepII(C))9

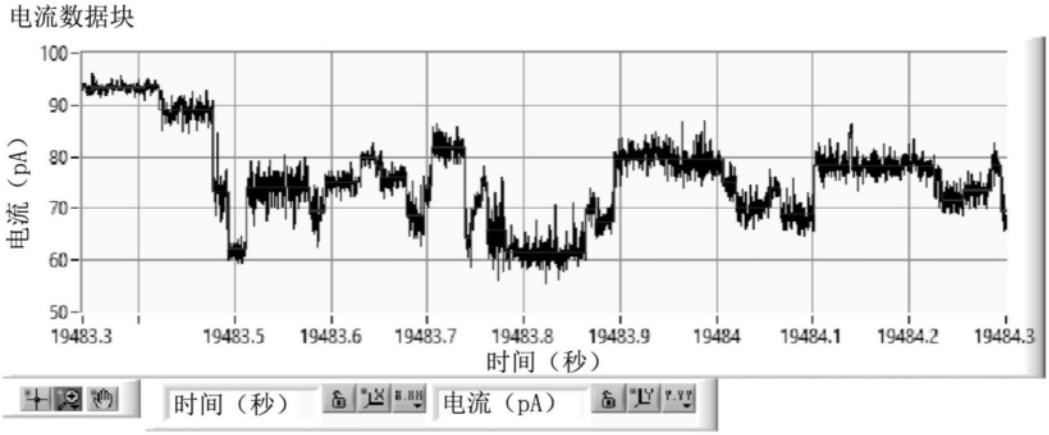


图37

Pro-Eco-CP1-(WT-Y51Q/F56Q)-Strep

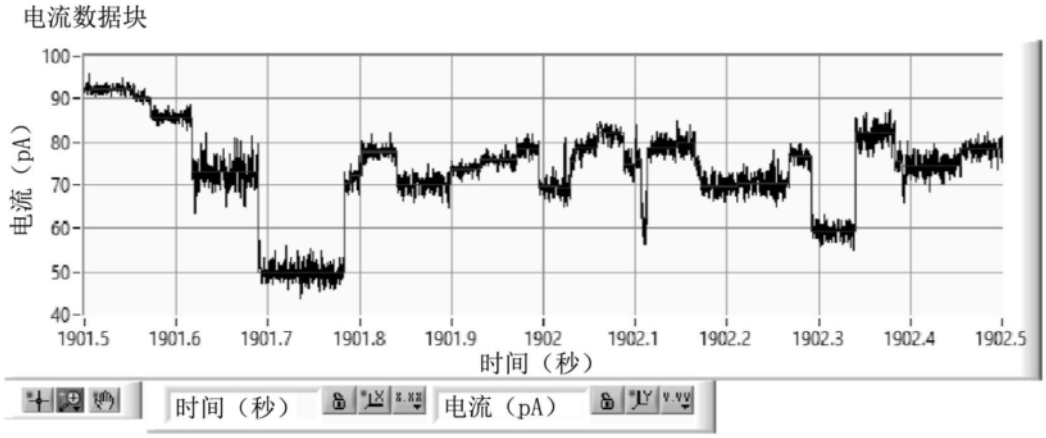


图38

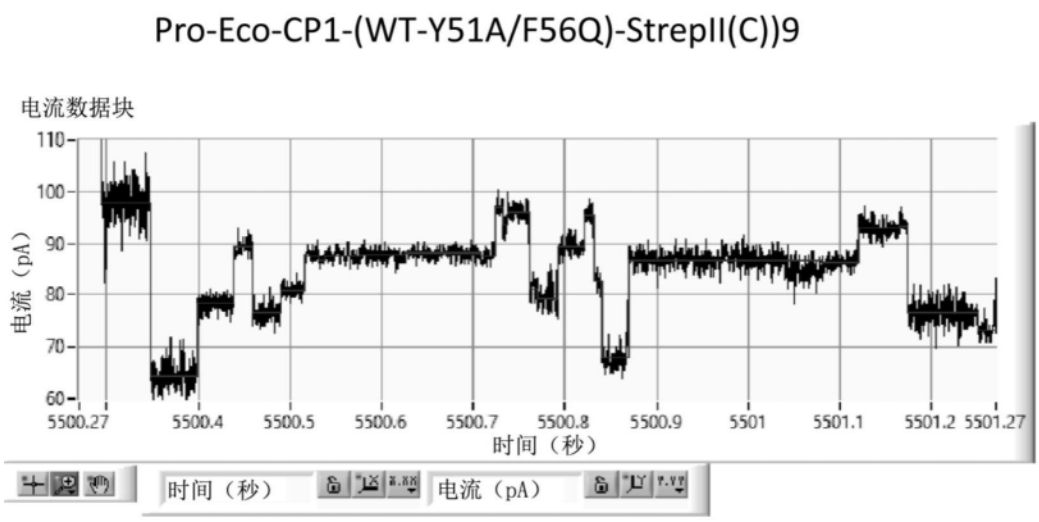

Pro-Eco-CP1-(WT-Y51A/F56Q)-StrepII(C))9
电流数据块
110
100
90
80
70
60
电流（pA）
5500.27
5500.4
5500.5
5500.6
5500.7
5500.8
5500.9
5501
5501.1
5501.2
5501.27
时间（秒）
时间（秒）
电流（pA）

图39

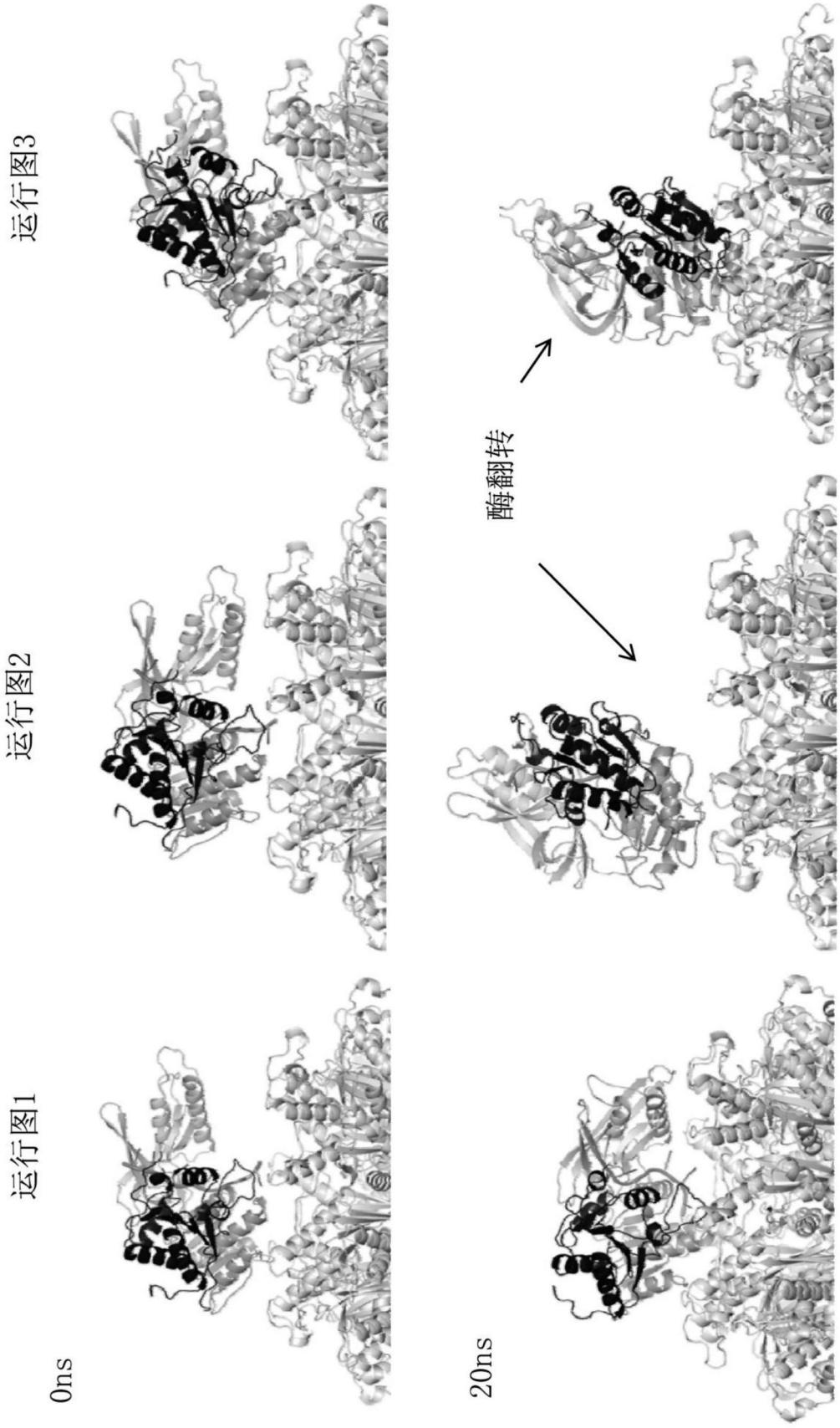


图40

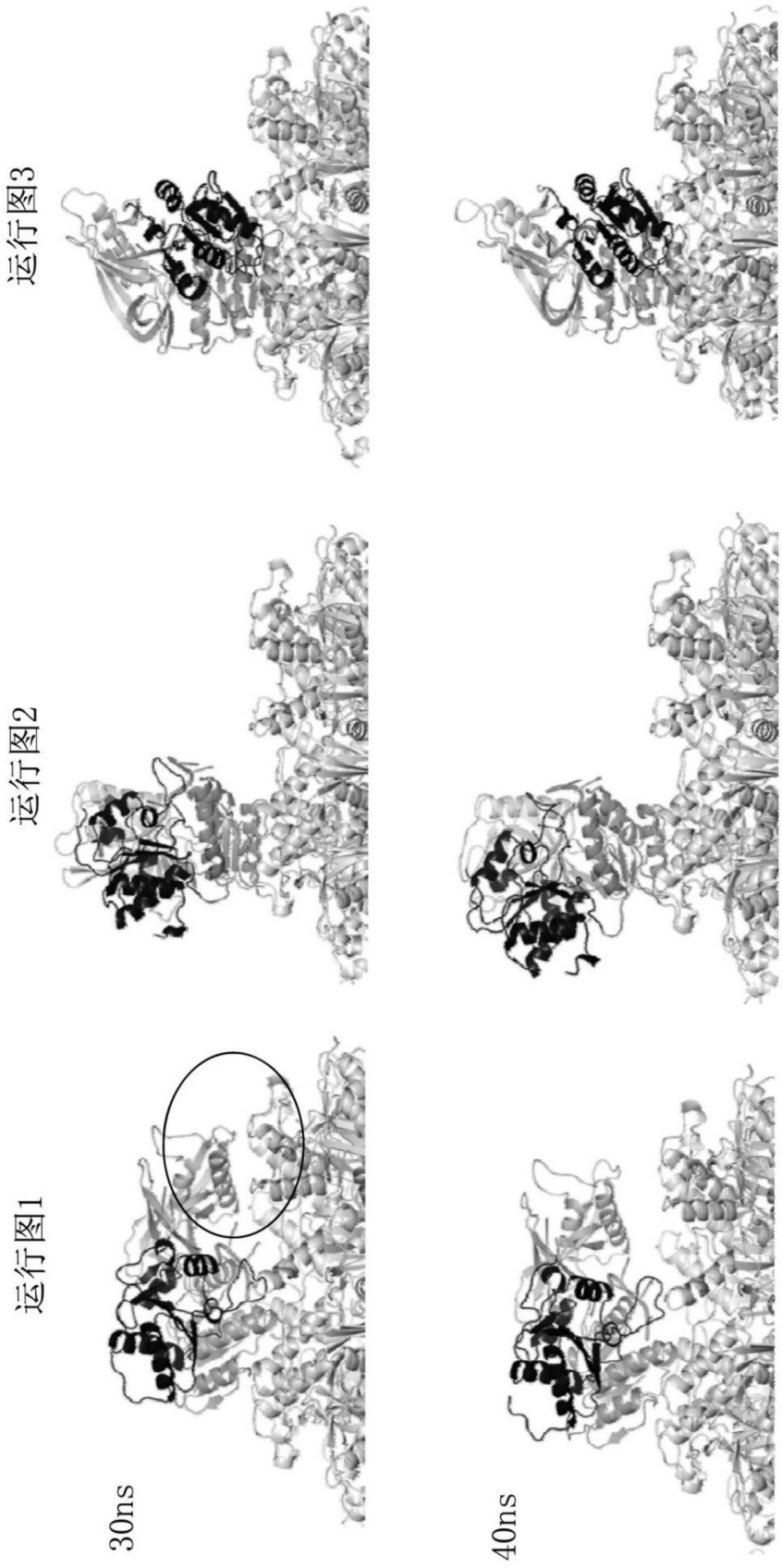


图41

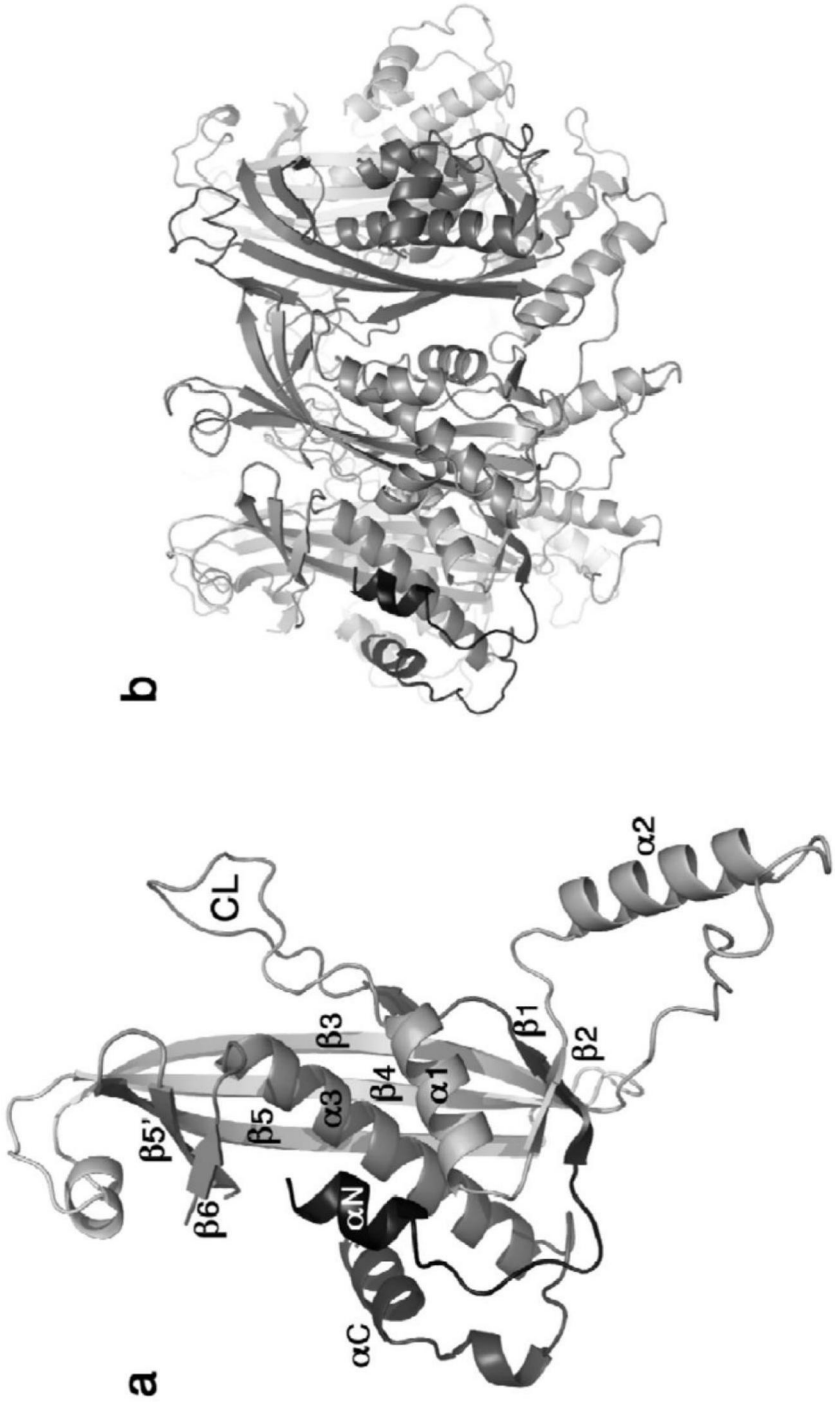


图42

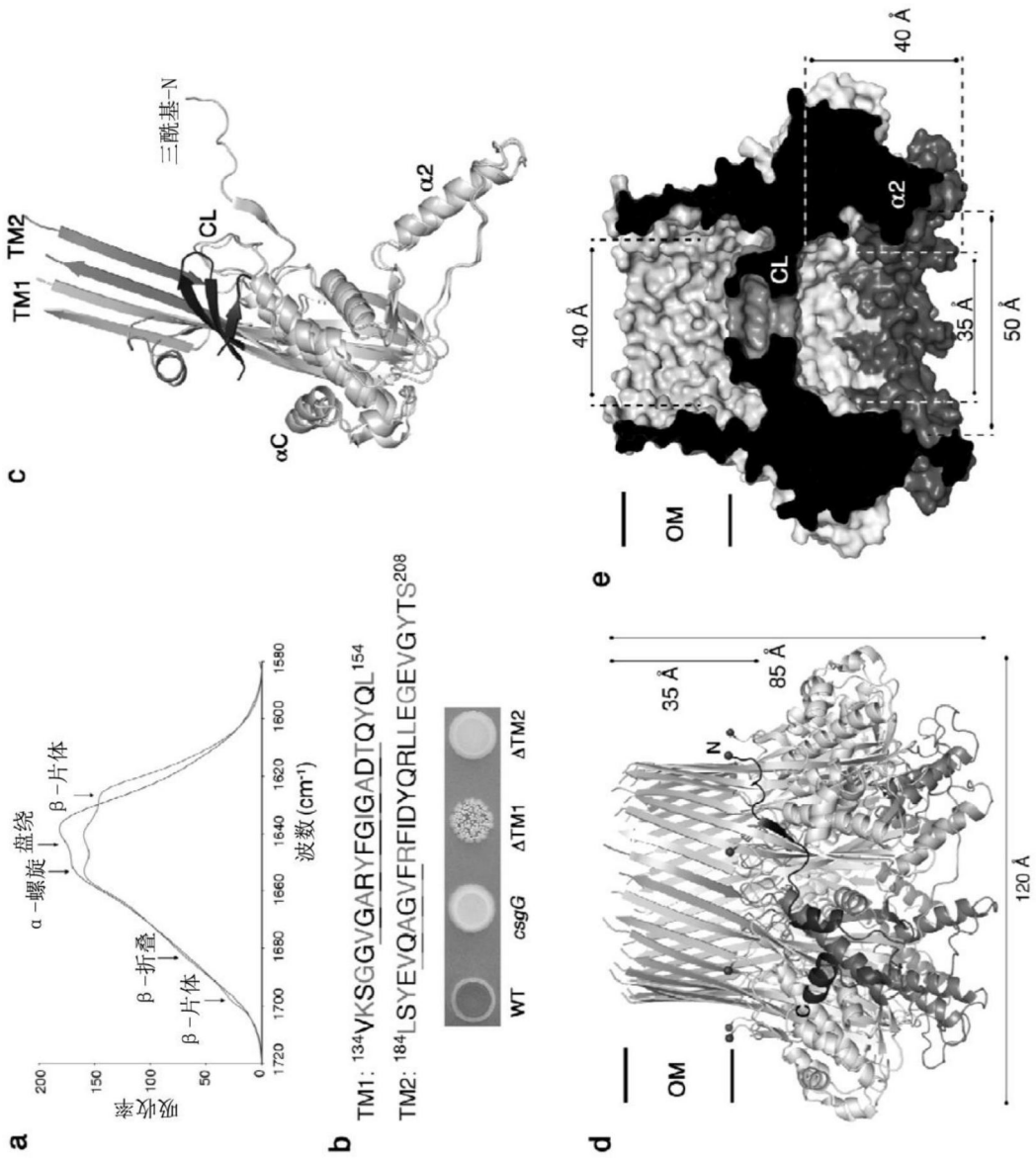

a
吸收率
波数(cm⁻¹)
α-螺旋
盘绕
β-片体
β-折叠
β-片体
b
TM1: 134VKSGGVGARYFGIGADTQYQL154
TM2: 184LSYEVQAGVFRFIDYQRLLEGEVGYTS208
WT
csgG
ΔTM1
ΔTM2
c
TM1
TM2
CL
三酰基-N
α2
αC
d
OM
35 Å
85 Å
N
C
120 Å
e
OM
40 Å
CL
α2
40 Å
35 Å
50 Å

图43

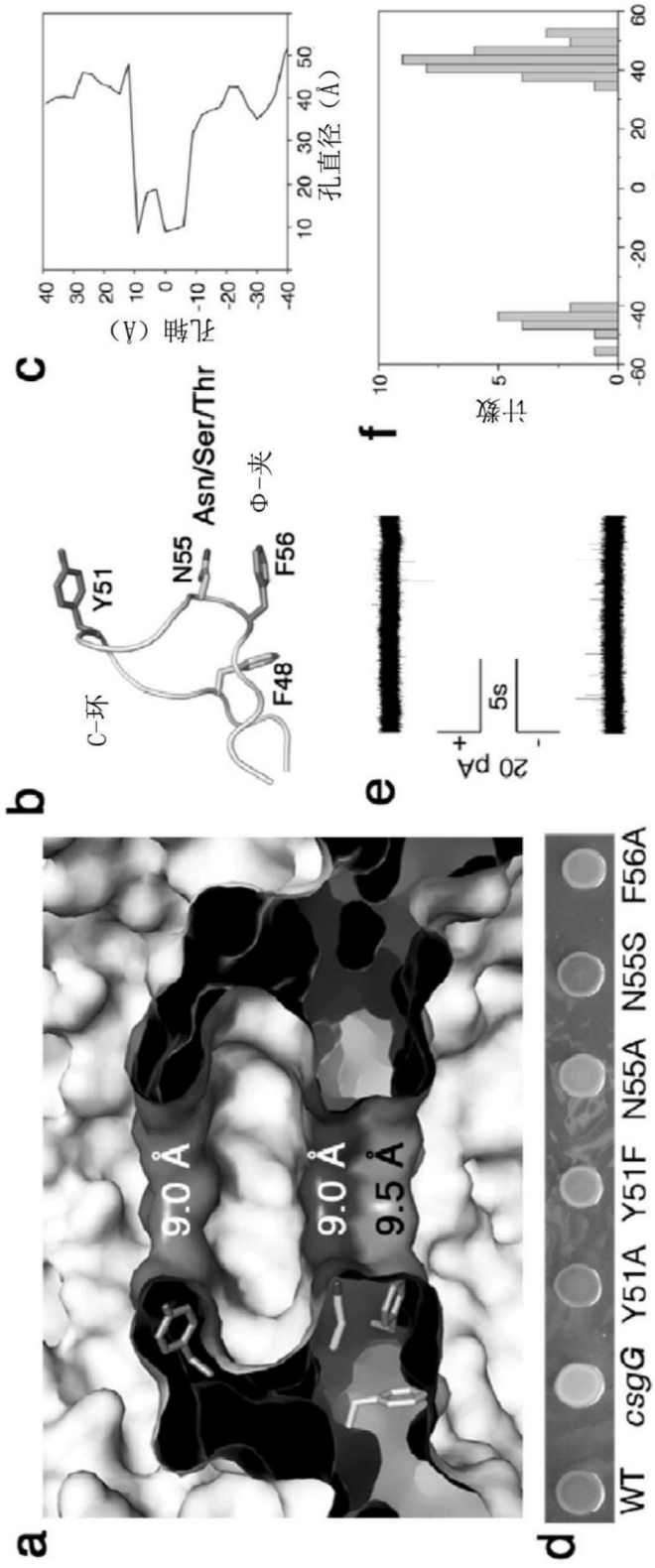

a
9.0 Å
9.0 Å
9.5 Å
b
C-环
Y51
N55
Asn/Ser/Thr
Φ-夹
F56
F48
c
孔轴（Å）
孔直径（Å）
d
WT
csgG
Y51A
Y51F
N55A
N55S
F56A
e
20 pA
5s
f
计数
pA

图44

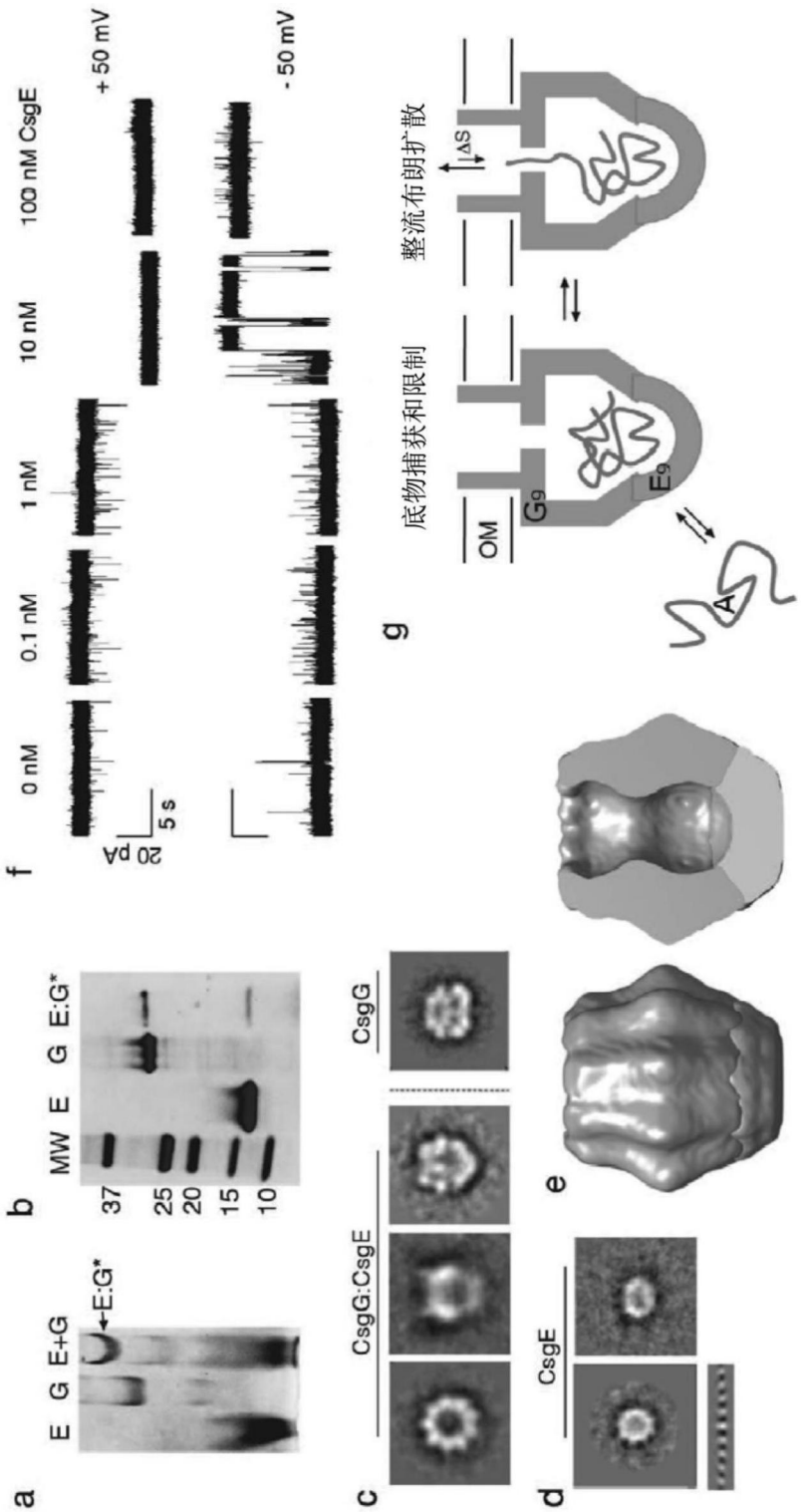

a
E G E+G
←E:G*
b
MW E G E:G*
37
25
20
15
10
c
CsgG:CsgE
CsgG
d
CsgE
e
f
0 nM
0.1 nM
1 nM
10 nM
100 nM CsgE
+ 50 mV
- 50 mV
20 pA
5 s
g
底物捕获和限制
整流布朗扩散
OM
G9
E9
ΔS
A

图45

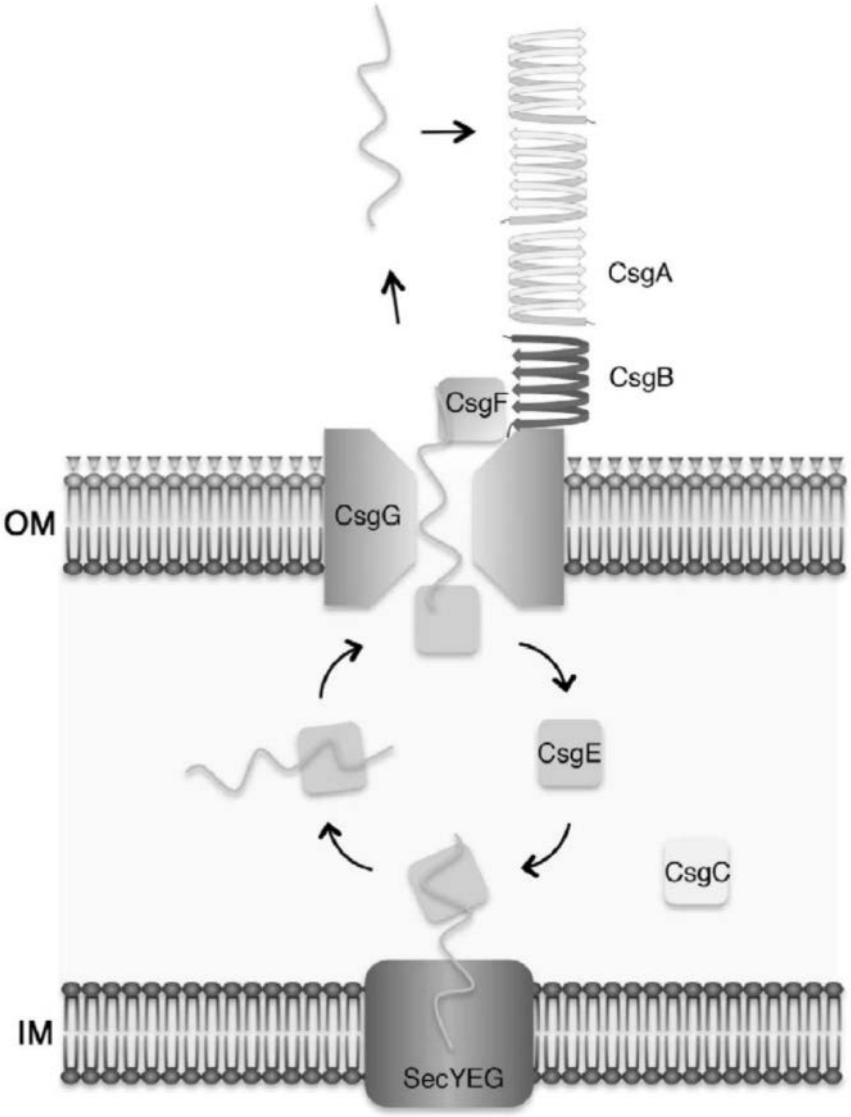


图46

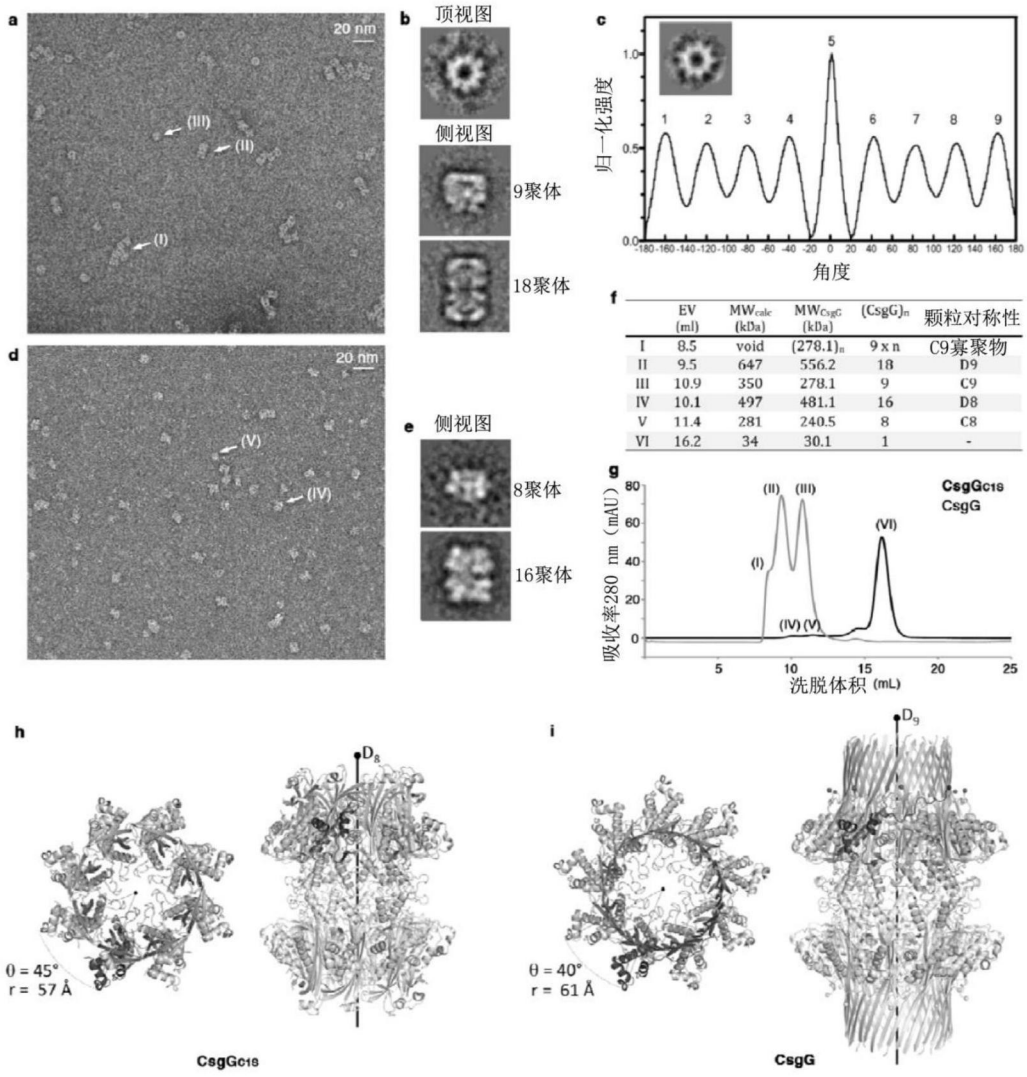

a
20 nm
(III)
(II)
(I)
b
顶视图
侧视图
9聚体
18聚体
c
归一化强度
角度
d
20 nm
(V)
(IV)
e
侧视图
8聚体
16聚体
f
EV (ml) | MWcalc (kDa) | MWCsgG (kDa) | (CsgG)n | 颗粒对称性
I | 8.5 | void | (278.1)n | 9 x n | C9寡聚物
II | 9.5 | 647 | 556.2 | 18 | D9
III | 10.9 | 350 | 278.1 | 9 | C9
IV | 10.1 | 497 | 481.1 | 16 | D8
V | 11.4 | 281 | 240.5 | 8 | C8
VI | 16.2 | 34 | 30.1 | 1 | -
g
CsgGc18
CsgG
吸收率280 nm（mAU）
洗脱体积 (mL)
h
θ = 45°
r = 57 Å
CsgGc18
i
θ = 40°
r = 61 Å
CsgG

图47

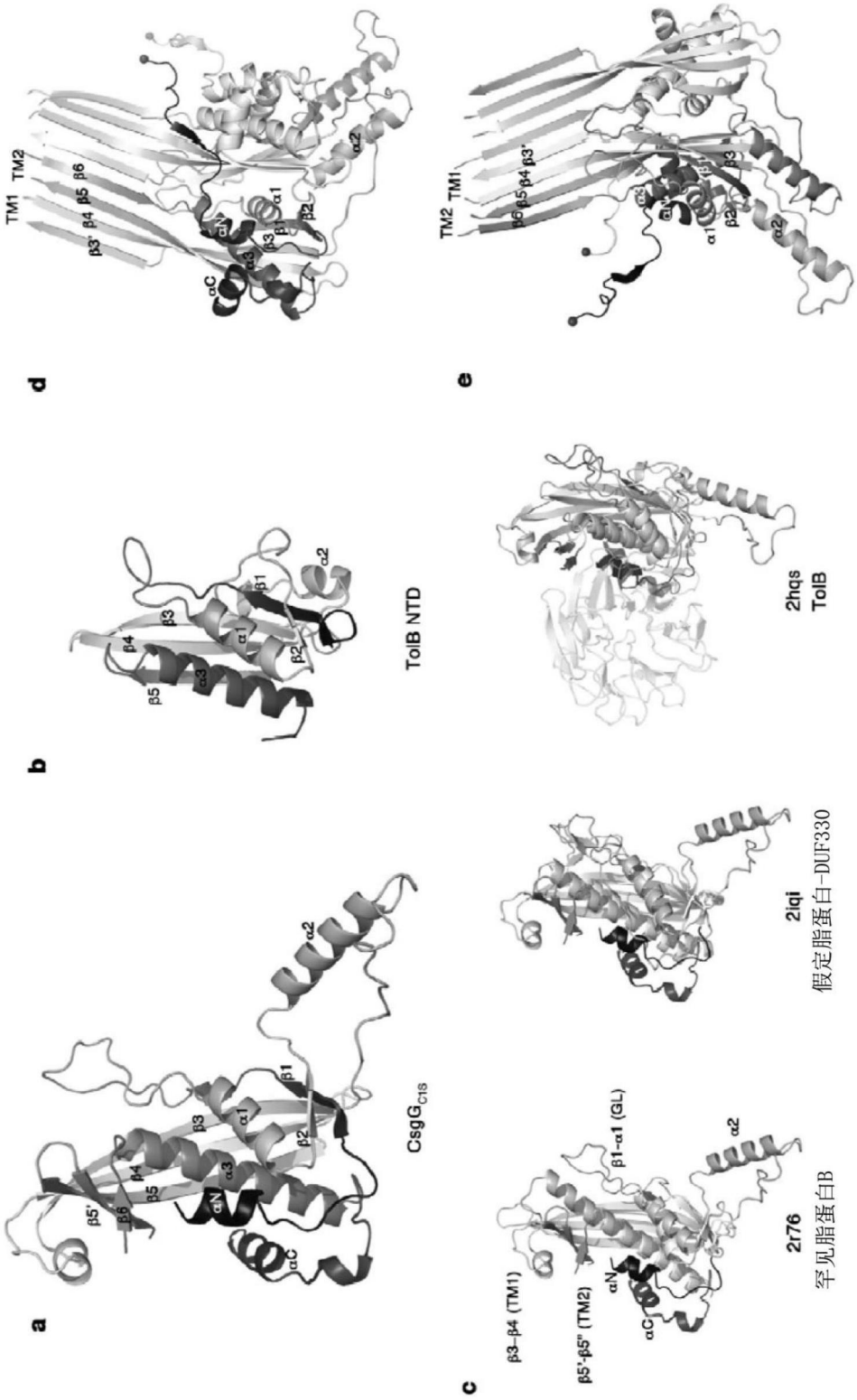

a
CsgG$_{C18}$
b
TolB NTD
c
2iqi
假定脂蛋白-DUF330
2r76
罕见脂蛋白B
d
e
2hqs
TolB

图48

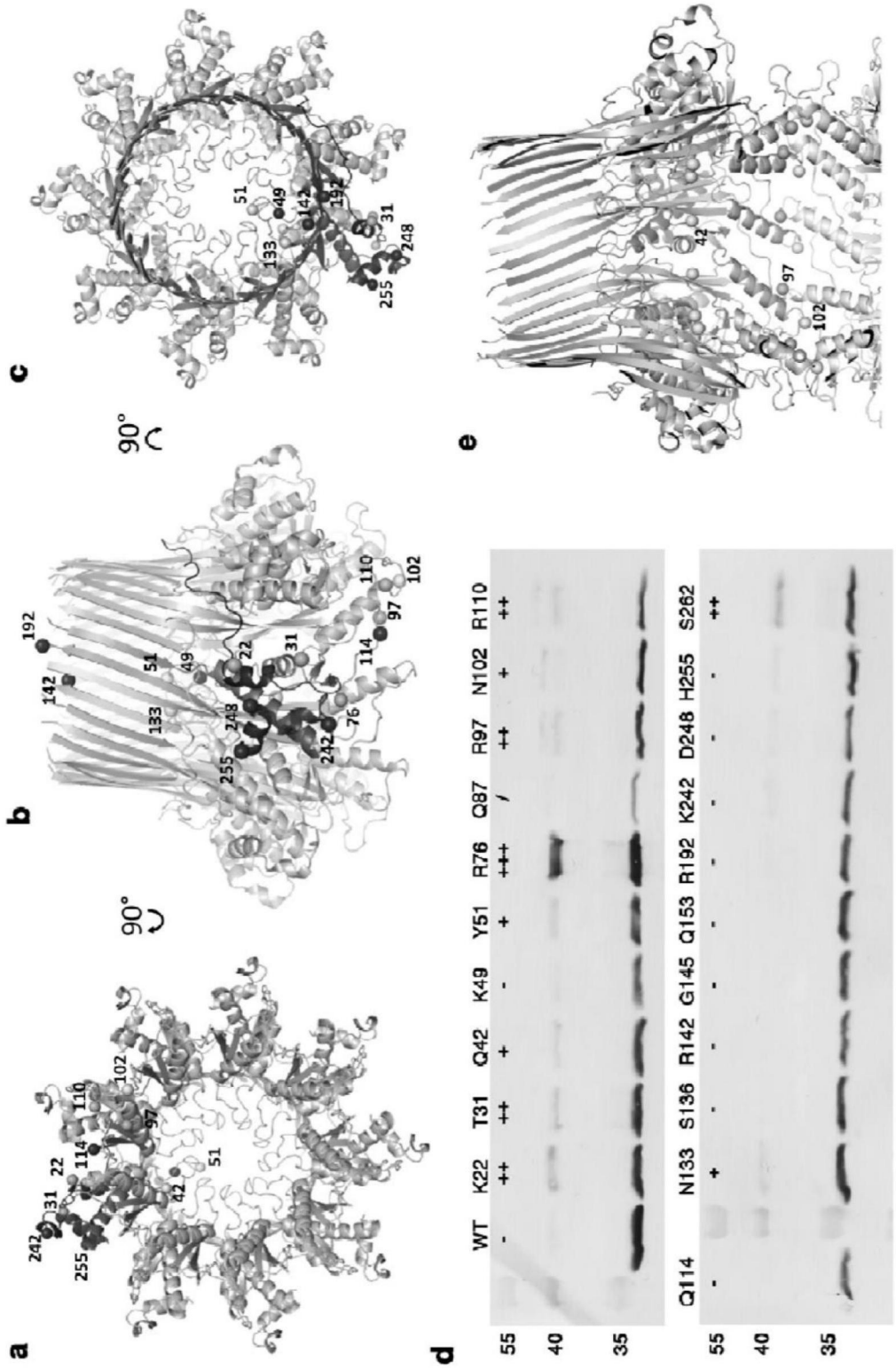

a
b
c
d
e
90°
90°
WT K22 T31 Q42 K49 Y51 R76 Q87 R97 N102 R110
Q114 N133 S136 R142 G145 Q153 R192 K242 D248 H255 S262
55
40
35

图49

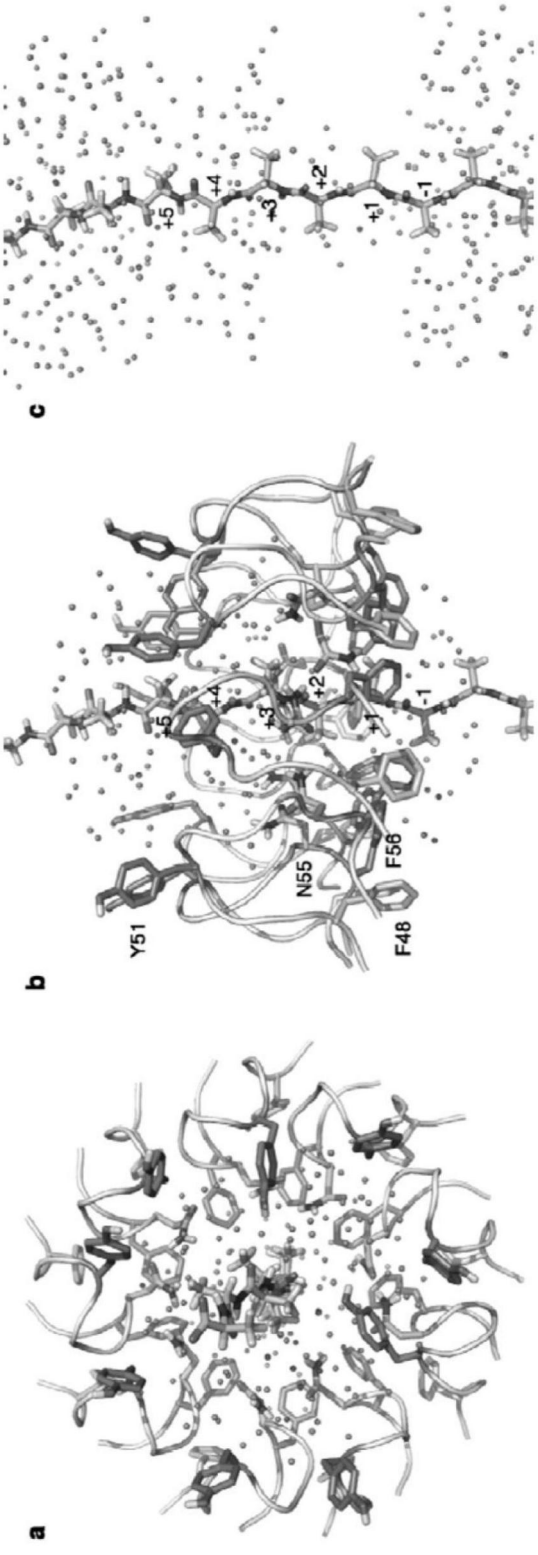


图50

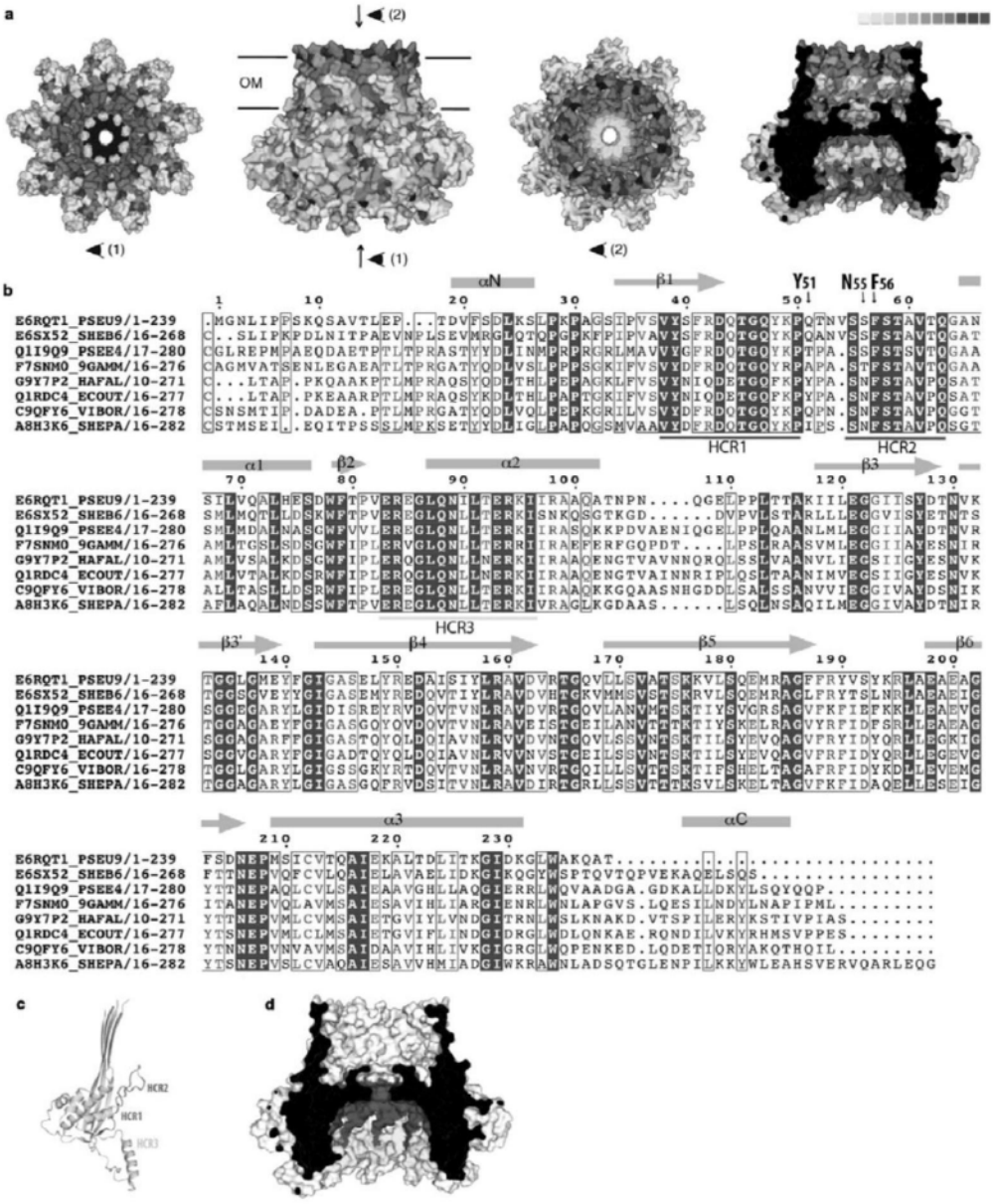

a
OM
(1)
(2)
b
αN
β1
Y51
N55 F56
HCR1
HCR2
α1
β2
α2
β3
HCR3
β3'
β4
β5
β6
αC
α3
E6RQT1_PSEU9/1-239
E6SX52_SHEB6/16-268
Q1I9Q9_PSEE4/17-280
F7SNM0_9GAMM/16-276
G9Y7P2_HAFAL/10-271
Q1RDC4_ECOUT/16-277
C9QFY6_VIBOR/16-278
A8H3K6_SHEPA/16-282
c
d

图51

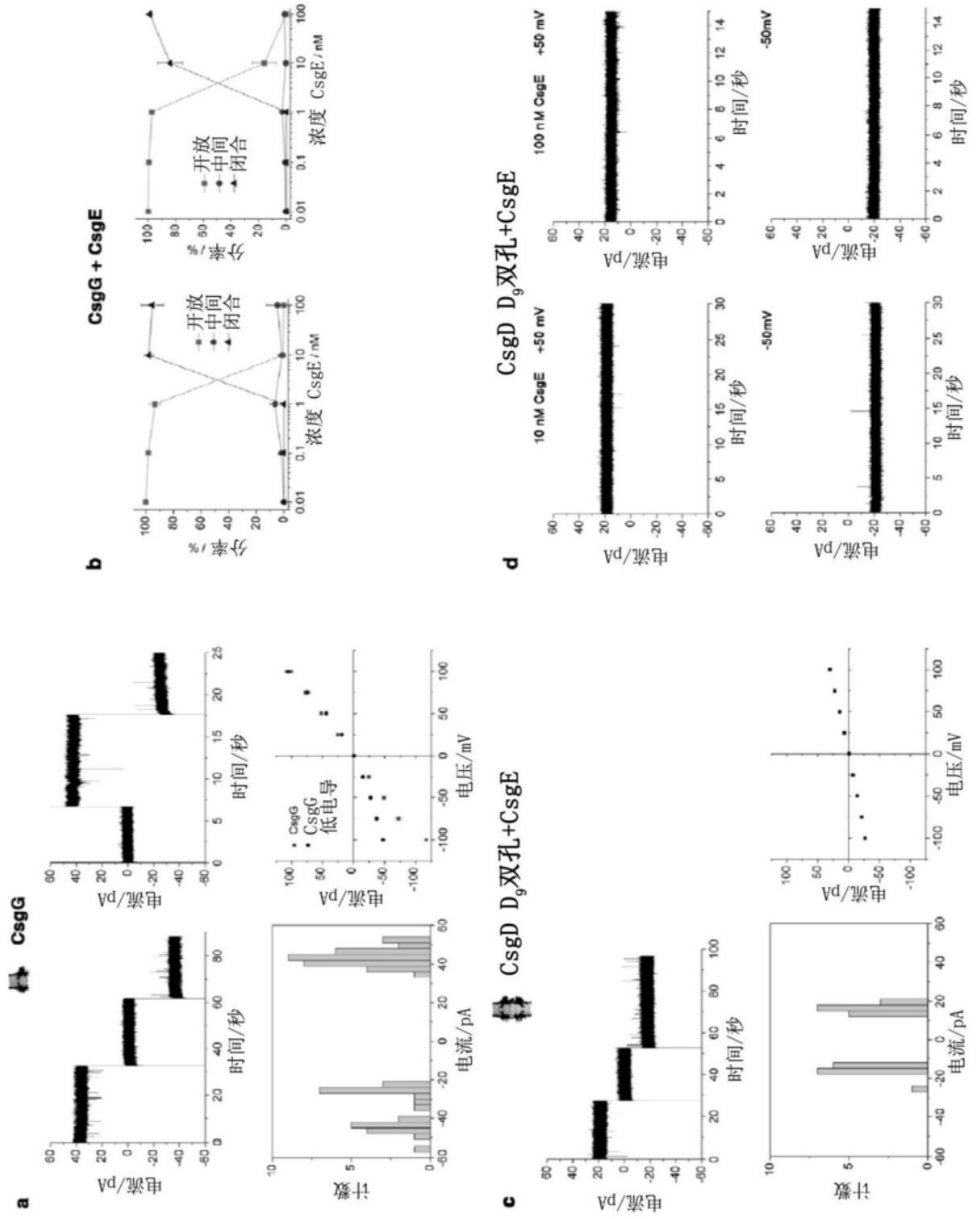

a
CsgG
电流/pA
时间/秒
计数
电流/pA
b
CsgG + CsgE
分率/%
开放
中间
闭合
浓度 CsgE/nM
c
CsgD D9双孔+CsgE
d
CsgD D9双孔+CsgE
10 nM CsgE +50 mV
100 nM CsgE +50 mV
-50mV
电压/mV
CsgG
CsgG
低电导

图52

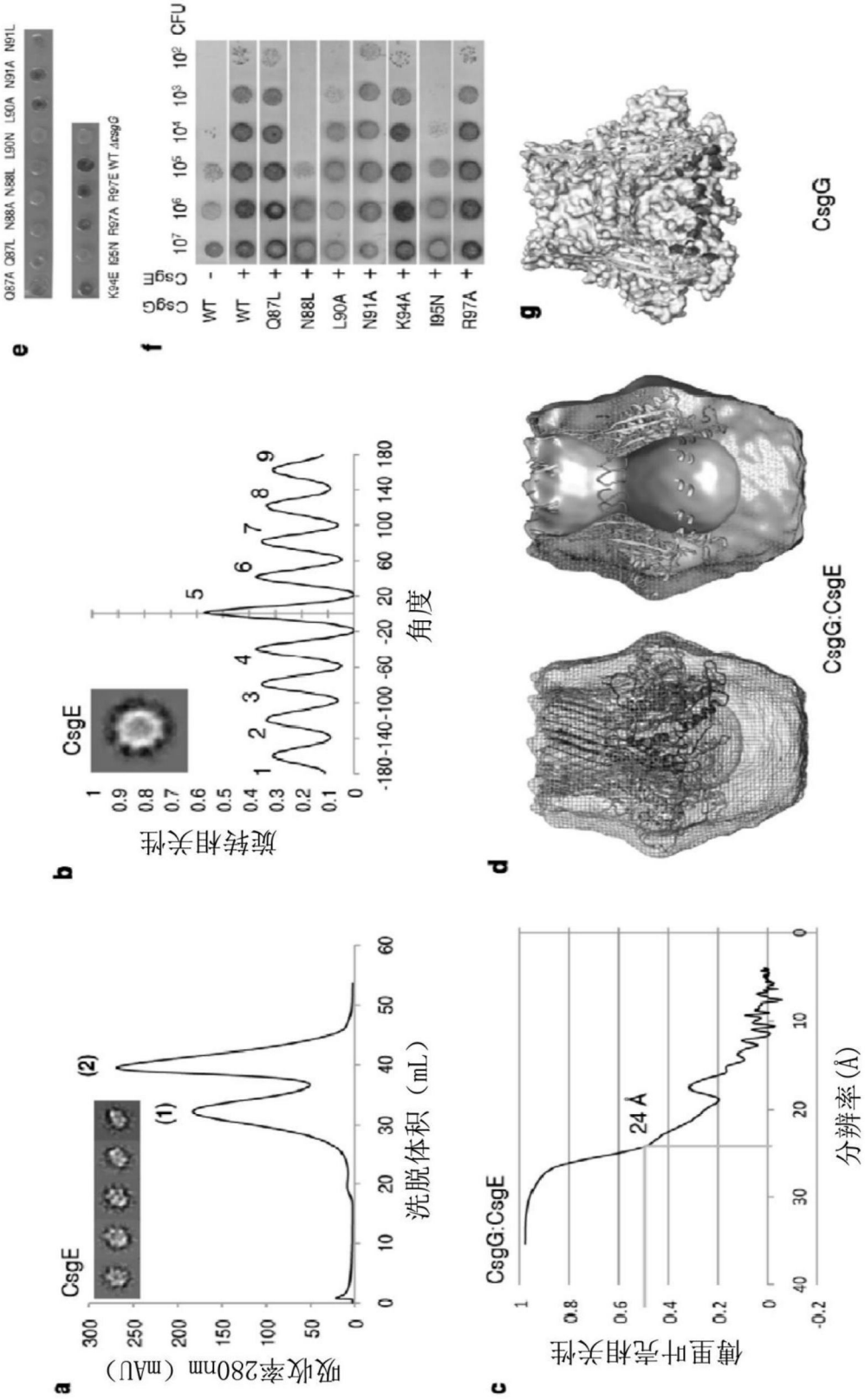

a
CsgE
吸收率280nm（mAU）
洗脱体积（mL）
(1)
(2)
b
CsgE
旋转相关性
角度
c
CsgG:CsgE
衍里叶壳相关性
24 Å
分辨率(Å)
d
CsgG:CsgE
e
Q67A Q67L N88A N88L L90N L90A N91A N91L
K94E I95N R97A R97E WT ΔcsgG
f
csgG
WT - WT + Q67L + N88L + L90A + N91A + K94A + I95N + R97A +
CFU
g
CsgG

图53

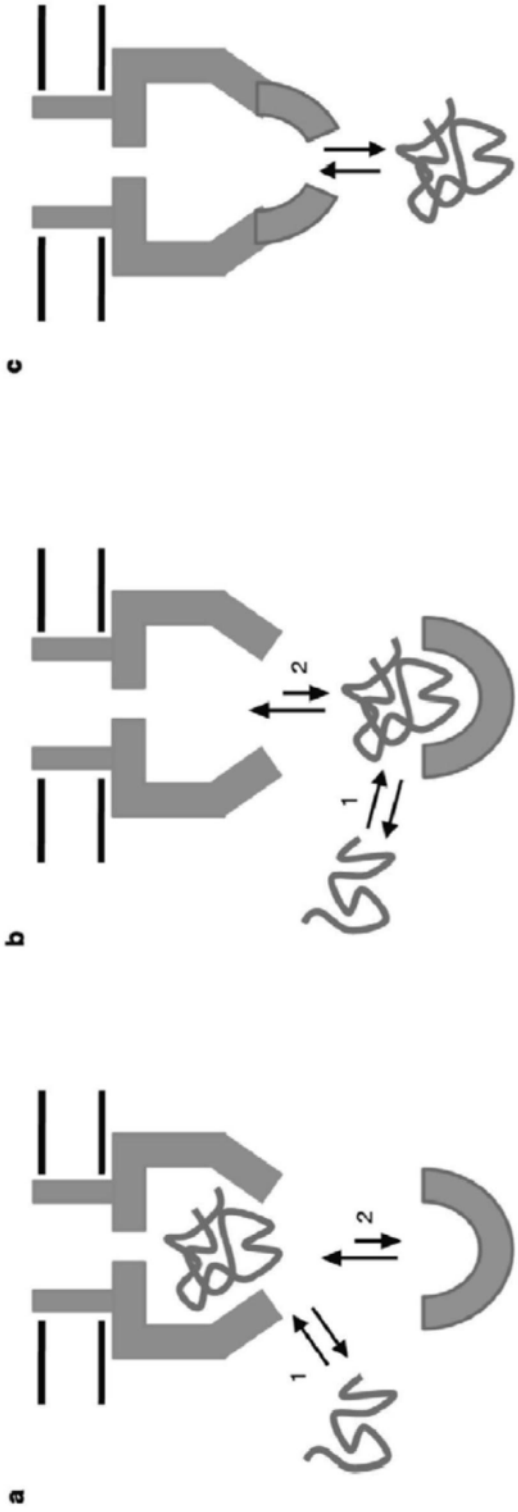


图54

a

数据收集和细化统计

	$CsgG_{CIS}$	CsgG
数据收集		
空间群	P1	C2
细胞大小		
a, *b*, *c* (Å)	101.3, 103.6, 141.7	161.9, 372.8, 161.9
α, *β*, *γ* (°)	111.3, 90.5, 118.2	90.0, 92.9, 90.0
分辨率 (Å)*	30-2.8 (2.9-2.8)	30-3.6 (3.7-3.6)
		30 – 3.6 (a*), –3.7 (b*), –3.8 (c*) †
*R*meas *	15.1 (81.8)	16.2 (90.6) †
$I/\sigma I$ *	9.82 (2.03)	6.80 (1.89) †
完整度(%)*	98.7 (98.3)	91.57 (27.26)
		99.9 (99.1) †
冗余*	11.2 (7.0)	4.4 (4.3)
Wilson B ($Å^2$)	46.7	101.0
细化		
分辨率 (Å)*	30-2.8 (2.9-2.8)	30-3.6 (3.7-3.6)
		30 – 3.6 (a*), –3.7 (b*), –3.8 (c*) †
反射数量*	112419 (11159)	102130 (11094)
R_{work}/R_{free}	0.1881 / 0.2337	0.3024 / 0.3542
原子数量		
蛋白	28853	34165
配体/离子	0	0
水	0	0
B-因子 ($Å^2$)		
蛋白	57.3	116.7
配体/离子		
水		
R. m. s. 偏差		
键长 (Å)	0.01	0.03
键角 (°)	1.31	1.87

$CsgG_{CIS}$和膜提取的CsgG(均从单个晶体收集）的数据统计；

*最高分辨率的壳在括号中示出；

†沿互逆方向a*、b*和c*的各向异性截断的校正值；

b

c

图55

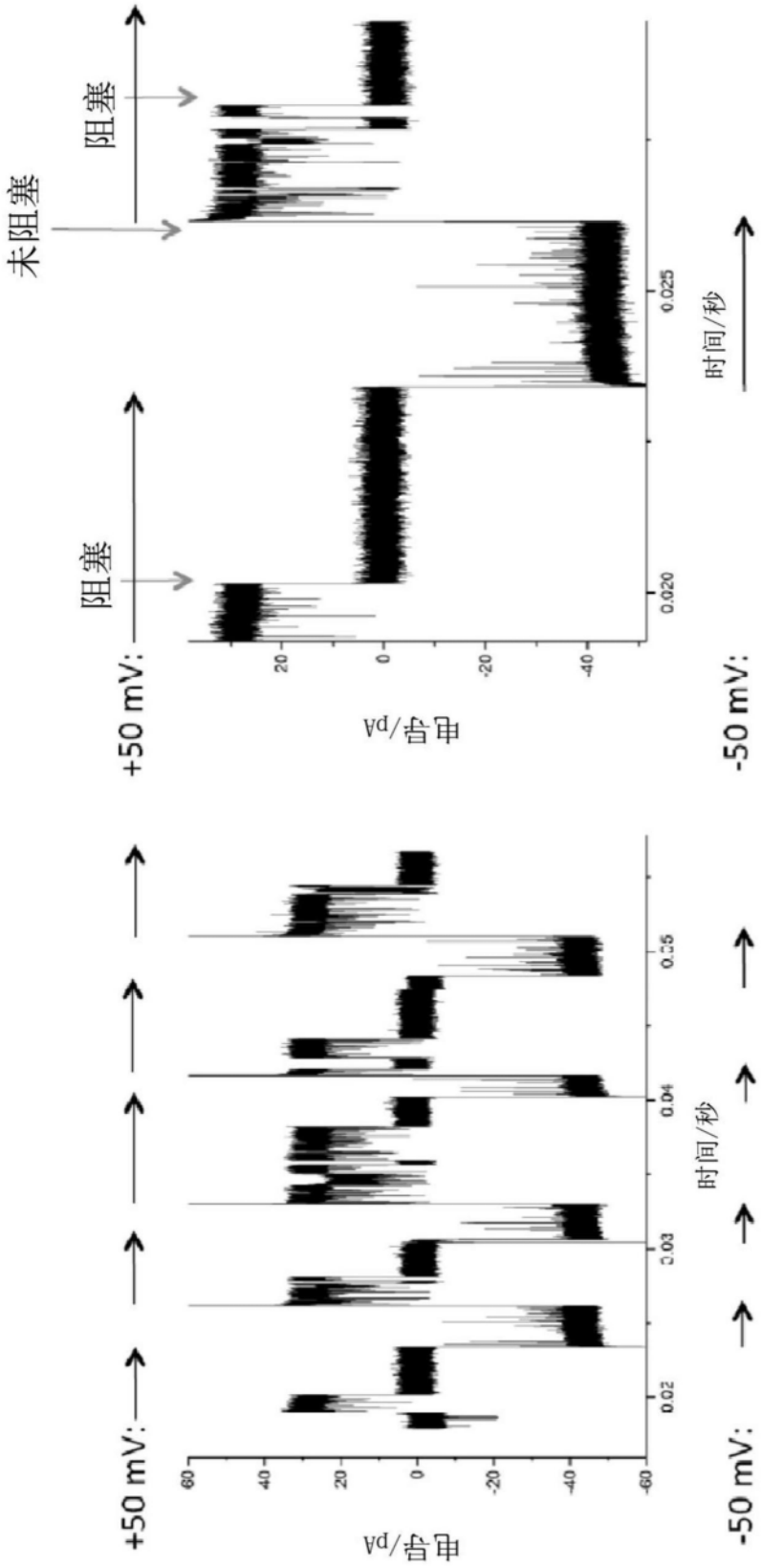

未阻塞
阻塞
阻塞
+50 mV:
-50 mV:
电导/pA
时间/秒
20
0
-20
-40
0.020
0.025
+50 mV:
-50 mV:
电导/pA
时间/秒
60
40
20
0
-20
-40
-60
0.02
0.03
0.04
0.25

图56

CsgG-△PYPA 的纯化和通道性质，
CsgG中P50Y51P52A53突变成GG
实验
日期：2015年4月7日
目标：CsgG收缩机构突变体的纯化
生产条件
菌株：BL21(DE3)细胞的PASK-IBA3+(amp)
中的CsgG-△PYPA C环突变
培养基：TB
诱导：OD600=0.6时O/N在 25° 下，
用无水四环素
缓冲液：50mM Tris pH 8.0,
150mM NaCl，1mM DTT，1mM EDTA,
0.1mg AEBSF/ml缓冲液，脱氧核糖核酸酶,
0.1mg/ml溶菌酶，亮肽酶素
链球菌色谱法
上样缓冲液：25mM Tris pH 8,
500mM NaCl，1mM EDTA，1mM DTT
洗脱缓冲液：100mM Tris pH 8.0,
500mM NaCl，1mM DTT，2.5mM脱硫生物素，
5mM LDAO，0.5% C8E4
蛋白印迹
抗链球菌
M: PageRuler预染蛋白阶梯
M
细胞溶解产物
1%肌氨酸洗涤
0.002% DDM洗涤
1%增溶缓冲液O/N
可溶性部
流经运行图1的抗链球菌柱
洗脱馏分
G9
G1
G1: CsgG-ΔPYPA单体
G9: CsgG-ΔPYPA九聚体
电流/pA
100
80
60
40
20
0
-20
7
8
9
时间/s
+50 mV下的CsgG-△PYPA电流迹线

图57