



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(11) BR 112013020411-7 B1



(22) Data do Depósito: 10/02/2012

(45) Data de Concessão: 08/09/2021

(54) Título: MONÔMERO DE MSP MUTANTE, CONSTRUTO, POLINUCLEOTÍDEO, PORO, KIT E APARELHO PARA CARACTERIZAR UMA SEQUÊNCIA DE ÁCIDO NUCLEICO ALVO, E, MÉTODO PARA CARACTERIZAR UMA SEQUÊNCIA DE ÁCIDO NUCLEICO ALVO

(51) Int.Cl.: G01N 33/487; C12Q 1/68.

(30) Prioridade Unionista: 11/02/2011 US 61/441718.

(73) Titular(es): OXFORD NANOPORE TECHNOLOGIES LIMITED.

(72) Inventor(es): JAMES CLARKE; ANDREW JOHN HERON; LAKMAL JAYASINGHE; JAYNE WALLACE; JAMES WHITE.

(86) Pedido PCT: PCT GB2012050301 de 10/02/2012

(87) Publicação PCT: WO 2012/107778 de 16/08/2012

(85) Data do Início da Fase Nacional: 09/08/2013

(57) Resumo: MONÔMERO DE Msp MUTANTE, CONSTRUTO, POLINUCLEOTÍDEO, PORO, KIT E APARELHO PARA CARACTERIZAR UMA SEQUÊNCIA DE ÁCIDO NUCLEICO ALVO, E, MÉTODO PARA CARACTERIZAR UMA SEQUÊNCIA DE ÁCIDO NUCLEICO ALVO. A invenção diz respeito às formas mutantes de Msp. A invenção também diz respeito à caracterização de ácido nucleico usando Msp.

MONÔMERO DE Msp MUTANTE, CONSTRUTO, POLINUCLEOTÍDEO, PORO, KIT E APARELHO PARA CARACTERIZAR UMA SEQUÊNCIA DE ÁCIDO NUCLEICO ALVO, E, MÉTODO PARA CARACTERIZAR UMA SEQUÊNCIA DE ÁCIDO NUCLEICO ALVO

Campo da Invenção

[001] A invenção diz respeito às formas mutantes de Msp. A invenção também diz respeito à caracterização de ácido nucleico usando Msp.

Fundamentos da Invenção

[002] A detecção de nanoporo é uma abordagem para detecção que depende da observação de eventos de ligação individuais entre as moléculas de analito e um receptor. Sensores de nanoporo podem ser gerados colocando um único poro de dimensões nanométricas em uma membrana isolante, e medindo o transporte iônico impulsionado por tensão através do poro na presença de moléculas de analito. A identidade de um analito é revelada por meio de sua assinatura da corrente distintiva, notavelmente a duração e extensão do bloco da corrente e a variância de níveis da corrente.

[003] Existe atualmente uma necessidade de tecnologias de sequenciamento de ácido nucleico rápidas e baratas (por exemplo, DNA ou RNA) através de uma faixa ampla de aplicações. Tecnologias existentes são lentas e caras basicamente em virtude de elas contarem com técnicas de amplificação para produzir grandes volumes de ácido nucleico e exigirem uma quantidade alta de produtos químicos fluorescentes para fins específicos para detecção de sinal. A detecção de nanoporo tem o potencial de prover sequenciamento de ácido nucleico rápido e barato reduzindo a quantidade de nucleotídeo e reagentes exigidos.

[004] Dois dos componentes essenciais de sequenciamento de ácidos nucleicos usando a detecção de nanoporo são (1) o controle de movimento do ácido nucleico através do poro e (2) a discriminação de nucleotídeos a medida que o polímero do ácido nucleico move-se através do poro. No passado, para

obter discriminação de nucleotídeo o ácido nucleico era passado através de um mutante de hemolisina. Isto forneceu assinaturas de correntes que mostraram ser dependentes da sequência. Mostrou-se também que um grande número de nucleotídeos contribui para a corrente observada, tornando um relacionamento direto entre corrente observada e sequência de ácido nucleico desafiador.

[005] Embora a faixa de corrente para discriminação de nucleotídeo tenha sido melhorada através de mutação do poro de hemolisina, um sistema de sequenciamento teria maior desempenho se as diferenças de corrente entre nucleotídeos pudessem ser melhoradas ainda mais. Além do mais, observou-se que quando os ácidos nucleicos movem-se através de um poro, alguns estados da corrente mostraram alta variância. Mostrou-se também que alguns poros de hemolisina mutante exibiram variância maior do que outros. Embora a variância desses estados possa conter informação específica de sequência, é desejável produzir poros que têm baixa variância para simplificar o sistema. É também desejável reduzir o número de nucleotídeos que contribui para a corrente observada.

[006] As diferentes formas de Msp são porinas de *Mycobacterium smegmatis*. MspA é um 157 kDa porina octamérica de *Mycobacterium smegmatis*. A estrutura de MspA foi bem documentada por pesquisadores (Gundlach, Proc Natl Acad Sci U S A. 2010 Sep 14; 107(37): 16060-5. Epub 2010 Aug 26). Alguns resíduos chave foram identificados e modificados para melhorar as propriedades do poro. Essas mutações foram realizadas para permitir que DNA migre através do poro de MspA. MspB, C e D são também formas conhecidas de Msp.

Sumário da Invenção

[007] Os inventores demonstraram surpreendentemente que mutantes inéditos de Msp apresentam melhores propriedades para estimar as características, tal como a sequência de ácidos nucleicos. Os mutantes

surpreendentemente apresentam melhor discriminação de nucleotídeo. Em particular, os mutantes surpreendentemente apresentam uma maior faixa de corrente, que torna fácil discriminá-los entre diferentes nucleotídeos, e uma menor variância de estados, que aumenta a razão de sinal para ruído. Além do mais, o número de nucleotídeos contribuindo para a corrente a medida que o ácido nucleico move através do poro é diminuído. Isto torna mais fácil identificar um relacionamento direto entre a corrente observada a medida que o ácido nucleico move através do poro e da sequência de ácido nucleico.

[008] Os inventores também mostraram surpreendentemente que Msp mostra propriedades de sequenciamento melhoradas quando o movimento do ácido nucleico através do poro é controlado por um DNA polimerase Phi29. Em particular, o acoplamento de Msp e DNA polimerase Phi29 resulta em três vantagens surpreendentes. Primeiro, o ácido nucleico move através do poro a uma taxa que é comercialmente viável permite ainda sequenciamento efetivo. Segundo, uma maior faixa de corrente é observada a medida que o ácido nucleico move através do poro permitindo que a sequência seja determinada mais facilmente. Terceiro, uma baixa variância de corrente é observada, aumentando por meio disso a razão de sinal para ruído.

[009] Dessa maneira, a invenção fornece a um monômero de Msp mutante compreendendo uma variante da sequência mostrada na SEQ ID NO: 2, em que a variante compreende pelo menos uma das seguintes mutações:

- (a) asparagina (N), serina (S), glutamina (Q) ou treonina (T) na posição 88
- (b) serina (S), glutamina (Q) ou tirosina (Y) na posição 90;
- (c) leucina (L) ou serina (S) na posição 105;
- (d) arginina (R) na posição 126;
- (e) serina (S) na posição 75;
- (f) serina (S) na posição 77;
- (g) arginina (R) na posição 59;

- (h) glutamina (Q), asparagina (N) ou treonina (T) na posição 75;
- (i) glutamina (Q), asparagina (N) ou treonina (T) na posição 77;
- (j) leucina (L) na posição 78;
- (k) asparagina (N) na posição 81;
- (l) asparagina (N) na posição 83;
- (m) serina (S) ou treonina (T) na posição 86;
- (n) fenilalanina (F), valina (V) ou leucina (L) na posição 87;
- (o) tirosina (Y), fenilalanina (F), valina (V), arginina (R), alanina (a), glicina (G) ou cisteína (C) na posição 88;
- (p) fenilalanina (F), valina (V) ou leucina (L) na posição 89;
- (q) leucina (L), fenilalanina (F), triptofano (W), histidina (H), treonina (T), glicina (G), alanina (a), valina (V), arginina (R), lisina (K), asparagina (N) ou cisteína (C) na posição 90;
- (r) serina (S), glutamina (Q), leucina (L), metionina (M), isoleucina (I), alanina (a), valina (V), glicina (G), fenilalanina (F), triptofano (W), tirosina (Y), histidina (H), treonina (T), arginina (R), lisina (K), asparagina (N) ou cisteína (C) na posição 91;
- (s) alanina (a) ou serina (S) na posição 92;
- (t) serina (S), alanina (a), treonina (T), glicina (G) na posição 93;
- (u) leucina (L) na posição 94;
- (v) valina (V) na posição 95;
- (w) arginina (R), ácido aspártico (D), valina (V), asparagina (N), serina (S) ou treonina (T) na posição 96;
- (x) serina (S) na posição 97;
- (y) serina (S) na posição 98;
- (z) serina (S) na posição 99;
- (aa) serina (S) na posição 100;
- (bb) fenilalanina (F) na posição 101;
- (cc) lisina (K), serina (S) ou treonina (T) na posição 102;

- (dd) alanina (a), glutamina (Q), asparagina (N), glicina (G) ou treonina (T) na posição 103;
- (ee) isoleucina na posição 104;
- (ff) tirosina (Y), alanina (a), glutamina (Q), asparagina (N), treonina (T), fenilalanina (F), triptofano (W), histidina (H), glicina (G), valina (V), arginina (R), lisina (K), prolina (P), ou cisteína (C) na posição 105;
- (gg) fenilalanina (F), isoleucina (I), valina (V) ou serina (S) na posição 106;
- (hh) prolina (P) ou serina (S) na posição 108;
- (ii) asparagina (N) na posição 118;
- (jj) serina (S) ou cisteína (C) na posição 103; e
- (kk) cisteína em uma ou mais das posições 10 a 15, 51 a 60, 136 a 139 e 168 a 172.

A invenção também fornece:

- um construto compreendendo dois ou mais monômeros covalentemente anexados derivados de Msp;
- um polinucleotídeo que codifica um mutante da invenção ou um construto da invenção;
- um poro homo-oligomérico derivado de Msp compreendendo monômeros mutantes idênticos da invenção;
- um poro hetero-oligomérico derivado de Msp compreendendo pelo menos um monômero mutante da invenção, em que pelo menos um dos oito monômeros difere dos outros;
- um método para caracterizar uma sequência de ácido nucleico alvo, compreendendo:
 - (a) colocar a sequência alvo em contato com um poro da invenção e uma proteína de ligação de ácido nucleico de maneira que a proteína controle o movimento da sequência alvo através do poro e

uma proporção dos nucleotídeos na sequência alvo interaja com o poro; e

- (b) medir a corrente que passa através do poro durante cada interação e, por meio disso, caracterizar a sequência alvo;
 - um estojo para sequenciar uma sequência de ácido nucleico alvo compreendendo (a) um poro da invenção e (b) uma enzima de manuseio de ácido nucleico;
 - um aparelho para sequenciar sequências de ácido nucleico alvos em uma amostra, compreendendo (a) uma pluralidade de poros da invenção e (b) uma pluralidade de enzimas de manuseio de ácido nucleico;
 - um método para caracterizar uma sequência de ácido nucleico alvo, compreendendo:
 - (a) colocar a sequência alvo em contato com um poro derivado de Msp e um DNA polimerase Phi29 de maneira tal que a polimerase controle o movimento da sequência alvo através do poro e uma proporção dos nucleotídeos na sequência alvo interaja com o poro; e
 - (b) medir a corrente que passa através do poro durante cada interação e, por meio disso, caracterizar a sequência alvo, em que as etapas (a) e (b) são realizadas com uma tensão aplicada através do poro;
 - um método para formar um sensor para caracterizar uma sequência de ácido nucleico alvo, compreendendo:
 - (a) colocar um poro derivado de Msp em contato com um DNA polimerase Phi29 na presença da sequência de ácido nucleico alvo; e
 - (b) aplicar uma tensão através do poro para formar um complexo entre o poro e a polimerase; e, por meio disso, formar um sensor para caracterizar a sequência de ácido nucleico alvo;

- um método para aumentar a taxa de atividade de um DNA polimerase Phi29, compreendendo:
 - (a) colocar a DNA polimerase Phi29 em contato com um poro derivado de Msp na presença de uma sequência de ácido nucleico; e
 - (b) aplicar uma tensão através do poro para formar um complexo entre o poro e a polimerase; e, por meio disso, aumentar a taxa de atividade de um DNA polimerase Phi29;
- um estojo para caracterizar uma sequência de ácido nucleico alvo compreendendo (a) um poro derivado de Msp e (b) um DNA polimerase Phi29; e
- um aparelho para caracterizar sequências de ácido nucleico alvos em uma amostra, compreendendo uma pluralidade de poros derivados de Msp e uma pluralidade de DNA polimerase Phi29s.

Descrição das Figuras

[0010] A Fig. 1 mostra o tempo de permanência médio de níveis individuais da corrente a medida que uma fita de DNA única transloca o nanoporo. Os dados são colados a partir de um número de moléculas únicas e são divididos em quartis por níveis da corrente.

[0011] A Fig. 2 mostra níveis e variância da corrente obtidos do uso de Phi29 em modo de abertura para mover uma fita de DNA (SEQ ID NO: 15) através do nanoporo MS-(NNNRRK)8.

[0012] A Fig. 3 mostra níveis e variância da corrente obtidos usando Phi29 em modo de abertura para mover uma fita de DNA (SEQ ID NO: 15) através do nanoporo HL-(mutante)7.

[0013] A Fig. 4 mostra os níveis da corrente para um canal de MspA único registrado a uma faixa de potenciais aplicados (-200 mV a 200 mV).

[0014] A Fig. 5 mostra a curva IV de níveis de poro aberto para o mutante de MspA de linha de base, MS-(B1)8. Cada linha representa um único poro.

[0015] A Fig. 6 mostra a curva IV de níveis de poro aberto para o mutante de MspA, MS-(B1-I105Y)₈. Cada linha representa um único poro.

[0016] A Fig. 7 mostra a curva IV de níveis de poro aberto para o mutante de MspA, MS-(B1-I105N)₈. Cada linha representa um único poro.

[0017] A Fig. 8 mostra a mudança na corrente entre um estado de alta condutância (275 pA) e um estado de baixa condutância (150 pA) para o poro MS-(B1-I105A)₈ a 180 mV.

[0018] A Fig. 9 mostra os níveis da corrente produzida quando DNA está aberto através do poro MS-(B1)₈ de linha de base. A faixa de corrente para esses eventos é ~ 30 pA.

[0019] A Fig. 10 mostra os níveis da corrente produzidos quando DNA está aberto através do poro MS-(B1-I10 A)₈ de linha de base. A faixa de corrente para esses eventos é ~ 40 pA.

[0020] A Fig. 11 mostra o projeto do substrato de DNA usado nos exemplos 9 e 12 e 15.

[0021] A Fig. 12 mostra o projeto do substrato de DNA usado nos exemplos 10 e 11.

[0022] A Fig. 13 mostra como o perfil de sequenciamento muda para a mesma sequência de DNA, quando mutações pontuais são feitas na sequência de monômero de MspA. Esses gráficos mostram a média do perfil dos níveis obtidos de múltiplos polinucleotídeos. A) Este gráfico mostra o perfil de sequenciamento para o poro MS-(B1)₈. B) Este gráfico mostra o perfil de sequenciamento para o poro MS-(B1-D90Q-D93S-1105A)₈. C) Este gráfico mostra o perfil de sequenciamento para o poro MS-(B 1 -D90Q-Q126R)₈. D) Este gráfico mostra o perfil de sequenciamento para o poro MS-(B1-L88N-D90Q-D91M)₈. E) Este gráfico mostra o perfil de sequenciamento para o poro MS-(B1-L88N-D90Q-D91S)₈. F) Este gráfico mostra o perfil de sequenciamento para o poro MS-(B1-G75S-G77S-L88N-Q126R)₈.

[0023] A Fig. 14 mostra o projeto do substrato de DNA usado em

Exemplo 13.

[0024] A Fig. 15 mostra um traço de evento de exemplo para a translocação controlada de RNA, mediada por DNA polimerase Phi29, através do poro MS-(B1)₈ de MspA mutante. Uma vista expandida da região salientada no traço superior é mostrada a seguir.

[0025] A Fig. 16 mostra inserção de poro na bicamada de lipídio. A) Mostra inserção de poro do MS-(BL)₈ oligomerizado a partir do monômero. B) Mostra inserção de poro do MS-(B1-B1)₄ oligomerizado a partir do dímero.

[0026] A Fig. 17 mostra um traço de evento de exemplo para a translocação controlada de DNA, mediada por um helicase, através do poro do mutante MS-(B1)₈ que foi produzido por oligomerização do monômero. Uma vista expandida, da região salientada no traço superior, é mostrada a seguir.

[0027] A Fig. 18 mostra um traço de evento de exemplo para a translocação controlada de DNA, mediada por um helicase, através do poro do mutante MS-(B1-B1)₄ que foi produzida por oligomerização do dímero. Uma vista expandida, da região salientada no traço superior, é mostrada a seguir.

[0028] A Fig. 19 mostra o projeto do substrato de DNA usado no exemplo 16.

[0029] A Fig. 20 mostra um traço de evento de exemplo para a translocação controlada de DNA contendo tanto citosina quanto 5-metilcitosina, mediada por um helicase, através do poro do mutante MS-(B1-L88N)₈. Uma vista expandida da região salientada no traço superior é mostrada a seguir.

Descrição da Listagem de Sequência

[0030] SEQ ID NO: 1 mostra a sequência de polinucleotídeo que codifica o monômero de MspA do NNN-RR mutante.

[0031] SEQ ID NO: 2 (também referida como “B1”) mostra a sequência de aminoácido da forma madura do mutante NNN-RRK do monômero de MspA. O mutante não tem a sequência sinal e a metionina amino terminal (codificada pelo códon de partida) e inclui as seguintes mutações: D90N, D91N, D93N, D118R, D134R e E139K. Essas mutações permitem a transição de DNA através do poro de MspA.

[0032] SEQ ID NO: 3 mostra a sequência de polinucleotídeo que codifica a DNA polimerase Phi29.

[0033] SEQ ID NO: 4 mostra a sequência de aminoácido da DNA polimerase Phi29.

[0034] SEQ ID NO: 5 mostra a sequência de polinucleotídeo otimizada por códon derivada do gene *sbcB* de *E. coli*. Ela codifica a enzima exonuclease I (EcoExo I) de *E. coli*.

[0035] SEQ ID NO: 6 mostra a sequência de aminoácido de enzima exonuclease I (EcoExo I) de *E. coli*.

[0036] SEQ ID NO: 7 mostra a sequência de polinucleotídeo otimizada por códon derivada do gene *xthA* de *E. coli*. Ela codifica a enzima exonuclease III de *E. coli*.

[0037] SEQ ID NO: 8 mostra a sequência de aminoácido da enzima exonuclease III de *E. coli*. Esta enzima realiza digestão distributiva de nucleosídeos 5' monofosfato de uma fita de DNA de fita dupla (DNAs) em uma direção 3' - 5'. Iniciação enzimática em uma fita exige uma saliência 5' de aproximadamente 4 nucleotídeos.

[0038] SEQ ID NO: 9 mostra a sequência de polinucleotídeo otimizada por códon derivada do gene *recJ* de *T. thermophilus*. Ela codifica a enzima *RecJ* de *T. thermophilus* (*TthRecJ-cd*).

[0039] SEQ ID NO: 10 mostra a sequência de aminoácido da enzima *RecJ* de *T. thermophilus* (*TthRecJ-cd*). Esta enzima realiza digestão processiva de nucleosídeos 5' monofosfato de DNAs em uma direção 5' - 3'.

Iniciação enzimática em uma fita exige pelo menos 4 nucleotídeos.

[0040] SEQ ID NO: 11 mostra a sequência de polinucleotídeo otimizada por códon derivada de bacteriófago lambda do gene *exo* (*redX*). Ela codifica o bacteriófago lambda exonuclease.

[0041] SEQ ID NO: 12 mostra a sequência de aminoácido do bacteriófago lambda exonuclease. A sequência é uma das três subunidades idênticas que montam em um trímero. A enzima realiza digestão altamente processiva de nucleotídeos de uma fita de DNAs, em uma direção 5'-3' (<http://www.neb.com/nebecomm/products/productM0262.asp>). Iniciação enzimática em uma fita preferencialmente exige um 5' saliência de aproximadamente 4 nucleotídeos com um fosfato 5'.

[0042] SEQ ID NOs: 13 a 15 mostram as sequências usadas no exemplo 2.

[0043] SEQ ID NOs: 16 a 18 mostram a sequências de aminoácido das formas maduras dos mutantes MspB, C e D respectivamente. As formas maduras não têm a sequência sinal.

[0044] SEQ ID NOs: 19 e 20 mostram as sequências usadas nos exemplos 9, 12 e 15.

[0045] SEQ ID NOs: 21 a 23 mostram as sequências usadas nos exemplos 10 e 11.

[0046] SEQ ID NOs: 24 a 27 mostram as sequências usadas no exemplo 13.

[0047] SEQ ID NO: 28 mostra a sequência de DNA do dímero da forma madura do mutante NNN-RRK do monômero de MspA usadas no exemplo 14.

[0048] SEQ ID NO: 29 mostra a sequência de proteína do dímero da forma madura de o mutante NNN-RRK do monômero de MspA usadas no exemplo 14.

[0049] SEQ ID NO: 30, 31 e 32 mostram as sequências usadas no

exemplo 16.

[0050] SEQ ID NO: 33 mostra a sequência ligante usada mostrada no construto mostrado na SEQ ID NO: 29.

Descrição Detalhada da Invenção

[0051] Deve-se entender que diferentes aplicações dos produtos e métodos revelados podem ser ajustadas às necessidades específicas na tecnologia. Deve-se também entender que a terminologia usada aqui tem apenas o propósito de descrever modalidades particulares da invenção, e não deve ser considerada limitante.

[0052] Além do mais, da maneira usada nesta especificação e nas reivindicações anexas, as formas singulares “um”, “uma”, e “o” “a” incluem referentes no plural a menos que o conteúdo dite claramente de outra forma. Assim, por exemplo, referência a “um mutante” inclui “mutantes”, referência a “uma substituição” inclui duas ou mais tais substituições, referência a “um poro” inclui dois ou mais poros como esses, referência a “uma sequência de ácido nucleico” inclui duas ou mais tais sequências, e similares.

[0053] Todas as publicações, patentes e pedidos de patente citados aqui, quer supra quer infra, estão aqui incorporados pela referência na sua íntegra.

Monômeros de Msp Mutantes

[0054] A presente invenção diz respeito a monômeros de Msp mutante. Os monômeros de Msp mutantes podem ser usados para formar os poros da invenção. Um monômero de Msp de mutante é um monômero cuja sequência varia daquela de um monômero de Msp tipo selvagem e que retém a capacidade de formar um poro. Métodos para confirmar a capacidade de monômeros mutantes de formar poros são bem conhecidos na tecnologia e são discutidos com mais detalhes a seguir.

[0055] Os monômeros mutantes têm melhores propriedades de leitura de nucleotídeo, isto é, apresentam melhor captura e discriminação de

nucleotídeo. Em particular, poros construídos dos monômeros mutantes capturam nucleotídeos e ácidos nucleicos mais facilmente do que o tipo selvagem. Além do mais, poros construídos dos monômeros mutantes apresentam uma maior faixa de corrente, que torna fácil discriminar diferentes nucleotídeos, e uma menor variância de estados, que aumenta a razão de sinal para ruído. Além do mais, o número de nucleotídeos contribuindo para a corrente a medida que o ácido nucleico move através de poros construídos dos mutantes é diminuído. Isto torna mais fácil identificar um relacionamento direto entre a corrente observada a medida que o ácido nucleico move através do poro e da sequência de ácido nucleico. As melhores propriedades de leitura de nucleotídeo dos mutantes são obtidas por meio de cinco mecanismos principais, a saber, por mudanças em:

- estéricos (aumentando ou diminuindo o tamanho de resíduos de aminoácido);
- carga (por exemplo, introduzindo carga +ve para interagir com a sequência de ácido nucleico);
- ligação de hidrogênio (por exemplo, introduzindo aminoácidos que podem ligar hidrogênio nos pares base);
- empilhamento Pi (por exemplo, introduzindo aminoácidos que interagem através de sistemas pi de elétrons deslocalizados); e/ou
- alteração da estrutura do poro (por exemplo, introduzindo aminoácidos que aumentam o tamanho do vestíbulo e/ou constrição).

[0056] Qualquer um ou mais desses cinco mecanismos pode ser responsável pelas propriedades melhoradas dos poros da invenção. Por exemplo, um poro da invenção pode exibir propriedades de leitura de nucleotídeo melhoradas em decorrência de estéricos alterados, ligação de hidrogênio alterada e uma estrutura alterada.

[0057] A introdução de resíduos volumosos, tais como fenilalanina (F), triptofano (W), tirosina (Y) ou histidina (H), aumenta os estéricos do

poro. A introdução de resíduos aromáticos, tais como fenilalanina (F), triptofano (W), tirosina (Y) ou histidina (H), também aumenta o empilhamento pi no poro. A introdução de resíduos volumosos ou aromáticos também altera a estrutura do poro, por exemplo, abrindo o poro e aumentando o tamanho do vestíbulo e/ou constrição. Isto é descrito com mais detalhes a seguir.

[0058] Um monômero mutante da invenção compreende uma variante da sequência mostrada na SEQ ID NO: 2. SEQ ID NO: 2 é o mutante NNN-RRK do monômero de MspA. Ele inclui as seguintes mutações: D90N, D91N, D93N, D118R, D134R e E139K. Uma variante da SEQ ID NO: 2 é um polipeptídeo que tem uma sequência de aminoácido que varia daquela da SEQ ID NO: 2 e que retém sua capacidade de formar um poro.

[0059] A variante compreende pelo menos uma das seguintes mutações:

- (a) asparagina (N), serina (S), glutamina (Q) ou treonina (T) na posição 88;
- (b) serina (S), glutamina (Q) ou tirosina (Y) na posição 90;
- (c) leucina (L) ou serina (S) na posição 105;
- (d) arginina (R) na posição 126;
- (e) serina (S) na posição 75;
- (f) serina (S) na posição 77;
- (g) arginina (R) na posição 59;
- (h) glutamina (Q), asparagina (N) ou treonina (T) na posição 75;
- (i) glutamina (Q), asparagina (N) ou treonina (T) na posição 77;
- (j) leucina (L) na posição 78;
- (k) asparagina (N) na posição 81;
- (l) asparagina (N) na posição 83;
- (m) serina (S) ou treonina (T) na posição 86;
- (n) fenilalanina (F), valina (V) ou leucina (L) na posição 87;

- (o) tirosina (Y), fenilalanina (F), valina (V), arginina (R), alanina (a), glicina (G) ou cisteína (C) na posição 88;
- (p) fenilalanina (F), valina (V) ou leucina (L) na posição 89;
- (q) leucina (L), fenilalanina (F), triptofano (W), histidina (H), treonina (T), glicina (G), alanina (a), valina (V), arginina (R), lisina (K), asparagina (N) ou cisteína (C) na posição 90;
- (r) serina (S), glutamina (Q), leucina (L), metionina (M), isoleucina (I), alanina (a), valina (V), glicina (G), fenilalanina (F), triptofano (W), tirosina (Y), histidina (H), treonina (T), arginina (R), lisina (K), asparagina (N) ou cisteína (C) na posição 91;
- (s) alanina (a) ou serina (S) na posição 92;
- (t) serina (S), alanina (a), treonina (T), glicina (G) na posição 93;
- (u) leucina (L) na posição 94;
- (v) valina (V) na posição 95;
- (w) arginina (R), ácido aspártico (D), valina (V), asparagina (N), serina (S) ou treonina (T) na posição 96;
- (x) serina (S) na posição 97;
- (y) serina (S) na posição 98;
- (z) serina (S) na posição 99;
- (aa) serina (S) na posição 100;
- (bb) fenilalanina (F) na posição 101;
- (cc) lisina (K), serina (S) ou treonina (T) na posição 102;
- (dd) alanina (a), glutamina (Q), asparagina (N), glicina (G) ou treonina (T) na posição 103;
- (ee) isoleucina na posição 104;
- (ff) tirosina (Y), alanina (a), glutamina (Q), asparagina (N), treonina (T), fenilalanina (F), triptofano (W), histidina (H), glicina (G), valina (V), arginina (R), lisina (K), prolina (P), ou cisteína (C) na posição 105;

- (gg) fenilalanina (F), isoleucina (I), valina (V) ou serina (S) na posição 106;
- (hh) prolina (P) ou serina (S) na posição 108;
- (ii) asparagina (N) na posição 118;
- (jj) serina (S) ou cisteína (C) na posição 103; e
- (kk) cisteína em uma ou mais das posições 10 a 15, 51 a 60, 136 a 139 e 168 a 172.

[0060] Em MspA tipo selvagem, resíduos 88 e 105 em cada monômero formam um anel hidrofóbico na constrição interna do poro. Os resíduos hidrofóbicos nas posições L88 e I105 assentam logo acima da constrição principal do poro, voltados para o canal aquoso. Mutação desses resíduos produz poros que têm correntes de poro aberto significativamente maiores na linha de base (SEQ ID NO: 2). As diferenças de corrente observadas quando mutações são feitas nestas posições são significativamente maiores do que seria esperado fazendo uma mutação simples. Este resultado surpreendentemente implica que mutações nessas posições podem ter um efeito na estrutura do canal sem ser apenas no ambiente local nestes resíduos. Embora tenha sido reportado que a linha de base da SEQ ID NO: 2 apresenta uma faixa ampla de condutância do poro, o motivo para isto não é bem entendido. Mutações nas posições L88 e I105 fazem com que o nível de corrente do poro dominante seja significativamente maior que o poro da linha de base. Além do mais, este estado de condutância maior é a conformação dominante do mutante, que é desejável para uma grande faixa de corrente e maior razão sinal para ruído.

[0061] A introdução de N, S, Q ou T na posição 88 (isto é, mutação (a) anterior) introduz na constrição interna do poro um aminoácido que pode ligar hidrogênio nos nucleotídeos em um ácido nucleico.

[0062] Resíduos 90 e 91 em cada monômero também formam parte da constrição interna do poro. Resíduo 118 em cada monômero está presente

no vestíbulo do poro. Resíduo 134 em cada monômero é parte da entrada para o poro.

[0063] A introdução de S, Q ou Y na posição 90 (isto é, mutação (b) anterior) introduz na constrição interna do poro um aminoácido que pode ligar hidrogênio nos nucleotídeos em um ácido nucleico.

[0064] A variante pode incluir qualquer número de mutações (a) a (kk), tal como 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ou mais das mutações. Combinações preferidas de mutações são discutidas a seguir. Os aminoácidos introduzidos na variante podem ser de ocorrência natural ou derivados de ocorrência não natural destes. Os aminoácidos introduzidos na variante podem ser D-aminoácidos.

[0065] Qualquer número de cisteínas pode ser introduzido na variante. Cisteínas são preferivelmente introduzidas em uma ou mais, tal como dois ou todas as posições de 90, 91 e 103. Essas posições podem ser usadas para anexação química de um adaptador molecular conforme discutido com mais detalhes a seguir. Qualquer número de cisteínas, tais como 2, 3, 4, 5, 6 ou mais cisteínas, pode ser introduzido nas posições 10 a 15, 51 a 60, 136 a 139 e 168 a 172. Essas posições estão presentes em regiões de laço não conservadas do poro e são assim usadas para anexar quimicamente uma proteína de ligação de ácido nucleico no poro conforme discutido com mais detalhes a seguir.

[0066] Em uma modalidade preferida, a variante compreende uma ou mais das substituições mostradas em (a) a (Z) a seguir. A variante pode incluir qualquer número das substituições em A a Z, tais como 1, 2, 3, 4 ou 5.

[0067] (A) A introdução de um ou mais de (i) serina (S) na posição 75, (ii) serina (S) na posição 77, (iii) asparagina (N) na posição 88, (iv) glutamina (Q) na posição 90 e (v) arginina (R) na posição 126. A variante pode incluir 1, 2, 3, 4 ou 5 dessas substituições. As vantagens de poros homooctaméricos incluindo todas quatro substituições em cada monômero são

mostradas na tabela 3 a seguir.

[0068] (B) A introdução de um ou mais de (i) glutamina (Q) na posição 90 e (ii) arginina (R) na posição 126. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 3 a seguir.

[0069] (C) A introdução de um ou mais de (i) asparagina (N) na posição 88, (ii) glutamina (Q) na posição 90 e (iii) arginina (R) na posição 126. A variante pode incluir 1, 2 ou 3 dessas substituições. As vantagens de poros homo-octaméricos incluindo todas três dessas substituições em cada monômero são mostradas na tabela 3 a seguir.

[0070] (D) A introdução de um ou mais de (i) serina (S) na posição 88 e (ii) glutamina (Q) na posição 90. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 3 a seguir.

[0071] (E) A introdução de um ou mais de (i) asparagina (N) na posição 88 e (ii) glutamina (Q) na posição 90. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 3 a seguir.

[0072] (F) A introdução de um ou mais de (i) glutamina (Q) na posição 90 e (ii) alanina (a) na posição 105. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0073] (G) A introdução de um ou mais de (i) serina (S) na posição 90 e (ii) serina (S) na posição 92. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0074] (H) A introdução de um ou mais de (i) treonina (T) na posição 88 e (ii) serina (S) na posição 90. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as

substituições em cada monômero são mostradas na tabela 2 a seguir.

[0075] (I) A introdução de um ou mais de (i) glutamina (Q) na posição 87 e (ii) serina (S) na posição 90. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0076] (J) A introdução de um ou mais de (i) tirosina (Y) na posição 89 e (ii) serina (S) na posição 90. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0077] (K) A introdução de um ou mais de (i) asparagina (N) na posição 88 e (ii) fenilalanina (F) na posição 89. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0078] (L) A introdução de um ou mais de (i) asparagina (N) na posição 88 e (ii) tirosina (Y) na posição 89. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0079] (M) A introdução de um ou mais de (i) serina (S) na posição 90 e (ii) alanina (a) na posição 92. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0080] (N) A introdução de um ou mais de (i) serina (S) na posição 90 e (ii) asparagina (N) na posição 94. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0081] (O) A introdução de um ou mais de (i) serina (S) na posição 90 e (ii) isoleucina (I) na posição 104. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0082] (P) A introdução de um ou mais de (i) ácido aspártico (D) na posição 88 e (ii) lisina (K) na posição 105. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0083] (Q) A introdução de um ou mais de (i) asparagina (N) na posição 88 e (ii) arginina (R) na posição 126. A variante pode incluir 1 ou 2 dessas substituições. As vantagens de poros homo-octaméricos incluindo ambas as substituições em cada monômero são mostradas na tabela 2 a seguir.

[0084] (R) Um ou mais de (i) asparagina (N) na posição 88, (ii) glutamina (Q) na posição 90 e (iii) arginina (R) na posição 91. A variante pode incluir 1, 2 ou 3 dessas substituições. As vantagens de poros homo-octaméricos incluindo todas três substituições em cada monômero são mostradas na tabela 2 a seguir.

[0085] (S) A introdução de um ou mais de (i) asparagina (N) na posição 88, (ii) glutamina (Q) na posição 90 e (iii) serina (S) na posição 91. A variante pode incluir 1, 2 ou 3 dessas substituições. As vantagens de poros homo-octaméricos incluindo todas três substituições em cada monômero são mostradas na tabela 2 a seguir.

[0086] (T) A introdução de um ou mais de (i) asparagina (N) na posição 88, (ii) glutamina (Q) na posição 90 e (iii) valina (V) na posição 105. A variante pode incluir 1, 2 ou 3 dessas substituições. As vantagens de poros homo-octaméricos incluindo todas três substituições em cada monômero são mostradas na tabela 2 a seguir.

[0087] (U) A introdução de um ou mais de (i) glutamina (Q) na posição 90, (ii) serina (S) na posição 93 e (iii) alaina (a) na posição 105. A variante pode incluir 1, 2 ou 3 dessas substituições. As vantagens de poros homo-octaméricos incluindo todas três substituições em cada monômero são mostradas na tabela 2 a seguir.

[0088] (V) A introdução de um ou mais de (i) fenilalanina (F),

triptofano (W), tirosina (Y) ou histidina (H) na posição 90, (ii) fenilalanina (F), triptofano (W), tirosina (Y) ou histidina (H) na posição 91 e (iii) fenilalanina (F), triptofano (W), tirosina (Y) ou histidina (H) na posição 105. A variante pode incluir 1, 2 ou 3 dessas substituições. A introdução desses resíduos aromáticos volumosos aumenta os estéricos e empilhamento Pi no vestíbulo e/ou constrição do poro. Ela também aumenta o tamanho do vestíbulo e/ou constrição (isto é, abre o poro).

[0089] (W) A introdução de um ou mais de (i) serina (S), treonina (T), glicina (G), alanina (a) ou valina (V) na posição 90, (ii) serina (S), treonina (T), glicina (G), alanina (a) ou valina (V) na posição 91 e (iii) serina (S), treonina (T), glicina (G), alanina (a) ou valina (V) na posição 105. A variante pode incluir 1, 2 ou 3 dessas substituições. A introdução de resíduos menores diminui os estéricos no vestíbulo e/ou constrição do poro.

[0090] (X) A introdução de serina (S), arginina (R), lisina (K) ou histidina (H) na posição 90 e/ou serina (S), arginina (R), lisina (K) ou histidina (H) na posição 91. A introdução de resíduos positivamente carregados (R, K ou H) aumenta as interações entre a constrição do poro e a sequência de ácido nucleico.

[0091] (Y) A introdução de serina (S), treonina (T), asparagina (N), glutamina (Q), tirosina (Y) ou histidina (H) na posição 90 e/ou serina (S), treonina (T), asparagina (N), glutamina (Q), tirosina (Y) ou histidina (H) na posição 91. A introdução desses resíduos aumenta a ligação de hidrogênio que ocorre entre a constrição do poro e a sequência de ácido nucleico. Ela também aumenta o tamanho do vestíbulo e/ou constrição (isto é, abrem o poro).

[0092] (Z) A introdução de cisteína em uma ou mais das posições 90, 91 e 103. Isto permite que grupos químicos sejam anexados no poro por meio de ligação de cisteína. Isto é discutido com mais detalhes anteriormente e a seguir.

[0093] Variantes preferidas incluem, mas sem limitações, aquelas

compreendendo pelo menos uma da(s) seguinte(s) substituição(s): L88N; L88S; L88Q; L88T; D90S; D90Q; D90Y; I105L; I105S; Q126R; G75S; G77S; G75S, G77S, L88N e Q126R; G75S, G77S, L88N, D90Q e Q126R; D90Q e Q126R; L88N, D90Q e Q126R; L88S e D90Q; L88N e D90Q; E59R; G75Q; G75N; G75S; G75T; G77Q; G77; G77S; G77T; I78L; S81N; T83N; N86S; N86T; I87F; I87V; I87L; L88N; L88S; L88Y; L88F; L88V; L88Q; L88T; I89F; I89V; I89L; N90S; N90Q; N90L; N90Y; N91S; N91Q; N91L; N91M; N91I; N91A; N91V; N91G; G92A; G92S; N93S; N93A; N93T; I94L; T95V; A96R; A96D; A96V; A96N; A96S; A96T; P97S; P98S; F99S; G100S; L101F; N102K; N102S; N102T; S103A; S103Q; S103N; S103G; S103T; V104I; I105Y; I105L; I105A; I105Q; I105N; I105S; I105T; T106F; T106I; T106V; T106S; N108P; N108S; D90Q e I105 A; D90S e G92S; L88T e D90S; I87Q e D90S; I89Y e D90S; L88N e I89F; L88N e I89Y; D90S e G92A; D90S e I94N; D90S e V104I; L88D e I105K; L88N e Q126R; L88N, D90Q e D91R; L88N, D90Q e D91S; L88N, D90Q e I105V; D90Q, D93S e I105A; N91Y; N90Y e 91G; N90G e N91Y; N90G e 91G; I05G; N90R; N91R; N90R e N91R; N90K; N91; N90K e N91K; N90Q e 91G; N90G e N91Q; N90Q e N91Q; R118N; N91C; N90C; N90W; N91W; N90K; N91K; N90R; N91R; N90S e N91S; N90Y e I105A; N90G e I105A; N90Q e I105A; N90S e I105A; L88A e I105A; L88S e I105S; L88N e I105N; N90G e 93G; N90G; N93G; N90G e 91A; I105K; I105R; I105V; I105P; I105W; L88R; L88A; L88G; L88N; N90R e I105A; N90S e I105A; L88A e I105A; L88S e I105S; L88N e I105N; L88C; S103C; e I105C.

[0094] Uma variante particularmente preferida compreende I105N. Poros construídos de monômeros mutantes compreendendo I105N têm uma corrente residual que é aumentada por aproximadamente 80 %.

[0095] A mudança na corrente em relação aos diferentes nucleotídeos é também aumentada. Isto reflete uma mudança na estrutura de poros construídos a partir de monômeros mutantes compreendendo I105N. Poros

como esses, portanto, têm uma maior capacidade de discriminar nucleotídeos.

[0096] Mutantes simples preferidos e suas vantagens quando usados em poros homo-octaméricos são mostrados na tabela 1 a seguir.

Tabela 1

Posição	Substituição	Vantagem(s)
E59	R	Maior captura de DNA
G75	Q	Menor variância
G75	N	Menor variância
G75	S	Menor variância
G75	T	Menor variância
G77	Q	Menor variância
G77	N	Menor variância
G77	S	Menor variância
G77	T	Menor variância
I78	L	Distribuição mais estreita de tamanhos de poro
S81	N	Barril beta mais estável
T83	N	Barril beta mais estável
N86	S	Maior faixa de DNA
N86	T	Maior faixa de DNA
I87	F	Menor variância
I87	V	Menor variância
I87	L	Menor variância
L88	N	Menor variância
L88	S	Menor variância
L88	Y	Perfil ruído-DNA alterado
L88	F	Perfil ruído-DNA alterado
L88	V	Menor variância
L88	Q	Menor variância
L88	T	Maior faixa de DNA Menor variância
I89	F	Maior faixa de DNA Reconhecimento de DNA alterado
I89	V	Distribuição mais estreita de tamanhos de poro
I89	L	Distribuição mais estreita de tamanhos de poro
N90	S	Menor variância Reconhecimento de DNA alterado
N90	Q	Maior corrente de poro Maior faixa de DNA Reconhecimento de DNA alterado
N90	L	Reconhecimento de DNA alterado
N90	Y	Reconhecimento de DNA alterado
N91	S	Reconhecimento de DNA alterado
N91	Q	Reconhecimento de DNA alterado
N91	L	Reconhecimento de DNA alterado
N91	M	Reconhecimento de DNA alterado
N91	I	Reconhecimento de DNA alterado
N91	A	Reconhecimento de DNA alterado
N91	V	Reconhecimento de DNA alterado
N91	G	Reconhecimento de DNA alterado
G92	A	Maior faixa de DNA
G92	S	Estabiliza mutações D90S
N93	S	Maior faixa de DNA
N93	A	Maior faixa de DNA
N93	T	Nível estável de corrente de poro aberto Maior faixa de DNA

Posição	Substituição	Vantagem(s)
I94	L	Maior faixa de DNA
T95	V	Nível estável de corrente de poro aberto Faixa de DNA alterada
A96	R	Maior captura de DNA
A96	D	Reconhecimento de DNA alterado
A96	V	Poros bons Variância do poro alterada
A96	N	Poros bons Variância do poro alterada
A96	S	Poros bons Variância do poro alterada
A96	T	Poros bons Variância do poro alterada
P97	S	Poros bons Variância do poro alterada
P98	S	Poros bons Variância do poro alterada
P99	S	Poros bons Variância do poro alterada
G100	S	Poros bons Variância do poro alterada
L101	F	Poros bons Variância do poro alterada
N102	K	Maior captura de DNA
N102	S	Perfil de variância do poro alterada Maior faixa de DNA
N102	T	Discriminação de DNA alterado
S103	A	Reconhecimento de DNA alterado
S103	Q	Maior faixa de DNA Reconhecimento de DNA alterado
S103	N	Maior faixa de DNA Reconhecimento de DNA alterado
S103	G	Maior faixa de DNA Reconhecimento de DNA alterado
S103	T	Discriminação de DNA alterado
V104	I	Discriminação de DNA alterado
I105	Y	Maior faixa de DNA Reconhecimento de DNA alterado
I105	L	Maior faixa de DNA Menor variância
I105	A	Maior faixa de DNA
I105	Q	Reconhecimento de DNA alterado
I105	N	Maior faixa de DNA
I105	S	Reconhecimento de DNA alterado
I105	T	Reconhecimento de DNA alterado
T106	F	Nível estável de corrente de poro aberto
T106	I	Reconhecimento de DNA alterado
T106	V	Reconhecimento de DNA alterado
T106	S	Maior faixa de DNA
N108	P	Nível estável de corrente de poro aberto variância de DNA alterada
N108	S	Nível estável de corrente de poro aberto variância de DNA alterada
Q126	R	Maior captura de DNA

[0097] Mutantes múltiplos preferidos e suas vantagens quando usados em poros homo-octaméricos são mostrados na tabela 2 a seguir.

Tabela 2

Mutante	Vantagem(s)
D90Q/I105A	Reconhecimento de DNA alterado
D90S/G92S	Reconhecimento de DNA alterado
L88T/D90S	Reconhecimento de DNA alterado
I87Q/D90S	Reconhecimento de DNA alterado
I89Y/D90S	Reconhecimento de DNA alterado
L88N/I89F	Reconhecimento de DNA alterado
L88N/I89Y	Reconhecimento de DNA alterado Maior faixa de DNA
D90S/G92A	Reconhecimento de DNA alterado
D90S/I94N	Reconhecimento de DNA alterado
D90S/V104I	Reconhecimento de DNA alterado
L88D/I105K	Reconhecimento de DNA alterado
L88N/Q126R	Menor variância Maior captura de DNA
L88N/D90Q/D91R	Discriminação de DNA alterado Maior captura de DNA
L88N/D90Q/D91S	Discriminação de DNA alterado
L88N/D90Q/ I105V	Discriminação de DNA alterado
D90Q-D93S- I105A	Discriminação de DNA alterado

[0098] Os mutantes acima de tudo preferidos e suas vantagens quando usados em poros homo-octaméricos são mostrados na tabela 3 a seguir.

Tabela 3 - Mutantes acima de tudo preferidos e suas vantagens

Mutante	Vantagem(s)
G75S/G77S/L88N/Q126R/(ONT Ref: B2C)	Corrente estável de poro aberto Maior captura de DNA Menor variância
G75S/G77S/L88N/D90Q/Q126R	Corrente estável de poro aberto Maior captura de DNA Menor variância Reconhecimento de DNA alterado
L88N	Menor variância
D90Q/Q126R	Maior corrente de poro Maior faixa de DNA Reconhecimento de DNA alterado Maior captura de DNA
L88N/D90Q/Q126R	Maior corrente de poro Maior faixa de DNA Reconhecimento de DNA alterado Maior captura de DNA
L88N/D90Q	Menor variância Reconhecimento de DNA alterado
D90S	Menor variância Reconhecimento de DNA alterado
D90Q	Maior corrente de poro Maior faixa de DNA Reconhecimento de DNA alterado
L88S	Menor variância
L88Q	Menor variância
L88N/D90Q	Maior corrente de poro Maior faixa de DNA Reconhecimento de DNA alterado
I105L	Menor variância Grande faixa de DNA
I105S	Grande faixa de DNA

[0099] Além das mutações específicas discutidas anteriormente, a variante pode incluir outras mutações. Em todo o comprimento da sequência

de aminoácido da SEQ ID NO: 2, uma variante será preferivelmente pelo menos 50 % homóloga à sequência com base na identidade do aminoácido. Mais preferivelmente, a variante pode ser pelo menos 55%, pelo menos 60 %, pelo menos 65%, pelo menos 70 %, pelo menos 75%, pelo menos 80 %, pelo menos 85%, pelo menos 90 % e mais preferivelmente pelo menos 95 %, 97 % ou 99 % homóloga com base na identidade do aminoácido à sequência de aminoácido da SEQ ID NO: 2 em toda a sequência. Podem existir pelo menos 80 %, por exemplo, pelo menos 85 %, 90 % ou 95 % de identidade do aminoácido em um trecho de 100 ou mais, por exemplo, 125, 150, 175 ou 200 ou mais, aminoácidos contíguos (“homologia difícil”).

[00100] Métodos padrões na tecnologia podem ser usados para determinar homologia. Por exemplo, o pacote UWGCG fornece o programa BESTFIT que pode ser usado para calcular homologia, por exemplo, usado em suas configurações padrões (Devereux et al (1984) *Nucleic Acids Research* **12**, p387-395). Os algoritmos PILEUP e BLAST podem ser usados para calcular homologia ou alinhar sequências (tal como identificar resíduos equivalentes ou sequências correspondentes (tipicamente em suas configurações padrões)), por exemplo, da maneira descrita em Altschul S. F. (1993) *J Mol Evol* 36:290-300; Altschul, S.F et al (1990) *J Mol Biol* 215:403-10.

[00101] Software para realizar análises BLAST é publicamente disponível através do National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo envolve primeiro identificar um par de sequências de maior pontuação (HSPs) identificando palavras curtas de comprimento W na sequência de consulta que tanto casam quanto satisfazem algum patamar de pontuação T de valor positivo quando alinhadas com uma do mesmo comprimento em uma sequência de base de dados. T é referido como o patamar de pontuação de palavra vizinha (Altschul et al, supra). Esses sucessos de palavra vizinha inicial agem como sementes para

iniciar buscas para encontrar HSP's contendo-os. Os sucessos de palavra são estendidos em ambas as direções ao longo de cada sequência até o ponto que a pontuação de alinhamento cumulativa pode ser aumentada. Extensões para os sucessos de palavra em cada direção são interrompidos quando: a pontuação de alinhamento cumulativa cai uma quantidade X de seu valor máximo obtido; a pontuação cumulativa vai para zero ou menos, devido ao acúmulo de um ou mais alinhamentos de resíduo de pontuação negativa; ou o final de qualquer sequência é atingido. Os parâmetros do algoritmo BLAST W, T e X determinam a sensibilidade e velocidade do alinhamento. O programa BLAST usa como padrões um comprimento de palavra (W) de 11, a matriz pontuação BLOSUM62 (vide Henikoff and Henikoff (1 92) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) alinhamentos (B) de 50, expectativa (E) de 10, M=5, N=4, e uma comparação de ambas as fitas.

[00102] O algoritmo BLAST realiza uma análise estatística da similaridade entre duas sequências; vide, por exemplo, Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5787. Uma medida de similaridade fornecida pelo algoritmo BLAST é a menor probabilidade de soma (P(N)), que fornece uma indicação da probabilidade pela qual um casamento entre duas sequências de aminoácido ocorreria por mudança. Por exemplo, uma sequência é considerada similar a outra sequência se a menor probabilidade de soma em comparação da primeira sequência com a segunda sequência for menos que cerca de 1, preferivelmente menos que cerca de 0,1, mais preferivelmente menos que cerca de 0,01, e acima de tudo preferivelmente menos que cerca de 0,001.

[00103] SEQ ID NO: 2 é o mutante NNN-RRK do monômero de MspA. A variante pode compreender qualquer uma das mutações nos monômeros MspB, C ou D comparadas com MspA. As formas maduras de MspB, C e D são mostradas em SEQ ID NOs: 16 a 18. Em particular, a variante pode compreender a substituição seguinte presente em MspB:

A138P. A variante pode compreender um ou mais das substituições seguintes presentes em MspC: A96G, N102E e A138P. A variante pode compreender uma ou mais das seguintes mutações presentes em MspD: Deleção de Gl, L2V, E5Q, L8V, D13G, W21A, D22E, K47T, I49H, I68V, D91G, A96Q, N102D, S103T, V104I, S136K e G141A. A variante pode compreender combinações de uma ou mais das mutações e substituições de Msp B, C e D.

[00104] Substituições de aminoácido podem ser feitas na sequência de aminoácido da SEQ ID NO: 2 além daquelas discutidas anteriormente, por exemplo, até 1, 2, 3, 4, 5, 10, 20 ou 30 substituições. Substituições conservativas substituem aminoácidos com outros aminoácidos de estrutura química similar, propriedades químicas similares ou volume de cadeia lateral similar. Os aminoácidos introduzidos podem ter polaridade, hidrofobicidade, basicidade, acidez, neutralidade ou carga similares aos aminoácidos que eles substituem. Alternativamente, a substituição conservativa pode introduzir outro aminoácido que é aromático ou alifático no lugar de um aminoácido aromático ou alifático pré-existente. Mudanças de aminoácido conservativo são bem conhecidas na tecnologia e podem ser selecionadas de acordo com as propriedades dos 20 aminoácidos principais, conforme definido na tabela 4 a seguir. Onde aminoácidos têm polaridade similar, isto pode também ser determinado por referência à escala de hidropatia para cadeias laterais de aminoácido na tabela 5.

Tabela 4 - Propriedades químicas de aminoácidos

Ala	alifático, hidrofóbico, neutro	Met	hidrofóbico, neutro
Cys	polar, hidrofóbico, neutro	Asn	polar, hidrofílico, neutro
Asp	polar, hidrofílico, carregado (-)	Pro	hidrofóbico, neutro
Glu	polar, hidrofílico, carregado (-)	Gln	polar, hidrofílico, neutro
Phe	aromático, hidrofóbico, neutro	Arg	polar, hidrofílico, carregado (+)
Gly	alifático, neutro	Ser	polar, hidrofílico, neutro
His	aromático, polar, hidrofílico, carregado (+)	Thr	polar, hidrofílico, neutro
Ile	alifático, hidrofóbico, neutro	Val	alifático, hidrofóbico, neutro
Lys	polar, hidrofóbico, carregado (+)	Trp	aromático, hidrofóbico, neutro
Leu	alifático, hidrofóbico, neutro	Tyr	aromático, polar hidrofóbico

Tabela 5 - Escala de hidropatia

Cadeia Lateral	Hidropatia
He	4,5
Val	4,2

Cadeia Lateral	Hidropatia
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

[00105] Um ou mais resíduos de aminoácido da sequência de aminoácido da SEQ ID NO: 2 podem adicionalmente ser deletados dos polipeptídeos descritos anteriormente. Até 1, 2, 3, 4, 5, 10, 20 ou 30 ou mais resíduos podem ser deletados.

[00106] Variantes podem incluir fragmentos da SEQ ID NO: 2. Tais fragmentos retêm atividade de formação de poro. Fragmentos podem ter pelo menos 50, 100, 150 ou 200 aminoácidos de comprimento. Tais fragmentos podem ser usados para produzir os poros da invenção. Um fragmento preferivelmente compreende o domínio de formação de poro da SEQ ID NO: 2. Fragmentos devem incluir um dos resíduos 88, 90, 91, 105, 118 e 134 da SEQ ID NO: 2. Tipicamente, fragmentos incluem todos dos resíduos 88, 90, 91, 105, 118 e 134 da SEQ ID NO: 2.

[00107] Um ou mais aminoácidos podem ser alternativamente ou adicionalmente adicionados aos polipeptídeos descritos anteriormente. Uma extensão pode ser fornecida no amino terminal ou carbóxi terminal da sequência de aminoácido da SEQ ID NO: 2 ou variante do polipeptídeo ou seu fragmento. A extensão pode ser bastante curta, por exemplo, de 1 a 10 aminoácidos de comprimento. Alternativamente, a extensão pode ser maior, por exemplo, até 50 ou 100 aminoácidos. Uma proteína veículo pode ser fundida com uma sequência de aminoácido de acordo com a invenção. Outras

proteínas de fusão são discutidas com mais detalhes a seguir. Uma variante pode ter uma metionina no amino terminal da SEQ ID NO: 2.

[00108] Conforme discutido anteriormente, uma variante é um polipeptídeo que tem uma sequência de aminoácido que varia daquela da SEQ ID NO: 2 e que retém sua capacidade de formar um poro. Uma variante tipicamente contém as regiões da SEQ ID NO: 2 que são responsáveis por formação de poro. A capacidade de formação de poro de Msp, que contém um barril β , é fornecida por folhas β em cada subunidade. Uma variante da SEQ ID NO: 2 tipicamente compreende as regiões em SEQ ID NO: 2 que formam folhas β . Uma ou mais modificações podem ser feitas nas regiões da SEQ ID NO: 2 que formam folhas β desde que a variante resultante retenha sua capacidade de formar um poro. Uma variante da SEQ ID NO: 2 preferivelmente inclui uma ou mais modificações, tais como substituições, adições ou deleções, em suas hélices α e/ou regiões de laço.

[00109] Os monômeros mutantes podem ser modificados para ajudar na sua identificação ou purificação, por exemplo, pela adição de resíduos de histidina (um his tag), resíduos de ácido aspártico (um asp tag), uma estreptavidina tag ou um flag tag, ou pela adição de uma sequência sinal para promover sua secreção de uma célula onde o polipeptídeo não conteve naturalmente uma sequência como essa. Uma alternativa para introduzir um tag genético é reagir quimicamente um tag em uma posição nativa ou desenvolvida no poro. Um exemplo disto seria reagir um gel-reagente de desligamento com uma cisteína desenvolvida de fora do poro. Isto foi demonstrado como um método para separar hetero-oligômeros de hemolisina (Chem Biol. 1997 Jul;4(7):497-505).

[00110] O monômero mutante pode ser marcado com um rótulo de revelação. O rótulo de revelação pode ser qualquer rótulo adequado que permite que o poro seja detectado. Rótulos adequados incluem, mas sem limitações, moléculas fluorescentes, radioisótopos, por exemplo, enzimas

1251, 35S, anticorpos, antígenos, polinucleotídeos e ligantes tais como biotina.

[00111] O monômero mutante pode ser feito sinteticamente ou por meio de recombinantes. Por exemplo, o poro pode ser sintetizado por tradução e transcrição *in vitro* (IVTT). A sequência de aminoácido do monômero mutante pode ser modificada para incluir aminoácidos de ocorrência não natural ou para aumentar a estabilidade do monômero. Quando o monômero mutante é produzido por meios sintéticos, tais aminoácidos podem ser introduzidos durante a produção. O monômero mutante pode também ser alterado depois de produção, tanto sintética quanto recombinante.

[00112] O monômero mutante pode também ser produzido usando aminoácidos D. Por exemplo, o monômero mutante pode compreender um mistura de aminoácidos L e aminoácidos D. Isto é convencional na tecnologia para produzir tais proteínas ou peptídeos.

[00113] O monômero mutante contém uma ou mais modificações específicas para facilitar a discriminação de nucleotídeo. O monômero mutante pode também conter outras modificações não específicas, desde que elas não interfiram na formação de poro. Inúmeras modificações de cadeia lateral não específicas são conhecidas na tecnologia e podem ser feitas nas cadeias laterais do monômero mutante. Tais modificações incluem, por exemplo, alquilação redutora de aminoácidos por reação com um aldeído seguido por redução com aBH₄, amidinação com metilacetimidato ou acilação com anidrido acético.

[00114] O monômero mutante pode ser produzido usando métodos padrões conhecidos na tecnologia.

[00115] Sequências de polinucleotídeo que codificam um monômero mutante podem ser derivadas e replicadas usando métodos padrões na tecnologia. Sequências como essas são discutidas com mais detalhes a seguir. Sequências de polinucleotídeo que codificam um monômero mutante podem

ser expressas em uma célula hospedeira bacteriana usando técnicas padrões na tecnologia. O monômero mutante pode ser produzido em uma célula por expressão *in situ* do polipeptídeo de um vetor de expressão recombinante. O vetor de expressão opcionalmente carrega um promotor indutível para controlar a expressão do polipeptídeo.

[00116] Um monômero mutante pode ser produzido em grande escala depois de purificação por qualquer sistema de cromatografia líquida de proteína a partir dos organismos de produção de poro ou após expressão recombinante conforme descrito a seguir. Sistema de cromatografia líquida de proteínas típico inclui sistemas FPLC, AKTA, o sistema Bio-Cad, o sistema Bio-Rad BioLogic e o sistema Gilson HPLC. O monômero mutante pode em seguida ser inserido em uma membrana de ocorrência natural ou artificial para uso de acordo com a invenção. Métodos para inserir poro nas membranas são discutidos a seguir.

[00117] Em algumas modalidades, o monômero mutante é modificado quimicamente. O monômero mutante pode ser modificado quimicamente de qualquer maneira e em qualquer lugar. O monômero mutante é preferivelmente modificado quimicamente por anexação de uma molécula a uma ou mais cisteínas (ligação de cisteína), anexação de uma molécula a uma ou mais lisinas, anexação de uma molécula a um ou mais aminoácidos não natural, modificação enzimática de um epítipo ou modificação de um terminal. Métodos adequados para realizar tais modificações são bem conhecidos na tecnologia. O monômero mutante pode ser modificado quimicamente pela anexação de qualquer molécula. Por exemplo, o monômero mutante pode ser modificado quimicamente por anexação de um corante ou um fluoróforo.

[00118] Em algumas modalidades, o monômero mutante é modificado quimicamente com um adaptador molecular que facilita a interação entre um poro compreendendo o monômero e um nucleotídeo alvo ou sequência de

ácido nucleico alvo. A presença do adaptador melhora a química hospedeiro-hóspede do poro e a sequência de nucleotídeo ou ácido nucleico e, por meio disso, melhora a capacidade de sequenciamento de poros formados a partir do monômero mutante. Os princípios da química hospedeiro-hóspede são bem conhecidos na tecnologia. O adaptador tem um efeito nas propriedades físicas ou químicas do poro que melhora sua interação com a sequência de nucleotídeo ou ácido nucleico. O adaptador pode alterar a carga do barril ou canal do poro ou interagir especificamente com a sequência de nucleotídeo ou ácido nucleico ou ligar neles, por meio disso facilitando sua interação com o poro.

[00119] O adaptador molecular é preferivelmente uma molécula cíclica, um ciclodextrina, uma espécie que é capaz de hibridização, um ligante ou interquelante de DNA, um peptídeo ou análogo de peptídeo, um polímero sintético, uma molécula planar aromática, uma molécula pequena positivamente carregada, ou uma molécula pequena capaz de ligação de hidrogênio.

[00120] O adaptador pode ser cíclico. Um adaptador cíclico preferivelmente tem a mesma simetria do poro. O adaptador preferivelmente tem simetria de oito vezes uma vez que Msp tipicamente tem oito subunidades em torno de um eixo central. Isto é discutido com mais detalhes a seguir.

[00121] O adaptador tipicamente interage com a sequência de nucleotídeo ou ácido nucleico por meio de química hospedeiro-hóspede. O adaptador é tipicamente capaz de interagir com a sequência de nucleotídeo ou ácido nucleico. O adaptador compreende um ou mais grupos químicos que são capazes de interagir com a sequência de nucleotídeo ou ácido nucleico. Um ou mais grupos químicos preferivelmente interagem com a sequência de nucleotídeo ou ácido nucleico por interações não covalentes, tais como interações hidrofóbicas, ligação de hidrogênio, forças de Van der Waal,

interações π -cátion e/ou forças eletrostáticas. Um ou mais grupos químicos que são capazes de interagir com a sequência de nucleotídeo ou ácido nucleico são preferivelmente positivamente carregados. Um ou mais grupos químicos que são capazes de interagir com a sequência de nucleotídeo ou ácido nucleico mais preferivelmente compreendem grupos amino. Os grupos amino podem ser anexados nos átomos de carbono primário, secundário ou terciário. O adaptador ainda mais preferivelmente compreende um anel de grupos amino, tal como um anel de 6, 7 ou 8 grupos amino. O adaptador acima de tudo preferivelmente compreende um anel de oito grupos amino. Um anel de grupos amino protonado pode interagir com grupos fosfato negativamente carregados na sequência de nucleotídeo ou ácido nucleico.

[00122] O posicionamento correto do adaptador dentro do poro pode ser facilitado por química hospedeiro-hóspede entre o adaptador e o poro compreendendo o monômero mutante. O adaptador preferivelmente compreende um ou mais grupos químicos que são capazes de interagir com um ou mais aminoácidos no poro. O adaptador mais preferivelmente compreende um ou mais grupos químicos que são capazes de interagir com um ou mais aminoácidos no poro por meio de interações não covalentes, tais como interações hidrofóbicas, ligação de hidrogênio, forças de Van der Waal, interações π -cátion e/ou forças eletrostáticas. Os grupos químicos que são capazes de interagir com um ou mais aminoácidos no poro são tipicamente hidroxilas ou aminas. Os grupos hidroxila podem ser anexados aos átomos de carbono primário, secundário ou terciário. Os grupos hidroxila podem formar ligações de hidrogênio com aminoácidos não carregados no poro. Qualquer adaptador que facilita a interação entre o poro e a sequência de nucleotídeo ou ácido nucleico pode ser usado.

[00123] Adaptadores adequados incluem, mas sem limitações, ciclodextrinas, peptídeos cíclicos e cucurbiturilas. O adaptador é preferivelmente uma ciclodextrina ou um derivado desta. A ciclodextrina ou

seu derivado pode ser qualquer daquelas reveladas em Eliseev, A. V., and Schneider, H-J. (1994) *J. Am. Chem. Soc.* 116, 6081-6088. O adaptador é mais preferivelmente heptaquis-6-amino- β -ciclodextrina (am7-pCD), 6-monodeóxi-6-monoamino- β -ciclodextrina (ami-CD) ou heptaquis-(6-deóxi-6-guanidino)-ciclodextrina (gu7-pCD). O grupo guanidino em gu7-pCD tem um pKa muito maior do que as amins primárias em am7-pCD e assim ele é mais positivamente carregado. Este adaptador de gu7-PCD pode ser usado para aumentar o tempo de permanência do nucleotídeo no poro, para aumentar a precisão da corrente residual medida, bem como para aumentar a taxa de detecção base a altas temperaturas ou baixas taxas de aquisição de dados.

[00124] Se um reticulador de 3-(2-piridiltio)propionato de succinimidila (SPDP) for usado conforme discutido com mais detalhes a seguir, o adaptador será preferivelmente heptaquis(6-deóxi-6-amino)-6-N-mono(2-piridil)ditiopropanoil- β -ciclodextrina (am₆amPDP 1- β CD).

[00125] Mais adaptadores adequados incluem γ -ciclodextrinas, que compreendem 8 unidades de açúcar (e portanto, têm simetria de oito vezes). A γ -ciclodextrina pode conter uma molécula ligante ou pode ser modificada para compreender todas ou mais das unidades de açúcar modificadas usadas nos exemplos de β -ciclodextrina discutidos anteriormente.

[00126] O adaptador molecular é preferivelmente anexado covalentemente ao monômero mutante. O adaptador pode ser anexado covalentemente ao poro usando qualquer método conhecido na tecnologia. O adaptador é tipicamente anexado por meio de ligação química. Se o adaptador molecular for anexado por meio de ligação de cisteína, uma ou mais cisteínas serão preferivelmente introduzidas no mutante por substituição. Os monômeros mutantes da invenção podem certamente compreender um resíduo de cisteína em uma ou mais das posições 88, 90, 91, 103 e 105. O monômero mutante pode ser modificado quimicamente por anexação de um adaptador molecular a uma ou mais, tais como 2, 3, 4 ou 5, dessas cisteínas.

[00127] Alternativamente, o monômero mutante pode ser modificado quimicamente por anexação de uma molécula a uma ou mais cisteínas introduzidas em outras posições. O adaptador molecular é preferivelmente anexado a uma ou mais das posições 90, 91 e 103 da SEQ ID NO: 2.

[00128] A reatividade dos resíduos de cisteína pode ser aumentada por modificação dos resíduos adjacentes. Por exemplo, os grupos básicos de resíduos de arginina, histidina ou lisina de flaqueamento mudarão o pKa do grupo tiol de cisteínas para aquele do grupo S⁻ mais reativo. A reatividade de resíduos de cisteína pode ser protegida por grupos protetores tiol tal como dTNB. Esses podem ser reagidos com um ou mais resíduos de cisteína do monômero mutante antes de um ligante ser anexado. A molécula pode ser anexada diretamente no monômero mutante. A molécula é preferivelmente anexada ao monômero mutante usando um ligante, tal como um reticulador químico ou um peptídeo ligante.

[00129] Reticuladores químicos adequados são bem conhecidos na tecnologia. Reticuladores preferidos incluem propanoato de 2,5-dioxopirrolidin-1-il 3-(piridin-2-ildissulfanil), butanoato de 2,5-dioxopirrolidin-1-il 4-(piridin-2-ildissulfanil) e octanoato de 2,5-dioxopirrolidin-1-il 8-(piridin-2-ildissulfanil). O reticulador acima de tudo preferido é 3-(2-piridiltio)propionato de succinimidila (SPDP). Tipicamente, a molécula é anexada covalentemente ao reticulador bifuncional antes de o complexo molécula/reticulador ser anexado covalentemente ao monômero mutante, mas é também possível anexar covalentemente o reticulador bifuncional ao monômero antes de o complexo reticulador/monômero bifuncional ser anexado na molécula.

[00130] O ligante é preferivelmente resistente a ditioneitol (DTT). Ligantes adequados incluem, mas sem limitações, ligantes a base de iodoacetamida e a base de maleimida.

[00131] Em outra modalidade, o monômero pode ser anexado a uma

proteína de ligação de ácido nucleico. Isto forma um sistema de sequenciamento modular que pode ser usado nos métodos de sequenciamento da invenção. Proteínas de ligação de ácido nucleico são discutidas a seguir.

[00132] A proteína de ligação de ácido nucleico é preferivelmente anexada covalentemente ao monômero mutante. A proteína pode ser anexada covalentemente ao poro usando qualquer método conhecido na tecnologia. O monômero e proteína podem ser quimicamente fundidos ou geneticamente fundidos. O monômero e proteína são geneticamente fundidos se o construto total for expresso de uma única sequência de polinucleotídeo. Fusão genética de um poro a uma proteína de ligação de ácido nucleico é discutido em pedido internacional PCT/GB09/001679 (publicado como WO 2010/004265).

[00133] Se a proteína de ligação de ácido nucleico for anexada por meio de ligação de cisteína, uma ou mais cisteínas serão preferivelmente introduzidas no mutante por substituição. Os monômeros mutantes da invenção podem certamente compreender resíduos de cisteína em uma ou mais das posições 10 a 15, 51 a 60, 136 a 139 e 168 a 172. Essas posições estão presentes em regiões de laço que têm baixa conservação entre homólogos indicando que mutações ou inserções podem ser toleradas. Elas são, portanto, adequadas para anexação de uma proteína de ligação de ácido nucleico. A reatividade de resíduos de cisteína pode ser aumentada por modificação conforme descrito anteriormente.

[00134] A proteína de ligação de ácido nucleico pode ser anexada diretamente no monômero mutante ou por meio de um ou mais ligantes. A molécula pode ser anexada no monômero mutante usando os ligantes de hibridização descritos em pedido internacional PCT/GB 10/000132 (publicado como WO 2010/086602). Alternativamente, ligantes de peptídeo podem ser usados. Ligantes de peptídeo são as sequências de aminoácido. O comprimento, flexibilidade e hidrofiliidade do ligante de peptídeo são tipicamente projetados de maneira tal que eles não perturbam as funções do

monômero e molécula. Ligantes de peptídeo flexíveis preferidos são trechos de 2 a 20, tal como 4, 6, 8, 10 ou 16, aminoácidos de serina e/ou glicina. Ligantes flexíveis mais preferidos incluem (SG)₁, (SG)₂, (SG)₃, (SG)₄, (SG)₅ e (SG)₈ em que S é serina e G é glicina. Ligantes rígidos preferidos são trechos de 2 a 30, tais como 4, 6, 8, 16 ou 24, aminoácidos de prolina. Ligantes rígidos mais preferidos incluem (P)₁₂ em que P é prolina.

[00135] O monômero mutante pode ser modificado quimicamente com um adaptador molecular e uma proteína de ligação de ácido nucleico.

Construtos

[00136] A invenção também diz respeito a um construto compreendendo dois ou mais monômeros anexados covalentemente derivados de Msp. O construto da invenção retém sua capacidade de formar um poro. Um ou mais construtos da invenção podem ser usados para formar poros para caracterizar, tal como sequenciar, sequências de ácidos nucleicos. O construto pode compreender 2, 3, 4, 5, 6, 7, 8, 9 ou 10 monômeros. Os dois ou mais monômeros podem ser os mesmos ou diferentes.

[00137] Os monômeros não precisam ser monômeros mutantes da invenção. Por exemplo, pelo menos um monômero pode compreender a sequência mostrada na SEQ ID NO: 2. Alternativamente, pelo menos um monômero pode compreender uma variante da SEQ ID NO: 2 que é pelo menos 50 % homóloga com SEQ ID NO: 2 em toda sua sequência com base na identidade do aminoácido, mas não incluem nenhuma das mutações específicas exigidas pelos monômeros mutantes da invenção. Mais preferivelmente, a variante pode ser pelo menos 55 %, pelo menos 60 %, pelo menos 65 %, pelo menos 70 %, pelo menos 75 %, pelo menos 80 %, pelo menos 85 %, pelo menos 90 % e mais preferivelmente pelo menos 95 %, 97 % ou 99 % homóloga com base na identidade do aminoácido com a sequência de aminoácido da SEQ ID NO: 2 em toda a sequência. Em uma modalidade preferida, pelo menos um monômero no construto é um monômero mutante

da invenção. Todos dos monômeros no construto podem ser um monômero mutante da invenção. Os monômeros mutantes podem ser os mesmos ou diferentes. Em uma modalidade preferida, o construto compreende dois monômeros e pelo menos um dos monômeros é um monômero mutante da invenção.

[00138] Os monômeros são preferivelmente fundidos geneticamente. Monômeros são geneticamente fundidos se o construto total for expresso a partir de uma única sequência de polinucleotídeo. As sequências de codificação dos monômeros podem ser combinadas de qualquer maneira para formar uma única sequência de polinucleotídeo que codifica o construto.

[00139] Os monômeros podem ser geneticamente fundidos em qualquer configuração. Os monômeros podem ser fundidos por meio de seus aminoácidos terminais. Por exemplo, o amino terminal de um monômero pode ser fundido ao carbóxi terminal de outro monômero. Se o construto for formado a partir da fusão genética de dois ou mais monômeros cada qual compreendendo a sequência mostrada na SEQ ID NO: 2 ou uma variante desta, o segundo monômero e monômeros subsequentes no construto (na direção amino para carbóxi) podem compreender uma metionina nas suas extremidades amino terminal (cada uma das quais é fundida no carbóxi terminal do monômero anterior). Por exemplo, se M for um monômero compreendendo a sequência mostrada na SEQ ID NO: 2 ou uma variante (sem uma metionina amino terminal) e mM for um monômero compreendendo a sequência mostrada na SEQ ID NO: 2 ou uma variante com um metionina amino terminal, o construto pode compreender a sequência M-mM, M-mM-mM ou M-mM-mM-mM. As presenças dessas metioninas tipicamente resultam da expressão do códon de partidas (isto é, ATGs) na extremidade 5' dos polinucleotídeos que codificam o segundo monômero ou monômeros subsequentes no polinucleotídeo que codificam todo o construto. O primeiro monômero no construto (na direção amino para carbóxi) pode

também compreender uma metionina (por exemplo, mM-mM, mM-mM-mM ou mM-mM-mM-mM).

[00140] Os dois ou mais monômeros podem ser geneticamente fundidos diretamente entre si. Os monômeros são preferivelmente fundidos geneticamente usando um ligante. O ligante pode ser projetado para restringir a mobilidade dos monômeros. Ligantes preferidos são as sequências de aminoácido (isto é, ligantes de peptídeo). Qualquer um dos ligantes de peptídeo discutidos anteriormente pode ser usado. O construto preferivelmente compreende a sequência mostrada na SEQ ID NO: 29 ou uma variante desta. Cada monômero em SEQ ID NO: 29 compreende a sequência mostrada na SEQ ID NO: 2 ou uma variante desta. O segundo monômero também compreende uma metionina no seu amino terminal conforme descrito anteriormente. Os dois monômeros são ligados por um ligante de peptídeo. Uma variante da SEQ ID NO: 29 pode variar da SEQ ID NO: 29 de qualquer uma das maneiras discutidas anteriormente com referência às variantes da SEQ ID NO: 2. O ligante pode também ser modificado ou substituído com um ligante de peptídeo discutido anteriormente.

[00141] Em outra modalidade preferida, os monômeros são quimicamente fundidos. Uma subunidade é quimicamente fundida em uma enzima se as duas partes forem quimicamente anexadas, por exemplo, por meio de um reticulador químico. Qualquer um dos reticuladores químicos discutidos anteriormente pode ser usado. O ligante pode ser anexado a um ou mais resíduos de cisteína introduzidos em um monômero mutante da invenção. Alternativamente, o ligante pode ser anexado a um terminal de um dos monômeros no construto.

[00142] Se um construto contiver diferentes monômeros, reticulação de monômeros em si mesmo pode ser impedida mantendo a concentração de ligante em um amplo excesso dos monômeros. Alternativamente, um arranjo “fechadura e chave” pode ser usado onde dois ligantes são usados. Apenas

uma extremidade de cada ligante pode reagir entre si para formar um maior ligante e as outras extremidades do ligante cada qual reage com um diferente monômero. Tais ligantes são descritos em Pedido internacional No. PCT/GB 10/000132 (publicado como WO 2010/086602).

Polinucleotídeos

[00143] A presente invenção também diz respeito a sequências de polinucleotídeo que codificam um monômero mutante da invenção. O monômero mutante pode ser qualquer daqueles discutidos anteriormente. A sequência de polinucleotídeo preferivelmente compreende uma sequência pelo menos 50 %, 60 %, 70 %, 80 %, 90 % ou 95 % homóloga com base na identidade de nucleotídeo com a sequência da SEQ ID NO: 1 em toda a sequência. Pode existir pelo menos 80 %, por exemplo, pelo menos 85 %, 90 % ou 95 % de identidade de nucleotídeo em um trecho de 300 ou mais, por exemplo, 375, 450, 525 ou 600 ou mais, nucleotídeos contíguos (“homologia difícil”). Homologia pode ser calculada conforme descrito anteriormente. A sequência de polinucleotídeo pode compreender uma sequência que difere da SEQ ID NO: 1 com base na degenerescência do código genético.

[00144] A presente invenção também diz respeito a sequências de polinucleotídeo que codificam qualquer um dos construtos geneticamente fundidos da invenção. O polinucleotídeo preferivelmente compreende duas ou mais sequências as mostrada na SEQ ID NO: 1 ou uma variante desta conforme descrito anteriormente. A sequência de polinucleotídeo preferivelmente compreende a sequência da SEQ ID NO: 28 ou uma sequência pelo menos 50 %, 60 %, 70 %, 80 %, 90 % ou 95 % homóloga com base na identidade de nucleotídeo com a sequência da SEQ ID NO: 28 em toda a sequência. Pode existir pelo menos 80 %, por exemplo, pelo menos 85 %, 90 % ou 95 % de identidade de nucleotídeo em um trecho de 600 ou mais, por exemplo, 750, 900, 1.050 ou 1.200 ou mais, nucleotídeos contíguos (“homologia difícil”). Homologia pode ser calculada conforme descrito

anteriormente. A sequência de polinucleotídeo pode compreender uma sequência que difere da SEQ ID NO: 28 com base na degenerescência do código genético.

[00145] Sequências de polinucleotídeo podem ser derivadas e replicadas usando métodos padrões na tecnologia. DNA cromossomal que codifica Msp tipo selvagem pode ser extraído de um organismo de produção de poro, tal como *Mycobacterium smegmatis*. O gene que codifica a subunidade do poro pode ser amplificado usando PCR envolvendo iniciadores específicos. A sequência amplificada pode em seguida passar por mutagênese direcionada ao sítio. Métodos adequados de sítio mutagênese direcionada são conhecidos na tecnologia e incluem, por exemplo, combinar reação de cadeia. Polinucleotídeos que codificam um construto da invenção podem ser feitos usando técnicas bem conhecidas, tais como aquelas descritas em Sambrook, J. and Russell, D. (2001). *Molecular Cloning: a Laboratory Manual*, 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[00146] A sequência de polinucleotídeo resultante pode em seguida ser incorporada em um vetor replicável recombinante tal como um vetor de clonagem. O vetor pode ser usado para replicar o polinucleotídeo em uma célula hospedeira compatível. Assim, a sequências de polinucleotídeo pode ser feita introduzindo um polinucleotídeo em um vetor replicável, introduzindo o vetor em uma célula hospedeira compatível, e crescendo a célula hospedeira em condições que realizam replicação do vetor. O vetor pode ser recuperado da célula hospedeira. Células hospedeiras adequadas para clonagem de polinucleotídeos são conhecidas na tecnologia e descritas com mais detalhes a seguir.

[00147] A sequência de polinucleotídeo pode ser clonada em vetor de expressão adequado. Em um vetor de expressão, a sequência de polinucleotídeo é tipicamente operavelmente ligada a uma sequência controle

que é capaz de fornecer a expressão da sequência de codificação pela célula hospedeira. Tais vetores de expressão podem ser usados para expressar uma subunidade do poro.

[00148] Os termos “operavelmente ligado” referem-se a uma justaposição em que os componentes descritos estão em um relacionamento que permite que eles funcionem de sua maneira pretendida. Uma sequência controle “operavelmente ligada” a uma sequência de codificação que é ligada de tal maneira que a expressão da sequência de codificação é obtida em condições compatíveis com as sequências controles. Múltiplas cópias das mesmas sequências, ou de diferentes sequências de polinucleotídeo podem ser introduzidas no vetor.

[00149] O vetor de expressão pode em seguida ser introduzido em uma célula hospedeira adequada. Assim, um monômero mutante ou construto da invenção pode ser produzido inserindo uma sequência de polinucleotídeo em um vetor de expressão, introduzindo o vetor em uma célula hospedeira bacteriana compatível, e crescendo a célula hospedeira em condições que realizam expressão da sequência de polinucleotídeo. O monômero ou construto recombinantemente expresso pode automontar em um poro na membrana da célula hospedeira. Alternativamente, o poro recombinante produzido desta maneira pode ser removido da célula hospedeira e inserido em outra membrana. Quando se produz poros compreendendo pelo menos duas subunidades diferentes, as diferentes subunidades podem ser expressas separadamente em diferentes células hospedeiras conforme descrito anteriormente, removidas das células hospedeiras e montadas em um poro em uma membrana separada, tal como uma membrana de célula de coelho.

[00150] Os vetores podem ser, por exemplo, vetores de plasmídeo, vírus ou fago fornecidos com uma origem de replicação, opcionalmente um promotor para a expressão da dita sequência de polinucleotídeo e opcionalmente um regulador do promotor. Os vetores podem conter um ou

mais genes marcadores selecionáveis, por exemplo, um gene resistente a tetraciclina. Promotores e outros sinais de regulação de expressão podem ser selecionados para ser compatíveis com a célula hospedeira para a qual o vetor de expressão é projetado. Um promotor de T7, trc, lac, ara ou L é tipicamente usado.

[00151] A célula hospedeira tipicamente expressa a subunidade do poro a um alto nível. Células hospedeiras transformadas com uma sequência de polinucleotídeo serão escolhidas para ser compatíveis com o vetor de expressão usado para transformar a célula. A célula hospedeira é tipicamente bacteriana e preferivelmente *Escherichia coli*. Qualquer célula com um lisógeno λ DE3, por exemplo, C41 (DE3), BL21 (DE3), JM109 (DE3), B834 (DE3), TUNER, Origami e Origami B, pode expressar um vetor compreendendo o promotor T7. Além das condições listadas anteriormente qualquer um dos métodos citados em Proc Natl Acad Sci U S A. 2008 Dec 30;105(52):20647-52 pode ser usado para expressar o proteínas Msp.

Poros

[00152] A invenção também diz respeito a vários poros. Os poros da invenção são ideais para caracterização, tal como sequenciamento, de sequências de ácido nucleico em virtude de eles poderem discriminar diferentes nucleotídeos com um alto grau de sensibilidade. Os poros podem surpreendentemente distinguir entre os quatro nucleotídeos em DNA e RNA. Os poros da invenção podem ainda distinguir entre nucleotídeos metilados e não metilados. A resolução base de poros da invenção é surpreendentemente alta. Os poros mostram separação quase completa de todos os quatro nucleotídeos de DNA. Os poros discriminam ainda mais entre monofosfato de deoxicidina (dCMP) e metil-dCMP com base no tempo de permanência no poro e a corrente que passa através do poro.

[00153] Os poros da invenção podem também discriminar entre diferentes nucleotídeos em um faixa de condições. Em particular, os poros

discriminação entre nucleotídeos em condições que são favoráveis à caracterização, tal como sequenciamento de ácidos nucleicos. A extensão na qual os poros da invenção podem discriminar diferentes nucleotídeos pode ser controlada alterando-se o potencial aplicado, a concentração de sal, o tampão, a temperatura e a presença de aditivos, tais como ureia, betaína e DTT. Isto permite que a função dos poros passe por um ajuste fino, particularmente durante sequenciamento. Isto é discutido com mais detalhes a seguir. Os poros da invenção podem também ser usados para identificar polímeros do ácido nucleico da interação com um ou mais monômeros sem ser em um nucleotídeo por base do nucleotídeo.

[00154] Um poro da invenção pode ser isolado, substancialmente isolado, purificado ou substancialmente purificado. Um poro da invenção é isolado ou purificado se ele ficar completamente sem nenhum outro componente, tais como lipídios ou outros poros. Um poro é substancialmente isolado se ele for misturado com veículos ou diluentes que não interferirão com seu uso visado. Por exemplo, um poro é substancialmente isolado ou substancialmente purificado se ele se apresentar em uma forma que compreende menos que 10 %, menos que 5 %, menos que 2 % ou menos que 1 % de outros componentes, tais como lipídios ou outros poros. Alternativamente, um poro da invenção pode estar presente em uma bicamada de lipídio.

[00155] Um poro da invenção pode estar presente como um poro individual ou único. Alternativamente, um poro da invenção pode estar presente em uma população homóloga ou heteróloga de dois ou mais poros.

Poros homo-oligoméricos

[00156] A invenção também diz respeito a um poro homo-oligomérico derivado de Msp compreendendo monômeros mutantes idênticos da invenção. O poro homo-oligomérico preferivelmente compreende um dos mutantes mostrados nas tabelas 1, 2 e 3. O poro homo-oligomérico da invenção é ideal

para caracterizar, tal como sequenciar, ácidos nucleicos. O poro homo-oligomérico da invenção pode ter qualquer das vantagens discutidas anteriormente. As vantagens de poros homo-oligoméricos específicos da invenção são indicadas nas tabelas 1, 2 e 3.

[00157] O poro homo-oligomérico pode conter qualquer número de monômeros mutantes. O poro tipicamente compreende 7, 8, 9 ou 10 monômeros mutantes idênticos. O poro preferivelmente compreende oito monômeros mutantes idênticos. Um ou mais, tais como 2, 3, 4, 5, 6, 7, 8, 9 ou 10, dos monômeros mutantes são preferivelmente modificados quimicamente conforme discutido anteriormente.

[00158] Métodos para fabricar poros são discutidos com mais detalhes a seguir.

Poros hetero-oligoméricos

[00159] A invenção também diz respeito a um poro hetero-oligomérico derivado de Msp compreendendo pelo menos um monômero mutante da invenção, em que pelo menos um dos oito monômeros difere dos outros. O poro hetero-oligomérico da invenção é ideal para caracterizar, tal como sequenciar, ácidos nucleicos. Poros hetero-oligoméricos podem ser produzidos usando métodos conhecidos na tecnologia (por exemplo, Protein Sci. 2002 Jul;11(7): 1813-24).

[00160] O poro hetero-oligomérico contém monômeros suficientes para formar o poro. Os monômeros podem ser de qualquer tipo. O poro tipicamente compreende 7, 8, 9 ou 10 monômeros. O poro preferivelmente compreende oito monômeros.

[00161] O poro pode compreender pelo menos um monômero compreendendo (a) a sequência mostrada na SEQ ID NO: 2 ou (b) uma variante desta que não teve uma mutação exigida pelos monômeros mutantes da invenção. Variantes adequadas são discutidas anteriormente. Nesta modalidade, os monômeros restantes são preferivelmente monômeros

mutantes da invenção. Consequentemente, o poro pode compreender 9, 8, 7, 6, 5, 4, 3, 2 ou 1 monômeros mutantes da invenção.

[00162] Em uma modalidade preferida, o poro compreende (a) um monômero mutante e (b) sete monômeros idênticos, em que o monômero mutante em (a) é diferente dos monômeros idênticos em (b). Os monômeros idênticos em (b) preferivelmente compreendem (i) a sequência mostrada na SEQ ID NO: 2 ou (ii) uma variante desta que não teve uma mutação presente nos monômeros mutantes da invenção.

[00163] Poros preferidos incluem, mas sem limitações, qualquer do seguinte:

- (a) Sete monômeros compreendendo a sequência mostrada na SEQ ID NO: 2 e um monômero mutante compreendendo a substituição N90R, N90, N90Y, N90Q, N90W ou N90C. Esses poros têm um único aminoácido estérico (Y ou W), um único aminoácido carregado (K ou R) ou um único aminoácido reativo (C) introduzidos na constrição interna.
- (b) Sete monômeros compreendendo a sequência mostrada na SEQ ID NO: 2 e um monômero mutante compreendendo a substituição N91R, N91, N91Y, N91Q, N91W ou N91C. Esses poros têm um único aminoácido estérico (Y ou W), um único aminoácido carregado (K ou R) ou um único aminoácido reativo (C) introduzidos na constrição interna.
- (c) Sete monômeros compreendendo a sequência mostrada na SEQ ID NO: 2 e um monômero mutante compreendendo a substituição L88C, S103C ou I105C. Esses poros têm um aminoácido reativo introduzido no poro.

[00164] Em outra modalidade preferida, todos os monômeros (isto é, 10, 9, 8 ou 7 dos monômeros) são monômeros mutantes da invenção e pelo menos um deles difere dos outros. Em uma modalidade mais preferida, o poro

compreende oito monômeros mutantes da invenção e pelo menos um deles difere dos outros.

[00165] Em todas as modalidades discutidas anteriormente, um ou mais, tais como 2, 3, 4, 5, 6, 7, 8, 9 ou 10, dos monômeros mutantes são preferivelmente modificados quimicamente conforme discutido anteriormente. Poros preferidos (a) a (c) anteriores são preferivelmente modificados quimicamente por anexação de uma molécula a uma ou mais das cisteínas introduzidas.

[00166] Métodos para fabricar poros são discutidos com mais detalhes a seguir.

Poros contendo construto

[00167] A invenção também diz respeito a um poro compreendendo pelo menos um construto da invenção. Um construto da invenção compreende dois ou mais monômeros anexados covalentemente derivados de Msp. Em outras palavras, um construto deve conter mais que um monômero. O poro contém construtos suficientes e, se necessário, monômeros para formar o poro. Por exemplo, um poro octamérico pode compreender (a) dois construtos cada qual compreendendo quatro monômeros ou (b) um construto compreendendo dois monômeros e seis monômeros que não formam parte de um construto. Pelo menos dois dos monômeros no poro são na forma de um construto da invenção. Os monômeros podem ser de qualquer tipo. O poro tipicamente compreende 7, 8, 9 ou 10 monômeros no total (pelo menos dois dos quais devem ser em um construto). O poro preferivelmente compreende oito monômeros (pelo menos dois dos quais devem ser em um construto).

[00168] Um poro tipicamente contém (a) um construto compreendendo dois monômeros e (b) 5, 6, 7 ou 8 monômeros. O construto pode ser qualquer daqueles discutidos anteriormente. Os monômeros podem ser qualquer daqueles discutidos anteriormente, incluindo monômeros mutantes da invenção.

[00169] Outro poro típico compreende mais que um construto da invenção, tal como dois, três ou quatro construtos da invenção. Poros como esses compreendem adicionalmente monômeros suficientes para formar o poro. O monômero pode ser qualquer daqueles discutidos anteriormente. Um poro adicional da invenção compreende apenas construtos compreendendo 2 monômeros, por exemplo, um poro pode compreender 4, 5, 6, 7 ou 8 construtos compreendendo 2 monômeros. Um poro específico de acordo com as invenções compreende quatro construtos cada qual compreendendo dois monômeros. Os construtos podem oligomerizar em um poro com uma estrutura de maneira tal que apenas um monômero de um construto contribua para o barril ou vestíbulo do poro. Tipicamente, os outros monômeros do construto serão de fora do barril ou vestíbulo do poro. Por exemplo, poros da invenção podem compreender 5, 6, 7 ou 8 construtos compreendendo 2 monômeros onde o barril ou vestíbulo compreende 8 monômeros.

[00170] Mutações podem ser introduzidas no construto conforme descrito anteriormente. As mutações podem ser alternadas, isto é, as mutações são diferentes para cada monômero em um dos construtos do monômero e os construtos são montados como um homo-oligômero resultando em modificações alternadas. Em outras palavras, monômeros compreendendo MutA e MutB são fundidos e montados para formar um poro A-B:A-B:A-B:A-B. Alternativamente, as mutações podem ser vizinhas, isto é, mutações idênticas são introduzidas em dois monômeros em um construto e isto é em seguida oligomerizado com diferentes monômeros mutantes. Em outras palavras, monômeros compreendendo MutA são fundidos seguido por oligomerização com monômeros contendo MutB para formar A-A:B:B:B:B:B:B.

[00171] Um ou mais dos monômeros da invenção em um poro contendo construto podem ser modificados quimicamente conforme discutido anteriormente.

Métodos de identificar um nucleotídeo individual

[00172] A presente invenção também diz respeito a métodos de caracterizar um nucleotídeo individual. Os métodos compreendem colocar o nucleotídeo em contato com um poro da invenção de maneira que o nucleotídeo interaja com o poro e medir a corrente que passa através do poro durante a interação e, por meio disso, caracterizar o nucleotídeo. A invenção, portanto, envolve a detecção de nanoporo de um nucleotídeo individual. A invenção também diz respeito a métodos de identificar um nucleotídeo individual compreendendo medir a corrente que passa através do poro durante a interação e, por meio disso, determinar a identidade do nucleotídeo. Qualquer um dos poros da invenção pode ser usado. O poro da invenção é preferivelmente modificado quimicamente com um adaptador molecular conforme discutido anteriormente.

[00173] O nucleotídeo está presente se a corrente passar através do poro de uma maneira específica para o nucleotídeo (isto é, se uma corrente distintiva associada com o nucleotídeo que passa através do poro for detectada). O nucleotídeo está ausente se a corrente não passar através do poro de uma maneira específica para o nucleotídeo.

[00174] A invenção pode ser usada para diferenciar nucleotídeos de estrutura similar com base nos diferentes efeitos que eles têm na corrente que passa através de um poro. Nucleotídeos individuais podem ser identificados no nível de molécula único da sua amplitude de corrente quando eles interagem com o poro. A invenção pode também ser usada para determinar se um nucleotídeo particular está ou não presente em uma amostra. A invenção pode também ser usada para medir a concentração de um nucleotídeo particular em uma amostra.

[00175] Os métodos podem ser realizados usando qualquer sistema de membrana/poro adequado no qual um poro da invenção é inserido em uma membrana. Os métodos são tipicamente realizados usando (i) uma membrana

artificial compreendendo um poro da invenção, (ii) uma membrana de ocorrência natural isolada compreendendo um poro da invenção, ou (iii) uma célula que expressa um poro que foi modificado de acordo com a invenção. Os métodos são preferivelmente realizados usando uma membrana artificial. A membrana pode compreender outras proteínas transmembranas e/ou intramembranas bem como outras moléculas além do poro da invenção.

[00176] A membrana forma uma barreira até o fluxo de íons, nucleotídeos e ácidos nucleicos. Qualquer membrana pode ser usada de acordo com a invenção. Membranas adequadas são bem conhecidas na tecnologia. A membrana é preferivelmente uma camada anfifílica. Uma camada anfifílica é uma camada formada das moléculas anfifílicas, tais como fosfolipídios, que têm propriedades tanto hidrofílicas quanto lipofílicas. Os anfifilos podem ser sintéticos ou de ocorrência natural. A camada anfifílica pode ser uma monocamada ou uma bicamada. Anfifilos de ocorrência não natural e anfifilos que formam uma monocamada são conhecidos na tecnologia e incluem, por exemplo, copolímeros blocos (*Gonzalez-Perez et al*, *Langmuir*, **2009**, 25, 10447-10450).

[00177] A membrana pode ser uma bicamada de lipídio. Bicamadas de lipídio adequadas para uso de acordo com a invenção podem ser feitas usando métodos conhecidos na tecnologia. Por exemplo, membranas de bicamada de lipídio podem ser formadas usando o método de Montal e Mueller (1972). Bicamadas de lipídio podem também ser formadas usando o método descrito no pedido internacional No. PCT/GB08/000563.

[00178] O método da invenção pode ser realizado usando bicamadas de lipídio formadas de qualquer lipídio da membrana incluindo, mas sem limitações, fosfolipídios, glicolipídios, colesterol, ácido micólico e misturas destes. Qualquer um dos lipídios descritos no pedido internacional No. PCT/GB08/000563 pode ser usado.

[00179] Em outra modalidade preferida, a membrana é uma camada

em estado sólido. Uma camada em estado sólido não é de origem biológica. Em outras palavras, uma camada em estado sólido não é derivada de um ambiente biológico ou isolado dele, tal como um organismo ou célula, ou uma versão fabricada sinteticamente de uma estrutura biologicamente disponível. Camadas em estado sólido podem ser formadas de materiais tanto orgânicos quanto inorgânicos incluindo, mas sem limitações, materiais microeletrônicos, materiais isolantes tais como Si_3N_4 , Al_2O_3 , e SiO , polímeros orgânicos e inorgânicos tal como poliamida, plásticos tal como Teflon® ou elastômeros tal como borracha de silicone de cura por adição de dois componentes, e vidros. A camada em estado sólido pode ser formada de camadas monoatômicas, tal como grafeno, ou camadas que têm apenas alguns átomos de espessura. Camadas de grafeno adequadas são reveladas no pedido internacional No. PCT/US2008/010637 (publicado como WO 2009/035647). Uma camada anfifílica pode ser formada através de um poro em estado sólido. Isto pode ser descrito na tecnologia como formação de poro híbrido (*Hall et al*, Nat Nanotechnol., **2010**, 5, 874-877).

[00180] Métodos são conhecidos na tecnologia para inserir poros nas membranas, tais como bicamadas de lipídio. Por exemplo, o poro pode ser suspenso em uma forma purificada em uma solução contendo uma bicamada de lipídio de maneira tal que ele difunda na bicamada de lipídio e seja inserido ligando na bicamada de lipídio e montando em um estado funcional. Alternativamente, o poro pode ser diretamente inserido na membrana usando o método “pegar e colocar” descrito em M.A. Holden, H. Bayley. *J. Am. Chem. Soc.* 2005, 127, 6502-6503 e Pedido internacional No. PCT/GB2006/001057 (publicado como WO 2006/100484).

[00181] Os métodos da invenção são tipicamente realizados *in vitro*.
Nucleotídeo individual

[00182] Um nucleotídeo individual é um nucleotídeo único. Um nucleotídeo individual é um que não é ligado a outro nucleotídeo ou ácido

nucleico por uma ligação de nucleotídeo. Uma ligação de nucleotídeo envolve um dos grupos de fosfato de um nucleotídeo sendo ligado ao grupo de açúcar de outro nucleotídeo. Um nucleotídeo individual é tipicamente um que não é ligado por uma ligação de nucleotídeo a outra sequência de ácido nucleico de pelo menos 5, pelo menos 10, pelo menos 20, pelo menos 50, pelo menos 100, pelo menos 200, pelo menos 500, pelo menos 1.000 ou pelo menos 5.000 nucleotídeos. Por exemplo, o nucleotídeo individual foi digerido de uma sequência de polinucleotídeo alvo, tal como uma fita de DNA ou RNA.

[00183] Os métodos da invenção podem ser usados para identificar qualquer nucleotídeo. O nucleotídeo pode ser de ocorrência natural ou artificial. Um nucleotídeo tipicamente contém uma nucleobase, um açúcar e pelo menos um grupo fosfato. A nucleobase é tipicamente heterocíclico. Nucleobases adequadas incluem purinas e pirimidinas e mais especificamente adenina, guanina, timina, uracila e citosina. O açúcar é tipicamente um açúcar pentose. Açúcares adequados incluem, mas sem limitações, ribose e deoxirribose. O nucleotídeo é tipicamente um ribonucleotídeo ou deoxirribonucleotídeo. O nucleotídeo tipicamente contém um monofosfato, difosfato ou trifosfato.

[00184] Nucleotídeos adequados incluem, mas sem limitações, monofosfato de adenosina (AMP), difosfato de adenosina (ADP), trifosfato de adenosina (ATP), monofosfato de guanosina (GMP), difosfato de guanosina (GDP), trifosfato de guanosina (GTP), monofosfato de timidina (TMP), difosfato de timidina (TDP), trifosfato de timidina (TTP), monofosfato de uridina (UMP), difosfato de uridina (UDP), trifosfato de uridina (UTP), monofosfato de citidina (CMP), difosfato de citidina (CDP), trifosfato de citidina (CTP), monofosfato de adenosina cíclica (cAMP), monofosfato de guanosina cíclica (cGMP), monofosfato de deoxiadenosina (dAMP), difosfato de deoxiadenosina (dADP), trifosfato de deoxiadenosina (dATP),

monofosfato de deoxiguanosina (dGMP), difosfato de deoxiguanosina (dGDP), trifosfato de deoxiguanosina (dGTP), monofosfato de deoxitimidina (dTMP), difosfato de deoxitimidina (dTDP), trifosfato de deoxitimidina (dTTP), monofosfato de deoxiuridina (dUMP), difosfato de deoxiuridina (dUDP), trifosfato de deoxiuridina (dUTP), monofosfato de deoxicitidina (dCMP), difosfato de deoxicitidina (dCDP) e trifosfato de deoxicitidina (dCTP). O nucleotídeo é preferivelmente AMP, TMP, GMP, UMP, dAMP, dTMP, dGMP ou dCMP.

[00185] O nucleotídeo pode ser derivado da digestão de uma sequência de ácido nucleico tais como ácido ribonucleico (R A) ou ácido deoxiribonucleico. Sequências de ácido nucleico podem ser digeridas usando qualquer método conhecido na tecnologia. Métodos adequados incluem, mas sem limitações, aqueles que usam enzimas ou catalisadores. Digestão catalítica de ácidos nucleicos é revelada em *Deck et al.*, *Inorg. Chem*, 2002; 41 : 669-677.

[00186] Nucleotídeos individuais de uma única sequência de ácido nucleico podem ser colocados em contato com o poro de uma maneira sequencial a fim de sequenciar todo o ácido nucleico ou parte dele.

[00187] Sequenciamento de ácidos nucleicos é discutido com mais detalhes a seguir.

[00188] O nucleotídeo é tipicamente não modificado, tal como quando o nucleotídeo é derivado da digestão de uma sequência de ácido nucleico. Alternativamente, o nucleotídeo pode ser modificado ou danificado. O nucleotídeo é tipicamente metilado ou oxidizado. O nucleotídeo pode ser marcado com um rótulo de revelação. O rótulo de revelação pode ser qualquer rótulo adequado que permite que o nucleotídeo seja detectado. Rótulos adequados incluem moléculas fluorescentes, radioisótopos, por exemplo, ¹²⁵I, ³⁵S, e ligantes tal como biotina.

[00189] O nucleotídeo está tipicamente presente em qualquer amostra

biológica adequada. Amostras biológicas adequadas são discutidas anteriormente.

Interação entre o poro e nucleotídeo

[00190] O nucleotídeo pode ser colocado em contato com o poro em qualquer lado da membrana. O nucleotídeo pode ser introduzido no poro em qualquer lado da membrana. O nucleotídeo pode ser colocado em contato com o lado da membrana que permite que o nucleotídeo passe através do poro para o outro lado da membrana. Por exemplo, o nucleotídeo é colocado em contato com uma extremidade do poro, que em seu ambiente nativo permite a entrada de íons ou molécula pequenas, tais como nucleotídeos, no barril ou canal do poro de maneira tal que o nucleotídeo pode passar através do poro. Em tais casos, o nucleotídeo interage com o poro e/ou adaptador a medida que ele passa através da membrana através do barril ou canal do poro. Alternativamente, o nucleotídeo pode ser colocado em contato com o lado da membrana que permite que o nucleotídeo interaja com o poro por meio de adaptador ou em junto com ele, dissocie do poro e permaneça do mesmo lado da membrana. A presente invenção diz respeito a poros nos quais a posição do adaptador é fixa. Em decorrência, o nucleotídeo é preferivelmente colocado em contato com a extremidade do poro que permite que o adaptador interaja com o nucleotídeo.

[00191] O nucleotídeo pode interagir com o poro de qualquer maneira e em qualquer lugar. Conforme discutido anteriormente, o nucleotídeo preferivelmente se liga reversivelmente ao poro por meio de adaptador ou junto com ele. O nucleotídeo acima de tudo preferivelmente se liga reversivelmente ao poro por meio de adaptador ou junto com ele a medida que ele passa através do poro através da membrana. O nucleotídeo pode também se ligar reversivelmente ao barril ou canal do poro por meio de adaptador ou junto com ele a medida que ele passa através do poro através da membrana.

[00192] Durante a interação entre o nucleotídeo e o poro, o nucleotídeo afeta a corrente que passa através do poro de uma maneira específica para este nucleotídeo. Por exemplo, um nucleotídeo particular reduzirá a corrente que passa através do poro por um período de tempo médio particular e a uma extensão particular. Em outras palavras, a corrente que passa através do poro é distintiva para um nucleotídeo particular. Experimentos controles podem ser realizados para determinar o efeito que um nucleotídeo particular tem na corrente que passa através do poro. Resultados de realizar o método da invenção em uma amostra teste podem em seguida ser comparados com aqueles derivados de um experimento controle como esse a fim de identificar um nucleotídeo particular na amostra ou determinar se um nucleotídeo particular está presente na amostra. A frequência na qual a corrente que passa através do poro é afetada de uma maneira indicativa de um nucleotídeo particular pode ser usada para determinar a concentração deste nucleotídeo na amostra. A razão de diferentes nucleotídeos em uma amostra pode também ser calculada. Por exemplo, a razão de dCMP para metil-dCMP pode ser calculada.

Aparelho

[00193] Os métodos podem ser realizados usando qualquer um aparelho que seja adequado para investigar um sistema de membrana/poro no qual um poro da invenção é inserido em uma membrana. O método pode ser realizado usando qualquer um aparelho que seja adequado para a detecção de nanoporo. Por exemplo, um aparelho compreende uma câmara compreendendo uma solução aquosa e uma barreira que separa a câmara em duas seções. A barreira tem uma abertura na qual a membrana contendo o poro é formada. O nucleotídeo pode ser colocado em contato com o poro introduzindo o nucleotídeo na câmara. O nucleotídeo pode ser introduzido em qualquer das duas seções da câmara.

[00194] Os métodos podem ser realizados usando um aparelho descrito

no pedido internacional No. PCT/GB08/000562.

[00195] Os métodos da invenção envolvem medir a corrente que passa através do poro durante a interação com o nucleotídeo. Portanto, um aparelho também compreende um circuito elétrico capaz de aplicar um potencial e medir um sinal elétrico através da membrana e poro. Os métodos podem ser realizados usando um patch clamp ou uma voltage clamp. Os métodos preferivelmente envolvem o uso de uma voltage clamp.

Amostra

[00196] O nucleotídeo está presente em qualquer amostra adequada. A invenção é tipicamente realizada em uma amostra que é conhecida para conter ou suspeita de conter o nucleotídeo. A invenção pode ser realizada em uma amostra que contém um ou mais nucleotídeos cuja identidade é desconhecida. Alternativamente, a invenção pode ser realizada em uma amostra para confirmar a identidade de um ou mais nucleotídeos cuja presença na amostra é conhecida ou esperada.

[00197] A amostra pode ser uma amostra biológica. A invenção pode ser realizada *in vitro* em uma amostra obtida ou extraída de qualquer organismo ou micro-organismo. O organismo ou micro-organismo é tipicamente procariótico ou eucariótico e tipicamente pertence a um dos cinco reinos: plantae, animalia, fungi, monera e protista. A invenção pode ser realizada *in vitro* em uma amostra obtida ou extraída de qualquer vírus. A amostra é preferivelmente uma amostra fluida. A amostra tipicamente compreende um fluido corpóreo do paciente. A amostra pode ser urina, linfa, saliva, muco ou fluido amniótico, mas, é preferivelmente sangue, plasma ou soro. Tipicamente, a amostra é de origem humana, mas, alternativamente, ela pode ser de outro animal mamífero tal como de animais criados comercialmente em fazenda tais como cavalos, gado, ovelhas ou porcos ou pode alternativamente ser animais domésticos tal como gatos ou cães. Alternativamente uma amostra de origem de planta é tipicamente obtida de

um colheita comercial, tais como um cereal, legume, fruta ou vegetal, por exemplo, trigo, cevada, aveias, canola, milho, soja, arroz, bananas, maçãs, tomates, batatas, uvas, tabaco, feijão, ervilhas, cana de açúcar, cacau, algodão, chá, café.

[00198] A amostra pode ser uma amostra não biológica. A amostra não biológica é preferivelmente uma amostra fluida. Exemplos de uma amostra não biológica incluem fluidos cirúrgicos, água tal como água potável, água do mar ou água do rio, e reagentes para testes de laboratório.

[00199] A amostra é tipicamente processada antes de ser ensaiada, por exemplo, por centrifugação ou por passagem através de uma membrana que filtra moléculas ou células indesejadas, tais como células de sangue vermelho. A amostra pode ser medida imediatamente ao ser tirada. A amostra pode também ser tipicamente armazenada antes do ensaio, preferivelmente abaixo de -70°C .

Condições

[00200] Os métodos da invenção envolvem a medição de uma corrente que passa através do poro durante interação com o nucleotídeo. As condições adequadas para medir correntes iônicas através de poros da proteína transmembrana são conhecidas na tecnologia e reveladas no exemplo. O método é realizado com uma tensão aplicada através da membrana e poro. A tensão usada é tipicamente de -400 mV a $+400\text{ mV}$. A tensão usada é preferivelmente em uma faixa com um limite inferior selecionado de -400 mV , -300 mV , -200 mV , -150 mV , -100 mV , -50 mV , -20 mV e 0 mV e um limite superior independentemente selecionado de $+10\text{ mV}$, $+20\text{ mV}$, $+50\text{ mV}$, $+100\text{ mV}$, $+150\text{ mV}$, $+200\text{ mV}$, $+300\text{ mV}$ e $+400\text{ mV}$. A tensão usada é mais preferivelmente na faixa de 100 mV a 240 mV e acima de tudo preferivelmente na faixa de 160 mV a 240 mV . É possível aumentar a discriminação entre diferentes nucleotídeos por um poro da invenção usando um maior potencial aplicado.

[00201] Os métodos são tipicamente realizados na presença de qualquer sal de cloreto de metal alcalino. No aparelho exemplar discutido anteriormente, o sal está presente na solução aquosa na câmara. Cloreto de potássio (KCl), cloreto de sódio (NaCl) ou cloreto de céscio (CsCl) são tipicamente usados. KCl é preferido. A concentração de sal é tipicamente de 0,1 a 2,5 M, de 0,3 a 1,9 M, de 0,5 a 1,8 M, de 0,7 a 1,7 M, de 0,9 a 1,6 M ou de 1 M a 1,4 M. A concentração de sal é preferivelmente de 10 mM a 1 M. Alta concentração de sais fornecem uma alta razão sinal para ruído e permite que correntes indicativas da presença de um nucleotídeo sejam identificadas contra o fundo de flutuações de corrente normais. Concentração mais baixa de sais pode ser usada se detecção de nucleotídeo for realizada na presença de uma enzima, tal como durante sequenciamento de ácidos nucleicos. Isto é discutido com mais detalhes a seguir.

[00202] Os métodos são tipicamente realizados na presença de um tampão. No aparelho exemplar discutido anteriormente, o tampão está presente na solução aquosa na câmara. Qualquer tampão pode ser usado no método da invenção. Um tampão adequado é tampão Tris-HCl. Os métodos são tipicamente realizados a um pH de 4,0 a 12,0, de 4,5 a 10,0, de 5,0 a 9,0, de 5,5 a 8,8, de 6,0 a 8,7 ou de 7,0 a 8,8 ou 7,5 a 8,5. O pH usado é preferivelmente cerca de 7,5.

[00203] Os métodos são tipicamente realizados de 0°C a 100°C, de 15°C a 95°C, de 16°C a 90°C, de 17°C a 85°C, de 18°C a 80°C, 19°C a 70°C, ou de 20°C a 60°C. Os métodos podem ser realizados à temperatura ambiente. Os métodos são preferivelmente realizados a uma temperatura que suporta função enzimática, tal como cerca de 37°C.

Métodos de caracterizar ácidos nucleicos

[00204] A presente invenção também diz respeito a métodos de caracterizar uma sequência de ácido nucleico alvo. Uma ou mais características da sequência de ácido nucleico alvo podem ser determinadas.

O método pode envolver medir dois, três, quatro ou cinco ou mais características da sequência de ácido nucleico alvo. Uma ou mais características são preferivelmente selecionadas (i) do comprimento da sequência de ácido nucleico alvo, (ii) da identidade da sequência de ácido nucleico alvo, (iii) da sequência da sequência de ácido nucleico alvo, (iv) da estrutura secundária da sequência de ácido nucleico alvo e (v) se a sequência de ácido nucleico alvo é ou não modificada. Qualquer combinação de (i) a (v) pode ser determinada de acordo com a invenção.

[00205] Para (i), o comprimento da sequência de ácido nucleico pode ser medido usando o número de interações entre a sequência de ácido nucleico alvo e o poro.

[00206] Para (ii), a identidade da sequência de ácido nucleico pode ser medida de inúmeras maneiras.

[00207] A identidade da sequência de ácido nucleico pode ser medida junto com medições da sequência da sequência de ácido nucleico alvo ou sem medições da sequência da sequência de ácido nucleico alvo. O primeiro é direto; o ácido nucleico é sequenciado e, por meio disso, identificado. O último pode ser feito de diversas maneiras. Por exemplo, a presença de um motivo particular na sequência de ácido nucleico pode ser medida (sem medir a sequência remanescente do polinucleotídeo). Alternativamente, as medições de um sinal elétrico particular no método pode identificar a sequência de ácido nucleico alvo proveniente de uma fonte particular.

[00208] Para (iii), a sequência da sequência de ácido nucleico pode ser determinada conforme descrito previamente. Métodos de sequenciamento adequados, particularmente aqueles usando medições elétricas, são descritos em 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, e pedido internacional WO 2000/28312.

[00209] Para (iv), a estrutura secundário pode ser medida em uma

variedade de maneiras. Por exemplo, a estrutura secundária pode ser medida usando uma mudança em tempo de permanência ou uma mudança na corrente que passa através do poro.

[00210] A invenção também diz respeito a um método de estimar a sequência de uma sequência de ácido nucleico alvo. A invenção adicionalmente diz respeito a um método de sequenciamento uma sequência de ácido nucleico alvo.

[00211] Um ácido nucleico é uma macromolécula compreendendo dois ou mais nucleotídeos. Os nucleotídeos podem ser qualquer daqueles discutidos anteriormente.

[00212] Em uma modalidade, o método compreende (a) colocar a sequência alvo em contato com um poro da invenção e uma proteína de ligação de ácido nucleico de maneira que a proteína controle o movimento da sequência alvo através do poro e uma proporção dos nucleotídeos na sequência alvo interage com o poro e (b) medir a corrente que passa através do poro durante cada interação e, por meio disso, caracterizar, tal como estimar, a sequência ou sequenciamento da sequência alvo. Conseqüentemente, o método envolve a detecção de nanoporo de uma proporção dos nucleotídeos em uma sequência de ácido nucleico alvo a medida que os nucleotídeos passam através do barril ou canal a fim de caracterizar, tal como sequenciar, a sequência alvo.

[00213] Em outra modalidade, o método compreende (a) colocar a sequência alvo em contato com um poro da invenção e uma exonuclease de maneira tal que a exonuclease digira um nucleotídeo individual de uma extremidade da sequência alvo; (b) colocar o nucleotídeo em contato com o poro de maneira que o nucleotídeo interaja com o adaptador; (c) medir a corrente que passa através do poro durante a interação e, por meio disso, caracterizar o nucleotídeo; e (d) repetir as etapas (a) a (c) na mesma extremidade da sequência alvo e, por meio disso, caracterizar a sequência

alvo. Conseqüentemente, o método envolve a detecção de nanoporo de uma proporção dos nucleotídeos em uma seqüência de ácido nucleico alvo de uma maneira sucessiva a fim de caracterizar a seqüência alvo. Em uma modalidade preferida, o método diz respeito ao sequenciamento da seqüência de ácido nucleico alvo e a etapa (a) compreende determinar a identidade do nucleotídeo. Nucleotídeos individuais são descritos anteriormente.

[00214] Os poros da invenção são particularmente adequados para esses métodos em virtude de eles exibirem melhor discriminação de nucleotídeo. Em particular, eles apresentam uma maior faixa de corrente, que torna fácil discriminá-los entre diferentes nucleotídeos, e uma menor variância de estados, que aumenta a razão de sinal para ruído. Além do mais, em relação à primeira modalidade, o número de nucleotídeos contribuindo para a corrente a medida que o ácido nucleico move através do poro é diminuído. Isto torna mais fácil identificar um relacionamento direto entre a corrente observada a medida que o ácido nucleico move através do poro e da seqüência de ácido nucleico. Os poros da invenção são preferivelmente modificados quimicamente com (1) um adaptador molecular e/ou (2) a proteína de ligação de ácido nucleico ou exonuclease conforme discutido anteriormente.

[00215] Toda ou apenas parte da seqüência de ácido nucleico alvo pode ser caracterizada, tal como sequenciada, usando este método. A seqüência de ácido nucleico pode ter qualquer comprimento. Por exemplo, a seqüência de ácido nucleico pode ter pelo menos 10, pelo menos 50, pelo menos 100, pelo menos 150, pelo menos 200, pelo menos 250, pelo menos 300, pelo menos 400 ou pelo menos 500 nucleotídeos de comprimento. A seqüência de ácido nucleico pode ter 1.000 ou mais nucleotídeos ou 5.000 ou mais nucleotídeos de comprimento. A seqüência de ácido nucleico pode ser de ocorrência natural ou artificial. Por exemplo, o método pode ser usado para verificar a seqüência de um oligonucleotídeo fabricado. Os métodos são

tipicamente realizados *in vitro*.

[00216] Os métodos podem ser realizados usando qualquer sistema de membrana/poro adequado no qual um poro é inserido em uma membrana. Os métodos são tipicamente realizados usando qualquer um dos sistemas, aparelho ou condições reveladas anteriormente.

[00217] Conforme mencionado anteriormente, boa discriminação de nucleotídeo pode ser obtida a baixa concentração de sais se a temperatura for aumentada. Além de aumentar a temperatura da solução, existem inúmeras outras estratégias que podem ser empregadas para aumentar a condutância da solução, mantendo ao mesmo tempo condições que são adequadas para atividade enzimática. Uma estratégia como essa é para usar a bicamada de lipídio para dividir duas diferentes concentrações de solução de sal, uma baixa concentração de sal de sal no lado da enzima e uma alta concentração no lado oposto. Um exemplo desta abordagem é usar 200 mM de KCl no lado cis da membrana e 500 mM de Cl na câmara trans. Nestas condições, espera-se que a condutância através do poro seja grosseiramente equivalente a 400 mM de Cl em condições normais, e a enzima apenas experimente 200 mM se colocada no lado cis. Outro possível benefício de usar condições de sal assimétricas é o gradiente osmótico induzido através do poro. Este fluxo líquido de água pode ser usado para empurrar nucleotídeos no poro para detecção. Um efeito similar pode ser obtido usando um osmólito neutro, tais como sacarose, glicerol ou PEG. Outra possibilidade é usar uma solução com níveis relativamente de baixos de KCl e contarem com uma carga adicional carregando espécie que seja menos disruptiva para atividade enzimática.

[00218] A sequência alvo sendo analisada pode ser combinada com produtos químicos de proteção conhecidos para impedir que a sequência sofra ação da proteína de ligação ou exonuclease enquanto na solução volumosa. O poro pode em seguida ser usado para remover o produto químico de proteção. Isto pode ser obtido tanto usando grupos de proteção que são não hibridizados

pelo poro, proteína de ligação ou enzima sob um potencial aplicado (WO 2008/124107) ou usando produtos químicos de proteção que são removidos pela proteína de ligação ou enzima quando mantidos em proximidade imediata do poro (*J Am Chem Soc.* 2010 Dec 22;132(50):17961-72).

[00219] *Sequenciamento de fita*

[00220] Sequenciamento de fita envolve a translocação controlada e em etapas de polímeros do ácido nucleico através de um poro. Poros da invenção podem ser usados em sequenciamento de fita. Um método da invenção usa uma proteína de ligação de ácido nucleico para controlar o movimento da sequência alvo através do poro. Exemplos de tais proteínas incluem, mas sem limitações, enzimas de manuseio de ácido nucleico, tal como nucleases, polimerases, topoisomerases, ligases e helicases, e proteínas de ligação não catalíticas tais como aquelas classificadas por SCOP (Structural Classification of Proteins) sob a superfamília de proteína de ligação de ácido nucleico (50249). A proteína de ligação pode ser proteína de ligação de fita simples (SSB).

[00221] Um ácido nucleico é uma macromolécula compreendendo dois ou mais nucleotídeos. O ácido nucleico ligado pela proteína pode compreender qualquer combinação de qualquer nucleotídeo. Os nucleotídeos podem ser qualquer daqueles discutidos anteriormente. O ácido nucleico pode ser ácido deoxiribonucleico (DNA) ou ácido ribonucleico (RNA). O ácido nucleico pode ser qualquer ácido nucleico sintético conhecido na tecnologia, tal como ácido nucleico peptídeo (PNA), ácido nucleico glicerol (GNA), ácido nucleico treose (TNA), ácido nucleico bloqueado (LNA) ou outros polímeros sintéticos com cadeias laterais de nucleotídeo. O ácido nucleico ligado pela proteína pode ser fita simples, tal como DNAc, RNA, GNA, TNA ou LNA, ou de fita dupla, tal como DNA. Proteínas que ligam ácidos nucleicos de fita simples podem ser usadas para sequenciar DNA de fita dupla desde que o DNA de fita dupla seja dissociado em uma fita simples antes de

ele se ligado pela proteína.

[00222] A proteína de ligação de ácido nucleico é preferivelmente uma enzima de manuseio de ácido nucleico. Uma enzima de manuseio de ácido nucleico é um polipeptídeo que é capaz de interagir com e modificar pelo menos uma propriedade de um ácido nucleico. A enzima pode modificar o ácido nucleico clivando-o para formar nucleotídeos individuais ou cadeias curtas de nucleotídeos, tais como di- ou trinucleotídeos. A enzima pode modificar o ácido nucleico orientando-o ou movendo-o para uma posição específica. A enzima de manuseio de ácido nucleico não precisa exibir atividade enzimática desde que ela seja capaz de ligar a sequência alvo e controlar seu movimento através do poro. Por exemplo, a enzima pode ser modificada para remover sua atividade enzimática ou pode ser usada em condições que a impede de agir como uma enzima. Tais condições são discutidas com mais detalhes a seguir.

[00223] A enzima de manuseio de ácido nucleico é preferivelmente derivada de uma enzima nucleolítica. A enzima de manuseio de ácido nucleico usada no construto da enzima é mais preferivelmente derivada de um membro de qualquer um dos grupos de classificação enzimática (EC) 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 e 3.1.31. A enzima pode ser qualquer uma daquelas reveladas no pedido internacional No. PCT/GB 10/000133 (publicado como WO 2010/086603).

[00224] Enzimas preferidas são polimerases, exonucleases, helicases e topoisomerasas, tais como girases. Enzimas adequadas incluem, mas sem limitações, exonuclease I de *E. coli* (SEQ ID NO: 6), enzima exonuclease III de *E. coli* (SEQ ID NO: 8), RecJ de *T. thermophilus* (SEQ ID NO: 10) e bacteriófago lambda exonuclease (SEQ ID NO: 12) e suas variantes. Três subunidades compreendendo a sequência mostrada na SEQ ID NO: 10 ou uma variante desta interação para formar um trímero exonuclease. A enzima é preferivelmente com base na DNA polimerase Phi29 (SEQ ID NO: 4).

[00225] Uma variante da SEQ ID NOs: 4, 6, 8, 10 ou 12 é uma enzima que tem uma sequência de aminoácido que varia daquela da SEQ ID NO: 4, 6, 8, 10 ou 12 e que retém a capacidade de ligação a nucleico. A variante pode incluir modificações que facilitam a ligação do ácido nucleico e/ou facilitam sua atividade a alta concentração de sais e/ou temperatura ambiente.

[00226] Em todo o comprimento da sequência de aminoácido da SEQ ID NO: 4, 6, 8, 10 ou 12, uma variante será preferivelmente pelo menos 50 % homóloga à sequência com base na identidade do aminoácido. Mais preferivelmente, a polipeptídeo variante pode ser pelo menos 55 %, pelo menos 60 %, pelo menos 65 %, pelo menos 70 %, pelo menos 75 %, pelo menos 80 %, pelo menos 85 %, pelo menos 90 % e mais preferivelmente pelo menos 95 %, 97 % ou 99 % homóloga com base na identidade do aminoácido à sequência de aminoácido da SEQ ID NO: 4, 6, 8, 10 ou 12 em toda a sequência. Pode existir pelo menos 80 %, por exemplo, pelo menos 85 %, 90 % ou 95 % de identidade do aminoácido em um trecho de 200 ou mais, por exemplo, 230, 250, 270 ou 280 ou mais, aminoácidos contíguos (“homologia difícil”). Homologia é determinada conforme descrito anteriormente. A variante pode diferir da sequência tipo selvagem de qualquer uma das maneiras discutidas anteriormente com referência à SEQ ID NO: 2. A enzima pode ser anexada covalentemente ao poro conforme discutido anteriormente.

[00227] Não é exigido que a enzima esteja em uma proximidade tão imediata do lúmen do poro como para o sequenciamento do nucleotídeo individual já que não existe potencial para desordem na série na qual os nucleotídeos atingem a fração de proteção do poro.

[00228] As duas estratégias para sequenciamento de DNA de fita simples são a translocação do DNA através do nanoporo, tanto cis para trans e trans para cis, quanto com ou contra um potencial aplicado. O mecanismo mais vantajoso para sequenciamento de fita é a translocação controlada de fita simples DNA através do nanoporo sob um potencial aplicado.

[00229] Exonucleases que agem progressiva ou processivamente na DNA de fita dupla podem ser usadas no lado cis do poro para alimentar a fita simples remanescente através de um potencial aplicado ou o translado sob um potencial reverso. Igualmente, um helicase que desenrola do DNA de fita dupla pode também ser usado de uma maneira similar. Existem também possibilidades para sequenciar aplicações que exigem translocação da fita contra um potencial aplicado, mas o DNA deve ser primeiro “pego” pela enzima sob um potencial reverso ou nenhum potencial. Com o potencial em seguida mudado de volta após a ligação a fita passará de cis para trans através do poro e será mantida em uma conformação estendida pela passagem da corrente. As exonucleases de DNA de fita simples ou polimerases dependentes de DNA de fita simples podem agir como motores moleculares para empurrar a fita simples recentemente translocada de volta através do poro de uma maneira controlada em etapas, trans para cis, contra o potencial aplicado.

Métodos baseados em exonuclease

[00230] Em uma modalidade, o método para caracterizar uma sequência de ácido nucleico alvo envolve colocar a sequência alvo em contato com uma enzima exonuclease. Qualquer das enzimas exonucleases discutidas anteriormente podem ser usadas no método. Os nucleotídeos individuais liberam exonuclease a partir de uma extremidade da sequência alvo. A enzima pode ser anexada covalentemente ao poro conforme discutido anteriormente.

[00231] Exonucleases são enzimas que tipicamente seguram uma extremidade de uma sequência de ácido nucleico e digerem a sequência um nucleotídeo por vez daquela extremidade. O exonuclease pode digerir o ácido nucleico na direção 5' a 3' ou direção 3' a 5'. A extremidade do ácido nucleico na qual o exonuclease se liga é tipicamente determinada através da escolha de enzima usada e/ou usando métodos conhecidos na tecnologia. Grupos hidroxila ou estruturas de tampa em qualquer extremidade da sequência de

ácido nucleico podem tipicamente ser usados para impedir ou facilitar a ligação da exonuclease em uma extremidade particular da sequência de ácido nucleico.

[00232] O método envolve colocar em contato a sequência de ácido nucleico com a exonuclease de maneira que os nucleotídeos sejam digeridos a partir da extremidade do ácido nucleico a uma taxa que permite a caracterização ou identificação de uma proporção de nucleotídeos conforme discutido anteriormente. Métodos para fazer isto são bem conhecidos na tecnologia. Por exemplo, degradação Edman é usada para digerir sucessivamente aminoácidos únicos da extremidade de polipeptídeo de maneira tal que eles possam ser identificados usando Cromatografia Líquida de Alto Desempenho (HPLC). Um método homólogo pode ser usado na presente invenção.

[00233] A taxa na qual a exonuclease funciona é tipicamente mais baixa do que a taxa ideal de um exonuclease tipo selvagem. Uma taxa adequada de atividade da exonuclease no método da invenção envolve digestão de 0,5 a 1.000 nucleotídeos por segundo, de 0,6 a 500 nucleotídeos por segundo, 0,7 a 200 nucleotídeos por segundo, de 0,8 a 100 nucleotídeos por segundo, de 0,9 a 50 nucleotídeos por segundo ou 1 a 20 ou 10 nucleotídeos por segundo. A taxa é preferivelmente 1, 10, 100, 500 ou 1.000 nucleotídeos por segundo. Uma taxa adequada de atividade exonuclease pode ser obtida de várias maneiras. Por exemplo, exonucleases variantes com uma taxa ideal reduzida de atividade podem ser usadas de acordo com a invenção.

Msp e DNA polimerase Phi29

[00234] Em uma modalidade preferida, caracterização, tal como sequenciamento de fita, é realizada usando um poro derivado de Msp e um DNA polimerase Phi29. O método compreende (a) colocar a sequência alvo em contato com um poro derivado de Msp e um DNA polimerase Phi29 de maneira tal que a polimerase controle o movimento da sequência alvo através

do poro e uma proporção dos nucleotídeos na sequência alvo interaja com o poro e (b) medir a corrente que passa através do poro durante cada interação e, por meio disso, caracterizar, tal como determinar a sequência, da sequência alvo, em que as etapas (a) e (b) são realizadas com uma tensão aplicada através do poro. Quando a sequência alvo é colocada em contato com uma DNA polimerase Phi29 e um poro derivado de Msp, a sequência alvo primeiramente forma um complexo com a DNA polimerase Phi29. Quando a tensão é aplicada através do poro, complexa sequência alvo/DNA polimerase Phi29 forma um complexo com o poro e controla o movimento da sequência alvo através do poro.

[00235] Esta modalidade tem três vantagens surpreendentes. Primeiramente, a sequência alvo move através do poro a uma taxa que é comercialmente viável e permite ainda sequenciamento efetivo. A sequência alvo move através do poro de Msp mais rapidamente do que ele através de um poro de hemolisina. Em segundo lugar, uma maior faixa de corrente é observada à medida que o ácido nucleico move através do poro que permitindo a sequência seja determinada mais facilmente. Em terceiro lugar, uma menor variância de corrente é observada quando o poro e polimerase específicos são usados juntos aumentando por meio disso a razão de sinal para ruído.

[00236] Qualquer sequência de ácido nucleico descrita anteriormente pode ser caracterizada ou sequenciada. Pelo menos uma porção da sequência de ácido nucleico é preferivelmente de fita dupla.

[00237] O poro pode ser qualquer um dos poros discutidos anteriormente. O poro é preferivelmente um poro da invenção. O poro pode compreender oito monômeros compreendendo a sequência mostrada na SEQ ID NO: 2, 16, 17 ou 18 ou uma variante desta. O poro não tem de incluir nenhuma das mutações da invenção.

[00238] DNA polimerase Phi29 tipo selvagem tem atividade

polimerase e exonuclease. Ele pode também abrir ácidos nucleicos de fita dupla nas condições corretas. Conseqüentemente, a enzima pode funcionar de três modos. Isto é discutido com mais detalhes a seguir.

[00239] O DNA polimerase Phi29 pode compreender a seqüência mostrada na SEQ ID NO: 4 ou uma variante desta. Uma variante da SEQ ID NOs: 4 é uma enzima que tem uma seqüência de aminoácido que varia daquela da SEQ ID NO: 4 e que retém atividade de ligação de ácido nucleico. A variante deve funcionar em pelo menos um dos três modos discutidos a seguir. Preferivelmente, a variante funciona em todos três modos. A variante pode incluir modificações que facilitam o manuseio do ácido nucleico e/ou facilitam sua atividade a alta concentração de sais e/ou temperatura ambiente.

[00240] Em todo o comprimento da seqüência de aminoácido da SEQ ID NO: 4, uma variante será preferivelmente pelo menos 40 % homóloga à seqüência com base na identidade do aminoácido. Mais preferivelmente, o polipeptídeo variante pode ser pelo menos 50 %, pelo menos 55 %, pelo menos 60 %, pelo menos 65 %, pelo menos 70 %, pelo menos 75 %, pelo menos 80 %, pelo menos 85 %, pelo menos 90 % e mais preferivelmente pelo menos 95 %, 97 % ou 99 % homólogo com base na identidade do aminoácido à seqüência de aminoácido da SEQ ID NO: 4 em toda a seqüência. Pode existir pelo menos 80 %, por exemplo, pelo menos 85 %, 90 % ou 95 % de identidade do aminoácido em um trecho de 200 ou mais, por exemplo, 230, 250, 270 ou 280 ou mais, aminoácidos contíguos (“homologia difícil”). Homologia é determinada conforme descrito anteriormente. A variante pode diferir da seqüência tipo selvagem de qualquer uma das maneiras discutidas anteriormente com referência à SEQ ID NO: 2. A enzima pode ser anexada covalentemente ao poro conforme discutido anteriormente.

[00241] Qualquer um dos sistemas, aparelho ou condições discutidos anteriormente podem ser usados de acordo com esta modalidade preferida. A concentração de sal é tipicamente de 0,15 M a 0,6 M. O sal é preferivelmente

KC1.

[00242] O método pode ser realizado em um dos três modos preferidos com base nos três modos da DNA polimerase Phi29. Cada modo inclui um método de leitura de prova da sequência. Primeiramente, o método é preferivelmente realizado usando a DNA polimerase P 29 como uma polimerase. Nesta modalidade, as etapas (a) e (b) são realizadas na presença de nucleotídeos livres e um cofator de enzima de maneira tal que a polimerase mova a sequência alvo através do poro contra o campo resultante da tensão aplicada. A sequência alvo move na direção 5' para 3'. Os nucleotídeos livres podem ser um ou mais de qualquer um dos nucleotídeos individuais discutidos anteriormente. O cofator de enzima é um fator que permite que a DNA polimerase Phi29 funcione tanto como uma polimerase quanto uma exonuclease. O cofator de enzima é preferivelmente um cátion metálico divalente. O cátion metálico divalente é preferivelmente Mg, Mn, Ca ou Co. O cofator de enzima é acima de tudo preferivelmente Mg²⁺. O método de preferência compreende adicionalmente (c) remover os nucleotídeos livres de maneira tal que a polimerase mova a sequência alvo através do poro com o campo resultante da tensão aplicada (isto é, na direção 3' para 5') e uma proporção dos nucleotídeos na sequência alvo interage com o poro e (d) medir a corrente que passa através do poro durante cada interação e, por meio disso, realizar a leitura de prova da sequência da sequência alvo obtida na etapa (b), em que as etapas (c) e (d) são também realizadas com uma tensão aplicada através do poro.

[00243] Em segundo lugar, o método é preferivelmente realizado usando a DNA polimerase Phi29 como uma exonuclease. Nesta modalidade, em que as etapas (a) e (b) são realizadas na ausência de nucleotídeos livres e na presença de um cofator de enzima de maneira tal que a polimerase mova a sequência alvo através do poro com o campo resultante da tensão aplicada. A sequência alvo move na direção 3' para 5'. O método de preferência

compreende adicionalmente (c) adicionar nucleotídeos livres de maneira tal que a polimerase mova a sequência alvo através do poro contra o campo resultante da tensão aplicada (isto é, na direção 5' para 3') e uma proporção dos nucleotídeos na sequência alvo interaja com o poro e (d) medir a corrente que passa através do poro durante cada interação e, por meio disso, realizar a leitura de prova da sequência da sequência alvo obtida na etapa (b), em que as etapas (c) e (d) são também realizadas com uma tensão aplicada através do poro.

[00244] Em terceiro lugar, o método é preferivelmente realizado usando a DNA polimerase Phi29 em modo de abertura. Nesta modalidade, as etapas (a) e (b) são realizadas na ausência de nucleotídeos livres e na ausência de um cofator de enzima de maneira tal que a polimerase controle o movimento da sequência alvo através do poro com o campo resultante da tensão aplicada (a medida que ela está aberto). Nesta modalidade, a polimerase age como um freio impedindo a sequência alvo de mover através do poro muito rapidamente sob a influência da tensão aplicada. O método de preferência compreende adicionalmente (c) reduzir a tensão aplicada através do poro de maneira tal que a sequência alvo move através do poro na direção oposta daquela nas etapas (a) e (b) (isto é, a medida que ela reanela) e uma proporção dos nucleotídeos na sequência alvo interage com o poro e (d) medir a corrente que passa através do poro durante cada interação e, por meio disso, realizar a leitura de prova da sequência da sequência alvo obtida na etapa (b), em que as etapas (c) e (d) são também realizadas com uma tensão aplicada através do poro.

[00245] A invenção também diz respeito a um método para formar um sensor para sequenciar uma sequência de ácido nucleico alvo, compreendendo (a) colocar um poro derivado de Msp em contato com um DNA polimerase Phi29 na presença da sequência de ácido nucleico alvo e (b) aplicar uma tensão através do poro para formar um complexo entre o poro e a polimerase

e, por meio disso, formar um sensor para sequenciar a sequência de ácido nucleico alvo. A invenção adicionalmente diz respeito a um método para aumentar a taxa de atividade de um DNA polimerase Phi29, compreendendo colocar a DNA polimerase Phi29 em contato com um poro derivado de Msp na presença de uma sequência de ácido nucleico e aplicar uma tensão através do poro para formar um complexo entre o poro e a polimerase e, por meio disso, aumentar a taxa de atividade de um DNA polimerase Phi29.

Estojos

[00246] A presente invenção também diz respeito a estojos para caracterizar, tal como sequenciar, uma sequência de ácido nucleico alvo. Um estojo compreende (a) um poro da invenção e (b) uma enzima de manuseio de ácido nucleico. Outro estojo compreende (a) um poro derivado de Msp e (b) um DNA polimerase Phi29. Qualquer uma das modalidades discutidas anteriormente com referência aos métodos da invenção são igualmente aplicáveis aos estojos da invenção.

[00247] Os estojos da invenção podem adicionalmente compreender um ou mais outros reagentes ou instrumentos que permitem que qualquer uma das modalidades supramencionadas sejam realizadas. Tais reagentes ou instrumentos incluem um ou mais do seguinte: tampão(s) adequado(s) (soluções aquosas), meios de obter uma amostra de um sujeito (tal como um vaso ou um instrumento compreendendo uma agulha), meios para amplificar e/ou expressar sequências de polinucleotídeo, uma membrana conforme supradefinido, ou tensão, ou aparelho de *patch clamp*. Reagentes podem estar presentes no estojo em um estado seco de maneira tal que uma amostra fluida ressuspenda os reagentes. O estojo pode também, opcionalmente, compreender instruções para permitir que o estojo seja usado no método da invenção ou detalhes relativos para quais pacientes o método pode ser usado. O estojo pode, opcionalmente, compreender nucleotídeos.

Aparelho

[00248] A invenção também diz respeito a um aparelho para caracterizar, tal como sequenciar, sequências de ácido nucleico alvos em uma amostra. O aparelho pode compreender (a) uma pluralidade de poros da invenção e (b) uma pluralidade de enzimas de manuseio de ácido nucleico. Alternativamente, a invenção pode compreender uma pluralidade de poros derivados de Msp e uma pluralidade de DNA polimerase Phi29s. O aparelho pode ser qualquer aparelho convencional para análise de analito, tal como um arranjo ou um chipe.

[00249] O aparelho preferivelmente compreende:

um dispositivo sensor que é capaz de suportar a pluralidade de poros e ser operável para realizar caracterização ou sequenciamento de ácido nucleico usando os poros e enzimas;

- pelo menos um reservatório para manter material para realizar a caracterização ou sequenciamento;

- um sistema fluídico configurado para suprir controlavelmente o material de pelo menos um reservatório para o dispositivo sensor; e

- uma pluralidade de recipientes para receber amostras respectivas, o sistema fluídico sendo configurado para suprir as amostras seletivamente dos recipientes para o dispositivo sensor. O aparelho pode ser qualquer um daquele descrito no pedido internacional No. PCT/GB 10/000789 (publicado como WO 2010/122293), Pedido internacional No. PCT/GB 10/002206 (ainda não publicado) ou Pedido internacional No. PCT/US99/25679 (publicado como WO 00/28312). Os seguintes exemplos ilustram a invenção:

Exemplo 1

[00250] Homo-oligômeros são poros onde todas as unidades de monômero são idênticas. Como as unidades de monômero se automontarão, esses são os construtos mais simples de produzir. Nossas estratégias para melhorar as propriedades da leitora base podem ser divididos nas categorias:

- Estéricos (*aumentando ou diminuindo o tamanho dos resíduos de aminoácido*)
- Carga (*introduzindo carga +ve para interagir com DNA*)
- Ligação de hidrogênio (*resíduos que podem ligar hidrogênio aos pares bases*)
- Empilhamento Pi (*aminoácidos que interagem através de sistemas pi de elétrons deslocalizados*)

[00251] **Estéricos aumentados / Empilhamento Pi (todo fundo NNN-RRK):**

Estéricos - substituição de resíduos com volume (por exemplo, *Fenilalanina, Triptofano, Tirosina, Histidina*)

Empilhamento Pi - substituição de resíduos aromáticos (por exemplo, *Fenilalanina, Triptofano, Tirosina, Histidina*)

[00252] Em todas as tabelas seguintes (6-11), as mutações feitas para SEQ ID NO: 2 são mostradas. B1 = SEQ ID NO: 2.

Tabela 6

MS-(B1-D91Y) ₈	Tirosina volumosa na constrição
MS-(B1-D90 _g /D91Y) ₈	Tirosina volumosa na constrição, barril reduzido
MS-(B1-D90Y/D91G) ₈	Tirosina volumosa na constrição, barril reduzido
MS-(B1-105Y) ₈	Mutação imediatamente acima para aumentar o tamanho

[00253] **Estéricos Diminuídos** - substituição para resíduos com tamanho menor (por exemplo, *Serina, Treonina, Glicina, Alanina, Valina*)

Tabela 7

MS-(B1-D90 _g /D91Y) ₈	Redução de estéricos no barril
MS-(B1-105A) ₈	Mutação imediatamente acima para reduzir o tamanho
MS-(B1-105G) ₈	Mutação imediatamente acima para reduzir o tamanho

[00254] **Carga** -substituição para resíduos com carga positiva (por exemplo, *Arginina, Lisina, Histidina*)

Tabela 8

MS-(B1-D90R) ₈	Arginina carregada na constrição
MS-(B1-D91R) ₈	Arginina carregada na constrição
MS-(B1-D90R/D91R) ₈	Arginina dupla na constrição
MS-(B1-D90K) ₈	Lisina carregada na constrição
MS-(B1-D91K) ₈	Lisina carregada na constrição
MS-(B1-D90K/D91K) ₈	Lisina dupla na constrição

[00255] **Ligação de hidrogênio** - substituição para resíduos com

capacidade de ligação (por exemplo, *Asparagina*, *Glutamina*, *Tirosina*, *Histidina*)

Tabela 9

MS-(B1-D90Q) ₈	Glutamina na constrição
MS-(B1-D91Q) ₈	Glutamina na constrição
MS-(B1-D90Q/D91G) ₈	Glutamina na constrição, redução de tamanho
MS-(B1-D90 _g /D91Q) ₈	Glutamina na constrição, redução de tamanho
MS-(B1-D90Q/D91Q) ₈	Glutamina dupla na constrição

Tabela 10

MS-(B1-D118N) ₈	Remoção da carga no meio do lúmen
MS-(B1-D118A) ₈	Remoção da carga no meio do lúmen

[00256] Homo-oligômeros podem também ser modificados para conter grupo reativo, que pode em seguida ser modificado quimicamente.

Tabela 11

MS-(B1-D91C) ₈	Adição de cisteína no barril
MS-(B1-D90C)	Adição de cisteína no barril

Exemplo 2

[00257] Diferentes unidades de monômero podem ser combinadas para criar poros oligômero inéditos. Quando o oligômero contém mais que uma diferente subunidade (por exemplo, $MS-(MutA)_6(MutB)_j(MutC)_1$), o poro é um hetero-oligômero. Hetero-oligômeros tipicamente têm apenas uma unidade modificada (por exemplo, $MS-(MutA)_7(MutB)_1$). Outras razões de hetero-oligômeros podem também ser formadas (por exemplo, $MS-(MutA)_6(MutB)_2$). Subunidades podem também incluir SEQ ID NO: 2.

[00258] A vantagem de hetero-oligômeros é que uma única mudança química pode ser feita no poro (sem ser introduzindo uma mudança para cada unidade de monômero). Isto é uma mudança menos drástica para a estrutura do que um homo-oligômero e pode permitir que resíduos sejam introduzidos no poro a uma posição que não funcionou para um homo-oligômero. Um único resíduo interagindo com o DNA pode ser benéfico comparado com múltiplas unidades (por exemplo, um único *Arg* em um hetero-octâmero, comparado com oito *Arg* em um octâmero). Mutantes podem também ser combinados para produzir diferentes efeitos no mesmo resíduo, um exemplo disto seria reduzir o tamanho de sete unidades, aumentando ao mesmo tempo

o tamanho de um (por exemplo, MS-(D90_g)₈(D90Y)₁).

[00259] Regras de projeto de mutante serão similares às apresentadas anteriormente para homo-oligômeros.

[00260] Introdução de um Resíduo Estérico Único

Tabela 12

MS-(B1) ₇ (B1-D90Y) ₁	Adição de tirosina única no barril
MS-(B1) ₇ (B1-D91Y) ₁	Adição de tirosina única no barril
MS-(B1) ₇ (B1-D90W) ₁	Adição de triptofano única no barril
MS-(B1) ₇ (B1-D91W) ₁	Adição de triptofano única no barril

[00261] Introdução de um Resíduo Carregado Único

Tabela 13

MS-(B1) ₇ (B1-D90K) ₁	Adição de lisina única no barril
MS-(B1) ₇ (B1-D91K) ₁	Adição de lisina única no barril
MS-(B1) ₇ (B1-D90R) ₁	Adição de lisina única no barril
MS-(B1) ₇ (B1-D91R) ₁	Adição de lisina única no barril

[00262] Introdução de um Resíduo Reativo Único

Tabela 14

MS-(B1) ₇ (B1-D90C) ₁	Adição de cisteína única no barril
MS-(B1) ₇ (B1-D90C) ₁	Adição de cisteína única no barril

Exemplo 3

[00263] Introdução de um Resíduo Reativo Único para Modificação Química.

Tabela 15

MS-(B1) ₇ (B1-D91C) ₁	Adição de cisteína única no barril
MS-(B1) ₇ (B1-D118C) ₁	Adição de cisteína única no lúmen
MS-(B1) ₇ (B1-G54) ₁	Adição de cisteína única para anexação de proteína de ligação de ácido nucleico

Exemplo 4

[00264] As tabelas seguintes sumarizam os poros mutantes da invenção. A primeira diz respeito a homo-oligômeros e a segunda diz respeito a hetero-oligômeros.

Tabela 16

Projeto_tipo_nome	Mutante_Resumido	Mutante_Total
Mutante (Homo)	MS-(B1)8	MS-(D90N/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90Y)8	MS-(D90Y/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90R)8	MS-(D90R/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90K)8	MS-(D90K/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D118N)8	MS-(D90N/D91N/D93N/D118N/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90Q)8	MS-(D90Q/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D91Q)8	MS-(D90N/D91Q/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D91Y)8	MS-(D90N/D91Y/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90G/D91Y)8	MS-(D90G/D91Y/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D91R)8	MS-(D90N/D91R/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90R/D91R)8	MS-(D90R/D91R/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D91K)8	MS-(D90N/D91K/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90K/D91K)8	MS-(D90K/D91K/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105A)8	MS-(D90N/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105G)8	MS-(D90N/D91N/D93N/I105G/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105Y)8	MS-(D90N/D91N/D93N/I105Y/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105N)8	MS-(D90N/D91N/D93N/I105N/D118R/D134R/E139K)8
Chemical (Homo)	MS-(B1-D91C)8	MS-(D90N/D91C/D93N/D118R/D134R/E139K)8
Chemical (Homo)	MS-(B1-D90C)8	MS-(D90C/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90Y/D91G)8	MS-(D90Y/D91G/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90G/D91G)8	MS-(D90G/D91G/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90G/D93G)8	MS-(D90G/D91N/D93G/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90G)8	MS-(D90G/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D91G)8	MS-(D90N/D91G/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D93G)8	MS-(D90N/D91N/D93G/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90G/D91A)8	MS-(D90G/D91A/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90S)8	MS-(D90S/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D91S)8	MS-(D90N/D91S/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90S/D91S)8	MS-(D90S/D91S/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105K)8	MS-(D90N/D91N/D93N/I105K/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105R)8	MS-(D90N/D91N/D93N/I105R/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105V)8	MS-(D90N/D91N/D93N/I105V/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105L)8	MS-(D90N/D91N/D93N/I105L/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105P)8	MS-(D90N/D91N/D93N/I105P/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105W)8	MS-(D90N/D91N/D93N/I105W/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105S)8	MS-(D90N/D91N/D93N/I105S/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105T)8	MS-(D90N/D91N/D93N/I105T/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-I105Q)8	MS-(D90N/D91N/D93N/I105Q/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88R)8	MS-(L88R/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88A)8	MS-(L88A/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88Y)8	MS-(L88Y/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88G)8	MS-(L88G/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88N)8	MS-(L88N/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88Q)8	MS-(L88Q/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90Y/I105A)8	MS-(D90Y/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90G/I105A)8	MS-(D90G/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90Q/I105A)8	MS-(D90Q/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90R/I105A)8	MS-(D90R/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-D90S/I105A)8	MS-(D90S/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88A/I105A)8	MS-(L88A/D90N/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88S/I105S)8	MS-(L88S/D90N/D91N/D93N/I105S/D118R/D134R/E139K)8
Mutante (Homo)	MS-(B1-L88N/I105N)8	MS-(L88N/D90N/D91N/D93N/I105N/D118R/D134R/E139K)8

Tabela 17

Projeto_Tipo	Mutante_Resumido	Mutante_Total
Mutante (Hetero)	MS-(B1)7(B1-D90K)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90K/D91N/D93N/D118R/D134R/E139K)1
Mutante (Hetero)	MS-(B1)7(B1-D90R)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90R/D91N/D93N/D118R/D134R/E139K)1
Mutante (Hetero)	MS-(B1)7(B1-D90Y)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90Y/D91N/D93N/D118R/D134R/E139K)1

Projeto_Tipo	Mutante_Resumido	Mutante_Total
Mutante (Hetero)	MS-(B1)7(B1-D90Q)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90Q/D91N/D93N/D118R/D134R/E139K)1
Mutante (Hetero)	MS-(B1)7(B1-D91Q)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91Q/D93N/D118R/D134R/E139K)1
Mutante (Hetero)	MS-(B1)7(B1-D91K)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91K/D93N/D118R/D134R/E139K)1
Mutante (Hetero)	MS-(B1)7(B1-D91R)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91R/D93N/D118R/D134R/E139K)1
Mutante (Hetero)	MS-(B1)7(B1-D91Y)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91Y/D93N/D118R/D134R/E139K)1
Mutante (Hetero)	MS-(B1)7(B1-D90W)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90W/D91N/D93N/D118R/D134R/E139K)1
Mutante (Hetero)	MS-(B1)7(B1-D91W)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91W/D93N/D118R/D134R/E139K)1
Química (Hetero)	MS-(B1)7(B1-D90C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90C/D91N/D93N/D118R/D134R/E139K)1
Química (Hetero)	MS-(B1)7(B1-D91C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91C/D93N/D118R/D134R/E139K)1
Química (Hetero)	MS-(B1)7(B1-L88C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91C/D93N/D118R/D134R/E139K)1
Química (Hetero)	MS-(B1)7(B1-S103C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91C/D93N/D118R/D134R/E139K)1
Química (Hetero)	MS-(B1)7(B1-I105C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91C/D93N/D118R/D134R/E139K)1

Exemplo 5 - MspA comparado com HL

[00265] Combinamos DNA polimerase Phi29 (DNAP) como um motor molecular com um nanoporo de MspA do mutante para permitir movimento controlado de uma fita de DNA através do poro. Uma tensão foi aplicada através do poro e uma corrente foi gerada a partir do movimento de íons em uma solução de sal de qualquer lado do nanoporo. A medida que o DNA moveu através do poro, a corrente iônica através do poro mudou com relação ao DNA. Esta informação mostrou ser dependente da sequência.

[00266] Comparamos uma forma mutante de hemolisina com MspA, em particular MS-(B1)g. A faixa de corrente é maior para MspA comparada com hemolisina (HL). Além do mais, a faixa de corrente é também maior para MspA quando uma fita de DNA é enfiada no poro.

[00267] Mostramos que existem inúmeros recursos surpreendentes com MspA que não foram previstos ligando o MspA e o Phi29 DNAP entre si. As diferenças principais são:

1. Movimento da fita mais rápido (Modo de abertura) comparado com HL.

2. Aumento da faixa de corrente quando movimenta uma fita

através do poro.

[00269] 3. Diminuição da variância dos níveis da corrente comparada com mutantes HL.

Movimento da fita mais rápido

[00270] Um molde de DNAss 134mer (SEQ ID NO: 13) foi hibridizado em um 84mer DNAss (SEQ ID NO: 14) para formar um molde de DNAds 84mer com uma saliência 5' de DNAss 50mer. Esta fita moveu através do mutante de MspA MS-(B1)₈ e o mutante de hemolisina usando Phi29 DNAP em Modo de abertura. Duas corridas foram adquiridas; um a KCl 400 mM e a outra a KCl 600 mM, todas à temperatura ambiente com Hepes 10 mM, pH 8,0, EDTA 1 mM, DTT 1mM. O potencial aplicado foi otimizado para cada construto mutante; HL foi corrido a 220 mV e o MspA a 180 mV.

[00271] Níveis da corrente foram extraídos como eventos do DNA no estado ligado a enzima, esses eventos foram indexados e o nível de corrente, duração e variância do evento registrados.

[00272] Para abertura de todas as corridas, a velocidade de abertura não foi consistente através da fita.

[00273] Isto pode mostrado calculando a média da duração do evento, dividida por quartos de índice do evento (Fig. 1). O primeiro quarto forneceu eventos que tiveram uma duração muito maior que os quartos seguintes, isto foi verdadeiro tanto para HL quanto para MspA. Para o primeiro quarto, a média do comprimento do evento foi a menor para MspA a KCl 400 mM e a menor para HL a 600 mM. Entretanto, em Q2, Q3 e Q4, o MspA produziu eventos pequenos para ambas condições de sal. Considerando que o sinal para ruído é suficiente, eventos pequenos são desejáveis uma vez que eles indicam um rápido movimento da fita de DNA através do poro, aumentando assim a produção experimental.

Maior Faixa de Corrente e Menor variância

[00274] Nos experimentos de nanoporo descritos aqui, os níveis da corrente dependem basicamente da concentração de sal, da tensão aplicada, e temperatura. O HL e o MS-(B1)₈ mutantes MspA foram comparados em modo de abertura usando DNA polimerase Phi29 com condições físicas de ajuste de: KCl 600 mM, Hepes 10 mM, EDTA 1 mM, DTT 1 mM, pH 8,0, +220 mV. O DNA usado neste experimento foi um grampo 100mer com uma saliência 5' de fita simples de 34mer (SEQ ID NO: 15). As corridas foram conduzidas à temperatura ambiente.

[00275] Níveis da corrente foram extraídos dos eventos do DNA no estado ligado a enzima, esses eventos foram indexados e o nível de corrente, duração e variância do evento registrados (Figuras 2 e 3).

[00276] Fica claro a partir desses experimentos, que o mutante de MspA dá uma faixa de corrente significativamente maior de aproximadamente 50 pA comparada com o mutante HL onde a faixa é aproximadamente 20 pA (Figuras 2 e 3). Uma grande faixa de corrente é vantajosa já que ela fornecerá um maior sinal para ruído e tornará mais fácil distinguir estados da corrente distintos. Isto é de benefício particular para sequenciar aplicações quando as bases N podem contribuir para o sinal de corrente, levando a possíveis estados da corrente 4^N .

[00277] A variância dos estados é também reduzida para o mutante de MspA comparada com o HL. Isto é mostrado pelo desvio padrão dos eventos nos traços anteriores (Figuras 2 e 3). Para as fitas anteriores, a média do desvio padrão através de todos os eventos para o fita MspA foi 3,6 comparada a 4,5 para HL. Baixa variância de estados é desejável para permitir estimativas exatas do nível de corrente do evento.

Exemplo 6 - Comparação da corrente de poro aberto da linha de base MS-(B1)₈ com os mutantes MS-(B1-I105)₈

[00278] Os níveis da corrente de poros MspA podem ser controlados mutando a posição I105 na proteína. Demonstramos que a corrente de poro

aberto pode ser aumentada acima de 80 % em decorrência de uma única mutação do monômero de MspA. Canais únicos foram inseridos em uma membrana de lipídio nas seguintes condições: KCl 400 mM, Hepes 10 mM, pH 8,0, temperatura ambiente. O nível de corrente de poro aberto foi registrado acima de uma faixa de potenciais aplicados de -200 mV a 200 mV para produzir uma curva IV. O experimento foi repetido para inúmeros poros para avaliar a distribuição da amostra. Um exemplo dos dados de uma corrida da curva IV pode ser visto (Fig. 4).

[00279] Em nossos experimentos, o mutante MS-(B1)₈ da linha de base produz poro que têm uma corrente de poro aberto de aproximadamente 150 pA a +160 mV (A Fig. 5).

[00280] O experimento foi repetido com o mutante MS-(B1-I105Y)₈ que exibiu um grande número de poros com uma maior corrente residual. Para esses canais, a corrente de poro aberto foi aproximadamente 200 pA a +160 mV (Fig. 6).

[00281] O experimento foi repetido com o mutante MS-(B1-I105N)₈ que exibiu duas principais distribuições de níveis da corrente. Dez de dezesseis poros deram uma maior corrente residual em uma distribuição apertada. Para esses canais, a corrente de poro aberto foi aproximadamente 280 pA a +160 mV (Fig. 7).

Exemplo 7 - um poro de MS-(B1-I105A)₈ que muda espontaneamente a condutância

[00282] Observou-se que poros de mutante MspA mudam espontaneamente a condutância durante experimentos de registro elétrico.

[00283] Medições elétricas foram adquiridas da maneira descrita no exemplo 6, usando o poro do mutante MS-(B1-II 05 A)₈.

[00284] Um poro do mutante MspA único é capaz de alternar entre estados de condutâncias espontaneamente alto e baixo (Fig. 8). Isto sugere que as mutações para o MspA permite mudanças conformacionais que são

raramente observadas no poro MS-(B1)₈ de linha de base. É possível que mutações na posição I105 estabilize o estado de alta condutância do poro.

Exemplo 8 - Comparação de correntes de DNA quando se movimentam DNA através dos poros de MS-(B1)₈ de linha de base comparados com os poros de MS-(B1-I1Q5A)₈

[00285] O Poro de MS-(B1)₈ e os poros de MS-(B1-I105N)₈ foram comparados em modo de abertura usando DNA polimerase Phi29 com condições físicas de ajuste de: KCl 400 mM, Hepes 10 mM, EDTA 1 mM, DTT 1 mM, pH 8,0, +180 mV. O DNA usado neste experimento foi um grampo de 100mer com uma saliência 5' de fita simples de 34mer (SEQ ID NO: 15). As corridas foram conduzidas à temperatura ambiente.

[00286] Níveis da corrente foram extraídos como eventos do DNA no estado ligado a enzima, esses eventos foram indexados e o nível de corrente, duração e variância do evento registrados.

[00287] O spread de níveis da corrente da fita de DNA movendo através do mutante MS-(BL)₈ foi ~ 30 pA nessas condições (Fig. 9). O mesmo experimento foi repetido usando o mutante MS-(B1-I105A)₈, os níveis da corrente exibiram uma faixa de ~ 40 pA para a mesma fita de DNA (Fig. 10). A maior faixa de corrente do mutante MS-(I105A)₈ é desejável para discriminar as combinações de nucleotídeos no nanoporo.

Exemplo 9 - Comparação ruído do sinal da linha de base de MS-(B1)₈ com os mutantes MS-(B1-L88N)₈

[00288] Os níveis de ruído de poro de MspA podem ser controlados mutando a posição L88 na sequência de monômero de MspA. Demonstrou-se que o nível de ruído pode ser reduzido em 19 % em decorrência de produzir uma mutação simples no monômero de MspA.

[00289] Este exemplo compara o poro de MS-(B1)₈ e os poros de MS-(B1-L88N)₈ em modo de translocação, usando um helicase para controlar o movimento de fitas de DNA intactas através de um nanoporo.

Materiais

[00290] Iniciadores de nucleotídeos foram projetados para amplificar um fragmento -400 bp de PhiX174. Cada das extremidades 5' desses iniciadores de nucleotídeos incluíram uma região não complementar de 50 nucleotídeos, tanto um trecho homopolimérico quanto seções de unidades de repetição de 10 nucleotídeos homopoliméricos. Esses serviram como os identificadores para translocação controlada da fita através de um nanoporo, bem como determinando a direcionalidade de translocação. Além do mais, a extremidade 5', do iniciador de nucleotídeo direto foi “capeada” para incluir quatro nucleotídeos 2'-O-Metil-Uracil (mU) e a extremidade 5'- do iniciador de nucleotídeo inverso foi quimicamente fosforilado. Essas modificações de iniciador de nucleotídeo permitem então a digestão controlada predominantemente apenas da fita antissentido, usando lambda exonuclease. O capeamento mU protege a fita sentido de digestão de nuclease enquanto o P04 na 5'- da fita antissentido promove-a. Portanto, após incubação com lambda exonuclease apenas a fita sentido do duplex permanece intacta, agora como DNA de fita simples (DNAss). O DNAss gerado foi então PAGE purificado conforme previamente descrito.

[00291] O projeto do substrato de DNA usado neste experimento é mostrado na Fig. 11 (SEQ ID NOs: 19 e 20 (sequências e tags apresentados a seguir)). O substrato de DNA consiste em uma seção 400base de DNAss de PhiX, com uma leitora 50T 5'- para ajudar na captura pelo nanoporo. Anelado a esta fita logo após a leitora 50T ser um iniciador de nucleotídeo contendo um tag colesterol 3' (3' Colesteril-TEG) para enriquecer o DNA na superfície da bicamada, e assim aumentar a eficiência da captura.

SEQ ID NO: 19

```
mUmUmUmUTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TTTTTTTTTTGGTTGTTTCTGTTGGTGCTGATATTGCTTTTGATGCCGA
CCCTAAATTTTTGCCTGTTTGGTTCGCTTTGAGTCTTCTTCGGTTC
```

CGACTACCCTCCCGACTGCCTATGATGTTTATCCTTTGAATGGTCG
 CCATGATGGTGGTTATTATACCGTCAAGGACTGTGTGACTATTGAC
 GTCCTTCCCCGTACGCCGGGCAATAACGTTTATGTTGGTTTCATGG
 TTTGGTCTAACTTTACCGCTACTAAATGCCGCGGATTGGTTTCGCT
 GAATCAGGTTATTAAAGAGATTATTTGTCTCCAGCCACTTAAGTGA
 GGTGATTTATGTTTGGTGCTATTGCTGGCGGTATTGCTTCTGCTCTT
 GCTGGTGGCGCCATGTCTAAATTGTTTGGAGGCGGTC

SEQ ID NO: 20 (mais 3' Colesteril-TEG tag)

GCAATATCAGCACCAACAGAAACAACCTTTTTTTTTTTTTTTTTTTTTT
 TTTTTTTTTT/3CholTEG/

Método Experimental

Solução tamponada: NaCl 400 mM, Hepes 10 mM pH 8,0,
 ATP 1 mM, MgCl₂ 1 mM, DTT 1 mM

Nanoporos: MspA MS(BL)₈; MspA MS(B1-L88N)₈

Enzima: Helicase

[00292] Medições elétricas foram adquiridas de nanoporos de MspA único inseridos em bicamadas de lipídio 1,2-difitanoil-glicero-3-fosfocolina (Avanti Polar Lipids). As bicamadas foram formadas através de aberturas de - 100 µm de diâmetro em filmes PTFE de 20 µm de espessura (em câmaras custom Delrin) por meio da técnica Montal-Mueller, separando duas soluções tamponadas de 1 mL. Todos os experimentos foram realizados em o solução tamponada determinada. Correntes de canal único foram medidas em amplificadores Axopatch 200B (Molecular Devices) equipados com 1440A digitalizadores. Eletrodos Ag/AgCl foram conectados nas soluções tamponadas de maneira que o compartimento cis (no qual tanto nanoporo quanto enzima/DNA são adicionados) seja conectado na terra do pré-amplificador Axopatch, e o compartimento trans seja conectado no eletrodo ativo do pré-amplificador.

[00293] Após obter um único poro tanto de MS(B1)₈ quanto de

MS(B1-L88N)₈ na bicamada, polinucleotídeo do DNA (SEQ ID NOs: 19 e 20) e helicase foram adicionados a 100 µL de tampão e pré-incubados por 5 minutos (DNA = 1,5 nM, Enzima = 1 µM). Esta mistura de pré-incubação foi adicionada a 900 µL de tampão no compartimento cis da câmara de eletrofisiologia para iniciar a captura dos complexos helicase-DNA no nanoporo de MspA (para dar concentrações finais de DNA = 0,15 nM, Enzima = 0,1 µM). A atividade Helicase ATPase foi iniciada conforme exigido pela adição de metal divalente (MgCl₂ 1 mM) e NTP (ATP 1 mM) no compartimento cis. Experimentos foram realizados a um potencial constante de +140 mV. Níveis da corrente foram extraídos como eventos do DNA no estado ligado a enzima, esses eventos foram indexados e o nível de corrente, duração e variância do evento registrados.

[00294] Usando o poro de MspA MS-(B1)₈, 31,08 % dos eventos detectados tiveram um desvio padrão >2,0 a um potencial aplicado de +140 mV (dados adicionais sumarizados na tabela 18). O experimento foi repetido com o mutante MS-(B1-L88N)₈ onde apenas 12,38 % dos eventos detectados exibiram um desvio padrão de >2,0 a um potencial aplicado de +140 mV (dados adicionais sumarizados na tabela 18). Portanto, a mutação pontual em L88 na sequência de monômero de MspA tem reduzido a faixa de ruído observada em 19 %.

Tabela 18

Propriedade	Poro	
	MS-(B1) ₈	MS-(B1-L88N) ₈
Média S.D.	2.30	1.79
Mediano S.D.	1.57	1.48
% de D. P. > 2	31.08	12.38

Exemplo 10 - Comparação de ruído do sinal da linha de base de MS-(B1)₈ como os mutantes MS-(B1-L88N)₈, MS-(B1-L88S)₈ e MS-(B1-L88Q)₈

[00295] Os níveis de ruído de poro de MspA podem ser alterados mutando a posição de L88 na proteína. Demonstrou-se que o nível de ruído pode ser reduzido em decorrência de produzir uma mutação simples no monômero de MspA.

[00296] Este exemplo compara o poro de MS-(B1)₈ com os poros de MS-(B1-L88N)₈, MS-(B1-L88S)₈ e MS-(B1-L88Q)₈ em modo de abertura, usando DNA polimerase Phi29 para controlar o movimento de fitas de DNA intactas através de um nanoporo. O projeto do substrato de DNA usado em todos os experimentos descritos neste exemplo é mostrado na Fig. 12 (SEQ ID NOs: 21, 22 e 23). SEQ ID NO: 23 foi marcado com um IDT Int Spacer 9 (iSp9) e 3' Colesteril-TEG (3CholTEG) conforme mostrado a seguir. As corridas foram conduzidas à temperatura ambiente a um potencial aplicado de +180 mV.

SEQ ID NO: 23:

CAGCGATGGAGATAC/iSp9//3 Cho 1TEG/

Método Experimental

Solução tamponada: 400 mM KCl, Hepes 10 mM pH 8,0, EDTA 1 mM, DTT 1 mM

Nanoporos: MS(B1)₈ MspA;

MspA MS(B1-L88N)₈;

MS(B1-L88S)₈ MspA;

MS(B1-L88Q)₈ MspA;

Enzima: DNA polimerase Phi29 SEQ ID NO: 4

[00297] Medições elétricas foram adquiridas da maneira descrita no exemplo 9. Após obter um único poro tanto de MS(B1)₈, MS(B1-L88N)₈, MS(B1-L88S)₈ quanto de MS(B1-L88Q)₈ na bicamada, polinucleotídeo do DNA (SEQ ID NOs: 21, 22 e 23) e DNA polimerase Phi29 foram adicionados a 100 µL de tampão e pré-incubados por 5 minutos. Esta mistura de pré-incubação foi adicionada a 900 µL de tampão no compartimento cis da câmara de eletrofisiologia para iniciar captura dos complexos polimerase-DNA no nanoporo de MspA (para dar concentrações finais de DNA = 0,5 nM, Enzima = 0,1 µM). Experimentos foram realizados a um potencial constante de +180 mV. Níveis da corrente observados quando o DNA está no

estado ligado a enzima foram indexados e o nível de corrente, sua duração e variância foram registrados.

[00298] Nos experimentos, a linha de base de mutante MS-(B1)₈ exibiu altos níveis de ruído (76,15 % de desvios padrões > 2,0, vide tabela 19) a +180 mV. Os outros três mutantes testados, (MS-(B1-L88N)₈, MS-(B1-L88S)₈ e MS-(B1-L88Q)₈) que tiveram uma única mutação pontual na posição L88, todos observaram níveis de ruído mais baixos (vide tabela 19) do que o poro da linha de base com relação a mesma fita de sequência de DNA. Portanto, foi possível reduzir o ruído do sinal aplicando mutações pontuais na posição L88 em sequência de monômero de MspA.

Tabela 19

Poros	D.P. da média	D.P. da mediana	% de D.P. > 2
MS-(B1) ₈	3.26	2.89	76.15
MS-(B1-L88N) ₈	3.22	2.60	74.18
MS-(B1-L88S) ₈	3.12	2.33	71.71
MS-(B1-L88Q) ₈	3.30	2.46	74.19

Exemplo 11 - Comparação da faixa de sinal geral da linha de base de MS-(B1)₈ com outros mutantes MspA

[00299] A faixa de sinal de poro de MspA pode ser aumentada mutando várias posições na sequência do monômero da proteína MspA.

[00300] Este exemplo compara o poro de MS-(B1)₈ com os seguintes poros - MS-(B1-D90Q)₈, MS-(B1-I105L)₈, MS-(B1-I105Y)₈, MS-(B1-I89Y-D90S)₈, MS-(B1-N86T)₈ e poros MS-(B1-S103G)₈ em modo de abertura, usando um DNA polimerase Phi29 para controlar o movimento de fitas de DNA intactas através de um nanoporo. O projeto do substrato de DNA, usado em todos os experimentos descritos neste exemplo, é mostrado na Fig. 12 (SEQ ID NOs: 21, 22 e 23). SEQ ID NO: 23, marcado com iSp9 e 3CholTEG é mostrado anteriormente. As corridas foram conduzidas à temperatura ambiente a um potencial aplicado de + 180 mV. Níveis da corrente observados quando o DNA está no estado ligado a enzima foram indexados e o nível de corrente, sua duração e variância foram registrados.

Método Experimental

Solução tamponada: 400 mM KCl, Hepes 10 mM pH 8,0, EDTA 1 mM, DTT 1 mM

Nanoporos: MS(B1)₈ MspA;

MS(B1-D90Q)₈ MspA;

MS-(B1-I105L)₈ MspA;

MS-(B1-I105Y)₈ MspA;

MS-(B1-I89Y-D90S)₈ MspA;

MS-(B1-N86T)₈ MspA;

MS-(B1-S103G)₈ MspA;

Enzima: DNA polimerase Phi29 SEQ ID NO: 4

[00301] Medições elétricas foram adquiridas da maneira descrita no exemplo 10. Após obter um único poro tanto de MS(B1)₈, MS(B1-D90Q)₈, MS(B1-I105L)₈, MS(B1-I105Y)₈, MS-(B1-I189Y-D90S)₈, MS-(B1-N86T)₈ quanto de MS-(B1-S103G)₈ na bicamada, polinucleotídeo do DNA (SEQ ID NOs: 21, 22 e 23) e DNA polimerase Phi29 foram adicionados a 100 µL de tampão e pré-incubados por 5 minutos. Esta mistura de pré-incubação foi adicionada a 900 µL de tampão no compartimento cis da câmara de eletrofisiologia para iniciar captura dos complexos polimerase-DNA no nanoporo de MspA (para dar concentrações finais de DNA = 0,5 nM, Enzima = 0,1 µM). Experimentos foram realizados a um potencial constante de +180 mV. Níveis da corrente observados quando o DNA está no estado ligado a enzima foram indexados e o nível de corrente, sua duração e variância foram registrados.

[00302] Nos experimentos, o mutante MS-(B1) da linha de base ₈ exibiu uma faixa máxima de 35 pA a +180 mV (Tabela 20). Os outros 6 mutantes testados (MS-(B1-D90Q)₈, MS-(B1-I105L)₈, MS-(B1-I105Y)₈, MS-(B1-I89Y-D90S)₈, MS-(B1-N86T)₈ e MS-(B1-S103G)₈) todos observaram uma maior faixa máxima do que o poro da linha de base (Vide Tabela 20) com relação a mesma fita de sequência de DNA. Portanto, foi possível

aumentar a faixa de sinal aplicando mutações pontuais em vários locais na sequência de monômero de MspA.

Tabela 20

No. de entrada	Poros Mutante	Faixa (pA)
1	MS(B1)8	34
2	MS-(B1-D90Q)8	70
3	MS-(B1-I105L)8	42
4	MS-(B1-I105Y)8	45
5	MS-(B1-I89Y-D90S)8	67
6	MS-(B1-N86T)8	58
7	MS-(B1-S103G)8	54

Exemplo 12 - Comparação de perfil de sequenciamento geral da linha de base de MS-(B1)₈ com outros mutantes MspA

[00303] O perfil de sequenciamento de poro de MspA pode ser controlado produzindo mutação em uma variedade de posições na sequência do monômero da proteína MspA.

[00304] Este exemplo compara o poro de MS-(B1)₈ com MS-(B1-D90Q-D93S-I105A)₈, MS-(B1-D90Q-Q126R)₈, MS-(B1-L88N-D90Q-D91M)₈, MS-(B1-L88N-D90Q-D91 S)₈ e poros MS-(B1-G75S-G77S-L88N-Q126R)₈ em modo de translocação, usando um helicase para controlar o movimento das fitas de DNA intactas através de um nanoporo.

Método Experimental

Solução tamponada: NaCl 400 mM, Hepes 10 mM pH 8,0, ATP 1 mM, MgCl₂ 1 mM, DTT 1 mM

Nanoporos: MS(B1)₈ MspA;

MS(B1-D90Q-D93S-I105A)₈ MspA;

MS(B1-D90Q-Q126R)₈ MspA;

MS(B1-L88N-D90Q-D91M)₈ MspA;

MS(B1-L88N-D90Q-D91S)₈ MspA;

MS(B1-G75S-G77S-L88N-Q126R)₈ MspA;

Enzima: Helicase

[00305] A configuração experimental foi realizada da maneira descrita no exemplo 9. Após obter um único poro tanto de MS-(B1)₈, MS-(B1-D90Q-D93S-I105A)₈, MS-(B1-D90Q-Q126R), MS-(B1-L88N-D90Q-D91M)₈, MS-

(B1-L88N-D90Q-D91 S)₈ quanto de MS-(B1-G75S-G77S-L88N-Q126R)₈ na bicamada, polinucleotídeo do DNA (SEQ ID NOs: 19 e 20 (sequência e tags mostrados anteriormente)) e helicase foram adicionados a 100 µL de tampão e pré-incubados por 5 minutos (DNA = 1,5 nM, Enzima = 1 µM). Esta mistura de pré-incubação foi adicionada a 900 µL de tampão no compartimento cis da câmara de eletrofisiologia para iniciar captura dos complexos helicase-DNA no nanoporo de MspA (para dar concentrações finais de DNA = 0,15 nM, Enzima = 0,1 µM). Atividade Helicase ATPase foi iniciada conforme exigido pela adição de metal divalente (MgCl₂ 1 mM) e NTP (ATP 1 mM) a o compartimento cis. Experimentos foram realizados a um potencial constante de +140 mV. Níveis da corrente observados quando o DNA estava no estado ligado a enzima foram indexados e o nível de corrente, sua duração e variância foram registrados.

[00306] Nos experimentos, o mutante MS-(B 1)₈ da linha de base produziram o perfil de sequenciamento mostrado na Fig. 13a. O experimento foi repetido com os seguintes mutantes MS-(B1-D90Q-D93S-I105A)₈, MS-(B1-D90Q-Q126R), MS-(B1-L88N-D90Q-D91M)₈, MS-(B1-L88N-D90Q-D91S)₈ e MS-(B1-G75S-G77S-L88N-Q126R)₈, que exibiram uma variedade de diferentes perfis de sequenciamento (vide Fig. 13 b-f). Portanto, produzindo mutações pontuais a uma variedade de posições na sequência de monômero de MspA é possível alterar o perfil de sequenciamento que é detectado.

Exemplo 13 - Análise de uma sequência de fita de RNA usando o poro da linha de base MS-(B1)₈

[00307] Este exemplo descreve como o poro da linha de base de MspA MS-(B1)₈ combinado com a DNA polimerase Phi29 pode ser usado para sequenciar uma fita de RNA.

[00308] Este exemplo usa o poro de MS-(B1)₈ em modo de abertura, usando um DNA polimerase Phi29 para controlar o movimento de fitas de

RNA intactas através de um nanoporo. O projeto substrato híbrido de RNA/DNA usado neste experimento é mostrado na Fig. 14 (SEQ ID NOs: 24 e 25). SEQ ID NOs: 24 e 25 serão apresentados a seguir (RNA em negrito). As corridas foram conduzidas à temperatura ambiente a um potencial aplicado de +180 mV.

SEQ ID NO: 24:

5'OH-CCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCUA**UUCUG**
UUUAUGUUUCUUGUUUGU – 3'OH

SEQ ID NO: 25 (mais colesterol tag):

5'Phos-**UAUUCUGUUUAUGUUUCUUGUUUGUUAGCCCCCUUGA**
UAAGACAAAUACAAAGAACAAA-3'Chol

Materiais

[00309] A fim de sintetizar a fita híbrida de RNA/DNA (120 mer de comprimento), foi necessário ligar a SEQ ID NOs: 24 e 25 entre si. Isto foi obtido usando a fita adaptadora de DNA complementar SEQ ID NO: 26 para ligar as duas fitas na proximidade imediata, onde elas foram subsequentemente ligadas entre si formando SEQ ID NO: 27 híbrido de DNA/RNA 120mer.

SEQ NO: 27 (mais colesterol tag; RNA em negrito):

5'OH-CCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCUA**UUCUG**
UUUAUGUUUCUUGUUUGUUAUUCUGUUUAUGUUUCUUGUUUG
UUAGCCCCCUUGAUAAGACAAAUACAAAGAACAAA-3'Chol

Método Experimental

Solução tamponada: 400 mM KCl, Hepes 10 mM pH 8,0, EDTA 1 mM, DTT 1 mM

Nanoporo: MS(B1)₈ MspA;

Enzima: DNA polimerase Phi29 SEQ ID NO: 4

[00310] Medições elétricas foram adquiridas da maneira descrita no exemplo 10. Após obter um único poro de MS(B1)₈ na bicamada,

polinucleotídeo do DNA (SEQ ID NOs: 24 e 25) e DNA polimerase Phi29 foram adicionados a 100 µL de tampão e pré-incubados por 5 minutos. Esta mistura de pré-incubação foi adicionada a 900 µL de tampão no compartimento cis da câmara de eletrofisiologia para iniciar captura dos complexos polimerase-DNA no nanoporo de MspA (para dar concentrações finais de DNA = 0.2 nM, Enzima = 0,2 µM). Experimentos foram realizados a um potencial constante de +180 mV. Níveis da corrente foram extraídos como eventos do DNA no estado ligado a enzima, esses eventos foram indexados e o nível de corrente, duração e variância do evento registrados.

[00311] Nos experimentos, observou-se que o mutante MS-(B1)₈ da linha de base, combinado com DNA polimerase Phi29 como um motor molecular, detecta níveis distintos da corrente a medida que a fita de RNA foi enfiada através do poro. Esses sinais da corrente foram então usados para determinar a sequência do alvo. Eventos de translocação de RNA típicos, em DNA polimerase Phi29 modo de abertura, são mostrados na Fig. 15.

Exemplo 14 - Dímero e oligomerização de MspA para formar poros

[00312] Este exemplo descreve a preparação e oligomerização do dímero de MspA.

Preparação do Dímero

[00313] Proteína monomérica NNNRRK MspA consiste em aminoácido de 184 resíduos. Um único polipeptídeo foi projetado para produzir uma versão dimérica de proteína MspA-NNNRRK.

[00314] A sequência de DNA que codifica o polipeptídeo MspA-NNNRRK de 184 resíduos foi ligada a uma segunda sequência de DNA que codifica a cadeia de polipeptídeo idêntica por meio de uma sequência ligante de DNA curta. A sequência ligante de DNA codifica SGSGSGDDDDDDDDSGSGSS (SEQ ID NO: 33; mostrada como -(SG)₃-D₈-(SG)₂(SS)-). Um códon iniciador (ATG) foi adicionado imediatamente antes da primeira base e um DNA que codifica dois códons de parada

(TAATAG) foi adicionado após a última base. Portanto, toda a sequência de DNA que codifica MspA-NNNRRK-(SG) -D8-(SG)2(SS)-MspA-NNNRRK é mostrada na SEQ ID NO: 28.

[00315] O DNA foi sintetizado a GenScript USA Inc e clonado em um vetor pT7 com propósitos de expressão.

[00316] A proteína foi gerada por transcrição e tradução *in vitro* (IVTT) acopladas usando um *E. coli* T7-S30 sistema de extrato para DNA circular (Promega).

[00317] A mistura de aminoácido 1 mM completa menos cisteína e a mistura de aminoácido 1 mM completa menos metionina foram misturadas em volumes iguais para obter a solução de aminoácido funcional para gerar altas concentrações das proteínas. A mistura de aminoácido (2,5,0 µL), solução da pré-mistura (10 µL), [35S]L-metionina (0,5 µL) e rifampicina (2 µL, 50 mg/mL) foram misturados com DNA do plasmídeo (4 µL, 400 ng/mL) e extrato T7 S30 (7,5 µL). Síntese foi realizada por 90 minutos a 37°C para gerar 25 µL de proteínas IVTT para monômero e dímero MspA-NNNRRK. Após a reação, amostras foram centrifugadas a 25.000 g por 10 minutos e o sobrenadante foi descartado. O precipitado foi lavado com 100 µL de MBSA (MOPS 10 mM, NaCl 150 mM, pH 7,4 contendo 1 mg/mL BSA) e ressuspenso em 25 µL de amostra tampão Lamellae. As amostras foram submetidas a SDS-PAGE em um gel 10 %. O gel foi seco a 80°C por 45 minutos e expostas a filme de raios X por 2 horas. O gel mostrou 2 bandas distintas, uma correspondente ao dímero MspA e uma ao monômero de MspA.

Oligomerização de Monômero e Dímero

[00318] Expressão do dímero e, separadamente, do monômero foi realizada na presença de vesículas de lipídio sintético para facilitar oligomerização. Uma mistura de lipídio de cinco componentes foi usada (PS: SM: PE: PC: Colesterol em razão 10:10:20:30:30, 25 Mg/mL). 50 µL, de

mistura de lipídio foram centrifugados a 25.000 g por 10 minutos em um tubo Eppendorf de 1,5 mL e o sobrenadante foi descartado. A mistura de aminoácido completa 1 mM menos cisteína e a mistura de aminoácido 1 mM completa menos metionina foram misturadas em volumes iguais para obter a solução de aminoácido funcional exigida para gerar altas concentrações das proteínas. O precipitado da membrana foi ressuspensão com mistura de aminoácido (10,0 µL), solução da pré-mistura (40 µL), [3 S]L-metionina e rifampicina (2 µL, 50 mg/mL). DNA do plasmídeo (16 µL, 400 ng/mL) e extrato de T7 S30 (30,0 µL) foram adicionados para iniciar a síntese. A síntese foi realizada por 90 minutos a 37°C para gerar 100 µL, de proteína IVTT. Amostra da reação de IVTT foi centrifugada (25.000 g, 10 minutos) e o precipitado da membrana resultante foi lavado com MBSA e submetido a eletroforese gel de SDS-poliacrilamida em um gel 7,5 %. O gel foi seco em um papel watman 3M a 50°C por 3 horas e exposto a filme de raios X por 2 horas. O gel mostrou 8 bandas distintas para o dímero de MspA oligomerizado, todos dos quais migraram mais lentamente em SDS PAGE do que o monômero oligomerizado.

Purificação da proteína para experimentos de bicamada

[00319] Três bandas da proteína do dímero oligomerização experimento foram excisadas do gel e purificadas. Usando o autoradiograma como o molde, bandas foram cortadas e reidratadas em tampão (150 a 200 µL de Tris.HCl 25 mM, pH 8,0). O papel foi removido e o pedaço de gel foi triturado usando um pilão. A lama foi filtrada através de uma coluna QIAshredder (Qiagen) por centrifugação a 25.000 x g por 10 minutos. A proteína resultante da terceira banda do nível do monômero foi em seguida usada nos experimentos de eletrofisiologia descritos no exemplo 15.

Exemplo 15 - Comparação do MS-(B1)₈ oligomerizado a partir do monômero com o MS-(B1-BD4 oligomerizado a partir do Dímero

[00320] Este exemplo compara o poro de MS-(B1)₈ oligomerizado a

partir do monômero (SEQ ID NO: 2) com o poro de MS-(B1-B1)4 oligomerizado a partir do dímero (SEQ ID NO: 29) em modo de translocação, usando um helicase para controlar o movimento de fitas de DNA intactas (SEQ ID NOs: 19 e 20 (sequência e tags mostrados anteriormente)) através de um nanoporo.

Método Experimental

Solução tamponada: NaCl 400 mM, Hepes 10 mM pH 8,0, ATP 1 mM, MgCl₂ 1 mM, DTT 1 mM

Nanoporos: MS-(B1)8;
MS-(B1-B1)4

Enzima: Helicase

[00321] Medições elétricas foram adquiridas usando chips de silício de 128 poços (formato 75 µm de diâmetro, profundidade 20 µm e 250 µm de passo) que foram plaqueados com prata (WO 2009/077734). Os chips foram inicialmente lavados com 20 mL de etanol, em seguida 20 mL de dH₂O, em seguida 20 mL de etanol antes de tratamento com plasma CF₄. Os chips usado foram em seguida pré-tratados por revestimento por imersão, seladora a vácuo e armazenados a 4°C. Antes do uso os chips foram aquecidos naturalmente à temperatura ambiente por pelo menos 20 minutos.

[00322] Bicamadas foram formadas passando uma série de cargas de 3,6 mg/mL lipídio de 1,2-difitanoil-glicero-3-fosfocolina (DPhPC, Avanti Polar Lipids, AL, USA) dissolvidos em KCl 1 M, Tris 10 mM, pH 7,5, a 0,45 µL/β através do chipe. Inicialmente, uma carga de lipídio (250 µL) escoou através do chipe, seguido por um 100 µL de carga de ar. Duas cargas adicionais de 155 µL e 150 µL de solução de lipídio, cada qual separado por uma carga de 100 µL de ar foi em seguida passado pelo chipe. Após formação da bicamada a câmara foi fluxada com 3 mL de tampão a um vazão de 3 µL/s. Registro elétrico da formação de bicamada foi realizado a 10 kHz com uma capacitância de integração de 1,0 pF.

[00323] Uma solução do nanoporo biológico foi preparada usando tanto o poro de MS-(B1)₈ oligomerizado a partir do monômero ou o poro de MS-(B1-B1)₄ oligomerizado a partir do dímero em Tris 10 mM, EDTA 1 mM, pH 8,0. Um potencial de manutenção de +180 mV foi aplicado e a solução escoou pelo chipe e os poros entraram naturalmente nas bicamadas. A taxa de amostragem e a capacitância de integração foram em seguida mantidas a 10 kHz e 1,0 pF respectivamente e o potencial aplicado reduzido a zero.

[00324] Um programa controle, que aplicou um potencial de manutenção de + 180 mV, foi corrido. Polinucleotídeo do DNA (SEQ ID NOs: 19 e 20) e helicase foram pré-incubados por 5 minutos. Esta mistura de pré-incubação (que incluiu MgCl₂ e ATP) foi em seguida escoada pelo chipe para iniciar a captura dos complexos helicase-DNA no nanoporo de MspA (para dar concentrações finais de DNA = 1,5 nM, Enzima = 10 nM). Experimentos foram realizados a um potencial constante de +180 mV. Níveis da corrente foram extraídos como eventos do DNA no estado ligado a enzima. Esses eventos foram indexados e o nível de corrente, duração e variância do evento registrado.

[00325] Nos experimentos, o poro do mutante MS-(B1-B1)₄ da linha de base formado a partir da oligomerização do dímero inserido nas bicamadas de lipídio tão efetivamente quanto o poro de MS(B1)₈ formado a partir da oligomerização do monômero (vide Fig. 16 mostrando inserção de poro para o MS(B1)₈ e MS-(B1-B1)₄). Quando os poros oligomerizados do monômero e dímero foram combinados com uma helicase como um motor molecular, foi possível detectar níveis distintos da corrente a medida que a fita de DNA foi enfiada através do poro. Eventos de translocação típicos, em modo de translocação helicase, são mostrados na Fig. 17 para o poro de MS-(B1)₈ formados a partir da oligomerização do monômero e a Fig. 18 para o poro de MS-(B1-B1)₄ formado a partir da oligomerização do dímero. Portanto,

observou-se que o poro do mutante MS-(B1-B1)₄ oligomerizado a partir da unidade do dímero era um poro tão bom quanto o poro do mutante MS-(B1)₈ oligomerizado a partir da unidade do monômero.

Exemplo 16 - Uso do poro do MspA do mutante MS-(B1-L88N)₈ para distinguir 5-metilcitosina de citosina

[00326] Este exemplo descreve como o poro do MspA do mutante MS-(B1-L88N)₈ pode ser usado para distinguir citosina de sua base epigeneticamente modificada 5-metilcitosina. Os projetos do substrato de DNA usados neste experimento são mostrados na Fig. 19 e têm as seguintes sequências:

TTTTTTTTT/idSp/TTTTTTTTTmCTTTTTTTTCTTTTTTTT
mCGTTTTTTTTTCGTTTTTTTTTGATCTCCATCGCTGCCCCCTTTTTCC
CCTTTTTT (que é SEQ ID NO: 30 com 9 T nucleotídeos em um IDT Int d
Spacer (idSp) na extremidade 5'). mC representa 5-metilcitosina
GGCAGCGATGGAGATACTTGAGGCGAGCGGTCAA (SEQ ID NO: 31)
e 5CholTEG/TTGACCGCTCGCCTC (SEQ ID NO: 32 com um 5' Colesteril-
TEG tag).

Materiais

[00327] A fim de formar o construto da fita de DNA mostrado na Fig. 19 foi necessário hibridizar SEQ ID NO: 30, 31 e 32 juntos. Isto foi realizado pré-incubando todas três fitas ao mesmo tempo.

Método Experimental

Solução tamponada: KCl 1 M, Hepes 10 mM pH 8,0, ATP 1 mM, MgCl₂ 1 mM, DTT 1 mM

Nanoporos: MspA MS(B1-L88N)₈

Enzima: Helicase

[00328] A configuração experimental foi realizada da maneira descrita no exemplo 9. Após obter um único poro de MS-(B1-L88N)₈, na bicamada, polinucleotídeo do DNA (SEQ ID NOs: 30, 31 e 32) e helicase foram

adicionados a 50 μ L de tampão e pré-incubados por 5 minutos (DNA = 5 nM, Enzima = 100 nM). Esta mistura de pré-incubação foi adicionada a 950 μ L de tampão no compartimento cis da câmara de eletrofisiologia para iniciar a captura dos complexos helicase-DNA no nanoporo de MspA (para dar concentrações finais de DNA = 5 nM, Enzima = 100 nM). A atividade Helicase ATPase foi iniciada conforme exigido pela adição de metal divalente (MgCl₂ 1 mM) e NTP (ATP 1 mM) a o compartimento cis. Experimentos foram realizados a um potencial constante de +120 mV. Níveis da corrente foram extraídos como eventos do DNA no estado ligado a enzima. Esses eventos foram indexados e o nível de corrente, duração e variância do evento registrados.

[00329] Nos experimentos observou-se que citosina e 5-metilcitosina produziram diferentes níveis da corrente quando translocadas através do poro de MS-(B1-L88N)₈ sob o controle de um helicase (vide Fig. 20). Portanto, usando esta forma mutada de MspA foi possível distinguir citosina de sua base epigeneticamente modificada de 5-metilcitosina.

REIVINDICAÇÕES

1. Monômero de Msp mutante compreendendo uma variante da sequência mostrada na SEQ ID NO: 2, caracterizado pelo fato de que a variante compreende a seguinte substituição: L88N.

2. Monômero de Msp mutante de acordo com a reivindicação 1, caracterizado pelo fato de que a variante ainda compreende pelo menos uma das seguintes substituições: G75S, G77S e Q126R.

3. Monômero Msp mutante de acordo com a reivindicação 2, caracterizado pelo fato de que a variante compreende as seguintes mutações: G75S, G77S e Q126R.

4. Monômero de Msp mutante de acordo com qualquer uma das reivindicações 1 a 3, caracterizado pelo fato de que a variante ainda compreende pelo menos uma das seguintes mutações:

- (a) serina (S), glutamina (Q) ou tirosina (Y) na posição 90;
- (b) leucina (L) ou serina (S) na posição 105;
- (c) arginina (R) na posição 59;
- (d) leucina (L) na posição 78;
- (e) asparagina (N) na posição 81;
- (f) asparagina (N) na posição 83;
- (g) serina (S) ou treonina (T) na posição 86;
- (h) fenilalanina (F), valina (V) ou leucina (L) na posição 87;
- (i) fenilalanina (F), valina (V) ou leucina (L) na posição 89;
- (j) leucina (L), fenilalanina (F), triptofano (W), histidina (H), treonina (T), glicina (G), alanina (A), valina (V), arginina (R), lisina (K), asparagina (N) ou cisteína (C) na posição 90;
- (k) serina (S), glutamina (Q), leucina (L), metionina (M), isoleucina (I), alanina (A), valina (V), glicina (G), fenilalanina (F), triptofano (W), tirosina (Y), histidina (H), treonina (T), arginina (R), lisina (K), asparagina (N) ou cisteína (C) na posição 91;

- (l) alanina (A) ou serina (S) na posição 92;
- (m) serina (S), alanina (a), treonina (T), glicina (G) na posição 93;
- (n) leucina (L) na posição 94;
- (o) valina (V) na posição 95;
- (p) arginina (R), ácido aspártico (D), valina (V), asparagina (N), serina (S) ou treonina (T) na posição 96;
- (q) serina (S) na posição 97;
- (r) serina (S) na posição 98;
- (s) serina (S) na posição 99;
- (t) serina (S) na posição 100;
- (u) fenilalanina (F) na posição 101;
- (v) lisina (K), serina (S) ou treonina (T) na posição 102;
- (w) alanina (A), glutamina (Q), asparagina (N), glicina (G) ou treonina (T) na posição 103;
- (x) isoleucina na posição 104;
- (y) tirosina (Y), alanina (A), glutamina (Q), asparagina (N), treonina (T), fenilalanina (F), triptofano (W), histidina (H), glicina (G), valina (V), arginina (R), lisina (K), prolina (P), ou cisteína (C) na posição 105;
- (z) fenilalanina (F), isoleucina (I), valina (V) ou serina (S) na posição 106;
- (aa) prolina (P) ou serina (S) na posição 108;
- (bb) asparagina (N) na posição 118;
- (cc) serina (S) ou cisteína (C) na posição 103; e
- (dd) cisteína em uma ou mais das posições 10 a 15, 51 a 60, 136 a 139 e 168 a 172,

em que o monômero de Msp mutante retém a capacidade de formar um poro.

5. Monômero de Msp mutante, de acordo com qualquer uma

das reivindicações 1 a 4, caracterizado pelo fato de que a variante compreende um ou mais das substituições seguintes:

- (a) serina (S) na posição 75, serina (S) na posição 77, asparagina (N) na posição 88, glutamina (Q) na posição 90 e arginina (R) na posição 126;
- (b) um ou mais de (i) glutamina (Q) na posição 90 e (ii) alanina (A) na posição 105;
- (c) um ou mais de (i) serina (S) na posição 90 e (ii) serina (S) na posição 92;
- (d) um ou mais de (i) glutamina (Q) na posição 87 e (ii) serina (S) na posição 90;
- (e) um ou mais de (i) tirosina (Y) na posição 89 e (ii) serina (S) na posição 90;
- (f) um ou mais de (i) serina (S) na posição 90 e (ii) alanina (A) na posição 92;
- (g) um ou mais de (i) serina (S) na posição 90 e (ii) asparagina (N) na posição 94;
- (h) um ou mais de (i) serina (S) na posição 90 e (ii) isoleucina (I) na posição 104;
- (i) um ou mais de (i) glutamina (Q) na posição 90, (ii) serina (S) na posição 93 e (iii) alanina (A) na posição 105;
- (j) um ou mais de (i) fenilalanina (F), triptofano (W), tirosina (Y) ou histidina (H) na posição 90, (ii) fenilalanina (F), triptofano (W), tirosina (Y) ou histidina (H) na posição 91 e (iii) fenilalanina (F), triptofano (W), tirosina (Y) ou histidina (H) na posição 105;
- (k) um ou mais de (i) serina (S), treonina (T), glicina (G), alanina (A) ou valina (V) na posição 90, (ii) serina (S), treonina (T), glicina (G), alanina (A) ou valina (V) na posição 91 e (iii) serina (S), treonina (T), glicina (G), alanina (A) ou valina (V) na

posição 105;

- (l) serina (S), arginina (R), lisina (K) ou histidina (H) na posição 90 e/ou serina (S), arginina (R), lisina (K) ou histidina (H) na posição 91;
- (m) serina (S), treonina (T), asparagina (N), glutamina (Q), tirosina (Y) ou histidina (H) na posição 90 e/ou serina (S), treonina (T), asparagina (N), glutamina (Q), tirosina (Y) ou histidina (H) na posição 91; e
- (n) cisteína em uma ou mais das posições 90, 91 e 103,

e em que o monômero de Msp mutante retém a capacidade de formar um poro.

6. Monômero de Msp mutante, de acordo com qualquer uma das reivindicações 1 a 5, caracterizado pelo fato de que a variante compreende pelo menos uma da(s) seguinte(s) substituição(s):

- (i) N90S;
- (ii) N90Q;
- (iii) N90Y;
- (iv) I105L;
- (v) I105S;
- (vi) G75S, G77S, L88N, N90Q e Q126R;
- (vii) E59R;
- (viii) I78L;
- (ix) S81N;
- (x) T83N;
- (xi) N86S;
- (xii) N86T;
- (xiii) I87F;
- (xiv) I87V;
- (xv) I87L;

- (xvi) I89F;
- (xvii) I89V;
- (xviii) I89L;
- (xix) N90L;
- (xx) N91S;
- (xxi) N91Q;
- (xxii) N91L;
- (xxiii) N91M;
- (xxiv) N91I;
- (xxv) N91A;
- (xxvi) N91V;
- (xxvii) N91G;
- (xxviii) G92A;
- (xxix) G92S;
- (xxx) N93S;
- (xxxi) N93A;
- (xxxii) N93T;
- (xxxiii) I94L;
- (xxxiv) T95V;
- (xxxv) A96R;
- (xxxvi) A96D;
- (xxxvii) A96V;
- (xxxviii) A96N;
- (xxxix) A96S;
- (xl) A96T;
- (xli) P97S;
- (xlii) P98S;
- (xliii) F99S;
- (xliv) G100S;

- (xlv) L101F;
- (xlvi) N102K;
- (xlvii) N102S;
- (xlviii) N102T;
- (xlix) S103A;
- (l) S103Q;
- (li) S103N;
- (lii) S103G;
- (liii) S103T;
- (liv) V104I;
- (lv) I105Y;
- (lvi) I105L;
- (lvii) I105A;
- (lviii) I105Q;
- (lix) I105N;
- (lx) I105S;
- (lxi) I105T;
- (lxii) T106F;
- (lxiii) T106I;
- (lxiv) T106V;
- (lxv) T106S;
- (lxvi) N108P;
- (lxvii) N108S;
- (lxviii) N90Q e I105A;
- (lxix) N90S e G92S;
- (lxx) I87Q e N90S;
- (lxxi) I89Y e N90S;
- (lxxii) N90S e G92A;
- (lxxiii) N90S e I94N;

- (lxxiv) N90S e V104I;
- (lxxv) N90Q, D93S e I105A;
- (lxxvi) N91Y;
- (lxxvii) N90Y e N91G;
- (lxxviii) N90G e N91Y;
- (lxxix) N90G e N91G;
- (lxxx) I105G;
- (lxxxii) N90R;
- (lxxxiii) N91R;
- (lxxxiv) N90R e N91R;
- (lxxxv) N90K;
- (lxxxvi) N91K;
- (lxxxvii) N90K e N91K;
- (lxxxviii) N90Q e N91G;
- (lxxxix) N90G e N91Q;
- (lxxxix) N90Q e N91Q;
- (xc) R118N;
- (xci) N91C;
- (xcii) N90C;
- (xciii) N90W;
- (xciv) N91W;
- (xcv) N90K;
- (xcvi) N91K;
- (xcvii) N90R;
- (xcviii) N91R;
- (xcix) N90S e N91S;
- (c) N90Y e I105A;
- (ci) N90G e I105A;
- (cii) N90Q e I105A;

- (ciii) N90S e I105A;
- (civ) N90G e N93G;
- (cv) N90G;
- (cvi) N93G;
- (cvii) N90G e N91A;
- (cviii) I105K;
- (cix) I105R;
- (cx) I105V;
- (cxi) I105P;
- (cxii) I105W;
- (cxiii) N90R e I105A;
- (cxiv) N90S e I105A;
- (cxv) S103C; e
- (cxvi) I105C

e em que o monômero de Msp mutante retém a capacidade de formar um poro.

7. Monômero de Msp mutante, de acordo com qualquer uma das reivindicações 1 a 6, caracterizado pelo fato de que: (i) o mutante é modificado quimicamente por anexação de uma molécula a uma ou mais cisteínas, anexação de uma molécula a uma ou mais lisinas, anexação de uma molécula a um ou mais aminoácidos não naturais, modificação enzimática de um epítipo ou modificação de um terminal; (ii) o mutante é modificado quimicamente por anexação de uma molécula a uma ou mais cisteínas e a uma ou mais cisteínas foram introduzidas no mutante por substituição; (iii) o mutante é modificado quimicamente por anexação de uma molécula a uma ou mais cisteínas, anexação de uma molécula a uma ou mais lisinas ou anexação de uma molécula a um ou mais aminoácidos não naturais e a molécula é (a) um adaptador molecular que facilita a interação entre um poro compreendendo o monômero e um nucleotídeo alvo ou sequência de ácido

nucleico alvo ou (b) uma proteína de ligação de ácido nucleico; (iv) o mutante é modificado quimicamente por anexação de uma molécula a uma ou mais cisteínas, anexação de uma molécula a uma ou mais lisinas ou anexação de uma molécula a um ou mais aminoácidos não naturais e a anexação é por meio de um ligante ou (v) o mutante é quimicamente modificado por anexação de uma molécula a uma ou mais cisteínas, anexação de uma molécula a uma ou mais lisinas ou anexação de uma molécula a um ou mais aminoácidos não naturais e a molécula é anexada a uma ou mais das posições 90, 91 e 103 da SEQ ID NO: 2.

8. Construto, caracterizado pelo fato de que compreende dois ou mais monômeros anexados covalentemente derivados de Msp, em que pelo menos um dos monômeros é um monômero mutante como definido em qualquer uma das reivindicações 1 a 7.

9. Construto, de acordo com a reivindicação 8, caracterizado pelo fato de que (a) os dois ou mais monômeros são os mesmos ou diferentes; (b) pelo menos um monômero compreende a sequência mostrada na SEQ ID NO: 2; (c) o construto compreende dois monômeros e pelo menos um dos monômeros é um monômero mutante conforme definido em qualquer uma das reivindicações 1 a 7; (d) os monômeros são geneticamente fundidos; ou (e) os monômeros são anexados por meio de um ligante.

10. Polinucleotídeo, caracterizado pelo fato de que codifica um mutante, como definido na reivindicação 1 ou um construto como definido na reivindicação 8, em que o polinucleotídeo é selecionado a partir do grupo que consiste na SEQ ID NO: 1, em que o códon para leucina na posição 88 da proteína de SEQ ID NO: 2 é substituído por um códon para asparagina, e suas sequências degeneradas que codificam para a referida proteína.

11. Poro homo-oligomérico, caracterizado pelo fato de que é derivado de Msp compreendendo monômeros mutantes idênticos como definido em qualquer uma das reivindicações 1 a 7.

12. Poro homo-oligomérico, de acordo com a reivindicação 11, caracterizado pelo fato de que o poro compreende oito monômeros mutantes idênticos como definidos em qualquer uma das reivindicações 1 a 7.

13. Poro hetero-oligomérico derivado de Msp, compreendendo pelo menos um monômero mutante, como definido em qualquer uma das reivindicações 1 a 7, caracterizado pelo fato de que pelo menos um dos oito monômeros difere dos outros.

14. Poro hetero-oligomérico, de acordo com a reivindicação 13, caracterizado pelo fato de que (i) o poro compreende oito monômeros mutantes como definido em qualquer uma das reivindicações 1 a 7 e pelo menos um deles difere dos outros; (ii) o poro compreende pelo menos um monômero compreendendo a sequência mostrada na SEQ ID NO: 2; (iii) o poro compreende (a) um monômero mutante e (b) sete monômeros idênticos, em que o monômero mutante em (a) é diferente dos monômeros idênticos em (b); ou (iv) o poro compreende (a) sete monômeros compreendendo a sequência mostrada na SEQ ID NO: 2 e um monômero mutante como definido em qualquer uma das reivindicações 1 a 7, ainda compreendendo a substituição N90R, N90K, N90Y, N90Q, N90W ou N90C; (b) sete monômeros compreendendo a sequência mostrada na SEQ ID NO: 2 e um monômero mutante como definido em qualquer uma das reivindicações 1 a 7 compreendendo a substituição N91R, N91K, N91Y, N91Q, N91W ou N91C; ou (c) sete monômeros compreendendo a sequência mostrada na SEQ ID NO: 2 e um monômero mutante como definido em qualquer uma das reivindicações 1 a 7 compreendendo a substituição L88C, S103C ou I105C.

15. Poro, caracterizado pelo fato de que compreende pelo menos um construto como definido nas reivindicações 8 ou 9.

16. Poro, de acordo com a reivindicação 15, caracterizado pelo fato de que compreende

- quatro construtos, cada qual compreendendo dois

monômeros e pelo menos um dos monômeros é um mutante como definido em qualquer uma das reivindicações 1 a 7; ou

- um construto compreendendo um monômero mutante conforme definido em qualquer uma das reivindicações 1 a 7 e seis monômeros cada qual compreendendo (i) a sequência mostrada na SEQ ID NO: 2 ou (ii) uma variante da SEQ ID NO: 2 conforme definido em qualquer uma das reivindicações 1 a 7.

17. Método para caracterizar uma sequência de ácido nucleico alvo, caracterizado pelo fato de que compreende:

(a) colocar a sequência alvo em contato com um poro, como definido em qualquer uma das reivindicações 11 a 16 e uma proteína de ligação de ácido nucleico de maneira que a proteína controle o movimento da sequência alvo através do poro e uma proporção dos nucleotídeos na sequência alvo interaja com o poro; e

(b) medir a corrente que passa através do poro durante cada interação e, por meio disso, caracterizar a sequência alvo.

18. Método, de acordo com a reivindicação 17, caracterizado pelo fato de que a caracterização a sequência de ácido nucleico alvo compreende estimar a sequência ou sequenciar a sequência de ácido nucleico alvo.

19. Kit para caracterizar uma sequência de ácido nucleico alvo, caracterizado pelo fato de que compreende (a) um poro como definido em qualquer uma das reivindicações 11 a 16 e (b) uma enzima de manuseio de ácido nucleico.

20. Aparelho para caracterizar sequências de ácido nucleico alvos em uma amostra, caracterizado pelo fato de que o aparelho compreende uma câmara compreendendo uma solução aquosa e uma barreira que separa a câmara em duas seções, em que a barreira tem uma abertura na qual uma membrana é formada e em que um poro como definido em qualquer uma das

reivindicações 11 a 16 é inserido dentro da membrana.

Fig. 1

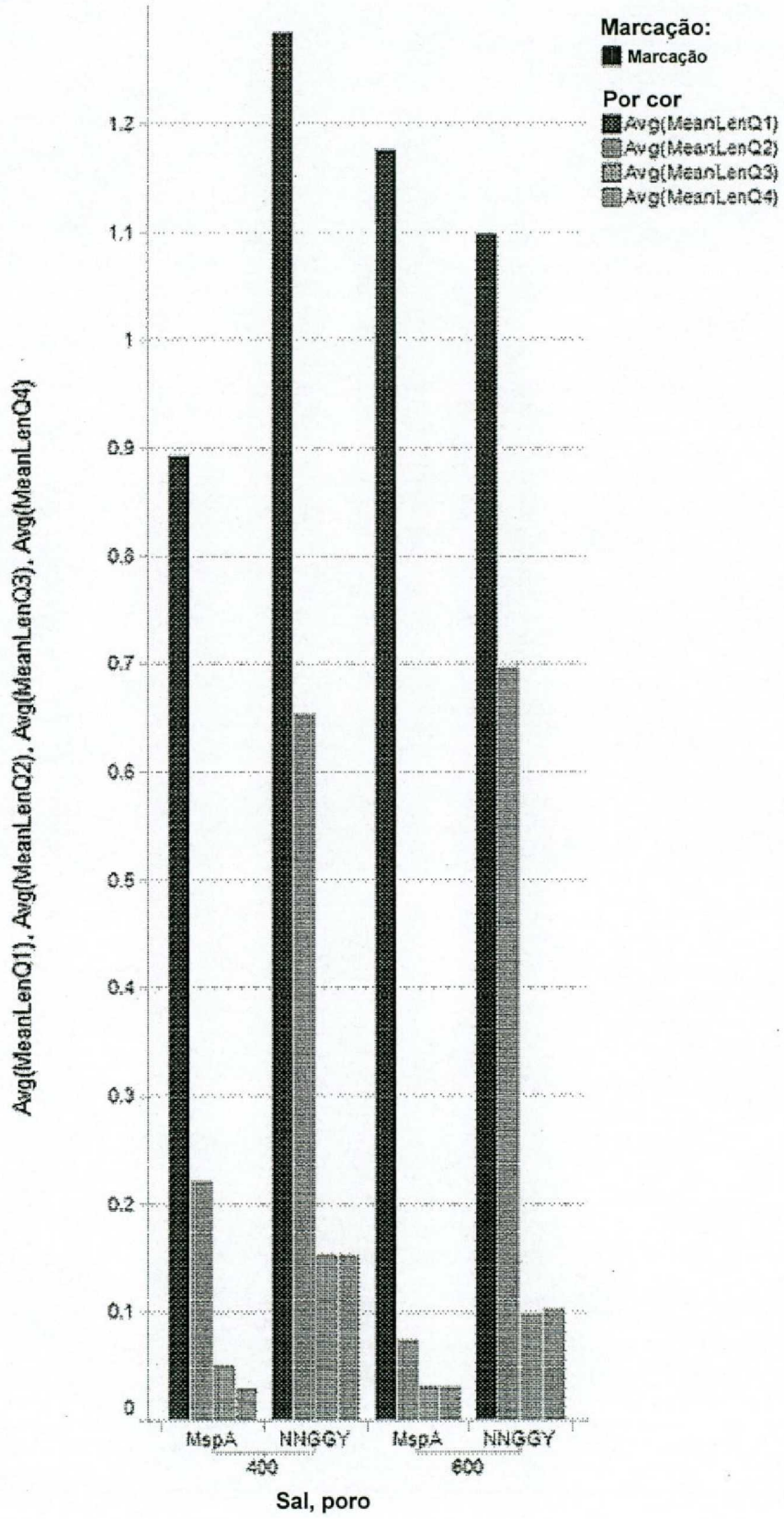


Fig. 2

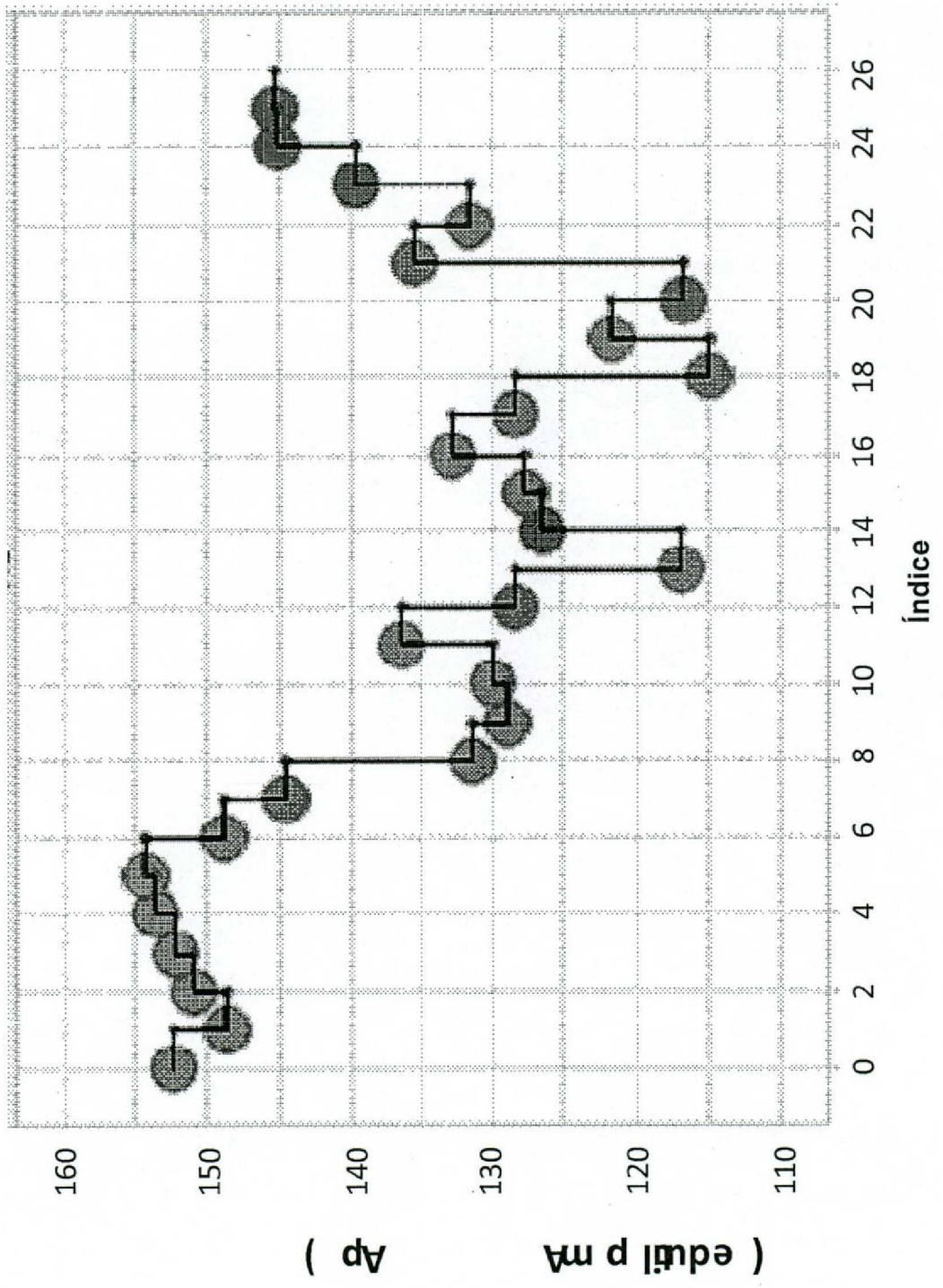


Fig. 3

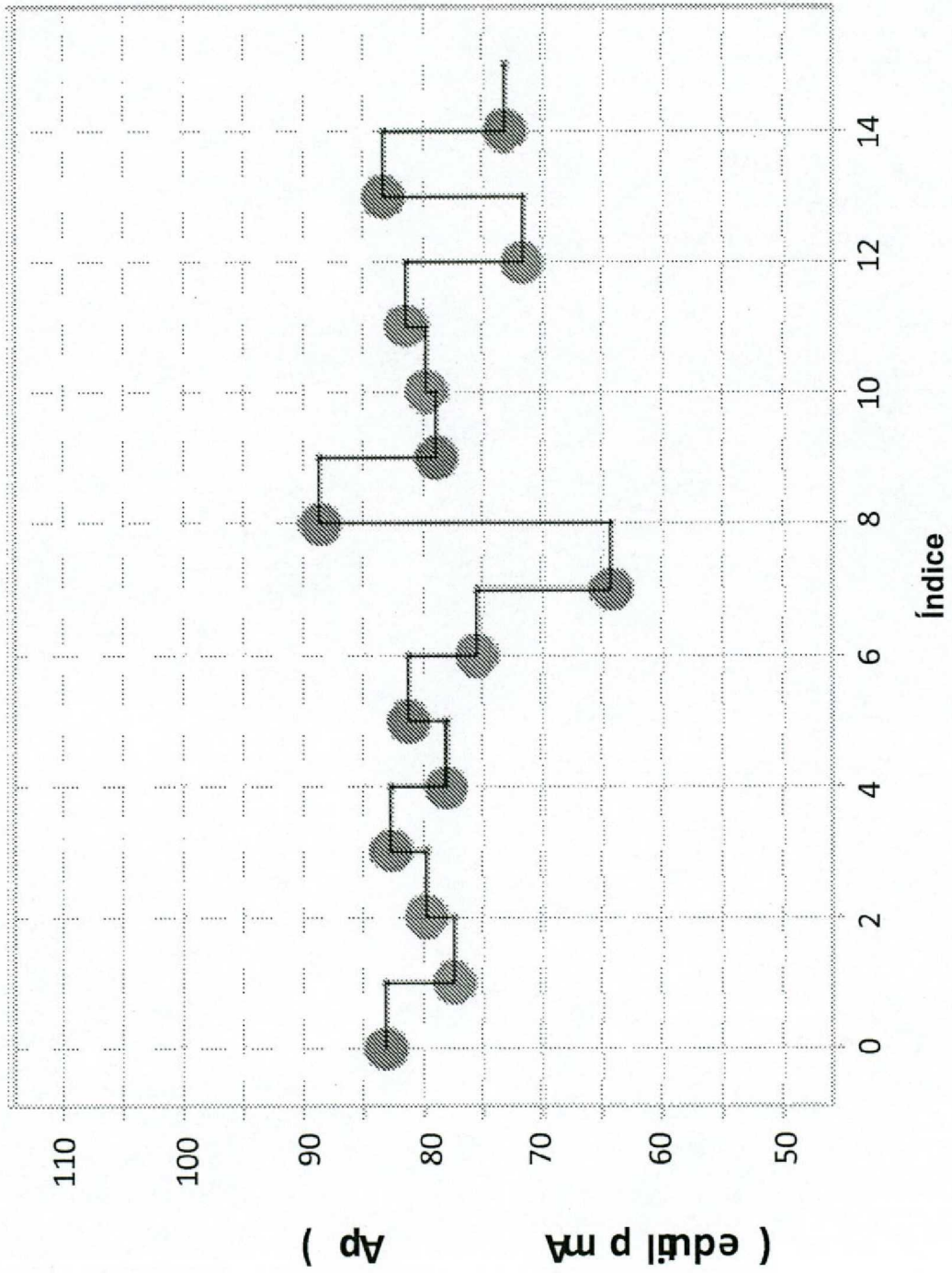


Fig. 4

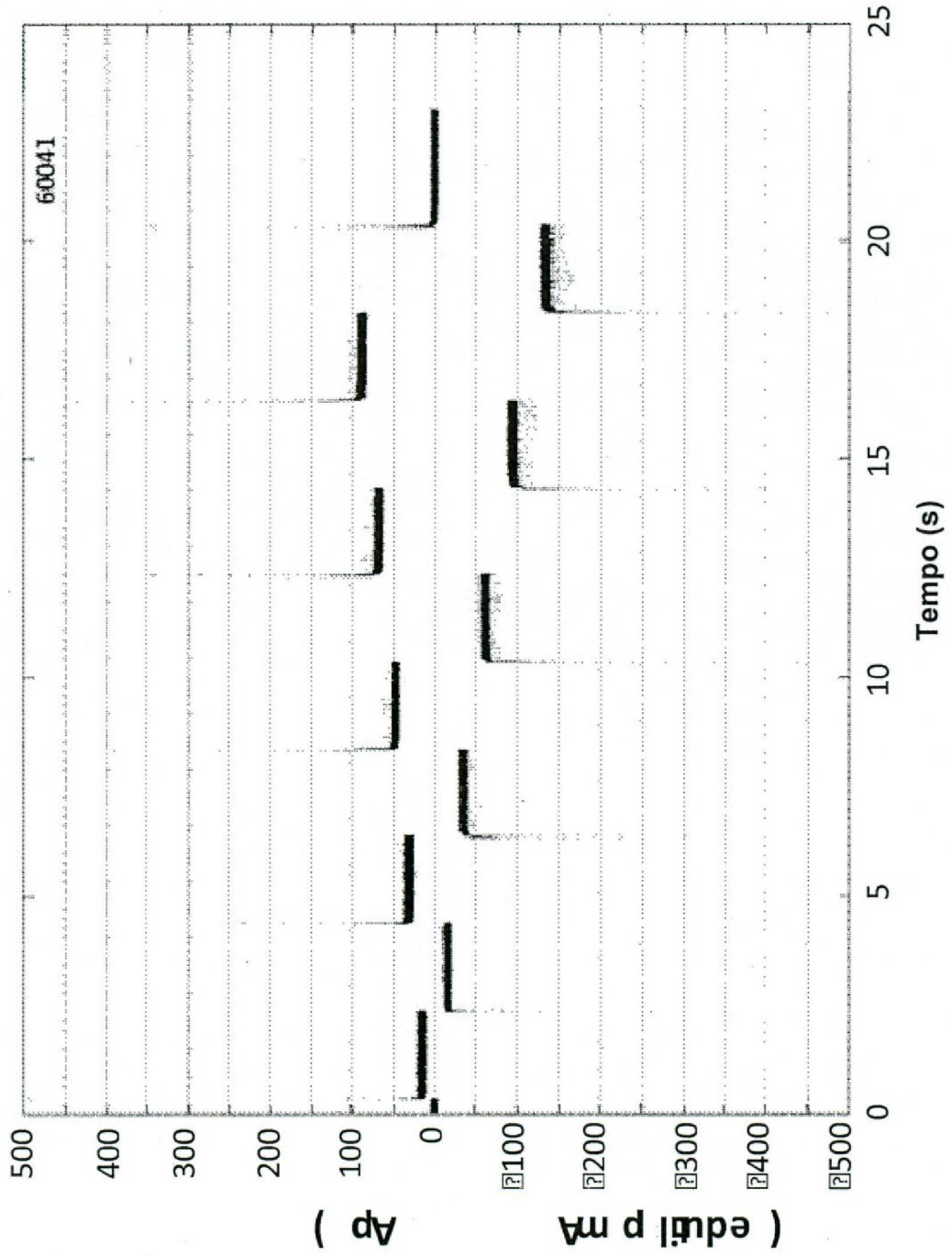


Fig. 5

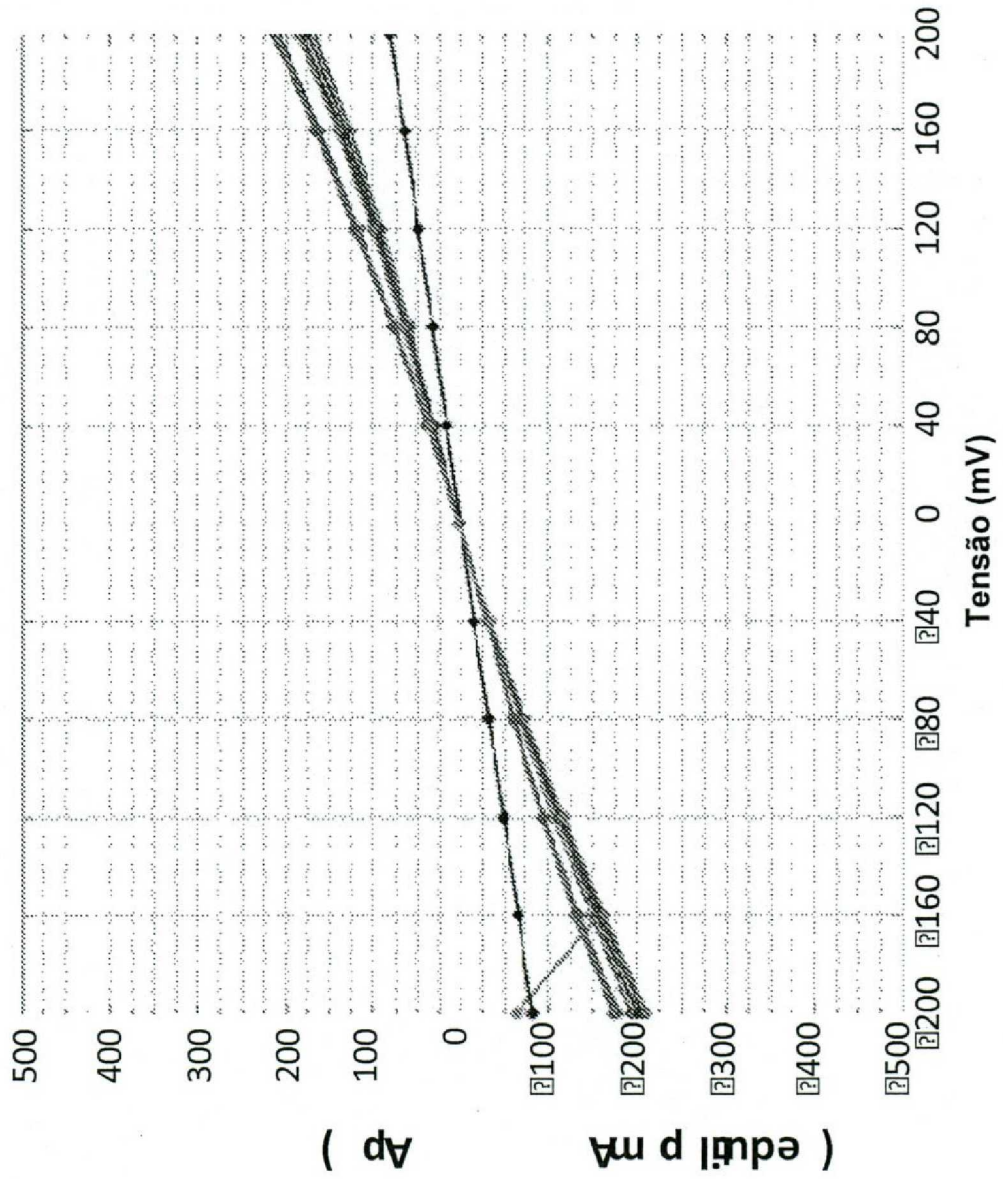


Fig. 6

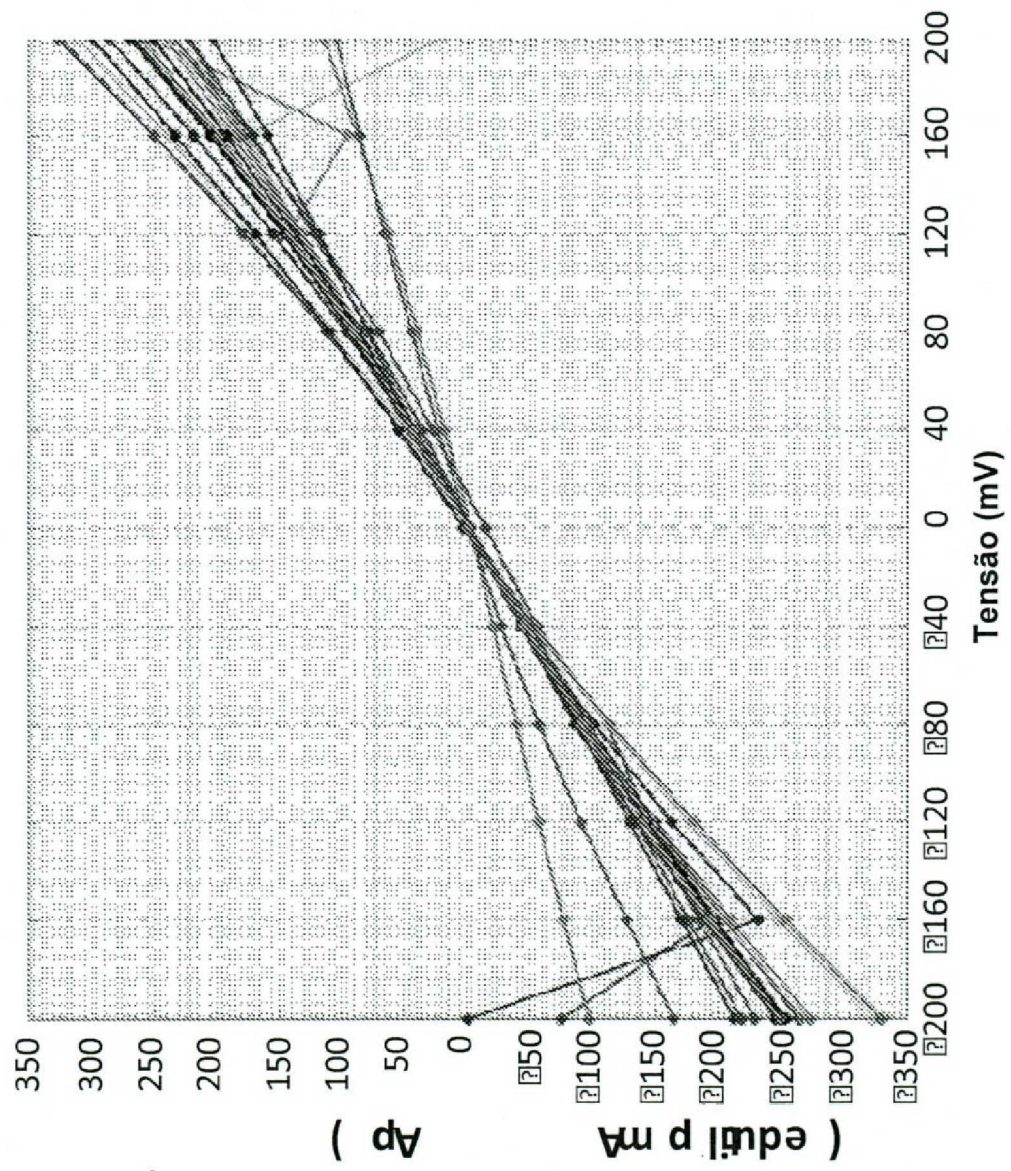


Fig. 7

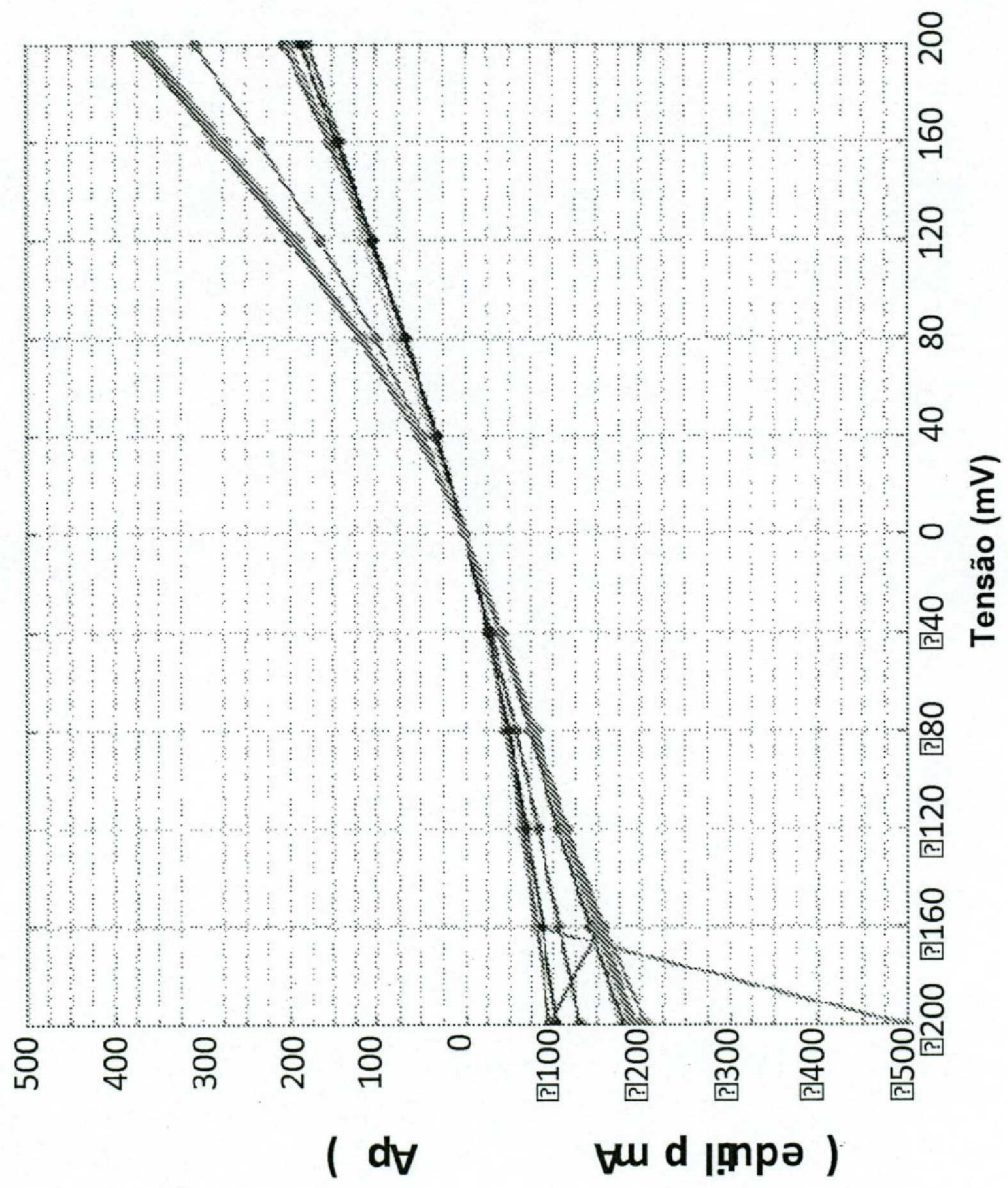


Fig. 8

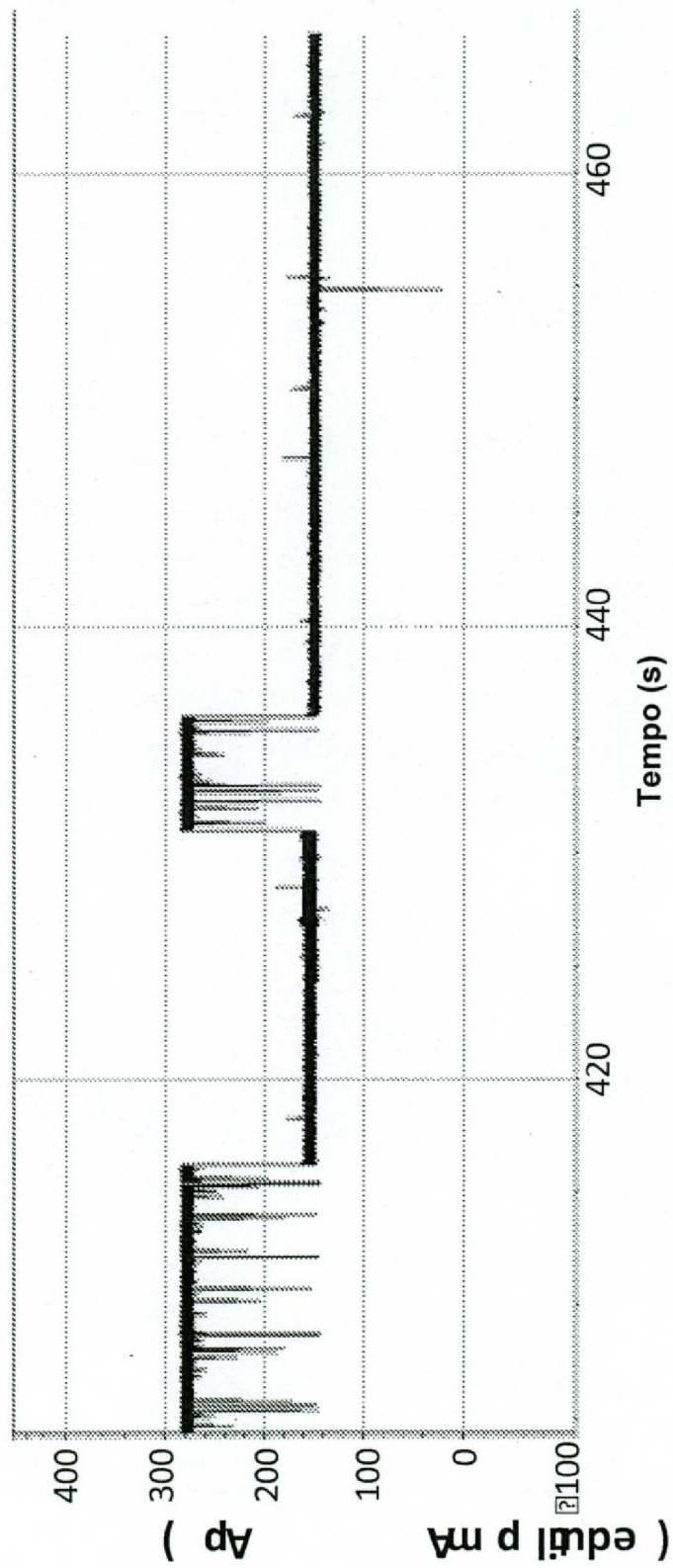


Fig. 9

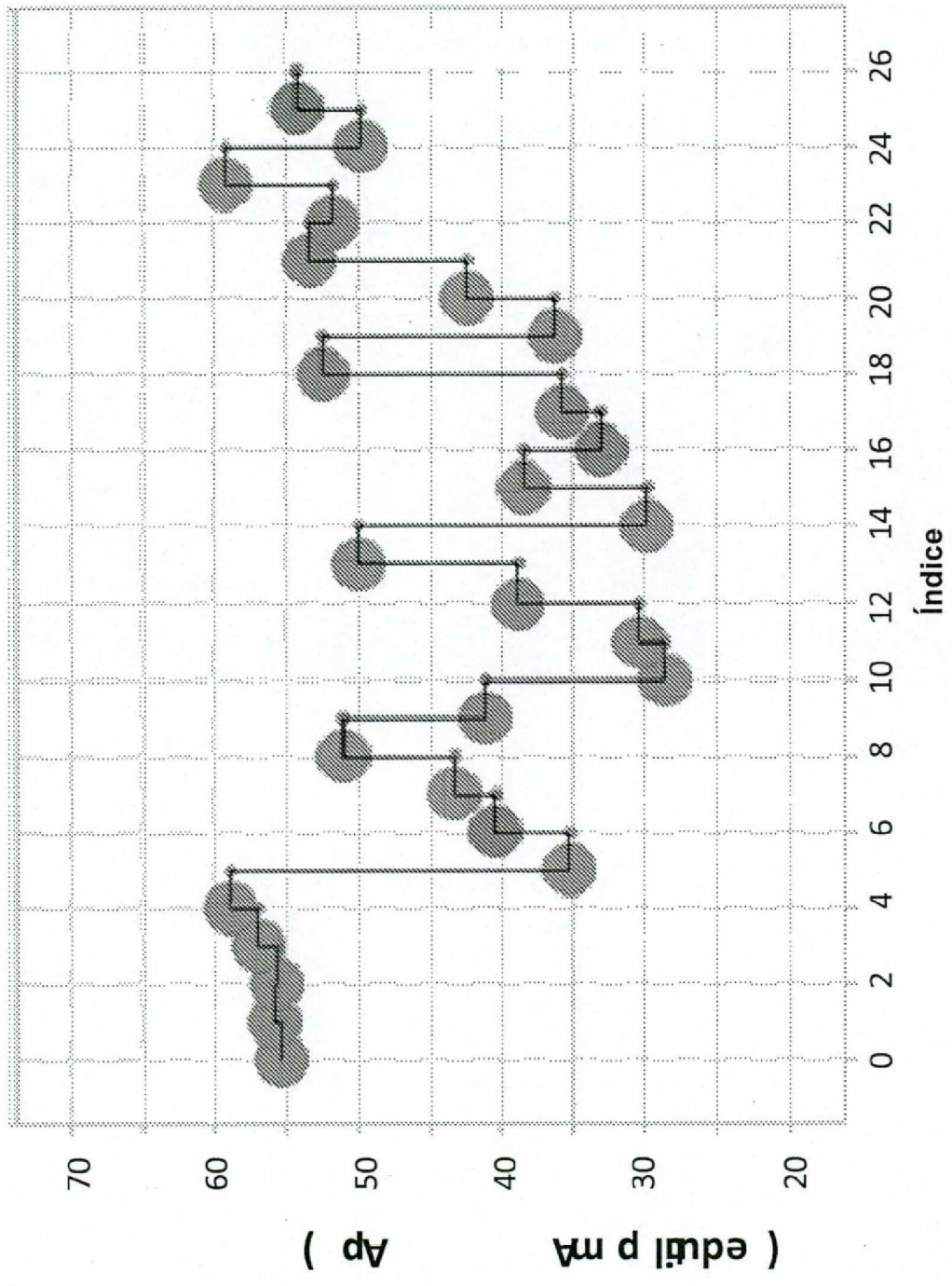


Fig. 10

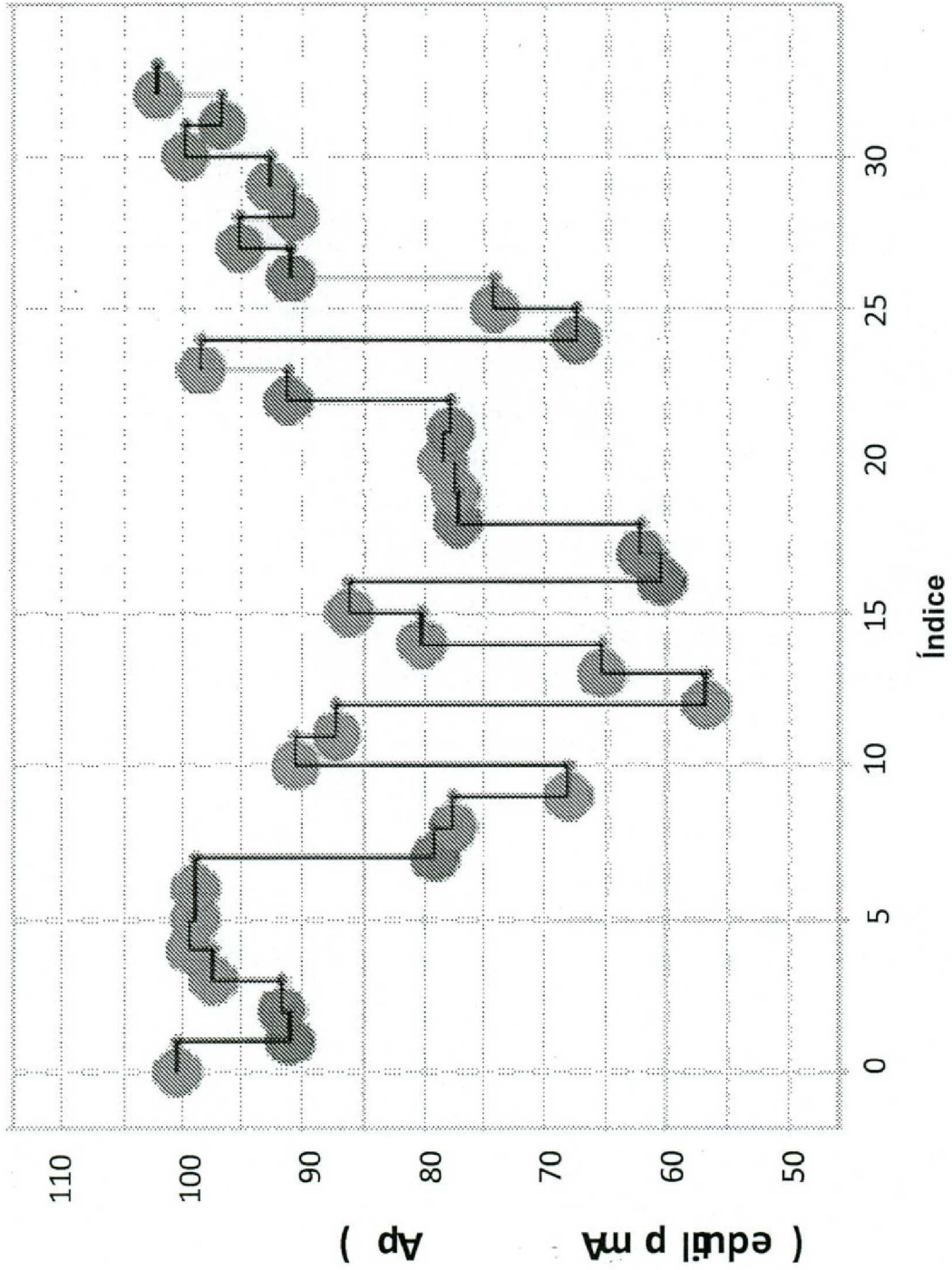


Fig. 11

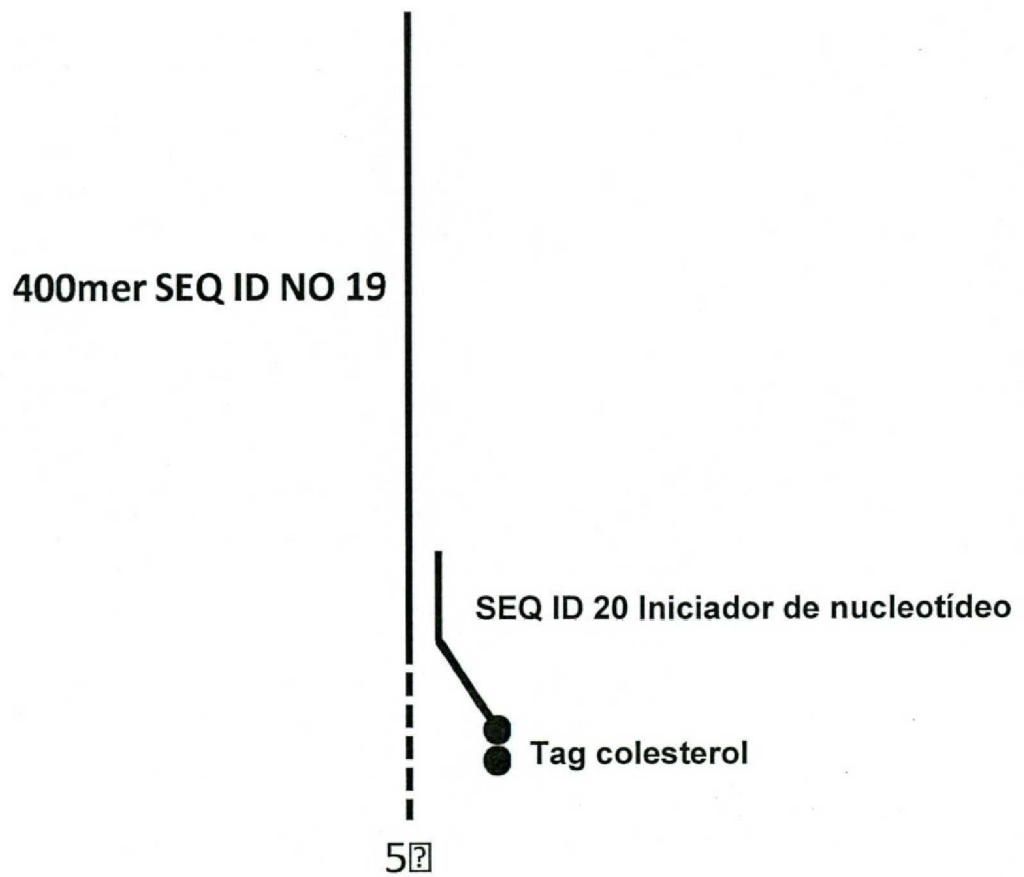


Fig. 12

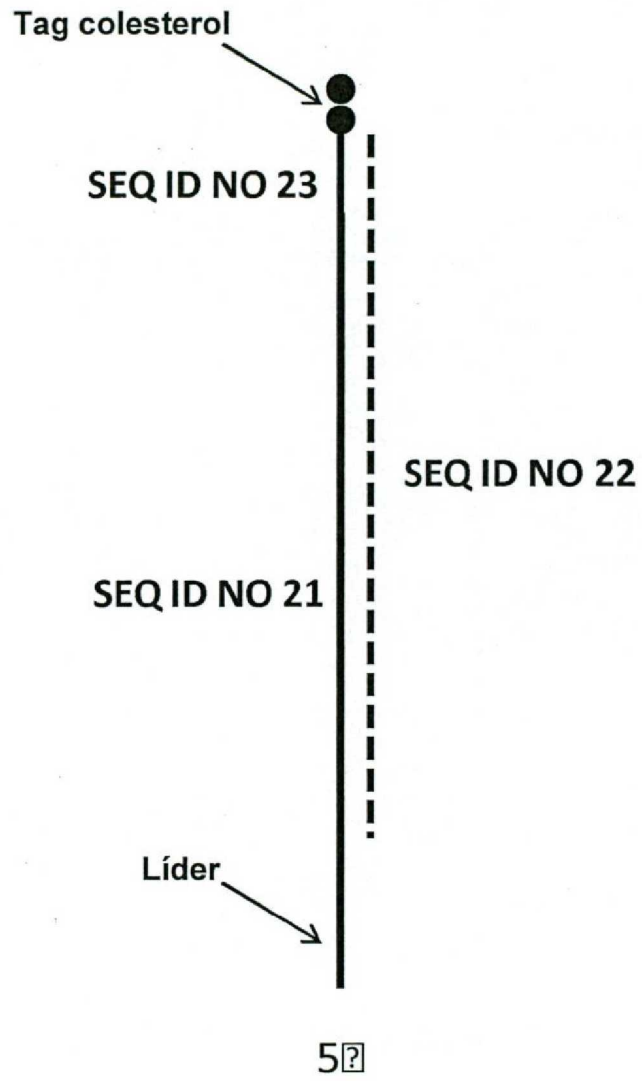
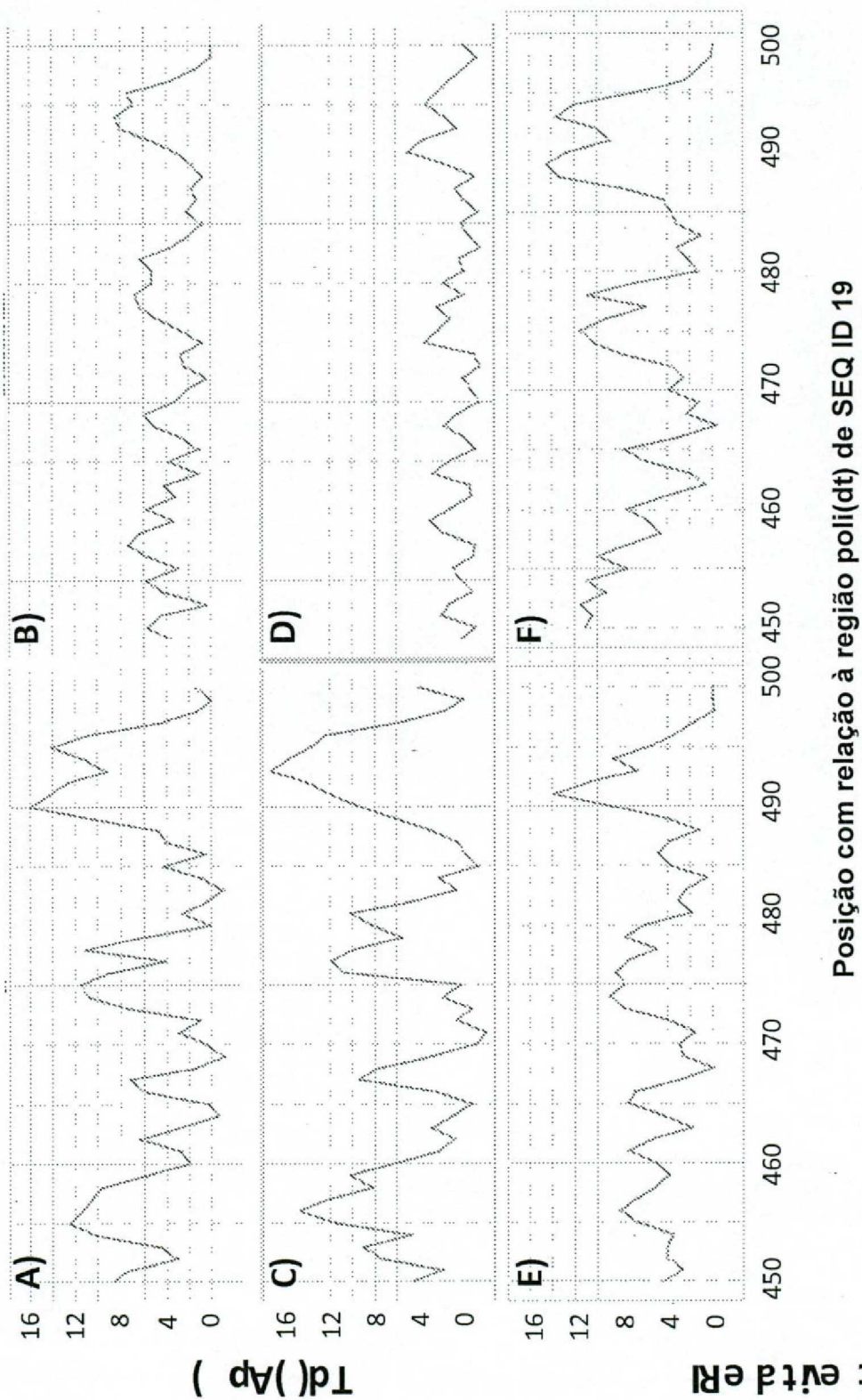


Fig. 13



opelt d evitã eRI

Fig. 14

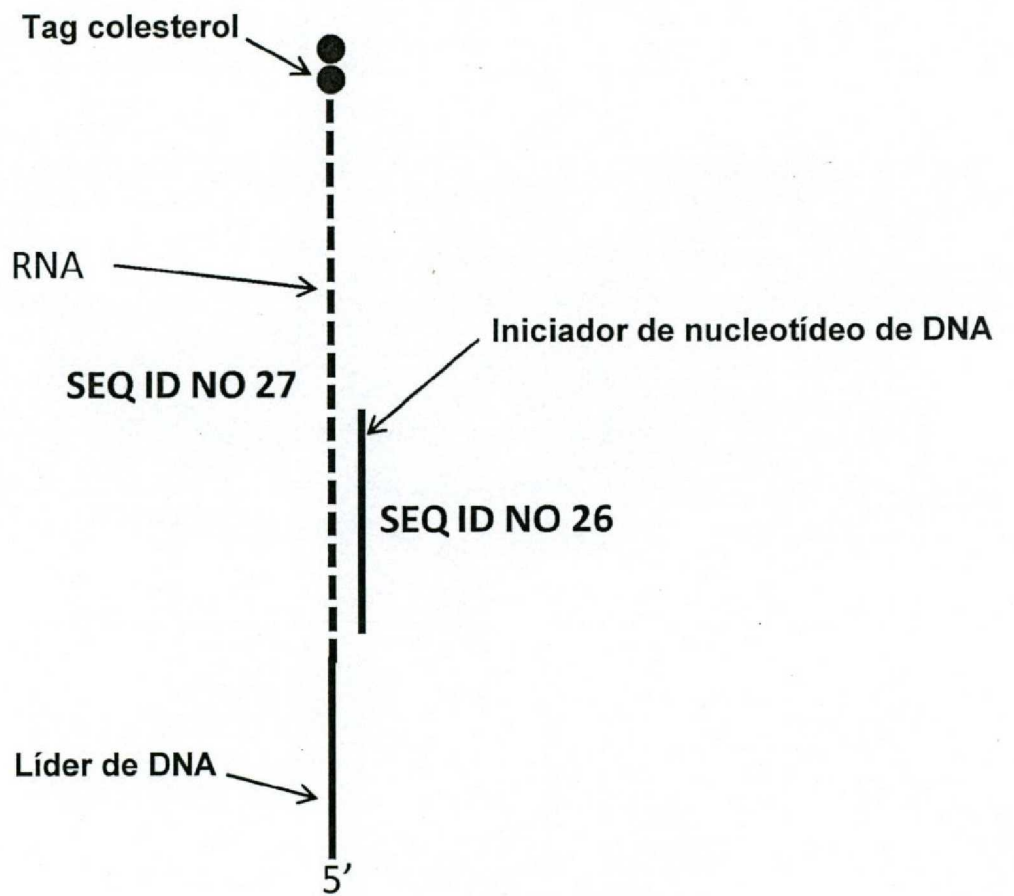


Fig. 15

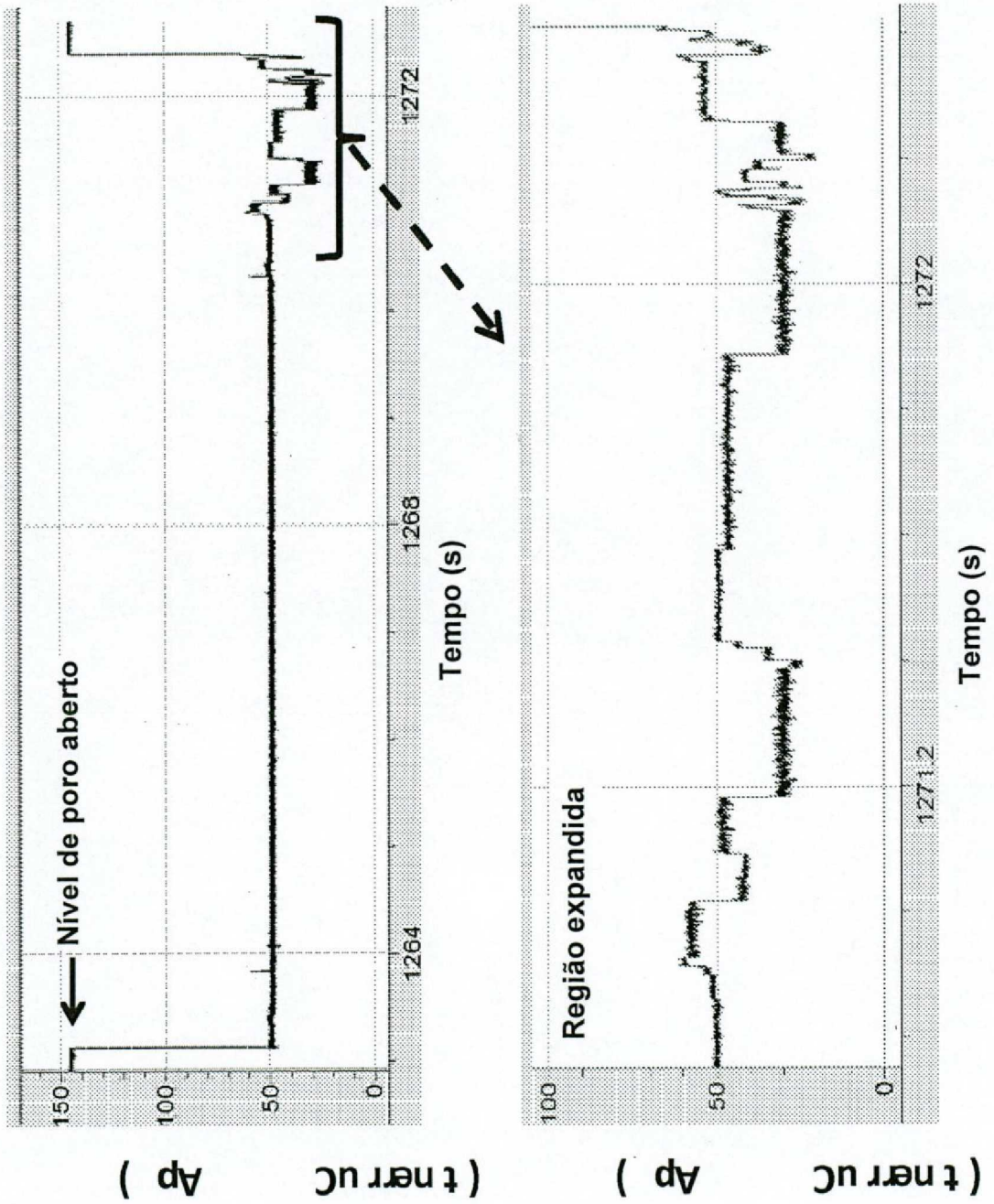


Fig. 16

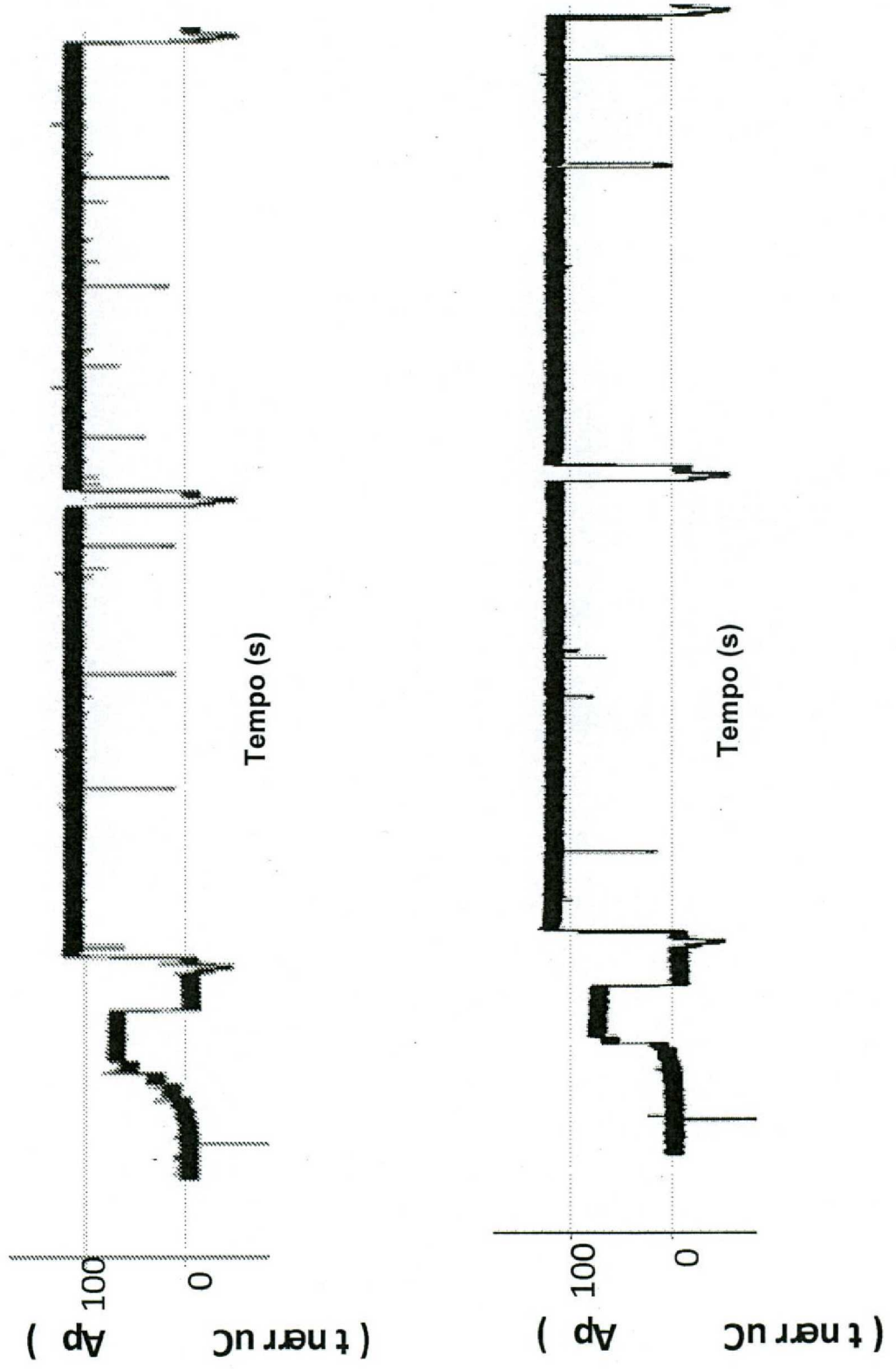


Fig. 17

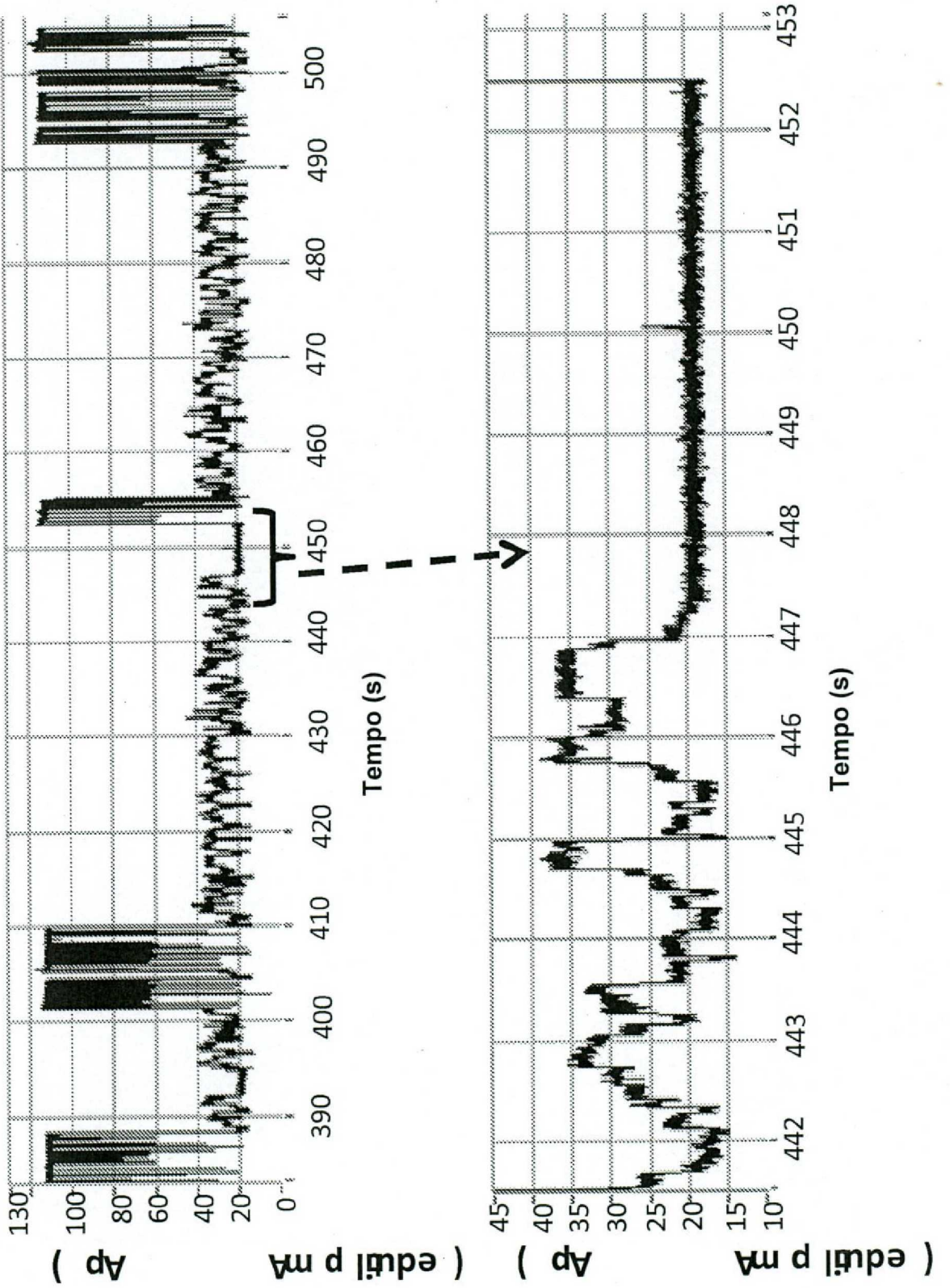


Fig. 18

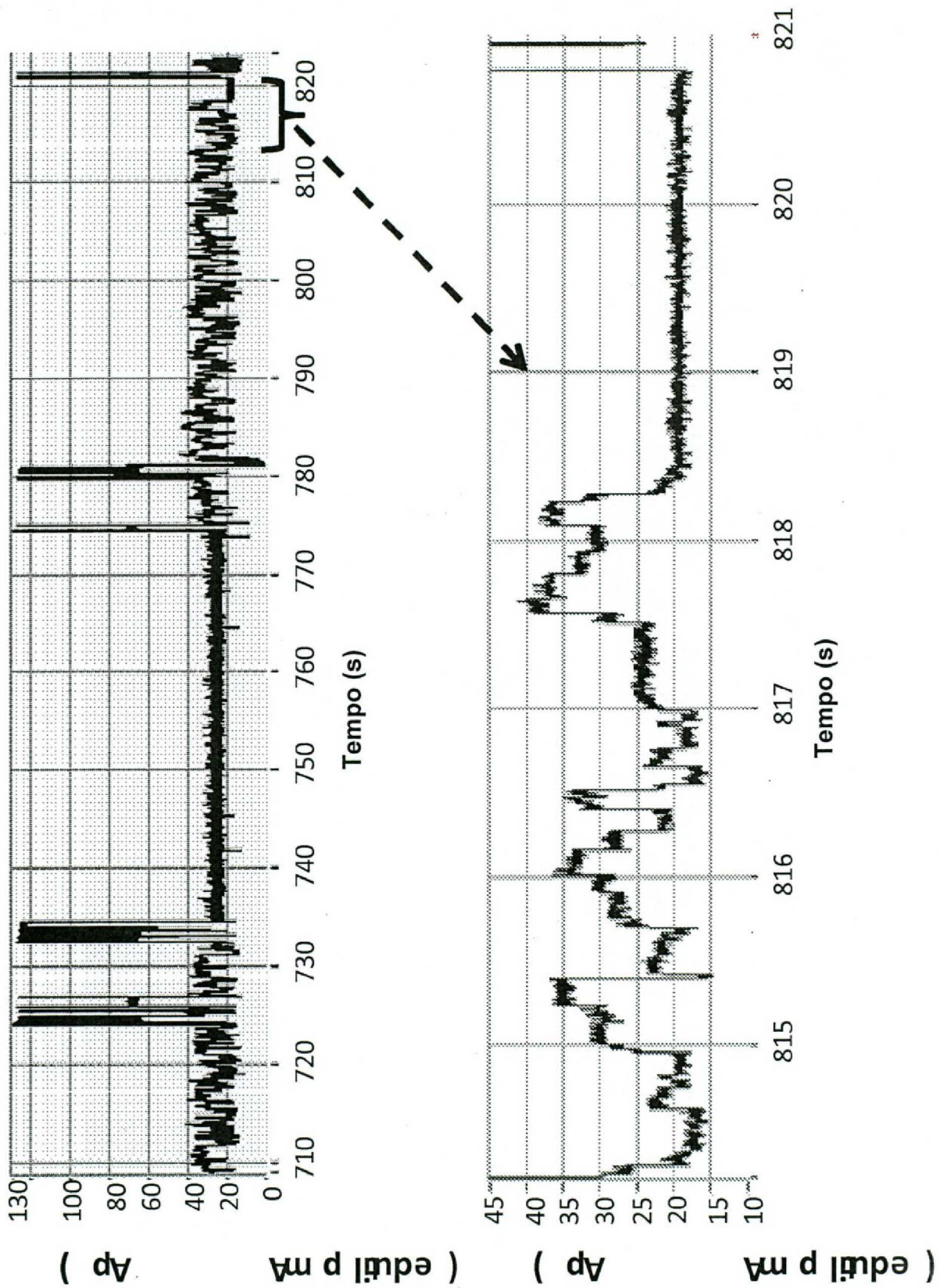


Fig. 19

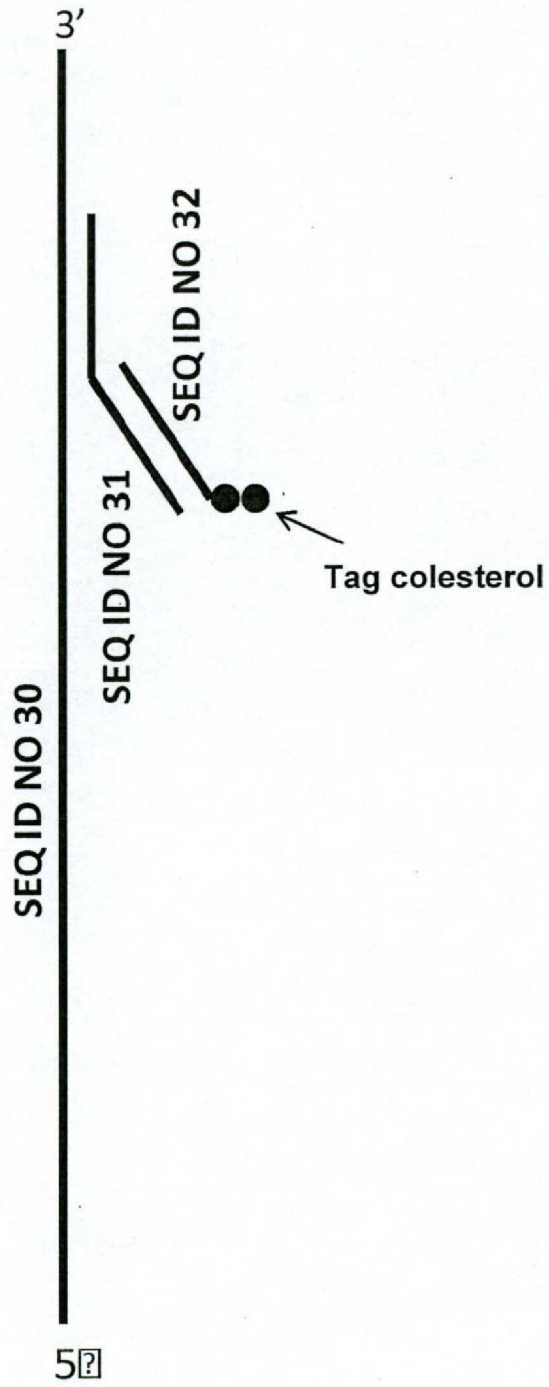


Fig. 20

