



(19) 대한민국특허청(KR)
 (12) 공개특허공보(A)

(11) 공개번호 10-2008-0083281
 (43) 공개일자 2008년09월17일

(51) Int. Cl.

G01N 27/00 (2006.01) C12Q 1/68 (2006.01)

G01N 33/483 (2006.01)

(21) 출원번호 10-2008-7014358

(22) 출원일자 2008년06월13일

심사청구일자 **없음**

번역문제출일자 2008년06월13일

(86) 국제출원번호 PCT/GB2006/004265

국제출원일자 2006년11월15일

(87) 국제공개번호 WO 2007/057668

국제공개일자 2007년05월24일

(30) 우선권주장

0523282.2 2005년11월15일 영국(GB)

(71) 출원인

아이시스 이노베이션 리미티드

영국 오엑스2 7에스쥐 옥스퍼드주 셤머타운 에워트 플레이스 에워트 하우스 내 영국법인체

(72) 발명자

베일리, 허건

영국 오엑스1 3티에이 옥스포드 맨스필드 로드 12 유니버시티오브 옥스포드 케미스트리 리서치 래버토리

애스티어, 얀

영국 오엑스1 3티에이 옥스포드 맨스필드 로드 12 유니버시티오브 옥스포드 케미스트리 리서치 래버토리

브라하, 오릿

영국 오엑스1 3티에이 옥스포드 맨스필드 로드 12 유니버시티오브 옥스포드 케미스트리 리서치 래버토리

(74) 대리인

양영준, 양영환

전체 청구항 수 : 총 16 항

(54) 공극을 사용한 방법**(57) 요 약**

본 발명은 (a) 개별 뉴클레오티드를 막횡단 단백질 공극과 접촉시켜 상기 뉴클레오티드가 상기 공극과 상호작용하도록 하는 단계, 및 (b) 상기 상호작용 동안에 공극을 통과하는 전류를 측정하여 상기 뉴클레오티드의 정체를 결정하는 단계를 포함하는, 개별 뉴클레오티드를 식별하는 방법에 관한 것이다. 본 발명은 핵산 서열의 서열화 방법 및 그에 관련된 키트에 관한 것이다.

특허청구의 범위

청구항 1

(a) 개별 뉴클레오티드를 막횡단 단백질 공극과 접촉시켜 상기 뉴클레오티드가 상기 공극과 상호작용하도록 하는 단계; 및

(b) 상기 상호작용 동안에 공극을 통과하는 전류를 측정하여 상기 뉴클레오티드의 정체(identity)를 결정하는 단계를 포함하는, 개별 뉴클레오티드의 식별 방법.

청구항 2

제1항에 있어서, 상호작용이 공극의 통로에 가역적으로 결합하는 뉴클레오티드를 포함하는 것인 방법.

청구항 3

제1항 또는 제2항에 있어서, 공극이 서열 2에 나타낸 α -헤모리신 또는 그의 변형체(variant)인 것인 방법.

청구항 4

제3항에 있어서, 변형체가 (M113R)₇인 것인 방법.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 공극이 뉴클레오티드와 공극 간의 상호작용을 촉진하는 분자 연결기(adaptor)를 포함하는 것인 방법.

청구항 6

제5항에 있어서, 분자 연결기가 헵타카스-6-아미노- β -시클로덱스트린 (am₇- β -CD)인 것인 방법.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 개별 뉴클레오티드가 모노포스페이트, 디포스페이트 또는 트리포스페이트인 것인 방법.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 개별 뉴클레오티드가 리보뉴클레오티드인 것인 방법.

청구항 9

제8항에 있어서, 단계 (a) 전에 리보핵산(RNA) 서열을 소화시켜 개별 뉴클레오티드를 제공하는 단계를 더 포함하는 것인 방법.

청구항 10

제1항 내지 제7항 중 어느 한 항에 있어서, 개별 뉴클레오티드가 데옥시리보뉴클레오티드인 것인 방법.

청구항 11

제10항에 있어서, 단계 (a) 전에 데옥시리보핵산(DNA) 서열을 소화시켜 개별 뉴클레오티드를 제공하는 단계를 더 포함하는 것인 방법.

청구항 12

제9항 또는 제11항에 있어서, RNA 또는 DNA 서열의 전체 또는 일부의 정체를 결정할 수 있는 연속적인 방식으로 상기 서열의 개별 뉴클레오티드의 하나 이상을 공극과 접촉시키는 것인 방법.

청구항 13

개별 뉴클레오티드를 식별하기 위한 공극의 용도.

청구항 14

- (a) 진행성(processive) 엑소뉴클레아제를 사용하여 표적 핵산 서열의 한쪽 말단으로부터 개별 뉴클레오티드를 소화시키는 단계;
- (b) 상기 뉴클레오티드를 막횡단 단백질 공극과 접촉시켜 상기 뉴클레오티드가 상기 공극과 상호작용하도록 하는 단계;
- (c) 상기 상호작용 동안에 공극을 통과하는 전류를 측정하여 상기 뉴클레오티드의 정체를 결정하는 단계; 및
- (d) 핵산 서열의 동일한 말단에서 단계 (a) 내지 (c)를 반복하여 핵산의 서열을 결정하는 단계를 포함하는, 표적 핵산 서열의 서열화 방법.

청구항 15

- 시클로덱스트린; 및
 - 진행성 엑소뉴클레아제
- 를 포함하는, 핵산을 서열화하기 위한 키트.

청구항 16

제15항에 있어서, 막횡단 단백질 공극을 더 포함하는 키트.

명세서

기술 분야

<1> 본 발명은 막횡단 공극을 사용한 개별 뉴클레오티드 및 기타 포스페이트 함유 잔기의 식별에 관한 것이다. 특히, 본 발명은 막횡단 공극을 사용한 표적 핵산의 서열화에 관한 것이다.

배경 기술

<2> DNA의 서열화를 위해 현재 사용되는 방법은 형광 ddXTP, dXTP, 프라이머 및 폴리머라제와 같은 여러 고가의 시약을 사용하는 것을 포함한다. 이러한 방법은 정교한 장치를 필요로 하며, 이는 면허를 가진 기술자에 의해 작동될 필요가 있다. 또한, 이러한 방법은 1000 미만의 뉴클레오티드 길이의 서열에 한정된다.

<3> 비용을 줄이고 방법을 단순화하고 실험실 밖에서 서열화를 수행하기 위하여 다른 서열화 방법이 고려되었다. 주기 연장, 폴리머라제 판독, 엑소뉴클레아제 서열화 및 DNA 마이크로-어레이가 고려되어온 방법이다 [Braslavsky, I., B., Herbert 등, (2003), PNAS 100(7): 3960-3964]. 이러한 방법들이 포괄적으로 검토되었다 [Marziali,A. 및 M.Akeson (2001), Ann.Rev.Biomed. Eng. 3: 195-223] 참조.

<4> DNA 서열화의 한가지 가능한 방법은, 나노공극을 통해 DNA의 단일 가닥을 관통시키고, 가닥이 관통될 때 공극을 통해 흐르는 이온 전류에서의 변동으로부터 그의 서열을 식별하는 것을 기초로 한다 [Kasianowicz,J. J., E. Brandin 등 (1996), Proc.Natl.Acad.Sci. 93: 13770-13773]. 두 번째 가능한 접근은 엑소뉴클레아제 서열화이다 [Chan,E.Y. (2005), Mutat.Res. 573: 13-40]. 이 방법은 DNA를 한번에 하나의 뉴클레오티드씩 소화시키고 ([Dapprich,J. (1999), Cytomet. 36: 163-168]; 및 [Matsuura,S.-I., J.Komatsu, 등 (2001), Nuc.Ac.Res. 29 (16): e79]), 이어서 각각의 방출된 뉴클레오티드를 식별하는 것을 포함한다. 그러나, 이러한 방법은 엑소뉴클레아제에 의해 DNA로부터 일단 방출된 후 뉴클레오티드의 소화 또는 변형에 앞서서 DNA의 변형을 필요로 한다. 뉴클레오티드가 효소에 의해 방출될 때 단일 분자 수준에서 뉴클레오티드를 식별하는데 있어서의 곤란함에 의하여 엑소뉴클레아제 서열화의 발전이 일방적으로 방해받고 있다. 연구자들은 형광 표지화를 사용하여 뉴클레오티드를 식별하고자 시도하고 있으나 성공은 제한적이다.

<5> 확률적 탐지(stochastic sensing)는 나노미터 크기 공극을 절연 지질 이층 막에 놓고 공극을 통한 이온 전달을 측정하는 것과 연관된다. 분석물이 공극 내에서 결합 부위와 상호작용할 때, 이온 전류에서의 변화가 검출된다 [Braha,O., B. Walker 등 (1997), Chem. & Biol. 4: 497-505; 및 Bayley, H. 및 P.S.Cremer (2001), Nature 413: 226-230]. 각각의 결합 사건으로부터 얻어지는 전류 블록의 정도 및 지속기간이 분석물의 정체(identit

y)를 드러낼 수 있다. 결합 사건의 빈도는 분석물 농도를 드러낼 수 있다. 단백질 돌연변이, 화학적 변형에 의해, 그리고 문자 연결기(adaptor) 및 담체의 사용에 의해 공극 내에서 다양한 결합 부위가 생성될 수 있다 [Gu,L.-Q., O. Braha 등 (1999), Nature 398: 686-690; 및 Braha, O., J. Webb. 등 (2005), Chem.Phys.Chem. 6: 889-892].

<6> 발명의 요약

<7> 놀랍게도, 개별 뉴클레오티드가 막횡단 공극과 상호작용할 때 그들의 전류 진폭으로부터 단일 분자 수준에서 개별 뉴클레오티드가 식별될 수 있다는 것이 증명되었다. 따라서, 개별 뉴클레오티드를 식별하고 엑소뉴클레아제 서열화를 통해 핵산 서열을 서열화하기 위하여 확률적 탐지가 사용될 수 있다.

<8> 따라서, 본 발명은

<9> (a) 개별 뉴클레오티드를 막횡단 단백질 공극과 접촉시켜 상기 뉴클레오티드가 상기 공극과 상호작용하도록 하는 단계; 및

<10> (b) 상기 상호작용 동안에 공극을 통과하는 전류를 측정하여 상기 뉴클레오티드의 정체를 결정하는 단계를 포함하는, 개별 뉴클레오티드의 식별 방법을 제공한다.

<11> 본 발명은

<12> - (a) 진행성(processive) 엑소뉴클레아제를 사용하여 핵산 표적 서열의 한쪽 말단으로부터 개별 뉴클레오티드를 소화시키는 단계;

<13> (b) 상기 뉴클레오티드를 막횡단 단백질 공극과 접촉시켜 상기 뉴클레오티드가 상기 공극과 상호작용하도록 하는 단계;

<14> (c) 상기 상호작용 동안에 공극을 통과하는 전류를 측정하여 상기 뉴클레오티드의 정체를 결정하는 단계; 및

<15> (d) 핵산 서열의 동일한 말단에서 단계 (a) 내지 (c)를 반복하여 핵산의 서열을 결정하는 단계를 포함하는, 표적 핵산 서열의 서열화 방법; 및

<16> - 시클로텍스트린; 및

<17> - 진행성 엑소뉴클레아제

<18> 를 포함하는 핵산의 서열화를 위한 키트를 제공한다.

<19> 본 발명의 서열화 방법은 단일 분자 수준에서 빠르고 간단한 DNA 서열화 방법이다. 또한, 형광단과 같은 비싼 시약의 사용을 포함하지 않기 때문에 DNA의 저렴한 서열화 방법이다.

발명의 상세한 설명

<38> **개별 뉴클레오티드의 식별 방법**

<39> 첫 번째 구현양태에서, 본 발명은 개별 뉴클레오티드를 막횡단 단백질 공극과 접촉시켜 상기 뉴클레오티드가 상기 공극과 상호작용하도록 하는 단계 및 상기 상호작용 동안에 공극을 통과하는 전류를 측정하여 뉴클레오티드의 정체를 결정하는 단계를 포함하는, 개별 뉴클레오티드의 식별 방법에 관한 것이다. 따라서, 본 발명은 개별 뉴클레오티드의 확률적 탐지와 연관된다. 본 발명은 막횡단 단백질 공극을 통과하는 전류에 미치는 상이한 영향을 기초로 하여 유사한 구조의 뉴클레오티드를 구별하기 위해 사용될 수 있다. 또한, 본 발명은 특정한 뉴클레오티드가 샘플에 존재하는지 아닌지의 여부를 결정하기 위해 사용될 수 있다. 본 발명은 또한 샘플 중에서 특정한 뉴클레오티드의 농도를 측정하기 위해 사용될 수 있다.

<40> 본 발명에 따른 개별 뉴클레오티드는 단일 뉴클레오티드이다. 개별 뉴클레오티드는 뉴클레오티드 결합에 의해 다른 폴리뉴클레오티드에 결합되지 않는 것이다. 뉴클레오티드 결합은 다른 뉴클레오티드의 당 기에 결합된 뉴클레오티드의 포스페이트 기의 하나를 포함한다. 개별 뉴클레오티드는 전형적으로 뉴클레오티드 결합에 의하여 적어도 5, 적어도 10, 적어도 20, 적어도 50, 적어도 100, 적어도 200, 적어도 500, 적어도 1000 또는 적어도 5000 뉴클레오티드의 다른 폴리뉴클레오티드 서열에 결합되지 않는 것이다. 예를 들어, 개별 뉴클레오티드는 DNA 또는 RNA 가닥과 같은 표적 폴리뉴클레오티드 서열로부터 소화되었다. 그러나, 개별 뉴클레오티드는 다른 화학 기, 예컨대 형광 분자 또는 방사성동위원소를 함유한 화학 기, 예를 들어 ^{125}I , ^{35}S 에 결합되거나 부착될 수

도 있다. 본 발명에 따른 식별을 위한 뉴클레오티드의 유형을 이하에서 더욱 상세히 언급한다.

<41> 막횡단 단백질 공극이 막에 삽입되는 적절한 막/공극 체계를 사용하여 방법을 수행할 수 있다. 방법은 전형적으로 (i) 천연 또는 재조합 막횡단 단백질 공극을 포함한 인공 막, (ii) 재조합 막횡단 단백질 공극을 포함하는 단리된 천연 막, (iii) 막횡단 단백질 공극을 포함한 단리된 천연 막, 또는 (iv) 천연 또는 재조합 막횡단 단백질 공극을 발현하는 세포를 사용하여 수행된다. 방법은 바람직하게는 인공 막을 사용하여 수행된다. 막은 막횡단 단백질 공극에 추가로 다른 막횡단 및/또는 막 내부 단백질뿐만 아니라 다른 분자를 포함할 수 있다.

<42> 본 발명의 방법은 전형적으로 시험관내에서 수행된다.

막

<44> 막은 이온 및 뉴클레오티드의 흐름에 대한 장벽을 형성한다. 막은 바람직하게는 지질 이층이다. 당 기술분야에 공지된 방법을 사용하여 본 발명에 따라 사용하기 위한 지질 이층을 형성할 수 있다. 예를 들어, 몬탈(Montal) 및 뮤엘러(Mueller) (1972)의 방법을 사용하여 지질 이층 막을 형성할 수 있다. 이에 한정되지 않지만 인지질, 당지질, 콜레스테롤 및 이들의 혼합물을 포함하는 막 지질로부터 형성된 지질 이층을 사용하여 본 발명의 방법을 수행할 수도 있다. 지질 이층은 바람직하게 1,2-디피타노일-sn-글리세로-3-포스포콜린으로부터 형성된다.

<45> 공극을 지질 이층과 같은 막 안에 삽입하기 위해 당 기술분야에서 방법이 공지되어 있다. 예를 들어, 용액이 지질 이층으로 확산되고 지질 이층에 결합하고 기능성 상태로 조립됨으로써 삽입되도록 공극을 지질 이층을 함유한 용액 중에서 정제된 형태로 혼탁할 수도 있다. 대안적으로, 문헌 [M.A.Holden, H.Bayley. J.Am.Chem.Soc. 2005, 127, 6502-6503]에 기재된 방법을 사용하여 공극을 막 안에 직접적으로 삽입할 수도 있다.

막횡단 단백질 공극

<47> 막횡단 단백질 공극을 사용하여 본 발명의 방법이 수행된다. 막횡단 단백질 공극은 전기화학 구배를 따라 막의 한쪽 면으로부터 다른 면으로 이온이 흐르게 하는 폴리펩티드이다. 공극은 바람직하게는 뉴클레오티드가 전기화학 구배를 따라 막의 한쪽 면으로부터 다른쪽으로 흐르도록 한다.

<48> 공극은 전형적으로 올리고머이다. 공극은 바람직하게는 몇 개의 반복되는 소단위로 이루어진다. 공극은 바람직하게는 펜타머 또는 헵타머이다. 공극은 전형적으로 이온이 흐를 수 있는 통 또는 통로를 포함한다.

<49> 공극의 통 또는 통로는 전형적으로 뉴클레오티드와의 상호작용을 촉진하는 아미노산을 포함한다. 본 발명에 따라 사용하기 위한 공극은 전형적으로 하나 이상의 양 하전된 아미노산, 예컨대 아르기닌, 리신 또는 히스티дин을 포함한다. 이러한 양 하전된 아미노산은 바람직하게는 통 또는 통로의 수축 근처에 위치한다. 이러한 아미노산은 전형적으로 뉴클레오티드에서 포스페이트 기와의 상호작용에 의해 또는 뉴클레오티드에서 염기와의 p-양이온 상호작용에 의해 공극과 뉴클레오티드 사이의 상호작용을 촉진한다. 공극은 바람직하게는 통 또는 통로의 수축 부분에 가까이 위치한 아르기닌, 리신 또는 히스티딘과 같은 양 하전된 아미노산의 고리를 갖는다. 각각의 공극 소단위에 의하여 각각의 양 하전된 아미노산이 전형적으로 제공된다.

<50> 본 발명에 따른 사용을 위해 적절한 공극은 이에 한정되지 않지만 α -헤모리신, 포린 및 류코시딘을 포함한다.

<51> 본 발명에서 사용하기 위해 바람직한 공극은 α -헤모리신 또는 그의 변형체(variant)이다. α -헤모리신 공극은 7개의 동일한 소단위(헵타머)로 형성된다. α -헤모리신의 1개 소단위의 서열을 서열 2에 나타낸다. 변형체는 7개 소단위의 하나 이상이 서열 2로부터 변형되고 공극 활성을 보유하는 아미노산 서열을 갖는 것인 헵타머 공극이다. 변형체 α -헤모리신에서 소단위의 1, 2, 3, 4, 5, 6 또는 7개가 서열 2로부터 변형된 아미노산 서열을 가질 수도 있다. 변형체 공극 내의 7개 소단위는 전형적으로 동일하지만 상이할 수도 있다.

<52> α -헤모리신의 바람직한 변형체는 통 또는 통로의 수축 부분 근처에 위치한 하나 이상의 양 하전된 아미노산, 예컨대 아르기닌, 리신 또는 히스티딘을 갖는다. 공극은 바람직하게는 통 또는 통로의 수축 부분에 위치한 아르기닌, 리신 또는 히스티딘과 같은 4, 5, 6 또는 바람직하게는 7개 양 하전된 아미노산의 고리를 갖는다. 고리에서 각각의 아미노산은 전형적으로 각각의 변형체 소단위에 의해 제공된다. 변형체는 전형적으로 각각의 소단위의 위치 113에서 양 하전된 아미노산을 포함한다. 본 발명에서 사용하기 위한 공극은 바람직하게는 서열 4에 나타낸 것과 같은 7개 소단위를 포함하는 α -헤모리신(M113K)₇이나 또는 바람직하게는 서열 6에 나타낸 7개 소단위를 포함하는 α -헤모리신(M113H)₇ 또는 가장 바람직하게는 서열 8에 나타낸 7개 소단위를 포함하는 α -헤모리신(M113R)₇이다.

<53> 변형체는 유기체, 예를 들어 스타필로코쿠스 세균에 의해 발현된 천연 변형체일 수도 있다. 변형체는 재조합 기술에 의해 생성된 비-천연 변형체를 포함한다. 서열 2의 아미노산 서열의 전체 길이에 걸쳐, 변형체의 소단위는 바람직하게는 아미노산 동일성을 기초로 하여 서열에 대해 적어도 50% 상동성이다. 더욱 바람직하게는, 소단위 폴리펩티드는 전체 서열에 걸쳐서 서열 2의 아미노산 서열에 대해 아미노산 동일성을 기초로 하여 적어도 55%, 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90% 및 더욱 바람직하게는 적어도 95%, 97% 또는 99%일 수도 있다. 200개 이상, 예를 들어 230, 250, 270 또는 280 또는 그 이상의 연속 아미노산의 연장에 걸쳐 적어도 80%, 예를 들어 적어도 85%, 90% 또는 95% 아미노산 동일성이 존재할 수도 있다 ("엄격한 상동성").

<54> 서열 2의 아미노산 서열에 대해 아미노산 치환, 예를 들어 1, 2, 3, 4, 5, 10, 20 또는 30 이하의 치환이 이루어질 수도 있다. 예를 들어 하기 표에 따라 보존적 치환이 이루어질 수도 있다. 서로에 대해 두 번째 세로줄의 동일한 블록, 바람직하게는 세 번째 세로줄의 동일한 선에 있는 아미노산들이 서로 치환될 수도 있다:

비-방향족	비-극성	GAP
		ILV
	극성-비하전	CSTM
		NQ
방향족	극성-하전	DE
		HKR
		HFWY

<56> 서열 2의 아미노산 서열의 하나 이상의 아미노산 잔기는 대안적으로 또는 추가로 결실될 수도 있다. 1, 2, 3, 4, 5, 10, 20 또는 30개까지의 잔기 또는 그 이상이 결실될 수도 있다.

<57> 변형체는 서열 2의 단편으로 이루어진 소단위를 포함할 수도 있다. 이러한 단편은 공극 형성 활성을 보유한다. 단편은 적어도 50, 100, 200 또는 250개 아미노산 길이일 수도 있다. 키메라 공극을 생성하기 위해 이러한 단편이 사용될 수도 있다. 단편은 바람직하게는 서열 2의 공극 형성 도메인을 포함한다.

<58> 변형체는 서열 2의 단편 또는 부분을 포함하는 키메라 단백질 공극을 포함한다. 각각 서열 2의 단편 또는 부분을 포함하는 소단위로부터 키메라 단백질 공극이 형성된다. 서열 2의 단편 또는 부분에 의하여 키메라 단백질 공극의 공극 또는 통로 부분이 전형적으로 형성된다.

<59> 하나 이상의 아미노산이 상기 기재된 폴리펩티드에 대안적으로 또는 추가로 첨가될 수도 있다. 서열 2의 아미노산 서열 또는 그의 폴리펩티드 변형체 또는 단편의 N-말단 또는 C-말단에서 연장부가 제공될 수도 있다. 연장부는 상당히 짧고, 예를 들어 1 내지 10개 아미노산 길이일 수도 있다. 대안적으로, 연장부는 더욱 길 수도 있고, 예를 들어 50 또는 100개까지의 아미노산일 수도 있다. 담체 단백질이 본 발명에 따른 아미노산 서열에 융합될 수도 있다.

<60> 상동성을 결정하기 위하여 당 기술분야의 표준 방법이 사용될 수도 있다. 예를 들어, UWGCG 패키지는 예를 들어 디풀트 세팅에서 사용되는 상동성을 계산하기 위해 사용될 수 있는 BESTFIT 프로그램을 제공한다 [Devereux 등 (1984) Nucleic Acids Research 12, p387-395]. 예를 들어 문헌 [Altschul S.F. (1993) J Mol Evol 36:290-300; Altschul,S.F. 등 (1990) J Mol Biol 215:403-10]에 기재된 바와 같이, 상동성을 계산하거나 서열을 정렬시키기 위하여 (예컨대 (전형적으로, 그들의 디풀트 세팅에서) 균등한 잔기 또는 상응하는 서열을 식별하기 위하여) PILEUP 및 BLAST 알고리듬이 사용될 수도 있다.

<61> BLAST 분석을 수행하기 위한 소프트웨어는 국립 바이오테크놀로지 정보 센터를 통해 공개적으로 입수 가능하다 (<http://www.ncbi.nlm.nih.gov/>). 이러한 알고리듬은, 데이터베이스 서열에서 동일한 길이의 단어와 정렬될 때, 일부 포지티브 값 역치 스코어 T와 일치하거나 이를 만족시키는 의문 서열에서 길이 W의 짧은 단어를 식별함으로써 고 스코어 서열 쌍(HSP)을 먼저 식별하는 것을 포함한다. T는 이웃한 단어 스코어 역치로 일컬어진다 (Altschul 등, 상동). 이러한 처음의 이웃 단어 적중은 이를 함유하는 HSP를 찾아내기 위한 조사를 시작하기 위한 핵심으로 작용한다. 누적 정렬 스코어가 증가하는 한, 각각의 서열을 따라 양쪽 방향에서 단어 적중이 연장된다. 누적 정렬 스코어가 최대 달성을 값으로부터 X 량 만큼씩 떨어질 때 각각의 방향에서 단어 적중을 위한 연장이 중단되고; 하나 이상의 네가티브-스코어 잔기 정렬의 축적에 기인하여 누적 스코어는 제로 또는 그 미만이 되거나; 또는 어느 하나의 서열의 말단에 도달된다. BLAST 알고리듬 매개변수 W, T 및 X는 정렬의 감도 및

속도를 결정한다. BLAST 프로그램은 11의 단어 길이(W), 50의 BLOSUM62 스코어 행렬 ([Henikoff and Henikoff (1992) Proc.Natl.Acad.Sci. USA 89: 10915-10919] 참조) 정렬 (B), 10의 예상치(E), M=5, N=4 및 양쪽 가닥의 비교를 디폴트로서 사용한다.

- <62> BLAST 알고리듬은 2개 서열 간의 유사성을 통계적으로 분석한다; 예를 들어 문헌 [Karlin and Altschul (1993) Proc.Natl.Acad.Sci. USA 90:5873-5787] 참조. BLAST 알고리듬에 의해 제공된 유사성의 한가지 측정은 최소 합계 확률(P(N))이고, 이것은 2개의 아미노산 서열 간의 일치가 우연히 발생할 확률을 표시하는 것이다. 예를 들어, 두 번째 서열에 대한 첫 번째 서열의 비교에서 최소 합계 확률이 약 1 미만, 바람직하게는 약 0.1 미만, 더욱 바람직하게는 약 0.01 미만, 가장 바람직하게는 약 0.001 미만이라면 서열이 다른 서열과 유사한 것으로 간주된다.
- <63> 예를 들어 그들의 식별 또는 정체를 돋기 위해 히스티딘 잔기를 첨가함으로써, 또는 폴리펩티드가 이러한 서열을 자연적으로 함유하지 않는 세포로부터 그들의 분비를 촉진하기 위해 시그널 서열을 첨가함으로써, 본 발명에 따라 사용되는 공극이 변형될 수도 있다. 고체 지지체에 부착하기 위해 적절한 형태로 폴리펩티드를 제공하는 것이 바람직 할 수도 있다. 예를 들어, 공극을 막에 삽입하기 위해 공극이 고체 지지체에 부착될 수도 있다.
- <64> 공극을 보이는 표지로 표지화할 수도 있다. 보이는 표지는 공극이 검출될 수 있도록 하는 적절한 표지일 수도 있다. 적절한 표지는 이에 한정되지 않지만 형광 분자, 방사성 동위원소, 예를 들어 ¹²⁵I, ³⁵S, 효소, 항체, 폴리뉴클레오티드 및 링커, 예컨대 비오틴을 포함한다.
- <65> 공극-생성 유기체, 예컨대 스타필로코쿠스 아우레우스로부터 공극이 단리될 수도 있거나 또는 합성적으로 또는 재조합 수단에 의해 만들어질 수 있다. 예를 들어, 시험관내 번역 전사에 의해 공극을 합성할 수도 있다. 비-천연 아미노산을 포함하기 위해 또는 화합물의 안정성을 증가시키기 위해 공극의 아미노산 서열을 변형시킬 수도 있다. 공극이 합성 수단에 의해 생성될 때, 생성 동안에 그러한 아미노산이 도입될 수도 있다. 합성 또는 재조합 생성에 따라 공극을 변형시킬 수도 있다.
- <66> D-아미노산을 사용하여 공극을 생성할 수도 있다. 이러한 경우에, 아미노산이 C에서 N 배향으로 역 서열로 연결될 것이다. 이것은 단백질 또는 웨პ티드를 생성하기 위한 기술 분야에서 통상적이다.
- <67> 다수의 측쇄 변형이 당 기술분야에 공지되어 있고 공극의 측쇄로 만들어질 수도 있다. 이러한 변형은 예를 들어 알데히드와의 반응에 이어서 NaBH₄와의 환원에 의한 환원적 알킬화, 메틸아세트이미데이트와의 아미드화 또는 아세트 안히드라이드와의 아실화에 의하여 아미노산의 변형을 포함한다.
- <68> 당 기술분야에 공지된 표준 방법을 사용하여 재조합 막횡단 공극이 생성될 수 있다. 공극을 코드화하는 핵산 서열이 단리될 수 있고 당 기술분야에서의 표준 방법을 사용하여 복제될 수 있다. 당 기술분야의 표준 기술을 사용하여 공극을 코드화하는 핵산 서열이 세균 숙주 세포에서 발현될 수도 있다. 재조합 발현 벡터로부터 폴리펩티드의 제자리(*in-situ*) 발현에 의하여 공극이 세포 내로 도입될 수 있다. 폴리펩티드의 발현을 제어하기 위하여 발현 벡터는 임의로 유도가능한 프로모터를 보유한다.
- <69> 당 기술분야의 표준 방법을 사용하여 공극을 코드화하는 핵산 서열을 단리하고 복제할 수 있다. 스타필로코쿠스 아우레우스와 같은 공극-생성 유기체로부터 염색체 DNA를 추출할 수 있다. 특정한 프로모터와 연관된 PCR을 사용하여 공극을 코드화하는 유전자를 증폭할 수도 있다. 증폭된 서열을 클론화 벡터와 같은 재조합 복제가능한 벡터 내에 혼입할 수도 있다. 친화성 숙주 세포에서 핵산을 복제하기 위해 벡터를 사용할 수도 있다. 따라서, 복제가능한 벡터 내에 공극을 코드화하는 폴리뉴클레오티드를 도입하고, 벡터를 친화성 숙주 세포 내에 도입하고, 벡터의 복제를 일으키는 조건 하에서 숙주 세포를 생육시킴으로써 공극을 코드화하는 핵산 서열이 만들어질 수도 있다. 숙주 세포로부터 벡터가 회수될 수도 있다. 공극을 코드화하는 폴리뉴클레오티드의 클론화를 위해 적절한 숙주 세포가 당 기술분야에 공지되어 있고 이하에서 더욱 상세히 설명된다.
- <70> 공극을 코드화하는 핵산 서열이 적절한 발현 벡터 내로 클론화될 수도 있다. 발현 벡터에서, 공극을 코드화하는 핵산 서열은 전형적으로 숙주 세포에 의해 코드화 서열의 발현을 제공할 수 있는 대조 서열에 작동적으로 연결된다. 공극을 발현하기 위하여 이러한 발현 벡터가 사용될 수도 있다.
- <71> 용어 "작동적으로 연결된"이란, 기재된 성분들이 원하는 방식으로 기능할 수 있는 관계로 있는 존재하는 병렬배치(juxtaposition)를 가리킨다. 코드화 서열에 "작동적으로 연결된" 대조 서열은, 대조 서열과 친화성인 조건 하에서 코드화 서열의 발현이 달성되는 방식으로 결찰된다. 동일하거나 상이한 공극 유전자의 다수의 카피가 벡터 내로 도입될 수도 있다.

- <72> 이어서 발현 벡터가 적절한 숙주 세포 내로 도입될 수도 있다. 따라서, 공극을 코드화하는 핵산 서열을 발현 벡터 내에 도입하고, 벡터를 친화성 세균 숙주 세포 내에 도입하고, 공극을 코드화하는 핵산 서열의 발현을 일으키는 조건 하에서 숙주 세포를 생육시킴으로써 생성되는 세포에서 본 발명의 방법을 수행할 수도 있다. 대안적으로, 이러한 방식으로 생성된 재조합 공극이 세균 숙주 세포로부터 단리되고 다른 막으로 삽입될 수도 있다.
- <73> 벡터는 예를 들어 복제 기원, 임의로 핵산 서열의 발현을 위한 프로모터 및 임의로 프로포터의 조절인자와 함께 제공된 플라스미드, 바이러스 또는 과지 벡터일 수도 있다. 벡터는 하나 이상의 선택가능한 마커 유전자, 예를 들어 테트라사이클린 내성 유전자를 함유할 수도 있다. 발현 벡터가 고안되는 숙주 세포와 친화성이 되도록 프로모터 및 다른 발현 조절 시그널이 선택될 수도 있다. T7, trc, lac, ara 또는 λ_L프로모터가 전형적으로 사용된다.
- <74> 숙주 세포는 전형적으로 높은 수준으로 공극을 발현한다. 세포를 형질전환하기 위해 사용된 발현 벡터와 친화성이 되도록, 공극을 코드화하는 핵산 서열로 형질전환된 숙주 세포를 선택한다. 숙주 세포는 전형적으로 세균 성이고 바람직하게는 에스케리키아 콜리이다. λ DE3 라이소젠, 예를 들어 C41 (DE 3), BL21 (DE3), JM109(DE3), B834 (DE3), TUNER, 오리가미(Origami) 및 오리가미 B를 가진 세포를 T7 프로모터를 포함한 벡터를 발현할 수 있다.
- <75> 공극-생성 유기체로부터 단백질 액체 크로마토그래피 시스템에 의해 정제 후에 또는 상기 기재된 것과 같은 재조합 발현 후에 대규모로 공극을 제조할 수 있다. 전형적인 단백질 액체 크로마토그래피 시스템은 FPLC, AKTA 시스템, 바이오-캐드 시스템, 바이오-래드 바이오로직(Bio-Rad BioLogic) 시스템 및 길슨(Gilson) HPLC 시스템을 포함한다. 본 발명에 따른 사용을 위하여 천연 또는 재조합-생성된 공극을 천연 또는 인공 막 내에 삽입할 수 있다.
- <76> 본 발명의 방법은 상기 기재된 공극의 어느 하나를 사용할 수도 있다.
- <77> **공극과 뉴클레오티드 간의 상호작용**
- <78> 뉴클레오티드는 막의 어느 한 면 위에서 공극과 접촉될 수 있다. 뉴클레오티드는 막의 어느 한 면 위에서 공극에 도입될 수도 있다. 뉴클레오티드는 바람직하게는 이온이 공극에 들어가서 전기화학 구배를 따라 막을 통해 흐르도록 막의 한 면 위에서 공극과 접촉된다. 뉴클레오티드는 바람직하게는 뉴클레오티드가 공극을 통해 막의 다른쪽 면으로 통과되도록 하는 막의 한 면과 접촉된다. 예를 들어, 뉴클레오티드는 원래의 환경에서 이온 또는 작은 분자, 예컨대 뉴클레오티드가 공극의 통 또는 통로 안으로 들어가서 이온 또는 작은 분자가 공극을 통해 통과될 수 있도록 공극의 말단과 접촉된다.
- <79> 뉴클레오티드는 어떠한 방식으로 어떠한 부위에서 공극과 상호작용할 수도 있다. 뉴클레오티드는 바람직하게는 공극에 가역적으로 결합된다. 뉴클레오티드는 더욱 바람직하게는 공극의 통 또는 통로에 가역적으로 결합한다. 뉴클레오티드는 막을 거쳐 공극을 통해 통과될 때 가장 바람직하게는 공극의 통로 또는 통에 가역적으로 결합한다.
- <80> 뉴클레오티드와 공극 간의 상호작용 동안에, 뉴클레오티드가 그 뉴클레오티드에 특이적인 방식으로 공극을 통해 전류가 흐르는데 영향을 미친다. 예를 들어, 특정한 뉴클레오티드는 특정한 기간 동안 특정한 정도까지 공극을 통해 흐르는 전류를 감소시킨다. 특정한 뉴클레오티드가 공극을 통해 흐르는 전류에 미치는 영향을 결정하기 위하여 대조 실험을 수행할 수도 있다. 샘플에서 특정한 뉴클레오티드를 식별하거나 또는 특정한 뉴클레오티드가 샘플에 존재하는지의 여부를 결정하기 위하여, 시험 샘플에 대해 본 발명의 방법을 수행하는 것으로부터의 결과를, 대조 실험으로부터 유래된 결과와 비교할 수 있다. 샘플 내의 뉴클레오티드의 농도를 결정하기 위하여, 특정한 뉴클레오티드를 나타내는 방식으로 공극을 통해 흐르는 전류가 영향을 미치는 주파수를 사용할 수 있다.
- <81> **장치**
- <82> 막횡단 단백질 공극이 막 안으로 삽입되는 막/공극 시스템을 조사하기 위해 적절한 장치를 사용하여 방법을 수행할 수도 있다. 확률적 탐지를 위해 적절한 장치를 사용하여 방법을 수행할 수도 있다. 예를 들어, 장치는 수용액을 포함한 챔버 및 챔버를 2개의 구역으로 분리하는 장벽을 포함한다. 장벽은 공극을 포함한 막이 형성되는 구멍을 갖는다. 뉴클레오티드를 챔버 내에 도입함으로써 뉴클레오티드를 공극과 접촉할 수도 있다. 뉴클레오티드를 챔버의 2개 구획의 어느 하나에 도입할 수도 있다.
- <83> 본 발명의 방법은 뉴클레오티드와의 상호작용 동안에 공극을 통과하는 전류를 측정하는 단계를 포함한다. 따라

서, 장치는 막 및 공극을 거쳐 전기 신호를 적용하고 측정할 수 있는 전기 회로를 포함한다. 패치 클램프 또는 전압 클램프를 사용하여 방법을 수행할 수도 있다. 방법은 바람직하게는 패치 클램프의 사용과 연관된다. 실시예는 패치 클램프 방법을 수행하는 한가지 방법을 개시하고 있다.

<84>

분자 연결기

<85>

막횡단 공극은 바람직하게는 공극과 뉴클레오티드 간의 상호작용을 촉진시키는 분자 연결기를 포함한다. 연결기는 전형적으로 뉴클레오티드와의 상호작용을 개선하는 공극의 물리적 또는 화학적 성질에 영향을 미친다. 연결기는 전형적으로 공극의 통 또는 통로의 하전을 변경시키거나 뉴클레오티드와 특이적으로 상호작용하거나 결합하고, 이에 의해 공극과의 상호작용을 촉진한다. 연결기는 바람직하게는 뉴클레오티드 위의 하나 이상의 포스페이트 기와 상호작용하거나 p-양이온 상호작용에 의하여 뉴클레오티드에서의 염기와 상호작용한다. 연결기는 뉴클레오티드와 공극 간의 상호작용을 매개할 수도 있다. 예를 들어, 뉴클레오티드는 연결기를 통해 공극에 가역적으로 결합할 수도 있다. 대안적으로, 연결기는 공극과 함께 뉴클레오티드와 상호작용할 수도 있다. 예를 들어, 뉴클레오티드는 공극과 연결기 양쪽 모두에 가역적으로 결합할 수도 있다. 연결기는 이들이 뉴클레오티드와 상호작용할 수 있도록 바람직하게는 통 또는 통로를 수축시킨다.

<86>

연결기 자체는 공극과 가역적으로 상호작용할 수도 있고 따라서 공극의 통 또는 통로 안 및 밖으로 이동할 수도 있다. 대안적으로, 연결기는 이들이 떠날 수 없는 공극의 통 또는 통로에 공유 결합될 수도 있다.

<87>

연결기는 전형적으로 아미노기의 고리를 갖는다. 연결기는 바람직하게는 7개 아미노 기의 고리를 갖는다. 아미노기의 고리는 공극의 통 또는 통로의 수축 부분에서 양 전하 아미노산의 고리와 조합되어 뉴클레오티드와 상호작용할 수도 있다.

<88>

하나의 적절한 연결기는 시클로덱스트린이다. 연결기는 바람직하게는 헵타키스-6-아미노- β -시클로덱스트린 ($\text{am}_7\text{-}\beta\text{-CD}$)이다.

<89>

뉴클레오티드

<90>

뉴클레오티드를 식별하기 위하여 본 발명의 방법이 사용될 수 있다. 뉴클레오티드는 천연 또는 인공일 수 있다. 뉴클레오티드는 전형적으로 뉴클레오염기, 당 및 적어도 하나의 포스페이트 기를 함유한다. 뉴클레오염기는 전형적으로 혜데로고리이다. 적절한 뉴클레오염기는 푸린 및 페리미딘을 포함하고, 더욱 구체적으로 아데닌, 구아닌, 티민, 우라실 및 시토신을 포함한다. 당은 전형적으로 펜토스 당이다. 적절한 당은 이에 한정되지 않지만 리보스 및 데옥시리보스를 포함한다. 뉴클레오티드는 전형적으로 리보뉴클레오티드 또는 데옥시리보뉴클레오티드이다. 뉴클레오티드는 전형적으로 모노포스페이트, 디포스페이트 또는 트리포스페이트를 포함한다.

<91>

적절한 뉴클레오티드는 이에 한정되지 않지만 아데노신 모노포스페이트(AMP), 아데노신 디포스페이트 (ADP), 아데노신 트리포스페이트(ATP), 구아노신 모노포스페이트(GMP), 구아노신 디포스페이트(GDP), 구아노신 트리포스페이트 (GTP), 티미딘 모노포스페이트(TMP), 티미딘 디포스페이트(TDP), 티미딘 트리포스페이트(TTP), 우리딘 모노포스페이트(UMP), 우리딘 디포스페이트(UDP), 우리딘 트리포스페이트(UTP), 시티딘 모노포스페이트(CMP), 시티딘 디포스페이트(CDP), 시티딘 트리포스페이트(CTP), 시클릭 아데노신 모노포스페이트(cAMP), 시클릭 구아노신 모노포스페이트(cGMP), 데옥시아데노신 모노포스페이트(dAMP), 데옥시아데노신 디포스페이트(dADP), 데옥시아데노신 트리포스페이트(dATP), 데옥시구아노신 모노포스페이트(dGMP), 데옥시구아노신 디포스페이트 (dGDP), 데옥시구아노신 트리포스페이트(dGTP), 데옥시티미딘 모노포스페이트 (dTTP), 데옥시티미딘 디포스페이트(dTDP), 데옥시티미딘 트리포스페이트(dTTP), 데옥시우리딘 모노포스페이트 (dUMP), 데옥시우리딘 디포스페이트(dUDP), 데옥시우리딘 트리포스페이트(dUTP), 데옥시시티딘 모노포스페이트(dCMP), 데옥시시티딘 디포스페이트(dCDP) 및 데옥시시티딘 트리포스페이트(dCTP)를 포함한다. 뉴클레오티드는 바람직하게는 AMP, TMP, GMP, UMP, dAMP, dTMP, dGMP 또는 dCMP이다.

<92>

뉴클레오티드는 리보핵산(RNA) 또는 데옥시리보핵산과 같은 핵산 서열의 소화로부터 유래될 수도 있다. 단일 핵산 서열로부터의 개별 뉴클레오티드는 핵산의 전체 또는 일부를 서열화하기 위하여 연속적인 방식으로 공극과 접촉될 수도 있다. 본 발명의 두 번째 구현양태에 따른 핵산의 서열화는 이하에서 더욱 상세히 설명된다.

<93>

예컨대 뉴클레오티드가 핵산 서열의 소화로부터 유래될 때, 뉴클레오티드는 전형적으로 변형되지 않는다. 대안적으로, 뉴클레오티드가 변형되거나 손상될 수도 있다. 뉴클레오티드는 전형적으로 메틸화된다. 뉴클레오티드는 드러나는 표지로 표지화될 수도 있다. 드러나는 표지는 뉴클레오티드가 검출될 수 있도록 하기에 적절한 표

지일 수도 있다. 적절한 표지는 형광 분자, 방사성 동위원소, 예를 들어 ^{125}I , ^{35}S , 및 비오틴과 같은 링커를 포함한다.

<94> 뉴클레오티드는 전형적으로 적절한 생물학적 샘플에 존재한다. 본 발명은 전형적으로 하나 이상의 뉴클레오티드를 함유하는 것으로 알려져 있거나 함유하는 것으로 추측되는 샘플에서 수행된다. 본 발명은 정체가 알려져 있지 않은 하나 이상의 뉴클레오티드를 함유하는 샘플에서 수행될 수도 있다. 대안적으로, 본 발명은 샘플 중에 존재하는 것으로 알려져 있거나 예상되는 하나 이상의 뉴클레오티드의 정체를 확인하기 위해 샘플에서 수행될 수도 있다. 본 발명은 유기체 또는 미생물로부터 수득되거나 추출되는 샘플에서 시험관내에서 수행될 수 있다. 유기체 또는 미생물은 전형적으로 원핵 또는 진핵이고, 전형적으로 5개 계 중의 하나에 속한다: 식물, 동물, 진균, 모네라 및 원생생물. 본 발명은 바이러스로부터 수득되거나 추출된 샘플에서 시험관내에서 수행될 수도 있다. 샘플은 바람직하게는 유체 샘플이다. 샘플은 전형적으로 환자의 체액을 포함한다. 샘플은 뇨, 림프, 타액, 점액 또는 양막액일 수도 있고, 바람직하게는 혈액, 혈장 또는 혈청이다. 전형적으로, 샘플은 인간 기원이지만, 대안적으로 이것은 다른 포유동물, 예컨대 말, 소, 양, 또는 돼지와 같은 통상적인 농장 동물로부터 유래될 수도 있거나, 또는 대안적으로 고양이 또는 개와 같은 애완동물로부터 유래될 수도 있다.

<95> 샘플은 전형적으로 분석 이전에, 예를 들어 원심분리에 의해 처리되거나, 또는 원하지 않는 분자 또는 세포, 예컨대 적혈구를 걸러내는 막을 통해 통과시킴으로써 처리된다. 샘플을 취한 후 즉시 측정할 수도 있다. 또한, 샘플을 전형적으로 분석 이전에 바람직하게는 $-70\text{ }^{\circ}\text{C}$ 미만에서 보관할 수도 있다.

조건

<97> 본 발명의 방법은 뉴클레오티드와의 상호작용 동안에 공극을 통과하는 전류를 측정하는 단계를 포함한다. 막횡단 단백질 공극을 통한 이온 전류를 측정하기 위해 적절한 조건은 당 기술분야에 공지되어 있고 실시예에 개시되어 있다. 막 및 공극을 가로질러 적용된 전압으로 방법을 수행한다. 사용된 전압은 전형적으로 +50 mV 내지 +200 mV이다. 사용된 전압은 바람직하게는 +70 mV 내지 +150 mV, +85 mV 내지 +145 mV, 또는 +100 mV 내지 +140 mV이다. 사용된 전압은 바람직하게는 데옥시-리보 뉴클레오티드 5' 모노포스페이트, 예컨대 dAMP, dTMP, dGMP 및 dCMP에 대해 약 +130 mV이고, 리보뉴클레오티드 5' 모노포스페이트, 예컨대 AMP, TMP, GMP 및 UMP에 대해 +110 mV이다.

<98> 알칼리 금속 클로라이드 염의 존재하에서 방법을 수행한다. 상기 언급된 일례의 장치에서, 염은 챔버에서 수용액으로 존재한다. 염화칼륨(KCl), 염화나트륨(NaCl) 또는 염화세슘(CsCl)이 전형적으로 사용된다. KCl이 바람직하다. 염 농도는 전형적으로 0.1 내지 2M, 0.3 내지 1.9M, 0.5 내지 1.8M, 0.7 내지 1.7M, 0.9 내지 1.6M, 또는 1M 내지 1.4M이다. 염 농도는 바람직하게는 약 1M이다.

<99> 방법은 전형적으로 완충액의 존재하에 수행된다. 상기 언급된 일례의 장치에서, 완충액은 챔버에서 수용액으로 존재한다. 본 발명의 방법에서 어떠한 완충액이라도 사용될 수 있다. 하나의 적절한 완충액은 트리스-HCl 완충액이다. 방법은 전형적으로 7.5 내지 12.0, 7.6 내지 11.0, 7.7 내지 10.0, 7.8 내지 9.5, 8.0 내지 9.0, 또는 8.0 내지 8.5의 pH에서 수행된다. 사용된 pH는 바람직하게는 약 8.0이다.

<100> 방법은 전형적으로 14 $^{\circ}\text{C}$ 내지 100 $^{\circ}\text{C}$, 15 $^{\circ}\text{C}$ 내지 90 $^{\circ}\text{C}$, 16 $^{\circ}\text{C}$ 내지 80 $^{\circ}\text{C}$, 17 $^{\circ}\text{C}$ 내지 70 $^{\circ}\text{C}$, 18 $^{\circ}\text{C}$ 내지 60 $^{\circ}\text{C}$, 19 $^{\circ}\text{C}$ 내지 50 $^{\circ}\text{C}$, 또는 20 $^{\circ}\text{C}$ 내지 40 $^{\circ}\text{C}$ 의 온도에서 수행된다. 방법은 전형적으로 실온에서 수행된다.

<101> 방법은 dAMP, dTMP, dGMP 및 dCMP와 같은 데옥시-리보 뉴클레오티드 5' 모노포스페이트에 대해 pH 8.0, 1M KCl에서 +130 mV에서, 그리고 AMP, TMP, GMP 및 UMP와 같은 리보 뉴클레오티드 5' 모노포스페이트에 대해 pH 8.0, 1M KCl에서 +110 mV에서 바람직하게 수행된다.

핵산의 서열화 방법

<102> 두 번째 구현양태에서, 본 발명은 (a) 진행성 엑소뉴클레아제를 사용하여 표적 핵산 서열의 한쪽 말단으로부터 개별 뉴클레오티드를 소화시키는 단계; (b) 상기 뉴클레오티드를 막횡단 단백질 공극과 접촉시켜 상기 뉴클레오티드가 상기 공극과 상호작용하도록 하는 단계; (c) 상기 상호작용 동안에 공극을 통해 통과하는 전류를 측정하여 상기 뉴클레오티드의 정체를 결정하는 단계; 및 (d) 핵산 서열의 동일한 말단에서 단계 (a) 내지 (c)를 반복하여 핵산의 서열을 결정하는 단계를 포함하는, 표적 핵산 서열의 서열화 방법에 관한 것이다. 따라서, 두 번째 구현양태는 핵산의 서열화를 위하여 연속적인 방식으로 핵산 서열의 단일 뉴클레오티드의 확률적 탐지와 연관된다. 두 번째 구현양태의 방법을 사용하여 핵산의 전체 또는 일부를 서열화할 수도 있다. 핵산은 천연 또

는 인공일 수도 있다. 예를 들어, 제조된 올리고뉴클레오티드의 서열을 입증하기 위하여 두 번째 구현양태의 방법을 사용할 수도 있다. 두 번째 구현양태의 방법을 시험관내에서 전형적으로 수행한다.

<104> 두 번째 구현양태의 방법의 단계 (b) 및 (c)는 상기 언급된 첫 번째 구현양태의 방법에서 수행되는 단계와 일반적으로 동일하다. 첫 번째 구현양태에 관한, 특히 첫 번째 구현양태에서 사용될 수도 있는 막, 장치, 공극, 분자 연결기, 뉴클레오티드 및 조건에 관한 상기 언급 내용의 전부는 두 번째 구현양태에도 동일하게 적용된다. 두 번째 구현양태에서 핵산은 전형적으로 첫 번째 구현양태를 위해 상기 언급된 생물학적 샘플에서 전형적으로 존재한다. 서열이 알려져 있지 않은 하나 이상의 핵산을 함유하는 샘플에서 두 번째 구현양태의 방법을 수행할 수도 있다. 대안적으로, 샘플에서의 존재가 알려져 있거나 예상되는 핵산의 정체를 입증하기 위하여 샘플에서 두 번째 구현양태의 방법을 수행할 수도 있다. 두 번째 구현양태의 방법을 사용하여 서열화하기 이전에 핵산 서열을 전형적으로 증폭시킨다.

진행성 엑소뉴클레아제

<106> 두 번째 구현양태의 방법은 핵산 서열을 진행성 엑소뉴클레아제와 접촉시켜 핵산의 한쪽 말단으로부터 개별 뉴클레오티드를 방출시키는 것을 포함한다. 진행성 엑소뉴클레아제는 전형적으로 핵산 서열의 한쪽 말단에 매달리는 효소이고 한쪽 말단으로부터 한번에 하나의 뉴클레오티드씩 서열을 소화시킨다. 진행성 엑소뉴클레아제는 5'에서 3' 방향으로 또는 3'에서 5' 방향으로 핵산을 소화시킬 수 있다. 당 기술분야에서 사용된 효소를 선택하고/하거나 당 기술분야에 공지된 방법을 사용하여 진행성 엑소뉴클레아제가 결합하는 핵산의 말단이 전형적으로 결정된다. 핵산 서열의 특정한 말단에 진행성 엑소뉴클레아제의 결합을 막거나 촉진하기 위하여 핵산 서열의 한쪽 말단에서 히드록실 기 또는 캡 구조가 전형적으로 사용될 수도 있다.

<107> 본 발명의 방법에서 진행성 엑소뉴클레아제 효소가 사용될 수도 있다. 본 발명의 방법에서 사용하기 위해 바람직한 효소는 람다 엑소뉴클레아제이다. 람다 엑소뉴클레아제의 하나의 소단위의 서열을 서열 9에 나타낸다. 3 개의 동일한 소단위가 상호작용하여 삼량체 엑소뉴클레아제를 형성한다. 람다 엑소뉴클레아제의 변형체는 서열 9의 서열로부터 변경된 아미노산 서열을 갖고 진행성 엑소뉴클레아제 활성을 보유하는 폴리펩티드 소단위로 형성된 효소이다. 변형체는 상기 서열 2의 변형체에 대해 언급된 것과 동일한 정도로 동일한 방식으로 서열 9로부터 변형될 수도 있다. 변형체는 바람직하게는 핵산에 결합하고 핵산을 소화시키는 원인이 되는 도메인을 포함한다 (촉매적 도메인). 변형체는 바람직하게는 야생형 효소에 비해 감소된 속도의 효소 활성 및/또는 높은 염 내성을 갖는다. 막횡단 단백질 공극의 생성을 위해 상기 언급된 어느 방법을 사용하여 진행성 엑소뉴클레아제가 생성될 수도 있다.

<108> 두 번째 구현양태의 방법은 본 발명의 첫 번째 구현양태에 따라 개별 뉴클레오티드의 식별을 가능하게 하는 속도로 핵산의 말단으로부터 뉴클레오티드가 소화되도록 핵산 서열을 진행성 엑소뉴클레아제와 접촉시키는 것을 포함한다. 이것을 수행하기 위한 방법이 당 기술분야에 공지되어 있다. 예를 들어, 고성능 액체 크로마토그래피(HPLC)를 사용하여 식별될 수 있도록 폴리펩티드의 말단으로부터 하나의 아미노산을 연속적으로 소화시키기 위해 에드만(Edman) 소화가 사용된다. 본 발명에서 상동성 방법이 사용될 수도 있다.

<109> 진행성 엑소뉴클레아제가 바람직하게는 막횡단 단백질 공극에 공유 부착된다. 공극에 대한 진행성 엑소뉴클레아제의 공유 부착 방법이 당 기술분야에 공지되어 있다.

<110> 두 번째 구현양태의 방법에서 진행성 엑소뉴클레아제가 작용해야 하는 속도는 전형적으로 야생형 진행성 엑소뉴클레아제의 최적 속도보다 더 느린다. 두 번째 구현양태의 방법에서 진행성 엑소뉴클레아제의 적절한 활성 속도는 0.5 내지 1000 뉴클레오티드/초, 0.6 내지 500 뉴클레오티드/초, 0.7 내지 200 뉴클레오티드/초, 0.8 내지 100 뉴클레오티드/초, 0.9 내지 50 뉴클레오티드/초 또는 1 내지 20 또는 10 뉴클레오티드/초의 소화를 포함한다. 속도는 바람직하게는 1, 10, 100, 500 또는 1000 뉴클레오티드/초이다. 진행성 엑소뉴클레아제 활성의 적절한 속도는 다양한 방식으로 달성될 수 있다. 예를 들어, 본 발명에 따르면, 감소된 최적 활성 속도를 가진 변형체 진행성 엑소뉴클레아제가 사용될 수도 있다.

<111> 진행성 엑소뉴클레아제의 활성은 전형적으로 pH 의존성이고, 따라서 pH가 감소됨에 따라 그들의 활성이 저하된다. 따라서, 두 번째 구현양태의 방법은 전형적으로 pH 7.5 내지 8.0, 또는 7.7 내지 8.0에서 수행된다. 사용된 pH는 바람직하게는 약 8.0이다.

<112> 진행성 엑소뉴클레아제의 활성 속도는 전형적으로 염 농도가 상승함에 따라 저하된다. 그러나, 매우 높은 염 농도는 전형적으로 효소의 활성에 해로운 효과를 미친다. 효소의 속도를 제한하는 다른 방법은 활성에 악영향을 미치지 않으면서 효소의 활성 속도를 감소시키는 염 농도에서 두 번째 구현양태의 방법을 수행하는 것이다.

예를 들어, 0.5 내지 1M의 염 농도에서 두 번째 구현양태의 방법을 수행할 수도 있다. 염 농도는 바람직하게는 약 1M이다.

<113> 키트

세 번째 구현양태에서, 본 발명은 본 발명의 두 번째 구현양태를 수행하기 위해 사용될 수 있는 키트에 관한 것이다. 따라서, 핵산을 서열화하기 위하여 키트가 적절하다. 키트는 시클로텍스트린 및 진행성 엑소뉴클레아제를 포함한다. 시클로텍스트린은 바람직하게는 헵타키스-6-아미노- β -시클로텍스트린이다. 진행성 엑소뉴클레아제는 두 번째 구현양태에 관해 상기 언급된 어느 것일 수도 있다. 키트는 바람직하게는 막횡단 단백질 공극을 더 포함한다. 공극은 첫 번째 구현양태에 관해 상기 언급된 것의 어느 것일 수도 있다.

<115> 키트는 추가로 상기 언급된 방법의 어느 구현양태라도 수행할 수 있도록 하는 하나 이상의 다른 시약 또는 장치를 포함할 수도 있다. 이러한 시약 또는 장치는 다음 중의 하나 이상을 포함한다: 적절한 완충액(들) (수용액), 환자로부터 샘플을 수득하기 위한 수단 (예컨대 바늘을 포함한 용기 또는 기구), 핵산 서열을 증폭시키기 위한 수단, 상기 정의된 막 또는 전압 또는 패치 클램프 장치. 시약은 유체 샘플이 시약을 재현탁시키도록 건조 상태에서 키트에 존재할 수도 있다. 키트는 임의로 본 발명의 방법에서 키트가 사용될 수 있도록 하는 지시 또는 환자를 위해 사용될 수 있는 방법에 관한 세부사항을 포함할 수도 있다. 키트는 임의로 뉴클레오티드를 포함할 수도 있다.

<116> 하기 실시예는 본 발명을 예증한다:

실시예

<117> α -헤모리신(M113R)₇ 변이체 (도 1A)의 이온 전도 경로의 크기를, 검출되어지는 뉴클레오티드의 크기에 더욱 가깝게 하기 위하여, 공극의 수축 부분 근처에 시클로텍스트린을 조화시킴으로써 나노공극의 직경을 감소시켰다. 주요 위치에 7개 아미노기를 가진 헵타키스-6-아미노- β -시클로텍스트린(am₇- β CB) (도 1B)가 사용되었다. 단백질 변이체 위의 위치 113에서 7개 아르기닌과 함께 시클로텍스트린이 공극 내에 있을 때 (도 1C), 한 면 위에서 7개 아미노 기의 1개 고리와 7개 아르기닌 기의 두 번째 고리가 공극을 통한 통로의 가장 좁은 부위에서 상호 간에 가장 짧은 거리 내에 존재한다. 아미노/아르기닌 고리 구조는 포스페이트 기를 가역적으로 결합하는 성질을 갖고 이에 의해 공극 내에서 XMP 및 dXMP를 5 내지 30 ms 동안 고정시킨다. 이러한 결합 사건은 전류 증폭에서의 변화를 통해 명백하게 검출할 수 있다.

<118> 1. 재료 및 방법

<119> 이전에 기재된 바와 같이 α -헤모리신 변이체(M113R)₇를 발현시키고 정제하였다 [Cheley, S., L.-Q. Gu 등 (2002), Chem & Biol. 9: 829-838].

<120> 화합물

<121> 1,2-디피타노일-sn-글리세로-3-포스포콜린은 아반티 폴라 리피드 인코포레이티드(Avanti Polar Lipids Inc.)으로부터 구입되었다. 웬탄은 JT 베이커로부터 구입되고 헥사데칸 99+%는 시그마-알드리치로부터 구입되었다. 헵타키스(6-데옥시-6-아미노)- β -시클로텍스트린 HCl>99%는 시클로랩(CYCLOLAB) Ltd. (헝가리 부다페스트)로부터 구입되었다. 2-데옥시-구아노신 5' 모노포스페이트 소듐 염 99%는 아크로스(Acros)로부터 구입되고, 2-데옥시-시토신 5' 모노포스페이트 디소듐 염>95%, 2-데옥시-티미딘 5' 모노포스페이트 디-소듐 염>97% 및 2-데옥시-아데노신 5' 모노포스페이트 디소듐 염>95%는 플루카(Fluka)로부터 구입되었다. 우리던 5' 모노포스페이트 디-소듐 염 99% 및 시토신 5' 모노포스페이트 산>98%는 플루카로부터 구입되었다. 아데노신 5' 모노포스페이트 산 99% 및 구아노신 5' 모노포스페이트 디-소듐 염 97%는 아크로스로부터 구입되었다. 트리즈마 염기 99.9%는 시그마-알드리치로부터 구입되고, 농축된 HCl 분석 시약 등급은 피셔 사이언티픽(Fisher Scientific)으로부터 구입되었다. 염화칼륨 99% 및 염화나트륨 99.9%는 시그마-알드리치로부터 구입되었다. 브롬화칼륨 99.5% 및 염화세슘 99%는 플루카로부터 구입되었다.

<122> 장치

<123> 액손 인스트루먼츠(Axon instruments)로부터의 패치 클램프 증폭기 액소패치(Axopatch) 200B를 디지데이터(Digidata) 1200 A/D 전환장치 (액손 인스트루먼츠)가 장착된 컴퓨터와 함께 사용하였다. 테플론 챔버를 사용하였다. 데이터를 p클램프 9.2에 수집하고 클램프피트(Clampfit) 9.0에서 분석하였다. 마이크로칼 오리진

(Microcal Origin) 6.0에 의해 도표 및 그래프를 수득하고 개인 계산기에서 적분을 시행하였다.

<124> 실험 조건

<125> 트랜스 및 시스 챔버를 분리시키는 20 μm 폴리카르보네이트 필름 (미국 펜실바니아 맬버른 굿펠로우 (Goodfellow)로부터의 20 μm 두께)에서 100 내지 150 μm 직경 구멍 위에 몬탈 및 뒤엘러 (1972)의 방법에 의해 1,2-디피타노일-sn-글리세로-3-포스포콜린으로부터 지질 이중 막을 형성하였다. 챔버의 시스 면은 바닥에 있고 챔버의 트랜스 면은 상부 단으로 연결되었다. 포텐셜은 트랜스 면 전극의 포텐셜 값을 가리킨다. 연결기 분자를 트랜스 면에 첨가하고, α -헤모리신 변이체 및 분석물 분자를 시스 면에 첨가하였다. dXMP 실험을 +130mV에서 수행하고, XMP 실험을 +110mV에서 수행하였다. 여기에 보고된 모든 실험은 1M KCl 중에서 pH 8.0 트리스-HCl 25mM에서 수득되었다. 매일 뉴클레오티드 용액의 새로운 분취액을 사용하였다. 다른 언급이 없는 한 실온 22.5 ± 2 °C에서 실험을 수행하였다.

<126> 2. 결과

<127> (M113R)₇/헵타키스 6 아미노 β -시클로덱스트린에 의해 형성된 2-데옥시-뉴클레오티드 5' 모노포스페이트 부분 블록 호모헵타머 공극

<128> 트랜스 면으로부터 적용된 $\text{am}_7\text{-}\beta\text{CD}$ 와 함께 $\alpha\text{HL-M113R}$ 로부터 형성된 호모-헵타머 공극에서 단일-통로 기록을 수행하였다 (도 2A 및 2B). $\text{am}_7\text{-}\beta\text{CD}$ 의 부재 하에서, 공극은 pH 8.0 트리스-HCl 25mM 완충액 중에서 1M KCl 중에서 145±5 pA (+130 mV)의 단일 전류 (L1)에 의하여 영구적으로 개방된 채로 유지되었다 (도 2A 및 2B, B). 트랜스 챔버에서 40 μM $\text{am}_7\text{-}\beta\text{CD}$ 의 첨가는 65±6 pA (도 2A, B 및 도 2B, C에서 L2)의 전류 수준까지 가역적인 차단 사건을 유도한다. 시스 챔버에 5 μM dCMP를 첨가할 때, 전류 수준 L2로부터 비롯된 22±1 pA (도 2A, C 및 도 2B, D에서 L3)에서 세 번째 전류 수준이 관찰된다. L3은 $\text{am}_7\text{-}\beta\text{CD}$ 에 의해 (M113R)₇의 착물에 대한 dCMP의 결합을 나타낸다. 챔버의 시스 면 대신에 트랜스 면으로 300 μM 이하의 dXMP를 첨가하여도 시클로덱스트린 결합 전도성 상태 (도시되지 않음)의 변경을 유도하지 않았다.

<129> 상기 기재된 실험 조건에서, $\alpha\text{HL-M113R}$ 단일 나노공극의 존재하에서 비변형 β -시클로덱스트린이 트랜스 챔버 (40 μM)에 첨가될 때, 시클로덱스트린 결합에 기인한 전류 차단 사건이 관찰되었다. 그러나, 트랜스 또는 시스 챔버 (도시되지 않음)에 dXMP (300 μM 이하)가 첨가될 때, 더 이상의 전류 차단 사건이 관찰되지 않았다.

<130> 트랜스 챔버에서 $\text{am}_7\text{-}\beta\text{CD}$ 의 부재하에서, 300 μM dGMP 또는 dTMP의 최소 농도가 시스 챔버 (도시되지 않음)에 첨가될 때 차단 사건 (<1ms)이 관찰되었다. 5 μM dXMP 또는 XMP에서 실험의 시간에 걸쳐 이러한 사건이 관찰되지 않는다.

<131> 야생형 $\alpha\text{-HL}$ 단일 통로를 통해 전류를 측정하면서 트랜스 챔버에서 $\text{am}_7\text{-}\beta\text{CD}$ 를 첨가하면 시클로덱스트린 결합 사건이 유도되지만, 시스 또는 트랜스 챔버에 dXMP를 첨가할 때 더 이상의 전류의 변경이 관찰되지 않았다.

<132> (M113R)₇/헵타키스 6 아미노 β -시클로덱스트린에 의해 형성된 호모헵타머 공극의 부분 블록의 진폭으로부터 2-데옥시-뉴클레오티드 5' 모노포스페이트가 식별될 수 있다.

<133> 일시적 착물 (M113R)₇/ $\text{am}_7\text{-}\beta\text{CD}$ 의 부분 블록은 dXMP가 시스 챔버에 첨가되었는지에 의존하여 진폭에서 차이를 나타내었다 (도 3). 시스 면으로 dGMP (5 μM)를 첨가하면 16pA의 전류 수준에 대해 평균 차단을 나타내었다 (도 3A). 도 3A에 나타낸 결과의 모든 점 진폭 히스토그램은 dGMP의 구조와 함께 트레이스의 오른쪽에 나타낸다. 다른 뉴클레오티드들은 모두 dTMP에 대해 도 3B, dAMP에 대해 3C, 및 dCMP에 대해 3D에 나타낸 것과 같이 상이한 진폭을 나타낸다. 4개의 dXMP 중에서 dGMP가 대부분의 전류를 차단한다.

<134> 독립적인 실험으로부터의 전류 진폭은 단백질 나노공극의 특성으로부터 비롯된 일부 변형을 나타내었다. +130 mV에서 (M113R)₇에 대해 평균 전류는 139 pA이지만, 일부 통로는 147 pA 만큼 높고 131 pA 만큼 낮은 전류를 나타낸다. 따라서, 전류 트레이스를 비교하기 위하여 0 전류와 65 pA로 설정된 (M113R)₇/ $\text{am}_7\text{-}\beta\text{CD}$ 전류 수준 간에 상이한 실험으로부터 전류 트레이스를 표준화하였다.

<135> 3개의 독립적인 실험으로부터 500회 사건에 걸쳐 각각의 dXMP의 체류 시간 (τ_{off})을 계산하였다 (표 1).

표 1

<136>

3개의 독립적 측정으로부터 평균을 낸 dGMP, dTMP, dAMP 및 dCMP의 체류 시간(ms)				
	dGMP	dTMP	dAMP	dCMP
τ_{off} (ms)	9.8±0.2	19.8±0.8	7.1±0.2	10.5±0.4

<137>

시클로덱스트린 전류 수준

<138>

pH 8.0에서, 변이체 ($M113R_7$)은 단백질 통로가 비어있을 때 2개의 전류 수준 $L1/L1'$ 를 나타낸다 (도 4). 시클로덱스트린 연결기는 단백질이 존재하는 전류 수준 $L1/L1'$ 와 무관하게 단백질에 결합할 수 있다.

<139>

($M113R_7$) 나노공극의 전류 수준을 기록할 때 2개의 전류 수준이 관찰되었다 (도 4). 도 4의 삽입에 나타낸 것과 같이 $L1$ 은 주된 전류 수준이다. 나노공극에 대한 $am_7-\beta CD$ 의 결합은 $L2$ 및 $L2'$ 에 의해 표시되는 2개의 전류 수준을 유도한다 (3개의 수준이 pH 7.5에서 관찰된다. 도시되지 않음). 단백질 나노공극에 대한 $am_7-\beta CD$ 의 결합은 전류 수준 $L1$ 또는 $L1'$ 과 독립적으로 발생한다. $L2$ 는 $am_7-\beta CD$ 가 ($M113R_7$)에 결합될 때 관찰되는 주된 전도성 수준이고, 이것은 명백한 상관관계를 갖지 않은 빈 나노공극의 $L1$ 및 $L1'$ 전도성 수준으로부터 비롯된다. $am_7-\beta CD$ 이 ($M113R_7$)에 결합될 때 관찰되는 전류 수준 $L2'$ 은 시클로덱스트린 연결기가 결합될 때의 전도성의 15 % 미만에 이르른다 (도 4의 삽입 참조).

<140>

뉴클레오티드 결합 사건은 때때로 이들이 비롯되는 전류 수준 $L2$ 및 $L2'$ 의 결과로서 진폭을 변화시킨다. dXMP 결합 사건이 비롯되는 $L2$ 또는 $L2'$ 에 의존하여 0.5pA 범위가 관찰되었다 (도시되지 않음). 이것은 뉴클레오티드 결합 사건 히스토그램의 겹침을 증가시킨다 (도 5).

<141>

도 3에 기재된 결합된 시클로덱스트린 전류 수준 $L2'$ 로부터 유래된 각각의 분석물 결합 사건을 제거하기 위하여 각각의 기록을 수동으로 분석할 수 있다. 도 5는 변형되지 않은 dXMP 검출 전류 기록으로부터 수득된 단일 사건 분석 히스토그램 (도 5A)과 수준 $L2'$ (도 4)로부터 유래하는 분석물 결합 사건이 제거되어진 동일한 기록 간의 차이를 나타낸다. 2개의 히스토그램은 dGMP, dTMP, dAMP 및 dCMP에 상응하는 것과 동일한 4개의 피크를 나타낸다. $L2'$ 로부터 유래된 분석물 결합 사건이 제거되었기 때문에 도 5A에서 피크의 진폭이 5B에서보다 크고, 따라서 히스토그램이 몇몇 사건으로부터 비롯된다. 각각의 피크 간의 분리는 5A에서보다 5B에서 더 양호한 것으로 보인다. 그러나, 기록으로부터 이러한 사건을 제거하면 각각의 피크가 완전히 분리되지 않는다 (도 5B). 그 결과, 도 4에 나타낸 시클로덱스트린 전류 수준 $L2$ 및 $L2'$ 는 단일 사건 분석 히스토그램에서 고려되지 않고 이후에 통계가 기록된다.

<142>

일시적 착물 ($M113R_7$)/헵타키스 6 아미노 β -시클로덱스트린에 의해 형성된 호모헵타머 공극의 부분 블록의 진폭으로부터 2-데옥시 뉴클레오티드 5' 모노포스페이트가 식별될 수 있다.

<143>

($M113R_7$)/ $am_7-\beta CD$ 공극의 부분 블록은 dXMP가 시스 챔버에 첨가되는지에 의존하여 진폭에서 상이한 것으로 입증되었다. dGMP, dTMP, dAMP 및 dCMP가 시스 챔버에 동시에 첨가될 때 상이한 진폭이 분해될 수 있다 (도 5).

<144>

도 6A는 단일 실험으로부터 모든 4개의 뉴클레오티드의 혼합 용액에 대한 전류 진폭을 나타낸다. 각각의 dXMP의 진폭 분포를 도시하기 위하여 색을 가진 띠를 기록된 전류 트레이스 위에 겹친다. 도 6B는 각각의 뉴클레오티드에 의해 발생된 진폭 분포인 것으로 추측되는 8000개 사건의 전류 트레이스의 진폭 히스토그램을 나타낸다. 진폭 히스토그램을 가우시안 분포와 겹친다. 이러한 실험에 의해 주어진 피크 전류 값으로부터, 각각의 뉴클레오티드의 분포를 독립적으로 조화시키고 평균화하는 것으로부터 수득되는 분포 평균 값 및 σ 값으로서 적합도를 수득한다. 전류 트레이스를 가우시안 분포와 조화시키는 것으로부터 각각의 뉴클레오티드에 대한 식별 가능성이 입증되었다.

<145>

통계적 방법

<146>

트랜스 면 위에서 $am_7-\beta CD$ 의 존재 하에서 ($M113R_7$)의 전류 트레이스와 시스 면 위에서 분석물 뉴클레오티드의 전류 트레이스를 300 Hz (저 통과 가우시안 필터)에서 디지털 여과하고, 모든 점 진폭 히스토그램을 구축하였다. 이러한 히스토그램은 $am_7-\beta CD$ 가 ($M113R_7$) α -헵토리신 변이체에 결합될 때 관찰되는 전류 진폭에 상응하는 큰 피크를 나타낸다 (도 1에서 L2에 상응). 이러한 전류 진폭은 하나의 실험으로부터 다른 실험 간에

단백질 통로 사이에서 5% 이내로 변경된다. 이러한 이유 때문에, 모든 점 진폭 히스토그램을 0 전류와 65pA에서 설정된 주요 시클로텍스트린 피크 사이에서 표준화하였다. 표준화된 히스토그램에서, 뉴클레오티드 피크가 가우시안 분포에 조화되었다. 각각 1000개 사건을 함유하는 적어도 3개의 독립적인 실험으로부터 동일한 뉴클레오티드의 평균 및 시그마 (σ) 값을 평균을 내었다 (표 2에 기재된 값).

표 2

0과 65pA 사이에서 표준화된 3개의 독립적 실험으로부터 각각의 뉴클레오티드의 평균 분포 값	
	평균
G	16.0, $\sigma=0.64$
T	17.4, $\sigma=0.41$
A	18.4, $\sigma=0.54$
C	20.0, $\sigma=0.51$

<147> 각각의 트레이스에서 적어도 3000개 뉴클레오티드 결합 사건에 걸쳐 동시에 존재하는 모든 4개의 뉴클레오티드를 가진 실험으로부터 각각의 염기의 관독에 대한 가능성을 결정하였다. 트레이스를 여과하고 (300 Hz 저 통과 가우시안 디지털 필터) 개별 뉴클레오티드 실험에 대해 상기 기재된 바와 같이 0과 65pA에서의 시클로텍스트린 피크 사이에서 표준화하였다. 각각의 뉴클레오티드의 피크 값을 5개의 독립적인 실험으로부터 평균내었다 (표 3).

표 3

5개 독립적 실험으로부터 모든 4개 뉴클레오티드가 존재하는 각각의 뉴클레오티드의 피크 값. 마지막 세로줄은 가우시안 분포의 겹침이 적분되는 각각의 피크의 평균값을 나타낸다.	
	평균
G(pA)	16.2±0.5
T(pA)	17.6±0.6
A(pA)	18.6±0.6
C(pA)	20.2±0.5

<150> 모든 4개의 뉴클레오티드가 존재하는 실험에서, 모든 4개의 뉴클레오티드로부터의 가우시안은 겹침을 갖는다. 이 뉴클레오티드의 가우시안 분포와 이웃 뉴클레오티드의 가우시안 분포 간에 겹침 수준으로부터 그 자체로서 또는 다른 뉴클레오티드로서 식별되는 1개의 뉴클레오티드로부터 결합 시그널에 대해 통계를 계산하였다.

<151> 각각의 피크 위치 (혼합된 뉴클레오티드와의 실험에 의해 주어짐) 및 각각의 분포에 대한 σ 값 (개별 뉴클레오티드 실험의 조화에 의해 주어짐)으로부터 2개의 가우시안 간의 교차 점을 계산한다. 이웃하는 가우시안과의 절편 값 너머에 있는 가우시안의 면적을 적분함으로써 정확성 확률이 제공된다 (도 6, 표 4). 표 4의 첫 번째 세로줄은 나노공극과 상호작용하는 뉴클레오티드이고, 첫 번째 가로줄은 상응하는 전류 진폭으로부터 관독되는 값이다.

표 4

첨가된 뉴클레오티드(수직)가 그 자체로서 또는 다른 뉴클레오티드(수평)로서 검출될 확률				
	G관독	T관독	A관독	C관독
G첨가	0.88	0.12	0	0
T첨가	0.06	0.83	0.11	0
A첨가	0	0.19	0.74	0.07
C첨가	0	0	0.06	0.94

3. 결론

<153> 나타낸 결과는 확률적 탐지가 단일 뉴클레오티드의 식별을 위해 유망한 대안임을 암시한다. 또한, 이것은 단일

분자 수준에서 저렴하고 빠르고 간단한 DNA 서열화 방법으로서 엑소뉴클레아제 서열화를 사용할 수 있음을 의미 한다. 또한, 엑소뉴클레아제 서열화는 형광단과 같은 값비싼 시약을 필요로 하지 않기 때문에 DNA의 저렴한 서열화 방법이다.

<155> 모든 점 히스토그램은 74 내지 94% 범위의 정확도에서 각각의 뉴클레오티드를 식별하기 위해 충분한 분석 방법이다 (표 4). XMP 및 dXMP의 체류 시간 값은 각각의 분석물을 더욱 구별하기 위한 조건에서 유사하다. 진폭 히스토그램으로부터 이끌어낸 통계는 도 4에 나타낸 시클로텍스트린 전류 수준을 상쇄시킴으로써 더욱 개선될 수 있다. 각각의 dXMP 간의 전류 진폭 차이는 약 1pA이다. 이러한 분해는 다음과 같은 다수의 매개변수에 의존된다.

전압 의존성

<157> 결합 사건은 전압 의존성이다. 50mV에서, 매우 적은 결합 사건이 관찰되고, 이것은 결합 부위로 dXMP 및 XMP를 보내기 위해 최소의 자계가 요구된다는 것을 제시한다. +150 및 +200 mV에서 진폭은 더 이상 뉴클레오티드를 구별하지 못한다. +130 mV는 테옥시 리보 뉴클레오티드 5' 모노포스페이트를 위해 최상의 전압임이 입증되었고, +110 mV는 리보 뉴클레오티드 5' 모노포스페이트를 위해 최상의 분해를 일으켰다.

염 농도

<159> 트리스-HCl pH 8 완충액 0.5, 1 및 2M KCl을 시험하였다. 모든 점 진폭 히스토그램으로부터, 1M KCl에서 피크 사이의 최상의 분해가 얻어졌다.

pH 의존성

<161> 전류 진폭은 pH 의존성이고, pH 7.5, 8.0, 8.2, 8.5, 9.0 및 9.5에서 트리스-HCl 완충액 1M HCl을 시험하였다. pH 8.0 및 그 이상에서 am₇-βCD의 결합 시에 2 개의 전류 수준을 관찰하였다 (도 4). pH 7.5에서 헵타키스 6 아미노 β-시클로텍스트린은 세 번째 전류 수준을 나타낸다 (도시되지 않음). 이것은 dTMP가 상이한 진폭을 가진 2가지 유형의 사건을 나타내도록 하고, 그 중 하나는 dGMP 사건의 범위이고 이것은 분해능의 손실을 유도한다. pH 9.5에서 뉴클레오티드 결합 사건이 더 이상 관찰되지 않는다. pH 8.0에서 최상의 피크 분리가 수득된다.

염 의존성

<163> dXMP와 XMP 간의 분해도는 NaCl 또는 CsCl에 의해보다 KCl에 의해 더욱 양호하다. 1M KCl은 2M KCl보다 더욱 양호한 분해능을 가져왔다. 각각의 결합 사건이 일시적 착물 (M113R)₇/am₇-βCD의 완벽한 블록을 유도하기 때문에, KBr의 사용은 상이한 뉴클레오티드들을 식별하지 못하였다.

온도 의존성

<165> 온도를 14 °C로 낮추거나 50 °C로 높히는 것은 dXMP/XMP의 검출을 방해하지 않았다. 그러나, 이것은 진폭 히스토그램의 분해도를 개선하지 못하였다.

다른 α-헤모리신 변이체

<167> (M113N)₇은 am₇-βCD를 결합하는 것으로 보이지만 뉴클레오티드 검출이 관찰되지 않았다. (M113F)₇ 및 (M113F/147K)₇는 am₇-βCD가 첨가되었는지 아닌지의 여부를 검출하지 못하였다. (M113K)₇를 동일한 조건에서 시험하였다. 이러한 경우에, 기록은 도 1에서와 매우 유사하다. (M113K)₇ 변이체 및 am₇-βCD에 의해 뉴클레오티드 결합이 검출되지만, 뉴클레오티드 간의 피크 분리는 (M113R)₇이 사용될 때에 비해 더 작았다.

리보뉴클레오티드 5' 모노포스페이트

<169> 이 방법에 의해 XMP가 성공적으로 식별되었으며, 각각의 염기 사이의 분해능은 U, A 및 C에 대해 1 pA 미만의 피크 분리와 함께 dXMP에 비해 열악하였다. 모든 점 히스토그램 전류 진폭은 dXMP에서와 동일한 정도로 나타나고, GMP는 가장 낮은 전류를 나타내고(최대 차단) 이어서 UMP, AMP 및 CMP가 최고의 전류 진폭을 나타내었다(최소 차단). XMP 식별을 위한 최적의 전압은 pH 8.0 1M KCl에서 +110 mV인 것으로 밝혀졌다.

메카니즘

<171> (M113R)₇/am₇-βCD 일시적 착물은 글루코스 포스페이트 (글루코스 1P 및 갈락토스 1P)를 결합하고 구별하는 것으

로 밝혀졌다. 이것은 한 쪽 위의 아르기닌 고리, 포스페이트 기 및 다른 쪽 위의 $\text{am}_7\text{-}\beta\text{CD}$ 로부터의 아민 고리 간에 강한 상호작용을 제시한다. 비변형 β -시클로덱스트린은 어떠한 검출을 나타내지 않았다. XMP와 dXMP 사이에서 차이가 거의 관찰되지 않았으며, 이는 히드록실 기가 결합에서 주요한 역할을 하지 않는다는 것을 제시한다.

도면의 간단한 설명

- <20> 도 1은 α -헤모리신 (M113R_7) 변이체 및 헵타키스-6-아미노- β -시클로덱스트린($\text{am}_7\text{-}\beta\text{CD}$)를 나타낸다. A - α -헤모리신 구조를 통해 절단된 화살촉모양, 위치 113을 화살표로 나타낸다. B - $\text{am}_7\text{-}\beta\text{CD}$ 의 공간충진된 구조. C - $\text{am}_7\text{-}\beta\text{CD}$ 와 α -헤모리신 (M113R_7)의 가능한 상호작용.
- <21> 도 2A는 dCMP 검출을 나타낸다. A - +130 mV에서 인지질 이층에 삽입된 단일 (M113R_7) 변이체의 전류 트레이스. L1은 비어있는 단백질 나노공극의 전류를 나타낸다. B - 트랜스 챔버에서 $40 \mu\text{M}$ $\text{am}_7\text{-}\beta\text{CD}$ 의 존재 하. L2는 $\text{am}_7\text{-}\beta\text{CD}$ 가 나노공극 내에서 임시로 결합할 때 관찰되는 전류 수준을 나타낸다. C - dCMP $5 \mu\text{M}$ 을 시스 챔버에 첨가한다. L3은 dCMP가 임시 착물 ($\text{M113R}_7\text{/am}_7\text{-}\beta\text{CD}$)에 결합할 때 관찰되는 전류 수준을 나타낸다.
- <22> 도 2B는 α -헤모리신 (α -HL) 공극과 헵타키스-(6-데옥시-6-아미노)- β -시클로덱스트린 ($\text{am}_7\text{-}\beta\text{CD}$) 및 dCMP의 상호작용을 나타낸다. A - Met-113을 Arg로 치환한 헵타머 α HL 공극 (7AHL)의 모델. 챔드로우 울트라 (ChemDraw Ultra)에서 발생된 단면에서 $\text{am}_7\text{-}\beta\text{CD}$ 의 모델은 Arg 측쇄로부터 반 테르 바스 거리에 인력으로 위치하고, 이것은 트랜스 면으로부터 공극에 들어갈 때 시클로덱스트린의 통과를 차단한다. $\text{am}_7\text{-}\beta\text{CD}$ 가 공극 내부에 존재할 때, 양 전하의 2개 고리, 시클로덱스트린에 의해 기여된 7개 주된 아미노산의 1개 고리, 및 7개 아르기닌 측쇄의 두 번째 고리가 $\sim 10 \text{ \AA}$ 에 의해 분리된다. 아미노시클로덱스트린은 양성자화 아미노기와의 이온성 상호작용에서 뉴클레오시드 모노포스페이트를 포스페이트 기와 결합하는 것으로 앞서 밝혀졌다. 이러한 착물의 전체 안정성은 뉴클레오티드 염기와 Arg 측쇄 간의 p-양이온 상호작용에 의해 증진될 수 있다. 포스페이트 기가 $\text{am}_7\text{-}\beta\text{CD}$ 의 양성자화 아민과 상호작용하고 시토신 고리가 Arg 측쇄의 구아니디늄 기와 상호작용하도록 dCMP 분자가 위치한다. B - +130 mV에서 단일 (M113R_7) 공극으로부터의 전류 트레이스. L1은 비어있는 단백질 나노공극을 통해 흐르는 전류를 나타내고, 오른쪽에 모델로서 표시된다. C - 트랜스 챔버에 $40 \mu\text{M}$ $\text{am}_7\text{-}\beta\text{CD}$ 를 첨가한 후의 전류 트레이스. L2는 $\text{am}_7\text{-}\beta\text{CD}$ 가 나노공극 내부에서 결합될 때 관찰되는 전류 수준을 나타낸다. D - 시스 챔버에 $5 \mu\text{M}$ dCMP를 첨가한 후의 전류 트레이스. L3은 dCMP가 ($\text{M113R}_7\text{/am}_7\text{-}\beta\text{CD}$) 착물에 결합할 때 관찰되는 전류 수준을 나타낸다.
- <23> 도 3은 dXMP 전류 진폭을 나타낸다. +130mV 포텐셜에서 인지질 이층에 삽입된 단일 (M113R_7) 공극의 전류 트레이스. $40 \mu\text{M}$ $\text{am}_7\text{-}\beta\text{CD}$ 가 트랜스 챔버에 존재한다. A - dGMP $5 \mu\text{M}$ 를 시스 챔버에 첨가한다. 전류 트레이스의 모든 점 히스토그램을 dGMP, dTMP(B), dAMP(C) 및 dCMP(D)와 함께 오른쪽에 나타낸다.
- <24> 도 4는 시클로덱스트린 전류 수준을 나타낸다. +130 mV에서 트랜스 챔버에 존재하는 $40 \mu\text{M}$ $\text{am}_7\text{-}\beta\text{CD}$ 와 함께 인지질 이층에 삽입된 단일 (M113R_7) 변이체의 전류 트레이스. L1 및 L1'는 비어있는 나노공극의 2개의 전류 수준을 나타내고, L2 및 L2'는 (M113R_7)에 $\text{am}_7\text{-}\beta\text{CD}$ 의 결합으로부터 얻어진 2개의 전류 수준을 나타낸다. 삽입은 전류 수준 L1, L1', L2 및 L2'에 상응하는 피크와 함께 전류 트레이스의 진폭 히스토그램을 나타낸다.
- <25> 도 5는 단일 사건 분석을 나타낸다. A는 동일한 용액 중에서 모든 4개 dXMP로부터의 L3 전류 수준의 단일 사건 분석 히스토그램을 나타낸다. B는 L2 단독으로부터 비롯된 L3의 단일 사건 분석 히스토그램을 나타낸다. $5 \mu\text{M}$ 의 dGMP, dAMP, dCMP 및 $10 \mu\text{M}$ 의 dTMP가 시스 챔버에 존재한다.
- <26> 도 6은 dXMP의 동시 검출을 나타낸다. A는 인지질 이층에 삽입된 단일 (M113R_7) 변이체의 전류 트레이스를 나타내고, +130 mV 포텐셜이 Ag/AgCl 전극 사이에 적용된다. 완충액은 1M KCl와 함께 트리스-HCl 25mM pH 8.0이다. $40 \mu\text{M}$ $\text{am}_7\text{-}\beta\text{CD}$ 가 트랜스 챔버에 존재한다. $5 \mu\text{M}$ 의 dGMP, dTMP, dAMP 및 dCMP가 시스 챔버에 첨가된다. 색을 가진 띠는 각각의 dXMP의 진폭 분포를 나타낸다. B는 8000개 결합 사건의 전류 트레이스로부터 모든

점 히스토그램을 나타낸다. 각각의 피크는 각각의 dXMP의 통계적 분포와 겹쳐진다.

<27> 도 7은 통계 방법을 나타낸다. 2개의 가우시안 분포 A 및 B가 교차점 I에서 중복된다. 교차점 I 너머의 가우시안 A의 면적을 적분하고 이것은 집단 B로서 식별되어지는 집단 A의 확률을 나타낸다.

서열 목록의 설명

<29> 서열 1은 α -헤모리신의 하나의 소단위를 코드화하는 폴리뉴클레오티드 서열을 나타낸다.

<30> 서열 2는 α -헤모리신의 하나의 소단위의 아미노산 서열을 나타낸다.

<31> 서열 3은 α -헤모리신 M113H의 1개 소단위를 코드화하는 폴리뉴클레오티드 서열을 나타낸다.

<32> 서열 4는 α -헤모리신 M113H의 1개 소단위의 아미노산 서열을 나타낸다.

<33> 서열 5는 α -헤모리신 M113K의 1개 소단위를 코드화하는 폴리뉴클레오티드 서열을 나타낸다.

<34> 서열 6은 α -헤모리신 M113K의 1개 소단위의 아미노산 서열을 나타낸다.

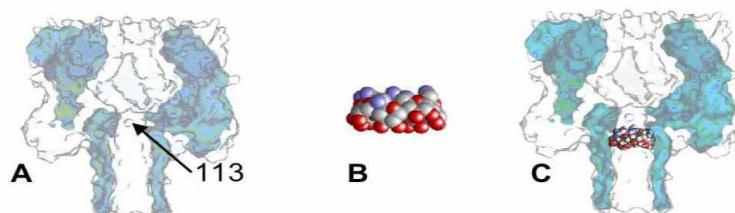
<35> 서열 7은 α -헤모리신 M113R의 1개 소단위를 코드화하는 폴리뉴클레오티드 서열을 나타낸다.

<36> 서열 8은 α -헤모리신 M113R의 1개 소단위의 아미노산 서열을 나타낸다.

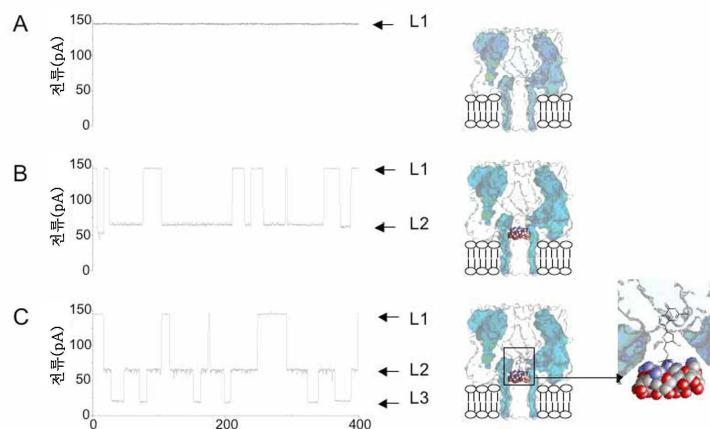
<37> 서열 9는 람다 엑소뉴클레아제의 아미노산 서열을 나타낸다. 서열은 삼량체로 조립되는 3개의 동일한 소단위 중의 하나이다.

도면

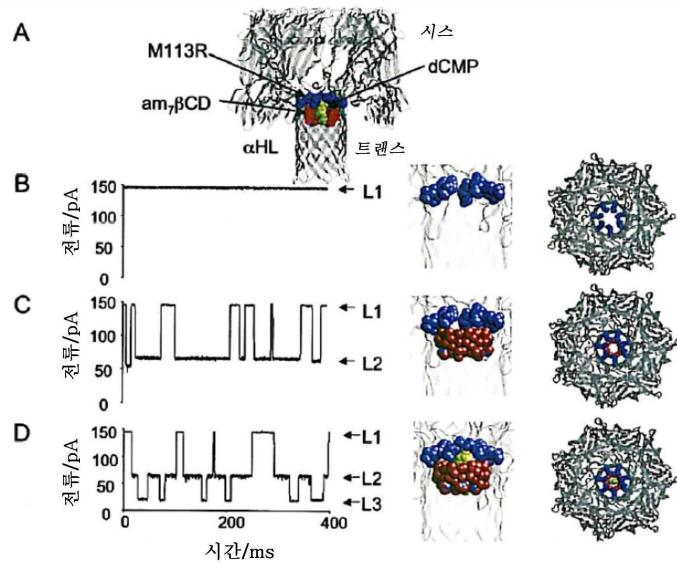
도면1



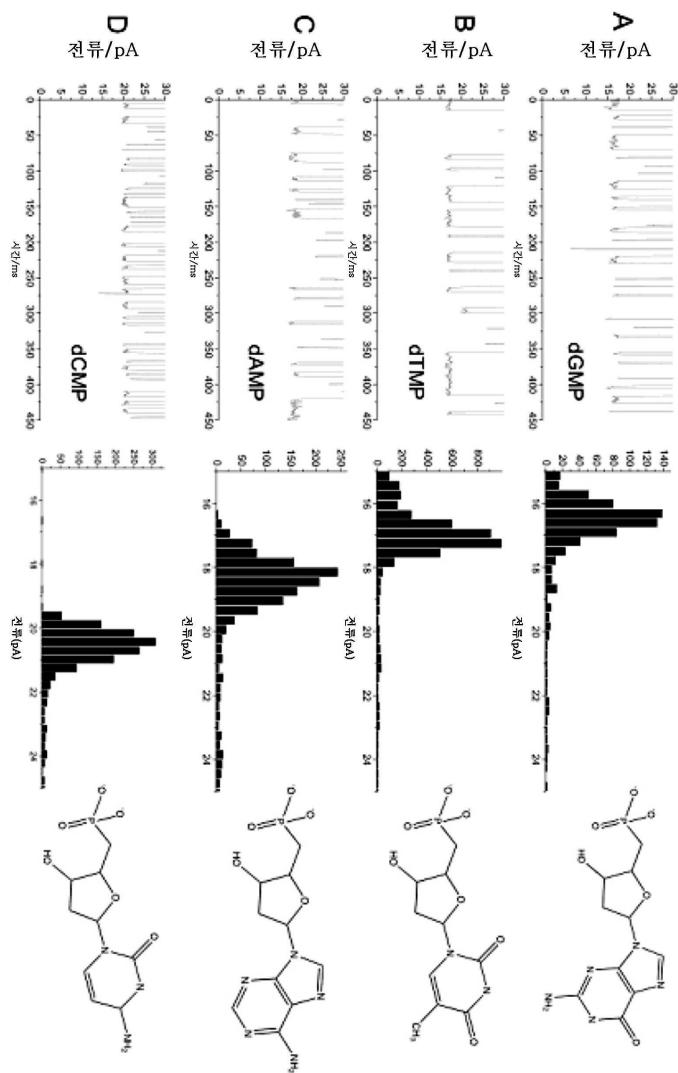
도면2A



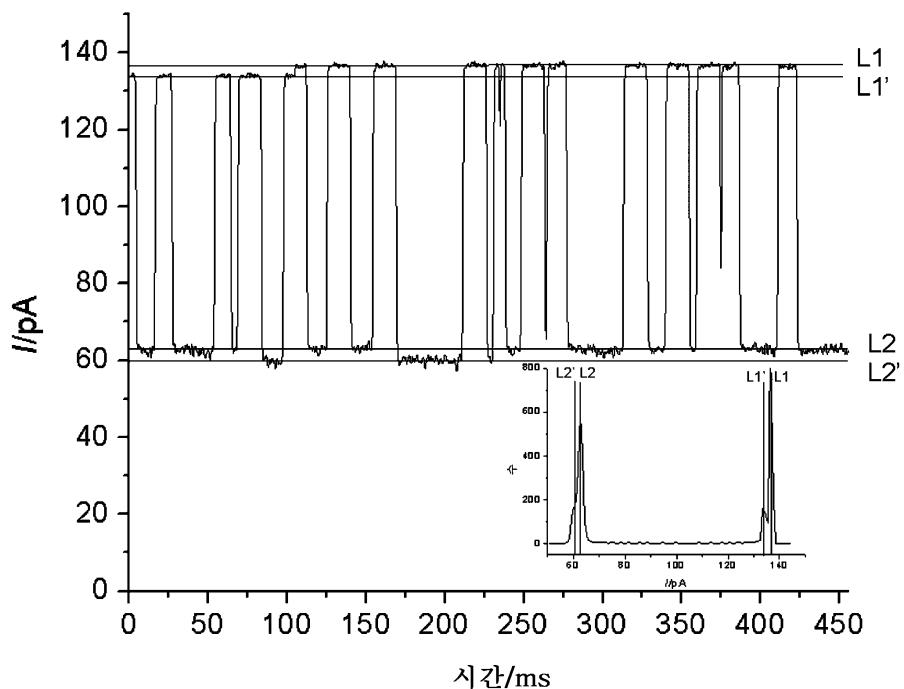
도면2B



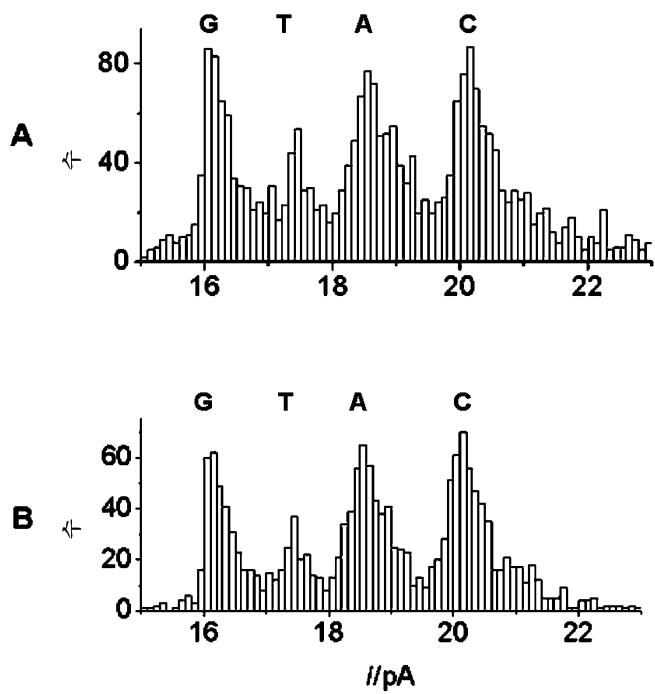
도면3



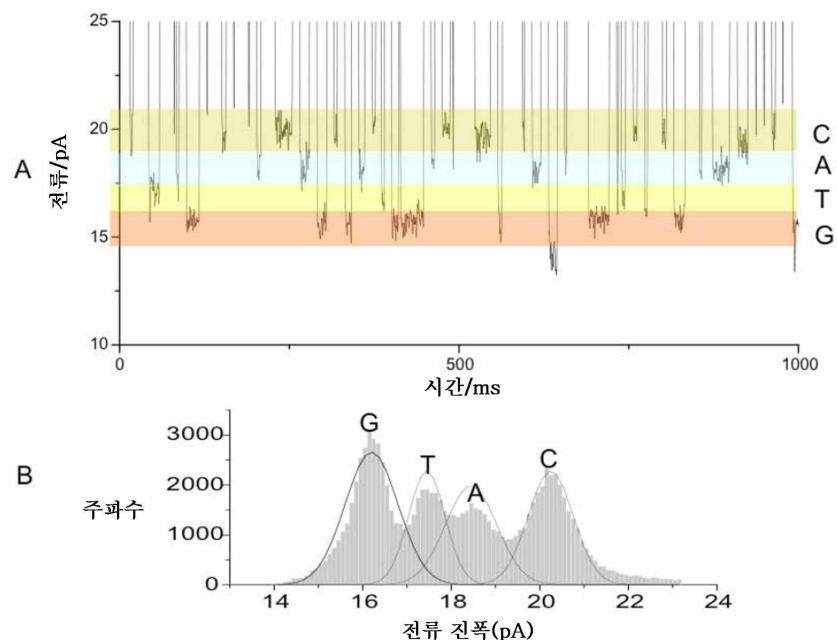
도면4



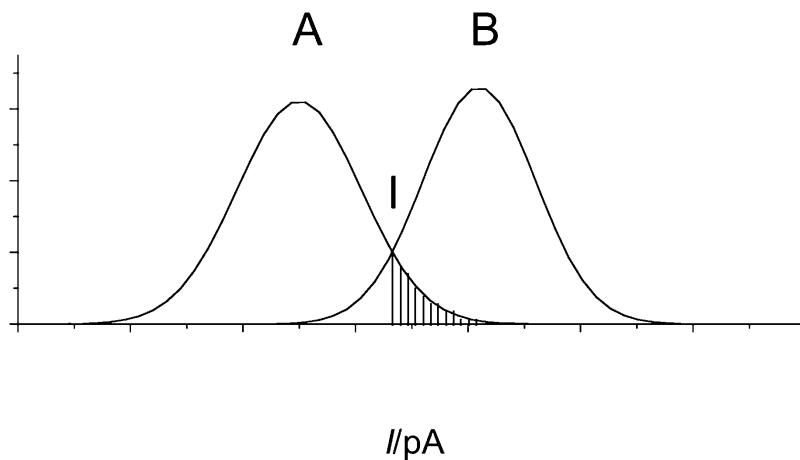
도면5



도면6



도면7



서 열 목 록

SEQUENCE LISTING

<110> ISIS INNOVATION LIMITED

<120> METHODS USING PORES

<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
<213> *Staphylococcus aureus*

<220>
<221> CDS
<222> (1)..(879)

<400> 1
gca gat tct gat att aat att aaa acc ggt act aca gat att gga agc 48
Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1 5 10 15

```

aat act aca gta aaa aca ggt gat tta gtc act tat gat aaa gaa aat 96
Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
          20           25           30

```

```

ggc atg cac aaa aaa gta ttt tat agt ttt atc gat gat aaa aat cac 144
Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His
          35           40           45

```

aat aaa aaa ctg cta gtt att aga acg aaa ggt acc att gct ggt caa 192
Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
50 55 60

tat aga gtt tat agc gaa gaa ggt gct aac aaa agt ggt tta gcc tgg 240
 Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
 65 70 75 80

```

cct tca gcc ttt aag gta cag ttg caa cta cct gat aat gaa gta gct      288
Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
          85           90           95

```

caa ata tct gat tac tat cca aga aat tcg att gat aca aaa gag tat 336
 Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
 100 105 110

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
 130 135 140 432

aca ctg aaa tat gtt caa cct gat ttc aaa aca att tta gag agc cca
 Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro
 145 150 155 160 480

act gat aaa aaa gta ggc tgg aaa gtg ata ttt aac aat atg gtg aat
 Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn
 165 170 175 528

caa aat tgg gga cca tat gat aga gat tct tgg aac ccg gta tat ggc
 Gln Asn Trp Gly Pro Tyr Asp Arg Ser Trp Asn Pro Val Tyr Gly
 180 185 190 576

aat caa ctt ttc atg aaa act aga aat ggt tct atg aaa gca gca gat
 Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp
 195 200 205 624

aac ttc ctt gat cct aac aaa gca agt tct cta tta tct tca ggg ttt
 Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe
 210 215 220 672

tca cca gac ttc gct aca gtt att act atg gat aga aaa gca tcc aaa
 Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys
 225 230 235 240 720

caa caa aca aat ata gat gta ata tac gaa cga gtt cgt gat gat tac
 Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr
 245 250 255 768

caa ttg cat tgg act tca aca aat tgg aaa ggt acc aat act aaa gat
 Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp
 260 265 270 816

aaa tgg aca gat cgt tct tca gaa aga tat aaa atc gat tgg gaa aaa
 Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys
 275 280 285 864

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			

<210> 3
 <211> 1300
 <212> DNA
 <213> Artificial sequence

<220>

<223> M113H alpha hemolysin mutant

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature
<222> (1169)..(1169)

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

<220>

<221> misc_feature

<222> (1171)..(1171)

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

<220>

<221> misc_feature

<222> (1175)..(1176)

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

<220>

<221> misc_feature

<222> (1180)..(1180)

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

<220>

<221> misc_feature

<222> (1182)..(1184)

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

<220>

<221> misc_feature

<222> (1188)..(1188)

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

<220>

<221> misc_feature

<222> (1191)..(1191)

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

<220>

<221> misc_feature

<222> (1193)..(1193)

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

<220>

<221> misc_feature

<222> (1196)..(1196)

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

<220>

<221> misc_feature

<222> (1201)..(1201)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature

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

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

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

tnn gcg acc ggt act aca gat att gga agc aat act aca gta aaa aca 104
 Xaa Ala Thr Gly Thr 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 gataat gaa gta gct caa ata tct gat tac tat 344
 Gln Leu Gln Leu Pro 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 Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp Thr Ser Lys Ile Gly Gly	120	125	130	440	
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 ataattata acttagccg aagctggatc 980
 cggtcgctac naanccnaa ngnagctgan ttgnctgctg cccccctgac natactagca 1040
 nacccttgg gnccctaacg ggtctgnggg gttttgctg aangngnact tttccgnan 1100
 tcnnccggn ccccccnngt gaaatccnaa ncccnaacn ggngntgnta ncaanttan 1160
 tggncntna nttnnaaan cnnttaantt ngnaanccc nttnncnan ggcnaannnn 1220
 nanccttna naaaaancc nnngggggg ttcnntnn annccnttn aangggccc 1280
 cnnggggnaa nnntnggggn 1300

<210> 4
 <211> 293
 <212> PRT
 <213> Artificial sequence

<220>
 <221> misc_feature
 <222> (5)..(5)
 <223> The 'Xaa' at location 5 stands for Leu, or Phe.

<220>
 <221> misc_feature
 <222> (7)..(7)
 <223> The 'Xaa' at location 7 stands for Tyr, Trp, Cys, Ser, Leu, or
 Phe.

<220>
 <223> M113H alpha hemolysin mutant

<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															
															45

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

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

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

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

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

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

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

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

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

Asn	Gln	Leu	Phe	Met	Lys	Thr	Arg	Asn	Gly	Ser	Met	Lys	Ala	Ala	Asp
195															
															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> Artificial sequence

<220>
 <223> M113K alpha hemolysin mutant

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

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

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

<220>
 <221> misc_feature

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature
<222> (1031)..(1031)

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

<220>

<221> misc_feature

<222> (1041)..(1041)

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

<220>

<221> misc_feature

<222> (1052)..(1052)

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

<220>

<221> misc_feature

<222> (1067)..(1067)

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

<220>

<221> misc_feature

<222> (1083)..(1083)

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

<220>

<221> misc_feature

<222> (1085)..(1085)

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

<220>

<221> misc_feature

<222> (1087)..(1087)

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

<220>

<221> misc_feature

<222> (1097)..(1098)

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

<220>

<221> misc_feature

<222> (1100)..(1100)

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

<220>

<221> misc_feature

<222> (1103)..(1104)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature
<222> (1214)..(1214)

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

<220>

<221> misc_feature

<222> (1217)..(1221)

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

<220>

<221> misc_feature

<222> (1223)..(1223)

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

<220>

<221> misc_feature

<222> (1229)..(1229)

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

<220>

<221> misc_feature

<222> (1231)..(1231)

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

<220>

<221> misc_feature

<222> (1238)..(1238)

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

<220>

<221> misc_feature

<222> (1241)..(1243)

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

<220>

<221> misc_feature

<222> (1255)..(1256)

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

<220>

<221> misc_feature

<222> (1258)..(1260)

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

<220>

<221> misc_feature

<222> (1262)..(1264)

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

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

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

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

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

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

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

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

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

<400> 5
gttctgttta actttaagaa gggagatata catatgag cag att ctg ata ttн 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 Xaa Ala Thr Gly Thr Thr Asp Ile Gly Ser Asn Thr Thr Val Lys Thr 10 15 20	104
ggt gat tta gtc act tat gat aaa gaa aat ggc atg cac aaa aaa gta Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn Gly Met His Lys Lys Val 25 30 35	152
ttt tat agt ttt atc gat gat aaa aat cac aat aaa aaa ctg cta gtt Phe Tyr Ser Phe Ile Asp Asp Lys Asn His Asn Lys Lys Leu Leu Val 40 45 50	200
att aga aca aaa ggt acc att gct ggt caa tat aga gtt tat agc gaa Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln Tyr Arg Val Tyr Ser Glu 55 60 65 70	248
gaa ggt gct aac aaa agt ggt tta gcc tgg cct tca gcc ttt aag gta Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp Pro Ser Ala Phe Lys Val 75 80 85	296
cag ttg caa cta cct gat aat gaa gta gct caa ata tct gat tac tat Gln Leu Gln Leu Pro Asp Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr 90 95 100	344
ccg cgg aat tcg att gat aca aaa gag tat aaa agt acg tta acg tac Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr Lys Ser Thr Leu Thr Tyr 105 110 115	392
gga ttc aac ggt aac ctt act ggt gat gat act agt aaa att gga ggc Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp Thr Ser Lys Ile Gly Gly 120 125 130	440
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	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
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
cggctgtac naanccnaa ngnagctgan ttgnctgctg ccccccgtac natactagca		1040
nacccttgg gnccctaact ggtctgnggg gttttgctg aangngnact tttccgnan		1100
tcnncccggn ccccccnngt gaaatccnaa nccccnaacn ggngntgnta ncaanttan		1160
tggnnctna nttnnaan cnnttaantt ngnaancccc nttnncnan ggcnaannnn		1220
nanccttna naaaaaancc nnngggggg ttcnntnnn annncnttn aangggcccc		1280
cnngggnaa nnntnggggn		1300

<211> 293
<212> PRT
<213> Artificial sequence

<220>
<221> misc_feature
<222> (5)..(5)
<223> The 'Xaa' at location 5 stands for Leu, or Phe.

<220>
<221> misc_feature
<222> (7)..(7)
<223> The 'Xaa' at location 7 stands for Tyr, Trp, Cys, Ser, Leu, or
Phe.

<220>
<223> M113K alpha hemolysin mutant

<400> 6

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
-----	-----	-----

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> Artificial sequence

<220>
<223> M113R alpha hemolysin mutant

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

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

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

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

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

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

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

<220>

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature
<222> (1083)..(1083)

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

<220>

<221> misc_feature

<222> (1085)..(1085)

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

<220>

<221> misc_feature

<222> (1087)..(1087)

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

<220>

<221> misc_feature

<222> (1097)..(1098)

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

<220>

<221> misc_feature

<222> (1100)..(1100)

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

<220>

<221> misc_feature

<222> (1103)..(1104)

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

<220>

<221> misc_feature

<222> (1110)..(1110)

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

<220>

<221> misc_feature

<222> (1117)..(1117)

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

<220>

<221> misc_feature

<222> (1128)..(1128)

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

<220>

<221> misc_feature

<222> (1131)..(1131)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature

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

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

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

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

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

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

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

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

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

<220>
<221> misc_feature
<222> (1231)..(1231)

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

<220>

<221> misc_feature

<222> (1238)..(1238)

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

<220>

<221> misc_feature

<222> (1241)..(1243)

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

<220>

<221> misc_feature

<222> (1255)..(1256)

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

<220>

<221> misc_feature

<222> (1258)..(1260)

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

<220>

<221> misc_feature

<222> (1262)..(1264)

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

<220>

<221> misc_feature

<222> (1267)..(1267)

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

<220>

<221> misc_feature

<222> (1270)..(1270)

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

<220>

<221> misc_feature

<222> (1273)..(1273)

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

<220>

<221> misc_feature

<222> (1282)..(1283)

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

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

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

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

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

tnn gcg acc ggt act aca gat att gga agc aat act aca gta aaa aca 104
 Xaa Ala Thr Gly Thr 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 Gln Leu Gln Leu Pro Asp Asn Glu Val Ala Gln Ile Ser Asp Tyr Tyr	90	95	100	344
ccg cgg aat tcg att gat aca aaa gag tat aga agt acg tta acg tac Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr Arg Ser Thr Leu Thr Tyr	105	110	115	392
gga ttc aac ggt aac ctt act ggt gat gat act agt aaa att gga ggc Gly Phe Asn Gly Asn Leu Thr Gly Asp Asp Thr Ser Lys Ile Gly Gly	120	125	130	440
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	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	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 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

tgttaanttat ttgtacatgt acaaataaat ataatttata acttagccg aagctggatc 980

cggctgctac naanccnaa ngnagctgan ttgnctgctg cccccctgac natactagca 1040

nacccttgg gncctaagc ggtctgnggg gttttgctg aangnact tttccgnan 1100

tcnncggc ccccccnggt gaaatccnaa nccccnaacn ggngntgnta ncaantttan 1160

tggnnncntna nt ttnaaan cnnttaantt ngnaancccc nt ttncnan ggcnaannnn 1220

nanccttna naaaaaancc nnnggggggg tttcnntnnn annnccnttn aangggcccc 1280

cnngggnaa nnntnggggn 1300

<210> 8
<211> 293
<212> PRT
<213> Artificial sequence

<220>
<221> misc_feature
<222> (5)..(5)
<223> The 'Xaa' at location 5 stands for Leu, or Phe.

<220>
<221> misc_feature
<222> (7)..(7)
<223> The 'Xaa' at location 7 stands for Tyr, Trp, Cys, Ser, Leu, or

Phe.

<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							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							160	

Thr	Asp	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> Artificial sequence

<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