



República Federativa do Brasil  
Ministério do Desenvolvimento, Indústria  
e do Comércio Exterior  
Instituto Nacional da Propriedade Industrial.

(21) PI0618391-3 A2



(22) Data de Depósito: 15/11/2006  
(43) Data da Publicação: 30/08/2011  
(RPI 2121)

(51) Int.Cl.:  
C12Q 1/68

(54) Título: MÉTODO PARA IDENTIFICAR UM NUCLEOTÍDEO INDIVIDUAL, MÉTODO PARA SEQUENCIAR UMA SEQUÊNCIA DE ÁCIDO NUCLEICO ALVO, E, KIT PARA SEQUENCIAR UM ÁCIDO NUCLEICO

(30) Prioridade Unionista: 15/11/2005 GB 0523282.2

(73) Titular(es): Isis Innovation Limited

(72) Inventor(es): Hagan Bayley, Orit Braha, Yann Astier

(74) Procurador(es): Momsen, Leonardos & CIA.

(86) Pedido Internacional: PCT GB2006004265 de 15/11/2006

(87) Publicação Internacional: WO 2007/057668de 24/05/2007

(57) Resumo: MÉTODO PARA IDENTIFICAR UM NUCLEOTÍDEO INDIVIDUAL, MÉTODO PARA SEQUENCIAR UMA SEQUÊNCIA DE ÁCIDO NUCLEICO ALVO, E, KIT PARA SEQUENCIAR UM ÁCIDO NUCLEICO. A invenção refere-se a um método de identificação de um nucleotídeo individual, compreendendo (a) contatar o nucleotídeo com um poro de proteína de transmembrana de modo que o nucleotídeo interaja com o poro e (b) medir a corrente passando através do poro durante a interação e deste modo determinar a identidade do nucleotídeo. A invenção também se refere a um método de seqüenciamento de sequências de ácido nucleico e aos kits relacionados com o mesmo.

“MÉTODO PARA IDENTIFICAR UM NUCLEOTÍDEO INDIVIDUAL, MÉTODO PARA SEQÜENCIAR UMA SEQÜÊNCIA DE ÁCIDO NUCLEICO ALVO, E, KIT PARA SEQÜENCIAR UM ÁCIDO NUCLEICO”

5 **Campo da Invenção**

A invenção refere-se à identificação de nucleotídeos individuais e de outros grupos contendo fosfato usado poros de transmembrana. Em particular, a invenção refere-se ao seqüenciamento de ácidos nucleicos alvo usando poros de transmembrana.

10 **Fundamentos da invenção**

O método corrente para seqüenciamento de DNA envolve numerosos reagentes caros tais como ddXTPs fluorescentes, dXTPs, iniciadores e polimerase. Este método requer equipamento sofisticado, que necessita ser operado por um técnico qualificado. Também, este método é limitado às seqüências menores do que um mil nucleotídeos em comprimento.

15 Outros métodos de seqüenciamento têm sido considerados com o objetivo de reduzir o custo, simplificar o método, e permitir que o seqüenciamento ocorra fora do laboratório. Extensão de ciclo, leitura de polimerase, seqüenciamento por exonuclease, e micro-arranjos de DNA são métodos que têm sido considerados (Braslavsky, I., B. Herbert, et al. (2003), *PNAS* 100(7): 3960-3964). Estes métodos têm sido revistos de modo abrangente (Marziali, A. e M. Akeson (2001), *Ann. Rev. Biomed. Eng.* 3: 195-223).

20 Um método potencial de seqüenciamento de DNA é baseado na passagem de uma fita única de DNA através de um nanoporo e identificação de sua seqüência da variação na corrente iônica fluindo através do poro à medida que a fita é passada através do nanoporo (Kasianowicz, J. J., E. Brandin, et al. (1996), *Proc. Natl. Acad. Sci.* 93: 13770-13773). Uma 25 segunda abordagem potencial é seqüenciamento por exonuclease (Chan, E. Y.

(2005), Mutat. Res. **573**: 13-40). Este método envolve digestão de DNA um nucleotídeo por vez (Dapprich, J. (1999), Cytomet. **36**: 163-168; e Matsuura, S.-I., J. Komatsu, et al. (2001), Nuc. Ac. Res. **29**(16): e79) e então identificação de cada um dos nucleotídeos liberados. Contudo, estes métodos 5 requerem modificação do DNA antes da digestão ou modificação dos nucleotídeos uma vez tenham sido liberados do DNA pela exonuclease. O desenvolvimento do seqüenciamento por exonuclease está correntemente sendo retardado pela dificuldade de identificação dos nucleotídeos em nível molecular individual à medida que são liberados pela enzima. Investigadores 10 têm tentado identificar os nucleotídeos usando marcação fluorescente com sucesso limitado.

Detecção estocástica envolve o posicionamento de um poro nanometricamente dimensionado em uma membrana de bicamada lipídica isolante e medição do transporte iônico através do poro. Quando um analito 15 interage com um sítio de ligação dentro do poro, uma mudança na corrente iônica é detectada (Braha, O., B. Walker, et al. (1997), Chem. & Biol. **4**: 497-505; e Bayley, H. e P. S. Cremer (2001), Nature **413**: 226-230). A extensão e a duração do bloco de corrente resultante de cada evento de ligação podem revelar a identidade do analito. A freqüência dos eventos de ligação pode 20 revelar a concentração de analito. Vários sítios de ligação podem ser criados dentro do poro por meio de mutação de proteína, modificação química, e pelo uso de adaptadores e carreadores moleculares (Gu, L.-Q., O. Braha, et al. (1999), Nature **398**: 686-690; r Braha, O., J. Webb, et al. (2005), Chem. Phys. Chem. **6**: 889-892).

## 25 Sumário da Invenção

Tem sido surpreendentemente demonstrado que nucleotídeos individuais podem ser identificados no nível molecular individual de sua amplitude de corrente quando interagem com um poro de transmembrana. Então, detecção estocástica pode ser usada para identificar nucleotídeos

individuais e para seqüenciar seqüências de ácido nucleico via seqüenciamento por exonuclease.

Conseqüentemente, a invenção proporciona um método para identificar um nucleotídeo individual, compreendendo:

5 (a) contatar o nucleotídeo com um poro de proteína de transmembrana de modo que o nucleotídeo interaja com o poro; e

(b) medir a corrente passando através do poro durante a interação e deste modo determinar a identidade do nucleotídeo.

a invenção adicionalmente proporciona:

10 - um método de seqüenciar uma seqüência de ácido nucleico alvo, compreendendo:

(a) digerir um nucleotídeo individual de uma extremidade da seqüência alvo usando uma exonuclease processiva;

15 (b) contatar o nucleotídeo com um poro de proteína de transmembrana de modo que o nucleotídeo interaja com o poro;

(c) medir a corrente passando através do poro durante a interação e deste modo determinar a identidade do nucleotídeo; e

20 (d) repetir as etapas (a) a (c) na mesma extremidade da seqüência de ácido nucleico e determinar deste modo a seqüência de ácido nucleico; e

- um kit para seqüenciar um ácido nucleico, compreendendo:

- uma ciclodextrina; e

- uma exonuclease processiva.

O método de seqüenciar da invenção é um método de seqüenciamento de DNA simples e rápido em nível de molécula individual. 25 Também é um método de seqüenciamento de DNA barato porque não envolve o uso de reagentes caros, tais como fluoroforos.

### **Descrição das Figuras**

Figura 1 mostra o mutante de  $\alpha$ -hemolisina (M113R), e

heptaquis-6-amino- $\beta$ -ciclodextrina (am<sub>7</sub>- $\beta$ CD). A - corte sagital através da estrutura de  $\alpha$ -hemolisina, posição 113 é indicado pela seta. B - estrutura espaço-cheia de am<sub>7</sub>- $\beta$ CD. C -possível interação de am<sub>7</sub>- $\beta$ CD com  $\alpha$ -hemolisina (M113R)<sub>7</sub>

5 Figura 2A mostra detecção de dCMP. A - Traço de corrente de mutante individual (M113R)<sub>7</sub> inserido em uma bicamada fosfolipídica a +130 mV. L1 identifica a corrente do nanoporo de proteína não ocupado. B - na presença de am<sub>7</sub>- $\beta$ CD 40  $\mu$ M na câmara trans. L2 indica o nível de corrente observado quando am<sub>7</sub>- $\beta$ CD se liga temporariamente dentro do nanoporo. C - 10 dCMP 5  $\mu$ M é agora adicionado na câmara cis. L3 mostra o nível de corrente que é observado quando dCMP se liga no complexo temporário (M113R)<sub>7</sub>/am<sub>7</sub>- $\beta$ CD.

15 Figura 2B mostra a interação do poro de  $\alpha$ -hemolisina ( $\alpha$ HL) com heptaquis-(6-desoxi-6-amino)- $\beta$ -ciclodextrina (am<sub>7</sub> $\beta$ CD) e dCMP. A - Modelo do poro  $\alpha$ HL heptamérico (7AHL), no qual Met-113 tem sido substituída por Arg. Um modelo de am<sub>7</sub> $\beta$ CD em seção transversal gerado em ChemDraw Ultra tem sido posicionado manualmente nas distâncias de van der Waals das cadeia laterais de Arg, que bloqueia a passagem da ciclodextrina quando ela entra no poro a partir do lado *trans*. Quando 20 am<sub>7</sub> $\beta$ CD está presente dentro do poro, dois anéis de carga positiva, um anel de sete grupos amino primários contribuídos pela ciclodextrina, e um segundo anel de cadeias laterais de sete argininas, são separados por ~10 $\text{\AA}$  (1,0 nm). Tem sido previamente mostrado que aminociclodextrinas se ligam em nucleosídeo monofosfatos com o grupo fosfato em uma interação iônica com 25 os grupos amino protonados. É possível que a estabilidade total de tais complexos seja intensificada por interações de cátion-p entre as bases de nucleotídeo e as cadeias laterais de Arg. A molécula de dCMP é posicionada de modo que o grupo fosfato interaja com as aminas protonadas de am<sub>7</sub> $\beta$ CD e o anel de citosina interaja com os grupos guanidínio das cadeias laterais de

Arg. B - Traço de corrente de um poro único (M113R)<sub>7</sub> a +130 mV. L1 identifica a corrente fluindo através do nanoporo de proteína não ocupado, que é mostrado como um modelo à direita. C - Traço de corrente após a adição de am<sub>7</sub>βCD 40 μM na câmara *trans*. L2 indica o nível de corrente observado quando am<sub>7</sub>βCD é ligado dentro do nanoporo. D - Traço de corrente após a adição de dCMP 5 μM na câmara *cis*. L3 mostra o nível de corrente que é observado quando dCMP se liga no complexo (M113R)<sub>7</sub>•am<sub>7</sub>βCD.

Figura 3 mostra amplitudes de corrente de dXMP. Traço de corrente de poro individual (M113R)<sub>7</sub> inserido em uma bicamada fosfolipídica, a potencial de +130 mV. am<sub>7</sub>-βCD 40 μM está presente na câmara *trans*. A - dGMP 5 μM é adicionado na câmara *cis*. O histograma de todos os pontos do traço de corrente é mostrado à direita junto com as estruturas de dGMP, dTMP (B), dAMP (C), e dCMP (D).

Figura 4 mostra os níveis de corrente de ciclodextrina. Traço de corrente de mutante individual (M113R)<sub>7</sub> inserido em uma bicamada fosfolipídica com am<sub>7</sub>-βCD 40 μM presente na câmara *trans* a +130 mV. L1 e L1' indicam dois níveis de corrente do nanoporo não ocupado, e L2 e L2' mostram dois níveis de corrente resultantes da ligação de am<sub>7</sub>-βCD em (M113R)<sub>7</sub>. O inserto mostra o histograma de amplitude do traço de corrente com os picos correspondendo aos níveis de corrente L1, L1', L2, e L2'.

Figura 5 mostra análise de evento único. A mostra um histograma de análise de evento único do nível de corrente L3 de todos os quatro dXMP na mesma solução. B mostra um histograma de análise de evento único de L3 originário de apenas L2. 5 μM de dGMP, dAMP, dCMP, e 10 μM de dTMP estão presentes na câmara *cis*.

Figura 6 mostra a detecção simultânea de dXMP. A mostra o traço de corrente de um mutante individual (M113R)<sub>7</sub> inserido em uma bicamada fosfolipídica, potencial de +130 mV é aplicado em eletrodos de

Ag/AgCls. O tampão é Tris-HCl 25 mM pH 8,0 com KCl 1M. am- $\beta$ CD 40  $\mu$ M está presente na câmara trans. 5  $\mu$ M de dGMP, dTMP, dAMP, e dCMP são adicionados na câmara cis. As bandas coloridas ilustram a distribuição de amplitude de cada dXMP. B mostra um histograma de todos os pontos de um traço de corrente de 8000 eventos de ligação. Cada pico está superposto com a distribuição estatística de cada dXMP.

Figura 7 mostra o método estatístico. Duas distribuições gaussianas A e B se sobrepõem no ponto de interseção I. A área de Gaussiana A além do ponto de interseção I está integrada e representa a probabilidade de 10 população A ser identificada como população B.

### **Descrição da Listagem de Seqüências**

SEQ ID NO: 1 mostra a seqüência de polinucleotídeo que codifica uma subunidade de  $\alpha$ -hemolisina.

SEQ ID NO: 2 mostra a seqüência de aminoácidos da uma 15 subunidade de  $\alpha$ -hemolisina.

SEQ ID NO: 3 mostra a seqüência de polinucleotídeo que codifica uma subunidade de  $\alpha$ -hemolisina M113H.

SEQ ID NO: 4 mostra a seqüência de aminoácidos da uma subunidade de  $\alpha$ -hemolisina M113H.

SEQ ID NO: 5 mostra a seqüência de polinucleotídeo que codifica uma subunidade de  $\alpha$ -hemolisina M113K.

SEQ ID NO: 6 mostra a seqüência de aminoácidos da uma subunidade de  $\alpha$ -hemolisina M113K.

SEQ ID NO: 7 mostra a seqüência de polinucleotídeo que codifica uma subunidade de  $\alpha$ -hemolisina M113R.

SEQ ID NO: 8 mostra a seqüência de aminoácidos da uma subunidade de  $\alpha$ -hemolisina M113R.

SEQ ID NO: 9 mostra a seqüência de aminoácidos de lambda exonuclease. A seqüência é uma das três subunidades idênticas que montam-

se em um trímero.

### **Descrição Detalhada da Invenção**

#### **Método para identificar um nucleotídeo individual**

Em uma primeira modalidade, a presente invenção refere-se a 5 um método para identificar um nucleotídeo individual compreendendo contatar o nucleotídeo com um poro de proteína de transmembrana de modo que o nucleotídeo interaja com o poro e medir a corrente passando através do poro durante a interação e deste modo determinar a identidade do nucleotídeo. A invenção portanto envolve a detecção estocástica de um nucleotídeo 10 individual. A invenção pode ser usada para diferenciar nucleotídeos de estrutura similar baseando-se nos efeitos diferentes que têm sobre a corrente passando através de um poro de proteína de transmembrana. A invenção também pode ser usada para determinar se um nucleotídeo particular está ou 15 não presente em uma amostra. A invenção também pode ser usada para medir a concentração de um nucleotídeo particular em uma amostra.

Um nucleotídeo individual de acordo com a invenção é um nucleotídeo individual. Um nucleotídeo individual é um que não está ligado em outro polinucleotídeo por uma ligação nucleotídica. Uma ligação nucleotídica envolve um dos grupos fosfato de um nucleotídeo sendo ligado 20 no grupo açúcar de outro nucleotídeo. Um nucleotídeo individual é tipicamente um que não está ligado por uma ligação nucleotídica em outra seqüência de polinucleotídeo de pelo menos 5, pelo menos 10, pelo menos 20, pelo menos 50, pelo menos 100, pelo menos 200, pelo menos 500, pelo menos 1000 ou pelo menos 5000 nucleotídeos. Por exemplo, o nucleotídeo 25 individual tem sido digerido de uma seqüência de polinucleotídeo alvo, tal como uma fita de DNA ou RNA. O nucleotídeo individual contudo pode ser ligado ou fixado em outros grupos químicos, tais como moléculas fluorescentes ou grupos químicos contendo radioisótopos, e.g.  $^{125}\text{I}$ ,  $^{35}\text{S}$ . Os tipos de nucleotídeos para identificação de acordo com a invenção são

discutidos com mais detalhe abaixo.

O método pode ser realizado usando qualquer sistema de poro/membrana adequado no qual um poro de proteína de transmembrana é inserido em uma membrana. O método é tipicamente realizado usando (i) uma membrana artificial compreendendo um poro de proteína de transmembrana recombinante ou naturalmente ocorrente, (ii) uma membrana naturalmente ocorrente, isolada compreendendo um poro de proteína de transmembrana recombinante, (iii) uma membrana naturalmente ocorrente, isolada compreendendo um poro de proteína de transmembrana ou (iv) uma célula expressando um poro de proteína de transmembrana recombinante ou naturalmente ocorrente. O método é preferivelmente realizado usando uma membrana artificial. A membrana pode compreender outras proteínas de transmembrana e/ou intramembrana bem como outras moléculas em adição ao poro de proteína de transmembrana.

15 O método da invenção é tipicamente realizado *in vitro*.

#### *Membrana*

A membrana forma uma barreira ao fluxo de íons e nucleotídeos. A membrana é preferivelmente uma bicamada lipídica preferivelmente uma bicamada lipídica. Bicamadas lipídicas adequadas para uso de acordo com a invenção podem fazer uso de métodos conhecidos na técnica. Por exemplo, membranas de bicamada lipídica podem ser formadas usando método de Montal e Mueller (1972). O método da invenção pode ser realizado usando bicamadas lipídicas formadas de qualquer lipídeo de membrana incluindo, mas não limitado a, fosfolipídeos, glicolipídeos, colesterol e suas misturas. A bicamada lipídica é preferivelmente formada de 1,2-difitanoil-*sn*-glicero-3-fosfocolina.

São conhecidos na técnica métodos para inserir poros em membranas, tais como bicamadas lipídicas. Por exemplo, o poro pode ser suspenso em uma forma purificada em uma solução contendo uma bicamada

lipídica de tal modo ele se difunda na bicamada lipídica e seja inserido por ligação na bicamada lipídica e montagem em um estado funcional. Alternativamente, o poro pode ser diretamente inserido na membrana usando o método descrito em M.A. Holden, H. Bayley. *J. Am. Chem. Soc.* 2005, **127**, 5 6502-6503.

#### *Poro de proteína de transmembrana*

O método da invenção é realizado usando um poro de proteína de transmembrana. Um poro de proteína de transmembrana é um polipeptídeo que permite que íons fluam de um lado da membrana para o outro ao longo de 10 um gradiente eletroquímico. O poro preferivelmente permite que o nucleotídeo flua de um lado da membrana para o outro ao longo de um gradiente eletroquímico.

O poro é tipicamente um oligômero. O poro é preferivelmente feito de várias subunidades repetidas. O poro é preferivelmente pentamérico 15 ou heptamérico. O poro tipicamente compreende um barril ou canal através do qual os íons podem fluir.

O barril ou canal do poro tipicamente compreende aminoácidos que facilitam a interação com o nucleotídeo. Um poro para uso de acordo com a invenção tipicamente compreende um ou mais aminoácidos 20 positivamente carregados, tais como arginina, lisina ou histidina. Estes aminoácidos positivamente carregados estão preferivelmente localizados próximos da constrição do barril ou canal. Estes aminoácidos tipicamente facilitam a interação entre o poro e o nucleotídeo pela interação com os grupos fosfato no nucleotídeo ou pela interação de cátion-p com a base no 25 nucleotídeo. O poro preferivelmente possui um anel de aminoácidos positivamente carregados, tais como arginina, lisina ou histidina, localizado próximos da constrição do barril ou canal. Cada aminoácido positivamente carregado é tipicamente proporcionado por cada uma das subunidades de poro.

Poros adequados para uso de acordo com a invenção incluem, mas não são limitados a,  $\alpha$ -hemolisina, purinas e leucocidinas.

O poro preferido para uso na invenção é  $\alpha$ -hemolisina ou uma sua variante. O poro de  $\alpha$ -hemolisina é formado de sete subunidades idênticas (heptamérico). A seqüência de uma subunidade de  $\alpha$ -hemolisina é mostrada em SEQ ID NO: 2. Uma variante é um poro heptamérico no qual uma ou mais das sete subunidades possui uma seqüência de aminoácidos que varia daquela da SEQ ID NO: 2 e que retém atividade de poro. 1, 2, 3, 4, 5, 6 ou 7 das subunidades em uma variante de  $\alpha$ -hemolisina podem possuir uma seqüência de aminoácidos que varia daquela da SEQ ID NO: 2. As sete subunidades dentro de um poro variante são tipicamente idênticas mas podem ser diferentes.

Uma variante preferida de  $\alpha$ -hemolisina possui um ou mais aminoácidos positivamente carregados, tais como arginina, lisina ou histidina, localizados próximos do barril ou canal. O poro preferivelmente possui um anel de 4, 5, 6 ou preferivelmente 7 aminoácidos positivamente carregados, tais como arginina, lisina ou histidina, localizados próximos da constrição do barril ou canal. Cada aminoácido no anel é tipicamente proporcionado por cada uma das subunidades variantes. Variantes tipicamente incluem um aminoácido positivamente carregado na posição 113 de cada subunidade. O poro para uso na invenção é preferivelmente  $\alpha$ -hemolisina (M113K)<sub>7</sub> que compreende sete subunidades como mostrado em SEQ ID NO: 4 ou preferivelmente  $\alpha$ -hemolisina (M113H)<sub>7</sub> que compreende sete subunidades como mostrado em SEQ ID NO: 6 ou mais preferivelmente  $\alpha$ -hemolisina (M113R)<sub>7</sub> que compreende sete subunidades como mostrado em SEQ ID NO: 8.

A variante pode ser uma variante naturalmente ocorrente que é expressada por um organismo, por exemplo por uma bactéria *Staphylococcus*. Variantes também incluem variantes não-naturalmente

ocorrentes produzidas por tecnologia recombinante. Sobre o comprimento inteiro da seqüência de aminoácidos de SEQ ID NO: 2, uma subunidade de uma variante preferivelmente será pelo menos 50% homóloga àquela seqüência baseada na identidade de aminoácido. Mais preferivelmente, o 5 polipeptídeo de subunidade 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% homólogo baseado na identidade de aminoácido em relação à seqüência de aminoácidos de SEQ ID NO: 2 sobre o comprimento inteiro. Pode haver pelo 10 menos 80%, por exemplo pelo menos 85%, 90% ou 95%, de identidade de aminoácido sobre um segmento de 200 ou mais, por exemplo 230, 250, 270 ou 280 ou mais, aminoácidos contíguos (“homologia dura”).

Substituições de aminoácido podem ser feitas na seqüência de aminoácidos de SEQ ID NO: 2, por exemplo até 1, 2, 3, 4, 5, 10, 20 ou 30 15 substituições. Substituições conservativas podem ser feitas, por exemplo, de acordo com a seguinte tabela. Aminoácidos no mesmo bloco na segunda coluna e preferivelmente na mesma linha na terceira coluna podem ser substituintes uns dos outros:

NÃO-AROMÁTICO	Não-polar	G A P
		I L V
	Polar – não carregado	C S T M
		N Q
	Polar – carregado	D E
		H K R
AROMÁTICO		H F W Y

Um ou mais resíduos de aminoácido da seqüência de 20 aminoácidos de SEQ ID NO: 2 podem ser alternativamente ou adicionalmente deletados. Até 1, 2, 3, 4, 5, 10, 20 ou 30 resíduos podem ser deletados, ou mais.

Variantes podem incluir subunidades feitas de fragmentos de SEQ ID NO: 2. Tais fragmentos retêm atividade de formação de poro. 25 Fragmentos podem ser de pelo menos 50, 100, 200 ou 250 aminoácidos em

comprimento. Tais fragmentos podem ser usados para produzir poros quiméricos. Um fragmento preferivelmente compreende o domínio de formação de poro de SEQ ID NO: 2.

Variantes incluem poros de proteína quiméricos compreendendo fragmentos ou porções de SEQ ID NO: 2. Poros de proteína quiméricos são formados das subunidades cada uma compreendendo fragmentos ou porções de SEQ ID NO: 2. O poro ou parte de canal de um poro de proteína quimérico é tipicamente formado pelos fragmentos ou porções de SEQ ID NO: 2.

Um ou mais aminoácidos podem ser alternativa ou adicionalmente adicionados nos polipeptídeos descritos acima. Uma extensão pode ser proporcionada na terminação-N ou na terminação-C da seqüência de aminoácidos de SEQ ID NO: 2 ou da variante de polipeptídeo ou de seu fragmento. A extensão pode ser bastante curta, por exemplo de 1 a 10 aminoácidos em comprimento. Alternativamente, a extensão pode ser mais longa, por exemplo de até 50 ou 100 aminoácidos. Uma proteína carreadora pode ser fusionada em uma seqüência de aminoácidos de acordo com a invenção.

Métodos padrão na técnica podem ser usados para determinar homologia. Por exemplo o UWGCG Package proporciona o programa BESTFIT que pode ser usado para calcular homologia, por exemplo usado em suas configurações pré-estabelecidas (Devereux *et al* (1984) *Nucleic Acids Research* 12, p387-395). Os algoritmos PILEUP e BLAST podem ser usados para calcular homologia ou alinhar seqüências (tal como identificar resíduos equivalentes ou seqüências correspondentes (tipicamente em suas configurações pré-estabelecidas)), por exemplo como descrito em Altschul S. F. (1993) *J Mol Evol* 36:290-300; Altschul, S.F *et al* (1990) *J Mol Biol* 215:403-10.

Programa de computador para realizar análises BLAST está

publicamente disponível na National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo envolve primeiro identificação de pares de seqüências de escore alto (HSPs) pela identificação de palavra de comprimento curto  $W$  na seqüência em questão que quer combina quer satisfaz algum escore limite positivamente avaliado  $T$  quando alinhada com a palavra de mesmo comprimento em uma seqüência de banco de dados.  $T$  é referido ao limite de escore de palavra de vizinhança (Altschul *et al*, supra). Estes sucessos de palavra de vizinhança inicial atuam como sementes para iniciação de pesquisas para encontrar HSP's contendo-as. Os sucessos de palavra são estendidos em ambas as direções ao longo de cada seqüência até que o escore de alinhamento cumulativo possa ser aumentado. Extensões para os sucessos de palavra em cada direção são interrompidas quando: o escore de alinhamento cumulativo falha pela quantidade  $X$  de seu valor máximo alcançado; o escore cumulativo vai a zero ou abaixo, devido ao acúmulo de um ou mais alinhamentos de resíduo de escore negativo; ou a extremidade de qualquer seqüência é alcançada. Os parâmetros  $W$ ,  $T$  e  $X$  do algoritmo BLAST determinam a sensibilidade e a velocidade do alinhamento. O programa BLAST usa como valores padrão um comprimento de palavra ( $W$ ) de 11, os alinhamentos de matriz de escore BLOSUM62 (veja Henikoff e Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) (B) de 50, expectativa (E) de 10, M=5, N=4, e uma comparação de ambas as fitas.

O algoritmo BLAST realiza uma análise estatística da similaridade entre duas seqüências; veja e.g., Karlin e Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90: 5873-5787. Uma medição de similaridade proporcionada pelo algoritmo BLAST é a probabilidade de soma mais baixa ( $P(N)$ ), que proporciona uma indicação da probabilidade pela qual uma combinação entre duas seqüências de aminoácidos ocorreria por acaso. Por exemplo, uma seqüência é considerada similar a outra seqüência se a probabilidade de soma mais baixa em comparação da primeira seqüência com

a segunda seqüência é menor do que cerca de 1, preferivelmente menor do que cerca de 0,1, mais preferivelmente menor do que cerca de 0,01, e muito mais preferivelmente menor do que cerca de 0,001.

- Poros usados de acordo com a invenção podem ser
- 5 modificados por exemplo pela adição de resíduos de histidina para ajudar na sua identificação ou purificação ou pela adição de uma seqüência de sinal para promover sua secreção de uma célula onde o polipeptídeo não contém naturalmente uma tal seqüência. Pode ser desejável proporcionar os polipeptídeos em uma forma adequada para fixação em um suporte sólido.
- 10 Por exemplo, o poro pode ser fixado em um suporte sólido com o objetivo de inserir o poro na membrana.

Um poro pode ser marcado com um marcador de revelação. O marcador de revelação pode ser qualquer marcador adequado que permite que o poro seja detectado. Marcadores adequados incluem, mas não são limitados a, moléculas fluorescentes, radioisótopos, e.g.  $^{125}\text{I}$ ,  $^{35}\text{S}$ , enzimas, anticorpos, 15 polinucleotídeos e ligantes tal como biotina.

O poro pode ser isolado de um organismo produtor de poros, tal como *Staphylococcus aureus*, ou feito sinteticamente ou por meio recombinante. Por exemplo, o poro pode ser sintetizado por transcrição de 20 tradução *in vitro*. A seqüência de aminoácidos do poro pode ser modificada para incluir aminoácidos não-naturalmente ocorrentes ou para aumentar a estabilidade do composto. Quando os poros são produzidos por meio sintético, tais aminoácidos podem ser introduzidos durante a produção. Os poros também podem ser modificados após a produção quer sintética quer 25 recombinante.

Os poros também podem ser produzidos usando D-aminoácidos. Em tais casos os aminoácidos serão ligados em seqüência reversa na orientação de C para N. Isto é convencional na técnica para produzir tais proteínas ou peptídeos.

Numerosas modificações de cadeia lateral são conhecidas na técnica e podem ser feitas nas cadeias laterais dos poros. Tais modificações incluem, por exemplo, modificações de aminoácidos por alquilação redutiva pela reação com um aldeído seguida por redução com NaBH<sub>4</sub>, amidinação com metilacetimidato ou acilação com anidrido acético.

Um poro de transmembrana recombinante pode ser produzido usando métodos padrão conhecidos na técnica. Seqüências de ácido nucleico codificadoras de um poro podem ser isoladas e replicadas usando métodos padrão na técnica. Seqüências de ácido nucleico codificadoras de um poro podem ser expressadas em uma célula hospedeira bacteriana usando técnicas padrão na técnica. O poro pode ser introduzido em uma célula por expressão *in situ* do polipeptídeo de um vetor de expressão recombinante. O vetor de expressão opcionalmente traz um promotor induzível para controlar a expressão do polipeptídeo.

Seqüências de ácido nucleico codificadoras de um poro podem ser isoladas e replicadas usando métodos padrão na técnica. DNA cromossômico pode ser extraído de um organismo produtor de poros, tal como *Staphylococcus aureus*. O gene codificador do poro pode ser amplificado usando PCR envolvendo iniciadores específicos. A seqüência amplificada pode ser então incorporada em um vetor replicável recombinante tal como um vetor de clonagem. O vetor pode ser usado para replicar o ácido nucleico em uma célula hospedeira compatível. Assim as seqüências de ácido nucleico codificadoras de um poro podem ser preparadas pela introdução do vetor em uma célula hospedeira compatível, e crescimento da célula hospedeira sob condições que acarretam a replicação do vetor. O vetor pode ser recuperado da célula hospedeira. Células hospedeiras adequadas para clonagem de polinucleotídeos codificadores de um poro são conhecidas na técnica e descritas em mais detalhe abaixo.

A seqüência de ácido nucleico codificadora de um poro pode

ser clonada em um vetor de expressão adequado. Em um vetor de expressão, a seqüência de ácido nucleico codificadora de um poro é tipicamente operacionalmente ligada em uma seqüência de controle que é capaz de proporcionar a expressão da seqüência codificadora pela célula hospedeira.

5 Tais vetores de expressão podem ser usados para expressar um poro.

O termo “operacionalmente ligado” refere-se a uma justaposição na qual os componentes descritos estão em relação permitindo que funcionem em sua maneira intencionada. Uma seqüência de controle “operacionalmente ligada” em uma seqüência codificadora é ligada em uma 10 tal maneira que a expressão da seqüência codificadora é alcançada sob condições compatíveis com as seqüências de controle. Cópias múltiplas dos mesmos genes de poro ou de genes de poro diferentes podem ser introduzidas no vetor.

O vetor de expressão pode ser introduzido em uma célula 15 hospedeira adequada. Assim o método da invenção pode ser realizado em uma célula produzida pela introdução de uma seqüência de ácido nucleico codificadora de um poro em um vetor de expressão, introdução do vetor em uma célula hospedeira bacteriana compatível, e crescimento da célula hospedeira sob condições que acarretam a expressão da seqüência de ácido 20 nucleico codificadora do poro. Alternativamente, o poro recombinante produzido nesta maneira pode ser isolado da célula hospedeira bacteriana e inserido em outra membrana.

Os vetores podem ser por exemplo, vetores plasmídeo, vírus 25 ou fago proporcionado com uma origem de replicação, opcionalmente um promotor para a expressão de citada seqüência de ácido nucleico e opcionalmente um regulador do promotor. Os vetores podem conter um ou mais genes marcadores selecionáveis, por exemplo um gene de resistência à tetraciclina. Promotores e outros sinais de regulação de expressão podem ser selecionados para serem compatíveis com a célula hospedeira para que o

vetor de expressão seja planejado. Um promotor T7, *trc*, *lac*, *ara* ou  $\lambda_L$  é tipicamente usado.

A célula hospedeira tipicamente expressa o poro em um nível alto. Células hospedeiras transformadas com uma seqüência de ácido nucleico 5 codificadora de um poro serão escolhidas para serem 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  $\lambda$  DE3, por exemplo C41 (DE3), BL21 (DE3), JM109 (DE3), B834 (DE3), TUNER, Origami e Origami B, pode expressar um vetor 10 compreendendo o promotor T7.

Um poro pode ser produzido em grande escala após a purificação por qualquer sistema de cromatografia líquida de proteína dos organismos produtores de poros ou após expressão recombinante como descrito acima. Sistemas de cromatografia líquida de proteína típicos incluem 15 sistemas FPLC, AKTA, o sistema Bio-Cad, o sistema Bio-Rad BioLogic e o sistema Gilson HPLC. O poro naturalmente ocorrente ou recombinantemente produzido pode ser então inserido em uma membrana naturalmente ocorrente ou artificial para uso de acordo com a invenção.

O método da invenção pode empregar qualquer um dos poros 20 descritos acima.

#### *Interação entre o poro e nucleotídeo*

O nucleotídeo pode ser contatado com o poro sobre qualquer lado da membrana. O nucleotídeo pode ser introduzido no poro sobre qualquer lado da membrana. O nucleotídeo é preferivelmente contatado com 25 o poro sobre um lado da membrana que permite que íons entrem no poro e fluam através da membrana ao longo de um gradiente eletroquímico. O nucleotídeo é preferivelmente contatado com um lado da membrana que permite que o nucleotídeo passe através do poro para o outro lado da membrana. Por exemplo, o nucleotídeo é contatado com uma extremidade do

poro que em seu ambiente nativo permite a entrada de íons ou moléculas pequenas, tais como nucleotídeos, para dentro do barril ou canal do poro de tal modo que íons ou moléculas pequenas possam passar através do poro.

O nucleotídeo pode interagir com o poro em qualquer maneira e em qualquer sítio. O nucleotídeo preferivelmente se liga reversivelmente no poro. O nucleotídeo mais preferivelmente se liga reversivelmente no barril ou no canal do poro. O nucleotídeo mais preferivelmente se liga reversivelmente no canal ou barril do poro à medida que passa através do poro através da membrana.

Durante a interação entre o nucleotídeo e o poro, o nucleotídeo afeta a corrente fluindo através do poro em uma maneira específica para aquele nucleotídeo. Por exemplo, um nucleotídeo particular reduzirá a corrente fluindo através do poro por um período de tempo particular e em uma extensão particular. Experimentos de controle podem ser realizados para determinar o efeito que um nucleotídeo particular tem sobre a corrente fluindo através do poro. Resultados da realização do método da invenção sobre uma amostra de teste podem ser então comparados com aqueles derivados de um tal experimento de controle com o objetivo de identificar um nucleotídeo particular na amostra ou determinar se um nucleotídeo particular está presente na amostra. A freqüência na qual a corrente fluindo através do poro é afetada em uma maneira indicativa de um nucleotídeo particular pode ser usada para determinar a concentração daquele nucleotídeo na amostra.

#### *Aparelho*

O método pode ser realizado usando qualquer aparelhagem que é adequada para investigar um sistema de membrana/poro no qual um poro de proteína de transmembrana é inserido em uma membrana. O método pode ser realizado usando qualquer aparelhagem que é adequada para detecção estocástica. Por exemplo, a aparelhagem compreende uma câmara compreendendo uma solução aquosa e uma barreira que separa a câmara em

duas seções. A barreira possui uma abertura na qual a membrana compreendendo o poro é formada. O nucleotídeo pode ser contatado com o poro pela introdução do nucleotídeo na câmara. O nucleotídeo pode ser introduzido em qualquer uma das duas seções da câmara.

5 O método da invenção envolve uma corrente passando através do poro durante a interação com o nucleotídeo. Portanto a aparelhagem também compreende um circuito elétrico capaz de aplicar e medir um sinal elétrico através da membrana e poro. O método pode ser realizado usando um *patch clamp* ou um *voltage clamp*. O método preferivelmente envolve o uso  
10 de um *patch clamp*. O Exemplo descreve um meio de realizar um método de *patch clamp*.

#### *Adaptador Molecular*

A transmembrana poro preferivelmente compreende um adaptador molecular que facilita a interação entre o poro e o nucleotídeo. O adaptador tipicamente possui um efeito sobre as propriedades físicas ou químicas do poro que melhora sua interação com o nucleotídeo. O adaptador tipicamente altera a carga do barril ou canal do poro ou especificamente interage com ou se liga no nucleotídeo facilitando deste modo sua interação com o poro. O adaptador preferivelmente interage com um ou mais grupos  
20 fosfato sobre o nucleotídeo ou interage com a base no nucleotídeo por interação de cátion-p. O adaptador pode mediar a interação entre o nucleotídeo e o poro. Por exemplo, o nucleotídeo pode se ligar reversivelmente no poro via o adaptador. Alternativamente, o adaptador pode interagir com o nucleotídeo conjuntamente com o poro. Por exemplo, o  
25 nucleotídeo pode se ligar reversivelmente em ambos o poro e o adaptador. O adaptador preferivelmente constringe o barril ou canal de modo que ele pode interagir com o nucleotídeo.

O próprio adaptador pode reversivelmente interagir com o poro e pode portanto se mover para dentro e para fora do barril ou do canal do

poro. Alternativamente, o adaptador pode ser covalentemente ligado no barril ou canal do poro de modo que ele não pode sair.

O adaptador tipicamente possui um anel de grupos amino. O adaptador preferivelmente possui um anel de sete grupos amino. Este anel de grupos amino pode interagir com o nucleotídeo em combinação com um anel de aminoácidos positivamente carregados na constrição do barril ou canal do poro.

Um adaptador adequado é ciclodextrina. O adaptador é preferivelmente heptaquis-6-amino- $\beta$ -ciclodextrina (am<sub>7</sub>- $\beta$ -CD).

#### 10 *Nucleotídeo*

O método da invenção pode ser usado para identificar qualquer nucleotídeo. O nucleotídeo pode ser naturalmente ocorrente ou artificial. Um nucleotídeo tipicamente contém uma nucleobase, um açúcar e pelo menos um grupo fosfato. A nucleobase é tipicamente heterocíclica. 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 não são limitados a, ribose e desoxirribose. O nucleotídeo é tipicamente um ribonucleotídeo ou desoxirribonucleotídeo. O nucleotídeo tipicamente contém um monofosfato, difosfato ou trifosfato.

20 Nucleotídeos adequados incluem, mas não são limitados a, adenosina monofosfato (AMP), adenosina difosfato (ADP), adenosina trifosfato (ATP), guanosina monofosfato (GMP), guanosina difosfato (GDP), guanosina trifosfato (GTP), timidina monofosfato (TMP), timidina difosfato (TDP), timidina trifosfato (TTP), uridina monofosfato (UMP), uridina difosfato (UDP), uridina trifosfato (UTP), citidina monofosfato (CMP), citidina difosfato (CDP), citidina trifosfato (CTP), adenosina monofosfato cíclico (cAMP), guanosina monofosfato cíclico (cGMP), desoxiadenosina monofosfato (dAMP), desoxiadenosina difosfato (dADP), desoxiadenosina trifosfato (dATP), desoxiguanosina monofosfato (dGMP), desoxiguanosina

5 difosfato (dGDP), desoxiguanosina trifosfato (dGTP), desoxitimidina monofosfato (dTDP), desoxitimidina difosfato (dTDP), desoxitimidina trifosfato (dTTP), desoxiuridina monofosfato (dUMP), desoxiuridina difosfato (dUDP), desoxiuridina trifosfato (dUTP), desoxicitidina monofosfato (dCMP), desoxicitidina difosfato (dCDP) e desoxicitidina trifosfato (dCTP). O nucleotídeo é preferivelmente AMP, TMP, GMP, UMP, dAMP, dTMP, dGMP ou dCMP.

10 O nucleotídeo pode ser derivado da digestão de uma seqüência de ácido nucleico tal como ácido ribonucleico (RNA) ou ácido desoxirribonucleico. Nucleotídeos individuais de uma única seqüência de ácido nucleico podem ser contatados com o poro em uma maneira seqüencial com o objetivo de seqüenciar todo o ou parte do ácido nucleico. Seqüenciamento de ácidos nucleicos de acordo com a segunda modalidade da invenção é discutido com mais detalhe abaixo.

15 O nucleotídeo é tipicamente não modificado, tal como quando o nucleotídeo é derivado da digestão de uma seqüência de ácido nucleico. Alternativamente, o nucleotídeo pode ser modificado ou danificado. O nucleotídeo é tipicamente metilado. O nucleotídeo pode ser marcado com um marcador de revelação. O marcador de revelação pode ser qualquer marcador 20 adequado que permite que o nucleotídeo seja detectado. Marcadores adequados incluem moléculas fluorescentes, radioisótopos, e.g.  $^{125}\text{I}$ ,  $^{35}\text{S}$ , e ligantes tal com biotina.

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

invenção pode ser realizada *in vitro* sobre uma amostra obtida de ou extraída de qualquer organismo ou microorganismo. O organismo ou microorganismo é tipicamente procariótico ou eucariótico e tipicamente pertence a um dos cinco reinos: plantas, animais, fungos, moneras e protistas. A invenção pode 5 ser realizada *in vitro* sobre uma amostra obtida de ou extraída de qualquer vírus. A amostra é preferivelmente uma amostra de fluido. A amostra tipicamente compreende um fluido corporal 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 10 alternativamente pode ser de outro animal mamífero tal como de animais comercialmente criados tais como cavalos, gado bovino, ovelhas ou porcos ou pode ser alternativamente animais de estimação tais como cães ou gatos.

A amostra é tipicamente processada antes de ser ensaiada, por exemplo por centrifugação ou pela passagem através de uma membrana que 15 filtra moléculas ou células desconhecidas, tais como células vermelhas do sangue. A amostra pode ser medida imediatamente após ter sido tirada. A amostra também pode ser tipicamente armazenada antes do ensaio, preferivelmente abaixo de -70°C.

#### *Condições*

20 O método da invenção envolve a medição de uma corrente passando através do poro durante a interação com o nucleotídeo. Condições adequadas para medir correntes iônicas através dos poros de proteína de transmembrana são conhecidas na técnica e descritas no Exemplo. O método é realizado com uma voltagem aplicada através da membrana e do poro. A 25 voltagem usada é tipicamente de +50 mV a +200 mV. A voltagem usada é preferivelmente de +70 mV a +150 mV, de +85 mV a +145 mV ou de +100 mV a +140 mV. A voltagem usada é preferivelmente de cerca de +130 mV para desoxribonucleotídeos 5' monofosfato, tais como dAMP, dTMP, dGMP e dCMP, e +110 mV para ribonucleotídeos 5' monofosfato, tais como AMP,

### TMP, GMP e UMP.

O método é realizado na presença de qualquer sal de cloreto de metal alcalino. Na aparelhagem exemplar discutida acima, o sal está presente na solução aquosa na câmara. Cloreto de potássio (KCl), cloreto de sódio (NaCl) ou cloreto de césio (CsCl) é tipicamente usado. KCl é preferido. A concentração de sal é tipicamente de 0,1 a 2M, de 0,3 a 1,9M, de 0,5 a 1,8M, de 0,7 a 1,7M, de 0,9 a 1,6M ou de 1M a 1,4M. A concentração de sal é preferivelmente de cerca de 1M.

O método é tipicamente realizado na presença de um tampão.

10 Na aparelhagem exemplar discutida acima, o tampão está presente na solução aquosa na câmara. Qualquer tampão pode ser usado no método da invenção. Um tampão adequado é o tampão Tris-HCl. O método é tipicamente realizado em um pH de 7,5 a 12,0, de 7,6 a 11,0, de 7,7 a 10,0, de 7,8 a 9,5, de 8,0 a 9,0 ou de 8,0 a 8,5. O pH usado é preferivelmente de cerca de 8,0.

15 O método é tipicamente realizado a de 14°C a 100°C, de 15°C a 90°C, de 16°C a 80°C, de 17°C a 70°C, de 18°C a 60°C, 19°C a 50°C, ou de 20°C a 40°C. O método é preferivelmente realizado na temperatura ambiente.

20 O método é preferivelmente realizado a +130mV a pH 8,0, KCl 1M para desoxirribonucleotídeos 5' monofosfato, tais como dAMP, dTMP, dGMP e dCMP, e a +110mV a pH 8,0, KCl 1M para ribonucleotídeos 5' monofosfato, tais como AMP, TMP, GMP e UMP.

### Método de seqüenciamento de ácidos nucleicos

Em uma segunda modalidade, a invenção refere-se a um método de seqüenciar uma seqüência de ácido nucleico alvo, compreendendo 25 (a) digerir um nucleotídeo individual de uma extremidade da seqüência alvo usando uma exonuclease processiva; (b) contatar o nucleotídeo com um poro de proteína de transmembrana de modo que o nucleotídeo interaja com o poro; (c) medir a corrente passando através do poro durante a interação e deste modo determinar a identidade do nucleotídeo; e (d) repetir as etapas (a)

a (c) na mesma extremidade da seqüência de ácido nucleico e determinar deste modo a seqüência de ácido nucleico. Então, a segunda modalidade envolve a detecção estocástica de cada nucleotídeo individual de uma seqüência de ácido nucleico em uma maneira sucessiva com o objetivo de 5 seqüenciar o ácido nucleico. Todo o ou apenas parte do ácido nucleico pode ser seqüenciado usando o método da segunda modalidade. O ácido nucleico pode ser naturalmente ocorrente ou artificial. Por exemplo, o método da segunda modalidade pode ser usado para verificar uma seqüência de um oligonucleotídeo manufaturado. O método da segunda modalidade é 10 tipicamente realizado *in vitro*.

Etapas (b) e (c) do método da segunda modalidade são geralmente idênticas às etapas realizadas no método da primeira modalidade discutida acima. Toda a discussão acima referente à primeira modalidade, e 15 em particular referente às membranas, aparelhagem, poros, adaptadores moleculares, nucleotídeos e condições que podem ser usados na primeira modalidade, igualmente se aplicam à segunda modalidade. O ácido nucleico na segunda modalidade está tipicamente presente em qualquer amostra biológica como discutida acima para a primeira modalidade. O método da segunda modalidade pode ser realizado sobre uma amostra que contém um ou 20 mais ácidos nucleicos cuja seqüência é desconhecida. Alternativamente o método da segunda modalidade pode ser realizado sobre uma amostra para confirmar a identidade dos ácidos nucleicos cuja presença na amostra é conhecida ou esperada. A seqüência de ácido nucleico é tipicamente amplificada antes de ser seqüenciada usando o método da segunda 25 modalidade.

#### *Exonuclease processiva*

O método da segunda modalidade envolve contatar a seqüência de ácido nucleico com uma exonuclease processiva para liberar nucleotídeos individuais de uma extremidade do ácido nucleico. Exonucleases

processivas são enzimas que tipicamente agarram-se sobre uma extremidade da seqüência de ácido nucleico e digerem a seqüência um nucleotídeo por vez naquela extremidade. A exonuclease processiva pode digerir o ácido nucleico na direção de 5' para 3' ou na direção de 3' para 5'. A extremidade do ácido nucleico na qual a exonuclease processiva se liga é tipicamente determinada através da escolha da enzima usada e/ou usando métodos conhecidos na técnica. Grupos hidroxila ou estruturas *cap* em qualquer extremidade da seqüência de ácido nucleico podem ser tipicamente usadas para prevenir ou facilitar a ligação da exonuclease processiva em uma extremidade particular da seqüência de ácido nucleico.

Qualquer enzima exonuclease processiva pode ser usada no método da invenção. A enzima preferida para uso no método da invenção é lambda exonuclease. A seqüência de uma subunidade de lambda exonuclease é mostrada em SEQ ID NO: 9. As três subunidades idênticas interagem para formar uma exonuclease trimérica. Variantes de lambda exonuclease são enzimas formadas de subunidades de polipeptídeo que possuem uma seqüência de aminoácidos que varia daquela da SEQ ID NO: 9 e que retêm a atividade de exonuclease processiva. As variantes podem variar de SEQ ID NO: 9 na mesma maneira e na mesma extensão que como discutido para as variantes de SEQ ID NO: 2 acima. Uma variante preferivelmente compreende os domínios responsáveis pela ligação no ácido nucleico e para digerir o ácido nucleico (domínio catalítico). Uma variante preferivelmente possui uma taxa reduzida de atividade enzimática e/ou tolerância maior a sal comparada com a enzima de tipo selvagem. A exonuclease processiva pode ser produzida usando qualquer um dos métodos discutidos acima para a produção de poros de proteína de transmembrana.

O método da segunda modalidade envolve contatar a seqüência de ácido nucleico com a exonuclease processiva de modo que os nucleotídeos seja digeridos da extremidade do ácido nucleico em uma

velocidade que permite a identificação de cada nucleotídeo individual de acordo com a primeira modalidade da invenção. Métodos para realizar isso são bem conhecidos na técnica. Por exemplo, degradação de Edman é usada para sucessivamente digerir aminoácidos individuais da extremidade de 5 polipeptídeo de tal modo que podem ser identificados usando Cromatografia Líquida de Desempenho Alto (HPLC). Um método homólogo pode ser usado na presente invenção.

A exonuclease processiva é preferivelmente covalentemente ligada no poro de proteína de transmembrana. Métodos para covalentemente 10 ligar a exonuclease processiva no poro são bem conhecidos na técnica.

A velocidade na qual a exonuclease processiva tem que funcionar no método da segunda modalidade é tipicamente mais lenta do que a velocidade ótima de uma exonuclease processiva de tipo selvagem. Uma velocidade adequada de atividade da exonuclease processiva no método da 15 segunda modalidade envolve digestão 0,5 a 1000 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 velocidade é preferivelmente 1, 10, 100, 500 ou 1000 nucleotídeos por segundo. Uma velocidade adequada de 20 atividade de exonuclease processiva pode ser alcançada em várias maneiras. Por exemplo, exonucleases processivas variantes com uma velocidade ótima reduzida de atividade podem ser usadas de acordo com a invenção.

A atividade de exonucleases processivas é tipicamente dependente de pH de tal modo que sua atividade cai à medida que o pH é 25 reduzido. Então, o método da segunda modalidade é tipicamente realizado em um pH de 7,5 a 8,0 ou de 7,7 a 8,0. O pH usado é preferivelmente de cerca de 8,0.

A velocidade de atividade de exonucleases processivas tipicamente cai à medida que a concentração de sal aumenta. Contudo,

concentrações de sal muito altas tipicamente possuem um efeito prejudicial sobre a atividade da enzima. Outro modo de limitar a velocidade da enzima é realizar o método da segunda modalidade em uma concentração de sal que reduz a velocidade da atividade da enzima sem afetar prejudicialmente sua atividade. Por exemplo, o método da segunda modalidade pode ser realizado em uma concentração de sal de 0,5 a 1M. A concentração de sal é preferivelmente de cerca de 1M.

### Kits

Em uma terceira modalidade, a invenção também se refere a kits que podem ser usados para realizar a segunda modalidade da invenção. Os kits são portanto adequados para seqüenciar ácidos nucleicos. Os kits compreendem uma ciclodextrina e uma exonuclease processiva. A ciclodextrina é preferivelmente heptaquis-6-amino- $\beta$ -ciclodextrina. A exonuclease processiva pode ser qualquer uma daquelas discutidas acima com referência à segunda modalidade. O kit preferivelmente adicionalmente compreende um poro de proteína de transmembrana. O poro pode ser qualquer um daqueles discutidos acima com referência à primeira modalidade.

O kit pode adicionalmente compreender um ou mais outros reagentes ou instrumentos que permitem que qualquer uma das modalidades do método mencionado acima seja realizada. Tais reagentes ou instrumentos incluem um ou mais dos seguintes: tampão(ões) adequado(s) (soluções aquosas), meio para obter uma amostra de um indivíduo (tal como um vaso ou um instrumento compreendendo uma agulha), meio para amplificar a seqüência de ácidos nucleicos, uma membrana como definida acima ou aparelhagem de *voltage clamp* ou *patch clamp*. Reagentes podem estar presentes no kit em um estado seco de tal modo que uma amostra de fluido ressuspense os reagentes. O kit também, opcionalmente, compreende instruções para permitir que o kit seja usado no método da invenção ou

detalhes relacionados com quais pacientes o método pode ser usado. O kit pode, opcionalmente, compreender nucleotídeos.

O seguinte Exemplo ilustra a invenção:

### Exemplo

5 Com o objetivo de trazer o tamanho do percurso de condução iônica do mutante de  $\alpha$ -hemolisina (M113R)<sub>7</sub> (Figura 1A) para mais próximo do tamanho do nucleotídeo a ser detectado, o diâmetro do nanoporo foi reduzido pela acomodação de uma ciclodextrina próxima da constrição do poro. A heptaquis-6-amino- $\beta$ -ciclodextrina (am<sub>7</sub>- $\beta$ CD) (Figura 1B), que 10 possui sete grupos amino nas posições primárias, foi usada. Quando a ciclodextrina está dentro de um poro (Figura 1C), conjuntamente com as sete argininas na posição 113 sobre o mutante de proteína, um anel de sete grupos amino sobre um lado, e um segundo anel de sete grupos arginina estão 15 presentes dentro de uma distância curta de um em relação ao outro na área mais estreita da passagem através do poro. Esta estrutura de anel de amino/arginina possui a propriedade de ligar reversivelmente grupos fosfato imobilizando assim o XMP e dXMP no poro por 5 a 30 ms. Estes eventos de ligação são claramente detectáveis através da mudança resultante na amplitude de corrente.

### 1. Material e Métodos

Mutante de  $\alpha$ -hemolisina (M113R)<sub>7</sub> foi expressado e purificado como previamente descrito (Cheley, S., L.-Q. Gu, et al. (2002), Chem. & Biol. **9**: 829-838).

### Produtos químicos

25 1,2-difitanoil-*sn*-glicero-3-fosfocolina de Avanti Polar Lipids Inc. Pentano foi adquirido de JT Baker, e hexadecano 99+% de Sigma-Aldrich. Heptaquis(6-desoxi-6-amino)- $\beta$ -ciclodextrina.HCl >99% foi adquirido de CYCLOLAB Ltd Budapest, Hungria. Sal de sódio de 2-desoxi-guanosina 5' monofosfato 99% foi adquirido de Acros, sal de dissódio de 2-

desoxi-citosina 5' monofosfato >95%, sal de dissódio de 2-desoxi-timidina 5' monofosfato >97%, e sal de dissódio de 2-desoxi-adenosina 5' monofosfato >95% de Fluka. Sal de dissódio de uridina 5' monofosfato 99%, e ácido citosina 5' monofosfato >98% foram adquiridos de Fluka. Ácido adenosina 5' monofosfato 99%, e sal de dissódio de guanosina 5' monofosfato 97% foram adquiridos de Acros. Base Trizma 99,9% foi adquirida de Sigma-Aldrich, e HCl concentrado de grau reagente analítico de Fisher Scientific. Cloreto de potássio 99%, e cloreto de sódio 99,9% foram adquiridos de Sigma-Aldrich. Brometo de potássio 99,5% e cloreto de césio 99% foram adquiridos de Fluka.

### Equipamento

Um amplificador de *patch clamp* Axopatch 200B de Axon Instruments foi usado com um computador equipado com um conversor Digidata 1200 A/D (Axon instruments). Uma câmara de Teflon foi usada. Dados foram coluídos em pClamp 9.2, e analisados em Clampfit 9.0. Plotagens e gráficos foram obtidos com Microcal Origin 6.0, e integração foi calculada em uma calculadora pessoal.

### Condições Experimentais

Membranas de bicamada lipídica foram formadas de 1,2-difitanoil-*sn*-glicero-3-fosfocolina pelo método de Montal e Mueller (1972), em um orifício de diâmetro de 100-150 µm em um filme de policarbonato de 20 µm (espessura de 20 µm da Goodfellow, Malvern PA ) separando as câmaras *trans* e *cis*. O lado *cis* da câmara foi um terra, e o lado *trans* da câmara foi conectado no estágio cabeça. O potencial refere-se ao valor potencial do eletrodo do lado *trans*. A molécula adaptadora foi adicionada no lado *trans*, as moléculas de mutante de  $\alpha$ -hemolisina de analito foram adicionadas no lado *cis*. Experimentos de dXMP foram realizados a +130mV, experimentos de XMP a +110mV. Todos os experimentos aqui relatados foram obtidos em Tris-HCl 25 mM de pH 8,0 em KCl 1 M. Alíquotas frescas

de solução de nucleotídeo foram usadas diariamente. Experimentos foram realizados na temperatura ambiente de 22,5 +/-2°C a não ser se indicado de outro modo.

## **2. Resultados**

5 2-desoxi-nucleotídeo 5' monofosfatos parcialmente bloqueiam poros homohexaméricos formados por (M113R)<sub>7</sub>/ heptaquis 6 amino β-cyclodextrina

Registros de canal único foram realizados nos poros homohexaméricos formados de  $\alpha$ HL-M113R com um am<sub>7</sub>- $\beta$ CD aplicado no lado *trans* (Figura 2A e 2B). Na ausência de am<sub>7</sub>- $\beta$ CD, o poro permaneceu permanentemente aberto (Figura 2A e 2B, B) com uma corrente unitária (L1) de  $145 \pm 5$  pA (+130 mV) em KCl 1M em tampão Tris-HCl 25 mM de pH 8,0. A adição de am<sub>7</sub>- $\beta$ CD 40  $\mu$ M na câmara *trans* sozinho acarreta eventos de bloqueio reversível em um nível de corrente de  $65 \pm 5$  pA (L2 em Figura 2A, B e Figura 2B, C). Sob adição de dCMP 5  $\mu$ M na câmara *cis*, é observado um terceiro nível de corrente a  $22 \pm 1$  pA (L3 em Figura 2A, C e Figura 2B, D) originário do nível de corrente L2. L3 representa a ligação de dCMP no complexo de (M113R)<sub>7</sub> com am<sub>7</sub>- $\beta$ CD. Adição de dXMP a até 300  $\mu$ M no lado *trans* em vez de no lado *cis* da câmara não acarretou qualquer alteração dos estados de condutância de ligação de ciclodextrina (não mostrado).

Nas condições experimentais descritas acima, eventos de bloqueio de corrente devido à ligação de ciclodextrina foram observados quando  $\beta$ -ciclodextrina não modificada foi adicionada na câmara *trans* (40  $\mu$ M) na presença de um nanoporo individual de  $\alpha$ HL-M113R. Contudo, no não foram observados outros eventos de bloqueio de corrente, quando dXMPs (até 300  $\mu$ M) foram adicionados em qualquer câmara *trans* ou *cis* (não mostrado).

Na ausência de am<sub>7</sub>- $\beta$ CD na câmara *trans*, eventos de bloqueio (<1ms) foram observados quando dGMP ou dTMP em concentrações

mínimas de 300  $\mu\text{M}$  foram adicionados na câmara *cis* (não mostrado). Estes eventos não são observados durante a escala de tempo do experimento com dXMP ou XMP a 5  $\mu\text{M}$ .

Adição de am<sub>7</sub>- $\beta$ CD na câmara *trans* enquanto se media a corrente através de um canal individual de  $\alpha$ -HL de tipo selvagem acarretou eventos de ligação de ciclodextrina, mas outras alterações da corrente não foram observadas quando foi adicionado dXMP em qualquer uma câmara *cis* ou *trans*.

2-desoxi-nucleotídeo 5' monofosfato pode ser identificado da amplitude do bloqueio parcial de poros homo-heptaméricos por (M113R)<sub>7</sub>/ heptaquis 6 amino  $\beta$ -ciclodextrina

O bloqueio parcial do complexo transiente (M113R)<sub>7</sub>/ am<sub>7</sub>- $\beta$ CD diferiu em amplitude dependendo de qual dXMP foi adicionado na câmara *cis* (Figura 3). Adição de dGMP (5  $\mu\text{M}$ ) no lado *cis* exibiu um bloqueio médio para um nível de corrente de 16 pA (Figura 3A). O histograma de amplitude de todos os pontos do traço mostrado em Figura 3A é mostrado à direita do traço juntamente com a estrutura de dGMP. Todos os outros nucleotídeos exibem amplitudes diferentes como mostrado em Figura 3B para dTMP, 3C para dAMP, e 3D para dCMP. Dentre os quatro dXMP, dGMP bloqueia a maior corrente.

As amplitudes de corrente de experimentos independentes exibiram algumas variações que originaram-se de individualidades do nanoporo de proteína. A corrente média, a +130 mV, para um (M113R)<sub>7</sub> é 139 pA, mas alguns canais exibem correntes tão altas quanto 147 pA e tão baixas quanto 131 pA. Portanto, para comparar os traços de corrente os traços de corrente foram normalizados de experimentos diferentes entre corrente 0 e o nível de corrente de (M113R)<sub>7</sub>/am<sub>7</sub>- $\beta$ CD ajustado a 65 pA.

O tempo de residência ( $\tau_{\text{off}}$ ) de cada dXMP foi calculado sobre 500 eventos de cada um de 3 experimentos independentes

(Tabela 1).

Tabela 1. Tempo de residência em ms de dGMP, dTMP, dAMP, e dCMP tomado na média de três medições independentes.

	dGMP	dTMP	dAMP	dCMP
$\tau_{off}$ (ms)	9,8 $\pm$ 0,2	19,8 $\pm$ 0,8	7,1 $\pm$ 0,2	10,5 $\pm$ 0,4

Níveis de corrente de ciclodextrina

5 Em pH 8,0, o mutante (M113R)<sub>7</sub> exibe dois níveis de corrente L1/L1' quando o canal de proteína não está ocupado (Figura 4). O adaptador ciclodextrina pode se ligar na proteína independentemente do nível de corrente L1/L1' no qual a proteína está.

10 Dois níveis de corrente são observados quando se registra o nível de corrente do nanoporo (M113R)<sub>7</sub> (Figura 4). L1 é o nível de corrente principal, como mostrado no inserto da Figura 4. A ligação de am<sub>7</sub>- $\beta$ CD no nanoporo acarreta dois níveis de corrente representados por L2 e L2' (três níveis são observados em pH 7,5, não mostrado). A ligação de am<sub>7</sub>- $\beta$ CD no nanoporo de proteína ocorre independentemente do nível de corrente L1 ou 15 L1'. L2 é o nível de condutância principal observado quando am<sub>7</sub>- $\beta$ CD é ligado em (M113R)<sub>7</sub> e se origina de ambos os níveis de condutância L1 e L1' do nanoporo vazio sem correlação aparente. O nível de corrente L2' observado quando am<sub>7</sub>- $\beta$ CD é ligado em (M113R)<sub>7</sub> é responsável por menos 20 do que 15% da condutância quando o adaptador ciclodextrina é ligado (veja inserto Figura 4).

25 Os eventos de ligação de nucleotídeo algumas vezes variam como um resultado do nível de corrente L2 e L2' que originam. Um deslocamento de 0,5pA foi observado dependendo de qual de L2 ou L2' o evento de ligação de dXMP se origina (não mostrado). Acarreta uma sobreposição aumentada do histograma de evento de ligação de nucleotídeo (Figura 5).

É possível analisar manualmente cada registro com o objetivo de remover cada evento de ligação de analito proveniente do nível de corrente

de ciclodextrina ligada L2' descrito em Figura 3. Figura 5 mostra a diferença entre o histograma de análise de evento individual obtido de um registro de corrente de detecção de dXMP não modificado (Figura 5A), e o mesmo registro onde eventos de ligação de analito provenientes de nível L2' (Figura 5) têm sido removidos. Os dois histogramas exibem os mesmos quatro picos correspondendo um dGMP, dTMP, dAMP, e dCMP. A amplitude dos picos em Figura 5A é maior do que em 5B porque os eventos de ligação de analito provenientes de L2' têm sido removidos, portanto o histograma origina-se de menos eventos. A separação entre cada pico parece ser melhor em 5B do que em 5A. Contudo, remoção destes eventos de registro não deu uma separação completa de cada pico (Figura 5B). Como um resultado, os níveis de corrente de ciclodextrina L2 e L2' mostrados em Figura 4 não foram levados em consideração nos histogramas de análise de evento individual, e na estatística resultante relatada depois.

15 2-desoxi nucleotídeo 5' monofosfatos podem ser identificados da amplitude do bloqueio parcial de poros homo-heptamérico formados pelo complexo transiente (M113R)<sub>7</sub>/ heptaquis 6 amino β-ciclodextrina

O bloqueio parcial de poros (M113R)<sub>7</sub>/am<sub>7</sub>-βCD mostrou-se diferir em amplitude dependendo de qual dXMP foi adicionado na câmara *cis*. 20 As amplitudes diferentes puderam ser resolvidas quando dGMP, dTMP, dAMP, e dCMP foram adicionadas na câmara *cis* simultaneamente (Figura 5).

25 Figura 6A mostra a amplitude de corrente para uma solução mista de todos os 4 nucleotídeos de um experimento individual. Bandas coloridas estão superpostas sobre o traço de corrente registrado com o objetivo de ilustrar a distribuição de amplitude de cada dXMP. Figura 6B mostra um histograma de amplitude de um traço de corrente de 8000 eventos assumido ser a distribuição de amplitude gerada por cada nucleotídeo. O histograma de amplitude está superposto com distribuições gaussianas. O ajuste é obtido do valor de corrente de pico dado por este experimento como o

valor médio de distribuição e o valor de  $\sigma$  que foi obtido do ajuste e da média da distribuição de cada nucleotídeo independentemente (Tabela 2). Do ajuste dos traços de corrente com distribuições gaussianas, as probabilidades de identificação de cada nucleotídeo foram estabelecidas.

5 Métodos estatísticos

Traços de corrente de  $(M113R)_7$  na presença de  $am_7\beta\text{CD}$  sobre o lado *trans* e um dos nucleotídeos de analito sobre o lado *cis* foram digitalmente filtrados a 300Hz (filtro gaussiano de passagem baixa), e um histograma de amplitude de todos os pontos foi construído. Estes histogramas 10 exibem um pico grande correspondendo à amplitude de corrente que é observado quando o  $am_7\beta\text{CD}$  se liga em mutante de  $\alpha$ -hemolisina  $(M113R)_7$  (correspondendo a L2 em Figura 1). Esta amplitude de corrente varia entre canais de proteína dentro de 5% de um experimento para outro. Por esta razão, os histogramas de amplitude de todos os pontos foram normalizados 15 entre corrente 0, e o pico principal de ciclodextrina ajustado a 65 pA. No histograma normalizado, o pico de nucleotídeo foi ajustado para uma distribuição gaussiana. Os valores médio e de sigma ( $\sigma$ ) do mesmo nucleotídeo foram tomados na média de pelo menos 3 experimentos independentes cada um contendo 1000 eventos (valores listados em tabela).

20 **Tabela 2.** Valores médios das distribuições de cada nucleotídeo de três experimentos independentes todos normalizados entre 0 e 65 pA.

	Média
G	16,0, $\sigma = 0,64$
T	17,4, $\sigma = 0,41$
A	18,4, $\sigma = 0,54$
C	20,0, $\sigma = 0,51$

As probabilidades para a leitura de cada base foram determinadas de experimentos com todos os 4 nucleotídeos presentes simultaneamente sobre pelo menos 3000 eventos de ligação de nucleotídeo 25 em cada traço. Os traços foram filtrados (filtro digital gaussiano de passagem baixa a 300Hz) e normalizados entre 0 e o pico de ciclodextrina a 65 pA

como descrito acima para os experimentos de nucleotídeo individual. Os valores de pico de cada nucleotídeo foram tomados na média de 5 experimentos independentes 5 (Tabela 3).

**Tabela 3.** Valores de pico de cada nucleotídeo onde todos os 4 nucleotídeos estão presentes de 5 experimentos independentes. A última coluna exibe o valor médio de cada pico para o qual a sobreposição das distribuições gaussianas está integrada.

	Média
G(pA)	16,2±0,5
T(pA)	17,6±0,6
A(pA)	18,6±0,6
C(pA)	20,2±0,5

Em experimentos onde todos os 4 nucleotídeos estão presentes, as gaussianas de todos os 4 nucleotídeos possuem uma sobreposição. A estatística foi calculada para o sinal de ligação de um nucleotídeo a ser identificado como o próprio ou outro nucleotídeo do nível de sobreposição entre a distribuição gaussiana deste nucleotídeo e aquela de seus vizinhos.

O ponto de interseção entre as duas gaussianas é calculado das respectivas posições de pico (dada pelo experimento com os nucleotídeos mistos), e os valores de  $\sigma$  para cada distribuição (dados pelo ajuste dos experimentos de nucleotídeo individual). A probabilidade de acurácia é dada pela integração da área da gaussiana que está além do valor de intercessão com gaussiana vizinha (Figura 6, Tabela 4). Na primeira coluna de Tabela 4 está o nucleotídeo que interage com o nanoporo, e a primeira linha é o que é lido da correspondente amplitude de corrente.

**Tabela 4.** Mostra a probabilidade do nucleotídeo adicionado (vertical) a ser detectado como ele mesmo ou outro nucleotídeo (horizontal).

	G <sub>lido</sub>	T <sub>lido</sub>	A <sub>lido</sub>	C <sub>lido</sub>
G <sub>adicionado</sub>	0,88	0,12	0	0
T <sub>adicionado</sub>	0,06	0,83	0,11	0
A <sub>adicionado</sub>	0	0,19	0,74	0,07
C <sub>adicionado</sub>	0	0	0,06	0,94

### 3. Conclusão

Os resultados apresentados indicam que a detecção estocástica é uma alternativa promissora para a identificação de nucleotídeos individuais. Também significa que o seqüenciamento por exonuclease pode ser usado 5 como um método de seqüenciamento de DNA simples, barato e rápido no nível de molécula individual. Seqüenciamento por exonuclease também é um método de seqüenciamento de DNA barato porque não requer reagentes caros, tais como fluoroforos.

Os histogramas de todos os pontos são um método de análise 10 suficiente para identificar cada nucleotídeo com uma acurácia variando de 74 a 94% (Tabela 4). Os valores de tempo de residência de XMP e dXMP também são muito similares nas condições para adicionalmente diferenciar cada analito. A estatística obtida dos histogramas de amplitude pode ser adicionalmente melhorada pela compensação dos níveis de corrente de 15 ciclodextrina como mostrado em Figura 4. A diferença de amplitude de corrente entre cada dXMP é cerca de 1 pA. Esta resolução depende de numerosos parâmetros como segue.

#### Dependência de voltagem

Os eventos de ligação são dependentes de voltagem. A 50 mV, 20 são observados muito poucos eventos de ligação, sugerindo que um campo mínimo é requerido para conduzir o dXMP e XMP para o sítio de ligação. A +150 e +200 mV as amplitudes não mais permitem diferenciar os nucleotídeos. +130 mV mostrou ser a melhor voltagem para desoxirribonucleotídeos 5' monofosfato, e +110 mV deu a melhor resolução 25 para os ribonucleotídeos 5' monofosfato.

#### Concentração de sal

Tampão Tris-HCl de pH 8, KCl 0,5, 1, e 2M KCl foram testados. Dos histogramas de amplitude de todos os pontos, a melhor resolução entre os picos foi obtida em KCl a 1M.

### Dependência de pH

As amplitudes de corrente são dependentes de pH, tampão Tris-HCl, KCl 1M em pH 7,5, 8,0, 8,2, 8,5, 9,0, e 9,5 foram testados. Em pH 8,0 e acima foram observados dois níveis de corrente sob ligação de am<sub>7</sub>-βCD (Figura 4). Em pH 7,5 a heptaquis-6 amino-β-ciclodextrina exibe um terceiro nível de corrente (não mostrado). Faz dTMP exibir dois tipos de eventos com amplitudes diferentes, um dos quais está dentro da faixa de eventos de dGMP, levando assim a uma perda de resolução. Em pH 9,5, os eventos de ligação de nucleotídeo não são mais observados. A melhor separação e pico é obtida em pH 8,0.

### Dependência de sal

A resolução entre dXMPs e XMPs é melhor com KCl do que com NaCl ou CsCl. KCl 1M dá resolução melhor do que KCl 2M. O uso de KBr não permitiu a identificação de nucleotídeos diferentes porque cada evento de ligação acarretou um bloqueio completo do complexo transiente (M113R)<sub>7</sub>/am<sub>7</sub>-βCD.

### Dependência da temperatura

Abaixamento da temperatura para 14°C ou aumento dela para 50°C não interferiu com a detecção de dXMP/XMP. Contudo, não melhorou a resolução dos histogramas de amplitude.

### Outros mutantes de α-hemolisina

Foi visto que (M113N)<sub>7</sub> se liga em am<sub>7</sub>-βCD mas não foi observada detecção de nucleotídeo. (M113F)<sub>7</sub> e (M113F/147K)<sub>7</sub> não deram detecção de se am<sub>7</sub>-βCD foi adicionado ou não. (M113K)<sub>7</sub> foi testado nas mesmas condições. Neste caso, o registro é muito similar àquele da Figura 1. A ligação de nucleotídeo foi detectada com mutante (M113K)<sub>7</sub> e am<sub>7</sub>-βCD mas a separação de picos entre os nucleotídeos foi menor do que quando (M113R)<sub>7</sub> foi usado.

### Ribonucleotídeos 5' monofosfato

XMP foram bem sucedidamente identificados com este método, a resolução entre cada base foi inferior àquela de dXMP com separações de picos menores do que 1 pA para U, A, e C. As amplitudes de corrente de histograma de todos os pontos parecem da mesma ordem que aquelas de dXMPs, com GMP exibindo a corrente mais baixa (bloqueio mais alto), seguido por UMP, AMP, e CMP com a amplitude de corrente mais elevada (bloqueio mais baixo). A voltagem ótima para identificação de XMP verificada foi de +110mV em KCl 1M, pH 8,0.

#### Mecanismo

O complexo transiente (M113R)<sub>7</sub>/am<sub>7</sub>- $\beta$ CD também tem mostrado que se liga e diferencia glicose fosfatos (glicose 1P e galactose 1P). Sugere interação entre o anel de arginina sobre o lado, o grupo fosfato e o anel de amina de am<sub>7</sub>- $\beta$ CD sobre o outro lado.  $\beta$ -ciclodextrina não modificada não deu qualquer detecção. Pouca diferença é observada entre XMP e dXMP sugerindo que os grupos hidroxila não desempenham um papel grande na ligação.

## LISTAGEM DE SEQÜÊNCIAS

<110> ISIS INNOVATION LIMITED

120 MÉTODOS USANDO POROS

<130> N.95303A SER

<150> GB 0523282.2

<151> 2005-11-15

<160> 9

<170> PatentIn version 3.2

<210> 1

<211> 882

<212> DNA

<212> DNA

52203

<221> GDS

<221> CDS

<400> 1

gca gat tct gat att aat att aaa acc ggt act aca gat att gga agc  
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser  
1 5 10 15

48

```

aat act aca gta aaa aca ggt gat tta gtc act tat gat aaa gaa aat
Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
          20          25          30

```

96

ggc atg cac aaa aaa gta ttt tat agt ttt atc gat gat aaa aat cac  
Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
35 40 45

144

```

aat aaa aaa ctg cta gtt att aga acg aaa ggt acc att gct ggt caa
Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
      50           55           60

```

192

```

tat aga gtt tat agc gaa gaa ggt gct aac aaa agt ggt tta gcc tgg
Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65           70           75           80

```

240

```

cct tca gcc ttt aag gta cag ttg caa cta cct gat aat gaa gta gct
Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
85          90          95

```

288

```

caa ata tct gat tac tat cca aga aat tcg att gat aca aaa gag tat
Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
          100          105          110

```

336

atg agt act tta act tat gga ttc aac ggt aat gtt act ggt gat gat  
Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp  
115 120 125

384

```

aca gga aaa att ggc ggc ctt att ggt gca aat gtt tcg att ggt cat
Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His
130          135          140

```

aca ctg aaa tat gtt caa cct gat ttc aaa aca att tta gag agc cca	480
Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro	
145 150 155 160	
act gat aaa aaa gta ggc tgg aaa gtg ata ttt aac aat atg gtg aat	528
Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn	
165 170 175	
caa aat tgg gga cca tat gat aga gat tct tgg aac ccg gta tat ggc	576
Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly	
180 185 190	
aat caa ctt ttc atg aaa act aga aat ggt tct atg aaa gca gca gat	624
Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp	
195 200 205	
aac ttc ctt gat cct aac aaa gca agt tct cta tta tct tca ggg ttt	672
Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe	
210 215 220	
tca cca gac ttc gct aca gtt att act atg gat aga aaa gca tcc aaa	720
Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys	
225 230 235 240	
caa caa aca aat ata gat gta ata tac gaa cga gtt cgt gat gat tac	768
Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr	
245 250 255	
caa ttg cat tgg act tca aca aat tgg aaa ggt acc aat act aaa gat	816
Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp	
260 265 270	
aaa tgg aca gat cgt tct tca gaa aga tat aaa atc gat tgg gaa aaa	864
Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys	
275 280 285	
gaa gaa atg aca aat taa	882
Glu Glu Met Thr Asn	
290	
<210> 2	
<211> 293	
<212> PRT	
<213> Staphylococcus aureus	
<400> 2	
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser	
1 5 10 15	
Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn	
20 25 30	
Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His	
35 40 45	
Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln	
50 55 60	

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
100 105 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp  
115 120 125

Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His  
130 135 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp  
260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
275 280 285

Glu Glu Met Thr Asn  
290

```
<211> 1300
<212> DNA
<213> Seqüênci a artificial

<220>
<223> mutante de alfa hemolisina M113H

<220>
<221> CDS
<222> (39)..(920)

<220>
<221> característica_misc
<222> (53)..(53)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (56)..(56)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (58)..(59)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (926)..(926)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (991)..(991)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (994)..(994)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (998)..(998)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1001)..(1001)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1003)..(1003)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1010)..(1010)
<223> n is a, c, g, or t

<220>
```

```
<221>  caracteristica_misc
<222>  (1014)..(1014)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1031)..(1031)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1041)..(1041)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1052)..(1052)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1067)..(1067)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1083)..(1083)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1085)..(1085)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1087)..(1087)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1097)..(1098)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1100)..(1100)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1103)..(1104)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1110)..(1110)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1117)..(1117)
<223>  n is a, c, g, or t
```

```
<220>
<221> característica_misc
<222> (1128)..(1128)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1131)..(1131)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1136)..(1136)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1140)..(1140)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1143)..(1143)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1145)..(1145)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1148)..(1148)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1151)..(1151)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1155)..(1155)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1160)..(1160)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1164)..(1165)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1167)..(1167)
<223> n is a, c, g, or t

<220>
<221> característica_misc
```

```
<222> (1169)..(1169)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1171)..(1171)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1175)..(1176)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1180)..(1180)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1182)..(1184)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1188)..(1188)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1191)..(1191)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1193)..(1193)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1196)..(1196)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1201)..(1201)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1206)..(1206)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1208)..(1208)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1210)..(1210)
<223> n is a, c, g, or t
```

```
<220>
<221>  característica_misc
<222>  (1214)..(1214)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1217)..(1221)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1223)..(1223)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1229)..(1229)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1231)..(1231)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1238)..(1238)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1241)..(1243)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1255)..(1256)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1258)..(1260)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1262)..(1264)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1267)..(1267)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1270)..(1270)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1273)..(1273)
```

```

<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1282)..(1283)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1288)..(1288)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1291)..(1293)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1295)..(1295)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1300)..(1300)
<223> n is a, c, g, or t

<400> 3
gttctgttta actttaagaa gggagatata catatgag cag att ctg ata ttn acn      56
      Gln Ile Leu Ile Xaa Thr
      1                      5

tnn gcg acc ggt act aca gat att gga agc aat act aca gta aaa aca      104
Xaa Ala Thr Gly Thr Asp Ile Gly Ser Asn Thr Thr Val Lys Thr
      10                     15                      20

gg t gat tta gtc act tat gat aaa gaa aat ggc atg cac aaa aaa gta      152
Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn Gly Met His Lys Lys Val
      25                     30                      35

ttt tat agt ttt atc gat gat aaa aat cac aat aaa aaa ctg cta gtt      200
Phe Tyr Ser Phe Ile Asp Asp Lys Asn His Asn Lys Lys Leu Leu Val
      40                     45                      50

att aga aca aaa ggt acc att gct ggt caa tat aga gtt tat agc gaa      248
Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu
      55                     60                      65                      70

gaa ggt gct aac aaa agt ggt tta gcc tgg cct tca gcc ttt aag gta      296
Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp Pro Ser Ala Phe Lys Val
      75                     80                      85

cag ttg caa cta cct gat aat gaa gta gct caa ata tct gat tac tat      344
Gln Leu Gln Leu Pro Asp Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr
      90                     95                      100

ccg cgg aat tcg att gat aca aaa gag tat cac agt acg tta acg tac      392
Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr His Ser Thr Leu Thr Tyr
      105                    110                     115

gga ttc aac ggt aac ctt act ggt gat gat act agt aaa att gga ggc      440
Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp Thr Ser Lys Ile Gly Gly
      120                    125                     130

```

ctt att ggg gcc cag gtt tcc cta ggt cat aca ctt aag tat gtt caa Leu Ile Gly Ala Gln Val Ser Leu Gly His Thr Leu Lys Tyr Val Gln 135 140 145 150	488
cct gat ttc aaa aca att ctc gag agc cca act gat aaa aaa gta ggc Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly 155 160 165	536
tgg aaa gtg ata ttt aac aat atg gtg aat caa aat tgg gga cca tac Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr 170 175 180	584
gat cga gat tct tgg aac ccg gta tat ggc aat caa ctt ttc atg aag Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys 185 190 195	632
act aga aat ggt tct atg aaa gca gca gat aac ttc ctt gat cct aac Thr Arg Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn 200 205 210	680
aaa gca agt tcc cta tta tct tca ggg ttt tca cca gac ttc gct aca Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr 215 220 225 230	728
gtt att act atg gat aga aaa gca tcc aaa caa caa aca aat ata gat Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp 235 240 245	776
gta ata tac gaa cga gtt cgt gat gat tac caa ttg cat tgg act tca Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser 250 255 260	824
cca aat tgg aaa ggt acc aat act aaa gat aaa tgg aca gat cgt tct Pro Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Thr Asp Arg Ser 265 270 275	872
tca gaa aga tat aaa atc gat tgg gaa aaa gaa gaa atg aca aat taa Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn 280 285 290	920
tgtaanttat ttgtacatgt acaaataaat ataatttata actttagccg aagctggatc cggtcgctac naanccnaa ngnagctgan ttgnctgctg ccccccgtac natactagca nacccttgg gnccctaacg ggtctgnggg gttttgctg aangngnact tttccgnan tcnncccggn ccccccnnggt gaaatccnaa nccccnaacn ggngntgnta ncaantttan tggnnncntna nttnnaaan cnnttaantt ngnaancccc nttnncnan ggcnaannnn nancctttna naaaaaancc nnnggggggg tttcnntnnn annncnttn aangggcccc cnnggggnaa nnntnggggn	980 1040 1100 1160 1220 1280 1300

<210> 4  
 <211> 293  
 <212> PRT  
 <213> Seqüênci a artificial  
  
 <220>  
 <221> caracteristica\_mis

<222> (5)..(5)  
 <223> O 'Xaa' na localizaçao 5 representa Leu, ou Phe.  
  
 <220>  
 <221> caracteristica\_misc  
 <222> (7)..(7)  
 <223> O 'Xaa' na localizaçao 7 representa Tyr, Trp, Cys, Ser, Leu, ou Phe.

<220>  
 <223> mutante de alfa hemolisina M113H

<400> 4

Gln Ile Leu Ile Xaa Thr Xaa Ala Thr Gly Thr Thr Asp Ile Gly Ser  
 1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
 20 25 30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
 35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
 50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
 65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
 85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
 100 105 110

His Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp  
 115 120 125

Thr Ser Lys Ile Gly Gly Leu Ile Gly Ala Gln Val Ser Leu Gly His  
 130 135 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
 145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
 165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
 180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp

195

200

205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
245 250 255

Gln Leu His Trp Thr Ser Pro Asn Trp Lys Gly Thr Asn Thr Lys Asp  
260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
275 280 285

Glu Glu Met Thr Asn  
290

<210> 5  
<211> 1300  
<212> DNA  
<213> Seqüênci a artificial

<220>  
<223> M113K alfa hemolisina mutante

<220>  
<221> CDS  
<222> (39)..(920)

<220>  
<221> característica\_misc  
<222> (53)..(53)  
<223> n is a, c, g, or t

<220>  
<221> característica\_misc  
<222> (56)..(56)  
<223> n is a, c, g, or t

<220>  
<221> característica\_misc  
<222> (58)..(59)  
<223> n is a, c, g, or t

<220>  
<221> característica\_misc  
<222> (926)..(926)  
<223> n is a, c, g, or t

<220>  
<221> característica\_misc  
<222> (991)..(991)

```
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (994)..(994)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (998)..(998)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1001)..(1001)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1003)..(1003)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1010)..(1010)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1014)..(1014)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1031)..(1031)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1041)..(1041)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1052)..(1052)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1067)..(1067)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1083)..(1083)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1085)..(1085)
<223>  n is a, c, g, or t
```

```
<221>  característica_misc
<222>  (1087)..(1087)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1097)..(1098)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1100)..(1100)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1103)..(1104)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1110)..(1110)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1117)..(1117)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1128)..(1128)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1131)..(1131)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1136)..(1136)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1140)..(1140)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1143)..(1143)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1145)..(1145)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1148)..(1148)
<223>  n is a, c, g, or t
```

```
<220>
<221>  característica_misc
<222>  (1151)..(1151)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1155)..(1155)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1160)..(1160)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1164)..(1165)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1167)..(1167)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1169)..(1169)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1171)..(1171)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1175)..(1176)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1180)..(1180)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1182)..(1184)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1188)..(1188)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1191)..(1191)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
```

```
<222> (1193)..(1193)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1196)..(1196)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1201)..(1201)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1206)..(1206)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1208)..(1208)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1210)..(1210)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1214)..(1214)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1217)..(1221)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1223)..(1223)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1229)..(1229)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1231)..(1231)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1238)..(1238)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1241)..(1243)
<223> n is a, c, g, or t
```

```

<220>
<221>  característica_misc
<222>  (1255)..(1256)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1258)..(1260)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1262)..(1264)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1267)..(1267)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1270)..(1270)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1273)..(1273)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1282)..(1283)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1288)..(1288)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1291)..(1293)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1295)..(1295)
<223>  n is a, c, g, or t

<220>
<221>  característica_misc
<222>  (1300)..(1300)
<223>  n is a, c, g, or t

<400> 5
gttctgttta actttaagaa gggagatata catatgag cag att ctg ata ttn acn
Gln Ile Leu Ile Xaa Thr
1 5

tnn gcg acc ggt act aca gat att gga agc aat act aca gta aaa aca
Xaa Ala Thr Gly Thr Asp Ile Gly Ser Asn Thr Thr Val Lys Thr
10 20

104

```

ggt gat tta gtc act tat gat aaa gaa aat ggc atg cac aaa aaa gta	152
Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn Gly Met His Lys Lys Val	
25 30 35	
ttt tat agt ttt atc gat gat aaa aat cac aat aaa aaa ctg cta gtt	200
Phe Tyr Ser Phe Ile Asp Asp Lys Asn His Asn Lys Lys Leu Leu Val	
40 45 50	
att aga aca aaa ggt acc att gct ggt caa tat aga gtt tat agc gaa	248
Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu	
55 60 65 70	
gaa ggt gct aac aaa agt ggt tta gcc tgg cct tca gcc ttt aag gta	296
Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp Pro Ser Ala Phe Lys Val	
75 80 85	
cag ttg caa cta cct gat aat gaa gta gct caa ata tct gat tac tat	344
Gln Leu Gln Leu Pro Asp Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr	
90 95 100	
ccg cgg aat tcg att gat aca aaa gag tat aaa agt acg tta acg tac	392
Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr Lys Ser Thr Leu Thr Tyr	
105 110 115	
gga ttc aac ggt aac ctt act ggt gat gat act agt aaa att gga ggc	440
Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp Thr Ser Lys Ile Gly Gly	
120 125 130	
ctt att ggg gcc cag gtt tcc cta ggt cat aca ctt aag tat gtt caa	488
Leu Ile Gly Ala Gln Val Ser Leu Gly His Thr Leu Lys Tyr Val Gln	
135 140 145 150	
cct gat ttc aaa aca att ctc gag agc cca act gat aaa aaa gta ggc	536
Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly	
155 160 165	
tgg aaa gtg ata ttt aac aat atg gtg aat caa aat tgg gga cca tac	584
Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr	
170 175 180	
gat cga gat tct tgg aac ccg gta tat ggc aat caa ctt ttc atg aag	632
Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys	
185 190 195	
act aga aat ggt tct atg aaa gca gca gat aac ttc ctt gat cct aac	680
Thr Arg Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn	
200 205 210	
aaa gca agt tcc cta tta tct tca ggg ttt tca cca gac ttc gct aca	728
Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr	
215 220 225 230	
gtt att act atg gat aga aaa gca tcc aaa caa caa aca aat ata gat	776
Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp	
235 240 245	
gta ata tac gaa cga gtt cgt gat gat tac caa ttg cat tgg act tca	824
Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser	
250 255 260	
cca aat tgg aaa ggt acc aat act aaa gat aaa tgg aca gat cgt tct	872
Pro Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Thr Asp Arg Ser	

265

270

275

tca gaa aga tat aaa atc gat tgg gaa aaa gaa gaa atg aca aat taa	920																
Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Glu Met Thr Asn																	
280	285	290		tgtaanttat ttgtacatgt acaaataaat ataattata acttagccg aagctggatc	980	cggctgctac naanccnaa ngnagctgan ttgnctgctg ccccccgtac nataactagca	1040	nacccttgg gnccctaacg ggtctgnggg gttttgctg aangngnact tttccgnnan	1100	tcnnccggn ccccccnggt gaaatccnaa ncccnnaacn ggnngntgnta ncaantttan	1160	tggnnncntna ntttnnaaan cnnttaantt ngnaancccc nttnncnan ggcnaannnn	1220	nancctttaa naaaaaancc nnnggggggg tttcnntnnn annncnttn aangggcccc	1280	cnnggggnnaa nnntnggggn	1300
290																	
tgtaanttat ttgtacatgt acaaataaat ataattata acttagccg aagctggatc	980																
cggctgctac naanccnaa ngnagctgan ttgnctgctg ccccccgtac nataactagca	1040																
nacccttgg gnccctaacg ggtctgnggg gttttgctg aangngnact tttccgnnan	1100																
tcnnccggn ccccccnggt gaaatccnaa ncccnnaacn ggnngntgnta ncaantttan	1160																
tggnnncntna ntttnnaaan cnnttaantt ngnaancccc nttnncnan ggcnaannnn	1220																
nancctttaa naaaaaancc nnnggggggg tttcnntnnn annncnttn aangggcccc	1280																
cnnggggnnaa nnntnggggn	1300																

&lt;210&gt; 6

&lt;211&gt; 293

&lt;212&gt; PRT

&lt;213&gt; Seqüência artificial

&lt;220&gt;

&lt;221&gt; característica\_misc

&lt;222&gt; (5)..(5)

&lt;223&gt; O 'Xaa' na localização 5 representa Leu, or Phe.

&lt;220&gt;

&lt;221&gt; característica\_misc

&lt;222&gt; (7)..(7)

&lt;223&gt; O 'Xaa' na localização 7 representa Tyr, Trp, Cys, Ser, Leu, or Phe.

&lt;220&gt;

&lt;223&gt; M113K alfa hemolisina mutante

&lt;400&gt; 6

Gln Ile Leu Ile Xaa Thr Xaa Ala Thr Gly Thr Thr Asp Ile Gly Ser					
1	5		10		15
	10		15		
	15				

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn			
20	25		30
	30		

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His			
35	40		45
	45		

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln			
50	55		60
	60		

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp					
65	70		75		80
	75		80		
	80				

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala	
---	--

85

90

95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
 100 105 110

Lys Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp  
 115 120 125

Thr Ser Lys Ile Gly Gly Leu Ile Gly Ala Gln Val Ser Leu Gly His  
 130 135 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
 145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
 165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
 180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
 195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
 210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
 225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
 245 250 255

Gln Leu His Trp Thr Ser Pro Asn Trp Lys Gly Thr Asn Thr Lys Asp  
 260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
 275 280 285

Glu Glu Met Thr Asn  
 290

<210> 7  
 <211> 1300  
 <212> DNA  
 <213> Seqüência artificial

<220>  
 <223> M113R alfa hemolisina mutante

```
<220>
<221> CDS
<222> (39)..(920)

<220>
<221> característica_misc
<222> (53)..(53)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (56)..(56)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (58)..(59)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (926)..(926)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (991)..(991)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (994)..(994)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (998)..(998)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1001)..(1001)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1003)..(1003)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1010)..(1010)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1014)..(1014)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1031)..(1031)
```

```
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1041)..(1041)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1052)..(1052)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1067)..(1067)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1083)..(1083)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1085)..(1085)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1087)..(1087)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1097)..(1098)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1100)..(1100)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1103)..(1104)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1110)..(1110)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1117)..(1117)
<223> n is a, c, g, or t

<220>
<221> caracteristica_misc
<222> (1128)..(1128)
<223> n is a, c, g, or t

<220>
```

```
<221> característica_misc
<222> (1131)..(1131)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1136)..(1136)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1140)..(1140)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1143)..(1143)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1145)..(1145)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1148)..(1148)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1151)..(1151)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1155)..(1155)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1160)..(1160)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1164)..(1165)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1167)..(1167)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1169)..(1169)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1171)..(1171)
<223> n is a, c, g, or t
```

```
<220>
<221> característica_misc
<222> (1175)..(1176)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1180)..(1180)
<223> n is a, c, g, or t

.
.
.

<220>
<221> característica_misc
<222> (1182)..(1184)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1188)..(1188)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1191)..(1191)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1193)..(1193)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1196)..(1196)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1201)..(1201)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1206)..(1206)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1208)..(1208)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1210)..(1210)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1214)..(1214)
<223> n is a, c, g, or t

<220>
<221> característica_misc
```

```
<222> (1217)..(1221)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1223)..(1223)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1229)..(1229)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1231)..(1231)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1238)..(1238)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1241)..(1243)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1255)..(1256)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1258)..(1260)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1262)..(1264)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1267)..(1267)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1270)..(1270)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1273)..(1273)
<223> n is a, c, g, or t

<220>
<221> característica_misc
<222> (1282)..(1283)
<223> n is a, c, g, or t
```

```

<220>
<221>  caracteristica_misc
<222>  (1288)..(1288)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1291)..(1293)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1295)..(1295)
<223>  n is a, c, g, or t

<220>
<221>  caracteristica_misc
<222>  (1300)..(1300)
<223>  n is a, c, g, or t

<400>  7
gttctgtta actttaagaa gggagatata catatgag cag att ctg ata ttu acn      56
          Gln Ile Leu Ile Xaa Thr
          1           5

tnn gcg acc ggt act aca gat att gga agc aat act aca gta aaa aca      104
Xaa Ala Thr Gly Thr Asp Ile Gly Ser Asn Thr Thr Val Lys Thr
          10          15          20

ggt gat tta gtc act tat gat aaa gaa aat ggc atg cac aaa aaa gta      152
Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn Gly Met His Lys Lys Val
          25          30          35

ttt tat agt ttt atc gat gat aaa aat cac aat aaa aaa ctg cta gtt      200
Phe Tyr Ser Phe Ile Asp Asp Lys Asn His Asn Lys Lys Leu Leu Val
          40          45          50

att aga aca aaa ggt acc att gct ggt caa tat aga gtt tat agc gaa      248
Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu
          55          60          65          70

gaa ggt gct aac aaa agt ggt tta gcc tgg cct tca gcc ttt aag gta      296
Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp Pro Ser Ala Phe Lys Val
          75          80          85

cag ttg caa cta cct gat aat gaa gta gct caa ata tct gat tac tat      344
Gln Leu Gln Leu Pro Asp Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr
          90          95          100

ccg cgg aat tcg att gat aca aaa gag tat aga agt acg tta acg tac      392
Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr Arg Ser Thr Leu Thr Tyr
          105         110         115

gga ttc aac ggt aac ctt act ggt gat gat act agt aaa att gga ggc      440
Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp Thr Ser Lys Ile Gly Gly
          120         125         130

ctt att ggg gcc cag gtt tcc cta ggt cat aca ctt aag tat gtt caa      488
Leu Ile Gly Ala Gln Val Ser Leu Gly His Thr Leu Lys Tyr Val Gln
          135         140         145         150

cct gat ttc aaa aca att ctc gag agc cca act gat aaa aaa gta ggc      536
Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro Thr Asp Lys Lys Val Gly

```

155	160	165	
tgg aaa gtg ata ttt aac aat atg gtg aat caa aat tgg gga cca tac Trp Lys Val Ile Phe Asn Asn Met Val Asn Gln Asn Trp Gly Pro Tyr 170	175	180	584
gat cga gat tct tgg aac ccg gta tat ggc aat caa ctt ttc atg aag Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly Asn Gln Leu Phe Met Lys 185	190	195	632
act aga aat ggt tct atg aaa gca gca gat aac ttc ctt gat cct aac Thr Arg Asn Gly Ser Met Lys Ala Ala Asp Asn Phe Leu Asp Pro Asn 200	205	210	680
aaa gca agt tcc cta tta tct tca ggg ttt tca cca gac ttc gct aca Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe Ser Pro Asp Phe Ala Thr 215	220	225	728
gtt att act atg gat aga aaa gca tcc aaa caa caa aca aat ata gat Val Ile Thr Met Asp Arg Lys Ala Ser Lys Gln Gln Thr Asn Ile Asp 235	240	245	776
gta ata tac gaa cga gtt cgt gat gat tac caa ttg cat tgg act tca Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr Gln Leu His Trp Thr Ser 250	255	260	824
cca aat tgg aaa ggt acc aat act aaa gat aaa tgg aca gat cgt tct Pro Asn Trp Lys Gly Thr Asn Thr Lys Asp Lys Trp Thr Asp Arg Ser 265	270	275	872
tca gaa aga tat aaa atc gat tgg gaa aaa gaa gaa atg aca aat taa Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys Glu Met Thr Asn 280	285	290	920
tgtaanttat ttgtacatgt acaaataaat ataatttata acttagccg aagctggatc cggtcgctac naanccnaa ngnagctgan ttgnctgctg ccccccgtac natactagca nacccttgg gnccctaacg ggtctgnggg gttttgctg aangngnact tttccgnnan tcnncccggn ccccccnggt gaaatccnaa ncccnnaacn ggngntgnta ncaantttan tggnnncntna nttnnaaan cnnttaantt ngnaancccc nttnncnan ggcnaannnn nancctttna naaaaaancc nnnggggggg tttcnntnnn annncnttn aangggcccc cnnggggnnaa nnntnggggn			980 1040 1100 1160 1220 1280 1300

<210> 8  
 <211> 293  
 <212> PRT  
 <213> Seqüênci a artificial

<220>  
 <221> característica\_mis  
 <222> (5)..(5)  
 <223> O 'Xaa' na localizaçao 5 representa Leu, or Phe.  
  
 <220>  
 <221> característica\_mis  
 <222> (7)..(7)  
 <223> O 'Xaa' na localizaçao 7 representa Tyr, Trp, Cys, Ser, Leu, or

Phe.

<220>

<223> M113R alfa hemolisina mutante

<400> 8

Gln Ile Leu Ile Xaa Thr Xaa Ala Thr Gly Thr Thr Asp Ile Gly Ser  
1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
20 25 30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
35 40 45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr  
100 105 110

Arg Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp  
115 120 125

Thr Ser Lys Ile Gly Gly Leu Ile Gly Ala Gln Val Ser Leu Gly His  
130 135 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
180 185 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
195 200 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
 225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
 245 250 255

Gln Leu His Trp Thr Ser Pro Asn Trp Lys Gly Thr Asn Thr Lys Asp  
 260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys  
 275 280 285

Glu Glu Met Thr Asn  
 290

<210> 9  
 <211> 228  
 <212> PRT  
 <213> Seqüênci a artif icial

<220>  
 <223> Lambda exonuclease

<400> 9

Ser His Met Thr Pro Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp Val  
 1 5 10 15

Arg Ala Val Glu Gln Gly Asp Asp Ala Trp His Lys Leu Arg Leu Gly  
 20 25 30

Val Ile Thr Ala Ser Glu Val His Asn Val Ile Ala Lys Pro Arg Ser  
 35 40 45

Gly Lys Lys Trp Pro Asp Met Lys Met Ser Tyr Phe His Thr Leu Leu  
 50 55 60

Ala Glu Val Cys Thr Gly Val Ala Pro Glu Val Asn Ala Lys Ala Leu  
 65 70 75 80

Ala Trp Gly Lys Gln Tyr Glu Asn Asp Ala Arg Thr Leu Phe Glu Phe  
 85 90 95

Thr Ser Gly Val Asn Val Thr Glu Ser Pro Ile Ile Tyr Arg Asp Glu  
 100 105 110

Ser Met Arg Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser Asp Gly Asn  
 115 120 125

Gly Leu Glu Leu Lys Cys Pro Phe Thr Ser Arg Asp Phe Met Lys Phe

130

135

140

Arg Leu Gly Gly Phe Glu Ala Ile Lys Ser Ala Tyr Met Ala Gln Val  
145 150 155 160

Gln Tyr Ser Met Trp Val Thr Arg Lys Asn Ala Trp Tyr Phe Ala Asn  
165 170 175

Tyr Asp Pro Arg Met Lys Arg Glu Gly Leu His Tyr Val Val Ile Glu  
180 185 190

Arg Asp Glu Lys Tyr Met Ala Ser Phe Asp Glu Ile Val Pro Glu Phe  
195 200 205

Ile Glu Lys Met Asp Glu Ala Leu Ala Glu Ile Gly Phe Val Phe Gly  
210 215 220

Glu Gln Trp Arg  
225

## REIVINDICAÇÕES

1. Método para identificar um nucleotídeo individual, caracterizado pelo fato de compreender:

5 (a) contatar o nucleotídeo com um poro de proteína de transmembrana de modo que o nucleotídeo interaja com o poro; e

(b) medir a corrente passando através do poro durante a interação e deste modo determinar a identidade do nucleotídeo.

10 2. Método de acordo com a reivindicação 1, caracterizado pelo fato de que a interação envolve o nucleotídeo reversivelmente ligando no canal do poro.

15 3. Método de acordo com a reivindicação 1 ou 2, caracterizado pelo fato de que o poro é (a)  $\alpha$ -hemolisina formada por sete subunidades idênticas como mostrado em SEQ ID NO: 2; ou (b) uma variante da mesma em que uma ou mais das sete subunidades possui pelo menos 50% de homologia à SEQ ID NO: 2 baseada na identidade de aminoácido sobre a seqüência inteira e retém atividade de poro.

4. Método de acordo com a reivindicação 3, caracterizado pelo fato de que uma ou mais das sete subunidades é (M113R)<sub>7</sub>.

20 5. Método de acordo com qualquer uma das reivindicações 1 a 4, caracterizado pelo fato de que o poro compreende um adaptador molecular que facilita a interação entre o nucleotídeo e o poro.

6. Método de acordo com a reivindicação 5, caracterizado pelo fato de que o adaptador molecular é uma heptaquis-6-amino- $\beta$ -ciclodextrina (am<sub>7</sub>- $\beta$ -CD).

25 7. Método de acordo com qualquer uma das reivindicações 1 a 6, caracterizado pelo fato de que o nucleotídeo individual é um monofosfato, difosfato ou trifosfato.

8. Método de acordo com qualquer uma das reivindicações 1 a 7, caracterizado pelo fato de que o nucleotídeo individual é um

ribonucleotídeo ou desoxirribonucleotídeo.

9. Método de acordo com a reivindicação 8, caracterizado pelo fato de adicionalmente compreender antes da etapa (a) a etapa de digerir uma seqüência de ácido ribonucleico (RNA) para proporcionar o nucleotídeo individual.

10. Método de acordo com a reivindicação 8, caracterizado pelo fato de adicionalmente compreender antes da etapa (a) a etapa de digerir uma seqüência de ácido desoxirribonucleico (DNA) para proporcionar o nucleotídeo individual.

11. Método de acordo com a reivindicação 9 ou 10, caracterizado pelo fato de que mais do que um dos nucleotídeos individuais da seqüência de RNA ou DNA são contatados com o poro em uma maneira seqüencial de tal modo que a identidade de toda ou de parte da seqüência pode ser determinada.

12. Método para seqüenciar uma seqüência de ácido nucleico alvo, caracterizado pelo fato de compreender:

(a) digerir um nucleotídeo individual de uma extremidade da seqüência alvo usando uma exonuclease processiva;

(b) contatar o nucleotídeo com um poro de proteína de transmembrana de modo que o nucleotídeo interaja com o poro;

(c) medir a corrente passando através do poro durante a interação e deste modo determinar a identidade do nucleotídeo; e

(d) repetir as etapas (a) a (c) na mesma extremidade da seqüência de ácido nucleico e determinar deste modo a seqüência de ácido nucleico.

13. Kit para seqüenciar um ácido nucleico, caracterizado pelo fato de compreender:

- uma ciclodextrina; e
- uma exonuclease processiva.

14. Kit de acordo com a reivindicação 13, caracterizado pelo fato de adicionalmente compreender um poro de proteína de transmembrana.

Fig. 1

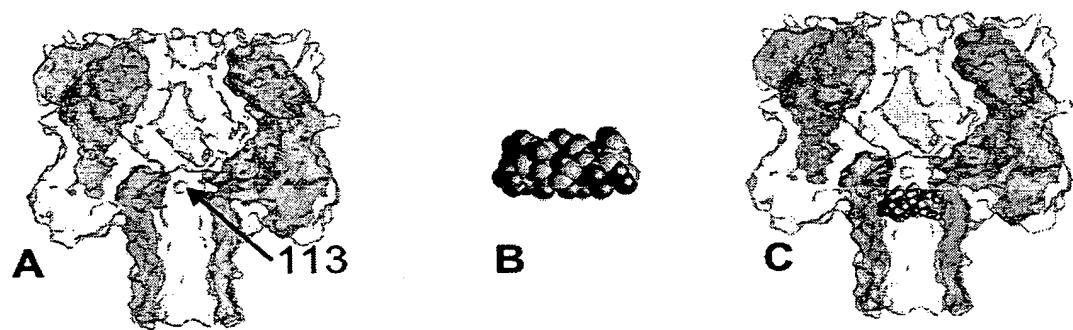


Fig.2A.

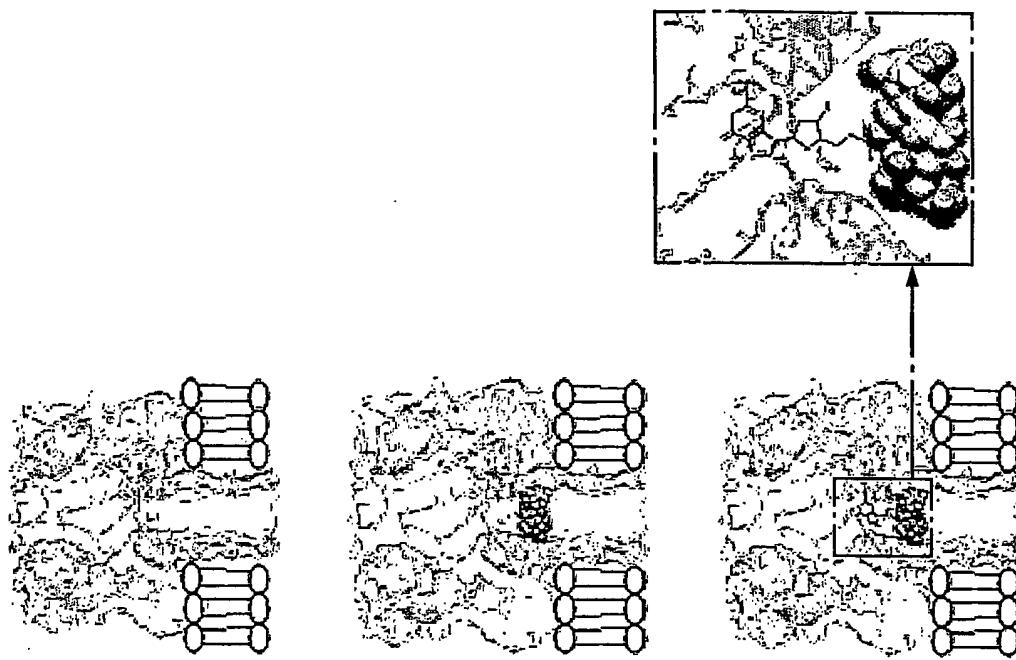
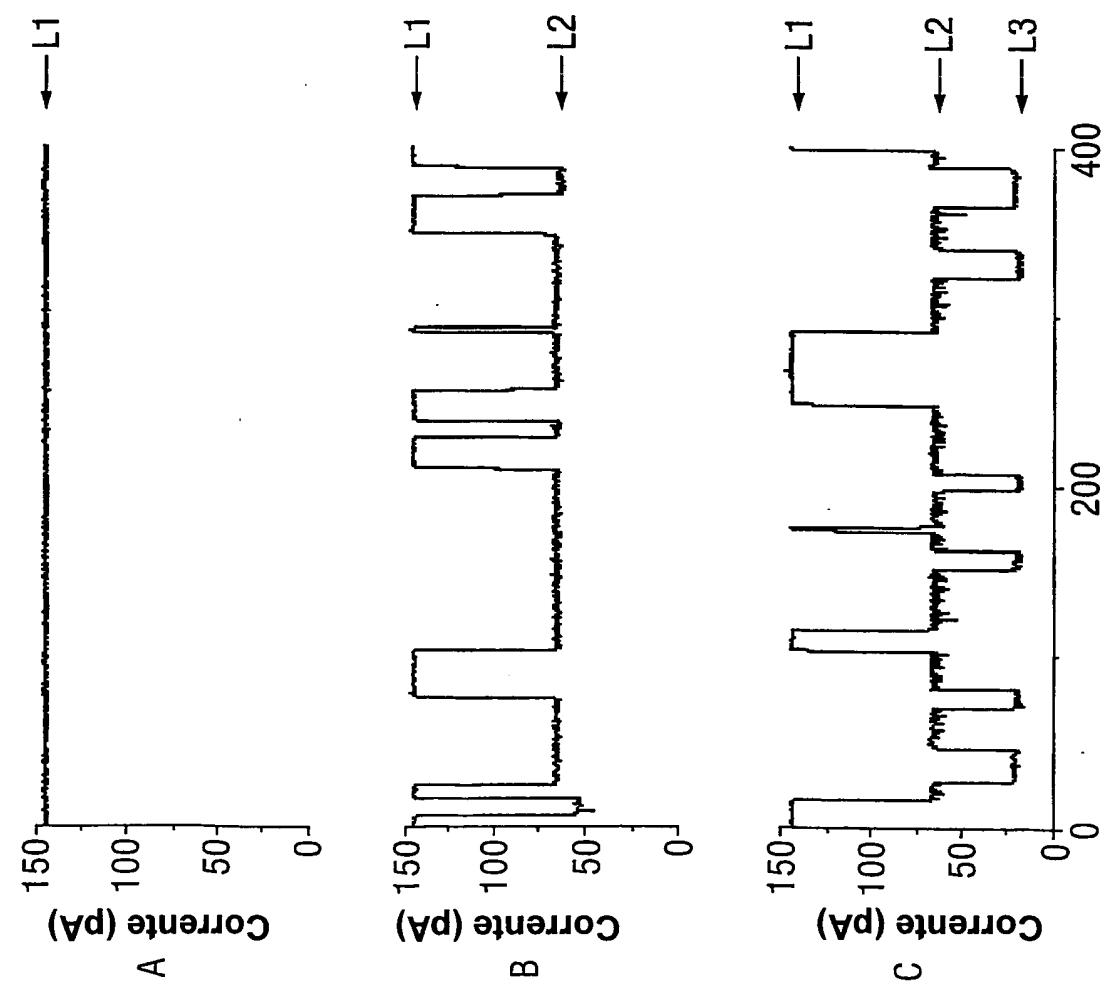


Fig. 2B.

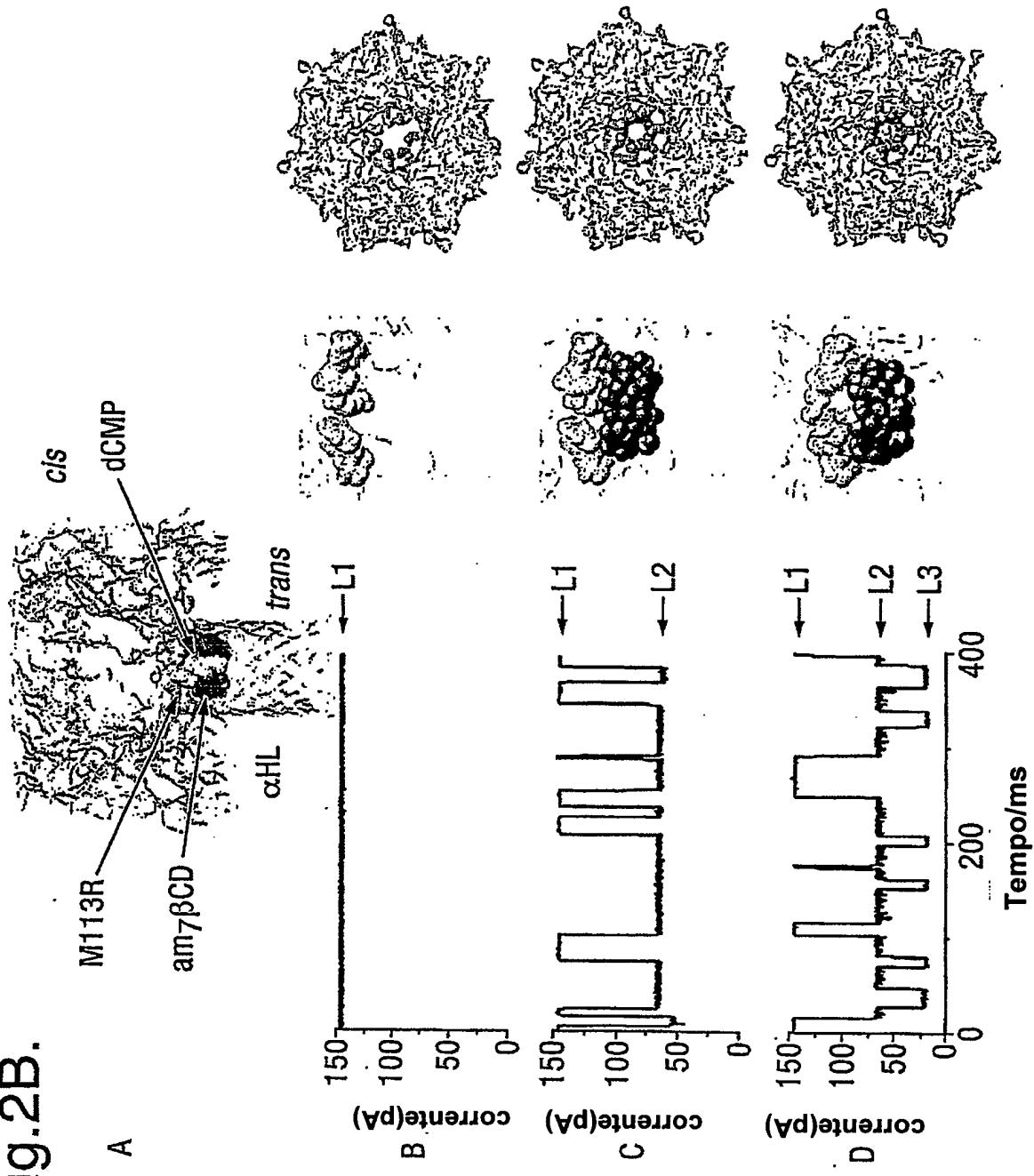


Fig.3.

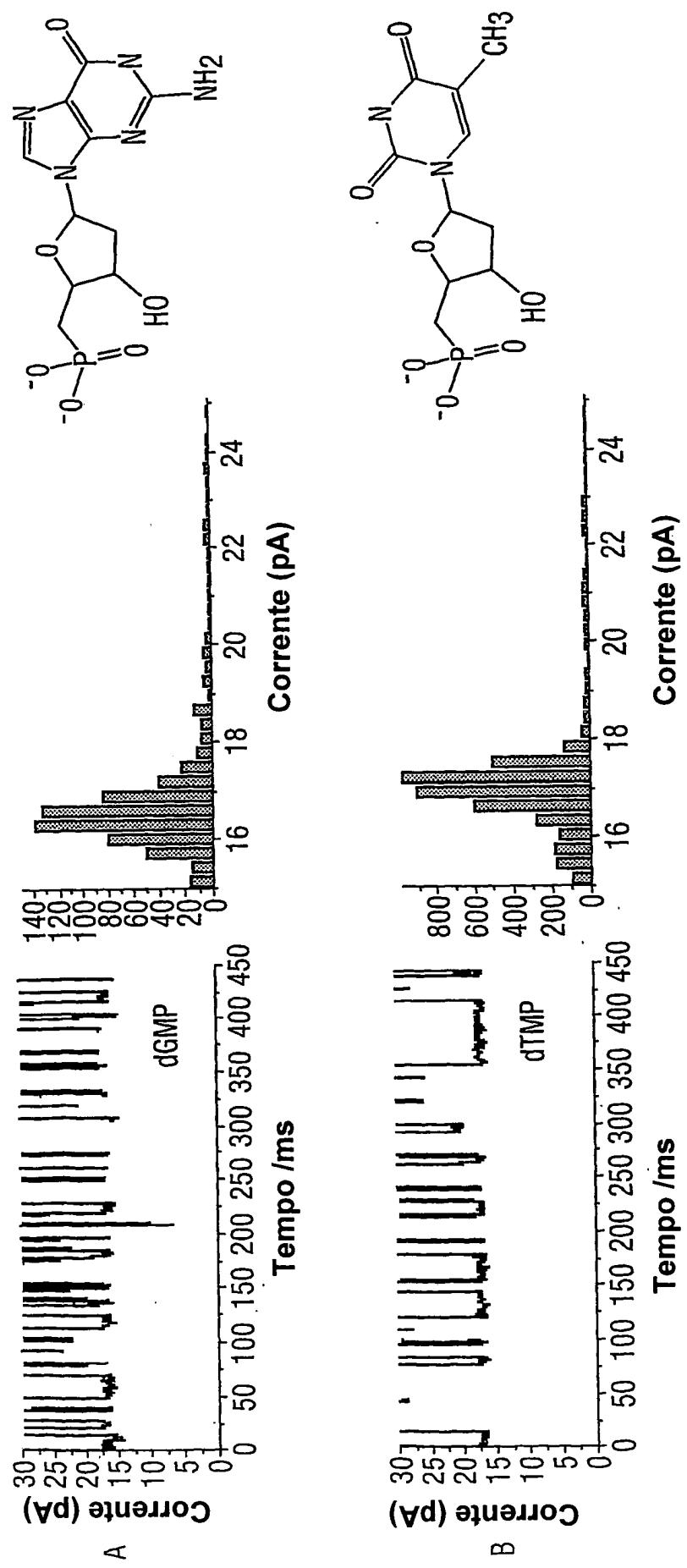


Fig.3 (Cont.).

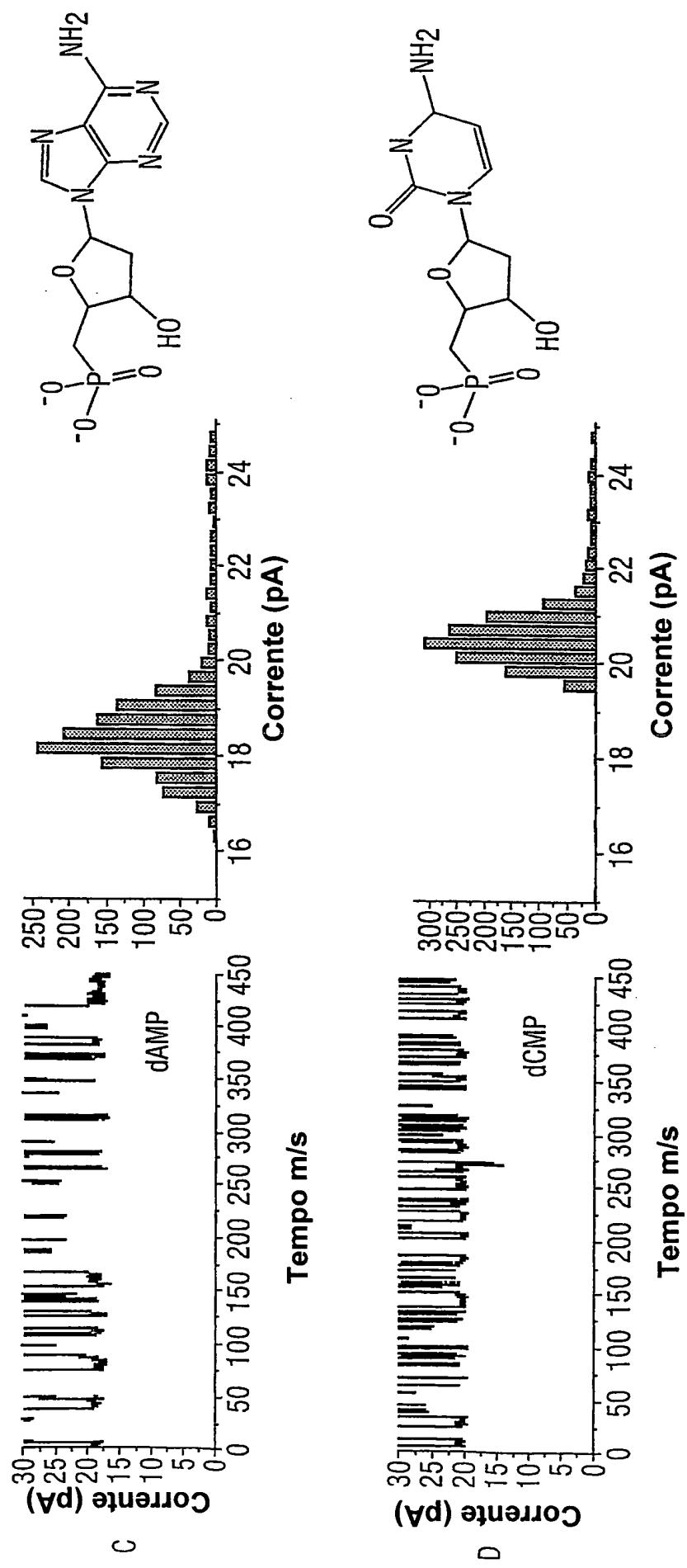


Fig. 4

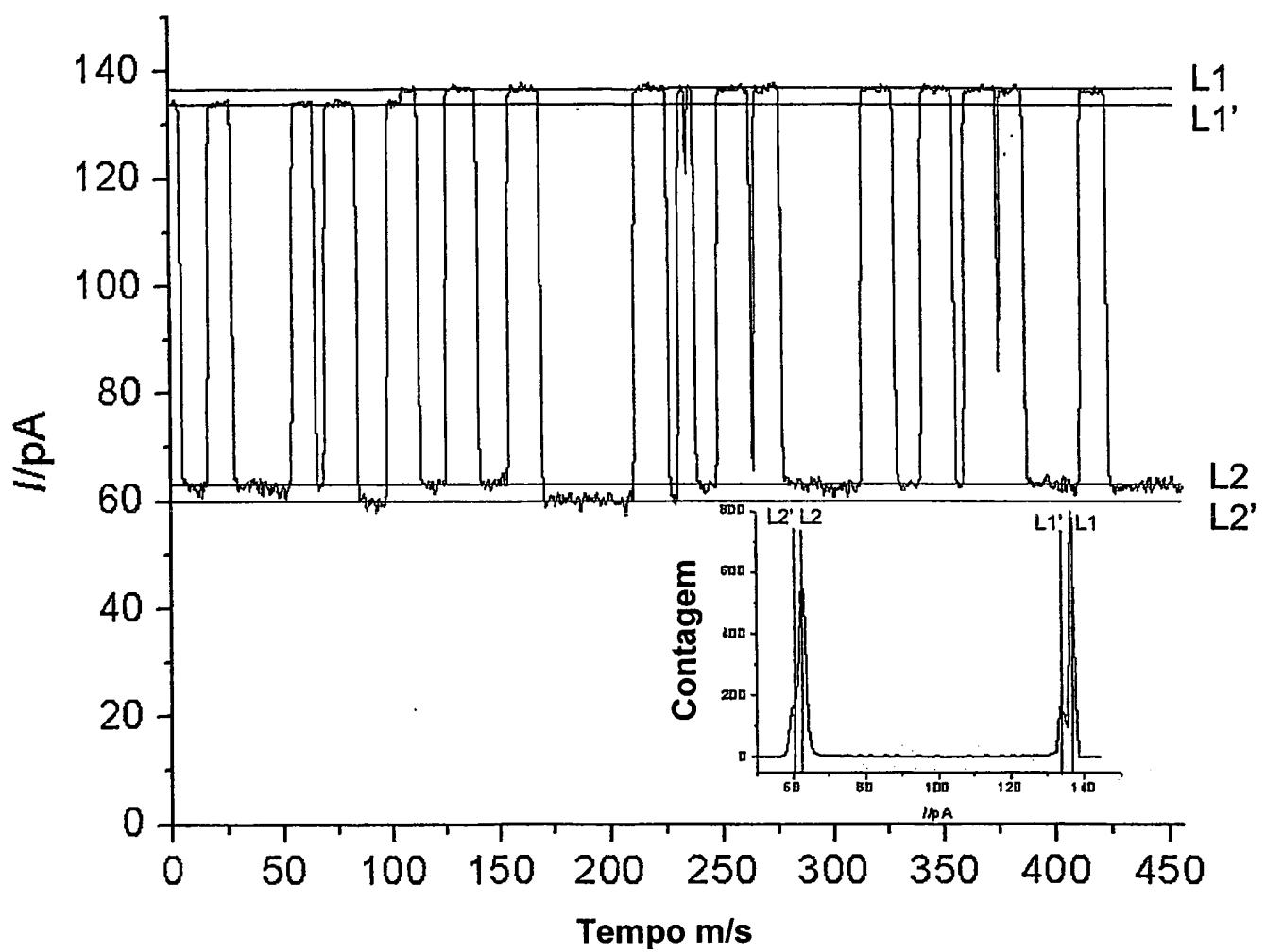


Fig. 5

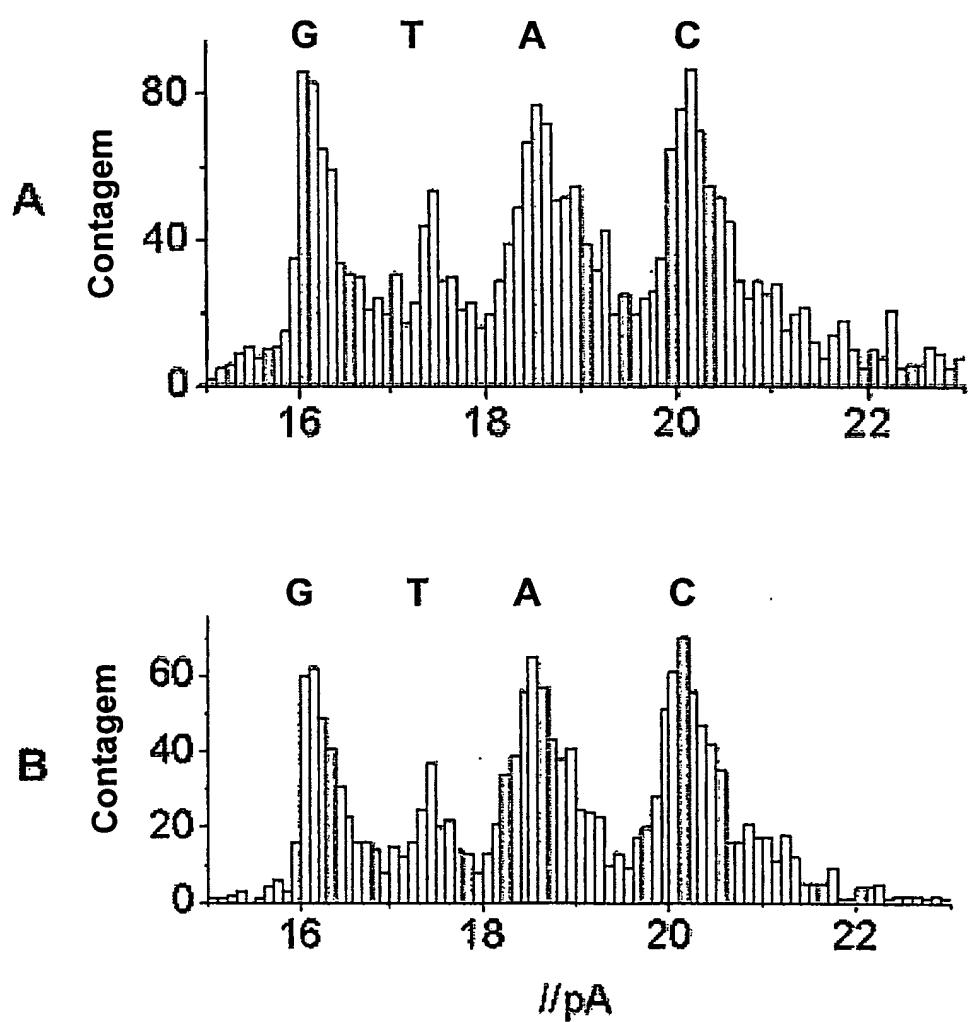


Fig. 6

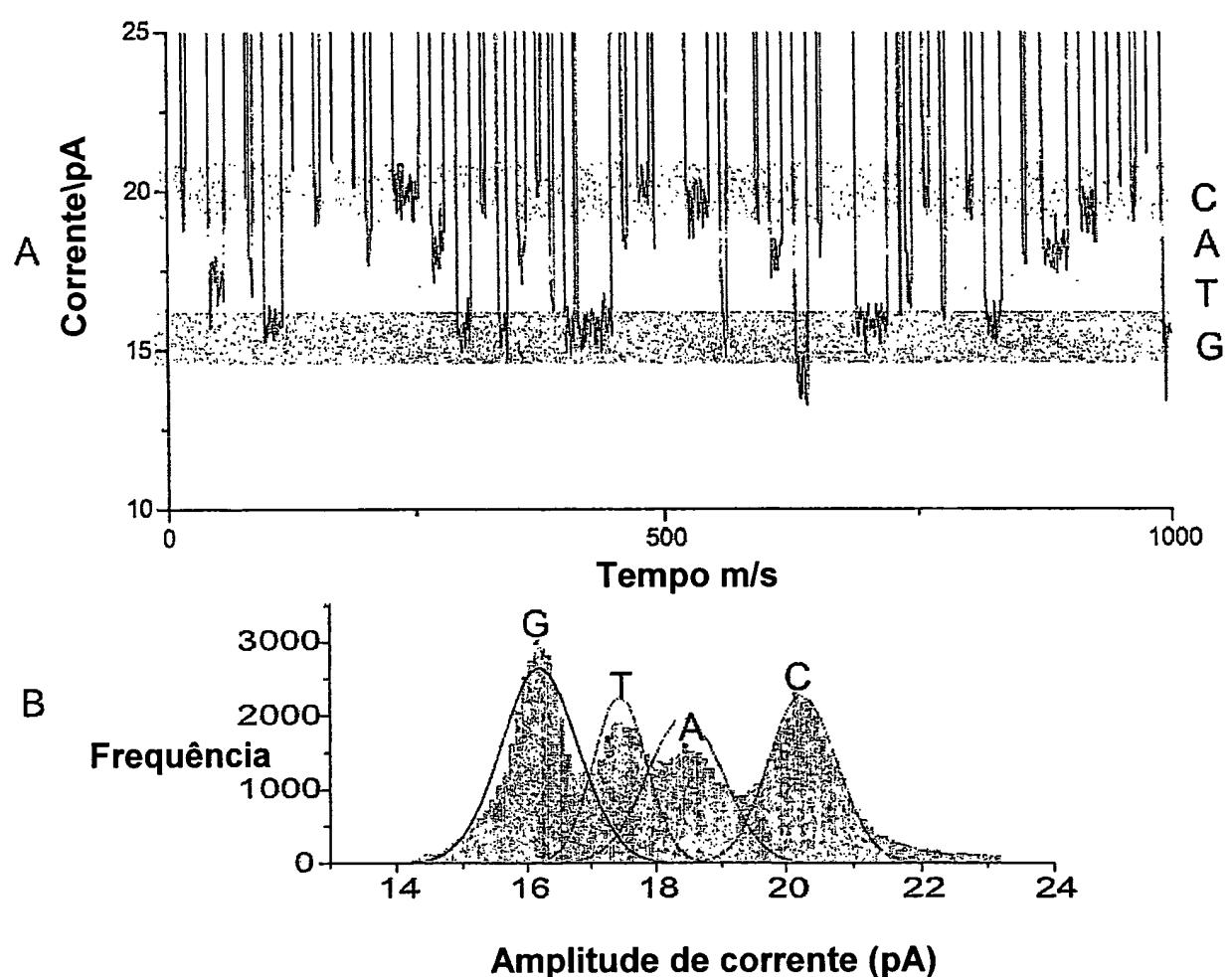
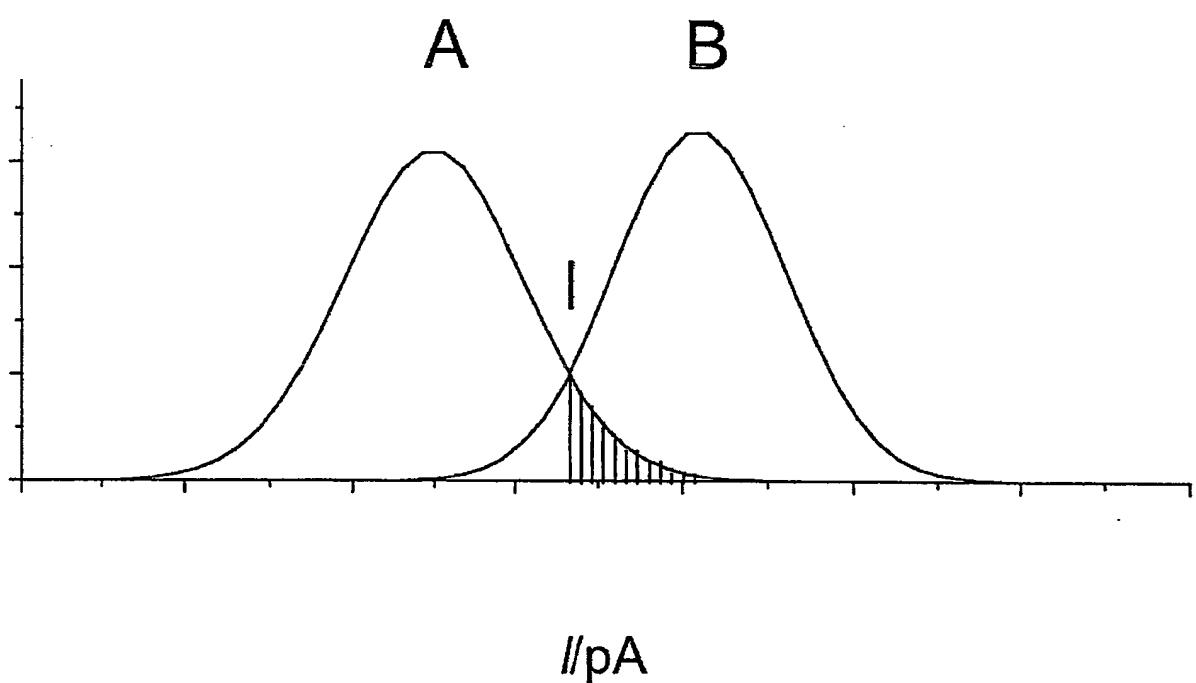


Fig. 7



RESUMO

“MÉTODO PARA IDENTIFICAR UM NUCLEOTÍDEO INDIVIDUAL,  
MÉTODO PARA SEQÜENCIAR UMA SEQÜÊNCIA DE ÁCIDO  
NUCLEICO ALVO, E, KIT PARA SEQÜENCIAR UM ÁCIDO  
5 NUCLEICO”

A invenção refere-se a um método de identificação de um nucleotídeo individual, compreendendo (a) contatar o nucleotídeo com um poro de proteína de transmembrana de modo que o nucleotídeo interaja com o poro e (b) medir a corrente passando através do poro durante a interação e 10 deste modo determinar a identidade do nucleotídeo. A invenção também se refere a um método de seqüenciamento de seqüências de ácido nucleico e aos kits relacionados com o mesmo.